

l'intègre

Sous la direction de
Christian Gautier
André Warusfel
Bruno Caminade
Gonzague de Monicault
Serge Nicolas

MATHÉMATIQUES

TOUT-EN-UN • ECS 2^e année

**NOUVELLE
ÉDITION**

- ▶ Un cours complet
- ▶ De nombreux exercices et problèmes
- ▶ Toutes les solutions détaillées en fin d'ouvrage

**PRÉPAS
COMMERCIALES**

DUNOD

Mathématiques

TOUT-EN-UN • ECS 2^e ANNÉE

Cours et exercices

Consultez nos parutions sur dunod.com

Dunod Éditeur, édition de livres, Microsoft Press, ETSF, Ediscience, InterEditions

http://www.dunod.com/

Recherche OK

Édiscience
ETSF
InterEditions
Microsoft Press

Donnez un nouveau souffle à votre vie...

Sciences et Techniques Informatique Gestion et Management Sciences Humaines

Accueil Contacts

Interviews

Reinventer les RH : urgence !
Giles Verrier

Ramses 2008 : exigez la nouvelle formule !
Thierry de Montbrial

toutes les Interviews
Club Enseignants
Inscrivez-vous!

Événements

Découvrez le vidéoBlog
Profession dirigeant

En librairie ce mois-ci

Développement personnel et coaching : découvrez le NOUVEAU SITE
intereditions.com !
les librairies

Sciences et Techniques Informatique Gestion et Management Sciences Humaines

- Nouveautés - Nouveautés - Nouveautés - Nouveautés -

Bacchus 2008
Enjeux, stratégies et pratiques dans la filière vitivinicole
Jean-Pierre Couderc, Hervé Hannin, François d'Hautleville, Etienne Montaigne

Profession dirigeant
De la conception du changement à l'action
Gérard Roth, Michal Kurtyka

PYTHON
Petit guide à l'usage du développeur agile
Tarek Ziadé

150 petites expériences de psychologie de sport
pour mieux comprendre les champions... et les autres
Yvan Paquet, Pascal Legrain, Elisabeth Rosnet, Stéphane Rusinek

Acheter Mon panier

LES BIBLIOTHÈQUES DES MÉTIERS

- Bibliothèque du DSI
- Gestion industrielle
- Métiers de la vigne et du vin
- Marketing et Communication
- Directeur d'établissement social et médico-social
- Toutes les bibliothèques

LES NEWSLETTERS

- Action sociale
- Psychologie
- Développement personnel et Bien-être
- Entreprise
- Expertise comptable
- Informatique et NTIC
- Industrie
- Toutes les newsletters

bibliothèques des métiers newsletters MicrosoftPress ediscience.net expert-sup.com

Notice légale

Mathématiques

Tout-en-un • ECS 2^e année

Cours et exercices corrigés

Sous la direction de

Christian Gautier et **André Warusfel**

Bruno Caminade

Professeur au lycée militaire de Saint-Cyr-l'École

Serge Nicolas

Professeur au lycée HENRI IV à Paris

Prépas commerciales

DUNOD

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

© Dunod, Paris, 2008
ISBN 978-2-10-053975-8

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements

d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.

Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du

droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).



Table des matières



Préface	vii
Chapitre 1 Compléments d'algèbre linéaire	1
1 Somme directe de sous-espaces, sous-espaces stables	1
2 Réduction des endomorphismes	7
3 Réduction d'une matrice	11
Chapitre 2 Algèbre bilinéaire	24
1 Produit scalaire	24
2 Espaces euclidiens	35
3 Endomorphismes symétriques	43
Chapitre 3 Intégration sur un intervalle quelconque	59
1 Définitions	59
2 Propriétés des intégrales convergentes	64
3 Cas des fonctions positives	71
4 Cas des fonctions de signe quelconque	77
Chapitre 4 Éléments de topologie de \mathbb{R}^n	88
1 Rappels sur \mathbb{R}^p	88
2 Distance euclidienne	91
3 Ouverts et fermés	94
4 Parties convexes	98
Chapitre 5 Fonctions de n variables – Continuité	104
1 Graphe d'une fonction	104

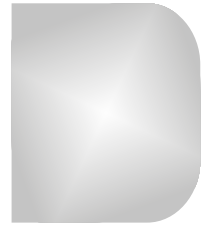
Table des matières

2	Continuité d'une fonction de \mathbb{R}^n dans \mathbb{R}	108
3	Opérations sur les fonctions continues	111
4	Propriétés des fonctions continues	116
Chapitre 6	Fonctions de n variables : calcul différentiel	125
1	Calcul différentiel d'ordre 1	125
2	Calcul différentiel d'ordre 2	135
Chapitre 7	Extremums	148
1	Extremums sur un ouvert	148
2	Extremums sous contrainte d'égalités linéaires	161
Chapitre 8	Variables aléatoires réelles discrètes	171
1	Généralités sur les variables aléatoires réelles	171
2	Séries doubles convergentes	179
3	Indépendance de variables aléatoires réelles discrètes	185
4	Espérance et conditionnement des variables discrètes	186
Chapitre 9	Vecteurs aléatoires discrets	202
1	Couples de variables aléatoires réelles discrètes	202
2	Variable aléatoire fonction d'un vecteur discret	211
3	Vecteurs aléatoires discrets à valeurs dans \mathbb{R}^n	221
Chapitre 10	Variables aléatoires réelles à densité	238
1	Définition des variables aléatoires réelles à densité	238
2	Moments d'une variable aléatoire à densité	246
3	Les lois usuelles	255
Chapitre 11	Convergences	284
1	Convergence en probabilité	285
2	Lois des grands nombres	295
3	Convergence en loi	298
4	Convergence en loi et approximations classiques	308
Chapitre 12	Estimation	318
1	Échantillons d'une loi de probabilité	319
2	Estimateurs	327
3	Suites d'estimateurs	333
4	Estimation par intervalles de confiance	337
5	Statistiques bivariées	347

Table des matières

Chapitre 13 Interventions informatiques	364
1 Récursivité	364
2 Gestion de listes à une dimension	380
3 Simulations de lois réelles discrètes	390
4 Simulations de lois réelles à densité	401
5 Estimation	412
Annexes Tables des lois usuelles	420
Solution des exercices	428
Chapitre 1	429
Chapitre 2	446
Chapitre 3	464
Chapitre 4	486
Chapitre 5	496
Chapitre 6	511
Chapitre 7	524
Chapitre 8	539
Chapitre 9	564
Chapitre 10	575
Chapitre 11	595
Chapitre 12	604
Index	620

Préface



Cet ouvrage est le deuxième de la série « Tout-en-un » consacré aux classes préparatoires au haut enseignement commercial. Il est destiné aux étudiants de seconde année de la filière scientifique.

Couvrant la totalité des résultats au programme, il contient tout ce qui est nécessaire pour la série « économique », au prix de certaines coupes évidentes ; cela dit, un autre volume couvrant les deux années de celle-ci est en cours de rédaction, et sera publié en juin 2006.

Rappelons dans quel esprit notre cours a été conçu et réalisé. Le rôle d'un professeur est, tout particulièrement en classes préparatoires, de construire son propre cours à partir de ses connaissances, de ses expériences et de ses lectures. Par suite ce livre n'est en aucun cas un modèle qui fournirait un cours prêt à l'emploi. En direction des enseignants, justement exigeants quant à leur liberté pédagogique, notre but a donc été humble : fournir à nos collègues quelques points de réflexion, quelques suggestions quant aux choix des propositions à invoquer et de leurs démonstrations.

S'il n'est pas un cours à l'usage des enseignants, il n'est pas davantage un texte dans lequel un élève découvrirait seul la partie mathématique des programmes de seconde année (à quelques exceptions près, dues à des cas d'isolement imparable). Il s'agit avant tout de donner aux étudiants un **ouvrage de référence**. Ce livre est à consulter de manière essentiellement ponctuelle, par exemple à la sortie d'un cours, pour trouver une vision autre permettant peut-être d'éclairer, par ses différences, l'exposé de parties plus délicates que d'autres, et aussi pour préparer une interrogation orale, la rédaction d'un devoir libre ou un contrôle. Il servira à vérifier, avec la plus grande précision possible, une définition, l'énoncé d'une proposition, une formule, et à se servir des nombreuses remarques mises au détour des points un peu subtils.

L'introduction de nouveaux concepts concernant le calcul des probabilités, introduits de façon plutôt abstraite, a ses avantages scientifiques évidents ; mais répétons qu'il serait stupide de penser que cela n'implique pas de réelles difficultés pédagogiques et techniques qu'on peut difficilement cacher sous le tapis.

Préface

Comme dans l'ouvrage de première année, a été préparée une copieuse liste d'exercices. Leurs énoncés sont classés par chapitre (à l'exception du dernier, qui n'en comporte pas). Comme précédemment, pour des raisons pédagogiques - ne pas laisser le lecteur devant la trop grande facilité de lire tout de suite une solution dès la première difficulté rencontrée -, les corrigés ont été regroupés sur le site de Dunod, où ils seront disponibles pour nos lecteurs dès la rentrée scolaire

Rappelons la technique très simple : une fois parvenu sur le site internet www.dunod.com, il suffit de cliquer successivement sur les items « sciences et techniques », « mathématiques », « classes préparatoires », l'icône de ce livre et enfin les « compléments en ligne ». Dès lors, le lecteur est prié d'entrer un mot de passe, à partir du livre qu'il a en main, et enfin de cliquer sur « corrigés ». Ceux-ci sont au format Pdf d'Adobe ; on peut s'y diriger par exemple à l'aide de la commande CTL-F pour retrouver un mot clé ou une expression (exemple : exercice 6.15) figurant avec certitude dans la solution cherchée. Toute partie sélectionnée à la souris peut être enregistrée séparément ou imprimée selon les techniques usuelles de Windows ou d'Apple.

Pour ce livre comme pour le précédent, nous redisons que nous serons toujours très intéressés par toutes les réactions des lecteurs de ce livre, étudiants ou enseignants : c'est par un travail commun, où les éléments extérieurs ont toute leur place, que des ouvrages comme celui-ci peuvent rapidement trouver une forme définitive répondant aux attentes légitimes fortes des étudiants et aux grandes ambitions des auteurs cherchant à mettre entre les mains de tous un outil scientifique et pédagogique de qualité.

Christian GAUTIER et André WARUSFEL

Compléments d'algèbre linéaire

1

Tous les espaces vectoriels mentionnés sont de dimension finie.

1. Somme directe de sous-espaces, sous-espaces stables

1.1 Somme directe de deux sous-espaces

Conformément à ce qu'exige le programme, commençons par quelques rappels de première année.

Définition 1

Soient F_1, F_2, \dots, F_n des sous-espaces vectoriels d'un espace E . On appelle somme des sous-espaces F_1, F_2, \dots, F_n l'ensemble des vecteurs de la forme

$$x_1 + x_2 + \dots + x_n$$

où $x_1 \in F_1, x_2 \in F_2, \dots, x_n \in F_n$.

Cet ensemble se note $F_1 + F_2 + \dots + F_n$ ou encore $\sum_{k=1}^n F_k$.

Définition 2

Soient F et G deux sous-espaces vectoriels d'un espace vectoriel E . La somme $F + G$ est dite directe si pour tout élément x de $F + G$, il existe un et un seul couple $(x_1, x_2) \in F \times G$ tel que $x = x_1 + x_2$.

Proposition 1

Soient F et G deux sous-espaces vectoriels d'un espace vectoriel E . La somme $F + G$ est **directe** si, et seulement si, $F \cap G = \{0\}$.

Définition 3

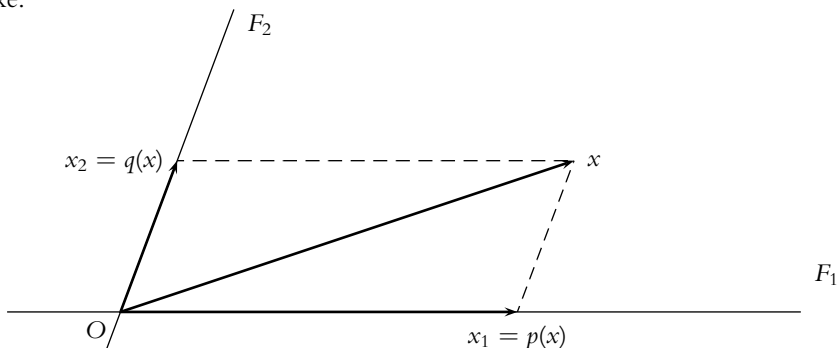
Soient F et G deux sous-espaces vectoriels d'un espace vectoriel E . F et G sont dits supplémentaires si $E = F \oplus G$.

Définition 4

Soient F_1, F_2 deux sous-espaces supplémentaires d’un espace vectoriel E . On appelle **projection** sur F_1 parallèlement à F_2 l’application p qui à tout vecteur x de E s’écrivant sous la forme $x = x_1 + x_2$ où $x_1 \in F_1$ et $x_2 \in F_2$ associe le vecteur $p(x) = x_1$.

Exemples

1. Si l’on note q la projection sur F_2 parallèlement à F_1 , on vérifie facilement que $p + q = \text{Id}_E$ et $p \circ q = q \circ p = 0$. Les projections p et q s’appellent des projections associées.
2. Traçons la figure dans le cas où F_1 et F_2 sont deux droites vectorielles du plan et x un vecteur fixé.



Proposition 2

Toute projection p est linéaire et vérifie $p \circ p = p$.

Preuve

Soit p une projection sur F_1 parallèlement à F_2 .

Montrons que p est linéaire. Soient x, y deux éléments de E avec

$$x = x_1 + x_2 \quad \text{et} \quad y = y_1 + y_2$$

où x_1, y_1 sont deux vecteurs de F_1 et x_2, y_2 deux vecteurs de F_2 . Soient λ, μ deux scalaires, on a alors

$$\begin{aligned} \lambda x + \mu y &= \lambda(x_1 + x_2) + \mu(y_1 + y_2) \\ &= (\lambda x_1 + \mu y_1) + (\lambda x_2 + \mu y_2). \end{aligned}$$

Comme F_1 et F_2 sont des sous-espaces, $\lambda x_1 + \mu y_1 \in F_1$ et $\lambda x_2 + \mu y_2 \in F_2$. On a donc

$$\begin{aligned} p(\lambda x + \mu y) &= \lambda x_1 + \mu y_1 \\ &= \lambda p(x) + \mu p(y) \end{aligned}$$

c’est-à-dire que p est linéaire.

Vérifions maintenant que $p \circ p = p$. Soit x un vecteur de E avec

$$x = \underbrace{x_1}_{\in F_1} + \underbrace{x_2}_{\in F_2}.$$

On a $p(x) = x_1$. Comme

$$x_1 = \underbrace{x_1}_{\in F_1} + \underbrace{0}_{\in F_2}$$

on a $(p \circ p)(x) = p(x_1) = x_1$. D'où $p \circ p = p$. □

Définition 5

On appelle **projecteur** toute application $p \in \mathcal{L}(E)$ telle que $p \circ p = p$.

Nous avons donc prouvé que toute projection était un projecteur. Prouvons maintenant la réciproque.

Proposition 3

Tout projecteur p réalise une projection sur $\Im m(p)$ parallèlement à $\text{Ker}(p)$.

Preuve

Montrons d'abord que $\Im m(p)$ et $\text{Ker}(p)$ sont supplémentaires. Nous allons prouver que pour tout vecteur x de E , il existe un unique couple $(x_1, x_2) \in \Im m(p) \times \text{Ker}(p)$ tel que $x = x_1 + x_2$.

Soit $x \in E$. Supposons avoir $x_1 \in \Im m(p)$ et $x_2 \in \text{Ker}(p)$ tels que $x = x_1 + x_2$. Fixons $y \in E$ tel que $x_1 = p(y)$. On a alors $x = p(y) + x_2$. En appliquant p on trouve

$$\begin{aligned} p(x) &= p(p(y)) + p(x_2) \\ &= (p \circ p)(y) + 0 \\ &= p(y) \\ &= x_1. \end{aligned}$$

On en déduit que nécessairement $x_1 = p(x)$ et $x_2 = x - p(x)$. Autrement dit l'écriture, si elle existe, est unique. Réciproquement pour tout $x \in E$, le vecteur $p(x) \in \Im m(p)$ et le vecteur $x - p(x) \in \text{Ker}(p)$ puisque $p(x - p(x)) = p(x) - p(p(x)) = p(x) - p(x) = 0$. La décomposition $x = p(x) + x - p(x)$ est donc une solution qui convient.

D'après ce qui précède, pour tout vecteur x de E , on a

$$x = \underbrace{p(x)}_{\in \Im m(p)} + \underbrace{x - p(x)}_{\in \text{Ker}(p)}$$

et $p(x)$ est bien la projection de x sur $\Im m(p)$ parallèlement à $\text{Ker}(p)$. □

En résumé nous pouvons affirmer

Proposition 4

Soit $p \in E^E$; p est un projecteur si, et seulement si, p est une projection.

1.2 Somme directe de n sous-espaces

Définition 6

Soient F_1, F_2, \dots, F_n des sous-espaces vectoriels d'un espace E et $F = \sum_{k=1}^n F_k$. On dit que F_1, F_2, \dots, F_n sont en somme directe si, et seulement si, pour tout vecteur $x \in F$, il existe une unique façon d'écrire

$$x = x_1 + x_2 + \dots + x_n$$

avec $x_1 \in F_1, x_2 \in F_2, \dots, x_n \in F_n$.

On note alors $F = F_1 \oplus F_2 \oplus \dots \oplus F_n$ ou encore $F = \bigoplus_{k=1}^n F_k$.

► Remarque

On note que si l'on a deux sous-espaces, on retrouve la définition de la somme directe de deux sous-espaces.

Proposition 5

Soient F_1, F_2, \dots, F_n des sous-espaces vectoriels d'un espace E , munis respectivement d'une base $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_n$, et $F = \sum_{k=1}^n F_k$.

Les propositions suivantes sont équivalentes.

- (i) F_1, F_2, \dots, F_n sont en somme directe.
- (ii) $\forall (x_1, x_2, \dots, x_n) \in F_1 \times F_2 \times \dots \times F_n,$

$$x_1 + x_2 + \dots + x_n = 0 \Rightarrow x_1 = x_2 = \dots = x_n = 0$$

- (iii) La famille obtenue en juxtaposant les vecteurs de $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_n$ est une base de F .

- (iv) $\dim F = \sum_{k=1}^n \dim(F_k)$.

Preuve

Notons $\mathcal{B}_k = (e_1^k, e_2^k, \dots, e_{n_k}^k)$ (où $k \in \llbracket 1, n \rrbracket$).

(i) \Rightarrow (ii) Soient $(x_1, x_2, \dots, x_n) \in F_1 \times F_2 \times \dots \times F_n$ tel que

$$x_1 + x_2 + \dots + x_n = 0$$

On remarque que l'on a également $0 + 0 + \dots + 0 = 0$ où $0 \in F_k$ pour $k \in \llbracket 1, n \rrbracket$. Comme F_1, F_2, \dots, F_n sont en somme directe, par unicité de l'écriture, il en résulte que, pour tout $k \in \llbracket 1, n \rrbracket$, $x_k = 0$.

(ii) \Rightarrow (iii) Il s'agit de prouver que la famille $\mathcal{B} = (e_j^k)_{\substack{1 \leq k \leq n \\ 1 \leq j \leq n_k}}$ est une base de F . Remarquons déjà que ces vecteurs sont bien des éléments de F .

Montrons que cette famille est génératrice. Soit $x \in F$, par définition de F , il existe $(x_1, x_2, \dots, x_n) \in F_1 \times F_2 \times \dots \times F_n$ tel que $x = x_1 + x_2 + \dots + x_n$. Or, pour $k \in \llbracket 1, n \rrbracket$, comme $x_k \in F_k$, il existe des scalaires $a_1^k, a_2^k, \dots, a_{n_k}^k$ tels que $x_k = \sum_{j=1}^{n_k} a_j^k e_j^k$.

On en déduit que $x = \sum_{k=1}^n \sum_{i=1}^{n_k} a_i^k e_i^k$. Donc x est bien combinaison linéaire des vecteurs de \mathcal{B} .

Montrons que cette famille est libre. Soient $(a_i^k)_{\substack{1 \leq k \leq n \\ 1 \leq i \leq n_k}}$ des scalaires tels que $\sum_{k=1}^n \sum_{i=1}^{n_k} a_i^k e_i^k = 0$. Posons,

pour $k \in \llbracket 1, n \rrbracket$, $x_k = \sum_{i=1}^{n_k} a_i^k e_i^k$, nous avons donc

$$\sum_{k=1}^n x_k = 0 \quad \text{avec } (x_1, x_2, \dots, x_n) \in F_1 \times F_2 \times \dots \times F_n.$$

D'après la propriété vérifiée par F_1, F_2, \dots, F_n , cela entraîne que pour $k \in \llbracket 1, n \rrbracket$, $x_k = 0$. Or pour $k \in \llbracket 1, n \rrbracket$, \mathcal{B}_k est une famille libre, donc, $\forall k \in \llbracket 1, n \rrbracket, \forall i \in \llbracket 1, n_k \rrbracket, a_i^k = 0$. Ce qui prouve que \mathcal{B} est une famille libre. La famille \mathcal{B} étant libre et génératrice de F , elle est une base de F .

(iii) \Rightarrow (i) Soit $x \in F$ tel que

$$x = x_1 + x_2 + \dots + x_n = x'_1 + x'_2 + \dots + x'_n$$

où $(x_1, \dots, x_n) \in F_1 \times \dots \times F_n$ et $(x'_1, \dots, x'_n) \in F_1 \times \dots \times F_n$. Montrons que pour $k \in \llbracket 1, n \rrbracket$, $x_k = x'_k$. Remarquons d'abord que

$$\sum_{k=1}^n x_k - x'_k = 0.$$

Soit $k \in \llbracket 1, n \rrbracket$, le vecteur $x_k - x'_k \in F_k$ (car F_k est un sous-espace vectoriel). Il existe donc des scalaires $a_i^k, a_2^k, \dots, a_{n_k}^k$ tels que $x_k - x'_k = \sum_{i=1}^{n_k} a_i^k e_i^k$. On a alors

$$\sum_{k=1}^n \sum_{i=1}^{n_k} a_i^k e_i^k = 0.$$

Comme la famille obtenue en juxtaposant les bases $\mathcal{B}_1, \dots, \mathcal{B}_n$ est une base de F , elle est en particulier libre. On en déduit que $\forall k \in \llbracket 1, n \rrbracket, \forall i \in \llbracket 1, n_k \rrbracket, a_i^k = 0$. Par conséquent, pour $k \in \llbracket 1, n \rrbracket$, $x_k - x'_k = 0$.

(iii) \Rightarrow (iv) C'est immédiat.

(iv) \Rightarrow (iii) Il est clair que la famille \mathcal{B} obtenue en juxtaposant les familles $\mathcal{B}_1, \dots, \mathcal{B}_n$ est génératrice de F . De plus, elle compte $\sum_{k=1}^n \dim(F_k)$ vecteurs, soit $\dim(F)$ vecteurs. C'est donc une base de F . □

1.3 Sous-espaces stables

Définition 7

Soit $u \in \mathcal{L}(E)$ et F un sous-espace vectoriel de E .
On dit que F est stable par u si pour tout $x \in F$, $u(x) \in F$.

Exemples

1. Pour tout endomorphisme u de E , $\text{Ker}(u)$ est stable par u . En effet, pour tout $x \in \text{Ker}(u)$, on a $u(u(x)) = u(0) = 0$ et donc $u(x) \in \text{Ker}(u)$.
2. Plus généralement, pour tout $u \in \mathcal{L}(E)$ et tout $\lambda \in \mathbb{K}$, $E_\lambda(u)$ est stable par u .
En effet, pour $x \in E_\lambda(u)$,

$$u(u(x)) = u(\lambda x) = \lambda u(x)$$

et donc $u(x) \in E_\lambda(u)$.

3. Si u et v sont deux endomorphismes de E tels que $uv = vu$, alors $\text{Ker}(u)$ et $\mathfrak{Im}(u)$ sont stables par v . En effet, soit $x \in \text{Ker}(u)$, alors $u(v(x)) = v(u(x)) = v(0) = 0$ et donc $v(x) \in \text{Ker}(u)$. De même soit $y \in \mathfrak{Im}(u)$, il existe $x \in E$ tel que $y = u(x)$. On a alors $v(y) = v(u(x)) = u(v(x))$ et donc $v(y) \in \mathfrak{Im}(u)$.

Soient F_1, F_2, \dots, F_n des sous-espaces de E tels que $E = \bigoplus_{k=1}^n F_k$. Pour $k \in \llbracket 1, n \rrbracket$, fixons $\mathcal{B}_k = (e_1^k, \dots, e_{n_k}^k)$ une base de F_k et notons \mathcal{B} la base de E obtenue en juxtaposant $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_n$. Soit u un endomorphisme de E qui laisse stable chacun de ces sous-espaces, la matrice A de u dans la base \mathcal{B} a la forme suivante :

$$A = \left(\begin{array}{c|c|c|c} A_1 & (0) & \dots & (0) \\ \hline (0) & A_2 & \ddots & \vdots \\ \hline \vdots & \ddots & \ddots & (0) \\ \hline (0) & \dots & (0) & A_n \end{array} \right)$$

où pour $k \in \llbracket 1, n \rrbracket$, la matrice A_k est de taille $n_k \times n_k$. On dit que A est diagonale par blocs.

Plus précisément, A_k est la matrice dans la base \mathcal{B}_k de l’endomorphisme de F_k induit par u .

Réciproquement, si la matrice d’un endomorphisme u de E dans la base \mathcal{B} est de cette forme, alors les sous-espaces vectoriels F_1, F_2, \dots, F_n sont stables par u .

Exemple

Soit s une symétrie d’un espace vectoriel E . On sait que

$$E = \text{Ker}(s - \text{Id}_E) \oplus \text{Ker}(s + \text{Id}_E) = E_1 \oplus E_{-1}.$$

De plus E_1, E_2 sont stables par s et par définition $s|_{E_1} = \text{Id}_{E_1}$ et $s|_{E_{-1}} = -\text{Id}_{E_{-1}}$. En posant $n = \dim(E)$ et $k = \dim(E_1)$, on en déduit qu’il existe une base de E dans laquelle la matrice de s s’écrit

$$\left(\begin{array}{c|c} I_k & (0) \\ \hline (0) & -I_{n-k} \end{array} \right)$$

2. Réduction des endomorphismes

2.1 Éléments propres

Commençons par quelques rappels

Définition 8

Soit $u \in \mathcal{L}(E)$.

- (i) Un scalaire $\lambda \in \mathbb{K}$ est dit valeur propre de u s'il existe $x \in E$ non nul tel que $u(x) = \lambda x$.
L'ensemble des valeurs propres de u s'appelle spectre de u et est noté $\text{Sp}(u)$.
- (ii) Un vecteur $x \in E$ est appelé vecteur propre de u si $x \neq 0$ et s'il existe $\lambda \in \mathbb{K}$ tel que $u(x) = \lambda x$.
Ce scalaire λ est unique et on dit alors que x est un vecteur propre de u associé à la valeur propre de λ .
- (iii) L'ensemble des valeurs propres de u s'appelle spectre de u et se note $\text{Sp}(u)$.

► Remarque

Prouvons l'unicité du scalaire. Soit x un vecteur propre de u tel que $u(x) = \lambda_1 x = \lambda_2 x$, alors $(\lambda_2 - \lambda_1)x = 0$. Comme $x \neq 0$, nécessairement $\lambda_1 - \lambda_2 = 0$, autrement dit $\lambda_1 = \lambda_2$.

Définition 9

Soient $u \in \mathcal{L}(E)$, on note E_λ l'ensemble suivant :

$$E_\lambda = \{x \in E \mid u(x) = \lambda x\}.$$

Si λ est une valeur propre de u , E_λ s'appelle le sous-espace propre de u associé à la valeur propre λ .

► Remarques

- En fait E_λ dépend également de u , en cas d'ambiguïté on notera $E_\lambda(u)$.
- Pour $\lambda \in \mathbb{K}$, E_λ est l'ensemble des vecteurs propres de u associés à la valeur propre λ auquel on ajoute le vecteur nul. En particulier, λ est valeur propre si, et seulement si, $E_\lambda \neq \{0\}$.
- On sait que $E_\lambda = \text{Ker}(u - \lambda \text{Id}_E)$, donc E_λ est un sous-espace vectoriel de E et

$$\begin{aligned} \lambda \text{ est une valeur propre} &\Leftrightarrow E_\lambda \neq \{0\} \Leftrightarrow u - \lambda \text{Id}_E \text{ n'est pas injectif} \\ &\Leftrightarrow u - \lambda \text{Id}_E \text{ n'est pas bijectif.} \end{aligned}$$

En particulier, 0 est valeur propre si, et seulement si, u n'est pas bijectif.

Proposition 6

Soient $u \in \mathcal{L}(E)$, $\lambda_1, \lambda_2, \dots, \lambda_k$ des valeurs propres de u distinctes deux à deux et x_1, x_2, \dots, x_k des vecteurs propres associés. La famille (x_1, x_2, \dots, x_k) est une famille libre.

Preuve

Prouvons-le par récurrence sur k .

C'est vrai pour $k = 1$ puisqu'un vecteur propre n'est pas nul.

Supposons la propriété vérifiée au rang k et montrons qu'elle est vraie au rang $k + 1$. Soit $(x_1, x_2, \dots, x_{k+1})$ des vecteurs propres de u associés à des valeurs propres distinctes $\lambda_1, \lambda_2, \dots, \lambda_{k+1}$. Supposons avoir des scalaires a_1, a_2, \dots, a_{k+1} tels que

$$\sum_{i=1}^{k+1} a_i x_i = 0. \tag{1}$$

En composant cette égalité par u , on obtient $\sum_{i=1}^{k+1} a_i u(x_i) = 0$ soit

$$\sum_{i=1}^{k+1} a_i \lambda_i x_i = 0. \tag{2}$$

En effectuant $\lambda_{k+1}(1) - (2)$, on trouve

$$\sum_{i=1}^k a_i (\lambda_{k+1} - \lambda_i) x_i = 0.$$

Mais d'après l'hypothèse de récurrence, la famille (x_1, \dots, x_k) est libre. On en déduit que $a_i (\lambda_{k+1} - \lambda_i) = 0$ pour $i \in \llbracket 1, k \rrbracket$. Comme les valeurs propres sont distinctes $\lambda_{k+1} - \lambda_i \neq 0$ et donc $a_i = 0$ pour $i \in \llbracket 1, k \rrbracket$. Il reste donc finalement $a_{k+1} x_{k+1} = 0$. Comme $x_{k+1} \neq 0$, puisque x_{k+1} est un vecteur propre, on a également $a_{k+1} = 0$. Cela prouve que la famille (x_1, \dots, x_{k+1}) est libre. La propriété est donc prouvée par récurrence. \square

Corollaire 1

Les sous-espaces propres d'un endomorphisme d'un espace de dimension finie sont en somme directe.

Preuve

Soit u un endomorphisme de E , si les sous-espaces propres de u n'étaient pas en somme directe, il existerait des vecteurs propres de u , x_1, x_2, \dots, x_k , associés à des valeurs propres distinctes, tels que $x_1 + x_2 + \dots + x_k = 0$. En particulier la famille (x_1, \dots, x_k) serait liée. Cela est en contradiction avec la proposition précédente. Les sous-espaces propres de u sont donc bien en somme directe. \square

2.2 Polynôme d'un endomorphisme

Définition 10

Soit $P \in \mathbb{K}[X]$, tel que $P = \sum_{k=0}^n a_k X^k$ et u un endomorphisme de E . On note $P(u)$ l'endomorphisme défini par

$$P(u) = \sum_{k=0}^n a_k u^k.$$

Proposition 7

Soient $P, Q \in \mathbb{K}[X]$, $\lambda, \mu \in \mathbb{K}$ et $u \in \mathcal{L}(E)$, on a les égalités

$$\begin{aligned}(\lambda P + \mu Q)(u) &= \lambda P(u) + \mu Q(u) \\ (PQ)(u) &= P(u) \circ Q(u).\end{aligned}$$

Preuve

C'est une conséquence de la structure d'algèbre de $\mathcal{L}(E)$. □

Définition 11

On dit qu'un polynôme P est un polynôme annulateur d'un endomorphisme u si $P(u) = 0$.

Exemples

1. Si u est une homothétie de rapport λ , autrement dit si $u = \lambda \text{Id}_E$, alors le polynôme $P(X) = X - \lambda$ est annulateur de u , puisque $P(u) = u - \lambda \text{Id}_E = 0_E$.
2. Le polynôme $Q(X) = X^2 - X$ est annulateur de toute projection. En effet, si p est un projecteur, alors $Q(p) = p^2 - p = p - p = 0$.
3. Le polynôme $R(X) = X^2 - 1$ est annulateur de toute symétrie. En effet, si s est une symétrie, alors $R(s) = s^2 - \text{Id}_E = \text{Id}_E - \text{Id}_E = 0$.

Proposition 8

Tout endomorphisme d'un espace de dimension finie admet un polynôme annulateur non nul.

Preuve

Soit $u \in \mathcal{L}(E)$. Notons $p = \dim(E)$. La famille $(\text{Id}_E, u, u^2, \dots, u^{p^2})$ est une famille de vecteurs de $\mathcal{L}(E)$ qui compte $p^2 + 1$ éléments. Comme $\dim(\mathcal{L}(E)) = p^2$, cette famille est liée, c'est-à-dire qu'il existe des scalaires a_0, a_1, \dots, a_{p^2} , non tous nuls, tels que $\sum_{k=0}^{p^2} a_k u^k = 0$. Si l'on pose $P(X) = \sum_{k=0}^{p^2} a_k X^k$, on a donc $P(u) = 0$, avec $P \neq 0$ puisque les a_k sont non tous nuls. □

Proposition 9

Soit $u \in \mathcal{L}(E)$ qui admet P comme polynôme annulateur. Si λ est une valeur propre de u , alors $P(\lambda) = 0$.

Preuve

Soit x un vecteur propre de u associé à λ . On a donc $u(x) = \lambda x$. Prouvons par récurrence que

$$\forall n \in \mathbb{N}, u^n(x) = \lambda^n x.$$

Pour $n = 0$, on a $u^0(x) = \text{Id}(x) = x$ qui est bien égal à $\lambda^0 x = x$.
Supposons que $u^n(x) = \lambda^n x$, alors

$$u^{n+1}(x) = u(u^n(x)) = u(\lambda^n x) = \lambda^n u(x) = \lambda^n \cdot \lambda x = \lambda^{n+1} x$$

ce qui achève la récurrence.

Notons $P = \sum_{k=0}^n a_k X^k$. On a

$$P(u)(x) = \left(\sum_{k=0}^n a_k u^k \right) (x) = \sum_{k=0}^n a_k u^k(x) = \sum_{k=0}^n a_k \lambda^k x = \left(\sum_{k=0}^n a_k \lambda^k \right) \cdot x = P(\lambda) \cdot x.$$

Mais par ailleurs, comme $P(u) = 0$, on a également $P(u)(x) = 0$. On en déduit que $P(\lambda) \cdot x = 0$. Or x étant un vecteur propre, $x \neq 0$, et donc $P(\lambda) = 0$. \square

Exemples

1. Les valeurs propres d’une projection sont donc racines du polynôme $X^2 - X$. Une projection n’admet donc au plus que deux valeurs propres à savoir 0 et 1.
2. De même les valeurs propres d’une symétrie sont racines du polynôme $X^2 - 1$. Autrement dit les valeurs propres possibles d’une symétrie sont 1 et -1 .
3. Soit Δ , l’endomorphisme de $\mathbb{R}_n[X]$, qui à tout polynôme P de $\mathbb{R}_n[X]$ associe P' . Le polynôme X^{n+1} est annulateur de Δ , puisque pour $P \in \mathbb{R}_n[X]$, on a $\Delta^{n+1}(P) = P^{(n+1)} = 0$ (car $\deg(P) \leq n$). Comme la seule racine de X^{n+1} est 0, 0 est la seule valeur propre possible de Δ . De plus 0 est bien valeur propre puisque $\Delta(1) = 0$.
Il en résulte que $\text{Sp}(\Delta) = \{0\}$.

Proposition 10

Tout endomorphisme d’un espace de dimension finie admet un nombre fini de valeurs propres.

Preuve

Cela a déjà été prouvé en première année. Donnons-en une autre démonstration.
Soit $u \in \mathcal{L}(E)$. On sait qu’il existe un polynôme P non nul annulateur de u . D’après la proposition précédente, toute valeur propre de u est une racine de P . Comme P a un nombre fini de racines, puisqu’il est non nul, nécessairement u admet un nombre fini de valeurs propres. \square

2.3 Endomorphismes diagonalisables

Rappelons la définition de la diagonalisabilité :

Définition 12

Un endomorphisme $u \in \mathcal{L}(E)$ est dit diagonalisable s’il existe une base de E formée de vecteurs propres de u , c’est-à-dire s’il existe une base (e_1, e_2, \dots, e_n) de E telle que pour tout $i \in \llbracket 1, n \rrbracket$, e_i soit un vecteur propre de u .

Proposition 11

Soit $u \in \mathcal{L}(E)$ de valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_k$. Les assertions suivantes sont équivalentes

- (i) u est diagonalisable ;
- (ii) $\dim(E) = \sum_{i=1}^k \dim(E_{\lambda_i})$;
- (iii) E est somme directe des sous-espaces propres de u .

Preuve

(i) \Rightarrow (iii) Si u est diagonalisable, il existe une base de vecteurs propres de E . Tout vecteur de E est donc combinaison linéaire de vecteurs propres. On en déduit que $E = \sum_{i=1}^k E_{\lambda_i}$. Or d'après la proposition 1, cette somme est directe, donc E est bien somme directe des sous-espaces propres de E .

(iii) \Rightarrow (i) Fixons pour $i \in \llbracket 1, k \rrbracket$, \mathcal{B}_i une base de E_{λ_i} . \mathcal{B}_i est constitué de vecteurs propres de u . On sait que la famille \mathcal{B} obtenue en juxtaposant $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_k$ est une base de E . Or cette base est formée de vecteurs propres de u . Cet endomorphisme est diagonalisable.

(ii) \Rightarrow (iii) Posons $F = \bigoplus_{i=1}^k E_{\lambda_i}$. Comme les sous-espaces propres sont en somme directe,

$$\dim(F) = \sum_{i=1}^k \dim(E_{\lambda_i}).$$

D'après l'hypothèse on a donc $\dim(F) = \dim(E)$, et par conséquent $F = E$. On

$$\text{obtient finalement } E = \bigoplus_{i=1}^k E_{\lambda_i}.$$

(iii) \Rightarrow (ii) Comme $E = \bigoplus_{i=1}^k E_{\lambda_i}$, on a $\dim(E) = \sum_{i=1}^k \dim(E_{\lambda_i})$. □

3. Réduction d'une matrice

3.1 Similitude

Rappelons quelques résultats autour de la similitude.

Définition 13

Deux matrices A et B de $\mathcal{M}_n(\mathbb{K})$ sont dites semblables s'il existe une matrice inversible $P \in \mathcal{M}_n(\mathbb{K})$ telle que $B = P^{-1}AP$.

► Remarque

Pour toutes matrices $A, B \in \mathcal{M}_n(\mathbb{K})$ et $P \in \mathcal{GL}_n(\mathbb{K})$, on a

1. $B = P^{-1}AP \Leftrightarrow A = PBP^{-1}$.
2. $\forall n \in \mathbb{N} \quad B^n = P^{-1}A^nP$ (formule valable pour $n \in \mathbb{Z}$ si A et B sont inversibles).

Proposition 12

Deux matrices sont semblables si, et seulement si, elle représentent le même endomorphisme dans des bases éventuellement différentes.

3.2 Éléments propres

Définition 14

Soit $A \in \mathcal{M}_n(\mathbb{K})$. Un scalaire $\lambda \in \mathbb{K}$ est dit valeur propre de A s’il existe un vecteur colonne $X \in \mathcal{M}_{n,1}(\mathbb{K})$ non nul telle que $AX = \lambda X$. Un tel vecteur colonne est alors appelé vecteur propre de A , associé à la valeur propre λ . L’ensemble des valeurs propres de A s’appelle le spectre de A et est noté $\text{Sp}(A)$.

Exemple

Soit $A = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix} \in \mathcal{M}_3(\mathbb{R})$.

On sait que pour rechercher les valeurs propres de A , on triangule la matrice $A - \lambda I_3$ (où $\lambda \in \mathbb{R}$) :

$$\begin{aligned} \begin{pmatrix} 3-\lambda & 1 & 1 \\ 1 & 3-\lambda & 1 \\ 1 & 1 & 3-\lambda \end{pmatrix} &\rightarrow \begin{pmatrix} 1 & 1 & 3-\lambda \\ 1 & 3-\lambda & 1 \\ 3-\lambda & 1 & 1 \end{pmatrix} \quad L_1 \leftrightarrow L_3 \\ &\rightarrow \begin{pmatrix} 1 & 1 & 3-\lambda \\ 0 & 2-\lambda & \lambda-2 \\ 0 & \lambda-2 & -\lambda^2+6\lambda-8 \end{pmatrix} \quad \begin{array}{l} L_2 \leftarrow L_2 - L_1 \\ L_3 \leftarrow L_3 + (\lambda-3)L_1 \end{array} \\ &\rightarrow A(\lambda) = \begin{pmatrix} 1 & 1 & 3-\lambda \\ 0 & 2-\lambda & \lambda-2 \\ 0 & 0 & -\lambda^2+7\lambda-10 \end{pmatrix} \quad L_3 \leftarrow L_3 + L_2 \end{aligned}$$

Les valeurs propres sont donc les solutions de $(2-\lambda)(-\lambda^2+7\lambda-10) = 0$. 2 est racine évidente de $-\lambda^2+7\lambda-10$, donc l’autre racine est $7-2 = 5$. On en déduit que $\text{Sp}(A) = \{2, 5\}$.

De plus, en remarquant que

$$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \\ 0 \end{pmatrix},$$

on peut affirmer que le vecteur $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ est un vecteur propre de A associé à la valeur propre

5 et que le vecteur $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ est un vecteur propre de A associé à la valeur propre 2.

Définition 15

Pour $A \in \mathcal{M}_n(\mathbb{K})$ et $\lambda \in \mathbb{K}$, on note

$$E_\lambda = \{X \in \mathcal{M}_{n,1}(\mathbb{K}) \mid AX = \lambda X\}.$$

Si λ est une valeur propre de A , E_λ s'appelle le sous-espace propre de A associé à la valeur propre λ .

► **Remarques**

- On remarquera que l'équation $AX = \lambda X$ équivaut $(A - \lambda I_n)X = 0$.
- L'ensemble E_λ dépend de A . En cas d'ambiguïté, on notera cet ensemble $E_\lambda(A)$.
- On prouve facilement que E_λ est un sous-espace vectoriel de $\mathcal{M}_{n,1}(\mathbb{K})$. De même $E_\lambda \neq \{0_E\}$ si et seulement si λ est une valeur propre et dans ce cas E_λ est l'ensemble des vecteurs propres de A associés à la valeur propre λ auquel on rajoute le vecteur nul.

Exemple

Reprenons la matrice $A = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$. Déterminons une base de E_2 et E_5 .

Soit $X \in \mathcal{M}_{3,1}(\mathbb{R})$ avec $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, on a

$$\begin{aligned} X \in E_2 &\iff (A - 2I_3)X = 0 \iff \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0 \\ &\iff x + y + z = 0 \iff x = -y - z. \end{aligned}$$

On en déduit que

$$\begin{aligned} E_2 &= \left\{ \begin{pmatrix} -y - z \\ y \\ z \end{pmatrix} \mid y, z \in \mathbb{R} \right\} \\ &= \left\{ y \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \mid y, z \in \mathbb{R} \right\} \\ &= \text{Vect} \left(\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right). \end{aligned}$$

Les deux vecteurs étant non colinéaires, $\left(\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right)$ est une base de E_2 .

De même

$$\begin{aligned} X \in E_5 &\iff (A - 5I_3)X = 0 \iff \begin{cases} x + y - 2z = 0 \\ -3y + 3z = 0 \end{cases} \\ &\iff \begin{cases} x = z \\ y = z \end{cases} \end{aligned}$$

et donc $E_5 = \text{Vect} \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right)$. Ce vecteur étant non nul, $\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right)$ est une base de E_5 .

Proposition 13

Soient $A \in \mathcal{M}_n(\mathbb{K})$, E un espace de dimension n muni d’une base $\mathcal{B} = (e_1, \dots, e_n)$ et u un endomorphisme de E de matrice A dans la base \mathcal{B} . Soit φ l’application

$$\begin{cases} \mathcal{M}_{n,1}(\mathbb{K}) & \rightarrow & E \\ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} & \mapsto & \sum_{k=1}^n x_k e_k \end{cases}$$

φ réalise un isomorphisme de $\mathcal{M}_{n,1}(\mathbb{K})$ sur E qui induit une bijection de $E_\lambda(A)$ sur $E_\lambda(u)$ pour tout $\lambda \in \mathbb{K}$.

Preuve

On prouve sans difficulté que φ est une application linéaire bijective.

De plus, pour $\lambda \in \mathbb{K}$ et $X \in \mathcal{M}_{n,1}(\mathbb{K})$,

$$\begin{aligned} X \in E_\lambda(A) &\Leftrightarrow AX = \lambda X \Leftrightarrow u(\varphi(X)) = \lambda\varphi(X) \\ &\Leftrightarrow \varphi(X) \in E_\lambda(u). \end{aligned}$$

□

► **Remarques**

- On déduit de ce théorème que les valeurs propres de A sont les valeurs propres de u et pour tout $\lambda \in \mathbb{K}$, $\dim(E_\lambda(u)) = \dim(E_\lambda(A))$.
- Cette proposition permet de donner une version « matricielle » de tous les résultats donnés sur les endomorphismes.
On peut par exemple affirmer que les sous-espaces propres d’une matrice sont en somme directe.

Proposition 14

Soit $A \in \mathcal{M}_n(\mathbb{K})$, un scalaire λ est une valeur propre de A si, et seulement si, $A - \lambda I_n$ n’est pas inversible.

Preuve

Soit u l’endomorphisme de \mathbb{K}^n canoniquement associé à A , on a alors

$$\begin{aligned} \lambda \text{ est valeur propre de } A &\Leftrightarrow \lambda \text{ est valeur propre de } u \\ &\Leftrightarrow (u - \lambda \text{Id}_E) \text{ n'est pas bijectif} \\ &\Leftrightarrow (A - \lambda I_n) \text{ n'est pas inversible.} \end{aligned}$$

□

► **Remarque**

En particulier 0 est valeur propre de A si et seulement si A n’est pas inversible.

Corollaire 2

Les valeurs propres d’une matrice triangulaire sont ses coefficients diagonaux.

Preuve

Soit $A \in \mathcal{M}_n(\mathbb{K})$ une matrice triangulaire supérieure, avec

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ 0 & a_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1,n} \\ 0 & \dots & 0 & a_{n,n} \end{pmatrix}.$$

On a alors

$$A - \lambda I_n = \begin{pmatrix} a_{1,1} - \lambda & a_{1,2} & \dots & a_{1,n} \\ 0 & a_{2,2} - \lambda & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1,n} \\ 0 & \dots & 0 & a_{n,n} - \lambda \end{pmatrix}.$$

Cette matrice est non inversible si, et seulement si, l'un de ses coefficients diagonaux est nul, c'est-à-dire si, et seulement si il existe un entier $i \in \llbracket 1, n \rrbracket$ tel que $\lambda = a_{i,i}$. Les valeurs propres de A sont donc les scalaires $a_{1,1}, a_{2,2}, \dots, a_{n,n}$.

La démonstration est analogue pour une matrice triangulaire inférieure. □

3.3 Polynôme d'une matrice**Définition 16**

Soient $P \in \mathbb{K}[X]$, tel que $P = \sum_{k=0}^n a_k X^k$ et A une matrice de $\mathcal{M}_n(\mathbb{K})$. On note $P(A)$ la matrice définie par

$$P(A) = \sum_{k=0}^n a_k A^k.$$

Proposition 15

Soient $P, Q \in \mathbb{K}[X]$, $\lambda, \mu \in \mathbb{K}$ et $A \in \mathcal{M}_n(\mathbb{K})$, on a les égalités

$$\begin{aligned} (\lambda P + \mu Q)(A) &= \lambda P(A) + \mu Q(A) \\ (PQ)(A) &= P(A) \times Q(A). \end{aligned}$$

Preuve

C'est une conséquence de la structure d'algèbre de $\mathcal{M}_n(\mathbb{K})$. □

Proposition 16

Soient $P \in \mathbb{K}[X]$, u un endomorphisme de E muni d'une base \mathcal{B} et A la matrice de u dans la base \mathcal{B} . On a $\mathcal{M}_{\mathcal{B}}(P(u)) = P(A)$.

Preuve

D'après les propriétés connues sur les matrices, pour tout $k \in \mathbb{N}$, on a :

$\mathcal{M}_{\mathcal{B}}(u^k) = \mathcal{M}_{\mathcal{B}}(u)^k = A^k$. Si l'on note $P(X) = \sum_{k=0}^n a_k X^k$, il s'ensuit que

$$\mathcal{M}_{\mathcal{B}}(P(u)) = \mathcal{M}_{\mathcal{B}}\left(\sum_{k=0}^n a_k u^k\right) = \sum_{k=0}^n \mathcal{M}_{\mathcal{B}}(u^k) = \sum_{k=0}^n a_k A^k = P(A). \quad \square$$

Définition 17

On dit qu'un polynôme P est un polynôme annulateur d'une matrice A si $P(A) = 0$.

Exemples

1. Un polynôme annulateur permet de déterminer la puissance $n^{\text{ème}}$ d'une matrice.

Considérons la matrice $A = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}$. On a $A^2 = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix}$. On remarque que $A^2 - 8A = -15I$. Le polynôme $P(X) = X^2 - 8X + 15$ est donc annulateur de A . Pour calculer A^n , on va déterminer le reste de la division de X^n par P . Pour tout $n \in \mathbb{N}$, on sait qu'il existe deux réels a_n, b_n et un polynôme Q_n tels que $X^n = P(X)Q_n(X) + a_n X + b_n$. Comme 3 et 5 sont les racines de P , on a

$$\begin{cases} 3a_n + b_n = 3^n \\ 5a_n + b_n = 5^n \end{cases}.$$

En résolvant ce système, on trouve que $a_n = \frac{5^n - 3^n}{2}$ et $b_n = \frac{5 \cdot 3^n - 3 \cdot 5^n}{2}$.

De plus, en évaluant l'égalité $X^n = P(X)Q_n(X) + a_n X + b_n$ en A , on trouve que $A^n = P(A)Q_n(A) + a_n A + b_n I = a_n A + b_n I$ (puisque $P(A) = 0$).

On en déduit que, pour $n \in \mathbb{N}$,

$$\begin{aligned} A^n &= \frac{5^n - 3^n}{2} A + \frac{5 \cdot 3^n - 3 \cdot 5^n}{2} I \\ &= \frac{1}{2} \begin{pmatrix} 3^n + 5^n & 5^n - 3^n \\ 5^n - 3^n & 3^n + 5^n \end{pmatrix}. \end{aligned}$$

Notons également que le polynôme annulateur de A permet de prouver que A est inversible et d'exprimer A^{-1} en fonction de A . En effet, on peut écrire que $A \times \frac{1}{15}(-A + 8I) = I$, donc A est inversible et $A^{-1} = \frac{1}{15}(-A + 8I)$.

Plus généralement si un polynôme annulateur d'une matrice a un coefficient constant non nul, alors la matrice est inversible et l'on peut exprimer son inverse comme un polynôme en la matrice.

2. Soit $B = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. On a

$$B^2 = \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix} \quad \text{et} \quad B^4 = (B^2)^2 = \begin{pmatrix} -4 & 0 \\ 0 & -4 \end{pmatrix}.$$

On en déduit que le polynôme $P(X) = X^4 + 4$ est un polynôme annulateur de B .

Proposition 17

Si $P \in \mathbb{K}[X]$ est un polynôme annulateur d'une matrice $A \in \mathcal{M}_n(\mathbb{K})$, les valeurs propres de A sont racines de P .

Preuve

Soit u l'endomorphisme de \mathbb{R}^n canoniquement associé à A et λ une valeur propre de A . On sait que λ est une valeur propre de u . Mais comme $P(A) = 0$, on a $P(u) = 0$, et donc λ est une racine de P . \square

► Remarque

On aurait pu également prouver que si $X \in E_\lambda(A)$, alors pour $n \in \mathbb{N}$, $A^n X = \lambda^n X$ et en déduire que $P(A)X = P(\lambda)X$.

Exemple

En reprenant l'exemple précédent, on en déduit que les valeurs propres de la matrice $B = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ sont racines de $X^4 + 4$. La matrice B n'a donc pas de valeur propre réelle. Sur \mathbb{C} , on a

$$\text{Sp}(B) \subset \left\{ \sqrt{2}e^{\frac{i\pi}{4}}, \sqrt{2}e^{\frac{3i\pi}{4}}, \sqrt{2}e^{\frac{5i\pi}{4}}, \sqrt{2}e^{\frac{7i\pi}{4}} \right\}.$$

Comme on sait que B possède au plus deux valeurs propres, on vérifie sur cet exemple que toutes les racines d'un polynôme annulateur de B ne sont pas nécessairement des valeurs propres de B .

3.4 Matrices diagonalisables**Définition 18**

Une matrice $A \in \mathcal{M}_n(\mathbb{K})$ est dite **diagonalisable** s'il existe une matrice $P \in \mathcal{GL}_n(\mathbb{K})$ telle que $P^{-1}AP$ soit une matrice diagonale.

► Remarques

- Autrement dit une matrice est diagonalisable si elle est semblable à une matrice diagonale.
- Une matrice de $\mathcal{M}_n(\mathbb{R})$ peut être diagonalisable si elle est considérée comme une matrice de $\mathcal{M}_n(\mathbb{C})$ sans l'être en tant que matrice de $\mathcal{M}_n(\mathbb{R})$. En cas d'ambiguïté, on parlera de matrice diagonalisable sur \mathbb{R} ou sur \mathbb{C} .

Proposition 18

Soit $u \in \mathcal{L}(E)$, \mathcal{B}_1 une base de E et $A = \mathcal{M}_{\mathcal{B}_1}(u)$.

1. Soit \mathcal{B}_2 une base de vecteurs propres de u associés aux valeurs propres $\lambda_1, \dots, \lambda_n$. En notant P la matrice de passage de \mathcal{B}_1 à \mathcal{B}_2 , on a $P^{-1}AP = \text{Diag}(\lambda_1, \dots, \lambda_n)$.
2. Soit P une matrice de $\mathcal{GL}_n(\mathbb{K})$ telle que $P^{-1}AP = \text{Diag}(\lambda_1, \dots, \lambda_n)$. On note \mathcal{B}_2 la base de E telle que P soit la matrice de passage de \mathcal{B}_1 à \mathcal{B}_2 . Cette base \mathcal{B}_2 est une base de vecteurs propres de u associés aux valeurs propres $\lambda_1, \dots, \lambda_n$.

En particulier u est diagonalisable si, et seulement si, A est diagonalisable.

Preuve

1. Soit \mathcal{B}_2 une telle base. On sait que $\mathcal{M}_{\mathcal{B}_2}(u) = \text{Diag}(\lambda_1, \dots, \lambda_n)$. Mais on sait aussi que si l'on note P la matrice de passage de \mathcal{B}_1 à \mathcal{B}_2 , on a $\mathcal{M}_{\mathcal{B}_2}(u) = P^{-1}AP$. D'où le résultat.

2. Soit P une telle matrice et \mathcal{B}_2 la base de E telle que P soit la matrice de passage de \mathcal{B}_1 à \mathcal{B}_2 . On sait alors que $\mathcal{M}_{\mathcal{B}_2}(u) = P^{-1}AP$. On a donc $\mathcal{M}_{\mathcal{B}_2}(u) = \text{Diag}(\lambda_1, \dots, \lambda_n)$, autrement dit \mathcal{B}_2 est une base de E formée de vecteurs propres de u associés aux valeurs propres $\lambda_1, \dots, \lambda_n$. \square

Théorème 1

Soit A une matrice de $\mathcal{M}_n(\mathbb{K})$, les propositions suivantes sont équivalentes :

- (i) A est diagonalisable ;
- (ii) $\mathcal{M}_{n,1}(\mathbb{K})$ est somme directe des sous-espaces propres de A ;
- (iii) La somme des dimensions des sous-espaces propres de A est égale à n .

Preuve

C’est une conséquence de la proposition 13. \square

Exemples

1. La matrice $A = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$ admet deux valeurs propres 2 et 5.

Comme $\dim(E_2 \oplus E_5) = 3 = \dim(\mathcal{M}_{3,1}(\mathbb{R}))$, on a $\mathcal{M}_{3,1}(\mathbb{R}) = E_2 \oplus E_5$ et donc A est diagonalisable.

2. La matrice

$$B = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

n’admet pas de valeur propre réelle, donc B n’est pas diagonalisable sur \mathbb{R} . Déterminons les valeurs propres de B sur \mathbb{C} .

La matrice $\begin{pmatrix} 1 - \lambda & -1 \\ 1 & 1 - \lambda \end{pmatrix}$ est non inversible si, et seulement si, $(1 - \lambda)^2 + 1 = 0$, soit encore $(1 - \lambda - i)(1 - \lambda + i) = 0$. Donc $\text{Sp}(B) = \{1 - i, 1 + i\}$. Comme B possède 2 valeurs propres, B est bien diagonalisable sur \mathbb{C} .

1. Déterminer les éléments propres de la matrice

$$A = \begin{pmatrix} 1 & 1 & \dots & \dots & \dots & 1 & 1 \\ 1 & 0 & \dots & \dots & \dots & 0 & 1 \\ \vdots & \vdots & & & & \vdots & \vdots \\ \vdots & \vdots & & (0) & & \vdots & \vdots \\ \vdots & \vdots & & & & \vdots & \vdots \\ 1 & 0 & \dots & \dots & \dots & 0 & 1 \\ 1 & 1 & \dots & \dots & \dots & 1 & 1 \end{pmatrix} \in \mathcal{M}_n(\mathbb{R})$$

2. Soit $n \in \mathbb{N}^*$ et $a \in \mathbb{R}$, déterminer les éléments propres de la matrice

$$A = \begin{pmatrix} a & 1 & \dots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & a \end{pmatrix} \in \mathcal{M}_n(\mathbb{R})$$

3. **(Oral ESCP)** Soit E l'espace vectoriel des fonctions continues sur \mathbb{R} , à valeurs réelles. Soit $a > 0$ un réel donné. A tout $f \in E$ on associe la fonction $T_a(f)$ définie pour tout x réel par :

$$T_a(f)(x) = \frac{1}{2a} \int_{x-a}^{x+a} f(t) dt$$

1. Montrer que pour tout $f \in E$, $T_a(f)$ est bien définie et est de classe C^1 sur \mathbb{R} .
2. Montrer que $T_a(f)$ est constante si, et seulement si, f est périodique de période $T = 2a$.
3. Montrer que l'application T_a est un endomorphisme de E . Déterminer son noyau. T_a est-il surjectif?
4. Soit $n \geq 2$ un entier naturel et $\mathbb{R}_n[X]$ l'espace vectoriel des fonctions polynômes de degré inférieur ou égal à n . Montrer que la restriction de T_a à $\mathbb{R}_n[X]$ est un endomorphisme de $\mathbb{R}_n[X]$.
On notera encore T_a cette restriction.
5. a) Montrer que la matrice associée à T_a dans la base canonique de $\mathbb{R}_n[X]$ est triangulaire supérieure. En déduire les valeurs propres de T_a . Cet endomorphisme est-il diagonalisable?
b) Soit $f \in \mathbb{R}_n[X]$. Montrer que si le degré de f est égal à 2, f n'est pas vecteur propre de T_a .
c) Montrer que si f est vecteur propre de T_a , sa dérivée f' l'est également. En déduire les sous-espaces propres de T_a .

4. Soit E un espace vectoriel sur \mathbb{K} de dimension finie.

1. Soient u_1, u_2 deux endomorphismes de E .
Montrer que $\dim(\text{Ker}(u_1 \circ u_2)) \leq \dim(\text{Ker}(u_1)) + \dim(\text{Ker}(u_2))$.

2. Généraliser le résultat précédent en prouvant que quels que soient les endomorphismes u_1, u_2, \dots, u_n de E , on a

$$\dim(\text{Ker}(u_1 \circ u_2 \circ \dots \circ u_n)) \leq \sum_{k=1}^n \dim(\text{Ker}(u_k)).$$

3. On suppose dans cette question que $\mathbb{K} = \mathbb{C}$ et on considère $u \in \mathcal{L}(E)$ tel que $u^3 = \text{Id}_E$. Montrer que u est diagonalisable.
 4. Plus généralement, soit $u \in \mathcal{L}(E)$ tel qu’il existe un polynôme annulateur de u à racines simples. Montrer que u est diagonalisable.
 5. Réciproquement, soit u un endomorphisme de E diagonalisable. Montrer que u admet un polynôme annulateur à racines simples.

5. Soit f l’endomorphisme de \mathbb{R}^3 dont la matrice dans la base canonique est

$$A = \begin{pmatrix} 1 & -12 & 2 \\ 1 & 1 & 1 \\ 4 & 8 & 3 \end{pmatrix}.$$

1. Déterminer les valeurs propres de f .
 2. Montrer que $\text{Ker}(f - 3 \text{Id})^2$ et $\text{Ker}(f + \text{Id})$ sont supplémentaires.
 3. Déterminer une base \mathcal{B}' de \mathbb{R}^3 dans laquelle la matrice de f est

$$B = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}.$$

4. Déterminer B^n pour $n \in \mathbb{N}$.
 5. En déduire la valeur de A^n .

6. (Oral ESCP) On considère la matrice $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2(\mathbb{Z})$. On suppose qu’il existe un entier $n \geq 2$ tel que $A^n = I_2$, où I_2 désigne la matrice identité de $\mathcal{M}_2(\mathbb{R})$. Le but de cet exercice est de montrer que $A^{12} = I_2$.

On note σ l’ensemble des valeurs propres (réelles ou complexes) de A .

1. Montrer que $\lambda \in \sigma$ si et seulement si $\lambda^2 - (a + d)\lambda + (ad - bc) = 0$.
 En déduire que σ n’est pas vide.
 On admettra que la matrice A vérifie la relation : $A^2 - (a + d)A + (ad - bc)I_2 = 0$ (*).
 2. Montrer que σ vérifie l’une, et l’une seulement, des deux propositions suivantes :
 a) $\sigma \subset \{-1, 1\}$
 b) il existe un entier p tel que $1 \leq p < n/2$ et $\sigma = \{e^{-2ip\pi/n}, e^{2ip\pi/n}\}$.
 Que peut-on dire, dans ce cas, du nombre $2 \cos(2p\pi/n)$?
 3. On suppose que $\text{card}(\sigma) = 2$. En étudiant les différents cas, montrer que $A^{12} = I_2$.
 4. On suppose que $\sigma = \{1\}$ et que $A \neq I_2$.

- a) En utilisant la relation (\star) , montrer que $\text{Ker}(A - I_2) = \mathfrak{S}(A - I_2)$
 b) En déduire que A est semblable à :

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

- c) Calculer T^k , pour $k \geq 1$. En déduire une contradiction.
 5. Montrer que si $\sigma = \{-1\}$ et $A \neq -I_2$, on arrive également à une contradiction.
 Conclure.

7. (Intersection d'hyperplans) Soit E un espace vectoriel sur \mathbb{K} de dimension n .

Soient $\varphi_1, \varphi_2, \dots, \varphi_p$ des formes linéaires sur E telles que la famille $(\varphi_1, \dots, \varphi_p)$ soit une famille libre de $\mathcal{L}(E, \mathbb{K})$ (en particulier $p \leq n$).

1. On définit l'application f par :

$$f : \begin{cases} E & \rightarrow \mathbb{K}^p \\ x & \mapsto (\varphi_1(x), \dots, \varphi_p(x)) \end{cases}$$

Montrer que $\text{rg}(f) = p$.

2. Montrer que $\text{Ker}(f) = \bigcap_{k=1}^p \text{Ker}(\varphi_k)$.

3. En déduire que

$$\dim \left(\bigcap_{k=1}^p \text{Ker}(\varphi_k) \right) = n - p.$$

8. (Réduction d'un endomorphisme nilpotent) Soit u un endomorphisme d'un \mathbb{C} -

espace vectoriel E tel que $u^k = 0$ et $u^{k-1} \neq 0$ (où $k \in \mathbb{N}^*$).

Soit $x_0 \in E$ tel que $u^{k-1}(x_0) \neq 0$.

1. Prouver que $(x_0, u(x_0), \dots, u^{k-1}(x_0))$ est une famille libre.
 On note pour la suite $F = \text{Vect}(x_0, \dots, u^{k-1}(x_0))$.
 2. Montrer que F est stable par u .
 3. Montrer qu'il existe une forme linéaire φ sur E telle que $\varphi(u^{k-1}(x_0)) \neq 0$.
 4. Montrer que la famille $(\varphi, \varphi \circ u, \dots, \varphi \circ u^{k-1})$ est une famille libre de $E^* = \mathcal{L}(E, \mathbb{C})$.
 5. On note

$$G = \bigcap_{i=0}^{k-1} \text{Ker}(\varphi \circ u^i).$$

Montrer que G est stable par u .

6. Montrer que F et G sont supplémentaires.
 (On utilisera le résultat de l'exercice précédent)

7. En déduire par récurrence sur la dimension de E qu'il existe une base de E dans laquelle la matrice de u est de la forme

$$\begin{pmatrix} 0 & \varepsilon_1 & & (0) \\ & 0 & \varepsilon_2 & \\ & & \ddots & \ddots \\ (0) & & & \ddots & \varepsilon_n \\ & & & & 0 \end{pmatrix}$$

où pour $i \in \llbracket 1, n \rrbracket$, $\varepsilon_i \in \{0, 1\}$.

- 9. (Oral ESCP)** On note $\mathcal{M}_3(\mathbb{R})$ l'espace vectoriel réel des matrices carrées d'ordre 3 à coefficients réels. On considère la matrice A définie par :

$$A = \begin{pmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix}$$

- Déterminer la matrice $B = A^2 + 2I$. On admettra que B est diagonalisable.
- Montrer que $B^2 = B + 2I$.
- Déterminer les valeurs propres de B . En déduire les sous-espaces propres associés.
- Vérifier que si λ est une valeur propre de A , alors $\lambda^2 + 2$ est une valeur propre de B . En déduire que A n'est pas diagonalisable dans $\mathcal{M}_3(\mathbb{R})$.
- Montrer que B est inversible et exprimer B^{-1} en fonction des matrices B et I .
- On s'intéresse maintenant aux puissances de B .
 - On pose, pour tout $n \geq 2$, $X^n = (X^2 - X - 2)Q_n(X) + R_n(X)$ où Q_n et R_n sont deux polynômes tels que $\deg(R_n) < 2$.
On note $R_n(X) = a_n X + b_n$. Déterminer le couple (a_n, b_n) .
 - En déduire l'expression de B^n en fonction de I , B et n , pour $n \geq 0$.
 - Montrer que l'expression de B^n en fonction de I , de B et de n , qui a été obtenue pour $n \geq 0$, est encore valable pour les entiers négatifs.

- 10.** 1. Soit $n \geq 2$ un entier naturel et A une matrice carrée d'ordre n sur \mathbb{C} telle que, pour tout $i \in \llbracket 1, \dots, n \rrbracket$

$$|a_{i,i}| > \sum_{j \neq i} |a_{i,j}|.$$

Montrer que la matrice A est inversible (on pourra raisonner par l'absurde et considérer une colonne X non nulle telle que $AX = 0$).

2. Soit A une matrice carrée d'ordre n quelconque. Soit λ une valeur propre de A . Montrer que :

$$\lambda \in \bigcup_{i=1}^n D(a_{i,i}, \sum_{j \neq i} |a_{i,j}|),$$

où pour $\alpha \in \mathbb{C}$, $R > 0$, $D(\alpha, R) = \{z \in \mathbb{C} / |z - \alpha| \leq R\}$.

3. Soit $n \geq 2$ et A la matrice :

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & 0 & 1 \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}.$$

- a) Montrer que si λ est une valeur propre réelle de A , alors $|\lambda| \leq 2$.
On pose alors $\lambda = 2 \cos(\theta)$, $\theta \in [0, \pi]$.
- b) Déterminer les valeurs propres et les vecteurs propres de A .

11. (Carrés magiques) Soit E l'ensemble des matrices carrées réelles d'ordre n , ($n \geq 2$) formé des matrices $A = (a_{ij})$ telles qu'il existe un réel unique noté $m(A)$ vérifiant la propriété suivante

$$\forall (i, j) \in \{1, \dots, n\}^2, \quad \sum_{k=1}^n a_{i,k} = \sum_{k=1}^n a_{k,j} = m(A).$$

On considère en outre la matrice $J = (j_{p,q})$ de E définie par :

$$\text{pour tout } (p, q) \in \llbracket 1, n \rrbracket^2, j_{p,q} = 1.$$

- 1. Montrer que E est un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{R})$.
Calculer $A.J$ et $J.A$ pour $A \in E$.
En déduire que E est stable par la multiplication des matrices et que l'application $m : A \mapsto m(A)$ est une application linéaire de E sur \mathbb{R} .
- 2. Soit G la droite vectorielle engendrée par J et soit $H = \text{Ker } m$ le noyau de m .
Montrer que $E = G \oplus H$ (on pourra considérer la matrice $A - \left(\frac{m(A)}{n}\right)J$, pour $A \in E$).
- 3. Pour tout couple $(k, l) \in \{2, \dots, n\}^2$, on considère la matrice $H^{k,l}$ dont tous les éléments sont nuls exceptés :

$$h_{1,1}^{k,l} = h_{k,l}^{k,l} = 1, \text{ et } h_{1,l}^{k,l} = h_{k,1}^{k,l} = -1.$$

Montrer que pour tout couple $(k, l) \in \{2, \dots, n\}^2$, $H^{k,l}$ est élément de H et que l'ensemble des matrices $(H^{k,l})$ forme une base de H (si $A = (a_{ij}) \in H$, on pourra considérer la matrice $A' = \sum_{2 \leq k, l \leq n} a_{k,l} H^{k,l}$).
En déduire la dimension de E .

Algèbre bilinéaire

2

Les espaces vectoriels de ce chapitre sont des espaces vectoriels sur \mathbb{R} .

1. Produit scalaire

1.1 Produit scalaire

Définition 1

On appelle produit scalaire sur E , toute application φ bilinéaire sur E , symétrique et définie positive, c'est-à-dire toute application φ de $E \times E$ vers \mathbb{R} telle que

1. $\forall (x, y, z) \in E^3, \forall (\lambda, \mu) \in \mathbb{R}^2, \varphi(\lambda x + \mu y, z) = \lambda \varphi(x, z) + \mu \varphi(y, z)$.
2. $\forall (x, y, z) \in E^3, \forall (\lambda, \mu) \in \mathbb{R}^2, \varphi(x, \lambda y + \mu z) = \lambda \varphi(x, y) + \mu \varphi(x, z)$.
3. $\forall (x, y) \in E^2, \varphi(x, y) = \varphi(y, x)$.
4. $\forall x \in E, \varphi(x, x) \geq 0$.
5. $\forall x \in E, \varphi(x, x) = 0 \Leftrightarrow x = 0$.

On note souvent $\varphi(x, y)$ sous la forme $\langle x, y \rangle$, $\langle x|y \rangle$, $(x|y)$ ou encore $x \cdot y$.

► Remarque

Plus généralement, si φ est une application de $E \times E \rightarrow \mathbb{R}$, on dit que φ est

- bilinéaire si elle vérifie les deux premiers points.
- symétrique si elle vérifie le troisième point.
- positive si elle vérifie le quatrième point.
- définie si elle vérifie le cinquième point.

Proposition 1

Une application φ de $E \times E$ vers \mathbb{R} est un produit scalaire si, et seulement si,

1. $\forall (x, y, z) \in E^3, \forall (\lambda, \mu) \in \mathbb{R}^2, \varphi(\lambda x + \mu y, z) = \lambda \varphi(x, z) + \mu \varphi(y, z)$.
2. $\forall (x, y) \in E^2, \varphi(x, y) = \varphi(y, x)$.
3. $\forall x \in E, \varphi(x, x) \geq 0$.
4. $\forall x \in E, \varphi(x, x) = 0 \Leftrightarrow x = 0$.

Preuve

Si φ est un produit scalaire, il est clair que φ vérifie les propriétés annoncées (puisque l'on reconnaît quatre des cinq propriétés vérifiées par un produit scalaire).

Réciproquement si φ vérifie ces quatre propriétés, il reste donc à prouver le deuxième point. Soient $x, y, z \in E$ et $\lambda, \mu \in \mathbb{K}$, on a

$$\begin{aligned} \varphi(x, \lambda y + \mu z) &= \varphi(\lambda y + \mu z, x) \\ &= \lambda \varphi(y, x) + \mu \varphi(z, x) \\ &= \lambda \varphi(x, y) + \mu \varphi(x, z). \end{aligned}$$

□

► **Remarques**

- Dans toute la suite E désignera un espace vectoriel muni d'un produit scalaire.
- Un produit scalaire sur E induit naturellement un produit scalaire sur tout sous-espace vectoriel de E .

Exemples

1. Soit φ l'application définie par

$$\varphi : \begin{cases} \mathbb{R}^2 \times \mathbb{R}^2 & \rightarrow \mathbb{R} \\ ((x_1, y_1), (x_2, y_2)) & \mapsto x_1x_2 + y_1y_2. \end{cases}$$

Montrons que φ est un produit scalaire.

- Soient $(x_1, y_1), (x_2, y_2), (x_3, y_3) \in \mathbb{R}^2$ et $\lambda, \mu \in \mathbb{R}$, on a

$$\begin{aligned} \varphi(\lambda(x_1, y_1) + \mu(x_2, y_2), (x_3, y_3)) &= \varphi((\lambda x_1 + \mu x_2, \lambda y_1 + \mu y_2), (x_3, y_3)) \\ &= (\lambda x_1 + \mu x_2)x_3 + (\lambda y_1 + \mu y_2)y_3 \\ &= \lambda x_1x_3 + \mu x_2x_3 + \lambda y_1y_3 + \mu y_2y_3 \\ &= \lambda(x_1x_3 + y_1y_3) + \mu(x_2x_3 + y_2y_3) \\ &= \lambda \varphi((x_1, y_1), (x_3, y_3)) + \mu \varphi((x_2, y_2), (x_3, y_3)). \end{aligned}$$

- Soient $(x_1, y_1), (x_2, y_2)$, on a

$$\varphi((x_1, y_1), (x_2, y_2)) = x_1x_2 + y_1y_2 = x_2x_1 + y_2y_1 = \varphi((x_2, y_2), (x_1, y_1)).$$

- Soit $(x, y) \in \mathbb{R}^2$, on a $\varphi((x, y), (x, y)) = x^2 + y^2 \geq 0$.
- Enfin, si (x, y) est un couple de réel tel que $\varphi((x, y), (x, y)) = 0$, alors $x^2 + y^2 = 0$ et donc $x = y = 0$.

Ce produit scalaire s'appelle le produit scalaire canonique de \mathbb{R}^2 .

2. Plus généralement, on montre que l'application

$$\begin{cases} \mathbb{R}^n \times \mathbb{R}^n & \rightarrow \mathbb{R} \\ ((x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_n)) & \mapsto \sum_{k=1}^n x_k y_k \end{cases}$$

est un produit scalaire sur \mathbb{R}^n . Ce produit scalaire est appelé produit scalaire canonique de \mathbb{R}^n . Sauf mention contraire, c'est le produit scalaire dont sera muni \mathbb{R}^n dans la suite.

3. Soit $n \in \mathbb{N}$, l'application définie par

$$\psi : \begin{cases} \mathbb{R}_n[X] \times \mathbb{R}_n[X] & \rightarrow \mathbb{R} \\ (P, Q) & \mapsto \int_{-1}^1 P(t)Q(t)dt \end{cases}$$

est un produit scalaire sur $\mathbb{R}_n[X]$. Admettons la bilinéarité et la symétrie de ψ , qui sont immédiates.

Prouvons les deux derniers points.

- Soit $P \in \mathbb{R}_n[X]$, alors $\psi(P, P) = \int_{-1}^1 P(t)^2 dt$. Or, pour $t \in [-1, 1]$, $P(t)^2 \geq 0$, donc par positivité de l'intégrale, $\psi(P, P) \geq 0$.
- Enfin si P est un polynôme de $\mathbb{R}_n[X]$ tel que $\psi(P, P) = 0$, alors on a $\int_{-1}^1 P(t)^2 dt = 0$. On sait que l'intégrale d'une fonction positive continue est nulle si, et seulement si, cette fonction est nulle. Comme $t \mapsto P(t)^2$ est continue sur $[-1, 1]$, on en déduit que pour tout $t \in [-1, 1]$, $P(t)^2 = 0$ et donc $P(t) = 0$. Le polynôme P admettant une infinité de racines, il est nul.

4. Plus généralement, si a et b sont deux réels tels que $a < b$, l'application

$$\begin{cases} \mathcal{C}([a, b], \mathbb{R}) \times \mathcal{C}([a, b], \mathbb{R}) & \rightarrow \mathbb{R} \\ (f, g) & \mapsto \int_a^b f(t)g(t)dt \end{cases}$$

est un produit scalaire sur $\mathcal{C}([a, b], \mathbb{R})$, ensemble des fonctions réelles définies et continues sur l'intervalle $[a, b]$.

1.2 Norme euclidienne

Définition 2

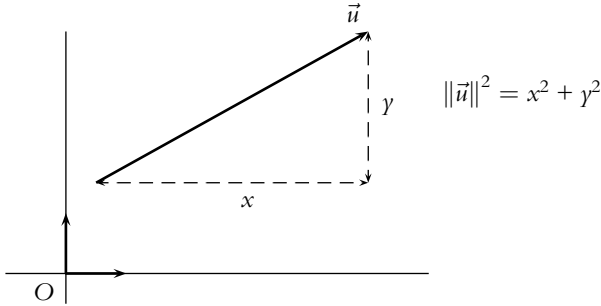
On appelle norme euclidienne associée au produit scalaire de E l'application qui à tout vecteur x de E associe le réel $\sqrt{\langle x, x \rangle}$. Ce réel s'appelle norme de x et est noté $\|x\|$.

Exemples

1. Pour la structure euclidienne canonique, on a $\|(3, -1)\| = \sqrt{3^2 + (-1)^2} = \sqrt{10}$.
2. Pour le produit scalaire $\langle P, Q \rangle = \int_0^1 P(t)Q(t)dt$, on a

$$\|X - 1\| = \sqrt{\int_0^1 (t - 1)^2 dt} = \sqrt{\left[\frac{(t - 1)^3}{3} \right]_0^1} = \sqrt{\frac{1}{3}} = \frac{\sqrt{3}}{3}.$$

3. Dans le plan rapporté à un repère orthonormé, la norme euclidienne d'un vecteur de coordonnées (x, y) est justement égal à $\|(x, y)\|$:



Théorème 1 (Inégalité de Cauchy-Schwarz)

Pour tout couple (x, y) de vecteurs de E , on a l'inégalité :

$$|\langle x, y \rangle| \leq \|x\| \|y\|,$$

avec égalité si, et seulement si, x et y sont colinéaires.

Preuve

Soient x, y deux vecteurs de E .

Si $y = 0$, l'inégalité est vérifiée. C'est en fait une égalité, et on a bien x et y qui sont colinéaires.

Supposons maintenant que y soit non nul. On sait que, pour tout réel λ , $\|x + \lambda y\|^2 \geq 0$, or

$$\begin{aligned} \|x + \lambda y\|^2 &= \langle x + \lambda y, x + \lambda y \rangle = \langle x, x \rangle + \lambda \langle x, y \rangle + \lambda \langle y, x \rangle + \lambda^2 \langle y, y \rangle \\ &= \lambda^2 \|y\|^2 + 2\lambda \langle x, y \rangle + \|x\|^2 \end{aligned}$$

Cette expression qui est un trinôme en λ , car $\|y\|^2 \neq 0$, est donc de signe constant sur \mathbb{R} . On en déduit que son discriminant réduit est négatif, c'est-à-dire que

$$\langle x, y \rangle^2 - \|x\|^2 \|y\|^2 \leq 0,$$

soit $\langle x, y \rangle^2 \leq \|x\|^2 \|y\|^2$, ou encore $|\langle x, y \rangle| \leq \|x\| \|y\|$.

Si cette inégalité est une égalité, alors $\Delta' = 0$ et donc il existe un réel λ_0 tel que $\|x + \lambda_0 y\| = 0$, soit encore $x = -\lambda_0 y$. Les vecteurs x et y sont colinéaires.

Réciproquement, si x et y sont colinéaires, alors il existe un réel λ_0 tel que $x = \lambda_0 y$ (puisque $y \neq 0$). On a alors $\|x - \lambda_0 y\|^2 = 0$. On en déduit que $-\lambda_0$ est racine du trinôme. Son discriminant Δ' est donc nul, c'est-à-dire que l'inégalité est une égalité. \square

► Remarque

L'inégalité reste vraie pour une forme bilinéaire symétrique positive qui n'est pas définie.

Exemples

1. Cette inégalité appliquée à deux vecteurs quelconques (a_1, \dots, a_n) et (b_1, \dots, b_n) de \mathbb{R}^n donne

$$\left| \sum_{k=1}^n a_k b_k \right| \leq \sqrt{\sum_{k=1}^n a_k^2} \sqrt{\sum_{k=1}^n b_k^2}.$$

2. De même, quelles que soient les fonctions réelles f et g continues sur $[a, b]$, on a

$$\left| \int_a^b f(t)g(t)dt \right| \leq \sqrt{\int_a^b f(t)^2 dt} \sqrt{\int_a^b g(t)^2 dt}.$$

Proposition 2

La norme vérifie

1. $\forall x \in E, \|x\| = 0 \Leftrightarrow x = 0$.
2. $\forall \lambda \in \mathbb{R}, \forall x \in E, \|\lambda x\| = |\lambda| \|x\|$.
3. $\forall (x, y) \in E^2, \|x + y\| \leq \|x\| + \|y\|$ (inégalité triangulaire).

Preuve

1. C'est une conséquence directe de la dernière propriété d'un produit scalaire.
2. Soit $\lambda \in \mathbb{R}$ et $x \in E$, on a

$$\|\lambda x\| = \sqrt{\varphi(\lambda x, \lambda x)} = \sqrt{\lambda^2 \varphi(x, x)} = |\lambda| \|x\|.$$

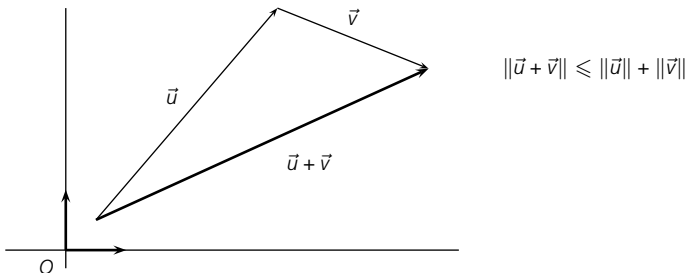
3. Soient $x, y \in E$, on a d'après l'inégalité de Cauchy-Schwarz

$$\begin{aligned} \|x + y\|^2 &= \|x\|^2 + 2 \langle x, y \rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 \\ &\leq (\|x\| + \|y\|)^2, \end{aligned}$$

d'où le résultat. □

► **Remarque**

La dernière propriété est une généralisation de l'inégalité triangulaire



Proposition 3

Soit $(x, y) \in E^2$, on a les égalités

$$\begin{aligned} \|x + y\|^2 &= \|x\|^2 + 2 \langle x, y \rangle + \|y\|^2 \\ \langle x, y \rangle &= \frac{1}{2} \left(\|x + y\|^2 - \|x\|^2 - \|y\|^2 \right). \end{aligned}$$

Preuve

La bilinéarité et la symétrie du produit scalaire permettent d'écrire

$$\begin{aligned}\|x + y\|^2 &= \langle x + y, x + y \rangle = \langle x, x + y \rangle + \langle y, x + y \rangle \\ &= \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \\ &= \|x\|^2 + 2 \langle x, y \rangle + \|y\|^2.\end{aligned}$$

La seconde égalité est une conséquence directe de la première. \square

Exemple

La seconde égalité permet de déterminer si une application donnée N de E vers \mathbb{R}_+ est une norme euclidienne associée à un produit scalaire.

1.3 Orthogonalité**Vecteurs orthogonaux****Définition 3**

Deux vecteurs x, y de E sont dits orthogonaux si $\langle x, y \rangle = 0$. On notera alors $x \perp y$.

Exemples

1. Les vecteurs $(1, 2)$ et $(-4, 2)$ sont orthogonaux puisque $1 \times (-4) + 2 \times 2 = 0$.
2. Les vecteurs de la base canonique de \mathbb{R}^n sont orthogonaux deux à deux.

Théorème 2 (Théorème de Pythagore)

Soient x, y deux vecteurs orthogonaux de E , on a $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Preuve

En effet, si $\langle x, y \rangle = 0$, alors

$$\|x + y\|^2 = \|x\|^2 + 2 \langle x, y \rangle + \|y\|^2 = \|x\|^2 + \|y\|^2.$$

 \square **Définition 4**

Une famille (e_1, e_2, \dots, e_n) de vecteurs de E est dite orthogonale si

$$\forall (i, j) \in \llbracket 1, n \rrbracket^2, i \neq j \Rightarrow \langle e_i, e_j \rangle = 0.$$

Exemples

1. La base canonique de \mathbb{R}^n est une famille orthogonale.
2. Soit n un entier naturel non nul. Définissons pour $k \in \llbracket 1, n \rrbracket$, la fonction f_k sur \mathbb{R} par $f_k(x) = \cos(kx)$. La famille (f_1, f_2, \dots, f_n) est une famille de l'espace vectoriel des fonctions réelles continues. Montrons que c'est une famille orthogonale pour le produit scalaire

$$\langle f, g \rangle = \int_0^{2\pi} f(t)g(t)dt.$$

Soient k_1 et k_2 deux entiers distincts de l'intervalle $\llbracket 1, n \rrbracket$, on a

$$\begin{aligned} \langle f_{k_1}, f_{k_2} \rangle &= \int_0^{2\pi} \cos(k_1 t) \cos(k_2 t) dt \\ &= \int_0^{2\pi} \frac{1}{2} (\cos(k_1 t + k_2 t) + \cos(k_1 t - k_2 t)) dt \\ &= \frac{1}{2} \int_0^{2\pi} \cos((k_1 + k_2)t) + \cos((k_1 - k_2)t) dt \\ &= \frac{1}{2} \left[\frac{\sin((k_1 + k_2)t)}{k_1 + k_2} + \frac{\sin((k_1 - k_2)t)}{k_1 - k_2} \right]_0^{2\pi} \quad (\text{car } k_1 + k_2 \neq 0 \text{ et } k_1 - k_2 \neq 0) \\ &= 0, \end{aligned}$$

et donc $f_{k_1} \perp f_{k_2}$.

Proposition 4

Toute famille orthogonale ne contenant pas le vecteur nul est libre.

Preuve

Soit (e_1, e_2, \dots, e_n) une telle famille et soient a_1, a_2, \dots, a_n des réels tels que $\sum_{k=1}^n a_k e_k = 0$. Pour $i \in \llbracket 1, n \rrbracket$, on a donc

$$\left\langle e_i, \sum_{k=1}^n a_k e_k \right\rangle = \sum_{k=1}^n a_k \langle e_i, e_k \rangle = a_i \langle e_i, e_i \rangle = a_i \|e_i\|^2 = 0$$

Comme $\|e_i\| \neq 0$, on en déduit que $a_i = 0$. La famille (e_1, e_2, \dots, e_n) est bien une famille libre. \square

Définition 5

Une famille (e_1, e_2, \dots, e_n) de vecteurs de E est dite orthonormale (ou orthonormée) si elle est orthogonale et si pour tout $k \in \llbracket 1, n \rrbracket$, $\|e_k\| = 1$.

► Remarque

D'après la proposition 4, on peut donc dire que toute famille orthonormale est libre.

Proposition 5

Soit (e_1, e_2, \dots, e_n) une famille orthonormale de E , pour tout vecteur $x \in \text{Vect}(e_1, e_2, \dots, e_n)$, on a

$$x = \sum_{k=1}^n \langle x, e_k \rangle e_k.$$

Preuve

Soient a_1, a_2, \dots, a_n les réels tels que $x = \sum_{k=1}^n a_k e_k$ (ces réels sont uniques puisque une famille orthonormale

est libre). Soit $i \in \llbracket 1, n \rrbracket$, on a

$$\langle x, e_i \rangle = \left\langle \sum_{k=1}^n a_k e_k, e_i \right\rangle = \sum_{k=1}^n a_k \langle e_k, e_i \rangle = a_i$$

d'où le résultat. □

Théorème 3 (Orthonormalisation de Schmidt)

Soit (e_1, e_2, \dots, e_n) une famille libre de vecteurs de E .

Posons $f_1 = \frac{e_1}{\|e_1\|}$, et pour $k \in \llbracket 2, n \rrbracket$

$$f_k = \frac{1}{\left\| e_k - \sum_{i=1}^{k-1} \langle e_k, f_i \rangle f_i \right\|} \cdot \left(e_k - \sum_{i=1}^{k-1} \langle e_k, f_i \rangle f_i \right).$$

Alors la famille (f_1, f_2, \dots, f_n) est orthonormée et de plus,

$$\forall k \in \llbracket 1, n \rrbracket, \quad \text{Vect}(e_1, e_2, \dots, e_k) = \text{Vect}(f_1, f_2, \dots, f_k).$$

Preuve

Prouvons par une récurrence forte sur k que (f_1, f_2, \dots, f_k) forme une famille orthonormale telle que $\text{Vect}(f_1, \dots, f_k) = \text{Vect}(e_1, \dots, e_k)$.

Pour $k = 1$, on a $e_1 \neq 0$ puisque c'est un vecteur d'une famille libre. Le vecteur f_1 est donc bien défini et $\text{Vect}(f_1) = \text{Vect}\left(\frac{1}{\|e_1\|} e_1\right) = \text{Vect}(e_1)$. Comme $\|f_1\| = \left\| \frac{1}{\|e_1\|} e_1 \right\| = \frac{\|e_1\|}{\|e_1\|} = 1$, la famille (f_1) est bien orthonormale.

Supposons la propriété vraie jusqu'au rang k .

Remarquons déjà que $e_{k+1} \notin \text{Vect}(e_1, \dots, e_k) = \text{Vect}(f_1, \dots, f_k)$ et donc

$$\left\| e_{k+1} - \sum_{i=1}^k \langle e_{k+1}, f_i \rangle f_i \right\| \neq 0.$$

Autrement dit f_{k+1} est bien défini. De plus $\|f_{k+1}\| = 1$ par construction et pour $j \in \llbracket 1, k \rrbracket$,

$$\begin{aligned} \left\| e_{k+1} - \sum_{i=1}^k \langle e_{k+1}, f_i \rangle f_i \right\| \times \langle f_{k+1}, f_j \rangle &= \left\langle \left(e_{k+1} - \sum_{i=1}^k \langle e_{k+1}, f_i \rangle f_i \right), f_j \right\rangle \\ &= \langle e_{k+1}, f_j \rangle - \sum_{i=1}^k \langle e_{k+1}, f_i \rangle \langle f_i, f_j \rangle \\ &= \langle e_{k+1}, f_j \rangle - \langle e_{k+1}, f_j \rangle = 0 \end{aligned}$$

donc $\langle f_{k+1}, f_j \rangle = 0$, ce qui prouve que la famille $(f_1, f_2, \dots, f_{k+1})$ est orthonormale. Enfin, on a clairement $f_{k+1} \in \text{Vect}(e_1, \dots, e_{k+1})$ d'où $\text{Vect}(f_1, \dots, f_{k+1}) \subset \text{Vect}(e_1, \dots, e_{k+1})$. La famille $(e_1, e_2, \dots, e_{k+1})$ étant libre, $\dim(\text{Vect}(e_1, \dots, e_{k+1})) = k + 1$. Mais la famille (f_1, \dots, f_{k+1}) est elle-même libre car orthonormale. Or son cardinal est également $k + 1$, c'est donc une base de $\text{Vect}(e_1, \dots, e_{k+1})$ et en particulier $\text{Vect}(f_1, \dots, f_{k+1}) = \text{Vect}(e_1, \dots, e_{k+1})$. □

► Remarques

- Si la famille (e_1, e_2, \dots, e_n) est orthonormale, le procédé d'orthonormalisation de Schmidt ne modifie pas cette famille.

- Si $\lambda_1, \lambda_2, \dots, \lambda_n$ sont des réels strictement positifs, les procédés d'orthonormalisation de Schmidt appliqués à (e_1, e_2, \dots, e_n) et à $(\lambda_1 e_1, \lambda_2 e_2, \dots, \lambda_n e_n)$ aboutissent à la même famille orthonormale.
- On peut remarquer en particulier que pour tout réel λ strictement positif, $\frac{e}{\|e\|} = \frac{\lambda e}{\|\lambda e\|}$, propriété dont on se servira pour simplifier les calculs.
Par exemple, le normalisé de $\frac{17}{201}(2, -2, 0)$ est aussi le normalisé de $(1, -1, 0)$ et c'est donc $\frac{1}{\sqrt{2}}(1, -1, 0)$.

Exemples

1. Soit $(1, X, X^2)$ la base canonique de $\mathbb{R}_2[X]$ muni du produit scalaire donné en début de chapitre. Déterminons (Q_0, Q_1, Q_2) la famille obtenue par le procédé d'orthonormalisation de Schmidt appliqué à $(1, X, X^2)$.

On a

$$\|1\|^2 = \int_{-1}^1 1^2 dt = \int_{-1}^1 1 dt = 2$$

d'où $Q_0 = \frac{1}{\sqrt{2}}$. Ensuite

$$\langle X, Q_0 \rangle = \frac{1}{\sqrt{2}} \int_{-1}^1 t dt = \frac{1}{\sqrt{2}} \left[\frac{t^2}{2} \right]_{-1}^1 = 0$$

On en déduit que

$$\|X - \langle X, Q_0 \rangle \times Q_0\|^2 = \|X\|^2 = \int_{-1}^1 t^2 dt = \left[\frac{t^3}{3} \right]_{-1}^1 = \frac{2}{3}$$

d'où $Q_1 = \sqrt{\frac{3}{2}}X$.

Enfin

$$\langle X^2, Q_0 \rangle = \frac{1}{\sqrt{2}} \int_{-1}^1 t^2 dt = \frac{1}{\sqrt{2}} \left[\frac{t^3}{3} \right]_{-1}^1 = \frac{\sqrt{2}}{3}$$

$$\langle X^2, Q_1 \rangle = \sqrt{\frac{3}{2}} \int_{-1}^1 t^3 dt = 0$$

d'où

$$\begin{aligned} \|X^2 - \langle X^2, Q_0 \rangle Q_0 - \langle X^2, Q_1 \rangle Q_1\|^2 &= \|X^2 - \frac{1}{3}\|^2 = \int_{-1}^1 \left(t^2 - \frac{1}{3} \right)^2 dt \\ &= \int_{-1}^1 \left(t^4 - \frac{2}{3}t^2 + \frac{1}{9} \right) dt = \frac{2}{5} - \frac{4}{9} + \frac{2}{9} \\ &= \frac{18 - 20 + 10}{45} = \frac{8}{45} \end{aligned}$$

et finalement $Q_2 = \sqrt{\frac{45}{8}} \left(X^2 - \frac{1}{3} \right) = \frac{\sqrt{10}}{4} (3X^2 - 1)$.

2. Recherchons une base orthonormale du sous-espace A de \mathbb{R}^3 défini par l'équation $3x - y + 2z = 0$. On a

$$\begin{aligned} A &= \{(x, 3x + 2z, z) \mid x, z \in \mathbb{R}\} \\ &= \text{Vect}((1, 3, 0), (0, 2, 1)). \end{aligned}$$

Appliquons le procédé d'orthonormalisation de Schmidt à la famille $(e_1, e_2) = ((1, 3, 0), (0, 2, 1))$. On a comme premier vecteur

$$f_1 = \frac{1}{\|e_1\|} e_1 = \frac{1}{\sqrt{10}}(1, 3, 0)$$

et pour le second vecteur

$$\begin{aligned} e_2 - \langle e_2, f_1 \rangle f_1 &= (0, 2, 1) - \frac{1}{\sqrt{10}} \langle (0, 2, 1), (1, 3, 0) \rangle \cdot \frac{1}{\sqrt{10}}(1, 3, 0) = (0, 2, 1) - \frac{3}{5}(1, 3, 0) \\ &= \frac{1}{5}(-3, 1, 5) \end{aligned}$$

d'où $f_2 = \frac{1}{\sqrt{35}}(-3, 1, 5)$.

Une base orthonormale de A est donc par exemple

$$\left(\frac{1}{\sqrt{10}}(1, 3, 0), \frac{1}{\sqrt{35}}(-3, 1, 5) \right).$$

Sous-espaces orthogonaux

Définition 6

Soient F et G deux sous-espaces vectoriels de E , on dit que F et G sont orthogonaux si

$$\forall (x, y) \in F \times G, \langle x, y \rangle = 0.$$

On écrira alors $F \perp G$.

Proposition 6

Soient F et G deux sous-espaces vectoriels de E tels que

$$F = \text{Vect}(f_1, f_2, \dots, f_n), \quad G = \text{Vect}(g_1, g_2, \dots, g_p).$$

Les propriétés suivantes sont équivalentes

- (i) $F \perp G$;
- (ii) $\forall (i, j) \in \llbracket 1, n \rrbracket \times \llbracket 1, p \rrbracket, \langle f_i, g_j \rangle = 0$.

Preuve

(i) \Rightarrow (ii) C'est immédiat.

(ii) \Rightarrow (i) Soient $x \in F$ et $y \in G$, tels que $x = \sum_{i=1}^n a_i f_i$ et $y = \sum_{j=1}^p b_j g_j$, on a

$$\langle x, y \rangle = \left\langle \sum_{i=1}^n a_i f_i, \sum_{j=1}^p b_j g_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^p a_i b_j \langle f_i, g_j \rangle = 0.$$

Il en résulte que $F \perp G$. □

Exemples

1. Dans \mathbb{R}^3 , les sous-espaces $F = \text{Vect}((1, 3, -4))$ et $G = \text{Vect}((3, -1, 0), (0, 4, 3))$ sont orthogonaux puisque

$$\langle (1, 3, -4), (3, -1, 0) \rangle = 3 - 3 = 0 \quad \text{et} \quad \langle (1, 3, -4), (0, 4, 3) \rangle = 12 - 12 = 0.$$

2. Soit n un entier naturel non nul. On définit pour $k \in \llbracket 1, n \rrbracket$ f_k et g_k sur \mathbb{R} par

$$f_k(x) = \cos(kx) \quad \text{et} \quad g_k(x) = \sin(kx)$$

Montrons que $\text{Vect}(f_1, \dots, f_n)$ et $\text{Vect}(g_1, \dots, g_n)$ sont deux sous-espaces vectoriels de $\mathcal{C}(\mathbb{R}, \mathbb{R})$ pour le produit scalaire

$$\langle f, g \rangle = \int_0^{2\pi} f(t)g(t)dt.$$

Pour tout couple $(i, j) \in \llbracket 1, n \rrbracket^2$, nous avons

$$\begin{aligned} \langle f_i, g_j \rangle &= \int_0^{2\pi} f_i(t)g_j(t)dt \\ &= \int_0^{2\pi} \cos(it) \sin(jt)dt \\ &= \frac{1}{2} \int_0^{2\pi} \sin((i+j)t) + \sin((i-j)t)dt \end{aligned}$$

Deux cas se présentent. Si $i = j$, alors $\sin((i-j)t) = \sin(0) = 0$ et on l'obtient

$$\begin{aligned} \langle f_i, g_j \rangle &= \frac{1}{2} \left[-\frac{\cos((i+j)t)}{i+j} \right]_0^{2\pi} = -\frac{1}{i+j} + \frac{1}{i+j} \\ &= 0; \end{aligned}$$

si $i \neq j$,

$$\begin{aligned} \langle f_i, g_j \rangle &= \frac{1}{2} \left[-\frac{\cos((i+j)t)}{i+j} - \frac{\cos((i-j)t)}{i-j} \right]_0^{2\pi} \\ &= 0. \end{aligned}$$

Dans tous les cas $\langle f_i, g_j \rangle = 0$, d'où $\text{Vect}(f_1, \dots, f_n) \perp \text{Vect}(g_1, \dots, g_n)$.

Définition 7

Soit A une partie de E , on pose

$$A^\perp = \{x \in E \mid \forall a \in A \langle x, a \rangle = 0\}.$$

L'ensemble A^\perp s'appelle orthogonal de A .

► Remarque

Si F et G sont deux sous-espaces vectoriels de E alors $F \perp G$ équivaut à $G \subset F^\perp$, mais aussi à $F \subset G^\perp$.

Exemples

1. On a $\{0_E\}^\perp = E$ et $E^\perp = \{0_E\}$.
2. Déterminons dans \mathbb{R}^2 , $\{(-1, 3)\}^\perp$. Soit $(x, y) \in \mathbb{R}^2$, on a

$$(x, y) \in \{(-1, 3)\}^\perp \Leftrightarrow \langle (x, y), (-1, 3) \rangle \Leftrightarrow -x + 3y = 0 \Leftrightarrow x = 3y.$$

On en déduit que $\{(-1, 3)\}^\perp = \text{Vect}(3, 1)$.

Proposition 7

Pour toute partie A de E , A^\perp est un sous-espace vectoriel de E .

Preuve

- Par définition A^\perp est bien une partie de E .
- A^\perp est non vide. En effet, pour tout $a \in A$, $\langle 0, a \rangle = 0$, d'où $0 \in A^\perp$.
- Soit $\lambda \in \mathbb{R}$, et $x, y \in A^\perp$, on a pour tout $a \in A$,

$$\langle a, \lambda x + y \rangle = \lambda \langle a, x \rangle + \langle a, y \rangle = 0$$

d'où $\lambda x + y \in A^\perp$. □

2. Espaces euclidiens

2.1 Premières propriétés

Définition 8

On appelle espace euclidien tout espace vectoriel réel de dimension finie muni d'un produit scalaire.

Dans toute la suite E désignera un espace euclidien.

Proposition 8

Tout espace euclidien admet une base orthonormale.

Preuve

Soit E un espace euclidien. En particulier E est un espace de dimension finie, donc il possède une base \mathcal{B} . En appliquant le procédé d'orthonormalisation de Schmidt à cette base, on trouve une famille orthonormale \mathcal{B}' telle que $\text{Vect}(\mathcal{B}) = \text{Vect}(\mathcal{B}')$. La famille \mathcal{B}' est donc à la fois libre et génératrice de E . C'est une base orthonormale de E . □

Théorème 4

Toute famille orthonormale de E peut être complétée en une base orthonormale de E .

Preuve

Soit $\mathcal{F} = (e_1, e_2, \dots, e_k)$ une famille orthonormale. On sait que cette famille est libre. On peut donc la compléter en une base $\mathcal{B} = (e_1, \dots, e_k, e_{k+1}, \dots, e_n)$ de E . En appliquant le procédé d'orthonormalisation de Schmidt à la famille \mathcal{B} , on obtient une famille \mathcal{B}' qui est une base orthonormale de E . Or il se trouve que le procédé ne modifie pas e_1, e_2, \dots, e_k . La famille \mathcal{B}' est donc une base orthonormale de E qui complète la famille \mathcal{B} . \square

Théorème 5

Soit (e_1, e_2, \dots, e_n) une base orthonormée de E . Pour tout $x \in E$,

$$x = \sum_{k=1}^n \langle x, e_k \rangle e_k \quad \text{et} \quad \|x\|^2 = \sum_{k=1}^n \langle x, e_k \rangle^2.$$

Preuve

La première propriété a déjà été prouvée.

Un petit calcul prouve la seconde égalité

$$\begin{aligned} \|x\|^2 &= \left\langle \sum_{i=1}^n \langle x, e_i \rangle e_i, \sum_{j=1}^n \langle x, e_j \rangle e_j \right\rangle = \sum_{i=1}^n \langle x, e_i \rangle \left\langle e_i, \sum_{j=1}^n \langle x, e_j \rangle e_j \right\rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n \langle x, e_i \rangle \langle x, e_j \rangle \underbrace{\langle e_i, e_j \rangle}_{\substack{0 \text{ si } i \neq j \text{ et } 1 \text{ sinon}}} \\ &= \sum_{i=1}^n \langle x, e_i \rangle^2. \end{aligned}$$

\square

Proposition 9

Soit $u \in \mathcal{L}(E)$, $\mathcal{B} = (e_1, \dots, e_n)$ une base orthonormale de E et $A = (a_{i,j}) = \mathcal{M}_{\mathcal{B}}(u)$. On a

$$a_{i,j} = \langle e_i, u(e_j) \rangle.$$

Preuve

En effet pour $j \in \llbracket 1, n \rrbracket$, on a $u(e_j) = \sum_{i=1}^n \langle e_i, u(e_j) \rangle e_i$. \square

Proposition 10

Soient \mathcal{B}_1 et \mathcal{B}_2 deux bases orthonormales de E . Si P désigne la matrice de changement de base de \mathcal{B}_1 vers \mathcal{B}_2 , on a ${}^t P P = I$.

Preuve

Posons $\mathcal{B}_1 = (e_1, \dots, e_n)$, $\mathcal{B}_2 = (f_1, \dots, f_n)$, $P = (a_{i,j})_{1 \leq i,j \leq n}$ et $C = (c_{i,j})_{1 \leq i,j \leq n} = {}^t P P$. Compte tenu de l'égalité $f_j = \sum_{i=1}^n \langle e_i, f_j \rangle e_i$, on a $a_{i,j} = \langle e_i, f_j \rangle$. On en déduit que pour $(i, j) \in \llbracket 1, n \rrbracket^2$,

$$c_{i,j} = \sum_{k=1}^n a_{k,i} a_{k,j} = \sum_{k=1}^n \langle e_k, f_i \rangle \langle e_k, f_j \rangle = \langle f_i, f_j \rangle = \delta_{i,j},$$

c'est-à-dire $C = I_n$. \square

Définition 9

On appelle matrice orthogonale toute matrice carrée P telle que ${}^tPP = I$.

► Remarque

Une matrice est orthogonale si, et seulement si, ses vecteurs colonnes forment une famille orthonormale de $\mathcal{M}_{n,1}(\mathbb{R})$ muni de sa structure euclidienne canonique.

Exemple

La matrice $\begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$ est orthogonale.

Théorème 6 (Expression matricielle du produit scalaire dans une base orthonormale)

Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base orthonormale de E .

Soient $x = \sum_{k=1}^n x_k e_k$ et $y = \sum_{k=1}^n y_k e_k$ deux vecteurs de E .

Posons $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$ et $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$.

On a, en identifiant les éléments de $\mathcal{M}_{1,1}(\mathbb{R})$ aux réels,

$$\langle x, y \rangle = {}^tXY \quad \text{et} \quad \|x\|^2 = {}^tXX.$$

Preuve

Un calcul montre que ${}^tXY = \left(\sum_{k=1}^n x_k y_k \right)$, ce qui prouve les deux égalités. \square

Définition 10

Soit $\mathcal{B} = (e_1, e_2, \dots, e_n)$ une base de E , on appelle matrice du produit scalaire \langle, \rangle dans la base \mathcal{B} , la matrice $A = (a_{i,j}) \in \mathcal{M}_n(\mathbb{R})$ définie par $a_{i,j} = \langle e_i, e_j \rangle$.

► Remarques

- La matrice d'un produit scalaire est symétrique.
- La base \mathcal{B} est orthogonale si, et seulement si, la matrice du produit scalaire dans la base \mathcal{B} est diagonale. Elle est orthonormale si, et seulement si, la matrice du produit scalaire est la matrice identité.

Exemple

La matrice du produit scalaire canonique de \mathbb{R}^3 dans la base $((1, 0, -1), (1, 0, 3), (1, 2, 0))$ est

$$\begin{pmatrix} 2 & -2 & 1 \\ -2 & 10 & 1 \\ 1 & 1 & 5 \end{pmatrix}.$$

Théorème 7 (Expression matricielle du produit scalaire dans une base quelconque)

Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base de E muni d'un produit scalaire dont la matrice est notée A .

Soient $x = \sum_{k=1}^n x_k e_k$ et $y = \sum_{k=1}^n y_k e_k$ deux vecteurs de E .

Posons $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$ et $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$.

On a, en identifiant les éléments de $\mathcal{M}_{1,1}(\mathbb{R})$ aux réels,

$$\langle x, y \rangle = {}^t XAY \quad \text{et} \quad \|x\|^2 = {}^t XAX.$$

Preuve

Posons $AY = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$. On a pour $i \in \llbracket 1, n \rrbracket$, $b_i = \sum_{j=1}^n \langle e_i, e_j \rangle y_j$. On en déduit que

$${}^t XAY = \sum_{i=1}^n x_i b_i = \sum_{i=1}^n x_i \left(\sum_{j=1}^n \langle e_i, e_j \rangle y_j \right) = \sum_{i=1}^n \sum_{j=1}^n x_i y_j \langle e_i, e_j \rangle.$$

Mais par ailleurs, par bilinéarité du produit scalaire

$$\langle x, y \rangle = \left\langle \sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n x_i y_j \langle e_i, e_j \rangle.$$

Le résultat annoncé en découle. □

► **Remarque**

On peut noter que la première égalité est caractéristique de A . Autrement dit, si A et B sont deux matrices telles que ${}^t XAY = {}^t XBY$ pour $X, Y \in \mathcal{M}_{n,1}(\mathbb{R})$ alors $A = B$.

En effet, si l'on note (X_1, \dots, X_n) la base canonique de $\mathcal{M}_{n,1}(\mathbb{R})$, on a alors pour $(i, j) \in \llbracket 1, n \rrbracket^2$, ${}^t X_i A X_j = a_{ij} = {}^t X_i B X_j = b_{ij}$.

La seconde égalité est caractéristique de A si l'on exige que la matrice soit symétrique. En effet soient A et B sont deux matrices symétriques de $\mathcal{M}_n(\mathbb{R})$ telles que pour $X \in \mathcal{M}_{n,1}(\mathbb{R})$, ${}^t XAX = {}^t XBX$. Pour $i \in \llbracket 1, n \rrbracket$, ${}^t X_i A X_i = a_{ii} = {}^t X_i B X_i = b_{ii}$, et pour $(i, j) \in \llbracket 1, n \rrbracket^2$

$$\begin{aligned} {}^t (X_i + X_j) A (X_i + X_j) &= a_{ji} + a_{ij} + a_{ii} + a_{jj} \\ \text{et } {}^t (X_i + X_j) B (X_i + X_j) &= b_{ji} + b_{ij} + b_{ii} + b_{jj} \end{aligned}$$

Comme A et B sont symétriques, que $a_{ij} = b_{i,j}$ et $a_{jj} = b_{j,j}$, on en déduit que $a_{ij} = b_{ij}$.

Théorème 8

Soient \mathcal{B}_1 et \mathcal{B}_2 deux bases de E , A la matrice du produit scalaire dans la base \mathcal{B}_1 et B la matrice du produit scalaire dans la base \mathcal{B}_2 .

En notant $P = P_{\mathcal{B}_1, \mathcal{B}_2}$, on a la relation $B = {}^t P A P$.

Preuve

Posons $n = \dim(E)$.

Soient x, y deux vecteurs de E dont les coordonnées dans la base \mathcal{B}_1 sont $X, Y \in \mathcal{M}_{n,1}(\mathbb{R})$ et les coordonnées dans la base \mathcal{B}_2 sont $X', Y' \in \mathcal{M}_{n,1}(\mathbb{R})$. D'après la proposition précédente, on peut écrire que

$$\langle x, y \rangle = {}^t XAY = {}^t X'BY'.$$

Comme on a également $X = PX'$ et $Y = PY'$, on en déduit que

$$\langle x, y \rangle = {}^t(PX')A(PY') = {}^tX'({}^tPAP)Y'.$$

Donc quels que soient $X', Y' \in \mathcal{M}_{n,1}(\mathbb{R})$, ${}^tX'BY' = X'({}^tPAP)Y'$.

Il en résulte que $B = {}^tPAP$. □

2.2 Supplémentaire orthogonal**Proposition 11**

Si $F = \text{Vect}(e_1, \dots, e_n)$, alors, pour $x \in E$,

$$x \in F^\perp \Leftrightarrow \forall i \in \llbracket 1, n \rrbracket, \langle x, e_i \rangle = 0.$$

Preuve

Supposons que $x \in F^\perp$. Soit $i \in \llbracket 1, n \rrbracket$. Comme $e_i \in F$, on en déduit que $\langle x, e_i \rangle = 0$.

Réciproquement, supposons que $\forall i \in \llbracket 1, n \rrbracket, \langle x, e_i \rangle = 0$. Soit $y \in F$, par définition il existe des réels

a_1, a_2, \dots, a_n tels que $\sum_{i=1}^n a_i e_i$, on a

$$\langle x, y \rangle = \left\langle x, \sum_{i=1}^n a_i e_i \right\rangle = \sum_{i=1}^n a_i \langle x, e_i \rangle = 0. \quad \square$$

Théorème 9

Soit F un sous-espace vectoriel de E , les sous-espaces F et F^\perp sont supplémentaires. F^\perp s'appelle le supplémentaire orthogonal de F .

Preuve

Soit $x \in F \cap F^\perp$. Par définition, on a $x \perp x$, c'est-à-dire $\|x\|^2 = 0$, soit $x = 0$. On en déduit que $F \cap F^\perp = \{0\}$.

Soit (e_1, \dots, e_k) une base orthonormale de F que l'on complète en une base orthonormale de E , (e_1, e_2, \dots, e_n) .

D'après la proposition précédente $e_{k+1}, \dots, e_n \in F^\perp$. Pour tout vecteur $x \in E$, il existe des réels a_1, \dots, a_n tels

que $x = \sum_{i=1}^n a_i e_i$. En posant $y = \sum_{i=1}^k a_i e_i$ et $z = \sum_{i=k+1}^n a_i e_i$, on a $y \in F$, $z \in F^\perp$ et $x = y + z$. Autrement dit

$$E = F + F^\perp.$$

Ces deux points permettent de conclure que F et F^\perp sont supplémentaires. □

► Remarque

On en déduit en particulier que $\dim(F^\perp) = \dim(E) - \dim(F)$.

2.3 Projection orthogonale

Définition 11

Soit F un sous-espace vectoriel de E , on appelle projection orthogonale sur F la projection sur F parallèlement à F^\perp . Cette projection est notée p_F .

► Remarque

Le projecteur associé à p_F est p_{F^\perp} .

Dans la suite de la partie 2.3, F désignera un sous-espace vectoriel fixé de E et on notera $p = p_F$.

Proposition 12

Soit $x \in E$, on a pour tout $y \in E$

$$y = p(x) \Leftrightarrow \begin{cases} y \in F \\ x - y \in F^\perp \end{cases}$$

Preuve

- Supposons que $y = p(x)$. On a alors $y \in \Im(p)$, c'est-à-dire $y \in F$. De plus, d'après les propriétés connues sur les projecteurs, $x = \underbrace{x - p(x)}_{\in F^\perp} + \underbrace{p(x)}_{\in F}$. Autrement dit $x - p(x) \in F^\perp$, soit $x - y \in F^\perp$.
 - Supposons que y vérifie les deux propriétés annoncées. On a alors $x = \underbrace{x - y}_{\in F^\perp} + \underbrace{y}_{\in F}$.
- Cela suffit pour conclure que $y = p(x)$. □

Théorème 10

Soit $x \in E$, on a, pour tout $y \in F$,

$$y = p(x) \Leftrightarrow \|x - y\| = \min_{z \in F} \|x - z\|.$$

Autrement dit, $p(x)$ est l'unique vecteur de F qui minimise la distance de x à un vecteur quelconque de F .

On dit que $p(x)$ est la meilleure approximation de x dans F .

On appelle distance de x à F , et on note $d(x, F)$, la norme $\|x - p(x)\|$.

Preuve

Soit $y \in F$. On sait que $x - p(x) \in F^\perp$. De plus, comme $y \in F$ et $p(x) \in F$, on a $p(x) - y \in F$ (en effet F est un sous-espace vectoriel). On en déduit que $x - p(x) \perp p(x) - y$.

D'après le théorème de Pythagore, on a donc

$$\|x - y\|^2 = \|x - p(x) + p(x) - y\|^2 = \|x - p(x)\|^2 + \|p(x) - y\|^2 \geq \|x - p(x)\|^2.$$

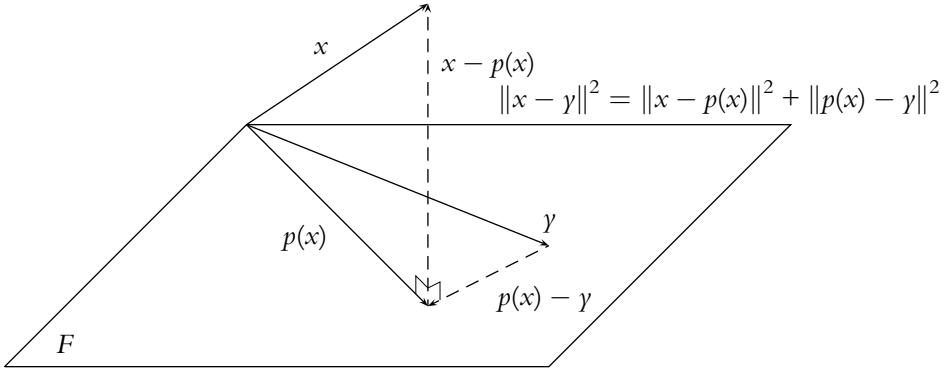
Cela prouve que

$$y = p(x) \Rightarrow \|x - y\| = \min_{z \in F} \|x - z\|$$

Réciproquement, si $y \neq p(x)$, on a alors

$$\|x - y\|^2 = \|x - p(x) + p(x) - y\|^2 = \|x - p(x)\|^2 + \|p(x) - y\|^2 > \|x - p(x)\|^2.$$

et donc $\|x - y\| > \min_{z \in F} \|x - z\|$. □



Signalons enfin

Proposition 13

Si (e_1, \dots, e_k) est une base orthonormée de F , alors

$$p(x) = \sum_{i=1}^k \langle x, e_i \rangle e_i.$$

Preuve

Complétons (e_1, \dots, e_k) en une base (e_1, \dots, e_n) orthonormale de l'espace E .

On sait que $x = \sum_{i=1}^n \langle x, e_i \rangle e_i$, ce que l'on écrit

$$x = \sum_{i=1}^k \langle x, e_i \rangle e_i + \sum_{i=k+1}^n \langle x, e_i \rangle e_i,$$

où $\sum_{i=1}^k \langle x, e_i \rangle e_i \in F$ et $\sum_{i=k+1}^n \langle x, e_i \rangle e_i \in F^\perp$. On en déduit que $p(x) = \sum_{i=1}^k \langle x, e_i \rangle e_i$. □

2.4 Problèmes des moindres carrés

Théorème 11

Soit $A \in \mathcal{M}_{n,p}(\mathbb{R})$ une matrice de rang p et $B \in \mathcal{M}_{n,1}(\mathbb{R})$. Il existe un et un seul vecteur colonne $X \in \mathcal{M}_{p,1}(\mathbb{R})$ qui minimise la norme $\|AX - B\|$.

Ce vecteur est la solution de l'équation $'AAX = 'AB$.

Preuve

Notons f l'application linéaire suivante

$$\begin{cases} \mathcal{M}_{p,1}(\mathbb{R}) & \rightarrow & \mathcal{M}_{n,1}(\mathbb{R}) \\ X & \mapsto & AX \end{cases}$$

On a alors

$$\inf_{X \in \mathcal{M}_{p,1}(\mathbb{R})} \|AX - B\| = \inf_{X \in \mathcal{M}_{p,1}(\mathbb{R})} \|f(X) - B\|.$$

Mais il est clair que

$$\inf_{X \in \mathcal{M}_{p,1}(\mathbb{R})} \|f(X) - B\| = \inf_{Y \in \Im m(f)} \|Y - B\|.$$

On en déduit que pour $X_0 \in \mathcal{M}_{p,1}(\mathbb{R})$, on a

$$\|AX_0 - B\| = \inf_{X \in \mathcal{M}_{p,1}(\mathbb{R})} \|AX - B\| \Leftrightarrow f(X_0) = p_{\Im m(f)}(B).$$

Un tel X_0 existe puisque $p_{\Im m(f)}(B) \in \Im m(f)$. De plus X_0 est unique, car f est injective puisque $\text{rg}(f) = p = \dim(\mathcal{M}_{p,1}(\mathbb{R}))$.

Enfin, on sait que

$$\begin{aligned} f(X_0) = p_{\Im m(f)}(B) &\Leftrightarrow \begin{cases} f(X_0) \in \Im m(f) \\ f(X_0) - B \in \Im m(f)^\perp \end{cases} \\ &\Leftrightarrow f(X_0) - B \in \Im m(f)^\perp \\ &\Leftrightarrow \forall Y \in \Im m(f), \langle AX_0 - B, Y \rangle = 0 \\ &\Leftrightarrow \forall X \in \mathcal{M}_{p,1}(\mathbb{R}), \langle AX_0 - B, AX \rangle = 0 \\ &\Leftrightarrow \forall X \in \mathcal{M}_{p,1}(\mathbb{R}), {}^t(AX)(AX_0 - B) = 0 \\ &\Leftrightarrow \forall X \in \mathcal{M}_{p,1}(\mathbb{R}), {}^tX {}^tA(AX_0 - B) = 0 \\ &\Leftrightarrow {}^tA(AX_0 - B) \in \mathcal{M}_{p,1}(\mathbb{R})^\perp. \end{aligned}$$

Or le vecteur ${}^tA(AX_0 - B)$ est un vecteur de $\mathcal{M}_{p,1}(\mathbb{R})$. La dernière condition équivaut donc à ${}^tA(AX_0 - B) = 0$, soit encore ${}^tAAX_0 = {}^tAB$. □

Exemple

Soient

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 3 \\ -2 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix}.$$

La matrice A est de rang 2 (elle est déjà échelonnée).

Recherchons le vecteur X de $\mathcal{M}_{2,1}(\mathbb{R})$ qui minimise $\|AX - B\|$.

Un calcul donne ${}^tAA = \begin{pmatrix} 5 & -2 \\ -2 & 10 \end{pmatrix}$, le vecteur $X = \begin{pmatrix} x \\ y \end{pmatrix}$ est donc la solution de l'équation

$$\begin{aligned} \begin{pmatrix} 5 & -2 \\ -2 & 10 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 1 & 0 & -2 \\ 0 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix} \Leftrightarrow \begin{cases} 5x - 2y = -3 \\ -2x + 10y = -7 \end{cases} \\ \Leftrightarrow \begin{cases} 5x - 2y = -3 \\ 46y = -41 \end{cases} &L_2 \leftarrow 5L_2 + 2L_1 \Leftrightarrow \begin{cases} 5x - 2y = -3 \\ y = -\frac{41}{46} \end{cases} &L_2 \leftarrow \frac{1}{46}L_2 \\ \Leftrightarrow \begin{cases} x = -\frac{22}{23} \\ y = -\frac{41}{46} \end{cases} &L_1 \leftarrow \frac{1}{5}(L_1 + 2L_2) \end{aligned}$$

Le vecteur recherché est donc $\begin{pmatrix} -\frac{22}{23} \\ -\frac{41}{46} \end{pmatrix}$.

3. Endomorphismes symétriques

3.1 Premières propriétés

Définition 12

Un endomorphisme u d'un espace euclidien E est dit symétrique si, pour tout $(x, y) \in E^2$, on a

$$\langle x, u(y) \rangle = \langle u(x), y \rangle.$$

Proposition 14

Soient $\mathcal{B} = (e_1, \dots, e_n)$ une base de E et u un endomorphisme de E ,

$$u \text{ est symétrique} \Leftrightarrow \forall (i, j) \in \llbracket 1, n \rrbracket^2, \langle u(e_i), e_j \rangle = \langle e_i, u(e_j) \rangle.$$

Preuve

\Rightarrow C'est immédiat.

\Leftarrow Supposons que u vérifie la propriété. Soit $(x, y) \in E^2$, avec

$$x = \sum_{k=1}^n a_k e_k \quad \text{et} \quad y = \sum_{k=1}^n b_k e_k.$$

On a alors

$$\begin{aligned} \langle x, u(y) \rangle &= \left\langle \sum_{i=1}^n a_i e_i, u \left(\sum_{j=1}^n b_j e_j \right) \right\rangle = \left\langle \sum_{i=1}^n a_i e_i, \sum_{j=1}^n b_j u(e_j) \right\rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i b_j \langle e_i, u(e_j) \rangle = \sum_{j=1}^n \sum_{i=1}^n a_i b_j \langle u(e_i), e_j \rangle \\ &= \sum_{j=1}^n b_j \left\langle \sum_{i=1}^n a_i u(e_i), e_j \right\rangle = \left\langle \sum_{i=1}^n a_i u(e_i), \sum_{j=1}^n b_j e_j \right\rangle \\ &= \left\langle u \left(\sum_{i=1}^n a_i e_i \right), \sum_{j=1}^n b_j e_j \right\rangle \\ &= \langle u(x), y \rangle. \end{aligned}$$

□

Théorème 12

Soit u un endomorphisme d'un espace euclidien E .

1. Si u est symétrique, alors la matrice de u dans toute base orthonormée est symétrique.
2. S'il existe une base orthonormée dans laquelle la matrice de u est symétrique, alors u est symétrique.

Preuve

1. Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base orthonormale de E et $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ la matrice de u dans la base \mathcal{B} . On sait que $a_{ij} = \langle e_i, u(e_j) \rangle$. Comme u est symétrique, on a $a_{ij} = \langle e_i, u(e_j) \rangle = \langle u(e_i), e_j \rangle = a_{ji}$. La matrice A est symétrique.
2. Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base orthonormale de E telle que $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ la matrice de u dans la base \mathcal{B} soit symétrique. Comme $a_{ij} = \langle e_i, u(e_j) \rangle$, on en déduit que $a_{ij} = \langle e_i, u(e_j) \rangle = a_{ji} = \langle e_j, u(e_i) \rangle = \langle u(e_i), e_j \rangle$. D'après la proposition 14, u est symétrique. \square

3.2 Réduction d'un endomorphisme symétrique

Théorème 13

Toute matrice symétrique réelle admet une valeur propre réelle.

Preuve

Nous admettons ce résultat, qui fait l'objet de l'exercice 24 du chapitre 7 \square

Théorème 14

Tout endomorphisme symétrique possède au moins une valeur propre réelle.

Preuve

Soit u un endomorphisme symétrique de E et A la matrice de u dans une base orthonormale. On sait que A est symétrique, on en déduit d'après la proposition précédente que A possède une valeur propre réelle. Il en est donc de même de u . \square

Théorème 15

Les sous-espaces propres d'un endomorphisme symétrique sont orthogonaux deux à deux.

Preuve

Soient u un endomorphisme symétrique de E , $x \in E_\lambda$ et $y \in E_\mu$ où λ et μ sont deux réels distincts. On a

$$\langle x, u(y) \rangle = \langle x, \mu y \rangle = \mu \langle x, y \rangle$$

et $\langle u(x), y \rangle = \langle \lambda x, y \rangle = \lambda \langle x, y \rangle$

et comme u est symétrique, on en déduit que $(\lambda - \mu) \langle x, y \rangle = 0$. Mais $\lambda - \mu \neq 0$, donc $\langle x, y \rangle = 0$. Il en résulte que $E_\lambda \perp E_\mu$. \square

Théorème 16

Soit u un endomorphisme d'un espace euclidien, les propositions suivantes sont équivalentes

- (i) u est symétrique.
- (ii) Il existe une base orthonormée dans laquelle la matrice de u est diagonale.

Preuve

(i) \Rightarrow (ii) Raisonnons par récurrence « forte » sur la dimension de l'espace euclidien.

Si $n = 1$, alors la matrice de u dans n'importe quelle base orthonormée de l'espace est diagonale.

Supposons la propriété vraie jusqu'au rang n . Soit E un espace de dimension $n + 1$ et u un endomorphisme symétrique de E .

Comme u est symétrique, il possède au moins une valeur propre réelle λ .

Si $E_\lambda = E$, alors $u = \lambda \text{Id}_E$ et donc la matrice de u est diagonale dans n'importe quelle base orthonormale de E .

Excluons dorénavant ce cas. E_λ étant stable par u , u induit un endomorphisme u_1 de E_λ .

Montrons que E_λ^\perp est lui-même stable par u . Soit $x \in E_\lambda^\perp$, on a pour $y \in E_\lambda$

$$\langle y, u(x) \rangle = \langle u(y), x \rangle = 0$$

puisque $u(y) \in E_\lambda$. Donc $u(x) \in E_\lambda^\perp$. On en déduit que u induit un endomorphisme u_2 de E_λ^\perp . L'endomorphisme u étant symétrique, u_2 est aussi symétrique.

Soit \mathcal{B}_1 une base orthonormale de E_λ , la matrice de u_1 dans la base \mathcal{B}_1 est égale à λI_p (où $p = \dim(E_\lambda)$).

Comme $E = E_\lambda \oplus E_\lambda^\perp$ et $n + 1 > \dim(E_\lambda) > 0$, on a $0 < \dim(E_\lambda^\perp) < n + 1$. D'après l'hypothèse de récurrence, il existe donc une base \mathcal{B}_2 orthonormale de E_λ^\perp telle que la matrice de u_2 dans la base \mathcal{B}_2 soit diagonale. En considérant la famille \mathcal{B} obtenue en juxtaposant les familles \mathcal{B}_1 et \mathcal{B}_2 , on obtient une base orthonormale de E , dans laquelle la matrice de u est diagonale.

(ii) \Rightarrow (i) Si la matrice de u dans une base orthonormale est diagonale, elle est en particulier symétrique. Donc u est lui-même symétrique. \square

3.3 Réduction d'une matrice symétrique**Théorème 17**

Soit A une matrice de $\mathcal{M}_n(\mathbb{R})$, les propositions suivantes sont équivalentes

- (i) A est une matrice symétrique.
- (ii) Il existe une matrice orthogonale P telle que $P^{-1}AP$ soit diagonale.

Preuve

(i) \Rightarrow (ii) Soit u l'endomorphisme canoniquement associé à A . Comme la base canonique de \mathbb{R}^n est orthonormale et que la matrice de u dans cette base est symétrique, puisque c'est la matrice A , on en déduit que u est un endomorphisme symétrique. On sait alors qu'il existe une base orthonormale \mathcal{B} formée de vecteurs propres de u telle que la matrice de u dans la base \mathcal{B} soit diagonale. Notons P la matrice de passage de la base canonique de \mathbb{R}^n vers la base \mathcal{B} . On sait que P est orthogonale, et de plus $P^{-1}AP$ est diagonale, puisque c'est la matrice de u dans la base \mathcal{B} .

(ii) \Rightarrow (i) Soit P une telle matrice. On a alors $P^{-1}AP = {}^tPAP$ qui est diagonale. La transposée d'une matrice diagonale est égale à elle-même, or on a

$${}^t({}^tPAP) = {}^tP {}^tA {}^t({}^tP) = P^{-1} {}^tAP$$

on en déduit que ${}^tPAP = {}^tP {}^tAP$. En multipliant par P à gauche et tP à droite, on obtient $A = {}^tA$. Autrement dit, A est symétrique. \square

Exemple

Reprenons l'exemple de la matrice $A = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$ donnée dans le chapitre précédent.

On a vu que $\text{Sp}(A) = \{2, 5\}$.

Nous avons également montré que $\left(\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right)$ était une base de E_2 et que $\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right)$ était une base de E_5 .

On vérifie sur cet exemple que E_2 et E_5 sont bien orthogonaux puisque

$$\left\langle \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\rangle = -1 + 1 = 0 \quad \text{et} \quad \left\langle \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\rangle = -1 + 1 = 0.$$

La famille $(e_1, e_2, e_3) = \left(\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right)$ est donc une base de $\mathcal{M}_{3,1}(\mathbb{R})$

formée de vecteurs propres de A . Mais cette famille n'est pas orthonormale. Pour déterminer une base orthonormale de vecteurs propres de A on applique le procédé d'orthonormalisation de Schmidt à chacune des bases des sous-espaces propres de A .

On a

$$f_1 = \frac{1}{\|e_1\|} \cdot e_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

et $e_2 - \langle e_2, f_1 \rangle f_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}.$

Soit $f_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$. On a enfin $f_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Une base orthonormale de vecteurs propres de A est donc

$$\left(\frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}, \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right).$$

Théorème 18

Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique réelle. Si (X_1, \dots, X_n) est une base orthonormée de vecteurs propres de A telle que, pour $k \in \llbracket 1, n \rrbracket$, X_k soit associé à la valeur propre λ_k , alors

$$A = \sum_{k=1}^n \lambda_k X_k X_k^t$$

Preuve

Notons $B = \sum_{k=1}^n \lambda_k X_k \text{ } ^t X_k$. Pour $i \in \llbracket 1, n \rrbracket$, nous avons

$$\begin{aligned} BX_i &= \left(\sum_{k=1}^n \lambda_k X_k \text{ } ^t X_k \right) X_i = \sum_{k=1}^n \lambda_k X_k \text{ } ^t X_k X_i = \sum_{k=1}^n \lambda_k X_k \langle X_k, X_i \rangle \\ &= \lambda_i X_i \|X_i\|^2 = \lambda_i X_i, \end{aligned}$$

et $AX_i = \lambda_i X_i$. Autrement dit, les endomorphismes canoniquement associés à A et B sont égaux sur une base. Ils sont donc égaux et $A = B$. \square

3.4 Forme quadratique sur \mathbb{R}^n

Dans cette partie, \mathbb{R}^n est muni de sa structure euclidienne canonique.

Définition 13

Soit q une application de \mathbb{R}^n dans \mathbb{R} . On dit que q est une forme quadratique sur \mathbb{R}^n s'il existe une forme bilinéaire φ sur \mathbb{R}^n telle que

$$\forall x \in \mathbb{R}^n, q(x) = \varphi(x, x).$$

Théorème 19

Soit q une application de \mathbb{R}^n vers \mathbb{R} . On note pour $x \in \mathbb{R}^n$, X le vecteur colonne des coordonnées de x dans la base canonique.

Les propositions suivantes sont équivalentes :

- (i) Il existe un endomorphisme symétrique u de \mathbb{R}^n tel que pour $x \in \mathbb{R}^n$, $q(x) = \langle x, u(x) \rangle$.
- (ii) Il existe une matrice symétrique réelle $A \in \mathcal{M}_n(\mathbb{R})$ telle que pour $x \in \mathbb{R}^n$, $q(x) = \text{ } ^t X A X$.
- (iii) q est une forme quadratique.

La matrice A et l'endomorphisme symétrique u sont alors uniques et u est l'endomorphisme canoniquement associé à A .

On dit que A est la matrice de q ou que q est la forme quadratique associée à la matrice A .

Preuve

(i) \Rightarrow (ii) Soit A la matrice de u dans la base canonique. Comme u est symétrique et que la base canonique est orthonormale, A est symétrique. Par ailleurs, pour $x \in \mathbb{R}^n$, les coordonnées de $u(x)$ dans la base canonique sont données par le vecteur colonne AX . D'après les propriétés connues sur l'expression du produit scalaire, on sait que

$$q(x) = \langle x, u(x) \rangle = \text{ } ^t X A X.$$

(ii) \Rightarrow (i) Soit u l'endomorphisme canoniquement associé à A , qui est symétrique puisque A est symétrique. On sait que pour $x \in \mathbb{R}^n$, $\langle x, u(x) \rangle = \text{ } ^t X A X$, le résultat en découle immédiatement.

(ii) \Rightarrow (iii) Soit φ l'application qui à tout couple $(x, y) \in (\mathbb{R}^n)^2$ associe le réel tXAY . On vérifie sans difficulté que φ est une forme bilinéaire. Or par construction

$$\forall x \in \mathbb{R}^n, \quad q(x) = \varphi(x, x).$$

(iii) \Rightarrow (ii) Soit φ la forme bilinéaire sur \mathbb{R}^n telle que $\forall x \in \mathbb{R}^n, q(x) = \varphi(x, x)$. Notons (e_1, \dots, e_n) la base canonique de \mathbb{R}^n . Par bilinéarité nous avons pour tout vecteur $x = \sum_{k=1}^n x_k e_k$,

$$\begin{aligned} q(x) &= \varphi(x, x) = \varphi\left(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n x_j e_j\right) = \sum_{i=1}^n \sum_{j=1}^n x_i x_j \varphi(e_i, e_j) \\ &= \sum_{i=1}^n x_i^2 \varphi(e_i, e_i) + \sum_{1 \leq i < j \leq n} x_i x_j (\varphi(e_i, e_j) + \varphi(e_j, e_i)). \end{aligned}$$

Soit A la matrice telle que $a_{ij} = \frac{1}{2}(\varphi(e_i, e_j) + \varphi(e_j, e_i))$ pour $(i, j) \in \llbracket 1, n \rrbracket^2$ (en particulier $a_{ii} = \varphi(e_i, e_i)$ pour $i \in \llbracket 1, n \rrbracket$). Cette matrice est symétrique, et pour tout $x \in \mathbb{R}^n$ tel que $x = \sum_{k=1}^n x_k e_k$,

$$\begin{aligned} {}^tXAX &= \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{1 \leq i < j \leq n} x_i x_j (a_{ij} + a_{ji}) \\ &= \sum_{i=1}^n x_i^2 \varphi(e_i, e_i) + \sum_{1 \leq i < j \leq n} x_i x_j (\varphi(e_i, e_j) + \varphi(e_j, e_i)) = q(x). \end{aligned}$$

L'unicité de A découle de la remarque du théorème 7.

De même si u_1 et u_2 sont deux endomorphismes symétriques de \mathbb{R}^n qui conviennent, d'après la démonstration, les matrices A et B de u_1 et u_2 dans la base canonique sont solutions du second point. On en déduit que $A = B$. Autrement dit, $u_1 = u_2$ et l'endomorphisme φ est lui aussi unique.

Enfin la démonstration prouve que u est l'endomorphisme canoniquement associé à A . □

► **Remarque**

On peut donc dire qu'une application q de \mathbb{R}^n vers \mathbb{R} est une forme quadratique sur \mathbb{R}^n si et seulement s'il existe des réels $(a_{ij})_{1 \leq i \leq j \leq n}$ tels que

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n, \quad q(x_1, \dots, x_n) = \sum_{1 \leq i \leq j \leq n} a_{ij} x_i x_j$$

auquel cas la matrice de q s'écrit

$$\begin{pmatrix} a_{1,1} & \frac{a_{1,2}}{2} & \dots & \frac{a_{1,n}}{2} \\ \frac{a_{1,2}}{2} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{a_{n-1,n}}{2} \\ \frac{a_{1,n}}{2} & \dots & \frac{a_{n-1,n}}{2} & a_{n,n} \end{pmatrix}$$

Si l'on considère par exemple l'application q définie sur \mathbb{R}^3 par

$$q(x, y, z) = 2x^2 + z^2 + 2xy + 4yz$$

on reconnaît la forme quadratique associée à la matrice

$$\begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 1 \end{pmatrix}.$$

Proposition 15

Soit q une forme quadratique sur \mathbb{R}^n associée à l'endomorphisme symétrique u .
Soit $\mathcal{B} = (v_1, v_2, \dots, v_n)$ une base orthonormale de vecteurs propres de u associés aux valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_n$.

Si x est un vecteur de \mathbb{R}^n de coordonnées (x_1, \dots, x_n) dans la base \mathcal{B} , alors

$$q(x) = \sum_{k=1}^n \lambda_k x_k^2.$$

Preuve

On a

$$\begin{aligned} q(x) &= \langle x, u(x) \rangle = \left\langle \sum_{k=1}^n x_k v_k, u \left(\sum_{k=1}^n x_k v_k \right) \right\rangle \\ &= \left\langle \sum_{k=1}^n x_k v_k, \sum_{k=1}^n x_k u(v_k) \right\rangle = \left\langle \sum_{k=1}^n x_k v_k, \sum_{k=1}^n x_k \lambda_k v_k \right\rangle \\ &= \sum_{k=1}^n \lambda_k x_k^2 \quad (\text{puisque } (v_1, \dots, v_n) \text{ est orthonormale}). \end{aligned}$$

□

Proposition 16

Soit q une forme quadratique sur \mathbb{R}^n associée à l'endomorphisme symétrique u .

1. Si toutes les valeurs propres de u sont positives, alors pour $x \in \mathbb{R}^n$, $q(x) \geq 0$.
2. Si toutes les valeurs propres de u sont strictement positives, alors pour $x \in \mathbb{R}^n$ non nul, $q(x) > 0$.
3. Si toutes les valeurs propres de u sont négatives, alors pour $x \in \mathbb{R}^n$, $q(x) \leq 0$.
4. Si toutes les valeurs propres de u sont strictement négatives, alors pour $x \in \mathbb{R}^n$ non nul, $q(x) < 0$.
5. Si u a des valeurs propres de signes contraires, q n'est pas de signe constant.

Preuve

On reprend les notations de la proposition précédente.

1. Posons $m = \min_{i \in [1, n]} (\lambda_i)$, on a donc $m \geq 0$. On en déduit que

$$q(x) = \sum_{k=1}^n \lambda_k x_k^2 \geq \sum_{k=1}^n m x_k^2 = m \left(\sum_{k=1}^n x_k^2 \right) \geq 0.$$

2. Cette fois-ci, $m > 0$ et donc

$$q(x) = \sum_{k=1}^n \lambda_k x_k^2 \geq \sum_{k=1}^n m x_k^2 = m \left(\sum_{k=1}^n x_k^2 \right) > 0.$$

3. Posons $M = \max_{i \in [1, n]} (\lambda_i)$, on a donc $M \leq 0$. On en déduit que

$$q(x) = \sum_{k=1}^n \lambda_k x_k^2 \leq \sum_{k=1}^n M x_k^2 = M \left(\sum_{k=1}^n x_k^2 \right) \leq 0.$$

4. Cette fois-ci, $M < 0$ et donc

$$q(x) = \sum_{k=1}^n \lambda_k x_k^2 \leq \sum_{k=1}^n M x_k^2 = M \left(\sum_{k=1}^n x_k^2 \right) < 0.$$

5. Pour $i \in \llbracket 1, n \rrbracket$ on a $q(v_i) = \lambda_i$, le résultat s'ensuit. □

Exemple

Reprenons la forme quadratique q définie sur \mathbb{R}^3 par $q(x, y, z) = 2x^2 + z^2 + 2xy + 4yz$. Pour étudier son signe, nous cherchons les valeurs propres de la matrice $A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 1 \end{pmatrix}$.

Soit $\lambda \in \mathbb{R}$, on triangule $A - \lambda I_3$

$$\begin{aligned} \begin{pmatrix} 2-\lambda & 1 & 0 \\ 1 & -\lambda & 2 \\ 0 & 2 & 1-\lambda \end{pmatrix} &\rightarrow \begin{pmatrix} 1 & -\lambda & 2 \\ 2-\lambda & 1 & 0 \\ 0 & 2 & 1-\lambda \end{pmatrix} & L_1 \leftrightarrow L_2 \\ &\rightarrow \begin{pmatrix} 1 & -\lambda & 2 \\ 0 & -\lambda^2 + 2\lambda + 1 & 2\lambda - 4 \\ 0 & 2 & 1-\lambda \end{pmatrix} & L_2 \leftarrow L_2 + (\lambda - 2)L_1 \\ &\rightarrow \begin{pmatrix} 1 & -\lambda & 2 \\ 0 & 2 & 1-\lambda \\ 0 & -\lambda^2 + 2\lambda + 1 & 2\lambda - 4 \end{pmatrix} & L_2 \leftrightarrow L_3 \\ &\rightarrow \begin{pmatrix} 1 & -\lambda & 2 \\ 0 & 2 & 1-\lambda \\ 0 & 0 & -\lambda^3 + 3\lambda^2 + 3\lambda - 9 \end{pmatrix} & L_3 \leftarrow 2L_3 + (\lambda^2 - 2\lambda - 1)L_2 \end{aligned}$$

Les valeurs propres sont les racines de $-\lambda^3 + 3\lambda^2 + 3\lambda - 9$, or 3 est une racine et

$$-\lambda^3 + 3\lambda^2 + 3\lambda - 9 = (\lambda - 3)(-\lambda^2 + 9) = (\lambda - 3)(\sqrt{3} + \lambda)(\sqrt{3} - \lambda).$$

On en déduit que $\text{Sp}(A) = \{-\sqrt{3}, \sqrt{3}, 3\}$. En particulier, A possède deux valeurs propres de signes opposés. La forme quadratique q change de signe sur \mathbb{R}^3 .

En réalité, il n'est pas nécessaire de rechercher les valeurs propres de la matrice associée à la forme quadratique pour déterminer son signe. Il suffit d'exprimer la forme quadratique comme combinaison linéaire de carrés de formes linéaires indépendantes. Une façon de déterminer une telle décomposition est la méthode de Gauss. Donnons-nous une forme quadratique q sur \mathbb{R}^n s'écrivant

$$q(x) = \sum_{1 \leq i \leq j \leq n} a_{i,j} x_i x_j$$

et donc de matrice $A = \begin{pmatrix} a_{1,1} & \frac{a_{1,2}}{2} & \cdots & \frac{a_{1,n}}{2} \\ \frac{a_{1,2}}{2} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{a_{n-1,n}}{2} \\ \frac{a_{1,n}}{2} & \cdots & \frac{a_{n-1,n}}{2} & a_{n,n} \end{pmatrix}.$

- Supposons que l'un des coefficients diagonaux de A est non nul, par exemple $a_{1,1}$. On peut alors écrire q sous la forme

$$\begin{aligned} q(x_1, \dots, x_n) &= a_{1,1}x_1^2 + x_1 \left(\sum_{2 \leq j \leq n} a_{1,j}x_j \right) + \sum_{2 \leq i < j \leq n} a_{i,j}x_i x_j \\ &= a_{1,1}x_{1,1}^2 + x_1 \varphi_1(x_2, \dots, x_n) + q_1(x_2, \dots, x_n) \end{aligned}$$

où θ_1 est une forme linéaire sur \mathbb{R}^{n-1} et q_1 est une forme quadratique sur \mathbb{R}^{n-1} . En poursuivant, nous obtenons :

$$\begin{aligned} q(x_1, \dots, x_n) &= a_{1,1} \left(x_1 + \frac{\theta_1(x_2, \dots, x_n)}{2a_{1,1}} \right)^2 - \frac{\theta_1(x_2, \dots, x_n)^2}{2a_{1,1}} + q_1(x_2, \dots, x_n) \\ &= a_{1,1}\theta_2(x_1, \dots, x_n)^2 + q_2(x_2, \dots, x_n) \end{aligned}$$

où θ_2 est une forme linéaire sur \mathbb{R}^n et q_2 une forme quadratique sur \mathbb{R}^{n-1} .

- Si tous les coefficients diagonaux sont nuls, supposons par exemple que $a_{1,2} \neq 0$, on peut alors écrire que

$$\begin{aligned} q(x_1, \dots, x_n) &= a_{1,2}x_1x_2 + x_1 \left(\sum_{j=3}^n a_{1,j}x_j \right) + x_2 \left(\sum_{j=3}^n a_{2,j}x_j \right) + \sum_{3 \leq i < j \leq n} a_{i,j}x_i x_j \\ &= a_{1,2}x_1x_2 + x_1\theta_1(x_3, \dots, x_n) + x_2\theta_2(x_3, \dots, x_n) + q_1(x_3, \dots, x_n) \end{aligned}$$

où θ_1, θ_2 sont des formes linéaires sur \mathbb{R}^{n-2} et q_1 une forme quadratique sur \mathbb{R}^{n-2} . On poursuit la transformation en écrivant :

$$\begin{aligned} q(x_1, \dots, x_n) &= a_{1,2} \left(x_1 + \frac{\theta_2(x_3, \dots, x_n)}{a_{1,2}} \right) \left(x_2 + \frac{\theta_1(x_3, \dots, x_n)}{a_{1,2}} \right) \\ &\quad + \left(q_1(x_3, \dots, x_n) - \frac{\theta_1(x_3, \dots, x_n)\theta_2(x_3, \dots, x_n)}{a_{1,2}} \right) \\ &= \frac{a_{1,2}}{4} \left[\left(x_1 + x_2 + \frac{\theta_1 + \theta_2}{a_{1,2}} \right)^2 - \left(x_1 - x_2 + \frac{\theta_2 - \theta_1}{a_{1,2}} \right)^2 \right] \\ &\quad + \left(q_1 - \frac{\theta_1\theta_2}{a_{1,2}} \right) \\ &= \frac{a_{1,2}}{4} (\theta_3(x_1, \dots, x_n)^2 - \theta_4(x_1, \dots, x_n)^2) + q_2(x_3, \dots, x_n) \end{aligned}$$

où θ_3, θ_4 sont des formes linéaires sur \mathbb{R}^n et q_2 une forme quadratique en x_3, \dots, x_n .

Dans tous les cas, nous avons exprimé q à l'aide de carré(s) de forme(s) linéaire(s) et d'une forme quadratique en x_2, \dots, x_n ou x_3, \dots, x_n . En répétant l'opération sur la

forme quadratique restante, de proche en proche, on exprimera donc q comme une combinaison linéaire d'au plus n carrés de formes linéaires. Il est aisé de prouver que les formes linéaires obtenues sont linéairement indépendantes.

À ce stade nous pouvons écrire q sous la forme

$$q = \lambda_1 \theta_1^2 + \lambda_2 \theta_2^2 + \dots + \lambda_k \theta_k^2$$

où les réels $\lambda_1, \dots, \lambda_k$ sont non nuls.

- Si ces réels sont tous positifs, naturellement, la forme quadratique q est-elle même positive. Si de plus $k = n$, alors on sait d'après l'exercice 7 que

$$\dim \left(\bigcap_{i=1}^n \text{Ker}(\varphi_i) \right) = n - n = 0 \quad \text{et donc} \quad \bigcap_{i=1}^n \text{Ker}(\varphi_i) = \{0\}.$$

On en déduit que pour tout n -uplet non nul $x = (x_1, \dots, x_n)$, il existe au moins un $i \in \llbracket 1, n \rrbracket$ tel que $\theta_i(x) \neq 0$ et donc pour tout $x \in \mathbb{R}^n$ non nul, $q(x) > 0$.

On traite d'une façon analogue le cas où ces réels sont tous négatifs.

- S'il existe deux de ces réels de signe contraire, par exemple $\lambda_1 > 0$ et $\lambda_2 < 0$, alors on sait qu'il existe $y, z \in \mathbb{R}^n$ tels que

$$\begin{cases} \theta_2(y) = \theta_3(y) = \dots = \theta_k(y) = 0 \\ \theta_1(y) = 1 \end{cases} \quad \text{et} \quad \begin{cases} \theta_1(z) = \theta_3(z) = \dots = \theta_k(z) = 0 \\ \varphi_2(z) = 1 \end{cases}$$

(puisque l'application $\begin{cases} \mathbb{R}^n & \rightarrow & \mathbb{R}^k \\ x & \mapsto & (\theta_1(x), \dots, \theta_k(x)) \end{cases}$ est surjective). Ces vecteurs y, z ainsi définis sont tels que $q(y) = \lambda_1 > 0$ et $q(z) = \lambda_2 < 0$. Autrement dit q change de signe.

Repartons de la forme quadratique $q(x, y, z) = 2x^2 + z^2 + 2xy + 4yz$. En suivant la méthode de Gauss, nous obtenons :

$$\begin{aligned} q(x, y, z) &= 2 \left(x + \frac{y}{2} \right)^2 - \frac{y^2}{2} + z^2 + 4yz \\ &= 2 \left(x + \frac{y}{2} \right)^2 + (z + 2y)^2 - 4y^2 - \frac{y^2}{2} \\ &= 2 \left(x + \frac{y}{2} \right)^2 + (z + 2y)^2 - \frac{9}{2}y^2 \end{aligned}$$

et l'on retrouve que q n'est pas de signe constant.

1. Montrer qu'il n'existe pas de matrice symétrique $A \in \mathcal{M}_n(\mathbb{R})$ telle que $A^k = 0$ et $A^{k-1} \neq 0$, avec $k \geq 2$.

2. Soient F et G deux sous-espaces vectoriels supplémentaires d'un espace euclidien E . Montrer que $E = F^\perp \oplus G^\perp$.

3. Soit $E = \mathbb{R}_3[X]$ muni de sa structure euclidienne canonique. On pose

$$F = \{P \in E \mid P(1) = 0\}.$$

1. Montrer que F est un sous-espace vectoriel de E .
2. Déterminer une base orthonormale de F .
3. En déduire le projeté orthogonal de X sur F .

4. Soit p un projecteur d'un espace euclidien E . Prouver que p est un projecteur orthogonal si et seulement si p est symétrique.

5. Soit

$$A = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -1 & -1 \end{pmatrix}$$

1. A est-elle diagonalisable ?
2. Déterminer une base orthonormale de vecteurs propres de A .

6. Soit $A \in \mathcal{GL}_n(\mathbb{R})$ et f l'endomorphisme de \mathbb{R}^n canoniquement associé. On pose $B = {}^tAA$ et on note g l'endomorphisme canoniquement associé à B . \mathbb{R}^n est muni de sa structure euclidienne canonique.

1. Montrer que l'application $\begin{cases} \mathbb{R}^n \times \mathbb{R}^n & \rightarrow \mathbb{R} \\ (x, y) & \mapsto \langle x, g(y) \rangle \end{cases}$ est un produit scalaire sur \mathbb{R}^n .
2. Montrer que g est un endomorphisme symétrique de \mathbb{R}^n .
3. En déduire qu'il existe une base orthonormale (e_1, e_2, \dots, e_n) de \mathbb{R}^n telle que la famille $(f(e_1), f(e_2), \dots, f(e_n))$ soit orthogonale.

7. Soit p un projecteur d'un espace euclidien E .

1. Montrer que quels que soient les réels a_1, a_2, \dots, a_n , on a

$$\sum_{k=1}^n |a_k| \leq \sqrt{n} \sqrt{\sum_{k=1}^n a_k^2}.$$

2. En déduire que l'ensemble

$$\left\{ \frac{\|p(x)\|}{\|x\|} \mid x \in E \setminus \{0\} \right\}$$

est majoré.

On pourra travailler dans une base orthonormale de E .

3. On pose $\|p\| = \sup_{x \in E \setminus \{0\}} \frac{\|p(x)\|}{\|x\|}$.

Montrer que si $p \neq 0$, alors $\|p\| \geq 1$.

4. Montrer que p est un projecteur orthogonal si et seulement si $\|p\| = 1$.

8. On se place dans \mathbb{R}^4 muni de sa structure euclidienne canonique. On considère F le plan d'équation $-x + y + z = 0$. Déterminer les matrices A et B des projections orthogonales sur F et F^\perp .

9. (Endomorphismes orthogonaux) Soit f un endomorphisme d'un espace euclidien E . On dit que f est une isométrie, ou que f est un endomorphisme orthogonal, si pour tout $x \in E$, $\|f(x)\| = \|x\|$.

1. Montrer que si f est une isométrie de E , f est bijective et f^{-1} est une isométrie.
2. Un projecteur orthogonal est-il un endomorphisme orthogonal?
3. Soit f une application de E vers E , montrer que

$$f \text{ est un isométrie} \iff \forall (x, y) \in E^2, \langle f(x), f(y) \rangle = \langle x, y \rangle.$$

4. Soit $f \in L(E)$. Montrer que f est une isométrie, si, et seulement si, la matrice de f dans une base orthonormale est orthogonale
5. On suppose maintenant et pour la suite que f est une isométrie de E . Montrer que si F est un sous-espace vectoriel stable par f , F^\perp est également stable par f .
6. Montrer que $f + f^{-1}$ est un endomorphisme symétrique de E .
7. Soit x un vecteur propre de $f + f^{-1}$, montrer que $\text{Vect}(x, f(x))$ est stable par f .
8. Montrer que si E est de dimension 2 et si f n'admet pas de valeur propre, il existe une base orthonormale \mathcal{B} et un réel θ tels que la matrice de f dans la base \mathcal{B} soit

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$
9. En déduire par récurrence sur la dimension de E qu'il existe une base orthonormale de E dans laquelle la matrice de f est de la forme

$$\begin{pmatrix} \varepsilon_1 & & & & & \\ & \ddots & & & & \\ & & \varepsilon_k & & & \\ & & & R_1 & & \\ & & & & \ddots & \\ 0 & & & & & R_p \end{pmatrix}$$

où pour $i \in \llbracket 1, k \rrbracket$, $\varepsilon_i = \pm 1$ et pour $j \in \llbracket 1, p \rrbracket$, R_j est de la forme $\begin{pmatrix} \cos(\theta_j) & -\sin(\theta_j) \\ \sin(\theta_j) & \cos(\theta_j) \end{pmatrix}$.

10. Soit u l'application linéaire de \mathbb{R}^3 dont la matrice dans la base canonique est

$$\frac{1}{3} \begin{pmatrix} 2 & 1 & 2 \\ 1 & 2 & -2 \\ -2 & 2 & 1 \end{pmatrix}.$$

- Montrer que u est une isométrie.
- Déterminer une base orthonormale de \mathbb{R}^3 dans laquelle la matrice de u est de la forme annoncée à la question 8.

10. (Oral ESCP 1999) On se place dans \mathbb{R}^4 muni de sa structure euclidienne canonique. On pose

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}.$$

- Calculer A^2 et B^2 . En déduire les valeurs propres de A et B .
- Déterminer les sous-espaces propres des endomorphismes a et b canoniquement associés à A et B .
- Montrer qu'il existe une base orthonormée de \mathbb{R}^4 dans laquelle les matrices de a et de b sont diagonales.

11. (d'après EML 2002) Soit E un espace vectoriel euclidien de dimension n , dont le produit scalaire est noté $\langle \cdot, \cdot \rangle$.

L'objectif du problème est d'étudier les endomorphismes u de E tels que

$$\forall x \in E, \langle u(x), x \rangle = 0.$$

Les endomorphismes vérifiant cette propriété sont appelés endomorphismes antisymétriques.

Partie I : étude d'un exemple

Dans cette partie, E est l'espace vectoriel des fonctions polynômes à coefficients réels, de degré inférieur ou égal à 2. On rappelle que $(1, X, X^2)$ est une base de E .

On considère l'application $\varphi : E^2 \rightarrow \mathbb{R}$ définie pour tout couple (P, Q) d'éléments de E par :

$$\varphi(P, Q) = P(0)Q(0) + P(1)Q(1) + P(-1)Q(-1).$$

- Vérifier que φ est un produit scalaire.
Dans cette première partie, on considère que E est muni de ce produit scalaire.
- On considère l'endomorphisme u de E défini pour tout P de E par

$$u(P) = 2P'(0)X^2 - (P(1) + P(-1))X.$$

- Vérifier : $\forall P \in E, 2P'(0) - P(1) + P(-1) = 0$.
- En déduire que u est un endomorphisme antisymétrique de l'espace vectoriel euclidien E .

3. Soient $P_1 = \frac{1}{2}(X^2 + X)$ et $P_2 = \frac{1}{2}u(P_1)$.

- a) Vérifier que P_1 est un vecteur propre de u^2 et que la famille (P_1, P_2) est orthonormale.
- b) Déterminer une base de $\text{Ker}(u)$.
- c) Déterminer une base orthonormale \mathcal{B} de E et un nombre réel a tels que la matrice associée à u relativement à cette base soit $\begin{pmatrix} 0 & -a & 0 \\ a & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Partie II : Caractérisations des endomorphismes antisymétriques

Soit u un endomorphisme de E .

4. Pour tout couple (x, y) de E^2 , développer $\langle u(x + y), x + y \rangle$.
 En déduire que u est un endomorphisme antisymétrique si, et seulement si,

$$\forall (x, y) \in E^2, \langle u(x), y \rangle = -\langle x, u(y) \rangle.$$

5. On suppose dans cette question que la dimension n de E est non nulle.
 Soient $\mathcal{B} = (e_1, e_2, \dots, e_n)$ une base orthonormale de E , et $M = (m_{i,j})_{1 \leq i, j \leq n}$ la matrice associée à u relativement à la base \mathcal{B} .
- a) Montrer : $\forall (i, j) \in \llbracket 1, n \rrbracket^2, m_{i,j} = \langle e_i, u(e_j) \rangle$.
 - b) En déduire que u est un endomorphisme antisymétrique si, et seulement si, la matrice M associée à u relativement à la base \mathcal{B} vérifie ${}^tM = -M$.

Partie III : Propriétés générales des endomorphismes antisymétriques

Soit u un endomorphisme antisymétrique non nul de E .

On pourra utiliser la caractérisation obtenue dans la question II.4.

- 6. Soit λ un nombre réel. Montrer que si λ est valeur propre de u , alors $\lambda = 0$.
- 7. Montrer que $\mathfrak{Sm}(u)$ et $\text{Ker}(u)$ sont orthogonaux et supplémentaires dans E .
 En déduire que $\text{Ker}(u) = \text{Ker}(u^2)$.
- 8. Montrer que u^2 est un endomorphisme symétrique de E et que toute valeur propre de u^2 est négative ou nulle.
- 9. a) Montrer que u^2 admet au moins une valeur propre non nulle.
 Soient x un vecteur propre de u^2 associé à une valeur propre non nulle, et F le sous-espace vectoriel de E engendré par $(x, u(x))$.
 b) Montrer que F est un plan vectoriel stable par u .
 c) Montrer que F^\perp , le supplémentaire orthogonal de F , est stable par u .
 d) On munit F^\perp du produit scalaire $\langle \cdot, \cdot \rangle_1$ défini pour tout couple (x, y) d'éléments de F^\perp par

$$\langle x, y \rangle_1 = \langle x, y \rangle.$$

On définit l'endomorphisme u_1 de F^\perp par : $\forall x \in F^\perp, u_1(x) = u(x)$.

Montrer que u_1 est un endomorphisme antisymétrique de F^\perp et que $\mathfrak{Sm}(u) = F \oplus \mathfrak{Sm}(u_1)$.

10. Montrer que le rang d'un endomorphisme antisymétrique est pair. On pourra faire une récurrence sur la dimension de E .

Partie IV : Application

Dans cette partie, E est un espace vectoriel euclidien de dimension 4 et $\mathcal{B} = (e_1, e_2, e_3, e_4)$ est une base orthonormale de E .

Soit u l'endomorphisme de E associé, relativement à la base \mathcal{B} , à la matrice

$$A = \begin{pmatrix} 0 & 4 & 1 & -1 \\ -4 & 0 & -1 & -1 \\ -1 & 1 & 0 & -5 \\ 1 & 1 & 5 & 0 \end{pmatrix}.$$

11. Montrer que u est un endomorphisme antisymétrique de E .
Vérifier que le vecteur $f_1 = e_1 + e_2 - e_3$ est vecteur propre de u^2 .
12. Soit F le sous-espace vectoriel de E engendré par la famille $(f_1, u(f_1))$. Déterminer une base orthonormale de F et une base orthonormale de F^\perp .
13. En déduire une base orthonormale \mathcal{B}_0 de E et deux nombres réels a et b tels que la matrice

associée à u relativement à \mathcal{B}_0 soit
$$\begin{pmatrix} 0 & -a & 0 & 0 \\ a & 0 & 0 & 0 \\ 0 & 0 & 0 & -b \\ 0 & 0 & b & 0 \end{pmatrix}.$$

12. Soit E un espace vectoriel euclidien de dimension n et f, g, h trois endomorphismes de E tels que

$$f \circ g = h, \quad g \circ h = f, \quad h \circ f = g.$$

1. a) Comparer les images de ces trois endomorphismes ainsi que leurs noyaux.
b) Montrer que $f^2 = g^2 = h^2$ et, en calculant $h^2 \circ f \circ h^2$, montrer que $f^5 = f$.
c) Montrer que les noyaux des puissances de f sont tous égaux et en déduire que $\Im f$ et $\text{Ker } f$ sont supplémentaires.
2. On suppose maintenant que f, g, h sont de rang n .
 - a) Quelles sont les valeurs propres possibles de f ?
Montrer que les sous-espaces propres de f (s'il y en a) sont stables par g et par h .
On suppose désormais que f est un endomorphisme symétrique.
 - b) Montrer que $f^2 = I_E$ (l'endomorphisme identité de E) et que f, g, h commutent deux à deux.
 - c) En déduire que f, g, h admettent une base commune de vecteurs propres.

13. (Polynômes de Tchebychev) 1. Soit $P \in \mathbb{R}[X]$. Montrer que l'intégrale

$$\int_{-1}^1 \frac{P(t)}{\sqrt{1-t^2}} dt \text{ converge.}$$

2. Soit $n \in \mathbb{N}^*$. On pose pour $P, Q \in \mathbb{R}_n[X]$,

$$\langle P, Q \rangle = \int_{-1}^1 \frac{P(t)Q(t)}{\sqrt{1-t^2}} dt.$$

Montrer que cette application est un produit scalaire sur $\mathbb{R}_n[X]$.

3. On définit la suite de polynômes (T_k) par

$$\begin{cases} T_0 = 1, & T_1 = X \\ \forall k \in \mathbb{N}, & T_{k+2} = 2XT_{k+1} - T_k. \end{cases}$$

Montrer que pour $k \in \mathbb{N}$, $T_k(\cos(x)) = \cos(kx)$.

4. Montrer que la famille (T_0, T_1, \dots, T_n) est une famille orthogonale de $R_n[X]$.
5. Calculer $\|T_k\|$ pour $k \in \llbracket 0, n \rrbracket$.

14. Soit $E = \mathcal{M}_3(\mathbb{R})$ l'espace vectoriel des matrices carrées $(3, 3)$ à coefficients réels. On définit la trace d'une matrice carrée par la somme de ses termes diagonaux. Si $A \in \mathcal{M}_3(\mathbb{R})$, on la notera $\text{tr}(A)$.

1. On considère l'application φ de $E \times E$ dans \mathbb{R} définie par

$$\varphi(A, B) = \langle A, B \rangle = \text{tr}(A {}^t B).$$

Montrer que φ définit un produit scalaire sur E .

2. **a)** On pose $J = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. Déterminer l'orthogonal de J .
- b)** Soit F le sous-espace vectoriel engendré par J , et F^\perp son orthogonal. Soit $A = (a_{i,j})$ une matrice de E . On note A' la projection orthogonale de A sur F et A'' sa projection orthogonale sur F^\perp . Déterminer A' et A'' .
- c)** Application. Déterminer les projetés orthogonaux sur F et F^\perp de la matrice identité I_3 .

Intégration sur un intervalle quelconque

3

1. Définitions

En première année, on a défini $\int_a^b f(t) dt$, pour une fonction continue ou continue par morceaux sur $[a, b]$. Dans ce chapitre, étant donnée une fonction f continue seulement sur $]a, b[$, on cherche à donner un sens à $\int_a^b f(t) dt$.

1.1 Intégration sur un intervalle semi-ouvert

Définition 1

Soit f une fonction continue sur l'intervalle $[a, b[$ ($-\infty < a < b \leq +\infty$).

On dit que l'intégrale $\int_a^b f(t) dt$ converge ou est convergente si la fonction $x \mapsto \int_a^x f(t) dt$, qui est définie sur $[a, b[$ possède une limite finie en b .

On pose alors $\int_a^b f(t) dt = \lim_{x \rightarrow b} \int_a^x f(t) dt$.

Si l'intégrale n'est pas convergente, on dit qu'elle est divergente.

L'intégrale $\int_a^b f(t) dt$ est dite impropre (ou généralisée). On dit aussi qu'il y a une impropriété en b (ou que l'intégrale est généralisée en b).

► Remarques

- Si f est continue sur $[a, b]$, on retrouve la définition de l'intégrale d'une fonction continue sur le segment $[a, b]$.
- Si b est fini et si f possède un prolongement par continuité en b , l'intégrale $\int_a^b f(t) dt$ converge, car en notant \tilde{f} ce prolongement, on obtient

$$\lim_{x \rightarrow b} \int_a^x f(t) dt = \lim_{x \rightarrow b} \int_a^x \tilde{f}(t) dt = \int_a^b \tilde{f}(t) dt.$$

On dit qu'on a une « fausse impropriété » en b .

- Si $b = +\infty$, l'intégrale $\int_a^b f(t) dt$ est toujours impropre.

Proposition 1

Si f est une fonction continue sur l'intervalle $[a, b[$ ($-\infty < a < b \leq +\infty$), l'intégrale $\int_a^b f(t) dt$ est convergente si, et seulement si, toute primitive F de f possède une limite finie en b et l'on a alors

$$\int_a^b f(t) dt = \lim_{x \rightarrow b} F(x) - F(a).$$

Preuve

C'est évident car, pour tout $x \in [a, b[$, $\int_a^x f(t) dt = F(x) - F(a)$. □

► **Remarques**

- Dans le cas où on connaît une primitive de f , on peut ainsi déterminer la nature convergente ou divergente de l'intégrale et la calculer.
- Il résulte de cette proposition que si $c \in]a, b[$, l'intégrale $\int_a^b f(t) dt$ converge si, et seulement si, $\int_c^b f(t) dt$ converge. La nature de l'intégrale $\int_a^b f(t) dt$ ne dépend donc que du comportement de f au voisinage de l'impropre b .

Exemples

1. La fonction $t \mapsto e^{-t}$ est continue sur $[0, +\infty[$ et a pour primitive la fonction $t \mapsto -e^{-t}$. Comme $\lim_{t \rightarrow +\infty} -e^{-t} = 0$, l'intégrale $\int_0^{+\infty} e^{-t} dt$ converge et $\int_0^{+\infty} e^{-t} dt = \lim_{t \rightarrow +\infty} -e^{-t} + e^0 = 1$.
2. L'intégrale $\int_0^{+\infty} \sin t dt$ diverge, car la fonction $-\cos$, primitive de \sin n'a pas de limite en $+\infty$.

Reste d'une intégrale convergente

Définition 2

Soit f une fonction continue sur l'intervalle $[a, b[$ ($-\infty < a < b \leq +\infty$).

On suppose que l'intégrale $\int_a^b f(t) dt$ est convergente. On appelle reste de cette intégrale impropre l'intégrale $\int_x^b f(t) dt$, où $x \in]a, b[$.

Proposition 2

Soit f une fonction continue sur l'intervalle $[a, b[$ ($-\infty < a < b \leq +\infty$) telle que l'intégrale $\int_a^b f(t) dt$ converge. Le reste $\int_x^b f(t) dt$, où $x \in]a, b[$ a pour limite 0 quand x tend vers b .

Preuve

On note F une primitive de f sur $[a, b[$. On a par définition, pour $x \in [a, b[$,

$$\int_x^b f(t) dt = \lim_b F - F(x).$$

On a donc

$$\lim_{x \rightarrow b} \int_x^b f(t) dt = \lim_b F - \lim_b F = 0. \quad \square$$

On définit de même l'intégrale d'une fonction continue sur $]a, b]$.

Définition 3

Soit f une fonction continue sur l'intervalle $]a, b]$ ($-\infty \leq a < b < +\infty$).

On dit que l'intégrale $\int_a^b f(t) dt$ est convergente si la fonction $x \mapsto \int_x^b f(t) dt$, qui est définie sur $]a, b]$ possède une limite finie en a .

On pose alors $\int_a^b f(t) dt = \lim_{x \rightarrow a} \int_x^b f(t) dt$.

On a comme précédemment :

Proposition 3

Si f est une fonction continue sur l'intervalle $]a, b]$ ($-\infty \leq a < b < +\infty$), l'intégrale $\int_a^b f(t) dt$ est convergente si, et seulement si, toute primitive F de f possède une limite finie en a . En cas de convergence, on a

$$\int_a^b f(t) dt = F(b) - \lim_{x \rightarrow a} F(x).$$

Exemple

Montrons la convergence de $\int_0^1 \ln t dt$, qui est impropre car \ln est continue sur $]0, 1]$, mais n'est pas définie en 0.

On obtient pour $x \in]0, 1[$, en intégrant par parties,

$$\int_x^1 \ln t dt = [t \ln t]_x^1 - \int_x^1 dt = x \ln x - 1 + x.$$

Comme $\lim_{x \rightarrow 0} x \ln x = 0$, on en déduit que $\int_0^1 \ln t dt$ converge et $\int_0^1 \ln t dt = -1$.

► **Remarque**

Comme précédemment, si f est continue sur $]a, b[$, le reste de l'intégrale convergente $\int_a^b f(t) dt$, qui est $\int_a^x f(t) dt$, tend vers 0 quand x tend vers a .

1.2 Intégrale d'une fonction sur un intervalle ouvert

Définition 4

Soit f continue sur $]a, b[$ ($-\infty \leq a < b \leq +\infty$), $c \in]a, b[$. On dit que l'intégrale $\int_a^b f(t) dt$ converge si $\int_a^c f(t) dt$ et $\int_c^b f(t) dt$ convergent et l'on pose, en cas de convergence,

$$\int_a^b f(t) dt = \int_a^c f(t) dt + \int_c^b f(t) dt.$$

Cette définition ne dépend pas du choix du point c comme le montre la proposition suivante.

Proposition 4

Soit f continue sur $]a, b[$ ($-\infty \leq a < b \leq +\infty$). L'intégrale $\int_a^b f(t) dt$ converge si, et seulement si, toute primitive F de f sur $]a, b[$ possède une limite finie en a et b et en cas de convergence, on a

$$\int_a^b f(t) dt = \lim_b F - \lim_a F.$$

Preuve

Cela résulte des propositions 1 et 3. □

Exemple

La fonction $t \mapsto \frac{1}{1+t^2}$ a pour primitive Arctan qui possède une limite finie en $-\infty$ et $+\infty$.

Ainsi $\int_{-\infty}^{+\infty} \frac{1}{1+t^2} dt$ converge et

$$\int_{-\infty}^{+\infty} \frac{1}{1+t^2} dt = \lim_{+\infty} \text{Arctan} - \lim_{-\infty} \text{Arctan} = \frac{\pi}{2} + \frac{\pi}{2} = \pi.$$

Intégrale d'une fonction continue sur un intervalle sauf en un nombre fini de points

Définition 5

Soit $f : [a, b] \rightarrow \mathbb{R}$.

On suppose qu'il existe une subdivision $a_0 = a < a_1 < \dots < a_p = b$ de $[a, b]$ telle que f soit définie et continue sur chaque intervalle $]a_k, a_{k+1}[$ ($0 \leq k \leq p - 1$).

On dit que l'intégrale $\int_a^b f(t) dt$ converge si, pour $0 \leq k \leq p - 1$, $\int_{a_k}^{a_{k+1}} f(t) dt$ converge.

En cas de convergence, on pose

$$\int_a^b f(t) dt = \sum_{k=0}^{p-1} \int_{a_k}^{a_{k+1}} f(t) dt.$$

► Remarques

- Dans la définition précédente, le résultat ne dépend pas de la subdivision choisie. Supposons qu'on ajoute un point c à la subdivision $\sigma = (a_0 = a < a_1 < \dots < a_p = b)$. Soit $q \in \llbracket 0, p - 1 \rrbracket$ un entier tel que $c \in]a_q, a_{q+1}[$.

La fonction f est continue sur $]a_q, c]$ et $[c, a_{q+1}[$. L'intégrale $\int_{a_q}^{a_{q+1}} f(t) dt$ converge si, et seulement si, $\int_{a_q}^c f(t) dt$ et $\int_c^{a_{q+1}} f(t) dt$ convergent et en cas de convergence, on a

$$\int_{a_q}^{a_{q+1}} f(t) dt = \int_{a_q}^c f(t) dt + \int_c^{a_{q+1}} f(t) dt$$

et donc

$$\sum_{k=0}^{p-1} \int_{a_k}^{a_{k+1}} f(t) dt = \sum_{k=0}^{q-1} \int_{a_k}^{a_{k+1}} f(t) dt + \int_{a_q}^c f(t) dt + \int_c^{a_{q+1}} f(t) dt + \sum_{k=q+1}^{p-1} \int_{a_k}^{a_{k+1}} f(t) dt.$$

Il en est de même si on ajoute un nombre fini de points à la subdivision σ . Si l'on a deux subdivisions σ et σ' , on considérant la réunion des deux subdivisions, on voit que le résultat est le même pour σ et σ' .

- Si la fonction f est continue par morceaux sur $[a, b]$ et si $\sigma = (a_0 = a < a_1 < \dots < a_p = b)$ est une subdivision adaptée à f , les intégrales $\int_{a_k}^{a_{k+1}} f(t) dt$ ($0 \leq k \leq p - 1$) convergent car $f|_{]a_k, a_{k+1}[}$ possède un

prolongement par continuité en a_k et a_{k+1} . Ainsi $\int_a^b f(t) dt$ converge et

$$\int_a^b f(t) dt = \sum_{k=0}^{p-1} \int_{a_k}^{a_{k+1}} f(t) dt = \sum_{k=0}^{p-1} \int_{a_k}^{a_{k+1}} f_k(t) dt,$$

où f_k est le prolongement de $f|_{]a_k, a_{k+1}[}$ à $[a_k, a_{k+1}]$. On retrouve l'expression de l'intégrale d'une fonction en escalier sur $[a, b]$.

Exemples

1. Pour démontrer l'existence de $\int_0^{+\infty} \frac{\ln t}{t^2 - 1} dt$, il faut prouver la convergence de

$\int_0^1 \frac{\ln t}{t^2 - 1} dt$ et de $\int_1^{+\infty} \frac{\ln t}{t^2 - 1} dt$, ces deux intégrales étant généralisées aux deux extrémités.

2. L'intégrale $\int_{-1}^1 \frac{1}{t} dt$ n'existe pas. En effet, il y a une impropreté en 0. On a

$$\lim_{x \rightarrow 0} \int_x^1 \frac{1}{t} dt = \lim_{x \rightarrow 0} -(\ln x) = +\infty$$

donc $\int_0^1 \frac{1}{t} dt$ diverge et il en est de même de $\int_{-1}^0 \frac{1}{t} dt$.

1.3 Un exemple fondamental : les intégrales de Riemann

Théorème 1

Soit α un réel quelconque.

Si a est un réel strictement positif, l'intégrale $\int_0^a \frac{1}{t^\alpha} dt$ converge si, et seulement si, $\alpha < 1$;

l'intégrale $\int_a^{+\infty} \frac{1}{t^\alpha} dt$ converge si, et seulement si, $\alpha > 1$.

Plus généralement, si a et b sont des réels tels que $a < b$, $\int_a^b \frac{1}{(t-a)^\alpha} dt$ converge si, et seulement si, $\alpha < 1$.

Preuve

Pour $\alpha \neq 1$, une primitive de $t \mapsto \frac{1}{t^\alpha}$ est la fonction $F : t \mapsto \frac{1}{(1-\alpha)t^{\alpha-1}}$. En $+\infty$, F a une limite finie si, et seulement si, $\alpha - 1 > 0$ soit $\alpha > 1$, donc $\int_a^{+\infty} \frac{1}{t^\alpha} dt$ converge si, et seulement si, $\alpha > 1$. En 0, F a une limite finie si, et seulement si, $\alpha - 1 < 0$ soit $\alpha < 1$, donc $\int_a^{+\infty} \frac{1}{t^\alpha} dt$ converge si, et seulement si, $\alpha > 1$.

Comme la fonction \ln , primitive de $t \mapsto \frac{1}{t}$, a une limite infinie en 0 et en $+\infty$, les intégrales $\int_0^a \frac{1}{t} dt$ et $\int_a^{+\infty} \frac{1}{t} dt$ divergent.

On montre de même qu'une primitive de $t \mapsto \frac{1}{(t-a)^\alpha}$ a une limite finie en a si, et seulement si, $\alpha < 1$. \square

2. Propriétés des intégrales convergentes

Les théorèmes qui suivent sont des versions « intégrales impropres » des propriétés des intégrales sur un segment.

2.1 Linéarité

Théorème 2

Soit f et g deux fonctions définies et continues sur l'intervalle $[a, b]$ ($-\infty \leq a < b \leq +\infty$) sauf en un nombre fini de points, λ et μ deux réels. Si les intégrales $\int_a^b f(t) dt$ et $\int_a^b g(t) dt$

convergent, alors $\int_a^b (\lambda f + \mu g)(t) dt$ converge et

$$\int_a^b (\lambda f + \mu g)(t) dt = \lambda \int_a^b f(t) dt + \mu \int_a^b g(t) dt.$$

Preuve

Soit $a_0 = a < a_1 < \dots < a_p = b$ une subdivision de $[a, b]$ telle que f et g soient définies et continues sur chacun des intervalles $]a_k, a_{k+1}[$. La fonction $\lambda f + \mu g$ est elle aussi continue sur $]a_k, a_{k+1}[$. Par définition, pour tout $k \in [0, p - 1]$, les intégrales $\int_{a_k}^{a_{k+1}} f(t) dt$ et $\int_{a_k}^{a_{k+1}} g(t) dt$ convergent. Des primitives F et G de f et g respectivement sur $]a_k, a_{k+1}[$ ont donc des limites finies en a_k^+ et a_{k+1}^- . La fonction $\lambda F + \mu G$, qui est une primitive de $\lambda f + \mu g$, a donc elle aussi des limites finies en a_k^+ et a_{k+1}^- . D'après les propositions 1 et 3, l'intégrale $\int_{a_k}^{a_{k+1}} (\lambda f + \mu g)(t) dt$ converge et

$$\begin{aligned} \int_{a_k}^{a_{k+1}} (\lambda f + \mu g)(t) dt &= \lim_{a_{k+1}^-} (\lambda F + \mu G) - \lim_{a_k^+} (\lambda F + \mu G) \\ &= \lambda (\lim_{a_{k+1}^-} F - \lim_{a_k^+} F) + \mu (\lim_{a_{k+1}^-} G - \lim_{a_k^+} G) \\ &= \lambda \int_{a_k}^{a_{k+1}} f(t) dt + \mu \int_{a_k}^{a_{k+1}} g(t) dt. \end{aligned}$$

On en déduit que $\int_a^b (\lambda f + \mu g)(t) dt$ converge et que

$$\begin{aligned} \int_a^b (\lambda f + \mu g)(t) dt &= \sum_{k=0}^{p-1} \int_{a_k}^{a_{k+1}} (\lambda f + \mu g)(t) dt \\ &= \sum_{k=0}^{p-1} \left(\lambda \int_{a_k}^{a_{k+1}} f(t) dt + \mu \int_{a_k}^{a_{k+1}} g(t) dt \right) \\ &= \lambda \sum_{k=0}^{p-1} \int_{a_k}^{a_{k+1}} f(t) dt + \mu \sum_{k=0}^{p-1} \int_{a_k}^{a_{k+1}} g(t) dt \\ &= \lambda \int_a^b f(t) dt + \mu \int_a^b g(t) dt. \end{aligned}$$

Corollaire 1

L'ensemble des fonctions définies et continues sur $]a, b[$, sauf en un nombre fini de points, telles que $\int_a^b f(t) dt$ converge, forment un espace vectoriel et l'application $f \mapsto \int_a^b f(t) dt$ est une forme linéaire sur cet espace vectoriel.

Preuve

Le théorème qui précède montre en effet que cet ensemble est un sous-espace de l'ensemble des fonctions de $]a, b]$ dans \mathbb{R} et que l'application $f \mapsto \int_a^b f(t) dt$ est linéaire.

2.2 Relation de Chasles

Proposition 5

Soit f une fonction définie et continue sur un intervalle $[a, b]$, sauf en un nombre fini de points, $c \in]a, b[$. Si $\int_a^b f(t) dt$ converge, il en est de même de $\int_a^c f(t) dt$ et de $\int_c^b f(t) dt$ et

$$\int_a^b f(t) dt = \int_a^c f(t) dt + \int_c^b f(t) dt.$$

Preuve

Soit $a = a_0 < a_1 < \dots < a_p = b$ une subdivision de $[a, b]$ contenant le point c telle que f soit continue sur chaque intervalle $]a_k, a_{k+1}[$ ($0 \leq k \leq p-1$). Par définition, pour tout $k \in [0, p-1]$, l'intégrale $\int_{a_k}^{a_{k+1}} f(t) dt$ converge et

$$\int_a^b f(t) dt = \sum_{k=0}^{p-1} \int_{a_k}^{a_{k+1}} f(t) dt.$$

On suppose que $c = a_q$ où $q \in [0, p]$. Comme $a_0 < a_1 < \dots < a_q$ est une subdivision de $[a, c]$ et que $\int_{a_k}^{a_{k+1}} f(t) dt$ converge pour $k \in [0, q-1]$, par définition l'intégrale $\int_a^c f(t) dt$ converge et

$$\int_a^c f(t) dt = \sum_{k=0}^{q-1} \int_{a_k}^{a_{k+1}} f(t) dt.$$

On trouve de même que $\int_c^b f(t) dt$ converge et

$$\int_c^b f(t) dt = \sum_{k=q}^{p-1} \int_{a_k}^{a_{k+1}} f(t) dt.$$

Cela montre que

$$\int_a^b f(t) dt = \int_a^c f(t) dt + \int_c^b f(t) dt. \quad \square$$

2.3 Positivité

Théorème 3

Soit f une fonction définie et continue sur $]a, b[$ ($-\infty \leq a < b \leq +\infty$).

Si $\int_a^b f(t) dt$ converge et si f est positive sur $]a, b[$, alors on a $\int_a^b f(t) dt \geq 0$. Si de plus f

n'est pas la fonction nulle, on obtient $\int_a^b f(t) dt > 0$.

Preuve

Soit F une primitive de f sur $]a, b[$. Elle possède une limite finie en a et en b car $\int_a^b f(t) dt$ converge. Par hypothèse $F' = f \geq 0$, donc F est croissante. On en déduit que $\lim F \geq \lim_a F$ et donc que $\int_a^b f(t) dt \geq 0$.

Si $\int_a^b f(t) dt = 0$, on a $\lim_b F = \lim_a F$ et la fonction F est constante sur $]a, b[$. Sa dérivée f est nulle. Par contraposition, on obtient que, si f n'est pas la fonction nulle, l'intégrale $\int_a^b f(t) dt$ est strictement positive. \square

Corollaire 2

Soient f et g deux fonctions définies et continues sur $]a, b[$ ($-\infty \leq a < b \leq +\infty$). Si les intégrales $\int_a^b f(t) dt$ et $\int_a^b g(t) dt$ convergent et si $f \leq g$, alors on a

$$\int_a^b f(t) dt \leq \int_a^b g(t) dt.$$

Si de plus f n'est pas égale à g , on obtient

$$\int_a^b f(t) dt < \int_a^b g(t) dt.$$

Preuve

Il résulte du théorème 2 que $\int_a^b (g(t) - f(t)) dt$ converge. La fonction $g - f$ étant positive, on a $\int_a^b (g(t) - f(t)) dt \geq 0$, c'est-à-dire par linéarité de l'intégrale $\int_a^b g(t) dt - \int_a^b f(t) dt \geq 0$, ce qui est le résultat voulu.

Si f n'est pas égale à g , $g - f$ n'est pas nulle. On a donc $\int_a^b (g(t) - f(t)) dt > 0$ et $\int_a^b f(t) dt < \int_a^b g(t) dt$. \square

Le théorème 3 et le corollaire 2 peuvent se généraliser à des fonctions définies et continues sur $]a, b[$ sauf en un nombre fini de points.

Proposition 6

Soit f une fonction définie et continue sur $]a, b[$, sauf un nombre fini de points. Si $\int_a^b f(t) dt$ converge et si f est positive, on a $\int_a^b f(t) dt \geq 0$.

Preuve

Soit $a_0 = a < a_1 < \dots < a_p = b$ une subdivision de $[a, b]$ telle que, pour tout $k \in [0, p - 1]$, f soit continue sur $]a_k, a_{k+1}[$. Soit $k \in [0, p - 1]$. Par hypothèse, f est positive sur $]a_k, a_{k+1}[$ et $\int_{a_k}^{a_{k+1}} f(t) dt$ converge. On déduit du théorème 3 que $\int_{a_k}^{a_{k+1}} f(t) dt \geq 0$ et, en additionnant, on obtient

$$\int_a^b f(t) dt = \sum_{k=0}^{p-1} \int_{a_k}^{a_{k+1}} f(t) dt \geq 0. \quad \square$$

Corollaire 3

Soient f et g des fonctions définies et continues sur $]a, b[$, sauf un nombre fini de points. Si $\int_a^b f(t) dt$ et $\int_a^b g(t) dt$ convergent et si en tout $t \in]a, b[$ où f et g sont définies on a $f(t) \geq g(t)$, on dispose de l'inégalité

$$\int_a^b f(t) dt \leq \int_a^b g(t) dt.$$

Preuve

Il résulte du théorème 2 que $\int_a^b (g(t) - f(t)) dt$ converge. La fonction $g - f$ étant positive sur son ensemble de définition, on a $\int_a^b (g(t) - f(t)) dt \geq 0$, c'est-à-dire par linéarité de l'intégrale $\int_a^b g(t) dt - \int_a^b f(t) dt \geq 0$, ce qui est le résultat voulu. \square

2.4 Intégration par parties

Théorème 4

Soient u et v deux fonctions de classe \mathcal{C}^1 de $]a, b[$ dans \mathbb{R} ($-\infty \leq a < b \leq +\infty$). Si le produit uv possède une limite finie en a et b , les intégrales $\int_a^b u'(t)v(t) dt$ et $\int_a^b u(t)v'(t) dt$ ont même nature et, en cas de convergence, on dispose de l'égalité

$$\int_a^b u'(t)v(t) dt = \lim_b uv - \lim_a uv - \int_a^b u(t)v'(t) dt.$$

Preuve

Supposons que $\int_a^b u(t)v'(t) dt$ converge. Soit F une primitive de uv' sur $]a, b[$. On a $(uv - F)' = u'v + uv' - uv' = u'v$: $uv - F$ est une primitive de $u'v$. Par hypothèse uv et F ont des limites finies en a et b car $\int_a^b u(t)v'(t) dt$ converge, donc $uv - F$ a aussi une limite finie en b et en a . Ainsi, $\int_a^b u'(t)v(t) dt$ converge et

$$\begin{aligned} \int_a^b u'(t)v(t) dt &= \lim_b (uv - F) - \lim_a (uv - F) = \lim_b uv - \lim_a uv - (\lim_b F - \lim_a F) \\ &= \lim_b uv - \lim_a uv - \int_a^b u(t)v'(t) dt. \end{aligned}$$

On montre de même que la convergence de $\int_a^b u'(t)v(t) dt$ implique celle de $\int_a^b u(t)v'(t) dt$ et que

$$\int_a^b u(t)v'(t) dt = \lim_b uv - \lim_a uv - \int_a^b u'(t)v(t) dt. \quad \square$$

Exemple

Considérons, pour $n \in \mathbb{N}$, $I_n = \int_0^1 t(\ln t)^n dt$, intégrale présentant une impropriété en 0. Pour $n \geq 1$, on pose

$$v(t) = (\ln t)^n, \quad u'(t) = t \text{ et donc } u(t) = \frac{1}{2}t^2, \quad v'(t) = \frac{n}{t}(\ln t)^{n-1}.$$

Les fonctions u et v sont de classe \mathcal{C}^1 sur $]0, 1]$ et $\lim_0 uv = \lim_1 uv = 0$. On en déduit que

$I_n = \int_0^1 u'(t)v(t) dt$ a même nature que

$$\int_0^1 u(t)v'(t) dt = \int_0^1 \frac{n}{2}t(\ln t)^{n-1} dt = \frac{n}{2}I_{n-1}$$

et donc même nature que I_{n-1} . En cas de convergence, on obtient la relation

$$I_n = -\frac{n}{2}I_{n-1}.$$

Comme $I_0 = \int_0^1 t \, dt$ est une intégrale convergente, qui vaut $\frac{1}{2}$, une récurrence immédiate montre que I_n converge pour tout entier n et que, pour tout $n \in \mathbb{N}$,

$$I_n = \frac{n!(-1)^n}{2^n}.$$

► Remarque

Si uv n'a pas une limite finie en a et b , les deux intégrales n'ont pas nécessairement même nature. Cela n'empêche pas d'utiliser l'intégration par parties. On intègre par parties sur un intervalle $[x, y]$ avec $a < x < y < b$ et, à la fin du calcul, on fait tendre x vers a et y vers b .

Exemple

Considérons l'intégrale $\int_0^1 \frac{\ln t}{(t+1)^2} \, dt$, impropre en 0, dont la suite du calcul va montrer la convergence.

Les fonctions $u : t \mapsto -\frac{1}{1+t}$ et $v = \ln$ sont de classe \mathcal{C}^1 sur $]0, 1[$. Pour $x \in]0, 1[$, on obtient, en intégrant par parties,

$$\begin{aligned} \int_x^1 \frac{\ln t}{(t+1)^2} \, dt &= \int_x^1 u'(t)v(t) \, dt = \left[-\frac{\ln t}{t+1} \right]_x^1 + \int_x^1 \frac{1}{t(t+1)} \, dt \\ &= \frac{\ln x}{x+1} + \int_x^1 \left(\frac{1}{t} - \frac{1}{t+1} \right) \, dt = \frac{\ln x}{x+1} - \ln x - \ln 2 + \ln(1+x) \\ &= -\frac{x \ln x}{x+1} - \ln 2 + \ln(1+x). \end{aligned}$$

Comme $\lim_{x \rightarrow 0} x \ln x = 0$, on obtient $\lim_{x \rightarrow 0} \int_x^1 \frac{\ln t}{(t+1)^2} \, dt = -\ln 2$. Ainsi, l'intégrale

$\int_0^1 \frac{\ln t}{(t+1)^2} \, dt$ converge et sa valeur est $-\ln 2$. Mais la fonction $t \mapsto \frac{\ln t}{t+1}$ n'a pas de limite finie en 0, ce qui interdit d'appliquer directement le théorème 4.

2.5 Changement de variable

Théorème 5

Soient φ une fonction de classe \mathcal{C}^1 , strictement monotone, réalisant une bijection de $] \alpha, \beta [$ sur $] a, b [$ et $f :] a, b [\rightarrow \mathbb{R}$, continue. Les intégrales $\int_\alpha^\beta f(\varphi(t))\varphi'(t) \, dt$ et $\int_a^b f(t) \, dt$ ont même nature et en cas de convergence,

$$\begin{aligned} \int_a^b f(t) \, dt &= \int_\alpha^\beta f(\varphi(t))\varphi'(t) \, dt \text{ si } \varphi \text{ est croissante} \\ \int_a^b f(t) \, dt &= - \int_\alpha^\beta f(\varphi(t))\varphi'(t) \, dt \text{ si } \varphi \text{ est décroissante.} \end{aligned}$$

Preuve

Soit F une primitive de f sur $]a, b[$. La fonction $F \circ \varphi$ est alors une primitive de $(f \circ \varphi)\varphi'$ sur $]\alpha, \beta[$. Supposons que φ est croissante. On a alors $\lim_{\alpha} \varphi = a$ et $\lim_{\beta} \varphi = b$.

- Si $\int_a^b f(t) dt$ converge, F possède une limite finie en a et b et par le théorème de composition des limites $F \circ \varphi$ possède une limite finie en α et β . Ainsi, l'intégrale $\int_{\alpha}^{\beta} f(\varphi(t))\varphi'(t) dt$ converge et

$$\int_{\alpha}^{\beta} f(\varphi(t))\varphi'(t) dt = \lim_{\beta} (F \circ \varphi) - \lim_{\alpha} (F \circ \varphi) = \lim_b F - \lim_a F = \int_a^b f(t) dt.$$

- Si $\int_{\alpha}^{\beta} f(\varphi(t))\varphi'(t) dt$ converge, la fonction $F \circ \varphi$ possède une limite finie en α et β . En écrivant $F = (F \circ \varphi) \circ \varphi^{-1}$, on montre, toujours par le théorème de composition des limites, que F possède une limite finie en a et b et donc que $\int_a^b f(t) dt$ converge. L'égalité des intégrales se montre comme précédemment.

Le cas où φ est décroissante se traite de la même façon. On a alors $\lim_{\alpha} \varphi = b$ et $\lim_{\beta} \varphi = a$. □

Exemple

Considérons l'intégrale, impropre en -1 et 1 , $I = \int_{-1}^1 \frac{1}{(1+x^2)\sqrt{1-x^2}} dx$. La fonction $\varphi : t \mapsto \sin t$ est une bijection croissante de classe \mathcal{C}^1 de $]-\frac{\pi}{2}, \frac{\pi}{2}[$ sur $] -1, 1[$. Ainsi I a même nature que

$$J = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{(1 + \sin^2 t)\sqrt{1 - \sin^2 t}} \cos t dt = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{1 + \sin^2 t} dt.$$

La fonction $u \mapsto \text{Arctan } u$ est une bijection croissante de \mathbb{R} sur $]-\frac{\pi}{2}, \frac{\pi}{2}[$. On sait que $\sin^2 t = \frac{\tan^2 t}{1 + \tan^2 t}$ et que la dérivée de Arctan est $u \mapsto \frac{1}{1+u^2}$. Ainsi J a même nature que

$$K = \int_{-\infty}^{+\infty} \frac{1}{1 + \frac{u^2}{1+u^2}} \frac{1}{1+u^2} du = \int_{-\infty}^{+\infty} \frac{1}{1+2u^2} du.$$

La fonction $u \mapsto \frac{1}{1+2u^2}$ a pour primitive $u \mapsto \frac{1}{\sqrt{2}} \text{Arctan } u\sqrt{2}$. Cette primitive a pour limite $-\frac{\pi}{2\sqrt{2}}$ et $\frac{\pi}{2\sqrt{2}}$ en $-\infty$ et $+\infty$ respectivement. On en déduit que K converge et que $K = 2 \frac{\pi}{2\sqrt{2}} = \frac{\pi}{\sqrt{2}}$.

Il en résulte que I est convergente et que $I = \frac{\pi}{\sqrt{2}}$.

3. Cas des fonctions positives

Pour la plupart des fonctions, on ne sait pas expliciter de primitive. C'est pourquoi il est utile de disposer de méthodes permettant de montrer la convergence des intégrales impropres sans en calculer la valeur. Cette étude est à rapprocher de l'étude de la convergence d'une série. Dans cette section, on traite du cas où la fonction garde un signe constant au voisinage de l'impropriété.

3.1 Condition nécessaire et suffisante de convergence

Théorème 6

Soit f une fonction continue et positive sur $[a, b[$. L'intégrale $\int_a^b f(t) dt$ converge si, et seulement si, la fonction $x \mapsto \int_a^x f(t) dt$ est majorée sur $[a, b[$.

Preuve

Comme f est positive, la fonction $F : x \mapsto \int_a^x f(t) dt$, dont la dérivée est f , est croissante. Elle possède donc une limite finie ou infinie en b . Cette limite est finie si, et seulement si, F est majorée sur $[a, b[$. \square

► Remarque

On montre de même que si f est continue et positive sur $]a, b]$, l'intégrale $\int_a^b f(t) dt$ converge si, et seulement si, la fonction $x \mapsto \int_x^b f(t) dt$, qui décroît, est majorée sur $]a, b]$.

3.2 Critère de comparaison

Théorème 7

Soient f et g deux fonctions continues sur $[a, b[$. On suppose qu'il existe $c \in [a, b[$ tel que $0 \leq f(t) \leq g(t)$ pour tout $t \in [c, b[$.

Si l'intégrale $\int_a^b g(t) dt$ converge, alors $\int_a^b f(t) dt$ converge.

Si l'intégrale $\int_a^b f(t) dt$ diverge, alors $\int_a^b g(t) dt$ diverge.

Preuve

D'après les remarques suivant la proposition 1, la convergence de $\int_a^b f(t) dt$ (resp. $\int_a^b g(t) dt$) équivaut à celle de $\int_c^b f(t) dt$ (resp. $\int_c^b g(t) dt$).

Si $\int_c^b g(t) dt$ converge, la fonction $x \mapsto \int_c^x g(t) dt$ est majorée. Puisque $0 \leq f \leq g$ sur $[c, b]$, on a, pour tout $x \in [c, b]$, $\int_c^x f(t) dt \leq \int_c^x g(t) dt$. On en déduit que la fonction $x \mapsto \int_c^x f(t) dt$ est majorée et donc que l'intégrale $\int_c^b f(t) dt$ converge.

La deuxième affirmation se déduit de la première par contraposition. □

3.3 Équivalence et convergence

Théorème 8

Soient f et g deux fonctions continues sur $[a, b]$. On suppose qu'au voisinage de b , f est positive et $f(t) \sim g(t)$. Alors les intégrales $\int_a^b f(t) dt$ et $\int_a^b g(t) dt$ ont même nature.

Preuve

Comme f et g sont équivalentes au voisinage de b , g est également positive au voisinage de b et il existe $c \in [a, b]$ tel que $f(t) \geq 0$ et $g(t) \geq 0$ pour tout $t \in [c, b]$. Par définition des fonctions équivalentes, il existe $d \in [a, b]$ et $\varphi : [d, b] \rightarrow \mathbb{R}$ tels que $f(t) = g(t)\varphi(t)$ pour tout $t \in [d, b]$ et $\lim_{t \rightarrow b} \varphi(t) = 1$. Par définition de

la limite, il existe $d' \in [d, b]$ tel que, pour tout $t \in [d', b]$, $\frac{1}{2} \leq \varphi(t) \leq \frac{3}{2}$. Posons $d'' = \max(c, d)$. Pour tout $t \in [d'', b]$, on a

$$0 \leq \frac{1}{2}g(t) \leq f(t) \leq \frac{3}{2}g(t).$$

Si $\int_a^b g(t) dt$ converge, il en est de même de $\int_a^b \frac{3}{2}g(t) dt$. Comme $0 \leq f(t) \leq \frac{3}{2}g(t)$ sur $[d'', b]$, le théorème 7 montre que $\int_a^b f(t) dt$ converge.

Si $\int_a^b f(t) dt$ converge, il en est de même de $\int_a^b 2f(t) dt$. Comme $0 \leq g(t) \leq 2f(t)$ sur $[d'', b]$, le théorème 7 montre que $\int_a^b g(t) dt$ converge. □

3.4 Négligeabilité et convergence

Théorème 9

Soient f et g deux fonctions continues sur $[a, b]$. On suppose qu'au voisinage de b , f et g sont positives et que $f(t) = o(g(t))$.

Si $\int_a^b g(t) dt$ converge, alors $\int_a^b f(t) dt$ converge.

Si $\int_a^b f(t) dt$ diverge, alors $\int_a^b g(t) dt$ diverge.

Preuve

Il existe $c \in [a, b]$ tel que $f(t) \geq 0$ et $g(t) \geq 0$ pour tout $t \in [c, b]$. Par définition de la négligeabilité, il existe $d \in [a, b]$ et $\varepsilon : [d, b] \rightarrow \mathbb{R}$ tels que $f(t) = g(t)\varepsilon(t)$ pour tout $t \in [d, b]$ et $\lim_{t \rightarrow b} \varepsilon(t) = 0$. Par définition de la

limite, il existe $d' \in [d, b[$ tel que, pour tout $t \in [d', b[, |\varepsilon(t)| \leq 1$. Posons $d'' = \max(c, d)$. Pour tout $t \in [d'', b[,$ on a

$$0 \leq f(t) \leq g(t).$$

Le théorème 7 permet de conclure.

La deuxième implication s'en déduit par contraposition. □

► **Remarques**

- On peut énoncer des théorèmes semblables aux théorèmes 7, 8 et 9 pour des intégrales possédant une impropreté en a . Il faut comparer les fonctions au voisinage de a .
- Si dans les théorèmes 8 et 9, on remplace f et g positives par f et g négatives au voisinage de b , le résultat subsiste. En effet, $\int_a^b f(t) dt$ converge si, et seulement si, $\int_a^b (-f)(t) dt$ converge et, d'autre part, $f \underset{b}{\sim} g$ (respectivement $f = o(g)$) implique $(-f) \underset{b}{\sim} (-g)$ (respectivement $-f = o(-g)$).

Les théorèmes précédents sont souvent utilisés pour comparer avec des intégrales de Riemann.

Exemples

1. Posons $I_n = \int_0^{+\infty} \frac{1}{(1+t^2)^n} dt$. Au voisinage de $+\infty$, on a $\frac{1}{(1+t^2)^n} \sim \frac{1}{t^{2n}}$. On en déduit que I_n converge si et seulement si $2n > 1$, i.e. $n \in \mathbb{N}^*$.
2. Considérons l'intégrale $I = \int_0^{+\infty} e^{-t^2} dt$, impropre en $+\infty$. Au voisinage de $+\infty$, on a $e^{-t^2} = o\left(\frac{1}{t^2}\right)$, car $\lim_{t \rightarrow +\infty} t^2 e^{-t^2} = 0$. Comme les fonctions sont positives et que $\int_1^{+\infty} \frac{1}{t^2} dt$ converge, l'intégrale I converge.
3. Posons, pour tout réel α , $I_\alpha = \int_1^2 \frac{1}{|\ln t|^\alpha} dt$ et $J_\alpha = \int_2^{+\infty} \frac{1}{|\ln t|^\alpha} dt$. La fonction $f_\alpha : t \mapsto \frac{1}{|\ln t|^\alpha}$ est continue sur $]1, +\infty[$. Au voisinage de 1, on a $f_\alpha(t) \sim \frac{1}{(t-1)^\alpha}$, donc $\int_1^2 f_\alpha(t) dt$ converge si, et seulement si, $\alpha < 1$.
On sait que $\lim_{t \rightarrow +\infty} \frac{(\ln t)^\alpha}{t} = 0$. On a donc $\frac{1}{t} = o\left(\frac{1}{(\ln t)^\alpha}\right)$ au voisinage de $+\infty$.
Comme $\int_2^{+\infty} \frac{1}{t} dt$ diverge, il en est de même de J_α pour tout réel α . On en déduit que $\int_1^{+\infty} \frac{1}{|\ln t|^\alpha} dt$ diverge pour tout réel α .

3.5 Fonction Γ

Définition 6

La fonction Γ est la fonction définie sur $]0, +\infty[$ par

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt.$$

► **Remarque**

Justifions la définition, c'est-à-dire la convergence de $\int_0^{+\infty} t^{x-1}e^{-t} dt$ pour tout $x > 0$. La fonction $f : t \mapsto t^{x-1}e^{-t}$ est continue sur $]0, +\infty[$.

En 0, on a $f(t) \sim \frac{1}{t^{1-x}}$ donc, par comparaison à une intégrale de Riemann, $\int_0^1 t^{x-1}e^{-t} dt$ converge si $1-x < 1$, soit $x > 0$.

On sait, par croissances comparées, que $\lim_{t \rightarrow +\infty} e^{-t}t^{x+1} = 0$. On en déduit que $f(t) = o\left(\frac{1}{t^2}\right)$ et la convergence de $\int_1^{+\infty} t^{x-1}e^{-t} dt$, pour tout $x > 0$.

Finalement, $\int_0^{+\infty} t^{x-1}e^{-t} dt$ converge pour tout réel $x > 0$.

Proposition 7

On a, pour tout $x > 0$, $\Gamma(x+1) = x\Gamma(x)$ et, pour tout $n \in \mathbb{N}$, $\Gamma(n+1) = n!$.

Preuve

Soit $x > 0$, $\Gamma(x+1) = \int_0^{+\infty} t^x e^{-t} dt$. On intègre par parties en posant

$$u(t) = t^x, \quad v'(t) = e^{-t} \text{ et donc } u'(t) = xt^{x-1}, \quad v(t) = -e^{-t}.$$

Comme $\lim_{t \rightarrow 0} (uv)(t) = \lim_{t \rightarrow 0} -t^x e^{-t} = 0$, car $x > 0$ et $\lim_{t \rightarrow +\infty} (uv)(t) = \lim_{t \rightarrow +\infty} -t^x e^{-t} = 0$, par croissances comparées, on a

$$\Gamma(x+1) = \int_0^{+\infty} u(t)v'(t) dt = - \int_0^{+\infty} u'(t)v(t) dt = \int_0^{+\infty} xt^{x-1}e^{-t} dt = x\Gamma(x).$$

Comme $\Gamma(1) = \int_0^{+\infty} e^{-t} dt = [-e^{-t}]_0^{+\infty} = 1$, une récurrence immédiate montre que $\Gamma(n+1) = n!$ pour tout $n \in \mathbb{N}$. C'est vrai pour $n = 0$ et si $\Gamma(n+1) = n!$, alors

$$\Gamma(n+2) = (n+1)\Gamma(n+1) = (n+1)n! = (n+1)!,$$

ce qui termine la démonstration. □

Le résultat qui suit sera utilisé en probabilités.

Proposition 8 (Intégrale de Gauss)

On dispose de l'égalité

$$\int_0^{+\infty} e^{-t^2} dt = \frac{\sqrt{\pi}}{2}.$$

Preuve

Ce résultat fait l'objet des exercices 8 et ?? □

Corollaire 4

On dispose des égalités $\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}$ et $\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$.

Preuve

La première égalité résulte de la proposition 8 et de la parité de la fonction $t \mapsto e^{-t^2}$. La seconde s'obtient en faisant le changement de variable $u = \frac{t}{\sqrt{2}}$ dans $\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt$. □

3.6 Comparaison série-intégrale

Théorème 10

Soit $a \geq 0$ et f une fonction continue, positive et décroissante sur $[a, +\infty[$. Alors la série de terme général $f(n)$ et l'intégrale $\int_a^{+\infty} f(t) dt$ ont même nature.

Preuve

Posons $n_0 = [a] + 1$, où $[a]$ désigne la partie entière de a . Soit $k \geq n_0$ un entier naturel. Comme f est décroissante, on a, pour tout $t \in [k, k + 1]$, $f(k + 1) \leq f(t) \leq f(k)$. En intégrant ces inégalités sur $[k, k + 1]$, on en déduit

$$f(k + 1) \leq \int_k^{k+1} f(t) dt \leq f(k).$$

Soit $n > n_0$ un entier naturel. En sommant les inégalités obtenues pour k variant de n_0 à $n - 1$, on obtient

$$\sum_{k=n_0}^{n-1} f(k + 1) \leq \sum_{k=n_0}^{n-1} \int_k^{k+1} f(t) dt \leq \sum_{k=n_0}^{n-1} f(k),$$

c'est-à-dire

$$S_n - f(n_0) \leq \int_{n_0}^n f(t) dt \leq S_n - f(n) \leq S_{n-1},$$

où $S_n = \sum_{k=n_0}^n f(k)$ est la n -ième somme partielle de la série de terme général $f(n)$.

- Si la série de terme général $f(n)$ converge, ses sommes partielles sont majorées. Soit M un majorant. On a, pour tout $n \geq n_0$, $\int_{n_0}^n f(t) dt \leq S_{n-1} \leq M$. Puisque f est positive, la fonction $x \mapsto \int_{n_0}^x f(t) dt$ est croissante. On en déduit que, pour tout $x \geq n_0$,

$$\int_{n_0}^x f(t) dt \leq \int_{n_0}^{[x]+1} f(t) dt \leq M.$$

D'après le théorème 7, $\int_{n_0}^{+\infty} f(t) dt$ est convergente et donc $\int_a^{+\infty} f(t) dt$ est convergente.

- Réciproquement, si l'intégrale $\int_a^{+\infty} f(t) dt$ est convergente, il en est de même de $\int_{n_0}^{+\infty} f(t) dt$. La fonction $x \mapsto \int_{n_0}^x f(t) dt$ est majorée (par M) et on a, pour tout $n \geq n_0$,

$$S_n \leq f(n_0) + \int_{n_0}^n f(t) dt \leq f(n_0) + M.$$

La série de terme général $f(n)$ est à termes positifs. Ses sommes partielles sont majorées, donc elle converge. □

Exemple

Pour $\alpha > 0$, la fonction $t \mapsto \frac{1}{t^\alpha}$ est continue, positive et décroissante sur $[1, +\infty[$. On en déduit que la série $\sum \frac{1}{n^\alpha}$ et l'intégrale $\int_1^{+\infty} \frac{1}{t^\alpha} dt$ ont même nature. D'après le théorème 1, la série $\sum \frac{1}{n^\alpha}$ converge donc pour $\alpha > 1$.

► **Remarques**

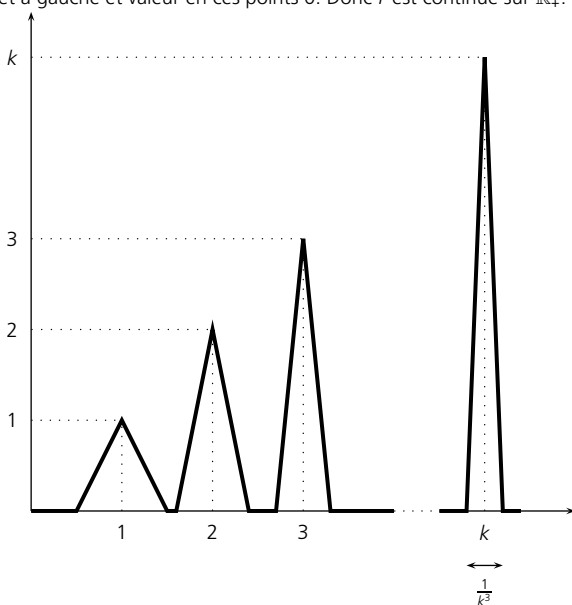
- Soit $a \in \mathbb{R}$ et $f : [a, +\infty[\rightarrow \mathbb{R}$. Même si f est positive, il n'est ni nécessaire, ni suffisant que $\lim_{+\infty} f = 0$ pour que $\int_a^{+\infty} f(t) dt$ converge.

L'exemple de la fonction $f : t \mapsto \frac{1}{t}$ montre que la condition n'est pas suffisante.

Il n'est même pas nécessaire que la fonction f soit bornée. Un exemple de fonction continue, positive, non bornée sur \mathbb{R}_+ telle que $\int_0^{+\infty} f(t) dt$ converge sera fourni par une fonction prenant de façon discontinue des valeurs très grandes sur des intervalles très petits.

Considérons la fonction $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ suivante : s'il existe $k \in \mathbb{N}^*$ tel que $t \in \left[k - \frac{1}{2k^3}, k + \frac{1}{2k^3} \right]$, alors $f(t) = k(-2k^3|t - k| + 1)$; sinon $f(t) = 0$.

La fonction f est continue et affine par morceaux sur chaque intervalle de la forme $\left[k - \frac{1}{2k^3}, k + \frac{1}{2k^3} \right]$; elle vérifie $f\left(k - \frac{1}{2k^3}\right) = f\left(k + \frac{1}{2k^3}\right) = 0$ et $f(k) = k$. Elle est continue, car nulle, sur chaque intervalle de la forme $\left] k + \frac{1}{2k^3}, k + 1 - \frac{1}{2(k+1)^3} \right[$. Enfin, elle est continue en chaque point $k - \frac{1}{2k^3}$ et $k + \frac{1}{2k^3}$, car elle a pour limite à droite et à gauche et valeur en ces points 0. Donc f est continue sur \mathbb{R}_+ .



On a, pour tout $k \in \mathbb{N}^*$,

$$\begin{aligned} \int_{k-\frac{1}{2k^3}}^{k+\frac{1}{2k^3}} f(t) dt &= \int_{k-\frac{1}{2k^3}}^{k+\frac{1}{2k^3}} k(-2k^3|t-k|+1) dt = \int_{-\frac{1}{2k^3}}^{\frac{1}{2k^3}} k(-2k^3|t|+1) dt \\ &= \frac{k}{2k^3} \int_{-1}^1 (-|u|+1) du = \frac{1}{k^2} \int_0^1 (-u+1) du = \frac{1}{2k^2}. \end{aligned}$$

On en déduit que, pour tout réel $x \geq 0$ et $n = \lfloor x \rfloor + 1$,

$$\int_0^x f(t) dt \leq \int_0^{n+\frac{1}{2n^3}} f(t) dt \leq \sum_{k=1}^n \frac{1}{2k^2} \leq \sum_{k=1}^{+\infty} \frac{1}{2k^2},$$

car la série à termes positifs $\sum \frac{1}{2n^2}$ converge.

La fonction $x \mapsto \int_0^x f(t) dt$ est majorée, donc $\int_0^{+\infty} f(t) dt$ est convergente. Pourtant la fonction f n'est pas majorée, puisque, pour tout $k \in \mathbb{N}^*$, $f(k) = k$.

- Par contre, si f possède une limite ℓ en $+\infty$, pour que $\int_a^{+\infty} f(t) dt$ converge, il faut nécessairement que $\ell = 0$. En effet, si $\ell \in]0, +\infty[\cup \{+\infty\}$, il existe $A \geq a$ et $k > 0$ (si $\ell \in \mathbb{R}$, on peut prendre $k = \frac{\ell}{2}$, sinon k quelconque) tels que $f(t) \geq k$ pour $t \geq A$. Comme $\int_a^{+\infty} k dt$ diverge, il en est de même de $\int_a^{+\infty} f(t) dt$. Si $\ell \in]-\infty, 0[\cup \{-\infty\}$, $\int_a^{+\infty} (-f(t)) dt$ diverge d'après ce qui précède et donc $\int_a^{+\infty} f(t) dt$ diverge.

4. Cas des fonctions de signe quelconque

4.1 Convergence absolue

Définition 7

Soit $f :]a, b[\rightarrow \mathbb{R}$ une fonction continue ($-\infty \leq a < b \leq +\infty$). On dit que l'intégrale $\int_a^b f(t) dt$ est absolument convergente si $\int_a^b |f(t)| dt$ est convergente.

Théorème 11

Soit $f :]a, b[\rightarrow \mathbb{R}$ une fonction continue ($-\infty \leq a < b \leq +\infty$). Si $\int_a^b f(t) dt$ est absolument convergente, elle est convergente. De plus, on dispose de l'inégalité

$$\left| \int_a^b f(t) dt \right| \leq \int_a^b |f(t)| dt.$$

Preuve

On pose

$$f_+ = \frac{1}{2}(|f| + f) \text{ et } f_- = \frac{1}{2}(|f| - f).$$

Ces fonctions sont positives et $f_+ + f_- = |f|$. On en déduit que $0 \leq f_+ \leq |f|$ et $0 \leq f_- \leq |f|$. De la convergence de $\int_a^b |f(t)| dt$ découle, d'après le théorème 7, la convergence de $\int_a^b f_+(t) dt$ et de $\int_a^b f_-(t) dt$.

Comme $f = f_+ - f_-$, le théorème 2 montre que $\int_a^b f(t) dt$ est convergente.

On a, pour $a < x < y < b$, $\left| \int_x^y f(t) dt \right| \leq \int_x^y |f(t)| dt$, d'après les propriétés des intégrales des fonctions continues sur un segment. On en déduit, en faisant tendre x vers a et y vers b , l'inégalité voulue. \square

► **Remarques**

- Si la fonction est continue sur $]a, b[$ sauf en nombre fini de points, on peut définir de la même manière l'absolue convergence. Le théorème 11 s'applique encore. Pour le démontrer, on applique le théorème de comparaison au voisinage de chaque impropreté.
- Pour démontrer l'absolue convergence, on peut appliquer les théorèmes 6, 7, 8 et 9 à $|f|$.

Exemple

Pour $\alpha > 1$, l'intégrale $\int_1^{+\infty} \frac{\sin t}{t^\alpha} dt$ est absolument convergente car, pour $t \geq 1$, $\left| \frac{\sin t}{t^\alpha} \right| \leq \frac{1}{t^\alpha}$ et $\int_1^{+\infty} \frac{1}{t^\alpha} dt$ converge.

4.2 Semi-convergence

Définition 8

Soit $f :]a, b[\rightarrow \mathbb{R}$ ($-\infty < a < b < +\infty$). Si l'intégrale $\int_a^b f(t) dt$ est convergente, mais pas absolument convergente, on dit qu'elle est semi-convergente.

Exemple

Considérons l'intégrale $I = \int_0^{+\infty} \frac{\sin t}{t} dt$.

- Montrons qu'elle converge. La fonction $t \mapsto \frac{\sin t}{t}$ est continue sur $]0, +\infty[$. On sait que $\lim_{t \rightarrow 0} \frac{\sin t}{t} = 1$, donc en 0 on a une fausse impropreté.

Pour démontrer la convergence de $\int_1^{+\infty} \frac{\sin t}{t} dt$, on intègre par parties. Pour $x \geq 1$, on obtient

$$\int_1^x \frac{\sin t}{t} dt = \left[-\frac{\cos t}{t} \right]_1^x - \int_1^x \frac{\cos t}{t^2} dt = -\frac{\cos x}{x} + \cos 1 - \int_1^x \frac{\cos t}{t^2} dt.$$

On a, pour $t \geq 1$, $0 \leq \left| \frac{\cos t}{t^2} \right| \leq \frac{1}{t^2}$. Comme $\int_1^{+\infty} \frac{1}{t^2} dt$ converge, il en est de même de $\int_1^{+\infty} \left| \frac{\cos t}{t^2} \right| dt$. Ainsi $\int_1^{+\infty} \frac{\cos t}{t^2} dt$ est absolument convergente et $\int_1^x \frac{\cos t}{t^2} dt$ a une limite finie quand x tend vers $+\infty$. Comme $\frac{\cos x}{x}$ a pour limite 0 quand x tend vers $+\infty$,

on en déduit que $\int_1^x \frac{\sin t}{t} dt$ a une limite finie en $+\infty$, c'est-à-dire que $\int_1^{+\infty} \frac{\sin t}{t} dt$ est convergente. On conclut que $\int_0^{+\infty} \frac{\sin t}{t} dt$ est convergente.

- Montrons que $\int_0^{+\infty} \frac{|\sin t|}{t} dt$ diverge.

Pour $n \in \mathbb{N}^*$, on a

$$\int_0^{n\pi} \frac{|\sin t|}{t} dt = \sum_{k=0}^{n-1} \int_{k\pi}^{(k+1)\pi} \frac{|\sin t|}{t} dt \geq \sum_{k=0}^{n-1} \frac{1}{(k+1)\pi} \int_{k\pi}^{(k+1)\pi} |\sin t| dt,$$

car pour tout $t \in [k\pi, (k+1)\pi]$, $\frac{|\sin t|}{t} \geq \frac{|\sin t|}{(k+1)\pi}$.

Comme $\int_{k\pi}^{(k+1)\pi} |\sin t| dt = \int_0^\pi \sin t dt = 2$, on obtient, avec un changement d'indice,

$$\int_0^{n\pi} \frac{|\sin t|}{t} dt \geq \frac{2}{\pi} \sum_{k=1}^n \frac{1}{k}.$$

Comme la série à termes positifs $\sum \frac{1}{k}$ diverge (série harmonique), on a $\lim_{n \rightarrow +\infty} \sum_{k=1}^n \frac{1}{k} = +\infty$. On en déduit que $\lim_{n \rightarrow +\infty} \int_0^{n\pi} \frac{|\sin t|}{t} dt = +\infty$ et la divergence de $\int_0^{+\infty} \frac{|\sin t|}{t} dt$.

► **Remarque**

Pour des fonctions ne gardant pas un signe constant au voisinage de l'impropreté, les théorèmes de la partie 3. ne s'appliquent pas. En particulier, si f et g sont continues sur $[a, b[$ et équivalentes en b , $\int_a^b f(t) dt$ et $\int_a^b g(t) dt$ n'ont pas nécessairement la même nature.

Exemple

Soit f et g définies sur $[1, +\infty[$ par

$$f(t) = \frac{\sin t}{\sqrt{t}} \text{ et } g(t) = \frac{\sin t}{\sqrt{t}} + \frac{|\sin t|}{t}.$$

Les fonctions f et g sont continues sur $[1, +\infty[$. Montrons qu'elles sont équivalentes en $+\infty$. Soit φ la fonction définie sur $[1, +\infty[$ par

$$\varphi(t) = \begin{cases} 1 + \frac{|\sin t|}{\sqrt{t} \sin t} & \text{si } \sin t \neq 0 \\ 1 & \text{si } \sin t = 0. \end{cases}$$

Elle vérifie $g(t) = f(t)\varphi(t)$ pour tout $t \geq 1$. L'inégalité $|\varphi(t) - 1| \leq \frac{1}{\sqrt{t}}$ pour tout $t \geq 1$ montre que $\lim_{t \rightarrow +\infty} \varphi(t) = 1$. On conclut que $f(t) \underset{t \rightarrow +\infty}{\sim} g(t)$.

Montrons que $\int_1^{+\infty} f(t) dt$ converge en intégrant de nouveau par parties. Pour $x \geq 1$, on obtient

$$\int_1^x f(t) dt = \left[-\frac{\cos t}{\sqrt{t}} \right]_1^x - \int_1^x \frac{\cos t}{2t^{3/2}} dt = -\frac{\cos x}{\sqrt{x}} + \cos 1 - \int_1^x \frac{\cos t}{2t^{3/2}} dt.$$

Comme $\left| \frac{\cos t}{2t^{3/2}} \right| \leq \frac{1}{2t^{3/2}}$, l'intégrale $\int_1^{+\infty} \frac{\cos t}{2t^{3/2}} dt$ converge. Ainsi $\int_1^x \frac{\cos t}{2t^{3/2}} dt$ a une limite finie quand x tend vers $+\infty$. Comme $\lim_{x \rightarrow +\infty} \frac{\cos x}{\sqrt{x}} = 0$, on en déduit que $\int_1^x f(t) dt$ possède une limite finie en $+\infty$ et donc que $\int_1^{+\infty} f(t) dt$ converge.

Si $\int_1^{+\infty} g(t) dt$ convergerait, $\int_1^{+\infty} (g(t) - f(t)) dt = \int_1^{+\infty} \frac{|\sin t|}{t} dt$ convergerait. Il a été démontré dans l'exemple qui suit la définition 4.2 que cette intégrale diverge. Donc $\int_1^{+\infty} g(t) dt$ diverge. Ainsi les deux intégrales n'ont pas même nature, bien que les deux fonctions soient équivalentes en $+\infty$.

1. Étudier la convergence des intégrales suivantes, en discutant éventuellement suivant les valeurs des paramètres α et β réels non nuls :

1. $\int_0^{+\infty} (x - \sqrt{x^2 + 1}) dx;$

2. $\int_0^{+\infty} \left(\frac{\text{Arctan } x}{x} \right)^2 dx;$

3. $\int_0^1 |1 - t^\alpha|^\beta dt;$

4. $\int_0^{+\infty} \frac{\ln x}{x^2 - 1} dx;$

5. $\int_0^{+\infty} \frac{\ln x}{x + e^{-x}} dx;$

6. $\int_0^{+\infty} \frac{x^\alpha}{1 + x^\beta} dx;$

7. $\int_0^{+\infty} \frac{\sin x}{x^\alpha} dx.$

2. (Intégrales de Bertrand) Soient α et β deux réels. Montrer que

$$\int_2^{+\infty} \frac{dx}{x^\alpha (\ln x)^\beta} \text{ converge } \iff \alpha > 1 \text{ ou } (\alpha = 1 \text{ et } \beta > 1).$$

$$\int_0^{\frac{1}{2}} \frac{dx}{x^\alpha (|\ln x|)^\beta} \text{ converge } \iff \alpha < 1 \text{ ou } (\alpha = 1 \text{ et } \beta < 1).$$

3. Calculer les intégrales suivantes, après avoir montré leur convergence ($n \in \mathbb{N}$) :

1. $\int_{-\infty}^{+\infty} \frac{1}{(x^2 + 2x + 3)^n} dx \text{ } (n \geq 1);$

2. $\int_0^{+\infty} x^n e^{-x} dx;$

3. $\int_0^{+\infty} \frac{\ln x}{(x + 1)^2} dx;$

4. $\int_0^{+\infty} \left(1 - x \text{Arctan } \frac{1}{x} \right) dx;$

5. $\int_0^{+\infty} \frac{1}{\sqrt{1 + e^x}} dx;$

4. On pose $I = \int_0^{+\infty} \frac{1}{1+t^4} dt$ et $J = \int_0^{\infty} \frac{t^2}{1+t^4} dt$.

1. Montrer que ces intégrales convergent et que $I = J$.
2. En effectuant le changement de variable $x = t - \frac{1}{t}$ dans l'intégrale $I + J$, calculer la valeur commune de I et J .

5. 1. Montrer que les intégrales suivantes

$$I = \int_0^{\frac{\pi}{2}} \ln \sin x \, dx \quad \text{et} \quad J = \int_0^{\frac{\pi}{2}} \ln \cos x \, dx$$

sont convergentes et égales.

2. En déduire que $I = \frac{1}{2} \int_0^{\frac{\pi}{2}} \ln \left(\frac{\sin 2x}{2} \right)$. En transformant cette dernière intégrale, en déduire la valeur de I .

6. Soit α un réel supérieur ou égal à 1. On pose

$$I = \int_0^{+\infty} e^{-\alpha t} \frac{\sin t}{t} dt \quad \text{et} \quad \forall n \in \mathbb{N} \quad I_n = \int_0^{+\infty} e^{-\alpha t} t^n dt.$$

1. Montrer que l'intégrale I est absolument convergente.
2. Montrer la convergence de I_n pour tout $n \in \mathbb{N}$. Calculer I_n .
3. a) Montrer que, pour $x \geq 0$ et $n \in \mathbb{N}$, on a

$$\left| \sin x - \left(x - \frac{x^3}{6} + \dots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} \right) \right| \leq \frac{x^{2n+2}}{(2n+2)!}.$$

b) En déduire que

$$I = \lim_{n \rightarrow +\infty} \left(\frac{1}{\alpha} - \frac{1}{3\alpha^3} + \dots + (-1)^n \frac{1}{(2n+1)\alpha^{2n+1}} \right).$$

4. a) Montrer que, pour $n \in \mathbb{N}^*$ et $x \in [0, 1]$, on a

$$\left| \sum_{k=0}^n (-1)^k \frac{x^{2k+1}}{2k+1} - \text{Arctan } x \right| \leq \frac{1}{2n+3}.$$

b) En déduire une expression simple de I en fonction de α .

7. On considère l'application $f : x \mapsto \int_x^{+\infty} \frac{\cos t}{t} dt$.

1. Montrer que f est définie et dérivable sur \mathbb{R}_+^* .
2. Montrer que, pour tout $x > 0$, $\left| f(x) + \frac{\sin x}{x} \right| \leq \frac{2}{x^2}$.
3. Montrer que, pour tout $x > 0$,

$$f(x) + \ln x = \int_x^1 \frac{\cos t - 1}{t} dt + \int_1^{+\infty} \frac{\cos t}{t} dt.$$

En déduire que $f(x) \underset{x \rightarrow 0}{\sim} -\ln x$.

4. En intégrant par parties, montrer la convergence et déterminer la valeur de l'intégrale $\int_0^{+\infty} f(x) dx$.

8. Calcul de l'intégrale de Gauss $\int_0^{+\infty} e^{-t^2} dt$ 1. Soit f la fonction définie par $f(x) = \int_0^1 \frac{e^{-x(1+t^2)}}{1+t^2} dt$. En utilisant l'inégalité de Taylor-Lagrange, démontrer que f est dérivable sur \mathbb{R} et que, pour tout réel x ,

$$f'(x) = - \int_0^1 e^{-x(1+t^2)} dt.$$

2. Soit g la fonction définie par $g(x) = f(x^2)$. Montrer que, pour tout réel x ,

$$g'(x) = -2e^{-x^2} \int_0^x e^{-t^2} dt, \text{ puis que } g(x) + \left(\int_0^x e^{-t^2} dt \right)^2 = \frac{\pi}{4}.$$

3. En déduire la valeur de $\int_0^{+\infty} e^{-t^2} dt$, puis celle de $\Gamma\left(\frac{1}{2}\right)$.

9. Soient a et b deux réels strictement positifs et $f \in \mathcal{C}^0(\mathbb{R}_+)$ telle que l'intégrale $\int_1^{+\infty} \frac{f(t)}{t} dt$ converge.

1. Montrer que, pour $\alpha > 0$, on a

$$\int_\alpha^{+\infty} \frac{f(bx) - f(ax)}{x} dx = \int_{ba}^{a\alpha} \frac{f(t)}{t} dt.$$

2. En déduire que l'intégrale $\int_0^{+\infty} \frac{f(bx) - f(ax)}{x} dx$ converge et vaut $f(0) \ln \frac{a}{b}$.
3. Montrer l'existence $\int_0^{+\infty} \frac{e^{-bx} - e^{-ax}}{x} dx$ et donner sa valeur.

4. Dédire de la question précédente la convergence et la valeur de

$$\int_0^1 \frac{t-1}{\ln t} dt.$$

5. Calculer $\int_0^{+\infty} \frac{\sin^3 t}{t^2} dt$, en linéarisant $\sin^3 t$.

10. Calcul de $I = \int_0^{+\infty} \frac{\sin t}{t} dt$ (dont l'existence a été démontrée dans le cours) 1.

Pour $n \in \mathbb{N}$, on pose

$$I_n = \int_0^{\frac{\pi}{2}} \frac{\sin((2n+1)t)}{\sin t} dt.$$

Justifier l'existence de I_n . Montrer que, pour tout $n \in \mathbb{N}^*$,

$$I_n - I_{n-1} = 0$$

(on rappelle que $\sin a - \sin b = 2 \sin \frac{a-b}{2} \cos \frac{a+b}{2}$). Calculer I_n pour tout $n \in \mathbb{N}$.

2. Montrer que, si f est une fonction de classe \mathcal{C}^1 sur l'intervalle $[a, b]$, alors

$J_n = \int_a^b f(t) \sin nt dt$ a pour limite 0 quand n tend vers $+\infty$ (lemme de Lebesgue).

3. Montrer que l'application définie sur $]0, \frac{\pi}{2}[$ par $f(t) = \frac{1}{t} - \frac{1}{\sin t}$ peut être prolongée en une fonction de classe \mathcal{C}^1 sur $[0, \frac{\pi}{2}]$.

4. En déduire que $I = \frac{\pi}{2}$.

11. (D'après ESCP 2002) Soit α un réel strictement supérieur à 1.

1. Montrer que l'intégrale $\int_0^{+\infty} \frac{dt}{(1+t^\alpha)^n}$ converge pour tout $n \in \mathbb{N}^*$.

On pose $u_n(\alpha) = \int_0^{+\infty} \frac{dt}{(1+t^\alpha)^n}$.

2. Montrer, à l'aide d'une intégration par parties, que pour tout entier naturel n non nul, on a

$$u_n(\alpha) = \alpha n (u_n(\alpha) - u_{n+1}(\alpha)).$$

Donner une expression de $u_n(\alpha)$ en fonction de $u_1(\alpha)$.

3. a) Étudier la monotonie de la suite $(u_n(\alpha))$. En déduire sa convergence.

b) Montrer que la série de terme général $\frac{u_n(\alpha)}{\alpha n}$ converge.

c) En déduire la valeur de la limite de la suite $(u_n(\alpha))$.

4. On pose, pour tout entier n non nul, $w_n(\alpha) = \ln(u_n(\alpha)) + \frac{\ln(n)}{\alpha}$.

- a) Démontrer que la série de terme général $(w_{n+1}(\alpha) - w_n(\alpha))$ est convergente (utiliser la formule de la question 2 puis un développement limité).
- b) En déduire l'existence d'un réel $K(\alpha) > 0$ tel que $u_n(\alpha)$ soit équivalent à $\frac{K(\alpha)}{n^{\frac{1}{\alpha}}}$, lorsque n tend vers l'infini.

12.

Soit f une fonction continue sur \mathbb{R}_+ telle que l'intégrale $\int_0^{+\infty} f^2(t) dt$ converge. On lui associe la fonction g définie sur \mathbb{R}_+ par

$$g(x) = \frac{1}{x} \int_0^x f(t) dt \text{ si } x > 0 \text{ et } g(0) = f(0).$$

- 1. Montrer que g est continue sur \mathbb{R}_+ . Exprimer $g'(x)$ en fonction de $f(x)$ et $g(x)$ pour tout $x > 0$.
- 2. Soit $x > 0$.

a) Montrer que

$$\int_0^x g^2(t) dt = -xg^2(x) + 2 \int_0^x f(t)g(t) dt.$$

On commencera par intégrer sur $[\varepsilon, x]$, avec $0 < \varepsilon < x$.

b) En déduire

$$\int_0^x g^2(t) dt \leq 2 \int_0^x f(t)g(t) dt, \text{ puis } \int_0^x g^2(t) dt \leq 4 \int_0^x f^2(t) dt.$$

c) Montrer que l'intégrale $\int_0^{+\infty} g^2(t) dt$ est convergente et que

$$\int_0^{+\infty} g^2(t) dt \leq 4 \int_0^{+\infty} f^2(t) dt.$$

3. Montrer que l'intégrale $\int_0^{+\infty} f(t)g(t) dt$ est absolument convergente et que

$$\int_0^{+\infty} g^2(t) dt = 2 \int_0^{+\infty} f(t)g(t) dt.$$

13.

(D'après ESCP 2002) Pour n entier naturel non nul, on définit l'application f_n de \mathbb{R}^+ dans \mathbb{R} par

$$\forall x \in \mathbb{R}^+, f_n(x) = \frac{1}{(x+1)(x+2)\dots(x+n)}.$$

1. Déterminer l'ensemble A des entiers naturels pour lesquels l'intégrale impropre $\int_0^{+\infty} f_n(x) dx$ est convergente. Lorsque n appartient à A , on pose

$$I_n = \int_0^{+\infty} f_n(x) dx.$$

2. a) À l'aide d'une majoration de $\frac{f_n(x)}{f_2(x)}$, montrer que la suite (I_n) est majorée par la suite de terme général $\frac{B}{n!}$ pour une valeur convenable de B .
 b) Préciser la nature de la série de terme général I_n .
 3. Pour $n \in \mathbb{N}^*$, on pose

$$H_n = \sum_{k=1}^n \frac{1}{k}.$$

- a) Montrer qu'au voisinage de $+\infty$, $H_n \sim \ln n$.
 b) Calculer $\frac{f_n'(x)}{f_n(x)}$ pour $x \in \mathbb{R}^+$ puis encadrer, à l'aide de termes de la suite $(H_k)_{k \in \mathbb{N}^*}$, la quantité $\frac{f_n'(x)}{f_n(x)}$ pour x élément de $[0, 1]$.
 En déduire un encadrement de $f_n(x)$ pour $x \in [0, 1]$.
 c) Montrer qu'au voisinage de $+\infty$,

$$\int_0^1 f_n(x) dx \sim \frac{1}{n! \ln n}.$$

4. Montrer que pour $n \in A$, $\int_0^1 f_n(x) dx = nI_{n+1}$ et en déduire un équivalent simple de I_n .

14.

1. Pour x et y réels strictement positifs, on pose

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt.$$

- a) Montrer la convergence de l'intégrale $B(x, y)$. Justifier que

$$B(x, y) = B(y, x).$$

- b) Déterminer une relation entre $B(x+1, y)$ et $B(x, y+1)$. En déduire que

$$B(x+1, y) = \frac{x}{x+y} B(x, y).$$

- c) Calculer $B(n+1, y)$ pour $y > 0$ et n dans \mathbb{N} .

2. Soit n un entier naturel non nul.

a) Montrer que, pour $t \geq 0$, on a $1 - t \leq e^{-t}$. En déduire que, pour tout $t \in [0, n]$, on a

$$\left(1 - \frac{t}{n}\right)^n \leq e^{-t}.$$

b) Montrer que, pour tout $t \in [0, n]$, on a

$$\left(1 - \frac{t^2}{n}\right) e^{-t} \leq \left(1 - \frac{t}{n}\right)^n.$$

On distinguera les cas $t \in [0, \sqrt{n}[$ et $t \in [\sqrt{n}, n]$.

c) Des questions a et b, déduire un encadrement de $\int_0^n \left(1 - \frac{t}{n}\right)^n t^{x-1} dt$ pour $x > 0$.
Conclure que

$$\lim_{n \rightarrow +\infty} \int_0^n \left(1 - \frac{t}{n}\right)^n t^{x-1} dt = \Gamma(x).$$

d) Exprimer $\int_0^n \left(1 - \frac{t}{n}\right)^n t^{x-1} dt$ en fonction de $B(n+1, x)$. En déduire que

$$\Gamma(x) = \lim_{n \rightarrow +\infty} \frac{n^x n!}{x(x+1) \dots (x+n)}.$$

3. a) Montrer que, quand l'entier n tend vers $+\infty$, on a

$$B(n+1, y) \sim \frac{\Gamma(y)}{n^y}.$$

En déduire que

$$B(x, y) \underset{x \rightarrow +\infty}{\sim} \frac{\Gamma(y)}{x^y}.$$

b) Montrer que, pour x et y strictement positifs,

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

On pourra, en utilisant le résultat de la question 2, montrer

$$\frac{\Gamma(x+y)}{\Gamma(x)} = \lim_{n \rightarrow +\infty} \frac{B(x+n, y)n^y}{B(x, y)}.$$

Éléments de topologie de \mathbb{R}^n

4

Dans ce chapitre et les trois suivants, n désigne un entier naturel non nul. Un élément de \mathbb{R}^n sera noté M ou X , voire x s'il n'y a pas de risque de confondre avec un réel ou, pour préciser, (x_1, \dots, x_n) . Dans les cas $n = 2$ et $n = 3$ on notera plutôt (x, y) ou (x, y, z) . Le vecteur nul de \mathbb{R}^n sera noté 0 ou, s'il y a risque d'ambiguïté, $(0, \dots, 0)$. Ce chapitre généralise au cas de \mathbb{R}^n les définitions données en première année dans le cas particulier $n = 2$.

1. Rappels sur \mathbb{R}^n

1.1 Droites et segments de \mathbb{R}^n

Définition 1

Si A et B sont deux éléments de \mathbb{R}^n , le vecteur \overrightarrow{AB} est l'élément de \mathbb{R}^n égal à $B - A$.

Définition 2

Soient A et U deux éléments de \mathbb{R}^n , U non nul. On appelle droite ou droite affine passant par A et de vecteur directeur U et on note $d_{A,U}$, l'ensemble des éléments M de \mathbb{R}^n tels que le vecteur \overrightarrow{AM} soit proportionnel au vecteur U .

► Remarque

Le terme de droite affine vise à distinguer $d_{A,U}$ d'une droite vectorielle. Notons que si $A = 0$, $d_{A,U}$ est la droite vectorielle engendrée par U .

Proposition 1 (paramétrage)

Soit A et U deux éléments de \mathbb{R}^n , U non nul. La droite $d_{A,U}$ est l'ensemble des éléments M de \mathbb{R}^n qui s'écrivent

$$M = A + tU, \quad t \in \mathbb{R}.$$

L'application $t \mapsto A + tU$ réalise une bijection de \mathbb{R} sur $d_{A,U}$, appelée paramétrage de $d_{A,U}$.

Preuve

Comme U est non nul, les vecteurs \vec{AB} et U sont proportionnels si et seulement s'il existe $t \in \mathbb{R}$ tel que $\vec{AB} = tU$, c'est-à-dire $B - A = tU$ ou encore $B = A + tU$. Puisque U est non nul, (U) est une base de $\text{Vect}(U)$ et t est unique. Ainsi, pour tout $M \in d_{A,U}$, il existe un unique $t \in \mathbb{R}$ tel que $M = A + tU$. L'application $t \mapsto A + tU$ est donc une bijection de \mathbb{R} sur $d_{A,U}$. \square

► Remarque

Si $A = (a_1, \dots, a_n)$ et $U = (u_1, \dots, u_n)$ les composantes de M sont donc $a_1 + tu_1, \dots, a_n + tu_n$.

Définition 3

Soit A et B deux éléments de \mathbb{R}^n .

Le segment $[A, B]$ est l'ensemble des points M tels qu'il existe $t \in [0, 1]$ vérifiant

$$\vec{AM} = t\vec{AB}.$$

Proposition 2

Si A et B sont deux éléments distincts de \mathbb{R}^n , l'application $t \mapsto (1-t)A + tB$ réalise une bijection de $[0, 1]$ sur $[A, B]$.

Preuve

On peut écrire, pour tout $t \in [0, 1]$,

$$(1-t)A + tB = A + t(B-A) = A + t\vec{AB}.$$

D'après la proposition 1, l'application considérée dans l'énoncé est injective et par définition, l'image est en $[A, B]$, donc elle est surjective. \square

1.2 Hyperplan affine

On rappelle qu'un hyperplan de \mathbb{R}^n est un sous-espace vectoriel de dimension $n-1$ (pour $n=2$, c'est une droite vectorielle et pour $n=3$, on dit un plan) et que les hyperplans de \mathbb{R}^n sont les noyaux des formes linéaires, c'est-à-dire des applications linéaires de \mathbb{R}^n dans \mathbb{R} , non nulles. Une forme linéaire de \mathbb{R}^n est de la forme

$$(x_1, \dots, x_n) \mapsto \sum_{k=1}^n u_k x_k,$$

où u_1, u_2, \dots, u_n sont des constantes réelles, ce qui peut encore s'écrire

$$X \mapsto \langle X, U \rangle,$$

où $X = (x_1, \dots, x_n)$ et $U = (u_1, \dots, u_n)$. On en déduit la proposition suivante.

Proposition 3

Soit H une partie de \mathbb{R}^n .

Alors H est un hyperplan si, et seulement si, il existe des réels u_1, u_2, \dots, u_n non tous nuls tels que l'on ait

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n, \quad \left((x_1, \dots, x_n) \in H \iff \sum_{k=1}^n u_k x_k = 0 \right).$$

De manière équivalente H est un hyperplan si et seulement si il existe $U \in \mathbb{R}^n \setminus \{0\}$ tel que

$$\forall X \in \mathbb{R}^n, \quad X \in H \iff \langle X, U \rangle = 0.$$

Preuve

Cela résulte des rappels. Les réels u_1, u_2, \dots, u_n ne sont pas tous nuls et donc U n'est pas nul car la forme

linéaire $(x_1, \dots, x_n) \mapsto \sum_{k=1}^n u_k x_k$ n'est pas nulle. □

Définition 4

Soit A un élément de \mathbb{R}^n , H un hyperplan de \mathbb{R}^n . On appelle hyperplan affine de direction H passant par A l'ensemble des éléments de \mathbb{R}^n qui s'écrivent $A + U$, où U est un élément de H .

► **Remarques**

- Cela signifie que M appartient à l'hyperplan affine de direction H passant par A si, et seulement si, $M - A \in H$.
- Pour $n = 2$, on obtient une droite affine. Pour $n = 3$, on parle de plan affine.

Proposition 4

Soit \mathcal{H} un sous-ensemble de \mathbb{R}^n . La partie \mathcal{H} est un hyperplan affine de \mathbb{R}^n si, et seulement si il existe $(u_1, \dots, u_n) \in \mathbb{R}^n \setminus \{0\}$ et $c \in \mathbb{R}$ tels

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n \quad \left((x_1, \dots, x_n) \in \mathcal{H} \iff u_1 x_1 + \dots + u_n x_n + c = 0 \right).$$

De manière équivalente, \mathcal{H} est un hyperplan affine de \mathbb{R}^n si, et seulement si il existe $U \in \mathbb{R}^n \setminus \{0\}$ et $c \in \mathbb{R}$ tels

$$\forall X \in \mathbb{R}^n, \quad X \in \mathcal{H} \iff \langle X, U \rangle + c = 0.$$

Preuve

- Supposons que \mathcal{H} est l'hyperplan affine de direction H contenant $A = (a_1, \dots, a_n)$. Notons $\sum_{k=1}^n u_k x_k = 0$ une équation de H . Par définition, $M = (x_1, \dots, x_n)$ appartient à \mathcal{H} si, et seulement si, $M - A \in H$, c'est-à-dire si $\sum_{k=1}^n u_k (x_k - a_k) = 0$ ou encore $\sum_{k=1}^n u_k x_k - \sum_{k=1}^n u_k a_k = 0$. C'est le résultat cherché avec $c = -\sum_{k=1}^n u_k a_k$.
- Supposons réciproquement que \mathcal{H} a pour équation $\sum_{k=1}^n u_k x_k + c = 0$. Comme les u_k ne sont pas tous nuls, \mathcal{H} n'est pas vide : si $u_i \neq 0$, il contient le point $A = (a_1, \dots, a_n)$ défini par $a_i = -\frac{c}{u_i}$ et $a_j = 0$ si

$j \neq i$. On a alors $M = (x_1, \dots, x_n) \in \mathcal{H}$ si, et seulement si, $\sum_{k=1}^n u_k x_k = -c = \sum_{k=1}^n u_k a_k$, c'est-à-dire si $\sum_{k=1}^n a_k(x_k - a_k) = 0$. En notant H l'hyperplan vectoriel d'équation $\sum_{k=1}^n u_k x_k = 0$, M appartient à \mathcal{H} si, et seulement si, $M - A$ appartient à H et \mathcal{H} est l'hyperplan affine de direction H passant par A . \square

► **Remarque**

De l'équation de l'hyperplan affine, $\sum_{k=1}^n u_k x_k + c = 0$, on peut déduire une équation de sa direction $\sum_{k=1}^n u_k x_k = 0$.

Réciproquement, si $\sum_{k=1}^n u_k x_k = 0$ est une équation de la direction, l'hyperplan possède une équation de la forme

$$\sum_{k=1}^n u_k x_k + c = 0, \text{ où } c \in \mathbb{R}.$$

Exemple

Cherchons une équation du plan affine \mathcal{H} de \mathbb{R}^3 de direction $H = \text{Vect}(U, V)$, où $U = (1, 1, 0)$ et $V = (1, 0, 1)$, passant par $A = (-1, 1, -1)$. Si une équation de H est $ux + vy + wz = 0$, on doit avoir $u + v = 0$ et $u + w = 0$ et donc $v = w = -u$. Une équation de H est donc $x - y - z = 0$. Il existe $c \in \mathbb{R}$ tel que \mathcal{H} ait une équation de la forme $x - y - z + c = 0$. Comme $A \in \mathcal{H}$, $-1 - 1 + 1 + c = 0$, $c = 1$ et l'équation cherchée est $x - y - z + 1 = 0$.

2. Distance euclidienne

2.1 Structure euclidienne de \mathbb{R}^n

Rappel : l'application qui, à deux éléments $U = (u_1, \dots, u_n)$ et $V = (v_1, \dots, v_n)$ de \mathbb{R}^n , associe le réel $u_1 v_1 + \dots + u_n v_n$, noté $\langle U, V \rangle$ est un produit scalaire sur \mathbb{R}^n qui munit \mathbb{R}^n de sa structure euclidienne canonique. En particulier, la norme canonique du vecteur U est définie par

$$\|U\| = \sqrt{\langle U, U \rangle} = \sqrt{\sum_{i=1}^n x_i^2}.$$

Proposition 5

Pour tout $X = (x_1, \dots, x_n) \in \mathbb{R}^n$, on dispose des inégalités

$$\sup_{i \in [1, n]} |x_i| \leq \|X\| \leq \sqrt{n} \sup_{i \in [1, n]} |x_i|.$$

Preuve

Soit $j \in [1, n]$ tel que $|x_j| = \sup_{i \in [1, n]} |x_i|$. On a clairement

$$\|X\| = \sqrt{\sum_{i=1}^n x_i^2} \geq \sqrt{x_j^2} = |x_j| \text{ et}$$

$$\|X\| = \sqrt{\sum_{i=1}^n x_i^2} \leq \sqrt{\sum_{i=1}^n x_j^2} \leq \sqrt{nx_j^2} = \sqrt{n}|x_j|.$$

\square

2.2 Distance euclidienne

Définition 5

Pour A et B éléments de \mathbb{R}^n , on appelle distance euclidienne entre A et B et on note $d(A, B)$ le réel

$$\|\vec{AB}\| = \|B - A\|.$$

Si $A = (a_1, \dots, a_n)$ et $B = (b_1, \dots, b_n)$, on a donc

$$d(A, B) = \sqrt{(a_1 - b_1)^2 + \dots + (a_n - b_n)^2}.$$

Proposition 6

La distance euclidienne vérifie les propriétés suivantes :

1. $\forall (A, B) \in (\mathbb{R}^n)^2, \quad d(A, B) = 0 \iff A = B$;
2. $\forall (A, B) \in (\mathbb{R}^n)^2, \quad d(A, B) = d(B, A)$;
3. $\forall (A, B, C) \in (\mathbb{R}^n)^3, \quad d(A, C) \leq d(A, B) + d(B, C)$.

Preuve

Ces propriétés résultent immédiatement de celles de la norme. Pour $(A, B, C) \in (\mathbb{R}^n)^3$, on a :

$$d(A, B) = 0 \iff \|B - A\| = 0 \iff B - A = (0, \dots, 0) \iff A = B;$$

$$d(A, B) = \|B - A\| = \|-(A - B)\| = \|A - B\| = d(B, A);$$

$$d(A, C) = \|C - A\| = \|C - B + B - A\| \leq \|C - B\| + \|B - A\| \leq d(B, C) + d(A, B). \quad \square$$

Corollaire 1

On a, pour $(A, B, C) \in (\mathbb{R}^n)^3, \quad |d(A, B) - d(B, C)| \leq d(A, C)$.

Preuve

On a, par l'inégalité triangulaire,

$$d(A, B) \leq d(A, C) + d(B, C) \text{ et donc } d(A, B) - d(B, C) \leq d(A, C).$$

On obtient de même $d(B, C) - d(A, B) \leq d(A, C)$, d'où l'on déduit

$$|d(A, B) - d(B, C)| \leq d(A, C). \quad \square$$

Proposition 7

Si $A = (a_1, \dots, a_n)$ et $B = (b_1, \dots, b_n)$ on a

$$\max_{1 \leq i \leq n} |b_i - a_i| \leq d(A, B) \leq \sqrt{n} \max_{1 \leq i \leq n} |b_i - a_i|.$$

Preuve

Cette proposition découle clairement des inégalités démontrées dans la proposition 5. On prend $X = B - A$. \square

2.3 Boules

Définition 6

Soit A un élément de \mathbb{R}^n .

On appelle boule ouverte de centre A et de rayon r (r réel strictement positif) et l'on note $B(A, r)$ l'ensemble des éléments M de \mathbb{R}^n tels que

$$d(A, M) < r.$$

On appelle boule fermée de centre A et de rayon r (r réel positif ou nul) et on note $B_f(A, r)$ l'ensemble des points M de \mathbb{R}^n tels que

$$d(A, M) \leq r.$$

► Remarques

- Si $r = 0$, la boule fermée $B_f(A, r)$ est réduite à $\{A\}$.
- Si $n = 2$, les boules sont des disques; si $n = 3$, on trouve des boules au sens classique du terme.
- Soient $A = (a_1, \dots, a_n)$ et $M = (x_1, \dots, x_n)$. Il résulte de la proposition 7 que si M appartient à $B_f(A, r)$, alors

$$\max_{1 \leq i \leq n} |x_i - a_i| \leq r.$$
 Ceci signifie que toute boule de centre A est incluse dans un hypercube de centre A .
 De la même proposition, on déduit que si $\max_{1 \leq i \leq n} |x_i - a_i| \leq \frac{r}{\sqrt{n}}$, alors M appartient à $B_f(A, r)$.

2.4 Parties bornées

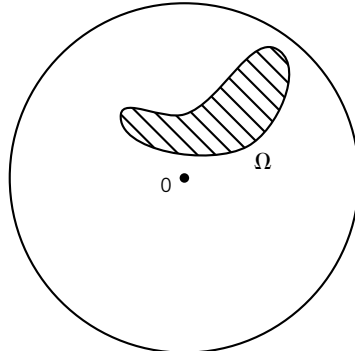
Définition 7

Une partie Ω de \mathbb{R}^n est dite bornée s'il existe un réel K tel que

$$\forall M \in \Omega \quad \|M\| \leq K.$$

► Remarque

L'inégalité $\|M\| \leq K$ peut s'écrire $d(0, M) \leq K$. Une partie de \mathbb{R}^n est donc bornée si elle est contenue dans une boule fermée de centre 0.



Exemples

1. Toute boule de \mathbb{R}^n est bornée.
En effet, en utilisant l'inégalité triangulaire, on obtient, pour tout point M de $B_f(A, r)$,

$$d(0, M) \leq d(0, A) + d(A, M) \leq d(0, A) + r \leq \|A\| + r,$$

ce qui montre que $B_f(A, r) \subset B_f(0, \|A\| + r)$ et a fortiori $B(A, r) \subset B_f(0, \|A\| + r)$.

2. Tout segment $[A, B]$ de \mathbb{R}^n est borné.
En effet, pour tout $t \in [0, 1]$, on a

$$\|(1-t)A + tB\| \leq (1-t)\|A\| + t\|B\| \leq \|A\| + \|B\|.$$

3. Si (a_1, \dots, a_n) et (b_1, \dots, b_n) sont deux n -listes telles que pour tout i , $a_i \leq b_i$, le pavé $\mathcal{R} = [a_1, b_1] \times \dots \times [a_n, b_n]$ de \mathbb{R}^n est borné car si $M = (x_1, \dots, x_n) \in \mathcal{R}$, alors, pour $1 \leq i \leq n$, $|x_i| \leq \max(|a_i|, |b_i|)$, donc

$$\|M\| \leq \sqrt{n} \max_{1 \leq i \leq n} \max(|a_i|, |b_i|).$$

4. La réunion de deux parties bornées et plus généralement d'un nombre fini de parties bornées de \mathbb{R}^n est bornée.

Soit $p \in \mathbb{N}^*$, $\Omega_1, \dots, \Omega_p$ des parties bornées de \mathbb{R}^n . Pour tout $i \in \llbracket 1, p \rrbracket$, il existe $K_i \in \mathbb{R}$ tel que $\|M\| \leq K_i$ pour tout M de Ω_i . On a, pour tout M de $\bigcup_{i=1}^p \Omega_i$, $\|M\| \leq \max(K_1, \dots, K_p)$.

3. Ouverts et fermés

3.1 Parties ouvertes

Définition 8

On dit qu'une partie Ω de \mathbb{R}^n est une partie ouverte de \mathbb{R}^n ou un ouvert de \mathbb{R}^n si $\Omega = \emptyset$ ou si, pour tout $M_0 \in \Omega$, il existe $r > 0$ tel que $B(M_0, r) \subset \Omega$.

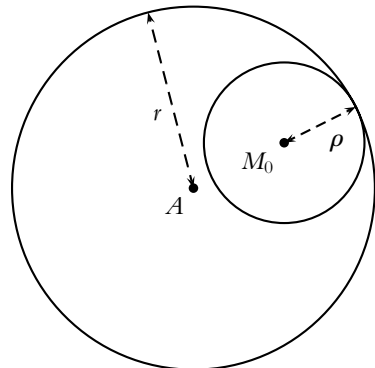
Exemples

1. L'ensemble \mathbb{R}^n est ouvert.
2. Toute boule ouverte $B(A, r)$ est un ouvert de \mathbb{R}^n .

Soit M_0 dans $B(A, r)$. On a $d(A, M_0) < r$ et $d(A, M) \leq d(A, M_0) + d(M_0, M)$ pour tout M de \mathbb{R}^n , par l'inégalité triangulaire. On en déduit que

$$d(M_0, M) < r - d(A, M_0) \implies d(A, M) < r.$$

Le réel $\rho = r - d(A, M_0)$ est strictement positif et $B(M_0, \rho) \subset B(A, r)$.



3. Si $(u_1, \dots, u_n) \in \mathbb{R}^n \setminus \{0\}$ et $c \in \mathbb{R}$, l'ensemble \mathcal{E} des éléments $M = (x_1, \dots, x_n)$ de \mathbb{R}^n tels que

$$u_1x_1 + \dots + u_nx_n + c > 0$$

est un ouvert. Cet ensemble est un demi-espace (ouvert) limité par l'hyperplan d'équation $u_1x_1 + \dots + u_nx_n + c = 0$ (pour $n = 2$, on dit un demi-plan). En notant $U = (u_1, \dots, u_n)$, M appartient à \mathcal{E} si et seulement si $\langle U, M \rangle + c > 0$. Si $M_0 \in \mathcal{E}$, on a, pour tout $M \in \mathbb{R}^n$,

$$\langle U, M \rangle + c = \langle U, M_0 \rangle + c + \langle U, M - M_0 \rangle,$$

par linéarité du produit scalaire, et $M \in \mathcal{E}$ si $|\langle U, M - M_0 \rangle| < \langle U, M_0 \rangle + c$. Comme $|\langle U, M - M_0 \rangle| \leq \|U\| \|M - M_0\|$, cela est réalisé si

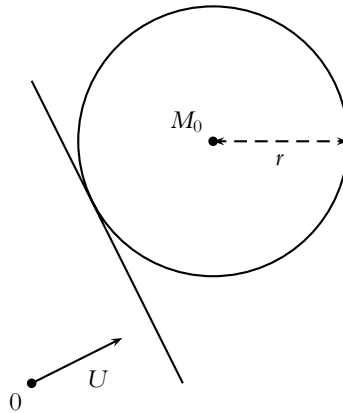
$$d(M_0, M) = \|M - M_0\| < \frac{\langle U, M_0 \rangle + c}{\|U\|}.$$

L'ensemble \mathcal{E} contient donc la boule ouverte de centre M_0 et de rayon $r = \frac{\langle U, M_0 \rangle + c}{\|U\|}$.

L'ensemble des éléments de \mathbb{R}^n qui vérifient $u_1x_1 + \dots + u_nx_n + c < 0$ est également ouvert. Pour le voir, il suffit d'écrire son équation

$$(-u_1)x_1 + \dots + (-u_n)x_n + (-c) > 0.$$

Pour $n = 2$ nous avons l'illustration suivante :



4. Si (a_1, \dots, a_n) et (b_1, \dots, b_n) sont deux n -listes telles que pour tout i , $a_i < b_i$, le pavé (ouvert) $\mathcal{R} =]a_1, b_1[\times \dots \times]a_n, b_n[$ de \mathbb{R}^n est une partie ouverte de \mathbb{R}^n . En effet, soit $M_0 = (u_1, \dots, u_n) \in \mathcal{R}$, $r = \min_{1 \leq i \leq n} \min(b_i - u_i, u_i - a_i)$.

Si $M = (x_1, \dots, x_n) \in B(M_0, r)$, on a, pour $1 \leq i \leq n$,

$$|x_i - u_i| \leq d(M_0, M) < r \leq \min(b_i - u_i, u_i - a_i)$$

et donc $a_i < x_i < b_i$, ce qui montre que

$$B(M_0, r) \subset \mathcal{R}.$$

Proposition 8

Toute réunion de parties ouvertes de \mathbb{R}^n est une partie ouverte.

L'intersection d'un nombre fini de parties ouvertes de \mathbb{R}^n est une partie ouverte.

Preuve

- Soit I un ensemble et $(\Omega_i)_{i \in I}$ une famille d'ouverts de \mathbb{R}^n , indexés par I . Soit $M \in \bigcup_{i \in I} \Omega_i$ et $i \in I$ tel que $M \in \Omega_i$. L'ensemble Ω_i étant ouvert, il existe $r > 0$ tel que $B(M, r) \subset \Omega_i$ et a fortiori on a

$$B(M, r) \subset \bigcup_{i \in I} \Omega_i.$$

- Soient $\Omega_1, \dots, \Omega_p$ ($p \in \mathbb{N}^*$) des ouverts de \mathbb{R}^n , $M \in \bigcap_{i=1}^p \Omega_i$. Comme chaque Ω_i est ouvert, il existe, pour tout $i \in [1, p]$, un réel $r_i > 0$ tel que $B(M, r_i) \subset \Omega_i$. On pose $r = \min(r_1, \dots, r_p)$. On a alors $r > 0$ et, pour $i \in [1, p]$,

$$B(M, r) \subset B(M, r_i) \subset \Omega_i$$

et donc

$$B(M, r) \subset \bigcap_{i=1}^p \Omega_i.$$

□

Exemple

Le sous-ensemble de \mathbb{R}^2 constitué des couples (x, y) vérifiant

$$x > 0, \quad y > 0 \quad \text{et} \quad x + y < 1$$

est un ouvert. C'est l'intersection de trois demi-plans ouverts. C'est l'intérieur du triangle limité par les droites d'équation $x = 0$, $y = 0$ et $x + y = 1$. Il a pour sommets les points $(0, 0)$, $(0, 1)$ et $(1, 0)$.

3.2 Parties fermées

Définition 9

On dit qu'une partie Ω de \mathbb{R}^n est une partie fermée ou est un fermé de \mathbb{R}^n si son complémentaire dans \mathbb{R}^n est un ouvert.

Exemples

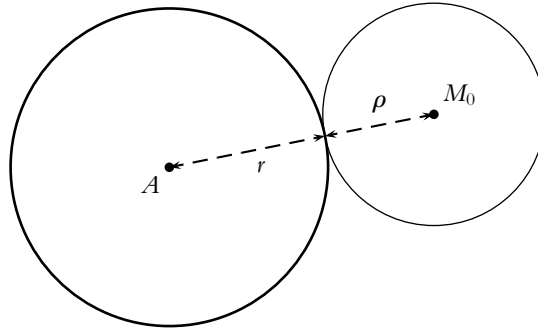
1. L'ensemble vide et \mathbb{R}^n sont des fermés car leur complémentaire est ouvert. Ces ensembles sont donc ouverts et fermés.
2. Toute boule fermée $B_f(A, r)$ de \mathbb{R}^n est un fermé.
On considère le complémentaire Ω de cette boule fermée. On a donc

$$\Omega = \{M \in \mathbb{R}^n, \quad d(A, M) > r\}.$$

Si $M_0 \in \Omega$, on pose $\rho = d(A, M_0) - r > 0$. On a alors, pour $M \in B(M_0, \rho)$,

$$d(A, M) \geq d(A, M_0) - d(M_0, M) > d(A, M_0) - \rho \geq r,$$

ce qui montre que $B(M_0, \rho) \subset \Omega$: l'ensemble Ω est ouvert, donc son complémentaire $B_f(A, r)$ est fermé.



En particulier, un singleton est un fermé (c'est une boule fermée de rayon nul).

On en déduit que, pour tout point $A \in \mathbb{R}^n$, l'ensemble $\mathbb{R}^n \setminus \{A\}$, qui est le complémentaire d'un fermé, est un ouvert.

3. Si $(u_1, \dots, u_n) \in \mathbb{R}^n \setminus \{0\}$ et $c \in \mathbb{R}$, l'ensemble \mathcal{E} des éléments (x_1, \dots, x_n) de \mathbb{R}^n tels que

$$u_1x_1 + \dots + u_nx_n + c \geq 0$$

est un fermé. Cet ensemble est un demi-espace (fermé) limité par l'hyperplan d'équation $u_1x_1 + \dots + u_nx_n + c = 0$.

En effet, c'est le complémentaire de l'ensemble d'équation

$$u_1x_1 + \dots + u_nx_n + c < 0$$

dont il a été démontré page 94 qu'il est ouvert.

De même, l'ensemble d'équation

$$u_1x_1 + \dots + u_nx_n + c \leq 0$$

est un fermé.

Proposition 9

Toute intersection de parties fermées de \mathbb{R}^n est une partie fermée.

La réunion d'un nombre fini de parties fermées de \mathbb{R}^n est une partie fermée.

Preuve

Ces propriétés résultent de celles des ouverts.

- Soit I un ensemble et $(\Omega_i)_{i \in I}$ une famille de fermés de \mathbb{R}^n , indexés par I . Pour tout $i \in I$, l'ensemble $\mathbb{R}^n \setminus \Omega_i$ est ouvert. On en déduit, d'après la proposition 8, que leur réunion

$$\bigcup_{i \in I} (\mathbb{R}^n \setminus \Omega_i) = \mathbb{R}^n \setminus \bigcap_{i \in I} \Omega_i$$

est ouverte. Le complémentaire $\bigcap_{i \in I} \Omega_i$ est fermé.

- De même, si $\Omega_1, \dots, \Omega_p$ sont des fermés ($p \in \mathbb{N}^*$), chaque ensemble $\mathbb{R}^n \setminus \Omega_i$ est ouvert et leur intersection

$$\bigcap_{i=1}^p (\mathbb{R}^n \setminus \Omega_i) = \mathbb{R}^n \setminus \bigcup_{i=1}^p \Omega_i$$

est ouverte, ce qui montre que $\bigcup_{i=1}^p \Omega_i$ est fermé. □

Exemples

1. Pour $A \in \mathbb{R}^n$ et $r > 0$, la sphère de centre A et de rayon r , $S = \{M \in \mathbb{R}^n \mid d(A, M) = r\}$ est un fermé de \mathbb{R}^n . On a en effet

$$S = B_f(A, r) \setminus B(A, r) = B_f(A, r) \cap (\mathbb{R}^n \setminus B(A, r)),$$

donc S est une intersection de fermés.

2. Si (a_1, \dots, a_n) et (b_1, \dots, b_n) sont deux n -listes telles que pour tout i , $a_i \leq b_i$, le pavé fermé $\mathcal{R} = [a_1, b_1] \times \dots \times [a_n, b_n]$ de \mathbb{R}^n est une partie fermée de \mathbb{R}^n . En effet, c'est l'intersection des $2n$ demi-espaces $x_i \geq a_i$ et $x_i \leq b_i$ qui sont fermés d'après les exemples de la page 96.

4. Parties convexes

Définition 10

On dit qu'une partie Ω de \mathbb{R}^n est convexe, si pour tous éléments A et B de Ω , le segment $[A, B]$ est inclus dans Ω .

Exemples

1. Pour tout couple (A, U) d'éléments de \mathbb{R}^n , U non nul, la droite $d_{A,U}$ est convexe. Soit $M = A + sU$ et $N = A + tU$ deux éléments de $d_{A,U}$ (s et t réels quelconques), $\lambda \in [0, 1]$. On a alors

$$(1 - \lambda)M + \lambda N = (1 - \lambda)(A + sU) + \lambda(A + tU) = A + ((1 - \lambda)s + \lambda t)U \in d_{A,U}.$$

Cela montre que $[M, N] \subset d_{A,U}$ et donc que $d_{A,U}$ est convexe.

2. Un segment $[A, B]$ est convexe. Soit C et D deux éléments de $[A, B]$. Il existe deux réels s et t dans $[0, 1]$ tels que $C = (1 - s)A + sB$ et $D = (1 - t)A + tB$. Si $M \in [C, D]$ il existe $\lambda \in [0, 1]$ tel que $M = (1 - \lambda)C + \lambda D$. On a alors

$$\begin{aligned} M &= (1 - \lambda)((1 - s)A + sB) + \lambda((1 - t)A + tB) \\ &= ((1 - \lambda)(1 - s) + \lambda(1 - t))A + ((1 - \lambda)s + \lambda t)B. \end{aligned}$$

Les réels $(1 - \lambda)(1 - s) + \lambda(1 - t)$ et $(1 - \lambda)s + \lambda t$ sont positifs et ont pour somme 1, donc M appartient à $[A, B]$.

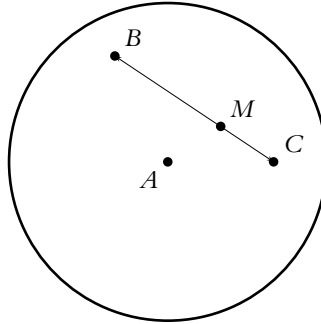
3. Une boule est convexe. Considérons la boule ouverte $B(A, r)$, B et C deux points de $B(A, r)$. Si M appartient à $[B, C]$, il existe $t \in [0, 1]$ tel que $M = (1 - t)B + tC$. On a alors

$$M - A = (1 - t)(B - A) + t(C - A).$$

On en déduit que

$$\begin{aligned} d(A, M) &= \|M - A\| \leq (1 - t)\|B - A\| + t\|C - A\| \\ &\leq (1 - t)d(A, B) + td(A, C) < (1 - t)r + tr, \end{aligned}$$

car B et C appartiennent à la boule $B(A, r)$. Cela montre que $d(A, M) < r$ et donc $M \in B(A, r)$: la boule ouverte $B(A, r)$ est convexe.

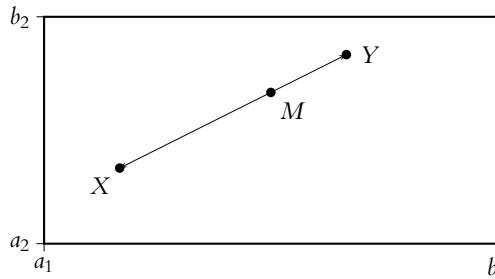


On démontre de la même façon qu'une boule fermée est convexe.

4. Si (a_1, \dots, a_n) et (b_1, \dots, b_n) sont deux n -listes telles que pour tout i , $a_i \leq b_i$, le pavé $\mathcal{R} = [a_1, b_1] \times \dots \times [a_n, b_n]$ de \mathbb{R}^n est convexe. En effet, soient $X = (x_1, \dots, x_n)$ et $Y = (y_1, \dots, y_n)$ deux éléments de \mathcal{R} , $Z = (z_1, \dots, z_n) = tX + (1-t)Y$ ($t \in [0, 1]$) un élément du segment $[X, Y]$. Pour tout $i \in \llbracket 1, n \rrbracket$,

$$z_i = tx_i + (1-t)y_i \in [a_i, b_i] \text{ car } x_i \in [a_i, b_i] \text{ et } y_i \in [a_i, b_i].$$

Ainsi $Z \in \mathcal{R}$ et $[X, Y] \subset \mathcal{R}$.



5. Si $(u_1, \dots, u_n) \in \mathbb{R}^n \setminus \{0\}$ et $c \in \mathbb{R}$, l'ensemble \mathcal{E} des éléments (x_1, \dots, x_n) de \mathbb{R}^n tels que $u_1x_1 + \dots + u_nx_n + c \geq 0$ est convexe. Cet ensemble est un demi-espace limité par l'hyperplan d'équation $u_1x_1 + \dots + u_nx_n + c = 0$. Soient $X = (x_1, \dots, x_n)$ et $Y = (y_1, \dots, y_n)$ deux éléments de \mathcal{E} , $t \in [0, 1]$. On pose $Z = (1-t)X + tY = (z_1, \dots, z_n)$. On a alors, pour $i \in \llbracket 1, n \rrbracket$, $z_i = (1-t)x_i + ty_i$. On en déduit que

$$u_1z_1 + \dots + u_nz_n + c = (1-t)(u_1x_1 + \dots + u_nx_n + c) + t(u_1y_1 + \dots + u_ny_n + c) \geq 0,$$

car $u_1x_1 + \dots + u_nx_n + c$, $u_1y_1 + \dots + u_ny_n + c$, $1-t$ et t sont positifs. Ainsi $Z \in \mathcal{E}$ et donc $[X, Y] \subset \mathcal{E}$.

On aurait le même résultat pour les ensembles de \mathbb{R}^n définis par

$$u_1x_1 + \dots + u_nx_n + c \leq 0 \text{ ou } u_1x_1 + \dots + u_nx_n + c > 0 \text{ ou } u_1x_1 + \dots + u_nx_n + c < 0.$$

Pour le cas $n = 2$, on pourra se reporter au livre de première année (chapitre 26, exercices pages 541 à 542).

1. Soit \mathcal{D} une droite affine de \mathbb{R}^n de vecteur directeur U , \mathcal{H} un hyperplan affine de \mathbb{R}^n de direction H .

1. Montrer que si $U \in H$, on a $\mathcal{D} \subset \mathcal{H}$ ou $\mathcal{D} \cap \mathcal{H} = \emptyset$.
2. Montrer que si U n'appartient pas à H , $\mathcal{D} \cap \mathcal{H}$ est un singleton.

2. 1. Montrer qu'étant donnés trois éléments A, B, C de \mathbb{R}^3 non alignés (*i.e.* n'appartenant pas à une même droite affine), il existe un unique plan affine contenant A, B, C .

2. Montrer que l'intersection de deux plans affines de \mathbb{R}^3 , de directions distinctes, est une droite affine.
3. Montrer que réciproquement toute droite affine de \mathbb{R}^3 est l'intersection de deux plans affines.
4. En déduire que toute droite affine de \mathbb{R}^3 possède un système d'équations du type :

$$\begin{cases} ux + vy + wz + c = 0 \\ u'x + v'y + w'z + c = 0. \end{cases}$$

Déterminer un tel système d'équations pour la droite passant par $A = (-1, 1, 2)$ de vecteur directeur $U = (1, 1, -1)$.

3. Dans \mathbb{R}^3 , on considère le plan affine \mathcal{P} dont une équation est

$$ux + vy + wz + h = 0.$$

1. On considère la droite \mathcal{D} passant par $A = (x_0, y_0, z_0)$ et de vecteur directeur $U = (a, b, c)$. Déterminer l'intersection de \mathcal{P} et \mathcal{D} .
2. Même question si la droite \mathcal{D} est donnée par le système d'équations

$$\begin{cases} u_1x + v_1y + w_1z + h_1 = 0 \\ u_2x + v_2y + w_2z + h_2 = 0. \end{cases}$$

4. Montrer que, pour tout couple (U, V) d'éléments de \mathbb{R}^n , on a

$$\|U + V\|^2 + \|U - V\|^2 = 2(\|U\|^2 + \|V\|^2).$$

5. Soit $p \in \mathbb{N}^*$, X_1, X_2, \dots, X_p des éléments de \mathbb{R}^n tels que, pour tout couple (i, j) d'indices distincts de $\llbracket 1, p \rrbracket$, on ait

$$\|X_i - X_j\| \geq 2.$$

1. Vérifier que

$$\sum_{1 \leq i < j \leq p} \|X_i - X_j\|^2 = p \sum_{i=1}^p \|X_i\|^2 - \left\| \sum_{i=1}^p X_i \right\|^2.$$

2. Soit B une boule fermée de rayon r contenant X_1, \dots, X_p . Montrer que

$$r \geq \sqrt{\frac{2(p-1)}{p}}.$$

6. Montrer que les sous-ensembles suivants de \mathbb{R}^3 sont ouverts et bornés :

1. $\{(x, y, z) \in \mathbb{R}^3 \mid 1 < \|(x, y, z)\| < 2\}$;
2. $\{(x, y, z) \in \mathbb{R}^3 \mid x + y + z < 1, x > 0, y > 0, z > 0\}$;
3. $\{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 - z^2 < 0, z \in]0, 1[\}$.

7. Montrer que tout sous-espace vectoriel de \mathbb{R}^n est un fermé de \mathbb{R}^n .

8. Soit $(A_p)_{p \in \mathbb{N}}$ une suite d'éléments de \mathbb{R}^n , c'est-à-dire une application de \mathbb{N} dans \mathbb{R}^n et A un élément de \mathbb{R}^n . On dit que la suite $(A_p)_{p \in \mathbb{N}}$ converge vers A si la suite $(d(A_p, A))_{p \in \mathbb{N}}$ converge vers 0.

Soit F une partie de \mathbb{R}^n . Montrer que F est fermée si et seulement, pour toute suite $(A_p)_{p \in \mathbb{N}}$ d'éléments de F convergeant vers $A \in \mathbb{R}^n$, la limite A appartient à F .

9. Soit Ω une partie de \mathbb{R}^n . On considère l'ensemble $\overline{\Omega}$, appelé *adhérence* de Ω , des points M de \mathbb{R}^n tels que, pour tout $r > 0$, la boule $B(M, r)$ ait une intersection non vide avec Ω et l'ensemble $\overset{\circ}{\Omega}$, appelé *intérieur* de Ω , des points M de \mathbb{R}^n tels qu'il existe $r > 0$ vérifiant $B(M, r) \subset \Omega$.

1. Montrer que $\overline{\Omega}$ est un fermé de \mathbb{R}^n qui contient Ω .

Montrer que $\overline{\Omega}$ est le plus petit fermé contenant Ω , c'est-à-dire que, si F un fermé de \mathbb{R}^n tel que $\Omega \subset F$, alors $\overline{\Omega} \subset F$.

2. Déterminer $\overline{\Omega}$ dans les cas suivants :

- a) $\Omega = B(A, r)$ ($A \in \mathbb{R}^2, r > 0$) ;
- b) $\Omega = \{M \in \mathbb{R}^n \mid d(M, A) > r\}$;
- c) $\Omega = \{(x_1, \dots, x_n) \in \mathbb{R}^n, u_1 x_1 + \dots + u_n x_n + c > 0\}$ (u_1, \dots, u_n, c réels, $(u_1, \dots, u_n) \neq (0, \dots, 0)$).

3. Montrer que $\overset{\circ}{\Omega}$ est un ouvert de \mathbb{R}^n qui contient Ω .

Montrer que $\overset{\circ}{\Omega}$ est le plus grand ouvert inclus dans Ω , c'est-à-dire que, si U un ouvert de \mathbb{R}^n tel que $U \subset \Omega$, alors $U \subset \overset{\circ}{\Omega}$.

4. Déterminer $\overset{\circ}{\Omega}$ dans les cas suivants :

- a) $\Omega = B_f(A, r)$ ($A \in \mathbb{R}^2, r > 0$);
 b) $\Omega = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid u_1x_1 + \dots + u_nx_n + c \geq 0\}$ (où u_1, \dots, u_n sont des réels non tous nuls et c un réel).

10. Soit Ω une partie de \mathbb{R}^n . On appelle frontière de Ω et on note $\text{Fr}(\Omega)$ l'ensemble des points M de \mathbb{R}^n tels que, pour tout $r > 0$, la boule $B(M, r)$ ait une intersection non vide avec Ω et avec le complémentaire de Ω .

1. Montrer que $\text{Fr}(\Omega)$ est un fermé de \mathbb{R}^n .
2. Déterminer la frontière des ensembles suivants :

- a) $\Omega = B(A, r)$ ($A \in \mathbb{R}^2 \mid r > 0$);
 b) $\Omega = \{M \in \mathbb{R}^n \mid \|M\| = 1\}$;
 c) $\Omega = \{(x, y, z) \in \mathbb{R}^3, -1 < x \leq 1\}$.

3. Montrer que la frontière de Ω est incluse dans Ω si et seulement si Ω est fermé.

11. Soit Ω une partie non-vide de \mathbb{R}^n . Pour tout $x \in \mathbb{R}^n$, on pose

$$d(x, \Omega) = \inf_{a \in \Omega} d(x, a).$$

1. Pour $\Omega \subset \mathbb{R}^n$ et $p \in \mathbb{N}^*$, on pose

$$\Omega_p = \left\{ x \in \Omega \mid d(x, \Omega) < \frac{1}{p} \right\}.$$

Montrer que Ω_p est un ouvert de \mathbb{R}^n et que $\bigcap_{p \in \mathbb{N}^*} \Omega_p = \overline{\Omega}$ (l'ensemble $\overline{\Omega}$ est défini dans l'exercice 9).

2. En déduire que tout fermé de \mathbb{R}^n est intersection d'une famille dénombrable d'ouverts.
3. Montrer que tout ouvert de \mathbb{R}^n est réunion d'une famille dénombrable de fermés.
4. Déterminer Ω_p quand Ω est une boule ouverte, puis lorsque Ω est une droite.

12. 1. Soit C une partie convexe de \mathbb{R}^n . Montrer que, si p est un entier naturel non nul, A_1, A_2, \dots, A_p des éléments de C et $\lambda_1, \lambda_2, \dots, \lambda_p$ des réels positifs tels que $\sum_{i=1}^p \lambda_i = 1$,

alors $\sum_{i=1}^p \lambda_i A_i$ appartient à C .

2. Soit Ω une partie quelconque de \mathbb{R}^n . On appelle *enveloppe convexe* de Ω et on note C l'ensemble des éléments M de \mathbb{R}^n pour lesquels il existe $p \in \mathbb{N}^*$, A_1, A_2, \dots, A_p dans Ω et $\lambda_1, \lambda_2, \dots, \lambda_p$ dans \mathbb{R}_+ tels que

$$\sum_{i=1}^p \lambda_i = 1 \quad \text{et} \quad M = \sum_{i=1}^p \lambda_i A_i.$$

- a) Montrer que C est un convexe contenant Ω .
 b) Montrer que C est le plus petit convexe de \mathbb{R}^n contenant Ω , c'est-à-dire que si C' est un convexe contenant Ω , on a $C \subset C'$.

13. Fonctions convexes Dans tout l'exercice, Ω désigne un ensemble convexe de \mathbb{R}^n et f une fonction définie sur Ω à valeurs réelles. La fonction f est dite convexe sur Ω si

$$\forall (x, y) \in \Omega^2 \quad \forall \lambda \in [0, 1] \quad f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

1. Montrer que les fonctions $x \mapsto \|x\|^2$ et $x \mapsto e^{\|x\|^2}$ sont convexes sur \mathbb{R}^n .
2. Soit f la fonction définie sur \mathbb{R}^2 par $f(u, v) = u^2 + v^2 - uv$.
Montrer que, pour $(u, v, u', v') \in \mathbb{R}^4$ et $\lambda \in [0, 1]$, on a

$$\begin{aligned} f(\lambda(u, v) + (1 - \lambda)(u', v')) &= \lambda f(u, v) + (1 - \lambda)f(u', v') \\ &\quad - \lambda(1 - \lambda)f(u - u', v - v'). \end{aligned}$$

En déduire que f est convexe.

3. Soient x et y deux points de Ω et g la fonction, à valeurs dans \mathbb{R} , définie sur $[0, 1]$ par

$$g(t) = f(tx + (1 - t)y).$$

Montrer que f est convexe sur Ω si et seulement si, pour tout choix de (x, y) , la fonction g est convexe sur $[0, 1]$.

4. Montrer que f est convexe sur Ω si et seulement si $C = \{(x, t) \in \Omega \times \mathbb{R} \mid f(x) \leq t\}$ est une partie convexe de \mathbb{R}^{n+1} .

Fonctions de n variables

Continuité

5

Toutes les fonctions considérées dans ce chapitre seront définies sur une partie de \mathbb{R}^n et à valeurs dans \mathbb{R} .

1. Graphe d'une fonction

Définition 1

Soit f une fonction définie sur $D \subset \mathbb{R}^n$, à valeurs dans \mathbb{R} . On appelle graphe de f l'ensemble

$$\{(x_1, \dots, x_n, x_{n+1}) \in \mathbb{R}^{n+1} \mid (x_1, \dots, x_n) \in D \text{ et } x_{n+1} = f(x_1, \dots, x_n)\}.$$

► Remarques

- Le graphe d'une fonction $f : D \rightarrow \mathbb{R}$, où $D \subset \mathbb{R}^n$ est une hypersurface de \mathbb{R}^{n+1} .
- Pour $n = 2$, le graphe est une surface de \mathbb{R}^3 . Dans l'espace usuel rapporté à un repère orthonormal $(O, \vec{i}, \vec{j}, \vec{k})$ on le représente par l'ensemble des points de coordonnées $(x, y, f(x, y))$, le couple (x, y) décrivant D . Pour $n \geq 3$, on ne peut représenter directement le graphe dans l'espace usuel.

Définition 2

La fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction affine de n variables s'il existe $(u_1, \dots, u_n, c) \in \mathbb{R}^{n+1}$ tel que, pour tout $(x_1, \dots, x_n) \in \mathbb{R}^n$,

$$f(x_1, \dots, x_n) = u_1x_1 + \dots + u_nx_n + c.$$

Proposition 1

Pour tout $(u_1, \dots, u_n) \in \mathbb{R}^n$, le graphe de la fonction affine $(x_1, \dots, x_n) \mapsto u_1x_1 + \dots + u_nx_n$ est un hyperplan vectoriel de \mathbb{R}^{n+1} .

Preuve

Le graphe de la fonction $f : (x_1, \dots, x_n) \mapsto u_1x_1 + \dots + u_nx_n$ est

$$\{(x_1, \dots, x_n, x_{n+1}) \in \mathbb{R}^{n+1} \mid u_1x_1 + \dots + u_nx_n - x_{n+1} = 0\}.$$

On reconnaît l'équation d'un hyperplan vectoriel de \mathbb{R}^{n+1} . □

Proposition 2

Pour tout $(u_1, \dots, u_n, c) \in \mathbb{R}^{n+1}$, le graphe de la fonction affine

$$(x_1, \dots, x_n) \longmapsto u_1x_1 + \dots + u_nx_n + c$$

est un hyperplan affine de \mathbb{R}^{n+1} .

Preuve

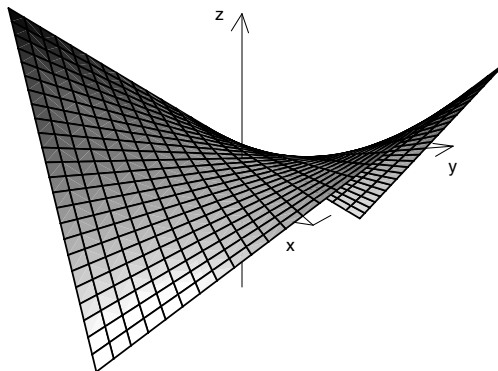
Le graphe \mathcal{G} de la fonction affine a pour équation

$$u_1x_1 + \dots + u_nx_n - x_{n+1} + c = 0.$$

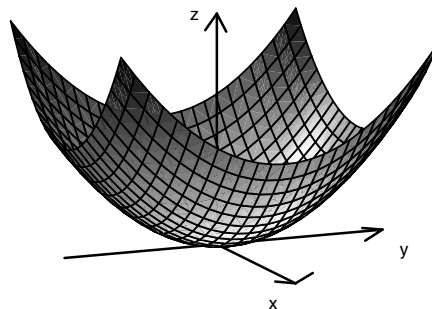
Il résulte de la proposition 4 du chapitre 4 que \mathcal{G} est un hyperplan affine. □

Exemples

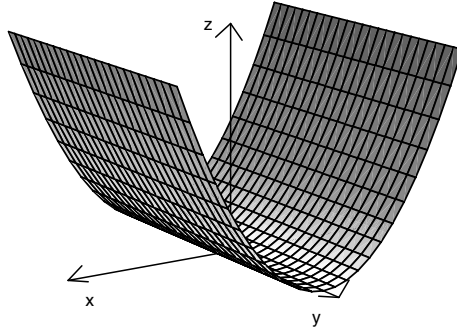
1. Soit f la fonction définie sur \mathbb{R}^2 par $f(x, y) = xy$. Le graphe de f est la surface de \mathbb{R}^3 d'équation $z = xy$. Elle est appelée paraboloides hyperbolique.



2. Soit f la fonction définie sur \mathbb{R}^n par $f(X) = \|X\|^2$. Le graphe est une surface de révolution autour de la droite engendrée par le dernier vecteur de la base canonique de \mathbb{R}^{n+1} . Pour $n = 2$, on obtient la surface d'équation $z = x^2 + y^2$, appelée paraboloides de révolution.



3. Soit f la fonction définie sur \mathbb{R}^2 par $f(x, y) = x^2$. Le graphe de f est la surface de \mathbb{R}^3 d'équation $z = x^2$. C'est un cylindre.



Lignes de niveau

Définition 3

Soit f une fonction définie sur $D \subset \mathbb{R}^n$, à valeurs dans \mathbb{R} . Pour tout réel λ , on appelle ligne de niveau λ l'ensemble des éléments de D tels que

$$f(x_1, \dots, x_n) = \lambda.$$

► Remarque

La ligne de niveau λ est l'intersection du graphe de f avec l'hyperplan d'équation $x_{n+1} = \lambda$. C'est donc l'intersection d'une surface et d'un hyperplan. Étudier les lignes de niveau d'une fonction est une manière de visualiser son graphe.

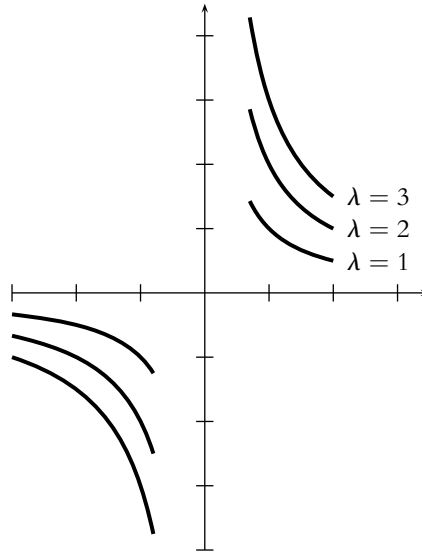
Exemples

- Soit f la fonction affine définie par $f(x_1, \dots, x_n) = u_1 x_1 + \dots + u_n x_n + c$.
Si $(u_1, \dots, u_n) = (0, \dots, 0)$, la fonction f est constante. La ligne de niveau c est \mathbb{R}^n . Pour $\lambda \neq c$, la ligne de niveau λ est l'ensemble vide.
Si $(u_1, \dots, u_n) \neq 0$, pour tout $\lambda \in \mathbb{R}$ la ligne de niveau λ a pour équation

$$u_1 x_1 + \dots + u_n x_n - \lambda = 0.$$

C'est un hyperplan affine de \mathbb{R}^n .

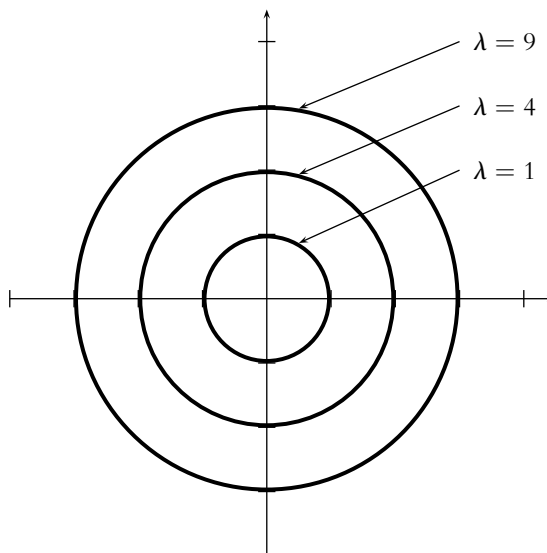
- Soit f la fonction définie sur \mathbb{R}^2 par $f(x, y) = xy$.
Si $\lambda = 0$, $xy = \lambda$ équivaut à $x = 0$ ou $y = 0$. La ligne de niveau 0 est la réunion de deux droites.
Si $\lambda \neq 0$, la ligne de niveau λ , d'équation $xy = \lambda$, est une hyperbole. C'est la courbe représentative de la fonction $x \mapsto \frac{\lambda}{x}$.



3. Soit f la fonction définie sur \mathbb{R}^n par $f(X) = \|X\|^2$.
 Si $\lambda < 0$, la ligne de niveau λ de f est l'ensemble vide : le graphe de f est inclus dans le demi-espace $x_{n+1} \geq 0$.
 Si $\lambda \geq 0$, la ligne de niveau λ a pour équation

$$\|X\| = \sqrt{\lambda}.$$

On obtient donc une hypersphère de centre $(0, \dots, 0)$ et de rayon $\sqrt{\lambda}$. Si $n = 2$, on obtient un cercle ; si $n = 3$, on a une sphère.



4. Soit f la fonction définie sur \mathbb{R}^2 par $f(x, y) = x^2$.
 Si $\lambda < 0$, la ligne de niveau λ de f est l'ensemble vide : le graphe de f est inclus dans le demi-espace $z \geq 0$.
 Si $\lambda \geq 0$, la ligne de niveau λ a pour équation $|x| = \sqrt{\lambda}$. On obtient la réunion de deux droites, confondues si $\lambda = 0$.

2. Continuité d'une fonction de \mathbb{R}^n dans \mathbb{R}

Dans toute cette partie, les fonctions seront définies sur une partie D de \mathbb{R}^n et à valeurs dans \mathbb{R} .

Définition 4

Soit $f : D \rightarrow \mathbb{R}$, $M_0 \in D$, ℓ un réel.

On dit que f admet pour limite ℓ en M_0 si,

$$\forall \varepsilon > 0 \quad \exists \eta > 0 \quad \forall M \in D \quad (d(M_0, M) \leq \eta \implies |f(M) - \ell| \leq \varepsilon).$$

On dit que f est continue en M_0 si,

$$\forall \varepsilon > 0 \quad \exists \eta > 0 \quad \forall M \in D \quad (d(M_0, M) \leq \eta \implies |f(M) - f(M_0)| \leq \varepsilon),$$

c'est-à-dire si f admet pour limite $f(M_0)$ en M_0 .

On dit que f est continue sur D si f est continue en M_0 pour tout point M_0 de D .

► Remarques

- Substantiellement, la définition est la même que pour la continuité d'une fonction d'une variable : $f(M)$ peut être obtenu aussi proche que l'on veut de $f(M_0)$, à condition que l'on prenne M assez proche de M_0 .
- La condition $d(M_0, M) \leq \eta$ peut encore s'écrire $\|M - M_0\| \leq \eta$ ou $M \in B_\eta(M_0)$.

Exemple

Une fonction constante sur D est continue sur D .

Proposition 3

Les projections

$$\begin{aligned} \mathbb{R}^n &\longrightarrow \mathbb{R} \\ p_i : (x_1, \dots, x_n) &\longmapsto x_i, \end{aligned}$$

où $1 \leq i \leq n$ sont continues sur \mathbb{R}^n .

La fonction $M \mapsto \|M\|$ est continue sur \mathbb{R}^n .

Preuve

Si $M_0 = (u_1, \dots, u_n)$ et $M = (x_1, \dots, x_n)$, on a

$$|p_i(M) - p_i(M_0)| = |x_i - u_i| \leq d(M_0, M).$$

Ainsi, on obtient pour tout $\varepsilon > 0$,

$$\forall M \in \mathbb{R}^n \quad (d(M_0, M) \leq \varepsilon \implies |\rho_i(M) - \rho_i(M_0)| \leq \varepsilon)$$

et ρ_i est continue en M_0 .

On a, pour M et M_0 dans \mathbb{R}^n ,

$$\|M\| - \|M_0\| \leq \|M - M_0\| \leq d(M_0, M),$$

par l'inégalité triangulaire. On en déduit que pour tout $\varepsilon > 0$,

$$d(M_0, M) \leq \varepsilon \implies \|M\| - \|M_0\| \leq \varepsilon,$$

ce qui montre la continuité de l'application $M \longmapsto \|M\|$ en tout point M_0 de \mathbb{R}^n . □

Donnons quelques exemples supplémentaires de fonctions continues.

Exemples

1. Pour tout $A \in \mathbb{R}^n$, la fonction $f : M \longmapsto d(A, M)$ est continue sur \mathbb{R}^n .

En effet, pour M_0 et M dans \mathbb{R}^n , on a

$$|f(M) - f(M_0)| = |d(A, M) - d(A, M_0)| \leq d(M_0, M),$$

par inégalité triangulaire et donc, pour tout $\varepsilon > 0$, on a

$$d(M_0, M) \leq \varepsilon \implies |f(M) - f(M_0)| \leq \varepsilon,$$

ce qui montre la continuité de f en M_0 pour tout M_0 de \mathbb{R}^n .

2. Pour tout $(u_1, \dots, u_n, c) \in \mathbb{R}^{n+1}$, la fonction affine

$$(x_1, \dots, x_n) \longmapsto u_1 x_1 + \dots + u_n x_n + c$$

est continue sur \mathbb{R}^n .

Soit $M_0 = (a_1, \dots, a_n)$ et $M = (x_1, \dots, x_n)$ deux points de \mathbb{R}^n . On a alors

$$\begin{aligned} |f(M) - f(M_0)| &= |u_1(x_1 - a_1) + \dots + u_n(x_n - a_n)| \\ &\leq |u_1||x_1 - a_1| + \dots + |u_n||x_n - a_n| \\ &\leq (|u_1| + \dots + |u_n|)d(M_0, M). \end{aligned}$$

Si $u_1 = \dots = u_n = 0$, f est constante, sinon on a, pour tout $\varepsilon > 0$,

$$d(M_0, M) \leq \frac{\varepsilon}{|u_1| + \dots + |u_n|} \implies |f(M) - f(M_0)| \leq \varepsilon.$$

La fonction f est continue en M_0 pour tout $M_0 \in \mathbb{R}^n$.

3. Soit f la fonction définie sur \mathbb{R}^3 par

$$f(x, y, z) = \frac{xy + yz + zx}{\sqrt{x^2 + y^2 + z^2}} \quad \text{si } (x, y, z) \neq (0, 0, 0) \quad \text{et } f(0, 0, 0) = 0.$$

Montrons que f est continue en $(0, 0, 0)$. On sait que, pour tout $(x, y, z) \in \mathbb{R}^3$, les réels $|x|$, $|y|$ et $|z|$ sont inférieurs ou égaux à $\sqrt{x^2 + y^2 + z^2}$. On en déduit que si $(x, y, z) \neq (0, 0, 0)$,

on a

$$\begin{aligned} |f(x, y, z) - f(0, 0, 0)| &= |f(x, y, z)| \leq \frac{|x||y| + |y||z| + |z||x|}{\sqrt{x^2 + y^2 + z^2}} \\ &\leq \frac{3(x^2 + y^2 + z^2)}{\sqrt{x^2 + y^2 + z^2}} \leq 3\sqrt{x^2 + y^2 + z^2} \\ &\leq 3d((x, y, z), (0, 0, 0)). \end{aligned}$$

Ainsi, pour tout $\varepsilon > 0$, on a

$$d((x, y, z), (0, 0, 0)) \leq \frac{\varepsilon}{3} \implies |f(x, y, z) - f(0, 0, 0)| \leq \varepsilon$$

et f est continue en $(0, 0, 0)$.

4. Donnons un exemple de fonction non continue.

Soit a un réel et f définie sur \mathbb{R}^3 par

$$f(x, y, z) = \frac{xy}{x^2 + y^2 + z^2} \quad \text{si } (x, y, z) \neq (0, 0, 0) \quad \text{et } f(0, 0, 0) = a.$$

Montrons que, quel que soit le choix de a , f n'est pas continue en $(0, 0, 0)$. On a, pour x non nul

$$f(x, x, 0) = \frac{1}{2} \quad \text{et } f(x, 0, 0) = 0.$$

Comme il existe des points de \mathbb{R}^3 de la forme $(x, x, 0)$ et $(x, 0, 0)$ aussi proche que l'on veut de $(0, 0, 0)$, on en déduit que f n'est pas continue en $(0, 0, 0)$.

Si en effet f est continue en $(0, 0, 0)$, en prenant $\varepsilon = \frac{1}{6}$, on obtient l'existence de η tel que

$$\sqrt{x^2 + y^2 + z^2} \leq \eta \implies |f(x, y, z) - a| \leq \frac{1}{6}.$$

En particulier, en prenant $|x| \leq \frac{\eta}{\sqrt{2}}$, on obtient

$$|f(x, x, 0) - a| = \left| \frac{1}{2} - a \right| \leq \frac{1}{6} \quad \text{et} \quad |f(x, 0, 0) - a| = |a| \leq \frac{1}{6}.$$

Ces deux relations sont incompatibles car

$$\left| \frac{1}{2} - a \right| + |a| \geq \frac{1}{2}.$$

Quel que soit a , f n'est pas continue en $(0, 0, 0)$. Autrement dit, la fonction f définie sur $\mathbb{R}^3 \setminus \{(0, 0, 0)\}$ par $f(x, y, z) = \frac{xy}{x^2 + y^2 + z^2}$ n'est pas prolongeable par continuité en $(0, 0, 0)$.

Proposition 4

Soit $f : D \rightarrow \mathbb{R}$ et $M_0 \in D$.

Si f est continue en M_0 , il existe $\eta > 0$ tel que f soit définie et bornée sur $D \cap B_f(M_0, \eta)$.

Preuve

En prenant la définition de la limite avec $\varepsilon = 1$, on obtient l'existence de $\eta > 0$ tel que

$$\forall M \in D \quad (d(M_0, M) \leq \eta \implies |f(M) - f(M_0)| \leq 1).$$

Si $M \in D \cap B_r(M_0, \eta)$, on a $|f(M) - f(M_0)| \leq 1$ et donc $|f(M)| \leq |f(M_0)| + 1$. □

Proposition 5

Soit $f : D \longrightarrow \mathbb{R}$ et $M_0 \in D$.

Si f est continue en M_0 et si $f(M_0) \neq 0$, il existe $\eta > 0$ tel que $|f|$ soit minorée par $\frac{|f(M_0)|}{2}$ et donc ne s'annule pas sur $D \cap B_r(M_0, \eta)$.

Preuve

Prenons $\varepsilon = \frac{|f(M_0)|}{2}$. Il existe $\eta > 0$ tel que

$$\forall M \in D \quad \left(d(M_0, M) \leq \eta \implies |f(M) - f(M_0)| \leq \frac{|f(M_0)|}{2} \right).$$

Pour $M \in D \cap B_r(M_0, \eta)$, on a, d'après l'inégalité triangulaire,

$$|f(M)| \geq |f(M_0)| - |f(M) - f(M_0)| \geq |f(M_0)| - \frac{|f(M_0)|}{2} \geq \frac{|f(M_0)|}{2} > 0.$$

La fonction f ne s'annule pas sur $D \cap B_r(M_0, \eta)$. □

3. Opérations sur les fonctions continues

Théorème 1

Soient f et g deux fonctions définies sur le même sous-ensemble D de \mathbb{R}^n , à valeurs dans \mathbb{R} , M_0 un point de D .

Si f et g sont continues en M_0 , alors les fonctions $f + g$, λf ($\lambda \in \mathbb{R}$) et f/g sont continues en M_0 .

Si, de plus, $g(M_0)$ n'est pas nul, les fonctions $\frac{1}{g}$ et $\frac{f}{g}$ sont définies sur l'intersection de D et d'une boule de centre M_0 de rayon strictement positif et sont continues en M_0 .

Preuve

- On a, pour tout $M \in D$,

$$|(f + g)(M) - (f + g)(M_0)| \leq |f(M) - f(M_0)| + |g(M) - g(M_0)|.$$

Soit $\varepsilon > 0$. Il existe $\eta_1 > 0$ et $\eta_2 > 0$ tels que, pour tout $M \in D$,

$$d(M_0, M) \leq \eta_1 \implies |f(M) - f(M_0)| \leq \frac{\varepsilon}{2}$$

et

$$d(M_0, M) \leq \eta_2 \implies |g(M) - g(M_0)| \leq \frac{\varepsilon}{2}.$$

On a donc, pour $M \in D$,

$$d(M_0, M) \leq \min(\eta_1, \eta_2) \implies |(f+g)(M) - (f+g)(M_0)| \leq \varepsilon$$

et $f+g$ est continue en M_0 .

- Si $\lambda = 0$ la fonction λf est nulle donc continue en M_0 .
Si $\lambda \neq 0$ et $\varepsilon > 0$, il existe, par continuité de f , $\eta > 0$ tel que, pour tout $M \in D$,

$$d(M_0, M) \leq \eta \implies |f(M) - f(M_0)| \leq \frac{\varepsilon}{|\lambda|}.$$

Si $d(M_0, M) \leq \eta$, on a donc

$$|\lambda f(M) - \lambda f(M_0)| = \lambda |f(M) - f(M_0)| \leq |\lambda| \frac{\varepsilon}{|\lambda|} \leq \varepsilon.$$

La fonction λf est donc continue en M_0 .

- Pour fg , on écrit, pour tout $M \in D$,

$$\begin{aligned} |(fg)(M) - (fg)(M_0)| &= |(f(M) - f(M_0))g(M) + (g(M) - g(M_0))f(M_0)| \\ &\leq |f(M) - f(M_0)||g(M)| + |g(M) - g(M_0)||f(M_0)|. \end{aligned}$$

Comme g est continue en M_0 , elle est bornée au voisinage de M_0 . Il existe η_1 et $K > 0$ tels que, pour tout $M \in D$,

$$d(M_0, M) \leq \eta_1 \implies |g(M)| \leq K.$$

On a, pour $d(M_0, M) \leq \eta_1$,

$$|(fg)(M) - (fg)(M_0)| \leq |f(M) - f(M_0)|K + |g(M) - g(M_0)||f(M_0)|.$$

Par continuité de f et g en M_0 , il existe $\eta_2 > 0$ et $\eta_3 > 0$ tels que

$$\begin{aligned} d(M_0, M) \leq \eta_2 &\implies |f(M) - f(M_0)| \leq \frac{\varepsilon}{2K} \\ \text{et } d(M_0, M) \leq \eta_3 &\implies |g(M) - g(M_0)| \leq \frac{\varepsilon}{2|f(M_0)| + 1}. \end{aligned}$$

On a alors, pour tout $M \in D$,

$$d(M_0, M) \leq \min(\eta_1, \eta_2, \eta_3) \implies |(fg)(M) - (fg)(M_0)| \leq \varepsilon,$$

ce qui prouve la continuité de fg en M_0 .

- Si $g(M_0) \neq 0$, il existe d'après la proposition 5 $\eta_1 > 0$ tel que, pour tout $M \in B_r(M_0, \eta_1)$,

$$|g(M)| \geq \frac{|g(M_0)|}{2} > 0.$$

Les fonctions $\frac{1}{g}$ et $\frac{f}{g}$ sont donc définies sur $D \cap B_r(M_0, \eta_1)$.

Pour $M \in D \cap B_r(M_0, \eta_1)$, on peut écrire

$$\left| \frac{1}{g(M)} - \frac{1}{g(M_0)} \right| = \frac{|g(M) - g(M_0)|}{|g(M)||g(M_0)|} \leq \frac{2}{(g(M_0))^2} |g(M) - g(M_0)|.$$

Par continuité de g en M_0 , on peut trouver $\eta_2 > 0$ tel que, pour tout $M \in D$,

$$d(M_0, M) \leq \eta_2 \implies |g(M) - g(M_0)| \leq \frac{\varepsilon (g(M_0))^2}{2}.$$

On alors, pour tout $M \in D$,

$$d(M_0, M) \leq \min(\eta_1, \eta_2) \implies \left| \frac{1}{g(M)} - \frac{1}{g(M_0)} \right| \leq \varepsilon,$$

ce qui montre la continuité de $\frac{1}{g}$ en M_0 .

Enfin, la fonction $\frac{f}{g}$ est continue en M_0 comme produit des deux fonctions $\frac{1}{g}$ et f qui sont toutes deux continues en M_0 . \square

Corollaire 1

Si f et g sont deux fonctions définies et continues sur le sous-ensemble D de \mathbb{R}^n , les fonctions $f + g$, λf ($\lambda \in \mathbb{R}$) et $f \cdot g$ sont continues sur D .

Si, de plus, g ne s'annule pas sur D , les fonctions $\frac{1}{g}$ et $\frac{f}{g}$ sont continues sur D .

Preuve

Ce corollaire résulte de manière immédiate du théorème précédent, appliqué en M_0 , pour tout M_0 de D . \square

Exemples

- Pour tout $i \in \llbracket 1, n \rrbracket$, la fonction $(x_1, \dots, x_n) \mapsto x_i$ est continue sur \mathbb{R}^n . Il s'ensuit que, pour tous entiers naturels k_1, \dots, k_n , la fonction $(x_1, \dots, x_n) \mapsto x_1^{k_1} \dots x_n^{k_n}$ qui s'obtient comme produit de telles fonctions est continue sur \mathbb{R}^n .
On en déduit que les fonctions polynomiales en (x_1, \dots, x_n) , c'est-à-dire les combinaisons linéaires de fonctions $(x_1, \dots, x_n) \mapsto x_1^{k_1} \dots x_n^{k_n}$ sont continues sur \mathbb{R}^n .
- Le quotient de deux fonctions polynomiales dont le dénominateur ne s'annule pas est une fonction continue sur \mathbb{R}^n .

Composition

Théorème 2

Soient $f : D \rightarrow \mathbb{R}$ une fonction définie sur une partie D de \mathbb{R}^n , $\varphi : I \rightarrow \mathbb{R}$, où I est un intervalle de \mathbb{R} , et M_0 un élément de D . On suppose que $f(D) \subset I$.

Si f est continue en M_0 et φ continue en $f(M_0)$, alors la fonction $\varphi \circ f$ est continue en M_0 .

Posons $y_0 = f(M_0)$. Soit $\varepsilon > 0$. Comme φ est continue en y_0 , il existe $\eta > 0$ tel que, pour tout $y \in I$,

$$|y - y_0| \leq \eta \implies |\varphi(y) - \varphi(y_0)| \leq \varepsilon.$$

Comme f est continue en M_0 , on peut ensuite trouver $\eta' > 0$ tel que, pour tout $M \in D$,

$$d(M_0, M) \leq \eta' \implies |f(M) - f(M_0)| \leq \eta.$$

Si $M \in D$, on sait que $f(M) \in I$. On obtient donc, pour tout $M \in D$,

$$d(M_0, M) \leq \eta' \implies |\varphi(f(M)) - \varphi(f(M_0))| \leq \varepsilon,$$

ce qui montre la continuité de $\varphi \circ f$ en M_0 .

Corollaire 2

Soient D une partie de \mathbb{R}^n , I un intervalle de \mathbb{R} , $f : D \rightarrow \mathbb{R}$, $\varphi : I \rightarrow \mathbb{R}$ des fonctions continues. On suppose que $f(D) \subset I$. Alors $\varphi \circ f$ est continue sur D .

Preuve

Ce corollaire résulte immédiatement du théorème qui précède, appliqué en M_0 pour tout M_0 de D . □

Corollaire 3

Si φ est une fonction continue sur l'intervalle I de \mathbb{R} , alors pour tout $i \in \llbracket 1, n \rrbracket$, la fonction $f_i : (x_1, \dots, x_n) \mapsto \varphi(x_i)$ est continue sur $\mathbb{R}^{i-1} \times I \times \mathbb{R}^{n-i}$.

Preuve

En effet, f_i peut s'écrire $f_i = \varphi \circ p_i$, où $p_i : (x_1, \dots, x_n) \mapsto x_i$ est continue sur \mathbb{R}^n . □

Exemples

1. Si $f : D \rightarrow \mathbb{R}$ est continue, la fonction $(x_1, \dots, x_n) \mapsto |f(x_1, \dots, x_n)|$ est continue sur D .
2. Si $f : D \rightarrow \mathbb{R}$ est continue et positive, la fonction

$$(x_1, \dots, x_n) \mapsto \sqrt{f(x_1, \dots, x_n)}$$

est continue sur D .

Théorème 3

Soient D une partie de \mathbb{R}^n , f une application de D dans \mathbb{R} , I un intervalle de \mathbb{R} , u_1, \dots, u_n des applications de I dans \mathbb{R} telles que, pour tout $t \in I$, le n -uplet $(u_1(t), \dots, u_n(t))$ appartienne à D , $t_0 \in I$.

Alors si u_1, \dots, u_n sont continues en t_0 et si f est continue en $(u_1(t_0), \dots, u_n(t_0))$, la fonction $\varphi : I \rightarrow \mathbb{R}$ définie par

$$\varphi(t) = f(u_1(t), \dots, u_n(t))$$

est continue en t_0 .

Preuve

Il résulte des hypothèses que φ est définie sur I . Posons $M_0 = (u_1(t_0), \dots, u_n(t_0))$ et pour tout $t \in I$, $M(t) = (u_1(t), \dots, u_n(t))$. Soit $\varepsilon > 0$. La fonction f étant continue en M_0 , il existe $\eta > 0$ tel que, pour tout $M \in D$,

$$d(M_0, M) \leq \eta \implies |f(M) - f(M_0)| \leq \varepsilon.$$

Les fonctions u_1, \dots, u_n étant continues en t_0 , il existe η_1, \dots, η_n , réels strictement positifs tels que, pour tout $t \in I$ et tout $i \in \llbracket 1, n \rrbracket$,

$$|t - t_0| \leq \eta_i \implies |u_i(t) - u_i(t_0)| \leq \frac{\eta_i}{\sqrt{n}}.$$

Si t appartient à I et vérifie $|t - t_0| \leq \min(\eta_1, \dots, \eta_n)$, on a $M(t) = (u_1(t), \dots, u_n(t)) \in D$ et $d(M_0, M(t)) \leq \eta$ d'après la proposition 7 du chapitre 4. On en déduit que

$$|\varphi(t) - \varphi(t_0)| = |f(M(t)) - f(M_0)| \leq \varepsilon.$$

On a donc, pour tout $t \in I$,

$$|t - t_0| \leq \min(\eta_1, \dots, \eta_n) \implies |\varphi(t) - \varphi(t_0)| \leq \varepsilon,$$

ce qui montre la continuité de φ en t_0 . \square

Corollaire 4

Soient D une partie de \mathbb{R}^n , f une application de D dans \mathbb{R} , I un intervalle de \mathbb{R} , u_1, \dots, u_n des applications de I dans \mathbb{R} telles que, pour tout $t \in D$, le n -uplet $(u_1(t), \dots, u_n(t))$ appartienne à D .

Alors si u_1, \dots, u_n sont continues sur I et si f est continue sur D , la fonction $\varphi : I \longrightarrow \mathbb{R}$ définie par $\varphi(t) = f(u_1(t), \dots, u_n(t))$ est continue sur I .

Preuve

Ce corollaire résulte immédiatement du théorème qui précède, que l'on l'applique en t_0 pour tout t_0 de I . \square

Exemple

Soit D un ouvert de \mathbb{R}^n , $f : D \longrightarrow \mathbb{R}$, $M_0 \in D$. Si la fonction f est continue en M_0 alors, pour tout vecteur U de \mathbb{R}^n , la fonction

$$f_U : t \longmapsto f(M_0 + tU)$$

est continue en 0, c'est-à-dire a pour limite $f(M_0)$ en 0.

En effet, D contient une boule ouverte $B(M_0, r)$ et f_U est définie au voisinage de 0 (pour $|t||U| \leq r$).

Si on pose $M_0 = (a_1, \dots, a_n)$ et $U = (\alpha_1, \dots, \alpha_n)$, on obtient

$$f_U(t) = f(a_1 + \alpha_1 t, \dots, a_n + \alpha_n t)$$

et le théorème s'applique avec $u_i : t \longmapsto a_i + \alpha_i t$, qui est continue car affine.

Étudier la continuité de la fonction f_U en 0, c'est faire tendre le point M vers M_0 , en se déplaçant sur la droite affine $d_{M_0, U}$.

La continuité en 0 de f_U pour tout vecteur U ne suffit pas à assurer la continuité de f en M_0 (cf exemple page 649 du livre de première année).

► Remarque

Le théorème permet de montrer la continuité de fonctions composées mais il permet aussi de montrer qu'une application $f : D \longrightarrow \mathbb{R}$, où $D \subset \mathbb{R}^n$, n'est pas continue en $M_0 = (a_1, \dots, a_n)$. Si on peut trouver des fonctions u_1, \dots, u_n définies sur un même intervalle I contenant t_0 , ayant pour limite respective a_1, \dots, a_n en t_0 et telles que la fonction $t \longmapsto f(u_1(t), \dots, u_n(t))$ n'ait pas pour limite $f(M_0)$ en t_0 , on conclut que f n'est pas continue en M_0 .

Exemple

Soient $a \in \mathbb{R}$ et f définie sur \mathbb{R}^3 par

$$f(x, y, z) = \begin{cases} \frac{xyz}{x^2 + y^4 + z^4} & \text{si } (x, y, z) \neq (0, 0, 0) \\ a & \text{si } (x, y, z) = (0, 0, 0). \end{cases}$$

On a pour tout $t \in \mathbb{R}^*$, $f(t, t, t) = \frac{t^3}{t^2 + t^4 + t^4} = \frac{t}{1 + 2t^2}$. On en déduit que $\lim_{t \rightarrow 0} f(t, t, t) = 0$.

On obtient de même $f(t^2, t, t) = \frac{t^4}{t^4 + t^4 + t^4} = \frac{1}{3}$ et $\lim_{t \rightarrow 0} f(t^2, t, t) = \frac{1}{3}$. Aucune valeur de a ne peut rendre la fonction continue en $(0, 0, 0)$, sinon les deux limites précédentes seraient égales à $f(0, 0, 0)$.

4. Propriétés des fonctions continues

Théorème 4

Si f une application continue de \mathbb{R}^n dans \mathbb{R} , l'image réciproque par f d'un intervalle ouvert (respectivement fermé) de \mathbb{R} est un ouvert (respectivement un fermé) de \mathbb{R}^n .

Preuve

- Soit I un intervalle ouvert de \mathbb{R} , $A \in f^{-1}(I)$. On a donc $f(A) \in I$ et comme I est un intervalle ouvert, il existe $\varepsilon > 0$ tel que $[f(A) - \varepsilon, f(A) + \varepsilon] \subset I$. Par continuité de f en A , il existe η tel que, pour tout $M \in \mathbb{R}^n$,

$$d(A, M) \leq \eta \implies |f(M) - f(A)| \leq \varepsilon.$$

Si M appartient à $B(A, \eta)$ alors $f(M) \in [f(A) - \varepsilon, f(A) + \varepsilon] \subset I$ et $f(M)$ appartient à I . Ainsi $f^{-1}(I)$ contient $B(A, \eta)$ et $f^{-1}(I)$ est ouvert.

- Soit $I = [\alpha, \beta]$ ($\alpha \leq \beta$) un intervalle fermé de \mathbb{R} . On montre que $f^{-1}(I)$ est fermé, c'est-à-dire que $\mathbb{R}^n \setminus f^{-1}(I)$ est ouvert. On a

$$\mathbb{R}^n \setminus f^{-1}(I) = f^{-1}(\mathbb{R} \setminus I) = f^{-1}(]-\infty, \alpha[\cup]\beta, +\infty[) = f^{-1}(]-\infty, \alpha[) \cup f^{-1}(]\beta, +\infty[).$$

D'après ce qui précède, $f^{-1}(]-\infty, \alpha[)$ et $f^{-1}(]\beta, +\infty[)$ sont des ouverts de \mathbb{R}^n , donc leur réunion est un ouvert et $f^{-1}(I)$ est fermé. □

Corollaire 5

Soient f une fonction continue sur \mathbb{R}^n , α un réel.

Les ensembles

$$\begin{aligned} &\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid f(x_1, \dots, x_n) < \alpha\} \text{ et} \\ &\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid f(x_1, \dots, x_n) > \alpha\} \end{aligned}$$

sont des ouverts de \mathbb{R}^n .

Les ensembles

$$\begin{aligned} &\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid f(x_1, \dots, x_n) = \alpha\}, \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid f(x_1, \dots, x_n) \leq \alpha\} \\ &\text{et } \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid f(x_1, \dots, x_n) \geq \alpha\} \end{aligned}$$

sont des fermés de \mathbb{R}^n .

Preuve

Les affirmations concernant les deux premiers ensembles résultent directement du théorème, car il s'agit des images réciproques des intervalles ouverts $] -\infty, \alpha[$ et $]\alpha, +\infty[$.

L'ensemble $\{(x_1, \dots, x_n) \in \mathbb{R}^n, f(x_1, \dots, x_n) = \alpha\}$ est fermé car c'est l'image réciproque de l'intervalle fermé $[\alpha, \alpha]$.

Enfin les deux derniers ensembles sont fermés car ce sont les complémentaires des ouverts

$$\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid f(x_1, \dots, x_n) > \alpha\} \text{ et}$$

$$\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid f(x_1, \dots, x_n) < \alpha\}$$

respectivement. □

► **Remarque**

Utiliser ce corollaire est souvent la méthode la plus simple pour démontrer qu'un sous-ensemble de \mathbb{R}^n est ouvert ou fermé.

Exemples

1. Pour tout $(u_1, \dots, u_n) \in \mathbb{R}^n \setminus \{(0, \dots, 0)\}$,

le demi-espace $\left\{ (x_1, \dots, x_n) \in \mathbb{R}^n \mid \sum_{k=1}^n u_k x_k + c > 0 \right\}$ est un ouvert de \mathbb{R}^n ,

le demi-espace $\left\{ (x_1, \dots, x_n) \in \mathbb{R}^n \mid \sum_{k=1}^n u_k x_k + c \geq 0 \right\}$ est un fermé de \mathbb{R}^n .

En effet, la fonction affine $(x_1, \dots, x_n) \mapsto \sum_{k=1}^n u_k x_k + c$ est continue sur \mathbb{R}^n .

2. Soit $A \in \mathbb{R}^n$. La fonction $f : M \mapsto d(A, M)$ est continue sur \mathbb{R}^n . On peut en déduire que, pour $r > 0$,

$$B(A, r) = \{M \in \mathbb{R}^n \mid f(M) < r\} \text{ est un ouvert de } \mathbb{R}^n,$$

$$B_f(A, r) = \{M \in \mathbb{R}^n \mid f(M) \leq r\} \text{ est un fermé de } \mathbb{R}^n.$$

On obtient une nouvelle démonstration de résultats obtenus dans le chapitre précédent.

Théorème 5 (admis)

Soient F une partie fermée et bornée de \mathbb{R}^n , $f : F \rightarrow \mathbb{R}$ une application continue. Alors, la fonction f est bornée sur F et atteint ses bornes. Il existe donc A et B dans F tels que

$$f(A) = \inf_{M \in F} f(M) \text{ et } f(B) = \sup_{M \in F} f(M).$$

Exemple

Si f est une application continue de \mathbb{R}^n dans \mathbb{R} , f est bornée et atteint ses bornes sur toute boule fermée $B_f(A, r)$ ou sur la sphère $S = \{M \in \mathbb{R}^n \mid \|M\| = 1\}$.

► **Remarque**

Si la fonction f est définie sur \mathbb{R}^n tout entier et non sur un fermé borné et est de plus minorée (resp. majorée), on peut parfois démontrer que f possède un minimum (resp. un maximum) en montrant que la borne inférieure (resp. supérieure) de f sur \mathbb{R}^n est égale à sa borne inférieure (resp. supérieure) sur un certain ensemble fermé borné où elle est atteinte.

Exemple

Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ définie par

$$f(x, y) = (2x^2 - y^2)e^{-(x^2+y^2)}.$$

On a, pour tout $(x, y) \in \mathbb{R}^2$,

$$|f(x, y)| \leq 2(x^2 + y^2)e^{-(x^2+y^2)} \leq 2\|(x, y)\|^2 e^{-\|(x, y)\|^2}.$$

L'étude de la fonction $t \longmapsto t^2 e^{-t^2}$ montre que son maximum sur \mathbb{R}_+ est atteint en 1 et vaut $\frac{1}{e}$. On a donc, pour tout $(x, y) \in \mathbb{R}^2$,

$$|f(x, y)| \leq \frac{2}{e}.$$

La fonction f est bornée. On pose $M = \sup_{(x,y) \in \mathbb{R}^2} f(x, y)$ et $m = \inf_{(x,y) \in \mathbb{R}^2} f(x, y)$.

On a $M > 0$, car $f(1, 0) = 2e^{-1} > 0$. Comme $\lim_{t \rightarrow +\infty} 2t^2 e^{-t^2} = 0$, il existe $R > 0$ tel que, pour tout réel $t \geq R$, $2t^2 e^{-t^2} \leq \frac{M}{2}$. On a donc, pour tout $(x, y) \in \mathbb{R}^2$, tel que $\|(x, y)\| \geq R$,

$$f(x, y) \leq 2\|(x, y)\|^2 e^{-\|(x, y)\|^2} \leq \frac{M}{2}.$$

On en déduit que

$$M = \sup_{\|(x,y)\| \leq R} f(x, y) = \sup_{(x,y) \in B_f(0,R)} f(x, y).$$

En effet M est clairement un majorant de f sur $B_f(0, R)$ et si f possédait sur $B_f(0, R)$ un majorant $M' < M$, f serait majorée sur \mathbb{R}^2 par $\max\left(M', \frac{M}{2}\right) < M$, ce qui est contraire à la définition de M .

Puisque f est continue et que $B_f(0, R)$ est un fermé borné, f atteint ses bornes sur $B_f(0, R)$. Ainsi il existe $(x_0, y_0) \in B_f(0, R)$ tel que $M = f(x_0, y_0)$.

On montrerait de même que $m < 0$ et qu'il existe R' tel que $\|(x, y)\| \geq R'$ implique $f(x, y) \geq \frac{m}{2}$. Ainsi m est la borne inférieure de f sur $B_f(0, R')$ et est donc atteinte.

Pour le cas $n = 2$, on pourra se reporter au livre de première année (chapitre 27, exercices pages 557 à 559).

1. Étudier et tracer les lignes de niveau de l'application $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dans les cas suivants :

1. $f(x, y) = 3x + 2y$;
2. $f(x, y) = 3|x| + 2|y|$;
3. $f(x, y) = 9x^2 + 4y^2$;
4. $f(x, y) = 9x^2 - 4y^2$.

2. Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ la fonction définie par

$$f(x, y) = xy e^{-x-y}.$$

Pour tout réel λ , on note Γ_λ la ligne de niveau λ de f , c'est-à-dire l'ensemble des éléments $(x, y) \in \mathbb{R}^2$ tels que $f(x, y) = \lambda$.

1. Soit φ la fonction définie sur \mathbb{R} par

$$\varphi(x) = x e^{-x}.$$

- a) Étudier les variations de φ . Discuter selon les valeurs de λ le nombre de solutions de l'équation $\varphi(x) = \lambda$.
- b) Montrer que la restriction ψ de φ à $] -\infty, 0[$ définit une bijection de $] -\infty, 0[$ sur lui-même. Montrer que la bijection réciproque ψ^{-1} est de classe \mathcal{C}^1 .

Pour tout $\lambda \neq 0$, on note Γ_λ^+ (respectivement Γ_λ^-) l'intersection de Γ_λ et du demi-plan d'équation $x > 0$ (respectivement $x < 0$).

2. Soit $\lambda > 0$.

- a) Montrer que Γ_λ^+ est non vide si, et seulement si, $\lambda \leq \frac{1}{e^2}$.
Montrer qu'alors Γ_λ^+ est une partie bornée de \mathbb{R}^2 .
- b) Établir que Γ_λ^- peut être caractérisée par une équation de la forme $y = g_\lambda(x)$ où g_λ est une fonction de classe \mathcal{C}^1 sur $] -\infty, 0[$, qu'on exprimera en fonction de φ et ψ^{-1} . Préciser les variations de g_λ et ses limites en $-\infty$ et 0 .

3. On note (X, Y) les coordonnées de $(x, y) \in \mathbb{R}^2$ dans la base $((1, 1), (-1, 1))$.

- a) Pour tout réel λ , caractériser Γ_λ par une équation de la forme $F(X, Y) = \lambda$.
- b) On suppose $\lambda < 0$. Montrer que Γ_λ^- se caractérise par une équation de la forme $Y = G_\lambda(X)$ où G_λ est une fonction de classe \mathcal{C}^∞ sur \mathbb{R} dont on précisera les variations et les branches infinies à l'aide de φ . Comment obtient-on Γ_λ^+ ?
- c) On suppose $0 < \lambda \leq \frac{1}{e^2}$. Caractériser Γ_λ^+ par une équation de la forme $Y = \pm G_\lambda(X)$, où G_λ est une fonction définie sur une partie de \mathbb{R}^+ dont on étudiera les variations à l'aide de φ .

4. Indiquer sur un graphique l'allure des diverses lignes de niveau de la fonction f .

3. Étudier la continuité en $(0, 0)$ des fonctions suivantes.

1. $f(x, y) = \begin{cases} \frac{xy^4}{x^4 + y^6} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{sinon.} \end{cases}$
2. $f(x, y) = \begin{cases} \frac{x^3 + y^3}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{sinon.} \end{cases}$
3. $f(x, y) = \begin{cases} \frac{\sin(xy)}{|x| + |y|} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{sinon.} \end{cases}$
4. $f(x, y) = \begin{cases} xy \ln(x^2 + y^2) & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{sinon.} \end{cases}$
5. $f(x, y) = \begin{cases} \frac{x \sin y - y \sin x}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{sinon.} \end{cases}$

4. Étudier la continuité des fonctions suivantes.

1. $f(x, y) = \begin{cases} ye^{\text{Arctan } \frac{x}{y}} & \text{si } y \neq 0 \\ 0 & \text{sinon.} \end{cases}$
2. $f(x, y) = \begin{cases} \frac{\ln(1 + x^2 y^2)}{y^2} & \text{si } y \neq 0 \\ x^2 & \text{sinon.} \end{cases}$
3. $f(x, y) = \begin{cases} (x + y) \sin \frac{1}{x} \sin \frac{1}{y} & \text{si } xy \neq 0 \\ 0 & \text{sinon.} \end{cases}$
4. $f(x, y) = \begin{cases} y - \varphi(x) & \text{si } y \leq \varphi(x) \\ 0 & \text{sinon,} \end{cases}$
où φ est une fonction continue de \mathbb{R} dans \mathbb{R} .

5. 1. Soient p, q, r trois réels strictement positifs et f la fonction définie sur \mathbb{R}^3 par

$$f(x, y, z) = \frac{|x|^p |y|^q |z|^r}{\sqrt{x^2 + y^2 + z^2}} \quad \text{si } (x, y, z) \neq (0, 0, 0) \quad \text{et} \quad f(0, 0, 0) = 0.$$

- a) Montrer que $|f(x, y, z)| \leq \|(x, y, z)\|^{p+q+r-1}$; calculer $f(x, x, x)$ pour $x \neq 0$.
 - b) En déduire les valeurs de p, q, r pour lesquelles f est continue en $(0, 0, 0)$.
2. Soient p, q, r, s, t, u quatre réels strictement positifs et g la fonction définie sur \mathbb{R}^3 par

$$g(x, y, z) = \frac{|x|^p |y|^q |z|^r}{|x|^s + |y|^t + |z|^u} \quad \text{si } (x, y, z) \neq (0, 0, 0) \quad \text{et} \quad g(0, 0, 0) = 0.$$

Calculer, pour $x > 0$, $g\left(x^{\frac{1}{s}}, x^{\frac{1}{t}}, x^{\frac{1}{u}}\right)$.

Déterminer les valeurs de p, q, r, s, t, u pour lesquelles g est continue en $(0, 0, 0)$.

6. Soit $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ l'application définie par

$$\forall (x, y, z) \in \mathbb{R}^3, f(x, y, z) = \max_{t \in [0,1]} (xt^2 + yt + z).$$

Donner l'expression de $f(x, y, z)$. Montrer que f est continue sur \mathbb{R}^3 .

7. Soit $C = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$ et f une application continue de C dans \mathbb{R} .

1. Soit $g : [0, 2\pi] \rightarrow \mathbb{R}$ définie par

$$g(t) = f(\cos t, \sin t) - f(-\cos t, -\sin t).$$

Montrer que g est continue. Comparer les signes de $g(0)$ et $g(\pi)$.

2. En déduire qu'il existe $(x, y) \in C$ tel que $f(x, y) = f(-x, -y)$.

8. Soient f une fonction définie et continue sur une partie convexe Ω de \mathbb{R}^n .

1. Soit $(x, y) \in \Omega^2$. Montrer que l'application $\varphi : t \mapsto f((1-t)x + ty)$ est définie et continue sur $[0, 1]$.

2. En déduire que $f(\Omega)$ est un intervalle de \mathbb{R} .

9. Soient $s \in \mathbb{R}_+^*$, $K = \left\{ x = (x_1, \dots, x_n) \in \mathbb{R}_+^n \mid \sum_{i=1}^n x_i = s \right\}$ et $f : K \rightarrow \mathbb{R}$ définie

$$\text{par } f(x) = \prod_{i=1}^n x_i.$$

1. Montrer que K est fermé et borné. En déduire que $f(K)$ possède un maximum.

2. Déterminer ce maximum en utilisant l'inégalité $\sqrt{xy} \leq \frac{x+y}{2}$, valable pour x et y positifs et en raisonnant par récurrence sur n . En quel point ce maximum est-il atteint?

3. En déduire que, pour $(x_1, \dots, x_n) \in \mathbb{R}_+^n$, on a

$$\left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{i=1}^n x_i.$$

10. Soit $\varphi : [0, 1] \rightarrow \mathbb{R}$ continue, $n \in \mathbb{N}$.

On pose, pour tout $a = (a_0, a_1, \dots, a_n) \in \mathbb{R}^{n+1}$,

$$f(a) = \sup_{t \in [0,1]} \left| \varphi(t) - \sum_{k=0}^n a_k t^k \right|.$$

1. Justifier l'existence de $f(a)$.

2. Soient a et b dans \mathbb{R}^{n+1} . Montrer que, pour tout $t \in [0, 1]$, on a

$$\begin{aligned} \left| \varphi(t) - \sum_{k=0}^n a_k t^k \right| &\leq \left| \varphi(t) - \sum_{k=0}^n b_k t^k \right| + \sum_{k=0}^n |b_k - a_k| t^k \\ &\leq f(b) + \sum_{k=0}^n |b_k - a_k| \\ &\leq f(b) + \sqrt{n+1} \|b - a\|. \end{aligned}$$

En déduire que $f(a) \leq f(b) + \sqrt{n+1} \|b - a\|$, puis que

$$|f(b) - f(a)| \leq \sqrt{n+1} \|b - a\|.$$

Montrer que f est continue.

3. On définit la fonction $g : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ par

$$g(a) = \sup_{t \in [0,1]} \left| \sum_{k=0}^n a_k t^k \right|.$$

On remarquera que la fonction g est continue sur \mathbb{R}^{n+1} car c'est un cas particulier de la fonction f correspondant à $\varphi = 0$.

- a) Montrer que $g(a) > 0$ pour tout $a \neq 0$. On pose $\mu = \inf_{\|a\|=1} g(a)$. Justifier l'existence de cette borne inférieure. Montrer qu'elle est atteinte. En déduire que $\mu > 0$.
Montrer que, pour tout $a \in \mathbb{R}^{n+1}$, on a $g(a) \geq \mu \|a\|$.
- b) En déduire que, pour tout $a \in \mathbb{R}^{n+1}$,

$$f(a) \geq g(a) - K \geq \mu \|a\| - K,$$

où $K = \sup_{t \in [0,1]} |\varphi(t)|$. Montrer que $\lim_{\|a\| \rightarrow +\infty} f(a) = +\infty$.

4. Déduire de ce qui précède que f possède un minimum.

Si le minimum de f est atteint en a , le polynôme $\sum_{k=0}^n a_k X^k$ est appelé un polynôme de meilleure approximation de φ .

11. Théorème de d'Alembert Soit P un polynôme non constant de $\mathbb{C}[X]$. On veut démontrer que P possède au moins une racine sur \mathbb{C} .

1. On note $m = \inf_{z \in \mathbb{C}} |P(z)|$.

- a) Justifier l'existence de m .
- b) Soit p le degré de P , a_p son coefficient dominant. Montrer qu'il existe $r > 0$ tel que, pour tout $z \in \mathbb{C}$

$$|z| \geq r \implies |P(z)| \geq \frac{1}{2} |a_p| |z|^p.$$

En déduire qu'il existe $r' > 0$ tel que pour tout $z \in \mathbb{C}$

$$|z| \geq r' \implies |P(z)| \geq m + 1.$$

c) Soit f la fonction

$$\begin{aligned} \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto |P(x + iy)|. \end{aligned}$$

Montrer que f est continue sur \mathbb{R}^2 . Montrer que f atteint sa borne inférieure. En déduire qu'il existe $z_0 \in \mathbb{C}$ tel que $|P(z_0)| = m$.

2. On suppose que $P(z_0) \neq 0$. On considère le polynôme Q défini par $Q(z) = \frac{P(z_0 + z)}{P(z_0)}$

pour tout $z \in \mathbb{C}$. On pose $Q(z) = \sum_{i=0}^p b_i z^i$.

a) Montrer que $b_0 = 1$ et $b_p \neq 0$.

On note k le plus petit indice $i \geq 1$ tel que $b_i \neq 0$ et ω un nombre complexe tel que $\omega^k = -\frac{1}{b_k}$.

b) Montrer que, pour tout réel $x \in]0, 1[$, on a

$$Q(\omega x) = 1 - x^k + \sum_{i=k+1}^n b_i \omega^i x^i, \text{ puis } |Q(\omega x)| \leq 1 - x^k + \sum_{i=k+1}^n |b_i \omega^i| x^i.$$

En déduire que $|Q(\omega x)| < 1$ pour x assez petit.

c) Déduire de ce qui précède que $P(z_0) = 0$. Conclure.

12.

1. Soit $x_0 \in \mathbb{R}^n$. Démontrer que l'application définie sur \mathbb{R}^n par

$$f(x) = d(x_0, x)$$

est continue.

2. Soit Ω une partie non vide de \mathbb{R}^n . Pour tout $x \in \mathbb{R}^n$, on pose

$$d(x, \Omega) = \inf_{u \in \Omega} d(x, u).$$

Montrer que, pour tous points x et x' de \mathbb{R}^n , on a

$$|d(x, \Omega) - d(x', \Omega)| \leq d(x, x').$$

En déduire l'application $x \mapsto d(x, \Omega)$ est continue sur \mathbb{R}^n .

3. a) Montrer que si Ω est un fermé borné, alors, pour tout $x \in \mathbb{R}^n$, il existe $y \in \Omega$ tel que $d(x, \Omega) = d(x, y)$.

b) On suppose seulement que Ω est fermé. Soit $x \in \mathbb{R}^n$ et $\alpha = d(x, \Omega)$. En écrivant que, pour tout $u \in \Omega$, $d(x, u) \geq \|u\| - \|x\|$, montrer que

$$\alpha = \inf_{u \in B \cap \Omega} d(x, u),$$

où B est la boule fermée de centre 0 et de rayon $\alpha + 1 + \|x\|$.

En déduire qu'il existe $y \in \Omega$ tel que

$$d(x, \Omega) = d(x, y).$$

4. Soit Ω un fermé et Ω' un fermé borné de \mathbb{R}^n . On suppose que Ω et Ω' sont disjoints. Montrer que $\inf_{(x,y) \in \Omega \times \Omega'} d(x, y) > 0$.

13. Soit C une partie convexe, fermée, non vide de \mathbb{R}^n . Pour x élément de \mathbb{R}^n , on définit la distance de x à C , notée $d(x, C)$, par

$$d(x, C) = \inf_{u \in C} d(x, u).$$

1. Soit $x \in C$. Montrer qu'il existe $y \in C$ tel que

$$d(x, C) = d(x, y)$$

(se reporter à la question 3 de l'exercice 12).

2. Montrer qu'un tel y est unique. S'il en existe un autre y' , on montrera que

$$\|x - y\|^2 + \|x - y'\|^2 = 2 \left\| x - \frac{1}{2}(y + y') \right\|^2 + \frac{1}{2} \|y - y'\|^2.$$

On pose $y = \text{proj}_C(x)$ et on appelle projection sur C l'application qui à x associe y .

3. Soit C un sous-espace vectoriel de \mathbb{R}^n .

- a) Montrer que C est un convexe, fermé, non vide de \mathbb{R}^n .
 b) Montrer que proj_C coïncide avec la projection orthogonale sur C .

4. Soit x un élément de \mathbb{R}^n . On pose $y = \text{proj}_C(x)$.

- a) Soit u un élément de C . En remarquant que, pour tout $t \in [0, 1]$

$$\|y - x\| \leq \|(1 - t)y + tu - x\|,$$

montrer que

$$\langle y - x, u - y \rangle \geq 0.$$

- b) Montrer réciproquement que si y est un élément de C tel que, pour tout $u \in C$,

$$\langle y - x, u - y \rangle \geq 0,$$

alors $y = \text{proj}_C(x)$.

Fonctions de n variables : calcul différentiel

6

Dans ce chapitre les fonctions seront définies sur un ouvert de \mathbb{R}^n et à valeurs dans \mathbb{R} .

1. Calcul différentiel d'ordre 1

1.1 Dérivées partielles d'ordre 1

Définitions

Définition 1

Soit $f : D \rightarrow \mathbb{R}$, où D est un ouvert de \mathbb{R}^n , $A = (a_1, \dots, a_n) \in D$. Pour $i \in \llbracket 1, n \rrbracket$, on note

$$D_i = \{x \in \mathbb{R} \mid (a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n) \in D\}$$

Les applications partielles de f en A sont les applications

$$\begin{aligned} D_i &\longrightarrow \mathbb{R} \\ f_i : x &\longmapsto f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n). \end{aligned}$$

► Remarque

Par définition d'un ouvert, on peut trouver une boule ouverte $B(A, r)$ incluse dans D . Si $|x - a_i| < r$, alors

$$d(A, (a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)) = |x - a_i| \leq r,$$

donc $(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)$ appartient à D et x est dans D_i . L'application f_i est donc définie au voisinage de a_i .

Définition 2

Soit D un ouvert de \mathbb{R}^n , $f : D \rightarrow \mathbb{R}$, $A = (a_1, \dots, a_n)$ un point de D .

Si la fonction partielle $f_i : x \mapsto f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)$ est dérivable en a_i , le

nombre dérivé s'appelle dérivée partielle de f par rapport à x_i en A et se note $\frac{\partial f}{\partial x_i}(A)$.

► **Remarque**

Pratiquement, on fixe $n - 1$ variables qui jouent le rôle de paramètres et l'on dérive par rapport à la variable restante.

Définition 3

Soit D un ouvert de $\mathbb{R}^n, f : D \longrightarrow \mathbb{R}, i \in \llbracket 1, n \rrbracket$.

Si en tout point M de D, f possède des dérivées partielles par rapport à x_i , on note $\frac{\partial f}{\partial x_i}$ l'application

$$\begin{cases} D & \longrightarrow & \mathbb{R} \\ M & \longmapsto & \frac{\partial f}{\partial x_i}(M). \end{cases}$$

Exemples

1. Soit $(u_1, \dots, u_n, c) \in \mathbb{R}^{n+1}$ et f la fonction affine définie par

$$f(x_1, \dots, x_n) = u_1x_1 + \dots + u_nx_n + c.$$

En tout point $A = (a_1, \dots, a_n)$ de \mathbb{R}^n , f possède des dérivées partielles qui vérifient $\frac{\partial f}{\partial x_i}(A) = u_i$.

2. En tout point $A = (a_1, \dots, a_n)$ différent de $(0, \dots, 0)$, la fonction f définie sur \mathbb{R}^n par $f(X) = \|X\| = \sqrt{x_1^2 + \dots + x_n^2}$ possède des dérivées partielles par rapport à chacune des variables

$$\frac{\partial f}{\partial x_i}(A) = \frac{a_i}{\sqrt{a_1^2 + \dots + a_n^2}} = \frac{a_i}{\|A\|}.$$

En $(0, \dots, 0)$, la fonction ne possède pas de dérivée partielle. En effet, la fonction partielle en $(0, \dots, 0), f_i : x_i \mapsto |x_i|$ n'est pas dérivable en 0.

3. Si la fonction φ est dérivable sur \mathbb{R} , la fonction f définie sur \mathbb{R}^2 par $f(x, y) = \varphi(xy^2)$ possède des dérivées partielles en tout point de \mathbb{R}^2

$$\frac{\partial f}{\partial x}(x, y) = y^2\varphi'(xy^2) \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) = 2xy\varphi'(xy^2).$$

4. La fonction f définie par

$$f(x, y, z) = \begin{cases} \frac{xyz}{x^2 + y^4 + z^4} & \text{si } (x, y, z) \neq (0, 0, 0) \\ 0 & \text{si } (x, y, z) = (0, 0, 0). \end{cases}$$

possède des dérivées partielles nulles en $(0, 0, 0)$ puisque les applications partielles en $(0, 0, 0)$ sont nulles. Néanmoins, comme il a été démontré précédemment (cf exemple page 116), elle n'est pas continue en $(0, 0, 0)$.

Définition 4

Soit D un ouvert de \mathbb{R}^n , $f : D \rightarrow \mathbb{R}$ et A un point de D . Si f admet des dérivées partielles par rapport aux n variables en A , on appelle gradient de f en A le vecteur de \mathbb{R}^n

$$\left(\frac{\partial f}{\partial x_1}(A), \dots, \frac{\partial f}{\partial x_n}(A) \right).$$

On le note ∇f_A ou $\nabla f(A)$.

Exemples

1. Soit $U \in \mathbb{R}^n$ et $c \in \mathbb{R}$. Si f est la fonction affine $X \mapsto \langle X, U \rangle + c$, on a, pour tout $X \in \mathbb{R}^n$, $\nabla f_X = U$.
2. Si f est définie sur \mathbb{R}^n par $f(X) = \|X\|^2$, on a pour tout $X \in \mathbb{R}^n$,

$$\nabla f_X = 2X,$$

car pour tout $1 \leq i \leq n$, l'application partielle $x_i \mapsto \sum_{k=1}^n x_k^2$ a pour dérivée $x_i \mapsto 2x_i$.

Dérivées directionnelles en A **Définition 5**

Soit D un ouvert de \mathbb{R}^n , $f : D \rightarrow \mathbb{R}$, A un point de D , U un vecteur unitaire.

Si la fonction φ définie par $\varphi(t) = f(A + tU)$ est dérivable en 0, le nombre dérivé est appelé dérivée de f en A dans la direction U et est notée $f'_U(A)$.

Exemples

1. Soit $f : D \rightarrow \mathbb{R}$, $A \in D$. Si on prend $U = e_i$, i -ième vecteur de la base canonique de \mathbb{R}^n , on obtient $\varphi(t) = f(A + te_i) = f_i(a_i + t)$. La fonction f a une dérivée en A dans la direction e_i si, et seulement si, elle possède en A une dérivée partielle par rapport à x_i et l'on a alors $f'_i(A) = \frac{\partial f}{\partial x_i}(A)$.
2. Soit $f : (x_1, \dots, x_n) \mapsto \alpha_1 x_1 + \dots + \alpha_n x_n + c$ une fonction affine, $A = (a_1, \dots, a_1)$ et $U = (u_1, \dots, u_n)$. On a, pour tout $t \in \mathbb{R}$,

$$\varphi(t) = \alpha_1(a_1 + tu_1) + \dots + \alpha_n(a_n + tu_n) + c$$

et donc $f'_U(A) = \alpha_1 u_1 + \dots + \alpha_n u_n$.

Opérations sur les dérivées partielles**Proposition 1**

Soient D un ouvert de \mathbb{R}^n , $f, g : D \rightarrow \mathbb{R}$.

Si f et g possèdent sur D des dérivées partielles par rapport à la i -ième variable x_i , alors $f + g, fg, \lambda f$ ($\lambda \in \mathbb{R}$) possèdent aussi des dérivées partielles par rapport à x_i sur D et l'on a

$$\frac{\partial(f+g)}{\partial x_i} = \frac{\partial f}{\partial x_i} + \frac{\partial g}{\partial x_i}, \quad \frac{\partial(\lambda f)}{\partial x_i} = \lambda \frac{\partial f}{\partial x_i}, \quad \frac{\partial(fg)}{\partial x_i} = f \frac{\partial g}{\partial x_i} + g \frac{\partial f}{\partial x_i}.$$

Si g ne s'annule pas sur D , alors $\frac{1}{g}$ et $\frac{f}{g}$ possèdent des dérivées partielles par rapport à x_i et l'on a

$$\frac{\partial}{\partial x_i} \left(\frac{1}{g} \right) = \frac{-1}{g^2} \frac{\partial g}{\partial x_i} \quad \text{et} \quad \frac{\partial}{\partial x_i} \left(\frac{f}{g} \right) = \frac{1}{g^2} \left(\frac{\partial f}{\partial x_i} g - f \frac{\partial g}{\partial x_i} \right).$$

Preuve

Ces propriétés découlent immédiatement de celles des dérivées des fonctions d'une variable réelle, puisque la dérivée partielle par rapport à x_i est la dérivée d'une fonction d'une variable, obtenue en fixant les $n - 1$ autres

Développement limité d'ordre 1

Définition 6

Soit D un ouvert de \mathbb{R}^n , $f : D \rightarrow \mathbb{R}$, $A \in D$.

S'il existe une application linéaire ℓ de \mathbb{R}^n dans \mathbb{R} et une fonction ε tendant vers 0 en $(0, \dots, 0)$ telle, pour tout $H \in \mathbb{R}^n$ tel que $A + H \in D$,

$$f(A + H) = f(A) + \ell(H) + \|H\| \varepsilon(H),$$

on dit que f possède un développement limité d'ordre 1 en A .

Exemples

- Supposons que f est une application affine. Il existe $(u_1, \dots, u_n, c) \in \mathbb{R}^{n+1}$ tels que pour tout $X = (x_1, \dots, x_n) \in \mathbb{R}^n$,

$$f(X) = \sum_{i=1}^n u_i x_i + c.$$

En notant U le vecteur (u_1, \dots, u_n) , on obtient, pour tout $X \in \mathbb{R}^n$, $f(X) = \langle U, X \rangle + c$. Pour A et H dans \mathbb{R}^n , on a donc

$$f(A + H) = \langle U, A + H \rangle + c = f(A) + \langle U, H \rangle = f(A) + \ell(H),$$

où ℓ est l'application linéaire $H \mapsto \langle U, H \rangle$. La fonction f possède un développement limité d'ordre 1, la fonction ε étant nulle.

- Soit f la fonction définie sur \mathbb{R}^n par $f(X) = \|X\|^2$. Montrons que f possède un développement limité d'ordre 1 au voisinage de tout point $A \in \mathbb{R}^n$. On a, pour $H \in \mathbb{R}^n$,

$$f(A + H) = \|A + H\|^2 = \|A\|^2 + 2\langle A, H \rangle + \|H\|^2 = f(A) + \ell(H) + \|H\| \varepsilon(H),$$

où ℓ est l'application linéaire définie par $\ell(H) = 2\langle A, H \rangle$ et $\varepsilon(H) = \|H\|$; on a $\lim_{H \rightarrow (0, \dots, 0)} \varepsilon(H) = 0$ donc f possède un développement limité d'ordre 1 en A .

Proposition 2

Soit D un ouvert de \mathbb{R}^n , $f : D \rightarrow \mathbb{R}$, $A \in D$. On suppose que f possède un développement limité d'ordre 1 au voisinage de A donné par

$$f(A + H) = f(A) + \ell(H) + \|H\| \varepsilon(H).$$

Alors f est continue en A et possède en A des dérivées partielles d'ordre 1 par rapport à chacune des variables; si ℓ est définie par $\ell(h_1, \dots, h_n) = u_1 h_1 + \dots + u_n h_n$, on a pour tout i ,

$$u_i = \frac{\partial f}{\partial x_i}(A).$$

Le développement limité à l'ordre 1 est donc unique.

Preuve

On a

$$\lim_{H \rightarrow (0, \dots, 0)} f(A + H) = \lim_{H \rightarrow (0, \dots, 0)} f(A) + \ell(H) + \|H\| \varepsilon(H) = f(A),$$

car l'application linéaire ℓ est continue donc sa limite en $(0, \dots, 0)$ est nulle. La fonction f est donc continue en A .

La i -ième fonction partielle en $A = (a_1, \dots, a_n)$ vérifie

$$f_i(a_i + h_i) = f(A + (0, \dots, 0, h_i, 0, \dots, 0)) = f(A) + u_i h_i + |h_i| \varepsilon(0, \dots, h_i, 0, \dots, 0).$$

Ainsi la fonction f_i possède un développement limité d'ordre 1 au voisinage de a_i . Elle est donc dérivable en a_i de nombre dérivé u_i . Par définition $\frac{\partial f}{\partial x_i}(A)$ existe et est égal à u_i . \square

► Remarque

L'existence de dérivées partielles ne suffit pas à assurer l'existence d'un développement limité. On a vu (dernier exemple page 126) qu'elle n'assurait même pas la continuité.

Corollaire 1

Si f est une fonction définie sur l'ouvert D de \mathbb{R}^n , qui possède en $A \in D$ un développement limité d'ordre 1, ce développement limité s'écrit nécessairement

$$f(A + H) = f(A) + \langle \nabla f_A, H \rangle + \|H\| \varepsilon(H),$$

avec $\lim_{H \rightarrow 0} \varepsilon(H) = 0$.

Preuve

Cela résulte de la proposition précédente puisque, avec les notations précédentes,

$$\ell(H) = \sum_{i=1}^n u_i h_i = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(A) h_i = \langle \nabla f_A, H \rangle. \quad \square$$

Définition 7

Soit D un ouvert de \mathbb{R}^n , $f : D \rightarrow \mathbb{R}$, $A \in D$. Si f possède un développement limité d'ordre 1 en A , la fonction

$$X \mapsto f(A) + \langle \nabla f_A, X - A \rangle,$$

c'est-à-dire

$$(x_1, \dots, x_n) \mapsto f(A) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(A) (x_i - a_i)$$

est appelée meilleure approximation affine de f au voisinage de A .

L'hyperplan de \mathbb{R}^{n+1} d'équation

$$x_{n+1} = f(A) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(A)(x_i - a_i)$$

est appelé hyperplan tangent au graphe de f en A .

► **Remarque**

Si f possède un développement limité au voisinage de A , f est la somme au voisinage de A de sa meilleure approximation affine et du terme $\|H\|\varepsilon(H)$ négligeable devant $\|H\|$.

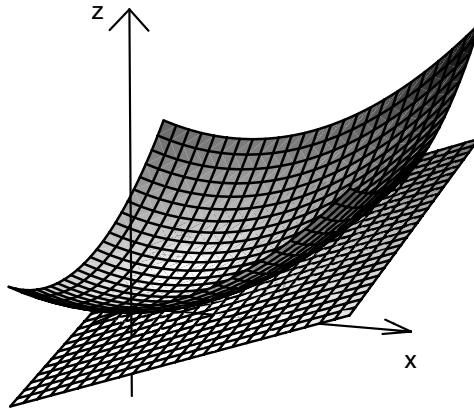
Exemple

Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est définie par $f(X) = \|X\|^2$, la meilleure approximation affine de f en A est

$$X \mapsto \|A\|^2 + 2\langle A, X - A \rangle = -\|A\|^2 + 2\langle A, X \rangle.$$

Un équation de l'hyperplan tangent en $A = (a_1, \dots, a_n)$ est

$$x_{n+1} = 2 \sum_{i=1}^n a_i x_i - \sum_{i=1}^n a_i^2.$$



1.2 Fonctions de classe \mathcal{C}^1

Définition - Opérations sur les fonctions de classe \mathcal{C}^1

Définition 8

Soit D un ouvert de \mathbb{R}^n , $f : D \rightarrow \mathbb{R}$. On dit que f est de classe \mathcal{C}^1 sur D si elle admet en tout point de D des dérivées partielles d'ordre 1 et si les fonctions $\frac{\partial f}{\partial x_i}$ ($1 \leq i \leq n$) sont continues sur D .

Proposition 3

Soit D un ouvert de \mathbb{R}^n , $f, g : D \rightarrow \mathbb{R}$.

Si f et g sont de classe \mathcal{C}^1 sur D , alors $f + g$, λf ($\lambda \in \mathbb{R}$) et fg sont de classe \mathcal{C}^1 sur D .

Si de plus g ne s'annule pas sur D , $\frac{1}{g}$ et $\frac{f}{g}$ sont de classe \mathcal{C}^1 .

Preuve

Cette proposition découle immédiatement de la proposition 1 et des propriétés de la continuité des fonctions de n variables. \square

Exemples

1. Les fonctions polynomiales sont de classe \mathcal{C}^1 sur \mathbb{R}^n et les quotients de fonctions polynomiales sont de classe \mathcal{C}^1 sur leur ensemble de définition.
2. Si f est de classe \mathcal{C}^1 sur l'intervalle ouvert I de \mathbb{R} , la fonction $g : I \times \mathbb{R} \rightarrow \mathbb{R}$ définie par $g(x, y) = f(x)$ est de classe \mathcal{C}^1 . En effet, on a, pour $(x, y) \in I \times \mathbb{R}$,

$$\frac{\partial g}{\partial x}(x, y) = f'(x) \quad \text{et} \quad \frac{\partial g}{\partial y}(x, y) = 0.$$

De même, la fonction $(x, y) \mapsto f(y)$ est \mathcal{C}^1 sur $\mathbb{R} \times I$.

Proposition 4

Soient $f : D \rightarrow \mathbb{R}$ une fonction définie sur un ouvert D de \mathbb{R}^n , $\varphi : I \rightarrow \mathbb{R}$, où I est un intervalle de \mathbb{R} . On suppose que $f(D) \subset I$.

Si f est de classe \mathcal{C}^1 sur D et si φ est de classe \mathcal{C}^1 sur I , alors la fonction $\varphi \circ f$ est de classe \mathcal{C}^1 sur D .

On a, pour $1 \leq i \leq n$,

$$\frac{\partial}{\partial x_i}(\varphi \circ f) = (\varphi' \circ f) \times \frac{\partial f}{\partial x_i}$$

Preuve

Cette proposition découle du théorème de dérivation des fonctions composées d'une variable qu'on applique à la fonction $x_i \mapsto (\varphi \circ f)(x_1, \dots, x_n)$. \square

Existence d'un développement limité d'ordre 1**Théorème 1 (admis)**

Soient D un ouvert de \mathbb{R}^n , f une fonction de classe \mathcal{C}^1 sur D , A un élément de D .

Alors f possède au voisinage de A un développement limité. Il existe une fonction ε ayant pour limite 0 en $(0, \dots, 0)$ telle que, pour tout H de \mathbb{R}^n tel que $A + H \in D$,

$$f(A + H) = f(A) + \langle \nabla f_A, H \rangle + \|H\| \varepsilon(H).$$

► **Remarque**

Si on écrit $A = (a_1, \dots, a_n)$ et $H = (h_1, \dots, h_n)$ le développement limité d'ordre 1 devient

$$f(a_1 + h_1, \dots, a_n + h_n) = f(a_1, \dots, a_n) + \sum_{i=1}^n h_i \frac{\partial f}{\partial x_i}(a_1, \dots, a_n) + \sqrt{\sum_{i=1}^n h_i^2} \varepsilon(h_1, \dots, h_n).$$

Exemple

Soit $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ définie par $f(x, y, z) = xyz$. Cette fonction est de classe \mathcal{C}^1 car polynomiale. Déterminons son développement limité d'ordre 1 en $A = (1, 1, 1)$. On a, pour tout $(x, y, z) \in \mathbb{R}^3$,

$$\frac{\partial f}{\partial x}(x, y, z) = yz, \quad \frac{\partial f}{\partial y}(x, y, z) = xz, \quad \frac{\partial f}{\partial z}(x, y, z) = xy.$$

On en déduit $f(A) = 1$, $\nabla f_A = (1, 1, 1)$ et le développement limité

$$f(1 + h, 1 + k, 1 + \ell) = 1 + h + k + \ell + \sqrt{h^2 + k^2 + \ell^2} \varepsilon(h, k, \ell),$$

où $\lim_{(h,k,\ell) \rightarrow (0,0,0)} \varepsilon(h, k, \ell) = 0$.

Corollaire 2

Si f est une fonction de classe \mathcal{C}^1 sur l'ouvert D de \mathbb{R}^n , f est continue sur D .

Preuve

D'après le théorème 1, f possède en tout point A de D un développement limité d'ordre 1. On en déduit, d'après la proposition 2, que f est continue en A . □

Dérivée d'une fonction composée

Théorème 2

Soit D un ouvert de \mathbb{R}^n , f une fonction de classe \mathcal{C}^1 sur D , u_1, \dots, u_n des fonctions dérivables sur un intervalle I de \mathbb{R} et vérifiant, pour tout $t \in I$, $(u_1(t), \dots, u_n(t)) \in D$.

La fonction g de I dans \mathbb{R} définie par

$$g(t) = f(u_1(t), \dots, u_n(t))$$

est dérivable sur I et pour tout $t \in I$,

$$g'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(u_1(t), \dots, u_n(t)) u_i'(t).$$

Preuve

Soit $t_0 \in I$, $a_i = u_i(t_0)$, $A = (a_1, \dots, a_n)$. Pour montrer que g est dérivable en t_0 , on montre qu'elle possède un développement limité d'ordre 1 au voisinage de t_0 .

Comme les fonctions u_i sont dérivables en t_0 , elle possèdent des développements limités d'ordre 1 au voisinage de t_0 . Il existe des fonctions ε_i ayant pour limite 0 en 0, telles que pour tout réel h vérifiant $t_0 + h \in I$,

$$u_i(t_0 + h) = a_i + hu_i'(t_0) + h\varepsilon_i(h).$$

D'après le théorème 1, il existe une fonction ε ayant pour limite 0 en $(0, \dots, 0)$ telle que, pour $K = (k_1, \dots, k_n) \in \mathbb{R}^n$, vérifiant $A + K \in D$,

$$f(A + K) = f(A) + \sum_{k=1}^n \frac{\partial f}{\partial x_i}(A)k_i + \|K\|\varepsilon(K).$$

Si $t_0 + h \in I$, $(u_1(t_0 + h), \dots, u_n(t_0 + h)) \in D$ et on obtient, en posant, pour $1 \leq i \leq n$, $k_i = hu'_i(t_0) + h\varepsilon_i(h)$,

$$\begin{aligned} g(t_0 + h) &= f(u_1(t_0 + h), \dots, u_n(t_0 + h)) \\ &= f(a_1 + hu'_1(t_0) + h\varepsilon_1(h), \dots, a_n + hu'_n(t_0) + h\varepsilon_n(h)) \\ &= f(A) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(A)k_i + \|K\|\varepsilon(K). \end{aligned}$$

Cela peut s'écrire encore

$$g(t_0 + h) = g(t_0) + h \sum_{i=1}^n u'_i(t_0) \frac{\partial f}{\partial x_i}(A) + R(h),$$

où

$$R(h) = h \sum_{i=1}^n \varepsilon_i(h) \frac{\partial f}{\partial x_i}(A) + \|(hu'_1(t_0) + h\varepsilon_1(h), \dots, hu'_n(t_0) + h\varepsilon_n(h))\| \varepsilon(K).$$

Pour $h \neq 0$, on a

$$\frac{R(h)}{h} = \sum_{i=1}^n \varepsilon_i(h) \frac{\partial f}{\partial x_i}(A) + \frac{|h|}{h} \|(u'_1(t_0) + \varepsilon_1(h), \dots, u'_n(t_0) + \varepsilon_n(h))\| \varepsilon(K).$$

Quand h tend vers 0, chaque $\varepsilon_i(h)$ tend vers 0, K tend vers $(0, \dots, 0)$, donc $\varepsilon(K)$ tend vers 0 et finalement, $\lim_{h \rightarrow 0} \frac{R(h)}{h} = 0$. On a donc $R(h) = o(h)$ et

$$g(t_0 + h) = g(t_0) + h \sum_{i=1}^n u'_i(t_0) \frac{\partial f}{\partial x_i}(A) + o(h).$$

La fonction g possède un développement limité d'ordre 1 en t_0 . Elle est donc dérivable en t_0 et le nombre dérivée est le coefficient de h

$$g'(t_0) = \sum_{i=1}^n u'_i(t_0) \frac{\partial f}{\partial x_i}(A). \quad \square$$

Exemple

Si $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ est de classe \mathcal{C}^1 sur \mathbb{R}^3 , la fonction $t \mapsto f(t^2, e^t, t)$ est de classe \mathcal{C}^1 sur \mathbb{R} , de dérivée

$$t \mapsto 2t \frac{\partial f}{\partial x}(t^2, e^t, t) + e^t \frac{\partial f}{\partial y}(t^2, e^t, t) + \frac{\partial f}{\partial z}(t^2, e^t, t).$$

Existence des dérivées directionnelles

Théorème 3

Soit f une fonction de classe \mathcal{C}^1 sur un ouvert D de \mathbb{R}^n , $A \in D$ et U un vecteur unitaire de \mathbb{R}^n .

La fonction f possède une dérivée dans la direction U qui est égale à

$$\langle \nabla f_A, U \rangle.$$

Preuve

On pose $A = (a_1, \dots, a_n)$, $U = (\alpha_1, \dots, \alpha_n)$. On considère la fonction $\varphi : t \mapsto f(A + tU)$ qui est définie au voisinage de 0. On peut écrire

$$\varphi(t) = f(A + tU) = f(a_1 + t\alpha_1, \dots, a_n + t\alpha_n) = f(u_1(t), \dots, u_n(t)),$$

où les fonctions u_i sont définies par $u_i(t) = a_i + t\alpha_i$. Les fonctions u_i étant dérivables sur \mathbb{R} , on en déduit que φ est dérivable en 0 et

$$\varphi'(0) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(A) u_i'(0) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(A) \alpha_i = \langle \nabla f_A, U \rangle. \quad \square$$

► **Remarque**

On a donc $|f'_i(A)| \leq \|\nabla f_A\|$, d'après l'inégalité de Cauchy-Schwarz, avec égalité si U est colinéaire à ∇f_A . Le vecteur ∇f_A donne la direction dans laquelle la variation de f au voisinage de A est maximale.

Formule des accroissements finis

Théorème 4

Soient Ω un ouvert de \mathbb{R}^n , $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^1 , A et B deux éléments de \mathbb{R}^n tels que $[A, B] \subset \Omega$. Alors il existe $C \in]A, B[$ tel que

$$f(B) - f(A) = \langle B - A, \nabla f_C \rangle.$$

Preuve

On note $A = (a_1, \dots, a_n)$ et $B = (b_1, \dots, b_n)$. Pour tout $t \in [0, 1]$, $(1 - t)A + tB \in [A, B] \subset \Omega$. On peut donc définir $\varphi : [0, 1] \rightarrow \mathbb{R}$ par $\varphi(t) = f((1 - t)A + tB)$. On a pour tout $t \in [0, 1]$, $\varphi(t) = f(u_1(t), \dots, u_n(t))$, où pour tout $i \in [1, n]$, $u_i(t) = (1 - t)a_i + tb_i$. Les fonctions u_i et f sont de classe \mathcal{C}^1 donc φ est de classe \mathcal{C}^1 et, pour tout $t \in [0, 1]$,

$$\begin{aligned} \varphi'(t) &= \sum_{i=1}^n u_i'(t) \frac{\partial f}{\partial x_i}((1 - t)A + tB) = \sum_{i=1}^n (b_i - a_i) \frac{\partial f}{\partial x_i}((1 - t)A + tB) \\ &= \langle B - A, \nabla f_{(1-t)A+tB} \rangle. \end{aligned}$$

La fonction φ étant de classe \mathcal{C}^1 sur $[0, 1]$, on peut appliquer la formule des accroissements finis à φ sur $[0, 1]$. Il existe $t_0 \in]0, 1[$ tel que $\varphi(1) - \varphi(0) = \varphi'(t_0)$. En posant $C = (1 - t_0)A + t_0B$, on obtient le résultat voulu, puisque $\varphi(0) = f(A)$ et $\varphi(1) = f(B)$. □

► **Remarque**

Si Ω est un ouvert convexe, on peut considérer des points A et B quelconques de Ω .

Exemples

1. Considérons une fonction f affine sur \mathbb{R}^n . Il existe $U \in \mathbb{R}^n$ et $c \in \mathbb{R}$ tels que, pour tout $X \in \mathbb{R}^n$,

$$f(X) = \langle X, U \rangle + c.$$

On a $\nabla f_X = U$ pour tout $X \in \mathbb{R}^n$. Pour A et B dans \mathbb{R}^n , on obtient

$$f(B) - f(A) = \langle B - A, U \rangle = \langle B - A, \nabla f_C \rangle$$

pour C quelconque.

2. Considérons la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par $f(X) = \|X\|^2$. On a, pour tout $X \in \mathbb{R}^n$, $\nabla f_X = 2X$. Pour A et B dans \mathbb{R}^n ,

$$f(B) - f(A) = \|B\|^2 - \|A\|^2 = \langle B - A, B + A \rangle = \langle B - A, \nabla f_C \rangle,$$

où $C = \frac{1}{2}(A + B)$ est le milieu du segment $[A, B]$.

2. Calcul différentiel d'ordre 2

2.1 Dérivées partielles d'ordre 2

Définition 9

Soient $f : D \rightarrow \mathbb{R}$, où D est un ouvert de \mathbb{R}^n , $A \in D$, $1 \leq i, j \leq n$. On suppose que $\frac{\partial f}{\partial x_i}$ est définie sur D . Si $\frac{\partial f}{\partial x_i}$ possède une dérivée partielle par rapport à x_j en A , celle-ci est notée $\frac{\partial^2 f}{\partial x_j \partial x_i}(A)$.

Si, en tout point $M \in D$, $\frac{\partial f}{\partial x_i}$ possède une dérivée partielle par rapport à x_j , on note $\frac{\partial^2 f}{\partial x_j \partial x_i}$ la fonction $M \mapsto \frac{\partial^2 f}{\partial x_j \partial x_i}(M)$.

Les fonctions $\frac{\partial^2 f}{\partial x_j \partial x_i}$, si elles existent, sont appelées dérivées partielles d'ordre 2.

2.2 Fonctions de classe \mathcal{C}^2

Définition - Opérations sur les fonctions de classe \mathcal{C}^2

Définition 10

Soient D un ouvert de \mathbb{R}^n , $f : D \rightarrow \mathbb{R}$. On dit que f est de classe \mathcal{C}^2 sur D si elle admet en tout point de D des dérivées partielles d'ordre 2 et si les fonctions $\frac{\partial^2 f}{\partial x_i \partial x_j}$ ($1 \leq i, j \leq n$) sont continues sur D .

► Remarque

Il revient au même de dire que, pour tout $i \in [1, n]$, $\frac{\partial f}{\partial x_i}$ est de classe \mathcal{C}^1 .

Proposition 5

Soient D un ouvert de \mathbb{R}^n , $f, g : D \rightarrow \mathbb{R}$.

Si f et g sont de classe \mathcal{C}^2 sur D , alors $f + g$, λf ($\lambda \in \mathbb{R}$) et $f g$ sont de classe \mathcal{C}^2 sur D .

Si de plus g ne s'annule pas sur D , $\frac{1}{g}$ et $\frac{f}{g}$ sont de classe \mathcal{C}^2 .

Preuve

Par définition, pour $1 \leq i \leq n$, les fonctions $\frac{\partial f}{\partial x_i}$ et $\frac{\partial g}{\partial x_i}$ sont définies sur D et de classe \mathcal{C}^1 . D'après la proposition 1, $f + g$, λf et fg ont des dérivées partielles d'ordre 1 sur D et pour $1 \leq i \leq n$, on a

$$\frac{\partial(f+g)}{\partial x_i} = \frac{\partial f}{\partial x_i} + \frac{\partial g}{\partial x_i}, \quad \frac{\partial(\lambda f)}{\partial x_i} = \lambda \frac{\partial f}{\partial x_i}, \quad \frac{\partial(fg)}{\partial x_i} = f \frac{\partial g}{\partial x_i} + g \frac{\partial f}{\partial x_i}.$$

Il résulte alors de la proposition 3 que ces dérivées partielles sont de classe \mathcal{C}^1 et donc, par définition, que $f + g$, λf et fg sont de classe \mathcal{C}^2 .

Si g ne s'annule pas, $\frac{1}{g}$ et $\frac{f}{g}$ possèdent des dérivées partielles d'ordre 1, d'après la proposition 1 et, pour $1 \leq i \leq n$,

$$\frac{\partial}{\partial x_i} \left(\frac{1}{g} \right) = -\frac{1}{g^2} \frac{\partial g}{\partial x_i} \quad \text{et} \quad \frac{\partial}{\partial x_i} \left(\frac{f}{g} \right) = \frac{1}{g^2} \left(\frac{\partial f}{\partial x_i} g - f \frac{\partial g}{\partial x_i} \right).$$

Il résulte encore de la proposition 3 que ces dérivées partielles sont de classe \mathcal{C}^1 et donc que $\frac{1}{g}$ et $\frac{f}{g}$ sont de classe \mathcal{C}^2 . □

Exemples

1. Les fonctions polynomiales sont de classe \mathcal{C}^2 sur \mathbb{R}^n et les quotients de fonctions polynomiales sont de classe \mathcal{C}^2 sur leur ensemble de définition.
2. Si $f : I \rightarrow \mathbb{R}$, où I est un intervalle de \mathbb{R} , est de classe \mathcal{C}^2 , les fonctions $g : (x, y) \mapsto f(x)$ et $h : (x, y) \mapsto f(y)$ sont de classe \mathcal{C}^2 sur $I \times \mathbb{R}$ et $\mathbb{R} \times I$ respectivement.

Proposition 6

Soient $f : D \rightarrow \mathbb{R}$ une fonction définie sur un ouvert D de \mathbb{R}^n , $\varphi : I \rightarrow \mathbb{R}$, où I est un intervalle de \mathbb{R} . On suppose que $f(D) \subset I$.

Si f est de classe \mathcal{C}^2 sur D et si φ est de classe \mathcal{C}^2 sur I , alors la fonction $\varphi \circ f$ est de classe \mathcal{C}^2 sur D .

Preuve

D'après la proposition 4, la fonction $\varphi \circ f$ est de classe \mathcal{C}^1 et, pour tout $i \in [1, n]$,

$$\frac{\partial}{\partial x_i} (\varphi \circ f) = (\varphi' \circ f) \frac{\partial f}{\partial x_i}.$$

Toujours d'après la proposition 4, la fonction $\varphi' \circ f$ est de classe \mathcal{C}^1 . Il en est de même de $\frac{\partial f}{\partial x_i}$ par définition d'une fonction de classe \mathcal{C}^2 . Leur produit $\frac{\partial}{\partial x_i} (\varphi \circ f)$ est de classe \mathcal{C}^1 . Cela est vrai pour tout $i \in [1, n]$ donc $\varphi \circ f$ est de classe \mathcal{C}^2 . □

Exemple

Soit $\alpha \in \mathbb{R}$. Si f est de classe \mathcal{C}^2 sur D , ouvert de \mathbb{R}^n , et strictement positive sur D , alors la fonction f^α est de classe \mathcal{C}^2 sur D , comme on le voit en écrivant $f^\alpha = \varphi \circ f$, où $\varphi : \mathbb{R}_+^* \rightarrow \mathbb{R}$ est définie par $\varphi(x) = x^\alpha$.

On obtient, pour $1 \leq i \leq n$,

$$\frac{\partial f^\alpha}{\partial x_i} = \alpha f^{\alpha-1} \frac{\partial f}{\partial x_i},$$

puis pour $1 \leq i, j \leq n$,

$$\frac{\partial^2 f^\alpha}{\partial x_j \partial x_i} = \alpha(\alpha - 1)f^{\alpha-2} \frac{\partial f}{\partial x_j} \frac{\partial f}{\partial x_i} + \alpha f^{\alpha-1} \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Théorème de Schwarz

Théorème 5

Soit D un ouvert de \mathbb{R}^2 , $f : D \rightarrow \mathbb{R}$ une application possédant des dérivées partielles $\frac{\partial^2 f}{\partial x \partial y}$ et $\frac{\partial^2 f}{\partial y \partial x}$ continues en A . On a alors

$$\frac{\partial^2 f}{\partial x \partial y}(A) = \frac{\partial^2 f}{\partial y \partial x}(A).$$

Preuve

On pose $A = (a, b)$. Comme D est ouvert, il existe $r > 0$ tel que $B(A, r) \subset D$. Si $0 < h, k < \frac{r}{\sqrt{2}}$, alors $[a, a+h] \times [b, b+k]$ est inclus dans D (cf les remarques de la page 93). On pose

$$\Delta(h, k) = f(a+h, b+k) - f(a+h, b) - f(a, b+k) + f(a, b).$$

La fonction $\varphi : x \mapsto f(x, b+k) - f(x, b)$ est dérivable sur $[a, a+h]$ et sa dérivée est

$$\varphi' : x \mapsto \frac{\partial f}{\partial x}(x, b+k) - \frac{\partial f}{\partial x}(x, b).$$

On peut lui appliquer la formule des accroissements finis : il existe $c \in [a, a+h]$ tel que

$$\Delta(h, k) = \varphi(a+h) - \varphi(a) = h\varphi'(c) = h \left(\frac{\partial f}{\partial x}(c, b+k) - \frac{\partial f}{\partial x}(c, b) \right).$$

La fonction $y \mapsto \frac{\partial f}{\partial x}(c, y)$ est dérivable sur $[b, b+k]$ de dérivée $y \mapsto \frac{\partial^2 f}{\partial y \partial x}(c, y)$. On peut donc lui appliquer le théorème des accroissements finis : il existe $d \in [b, b+k]$ tel que

$$\Delta(h, k) = hk \frac{\partial^2 f}{\partial y \partial x}(c, d).$$

En considérant la fonction $\Psi : y \mapsto f(a+h, y) - f(a, y)$, on montre de même qu'il existe $c' \in [a, a+h]$ et $d' \in [b, b+k]$ tels que

$$\Delta(h, k) = hk \frac{\partial^2 f}{\partial x \partial y}(c', d').$$

On en déduit que

$$\frac{\partial^2 f}{\partial y \partial x}(c, d) = \frac{\partial^2 f}{\partial x \partial y}(c', d'). \quad (*)$$

Quand h et k tendent vers 0, (c, d) et (c', d') tendent vers $(a, b) = A$. Les fonctions $\frac{\partial^2 f}{\partial y \partial x}$ et $\frac{\partial^2 f}{\partial x \partial y}$ étant continues en A , on obtient par passage à la limite dans $(*)$, l'égalité

$$\frac{\partial^2 f}{\partial x \partial y}(A) = \frac{\partial^2 f}{\partial y \partial x}(A). \quad \square$$

Corollaire 3

Soit D un ouvert de \mathbb{R}^n , $A \in D$, i et j des entiers entre 1 et n , $f : D \rightarrow \mathbb{R}$ une application possédant des dérivées partielles $\frac{\partial^2 f}{\partial x_i \partial x_j}$ et $\frac{\partial^2 f}{\partial x_j \partial x_i}$ continues en A . On a alors

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(A) = \frac{\partial^2 f}{\partial x_j \partial x_i}(A).$$

Preuve

Soient $A = (a_1, \dots, a_n)$,

$$\Delta = \{(x, y) \in \mathbb{R}^2 \mid (a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_{j-1}, y, a_{j+1}, \dots, a_n) \in D\}.$$

L'ensemble Δ est un ouvert de \mathbb{R}^2 . On considère $\varphi : \Delta \rightarrow \mathbb{R}$ définie par

$$\varphi(x, y) = f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_{j-1}, y, a_{j+1}, \dots, a_n).$$

Les dérivées partielles $\frac{\partial^2 \varphi}{\partial x \partial y}$ et $\frac{\partial^2 \varphi}{\partial y \partial x}$ sont égales respectivement à

$$(x, y) \mapsto \frac{\partial^2 f}{\partial x_i \partial x_j}(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_{j-1}, y, a_{j+1}, \dots, a_n)$$

et

$$(x, y) \mapsto \frac{\partial^2 f}{\partial x_j \partial x_i}(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_{j-1}, y, a_{j+1}, \dots, a_n).$$

Elles sont continues en (a_i, a_j) , puisque $\frac{\partial^2 f}{\partial x_i \partial x_j}$ et $\frac{\partial^2 f}{\partial x_j \partial x_i}$ sont continues en A . On déduit du théorème 5 que

$$\frac{\partial^2 \varphi}{\partial x \partial y}(a_i, a_j) = \frac{\partial^2 \varphi}{\partial y \partial x}(a_i, a_j), \text{ c'est-à-dire que } \frac{\partial^2 f}{\partial x_i \partial x_j}(A) = \frac{\partial^2 f}{\partial x_j \partial x_i}(A). \quad \square$$

Corollaire 4

Soit D un ouvert de \mathbb{R}^n , $f : D \rightarrow \mathbb{R}$ une application de classe \mathcal{C}^2 . Alors, pour tous i et j de $\llbracket 1, n \rrbracket$, on a

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Preuve

Cela résulte du corollaire 3 appliqué en tout point $A \in D$. □

Exemples

1. Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = 2x^2y + y^2 - x^3y^4.$$

On obtient

$$\frac{\partial f}{\partial x}(x, y) = 4xy - 3x^2y^4 \text{ et } \frac{\partial f}{\partial y}(x, y) = 2x^2 + 2y - 4x^3y^3$$

puis

$$\frac{\partial^2 f}{\partial y \partial x}(x, y) = \frac{\partial^2 f}{\partial x \partial y}(x, y) = 4x - 12x^2y^3.$$

2. Donnons un contre-exemple. Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ la fonction définie par

$$f(x, y) = \begin{cases} \frac{x^3 y}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{sinon.} \end{cases}$$

La fonction f est de classe \mathcal{C}^2 sur $\mathbb{R}^2 \setminus \{(0, 0)\}$ car quotient de polynômes, et on trouve, pour $(x, y) \neq (0, 0)$,

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= \frac{y(3x^2(x^2 + y^2) - x^3(2x))}{(x^2 + y^2)^2} = \frac{x^2 y(x^2 + 3y^2)}{(x^2 + y^2)^2} \\ \frac{\partial f}{\partial y}(x, y) &= \frac{x^3(x^2 + y^2 - 2y^2)}{(x^2 + y^2)^2} = \frac{x^3(x^2 - y^2)}{(x^2 + y^2)^2}. \end{aligned}$$

On a d'autre part $\frac{\partial f}{\partial x}(0, 0) = \frac{\partial f}{\partial y}(0, 0) = 0$, car les applications partielles en 0 sont nulles.

On en déduit que

$$\begin{aligned} \frac{\partial^2 f}{\partial y \partial x}(0, 0) &= \lim_{y \rightarrow 0} \frac{1}{y} \left(\frac{\partial f}{\partial x}(0, y) - \frac{\partial f}{\partial x}(0, 0) \right) = \lim_{y \rightarrow 0} 0 = 0 \\ \frac{\partial^2 f}{\partial x \partial y}(0, 0) &= \lim_{x \rightarrow 0} \frac{1}{x} \left(\frac{\partial f}{\partial y}(x, 0) - \frac{\partial f}{\partial y}(0, 0) \right) = \lim_{x \rightarrow 0} \frac{x - 0}{x} = 1. \end{aligned}$$

On a donc

$$\frac{\partial^2 f}{\partial x \partial y}(0, 0) \neq \frac{\partial^2 f}{\partial y \partial x}(0, 0).$$

On en déduit que $\frac{\partial^2 f}{\partial x \partial y}$ et $\frac{\partial^2 f}{\partial y \partial x}$ ne sont pas toutes deux continues en $(0, 0)$. La fonction f n'est pas de classe \mathcal{C}^2 sur \mathbb{R}^2 . Par contre, elle est de classe \mathcal{C}^1 , car on montre aisément que les dérivées partielles d'ordre 1 sont continues en $(0, 0)$.

Hessienne

Définition 11

Soient D un ouvert de \mathbb{R}^n et $f : D \rightarrow \mathbb{R}$ une application de classe \mathcal{C}^2 . On appelle hessienne de f en un point $A \in D$ la matrice

$$H_A = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(A) \right)_{1 \leq i, j \leq n} \in \mathcal{M}_n(\mathbb{R}).$$

► Remarque

Il résulte du corollaire 3 que la hessienne de f en A est une matrice symétrique. On note q_A la forme quadratique de \mathbb{R}^n dont la matrice est la hessienne de f en A . On a donc, pour tout $X = (x_1, \dots, x_n) \in \mathbb{R}^n$,

$$q_A(X) = \sum_{1 \leq i, j \leq n} \frac{\partial^2}{\partial x_i \partial x_j}(A) x_i x_j.$$

Dans le cas $n = 2$, on note traditionnellement (notations de Monge)

$$r = \frac{\partial^2 f}{\partial x^2}, \quad s = \frac{\partial^2 f}{\partial x \partial y}(A) = \frac{\partial^2 f}{\partial y \partial x}(A) \quad \text{et} \quad t = \frac{\partial^2 f}{\partial y^2}(A).$$

Ainsi, on obtient

$$q_A(x, y) = rx^2 + 2sxy + ty^2.$$

Exemple

Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est définie par $f(X) = \|X\|^2$, on a pour $X = (x_1, \dots, x_n)$ et $i \in \llbracket 1, n \rrbracket$,

$$\frac{\partial f}{\partial x_i}(X) = 2x_i.$$

On en déduit que, pour tout $(i, j) \in \llbracket 1, n \rrbracket^2$,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(X) = 2 \quad \text{si} \quad i = j \quad \text{et} \quad 0 \quad \text{sinon.}$$

La matrice hessienne de f en tout point de \mathbb{R}^n est $2I_n$.

2.3 Étude locale des fonctions de classe \mathcal{C}^2

Égalité de Taylor-Lagrange à l'ordre 1

Théorème 6

Soient D un ouvert de \mathbb{R}^n , $f : D \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 . Si A et H sont deux éléments de \mathbb{R}^n tels que $[A, A + H] \subset D$, il existe $\theta \in]0, 1[$ tel que

$$f(A + H) = f(A) + \langle \nabla f_A, H \rangle + \frac{1}{2} q_{A+\theta H}(H).$$

Preuve

Pour tout $t \in [0, 1]$, $A + tH \in [A, A + H] \subset D$. On peut donc définir $\varphi : [0, 1] \rightarrow \mathbb{R}$ par $\varphi(t) = f(A + tH)$. Si $A = (a_1, \dots, a_n)$ et $H = (h_1, \dots, h_n)$, on obtient

$$\varphi(t) = f(a_1 + th_1, \dots, a_n + th_n).$$

La fonction f étant de classe \mathcal{C}^1 , ainsi que les fonctions $t \mapsto a_i + th_i$ ($1 \leq i \leq n$), il résulte du théorème 2 que φ est de classe \mathcal{C}^1 et que, pour $t \in [0, 1]$,

$$\varphi'(t) = \sum_{i=1}^n h_i \frac{\partial f}{\partial x_i}(A + tH) = \langle \nabla f_{A+tH}, H \rangle.$$

Puisque f est de classe \mathcal{C}^2 , les fonctions $\frac{\partial f}{\partial x_i}$ sont de classe \mathcal{C}^1 . En appliquant de nouveau le théorème 2, on obtient que φ' est dérivable et que, pour tout $t \in [0, 1]$,

$$\varphi''(t) = \sum_{i=1}^n h_i \sum_{j=1}^n h_j \frac{\partial^2 f}{\partial x_j \partial x_i}(A + tH) = \sum_{1 \leq i, j \leq n} h_i h_j \frac{\partial^2 f}{\partial x_j \partial x_i}(A + tH) = q_{A+tH}(H).$$

Comme f est de classe \mathcal{C}^2 , les dérivées partielles d'ordre 2 sont continues et φ'' est continue. La fonction φ est donc de classe \mathcal{C}^2 sur $]0, 1[$; on peut lui appliquer la formule de Taylor-Lagrange à l'ordre 1 entre 0 et 1. Il existe $\theta \in]0, 1[$ tel que

$$\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(\theta).$$

D'après ce qui précède, on a

$$\varphi(1) = f(A + H), \quad \varphi(0) = f(A), \quad \varphi'(0) = \langle \nabla f_A, H \rangle \quad \text{et} \quad \varphi''(\theta) = q_{A+\theta H}(H).$$

On obtient donc l'égalité voulue. □

► **Remarques**

- Ce théorème généralise la formule de Taylor-Lagrange à l'ordre 1 pour les fonctions d'une variable qui peut s'énoncer : si f est de classe \mathcal{C}^2 sur un intervalle I et si a et $a + h$ appartiennent à I , il existe $\theta \in]0, 1[$ tel que

$$f(a + h) = f(a) + hf'(a) + \frac{h^2}{2}f''(a + \theta h)$$

(cf pages 559-560 du livre de première année).

- Si D est convexe, on peut prendre A et $A + H$ quelconques dans D .

Exemple

La fonction $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ définie par $f(x, y) = xy + e^y$ est de classe \mathcal{C}^2 sur \mathbb{R}^2 . Pour tout $(x, y) \in \mathbb{R}^2$, on a

$$\frac{\partial f}{\partial x}(x, y) = y \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) = x + e^y,$$

puis

$$\frac{\partial^2 f}{\partial x^2}(x, y) = 0, \quad \frac{\partial^2 f}{\partial x \partial y}(x, y) = \frac{\partial^2 f}{\partial y \partial x}(x, y) = 1, \quad \frac{\partial^2 f}{\partial y^2}(x, y) = e^y.$$

Prenons $A = (0, 0)$ et $H = (h, k)$. On a alors $f(A) = 1$, $\nabla f_A = (0, 1)$ et la hessienne au point $A + \theta H = (\theta h, \theta k)$ est $\begin{pmatrix} 0 & 1 \\ 1 & e^{\theta k} \end{pmatrix}$. Il existe donc $\theta \in]0, 1[$ tel que

$$f(h, k) = 1 + k + 2hk + k^2 e^{\theta k}.$$

Développement limité d'ordre 2

Théorème 7

Soit D un ouvert de \mathbb{R}^n , $f : D \longrightarrow \mathbb{R}$ une application de classe \mathcal{C}^2 , $A \in D$. On note q_A la forme quadratique de \mathbb{R}^n dont la matrice est la hessienne de f en A .

Il existe ε telle que, pour tout $H \in \mathbb{R}^n$ tel que $A + H \in D$,

$$f(A + H) = f(A) + \langle \nabla f_A, H \rangle + \frac{1}{2}q_A(H) + \|H\|^2 \varepsilon(H)$$

et $\lim_{H \rightarrow (0, \dots, 0)} \varepsilon(H) = 0$.

Preuve

Soit ε la fonction définie sur $\Omega = \{H \in \mathbb{R}^n, A + H \in D\}$ par

$$\varepsilon(H) = \begin{cases} 0 & \text{si } H = (0, \dots, 0) \\ \frac{1}{\|H\|^2} \left(f(A + H) - f(A) - \langle \nabla f_A, H \rangle - \frac{1}{2} q_A(H) \right) & \text{si } H \neq (0, \dots, 0). \end{cases}$$

Il faut démontrer que $\lim_{H \rightarrow (0, \dots, 0)} \varepsilon(H) = 0$. Puisque D est ouvert, il existe $r > 0$ tel que $B(A, r) \subset D$. Soit $\|H\| \leq r$.

Pour tout $t \in [0, 1]$, $\|tH\| \leq r$ donc $A + tH \in D$ et $[A, A + H] \subset D$. D'après le théorème 6, il existe $\theta \in]0, 1[$ tel que

$$f(A + H) = f(A) + \langle \nabla f_A, H \rangle + \frac{1}{2} q_{A+\theta H}(H).$$

Si $H \neq (0, \dots, 0)$, on en déduit

$$\varepsilon(H) = \frac{1}{2\|H\|^2} (q_{A+\theta H}(H) - q_A(H)).$$

En posant $A = (a_1, \dots, a_n)$ et $H = (h_1, \dots, h_n)$, on obtient

$$\varepsilon(H) = \frac{1}{2\|H\|^2} \sum_{1 \leq i, j \leq n} h_i h_j \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(A + \theta H) - \frac{\partial^2 f}{\partial x_i \partial x_j}(A) \right).$$

Comme, pour tout $i \in \llbracket 1, n \rrbracket$, $|h_i| \leq \|H\|$, on en déduit que

$$|\varepsilon(H)| \leq \frac{1}{2} \sum_{1 \leq i, j \leq n} \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(A + \theta H) - \frac{\partial^2 f}{\partial x_i \partial x_j}(A) \right|.$$

Quand H tend vers 0, $A + \theta H$ tend vers A et, les dérivées partielles secondes de f étant continues sur D , chaque terme de la somme tend vers 0. On a donc $\lim_{H \rightarrow (0, \dots, 0)} \varepsilon(H) = 0$. □

► **Remarque**

Ce théorème doit être rapproché de la formule de Taylor-Young à l'ordre 2 pour les fonctions d'une variable.

Exemple

Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $f(x, y) = \ln(e^x + e^y)$. La fonction f est de classe \mathcal{C}^2 . Pour $(x, y) \in \mathbb{R}^2$, on obtient

$$\frac{\partial f}{\partial x}(x, y) = \frac{e^x}{e^x + e^y} \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) = \frac{e^y}{e^x + e^y},$$

$$\frac{\partial^2 f}{\partial x^2}(x, y) = \frac{e^x e^y}{(e^x + e^y)^2}, \quad \frac{\partial^2 f}{\partial x \partial y}(x, y) = \frac{\partial^2 f}{\partial y \partial x}(x, y) = \frac{-e^x e^y}{(e^x + e^y)^2}, \quad \frac{\partial^2 f}{\partial y^2}(x, y) = \frac{e^x e^y}{(e^x + e^y)^2}.$$

On prend $A = (0, 0)$. On obtient $f(0, 0) = \ln(2)$, $\nabla f_{(0,0)} = \left(\frac{1}{2}, \frac{1}{2}\right)$. La hessienne en $(0, 0)$

est $\begin{pmatrix} \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{pmatrix}$. On obtient, pour $(h, k) \in \mathbb{R}^2$,

$$f(h, k) = \ln 2 + \frac{1}{2}(h + k) + \frac{1}{8}(h^2 - 2hk + k^2) + (h^2 + k^2)\varepsilon(h, k),$$

avec $\lim_{(h,k) \rightarrow (0,0)} \varepsilon(h, k) = 0$.

Dérivée seconde directionnelle en A

Définition 12

Soient D un ouvert de \mathbb{R}^n , $f : D \rightarrow \mathbb{R}$, A un point de D , U un vecteur unitaire. Si la fonction φ définie par $\varphi(t) = f(A + tU)$ est deux fois dérivable en 0, le nombre dérivé seconde en 0 est appelé dérivée seconde de f en A dans la direction U et est noté $f''_U(A)$.

Proposition 7

Soient D un ouvert de \mathbb{R}^n , $f : D \rightarrow \mathbb{R}$ une application de classe \mathcal{C}^2 , $A \in D$, U un vecteur unitaire. La fonction f possède une dérivée seconde en A dans la direction U , égale à $q_A(U)$.

Preuve

On pose $A = (a_1, \dots, a_n)$, $U = (\alpha_1, \dots, \alpha_n)$. On considère la fonction $\varphi : t \mapsto f(A + tU)$ qui est définie au voisinage de 0. On peut écrire

$$\varphi(t) = f(A + tU) = f(a_1 + t\alpha_1, \dots, a_n + t\alpha_n) = f(u_1(t), \dots, u_n(t)),$$

où les fonctions u_i sont définies par $u_i(t) = a_i + t\alpha_i$. Les fonctions u_i étant dérivables sur \mathbb{R} , on en déduit que φ est dérivable au voisinage de 0 et

$$\varphi'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(A + tU)u'_i(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(A + tU)\alpha_i.$$

Les fonctions $\frac{\partial f}{\partial x_i}$ étant de classe \mathcal{C}^1 , on trouve de même que la fonction φ' est dérivable et

$$\varphi''(t) = \sum_{i=1}^n \alpha_i \left(\sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(A + tU)\alpha_j \right).$$

On obtient en particulier

$$\varphi''(0) = \sum_{1 \leq i, j \leq n} \frac{\partial^2 f}{\partial x_j \partial x_i}(A)\alpha_i\alpha_j = q_A(U). \quad \square$$

Pour le cas $n = 2$, on pourra se reporter au livre de première année (chapitre 28, exercices page 570 à 572).

1. Déterminer les dérivées partielles d'ordre 1 et 2 des fonctions $f : D \longrightarrow \mathbb{R}$, où D est un ouvert de \mathbb{R}^2 à déterminer

1. $f(x, y) = x^2\sqrt{y}$;
2. $f(x, y) = \ln(x + \sqrt{x^2 + y^2})$;
3. $f(x, y) = (x + y)^y$.

2. Déterminer les dérivées partielles d'ordre 1 et 2 des fonctions $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$ définies par

1. $f(x, y, z) = e^{x^2+y^2+z^2}$;
2. $f(x, y, z) = xy + yz + zx$;
3. $f(x, y, z) = \frac{xyz}{1 + x^2 + y^2}$.

3. Pour chacune des fonctions suivantes, étudier la continuité, l'existence de dérivées partielles d'ordre 1, le caractère \mathcal{C}^1 .

1. $f(x, y) = \frac{\sin x^2 + \sin y^2}{\sqrt{x^2 + y^2}}$ si $(x, y) \neq (0, 0)$ et $f(0, 0) = 0$;
2. $g(x, y) = \frac{|xy|^\alpha}{x^2 + y^2}$ si $(x, y) \neq (0, 0)$ et $g(0, 0) = 0$ ($\alpha \in \mathbb{R}$);

4. Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ de classe \mathcal{C}^1 telle que

$$f(0, 1) = 0, \quad \frac{\partial f}{\partial x}(0, 1) = 1 \quad \text{et} \quad \frac{\partial f}{\partial y}(0, 1) = 2.$$

1. Écrire le développement limité de f d'ordre 1 au voisinage de $(0, 1)$.
2. En déduire le développement limité d'ordre 1 au voisinage de 0 des fonctions $t \longmapsto f(-2t, e^t)$ et $t \longmapsto f\left(t, \frac{e^t + e^{-t}}{2}\right)$. Calculer

$$\lim_{t \rightarrow 0} \frac{f(-2t, e^t)}{f\left(t, \frac{e^t + e^{-t}}{2}\right)}.$$

5. Montrer que la fonction $X \longmapsto \frac{X}{\|X\|^2}$ est de classe \mathcal{C}^1 sur l'ouvert $\mathbb{R}^n \setminus \{(0, \dots, 0)\}$. Déterminer son développement limité d'ordre 1 au voisinage de X_0 .

6. Soit f une fonction continue sur \mathbb{R} . Pour $(x, y) \in \mathbb{R}^2$, $x \neq 0$, on pose

$$g(x, y) = \frac{1}{x} \int_x^{xy} f(t) dt.$$

1. Montrer que g peut être prolongée en une fonction continue sur \mathbb{R}^2 .
2. Étudier l'existence de dérivées partielles d'ordre 1 pour g .
3. Pour quelles fonctions f la fonction g est-elle indépendante de y ?

7. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application de classe \mathcal{C}^1 et $\alpha \in \mathbb{R}$. On dit que f est positivement homogène de degré α si

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n \quad \forall t > 0 \quad f(tx_1, \dots, tx_n) = t^\alpha f(x_1, \dots, x_n).$$

1. Montrer que, si f est positivement homogène de degré α , ses dérivées partielles sont positivement homogènes de degré $\alpha - 1$.
2. Montrer que si f est positivement homogène de degré α , on a

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n \quad \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}(x_1, \dots, x_n) = \alpha f(x_1, \dots, x_n). \quad (*)$$

3. On suppose réciproquement que f vérifie (*). Montrer que, pour tout $(x_1, \dots, x_n) \in \mathbb{R}^n$, l'application $\varphi : t \mapsto f(tx_1, \dots, tx_n)$ vérifie l'équation différentielle

$$\forall t > 0 \quad \varphi'(t) = \frac{\alpha}{t} \varphi(t).$$

En déduire que f est positivement homogène de degré α .

8. (D'après ESCP 2000) Soit a un réel strictement positif. On note $I =]-a, a[$. On considère une application $f : I^2 \rightarrow I$ de classe \mathcal{C}^1 sur I^2 . On suppose qu'il existe un réel $k \in [0, 1[$ tel que, pour tout $(x, y) \in I^2$,

$$\left| \frac{\partial f}{\partial x}(x, y) \right| + \left| \frac{\partial f}{\partial y}(x, y) \right| \leq k.$$

1. Soient (x, y) et (x', y') deux couples de I^2 . En utilisant la formule des accroissements finis, montrer que

$$|f(x, y) - f(x', y')| \leq k \max(|x - x'|, |y - y'|).$$

2. Soient $(\alpha, \beta) \in I^2$ et $(u_n)_{n \in \mathbb{N}}$ la suite définie par

$$u_0 = \alpha, \quad u_1 = \beta, \quad \text{et} \quad \forall n \in \mathbb{N}, \quad u_{n+2} = f(u_{n+1}, u_n).$$

Pour $n \in \mathbb{N}$, on pose $a_n = \max(|u_{n+2} - u_{n+1}|, |u_{n+1} - u_n|)$.

- a) Montrer que, pour tout $n \in \mathbb{N}$, $a_{n+1} \leq ka_n$.
 - b) Montrer que la série $\sum a_n$ est convergente.
 - c) Montrer que la suite (u_n) est convergente.
3. Montrer que la limite de la suite (u_n) est indépendante du couple (α, β) .

9. Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = \begin{cases} \frac{xy(x^2 - y^2)}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{sinon.} \end{cases}$$

1. Montrer que f est de classe \mathcal{C}^1 sur \mathbb{R}^2 .
2. Comparer $\frac{\partial^2 f}{\partial x \partial y}(0, 0)$ et $\frac{\partial^2 f}{\partial y \partial x}(0, 0)$. Que peut-on en déduire ?

10. Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une application de classe \mathcal{C}^2 . On pose, pour $(x, y, h) \in \mathbb{R}^3$,

$$\Delta(x, y, h) = f(x + h, y + h) + f(x - h, y - h) - f(x + h, y - h) - f(x - h, y + h).$$

1. On suppose que, pour tout $(x, y, h) \in \mathbb{R}^3$, on a $\Delta(x, y, h) = 0$. Montrer que

$$\frac{\partial^2 f}{\partial x \partial y} = 0.$$

Pour cela, on pourra, (x, y) étant fixé, écrire le développement limité d'ordre 2 en 0 de $h \mapsto \Delta(x, y, h)$.

En déduire qu'il existe des fonctions $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 telles que

$$\forall (x, y) \in \mathbb{R}^2 \quad f(x, y) = \varphi(x) + \psi(y).$$

2. Étudier la réciproque.

11. Déterminer les fonctions $f : D \rightarrow \mathbb{R}$, où D est un ouvert de \mathbb{R}^2 à déterminer, vérifiant :

1.
$$\begin{cases} \frac{\partial f}{\partial x}(x, y) = \frac{2+x}{y} \\ \frac{\partial f}{\partial y}(x, y) = \frac{2+y}{x} ; \end{cases}$$
2.
$$\begin{cases} \frac{\partial f}{\partial x}(x, y) = \frac{1-y}{(x+y+1)^2} \\ \frac{\partial f}{\partial y}(x, y) = \frac{2+x}{(x+y+1)^2} ; \end{cases}$$
3.
$$\begin{cases} \frac{\partial f}{\partial x}(x, y) = \frac{y^2}{(x+y)^2} \\ \frac{\partial f}{\partial y}(x, y) = \frac{x^2}{(x+y)^2} ; \end{cases}$$
4.
$$\begin{cases} \frac{\partial f}{\partial x}(x, y) = 2x + \frac{1}{y} \\ \frac{\partial f}{\partial y}(x, y) = 2y - \frac{x}{y^2}. \end{cases}$$

12. On considère l'équation différentielle (E) :

$$x^2 \frac{\partial f}{\partial x^2} + 2xy \frac{\partial f}{\partial x \partial y} + y^2 \frac{\partial f}{\partial y^2} = 0,$$

où l'inconnue f est une fonction de classe \mathcal{C}^2 sur $\mathbb{R}_+^* \times \mathbb{R}$.

1. Si f est une fonction de classe \mathcal{C}^2 sur $\mathbb{R}_+^* \times \mathbb{R}$, on définit $g : \mathbb{R}_+^* \times \mathbb{R} \rightarrow \mathbb{R}$ par

$$g(u, v) = f(u, uv).$$

Justifier l'existence de $\frac{\partial^2 g}{\partial u^2}$. La calculer.

2. En déduire que f est solution de (E) si et seulement si $\frac{\partial^2 g}{\partial u^2} = 0$.

Montrer qu'il en est ainsi si et seulement s'il existe deux fonctions k et ℓ de classe \mathcal{C}^2 sur \mathbb{R} telles que

$$\forall (u, v) \in \mathbb{R}_+^* \times \mathbb{R} \quad g(u, v) = uk(v) + \ell(v).$$

3. En déduire les solutions de (E).

13. Soit $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^1 , où Ω est un ouvert convexe de \mathbb{R}^n .

Pour $(X, Y) \in \Omega^2$, on définit $\varphi_{X,Y} : [0, 1] \rightarrow \mathbb{R}$ par

$$\varphi_{X,Y}(t) = f(tX + (1-t)Y).$$

On rappelle le résultat de l'exercice 13 du chapitre 4 : la fonction f est convexe sur Ω si et seulement si, pour tout $(X, Y) \in \Omega^2$, la fonction $\varphi_{X,Y}$ est convexe sur $[0, 1]$.

1. Montrer que, pour tout $(X, Y) \in \Omega^2$, $\varphi_{X,Y}$ est de classe \mathcal{C}^1 . Calculer $\varphi'_{X,Y}$.

2. On suppose que f est convexe. En utilisant la convexité de $\varphi_{X,Y}$, montrer que, pour tout $(X, Y) \in \Omega^2$,

$$f(X) \geq f(Y) + \langle \nabla f_Y, X - Y \rangle.$$

3. On suppose réciproquement que, pour tout $(X, Y) \in \Omega^2$,

$$f(X) \geq f(Y) + \langle \nabla f_Y, X - Y \rangle.$$

Montrer que, pour u et t dans $[0, 1]$, on a

$$\varphi_{X,Y}(u) \geq \varphi_{X,Y}(t) + (u-t)\varphi'_{X,Y}(t).$$

Montrer que $\varphi_{X,Y}$ est convexe. Conclure.

4. On suppose que f est de classe \mathcal{C}^2 . Montrer que, pour tout $(X, Y) \in \Omega^2$, la fonction $\varphi_{X,Y}$ est de classe \mathcal{C}^2 . Calculer φ'' .

En déduire que f est convexe si, et seulement si, pour tout $A \in \Omega$ et $H \in \mathbb{R}^n$, on a $q_A(H) \geq 0$.

Extremums

7

1. Extremums sur un ouvert

1.1 Définitions

Définition 1

Soit Ω une partie de \mathbb{R}^n , $f : \Omega \rightarrow \mathbb{R}$, A un élément de Ω .

On dit que f admet un maximum absolu en A si

$$\forall X \in \Omega, f(X) \leq f(A).$$

On dit que f admet un minimum absolu en A si

$$\forall X \in \Omega, f(X) \geq f(A).$$

On dit que f admet en A un extremum absolu si elle admet en A un maximum ou un minimum absolu.

Définition 2

Soit Ω une partie de \mathbb{R}^n , $f : \Omega \rightarrow \mathbb{R}$, A un élément de Ω .

On dit que f admet en A un maximum local (ou relatif) s'il existe $r > 0$ tel que

$$\forall X \in B(A, r) \cap \Omega, f(X) \leq f(A).$$

On dit que f admet en A un minimum local (ou relatif) s'il existe $r > 0$ tel que

$$\forall X \in B(A, r) \cap \Omega, f(X) \geq f(A).$$

On dit que f admet en A un extremum local si elle admet en A un maximum ou un minimum local.

► **Remarques**

- Si f admet en A un extremum global, elle admet en A un extremum local.
- Si les inégalités précédentes sont strictes pour $X \neq A$, on dit qu'on a un extremum strict.
- Dans la suite, on supposera que Ω est un ouvert.

1.2 Condition nécessaire du premier ordre**Théorème 1**

Soient Ω un ouvert de \mathbb{R}^n , $f : \Omega \rightarrow \mathbb{R}$, possédant sur Ω des dérivées partielles d'ordre 1, et A un point de Ω . Si f admet un extremum local en A , alors $\nabla f_A = 0$.

Preuve

Supposons que f admette en A un maximum local. Il existe $r > 0$ tel que

$$\forall X \in B(A, r) \cap \Omega \quad f(X) \leq f(A).$$

Comme Ω est ouvert, il existe $r' > 0$ tel que $B(A, r') \subset \Omega$. En posant $\rho = \min(r, r')$, on a donc

$$\forall X \in B(A, \rho) \quad f(X) \leq f(A).$$

Posons $A = (a_1, \dots, a_n)$ et considérons, pour $i \in [1, n]$, la i -ème fonction partielle

$$f_i : x \mapsto f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n).$$

Si on considère, pour $x \in]a_i - \rho, a_i + \rho[$, $X = (a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)$, on a $d(X, A) = |x - a_i| < \rho$, donc $X \in B(A, \rho)$. Ainsi f_i est définie sur $]a_i - \rho, a_i + \rho[$ et on a, pour tout $x \in]a_i - \rho, a_i + \rho[$, $f(x) \leq f(A)$, c'est-à-dire

$$f_i(x) \leq f_i(a_i).$$

La fonction f_i définie sur $]a_i - \rho, a_i + \rho[$ possède en a_i un extremum. On en déduit que $f'_i(a_i) = 0$ (cf théorème 14 du chapitre 20 du livre de première année). Mais, par définition, $f'_i(a_i) = \frac{\partial f}{\partial x_i}(A)$. On a donc,

$$\forall i \in [1, n] \quad \frac{\partial f}{\partial x_i}(A) = 0, \quad \text{i.e.} \quad \nabla f_A = 0. \quad \square$$

► **Remarques**

- Ce théorème est l'équivalent dans \mathbb{R}^n du théorème mentionné dans la démonstration pour les fonctions définies sur une partie de \mathbb{R} .
- L'hypothèse Ω ouvert est essentielle. Si Ω est la boule fermée de centre 0 et de rayon 1, la fonction $f : X \mapsto \|X\|^2$ possède un maximum local en tout point X de norme 1. Mais en tel point le gradient de f n'est pas nul car, pour tout $X \in \Omega$, $\nabla f_X = 2X$.

Définition 3

Soient Ω un ouvert de \mathbb{R}^n , $f : \Omega \rightarrow \mathbb{R}$, possédant sur Ω des dérivées partielles d'ordre 1, et A un point de Ω . Si $\nabla f_A = 0$, on dit que A est un point critique de f .

Proposition 1

Soient Ω un ouvert de \mathbb{R}^n , $f : \Omega \rightarrow \mathbb{R}$ une application de classe \mathcal{C}^1 . Si A est un point critique de f , toutes les dérivées directionnelles en A sont nulles

Preuve

Comme f est de classe \mathcal{C}^1 , il résulte du théorème 3 du chapitre 6 que pour tout vecteur U de \mathbb{R}^n , f possède une dérivée dans la direction U égale à $\langle \nabla f_A, U \rangle = 0$. □

► **Remarques**

- Pour déterminer les extremums sur un ouvert d'une fonction possédant des dérivées partielles, on commence par déterminer ses points critiques.
- L'annulation du gradient n'est qu'une condition nécessaire d'extremum, comme le montre l'exemple de la fonction

$$f : (x, y) \mapsto x^2 - y^2.$$

Son gradient est nul en $(0, 0)$, mais $f(0, 0) = 0$ et pour tout réel x non nul, on a $f(x, 0) > 0$ et $f(0, x) < 0$, ce qui montre que dans toute boule de centre $(0, 0)$, f prend des valeurs plus petites et des valeurs plus grandes que 0, donc f n'admet pas d'extremum en $(0, 0)$.

Exemples

1. La fonction définie sur \mathbb{R}^2 par

$$f(x, y) = x^2 + xy + y^2 + x - y + 3$$

a pour dérivées partielles

$$\frac{\partial f}{\partial x}(x, y) = 2x + y + 1 \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) = x + 2y - 1.$$

Le seul point critique est $(-1, 1)$. En faisant un changement d'origine, on étudie, pour tout $(h, k) \in \mathbb{R}^2$,

$$f(-1 + h, 1 + k) = h^2 + hk + k^2 + 2.$$

Comme $h + hk + k^2 = \left(h + \frac{k}{2}\right)^2 + \frac{3k^2}{4} \geq 0$, on a $f(-1 + h, 1 + k) \geq 2$, i.e. $f(-1 + h, 1 + k) \geq f(-1, 1)$, ce qui montre que f admet un minimum global en $(-1, 1)$.

2. La fonction définie sur \mathbb{R}^2 par

$$f(x, y) = x^3 - y^3 + 3x^2 - 3y^2$$

a pour dérivées partielles

$$\frac{\partial f}{\partial x}(x, y) = 3x^2 + 6x \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) = -3y^2 - 6y.$$

Ses points critiques sont $(0, 0)$, $(-2, 0)$, $(0, -2)$ et $(-2, -2)$.

- On a $f(0, 0) = 0$ et au voisinage de 0,

$$f(h, 0) = h^3 + 3h^2 \sim 3h^2 \quad \text{et} \quad f(0, k) = -k^3 - 3k^2 \sim -3k^2.$$

Pour h et k assez petits et non nuls, on obtient $f(h, 0) > 0$ et $f(0, k) < 0$: la fonction n'admet pas d'extremum en $(0, 0)$.

- On obtient de même $f(-2, -2) = 0$ et, pour h et k assez petit non nuls,

$$f(-2 + h, -2) = -3h^2 + h^3 < 0 \quad \text{et} \quad f(-2, -2 + k) = 3k^2 - k^3 > 0 :$$

La fonction f n'admet pas d'extremum en $(-2, -2)$.

- En $A = (-2, 0)$, on étudie

$$f(-2 + h, k) = 4 - 3h^2 - 3k^2 + h^3 - k^3 = 4 - h^2(3 - h) - k^2(3 + k).$$

Si $\|(h, k)\| < 3$, on a $3 - h > 0$ et $3 + k > 0$ et donc $f(-2 + h, k) \leq 4$. Autrement dit, pour $(x, y) \in B(A, 3)$, on a $f(x, y) \leq f(-2, 0)$: la fonction f présente un maximum local en $(-2, 0)$.

- En $B = (0, -2)$, on étudie

$$f(h, -2 + k) = -4 + 3h^2 + 3k^2 + h^3 - k^3 = -4 + h^2(3 + h) + k^2(3 - k).$$

On obtient de même que, pour tout $(x, y) \in B(B, 3)$, on a $f(x, y) \geq f(0, -2)$: la fonction f présente un minimum local en $(0, -2)$.

Si f est définie sur un ensemble fermé borné, on sait que f possède un maximum et un minimum sur Ω (théorème 5 du chapitre 5). Mais, comme Ω n'est pas ouvert, on ne peut pas dire que ces extremums sont atteints en des points critiques de f . On peut alors considérer l'ensemble des points X de Ω pour lesquels il existe une boule ouverte de centre X incluse dans Ω . C'est un ouvert appelé intérieur de Ω (cf exercice 9 du chapitre 4), sur lequel ce qui précède s'applique.

Un extremum de f est donc obtenu

- soit en un point critique de l'intérieur de Ω ;
- soit en un point de Ω qui n'appartient pas à l'intérieur de Ω (on dit qu'il appartient à la frontière de Ω).

Dans le cas où Ω est la boule fermé $B_f(A, r)$, l'intérieur de Ω est la boule ouverte $B(A, r)$.

Exemple

Soit f la fonction définie sur $D = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$ par

$$f(x, y) = x^2 - xy + y^2.$$

Comme elle est continue, f possède un maximum et un minimum sur le fermé borné D (c'est le disque fermé de centre $(0, 0)$ et de rayon 1).

- Sur le disque ouvert, f possède des dérivées partielles

$$\frac{\partial f}{\partial x}(x, y) = 2x - y \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) = -x + 2y$$

et un seul point critique $(0, 0)$. On obtient $f(0, 0) = 0$.

- Sur le cercle, on peut poser $x = \cos \theta$, $y = \sin \theta$, $\theta \in [-\pi, \pi]$. On obtient

$$f(\cos \theta, \sin \theta) = 1 - \sin \theta \cos \theta = 1 - \frac{1}{2} \sin 2\theta.$$

Sur ce cercle, f possède un maximum $\frac{3}{2}$ et un minimum $\frac{1}{2}$.

Le minimum de f sur D vaut donc 0 est atteint en $(0, 0)$. Le maximum est $\frac{3}{2}$ et est atteint pour $\sin 2\theta = -1$, ce qui donne $\theta = -\frac{\pi}{4}$ ou $\theta = \frac{3\pi}{4}$; le maximum est donc atteint aux points $\left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$ et $\left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$.

1.3 Conditions du second ordre

Le théorème suivant donne une condition suffisante du second ordre.

Théorème 2

Soit Ω un ouvert de \mathbb{R}^n , $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^2 , A un point critique de f . Si la forme quadratique q_A associée à la hessienne en A prend des valeurs strictement négatives (respectivement strictement positives) pour tout vecteur non nul, alors f admet un maximum local (respectivement minimum local) en A .

Preuve

Soit $r > 0$ tel que $B(A, r) \subset \Omega$. La fonction f étant de classe \mathcal{C}^2 sur Ω , elle possède en A un développement limité d'ordre 2. Il existe $\varepsilon : \Omega \rightarrow \mathbb{R}$ telle que, pour tout $H \in \mathbb{R}^n$ tel que $\|H\| < r$,

$$f(A + H) = f(A) + \langle \nabla f_A, H \rangle + \frac{1}{2} q_A(H) + \|H\|^2 \varepsilon(H)$$

et $\lim_{H \rightarrow (0, \dots, 0)} \varepsilon(H) = 0$.

Soit H_A la matrice hessienne de f en A , $\mathcal{B} = (e_1, \dots, e_n)$ une base orthonormale de vecteurs propres de H_A , $(\lambda_1, \dots, \lambda_n)$ étant les valeurs propres associées. Pour tout vecteur H de \mathbb{R}^n , de coordonnées (h_1, \dots, h_n) dans la base \mathcal{B} , on obtient

$$q_A(H) = \sum_{i=1}^n \lambda_i h_i^2.$$

La forme quadratique q_A prend des valeurs strictement négatives (respectivement strictement positives) pour tout vecteur non nul si, et seulement si, toutes les valeurs propres de H sont strictement négatives (respectivement strictement positives). Supposons par exemple que les valeurs propres de H_A sont strictement négatives. On a avec les notations précédentes, pour tout $H \in \mathbb{R}^n$,

$$q_A(H) \leq \alpha \sum_{i=1}^n h_i^2 \leq \alpha \|H\|^2,$$

en notant α la plus grande valeur propre de H_A . On obtient, pour $\|H\| < r$,

$$f(A + H) = f(A) + \frac{1}{2} q_A(H) + \|H\|^2 \varepsilon(H) \leq f(A) + \left(\frac{1}{2} \alpha + \varepsilon(H)\right) \|H\|^2.$$

Comme $\lim_{H \rightarrow (0, \dots, 0)} \left(\frac{1}{2} \alpha + \varepsilon(H) \right) = \frac{1}{2} \alpha < 0$, on peut trouver un réel $\rho \in]0, r[$ tel que,

$$\forall H \in \mathbb{R}^n \quad \|H\| \leq \rho \implies \frac{1}{2} \alpha + \varepsilon(H) < 0.$$

On a alors, pour $\|H\| \leq \rho$, $f(A+H) \leq f(A)$ et même $f(A+H) < f(A)$ si $H \neq 0$. La fonction f admet en A un maximum local strict.

Si les valeurs propres de H_A sont strictement positives, on montre de même que f admet en A un minimum local strict en écrivant $q_A(H) \geq \beta \|H\|^2$, où β est la plus petite valeur propre de H_A . \square

Le théorème suivant donne une condition nécessaire du second ordre.

Théorème 3

Soient Ω un ouvert de \mathbb{R}^n , $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^2 , A un point critique de f .

Si f admet un maximum local (respectivement minimum local) en A alors q_A prend des valeurs négatives ou nulles (respectivement positives ou nulles).

Preuve

Supposons par exemple que f possède un maximum en A .

Soit $r > 0$ tel que $B(A, r) \subset \Omega$ et $f(X) \leq f(A)$ pour $X \in B(A, r)$. Si H est un vecteur non nul de \mathbb{R}^n et t un réel tel que $|t| < \frac{r}{\|H\|}$, alors $A + tH \in \Omega$ et $f(A + tH) \leq f(A)$. Mais on sait par ailleurs que

$$f(A + tH) = f(A) + \frac{1}{2} q_A(tH) + \|tH\|^2 \varepsilon(tH) = f(A) + \frac{1}{2} t^2 (q_A(H) + 2\|H\|^2 \varepsilon(tH)),$$

puisque A est un point critique. On en déduit que, pour $0 < |t| < \frac{r}{\|H\|}$, on a

$$q_A(H) + 2\|H\|^2 \varepsilon(tH) \leq 0.$$

En faisant tendre t vers 0, on obtient $q_A(H) \leq 0$. Cela est vrai pour tout $H \in \mathbb{R}^n$ (pour $H = 0$ c'est évident).

Dans le cas où f admet en A un minimum local, on montre de même que q_A est positive ou nulle. \square

► Remarque

Le théorème signifie que la forme quadratique q_A a des valeurs propres positives ou nulles si f admet un minimum en A et des valeurs propres négatives ou nulles si f admet en A un maximum.

Corollaire 1

Soient Ω un ouvert de \mathbb{R}^n , $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^2 , A un point critique de f .

S'il existe des vecteurs H_1 et H_2 tels que $q_A(H_1) > 0$ et $q_A(H_2) < 0$, la fonction f n'admet pas d'extremum en A .

Preuve

C'est la contraposée du théorème précédent. \square

► **Remarque**

Ce cas correspond à celui où q_A possède une valeur propre strictement positive et une valeur propre strictement négative.

Exemple

Soit $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ définie par $f(x, y, z) = xy + yz + zx - xyz$. La fonction f est de classe \mathcal{C}^2 . On vérifie que $A = (0, 0, 0)$ est point critique. Pour tout $X = (x, y, z)$ la hessienne au point X

est égale à $H_X = \begin{pmatrix} 0 & 1-z & 1-y \\ 1-z & 0 & 1-x \\ 1-y & 1-x & 0 \end{pmatrix}$ et en particulier $H_A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. On a donc, pour tout $(x, y, z) \in \mathbb{R}^3$,

$$q_A(x, y, z) = 2(xy + yz + zx).$$

on en déduit que $q_A(1, 1, 0) = 2$ et $q_A(1, -1, 0) = -2$. La fonction f n'admet pas d'extremum en A .

► **Remarque**

La condition énoncée dans le théorème 3 n'est pas suffisante. Dans le cas où q_A prend des valeurs positives ou nulles et s'annule pour d'autres valeurs que le vecteur nul, c'est-à-dire où les valeurs propres de la hessienne sont positives ou nulles sans être toutes strictement positives, on ne peut pas conclure.

Exemples

- Si f est définie sur \mathbb{R}^2 par $f(x, y) = x^4 - xy^3 + 3y^4$, on observe que les dérivées partielles d'ordre 1 et 2 en $(0, 0)$ sont nulles. Le point $(0, 0)$ est donc un point critique et $q_{(0,0)} = 0$. On ne peut pas conclure. Mais de l'inégalité $|uv| \leq \frac{1}{2}(u^2 + v^2)$ pour tout $(u, v) \in \mathbb{R}^2$, on déduit que, pour tout $(x, y) \in \mathbb{R}^2$,

$$|xy^3| \leq |xy|y^2 \leq \frac{1}{2}(x^2y^2 + y^4) \leq \frac{1}{2} \left(\frac{1}{2}(x^4 + y^4) + y^4 \right) \leq \frac{1}{4}(x^4 + 3y^4).$$

On en déduit que

$$f(x, y) = (x^4 + 3y^4) - xy^3 \geq \frac{3}{4}(x^4 + 3y^4) \geq 0 \geq f(0, 0).$$

La fonction f admet un minimum absolu en $(0, 0)$.

- Si f est définie sur \mathbb{R}^2 par $f(x, y) = x^2 + y^2 + 2xy + xy^3$, on observe que $(0, 0)$ est un point critique. La matrice hessienne au point $X = (x, y)$ est $H_X = \begin{pmatrix} 2 & 2 + 3y^2 \\ 2 + 3y^2 & 2 + 6xy \end{pmatrix}$ et en particulier $H_{(0,0)} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$. On en déduit que

$$q_{(0,0)}(x, y) = 2x^2 + 4xy + y^2 = 2(x + y)^2 \geq 0.$$

La forme quadratique $q_{(0,0)}$ ne prend que des valeurs positives ou nulles et s'annule pour les points $(x, -x)$. On ne peut pas conclure. Mais l'on observe que, pour tout $x \in \mathbb{R}^*$,

$$f(x, -x) = -x^4 < 0 \quad \text{et} \quad f(x, x) = 4x^2 + x^4 > 0.$$

Comme toute boule de centre $(0, 0)$ contient des points de la forme $(x, -x)$ et d'autres de la forme (x, x) , f n'admet pas d'extremum en $(0, 0)$.

1.4 Cas $n = 2$

Théorème 4

Soit Ω un ouvert de \mathbb{R}^2 , $f : \Omega \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 et A un point critique de f . On pose

$$r = \frac{\partial^2 f}{\partial x^2}(A), \quad s = \frac{\partial^2 f}{\partial x \partial y}(A), \quad t = \frac{\partial^2 f}{\partial y^2}(A) \quad (\text{notations de Monge}).$$

- Si $rt - s^2 > 0$, la fonction f admet un extremum local en A qui est un minimum si $r > 0$ et un maximum si $r < 0$.
- Si $rt - s^2 < 0$, la fonction f n'a pas un extremum en A .
- Si $rt - s^2 = 0$, on ne peut pas conclure.

Preuve

La hessienne de f au point A est $H_A = \begin{pmatrix} r & s \\ s & t \end{pmatrix}$. On a, pour tout réel λ ,

$$\lambda \in \text{Sp}(A) \iff H_A - \lambda I_2 = \begin{pmatrix} r - \lambda & s \\ s & t - \lambda \end{pmatrix} \text{ non inversible} \iff (r - \lambda)(t - \lambda) - s^2 = 0.$$

Les valeurs propres sont donc les solutions de l'équation

$$\lambda^2 - (r + t)\lambda + rt - s^2 = 0.$$

Le produit des valeurs propres est donc $rt - s^2$.

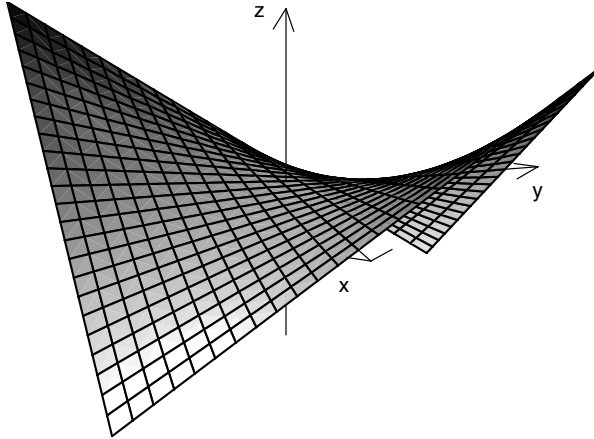
- Si $rt - s^2 > 0$, les deux valeurs propres ont même signe. Ce signe est celui de la somme des racines, soit $r + t$. On remarque que $rt > s^2$, donc $rt > 0$ et $r + t$ a le signe de r . Ainsi si $r > 0$, les deux valeurs propres sont strictement positives et $q_A(H) > 0$ pour tout vecteur H non nul : la fonction f admet un minimum local en A ; si $r < 0$, les deux valeurs propres sont strictement négatives et $q_A(H) < 0$ pour tout vecteur H non nul : la fonction f admet un maximum local en A .
- Si $rt - s^2 < 0$, les deux valeurs propres sont de signe contraire : la forme quadratique q_A prend des valeurs strictement positives et des valeurs propres strictement négatives ; la fonction f ne présente pas d'extremum local en A .
- Si $rt - s^2 = 0$, une des valeurs propres est nulle. La forme quadratique q_A s'annule pour d'autres valeurs que le vecteur nul : on ne peut pas conclure. \square

Définition 4

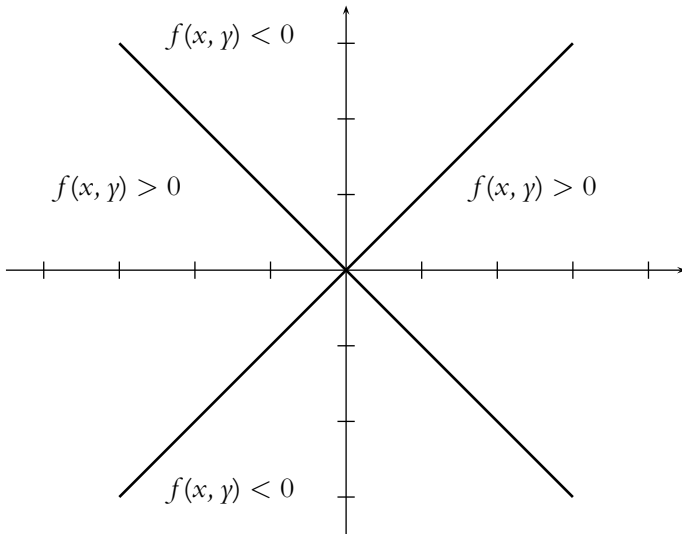
Dans le cas où les deux valeurs propres de la hessienne sont de signe contraire, on dit qu'on a un point selle ou que la fonction présente un col.

Exemple

Soit f la fonction définie sur \mathbb{R}^2 par $f(x, y) = x^2 - y^2$ dont le graphe est :



Le point $(0, 0)$ est un point critique, la hessienne en $(0, 0)$ est $\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$ dont les valeurs propres sont -2 et 2 . On a $f(0, 0) = 0$, $f(x, y) > 0$ si $|x| > |y|$ et $f(x, y) < 0$ si $|x| < |y|$. Le plan se divise en quatre quadrants sur lesquels la fonction garde un signe constant.



La situation décrite dans l'exemple précédent est générale. Soit f une fonction définie sur un ouvert Ω de \mathbb{R}^2 , A un point critique de f où la hessienne possède deux valeurs propres $\lambda_1 > 0$ et $\lambda_2 < 0$.

Soit $r > 0$ tel que $B(A, r) \subset \Omega$. Il existe une fonction ε ayant pour limite 0 en $(0, 0)$ telle que, si $H \in \mathbb{R}^2$ et $\|H\| < r$,

$$f(A + H) = f(A) + \frac{1}{2}q_A(H) + \|H\|^2\varepsilon(H).$$

Soit (U_1, U_2) une base orthonormale de vecteurs propres de la hessienne. Si H a pour coordonnées (h_1, h_2) dans cette base,

$$f(A + H) = f(A) + \frac{1}{2}(\lambda_1 h_1^2 + \lambda_2 h_2^2) + (h_1^2 + h_2^2)\varepsilon(H).$$

Posons $h_2 = th_1$. On a alors

$$f(A + H) = f(A) + h_1^2 \left(\frac{1}{2}(\lambda_1 + t^2\lambda_2) + (1 + t^2)\varepsilon(H) \right).$$

Si $\lambda_1 + t^2\lambda_2 \neq 0$, $f(A + H) - f(A)$ a, pour H assez petit, le signe de $\lambda_1 + t^2\lambda_2$.

Cette expression s'annule pour $|t| = \frac{\sqrt{\lambda_1}}{\sqrt{-\lambda_2}}$ (on rappelle que $\lambda_2 < 0$), soit

$|h_2| = \frac{\sqrt{\lambda_1}}{\sqrt{-\lambda_2}}|h_1|$. On obtient finalement :

- si $|h_2| < \frac{\sqrt{\lambda_1}}{\sqrt{-\lambda_2}}|h_1|$, on a $f(A + H) > f(A)$ pour $\|H\|$ assez petit ;
- si $|h_2| > \frac{\sqrt{\lambda_1}}{\sqrt{-\lambda_2}}|h_1|$, on a $f(A + H) < f(A)$ pour $\|H\|$ assez petit.

Le plan se divise en quatre quadrants limités par les droites d'équation $h_2 = \frac{\sqrt{\lambda_1}}{\sqrt{-\lambda_2}}h_1$

et $h_2 = -\frac{\sqrt{\lambda_1}}{\sqrt{-\lambda_2}}h_1$ sur lequel le signe de $f(A + H) - f(A)$ est constant pour H assez petit.

1.5 Extremums globaux

Si f est définie sur un ouvert Ω de \mathbb{R}^n , tout extremum global est un extremum local et donc un point critique. Pour montrer qu'un point critique A est un extremum global, on peut démontrer directement que le signe de $f(X) - f(A)$, où X décrit Ω est indépendant de X . Il est d'ailleurs souvent plus simple d'étudier le signe de $f(A + H) - f(A)$.

Exemple

Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $f(x, y) = x^2 + 2y^2 + xy - x + 3y + 3$. On vérifie que le seul point critique est $(1, -1)$. On a $f(1, -1) = 1$ et, pour $(h, k) \in \mathbb{R}^2$,

$$f(1 + h, -1 + k) = 1 + h^2 + hk + 2k^2 \geq 1 + h^2 - \frac{1}{2}(h^2 + k^2) + 2k^2 \geq 1 + \frac{1}{2}h^2 + \frac{3}{2}k^2.$$

Si $(h, k) \neq (0, 0)$, on a $f(1+h, -1k) > f(1, -1)$: la fonction f admet en $(-1, 1)$ un minimum global strict.

Une autre moyen de démontrer qu'on a un extremum global est de montrer que la forme quadratique q_B garde un signe constant, indépendant de $B \in \Omega$.

Théorème 5

Soit $f : \Omega \rightarrow \mathbb{R}$, où Ω est un ouvert convexe de \mathbb{R}^n . Si la forme quadratique q_B est positive (respectivement négative) en tout point de B de Ω , tout point critique de f est un minimum (respectivement un maximum) global.

Preuve

Traisons le cas où la forme q_B est toujours positive. Soit A un point critique de f . En appliquant l'égalité de Taylor-Lagrange à l'ordre 1 (théorème 6 du chapitre 6), on obtient que, pour tout $H \in \mathbb{R}^n$ tel que $A + H$ appartienne à Ω l'existence de $\theta \in]0, 1[$ tel que

$$f(A + H) = f(A) + \langle \nabla f_A, H \rangle + \frac{1}{2} q_{A+\theta H}(H) = f(A) + \frac{1}{2} q_{A+\theta H}(H),$$

car A est un point critique de f . Comme la forme quadratique $q_{A+\theta H}$ est positive, on a $f(A + H) \geq f(A)$ pour tout H : la fonction f admet en A un minimum global.

On obtient de même un maximum global si q_B est négative pour tout B . □

Exemple

Soit $f : (\mathbb{R}_+^*)^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = x^2 + y^2 + \frac{1}{x + y}.$$

On obtient, pour $(x, y) \in (\mathbb{R}_+^*)^2$,

$$\frac{\partial f}{\partial x}(x, y) = 2x - \frac{1}{(x + y)^2}, \quad \frac{\partial f}{\partial y}(x, y) = 2y - \frac{1}{(x + y)^2},$$

puis

$$r = \frac{\partial^2 f}{\partial x^2}(x, y) = 2 + \frac{2}{(x + y)^3}, \quad s = \frac{\partial^2 f}{\partial x \partial y}(x, y) = \frac{2}{(x + y)^3}, \quad t = \frac{\partial^2 f}{\partial y^2}(x, y) = 2 + \frac{2}{(x + y)^3}.$$

Un point critique vérifie $2x = 2y = \frac{1}{(x + y)^2} = \frac{1}{4x^2}$. Le seul point critique est $A = \left(\frac{1}{2}, \frac{1}{2}\right)$.

D'autre part, pour tout $(x, y) \in (\mathbb{R}_+^*)^2$, on a

$$rt - s^2 = 4 + \frac{4}{(x + y)^6} + \frac{8}{(x + y)^3} - \frac{4}{(x + y)^6} = 4 + \frac{8}{(x + y)^3} > 0.$$

De plus r est positif, donc la forme quadratique q_B est positive pour tout $B \in \Omega$. La fonction f possède donc un minimum absolu en A qui vaut $f(A) = \frac{3}{2}$.

Si on a démontré que la fonction f possède un extremum global sur l'ouvert Ω , en montrant par exemple qu'elle possède une borne supérieure (ou inférieure) qui est atteinte sur un fermé borné, on déterminera cet extremum en cherchant les points critiques de f .

Exemple

Soit $f : (\mathbb{R}_+^*)^2 \longrightarrow \mathbb{R}$ définie par

$$f(x, y) = x + y + \frac{1}{xy}.$$

La fonction est positive donc minorée. Posons $\alpha = \inf_{(x,y) \in (\mathbb{R}_+^*)^2} f(x, y)$. On note que $f(x, y)$ tend vers $+\infty$ quand $\|(x, y)\|$ tend vers $+\infty$ ou 0. On va en déduire que sa borne inférieure est atteinte.

Si $x > \alpha + 1$ ou $y > \alpha + 1$, alors $f(x, y) > \alpha + 1$.

Si $0 < x \leq \alpha + 1$ et $0 < y \leq \alpha + 1$, on a $f(x, y) > \frac{1}{xy} \geq \frac{1}{(\alpha + 1) \min(x, y)}$. On en déduit que $f(x, y) > \alpha + 1$ si $\min(x, y) < \frac{1}{(\alpha + 1)^2}$.

On a donc montré que si x et y ne sont pas dans le segment $\left[\frac{1}{(\alpha + 1)^2}, \alpha + 1 \right]$, on a $f(x, y) > \alpha + 1$. On en déduit que

$$\alpha = \inf_{(x,y) \in K} f(x, y) \text{ où } K = \left[\frac{1}{(\alpha + 1)^2}, \alpha + 1 \right]^2.$$

L'ensemble K est fermé car c'est un pavé fermé (cf démonstration page 98) et borné. La fonction f atteint donc sa borne inférieure sur K en un point (x_0, y_0) et l'on obtient

$$f(x_0, y_0) = \alpha = \inf_{(x,y) \in (\mathbb{R}_+^*)^2} f(x, y).$$

La fonction f admet donc un minimum global en (x_0, y_0) et comme $(\mathbb{R}_+^*)^2$ est ouvert, c'est un point critique de f . On vérifie que le seul point critique de f est $(1, 1)$. Ainsi f admet un minimum global en $(1, 1)$ qui vaut $f(1, 1) = 3$.

1.6 Position du graphe par rapport à l'hyperplan tangent

Soit f une fonction de classe \mathcal{C}^2 sur l'ouvert Ω de \mathbb{R}^n . Une équation du graphe de f est $x_{n+1} = f(x_1, \dots, x_n)$ et une équation de l'hyperplan tangent en $A = (a_1, \dots, a_n) \in \Omega$ est

$$x_{n+1} = f(A) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(A)(x_i - a_i).$$

Étudier la position du graphe par rapport à l'hyperplan tangent, c'est étudier le signe de

$$f(x_1, \dots, x_n) - f(A) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(A)(x_i - a_i),$$

c'est-à-dire en posant $(x_1, \dots, x_n) = X = A + H$ le signe de

$$f(A + H) - f(A) - \langle \nabla f_A, H \rangle.$$

Si cette quantité est positive, on dit que le graphe est au-dessus de l'hyperplan tangent ; si elle est négative, on dit que le graphe est au-dessous de l'hyperplan tangent.

En général, on fait une étude locale : on étudie ce signe pour H petit. On écrit le développement limité d'ordre 2 de f en A ,

$$f(A + H) = f(A) + \langle \nabla f_A, H \rangle + \frac{1}{2} q_A(H) + \|H\|^2 \varepsilon(H).$$

Il faut donc étudier le signe de $\frac{1}{2} q_A(H) + \|H\|^2 \varepsilon(H)$. C'est exactement l'étude qui a été faite dans la section 1.3 dans l'étude des extremums locaux. Des résultats obtenus, on déduit le théorème suivant.

Théorème 6

Soit f une fonction de classe \mathcal{C}^2 sur un ouvert Ω de \mathbb{R}^n , A un point de Ω .
 Si la forme quadratique q_A prend des valeurs strictement négatives (respectivement strictement positives) pour tout vecteur non nul, le graphe de f est au-dessous (respectivement au-dessus) de l'hyperplan tangent en A pour tout point proche de A .
 S'il existe des vecteurs H_1 et H_2 tels que $q_A(H_1) > 0$ et $q_A(H_2) < 0$, dans toute boule ouverte de centre A il existe des points où le graphe de f est au-dessous de l'hyperplan tangent en A et d'autres où le graphe de f est au-dessus de l'hyperplan tangent.
 Dans le cas $n = 2$, le premier cas correspond à $rt - s^2 > 0$ et le second à $rt - s^2 < 0$.

Exemple

Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $f(x, y) = x^3 - y^3 + 3xy + x + y$ et $A = (-1, 1)$. On a, pour $(x, y) \in \mathbb{R}^2$,

$$\frac{\partial f}{\partial x}(x, y) = 3x^2 + 3y + 1 \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) = -3y^2 + 3x + 1.$$

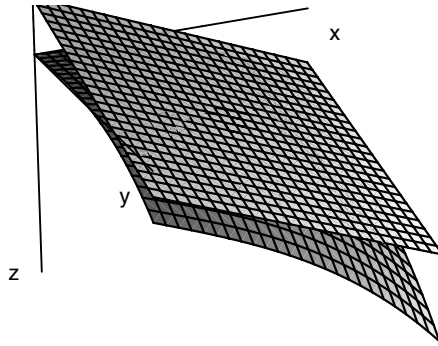
On en déduit $f(A) = -5$, $\frac{\partial f}{\partial x}(A) = 7$ et $\frac{\partial f}{\partial y}(A) = -5$. Une équation du plan tangent en A est donc

$$z = 7(x + 1) - 5(y - 1) - 5 = 7x - 5y + 7.$$

Comme

$$\frac{\partial^2 f}{\partial x^2}(x, y) = 6x \quad \frac{\partial^2 f}{\partial x \partial y}(x, y) = 3 \quad \frac{\partial^2 f}{\partial y^2}(x, y) = -6y,$$

les dérivées partielles secondes en A sont $r = -6$, $s = 3$ et $t = -6$. On a donc $rt - s^2 = 27$ et comme $r < 0$, le graphe est au-dessous du plan tangent pour des points proches de A .



► **Remarque**

Parfois, la différence $f(A + H) - f(A) - \langle \nabla f_A, H \rangle$ garde un signe constant, pour tout H tel que $A + H \in \Omega$.

Exemple

Soient a, b, c trois réels strictement positifs et $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ définie par

$$f(x, y, z) = ax^2 + by^2 + cz^2.$$

Pour tout $A = (x_0, y_0, z_0) \in \mathbb{R}^3$, on obtient $\nabla f_A = (2ax_0, 2by_0, 2cz_0)$. L'hyperplan tangent est l'ensemble des $(x, y, z, t) \in \mathbb{R}^4$ tel que

$$t = ax_0^2 + by_0^2 + cz_0^2 + 2ax_0(x - x_0) + 2by_0(y - y_0) + 2cz_0(z - z_0).$$

Pour étudier la position du graphe par rapport à l'hyperplan tangent, on étudie, pour $(h, k, \ell) \in \mathbb{R}^3$, le signe de

$$f(x_0 + h, y_0 + k, z_0 + \ell) - ax_0^2 - by_0^2 - cz_0^2 - 2ax_0h - 2by_0k - 2cz_0\ell = ah^2 + bk^2 + c\ell^2 \geq 0.$$

Le graphe est toujours au-dessus de son hyperplan tangent.

2. Extremums sous contrainte d'égalités linéaires

2.1 Définitions

Dans toute cette section, \mathcal{C} désigne l'ensemble des solutions d'un système linéaire

$$(S) \begin{cases} g_1(X) & = & b_1 \\ \vdots & \vdots & \vdots \\ g_p(X) & = & b_p, \end{cases}$$

où l'inconnue est un élément X de \mathbb{R}^n , et \mathcal{H} l'ensemble des solutions du système homogène associé. Nous rappelons le résultat vu en première année (page 326).

Proposition 2

Supposons que le système (S) admette des solutions, c'est-à-dire que \mathcal{C} n'est pas vide et considérons $X_0 \in \mathcal{C}$. Alors \mathcal{C} est l'ensemble des éléments de \mathbb{R}^n qui s'écrivent $X_0 + H$ avec H dans \mathcal{H} .

Définition 5

Soit $f : \Omega \rightarrow \mathbb{R}$, où Ω est un sous-ensemble de \mathbb{R}^n .

On dit que f admet un extremum local sous la contrainte \mathcal{C} en un point A si la restriction

de f à \mathcal{C} , i.e. $g : \begin{matrix} \Omega \cap \mathcal{C} & \longrightarrow & \mathbb{R} \\ X & \longmapsto & f(X) \end{matrix}$ admet un extremum local en A .

On dit que f admet un extremum global sous la contrainte \mathcal{C} en un point A si cette restriction g admet un extremum global en A .

2.2 Points critiques sous contrainte

Proposition 3

Si une fonction f de classe \mathcal{C}^1 sur un ouvert Ω de \mathbb{R}^n admet un extremum sous la contrainte \mathcal{C} en un point A , alors pour tout H de \mathcal{H} , la dérivée de f en A dans la direction H est nulle, c'est-à-dire $\langle \nabla f_A, H \rangle = 0$.

Preuve

Supposons que f possède un maximum sous la contrainte \mathcal{C} en A . Soit $H \in \mathcal{H}$ non nul.

Pour tout réel t , tH appartient à \mathcal{H} , donc d'après la proposition 2, $A + tH$ appartient à \mathcal{C} . De plus, puisque Ω est ouvert, il existe $r > 0$ tel que $B(A, r) \subset \Omega$. Si $\|tH\| < r$, c'est-à-dire si $|t| < \frac{r}{\|H\|}$, $A + tH$ appartient à $\Omega \cap \mathcal{C}$. On a donc

$$\forall t \in \left] -\frac{r}{\|H\|}, \frac{r}{\|H\|} \right[\quad f(A + tH) \leq f(A).$$

En considérant la fonction f_H définie par $f_H(t) = f(A + tH)$, on obtient

$$\forall t \in \left] -\frac{r}{\|H\|}, \frac{r}{\|H\|} \right[\quad f_H(t) \leq f_H(0).$$

La fonction f étant de classe \mathcal{C}^1 , on sait d'après le théorème 3 du chapitre 6 que f_H est dérivable en 0 et que son nombre dérivée, qui est la dérivée de f en A dans la direction H , vaut $\langle H, \nabla f_A \rangle$. Comme f_H possède un maximum en 0, on a $\langle H, \nabla f_A \rangle = 0$.

Si f possède un minimum local en A sous contrainte \mathcal{C} , on obtient le même résultat car f_H possède alors un minimum local en 0. □

Théorème 7

Si une fonction f de classe \mathcal{C}^1 sous un ouvert Ω de \mathbb{R}^n admet un extremum sous la contrainte \mathcal{C} en un point A , alors son gradient en A appartient à l'orthogonal de \mathcal{H} .

Preuve

D'après la proposition précédente, on a, pour tout $H \in \mathcal{H}$, $\langle H, \nabla f_A \rangle = 0$: le vecteur ∇f_A est orthogonal à \mathcal{H} . Ainsi ∇f_A appartient à \mathcal{H}^\perp . □

Définition 6

Soient f une fonction de classe \mathcal{C}^1 sur un ouvert Ω de \mathbb{R}^n , A un point de Ω . Si ∇f_A appartient à \mathcal{H}^\perp , on dit que A est un point critique de f dans l'optimisation sous contrainte \mathcal{C} .

Proposition 4

On a $\mathcal{H}^\perp = \text{Vect}(\nabla g_1, \dots, \nabla g_p)$.

Preuve

Soit $M = (m_{ij})_{\substack{1 \leq i \leq p \\ 1 \leq j \leq n}} \in \mathcal{M}_{p,n}(\mathbb{R})$ la matrice du système (S). Pour $i \in [1, p]$ et $X = (x_1, \dots, x_n) \in \mathbb{R}^n$, on a

$$g_i(X) = \sum_{j=1}^n m_{ij}x_j. \text{ On en déduit que}$$

$$\forall j \in [1, n] \quad \frac{\partial g_i}{\partial x_j}(X) = m_{ij}.$$

Ainsi $\nabla g_{iX} = (m_{i1}, \dots, m_{in})$ est indépendant de X . C'est le i -ième vecteur ligne de la matrice M . On le note ∇g_i (confondant ainsi la fonction $X \mapsto \nabla g_{iX}$ et la constante ∇g_{iX}). On a alors

$$\forall X \in \mathbb{R}^n \quad g_i(X) = \langle X, \nabla g_i \rangle.$$

On en déduit que

$$X \in \mathcal{H} \iff \forall i \in [1, p] \quad \langle X, \nabla g_i \rangle = 0 \iff \forall U \in \text{Vect}(\nabla g_1, \dots, \nabla g_p) \quad \langle X, U \rangle = 0,$$

cette dernière équivalence résultant de la linéarité du produit scalaire. Cela montre que $\mathcal{H} = (\text{Vect}(\nabla g_1, \dots, \nabla g_p))^\perp$ et donc $\mathcal{H}^\perp = \text{Vect}(\nabla g_1, \dots, \nabla g_p)$. \square

Corollaire 2

Soit f une fonction de classe \mathcal{C}^1 sur un ouvert Ω de \mathbb{R}^n , A un point de Ω . Alors A est un point critique de f dans l'optimisation sous contrainte \mathcal{C} si, et seulement si il existe $(\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$ tel que

$$\nabla f_A = \sum_{i=1}^p \lambda_i \nabla g_i.$$

► **Remarques**

- Si la famille $(\nabla g_1, \dots, \nabla g_p)$ est libre, c'est-à-dire si les formes linéaires (g_1, \dots, g_p) sont indépendantes, les scalaires $\lambda_1, \dots, \lambda_p$ sont uniques. Ils sont appelés multiplicateurs de Lagrange.
- On notera que la condition trouvée n'est que nécessaire, pas suffisante. Pour montrer que le « candidat extremum » est effectivement un extremum, il faut soit le montrer directement en étudiant le signe de $f(X) - f(A)$, pour $X \in \Omega \cap \mathcal{C}$, soit trouver un argument supplémentaire : se ramener par exemple à l'étude de f sur un fermé borné, ce qui montrera l'existence *a priori* d'un extremum.
- Si $A \in \mathcal{C}$, X appartient à \mathcal{C} s'il existe $H \in \mathcal{H}$ tel que $X = A + H$. Il faut donc étudier le signe de $f(A + H) - f(A)$ pour $H \in \mathcal{H}$. Si A est un point critique, $\nabla f_A \in \mathcal{H}^\perp$ et, si de plus f est de classe \mathcal{C}^2 , on obtient à partir de son développement limité d'ordre 2 en A ,

$$\forall H \in \mathcal{H}, \quad f(A + H) - f(A) = \frac{1}{2}q_A(H) + \|H\|^2 \varepsilon(H),$$

avec $\lim_{H \rightarrow 0} \varepsilon(H) = 0$. Si $q_A(H) < 0$ (resp. $q_A(H) > 0$) pour tout $H \in \mathcal{H}$, non nul, alors f possède un maximum (resp. un minimum) local en A .

Exemples

1. Soient $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ définie par $f(x, y, z) = \frac{1}{1 + 2x^2 + y^2 + z^2}$ et \mathcal{C} d'équation $g(x, y, z) = x - y + z = 5$. On a $\nabla g = (1, -1, 1)$. Si f possède un extremum relatif sous la contrainte \mathcal{C} en A , ∇f_A est colinéaire à $(1, -1, 1)$. Comme

$$\nabla f_{(x,y,z)} = \frac{-1}{(1 + 2x^2 + y^2 + z^2)^2} (4x, 2y, 2z),$$

il faut $4x = -2y = 2z$, soit $y = -2x$, $z = 2x$. Comme $A \in \mathcal{C}$, il faut de plus $x - y + z = 5$. On obtient $x = 1$, $y = -2$, $z = 2$ et $A = (1, -2, 2)$. Soit H un élément $(h, k, \ell) \in \mathcal{H}$, c'est-à-dire tel que $h - k + \ell = 0$, soit $\ell = -h + k$. On a alors

$$\begin{aligned} f(A + H) - f(A) &= \frac{1}{1 + 2(1 + h)^2 + (-2 + k)^2 + (2 - h + k)^2} - \frac{1}{1 + 2 + 4 + 4} \\ &= \frac{- (3h^2 + 2k^2 - 2hk)}{11(1 + 2(1 + h)^2 + (-2 + k)^2 + (2 - h + k)^2)}. \end{aligned}$$

Comme $-(3h^2 + 2k^2 - 2hk) = -(h - k)^2 - 2h^2 - k^2 \leq 0$, on a $f(A + H) - f(A) \leq 0$ pour tout $H \in \mathcal{H}$: la fonction f possède en A un minimum global sous la contrainte \mathcal{C} .

2. Soient $f : \mathbb{R}_+^n \rightarrow \mathbb{R}$ définie par $f(x_1, \dots, x_n) = \prod_{i=1}^n x_i$, $s \in \mathbb{R}_+^*$ et \mathcal{C} d'équation

$g(x_1, \dots, x_n) = \sum_{i=1}^n x_i = s$. Étudions les extremums de f sous la contrainte \mathcal{C} . On note que f n'est pas définie sur un ouvert.

Considérons $\Gamma = \{(x_1, \dots, x_n) \in \mathbb{R}_+^n \mid x_1 + \dots + x_n = s\}$. L'ensemble est borné car, si $(x_1, \dots, x_n) \in \Gamma$, alors $0 \leq x_i \leq s$ pour tout i et donc $\|(x_1, \dots, x_n)\| \leq s\sqrt{n}$. D'autre part, cet ensemble est fermé, car c'est l'intersection des demi-espaces fermés $x_i \geq 0$ et de

l'hyperplan $\sum_{i=1}^n x_i = s$. La fonction f possède donc un minimum et un maximum sur Γ .

Si le maximum de f est atteint en A , alors $A \in (\mathbb{R}_+^*)^n$ car le maximum de f sur Γ n'est manifestement pas nul. On en déduit que $A \in (\mathbb{R}_+^*)^n \cap \mathcal{C}$. Par définition, la restriction de f à $(\mathbb{R}_+^*)^n$ présente en A un maximum absolu sous la contrainte \mathcal{C} .

La fonction f étant de classe \mathcal{C}^1 sur $(\mathbb{R}_+^*)^n$, A est un point critique de f sous la contrainte \mathcal{C} .

On pose $A = (a_1, a_2, \dots, a_n)$. Comme $\frac{\partial f}{\partial x_i}(A) = \prod_{j \neq i} a_j = \frac{f(A)}{a_i}$ et $\nabla g = (1, \dots, 1)$, ∇f_A

est colinéaire à ∇g si $a_1 = \dots = a_n$ et donc $a_1 = \dots = a_n = \frac{s}{n}$. Le maximum de f sur Γ

est $f(A) = \left(\frac{s}{n}\right)^n$. On en déduit que, pour tout $(x_1, \dots, x_n) \in \Gamma$, $\prod_{i=1}^n x_i \leq \left(\frac{s}{n}\right)^n$ et donc

$\left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} \leq \frac{s}{n} \leq \frac{1}{n} \sum_{i=1}^n x_i$. On retrouve ainsi l'inégalité entre les moyennes arithmétique et géométrique.

La fonction n'ayant qu'un seul point critique sous la contrainte \mathcal{C} , le minimum de f sur Γ n'est pas atteint en un point de $(\mathbb{R}_+^*)^n$. En fait, il est manifeste que ce minimum vaut 0. Il est atteint sur le bord de Γ .

Donnons un exemple dans le cas où \mathcal{C} est défini par plus d'une équation.

Exemple

Déterminer les extremums de la fonction f définie sur \mathbb{R}^4 par

$$f(x, y, z, t) = x^2 + y^2 + z^2 + t^2$$

sous la contrainte $\begin{cases} g(x, y, z, t) = x + y + z - t = 3 \\ h(x, y, z, t) = 2x - y + z + t = -6 \end{cases}$. Si $A = (x, y, z, t)$, on a $\nabla f_A = 2(x, y, z, t)$. D'autre part, $\nabla g = (1, 1, 1, -1)$ et $\nabla h = (2, -1, 1, 1)$. Si A est un point critique sous la contrainte \mathcal{C} , il existe donc $(\lambda, \mu) \in \mathbb{R}^2$ tel que

$$(x, y, z, t) = \lambda(1, 1, 1, -1) + \mu(2, -1, 1, 1) = (\lambda + 2\mu, \lambda - \mu, \lambda + \mu, -\lambda + \mu).$$

En écrivant que $A \in \mathcal{C}$, on obtient

$$x + y + z - t = 4\lambda + \mu = 3 \quad \text{et} \quad 2x - y + z + t = \lambda + 7\mu = -6$$

soit $\lambda = 1$ et $\mu = -1$, puis $A = (-1, 2, 0, -2)$. On vérifie que, pour $H = (h, k, \ell, m) \in \mathbb{R}^4$, $q_M(H) = 2h^2 + 2k^2 + 2\ell^2 + 2m^2 > 0$ si $H \neq 0$. La fonction f possède en A un minimum global sous la contrainte \mathcal{C} . Elle ne possède pas de maximum sous la contrainte \mathcal{C} .

1. Soit f définie sur \mathbb{R}^2 par

$$f(x, y) = (x^2 - y)(2x^2 - y).$$

1. Montrer que, pour tout $\theta \in \mathbb{R}$ la fonction définie sur \mathbb{R} par

$$f_\theta(t) = f(t \cos \theta, t \sin \theta)$$

admet un minimum local en 0.

2. Montrer que f n'admet pas de minimum local en $(0, 0)$.

2. Déterminer les extremums sur \mathbb{R}^2 des fonctions

1. $f : (x, y) \mapsto x^2 + xy + y^2 + 2x - 2y;$

2. $f : (x, y) \mapsto x^3 y^2 (1 - x - y);$

3. $f : (x, y) \mapsto x^3 + 3xy^2 + 15x + 12y;$

4. $f : (x, y) \mapsto x^4 + y^4 - 2x^2 - 2y^2 + 4xy.$

3. Étudier les extremums de la fonction f définie sur \mathbb{R}^3 par

$$f(x, y, z) = (x + z^2)e^{x(y^2 + z^2 + 1)}.$$

4. 1. Soit f la fonction définie sur $\mathbb{R}_+^* \times \mathbb{R}$ par $f(x, y) = x((\ln x)^2 + y^2)$. Examiner la position du graphe de f par rapport au plan tangent au voisinage des points critiques.

2. Même question pour la fonction définie sur \mathbb{R}^2 par

$$f(x, y) = x^2 y^3 (1 + 3x + 2y).$$

3. Même question pour la fonction définie sur $\left] -\frac{\pi}{2}, \frac{\pi}{2} \right[$ par

$$f(x, y) = \sin x \sin y \sin(x + y).$$

5. Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = x^3 - 3x(1 + y^2).$$

1. Déterminer les extremums locaux de f .

2. Soit $D = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$.

a) Montrer que f possède un maximum M et un minimum m sur D et qu'ils sont atteints sur $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$.

b) Étudier la fonction, $t \mapsto f(\cos t, \sin t)$. Déterminer M et m .

6. (ESCP 2002) Soit f la fonction de \mathbb{R}^2 dans \mathbb{R} définie par

$$f(x, y) = \begin{cases} (x^2 + y^2)^x & \text{si } (x, y) \neq (0, 0) \\ 1 & \text{si } (x, y) = (0, 0). \end{cases}$$

1. a) Montrer que f est continue sur \mathbb{R}^2 .
b) Étudier l'existence des dérivées partielles de f .
2. On pose, pour $x \in \mathbb{R}$, $h(x) = f(x, 0) - 1$.
 - a) Étudier les variations de h .
 - b) En déduire que f n'admet pas d'extremum en $(0, 0)$.
3. Déterminer les points critiques de f .
4. a) Montrer que

$$\forall x \geq 0 \quad f(x, y) \geq h(x) + 1.$$
 - b) En déduire que f admet en $\left(\frac{1}{e}, 0\right)$ un minimum local.
5. On pose $g(x) = f(x, 1) - f(0, 1)$.
 - a) Montrer que $g(x)$ est du signe de x .
 - b) En déduire que f n'admet pas d'extremum local en $(0, 1)$.

7. Soit A une matrice symétrique d'ordre n . On considère la fonction définie sur $\mathbb{R}^n \setminus \{0\}$ par

$$\varphi(x) = \frac{{}^t X A X}{\|x\|^2},$$

où X est le vecteur-colonne des coordonnées de x dans la base canonique de \mathbb{R}^n . On note $S = \{x \in \mathbb{R}^n, \|x\| = 1\}$ la sphère unité de \mathbb{R}^n .

1. Montrer qu'il existe $x_0 \in S$ tel que

$$\varphi(x_0) = \max_{x \in S} \varphi(x).$$

2. Montrer que

$$\varphi(x_0) = \max_{x \in \mathbb{R}^n \setminus \{0\}} \varphi(x).$$

3. Montrer que φ est de classe \mathcal{C}^1 . En écrivant que x_0 est un point critique de φ , montrer que x_0 est vecteur propre de A .

8. (D'après ESSEC 2002) 1. Soit $U = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 < x_2 < x_3\}$. Montrer que l'ensemble U est ouvert et convexe.

On étudie les extremums de la fonction F définie sur U par

$$F(x) = \sum_{i=1}^3 x_i^2 - 2 \sum_{1 \leq i < j \leq 3} \ln(x_j - x_i).$$

2. À tout $a = (a_1, a_2, a_3) \in U$, on associe le polynôme

$$P = (X - a_1)(X - a_2)(X - a_3).$$

a) Montrer que, pour tout réel x distinct de a_1, a_2, a_3 ,

$$\frac{P'(x)}{P(x)} = \sum_{j=1}^3 \frac{1}{x - a_j}.$$

b) À l'aide de la formule de Taylor-Young, déterminer le développement limité d'ordre 2 au voisinage de a_i de $(x - a_i)P'(x) - P(x)$ et de $(x - a_i)P(x)$. En déduire la limite quand x tend vers a_i de $\frac{P'(x)}{P(x)} - \frac{1}{x - a_i}$. Montrer que

$$\sum_{\substack{j=1 \\ j \neq i}}^3 \frac{1}{a_i - a_j} = \frac{P''(a_i)}{2P'(a_i)}.$$

c) Exprimer les dérivées partielles de F en (x_1, x_2, x_3) en fonction de x_1, x_2, x_3 , puis démontrer que a est point critique si, et seulement si, $2XP' - P''$ admet pour racine a_1, a_2, a_3 .

d) Montrer que ceci est réalisé si et seulement si $2XP' - P'' = 6P$.

e) Montrer qu'il existe un unique polynôme unitaire de degré 3 tel que $2XP' - P'' = 6P$. En déduire que F admet un seul point critique a que l'on précisera.

3. Soient a le point critique de F , $x \in U$ et ψ la fonction définie sur $[0, 1]$ par

$$\psi(t) = F(tx + (1 - t)a).$$

a) Montrer que l'application Ψ est convexe sur $[0, 1]$ et que $\Psi'(0) = 0$.

b) En déduire que F admet un minimum absolu en a .

4. Calculer le minimum $F(a)$ de F .

9. Soit $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ définie par

$$f(x, y, z) = x^2 - 2xy + yz + y - z.$$

1. Déterminer les points critiques de f , les extremums, les points-selles.

2. Déterminer les points critiques et les extremums sous la contrainte

$$\begin{cases} 2x - y = 1 \\ x + z = 1. \end{cases}$$

10. 1. On note Δ l'ensemble $\{(x, y) \in \mathbb{R}^2 \mid x \geq 0, y \geq 0, -2x + y + 1 \geq 0, x - 2y + 1 \geq 0\}$ et g la fonction définie sur Δ par

$$g(x, y) = 3x - y + 4.$$

a) Représenter graphiquement Δ .

- b) Montrer que Δ est une partie fermée et bornée de \mathbb{R}^2 .
 c) Montrer que g admet un maximum sur Δ .
 d) Ce maximum peut-il être atteint en un point de l'ensemble Δ' défini par

$$\Delta' = \{(x, y) \in \mathbb{R}^2 \mid x > 0, y > 0, -2x + y + 1 > 0, x - 2y + 1 > 0\} ?$$

- e) Déterminer l'ensemble des points de Δ où ce maximum est atteint.
 2. Dans cette question, on identifiera \mathbb{R}^4 et $\mathcal{M}_{4,1}(\mathbb{R})$. On considère la matrice

$$A = \begin{pmatrix} 2 & -1 & 1 & 0 \\ -1 & 2 & 0 & 1 \end{pmatrix} \text{ et la matrice colonne } B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

On note \mathcal{C} l'ensemble

$$\left\{ X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4 \mid x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0 \text{ et } AX = B \right\}.$$

- a) Montrer que $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathcal{C}$ si, et seulement si, x_1, x_2, x_3 et x_4 satisfont

$$(x_1, x_2) \in \Delta, \quad x_3 = -2x_1 + x_2 + 1, \quad x_4 = x_1 - 2x_2 + 1.$$

- b) On considère $W = (2, 1, 4, 3) \in \mathbb{R}^4$ et la fonction f définie sur \mathcal{C} par $f(X) = \langle X, W \rangle$.

Montrer que $f(X) = g(x_1, x_2)$ pour tout élément $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathcal{C}$.

Déterminer l'ensemble des points en lesquels f atteint son maximum sur \mathcal{C} .

- 11.** Soient $n \in \mathbb{N}^*$, $\alpha_1, \dots, \alpha_n$ des réels strictement positifs tels que $\sum_{i=1}^n \alpha_i = 1$. On considère les fonctions f et g définies sur \mathbb{R}_+^n par

$$f(x_1, \dots, x_n) = \prod_{i=1}^n x_i^{\alpha_i} \text{ et } g(x_1, \dots, x_n) = \sum_{i=1}^n \alpha_i x_i.$$

On pose $\Gamma = \{(x_1, \dots, x_n) \in \mathbb{R}_+^n \mid g(x_1, \dots, x_n) = 1\}$.

1. Montrer que f possède un maximum M sur Γ et que celui-ci est atteint sur $\Gamma \cap (\mathbb{R}_+^*)^n$.
2. Déterminer les points critiques de f sur $(\mathbb{R}_+^*)^n$ sous la contrainte $g(x_1, \dots, x_n) = 1$. Montrer que $M = f(1, \dots, 1) = 1$.
3. En déduire que

$$\forall (x_1, \dots, x_n) \in \mathbb{R}_+^n, \quad \prod_{i=1}^n x_i^{\alpha_i} \leq \sum_{i=1}^n \alpha_i x_i.$$

12. Soit n un entier supérieur ou égal à 2 et f la fonction définie sur $(\mathbb{R}_+^*)^n$ par

$$f(x_1, \dots, x_n) = \sum_{k=1}^n x_k^4.$$

Minimiser f sous la contrainte $x_1 + \dots + x_n = n$.

13. Soient α, β, γ trois réels strictement positifs, ABC un triangle (les sommets du triangle sont identifiés à des éléments de \mathbb{R}^2). On note $a = BC$, $b = CA$ et $c = AB$ les longueurs des côtés du triangle. Pour tout M intérieur (au sens large) au triangle, on pose

$$f(M) = x^\alpha y^\beta z^\gamma,$$

où x, y et z sont les distances respectives de M aux côtés (BC) , (CA) et (AB) du triangle. On cherche à trouver le maximum de f .

1. Montrer que, pour tout point M , on a $ax + by + cz = S$, où S est le double de l'aire du triangle.

Montrer réciproquement que, pour tout triplet (x, y, z) de réels positifs tels que $ax + by + cz = S$, il existe un point M intérieur au triangle dont les distances respectives aux trois côtés du triangle sont x, y, z .

On est donc ramené à maximaliser la fonction $g : (\mathbb{R}_+)^3 \rightarrow \mathbb{R}$ définie par

$$g(x, y, z) = x^\alpha y^\beta z^\gamma$$

sous la contrainte $ax + by + cz = S$.

2. Montrer que $\{(x, y, z) \in (\mathbb{R}_+^*)^3, ax + by + cz = S\}$ est un fermé borné de \mathbb{R}^3 . En déduire l'existence du maximum.

3. Montrer que si ce maximum est atteint en (x_0, y_0, z_0) , on a $x_0 y_0 z_0 \neq 0$. On va donc se placer sur l'ouvert $\Omega = (\mathbb{R}_+^*)^3$ de \mathbb{R}^3 .

4. Écrire la condition nécessaire d'optimisation sous contrainte. En déduire qu'il existe dans Ω un point critique unique sous la contrainte $ax + by + cz = S$. Le déterminer. Préciser la valeur du maximum cherché.

Variables aléatoires réelles discrètes

8

Les simulations de variables aléatoires discrètes suivant une loi donnée ont été regroupées dans le dernier chapitre, consacré aux interventions informatiques, avec les simulations de variables aléatoires à densité. La théorie de ces simulations est présentée dans le chapitre portant sur l'estimation.

On trouvera ci-dessous des rappels et quelques extensions de résultats développés dans le livre de première année, dont certains portent sur les variables aléatoires quelconques. On trouvera également, en fin d'ouvrage, un tableau récapitulatif et des tables numériques relatifs aux principales variables aléatoires discrètes classiques. Ces variables aléatoires doivent être bien maîtrisées à l'issue de la première années. On trouvera aussi des exercices concernant d'autres variables aléatoires discrètes classiques.

Dans cet ouvrage, conformément à ce que nous avons écrit dans le livre de première année, nous utiliserons les notations [...] (comme $[X \in A]$) pour écrire un événement mais, comme nous écrivons $f(x, y)$ et non $f((x, y))$, nous garderons le symbolisme plus clair $P(X \in A)$ pour écrire la probabilité de ce même événement, au lieu du bien pédant $P([X \in A])$. Cela dit, la probabilité $P(|X| > 1)$ sera naturellement toujours écrite sous la forme

$$P(|X| > 1) = P([X > 1] \cup [X < -1])$$

puisque, dans ces conditions, la référence ensembliste est impérative.

Cette simplification est d'ailleurs conforme à certaines rubriques du programme de première année, dans sa définition de la fonction de répartition F_X ou de la loi géométrique par exemple, bien que l'on y trouve aussi une écriture des plus discutables telle que $P[X = x]$.

1. Généralités sur les variables aléatoires réelles

Rappelons qu'une variable aléatoire réelle X sur l'espace probabilisé (Ω, \mathcal{T}, P) est une application de Ω dans \mathbb{R} telle que, pour tout borélien $B \in \mathcal{B}$, on dispose de la relation $X^{-1}(B) \in \mathcal{T}$ (en fait, il faut et il suffit que cela soit vérifié pour tout intervalle de \mathbb{R} , et même pour tout intervalle semi-ouvert $B =]a, b]$, et même pour toute demi-droite $B =]-\infty, b]$).

Une présentation mathématique moderne satisfaisante du calcul des probabilités ne peut que s'appuyer sur la complexe théorie de la mesure, mais nous ne soulèverons pas ici davantage ce point.

Rappelons qu'un élément ω de Ω est appelé un *possible*, une *épreuve* ou une *éventualité* : en général ce n'est pas événement – auquel cas il serait *élémentaire* –, mais il admet une image $X(\omega)$ par la variable aléatoire X .

Nous admettrons que les opérations usuelles (par exemple la somme ou le produit) portant sur des variables aléatoires réelles définissent encore des variables aléatoires réelles.

1.1 Indépendance de variables aléatoires réelles

Rappelons deux définitions données dans le livre de première année à propos de variables aléatoires réelles discrètes.

Définition 1

Pour tout n -uplet (X_1, \dots, X_n) de variables aléatoires réelles discrètes de l'espace probabilisé (Ω, \mathcal{T}, P) tel que, pour tout $(x_1, \dots, x_n) \in X_1(\Omega) \times \dots \times X_n(\Omega)$,

$$P([X_1 = x_1] \cap \dots \cap [X_n = x_n]) = P(X_1 = x_1) \cdots P(X_n = x_n)$$

les variables aléatoires X_1, \dots, X_n sont dites **mutuellement indépendantes**.

Définition 2

Pour toute suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles discrètes de l'espace probabilisé (Ω, \mathcal{T}, P) telle que pour toute partie finie I de \mathbb{N} , les variables aléatoires réelles discrètes X_i où i décrit I sont mutuellement indépendantes, la suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles discrètes est appelée **suite de variables aléatoires mutuellement indépendantes**.

Rappelons également une propriété très importante, aussi démontrée dans ce livre.

Proposition 1

Pour tout couple (X_1, X_2) de variables aléatoires réelles discrètes de l'espace probabilisé (Ω, \mathcal{T}, P) , les variables aléatoires réelles discrètes X_1 et X_2 sont indépendantes

- si, et seulement si, pour tout couple (i, j) de $X_1(\Omega) \times X_2(\Omega)$, les événements $[X_1 = i]$ et $[X_2 = j]$ sont indépendants ;
- si, et seulement si, pour toute partie A de $X_1(\Omega)$ et pour toute partie B de $X_2(\Omega)$, les événements $[X_1 \in A]$ et $[X_2 \in B]$ sont indépendants.

Ces définitions et propriétés sont étendues comme suit aux variables aléatoires réelles quelconques, sous des formes un peu plus abstraites.

Définition 3

Un n -uplet (X_1, X_2, \dots, X_n) de variables aléatoires réelles définies sur un espace probabilisé (Ω, \mathcal{T}, P) est dite formée de variables aléatoires réelles mutuellement indépendantes si, pour tout n -uplet (A_1, A_2, \dots, A_n) de boréliens, on dispose de l'égalité

$$P\left(\bigcap_{i=1}^n [X_i \in A_i]\right) = \prod_{i=1}^n P(X_i \in A_i).$$

En particulier, si l'on note F_i la fonction de répartition de X_i , on dispose, pour tout $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, de la relation

$$P\left(\bigcap_{i=1}^n [X_i \leq x_i]\right) = \prod_{i=1}^n F_i(x_i).$$

Nous admettrons que, comme dans le cas discret, **cette dernière propriété est caractéristique des suites de variables mutuellement indépendantes.**

Une démonstration possible repose sur la notion d'indépendance des tribus \mathcal{A}_i associées aux variables X_i . Vérifions simplement, à titre d'exemple, un cas très particulier qui repose sur le fait que, si U et V sont des événements indépendants, il en est de même de \overline{U} et de \overline{V} : si X et Y vérifient cette propriété, alors les événements $[X > x]$ et $[Y > y]$ sont indépendants. On peut d'ailleurs vérifier directement les égalités

$$\begin{aligned} P([X > x] \cap [Y > y]) &= P(X > x) + P(Y > y) - P([X > x] \cup [Y > y]) \\ &= 2 - P(X \leq x) - P(Y \leq y) - P([X > x] \cup [Y > y]), \\ P([X > x] \cup [Y > y]) &= 1 - P([X \leq x] \cap [Y \leq y]) \\ &= 1 - P(X \leq x) P(Y \leq y) \end{aligned}$$

d'où $P([X > x] \cap [Y > y]) = (1 - P(X \leq x)) (1 - P(Y \leq y)) = P(X > x) P(Y > y)$.

De la même manière, $P([X > x] \cap [Y \leq y]) = P(X > x) P(Y \leq y)$ car \overline{U} et V sont indépendants comme l'étaient U et V etc. Le lecteur pourra s'exercer à vérifier de façon analogue (mais plus pénible) l'indépendance des événements $[a < X \leq b]$ et $[c < Y \leq d]$ et ainsi de suite.

Définition 4

Une famille dénombrable (X_i) de variables aléatoires réelles définies sur un espace probabilisé (Ω, \mathcal{T}, P) est dite formée de variables aléatoires réelles mutuellement indépendantes si toutes ses sous-familles finies sont formées de variables aléatoires réelles mutuellement indépendantes.

Dans la pratique, ces définitions très lourdes sont fort peu utilisées; elles n'interviennent que dans l'énoncé de théorèmes essentiels sur des suites (X_n) de variables aléatoires supposées mutuellement indépendantes.

1.2 Fonctions de répartition

Définition 5

La fonction de répartition F d'une variable aléatoire X est la fonction de \mathbb{R} dans $[0, 1]$ définie par l'égalité

$$F(x) = P(X \leq x).$$

Théorème 1

Une fonction de répartition F d'une variable aléatoire X est

- croissante [au sens large comme on le dit parfois, c'est-à-dire telle que $x < y$ implique $F(x) \leq F(y)$],
- continue à droite [c'est-à-dire telle que la limite $F(x + 0)$ de $F(x + h)$ lorsque h tend vers 0 par valeurs positives est égale à $F(x)$],
- telle que $\lim_{-\infty} F = 0$ et $\lim_{+\infty} F = 1$.

Pour tout réel x , la limite de $F(x - h)$ lorsque h tend vers 0 par valeurs strictement positives, notée $F(x - 0)$ ou $F(x^-)$, vaut $F(x) - P(X = x)$. En ce point, F est continue si, et seulement si, $P(X = x) = F(x) - F(x - 0) = 0$.

Preuve

- La croissance résulte simplement de l'inclusion $[X \leq x] \subset [X \leq y]$.
- Les autres propriétés peuvent être vues comme conséquences d'un lemme unique selon lequel la probabilité $P(A_n)$ d'une suite décroissante d'événements d'intersection vide tend vers 0 lorsque n tend vers l'infini, résultant trivialement de la propriété de la limite monotone.

(i) Pour démontrer la continuité à droite, il suffit de poser

$$A_n = \left[x < X \leq x + \frac{1}{n+1} \right]$$

de probabilité $F\left(x + \frac{1}{n+1}\right) - F(x)$. Il en résulte que la suite $F\left(x + \frac{1}{n+1}\right)$ tend vers $F(x)$; comme l'on sait que $F(x + 0)$ existe par suite de la croissance de F , il en résulte que $F(x + 0) = F(x)$.

(ii) Pour démontrer que $\lim_{-\infty} F = 0$, il suffit de poser

$$A_n = [X \leq -n]$$

de probabilité $F(-n)$. Comme l'intersection des A_n est vide, la suite $(F(-n))$ tend vers 0. Comme la croissance de F implique l'existence de $\lim_{-\infty} F$, cette limite est donc égale à 0.

(iii) Pour démontrer que $\lim_{+\infty} F = 1$, il suffit de même de poser

$$A_n = [X > n]$$

de probabilité $1 - F(n)$.

(iv) Pour démontrer que $P(X = x) = F(x) - F(x - 0)$, il suffit de même de poser

$$A_n = \left[x - \frac{1}{n+1} < X < x \right]$$

de probabilité $F(x) - P(X = x) - F\left(x - \frac{1}{n+1}\right)$.

□

1.3 Le théorème fondamental d'existence

Si une fonction réelle F vérifie les conditions énumérées plus haut, on dit qu'il s'agit d'une *fonction de répartition*. Nous admettrons le théorème suivant, d'importance pratique faible mais évidemment fondamental quant à la théorie : toute fonction de ce type est la fonction de répartition d'au moins une variable aléatoire.

Théorème 2 (Définition d'une probabilité à partir d'une fonction de répartition)

Pour toute fonction réelle F définie sur \mathbb{R} et vérifiant les propriétés suivantes

- F est croissante (au sens large),
- F est continue à droite en tout point, c'est-à-dire que pour tout réel x , $F(x+h)$ tend vers $F(x)$ lorsque h tend vers 0 par valeurs positives,
- $\lim_{-\infty} F = 0$,
- $\lim_{+\infty} F = 1$,

il existe une unique probabilité P définie sur l'espace probablisable $(\mathbb{R}, \mathcal{B})$, où \mathcal{B} est la tribu des boréliens de \mathbb{R} , vérifiant l'égalité $P(]-\infty, x]) = F(x)$.

Dans ces conditions, il existe au moins une variable aléatoire réelle X dont F est la fonction de répartition.

► Remarques

- Rappelons que la tribu (ou σ -algèbre) des boréliens est la plus petite tribu contenant les intervalles ouverts $]a, b[$ de \mathbb{R} .
- Les points de discontinuité de F forment un ensemble au plus dénombrable; en effet, pour tout entier n , l'ensemble des points tels que $F(x) - F(x-0) = P(X = x) \leq \frac{1}{n}$ est formé d'au plus n points, et l'on sait qu'une réunion dénombrable d'ensembles au plus dénombrables est au plus dénombrable.
- Le lecteur pourra par exemple démontrer facilement ce théorème dans le cas, très particulier, où l'ensemble des points de discontinuité est inclus dans \mathbb{N} .

Théorème 3

Soit F la fonction de répartition d'une variable aléatoire réelle X et (a, b) un couple de réels vérifiant $a < b$. On dispose alors des neuf propriétés

- $P(X \in \{a\}) = P(X = a) = F(a) - F(a-0)$;
- $P(X \in]-\infty, a]) = P(X \leq a) = F(a)$;
- $P(X \in]a, +\infty]) = P(X > a) = 1 - F(a)$;
- $P(X \in]-\infty, a]) = P(X < a) = F(a) - P(X = a) = F(a-0)$;
- $P(X \in [a, +\infty]) = P(X \geq a) = P(X = a) + 1 - F(a) = 1 - F(a-0)$;
- $P(X \in]a, b]) = P(a < X \leq b) = F(b) - F(a)$;
- $P(X \in [a, b]) = P(a \leq X < b) = F(b) - F(a) - P(X = b) + P(X = a)$
 $= F(b-0) - F(a-0)$;
- $P(X \in]a, b]) = P(a < X < b) = F(b) - P(X = b) - F(a) = F(b-0) - F(a)$;
- $P(X \in [a, b]) = P(a \leq X \leq b) = P(X = a) + F(b) - F(a) = F(b) - F(a-0)$.

Preuve

La première et la quatrième propriété découlent de l'égalité

$$P(X = x) = F(x) - F(x - 0).$$

La seconde est la définition même de F .

La troisième et la cinquième reposent sur le lien entre les probabilités d'un événement et de son contraire.

Enfin la sixième traduit la réunion disjointe

$$[X \leq a] \cup [a < X \leq b] = [X \leq b].$$

Les autres en découlent aussitôt. □

► **Remarques**

- Les propriétés ci-dessus montrent que les probabilités pour que X appartienne à l'un quelconque des intervalles de \mathbb{R} peuvent s'exprimer à partir des valeurs de F .
Bien que nous ne puissions le démontrer dans le cadre de ce cours, cela implique que la probabilité de tout borélien $B \in \mathcal{B}$ est connue dès que F est connue.
- La condition $a < b$ est essentielle : par exemple, si a est un point de discontinuité de F , on a $P(]a, a]) = 0$ et non $F(a - 0) - F(a)$, qui est d'ailleurs strictement négatif.

Corollaire 1

Soit F la fonction de répartition d'une variable aléatoire réelle X et c un réel positif ou nul. On dispose alors des relations

$$P(|X| \leq c) = P(-c \leq X \leq c) = F(c) - F(-c) + P(X = -c) = F(c) - F(-c - 0).$$

Si de plus c n'est pas nul, alors

$$P(|X| < c) = P(-c < X < c) = F(c) - F(-c) - P(X = c) = F(c - 0) - F(-c).$$

Corollaire 2

Soit X une variable aléatoire réelle de fonction de répartition F . Alors $Y = |X|$ est une variable aléatoire réelle, dont la fonction de répartition G , définie par $G(y) = P(Y \leq y)$, vérifie $G(y) = 0$ pour $y < 0$ et sinon

$$G(y) = F(y) - F(-y) + P(X = -y).$$

Proposition 2

Soient X une variable aléatoire réelle de fonction de répartition F et a un réel. Alors $Y = aX$ est une variable aléatoire réelle, dont la fonction de répartition G , définie par $G(y) = P(Y \leq y)$, vérifie les propriétés

- si $a = 0$: $G(y) = 0$ si $y < 0$, $G(y) = 1$ sinon ;
- si $a > 0$: $G(y) = F\left(\frac{y}{a}\right)$;
- si $a < 0$: $G(y) = P\left(X = \frac{y}{a}\right) + 1 - F\left(\frac{y}{a}\right) = 1 - F\left(\frac{y}{a} - 0\right)$.

Preuve

Que Y soit une variable aléatoire réelle vérifiant les propriétés annoncées traduit simplement des égalités ensemblistes presque évidentes, sauf peut-être la dernière, où $a < 0$

$$\{\omega \mid aX(\omega) \leq y\} = \left[X \geq \frac{y}{a} \right] = \left[X = \frac{y}{a} \right] \cup \left[X > \frac{y}{a} \right] = \left[X = \frac{y}{a} \right] \cup \left(\Omega \setminus \left[X \leq \frac{y}{a} \right] \right). \quad \square$$

► Remarque

On peut étudier de même les variables définies par $Y = aX^n$ avec $n \in \mathbb{N}$ et, avec un peu plus de précautions, $n \in \mathbb{Z}$.

Exercice 1. (Maximum et minimum de variables uniformes indépendantes)

Soit (X_1, X_2, \dots, X_n) une famille finie de variables aléatoires réelles mutuellement indépendantes à valeurs entières suivant la loi uniforme sur $\llbracket 1, N \rrbracket$, et $S = \sup_{k \in \llbracket 1, n \rrbracket} (X_k)$, $I = \inf_{k \in \llbracket 1, n \rrbracket} (X_k)$. Étudier les fonctions de répartition et les lois de probabilité des deux variables aléatoires discrètes S et I .

Le cas particulier $n = 2$ a déjà été étudié dans le dernier chapitre du livre de première année.

- Soit $1 \leq h \leq N$. Par définition, S prend ses valeurs également dans $\llbracket 1, N \rrbracket$ et sa fonction de répartition F_S vérifie

$$F_S(h) = P(S \leq h) = P(\{\omega \mid \forall i \in \llbracket 1, n \rrbracket \quad X_i(\omega) \leq h\}) = \prod_{i=1}^n P(X_i \leq h).$$

On trouve donc $F_S(h) = \left(\frac{h}{N}\right)^n$.

Intuitivement, ce résultat peut être retrouvé de la manière suivante : l'ensemble des N^n résultats possibles, tous équiprobables, peut-être vu comme un « hypercube » de \mathbb{R}^n dont le côté est formé de N points. Ceux qui satisfont à l'inégalité $S \leq h$ forment un hypercube inclus dans un « coin » du précédent, dont le côté est formé de h points, d'où le résultat.

Par suite,

$$P(S = 1) = F_S(1) = \frac{1}{N^n}$$

et

$$P(S = h + 1) = F_S(h + 1) - F_S(h) = \frac{(h + 1)^n - h^n}{N^n}$$

pour $1 \leq h < N$.

- Le cas de I est très légèrement différent. Ici nous devons passer par l'événement contraire $[I > h] = \bigcap_{i=1}^n [X_i > h]$. Par suite

$$F_I(h) = 1 - \prod_{i=1}^n \frac{N - h}{N} = 1 - \left(\frac{N - h}{N}\right)^n.$$

Intuitivement, ce résultat peut être retrouvé de la manière suivante : l'ensemble des N^n résultats possibles qui satisfont à l'inégalité $I \leq h$ forment le complémentaire d'un hypercube inclus dans un « coin » du précédent, dont le côté est formé de $N - h + 1$ points, d'où le résultat.

Le calcul de la loi de probabilité de I s'en déduit aussitôt.

1.4 Espérance et variance des variables aléatoires réelles

La définition de moments d'une variable aléatoire réelle a été rencontrée dans le livre de première année dans le cadre des variables discrètes ; une autre sera donnée plus loin dans le cadre des variables à densité.

Il existe une définition plus générale, couvrant ces deux cas, concernant les variables aléatoires **mesurables**, reposant sur la notion délicate d'**intégration par rapport à une mesure de probabilité**, qui dépasse très largement le niveau de la classe (voir par exemple le chapitre 9, page 55, de **L'essentiel en théorie des probabilités** par Jean Jacod et Philippe Protter, éditions Cassini 2003).

On ne donnera donc pas ici de définitions concernant par exemple l'espérance ou la variance d'une variable aléatoire réelle « quelconque ». Par contre, conformément au programme, nous admettrons les théorèmes suivants qui peuvent paraître comme raisonnables compte tenu de l'étude des variables aléatoires réelles discrètes.

Théorème 4

Soient X une variable aléatoire réelle admettant une espérance mathématique et a un réel ; alors le produit aX est une variable aléatoire réelle admettant une espérance $E(aX) = aE(X)$.

Théorème 5

Soient X et Y deux variables aléatoires réelles admettant des espérances mathématiques ; alors leur somme $X + Y$ est une variable aléatoire réelle admettant une espérance $E(X + Y) = E(X) + E(Y)$.

► Remarque

On traduit ces deux théorèmes en disant que l'espérance est *linéaire*.

Théorème 6

Soient X et Y deux variables aléatoires réelles admettant des espérances mathématiques et telles que $P(X > Y) = 0$ (on dit que $X \leq Y$ **presque sûrement**).

On dispose alors de l'inégalité $E(Y) \leq E(X)$.

► Remarque

On traduit ce théorème en disant que l'espérance est *croissante* ou *positive*.

Corollaire 3

Soient X et Y deux variables aléatoires réelles admettant des espérances mathématiques et telles que $P(X = Y) = 1$ (on dit que $X = Y$ **presque sûrement**).

On dispose alors de l'égalité $E(Y) = E(X)$.

Le dernier de la liste n'est pas le moindre.

Théorème 7

Soient X et Y deux variables aléatoires réelles **indépendantes** admettant des variances ; alors leur somme $X + Y$ est une variable aléatoire réelle admettant une variance $V(X + Y) = V(X) + V(Y)$.

2. Séries doubles convergentes

Une grande partie du matériel ci-dessous a déjà été présentée dans le livre de première année. Elle était en effet déjà indispensable dès que l'on sortait du strict cadre des variables aléatoires finies.

2.1 Suites et séries doubles

Définition 6

Une application u de \mathbb{N}^2 dans \mathbb{R} , dont le terme général est noté $u(i, j)$ ou plus commodément $u_{i,j}$, est appelée une **suite double**.

Elle est dite **positive** si l'on a $u_{i,j} \geq 0$ pour tout couple (i, j) d'entiers.

On appelle **série double** $\sum u_{i,j}$ associée à la suite double u l'ensemble, noté $\sum u_{i,j}$, des nombres de la forme

$$\sum_{i \in A} \sum_{j \in B} u_{i,j}$$

où A et B sont des parties finies arbitraires de \mathbb{N} .

Elle est dite **positive** si u est positive.

Toute série double définit des séries au sens usuel en choisissant $A = \{i\}$ ou $B = \{j\}$, notées respectivement $\sum_j u_{i,j}$ et $\sum_i u_{i,j}$.

L'étude de la série double est essentiellement constituée de l'étude de la nature de ces séries et des rapports entre leurs sommes éventuelles.

► **Remarque**

Une telle série est parfois notée $\sum u_j^i$.

Proposition 3

Toute suite (resp. série) double est la différence de deux suites (resp. séries) positives.

Preuve

Il suffit de poser $u_{ij} = v_{ij} - w_{ij}$ où

$$v_{ij} = \frac{|u_{ij}| + u_{ij}}{2} = \max(u_{ij}, 0), \quad w_{ij} = \frac{|u_{ij}| - u_{ij}}{2} = -\min(u_{ij}, 0).$$

□

2.2 Sommation par paquets

Rappelons le résultat suivant, établi dans le livre de première année.

Théorème 8 (Somme par paquets)

Soit $\sum u_n$ une série **absolument convergente**, f une application de \mathbb{N} dans \mathbb{N} .

Pour m et n dans \mathbb{N} , on pose $v_{n,m} = u_n$ si $f(n) = m$ et 0 sinon.

Pour tout entier naturel m , la série de terme général $v_{n,m}$ est convergente ; on note v_m sa somme. Alors, la série de terme général v_m est absolument convergente et l'on a

$$\sum_{m=0}^{+\infty} v_m = \sum_{n=0}^{+\infty} u_n.$$

Preuve

On a, pour tout $(m, n) \in \mathbb{N}^2$, $|v_{n,m}| \leq |u_n|$. Par hypothèse, la série $\sum |u_n|$ est convergente. D'après le théorème de comparaison des séries à termes positifs, pour tout $m \in \mathbb{N}$ la série $\sum |v_{n,m}|$ converge (la variable de sommation est n), donc $\sum v_{n,m}$ converge absolument. Soit v_m sa somme.

On note V_M et U_N respectivement les sommes partielles d'indice M et N des séries $\sum v_m$ et $\sum u_n$. Par linéarité de l'application qui à une série convergente associe sa somme, on a

$$V_M = \sum_{k=0}^M v_k = \sum_{n=0}^{+\infty} (v_{n,0} + \dots + v_{n,M}).$$

Puisque $v_{n,m} = u_n$ si $f(n) = m$ et 0 sinon, on obtient, puisqu'un terme au plus dans cette somme n'est pas nul,

$$v_{n,0} + \dots + v_{n,M} = u_n \quad \text{si } f(n) \leq M \quad \text{et } 0 \text{ sinon.}$$

Si M est supérieur ou égal à $\max(f(0), f(1), \dots, f(N))$, on a alors $v_{n,0} + \dots + v_{n,M} = u_n$ pour $n \leq N$. On obtient donc

$$V_M = U_N + \sum_{n=N+1}^{+\infty} (v_{n,0} + \dots + v_{n,M}) = U_N + \sum_{\substack{n \geq N+1 \\ f(n) \leq M}} u_n$$

puis l'inégalité

$$|V_M - U_N| \leq \sum_{n=N+1}^{+\infty} |u_n|.$$

Soit U la somme de la série $\sum u_n$ et $\varepsilon > 0$. Puisque la série $\sum |u_n|$ converge, il existe N tel que le reste d'indice N vérifie $\sum_{n=N+1}^{+\infty} |u_n| \leq \varepsilon$. On en déduit

$$|U - U_N| = \left| \sum_{n=N+1}^{+\infty} u_n \right| \leq \sum_{n=N+1}^{+\infty} |u_n| \leq \varepsilon.$$

Pour $M \geq \max(f(0), \dots, f(N))$, on obtient

$$|V_M - U| \leq |V_M - U_N| + |U_N - U| \leq 2\varepsilon.$$

Cela montre que $\lim_{M \rightarrow +\infty} V_M = U$. Ainsi, la série $\sum v_m$ converge et sa somme est U .

On peut noter que $\sum v_m$ est même absolument convergente, car les sommes partielles de $\sum |v_m|$ sont majorées. En effet, d'après ce qui précède, on a pour tout entier M , les inégalités

$$\sum_{k=0}^M |v_k| \leq \sum_{n=0}^{+\infty} (|v_{n,0}| + \dots + |v_{n,M}|) \leq \sum_{n=0}^{+\infty} |u_n|. \quad \square$$

► Remarque

Éclairons la signification de ce théorème : v_m représente la somme des u_n pour tous les n tels que $f(n) = m$ (un tel ensemble d'entiers n est l'un des paquets qui donne son nom au théorème); on peut écrire $v_m = \sum_{n \in f^{-1}(m)} u_n$.

Calculer la somme de la série $\sum u_n$ revient à sommer les sommes v_m des différents paquets. Il s'agit donc d'un théorème d'**associativité**.

Corollaire 4

Si la série $\sum u_n$ converge absolument, alors pour toute bijection σ de \mathbb{N} , la série $\sum u_{\sigma(n)}$ converge absolument et

$$\sum_{n=0}^{+\infty} u_n = \sum_{n=0}^{+\infty} u_{\sigma(n)}.$$

Preuve

On pose $f = \sigma^{-1}$. On reprend les notations du théorème précédent. On a $v_{n,m} = u_n$ si $f(n) = m$, c'est-à-dire si $n = \sigma(m)$, et $v_{n,m} = 0$ sinon. On en déduit que $v_m = u_{\sigma(m)}$ et le corollaire découle immédiatement du théorème 8. \square

► Remarques

- Soit A une partie dénombrable de \mathbb{R} . On peut écrire les éléments de A sous forme d'une suite de réels $(u_n)_{n \in \mathbb{N}}$. On obtient $A = \{u_n \mid n \in \mathbb{N}\}$. Changer la numérotation des éléments de A , c'est écrire $A = \{u_{\sigma(n)} \mid n \in \mathbb{N}\}$, où σ est une bijection de \mathbb{N} . Si la série $\sum u_n$ converge absolument, on a pour tout σ , $\sum_{n=0}^{+\infty} u_n = \sum_{n=0}^{+\infty} u_{\sigma(n)}$.

Cette somme peut s'écrire $\sum_{a \in A} a$, car elle ne dépend que de A et non de la numérotation de ses éléments.

- Ce théorème est faux pour une série semi-convergente. On peut modifier la somme d'une série semi-convergente car celle-ci ne dépend pas intrinsèquement de l'ensemble des u_n mais de leur ordre. On peut même trouver σ pour que la somme de la série soit un nombre donné à l'avance, ou tel que la série devienne divergente.

Proposition 4

Pour qu'une série $\sum u_n$ soit absolument convergente, il faut et il suffit qu'il existe un majorant M commun à toutes les sommes $\sum_{n \in A} |u_n|$ où A décrit l'ensemble des parties finies de \mathbb{N} .

Preuve

La condition nécessaire est évidente, en posant $M = \sum_{n \in \mathbb{N}} |u_n|$.

Inversement, si M majore toutes les sommes finies $\sum_{n \in A} |u_n|$, il majore en particulier les réels de la suite croissante

$$S_N = \sum_{n=0}^N |u_n|, \text{ d'où l'existence de } \sum_{n=0}^{+\infty} |u_n|. \quad \square$$

Application aux séries doubles à termes quelconques

Fubini

Soit $\sum u_{i,j}$ une série double.

Si, pour tout entier i , la série $\sum_j u_{i,j}$ converge **absolument** et est de somme a_i , si la série

$\sum_j (\sum_i |u_{i,j}|)$ converge, alors la série $\sum a_i$ converge **absolument** vers une somme S et,

pour tout entier j , la série $\sum_i u_{i,j}$ converge **absolument** vers un réel b_j tel que la série

$\sum b_j$ converge **absolument** vers S , ce que l'on écrit

$$\sum_i \sum_j u_{i,j} = \sum_j \sum_i u_{i,j}.$$

On dit alors que la série double $\sum u_{i,j}$ est **absolument convergente**, et de somme S .

Preuve

Le carré \mathbb{N}^2 est dénombrable car il existe une bijection s de \mathbb{N}^2 sur \mathbb{N} , par exemple

$$s(i, j) = 2^i(2j + 1) - 1.$$

Si $k = s(i, j)$, nous noterons $\widehat{u}_k = u_{i,j}$. Les convergences absolues conduisant aux sommes a_i , b_j et S montrent que la série $\sum \widehat{u}_k$ est absolument convergente, puisque les sommes formées d'un nombre fini des $|u_{i,j}|$ sont toutes majorées par le réel $\sum_i \sum_j |u_{i,j}|$. Soit S sa somme.

On peut donc appliquer le théorème de sommation par paquets. Dans un premier temps, avec les notations ci-dessus, nous choisirons l'application f définie par $f(k) = f(s(i, j)) = i$; les sommes de ces paquets sont alors exactement les a_i , et leur somme est S . Prenant ensuite l'application f définie par $f(k) = j$, on obtient les b_j et la même conclusion. □

► **Remarque**

La condition ci-dessus est suffisante pour que les sommes des séries $\sum a_i$ et $\sum b_j$ existent et soient égales, mais non nécessaire comme le montre l'exemple $u_{i,j} = \frac{(-1)^{i+j}}{(i+1)(j+1)}$ où les a_i et b_j existent pour tout couple (i, j) et ont même somme $S = (\ln 2)^2$, mais sont définis par des séries non absolument convergentes.

Exercice 2.

Étudier si la série $\sum u_{i,j}$, où

$$u_{i,2k} = \frac{1}{(i+1)(k+1)(k+2)}, \quad u_{i,2k+1} = -u_{i,2k}$$

est absolument convergente (on pourra essayer de calculer les a_i et les b_j).

- Il est clair que $a_i = \sum_j u_{i,j}$ existe pour tout i et vaut 0. Il existe donc bien

$$S = \sum_i a_i = 0 \text{ (et cette série simple est même absolument convergente).}$$

Par contre il n'existe pas de b_j puisque, par exemple

$$\sum_{i=0}^N u_{i,2k} = \frac{1}{(k+1)(k+2)} \sum_{i=0}^N \frac{1}{i+1}$$

tend vers l'infini avec N (il en irait de même pour b_{2k+1}).

Par suite la série n'est pas absolument convergente.

- On peut en déduire que la série $\sum_i a'_i$, où $a'_i = \sum_j |u_{i,j}|$ ou bien n'a pas de sens parce que l'un des a'_i est infini, ou bien diverge.

En effet, pour tout i , l'égalité $\frac{1}{(k+1)(k+2)} = \frac{1}{k+1} - \frac{1}{k+2}$ montre qu'il existe a'_i , et que l'on a $a'_i = \frac{2}{i+1}$, ce qui est conforme à l'affirmation ci-dessus.

► Remarque

Comme on vient de le voir, même s'il est facile de s'y tromper, il ne suffit pas que la série de terme général $|a_i|$ soit convergente, le théorème de Fubini demandant, de manière plus exigeante, que la série de terme général a'_i soit (absolument) convergente, avec $a'_i = \sum_j |u_{i,j}| \geq |a_i|$: la convergence de $\sum a'_i$ implique donc celle de $\sum |a_i|$, sans que l'on puisse dire quoi que ce soit en sens inverse.

Proposition 5

Toute série double absolument convergente est la différence de deux séries doubles positives convergentes.

Preuve

Il suffit, ici encore, de poser $u_{i,j} = v_{i,j} - w_{i,j}$ où $v_{i,j} = \frac{|u_{i,j}| + u_{i,j}}{2}$, $w_{i,j} = \frac{|u_{i,j}| - u_{i,j}}{2}$. En effet l'absolue convergence d'une série réelle $\sum x_n$ signifie la convergence de la série $\sum |x_n|$ et donc celles de $\frac{|x_n| + x_n}{2}$ et de $\frac{|x_n| - x_n}{2}$. □

Proposition 6

Pour qu’une série double $\sum u_{i,j}$ soit absolument convergente, il faut et il suffit qu’il existe un majorant M commun à toutes les sommes $\sum_{i \in A} \sum_{j \in B} |u_{i,j}|$ où A et B décrivent respectivement les ensembles des parties finies de I et de J .

Preuve

Ce n’est que la traduction en termes de séries doubles, grâce à la bijection s introduite plus haut, de la proposition 4. □

Application aux séries doubles à termes positifs

Le corollaire ci-après est une conséquence immédiate du théorème de Fubini.

Corollaire 6

Soit $\sum u_{i,j}$ une série double positive.

Si, pour tout entier i , la série $\sum_j u_{i,j}$ converge vers un réel a_i , et si la série $\sum a_i$ converge vers un réel S , alors, pour tout entier j , la série $\sum_i u_{i,j}$ converge vers un réel b_j tel que la série $\sum b_j$ converge vers S .

La série double $\sum u_{i,j}$ est alors absolument convergente.

► **Remarques**

- Bien que le programme ne contienne que l’énoncé du corollaire ci-dessus, l’étude du précédent est pourtant indispensable, comme va le montrer l’étude de la formule de l’espérance totale.
- Il est facile d’énoncer et de démontrer une proposition analogue où les indices i et j décrivent des parties arbitraires I et J de \mathbb{N} . Si l’une de ces deux parties est finie, le résultat est naturellement immédiat.

Proposition 7

Soient deux suites doubles vérifiant les relations

$$\forall (i, j), \quad 0 \leq v_{i,j} \leq u_{i,j}.$$

Alors, pour que la série $\sum v_{i,j}$ soit absolument convergente, il suffit que la série $\sum u_{i,j}$ le soit.

Preuve

Gardons les notations précédentes. Pour tout i , nous disposons de la majoration

$$\sum_{j=0}^{+\infty} v_{i,j} \leq \sum_{j=0}^{+\infty} u_{i,j} = a_i$$

qui montre que cette série à termes positifs est convergente, donc absolument convergente, de somme positive majorée par a_i . Il en résulte aussitôt que sa somme est le terme général, majoré par S , d’une série positive, et l’absolue convergence de $\sum v_{i,j}$.

Cette condition, qui généralise un résultat bien connu de la théorie des séries ordinaires, n’est naturellement pas nécessaire. □

3. Indépendance de variables aléatoires réelles discrètes

Les généralités sur l'indépendance d'une suite de variables aléatoires réelles s'appliquent naturellement au cas des variables aléatoires discrètes, c'est-à-dire d'image au plus dénombrable.

On peut toutefois y être souvent plus précis, comme le montre l'exemple ci-dessous : la propriété qui y est prouvée est en fait valable dans le cas général, mais la démonstration est alors bien plus complexe.

Exercice 3.

Soit (X, Y, Z) un triplet de valeurs aléatoires réelles discrètes mutuellement indépendantes. Montrer que $X + Y$ et Z sont indépendantes.

Généraliser au cas d'une suite (X_n) de variables aléatoires réelles discrètes mutuellement indépendantes.

Soient respectivement $\{x_i \mid i \in I\}$ et $\{y_j \mid j \in J\}$ les images au plus dénombrables de X et Y . On dispose par exemple de l'égalité

$$\Omega = \bigcup_{i \in I} \bigcup_{j \in J} ([X = x_i] \cap [Y = y_j]).$$

Soit u un réel. L'ensemble des $\omega \in \Omega$ tels que $X(\omega) + Y(\omega) \leq u$ est l'événement

$$\bigcup_{i \in I} \bigcup_{j \in J_i} ([X = x_i] \cap [Y = y_j])$$

où $J_i = \{j \mid y_j \leq u - x_i\}$. La réunion ci-dessus étant au plus dénombrable, cela montre au passage que $X + Y$ est bien une variable aléatoire.

Cette réunion étant disjointe, X et Y étant indépendantes, on dispose donc par sommation par paquets de l'égalité

$$P(X + Y \leq u) = \sum_{i \in I} \sum_{j \in J_i} P(X = x_i) P(Y = y_j).$$

Soit v un réel. L'ensemble des $\omega \in \Omega$ tels que $X(\omega) + Y(\omega) \leq u$ et $Z(\omega) \leq v$ est l'événement

$$\bigcup_{i \in I} \bigcup_{j \in J_i} ([X = x_i] \cap [Y = y_j] \cap [Z \leq v]).$$

Ici encore, cette réunion étant disjointe, X , Y et Z étant mutuellement indépendantes, on dispose des égalités

$$\begin{aligned} P([X + Y \leq u] \cap [Z \leq v]) &= \sum_{i \in I} \sum_{j \in J_i} P(X = x_i) P(Y = y_j) P(Z \leq v) \\ &= P(X + Y \leq u) P(Z \leq v) \end{aligned}$$

qu'il fallait démontrer.

Une généralisation possible est la suivante : si les (X_k) sont mutuellement indépendantes, les variables aléatoires $X_0 + X_1 + \dots + X_n$ et X_{n+1} sont indépendantes, comme le montre une récurrence sur $n \geq 2$.

4. Espérance et conditionnement des variables discrètes

Reprenons ici l'introduction faite en première année des valeurs typiques des variables aléatoires discrètes, surtout vue du côté des variables finies, à la lumière des applications du théorème sur les séries doubles convergentes.

L'espace probabilisé sous-jacent aux variables aléatoires figurant dans ce cours de probabilités sera généralement noté (Ω, \mathcal{T}, P) où \mathcal{T} est une tribu, ou encore (Ω, \mathcal{A}, P) où \mathcal{A} est une σ -algèbre (les deux termes sont équivalents).

4.1 Espérance et variance

Généralisons très légèrement la définition donnée en première année de l'éventuelle espérance mathématique d'une variable aléatoire discrète pour $I = \mathbb{N}$.

Définition 7

Soit X une variable aléatoire réelle discrète sur l'espace probabilisé (Ω, \mathcal{T}, P) , d'image $X(\Omega) = \{x_i \mid i \in I\}$ où I est une partie de \mathbb{Z} . L'espérance $E(X)$ est définie comme la somme

$$E(X) = \sum_{i \in I} x_i P(X = x_i)$$

si cette dernière est une somme d'un nombre fini de termes ou, plus généralement, la somme d'une série **absolument convergente**.

Si $I = \mathbb{Z}$, on écrit encore parfois

$$E(X) = \sum_{i=-\infty}^{+\infty} x_i P(X = x_i)$$

(rappelons qu'un tel abus de notation de la somme d'une série n'est admis qu'en cas d'absolue convergence).

Cette définition s'étend aussitôt au cas où I est un ensemble quelconque au plus dénombrable.

Une extension analogue à celle de l'espérance est aussitôt étendue aux variables aléatoires réelles discrètes admettant une variance.

► Remarque

Bien entendu, l'égalité écrite exige la convergence absolue de la série, c'est-à-dire la convergence de la série $\sum |x_i| P(X = x_i)$, c'est-à-dire encore l'existence de $E(|X|)$, liée à $E(X)$ par l'inégalité évidente

$$|E(X)| \leq E(|X|).$$

Généralisons là encore un théorème rencontré dans le livre de première année.

Théorème 9 (Théorème de transfert)

Pour toute variable aléatoire réelle discrète d'image $X(\Omega) = \{x_i \mid i \in I\}$ (I au plus dénombrable) et pour toute fonction f de $X(\Omega)$ à valeurs dans \mathbb{R} , l'application $Y = f \circ X$, notée abusivement $Y = f(X)$, est une variable aléatoire de même type.

Pour qu'elle admette une espérance mathématique, il suffit que la série $\sum f(x_i)P(X = x_i)$ soit absolument convergente, et l'on dispose alors de l'égalité

$$E(Y) = E(f \circ X) = E(f(X)) = \sum_{i \in I} f(x_i)P(X = x_i).$$

Preuve

Montrons d'abord que l'image $\cup = Y(\Omega)$, image par f de l'ensemble au plus dénombrable $X(\Omega)$ est elle-même au plus dénombrable.

En effet, pour tout réel $y \in \cup$, on peut choisir un indice $i \in I$, noté $\varphi(y)$, vérifiant $y = f(x_i)$. L'application φ de \cup vers I ainsi définie est clairement injective et d'image $\varphi(\cup)$ au plus dénombrable car incluse dans I . Or \cup est en bijection avec cette image, et donc au plus dénombrable lui-même.

Pour $y \in \cup$, nous noterons I_y l'ensemble $f^{-1}(y) : I$ est la réunion disjointe des I_y . Par définition, $I_y = \{i \in I \mid f(x_i) = y\}$, d'où

$$P(Y = y) = P\left(\bigcup_{i \in I_y} [X = x_i]\right) = \sum_{i \in I_y} P(X = x_i),$$

$$y P(Y = y) = y \sum_{i \in I_y} P(X = x_i) = \sum_{i \in I_y} f(x_i)P(X = x_i).$$

Si I_y est fini, nous avons affaire à des sommes finies. Sinon, dans ces égalités, nous avons *a priori* affaire à des sommes de séries convergentes. En fait elle sont même absolument convergentes : la première, parce qu'elle est à termes positifs ; pour la seconde, la question pourrait se poser mais, les différents $f(x_i)$ étant égaux entre eux, ses termes ont tous même signe (celui de y).

Si la série de l'énoncé est absolument convergente, on peut lui appliquer le théorème de sommation par paquets en regroupant les indices i dans les différents sous-ensembles I_z de I . On dispose alors des égalités

$$\sum_{i \in I} |f(x_i)| P(X = x_i) = \sum_{y \in \cup} \sum_{i \in I_y} |f(x_i)| P(X = x_i) = \sum_{y \in \cup} |y| \sum_{i \in I_y} P(X = x_i)$$

$$= \sum_{y \in \cup} |y| P(Y = y),$$

$$\sum_{i \in I} f(x_i) P(X = x_i) = \sum_{y \in \cup} \sum_{i \in I_y} f(x_i) P(X = x_i) = \sum_{y \in \cup} y \sum_{i \in I_y} P(X = x_i)$$

$$= \sum_{y \in \cup} y P(Y = y)$$

et l'existence et la valeur de l'espérance de $Y = f(X)$ en découlent. □

Exercice 4.

Soit (X_n) une suite de variables aléatoires deux à deux indépendantes et de même loi admettant une variance σ^2 . Calculer l'espérance de la variable aléatoire définie par

$$Y_n = \frac{1}{n-1} \sum_{k=1}^{n-1} (X_k - X_n)^2.$$

Les variables $X_k - X_n$ sont centrées puisque X_k et X_n ont la même espérance. Par suite l'espérance de $(X_k - X_n)^2$ est égale à la variance $V(X_k - X_n)$, soit à $V(X_k) + V(X_n) = 2\sigma^2$ par indépendance de X_k et de X_n .

Par suite l'espérance demandée de Y_n vaut $2\sigma^2$.

► **Remarque**

Cet exercice fournit théoriquement une façon d'estimer un écart type inconnu à partir de simulations des Y_n . Lorsque l'on connaît l'espérance commune μ , une meilleure approximation de σ est donnée par l'expression

$$\frac{1}{\sqrt{n}} \sqrt{\sum_{k=1}^n (X_k - \mu)^2}$$

très utilisée dans la pratique (pour une justification, voir l'énoncé du théorème de la limite centrée).

Proposition 8

Soit X une variable aléatoire discrète d'image $X(\Omega) = \{x_i \mid i \in I\}$ admettant une variance $V(X)$. Alors cette variance est donnée par l'égalité

$$V(X) = \frac{1}{2} \sum_{(i,j)} (x_i - x_j)^2 P(X = x_i) P(X = x_j).$$

► **Remarque**

Le lecteur trouvera dans le livre de première année une première démonstration de cette égalité dans le cas particulier où I est fini, et pourra s'exercer à la généraliser au cas dénombrable.

Celle qui suit a été détaillée de façon à rappeler comment vérifier, dans un cas particulier, les hypothèses du théorème des séries doubles ci-dessus.

Preuve

Soit $u_{i,j} = (x_i - x_j)^2 P(X = x_i) P(X = x_j)$. Fixant i et notant $M_2(X)$ le moment d'ordre 2 de X qui existe d'après l'hypothèse d'existence de la variance, il vient

$$\begin{aligned} a_i &= \sum_j u_{i,j} = P(X = x_i) \sum_j (x_i^2 - 2x_i x_j + x_j^2) P(X = x_j) \\ &= P(X = x_i) (x_i^2 - 2x_i E(X) + M_2(X)). \end{aligned}$$

Cette série de somme a_i est bien entendu absolument convergente puisque $\sum x_i P(X = x_i)$ l'est.

La somme des $|u_{i,j}|$ à i fixée se calcule de la même manière : on trouve

$$a'_i = P(X = x_i) (x_i^2 + 2|x_i|E(|X|) + M_2(X))$$

et la série $\sum a'_i$ est convergente (sa somme est d'ailleurs égale à $2M_2(X) + 2E(|X|)^2$).

Par suite, le théorème sur les séries doubles s'applique, et l'on obtient effectivement

$$\sum_{i,j} u_{i,j} = \sum_i a_i = M_2(X) - 2E(X)^2 + M_2(X) = 2V(X).$$

□

Exercice 5.

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires positives définies sur le même espace probabilisé (Ω, \mathcal{T}, P) , (S_n) la suite définie par $S_0 = 0$ et $S_n = X_1 + X_2 + \dots + X_n$ si $n > 1$, les (F_n) étant les fonctions de répartition des (S_n) .

- a) Pour tout réel x positif ou nul, on note A la variable aléatoire discrète d'image incluse dans \mathbb{N} définie par la relation « n est le plus petit entier tel que $S_n > x$ », c'est-à-dire encore pour $n \geq 1$ par

$$[A = n] = [S_{n-1} \leq x] \cap [S_n > x].$$

Si $S_n \leq x$ pour tout n , on posera $A = 0$. Calculer la loi de A .

- b) Montrer que, pour que A ait une espérance, il suffit que la série $\sum F_n(x)$ soit convergente. Calculer alors cette espérance.
 c) Étudier le cas où X_n est la variable certaine égale à 2^{-n} et $x = 1$.

- Rappelons un lemme souvent utile : si E et F sont deux événements vérifiant $E \subset F$, alors on dispose des égalités

$$F = E \cup (F \setminus E) \quad \text{et} \quad P(F \setminus E) = P(F) - P(E)$$

[dans le cas général, on a seulement $P(F \setminus E) \geq P(F) - P(E)$].

Puisque les X_n sont positives, il en résulte que $[S_n \leq x]$ est inclus dans $[S_{n-1} \leq x]$, et donc que pour tout entier n strictement positif

$$P(A = n) = P([S_{n-1} \leq x] \cap [S_n > x]) = P(S_{n-1} \leq x) - P(S_n \leq x) = F_{n-1}(x) - F_n(x).$$

Il en résulte que, pour tout entier N , on dispose des égalités

$$\sum_{n=1}^N n P(A = n) = \sum_{n=1}^N n (F_{n-1}(x) - F_n(x)) = \left[\sum_{n=1}^N F_{n-1}(x) \right] - N F_N(x)$$

dont la dernière se prouve aisément par un changement d'indice.

On notera que $F_0(x) = 1$ et que la suite $(F_n(x))$ étant décroissante et minorée admet une limite $\lambda \in [0, 1]$.

Puisque, pour tout $N > 0$ on a $\sum_{n=1}^N P(A = n) = \sum_{n=1}^N (F_{n-1}(x) - F_n(x)) = 1 - F_N(x)$,

il en résulte l'égalité

$$P(A = 0) = 1 - \sum_{n=1}^{+\infty} P(A = n) = \lambda.$$

- Supposons la série $\sum F_n(x)$ convergente.

Les $F_n(x)$ étant positives, l'expression $\sum_{n=0}^N n P(A = n) = \sum_{n=1}^N n P(A = n)$ est donc

majorée par le nombre $\sum_{n=1}^{+\infty} F_{n-1}(x)$. La série dont on veut trouver la somme $E(A)$ étant à termes positifs, cette majoration indépendante de N implique sa convergence (absolue), et il vient

$$\lim_{n \rightarrow +\infty} N F_N(x) = \sum_{n=1}^{+\infty} F_{n-1}(x) - E(A) = \mu \geq 0.$$

Si l'on avait $\mu > 0$, il en résulterait que $F_n(x) \sim \frac{\mu}{n}$ et $\sum F_n(x)$ serait divergente ce qui n'est pas. Donc $\mu = 0$ et

$$E(A) = \sum_{n=1}^{+\infty} F_{n-1}(x) = \sum_{n=0}^{+\infty} F_n(x).$$

- On voit aussitôt que $S_n = 1 - \frac{1}{2^n} < 1 = x$, d'où $F_n(x) = 1$ pour tout entier n , soit $\lambda = 1$, $P(A = 0) = 1$ et $P(A = n + 1) = 0$. Ici, il existe donc $E(A)$, égale à 0. Par suite, l'existence de l'espérance $E(A)$ n'implique pas la convergence de $\sum F_n(x)$.

4.2 Espérance conditionnelle

Définition 8

Soit X une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{T}, P) et d'image au plus dénombrable $X(\Omega) = \{x_i \mid i \in I\}$. Pour tout événement A de \mathcal{T} de probabilité non nulle, on appelle **espérance de X conditionnée par A** , et l'on note $E(X \mid A)$, l'éventuelle espérance de X considérée comme définie sur l'espace probabilisé $(\Omega, \mathcal{T}, P_A)$.

► Remarque

Attention de ne pas confondre le bloc « $X \mid A$ » avec un événement : une remarque analogue a déjà été faite, en première année, au sujet des deux notations traditionnelles des probabilités conditionnelles, à savoir

$$P_A(E) = P(E \mid A) = \frac{P(E \cap A)}{P(A)}.$$

Proposition 9

Avec les notations ci-dessus, on dispose de l'égalité

$$E(X \mid A) = \sum_{i \in I} x_i P_A(X = x_i) = \frac{1}{P(A)} \sum_{i \in I} x_i P([X = x_i] \cap A).$$

Preuve

C'est la définition de l'espérance d'une variable donnée dans la section précédente. □

► **Remarques**

- Il est abusif d'écrire encore X la variable aléatoire associée à P_A comme l'était la variable initiale, mais cet usage est universel.
- Bien entendu, l'égalité obtenue exige la convergence absolue de la série, c'est-à-dire la convergence de la série $\sum |x_i| P_A(X = x_i)$, c'est-à-dire encore l'existence de $E(|X| | A)$.

Proposition 10

Avec les notations ci-dessus, l'existence de $E(X)$ implique celles de $E(|X|)$ et $E(X | A)$, et l'on dispose de l'inégalité

$$|E(X | A)| \leq \frac{1}{P(A)} E(|X|).$$

Preuve

Les inégalités $0 \leq |x_i| P_A(X = x_i) \leq \frac{|x_i|}{P(A)} P(X = x_i)$ montrent que la série $\sum x_i P_A(X = x_i)$ est absolument convergente.

De même, on dispose donc, pour toute partie finie J de I , de l'inégalité

$$\left| \sum_{j \in J} x_j P_A(X = x_j) \right| \leq \sum_{j \in J} \frac{|x_j|}{P(A)} P(X = x_j)$$

qui permet de comparer les sommes des deux séries. Le résultat annoncé en découle. □

4.3 Formule de l'espérance totale

Rappelons que, dans les ouvrages de cette collection, une famille d'événements $(A_j)_{j \in J}$ est un système complet si J est au plus dénombrable, si la réunion de ces éléments est Ω et s'ils sont deux à deux disjoints.

Le théorème suivant, très important, repose essentiellement sur le théorème de Fubini.

Théorème 10

Soit (Ω, \mathcal{T}, P) un espace probabilisé, X une variable aléatoire réelle discrète d'image $X(\Omega) = \{x_i | i \in I\}$ avec I dénombrable, et $(A_n)_{n \in \mathbb{N}}$ un système complet d'événements.

On suppose que

- Pour tout indice n tel que $P(A_n) > 0$, X admet une espérance pour la probabilité P_{A_n} , notée $E(X | A_n)$.
- La série positive $\sum e_n$ de terme général e_n égal à 0 si $P(A_n) = 0$ et à $\sum_i |x_i| P_{A_n}(X = x_i) P(A_n)$ sinon, est convergente.

Alors les nombres $E(X)$ et $\sum_n^* E(X | A_n) P(A_n)$, où l'astérisque signifie que l'on se limite aux indices n tels que $P(A_n) > 0$, existent et sont égaux

$$E(X) = \sum_n^* E(X | A_n) P(A_n).$$

Preuve

Notons d’abord que, par la définition des espérances conditionnelles, la première condition signifie justement que la série e_n existe pour tout n .

Soit la suite double définie par $u_{i,n} = x_i P([X = x_i] \cap A_n)$. Si $P(A_n) = 0$ alors $u_{i,n} = 0$. Si l’on fixe n tel que $P(A_n) > 0$, on voit que la somme sur i décrivant I des $u_{i,n}$ existe et vaut

$$\sum_{i \in I} u_{i,n} = P(A_n) \left[\sum_{i \in I} x_i P_{A_n}(X = x_i) \right] = E(X \mid A_n) P(A_n).$$

En sens opposé, si l’on fixe i dans I et l’on somme sur n décrivant \mathbb{N} , on trouve

$$\sum_{n=0}^{+\infty} u_{i,n} = x_i \sum_{n=0}^{+\infty} P([X = x_i] \cap A_n) = x_i P\left(\bigcup_{n \in \mathbb{N}} [X = x_i] \cap A_n\right) = x_i P(X = x_i).$$

Reste à traduire la fin de l’hypothèse : la seconde condition de l’énoncé du théorème affirme en fait qu’il existe $\sum_n \sum_i |u_{i,n}|$. La somme $\sum_{i,n} u_{i,n}$ existe donc bien, et les deux manières dont nous disposons pour la calculer donnent exactement les deux membres de l’égalité à démontrer. □

► **Remarques**

- Cette formule est naturellement à rapprocher d’une formule sur les probabilités conditionnelles d’un événement E généralisant le théorème des probabilités totales, à savoir

$$P(E) = \sum_n P(E \mid A_n) P(A_n).$$

- Remplacer dans la seconde condition

$$\sum_i |x_i| P_{A_n}(X = x_i) \quad \text{par} \quad \sum_i x_i P_{A_n}(X = x_i),$$

même si l’on suppose la convergence absolue de $\sum e_n$, pourrait conduire à des erreurs : voir l’exercice 18.

Proposition 11

Soit (Ω, \mathcal{T}, P) un espace probabilisé, X une variable aléatoire réelle discrète d’image $X(\Omega) = \{x_i \mid i \in I\}$ avec I dénombrable, et $(A_n)_{n \in \mathbb{N}}$ un système complet d’événements.

On suppose que X admet une espérance $E(X)$. Si l’on se limite aux indices n tels que $P(A_n) > 0$, les nombres $E(X \mid A_n)$ existent alors, la série $\sum^* E(X \mid A_n) P(A_n)$ est absolument convergente et admet $E(X)$ comme somme.

Preuve

La proposition 10 implique que l’existence de $E(X)$ implique celles des $E(X \mid A_n)$. La proposition 7, l’inégalité $|E(X \mid A_n) P(A_n)| \leq E(|X|) P(A_n)$ et l’existence de $E(|X|)$ montrent que la série $\sum^* E(X \mid A_n) P(A_n)$ est absolument convergente, et il suffit alors d’utiliser le théorème 10. □

► **Remarques**

- Le théorème précédent pourrait donc être réécrit sous forme de condition nécessaire et suffisante.
- Le lecteur pourra regarder le théorème de transfert comme résultant de la formule de l’espérance totale pour le système complet d’événements formé par les $[\varphi(X) = y_j]$.

Exercice 6.

Soit X une variable aléatoire définie sur l'espace probabilisé \mathbb{R} muni de la tribu \mathcal{B} des boréliens et suivant la loi de Poisson de paramètre $\lambda > 0$. Soit Y une autre variable aléatoire sur le même espace et suivant la loi géométrique de paramètre p . Posant $A_n = [Y = n]$, étudier la formule de l'espérance totale dans le cas où les deux variables sont indépendantes. L'indépendance se traduit par les égalités suivantes, où $n \geq 1$ pour la première

$$\begin{aligned} P([X = k] \cap A_n) &= P([X = k] \cap [Y = n]) = P(X = k) P(Y = n) \\ &= g_k h_n = e^{-\lambda} \frac{\lambda^k}{k!} p(1-p)^{n-1}, \\ P([X = k] \cap A_0) &= P([X = k] \cap [Y = 0]) = P(X = k) P(Y = 0) = 0 \end{aligned}$$

pour $n = 0$.

Puisque seule la probabilité $P(A_0)$ est nulle, pour $n \geq 1$ on a

$$E(X | A_n) = \frac{1}{P(A_n)} \sum_{k=0}^{+\infty} k g_k h_n = \sum_{k=0}^{+\infty} k g_k = E(X).$$

Il en résulte que

$$\sum_{n \in \mathbb{N}^*} E(X | A_n) P(A_n) = \sum_{n \in \mathbb{N}^*} E(X) P(A_n) = E(X) \sum_{n \in \mathbb{N}^*} P(A_n) = E(X).$$

► Remarque

Ici la définition de l'espace probabilisé, celles des valeurs précises de g_k et de h_n n'ont évidemment joué aucun rôle, la point essentiel restant valable pour tous les couples de variables aléatoires (X, Y) indépendantes.

1. (ESCP-EAP 2001-2002) 1. Soit (Ω, \mathcal{T}, P) un espace probabilisé, et soit X une variable aléatoire réelle prenant ses valeurs dans \mathbb{N} .

a) Montrer que, pour tout $n \in \mathbb{N}^*$,

$$\sum_{k=0}^n k P(X = k) = \sum_{k=0}^{n-1} P(X > k) - n P(X > n).$$

b) On suppose que la variable aléatoire X admet une espérance $E(X)$. Montrer que, pour

tout entier naturel n , $0 \leq n P(X > n) \leq \sum_{k=n+1}^{+\infty} k P(X = k)$.

En déduire que la série de terme général $P(X > n)$ converge, et que

$$\sum_{n=0}^{+\infty} P(X > n) = E(X).$$

c) On suppose que la série de terme général $P(X > n)$ converge. Montrer que la série de terme général $n P(X = n)$ converge, et que X admet une espérance.

d) Énoncer le théorème qui vient d'être établi, en faisant intervenir la fonction de répartition de X .

2. Montrer que, si X admet une variance, $E(X^2) = \sum_{k=0}^{+\infty} (2k + 1) P(X > k)$.

3. Un horticulteur plante n oignons de narcisse dans un jardin ($n \in \mathbb{N}^*$). Chaque oignon est susceptible de fleurir au printemps et donne une fleur avec la probabilité p ; de plus, s'il donne une fleur une année, il refleurit de manière certaine les années suivantes, mais s'il n'en donne pas, cela n'influe en rien sur ce qui est susceptible de se passer les années suivantes.

Pour $j \in \{1, 2, \dots, n\}$, on note j le nombre aléatoire d'années nécessaires au narcisse numéro j pour produire une première fleur. On suppose que X_1, X_2, \dots, X_n sont des variables aléatoires, et qu'elles sont mutuellement indépendantes.

On note X le nombre aléatoire d'années au bout desquelles le jardin sera pour la première fois, fleuri des n narcisses.

a) Exprimer X en fonction de X_1, X_2, \dots, X_n , et calculer $P(X > k)$, pour tout $k \in \mathbb{N}$.

b) En déduire que X admet une espérance et exprimer cette espérance sous la forme de la somme d'une série.

c) Étudier la variance de X .

2. (Fonction génératrice) Soit (Ω, \mathcal{T}, P) un espace probabilisé, et soit X une variable aléatoire réelle prenant ses valeurs dans \mathbb{N} . On appelle fonction génératrice de X la fonction

\mathcal{G}_X définie par $\mathcal{G}_X(t) = \sum_{k=0}^{+\infty} P(X = k) t^k$. Remarquons que si la variable aléatoire X est finie, les termes $p_k = P(X = k)$ sont nuls à partir d'un certain rang, et la fonction \mathcal{G}_X est un polynôme.

1. a) Montrer que la série de terme général $P(X = n) t^n$ est absolument convergente pour tout réel t de l'intervalle $[-1, 1]$.

b) Quel peut être l'ensemble de définition de \mathcal{G}_X ?

2. a) Exprimer $\mathcal{G}_X(t)$ comme espérance d'une variable aléatoire.

- b) Montrer que si X et Y sont deux variables aléatoires indépendantes prenant leurs valeurs dans \mathbb{N} , alors pour tout t de $[-1, 1]$,

$$\mathcal{G}_{X+Y}(t) = \mathcal{G}_X(t) \cdot \mathcal{G}_Y(t).$$

3. a) Montrer que

$$\forall t \in [0, 1[, \quad \mathcal{T}(t) = \frac{\mathcal{G}_X(t) - \mathcal{G}_X(1)}{t - 1} = \sum_{k=1}^{+\infty} p_k (1 + t + t^2 + \dots + t^{k-1}).$$

En déduire que la fonction \mathcal{T} est croissante sur $[0, 1[$.

- b) On suppose que X admet une espérance. Prouver que

$$\forall t \in [0, 1[, \quad \mathcal{T}(t) \leq E(X).$$

- c) En déduire que \mathcal{G}_X est dérivable à gauche en 1. Quelle relation existe-t-il entre $E(X)$ et la dérivée à gauche de \mathcal{G}_X en 1 ?
 d) On suppose maintenant que \mathcal{G}_X est dérivable à gauche en 1. Prouver que

$$\forall n \in \mathbb{N}^*, \quad \sum_{k=1}^n k p_k \leq (\mathcal{G}_X)'_g(1).$$

En déduire que X possède une espérance. Quelle relation entre $(\mathcal{G}_X)'_g(1)$ et $E(X)$ peut-on en déduire ?

- e) Conclure.
 f) On suppose que X admet une variance et que \mathcal{G}_X est deux fois dérivable à gauche en 1. Calculer la variance de X en fonction de $(\mathcal{G}_X)'_g(1)$ et de $(\mathcal{G}_X)''_g(1)$.
 4. Dans chacun des cas suivants, déterminer la fonction génératrice de la variable aléatoire X , et retrouver ainsi la valeur de $E(X)$ et celle de $V(X)$.
 a) X suit la loi binomiale de paramètre (n, p) .
 b) X suit la loi uniforme sur $\llbracket 1, n \rrbracket$.
 c) X suit la loi géométrique de paramètre p .
 d) X suit la loi de Poisson de paramètre λ .

3. (Loi de Pascal et loi binomiale négative) On considère un processus binomial de paramètre p , c'est à dire une suite d'épreuves indépendantes telles que chaque épreuve conduit à un succès avec la probabilité p , et à un échec avec la probabilité $q = 1 - p$.

Dans l'exercice, on pourra utiliser la formule $\sum_{n=r-1}^{+\infty} \binom{n}{r-1} q^{n-r+1} = \frac{1}{(1-q)^r}$, valable pour tout q de $[0, 1[$, et tout entier naturel r non nul.

1. Soit r un entier strictement positif. On note X_r le rang d'apparition du $r^{\text{ième}}$ succès.
 a) Que peut-on dire de X_1 ?
 b) Quel est l'ensemble V des valeurs que X_r peut prendre ?
 c) Pour tout k élément de V , calculer $P(X_r = k)$.
 2. a) Montrer que $\sum_{k=r}^{+\infty} P(X_r = k) = 1$. En déduire que X_r est une variable aléatoire. On dit que X_r suit la loi de Pascal de paramètre (r, p) .

- b) Montrer que $E(X_r) = \frac{r}{p}$.
- c) Montrer que $V(X_r) = \frac{rq}{p^2}$ (on pourra commencer par calculer l'espérance de $X_r(X_r - 1)$).
- d) Déterminer la fonction génératrice de X_r . Retrouver l'espérance et la variance de X_r .
3. On note Y_r le nombre d'échecs précédant le r -ième succès. On dit que Y_r suit une loi binomiale négative.
- a) Quelle relation y a-t-il entre X_r et Y_r ?
- b) En déduire la loi de Y_r , son espérance et sa variance.
- c) Pour tout x réel et tout entier k , on définit le coefficient du binôme généralisé par

$$\binom{x}{k} = \frac{x(x-1)\dots(x-k+1)}{k!}. \text{ Montrer que}$$

$$\binom{k+r-1}{k} = (-1)^k \binom{-r}{k}.$$

En déduire que $P(Y_r = k)$ peut s'écrire $\binom{-r}{k} p^r (-q)^k$, ce qui justifie l'appellation de loi binomiale négative.

- d) Déterminer la fonction génératrice de Y_r . Retrouver l'espérance et la variance de Y_r .

4. (EDHEC 1999) Une urne contient une boule noire et $(n - 1)$ boules blanches, n désignant un entier supérieur ou égal à 2. On vide l'urne en effectuant des tirages d'une boule de la manière suivante : le premier tirage s'effectue sans remise, le deuxième s'effectue avec remise, le troisième s'effectue sans remise, le quatrième s'effectue avec remise... D'une manière générale, les tirages d'ordre impair s'effectuent sans remise et les tirages d'ordre pair s'effectuent avec remise de la boule tirée.

1. a) Quel est le nombre total N de tirages effectués lors de cette épreuve ?
 b) Pour j élément de $\llbracket 1, n - 1 \rrbracket$, combien reste-t-il de boules avant le $(2j)$ -ème tirage ? Combien en reste-t-il avant le $(2j + 1)$ -ème tirage ?

On désigne par X_k la variable aléatoire qui vaut 1 si la boule noire est obtenue au k -ème tirage (que ce soit la première fois ou non) et 0 sinon. On désigne par X la variable aléatoire égale au nombre d'apparitions de la boule noire lors de cette épreuve.

2. a) Calculer $P(X_1 = 1)$, $P(X_2 = 1)$.
 b) Pour tout entier naturel j de $\llbracket 1, n - 1 \rrbracket$, calculer $P(X_{2j+1} = 1)$ et $P(X_{2j} = 1)$.
 c) En déduire la loi suivie par toutes les variables X_k .
3. Pour tout j élément de $\llbracket 1, n \rrbracket$, on note U_j l'événement « On obtient la boule noire pour la première fois au $(2j - 1)$ -ème tirage ».
- a) En considérant l'état de l'urne avant le $(2n - 2)$ -ème tirage, montrer que $P(U_n) = 0$.
 Montrer que $\forall j \in \llbracket 1, n \rrbracket, P(U_j) = \frac{n-j}{n(n-1)}$.
- b) Exprimer l'événement $(X = 1)$ en fonction des U_j , puis en déduire la valeur de $P(X = 1)$.
- c) Montrer que $P(X = n) = \frac{1}{n!}$.

4. Montrer que $X = \sum_{k=1}^{2n-1} X_k$, puis en déduire l'espérance de X .

5. Soit i un entier naturel compris entre 0 et $n - 2$.

a) Pour tout j de $\llbracket 1, 2n - 2j - 2 \rrbracket$, donner la valeur de

$$P(X_{2i+j+1} = 1 \mid X_{2i+1} = 1).$$

b) En déduire que $\forall j \in \llbracket 1, 2n - 2j - 2 \rrbracket$, $\text{Cov}(X_{2i+1}, X_{2i+j+1}) = \frac{-1}{n^2}$.

6. Soit i un entier naturel compris entre 1 et $n - 1$.

a) Montrer que, pour tout k de $\llbracket 1, n - i - 1 \rrbracket$, $P(X_{2i+2k} = 1 \mid X_{2i=1}) = \frac{1}{n-i}$.

b) Montrer que, pour tout k de $\llbracket 0, n - i - 1 \rrbracket$, $P(X_{2i+2k+1} = 1 \mid X_{2i=1}) = \frac{1}{n-i}$.

c) En déduire que : $\forall j \in \llbracket 1, 2n - 2i - 1 \rrbracket$, $\text{Cov}(X_{2i}, X_{2i+j}) = \frac{i}{n^2(n-i)}$.

7. Montrer que la variance de X est $V(X) = \frac{(2n+1)(n-1)}{n^2} - \frac{2}{n} \sum_{j=1}^{n-1} \frac{1}{j}$.

5. (ESCP-EAP 2000) Dans cet exercice, on admettra les propriétés suivantes :

– Soient $\sum u_n$ et $\sum v_n$ deux séries à termes positifs convergentes de sommes respectives U et V . La série de terme général $w_n = \sum_{i=0}^n u_i v_{n-i}$ (série produit) est convergente et a pour somme $W = UV$.

– Si de plus, la série de terme général $w_n t^n$ est convergente pour tout $t \in [0, 1[$, alors la fonction $t \mapsto \sum_{k=0}^{+\infty} w_n t^n$ est deux fois dérivable sur cet intervalle, les fonctions dérivées seconde et

première étant obtenues en dérivant terme à terme la somme $\sum_{k=0}^{+\infty} w_n t^n$.

On effectue une suite de lancers indépendants d'une pièce de monnaie donnant pile (P) avec la probabilité α et face (F) avec la probabilité $\beta = 1 - \alpha$. ($\alpha \in]0, 1[$).

On s'intéresse à l'apparition de deux piles consécutifs. On note S_k l'événement « deux piles consécutifs sont apparus au $(k - 1)$ -ème et au k -ème tirage ».

Chaque pile ne peut servir qu'à une seule série de deux piles consécutifs. Ainsi, dans la succession F P F P P P P, seuls sont réalisés les événements S_5 et S_7 et non S_6 .

On note G_k l'événement « deux piles consécutifs sont apparus pour la première fois au $(k - 1)$ -ème et au k -ème tirage ».

1. a) On pose $a_k = P(S_k)$ (probabilité de l'événement S_k). Soit P_k l'événement « on obtient pile au k -ème tirage ».

En écrivant que $(P_{k-1} \cap P_k) \subset (S_{k-1} \cup S_k)$, montrer que, pour tout $k \geq 2$, $\alpha^2 = a_k + \alpha a_{k-1}$.

b) En déduire la valeur de a_k pour tout $k \geq 2$.

2. On pose $b_k = P(G_k)$ (probabilité de l'événement G_k).

Montrer que pour tout $k \geq 2$, alors $a_k = \sum_{i=1}^{k-1} a_i b_{k-i} + b_k$.

3. a) Montrer que la série de terme général $a_k t^k$ est convergente pour tout t de $[0, 1[$ et calculer sa somme $F(t)$.

b) Montrer de même que la série de terme général $b_k t^k$ est convergente et exprimer sa somme H en fonction de F .

c) Expliciter $H(t)$ pour $t \in [0, 1[$.

4. Montrer que H admet une limite lorsque t tend vers 1^- .

5. a) En admettant que $H(1) = \lim_{t \rightarrow 1^-} H(t)$, que $H'(1) = \lim_{t \rightarrow 1^-} H'(t)$ et que $H''(1) = \lim_{t \rightarrow 1^-} H''(t)$, montrer que la suite $(b_k)_{k \geq 1}$ permet de définir la loi d'une variable aléatoire X à valeurs dans \mathbb{N}^* .

b) Montrer que X admet une espérance et une variance et donner les valeurs de ces moments.

6. On pose $a_{n,p} = \frac{1}{n^2 - p^2}$ si $n \neq p$ et $a_{n,n} = 0$.

Calculer $\sum_{n=0}^{+\infty} \sum_{p=0}^{+\infty} a_{n,p}$ et $\sum_{p=0}^{+\infty} \sum_{n=0}^{+\infty} a_{n,p}$.

Que peut-on en déduire ?

7. Mêmes questions que dans l'exercice précédent avec la suite double $(a_{n,p})_{(n,p) \in (\mathbb{N}^*)^2}$ définie par

$$a_{n,p} = \frac{1}{n+1} \left(\frac{n}{n+1} \right)^m - \frac{1}{n+2} \left(\frac{n+1}{n+2} \right)^m.$$

8. Soit $(a_{n,p})_{(n,p) \in \mathbb{N}^2}$ la suite double définie par

$\forall n \in \mathbb{N}, a_{n,n} = 1, a_{2n,2n+1} = a_{2n+1,2n} = -1$ et $a_{n,p} = 0$ dans les autres cas.

Montrer que $\sum_{n=0}^{+\infty} \sum_{p=0}^{+\infty} a_{n,p} = \sum_{p=0}^{+\infty} \sum_{n=0}^{+\infty} a_{n,p}$, bien que la série double $\sum a_{n,p}$ ne soit pas absolument convergente.

9. Montrer que la série double $\sum a_{i,j}$ définie par

$$\forall (i,j) \in \mathbb{N}^2, a_{i,j} = \frac{(i+j)}{i!j!2^{i+j}}$$

est convergente de somme e .

10. Soit x un réel tel que $|x| < \frac{1}{2}$. Montrer que la série double de terme général $a_{i,j} = \frac{(i+j)!}{i!j!} x^{i+j}$ est absolument convergente. Calculer sa somme.

11. Calculer $\sum_{p=1}^{+\infty} \sum_{q=p}^{+\infty} \frac{(-1)^p}{q^3}$ en fonction de $\sum_{n=1}^{+\infty} \frac{1}{n^3}$.

12. Montrer que pour $x \in \mathbb{R}$, $|x| < 1$, on a l'égalité

$$\sum_{n=0}^{+\infty} \frac{x^{2n+1}}{1-x^{2n+1}} = \sum_{n=1}^{+\infty} \frac{x^n}{1-x^{2n}}.$$

On utilisera l'égalité $\frac{1}{1-\gamma} = \sum_{k=0}^{+\infty} \gamma^k$ pour tout $\gamma \in]-1, 1[$ et une série double.

13. Soient a, b, c trois entiers naturels non nuls et x un nombre réel vérifiant $|x| < 1$. En utilisant des séries doubles convenables, montrer que

$$\begin{aligned} \sum_{n=0}^{+\infty} \frac{x^{bn}}{1-x^{an+c}} &= \sum_{n=0}^{+\infty} \frac{x^{cn}}{1-x^{an+b}}, \\ \sum_{n=0}^{+\infty} \frac{x^{bn}}{1+x^{an+c}} &= \sum_{n=0}^{+\infty} \frac{(-1)^n x^{cn}}{1-x^{an+b}}. \end{aligned}$$

14. 1. Montrer que, pour tout entier naturel non nul m , on a

$$\sum_{n=1}^{+\infty} \frac{\sqrt{m}}{(m+n)\sqrt{n}} \leq \int_0^{+\infty} \frac{1}{(x+1)\sqrt{x}} dx.$$

2. Soient $(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ deux suites de réels positifs telles que les séries $\sum a_n^2$ et $\sum b_n^2$ convergent.

a) Montrer que pour tout $n \in \mathbb{N}^*$,

$$\sum_{1 \leq i, j \leq n} \frac{a_i b_j}{i+j} \leq \left(\sum_{1 \leq i, j \leq n} \frac{\sqrt{i} a_i^2}{\sqrt{j}(i+j)} \right)^{\frac{1}{2}} \left(\sum_{1 \leq i, j \leq n} \frac{\sqrt{j} b_j^2}{\sqrt{i}(i+j)} \right)^{\frac{1}{2}}.$$

b) En déduire que la série double $\left(\frac{a_i b_j}{i+j} \right)_{(i,j) \in (\mathbb{N})^2}$ est absolument convergente.

Démontrer l'inégalité

$$\sum_{(i,j) \in (\mathbb{N})^2} \frac{a_i b_j}{i+j} \leq \pi \left(\sum_{n=0}^{+\infty} a_n^2 \right)^{\frac{1}{2}} \left(\sum_{n=0}^{+\infty} b_n^2 \right)^{\frac{1}{2}}.$$

- 15.** 1. Soit $\sum a_{i,j}$ une série double absolument convergente de somme S . En utilisant le théorème de sommation par paquets, montrer que

$$S = \sum_{n=0}^{+\infty} \sum_{(i,j) \in I_n} a_{i,j},$$

où, pour tout entier naturel n , $I_n = \{(i, j) \in \mathbb{N}^2 \mid i + j = n\}$.

2. Soit x et y deux réels. Montrer que la série double de terme général

$$a_{i,j} = \frac{x^i y^j}{(i+j)!}$$

est absolument convergente. Calculer sa somme.

- 16.** 1. En utilisant le théorème de sommation par paquets, montrer que pour $x \in \mathbb{R}$, $|x| < 1$, on a l'égalité

$$\sum_{n=1}^{+\infty} \frac{x^n}{1-x^n} = \sum_{n=1}^{+\infty} d(n)x^n,$$

où $d(n)$ est le nombre de diviseurs positifs de n .

2. Soit un entier $p \geq 2$. En considérant la série double $\sum \frac{1}{i^p j^p}$, montrer que

$$\left(\sum_{i=1}^{+\infty} \frac{1}{i^p} \right)^2 = \sum_{n=1}^{+\infty} \frac{d(n)}{n^p}.$$

- 17.** Soit α un réel positif. On considère la série double $(a_{n,p})_{(n,p) \in (\mathbb{N}^*)^2}$ définie par $a_{n,p} = \frac{1}{(n^2 + p^2)^\alpha}$. On pose, pour tout $k \in \mathbb{N}^*$, $u_k = \sum_{\substack{1 \leq n \leq k \\ 1 \leq p \leq k}} a_{n,p}$.

1. Montrer que, pour tout entier $k \in \mathbb{N}^*$, on a

$$\frac{2k+1}{2^\alpha(k+1)^{2\alpha}} \leq u_{k+1} - u_k \leq \frac{2k+1}{(k+1)^{2\alpha}}.$$

2. On suppose que $\alpha \leq \frac{1}{2}$. Montrer que la série de terme général $u_{k+1} - u_k$ diverge. En déduire que la suite $(u_k)_{k \in \mathbb{N}^*}$ a pour limite $+\infty$. Montrer que la série double $\sum a_{n,p}$ diverge.
3. On suppose que $\alpha > \frac{1}{2}$. Montrer que la suite $(u_k)_{k \in \mathbb{N}^*}$ converge. En déduire que la série double $\sum a_{n,p}$ converge.

18. 1. Soit la suite (a_n) définie pour $n > 0$ par $a_n = \frac{1}{n(n+1)}$. Déterminer la nature et la somme éventuelle de chacune des séries $\sum a_n$ et $\sum na_n$.
2. Soit un espace probabilisé dénombrable $(\Omega, \mathcal{P}(\Omega), P)$. Montrer qu'il existe sur cet espace une variable aléatoire X , d'image \mathbb{Z}^* , vérifiant

$$P(X = n) = \frac{1}{2} a_{|n|}.$$

3. Que peut-on dire de la famille d'événements (A_n) avec $A_n = [X = n] \cap [X = -n]$ pour $n \in \mathbb{N}^*$?
4. Montrer que le symbole $E(X | A_n)$ a un sens pour tout $n \in \mathbb{N}^*$ et calculer sa valeur.
5. Montrer que la série de terme général $E(X | A_n)P(A_n)$ est absolument convergente et calculer sa somme.
6. La variable X admet-elle une espérance ?

Autres exercices

On cherchera également avec profit les problèmes et exercices de probabilité proposés aux concours suivants dans diverses séries :

EDHEC 1995 (Problème d'ascenseur), ESLSCA 1995 (Jeu à trois joueurs), ESLSCA 1995 (Rangement dans trois cases), EDHEC 2002 (Gestion de stock), ESCP-EAP 2000, ESCP-EAP 2001, ESCP-EAP 2002, EDHEC 1997 (Temps d'attente), HEC 2004 (Paradoxe de Walter Penney), EDHEC 1994 (Marche aléatoire avec barrière absorbante), Ecricome 1992 (Chaîne de Markov), Ecricome 1995 (La ruine d'un joueur), ESSEC 1991 (Le problème du collectionneur).

Vecteurs aléatoires discrets

9

Le présent chapitre est largement consacré à des rappels de définitions et propriétés rencontrées et démontrées dans le premier tome du présent ouvrage. On y présentera aussi quelques compléments et généralisations.

1. Couples de variables aléatoires réelles discrètes

Voir en particulier le chapitre 31 du livre de première année.

1.1 Généralités

Définition 1

Soit (Ω, \mathcal{T}) un espace probabilisable.

Tout couple $X = (X_1, X_2)$ de variables aléatoires réelles tel que X_1 et X_2 soient des variables aléatoires réelles discrètes est appelé **un couple de variables aléatoires réelles discrètes**, ou **couple discret** par abréviation.

► Remarque

Rappelons encore une fois que la notation \mathcal{T} désigne une tribu, encore souvent appelée « σ -algèbre » (et alors souvent notée \mathcal{A}). Nous utiliserons indistinctement ces deux notations.

Exemple

Considérons une urne contenant 10 boules noires numérotées de 1 à 10, et une boule blanche. On effectue des tirages successifs avec remise ; on note X_1 le rang de la première boule blanche tirée, et X_2 le nombre de boules de numéro pair tirées avant l'obtention de cette première boule blanche. On définit ainsi un vecteur aléatoire $X = (X_1, X_2)$, avec $X_1(\Omega) = \mathbb{N}^*$ et $X_2(\Omega) = \mathbb{N}$. Alors l'ensemble $X(\Omega)$ est égal à l'ensemble des couples (i, j) de $\mathbb{N}^* \times \mathbb{N}$ tels que $j < i$.

Proposition 1

Pour tout couple $X = (X_1, X_2)$ de variables aléatoires réelles discrètes, la famille d'événements

$$([X_1 = x_i] \cap [X_2 = y_j])_{(x_i, y_j) \in X_1(\Omega) \times X_2(\Omega)},$$

encore notée

$$([X_1 = x_i, X_2 = y_j])_{(x_i, y_j) \in X_1(\Omega) \times X_2(\Omega)}$$

$$(X_1 = x_i, X_2 = y_j)_{(x_i, y_j) \in X_1(\Omega) \times X_2(\Omega)}$$

$$([X = (x_i, y_j)])_{(x_i, y_j) \in X_1(\Omega) \times X_2(\Omega)},$$

est un système complet d'événements de (Ω, \mathcal{T}) .

Preuve

Cette démonstration a déjà été faite dans l'ouvrage de première année. Nous la reprenons pour mémoire.

Remarquons que $[X_1 = x_i, X_2 = y_j] = \{\omega \in \Omega \mid X_1(\omega) = x_i \text{ et } X_2(\omega) = y_j\}$, soit encore $[X_1 = x_i] \cap [X_2 = y_j]$. Comme X_1 et X_2 sont des variables aléatoires réelles, $[X_1 = x_i]$ et $[X_2 = y_j]$ sont des éléments de \mathcal{T} , et $[X_1 = x_i, X_2 = y_j]$ est un élément de \mathcal{T} .

Soient (x_i, y_j) et (x'_i, y'_j) des éléments distincts de $X_1(\Omega) \times X_2(\Omega)$. Si $x_i \neq x'_i$ ou $y_j \neq y'_j$, alors $[X_1 = x_i] \cap [X_1 = x'_i] = \emptyset$ ou $[X_2 = y_j] \cap [X_2 = y'_j] = \emptyset$. On en déduit que $[X_1 = x_i, X_2 = y_j] \cap [X_1 = x'_i, X_2 = y'_j] = \emptyset$. Comme

$$\bigcup_{(x_i, y_j) \in X_1(\Omega) \times X_2(\Omega)} [X_1 = x_i, X_2 = y_j] = \left[\bigcup_{x_i \in X_1(\Omega)} [X_1 = x_i] \right] \cap \left[\bigcup_{y_j \in X_2(\Omega)} [X_2 = y_j] \right] = \Omega$$

on en déduit que $([X_1 = x_i, X_2 = y_j])_{(x_i, y_j) \in X_1(\Omega) \times X_2(\Omega)}$ est un système complet d'événements de (Ω, \mathcal{T}) . \square

Définition 2

Pour tout couple $X = (X_1, X_2)$ de variables aléatoires réelles discrètes, le système complet d'événements $([X_1 = x_i, X_2 = y_j])_{(x_i, y_j) \in X_1(\Omega) \times X_2(\Omega)}$ est appelé le **système complet d'événements associé au vecteur aléatoire** $X = (X_1, X_2)$.

La tribu \mathcal{A}_X engendrée par le système complet d'événements associé à X est appelée la **tribu associée à** X .

Proposition 2

Pour tout couple $X = (X_1, X_2)$ de variables aléatoires réelles discrètes, la tribu associée à X_1 et la tribu associée à X_2 sont incluses dans la tribu associée à $X = (X_1, X_2)$:

$$\mathcal{A}_{X_1} \subset \mathcal{A}_{(X_1, X_2)}, \quad \mathcal{A}_{X_2} \subset \mathcal{A}_{(X_1, X_2)}.$$

Preuve

Soit A un élément de \mathcal{A}_{X_1} . Il existe alors une partie I de $X_1(\Omega)$ telle que $A = [X_1 \in I]$

$$A = \bigcup_{x_i \in I, y_j \in X_2(\Omega)} [X_1 = x_i, X_2 = y_j].$$

Donc A est un élément de $\mathcal{A}_{(X_1, X_2)}$, et $\mathcal{A}_{X_1} \subset \mathcal{A}_{(X_1, X_2)}$. \square

► **Remarque**

La réunion $\mathcal{A}_{X_1} \cup \mathcal{A}_{X_2}$ est donc incluse dans $\mathcal{A}_{(X_1, X_2)}$, mais ne lui est pas égale dans le cas général, car une réunion de tribus n'est en général pas une tribu.

1.2 Loi conjointe et lois marginales

Définition 3

Pour tout couple $X = (X_1, X_2)$ de variables aléatoires réelles discrètes sur l'espace probabilisé (Ω, \mathcal{T}, P) , l'application

$$\widehat{P}_X : \begin{cases} X_1(\Omega) \times X_2(\Omega) & \longrightarrow \mathbb{R} \\ (x_i, y_j) & \longmapsto P(X_1 = x_i, X_2 = y_j) \end{cases}$$

est appelée la **loi du couple** $X = (X_1, X_2)$ ou la **loi conjointe des variables aléatoires** X_1, X_2 .

Exemple

Reprenons l'exemple ci-dessus, où l'on tire successivement et avec remise une boule d'une urne contenant 10 boules noires numérotées de 1 à 10, et une boule blanche, et où l'on note X_1 le rang de la première boule blanche tirée, et X_2 le nombre de boules de numéro pair tirées avant l'obtention de cette première boule blanche.

Alors $\widehat{P}_X(i, j) = P(X_1 = i, X_2 = j) = P(X_1 = i)P_{|X_1=i}(X_2 = j)$. On a clairement $\widehat{P}_X(0, j) = 0$ et plus généralement, si $j \geq i$, $\widehat{P}_X(i, j) = 0$.

Dans le cas où $0 < j < i$, on obtient

$$\widehat{P}_X(i, j) = \left(\frac{10}{11}\right)^{i-1} \times \frac{1}{11} \times \binom{i-1}{j} \times \left(\frac{1}{2}\right)^{i-1} = \frac{1}{11} \left(\frac{5}{11}\right)^{i-1} \binom{i-1}{j}.$$

► **Remarque**

Il est clair que, pour tout couple discret $X = (X_1, X_2)$ de variables aléatoires réelles sur l'espace probabilisé (Ω, \mathcal{T}, P) de loi \widehat{P}_X , alors

$$\begin{aligned} 1 &= \sum_{(x_i, y_j) \in X(\Omega)} \widehat{P}_X(x_i, y_j) = \sum_{x_i \in X_1(\Omega), y_j \in X_2(\Omega)} P(X_1 = x_i, X_2 = y_j) \\ &= \sum_{x_i \in X_1(\Omega)} \sum_{y_j \in X_2(\Omega)} P(X_1 = x_i, X_2 = y_j). \end{aligned}$$

Définition 4

Soit $X = (X_1, X_2)$ un couple de variables aléatoires réelles discrètes sur l'espace probabilisé (Ω, \mathcal{T}, P) . La fonction de répartition de X est la fonction de deux variables F_X définie par $F_X(x, y) = P(X_1 \leq x, X_2 \leq y)$.

► **Remarques**

- Une telle fonction de répartition est « constante par rectangles », ce qui généralise les fonctions constantes par intervalles. Ses valeurs sont comprises entre 0 et 1. Ses applications partielles sont constantes par intervalles et croissantes.
- Dans le cas des couples discrets, les fonctions de répartition sont assez peu utilisées. Mais leur définition s'étend aux couples de variables aléatoires réelles à densité, où leur utilisation est importante.

Exemple

Soient X_1 et X_2 des variables de Bernoulli indépendantes de paramètres respectifs p_1 et p_2 . La fonction de répartition du couple $X = (X_1, X_2)$ est donnée par

$$\begin{cases} F_X(x, y) = 0 & \text{si } x < 0 & \text{ou } y < 0 \\ F_X(x, y) = (1 - p_1)(1 - p_2) & \text{si } 0 \leq x < 1 & \text{et } 0 \leq y < 1 \\ F_X(x, y) = 1 - p_1 & \text{si } 0 \leq x < 1 & \text{et } 1 \leq y \\ F_X(x, y) = 1 - p_2 & \text{si } 1 \leq x & \text{et } 0 \leq y < 1 \\ F_X(x, y) = 1 & \text{si } 1 \leq x & \text{et } 1 \leq y. \end{cases}$$

Définition 5

Pour tout couple $X = (X_1, X_2)$ de variables aléatoires réelles discrètes sur l'espace probabilisé (Ω, \mathcal{T}, P) , l'application

$$\widehat{P}_{X_1} : \begin{cases} X_1(\Omega) & \longrightarrow \mathbb{R} \\ x_i & \longmapsto P(X_1 = x_i) \end{cases}$$

est appelée la **loi marginale de X_1** .

L'application

$$\widehat{P}_{X_2} : \begin{cases} X_2(\Omega) & \longrightarrow \mathbb{R} \\ y_j & \longmapsto P(X_2 = y_j) \end{cases}$$

est appelée la **loi marginale de X_2** .

► Remarques

- On reconnaît ici les lois respectives des variables X_1 et X_2 .
- On peut naturellement vérifier dans chaque cas particulier que

$$1 = \sum_{x_i \in X_1(\Omega)} P(X_1 = x_i) = \sum_{y_j \in X_2(\Omega)} P(X_2 = y_j).$$

- Il est clair que pour tout couple $X = (X_1, X_2)$ de variables aléatoires réelles discrètes, la loi marginale de X_1 vérifie l'égalité

$$\forall x_i \in X_1(\Omega), \quad P(X_1 = x_i) = \sum_{y_j \in X_2(\Omega)} P(X_1 = x_i, X_2 = y_j).$$

$$\text{On note alors } p_{i.} = P(X_1 = x_i) = \sum_{y_j \in X_2(\Omega)} P(X_1 = x_i, X_2 = y_j).$$

La loi marginale de X_2 vérifie l'égalité

$$\forall y_j \in X_2(\Omega), \quad P(X_2 = y_j) = \sum_{x_i \in X_1(\Omega)} P(X_1 = x_i, X_2 = y_j).$$

$$\text{On note de même } p_{.j} = P(X_2 = y_j) = \sum_{x_i \in X_1(\Omega)} P(X_1 = x_i, X_2 = y_j).$$

Lorsque c'est possible, on présente une loi conjointe et les lois marginales correspondantes dans un tableau, analogue à ceux qui figurent ci-dessous, où $p_{i,j} = P(X_1 = x_i, X_2 = y_j)$.

$X_1 \backslash X_2$	y_1	y_2	\cdots	y_j	\cdots	y_n		$X_1 \backslash X_2$	y_1	y_2	\cdots	y_j	\cdots	
x_1	$p_{1,1}$	$p_{1,2}$	\cdots	$p_{1,j}$	\cdots	$p_{1,n}$	$p_{1,\cdot}$	x_1	$p_{1,1}$	$p_{1,2}$	\cdots	$p_{1,j}$	\cdots	$p_{1,\cdot}$
x_2	$p_{2,1}$	$p_{2,2}$	\cdots	$p_{2,j}$	\cdots	$p_{2,n}$	$p_{2,\cdot}$	x_2	$p_{2,1}$	$p_{2,2}$	\cdots	$p_{2,j}$	\cdots	$p_{2,\cdot}$
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots	\vdots	\vdots	\vdots		\vdots		\vdots
x_i	$p_{i,1}$	$p_{i,2}$	\cdots	$p_{i,j}$	\cdots	$p_{i,n}$	$p_{i,\cdot}$	x_i	$p_{i,1}$	$p_{i,2}$	\cdots	$p_{i,j}$	\cdots	$p_{i,\cdot}$
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots	\vdots	\vdots	\vdots		\vdots		\vdots
x_n	$p_{n,1}$	$p_{n,2}$	\cdots	$p_{n,j}$	\cdots	$p_{n,n}$	$p_{n,\cdot}$	\vdots	\vdots	\vdots		\vdots		\vdots
	$p_{\cdot,1}$	$p_{\cdot,2}$	\cdots	$p_{\cdot,j}$	\cdots	$p_{\cdot,n}$	1		$p_{\cdot,1}$	$p_{\cdot,2}$	\cdots	$p_{\cdot,j}$	\cdots	1

Définition 6

Pour tout couple $X = (X_1, X_2)$ de variables aléatoires réelles discrètes sur l'espace probabilisé (Ω, \mathcal{T}, P) admettant une espérance, on définit le **vecteur espérance** $E(X)$ du vecteur aléatoire X par l'égalité $E(X) = (E(X_1), E(X_2))$.

Exemple

Soit n un entier strictement positif. Une urne contient n boules noires et $2n$ boules blanches. on tire successivement deux boules dans cette urne, soit avec remise, ce qui constituera un premier mode de tirage, soit sans remise, ce sera alors notre deuxième mode de tirage.

Notons X_1 la variable aléatoire égale à 0 si la première boule tirée est noire et à 1 si elle est blanche. Notons également X_2 la variable aléatoire égale à 0 si la deuxième boule tirée est noire et à 1 si elle est blanche.

Tout calcul fait, la loi du vecteur aléatoire $X = (X_1, X_2)$ est donnée dans les tableaux suivants

Tirage avec remise

Tirage sans remise

$X_1 \backslash X_2$	0	1	loi de X_1
0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$
1	$\frac{2}{9}$	$\frac{4}{9}$	$\frac{2}{3}$
loi de X_2	$\frac{1}{3}$	$\frac{2}{3}$	

$X_1 \backslash X_2$	0	1	loi de X_1
0	$\frac{n-1}{9n-3}$	$\frac{2n}{9n-3}$	$\frac{1}{3}$
1	$\frac{2n}{9n-3}$	$\frac{4n-2}{9n-3}$	$\frac{2}{3}$
loi de X_2	$\frac{1}{3}$	$\frac{2}{3}$	

Remarquons que dans les deux cas, les lois marginales sont les mêmes, c'est-à-dire que X_1 et X_2 ont les mêmes lois dans les deux modes de tirages. Le vecteur moyen

est également le même : c'est le vecteur $\bar{X} = \left(\frac{2}{3}, \frac{2}{3}\right)$. Cependant les lois conjointes ne sont pas identiques.

Cet exemple est un exemple de plus du fait que l'on ne peut pas déduire la loi conjointe des lois marginales.

1.3 Lois conditionnelles

Définition 7

Pour tout couple $X = (X_1, X_2)$ de variables aléatoires réelles discrètes et tout y_j de $X_2(\Omega)$ tel que $P(X_2 = y_j) \neq 0$, l'application

$$\begin{cases} X_1(\Omega) & \longrightarrow \mathbb{R} \\ x_i & \longmapsto \frac{P(X_1 = x_i, X_2 = y_j)}{P(X_2 = y_j)} = P_{[X_2=y_j]}(X_1 = x_i) = P(X_1 = x_i | X_2 = y_j) \end{cases}$$

est appelée la **loi conditionnelle sachant $[X_2 = j]$ de X_1** .

Pour tout i de $X_1(\Omega)$ tel que $P(X_1 = x_i) \neq 0$, l'application

$$\begin{cases} X_2(\Omega) & \longrightarrow \mathbb{R} \\ y_j & \longmapsto \frac{P(X_1 = x_i, X_2 = y_j)}{P(X_1 = x_i)} = P_{[X_1=x_i]}(X_2 = y_j) = P(X_2 = y_j | X_1 = x_i) \end{cases}$$

est appelée la **loi conditionnelle sachant $[X_1 = x_i]$ de X_2** .

► Remarque

On sait que $Q = P_{[X_2=y_j]}$ est une probabilité, et $([X_1 = x_i])_{x_i \in X_1(\Omega)}$ un système complet d'événements, d'où

$$1 = \sum_{x_i \in X_1(\Omega)} P_{[X_2=y_j]}(X_1 = x_i) = \sum_{x_i \in X_1(\Omega)} P(X_1 = x_i | X_2 = y_j).$$

De même $P_{[X_1=x_i]}$ est une probabilité, et $([X_2 = y_j])_{y_j \in X_2(\Omega)}$ un système complet d'événements, d'où

$$1 = \sum_{y_j \in X_2(\Omega)} P_{[X_1=x_i]}(X_2 = y_j) = \sum_{y_j \in X_2(\Omega)} P(X_2 = y_j | X_1 = x_i).$$

Si la loi conjointe est présentée dans un tableau, on obtient par exemple la loi conditionnelle de X_2 sachant $[X_1 = x_i]$ sur la ligne (ou colonne) correspondant à $X_1 = x_i$, en divisant par la valeur marginale correspondante ($P(X_1 = x_i) = p_{i,\cdot}$), comme indiqué ci-dessous

		y_1	y_2	\cdots	y_j	\cdots	
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	
$P_{[X_1=x_i]}(X_2 = y_j)$	$\frac{p_{i,1}}{p_{i,\cdot}}$	$\frac{p_{i,2}}{p_{i,\cdot}}$	\dots	$\frac{p_{i,j}}{p_{i,\cdot}}$	\dots	1	
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	

Exemple

Cet exemple a été traité dans l’ouvrage de première année, comme exercice corrigé. Nous le reprenons, compte tenu de son importance.

Soient X et Y deux variables aléatoires réelles discrètes telles que Y suive la loi de Poisson de paramètre λ , où λ est un réel strictement positif, et que la loi conditionnelle de X sachant $[Y = m]$ soit la loi binomiale de paramètre (m, p) pour tout entier m , où p est un réel de $]0, 1[$. Ces données permettent de déterminer la loi de (X, Y) et la loi de X .

Pour tout entier m , par définition de la loi conditionnelle et de la loi binomiale de paramètre (m, p) . Par définition

$$\forall k \in \llbracket 0, m \rrbracket, \quad P_{[Y=m]}(X = k) = \frac{P(X = k, Y = m)}{P(Y = m)}.$$

Donc, pour tout entier m et pour tout entier k de $\llbracket 0, m \rrbracket$

$$P(X = k, Y = m) = P_{[Y=m]}(X = k)P(Y = m) = \binom{m}{k} p^k (1 - p)^{m-k} \frac{\lambda^m}{m!} e^{-\lambda}.$$

On obtient donc ainsi la loi conjointe

$$P(X = k, Y = m) = \binom{m}{k} p^k (1 - p)^{m-k} \frac{\lambda^m}{m!} e^{-\lambda} \text{ si } k \leq m \text{ et } P(X = k, Y = m) = 0 \text{ si } k > m.$$

Comme $([Y = m])_{m \in \mathbb{N}}$ est un système complet d’événements, pour tout entier k , on a

$$P(X = k) = \sum_{m=0}^{+\infty} P(X = k, Y = m).$$

Par suite

$$\begin{aligned} P(X = k) &= \sum_{m=k}^{+\infty} p^k (1 - p)^{m-k} \frac{\lambda^m}{k!(m-k)!} e^{-\lambda} \\ &= (p\lambda)^k \frac{1}{k!} e^{-\lambda} \sum_{m=k}^{+\infty} \frac{((1-p)\lambda)^{m-k}}{(m-k)!} \\ &= (p\lambda)^k \frac{1}{k!} e^{-\lambda} e^{(1-p)\lambda} \\ &= \frac{(p\lambda)^k}{k!} e^{-p\lambda}. \end{aligned}$$

On en déduit que X suit la loi de Poisson de paramètre $p\lambda$.

Espérance conditionnelle

L’espérance de la loi conditionnelle de X_2 sachant que $X_1 = x_i$ est donnée par l’égalité

$$E(X_2 | X_1 = x_i) = \sum_j y_j P(X_2 = y_j | X_1 = x_i) = \sum_j y_j P_{[X_1=x_i]}(X_2 = y_j)$$

que l’on note encore parfois $E_{[X_1=x_i]}(X_2)$.

On constate que cette espérance dépend des valeurs prises par X_1 puisque $E(X_2|X_1 = x_i)$ est fonction de x_i et peut s'écrire sous la forme $\psi(x_i)$.

Définition 8

Pour tout vecteur aléatoire discret $X = (X_1, X_2)$ on appelle **variable aléatoire espérance conditionnelle** de X_2 sachant X_1 , et l'on note $E(X_2|X_1)$ ou encore $E_{X_1}(X_2)$, la variable aléatoire réelle discrète qui prend les valeurs $E(X_2|X_1 = x_i)$ avec les probabilités $P(X_1 = x_i)$.

► Remarques

- Plus précisément $E_{X_1}(X_2)$ associe $E(X_2|X_1 = X_1(\omega))$ à un $\omega \in \Omega$.
- On a clairement, avec les notations ci-dessus, $E_{X_1}(X_2) = E(X_2|X_1) = \psi(X_1)$.
- Dans ce qui précède, on suppose que toutes les espérances écrites existent.
- Ces définitions s'étendent à des variables aléatoires réelles à densité.

Théorème de l'espérance totale

Théorème 1

Soit $X = (X_1, X_2)$ un vecteur aléatoire réel discret. Si toutes les espérances écrites existent, on a

$$E(E(X_2|X_1)) = E(X_2).$$

Preuve

On reconnaît le théorème de l'espérance totale (voir chapitre 8).

En supposant que toutes les espérances écrites existent, que toutes les séries pouvant intervenir convergent absolument, et en utilisant le théorème de sommation par paquets, la démonstration du résultat annoncé peut prendre la forme suivante

$$\begin{aligned} E(E(X_2|X_1)) &= \sum_i E(X_2|X_1 = x_i)P(X_1 = x_i) \\ &= \sum_i \left(\sum_j y_j P(X_2 = y_j|X_1 = x_i) \right) P(X_1 = x_i) \\ &= \sum_j y_j \sum_i P(X_2 = y_j|X_1 = x_i)P(X_1 = x_i) \\ &= \sum_j y_j P(X_2 = y_j) = E(X_2). \end{aligned}$$

□

Exemple

Reprenons l'exemple du 1.1, où l'on considère une urne contenant 10 boules noires numérotées de 1 à 10, et une boule blanche. On effectue des tirages successifs avec remise ; on note X_1 le rang de la première boule blanche tirée, et X_2 le nombre de boules de numéro pair tirées avant l'obtention de cette première boule blanche.

L'énoncé signifie que X_1 suit la loi géométrique de paramètre $\frac{1}{11}$, et que la loi conditionnelle de X_2 sachant $X_1 = i$ est la loi binomiale de paramètre $\left(i - 1, \frac{1}{2}\right)$.

Donc, $E(X_2|X_1 = i) = \frac{i-1}{2}$ et $P(X_1 = i) = \left(\frac{10}{11}\right)^{i-1} \frac{1}{11}$.

On en déduit

$$E(X_2) = \sum_{i=1}^{+\infty} \frac{i-1}{2} \left(\frac{10}{11}\right)^{i-1} \frac{1}{11} = \frac{1}{22} \sum_{k=0}^{+\infty} k \left(\frac{10}{11}\right)^k = \frac{1}{22} \times \frac{10}{11} \times (11)^2 = 5.$$

On a ainsi pu calculer l'espérance de X_2 sans en déterminer explicitement la loi.

1.4 Indépendance, généralités

Rappelons pour mémoire des définitions et résultats déjà rencontrés (et parfois démontrés) dans l'ouvrage de première année.

Définition 9

Pour tout couple (X_1, X_2) de variables aléatoires réelles discrètes sur l'espace probabilisé (Ω, \mathcal{T}, P) , les variables aléatoires réelles discrètes X_1, X_2 sont dites **indépendantes** si

$$\forall (x_i, y_j) \in X_1(\Omega) \times X_2(\Omega), \quad P(X_1 = x_i, X_2 = y_j) = P(X_1 = x_i) P(X_2 = y_j).$$

► Remarques

- Les variables aléatoires réelles discrètes X_1 et X_2 sont indépendantes si, et seulement si, la loi conditionnelle sachant $[X_2 = y_j]$ de X_1 est égale à la loi de X_1 pour tout y_j de $X_2(\Omega)$ et la loi conditionnelle sachant $[X_1 = x_i]$ de X_2 est égale à la loi de X_2 pour tout x_i de $X_1(\Omega)$.
- Dans le cas de variables aléatoires réelles discrètes indépendantes, la donnée des lois marginales permet de connaître la loi conjointe.

Proposition 3

Pour tout couple (X_1, X_2) de variables aléatoires réelles discrètes sur l'espace probabilisé (Ω, \mathcal{T}, P) , les variables aléatoires X_1 et X_2 sont indépendantes si, et seulement si, pour tout couple (x_i, y_j) de $X_1(\Omega) \times X_2(\Omega)$, les événements $[X_1 = x_i]$ et $[X_2 = y_j]$ sont indépendants, c'est-à-dire si, et seulement si, pour toute partie A de $X_1(\Omega)$ et pour toute partie B de $X_2(\Omega)$, les événements $[X_1 \in A]$ et $[X_2 \in B]$ sont indépendants.

Proposition 4

Tout couple discret (X_1, X_2) de variables aléatoires réelles est formé de variables aléatoires indépendantes si, et seulement si, les tribus associées \mathcal{A}_{X_1} et \mathcal{A}_{X_2} sont indépendantes.

Proposition 5

Pour tout couple d'événements (A, B) de l'espace probabilisé (Ω, \mathcal{T}, P) , les événements A et B sont indépendants si, et seulement si, les variables aléatoires indicatrices de A et de B sont indépendantes.

Proposition 6

Pour tout couple $X = (X_1, X_2)$ de variables aléatoires réelles discrètes sur l'espace probabilisé (Ω, \mathcal{T}, P) , toute fonction f de $X_1(\Omega)$ à valeurs dans \mathbb{R} et toute fonction g de $X_2(\Omega)$ à valeurs dans \mathbb{R} , si les variables aléatoires réelles X_1 et X_2 sont indépendantes, alors $f(X_1)$ et $g(X_2)$ sont des variables aléatoires réelles discrètes indépendantes.

► Remarque

Attention : la réciproque peut être fautive si on ne l'exprime pas correctement.

2. Variable aléatoire fonction d'un vecteur discret**2.1 Généralités**

Ici encore, nous nous bornons à rappeler des résultats démontrés et abondamment illustrés dans le premier tome de cet ouvrage.

Théorème 2

Pour tout couple (X_1, X_2) de variables aléatoires réelles discrètes et pour toute fonction g définie sur $X_1(\Omega) \times X_2(\Omega)$ à valeurs dans \mathbb{R} , l'application

$$Z : \begin{cases} \Omega & \longrightarrow \mathbb{R} \\ \omega & \longmapsto g(X_1(\omega), X_2(\omega)) \end{cases}$$

est une variable aléatoire réelle discrète.

Proposition 7

Pour tout couple (X_1, X_2) de variables aléatoires réelles discrètes de l'espace probabilisable (Ω, \mathcal{T}) et pour toute fonction g définie sur $X_1(\Omega) \times X_2(\Omega)$, la tribu associée à la variable aléatoire réelle discrète $Z = g(X_1, X_2)$ est incluse dans la tribu associée au couple (X_1, X_2) .

$$\mathcal{A}_Z \subset \mathcal{A}_{(X_1, X_2)}.$$

Théorème 3

Pour tout couple (X_1, X_2) de variables aléatoires réelles discrètes et pour toute fonction g définie sur $X_1(\Omega) \times X_2(\Omega)$ à valeurs dans \mathbb{R} , la loi de probabilité de la variable aléatoire réelle discrète $Z = g(X_1, X_2)$ est l'application

$$Z(\Omega) \longrightarrow \mathbb{R} \\ z_i \longmapsto \sum_{\substack{(x_1, x_2) \in X_1(\Omega) \times X_2(\Omega) \\ g(x_1, x_2) = z_i}} P(X_1 = x_1, X_2 = x_2) .$$

2.2 Somme de deux variables aléatoires réelles discrètes

Proposition 8

Pour tout couple (X_1, X_2) de variables aléatoires réelles discrètes, l'application

$$Z : \begin{cases} \Omega & \longrightarrow \mathbb{R} \\ \omega & \longmapsto X_1(\omega) + X_2(\omega) \end{cases}$$

est une variable aléatoire réelle discrète, et

$$P(X_1 + X_2 = s) = \sum_{\substack{x_1 \in X_1(\Omega) \\ s - x_1 \in X_2(\Omega)}} P(X_1 = x_1, X_2 = s - x_1).$$

Si X_1 et X_2 sont des variables aléatoires réelles discrètes indépendantes, alors

$$P(X_1 + X_2 = s) = \sum_{\substack{(x_1, x_2) \in X_1(\Omega) \times X_2(\Omega) \\ x_1 + x_2 = s}} P(X_1 = x_1) P(X_2 = x_2).$$

ou encore

$$P(X_1 + X_2 = s) = \sum_{\substack{x_1 \in X_1(\Omega) \\ s - x_1 \in X_2(\Omega)}} P(X_1 = x_1) P(X_2 = s - x_1)$$

Preuve

Il s'agit simplement d'appliquer les définitions et les résultats ci-dessus. □

► Remarques

- En particulier si X_1 et X_2 sont des variables aléatoires réelles indépendantes et si $X_1(\Omega) = X_2(\Omega) = \mathbb{N}$, alors pour tout entier naturel n

$$P(X + Y = n) = \sum_{k+\ell=n} P(X = k) P(Y = \ell) = \sum_{k=0}^n P(X = k) P(Y = n - k).$$

- On dit que la loi de $X_1 + X_2$ est le **produit de convolution** des lois de X_1 et X_2 . Cela sera à rapprocher du produit de convolution de deux lois de variables aléatoires réelles à densité indépendantes.

Somme de deux variables de Poisson

Théorème 4

Pour tout couple (X_1, X_2) de variables aléatoires réelles discrètes **indépendantes** suivant respectivement deux lois de Poisson de paramètres λ_1 et λ_2 la somme $X_1 + X_2$ suit une loi de Poisson de paramètre $\lambda_1 + \lambda_2$.

Preuve

Voir le tome 1, p 855. □

Somme de deux variables binomiales de paramètres (n_1, p) et (n_2, p)

Théorème 5

Pour tout couple (X_1, X_2) de variables aléatoires discrètes **indépendantes** X_1 et X_2 suivant respectivement deux lois binomiales de paramètres (n_1, p) et (n_2, p) la somme $X_1 + X_2$ suit une loi binomiale de paramètre $(n_1 + n_2, p)$.

Preuve

Voir le tome 1, p 854. □

2.3 Espérance d'une fonction d'un vecteur aléatoire discret

On sait que, si X et Y sont des variables aléatoires discrètes, alors, pour toute fonction g , $g(X, Y)$ est une variable aléatoire discrète.

D'autre part, une variable aléatoire réelle discrète Z (avec $Z(\Omega) = \{z_1, \dots, z_n, \dots\}$) admet une espérance si, et seulement si, la série $\sum_{k=1}^{+\infty} z_k P(Z = z_k)$ est absolument

convergente. On a alors $E(Z) = \sum_{k=1}^{+\infty} z_k P(Z = z_k)$.

Pour calculer l'espérance de $Z = g(X, Y)$, il n'est pas nécessaire d'en déterminer la loi, car l'on dispose d'un théorème, dit **théorème de transfert**.

Théorème 6

Soit $X = (X, Y)$ un couple de variables aléatoires réelles discrètes admettant chacune une espérance, et soit g une fonction réelle de deux variables réelles. On suppose que la variable aléatoire réelle discrète $Z = g(X, Y)$ admet une espérance. On a alors

$$E(Z) = \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} g(x, y) P(X = x, Y = y).$$

Preuve

Notons $X(\Omega) = \{x_j | j \in \mathbb{N}\}$, $Y(\Omega) = \{y_k | k \in \mathbb{N}\}$ et $Z(\Omega) = \{z_i | i \in \mathbb{N}\}$.

Il est clair que $Z(\Omega) = g(X(\Omega) \times Y(\Omega))$.

La loi de Z est donnée par $P(Z = z_i) = \sum_{\substack{(x_j, y_k) \in X(\Omega) \times Y(\Omega) \\ g(x_j, y_k) = z_i}} P(X = x_j, Y = y_k)$.

Comme, par hypothèse, Z admet une espérance,

$$E(Z) = \sum_{i=1}^{+\infty} z_i P(Z = z_i)$$

cette série étant absolument convergente.

Par suite, en utilisant judicieusement le théorème de sommation par paquets

$$\begin{aligned}
 E(Z) &= \sum_{i=1}^{+\infty} z_i \sum_{\substack{(x_j, y_k) \in X(\Omega) \times Y(\Omega) \\ g(x_j, y_k) = z_i}} P(X = x_j, Y = y_k) \\
 &= \sum_{i=1}^{+\infty} \sum_{\substack{(x_j, y_k) \in X(\Omega) \times Y(\Omega) \\ g(x_j, y_k) = z_i}} z_i P(X = x_j, Y = y_k) \\
 &= \sum_{(x_j, y_k) \in X(\Omega) \times Y(\Omega)} g(x_j, y_k) P(X = x_j, Y = y_k) \\
 &= \sum_{j=1}^{+\infty} \sum_{k=1}^{+\infty} g(x_j, y_k) P(X = x_j, Y = y_k). \quad \square
 \end{aligned}$$

Dans le cas où il existe des réels a et b tels que $g(x, y) = ax + by$, on dispose du théorème suivant

Théorème 7

Pour tout couple $X = (X, Y)$ de variables aléatoires réelles discrètes admettant chacune une espérance et tout couple (a, b) de réels, alors la variable aléatoire réelle discrète $aX + bY$ admet une espérance, et $E(aX + bY) = aE(X) + bE(Y)$.

Preuve

Ce résultat a été démontré dans le premier tome de cet ouvrage (voir pp. 778 à 780). □

2.4 Croissance de l'espérance

Proposition 9

Toute variable aléatoire réelle discrète positive admettant une espérance nulle est presque sûrement égale à 0 (c'est-à-dire que $P(X = 0) = 1$).

Preuve

Ce résultat a été démontré dans le premier tome du présent ouvrage (voir p. 778). □

Proposition 10

Soit X une variable aléatoire réelle discrète admettant une espérance. Si $X = 0$ presque sûrement (c'est-à-dire si $P(X = 0) = 1$), alors $E(X) = 0$.

Preuve

Soit $X(\Omega) = \{0, x_1, x_2, \dots, x_n, \dots\}$.

Pour tout entier naturel n , on pose $E_n = 0 \times P(X = 0) + \sum_{k=1}^n x_k P(X = x_k)$.

Or l'on sait que $P(X \neq 0) = 0$. Donc, pour tout $x_k \neq 0$, $P(X = x_k) = 0$.

Il en résulte que, pour tout n , $E_n = 0$, et par suite, $\lim_{n \rightarrow +\infty} E_n(X) = E(X) = 0$. □

Corollaire 1

Soient X et Y deux variables aléatoires réelles discrètes admettant une espérance. Si $X = Y$ presque sûrement (c'est-à-dire si $P(X = Y) = 1$), alors $E(X) = E(Y)$.

Preuve

La condition $P(X = Y) = 1$ équivaut à la condition $P(X - Y = 0) = 1$, ce qui implique d'après ce qui précède que $E(X - Y) = 0$. Le résultat annoncé en découle d'après la linéarité de l'espérance. \square

Proposition 11

Soit X une variable aléatoire réelle discrète admettant une espérance. Si $X \geq 0$ presque sûrement (c'est-à-dire si $P(X \geq 0) = 1$), alors $E(X) \geq 0$.

Preuve

Soit $X(\Omega) = \{\dots, x_{-p}, \dots, x_{-1}, x_0 = 0, x_1, x_2, \dots, x_n, \dots\}$, où x_i est du signe de i . Soit $I = \{x_0, x_1, x_2, \dots, x_n, \dots\}$. On sait que $P(X \in I) = 1$, et donc que, pour tout $k < 0$, $P(X = x_k) = 0$.

Quels que soient les entiers positifs p et q , on pose

$$E_{p,q} = \sum_{k=-p}^q x_k P(X = x_k) = \sum_{k=0}^q x_k P(X = x_k).$$

Il en résulte $E(X) = \lim_{\substack{p \rightarrow +\infty \\ q \rightarrow +\infty}} E_{p,q} = \sum_{k=0}^{+\infty} x_k P(X = x_k) \geq 0$. \square

Corollaire 2

Soient X et Y deux variables aléatoires réelles discrètes admettant une espérance. Si $X \leq Y$ presque sûrement (c'est-à-dire si $P(X \leq Y) = 1$), alors $E(X) \leq E(Y)$.

Preuve

La condition $P(X \leq Y) = 1$ équivaut à la condition $P(Y - X \geq 0) = 1$, ce qui implique d'après ce qui précède que $E(Y - X) \geq 0$.

Le résultat annoncé en découle d'après la linéarité de l'espérance. \square

Corollaire 3

Soit X une variable aléatoire réelle presque sûrement positive admettant une espérance. Alors, quelle que soit la variable aléatoire réelle discrète Y admettant une espérance, $E(X + Y) \geq E(Y)$.

Preuve

Ce résultat résulte des résultats ci-dessus et de la linéarité de l'espérance. \square

2.5 Variance d'une somme, covariance

Rappels des propriétés de la variance

- Rappelons qu'une variable aléatoire réelle discrète X admet une variance si, et seulement si, X admet un moment d'ordre 2. Elle admet alors une espérance, et la variance $V(X)$ est donnée par

$$V(X) = E((X - E(X))^2).$$

La variance est positive ou nulle, et l'on définit l'**écart-type** $\sigma(X)$ comme la racine carrée de la variance

$$\sigma(X) = \sqrt{V(X)}.$$

- Si $V(X) = 0$, alors il existe un réel $a (= E(X))$ tel que $X = a$ presque sûrement (c'est-à-dire que $P(X = a) = 1$).
- Rappelons aussi que $V(aX + b) = a^2 V(X)$.
- D'autre part :

Proposition 12

Soit X une variable aléatoire admettant une variance. Alors, quel que soit le réel a

$$V(X) = E((X - a)^2) - (E(X) - a)^2.$$

Preuve

On a successivement

$$\begin{aligned} E((X - a)^2) &= E((X - E(X) + E(X) - a)^2) \\ &= E((X - E(X))^2) + 2E((X - E(X))(E(X) - a)) + E((E(X) - a)^2) \\ &= V(X) + (E(X) - a)^2 \quad (\text{car } E(X - E(X)) = 0 \text{ et } E(X) - a \in \mathbb{R}). \end{aligned}$$

□

Corollaire 4

Soit X une variable aléatoire admettant une variance. La valeur minimum de $E((X - a)^2)$ pour a constante réelle est $V(X)$. Cette valeur minimum est atteinte pour $a = E(X)$.

Preuve

Ce résultat est une conséquence claire de la proposition ci-dessus.

□

Corollaire 5

Soit X une variable aléatoire admettant une variance. On a alors

$$V(X) = E(X^2) - (E(X))^2.$$

Preuve

Cette formule classique, communément appelée **formule de Koenig-Huygens** résulte clairement du développement de

$$V(X) = E[(X - E(X))^2] = E(X^2 - 2E(X)X + (E(X))^2).$$

□

Variance d'une somme, covariance

Considérons des variables aléatoires réelles discrètes X et Y admettant une variance. Alors le produit XY admet une espérance (voir premier tome p. 863), et le calcul de la variance de la somme $X + Y$ donne

$$\begin{aligned}
 V(X + Y) &= E((X + Y)^2) - (E(X + Y))^2 \\
 &= E(X^2 + 2XY + Y^2) - (E(X) + E(Y))^2 \\
 &= E(X^2) + 2E(XY) + E(Y^2) - (E(X))^2 - 2E(X)E(Y) - (E(Y))^2 \\
 &= E(X^2) - (E(X))^2 + E(Y^2) - (E(Y))^2 + 2(E(XY) - E(X)E(Y)) \\
 &= V(X) + V(Y) + 2(E(XY) - E(X)E(Y)) \\
 &= V(X) + V(Y) + 2E((X - E(X))(Y - E(Y))).
 \end{aligned}$$

Définition 10

Pour tout couple (X, Y) de variables aléatoires réelles discrètes admettant une variance, on appelle **covariance** de X et Y , et l'on note $\text{Cov}(X, Y)$, le réel

$$\text{Cov}(X, Y) = E(X - E(X))(Y - E(Y)) = E(XY) - E(X)E(Y).$$

De cette définition et du calcul ci-dessus, résulte le

Théorème 8

Pour tout couple (X, Y) de variables aléatoires réelles discrètes admettant une variance

$$V(X + Y) = V(X) + V(Y) + 2\text{Cov}(X, Y).$$

2.6 Propriétés de la covariance

Proposition 13

La covariance est une forme bilinéaire symétrique positive de l'espace vectoriel des variables aléatoires admettant un moment d'ordre 2.

Preuve

Il s'agit d'une reformulation en termes d'algèbre bilinéaire des propriétés démontrées dans le premier tome p. 865. \square

Corollaire 6

Pour toute variable aléatoire réelle discrète X admettant une variance, on a $\text{Cov}(X, X) = V(X)$. De plus, $\text{Cov}(X, X) = 0$ si, et seulement si, X est constante presque sûrement (c'est-à-dire $P(X = E(X)) = 1$).

Preuve

Cela résulte directement de ce qui précède. □

► **Remarque**

Il en résulte que la forme bilinéaire Cov n'est pas « définie » (au sens de l'algèbre bilinéaire).

Proposition 14

Pour tout couple (X, Y) de variables aléatoires discrètes **indépendantes** admettant une variance

$$E(XY) = E(X)E(Y), \quad V(X + Y) = V(X) + V(Y), \quad \text{Cov}(X, Y) = 0.$$

Preuve

Ces trois résultats sont étroitement liés. La démonstration a été faite dans le premier tome p. 866. □

► **Remarque**

Insistons sur le fait que la réciproque de cette proposition est fausse.

Matrice des variances-covariances

Soit (X, Y) un couple de variables aléatoires réelles discrètes admettant une variance, et a et b des réels.

- Le calcul de $V(aX + bY)$ conduit à

$$\begin{aligned} V(aX + bY) &= V(aX) + V(bY) + 2 \text{Cov}(aX, bY) \\ &= a^2 V(X) + b^2 V(Y) + 2ab \text{Cov}(X, Y) \end{aligned}$$

ce qui peut s'écrire matriciellement sous la forme

$$V(aX + bY) = (a \quad b) \begin{pmatrix} V(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & V(Y) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}.$$

- De même, a' et b' désignant aussi des réels, le calcul de $\text{Cov}(aX + bY, a'X + b'Y)$ conduit aux égalités

$$\begin{aligned} \text{Cov}(aX + bY, a'X + b'Y) &= a \text{Cov}(X, a'X + b'Y) + b \text{Cov}(Y, a'X + b'Y) \\ &= aa' V(X) + ab' \text{Cov}(X, Y) + ba' \text{Cov}(X, Y) + bb' V(Y) \\ &= aa' V(X) + (ab' + a'b) \text{Cov}(X, Y) + bb' V(Y) \end{aligned}$$

ce qui peut s'écrire matriciellement sous la forme

$$\text{Cov}(aX + bY, a'X + b'Y) = (a \quad b) \begin{pmatrix} V(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & V(Y) \end{pmatrix} \begin{pmatrix} a' \\ b' \end{pmatrix}.$$

Définition 11

Pour tout couple (X, Y) de variables aléatoires réelles discrètes admettant une variance, la **matrice des variances covariances** de (X, Y) est la matrice

$$\begin{pmatrix} V(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & V(Y) \end{pmatrix}.$$

2.7 Corrélation linéaire**Coefficient de corrélation linéaire****Théorème 9**

Pour tout couple (X, Y) de variables aléatoires réelles discrètes admettant une variance

$$\left(\text{Cov}(X, Y)\right)^2 \leq V(X)V(Y).$$

Preuve

Le résultat s'obtient par simple application de l'inégalité de Cauchy-Schwarz (cf. remarque page 32). \square

Théorème 10

Pour tout couple (X, Y) de variables aléatoires réelles discrètes admettant une variance et vérifiant $V(X) \neq 0$, alors $(\text{Cov}(X, Y))^2 = V(X)V(Y)$ si, et seulement si, il existe deux réels a et b tels que $Y = aX + b$ presque sûrement (c'est-à-dire tels que $P(Y = aX + b) = 1$).

Preuve

Cette démonstration ressemble beaucoup à l'étude du cas d'égalité dans l'inégalité de Cauchy-Schwarz pour un produit scalaire.

Soit λ un réel. On a $V(\lambda X + Y) = \lambda^2 V(X) + 2\lambda \text{Cov}(X, Y) + V(Y)$; $V(\lambda X + Y)$ est donc un trinôme en λ , de discriminant réduit $\Delta' = \text{Cov}(X, Y)^2 - V(X)V(Y)$.

Par hypothèse, ce discriminant est nul. Il existe donc un réel λ_0 (et un seul), tel que $V(\lambda_0 X + Y) = 0$, c'est-à-dire tel que la variable aléatoire $\lambda_0 X + Y = 0$ soit quasi certaine (c'est-à-dire telle qu'il existe un réel μ_0 tel que $P(\lambda_0 X + Y = \mu_0) = 1$).

Le résultat annoncé en découle, avec $a = -\lambda_0$ et $b = \mu_0$. \square

Définition 12

Pour tout couple (X, Y) de variables aléatoires réelles discrètes admettant chacune une variance non nulle, le **coefficient de corrélation linéaire** de X et Y est le réel $\rho_{X,Y}$

défini par $\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$.

► **Remarques**

- Rappelons que $\sigma(X)$ est l'écart-type de X , c'est-à-dire la racine carrée de $V(X)$.
- Le coefficient de corrélation linéaire joue le même rôle vis-à-vis de la covariance que le cosinus par rapport au produit scalaire classique dans un espace euclidien.
- Si X et Y sont indépendantes, leur coefficient de corrélation linéaire est nul. Mais la réciproque est fautive.

Théorème 11

Pour tout couple (X, Y) de variables aléatoires réelles discrètes admettant des variances non nulles

$$-1 \leq \rho_{X,Y} \leq 1.$$

Preuve

Ce résultat découle directement de ce qui précède. □

2.8 Régression

- Soit (X, Y) un couple de variables aléatoires réelles. Si X et Y ne sont pas indépendantes, la connaissance de la valeur prise par X (par exemple) change la connaissance que l'on peut avoir des probabilités des valeurs prises par Y .

Lorsque l'on peut admettre que le phénomène aléatoire représenté par X peut, dans une certaine mesure, servir à prédire celui représenté par Y , on est conduit à rechercher une fonction φ , telle que $\varphi(X)$ serve à évaluer les valeurs prises par Y .

Un cas particulier, rencontré ci-dessus pour des variables aléatoires réelles discrètes admettant des variances, est le cas où elles sont **corrélées linéairement**, c'est-à-dire telles que $(\text{Cov}(X)(Y))^2 = V(X)V(Y)$, c'est-à-dire $|\rho_{X,Y}| = 1$, où l'on sait donc qu'il existe des réels a et b tels que $P(Y = aX + b) = 1$.

Mais dans le cas général, on n'aura pas la possibilité d'obtenir une formule de prévision aussi précise. On cherchera alors la fonction φ de telle manière que $Y = \varphi(X) + \varepsilon$, avec $E(Y) = E(\varphi(X))$. La variable aléatoire ε représente l'erreur commise en remplaçant Y par $\varphi(X)$. Elle sera centrée ($E(\varepsilon) = 0$) et on la supposera non corrélée linéairement avec X . Pour minimiser l'erreur de prévision, on cherchera à minimiser la variance de ε .

- Le cas le plus important en pratique est celui où φ est une fonction affine. C'est le seul que nous aborderons dans cet ouvrage.

Théorème 12

Pour tout couple (X, Y) de variables aléatoires réelles discrètes admettant une variance non nulle, il existe une variable aléatoire centrée ε , non corrélée linéairement avec X , et telle que

$$Y - E(Y) = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} (X - E(X)) + \varepsilon.$$

De plus $V(\varepsilon) = (1 - \rho_{X,Y}^2)V(Y)$.

Preuve

Soient donc deux variables aléatoires X et Y . On cherche à déterminer des réels a et b tels que $Y = aX + b + \varepsilon$ où ε est une variable aléatoire telle que $E(\varepsilon) = 0$ et $\text{Cov}(\varepsilon, X) = 0$.

En prenant l'espérance de chaque membre, on obtient

$$E(Y) = aE(X) + b + E(\varepsilon) = aE(X) + b,$$

et par suite

$$Y - E(Y) = a(X - E(X)) + \varepsilon.$$

D'autre part, en multipliant par $(X - E(X))$ et en prenant l'espérance, on obtient

$$\begin{aligned} E\left((X - E(X))(Y - E(Y))\right) &= aE\left((X - E(X))^2\right) + E\left(\varepsilon(X - E(X))\right), \\ \text{Cov}(X, Y) &= aV(X) + \text{Cov}(\varepsilon, X) \quad (\text{car } E(\varepsilon) = 0), \\ a &= \frac{\text{Cov}(X, Y)}{V(X)} = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X}. \end{aligned}$$

Réciproquement si on définit ε par

$$Y = E(Y) + \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} (X - E(X)) + \varepsilon,$$

on vérifie que $E(\varepsilon) = 0$ et que ε et X ne sont pas corrélées linéairement.

En prenant la variance de chaque membre, on obtient

$$V(Y) = \rho_{X,Y}^2 \frac{\sigma_Y^2}{\sigma_X^2} V(X) + V(\varepsilon) = \rho_{X,Y}^2 V(X) + V(\varepsilon).$$

Le résultat annoncé en découle. □

► **Remarques**

- La droite d'équation $y = ax + b = E(Y) + \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} (x - E(X))$ est appelée **droite de régression de y en x** . Il va de soi que l'on définirait de façon analogue, et avec des calculs semblables, la droite de régression de x en y .
- La variance de la variable aléatoire ε est parfaitement définie à partir des variables aléatoires X et Y . On ne peut donc pas espérer améliorer la précision de la corrélation en diminuant la variance de ε . Tout au plus pourra-t-on, par exemple, préférer celle des deux droites de régression qui correspondra à la plus faible des deux variances $V(X)$ et $V(Y)$.
- Dans la pratique, il est fréquent que l'on ne connaisse X et Y qu'à travers une série de réalisations du vecteur aléatoire (X, Y) . On sera alors conduit à estimer les coefficients des droites de régression (voir le chapitre consacré à l'estimation).

3. Vecteurs aléatoires discrets à valeurs dans \mathbb{R}^n

3.1 Généralités

Nous noterons, pour tout entier k de $\llbracket 1, n \rrbracket$, ϖ_k la projection de \mathbb{R}^n dans \mathbb{R} définie pour tout n -uplet (x_1, \dots, x_n) de \mathbb{R}^n par l'égalité $\varpi_k(x_1, \dots, x_n) = x_k$.

Définition 13

Soit (Ω, \mathcal{T}) un espace probabilisable.

Toute application X de Ω à valeurs dans \mathbb{R}^n , où n est un entier naturel supérieur ou égal à 2, et telle que pour tout entier k de $\llbracket 1, n \rrbracket$, l'application $X_k = \pi_k(X)$ soit une variable aléatoire réelle sur (Ω, \mathcal{T}) est appelée **vecteur aléatoire sur (Ω, \mathcal{T})** .

On note alors $X = (X_1, X_2, \dots, X_n)$ et X_k est appelée la k -ième **variable aléatoire réelle coordonnée**.

Si les variables aléatoires X_k sont toutes discrètes, on dit que le vecteur X est un **vecteur aléatoire réel discret**.

L'essentiel des notions et des propriétés que nous rencontrerons dans cette section sont des extensions naturelles des notions et propriétés relatives aux couples discrets, qui ont été rencontrées dans les deux précédentes sections, et ne nécessitent pas l'introduction de concepts vraiment nouveaux.

Si une démonstration est une simple adaptation de la démonstration dans le cas de couples de variables aléatoires réelles discrètes, nous nous bornerons à le signaler, sans développer cette démonstration.

Proposition 15

Pour tout vecteur aléatoire discret $\mathbf{X} = (X_1, X_2, \dots, X_n)$, la famille d'événements

$$([X_1 = x_1] \cap [X_2 = x_2] \cap \dots \cap [X_n = x_n])_{(x_1, x_2, \dots, x_n) \in X_1(\Omega) \times X_2(\Omega) \times \dots \times X_n(\Omega)},$$

encore notée

$$(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)_{(x_1, x_2, \dots, x_n) \in X_1(\Omega) \times X_2(\Omega) \times \dots \times X_n(\Omega)}$$

$$(\mathbf{X} = (x_1, x_2, \dots, x_n))_{(x_1, x_2, \dots, x_n) \in X_1(\Omega) \times X_2(\Omega) \times \dots \times X_n(\Omega)},$$

est un système complet d'événements de (Ω, \mathcal{T}) appelé **le système complet d'événements associé au vecteur aléatoire $\mathbf{X} = (X_1, X_2, \dots, X_n)$** .

La tribu $\mathcal{A}_{\mathbf{X}}$ engendrée par ce système complet est appelée la tribu associée à \mathbf{X} .

Preuve

Simple extension de la démonstration faite dans le cas $n = 2$. □

3.2 Loi conjointe et lois marginales

Définition 14

Pour tout n -uplet $\mathbf{X} = (X_1, X_2, \dots, X_n)$ de variables aléatoires réelles discrètes sur l'espace probabilisé (Ω, \mathcal{T}, P) , l'application

$$\hat{P}_{\mathbf{X}} : \begin{cases} X_1(\Omega) \times X_2(\Omega) \times \dots \times X_n(\Omega) & \longrightarrow & \mathbb{R} \\ (x_1, x_2, \dots, x_n) & \longmapsto & P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \end{cases}$$

est appelée la **loi du vecteur \mathbf{X}** ou la **loi conjointe des variables aléatoires X_1, X_2, \dots, X_n** .

► **Remarque**

Il est clair que, pour tout n -uplet $\mathbf{X} = (X_1, X_2, \dots, X_n)$ de variables aléatoires réelles discrètes sur l'espace probabilisé (Ω, \mathcal{T}, P) de loi $\widehat{P}_{\mathbf{X}}$, alors

$$\begin{aligned} 1 &= \sum_{(x_1, x_2, \dots, x_n) \in \mathbf{X}(\Omega)} \widehat{P}_{\mathbf{X}}(x_1, x_2, \dots, x_n) \\ &= \sum_{x_1 \in X_1(\Omega)} \sum_{x_2 \in X_2(\Omega)} \cdots \sum_{x_n \in X_n(\Omega)} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n). \end{aligned}$$

Définition 15

Soit $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un n -uplet de variables aléatoires réelles discrètes sur l'espace probabilisé (Ω, \mathcal{T}, P) . La fonction de répartition de \mathbf{X} est la fonction de n variables $F_{\mathbf{X}}$ définie par $F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$.

► **Remarques**

- Une telle fonction de répartition est « constante par pavés », ce qui généralise les fonctions constantes par rectangles. Ses valeurs sont comprises entre 0 et 1. Ses applications partielles sont constantes par intervalles et croissantes.
- L'utilisation d'une telle fonction de répartition est très malaisée dès que n est supérieur ou égal à 3. Cette notion n'est alors pratiquement pas utilisée.

Définition 16

Pour tout n -uplet $\mathbf{X} = (X_1, X_2, \dots, X_n)$ de variables aléatoires réelles discrètes sur l'espace probabilisé (Ω, \mathcal{T}, P) , et pour tout entier naturel k de $\llbracket 1, n \rrbracket$, l'application

$$\widehat{P}_{X_k} : \begin{cases} X_k(\Omega) & \longrightarrow \mathbb{R} \\ x & \longmapsto P(X_k = x) \end{cases}$$

est appelée la **loi marginale de X_k** .

► **Remarques**

- On reconnaît ici la loi de X_k .
- Il est clair que, pour tout n -uplet $\mathbf{X} = (X_1, X_2, \dots, X_n)$ de variables aléatoires réelles discrètes, la loi marginale de X_k vérifie pour tout $x_k \in X_k(\Omega)$

$$\begin{aligned} P(X_k = x_k) &= \sum_{x_1 \in X_1(\Omega)} \cdots \sum_{x_{k-1} \in X_{k-1}(\Omega)} \sum_{x_{k+1} \in X_{k+1}(\Omega)} \cdots \\ &\cdots \sum_{x_n \in X_n(\Omega)} P(X_1 = x_1, \dots, X_{k-1} = x_{k-1}, X_{k+1} = x_{k+1}, \dots, X_n = x_n). \end{aligned}$$

Les définitions qui suivent sont de simples généralisations au cas de n variables aléatoires des définitions déjà données dans le cas de deux variables.

Définition 17

Pour tout vecteur aléatoire $\mathbf{X} = (X_1, X_2, \dots, X_n)$ sur l'espace probabilisé (Ω, \mathcal{T}, P) , tel que toutes les variables aléatoires composantes X_k admettent une espérance, le **vecteur espérance de \mathbf{X}** est le vecteur $E(\mathbf{X}) = (E(X_1), E(X_2), \dots, E(X_n))$.

Définition 18

Pour tout vecteur aléatoire $\mathbf{X} = (X_1, X_2, \dots, X_n)$ sur l'espace probabilisé (Ω, \mathcal{T}, P) , tel que toutes les variables aléatoires composantes X_k admettent une variance, la **matrice des variances-covariances de \mathbf{X}** est la matrice carrée $V(\mathbf{X})$, d'ordre n et de terme général $\alpha_{i,j} = \text{Cov}(X_i, X_j)$.

► **Remarques**

- Insistons sur le fait que, si l'une des variables composantes n'admet pas d'espérance, il en est de même pour le vecteur \mathbf{X} . De même, si l'une des variables composantes n'admet pas de variance, le vecteur \mathbf{X} n'admet pas de matrice des variances-covariances.
- La matrice $V(\mathbf{X})$ des variances-covariances d'un vecteur aléatoire $\mathbf{X} = (X_1, X_2, \dots, X_n)$ s'écrit (lorsqu'elle existe) sous la forme

$$V(\mathbf{X}) = \begin{pmatrix} V(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & V(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & V(X_n) \end{pmatrix}.$$

C'est une matrice symétrique positive.

Lois conditionnelles

Considérons un vecteur aléatoire réel $\mathbf{X} = (X_1, X_2, \dots, X_n)$ de dimension n sur l'espace probabilisé (Ω, \mathcal{T}, P) . Comme dans le cas de deux variables, on peut définir de manière analogue, des lois conditionnelles. Mais les possibilités sont beaucoup plus nombreuses.

- L'entier k_1 étant donné, si l'on conditionne par un événement de la forme $[X_{k_1} = x_1]$, la loi de \mathbf{X} conditionnée par $[X_{k_1} = x_1]$ est la loi d'un vecteur aléatoire de dimension $n - 1$.
- Les entiers k_1 et k_2 étant donnés, si l'on conditionne par un événement de la forme $[X_{k_1} = x_1, X_{k_2} = x_2]$, la loi de \mathbf{X} conditionnée par $[X_{k_1} = x_1, X_{k_2} = x_2]$ est la loi d'un vecteur aléatoire de dimension $n - 2$.
- Et ainsi de suite. . .
- Pour obtenir une variable aléatoire réelle, il suffit de conditionner par l'événement $[X_{k_1} = x_1, X_{k_2} = x_2, \dots, X_{k_{n-1}} = x_{n-1}]$.

Il serait fastidieux (et sans intérêt) de donner des formules générales. On se contentera d'appliquer les principes généraux dans chaque cas particulier.

3.3 Indépendance

Un n -uplet (X_1, X_2, \dots, X_n) de variables aléatoires réelles discrètes de l'espace probabilisé (Ω, \mathcal{T}, P) étant donné, on peut envisager l'indépendance entre les variables X_k de plusieurs façons.

Définition 19

Les variables aléatoires X_1, X_2, \dots, X_n sont dites **indépendantes deux à deux** si, quels que soient les entiers i et j , éléments différents de $\llbracket 1, n \rrbracket$, les variables aléatoires X_i et X_j sont indépendantes.

► Remarque

Si les variables aléatoires X_1, X_2, \dots, X_n sont indépendantes deux à deux, alors la matrice $V(X)$ des variances-covariances du vecteur aléatoire $\mathbf{X} = (X_1, X_2, \dots, X_n)$ est une matrice diagonale. La réciproque est fautive.

Définition 20

Soit (X_1, X_2, \dots, X_n) un n -uplet de variables aléatoires réelles discrètes de l'espace probabilisé (Ω, \mathcal{T}, P) . Les variables aléatoires X_1, \dots, X_n sont dites **mutuellement indépendantes** si, pour tout $(x_1, x_2, \dots, x_n) \in X_1(\Omega) \times X_2(\Omega) \times \dots \times X_n(\Omega)$,

$$P([X_1 = x_1] \cap [X_2 = x_2] \cap \dots \cap [X_n = x_n]) = P(X_1 = x_1) P(X_2 = x_2) \dots P(X_n = x_n).$$

► Remarques

- Il est clair que, si les variables aléatoires X_1, X_2, \dots, X_n sont mutuellement indépendantes, alors elles sont indépendantes deux à deux. La réciproque est fautive.
- Il est clair que, si les variables aléatoires X_1, X_2, \dots, X_n sont mutuellement indépendantes, alors, pour tout entier k de $\llbracket 1, n \rrbracket$, les variables aléatoires X_1, X_2, \dots, X_k sont mutuellement indépendantes.

Définition 21

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles discrètes de l'espace probabilisé (Ω, \mathcal{T}, P) telle que, pour toute partie finie I de \mathbb{N} , les variables aléatoires réelles discrètes X_i où i décrit I sont mutuellement indépendantes. Alors la suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles discrètes est dite une **suite de variables aléatoires réelles mutuellement indépendantes**.

Lemme des coalitions

Nous admettrons les résultats suivants

Proposition 16

Soit (X_1, X_2, \dots, X_n) un n -uplet de variables aléatoires réelles discrètes mutuellement indépendantes de l'espace probabilisé (Ω, \mathcal{T}, P) .

Alors, quel que soit l'entier p de $\llbracket 2, n-1 \rrbracket$, les σ -algèbres (tribus) respectivement associées à (X_1, X_2, \dots, X_p) et (X_{p+1}, \dots, X_n) sont indépendantes.

Ce résultat a partiellement été démontré au tome 1 p. 737.

Théorème 13

Soit (X_1, X_2, \dots, X_n) un n -uplet de variables aléatoires réelles discrètes mutuellement indépendantes de l'espace probabilisé (Ω, \mathcal{T}, P) .

Alors, quels que soient l'entier p de $\llbracket 2, n-1 \rrbracket$ et les fonctions φ de \mathbb{R}^p dans \mathbb{R} et ψ de \mathbb{R}^{n-p} dans \mathbb{R} , les variables aléatoires $\varphi(X_1, X_2, \dots, X_p)$ et $\psi(X_{p+1}, \dots, X_n)$ sont indépendantes.

Application aux variables binomiales

Théorème 14

Soit (X_1, X_2, \dots, X_m) un m -uplet de variables aléatoires mutuellement indépendantes telles que, pour tout k de $\llbracket 1, m \rrbracket$, X_k suit la loi binomiale de paramètre (n_k, p) .

Alors la variable aléatoire $Z_m = X_1 + X_2 + \dots + X_m$ suit la loi binomiale de paramètre $(n_1 + n_2 + \dots + n_m, p)$.

Preuve

Prouvons ce résultat par récurrence.

- On sait que la somme de deux variables aléatoires indépendantes X_1 et X_2 suivant respectivement les lois binomiales de paramètres (n_1, p) et (n_2, p) est une variable aléatoire Y qui suit une loi binomiale de paramètre $(n_1 + n_2, p)$ (voir tome 1, p. 853). La propriété annoncée est donc vraie pour $m = 2$.
- On suppose que la propriété est vraie au rang m ; on la démontre au rang $m + 1$.
Considérons alors $Z_{m+1} = X_1 + X_2 + \dots + X_m + X_{m+1} = Z_m + X_{m+1}$. Par hypothèse de récurrence, Z_m suit une loi binominale de paramètre $(n_1 + \dots + n_m, p)$. D'après le théorème admis, les variables aléatoires Z_m et X_{m+1} sont indépendantes. Donc, d'après la propriété pour $m = 2$, on peut conclure que Z_{m+1} suit la loi binomiale de paramètre $(n_1 + n_2 + \dots + n_m + n_{m+1}, p)$.
La propriété annoncée est donc héréditaire.
- Le résultat en découle par application du principe de récurrence. □

Application aux variables de Poisson

Théorème 15

Soit (X_1, X_2, \dots, X_m) un m -uplet de variables aléatoires mutuellement indépendantes telles que, pour tout k de $\llbracket 1, m \rrbracket$, X_k suit la loi de Poisson de paramètre λ_k .

Alors la variable aléatoire $Z_m = X_1 + X_2 + \dots + X_m$ suit la loi binomiale de paramètre $\lambda_1 + \lambda_2 + \dots + \lambda_m$.

Preuve

Prouvons ce résultat par récurrence.

- On sait que la somme de deux variables aléatoires indépendantes X_1 et X_2 suivant respectivement des lois de Poisson de paramètres λ_1 et λ_2 est une variable aléatoire Y qui suit une loi de Poisson de paramètre $\lambda_1 + \lambda_2$ (voir tome 1 p. 854). La propriété annoncée est donc vraie pour $m = 2$.
- On suppose que la propriété est vraie au rang m ; on la démontré au rang $m + 1$.
Considérons alors $Z_{m+1} = X_1 + X_2 + \dots + X_m + X_{m+1} = Z_m + X_{m+1}$. Par hypothèse de récurrence, Z_m suit une loi de Poisson de paramètre $\lambda_1 + \dots + \lambda_m$. D'après le théorème admis, les variables aléatoires Z_m et X_{m+1} sont indépendantes. Donc, d'après la propriété pour $m = 2$, on peut conclure que Z_{m+1} suit la loi de Poisson de paramètre $\lambda_1 + \lambda_2 + \dots + \lambda_m + \lambda_{m+1}$.
La propriété annoncée est donc héréditaire.
- Le résultat en découle par application du principe de récurrence. □

3.4 Variables aléatoires fonctions d'un vecteur aléatoire discret

Généralités

Ici encore, nous nous bornons à étendre des résultats qui ont été rencontrés dans le cas de deux variables, les preuves s'effectuant de manière analogue.

Théorème 16

Pour tout vecteur aléatoire réel discret $\mathbf{X} = (X_1, X_2, \dots, X_n)$ de dimension n sur l'espace probabilisé (Ω, \mathcal{T}, P) et toute fonction g définie sur $X_1(\Omega) \times X_2(\Omega) \times \dots \times X_n(\Omega)$ à valeurs dans \mathbb{R} , alors $g \circ \mathbf{X}$, que l'on note usuellement $g(\mathbf{X})$, est une variable aléatoire discrète.

Proposition 17

Pour tout n -uplet (X_1, X_2, \dots, X_n) de variables aléatoires réelles discrètes de l'espace probabilisable (Ω, \mathcal{T}) et toute fonction g définie sur le produit $X_1(\Omega) \times X_2(\Omega) \times \dots \times X_n(\Omega)$, la tribu associée à la variable aléatoire réelle discrète définie par $Z = g(X_1, X_2, \dots, X_n)$ est incluse dans la tribu associée au n -uplet (X_1, X_2, \dots, X_n) .

$$\mathcal{A}_Z \subset \mathcal{A}_{(X_1, X_2, \dots, X_n)}.$$

Théorème 17

Pour tout n -uplet (X_1, X_2, \dots, X_n) de variables aléatoires réelles discrètes et toute fonction g définie sur $X_1(\Omega) \times \dots \times X_n(\Omega)$ à valeurs dans \mathbb{R} , la loi de probabilité de la variable aléatoire réelle discrète $Z = g(X_1, X_2, \dots, X_n)$ est l'application \hat{P}_Z de $Z(\Omega)$ dans \mathbb{R} définie par

$$\hat{P}_Z(z_i) = \sum_{\substack{(x_{i_1}, x_{i_2}, \dots, x_{i_n}) \in X_1(\Omega) \times X_2(\Omega) \times \dots \times X_n(\Omega) \\ g(x_{i_1}, x_{i_2}, \dots, x_{i_n}) = z_i}} P(X_1 = x_{i_1}, X_2 = x_{i_2}, \dots, X_n = x_{i_n}).$$

Théorème de transfert

Le théorème de transfert, d'abord rencontré pour les fonctions d'une variable aléatoire (voir premier tome p. 784), puis pour les fonctions de deux variables aléatoires (voir premier tome p. 860 et début de ce chapitre) s'étend, sans autre difficulté que celles des notations, aux fonctions d'un vecteur aléatoire réel discret de dimension quelconque.

Théorème 18

Soit (X_1, X_2, \dots, X_n) un n -uplet de variables aléatoires réelles discrètes admettant chacune une espérance et g une fonction réelle de n variables réelles. On suppose que la variable aléatoire réelle discrète $Z = g(X_1, X_2, \dots, X_n)$ admet une espérance. Alors la valeur $E(Z)$

de cette dernière est égale à

$$\sum_{(x_1, x_2, \dots, x_n) \in X_1(\Omega) \times X_2(\Omega) \times \dots \times X_n(\Omega)} g(x_1, x_2, \dots, x_n) P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n).$$

3.5 Combinaisons linéaires de variables aléatoires réelles discrètes

Théorème 19

Soit (X_1, X_2, \dots, X_n) un n -uplet de variables aléatoires réelles discrètes et \mathcal{X} la matrice

colonne $\mathcal{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$.

Soient a_1, a_2, \dots, a_n des réels et \mathbf{A} la matrice ligne $\mathbf{A} = (a_1 \ a_2 \ \dots \ a_n)$.

On note Z la variable aléatoire réelle définie par

$$Z = \mathbf{A}\mathcal{X} = \sum_{k=1}^n a_k X_k.$$

Alors

- $E(Z) = \mathbf{A}E(\mathcal{X}) = \sum_{k=1}^n a_k E(X_k)$
- $V(Z) = \mathbf{A}V(\mathcal{X})' \mathbf{A} = \sum_{k=1}^n a_k^2 V(X_k) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n a_i a_j \text{Cov}(X_i, X_j)$
 $= \sum_{k=1}^n a_k^2 V(X_k) + 2 \sum_{i=2}^n \sum_{j=1}^{i-1} a_i a_j \text{Cov}(X_i, X_j).$

Preuve

Ces résultats sont triviaux. Il suffit d'appliquer les définitions et les propriétés de l'espérance et de la variance.

Rappelons toutefois que $E(\mathcal{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix}$

et que $V(\mathcal{X}) = \begin{pmatrix} V(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & V(X_2) & \dots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \dots & V(X_n) \end{pmatrix}$.

□

► Remarque

Dans la plupart des ouvrages de probabilités et statistiques, pour une matrice \mathbf{A} , sa transposée est notée \mathbf{A}' . On rencontrera alors la deuxième formule du théorème ci-dessus sous la forme $V(Z) = \mathbf{A}V(\mathcal{X})\mathbf{A}'$.

Cas d'une somme**Corollaire 7**

Soit (X_1, X_2, \dots, X_n) un n -uplet de variables aléatoires réelles discrètes.

Alors $E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$ et

$$\begin{aligned} V(X_1 + X_2 + \dots + X_n) &= \sum_{k=1}^n V(X_k) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \\ &= \sum_{k=1}^n V(X_k) + 2 \sum_{j < i} \text{Cov}(X_i, X_j). \end{aligned}$$

Preuve

Il s'agit de la simple application du théorème ci-dessus pour $a_1 = a_2 = \dots = a_n = 1$. □

Application à la loi binomiale

Soit $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un n -uplet de variables aléatoires X_k suivant toutes une même loi de Bernoulli de paramètre p . On suppose de plus que les variables X_k sont deux à deux indépendantes.

On considère alors la variable aléatoire $Y = X_1 + X_2 + \dots + X_n$. On sait que Y suit la loi binomiale de paramètre (n, p) , que, pour tout k de $\llbracket 1, n \rrbracket$, $E(X_k) = p$ et $V(X_k) = p(1 - p)$, et enfin que, pour tout couple (i, j) de $\llbracket 1, n \rrbracket$ avec $i \neq j$, $\text{Cov}(X_i, X_j) = 0$.

On retrouve ainsi $E(Y) = np$ et $V(Y) = np(1 - p)$.

Application à la loi hypergéométrique

Considérons une urne contenant a boules blanches, et $N - a$ boules noires. On extrait sans remise n boules de cette urne, et l'on note Z le nombre de boules blanches obtenues.

On sait (voir tome 1 p. 801) que Z suit une loi hypergéométrique de paramètre $(N, n, \frac{a}{N})$.

On sait aussi que, pour $\max\{0, n - N + a\} \leq k \leq \min\{n, a\}$, on a

$$P(Z = k) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}.$$

Supposons que les a boules blanches soient numérotées de 1 à a , et, pour $i \in \llbracket 1, a \rrbracket$, notons X_i la variable de Bernoulli qui vaut 1 si la boule blanche numéro i fait partie des n boules tirées, et 0 sinon. Il est clair, dans ces conditions, que $Z = X_1 + X_2 + \dots + X_a$.

- Pour tout k de $\llbracket 1, a \rrbracket$, on a clairement $P(X_k = 1) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$, et donc

$$E(X_k) = \frac{n}{N}, \quad V(X_k) = \frac{n}{N} \frac{N-n}{N}.$$

- D'autre part, pour i et j éléments différents de $\llbracket 1, a \rrbracket$, la variable aléatoire $X_i X_j$ est une variable de Bernoulli.

$$\text{Il est clair que } P(X_i X_j = 1) = P(X_i = 1, X_j = 1) = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}.$$

- Il en résulte alors que

$$\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) = \frac{n(n-1)}{N(N-1)} - \frac{n^2}{N^2} = -\frac{n(N-n)}{N^2(N-1)}.$$

- Le calcul de la variance de Z donne alors

$$V(Z) = V\left(\sum_{k=1}^a X_k\right) = \sum_{k=1}^a V(X_k) + \sum_{i \neq j} \text{Cov}(X_i, X_j).$$

Dans cette formule, les variances (au nombre de a) sont toutes égales à $\frac{n}{N} \frac{N-n}{N}$.

Les covariances sont toutes égales à $-\frac{n(N-n)}{N^2(N-1)}$, et il y en a autant que de couples d'entiers différents, éléments de $\llbracket 1, a \rrbracket$, c'est-à-dire $a(a-1)$. On obtient alors

$$V(Z) = a \frac{n}{N} \frac{N-n}{N} - a(a-1) \frac{n(N-n)}{N^2(N-1)} = n \frac{N-n}{N-1} \frac{a}{N} \left(1 - \frac{a}{N}\right).$$

À ce stade, il est d'usage de poser $p = \frac{a}{N}$ et $q = 1 - p$. On obtient alors

$$\boxed{V(Z) = npq \frac{N-n}{N-1}}.$$

► **Remarque**

Ce résultat est à rapprocher de la variance de la loi binomiale de paramètre (n, p) .

Le coefficient $\frac{N-n}{N-1}$ est le **coefficient d'exhaustivité**. Si n est petit devant N , ce coefficient est voisin de 1, et l'on peut utiliser la loi binomiale $\mathcal{B}(n, p)$ comme approximation de la loi hypergéométrique $\mathcal{H}(N, n, p)$.

3.6 L'exemple de la loi multinomiale

Vecteur aléatoire de Bernoulli

Soient p_1, p_2, \dots, p_m m réels positifs dont la somme est égale à 1 et X_1, X_2, \dots, X_m m variables aléatoires telles que, pour tout k de $\llbracket 1, m \rrbracket$, X_k suit la loi de Bernoulli de paramètre p_k , et que $X_1 + X_2 + \dots + X_m = 1$. Cette dernière condition implique qu'une, et une seule, variable X_k prenne la valeur 1, les autres prenant la valeur 0.

On dit alors que le vecteur $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix}$, suit une loi de Bernoulli généralisée de dimension m , de paramètre $\vec{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{pmatrix}$.

Compte tenu de la définition des lois marginales, on a clairement $E(\mathbf{X}) = \vec{p}$.

Pour tout k de $\llbracket 1, m \rrbracket$, on a $V(X_k) = p_k(1 - p_k)$.

Pour tout couple (i, j) d'éléments distincts de $\llbracket 1, m \rrbracket$, la variable aléatoire $X_i X_j$ est nulle, car une et une seule des variables composantes est égale à 1, les autres étant égales à 0.

Donc $\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) = -E(X_i)E(X_j) = -p_i p_j$.

Il en résulte que la matrice des variances-covariances de \mathbf{X} est

$$V(\mathbf{X}) = \begin{pmatrix} p_1(1 - p_1) & -p_1 p_2 & \cdots & -p_1 p_m \\ -p_2 p_1 & p_2(1 - p_2) & \cdots & -p_2 p_m \\ \vdots & \vdots & \ddots & \vdots \\ -p_m p_1 & -p_m p_2 & \cdots & p_m(1 - p_m) \end{pmatrix}.$$

► Remarque

Un tel vecteur de Bernoulli intervient, par exemple lorsque l'on dispose d'une urne contenant des boules numérotées k en proportion p_k (pour $k \in \llbracket 1, m \rrbracket$), et que l'on tire une boule de cette urne. Le numéro de la boule tirée correspond au numéro de la composante qui prend la valeur 1.

Vecteur aléatoire multinomial

Considérons l'urne faisant l'objet de la remarque ci-dessus, et effectuons n tirages, avec remise, dans cette urne. Pour $k \in \llbracket 1, m \rrbracket$, on compte alors le nombre Y_k de boules

numérotées k ainsi obtenues. On obtient ainsi le vecteur aléatoire $\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix}$, tel

que, pour tout $k \in \llbracket 1, m \rrbracket$, $Y_k \in \llbracket 0, n \rrbracket$ et $Y_1 + Y_2 + \dots + Y_m = n$.

Il est clair que la loi marginale de Y_k est la loi binomiale de paramètre (n, p_k) .

On dit alors que \mathbf{Y} suit la loi multinomiale de paramètre (n, \vec{p}) , avec $\vec{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{pmatrix}$.

- Déterminons la loi du vecteur multinomial \mathbf{Y} .

Soit $(n_1, n_2, \dots, n_m) \in \llbracket 0, n \rrbracket$ tel que $n_1 + n_2 + \dots + n_m = n$.

Pour déterminer la probabilité $\varpi = P\left(\mathbf{Y} = \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_m \end{pmatrix}\right)$, considérons successive-

ment les numéros des boules tirées.

L'ordre des tirages n'intervient pas dans le nombre de boules de chaque catégorie tirées à la fin du processus. De plus, les tirages sont indépendants.

Au cours des n tirages, il y a $\binom{n}{n_1}$ façons de tirer n_1 boules numérotées 1, chaque tirage intervenant avec la probabilité p_1 .

La probabilité de tirer n_1 boules numérotées 1 au cours des n tirages est donc

$$\varpi_1 = \frac{n!}{n_1!(n - n_1)!} p_1^{n_1}.$$

Une fois ces n_1 boules numérotées 1 obtenues, il reste $n - n_1$ boules à tirer. La probabilité ϖ_2 de tirer n_2 boules numérotées 2 se calcule alors de la même façon.

On trouve $\varpi_2 = \frac{(n - n_1)!}{n_2!(n - n_1 - n_2)!} p_2^{n_2}$.

On procède de même pour calculer $\varpi_3 = \frac{(n - n_1 - n_2)!}{n_3!(n - n_1 - n_2 - n_3)!} p_3^{n_3}$, puis

$\varpi_4 = \frac{(n - n_1 - n_2 - n_3)!}{n_4!(n - n_1 - n_2 - n_3 - n_4)!} p_4^{n_4}$, et ainsi de suite jusqu'à $\varpi_m = p_m^{n_m}$ (il n'y a plus qu'une façon de choisir les rangs des n_m boules numérotées m parmi les n_m emplacements qui restent à ce stade du raisonnement).

On obtient alors, après calcul

$$\varpi = \frac{n!}{n_1!n_2!\dots n_m!} p_1^{n_1} p_2^{n_2} \dots p_m^{n_m}.$$

- On a remarqué que, pour tout $k \in \llbracket 0, m \rrbracket$, la loi marginale de Y_k est la loi binomiale de paramètre (n, p_k) , dont l'espérance est $E(Y_k) = np_k$. Il en résulte que le vecteur espérance de \mathbf{Y} est donné par $E(\mathbf{Y}) = n\vec{p}$.

- Pour calculer la matrice des covariances de \mathbf{Y} , considérons \mathbf{Y} comme la somme de n vecteurs aléatoires indépendants $\mathbf{X}^{(k)}$ suivant une loi de Bernoulli généralisée de paramètre \vec{p} .

Notons ainsi $\mathbf{Y} = \mathbf{X}^{(1)} + \mathbf{X}^{(2)} + \dots + \mathbf{X}^{(n)}$.

Remarquons que, compte tenu de la linéarité de l'espérance, on retrouve aisément le vecteur espérance de \mathbf{Y} .

La loi marginale de Y_k étant pour tout $k \in \llbracket 0, m \rrbracket$ la loi binomiale de paramètre (n, p_k) , le terme diagonal correspondant de la matrice des covariances est $np_k(1 - p_k)$.

Pour i et j éléments différents de $k \in \llbracket 0, m \rrbracket$, on a $Y_i = \sum_{k=1}^n X_i^{(k)}$ et $Y_j = \sum_{s=1}^n X_j^{(s)}$.

D'après la bilinéarité de la covariance, on a alors

$$\text{Cov}(Y_i, Y_j) = \sum_{k=1}^n \sum_{s=1}^n \text{Cov}(X_i^{(k)}, X_j^{(s)}).$$

Dans cette somme double, lorsque k et s sont différents, $X_i^{(k)}$ et $X_j^{(s)}$ sont des composantes de deux vecteurs indépendants. Elles sont donc indépendantes, et leur covariance est nulle.

Il reste alors $\text{Cov}(Y_i, Y_j) = \sum_{k=1}^n \text{Cov}(X_i^{(k)}, X_j^{(k)})$.

$X_i^{(k)}$ et $X_j^{(k)}$ étant deux composantes distinctes d'un même vecteur de Bernoulli, on a $\text{Cov}(X_i^{(k)}, X_j^{(k)}) = -p_i p_j$ (voir ci-dessus), et donc $\text{Cov}(Y_i, Y_j) = -np_i p_j$.

La matrice des covariances de \mathbf{Y} est donc

$$V(\mathbf{Y}) = \begin{pmatrix} np_1(1 - p_1) & -np_1 p_2 & \cdots & -np_1 p_m \\ -np_2 p_1 & np_2(1 - p_2) & \cdots & -np_2 p_m \\ \vdots & \vdots & \ddots & \vdots \\ -np_m p_1 & -np_m p_2 & \cdots & np_m(1 - p_m) \end{pmatrix}$$

3.7 Remarque

Le programme de la classe ne prévoit rien à propos des vecteurs aléatoires dont les composantes sont des variables aléatoires réelles à densité. Or nous serons conduits à estimer des paramètres de lois à densités, et donc à utiliser des réalisations de tels vecteurs aléatoires.

Il nous suffira de considérer que les définitions et propriétés rencontrées dans le présent chapitre, et qui sont nécessaires au chapitre du présent ouvrage consacré à l'estimation s'étendent généralement sans problème, parfois au prix de quelques hypothèses supplémentaires, aux vecteurs aléatoires à densité.

1. Soient X et Y des variables aléatoires réelles dont le coefficient de corrélation est noté $\rho_{X,Y}$. Soient a et b deux réels non nuls, et b et d des réels. Prouver que $\rho_{aX+b,cY+d} = \rho_{X,Y}$.
2. Soient X une variable aléatoire réelle, a un réel non nul et b un réel.
 - a) Calculer $\rho_{X,aX+b}$.
 - b) Quelle est la valeur du coefficient de corrélation de X et Y si Y est une fonction affine non constante de X ?
3. Soient X et Y des variables aléatoires admettant des variances non nulles. Soit

$$Z = \left(\frac{1}{\sigma_Y}\right) Y - \left(\frac{\rho_{X,Y}}{\sigma_X}\right) X.$$

- a) Calculer $V(Z)$.
 - b) Que peut-on en déduire si $|\rho_{X,Y}| = 1$?
2. Soit (X, Y) un couple de variables aléatoires discrètes dont la loi est définie dans le tableau ci-dessous.

X \ Y	1	2	3	4
1	0,08	0,04	0,16	0,12
2	0,04	0,02	0,08	0,06
3	0,08	0,04	0,16	0,12

1.
 - a) Déterminer les lois marginales de ce couple.
 - b) Les lois de X et Y sont-elles indépendantes?
 - c) Calculer la covariance du couple (X, Y) .
2.
 - a) Déterminer les lois conditionnelles de X sachant que $Y = 2$ et de Y sachant que $X \in \{1, 3\}$.
 - b) Déterminer la loi de la variable aléatoire $E(Y | X)$.
 - c) Calculer $E(E(Y | X))$ et comparer à $E(Y)$.
3. Déterminer la loi du couple $(\inf(X, Y), \sup(X, Y))$.

3. Soit (X, Y) un couple de variables aléatoires discrètes dont la loi est définie dans le tableau ci-dessous.

X \ Y	1	2	3	4
1	0	0	0	0,3
2	0,2	0	0	0
3	0	0	0,1	0
4	0,3	0,1	0	0

1.
 - a) Déterminer les lois marginales de ce couple.
 - b) Les lois de X et Y sont-elles indépendantes?
 - c) Calculer la covariance du couple (X, Y) .
2.
 - a) Déterminer les lois conditionnelles de X sachant que $Y = 2$ et de Y sachant que $X \in \{1, 4\}$.
 - b) Déterminer la loi de la variable aléatoire $E(X | Y)$.

c) Calculer $E(E(X | Y))$ et comparer à $E(X)$.

3. Déterminer la loi du couple $(\inf(X, Y), \sup(X, Y))$.

4. On considère un couple (X, Y) de variables aléatoires réelles à valeurs dans \mathbb{N} , pour lesquels il existe un réel λ tel que la loi de (X, Y) soit définie par

$$\forall (j, k) \in \mathbb{N}^2, \quad P([X = j] \cap [Y = k]) = \frac{(j+k)\lambda^{j+k}}{e^j k!}.$$

1. Déterminer λ .
2. Déterminer la loi de X .
3. Les variables aléatoires X et Y sont-elles indépendantes ?
4. Calculer $E(2^{X+Y})$.

5. On donne $\forall (j, k) \in \mathbb{N}^2, P([X = j] \cap [Y = k]) = \frac{1}{e^j k!} \times \frac{1}{2^{j+1}}$.

1. Montrer que l'on définit ainsi la loi conjointe de deux variables aléatoires X et Y à valeurs dans \mathbb{N} .
2. Déterminer la loi de X et celle de Y .
3. Les variables aléatoires X et Y sont-elles indépendantes ?
4. Calculer $E(X) V(X), E(Y)$ et $V(Y)$.

6. On donne trois réels positifs p_1, p_2 et p_3 de somme 1. On définit alors

$$\forall (j, k) \in \mathbb{N}^2, \quad P([X = j] \cap [Y = k]) = \frac{n! p_1^j p_2^k p_3^{n-j-k}}{j! k! (n-j-k)!}.$$

1. Montrer que l'on définit ainsi la loi conjointe de deux variables aléatoires X et Y à valeurs dans \mathbb{N} .
2. Déterminer le loi de X , son espérance et sa variance.
3. Déterminer la loi conditionnelle de Y sachant que $[X = j]$.

7. Dans un sac, il y a $n - 2$ boules noires et 2 boules blanches. On tire les boules une à une, sans remise. Soit X la variable aléatoire égale au rang de la première boule blanche tirée, et Y la variable aléatoire égale au rang de la deuxième boule blanche tirée.

1. Déterminer la loi du couple (X, Y) .
2. En déduire les lois de X et de Y .
3. Calculer le coefficient de corrélation du couple (X, Y) .
4. Donner les équations des droites de régression.

8. Soient p, ϖ deux réels de l'intervalle $[0, 1]$, et m un entier naturel non nul. On considère une variable X suivant la loi binomiale de paramètre (m, p) . Soit Y une variable aléatoire à valeurs dans \mathbb{N} telle que, pour tout entier k la loi conditionnelle de Y sachant que $X = k$ soit la loi binomiale de paramètre (k, ϖ) .

Déterminer la loi de Y .

9. Soient p et q des réels de l'intervalle $[0, 1]$, et λ un réel strictement positif. On considère deux variables aléatoires réelles X et Y telles que :

- X suit la loi de Poisson de paramètre λ
- Si $X = 0$, alors $Y = 0$ et, pour $n \in \mathbb{N}^*$, la loi conditionnelle $(Y \mid X = n)$ est la loi binomiale de paramètre (n, p) .

1. Déterminer la loi du couple (X, Y) .
2. Déterminer la loi de Y .

10. 1. Soit p un réel de l'intervalle $[0, 1]$. On pose $q = 1 - p$. On considère deux variables aléatoires X et Y indépendantes, à valeurs dans \mathbb{N} , et de même loi

$$\forall n \in \mathbb{N}, \quad P(X = n) = P(Y = n) = p q^n.$$

Montrer que les variables aléatoires $V = \inf(X, Y)$ et $W = X - Y$ sont indépendantes.

2. Réciproquement, soient X et Y deux variables indépendantes à valeurs dans \mathbb{N} , de même loi, telles que, pour tout entier naturel n , $P(X = n) \neq 0$, et telles que les variables aléatoires $V = \inf(X, Y)$ et $W = X - Y$ soient indépendantes. On pose $k = \frac{P(W = 1)}{P(W = 0)}$.

En calculant de deux façons le quotient $\frac{P([X = n + 1] \cap [Y = n])}{P([X = n] \cap [Y = n])}$ pour $n \in \mathbb{N}$, déterminer les lois de X et Y .

11. Soient X et Y deux variables de Poisson de paramètres respectifs λ et μ .

1. Déterminer la loi conditionnelle de X sachant que la somme $S = X + Y$ prend la valeur $S = s$.
2. En déduire la loi de la variable aléatoire $E(X \mid S)$.
3. Calculer $E(E(X \mid S))$. Quelle valeur retrouve-t-on ainsi ?

12. Soient a un entier naturel non nul, et n un entier naturel supérieur ou égal à 2. À un péage d'autoroute comportant n guichets, na voitures se présentent. Chaque conducteur choisit un guichet au hasard, de manière équiprobable. Les choix des automobilistes sont supposés indépendants entre eux. On note X_i le nombre de voitures ayant passé par le guichet numéro i .

1. Déterminer la loi, l'espérance et la variance de la variable aléatoire X_i .
2. a) Calculer $V(X_1 + X_2 + \dots + X_n)$.
b) En déduire $\text{Cov}(X_i, X_j)$ où i et j sont des entiers distincts quelconques de $\llbracket 1, n \rrbracket$.
3. a) Calculer le coefficient de corrélation linéaire de X_i et X_j .
b) Commenter le cas $n = 2$.

13. Le nombre N de personnes de présentant à un bureau de poste suit une loi de Poisson de paramètre λ .

Une personne vient au bureau de poste pour poster un envoi avec la probabilité p ($p \in]0, 1[$), et pour une autre opération avec la probabilité $q = 1 - p$.

On suppose que chaque personne n'effectue qu'une opération, et qu'elles font ces opérations indépendamment les unes des autres.

On note X le nombre des personnes qui viennent pour poster un envoi, et Y le nombre de celles qui viennent pour une autre opération.

1. Soit j un entier naturel. Quelle est la loi de X sachant que $[N = j]$?
2. Déterminer la loi du vecteur (X, N) .
3. **a)** En déduire la loi de X .
b) Donner les valeurs de $E(X)$ et de $V(X)$.
4. Montrer que X et Y sont indépendantes.
5. Calculer $\text{Cov}(X, N)$ (on pourra remarquer que $N = X + Y$).
6. Calculer $\rho_{X,N}$.

14. On considère 3 boîtes et une infinité de jetons. On place successivement chacun des jetons, au hasard dans l'une des trois boîtes. On suppose que chaque boîte est vide au départ, et de capacité illimitée. On suppose également qu'à chaque fois qu'un jeton est placé, il l'est de façon équiprobable dans chaque boîte.

Soit Y le nombre de jetons placés lorsque, pour la première fois, deux boîtes exactement sont occupées par au moins un jeton.

Soit Z le nombre de jetons placés lorsque, pour la première fois, les trois boîtes contiennent chacune au moins un jeton.

1. Calculer, pour $k \in \mathbb{N}^*$, la probabilité $P(Y = k)$.
2. **a)** Calculer, pour $(k, \ell) \in (\mathbb{N}^*)^2$ la probabilité conditionnelle $P(Z = \ell \mid Y = k)$.
b) En déduire la loi de Z .
3. **a)** Calculer $E(Z \mid Y = k)$. En déduire la loi de la variable aléatoire $E(Z \mid Y)$.
b) Calculer $E(Z)$ de deux façons différentes.

15. Dans une urne, on a placé des boules rouges en proportion p et des boules vertes en proportion $q = 1 - p$. On effectue n tirages avec remise dans cette urne, et l'on note R_n le nombre de boules rouges tirées et V_n le nombre de boules vertes tirées. On note alors $D_n = R_n - V_n$.

1. Déterminer la loi de D_n , son espérance et sa variance.
2. On note B_i la variable aléatoire qui prend la valeur 1 si le $i^{\text{ème}}$ tirage amène une boule rouge, et 0 sinon.
Soit n un entier naturel non nul. Déterminer la matrice des variances-covariances du vecteur $\mathbf{D} = (D_1, D_2, \dots, D_n)$.

Autres exercices

On cherchera également avec profit les problèmes et exercices de probabilité proposés aux concours suivants dans diverses séries :

HEC oral 1998, 1991 ; ESCP oral 1986 ; ESCP-EAP 2000, 2001, 2004 ; HEC 1990, 1991, 1999.

Variables aléatoires réelles à densité

10

Ce chapitre présente la seconde famille des lois de variables aléatoires réelles, à savoir celle dont l'image est \mathbb{R} (parfois \mathbb{R}_+) et que l'on ne peut donc plus définir par la donnée d'une famille de probabilités de la forme $P(X = x)$ (car elles sont alors toutes nulles) mais par une fonction de répartition $F(x) = P(X \leq x)$, donnée sous forme d'une intégrale de fonction simple.

1. Définition des variables aléatoires réelles à densité

Les variables aléatoires réelles à densité forment, avec les variables discrètes, le deuxième cas particulier très important de la notion générale de variable aléatoire réelle. Les résultats fondamentaux les concernant s'inspirent souvent du cas discret, à condition de remplacer par exemple des sommes de séries par des intégrales, souvent généralisées.

1.1 Fonctions de répartition et densités

Définition 1

On appelle **variable aléatoire réelle à densité** toute variable aléatoire réelle X dont la fonction de répartition F est continue sur \mathbb{R} et est C^1 sur \mathbb{R} privé d'un nombre fini de points.

Toute fonction positive ou nulle f telle que $f(x) = F'(x)$ **sauf peut-être en un nombre fini de points** est appelé une **densité** de f .

► Remarques

- La définition ci-dessus peut être précisée en disant que, si a est un point n'appartenant pas à l'ensemble fini des points retirés de \mathbb{R} , alors il existe un réel $h > 0$ tel que, sur $]a - h, a + h[$, F' existe et est continue, c'est-à-dire que F est de classe C^1 sur cet intervalle.

Toutefois, il serait inexact de dire que F est de classe C^1 par morceaux sur \mathbb{R} , comme le montre l'exemple ci-dessous.

- Les fonctions de répartition des variables aléatoires réelles discrètes étant discontinues en au moins un point, aucune d'elles ne peut être aussi une variable aléatoire réelle à densité.

Théorème 1

Toute variable aléatoire réelle à densité X et tout réel x vérifient l'égalité

$$P(X = x) = 0.$$

Preuve

Cela résulte de la continuité de F et de l'égalité $P(X = x) = F(x) - F(x - 0)$. □

► Remarques

- On rappelle qu'une fonction de répartition est nécessairement croissante (donc de dérivée positive ou nulle là où elle est dérivable) et de limites respectives 0 et 1 en $-\infty$ et $+\infty$.
- Lorsque plusieurs variables aléatoires apparaissent dans un même texte, il est alors préférable de noter $F = F_X$ et $f = f_X$; si le contexte est clair, la présence des indices X est superflue. Il est toutefois bon de remarquer que, si X détermine toujours sa fonction de répartition, il n'en va pas de même pour ses densités; à la différence de F_X , la notation f_X est donc ambiguë et désigne plutôt une classe de fonctions que l'une d'entre elles en particulier.
- Pour aider à la compréhension du concept assez délicat de densité, on dit parfois que, pour un petit accroissement Δx de la variable, $f(x)\Delta x$ est approximativement la probabilité pour que $X \in [x, x + \Delta x]$. Cela permet par exemple de mieux comprendre pourquoi certaines formules concernant les variables à densité, à l'aide d'intégrales, peuvent paraître comme proches des formules analogues portant sur les variables discrètes où n'intervient que des sommes finies ou dénombrables.

Exemples

1. L'exemple le plus simple est celui de F définie par $F(x) = 0$ si $x \leq 0$, $F(x) = x$ si $0 \leq x \leq 1$ et $F(x) = 1$ si $x \geq 1$ (les valeurs de $F(0)$ et $F(1)$ sont définies de manière surabondante mais cohérente).

Toute fonction f égale à 1 sur le segment $[0, 1]$ éventuellement privé d'un nombre fini de points, nulle sur les demi-droites $]-\infty, 0[$ et $]1, +\infty[$ éventuellement privées d'un nombre fini de points et de valeurs positives ou nulles arbitraires ailleurs est l'une des densités de la variable aléatoire X .

On notera que, même dans ce cas très simple, cette variable admet une infinité de densités distinctes : c'est d'ailleurs toujours le cas.

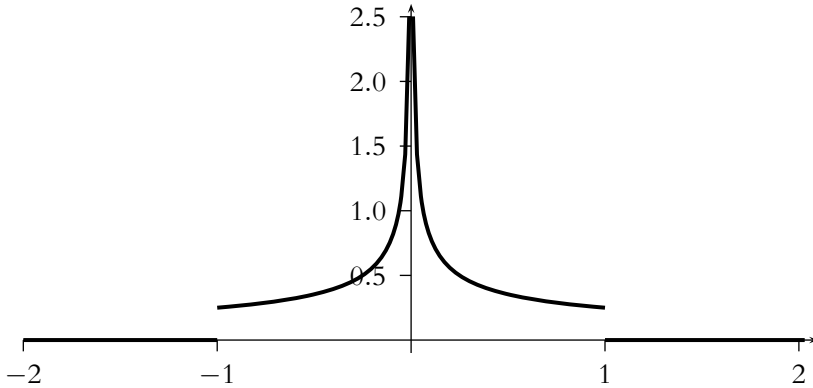
2. Voici un autre exemple qui montre qu'une densité peut ne pas être bornée : on pose

$$F(x) = 0 \text{ si } x < -1, F(x) = \frac{1 - \sqrt{-x}}{2} \text{ si } -1 \leq x \leq 0, F(x) = \frac{1 + \sqrt{x}}{2} \text{ si } 0 \leq x \leq 1 \text{ et } F(x) = 1 \text{ si } x > 1.$$

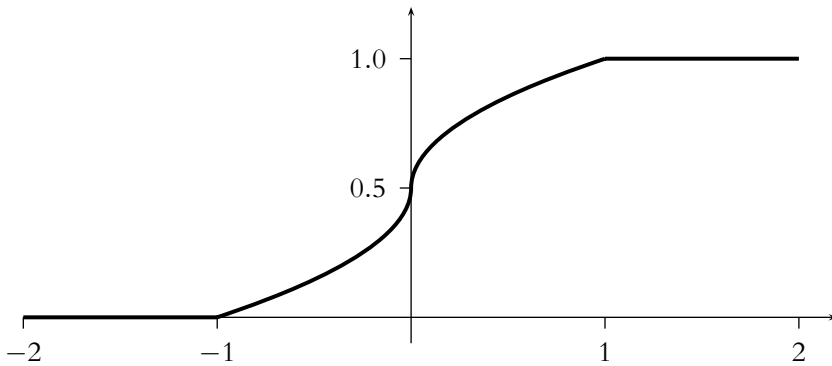
Une densité associée à f est, par exemple, la fonction définie par $f(x) = 0$ si $|x| > 1$,

$$f(x) = \frac{1}{4\sqrt{|x|}} \text{ si } 0 < |x| \leq 1 \text{ et } f(0) = 0 \text{ sinon, non bornée en 0.}$$

Le lecteur est invité à noter que F admet en -1 et 1 des dérivées à gauche et à droite qu'il calculera.



Exemple de densité non bornée définie par $f(x) = \frac{1}{4\sqrt{|x|}}$ sur $[-1, 1]$ privé de 0 et $f(x) = 0$ ailleurs



Fonction de répartition associée à la densité précédente : segments et arcs de paraboles

Proposition 1

Si f est une densité d'une variable aléatoire réelle X , l'ensemble de ses densités est celui des fonctions positives ou nulle g telles que $f(x) = g(x)$ sauf peut-être en un nombre fini de points.

Preuve

C'est clair. □

► **Remarque**

Dire qu'une densité est une fonction positive ou nulle « qui ne diffère de F' qu'au plus en un nombre fini de points » est impropre, la fonction F' n'étant pas définie sur \mathbb{R} entier ; par contre dire que « $f(x) = F'(x)$ sauf peut-être en un nombre fini de points » est correct.

Théorème 2

Soit f une densité d'une variable aléatoire réelle X de fonction de répartition F . On dispose alors des égalités

$$\int_{-\infty}^{+\infty} f(t) dt = 1, \quad F(x) = \int_{-\infty}^x f(t) dt \quad \text{pour tout réel } x.$$

Preuve

La première égalité découlant de la seconde en faisant tendre x vers $+\infty$, il suffit de démontrer cette dernière.

Soit x un réel et a un réel inférieur ou égal à x ; il existe une suite finie - peut-être vide - de réels (a_1, a_2, \dots, a_p) vérifiant les relations $a < a_1 < a_2 < \dots < a_p < x$ telle que F soit dérivable et de dérivée continue en dehors peut-être de a , de x et des points a_i .

Posant $a_0 = a$ et $a_{p+1} = x$, on dispose des relations

$$\int_u^v f(t) dt = \int_u^v F'(t) dt = F(v) - F(u) \leq F(a_{i+1}) - F(a_i)$$

sur tout segment $[u, v]$ inclus dans l'un des intervalles ouverts $]a_i, a_{i+1}[$ avec $0 \leq i \leq p$. La majoration ci-dessus et la positivité de f montrent que, faisant tendre u vers a_i et v vers a_{i+1} , le premier membre permet de définir une intégrale généralisée convergente; la continuité de F implique alors l'égalité

$$F(a_{i+1}) - F(a_i) = \lim_{\substack{u \rightarrow a_i \\ v \rightarrow a_{i+1}}} \int_u^v f(t) dt.$$

On sait que, l'on note cette égalité sous la forme

$$F(a_{i+1}) - F(a_i) = \int_{a_i}^{a_{i+1}} f(t) dt$$

et que le théorème de Chasles donne finalement

$$F(x) = F(a) + \int_a^x f(t) dt.$$

Il suffit alors de faire tendre a vers $-\infty$ pour obtenir l'égalité qu'il fallait démontrer. \square

Corollaire 1

Soient f une densité d'une variable aléatoire réelle X , a et b deux nombres réels tels que $a < b$. On dispose alors des égalités

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = \int_a^b f(t) dt$$

Corollaire 2

Soient f une densité d'une variable aléatoire réelle **positive** X de fonction de répartition F et x un nombre réel positif ou nul. On dispose alors des égalités

$$P(X \leq x) = F(x) = \int_0^x f(t) dt.$$

Corollaire 3

Soient f une densité d'une variable aléatoire réelle X de fonction de répartition F et x un nombre réel. On dispose alors des égalités

$$P(X \geq x) = 1 - F(x) = \int_x^{+\infty} f(t) dt.$$

Preuve

Il suffit de remarquer que les événements $[X \geq x]$ et $[X > x]$ ont même probabilité puisque $P(X = x) = 0$. \square

Théorème 3

Toute fonction réelle positive f continue sur \mathbb{R} sauf peut-être en un nombre fini de points et vérifiant l'égalité

$$\int_{-\infty}^{+\infty} f(t) dt = 1$$

est une densité d'une variable aléatoire réelle à densité X .

► **Remarque**

Attention à ne pas dire « la densité », puisque cette dernière n'est pas définie de manière unique.

Preuve

Le raisonnement ci-dessus implique que la fonction F définie par l'égalité $F(x) = \int_{-\infty}^x f(t) dt$ a un sens pour tout réel x , est dérivable sauf peut-être en un nombre fini de points, est croissante et admet respectivement 0 et 1 comme limites en $-\infty$ et $+\infty$. Pour que F soit une fonction de répartition d'une unique variable aléatoire réelle à densité, il suffit de prouver que F est continue à gauche : en fait elle est continue en tout point (même de discontinuité de f), ce qui prouve le théorème. \square

Théorème 4

Pour toute variable aléatoire X à densité f , la variable aléatoire réelle $Y = |X|$ admet pour densité la fonction g définie par $g(x) = 0$ pour $x \leq 0$ et sinon par l'égalité

$$g(x) = f(x) + f(-x).$$

Preuve

Nous savons que, pour tout réel $x > 0$, on dispose des égalités

$$G(x) = P(Y \leq x) = P(|X| \leq x) = F(x) - F(-x) + P(X = -x) = F(x) - F(-x).$$

Sauf peut-être en un nombre fini de points, le dernier membre est dérivable et vérifie les égalités $G'(x) = F'(x) + F'(-x) = f(x) + f(-x)$. \square

1.2 Premier théorème de transfert

Moyennant certaines précautions (pas toujours observées dans la pratique ou les textes officiels), il est possible de calculer une densité pour une variable aléatoire $Y = \varphi(X)$ où X est elle-même à densité. Nous donnerons ci-dessous la forme la plus simple

de ces résultats, connus sous le nom de *premier théorème de transfert* (le second concerne l'espérance mathématique de Y).

Théorème 5

Soit X une variable aléatoire réelle à densité f et φ une fonction de classe \mathcal{C}^1 de $X(\Omega)$ dans \mathbb{R} admettant en tout point une dérivée strictement positive.

Alors $Y = \varphi(X) = \varphi \circ X$ est une variable aléatoire réelle admettant une densité nulle en dehors de l'intervalle $\varphi(X(\Omega))$ et donnée sur l'intérieur de cet intervalle par l'égalité

$$g(y) = (f \circ \varphi^{-1}(y)) \frac{d}{dy} \varphi^{-1}(y).$$

Si φ admet en tout point une dérivée strictement négative, il faut remplacer $\frac{d}{dy} \varphi^{-1}(y)$ par son opposé, qui est aussi sa valeur absolue.

► **Remarques**

- Le résultat se retient plus facilement sous la forme $g(y) dy = f(x) dx$ puisque

$$g(y) = f(x) \frac{dx}{dy}.$$

L'égalité $g(y) dy = f(x) dx$ peut d'ailleurs se retrouver très simplement en écrivant que les fonctions de répartition F et G vérifient l'identité

$$F(x) = G(y) \quad [= G(\varphi(x))]$$

et il ne reste plus qu'à différencier les deux membres de cette égalité pour trouver la relation désirée.

- Le programme affirme qu'il suffit que φ soit strictement monotone et de classe \mathcal{C}^1 . C'est inexact comme le montre l'exemple de la fonction définie par $\varphi(t) = t - \sin t$, qui se comporte en une infinité de points comme la fonction $x \mapsto x^3$, en lesquels la dérivée de φ^{-1} n'existe pas.

Preuve

Puisque la dérivée de φ est strictement positive, la fonction φ est strictement croissante sur son intervalle de définition.

Pour tous les réels $y \in Y(\Omega)$ et $x = \varphi^{-1}(y)$ tel que $\varphi(x) = y$, l'ensemble $\{\omega \mid Y(\omega) \leq y\}$ est égal à l'ensemble $\{\omega \mid \varphi \circ X(\omega) \leq y\} = \{\omega \mid X(\omega) \leq x\} = [X \leq x]$. Par suite Y est une variable aléatoire.

Alors φ^{-1} admet en y une dérivée égale à $\frac{d}{dy} \varphi^{-1}(y) = \frac{1}{\varphi'(x)}$, d'où

$$F_Y(y) = P(Y \leq y) = P(X \leq x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^{\varphi^{-1}(y)} f(t) dt.$$

Le calcul de la dérivée par rapport à y de cette intégrale par rapport à sa limite supérieure donne le résultat annoncé.

Si φ' est strictement négative, les relations ci-dessus deviennent

$$F_Y(y) = P(Y \leq y) = P(X \geq x) = \int_x^{+\infty} f(t) dt = \int_{\varphi^{-1}(y)}^{+\infty} f(t) dt.$$

Le calcul de la dérivée par rapport à y de cette intégrale par rapport à sa limite inférieure donne le résultat annoncé. □

Corollaire 4

Soit X une variable aléatoire réelle à densité f et (a, b) un couple de réels vérifiant l'inégalité $a > 0$. Alors $Y = aX + b$ est une variable aléatoire réelle admettant sur \mathbb{R} une densité donnée par l'égalité

$$g(y) = \frac{1}{a} f\left(\frac{y-b}{a}\right).$$

► **Remarques**

- Si $a < 0$, il faut remplacer $g(y)$ par son opposé.
- Si $a = 0$, Y est la variable aléatoire certaine b , discrète mais n'admet aucune densité puisque la fonction de répartition, constante sur les deux demi-droites $]-\infty, b[$ et $]b, +\infty[$ où elle vaut respectivement 0 et 1, est discontinue en b .
- Si s est un réel strictement positif, une densité en y de la variable aléatoire définie par $Y = \frac{X-m}{s}$ (ou encore $X = sY + m$) est $sf(sy + m)$. Cette remarque est par exemple utile pour l'étude des variables normales de Laplace-Gauss.

Proposition 2

Soit X une variable aléatoire réelle à densité f . Alors $Y = X^2$ est une variable aléatoire réelle admettant une densité nulle sur \mathbb{R}_- et donnée pour $y > 0$ par l'égalité

$$g(y) = \frac{1}{2\sqrt{y}} (f(\sqrt{y}) + f(\sqrt{-y})).$$

Preuve

Il n'y a rien à prouver si $y \leq 0$. Sinon, posant $x = \sqrt{y}$, on dispose des égalités

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq x^2) = P(|X| \leq x) = \int_{-x}^x f(t) dt \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f(t) dt = \int_0^{\sqrt{y}} f(t) dt - \int_0^{-\sqrt{y}} f(t) dt. \end{aligned}$$

Le calcul des dérivées par rapport à y de ces intégrales par rapport à leurs bornes supérieures donne le résultat annoncé. ◻

► **Remarques**

- Le cas $Y = |X|$ a été traité plus haut ; on dispose alors pour $x \geq 0$ de l'égalité

$$F_{|X|}(x) = \int_{-x}^x f(t) dt.$$

- On peut étudier de même les variables définies par $Y = aX^n$ avec $n \in \mathbb{N}$ et, avec un peu plus de précautions, $n \in \mathbb{Z}$.

1.3 Somme de deux variables indépendantes à densité

Conformément au programme, nous admettrons le théorème suivant.

Théorème 6

Soient X et Y deux variables aléatoires réelles **indépendantes** admettant respectivement des densités f et g . Si l'on note

$$h(z) = \int_{-\infty}^{+\infty} f(t)g(z-t) dt$$

dont on suppose l'existence et la continuité par rapport à z sauf peut-être en un nombre fini de points, alors la fonction h ainsi définie et éventuellement complétée par quelques valeurs positives arbitraires est une densité de la variable aléatoire réelle définie par la somme $Z = X + Y$.

C'est notamment le cas lorsque l'une des deux densités f ou g est bornée.

► Remarques

- Puisque f et g sont positives, l'intégrale définissant $h(z)$ est absolument convergente dès qu'elle est convergente.
- Il est possible d'échanger les rôles de f et de g dans cette égalité comme le montre le changement de variable $u = z - t$.
- Une fonction h ainsi définie s'appelle **produit de convolution** de f et g , noté $f * g$. Toutefois, h pouvant être remplacé par n'importe quelle fonction qui ne diffère d'elle qu'en un nombre fini de points, ce produit n'est pas unique ; il s'agit encore ici d'une classe d'équivalence de fonctions plutôt que d'une fonction particulière.
- On pourra comparer cette formule à celle qui donne la loi de la somme de deux variables aléatoires discrètes indépendantes X et Y comme somme d'une série de la forme

$$P(Z = z) = P(X + Y = z) = \sum_{i \in I} P(X = x_i) P(Y = z - x_i).$$

Conformément à une remarque intuitive faite plus haut, on peut interpréter cette égalité en disant que, pour un petit accroissement Δt de la variable t , $f(t)g(z-t)\Delta t$ est, approximativement, la probabilité pour que X reste comprise entre t et $t + \Delta t$ et que $X + Y$ prenne une valeur voisine de z .

- Dans certains cas de forte régularité de f et de g , on a même une expression de la fonction de répartition H de $Z = X + Y$

$$H(z) = P(X + Y \leq z) = \int_{-\infty}^{+\infty} f(t)G(z-t) dt$$

(ou une autre, équivalente, où f et g échangent leurs rôles). Une justification de cette remarque passerait par une dérivation par rapport à z sous un signe intégral, opération risquée qui n'est légitime que sous des « bonnes » conditions.

- Le cadre naturel d'étude des fonctions h et H est celui des *intégrales doubles*, qui est largement en dehors du programme de la classe. Avec ce formalisme, la probabilité pour que $X + Y$ prenne une valeur dans une partie A suffisamment régulière du plan \mathbb{R}^2 est donnée par l'égalité

$$P(X + Y \in A) = \iint_A f(x)g(y) dx dy.$$

On peut retrouver ainsi les valeurs de $H(z)$ et de $h(z)$ en prenant pour A le demi-plan défini par $x + y \leq z$.

Corollaire 5

Si de plus X et Y sont à valeurs positives ou nulles, alors h est définie sur \mathbb{R}_+ par l'égalité

$$h(z) = \int_0^z f(t)g(z-t) dt$$

et est nulle sur \mathbb{R}_- .

Preuve

En effet, les densités f et g sont nulles sur \mathbb{R}_- . □

Cette remarque sera par exemple utile dans l'étude des lois gamma ou des lois normales.

2. Moments d'une variable aléatoire à densité

Les concepts d'espérance et de variance, déjà rencontrés au niveau des variables aléatoires réelles discrètes, possèdent leurs analogues pour les variables à densité.

Définition 2

Pour tout entier strictement positif n et toute variable aléatoire réelle X admettant une densité f , le **moment d'ordre n** de X est donné par l'égalité

$$M_n(X) = \int_{-\infty}^{+\infty} t^n f(t) dt$$

si cette intégrale est **absolument convergente**.

► **Remarques**

- Cette définition a un sens dans la mesure où remplacer f par une autre densité de X ne modifie, ni la nature de la série, ni sa somme éventuelle.
- L'existence de $M_n(X)$ équivaut à celle de $M_n(|X|)$, égale à $\int_{-\infty}^{+\infty} |t|^n f(t) dt$, et vérifiée s'il y a lieu l'inégalité $|M_n(X)| \leq M_n(|X|)$, qui est une égalité si n est pair.
- On pourra comparer cette valeur à celle du moment d'ordre n d'une variable aléatoire réelle discrète X d'image au plus dénombrable $\{x_i \mid i \in I\}$ qui vérifie l'égalité

$$M_n(X) = \sum_{i \in I} x_i^n P(X = x_i).$$

Proposition 3

Soient n un entier strictement positif, X et Y deux variables aléatoires réelles admettant respectivement comme densités f et g vérifiant l'inégalité $f \leq g$. Alors, pour qu'il existe $M_n(X)$, il suffit qu'il existe $M_n(Y)$.

Preuve

Supposons en effet l'existence de $M_n(Y)$, donc celle de $M_n(|Y|)$; pour tout couple (A, B) de réels positifs ou nul, on dispose des inégalités

$$\int_{-A}^B |t|^n f(t) dt \leq \int_{-A}^B |t|^n g(t) dt \leq M_n(|Y|)$$

et faire tendre indépendamment A et B vers $+\infty$ implique l'existence de $M_n(|X|)$, donc celle de $M_n(X)$. (En fait $f = g$ presque partout). \square

2.1 Espérance mathématique

Définition 3

L'espérance mathématique éventuelle d'une variable aléatoire réelle X admettant une densité f est

$$E(X) = M_1(X) = \int_{-\infty}^{+\infty} t f(t) dt$$

si l'intégrale est absolument convergente.

► Remarques

- L'espérance n'existe pas toujours : c'est le cas pour la *densité de Cauchy* définie par $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$.
- On pourra comparer cette définition à celle qui donne l'espérance d'une variable aléatoire réelle finie

$$E(X) = \sum_{i=1}^p x_i P(X = x_i)$$

où la sommation porte sur les images x_i de X .

Théorème 7

Soit X une variable aléatoire réelle admettant une fonction **paire** f pour densité. Alors l'existence de l'espérance mathématique $E(X)$ de X équivaut à celle de l'intégrale

$$\int_0^{+\infty} t f(t) dt \text{ et l'on dispose de l'égalité}$$

$$E(X) = 0.$$

Preuve

Il suffit de remarquer que, pour tout couple (A, B) de réels positifs ou nuls, on dispose des égalités

$$\int_{-A}^B |t| f(t) dt = \int_0^B t f(t) dt + \int_0^A t f(-t) dt = \int_0^B t f(t) dt + \int_0^A t f(t) dt,$$

$$\int_{-A}^B t f(t) dt = \int_0^B t f(t) dt - \int_0^A t f(-t) dt = \int_0^B t f(t) dt - \int_0^A t f(t) dt$$

et de faire tendre indépendamment A et B vers $+\infty$. \square

Théorème 8

Soient X une variable aléatoire réelle admettant une densité f et une espérance mathématique $E(X)$, et a un réel. Alors $Y = X + a$ admet une espérance mathématique vérifiant

$$E(Y) = E(X + a) = E(X) + a.$$

Preuve

La fonction de répartition de Y est la fonction G définie par

$$G(y) = P(Y \leq y) = P(X \leq y - a) = F(y - a).$$

Par suite une densité de Y est la fonction g définie par $g(y) = f(y - a)$.

Pour tout couple (A, B) de réels positifs ou nuls, le changement de variable $y = a + u$ montre que l'on dispose des relations

$$\begin{aligned} \int_{-A}^B |y| g(y) dy &= \int_{-A}^B |y| f(y - a) dy = \int_{-A-a}^{B-a} |a + u| f(u) du \\ &\leq \int_{-A-a}^{B-a} (|a| + |u|) f(u) du \leq |a| + \int_{-\infty}^{+\infty} |u| f(u) du = |a| + E(|X|) \end{aligned}$$

et il suffit de faire tendre A et B indépendamment vers $+\infty$ pour voir que Y admet une espérance. Celle-ci est alors la limite de

$$\int_{-A}^B y g(y) dy = \int_{-A}^B y f(y - a) dy = \int_{-A-a}^{B-a} (a + u) f(u) du$$

c'est-à-dire $a + E(X)$. □

Théorème 9

Soient X une variable aléatoire réelle admettant une densité f et une espérance mathématique $E(X)$, et a un réel. Alors $Y = aX$ admet une espérance mathématique vérifiant

$$E(Y) = E(aX) = aE(X).$$

Preuve

C'est clair si $a = 0$.

Soit d'abord $a > 0$; Y admet comme densité $g(y) = \frac{1}{a} f\left(\frac{y}{a}\right)$ d'où, par le changement de variable $t = \frac{y}{a}$, les égalités

$$\begin{aligned} \int_{-\infty}^{+\infty} |t| g(t) dt &= \frac{1}{a} \int_{-\infty}^{+\infty} |t| f\left(\frac{t}{a}\right) dt = a \int_{-\infty}^{+\infty} |u| f(u) du = aE(|X|), \\ E(aX) = E(Y) &= \int_{-\infty}^{+\infty} t g(t) dt = \frac{1}{a} \int_{-\infty}^{+\infty} t f\left(\frac{t}{a}\right) dt = a \int_{-\infty}^{+\infty} u f(u) du = aE(X). \end{aligned}$$

Le cas $a < 0$ est très peu différent : simplement $g(y)$ doit être multiplié par -1 et les limites de l'intégrale généralisée changées en leurs opposées, ce qui donne le même résultat. □

► **Remarque**

Si l'emploi direct d'un changement de variable dans une intégrale généralisée apparaît comme trop brutal, il est possible d'opérer plus prudemment comme dans la preuve du théorème précédent.

Proposition 4

Si X est une variable aléatoire réelle à densité admettant un moment d'ordre n , alors pour tout réel a la variable aléatoire aX admet un moment d'ordre n donné par l'égalité

$$M_n(aX) = a^n M_n(a).$$

Preuve

Il suffit de reprendre la démonstration ci-dessus du cas particulier $n = 1$ qui s'étend aussitôt en distinguant encore les trois cas $a = 0$, $a > 0$ et $a < 0$. □

Corollaire 6

Soient X une variable aléatoire réelle admettant une densité f et une espérance mathématique $E(X)$, et (a, b) un couple de réels. Alors $Y = aX + b$ admet une espérance mathématique vérifiant

$$E(Y) = E(aX + b) = aE(X) + b.$$

Preuve

Cela résulte aussitôt des deux théorèmes précédents. \square

Nous admettrons la propriété suivante, qui est très importante.

Théorème 10

Soient X et Y deux variables aléatoires réelles à densité **indépendantes** admettant des espérances mathématiques. Alors $Z = XY$ admet une espérance mathématique vérifiant

$$E(Z) = E(XY) = E(X)E(Y).$$

► Remarque

Cela revient à démontrer l'égalité formelle

$$\int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} xy f(x)g(y) dy \right] dx = \left[\int_{-\infty}^{+\infty} x f(x) dx \right] \left[\int_{-\infty}^{+\infty} y f(y) dx \right].$$

► Remarque

Rappelons que, dans le chapitre consacré aux variables aléatoires réelles et aux variables aléatoires « quelconques » nous avons aussi admis que, si X (et éventuellement Y) a (ont) une (des) espérance(s) mathématique(s), et si a est un réel, on dispose alors des égalités

$$E(aX) = aE(X), \quad E(X + Y) = E(X) + E(Y)$$

cela étant en particulier vrai dans le cadre des variables à densité.

2.2 Second théorème de transfert

Le théorème suivant est une extension, assez délicate, d'un théorème de transfert déjà étudié dans le cas d'une variable discrète.

Théorème 11

Soient X une variable aléatoire réelle à densité f dont l'image $X(\Omega)$ est un intervalle I de \mathbb{R} et φ une fonction de I dans \mathbb{R} , continue sauf en un nombre fini de points.

Alors $Y = \varphi(X) = \varphi \circ X$ est une variable aléatoire réelle. Elle admet une espérance mathématique si, et seulement si, l'intégrale

$$\int_I |\varphi(t)|f(t) dt$$

est convergente. Cette espérance est alors donnée par l'égalité

$$E(\varphi(X)) = \int_I \varphi(t)f(t) dt.$$

Preuve

Si I admet a et b comme extrémités avec $-\infty \leq a < b \leq +\infty$, toute intégrale $\int_I h(t) dt$ est une abréviation pour $\int_a^b h(t) dt$.

Comme nous y invite le programme, nous ne démontrerons ce résultat que pour φ de classe C^1 avec $\varphi'(x) > 0$ en tout point de I (d'ailleurs la preuve du théorème général dépasse clairement le niveau de ce livre).

D'après la démonstration du premier théorème de transfert, nous savons qu'alors Y est effectivement une variable aléatoire dont nous connaissons la densité sur l'intérieur de I

$$g(y) = (f \circ \varphi^{-1}(y)) \frac{d}{dy} \varphi^{-1}(y) = f(x) \frac{dx}{dy}.$$

L'image $J = \varphi(I)$ de l'intervalle I par φ est ici un intervalle de \mathbb{R} , dont nous noterons - abusivement - $\varphi(a)$ et $\varphi(b)$ les extrémités.

Soient un réel c vérifiant les inégalités $-\infty \leq a < c < b \leq +\infty$, $d = \varphi(c)$, x un point de I et $y = \varphi(x)$. Le changement de variables monotone $u = \varphi(t)$, légitimé par les conditions fortes imposées à φ , justifie l'égalité

$$\int_d^y |u| g(u) du = \int_{\varphi(c)}^{\varphi(x)} |u| g(u) du = \int_c^x |\varphi(t)| f(t) dt.$$

L'existence de $E(Y)$ équivaut à l'intégrabilité de $|u| g(u)$ sur $[d, y]$ lorsque y tend vers $\varphi(a)$ puis vers $\varphi(b)$, et l'espérance est alors égale à l'intégrale $\int_I u g(u) du$.

L'égalité des intégrales sur $[d, y]$ et $[c, x]$ implique que cette existence équivaut à l'intégrabilité de $|\varphi(t)| f(t)$ sur $[c, x]$ lorsque x tend vers a puis vers b , c'est-à-dire à la convergence absolue de l'intégrale $\int_I \varphi(t) f(t) dt$. Dans ce cas, il en résulte l'égalité cherchée

$$E(\varphi(X)) = \int_a^b \varphi(t) f(t) dt. \quad \square$$

► **Remarques**

- La démonstration ci-dessus introduit, un instant, les symboles $\varphi(a)$ et $\varphi(b)$, qui ne sont pas nécessairement des nombres (pas plus d'ailleurs que a et b eux-mêmes). Le lecteur la complétera sans difficulté pour la rendre totalement correcte ; mais il a semblé que l'introduction de ces deux symboles aiderait à la compréhension intuitive du mécanisme de la preuve.
- Le cas $\varphi'(x) < 0$ en tout point se traite de la même manière.
- Formellement, ce théorème s'exprime comme un simple changement de variable dans une intégrale.
- Le lecteur pourra reprendre la démonstration du cas particulier de ce théorème donnée page ?? où X est une variable aléatoire réelle discrète, et observer la ressemblance formelle entre les deux résultats, l'un exprimé sous forme de série, l'autre d'intégrale.

Nous avons admis le théorème suivant, donné dans un cadre plus général.

Théorème 12

Soient X et Y deux variables aléatoires réelles admettant des espérances mathématiques et telles que $P(X > Y) = 0$ (on dit que $X \leq Y$ **presque sûrement**).

On dispose alors de l'inégalité $E(Y) \leq E(X)$.

► **Remarque**

On peut noter que ce résultat est presque évident si $X = 0$: il résulte alors des relations $P(Y < 0) = 0$ et

$$E(Y) = \int_{-\infty}^{+\infty} t g(t) dt = \int_0^{+\infty} t g(t) dt \geq 0.$$

Le cas général se ramène alors à celui-là en admettant, comme nous l'avons fait plus haut, que l'espérance de la différence $Y - X$ est égale à la différence de leurs espérances.

2.3 Variance et écart-type

Définir une variance pour une variable à densité consiste à rester près du cas simple des variables discrètes, en étendant convenablement la relation de *Kœnig-Huygens*

Définition 4

Une variable aléatoire réelle à densité X est dite admettre une **variance** $V(X)$ si elle admet une espérance mathématique $E(X) = m$ et un moment $M_2(X)$ d'ordre deux.

Cette variance est alors donnée par la formule de Kœnig-Huygens

$$V(X) = M_2(X) - E(X)^2.$$

Théorème 13

Si X est une variable aléatoire réelle à densité admettant une variance $V(X)$, alors pour tout réel a la variable aléatoire aX admet une variance donnée par l'égalité

$$V(aX) = a^2 V(X).$$

Preuve

Cela résulte des égalités $E(aX) = aE(X)$ et $M_2(aX) = a^2 M_2(X)$. □

Théorème 14

Soit X une variable aléatoire réelle admettant une densité f . Alors X admet une variance si, et seulement si, elle admet une espérance mathématique $E(X) = m$ et si l'intégrale

$$\int_{-\infty}^{+\infty} (t - m)^2 f(t) dt$$

est convergente. On dispose alors de l'égalité

$$V(X) = \int_{-\infty}^{+\infty} (t - m)^2 f(t) dt$$

Preuve

Supposons que m et l'intégrale existent. On note S sa valeur. Soient A et B des réels positifs ou nuls. On dispose des relations

$$\begin{aligned} \int_{-A}^B t^2 f(t) dt &= \int_{-A}^B (t - m)^2 f(t) dt + 2m \int_{-A}^B t f(t) dt - m^2 \int_{-A}^B f(t) dt \\ &\leq \int_{-A}^B (t - m)^2 f(t) dt + 2|m| \int_{-A}^B |t| f(t) dt \leq S + 2|m| E(|X|) \end{aligned}$$

qui prouvent l'existence de $M_2(X)$.

Il ne reste plus qu'à faire tendre indépendamment A et B vers $+\infty$ dans les égalités

$$\int_{-A}^B t^2 f(t) dt = \int_{-A}^B (t - m)^2 f(t) dt + 2m \int_{-A}^B t f(t) dt - m^2 \int_{-A}^B f(t) dt$$

pour obtenir la valeur annoncée de $V(X)$ puisque

$$V(X) + m^2 = M_2(X) = \int_{-\infty}^{+\infty} (t - m)^2 f(t) dt + 2m^2 - m^2.$$

Inversement, si la variance existe, il en est de même de S et $S = V(X)$, par des calculs pratiquement identiques à ceux qui précèdent, la seule différence résultant dans la majoration légèrement plus complexe

$$\int_{-A}^B (t - m)^2 f(t) dt \leq M_2(X) + 2|m|E(|X|) + m^2. \quad \square$$

► **Remarques**

- L'intégrale donnant $V(X)$ est absolument convergente dès qu'elle est convergente.
- On pourra comparer cette valeur à celle d'une variable aléatoire réelle discrète X d'image au plus dénombrable $\{x_i \mid i \in I\}$ qui vérifie l'égalité

$$V(X) = \sum_{i \in I} (x_i - m)^2 P(X = x_i).$$

Corollaire 7

La variance d'une variable aléatoire réelle X admettant une espérance mathématique est, si elle existe, l'espérance mathématique de la variable aléatoire réelle $(X - E(X))^2$.

Preuve

Cela résulte aussitôt de la valeur de $V(X)$ comme intégrale généralisée. □

Théorème 15

Si X est une variable aléatoire réelle à densité admettant une variance $V(X)$, alors pour tout couple (a, b) de réels la variable aléatoire $aX + b$ admet une variance donnée par l'égalité

$$V(aX + b) = a^2 V(X).$$

Preuve

Soit d'abord $a = 1$; alors les égalités $E(X + b) = E(X) + b$ et $V(X) = E(X - E(X))^2$ montrent que $V(X + b) = V(X)$ (*invariance par translation*).

Si $a = 0$, le résultat est clair car $aX + b$ est alors une variable certaine (mais il faut noter qu'elle n'est plus à densité mais discrète).

Si enfin $a \neq 0$, il suffit d'écrire $Y = a \left(X + \frac{b}{a} \right)$ pour pouvoir conclure. □

Théorème 16

Toute variance de variable aléatoire réelle à densité est strictement positive.

Preuve

Supposons qu'il existe une variable aléatoire réelle à densité X admettant une variance nulle, c'est-à-dire vérifiant les égalités

$$V(X) = \int_{-\infty}^{+\infty} (x - m)^2 f(t) dt = 0.$$

Soit (a, b) un couple de réels vérifiant $a < m < b$. On en déduit

$$0 \leq F(a) = \int_{-\infty}^a f(t) dt \leq \frac{1}{(m-a)^2} \int_{-\infty}^a (x-m)^2 f(t) dt \leq \frac{V(X)}{(m-a)^2} = 0$$

d'où $F(a) = 0$.

On montre de la même manière que $F(b) = 1$, ce qui contredit la continuité de F en m . \square

Définition 5

Toute variance $V(X)$ de variable aléatoire réelle à densité admet une racine carrée notée strictement positive, appelé écart-type de X , $\sigma(X) = \sqrt{V(X)}$.

Proposition 5

Si X est une variable aléatoire réelle à densité admettant un écart-type $\sigma(X)$, alors pour tout couple (a, b) de réels la variable aléatoire $aX + b$ admet un écart-type donné par l'égalité

$$\sigma(aX + b) = |a| \sigma(X).$$

Preuve

C'est un simple corollaire du calcul de $V(aX + b)$ fait ci-dessus. \square

► Remarque

Rappelons qu'à cause du cas particulier $a = 0$, on ne peut pas dire que $aX + b$ est une variable à densité.

Exercice 1.

Soit X une variable aléatoire réelle admettant une densité f et une variance $V(X)$. Calculer

$$\frac{1}{2} \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} (x - y)^2 f(x) f(y) dx \right] dy.$$

Le lecteur est prié de vérifier que chacune des intégrales figurant dans la solution de cet exercice est bien absolument convergente.

Cela dit, un calcul formel donne les égalités

$$\begin{aligned} \int_{-\infty}^{+\infty} (x - y)^2 f(x) f(y) dx &= \int_{-\infty}^{+\infty} x^2 f(x) f(y) dx \\ -2 \int_{-\infty}^{+\infty} xy f(x) f(y) dx + \int_{-\infty}^{+\infty} y^2 f(x) f(y) dx &= (M_2(X) - 2ym + y^2) f(y), \\ \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} (x - y)^2 f(x) f(y) dx \right] dy &= M_2(X) - 2m^2 + M_2(X) = 2V(X). \end{aligned}$$

► **Remarque**

On pourra comparer cette propriété à celle de la variance d'une variable aléatoire discrète X d'image au plus dénombrable $\{x_i \mid i \in I\}$ qui vérifie l'égalité

$$V(X) = \frac{1}{2} \sum_{(i,j) \in I^2} (x_i - x_j)^2 P(X = x_i) P(X = x_j).$$

2.4 Inégalités de Markov et Bienaymé-Tchebychev

Il s'agit ici encore d'extensions d'inégalités théoriques importantes (même si elles sont en général inopérantes dans la pratique) rencontrées en première année dans le cadre des variables discrètes.

Théorème 17 (Inégalité de Markov)

Soit X une variable aléatoire réelle **positive** admettant une densité f et une variance $E(X)$. Alors, pour tout réel $a > 0$, on dispose de l'inégalité

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Preuve

Cela résulte des relations

$$E(X) = \int_0^{+\infty} t f(t) dt \geq \int_a^{+\infty} a f(t) dt = a (1 - F(a)) = a P(X > a) = a P(X \geq a). \quad \square$$

Théorème 18 (Inégalité de Bienaymé-Tchebychev)

Soit X une variable aléatoire réelle admettant une densité f , une espérance mathématique $E(X) = m$ et une variance $V(X)$.

Alors, pour tout réel $a > 0$, on dispose de l'inégalité

$$P(|X - m| \geq a) \leq \frac{V(X)}{a^2}.$$

Preuve

Il suffit de substituer dans l'inégalité de Markov le couple $(|X - m|^2, a^2)$ au couple (X, a) . □

Corollaire 8

Soit X une variable aléatoire réelle admettant une densité f , une espérance mathématique $E(X) = m$ et un écart-type $\sigma(X)$.

Alors, pour tout réel $\lambda > 0$, on dispose de l'inégalité

$$P(|X - m| \geq \lambda \sigma(X)) \leq \frac{1}{\lambda^2}.$$

Preuve

Il suffit de poser $a = \lambda \sigma$ dans l'inégalité de Bienaymé-Tchebychev. □

► **Remarque**

Ainsi, pour $\lambda = 3$, la probabilité de s'écarter de plus de trois écart-types est inférieure à 0,12. En fait, elle est généralement encore beaucoup plus petite (par exemple de l'ordre de 0,003 pour une loi normale), ce qui fait dire que, pour les lois usuelles, les valeurs prises par la variable X sont pratiquement toutes dans le segment $[m - 3\sigma, m + 3\sigma]$.

2.5 Variable centrée réduite associée à une variable à densité

Définition 6

Soit X une variable aléatoire réelle à densité admettant une espérance mathématique $E(X) = \mu$ et un écart-type $\sigma = \sqrt{V(X)}$.

Alors $Y = \frac{1}{\sigma}(X - \mu)$ est une variable aléatoire réelle à densité, appelée variable **centrée réduite** associée à X .

Cette définition est rendue possible par le corollaire 4, qui donne d'ailleurs la densité de Y en fonction de celle de X , et le théorème 16.

3. Les lois usuelles

On trouvera ci-dessous une description de six des principales lois de variables aléatoires réelles à densité. Il en existe bien d'autres : lois de Pareto et lois Beta (cas particuliers : lois de l'arc sinus et lois de Yule), loi lognormale, loi de Weibull (cas particulier : loi de Rayleigh), loi de Gompertz, loi de Gumbel, loi de Student, loi logistique, loi de Maxwell, loi de Lorentz, loi de Khintchine... dont on trouvera les propriétés dans les ouvrages spécialisés notamment destinés aux statisticiens qui en font grand usage.

3.1 La loi uniforme

La loi uniforme sur un segment est la plus simple de toutes les lois de variables aléatoires réelles à densité.

Soient a et b deux réels vérifiant l'inégalité $a < b$.

Rappelons l'égalité $\int_a^b dx = b - a$.

Définition 7

Une variable aléatoire réelle X admettant pour densité une fonction f égale à $\frac{1}{b-a}$ sur $[a, b]$ et nulle ailleurs est dite suivre **la loi uniforme sur $[a, b]$** , ce que l'on note

$$X \hookrightarrow \mathcal{U}[a, b].$$

► **Remarque**

L'existence d'une telle loi est justifiée par le rappel précédent.

Proposition 6

Si g est une densité d'une variable aléatoire réelle suivant la loi $\mathcal{U}[0, 1]$, alors la fonction f définie par

$$f(x) = \frac{1}{b-a} g\left(\frac{x-a}{b-a}\right)$$

est une densité d'une variable aléatoire réelle suivant la loi $\mathcal{U}[a, b]$.

Preuve

C'est clair par le changement de variable $x \mapsto \frac{x-a}{b-a}$. □

► **Remarque**

On dispose donc des relations équivalentes

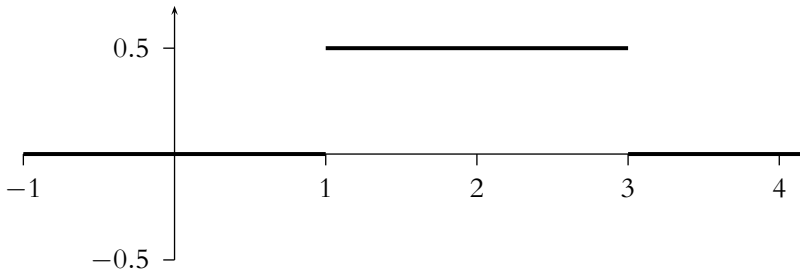
$$\begin{aligned} X \hookrightarrow \mathcal{U}[0, 1] &\iff a + (b-a)X \hookrightarrow \mathcal{U}[a, b], \\ Y \hookrightarrow \mathcal{U}[u, v] &\iff \frac{1}{b-a}(X-a) \hookrightarrow \mathcal{U}[0, 1]. \end{aligned}$$

Proposition 7

Si f est la densité donnée dans la définition 7 ci-dessus d'une variable aléatoire réelle suivant la loi $\mathcal{U}[a, b]$, son graphe est formé d'une demi-droite $] -\infty, a[\times \{0\}$, du segment $[a, b] \times \left\{ \frac{1}{b-a} \right\}$ et de la demi-droite $] b, +\infty[\times \{0\}$.

Preuve

C'est clair. □



Densité de la loi $\mathcal{U}[1, 3]$

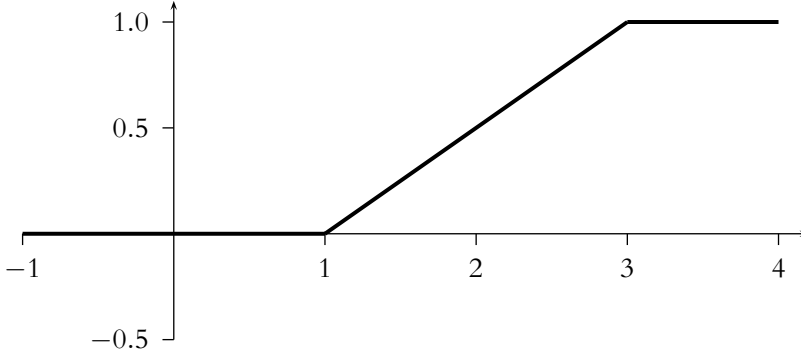
Théorème 19

La fonction de répartition d'une variable aléatoire réelle suivant la loi uniforme $\mathcal{U}[a, b]$ vaut 0 sur $] -\infty, a[$, 1 sur $] b, +\infty[$ et vérifie, pour tout réel $x \in [a, b]$, l'égalité

$$F(x) = P(X \leq x) = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}.$$

Preuve

Cela traduit la relation entre densité et fonction de répartition. □



Fonction de répartition de la loi $\mathcal{U}[1, 3]$

Théorème 20

Toute variable aléatoire réelle X suivant une loi uniforme de paramètre $[a, b]$ admet des moments de tous ordres, donnés par les égalités

$$M_n(X) = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}$$

Preuve

En effet

$$M_n(X) = \int_a^b x^n \frac{1}{b-a} dx = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}$$
□

Théorème 21

Toute variable aléatoire réelle X suivant une loi uniforme de paramètre $[a, b]$ admet une espérance mathématique égale à $\frac{a+b}{2}$.

Preuve

Cela résulte de la formule précédente puisque $M_1(X) = \frac{a+b}{2}$. Cela résulte d'ailleurs aussi de l'axe de symétrie d'équation $x = \frac{a+b}{2}$ du graphe de f . □

Théorème 22

Toute variable aléatoire réelle X suivant une loi uniforme de paramètre $[a, b]$ admet une variance égale à $\frac{(b-a)^2}{12}$.

Preuve

Cela résulte de la même formule, puisque

$$V(X) = M_2(X) - E(X)^2 = \frac{1}{12}(b^2 - 2ab + a^2) = \frac{(b-a)^2}{12}. \quad \square$$

Exercice 2. (Maximum et minimum de variables uniformes indépendantes)

Soient (X_1, X_2, \dots, X_n) une famille finie de variables aléatoires réelles à densité mutuellement indépendantes suivant la loi uniforme sur $[a, b]$, et $S = \sup_{k \in \llbracket 1, n \rrbracket} (X_k)$, $I = \inf_{k \in \llbracket 1, n \rrbracket} (X_k)$. Étudier les fonctions de répartition des deux variables aléatoires discrètes S et I .

- Commençons par le cas particulier $[0, 1]$. La démonstration est exactement celle du cas analogue des variables discrètes (exercice 1 du chapitre consacré à ces variables). Pour tout x de $[0, 1]$,

$$F_S(x) = P(S \leq x) = P(\{\omega \mid \forall i \in \llbracket 1, n \rrbracket \quad X_i(\omega) \leq x\}) = \prod_{i=1}^n P(X_i \leq x) = x^n.$$

Il en va exactement de même pour la variable aléatoire I , toujours en passant par l'événement contraire $[I > x]$

$$F_I(x) = P(I \leq x) = 1 - (1 - x)^n.$$

► **Remarque**

L'inclusion $[S \leq x] \subset [I \leq x]$ conduit à l'inégalité $x^n + (1 - x)^n \leq 1$ valable sur $[0, 1]$, inégalité que redonne aussitôt le développement de $(x + (1 - x))^n$ par la formule du binôme.

- Le cas général est alors immédiat en utilisant le changement de variable $t \mapsto a + (b - a)t$. On trouve ainsi sur $[a, b]$

$$F_S(x) = P(S \leq x) = \left(\frac{x - a}{b - a}\right)^n, \quad F_I(x) = P(I \leq x) = 1 - \left(\frac{b - x}{b - a}\right)^n.$$

3.2 La loi exponentielle

Soit un réel **strictement positif** c .

Rappelons l'égalité

$$\int_0^{+\infty} c e^{-ct} dt = \left[- e^{-ct} \right]_0^{+\infty} = 1.$$

Définition 8

Une variable aléatoire réelle X admettant pour densité une fonction f nulle sur \mathbb{R}_-^* et définie, pour tout réel $x \geq 0$, par l'égalité

$$f(x) = c e^{-cx}$$

est dite suivre **la loi exponentielle de paramètre** (c), ce que l'on note

$$X \hookrightarrow \mathcal{E}(c).$$

► **Remarque**

L'existence d'une telle loi est justifiée par le rappel précédent.

Proposition 8

Si g est une densité d'une variable aléatoire réelle suivant la loi $\mathcal{E}(1)$, alors la fonction f définie par

$$f(x) = cg(cx)$$

est une densité d'une variable aléatoire réelle suivant la loi $\mathcal{E}(c)$.

Preuve

C'est clair par le changement de variable $x \mapsto cx$. □

► **Remarque**

On dispose donc des relations équivalentes

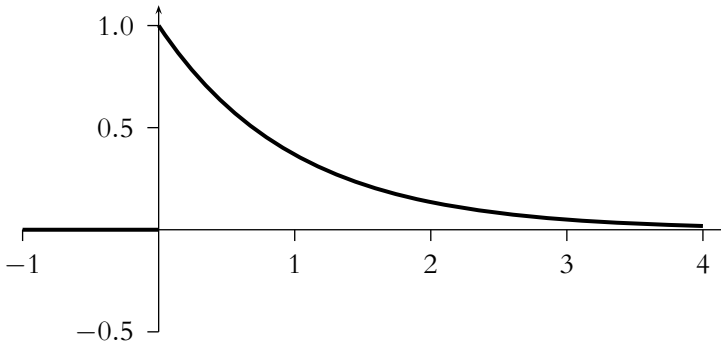
$$\begin{aligned} X \hookrightarrow \mathcal{E}(c) &\iff cX \hookrightarrow \mathcal{E}(1), \\ Y \hookrightarrow \mathcal{E}(1) &\iff \frac{1}{c}Y \hookrightarrow \mathcal{E}(c). \end{aligned}$$

Proposition 9

Si f est la densité d'une variable aléatoire réelle suivant la loi $\mathcal{E}(c)$, son graphe admet l'axe des abscisses comme asymptote et f est strictement décroissante sur \mathbb{R}_+ ; la demi-tangente au sommet $(0, c)$ a pour pente $-c^2$. Il n'y a pas de tangente d'inflexion.

Preuve

Cela résulte de l'étude de la dérivée de f , égale à $-c^2 e^{-cx}$, et de l'étude de la dérivée seconde de f , égale à $c^3 e^{-cx}$. □



Densité de la loi $\mathcal{E}(1)$

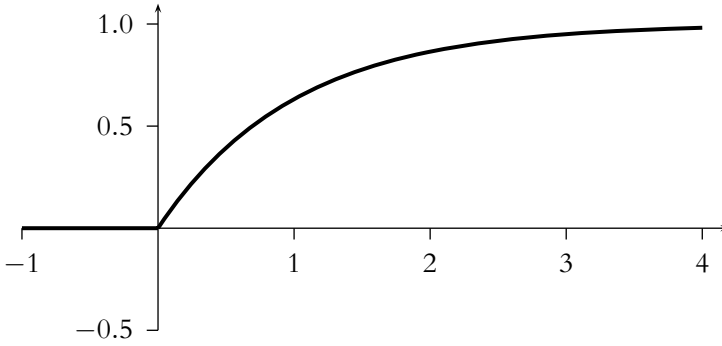
Théorème 23

La fonction de répartition d'une variable aléatoire réelle suivant la loi exponentielle $\mathcal{E}(c)$ est nulle sur \mathbb{R}_- et vérifie, pour tout réel $x \geq 0$, l'égalité

$$F(x) = P(X \leq x) = \int_0^x c e^{-ct} dt = 1 - e^{-cx}.$$

Preuve

Cela traduit la relation entre densité et fonction de répartition. □



Fonction de répartition de la loi $\mathcal{E}(1)$

Définition 9

On dit qu'une variable aléatoire réelle X associée à une probabilité P suit une loi **sans mémoire** si elle est positive ou nulle et si, pour tout couple (x, y) de réels positifs ou nuls, on dispose de l'égalité

$$P(X > x + y) = P(X > x) P(X > y).$$

En termes de fonction de répartition F , cela équivaut aux égalités

$$F(x + y) = 1 - (1 - F(x)) (1 - F(y)) = F(x) + F(y) - F(x) F(y).$$

► **Remarque**

Cette définition équivaut, si $P(X > x) > 0$, aux égalités

$$P(X > x + y \mid X > x) = P_{X > x}(X > x + y) = P(X > y).$$

Proposition 10

Toute variable certaine nulle obéit à une loi sans mémoire.

Preuve

En effet, elle ne prend que la valeur 0 qui est positive ou nulle ; de plus $F(x) = 1$ sur \mathbb{R}_+ , et $0 = 0 \times 0$. □

Proposition 11

Toute variable aléatoire réelle suivant une loi exponentielle obéit à une loi sans mémoire.

Preuve

En effet, elle ne prend que des valeurs strictement positives, et $F(x) = 1 - e^{-\alpha x}$ vérifie les conditions de la définition puisque $e^{-\alpha(x+y)} = e^{-\alpha x} e^{-\alpha y}$. \square

Lemme

Soit π une fonction de \mathbb{R}_+ dans $[0, 1]$ dont la restriction à \mathbb{R}_+^* n'est pas la fonction nulle et vérifiant l'égalité $\pi(x+y) = \pi(x)\pi(y)$ pour tout couple (x, y) de réels positifs ou nuls. Il existe alors un réel $a \in]0, 1]$ tel que, pour tout $x \geq 0$, on ait

$$\pi(x) = a^x.$$

Preuve

Notons que, par une récurrence immédiate sur l'entier positif n , on dispose pour tout réel $x \geq 0$ de l'égalité $\pi(nx) = (\pi(x))^n$.

Puisque π ne prend que des valeurs comprises entre 0 et 1, la fonction π vérifie la relation $\pi(x+y) \leq \pi(x)$ et est donc décroissante.

Puisque π n'est pas nulle sur \mathbb{R}_+^* , il existe un réel $t > 0$ tel que $0 < \pi(t) = \pi(t)\pi(0)$, d'où $\pi(0) = 1$.

Soit $a = \pi(1) \in [0, 1]$. Si l'on avait $a = 0$, il en résulterait que $\pi\left(\frac{1}{n}\right) = \sqrt[n]{a} = 0$ et $\pi(n) = a^n = 0$ pour tout entier $n > 0$, d'où $\pi(x) = 0$ pour tout $x > 0$ par décroissance de la fonction, ce qui est écarté par l'énoncé.

Donc $a \in]0, 1]$. On a $\pi(n) = a^n$ pour tout $n \in \mathbb{N}$. Si m est un entier naturel non nul et $r = \frac{n}{m}$, il vient

$$\pi(r)^m = \pi(nr) = \pi(n) = a^n = (a^r)^m$$

d'où $\pi(r) = a^r$ pour tout rationnel r positif ou nul.

Pour tout réel $x \geq 0$, il existe une suite (r_n) de rationnels positifs croissant vers x ; la suite décroissante $(\pi(r_n))$, vérifiant l'inégalité $\pi(r_n) \geq \pi(x)$, admet une limite ℓ vérifiant $\ell \geq \pi(x)$; de même une suite (s_n) de rationnels décroissant vers x est telle que $\pi(s_n)$ admet une limite μ vérifiant $\mu \leq \pi(x)$. Or $\pi(r_n) = a^{r_n}$, d'où $\ell = a^x$, de même $\mu = a^x$ et enfin $\pi(x) = a^x$. \square

► Remarques

- Si l'on ajoute aux hypothèses que π est continue, il suffit de vérifier que $\pi(r) = a^r$ pour tout rationnel r positif ou nul pour pouvoir conclure.
- Pour $a = 1$, on trouve $\pi(x) = 1 = e^{-0x}$; sinon, il existe un réel $c > 0$ tel que $a = e^{-c}$, d'où $\pi(x) = e^{-\alpha x}$.

Théorème 24

Pour qu'une aléatoire réelle X suive une loi sans mémoire, il faut et il suffit

- ou bien qu'elle soit discrète et certaine nulle,
- ou bien qu'elle soit à densité et obéisse à une loi exponentielle.

Preuve

Les conditions suffisantes ont été vues plus haut.

Soit X une variable à valeurs positive ou nulles.

Supposons que pour tout couple (x, y) de réels positifs ou nuls soit vérifiée la condition $1 - F(x+y) = (1 - F(x))(1 - F(y))$. Posant $\pi(x) = 1 - F(x) = P(X > x) \in [0, 1]$, il vient $0 \leq \pi(x) \leq 1$ et $\pi(x+y) = \pi(x)\pi(y)$.

Si $\pi(x)$ était nulle sur \mathbb{R}_+^* , on aurait $F(x) = 1$ sur \mathbb{R}_+^* , et même sur \mathbb{R}_+ par continuité à droite, et $F(x) = P(X \leq x) = 0$ sur \mathbb{R}_-^* : X est la variable certaine nulle.

Écartons ce cas ; on est donc dans le cadre de l'application du lemme et X vérifie une relation de la forme $\pi(x) = a^x$ pour tout $x \geq 0$ avec $a = \pi(1) \in]0, 1]$.

Si a était égale à 1, il en résulterait $F(x) = 1 - \pi(x) = 0$ pour tout $x \geq 0$, ce qui contredirait l'égalité $\lim_{+\infty} F = 1$.
Donc $a < 1$ et l'on peut poser $a = e^{-c}$ avec $c > 0$, ce qui définit effectivement une loi exponentielle de paramètre $c = -\ln(1 - F(1))$. \square

► **Remarque**

Si l'on met à part la loi certaine nulle, la loi exponentielle est donc caractérisée par l'absence de mémoire.

Théorème 25

Toute variable aléatoire réelle X suivant une loi exponentielle de paramètre c admet des moments de tous ordres, donnés par les égalités

$$M_n(X) = \frac{n!}{c^n}.$$

Preuve

Pour l'existence, il suffit de vérifier, par exemple en prenant des logarithmes, que l'inégalité suivante

$$t^n \exp(-ct) < \exp(-ct/2)$$

est vraie pour tout entier n et tout réel $t > 1$ assez grand puisqu'elle équivaut à la relation

$$\frac{t}{\ln t} > 2 \frac{n}{c}.$$

Mais cela résulte aussi de la définition de la fonction Γ , selon laquelle

$$M_n(X) = \int_0^{+\infty} x^n c e^{-cx} dx = \frac{1}{c^n} \int_0^{+\infty} y^n e^{-y} dy = \frac{\Gamma(n+1)}{c^n} = \frac{n!}{c^n}. \quad \square$$

Théorème 26

Toute variable aléatoire réelle X suivant une loi exponentielle de paramètre c admet une espérance mathématique égale à $\frac{1}{c}$.

Preuve

Cela résulte de la formule précédente puisque $M_1(X) = \frac{1}{c}$. \square

Théorème 27

Toute variable aléatoire réelle X suivant une loi exponentielle de paramètre c admet une variance égale à $\frac{1}{c^2}$.

Preuve

Cela résulte de la même formule, puisque

$$V(X) = M_2(X) - E(X)^2 = \frac{2}{c^2} - \left(\frac{1}{c}\right)^2 = \frac{1}{c^2}. \quad \square$$

3.3 Les lois gamma

Soient deux réels **strictement positifs** b et ν .

L'existence de l'intégrale $\int_0^{+\infty} t^{\nu-1} e^{-bt} dt$ peut être établie par un changement de variable simple. On obtient l'égalité $\int_0^{+\infty} t^{\nu-1} e^{-bt} dt = \frac{\Gamma(\nu)}{b^\nu}$ à partir de l'égalité $\int_0^{+\infty} t^{\nu-1} e^{-u} du = \Gamma(\nu)$.

Si ν est entier, on sait que $\Gamma(\nu) = (\nu - 1)!$. Nous avons également vu en exercice qu'une autre valeur intéressante de la fonction Γ était $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, d'où l'on peut déduire facilement les valeurs de $\Gamma\left(n + \frac{1}{2}\right)$ pour n entier.

Définition 10

Une variable aléatoire réelle X admettant pour densité une fonction f nulle sur \mathbb{R}_- et définie, pour tout réel $x > 0$, par l'égalité

$$f(x) = \frac{b^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-bx}$$

est dite suivre **la loi Gamma de paramètre** (b, ν) , ce que l'on note

$$X \hookrightarrow \Gamma(b, \nu).$$

Le nombre b est dit *paramètre d'échelle*.

► Remarques

- L'existence d'une telle loi est justifiée par la remarque précédente.
- Pour $\nu = 1$, on retrouve la **loi exponentielle** $\mathcal{E}(b)$ de paramètre b .
On remarquera que la densité donnée plus haut pour cette dernière diffère de celle-ci au point 0, puisqu'elle y prend la valeur 1 et non 0. Cela est sans importance, une densité n'étant, comme on le sait, que définie à un nombre fini de valeurs près.

Définition 11

Une variable aléatoire réelle X suivant la loi $\Gamma(1, \nu)$ est également dite suivre **la loi gamma standard de paramètre** ν , ce que l'on note

$$X \hookrightarrow \gamma(\nu).$$

► Remarque

On oublie parfois le mot standard, la différence étant assez claire quant aux paramètres du type général et du type particulier.

Proposition 12

Si g est une densité d'une variable aléatoire réelle suivant la loi standard $\gamma(\nu)$, alors la fonction f définie par

$$f(x) = b g(bx)$$

est une densité d'une variable aléatoire réelle suivant la loi $\Gamma(b, \nu)$.

Preuve

C'est clair par le changement de variable $x \mapsto bx$. □

► **Remarques**

- On dispose donc des relations équivalentes

$$X \mapsto \Gamma(b, \nu) \iff bX \mapsto \gamma(\nu),$$

$$Y \mapsto \gamma(\nu) \iff \frac{1}{b} Y \mapsto \Gamma(b, \nu).$$

- On notera que ces relations diffèrent de celle qui figure au programme.
- Ainsi les lois γ standard, cas particuliers des lois Γ , permettent de les engendrer par une opération simple.
- En dehors d'elles et des lois exponentielles, il existe encore une autre famille importante de cas particulier des lois Γ : ce sont les lois dites de χ_k^2 à k degrés de liberté (lire : khi-k deux, voire khi deux), pour $b = \frac{1}{2}$ et $\nu = \frac{k}{2}$ où k est entier, d'espérance mathématique $m = k$ et de variance $V = 2k$.

Proposition 13

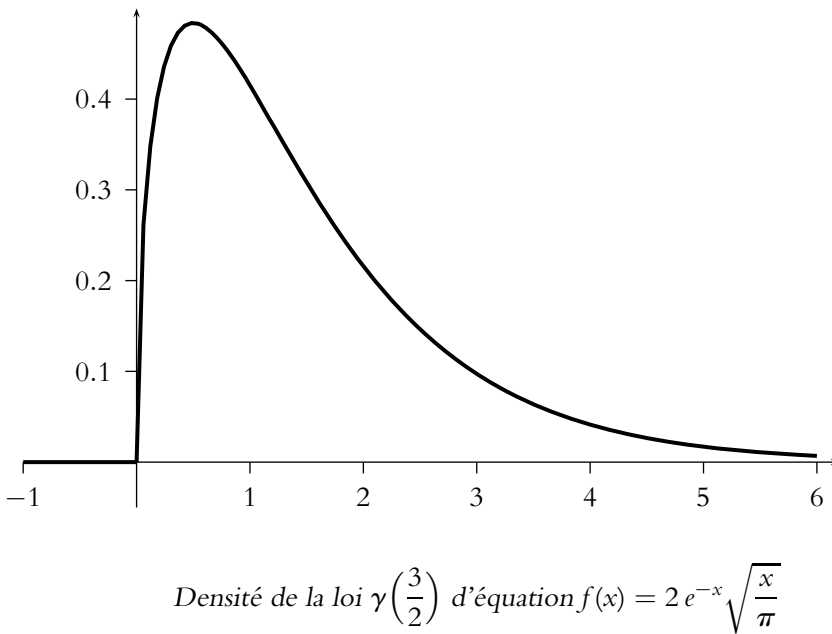
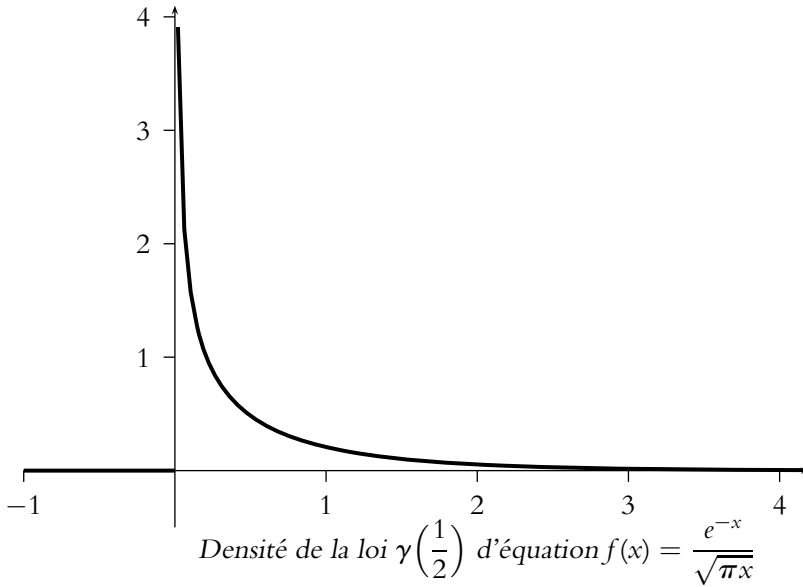
Si f est la densité d'une variable aléatoire réelle suivant la loi $\Gamma(b, \nu)$, son graphe admet l'axe des abscisses comme asymptote et

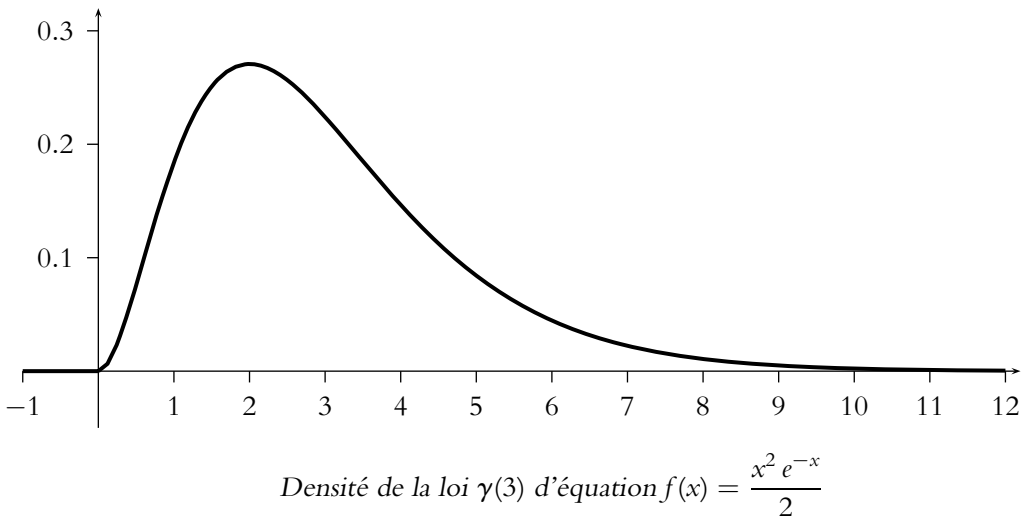
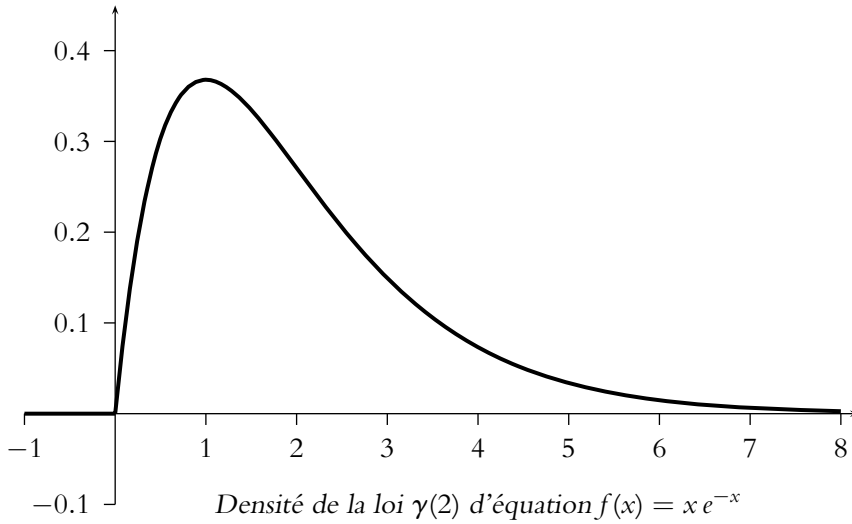
- pour $0 < \nu < 1$, l'axe des ordonnées est également asymptote, et f est strictement décroissante sur \mathbb{R}_+^* ;
 - pour $\nu = 1$, le graphe admet le point de coordonnées $(0, 1)$ comme point limite, et f , prolongée par $f(0) = 1$ où la demi-tangente à droite a pour pente $-b^2$, est strictement décroissante sur \mathbb{R}_+ (loi exponentielle) ;
 - pour $\nu > 1$, le graphe admet le point de coordonnées $(0, 0)$ comme point limite, et f , prolongée par $f(0) = 0$, est strictement croissante sur $\left[0, \frac{\nu - 1}{b}\right]$ puis strictement décroissante ; la demi-tangente à droite à l'origine a une pente infinie si $1 < \nu < 2$, égale à $\frac{b^\nu}{\Gamma(\nu)}$ si $\nu = 2$ et nulle si $\nu > 2$.
- Il y a une tangente d'inflexion si $1 < \nu < 2$ et deux si $\nu > 2$.

Preuve

Cela résulte de l'étude de la dérivée de f , égale à $\frac{b^\nu}{\Gamma(\nu)} (\nu - 1 - bx) x^{\nu-2} e^{-bx}$, et de celle de la limite de $\frac{f(x)}{x}$.

Enfin le nombre de points d'inflexion du graphe se déduit assez facilement de l'étude de la dérivée seconde de f , qui admet un facteur du second degré en x . □





Théorème 28

La fonction de répartition d'une variable aléatoire réelle suivant la loi $\Gamma(b, \nu)$ est nulle sur \mathbb{R}_- et vérifie, pour tout réel $x > 0$, l'égalité

$$F(x) = P(X \leq x) = \frac{b^\nu}{\Gamma(\nu)} \int_0^x t^{\nu-1} e^{-bt} dt.$$

Preuve

Cela traduit la relation entre densité et fonction de répartition. □

Définition 12

Une variable aléatoire réelle suivant une loi Γ de paramètre (b, n) avec n entier, est dite suivre **la loi d'Erlang** de même paramètre.

Théorème 29

La fonction de répartition d'une variable aléatoire réelle suivant une loi d'Erlang de paramètre (b, n) , vérifie l'égalité

$$F(x) = P(X \leq x) = 1 - \sum_{k=0}^{n-1} \frac{b^k x^k}{k!} \exp(-bx).$$

Preuve

La vérification, par récurrence sur n , consiste simplement à dériver le second membre. □

Exercice 3.

Soit (X_n) une suite de variables aléatoires à densité mutuellement indépendantes suivant une loi exponentielle $\mathcal{E}(1)$. On note $S_n = X_0 + X_1 + \dots + X_n$ et F_n la loi de répartition de S_n .

Pour tout réel x positif ou nul, on note A la variable aléatoire discrète d'image incluse dans \mathbb{N} définie par la relation « n est le plus petit entier tel que $S_n > x$ », c'est-à-dire encore pour $n \geq 1$ par

$$[A = n] = [S_{n-1} \leq x] \cap [S_n > x].$$

Si $S_n \leq x$ pour tout n , on posera $A = 0$.

Alors A suit une loi de Poisson de paramètre x décalée d'un indice.

Nous admettons le résultat du théorème de la page 370.

L'exercice de la page 189 montre que, pour $n \geq 1$, la loi de A vérifie l'égalité

$$P(A = n) = P([S_{n-1} \leq x] \cap [S_n > x]) = P(S_{n-1} \leq x) - P(S_n \leq x) = F_{n-1}(x) - F_n(x).$$

Or nous savons que

$$F_n(x) = P(S_n \leq x) = 1 - \sum_{k=0}^{n-1} \frac{x^k}{k!} \exp(-x).$$

Par suite $P(A = n) = \frac{x^{n-1}}{(n-1)!} e^{-x}$, ce qu'il fallait démontrer.

Théorème 30

Toute variable aléatoire réelle X suivant une loi γ standard de paramètre ν admet des moments de tous ordres, donnés par les égalités

$$M_n(X) = \frac{\Gamma(\nu + n)}{\Gamma(\nu)} = \prod_{m=0}^{n-1} (\nu + m).$$

Preuve

Cela résulte de l'existence des intégrales $\int_0^{+\infty} t^{\nu+n-1} e^{-t} dt = \Gamma(\nu+n)$. Les valeurs des moments se justifient à partir de la relation $\Gamma(x+1) = x\Gamma(x)$ par récurrence sur n . □

Théorème 31

Toute variable aléatoire réelle X suivant une loi γ standard de paramètre ν admet une espérance mathématique égale à ν .

Preuve

Cela résulte de la formule précédente puisque $M_1(X) = \nu$. □

Théorème 32

Toute variable aléatoire réelle X suivant une loi γ standard de paramètre ν admet une variance égale à ν .

Preuve

Cela résulte de la même formule, puisque

$$V(X) = M_2(X) - E(X)^2 = \nu(\nu+1) - \nu^2 = \nu. \quad \square$$

Théorème 33

Toute variable aléatoire réelle X suivant une loi Γ admet des moments de tous ordres, une espérance mathématique égale à $\frac{\nu}{b}$ et une variance égale à $\frac{\nu}{b^2}$.

Preuve

Cela résulte du fait que bX suit la loi standard de paramètre ν et des théorèmes précédents. □

3.4 La loi de Laplace

Cette loi est la première essayée par les mathématiciens pour rendre compte de la dispersion d'erreurs de mesures indépendantes autour d'une valeur moyenne nulle. Elle fut ensuite remplacée par la seconde loi de Laplace, c'est-à-dire la loi normale de Laplace-Gauss, qui en son sommet ne possède pas de point anguleux, mais une tangente parallèle à l'axe de abscisses.

Son étude peut se ramener à celle de la loi exponentielle, grâce à laquelle elle a pu être définie ; le lecteur est prié de procéder à cette adaptation. Il trouvera ci-dessous une étude directe de cette loi.

Elle figure dans ce cours notamment parce que sa simulation à partir de celle de la loi exponentielle pose un problème intéressant.

Soit un réel **strictement positif** c .

Rappelons l'égalité

$$\int_0^{+\infty} c e^{-at} dt = \left[-e^{-at} \right]_0^{+\infty} = 1.$$

Définition 13

Une variable aléatoire réelle X admettant pour densité une fonction f définie, pour tout réel x , par l'égalité

$$f(x) = \frac{c}{2} e^{-c|x|}$$

est dite suivre **la loi de Laplace de paramètre** (c), ce que l'on note

$$X \hookrightarrow \mathcal{L}(c).$$

► Remarque

L'existence d'une telle loi est justifiée par le rappel précédent.

Proposition 14

Si g est une densité d'une variable aléatoire réelle suivant la loi $\mathcal{L}(1)$, alors la fonction f définie par

$$f(x) = cg(cx)$$

est une densité d'une variable aléatoire réelle suivant la loi $\mathcal{L}(c)$.

Preuve

C'est clair par le changement de variable $x \mapsto cx$. □

► Remarque

On dispose donc des relations équivalentes

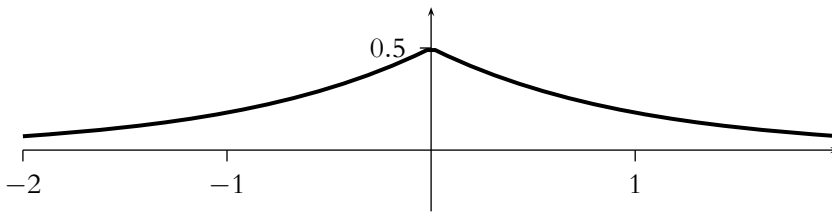
$$\begin{aligned} X \hookrightarrow \mathcal{L}(c) &\iff cX \hookrightarrow \mathcal{L}(1), \\ Y \hookrightarrow \mathcal{L}(1) &\iff \frac{1}{c}Y \hookrightarrow \mathcal{L}(c). \end{aligned}$$

Proposition 15

Si f est la densité d'une variable aléatoire réelle suivant la loi $\mathcal{L}(c)$, son graphe est symétrique par rapport à l'axe des ordonnées, admet l'axe des abscisses comme asymptote et f est strictement décroissante sur \mathbb{R}_+ ; les demi-tangentes au sommet $\left(0, \frac{c}{2}\right)$ ont pour pentes $\pm \frac{c^2}{2}$. Il n'y pas de tangente d'inflexion.

Preuve

Cela résulte de l'étude sur \mathbb{R}_+ de la dérivée de f , égale à $-\frac{c^2}{2}e^{-cx}$, et de l'étude de la dérivée seconde de f , égale à $\frac{c^3}{2}e^{-cx}$. □



Densité de la loi $\mathcal{L}(1)$

Théorème 34

La fonction de répartition d'une variable aléatoire réelle suivant la loi de Laplace $\mathcal{L}(c)$ vérifie, pour tout réel $x \geq 0$, l'égalité

$$F(x) = P(X \leq x) = \int_{-\infty}^x \frac{c}{2} e^{-c|t|} dt.$$

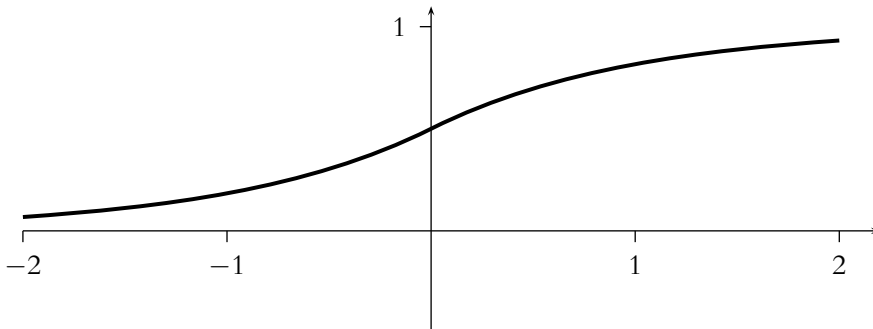
Preuve

Cela traduit la relation entre densité et fonction de répartition. □

► **Remarque**

Une telle densité est paire ; la fonction de répartition vérifie donc l'égalité $F(-x) = 1 - F(x)$.

En fait $F(x) = \frac{e^{cx}}{2}$ si $x \leq 0$, et $1 - \frac{e^{-cx}}{2}$ si $x \geq 0$.



Fonction de répartition de la loi $\mathcal{L}(1)$

Théorème 35

Toute variable aléatoire réelle X suivant une loi de Laplace de paramètre c admet des moments de tous ordres, donnés par les égalités

$$M_{2n}(X) = \frac{(2n)!}{c^{2n}}, \quad M_{2n+1} = 0.$$

Preuve

Pour l'existence, il suffit de vérifier, par exemple en prenant des logarithmes, que l'inégalité suivante

$$t^n \exp(-ct) < \exp(-ct/2)$$

est vraie pour tout entier n et tout réel $t > 1$ assez grand puisqu'elle équivaut à la relation

$$\frac{t}{\ln t} > 2 \frac{n}{c}.$$

Leur calcul dépend notamment de la définition de la fonction Γ , selon laquelle

$$M_{2n}(X) = 2 \int_0^{+\infty} x^{2n} \frac{c}{2} e^{-cx} dx = \frac{1}{c^{2n}} \int_0^{+\infty} y^{2n} e^{-y} dy = \frac{\Gamma(2n+1)}{c^{2n}} = \frac{(2n)!}{c^{2n}}.$$

Enfin le cas de M_{2n+1} est évident par parité dès que l'on sait que les moments de tous ordres existent. □

Théorème 36

Toute variable aléatoire réelle X suivant une loi de Laplace de paramètre c admet une espérance mathématique nulle.

Preuve

Cela résulte de la formule précédente puisque $M_1(X) = 0$. □

Théorème 37

Toute variable aléatoire réelle X suivant une loi de Laplace de paramètre c admet une variance égale à $\frac{2}{c^2}$.

Preuve

Cela résulte de la même formule, puisque

$$V(X) = M_2(X) - E(X)^2 = M_2(X) = \frac{2}{c^2}. \quad \square$$

3.5 La loi de Cauchy

Rappelons l'égalité

$$\int_0^{+\infty} \frac{dt}{1+t^2} = \left[\text{Arctan } t \right]_0^{+\infty} = \frac{\pi}{2}.$$

Définition 14

Une variable aléatoire réelle X admettant pour densité une fonction f définie, pour tout réel x , par l'égalité

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

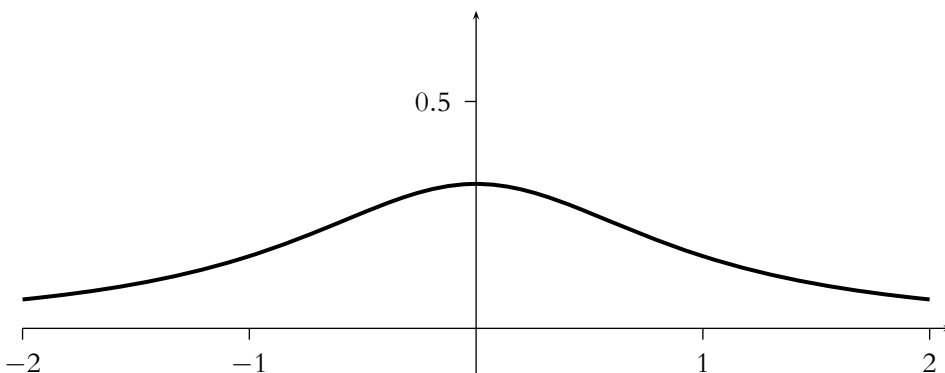
est dite suivre **la loi de Cauchy**.

Proposition 16

Si f est la densité d'une variable aléatoire réelle suivant la loi de Cauchy, son graphe est symétrique par rapport à l'axe des ordonnées, admet l'axe des abscisses comme asymptote et f est strictement décroissante sur \mathbb{R}_+ . Il y a deux tangentes d'inflexion.

Preuve

Cela résulte de l'étude de la dérivée de f , égale à $\frac{2}{\pi} \frac{-x}{(1+x^2)^2}$, et de l'étude de la dérivée seconde de f , égale à $\frac{2}{\pi} \frac{3x^2 - 1}{(1+x^2)^3}$. □



Densité de la loi de Cauchy

Théorème 38

La fonction de répartition d'une variable aléatoire réelle suivant la loi de Cauchy vérifie, pour tout réel x , l'égalité

$$F(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\pi} \frac{dt}{1+t^2} = \frac{1}{2} + \frac{1}{\pi} \operatorname{Arctan} x.$$

Preuve

Cela traduit la relation entre densité et fonction de répartition. □

► **Remarque**

Une telle densité est paire ; la fonction de répartition vérifie donc l'égalité $F(-x) = 1 - F(x)$.

Théorème 39

Aucune variable aléatoire réelle suivant une loi de Cauchy ne possède de moments.

Preuve

Il suffit de vérifier la minoration, pour tout entier $n \geq 1$ et tout réel $A \geq 1$

$$\int_1^A \frac{x^n dx}{1+x^2} \geq \int_0^A \frac{x dx}{1+x^2} = \ln \sqrt{\frac{1+A^2}{2}}.$$

□

Corollaire 9

Aucune variable aléatoire réelle suivant une loi de Cauchy ne possède d'espérance mathématique ni de variance.

3.6 La loi normale de Laplace-Gauss

Le dernier exemple de loi à densité est le plus complexe, mais le plus important sur le plan théorique à cause du *théorème de la limite centrée* qui sera exposé dans le chapitre sur les convergences en calcul des probabilités. C'est Laplace qui, après avoir donné un premier exemple vu ci-dessus, commença à étudier cette loi mais c'est Gauss qui en fit une étude rigoureuse et complète.

La loi normale centrée réduite

Rappelons qu'un changement de variable simple donne l'égalité $\int_{-\infty}^{+\infty} \exp\left(-\frac{t^2}{2}\right) dt = \sqrt{2\pi}$ à partir de l'égalité $\int_0^{+\infty} \exp(-u^2) du = \frac{\sqrt{\pi}}{2}$.

Définition 15

Une variable aléatoire réelle X admettant pour densité une fonction f définie pour tout réel x par l'égalité

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

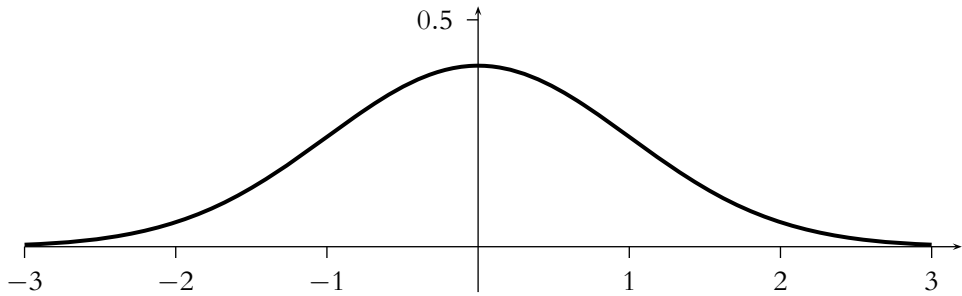
est dite suivre **la loi normale centrée réduite**, ce que l'on note

$$X \hookrightarrow \mathcal{N}(0, 1).$$

► **Remarque**

L'existence d'une telle loi est justifiée par le rappel précédent.

Le graphe de f est symétrique par rapport à l'axe des ordonnées ; celui des abscisses en est une asymptote. Le maximum de f vaut $f(0) = \frac{1}{\sqrt{2\pi}} \approx 0,3989$. Les dérivées successives de $\exp\left(-\frac{x^2}{2}\right)$ étant $-x \exp\left(-\frac{x^2}{2}\right)$ et $(x^2 - 1) \exp\left(-\frac{x^2}{2}\right)$, on voit que ce graphe admet deux points d'inflexion aux points d'abscisse 1 et -1 et d'ordonnée $\frac{1}{\sqrt{2\pi e}} \approx 0,2420$ (ce nombre est aussi la valeur absolue des pentes des tangentes d'inflexion, qui coupent donc l'axe des abscisses aux points d'abscisse 2 et -2).



Densité de la loi normale réduite d'équation $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

Théorème 40

La fonction de répartition d'une variable aléatoire réelle suivant une loi normale centrée réduite vérifie l'égalité

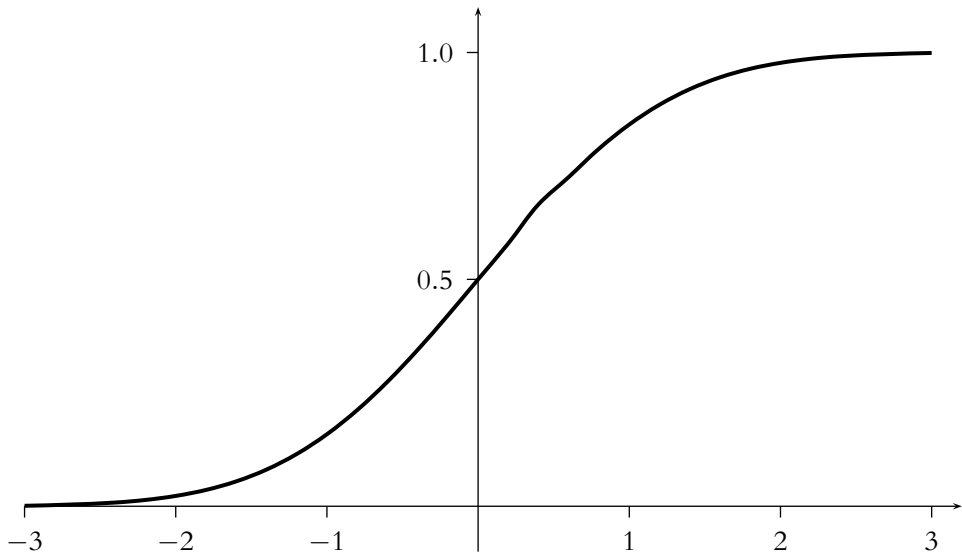
$$F(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt.$$

Preuve

Cela traduit la relation entre densité et fonction de répartition. □

► **Remarque**

Une telle densité est paire ; la fonction de répartition vérifie donc l'égalité $F(-x) = 1 - F(x)$.



Fonction de répartition de la loi normale centrée réduite

Théorème 41

Toute variable aléatoire réelle suivant une loi normale centrée réduite admet des moments de tous ordres.

Preuve

Il suffit de vérifier, par exemple en prenant des logarithmes, que l'inégalité suivante

$$t^n \exp\left(-\frac{t^2}{2}\right) < \exp(-t)$$

est vraie pour tout entier n et tout réel $t > 1$ assez grand puisqu'elle équivaut à la relation

$$\frac{t(t-2)}{2 \ln t} > 2n. \quad \square$$

Théorème 42

Toute variable aléatoire réelle suivant une loi normale centrée réduite admet une espérance mathématique nulle.

Preuve

Cela résulte du théorème précédent et de la parité de f . □

Théorème 43

Toute variable aléatoire réelle suivant une loi normale centrée réduite admet une variance égale à 1.

Preuve

Posons $S(x) = \int_0^x t \exp\left(-\frac{t^2}{2}\right) dt = 1 - \exp\left(-\frac{x^2}{2}\right)$, ce qui se vérifie instantanément en dérivant le second membre, et $T(x) = x S(x) - \int_0^x S(t) dt$.

On dispose des égalités

$$T(x) = x \left[1 - \exp\left(-\frac{x^2}{2}\right)\right] + \int_0^x \left[\exp\left(-\frac{t^2}{2}\right) - 1\right] dt = -x \exp\left(-\frac{x^2}{2}\right) + \int_0^x \exp\left(-\frac{t^2}{2}\right) dt$$

ainsi que de $T'(x) = x S'(x) = x^2 \exp\left(-\frac{x^2}{2}\right)$.

Par suite la variance $V(X) = M_2(X)$ est la limite quand A et B tendent indépendamment vers $+\infty$ de $\frac{1}{\sqrt{2\pi}} (T(B) - T(-A))$, soit encore la limite quand B tend vers $+\infty$ de $\sqrt{\frac{2}{\pi}} T(B)$, soit enfin

$$\sqrt{\frac{2}{\pi}} \int_0^{+\infty} \exp\left(-\frac{t^2}{2}\right) dt = \sqrt{\frac{2}{\pi}} \frac{\sqrt{2\pi}}{2} = 1. \quad \square$$

► Remarque

On peut retrouver ce résultat par une double intégration par parties.

La loi normale générale

Rappelons que si X est une variable aléatoire réelle admettant une densité f et si $Y = \frac{1}{s}(X - m)$ avec $s > 0$, alors Y est une variable aléatoire admettant pour densité la fonction g définie par $g(y) = sf(sy + m)$, ce qui équivaut à $f(x) = \frac{1}{s}g\left(\frac{x - m}{s}\right)$.

Dans ce qui suit, nous supposerons que Y suit une loi normale centrée réduite, et poserons $m = \mu$, $s = \sigma > 0$.

Définition 16

Soit μ un réel, et σ un réel strictement positif. Une variable aléatoire réelle X admettant pour densité une fonction f définie pour tout réel x par l'égalité

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

est dite suivre **la loi normale** de paramètre (μ, σ^2) , ce que l'on note

$$X \hookrightarrow \mathcal{N}(\mu, \sigma^2).$$

► Remarques

- L'existence d'une telle loi est justifiée par le rappel précédent.
- Certains auteurs utilisent (μ, σ) comme paramètre ; nous verrons plus loin, lors de l'étude de la stabilité des variables normales par addition (page 278), la raison de l'introduction, *a priori* bizarre, d'un carré dans le paramètre.

Théorème 44

La fonction de répartition d'une variable aléatoire réelle suivant une loi normale de paramètre (μ, σ^2) vérifie l'égalité

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right) dt.$$

Preuve

Cela traduit la relation entre densité et fonction de répartition. □

Théorème 45

Toute variable aléatoire réelle suivant une loi normale centrée réduite admet des moments de tous ordres.

Preuve

Ce théorème se démontre par la même méthode que celle qui a été employée pour les variables normales centrées réduites. □

Théorème 46

Toute variable aléatoire réelle suivant une loi normale de paramètre (μ, σ^2) admet une espérance mathématique égale à μ .

Preuve

Cela résulte du fait que $Y = \frac{1}{\sigma}(X - \mu)$ a pour densité

$$g(y) = \sigma f(\sigma y + \mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$$

et suit donc une loi normale centrée réduite qui admet 0 comme espérance, d'où

$$E(X) = E(\mu + \sigma Y) = \mu. \quad \square$$

Théorème 47

Toute variable aléatoire réelle suivant une loi normale de paramètre (μ, σ^2) admet une variance égale à σ^2 .

Preuve

Ici encore on utilise le fait que Y admet 1 pour variance, d'où

$$V(X) = V(\mu + \sigma Y) = \sigma^2. \quad \square$$

Corollaire 10

La variable aléatoire centrée réduite Y d'une variable aléatoire réelle X qui suit une loi normale de paramètre (μ, σ^2) vérifie l'égalité

$$Y = \frac{1}{\sigma}(X - \mu).$$

Preuve

Cela résulte des deux théorèmes précédents. □

► Remarque

On peut donc écrire les relations équivalentes

$$X \hookrightarrow \mathcal{N}(\mu, \sigma^2) \iff \frac{1}{\sigma}(X - \mu) \hookrightarrow \mathcal{N}(0, 1),$$

$$Y \hookrightarrow \mathcal{N}(0, 1) \iff \mu + \sigma Y \hookrightarrow \mathcal{N}(\mu, \sigma^2).$$

3.7 Théorèmes de stabilité par rapport à l'addition

En fin de ce chapitre, nous avons regroupé deux théorèmes de stabilité de lois à densité par rapport à l'addition, complétant le fait que la somme de deux variables discrètes indépendantes suivant des lois de Poisson de paramètres λ et λ' suit une loi de Poisson de paramètre $\lambda + \lambda'$, et que celle de deux variables discrètes indépendantes

suyant des lois binomiales de paramètres (n, p) et (n', p) suit une loi binomiale de paramètre $(n + n', p)$.

Les lois gamma

Théorème 48

La somme de deux variables aléatoires réelles **indépendantes** suivant des lois Γ de paramètres respectifs (b, ν) et (b, ν') , de même paramètre d'échelle, suit elle-même une loi Γ de paramètre $(b, \nu + \nu')$.

Preuve

Nous admettons que nous pouvons nous placer dans le cadre du théorème 6. La propriété fondamentale de l'exponentielle et le changement de variable $t = zu$ donnent les égalités

$$h(z) = \frac{b^{\nu+\nu'} e^{-bz}}{\Gamma(\nu)\Gamma(\nu')} \int_0^z t^{\nu-1} (z-t)^{\nu'-1} dt = \lambda \frac{b^{\nu+\nu'}}{\Gamma(\nu)\Gamma(\nu')} z^{\nu+\nu'-1} e^{-bz}$$

avec $\lambda = \int_0^1 u^{\nu-1} (1-u)^{\nu'-1} du$ indépendant de z pour tout $z > 0$.

Puisque h est une densité proportionnelle à $z^{\nu+\nu'-1} e^{-bz}$, c'est bien une densité de loi Γ de paramètre $(b, \nu + \nu')$ \square

► Remarque

En fait, nous savons que $\int_0^{+\infty} h(z) dz = 1$. Par suite $\lambda = \frac{\Gamma(\nu)\Gamma(\nu')}{\Gamma(\nu + \nu')}$ (ce qui peut d'ailleurs être démontré directement). Le second membre se note $B(\nu, \nu')$ et porte le nom de *fonction beta*.

Théorème 49

La loi de la variable aléatoire réelle $S_n = X_1 + X_2 + \dots + X_n$, somme de n variables aléatoires réelles **mutuellement indépendantes** suivant la loi exponentielle $\mathcal{E}(b)$, est la loi Γ de paramètre (b, n) (loi d'Erlang).

Preuve

Cela résulte du théorème ci-dessus et du fait (admis) que la somme S_{n-1} est indépendante de X_n . \square

► Remarque

Ce résultat joue un rôle important pour l'écriture d'algorithmes de simulation de lois Γ où ν est entier.

Les lois normales

Le calcul est plus délicat en ce qui concerne les lois normales, mais leur importance mérite que nous donnions ici une preuve, calculatoire, de leur stabilité par addition.

Lemme

Soient (a, b, c) trois réels vérifiant $a > 0$. On dispose alors de l'égalité

$$\int_{-\infty}^{+\infty} \exp(-at^2 - 2bt - c) dt = e^{-d} \sqrt{\frac{\pi}{a}}$$

avec $d = c - \frac{b^2}{a}$.

Preuve

Cela résulte des égalités

$$\int_{-\infty}^{+\infty} \exp(-au^2) du = \sqrt{\frac{\pi}{a}} \quad \text{et} \quad at^2 + 2bt + c = au^2 + d \quad \text{où} \quad u = t + \frac{b}{a}.$$

□

Théorème 50

La somme de deux variables aléatoires réelles **indépendantes** suivant des lois normales de paramètres respectifs (μ, σ^2) et (μ', σ'^2) suit elle-même une loi normale de paramètre $(\mu + \mu', \sigma^2 + \sigma'^2)$.

Preuve

Nous admettons que nous pouvons nous placer dans le cadre du théorème 6. La propriété fondamentale de l'exponentielle donne les égalités

$$\begin{aligned} h(z) &= \int_{-\infty}^{+\infty} f(t) f'(z-t) dt = \frac{1}{2\pi\sigma\sigma'} \int_{-\infty}^{+\infty} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2} - \frac{(z-t-\mu')^2}{2\sigma'^2}\right) dt \\ &= \frac{1}{2\pi\sigma\sigma'} \int_{-\infty}^{+\infty} \exp(-at^2 - 2bt - c) dt = \frac{1}{2\pi\sigma\sigma'} e^{-d} \sqrt{\frac{\pi}{a}}, \end{aligned}$$

avec $a = \frac{\sigma^2 + \sigma'^2}{2\sigma^2\sigma'^2}$, $b = \frac{-\mu\sigma'^2 + (\mu' - z)\sigma^2}{2\sigma^2\sigma'^2}$ et $c = \frac{\mu^2\sigma'^2 + (z - \mu')^2\sigma^2}{2\sigma^2\sigma'^2}$ d'où, après simplification, $d = \frac{(z - \mu - \mu')^2}{2(\sigma^2 + \sigma'^2)}$.

Le résultat découle alors du lemme.

□

► **Remarques**

- C'est ce théorème qui fait que le programme préfère écrire (μ, σ^2) comme paramètre plutôt que (μ, σ) .
- Si l'on sait que la somme est une variable aléatoire réelle suivant une loi normale, son paramètre est clairement $(E(X + X'), V(X + X')) = (\mu + \mu', \sigma^2 + \sigma'^2)$ puisqu'alors $V(X + X') = V(X) + V(X')$.
- Une variante consiste à se ramener au cas $\mu = \mu' = 0$ et à mettre $h(z)$ sous la forme

$$h(z) = \frac{1}{2\pi\sigma\sigma'} \exp\left(-\frac{z^2}{2(\sigma^2 + \sigma'^2)}\right) \int_{-\infty}^{+\infty} \exp\left[-\frac{\left(t\sqrt{\sigma^2 + \sigma'^2} - z\frac{\sigma^2}{\sqrt{\sigma^2 + \sigma'^2}}\right)^2}{2\sigma^2\sigma'^2}\right] dt$$

où l'intégrale vaut $\sigma\sigma' \sqrt{\frac{2\pi}{\sigma^2 + \sigma'^2}}$, soit

$$h(z) = \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma'^2)}} \exp\left(-\frac{z^2}{2(\sigma^2 + \sigma'^2)}\right).$$

- Il existe des démonstrations plus élégantes, mais faisant intervenir des notions non abordées dans ce livre.

1. Soit f une densité continue d'une variable X admettant une espérance mathématique. Caractériser la valeur du réel a pour laquelle est minimum la fonction définie par

$$\varphi(a) = \int_{-\infty}^{+\infty} |t - a| f(t) dt.$$

2. Soit f une densité continue d'une variable X admettant une variance. Caractériser la valeur du réel a pour laquelle est minimum la fonction définie par

$$\psi(a) = \int_{-\infty}^{+\infty} (t - a)^2 f(t) dt.$$

3. **1.** Soit X une variable aléatoire à densité prenant ses valeurs dans \mathbb{R}_+ et admettant une espérance notée $E(X)$. Montrer que $\lim_{n \rightarrow +\infty} n(1 - F_X(x)) = 0$.

2. Soit X une variable aléatoire à densité prenant ses valeurs dans \mathbb{R}_+ . Montrer que X admet une espérance si, et seulement si, $\int_0^{+\infty} (1 - F_X(x)) dx$ converge et que l'on a alors

$$E(X) = \int_0^{+\infty} (1 - F_X(x)) dx.$$

3. Soit X une variable aléatoire admettant une densité f continue sur \mathbb{R} et telle que $E(X^2)$ existe.

a) Montrer que $\lim_{x \rightarrow +\infty} P(|X| \geq x) = 0$.

b) Montrer que l'intégrale $\int_0^{+\infty} t P(|X| \geq t) dt$ converge et est égale à $\frac{E(X^2)}{2}$.

4. **1.** On considère la fonction f_X définie par $f_X(x) = \mathbb{1}_{\mathbb{R}_+^*}(x) \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}(\ln x - 1)^2}$.

a) Montrer que f_X est une densité et exprimer la fonction de répartition de X à l'aide de la fonction de répartition d'une loi normale centrée réduite.

b) Calculer $E(X)$ et $V(X)$.

2. On considère une variable aléatoire Y qui suit une loi normale de paramètre (m, σ^2) . On pose $X = e^Y$.

a) Déterminer une densité de X (on dit que X suit la loi Log-normale de paramètre (m, σ^2)) ($\mathcal{LN}(m, \sigma^2)$).

b) Calculer $E(X)$ et $V(X)$.

3. a) Soit X une variable aléatoire suivant une loi Log-normale de paramètre (m, σ^2) . On considère un réel a et un réel non nul b . Quelle est la loi suivie par $Z = aX^b$?

b) Donner l'espérance et la variance de Z .

5. Soit λ un réel. On considère la fonction f_X définie par $f_X(x) = \mathbb{1}_{]1, +\infty[}(x) \frac{\lambda}{x^2 - \frac{1}{9}}$.

Déterminer λ tel que f_X soit la densité d'une variable aléatoire X dont on étudiera l'espérance.

6. Soit X une variable uniforme sur $[0, 1]$. Déterminer la loi, et, si elles existent, l'espérance et la variance de $Y = \sin \pi X$.

7. Soit X une variable uniforme sur $\left] \frac{-\pi}{2}, \frac{\pi}{2} \right[$. Déterminer la loi, et, si elles existent, l'espérance et la variance de $Y = \tan X$.

8. **(Loi de Weibull)** Soit F la fonction définie par $F(x) = \mathbb{1}_{\mathbb{R}_+^*}(x) \left(1 - e^{-\left(\frac{x}{\beta}\right)^\alpha}\right)$.

1. Montrer que F est une fonction de répartition.
2. Soit X une variable aléatoire admettant F comme fonction de répartition. Déterminer une densité de X .
3. Calculer $E(X)$ et $V(X)$.
4. Pour tout entier naturel n supérieur ou égal à 1, calculer $E(X^n)$.

9. **(Loi de Rayleigh)** Soit f la fonction définie par $f(x) = \mathbb{1}_{\mathbb{R}_+}(x) x e^{-\frac{x^2}{2}}$.

1. Montrer que f est une densité.
2. Soit X une variable aléatoire admettant f pour densité. Déterminer la fonction de répartition de X .
3. Calculer $E(X)$ et $V(X)$.
4. On pose $Y = X^2$. Déterminer la loi de Y .

10. **(Loi de Pareto)** Soient α et a des réels strictement positifs, et x_0 et λ des réels. On considère la fonction f définie par $f(x) = \mathbb{1}_{]x_0+a, +\infty[}(x) \lambda \left(\frac{a}{x-x_0}\right)^{\alpha+1}$.

1. a) Déterminer λ pour que f soit une densité d'une variable aléatoire X . On dit alors que X suit une loi de Pareto de paramètre (α, a, x_0) .
 b) Déterminer la fonction de répartition de X .
 c) Étudier l'existence et la valeur éventuelle de $E(X)$.
 d) Étudier l'existence et la valeur éventuelle de $V(X)$.
2. Soient X une variable aléatoire suivant une loi de Pareto de paramètre (α, a, x_0) et des réels $r > 0$ et s .
 Quelle est la loi suivie par la variable aléatoire $Y = rX + s$?
3. a) Soient X une variable aléatoire suivant une loi exponentielle de paramètre strictement positif μ et deux réels $\beta > 0$ et $\gamma > 1$.
 Quelle est la loi de la variable aléatoire $Y = \beta \cdot \gamma^X$?
 b) Étudier la réciproque de la propriété ainsi démontrée.
 c) Soient Z une variable aléatoire qui suit une loi de Pareto de paramètre $(\alpha, a, 0)$, et c un réel strictement positif.
 Quelle est la loi de la variable aléatoire Z^c ?

11. (Loi beta de première espèce) Dans tout l'exercice, n et m sont des entiers naturels non nuls. Pour tout réel t de l'intervalle $[0, 1]$, on pose $B(t, n, m) = \int_0^t u^{n-1}(1-u)^{m-1} du$ et $\beta(n, m) = B(1, n, m)$.

1. **a)** Prouver que $\beta(n, m) = \beta(m, n)$ et que, pour $m \geq 2$, $\beta(n, m) = \frac{m-1}{n} \beta(n+1, m-1)$.
b) En déduire $\beta(n, m)$.
2. On considère la fonction $f_{n,m}(x) = \mathbb{1}_{]0,1[}(x) \frac{1}{\beta(n, m)} x^{n-1}(1-x^{m-1})$.
a) Montrer que $f_{n,m}$ est une densité de probabilité.
b) Soit X une variable aléatoire admettant $f_{n,m}$ comme densité. Après en avoir justifié l'existence, calculer $E(X)$ et $V(X)$.

12. Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes suivant toutes la loi uniforme sur $[0, 1]$.

1. Déterminer et représenter graphiquement une densité de $X_1 + X_2$.
2. Déterminer et représenter graphiquement une densité de $X_1 + X_2 + X_3$.
3. Déterminer et représenter graphiquement une densité de $X_1 + X_2 + X_3 + X_4$.
4. Les résultats des questions précédentes vous inspirent-ils une remarque. Si oui, laquelle ?

13. Dans cet exercice, on pourra utiliser la loi normale.

1. **a)** Calculer $I = \int_0^1 e^{-x^2+2x-1} dx$.
b) Calculer $I = \int_0^1 e^{-x^2+2x} dx$.
c) Calculer $I = \int_0^1 e^{-4x^2+4x-1} dx$.
d) Calculer $I = \int_0^1 e^{-4x^2+4x-1} dx$.
2. **a)** Calculer $I = \int_{-\infty}^{+\infty} e^{-x^2+2x-1} dx$.
b) Calculer $I = \int_{-\infty}^{+\infty} e^{-x^2+2x} dx$.
c) Calculer $I = \int_{-\infty}^{+\infty} e^{-4x^2+4x-1} dx$.
d) Calculer $I = \int_{-\infty}^{+\infty} e^{-4x^2+4x-1} dx$.
e) Soient a un réel strictement positif et b et c des réels. Calculer $I = \int_{-\infty}^{+\infty} e^{-ax^2+bx+c} dx$.

14. Dans cet exercice, on utilisera une table numérique de la fonction de répartition de la loi normale centrée réduite.

- Soit X une variable aléatoire suivant la loi normale centrée réduite. Calculer des valeurs approchées des probabilités suivantes
 - $P(X \leq 1,36)$; $P(X = 1,36)$; $P(X > 1,36)$.
 - $P(X \leq -0,86)$; $P(X > -0,86)$.
 - $P(-0,86 \leq X \leq 1,36)$; $P(0,86 \leq X \leq 1,36)$; $P(-1,36 \leq X \leq -0,86)$.
- La variable X suivant toujours la loi normale centrée réduite, déterminer une valeur approchée de x tel que la condition donnée soit remplie (on pourra avoir à effectuer une interpolation linéaire)
 - $P(X \leq x) = 0,9474$; $P(X \leq x) = 0,0537$.
 - $P(X \leq x) = 0,9258$; $P(X \leq x) = 0,1778$.
- Soit X une variable aléatoire suivant la loi normale $\mathcal{N}(5,9)$. Calculer des valeurs approchées des probabilités suivantes
 - $P(X \leq 1,36)$; $P(X = 1,36)$; $P(X > 1,36)$.
 - $P(X \leq -0,86)$; $P(X > -0,86)$.
 - $P(-0,86 \leq X \leq 1,36)$; $P(0,86 \leq X \leq 1,36)$; $P(-1,36 \leq X \leq -0,86)$.
- La variable X suivant toujours la loi normale $\mathcal{N}(5,9)$, déterminer une valeur approchée de x tel que la condition donnée soit remplie (on pourra avoir à effectuer une interpolation linéaire)
 - $P(X \leq x) = 0,9474$; $P(X \leq x) = 0,0537$.
 - $P(X \leq x) = 0,9258$; $P(X \leq x) = 0,1778$.

15. On note $\mathcal{C}(t)$ la loi d'une variable aléatoire de densité définie par

$$f(x) = \frac{1}{\pi} \frac{t}{t^2 + x^2}.$$

- Montrer que la somme de deux variables indépendantes suivant respectivement les lois $\mathcal{C}(u)$ et $\mathcal{C}(v)$ suit la loi $\mathcal{C}(u+v)$. On pourra déterminer (A, B, C, D) tel que

$$\frac{1}{(u^2 + x^2)(v^2 + [z - x]^2)} = \frac{Ax + B}{u^2 + x^2} + \frac{C(z - x) + D}{v^2 + (z - x)^2}.$$

- Que peut-on dire de la moyenne arithmétique de variables aléatoires mutuellement indépendantes suivant toutes la loi $\mathcal{C}(t)$?

Autres exercices

On cherchera également avec profit les problèmes et exercices de probabilité proposés aux concours suivants dans diverses séries :

ESCP-EAP 2002, 2004; ESCP 1990, 2002; HEC 1990; EDHEC 1998.

Convergences

11

Dès l'instant où le programme prévoit l'introduction des notions de convergence en probabilité et en loi, et dans le but de rendre ces notions opérationnelles, nous avons jugé utile d'en étudier les propriétés élémentaires. Il n'est pas question de se contenter d'apprendre par cœur les résultats donnés. Au contraire, on devra considérer les questions traitées comme autant d'exercices visant à approfondir les notions attachées aux suites de variables aléatoires réelles, notions qui sont utilisées dans l'échantillonnage et l'estimation.

Dans ce chapitre, nous considérons des variables aléatoires X_n définies sur un même espace probabilisé (Ω, \mathcal{T}, P) , et dépendant d'un entier n . Nous envisageons d'étudier le comportement de la suite (X_n) lorsque n tend vers l'infini.

Pour mener à bien cette étude, nous aurons besoin de critères permettant de comparer deux variables aléatoires, et de dire dans quelle mesure on peut les considérer comme assez proches pour que l'une puisse être considérée comme une bonne approximation de l'autre.

Dans le cas d'éléments de \mathbb{R} , \mathbb{R}^n ou d'ensembles de fonctions d'un intervalle I de \mathbb{R}^p dans \mathbb{R} , un tel critère d'approximation peut être fourni par l'utilisation de normes, comme par exemple

$$|x|, \quad \sqrt{\sum_{k=1}^n x_k^2}, \quad \sup_{k \in \llbracket 1, n \rrbracket} |x_k|, \quad \sup_{x \in I} |f(x) - g(x)|$$

ou, si f et g sont des fonctions continues sur I , $\sqrt{\int_a^b (f(t) - g(t))^2 dt} \dots$

Mais, dans le cas de variables aléatoires, la situation est plus compliquée. Nous envisagerons deux façons d'aborder le problème, qui débouchent chacune sur une notion de convergence à la base d'un théorème très important des probabilités.

1. Convergence en probabilité

1.1 Analyse du problème

Une variable aléatoire réelle X définie sur un espace probabilisé (Ω, \mathcal{T}, P) étant une application de Ω dans \mathbb{R} , il serait naturel de considérer qu'une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires définies sur (Ω, \mathcal{T}, P) converge vers une variable aléatoire X définie sur le même espace probabilisé si, et seulement si, pour tout élément ω de Ω , la suite réelle $(X_n(\omega))_{n \in \mathbb{N}}$ converge vers le réel $X(\omega)$, c'est-à-dire si la fonction X_n converge simplement en tout point vers la fonction X .

Malheureusement, cette définition, qui ne fait d'ailleurs pas intervenir la probabilité sous-jacente, est très restrictive et conduit à des résultats parfois contraires à ce qu'une bonne intuition permettrait d'attendre.

Si l'on note X_n la variable indicatrice de l'intervalle $]0, \frac{1}{n+1}]$ pour la probabilité uniforme sur le segment $[0, 1]$, elle admet comme limite au sens ci-dessus la variable certaine nulle $X = 0$ comme on le voit en faisant décrire $[0, 1]$ au réel ω .

Par contre la variable indicatrice Y_n de l'intervalle $[0, \frac{1}{n+1}]$ n'a pas la même limite en ce sens que X_n , bien que les deux intervalles ne diffèrent que par le point 0 tel que $P(\{0\}) = 0$ et $Y_n(0) = 1$ pour tout n .

En fait, ces deux suites convergent bien vers 0, mais pour la seconde seulement au sens de la *convergence presque sûre* définie plus loin.

Pour « adoucir » un peu la condition ci-dessus, on peut envisager de *négliger* l'ensemble des ω pour lesquels la suite $(X_n(\omega))_{n \in \mathbb{N}}$ ne converge pas vers le réel $X(\omega)$ dès lors que sa probabilité est nulle. La probabilité de l'ensemble des ω pour lesquels la suite réelle $(X_n(\omega))_{n \in \mathbb{N}}$ converge vers le réel $X(\omega)$ est alors égale à 1. Comme un événement de probabilité nulle est dit *presque impossible*, et qu'un événement de probabilité 1 est dit *presque certain*, on dira que la convergence obtenue est *presque sûre*.

Définition 1

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur un même espace probabilisé (Ω, \mathcal{T}, P) , et soit X une variable aléatoire elle aussi définie sur le même espace.

On dit que la suite $(X_n)_{n \in \mathbb{N}}$ **converge presque sûrement** vers la variable aléatoire X , et l'on note $X_n \xrightarrow{p.s.} X$ si

$$P(\{\omega \in \Omega \mid \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\}) = 1.$$

► Remarques

- On parle également de convergence *forte*.
- Cette notion de convergence presque sûre est fondamentale, mais ne figure pas au programme. Les quelques passages de ce chapitre où elle intervient doivent donc être considérés comme fournissant des exercices féconds sur la théorie des variables aléatoires.

- Les objectifs principaux étant l’initiation à l’estimation et le théorème de la limite centrée, le programme comporte seulement deux notions de convergence totalement différentes. L’une d’entre elles procède d’un nouvel affaiblissement des conditions de la convergence presque sûre. On obtient ainsi la notion de convergence stochastique, qui fait l’objet de la suite de cette section.

1.2 Définition

Lorsque la suite réelle $(X_n(\omega))_{n \in \mathbb{N}}$ converge vers le réel $X(\omega)$, on doit s’attendre à ce que pour tout réel strictement positif ε , il existe un entier n_0 tel que, pour tout entier n supérieur à n_0 , $|X_n(\omega) - X(\omega)|$ soit inférieur à ε . Dans la droite ligne du raisonnement tenu au 1.1.1, on peut ainsi avoir l’idée d’exprimer que la probabilité de tout événement de la forme $[|X_n - X| > \varepsilon]$, où ε est un réel strictement positif, tend vers 0 quand n tend vers l’infini.

Définition 2

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur un même espace probabilisé (Ω, \mathcal{T}, P) , et soit X une variable aléatoire elle aussi définie sur le même espace.

On dit que la suite $(X_n)_{n \in \mathbb{N}}$ **converge en probabilité** (ou **stochastiquement**) vers la variable aléatoire X , et l’on note $X_n \xrightarrow{\mathcal{P}} X$, si pour tout réel strictement positif ε , la probabilité de l’événement $(|X_n - X| > \varepsilon)$ tend vers 0 quand n tend vers l’infini.

$$X_n \xrightarrow{\mathcal{P}} X \iff \forall \varepsilon \in \mathbb{R}_+^* \quad \lim_{n \rightarrow +\infty} P(|X_n - X| > \varepsilon) = 0.$$

► Remarques

- On parle également de convergence *stochastique*.
- La condition $\lim_{n \rightarrow +\infty} P(|X_n - X| > \varepsilon) = 0$ équivaut à la condition $\lim_{n \rightarrow +\infty} P(|X_n - X| \leq \varepsilon) = 1$.
- Il résulte clairement de la définition que $|X_n| \xrightarrow{\mathcal{P}} 0$ si, et seulement si, $X_n \xrightarrow{\mathcal{P}} 0$, et que $X_n - X \xrightarrow{\mathcal{P}} 0$ si, et seulement si, $X_n \xrightarrow{\mathcal{P}} X$.
- Dans le cas des convergences de suites de variables aléatoires, les notations rencontrées sont nombreuses et variées. On rencontre en particulier pour la convergence en probabilité : $X_n \xrightarrow{\mathcal{P}} X$, $X_n \xrightarrow{\mathcal{S}} X$, $X_n \xrightarrow{\mathcal{E}} X$, ou encore $X_n \xrightarrow{\mathcal{S}} X$, $X_n \xrightarrow{\mathcal{E}} X$, voire $X_n \xrightarrow[\text{en probabilité}]{} X$, $X_n \xrightarrow[n \rightarrow +\infty]{} X$ ou $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{S}} X$.

Exemples

1. On considère n variables aléatoires indépendantes $(B_k)_{1 \leq k \leq n}$ suivant une loi de Bernoulli de paramètre p . On note alors $Y_n = \sum_{k=1}^n B_k$ et $X_n = \frac{1}{n} Y_n$. La variable aléatoire Y_n suit une loi binomiale de paramètres n et p . Elle prend ses valeurs dans $[0, n]$, son espérance est np et sa variance $np(1-p)$. Pour tout réel ε strictement positif, l’événement $A = [|X_n - p| > \varepsilon]$, est égal à l’événement $[|Y_n - np| > n\varepsilon]$. Calculons alors la variance de Y_n

$$V(Y_n) = E((Y_n - np)^2) = \sum_{k=0}^n (k - np)^2 P(Y_n = k).$$

Tous les termes de cette somme sont positifs. Donc, en restreignant la somme aux éléments de l'événement A , on obtient

$$V(Y_n) \geq \sum_{k \in A} (k - np)^2 P(Y_n = k).$$

Mais, pour tous les éléments k de A , $(k - np)^2 \geq (n\varepsilon)^2$. Donc

$$V(Y_n) \geq (n\varepsilon)^2 \sum_{k \in A} P(Y_n = k)$$

c'est-à-dire $V(Y_n) \geq (n\varepsilon)^2 P(A)$.

On en conclut que $P(|X_n - p| > \varepsilon) \leq \frac{V(Y_n)}{n^2 \varepsilon^2}$ soit $P(|X_n - p| > \varepsilon) \leq \frac{np(1-p)}{n^2 \varepsilon^2}$. Il en résulte clairement que $\lim_{n \rightarrow +\infty} P(|X_n - p| > \varepsilon) = 0$, et donc que la suite (X_n) converge en probabilité vers la variable aléatoire constante égale à p (on écrit $X_n \xrightarrow{\mathcal{P}} p$).

NB : ce type de minoration sera repris plus loin pour justifier l'inégalité de Bienaymé-Tchebychev, partiellement établie en première année.

- On considère une suite de variable aléatoires (U_n) suivant une loi uniforme à densité sur l'intervalle $[0,1]$. On note X_n la variable aléatoire égale à la plus grande des valeurs prises par les $(U_k)_{1 \leq k \leq n}$ ($X_n = \max_{k \in \llbracket 1, n \rrbracket} (U_k)$).

Soit ε strictement positif, considérons l'événement

$$B = [|X_n - 1| > \varepsilon] = [X_n < 1 - \varepsilon].$$

Pour qu'il soit réalisé, il faut et il suffit que toutes les variables aléatoires U_k prennent une valeur strictement inférieure à $1 - \varepsilon$. Comme ces variables aléatoires sont indépendantes, on a

$$P(|X_n - 1| > \varepsilon) = P\left[\bigcap_{k=1}^n (U_k < 1 - \varepsilon)\right] = \prod_{k=1}^n P(U_k < 1 - \varepsilon) = (1 - \varepsilon)^n.$$

Comme $0 \leq 1 - \varepsilon < 1$, il en résulte que $\lim_{n \rightarrow +\infty} P(|X_n - 1| > \varepsilon) = 0$, et donc que $X_n \xrightarrow{\mathcal{P}} 1$.

1.3 Exercice corrigé

Cette sous-section est consacrée à la démonstration d'un résultat, qui peut paraître évident, au regard du raisonnement que nous avons tenu. Cet exercice est exemplaire à de nombreux égards. De nombreuses démonstrations relatives aux convergences de suites de variables aléatoires peuvent en effet être présentées sous la même forme.

Exercice 1.

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur un même espace probabilisé (Ω, \mathcal{T}, P) , et soit X une variable aléatoire elle aussi définie sur le même espace.

On suppose que $X_n \xrightarrow[p.s.]{} X$. Prouver que $X_n \xrightarrow{\mathcal{P}} X$.

Soient un réel $\varepsilon > 0$, $A = \{\omega \in \Omega \mid X_n(\omega) \rightarrow X(\omega)\}$ et, pour tout entier n , $A_n = [|X - X_n| \leq \varepsilon]$, $B_n = \bigcap_{m=n}^{+\infty} A_m \subset A_n$, tous éléments de \mathcal{T} .

Pour majorer $P(A) = 1$, montrons l'inclusion $A \subset \bigcup_{n=0}^{+\infty} B_n$.

Soit donc ω un élément quelconque de A , il existe un n tel que, pour tout $m \geq n$, on ait la majoration $|X(\omega) - X_m(\omega)| \leq \varepsilon$, d'où la relation $\omega \in B_n$. L'inclusion annoncée en résulte.

La suite (B_n) étant croissante, d'après le théorème de la limite monotone

$$\lim_{n \rightarrow +\infty} P(B_n) = P\left(\bigcup_{k=0}^{+\infty} B_k\right) \geq P(A) = 1.$$

Par suite $\lim_{n \rightarrow +\infty} P(B_n) = 1$ et, *a fortiori*, $\lim_{n \rightarrow +\infty} P(A_n) = 1$.

► **Remarques**

- Il est à noter que la réciproque est fautive. Considérons par exemple encore l'univers $\Omega = [0, 1]$ muni de l'algèbre \mathcal{A} des boréliens et de la probabilité uniforme P . L'espérance mathématique de la variable indicatrice X d'un segment $[a, b]$ inclus dans $[0, 1]$ est alors égale à $P(X = 1)$, donc à $b - a$.
 Pour tout entier naturel non nul n , on appelle X_n la variable indicatrice de l'intervalle $I_n = \left] \frac{n - 2^p}{2^p}, \frac{n - 2^p + 1}{2^p} \right]$, où $p = \left\lfloor \frac{\ln n}{\ln 2} \right\rfloor$ (c'est-à-dire la partie entière du logarithme de base 2 de n).
 La suite (I_n) peut être découpée en paquets consécutifs pour lesquels p est donné et n varie de 2^p à $2^p + 2^p - 1$; les 2^p intervalles de I_n ainsi définis forment une partition de $]0, 1]$. Pour p fixé, tout $\omega \in]0, 1]$ appartient à l'un de ces I_n et à un seul; l'ensemble des valeurs de $X_n(\omega)$ est donc une suite de 0 et de 1 telle que, pour tout entier N , il existe un $u > N$ tel que $X_u(\omega) = 0$ et un $v > N$ tel que $X_v(\omega) = 1$. Par conséquent la suite $(X_n(\omega))$ est divergente pour tout $\omega > 0$, et elle ne peut converger presque sûrement vers la variable certaine nulle $X = 0$. Pourtant, comme X_n est la variable indicatrice positive d'un intervalle dont la longueur 2^{-p} tend vers zéro, alors l'espérance de X_n tend vers 0 et X_n converge en probabilité vers la variable certaine nulle (voir le a) du 1.7), ce qui démontre la fausseté de la réciproque.

1.4 Problème de l'unicité de la limite

Dans le cas de limites de suites réelles, de suites de vecteurs, ou de suites de fonctions, si une suite converge, sa limite est unique. Dans le cas de la convergence en probabilité (stochastique) de suites de variables aléatoires, la situation est plus compliquée. On dispose seulement du théorème suivant

Théorème 1

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur un même espace probabilisé (Ω, \mathcal{T}, P) , et soient X et X' des variables aléatoires définies sur le même espace et telles que $X_n \xrightarrow{P} X$ et $X_n \xrightarrow{P} X'$. Alors $P(X \neq X') = 0$.

Preuve

Pour tout réel ε strictement positif, comme $|X - X'| \leq |X - X_n| + |X_n - X'|$, l'événement $[|X - X'| > 2\varepsilon]$ implique l'événement $[|X - X_n| > \varepsilon] \cup [|X_n - X'| > \varepsilon]$ (en effet, la somme de deux réels positifs ne peut pas être supérieure à 2ε sans qu'au moins l'un des deux soit supérieur à ε).

Il en résulte que, pour tout entier n , on dispose des relations

$$[|X - X'| > 2\varepsilon] \subset ([|X_n - X| > \varepsilon] \cup [|X_n - X'| > \varepsilon]),$$

$$P(|X - X'| > 2\varepsilon) \leq P(|X_n - X| > \varepsilon) + P(|X_n - X'| > \varepsilon).$$

Soit maintenant un réel $\delta > 0$. Il existe au moins un entier n tel que l'on ait à la fois

$$P(|X_n - X| > 2\varepsilon) < \delta, \quad P(|X_n - X'| > 2\varepsilon) < \delta.$$

Donc, $\forall \delta > 0, P(|X - X'| > 2\varepsilon) < 2\delta$, d'où $\forall \varepsilon > 0, P(|X - X'| > 2\varepsilon) = 0$.

D'autre part, pour tout $\omega \in \Omega$ tel que $X(\omega) \neq X'(\omega)$, il existe un entier n tel que $n|X(\omega) - X'(\omega)| > 1$, d'où l'inclusion

$$[X \neq X'] \subset \bigcup_{n=1}^{+\infty} \left[|X - X'| > \frac{1}{n} \right].$$

Comme l'inclusion inverse est une évidence, on peut exprimer l'événement $[X \neq X']$ sous la forme

$$[X \neq X'] = \bigcup_{n=1}^{+\infty} \left[|X - X'| > \frac{1}{n} \right].$$

Or, pour $n > n'$, $\left[|X - X'| > \frac{1}{n} \right] \subset \left[|X - X'| > \frac{1}{n'} \right]$. Il en résulte, d'après la propriété de la limite monotone, que

$$P(X \neq X') = P \left[\bigcup_{n \in \mathbb{N}^*} \left(|X - X'| > \frac{1}{n} \right) \right] = \lim_{n \rightarrow +\infty} P \left(|X - X'| > \frac{1}{n} \right) = 0.$$

La propriété annoncée est ainsi démontrée. □

► **Remarque**

Deux variables aléatoires X et X' telles que $P(X \neq X') = 0$ sont dites **presque sûrement égales pour la probabilité P** (et l'on écrit P-p.s. égales).

Le théorème ci-dessus peut s'énoncer

Une limite d'une suite de variables aléatoires (X_n) qui converge en probabilité est P-p.s. unique.

1.5 Composition par une fonction continue

Théorème 2

Soit f une fonction continue sur \mathbb{R} . Si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires réelles qui convergent en probabilité vers X , alors la suite $(f(X_n))_{n \in \mathbb{N}}$ converge en probabilité vers $f(X)$.

Preuve

Pour démontrer ce théorème, nous *admettrons* que la continuité de f implique que, pour tout réel strictement positif ε , quel que soit l'entier naturel k , il existe un réel strictement positif α tels que, quels que soit le couple (x, y) de réels, on ait

$$[|x| \leq k \text{ et } |f(y) - f(x)| > \varepsilon] \implies [|y - x| > \alpha].$$

(Une démonstration de cette propriété est proposée dans l'exercice ??). La suite d'événements $(\{|X| > n\})_{n \in \mathbb{N}}$ est décroissante (en effet, $n > n' \Rightarrow \{|X| > n\} \subset \{|X| > n'\}$).

Il est d'autre part clair que $\bigcap_{n=0}^{+\infty} \{|X| > n\} = \emptyset$.

Du théorème de la limite monotone on déduit que $\lim_{n \rightarrow +\infty} P(|X| > n) = 0$.

Soient alors un entier k , et deux réels strictement positifs δ et ε .

D'après la propriété admise, il existe un réel strictement positif α tel que, si $|X| \leq k$ et $|f(X_n) - f(X)| > \varepsilon$, alors $|X_n - X| > \alpha$. On en déduit qu'il existe un réel strictement positif α tel que

$$\{|f(X_n) - f(X)| > \varepsilon\} \subset (\{|X - X_n| > \alpha\} \cup \{|X| > k\})$$

et donc tel que

$$P(\{|f(X_n) - f(X)| > \varepsilon\}) \leq P(\{|X - X_n| > \alpha\}) + P(\{|X| > k\}).$$

Or

- D'après ce qui précède, il existe un entier k tel que $P(|X| > k) < \delta$.
- Comme $X_n \xrightarrow{\mathcal{L}} X$, alors pour tout α , il existe un entier n_0 tel que, pour tout entier n supérieur à n_0 , on ait $P(|X - X_n| > \alpha) < \delta$.
- On en déduit donc que, quel que soit le réel strictement positif δ , il existe un entier n_0 tel que, pour tout entier n supérieur à n_0 , on ait

$$0 \leq P(|f(X_n) - f(X)| > \varepsilon) < 2\delta.$$

La propriété annoncée en résulte. □

Conséquences

Soient $(X_{1,n}), (X_{2,n}), \dots, (X_{p,n})$ p suites de variables aléatoires réelles qui convergent en probabilité respectivement vers des variables aléatoires réelles X_1, X_2, \dots, X_p . On dit alors que la suite de vecteurs aléatoires $(\mathcal{X}_n) = (X_{1,n}, X_{2,n}, \dots, X_{p,n})$ converge en probabilité vers le vecteur aléatoire $\mathcal{X} = (X_1, X_2, \dots, X_p)$.

On dispose alors du théorème

Théorème 3

Soit φ une fonction continue de \mathbb{R}^p dans \mathbb{R} . Si $(\mathcal{X}_n)_{n \in \mathbb{N}}$ est une suite de vecteurs aléatoires qui converge en probabilité vers le vecteur aléatoire \mathcal{X} , alors la suite $(\varphi(\mathcal{X}_n))_{n \in \mathbb{N}}$ converge en probabilité vers $\varphi(\mathcal{X})$.

Preuve

La démonstration de ce théorème est exactement la même que celle du théorème ci-dessus, à cela près que la valeur absolue est remplacée (lorsque nécessaire) par la norme euclidienne sur \mathbb{R}^p (par exemple), les intervalles fermés bornés devenant simplement des « fermés bornés » de \mathbb{R}^p . La propriété admise doit également être adaptée de la même façon, et reste vraie. □

Corollaire 1

Quelles que soient les suites $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles, et quels que soient les réels λ et μ , alors, si $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ convergent en probabilité respectivement vers X et vers Y , alors

- $(\lambda X_n + \mu Y_n)$ converge en probabilité vers $\lambda X + \mu Y$,

- $(X_n Y_n)$ converge en probabilité vers XY ,
et, si $P(Y = 0) = 0$,
- $\frac{X_n}{Y_n}$ converge en probabilité vers $\frac{X}{Y}$.

Preuve

Pour les deux premiers points, il suffit d'appliquer le théorème ci-dessus avec les fonctions φ définies respectivement par $\varphi(x, y) = x + y$ et $\varphi(x, y) = xy$, qui sont des fonctions de \mathbb{R}^2 dans \mathbb{R} continues sur leur domaine de définition. Nous admettrons le troisième point. □

► **Remarque**

Il résulte en particulier de ce résultat que

$$X_n \xrightarrow{P} X \text{ si, et seulement si, } X_n - X \xrightarrow{P} 0.$$

1.6 Inégalités de Markov et Bienaymé-Tchebychev

Pour démontrer une convergence en probabilité, on a souvent recours à des inégalités. Les règles élémentaires des probabilités fournissent de telles inégalités (comme celles que nous avons utilisées dans les démonstrations des théorèmes de cette section).

Parmi les inégalités probabilistes qui jouent un grand rôle théorique, on distingue les inégalités de Markov et de Bienaymé-Tchebychev. Ces inégalités qui servent essentiellement à démontrer des convergences en probabilité, sont numériquement très mauvaises. Elles ont déjà été introduites, dans un cas particulier, dans le livre de première année, et dans le chapitre sur les variables réelles à densité.

Théorème 4 (Inégalité de Markov)

Soit X une variable aléatoire réelle **positive** admettant une espérance. Pour tout réel ε strictement positif, on a

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

Preuve

Considérons l'événement $D = [X \geq \varepsilon]$, et envisageons deux cas suivant que X est discrète ($X(\Omega) = \{x_0, x_1, \dots, x_n, \dots\}$), ou à densité (soit f_X l'une de ses densités).

- Si X est discrète, on a $E(X) = \sum_{k \geq 0} x_k P(X = x_k)$.

Comme X est supposée positive, on a $E(X) \geq \sum_{\{k | x_k \geq \varepsilon\}} x_k P(X = x_k) \geq \varepsilon P(D)$.

- Si X est à densité (cas déjà traité), on a $E(X) = \int_{-\infty}^{+\infty} t f_X(t) dt$.

Comme X est supposée positive, on a

$$E(X) = \int_0^{+\infty} t f_X(t) dt \geq \int_{\varepsilon}^{+\infty} t f_X(t) dt \geq \varepsilon \int_{\varepsilon}^{+\infty} f_X(t) dt$$

puis $E(X) \geq \varepsilon P(D)$. □

Corollaire 2

Soit X une variable aléatoire réelle positive admettant une espérance. Pour tout réel ε strictement positif, on a

$$P(X > \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

Preuve

Ce résultat découle directement de l'implication $[X > \varepsilon] \subset [X \geq \varepsilon]$. □

Corollaire 3

Soit X une variable aléatoire admettant une variance. Alors, quel que soit le réel strictement positif ε , $P(|X| \geq \varepsilon) \leq \frac{E(X^2)}{\varepsilon^2}$.

Preuve

Il suffit d'appliquer le résultat précédent à la variable aléatoire X^2 et au réel strictement positif ε^2 , en remarquant que $P(X^2 \geq \varepsilon^2) = P(|X| \geq \varepsilon)$, et que, si $V(X)$ existe, alors $E(X^2)$ existe. □

► **Remarque**

Pour tout vecteur aléatoire V , en appliquant le résultat ci-dessus à la variable aléatoire $\|V\|$ (et en supposant qu'elle admette une espérance), on conclut que, pour tout réel strictement positif ε , $P(\|V\| > \varepsilon) \leq \frac{E(\|V\|)}{\varepsilon}$.

Théorème 5 (Inégalité de Bienaymé-Tchebychev)

Soit X une variable aléatoire réelle admettant une variance σ_X^2 . Alors

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\sigma_X^2}{\varepsilon^2}.$$

Preuve

Il suffit d'appliquer l'inégalité de Markov à la variable aléatoire $(X - E(X))$ et au réel strictement positif ε^2 , en remarquant que $P[(X - E(X))^2 \geq \varepsilon^2] = P[|X - E(X)| \geq \varepsilon]$, et que $E[(X - E(X))^2] = \sigma_X^2$. □

1.7 Conditions suffisantes de convergence en probabilité

Proposition 1

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur un même espace probabilisé (Ω, \mathcal{T}, P) . Alors

- (i) si $\lim_{n \rightarrow +\infty} E(X_n) = 0$ et si les variables X_n sont positives, alors $X_n \xrightarrow{P} 0$,
- (ii) si $\lim_{n \rightarrow +\infty} E(X_n) = 0$ et $\lim_{n \rightarrow +\infty} V(X_n) = 0$, alors $X_n \xrightarrow{P} 0$.

Preuve

- (i) En appliquant l'inégalité de Markov à la variable aléatoire positive X_n , on obtient, pour tout réel ε strictement positif

$$P(X_n \geq \varepsilon) \leq \frac{E(X_n)}{\varepsilon}.$$

Comme X_n est positif et $\lim_{n \rightarrow +\infty} E(X_n) = 0$, il en résulte que $\lim_{n \rightarrow +\infty} P(|X_n| > \varepsilon) = 0$ pour tout $\varepsilon > 0$, et donc que $X_n \xrightarrow{\mathcal{P}} 0$.

- (ii) En appliquant l'inégalité de Markov sous la forme du deuxième corollaire à la variable aléatoire X_n , on obtient, pour tout réel ε strictement positif

$$P(|X_n| \geq \varepsilon) \leq \frac{E(X_n^2)}{\varepsilon^2}.$$

D'après la formule de Kœnig-Huygens, $E(X_n^2) = V(X_n) + [E(X_n)]^2$. Les hypothèses conduisent alors à $\lim_{n \rightarrow +\infty} E(X_n^2) = 0$. On conclut alors que $\lim_{n \rightarrow +\infty} P(|X_n| > \varepsilon) = 0$ pour tout $\varepsilon > 0$, et donc que $X_n \xrightarrow{\mathcal{P}} 0$. \square

Corollaire 4

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur un même espace probabilisé (Ω, \mathcal{T}, P) , et soit X une variable aléatoire réelle définie sur le même espace. Alors

- (i) $\lim_{n \rightarrow +\infty} E(|X_n - X|) = 0 \implies X_n \xrightarrow{\mathcal{P}} X$,
- (ii) $\lim_{n \rightarrow +\infty} E(X_n - X) = 0$ et $\lim_{n \rightarrow +\infty} V(X_n - X) = 0 \implies X_n \xrightarrow{\mathcal{P}} X$,
- (iii) $\lim_{n \rightarrow +\infty} E(|X_n - X|^2) = 0 \implies X_n \xrightarrow{\mathcal{P}} X$.

Preuve

- (i) D'après le a) de la proposition ci-dessus, $\lim_{n \rightarrow +\infty} E(|X_n - X|) = 0 \implies |X_n - X| \xrightarrow{\mathcal{P}} 0$.
Donc, d'après la remarque de la définition 2, $\lim_{n \rightarrow +\infty} E(|X_n - X|) = 0 \implies X_n - X \xrightarrow{\mathcal{P}} 0$. On en conclut que $X_n \xrightarrow{\mathcal{P}} X$.
- (ii) D'après le b) de la proposition ci-dessus, les hypothèses entraînent $X_n - X \xrightarrow{\mathcal{P}} 0$, ce qui implique comme au a) que $X_n \xrightarrow{\mathcal{P}} X$.
- (iii) D'après la formule de Kœnig-Huygens, $E(|X_n - X|^2) = V(X_n - X) + [E(X_n - X)]^2$.
Ainsi, $E(|X_n - X|^2)$ étant la somme de deux nombres positifs ne peut tendre vers 0 que si chacun d'eux tend vers 0. On en conclut que $\lim_{n \rightarrow +\infty} E(X_n - X) = 0$ et $\lim_{n \rightarrow +\infty} V(X_n - X) = 0$ et l'on termine en appliquant le **b)**

► **Remarques**

- Ne pas perdre de vue que $\lim_{n \rightarrow +\infty} E(X_n - X) = 0$ si, et seulement si, $\lim_{n \rightarrow +\infty} E(X_n) = E(X)$, mais que cela n'est pas si simple pour la variance.
- Il est important de remarquer que les résultats que nous venons d'obtenir ne concernent que des **conditions suffisantes**. On peut par exemple, trouver des suites $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires qui convergent en probabilité vers une variable X sans pour autant que $E(X_n)$ tende vers $E(X)$, ou même sans que $E(X_n)$ n'admette de limite finie.
 - Considérons par exemple la variable aléatoire X_n définie par $P(X_n = 0) = 1 - \frac{1}{n}$ et $P(X_n = n) = \frac{1}{n}$. On vérifiera sans peine que pour tout entier naturel n non nul $E(X_n) = 1$, et que pourtant $X_n \xrightarrow{\mathcal{P}} 0$.
 - Considérons de même la variable aléatoire Y_n définie par $P(Y_n = 0) = 1 - \frac{1}{n}$ et $P(Y_n = n^2) = \frac{1}{n}$. On vérifiera sans peine que pour tout entier naturel n non nul on a $E(Y_n) = n$, donc que $\lim_{n \rightarrow +\infty} E(Y_n) = +\infty$, et que pourtant $X_n \xrightarrow{\mathcal{P}} 0$.

- Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles.
 - Si $\lim_{n \rightarrow +\infty} E(|X_n - X|) = 0$, on dit que $(X_n)_{n \in \mathbb{N}}$ converge en moyenne vers X .
 - Si $\lim_{n \rightarrow +\infty} E(|X_n - X|^2) = 0$, on dit que $(X_n)_{n \in \mathbb{N}}$ converge en moyenne quadratique vers X .
 - Plus généralement, si $\lim_{n \rightarrow +\infty} E(|X_n - X|^p) = 0$, on dit que $(X_n)_{n \in \mathbb{N}}$ converge en moyenne d'ordre p vers X .

1.8 Exercices corrigés

Exercice 2.

Soit (U_n) une suite de variables aléatoires de même loi uniforme sur $[0, 1]$, et indépendantes. On considère la variable X_n définie par $X_n = \min(U_1, U_2, \dots, U_n)$. Montrer que la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers la variable constante nulle.

- Pour $\varepsilon \geq 1$, $P(X_n > \varepsilon) = 0$.
- Pour $0 < \varepsilon < 1$, on a $[X_n > \varepsilon] = \bigcap_{k=0}^n [U_k > \varepsilon]$ (le plus petit d'un ensemble de nombres est supérieur à ε si, et seulement si, tous ces nombres sont supérieurs à ε).

Comme les U_i sont indépendantes, il en résulte que $P(X_n > \varepsilon) = \prod_{k=1}^n P(U_k > \varepsilon)$.

Or, pour tout k , $P(U_k > \varepsilon) = 1 - P(U_k \leq \varepsilon) = 1 - \varepsilon$.

Donc, $P(X_n > \varepsilon) = (1 - \varepsilon)^n$.

Comme $0 < \varepsilon < 1$, on a $0 < 1 - \varepsilon < 1$, et donc $\lim_{n \rightarrow +\infty} (1 - \varepsilon)^n = 0$.

On conclut donc que $\lim_{n \rightarrow +\infty} P(X_n > \varepsilon) = 0$ pour tout $\varepsilon > 0$, ce qui prouve que

$$\boxed{X_n \xrightarrow{\mathcal{P}} 0.}$$

Exercice 3.

- a) Soit X une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{A}, P) . On note F la fonction de répartition de X .

Pour tout réel x , on note $\mathbf{1}_{]-\infty, x]}$ la variable indicatrice de l'intervalle $] -\infty, x]$. On note aussi $Y = \mathbf{1}_{]-\infty, x]}(X)$ (notation usuelle, rappelons le, pour $Y = \mathbf{1}_{]-\infty, x]} \circ X$).

Quelle est la signification de la variable Y , et quelle loi suit-elle ?

- b) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi que X . Pour

tout réel x , on note $F_{n,x}$ la variable aléatoire définie par $F_{n,x} = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{]-\infty, x]}(X_k)$.

Que représente la variable aléatoire $F_{n,x}$? Montrer qu'elle converge en probabilité vers la variable certaine égale à $F(x)$.

- a) Y est la variable aléatoire qui prend la valeur 1 si X prend une valeur de l'intervalle $] -\infty, x]$, et 0 dans le cas contraire. Elle suit donc une loi de Bernoulli de paramètre $P(X \in] -\infty, x]) = F(x)$.

On a donc $E(Y) = F(x)$, et $V(X) = F(x)(1 - F(x))$.

- b) $F_{n,x}$ représente la proportion des variables (X_1, \dots, X_n) qui prennent leur valeur dans l'intervalle $] -\infty, x]$.

On a clairement $E(F_{n,x}) = F(x)$ et $V(F_{n,x}) = \frac{1}{n}F(x)(1 - F(x))$.

Donc, $\lim_{n \rightarrow +\infty} E(F_{n,x}) = F(x)$ et $\lim_{n \rightarrow +\infty} V(F_{n,x}) = 0$.

Il en résulte que (condition suffisante de convergence) $F_{n,x} \xrightarrow{\mathcal{P}} F(x)$.

2. Lois des grands nombres

2.1 Loi faible des grands nombres

Théorème 6

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur un même espace probabilisé (Ω, \mathcal{T}, P) , de même espérance $E(X_n) = m$ et de même variance $V(X_n) = \sigma^2$, et deux à deux non corrélées.

Alors la variable aléatoire Y_n définie par $Y_n = \frac{1}{n} \sum_{k=1}^n X_k$ converge en probabilité vers la variable certaine égale à m .

Preuve

Pour appliquer l'inégalité de Bienaymé-Tchebychev, calculons l'espérance et la variance de Y_n .

$$E(Y_n) = \frac{1}{n} \sum_{k=1}^n E(X_k) = \frac{1}{n} \sum_{k=1}^n m = \frac{nm}{n} = m$$

$$V(Y_n) = V\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} \left[\sum_{k=1}^n V(X_k) + \sum_{k \neq k'} \text{Cov}(X_k, X_{k'}) \right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

En appliquant l'inégalité de Bienaymé-Tchebychev à la variable aléatoire Y_n , on obtient, pour tout réel strictement positif ε , $P(|Y_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$.

Il en résulte que $\lim_{n \rightarrow +\infty} P(|Y_n - m| \geq \varepsilon) = 0$, et donc que $Y_n \xrightarrow{\mathcal{P}} m$. \square

► Remarques

- Rappelons que deux variables aléatoires sont non corrélées lorsque leur covariance est nulle. Deux variables aléatoires indépendantes sont non corrélées, mais la réciproque n'est pas vraie.
- Le théorème ci-dessus n'oblige nullement les variables X_k à être indépendantes, ni à suivre la même loi, mais il s'applique naturellement lorsque les variables X_k sont deux à deux indépendantes et de même loi. On dit alors que pour tout n , le n -uplet $(X_k)_{k \in [1, n]}$ est *iid* (indépendant et identiquement distribué). On se reportera au chapitre consacré à l'estimation.
- Il existe d'autres formes de la loi des grands nombres. L'une d'entre elle en particulier s'applique au cas où les variables aléatoires X_k admettent simplement une espérance, mais suivent la même loi de probabilité. On montre alors que Y_n converge presque sûrement vers la variable certaine égale à m . On parle alors de **loi forte des grands nombres**. Ce résultat dépasse le cadre du programme.

- On rencontre çà et là l'opinion selon laquelle ces lois des grands nombres constituent une « justification » de la notion de probabilité. Une théorie mathématique n'a pas besoin de justification. Elle constitue un développement rationnel de résultats reliés entre eux et obtenus à partir d'un système d'axiomes retenus *a priori*. Tout au plus remarquera-t-on que les résultats ainsi obtenus sont en accord avec ceux que l'on obtient par l'expérimentation ou certaines simulations.
On peut sans doute en déduire que le système d'axiomes de la théorie des probabilités a été bien choisi, mais rien ne permet de penser que cette coïncidence partielle avec l'expérimentation sera confirmée par d'autres coïncidences, ni qu'avec un autre système d'axiomes, on ne serait pas arrivé au même résultat ou même globalement à de meilleurs résultats.

2.2 Théorème de Bernoulli

Théorème 7

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables de Bernoulli indépendantes et de même paramètre p ($p \in]0, 1[$), et soit Y_n la variable aléatoire définie par $Y_n = \frac{1}{n} \sum_{k=1}^n X_k$. Alors, pour tout $\varepsilon > 0$,

$$P(|Y_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$$

et Y_n converge en probabilité vers la variable certaine égale à p .

Preuve

Pour tout entier k de $[1, n]$, on a $E(X_k) = p$ et $V(X_k) = p(1 - p)$. En appliquant le résultat obtenu dans la démonstration du théorème ci-dessus, on obtient donc

$$P(|Y_n - p| \geq \varepsilon) \leq \frac{p(1 - p)}{n\varepsilon^2}.$$

Or l'on vérifie sans peine que le fonction $[x \mapsto x(1 - x)]$ admet, pour $x = \frac{1}{2}$ un maximum absolu égal à $\frac{1}{4}$.
Donc, pour tout p de l'intervalle $]0, 1[$, $p(1 - p) \leq \frac{12}{4}$, et donc, $P(|Y_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$. En faisant tendre n vers l'infini, on conclut que Y_n converge en probabilité vers la variable certaine égale à p . \square

► Remarque

Ce résultat n'est autre que la loi faible des grands nombres appliquée à une suite $(X_n)_{n \in \mathbb{N}}$ de variables de Bernoulli indépendantes et de même paramètre p . Il apporte quand même une précision concernant la majoration de $P(|Y_n - p| \geq \varepsilon)$ par $\frac{1}{4n\varepsilon^2}$, qui ne dépend pas de p . On pourra ainsi déterminer des intervalles contenant p avec une probabilité donnée.

2.3 Exercice corrigé

Exercice 4. L'aiguille de Buffon

Un parquet est constitué de lames parallèles de largeur d . On laisse tomber (au hasard) sur ce parquet une aiguille de longueur ℓ . Si l'on suppose de plus que $\ell < d$, l'aiguille peut, soit être complètement sur l'une des lames du parquet, soit chevaucher deux lames. Soit p la probabilité pour que l'aiguille rencontre deux lames du parquet. Le calcul théorique de p conduit à $p = \frac{2\ell}{\pi d}$ (résultat attribué à Georges Louis Leclerc comte de Buffon, et daté de

1777). On répète n fois l'expérience, de manière indépendante, et l'on note X_k la variable aléatoire égale à 1 si, au k -ème jet l'aiguille rencontre deux lames de parquet, et à 0 sinon. On note alors $Y_n = \sum_{k=1}^n X_k$ et $F_n = \frac{Y_n}{n}$.

Montrer que l'on peut trouver n tel que l'intervalle aléatoire $]F_n - 0,05; F_n + 0,05[$ contienne p avec une probabilité supérieure à 0,99. Donner une valeur pour n lorsque $\ell=2,2$ cm et $d=7$ cm.

La loi des grands nombres montre que F_n converge en probabilité vers la variable certaine égale à p . Donc, pour tout réel strictement positif ε , $\lim_{n \rightarrow \infty} P(|F_n - p| \geq \varepsilon) = 0$.

Donc, quel que soit δ , il existe n_0 tel que, pour tout entier n supérieur à n_0 , on ait $P(|F_n - p| \geq \varepsilon) \leq \delta$.

La condition de l'énoncé se traduit par $P(|F_n - p| \geq 0,05) \leq 0,01$, c'est-à-dire par $P(|Y_n - np| \geq 0,05n) \leq 0,01$.

Or, Y_n suit une loi binomiale de paramètres n et p . Son espérance est donc $E(Y_n) = np$. En appliquant l'inégalité de Bienaymé Tchebychev, on obtient

$$P(|Y_n - np| \geq 0,05n) = P(|Y_n - E(Y_n)| \geq 0,05n) \leq \frac{V(X_n)}{n^2(0,05)^2}.$$

$$\text{Soit } P(|Y_n - np| \geq 0,05n) \leq \frac{np(1-p)}{n^2(0,05)^2}.$$

Sachant que $p = \frac{2 \times 2,2}{\pi \times 7} \approx 0,2$, on a $V(Y_n) \approx 0,16n$.

La condition de l'énoncé sera remplie dès que $\frac{0,16n}{0,0025n^2} \leq 0,01$, c'est-à-dire $n \geq 6400$.

Si l'on ne connaît pas p , on applique directement le théorème de Bernoulli (on est dans le cas où ce théorème s'applique). On obtient alors
$$P(|F_n - p| \geq 0,05) \leq \frac{1}{4n \times 0,0025}.$$

La condition de l'énoncé sera donc remplie dès que $\frac{1}{4n \times 0,0025} \leq 0,01$, c'est-à-dire, après calcul, dès que $n \geq 10\,000$.

Il n'est pas étonnant que cette deuxième évaluation de n soit plus mauvaise que la première, le théorème de Bernoulli ayant été obtenu à partir de la loi des grands nombres en majorant la variance, pour toutes les valeurs de p .

En tout état de cause, ces minorations sont très grossières.

3. Convergence en loi

3.1 Définition

Considérons toujours une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles, et X une variable aléatoire réelle. Une autre façon d’aborder le problème de la convergence de la suite $(X_n)_{n \in \mathbb{N}}$ vers la variable aléatoire X est de considérer qu’une variable aléatoire réelle est définie par la donnée de sa fonction de répartition F_{X_n} (parfois notée plus simplement F_n), et d’envisager la convergence pour tout réel x de la suite $F_{X_n}(x)$ vers $F_X(x)$.

Mais, cette conception de la convergence d’une suite de variables aléatoires conduirait à des paradoxes difficiles à gérer. Par exemple, il semble naturel de considérer que si X_n est la fonction indicatrice du singleton $\left\{ \frac{1}{n} \right\}$ (c’est à dire la variable aléatoire certaine égale à $\frac{1}{n}$), alors la suite $(X_n)_{n \in \mathbb{N}}$ converge vers la variable aléatoire certaine égale à 0, ce qui ne serait pas le cas avec cette façon de voir (ainsi que le lecteur s’en convaincra aisément).

On donne alors la définition plus souple suivante

Définition 3

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles. On note F_{X_k} la fonction de répartition de X_k .

On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers la variable aléatoire réelle X de fonction de répartition F_X si la suite $(F_{X_n}(x))$ converge vers $F_X(x)$ **pour tout x où F_X est continue.**

On note $X_n \xrightarrow{\mathcal{L}} X$.

► **Remarques**

- On parle également de convergence *faible*, *étroite* ou *en distribution*.
- Si $D_c(F_X)$ désigne le domaine de continuité de la fonction de répartition F_X de la variable aléatoire X , $X_n \xrightarrow{\mathcal{L}} X$ signifie que $\forall x \in D_c(F_X), \lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$.
- Notons $F_X(x-0)$ la limite de x de $F_X(t)$ lorsque t tend vers x par valeurs inférieures. Alors la continuité de F_X en x s’écrit $F_X(x) = F_X(x-0)$ puisque F_X est continue à droite, et équivaut donc aux relations

$$0 = F_X(x) - F_X(x-0) = P(X = x).$$

- En réalité, la notion de convergence en loi n’est pas relative aux variables aléatoires en elles mêmes, mais concerne les lois des variables aléatoires. C’est une notion de convergence dans l’ensemble des fonctions de répartition de variables aléatoires, c’est à dire des fonctions F croissantes sur \mathbb{R} , de classe \mathcal{C}^1 sur \mathbb{R} privé d’un nombre fini x_1, x_2, \dots, x_k d’éléments où elle admet une limite à droite et une limite à gauche, telles que $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Dans cet ensemble figurent des lois discrètes, des lois à densité, mais aussi des lois mixtes. On pourra rencontrer des suites de variables aléatoires de lois discrètes convergeant vers une loi continue, ou vers une loi mixte, ou même des suites de lois continues convergeant vers une loi discrète, ou mixte. . .

- On rencontre les notations analogues à celles de la convergence en probabilité, qui sont multiples et variées. On rencontre même des notations du genre $X_n \xrightarrow{\mathcal{L}} \mu$ où μ désigne une loi de probabilité, c’est-à-dire un objet mathématique qui n’est pas une variable aléatoire, mais désigne un ensemble de variables aléatoires.

Proposition 2

Soient (X_n) et (Y_n) deux suites de variables aléatoires réelles, de fonctions de répartition F_{X_k} et F_{Y_k} , convergeant en loi vers la même variable aléatoire réelle Z de fonction de répartition F_Z .

Alors, en tout point x où F_Z est continue, la suite numérique $(F_{X_n}(x) - F_{Y_n}(x))$ converge vers 0.

Preuve

Il suffit d'écrire

$$|F_{X_n}(x) - F_{Y_n}(x)| \leq |F_{X_n}(x) - F_Z(x)| + |F_Z(x) - F_{Y_n}(x)|. \quad \square$$

► **Remarques**

- Les hypothèses de la proposition n'impliquent pas que $X_n - Y_n$ converge en loi vers $Z - Z = 0$, comme le montre l'exemple suivant : soit X une variable de Bernoulli de paramètre $\frac{1}{2}$. La variable $Y = 1 - X$ est également une variable de Bernoulli de paramètre $\frac{1}{2}$, différente de X mais de même loi de répartition que X . Posons, pour tout entier n , $X_n = X$ et $Y_n = Y$. On a naturellement $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} X$, puisque X et Y ont la même loi, d'où l'existence de $Z = X$.
Mais $|X_n - Y_n| = |2X - 1| = 1$ car, si $X = 1$, alors $2X - 1 = 1$, et si $X = 0$, alors $2X - 1 = -1$. Par suite $F_{X_n - Y_n}(-1) = \frac{1}{2}$ et $X_n - Y_n$ ne converge pas en loi vers la variable aléatoire certaine nulle.
- Il en résulte que les convergences en loi respectives de (X_n) et (Y_n) vers X et Y n'impliquent pas la convergence en loi de $(X_n - Y_n)$ vers $X - Y$.
- La convergence en loi éventuelle de $(X_n + Y_n)$ vers $X + Y$ n'est pas davantage assurée : il suffit de prendre X comme dans le contre-exemple ci-dessus, $X_n = X$ convergeant en loi vers X , $Y_n = 1 - X_n = 1 - X$ convergeant également en loi vers X puisque $F_{1-X} = F_X$, d'où $X_n + Y_n = 1$, $X + Y = 2X$, bien que $F_{X_n + Y_n}(1) = 1$ ne tende pas vers $F_{2X}(1) = \frac{1}{2}$ alors que 1 est un point de continuité de F_{2X} . Pourtant la proposition ci-dessous donne un exemple très simple dans lequel cette convergence est assurée.

Proposition 3

Soient a un réel et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles convergeant en loi vers une variable aléatoire X . Alors la suite de variables aléatoires $(X_n + a)$ converge en loi vers $X + a$.

Preuve

Notant F, F_n, G et G_n les fonctions de répartitions respectives de $X, X_n, X + a$ et $X_n + a$, on dispose des relations

$$G(x) = P(X + a \leq x) = F(x - a), \quad G_n(x) = P(X_n + a \leq x) = F_n(x - a).$$

La fonction G possède comme points de discontinuité les images de ceux de F par la translation de vecteur a .

Si G est continue en x , c'est-à-dire si F est continue en $x - a$, alors

$$G_n(x) - G(x) = F_n(x - a) - F(x - a) \rightarrow 0$$

ce qui démontre la proposition. □

Exemples

1. Soit X_n une variable aléatoire uniforme sur l'intervalle $\left[-\frac{1}{n}, \frac{1}{n}\right]$. Sa fonction de répartition F_{X_n} est définie pour tout x réel par

$$\begin{cases} F_{X_n}(x) = 0 & \text{si } x \leq -\frac{1}{n} \\ F_{X_n}(x) = \frac{n}{2}x + \frac{1}{2} & \text{si } -\frac{1}{n} \leq x \leq \frac{1}{n} \\ F_{X_n}(x) = 1 & \text{si } \frac{1}{n} \leq x. \end{cases}$$

On constate clairement que, pour tout réel x strictement négatif, $\lim_{n \rightarrow +\infty} F_{X_n}(x) = 0$ et que pour tout réel x strictement positif, $\lim_{n \rightarrow +\infty} F_{X_n}(x) = 1$. Il en résulte que la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers la variable certaine égale à 0 (c'est-à-dire $X_n \xrightarrow{\mathcal{L}} 0$). On rencontre ici un exemple d'une suite de variables aléatoires à densité qui converge en loi vers une variable discrète.

2. En voici un autre exemple : X_n suit la loi exponentielle $\mathcal{E}(n)$, de fonction de répartition donnée par

$$F_n(x) = 0 \quad \text{si } x \leq 0, \quad F_n(x) = 1 - e^{-nx} \quad \text{sinon.}$$

Pour la même raison que ci-dessus, cette suite converge en loi vers la variable certaine égale à 0 (c'est-à-dire $X_n \xrightarrow{\mathcal{L}} 0$).

On pourra noter au passage que les variables suivant une loi exponentielle ou certaine nulle sont les seules lois *sans mémoire*.

3. Soit X_n la variable aléatoire uniforme discrète sur l'ensemble $\left\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right\}$ [pour tout k de $\llbracket 1, n \rrbracket$, $P\left(\frac{k}{n}\right) = \frac{1}{n}$].

Soit x un élément de l'intervalle $]0, 1]$. En notant $[x]$ la partie entière de x , on a $[nx] \leq nx < [nx] + 1$, soit $\frac{[nx]}{n} \leq x < \frac{[nx]}{n} + 1$.

Il en résulte que la fonction de répartition de X_n est donnée, pour tout x réel par

$$\begin{cases} F_{X_n}(x) = 0 & \text{si } x \leq 0 \\ F_{X_n}(x) = \frac{[nx]}{n} & \text{si } 0 < x \leq 1 \\ F_{X_n}(x) = 1 & \text{si } 1 > x \end{cases}$$

Lorsque n tend vers l'infini, on constate alors que X_n converge en loi vers une variable aléatoire X dont la fonction de répartition est donnée, pour tout x réel, par

$$\begin{cases} F_X(x) = 0 & \text{si } x \leq 0 \\ F_X(x) = x & \text{si } 0 \leq x \leq 1 \\ F_X(x) = 1 & \text{si } 1 \leq x \end{cases}$$

c'est-à-dire la variable aléatoire uniforme à densité sur l'intervalle $[0, 1]$.

On rencontre ici un exemple d'une suite de variables aléatoires discrètes qui converge en loi vers une variable continue.

4. Soit X_n une variable aléatoire telle que $P\left(X_n = \frac{1}{n}\right) = 1 - \frac{1}{n}$ et $P(X_n = n) = \frac{1}{n}$. Sa fonction de répartition F_{X_n} est définie pour tout x réel par

$$\begin{cases} F_{X_n}(x) = 0 & \text{si } x \leq \frac{1}{n} \\ F_{X_n}(x) = 1 - \frac{1}{n} & \text{si } \frac{1}{n} \leq x < n \\ F_{X_n}(x) = 1 & \text{si } n \leq x \end{cases}$$

Remarquons que tout x négatif relève du premier cas, et que, pour x strictement positif, $\frac{1}{n} \leq x < n$ si, et seulement si, $n > x$ et $x \geq \frac{1}{n}$, c'est-à-dire pour $n > \max\left(x, \frac{1}{x}\right)$.

Comme $\lim_{n \rightarrow +\infty} 1 - \frac{1}{n} = 1$, on peut conclure que X_n converge en loi vers une variable aléatoire X dont la fonction de répartition F_X est telle que, pour tout x strictement négatif, $F_X(x) = 0$, et pour tout x strictement positif, $F_X(x) = 1$, c'est-à-dire la variable certaine égale à 0. (Remarquons que le cas $x = 0$ n'est pas à étudier car il correspond à une discontinuité de la fonction F_X).

Notons d'autre part, que $E(X) = 0$, et que $E(X_n) = 1 + \frac{1}{n} - \frac{1}{n^2}$, et donc que $\lim_{n \rightarrow +\infty} E(X_n) = 1$.

On rencontre donc ici un exemple d'une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles telles que $X_n \xrightarrow{\mathcal{L}} X$, mais que $\lim_{n \rightarrow +\infty} E(X_n) \neq E(X)$.

3.2 Conditions suffisantes pour une convergence en loi

Le théorème qui ouvre cette partie est, en fait, une condition nécessaire et suffisante de convergence en loi. Sa démonstration pourra être sautée en première lecture, mais elle est très instructive sur le type de majorations/minorations inséparables de beaucoup de manipulations théoriques en calcul des probabilités.

Théorème 8

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur un même espace probabilisé (Ω, \mathcal{T}, P) , et soit X une variable aléatoire elle aussi définie sur le même espace. On suppose que, pour tout entier n , $X(\Omega) = X_n(\Omega) = \mathbb{Z}$.

Dans ces conditions, $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si, et seulement si, pour tout entier relatif i , $\lim_{n \rightarrow \infty} P(X_n = i) = P(X = i)$.

Preuve

Insistons sur le fait que nous sommes dans le cas où $X(\Omega) = X_n(\Omega) = \mathbb{Z}$.

- Supposons que $X_n \xrightarrow{\mathcal{L}} X$. Alors, pour tout entier i , les réels $x = i - \frac{1}{2}$ et $y = i + \frac{1}{2}$, sont des points de continuité de F et donc tels que $F_n(x)$ et $F_n(y)$ convergent respectivement vers $F(x)$ et $F(y)$. Or $P(X_n = i) = F_n(y) - F_n(x)$ et $P(X = i) = F(y) - F(x)$. Le résultat annoncé en découle par passage à la limite.

- Réciproquement, supposons que pour tout entier relatif i , $\lim_{n \rightarrow \infty} P(X_n = i) = P(X = i)$. On note F la fonction de répartition de X et F_n la fonction de répartition de X_n , pour n quelconque.

Soit x un réel quelconque, et soit δ un réel strictement positif. Comme F est croissante, admet 0 comme limite en $-\infty$ et 1 en $+\infty$, il existe un couple de réels (A, B) vérifiant $A < x < B$, $F(A) < \delta$ et $F(B) > 1 - \delta$.

Par suite, $F_n(x) - F_n(A) = \sum_{k \in K} P(X_n = k)$, où K est l'ensemble fini des entiers de l'intervalle $]A, x]$, converge

vers $\sum_{k \in K} P(X = k) = F(x) - F(A)$, et il existe donc un entier n_0 tel que, pour tout entier $n \geq n_0$ on ait

$$F_n(x) - F_n(A) > F(x) - F(A) - \delta,$$

d'où $F_n(x) > F_n(A) + F(x) - F(A) - \delta > 0 + F(x) - 2\delta = F(x) - 2\delta$.

De même, $F_n(x) - F_n(B) = -\sum_{\ell \in L} P(X_n = \ell)$, où L est l'ensemble fini des entiers de l'intervalle $]x, B]$, converge

vers $-\sum_{\ell \in L} P(X = \ell) = F(x) - F(B)$, et il existe un entier n_1 tel que, pour tout entier $n \geq n_1$, on ait

$$F_n(x) - F_n(B) < F(x) - F(B) + \delta,$$

d'où $F_n(x) < F_n(B) + F(x) - F(B) + \delta < 1 + F(x) + 2\delta - 1 = F(x) + 2\delta$.

Il en résulte que, si l'on pose $N = \max(n_0, n_1)$, pour tout entier $n \geq N$, on a $|F_n(x) - F(x)| < 2\delta$ ce qui prouve bien que $X_n \xrightarrow{\mathcal{L}} X$. □

► **Remarques**

- La propriété reste vraie si l'image commune à X et aux X_n n'est plus nécessairement égale à \mathbb{Z} , mais seulement à une partie I de \mathbb{Z} . La démonstration se ramène aussitôt à la précédente en fixant comme nulles les probabilités aux points de $\mathbb{Z} \setminus I$.
- Si I est une partie finie de \mathbb{Z} ou une partie de \mathbb{N} , la démonstration précédente peut être nettement simplifiée, en ne s'appuyant alors que sur le fait que la somme d'une suite finie de fonctions admettant des limites a pour limite la somme de ces limites.
- Dans le cas de variables aléatoires X_n et X prenant leurs valeurs dans \mathbb{R} et admettant des densités **continues** respectivement notées f_n et f , alors, $X_n \xrightarrow{\mathcal{L}} X$ si, et seulement si, $\lim_{n \rightarrow +\infty} f_n(t) = f(t)$ pour presque tout réel t . On pourrait démontrer ce théorème en suivant les grandes lignes de la précédente, mais on aurait à utiliser une propriété concernant l'intégrale d'une suite de fonctions continues de limite nulle, dont ne disposent pas les étudiants des classes préparatoires économiques.
De toutes façons, la portée de ce théorème y serait limitée du fait que beaucoup des variables aléatoires que l'on y rencontre n'admettent que des densités non continues.

En dépit de l'extrême simplicité de son énoncé, la proposition qui suit demande quelques précautions ; sa démonstration pourra naturellement être sautée en première lecture, d'autant plus que cette proposition ne figure pas explicitement au programme. Mais l'expérience prouve qu'elle est parfois utilisée de manière inconsciente au cours d'un exercice : sa présence ici permettra au moins de montrer combien la manipulation de la convergence en loi peut être parfois délicate.

Proposition 4 (Lemme de Slutsky)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires convergeant en loi vers une variable aléatoire X .

Alors la suite $(-X_n)$ converge en loi vers la variable $-X$.

Preuve

Rappelons d'abord que l'ensemble des points de discontinuité de F est au plus dénombrable. (En un tel point x , puisque F y est continue à droite, c'est que $s = F(x) - F(x - 0)$ est strictement positif, c'est-à-dire tel que $P(X = x)$ l'est aussi. Une probabilité étant comprise entre 0 et 1, l'ensemble E_n de tels x tels que $s > \frac{1}{n}$ est fini (de cardinal au plus n), et la réunion dénombrable des E_n est au plus dénombrable, ce qu'il fallait démontrer.)

Par suite, tout intervalle de la forme $[a, a + h[$, étant en bijection avec \mathbb{R} , contient nécessairement au moins un point de continuité de F (et même une infinité non dénombrable).

Étudions maintenant les rapports entre la fonction de répartition F de X et celle, notée G , de $-X$. Les égalités ensemblistes

$$[-X \leq x] = [-X < x] \cup [-X = x] = (\Omega \setminus [X \leq -x]) \cup [X = -x]$$

impliquent l'égalité $G(x) = 1 - F(-x) + P(X = -x)$.

Une relation analogue vaut pour les fonctions F_n et G_n associées à $-X_n$.

Les points de continuité de G sont exactement ceux de $x \mapsto F(-x)$ puisqu'ils sont caractérisés par $P(X = -x) = 0$. Soit x l'un d'entre eux. On dispose de l'égalité

$$G_n(x) - G(x) = F(-x) - F_n(-x) + P(X_n = -x).$$

Soit δ un réel strictement positif. Il existe un entier n_0 tel que $n \geq n_0$ implique $|F(-x) - F_n(-x)| < \delta$ et un réel $\alpha > 0$ tel que $F(-x) - F(-x - t) < \delta$ pour $0 \leq t < \alpha$.

L'ensemble des nombres $\varepsilon > 0$ tels que $-x - \varepsilon$ soit un point de discontinuité de F est dénombrable; il existe donc un $\varepsilon < \alpha$ tel que $-x - \varepsilon$ soit un point de continuité de F et vérifie l'inégalité $F(-x) - F(-x - \varepsilon) < \delta$.

Enfin il existe un entier $N \geq n_0$ tel que, pour tout entier $n \geq N$, on dispose de l'inégalité $F(-x - \varepsilon) - F_n(-x - \varepsilon) < \delta$. Soit donc un tel entier n .

- Majorant $G_n(x) - G(x)$ à l'aide de la relation

$$[X_n \leq -x] \supset ([X_n = -x] \cup [X_n \leq -x - \varepsilon]),$$

d'où $F_n(-x) - P(X_n = -x) \geq F_n(-x - \varepsilon) > F(-x - \varepsilon) - \delta$, il vient

$$G_n(x) - G(x) = F(-x) - F_n(-x) + P(X_n = -x) < F(-x) - F(-x - \varepsilon) + \delta < 2\delta.$$

- Minorant $G_n(x) - G(x)$ par l'inégalité $P(X_n = -x) \geq 0$, il vient

$$G_n(x) - G(x) = F(-x) - F_n(-x) + P(X_n = -x) \geq F(-x) - F_n(-x) > -\delta.$$

Le résultat en découle. □

Proposition 5

Soient a un réel et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires convergeant en loi vers une variable aléatoire X .

Alors la suite (aX_n) converge en loi vers aX .

► **Remarque**

Cette proposition étend la précédente.

Preuve

C'est clair si $a = 0$.

Si $a > 0$, la fonction de répartition de aX , égale à $F\left(\frac{X}{a}\right)$, possède comme points de discontinuité les images de ceux de F par l'homothétie de rapport a ; en dehors de ces points, la convergence de $F_{aX_n}(x)$ vers $F_{aX}(x)$ résulte simplement de la convergence de $F_{X_n}\left(\frac{X}{a}\right)$ vers $F_X\left(\frac{X}{a}\right)$.

Si $a < 0$, il suffit d'appliquer la proposition 4 et le résultat ci-dessus à la suite $(-X_n)$. □

► **Remarque**

Par suite $2X_n$ converge en loi vers $2X$, bien qu'en général les convergences en loi respectives de (X_n) et (Y_n) vers X et Y n'impliquent pas la convergence en loi de $(X_n + Y_n)$ vers $X + Y$.

Proposition 6

Soient a un réel, X une variable aléatoire dont la fonction de répartition F est continue, (X_n) une suite de variables aléatoires de fonctions de répartition (F_n) et (Y_n) une suite de variables aléatoires, toutes définies sur le même espace probabilisé (Ω, \mathcal{T}, P) .

Alors, si la suite (X_n) converge **en loi** vers X et si la suite (Y_n) converge **en probabilité** vers a , la suite $(X_n Y_n)$ converge **en loi** vers aX .

Ce résultat reste valable si l'on écarte l'hypothèse de continuité de F .

► **Remarque**

Cette proposition étend la précédente.

Preuve

(i) Supposons dans un premier temps $a > 0$.

Soient $x \geq 0$ et $\delta > 0$ deux réels. Fixons un $\varepsilon > 0$ strictement inférieur à a tel que l'on ait

$$F\left(\frac{x}{a+\varepsilon}\right) - F\left(\frac{x}{a}\right) > -\delta, \quad F\left(\frac{x}{a-\varepsilon}\right) - F\left(\frac{x}{a}\right) < \delta.$$

Il existe un entier N tel que, pour tout entier $n \geq N$, on ait

$$P(A_n) = P(|Y_n - a| > \varepsilon) < \delta,$$

$$F_n\left(\frac{x}{a+\varepsilon}\right) - F\left(\frac{x}{a+\varepsilon}\right) > -\delta \quad \text{et} \quad F_n\left(\frac{x}{a-\varepsilon}\right) - F\left(\frac{x}{a-\varepsilon}\right) < \delta.$$

Soit enfin $U_n = [X_n Y_n \leq x]$.

- Pour majorer $P(U_n)$, montrons l'inclusion

$$U_n \subset \left(\left[X_n \leq \frac{x}{a-\varepsilon} \right] \cup A_n \right).$$

En effet, si $X_n Y_n \leq x$ et $|Y_n - a| \leq \varepsilon$, alors

$$Y_n \geq a - \varepsilon > 0, \quad Y_n > 0 \quad \text{et} \quad X_n \leq \frac{x}{Y_n} \leq \frac{x}{a-\varepsilon}.$$

Il en résulte

$$P(U_n) \leq F_n\left(\frac{x}{a-\varepsilon}\right) + P(A_n) < F_n\left(\frac{x}{a-\varepsilon}\right) + \delta < F\left(\frac{x}{a-\varepsilon}\right) + 2\delta < F\left(\frac{x}{a}\right) + 3\delta.$$

- Pour minorer $P(U_n)$, montrons l'inclusion

$$\left[X_n \leq \frac{x}{a+\varepsilon} \right] \subset (U_n \cup A_n).$$

En effet, si $X_n \leq \frac{x}{a + \varepsilon}$ et $|Y_n - a| \leq \varepsilon$, alors

$$0 < a - \varepsilon \leq Y_n \leq a + \varepsilon, \quad Y_n > 0 \quad \text{et} \quad X_n Y_n \leq \frac{x Y_n}{a + \varepsilon} \leq x.$$

Il en résulte

$$P(U_n) \geq F_n\left(\frac{x}{a + \varepsilon}\right) - P(A_n) > F_n\left(\frac{x}{a + \varepsilon}\right) - \delta > F\left(\frac{x}{a + \varepsilon}\right) - 2\delta > F\left(\frac{x}{a}\right) - 3\delta.$$

Le résultat en découle, puisque la fonction de répartition de aX est $t \mapsto F\left(\frac{t}{a}\right)$.

Le cas $x < 0$ se traite de manière analogue, à condition de changer légèrement le choix de ε et de considérer les deux inclusions

$$U_n \subset \left(\left[X_n \leq \frac{x}{a + \varepsilon} \right] \cup A_n \right), \quad \left[X_n \leq \frac{x}{a - \varepsilon} \right] \subset (U_n \cup A_n).$$

- (ii) Si $a < 0$, il suffit d'appliquer le résultat précédent à la suite $(-Y_n)$.
- (iii) Supposons enfin $a = 0$. Nous allons montrer qu'alors $(X_n Y_n)$ converge en probabilité (et donc en loi) vers la variable certaine nulle 0.
Soit d'abord $\delta > 0$ un réel. Fixons un $\varepsilon > 0$ tel que l'on ait

$$F\left(-\frac{1}{\varepsilon}\right) < \delta, \quad F\left(\frac{1}{\varepsilon}\right) > 1 - \delta.$$

Il existe un entier N tel que, pour tout entier $n \geq N$, on ait

$$P(B_n) = P(|Y_n| > \varepsilon^2) < \delta,$$

$$F_n\left(-\frac{1}{\varepsilon}\right) - F\left(-\frac{1}{\varepsilon}\right) < \delta \quad \text{et} \quad F_n\left(\frac{1}{\varepsilon}\right) - F\left(\frac{1}{\varepsilon}\right) > -\delta.$$

Soit enfin $V_n = \{|X_n Y_n| > \varepsilon\}$.

Pour majorer le nombre positif $P(V_n)$, montrons l'inclusion

$$V_n \subset \left(B_n \cup \left[X_n < -\frac{1}{\varepsilon} \right] \cup \left[X_n > \frac{1}{\varepsilon} \right] \right).$$

En effet, si $|X_n Y_n| > \varepsilon$ et $|Y_n| \leq \varepsilon^2$, alors

$$|X_n| \varepsilon^2 \geq |X_n Y_n| > \varepsilon \quad \text{d'où} \quad |X_n| > \frac{1}{\varepsilon}.$$

Il en résulte

$$0 \leq P(V_n) \leq P(B_n) + F_n\left(-\frac{1}{\varepsilon}\right) + \left[1 - F_n\left(\frac{1}{\varepsilon}\right)\right] < 3\delta + F\left(-\frac{1}{\varepsilon}\right) + \left[1 - F\left(\frac{1}{\varepsilon}\right)\right] < 5\delta.$$

- (iv) Oublions maintenant la continuité de F . Dans la partie (i) de la preuve ci-dessus, elle intervenait trois fois : d'abord pour dire que F était continue en $\frac{x}{a}$ (mais ceci équivaut au fait que x est un point de continuité de $P(aX \leq x)$ et reste donc acquis), ensuite pour dire que F était continue en les points $\frac{x}{a - \varepsilon}$ et $\frac{x}{a + \varepsilon}$. Mais l'ensemble des $\varepsilon > 0$ tels que F est discontinue en l'un de ces deux points est au plus dénombrable : on peut donc toujours choisir ε de façon à vérifier les inégalités utilisées. Le reste est sans changements. Les autres parties de la preuve doivent être modifiées de manière analogue, et c'est possible comme un lecteur scrupuleux peut le vérifier sans grande peine. □

► **Remarque**

Il aurait naturellement été possible d'écartier la continuité de F dès le début, mais la preuve en aurait été compliquée de manière sans doute trop rébarbative.

3.3 Rapports avec la convergence en probabilité

Théorème 9

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur un même espace probabilisé (Ω, \mathcal{T}, P) , et soit X une variable aléatoire réelle définie sur le même espace. Alors

$$X_n \xrightarrow{P} X \implies X_n \xrightarrow{\mathcal{L}} X.$$

Preuve

Soient un entier n , un réel x et un réel $\varepsilon > 0$. On note F (resp F_n) la fonction de répartition de X (resp X_n).

- Pour majorer $F_n(x)$, montrons l'inclusion

$$[X_n \leq x] \subset ([X \leq x + \varepsilon] \cup [|X - X_n| > \varepsilon]).$$

En effet, si $X_n \leq x$ et $X > x + \varepsilon$, alors $|X - X_n| \geq X - X_n \geq X - x > \varepsilon$.
De l'inclusion ainsi démontrée, il résulte que

$$F_n(x) \leq F(x + \varepsilon) + P(|X - X_n| > \varepsilon).$$

- Pour minorer $F_n(x)$, montrons l'inclusion

$$[X \leq x - \varepsilon] \subset ([X_n \leq x] \cup [|X - X_n| > \varepsilon]).$$

En effet, si $X_n > x$ et $X \leq x - \varepsilon$, alors $|X - X_n| \geq X_n - X \geq X_n - x + \varepsilon > \varepsilon$.
De l'inclusion ainsi démontrée, il résulte que

$$F(x - \varepsilon) \leq F_n(x) + P(|X - X_n| > \varepsilon).$$

- Soit maintenant un réel $\delta > 0$. Supposons que x soit un point de continuité de F . On peut alors choisir ε tel que l'on ait

$$F(x + \varepsilon) < F(x) + \delta, \quad F(x - \varepsilon) > F(x) - \delta.$$

Il en résulte que

$$F(x) - \delta - P(|X - X_n| > \varepsilon) < F_n(x) < F(x) + \delta + P(|X - X_n| > \varepsilon).$$

Or il existe un N tel que, pour tout $n \geq N$, on ait $P(|X - X_n| > \varepsilon) < \delta$, d'où enfin

$$-2\delta < F_n(x) - F(x) < 2\delta.$$

Quel que soit δ , il existe donc N tel que, $n > N \implies |F_n(x) - F(x)| < 2\delta$, ce qui prouve que $X_n \xrightarrow{\mathcal{L}} X$. □

► **Remarques**

- Il est à noter que la réciproque est fautive, comme le montre l'exemple suivant déjà rencontré.
Soit X une variable de Bernoulli de paramètre $\frac{1}{2}$. La variable $Y = 1 - X$ est également une variable de Bernoulli de paramètre $\frac{1}{2}$. Posons, pour tout entier n , $X_n = X$. On a naturellement $X_n \xrightarrow{\mathcal{L}} X$, donc $X_n \xrightarrow{\mathcal{L}} Y$, puisque X et Y ont la même loi.

Mais $|X_n - Y| = |2X - 1| = 1$ (en effet, si $X = 1$, alors $|2X - 1| = 1$, et si $X = 0$, alors $|2X - 1| = 1$). Donc, pour $\varepsilon \in]0, 1[$, $P(|X_n - Y| > \varepsilon) = 1$ et ne tend donc pas vers 0. La suite $(X_n)_{n \in \mathbb{N}}$ ne converge donc pas en probabilité.

- Il résulte aussi du contre-exemple de la remarque 1 que la limite en loi n'est pas nécessairement unique. D'ailleurs, comme nous l'avons déjà remarqué, la variable limite n'intervient que par sa loi. On ne pourra donc pas avoir de « règles de calcul usuelles » analogues aux règles sur les suites réelles par exemple sur les limites en loi.
- Toutefois, dans certains cas, on dispose de réciproques partielles, et de règles de calcul partielles comme la proposition qui suit où certaines proches du lemme de Slutsky. Les démonstrations de ces propositions peuvent être considérées comme des exercices, mais les plus attentifs de nos lecteurs pourront remarquer que certains résultats serviront dans le chapitre estimation.

Proposition 7

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur un même espace probabilisé (Ω, \mathcal{T}, P) , qui converge en loi vers une variable constante a , alors elle converge aussi en probabilité vers a .

Preuve

Par définition, la fonction de répartition F de la variable certaine a vaut $F(x) = 0$ si $x < a$ et $F(x) = 1$ si $x \geq a$. Elle est continue sur \mathbb{R}^* .

Soient deux réels $\delta > 0$ et $\varepsilon > 0$. Il existe un entier N tel que, pour tout entier $n \geq N$, on ait

$$F_n(a - \varepsilon) < F(a - \varepsilon) + \delta = \delta, \quad F_n(a + \varepsilon) > F(a + \varepsilon) - \delta = 1 - \delta.$$

Pour majorer le nombre $P(|X_n - a| > \varepsilon)$, montrons l'inclusion

$$[|X_n - a| > \varepsilon] \subset ([X_n \leq a - \varepsilon] \cup [X_n > a + \varepsilon]).$$

En effet, si $|X_n - a| > \varepsilon$ et $X_n > a - \varepsilon$, alors $X_n > a + \varepsilon$.

Il en résulte donc

$$0 \leq P(|X_n - a| > \varepsilon) \leq F_n(a - \varepsilon) + 1 - F_n(a + \varepsilon) \leq 2\delta. \quad \square$$

3.4 Exercice corrigé

Exercice 5.

Soit $(U_n)_{n \in \mathbb{N}}$ une suite de variables indépendantes suivant toutes la loi uniforme sur l'intervalle $[0, 1]$. On note $M_n = \max(U_1, U_2, \dots, U_n)$ et $X_n = n(1 - M_n)$.

Étudier la convergence en loi de la suite $(X_n)_{n \in \mathbb{N}}$.

Il est devenu classique que la fonction de répartition de la variable aléatoire M_n est donnée (pour $n \geq 1$) par

$$\begin{cases} F_{M_n}(x) = 0 & \text{si } x \leq 0 \\ F_{M_n}(x) = x^n & \text{si } 0 \leq x \leq 1 \\ F_{M_n}(x) = 1 & \text{si } 1 \leq x. \end{cases}$$

D'autre part,
$$P(X_n \leq x) = P\left(M_n \geq 1 - \frac{x}{n}\right) = 1 - P\left(M_n \leq 1 - \frac{x}{n}\right).$$

Il en résulte que la fonction de répartition de X_n est donnée (pour $n \geq 1$) par

$$\begin{cases} F_{X_n}(x) = 0 & \text{si } x \leq 0 \\ F_{X_n}(x) = 1 - \left(1 - \frac{x}{n}\right)^n & \text{si } 0 \leq x \leq n \\ F_{X_n}(x) = 1 & \text{si } n \leq x. \end{cases}$$

- Lorsque x est négatif, il est clair que $\lim_{n \rightarrow \infty} F_{X_n}(x) = 0$.
- Soit un réel x positif. Dès que n est supérieur à x , l'expression de $F_{X_n}(x)$ est $1 - \left(1 - \frac{x}{n}\right)^n$, de sorte que

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \lim_{n \rightarrow \infty} \left[1 - \left(1 - \frac{x}{n}\right)^n\right] = 1 - e^{-x}.$$

On en déduit que la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers une variable aléatoire X dont la fonction de répartition est donnée par

$$\begin{cases} F_X(x) = 0 & \text{si } x \leq 0 \\ F_X(x) = 1 - e^{-x} & \text{si } 0 \leq x \end{cases}$$

Reconnaissant une loi exponentielle de paramètre 1, on conclut que

$$\boxed{X_n \xrightarrow{\mathcal{L}} X \quad \text{où } X \hookrightarrow \mathcal{E}(1)}$$

ce qui s'écrit parfois sous la forme abusive mais plus simple $X_n \xrightarrow{\mathcal{L}} \mathcal{E}(1)$.

4. Convergence en loi et approximations classiques

Conformément au programme, le théorème ci-dessous sera admis, sa démonstration exigeant un niveau mathématique qui dépasse le cadre de cet ouvrage.

4.1 Théorème de la limite centrée

Théorème 10 Théorème de la limite centrée

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur un même espace probabilisé (Ω, \mathcal{T}, P) , mutuellement indépendantes, de même loi admettant une variance. On

note $E(X_n) = \mu$, $\sigma = \sqrt{V(X_n)}$ supposé non nul, $S_n = \sum_{k=1}^n X_k$ et $S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$.

Alors la suite $(S_n^*)_{n \in \mathbb{N}}$ converge en loi vers une variable aléatoire suivant la loi normale centrée réduite $\mathcal{N}(0, 1)$.

Cela signifie que, dans les conditions et avec les notations de l'énoncé où les réels a et b sont arbitraires

$$\lim_{n \rightarrow +\infty} P(a < S_n^* \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

► **Remarques**

- Lorsque σ n'est pas connu, on montre qu'on peut le remplacer dans la définition de S_n^* par

$$\sqrt{\frac{\sum_{k=1}^n (X_k - \mu)^2}{n}}.$$

En effet, par définition, l'espérance mathématique de la somme $\sum_{k=1}^n (X_k - \mu)^2$ vaut $\sum_{k=1}^n V(X_k)$, c'est-à-dire $n\sigma^2$.

Ce remplacement signifie donc que l'on confond pratiquement cette somme de carrés avec son espérance, ce qui semble raisonnable. On obtient ainsi une formule très parlante

$$S_n^* = \frac{\sum (X_k - \mu)}{\sqrt{\sum (X_k - \mu)^2}}.$$

- Il existe de nombreuses versions de ce théorème, parfois appelé *théorème central limite* (TCL), ou encore *théorème-limite central* (TLC). La version que nous avons donnée est connue sous le nom de *théorème de Lindeberg-Lévy*.
- Le nom initial en a été donné par Pólya en 1920 sous la forme *zentraler Grenzwertsatz*. D'après la syntaxe allemande, il s'agit donc d'un « théorème-frontière » reconnu comme central, ce qui ne manque pas d'être légèrement contradictoire dans les termes et qui justifierait plutôt l'emploi de la forme TLC, même si TCL est encore souvent employé, notamment dans les pays anglo-saxons.
- Ce théorème est remarquable dans ce sens que, avec peu d'hypothèses, on arrive à un résultat très fort. Il met en évidence le rôle central joué par la loi de Gauss en statistiques et en probabilités.

4.2 Convergence de la loi hypergéométrique

Le résultat suivant a été démontré dans le premier tome de cet ouvrage.

Proposition 8

Soit $(X_N)_{N \in \mathbb{N}}$ une suite de variables aléatoires réelles discrètes telles que X_N suive la loi hypergéométrique de paramètre (N, n, p) . Alors, pour tout entier k de $\llbracket 0, n \rrbracket$, on a

$$\lim_{N \rightarrow +\infty} P(X_N = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

et la suite $(X_N)_{N \in \mathbb{N}}$ converge en loi vers une variable aléatoire X qui suit la loi binomiale de paramètre (n, p) .

Preuve

Rappelons que, si $q = 1 - p$,

$$P(X_N = k) = \frac{\binom{Np}{k} \binom{Nq}{n-k}}{\binom{N}{n}}.$$

Alors

$$P(X_N = k) = \frac{n!}{k!(n-k)!} \frac{(Np)(Np-1)\cdots(Np-k+1)(Nq)(Nq-1)\cdots(Nq-n+k+1)}{N(N-1)\cdots(N-n+1)}.$$

Or les produits

$$(Np)(Np-1)\cdots(Np-k+1)(Nq)(Nq-1)\cdots(Nq-n+k+1) \quad \text{et} \quad N(N-1)\cdots(N-n+1)$$

possèdent n facteurs, n étant indépendant de N . Donc

$$\lim_{N \rightarrow +\infty} \frac{(Np)(Np-1)\cdots(Np-k+1)(Nq)(Nq-1)\cdots(Nq-n+k+1)}{N(N-1)\cdots(N-n+1)} = p^k q^{n-k}.$$

Ainsi

$$\lim_{N \rightarrow +\infty} P(X_N = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

□

Dans la pratique, dès que $N \geq 10n$, on approche $\mathcal{H}(N, n, p)$ par $\mathcal{B}(n, p)$.

Exemple

Soit X une variable aléatoire suivant la loi hypergéométrique de paramètre $(100; 4; 0, 05)$, et intéressons nous à $P(X \geq 1)$.

- *Calcul exact*

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) = 1 - \frac{\binom{5}{0} \binom{95}{4}}{\binom{100}{4}} \\ &= 1 - \frac{95 \times 94 \times 93 \times 92}{4 \times 3 \times 2 \times 1} \times \frac{4 \times 3 \times 2 \times 1}{100 \times 99 \times 98 \times 97} \approx 0,188. \end{aligned}$$

- *Calcul approché*

Approchons la loi $\mathcal{H}(100; 4; 0, 05)$ par la loi $\mathcal{B}(4; 0, 05)$. On a alors

$$P(X \geq 1) = 1 - P(X = 0) \approx 1 - \binom{4}{0} \times (0,05)^0 (0,95)^4 \approx 1 - (0,95)^4 \approx 0,185.$$

4.3 Approximations de la loi binomiale

On rencontre souvent des lois binomiales. Il en existe des tables numériques, usuellement jusqu'à $n = 50$, et pour des valeurs de p comprises entre 0 et 0,5. Pour les valeurs de n supérieures à 50, on envisage deux cas.

Si n est grand alors que p est petit (cas des événements rares), on utilise l'approximation par la loi de Poisson

Proposition 9 (Approximation poissonnienne de certaines lois binomiales)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles discrètes suivant les lois binomiales de paramètres (n, p_n) . On suppose qu'il existe un réel non nul λ tel que $\lim_{n \rightarrow +\infty} np_n = \lambda$.

Alors, pour tout entier k de \mathbb{N} , on a

$$\lim_{n \rightarrow +\infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

La suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers une variable aléatoire X qui suit la loi de Poisson de paramètre λ .

Preuve

Pour tout entier k (fixé) et pour tout entier $n \geq k$

$$P(X_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{p_n^k}{k!} (n(n-1) \cdots (n-k+1)) e^{(n-k)\ln(1-p_n)}.$$

Or $n(n-1) \cdots (n-k+1) \sim n^k$ et $(n-k)\ln(1-p_n) \sim -np_n \sim -\lambda$.

Ainsi $P(X_n = k) \sim \frac{n^k p_n^k}{k!} e^{-\lambda}$ et $\lim_{n \rightarrow +\infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}$.

La conclusion en résulte par utilisation du théorème du 8. □

Dans la pratique, dès que $n \geq 30$, $p \leq 0,1$ et $np < 15$, on approche $\mathcal{B}(n, p)$ par $\mathcal{P}(np)$.

Si n est grand alors que p reste fixe, on utilise l'approximation normale.

Proposition 10 (Approximation normale de certaines loi binomiales)

Soient p un réel strictement compris entre 0 et 1 et $q = 1 - p$, (X_n) une suite de variables aléatoires réelles discrètes suivant la loi binomiale de paramètre (n, p) et (Y_n) une suite de variables aléatoires suivant la loi normale de paramètre (np, npq) .

Alors leurs variables centrées réduites X_n^* et Y_n^* sont telles que la loi de la seconde est indépendante de l'indice n et que la première converge en loi vers elle.

Preuve

Toute variable binomiale de paramètre (n, p) peut être considérée comme la somme de n variables de Bernoulli mutuellement indépendantes de même paramètre p . On a donc $X_n = \sum_{k=0}^n B_k$ où B_k désigne une variable aléatoire de Bernoulli de paramètre p , et où les B_k sont mutuellement indépendantes. Comme les variables de Bernoulli admettent une variance, on est dans le cadre de l'application du théorème de la limite centrée, et l'on en déduit que la variable centrée réduite $X_n^* = \frac{X_n - np}{\sqrt{npq}}$ converge en loi vers une variable aléatoire Z suivant la loi normale centrée réduite, de fonction de répartition continue en tout point.

On peut prendre pour Z la variable aléatoire centrée réduite Y_n^* , dont la loi est effectivement indépendante de n . Il en résulte que X_n^* converge en loi vers $Z = Y_n^*$. □

► **Remarques**

- Bien que l'on puisse écrire $X_n = \varphi(X_n^*)$ et $Y_n = \varphi(Y_n^*)$ avec la même application affine φ définie par $\varphi(t) = np + t\sqrt{npq}$, on ne peut pourtant en déduire que X_n converge en loi vers Y_n ; d'ailleurs cette dernière dépend aussi de l'indice n .

- Cela dit, on peut néanmoins en conclure intuitivement que, pour n assez grand, la loi de X_n n'est sans doute pas très éloignée de la loi de Y_n , c'est-à-dire de la loi normale de paramètre (np, npq) .
- On aura noté que l'on approche ici une loi discrète par une loi continue. On doit alors souvent utiliser la *correction de continuité*, et ainsi approcher $P(B = k)$ par $P(k - 0,5 < N < k + 0,5)$, où B suit une loi binomiale et N la loi normale appropriée.

Dans la pratique, dès que $n \geq 30$, $np \geq 5$ et $nq \geq 5$, on approche $\mathcal{B}(n, p)$ par $\mathcal{N}(np, npq)$.

Exemples

1. On considère une variable aléatoire X suivant la loi binomiale $\mathcal{B}(100; 0,05)$, et l'on s'intéresse à la probabilité pour que X prenne la valeur 2.

• *Calcul exact*

$$P(X = 2) = \binom{100}{2} (0,05)^2 (0,95)^{98} = 4950 \times 0,0025 \times (0,95)^{98} \approx 0,0812.$$

• *Calcul approché*

Nous sommes dans les conditions de l'approximation de la loi binomiale par la loi de Poisson. On approche donc la loi $\mathcal{B}(100; 0,05)$ par la loi $\mathcal{P}(100 \times 0,05) = \mathcal{P}(5)$. On

obtient $P(X = 2) \approx e^{-5} \times \frac{5^2}{2!} \approx 0,0843$.

2. Soit X une variable binomiale de paramètre $(900; 0,5)$. On cherche à calculer

$$P(0,45 \leq X \leq 0,55).$$

Le calcul avec les formules exactes nécessiterait l'utilisation d'un ordinateur, et ne donnerait de toutes façons qu'un résultat approché.

Mais nous sommes dans les conditions d'approximation de la loi binomiale par la loi normale. On approche la loi $\mathcal{B}(900; 0,5)$ par la loi $\mathcal{N}(450, 225)$, et l'on considérera que la variable $\frac{X - 450}{15}$ suit approximativement la loi normale centrée réduite. On obtient alors

$$P(0,45 \leq X \leq 0,55) = P\left(-3 \leq \frac{X - 450}{15} \leq 3\right) \approx 2\Phi(3) - 1 \approx 0,9973.$$

4.4 Approximation de la loi de Poisson

La loi de Poisson peut également être approchée par la loi normale.

Proposition 11 (Approximation normale de certaines lois de Poisson)

Soient α un réel strictement positif, X_n une variable aléatoire réelle discrète suivant la loi de Poisson de paramètre $n\alpha$ et Y_n une variable aléatoire suivant la loi normale de paramètre $(n\alpha, n\alpha)$.

Alors leurs variables centrées réduites X_n^* et Y_n^* sont telles que la loi de la seconde est indépendante de l'indice n et que la première converge en loi vers elle.

Preuve

Toute variable de Poisson de paramètre $(n\alpha)$ peut être considérée comme la somme de n variables de Poisson mutuellement indépendantes de même paramètre α . On a donc $X_n = \sum_{k=0}^n P_k$ où P_k désigne une variable aléatoire de Poisson de paramètre α , et où les P_k sont mutuellement indépendantes. Comme les variables de Poisson admettent une variance, on est dans le cadre de l'application du théorème de la limite centrée, et l'on en déduit que la variable centrée réduite $X_n^* = \frac{X_n - n\alpha}{\sqrt{n\alpha}}$ converge en loi vers une variable aléatoire Z suivant la loi normale centrée réduite, de fonction de répartition continue en tout point.

On peut prendre pour Z la variable aléatoire centrée réduite Y_n^* , dont la loi est effectivement indépendante de n . Il en résulte que X_n^* converge en loi vers $Z = Y_n^*$. □

► **Remarques**

- Bien que l'on puisse écrire $X_n = \varphi(X_n^*)$ et $Y_n = \varphi(Y_n^*)$ avec la même application affine φ définie par $\varphi(t) = n\alpha + t\sqrt{n\alpha}$, on ne peut pourtant en déduire que X_n converge en loi vers Y_n ; d'ailleurs cette dernière dépend aussi de l'indice n .
- Cela dit, on peut néanmoins en conclure intuitivement que, pour n assez grand, la loi de X_n n'est sans doute pas très éloignée de la loi de Y_n , c'est-à-dire de la loi normale de paramètre $(n\alpha, n\alpha)$.

Dans la pratique, dès que $\lambda \geq 18$, on approche la loi $\mathcal{P}(\lambda)$ par la loi normale $\mathcal{N}(\lambda, \lambda)$.

► **Remarque**

On aura noté que, comme dans le cas précédent, on approche ici une loi discrète par une loi continue. On doit alors souvent utiliser la *correction de continuité*.

Exemple

Dans la table de la loi de Poisson de paramètre $\lambda = 16$, on lit $P(X \leq 21) \approx 0,9107$.

En approchant X par une variable normale de paramètre $(16,16)$, on est conduit à considérer que la variable $\frac{X - 16}{4}$ suit une loi normale centrée réduite.

On obtient alors
$$P(X \leq 21) \approx P\left(\frac{X - 16}{4} \leq 1,25\right) \approx \Phi(1,25) \approx 0,8944.$$

L'approximation obtenue n'est pas catastrophique, bien que la valeur de λ soit faible.

► **Remarque**

Comme on le verra dans le chapitre consacré aux interventions informatiques et aux simulations, le théorème de la limite centrée permet d'envisager des simulations de la loi normale centrée réduite, par exemple en utilisant des variables de Bernoulli, ou des variables uniformes sur l'intervalle $[0,1]$.

1. On considère une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires et une variable aléatoire X , toutes définies sur un même espace probabilisé.

Prouver que

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} X \iff [\forall \varepsilon \in \mathbb{R}_+^*, \exists N(\varepsilon) \in \mathbb{N}, n \geq N(\varepsilon) \Rightarrow P(|X_n - X| > \varepsilon) \leq \varepsilon].$$

2. On considère une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires et une variable aléatoire X , toutes définies sur un même espace probabilisé.

1. Prouver que

$$X_n \xrightarrow{\mathcal{P}} X \iff \lim_{n \rightarrow +\infty} E \left(\frac{|X_n - X|}{1 + |X_n - X|} \right) = 0.$$

2. Reprendre la question précédente avec $E(f(|X_n - X|))$ au lieu de $E \left(\frac{|X_n - X|}{1 + |X_n - X|} \right)$, où f désigne une fonction continue, bornée, strictement croissante de \mathbb{R}_+ dans \mathbb{R} , telle que, de plus, $f(0) = 0$.

3. Pour tout entier naturel non nul n , on considère la variable aléatoire X_n telle que

$$P(X_n = -n) = P(X_n = n) = \frac{1}{2n^2} \text{ et } P(X_n = 0) = 1 - \frac{1}{n^2}.$$

1. Vérifier que X_n est bien une variable aléatoire. Calculer son espérance et sa variance.
2. Étudier la convergence en probabilité, en moyenne et en moyenne quadratique de la suite $(X_n)_{n \in \mathbb{N}^*}$ (voir la remarque 3 du 1.1.7).

4. Soit θ un réel positif. Soit une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires indépendantes, de même loi, admettant toutes comme densité la fonction f définie par

$$f(x) = \mathbf{1}_{[\theta, +\infty[}(x) e^{-(x-\theta)}.$$

On pose $m_n = \min_{k \in \llbracket 1, n \rrbracket} \{X_k\}$.

Étudier la convergence en probabilité de la suite $(m_n)_{n \in \mathbb{N}}$.

5. On considère une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires indépendantes suivant la même loi de Bernoulli de paramètre p . Soit alors la suite $(Y_n)_{n \in \mathbb{N}^*}$ des variables aléatoires Y_k définies pour $k > 0$ par $Y_k = X_{k-1} + X_k$, et soit $Z_n = \frac{1}{n} \sum_{k=1}^n Y_k$.

1. Déterminer la loi de Y_k , calculer son espérance et sa variance.
2. Calculer l'espérance et la variance de Z_n .
3. Étudier la convergence en loi de la suite $(Z_n)_{n \in \mathbb{N}}$. Peut-on utiliser la loi des grands nombres ?

6. On considère une variable aléatoire réelle X suivant la loi exponentielle de paramètre 1, la variable aléatoire $Y = e^{-X}$ et, pour tout entier naturel non nul n , la variable aléatoire réelle Y_n définie par $Y_n = e^{-X+1/n}$.

1. Montrer que Y est une variable aléatoire absolument continue dont on précisera la loi.
2. Prouver que $Y_n \xrightarrow{\mathcal{L}} Y$.

7. On considère une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires telles que

$$P(X_n = n) = \frac{1}{n} \text{ et } P(X_n = 0) = 1 - \frac{1}{n}.$$

On considère également une suite $(Y_n)_{n \in \mathbb{N}^*}$ de variables aléatoires suivant toutes la loi normale centrée réduite, et indépendante de la suite $(X_n)_{n \in \mathbb{N}^*}$.

1. Étudier la convergence en probabilité et en moyenne quadratique (voir la remarque 3 du 1.1.7) de la suite $(X_n)_{n \in \mathbb{N}^*}$.
2. Étudier la convergence en loi de la suite $(Z_n)_{n \in \mathbb{N}^*}$ telle que $Z_n = X_n + Y_n$.
3. Étudier la limite de la suite $(V(Z_n))_{n \in \mathbb{N}^*}$.

8. Soit M un réel strictement positif, et soit X une variable aléatoire réelle bornée par M (c'est à dire telle que, pour tout ω de Ω , $|X(\omega)| \leq M$).

1. Montrer que, pour tout entier naturel non nul, X admet un moment d'ordre n .
2. Prouver que, pour tout réel strictement positif a ,

$$\frac{E(X^2) - a^2}{M^2} \leq P(X \geq a) \leq \frac{E(|X|)}{a}$$

(on pourra s'intéresser à la variable indicatrice de l'événement $[X \geq a]$).

9. Soit σ un réel strictement positif. Pour tout entier naturel non nul n , on considère

la fonction f_n définie par $f_n(x) = \mathbb{1}_{\mathbb{R}_+}(x) \frac{n^2 x}{\sigma^2} \exp\left(-\frac{n^2 x^2}{2\sigma^2}\right)$.

1. Montrer que f_n est la densité d'une variable aléatoire.
2. Prouver que la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers une variable X que l'on précisera.

10. On lance n fois un dé honnête, et l'on note A l'événement « obtenir un as ou un six au cours d'un jet ». Soit X_k la variable aléatoire égale à 1 si l'événement A est réalisé au

k -ème jet, et à 0 sinon. On note enfin $Y_n = \frac{1}{n} \sum_{k=1}^n X_k$.

1. Étudier la convergence en probabilité de la suite $(Y_n)_{n \in \mathbb{N}^*}$.
2. On lance le dé 100 000 fois, et l'on trouve $Y_{100000} = 0,34$. Déterminer un majorant de la probabilité $P(|Y_n - E(Y_n)| > |0,34 - E(Y_n)|)$.

11. Soit X une variable aléatoire exponentielle de paramètre λ .

1. Utiliser l'inégalité de Bienaymé Tchebychev pour prouver que $\forall t \in \mathbb{R}_+^*, t^2 e^{-t} \leq e$.
2. En étudiant directement la fonction f définie par $f(x) = x^2 e^{-x}$, préciser et commenter ce résultat.

12. On lance n fois un dé honnête à six faces. En utilisant l'inégalité de Bienaymé Tchebychev, déterminer un majorant des valeurs de n pour lesquelles on a plus d'une chance sur deux d'obtenir une fréquence d'apparition de l'as qui s'écarte de moins de 10^{-2} de la valeur théorique $\frac{1}{6}$.

13. Soit θ un réel strictement positif, et, pour tout entier naturel n , soit X_n une variable aléatoire réelle qui suit une loi géométrique de paramètre $p_n = \frac{\theta}{n}$.

Étudier la convergence en loi de la suite $(Y_n)_{n \in \mathbb{N}^*}$ où $Y_n = \frac{X_n}{n}$.

14. Soit n un entier naturel non nul et a un réel. On considère la fonction f_n définie par $f_n(x) = \frac{an}{\pi(1+n^2x^2)}$.

1. Déterminer a pour que f_n soit une densité d'une variable aléatoire.
2. Soit X_n une variable aléatoire de densité f_n . Étudier les moments de X_n .
3. Étudier la convergence en loi de la suite $(X_n)_{n \in \mathbb{N}^*}$.

15. Soit n un entier naturel non nul. On sait qu'il existe $n!$ permutations de l'ensemble $\{1, 2, 3, \dots, n\}$. On note p_n le nombre de ces permutations qui ne présentent aucun point fixe.

On prend maintenant une de ces permutations au hasard, et l'on note X_n le nombre de points fixes de la permutation ainsi déterminée au hasard.

1. Calculer p_n .
2. Déterminer la loi de X_n .
3. Prouver que la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers une variable de loi classique.

16. On lance un dé équilibré. On gagne 1 euro si le résultat est pair, et on perd un euro si le résultat est impair. Majorer le nombre de lancers à effectuer pour que l'on ait au moins une probabilité 0,95 pour perdre au plus 20 euros (on comparera les résultats obtenus par approximation de loi, avec et sans correction de continuité, et par l'inégalité de Bienaymé-Tchebychev).

17. Soit X une variable aléatoire suivant la loi binomiale de paramètre $\left(n, \frac{1}{3}\right)$. On pose $F = \frac{X}{n}$. Déterminer n tel que $P(|F - E(F)| < 0,01) \geq 0,98$ (on comparera les résultats obtenus par approximation de loi, et par l'inégalité de Bienaymé-Tchebychev).

Autres exercices

On cherchera également avec profit les exercices proposés aux oraux d'HEC et de ESCP-EAP. Signalons entre autres ESCP-EAP oral 1990, 1997, 2001, 2002, 2003 ; HEC oral 1990.

Face à un phénomène aléatoire numérique, on cherche à connaître autant que possible la loi qui le régit et les valeurs des paramètres dont elle peut dépendre.

Or ce phénomène n'est en général accessible que par les réalisations de la variable aléatoire X qui le décrit. C'est donc par l'intermédiaire de plusieurs réalisations de cette variable aléatoire que l'on cherche à déterminer, éventuellement de manière approchée (on dit *estimer*), la loi de la variable X et ses paramètres éventuels. Lorsque l'on envisage ainsi n réalisations successives de la variable X , on dit que l'on a réalisé *un échantillon de taille n* (ou un n -échantillon) de la variable X . Il est alors d'usage de considérer que ces n réalisations successives de la même variable aléatoire X constituent une seule et même réalisation du vecteur aléatoire (X_1, X_2, \dots, X_n) , où les X_i sont des variables aléatoires de même loi (et de même paramètre) que X . On dit que X est la *loi parente* de l'échantillon envisagé.

Il existe de nombreuses façons de constituer des échantillons. Nous n'envisagerons que le cas de l'échantillonnage aléatoire simple, où les variables X_i sont **mutuellement indépendantes**. Pour rappeler qu'un échantillon (X_1, X_2, \dots, X_n) est constitué de variables aléatoires X_i mutuellement indépendantes et de même loi (avec le même paramètre), on parlera d'un échantillon *iid* (*indépendant et identiquement distribué*). La constitution d'échantillons plus compliqués, et plus sophistiqués sera par exemple envisagée dans la théorie des sondages.

Il arrive souvent, et c'est le seul cas qui sera envisagé dans cet ouvrage, que la forme de la loi de la variable parente X soit connue, et que l'on cherche seulement à en déterminer certains paramètres. On parle alors d'*estimation paramétrique*.

Dans le présent chapitre, on commence par étudier quelques propriétés d'un échantillon *iid* d'une loi parente connue, puis on s'intéressera au cas où la loi parente appartient à un ensemble dépendant d'un ou de plusieurs paramètres que l'on cherchera à estimer.

► Remarque

Treize simulations de lois classiques, ainsi que cinq applications concrètes de l'estimation, figurent sous forme de programmes en Turbo Pascal, dans le chapitre suivant, consacré aux interventions informatiques.

1. Échantillons d'une loi de probabilité

Les définitions qui suivent jouent un rôle crucial en statistique, théorique et appliquée.

1.1 Définitions

Définition 1

Soit \mathcal{L} une loi de probabilité sur un espace probabilisé (Ω, \mathcal{T}, P) . On appelle *échantillon de taille n* (ou *n -échantillon*) de la loi \mathcal{L} , une suite $\mathcal{E}_n = (X_1, X_2, \dots, X_n)$ de n variables aléatoires X_i suivant toutes la loi \mathcal{L} .

La loi \mathcal{L} est la *loi parente* de l'échantillon.

► Remarques

- La donnée de la suite finie (X_1, X_2, \dots, X_n) équivaut naturellement à la donnée du vecteur aléatoire (X_1, X_2, \dots, X_n) .
La variable aléatoire X_i se déduit donc de l'échantillon par projection sur le i -ième vecteur de la base canonique de \mathbb{R}^n .
- Pour tout possible ω de la tribu \mathcal{T} , la réalisation

$$(X_1(\omega), X_2(\omega), \dots, X_n(\omega)) = (x_1, x_2, \dots, x_n)$$

est un élément de \mathbb{R}^n , et c'est cet élément que l'on appelle souvent « échantillon ». Pour éviter les confusions, le vecteur aléatoire est parfois appelé *échantillon aléatoire*, (x_1, x_2, \dots, x_n) étant un *échantillon observé*.

- Nous rappelons que nous n'aurons à envisager que des échantillons dont les composantes sont des variables aléatoires mutuellement indépendantes.
Dans la suite, nous nous permettrons l'abus de langage qui consiste à appeler simplement « échantillon » ce que nous devrions appeler « échantillon *iid* ».
- Insistons sur le fait que, dans cette section, la loi \mathcal{L} est parfaitement définie.

1.2 Statistiques sur un échantillon

Définition 2

Soit $\mathcal{E}_n = (X_1, X_2, \dots, X_n)$ un n -échantillon d'une loi de probabilité \mathcal{L} . On appelle *statistique sur \mathcal{E}_n* toute variable aléatoire $Y = \varphi_n(X_1, X_2, \dots, X_n)$ où φ_n est une application de \mathbb{R}^n dans \mathbb{R} telle que Y soit une variable aléatoire.

► Remarques

- C'est une extension de l'écriture classique, mais assez abusive, qui fait que l'on note $f(X)$ la variable aléatoire $f \circ X$. De manière plus précise, pour tout $\omega \in \Omega$

$$Y(\omega) = \varphi_n(X_1(\omega), X_2(\omega), \dots, X_n(\omega)).$$

- Il est possible d'envisager des statistiques vectorielles avec des applications φ_n de \mathbb{R}^n dans \mathbb{R}^p . Ce ne sera pas le cas dans cet ouvrage, sauf mention expresse du contraire.
- Une statistique sur \mathcal{E}_n est une variable aléatoire définie sur le même espace probabilisé (Ω, \mathcal{T}, P) que la loi parente \mathcal{L} de l'échantillon \mathcal{E}_n .

Exemples

1. Le résultat du lancer d'un dé supposé bien équilibré, est régit par la loi uniforme discrète sur $\llbracket 1, 6 \rrbracket$. En lançant le dé 100 fois, on réalise un 100-échantillon $\mathcal{E}_{100} = (X_1, X_2, \dots, X_{100})$ de cette loi. Si l'on cherche à illustrer la loi des grands nombres, on est conduit à calculer la moyenne des résultats obtenus, c'est à dire à étudier la variable aléatoire $Y = \frac{1}{100} \sum_{k=1}^{100} X_k$.
 Cette statistique s'appelle la *moyenne empirique* de l'échantillon \mathcal{E}_{100} .
2. Deux réels a et b tels que $a < b$ étant donnés, on considère la variable aléatoire à densité X , suivant la loi uniforme sur un intervalle $[a, b]$. Soit $\mathcal{E}_n = (X_1, X_2, \dots, X_n)$ un n -échantillon de la loi de X .
 Les variables aléatoires $Y_n = \sup_{k \in \llbracket 1, n \rrbracket} X_k$ et $Z_n = \inf_{k \in \llbracket 1, n \rrbracket} X_k$ sont alors des statistiques sur le n -échantillon $\mathcal{E}_n = (X_1, X_2, \dots, X_n)$.
 Il peut être intéressant d'étudier le comportement de ces statistiques lorsque n tend vers l'infini.

1.3 Statistiques empiriques

Soit \mathcal{L} une loi de probabilité, et soit F la fonction de répartition d'une variable aléatoire X suivant la loi \mathcal{L} . Les valeurs typiques des paramètres usuels de la loi \mathcal{L} (espérance, variance ...) diffèrent des valeurs obtenues en utilisant un n -échantillon $\mathcal{E}_n = (X_1, X_2, \dots, X_n)$ de la loi \mathcal{L} , qui sont qualifiées d'*empiriques* (car obtenues à partir de réalisations de l'échantillon).

Fonction de répartition empirique

Définition 3

Soit $\mathcal{E}_n = (X_1, X_2, \dots, X_n)$ un n -échantillon d'une loi \mathcal{L} .

La statistique F_n définie pour tout x réel par

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{]-\infty, x]}(X_k)$$

porte le nom de *fonction de répartition empirique de la loi \mathcal{L} associée à l'échantillon \mathcal{E}_n* .

► **Remarques**

- Le symbole $\mathbb{1}_{]-\infty, x]}(X_k)$ s'écrit encore $(indic \circ X_k)(x)$ ou, comme on le fait habituellement, $indic(X_k)(x)$; c'est la variable aléatoire Y vérifiant $Y(\omega) = 1$ si $X_k(\omega) \leq x$, et 0 sinon.
- Une fonction de répartition d'échantillon est une statistique qui dépend du réel x . Ce n'est pas une fonction de répartition au sens usuel du terme. Insistons sur le fait que, pour x fixé, $F_n(x)$ est une variable aléatoire.
- La variable aléatoire $F_n(x)$ représente le pourcentage (aléatoire) des valeurs prises par les variables aléatoires composant l'échantillon et qui sont inférieures ou égales à x .
- On peut dire que $F_n(x)$ est une variable aléatoire dont chaque réalisation est une fonction en escalier dont les sauts sont des multiples de $\frac{1}{n}$.

Théorème 1

Pour tout réel x , la suite de variables aléatoires $F_n(x)$ converge en probabilité (donc en loi) vers la variable aléatoire constante égale à $F(x)$

$$\forall x \in \mathbb{R}, \quad F_n(x) \xrightarrow{p} F(x).$$

Preuve

La variable aléatoire $\mathbf{1}_{]-\infty, x]}(X_k)$ est une variable de Bernoulli qui prend la valeur 1 si, et seulement si, $X_k \in]-\infty, x]$, c'est à dire si, et seulement si, $X_k \leq x$, ce qui se produit avec la probabilité $F(x)$, puisque toutes les variables aléatoires X_k suivent la même loi \mathcal{L} de fonction de répartition F .

La variable aléatoire $F_n(x)$ est donc la moyenne de n variables de Bernoulli de même paramètre $F(x)$ indépendantes comme les variables X_k . D'après la loi faible des grands nombres, elle converge en probabilité (et donc en loi) vers la variable certaine égale à $F(x)$. \square

► Remarques

- La convergence de $F_n(x)$ vers $F(x)$ est en réalité bien plus forte que ce que le dit ce théorème. Elle est *presque sûre*, et *presque sûrement uniforme*. Mais ces notions dépassent notablement le cadre du programme. On retiendra que la convergence annoncée dans ce théorème est « très forte ».
- Ce théorème, ainsi que ceux qui sont évoqués à la première remarque, constituent l'argument fondamental qui justifie l'utilisation d'échantillons en statistiques.

Moyenne empirique**Définition 4**

Soit $\mathcal{E}_n = (X_1, X_2, \dots, X_n)$ un n -échantillon d'une loi \mathcal{L} .

La statistique \bar{X}_n définie par

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$$

porte le nom de *moyenne empirique de la loi \mathcal{L} associée à l'échantillon \mathcal{E}_n* .

► Remarque

Certains auteurs ont l'habitude d'écrire \bar{X}_n pour $E(X_n)$; cette notation peut donc parfois être ambiguë. Toutefois cette possibilité, bien que naturellement évoquée dans le livre de première année au moment de la définition de l'espérance, n'est jamais mise en œuvre dans cet ouvrage.

Théorème 2

Soit \bar{X}_n la moyenne empirique associée à un n -échantillon *iid* d'une loi \mathcal{L} d'espérance m et de variance σ^2 . Alors $E(\bar{X}_n) = m$ et $V(\bar{X}_n) = \frac{\sigma^2}{n}$.

Preuve

D'après la linéarité de l'espérance on a $E(\bar{X}_n) = \frac{1}{n} n E(X) = E(X) = m$.

Compte tenu du caractère quadratique de la variance, on a $V(\bar{X}_n) = \frac{1}{n^2} V\left(\sum_{k=1}^n X_k\right)$, c'est-à-dire, compte tenu de l'indépendance mutuelle des variables et l'égalité de leurs variances, $V(\bar{X}_n) = \frac{1}{n^2} n V(X) = \frac{\sigma^2}{n}$. \square

Comportement asymptotique de la moyenne empirique

D'après les résultats du chapitre sur les convergences, on dispose des résultats suivants

Théorème 3

La moyenne empirique \bar{X}_n associée à un n -échantillon *iid* d'une loi \mathcal{L} d'espérance m converge en probabilité vers la variable certaine égale à m

$$\bar{X}_n \xrightarrow{P} m.$$

Preuve

Il s'agit simplement de la loi des grands nombres. □

Théorème 4

Soit \bar{X}_n la moyenne empirique associée à un n -échantillon *iid* d'une loi \mathcal{L} d'espérance m et de variance σ^2 . Alors la variable centrée réduite associée à \bar{X}_n converge en loi vers une variable normale centrée réduite

$$\sqrt{n} \frac{\bar{X}_n - m}{\sigma} \xrightarrow{\mathcal{L}} Y, \quad \text{avec } Y \leftrightarrow \mathcal{N}(0, 1).$$

Preuve

Il s'agit simplement d'une application du théorème de la limite centrée, puisque la variance de \bar{X}_n est $\frac{\sigma^2}{n}$. □

Variance empirique

Définition 5

Soit $\mathcal{E}_n = (X_1, X_2, \dots, X_n)$ un n -échantillon d'une loi \mathcal{L} .

La statistique S_n^2 définie par

$$S_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

porte le nom de *variance empirique de la loi \mathcal{L} associée à l'échantillon \mathcal{E}_n* .

Théorème 5

Soit S_n^2 la variance empirique associée à un n -échantillon *iid* d'une loi \mathcal{L} d'espérance m et de variance σ^2 . Alors $E(S_n^2) = \frac{n-1}{n} \sigma^2$.

Preuve

Pour tout k de $[1, n]$, on a $X_k - m = (X_k - \bar{X}_n) + (\bar{X}_n - m)$.

$$\text{Donc } \sum_{k=1}^n (X_k - m)^2 = \sum_{k=1}^n (X_k - \bar{X}_n)^2 + \sum_{k=1}^n (\bar{X}_n - m)^2 + 2(\bar{X}_n - m) \sum_{k=1}^n (X_k - \bar{X}_n).$$

$$\text{Or } \sum_{k=1}^n (X_k - \bar{X}_n) = 0, \text{ d'où } \sum_{k=1}^n (X_k - m)^2 = \sum_{k=1}^n (X_k - \bar{X}_n)^2 + \sum_{k=1}^n (\bar{X}_n - m)^2.$$

Il en résulte que $S_n^2 = \left[\frac{1}{n} \sum_{k=1}^n (X_k - m)^2 \right] - (\bar{X}_n - m)^2$, puis que

$$E(S_n^2) = \left[\frac{1}{n} \sum_{k=1}^n E(X_k - m)^2 \right] - E((\bar{X}_n - m)^2) = \left[\frac{1}{n} \sum_{k=1}^n V(X_k) \right] - V(\bar{X}_n) = \sigma^2 - \frac{\sigma^2}{n}$$

et enfin le résultat annoncé. □

► Remarques

- Cette valeur est donc (légèrement) inférieure à la valeur σ^2 qui, *a priori*, aurait peut-être été attendue comme variance de l'échantillon comme m en était l'espérance.
- Ce résultat légitime l'ultime remarque du livre de première année et l'apparente bizarrerie du comportement d'un logiciel comme `Excel` ou de certaines calculettes quand on leur demande de calculer la variance d'une série statistique. Il y a là en effet une correction systématique de multiplication par $\frac{n}{n-1}$, dans la mesure où il pense que la série statistique sur laquelle il calcule est un échantillon et non la série elle-même.
- On remarquera que l'on connaît les valeurs typiques de la moyenne empirique, mais seulement l'espérance de la variance empirique, rien n'étant dit ici de la variance de la variance empirique.

1.4 Simulation de variables aléatoires réelles

On a souvent besoin de disposer d'échantillons de lois connues. Pour obtenir de tels échantillons, on peut, lorsque c'est possible, utiliser un modèle physique. C'est ainsi que l'utilisation d'une pièce de monnaie ou celle d'un dé à jouer permettent de simuler certaines variables de Bernoulli, uniformes, binomiales, géométriques... Mais ce type d'expérimentation physique est souvent assez limité, lourd et coûteux à mettre en œuvre. On a alors recours à une simulation informatique, qui consiste à élaborer un programme qui donne des nombres de manière aléatoire, en suivant la loi que l'on cherche à simuler.

Pour réaliser de tels programmes, l'outil essentiel est un *générateur de nombres au hasard*. On trouvera ci-dessous quelques brèves indications de méthodes et des exemples de simulation, mais il n'est pas question ici d'être exhaustif. Chaque cas particulier de simulation peut donner lieu à plusieurs méthodes particulières.

Générateurs pseudo-aléatoires

Tous les langages de programmation incorporent une fonction permettant d'engendrer des « nombres au hasard ». En Pascal, la fonction `RANDOM` s'utilise après initialisation par la fonction `RANDOMIZE`.

Utilisée sous la forme `RANDOM(n)`, elle simule la variable uniforme discrète sur $\llbracket 0, n-1 \rrbracket$.

Utilisée seule, elle simule la variable uniforme à densité sur l'intervalle $[0,1]$ (en fait, elle ne prend que des valeurs dans $[0, 1[$).

En réalité, on ne sait pas engendrer des nombres de manière absolument aléatoire. Les générateurs de nombres au hasard utilisent en fait des processus parfaitement déterministes (comme certaines suites récurrentes). C'est d'ailleurs la raison pour laquelle on utilise à leur endroit l'expression « générateurs pseudo-aléatoires ». Mais les résultats obtenus sont très proches de ce que l'on recherche, et les progrès effectués rendent l'approximation du processus aléatoire par le processus pseudo-aléatoire excellente.

En particulier, la fonction RANDOM ne peut pas simuler une variable aléatoire uniforme à densité sur l'intervalle $[0, 1]$, pour la bonne raison que les nombres fournis par l'ordinateur sont des décimaux et non pas des réels quelconques. Mais le nombre de décimales fourni est assez grand pour que l'on puisse admettre que l'approximation obtenue est assez bonne pour les usages que l'on veut en faire.

Cette fonction RANDOM respecte en particulier deux propriétés fondamentales de la loi continue uniforme sur $[0, 1]$

- (i) pour tous les réels a et b tels que $0 \leq a < b \leq 1$, on dispose de l'égalité

$$P(\text{RANDOM} \in [a, b]) = b - a;$$

- (ii) les appels successifs de RANDOM sont indépendants.

► **Remarque**

La fonction RANDOMIZE sert à initialiser la suite récurrente qui permet à l'ordinateur de fournir les résultats. Elle utilise le compteur de temps de la machine, de sorte que, lors de deux utilisations, à des moments forcément différents, le générateur est initialisé de deux manières différentes.

Si l'on n'utilise pas cette fonction, les résultats obtenus lors de deux utilisations sont les mêmes, ce qui n'est vraiment pas aléatoire.

Simulations élémentaires

L'utilisation simple d'un générateur de nombres au hasard permet de simuler beaucoup de lois élémentaires. Nous ne donnerons ici que quelques indications ; le lecteur est prié de se reporter, pour avoir des détails et des programmes explicites, à la dernière partie du dernier chapitre.

- *Loi uniforme discrète* sur $\llbracket n, p \rrbracket$ (avec $n < p$) : la variable $Y = n + \text{RANDOM}(p-n+1)$ suit la loi uniforme discrète sur $\llbracket n, p \rrbracket$.
- *Loi uniforme continue* sur $[a, b]$ (avec $a < b$) : la variable $Z = a + (b-a)\text{RANDOM}$ suit la loi uniforme continue sur $[a, b]$.
- *Loi de Bernoulli* de paramètre p (avec $0 < p < 1$) : si U est une variable uniforme sur l'intervalle $[0, 1]$, la variable X telle que $X = 1$ si, et seulement si, $U < p$ et $X = 0$ sinon suit la loi de Bernoulli de paramètre p .
- *Loi géométrique* de paramètre p : en utilisant une variable indicatrice de paramètre p et le calcul du temps d'attente du premier succès, on peut simuler cette loi.
- *Loi hypergéométrique* : la simulation de cette loi est plus délicate que les simulations précédentes, et nous renvoyons au dernier chapitre. Mais on peut également agir autrement. Invitons par exemple le lecteur à réfléchir aux questions suivantes

1. Considérons un tableau X à n cases : élaborer un programme permettant de remplir ce tableau avec n entiers **distincts** inférieurs ou égaux à N (avec bien sûr $n < N$).
2. Utiliser ce programme avec $N = a + b$ (où a et b sont des entiers strictement positifs) pour compter le nombre des entiers choisis qui sont inférieurs à a .
3. En déduire une simulation de la loi hypergéométrique de paramètre (N, n, p) .

Simulation de variables discrètes par fonctions indicatrices

La technique qui suit est très importante, et facile à mettre en œuvre informatiquement dès lors que l'on connaît les probabilités des différentes valeurs susceptibles d'être prises par la variable ainsi que ces valeurs elles-mêmes classées par ordre croissant. Nous supposons, pour simplifier, que ces valeurs notées (x_k) commencent à x_0 et poserons $p_k = P(X = x_k)$ et $q_{-1} = -\infty$, $q_0 = p_0$ et $q_{k+1} = q_k + p_{k+1}$ (si l'ensemble des valeurs est fini, il convient de préciser pour quelles valeurs de k les nombres p_k et q_k sont définis, mais c'est sans importance pour ce qui suit).

Naturellement on a aussi $q_k = p_0 + p_1 + \dots + p_k \in [0, 1]$ et l'une des deux relations $\bigcup_k]q_{k-1}, q_k] =]-\infty, 1[$ ou $\bigcup_k]q_{k-1}, q_k] =]-\infty, 1[$ (par exemple dans le cas où l'ensemble des x_k est fini).

Alors l'égalité $\psi(t) = \sum_k x_k \mathbf{1}_{]q_{k-1}, q_k]}(t)$ sur $]0, 1[$, complétée par $\psi(0) = \psi(1) = a$ avec $a < x_0$, a un sens (puisque au plus un terme de la somme est différent de 0) et définit sur $[0, 1]$ une fonction réelle telle que $\psi(t) = x_k$ soit équivalent à la relation $t \in]q_{k-1}, q_k]$ (ou $t \in]q_{k-1}, 1[$ si $q_k = 1$).

Soient U une variable continue de loi uniforme sur $[0, 1]$, telle que $P(U \leq s) = s$ pour $s \in [0, 1]$, et X une variable aléatoire discrète prenant les valeurs (x_k) comme plus haut avec les probabilités respectives (p_k) .

Alors la variable aléatoire $Y = \psi \circ U = \psi(U)$, encore définie par

$$Y = x_0 \mathbf{1}_{]-\infty, p_0]} + \sum_{k=1}^{+\infty} x_k \mathbf{1}_{]q_{k-1}, q_k]}$$

prend les mêmes valeurs que X et avec les mêmes probabilités respectives (p_k) puisque, au moins pour $k \geq 1$

$P(Y = x_k) = P(q_{k-1} < U \leq q_k) = P(U \leq q_k) - P(U \leq q_{k-1}) = q_k - q_{k-1} = p_k$
(le cas de x_0 se déduit de ces relations, par exemple en écrivant que $\sum p_k = 1$).

► Remarque

En fait on peut aussi avoir $Y = a$, mais avec une probabilité nulle, ce qui est donc sans importance.

Par suite, on peut utiliser Y pour simuler la variable aléatoire X .

Cette méthode permet de simuler toute loi discrète connue à partir de la loi uniforme U , mais certaines peuvent évidemment être simulées de manière plus simple (voir ci-dessus).

On trouvera par exemple, en fin de volume, comment simuler une loi de Poisson à l'aide de cette technique.

► **Remarque**

Le cas où les valeurs x_k n'ont pas d'élément minimum peut se ramener à celui-là, par des modifications techniques un peu lourdes à expliciter, mais faciles à deviner. Nous nous en tiendrons donc à ce qui a été supposé (exemple : variable aléatoire à image incluse dans \mathbb{N}).

Simulation de variables à densité par inversion de répartition

Soient U une variable aléatoire suivant la loi uniforme sur $[0, 1]$ et X une variable aléatoire dont la fonction de répartition, croissante et **supposée bijective** de \mathbb{R} sur $]0, 1[$, est notée F_X . On considère la variable aléatoire $Y = F_X^{-1} \circ U = F_X^{-1}(U)$. Elle prend ses valeurs dans l'intervalle $]0, 1[$, et, pour tout x de $]0, 1[$, on dispose des égalités

$$\begin{aligned} P(Y \leq x) &= P(F_X^{-1}(U) \leq x) = P(\{\omega \mid F_X^{-1}(U(\omega)) \leq x\}) \\ &= P(\{\omega \mid U(\omega) \leq F_X(x)\}) = P(U \leq F_X(x)) = F_X(x). \end{aligned}$$

Il en résulte que X et Y ont la même fonction de répartition F_X . Ainsi, pour tout n -échantillon (u_1, u_2, \dots, u_n) de la loi uniforme sur $[0, 1]$ (obtenu par exemple à l'aide d'un générateur de nombres au hasard), alors $(F_X^{-1}(u_1), F_X^{-1}(u_2), \dots, F_X^{-1}(u_n))$ est un n -échantillon de la loi de la variable X .

Cette méthode permet donc de simuler certaines lois à densité, à condition notamment que l'on sache calculer effectivement les valeurs de $F_X^{-1}(u)$.

► **Remarques**

- Cette méthode porte le nom d'**anamorphose**.
- Plus précisément, $Y(\omega)$ n'est pas défini si $U(\omega)$ vaut 0 ou 1, mais ces deux cas sont quasi-impossibles et peuvent donc être écartés.
- Pour une généralisation de cette méthode, due à Paul Lévy, voir l'exercice 2.
- Voici une variante de la preuve de la méthode d'anamorphose (mais qui ne se généralise pas aussi bien que celle qui figure ci-dessus).
Il suffit de remarquer que, **lorsque F_X est inversible**, la variable aléatoire $Z = F_X(X)$ suit une loi uniforme. En effet

$$F_Z(x) = P(Z \leq x) = P(F_X \circ X \leq x) = P(X \leq F_X^{-1}(x)) = F_X \circ F_X^{-1}(x) = x.$$

Cette technique sera, par exemple, employée dans le dernier chapitre pour simuler la *loi exponentielle*.

Simulation par utilisation de propriétés de stabilité

Pour simuler certaines variables aléatoires on peut utiliser, le cas échéant, le fait qu'elle est la somme de variables aléatoires indépendantes que l'on sait simuler. On trouvera ci-dessous quelques exemples importants, développés en fin d'ouvrage.

- Soit X une *variable binomiale* de paramètre (n, p) . On sait que X est la somme de n variables de Bernoulli indépendantes, de paramètre p .

- Soit X une *variable de Poisson* de paramètre entier n . On sait que X est la somme de n variables de Poisson indépendantes de paramètre 1.
- Soit X une *variable gamma* de paramètre (n, θ) où n est un entier strictement positif. On sait que X est la somme de n variables gamma indépendantes de paramètres 1 et θ . On sait d'autre part, que la loi gamma de paramètres 1 et θ est la loi exponentielle de paramètre θ .

Simulation par utilisation d'une convergence en loi

Pour simuler une variable aléatoire X , on peut également utiliser, le cas échéant, le fait qu'elle est la limite en loi d'une suite de variables aléatoires que l'on sait simuler.

Par exemple, d'après le théorème de la limite centrée, la somme de n variables uniformes sur l'intervalle $[0, 1]$ peut être approchée par une variable aléatoire suivant la loi normale de paramètres $\frac{n}{2}$ et $\frac{n}{12}$.

On peut en déduire une simulation de la loi normale centrée réduite, et aussi de toutes les autres lois normales.

2. Estimateurs

2.1 Position du problème

Dans la première section du présent chapitre, nous avons étudié les échantillons issus d'une variable aléatoire dont la loi était bien déterminée. Nous allons maintenant nous intéresser au cas où l'on est en présence d'une loi de probabilité dont certains paramètres sont inconnus.

La situation de référence est celle où l'on se trouve en face d'un phénomène dont les manifestations consistent en des résultats de mesures que l'on est incapable de prévoir. On fait alors le postulat suivant :

Les résultats observés sont les réalisations d'une variable aléatoire.

La loi de cette variable aléatoire X est inconnue. Le travail du statisticien consiste alors, à partir d'un échantillon de cette loi inconnue, à rechercher quelle loi théorique on peut retenir comme loi parente de l'échantillon.

Dans la grande majorité des cas, on suppose que l'échantillonnage a été effectuée de manière indépendante, et que les résultats successifs sont des réalisations de variables aléatoires suivant la même loi, de telle manière que l'échantillon obtenu soit un échantillon *iid*. On dit alors parfois que l'on est alors dans les conditions d'*échantillonnage aléatoire simple*.

Dans certains cas, on ne connaît rien *a priori* sur la loi de X . Mais souvent, des considérations sur les conditions de l'expérimentation permettent de préciser que la loi cherchée appartient à une certaine famille $(\mu_\theta)_{\theta \in \Theta}$ de lois μ_θ dépendant d'un paramètre θ réel ou vectoriel, dont on sait qu'il appartient à un certain ensemble Θ . C'est le seul cas que nous envisagerons dans cet ouvrage. La loi que l'on cherche à évaluer

est ainsi définie par son paramètre θ_0 . On dit que l'on est en présence d'un problème d'estimation paramétrique, et nous supposons qu'à deux valeurs distinctes θ et θ' du paramètre, correspondent deux lois distinctes μ_θ et $\mu_{\theta'}$.

Pour résumer, c'est à partir des réalisations d'un échantillon aléatoire que l'on doit évaluer (on dit aussi **estimer**) la valeur θ_0 du paramètre θ (réel ou vectoriel), qui permet de définir sans ambiguïté la loi μ_{θ_0} qui sera le meilleur choix possible comme loi parente de l'échantillon dans la famille $(\mu_\theta)_{\theta \in \Theta}$.

Dans la pratique, et parce que l'image d'un estimateur par une fonction ne conserve pas les qualités de l'estimateur initial (voir 1.3.2 exemple 2), on est souvent conduit à estimer non pas directement θ_0 mais une fonction de θ_0 (cette fonction pouvant être l'identité). Dans le présent ouvrage, nous nous bornerons au cas où cette fonction est à valeurs réelles.

2.2 Définitions

Définition 6

Soit g une fonction de Θ dans \mathbb{R} ; et soit $\mathcal{E}_n = (X_1, X_2, \dots, X_n)$ un n -échantillon d'une loi μ_θ . On appelle *estimateur* de $g(\theta_0)$ toute statistique φ_n sur l'échantillon \mathcal{E}_n prenant ses valeurs dans l'ensemble $g(\Theta)$ des valeurs possibles pour $g(\theta)$.

L'estimateur T_n est donc la variable aléatoire $T_n = \varphi_n(X_1, X_2, \dots, X_n)$

L'estimateur est une variable aléatoire dépendant de (X_1, X_2, \dots, X_n) . Comme les X_i suivent tous la loi μ_θ , l'estimateur T_n est une variable aléatoire dépendant de θ .

Les valeurs observées grâce auxquelles on cherchera à évaluer $g(\theta_0)$ sont des réalisations de cette variable aléatoire.

Définition 7

Soit $T_n = \varphi_n(X_1, X_2, \dots, X_n)$ un estimateur de $g(\theta_0)$. Une *estimation* de $g(\theta_0)$ est une réalisation $\varphi_n(x_1, x_2, \dots, x_n)$ de T_n où (x_1, x_2, \dots, x_n) est une réalisation de l'échantillon aléatoire observé (X_1, X_2, \dots, X_n) .

► Remarques

- De la même façon que dire qu'un décimal d est une valeur approchée de π n'a de sens que si l'on précise une incertitude sur l'approximation, dire que $\varphi_n(x_1, x_2, \dots, x_n)$ est une estimation de $g(\theta_0)$ ne prend son sens que si l'on précise les qualités de cette estimation. C'est ce que nous envisagerons dans la suite de cette section.
- L'estimation $\varphi_n(x_1, x_2, \dots, x_n)$ ne dépend que de l'échantillon observé.
- La définition ci-dessus est extrêmement générale. Elle ne donne aucune indication sur la manière de construire un estimateur. Mais souvent, l'interprétation que l'on peut donner du paramètre guide vers certains estimateurs, que l'on pourra considérer comme « naturels » (voir exemples ci-dessous).
- En pratique, on désigne la variable θ et la vraie valeur θ_0 que l'on cherche à évaluer par la même lettre θ . Il s'agit d'un abus commode, que nous commettrons souvent, le contexte permettant de comprendre de quoi l'on parle.

Exemples

1. On considère un dé dont on ne sait pas s'il est pipé ou non. On lance le dé n fois. Au k -ième jet du dé, on associe la variable aléatoire X_k qui prend la valeur 1 si le résultat obtenu est 6, et 0 dans les autres cas. Les variables X_k suivent une loi de Bernoulli de paramètre p dont (X_1, X_2, \dots, X_n) est un n -échantillon. Dans cet exemple, on prend $\theta = p$ et $\Theta = [0, 1]$.

La variable aléatoire $T_n = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$ (moyenne empirique de l'échantillon) prend ses valeurs dans $[0, 1]$. C'est l'estimateur le plus naturel de p ($= \theta$).

On peut envisager bien d'autres estimateurs de T_n . Par exemple

$$U_n = \frac{2}{n(n+1)} \sum_{k=1}^n kX_k, \quad V_n = \sum_{k=0}^{n-1} X_k X_{k-1}$$

sont des estimateurs de p_0 . En l'absence d'autres indications, on se demande pourquoi on pourrait les envisager, mais ils existent, comme beaucoup d'autres.

2. On sait que X est une variable aléatoire uniforme continue sur un intervalle $[a, b]$. On répète n fois l'expérience. À la $k^{\text{ème}}$ expérience, on associe la variable aléatoire X_k continue uniforme sur $[a, b]$. On pose $\theta = (a, b)$ (paramètre vectoriel) et $g(\theta) = b - a$ qui peut prendre toutes les valeurs réelles positives (on suppose que $a \leq b$). Un estimateur naturel de $g(\theta)$ peut être $T_n = \sup_{k \in \llbracket 1, n \rrbracket} X_k - \inf_{k \in \llbracket 1, n \rrbracket} X_k$.

Si l'on connaît bien la loi uniforme, on peut aussi considérer que l'estimateur

$$Y_n = \sqrt{\frac{12}{n} \sum_{k=1}^n (X_k - T_n)^2}$$

est aussi un estimateur naturel de $\theta = b - a$, mais il est bien entendu beaucoup moins facile à manipuler.

On pourrait envisager bien d'autres estimateurs de $b - a$. Par exemple

$$U_n = \frac{2}{n(n-1)} \sum_{j=1}^n \sum_{k=1}^n (X_j - X_k)^2$$

ou encore

$$V_n = \frac{2}{n} \left(\sum_{k=1}^n X_k - \inf_{k \in \llbracket 1, n \rrbracket} X_k \right).$$

La raison d'être et l'intérêt de ces estimateurs ne sont pas clairs *a priori*, mais ce sont des estimateurs de $b - a$, comme beaucoup d'autres.

2.3 Biais d'un estimateur

Pour construire un estimateur permettant d'obtenir des évaluations de bonne qualité du paramètre étudié, pour choisir entre deux estimateurs, il faut se donner des critères de qualité pour un estimateur. Si l'on veut estimer $g(\theta)$ par les valeurs prises par la variable aléatoire T_n , il faut que ces valeurs ne s'éloignent pas trop de $g(\theta)$.

Un estimateur étant une variable aléatoire, et l'espérance d'une variable aléatoire étant (lorsqu'elle existe) la principale caractéristique de tendance centrale, c'est tout naturellement que l'on s'intéressera à la différence entre l'espérance de cet estimateur et la vraie valeur du réel $g(\theta_0)$ que l'on cherche à évaluer (vraie valeur naturellement inconnue).

Définition 8

Soit T_n un estimateur de $g(\theta)$. Si T_n admet une espérance pour tout θ , on appelle *biais* de T_n le réel

$$b_{T_n}(\theta) = E_\theta(T_n) - g(\theta).$$

► **Remarques**

- Rappelons que la variable aléatoire T_n dépend *a priori* de θ , et que, par conséquent son espérance dépend de θ . C'est ainsi que cette espérance peut se noter parfois $E_\theta(T_n)$.
- Le biais d'un estimateur peut être positif ou négatif. Il est clair que l'on cherchera à obtenir un estimateur admettant un biais le plus faible possible (en valeur absolue), l'idéal étant un biais nul.
- Il arrive que l'on parle du biais de T_n en θ .
- La valeur importante du biais est bien entendu $b_{T_n}(\theta_0)$, qui est inconnue. Si l'on remplace ce paramètre par l'estimation qu'on en a faite, on obtient une valeur approchée du biais, si b_{T_n} est une fonction continue.

Définition 9

La variable aléatoire T_n est un *estimateur sans biais* de $g(\theta)$ si $b_{T_n}(\theta) = 0$, c'est-à-dire si $E_\theta(T_n) = g(\theta)$.

Exemples

1. Soit m l'espérance de la loi parente de l'échantillon, et considérons $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ (moyenne empirique de l'échantillon) comme un estimateur de m . Les résultats obtenus à la section précédente permettent de conclure que \bar{X}_n est un estimateur sans biais de m .
2. La variance empirique S_n^2 n'est pas un estimateur sans biais de la variance σ^2 de la loi parente. En effet, on a obtenu à la section précédente l'égalité $E(S_n^2) = \frac{n-1}{n} \sigma^2$. Le biais de la variance empirique, en tant qu'estimateur de la variance σ^2 est donc $E_{S_n^2}(\sigma^2) = \frac{n-1}{n} \sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}$.
3. On reprend l'exemple d'une variable aléatoire parente X uniforme continue sur un intervalle $[a, b]$, et de l'estimateur $T_n = \sup_{k \in \llbracket 1, n \rrbracket} X_k - \inf_{k \in \llbracket 1, n \rrbracket} X_k$ de $b - a$.

Alors $E(T_n) = E \left(\sup_{k \in \llbracket 1, n \rrbracket} X_k \right) - E \left(\inf_{k \in \llbracket 1, n \rrbracket} X_k \right)$.

La fonction de répartition de tous les X_k est $F_{X_k} = \frac{x-a}{b-a} \mathbf{1}_{[a,b]}$.

Posons $S_n = \sup_{k \in \llbracket 1, n \rrbracket} X_k$.

On sait (voir 2) que la fonction de répartition de S_n est alors $F_{S_n}(x) = \left(\frac{x-a}{b-a}\right)^n \mathbb{1}_{[a,b]}$.

$$\text{Donc } E(S_n) = \int_{-\infty}^{+\infty} nt \frac{(t-a)^{n-1}}{(b-a)^n} \mathbb{1}_{[a,b]}(t) dt = n \int_a^b t \frac{(t-a)^{n-1}}{(b-a)^n} dt$$

$$\text{soit, en posant } u = t - a, E(S_n) = \frac{n}{(b-a)^n} \int_0^{b-a} (u+a)u^{n-1} du = \frac{a+nb}{n+1}.$$

De façon analogue, en posant $I_n = \inf_{k \in \llbracket 1, n \rrbracket} X_k$, et après des calculs analogues, on trouve

$$E(I_n) = \frac{b+na}{n+1}.$$

$$\text{Alors } E(T_n) = E(S_n) - E(I_n) = \frac{(a+nb) - (b+na)}{n+1} = \frac{n-1}{n+1} (b-a).$$

Par suite T_n est donc un estimateur biaisé de $b-a$, le biais étant égal à $-\frac{2}{n+1} (b-a)$.

► Remarques

- On donne souvent le biais en valeur absolue. Ainsi dira-t-on, dans les exemples ci-dessus, que le biais de l'estimateur S_n^2 de σ^2 est $\frac{\sigma^2}{n}$ et que celui de T_n est $\frac{2}{n+1} (b-a)$.
- Rappelons que la lettre qui désigne le paramètre à estimer peut désigner tour à tour un élément quelconque de l'ensemble des paramètres, ou la vraie valeur que l'on cherche à estimer. Lorsque nous voulons insister sur le fait que l'on veut parler de la vraie valeur, nous la notons avec l'indice 0.
- Lorsqu'il n'est pas nul, le biais d'un estimateur dépend en général de n , et l'important est d'étudier son comportement lorsque n tend vers l'infini.

2.4 Risque quadratique d'un estimateur

La qualité d'un estimateur ne dépend pas seulement de la proximité de son espérance avec la vraie valeur du paramètre à estimer, mais aussi de la dispersion des valeurs qu'il prend autour de cette valeur à estimer.

Définition 10

Soit T_n un estimateur de $g(\theta)$. Si T_n admet un moment d'ordre 2 pour tout θ , on appelle *risque quadratique* ou *erreur quadratique moyenne* de T_n le réel

$$r_{T_n}(\theta) = E_{\theta} \left([T_n - g(\theta)]^2 \right).$$

► Remarque

L'expression « erreur quadratique moyenne » semble plus répandue que l'expression « risque quadratique », bien que cette dernière semble plus correcte.

Théorème 6

Le risque quadratique d'un estimateur est la somme de sa variance et du carré de son biais.

Preuve

En effet :

$$\begin{aligned} r_{T_n}(\theta) &= E_{\theta} \left([(T_n - E_{\theta}(T_n)) - (E_{\theta}(T_n) - g(\theta))]^2 \right) \\ &= E_{\theta} \left([T_n - E_{\theta}(T_n)]^2 \right) + E_{\theta} \left([E_{\theta}(T_n) - g(\theta)]^2 \right) \\ &\quad - 2 E_{\theta} \left([T_n - E_{\theta}(T_n)] [E_{\theta}(T_n) - g(\theta)] \right). \end{aligned}$$

Compte tenu de la linéarité de l'espérance, du fait que $E_{\theta}(T_n) - g(\theta) = b_{T_n}(\theta)$ est une constante et que la variable centrée $T_n - E_{\theta}(T_n)$ a une espérance nulle, on en déduit

$$r_{T_n}(\theta) = E_{\theta} \left([T_n - E_{\theta}(T_n)]^2 \right) + (b_{T_n}(\theta))^2 = V_{\theta}(T_n) + (b_{T_n}(\theta))^2$$

c'est-à-dire le résultat annoncé. □

Corollaire 1

Quand un estimateur est sans biais, son risque quadratique est égal à sa variance.

Preuve

Cela résulte immédiatement du théorème ci-dessus. □

Exemples

1. On suppose que X , loi parente d'échantillon suit une loi de Bernoulli de paramètre p .

Considérons $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ (moyenne empirique de l'échantillon) comme un estimateur de p . Les résultats obtenus plus haut permettent de conclure que le risque quadratique de l'estimateur \bar{X}_n est sa variance, c'est à dire $\frac{p(1-p)}{n}$.

2. On reprend l'exemple d'une variable aléatoire parente X uniforme continue sur un intervalle $[a, b]$. Considérons $S_n = \sup_{k \in \llbracket 1, n \rrbracket} X_k$ comme un estimateur de b .

On a calculé plus haut $E(S_n) = \frac{n}{(b-a)^n} \int_0^{b-a} (u+a)u^{n-1} du = \frac{a+nb}{n+1}$.

Par suite S_n est donc un estimateur biaisé de b . Son biais est négatif (ce que l'on aurait pu prévoir), et vaut $\frac{a-b}{n+1}$.

En supposant $a = 0$ et $b = 1$, on obtient $E(S_n'^2) = n \int_0^1 t^2 t^{n-1} dt = \frac{n}{n+2}$.

Ce qui conduit à

$$V(S_n') = E(S_n'^2) - (E(S_n'))^2 = \frac{n}{n+2} - \left(\frac{n}{n+1} \right)^2 = \frac{n}{(n+1)^2(n+2)}.$$

Revenant au cas général par la transformation $x \mapsto (b-a)x + a$, on obtient

$$V(S_n) = \frac{n(b-a)^2}{(n+1)^2(n+2)}.$$

On en déduit le risque quadratique de l'estimateur S_n de b

$$r_{S_n}(b) = \frac{n(b-a)^2}{(n+1)^2(n+2)} + \left(\frac{a-b}{n+1} \right)^2 = \frac{2(b-a)^2}{(n+1)(n+2)}.$$

► **Remarques**

- L'erreur quadratique moyenne (le risque quadratique) d'un estimateur dépend du paramètre à estimer, qui est inconnu. Pour en donner une valeur numérique approchée, on remplace ce paramètre par l'estimation qu'on en a faite. Il dépend aussi de n , et l'important, c'est d'étudier son comportement quand n tend vers l'infini.
- Il ne faut pas croire que, entre deux estimateurs, il faille systématiquement choisir celui dont le biais est le plus petit (en valeur absolue), éventuellement en priorité celui qui serait sans biais. En réalité, on est parfois conduit à préférer un estimateur biaisé, mais dont les valeurs sont très regroupées autour de la valeur moyenne, ce qui se traduit par une variance faible. C'est pourquoi l'erreur quadratique moyenne tient compte de la variance et du carré du biais.

3. Suites d'estimateurs

3.1 Généralités

De façon générale, un estimateur dépend non seulement des valeurs possibles du paramètre à estimer, mais aussi du nombre n de variables aléatoires figurant dans l'échantillon. Sauf cas très particulier, on définit donc, non un estimateur seul, mais une suite d'estimateurs. C'est la raison pour laquelle, dans la pratique, on désignera de la même façon un estimateur, et la suite d'estimateurs qui lui est associée. Cet abus de langage est sans importance pratique.

La loi des grands nombres, et plus généralement les propriétés de convergence, nous laissent entendre que souvent, plus n est grand, meilleure est l'approximation.

Définition 11

Une suite $(T_n)_{n \in \mathbb{N}^*}$ d'estimateurs de $g(\theta)$ est *asymptotiquement sans biais* si, pour tout θ de Θ , $\lim_{n \rightarrow \infty} E_\theta(T_n) = g(\theta)$.

► **Remarque**

On dit aussi que l'estimateur T_n est asymptotiquement sans biais.

Définition 12

Une suite $(T_n)_{n \in \mathbb{N}^*}$ d'estimateurs de $g(\theta)$ est *convergente* si, pour tout θ de Θ , la suite $(T_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers la variable certaine $g(\theta)$, c'est-à-dire si

$$\forall \varepsilon \in \mathbb{R}_+, \lim_{n \rightarrow \infty} P_\theta(|T_n - g(\theta)| > \varepsilon) = 0.$$

► **Remarque**

On dit aussi que l'estimateur T_n est convergent (ou *consistant*).

3.2 Convergence et risque quadratique

Théorème 7

Soit T_n un estimateur de $g(\theta)$. Si le risque quadratique de T_n tend vers 0 quand n tend vers l'infini, alors T_n est un estimateur convergent de $g(\theta)$.

Preuve

Soit ε un réel strictement positif quelconque.

Les événements $[|T_n - g(\theta)| > \varepsilon]$ et $[(T_n - g(\theta))^2 > \varepsilon^2]$ sont égaux. On obtient, avec le théorème de transfert

$$E[(T_n - g(\theta))^2] > \varepsilon^2 P(|T_n - g(\theta)| > \varepsilon) \geq 0.$$

Il en résulte que, si le risque quadratique $E[(T_n - g(\theta))^2]$ tend vers 0, il en est de même de $P(|T_n - g(\theta)| > \varepsilon)$, ce qui prouve bien le résultat annoncé. \square

► **Remarque**

Nous nous permettons l'abus de langage déjà signalé en disant que « l'estimateur est convergent », plutôt que « la suite d'estimateurs est convergente ».

On en déduit aisément deux corollaires concernant le cas particulier des estimateurs sans biais, et qui se démontrent aisément directement

Corollaire 2

Tout estimateur sans biais dont la variance tend vers zéro est convergent.

Preuve

Résulte directement du théorème précédent.

Donnons en une démonstration directe. D'après l'inégalité de Bienaymé-Tchebychev appliquée à T_n , on a, pour tout réel ε strictement positif

$$P_\theta(|T_n - g(\theta)| > \varepsilon) \leq \frac{V_\theta(T_n)}{\varepsilon^2}.$$

Comme par hypothèse $V_\theta(T_n)$ tend vers zéro lorsque n tend vers l'infini, il en résulte que $\lim_{n \rightarrow \infty} \frac{V_\theta(T_n)}{\varepsilon^2} = 0$.

Comme toute probabilité est positive, on en déduit que $\lim_{n \rightarrow \infty} P_\theta(|T_n - g(\theta)| > \varepsilon) = 0$, c'est-à-dire le résultat annoncé. \square

Corollaire 3

Tout estimateur asymptotiquement sans biais dont la variance tend vers zéro est convergent, c'est-à-dire que

$$\left[E_\theta(T_n) \xrightarrow{n \rightarrow +\infty} g(\theta) \text{ et } V_\theta(T_n) \xrightarrow{n \rightarrow +\infty} 0 \right] \implies \left[T_n \xrightarrow[n \rightarrow +\infty]{\mathcal{P}} g(\theta) \right].$$

Preuve

Résulte directement du théorème ci-dessus et du théorème 6.

Ce résultat est aussi une conséquence immédiate des conditions suffisantes de convergence en probabilité rencontrées au chapitre précédent. \square

Appliqués au cas de la moyenne empirique, ces résultats permettent de conclure

Corollaire 4

Si les μ_θ admettent une variance, la moyenne empirique est un estimateur convergent de l'espérance des μ_θ .

Preuve

Soient m et σ^2 l'espérance et la variance de la loi parente de l'échantillon. Pour tout θ , les résultats de la première section du présent chapitre s'appliquent, et si l'on note $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ la moyenne empirique de l'échantillon, on a $E(\bar{X}_n) = m$ (c'est à dire que l'estimateur est sans biais) et $V(\bar{X}_n) = \frac{\sigma^2}{n}$. Il en résulte que $V(\bar{X}_n)$ tend vers 0 quand n tend vers l'infini, ce qu'il fallait démontrer. \square

► Remarque

Rappelons que, quelle que soit la valeur du paramètre, \bar{X}_n converge en probabilité vers la variable certaine égale à m , et que la variable centrée réduite associée à \bar{X}_n converge en loi vers une variable normale centrée réduite (donc de loi indépendante du paramètre).

Exemples

1. On reprend l'exemple d'une variable aléatoire parente X uniforme continue sur un intervalle $[a, b]$. On a vu plus haut que l'estimateur $S_n = \sup_{k \in \llbracket 1, n \rrbracket} X_k$ de b est tel que

$$E(S_n) = \frac{a + nb}{n + 1} \text{ et } V(S_n^2) = \frac{n(b-a)^2}{(n+1)^2(n+2)}.$$

On a alors clairement $E(S_n) \xrightarrow{n \rightarrow +\infty} b$ et $V(T_n) \xrightarrow{n \rightarrow +\infty} 0$, ce qui prouve que S_n est un estimateur asymptotiquement sans biais et convergent de b .

2. Soit (X_1, X_2, \dots, X_n) un n -échantillon iid d'une variable aléatoire parente X suivant une loi de Poisson de paramètre λ inconnu. On sait que la moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ est un estimateur sans biais et convergent de l'espérance λ de la loi de X .

Soit x un réel connu. Si l'on cherche à estimer $e^{-\lambda x}$, il semble naturel d'envisager l'estimateur $T_n = e^{-x\bar{X}_n}$. Calculons donc $E(T_n)$.

Remarquons tout d'abord que $\bar{X}_n = \frac{1}{n} Y_n$, où $Y_n = \sum_{k=1}^n X_k$ suit la loi de Poisson de paramètre $n\lambda$.

On a donc, d'après le théorème de transfert

$$\begin{aligned} E(T_n) &= E\left(\exp\left(-\frac{x}{n} Y_n\right)\right) = \sum_{k=0}^{+\infty} \exp\left(-\frac{xk}{n}\right) P(Y_n = k) \\ &= \sum_{k=0}^{+\infty} \exp\left(-\frac{xk}{n}\right) \frac{(n\lambda)^k e^{-n\lambda}}{k!}. \end{aligned}$$

On obtient alors

$$E(T_n) = e^{-n\lambda} \sum_{k=0}^{+\infty} \frac{(n\lambda e^{-\frac{x}{n}})^k}{k!} = e^{-n\lambda} e^{n\lambda e^{-\frac{x}{n}}} = e^{-n\lambda \left(1 - e^{-\frac{x}{n}}\right)}.$$

Cela prouve que T_n n'est pas un estimateur sans biais de $e^{-\lambda x}$.

Mais on sait que $(1 - e^{-\frac{x}{n}}) \sim \frac{x}{n}$, et donc $-n\lambda(1 - e^{-\frac{x}{n}}) \xrightarrow[n \rightarrow +\infty]{} -\lambda x$, et donc que

$$E(T_n) \xrightarrow[n \rightarrow +\infty]{} e^{-\lambda x}.$$

Par suite T_n est un estimateur asymptotiquement sans biais de $e^{-\lambda x}$.

Calculons alors $V(T_n) = E(T_n^2) - (E(T_n))^2$. On remarque que $T_n^2 = e^{-\frac{2x}{n}} Y_n$.

Le calcul de $E(T_n^2)$ s'obtient donc, à partir de celui de $E(T_n)$ en remplaçant x par $2x$.

On obtient alors $V(T_n) = e^{-n\lambda(1 - e^{-\frac{2x}{n}})} - e^{-2n\lambda(1 - e^{-\frac{x}{n}})}$.

En étudiant comme plus haut les limites des deux termes de cette différences, on conclut sans peine que $\lim_{n \rightarrow +\infty} V(T_n) = 0$.

En conclusion, T_n est un estimateur asymptotiquement sans biais et convergent de $e^{-\lambda x}$.

3.3 Image par une fonction continue

Il est important de remarquer, grâce à l'un des exemples ci-dessus, que l'image par une fonction f d'un estimateur sans biais de α n'est pas en général un estimateur sans biais de $f(\alpha)$, et ceci, même si la fonction f est continue.

Toutefois, on dispose d'un théorème concernant l'image par une fonction continue d'un estimateur convergent.

Théorème 8

Soient T_n un estimateur convergent de $g(\theta)$ et f une fonction à valeurs réelles, continue sur $g(\Theta)$. Alors $f(T_n)$ est un estimateur convergent de $f(g(\theta))$.

Preuve

Soit ε un réel strictement positif quelconque. La fonction f étant continue sur $g(\Theta)$, il existe un réel strictement positif α tel que, si $|T_n - g(\theta)| \leq \alpha$, alors $|f(T_n) - f(g(\theta))| \leq \varepsilon$. Cela signifie que l'événement $[|T_n - g(\theta)| \leq \alpha]$ implique l'événement $[|f(T_n) - f(g(\theta))| \leq \varepsilon]$, et par conséquent

$$P(|T_n - g(\theta)| \leq \alpha) \leq P(|f(T_n) - f(g(\theta))| \leq \varepsilon) \leq 1.$$

Or T_n est un estimateur convergent de $g(\theta)$, ce qui peut s'exprimer par le fait que, pour tout réel strictement positif α , $\lim_{n \rightarrow +\infty} P(|T_n - g(\theta)| \leq \alpha) = 1$. Il en résulte, par encadrement, que $\lim_{n \rightarrow +\infty} P(|f(T_n) - f(g(\theta))| \leq \varepsilon) = 1$, ce qui prouve bien que $f(T_n)$ converge en probabilité vers $f(g(\theta))$. \square

Exemple

Considérons un n -échantillon *iid* de la loi uniforme sur un intervalle $[0, \theta]$, où θ est un réel strictement positif. On sait que la moyenne empirique \bar{X}_n de l'échantillon est un estimateur sans biais et convergent de l'espérance $\frac{\theta}{2}$ de la loi parente. On peut conclure du théorème ci-dessus que $Y_n = 2\bar{X}_n$ est un estimateur convergent de θ . On ne peut par contre pas affirmer qu'il est sans biais (ni même asymptotiquement sans biais).

4. Estimation par intervalles de confiance

4.1 Première approche

Le résultat d'une estimation est une valeur approchée du paramètre que l'on cherche à évaluer. Si l'on effectue une autre estimation, on n'obtiendra en général pas le même résultat. C'est pourquoi donner un tel résultat approché sans indication sur la précision de l'évaluation n'a pas grand intérêt. Plutôt que de donner une (ou plusieurs) estimations numériques, on cherchera, à partir de l'estimateur dont on dispose, à préciser un intervalle qui contiendra, avec une probabilité donnée, la valeur exacte du paramètre que l'on cherche à évaluer.

Définition 13

Soit $\mathcal{E}_n = (X_1, X_2, \dots, X_n)$ un n -échantillon issu d'une loi μ_θ . Soit α un réel quelconque de l'intervalle $]0, 1[$. On appelle *intervalle de confiance* pour le paramètre θ , *au risque α* (ou *au niveau de confiance $1-\alpha$*), tout intervalle de la forme $[I_n, S_n]$ où I_n et S_n sont des estimateurs de θ (c'est à dire des statistiques sur l'échantillon \mathcal{E}_n) tels que $P(\theta \in [I_n, S_n]) = 1 - \alpha$.

► Remarques

- Un intervalle de confiance est un intervalle dont les bornes sont aléatoires et qui contient, avec une probabilité donnée, la valeur θ que l'on cherche à évaluer. Cette valeur n'est pas aléatoire. Elle est seulement inconnue.
- Le nombre α est le risque qu'à l'issue d'une expérience la réalisation de l'intervalle de confiance ne contienne pas la valeur θ que l'on cherche à évaluer.
- C'est dans un souci de simplification que nous appelons simplement θ , et non $g(\theta)$ comme ci-dessus, le paramètre à estimer. Ce changement de notation n'a aucune conséquence pratique.

Le problème étant ainsi posé, on peut penser à utiliser l'inégalité de Bienaymé-Tchebychev, qui s'écrit, pour un estimateur T_n de θ et un réel α de l'intervalle $[0, 1]$

$$P\left(T_n \in \left[E(T_n) - \frac{\sigma(T_n)}{\sqrt{\alpha}}, E(T_n) + \frac{\sigma(T_n)}{\sqrt{\alpha}}\right]\right) \geq 1 - \alpha.$$

Ainsi, si T_n est un estimateur de θ , l'intervalle $\left[E(T_n) - \frac{\sigma(T_n)}{\sqrt{\alpha}}, E(T_n) + \frac{\sigma(T_n)}{\sqrt{\alpha}}\right]$ est un intervalle de confiance pour θ à un risque inférieur ou égal à α (ou à un niveau de confiance supérieur ou égal à $1 - \alpha$).

Exemple

On considère un dé à jouer non pipé, On le lance n fois, et l'on cherche à déterminer un intervalle dans lequel la fréquence T_n des résultats « six » se situe avec une probabilité supérieure à $1 - \alpha$. La variable T_n est la fréquence empirique et l'on a $E(T_n) = \frac{1}{6}$ et

$\sigma(T_n) = \frac{\sqrt{5}}{6\sqrt{n}}$. On obtient ainsi comme intervalle de confiance à un risque inférieur à α ,

l'intervalle $\left[\frac{1}{6} - \frac{\sqrt{5}}{6\sqrt{n\alpha}}; \frac{1}{6} + \frac{\sqrt{5}}{6\sqrt{n\alpha}}\right]$.

Ainsi, pour $\alpha = 0,05$, on a $P\left(T_n \in \left[\frac{1}{6} - \frac{5}{3\sqrt{n}}; \frac{1}{6} + \frac{5}{3\sqrt{n}}\right]\right) \geq 0,95$.

Pour $n = 100$ on obtient alors $P(T_n \in [0; 0,34]) \geq 0,95$.

► **Remarques**

- Dans l'exemple précédent, nous avons changé, pour des raisons évidentes, l'écriture d'un intervalle $[a, b]$ en $[a; b]$. Cela se produira encore souvent dans les pages qui suivent.
- Les encadrements obtenus à l'aide de la formule de Bienaymé-Tchebychev ne dépendent pas de la nature de la variable aléatoire T_n . L'obtention de l'intervalle de confiance n'utilise que le calcul de l'espérance et de la variance de T_n (et les bornes de l'intervalle de confiance obtenu sont des variables certaines). Parce qu'elle est une formule très générale, et ainsi qu'on aura pu le constater sur l'exemple ci-dessus, les résultats qu'elle permet d'obtenir sont de mauvaise qualité, d'où la nécessité d'affiner la réflexion.
Pour obtenir des encadrements plus précis, on envisage la possibilité de tenir compte de la répartition des valeurs prises par les variables aléatoires utilisées, et en particulier de la dispersion de ces valeurs dans l'ensemble support.

4.2 Intervalles de dispersion

Dans toute cette sous-section, X désigne une variable aléatoire à densité donnée. On appelle *support* de X son image $X(\Omega)$.

Définition 14

Soit X une variable aléatoire à densité dont le support est un intervalle I , et dont une densité s'annule au plus sur un ensemble fini d'éléments de I . Sa fonction de répartition F_X est alors continue, strictement croissante sur I . Elle induit donc une bijection de I sur un intervalle dont les bornes sont 0 et 1, et qui est fermé, semi-ouvert ou ouvert suivant la forme de l'intervalle I . On peut alors définir la fonction réciproque $F_X^{-1} = Q_X$. Cette fonction prend le nom de *fonction quantile* de X .

► **Remarques**

- Les variables aléatoires à densité usuelles satisfont aux conditions de cette définition. C'est ainsi que, pour la loi uniforme sur $[a, b]$, on a naturellement $I = [a, b]$; pour une loi exponentielle, on a $I = [0, +\infty[$; pour une loi normale, on a $I =]-\infty, +\infty[$; pour des lois gamma et Log-normale, on a $I =]0, +\infty[$; pour une loi de Pareto, on a $I =]x_0, +\infty[$; pour une loi Beta, on a $I =]0, 1[$. . .
- Sauf dans quelques cas particuliers (loi uniforme, loi exponentielle, loi de Weibull . . .), la fonction de répartition d'une variable aléatoire à densité ne s'exprime pas à l'aide des fonctions usuelles; on ne peut pas exprimer $F_X(x)$ sans signe d'intégration. Cela a deux conséquences importantes. Tout d'abord, on ne connaît la fonction F_X que par des valeurs approchées, que l'on recalcule au besoin à l'aide d'un ordinateur, ou que l'on lit dans une table numérique. Surtout, les valeurs de la fonction quantile ne sont elles-mêmes connues que par des valeurs approchées, que l'on détermine en général par lecture inverse des tables numériques de la fonction F_X , en utilisant éventuellement une interpolation linéaire.
- Dans le cas de la définition ci-dessus, la fonction quantile est définie par $Q_X(u) = t$ tel que $P(X \leq t) = u$. Pour $u = \frac{1}{2}$, on reconnaît en particulier la définition de la médiane (qui ici est sans ambiguïté). On reconnaît de manière analogue les définitions des quartiles, déciles, centiles . . .
- Si X n'est pas une variable à densité, ou si $X(\Omega)$ n'est pas un intervalle, on définit usuellement une fonction quantile par $Q(u) = \inf\{t \mid P(X \leq t) \geq u\}$. Mais cette définition ne fait pas l'unanimité (voir Tome 1, pp 901 et suivantes).

Définition 15

Soient X une variable aléatoire et α un réel strictement compris entre 0 et 1. On appelle *intervalle de dispersion* de niveau $1 - \alpha$ tout intervalle $[a, b]$ inclus dans $X(\Omega)$ tel que $P(X \in [a, b]) = 1 - \alpha$.

Comme $P(X \in [a, b]) = F_X(b) - F_X(a)$, si l'on pose $\beta = F_X(a)$, on a $a = Q_X(\beta)$ et $F_X(b) = 1 - \alpha + \beta$. Si β est un réel de l'intervalle $[0, \alpha]$, on obtient alors un intervalle de dispersion de niveau $1 - \alpha$ sous la forme $[Q_X(\beta), Q_X(1 - \alpha + \beta)]$ (avec $Q_X = F_X^{-1}$ dans les conditions de la première définition ci-dessus). On dispose alors de la proposition suivante

Proposition 1

Soient X une variable aléatoire à densité satisfaisant aux conditions de la définition ci-dessus et α un réel de l'intervalle $]0, 1[$. Alors, pour tout réel β de l'intervalle $[0, \alpha]$, l'intervalle $[F_X^{-1}(\beta), F_X^{-1}(1 - \alpha + \beta)]$ est un intervalle de dispersion de niveau $1 - \alpha$ pour la variable X .

► **Remarques**

- Dans la pratique, α est « petit » (en général inférieur à 0,1), de sorte qu'un intervalle de dispersion de niveau $1 - \alpha$ contient une forte proportion des valeurs prises par X . Dans ces conditions, plus on pourra trouver un intervalle de dispersion de niveau $1 - \alpha$ dont les bornes seront « voisines », moins les valeurs prises par X seront « dispersées ».
- Une variable aléatoire X donnée admet autant d'intervalles de dispersion de niveau $1 - \alpha$ qu'il y a de façons de choisir β dans l'intervalle $[0, \alpha]$, c'est à dire une infinité. On distingue alors l'intervalle de dispersion *symétrique* (pour $\beta = \frac{\alpha}{2}$), l'intervalle de dispersion *unilatéral inférieur* (pour $\beta = 0$), et l'intervalle de dispersion *unilatéral supérieur* pour $\beta = \alpha$, et l'on cherche à déterminer l'intervalle de dispersion dont la longueur est la plus faible (intervalle de dispersion *optimal*). On démontre que, dans le cas d'une variable aléatoire symétrique (c'est à dire dont la fonction de répartition admet un centre de symétrie) l'intervalle de dispersion optimal est l'intervalle de dispersion symétrique.
- Du fait que les valeurs utilisables de la fonction quantile sont des valeurs approchées des valeurs exactes, les intervalles de dispersion sont obtenus de manière approchée. On arrondit alors de manière à être sûr que l'intervalle de dispersion de niveau $1 - \alpha$ obtenu contienne *au moins* une proportion $1 - \alpha$ des valeurs prises par X , de telle façon que l'on aura $P(X \in [a, b]) \geq 1 - \alpha$. Mais on ne peut pas prendre cette relation comme définition, car alors, tout intervalle contenant $[a, b]$, en particulier I serait un intervalle de dispersion de niveau $1 - \alpha$, ce qui ôterait tout intérêt à la notion ou qui obligerait à la définir de manière plus compliquée. Les intervalles de dispersion de niveau $1 - \alpha$ ainsi obtenus sont en réalité des intervalles de dispersion de niveau « légèrement supérieur » à $1 - \alpha$.

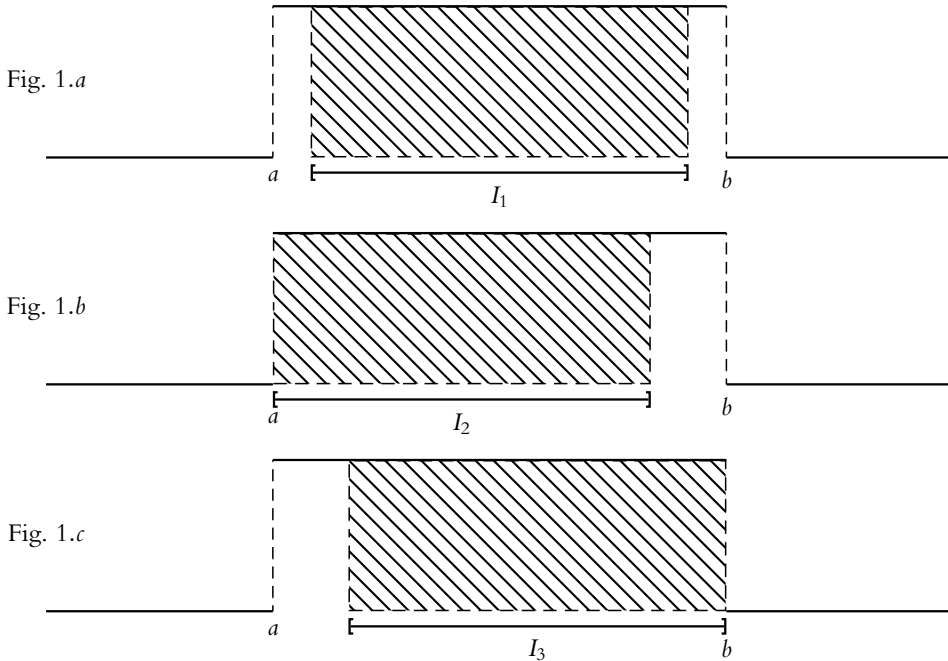
Exemples

1. Considérons une variable aléatoire X suivant la loi uniforme sur un intervalle fermé borné $I = [a, b]$. Une densité de X est la fonction constante égale à $\frac{1}{b - a}$ sur I , et nulle pour tout x n'appartenant pas à I . Sa fonction de répartition F_X est définie sur I par $F_X(x) = \frac{x - a}{b - a}$. La fonction quantile est alors définie par $Q_X(t) = a + t(b - a)$.

Pour $\alpha = 0,1$ (et donc $\frac{\alpha}{2} = 0,05$), on obtient

- (i) l'intervalle de dispersion symétrique $I_1 = [a + 0,05(b - a); a + 0,95(b - a)]$
- (ii) l'intervalle de dispersion unilatéral inférieur $I_2 = [a; a + 0,9(b - a)]$
- (iii) l'intervalle de dispersion unilatéral supérieur $I_3 = [a + 0,9(b - a); b]$.

Sur cet exemple particulièrement simple, il est clair que les intervalles de dispersion trouvés contiennent 90 % des valeurs prises par la variable aléatoire X (voir les figures qui suivent).



2. Considérons une variable aléatoire X suivant la loi exponentielle de paramètre 1. Une densité de X est la fonction f_X définie sur $[0, +\infty[$ par $f_X(x) = e^{-x}$, et nulle pour tout x n'appartenant pas à I . Sa fonction de répartition F_X est définie sur $I = [0, +\infty[$ par $F_X(x) = 1 - e^{-x}$. La fonction quantile est alors définie par $Q_X(t) = -\ln(1 - t)$.

Pour $\alpha = 0,1$ (et donc $\frac{\alpha}{2} = 0,05$), on obtient

- (i) l'intervalle de dispersion symétrique $I_1 = [-\ln 0,95; -\ln 0,05]$. En arrondissant $-\ln 0,95$ par défaut et $-\ln 0,05$ par excès, on pourra prendre $I_1 \approx [0,051; 2,996]$.
- (ii) l'intervalle de dispersion unilatéral inférieur $I_2 = [-\ln 1; -\ln 0,1]$. En arrondissant $-\ln 0,1$ par excès on pourra prendre $I_2 \approx [0; 2,303]$. Dans le cas de cet exemple, le choix naturel est celui de l'intervalle unilatéral inférieur.
- (iii) l'intervalle de dispersion unilatéral supérieur $I_3 = [-\ln 0,9; +\infty[$. On prendra $I_3 \approx [0,105; +\infty[$.

Ces différents intervalles de dispersion sont illustrés par les figures suivantes

Fig. 2.a

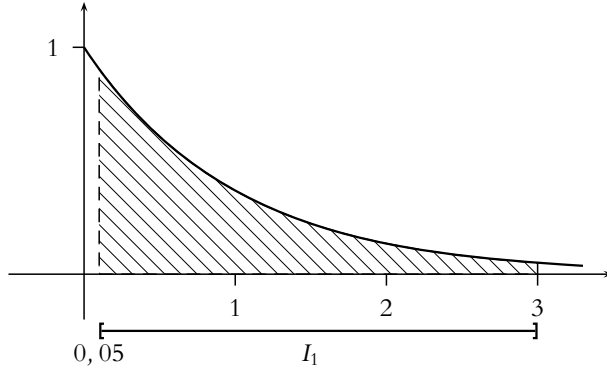


Fig. 2.b

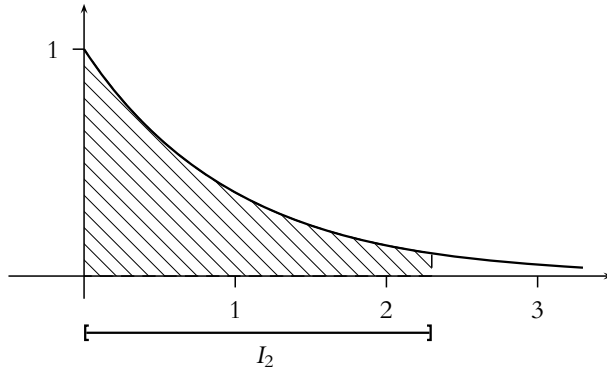
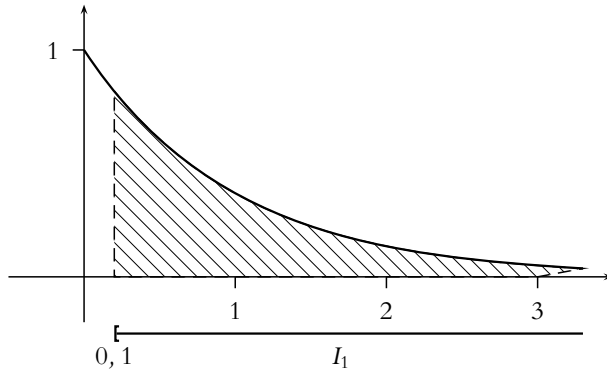


Fig. 2.c



3. Considérons une variable aléatoire X suivant la loi normale centrée réduite. Une densité de X est la fonction f_X définie sur $] -\infty, +\infty[$ par $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}$. Sa fonction de

répartition ne s'exprime pas autrement que par $F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2} dt$ (on note souvent Φ cette fonction de répartition). Pour utiliser la fonction réciproque de Φ , on aura recours à une table numérique.

Pour $\alpha = 0,1$ (et donc $\frac{\alpha}{2} = 0,05$), on obtient

- (i) l'intervalle de dispersion symétrique $I_1 = [\Phi^{-1}(0,05); \Phi^{-1}(0,95)]$. La table numérique donne $\Phi(1,64) \approx 0,9495$ et $\Phi(1,65) \approx 0,9505$. Par interpolation linéaire, et symétrie, nous prendrons $I_1 \approx [-1,645; 1,645]$.
- (ii) l'intervalle de dispersion unilatéral inférieur $I_2 = [-\infty; \Phi^{-1}(0,90)]$. La table numérique donne $\Phi(1,28) \approx 0,8997$ et $\Phi(1,29) \approx 0,9015$. Par interpolation linéaire, nous prendrons $I_2 \approx [-\infty; 1,287]$.
- (iii) l'intervalle de dispersion unilatéral supérieur $I_3 = [\Phi^{-1}(0,10); +\infty]$. Par un calcul symétrique du précédent, nous prendrons $I_3 \approx [-1,286; +\infty[$.

Ces résultats sont illustrés aux figures par les trois figures suivantes :

Fig. 3.a

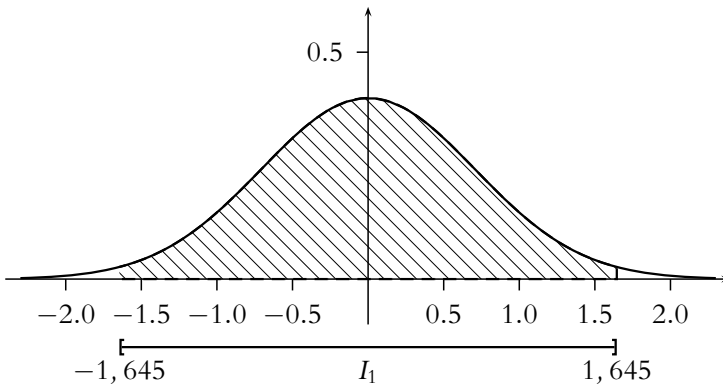


Fig. 3.b

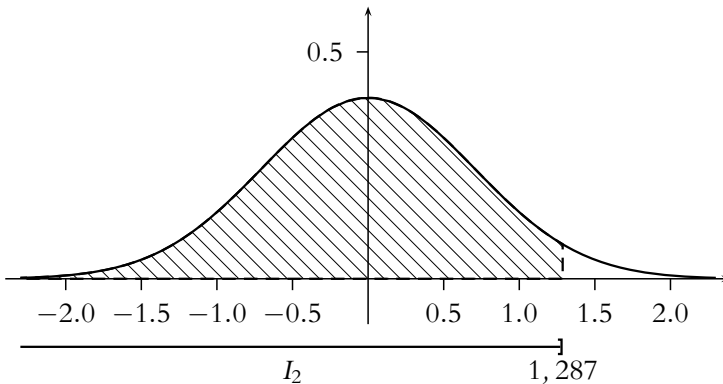
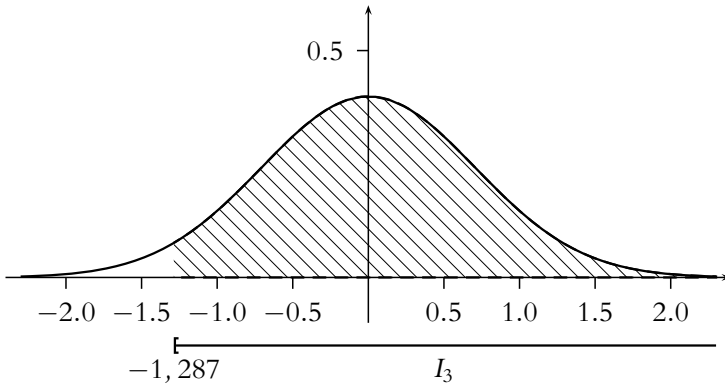


Fig. 3.c



4.3 Construction d'intervalles de confiance

Supposons que l'on cherche à estimer un paramètre θ . On commence par construire un estimateur T_n de θ (on choisira bien sûr le meilleur estimateur possible). Pour toute valeur de θ élément de Θ , T_n est une variable aléatoire dépendant de θ . On fait alors un choix (le même pour toutes les valeurs de θ) parmi les intervalles de dispersion de T_n (intervalle symétrique (le plus souvent), unilatéral, ...)

Une fois ce choix effectué, on détermine l'intervalle de dispersion de T_n au niveau $1 - \alpha$.

Bien entendu, les bornes de cet intervalle de dispersion dépendent *a priori* de θ . On obtient alors deux fonctions a et b de θ telles que $P(T_n \in [a(\theta), b(\theta)]) = 1 - \alpha$.

Le problème revient alors à déterminer un intervalle de confiance pour θ au risque α à partir de ces données. Nous décrivons ci-dessous des situations simples et qui se présentent fréquemment.

- Dans la pratique, lorsque μ_θ est une loi continue, les fonctions a et b sont continues strictement croissantes, donc bijectives. La condition $a(\theta) \leq T_n$ équivaut alors à $\theta \leq a^{-1}(T_n)$, et la condition $T_n \leq b(\theta)$ équivaut à $b^{-1}(T_n) \leq \theta$. Il en résulte que la condition $P(T_n \in [a(\theta), b(\theta)]) = 1 - \alpha$ équivaut à la condition $P(\theta \in [b^{-1}(\theta), a^{-1}(\theta)]) = 1 - \alpha$, et donc, dans ce cas, l'intervalle $[b^{-1}(\theta), a^{-1}(\theta)]$ est un intervalle de confiance pour θ au risque α .
- Parfois, on peut trouver une fonction bijective φ telle que la variable aléatoire $\varphi(T_n)$ suive une loi indépendante de θ . Dans ce cas, les intervalles de dispersion sont indépendants de θ . Pour α donné, il existe alors des constantes m et n telles que $P(\varphi(T_n) \in [m, n]) = 1 - \alpha$, ce qui équivaut à $P(T_n \in [\varphi^{-1}(m), \varphi^{-1}(n)]) = 1 - \alpha$. Alors, dans ce cas, l'intervalle $[\varphi^{-1}(m), \varphi^{-1}(n)]$ est un intervalle de confiance pour θ au risque α .
- On opère comme ci-dessus lorsqu'il existe une fonction bijective φ telle que la variable aléatoire $\varphi(T_n)$ converge en loi vers une variable aléatoire X qui suit une loi indépendante de θ . On veillera seulement à utiliser des valeurs de n assez grandes pour que l'on puisse considérer que X soit une approximation assez bonne de T_n .

Ce cas est fréquent. On le rencontre par exemple dès que l'on peut appliquer le théorème de la limite centrée (ou plus généralement dès que l'on est en présence d'une situation de convergence vers une loi à densité).

► **Remarques**

- Dans le cas où $\varphi(T_n)$ suit une loi indépendante de θ , la variable aléatoire T_n dépend *a priori* de θ . C'est sa loi qui n'en dépend pas (voir le premier exemple ci-dessous).
- Si $\varphi(T_n)$ suit (resp. converge en loi vers) une loi indépendante de θ , on dit que $\varphi(T_n)$ est une statistique *pivotale* (resp. *asymptotiquement pivotale*) pour le paramètre θ .
- Comme dans le cas des intervalles de dispersion, on peut en général construire plusieurs intervalles de confiance (souvent même une infinité de ces intervalles) pour un même paramètre et un même niveau de confiance.
- Les réalisations des intervalles de confiance sont obtenus de manière approchée. On arrondit alors de manière à être sûr que $P(\theta \in [I_n, S_n]) \geq 1 - \alpha$. On obtient alors un intervalle de confiance à un risque légèrement inférieur à α (ou à un niveau de confiance légèrement supérieur à $1 - \alpha$).

4.4 Exemples

Estimation de l'espérance d'une loi normale d'écart-type connu

Considérons un n -échantillon *iid* $\mathcal{E}_n = (X_1, X_2, \dots, X_n)$ issu d'une loi normale d'espérance m et d'écart type σ_0 supposé connu.

Soit alors \bar{X}_n la moyenne empirique de cet échantillon. Comme \bar{X}_n est combinaison linéaire de variables normales, il suit lui-même une loi normale. De plus on connaît son espérance et sa variance : \bar{X}_n suit la loi normale $\mathcal{N}\left(m, \frac{\sigma_0^2}{n}\right)$. Il en résulte que

$$\varphi(\bar{X}_n) = \frac{\bar{X}_n - m}{\sigma_0/\sqrt{n}} \text{ suit la loi normale centrée réduite } \mathcal{N}(0, 1).$$

Nous sommes donc dans la situation où la variable aléatoire $\varphi(\bar{X}_n)$, qui dépend de m , suit une loi qui ne dépend pas de m . L'intervalle de dispersion symétrique de niveau $1 - \alpha$ de la loi normale centrée réduite est

$$I_\alpha = \left[\Phi^{-1}\left(\frac{\alpha}{2}\right), \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right] = \left[1 - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right].$$

Notons $t_\alpha = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$. L'intervalle de dispersion symétrique $[1 - t_\alpha, t_\alpha]$ de niveau $1 - \alpha$ pour la loi normale centrée réduite permet de conclure qu'un intervalle de confiance au niveau de confiance $1 - \alpha$ (ou au risque α) de l'espérance m d'une loi normale X d'écart type connu σ_0 est $\left[\bar{X}_n - \frac{t_\alpha \sigma_0}{\sqrt{n}}, \bar{X}_n + \frac{t_\alpha \sigma_0}{\sqrt{n}} \right]$.

Les valeurs usuelles de α sont 0,1 (niveau de confiance 0,9), 0,05 (niveau de confiance 0,95) et 0,01 (niveau de confiance 0,99). Nous calculons ci-dessous les valeurs correspondantes de t_α .

- Pour $\alpha = 0,1; 1 - \frac{\alpha}{2} = 0,95$, on lit dans la table $\Phi(1,64) \approx 0,9495$ et $\Phi(1,65) \approx 0,9505$, ce qui donne $t_{0,10} = \Phi^{-1}(0,95) \approx 1,645$.

- Pour $\alpha = 0,05; 1 - \frac{\alpha}{2} = 0,975$, on lit dans la table $\Phi(1,96) \approx 0,975$, ce qui donne $t_{0,05} = \Phi^{-1}(0,975) \approx 1,96$.
- Pour $\alpha = 0,01; 1 - \frac{\alpha}{2} = 0,995$, on lit dans la table $\Phi(2,57) \approx 0,9949$ et $\Phi(2,58) \approx 0,9951$, ce qui donne $t_{0,01} = \Phi^{-1}(0,995) \approx 2,575$.

► **Remarque**

La situation ci-dessus est un cas d'école. Il est en effet bien rare que l'on connaisse l'écart type d'une variable sans en connaître l'espérance. Dans la pratique, \bar{X}_n et \bar{S}_n désignant respectivement la moyenne empirique et la variance empirique corrigée, on utilisera la statistique $\sqrt{n} \frac{\bar{X}_n - m}{\bar{S}_n}$ qui suit une *loi de Student* de paramètre $n - 1$, unimodale et symétrique qui ne dépend pas de m . On peut utiliser comme ci-dessus les intervalles de dispersion de cette loi (qui n'est pas au programme de la classe) pour déterminer des intervalles de confiance pour m .

Exemple

Pour $n = 100$, une réalisation de \bar{X}_n égale à 1,56 et une loi normale d'écart type 0,5. On obtient alors les intervalles de confiance suivants

- Pour $\alpha = 0,1 : I_{0,1} = [1,47775; 1,64225]$. Mais, compte tenu du fait que les valeurs obtenues sont des valeurs approchées, il est tout à fait inutile (et sans signification) de conserver autant de chiffres après la virgule. On arrondit de façon à agrandir l'intervalle de confiance, et l'on gardera $I_{0,1} = [1,47; 1,65]$.
- Pour $\alpha = 0,05$: on gardera $I_{0,05} = [1,46; 1,66]$.
- Pour $\alpha = 0,01$: on gardera $I_{0,01} = [1,43; 1,69]$.

Estimation du paramètre d'une variable de Bernoulli

Considérons un n -échantillon *iid* $\mathcal{E}_n = (X_1, X_2, \dots, X_n)$ issu d'une loi de Bernoulli de paramètre p et soit \bar{X}_n la moyenne empirique de cet échantillon. On sait, d'après le théorème de la limite centrée, que la variable aléatoire $\varphi(\bar{X}_n) = \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}}$ converge en loi vers une variable X suivant la loi normale centrée réduite $\mathcal{N}(0, 1)$. On se trouve donc dans le cas où une statistique $\varphi(\bar{X}_n)$ dépendant de p converge en loi vers une variable aléatoire X dont la loi ne dépend pas de p . On peut donc, pour un échantillon assez grand, utiliser l'intervalle de dispersion symétrique $[-t_\alpha, t_\alpha]$ de niveau $1 - \alpha$ de la loi $\mathcal{N}(0, 1)$ (où t_α est défini comme ci-dessus). On a alors

$$P\left(\left[-t_\alpha \leq \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \leq t_\alpha \right] \right) = 1 - \alpha.$$

Mais la fonction φ , bien qu'elle soit bien bijective, ne possède pas une fonction réciproque simple, permettant d'en déduire simplement un intervalle de confiance pour p . Les valeurs de la fonction φ^{-1} peuvent être déterminées au cas par cas à l'aide d'un ordinateur, mais elles ne figurent pas dans des tables numériques, et cela parce que l'on peut utiliser l'une des deux méthodes suivantes, permettant de se passer φ^{-1} .

- L'étude de la fonction $x \mapsto x(1-x)$ montre que, pour tout $p \in [0, 1]$, on a $\sqrt{p(1-p)} \leq \frac{1}{2}$.

Partant alors de l'encadrement ci-dessus, on obtient

$$P\left(\left[\bar{X}_n - t_\alpha \sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{X}_n + t_\alpha \sqrt{\frac{p(1-p)}{n}}\right]\right) = 1 - \alpha.$$

Or, comme $\sqrt{p(1-p)} \leq \frac{1}{2}$, on obtient

$$\left[\bar{X}_n - t_\alpha \sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{X}_n + t_\alpha \sqrt{\frac{p(1-p)}{n}}\right] \subset \left[\bar{X}_n - t_\alpha \frac{1}{2\sqrt{n}} \leq p \leq \bar{X}_n + t_\alpha \frac{1}{2\sqrt{n}}\right].$$

L'intervalle $I = \left[\bar{X}_n - t_\alpha \frac{1}{2\sqrt{n}}, \bar{X}_n + t_\alpha \frac{1}{2\sqrt{n}}\right]$ ainsi obtenu est un intervalle d'estimation de p à un niveau de confiance supérieur à $1 - \alpha$ (ou à un risque inférieur à α).

- On peut aussi remarquer que, d'après la loi des grands nombres, \bar{X}_n converge en probabilité vers la variable certaine égale à p . Comme la convergence en probabilité est conservée par une fonction continue, il en résulte que $\sqrt{\bar{X}_n(1-\bar{X}_n)}$ converge en probabilité vers $\sqrt{p(1-p)}$ et, compte tenu des propriétés des convergences en probabilité et en loi, on conclut que $\sqrt{n} \frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}$ converge en loi vers une variable X suivant la loi normale centrée réduite $\mathcal{N}(0, 1)$.

Il en résulte directement que l'intervalle

$$\left[\bar{X}_n - t_\alpha \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + t_\alpha \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\right]$$

est un intervalle de confiance pour p au risque α . On prouve même qu'il s'agit d'un intervalle de confiance minimal pour le niveau de confiance $1 - \alpha$.

Exemples

1. Pour $n = 1000$, une réalisation de \bar{X}_n égale à 0,54, et une loi de Bernoulli de paramètre p , on obtient, après arrondi, les intervalles de confiance suivants pour p
 - Pour $\alpha = 0,1$: $I_{0,1} = [0,513; 0,567]$ par la première méthode (confiance supérieure à 0,9), et $I'_{0,1} = [0,514; 0,566]$ par la seconde méthode.
 - Pour $\alpha = 0,05$: $I_{0,05} = [0,509; 0,571]$ par la première méthode (confiance supérieure à 0,95), et $I'_{0,05} = [0,509; 0,571]$ par la seconde méthode.
 - Pour $\alpha = 0,01$: $I_{0,01} = [0,499; 0,581]$ par la première méthode (confiance supérieure à 0,99), et $I'_{0,01} = [0,499; 0,581]$ par la seconde méthode.
2. Pour $n = 400$, une réalisation de \bar{X}_n égale à 0,83, et une loi de Bernoulli de paramètre p , on obtient, après arrondi, les intervalles de confiance suivants pour p
 - Pour $\alpha = 0,1$: $I_{0,1} = [0,788; 0,872]$ par la première méthode (confiance supérieure à 0,9), et $I'_{0,1} = [0,799; 0,861]$ par la seconde méthode.
 - Pour $\alpha = 0,05$: $I_{0,05} = [0,781; 0,879]$ par la première méthode (confiance supérieure à 0,95), et $I'_{0,05} = [0,793; 0,867]$ par la seconde méthode.

- Pour $\alpha = 0,01$: $I_{0,01} = [0,765; 0,895]$ par la première méthode (confiance supérieure à 0,99), et $I'_{0,01} = [0,781; 0,879]$ par la seconde méthode.

► **Remarques**

- Insistons sur le fait que ces estimations par intervalles ne donnent des résultats corrects que pour de grands échantillons.
- Le fait que les deux méthodes donnent des résultats très voisins pour des valeurs de p voisines de 0,5 était prévisible. Par contre, pour des valeurs de \bar{X}_n voisines de 1, ou de 0, les résultats deviennent notablement différents. La première méthode fournit même parfois des résultats aberrants (par exemple, pour $\bar{X}_n = 0,94$, cette méthode fournit pour $n = 400$, au risque 0,01, l'intervalle de confiance $[0,875; 1,005]$, ce qui n'est pas raisonnable...)

4.5 Perspectives

Ce qui précède n'est qu'une brève introduction au vaste chapitre de l'inférence statistique. Il existe beaucoup de manières de construire des estimateurs (méthode des moments, des moindres carrés, du maximum de vraisemblance...), et des méthodes permettant de juger de la qualité d'un estimateur (inégalités de Fréchet, Darmois, Cramer et Rao...). Il existe également de nombreuses lois de probabilité permettant de construire des variables aléatoires pivotales. Par exemple, les lois de Student permettent de construire une variable aléatoire pivotale pour l'espérance d'une loi normale dont on ne connaît pas la variance, les lois du χ^2 permettent de construire une variable pivotale pour la variance d'une loi normale dont on ne connaît pas l'espérance...

L'étude approfondie du présent chapitre devrait permettre d'aborder ce domaine d'étude dans de bonnes conditions dès l'entrée dans les écoles supérieures de commerce.

5. Statistiques bivariées

Lorsque l'on effectue un relevé de données statistiques dans une certaine population, il est rare que l'on se contente de mesurer un seul caractère. La plupart du temps, on relève beaucoup de données. La question se pose alors des liens qui peuvent exister entre les différents caractères ainsi mesurés.

Nous nous contenterons, comme première approche du phénomène, d'étudier le cas où l'on mesure deux caractères, et nous chercherons à estimer dans quelle mesure les deux caractères ainsi mesurés sont liés par une relation affine (de la forme $y = ax + b$), ce qui serait le cas par exemple si l'on mesure à la fois la taille (variable x) et le poids (variable y) des individus d'une population humaine, ou bien les notes en maths et en physique des élèves d'une classe de seconde.

5.1 Présentation des données

S'agissant des données relatives à des statistiques univariées (c'est à dire mesurant un seul caractère), nous renvoyons le lecteur au premier tome du présent ouvrage. En particulier, lorsque les observations sont regroupées en classes, on prendra garde à ce

que toutes les classes soient de même amplitude, et on représentera chaque classe par sa valeur médiane.

Dans le cas où l'on mesure deux caractères quantitatifs, les données consistent en un n -échantillon *iid* $\left((x_i, y_i)_{i \in \llbracket 1, n \rrbracket} \right)$ du vecteur aléatoire parent (X, Y) .

On peut aussi dire que ces données sont résumées par la donnée du couple $(\mathfrak{X}, \mathfrak{Y})$

$$\text{constitué par les vecteurs de } \mathbb{R}^n \text{ définis par } \mathfrak{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} \text{ et } \mathfrak{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}.$$

Pour présenter ces données, on peut regrouper les classes, essentiellement pour en réduire le nombre, et présenter ces données dans un tableau d'effectifs (figure 4.a. où $n_{i,j}$ représente l'effectif de la classe (x_i, y_j)) ou, en divisant par l'effectif total, un tableau de fréquences (figure 4.b.). Cette présentation a l'avantage de faire apparaître les *statistiques marginales* et les *statistiques conditionnelles*, ces mots ayant la même signification que dans le chapitre concernant les couples de variables aléatoires discrètes. Nous n'y reviendrons pas.

	γ_1	γ_2	\cdots	γ_j	\cdots	γ_n	
x_1	$n_{1,1}$	$n_{1,2}$	\cdots	$n_{1,j}$	\cdots	$n_{1,n}$	$n_{1,\cdot}$
x_2	$n_{2,1}$	$n_{2,2}$	\cdots	$n_{2,j}$	\cdots	$n_{2,n}$	$n_{2,\cdot}$
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
x_i	$n_{i,1}$	$n_{i,2}$	\cdots	$n_{i,j}$	\cdots	$n_{i,n}$	$n_{i,\cdot}$
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
x_n	$n_{n,1}$	$n_{n,2}$	\cdots	$n_{n,j}$	\cdots	$n_{n,n}$	$n_{n,\cdot}$
	$n_{\cdot,1}$	$n_{\cdot,2}$	\cdots	$n_{\cdot,j}$	\cdots	$n_{\cdot,n}$	n
	γ_1	γ_2	\cdots	γ_j	\cdots	γ_n	
x_1	$f_{1,1}$	$f_{1,2}$	\cdots	$f_{1,j}$	\cdots	$f_{1,n}$	$f_{1,\cdot}$
x_2	$f_{2,1}$	$f_{2,2}$	\cdots	$f_{2,j}$	\cdots	$f_{2,n}$	$f_{2,\cdot}$
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
x_i	$f_{i,1}$	$f_{i,2}$	\cdots	$f_{i,j}$	\cdots	$f_{i,n}$	$f_{i,\cdot}$
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
x_n	$f_{n,1}$	$f_{n,2}$	\cdots	$f_{n,j}$	\cdots	$f_{n,n}$	$f_{n,\cdot}$
	$f_{\cdot,1}$	$f_{\cdot,2}$	\cdots	$f_{\cdot,j}$	\cdots	$f_{\cdot,n}$	1

Dans la pratique, et si l'on ne s'intéresse ni aux statistiques marginales, ni aux statistiques conditionnelles, mais au(x) lien(s) éventuels qui peuvent exister entre les caractères étudiés, les ordinateurs et les logiciels modernes disposent d'assez de puissance

pour calculer directement avec les données brutes, sans traitement préalable. On présente les vecteurs \mathfrak{X} et \mathfrak{Y} sous forme de tableau, et l'on représente graphiquement les points (x_i, y_i) dans un repère cartésien. On obtient ainsi le **nuage de points** associé à la série statistique double étudiée.

Définition 16

Soit une série statistique double constituée d'une suite de n couples (x_i, y_i) . Le **nuage de points** associé à cette série statistique double est l'ensemble des points de coordonnées (x_i, y_i) dans un repère cartésien.

Si l'on note \bar{x} la moyenne des x_i et \bar{y} la moyenne des y_i , le point de coordonnées (\bar{x}, \bar{y}) est le **point moyen du nuage**.

Exemples

1. Le nuage de points associé à la série statistique double présentée dans le tableau suivant

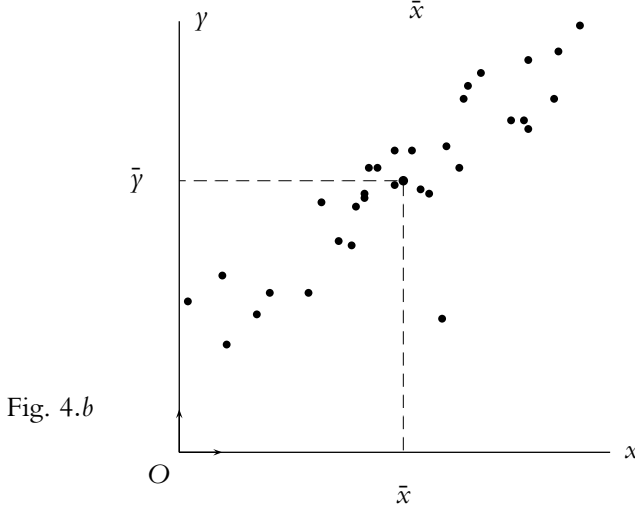
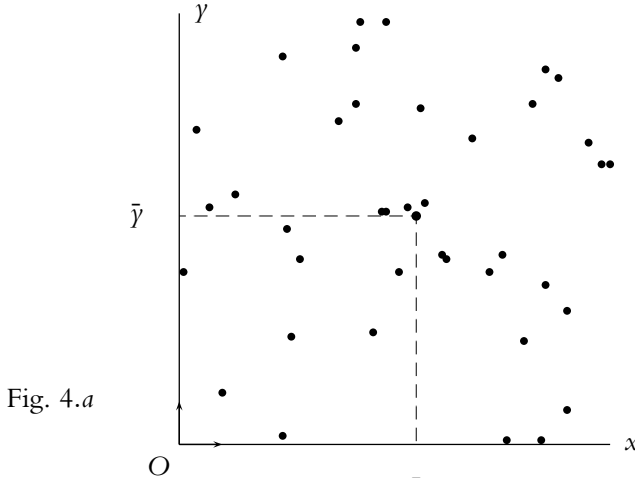
i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	4.08	4.53	3.05	2.43	9.99	2.48	3.68	5.07	8.47	1.02	7.61	0.08
y_i	7.87	2.56	0.23	0.23	6.49	4.98	7.50	4.03	3.66	1.16	0.06	3.96
i	13	14	15	16	17	18	19	20	21	22	23	24
x_i	8.03	1.25	2.40	4.75	6.22	9.00	5.26	5.57	6.82	4.08	4.53	8.96
y_i	2.37	5.78	9.02	5.36	4.29	0.84	5.53	7.79	7.05	7.87	2.56	3.05
i	25	26	27	28	29	30	31	32	33	34	35	36
x_i	2.43	9.99	2.48	3.68	5.07	8.47	1.02	7.61	0.08	8.03	2.82	9.53
y_i	0.23	6.49	4.98	7.50	4.03	3.66	1.16	0.06	3.96	2.37	4.33	6.99
i	37	38	39	40								
x_i	8.79	7.58	2.61	4.82								
y_i	6.53	4.35	2.54	9.75								

est présenté à la figure 4.a. Le point moyen, correspondant à $\bar{x} \approx 5,49$ et $\bar{y} \approx 5,33$, est également représenté sur cette figure.

2. Le nuage de points associé à la série statistique double présentée dans le tableau suivant

i	1	2	3	4	5	6	7	8
x_i	6.5	8.1	9.3	5.8	1.8	3.7	8.8	6.7
y_i	6.6	7.5	9.9	6	3.2	4.9	9.3	8.5
i	9	10	11	12	13	14	15	16
x_i	3	5	5.6	2.1	1.1	3.3	0.2	6.6
y_i	3.7	7	6.1	3.7	2.5	5.8	3.5	8.2
i	17	18	19	20	21	22	23	24
x_i	6.1	8.1	5	7.7	8.7	4.3	8	6.2
y_i	3.1	9.1	6.2	7.7	8.2	5.9	7.7	7.1
i	25	26	27	28	29	30	31	32
x_i	4	4.4	5.4	4.3	4.6	1	4.1	7
y_i	4.8	6.6	7.2	6	6.6	4.1	5.7	8.8

est présenté à la figure 4.b. Le point moyen, correspondant à $\bar{x} \approx 5,2$ et $\bar{y} \approx 6,3$, est également représenté sur cette figure.



Avec ces données, on peut naturellement calculer la *variance empirique* $s_x^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$ de l'échantillon $(x_i)_{i \in \llbracket 1, n \rrbracket}$ de la première loi marginale, la *variance empirique* $s_y^2 = \frac{1}{n} \sum_{k=1}^n (y_k - \bar{y})^2$ de l'échantillon $(y_i)_{i \in \llbracket 1, n \rrbracket}$ de la seconde loi marginale, ainsi que la *covariance empirique* $\text{Cov}(x, y) = \frac{1}{n} \sum_{k=1}^n x_k y_k - \bar{x} \bar{y}$ de l'échantillon vectoriel $((x_i, y_i)_{i \in \llbracket 1, n \rrbracket})$ de la loi conjointe.

Si les données sont regroupées en classes, des calculs analogues conduisent à ces résultats en utilisant les séries des fréquences.

Les propriétés de ces moments empiriques sont à rapprocher des propriétés des moments des couples de variables aléatoires discrètes. Nous n'y reviendrons pas.

5.2 Corrélation et droites des moindres carrés

En examinant les nuages de points des figures 4.a. et 4.b., on remarque que, sur la figure 4.b., le nuage s'allonge selon une direction privilégiée, ce qui ne semble pas être le cas pour le nuage de la figure 4.a.

D'où l'idée de rechercher s'il existe une droite ajustant « au mieux » le nuage de points associé à une série statistique double donnée.

Il existe plusieurs méthodes pour déterminer une telle droite.

La plus connue est la méthode dite **des moindres carrés**, qui consiste à chercher une droite par son équation $y = \varphi_{m_x, p}(x) = m_x x + p$ dans le repère du nuage, de telle

manière que $F(m_x, p) = \sum_{k=1}^n (y_k - \varphi_{m_x, p}(x_k))^2 = \sum_{k=1}^n (y_k - m_x x_k - p)^2$ soit minimum.

Il doit être clair dès à présent qu'une telle droite étant définie à l'aide d'une réalisation d'un échantillon *iid* d'un vecteur aléatoire (X, Y) , sera une estimation de la véritable droite de régression de Y en X telle que nous l'avons rencontrée au chapitre 9.

Définition 17

On considère une série statistique double constituée d'une suite de n couples (x_i, y_i) . On appelle droite des moindres carrés de y en x , ou droite de régression de y en x de la série statistique, ou du nuage associé, la droite d'équation $y = \varphi_{m_x, p}(x) = m_x x + p$ dans le repère du nuage, de telle manière que $F(m_x, p) = \sum_{k=1}^n (y_k - \varphi_{m_x, p}(x_k))^2 = \sum_{k=1}^n (y_k - m_x x_k - p)^2$ soit minimum.

► Remarque

On définirait de façon analogue la droite de régression de x en y comme la droite d'équation

$$x = \psi_{m_y, q}(y) = m_y y + q$$

dans le repère du nuage, de telle manière que $F(m_y, q) = \sum_{k=1}^n (x_k - \psi_{m_y, q}(y_k))^2 = \sum_{k=1}^n (x_k - m_y y_k - q)^2$ soit minimum.

Au chapitre 9, nous avons déterminé par son équation, la droite de régression de Y en X (et celle de X en Y). À partir de réalisations d'un échantillon du vecteur aléatoire (X, Y) , on obtiendra une estimation de cette droite de régression.

Pour déterminer une droite de régression d'un nuage, on peut envisager plusieurs méthodes.

- Une première méthode consiste à chercher le minimum éventuel de la fonction F , en utilisant les règles de détermination des extremums des fonctions de plusieurs variables. On trouve que ce minimum existe et est atteint pour $\frac{\partial F}{\partial m_x}(m_x, p) = \frac{\partial F}{\partial p}(m_x, p) = 0$, ce qui conduit au système d'équations

$$\begin{cases} \sum_{k=1}^n (y_k - m_x x_k - p) = \bar{y} - m_x \bar{x} - p = 0 \\ \sum_{k=1}^n x_k (y_k - m_x x_k - p) = 0. \end{cases}$$

On en déduit que la droite cherchée passe par le point moyen du nuage et admet

pour coefficient directeur $m_x = \frac{\sum_{k=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{k=1}^n (x_i - \bar{x})^2}$.

- Une seconde méthode consiste à considérer que m_x et p sont, s'il en existe, solution du système

$$\begin{cases} \gamma_1 = m_x x_1 + p \\ \gamma_2 = m_x x_2 + p \\ \dots & \dots \\ \gamma_i = m_x x_i + p \\ \dots & \dots \\ \gamma_n = m_x x_n + p \end{cases}$$

Posons $\mathfrak{U} = \begin{pmatrix} 1 \\ 1 \\ \cdot \\ 1 \\ \cdot \\ 1 \end{pmatrix}$.

Le système ci-dessus s'écrit sous forme vectorielle $\mathfrak{Y} = m_x \mathfrak{X} + p \mathfrak{U}$ ou sous forme

matricielle $\mathbf{A} \begin{pmatrix} m_x \\ p \end{pmatrix} = \mathfrak{Y}$ avec $\mathbf{A} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \cdot \\ x_i & 1 \\ \dots & \cdot \\ x_n & 1 \end{pmatrix}$.

Ce système admet une solution si, et seulement si, \mathfrak{Y} appartient au sous-espace vectoriel engendré par \mathfrak{X} et \mathfrak{U} .

À défaut de pouvoir annuler le vecteur $\mathbf{A} \begin{pmatrix} m_x \\ p \end{pmatrix} - \mathfrak{Y}$, on choisit de minimiser sa norme. On sait alors que cette norme est minimum si, et seulement si, $\mathbf{A} \begin{pmatrix} m_x \\ p \end{pmatrix}$ est le projeté orthogonal de \mathfrak{Y} sur l'espace vectoriel engendré par \mathfrak{X} et \mathfrak{U} . Comme les vecteurs \mathfrak{X} et \mathfrak{U} sont précisément les vecteurs colonnes de la matrice \mathbf{A} , cette orthogonalité se traduit par ${}^t\mathbf{A}\mathbf{A} \begin{pmatrix} m_x \\ p \end{pmatrix} = {}^t\mathbf{A}\mathfrak{Y}$, ce qui donne un système à deux équations et deux inconnues qui admet la solution

$$(m_x, p) = \left(\frac{\sum_{k=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{k=1}^n (x_i - \bar{x})^2}, \bar{y} - m_x \bar{x} \right).$$

- On peut aussi envisager d'utiliser les moments empiriques pour estimer les coefficients des droites de régression tels qu'ils ont été déterminés au chapitre 9.

En utilisant ces moments empiriques, et en posant $r_{x,y} = \frac{\text{Cov}(x, y)}{s_x s_y}$ (coefficient de corrélation empirique), on obtient $m_x = \frac{\text{Cov}(x, y)}{s_x^2} = r_{x,y} \frac{s_y}{s_x}$, ce qui correspond tout à fait aux résultats obtenus par les deux premières méthodes.

Théorème 9

Soit une série statistique double constituée d'une suite de n couples (x_i, y_i) . On note \bar{x} , \bar{y} , s_x et s_y les moyennes respectives et les écarts types respectifs des distributions marginales de la série statistique donnée, et l'on note $r_{x,y}$ le coefficient de corrélation de ces distributions marginales.

La droite de régression de y en x de cette série statistique double est la droite dont l'équation dans le repère du nuage est $y = \bar{y} + r_{x,y} \frac{s_y}{s_x} (x - \bar{x})$.

On obtient une équation de la droite de régression de x en y en intervertissant le rôle des lettres x et y .

► Remarques

- Les deux droites de régression passent par le point moyen du nuage, et $r_{x,y}$ est le cosinus de l'angle qu'ils forment.
- La notation \mathbf{A}' désigne la matrice transposée de \mathbf{A} , qui est aussi souvent notée \mathbf{A}^t . On rencontre souvent aussi la notation \mathbf{A}' . Cette notation de la transposée est très répandue dans les ouvrages de statistique.
- Il est important de constater que les droites de régression ainsi obtenues dépendent d'une réalisation d'un échantillon d'un vecteur aléatoire parent. Les résultats obtenus ne permettent pas de tirer quelque conclusion que ce soit concernant les variables aléatoires X et Y . Ainsi, les quatre nuages de points de l'exemple 2 ont été obtenus (par simulation sur **Excel**) à partir du même vecteur aléatoire parent (avec X et Y indépendants). Le fait que les coefficients de corrélation empiriques obtenus varient entre $-0,42$ et $0,39$ provient du fait que $n = 40$ est un effectif trop faible pour espérer une approximation fiable.

Exemple 1

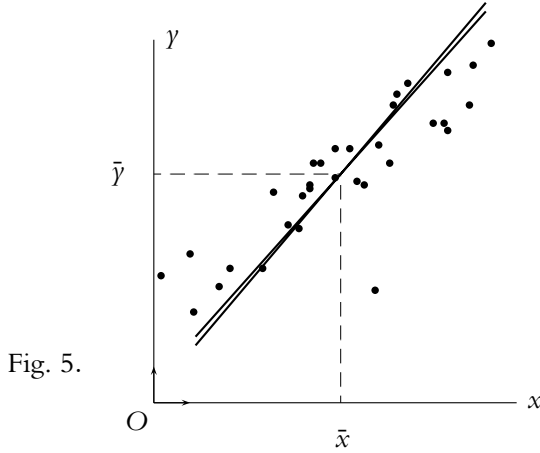


Fig. 5.

On a $\bar{x} \approx 5,2$, $\bar{y} \approx 6,3$, $r_{x,y} \approx 0,87$. Les coefficients des droites de régression sont approximativement 1,12 et 1,18.

Exemple 2

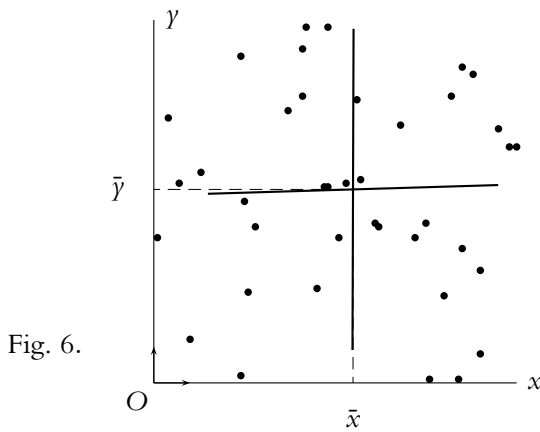
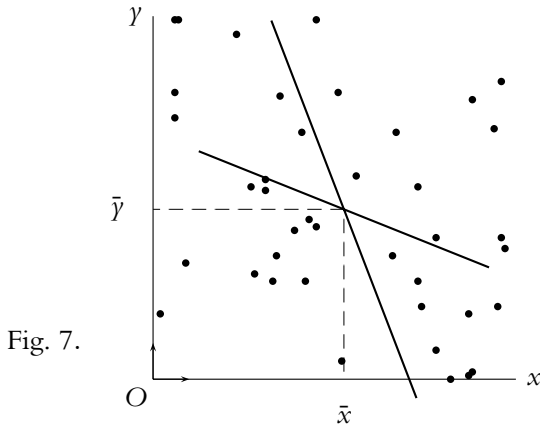
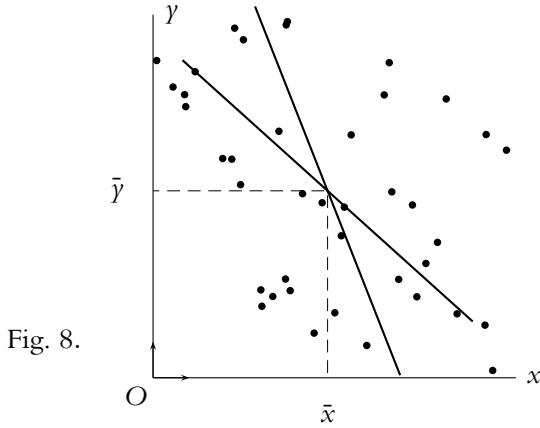


Fig. 6.

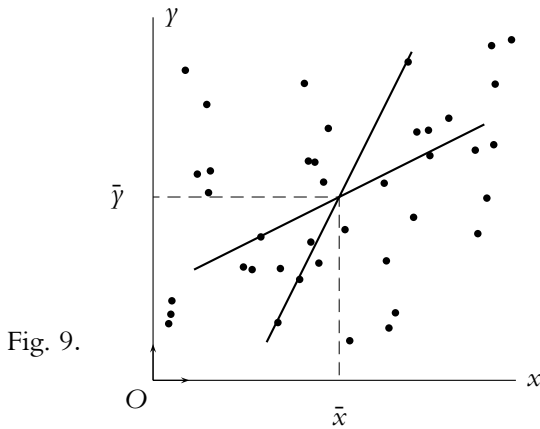
On a $\bar{x} \approx 5,49$, $\bar{y} \approx 5,33$, $r_{x,y} \approx 0,03$. Les coefficients des droites de régression sont approximativement 0,03 et 295.



On a $\bar{x} \approx 5,26$, $\bar{y} \approx 4,68$, $r_{x,y} \approx -0,38$. Les coefficients des droites de régression sont approximativement $-0,4$ et $-2,6$.



On a $\bar{x} \approx 4,81$, $\bar{y} \approx 5,15$, $r_{x,y} \approx -0,42$. Les coefficients des droites de régression sont approximativement $-0,9$ et $-2,54$.



On a $\bar{x} \approx 5,13$, $\bar{y} \approx 5,05$, $r_{x,y} \approx 0,39$. Les coefficients des droites de régression sont approximativement 0,5 et 2.

On aura bien compris que, si $\mathcal{E} = ((x_i, y_i)_{i \in \llbracket 1, n \rrbracket})$ est un échantillon du couple (X, Y) de variables aléatoires, les coefficients des droites de régression de l'échantillon sont des statistiques sur cet échantillon.

On dispose alors du théorème suivant, que nous admettrons.

Théorème 10

Les coefficients des droites de régression d'un échantillon *iid* d'un vecteur aléatoire (X, Y) sont des estimateurs sans biais des coefficients des droites de régression de X et Y .

► **Remarques**

- On utilise souvent les droites de régression obtenues à partir d'un échantillon pour estimer une valeur prise par Y connaissant la valeur correspondante prise par X . Il est clair qu'un tel calcul donne une estimation de la valeur cherchée, mais, en l'absence de calcul, et de raisonnements supplémentaires, rien ne permet de dire dans quelle mesure on peut faire confiance à une telle estimation.
- Il existe d'autres méthodes pour ajuster un nuage de points par une droite. On utilise par exemple la droite de Mayer. Pour construire une droite de Mayer, on partage le nuage en deux sous-nuages de même effectif (à peu près), et l'on joint les points moyens de ces deux sous-nuages. Une telle droite est facile à construire graphiquement, mais elle dépend du choix des deux sous-nuages. De plus, on peut la construire dans tous les cas. On s'abstiendra d'une telle construction dans les cas où le nuage de points ne serait pas visiblement allongé dans une direction privilégiée.
Pour théoriser l'utilisation des droites de Mayer, on pourrait envisager de les tracer toutes et de considérer l'enveloppe de la famille de droites ainsi construites.
- L'utilisation de la droite de régression de $\ln Y$ en X permet de réaliser un ajustement exponentiel d'un nuage de points.
- L'utilisation de la droite de régression de $\ln Y$ en $\ln X$ permet de réaliser un ajustement d'un nuage de points par une fonction puissance.

1. Soient c un réel strictement positif et U une variable aléatoire suivant la loi uniforme sur $]0, 1[$.

1. Déterminer la loi de la variable aléatoire X définie par $X = -\frac{1}{c} \ln(U)$.
2. Utiliser ce qui précède pour simuler une variable aléatoire suivant la loi exponentielle de paramètre c au moyen d'un programme Pascal.

2. Pseudo-inverse

Soit X une variable aléatoire dont la fonction de répartition est notée F_X . On considère la fonction F_X^* définie par $F_X^*(t) = \inf\{x \mid F_X(x) \geq t\}$ (F_X^* est appelée *pseudo-inverse* de F_X).

On utilisera les conventions suivantes : $\inf \mathbb{R} = -\infty$, $\inf \emptyset = +\infty$, $F_X(-\infty) = 0$ et $F_X(+\infty) = 1$.

1. a) Que peut-on dire de $F_X^*(t)$ pour $t < 0$? Même question pour $t > 1$.
 b) Déterminer F_X^* lorsque X est une variable uniforme discrète sur $\{0, 1, 2, 3\}$. Représenter graphiquement F_X et F_X^* sur un même graphique.
2. Montrer que
 - a) Pour tout réel x et tout réel $t \in]0, 1[$, alors

$$F_X^*(F_X(x)) \leq x, \quad F_X(F_X^*(t)) \geq t,$$

$$\{x \mid F_X(x) \geq t\} = [F_X^*(t), +\infty[, \quad F_X(x) \geq t \iff X \geq F_X^*(t).$$

- b) Si F_X est strictement croissante, alors $F_X^*(F_X(x)) = x$ pour tout réel x . Pouvez-vous donner un contre-exemple?
 - c) Si F_X est continue, alors $F_X(F_X^*(t)) = t$ pour tout t de l'intervalle $]0, 1[$. Pouvez-vous donner un contre-exemple?
3. Montrer que, si U suit la loi uniforme sur $]0, 1[$, alors $F_X^*(U)$ admet F_X comme fonction de répartition.
4. Montrer que, si F_X est continue et strictement croissante, alors
 - a) F_X est bijective de \mathbb{R} sur $]0, 1[$ et $F_X^* = F_X^{-1}$ (ici F_X^* est l'inverse de F_X);
 - b) $F_X(X)$ suit la loi uniforme sur $]0, 1[$.
5. Utiliser ce qui précède pour simuler une variable aléatoire exponentielle de paramètre λ au moyen d'un programme Pascal.

3. Démontrer que, si l'erreur quadratique moyenne (risque quadratique) d'un estimateur tend vers 0, alors cet estimateur est convergent.

Que penser de la réciproque de ce théorème?

4. (Questions confidentielles) Pour préserver la confidentialité des opinions individuelles, un sondage est effectué avec le protocole qui suit. Chaque personne sondée doit, avant de répondre « oui » ou « non » à chacune des questions, réaliser confidentiellement (elle seule connaît le résultat) une variable de Bernoulli de paramètre α ($\alpha \in]0, 1[$). Si le résultat est 1, la personne doit répondre à toutes les questions selon ses convictions, et si le résultat est 0, elle doit répondre à toutes les questions au contraire de ses opinions. On suppose que les

personnes sondées jouent parfaitement le jeu, et l'on note X_i la variable de Bernoulli égale à 1 si la i -ème personne sondée répond « oui », et 0 dans le cas contraire.

Soient n le nombre des personnes sondées et p la fréquence des opinions « oui » dans la population.

L'institut de sondage recueille une proportion q de « oui ».

1. a) Calculer q en fonction de p et de α .
 b) Donner un estimateur sans biais S_n de q . Cet estimateur est-il convergent ?
 c) Dédire de ce qui précède un estimateur T_n de p . Pourquoi peut-on conclure qu'il est convergent ?
2. a) Calculer l'espérance de T_n . Cet estimateur est-il sans biais ?
 b) Calculer la variance de T_n . Retrouver ainsi le fait que T_n est un estimateur convergent de p .
 c) Quelles contraintes contradictoires pèsent-elles sur le choix de α ? Comment choisiriez vous ce paramètre α ?
3. a) Pour $\alpha = \frac{1}{6}$ (jet d'un dé), $n = 1000$ et une réalisation $S_n = 425$, donner un intervalle de confiance de coefficient 0,95 pour p .
 b) Écrire un programme Pascal simulant le processus décrit ci-dessus pour $\alpha = \frac{1}{6}$ (jet d'un dé), $n = 1000$ et $q = 0,425$, et donnant 50 intervalles de confiance à 95 %.

5. (Estimation d'une population par capture-recapture) On considère une urne contenant un nombre inconnu N de boules blanches indiscernables. On cherche à estimer la valeur de N à l'aide d'un processus qui consiste à prendre un nombre connu m des boules contenues dans l'urne, à les marquer de façon à les rendre discernables des $N - m$ autres boules restées dans l'urne, puis à les remettre dans l'urne (qui contient alors $N - m$ boules blanches et m boules marquées).

1. a) On tire alors les boules une à une de l'urne avec remise. On note X la variable aléatoire égale au nombre de tirages nécessaires (et suffisants) à l'obtention de la première boule marquée.
 Quelle est la loi de X , son espérance et sa variance ?
 On considère un échantillon (X_1, X_2, \dots, X_n) de variables indépendantes suivant toutes la même loi que X . Construire un estimateur de N . Est-il sans biais, est-il convergent ?
 Quel est l'inconvénient de l'estimateur trouvé ci-dessus ?
 b) Pour éviter cet inconvénient, on considère un entier naturel non nul b , et l'on décide de tirer au maximum b boules. On définit ainsi une variable aléatoire Y par $Y(\Omega) = \llbracket 0, n \rrbracket$ telle que $\forall k \in \llbracket 1, b \rrbracket, P(Y = k) = P(X = k)$ et $P(Y = 0) = P(X > b)$.
 Calculer l'espérance de Y .
 Construire à l'aide de Y un estimateur de N . Montrer qu'il est convergent et asymptotiquement sans biais.
2. À chacun des tirages effectués dans l'urne on associe la variable aléatoire Z qui vaut 1 si la boule tirée est marquée, et 0 sinon.
 a) Quelle est la nature de la variable aléatoire Z ? Donner son espérance et sa variance.
 b) On considère un échantillon (Z_1, Z_2, \dots, Z_n) de variables indépendantes suivant toutes la même loi que Z , et l'on considère la statistique $S_n = Z_1 + Z_2 + \dots + Z_n$. Quelle est la loi de S_n ? Montrer que $E(S_n) = \frac{nm}{N}$, et donner $V(S_n)$.

- c) L'expression $T_n = \frac{nm}{S_n}$ définit-elle un estimateur de N ? Pourquoi?
- d) Montrer que $\frac{1}{S_n + 1}$ est une variable aléatoire. Calculer son espérance.
- e) Utiliser ce qui précède pour construire un estimateur asymptotiquement sans biais et convergent de N .

6. (Vraisemblance, cas discret) Soient X_θ une variable aléatoire réelle discrète dépendant d'un paramètre θ et (X_1, X_2, \dots, X_n) un échantillon *iid* de variables indépendantes suivant toutes la même loi que X_θ . Pour toute réalisation (x_1, x_2, \dots, x_n) de cet échantillon, la probabilité $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ dépend bien sûr de (x_1, x_2, \dots, x_n) , mais aussi de θ . On note cette probabilité

$$L(x_1, x_2, \dots, x_n, \theta) = P_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n).$$

La fonction L prend le nom de fonction *vraisemblance* (L pour *Likelihood*) de X_θ .

1. a) Que représente $L(x_1, x_2, \dots, x_n, \theta)$?
b) Utiliser les propriétés de l'échantillon pour exprimer $L(x_1, x_2, \dots, x_n, \theta)$ sous forme d'un produit.
2. Exprimer la fonction de vraisemblance des lois suivantes
 - a) Loi de Bernoulli de paramètre p .
 - b) Pour k donné, loi binomiale $\mathcal{B}(k, p)$ pour le paramètre p .
 - c) Loi géométrique de paramètre p .
3. Lorsque l'on cherche à estimer le paramètre θ à partir d'une réalisation (x_1, x_2, \dots, x_n) de l'échantillon (X_1, X_2, \dots, X_n) , il est naturel d'essayer déterminer la valeur de θ qui rend maximum la probabilité d'observer la réalisation (x_1, x_2, \dots, x_n) que l'on vient justement d'observer.
 - a) La fonction L est une fonction de plusieurs variables, que l'on suppose de classe \mathcal{C}^2 par rapport à θ . À quelles conditions doit-elle satisfaire pour que la fonction partielle associée à θ admette un maximum?
 - b) Montrer que la détermination de la (ou des) valeurs (éventuelles) de θ rendant maximum la fonction L pour un (x_1, x_2, \dots, x_n) donné équivaut à la détermination de la (ou des) valeurs (éventuelles) de θ rendant maximum la fonction $\ln(L)$ pour un (x_1, x_2, \dots, x_n) donné.
4. Une valeur $\hat{\theta}$ (éventuellement) déterminée ci-dessus pour laquelle le maximum de L est atteint dépend de la réalisation (x_1, x_2, \dots, x_n) . On appelle *estimateur du maximum de vraisemblance de θ* l'estimateur obtenu en remplaçant (x_1, x_2, \dots, x_n) par (X_1, X_2, \dots, X_n) dans l'expression de ce maximum.
 - a) Montrer que l'estimateur du maximum de vraisemblance du paramètre p d'une variable de Bernoulli est la fréquence empirique.
 - b) Déterminer l'estimateur du maximum de vraisemblance du paramètre p d'une variable binomiale $\mathcal{B}(k, p)$ dont le paramètre entier k est connu.
 - c) Déterminer l'estimateur du maximum de vraisemblance du paramètre p d'une loi géométrique.

7. Une machine fabrique des pièces en grande série. À chaque pièce on associe une de ses cotes, exprimée en mm. On définit ainsi une variable aléatoire X .

- On mesure la longueur de la pièce, et l'on suppose que X suit la loi normale $\mathcal{N}(\mu, \sigma^2)$, avec $\mu = 150$ et $\sigma = 0,21$. On tire au hasard et avec remise un échantillon de 400 pièces, et l'on note $\bar{\mu}$ la moyenne des longueurs des pièces de cet échantillon. Déterminer le réel positif h tel que $P(|\mu - \bar{\mu}| \leq h) = 0,95$.
- On mesure la largeur de la pièce, et l'on suppose que X suit la loi normale $\mathcal{N}(m, s^2)$, avec $m = 28,2$ et $s = 0,027$. Soit M la moyenne empirique d'un échantillon de n variables aléatoires indépendantes de même loi que X . Déterminer n pour que

$$P(28,195 \leq M \leq 28,205) = 0,95.$$

8. On considère un n -échantillon *iid* issu d'une loi normale $\mathcal{N}(m, \sigma^2)$ où m est inconnu et $\sigma = 200$.

- On pose $n = 36$. Déterminer un intervalle de confiance de m , avec le coefficient 0,90. Même question pour le coefficient 0,95 ; même question pour le coefficient 0,99. Commenter les résultats.
- Reprendre la première question avec $n = 400$.

9. Un relevé de vitesses pour un échantillon de véhicules pris au hasard sur une portion de route à grande circulation a donné les résultats suivants

Vitesse(km/h)	Effectif
]75; 80]	5
]80; 85]	10
]85; 90]	20
]90; 95]	36
]95; 100]	15
]100; 105]	8
]105; 110]	6

- Calculer la moyenne \bar{x} et l'écart type σ de cette distribution empirique.
- Proposer une estimation ponctuelle de la moyenne μ et de l'écart type s de la distribution des vitesses des véhicules empruntant cette voie.
- On suppose que la variable aléatoire qui, à tout échantillon de taille n associe la moyenne des vitesses de l'échantillon, suit la loi normale $\mathcal{N}\left(\mu, \frac{s}{\sqrt{n}}\right)$. Déterminer un intervalle de confiance de μ au coefficient 0,99.
- Pour quelles valeurs de n est-on sûr de connaître un intervalle de confiance de μ au coefficient 0,95, dont l'amplitude soit 0,5 ?

10. Un jour d'élection, un sondage « sortie des urnes » donne 51 % de votes en faveur du candidat A .

- On a interrogé 100 personnes. Donner un intervalle de confiance au coefficient 95 % pour la proportion p de votes en faveur de A .

2. Reprendre cette question sachant qu'en fait on a interrogé 1000 personnes.
3. Combien aurait-il fallu interroger de personnes pour que le résultat soit connu à 4 % près (c'est à dire que la longueur de l'intervalle de confiance de coefficient 0,95 trouvé soit 0,04) ?

11. On effectue des pesées avec une balance. On sait, pour l'avoir testée, que cette balance donne, pour un objet donné, des résultats qui suivent une loi normale dont la moyenne est la masse de l'objet pesé, et dont l'écart type est 1 g.

1. On a effectué 25 mesures d'un certain objet, et la somme des résultats est 30,25 g. Donner un intervalle de confiance à 90 % pour la masse de cet objet.
2. Reprendre ce qui précède pour 400 mesures dont la somme donne le résultat 484 g.

12. Soit (X_1, X_2, \dots, X_n) un n -échantillon *iid* issu d'une loi normale $\mathcal{N}(0, \sigma^2)$.

1. **a)** Pour tout i , calculer l'espérance de la variable aléatoire $|X_i|$.
b) En déduire un estimateur sans biais de σ .
2. **a)** Pour tout i , calculer la variance de la variable aléatoire $|X_i|$.
b) L'estimateur obtenu est-il convergent ?
3. On considère une réalisation d'un échantillon correspondant à $n = 100$, pour lequel la moyenne empirique des valeurs absolues a été calculée à 1,084. En déduire un intervalle de confiance de coefficient 0,95 pour σ .

13. Soit θ un réel strictement positif.

On considère la fonction f_θ définie pour tout réel x par $f_\theta(x) = \frac{2x}{\theta^2} \mathbb{1}_{[0, \theta]}(x)$.

1. **a)** Prouver que, pour tout θ , f_θ est une densité de probabilité.
b) Soit X_θ une variable aléatoire admettant f_θ pour densité. Calculer l'espérance et la variance de X_θ .
2. **a)** Soit (X_1, X_2, \dots, X_n) un n -échantillon *iid* issu de la loi de X_θ . Déduire de la question précédente un estimateur T_n sans biais et convergent de θ .
b) On considère la fonction de vraisemblance L (voir exercice 6) définie par

$$L(x_1, x_2, \dots, x_n, \theta) = \prod_{k=1}^n f_\theta(x_k).$$

Étudier son maximum éventuel.

- c)** En déduire que l'estimateur du maximum de vraisemblance (voir exercice 6) de θ est $M_n = \max(X_1, X_2, \dots, X_n)$.
3. **a)** Déterminer la loi de probabilité de M_n .
b) L'estimateur M_n est-il sans biais ?
c) Déduire de ce qui précède un estimateur sans biais \widehat{M}_n de θ .
d) L'estimateur \widehat{M}_n est-il convergent ?
e) Comparer les estimateurs T_n et \widehat{M}_n . Quel est le meilleur ?

4. En utilisant la loi de M_n , construire un intervalle de confiance de niveau 0,95, pour θ dans le cas où l'on a observé $\max(X_1, X_2, \dots, X_n) = 5$.

14. On considère la série statistique double définie par le tableau suivant

i	1	2	3	4	5	6	7	8
x_i	1	2	3	6	8	9	11	14
y_i	1	2	4	4	5	7	8	9

Représenter le nuage de points correspondant et son point moyen. Déterminer à la main, par une de leurs équations, les droites de régression de ce nuage, et les tracer.

15. En utilisant la commande ALEA() du logiciel EXCEL construire une feuille de calcul donnant deux séries de n valeurs pour des variables aléatoires X et Y indépendantes, les moyennes marginales, le coefficient de corrélation et les coefficients directeurs des droites de régression correspondant à l'échantillon obtenu, et comportant un graphique représentant le nuage de points.

Faire fonctionner plusieurs fois (en notant les résultats), pour $n = 50$, puis $n = 100$, puis $n = 1000$. Conclure.

16. Dans le tableau ci-dessous, on a fait figurer les valeurs expérimentales de la pression P d'une masse de gaz pour différentes valeurs du volume V . On sait que $PV^\gamma = C$, où γ et C sont des constantes dépendant des conditions de l'expérience.

i	1	2	3	4	5	6
v_i (en cm^3)	54,3	61,8	72,4	88,7	116,6	194
p_i (en kg/cm^2)	61,2	49,5	37,6	28,4	19,2	10,1

- Poser $X = \ln V$ et $Y = \ln P$ et déterminer les droites des moindres carrés de Y en X et de X en Y ainsi que le coefficient de corrélation.
- En déduire des estimations de C et γ .
- Estimer P pour $V = 100 \text{ cm}^3$, et V pour $P = 15 \text{ kg}/\text{cm}^2$.
- Quelle confiance peut-on faire à ces estimations? Pourquoi?

17. (Parabole des moindres carrés) On considère la série statistique double définie par le tableau ci-dessous

i	1	2	3	4	5	6	7	8	9	10	11
x_i	-5	-4	-3	-2	-1	0	1	2	3	4	5
y_i	23,2	31,4	39,8	50,2	62,9	76	92	105,7	122,8	131,7	151,1

1. En utilisant les mêmes méthodes que pour la droite des moindres carrés de Y en X , déterminer la parabole des moindres carrés de Y en X , c'est à dire la courbe dont l'équation est

$$y = f(x) = ax^2 + bx + c \text{ de telle manière que } \sum_{k=1}^{11} (y_i - f(x_i))^2 \text{ soit minimum.}$$

2. Représenter le nuage de points et la parabole obtenue sur une même figure.
3. Avec ces résultats, estimer la valeur de Y pour $X = 4, 5$.
4. Déterminer les droites des moindres carrés et les tracer sur le même graphique qu'au 2. Quelle estimation de Y obtient-on alors pour $X = 4, 5$?

Autres exercices

On cherchera également avec profit les exercices proposés aux différents concours. Signalons entre autres EDHEC 2000 ; ESCP oral 1997, 1998, 1999, 2001, 2002, 2003.

Ce chapitre traite de plusieurs interventions assez différentes de l'informatique dans le programme

- l'étude théorique d'une caractéristique intéressante de Pascal, la **récurtivité**, dont on verra tout l'intérêt mais aussi toutes les limites, sur cinq exemples d'efficacités très différentes ;
- en prolongement de la précédente, l'amorce sur trois exemples de l'étude de la très importante **gestion de listes**, avec des versions itératives et récursives selon les cas ;
- treize exemples itératifs de **simulation de variables aléatoires réelles** suivant des lois variées, à l'aide de la fonction RANDOM.
- enfin cinq exemples de techniques d'**estimation**.

1. Récurtivité

La récurtivité, proche mais différente de la récurrence, est partout présente en mathématiques. La façon la plus immédiate de la rencontrer se présente naturellement dans l'étude de suites de la forme

$$u_0 = a, \quad u_{n+1} = f(u_n)$$

dont le prototype est évidemment la factorielle, à laquelle d'ailleurs nous consacrons une partie que l'on pourrait étendre sans problème au cas général.

Presque tous les langages informatiques acceptent la traduction immédiate de telles définitions. Ce ne fut pas toujours le cas ; ainsi le premier logiciel scientifique qui permit de s'extraire des méandres des programmes travaillant directement sur les mémoires de l'ordinateur, le célèbre **Fortran** de John Warner Backus, ignorait cette possibilité (demander à un navigateur `ibm 704 fortran` permet d'en lire le tout premier manuel, datant du 16 octobre 1956). Il en fut de même de plusieurs de ses fils directs, comme le **Basic** de John Kemeny et Thomas Kurtz dans ses premières versions de 1963 : mais aujourd'hui un langage « moderne » comme **Visual Basic** permet naturellement la programmation récursive. D'autres langages comme **Lisp** et, naturellement, **Pascal** avaient intégré cette possibilité dès leur mise au jour.

Il n'y a pas de programmes récursifs (même si cette expression est parfois utilisée par extension) mais, dans le cas de Pascal, uniquement des procédures (et donc des fonctions) récursives.

Définition 1

Une procédure `proc(VAR x, ...)` est **récursive** si, dans la liste des instructions qui la définissent, on rencontre un (ou plusieurs appels) à la même procédure portant sur au moins une valeur du paramètre

```
PROCEDURE proc(VAR x, ...);

    BEGIN
        (...)
        proc(a, ...);
        (...)
        proc(b, ...);
        (...)
    END;
```

Le livre de première année a déjà brièvement présenté une telle procédure, doublement récursive, à propos du calcul de coefficients du binôme (sujet largement repris plus bas).

Disposer de la récursivité permet souvent une programmation plus rapide et agréable ; le côté plus court de la liste des instructions garantit d'une certaine façon un moindre risque d'erreurs, mais cela présente aussi de graves inconvénients.

En effet, le déroulement de l'exécution de la procédure se doit de garder trace de chacun des appels à elle-mêmes (et ils peuvent être très nombreux, par exemple dans le cas d'une procédure doublement récursive). S'il s'agit d'objets assez lourds, comme des matrices par exemple, l'encombrement peut très vite dépasser les possibilités de la mémoire (il y a des exemples, doublement récursifs, où l'algorithme est bloqué dès que l'on atteint des matrices d'ordre 4 ou 5).

D'ailleurs, dans leurs manuels, les auteurs des manuels présentant les différentes versions d'un langage comme Turbo Pascal sont très discrets sur la possibilité de recourir à la récursivité, et ne fournissent notamment aucune indication sur la *profondeur* admise pour de tels programmes sous tel système d'exploitation. L'expérience montre que le phénomène d'engorgement se produit souvent très vite.

En résumé, l'usage effectif de la récursion (thème théorique pourtant très fascinant) doit être restreint. Et cela d'autant plus qu'il est toujours possible de transformer une procédure récursive en procédure itérative (on appelle cela la **dérécursifier**) ; d'ailleurs, au niveau de la machine elle-même, il n'existe qu'un seul mode, le mode itératif.

Nous admettons ce théorème très important ; on en trouvera plusieurs exemples dans ce qui suit, mais la technique générale, à base de piles et de booléens par exemple, est trop complexe pour pouvoir être exposée ici.

Dans la plupart des cas au programme de la classe, cette dérécursification est pourtant généralement facile. Pour ne prendre comme exemple que celui dont nous étions partis, le calcul de $u_{n+1} = f(u_n)$, elle est même triviale sous la forme suivante

```
BEGIN
  u := a
  FOR i = 1 TO n DO u := f(u) ;
  WRITE (u)
END ;
```

► Remarque

Pour terminer cette trop brève introduction, nous ne résistons pas au plaisir de poser au lecteur le problème suivant, inventé par John Mac Carthy : définir directement la suite u dont les valeurs sont obtenues grâce à la fonction récursive ci-dessous

```
FUNCTION u (n : INTEGER) : INTEGER ;

BEGIN
  IF n > 100
    THEN u := n-10
    ELSE u := u(u(n+11))
  END ;
```

La réponse est surprenante : on a

$$u(n) = \max (n-10, 91)$$

d'où le nom de « programme 91 » qui lui est parfois donné. Il est tout à fait possible de démontrer cette étonnante égalité après coup, à la main, par une récurrence normale. Une meilleure façon de faire est sans doute d'entrer ce programme en machine, de le faire tourner un certain nombre de fois, et de suivre ses péripéties : la preuve indiquée vient alors assez naturellement.

En tout cas cet exemple frappant montre qu'interpréter une définition récursive n'est pas toujours un jeu mécanique pour enfants.

1.1 Factorielle

Ce sujet est toujours celui qui ouvre les cours sur la programmation récursive. D'ailleurs ni Turbo Pascal et *a fortiori* le Pascal standard n'acceptent le symbole $n!$; il est donc extrêmement intéressant de donner un algorithme itératif puis un algorithme récursif permettant de calculer ce nombre pour n entier positif ou nul. Tous deux sont triviaux.

```
FUNCTION Factorielle1 (n : INTEGER) : INTEGER ;
VAR f, i : INTEGER ;

BEGIN
  f := 1
  FOR i := 1 TO n DO f := f*i ;
  Factorielle1 := f
END ;
```

À cause des limitations du type INTEGER, cette fonction ne renvoie les valeurs exactes de $n!$ que jusqu'à $7! = 5040$. Pour $n > 7$, le résultat est égal à la

bonne valeur, diminuée d'un certain multiple de $2^{16} = 65536$ de façon à rentrer dans le cadre `INTEGER`; par exemple $n = 17$ renvoie -32768 au lieu de $355687428096000 = -32768 + 5427359438 \cdot 2^{16}$. Pour $n > 17$ on trouve 0 : c'est normal, puisque 2^{16} divise exactement $18!$.

► **Remarque**

Le lecteur pourra transformer la fonction ci-dessus de façon à donner à `f` et `Factorielle1` le type `REAL`; pour afficher le résultat, on peut utiliser la fonction `TRUNC (x+0.5)` (voir des exemples plus loin). Mais ici aussi, il y a une limite à la fiabilité du programme : rappelons que les valeurs absolues des `REAL` non nuls sont approximativement comprises entre $3 \cdot 10^{-39}$ et $7 \cdot 10^{38}$.

Le programme ainsi modifié tourne exactement jusqu'à $12! = 479001600$. Pour $13 \leq n \leq 33$, le résultat retourné est toujours le même : -2147483648 . Au delà, le programme s'arrête faute de place en mémoire; heureusement la présence du signe $-$ attire l'œil et invite à la méfiance...

Voici enfin une version récursive, très courte et très facile à écrire, toujours basée sur les égalités fondamentales

$$0! = 1, \quad n! = n * (n - 1)! \quad \text{pour } n > 0.$$

```
FUNCTION Factorielle2 (n : INTEGER) : INTEGER ;
    BEGIN
        IF n = 0
            THEN Factorielle2 := 1
            ELSE Factorielle2 := n*Factorielle2(n-1)
        END ;
```

Les résultats sont absolument identiques à ceux que donne le premier programme itératif en mode `INTEGER`.

Il est très facile de vérifier que ces deux versions font bien ce que l'on attend qu'elles fassent, en examinant calmement la modification des différentes variables au cours du déroulement de leur boucle : le lecteur est prié de la tracer lui-même.

► **Remarque**

Le lecteur pourra, ici encore, transformer la fonction ci-dessus de façon à donner à `Factorielle2` le type `REAL`. Le programme ainsi modifié tourne exactement jusqu'à $7! = 5040$. Pour $8 \leq n \leq 9$, aucun résultat n'est retourné, l'utilisateur étant renvoyé silencieusement vers le programme. Au delà, le programme s'arrête faute de place en mémoire.

Passer de `INTEGER` à `REAL` est assez catastrophique puisque, si la limite $n = 7$ est la même, au moins obtenait-on quelques résultats dont on peut tirer quelque chose jusqu'à 17 . De même, on voit qu'en mode `REAL`, passer de l'itératif au récursif est également très pénalisant. Rien n'aurait pu faire prévoir *a priori* ces dégradations.

1.2 Résolution dichotomique d'une équation

Reprenons un exemple donné dans le chapitre **Continuité sur un intervalle** du livre de première année. Le programme **itératif** ci-dessous calcule une valeur appro-

chée d'une racine de l'équation $\frac{e^x + e^{-x}}{10} = x$ dans l'intervalle $[0, 1]$. On trouve $x \approx 0,2041836$ qui est une réponse convenable.

Pour des raisons de plus grande simplicité de la liste d'instructions, ce programme se limite à un calcul unique, celui de $e^x + e^{-x} = 10x$ pour $0 < x < 1$; il serait naturellement immédiat de le transformer de façon à le rendre interactif, modifiant à la guise de l'utilisateur l'intervalle de recherche d'une solution. Il en sera de même pour sa version récursive.

```

PROGRAM Dichotomie1;
VAR min, max, t : REAL;

FUNCTION f(x : REAL) : REAL;
BEGIN
    f := -x + (exp(x) + exp(-x))/10
END;

PROCEDURE Racine(VAR c : REAL);
VAR a, b, eps, u, v, w : REAL;

BEGIN
    eps := 1e-8;
    a := min; b := max;
    u := f(a); v := f(b);
    REPEAT
        c := (a+b)/2; w := f(c);
        IF u*w < 0
            THEN BEGIN b := c; v := w END
            ELSE BEGIN a := c; u := w END
        UNTIL ABS(v-u) < eps;
    END;

BEGIN
    min := 0; max := 1;
    Racine(t);
    WRITELN(t)
END.

```

L'idée est naturelle : partant d'un segment, noté $[a, b]$ par exemple, pour lequel une fonction continue f prend des valeurs de signes différents, on le coupe en deux en calculant son milieu $c = \frac{a+b}{2}$ et en remplaçant, selon le signe de $f(a)f(c)$, le segment initial $[a, b]$ par $[a, c]$ ou $[c, b]$, de largeur deux fois moindre, et ce jusqu'à ce que cette largeur devienne inférieure à une constante eps donnée.

► **Remarques**

- La fonction f n'est pas supposée monotone sur $[a, b]$.
- L'inégalité $a < b$ n'est nullement nécessaire, comme on le voit en changeant f en $-f$. Mais la valeur approchée de la racine retournée peut être naturellement différente.

Cet algorithme itératif devient facilement, à peu de frais, un algorithme récursif. Voici donc une version **récursive** de la résolution de la même équation. Le lecteur est invité à regarder de près les quelques transformations subies par le programme ci-dessus, qui n'en altèrent ni l'esprit ni les qualités.

```

PROGRAM Dichotomie2;
VAR eps : REAL;

FUNCTION f(x : REAL) : REAL;

BEGIN
  f := -x + (exp(x) + exp(-x))/10
END;

FUNCTION Racine(a, b, u, v : REAL) : REAL;
VAR c, w : REAL;

BEGIN
  IF ABS(b-a) < eps
  THEN Racine := a
  ELSE
  BEGIN
    c := (a+b)/2; w := f(c);
    IF u*w < 0
    THEN Racine := Racine(a, c, u, w)
    ELSE Racine := Racine(c, b, w, v)
  END
END;

BEGIN
  eps := 1e-8;
  WRITELN (Racine(0, 1, f(0), f(1)))
END.

```

Comme pour la version itérative, on trouve le résultat tout à fait correct 0,2041836.

► Remarque

On peut transformer de la même manière en algorithme récursif l'algorithme itératif de calcul d'une intégrale présenté, lui aussi, dans le livre de première année, mais sans que l'on y trouve un gain particulier.

1.3 Exponentiation rapide

Une application de la récursivité, beaucoup moins triviale que l'étude précédente de la fonction factorielle, est donnée par le problème suivant, à l'air innocent : **trouver une méthode rapide de calcul de puissances**, plus précisément d'un x^n où x est un réel positif et n un entier positif (le cas général x^y ne rend pas dans ce cadre). La méthode naïve (avec une boucle FOR) est si évidente que nous ne la donnerons pas.

Pour des raisons de plus grande simplicité des listes d'instructions, les programmes ci-dessous se limitent à un calcul unique, celui de $(2, 5)^{96}$; il serait naturellement immédiat de les transformer de façon à les rendre interactifs, modifiant à la guise de l'utilisateur le nombre à porter à une certaine puissance entière ainsi que la valeur de l'exposant.

```

PROGRAM exprap ;

FUNCTION expr(x : real ; n : INTEGER) : REAL ;
VAR y : REAL ; m : INTEGER ;

BEGIN
  IF n = 0
    THEN expr := 1
    ELSE BEGIN
      m := n DIV 2 ;
      y := expr(x, m) ;
      y := y*y ;
      IF n > 2*m THEN y := x*y ;
      expr := y
    END
  END ;

BEGIN
  WRITE (expr(2.5, 96))
END.

```

La valeur 96 de l'exposant auquel on veut porter le nombre $\frac{5}{2}$ n'est pas là par hasard ; en effet, tel qu'il est écrit, ce programme a retourné une valeur tout à fait convenable, à savoir $(2, 5)^{96} \approx 1,59309 \cdot 10^{38}$. Toutefois une tentative analogue pour obtenir $(2, 5)^{97}$ conduit à un échec et à la publication d'un message d'erreur (*overflow*, c'est-à-dire dépassement des limites permises), en tout cas avec les versions du logiciel à la disposition des auteurs. Effectivement, on rappelle que le plus grand réel reconnu par Turbo Pascal est $2^{127} - 2^{87} \approx 1,70141 \cdot 10^{38}$, ce qui explique parfaitement le rôle limite du nombre 96.

```

PROGRAM rapexp ;

FUNCTION expr(x : real ; n : INTEGER) : REAL ;
VAR y : REAL ; m : INTEGER ;

BEGIN
  IF n = 0
    THEN expr := 1

```

```

ELSE BEGIN
    m := n DIV 2 ;
    y := expr(x*x, m) ;
    IF n > 2*m THEN y := x*y ;
    expr := y
END
END ;

BEGIN
    WRITE (expr(2.5, 96))
END.

```

Ce second programme semble un peu meilleur que le précédent : en écrivant $(x * x)^m$ plutôt que $(x^m)^2$, on gagne une ligne et le processus est encore plus lisible.

Malheureusement, si l'on trouve bien de manière analogue que $(2,5)^{63} \approx 1,17549 \cdot 10^{25}$, ce qui est tout à fait convenable, le calcul de $(2,5)^{64}$ conduit à un dépassement, et *a fortiori* pour 96. La comparaison entre ces deux limites, 96 et 63, est cruelle pour une toute petite modification qui paraissait pourtant être une bonne idée (recommandée dans quelques livres).

Il existe également une version itérative de l'exponentiation rapide. Comme elle est malheureusement soumise à la même limitation que `rapexp`, nous l'avons écrite pour $n = 63$ et $x = 2.5$.

```

PROGRAM x_puissance_n ;
VAR m, n : INTEGER ; a, x : REAL ;

BEGIN
    n := 63 ; x := 2.5 ;
    a := 1 ;
    WHILE n > 0 DO
        BEGIN
            m := n DIV 2 ;
            IF n > 2*m THEN a := a*x ;
            x := x*x ; n := m
        END ;
        WRITE (a)
    END.

```

Le lecteur est prié de démontrer, par un invariant de boucle convenable, que cet algorithme fait bien ce qu'il est censé faire dans le cas général. S'il connaît l'écriture binaire d'un entier, la tâche lui sera facilitée.

Dans le cas particulier choisi, on pourra noter la remarquable égalité

$$x^{63} = x^{32} x^{16} x^8 x^4 x^2 x$$

et regarder avec précision comment se déroule alors l'exécution du programme `x_puissance_n`.

Les tests sont presque triviaux à réaliser par la machine en raison de la manière particulière de stockage des entiers en mémoire ; ce qui est coûteux ce sont les multiplications. On peut montrer que les différentes exponentiations rapides présentées ici nécessitent toutes un nombre de multiplications de l'ordre de $\ln n$.

Dans le cas des exponentiations, la version récursive a un champ d'application plus vaste que celui de la version itérative, ce qui est loin d'être toujours le cas.

1.4 Suite de Fibonacci

Cette partie, très classique mais ne figurant pas explicitement au programme, n'est présente ici que comme entrée à la partie suivante, plus complexe, où sont également présentées des formes itératives, simplement récursives ou doublement récursives.

La suite de Fibonacci est définie par les formules récursives

$$f_0 = 0, \quad f_1 = 1, \quad f_{n+2} = f_{n+1} + f_n.$$

Ses premières valeurs sont donc

$$0 \quad 1 \quad 1 \quad 2 \quad 3 \quad 5 \quad 8 \quad 13 \quad 21 \quad 34 \quad \dots$$

Il existe une formule exacte donnant la valeur de f_n en fonction de n , à savoir

$$f_n = \frac{\varphi^n - (-\varphi)^{-n}}{\sqrt{5}}$$

où $\varphi = \frac{1 + \sqrt{5}}{2} \approx 1,618$ est le *nombre d'or*. C'est donc un entier voisin de $\frac{\varphi^n}{\sqrt{5}}$.

La suite (f_n) est croissante ; $f_{23} = 28657$ est le plus grand des termes de la suite inférieurs à $2^{15} = 32768$ puisque $f_{24} = 46368 = 65536 - 19168$.

Si l'on se restreint au type `INTEGER`, il est donc hors de question d'écrire des algorithmes de calcul des f_n avec un $n > 23$, en sachant que les résultats d'indices supérieurs ne seront pas exacts, mais donneront des valeurs différant de f_n par un multiple entier de $2^{16} = 65536$.

On trouvera ci-dessous quatre programmes, calculant tous f_{23} , deux récursifs et deux itératifs. Tous donnent la valeur exacte de f_{23} .

Pour des raisons de plus grande simplicité des listes d'instructions, les programmes ci-dessous se limitent donc à un calcul unique ; il serait naturellement immédiat de les transformer de façon à les rendre interactifs, modifiant à la guise de l'utilisateur la valeur de l'indice du terme de la suite à déterminer.

- Le premier est **doublement récursif**, puisqu'il faut deux appels à des f_i à chaque étape pour calculer un f_j . Il suit de très près la définition récursive vue plus haut.

En général une telle double récursivité est à regarder avec précautions, car il existe beaucoup d'étapes dans les calculs de f_{22} et f_{21} , eux-mêmes résultant des valeurs de f_{20} et f_{19} et ainsi de suite, certaines valeurs étant recalculées plusieurs fois.

Toutefois ces craintes légitimes sont apaisées par l'expérience du calcul, qui se déroule sans problème. Mais, si l'on essaie de calculer des f_n avec n beaucoup plus grand, il peut arriver que le programme bloque complètement le logiciel et même l'ordinateur, qu'il faut alors éteindre et relancer.

```
PROGRAM Fibonacci1 ;

    FUNCTION f1(n : INTEGER) : INTEGER ;

        BEGIN
            IF n < 2
                THEN f1 := n
                ELSE f1 := f1(n-1)+f1(n-2)
            END ;
        END ;
```

```
BEGIN
    WRITE (f1(23))
END.
```

Nous ne donnerons pas de preuve détaillée de la validité de ce programme, tellement cette dernière est claire.

► **Remarque**

On peut montrer que sa complexité (ici le nombre d'additions) est en f_n , c'est-à-dire de l'ordre de φ^n et donc un peu inférieure à 2^n . C'est une complexité très élevée.

- Le deuxième programme est **simplement récursif**. Il n'est pas évident que tout algorithme doublement récursif puisse être simplifié sous une forme de récursion simple, mais c'est le cas (en fait, on peut même le mettre sous forme itérative, comme on le verra plus bas sur le cas particulier de la suite de Fibonacci).

```
PROGRAM Fibonacci2 ;
VAR n, u, v : INTEGER ;

    PROCEDURE f2(VAR n, u, v : INTEGER) ;
    VAR w : INTEGER ;

        BEGIN
            IF n = 0
                THEN
                    BEGIN
                        u := 0 ; v := 1
                    END
                ELSE
                    BEGIN
                        n := n-1 ;
```



```

                f2(u, v) ;
                w := u ; u := v ; v := v+w
            END
        END ;

BEGIN
    n := 23 ;
    f2(u,v) ;
    WRITE (u)
END.

```

L'idée essentielle est la suivante : si au début d'une étape de la boucle on dispose de $u = f_{i-1}$ et de $v = f_i$, le remplacement en fin d'étape du couple (u, v) par le couple $(v, u + v)$ revient à modifier le couple (f_{i-1}, f_i) par le couple (f_i, f_{i+1}) et ainsi de suite. Le calcul d'un invariant de boucle est alors facile à mener, c'est lui qui prouvera que l'algorithme fait bien ce qu'il doit faire, et en un temps fini : nous laissons au lecteur la formalisation complète de cette preuve.

► **Remarques**

- On peut montrer que sa complexité est de l'ordre de n . C'est une complexité plutôt élevée, mais très inférieure à la précédente.
- On notera que dans la procédure, u et v sont passés en variable : cela est rendu nécessaire pour que `Fibonacci2(u,v)` puisse agir comme une fonction (on sait que Pascal ne connaît pas de fonctions à valeurs vectorielles).
- Le troisième programme est une version **itérative**, facile à écrire à partir du deuxième programme (et réciproquement) ; on y retrouve encore le rôle essentiel d'un couple (u, v) avec le même échange $(u, v) \leftrightarrow (v, u + v)$. La ressemblance est frappante, même après une lecture superficielle. Il n'est donc pas étonnant que les performances des deux algorithmes soient pratiquement identiques.

```

PROGRAM Fibonacci3 ;
VAR i, n, u, v, w : INTEGER ;

BEGIN
    n := 23 ;
    u := 0 ; v := 1 ;
    FOR i := 1 TO n DO
        BEGIN
            w := u ; u := v ; v := v+w
        END ;
        WRITE (u)
    END.

```

► **Remarque**

On peut montrer que, comme il a déjà été dit, sa complexité est en n comme l'était la précédente.

- Enfin le quatrième est encore **itératif**, a un air de famille avec les deux précédents, mais sa lecture est *a priori* plus difficile que celles des trois autres.

```

PROGRAM Fibonacci4;
VAR m, n, h, k, u, v, w : INTEGER;

BEGIN
  n := 23;
  u := 0; v := 1;
  h := 1; k := 0;
  WHILE n > 0 DO
    BEGIN
      m := n DIV 2;
      IF n > 2*m
      THEN
        BEGIN
          w := u*h; u := v*h+u*k+w; v := v*k+w
        END;
      n := m; w := h*h; h := 2*k*h+w; k := k*k+w
    END;
  WRITE (u)
END.

```

La preuve de la validité du programme repose sur l'égalité matricielle

$$\begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} f_{n-1} \\ f_n \end{pmatrix}.$$

Nous ne la donnerons pas ici, non qu'elle soit difficile, mais parce qu'elle peut être le sujet d'un exercice intéressant à développer par chacun.

► Remarques

- On peut montrer que la complexité est maintenant en $\ln n$. C'est une complexité excellente ; on passe de 2^n (le premier programme), à n (les deux suivants) et enfin à $\ln n$ (le dernier). À chaque fois le progrès serait tout à fait notable si l'on pouvait atteindre de grands nombres, par exemple $n > 100$. Mais les sévères limitations du Turbo Pascal font que le gain de temps est beaucoup moindre dans les limites où l'on est obligé de se tenir.
- Il est intéressant de transformer ces quatre programmes en remplaçant le type INTEGER par le type REAL, et de comparer les performances des uns et des autres. Cela ne demande qu'un peu de programmation, d'abord sur papier naturellement, puis au clavier pour expérimenter. Il ne faut toutefois pas en attendre d'améliorations notables.

1.5 Coefficients du binôme

Reprenons un thème déjà abordé dans le livre de première année : le calcul des $\binom{n}{p}$, et particulier celui de $\binom{25}{13} = 5200300$. Nous allons brièvement écrire six

programmes (deux itératifs et quatre récursifs) capables de satisfaire le calcul des $\binom{n}{p}$ pour de petites valeurs de n et p , mais dont un seul vient à bout de notre exemple numérique à l'air bien inoffensif.

Pour des raisons de plus grande simplicité des listes d'instructions, ces programmes se limitent à un calcul unique, celui de $\binom{25}{13}$; il serait naturellement immédiat de les transformer de façon à les rendre interactifs, modifiant à la guise de l'utilisateur les deux paramètres du coefficient du binôme à déterminer.

Algorithmes itératifs

Ces deux programmes s'appuient sur l'égalité

$$\binom{n}{p} = \frac{n}{p} \binom{n-1}{p-1}.$$

Le premier travaille dans le mode REAL, le second dans le mode INTEGER.

Le résultat est affiché comme un entier grâce à une propriété cachée de la fonction TRUNC, qui renvoie en principe un entier du langage, c'est-à-dire entre $-2^{15} = -32768$ et $2^{15} - 1 = 32767$, mais qui ici accepte contre les règles d'afficher normalement la valeur d'un réel tel que 5200300 qui est pratiquement un entier.

```
PROGRAM np1 ;
VAR n, p, i : INTEGER ; c : REAL ;

BEGIN
  n := 25 ; p := 13 ; c := 1 ;
  FOR i := n DOWNTO n-p+1 DO c := (c*i)/(n+1-i) ;
  WRITE (TRUNC (c+0.5))
END.
```

Ce programme affiche bien le nombre voulu, à savoir 5200300. Ce sera le seul avec la version de Turbo Pascal utilisée par les auteurs (naturellement, les autres programmes fonctionnent aussi, mais pour des valeurs plus modestes de n et de p). Notons qu'il a naturellement aussi ses limites.

```
PROGRAM np2 ;
VAR c, n, p, i : INTEGER ;

BEGIN
  n := 25 ; p := 13 ; c := 1 ;
  FOR i := n DOWNTO n-p+1 DO c := (c*i) DIV (n+1-i) ;
  WRITE (c)
END.
```

Pour ne pas passer en mode REAL, il a bien fallu remplacer la division usuelle x/y par la division entière $x \text{ DIV } y$.

En revanche, le nombre cherché étant de type INTEGER, il n'est pas indispensable d'utiliser TRUNC. Cela dit, on pourrait s'attendre à ce que le résultat soit en tout cas exact à un multiple près de 2^{16} . Or le programme retourne -2148 alors que le « moins mauvais » résultat attendu serait plutôt $22956 = 5200300 - 76 * 2^{16}$.

En fait, les erreurs dues à la limitation du type INTEGER remontent beaucoup plus haut, de manière précise à la quatrième étape de la boucle : jusque là, on avait bien trouvé les valeurs de $\binom{25}{p}$ pour $p \geq 22$, à savoir 25, 300 et 2300, mais le calcul de l'étape suivante donne en fait

$$(2300 * 22) \text{ DIV } 4 = 50600 \text{ DIV } 4 = -14936 \text{ DIV } 4 = -3734$$

au lieu du résultat normal $\binom{25}{21} = 12650$.

L'égalité dont nous sommes partis étant bien de type récursif, nous allons voir ce que donnent nos deux premiers programmes en utilisant la puissance de la récursivité.

Algorithmes simplement récursifs

Les deux programmes suivants s'appuient encore sur l'égalité

$$\binom{n}{p} = \frac{n}{p} \binom{n-1}{p-1}.$$

Le premier travaille dans le mode REAL, le second dans le mode INTEGER.

Ils mettent en œuvre des fonctions définies récursivement.

```
PROGRAM np3 ;

FUNCTION comb(n, p : INTEGER) : REAL ;

BEGIN
  IF p > n
    THEN comb := 0
    ELSE BEGIN
      IF 2*p > n
        THEN p := n-p ;
      IF p = 0
        THEN comb := 1
        ELSE comb := (n*comb(n-1, p-1))/p
      END
    END ;

BEGIN
  WRITE (TRUNC (comb(25,13)+0.5))
END.
```

On notera que l'on a ajouté, par rapport aux algorithmes itératifs, quelques précautions écartant des cas triviaux (on aurait d'ailleurs aussi pu ajouter un test sur la positivité de p) et, surtout, le remplacement systématique de p par $n-p$ si ce dernier lui est strictement inférieur.

Un message d'erreur indiquant une instruction invalide en calcul flottant (tel est le nom savant des opérations sur les réels de Turbo Pascal) fait que le programme se bloque *et ne retourne aucune réponse*.

Des expériences menées avec le logiciel dont nous disposons montrent que, pour $n = 25$, le programme fonction parfaitement de $p = 25$ jusqu'à $p = 18$ ($\binom{25}{18} = 480700$); pour $p = 17$ ou $p = 16$, l'exécution du programme n'indique rien de spécial sauf qu'elle renvoie silencieusement au texte du programme. Enfin le message d'erreur signalé apparaît dès $p = 15$.

Le suivant est quasiment identique, sauf que le type de `comb` est maintenant `INTEGER`, d'où naturellement la présence d'un `DIV`.

```
PROGRAM np4 ;

    FUNCTION comb(n, p : INTEGER) : INTEGER ;

    BEGIN
        IF p > n
            THEN comb := 0
            ELSE BEGIN
                IF 2*p > n
                    THEN p := n-p ;
                IF p = 0
                    THEN comb := 1
                    ELSE comb := (n*comb(n-1, p-1)) DIV p
                END
            END ;

    END ;

    BEGIN
        WRITE (comb(25,13))
    END.
```

On pourrait s'attendre à ce que le résultat soit au moins exact à un multiple près de 2^{16} . Or le programme retourne 2134 alors que le « moins mauvais » résultat attendu serait plutôt $22956 = 5200300 - 76 * 2^{16}$.

Essayant de comprendre ce qui se passe, on s'aperçoit qu'il y a déjà erreur dès le calcul de $\binom{25}{21}$, à la même étape que dans `np2` : on trouve en effet $\binom{25}{21} = -3733$ (valeur très proche du résultat partiel atteint dans le déroulement de `np2`, à savoir -3734) au lieu de 12650.

Algorithmes doublement récursifs

Les deux derniers programmes s'appuient sur l'égalité de Pascal

$$\binom{n}{p} = \binom{n-1}{p} + \binom{n-1}{p-1}.$$

Le premier travaille dans le mode REAL, le second dans le mode INTEGER.

Ils mettent encore en œuvre des fonctions définies récursivement.

```
PROGRAM np5 ;

    FUNCTION comb(n, p : INTEGER) : REAL ;

    BEGIN
        IF p > n
            THEN comb := 0
            ELSE
                BEGIN
                    IF 2*p > n
                        THEN p := n-p ;
                        IF p = 0
                            THEN comb := 1
                            ELSE comb := comb(n-1, p) + comb(n-1, p-1)
                        END
                END
            END ;

    BEGIN
        WRITE (TRUNC (comb(25,13)+0.5))
    END.
```

Ici, le programme ne se bloque pas comme dans np3, mais s'arrête sans message d'erreur en renvoyant simplement au texte du programme *et ne retourne naturellement aucune réponse*.

Le suivant est quasiment identique, sauf que le type de comb est maintenant INTEGER, d'où naturellement la présence d'un DIV.

```
PROGRAM np6 ;

    FUNCTION comb(n, p : INTEGER) : INTEGER ;

    BEGIN
        IF p > n
            THEN comb := 0
            ELSE BEGIN
                    IF 2*p > n
```

```

        THEN p := n-p ;
    IF p = 0
        THEN comb := 1
        ELSE comb := comb(n-1, p) + comb(n-1, p-1)
    END
END ;

END ;

BEGIN
    WRITE (comb(25,13))
END .

```

Le résultat renvoyé par l'exécution du programme est encore décevant à première vue, puisqu'il vaut 22956. Cela dit, il est en tout cas égal à $\binom{25}{13}$ à un multiple de 2^{16} près, ce qui est tout à fait normal.

Il faut noter que la double récursivité a été ici, d'un certain point de vue, plus efficace que la simple, en dépit de ce que l'intuition pourrait laisser penser.

Retenons donc de ces quelques exemples que, même sur des thèmes extrêmement simples figurant d'ailleurs explicitement au programme, dès que l'on sort du cadre de très petits nombres (25 et 13 ne sont pas si énormes) la fiabilité des outils logiciels trop classiques est très douteuse. En revanche, le recours à **Mathematica** ou à **Maple**, qui peuvent traiter de très grands entiers, permettrait d'aller beaucoup plus loin, mais il connaît aussi naturellement ses limites.

2. Gestion de listes à une dimension

Au programme de la classe figure une timide initiation au traitement de listes finies de données. Pour simplifier, ces listes sont ici supposées indexées par des entiers naturels i allant de 1 à un certain entier n pouvant atteindre 10000 par exemple. Ces entiers sont appelés *clefs* ou *rangs*, voire *indices*, et leur nature fait que l'on peut les comparer entre elles, ce qui est très précieux. Les valeurs de la liste, notées ici $A[i]$, seront supposées être des nombres (entiers ou réels), également comparables ce qui est également bien commode. Dans la pratique, il s'agit d'enregistrements de nature souvent plus complexe que celle de nombres.

Une théorie plus poussée aurait à travailler sur des données à plusieurs dimensions (comme des $A[i, j, k]$) et de types non numériques, éventuellement avec des clefs également non numériques (alphabétiques par exemple) ; mais ces questions ne seront pas abordées ici.

Dans toute cette section, A est une donnée initiale de type `tab`, c'est-à-dire `ARRAY[1..n] OF num`, où n est une constante entière définie dans le programme, supérieure ou égale à 2, et `num` est par exemple `INTEGER` ou `REAL`.

2.1 Recherche d'éléments extrémaux

Le but de cette partie est de rechercher, compte tenu des valeurs des éléments numériques $A[j]$ où j varie de 1 à n , l'un des $A[i]$ vérifiant l'inégalité $A[i] \geq A[j]$ pour toute clef j entre 1 et n . Une première procédure ci-dessous se limite à ce problème, la seconde montre comment obtenir à la fois un maximum et un minimum, avec un léger gain en complexité (un quart) sur la technique trop « naïve » consistant à mettre bout à bout deux versions du premier.

► Remarques

- Bien que présente dans des documents officiels, la phrase « recherche de la valeur et **du** rang d'un maximum d'une liste » est incorrecte, puisqu'il arrive qu'un maximum, par exemple, soit présent plusieurs fois dans la liste, avec des rangs différents.
- Bien que l'on puisse également déterminer assez facilement **tous** les éléments maximums, c'est-à-dire leur valeur commune et l'ensemble de leurs clefs, nous nous limiterons ci-dessous à la recherche de cette valeur et de l'une de ses clefs.

Recherche d'un élément maximum

L'algorithme est trivial; on peut montrer que l'exemple de la procédure ci-dessous est presque optimal, le nombre de ses $n - 1$ comparaisons ne pouvant être diminué comme le montre le cas d'une liste écrite en ordre décroissant.

Exemple

```
PROCEDURE Maximum(VAR max : num);
    VAR i, j : INTEGER;

    BEGIN
        i := 1; max := A[i];
        FOR j := 2 TO n DO
            BEGIN
                IF A[j] > max THEN
                    BEGIN
                        i := j; max := A[i]
                    END
                END
            END
        END ;
```

À la fin de la boucle, la valeur `max` est celle du maximum, et celle de `i` est celle de l'une de ses clefs (plus exactement la plus petite possible).

Il serait évidemment possible de ne déterminer que la clef `i`, puisque $M=A[i]$, mais le programme serait pratiquement aussi coûteux en instructions (sans compter qu'une comparaison du type $A[j] > A[i]$ est un tout petit peu plus complexe qu'un simple $A[j] > \text{max}$).

Changer le symbole d'inégalité $>$ en $<$ donnerait naturellement un élément minimum.

La complexité de ce tri « naïf » (c'est-à-dire essentiellement le nombre de tests ou d'affectations) est donc de l'ordre de n . De manière plus précise, elle vaut $n - 1$ comparaisons et, au moins deux affectations dans le cas le plus favorable (où le tableau initial est trié en ordre ascendant) et au plus $2n$ affectations dans le cas le plus défavorable (où le tableau initial est trié en ordre descendant), sans compter les $n - 1$ affectations résultant de la boucle FOR.

Maximum et minimum

Voici maintenant une procédure **récursive** donnant à la fois un maximum `max` et un minimum `min` (sans leurs clefs), avec une diminution sensible du nombre de comparaisons par à l'utilisation doublée du programme ci-dessus. Son étude est plus délicate, néanmoins elle permet une bonne initiation à quelques problèmes effectifs de la programmation, qui exige souvent une attention poussée.

Ici encore l'on suppose connu le tableau à trier `tab`, du type `ARRAY[1..n] OF num` où n est une constante définie dans le programme et `num` est `INTEGER` ou `REAL`.

On supposera vérifiée l'inégalité $g \leq d$ (c'est-à-dire au départ $1 \leq n$: cette condition doit être testée avant l'appel à la procédure, car on n'est jamais trop prudent en une telle matière). Ces deux variables `g` et `d`, passées par valeur, désignent simplement les rangs « origine » et « extrémité » d'un sous-tableau de `A` dont on recherche le maximum et le minimum ; il suffit alors de lancer `Tri_rapide(1, n, A)` pour obtenir le résultat cherché pour le tableau complet.

Le fond de l'algorithme est simple, même si ses détails paraissent complexes. L'idée est de lancer successivement `Tri_rapide(1, 2, A)` et `Tri_rapide(3, n, A)`, puis de recommencer jusqu'à tomber sur des cas triviaux où $d - g = 0$ ou $d - g = 1$ pour lesquels sont prévus des calculs directs. Le fait qu'il est exécuté en temps fini en découle.

Exemple

```
PROCEDURE Maximin(g, d : INTEGER ; VAR max, min : num) ;
  VAR p, q, r, s : num ;

  BEGIN
    IF d = g THEN
      BEGIN
        max := A[g] ; min := A[g]
      END ELSE
      BEGIN
        IF d = g+1 THEN
          BEGIN
            IF A[d] > A[g] THEN
              BEGIN
                max := A[d] ; min := A[g]
              END ELSE

```

```

BEGIN
    max := A[g] ; min := A[d]
    END
END ELSE
BEGIN
    Maximin(g, g+1, p, q) ; Maximin(g+2, d, r, s) ;
    IF p > r THEN max := p ELSE max := r ;
    IF q > s THEN min := s ELSE min := q
    END
END
END ;

```

À la lecture de ce texte, tout programmeur doit avoir comme réflexe de remarquer que l'étude à part du cas $d = g+1$ est inutile, à condition d'appeler $\text{Maximin}(g, g, p, q)$; $\text{Maximin}(g+1, d, r, s)$ au lieu de $\text{Maximin}(g, g+1, p, q)$; $\text{Maximin}(g+2, d, r, s)$, ou encore d'appeler $\text{Maximin}(g, c, p, q)$; $\text{Maximin}(c+1, d, r, s)$ où c est un nombre strictement compris entre g et d , par exemple voisin de leur demi-somme (on peut penser à prendre $c=(g+d) \text{ DIV } 2$). Une autre façon de faire consisterait à regrouper les deux cas spéciaux $d < g+2$.

Une étude précise montre toutefois que ces variantes, qui fonctionnent effectivement, sont un peu moins performantes que la procédure choisie (rien n'est évident en informatique).

Soit $C(n)$ le nombre de comparaisons entre des num (comme les $A[i]$) nécessaires à l'obtention du maximum et du minimum d'un tableau $A[1..n]$. On voit facilement que $C(1) = 0$, $C(2) = 1$, puis que

$$C(n) = C(2) + C(n-2) + 2 = C(n-2) + 3$$

pour $n \geq 3$.

Ainsi peut-on en déduire par récurrence que $C(2m) = 3m-2$ et que $C(2m+1) = 3m$, soit $C(n)$ voisin de $\frac{3n}{2}$ au lieu de $2(n-1)$ avec la procédure naïve. Ce gain d'un quart peut, pour de grandes valeurs de n , se révéler intéressant. Pour les variantes, on trouverait une relation $C(n) - C(n-2) = c > 3$, pour laquelle $C(n)$ est voisin de $\frac{cn}{2} > \frac{3n}{2}$.

Cela dit, pour éviter les problèmes d'encombrement mémoire inhérents à tout algorithme récursif, il faut naturellement le « dérécursifier », c'est-à-dire le transformer en un algorithme itératif. Nous n'entrerons pas dans l'étude de cette transformation, nous contentant de dire qu'elle est tout à fait possible, mais assez complexe à réaliser.

2.2 Tri par insertion et tri rapide

Dans la ligne de l'étude précédente, qui faisait extraire de la liste l'élément maximum (et éventuellement le minimum), il faut dire quelques mots – même si le programme ne le demande pas – de sa généralisation naturelle : partant d'une liste dans le désordre, comment la transformer en une liste formée des mêmes nombres $A[i]$, non plus dans l'ordre de leurs clefs, mais dans celui de leurs valeurs ?

Ce problème de tri est fondamental en informatique et il a reçu d'innombrables solutions, que l'on peut deviner à l'œuvre par exemple dans le maniement des messages électroniques par Internet ou toute application de gestion même élémentaire (listes d'élèves d'une classe rangés par ordre alphabétique ou en fonction de leur dernière note en mathématiques...).

Nous donnerons deux exemples, l'un de compréhension très simple sans être toutefois le plus naïf possible, l'autre beaucoup plus efficace même si nous ne pourrions pas le prouver ici.

Tri par insertion

Pour trier un tableau de taille n , on peut, de manière naïve, en rechercher l'élément maximum par $(n - 1)$ comparaisons et autant d'affectations, puis le mettre en dernière place par au plus $n - 1$ transpositions d'éléments contigus, puis continuer sur le sous-tableau formé des $n - 1$ premiers éléments et ainsi de suite.

La complexité de ce tri « naïf » (c'est-à-dire essentiellement le nombre de tests ou d'affectations) est de l'ordre de n^2 , ce qui n'est pas très bon, même dans le cas le plus favorable (où le tableau initial est trié en ordre descendant) et *a fortiori* dans le cas le plus défavorable (où le tableau initial est trié en ordre ascendant) ; le lecteur est prié de regarder lui-même les calculs, qui sont faciles.

L'algorithme itératif de *tri par insertion* présenté ci-dessous est l'un des plus simples qui soient ; il permet d'obtenir un nouveau tableau A tel que $A[i] \leq A[j]$ pour tout couple vérifiant $i < j$.

Ici encore l'on suppose connu le tableau à trier tab , du type `ARRAY[1..n] OF num` où n est une constante définie dans le programme et num est `INTEGER` ou `REAL`.

Exemple

```
PROCEDURE Tri_par_insertion(VAR A : tab);
    VAR i, j : INTEGER; a : num;

    BEGIN
        FOR i := 2 TO n DO
            BEGIN
                a := A[i]; j := i-1;
                WHILE (j > 0) AND (A[j] > a) DO
                    BEGIN
                        A[j+1] := A[j]; j := j-1
                    END;
            END;
        END;
```

```

        A[j+1] := a
    END
END ;

```

La structure du programme est simple : il y a deux boucles imbriquées. Le seul problème vient de ce que la variable $A[i]$ change tout le temps de valeur, ce qui est normal car le tableau A est justement déclaré comme passé par valeur : en effet la liste va être modifiée, et il serait étonnant qu'on lance une procédure de ce type sans utiliser son travail !

Le fait que le programme s'arrête en un temps fini est assuré par la stricte décroissance de la variable j dans la boucle `WHILE`, et le fait que la boucle extérieure est du type `FOR`.

L'invariant principal de cette boucle extérieure d'indice i variant de 2 à n est que le sous-tableau formé des $A[j]$ avec $j < i$ est déjà trié par ordre croissant au début de la boucle, ce qui est bien le cas pour $i = 2$.

Le nombre a ayant pour valeur la valeur initiale de $A[i]$ va être comparé à tous les $A[j]$ avec $j < i$, et éventuellement échangé avec $A[i-1]$, $A[i-2]$, $A[i-3]$ etc., de façon à être mis en fin de boucle

- soit au premier rang s'il est strictement inférieur à tous ceux qui le précédaient au début de la boucle,
- soit plus généralement au rang h s'il était supérieur ou égal à ceux de rang $k < h$ et strictement inférieur à ceux de rang $k > h$ pour $k < i$.

C'est cette double opération de comparaisons systématiques et de transpositions d'éléments deux à deux contigus de la liste qui fait l'objet de la boucle intérieure.

Il faut remarquer que le programme ne fonctionne que parce que, en Pascal, une évaluation négative du premier booléen p dans un test du genre « `WHILE p OU q` » bloque le processus, ce qui fait que q n'est alors pas évalué (dans ce cas présent, le fait que $j = 0$ implique justement que $A[j]$ n'est pas défini, ce qui aurait bloqué le système).

La complexité du tri par insertion (c'est-à-dire essentiellement le nombre de tests ou d'affectations) est de l'ordre de n^2 , ce qui n'est pas très bon. De manière plus précise, elle évolue approximativement entre $5n$ dans le cas le plus favorable, $\frac{3}{2}n^2$ dans le cas le plus défavorable où le tableau initial est trié en ordre descendant et, « en moyenne », c'est-à-dire par exemple si les $A[i]$ résultent d'appels à la fonction `RANDOM`, elle est voisine de $\frac{3}{4}n^2$.

► Remarques

- Ne pas oublier les parenthèses dans le texte de la boucle `WHILE` ; elles sont par contre inutiles dans le cas d'un booléen simple comme $i > 0$.
- Il existe évidemment des versions récursives de cette procédure, mais on ne gagne pas en clarté par rapport au code ci-dessus ; le lecteur est prié de les écrire lui-même à partir de l'invariant de la boucle extérieure.

Tri rapide

L'algorithme récursif qui suit, bien que d'apparence inoffensive, est assez subtil à décoder. C'est une version du célèbre *Quicksort* découvert par Charles Anthony Richard Hoare en 1962.

Ici encore `tab` est un type `ARRAY[1..n] OF num` où n est une constante définie dans le programme et `num` est `INTEGER` ou `REAL`.

Voici une procédure de tri rapide choisie parmi les plus simples (il en existe plusieurs versions, plus ou moins lourdes, plus ou moins efficaces, mais en gros comparables les unes aux autres).

Exemple

```
PROCEDURE Tri_rapide(g, d : INTEGER; VAR A : tab);
  VAR i, j : INTEGER; a, x : num;

  BEGIN
    i := g; j := d; x := A[(g+d) DIV 2];
    REPEAT
      WHILE A[i] < x DO i := i+1;
      WHILE x < A[j] DO j := j-1;
      IF i <= j THEN
        BEGIN
          a := A[i]; A[i] := A[j]; A[j] := a;
          i := i+1; j := j-1
        END
      UNTIL i > j;
      IF g < j THEN Tri_rapide (g, j, A);
      IF i < d THEN Tri_rapide (i, d, A)
    END;
```

Il suffit ensuite de lancer `Tri_rapide(1, n, A)` pour obtenir le résultat cherché.

Nous ne chercherons pas à justifier qu'elle trie effectivement le tableau `A` (dès lors détruit et remplacé par une nouvelle version, d'où son passage par valeur), mais recommandons au lecteur de le vérifier dans un cas particulier d'une dizaine de nombres, par un déroulement à la main sur papier d'un exemple choisi à cette fin.

Cela dit, il est bon de donner au moins une idée de son fonctionnement. Il s'agit de choisir un élément $p = A[k]$ dit *pivot*, puis de réarranger, par échanges successifs, le tableau de façon que les nouveaux `A[i]` tels que $i < h$ (où h est la nouvelle clef de p) vérifient $A[i] \leq p$ et ceux pour lesquels $j > h$ vérifient $A[j] \geq p$. Chacun des `A[i]` est alors inférieur ou égal à p et à chacun des `A[j]` et inversement; une fois cette partition faite, il ne reste plus qu'à trier séparément les deux sous-tableaux séparés par le pivot par deux appels récursifs.

La plus ou moins grande qualité des différentes versions du tri rapide repose essentiellement sur la technique de séparation utilisée.

On peut montrer que, même pour de grandes valeurs de n , le nombre de récursions reste raisonnable. On peut d'ailleurs, au prix de quelques efforts, réécrire cet algorithme sous forme itérative et non plus récursive, ce qui simplifie d'éventuels problèmes de mémoire.

Il est bon de savoir que dans certains cas défavorables - notamment si le tableau est déjà trié, ce qui est assez étonnant! - le nombre de comparaisons à effectuer est de l'ordre de n^2 , comme pour le tri par insertion (ou le tri par fusion, le tri à bulles et tous les algorithmes plus ou moins naïfs). Par contre, « en moyenne », c'est-à-dire par exemple si les $A[i]$ résultent d'appels à la fonction `RANDOM`, ce nombre tombe à un ordre de $n \ln_2 n$ (logarithme à base 2), ce qui est bien plus favorable : comparer les deux pour $n = 10000$, soit théoriquement un rapport d'environ 1 à 750!

Dans la pratique, il ne doit pas être employé en dessous de $n = 100$, car en cas contraire il est largement battu par un grand nombre d'algorithmes. C'est pourquoi, dans la majorité des implantations effectives de l'algorithme, il est imposé d'appeler le tri par insertion (par exemple) si n est inférieur à un certain seuil.

2.3 Recherche d'un élément dans une liste

Nous allons étudier en dernier le problème suivant : étant donnée une suite mathématique u définie sur un segment $[[1, n]]$ et un nombre γ , de même type que les u_i , rechercher s'il figure dans le tableau des u_i et, si possible, donner au moins une clef i telle que $\gamma = u_i$ (généralement la plus petite).

La situation est très différente selon que u est arbitraire (*recherche linéaire*) ou croissante (*recherche dichotomique*). Dans ce dernier cas, le temps d'exécution de l'algorithme est, en gros, divisé par $\frac{n}{\ln_2 n}$, c'est-à-dire environ 100 si $n = 1000$ (le logarithme est ici à base 2).

Cette situation où u est une suite numérique peut souvent être transposée à des cadres plus généraux. Il en est ainsi pour la recherche du nom d'un abonné au téléphone connaissant son numéro (*ctki?*), avec le cas favorable où les abonnés sont triés par ordre alphabétique : le lecteur pourra s'exercer à adapter les deux procédures ci-dessous à un tel cas.

Recherche linéaire

Ici encore `tab` est un type `ARRAY[1..n] OF num` où n est une constante définie dans le programme et `num` est `INTEGER` ou `REAL`.

Avant l'appel à la procédure ci-dessous, le programme doit aussi fixer la valeur de y , par exemple par un appel `READLN(y)`, supposée de type `num`.

Voici deux réponses possibles au problème de savoir si l'équation $A[i] = y$ a au moins une solution.

- La première est « très naïve » : faire parcourir à i le segment $[[1, n]]$ et noter si l'on constate, ou non, l'égalité cherchée. Dans le cas particulier où $y = A[1]$, on peut ainsi être conduit à effectuer n tests alors qu'un aurait suffi!

- La seconde est proche de la première, mais un peu plus subtile car elle corrige la remarque précédente. En voici une version intelligente dite à *sentinelle* : elle demande que le tableau A soit en fait de type ARRAY[1..n+1] et non plus ARRAY[1..n] ; la valeur A[n+1] sera définie par le programme comme égale à y (du coup l'équation A[i] = y a au moins une solution, peut-être factice).

Exemple

```
PROCEDURE Recherche_lin(VAR i : INTEGER) ;

    BEGIN
        i := 1 ; A[n+1] := y ;
        WHILE y <> A[i] DO i := i+1
    END ;
```

Le lancement de Recherche_lin(i) doit alors par exemple être suivi de l'écriture de la valeur de la clef i, qui est n+1 par convention si y n'apparaît pas dans le tableau et, sinon, la plus petite clef h telle que y = A[h]. Cette variable i peut aussi être utilisée telle quelle dans la suite du programme ; elle suffit pour savoir si la solution proposée au problème de recherche est effective, ou artificielle et devant être rejetée.

► **Remarque**

Pour des raisons de plus grande lisibilité du résultat, on peut ajouter un booléen i < n+1 qui, après la fin de la boucle, indique si la recherche a été fructueuse ou non, ou une instruction finale telle que

```
IF i = n+1 THEN i := 0
```

voire i := -1 qui rend la lecture de i plus claire.

La preuve de cet algorithme, très facile, repose sur l'affirmation suivante : au début d'une mise en œuvre d'une étape de la boucle WHILE, on dispose d'invariants de boucle, à savoir les relations

$$1 \leq i \leq n, \quad A[j] \neq y \quad \text{pour toute clef } j \leq i.$$

Les relations analogues restent vraies, sauf si $i = n$ ou si ($i < n$ et $A[i + 1] = y$), mais dans chacun de ces deux cas l'égalité $A[i + 1] = y$ met alors fin à la boucle. Cela arrive car i finit par atteindre une solution de l'équation à résoudre ou, au pire, la valeur n + 1.

Il est facile d'en déduire que, dans le pire des cas, le nombre de comparaisons à effectuer dans les tests est n + 1. On peut montrer que, même « en moyenne », c'est-à-dire par exemple si les A[i] résultent d'appels à la fonction RANDOM, ce nombre reste de l'ordre de n, plus précisément voisin de $\frac{n}{2} + 1$.

Recherche dichotomique

Cette partie est consacrée à un cas particulier très fréquent : **la liste est triée par ordre ascendant**, c'est-à-dire $A[i] \leq A[j]$ dès que $i < j$. Il existe alors une technique en $\ln n$ (et non plus n) pour résoudre le problème. Elle s'appelle **recherche**

dichotomique en français (en anglais *binary search*), parce qu'elle revient *grosso modo* à couper en deux la liste en un point médian, afin de choisir la « bonne » moitié, celle qui contient la valeur recherchée.

Ici encore `tab` est un type `ARRAY[1..n]` OF `num` où `n` est une constante définie dans le programme et `num` est `INTEGER` ou `REAL`.

Avant l'appel à la procédure ci-dessous, le programme doit aussi fixer la valeur de `y`, par exemple par un appel `READLN(y)`, supposée de type `num`.

Exemple

```
PROCEDURE Recherche_dicho(VAR i : INTEGER);
    VAR j, k : INTEGER; y : num;

    BEGIN
        i := 1; j := n;
        BEGIN
            WHILE i < j DO
                BEGIN
                    k := (i+j) DIV 2;
                    IF y > A[k] THEN i := k+1 ELSE j := k
                END
            END ;
            IF y <> A[i] THEN i := 0
        END ;
```

Le lancement de `Recherche_dicho(i)` doit alors par exemple être suivi de l'écriture de la valeur de la clef `i`, qui est 0 par convention si `y` n'apparaît pas dans le tableau et, sinon, la plus petite clef `h` telle que `y = A[h]`. Cette variable `i` peut aussi être utilisée telle quelle dans la suite du programme ; elle suffit pour savoir si la solution proposée au problème de recherche est effective, ou artificielle et devant être rejetée.

► Remarque

Pour des raisons de plus grande lisibilité du résultat, on peut ajouter un booléen qui, après la fin de la boucle, indique si la recherche a été fructueuse ou non.

La preuve détaillée de cet algorithme est raisonnablement facile, mais plus longue que la liste de ses instructions ! Nous nous contenterons d'en présenter les grandes lignes, laissant au lecteur le soin de préciser les détails.

Il démontrera d'abord, par exemple par examen des parités des paramètres, la propriété suivante

Lemme

Si trois entiers i, j et k vérifient les relations $i < j$ et $k = i + j \text{ DIV } 2$ (c'est-à-dire si k est la partie entière de la demi-somme de i et de j), ils vérifient également les inégalités

$$i \leq k \leq \frac{i+j}{2} < k+1 \leq j < j+1.$$

Cela fait, la preuve repose essentiellement sur l'affirmation suivante : au début d'une mise en œuvre d'une étape de la boucle WHILE, on dispose d'invariants de boucle, à savoir les relations

$$1 \leq i < j \leq n, \quad A[i-1] < \gamma, \quad A[j] \geq \gamma.$$

Si l'on note i' et j' les valeurs de i et de j à la fin de cette étape, les relations analogues restent vraies, sauf peut-être $i' < j'$: on peut avoir en effet $i' = j'$ mais cela met alors fin à la boucle. Ce qui finit par arriver : on dispose en effet de l'inégalité $j' - i' \leq j - i - 1$.

On peut montrer que cet algorithme a, même dans le cas le plus défavorable, un ordre en $\ln n$, ce qui ne peut être amélioré : il est donc pratiquement optimal.

► **Remarque**

Il existe aussi des versions récursives de ces deux algorithmes de recherche, mais sans intérêt particulier.

3. Simulations de lois réelles discrètes

Le programme de la classe demande explicitement d'écrire des fonctions Pascal permettant de simuler une variable aléatoire, discrète ou à densité, suivant une loi simple donnée.

Nous donnons ci-dessous treize exemples de telles simulations (discrètes et à densité mélangées). Dans un premier temps, on peut même se contenter de répondre à la question posée par un code de quelques lignes seulement, comme le montre le cas suivant consacré à la simulation de la loi uniforme discrète sur le segment $\llbracket 1, n \rrbracket$.

L'outil fondamental des approches pédagogiques sur la simulation et l'estimation est le recours aux fonctions RANDOM et RANDOM(n) de Pascal.

Si l'on note $a := \text{RANDOM}$ le nombre aléatoire renvoyé par le logiciel à un tel appel, et si l'on veut lancer une instruction donnée I avec une probabilité p ou ne rien faire en cas contraire avec la probabilité $1 - p$, il suffit de tester si a appartient à l'intervalle $[0, p[$, ou plus généralement à un intervalle $[\alpha, \alpha + p[$: si la réponse est *oui*, on met I en œuvre, sinon on reste en l'état. En effet, la probabilité pour que, par exemple, $a < p$ est justement égale à p .

Tel est le principe général ; rien n'interdit de compliquer la règle ci-dessus de manière évidente.

Bien entendu, dans le cadre de leur utilisation, des appels à RANDOM ne doivent être mises en jeu qu'après qu'un appel à RANDOMIZE ait été effectué en début de programme comme le montre l'extrait ci-dessous

Exemple

```

FUNCTION Unif(n : INTEGER) : INTEGER ;

    BEGIN
        RANDOMIZE ;
    
```

```

Unif := 1+RANDOM(n) ;
END ;

```

► Remarque

Voici un prototype de structure d'une procédure dans laquelle insérer des fonctions telles que la précédente

Exemple

```

PROCEDURE Machin_chose(...);
  VAR (...);

  BEGIN
    RANDOMIZE ;
    FOR i := 1 TO 100 DO
      BEGIN
        (...);
        WRITE(' ', k)
      END
    END
  END ;

```

La variable k (qui peut aussi être un x , ou même un nom de fonction) est envoyée à l'écran pour pouvoir être lue, par exemple aux fins de statistiques ; bien entendu, dans le cadre de vrais programmes, cette action simpliste devrait être remplacée par une utilisation adaptée au cadre dans lequel on utilise la simulation donnée.

Pour chacun des cas traités ci-dessous, le lecteur pourra extraire de telles fonctions qui constituent le cœur des procédures proposées.

► Remarques

- Si ces dernières sont plus complexes que celles dont il serait possible de se contenter pour répondre aux demandes du programme, c'est pour deux raisons.

La plus importante repose sur le fait qu'être capable de sortir une valeur numérique isolée n'a pas grand sens, il faut de toute manière pouvoir juger expérimentalement de la qualité de l'approximation fournie. C'est pourquoi nous avons écrit des procédures fournissant pour chaque ligne *cent valeurs* de la variable émulée, afin de pouvoir tester, par des calculs de fréquences ou de fréquences cumulées si l'algorithme semble bien fiable (ce sont donc des procédures et non des fonctions, car ces dernières ne peuvent renvoyer de résultat du type `ARRAY [1..100]`). Il est clair qu'une boucle de longueur 1000, voire 10000 fournirait des résultats plus précis. Mais l'on verra à propos de chaque exemple que le nombre 100, bien que faible, permet déjà des approximations honnêtes.

- On pourrait également, grâce à ces expériences simulées, essayer d'estimer espérance et variance de la loi concernée. Là aussi, ce nombre 100 devrait naturellement être largement redimensionné pour essayer d'avoir des estimations convenables.

- Ensuite il est apparu comme plus sage de préciser chaque fois de la manière la plus explicite possible les appels au générateur de nombres au hasard en notant systématiquement de la même manière le nombre aléatoire tiré

```
a := RANDOM
```

même si, comme le montrent les cinq lignes de code ci-dessus, on peut presque toujours se passer d'une telle variable et gagner en concision.

Cela étant, les procédures suivantes semblent assez courtes et clairement structurées pour que l'on puisse saisir leur signification sans effort d'abstraction excessif.

- Un autre intérêt d'introduire systématiquement la variable réelle a réside dans le fait qu'elle met en lumière qu'appeler `RANDOM` c'est simplement introduire une variable aléatoire A définie par des égalités de la forme $A(\omega) = a$ pour tout possible ω dans Ω ; le caractère magique de l'invocation quelque peu théâtrale de `RANDOM` disparaît ainsi.

3.1 Loi uniforme discrète

Cette loi, notée $X \hookrightarrow \mathcal{U}(n)$ où n est un entier strictement positif, est celle d'une variable qui peut prendre les valeurs entières k comprises entre 1 et n avec la loi de probabilité

$$P(X = k) = \frac{1}{n}.$$

Ses valeurs typiques sont, pour l'espérance et la variance, $m = \frac{n+1}{2}$ et $V = \frac{n^2-1}{12}$.

L'algorithme de simulation de cette loi est très simple : il repose sur la fonction `RANDOM(n)`, qui fournit un entier au hasard suivant la loi uniforme dans l'ensemble $\llbracket 0, n-1 \rrbracket$ (elle peut aussi s'écrire sous la forme `TRUNC(n*RANDOM)`).

Comme tous celles qui suivront, la procédure se limite à afficher à l'écran cent valeurs d'un nombre (ici un entier k , qui peut valoir 1, 2, ..., n). Ces cent valeurs définissent un vecteur aléatoire suivant la loi uniforme discrète ; si n est assez petit, ce nombre cent est assez grand pour que l'on puisse espérer que les entiers k affichés semblent tous aussi à peu près aussi probables les uns que les autres.

Exemple

```
PROCEDURE Uniforme_discrete(n : INTEGER) ;
    VAR a, i, k : INTEGER ;

    BEGIN
        RANDOMIZE ;
        FOR i := 1 TO 100 DO
            BEGIN
                a := RANDOM(n) ;
                k := 1+a ;
                WRITE(' ', k)
            END
        END ;
```

► Remarques

- Par exemple, pour $n = 5$, une expérience a donné 17 fois le nombre 1, 21 nombres 2, 22 nombres 3, 23 nombres 4 et 17 nombres 5, à comparer aux valeurs

$$P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = \frac{1}{5} = 0, 2.$$

Mais d'autres expériences peuvent être moins performantes : par exemple on a également relevé le vecteur (19, 20, 27, 17, 17) ; la probabilité d'un tel résultat moins satisfaisant n'est pas très faible, car le rapport $\frac{n}{100} = 0, 05$ est relativement encore un peu grand.

- Il est clair que l'on peut étendre cette procédure aux autres lois uniformes discrètes définies, non plus sur le segment $[1, n]$, mais sur un segment $[u, v]$ où u et v sont entiers en posant simplement $k := u + \text{RANDOM}(v-u+1)$, de valeurs typiques, pour l'espérance et la variance, $m = \frac{u+v}{2}$ et $V(X) = \frac{(v-u)(v-u+2)}{12}$.
- Un bon exercice pour voir si cette technique de simulation est bien comprise consiste, par exemple, à écrire une procédure simulant la variable aléatoire $\max(X, Y, Z)$, où X, Y et Z sont mutuellement indépendantes et suivent toutes trois une loi uniforme discrète sur $[1, n]$.

3.2 Loi de Bernoulli

Cette loi, notée $X \hookrightarrow \mathcal{B}(1, p)$ où p est un réel entre 0 et 1, est celle d'une variable qui peut prendre les valeurs entières 0 et 1 avec la loi de probabilité

$$P(X = 0) = 1 - p, \quad P(X = 1) = p.$$

Ses valeurs typiques sont, pour l'espérance et la variance, $m = p$ et $V = p(1 - p)$.

L'algorithme de simulation de cette loi est très simple : si A est une variable aléatoire uniforme à valeurs dans le segment $[0, 1[$, comme celles que fournit un recours au générateur de nombres au hasard de Pascal. On notera que, bien entendu, les événements $[A < p]$ et $[A \leq p]$ ont tous deux $F(p) = p$ comme probabilité puisque la probabilité de $\{p\}$ est nulle. Par suite la probabilité de l'événement $[A < p]$ est égale à p , ce qui était souhaité.

► **Remarque**

Cette remarque est fondamentale : elle sera sous-jacente à la justification poussée de toutes les procédures ici présentées.

Ici encore, ces cent valeurs définissent un vecteur aléatoire suivant la loi de Bernoulli et le nombre cent est assez grand pour que l'on puisse espérer trouver un nombre de 1 voisin de $100p$.

Une première simulation

Exemple

```
PROCEDURE Bernoulli1(p : REAL) ;
  VAR i, k : INTEGER ; a : REAL ;

  BEGIN
    RANDOMIZE ;
    FOR i := 1 TO 100 DO
      BEGIN
        a := RANDOM ;
        IF a < p THEN k := 1 ELSE k := 0 ;
        WRITE(' ', k)
      END
    END ;
```

► **Remarque**

Par exemple, pour $p = \frac{2}{3}$, une expérience a donné 35 fois le nombre 0 et 65 nombres 1, à comparer aux valeurs $P(X = 0) = \frac{1}{3} \approx 0,3333$ et $P(X = 1) = \frac{2}{3} \approx 0,6667$.

Une variante

La variante ci-dessous calcule toujours le même entier k mais, au lieu de l'écrire à l'écran, le comptabilise dans un tableau s . Avant de quitter le corps de la procédure, l'ensemble de ses valeurs est affiché à l'écran.

Il serait naturellement possible de déclarer ce tableau avec le mot réservé VAR dans la liste des paramètres de la procédure, ce qui permettrait d'utiliser ses valeurs même en dehors de son déroulement.

Exemple

```
PROCEDURE Bernoulli2(p : REAL) ;
  VAR i, k, m : INTEGER ; a : REAL ; s : ARRAY [0..1] OF INTEGER ;

  BEGIN
    RANDOMIZE ;
    FOR m := 0 TO 1 DO s[m] := 0 ;
    FOR i := 1 TO 100 DO
      BEGIN
        a := RANDOM ;
        IF a < p THEN k := 1 ELSE k := 0 ;
        s[k] := s[k]+1
      END ;
    FOR m := 0 TO 1 DO WRITE(' ', s[m])
  END ;
```

► **Remarque**

Bien entendu une même variante peut être apportée pour l'étude de toutes les autres lois, à condition d'élargir convenablement la taille du tableau en posant par exemple

```
s : ARRAY[0..10] OF INTEGER
```

3.3 Loi binomiale

Cette loi, notée $X \leftrightarrow \mathcal{B}(n, p)$, avec n entier strictement positif et p réel entre 0 et 1, est celle d'une variable qui peut prendre les valeurs entières k de 0 à n avec la loi de probabilité

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Ses valeurs typiques sont, pour l'espérance et la variance, $m = np$ et $V = np(1 - p)$.

► **Remarque**

Rappelons que les sommes finies de variables mutuellement indépendantes et suivant les lois $\mathcal{B}(n_i, p)$, de même p , suivent la loi $\mathcal{B}(\sum n_i, p)$.

L'algorithme de simulation de cette loi est simple : il s'agit simplement de la somme de n variables suivant la loi de Bernoulli $\mathcal{B}(1, p)$. Le code reprend donc simplement n fois de suite celui de Bernoulli que l'on vient de voir.

Exemple

```
PROCEDURE Binomiale(n : INTEGER ; p : REAL) ;
  VAR i, j, k : INTEGER ; a : REAL ;

  BEGIN
    RANDOMIZE ;
```

```

FOR i := 1 TO 100 DO
  BEGIN
    k := 0 ;
    FOR j := 1 TO n DO
      BEGIN
        a := RANDOM ;
        IF a < p THEN k := k+1
      END ;
    WRITE(' ', k) ;
  END
END ;

```

► **Remarques**

- On pourrait aussi imaginer une procédure récursive où $\mathcal{B}(n, p)$ appelle $\mathcal{B}(n - 1, p)$.
- Par exemple, pour $n = 5$ et $p = \frac{2}{3}$, une expérience a donné 0 fois le nombre 0, 5 nombres 1, 18 nombres 2, 37 nombres 3, 29 nombres 4 et 11 nombres 5, à comparer aux valeurs $P(X = 0) = \frac{1}{243} \approx 0,0041$, $P(X = 1) = \frac{10}{243} \approx 0,0412$, $P(X = 2) = \frac{40}{243} \approx 0,1646$, $P(X = 3) = \frac{80}{243} \approx 0,3292$, $P(X = 4) = \frac{80}{243} \approx 0,3292$ et $P(X = 5) = \frac{32}{243} \approx 0,1317$.

3.4 Loi géométrique

Cette loi, notée $X \hookrightarrow \mathcal{G}(p)$ où p est un réel entre 0 et 1, est celle d’une variable qui peut prendre les valeurs entières $k > 0$ avec la loi de probabilité

$$P(X = k) = p(1 - p)^{k-1}.$$

Ses valeurs typiques sont, pour l’espérance et la variance, $m = \frac{1}{p}$ et $V = \frac{1 - p}{p^2}$.

L’algorithme de simulation de cette loi repose sur le fait que la loi géométrique de paramètre p apparaît dans le cadre suivant : si l’on effectue une suite *a priori* infinie de tirages indépendants de Bernoulli, de même probabilité p , alors k est l’indice du premier tirage à donner un succès (on aurait $X = 0$ dans le cas, quasi impossible, où aucun tirage ne serait satisfaisant).

Exemple

```

PROCEDURE Geometrique(p : REAL) ;
  VAR i, k : INTEGER ; a, p : REAL ;

  BEGIN
    RANDOMIZE ;
    FOR i := 1 TO 100 DO
      BEGIN
        k := 0 ;
        REPEAT
          a := RANDOM ;

```

```

        k := k+1
    UNTIL a < p ;
    WRITE(' ', k)
END
END ;

```

► **Remarque**

Par exemple, pour $p = \frac{2}{3}$, une expérience a donné 71 fois le nombre 1, 20 nombres 2, 7 nombres 3, 1 nombre 4 et 1 nombre 5, à comparer aux valeurs $P(X = 1) = \frac{2}{3} \approx 0,6667$, $P(X = 2) = \frac{2}{9} \approx 0,2222$, $P(X = 3) = \frac{2}{27} \approx 0,0740$, $P(X = 4) = \frac{2}{81} \approx 0,0247$, $P(X = 5) = \frac{2}{243} \approx 0,0082$ et $P(X > 5) = \frac{1}{243} \approx 0,0041$.

3.5 Loi binomiale négative (de Pascal)

Cette loi, notée $X \hookrightarrow \mathcal{BN}(n, p)$ où $n > 0$ est un entier et p un réel entre 0 et 1, est celle d’une variable qui peut prendre les valeurs entières $k \geq n$ avec la loi de probabilité

$$P(X = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}.$$

Ses valeurs typiques sont, pour l’espérance et la variance, $m = \frac{n}{p}$ et $V = \frac{n(1-p)}{p^2}$.

Par définition, c’est la loi de la somme de n variables mutuellement indépendantes suivant la loi $\mathcal{G}(p)$, ce qui justifie par exemple aussitôt les valeurs typiques données ci-dessus.

La loi binomiale négative de paramètre (n, p) apparaît concrètement dans le cadre suivant : si l’on effectue une suite *a priori* infinie de tirages indépendants de Bernoulli de même probabilité p , alors k est l’indice du premier tirage à donner n succès.

Par suite les sommes finies de variables mutuellement indépendantes et suivant les lois $\mathcal{BN}(n_i, p)$, de même p , suivent la loi $\mathcal{BN}(\sum n_i, p)$.

La valeur de $P(X = k)$ et l’algorithme de simulation de cette loi découlent aussi presque trivialement de cette interprétation.

► **Remarques**

- Pour $n = 1$ on retrouve naturellement la loi géométrique.
- Cette loi et *a fortiori* sa simulation ne figurent pas au programme ; elles ont été placées ici comme un exemple, dont la maîtrise n’est naturellement pas exigible des candidats.

Exemple

```

PROCEDURE Binomiale_negative(n : INTEGER ; p : REAL) ;
    VAR i, j, k : INTEGER ; a, p : REAL ;

    BEGIN
        RANDOMIZE ;

```

```

FOR i := 1 TO 100 DO
  BEGIN
    k := 0 ;
    FOR j := 1 TO n DO
      REPEAT
        a := RANDOM ;
        k := k+1
      UNTIL a < p ;
      WRITE(' ', k)
    END
  END
END ;

```

► Remarques

- Il se peut théoriquement que l'un des boucles REPEAT ne s'arrête pas : mais on voit facilement que ce cas est quasi-impossible.
- Par exemple, pour $n = 5$ et $p = \frac{2}{3}$, une expérience a donné 16 fois le nombre 5, 25 nombres 6, 24 nombres 7, 18 nombres 8, 7 nombres 9, 6 nombres 10, 3 nombres 11 et 1 nombre 14, à comparer aux valeurs $P(X = 5) \approx 0,1317$, $P(X = 6) = P(X = 7) \approx 0,2195$, $P(X = 8) \approx 0,1707$, $P(X = 9) \approx 0,1138$, $P(X = 10) \approx 0,0683$, $P(X = 11) \approx 0,0379$, $P(X = 12) \approx 0,0199$, $P(X = 13) \approx 0,0099$, $P(X = 14) \approx 0,0048$, $P(X = 15) \approx 0,0022$ et $P(X = 16) \approx 0,0010$.

3.6 Loi hypergéométrique

Cette loi, notée $X \hookrightarrow \mathcal{H}(N, n, p)$ où N et n sont des entiers vérifiant $N \geq n > 0$ et p un rationnel entre 0 et 1 tel que Np soit entier, est celle d'une variable qui peut prendre les valeurs entières k comprises entre $\max(0, n - N(1 - p))$ et $\min(n, Np)$ avec la loi de probabilité

$$P(X = k) = \frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}}.$$

Ses valeurs typiques sont, pour l'espérance et la variance, $m = np$ et $V = np(1-p) \frac{N-n}{N-1}$.

Algorithme itératif

Un premier algorithme de simulation de cette loi repose sur le fait que la loi hypergéométrique de paramètres (N, n, p) apparaît dans le cadre suivant : si l'on place dans une urne N boules dont $b = Np$ sont blanches, et si l'on effectue n tirages sans remise, alors k est le nombre de boules blanches obtenues en fin de tirages.

Par conséquent, il suffit d'effectuer cent fois la tâche suivante : mettre les compteurs de la variable k à zéro, du nombre total de boules à N et le nombre b de boules blanches dans l'urne à Np . Ensuite, il n'y a plus qu'à appeler n fois le générateur de nombres au hasard et d'examiner si sa sortie aléatoire a est bien inférieure au rapport

du nombre b de boules blanches présentes dans l'urne après le tirage précédent, au nombre total t de boules restant dans l'urne. Si c'est exact, cela simulera la sortie d'une boule blanche ; il faut donc alors augmenter k d'une unité et diminuer à la fois b et t d'une unité. Sinon seul t est à modifier.

Dans le programme ci-dessous, nous avons du remplacer l'entier N par la lettre doublée "nn" puisque Pascal ne distingue malheureusement pas entre minuscules et majuscules.

► **Remarques**

- Cette simulation itérative n'est pas suggérée par le programme ; elle a été placée ici comme un exemple, dont la maîtrise n'est pas naturellement exigible des candidats.
- Il est facile de voir que, si p et k restent constants et si N tend vers l'infini en ne prenant que des valeurs entières telles que Np reste entier, la probabilité $P(X = k)$ a pour limite $\binom{n}{k} p^k (1 - p)^{n-k}$; en certain sens, la loi binomiale est donc une limite de la loi hypergéométrique.
Par contre, et en dépit du vocabulaire, il n'existe pas de lien de ce genre entre les lois géométrique et hypergéométrique : il s'agit en fait d'une allusion à des sommes de séries appelées *hypergéométriques* par Gauss qui interviennent dans la fonction génératrice de la loi, concept hors du programme de cette classe.

Exemple

```

PROCEDURE Hypergeometrique(nn, n : INTEGER ; p : REAL) ;
    VAR i, j, k, t : INTEGER ; a, b : REAL ;

    BEGIN
        RANDOMIZE ;
        FOR i := 1 TO 100 DO
            BEGIN
                k := 0 ; t := nn ; b := t*p ;
                FOR j := 1 TO n DO
                    BEGIN
                        a := RANDOM ;
                        IF t*a < b THEN
                            BEGIN
                                k := k+1 ; b := b-1
                            END ;
                        t := t-1 ;
                    END ;
                WRITE(' ', k)
            END
        END ;
    
```

► **Remarque**

Par exemple, pour $N = 9$, $n = 4$ et $p = \frac{2}{3}$, une expérience a donné 4 fois le nombre 1, 41 nombres 2, 45 nombres 3 et 10 nombres 4, à comparer aux valeurs $P(X = 1) = \frac{6}{126} \approx 0,0476$, $P(X = 2) = \frac{45}{126} \approx 0,3571$, $P(X = 3) = \frac{60}{126} \approx 0,4762$ et $P(X = 4) = \frac{15}{126} \approx 0,1190$.

Algorithme récursif

Cette seconde simulation, récursive cette fois et présentée sous la forme d'une fonction, est suggérée par le programme à titre d'exercice ; elle a été placée ici comme un exemple, dont la maîtrise n'est pas naturellement pas exigible des candidats.

Il repose sur le fait que, si $n = 0$, alors X est la variable certaine nulle. la récursion consiste à arriver de n à 0 par décrets successifs de 1. En effet, lors de la simulation du *dernier* tirage, il reste exactement $nn-n+1$ boules dans l'urne, dont $nn*p-k$ blanches si les tirages précédents en ont déjà sorti k . La probabilité de tirer encore une boule blanche est nulle si $k = nn*p$, et vaut $(nn*p-k)/(nn-n+1)$ sinon. Appelant alors la valeur aléatoire $a := \text{RANDOM}$, il suffit de suivre l'algorithme suivant : on augmentera de 1 la valeur de k si $(nn-n+1)*a < nn*p-k$, et on la gardera sinon.

► Remarque

Une fois dérécurivée — ce qui est ici presque trivial — elle donne la fonction au cœur de la procédure précédente ; les deux algorithmes sont donc fondamentalement les mêmes et ne diffèrent que par les lisibilité (subjectives) de leurs codes.

Exemple

```
FUNCTION Hyp(nn, n : INTEGER ; p : REAL) : INTEGER ;
  VAR k : INTEGER ; a : REAL ;

  BEGIN
    IF n = 0 THEN k := 0
      ELSE
        BEGIN
          a := RANDOM ;
          k := Hyp(nn, n-1, p) ;
          IF (nn-n+1)*a < nn*p-k THEN k := k+1
        END ;
      Hyp := k
    END ;
```

► Remarques

- Rappelons que recourir, par exemple cent fois, à cette fonction ne peut se faire qu'après avoir lancé un `RANDOMIZE` initial.
- Les expériences conduites avec l'un ou l'autre algorithme sont de fiabilité comparable.

3.7 Loi de Poisson

Cette loi, notée $X \leftrightarrow \mathcal{P}(h)$, de paramètre $h > 0$, est celle d'une variable qui peut prendre uniformément les valeurs entières $k \geq 0$ avec la loi de probabilité

$$P(X = k) = \exp(-h) \frac{h^k}{k!}.$$

Ses valeurs typiques sont, pour l'espérance et la variance, $m = h$ et $V = h$.

► **Remarque**

Rappelons que les sommes finies de variables mutuellement indépendantes et suivant les lois $\mathcal{P}(h_i)$ suivent la loi $\mathcal{P}(\sum h_i)$.

Simulation par fonctions indicatrices

La procédure ci-dessous de simulation de la loi de Poisson $\mathcal{P}(h)$, avec $h > 0$, repose sur la technique décrite plus haut pour une variable discrète à l'aide de combinaisons linéaires de fonctions indicatrices.

► **Remarque**

Cette simulation ne figure pas explicitement au programme; elle a été placée ici comme un exemple, dont la maîtrise n'est naturellement pas exigible des candidats.

Exemple

```

PROCEDURE Poisson1(h : REAL);
  VAR k, i : INTEGER; a, e, p, q : REAL;

  BEGIN
    RANDOMIZE;
    e := exp(-h);
    FOR i := 1 TO 100 DO
      BEGIN
        k := 0; p := e; q := p;
        a := RANDOM;
        WHILE a > q DO
          BEGIN
            k := k+1; p := p*h/k; q := q+p
          END;
        WRITE(' ', k)
      END
    END;
  END;

```

► **Remarque**

Par exemple, pour $h = 3$, une expérience a donné 4 fois le nombre 0, 18 nombres 1, 26 nombres 2, 18 nombres 3, 14 nombres 4, 12 nombres 5, 4 nombres 6, 2 nombres 7, 1 nombre 8 et 1 nombre 9, à comparer aux valeurs $P(X = 0) \approx 0,0498$, $P(X = 1) \approx 0,1494$, $P(X = 2) = P(X = 3) \approx 0,2240$, $P(X = 4) \approx 0,1680$, $P(X = 5) \approx 0,1008$, $P(X = 6) \approx 0,0504$, $P(X = 7) \approx 0,0216$, $P(X = 8) \approx 0,0081$, $P(X = 9) \approx 0,0027$, $P(X = 10) \approx 0,0008$ et $P(X > 10) \approx 0,0004$.

Une variante

La seconde procédure de simulation de la loi de Poisson $\mathcal{P}(h)$ repose sur le corollaire 3 dont il résulte que l'étude des sommes partielles d'une suite *a priori* infinie de variables aléatoires à densité mutuellement indépendantes suivant une loi exponentielle $\mathcal{E}(1)$ engendre une loi de Poisson. On verra plus bas en son temps comme simuler de telles variables, ce qui se traduit dans le programme ci-dessous par l'introduction de termes comme $x-\text{LN}(1-a)$.

► Remarque

Cette simulation ne figure pas au programme.

Exemple

```
PROCEDURE Poisson2(h : REAL) ;
  VAR k, i : INTEGER ; a, x : REAL ;

  BEGIN
    RANDOMIZE ;
    FOR i := 1 TO 100 DO
      BEGIN
        k := -1 ; x := 0 ;
        REPEAT
          a := RANDOM ;
          k := k+1 ; x := x-LN(1-a)
        UNTIL x > h ;
        WRITE(' ', k)
      END
    END ;
```

► Remarques

- Par exemple, pour $h = 3$, une expérience a donné 6 fois le nombre 0, 9 nombres 1, 27 nombres 2, 22 nombres 3, 21 nombres 4, 9 nombres 5, 4 nombres 6 et 2 nombres 7, à comparer aux valeurs données plus haut.
- Comme il sera expliqué lors de la simulation de la loi exponentielle, on aurait pu prendre ici la forme très légèrement plus simple $x := x - \text{LN}(a)$, voire même $x := x - \text{LN}(\text{RANDOM})$.

4. Simulations de lois réelles à densité

Les simulations des lois classiques à densité sont un peu moins simples que dans le cas discret, mais il est quand même possible d'en donner quelques unes qui restent faciles.

4.1 Loi uniforme à densité

Ces lois sont les plus simples des lois de probabilités à densité, celle-ci étant constante sur un certain intervalle, et nulle au-dehors.

Loi uniforme sur le segment $[0, 1]$

Cette loi, notée $X \leftrightarrow \mathcal{U}[0, 1]$, est celle d'une variable à densité qui peut prendre uniformément les valeurs réelles x comprises entre 0 et 1 avec la fonction de répartition définie sur $[0, 1]$ par

$$F(x) = P(X \leq x) = x$$

d'où la densité $f(x) = 1$.

Ses valeurs typiques sont, pour l'espérance et la variance, $m = \frac{1}{2}$ et $V = \frac{1}{12}$.

L'algorithme de simulation de cette loi, extrêmement simple, repose sur l'idée de tirer successivement, grâce à la fonction `RANDOM`, des nombres au hasard x , puisqu'ils suivent eux-mêmes une loi uniforme sur $[0, 1[$.

Exemple

```
PROCEDURE Uniforme_a_densite1;
    VAR i : INTEGER; a, x : REAL;

    BEGIN
        RANDOMIZE;
        FOR i := 1 TO 100 DO
            BEGIN
                a := RANDOM; x := a;
                WRITE(' ', TRUNC(1000*x))
            END
        END;
    END;
```

► Remarques

- Le plus simple serait d'écrire dans le programme `WRITE(' ', x)`, comme dans les programmes précédents, mais la lisibilité des résultats serait mauvaise, d'où l'idée de les multiplier par 1000 et d'en prendre la partie entière.

On obtient ainsi des nombres entiers généralement à trois chiffres - par exemple 631 -, qu'à la lecture on rétablit mentalement aussitôt comme étant la partie décimale du nombre $x = 0,631$ qui reste compris entre 0 et 1.

Cette complication résulte du choix fait par Pascal d'utiliser systématiquement la notation scientifique avec un exposant noté par la lettre E, qu'il convient de contourner pour que les sorties apparaissent ayant à peu près un même nombre de chiffres, restreint pour les commodités de lecture à vue.

Il existe d'autres solutions à ce problème de lisibilité à l'écran : on peut par exemple utiliser le « formatage » préconisé par le Pascal standard de Niklaus Wirth : ainsi l'instruction `WRITE(' ', x :4 :3)` donne directement 0,631 puisque cette chaîne de nombres contient 4 nombres dont 3 après la virgule. On pourrait également écrire `WRITE(' 0,' TRUNC(1000*x))`, ce qui aurait le même effet.

Tout cela vaut pour les programmes de simulation de variables à images non nécessairement entières.

- Un nombre x tel que `TRUNC(1000*x)=n` doit être compté, en toute rigueur, parmi ceux qui vérifient $X \leq \frac{n+1}{1000}$ et peut-être pas $X \leq \frac{n}{1000}$, mais la différence est minime ; elle ne le serait pas si l'on prendrait seulement `TRUNC(x)` ; dans ce cas il vaudrait mieux écrire à l'écran `1+TRUNC(x)` (voir l'exemple de la loi de Cauchy).
- L'exploitation des résultats donnés par cette procédure est un peu moins simple que pour les variables aléatoires discrètes. Ici, il faut se donner des valeurs $0 \leq t \leq 1$, compter les pourcentages des résultats x inférieurs ou égaux à t , et les comparer avec les valeurs $F(t) = t$.
- Par exemple, une expérience a donné 100 nombres $x \leq 1$, 49 nombres $x \leq 0,5$ et 10 nombres $x \leq 0,1$, à comparer aux valeurs $F(1) = 1$, $F(0,5) = 0,5$ et $F(0,1) = 0,1$.

Loi uniforme sur un segment $[u, v]$

Voici une extension concernant une variable suivant une loi uniforme à densité sur un segment $[u, v]$, avec $u < v$.

Elle peut se ramener systématiquement à la loi uniforme $\mathcal{U}[0, 1]$ puisque, si $X \leftrightarrow \mathcal{U}[0, 1]$, alors $u + (v - u)X \leftrightarrow \mathcal{U}[u, v]$, ce qui s'écrit sous les formes

équivalentes

$$\forall x \in [u, v], \quad P(u + (v - u)X \leq x) = P\left(X \leq \frac{x - u}{v - u}\right) = \frac{x - u}{v - u},$$

$$X \hookrightarrow \mathcal{U}[0, 1] \iff u + (v - u)X \hookrightarrow \mathcal{U}[u, v],$$

$$Y \hookrightarrow \mathcal{U}[u, v] \iff \frac{1}{v - u}(X - u) \hookrightarrow \mathcal{U}[0, 1].$$

Ses valeurs typiques sont, pour l'espérance et la variance, $m = \frac{u + v}{2}$ et $V = \frac{(v - u)^2}{12}$.

L'algorithme de simulation de cette loi est évident d'après ce qui précède.

► **Remarque**

Dans le cas particulier où u et v sont entiers, il est tentant de comparer les lois uniformes à densité et discrètes sur $[u, v]$. Si elles ont même espérance mathématique $E(X) = \frac{u+v}{2}$, ce qui est évident par symétrie, il n'en est pas de même de la variance.

Pour la loi discrète, elle vaut $V(X) = \frac{(v - u)(v - u + 2)}{12}$, alors qu'elle n'est que $V(X) = \frac{(v - u)^2}{12}$ pour la loi à densité.

Exemple

```
PROCEDURE Uniforme_a_densite2(VAR u, v : REAL) ;
    VAR i : INTEGER ; a, x : REAL ;

    BEGIN
        RANDOMIZE ;
        FOR i := 1 TO 100 DO
            BEGIN
                a := RANDOM ;
                x := u+(v-u)*a ;
                WRITE(' ', TRUNC(1000*x))
            END
        END ;
```

► **Remarques**

- Le plus simple serait d'écrire dans le programme `WRITE(' ', x)`, comme dans les programmes précédents, mais la lisibilité des résultats serait mauvaise, d'où l'idée de les multiplier par 1000 et d'en prendre la partie entière.
- On pourrait gagner une variable en supprimant `a` deux fois remplacée par `x`, mais ce serait sans doute au prix de la lisibilité de l'algorithme.
- Prenons l'exemple $u = 10, v = 60$. L'exploitation des résultats donnés par cette procédure est un peu moins simple que pour les variables aléatoires discrètes. Ici, il faut se donner des valeurs $10 \leq t \leq 60$, compter les pourcentages des résultats x inférieurs ou égaux à t , à comparer aux valeurs $F(t) = \frac{t - 10}{50}$ (égalité valide sur $[10, 60]$).
- Par exemple, une expérience a donné 100 nombres $x \leq 60$, 47 nombres $x \leq 35$ et 9 nombres $x \leq 15$, à comparer aux valeurs $F(60) = 1, F(35) = 0,5$ et $F(15) = 0,1$.

- Un bon exercice pour voir si cette technique de simulation est bien comprise consiste, par exemple, à écrire une procédure simulant la variable aléatoire $\min(X, Y, Z)$, où X, Y et Z sont mutuellement indépendantes et suivent toutes trois une loi uniforme sur $[u, v]$.

4.2 Loi exponentielle

Cette loi, notée $X \leftrightarrow \mathcal{E}(c)$, de paramètre $c > 0$, est celle d'une variable à densité qui peut prendre les valeurs réelles positives x avec la fonction de répartition définie sur \mathbb{R}_+ par

$$F(x) = P(X \leq x) = 1 - \exp(-cx)$$

d'où la densité $f(x) = c \exp(-cx)$.

Ses valeurs typiques sont, pour l'espérance et la variance, $m = \frac{1}{c}$ et $V = \frac{1}{c^2}$.

L'algorithme de simulation de la loi exponentielle $\mathcal{E}(h)$ repose sur la technique dite d'*anamorphose*.

L'idée est de tirer successivement, grâce à la fonction `RANDOM`, des nombres au hasard a suivant une loi uniforme sur $[0, 1[$, puis d'en déduire des nombres $x = -\frac{1}{c} \ln(1 - a)$.

Ces valeurs suivent approximativement la loi exponentielle de paramètre h puisque $y = F(x) = 1 - \exp(-cx)$ implique $x = F^{-1}(y) = -\frac{1}{c} \ln(1 - y)$.

► Remarque

Cette simulation ne figure explicitement qu'au programme de la voie économique, avec l'indication ci-dessus, sous la forme d'exercice suggéré; elle a été placée ici comme un exemple, dont la maîtrise n'est naturellement pas exigible des candidats.

Exemple

```
PROCEDURE Exponentielle(c : REAL);
  VAR i : INTEGER; a, x : REAL;

  BEGIN
    RANDOMIZE;
    FOR i := 1 TO 100 DO
      BEGIN
        a := RANDOM;
        x := - LN(1-a)/c;
        WRITE(' ', TRUNC(1000*x))
      END
    END;
  END;
```

► Remarques

- Le plus simple serait d'écrire dans le programme `WRITE(' ', x)`, comme dans des programmes précédents, mais la lisibilité des résultats serait mauvaise, d'où l'idée de les multiplier par 1000 et d'en prendre la partie entière.
- Si U est une variable aléatoire de loi uniforme sur $[0, 1]$, il en va évidemment de même pour $1 - U$; on aurait donc pu prendre la forme légèrement plus simple $x = -\frac{1}{c} \ln a$.

- L'exploitation des résultats donnés par cette procédure est un peu moins simple que pour les variables aléatoires discrètes. Ici, il faut se donner des valeurs $t > 0$, compter les pourcentages des résultats x inférieurs ou égaux à t , et les comparer avec les valeurs $F(t) = 1 - \exp(-ct)$.
- Par exemple, pour $c = 3$, une expérience a donné 95 nombres $x \leq 1$, 74 nombres $x \leq 0,5$ et 26 nombres $x \leq 0,1$, à comparer aux valeurs $F(1) = 1 - \exp(-3) \approx 0,95$, $F(0,5) = 1 - \exp(-1,5) \approx 0,78$ et $F(0,1) = 1 - \exp(-0,3) \approx 0,26$.
- On peut également se ramener systématiquement à la loi exponentielle $\mathcal{E}(1)$ puisque, si X a pour loi $\mathcal{E}(c)$, alors cX a pour loi $\mathcal{E}(1)$, ce qui s'écrit sous les formes équivalentes

$$\forall x \geq 0, \quad P(cX \leq x) = P\left(X \leq \frac{x}{c}\right) = 1 - \exp\left(-c \frac{x}{c}\right) = 1 - \exp(-x),$$

$$X \leftrightarrow \mathcal{E}(c) \iff cX \leftrightarrow \mathcal{E}(1),$$

$$Y \leftrightarrow \mathcal{E}(1) \iff \frac{1}{c}Y \leftrightarrow \mathcal{E}(c).$$

4.3 Loi de Cauchy

Cette loi est celle d'une variable à densité qui peut prendre les valeurs réelles x avec la fonction de répartition définie sur \mathbb{R} par

$$F(x) = P(X \leq x) = \frac{1}{2} + \frac{\text{Arctan } x}{\pi}$$

d'où la densité $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$.

Elle ne possède ni espérance (bien que sa densité soit paire), ni *a fortiori* de variance. L'algorithme de simulation de la loi de Cauchy repose sur la technique dite d'*anamorphose*.

L'idée est de tirer successivement, grâce à la fonction RANDOM, des nombres au hasard a suivant une loi uniforme sur $[0, 1[$, puis d'en déduire des nombres $x = -\cot \pi a$.

Ces valeurs suivent approximativement la loi de Cauchy puisque $y = F(x) = \frac{1}{2} + \frac{\text{Arctan } x}{\pi}$ implique $x = F^{-1}(y) = -\cot \pi y$.

► **Remarque**

Cette loi et *a fortiori* sa simulation ne figurent pas au programme; elles ont été placées ici comme un exemple, dont la maîtrise n'est naturellement pas exigible des candidats.

Exemple

```
PROCEDURE Cauchy ;
  VAR i, k : INTEGER ; a, pi, x : REAL ;

  BEGIN
    RANDOMIZE ;
    pi := 4*arctan(1) ;
    FOR i := 1 to 100 DO
      BEGIN
        a := RANDOM ;
        x := -COS(pi*a)/SIN(pi*a) ;
        k := TRUNC(x) ;
```



```

        IF x > 0 then k := k+1 ;
        WRITE(' ', k)
    END
END ;

```

► **Remarques**

- Le plus simple serait d'écrire dans le programme `WRITE(' ', x)`, comme dans des programmes précédents, mais la lisibilité des résultats serait mauvaise, d'où l'idée d'en prendre les parties entières augmentées de 1 puisque les nombres inférieurs à un entier k ont, sauf k lui-même, des parties entières inférieures ou égales à $k - 1$.
- L'exploitation des résultats donnés par cette procédure est un peu moins simple que pour les variables aléatoires discrètes. Ici, il faut se donner des valeurs $t > 0$, compter les pourcentages des résultats x inférieurs ou égaux à t , et les comparer avec les valeurs $F(t) = \frac{1}{2} + \frac{\text{Arctan } t}{\pi}$.
- Par exemple, une expérience a donné 17 nombres $x \leq -2$ (y compris l'entier surprenant -144 , correspondant à une valeur a très proche de 0), 25 nombres $x \leq -1$, 50 nombres $x \leq 0$, 73 nombres $x \leq 1$, 85 nombres $x \leq 2$, 88 nombres $x \leq 3$, 89 nombres $x \leq 4$, 91 nombres $x \leq 5$ et 92 nombres $x \leq 6$, à comparer aux valeurs $F(-2) \approx 0,15$, $F(-1) = \frac{1}{4} = 0,25$, $F(0) = \frac{1}{2} = 0,5$, $F(1) = \frac{3}{4} = 0,75$, $F(2) \approx 0,85$, $F(3) \approx 0,90$, $F(4) \approx 0,92$, $F(5) \approx 0,94$ et $F(6) \approx 0,95$.

4.4 Loi de Laplace

Cette loi est celle d'une variable à densité admettant comme paramètre un réel $c > 0$ qui peut prendre les valeurs réelles x avec la fonction de répartition définie sur \mathbb{R} par

$$F(x) = P(X \leq x) = \begin{cases} \frac{1}{2} \exp(cx) & \text{si } x < 0, \\ 1 - \frac{1}{2} \exp(-cx) & \text{sinon} \end{cases}$$

d'où la densité $f(x) = \frac{c}{2} \exp(-c|x|)$.

Ses valeurs typiques sont, pour l'espérance et la variance, $m = 0$ et $V = \frac{2}{c^2}$.

L'algorithme de simulation de la loi de Laplace repose sur la technique dite d'*anamorphose*.

L'idée est de tirer successivement, grâce à la fonction `RANDOM`, des nombres au hasard a suivant une loi uniforme sur $[0, 1]$, puis d'en déduire des nombres $x = \frac{1}{c} \ln 2a$ si $a < \frac{1}{2}$ et $x = -\frac{1}{c} \ln 2(1 - a)$ sinon.

Ces valeurs suivent approximativement la loi de Laplace puisque $y = F(x)$ implique $x = F^{-1}(y)$, avec $F^{-1}(y) = \frac{1}{c} \ln 2y < 0$ si $y < \frac{1}{2}$ et $F^{-1}(y) = -\frac{1}{c} \ln 2(1 - y) \geq 0$ sinon.

Une petite astuce algorithmique permet d'unifier légèrement les deux cas, on la lira ci-dessous.

► **Remarque**

Cette loi et *a fortiori* sa simulation ne figurent pas au programme ; elles ont été placées ici comme un exemple, dont la maîtrise n'est naturellement pas exigible des candidats.

Exemple

```

PROCEDURE Laplace(c : REAL) ;
  VAR i, k : INTEGER ; a, x : REAL ;

  BEGIN
    RANDOMIZE ;
    FOR i := 1 to 100 DO
      BEGIN
        a := RANDOM ;
        IF 2*a < 1 THEN x := LN(2*a)/c
          ELSE x := -LN(2-2*a)/c ;
        k := TRUNC(10*x) ;
        IF x > 0 then k := k+1 ;
        WRITE(' ', k)
      END
    END ;
  END ;

```

► Remarques

- Le plus simple serait d'écrire dans le programme `WRITE(' ', x)`, comme dans des programmes précédents, mais la lisibilité des résultats serait mauvaise, d'où l'idée d'en prendre les parties entières augmentées de 1 puisque les nombres inférieurs à un entier k ont, sauf k lui-même, des parties entières inférieures ou égales à $k - 1$.
- L'exploitation des résultats donnés par cette procédure est un peu moins simple que pour les variables aléatoires discrètes. Ici, il faut se donner des valeurs $t > 0$, compter les pourcentages des résultats x inférieurs ou égaux à t , et les comparer avec les valeurs $F(t)$.
- Par exemple, pour $c = 3$, une expérience a donné 3 nombres $x \leq -1$ (y compris un nombre égal à $-1,5$), 8 nombres $x \leq -0,5$, 23 nombres $x \leq -0,2$, 48 nombres $x \leq 0$, 69 nombres $x \leq 0,2$, 86 nombres $x \leq 0,5$, 95 nombres $x \leq 1$, 99 nombres $x \leq 1,5$ et 100 nombres $x \leq 1,6$, à comparer aux valeurs $F(-1,5) \approx 0,006$, $F(-1) \approx 0,025$, $F(-0,5) \approx 0,112$, $F(-0,2) \approx 0,274$, $F(0) = 0,5$, $F(0,2) \approx 0,726$, $F(0,5) \approx 0,888$, $F(1) \approx 0,975$, $F(1,5) \approx 0,994$ et $F(1,6) \approx 0,996$.

4.5 Loi gamma

Les différentes lois gamma jouent un rôle essentiel en statistique mathématique et appliquée.

Loi gamma standard

Cette loi, notée $X \hookrightarrow \gamma(n)$, de paramètre $n > 0$ supposé ici entier (voir le chapitre sur les variables à densité pour avoir la forme générale), est celle d'une variable à densité qui peut prendre les valeurs réelles positives x avec la fonction de répartition définie sur \mathbb{R}_+ par

$$F(x) = P(X \leq x) = \frac{1}{(n-1)!} \int_0^x t^{n-1} e^{-t} dt$$

d'où la densité $f(x) = \frac{1}{(n-1)!} t^{n-1} e^{-t}$.

Ses valeurs typiques sont, pour l'espérance et la variance, $m = n$ et $V = n$.

Nous ne donnerons effectivement d'algorithme de simulation de cette loi que pour n entier strictement positif (loi d'Erlang). Elle repose sur le théorème selon laquelle la somme de n variables aléatoires mutuellement indépendantes et suivant la loi exponentielle $\mathcal{E}(1)$ est une variable suivant la loi réduite $\gamma(n)$.

Exemple

```

PROCEDURE Gamma_Reduite(n : INTEGER) ;
    VAR i, j, k : INTEGER ; a, x : REAL ;

    BEGIN
        RANDOMIZE ;
        FOR i := 1 TO 100 DO
            BEGIN
                x := 0 ;
                FOR m := 1 TO n DO
                    BEGIN
                        a := RANDOM ; x := x-LN(1-a) ;
                    END ;
                k := 1+TRUNC(x) ;
                WRITE(' ', k)
            END
        END ;
    END ;

```

► **Remarque**

Pour $n = 6$, on peut commencer par calculer, à l'aide d'un programme Pascal par exemple, quelques valeurs de la fonction de répartition $F(x) = P(X \leq x)$ d'une loi gamma de paramètre $n = 6$. Posant $u_0 = 0$ et $u_n = F(n) - F(n-1)$ sinon, on obtient ainsi $u_1 \approx 0,001, u_2 \approx 0,016, u_3 \approx 0,067, u_4 \approx 0,131, u_5 \approx 0,169, u_6 \approx 0,170, u_7 \approx 0,145, u_8 \approx 0,110, u_9 \approx 0,075, u_{10} \approx 0,049, u_{11} \approx 0,029, u_{12} \approx 0,018, u_{13} \approx 0,009, u_{14} \approx 0,005, u_{15} \approx 0,003, u_{16} \approx 0,002, u_{17} \approx 0,000, u_{18} \approx 0,001$ et $0,000$ pour des indices supérieurs.

Ces nombres sont à rapprocher, par exemple, des résultats d'une expérience qui a donné 0 fois le nombre 0, 0 nombre 1, 1 nombre 2, 8 nombres 3, 10 nombres 4, 24 nombres 5, 16 nombres 6, 15 nombres 7, 5 nombres 8, 7 nombres 9, 6 nombres 10, 3 nombres 11, 2 nombres 12, 1 nombre 13, 1 nombre 14 et 1 nombre 16 (aucun au delà).

Loi gamma générale

Cette loi, notée $X \hookrightarrow \Gamma(b, n)$, de paramètre (b, n) où b (paramètre d'échelle) et n sont des réels strictement positifs avec, dans ce paragraphe, n entier (loi d'Erlang), est celle d'une variable à densité qui peut prendre les valeurs réelles positives x avec la fonction de répartition définie sur \mathbb{R}_+ par

$$F(x) = P(X \leq x) = \frac{b^n}{\Gamma(n)} \int_0^x t^{n-1} e^{-bt} dt$$

d'où la densité $f(x) = \frac{b^n}{\Gamma(n)} t^{n-1} e^{-bt}$.

Ses valeurs typiques sont, pour l'espérance et la variance, $m = \frac{n}{b}$ et $V = \frac{n}{b^2}$.

► **Remarques**

- Rappelons que les sommes finies de variables mutuellement indépendantes et suivant les lois $\Gamma(b, n_i)$, de même paramètre d'échelle b , suivent la loi $\Gamma(b, \sum n_i)$.
- Outre la loi standard, cette loi contient des cas particuliers intéressants : la loi exponentielle pour $n = 1$ et les lois dites de χ_k^2 à k degrés de liberté (lire : khi-k deux), pour $b = \frac{1}{2}$ et $n = \frac{k}{2}$ demi-entier. Ces dernières ne sont pas au programme ; elles jouent un rôle essentiel dans l'étude des tests statistiques. Leurs valeurs typiques sont, pour l'espérance et la variance, $m = k$ et $V = 2k$.

Le passage du cas standard à la loi générale s'obtient simplement en changeant $\mathcal{E}(1)$ en $\mathcal{E}\left(\frac{1}{b}\right)$, donc l'instruction $x := x\text{-LN}(1-a)$ en $x := x\text{-b*LN}(1-a)$.

Nous ne donnerons pas d'algorithme détaillé pour la simulation de cette loi, puisqu'il est clair que si X a pour loi $\Gamma(b, n)$, alors $Y = bX$ a pour loi $\gamma(n)$, ce qui s'écrit, grâce au changement de variables $u = bt$

$$\forall x \geq 0, \quad P(Y \leq x) = P\left(X \leq \frac{x}{b}\right) = \frac{b^n}{\Gamma(n)} \int_0^{x/b} t^{n-1} e^{-bt} dt = \frac{1}{\Gamma(n)} \int_0^x u^{n-1} e^{-u} du.$$

Il en résulte que

$$\begin{aligned} X \hookrightarrow \Gamma(b, n) &\iff bX \hookrightarrow \gamma(n), \\ Y \hookrightarrow \gamma(n) &\iff \frac{1}{b} Y \hookrightarrow \Gamma(b, n) \end{aligned}$$

ainsi que, en notant f et g les densités respectives de X et Y ,

$$f(x) = \frac{b^n}{\Gamma(n)} x^{n-1} e^{-bx} = \frac{b}{\Gamma(n)} (bx)^{n-1} e^{-bx} = b g(bx).$$

4.6 Loi normale (de Laplace-Gauss)

La simulation des lois normales est le dernier exemple de cette longue série. Son importance ne pourrait évidemment être surestimée.

Loi normale centrée réduite

Cette loi, notée $X \hookrightarrow \mathcal{N}(0, 1)$, est celle d'une variable à densité qui peut prendre les valeurs réelles x avec la fonction de répartition définie sur \mathbb{R} par

$$F(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt$$

d'où la densité $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$.

Ses valeurs typiques sont, pour l'espérance et la variance, $m = 0$ et $V = 1$.

L'algorithme de simulation de la loi normale centrée réduite repose sur le *théorème de la limite centrée*, selon laquelle la moyenne centrée réduite d'une suite (X_n) de variables aléatoires mutuellement indépendantes suivant une même loi uniforme

converge en loi vers une variable X suivant la loi normale centrée réduite

$$\sqrt{\frac{12}{n}} \sum_{j=1}^n X_j - \sqrt{3n} \xrightarrow{\mathcal{L}} X.$$

L'idée est de tirer successivement, grâce à la fonction RANDOM, n nombres au hasard a suivant une loi uniforme sur $[0, 1[$.

Pour des raisons de commodité, on peut se limiter à des entiers n multiples de 12 ; le programme ci-dessous prend $n = 12$.

► **Remarques**

- Depuis 1958, on remplace cette méthode de calcul par une autre due à G.E.P. Box et M.E. Muller, sur les détails de laquelle nous ne pouvons pas entrer ici.
- La simulation de cette loi figure au programme sous la forme d'un exemple, dont la maîtrise n'est naturellement pas exigible des candidats, suivant la loi de la limite centrée.

Exemple

```

PROCEDURE Normale_centree_reduite ;
    VAR i, j, k : INTEGER ; a, x : REAL ;

    BEGIN
        RANDOMIZE ;
        FOR i := 1 TO 100 DO
            BEGIN
                x := -6 ;
                FOR j := 1 TO 12 DO
                    BEGIN
                        a := RANDOM ;
                        x := x+a
                    END ;
                k := TRUNC(10*x) ;
                IF x > 0 THEN k := k+1 ;
                WRITE(' ', k)
            END
        END ;
    
```

► **Remarques**

- Le plus simple serait d'écrire dans le programme WRITE(' ', x), comme dans des programmes précédents, mais la lisibilité des résultats serait mauvaise, d'où l'idée d'en prendre les parties entières de $10*x$ augmentées de 1 puisque les nombres inférieurs à un entier k ont, sauf k lui-même, des parties entières inférieures ou égales à $k - 1$.
- L'exploitation des résultats donnés par cette procédure est un peu moins simple que pour les variables aléatoires discrètes. Ici, il faut se donner des valeurs t , compter les pourcentages des résultats x inférieurs ou égaux à t , et les comparer avec les valeurs $F(t)$ trouvées par exemple dans une table ou délivrée par un logiciel adéquat.
- Par exemple, une expérience a donné exclusivement des entiers entre -23 et 20 et des fréquences cumulées égales à 0 pour $t = -2, 5$, à 0,02 pour $t = -2$, à 0,09 pour $t = -1, 5$, à 0,19 pour $t = -1$, à 0,26 pour $t = -0, 7$, à 0,3 pour $t = -0, 5$, à 0,34 pour $t = -0, 4$, à 0,38 pour $t = -0, 3$, à 0,4 pour $t = -0, 2$, à 0,46 pour $t = -0, 1$, à 0,49 pour $t = 0$, à 0,53 pour $t = 0, 1$, à 0,57 pour $t = 0, 2$, à 0,59 pour $t = 0, 3$, à 0,64

pour $t = 0, 4$, à $0,66$ pour $t = 0, 5$, à $0,77$ pour $t = 0, 7$, à $0,85$ pour $t = 1$, à $0,97$ pour $t = 1, 5$ et à 1 pour $t = 2$ et $t = 2, 5$.

Elles sont à comparer aux valeurs $F(-2, 5) \approx 0, 01$, $F(-2) \approx 0, 02$, $F(-1, 5) \approx 0, 07$, $F(-1) \approx 0, 16$, $F(-0, 7) \approx 0, 24$, $F(-0, 5) \approx 0, 31$, $F(-0, 4) \approx 0, 34$, $F(-0, 3) \approx 0, 38$, $F(-0, 2) \approx 0, 42$, $F(-0, 1) \approx 0, 46$, $F(0) = 0, 5$, $F(0, 1) \approx 0, 54$, $F(0, 2) \approx 0, 58$, $F(0, 3) \approx 0, 62$, $F(0, 4) \approx 0, 66$, $F(0, 5) \approx 0, 69$, $F(0, 7) \approx 0, 76$, $F(1) \approx 0, 84$, $F(1, 5) \approx 0, 93$, $F(2) \approx 0, 98$ et $F(2, 5) \approx 0, 99$.

On voit que l'accord est satisfaisant ; il serait nettement plus affiné si l'on utilisait une suite de 108 variables aléatoires à densité X_j suivant la loi uniforme au lieu de 6 ; la formule à utiliser serait ici

$$P(X = k) \approx \frac{1}{3} \sum_{j=1}^{108} X_j - 18.$$

Loi normale générale

Cette loi, notée $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$, parfois notée $X \hookrightarrow \mathcal{N}(\mu, \sigma)$ avec $\sigma > 0$, est celle d'une variable à densité qui peut prendre les valeurs réelles x avec la fonction de répartition définie sur \mathbb{R} par

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

d'où la densité $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.

Ses valeurs typiques sont, pour l'espérance et la variance, $m = \mu$ et $V = \sigma^2$.

► **Remarque**

Rappelons que les sommes finies de variables mutuellement indépendantes et suivant les lois $\mathcal{N}(\mu_i, \sigma_i^2)$ suivent la loi $\mathcal{N}(\Sigma \mu_i, \Sigma \sigma_i^2)$.

Notons un instant F_0 la loi de répartition du cas particulier d'une variable suivant la loi normale centrée réduite. Un changement de variable immédiat dans l'intégrale montre que

$$F(\mu + \sigma x) = F_0(x).$$

Nous ne donnerons pas d'algorithme détaillé pour la simulation de cette loi, puisque, si X a pour loi $\mathcal{N}(\mu, \sigma^2)$, alors $\frac{1}{\sigma}(X - \mu)$ a pour loi $\mathcal{N}(0, 1)$, ce qui s'écrit sous les formes équivalentes

$$\forall x, \quad P\left(\frac{1}{\sigma}(X - \mu) \leq x\right) = P(X \leq \mu + \sigma x) = F(\mu + \sigma x),$$

$$X \hookrightarrow \mathcal{N}(\mu, \sigma^2) \iff \frac{1}{\sigma}(X - \mu) \hookrightarrow \mathcal{N}(0, 1),$$

$$Y \hookrightarrow \mathcal{N}(0, 1) \iff \mu + \sigma Y \hookrightarrow \mathcal{N}(\mu, \sigma^2).$$

► **Remarques**

- On pourra pourtant se reporter à la fin du chapitre, où une procédure générale de simulation d'une loi normale générale est incluse dans un programme d'estimation de l'espérance d'une variable normale d'écart-type connu.
- À la place de variables uniformes à densité sur $[0, 1]$, on aurait pu utiliser des variables de Bernoulli, binomiales etc.

Par exemple : pour n assez grand, le théorème de la limite centrée dit que si X_n est une variable aléatoire suivant la loi $\gamma(n)$, alors

$$\frac{X_n - n}{\sqrt{n}}$$

approche correctement la loi normale centrée réduite. Or nous disposons effectivement d'une simulation pour un tel X_n , ce qui fournit bien un autre algorithme satisfaisant pour la loi normale.

5. Estimation

Rappelons que l'outil fondamental des approches pédagogiques sur l'estimation et la simulation est le recours aux fonctions `RANDOM` et `RANDOM(n)` de Pascal.

Si l'on note $a := \text{RANDOM}$ le nombre aléatoire renvoyé par le logiciel à un tel appel, et si l'on veut lancer une instruction donnée I avec une probabilité p ou ne rien faire en cas contraire avec la probabilité $1 - p$, il suffit de tester si a appartient à l'intervalle $[0, p[$: si la réponse est *oui*, on met I en œuvre, sinon on reste en l'état. En effet, la probabilité pour que, par exemple, $a < p$ est justement égale à p .

Tel est le principe général ; rien n'interdit de compliquer la règle ci-dessus de manière évidente.

Bien entendu, dans le cadre de leur utilisation, des appels à `RANDOM` ne doivent être mises en jeu qu'après qu'un appel à `RANDOMIZE` ait été effectué en début de programme.

5.1 Estimation d'un nombre inconnu

Il s'agit ici de l'estimation d'un nombre inconnu p , que l'on suppose pour simplifier compris entre 0 et 1 (il pourra donc être considéré comme une probabilité). Il peut être tiré au hasard par Turbo Pascal ou, comme dans le programme écrit ci-dessous, être donné au départ.

Le principe est simple : le logiciel simule grâce à `RANDOM` un n -échantillon d'épreuves de Bernoulli de paramètre p (ici l'on prendra $n = 1000$). La moyenne des n résultats, tous égaux à 0 ou 1, est une estimation notée e de la valeur de p . Le tout est repris à l'identique k fois de suite (ici l'on prendra $k = 10$ et les différentes valeurs de e sont rassemblées dans un tableau final publié à l'écran). Tous ces nombres sont évidemment modifiables à volonté.

La liste des $e[m]$ où m varie de 1 à k est publiée en fin d'exécution du programme, qui donne ensuite la moyenne de ces $e[m]$ à côté de la « vraie » valeur de p .

On notera dans le programme ci-dessous le « format » des sorties des estimations de p : écrire `WRITE (x:4:3)` signifie que la valeur de la variable s doit être écrite à l'aide de 4 chiffres dont 3 après la virgule.

```
PROGRAM Estimp ;
VAR h, i, k, m, n : INTEGER ; a, p, s : REAL ;
    e : ARRAY [1..20] OF REAL ;
```

```

BEGIN
  RANDOMIZE ;
  p := 2/3 ; h := 10 ; n := 1000 ; s := 0 ;
  FOR m := 1 TO h DO
    BEGIN
      k := 0 ;
      FOR i := 1 TO n DO
        BEGIN
          a := RANDOM ;
          IF a < p THEN k := k+1 ;
        END ;
        e[m] := k/n ; s := s+e[m] ;
        WRITE ( ' ', e[m]:4:3)
      END ;
      s := s/h ;
      WRITELN ; WRITELN (s:4:3, ' ', p:4:3)
    END.

```

Chacune des 10 estimations est indexée par m ; alors $e[m]$ est la « réussite moyenne » parmi les $n = 1000$ appels à `RANDOM` (une réussite est caractérisée par le fait qu'un compteur k augmente d'une unité).

Ces moyennes, en principe proches de p , sont elles-mêmes ajoutées entre elles avant une division finale par $h = 10$ qui donne une « moyenne des moyennes » notée s et publiée à l'écran.

Par exemple, une expérience a donné les 10 nombres successifs $e[m] = 0,649 ; 0,671 ; 0,664 ; 0,692 ; 0,656 ; 0,663 ; 0,689 ; 0,679 ; 0,662$ et $0,661$, de moyenne $s = 0,669$, à comparer à $0,667$.

► Remarques

- Calculer la « moyenne des moyennes » revient en fait naturellement à remplacer $n = 1000$ par $n = 10000$. Toutefois le fait de publier aussi les dix $e[m]$ et non simplement s permet de se faire une idée de la dispersion de ces $e[m]$ autour de s .
- Le lecteur adaptera l'algorithme ci-dessus au cas où l'on sait que, pour certaines raisons, p décrit de manière uniforme un intervalle $[u, v]$ (penser au changement de variable $x \mapsto u + (v - u)x$).

5.2 Estimation d'un intervalle inconnu

Ce deuxième exemple ressemble au premier, à cela près que l'on y estime deux nombres, les extrémités (u, v) d'un intervalle $[u, v]$ aléatoire (ainsi que sa longueur $v-u$).

On y utilise encore une loi uniforme émulée par un `RANDOM`.

```

PROGRAM Estimi ;
VAR h, i, m, n : INTEGER ; a, p, q, r, s, t, u, v : REAL ;
    e, f : ARRAY[1..20] OF REAL ;

```



```

BEGIN
  RANDOMIZE ;
  u := -2/3 ; v := 23/4 ; h := 10 ; n := 1000 ; s := 0 ; t := 0 ;
  FOR m := 1 TO h DO
    BEGIN
      p := 0 ; q := 0 ;
      FOR i := 1 TO n DO
        BEGIN
          a := RANDOM ; r := u+(v-u)*a ;
          IF r < p THEN p := r ;
          IF r > q THEN q := r
        END ;
        e[m] := p ; s := s+e[m] ;
        f[m] := q ; t := t+f[m] ;
        WRITE (e[m]:4:3, ' ', f[m]:4:3) ;
        WRITE (f[m]-e[m]):4:3) ; WRITELN
      END ;
      s := s/h ; t := t/h ;
      WRITELN (s:4:3, ' ', t:4:3, ' ', (t-s):4:3) ;
      WRITELN (u:4:3, ' ', v:4:3, ' ', (v-u):4:3)
    END.

```

Ici, le changement de variable $x \mapsto u + (v - u)x$ donne des résultats uniformément répartis sur l'intervalle $[u, v]$ à estimer.

Chacune des 10 estimations est indexée par m ; alors $e[m]$ est la plus petite valeur fournie par les $n = 1000$ appels à `RANDOM` tandis que $f[m]$ en est la plus grande.

On notera que p et q sont deux variables, respectivement décroissante et croissante, destinées à obtenir les estimations cherchées de u et de v pour chaque valeur de l'index m .

Ces nombres $e[m]$ et $f[m]$, en principe proches de u et v , sont eux-mêmes ajoutés entre eux avant une division finale par $h = 10$ qui donne des « moyennes des estimations » notées s et t , publiées à l'écran (on peut faire à leur sujet la même remarque que sur la « moyenne des moyennes » de l'algorithme précédent).

Par exemple, une expérience a donné

- les 10 nombres successifs $e[m] = -0,656 ; -0,653 ; -0,665 ; -0,662 ; -0,643 ; -0,657 ; -0,665 ; -0,662 ; -0,650$ et $-0,662$, de moyenne $s = -0,657$, à comparer à $-0,667$;
- les 10 nombres successifs $f[m] = 5,742 ; 5,744 ; 5,745 ; 5,741 ; 5,746 ; 5,750 ; 5,742 ; 5,742 ; 5,740$ et $5,746$, de moyenne $t = 5,744$, à comparer à $5,750$;
- 10 estimations $f[m]-e[m]$ de $v-u$, qui s'en déduisent aussitôt et dont la moyenne est $6,401$, à comparer à $6,417$.

► Remarque

Naturellement les symboles « :4:3 » qui servent à préparer un affichage propre à l'écran sont à modifier en fonction des valeurs de u et de v , dont on a de toute manière une idée au moins grossière avant de lancer le programme d'estimation.

5.3 Intervalles de confiance pour un nombre inconnu

Les programmes suivants ont pour but de calculer un certain nombre d'intervalles de confiance symétriques censés contenir un nombre inconnu p avec une probabilité $1 - \alpha$ dont on sait simplement qu'il est compris entre 0 et 1 (comme dans 5.1), et de voir parmi eux le nombre d'intervalles qui ne le contiennent pas.

Les techniques sont celles du chapitre précédent : chacun des h intervalles construits est de la forme $[estim - d, estim + d]$ où $estim$ est l'estimation obtenue et d dépend de n et de α .

Utilisation d'une approximation normale de variables de Bernoulli

Dans ce cas, nous posons

$$[estim - d, estim + d] = \left[estim - \frac{t_\alpha}{2\sqrt{n}}, estim + \frac{t_\alpha}{2\sqrt{n}} \right]$$

qui résulte de l'approximation normale centrée réduite d'une somme de variables de Bernoulli de paramètre p mutuellement indépendantes, avec la majoration $\sqrt{p(1-p)} \leq \frac{1}{2}$.

Pour des raisons de plus grande simplicité de la liste d'instructions, ce programme se limite à un calcul unique ; il serait naturellement immédiat de le transformer de façon à le rendre interactif.

Dans l'algorithme ci-dessous, on a pris $\alpha = 0,1$, d'où $t_\alpha \approx 1,645$, et $h = 100$ (la valeur $h = 10$ utilisée plus haut ne laisserait pas assez de souplesse pour déterminer différents nombres d'intervalles « incorrects »). Le compteur entier c dénombre ces intervalles, qui ne contiennent pas p .

```
PROGRAM Intervalles_de_confiance_pour_un_nombre1 ;
VAR c, h, i, k, m, n : INTEGER ; a, d, p, t : REAL ;
    e : ARRAY[1..100] OF REAL ;

BEGIN
  RANDOMIZE ;
  c := 0 ; p := 2/3 ; h := 100 ; n := 1000 ;
  t := 1.645 ; d := t/(2*SQRT (n)) ;
  FOR m := 1 TO h DO
    BEGIN
      k := 0 ;
      FOR i := 1 TO n DO
```

```

BEGIN
  a := RANDOM ;
  IF a < p THEN k := k+1 ;
END ;
e[m] := k/n ;
IF ABS (e[m]-p) > d THEN c := c+1 ;
WRITE ( ' [', (e[m]-d):4:3, ' ; ', (e[m]+d):4:3, ' ] ' )
END ;
WRITELN ( c, ' ', p:4:3 )
END .

```

Par exemple, une expérience a donné 9 cas sur 100 où un intervalle ainsi construit ne contenait pas $p = 2/3$, par exemple $[0, 612; 0, 664]$. On notera que le rapport $\frac{c}{h} = \frac{9}{100} = 0,09$ est légèrement inférieur au risque $\alpha = 0,1$ accepté *a priori*.

Utilisation de Bienaymé-Tchebychev

Nous passons maintenant à une méthode beaucoup plus générale, figurant également dans le cours, mais dont l'intérêt pratique comme on le verra est faible.

Dans ce cas, nous posons

$$[estim - d, estim + d] = \left[estim - \frac{1}{2\sqrt{n\alpha}}, estim + \frac{1}{2\sqrt{n\alpha}} \right]$$

qui résulte de la considération de $\left[E(T_n) - \frac{\sigma(T_n)}{\sqrt{\alpha}}, E(T_n) + \frac{\sigma(T_n)}{\sqrt{\alpha}} \right]$ (application de l'inégalité de Bienaymé-Tchebychev) avec la majoration $\sigma = \sqrt{\frac{p(1-p)}{n}} \leq \frac{1}{2\sqrt{n}}$.

Pour des raisons de plus grande simplicité de la liste d'instructions, ce programme se limite à un calcul unique ; il serait naturellement immédiat de le transformer de façon à le rendre interactif.

L'application de Bienaymé-Tchebychev est bien trop grossière ; aussi un exemple raisonnable, avec $\alpha = 0,1$ et $h = 100$ par exemple comme dans la précédente estimation p , conduit en général à un résultat excessif : tous les intervalles obtenus contiennent effectivement p !

La majoration de $\sqrt{p(1-p)}$ n'arrange pas non plus les choses.

Dans l'algorithme ci-dessous, on a donc volontairement forcé le trait en portant le risque α de 0,1 à 0,2 et, surtout, le nombre h d'estimations de 100 à 1000. Le compteur entier c dénombre les intervalles ne contenant pas p .

```

PROGRAM Intervalles_de_confiance_pour_un_nombre2 ;
VAR c, h, i, k, m, n : INTEGER ; alpha, a, d, p : REAL ;
    e : ARRAY[1..1000] OF REAL ;

```

```

BEGIN
  RANDOMIZE ;
  c := 0 ; p := 2/3 ; h := 1000 ; n := 1000 ;
  alpha := 0.2 ; d := 1/(2*SQRT (n*alpha)) ;
  FOR m := 1 TO h DO
    BEGIN
      k := 0 ;
      FOR i := 1 TO n DO
        BEGIN
          a := RANDOM ;
          IF a < p THEN k := k+1
        END ;
        e[m] := k/n ;
        IF ABS (e[m]-p) > d THEN c := c+1 ;
        WRITE ( ' [', (e[m]-d):4:3, ' ; ', (e[m]+d):4:3, ' ] ' )
      END ;
    WRITELN ; WRITELN (c, ' ', p:4:3)
  END.

```

Par exemple, une expérience a donné 18 cas sur 1000 où un intervalle ainsi construit ne contenait pas $p = 2/3$, par exemple $[0, 669; 0, 739]$ ou même $[0, 582; 0, 652]$. On notera que, comme on pouvait s'y attendre, le rapport $\frac{c}{h} = \frac{18}{1000} = 0,018$ est très inférieur au risque $\alpha = 0,2$ accepté *a priori*.

5.4 Intervalles de confiance pour l'espérance d'une loi normale

Le programme suivant a pour but de calculer un certain nombre d'intervalles de confiance symétriques censés contenir l'espérance mathématique inconnue d'une variable suivant une loi normale $\mathcal{N}(\mu, \sigma^2)$ avec une probabilité $1 - \alpha$, et de voir parmi eux le nombre d'intervalles qui ne le contiennent pas. L'écart-type de cette loi est supposé connu, ou au moins majorable d'une manière assez précise.

La technique est celle du chapitre précédent : chacun des h intervalles construits est de la forme

$$[estim - d, estim + d] = \left[estim - \frac{t_\alpha \sigma}{\sqrt{n}}, estim + \frac{t_\alpha \sigma}{\sqrt{n}} \right]$$

qui résulte de l'étude de la fonction de répartition Φ de la loi normale.

Pour des raisons de plus grande simplicité de la liste d'instructions, ce programme se limite à un calcul unique ; il serait naturellement immédiat de le transformer de façon à le rendre interactif.

Dans l'algorithme ci-dessous, on a pris $\alpha = 0,1$, d'où $t_\alpha \approx 1,645$, $\mu = 3,14$, $\sigma = 0,2$ et $h = 100$. Le compteur entier c dénombre les intervalles qui ne contiennent pas μ .

L'écriture de la fonction `normal` est reprise de la simulation d'une variable aléatoire suivant une loi normale centrée réduite. La somme x des $n = 1000$ valeurs engendrées par autant d'appels de cette fonction, puis divisée par n , suit approximativement la loi normale centrée réduite. Pour obtenir une simulation d'une loi du type $\mathcal{N}(\mu, \sigma^2)$, il suffit de la transformer par l'application $\frac{x}{n} \mapsto \mu + \sigma \frac{x}{n}$.

```

PROGRAM Intervalles_de_confiance_pour_une_esperance ;
VAR c, h, i, m, n : INTEGER ; mu, sigma, d, t, x : REAL ;
    e : ARRAY[1..100] OF REAL ;

FUNCTION normal : REAL ;
VAR j : INTEGER ; a : REAL ;

    BEGIN
        a := -6 ;
        FOR j := 1 TO 12 DO a := a+RANDOM ;
        normal := a
    END ;

BEGIN
    RANDOMIZE ;
    c := 0 ; mu := 3.14 ; sigma := 0.2 ; h := 100 ; n := 1000 ;
    t := 1.645 ; d := (t*sigma)/SQRT (n) ;
    FOR m := 1 TO h DO
        BEGIN
            x := 0 ;
            FOR i := 1 TO n DO x := x+normal ;
            e[m] := mu+(sigma*x)/n ;
            IF ABS (e[m]-mu) > d THEN c := c+1 ;
            WRITE ( ' [', (e[m]-d):4:3, ' ; ', (e[m]+d):4:3, ' ] ' )
        END ;
    WRITELN ; WRITELN (c, ' ', mu:4:3)
END.

```

Par exemple, une expérience a donné 11 cas sur 100 où un intervalle ainsi construit ne contenait pas $\mu = 3,14$, par exemple $[3,142; 3,163]$. On notera que le rapport $\frac{c}{h} = \frac{11}{100} = 0,11$ est légèrement supérieur au risque $\alpha = 0,1$ accepté *a priori*, et non légèrement inférieur – ce qui peut aussi se produire – ; cela tient à ce que la fonction censée simuler une loi normale centrée réduite n'est pas parfaite (le nombre 12 devant sans aucun doute remplacé par un entier nettement plus grand).

► Remarque

Le lecteur curieux pourra réfléchir sur les rôles différents des trois boucles (ou types de boucles) présents dans ce dernier programme :

- L'indice m , qui varie de 1 à n , correspond à l'obtention de 100 estimations $e[m]$ et des 100 intervalles de confiance qui s'en déduisent.
- Au sein de chacune de ces estimations, l'indice i , qui varie de 1 à n , correspond à la construction de chacune des 1000 simulations de variable normale centrée réduite servant à déterminer $e[m]$.
- Enfin l'indice j de la fonction `normal`, qui varie de 1 à 12, correspond à l'entassement des valeurs de 12 variables mutuellement indépendantes suivant une loi uniforme sur $[0, 1]$.

Ainsi un programme plutôt court d'une vingtaine de lignes, d'aspect assez anodin, correspond dans les faits à plus d'un million d'appels à `RANDOM` ! Seule la rapidité des processeurs actuels masque admirablement ce fait.

Tables des lois usuelles

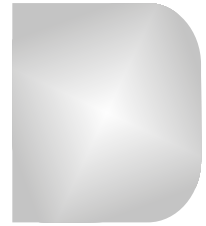


Tableau récapitulatif des variables discrètes classiques

Nom	Ensemble des valeurs	Loi	Espérance	Variance
loi uniforme $X \hookrightarrow \mathcal{U}(\llbracket 1, n \rrbracket)$	$\llbracket 1, n \rrbracket$	$P([X = k]) = \frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
loi de Bernoulli $X \hookrightarrow \mathcal{B}(1, p)$	$\{0, 1\}$	$P([X = 1]) = p$ $P([X = 0]) = q$	p	pq
loi binomiale $X \hookrightarrow \mathcal{B}(n, p)$	$\llbracket 0, n \rrbracket$	$P([X = k]) = \binom{n}{k} p^k q^{n-k}$	np	npq
loi géométrique $X \hookrightarrow \mathcal{G}(p)$	\mathbb{N}^*	$P([X = k]) = pq^{k-1}$	$\frac{1}{p}$	$\frac{q}{p^2}$
loi hypergéométrique $X \hookrightarrow \mathcal{H}(N, n, p)$	inclus dans $\llbracket 0, n \rrbracket$	$P([X = k]) = \frac{\binom{Np}{k} \binom{Nq}{n-k}}{\binom{N}{n}}$	np	$npq \frac{N-n}{N-1}$
loi de Poisson $X \hookrightarrow \mathcal{P}(\lambda)$	\mathbb{N}	$P([X = k]) = e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ

I. Loi binomiale

$k \backslash p$	0,1	0,2	0,3	0,4	0,5
0	0,5905	0,3277	0,1681	0,0778	0,0313
1	0,3281	0,4096	0,3602	0,2592	0,1563
2	0,0729	0,2048	0,3087	0,3456	0,3125
3	0,0081	0,0512	0,1323	0,2304	0,3125
4	0,0005	0,0064	0,0284	0,0768	0,1563
5		0,0003	0,0024	0,0102	0,0313
$n = 5$					
0	0,3487	0,1074	0,0282	0,0060	0,0010
1	0,3874	0,2684	0,1211	0,0403	0,0098
2	0,1937	0,3020	0,2335	0,1209	0,0439
3	0,0574	0,2013	0,2668	0,2150	0,1172
4	0,0112	0,0881	0,2001	0,2508	0,2051
5	0,0015	0,0264	0,1029	0,2007	0,2461
6	0,0001	0,0055	0,0368	0,1115	0,2051
7		0,0008	0,0090	0,0425	0,1172
8		0,0001	0,0015	0,0106	0,0439
9			0,0001	0,0016	0,0098
10				0,0001	0,0010
$n = 10$					
0	0,2059	0,0352	0,0047	0,0005	
1	0,3432	0,1319	0,0305	0,0047	0,0005
2	0,2669	0,2309	0,0916	0,0219	0,0032
3	0,1285	0,2501	0,1700	0,0634	0,0139
4	0,0428	0,1876	0,2186	0,1268	0,0417
5	0,0105	0,1032	0,2061	0,1859	0,0916
6	0,0019	0,0430	0,1472	0,2066	0,1527
7	0,0003	0,0138	0,0811	0,1771	0,1964
8		0,0035	0,0348	0,1181	0,1964
9		0,0007	0,0116	0,0612	0,1527
10		0,0001	0,0030	0,0245	0,0916
11			0,0006	0,0074	0,0417
12			0,0001	0,0016	0,0139
13				0,0003	0,0032
14					0,0005
15					
$n = 15$					

0	0,1216	0,0115	0,0008		
1	0,2701	0,0576	0,0068	0,0005	
2	0,2852	0,1369	0,0278	0,0031	0,0002
3	0,1901	0,2054	0,0716	0,0123	0,0011
4	0,0898	0,2182	0,1304	0,0350	0,0046
5	0,0319	0,1746	0,1789	0,0746	0,0148
6	0,0089	0,1091	0,1916	0,1244	0,0370
7	0,0020	0,0545	0,1643	0,1659	0,0739
8	0,0004	0,0222	0,1144	0,1797	0,1201
9	0,0001	0,0074	0,0654	0,1597	0,1602
10		0,0020	0,0308	0,1171	0,1762
11		0,0005	0,0120	0,0710	0,1602
12		0,0001	0,0039	0,0355	0,1201
13			0,0010	0,0146	0,0739
14			0,0002	0,0049	0,0370
15				0,0013	0,0148
16				0,0003	0,0046
17					0,0011
18					0,0002
19					
20					
$n = 20$					

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

II. Loi de Poisson

$k \backslash m$	1	2	3	4	5	6	7	8	9	10
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	
1	0,3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011	0,0005
2	0,1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050	0,0023
3	0,0613	0,1804	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150	0,0076
4	0,0153	0,0902	0,1680	0,1954	0,1755	0,1339	0,0912	0,0573	0,0337	0,0189
5	0,0031	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607	0,0378
6	0,0005	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911	0,0631
7	0,0001	0,0034	0,0216	0,0595	0,1044	0,1377	0,1490	0,1396	0,1171	0,0901
8		0,0009	0,0081	0,0298	0,0653	0,1033	0,1304	0,1396	0,1318	0,1126
9		0,0002	0,0027	0,0132	0,0363	0,0688	0,1014	0,1241	0,1318	0,1251
10			0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186	0,1251
11			0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970	0,1137
12			0,0001	0,0006	0,0034	0,0113	0,0263	0,0481	0,0728	0,0948
13				0,0002	0,0013	0,0052	0,0142	0,0296	0,0504	0,0729
14				0,0001	0,0005	0,0022	0,0071	0,0169	0,0324	0,0521
15					0,0002	0,0009	0,0033	0,0090	0,0194	0,0347
16						0,0003	0,0014	0,0045	0,0109	0,0217
17						0,0001	0,0006	0,0021	0,0058	0,0128
18							0,0002	0,0009	0,0029	0,0071
19							0,0001	0,0004	0,0014	0,0037
20								0,0002	0,0006	0,0019
21								0,0001	0,0003	0,0009
22									0,0001	0,0004
23										0,0002
24										0,0001
25										

$$P(X = k) = e^{-m} \frac{m^k}{k!}$$

Fonction de répartition de la loi de Poisson

$k \backslash m$	1	2	3	4	5	6	7	8	9	10
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	
1	0,7358	0,4060	0,1991	0,0916	0,0404	0,0174	0,0073	0,0030	0,0012	0,0005
2	0,9197	0,6767	0,4232	0,2381	0,1247	0,0620	0,0296	0,0138	0,0062	0,0028
3	0,9810	0,8571	0,6472	0,4335	0,2650	0,1512	0,0818	0,0424	0,0212	0,0103
4	0,9963	0,9473	0,8153	0,6288	0,4405	0,2851	0,1730	0,0996	0,0550	0,0293
5	0,9994	0,9834	0,9161	0,7851	0,6160	0,4457	0,3007	0,1912	0,1157	0,0671
6	0,9999	0,9955	0,9665	0,8893	0,7622	0,6063	0,4497	0,3134	0,2068	0,1301
7	1	0,9989	0,9881	0,9489	0,8666	0,7440	0,5987	0,4530	0,3239	0,2202
8		0,9998	0,9962	0,9786	0,9319	0,8472	0,7291	0,5925	0,4557	0,3328
9		1	0,9989	0,9919	0,9682	0,9161	0,8305	0,7166	0,5874	0,4579
10			0,9997	0,9972	0,9863	0,9574	0,9015	0,8159	0,7060	0,5830
11			0,9999	0,9991	0,9945	0,9799	0,9467	0,8881	0,8030	0,6968
12			1	0,9997	0,9980	0,9912	0,9730	0,9362	0,8758	0,7916
13				0,9999	0,9993	0,9964	0,9872	0,9658	0,9261	0,8645
14				1	0,9998	0,9986	0,9943	0,9827	0,9585	0,9165
15					0,9999	0,9995	0,9976	0,9918	0,9780	0,9513
16					1	0,9998	0,9990	0,9963	0,9889	0,9730
17						0,9999	0,9996	0,9984	0,9947	0,9857
18						1	0,9999	0,9993	0,9976	0,9928
19							1	0,9997	0,9989	0,9965
20								0,9999	0,9996	0,9984
21								1	0,9998	0,9993
22									0,9999	0,9997
23									1	0,9999
24										1
25										

$$F(k) = P(X \leq k) = \sum_{n=0}^k e^{-m} \frac{m^n}{n!}.$$

III. Loi de Laplace-Gauss

$\varphi(x)$	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2102	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \varphi(-x).$$

Fonction de répartition de la loi de Laplace-Gauss

$\Phi(x)$	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1	1	1	1	1	1	1	1	1	1

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = 1 - \Phi(-x), \quad \text{quand } x \text{ est grand } \Phi(x) \approx 1 - \frac{\varphi(x)}{x}$$

**Solution
des
exercices**

Chapitre 1

1 La matrice A est clairement de rang 2, donc on peut déjà conclure que 0 est valeur propre et que $\dim(E_0) = n - 2$.

De plus

$$\left\{ \begin{array}{l} x_1 + x_2 + \cdots + x_n = 0 \\ \quad x_1 + x_n = 0 \\ \quad \quad x_1 + x_n = 0 \\ \quad \quad \quad \vdots \\ \quad \quad \quad x_1 + x_n = 0 \\ x_1 + x_2 + \cdots + x_n = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} x_1 = -x_n \\ x_2 = -x_3 + \cdots - x_{n-1} \end{array} \right.$$

d'où

$$E_0 = \text{Vect} \left(\left(\begin{array}{c} -1 \\ 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \\ 1 \end{array} \right), \left(\begin{array}{c} 0 \\ -1 \\ 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ 0 \end{array} \right), \dots, \left(\begin{array}{c} 0 \\ -1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{array} \right) \right).$$

Pour les autres valeurs propres déterminons les réels λ non nuls tels que le système $(A - \lambda I)X = 0$ admette une solution non nulle. Résolvons ce système

$$\left\{ \begin{array}{l} x_1 + x_2 + \cdots + x_n = \lambda x_1 \\ \quad x_1 + x_n = \lambda x_2 \\ \quad \quad x_1 + x_n = \lambda x_3 \\ \quad \quad \quad \vdots \\ \quad \quad \quad x_1 + x_n = \lambda x_{n-1} \\ x_1 + x_2 + \cdots + x_n = \lambda x_n \end{array} \right.$$

Comme $\lambda \neq 0$ ce système équivaut à

$$\left\{ \begin{array}{l} x_2 = x_3 = x_4 = \cdots = x_{n-1} = \frac{1}{\lambda}(x_1 + x_n) \\ x_1 = x_n = \frac{1}{\lambda} \sum_{k=1}^n x_k \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} x_2 = x_3 = x_4 = \cdots = x_{n-1} = \frac{2}{\lambda}x_1 \\ x_n = x_1 \\ x_1 = \frac{1}{\lambda}(2x_1 + (n-2)\frac{2}{\lambda}x_1) \end{array} \right.$$

$$\Leftrightarrow \left\{ \begin{array}{l} x_2 = x_3 = x_4 = \cdots = x_{n-1} = \frac{2}{\lambda}x_1 \\ x_n = x_1 \\ (\lambda^2 - 2\lambda - 2(n-2))x_1 = 0 \end{array} \right.$$

Ce système admet une solution non nulle si, et seulement si, $\lambda^2 - 2\lambda - 2(n-2) = 0$. Nous avons $\Delta' = 1 + 2(n-2) = 2n-3$.

La matrice A possède donc deux autres valeurs propres $1 \pm \sqrt{2n-3}$. D'après la résolution du système on a, pour $\lambda = 1 \pm \sqrt{2n-3}$.

$$E_\lambda = \text{Vect} \left(\begin{pmatrix} \lambda \\ 2 \\ \vdots \\ 2 \\ \lambda \end{pmatrix} \right).$$

2 On remarque que

$$A - (a-1)I = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{pmatrix}$$

d'où $\text{rg}(A - (a-1)I) = 1$. D'après la formule du rang, $\dim(E_{a-1}(A)) = n-1$ (en particulier $a-1$ est une valeur propre de A). De plus, pour $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$,

$$X \in E_{a-1}(A) \Leftrightarrow x_1 + x_2 + \dots + x_n = 0 \Leftrightarrow x_1 = -x_2 + \dots - x_n,$$

d'où

$$E_1(A) = \text{Vect} \left(\begin{pmatrix} -1 \\ 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} -1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right)$$

Comme $\dim(E_{a-1}) = n-1$, il ne reste plus qu'éventuellement une seule autre valeur propre. Mais un peu d'intuition nous fait constater que

$$A \times \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} a+(n-1) \\ \vdots \\ a+(n-1) \end{pmatrix} = (a+n-1) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

donc $a+n-1$ est la seconde valeur propre de A et $E_{a+n-1}(A) = \text{Vect} \left(\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right)$.

3 1. Soit F une primitive de f (qui existe puisque f est continue), on a alors pour tout réel x :

$$T_a(f)(x) = \frac{1}{2a} (F(x+a) - F(x-a)).$$

Il en résulte que $T_a(f)$ est \mathcal{C}^1 comme somme et composée de fonctions \mathcal{C}^1 .

2. D'après la question précédente,

$$T_a(f)'(x) = \frac{1}{2a} (f(x+a) - f(x-a)).$$

- Si f admet $2a$ comme période, il en résulte immédiatement que $T(f)'(x) = 0$ pour tout réel x . Par voie de conséquence $T(f)$ est constante.
- Si $T(f)$ est constante alors, sa dérivée est nulle. Donc pour tout réel x , $f(x+a) = f(x-a)$, ce qui équivaut à

$$\forall x \in \mathbb{R}, f(x) = f(x+2a),$$

autrement dit f est périodique de période $2a$.

3. Remarquons déjà que pour tout $f \in E$, $T(f)$ est \mathcal{C}^1 sur \mathbb{R} et a fortiori continue sur \mathbb{R} , ce qui signifie que T est à valeurs dans E .

Soient $f, g \in E$ et $\lambda \in \mathbb{R}$, nous avons pour tout réel x :

$$\begin{aligned} T_a(\lambda f + g)(x) &= \frac{1}{2a} \int_{x-a}^{x+a} (\lambda f + g)(t) dt = \frac{1}{2a} \int_{x-a}^{x+a} \lambda f(t) + g(t) dt \\ &= \lambda \times \frac{1}{2a} \int_{x-a}^{x+a} f(t) dt + \frac{1}{2a} \int_{x-a}^{x+a} g(t) dt \\ &= \lambda T_a(f)(x) + T_a(g)(x) \\ &= (\lambda T_a(f) + T_a(g))(x) \end{aligned}$$

Il en résulte que $T_a(\lambda f + g) = \lambda T_a(f) + T_a(g)$ et donc T_a est un endomorphisme de E .

Soit $f \in E$, on sait déjà que si $f \in \text{Ker}(T_a)$ alors f est périodique de période $2a$.

Réciproquement, soit f une fonction périodique de E de période $2a$, alors $T_a(f)$ est constante et

$$T_a(f) = 0 \Leftrightarrow \int_{-a}^a f(t) dt = 0.$$

Autrement dit le noyau de T_a est constitué des applications périodiques f de E de période $2a$ et telles que $\int_{-a}^a f(t) dt = 0$.

D'après la question 1, pour tout $f \in E$, $T_a(f)$ est \mathcal{C}^1 sur \mathbb{R} . Or il existe des fonctions continues qui ne sont pas \mathcal{C}^1 , donc T_a n'est pas surjectif.

4. Soit $f \in \mathbb{R}_n[X]$ telle que pour tout réel x , $f(x) = \sum_{k=0}^n a_k x^k$, on a alors

$$\begin{aligned} \forall x \in \mathbb{R}, T_a(f)(x) &= \frac{1}{2a} \int_{x-a}^{x+a} \sum_{k=0}^n a_k t^k dt = \frac{1}{2a} \sum_{k=0}^n a_k \int_{x-a}^{x+a} t^k dt = \frac{1}{2a} \sum_{k=0}^n \left[\frac{t^{k+1}}{k+1} \right]_{x-a}^{x+a} \\ &= \frac{1}{2a} \sum_{k=0}^n a_k \frac{(x+a)^{k+1} - (x-a)^{k+1}}{k+1} \end{aligned}$$

Les termes de la somme pour $k \leq n-1$ sont clairement de degré inférieur ou égal à n . De plus le coefficient de degré $n+1$ de $(x+a)^{n+1}$ est x^{n+1} et celui de $(x-a)^{n+1}$ est aussi x^{n+1} , donc $(x+a)^{n+1} - (x-a)^{n+1}$ est aussi de degré inférieur à n . Par voie de conséquence $T_a(f)$ est une fonction polynomiale de degré inférieur ou égal à n .

5. a) Pour $k \in \llbracket 0, n \rrbracket$,

$$\begin{aligned} T_a(X^k)(x) &= \frac{1}{2a} \int_{x-a}^{x+a} t^k dt = \frac{1}{2a} \left(\frac{(x+a)^{k+1} - (x-a)^{k+1}}{k+1} \right) \\ &= \frac{1}{2a(k+1)} \sum_{i=0}^{k+1} \binom{k+1}{i} (x^i a^{k+1-i} - x^i (-a)^{k+1-i}). \end{aligned}$$

On remarque comme dans la question précédente que le terme de degré $k+1$ est nul, d'où

$$T_a(X^k)(x) = \frac{1}{2a(k+1)} \sum_{i=0}^k \binom{k+1}{i} (x^i a^{k+1-i} + (-1)^{k-i} x^i a^{k+1-i}).$$

Donc $T_a(f)$ est de degré inférieur ou égal à k et le terme de degré k est

$$\frac{1}{2a(k+1)} \binom{k+1}{k} 2x^k a = x^k.$$

La matrice de $T_a(f)$ est triangulaire supérieure avec des 1 sur la diagonale. On en déduit que $\text{Sp}(T_a) = \{1\}$. Si T_a était diagonalisable, comme 1 est son unique valeur propre, on aurait $T_a = \text{Id}_{\mathbb{R}_n[X]}$, or on vérifie aisément que $T_a(X^2) \neq X^2$. T_a n'est donc pas diagonalisable.

b) Soit f un polynôme tel que $f(x) = a_2 x^2 + a_1 x + a_0$ alors

$$\begin{aligned} T(f)(x) &= a_2 T(X^2)(x) + a_1 T(X)(x) + a_0 T(1)(x) = a_2 x^2 + \frac{a_2 a^2}{3} + a_1 x + a_0 \\ &= a_2 x^2 + a_1 x + \left(\frac{a_2 a^2}{3} + a_0 \right). \end{aligned}$$

On a donc $T(f) = f \Leftrightarrow \left(\frac{a_2 a^2}{3} + a_0 \right) = a_0 \Leftrightarrow a_2 = 0$. Il n'existe donc pas de polynôme de degré 2 qui soit vecteur propre de f . Le même calcul prouve d'ailleurs que $\mathbb{R}_1[X] \subset E_1(T_a)$.

c) Soit $f \in \mathbb{R}_n[X]$ tel que $T(f) = f$. Par définition,

$$\forall x \in \mathbb{R}, T(f')(x) = \frac{1}{2a} \int_{x-a}^{x+a} f'(t) dt = \frac{1}{2a} (f(x+a) - f(x-a)).$$

Or on sait que $T(f)'(x) = \frac{1}{2a} (f(x+a) - f(x-a))$. On en déduit que

$$\forall x \in \mathbb{R}, T(f')(x) = (T(f))'(x) = f'(x).$$

Soit $T(f') = f'$. En conclusion, si f est un vecteur propre de degré supérieur ou égal à 1 alors f' est également un vecteur propre.

Comme il n'y a pas de vecteur propre de T_a de degré 2, il ne peut y avoir de vecteur propre de T_a de degré supérieur à 2. En effet si f était un tel vecteur propre, alors $f^{(\text{deg}(f)-2)}$ serait lui-même vecteur propre et de degré 2 ce qui est absurde.

On en déduit que $E_1(T_a) \subset \mathbb{R}_1[X]$. D'après la question précédente, il en résulte que $E_1(T_a) = \mathbb{R}_1[X]$.

- 4 1. Soit u'_2 la restriction de u_2 à $\text{Ker}(u_1 \circ u_2)$. D'après la formule du rang appliquée à u'_2 nous avons

$$\text{rg}(u'_2) + \dim(\text{Ker}(u'_2)) = \dim(\text{Ker}(u_1 \circ u_2)).$$

Or $\text{Im}(u'_2) \subset \text{Ker}(u_1)$ et $\text{Ker}(u'_2) \subset \text{Ker}(u_1 \circ u_2)$, donc

$$\dim(\text{Ker}(u_1)) + \dim(\text{Ker}(u_2)) \geq \dim(\text{Ker}(u_1 \circ u_2)).$$

2. C'est une récurrence immédiate.
 3. On a $u^3 - \text{Id}_E = 0$, soit encore $(u - \text{Id}_E) \circ (u - j \text{Id}_E) \circ (u - j^2 \text{Id}_E) = 0$. On a donc $\dim(\text{Ker}(u^3 - \text{Id}_E)) = \dim(E)$ et d'après la question précédente

$$\dim(u - \text{Id}_E) + \dim(u - j \text{Id}_E) + \dim(u - j^2 \text{Id}_E) \geq \dim(E)$$

ce qui peut encore s'écrire $\dim(E_1(u)) + \dim(E_j(u)) + \dim(E_{j^2}(u)) \geq \dim(E)$.

Mais on sait également que $\dim(E_1(u)) + \dim(E_j(u)) + \dim(E_{j^2}(u)) \leq \dim(E)$, car ces trois sous-espaces vectoriels sont en somme directe. On en déduit que

$$\dim(E_1(u)) + \dim(E_j(u)) + \dim(E_{j^2}(u)) = \dim(E),$$

et donc que E est somme directe de sous-espaces propres de u (même si parmi les sous-espaces vectoriels E_1 , E_j et E_{j^2} , certains peuvent être réduits au vecteur nul). Il s'ensuit que u est diagonalisable.

4. Soit P un tel polynôme annulateur qui peut donc s'écrire sous la forme $P = \prod_{k=1}^n (X - \lambda_k)$

où les réels λ_k sont distincts deux à deux.

Par hypothèse $\dim(\text{Ker}(P(u))) = \dim(E)$. Mais d'après la question 2,

$$\sum_{k=1}^n \dim(\text{Ker}(u - \lambda_k \text{Id}_E)) \geq \dim(\text{Ker}(P(u))) = \dim(E).$$

Mais on remarque que $\text{Ker}(u - \lambda_k \text{Id}_E) = E_{\lambda_k}(u)$ et d'après le cours on sait que $E_{\lambda_1}(u)$, $E_{\lambda_2}(u)$, \dots , $E_{\lambda_n}(u)$ sont en somme directe, d'où

$$\sum_{k=1}^n \dim(E_{\lambda_k}(u)) \leq \dim(E).$$

Il s'ensuit que $\sum_{k=1}^n \dim(E_{\lambda_k}(u)) = \dim(E)$, ce qui entraîne la diagonalisabilité de u .

5. Soit $\lambda_1, \lambda_2, \dots, \lambda_n$ les valeurs propres de u . Posons $P = \prod_{k=1}^n (X - \lambda_k)$.

Nous allons montrer que $P(u) = 0$ en montrant que $P(u)$ est nul sur chacun des sous-espaces propres.

Soit $i \in \llbracket 1, n \rrbracket$, la restriction de $u - \lambda_i \text{Id}_E$ au sous-espace vectoriel $E_{\lambda_i}(u)$ est nul, par définition. A fortiori la restriction de $P(u)$ à $E_{\lambda_i}(u)$ est également nul.

L'espace E étant somme directe de $E_{\lambda_1}(u)$, $E_{\lambda_2}(u)$, \dots , $E_{\lambda_n}(u)$, $P(u)$ s'annule sur E .

5 1. Triangulons $A - \lambda I_3$:

$$\begin{aligned} & \begin{pmatrix} 1-\lambda & -12 & 2 \\ 1 & 1-\lambda & 1 \\ 4 & 8 & 3-\lambda \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1-\lambda & 1 \\ 1-\lambda & -12 & 2 \\ 4 & 8 & 3-\lambda \end{pmatrix} \quad L_1 \leftrightarrow L_2 \\ & \rightarrow \begin{pmatrix} 1 & 1-\lambda & 1 \\ 0 & -\lambda^2 + 2\lambda - 13 & 1+\lambda \\ 0 & 4+4\lambda & -1-\lambda \end{pmatrix} \quad \begin{array}{l} L_2 \leftarrow L_2 - (1-\lambda)L_1 \\ L_3 \leftarrow L_3 - 4L_1 \end{array} \\ & \rightarrow A(\lambda) = \begin{pmatrix} 1 & 1-\lambda & 1 \\ 0 & -\lambda^2 + 2\lambda - 13 & 1+\lambda \\ 0 & -\lambda^2 + 6\lambda - 9 & 0 \end{pmatrix} \quad L_3 \leftarrow L_3 + L_2 \end{aligned}$$

Cette matrice est non inversible si, et seulement si, $\lambda = -1$ ou $-\lambda^2 + 6\lambda - 9 = 0$. Le discriminant réduit Δ' vaut $9 - 9 = 0$, il y a une racine double 3. Les valeurs propres de f sont -1 et 3.

2. Déterminons une base de ces sous-espaces.

Soit $(x, y, z) \in \mathbb{R}^3$,

$$(x, y, z) \in \text{Ker}(f + \text{Id}) \Leftrightarrow A(-1) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0 \Leftrightarrow \begin{cases} x + 2y + z = 0 \\ -16y = 0 \\ -16y = 0 \end{cases} \Leftrightarrow \begin{cases} x = -z \\ y = 0 \end{cases}$$

d'où $E_{-1} = \text{Vect}((-1, 0, 1))$.

Pour déterminer $\text{Ker}(f - 3 \text{Id})^2$, on calcule la matrice de $(f - 3 \text{Id})^2$ dans la base canonique :

$$\begin{pmatrix} -2 & -12 & 2 \\ 1 & -2 & 1 \\ 4 & 8 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 64 & -16 \\ 0 & 0 & 0 \\ 0 & -64 & 16 \end{pmatrix},$$

il en résulte que $(x, y, z) \in \text{Ker}(f - 3 \text{Id})^2 \Leftrightarrow 64y - 16z = 0 \Leftrightarrow z = 4y$.

D'où $\text{Ker}(f - 3 \text{Id})^2 = \text{Vect}((1, 0, 0), (0, 1, 4))$.

Montrons que la famille $((-1, 0, 1), (1, 0, 0), (0, 1, 4))$ est une famille libre. Soient a, b, c trois réels, nous avons :

$$a(-1, 0, 1) + b(1, 0, 0) + c(0, 1, 4) = 0 \Leftrightarrow \begin{cases} -a + b = 0 \\ c = 0 \\ a + 4c = 0 \end{cases} \Leftrightarrow a = b = c = 0,$$

cette famille est donc libre. Comme elle compte trois vecteurs et que $\dim(\mathbb{R}^3) = 3$, c'est une base de \mathbb{R}^3 et $\text{Ker}(f + \text{Id})$ et $\text{Ker}(f - 3 \text{Id})^2$ sont supplémentaires.

3. Le premier vecteur \vec{e}_1 de \mathcal{B}' doit vérifier $f(\vec{e}_1) = -\vec{e}_1$, soit $\vec{e}_1 \in E_{-1}$. D'après ce qui précède, on peut choisir $\vec{e}_1 = (-1, 0, 1)$.

De même le second vecteur doit appartenir à E_3 . D'après la triangulation effectuée à la

première question,

$$(x, y, z) \in E_3 \Leftrightarrow A(3) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0 \Leftrightarrow \begin{cases} x - 2y + z = 0 \\ -16y + 4z = 0 \end{cases} \Leftrightarrow \begin{cases} x = -2y \\ z = 4y \end{cases}$$

On peut choisir comme deuxième vecteur de la base \mathcal{B}' le triplet $\vec{e}_2 = (-2, 1, 4)$.

Comme $(B - 3I)^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, le dernier vecteur est nécessairement dans

$\text{Ker}(f - 3\text{Id})^2$. Cherchons deux réels a et b tels que le vecteur $\vec{e}_3 = a(1, 0, 0) + b(0, 1, 4)$ vérifie $f(\vec{e}_3) = 3\vec{e}_3 + \vec{e}_2$,

$$f(\vec{e}_3) = 3\vec{e}_3 + \vec{e}_2 \Leftrightarrow a(1, 1, 4) + b(-4, 5, 20) = 3(a, b, 4b) + (-2, 1, 4)$$

$$\Leftrightarrow \begin{cases} a - 4b = 3a - 2 \\ a + 5b = 3b + 1 \\ 4a + 20b = 12b + 4 \end{cases} \Leftrightarrow \begin{cases} a + b = \frac{1}{2} \\ a + 2b = \frac{1}{2} \\ 4a + 8b = 4 \end{cases} \Leftrightarrow \begin{cases} a + b = \frac{1}{2} \\ b = \frac{1}{2} \end{cases}$$

$$\Leftrightarrow \begin{cases} a = 0 \\ b = \frac{1}{2} \end{cases}$$

le vecteur $\vec{e}_3 = (0, \frac{1}{2}, 2)$.

Pour conserver des vecteurs à coordonnées entières on peut multiplier les deux derniers vecteur par 2. On pose alors $\mathcal{B}' = ((-1, 0, 1), (-4, 2, 8), (0, 1, 4))$, dont on montre facilement que c'est une base, et dans cette base, par construction, la matrice de f est

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}.$$

4. On remarque que $B = M + N$ où

$$M = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix} \text{ et } N = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

On vérifie que $MN = NM$ et que $N^2 = 0$. D'après la formule du binôme de Newton, on obtient

$$B^n = \sum_{k=0}^n \binom{n}{k} N^k M^{n-k} = M^n + nM^{n-1}N = \begin{pmatrix} (-1)^n & 0 & 0 \\ 0 & 3^n & n3^{n-1} \\ 0 & 0 & 3^n \end{pmatrix}.$$

5. On sait qu'en posant P la matrice de passage de la base canonique vers la base \mathcal{B}' , c'est-à-dire

$$P = \begin{pmatrix} -1 & -4 & 0 \\ 0 & 2 & 1 \\ 1 & 8 & 4 \end{pmatrix},$$

on a la relation $B = P^{-1}AP$ et par suite $A^n = PB^nP^{-1}$.

Après calculs, on trouve $P^{-1} = \frac{1}{4} \begin{pmatrix} 0 & -16 & 4 \\ -1 & 4 & -1 \\ 2 & -4 & 2 \end{pmatrix}$, d'où

$$\begin{aligned} A^n &= \frac{1}{4} \begin{pmatrix} -1 & -4 & 0 \\ 0 & 2 & 1 \\ 1 & 8 & 4 \end{pmatrix} \begin{pmatrix} (-1)^n & 0 & 0 \\ 0 & 3^n & n3^{n-1} \\ 0 & 0 & 3^n \end{pmatrix} \begin{pmatrix} 0 & -16 & 4 \\ -1 & 4 & -1 \\ 2 & -4 & 2 \end{pmatrix} \\ &= \frac{1}{4} \begin{pmatrix} (-1)^{n+1} & -4 \cdot 3^n & -4n \cdot 3^{n-1} \\ 0 & 2 \cdot 3^n & 3^n + 2n \cdot 3^{n-1} \\ (-1)^n & 8 \cdot 3^n & 4 \cdot 3^n + 8n \cdot 3^{n-1} \end{pmatrix} \begin{pmatrix} 0 & -16 & 4 \\ -1 & 4 & -1 \\ 2 & -4 & 2 \end{pmatrix} \\ &= \frac{1}{4} \begin{pmatrix} (12 - 8n)3^{n-1} & 16(-1)^n + (-48 + 16n)3^{n-1} & 4(-1)^{n+1} + (12 - 8n)3^{n-1} \\ 4n \cdot 3^{n-1} & (12 - 8n)3^{n-1} & 4n \cdot 3^{n-1} \\ 16n \cdot 3^{n-1} & 16(-1)^{n+1} + (48 - 32n)3^{n-1} & 4(-1)^n + 16n \cdot 3^{n-1} \end{pmatrix} \\ &= \begin{pmatrix} (3 - 2n)3^{n-1} & 4(-1)^n + (-12 + 4n)3^{n-1} & (-1)^{n+1} + (3 - 2n)3^{n-1} \\ n \cdot 3^{n-1} & (3 - 2n)3^{n-1} & n \cdot 3^{n-1} \\ 4n \cdot 3^{n-1} & 4(-1)^{n+1} + (12 - 8n)3^{n-1} & (-1)^n + 4n \cdot 3^{n-1} \end{pmatrix}. \end{aligned}$$

6 1. Soit $\lambda \in \mathbb{C}$, la matrice $A - \lambda I_2$ est non inversible si, et seulement si, $(a - \lambda)(d - \lambda) - bc = 0$, soit $\lambda^2 - (a + d)\lambda + ad - bc = 0$. Un polynôme de degré 2 admettant toujours au moins une racine dans \mathbb{C} , σ n'est pas vide.

2. On sait que si λ est valeur propre de A alors λ^n est valeur propre de A^n . Compte tenu des hypothèses, on en déduit que toute valeur propre λ de A vérifie l'équation $\lambda^n = 1$. On peut d'ores et déjà écrire que

$$\text{Sp}(A) \subset \left\{ e^{\frac{2ik\pi}{n}} \mid k \in \llbracket 0, n-1 \rrbracket \right\}.$$

Deux cas se présentent :

– Il existe une valeur propre complexe non réelle, de la forme $\lambda = e^{2ip\pi/n}$ (où $p \in \llbracket 1, n-1 \rrbracket$).

On sait alors que $\bar{\lambda}$ (qui est distinct de λ) est aussi valeur propre et comme A ne peut avoir plus de deux valeurs propres, on a

$$\text{Sp}(A) = \left\{ e^{\frac{2ip\pi}{n}}, e^{-\frac{2ip\pi}{n}} \right\}$$

Si $\frac{n}{2} < p \leq n-1$ alors il suffit de poser $p' = n-p$ pour obtenir l'égalité sous la forme voulue.

De plus, $\lambda + \bar{\lambda} = 2 \cos(2p\pi/n)$, et comme les valeurs propres sont les racines du trinôme de la question 1, on en déduit que $2 \cos(2p\pi/n) = a + d$ (en particulier c'est un entier).

– Sinon, les valeurs propres sont nécessairement réelles et alors

$$\text{Sp}(A) \subset \{-1, 1\}.$$

3. Remarquons déjà que A est diagonalisable puisque A possède deux valeurs propres.

Là encore, deux cas se présentent

– $\sigma = \{-1, 1\}$, et alors il existe $P \in \mathcal{GL}_2(\mathbb{C})$ telle que $P^{-1}AP = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Il en résulte que

$$A^{12} = P \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}^{12} P^{-1} = P \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} P^{-1} = I_2.$$

$$-\sigma = \left\{ e^{\frac{2ip\pi}{n}}, e^{-\frac{2ip\pi}{n}} \right\} \text{ où } 1 \leq p < \frac{n}{2}.$$

On sait que l'on doit avoir $2 \cos(2p\pi/n) \in \mathbb{Z}$, comme $|\cos| \leq 1$, il y a cinq solutions

$$2 \cos\left(\frac{2p\pi}{n}\right) \in \{-2, -1, 0, 1, 2\}.$$

Les cas 2 et -2 correspondent à des valeurs propres réelles.

a. Si $\cos\left(\frac{2p\pi}{n}\right) = \frac{1}{2}$, alors $\frac{2p\pi}{n} = \frac{\pi}{3}$. Les valeurs propres sont alors $-j$ et $-j^2$.

$$\text{Il existe } P \in \mathcal{GL}_2(\mathbb{C}) \text{ telle que } P^{-1}AP = \begin{pmatrix} -j & 0 \\ 0 & -j^2 \end{pmatrix}.$$

Il en résulte que

$$A^{12} = P \begin{pmatrix} -j & 0 \\ 0 & -j^2 \end{pmatrix}^{12} P^{-1} = P \begin{pmatrix} j^{12} & 0 \\ 0 & j^{24} \end{pmatrix} P^{-1} = I_2.$$

b. Si $\cos\left(\frac{2p\pi}{n}\right) = 0$, alors $\frac{2p\pi}{n} = \frac{\pi}{2}$. Les valeurs propres sont alors i et $-i$.

$$\text{Il existe } P \in \mathcal{GL}_2(\mathbb{C}) \text{ telle que } P^{-1}AP = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

Il en résulte que

$$A^{12} = P \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}^{12} P^{-1} = P \begin{pmatrix} i^{12} & 0 \\ 0 & i^{12} \end{pmatrix} P^{-1} = I_2.$$

c. Si $\cos\left(\frac{2p\pi}{n}\right) = -\frac{1}{2}$, alors $\frac{2p\pi}{n} = \frac{2\pi}{3}$. Les valeurs propres sont alors j et j^2 .

$$\text{Il existe } P \in \mathcal{GL}_2(\mathbb{C}) \text{ telle que } P^{-1}AP = \begin{pmatrix} j & 0 \\ 0 & j^2 \end{pmatrix}.$$

Il en résulte que

$$A^{12} = P \begin{pmatrix} j & 0 \\ 0 & j^2 \end{pmatrix}^{12} P^{-1} = P \begin{pmatrix} j^{12} & 0 \\ 0 & j^{24} \end{pmatrix} P^{-1} = I_2.$$

4. a) Comme 1 est l'unique valeur propre de A , c'est l'unique racine du trinôme de la question 1. On en déduit que celle-ci se réécrit sous la forme $(\lambda - 1)^2 = 0$. En remplaçant dans la relation $(*)$, on obtient $(A - I_2)^2 = 0$. On en déduit que $\mathfrak{S}m(A - I_2) \subset \text{Ker}(A - I_2)$. De plus, d'après la formule du rang, $\dim(\text{Ker}(A - I_2)) + \dim(\mathfrak{S}m(A - I_2)) = 2$. On ne peut avoir $\dim(\text{Ker}(A - I_2)) = 2$ sinon on aurait $A = I_2$. On ne peut avoir non plus $\dim(\text{Ker}(A - I_2)) = 0$, sinon 0 ne serait pas valeur propre. Il en découle que $\dim(\text{Ker}(A - I_2)) = \dim(\mathfrak{S}m(A - I_2)) = 1$, et donc $\text{Ker}(A - I_2) = \mathfrak{S}m(A - I_2)$.
- b) Soit X un vecteur non nul de $\mathfrak{S}m(A - I_2)$, il existe $Y \in \mathcal{M}_2(\mathbb{C})$ tel que $X = AY - Y$. Nous avons les deux relations

$$AX = X \text{ et } AY = X + Y.$$

Montrons que les vecteurs X et Y sont linéairement indépendants. Soient a, b deux réels tels que $aX + bY = 0$, on a alors $(A - I_2)(aX + bY) = a(A - I_2)X + b(A - I_2)Y = 0$, soit $bX = 0$. Comme $X \neq 0$, $b = 0$. Par suite $a = 0$. La famille (X, Y) est libre, comme elle compte deux vecteurs, c'est une base de $\mathcal{M}_2(\mathbb{C})$.

De ce qui précède, on déduit qu’en posant P la matrice de passage de la base canonique de $\mathcal{M}_2(\mathbb{C})$ vers la base (X, Y) , on a la relation :

$$P^{-1}AP = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

- c) C’est un classique, on prouve facilement que $T^k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ (par exemple par récurrence). On en déduit que

$$T^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} = PA^nP^{-1} = PI_2P^{-1} = I_2,$$

ce qui est absurde. L’hypothèse $\sigma = \{1\}$ est donc fausse.

5. En reprenant un raisonnement analogue on prouve que A est semblable à la matrice $U = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$, on montre que $U^n \neq I_2$ et on aboutit à la même contradiction.

7 1. Supposons que $\text{rg}(f) \leq p - 1$, alors $\mathfrak{S}m(f)$ est contenu dans un hyperplan. Il existe donc des réels a_1, \dots, a_p non tous nuls tels que $\mathfrak{S}m(f)$ soit contenu dans l’ensemble

$$H = \{(x_1, \dots, x_p) \mid a_1x_1 + \dots + a_px_p = 0\}$$

Cela signifie que pour tout $x \in E$, $\sum_{k=1}^p a_k \varphi_k(x) = 0$, soit $\sum_{k=1}^p \varphi_k = 0$. Cela est en contradiction avec l’indépendance linéaire de $\varphi_1, \dots, \varphi_p$.

On en conclut que $\text{rg}(f) = p$.

2. C’est immédiat puisque $f(x) = 0$ si, et seulement si, pour $k \in \llbracket 1, p \rrbracket$, $\varphi_k(x) = 0$.
3. D’après la formule du rang appliquée à f , on a $\dim(\text{Ker}(f)) = \dim(E) - \text{rg}(f) = n - p$, d’où la relation annoncée puisque $\text{Ker}(f) = \bigcap_{k=1}^p \text{Ker}(\varphi_k)$.

8 1. C’est un classique, déjà vu en première année.

Soient $\lambda_0, \dots, \lambda_{k-1}$ des scalaires tels que $\sum_{i=0}^{k-1} \lambda_i u^i(x_0) = 0$. Supposons que ces scalaires ne soient pas tous nuls et notons j le plus petit indice i tel que $\lambda_i \neq 0$. On a alors

$$\begin{aligned} u^{k-j-1} \left(\sum_{i=j}^{k-1} \lambda_i u^i(x_0) \right) &= \sum_{i=j}^{k-1} \lambda_i u^{k-j-1+i}(x_0) \\ &= \sum_{i=k-1}^{2k-2-j} \lambda_{i-k+1+j} \underbrace{u^i(x_0)}_{=0 \text{ pour } i \geq k} \\ &= \lambda_j u^{k-1}(x_0), \end{aligned}$$

mais aussi $u^{k-j-1} \left(\sum_{i=j}^{k-1} \lambda_i u^i(x_0) \right) = 0$. On en déduit finalement que $\lambda_j u^{k-1}(x_0) = 0$. Le vecteur $u^{k-1}(x_0)$ étant non nul, on a $\lambda_j = 0$. Ceci est absurde. Les scalaires $\lambda_0, \dots, \lambda_{k-1}$ sont nuls et la famille $(x_0, u(x_0), \dots, u^{k-1}(x_0))$ est libre.

2. Soit $x \in F$, par définition il existe des scalaires $\lambda_0, \dots, \lambda_{k-1}$ tels que $x = \sum_{i=0}^{k-1} \lambda_i u^i(x_0)$. On a alors

$$u(x) = u \left(\sum_{i=0}^{k-1} \lambda_i u^i(x_0) \right) = \sum_{i=0}^{k-1} \lambda_i u^{i+1}(x_0) = \sum_{i=0}^{k-2} \lambda_i u^{i+1}(x_0) = \sum_{i=1}^{k-1} \lambda_{i-1} u^i(x_0) \in F.$$

Le sous-espace F est bien stable par u .

3. Soit H un supplémentaire de $\text{Vect}(u^{k-1}(x_0))$, comme $\dim(\text{Vect}(u^{k-1}(x_0))) = 1$, H est un hyperplan et il existe une forme linéaire φ non nulle telle que $H = \text{Ker}(\varphi)$. On ne peut avoir $\varphi(u^{k-1}(x_0)) = 0$, sinon φ serait nulle. L'application φ répond à la question.
4. Soient $\lambda_0, \dots, \lambda_{k-1}$ des scalaires tels que $\sum_{i=0}^{k-1} \lambda_i \varphi \circ u^i = 0$, on a pour tout $x \in E$,

$$\sum_{i=0}^{k-1} \lambda_i \varphi(u^i(x)) = 0.$$

En appliquant cette égalité à $x = u^{k-1}(x_0)$ et en retirant les termes nuls, il reste

$$\lambda_{k-1} \varphi(u^{k-1}(x_0)) = 0,$$

le vecteur étant non nul, $\lambda_{k-1} = 0$.

En appliquant l'égalité successivement à $u^{k-2}(x_0), u^{k-3}(x_0), \dots, x_0$, on trouve que

$$\lambda_{k-2} = \lambda_{k-3} = \dots = \lambda_0 = 0.$$

La famille $(\varphi, \varphi \circ u, \dots, \varphi \circ u^{k-1})$ est libre.

5. Soit $x \in G$, on a alors pour $i \in \llbracket 0, k-1 \rrbracket$, $\varphi \circ u^i(u(x)) = \varphi(u^{i+1}(x))$.

Deux cas se présentent

– Soit $i < k-1$ et alors, comme $x \in G$, $\varphi(u^{i+1}(x)) = 0$.

– Soit $i = k-1$ et alors $\varphi(u^{i+1}(x)) = \varphi(0) = 0$.

Dans tous les cas $\varphi \circ u^i(u(x)) = 0$. Il en découle que $u(x) \in G$ et que G est stable par u .

6. Montrons que $F \cap G = \{0\}$. Soit $x \in F \cap G$. Comme $x \in F$, il existe des scalaires

$$\lambda_0, \dots, \lambda_{k-1} \text{ tels que } x = \sum_{i=0}^{k-1} \lambda_i u^i(x_0).$$

Comme $x \in G$, on a en particulier $x \in \text{Ker}(\varphi \circ u^{k-1})$, soit

$$(\varphi \circ u^{k-1})(x) = 0 = (\varphi \circ u^{k-1}) \left(\sum_{i=0}^{k-1} \lambda_i u^i(x_0) \right) = \sum_{i=0}^{k-1} \lambda_i \varphi(u^{k-1+i}(x_0)) = \lambda_{k-1} \varphi(u^{k-1}(x_0)).$$

Comme $\varphi(u^{k-1}(x_0)) \neq 0$, $\lambda_{k-1} = 0$. En évaluant ensuite $(\varphi \circ u^i)(x)$ avec $i = k-2, i = k-3, \dots, i = 0$, on prouve que $\lambda_{k-2} = \lambda_{k-3} = \dots = \lambda_0 = 0$. Il s’ensuit que $x = 0$.

Cela prouve que $F \cap G = \{0\}$.

De plus, d’après l’exercice précédent $\dim(G) = \dim(E) - k$ et d’après la question 1, $\dim(F) = k$. On en déduit que $\dim(F) + \dim(G) = k + \dim(E) - k = \dim(E)$. Les sous-espaces vectoriels F et G sont supplémentaires.

7. Pour $n = \dim(E) = 1$, la matrice de u dans la base canonique s’écrit sous la forme (a) où $a \in \mathbb{C}$. La matrice de u^k est (a^k) , donc $a^k = 0$, d’où $a = 0$. La matrice de u dans la base canonique est bien sous la forme annoncée.

Supposons la propriété vérifiée pour tout espace E de dimension inférieur ou égal à n et montrons la propriété pour tout espace de dimension $n + 1$.

Soit u un endomorphisme nilpotent d’un espace E de dimension $n + 1$ tel que $u^k = 0$ et $u^{k-1} \neq 0$ (où $k \in \mathbb{N}^*$). On reprend les notations des questions précédentes. Deux cas se présentent :

- Nous avons $\dim(F) = \dim(E)$ et alors $F = E$. La base $(x_0, \dots, u^{k-1}(x_0))$ est une base de E et dans cette base, la matrice de u est

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix},$$

qui est bien sous la forme voulue.

- Nous avons $\dim(F) < \dim(E)$. Dans ce cas $1 \leq \dim(F) \leq n$ et $1 \leq \dim(G) \leq n$. Or F et G sont stables par u . L’endomorphisme u induit donc un endomorphisme de F et un endomorphisme de G , qui sont tous les deux nilpotents. D’après l’hypothèse de récurrence, il existe une base \mathcal{B}_1 de F et une base \mathcal{B}_2 de G telles que

$$\mathcal{M}_{\mathcal{B}_1}(u|_F) = \begin{pmatrix} 0 & \varepsilon_1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & \varepsilon_{k-1} \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix} \text{ et } \mathcal{M}_{\mathcal{B}_2}(u|_G) = \begin{pmatrix} 0 & \varepsilon_k & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & \varepsilon_n \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix}$$

où pour $k \in \llbracket 1, n \rrbracket$, $\varepsilon_k \in \{-1, 1\}$. Soit \mathcal{B} la base de E obtenue en juxtaposant les vecteurs de la base \mathcal{B}_1 et de la base \mathcal{B}_2 . Dans cette base, la matrice de u est

$$\mathcal{M}_{\mathcal{B}}(u) = \begin{pmatrix} 0 & \varepsilon_1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & \varepsilon_n \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix}.$$

qui est bien sous la forme désirée.

La propriété est donc prouvée par récurrence.

9 **1.** Un petit calcul donne $B = \begin{pmatrix} 0 & -1 & 1 \\ -1 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix}$. B étant symétrique réelle, elle est diagonalisable.

2. Il suffit d'effectuer le calcul.

3. Les valeurs propres de B sont racines de $X^2 - X - 2$. Les réels -1 et 2 sont racines évidentes de ce trinôme.

De plus, pour $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, nous avons

$$X \in E_{-1}(B) \Leftrightarrow (B + I)X = 0 \Leftrightarrow \begin{cases} x - y + z = 0 \\ -x + y - z = 0 \\ x - y + z = 0 \end{cases} \Leftrightarrow x = y - z,$$

d'où $E_{-1}(B) = \text{Vect} \left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right)$ (en particulier -1 est bien une valeur propre).

De même

$$X \in E_2(B) \Leftrightarrow (B - 2I)X = 0 \Leftrightarrow \begin{cases} -2x - y + z = 0 \\ -x - 2y - z = 0 \\ x - y - 2z = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} -2x - y + z = 0 \\ 0 - 3y - 3z = 0 \\ -3y - 3z = 0 \end{cases} \begin{matrix} L_2 \leftarrow 2L_2 - L_1 \\ L_3 \leftarrow 2L_3 + L_1 \end{matrix} \Leftrightarrow \begin{cases} x = z \\ y = -z \end{cases}$$

d'où $E_2(B) = \text{Vect} \left(\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right)$ (et 2 est bien une valeur propre de B).

4. Soit λ une valeurs propre de A et X un vecteur propre associé, on a alors

$$BX = (A^2 + 2I)X = A^2X + 2X = \lambda^2X + 2X = (\lambda^2 + 2)X,$$

et donc $\lambda^2 + 2$ est valeur propre de B .

Comme $\lambda^2 + 2 \geq 2$ pour tout réel λ et que $\text{Sp}(B) = \{-1, 2\}$, la seule valeur propre réelle possible de A est 0 . Si A était diagonalisable sur \mathbb{R} , elle serait semblable à la matrice nulle. Elle serait donc nulle, ce qui est absurde. A n'est pas diagonalisable sur \mathbb{R} .

5. De la relation de la question 2, on déduit que $B \times \frac{1}{2}(B - I) = I$. La matrice B est donc inversible et $B^{-1} = \frac{1}{2}(B - I)$.

6. a) En évaluant l'égalité en -1 et 2 , on trouve que a_n et b_n sont solutions d'un système linéaire que l'on résout :

$$\begin{cases} -a_n + b_n = (-1)^n \\ 2a_n + b_n = 2^n \end{cases} \Leftrightarrow \begin{cases} -a_n + b_n = (-1)^n \\ 3b_n = 2^n + 2(-1)^n \end{matrix} \quad L_2 \leftarrow L_2 + 2L_1$$

$$\Leftrightarrow \begin{cases} -a_n + b_n = (-1)^n \\ 3b_n = 2^n + 2(-1)^n \end{matrix} \quad L_2 \leftarrow L_2 + 2L_1 \Leftrightarrow \begin{cases} a_n = \frac{1}{3}(2^n - (-1)^n) \\ b_n = \frac{1}{3}(2^n + 2(-1)^n) \end{cases}$$

b) En évaluant l'égalité précédente en B , il vient pour $n \geq 2$,

$$B^n = \frac{1}{3}(2^n - (-1)^n)B + \frac{1}{3}(2^n + 2(-1)^n)I,$$

formule également valable pour $n = 0$ et $n = 1$.

c) Soit $n \in \mathbb{N}^*$, on vérifie que B^n a pour inverse le candidat proposé en effectuant le calcul

$$\begin{aligned} & B^n \times \left(\frac{1}{3}(2^{-n} - (-1)^{-n})B + \frac{1}{3}(2^{-n} + 2(-1)^{-n})I \right) \\ &= \left(\frac{1}{3}(2^n - (-1)^n)B + \frac{1}{3}(2^n + 2(-1)^n)I \right) \times \left(\frac{1}{3}(2^{-n} - (-1)^{-n})B + \frac{1}{3}(2^{-n} + 2(-1)^{-n})I \right) \\ &= \frac{1}{9} \left((1+1 - (-2)^{-n} - (-2)^n)B^2 + (1-2 - (-2)^n + 2(-2)^{-n} + 1-2 + 2(-2)^n - (-2)^{-n})B \right. \\ &\quad \left. + (1+4 + 2(-2)^n + 2(-2)^{-n})I \right) \\ &= \frac{1}{9} \left((2 - (-2)^{-n} - (-2)^n)(B + 2I) + (-2 + (-2)^n + (-2)^{-n})B \right. \\ &\quad \left. + (5 + 2(-2)^n + 2(-2)^{-n})I \right) \\ &= I. \end{aligned}$$

La formule est donc encore valable pour les entiers négatifs.

10 1. Soit $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ un tel vecteur. Notons k l'entier tel que $x_k = \max_{1 \leq i \leq n} |x_i|$.

L'égalité $AX = 0$ signifie que pour tout $i \in \llbracket 1, n \rrbracket$,

$$\sum_{j=1}^n a_{i,j}x_j = 0.$$

Cette égalité est en particulier vrai pour $i = k$, il en découle que

$$a_{k,k}x_k = \sum_{\substack{1 \leq j \leq n \\ j \neq k}} a_{k,j}x_j.$$

On en déduit que

$$|a_{k,k}x_k| = \left| \sum_{\substack{1 \leq j \leq n \\ j \neq k}} a_{k,j}x_j \right| \leq \sum_{\substack{1 \leq j \leq n \\ j \neq k}} |a_{k,j}x_j| \leq \sum_{\substack{1 \leq j \leq n \\ j \neq k}} |a_{k,j}x_k| \leq |x_k| \sum_{\substack{1 \leq j \leq n \\ j \neq k}} |a_{k,j}|.$$

Or $x_k \neq 0$ (puisque $X \neq 0$), il en résulte que

$$|a_{k,k}| \leq \sum_{\substack{1 \leq j \leq n \\ j \neq k}} |a_{k,j}|,$$

ce qui est en contradiction avec les hypothèses de l'énoncé. Il n'existe donc pas de vecteur colonne non nul X tel que $AX = 0$ et la matrice A est inversible.

2. Soit $\lambda \in \mathbb{C}$ tel que $\lambda \notin \bigcup_{i=1}^n D(a_{i,i}, \sum_{j \neq i} |a_{i,j}|)$, notons $B = (b_{i,j})$ la matrice $A - \lambda I_n$. Montrons que cette matrice est inversible grâce à la première question.

Soit $i \in \llbracket 1, n \rrbracket$, comme $\lambda \notin D(a_{i,i}, \sum_{j \neq i} |a_{i,j}|)$, nous avons

$$|b_{i,i}| = |a_{i,i} - \lambda| > \sum_{j \neq i} |a_{i,j}| = \sum_{j \neq i} |b_{i,j}|.$$

D'après la question 1, il en résulte que B est inversible. Le scalaire λ n'est donc pas valeur propre.

Il s'ensuit que

$$\text{Sp}(A) \subset \bigcup_{i=1}^n D(a_{i,i}, \sum_{j \neq i} |a_{i,j}|).$$

3. a) D'après la question précédente toute valeur propre λ de A est dans le disque $D(0, 1) \cup D(0, 2) = D(0, 2)$. Cela signifie que tout valeur propre λ vérifie $|\lambda| \leq 2$.

- b) On recherche un vecteur non nul $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ tel que $(A - 2 \cos(\theta)I)X = 0$, c'est-

à-dire tel que

$$\begin{cases} -2 \cos(\theta)x_1 + x_2 = 0 \\ x_1 - 2 \cos(\theta)x_2 + x_3 = 0 \\ x_2 - 2 \cos(\theta)x_3 + x_4 = 0 \\ \vdots \\ x_{n-2} - 2 \cos(\theta)x_{n-1} + x_n = 0 \\ x_{n-1} - 2 \cos(\theta)x_n = 0 \end{cases}$$

Posons $x_0 = 0$ et $x_{n+1} = 0$. La suite (x_0, \dots, x_{n+1}) vérifie

$$\forall k \in \llbracket 0, n-1 \rrbracket, x_k - 2 \cos(\theta)x_{k+1} + x_{k+2} = 0,$$

c'est-à-dire que c'est une suite récurrente linéaire d'ordre 2. Son équation caractéristique admet pour racines $e^{i\theta}$ et $e^{-i\theta}$.

Supposons dans un premier temps que ces deux racines soient distinctes, c'est-à-dire que $\theta \in]0, \pi[$. Il existe donc deux scalaires a et b tels que pour $k \in \llbracket 0, n+1 \rrbracket$,

$$x_k = ae^{ki\theta} + be^{-ki\theta}.$$

En exprimant ces égalités pour $k = 0$ et $k = n+1$, il vient

$$\begin{cases} a + b = 0 \\ ae^{(n+1)i\theta} + be^{-(n+1)i\theta} = 0 \end{cases}$$

Ce système est inversible si $e^{(n+1)i\theta} \neq e^{-(n+1)i\theta}$, auquel cas $a = b = 0$ et donc $x_k = 0$. Autrement dit pour un tel θ , λ n'est pas valeur propre.

Supposons que $e^{(n+1)i\theta} = e^{-(n+1)i\theta}$, ce qui équivaut à $2(n+1)\theta \in 2\pi\mathbb{Z}$. Compte tenu de la condition $\theta \in]0, \pi[$, il y a donc n valeurs possibles pour θ à savoir

$$\theta_p = \frac{p\pi}{n+1} \quad \text{où } p \in \llbracket 1, n \rrbracket.$$

Pour $\theta = \theta_p$, le système équivaut à $a = -b$, et alors il existe une solution non nulle au système $(A - 2 \cos(\theta)I)X = 0$, à savoir par exemple $x_k = \frac{1}{2i}(e^{ki\theta_p} - e^{-ki\theta_p}) = \sin(k\theta_p)$.

Il en résulte que pour $p \in \llbracket 1, n \rrbracket$, $2 \cos\left(\frac{p\pi}{n+1}\right)$ est une valeur propre et un vecteur propre associé est

$$\begin{pmatrix} \sin\left(\frac{p\pi}{n+1}\right) \\ \vdots \\ \sin\left(\frac{np\pi}{n+1}\right) \end{pmatrix}.$$

Nous avons déjà n valeurs propres, donc il ne peut y en avoir d'autres et les sous-espaces propres sont de dimension 1. On en conclut que

$$E_{2 \cos\left(\frac{p\pi}{n+1}\right)}(A) = \text{Vect} \begin{pmatrix} \sin\left(\frac{p\pi}{n+1}\right) \\ \sin\left(\frac{2p\pi}{n+1}\right) \\ \vdots \\ \sin\left(\frac{np\pi}{n+1}\right) \end{pmatrix}.$$

11 1. On vérifie les trois points

- E est une partie de $\mathcal{M}_n(\mathbb{R})$.
- E est non vide puisqu'il contient par exemple la matrice nulle.
- Soient $A = (a_{i,j})$ et $B = (b_{i,j})$ deux matrices de E et λ un réel. Notons $C = (c_{i,j})$ la matrice $\lambda A + B$ et montrons que $C \in E$.
Pour $i \in \llbracket 1, n \rrbracket$, on a

$$\sum_{k=1}^n c_{i,k} = \sum_{k=1}^n \lambda a_{i,k} + b_{i,k} = \lambda \sum_{k=1}^n a_{i,k} + \sum_{k=1}^n b_{i,k} = \lambda m(A) + m(B).$$

et pour $j \in \llbracket 1, n \rrbracket$

$$\sum_{k=1}^n c_{k,j} = \sum_{k=1}^n \lambda a_{k,j} + b_{k,j} = \lambda \sum_{k=1}^n a_{k,j} + \sum_{k=1}^n b_{k,j} = \lambda m(A) + m(B).$$

Donc $C \in E$.

E est bien un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{R})$.

Soit $A = (a_{i,j})$ une matrice de E , en notant $D = (d_{i,j})$ la matrice AJ , on a, pour tout $(i,j) \in \llbracket 1, n \rrbracket^2$,

$$d_{i,j} = \sum_{k=1}^n a_{i,k} j_{k,j} = \sum_{k=1}^n a_{i,k} = m(A).$$

Autrement dit $AJ = m(A)J$. On montrerait de même que $JA = m(A)J$.

Soient A, B deux matrices de E , d'après ce qui précède $JAB = m(A)JB = m(A)m(B)J$.

Si l'on note $C = (c_{i,j})$ la matrice AB , cela signifie que pour $j \in \llbracket 1, n \rrbracket$,

$$\sum_{k=1}^n c_{k,j} = m(A)m(B).$$

On montrerait de même, à l'aide du produit ABJ que pour $i \in \llbracket 1, n \rrbracket$,

$$\sum_{k=1}^n c_{i,k} = m(A)m(B).$$

On en déduit que $AB \in E$ (et d'ailleurs $m(AB) = m(A)m(B)$).

La linéarité de m a déjà été démontrée lorsque l'on a prouvé que E était un sous-espace vectoriel.

2. Soit $A \in G \cap H$. Il existe un réel λ tel que $A = \lambda J$. On a alors $m(A) = n\lambda$. Or $m(A) = 0$, donc $\lambda = 0$ et finalement $A = 0$. Il en résulte que $G \cap H = \{0\}$.

Montrons que $E = G + H$. Soit $A \in E$, on a $A = A - \frac{m(A)}{n}J + \frac{m(A)}{n}J$.

Or

$$m\left(A - \frac{m(A)}{n}J\right) = m(A) - \frac{m(A)}{n}m(J) = m(A) - m(A) = 0,$$

donc $A - \frac{m(A)}{n}J \in H$ et $\frac{m(A)}{n}J \in G$. Cela prouve que $E = G + H$.
Il s'ensuit que $E = G \oplus H$.

3. Pour fixer les idées écrivons la matrice $H_{k,l}$:

$$H_{k,l} = \begin{pmatrix} 1 & \dots & -1 & 0 & \dots & 0 \\ \vdots & 0 & \vdots & \vdots & & \vdots \\ -1 & \dots & 1 & \vdots & & \vdots \\ 0 & \dots & \dots & 0 & & \vdots \\ \vdots & & & & & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 \end{pmatrix}$$

(où les -1 sont en k^e ligne, première colonne et première ligne, l^e colonne).

La somme des coefficients d'une ligne ou d'une colonne quelconque de $H_{k,l}$ est nulle, ce qui signifie que $H_{k,l} \in H$.

Montrons que la famille $(H_{k,l})$ est libre. Soit $(a_{k,l})$ des réels tels que

$$\sum_{2 \leq k,l \leq n} a_{k,l} H_{k,l} = 0.$$

Soit $(i,j) \in \llbracket 2, n \rrbracket^2$, le seul coefficient non nul en i^e ligne, j^e colonne provient de $H_{i,j}$. On en déduit que $a_{i,j} = 0$. Cela prouve que $(H_{k,l})$ est une famille libre.

Montrons que cette famille est génératrice. Soit $A = (a_{i,j}) \in E$, suivons les conseils de l'énoncé en posant

$$A' = \sum_{2 \leq k,l \leq n} a_{k,l} H_{k,l}.$$

Montrons que $A' = A$. Pour $(i,j) \in \llbracket 2, n \rrbracket^2$, pour des raisons déjà évoquées, $a'_{i,j} = a_{i,j}$.
De plus

$$a'_{i,1} = - \sum_{k=2}^n a_{i,k} = a_{i,1} - m(A) = a_{i,1},$$

$$\text{et } a'_{1,j} = - \sum_{k=2}^n a_{k,j} = a_{1,j} - m(A) = a_{1,j}.$$

Il reste un coefficient, $a'_{1,1}$:

$$\begin{aligned} a'_{1,1} &= \sum_{2 \leq k, l \leq n} a_{k,l} = \sum_{k=2}^n \sum_{l=2}^n a_{k,l} \\ &= \sum_{k=2}^n -a_{k,1} = a_{1,1}. \end{aligned}$$

Nous avons bien $A' = A$. La famille $(H_{k,l})$ est donc génératrice et par conséquent, c'est une base de H .

Chapitre 2

1 Supposons que A soit une telle matrice. A étant symétrique, il existe une matrice inversible P telle que $P^{-1}AP = D$ où D est du type $D = \text{Diag}(\lambda_1, \dots, \lambda_n)$. On a donc $D^k = \text{Diag}(\lambda_1^k, \dots, \lambda_n^k) = P^{-1}A^kP = 0$. Il en résulte que $\lambda_1 = \dots = \lambda_n = 0$, et par voie de conséquence $A = 0$, ce qui contredit $A^{k-1} \neq 0$. Une telle matrice A n'existe donc pas.

2 Posons $n = \dim(E)$, on sait déjà que

$$\dim(F^\perp) + \dim(G^\perp) = n - \dim(F) + n - \dim(G) = 2n - n = n.$$

Montrons que $F^\perp \cap G^\perp = \{0\}$. Soit $x \in F^\perp \cap G^\perp$, comme $E = F \oplus G$, il existe $(y, z) \in F \times G$ tel que $x = y + z$. De $x \in F^\perp$, il résulte que $\langle x, y \rangle = 0$ et de $x \in G^\perp$, $\langle x, z \rangle = 0$. Par bilinéarité, il vient $\|x\|^2 = \langle x, x \rangle = \langle x, y \rangle + \langle x, z \rangle = 0$. Autrement dit $x = 0$. On en déduit que $E = F^\perp \oplus G^\perp$.

3 1. Vérifions les trois points :

- F est une partie de E par définition.
- F est non vide puisqu'il contient le polynôme nul.
- Soient $P, Q \in F$ et $\lambda \in \mathbb{R}$, on a

$$(\lambda P + Q)(1) = \lambda P(1) + Q(1) = 0,$$

d'où $\lambda P + Q \in E$.

On en déduit que F est un sous-espace vectoriel de E .

2. On a

$$\begin{aligned} F &= \{aX^3 + bX^2 + cX + d \mid a + b + c + d = 0\} \\ &= \{aX^3 + bX^2 + cX - a - b - c \mid a, b, c \in \mathbb{R}\} \\ &= \text{Vect}(X^3 - 1, X^2 - 1, X - 1). \end{aligned}$$

Une base de F est $(e_1, e_2, e_3) = (X^3 - 1, X^2 - 1, X - 1)$. On orthonormalise cette famille. Nous avons

$$f_1 = \frac{1}{\|X^3 - 1\|} (X^3 - 1) = \frac{1}{\sqrt{2}} (X^3 - 1).$$

Ensuite

$$\begin{aligned} e_2 - \langle e_2, f_1 \rangle f_1 &= X^2 - 1 - \frac{1}{2} \langle X^2 - 1, X^3 - 1 \rangle (X^3 - 1) \\ &= X^2 - 1 - \frac{1}{2} (X^3 - 1) = -\frac{1}{2} X^3 + X^2 - \frac{1}{2}, \end{aligned}$$

or

$$\left\| -\frac{1}{2} X^3 + X^2 - \frac{1}{2} \right\|^2 = \frac{1}{4} + 1 + \frac{1}{4} = \frac{3}{2}.$$

D'où

$$f_2 = \frac{1}{2} \sqrt{\frac{2}{3}} (-X^3 + 2X^2 - 1) = \frac{1}{\sqrt{6}} (-X^3 + 2X^2 - 1).$$

Enfin,

$$\begin{aligned} e_3 - \langle e_3, f_2 \rangle f_2 - \langle e_3, f_1 \rangle f_1 &= X - 1 - \frac{1}{6} \langle X - 1, -X^3 + 2X^2 - 1 \rangle (-X^3 + 2X^2 - 1) \\ &\quad - \frac{1}{2} \langle X - 1, X^3 - 1 \rangle (X^3 - 1) \\ &= X - 1 - \frac{1}{6} (-X^3 + 2X^2 - 1) - \frac{1}{2} (X^3 - 1) \\ &= -\frac{1}{3} X^3 - \frac{1}{3} X^2 + X - \frac{1}{3} = -\frac{1}{3} (X^3 + X^2 - 3X + 1). \end{aligned}$$

D'où par exemple $f_3 = \frac{1}{\sqrt{12}} (X^3 + X^2 - 3X + 1)$.

3. Si l'on note u ce projeté orthogonal, nous avons

$$\begin{aligned} u &= \langle X, f_1 \rangle f_1 + \langle X, f_2 \rangle f_2 + \langle X, f_3 \rangle f_3 \\ &= \frac{1}{2} \langle X, X^3 - 1 \rangle (X^3 - 1) + \frac{1}{6} \langle X, -X^3 + 2X^2 - 1 \rangle (-X^3 + 2X^2 - 1) \\ &\quad + \frac{1}{12} \langle X, X^3 + X^2 - 3X + 1 \rangle (X^3 + X^2 - 3X + 1) \\ &= -\frac{1}{4} (X^3 + X^2 - 3X + 1). \end{aligned}$$

4 – Supposons que p soit un projecteur orthogonal. Les sous-espaces $\text{Ker}(p)$ et $\mathfrak{Im}(p)$ sont alors des sous-espaces supplémentaires orthogonaux. Soit \mathcal{B}_1 une base orthonormale de $\text{Ker}(p)$, \mathcal{B}_2 une base orthonormale de $\mathfrak{Im}(p)$ et \mathcal{B} la famille obtenue en juxtaposant ces deux bases. \mathcal{B} est une base orthonormale de E et cette base est une base de vecteurs propres de p . On en déduit que p est symétrique.

- Supposons que p soit symétrique, alors on sait que les sous-espaces propres de p sont orthogonaux. Or $\text{Ker}(p) = E_0(p)$ et $\mathfrak{Im}(p) = E_1(p)$, donc $\text{Ker}(p)$ et $\mathfrak{Im}(p)$ sont des supplémentaires orthogonaux et p est un projecteur orthogonal.

5 1. A est diagonalisable car c'est une matrice symétrique réelle.

2. Triangulons $A - \lambda I_3$

$$\begin{aligned} \begin{pmatrix} -2-\lambda & 1 & 1 \\ 1 & -2-\lambda & -1 \\ 1 & -1 & -1-\lambda \end{pmatrix} &\rightarrow \begin{pmatrix} 1 & -1 & -1-\lambda \\ 1 & -2-\lambda & -1 \\ -2-\lambda & 1 & 1 \end{pmatrix} \quad L_1 \leftrightarrow L_3 \\ &\rightarrow \begin{pmatrix} 1 & -1 & -1-\lambda \\ 0 & -1-\lambda & \lambda \\ 0 & -1-\lambda & -1-3\lambda-\lambda^2 \end{pmatrix} \quad \begin{array}{l} L_2 \leftarrow L_2 - L_1 \\ L_3 \leftarrow L_3 + (2+\lambda)L_1 \end{array} \\ &\rightarrow A(\lambda) = \begin{pmatrix} 1 & -1 & -1-\lambda \\ 0 & -1-\lambda & \lambda \\ 0 & 0 & -1-4\lambda-\lambda^2 \end{pmatrix} \quad L_3 \leftarrow L_3 - L_1, \end{aligned}$$

les valeurs propres de A sont donc -1 et les solutions de l'équation $\lambda^2 + 4\lambda + 1 = 0$.
On a $\Delta' = 4 - 1 = \sqrt{3}$. On en déduit que

$$\text{Sp}(A) = \{-2 - \sqrt{3}, -1, -2 + \sqrt{3}\}.$$

Soit $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, on a

$$\begin{aligned} X \in E_{-2-\sqrt{3}} &\Leftrightarrow A(-2-\sqrt{3})X = 0 \Leftrightarrow \begin{cases} x - y + (1 + \sqrt{3})z = 0 \\ (1 + \sqrt{3})y - (2 + \sqrt{3})z = 0 \end{cases} \\ &\Leftrightarrow \begin{cases} x - y + (1 + \sqrt{3})z = 0 \\ y = \frac{2+\sqrt{3}}{1+\sqrt{3}}z \end{cases} \Leftrightarrow \begin{cases} x = y - (1 + \sqrt{3})z \\ y = \frac{2+\sqrt{3}}{1+\sqrt{3}}z \end{cases} \\ &\Leftrightarrow \begin{cases} x = \frac{y - (1 + \sqrt{3})z}{1} \\ y = \frac{(\sqrt{3}-1)(2+\sqrt{3})}{2}z \end{cases} \Leftrightarrow \begin{cases} x = \frac{y - (1 + \sqrt{3})z}{1} \\ y = \frac{1+\sqrt{3}}{2}z \end{cases} \\ &\Leftrightarrow \begin{cases} x = \frac{-1-\sqrt{3}}{2}z \\ y = \frac{1+\sqrt{3}}{2}z, \end{cases} \end{aligned}$$

D'où $E_{-2-\sqrt{3}}(A) = \text{Vect} \left(\begin{pmatrix} -1 \\ 1 \\ \sqrt{3}-1 \end{pmatrix} \right)$.

De même,

$$X \in E_{-1}(A) \Leftrightarrow A(-1)X = 0 \Leftrightarrow \begin{cases} x - y = 0 \\ -z = 0 \\ -z = 0 \end{cases} \Leftrightarrow \begin{cases} x = y \\ z = 0, \end{cases}$$

d'où $E_{-1}(A) = \text{Vect} \left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right)$.

Et enfin,

$$\begin{aligned} X \in E_{-2+\sqrt{3}} &\Leftrightarrow A(-2+\sqrt{3})X = 0 \Leftrightarrow \begin{cases} x - y + (1 - \sqrt{3})z = 0 \\ (1 - \sqrt{3})y - (2 - \sqrt{3})z = 0 \end{cases} \\ &\Leftrightarrow \begin{cases} x - y + (1 - \sqrt{3})z = 0 \\ y = \frac{2 - \sqrt{3}}{1 - \sqrt{3}}z \end{cases} \Leftrightarrow \begin{cases} x = y - (1 - \sqrt{3})z \\ y = \frac{2 - \sqrt{3}}{1 - \sqrt{3}}z \end{cases} \\ &\Leftrightarrow \begin{cases} x = \frac{y - (1 - \sqrt{3})z}{1} \\ y = \frac{(1 + \sqrt{3})(2 - \sqrt{3})}{-2}z \end{cases} \Leftrightarrow \begin{cases} x = \frac{y - (1 - \sqrt{3})z}{1} \\ y = \frac{1 - \sqrt{3}}{2}z \end{cases} \\ &\Leftrightarrow \begin{cases} x = \frac{-1 + \sqrt{3}}{2}z \\ y = \frac{1 - \sqrt{3}}{2}z, \end{cases} \end{aligned}$$

$$\text{D'où } E_{-2+\sqrt{3}}(A) = \text{Vect} \left(\begin{pmatrix} 1 \\ -1 \\ 1 + \sqrt{3} \end{pmatrix} \right).$$

Les sous-espaces propres de A étant orthogonaux, on sait déjà que la famille

$$\left(\begin{pmatrix} -1 \\ 1 \\ \sqrt{3} - 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 + \sqrt{3} \end{pmatrix} \right)$$

est une base orthogonale formée de vecteurs propres de A . En la normalisant, on trouve que la famille

$$\left(\frac{1}{\sqrt{6 - 2\sqrt{3}}} \begin{pmatrix} -1 \\ 1 \\ \sqrt{3} - 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{6 + 2\sqrt{3}}} \begin{pmatrix} 1 \\ -1 \\ 1 + \sqrt{3} \end{pmatrix} \right)$$

est une base orthonormale de $\mathcal{M}_{3,1}(\mathbb{R})$ formée de vecteurs propres de A .

6 1. Notons φ cette application. La bilinéarité de φ ne pose pas de problème.

Montrons que φ est symétrique. Soient $x, y \in \mathbb{R}^n$,

$$\varphi(x, y) = \langle x, g(y) \rangle = {}^t X \cdot ({}^t AAY) = {}^t X^t A \cdot AY = ({}^t AX) \cdot AY = \langle f(x), f(y) \rangle.$$

On montrerait de la même façon que $\varphi(y, x) = \langle f(y), f(x) \rangle$, d'où $\varphi(x, y) = \varphi(y, x)$. φ est symétrique.

Soit $x \in \mathbb{R}^n$, d'après le calcul précédent, $\varphi(x, x) = \langle f(x), f(x) \rangle = \|f(x)\|^2 \geq 0$. φ est positive.

Enfin, si $\varphi(x, x) = 0$ alors $\|f(x)\| = 0$, c'est-à-dire $f(x) = 0$. Comme A est inversible, f est bijective et donc $x = 0$. On en déduit que φ est définie.

- Pour $x, y \in \mathbb{R}^n$, $\langle x, g(y) \rangle = \varphi(x, y) = \varphi(y, x) = \langle y, g(x) \rangle = \langle g(x), y \rangle$. L'endomorphisme g est symétrique.
- L'endomorphisme g étant symétrique, il existe une base orthonormale (e_1, \dots, e_n) de \mathbb{R}^n formée de vecteurs propres de g . Notons $\lambda_1, \dots, \lambda_n$ les valeurs propres associées, on a alors pour tout couple $(i, j) \in \llbracket 1, n \rrbracket^n$,

$$\langle f(e_i), f(e_j) \rangle = \langle e_i, g(e_j) \rangle = \langle e_i, \lambda_j e_j \rangle = \lambda_j \langle e_i, e_j \rangle = \lambda_j \delta_{i,j}.$$

Il s'ensuit que la famille $(f(e_1), \dots, f(e_n))$ est orthogonale.

- 7** 1. Soit $x = (a_1, \dots, a_n) \in \mathbb{R}^n$, et $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ où $y_k = 1$ si $a_k \geq 0$ et $y_k = -1$ sinon. L'inégalité de Cauchy-Schwarz appliquée à x et y donne $|\langle x, y \rangle| \leq \|x\| \|y\|$, ce qui s'écrit

$$\sum_{k=1}^n |a_k| \leq \sqrt{n} \sqrt{\sum_{k=1}^n a_k^2}.$$

2. Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base orthonormale de E et x un vecteur de E , de coordonnées (a_1, \dots, a_n) dans la base \mathcal{B} . Nous avons

$$\|x\| = \sqrt{\sum_{k=1}^n a_k^2}$$

et

$$\begin{aligned} \|p(x)\| &= \left\| \sum_{k=1}^n a_k p(e_k) \right\| \leq \sum_{k=1}^n |a_k| \|p(e_k)\| \leq \max_{k \in \llbracket 1, n \rrbracket} \|p(e_k)\| \sum_{k=1}^n |a_k| \\ &\leq \max_{k \in \llbracket 1, n \rrbracket} \|p(e_k)\| \sqrt{n} \sqrt{\sum_{k=1}^n a_k^2}. \end{aligned}$$

On en déduit que pour $x \neq 0$,

$$\frac{\|p(x)\|}{\|x\|} \leq \frac{\max_{k \in \llbracket 1, n \rrbracket} \|p(e_k)\| \times \sqrt{n} \sqrt{\sum_{k=1}^n a_k^2}}{\sqrt{\sum_{k=1}^n a_k^2}} = \sqrt{n} \times \max_{k \in \llbracket 1, n \rrbracket} \|p(e_k)\|,$$

où le majorant est indépendant de x .

3. Si $p \neq 0$, comme $E = \text{Ker}(p) \oplus \mathfrak{Im}(p)$, $\dim(\mathfrak{Im}(p)) \neq 0$. Il existe donc un vecteur non nul x de $\mathfrak{Im}(p)$. Pour un tel vecteur $p(x) = x$, soit $\frac{\|p(x)\|}{\|x\|} = 1$, d'où $\|p\| \geq 1$.
4. Signalons que cette question exclut le cas $p = 0$.

– Supposons que p soit orthogonal. Les sous-espaces $\text{Ker}(p)$ et $\mathfrak{Im}(p)$ sont supplémentaires orthogonaux.

Soit $x \in E \setminus \{0\}$, avec $x = y + z$ où $(y, z) \in \text{Ker}(p) \times \mathfrak{Im}(p)$.

Si $z = 0$, alors $p(x) = z = 0$ et $\frac{\|p(x)\|}{\|x\|} = 0$. Sinon, nous avons :

$$\frac{\|p(x)\|}{\|x\|} = \frac{\|p(y+z)\|}{\|y+z\|} = \frac{\|z\|}{\sqrt{\|y\|^2 + \|z\|^2}} \leq \frac{\|z\|}{\|z\|} = 1.$$

Dans tous les cas $\frac{\|p(x)\|}{\|x\|} \leq 1$.

Il en découle que $\|p\| \leq 1$. Or on sait que $\|p\| \geq 1$. Il en résulte que $\|p\| = 1$.

- Supposons que p ne soit pas orthogonal, alors il existe $(x, y) \in \text{Ker}(p) \times \mathfrak{Im}(p)$ tel que $\langle x, y \rangle \neq 0$. On a alors pour tout réel λ ,

$$\|\lambda x + y\|^2 = \lambda^2 \|x\|^2 + 2\lambda \langle x, y \rangle + \|y\|^2.$$

En particulier au voisinage de 0, $\|\lambda x + y\|^2 - \|y\|^2 \sim 2\lambda \langle x, y \rangle$. Donc il existe λ_0 tel que $\|\lambda_0 x + y\|^2 < \|y\|^2$. En posant $z = \lambda_0 x + y$, on a alors

$$\frac{\|p(z)\|}{\|z\|} = \frac{\|y\|}{\|\lambda_0 x + y\|} > 1.$$

Il en résulte que $\|p\| > 1$.

- 8** On voit que $F^\perp = \text{Vect}(-1, 1, 1)$. On en déduit que si l'on note u la projection orthogonale sur F^\perp , on a

$$u(x, y, z) = \frac{1}{3} \langle (x, y, z), (-1, 1, 1) \rangle (-1, 1, 1) = \frac{-x + y + z}{3} (-1, 1, 1).$$

D'où

$$B = \frac{1}{3} \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix} \quad \text{et} \quad A = I_3 - B = \frac{1}{3} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix}.$$

- 9** 1. Soit $x \in E$ tel que $f(x) = 0$, alors $\|x\| = \|f(x)\| = 0$ et donc $x = 0$. Autrement dit f est injectif. Comme E est de dimension finie, f est bijective.
Soit $x \in E$, f étant une isométrie, $\|f(f^{-1}(x))\| = \|f^{-1}(x)\|$, soit encore $\|x\| = \|f^{-1}(x)\|$. L'endomorphisme f^{-1} est donc également une isométrie.

2. Une isométrie étant bijective, un projecteur qui est une isométrie doit vérifier $\text{Ker}(p) = \{0\}$. On a alors $\mathfrak{Im}(p) = E$ et $p = \text{Id}_E$. Réciproquement, Id_E est bien un projecteur orthogonal qui est une isométrie.
En conclusion, le seul projecteur orthogonal qui soit une isométrie est Id_E .
3. On prendra garde au fait que f n'est pas supposée linéaire dans cette question.
– Supposons que f vérifie la seconde propriété.
Remarquons dans un premier temps que

$$\forall x \in E, \|f(x)\|^2 = \langle f(x), f(x) \rangle = \langle x, x \rangle = \|x\|^2,$$

autrement dit, f conserve la norme.

Montrons ensuite la linéarité de f . Soient $(x, y) \in E^2$ et $\lambda \in \mathbb{R}$,

$$\begin{aligned} \|f(x+y) - f(x) - f(y)\|^2 &= \|f(x+y)\|^2 + \|f(x)\|^2 + \|f(y)\|^2 - 2\langle f(x+y), f(x) \rangle \\ &\quad - 2\langle f(x+y), f(y) \rangle + 2\langle f(x), f(y) \rangle \\ &= \|x+y\|^2 + \|x\|^2 + \|y\|^2 - 2\langle x+y, x \rangle \\ &\quad - 2\langle x+y, y \rangle + 2\langle x, y \rangle \\ &= \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle + \|x\|^2 + \|y\|^2 - 2\|x\|^2 - 2\langle y, x \rangle \\ &\quad - 2\langle x, y \rangle - 2\|y\|^2 + 2\langle x, y \rangle \\ &= 0. \end{aligned}$$

$$\begin{aligned} \text{et} \quad \|f(\lambda x) - \lambda f(x)\|^2 &= \|f(\lambda x)\|^2 - 2\lambda \langle f(\lambda x), f(x) \rangle + \lambda^2 \|f(x)\|^2 \\ &= \|\lambda x\|^2 - 2\lambda \langle \lambda x, x \rangle + \lambda^2 \|x\|^2 \\ &= \lambda^2 \|x\|^2 - 2\lambda^2 \langle x, x \rangle + \lambda^2 \|x\|^2 = 0. \end{aligned}$$

C'est-à-dire que $f(x + y) = f(x) + f(y)$ et $f(\lambda x) = \lambda f(x)$.

Il en résulte que f est une isométrie.

- Réciproquement si f est une isométrie, alors pour tout couple $(x, y) \in E^2$,

$$\begin{aligned} \langle f(x), f(y) \rangle &= \frac{1}{4} \left(\|f(x) + f(y)\|^2 - \|f(x) - f(y)\|^2 \right) = \frac{1}{4} \left(\|f(x + y)\|^2 - \|f(x - y)\|^2 \right) \\ &= \frac{1}{4} \left(\|x + y\|^2 - \|x - y\|^2 \right) = \langle x, y \rangle. \end{aligned}$$

4. – Supposons que f soit une isométrie. Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base orthonormale de E , dans laquelle la matrice de f est notée $A = (a_{i,j})$. Nous avons alors, d'après la question précédente, pour tout $(i, j) \in \llbracket 1, n \rrbracket^2$,

$$\langle f(e_i), f(e_j) \rangle = \langle e_i, e_j \rangle = \delta_{i,j} = \sum_{k=1}^n a_{k,i} a_{k,j}.$$

Cette relation signifie précisément que ${}^tAA = I_n$.

- Réciproquement, supposons que $\mathcal{B} = (e_1, \dots, e_n)$ soit une base orthonormale de E dans laquelle la matrice de f , notée $A = (a_{i,j})$ soit orthogonale.

Nous avons

$$\langle f(e_i), f(e_j) \rangle = \sum_{k=1}^n a_{k,i} a_{k,j} = \delta_{i,j}.$$

Autrement dit, la famille $(f(e_1), \dots, f(e_n))$ est elle aussi orthonormale. Soit x un vecteur de E s'écrivant $x = \sum_{k=1}^n \lambda_k e_k$, d'après les propriétés des familles orthonormales, on a

$$\|x\|^2 = \sum_{k=1}^n \lambda_k^2$$

et

$$\|f(x)\|^2 = \left\| \sum_{k=1}^n \lambda_k f(e_k) \right\|^2 = \sum_{k=1}^n \lambda_k^2 = \|x\|^2.$$

L'application f est une isométrie.

5. Soit $x \in F$, et $\gamma \in F^\perp$, alors

$$\langle x, f(\gamma) \rangle = \langle f(f^{-1}(x)), f(\gamma) \rangle = \langle f^{-1}(x), \gamma \rangle,$$

or F étant stable par f , f induit un endomorphisme de F qui est injectif. Comme F est de dimension finie, f induit un automorphisme de F . Autrement dit, pour tout $x \in F$, il existe $\gamma \in F$ tel que $f(\gamma) = x$, et alors $f^{-1}(x) = \gamma \in F$. Cela prouve que F est stable par f^{-1} . Il en résulte que $\langle x, f(\gamma) \rangle = \langle f^{-1}(x), \gamma \rangle = 0$. Autrement dit F^\perp est stable par f .

6. Soit $(x, y) \in E^2$,

$$\begin{aligned} \langle x, (f + f^{-1})(y) \rangle &= \langle x, f(y) + f^{-1}(y) \rangle = \langle x, f(y) \rangle + \langle x, f^{-1}(y) \rangle \\ &= \langle f(f^{-1}(x)), f(y) \rangle + \langle f^{-1}(f(y)), f^{-1}(y) \rangle \\ &= \langle f^{-1}(x), y \rangle + \langle f(x), y \rangle = \langle (f + f^{-1})(x), y \rangle. \end{aligned}$$

L'endomorphisme $f + f^{-1}$ est symétrique.

7. Il suffit de prouver que $f^2(x) \in \text{Vect}(x, f(x))$.

Soit λ le réel tel que $(f + f^{-1})(x) = \lambda x$, en appliquant f à l'égalité, nous obtenons

$$f^2(x) + x = \lambda f(x),$$

soit

$$f^2(x) = \lambda f(x) - x.$$

8. Soit $\mathcal{B} = (e_1, e_2)$ une base orthonormale de E et $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ la matrice de f dans cette base.

On sait que $\|f(e_1)\| = \|f(e_2)\| = 1$ et que $\langle f(e_1), f(e_2) \rangle = \langle e_1, e_2 \rangle = 0$. On en déduit que les réels a, b, c, d vérifient

$$\begin{cases} a^2 + c^2 = 1 \\ b^2 + d^2 = 1 \\ ab + cd = 0 \end{cases}$$

En particulier $a \in [-1, 1]$. Il existe donc un réel $\alpha \in [0, \pi]$ tel que $a = \cos(\alpha)$.

On a alors $c^2 = 1 - \cos^2(\alpha) = \sin^2(\alpha)$, soit $c = \pm \sin(\alpha)$.

Posons $\theta = \alpha$ si $c = \sin(\alpha)$ et $\theta = -\alpha$ si $c = -\sin(\alpha)$. On a alors $a = \cos(\theta)$ et $c = \sin(\theta)$.

Comme f n'a pas de vecteur propre, $c \neq 0$ et donc

$$d = -\frac{a}{c}b = -\cotan(\theta)b.$$

On en déduit que $b^2 + \cotan^2(\theta)b^2 = 1$, soit $b^2 = \sin^2(\theta)$.

- Si $b = -\sin(\theta)$, on a alors $d = \cotan(\theta)\sin(\theta) = \cos(\theta)$ et dans la base \mathcal{B} la matrice de f est

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix},$$

qui est bien sous la forme voulue.

- Si $b = \sin(\theta)$ alors $d = -\cos(\theta)$ et la matrice de f dans la base \mathcal{B} est

$$\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}.$$

Mais alors les valeurs propres de f sont les racines de

$$(\cos(\theta) - \lambda)(-\cos(\theta) - \lambda) - \sin^2(\theta) = \lambda^2 - 1$$

c'est-à-dire que $\text{Sp}(f) = \{-1, 1\}$, ce qui est absurde.

L'hypothèse $b = -\sin(\theta)$ était donc fautive.

9. Pour $\dim(E) = 1$, soit $\mathcal{B} = (e)$ une base orthonormée de E , et $A = (a)$ la matrice de f dans la base \mathcal{B} . Comme $\|f(e)\| = 1$, on a $a^2 = 1$, soit $a = \pm 1$. La matrice A est bien sous la forme souhaitée.

Supposons la propriété vérifiée pour tout espace euclidien de dimension inférieure à $n - 1$ (où $n \geq 2$).

Soit f une isométrie d'un espace euclidien E de dimension n . Deux cas se présentent

- Supposons que f admette un vecteur propre x (que l'on peut choisir normé) associé à la valeur propre λ . Notons $F = \text{Vect}(x)$. F est stable par f . De plus comme

$\|f(x)\| = \|x\| = 1$, nous avons $\|f(x)\| = \|\lambda x\| = |\lambda| \|x\| = |\lambda| = 1$. Autrement dit $\lambda = \pm 1$.

De plus f induit une isométrie sur F^\perp . Par hypothèse de récurrence, il existe une base orthonormale \mathcal{B}_1 de F^\perp telle que la matrice de $f|_{F^\perp}$ soit de la forme

$$\begin{pmatrix} \varepsilon_1 & & & & & \\ & \ddots & & & & \\ & & \varepsilon_k & & & \\ & & & R_1 & & \\ & & & & \ddots & \\ 0 & & & & & R_p \end{pmatrix}$$

où pour $i \in \llbracket 1, k \rrbracket$, $\varepsilon_i = \pm 1$ et pour $j \in \llbracket 1, p \rrbracket$, R_j est de la forme $\begin{pmatrix} \cos(\theta_j) & -\sin(\theta_j) \\ \sin(\theta_j) & \cos(\theta_j) \end{pmatrix}$.

En considérant la famille \mathcal{B} obtenue en juxtaposant x et \mathcal{B}_1 , \mathcal{B} est une base orthonormale de E et dans cette base la matrice de f est

$$\begin{pmatrix} \lambda & & & & & \\ & \varepsilon_1 & & & & \\ & & \ddots & & & \\ & & & \varepsilon_k & & \\ & & & & R_1 & \\ & & & & & \ddots \\ 0 & & & & & & R_p \end{pmatrix},$$

qui est bien sous la forme souhaitée.

- Supposons que f n'admette pas de vecteur propre. D'après ce qui précède, il existe un vecteur x non nul tel que $F = \text{Vect}(x, f(x))$ soit stable par f . f induit une isométrie de F qui n'admet pas de vecteur propre, donc il existe une base orthonormale \mathcal{B}_1 de F telle que

$$R = \mathcal{M}_{\mathcal{B}_1}(f|_F) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \text{ où } \theta \in \mathbb{R}.$$

De même f induit une isométrie de F^\perp . D'après l'hypothèse de récurrence il existe une base orthonormale \mathcal{B}_2 de F^\perp telle que la matrice de $f|_{F^\perp}$ soit de la forme

$$\begin{pmatrix} \varepsilon_1 & & & & & \\ & \ddots & & & & \\ & & \varepsilon_k & & & \\ & & & R_1 & & \\ & & & & \ddots & \\ 0 & & & & & R_p \end{pmatrix}.$$

Comme nous sommes curieux, recherchons la matrice de u dans cette base. Nous avons

$$\begin{aligned} u(-1, 1, 0) &= \frac{1}{3}(-1, 1, 4) = \frac{1}{3}(-1, 1, 0) + \frac{4}{3}(0, 0, 1) \\ \text{et } u(0, 0, 1) &= \frac{1}{3}(2, -2, 1) = -\frac{2}{3}(-1, 1, 0) + \frac{1}{3}(0, 0, 1). \end{aligned}$$

Il en résulte que $u(f_2) = \frac{1}{3}f_2 + \frac{2\sqrt{2}}{3}f_3$ et $u(f_3) = -\frac{2\sqrt{2}}{3}f_2 + \frac{1}{3}f_3$, d'où

$$\mathcal{M}_B(u) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & -\frac{2\sqrt{2}}{3} \\ 0 & \frac{2\sqrt{2}}{3} & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

où $\theta = \text{Arccos}\left(\frac{1}{3}\right)$.

10 1. On a $A^2 = 4A$ et $B^2 = 2B$. Or on sait que les valeurs propres d'une matrice sont racines de tout polynôme annulateur de cette matrice. On a donc $\text{Sp}(A) \subset \{0, 4\}$ et $\text{Sp}(B) \subset \{0, 2\}$.

Comme A et B sont symétriques réelles, elles sont diagonalisables. Elles ont au moins une valeur propre. De plus ces matrices n'étant pas scalaires, elles ont au moins deux valeurs propres.

Il en résulte que

$$\text{Sp}(A) = \{0, 4\} \quad \text{et} \quad \text{Sp}(B) = \{0, 2\}.$$

2. Nous avons

$$(x, y, z, t) \in E_0(a) \Leftrightarrow x + y + z + t = 0 \Leftrightarrow x = -y - z - t,$$

d'où $E_0(a) = \text{Vect}((-1, 1, 0, 0), (-1, 0, 1, 0), (-1, 0, 0, 1))$. On en déduit que $\dim(E_0(a)) = 3$, d'où $\dim(E_4(a)) = 1$. Or manifestement $a(1, 1, 1, 1) = (4, 4, 4, 4)$. On en déduit que $E_4(a) = \text{Vect}((1, 1, 1, 1))$.

De même

$$(x, y, z, t) \in E_0(b) \Leftrightarrow \begin{cases} x - z = 0 \\ y - t = 0 \end{cases} \Leftrightarrow \begin{cases} x = z \\ y = t \end{cases},$$

d'où $E_0(b) = \text{Vect}((1, 0, 1, 0), (0, 1, 0, 1))$, et enfin

$$(x, y, z, t) \in E_2(b) \Leftrightarrow \begin{cases} -x - z = 0 \\ -y - t = 0 \end{cases} \Leftrightarrow \begin{cases} x = -z \\ y = -t \end{cases},$$

d'où $E_2(b) = \text{Vect}((-1, 0, 1, 0), (0, -1, 0, 1))$.

3. Comme $\dim(E_4(a)) = 1$, un des vecteurs de cette base doit nécessairement être $e_1 = \frac{1}{2}(1, 1, 1, 1)$ (ou son opposé). Il se trouve que ce vecteur est dans $E_0(b)$. Nous devons donc choisir un second vecteur de $E_0(b)$ qui soit orthogonal au premier. Un peu d'intuition nous amène à choisir $e_2 = \frac{1}{2}(1, -1, 1, -1)$. Il se trouve que e_2 est aussi un élément de $E_0(a)$.

Il reste à choisir deux vecteurs qui forment une base orthonormale de $E_2(b)$ et qui soient également dans $E_0(a)$.

Il se trouve que les deux vecteurs trouvés à la question 2, et normalisés, répondent à la question

$$e_3 = \frac{1}{\sqrt{2}}(-1, 0, 1, 0) \quad \text{et} \quad e_4 = \frac{1}{\sqrt{2}}(0, -1, 0, 1).$$

La famille $\mathcal{B} = (e_1, e_2, e_3, e_4)$ étant la juxtaposition de bases orthonormales des sous-espaces supplémentaires orthogonaux $E_0(b)$ et $E_2(b)$, elle est une base orthonormale de \mathbb{R}^4 et elle est de plus formée de vecteurs qui sont à la fois vecteurs propres de a et de b . Dans cette base, les matrices de a et b sont diagonales.

11 1. – Vérifions la linéarité par rapport à la première variable.

Soient $P, Q, R \in E$ et $\lambda \in \mathbb{R}$,

$$\begin{aligned} \varphi(\lambda P + Q, R) &= (\lambda P + Q)(0)R(0) + (\lambda P + Q)(1)R(1) + (\lambda P + Q)(-1)R(-1) \\ &= (\lambda P(0) + Q(0))R(0) + (\lambda P(1) + Q(1))R(1) + (\lambda P(-1) + Q(-1))R(-1) \\ &= \lambda(P(0)R(0) + P(1)R(1) + P(-1)R(-1)) \\ &\quad + (Q(0)R(0) + Q(1)R(1) + Q(-1)R(-1)) \\ &= \lambda\varphi(P, R) + \varphi(Q, R). \end{aligned}$$

– La symétrie,

$$\begin{aligned} \varphi(P, Q) &= P(0)Q(0) + P(1)Q(1) + P(-1)Q(-1) = Q(0)P(0) + Q(1)P(1) + Q(-1)P(-1) \\ &= \varphi(Q, P). \end{aligned}$$

– La positivité

$$\varphi(P, P) = P(0)^2 + P(1)^2 + P(-1)^2 \geq 0.$$

– Montrons enfin que φ est définie. Soit $P \in E$ tel que $\varphi(P, P) = 0$, on a

$$\varphi(P, P) = P(0)^2 + P(1)^2 + P(-1)^2 = 0,$$

d'où $P(0) = P(1) = P(-1) = 0$. Autrement dit P a au moins trois racines, mais $d^o P < 2$, donc $P = 0$, ce qui prouve que φ est définie.

2. a) Soit $P \in E$ où $P = aX^2 + bX + c$, nous avons

$$2P'(0) - P(1) + P(-1) = 2b - (a + b + c) + (a - b + c) = 0.$$

b) Soit $P \in E$,

$$\begin{aligned} \langle \varphi(P), P \rangle &= u(P)(0)P(0) + u(P)(1)P(1) + u(P)(-1)P(-1) \\ &= 0 + (2P'(0) - (P(1) + P(-1)))P(1) + (2P'(0) + P(1) + P(-1))P(-1) \\ &= 2P'(0)P(1) - P(1)^2 - P(-1)P(1) + 2P'(0)P(-1) + P(1)P(-1) + P(-1)^2 \\ &= 2P'(0)(P(1) + P(-1)) + (P(-1) - P(1))(P(-1) + P(1)) \\ &= (P(1) + P(-1))(2P'(0) - P(1) + P(-1)) = 0. \end{aligned}$$

L'endomorphisme φ est bien antisymétrique.

3. a) Nous avons

$$\begin{aligned} u(u(P_1)) &= u\left(2 \times \frac{1}{2}X^2 - (1)X\right) = u(X^2 - X) \\ &= -2X^2 - (2)X = -4P_1. \end{aligned}$$

De plus $P_1 \neq 0$, donc P_1 est un vecteur propre de u^2 associé à la valeur propre -4 .
Par ailleurs

$$\begin{aligned} \langle P_1, P_2 \rangle &= \left\langle \frac{1}{2}(X^2 + X), \frac{1}{2}(X^2 - X) \right\rangle = 0 + 0 + 0 = 0 \\ \|P_1\|^2 = \langle P_1, P_1 \rangle &= \left\langle \frac{1}{2}(X^2 + X), \frac{1}{2}(X^2 + X) \right\rangle = 0 + 1 \times 1 + 0 = 1 \\ \|P_2\|^2 = \left\langle \frac{1}{2}(X^2 - X), \frac{1}{2}(X^2 - X) \right\rangle &= 0 + 0 + 1 \times 1 = 1. \end{aligned}$$

La famille (P_1, P_2) est bien orthonormale.

b) Soit $P = aX^2 + bX + c$, nous avons

$$P \in \text{Ker}(u) \Leftrightarrow \begin{cases} 2P'(0) = 0 \\ P(1) + P(-1) = 0 \end{cases} \Leftrightarrow \begin{cases} b = 0 \\ a + c = 0 \end{cases} \Leftrightarrow \begin{cases} b = 0 \\ a = -c \end{cases}$$

d'où $\text{Ker}(u) = \text{Vect}(X^2 - 1)$.

c) En posant $P_3 = X^2 - 1$, on a

$$\begin{aligned} \langle P_3, P_1 \rangle &= 0 \\ \langle P_3, P_2 \rangle &= 0 \\ \|P_3\|^2 = \langle P_3, P_3 \rangle &= 1. \end{aligned}$$

La famille (P_1, P_2, P_3) est orthonormale. Elle est donc libre, or elle est de cardinal 3 et $\dim(E) = 3$. On en déduit que c'est une base orthonormale de E .

De plus $u(P_1) = 2P_2$, $u(P_2) = \frac{1}{2}u^2(P_1) = \frac{1}{2} \times -4P_1 = -2P_1$ et $u(P_3) = 0$.

Il en résulte que dans cette base, la matrice de u est

$$\begin{pmatrix} 0 & -2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Cette matrice est sous la forme voulue (avec $a = 2$).

4. Nous avons

$$\langle u(x + y), x + y \rangle = \langle u(x) + u(y), x + y \rangle = \langle u(x), x \rangle + \langle u(x), y \rangle + \langle u(y), x \rangle + \langle u(y), y \rangle.$$

Prouvons maintenant l'équivalence :

– Supposons u antisymétrique, soit $(x, y) \in E^2$, nous avons $\langle u(x + y), x + y \rangle = 0$, par suite

$$\langle u(x), x \rangle + \langle u(x), y \rangle + \langle u(y), x \rangle + \langle u(y), y \rangle = 0,$$

soit

$$\langle u(x), y \rangle + \langle u(y), x \rangle = 0$$

ou encore $\langle u(x), y \rangle = -\langle u(y), x \rangle$.

– Réciproquement, supposons que u vérifie la propriété annoncée.

Soit $x \in E$, on a par hypothèse $\langle u(x), x \rangle = -\langle x, u(x) \rangle$. Le produit scalaire étant symétrique, il vient $2\langle u(x), x \rangle = 0$, soit $\langle u(x), x \rangle = 0$. L'endomorphisme u est bien symétrique.

5. a) C'est une propriété qui a été vu dans le cours.

b) Prouvons par une double implication :

– Si u est antisymétrique, alors pour tout couple $(i, j) \in \llbracket 1, n \rrbracket^2$,

$$m_{i,j} = \langle e_i, u(e_j) \rangle = -\langle u(e_i), e_j \rangle = -m_{i,j}.$$

Autrement dit ${}^tM = -M$.

– Supposons que ${}^tM = -M$. Nous avons donc pour tout $(i, j) \in \llbracket 1, n \rrbracket^2$,

$$\langle e_i, u(e_j) \rangle = -\langle u(e_i), e_j \rangle.$$

Soit $x \in E$ tel que $x = \sum_{k=1}^n a_k e_k$, nous avons

$$\begin{aligned} \langle u(x), x \rangle &= \left\langle u \left(\sum_{i=1}^n a_i e_i \right), \sum_{j=1}^n a_j e_j \right\rangle = \left\langle \sum_{i=1}^n a_i u(e_i), \sum_{j=1}^n a_j e_j \right\rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \langle u(e_i), e_j \rangle = \sum_{i=1}^n \sum_{j=1}^n -a_i a_j \langle e_i, u(e_j) \rangle \\ &= - \left\langle \sum_{i=1}^n a_i e_i, \sum_{j=1}^n a_j u(e_j) \right\rangle = -\langle x, u(x) \rangle. \end{aligned}$$

Autrement dit $\langle u(x), x \rangle = 0$ et u est antisymétrique.

6. Soit λ une valeur propre de u et x un vecteur propre associé, on a $\langle u(x), x \rangle = \langle \lambda x, x \rangle = \lambda \|x\|^2$. Or u étant antisymétrique $\langle u(x), x \rangle = 0$ et comme $\|x\| \neq 0$, nous avons $\lambda = 0$.

7. Par la formule du rang, nous savons déjà que $\dim(\Im(u)) + \dim(\text{Ker}(u)) = \dim(E)$.

Montrons que ces sous-espaces sont orthogonaux.

Soient $x \in \Im(u)$ et $y \in \text{Ker}(u)$. Par définition, il existe $z \in E$ tel que $u(z) = x$. On a alors d'après la caractérisation de l'antisymétrie de la question II.4,

$$\langle x, y \rangle = \langle u(z), y \rangle = -\langle z, u(y) \rangle = -\langle z, 0 \rangle = 0,$$

ce qui prouve le résultat.

Montrons maintenant que $\text{Ker}(u) = \text{Ker}(u^2)$.

Si $x \in \text{Ker}(u)$ alors $u^2(x) = u(u(x)) = u(0) = 0$, donc $x \in \text{Ker}(u^2)$. Autrement dit $\text{Ker}(u) \subset \text{Ker}(u^2)$.

Soit $x \in \text{Ker}(u^2)$, alors $u(u(x)) = 0$. C'est-à-dire que $u(x) \in \text{Ker}(u)$. Or on a également $u(x) \in \Im(u)$. Il s'ensuit que $u(x) = 0$, ce qui signifie que $x \in \text{Ker}(u)$, et donc $\text{Ker}(u^2) \subset \text{Ker}(u)$.

On peut conclure que $\text{Ker}(u) = \text{Ker}(u^2)$.

8. Pour tout $(x, y) \in E^2$,

$$\langle u^2(x), y \rangle = -\langle u(x), u(y) \rangle = -(-\langle x, u^2(y) \rangle) = \langle x, u^2(y) \rangle.$$

donc u^2 est symétrique.

Soit λ une valeur propre de u^2 et x un vecteur propre associé. On a d'une part

$$\langle u^2(x), x \rangle = \langle \lambda x, x \rangle = \lambda \|x\|^2$$

et d'autre part

$$\langle u^2(x), x \rangle = -\langle u(x), u(x) \rangle = -\|u(x)\|^2.$$

Comme $x \neq 0$, on en déduit que $\lambda = -\frac{\|u(x)\|^2}{\|x\|^2} \leq 0$.

9. a) Supposons que u^2 n'admette pas de valeur propre non nulle. Comme u^2 est diagonalisable, puisque symétrique, et que 0 est son unique valeur propre, on a $u^2 = 0$. Il en résulte alors que pour tout $x \in E$,

$$\langle u^2(x), x \rangle = -\langle u(x), u(x) \rangle = -\|u(x)\|^2 = 0.$$

Autrement dit pour tout $x \in E$, $u(x) = 0$, ce qui est absurde car u est supposé non nul. u^2 admet donc des valeurs propres non nulles.

b) Pour tout couple de réels (a, b) , $u(ax+bu(x)) = au(x)+bu^2(x) = au(x)+b\lambda x \in \text{Vect}(x, u(x))$.

Donc F est stable par u .

c) Soit $x \in F^\perp$, pour tout $y \in F$, on a

$$\langle u(x), y \rangle = -\langle x, u(y) \rangle = 0$$

puisque $u(y) \in F$ (F est stable par u). Il s'ensuit que $u(x) \in F^\perp$, c'est-à-dire que F^\perp est stable par u .

d) L'antisymétrie de u_1 découle immédiatement de celle de u .

Soit $y \in \mathfrak{Im}(u)$, il existe $z \in E$ tel que $y = u(z)$. Comme $E = F \oplus F^\perp$, il existe $(z_1, z_2) \in F \times F^\perp$ tel que $z = z_1 + z_2$. On a alors $y = u(z) = u(z_1 + z_2) = u(z_1) + u(z_2) = u(z_1) + u_1(z_2) = u(z_1) + u_1(z_2)$.

On en déduit que $\mathfrak{Im}(u) \subset F + \mathfrak{Im}(u_1)$.

Réciproquement, soit $y \in F + \mathfrak{Im}(u_1)$, y peut s'écrire sous la forme $y = \gamma_1 + u_1(\gamma_2)$ (où $\gamma_1 \in F$, $\gamma_2 \in F^\perp$). Par définition il existe deux scalaires a, b tels que

$$\gamma_1 = ax + bu(x) = a\frac{1}{\lambda}u^2(x) + bu(x) = u\left(\frac{a}{\lambda}u(x) + bx\right),$$

et finalement

$$y = u\left(\frac{a}{\lambda}u(x) + bx + \gamma_2\right) \in \mathfrak{Im}(u).$$

On en déduit que $F + \mathfrak{Im}(u_1) \subset \mathfrak{Im}(u)$ et donc $\mathfrak{Im}(u) = F + \mathfrak{Im}(u_1)$.

Enfin, comme $\mathfrak{Im}(u_1) \subset F^\perp$ et que $F \cap F^\perp = \{0\}$, F et $\mathfrak{Im}(u_1)$ sont en somme directe.

Autrement dit $\mathfrak{Im}(u) = F \oplus \mathfrak{Im}(u_1)$.

10. Soit u un endomorphisme antisymétrique d'un espace E de dimension 1 alors, comme la seule valeur propre possible de u est 0, on a $u = 0$. Le rang de u est nul et donc pair.

Supposons la propriété prouvée pour tout endomorphisme antisymétrique d'un espace de

dimension inférieure à n , prouvons la propriété pour tout endomorphisme d'un espace de dimension $n + 1$.

Soit u un tel endomorphisme. Si u est nul, le rang de u est nul et pair.

Si u n'est pas nul, alors d'après ce qui précède il existe un sous-espace F de E de dimension 2 tel que $\mathfrak{Im}(u) = F \oplus \mathfrak{Im}(u_1)$ où u_1 est un endomorphisme symétrique de F^\perp .

F est de dimension 2 car les vecteurs x et $u(x)$ de la question 9.a, ne peuvent être colinéaires. En effet 0 étant l'unique valeur propre possible de u , si x et $u(x)$ étaient colinéaires, on aurait $u(x) = 0x = 0$, et $u^2(x) = 0$ ce qui est absurde. La famille $(x, u(x))$ est donc une base de F .

Comme $\dim(F^\perp) < n + 1$, d'après l'hypothèse de récurrence, $\text{rg}(u_1)$ est pair. Or $\text{rg}(u) = \dim(F) + \text{rg}(u_1) = 2 + \text{rg}(u_1)$, donc $\text{rg}(u)$ est pair.

11. La base \mathcal{B} est orthonormale et ${}^tA = -A$ donc u est antisymétrique.

Un calcul donne

$$A^2 = \begin{pmatrix} -18 & 0 & -9 & -9 \\ 0 & -18 & -9 & 9 \\ -9 & -9 & -27 & 0 \\ -9 & 9 & 0 & -27 \end{pmatrix} \text{ et } A^2 \begin{pmatrix} 1 \\ 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -9 \\ -9 \\ 9 \\ 0 \end{pmatrix}.$$

f_1 est un vecteur propre de u^2 associé à la valeur propre -9 .

12. Comme nous l'avons déjà vu $(f_1, u(f_1))$ est une base de F . Il reste à l'orthonormaliser par le procédé de Schmidt. Tout d'abord $f_2 = u(f_1)$ a pour coordonnées $(3, -3, 0, -3)$ dans la base \mathcal{B} . On constate que

$$\langle f_1, f_2 \rangle = \langle (1, 1, -1, 0), (3, -3, 0, -3) \rangle = 3 - 3 = 0.$$

La famille est donc orthogonale. Par suite une base orthonormale de F est

$$\left(\frac{1}{\|f_1\|} f_1, \frac{1}{\|f_2\|} f_2 \right) = \left\langle \frac{1}{\sqrt{3}}(e_1 + e_2 - e_3), \frac{1}{\sqrt{3}}(e_1 - e_2 - e_4) \right\rangle.$$

Déterminons maintenant une base de F^\perp . Soit $u \in E$ de coordonnées (x, y, z, t) dans la base \mathcal{B} , on a

$$x \in F^\perp \Leftrightarrow \begin{cases} x + y - z = 0 \\ x - y - t = 0 \end{cases} \Leftrightarrow \begin{cases} z = x + y \\ t = x - y \end{cases}$$

d'où $F^\perp = \text{Vect}(e_1 + e_3 + e_4, e_2 + e_3 - e_4)$. Cette famille est également orthogonale, une base orthonormale de F^\perp est donc

$$\left(\frac{1}{\sqrt{3}}(e_1 + e_3 + e_4), \frac{1}{\sqrt{3}}(e_2 + e_3 - e_4) \right).$$

13. On vérifie que dans la base orthonormale de E formée de la juxtaposition des bases orthonormales de F et F^\perp , la matrice de u est :

$$\begin{pmatrix} 0 & -3 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 \\ 0 & 0 & -6 & 0 \end{pmatrix}.$$

- 12** 1. **a)** De $f \circ g = h$ il découle que $\Im m(h) \subset \Im m(f)$ et $\text{Ker}(g) \subset \text{Ker}(h)$. De même $\Im m(f) \subset \Im m(g)$, $\text{Ker}(h) \subset \text{Ker}(f)$, $\Im m(g) \subset \Im m(h)$ et $\text{Ker}(f) \subset \text{Ker}(g)$.
- b)** D'après les égalités supposées, on a $f^2 = fgh$, $g^2 = (hf)g = h(fg) = fgh$ et $h^2 = fgh$. D'où $f^2 = g^2 = h^2$.
De plus $h^2fh^2 = f^5$ mais également $h^2fh^2 = hgh^2 = hfn = gh = f$, donc $f = f^5$.
- c)** Soit $i \in \mathbb{N}^*$, montrons que $\text{Ker}(f^i) = \text{Ker}(f^{i+1})$.
Soit $x \in \text{Ker}(f^i)$, alors $f^{i+1}(x) = f(f^i(x)) = f(0) = 0$, donc $x \in \text{Ker}(f^{i+1})$.
Réciproquement, soit $x \in \text{Ker}(f^{i+1})$. Notons p le plus petit multiple de 5 supérieur ou égal à i . On alors $x \in \text{Ker}(f^p)$. En notant $k = \frac{p}{5}$, on a, compte tenu de $f^5 = f$, $x \in \text{Ker}(f^k)$, or $k \leq i$, donc $x \in \text{Ker}(f^i)$.
On en déduit que $\text{Ker}(f^i) = \text{Ker}(f^{i+1})$ pour tout $i \in \mathbb{N}^*$ ce qui répond à la question.
Montrons que $\Im(f) \cap \text{Ker}(f) = \{0\}$.
Soit $x \in \Im(f) \cap \text{Ker}(f)$. Comme $x \in \Im(f)$, il existe $y \in E$ tel que $x = f(y)$. Mais comme $x \in \text{Ker}(f)$, on a $f(x) = f^2(y) = 0$. Autrement dit $y \in \text{Ker}(f^2)$, d'où $y \in \text{Ker}(f)$ et $x = f(y) = 0$.
Sachant que $\dim \Im(f) + \dim \text{Ker}(f) = \dim E$, on a bien $\Im(f) \oplus \text{Ker}(f) = E$.
- 2. a)** Le polynôme $X^5 - X = X(X^4 - 1)$ est annulateur de f , d'où $\text{Sp}(f) \subset \{-1, 0, 1\}$.
Comme f est de rang n , 0 ne peut être valeur propre de f . Les valeurs propres possibles de f sont -1 et 1 .
De plus si $x \in E_\lambda(f)$, alors $f(g(x)) = h(x)$ et $g(f(x)) = g(\lambda x) = \lambda g(x) = \lambda hf(x) = \lambda^2 h(x)$.
Or $\lambda = \pm 1$, d'où $g(f(x)) = \lambda g(x) = h(x) = f(g(x))$, autrement dit $g(x) \in E_\lambda(f)$. Les sous-espaces propres de f sont bien stables par g .
De même $f(h(x)) = f(f(g(x))) = f^2(\lambda x) = \lambda f^2(x) = \lambda^3 x = \lambda x$, donc $h(x) \in E_\lambda(f)$. Les sous-espaces propres de f sont stables par h .
- b)** f étant symétrique est diagonalisable. Comme $\text{Sp}(f) \subset \{-1, 1\}$, il existe une base \mathcal{B} dans laquelle la matrice de f est de la forme $A = \text{Diag}(\pm 1, \pm 1, \dots, \pm 1)$. On a alors $\mathcal{M}_{\mathcal{B}}(f^2) = A^2 = \text{Diag}(1, 1, \dots, 1) = I_n$, d'où $f^2 = \text{Id}_E$.
On en déduit que $gf = g^2h = f^2h = h = fg$, $fh = f^2g = g = hf$, $hg = fg^2 = f = gh$.
- c)** Comme $f^2 = g^2 = h^2 = \text{Id}_E$, les endomorphismes f , g et h sont des symétries. Ils sont donc diagonalisables. On montre alors comme dans l'exercice 1.7 qu'ils admettent une base commune de vecteurs propres.

13 1. La fonction $t \mapsto \frac{P(t)}{\sqrt{1-t^2}}$ est continue sur $] -1, 1[$.

En 1, si $P(1) \neq 0$, alors

$$\frac{P(t)}{\sqrt{1-t^2}} \sim \frac{P(1)}{\sqrt{1-t}\sqrt{1+t}} \sim \frac{P(1)}{\sqrt{2}\sqrt{1-t}}$$

Or l'intégrale de $\frac{1}{\sqrt{1-t}}$ converge en 1 d'après le critère de Riemann. De plus $\frac{P(1)}{\sqrt{2}\sqrt{1-t}}$ est

de signe constant au voisinage de 1. On en déduit que l'intégrale $\int_0^1 \frac{P(t)}{\sqrt{1-t^2}} dt$ converge.

Si $P(1) = 0$, alors P peut se mettre sous la forme $P(t) = (1-t)Q$. On a alors

$$\frac{P(t)}{\sqrt{1-t^2}} = \frac{(1-t)Q(t)}{\sqrt{1-t}\sqrt{1+t}} = \frac{\sqrt{1-t}Q(t)}{\sqrt{1+t}}$$

fonction qui est prolongeable par continuité en 1. L'intégrale est donc convergente en 1 également dans ce cas.

La démonstration est analogue en -1 .

2. Des exemples analogues ont été vus dans le cours.

3. On le montre par une récurrence double sur k .

La vérification est immédiate pour $k = 0$ et $k = 1$.

Supposons que la propriété soit vraie aux rangs n et $n + 1$. On a alors pour tout réel x ,

$$\begin{aligned} T_{n+2}(\cos(x)) &= 2 \cos(x) T_{n+1}(\cos(x)) - T_n(\cos(x)) = 2 \cos(x) \cos((n+1)x) - \cos(nx) \\ &= \cos((n+1)x + x) + \cos((n+1)x - x) - \cos(nx) = \cos((n+2)x). \end{aligned}$$

La propriété est prouvée par récurrence.

4. Soit $(i, j) \in \llbracket 0, n \rrbracket$ avec $i \neq j$, alors

$$\langle T_i, T_j \rangle = \int_{-1}^1 \frac{T_i(t) T_j(t)}{\sqrt{1-t^2}} dt,$$

Posons $t = \cos(u)$ avec $u \in [0, \pi]$, on a $dt = -\sin(u) du = -\sqrt{1-t^2} du$, d'où

$$\begin{aligned} \langle T_i, T_j \rangle &= \int_{\pi}^0 -T_i(\cos(u)) T_j(\cos(u)) du = \int_0^{\pi} \cos(iu) \cos(ju) du \\ &= \frac{1}{2} \int_0^{\pi} \cos((i+j)u) + \cos((i-j)u) du \\ &= \frac{1}{2} \left[\frac{\sin((i+j)u)}{i+j} + \frac{\sin((i-j)u)}{i-j} \right]_0^{\pi} \\ &= 0. \end{aligned}$$

La famille est bien orthogonale.

5. En reprenant le calcul précédent, on trouve

$$\begin{aligned} \langle T_k, T_k \rangle &= \int_0^{\pi} T_k(\cos(u)) T_k(\cos(u)) du = \int_0^{\pi} \cos^2(ku) du \\ &= \frac{1}{2} \int_0^{\pi} \cos(2ku) + 1 du \\ &= \frac{1}{2} \left[\frac{\sin(2ku)}{2k} + u \right]_0^{\pi} \\ &= \frac{\pi}{2}. \end{aligned}$$

D'où $\|T_k\| = \sqrt{\frac{\pi}{2}}$.

- 14** 1. Rappelons que la trace est linéaire (cela a par exemple été vu en première année).

Soient $A, B, C \in \mathcal{M}_3(\mathbb{R})$ et $\lambda, \mu \in \mathbb{R}$, on a

$$\begin{aligned} \varphi(\lambda A + \mu B, C) &= \text{tr}((\lambda A + \mu B)^t C) = \text{tr}(\lambda A^t C + \mu B^t C) = \lambda \text{tr}(A^t C) + \mu \text{tr}(B^t C) \\ &= \lambda \varphi(A, C) + \mu \varphi(B, C). \end{aligned}$$

Montrons que $\varphi(A, B) = \varphi(B, A)$. Posons $A = (a_{i,j})$, $B = (b_{i,j})$ et $A^t B = (d_{i,j})$, on a

$$\varphi(A, B) = \text{tr}(A^t B) = \sum_{i=1}^n d_{i,i} = \sum_{i=1}^n \sum_{j=1}^n a_{i,j} b_{i,j} = \sum_{i=1}^n \sum_{j=1}^n b_{i,j} a_{i,j} = \text{tr}(B^t A) = \varphi(B, A).$$

De plus, $\varphi(A, A) = \sum_{i=1}^n \sum_{j=1}^n a_{i,j}^2 \geq 0$, donc φ est positive.

Enfin, si $\varphi(A, A) = 0$, alors $\sum_{i=1}^n \sum_{j=1}^n a_{i,j}^2 = 0$, et donc $\forall (i, j) \in \llbracket 1, n \rrbracket^2$, $a_{i,j} = 0$, autrement

dit $A = 0$ et φ est définie.

L'application φ est un produit scalaire.

2. a) Soit $A = (a_{i,j}) \in \mathcal{M}_3(\mathbb{R})$,

$$A \in J^\perp \Leftrightarrow \langle A, J \rangle = 0 \Leftrightarrow \sum_{i=1}^n \sum_{j=1}^n a_{i,j} = 0.$$

Autrement dit l'orthogonal de J est l'ensemble des matrices dont la somme des coefficients est égale à 0.

b) On sait que

$$A' = \left\langle A, \frac{1}{\|J\|} J \right\rangle \frac{1}{\|J\|} J = \frac{1}{\|J\|^2} \langle A, J \rangle J = \frac{1}{n^2} \left(\sum_{i=1}^n \sum_{j=1}^n a_{i,j} \right) J.$$

Par suite

$$A'' = A - \frac{1}{n^2} \left(\sum_{i=1}^n \sum_{j=1}^n a_{i,j} \right) J.$$

c) D'après la question précédente, le projeté orthogonal de I_3 sur F est $\frac{1}{n}J$ et le projeté orthogonal de I_3 sur F^\perp est $I_3 - \frac{1}{n}J$.

Chapitre 3

1. La fonction $x \mapsto (x - \sqrt{x^2 + 1})$ est continue sur \mathbb{R}_+ et négative. En $+\infty$, on

$$x - \sqrt{x^2 + 1} = \frac{-1}{x + \sqrt{x^2 + 1}} \sim -\frac{1}{2x}.$$

Comme $\int_1^{+\infty} -\frac{1}{2x} dx$ diverge, il en est de même de $\int_0^{+\infty} (x - \sqrt{x^2 + 1}) dx$.

2. La fonction $f : x \mapsto \left(\frac{\text{Arctan } x}{x} \right)^2$ est continue sur $]0, +\infty[$.

Comme $\text{Arctan } x \underset{x \rightarrow 0}{\sim} x$, on a $\lim_{x \rightarrow 0} f(x) = 1$. Ainsi $\int_0^1 f(x) dx$ converge. De

$\lim_{x \rightarrow +\infty} \text{Arctan } x = \frac{\pi}{2}$, on déduit $f(x) \underset{x \rightarrow +\infty}{\sim} \frac{\pi^2}{4x^2}$. Comme $\int_1^{+\infty} \frac{1}{x^2} dx$ converge, on en

déduit que $\int_1^{+\infty} f(x) dx$ converge.

Finalement $\int_0^{+\infty} \left(\frac{\text{Arctan } x}{x} \right)^2 dx$ converge.

3. La fonction $f : t \mapsto |1 - t^\alpha|^\beta$ est continue sur $]0, 1[$ et positive. Il peut y avoir une impropriété en 0 et en 1.

Si $\alpha > 0$, on a $\lim_{t \rightarrow 0} f(t) = 1$; il n'y a pas d'impropriété en 0. Si $\alpha < 0$, on a $\lim_{t \rightarrow 0} t^\alpha = +\infty$.

On en déduit que $|1 - t^\alpha| \sim t^\alpha$ et $f(t) \sim t^{\alpha\beta} \sim \frac{1}{t^{-\alpha\beta}}$. On en déduit que $\int_0^{\frac{1}{2}} f(t) dt$ converge si $-\alpha\beta < 1$, soit $\alpha\beta > -1$.

Pour l'étude au voisinage de 1, on pose $t = 1 - h$. On a $1 - t^\alpha = 1 - (1 - h)^\alpha \underset{h \rightarrow 0^+}{\sim} \alpha h$ et

$$f(t) \sim |\alpha h|^\beta \sim |\alpha|^\beta h^\beta \sim \frac{|\alpha|^\beta}{(1-t)^{-\beta}}.$$

Ainsi $\int_{\frac{1}{2}}^1 f(t) dt$ converge si $-\beta < 1$ soit $\beta > -1$.

Finalement $\int_0^1 f(t) dt$ converge pour ($\alpha > 0$ ou $\alpha < 0$ et $\alpha\beta > -1$) et $\beta > -1$.

4. La fonction $f : x \mapsto \frac{\ln x}{x^2 - 1}$ est définie sur $]0, 1[\cup]1, +\infty[$. Elle est positive. Comme

$\ln x \underset{x \rightarrow 1}{\sim} x - 1$, on a $\lim_{x \rightarrow +\infty} f(x) = \frac{1}{2}$ et la fonction peut être prolongée par continuité en 1.

En 0, on a $f(x) \sim -\ln x$. Comme l'intégrale $\int_0^1 -\ln x dx$ converge, il en est de même de

$$\int_0^1 f(x) dx.$$

En $+\infty$, on a

$$f(x) = o\left(\frac{1}{x^{\frac{3}{2}}}\right),$$

car $\lim_{x \rightarrow +\infty} \frac{\ln x}{x^{\frac{1}{2}}} = 0$, donc $\int_1^{+\infty} f(x) dx$ converge par comparaison à une intégrale de Riemann.

Finalement $\int_0^{+\infty} f(x) dx$ converge.

5. La fonction $f : x \mapsto \frac{\ln x}{x + e^{-x}}$ est continue sur \mathbb{R}_+^* .

On a $f(x) \underset{x \rightarrow 0}{\sim} \ln x$ et $\int_0^1 f(x) dx$ converge comme $\int_0^1 \ln x dx$.

En $+\infty$, on obtient $f(x) \sim \frac{\ln x}{x}$ et comme $\int_1^X \frac{\ln x}{x} dx = \frac{1}{2}(\ln X)^2$ tend vers $+\infty$ quand X

tend vers $+\infty$, $\int_1^{+\infty} f(x) dx$ diverge. Finalement $\int_0^{+\infty} \frac{\ln x}{x + e^{-x}} dx$ diverge.

6. La fonction $f : x \mapsto \frac{x^\alpha}{1 + x^\beta}$ est continue sur $]0, +\infty[$ et positive.

• Supposons $\beta > 0$. On a alors $\lim_{x \rightarrow 0} (1 + x^\beta) = 1$ et $f(x) \underset{x \rightarrow 0}{\sim} x^\alpha$: l'intégrale $\int_0^1 f(x) dx$ converge si $\alpha > -1$.

De même, on obtient $\lim_{x \rightarrow +\infty} x^\beta = +\infty$ et $f(x) \underset{x \rightarrow +\infty}{\sim} \frac{1}{x^{\beta-\alpha}}$. L'intégrale $\int_1^{+\infty} f(x) dx$ converge si $\beta - \alpha > 1$.

Finalement $\int_0^{+\infty} f(x) dx$ converge si $-1 < \alpha < \beta - 1$.

• Supposons $\beta < 0$. On a $\lim_{x \rightarrow 0} x^\beta = +\infty$ et $f(x) \underset{x \rightarrow 0}{\sim} \frac{1}{x^{\beta-\alpha}}$. L'intégrale $\int_0^1 f(x) dx$ converge si $\beta - \alpha < 1$.

On obtient de même $\lim_{x \rightarrow +\infty} x^\beta = 0$ et $f(x) \underset{x \rightarrow +\infty}{\sim} \frac{1}{x^{-\alpha}}$. L'intégrale converge $\int_1^{+\infty} f(x) dx$ si $-\alpha > 1$.

Finalement $\int_0^{+\infty} f(x) dx$ converge si, et seulement si, $\beta - 1 < \alpha < -1$.

• Supposons $\beta = 0$ et donc $f(x) = x^\alpha$. L'intégrale $\int_0^{+\infty} f(x) dx$ diverge car $\int_0^1 f(x) dx$ converge si $\alpha > -1$ et $\int_1^{+\infty} f(x) dx$ converge si $\alpha < -1$.

7. • La fonction $f : x \mapsto \frac{\sin x}{x^\alpha}$ est continue sur $]0, +\infty[$.

En 0, on a $f(x) \sim \frac{1}{x^{\alpha-1}}$ et $\int_0^1 f(x) dx$ converge pour $\alpha - 1 < 1$ soit $\alpha < 2$.

Si $\alpha > 0$, on montre que $\int_1^{+\infty} f(x) dx$ converge, en intégrant par parties. On obtient, pour $X > 1$,

$$\int_1^X f(x) dx = \left[-\frac{\cos x}{x^\alpha} \right]_1^X - \int_1^X \frac{\alpha \cos x}{x^{\alpha+1}} dx = -\frac{\cos X}{X^\alpha} + 1 - \int_1^X \frac{\alpha \cos x}{x^{\alpha+1}} dx.$$

De la majoration $\left| \frac{\alpha \cos x}{x^{\alpha+1}} \right| \leq \frac{1}{x^{\alpha+1}}$, avec $\alpha + 1 > 1$, on déduit la convergence absolue de $\int_1^{+\infty} \frac{\alpha \cos x}{x^{\alpha+1}}$. Ainsi $\int_1^X \frac{\alpha \cos x}{x^{\alpha+1}} dx$ a une limite finie quand X tend vers $+\infty$. Comme $\frac{\cos X}{X^\alpha}$ tend vers 0, on en déduit que $\int_1^X f(x) dx$ a une limite finie quand X tend vers $+\infty$.

Autrement dit, $\int_1^{+\infty} f(x) dx$ converge.

Si $\alpha = 0$, $\int_1^{+\infty} f(x) dx$ diverge car la fonction $-\cos$, primitive de f , n'a pas de limite en $+\infty$.

Si $\alpha < 0$, on a, pour tout $k \in \mathbb{N}^*$,

$$\int_{\frac{\pi}{6} + k\pi}^{\frac{5\pi}{6} + k\pi} f(x) dx \geq \int_{\frac{\pi}{6} + k\pi}^{\frac{5\pi}{6} + k\pi} \frac{1}{2} \left(\frac{\pi}{6} + k\pi \right)^{-\alpha} dx \geq \frac{\pi}{3} \left(\frac{\pi}{6} + k\pi \right)^{-\alpha},$$

donc $\int_{\frac{\pi}{6}+k\pi}^{\frac{5\pi}{6}+k\pi} f(x) dx$ tend vers $+\infty$ quand k tend vers $+\infty$. Cela montre que $\int_1^{+\infty} f(x) dx$ diverge, sinon on aurait, en notant F une primitive de f ,

$$\lim_{k \rightarrow +\infty} \int_{\frac{\pi}{6}+k\pi}^{\frac{5\pi}{6}+k\pi} f(x) dx = \lim_{k \rightarrow +\infty} F\left(\frac{\pi}{6} + k\pi\right) - F\left(\frac{5\pi}{6} + k\pi\right) = \lim_{+\infty} F - \lim_{+\infty} F = 0.$$

Finalement $\int_0^{+\infty} f(x) dx$ converge si $\alpha \in]0, 2[$.

2 1. On considère $f_{\alpha,\beta} : x \mapsto \frac{1}{x^\alpha (\ln x)^\beta}$ et $I_{\alpha,\beta} = \int_2^{+\infty} f_{\alpha,\beta}(x) dx$. La fonction $f_{\alpha,\beta}$ est continue sur $]2, +\infty[$ et positive.

Si $\alpha < 1$, on a $\lim_{x \rightarrow 0} x f_{\alpha,\beta}(x) = \lim_{x \rightarrow 0} \frac{x^{1-\alpha}}{(\ln x)^\beta} = +\infty$ car $1 - \alpha > 0$. On a donc, quand x tend vers $+\infty$, $\frac{1}{x} = o(f(x))$. Comme $\int_2^{+\infty} \frac{1}{x} dx$ diverge, il en est de même de $I_{\alpha,\beta}$.

Si $\alpha > 1$, on a $\lim_{x \rightarrow +\infty} x^{\frac{1+\alpha}{2}} f_{\alpha,\beta}(x) = \lim_{x \rightarrow +\infty} \frac{x^{\frac{1-\alpha}{2}}}{(\ln x)^\beta} = 0$ car $1 - \alpha < 0$. On a donc, quand x tend vers $+\infty$, $f_{\alpha,\beta}(x) = o\left(\frac{1}{x^{\frac{1+\alpha}{2}}}\right)$. Comme $\frac{1+\alpha}{2} > 1$, $\int_2^{+\infty} \frac{1}{x^{\frac{1+\alpha}{2}}} dx$ converge et il en est de même de $I_{\alpha,\beta}$.

Si $\alpha = 1$, en faisant le changement de variable $u = \ln x$, on obtient que $I_{\alpha,\beta}$ a même nature que $\int_{\ln 2}^{+\infty} \frac{du}{u^\beta}$. Elle converge donc si, et seulement si, $\beta > 1$.

2. Pour étudier la nature de $J_{\alpha,\beta} = \int_0^{\frac{1}{2}} \frac{1}{x^\alpha (|\ln x|)^\beta}$, on fait le changement de variable $u = \frac{1}{x}$.

Il montre que $J_{\alpha,\beta}$ a même nature que $\int_2^{+\infty} \frac{u^\alpha}{(\ln u)^\beta u^2} du = I_{2-\alpha,\beta}$. D'après ce qui précède, $J_{\alpha,\beta}$ converge si, et seulement si, $2 - \alpha > 1$ ou $2 - \alpha = 1$ et $\beta > 1$, i.e. $\alpha < 1$ ou $\alpha = 1$ et $\beta > 1$.

3 1. La fonction $f_n : x \mapsto \frac{1}{(x^2 + 2x + 3)^n}$ est continue et positive sur \mathbb{R} . Comme

$f_n(x) \underset{x \rightarrow +\infty}{\sim} \frac{1}{x^{2n}}$, l'intégrale I_n converge pour $n \geq 1$, car $2n > 1$. Le changement de variable

$t = x + 1$ donne $I_n = \int_{-\infty}^{+\infty} \frac{1}{(t^2 + 2)^n} dt$. On calcule I_n en intégrant par parties. On pose,

pour $n \geq 1$, $u'(t) = 1$, $v(t) = \frac{1}{(t^2 + 2)^n}$ et donc $u(t) = t$, $v'(t) = -\frac{2nt}{(t^2 + 2)^{n+1}}$. Comme uv a pour limite 0 en $\pm\infty$, on obtient

$$\begin{aligned} I_n &= \int_{-\infty}^{+\infty} u'(t)v(t) dt = - \int_{-\infty}^{+\infty} u(t)v'(t) dt = \int_{-\infty}^{+\infty} \frac{2nt^2}{(t^2 + 2)^{n+1}} dt \\ &= 2n \int_{-\infty}^{+\infty} \frac{(t^2 + 2) - 2}{(t^2 + 2)^{n+1}} dt = 2n(I_n - 2I_{n+1}), \end{aligned}$$

c'est-à-dire

$$I_{n+1} = \frac{2n-1}{4n} I_n.$$

On calcule I_1 :

$$I_1 = \int_{-\infty}^{\infty} \frac{1}{t^2+2} dt = \left[\frac{1}{\sqrt{2}} \operatorname{Arctan} \frac{t}{\sqrt{2}} \right]_{-\infty}^{+\infty} = \frac{\pi}{\sqrt{2}}.$$

On en déduit que, pour $n \geq 1$,

$$I_n = \prod_{k=1}^{n-1} \frac{2k-1}{4k} I_1 = \prod_{k=1}^{n-1} \frac{(2k-1)2k}{8k^2} \frac{\pi}{\sqrt{2}} = \frac{(2n-2)! \pi}{8^{n-1} ((n-1)!)^2 \sqrt{2}}.$$

2. La fonction $g_n : x \mapsto x^n e^{-x}$ est continue sur \mathbb{R}_+ et au voisinage de $+\infty$, $g_n(x) = o\left(\frac{1}{x^2}\right)$.

On en déduit que $J_n = \int_0^{+\infty} g_n(x) dx$ converge pour tout $n \in \mathbb{N}$. On la calcule en intégrant par parties. On a

$$I_0 = \int_0^{+\infty} e^{-x} dx = 1$$

et pour tout $n \in \mathbb{N}$,

$$I_{n+1} = \int_0^{+\infty} x^{n+1} e^{-x} dx = - \int_0^{+\infty} (n+1)x^n (-e^{-x}) dx = (n+1)I_n,$$

car

$$\lim_{x \rightarrow 0} x^{n+1} (-e^{-x}) = \lim_{x \rightarrow +\infty} x^{n+1} (-e^{-x}) = 0.$$

On en déduit que pour tout $n \in \mathbb{N}$, $I_n = n!$.

3. La fonction $f : x \mapsto \frac{\ln x}{(x+1)^2}$ est continue sur $]0, +\infty[$. Au voisinage de 0, elle est négative

et $f(x) \sim \ln x$. Comme $\int_0^1 \ln x dx$ converge (cela a été démontré dans le cours), $\int_0^1 f(x) dx$

converge également. Au voisinage de $+\infty$, on a $f(x) \sim \frac{\ln x}{x^2}$ et donc $f(x) = o\left(\frac{1}{x^{\frac{3}{2}}}\right)$. On

en déduit que $\int_1^{+\infty} f(x) dx$ converge puisque $\frac{3}{2} > 1$.

Pour calculer l'intégrale, on intègre par parties en posant $u(x) = \ln x$ et $v(x) = -\frac{1}{x+1}$.

Comme uv n'a pas une limite finie en 0, on intègre d'abord sur $[a, +\infty[$ pour $a > 0$. On obtient

$$\begin{aligned} \int_a^{+\infty} f(x) dx &= \left[-\frac{\ln x}{x+1} \right]_a^{+\infty} - \int_a^{+\infty} \frac{-1}{x(x+1)} dx \\ &= \frac{\ln a}{a+1} + \int_a^{+\infty} \left(\frac{1}{x} - \frac{1}{x+1} \right) dx = \frac{\ln a}{a+1} + \left[\ln \frac{x}{x+1} \right]_a^{+\infty} \\ &= \frac{\ln a}{a+1} - \ln \frac{a}{a+1} = -\frac{a \ln a}{a+1} + \ln(a+1). \end{aligned}$$

En faisant tendre a vers 0, on obtient $\int_0^{+\infty} \frac{\ln x}{(x+1)^2} dx = 0$.

On peut obtenir ce résultat différemment, par le changement de variable $t = \frac{1}{x}$. On obtient

$$\begin{aligned} \int_0^{+\infty} \frac{\ln x}{(x+1)^2} dx &= - \int_0^{+\infty} \frac{\ln \frac{1}{t}}{\left(\frac{1}{t}+1\right)^2} \left(-\frac{dt}{t^2}\right) \\ &= - \int_0^{+\infty} \frac{\ln t}{(t+1)^2} dt = - \int_0^{+\infty} \frac{\ln x}{(x+1)^2} dx \end{aligned}$$

et donc $\int_0^{+\infty} \frac{\ln x}{(x+1)^2} dx = 0$.

4. La fonction $g : x \mapsto 1 - x \operatorname{Arctan} \frac{1}{x}$ est continue sur \mathbb{R}_+^* . Comme $\lim_{x \rightarrow 0} \operatorname{Arctan} \frac{1}{x} = \frac{\pi}{2}$, on a $\lim_{x \rightarrow 0} g(x) = 1$: il y a une « fausse impropriété » en 0. En utilisant le développement limité de Arctan en 0, on obtient, quand x tend vers $+\infty$,

$$g(x) = x \left(\frac{1}{x} - \operatorname{Arctan} \frac{1}{x} \right) = x \left(\frac{1}{3x^3} + o\left(\frac{1}{x^3}\right) \right) \sim \frac{1}{3x^2}.$$

On en déduit que $\int_0^{+\infty} g(x) dx$ converge.

On intègre par parties. On écrit $g(x) = x \left(\frac{1}{x} - \operatorname{Arctan} \frac{1}{x} \right)$. On pose $u'(x) = x$, $v(x) = \frac{1}{x} - \operatorname{Arctan} \frac{1}{x}$, $u(x) = \frac{1}{2}x^2$, $v'(x) = -\frac{1}{x^2} + \frac{1}{x^2} \frac{1}{1 + \frac{1}{x^2}} = -\frac{1}{x^2(x^2 + 1)}$. Comme

$\lim_0 uv = \lim_{+\infty} uv = 0$, car $u(x)v(x) \underset{x \rightarrow +\infty}{\sim} -\frac{1}{6x}$, on a

$$\int_0^{+\infty} g(x) dx = - \int_0^{+\infty} u(x)v'(x) dx = \frac{1}{2} \int_0^{+\infty} \frac{1}{1+x^2} dx = \frac{\pi}{4}.$$

5. La fonction $h : x \mapsto \frac{1}{\sqrt{1+e^x}}$ est continue sur \mathbb{R}_+ . Comme $h(x) \underset{x \rightarrow +\infty}{\sim} e^{-\frac{x}{2}}$ et que $\int_0^{+\infty} e^{-\frac{x}{2}} dx$ converge, l'intégrale $\int_0^{+\infty} h(x) dx$ converge. En faisant le changement de variable $t = \sqrt{1+e^x}$ et donc $x = \ln(t^2 - 1)$, on obtient

$$\begin{aligned} \int_0^{+\infty} h(x) dx &= \int_{\sqrt{2}}^{+\infty} \frac{1}{t} \frac{2tdt}{t^2-1} = \int_{\sqrt{2}}^{+\infty} \frac{2}{t^2-1} dt = \int_{\sqrt{2}}^{+\infty} \left(\frac{1}{t-1} - \frac{1}{t+1} \right) dt \\ &= \left[\ln \frac{t-1}{t+1} \right]_{\sqrt{2}}^{+\infty} = -\ln \frac{\sqrt{2}-1}{\sqrt{2}+1} = 2 \ln(\sqrt{2}+1). \end{aligned}$$

- 4 1. Les fonctions $f : t \mapsto \frac{1}{1+t^4}$ et $g : t \mapsto \frac{t^2}{1+t^4}$ sont continues sur \mathbb{R}_+ . Comme en $+\infty$, $f(t) \sim \frac{1}{t^4}$ et $g(t) \sim \frac{1}{t^2}$, les intégrales convergent. La fonction $t \mapsto \frac{1}{t}$ étant une bijection décroissante de classe \mathcal{C}^1 de $]0, +\infty[$ sur $]0, +\infty[$, on obtient par le théorème de changement de variable

$$I = - \int_0^{+\infty} \frac{1}{1+\frac{1}{u^4}} \frac{-du}{u^2} = \int_0^{+\infty} \frac{u^2}{1+u^4} du = J.$$

2. Considérons $I + J = \int_0^{+\infty} \frac{1+t^2}{1+t^4} dt = \int_0^{+\infty} \frac{1+\frac{1}{t^2}}{t^2+\frac{1}{t^2}} dt$. La fonction $\varphi : t \mapsto t - \frac{1}{t}$ est strictement croissante et réalise une bijection de $]0, +\infty[$ sur \mathbb{R} . Comme, pour $t > 0$, $\varphi'(t) = 1 + \frac{1}{t^2}$ et $t^2 + \frac{1}{t^2} = (\varphi(t))^2 + 2$, on obtient

$$I + J = \int_0^{+\infty} \frac{\varphi'(t) dt}{(\varphi(t))^2 + 2} = \int_{-\infty}^{+\infty} \frac{du}{u^2 + 2} = \frac{1}{\sqrt{2}} \left[\operatorname{Arctan} \frac{u}{\sqrt{2}} \right]_{-\infty}^{+\infty} = \frac{\pi}{\sqrt{2}}.$$

On en déduit

$$I = J = \frac{\pi}{2\sqrt{2}}.$$

- 5 1. La fonction $f : x \mapsto \ln \sin x$ est continue sur $]0, \frac{\pi}{2}]$ et négative. En 0, on a $\sin x \sim x$ et donc $f(x) \sim \ln x$. Comme $\int_0^{\frac{\pi}{2}} \ln x dx$ converge, il en est de même de $\int_0^{\frac{\pi}{2}} \ln \sin x dx$. En faisant le changement de variable $t = \frac{\pi}{2} - x$, on obtient, comme $x \mapsto \frac{\pi}{2} - x$ est une bijection décroissante de l'intervalle $]0, \frac{\pi}{2}[$ sur lui-même,

$$I = - \int_0^{\frac{\pi}{2}} \ln \sin \left(\frac{\pi}{2} - t \right) (-dt) = \int_0^{\frac{\pi}{2}} \ln \cos t dt = J.$$

Ainsi J converge et $J = I$.

2. On en déduit que

$$\begin{aligned} 2I = I + J &= \int_0^{\frac{\pi}{2}} (\ln \sin x + \ln \cos x) dx = \int_0^{\frac{\pi}{2}} \ln \left(\frac{\sin(2x)}{2} \right) dx \\ &= \int_0^{\frac{\pi}{2}} \ln \sin(2x) dx - \frac{\pi \ln 2}{2}. \end{aligned}$$

En faisant le changement de variable $u = 2x$ dans l'intégrale, on obtient

$$2I = \frac{1}{2} \int_0^{\pi} \ln \sin u du - \frac{\pi \ln 2}{2} = \frac{1}{2} \left(I + \int_{\frac{\pi}{2}}^{\pi} \ln \sin u du \right) - \frac{\pi \ln 2}{2}.$$

En faisant le changement de variable $t = \pi - u$, on obtient $\int_{\frac{\pi}{2}}^{\pi} \ln \sin u \, du = \int_0^{\frac{\pi}{2}} \ln \sin t \, dt = I$

et donc $2I = I - \frac{\pi \ln 2}{2}$, c'est-à-dire

$$I = -\frac{\pi \ln 2}{2}.$$

6 1. La fonction $f : t \mapsto e^{-\alpha t} \frac{\sin t}{t}$ est continue sur \mathbb{R}_+^* et, pour tout $t > 0$, $|\sin t| \leq t$ donc $|f(t)| \leq e^{-\alpha t}$. Comme $\int_0^{+\infty} e^{-\alpha t} \, dt$ converge, I est absolument convergente.

2. On a, pour tout $n \in \mathbb{N}$, $e^{-\alpha t} t^n = o\left(\frac{1}{t^2}\right)$ quand t tend vers $+\infty$, donc I_n converge.

On a $I_0 = \left[\frac{e^{-\alpha t}}{-\alpha} \right]_0^{+\infty} = \frac{1}{\alpha}$ et, en intégrant par parties, pour $n \geq 1$,

$$I_n = \left[\frac{e^{-\alpha t} t^n}{-\alpha} \right]_0^{+\infty} - \int_0^{+\infty} \frac{e^{-\alpha t}}{-\alpha} n t^{n-1} \, dt = \frac{n}{\alpha} I_{n-1}.$$

On en déduit que, pour tout $n \in \mathbb{N}$,

$$I_n = \frac{n!}{\alpha^{n+1}}.$$

3. a) Cette inégalité résulte simplement de l'application de l'inégalité de Taylor-Lagrange.
b) On a d'après les deux premières questions

$$\begin{aligned} \left| I - \sum_{k=0}^n \frac{(-1)^k}{(2k+1)\alpha^{2k+1}} \right| &= \left| I - \sum_{k=0}^n (-1)^k \frac{I_{2k}}{(2k+1)!} \right| \\ &\leq \left| \int_0^{+\infty} e^{-\alpha t} \frac{\sin t}{t} \, dt - \sum_{k=0}^n \int_0^{+\infty} (-1)^k e^{-\alpha t} \frac{t^{2k}}{(2k+1)!} \, dt \right| \\ &\leq \int_0^{+\infty} \frac{e^{-\alpha t}}{t} \left| \sin t - \sum_{k=0}^n (-1)^k \frac{t^{2k+1}}{(2k+1)!} \right| \, dt. \end{aligned}$$

En utilisant l'inégalité de la question précédente, on obtient

$$\left| I - \sum_{k=0}^n \frac{(-1)^k}{(2k+1)\alpha^{2k+1}} \right| \leq \int_0^{+\infty} e^{-\alpha t} \frac{t^{2n+1}}{(2n+2)!} \, dt \leq \frac{I_{2n+1}}{(2n+2)!} \leq \frac{1}{(2n+2)\alpha^{2n+2}}.$$

Comme $\alpha > 1$, on en déduit que

$$I = \lim_{n \rightarrow +\infty} \sum_{k=0}^n \frac{(-1)^k}{(2k+1)\alpha^{2k+1}}.$$

4. a) Pour $x \in [0, 1]$ et $n \in \mathbb{N}^*$, on a

$$\begin{aligned} \sum_{k=0}^n (-1)^k \frac{x^{2k+1}}{2k+1} - \operatorname{Arctan} x &= \sum_{k=0}^n \int_0^x (-t^2)^k dt - \int_0^x \frac{1}{1+t^2} dt \\ &= \int_0^x \frac{1 - (-t^2)^{n+1}}{1+t^2} dt - \int_0^x \frac{1}{1+t^2} dt \\ &= \int_0^x \frac{-(-t^2)^{n+1}}{1+t^2} dt. \end{aligned}$$

On en déduit que

$$\left| \sum_{k=0}^n (-1)^k \frac{x^{2k+1}}{2k+1} - \operatorname{Arctan} x \right| \leq \int_0^x \frac{t^{2n+2}}{1+t^2} dt \leq \int_0^1 t^{2n+2} dt \leq \frac{1}{2n+3}.$$

b) De la question précédente, on déduit que, pour tout $x \in [0, 1]$,

$$\operatorname{Arctan} x = \lim_{n \rightarrow +\infty} \sum_{k=0}^n (-1)^k \frac{x^{2k+1}}{2k+1}.$$

En particulier, comme $\frac{1}{\alpha} \in [0, 1]$,

$$I = \lim_{n \rightarrow +\infty} \sum_{k=0}^n \frac{(-1)^k}{(2k+1)\alpha^{2k+1}} = \operatorname{Arctan} \frac{1}{\alpha}.$$

7

1. Pour tout $x > 0$ l'intégrale $\int_x^{+\infty} \frac{\cos t}{t} dt$ convergente. Pour le montrer, on intègre par parties. Pour $X > x$,

$$\int_x^X \frac{\cos t}{t} dt = \left[\frac{\sin t}{t} \right]_x^X + \int_x^X \frac{\sin t}{t^2} dt = \frac{\sin X}{X} - \frac{\sin x}{x} + \int_x^X \frac{\sin t}{t^2} dt.$$

Comme $\left| \frac{\sin t}{t^2} \right| \leq \frac{1}{t^2}$, l'intégrale $\int_x^{+\infty} \frac{\sin t}{t^2} dt$ est absolument convergente. Comme de plus, $\lim_{X \rightarrow +\infty} \frac{\sin X}{X} = 0$, $\int_x^X \frac{\cos t}{t} dt$ a une limite quand X tend vers $+\infty$. Ainsi $\int_x^{+\infty} \frac{\cos t}{t} dt$ converge et de plus

$$f(x) = -\frac{\sin x}{x} + \int_x^{+\infty} \frac{\sin t}{t^2} dt.$$

Si F est une primitive de $x \mapsto \frac{\cos x}{x}$ sur \mathbb{R}_+^* , on a, pour $x > 0$, $f(x) = \lim_{+\infty} F - F(x)$. On en déduit que f est dérivable sur \mathbb{R}_+^* et

$$f'(x) = -\frac{\cos x}{x}.$$

2. En intégrant de nouveau par parties, on obtient, pour $x > 0$,

$$\begin{aligned} f(x) + \frac{\sin x}{x} &= \int_x^{+\infty} \frac{\sin t}{t^2} dt = \left[\frac{-\cos t}{t^2} \right]_x^{+\infty} - \int_x^{+\infty} \frac{2 \cos t}{t^3} dt \\ &= \frac{\cos x}{x^2} - \int_x^{+\infty} \frac{2 \cos t}{t^3} dt \end{aligned}$$

puis

$$\left| f(x) + \frac{\sin x}{x} \right| \leq \frac{1}{x^2} + \int_x^{+\infty} \frac{2}{t^3} dt = \frac{1}{x^2} - \left[\frac{1}{t^2} \right]_x^{+\infty} \leq \frac{2}{x^2}.$$

3. On a, pour tout $x > 0$, $\ln x = \int_1^x \frac{1}{t} dt = - \int_x^1 \frac{1}{t} dt$ et donc

$$f(x) - \ln x = \int_x^1 \frac{\cos t}{t} dt + \int_1^{+\infty} \frac{\cos t}{t} dt - \int_x^1 \frac{1}{t} dt = \int_x^1 \frac{\cos t - 1}{t} dt + \int_1^{+\infty} \frac{\cos t}{t} dt.$$

La fonction $t \mapsto \frac{\cos t - 1}{t}$ est continue sur $]0, 1]$ et $\lim_{t \rightarrow 0} \frac{\cos t - 1}{t} = \cos'(0) = 0$. Ainsi

$\int_0^1 \frac{\cos t - 1}{t} dt$ converge et

$$\lim_{x \rightarrow 0} (f(x) + \ln x) = \int_0^1 \frac{\cos t - 1}{t} dt + \int_1^{+\infty} \frac{\cos t}{t} dt.$$

On a donc, au voisinage de 0, $f(x) + \ln x = o(\ln x)$ et donc $f(x) \sim -\ln x$.

4. La première question montre que f est de classe \mathcal{C}^1 sur \mathbb{R}_+^* . Si a et b sont deux réels tels que $0 < a < b$, on obtient en intégrant par parties

$$\begin{aligned} \int_a^b f(x) dx &= [xf(x)]_a^b - \int_a^b xf'(x) dx = bf(b) - af(a) + \int_a^b \cos x dx \\ &= bf(b) + \sin b - af(a) - \sin a. \end{aligned}$$

Il résulte de la majoration établie dans la question 2 que $|bf(b) + \sin b| \leq \frac{2}{b}$ et donc $\lim_{b \rightarrow +\infty} (f(b) + \sin b) = 0$. Comme $af(a) \sim -a \ln a$, on a $\lim_{a \rightarrow 0} af(a) = 0$. Il en résulte que $\int_a^b f(x) dx$ a pour limite 0 quand a tend vers 0 et b vers $+\infty$. Cela montre que $\int_0^{+\infty} f(x) dx$ converge et est égale à 0.

8 1. On a, pour tous réels x et h ,

$$\begin{aligned} \left| f(x+h) - f(x) + h \int_0^1 e^{-x(1+t^2)} dt \right| &= \left| \int_0^1 \left(\frac{e^{-(x+h)(1+t^2)} - e^{-x(1+t^2)}}{1+t^2} + he^{-x(1+t^2)} \right) dt \right| \\ &\leq \int_0^1 \frac{e^{x(1+t^2)}}{1+t^2} \left| e^{-h(1+t^2)} - 1 + h(1+t^2) \right| dt. \end{aligned}$$

En appliquant l'inégalité de Taylor-Lagrange à l'ordre 1 à la fonction exponentielle entre 0 et $-h(1+t^2)$, on obtient

$$\left| e^{-h(1+t^2)} - 1 + h(1+t^2) \right| \leq \frac{1}{2}(h(1+t^2))^2 M,$$

où M est un majorant de la fonction exponentielle entre 0 et $-h(1+t^2)$. On peut prendre $M = e^{|h|(1+t^2)}$. Pour $t \in [0, 1]$, on obtient

$$\left| e^{-h(1+t^2)} - 1 + h(1+t^2) \right| \leq \frac{1}{2}(h(1+t^2))^2 e^{|h|(1+t^2)} \leq 2h^2 e^{2|h|}.$$

On en déduit

$$\left| f(x+h) - f(x) + h \int_0^1 e^{-x(1+t^2)} dt \right| \leq 2h^2 e^{2|h|} \int_0^1 \frac{e^{x(1+t^2)}}{1+t^2} dt.$$

En divisant par $|h|$ et en faisant tendre h vers 0, on obtient, pour tout $x \in \mathbb{R}$,

$$f'(x) = - \int_0^1 e^{-x(1+t^2)} dt.$$

2. La fonction g est dérivable sur \mathbb{R} et, pour tout réel x ,

$$g'(x) = 2xf'(x^2) = -2x \int_0^1 e^{-x^2(1+t^2)} dt = -2e^{-x^2} \int_0^1 e^{-(xt)^2} x dt.$$

Le changement de variable $u = xt$ donne

$$g'(x) = -2e^{-x^2} \int_0^x e^{-u^2} du.$$

La fonction $h : x \mapsto g(x) + \left(\int_0^x e^{-t^2} dt \right)^2$ est dérivable sur \mathbb{R} et

$$h'(x) = g'(x) + 2e^{-x^2} \int_0^x e^{-t^2} dt = 0.$$

La fonction h est constante. Comme

$$h(0) = g(0) = f(0) = \int_0^1 \frac{1}{1+t^2} dt = \text{Arctan } 1 = \frac{\pi}{4},$$

on a, pour tout réel x , $h(x) = \frac{\pi}{4}$.

3. D'après la question précédente, pour tout réel $x \geq 0$, $\int_0^x e^{-t^2} dt = \sqrt{\frac{\pi}{4} - g(x)}$. On fait tendre x vers $+\infty$. De

$$0 \leq g(x) = \int_0^1 \frac{e^{-x^2(1+t^2)}}{1+t^2} dt \leq \int_0^1 e^{-x^2} dt = e^{-x^2},$$

on déduit $\lim_{x \rightarrow +\infty} g(x) = 0$ et donc

$$\int_0^{+\infty} e^{-t^2} dt = \sqrt{\frac{\pi}{4}} = \frac{\sqrt{\pi}}{2}.$$

En faisant le changement de variable $t = \sqrt{u}$, on obtient

$$\int_0^{+\infty} e^{-t^2} dt = \int_0^{+\infty} e^{-u} \frac{1}{2} u^{-\frac{1}{2}} du = \frac{1}{2} \Gamma\left(\frac{1}{2}\right).$$

On en déduit

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

9 1. Pour tout $\alpha > 0$, l'intégrale $\int_{\alpha\alpha}^{+\infty} \frac{f(t)}{t} dt$ converge, donc par le changement de variable $t = ax$, on obtient la convergence de $\int_a^{+\infty} \frac{f(ax)}{x} dx$ et l'égalité

$$\int_{\alpha\alpha}^{+\infty} \frac{f(t)}{t} dt = \int_{\alpha}^{+\infty} \frac{f(ax)}{ax} a dx = \int_{\alpha}^{+\infty} \frac{f(ax)}{x} dx.$$

On montre de même la convergence de $\int_b^{+\infty} \frac{f(bx)}{x} dx$ et l'égalité

$$\int_{\alpha}^{+\infty} \frac{f(bx)}{x} dx = \int_{b\alpha}^{+\infty} \frac{f(t)}{t} dt.$$

On en déduit la convergence de $\int_{\alpha}^{+\infty} \frac{f(bx) - f(ax)}{x} dx$ et l'égalité

$$\begin{aligned} \int_{\alpha}^{+\infty} \frac{f(bx) - f(ax)}{x} dx &= \int_{\alpha}^{+\infty} \frac{f(bx)}{x} dx - \int_{\alpha}^{+\infty} \frac{f(ax)}{x} dx \\ &= \int_{b\alpha}^{+\infty} \frac{f(t)}{t} dx - \int_{\alpha\alpha}^{+\infty} \frac{f(t)}{t} dt \\ &= \int_{b\alpha}^{\alpha\alpha} \frac{f(t)}{t} dt. \end{aligned}$$

2. Le résultat est évident si $a = b$. Supposons donc $a \neq b$. On remarque que, pour tout $\alpha > 0$,

$$f(0) \ln \frac{a}{b} = f(0) \int_{b\alpha}^{\alpha\alpha} \frac{1}{t} dt \text{ et donc}$$

$$\left| \int_{b\alpha}^{\alpha\alpha} \frac{f(t)}{t} dt - f(0) \ln \frac{a}{b} \right| \leq \int_{\alpha\alpha}^{b\alpha} \frac{|f(t) - f(0)|}{t} dt.$$

Soit $\varepsilon > 0$. Comme f est continue en 0, il existe $\eta > 0$ tel que $0 \leq t \leq \eta$ implique $|f(t) - f(0)| \leq \frac{\varepsilon}{\left| \ln \frac{a}{b} \right|}$. Soit $\alpha > 0$ tel que $\alpha < \frac{\eta}{\max(a, b)}$. On a alors, pour tout t entre $\alpha\alpha$

et $b\alpha$, $|t| \leq \eta$ et donc $|f(t) - f(0)| \leq \frac{\varepsilon}{\left|\ln \frac{a}{b}\right|}$. On en déduit que

$$\left| \int_{b\alpha}^{a\alpha} \frac{f(t)}{t} dt - f(0) \ln \frac{a}{b} \right| \leq \varepsilon.$$

Cela montre que

$$\lim_{\alpha \rightarrow 0} \int_{\alpha}^{+\infty} \frac{f(bx) - f(ax)}{x} dx = \lim_{\alpha \rightarrow 0} \int_{b\alpha}^{a\alpha} \frac{f(t)}{t} dt = f(0) \ln \frac{a}{b}.$$

Ainsi $\int_0^{+\infty} \frac{f(bx) - f(ax)}{x} dx$ converge et vaut $f(0) \ln \frac{a}{b}$.

3. On applique le résultat de la question 2 à la fonction $x \mapsto e^{-x}$. On en déduit que

$$\int_0^{+\infty} \frac{e^{-bx} - e^{-ax}}{x} dx = \ln \frac{a}{b}.$$

4. La fonction $t \mapsto -\ln t$ est de classe \mathcal{C}^1 . Elle réalise une bijection décroissante de $]0, 1[$ sur $]0, +\infty[$. On en déduit, par le changement de variable $u = -\ln t$, que l'intégrale

$\int_0^1 \frac{t-1}{\ln t} dt$ a même nature (et en cas de convergence une valeur opposée) que

$$\int_0^{+\infty} \frac{e^{-u} - 1}{-u} (-e^{-u}) du = \int_0^{+\infty} \frac{e^{-2u} - e^{-u}}{u} du.$$

D'après la question précédente, cette intégrale converge et vaut $\ln \frac{1}{2}$. On en déduit que

$\int_0^1 \frac{\ln t}{t-1} dt$ converge et vaut $\ln 2$.

5. On a, pour tout réel t ,

$$\begin{aligned} \sin^3 t &= \sin t \frac{1 - \cos(2t)}{2} = \frac{1}{2} (\sin t - \sin t \cos(2t)) \\ &= \frac{1}{2} \left(\sin t - \frac{1}{2} (\sin(3t) - \sin t) \right) = \frac{3}{4} \sin t - \frac{1}{4} \sin(3t). \end{aligned}$$

On peut écrire pour $t > 0$,

$$\frac{\sin^3 t}{t^2} = \frac{3}{4t} \left(\frac{\sin t}{t} - \frac{\sin(3t)}{3t} \right).$$

On considère la fonction f définie sur \mathbb{R}_+ par $f(t) = \frac{3 \sin t}{4t}$ si $t > 0$ et $f(0) = \frac{3}{4}$. On obtient une fonction continue sur \mathbb{R}_+ telle que $\int_1^{+\infty} \frac{f(t)}{t} dt$ converge, car, pour tout $t \geq 1$,

$\left| \frac{f(t)}{t} \right| \leq \frac{3}{4t^2}$. Comme, pour tout $t > 0$,

$$\frac{\sin^3 t}{t^2} = \frac{f(t) - f(3t)}{t},$$

il résulte de la question 1 que $\int_0^{+\infty} \frac{\sin^3 t}{t^2} dt$ converge et vaut $f(0) \ln 3 = \frac{3}{4} \ln 3$.

- 10** 1. Pour tout $n \in \mathbb{N}$, la fonction $f_n : t \mapsto \frac{\sin((2n+1)t)}{\sin t}$ est continue sur $\left]0, \frac{\pi}{2}\right[$ et a pour limite $2n+1$ en 0. On en déduit que l'intégrale $I_n = \int_0^{\frac{\pi}{2}} f_n(t) dt$ converge. Pour tout $n \in \mathbb{N}^*$, on a

$$\begin{aligned} I_n - I_{n-1} &= \int_0^{\frac{\pi}{2}} \frac{\sin((2n+1)t) - \sin((2n-1)t)}{\sin t} dt \\ &= \int_0^{\frac{\pi}{2}} 2 \frac{\sin t \cos(2nt)}{\sin t} dt = \int_0^{\frac{\pi}{2}} 2 \cos(2nt) dt = 0. \end{aligned}$$

La suite (I_n) est donc constante et pour $n \in \mathbb{N}$, $I_n = I_0 = \frac{\pi}{2}$.

2. En intégrant par parties, on obtient, pour $n \geq 1$,

$$J_n = \int_a^b f(t) \sin nt dt = \left[-\frac{f(t) \cos nt}{n} \right]_a^b + \int_a^b \frac{f'(t) \cos nt}{n} dt,$$

puis

$$|J_n| \leq \frac{|f(a)| + |f(b)|}{n} + \frac{1}{n} \int_a^b |f'(t)| dt.$$

On en déduit que $\lim_{n \rightarrow +\infty} J_n = 0$.

3. La fonction f est de classe \mathcal{C}^1 sur $\left]0, \frac{\pi}{2}\right[$. On a au voisinage de 0,

$$\begin{aligned} f'(t) &= -\frac{1}{t^2} + \frac{\cos t}{\sin^2 t} = \frac{-\sin^2 t + t^2 \cos t}{t^2 \sin^2 t} \\ &= \frac{-\left(t - \frac{t^3}{6} + o(t^4)\right)^2 + t^2 \left(1 - \frac{t^2}{2} + o(t^2)\right)}{t^2 \sin^2 t} \\ &= \frac{-t^2 + \frac{t^4}{3} + t^2 - \frac{t^4}{2} + o(t^4)}{t^4 + o(t^4)} = -\frac{1}{6} + o(1). \end{aligned}$$

On en déduit que $\lim_{t \rightarrow 0} f'(t) = -\frac{1}{6}$. D'après le théorème de prolongement des fonctions de classe \mathcal{C}^1 , la fonction f est prolongeable en une fonction de classe \mathcal{C}^1 sur $\left[0, \frac{\pi}{2}\right]$.

4. On pose $J_n = \int_0^{\frac{\pi}{2}} f(t) \sin(nt) dt$ où f est la fonction définie dans la question précédente, prolongée en une fonction de classe C^1 sur $[0, \frac{\pi}{2}]$. On obtient, pour tout entier naturel n ,

$$J_{2n+1} = \int_0^{\frac{\pi}{2}} \frac{\sin((2n+1)t)}{t} dt - I_{2n+1} = \int_0^{\frac{(2n+1)\pi}{2}} \frac{\sin u}{u} du - \frac{\pi}{2},$$

en faisant le changement de variable $u = (2n+1)t$ et en utilisant le résultat de la première question. Quand n tend vers $+\infty$, J_{2n+1} tend vers 0 d'après la question 3, et $\int_0^{\frac{(2n+1)\pi}{2}} \frac{\sin u}{u} du$ vers I . Par passage à la limite, on en déduit que

$$I = \frac{\pi}{2}.$$

11 1. La fonction à intégrer est continue et positive sur \mathbb{R}^+ et équivalente au voisinage de l'infini à $\frac{1}{t^{\alpha n}}$. Comme $\alpha > 1$ et $n \geq 1$, on a $\alpha n > 1$ et l'intégrale converge par comparaison à une intégrale de Riemann.

2. Pour $n \in \mathbb{N}^*$, en intégrant par parties, on obtient

$$\begin{aligned} u_n(\alpha) &= \int_0^{+\infty} \frac{1}{(1+t^\alpha)^n} dt = \left[\frac{t}{(1+t^\alpha)^n} \right]_0^{+\infty} - \int_0^{+\infty} \frac{-n\alpha t^\alpha}{(1+t^\alpha)^{n+1}} dt \\ &= n\alpha \int_0^{+\infty} \frac{1+t^\alpha-1}{(1+t^\alpha)^{n+1}} dt = n\alpha(u_n(\alpha) - u_{n+1}(\alpha)), \end{aligned}$$

c'est-à-dire

$$u_{n+1}(\alpha) = \frac{n\alpha - 1}{\alpha n} u_n(\alpha).$$

Par récurrence, on en déduit, pour tout $n \in \mathbb{N}^*$,

$$u_n(\alpha) = \prod_{k=1}^{n-1} \frac{(k\alpha - 1)}{k\alpha} u_1(\alpha) = \frac{\prod_{k=1}^{n-1} (k\alpha - 1)}{(n-1)! \alpha^{n-1}} u_1(\alpha).$$

3. a) Pour tout $t \geq 0$, $1+t^\alpha \geq 1$, donc pour tout n de \mathbb{N}^* , $\frac{1}{(1+t^\alpha)^{n+1}} \leq \frac{1}{(1+t^\alpha)^n}$ et en intégrant

$$u_{n+1}(\alpha) \leq u_n(\alpha).$$

La suite $(u_n(\alpha))_n$ est décroissante et minorée par 0, donc est convergente.

- b) On a, pour tout $n \in \mathbb{N}^*$,

$$\sum_{k=1}^n \frac{u_k(\alpha)}{k\alpha} = \sum_{k=1}^n (u_k(\alpha) - u_{k+1}(\alpha)) = u_1(\alpha) - u_{n+1}(\alpha).$$

On en déduit que

$$\lim_{n \rightarrow +\infty} \sum_{k=1}^n \frac{u_k(\alpha)}{k\alpha} = u_1(\alpha) - \lim_{n \rightarrow +\infty} u_n(\alpha).$$

Ainsi la série de terme général $\frac{u_n(\alpha)}{n\alpha}$ converge.

- c) Notons ℓ la limite de la suite $(u_n(\alpha))$. Si $\ell \neq 0$, alors $\frac{u_n(\alpha)}{n\alpha} \sim \frac{\ell}{n\alpha}$ et la série de terme général $\frac{u_n(\alpha)}{n\alpha}$ diverge. On a donc

$$\lim_{n \rightarrow \infty} u_n(\alpha) = 0.$$

4. a) Pour $n \geq 1$,

$$w_{n+1}(\alpha) - w_n(\alpha) = \ln \left(\frac{u_{n+1}(\alpha)}{u_n(\alpha)} \right) + \frac{1}{\alpha} \ln \left(1 + \frac{1}{n} \right).$$

De la question 2, on déduit que

$$\begin{aligned} w_{n+1}(\alpha) - w_n(\alpha) &= \ln \left(1 - \frac{1}{n\alpha} \right) + \frac{1}{\alpha} \ln \left(1 + \frac{1}{n} \right) \\ &= -\frac{1}{n\alpha} - \frac{1}{2n^2\alpha^2} + \frac{1}{\alpha} \left(\frac{1}{n} - \frac{1}{2n^2} \right) + o \left(\frac{1}{n^2} \right) \\ &= - \left(\frac{1}{\alpha^2} - \frac{1}{\alpha} \right) \frac{1}{2n^2} + o \left(\frac{1}{n^2} \right). \end{aligned}$$

La comparaison à une série de Riemann donne la convergence de la série de terme général $w_{n+1}(\alpha) - w_n(\alpha)$.

- b) Pour tout $n \geq 1$, on a

$$\sum_{k=1}^{n-1} (w_{k+1}(\alpha) - w_k(\alpha)) = w_n(\alpha) - w_1(\alpha).$$

Comme la série de terme général $w_{n+1}(\alpha) - w_n(\alpha)$ converge, la suite des sommes partielles converge et donc la suite $(w_n(\alpha))$ converge. On note L sa limite. Comme $w_n(\alpha) = \ln \left(u_n(\alpha)n^{\frac{1}{\alpha}} \right)$, on obtient par composition des limites, $\lim_{n \rightarrow +\infty} u_n(\alpha)n^{\frac{1}{\alpha}} = e^L$. On note $K(\alpha)$ cette limite. On a donc

$$u_n(\alpha) \sim \frac{K(\alpha)}{n^{\frac{1}{\alpha}}}.$$

- 12** 1. La fonction g est dérivable sur \mathbb{R}_+^* comme quotient de fonctions dérivables, donc continue. Il faut vérifier la continuité en 0. En notant F une primitive de f sur \mathbb{R}_+ , on a, pour tout $x > 0$, $g(x) = \frac{F(x) - F(0)}{x}$. On en déduit que

$$\lim_{x \rightarrow 0} g(x) = \lim_{x \rightarrow 0} \frac{F(x) - F(0)}{x} = F'(0) = f(0) = g(0)$$

et la continuité de g en 0. Pour tout $x > 0$,

$$g'(x) = -\frac{1}{x^2} \int_0^x f(t) dt + \frac{f(x)}{x} = \frac{f(x) - g(x)}{x}.$$

2. Soit $x > 0$.

a) Pour $\varepsilon \in]0, x[$, on peut intégrer par parties sur $[\varepsilon, x]$ car g est \mathcal{C}^1 sur \mathbb{R}_+^* . On obtient

$$\begin{aligned} \int_{\varepsilon}^x g^2(t) dt &= [tg^2(t)]_{\varepsilon}^x - \int_{\varepsilon}^x 2tg(t)g'(t) dt \\ &= xg^2(x) - \varepsilon g^2(\varepsilon) - 2 \int_{\varepsilon}^x g(t)(f(t) - g(t)) dt \\ &= xg^2(x) - \varepsilon g^2(\varepsilon) - 2 \int_{\varepsilon}^x g(t)f(t) dt + 2 \int_{\varepsilon}^x g^2(t) dt. \end{aligned}$$

En simplifiant, on obtient

$$\int_{\varepsilon}^x g^2(t) dt = -xg^2(x) + \varepsilon g^2(\varepsilon) + 2 \int_{\varepsilon}^x g(t)f(t) dt,$$

puis en faisant tendre ε vers 0,

$$\int_0^x g^2(t) dt = -xg^2(x) + 2 \int_0^x g(t)f(t) dt.$$

b) Puisque x est positif, on en déduit

$$\int_0^x g^2(t) dt \leq 2 \int_0^x g(t)f(t) dt$$

En appliquant l'inégalité de Cauchy-Schwarz, on obtient

$$\left(\int_0^x g^2(t) dt \right)^2 \leq 4 \left(\int_0^x g(t)f(t) dt \right)^2 \leq 4 \left(\int_0^x g^2(t) dt \right) \left(\int_0^x f^2(t) dt \right).$$

Si $\int_0^x g^2(t) dt$ n'est pas nulle, on obtient en simplifiant

$$\int_0^x g^2(t) dt \leq 4 \int_0^x f^2(t) dt,$$

inégalité qui est évidente si $\int_0^x g^2(t) dt = 0$.

c) Puisque f^2 est positive et que $\int_0^{+\infty} f^2(t) dt$ converge, on a pour tout $x > 0$,

$$\int_0^x g^2(t) dt \leq 4 \int_0^{+\infty} f^2(t) dt.$$

La fonction g^2 est positive et la fonction $x \mapsto \int_0^x g^2(t) dt$ est majorée, donc $\int_0^{+\infty} g^2(t) dt$ converge et par passage à la limite dans l'inégalité précédente, on obtient

$$\int_0^{+\infty} g^2(t) dt \leq 4 \int_0^{+\infty} f^2(t) dt.$$

3. Comme, pour tout réel x , on a par l'inégalité de Cauchy-Schwarz

$$\begin{aligned} \int_0^x |f(t)g(t)| dt &\leq \sqrt{\int_0^x f^2(t) dt} \sqrt{\int_0^x g^2(t) dt} \\ &\leq \sqrt{\int_0^{+\infty} f^2(t) dt} \sqrt{\int_0^{+\infty} g^2(t) dt}. \end{aligned}$$

La fonction $x \mapsto \int_0^x |f(t)g(t)| dt$ est majorée donc $\int_0^{+\infty} |f(t)g(t)| dt$ converge. On repart de l'égalité démontrée dans la question 2 : pour tout $x > 0$,

$$\int_0^x g^2(t) dt = -xg^2(x) + 2 \int_0^x g(t)f(t) dt.$$

Les deux intégrales ont une limite finie quand x tend vers $+\infty$. Il en est donc de même de leur différence $xg^2(x)$. Notons L la limite de cette expression. Si $L \neq 0$, alors en $+\infty$, $g^2(x) \sim \frac{L}{x}$, ce qui contredit la convergence de $\int_0^{+\infty} g^2(t) dt$. On a donc $L = 0$, et en faisant tendre x vers $+\infty$ dans l'égalité, on obtient

$$\int_0^{+\infty} g^2(t) dt = 2 \int_0^{+\infty} g(t)f(t) dt.$$

13 1. La fonction f_n est continue et positive sur \mathbb{R}_+ . Comme $f_n(x) \sim_{x \rightarrow +\infty} \frac{1}{x^n}$, l'intégrale I_n converge si, et seulement si, $n > 1$. Ainsi $A = \{n \in \mathbb{N}, n \geq 2\}$.

2. a) Si $n \geq 2$, on a pour tout $x \geq 0$,

$$\frac{f_n(x)}{f_2(x)} = \frac{1}{(x+3)(x+4)\dots(x+n)} \leq \frac{1}{3 \cdot 4 \cdot \dots \cdot n} \leq \frac{2}{n!}$$

d'où, $f_n(x) \leq \frac{2}{n!} f_2(x)$ et en intégrant

$$I_n \leq \frac{2I_2}{n!},$$

ce qui est le résultat voulu avec $B = 2I_2$.

b) La série de terme général I_n , qui est à termes positifs (car f_n est positive), est majorée par la série de terme général $\frac{B}{n!}$ qui est convergente, donc elle converge.

3. a) Par décroissance de la fonction $t \mapsto \frac{1}{t}$, on a pour $k \in \mathbb{N}^*$

$$\frac{1}{k+1} \leq \int_k^{k+1} \frac{dt}{t} \leq \frac{1}{k},$$

et en sommant de $k = 1$ à $n - 1$ pour $n \geq 1$, on obtient

$$\sum_{k=1}^{n-1} \frac{1}{k+1} \leq \int_1^n \frac{dt}{t} \leq \sum_{k=1}^n \frac{1}{k},$$

et donc

$$H_n - 1 \leq \ln n \leq H_n - \frac{1}{n},$$

c'est-à-dire

$$\ln n + \frac{1}{n} \leq H_n \leq \ln n + 1.$$

On en déduit que

$$H_n \sim \ln n.$$

b) Pour $x \geq 0$,

$$\frac{f'_n(x)}{f_n(x)} = \frac{d}{dx} [\ln f(x)] = - \sum_{k=1}^n \frac{1}{x+k}.$$

Ainsi pour $x \in [0, 1]$, on a

$$H_{n+1} - 1 = \sum_{k=1}^n \frac{1}{1+k} \leq -\frac{f'_n(x)}{f_n(x)} = \sum_{k=1}^n \frac{1}{x+k} \leq \sum_{k=1}^n \frac{1}{k} = H_n.$$

Puisque f_n est positive, on en déduit, pour $n \geq 1$ et $x \in [0, 1]$,

$$\frac{-f'_n(x)}{H_n} \leq f_n(x) \leq \frac{-f'_n(x)}{H_{n+1} - 1}.$$

c) En intégrant, compte tenu de l'égalité

$$\int_0^1 (-f'_n(x)) dx = f_n(0) - f_n(1) = \frac{1}{n!} - \frac{1}{(n+1)!} = \frac{n}{(n+1)!},$$

on obtient

$$\frac{n}{(n+1)!H_n} \leq \int_0^1 f_n(x) dx \leq \frac{n}{(n+1)!(H_{n+1} - 1)}.$$

D'après a, on a

$$\frac{n}{(n+1)!H_n} \sim \frac{1}{n!H_n} \sim \frac{1}{n! \ln n} \quad \text{et} \quad \frac{n}{(n+1)!(H_{n+1} - 1)} \sim \frac{1}{n! \ln(n+1)} \sim \frac{1}{n! \ln n}.$$

De l'encadrement précédent, on déduit donc

$$\int_0^1 f_n(x) dx \sim \frac{1}{n! \ln n}.$$

4. Pour $n \geq 2$, on peut écrire

$$\int_0^1 f_n(x) dx = \int_0^{+\infty} f_n(x) dx - \int_1^{+\infty} f_n(x) dx.$$

En faisant le changement de variable $t = x - 1$ dans la deuxième intégrale, on obtient

$$\begin{aligned} \int_0^1 f_n(x) dx &= \int_0^{+\infty} \frac{1}{(x+1)\dots(x+n)} dx - \int_0^{+\infty} \frac{1}{(t+2)\dots(t+n+1)} dt \\ &= \int_0^{+\infty} \left(\frac{1}{(x+1)\dots(x+n)} - \frac{1}{(x+2)\dots(x+n+1)} \right) dx \\ &= \int_0^{+\infty} \frac{n}{(x+1)\dots(x+n+1)} dx = nI_{n+1}. \end{aligned}$$

On a donc, pour $n \geq 3$,

$$I_n = \frac{1}{n-1} \int_0^1 f_{n-1}(x) dx \sim \frac{1}{n(n-1)! \ln(n-1)} \sim \frac{1}{n! \ln n}.$$

14 **1. a)** La fonction $f : t \mapsto t^{x-1}(1-t)^{y-1}$ est continue sur $]0, 1[$ et positive.

On a $f(t) \underset{t \rightarrow 0}{\sim} t^{x-1} \underset{t \rightarrow 0}{\sim} \frac{1}{t^{1-x}}$. Comme $1-x < 1$, $\int_0^{\frac{1}{2}} f(t) dt$ converge.

De même, $f(t) \underset{t \rightarrow 1}{\sim} \frac{1}{(1-t)^{1-y}}$, avec $1-y < 1$ donc $\int_{\frac{1}{2}}^1 f(t) dt$ converge.

L'intégrale $B(x, y)$ converge pour $x > 0$ et $y > 0$.

b) Le changement de variable $u = 1 - t$ permet de montrer que $B(x, y) = B(y, x)$. On intègre par parties

$$\begin{aligned} B(x+1, y) &= \int_0^1 t^x (1-t)^{y-1} dt = \underbrace{\left[\frac{t^x (1-t)^y}{-y} \right]_0^1}_{=0} + \frac{x}{y} \int_0^1 t^{x-1} (1-t)^y dt \\ &= \frac{x}{y} B(x, y+1). \end{aligned}$$

Mais on peut écrire

$$\begin{aligned} B(x, y+1) &= \int_0^1 t^{x-1} (1-t)^y dt = \int_0^1 t^{x-1} (1-t)^{y-1} (1-t) dt \\ &= \int_0^1 (t^{x-1} - t^x) (1-t)^{y-1} dt = B(x, y) - B(x+1, y). \end{aligned}$$

On en déduit que $B(x+1, y) = \frac{x}{y} (B(x, y) - B(x+1, y))$, c'est-à-dire que

$$B(x+1, y) = \frac{x}{x+y} B(x, y).$$

c) Pour $x \in \mathbb{N}$, cela fournit une relation de récurrence qui conduit à

$$B(n+1, \gamma) = \prod_{k=1}^n \frac{k}{k+\gamma} B(1, \gamma) = \frac{n!}{\gamma(\gamma+1)\dots(\gamma+n)},$$

car $B(1, \gamma) = \frac{1}{\gamma}$.

2. a) Un étude de fonction élémentaire montre que, pour tout $t \in [0, 1]$, $1 - t \leq e^{-t}$. Si $t \in [0, n]$, $\frac{t}{n} \in [0, 1]$ et donc $1 - \frac{t}{n} \leq e^{-\frac{t}{n}}$. En élevant à la puissance n , on obtient

$$\left(1 - \frac{t}{n}\right)^n \leq e^{-t}.$$

b) La relation à démontrer est évidente si $t \geq \sqrt{n}$ (le terme de gauche est négatif). Pour $t \in [0, \sqrt{n}]$, il faut démontrer

$$-t + \ln\left(1 - \frac{t^2}{n}\right) \leq n \ln\left(1 - \frac{t}{n}\right).$$

On étudie la fonction $f : t \mapsto n \ln\left(1 - \frac{t}{n}\right) + t - \ln\left(1 - \frac{t^2}{n}\right)$. Pour $t \in [0, \sqrt{n}]$, on a

$$f'(t) = \frac{-1}{1 - \frac{t}{n}} + 1 + \frac{\frac{2t}{n}}{1 - \frac{t^2}{n}} = \frac{t(n - 2t + t^2)}{(n-t)(n-t^2)} = \frac{t(n-1 + (t-1)^2)}{(n-t)(n-t^2)} \geq 0.$$

La fonction f est croissante et, comme $f(0) = 0$, positive, ce qui donne l'inégalité demandée.

c) Des deux questions précédentes résultent l'encadrement, pour $t \in [0, n]$,

$$\left(1 - \frac{t^2}{n}\right) e^{-t} t^{x-1} \leq \left(1 - \frac{t}{n}\right)^n t^{x-1} \leq e^{-t} t^{x-1},$$

puis en intégrant

$$\int_0^n \left(1 - \frac{t^2}{n}\right) e^{-t} t^{x-1} dt \leq \int_0^n \left(1 - \frac{t}{n}\right)^n t^{x-1} dt \leq \int_0^n e^{-t} t^{x-1} dt.$$

Par définition, on a $\lim_{n \rightarrow +\infty} \int_0^n e^{-t} t^{x-1} dt = \Gamma(x)$. En écrivant

$$\int_0^n \left(1 - \frac{t^2}{n}\right) e^{-t} t^{x-1} dt = \int_0^n e^{-t} t^{x-1} dt - \frac{1}{n} \int_0^n e^{-t} t^{x+1} dt,$$

on obtient $\lim_{n \rightarrow +\infty} \int_0^n \left(1 - \frac{t^2}{n}\right) e^{-t} t^{x-1} dt = \Gamma(x) - 0 \cdot \Gamma(x+1) = \Gamma(x)$. Par encadrement, on en déduit que

$$\lim_{n \rightarrow +\infty} \int_0^n \left(1 - \frac{t}{n}\right)^n t^{x-1} dt = \Gamma(x).$$

d) En faisant le changement de variable $u = \frac{t}{n}$, on obtient

$$\int_0^n \left(1 - \frac{t}{n}\right)^n t^{x-1} dt = n^x \int_0^1 (1-u)^n u^{x-1} du = n^x B(n+1, x)$$

et d'après la question 1,

$$\int_0^n \left(1 - \frac{t}{n}\right)^n t^{x-1} dt = \frac{n^x n!}{x(x+1)\dots(x+n)}.$$

On a donc

$$\Gamma(x) = \lim_{n \rightarrow +\infty} \frac{n^x n!}{x(x+1)\dots(x+n)}.$$

3. a) D'après la question 2, on a $\Gamma(y) = \lim_{n \rightarrow +\infty} n^y B(n+1, y)$ et donc $\Gamma(y) \sim n^y B(n+1, y)$ soit

$$B(n+1, y) \sim \frac{\Gamma(y)}{n^y}.$$

La fonction $x \rightarrow B(x, y)$ est une fonction décroissante, car pour tout $t \in]0, 1[$, $x \mapsto t^{x-1}$ décroît. En posant $n = \lfloor \rfloor x$, on obtient

$$B(n, y) \leq B(x, y) \leq B(n+1, y).$$

Quand x tend vers $+\infty$, on a $n \sim x$ et d'après ce qui précède,

$$B(n+1, y) \sim \frac{\Gamma(y)}{n^y} \sim \frac{\Gamma(y)}{x^y} \quad \text{et} \quad B(n, y) \sim \frac{\Gamma(y)}{(n-1)^y} \sim \frac{\Gamma(y)}{x^y}.$$

On en déduit que

$$B(x, y) \underset{x \rightarrow +\infty}{\sim} \frac{\Gamma(y)}{x^y}.$$

b) D'après la question 1, on a pour $n \geq 1$, $B(x+n, y) = \frac{x+n-1}{x+y+n-1} B(x+n-1, y)$. En réitérant cette relation, on obtient

$$B(x+n, y) = \frac{x(x+1)\dots(x+n-1)}{(x+y)(x+y-1)\dots(x+y+n-1)} B(x, y)$$

et donc

$$\begin{aligned} \frac{B(x+n, y)n^y}{B(x, y)} &= \frac{x(x+1)\dots(x+n-1)n^y}{(x+y)(x+y+1)\dots(x+y+n-1)} \\ &\underset{n \rightarrow +\infty}{\sim} \frac{x(x+1)\dots(x+n)n^{x+y}n!}{(x+y)(x+y+1)\dots(x+y+n)n^x n!}. \end{aligned}$$

En utilisant la relation démontrée dans la question 2.c, pour x et $x+y$, on en déduit que

$$\lim_{n \rightarrow +\infty} \frac{B(x+n, y)n^y}{B(x, y)} = \frac{\Gamma(x+y)}{\Gamma(x)}.$$

Mais par ailleurs, il résulte de a que

$$B(x+n, y) \underset{n \rightarrow +\infty}{\sim} \frac{\Gamma(y)}{(x+n)^y} \underset{n \rightarrow +\infty}{\sim} \frac{\Gamma(y)}{n^y},$$

ce qui montre que

$$\lim_{n \rightarrow +\infty} \frac{B(x+n, y)n^y}{B(x, y)} = \frac{\Gamma(y)}{B(x, y)}.$$

D'où l'égalité $\frac{\Gamma(x+y)}{\Gamma(x)} = \frac{\Gamma(y)}{B(x, y)}$ soit encore

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Chapitre 4

- 1** 1. Si $U \in H$ et $\mathcal{D} \cap \mathcal{H} \neq \emptyset$, on considère $A \in \mathcal{D} \cap \mathcal{H}$. Pour tout $M \in \mathcal{D}$, il existe $t \in \mathbb{R}$ tel que $M = A + tU$. Comme $tU \in H$, M appartient à \mathcal{H} et $\mathcal{D} \subset \mathcal{H}$.
2. Si $U \notin H$, on a $\text{Vect}(U) \oplus H = \mathbb{R}^n$. Soit $A \in \mathcal{D}$, $B \in \mathcal{H}$. Il existe $t \in \mathbb{R}$ et $h \in H$ tel que $B - A = tU + h$ et donc $A + tU = B - h$. Comme $A \in \mathcal{D}$, $A + tU \in \mathcal{D}$. De même, $B \in \mathcal{H}$ et $-h \in H$ donc $B - h \in \mathcal{H}$. On a donc $A + tU \in \mathcal{D} \cap \mathcal{H}$.
Si $\mathcal{D} \cap \mathcal{H}$ contient deux points M et N , $M - N$ appartient à $H \cap \text{Vect}(U) = \{0\}$, donc $M = N$: $\mathcal{D} \cap \mathcal{H}$ est un singleton.

- 2** 1. Le plan affine \mathcal{P} contient A, B et C si, et seulement si, $A \in \mathcal{P}$ et \overrightarrow{AB} et \overrightarrow{AC} appartiennent à la direction P de \mathcal{P} . Comme A, B, C ne sont pas alignés, \overrightarrow{AB} et \overrightarrow{AC} ne sont pas colinéaires et donc le seul plan affine \mathcal{P} contenant A, B et C est le plan affine passant par A et de direction $P = \text{Vect}(\overrightarrow{AB}, \overrightarrow{AC})$.
2. Soient \mathcal{P} et \mathcal{P}' deux plans affines de directions distinctes P et P' . Comme $P \neq P'$, $P + P'$ est de dimension strictement supérieure à celle de P , donc $P + P' = \mathbb{R}^3$. D'autre part, on a

$$\dim(P \cap P') = \dim P + \dim P' - \dim(P + P') = 2 + 2 - 3 = 1$$

donc $P \cap P'$ est une droite vectorielle $\text{Vect}(U)$.

Soit A un point de \mathcal{P} , A' un point de \mathcal{P}' . Il existe $(u, u') \in P \times P'$ tel que $A' - A = u + u'$. On a alors $A + u = A' - u'$. On note $B = A + u$. Le point B appartient à $\mathcal{P} \cap \mathcal{P}'$. Si $M \in \mathbb{R}^3$, M appartient à $\mathcal{P} \cap \mathcal{P}'$ si, et seulement si, $M - B \in P \cap P' = \text{Vect}(U)$. Ainsi $\mathcal{P} \cap \mathcal{P}'$ est la droite affine passant par B de vecteur directeur U .

3. Soit \mathcal{D} une droite affine passant par A de vecteur directeur U . On considère P et P' deux plans vectoriels distincts contenant U (on peut prendre $P = \text{Vect}(U, V)$ et $P' = \text{Vect}(U, V')$, où (U, V, V') est une base de \mathbb{R}^3), \mathcal{P} (resp. \mathcal{P}') le plan affine contenant A et de direction P (resp. P'). Comme $\mathcal{P} \cap \mathcal{P}'$ contient clairement \mathcal{D} , il résulte de la question 2, que $\mathcal{P} \cap \mathcal{P}' = \mathcal{D}$.
4. Comme tout plan affine possède une équation de la forme $ux + vy + wz + c = 0$, le résultat découle de la question 3.

Le point $M = (x, y, z)$ appartient à la droite \mathcal{D} contenant A et de vecteur directeur U si et seulement s'il existe $t \in \mathbb{R}$ tel que

$$\begin{cases} x = -1 + t \\ y = 1 + t \\ z = 2 - t. \end{cases}$$

Cela équivaut à

$$\begin{cases} y = 2 + x \\ z = 1 - x. \end{cases}$$

3 La direction P de \mathcal{P} a pour équation $ux + vy + wz = 0$.

1. Si $U \in P$, c'est-à-dire si $ua + vb + wc = 0$, soit $A \in \mathcal{P}$, i.e. $ux_0 + vy_0 + wz_0 + h = 0$ et $\mathcal{D} \subset \mathcal{P}$ soit $A \notin \mathcal{P}$ et $\mathcal{D} \cap \mathcal{P} = \emptyset$.

Supposons que $U \notin P$. Un point M appartient à \mathcal{D} s'il existe $t \in \mathbb{R}$ tel que

$$M = A + tU = (x_0 + ta, y_0 + tb, z_0 + tc).$$

Alors $M \in \mathcal{D} \cap \mathcal{P}$ si $u(x_0 + ta) + v(y_0 + tb) + w(z_0 + tc) + h = 0$, soit $t = -\frac{ux_0 + vy_0 + wz_0 + h}{ua + vb + wc}$.

Le point d'intersection de \mathcal{D} et \mathcal{P} est $A + tU$.

2. Les coordonnées (a, b, c) de U vérifie le système

$$\begin{cases} u_1a + v_1b + w_1c = 0 \\ u_2a + v_2b + w_2c = 0. \end{cases}$$

En combinant les deux équations, obtient

$$b(u_2v_1 - u_1v_2) + c(u_2w_1 - u_1w_2) = 0.$$

On peut prendre $b = w_1u_2 - w_2u_1$ et $c = u_1v_2 - u_2v_1$. On trouve alors $a = v_1w_2 - v_2w_1$ et donc

$$U = (v_1w_2 - v_2w_1, w_1u_2 - w_2u_1, u_1v_2 - u_2v_1).$$

La condition $U \in P$ s'écrit

$$u(v_1w_2 - v_2w_1) + v(w_1u_2 - w_2u_1) + w(u_1v_2 - u_2v_1) = 0.$$

Si $U \notin P$, l'intersection de \mathcal{D} et \mathcal{P} s'obtient en résolvant le système

$$\begin{cases} ux + vy + wz + h = 0 \\ u_1x + v_1y + w_1z + h_1 = 0 \\ u_2x + v_2y + w_2z + h_2 = 0. \end{cases}$$

On trouve

$$\begin{cases} x = \frac{(v_2w_1 - v_1w_2)h + (v_1w_2 - w_2v_1)h_1 + (w_1v_2 - v_2w_1)h_2}{u(v_1w_2 - v_2w_1) + v(w_1u_2 - w_2u_1) + w(u_1v_2 - u_2v_1)} \\ y = \frac{(u_2w_1 - w_1u_2)h + (u_1v_2 - v_2u_1)h_1 + (v_1w_2 - w_2v_1)h_2}{u(v_1w_2 - v_2w_1) + v(w_1u_2 - w_2u_1) + w(u_1v_2 - u_2v_1)} \\ z = \frac{(u_2v_1 - u_1v_2)h + (u_1v_2 - v_2u_1)h_1 + (v_1w_2 - w_2v_1)h_2}{u(v_1w_2 - v_2w_1) + v(w_1u_2 - w_2u_1) + w(u_1v_2 - u_2v_1)}. \end{cases}$$

4 On a pour U et V dans \mathbb{R}^n ,

$$\|U+V\|^2 + \|U-V\|^2 = \|U\|^2 + \|V\|^2 + 2\langle U, V \rangle + \|U\|^2 + \|V\|^2 - 2\langle U, V \rangle = 2(\|U\|^2 + \|V\|^2).$$

5 1. On a

$$\begin{aligned} \sum_{1 \leq i < j \leq p} \|X_i - X_j\|^2 &= \sum_{1 \leq i < j \leq p} (\|X_i\|^2 + \|X_j\|^2 - 2\langle X_i, X_j \rangle) \\ &= \sum_{i=1}^p (p-i)\|X_i\|^2 + \sum_{j=1}^p (j-1)\|X_j\|^2 - 2 \sum_{1 \leq i < j \leq p} \langle X_i, X_j \rangle \\ &= (p-1) \sum_{i=1}^p \|X_i\|^2 - 2 \sum_{1 \leq i < j \leq p} \langle X_i, X_j \rangle \\ &= p \sum_{i=1}^p \|X_i\|^2 - \left(\sum_{i=1}^p \|X_i\|^2 + 2 \sum_{1 \leq i < j \leq p} \langle X_i, X_j \rangle \right) \\ &= p \sum_{i=1}^p \|X_i\|^2 - \left\| \sum_{i=1}^p X_i \right\|^2. \end{aligned}$$

2. Notons X_0 le centre de la boule et posons, pour tout $i \in \llbracket 1, p \rrbracket$, $Y_i = X_i - X_0$. Les Y_i vérifient, pour $i \neq j$,

$$\|Y_i - Y_j\| = \|X_i - X_j\| \geq 2$$

et pour tout i , $\|Y_i\| \leq r$.

En appliquant l'inégalité démontrée dans la question 1 aux Y_i , on obtient

$$\sum_{1 \leq i < j \leq p} \|Y_i - Y_j\|^2 \leq p \sum_{i=1}^p \|Y_i\|^2 \leq p^2 r^2.$$

Par ailleurs, on a

$$\sum_{1 \leq i < j \leq p} \|Y_i - Y_j\|^2 \geq \sum_{1 \leq i < j \leq p} 4 \geq 4 \frac{p(p-1)}{2}.$$

On en déduit $p^2 r^2 \geq 2p(p-1)$ et

$$r \geq \sqrt{\frac{2(p-1)}{p}}.$$

6 1. L'ensemble $\Omega_1 = \{(x, y, z) \in \mathbb{R}^3, 1 < \|(x, y, z)\| < 2\}$ est égal à la différence $B(0, 2) \setminus B_f(0, 1)$. C'est l'intersection de deux ouverts : la boule ouverte $B(0, 2)$ et le complémentaire de la boule fermée $B_f(0, 1)$, donc c'est un ouvert. L'ensemble Ω_1 est borné car inclus dans $B(0, 2)$.

2. L'ensemble $\Omega_2 = \{(x, y, z) \in \mathbb{R}^3, x + y + z < 1, x > 0, y > 0, z > 0\}$ est un ouvert car c'est l'intersection de quatre demi-espaces ouverts.

Pour tout $(x, y, z) \in \Omega_2$, on a $0 < x < x + y + z < 1$ et de même, $0 < y < 1$ et $0 < z < 1$. On en déduit que $\|(x, y, z)\| \leq \sqrt{3}$ et Ω_2 est borné.

3. Soit $\Omega_3 = \{(x, y, z) \in \mathbb{R}^3, x^2 + y^2 - z^2, z \in]0, 1[\}$, $X_0 = (x_0, y_0, z_0) \in \Omega_3$ et $\rho > 0$ tel que $\rho < \min(z_0, 1 - z_0)$. Si $X = (x, y, z) \in B(X_0, \rho)$, on a $|x - x_0| \leq \|X - X_0\| < \rho$ et de même, $|y - y_0| < \rho$ et $|z - z_0| < \rho$. On en déduit que $z > z_0 - \rho > 0$ et $z < z_0 + \rho < 1$, donc $z \in]0, 1[$ et, d'autre part,

$$x^2 + y^2 - z^2 \leq (|x_0| + \rho)^2 + (|y_0| + \rho)^2 - (|z_0| - \rho)^2.$$

Comme $\lim_{\rho \rightarrow 0} (|x_0| + \rho)^2 + (|y_0| + \rho)^2 - (|z_0| - \rho)^2 = x_0^2 + y_0^2 - z_0^2 < 0$, on peut choisir ρ assez petit pour avoir $(|x_0| + \rho)^2 + (|y_0| + \rho)^2 - (|z_0| - \rho)^2 < 0$. On a alors $B(X_0, \rho) \subset \Omega_3$. Donc Ω_3 est ouvert.

Pour $(x, y, z) \in \Omega_3$, on a

$$\|(x, y, z)\|^2 = x^2 + y^2 + z^2 \leq 2z^2 \leq 2,$$

donc Ω_3 est borné.

- 7 Si ce sous-espace est \mathbb{R}^n ou $\{0\}$, c'est du cours (dans le deuxième cas, c'est un singleton).

Par ailleurs, il a été démontré dans le cours que tout hyperplan de \mathbb{R}^n est fermé. Tout sous-espace vectoriel F de \mathbb{R}^n de dimension $p \in \llbracket 1, n-1 \rrbracket$ est une intersection d'un nombre fini d'hyperplans. Pour le démontrer, on part d'une base (e_1, \dots, e_p) de F , qu'on complète en une base (e_1, \dots, e_n) de \mathbb{R}^n . Pour $p+1 \leq i \leq n$, on note H_i l'hyperplan engendré par les vecteurs de base différents de e_i . On a alors

$$H_{p+1} \cap \dots \cap H_n = \text{Vect}(e_1, \dots, e_p) = F.$$

Ainsi, F est une intersection d'un nombre fini de fermés donc est fermé.

- 8 Soit F une partie de \mathbb{R}^2 .

• Supposons que F est fermé et considérons une suite $(A_p)_{p \in \mathbb{N}}$ éléments de F qui converge vers A . Raisonnons par l'absurde et supposons que $A \notin F$. Comme le complémentaire de F est ouvert, on peut trouver $r > 0$ tel que la boule ouverte de centre A et de rayon r soit incluse dans le complémentaire de F . On a alors, pour tout $p \in \mathbb{N}$, $d(A, A_p) \geq r$ puisque $A_p \in F$. Cela est contradictoire avec la convergence de $(A_p)_{p \in \mathbb{N}}$ vers A . Donc A appartient à F .

• Supposons réciproquement que toute suite convergente d'éléments de F ait sa limite dans F . Montrons que F est fermé en raisonnant encore par l'absurde. Si F n'est pas fermé, son complémentaire n'est pas ouvert et on peut trouver $A \notin F$ tel que toute boule ouverte de centre A rencontre F . En particulier, pour tout $p \in \mathbb{N}^*$, il existe $A_p \in F$ tel que $d(A_p, A) < \frac{1}{p}$.

Par construction, $(A_p)_{p \in \mathbb{N}}$ est une suite d'éléments de F qui converge vers A , qui n'appartient pas à F . C'est contradictoire avec l'hypothèse. Ainsi, F est fermé.

- 9 1. Soit M un point qui n'appartient pas à $\overline{\Omega}$. Par définition, il existe $r > 0$ tel que la boule $B(M, r)$ ne rencontre pas Ω . Soit $N \in B(M, r)$. Puisque cette boule ouverte est un ouvert, on peut trouver $\rho > 0$ tel que la boule ouverte $B(N, \rho)$ soit incluse dans $B(M, r)$. La boule $B(N, \rho)$ ne rencontre pas non plus Ω . Par définition, le point N n'appartient donc pas à $\overline{\Omega}$. Ainsi, la boule ouverte $B(M, r)$ est incluse dans le complémentaire de $\overline{\Omega}$. Ce complémentaire est donc ouvert et $\overline{\Omega}$ est fermé.

Soit F un fermé tel que $\Omega \subset F$. Soit M tel que $M \notin F$. Comme le complémentaire de F est un ouvert, on peut trouver $r > 0$ tel que $B(M, r)$ soit incluse dans le complémentaire de F . La boule $B(M, r)$ ne rencontre pas F et a fortiori ne rencontre pas Ω . On en déduit que $M \notin \overline{\Omega}$. On a montré que si M n'est pas dans F , il n'est pas dans $\overline{\Omega}$. Cela montre que $\overline{\Omega} \subset F$.

2. a) On sait que $\overline{\Omega}$ contient $\Omega = B(A, r)$.

Si M appartient à la sphère de centre A et de rayon r , toute boule ouverte de centre M rencontre Ω . En effet, pour tout $\varepsilon \in]0, r[$, le point N défini par $N = M - \frac{\varepsilon}{r} \overrightarrow{AM}$ appartient à Ω et est à une distance ε de M . Donc M appartient à $\overline{\Omega}$.

Ainsi $\overline{\Omega}$ contient $B_f(A, r)$. Mais par ailleurs cette boule fermée est un fermé qui contient Ω . D'après la question 1, elle contient $\overline{\Omega}$, donc $\overline{\Omega} = B_f(A, r)$.

b) On montre comme dans le cas a que $\overline{\Omega}$ contient la sphère de centre A et de rayon r , donc contient $\{M \in \mathbb{R}^n, d(M, A) \geq r\}$. Mais cet ensemble qui est le complémentaire de $B(A, r)$ est fermé. Comme dans a, on conclut qu'il est égal à $\overline{\Omega}$.

c) Supposons que $M = (x_1, \dots, x_n)$ vérifie $u_1x_1 + \dots + u_nx_n + c = 0$.

Pour $h \in \mathbb{R}$, considérons $N = M + h(u_1, \dots, u_n) = (x_1 + hu_1, \dots, x_n + hu_n)$. On a

$$\sum_{i=1}^n u_i(x_i + hu_i) + c = h \sum_{i=1}^n u_i^2.$$

Si on prend $h > 0$, le point N appartient à Ω . Comme h est un réel quelconque strictement positif et que $d(M, N) = h\sqrt{\sum_{i=1}^n u_i^2}$ on peut trouver dans toute boule ouverte de centre M de tels points de Ω . On en déduit que M appartient à $\overline{\Omega}$ et donc que $\overline{\Omega}$ contient

$$\{(x_1, \dots, x_n) \in \mathbb{R}^n, u_1x_1 + \dots + u_nx_n + c \geq 0\}.$$

Comme cet ensemble est fermé et contient Ω , il contient aussi $\overline{\Omega}$ et finalement, il est égal à $\overline{\Omega}$.

3. Soit $M \in \overset{\circ}{\Omega}$. Par définition, il existe $r > 0$ tel que $B(M, r)$ soit inclus dans Ω . Soit $N \in B(M, r)$. Cette boule étant un ouvert, il existe $\rho > 0$ tel que $B(N, \rho)$ soit contenue dans $B(M, r)$ et a fortiori dans Ω . Ainsi $B(M, r)$ est inclus dans $\overset{\circ}{\Omega}$ et $\overset{\circ}{\Omega}$ est ouvert.

Si U est un ouvert inclus dans Ω , pour tout M de U , il existe $r > 0$ tel que $B(M, r)$ soit inclus dans U , et a fortiori dans Ω . Ainsi, M appartient à $\overset{\circ}{\Omega}$ et $U \subset \overset{\circ}{\Omega}$.

4. a) La boule ouverte $B(A, r)$ est incluse dans Ω , donc d'après la question 3, elle est incluse dans $\overset{\circ}{\Omega}$.

Si M appartient à la sphère de centre A et de rayon r , toute boule de centre M rencontre le complémentaire de Ω (le raisonnement est le même que dans 2.b), donc M n'est pas dans $\overset{\circ}{\Omega}$. Finalement $\overset{\circ}{\Omega} = B(A, r)$.

b) L'ensemble $\{(x_1, \dots, x_n) \in \mathbb{R}^n, u_1x_1 + \dots + u_nx_n + c > 0\}$ est un ouvert contenu dans Ω donc contenu dans $\overset{\circ}{\Omega}$. D'autre part, si le point $M = (x_1, \dots, x_n)$ vérifie $u_1x_1 + \dots + u_nx_n + c = 0$, toute boule de centre M rencontre le complémentaire de Ω (le raisonnement est le même que dans la question 2.c), donc M n'appartient pas

à $\mathring{\Omega}$. On en déduit que

$$\mathring{\Omega} = \{(x_1, \dots, x_n) \in \mathbb{R}^n, u_1 x_1 + \dots + u_n x_n + c > 0\}.$$

10 1. Si $M \notin \text{Fr}(\Omega)$, il existe une boule ouverte $B(M, r)$ incluse dans Ω ou dans le complémentaire de Ω . Comme cette boule ouverte est un ouvert, pour tout point N de $B(M, r)$, il existe une boule ouverte de centre N qui est incluse dans $B(M, r)$ et *a fortiori* dans Ω ou son complémentaire. Ainsi N n'appartient pas à $\text{Fr}(\Omega)$. La boule ouverte $B(M, r)$ est incluse dans le complémentaire $\text{Fr}(\Omega)$ qui est donc ouvert. Ainsi $\text{FR}(\Omega)$ est fermé.

2. a) Comme Ω est ouvert, si M appartient à Ω , il existe une boule ouverte $B(M, r)$ de centre M qui est incluse dans Ω . Elle ne rencontre pas le complémentaire de Ω , donc M n'appartient pas à $\text{Fr}(\Omega)$.

Comme l'ensemble $\Omega' = \{M \in \mathbb{R}^n, d(M, A) > r\}$ est également ouvert, on montre de même qu'un élément de Ω' n'appartient pas à $\text{Fr}(\Omega)$.

Enfin si $d(A, M) = r$, toute boule ouverte de centre M rencontre Ω et son complémentaire. Il suffit de considérer des points de la forme $M - \varepsilon \overrightarrow{AM}$ et $M + \varepsilon \overrightarrow{AM}$, avec ε assez petit.

On conclut que $\text{Fr}(\Omega)$ est la sphère de centre A de rayon r :

$$\text{Fr}(\Omega) = \{M \in \mathbb{R}^2, d(M, A) = r\}.$$

b) On montre comme dans l'exemple précédent, qu'un élément de

$$U = \{M \in \mathbb{R}^2, \|M\| < 1\} \text{ ou de } U' = \{M \in \mathbb{R}^2, \|M\| > 1\}$$

n'appartient pas à $\text{Fr}(\Omega)$.

Si $M \in \Omega$, toute boule ouverte de centre M contient M et des points du complémentaire de Ω . On conclut : $\text{Fr}(\Omega) = \Omega$.

c) On montre que $\Omega_1 = \{(x, y, z) \in \mathbb{R}^3, -1 < x < 1\}$ est ouvert. En effet si $(x, y, z) \in \Omega_1$ et $r = \min(1 - x, x + 1)$, la boule ouverte de centre (x, y, z) et de rayon r est incluse dans Ω_1 . Cette boule ne rencontre pas le complémentaire de Ω donc (x, y, z) n'appartient pas à $\text{Fr}(\Omega)$.

On montre de même que les ensembles

$$\Omega_2 = \{(x, y, z) \in \mathbb{R}^3, x < -1\}$$

$$\text{et } \Omega_3 = \{(x, y, z) \in \mathbb{R}^3, x > 1\}$$

sont ouverts. Ainsi, si (x, y, z) appartient à l'un de ces ensembles, il existe une boule ouverte de centre (x, y, z) qui ne rencontre pas Ω et (x, y, z) n'appartient pas à $\text{Fr}(\Omega)$.

Enfin si $x = -1$ et $(y, z) \in \mathbb{R}^2$, toute boule ouverte de centre $(-1, y, z)$ contient des points (u, v, w) tels que $u < -1$ et des points tels que $-1 < u < 1$. Il suffit de considérer les points $(-1 - \varepsilon, y, z)$ et $(-1 + \varepsilon, y, z)$ pour $\varepsilon > 0$ assez petit. Le point $(-1, y, z)$ appartient donc à $\text{Fr}(\Omega)$. Il en est de même de tout point $(1, y, z)$.

Conclusion : $\text{Fr}(\Omega) = \{(x, y, z) \in \mathbb{R}^3, x = -1 \text{ ou } x = 1\}$.

3. • Si la frontière de Ω est incluse dans Ω et si $M \notin \Omega$, M n'appartient pas à $\text{Fr}(\Omega)$. Il existe une boule ouverte de centre M dont l'intersection avec Ω ou son complémentaire est vide. Comme l'intersection avec le complémentaire de Ω contient M , la boule ne rencontre pas Ω . Elle est incluse dans son complémentaire, ce qui montre que celui-ci est ouvert et Ω fermé.

- Réciproquement, si Ω est fermé et si $M \notin \Omega$, il existe une boule ouverte de centre M qui est incluse dans le complémentaire de Ω . Elle ne rencontre pas Ω et $M \notin \text{Fr}(\Omega)$. On a donc $\text{Fr}(\Omega) \subset \Omega$.

11 1. Soit $x \in \Omega_p$. Considérons un réel ρ strictement positif tel que $\rho < \frac{1}{p} - d(x, \Omega)$.

On a alors $d(x, \Omega) < \frac{1}{p} - \rho$ et par définition de la borne inférieure, il existe $a \in \Omega$ tel que

$$d(x, a) < \frac{1}{p} - \rho.$$

Soit $y \in B(x, \rho)$. Par inégalité triangulaire, on obtient

$$d(y, a) \leq d(y, x) + d(x, a) < \rho + \frac{1}{p} - \rho \leq \frac{1}{p}.$$

On a *a fortiori* $d(y, \Omega) < \frac{1}{p}$ et $y \in \Omega_p$. La boule $B(x, \rho)$ est incluse dans Ω_p et Ω_p est un ouvert.

On a clairement, pour $x \in \mathbb{R}^n$, $x \in \bigcap_{p \in \mathbb{N}^*} \Omega_p$ si, et seulement, si $d(x, \Omega) = 0$.

Par définition de la borne inférieure, on a $d(x, \Omega) = 0$ si et seulement si, pour tout $\varepsilon > 0$, il existe $a \in \Omega$ tel que $d(x, a) < \varepsilon$, c'est-à-dire si et seulement si tout disque de centre a rencontre Ω , c'est-à-dire si et seulement si $x \in \overline{\Omega}$. On a donc

$$\overline{\Omega} = \bigcap_{p \in \mathbb{N}^*} \Omega_p.$$

2. D'après l'exercice 9, $\overline{\Omega}$ est le plus petit fermé contenant Ω . Si Ω est fermé, on a donc

$$\Omega = \overline{\Omega} = \bigcap_{p \in \mathbb{N}^*} \Omega_p,$$

ce qui montre que Ω est intersection d'un nombre dénombrables d'ouverts.

3. Soit Ω un ouvert, $\Omega' = \mathbb{R}^n \setminus \Omega$. L'ensemble Ω' est fermé donc, d'après la question 2, on a

$$\Omega' = \bigcap_{p \in \mathbb{N}^*} \Omega'_p \text{ et}$$

$$\Omega = \mathbb{R}^n \setminus \Omega' = \bigcup_{p \in \mathbb{N}^*} (\mathbb{R}^n \setminus \Omega'_p),$$

où les ensembles $\mathbb{R}^n \setminus \Omega'_p$, complémentaires d'ouverts, sont fermés.

4. • Supposons que $\Omega = B(x_0, r)$. Soit $x \in \Omega_p$. On a $d(x, \Omega) < \frac{1}{p}$. Il existe donc $a \in \Omega$ tel

que $d(x, a) < \frac{1}{p}$. On a alors, par inégalité triangulaire,

$$d(x, x_0) \leq d(x, a) + d(a, x_0) < \frac{1}{p} + r$$

et x appartient à $B(x_0, \frac{1}{p} + r)$.

Réciproquement, soit $x \in B(x_0, \frac{1}{p} + r)$, $a = x_0 + \frac{r}{\frac{1}{p} + r}(x - x_0)$. On a alors

$$d(x_0, a) = \|a - x_0\| = \frac{r}{\frac{1}{p} + r} \|x - x_0\| < r,$$

donc $a \in B(x_0, r) = \Omega$ et

$$d(x, a) = \|x - a\| = \frac{\frac{1}{p}}{\frac{1}{p} + r} \|x - x_0\| < \frac{1}{p}.$$

Comme $a \in \Omega$, on a *a fortiori* $d(x, \Omega) < \frac{1}{p}$ et $x \in \Omega_p$.

On a donc

$$\Omega_p = B(x_0, \frac{1}{p} + r).$$

• Supposons que Ω est la droite passant par a et de vecteur directeur u . Un point quelconque de Ω s'écrit $a + \lambda u$. Pour tout point $x \in \mathbb{R}^n$, on peut écrire $x - a = \alpha u + y$, où $y \in (\text{Vect}(u))^\perp$. On a alors, pour tout réel λ ,

$$\|x - (a + \lambda u)\|^2 = \|(\alpha - \lambda)u + y\|^2 = (\alpha - \lambda)^2 \|u\|^2 + \|y\|^2.$$

La distance de x au point $a + \lambda u$ est minimale si $\lambda = \alpha$. On a donc $d(x, \Omega) = \|y\|$ et $x \in \Omega_p$ si $\|y\| < \frac{1}{p}$. Ainsi les éléments de Ω_p sont ceux qui s'écrivent $b + y$, avec $b \in \Omega$ et y orthogonal à u de norme inférieure à $\frac{1}{p}$. L'ensemble Ω_p est une réunion de droites parallèles à Ω . L'intersection avec un hyperplan normal à u est une boule ouverte (de l'hyperplan), de rayon $\frac{1}{p}$. On obtient un cylindre.

12 1. On raisonne par récurrence sur p .

La propriété est évidente pour $p = 1$, car alors $\lambda_1 = 1$.

Supposons que la propriété est vérifiée au rang p et considérons $p + 1$ points A_1, \dots, A_{p+1}

de Ω et $p + 1$ réels positifs $\lambda_1, \dots, \lambda_{p+1}$ tels que $\sum_{i=1}^{p+1} \lambda_i = 1$.

Si $\lambda_{p+1} = 1$, alors $\sum_{i=1}^{p+1} \lambda_i A_i = A_{p+1}$ et il n'y a rien à démontrer. Sinon, on écrit

$$\sum_{i=1}^{p+1} \lambda_i A_i = (1 - \lambda_{p+1}) \left(\sum_{i=1}^p \frac{\lambda_i}{1 - \lambda_{p+1}} A_i \right) + \lambda_{p+1} A_{p+1}.$$

Comme $\sum_{i=1}^p \frac{\lambda_i}{1 - \lambda_{p+1}} = 1$, l'hypothèse de récurrence donne que $B = \sum_{i=1}^n \frac{\lambda_i}{1 - \lambda_{p+1}} A_i$ appartient à Ω . Le point

$$\sum_{i=1}^{p+1} \lambda_i A_i = (1 - \lambda_{p+1})B + \lambda_{p+1} A_{p+1}$$

appartient aussi à Ω par définition d'un ensemble convexe, ce qui termine la démonstration par récurrence.

2. a) Si $M \in \Omega$, alors $M = 1 \cdot M$ donc $M \in C : \Omega$ est inclus dans C .

Soient A et B deux points de C . Il existe des entiers naturels non nuls p et q , des réels positifs $\lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q$ et des éléments $A_1, \dots, A_p, B_1, \dots, B_q$ de Ω tels que

$$\sum_{i=1}^p \lambda_i = \sum_{j=1}^q \mu_j = 1, \quad A = \sum_{i=1}^p \lambda_i A_i \quad \text{et} \quad B = \sum_{j=1}^q \mu_j B_j.$$

Soit $M \in [A, B]$. Il existe $\alpha \in [0, 1]$ tel que $M = (1 - \alpha)A + \alpha B$. On obtient alors

$$M = \sum_{i=1}^p (1 - \alpha) \lambda_i A_i + \sum_{j=1}^q \alpha \mu_j B_j.$$

Comme $\sum_{i=1}^p (1 - \alpha) \lambda_i + \sum_{j=1}^q \alpha \mu_j = 1 - \alpha + \alpha = 1$, on en déduit que $M \in C$. Donc C est convexe.

b) Soit C' un convexe contenant Ω . Soit $A \in C$, des réels positifs $\lambda_1, \dots, \lambda_p$, des éléments

A_1, \dots, A_p de Ω tels que $\sum_{i=1}^p \lambda_i = 1$ et $A = \sum_{i=1}^p \lambda_i A_i$. Comme $\Omega \subset C'$, on a, pour

tout i , $A_i \in C'$ et donc d'après la question 1, A appartient à C' . On a donc $C \subset C' : C$ est donc le plus petit convexe contenant Ω .

13 1. Soient x et y deux éléments de \mathbb{R}^n , $\lambda \in [0, 1]$.

On a

$$\begin{aligned} \|\lambda x + (1 - \lambda)y\|^2 - (\lambda\|x\|^2 + (1 - \lambda)\|y\|^2) &= \lambda^2\|x\|^2 + (1 - \lambda)^2\|y\|^2 \\ &\quad + 2\lambda(1 - \lambda)\langle x, y \rangle - \lambda\|x\|^2 - (1 - \lambda)\|y\|^2 \\ &= \lambda(\lambda - 1)(\|x\|^2 + \|y\|^2 - 2\langle x, y \rangle) \\ &= \lambda(1 - \lambda)\|x - y\|^2 \leq 0 \end{aligned}$$

car $\lambda(\lambda - 1) \leq 0$. Donc la fonction $x \mapsto \|x\|^2$ est convexe.

En utilisant la croissance et la convexité de la fonction \exp , on en déduit

$$e^{\|\lambda x + (1 - \lambda)y\|^2} \leq e^{\lambda\|x\|^2 + (1 - \lambda)\|y\|^2} \leq \lambda e^{\|x\|^2} + (1 - \lambda)e^{\|y\|^2}.$$

Donc la fonction $x \mapsto e^{\|x\|^2}$ est convexe.

2. Pour $(u, v, u', v') \in \mathbb{R}^4$ et $\lambda \in [0, 1]$, on a

$$\begin{aligned} f(\lambda(u, v) + (1 - \lambda)(u', v')) &= f(\lambda u + (1 - \lambda)u', \lambda v + (1 - \lambda)v') \\ &= (\lambda u + (1 - \lambda)u')^2 + (\lambda v + (1 - \lambda)v')^2 \\ &\quad - (\lambda u + (1 - \lambda)u')(\lambda v + (1 - \lambda)v') \\ &= \lambda^2 f(u, v) + (1 - \lambda)^2 f(u', v') \\ &\quad + \lambda(1 - \lambda)(2uu' + 2vv' - uv' - u'v). \end{aligned}$$

On en déduit

$$\begin{aligned} & f(\lambda(u, v) + (1 - \lambda)(u', v')) - \lambda f(u, v) - (1 - \lambda)f(u', v') \\ &= (\lambda^2 - \lambda)(f(u, v) + f(u', v') - 2uu' - 2vv' + uu' + u'v) \\ &= (\lambda^2 - \lambda)((u - u')^2 + (v - v')^2 - uv - u'v' + uu' + u'v) \\ &= (\lambda^2 - \lambda)((u - u')^2 + (v - v')^2 - (u - u')(v - v')) \\ &= -\lambda(1 - \lambda)f(u - u', v - v'). \end{aligned}$$

On a, pour tout $(u, v) \in \mathbb{R}^2$, $f(u, v) = \left(u + \frac{1}{2}v\right)^2 + \frac{3}{4}v^2 \geq 0$. On en déduit que $f(\lambda(u, v) + (1 - \lambda)(u', v')) - \lambda f(u, v) - (1 - \lambda)f(u', v') \leq 0$. La fonction f est donc convexe.

3. • Supposons que f est convexe et fixons $(x, y) \in \Omega^2$. Soient t, t' et λ trois éléments de $[0, 1]$. On a

$$g(\lambda t + (1 - \lambda)t') = f[(\lambda t + (1 - \lambda)t')x + (1 - (\lambda t + (1 - \lambda)t'))y].$$

En remarquant que $1 - (\lambda t + (1 - \lambda)t') = \lambda(1 - t) + (1 - \lambda)(1 - t')$, on obtient

$$g(\lambda t + (1 - \lambda)t') = f[\lambda(tx + (1 - t)y) + (1 - \lambda)(t'x + (1 - t')y)]$$

et donc, en utilisant la convexité de f ,

$$\begin{aligned} g(\lambda t + (1 - \lambda)t') &\leq \lambda f(tx + (1 - t)y) + (1 - \lambda)f(t'x + (1 - t')y) \\ &\leq \lambda g(t) + (1 - \lambda)g(t'), \end{aligned}$$

ce qui montre que g est convexe sur $[0, 1]$.

- Supposons réciproquement que g est convexe sur $[0, 1]$ pour tout $(x, y) \in \Omega^2$. Soit $(x, y) \in \Omega^2$ et $t \in [0, 1]$. De la convexité de g , on déduit

$$g(t) = g(t \cdot 1 + (1 - t) \cdot 0) \leq tg(1) + (1 - t)g(0),$$

c'est-à-dire

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y).$$

Comme cela est vérifié pour tout $(x, y) \in \Omega^2$ et tout $t \in [0, 1]$, la fonction f est convexe.

4. • Supposons que f est convexe. Soit (x, t) et (y, u) dans C , $\lambda \in [0, 1]$. De la convexité de f , on déduit

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \leq \lambda t + (1 - \lambda)u$$

donc $(\lambda x + (1 - \lambda)y, \lambda t + (1 - \lambda)u) = \lambda(x, t) + (1 - \lambda)(y, u)$ appartient à C , ce qui montre que C est convexe.

- Supposons réciproquement que C est convexe et considérons $(x, y) \in \Omega^2$, $\lambda \in [0, 1]$. Les deux couples $(x, f(x))$ et $(y, f(y))$ appartiennent à C . Comme C est convexe, on en déduit que

$$\lambda(x, f(x)) + (1 - \lambda)(y, f(y)) = (\lambda x + (1 - \lambda)y, \lambda f(x) + (1 - \lambda)f(y))$$

appartient à C . Cela signifie que

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Donc f est convexe.

Chapitre 5

- 1** 1. La ligne de niveau λ est la droite affine d'équation $3x + 2y - \lambda = 0$. On obtient toutes les droites de vecteur directeur $(-2, 3)$.
2. Si $\lambda < 0$, la ligne de niveau λ est \emptyset . Si $\lambda = 0$, on obtient $\{(0, 0)\}$.
Si $\lambda > 0$, la ligne de niveau λ est la réunion de quatre segments : le segment défini par $x \geq 0, y \geq 0$ et $3x + 4y = \lambda$, qui est d'extrémités $\left(\frac{\lambda}{3}, 0\right)$ et $\left(0, \frac{\lambda}{4}\right)$, et les segments qui s'en déduisent en faisant des symétries par rapport aux droite $x = 0$ et $y = 0$. Ces segments forment un carré.
3. Pour $\lambda < 0$, la ligne de niveau λ est l'ensemble vide. Pour $\lambda = 0$, c'est $\{(0, 0)\}$.
Pour $\lambda > 0$, on obtient une ellipse d'équation $9x^2 + 4y^2 = \lambda$, courbe symétrique par rapport aux axes de coordonnées et bornée. On l'étudie en la mettant sous la forme $y = \pm \frac{1}{2} \sqrt{\lambda - 9x^2}$.
4. Pour $\lambda = 0$, la ligne de niveau λ est la réunion des droites d'équation $y = \pm \frac{3x}{2}$.
Si $\lambda \neq 0$, on obtient une hyperbole. Pour la tracer, on la met sous la forme $y = \pm \frac{1}{2} \sqrt{9x^2 - \lambda}$. La courbe possède deux asymptotes d'équation $y = \pm \frac{3x}{2}$.

- 2** 1. a) On a, pour tout réel x , $\varphi'(x) = (1 - x)e^{-x}$. La fonction φ croît sur $] - \infty, 1]$ et décroît sur $[1, +\infty[$. On a, d'autre part $\lim_{x \rightarrow -\infty} \varphi(x) = -\infty$ et $\lim_{x \rightarrow +\infty} \varphi(x) = 0$, par croissances comparées. On en déduit le tableau de variation.

x	$-\infty$	1	$+\infty$
$\varphi(x)$	$-\infty$	$\nearrow \frac{1}{e}$	$\searrow 0$

On en déduit que

- si $\lambda \leq 0$, l'équation $\varphi(x) = \lambda$ possède une unique solution (négative) ;
 - si $0 < \lambda < \frac{1}{e}$, l'équation $\varphi(x) = \lambda$ possède deux solutions (l'une dans $]0, 1[$, l'autre dans $]1, +\infty[$) ;
 - si $\lambda = \frac{1}{e}$, l'équation $\varphi(x) = \lambda$ possède une unique solution 1 ;
 - si $\lambda > \frac{1}{e}$, l'équation $\varphi(x) = \lambda$ n'a pas de solution.
- b) La fonction ψ est continue et strictement croissante sur $] - \infty, 0[$. Comme $\lim_{x \rightarrow -\infty} \psi(x) = -\infty$ et $\lim_{x \rightarrow 0} \psi(x) = 0$, ψ réalise une bijection de $] - \infty, 0[$ sur $] - \infty, 0[$. Comme ψ est \mathcal{C}^1 et ψ' ne s'annule pas, ψ^{-1} est dérivable et

$$(\psi^{-1})' = \frac{1}{\psi' \circ \psi^{-1}}.$$

On voit que $(\psi^{-1})'$ est continue donc ψ^{-1} est \mathcal{C}^1 .

2. a) • Supposons Γ_λ^+ non vide et considérons $(x, y) \in \Gamma_\lambda^+$. On a $x > 0$ et, comme $\lambda > 0$, on a aussi $y > 0$. On en déduit que $\lambda = f(x, y) = \varphi(x)\varphi(y) \leq \left(\frac{1}{e}\right)^2 = \frac{1}{e^2}$, d'après le tableau de variation de φ .

Réciproquement, si $0 < \lambda \leq \frac{1}{e^2}$, on a $0 < \sqrt{\lambda} \leq \frac{1}{e}$. D'après 1.a, il existe $x > 0$ tel que $\varphi(x) = \sqrt{\lambda}$. On en déduit $f(x, x) = (\varphi(x))^2 = \lambda$ et $(x, x) \in \Gamma_\lambda^+$. Ainsi Γ_λ^+ n'est pas vide.

• Soit $\lambda \in \left]0, \frac{1}{e^2}\right]$. Si $(x, y) \in \Gamma_\lambda^+$, on a $\varphi(y) \leq \frac{1}{e}$ et donc $\varphi(x) = \frac{\lambda}{\varphi(y)} \geq e\lambda$.

Comme $e\lambda \in \left]0, \frac{1}{e}\right]$, l'équation $\varphi(x) = e\lambda$ possède deux solutions strictement positives a et b , confondues si $\lambda = \frac{1}{e^2}$ et d'après le tableau de variation de φ , $\varphi(x) \geq e\lambda$ implique $x \in [a, b]$. On démontre de la même façon que $y \in [a, b]$, car x et y jouent des rôles symétriques. Ainsi Γ_λ^+ est borné.

- b) Si $(x, y) \in \Gamma_\lambda^-$, on a $\lambda > 0$ et $x < 0$ donc $y < 0$. On a donc $(x, y) \in \Gamma_\lambda^-$ si, et seulement si, $(x, y) \in (\mathbb{R}_-^*)^2$ et $\psi(x)\psi(y) = \lambda$. Ceci équivaut à $\psi(y) = \frac{\lambda}{\psi(x)}$ et donc comme ψ est une bijection de $] -\infty, 0[$ sur $] -\infty, 0[$ à $y = \psi^{-1}\left(\frac{\lambda}{\psi(x)}\right)$.

Cela montre que Γ_λ^- est la courbe représentative de la fonction g définie sur $] -\infty, 0[$ par

$$g(x) = \psi^{-1}\left(\frac{\lambda}{\psi(x)}\right) = \psi^{-1}\left(\frac{\lambda}{\varphi(x)}\right).$$

D'après la question 1, g est de classe \mathcal{C}^1 .

Comme ψ est croissante et $\lambda > 0$, la fonction $x \mapsto \frac{\lambda}{\psi(x)}$ décroît et g décroît. On a $\lim_{x \rightarrow -\infty} g(x) = 0$ et $\lim_{x \rightarrow 0} g(x) = -\infty$.

3. a) On a,

$$(x, y) = X(1, 1) + Y(-1, 1) = (X - Y, X + Y).$$

On en déduit $x + y = 2X$,

$$f(x, y) = (X + Y)(X - Y)e^{-2X}$$

et Γ_λ a pour équation

$$(X^2 - Y^2)e^{-2X} = \lambda \text{ ou encore } Y^2 = X^2 - \lambda e^{2X}.$$

- b) On suppose $\lambda < 0$.

• Si $(x, y) \in \Gamma_\lambda^-$, on a $x < 0$ et donc $y > 0$. On en déduit $Y = \frac{y - x}{2} > 0$. On a alors

$$(x, y) \in \Gamma_\lambda^- \iff Y^2 = X^2 - \lambda e^{2X} \iff Y = \sqrt{X^2 - \lambda e^{2X}}.$$

La fonction $G_\lambda : X \mapsto \sqrt{X^2 - \lambda e^{2X}}$ est définie sur \mathbb{R} et de classe C^∞ car, pour tout réel X , $X^2 - \lambda e^{2X} > 0$. On a

$$G'_\lambda(X) = \frac{2X - 2\lambda e^{2X}}{2G_\lambda(X)} = \frac{e^{2X}(\varphi(2X) - 2\lambda)}{2G_\lambda(X)}.$$

On en déduit que $G'_\lambda(X) = 0$ si, et seulement si, $\varphi(2X) = 2\lambda$ soit

$$X = \frac{1}{2}\psi^{-1}(2\lambda) = X_0,$$

car alors $X < 0$. Du tableau de variation de φ , on déduit $G'_\lambda(X) < 0$ si $X < X_0$ et $G'_\lambda(X) > 0$ si $X > X_0$.

On a, quand X tend vers $-\infty$,

$$G_\lambda(X) = \sqrt{X^2 + o(1)} = -X + o(1),$$

donc la droite d'équation $Y = -X$ i.e. $y = 0$ est asymptote en $-\infty$.

Quand X tend vers $+\infty$, on obtient

$$G_\lambda(X) \sim \sqrt{-\lambda}e^X.$$

Il y a donc une direction asymptotique $X = 0$ i.e. $y = -x$.

• Si $(x, y) \in \Gamma_\lambda^+$, on a $x > 0$, $y < 0$ et donc $Y < 0$ et Γ_λ^+ a pour équation $Y = -G_\lambda(X)$, donc Γ_λ^- est symétrique de Γ_λ^+ par rapport à la droite $Y = 0$, i.e. $y = x$.

- c) On suppose $0 < \lambda \leq \frac{1}{e^2}$. On a alors, si $(x, y) \in \Gamma_\lambda^+$, $x > 0$ et $y > 0$, donc $X = \frac{x+y}{2} > 0$. La relation $F_\lambda(X, Y) = \lambda$ s'écrit

$$Y = \pm \sqrt{X^2 - \lambda e^{2X}} = \pm G_\lambda(X).$$

Pour $\lambda > 0$, la fonction G_λ n'est pas définie sur \mathbb{R}_+^* en entier. Il faut $X^2 \geq \lambda e^{2X}$, soit $X \geq \sqrt{\lambda}e^X$, soit encore

$$\varphi(X) \geq \sqrt{\lambda}.$$

Comme $\sqrt{\lambda} \leq \frac{1}{e}$, l'équation $\varphi(X) = \sqrt{\lambda}$ possède deux solutions α et β

($0 < \alpha \leq 1 \leq \beta$), confondues si $\lambda = \frac{1}{e^2}$. La fonction G_λ est définie sur $[\alpha, \beta]$.

On a, pour $X \in]\alpha, \beta[$,

$$G'_\lambda(X) = \frac{2X - 2\lambda e^{2X}}{2G_\lambda(X)} = \frac{e^{2X}(\varphi(2X) - 2\lambda)}{2G_\lambda(X)}.$$

On a $G'_\lambda(X) = 0$ si, et seulement si, $\varphi(2X) = 2\lambda$. D'après la question 1, cette équation possède au plus deux solutions, donc G'_λ s'annule au plus deux fois sur $]\alpha, \beta[$. Comme $G_\lambda(\alpha) = G_\lambda(\beta) = 0$, la fonction G_λ n'est pas strictement monotone sur $[\alpha, \beta]$. Mais G_λ étant positive, la dérivée ne peut pas changer deux fois de signe. Ainsi, G'_λ change de signe une fois en un point qui correspond au maximum de G_λ .

4. Pour $\lambda \neq 0$, Γ_λ est la réunion de deux ensembles disjoints Γ_λ^+ et Γ_λ^- .
- Si $\lambda < 0$, Γ_λ^+ et Γ_λ^- sont deux courbes symétriques par rapport à la droite d'équation $y = x$, possédant comme asymptote oblique la droite $y = 0$ et comme direction asymptotique $y = -x$. Ces deux courbes sont situées dans les quadrants $x > 0, y < 0$ et $x < 0, y > 0$ respectivement.
 - Si $0 < \lambda \leq \frac{1}{e^2}$, Γ_λ^+ est une courbe fermée bornée, symétrique par rapport à la droite d'équation $y = x$, incluse dans le quadrant $x > 0, y > 0$. Cette courbe est réduite à un point si $\lambda = \frac{1}{e^2}$.
 - Pour $\lambda > 0$, Γ_λ^- est une courbe asymptote aux droites d'équation $x = 0$ et $y = 0$, située dans le quadrant $x < 0, y < 0$.
 - Γ_0 est la réunion des droites d'équation $x = 0$ et $y = 0$.

3

1. On a, pour $t \neq 0$,

$$f(t^2, t) = \frac{t^6}{t^8 + t^6} = \frac{1}{t^2 + 1}.$$

On en déduit que $\lim_{t \rightarrow 0} f(t, t^2) = 1$. La fonction f n'est pas continue en 0, sinon on aurait $\lim_{t \rightarrow 0} f(t, t^2) = f(0, 0) = 0$.

2. En utilisant $|x| \leq \|(x, y)\|$ et $|y| \leq \|(x, y)\|$, on obtient, pour $(x, y) \neq (0, 0)$,

$$|f(x, y)| \leq \frac{|x^3| + |y^3|}{x^2 + y^2} \leq \frac{\|(x, y)\|x^2 + \|(x, y)\|y^2}{x^2 + y^2} \leq \|(x, y)\|.$$

On en déduit $\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 0$ et la continuité de f en $(0, 0)$.

3. On a, pour $(x, y) \neq (0, 0)$,

$$|f(x, y)| \leq \frac{|\sin(xy)|}{|x| + |y|} \leq \frac{|xy|}{|x| + |y|} \leq |x| \leq \|(x, y)\|,$$

car pour tout réel t , $|\sin t| \leq |t|$. On en déduit $\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 0$ et la continuité de f en $(0, 0)$.

4. On a, pour $(x, y) \neq (0, 0)$,

$$|f(x, y)| = |x||y| |\ln(x^2 + y^2)| \leq \|(x, y)\|^2 |\ln \|(x, y)\|^2|.$$

Comme $\lim_{t \rightarrow 0} t \ln t = 0$, on a $\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 0$ et f est continue en $(0, 0)$.

5. On a, pour $(x, y) \neq (0, 0)$,

$$f(x, y) = \frac{x(\sin y - y) + y(x - \sin x)}{x^2 + y^2}.$$

Par l'inégalité de Taylor-Lagrange, on obtient, pour tout réel t ,

$$|\sin t - t| \leq \frac{t^2}{2}.$$

On a donc

$$\begin{aligned} |f(x, y)| &\leq \frac{|x| |\sin y - y| + |y| |x - \sin x|}{x^2 + y^2} \leq \frac{|x|y^2 + |y|x^2}{2(x^2 + y^2)} \\ &\leq \frac{\|(x, y)\|^3}{x^2 + y^2} \leq \|(x, y)\|. \end{aligned}$$

On en déduit $\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 0$ et la continuité de f en $(0, 0)$.

- 4** 1. Sur l'ouvert $\Omega = \{(x, y) \in \mathbb{R}^2, y \neq 0\}$, on a $f(x, y) = ye^{\text{Arctan } \frac{x}{y}}$ et f est continue en tout point de Ω , comme composée de fonctions continues. Reste à étudier la continuité de f en un point $(x_0, 0)$. Si $y \neq 0$, on a

$$|f(x, y)| = |y|e^{\text{Arctan } \frac{x}{y}} \leq |y|e^{\frac{\pi}{2}} \leq e^{\frac{\pi}{2}} d((x, y), (x_0, 0)).$$

Cette inégalité reste vérifiée si $y = 0$. Pour tout couple (x, y) ,

$$|f(x, y) - f(x_0, 0)| \leq e^{\frac{\pi}{2}} d((x, y), (x_0, 0)).$$

On en déduit la continuité de f en $(x_0, 0)$ et donc sur \mathbb{R}^2 .

2. Sur l'ouvert $\Omega = \{(x, y) \in \mathbb{R}^2, y \neq 0\}$, on a $f(x, y) = \frac{\ln(1 + x^2 y^2)}{y^2}$ et f est continue en tout point de Ω . Reste à étudier la continuité en un point $(x_0, 0)$. Si $y \neq 0$, on a

$$|f(x, y) - f(x_0, 0)| = \left| \frac{\ln(1 + x^2 y^2)}{y^2} - x_0^2 \right| \leq \left| \frac{\ln(1 + x^2 y^2)}{y^2} - x^2 \right| + |x^2 - x_0^2|.$$

L'inégalité de Taylor-Lagrange donne, pour tout $t \geq 0$,

$$|\ln(1 + t) - t| \leq \frac{t^2}{2}.$$

On en déduit

$$\left| \frac{\ln(1 + x^2 y^2)}{y^2} - x^2 \right| = \frac{|\ln(1 + x^2 y^2) - x^2 y^2|}{y^2} \leq \frac{x^4 y^4}{2y^2} \leq \frac{x^4 y^2}{2}$$

puis

$$|f(x, y) - f(x_0, 0)| \leq \frac{x^4 y^2}{2} + |x^2 - x_0^2|,$$

inégalité qui reste vérifiée si $y = 0$. Comme $\lim_{(x,y) \rightarrow (x_0,0)} \frac{x^4 y^2}{2} + |x^2 - x_0^2| = 0$, la fonction f est continue en $(x_0, 0)$ pour tout $x_0 \in \mathbb{R}$, donc f est continue sur \mathbb{R}^2 .

3. La fonction f est continue sur l'ouvert $\Omega = \{(x, y) \in \mathbb{R}^2, xy \neq 0\}$. D'autre part, on vérifie que, pour tout $(x, y) \in \mathbb{R}^2$,

$$|f(x, y)| \leq |x| + |y| \leq 2\|(x, y)\|,$$

donc f est continue en $(0, 0)$.

Pour $x_0 \in \mathbb{R}^*$ et $k \in \mathbb{N}^*$, on a

$$f\left(x_0, \frac{1}{\frac{\pi}{2} + 2k\pi}\right) = \left(x_0 + \frac{1}{\frac{\pi}{2} + 2k\pi}\right) \sin \frac{1}{x_0} \sin\left(\frac{\pi}{2} + 2k\pi\right) = \left(x_0 + \frac{1}{\frac{\pi}{2} + 2k\pi}\right) \sin \frac{1}{x_0}$$

et

$$\lim_{k \rightarrow +\infty} \left(x_0 + \frac{1}{\frac{\pi}{2} + 2k\pi}\right) \sin \frac{1}{x_0} = x_0 \sin \frac{1}{x_0}.$$

Comme $\frac{1}{\frac{\pi}{2} + 2k\pi}$ tend vers 0, on en déduit n'est pas continue en $(x_0, 0)$ si $\sin \frac{1}{x_0} \neq 0$.

Supposons que $\sin \frac{1}{x_0} = 0$. Pour $xy \neq 0$, on obtient

$$|f(x, y) - f(x_0, 0)| = |f(x, y)| \leq |x + y| \sin \frac{1}{x}.$$

Comme

$$\lim_{(x, y) \rightarrow (x_0, 0)} |x + y| \sin \frac{1}{x} = |x_0| \sin \frac{1}{x_0} = 0,$$

pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que, pour $(x, y) \in \mathbb{R}^2$,

$$d((x, y), (x_0, 0)) \leq \eta \text{ et } xy \neq 0 \implies |f(x, y) - f(x_0, 0)| \leq \varepsilon.$$

Comme si $xy = 0$, on a $f(x, y) = f(x_0, y_0)$, on obtient pour $(x, y) \in \mathbb{R}^2$,

$$d((x, y), (x_0, 0)) \leq \eta \implies |f(x, y) - f(x_0, 0)| \leq \varepsilon$$

et la continuité de f en $(x_0, 0)$.

On a donc montré que si $x_0 \neq 0$, la fonction est continue en $(x_0, 0)$ si, et seulement si, $\sin \frac{1}{x_0} = 0$. De même, si $y_0 \neq 0$, la fonction f est continue en $(0, y_0)$ si, et seulement si, $\sin \frac{1}{y_0} = 0$.

4. La fonction φ étant continue sur \mathbb{R} , la fonction $\psi : (x, y) \mapsto y - \varphi(x)$ est continue sur \mathbb{R}^2 . On en déduit que les ensembles

$$\Omega_1 = \{(x, y) \in \mathbb{R}^2, y < \varphi(x)\} = \psi^{-1}(] - \infty, 0[)$$

et

$$\Omega_2 = \{(x, y) \in \mathbb{R}^2, y > \varphi(x)\} = \psi^{-1}(]0, +\infty[)$$

sont des ouverts de \mathbb{R}^2 . Comme la fonction nulle et la fonction ψ sont continues sur \mathbb{R}^2 , la fonction f est continue en tout point de Ω_1 et en tout point de Ω_2 .

Il faut étudier la continuité de f en un point (x_0, y_0) tel que $y_0 = \varphi(x_0)$. On a alors $f(x_0, y_0) = 0$. Soit $(x, y) \in \mathbb{R}^2$. Si $y \leq \varphi(x)$, on obtient

$$|f(x, y) - f(x_0, y_0)| = |y - \varphi(x)| = |y - y_0 - \varphi(x) + \varphi(x_0)| \leq |y - y_0| + |\varphi(x) - \varphi(x_0)|.$$

Cette inégalité reste vraie si $y > \varphi(x)$ car alors $|f(x, y) - f(x_0, y_0)| = 0$.

De $\lim_{(x, y) \rightarrow (x_0, y_0)} |y - y_0| + |\varphi(x) - \varphi(x_0)| = 0$, on déduit $\lim_{(x, y) \rightarrow (x_0, y_0)} f(x, y) = f(x_0, y_0)$ et la continuité de f en (x_0, y_0) . La fonction f est donc continue sur \mathbb{R}^2 .

- 5** 1. a) Comme, pour tout $(x, y, z) \in \mathbb{R}^3$, $|x|$, $|y|$ et $|z|$ sont inférieurs à $\|(x, y, z)\|$, on obtient, pour $(x, y, z) \neq (0, 0, 0)$,

$$|f(x, y, z)| \leq \frac{\|(x, y, z)\|^{p+q+r}}{\|(x, y, z)\|} \leq \|(x, y, z)\|^{p+q+r-1}.$$

Pour tout $x \in \mathbb{R}^*$,

$$f(x, x, x) = \frac{|x|^{p+q+r}}{\sqrt{3x^2}} = \frac{1}{\sqrt{3}}|x|^{p+q+r-1}.$$

- b) • Si $p + q + r - 1 > 0$, on a $\lim_{t \rightarrow 0^+} t^{p+q+r-1} = 0$ et d'après l'inégalité $|f(x, y, z)| \leq \|(x, y, z)\|^{p+q+r-1}$, on a $\lim_{(x,y,z) \rightarrow (0,0,0)} f(x, y, z) = 0 = f(0, 0, 0)$: la fonction f est continue en $(0, 0, 0)$.
 • Si $x + y + z - 1 \leq 0$, on n'a pas $\lim_{t \rightarrow 0^+} t^{p+q+r-1} = 0$, donc $f(x, x, x)$ ne tend pas vers 0 lorsque x tend vers 0, bien que (x, x, x) tende vers $(0, 0, 0)$: f n'est pas continue en $(0, 0, 0)$.
 La fonction f est continue en $(0, 0, 0)$ si, et seulement si, $p + q + r - 1 > 0$.

2. On a, pour tout $x > 0$,

$$g\left(x^{\frac{1}{s}}, x^{\frac{1}{t}}, x^{\frac{1}{u}}\right) = \frac{x^{\frac{p}{s}} x^{\frac{q}{t}} x^{\frac{r}{u}}}{3x} = \frac{1}{3} x^{\frac{p}{s} + \frac{q}{t} + \frac{r}{u} - 1}.$$

- On montre comme dans la question précédente que, si $\frac{p}{s} + \frac{q}{t} + \frac{r}{u} - 1 \leq 0$, la fonction g n'est pas continue en $(0, 0, 0)$.
- Si $\frac{p}{s} + \frac{q}{t} + \frac{r}{u} - 1 > 0$, on écrit, pour $(x, y, z) \in \mathbb{R}^3 \setminus \{(0, 0, 0)\}$,

$$\begin{aligned} g(x, y, z) &= \frac{(|x|^s)^{\frac{p}{s}} (|y|^t)^{\frac{q}{t}} (|z|^u)^{\frac{r}{u}}}{|x|^s + |y|^t + |z|^u} \leq \frac{(|x|^s + |y|^t + |z|^u)^{\frac{p}{s} + \frac{q}{t} + \frac{r}{u}}}{|x|^s + |y|^t + |z|^u} \\ &= (|x|^s + |y|^t + |z|^u)^{\frac{p}{s} + \frac{q}{t} + \frac{r}{u} - 1}. \end{aligned}$$

Comme la fonction $h : (x, y, z) \mapsto |x|^s + |y|^t + |z|^u$ est continue en $(0, 0, 0)$, que $h(0, 0, 0) = 0$ et que la fonction $t \mapsto t^{\frac{p}{s} + \frac{q}{t} + \frac{r}{u} - 1}$ a pour limite 0 en 0, on obtient, par le théorème de composition des limites,

$$\lim_{(x,y,z) \rightarrow (0,0,0)} (|x|^s + |y|^t + |z|^u)^{\frac{p}{s} + \frac{q}{t} + \frac{r}{u} - 1} = 0$$

et a fortiori,

$$\lim_{(x,y,z) \rightarrow (0,0,0)} g(x, y, z) = 0.$$

La fonction g est continue en $(0, 0, 0)$ si, et seulement si, $\frac{p}{s} + \frac{q}{t} + \frac{r}{u} - 1 > 0$.

- 6** • Lorsque $x \neq 0$, la dérivée de la fonction trinôme

$$g_{x,y,z} : t \mapsto xt^2 + yt + z$$

s'annule en $-\frac{y}{2x}$ et $g_{x,y,z}\left(-\frac{y}{2x}\right) = z - \frac{y^2}{4}$. Si $x < 0$, la fonction $g_{x,y,z}$ atteint son maximum en ce point. Si $x < 0$ et $-\frac{y}{2x} \in]0, 1[$, c'est-à-dire si $0 < y < -2x$, on a donc

$$f(x, y, z) = z - \frac{y^2}{4x}.$$

Dans le cas où la dérivée s'annule hors de $]0, 1[$, la fonction $g_{x,y,z}$ est monotone sur $[0, 1]$ et son maximum est obtenu en 0 ou 1. Il en est de même si $x = 0$, car alors la fonction f est affine donc monotone sur \mathbb{R} tout entier. Enfin, si $x > 0$ et si $g'_{x,y,z}$ s'annule sur $]0, 1[$, ce point où $g'_{x,y,z}$ s'annule correspond à un minimum et le maximum de g est obtenu en 0 et 1. Finalement si la condition $0 < y < -2x$ n'est pas réalisée, on a

$$f(x, y, z) = \max(g_{x,y,z}(0), g_{x,y,z}(1)) = \max(z, x + y + z) = z + \max(0, x + y).$$

La fonction est donc définie sur \mathbb{R}^3 par

$$f(x, y, z) = \begin{cases} z - \frac{y^2}{4x} & \text{si } 0 < y < -2x \\ z + \max(0, x + y) & \text{si } y \leq 0 \text{ ou } y + 2x \geq 0. \end{cases}$$

• Pour étudier la continuité de f , on peut procéder comme dans les exercices précédents. Il faut essentiellement étudier la continuité aux points où la formule donnant $f(x, y, z)$ change. Il est plus simple de revenir à la définition de f .

Soit (x, y, z) et (x', y', z') deux éléments de \mathbb{R}^3 , $t \in [0, 1]$ tel que $f(x, y, z) = g_{x,y,z}(t)$. On a alors

$$\begin{aligned} f(x, y, z) &= xt^2 + yt + z = (x - x')t^2 + (y - y')t + (z - z') + x't^2 + y't + z' \\ &\leq |x - x'| + |y - y'| + |z - z'| + g_{x',y',z'}(t) \\ &\leq |x - x'| + |y - y'| + |z - z'| + f(x', y', z'). \end{aligned}$$

On a donc

$$f(x, y, z) - f(x', y', z') \leq |x - x'| + |y - y'| + |z - z'|$$

et en échangeant les rôles de (x, y, z) et (x', y', z') ,

$$f(x, y, z) - f(x', y', z') \leq |x - x'| + |y - y'| + |z - z'|.$$

On obtient finalement

$$\begin{aligned} |f(x, y, z) - f(x', y', z')| &\leq |x - x'| + |y - y'| + |z - z'| \\ &\leq 3\|(x, y, z) - (x', y', z')\|. \end{aligned}$$

On a donc, pour X et X' dans \mathbb{R}^3 ,

$$|f(X) - f(X')| \leq 3\|X - X'\|.$$

Ceci montre la continuité de f en tout point car, pour tout $\varepsilon > 0$, on a pour tout $(X, X') \in (\mathbb{R}^3)^2$,

$$\|X - X'\| \leq \frac{\varepsilon}{3} \implies |f(X) - f(X')| \leq \varepsilon.$$

7 1. Pour tout $t \in [0, 2\pi]$, $(\cos t, \sin t)$ et $(-\cos t, -\sin t)$ appartiennent à C donc $g(t)$ est définie. La fonction f est continue sur C et les fonctions \sin et \cos sont continues sur $[0, 2\pi]$ donc g est continue sur $[0, 2\pi]$ d'après le théorème de composition.

On a $g(0) = f(1, 0) - f(-1, 0)$ et $g(\pi) = f(-1, 0) - f(1, 0) = -g(0)$.

2. La fonction g est continue et prend des valeurs opposées en 0 et π . D'après le théorème des valeurs intermédiaires, il existe $t_0 \in [0, \pi]$ tel que $g(t_0) = 0$. En notant $X_0 = (\cos t_0, \sin t_0)$, point de C , on obtient $g(t_0) = f(X_0) - f(-X_0) = 0$ et donc $f(X_0) = f(-X_0)$.

8 1. Comme Ω est convexe, on a, pour tout $t \in [0, 1]$, $(1-t)x + ty \in [0, 1]$ donc φ est définie sur $[0, 1]$. Si $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$, alors

$$(1-t) + ty = ((1-t)x_1 + ty_1, \dots, (1-t)x_n + ty_n).$$

Comme les fonctions $t \mapsto (1-t)x_i + ty_i$ ($1 \leq i \leq n$) sont continues sur $[0, 1]$ et que f est continue sur Ω , il résulte du théorème de composition que f est continue sur $[0, 1]$.

2. L'image de $[0, 1]$ par l'application continue φ est un intervalle qui contient $\varphi(0) = f(x)$ et $\varphi(1) = f(y)$ et donc toute valeur entre $f(x)$ et $f(y)$. Mais on a $\varphi([0, 1]) \subset f(\Omega)$, car, pour tout $t \in [0, 1]$, $(1-t)x + ty \in \Omega$. On obtient donc que tout réel entre $f(x)$ et $f(y)$ appartient à $f(\Omega)$. Cela étant vrai pour tous éléments x et y de Ω , l'ensemble $f(\Omega)$ vérifie la propriété caractéristique d'un intervalle : $f(\Omega)$ est un intervalle de \mathbb{R} .

9 1. K est l'intersection des demi-espaces $x_1 \geq 0, \dots, x_n \geq 0$ et de l'hyperplan affine

$\sum_{i=1}^n x_i = s$ qui sont des fermés. Donc K , intersection d'un nombre fini de fermés, est un fermé.

Si $x = (x_1, \dots, x_n) \in K$, on a, pour $1 \leq i \leq n$, $0 \leq x_i \leq s$ et donc $\|x\| \leq \sqrt{ns}$. Ainsi K est borné.

La fonction f est continue car polynomiale. Elle est donc bornée sur K et atteint ses bornes. Donc $f(K)$ possède un maximum.

2. L'inégalité $\sqrt{xy} \leq \frac{x+y}{2}$ équivaut à $(\sqrt{x} - \sqrt{y})^2 \geq 0$; il y a égalité pour $x = y$.

Si $n = 1$ le maximum est s . Si $n = 2$, on a $x_1 x_2 \leq \left(\frac{x_1 + x_2}{2}\right)^2 \leq \left(\frac{s}{2}\right)^2$, avec égalité si $x_1 = x_2 = \frac{s}{2}$.

Supposons qu'au rang n , pour $s > 0$ quelconque on a $f(x) \leq \left(\frac{s}{n}\right)^n$ pour tout $x \in K$, avec égalité si, et seulement si, $x_1 = \dots = x_n = \frac{s}{n}$ et étudions la propriété au rang $n + 1$.

Soit $x = (x_1, \dots, x_{n+1}) \in (\mathbb{R}_+)^{n+1}$ tel que $\sum_{i=1}^{n+1} x_i = s$. On a alors $\sum_{i=1}^n x_i = s - x_{n+1}$ et par hypothèse de récurrence

$$f(x) = \prod_{i=1}^n x_i \times x_{n+1} \leq \left(\frac{s - x_{n+1}}{n}\right)^n x_{n+1} = \frac{(s - x_{n+1})^n x_{n+1}}{n^n}.$$

On étudie la fonction $\varphi : x \mapsto (s-x)^n x$ sur $[0, s]$. On obtient

$$\varphi'(x) = (s-x)^{n-1}(s-(n+1)x).$$

Le maximum de φ est atteint en $x = \frac{s}{n+1}$ et vaut $\frac{n^n s^{n+1}}{(n+1)^{n+1}}$.

On en déduit que, pour tout $x \in K$, $f(x) \leq \frac{s^{n+1}}{(n+1)^{n+1}}$, le maximum étant atteint si

$x_{n+1} = \frac{s}{n+1}$ et $\prod_{i=1}^n x_i = \left(\frac{s-x_{n+1}}{n}\right)^n$, ce qui par hypothèse de récurrence n'est réalisé

que pour $x_1 = \dots = x_n = \frac{s-x_{n+1}}{n} = \frac{s}{n+1}$.

La propriété est donc vérifiée au rang $n+1$, ce qui termine la démonstration par récurrence.

3. Il a été démontré dans la question précédente que, pour $(x_1, \dots, x_n) \in (\mathbb{R}_+)^n$,

$$\prod_{i=1}^n x_i \leq \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^n,$$

ce qui donne l'inégalité voulue en élevant à la puissance $\frac{1}{n}$.

- 10** 1. La fonction $t \mapsto \varphi(t) - \sum_{k=0}^n a_k t^k$ est continue donc bornée sur le segment $[0, 1]$.

Cela justifie l'existence de la borne supérieure.

2. Soit a et b dans \mathbb{R}^{n+1} . Pour tout $t \in [0, 1]$,

$$\begin{aligned} \left| \varphi(t) - \sum_{k=0}^n a_k t^k \right| &= \left| \varphi(t) - \sum_{k=0}^n b_k t^k + \sum_{k=0}^n (b_k - a_k) t^k \right| \\ &\leq \left| \varphi(t) - \sum_{k=0}^n b_k t^k \right| + \sum_{k=0}^n |b_k - a_k| t^k \\ &\leq f(b) + \sum_{k=0}^n |b_k - a_k|. \end{aligned}$$

En appliquant l'inégalité de Cauchy-Schwarz,

$$\sum_{k=0}^n |b_k - a_k| = \sum_{k=0}^n (1 \cdot |b_k - a_k|) \leq \sqrt{\sum_{k=0}^n 1} \sqrt{\sum_{k=0}^n (b_k - a_k)^2} \leq \sqrt{n+1} \|b - a\|.$$

On a donc, pour tout $t \in [0, 1]$,

$$\left| \varphi(t) - \sum_{k=0}^n a_k t^k \right| \leq f(b) + \sqrt{n+1} \|b - a\|.$$

Ainsi $f(b) + \sqrt{n+1}\|b-a\|$ majore la fonction $t \mapsto \left| \varphi(t) - \sum_{k=0}^n a_k t^k \right|$ et par définition de la borne supérieure, on a

$$f(a) \leq f(b) + \sqrt{n+1}\|b-a\|.$$

En échangeant les rôles de a et b , on montre de même que $f(b) \leq f(a) + \sqrt{n+1}\|b-a\|$ et finalement

$$|f(b) - f(a)| \leq \sqrt{n+1}\|b-a\|.$$

Soit $a \in \mathbb{R}^{n+1}$ et $\varepsilon > 0$. On a, pour tout $b \in \mathbb{R}^{n+1}$,

$$\|b-a\| \leq \frac{\varepsilon}{\sqrt{n+1}} \implies |f(b) - f(a)| \leq \varepsilon,$$

ce qui montre la continuité de f en a , pour tout $a \in \mathbb{R}^{n+1}$: la fonction f est continue sur \mathbb{R}^{n+1} .

3. a) Si $a \neq 0$, la fonction $t \mapsto \sum_{k=0}^n a_k t^k$ n'est pas identiquement nulle sur $[0, 1]$, donc

$g(a) \neq 0$ et $g(a) > 0$.

L'ensemble des éléments de \mathbb{R}^{n+1} de norme 1 est un fermé borné. La fonction g possède sur ce fermé borné un minimum qui est atteint. D'où l'existence de μ et de $a_0 \in \mathbb{R}^{n+1}$ tel que $\|a_0\| = 1$ et $\mu = g(a_0)$, ce qui montre que $\mu > 0$.

Soit $a \in \mathbb{R}^{n+1}$, non nul, $b = \frac{1}{\|a\|}a$. Alors b est de norme 1 donc $g(b) \geq \mu$ et comme, pour tout k , $a_k = \|a\|b_k$,

$$g(a) = \sup_{t \in [0,1]} \left| \sum_{k=0}^n a_k t^k \right| = \sup_{t \in [0,1]} \left\| \|a\| \sum_{k=0}^n b_k t^k \right\| = \|a\|g(b) \geq \|a\|\mu.$$

Par ailleurs, cette inégalité est évidente pour $a = 0$ car $g(0) = 0$.

- b) Soit $a \in \mathbb{R}^{n+1}$. On a, pour tout $t \in [0, 1]$, par inégalité triangulaire,

$$\left| \varphi(t) - \sum_{k=0}^n a_k t^k \right| \geq \left| \sum_{k=0}^n a_k t^k \right| - |\varphi(t)| \geq \left| \sum_{k=0}^n a_k t^k \right| - K.$$

La fonction $t \mapsto \sum_{k=0}^n a_k t^k$ étant continue atteint sa borne supérieure sur $[0, 1]$. En

appliquant l'inégalité précédente à un réel t tel que $g(a) = \sum_{k=0}^n a_k t^k$, on obtient

$$f(a) \geq \left| \varphi(t) - \sum_{k=0}^n a_k t^k \right| \geq \left| \sum_{k=0}^n a_k t^k \right| - K \geq g(a) - K \geq \mu\|a\| - K.$$

Comme $\lim_{\|a\| \rightarrow +\infty} (\mu\|a\| - K) = +\infty$, on a *a fortiori* $\lim_{\|a\| \rightarrow +\infty} f(a) = +\infty$.

4. La fonction f est positive sur \mathbb{R}^{n+1} donc possède une borne inférieure m .

Puisque $\lim_{\|a\| \rightarrow +\infty} f(a) = +\infty$, il existe $r > 0$ tel que, pour tout $a \in \mathbb{R}^{n+1}$, $\|a\| \geq r$ implique $f(a) \geq m + 1$. On a alors

$$m = \inf_{\|a\| \leq r} f(a) = \inf_{a \in B(0, r)} f(a).$$

Comme la boule fermée de centre 0 de rayon r est un fermé borné, la borne inférieure de f sur cette boule est un minimum, i.e il existe $a_1 \in B(0, r)$ tel que $f(a_1) = m$. Ainsi f possède un minimum sur \mathbb{R}^{n+1} atteint en a_1 .

- 11 1. a) L'ensemble $\{|P(z)|, z \in \mathbb{C}\}$ est une partie de \mathbb{R} minorée par 0. Elle possède donc une borne inférieure m .

- b) Écrivons $P(z) = \sum_{k=0}^p a_k z^k$, où $p \geq 1$ et $a_p \neq 0$. Pour tout $z \in \mathbb{C}^*$, on a, par inégalité triangulaire,

$$\begin{aligned} |P(z)| &\geq |a_p||z|^p - \left| \sum_{k=0}^{p-1} a_k z^k \right| \geq |a_p||z|^p - \sum_{k=0}^{p-1} |a_k||z|^k \\ &\geq |a_p||z|^p \left(1 - \sum_{k=0}^{p-1} \frac{|a_k|}{|a_p||z|^{p-k}} \right). \end{aligned}$$

Comme $\lim_{x \rightarrow +\infty} \left(1 - \sum_{k=0}^{p-1} \frac{|a_k|}{|a_p|x^{p-k}} \right) = 1$, il existe $r > 0$ tel que, pour tout $x \in \mathbb{R}$,

$$x \geq r \implies \left(1 - \sum_{k=0}^{p-1} \frac{|a_k|}{|a_p|x^{p-k}} \right) \geq \frac{1}{2}.$$

On en déduit que, pour tout $z \in \mathbb{C}$,

$$|z| \geq r \implies |P(z)| \geq \frac{1}{2}|a_p||z|^p.$$

De $\lim_{x \rightarrow +\infty} \frac{1}{2}|a_p|x^p = +\infty$, on déduit l'existence de $r' \geq r$ tel que, pour tout réel x ,

$$x \geq r' \implies \frac{1}{2}|a_p|x^p \geq m + 1.$$

On en déduit que, pour tout $z \in \mathbb{C}$,

$$|z| \geq r' \implies |P(z)| \geq \frac{1}{2}|a_p||z|^p \geq m + 1.$$

- c) En développant les expressions $(x + iy)^k$ et séparant partie réelle et partie imaginaire, on obtient des fonctions polynomiales R et S définies sur \mathbb{R}^2 telles que, pour tout $(x, y) \in \mathbb{R}^2$, $P(x + iy) = R(x, y) + iS(x, y)$. On en déduit

$$f(x, y) = \sqrt{(P(x, y))^2 + (Q(x, y))^2}.$$

La fonction $(x, y) \mapsto (P(x, y))^2 + (Q(x, y))^2$ est continue sur \mathbb{R}^2 , car polynomiale, et positive. On en déduit que f est continue sur \mathbb{R}^2 .

Par définition, on a $\inf_{(x,y) \in \mathbb{R}^2} f(x, y) = m$. Comme $\|(x, y)\| = |x + iy|$, il résulte de b que, pour tout $(x, y) \in \mathbb{R}^2$,

$$\|(x, y)\| \geq r' \implies f(x, y) \geq m + 1.$$

Il en résulte que

$$m = \inf_{(x,y) \in B} f(x, y),$$

où B est la boule fermée de centre 0 et de rayon r' . Comme B est un fermé borné, f atteint sa borne supérieure sur B . Il existe $(x_0, y_0) \in B$ tel que $f(x_0, y_0) = m$. En posant $z_0 = x_0 + iy_0$, on obtient

$$|P(z_0)| = f(x_0, y_0) = m.$$

2. a) On a $b_0 = Q(0) = \frac{P(z_0)}{P(z_0)} = 1$. D'autre part, Q est un polynôme de degré p , car pour tout $i \in \llbracket 0, p \rrbracket$, $(z + z_0)^i$ est de degré i . On en déduit que $b_p \neq 0$.
- b) Par définition de k , on a, pour $z \in \mathbb{C}$,

$$Q(z) = 1 + b_k z^k + \sum_{i=k+1}^p b_i z^i,$$

et donc, pour tout $x \in]0, 1[$,

$$Q(\omega x) = 1 + b_k \omega^k x^k + \sum_{i=k+1}^p b_i \omega^i x^i = 1 - x^k + \sum_{i=k+1}^p b_i \omega^i x^i.$$

On en déduit, par inégalité triangulaire,

$$|Q(\omega x)| \leq |1 - x^k| + \sum_{i=k+1}^p |b_i \omega^i| x^i \leq 1 - x^k + \sum_{i=k+1}^p |b_i \omega^i| x^i$$

car $1 - x^k \in \mathbb{R}_+$.

On a $\lim_{x \rightarrow 0} \sum_{i=k+1}^p |b_i \omega^i| x^{i-k} = 0$, donc il existe $\eta \in]0, 1[$ tel que, pour tout réel x ,

$$x \in]0, \eta[\implies \sum_{i=k+1}^p |b_i \omega^i| x^{i-k} \leq \frac{1}{2}.$$

On a alors, pour tout $x \in]0, \eta[$,

$$|Q(\omega x)| \leq 1 - x^k + \frac{1}{2} x^k \leq 1 - \frac{1}{2} x^k < 1.$$

- c) Par définition de m , on a pour tout $z \in \mathbb{C}$, $|P(z + z_0)| \geq m$ et $|P(z_0)| = m$. On en déduit que, pour tout $z \in \mathbb{C}$,

$$|Q(z)| \geq 1.$$

L'existence de x tel que $|Q(\omega x)| < 1$ conduit à une contradiction. On n'a donc pas $P(z_0) \neq 0$, mais $P(z_0) = 0$. Le polynôme P possède au moins une racine.

- 12 1. Par inégalité triangulaire, on a, pour x et y dans \mathbb{R}^n ,

$$|f(x) - f(y)| = |d(x_0, x) - d(x_0, y)| \leq d(x, y).$$

On en déduit que, pour tout $x \in \mathbb{R}^n$, $\lim_{y \rightarrow x} f(y) = f(x)$ et la continuité de f sur \mathbb{R}^n .

2. La définition de $d(x, \Omega)$ est justifiée par le fait que $\{d(x, u), u \in \Omega\}$ est une partie non vide de \mathbb{R} , minorée par 0.

On a, pour x et x' dans \mathbb{R}^n et $u \in \Omega$,

$$d(x, \Omega) \leq d(x, u) \leq d(x, x') + d(x', u)$$

et donc, pour tout $u \in \Omega$, $d(x', u) \geq d(x, \Omega) - d(x, x')$. Par définition de la borne inférieure, on en déduit

$$d(x', \Omega) \geq d(x, \Omega) - d(x, x'),$$

soit $d(x, \Omega) - d(x', \Omega) \leq d(x, x')$. En échangeant les rôles de x et x' , on obtient $d(x', \Omega) - d(x, \Omega) \leq d(x, x')$ et donc

$$|d(x, \Omega) - d(x', \Omega)| \leq d(x, x').$$

On en déduit que, pour tout $x \in \mathbb{R}^n$, on a $\lim_{x' \rightarrow x} d(x', \Omega) = d(x, \Omega)$. La fonction $x \mapsto d(x, \Omega)$ est continue sur \mathbb{R}^n .

3. a) La fonction $u \mapsto d(x, u)$ est continue d'après la question 1. On en déduit qu'elle est bornée et atteint ses bornes sur le fermé borné Ω . En particulier, il existe $\gamma \in \Omega$ tel que

$$d(x, \Omega) = \inf_{u \in \Omega} d(x, u) = d(x, \gamma).$$

- b) On a, pour tout $u \in \Omega$, $d(x, u) \geq \|u\| - \|x\|$.

On en déduit que si, de plus, $\|u\| \geq \alpha + 1 + \|x\|$, alors $d(x, u) \geq \alpha + 1$.

Posons $\beta = \inf_{u \in B \cap \Omega} d(x, u)$. Comme α minore $d(x, u)$ pour $u \in \Omega$, il minore $d(x, u)$ sur $B \cap \Omega$ et donc $\beta \geq \alpha$. Si on avait $\beta > \alpha$, comme sur $\Omega \setminus B$, $d(x, u)$ est minoré par $\alpha + 1$, $d(x, u)$ serait minoré par $\min(\alpha + 1, \beta) > \alpha$, ce qui contredit la définition de la borne inférieure. On a donc

$$\alpha = \inf_{u \in B \cap \Omega} d(x, u).$$

L'ensemble $B \cap \Omega$ est fermé, car c'est l'intersection de deux fermés. Il est borné car inclus dans B . La fonction continue $u \mapsto d(x, u)$ atteint donc sa borne inférieure sur $B \cap \Omega$. Il existe $\gamma \in B \cap \Omega$ tel que

$$d(x, \Omega) = \alpha = d(x, \gamma).$$

4. Soit $\gamma \in \Omega'$. On a, pour tout $x \in \Omega$,

$$d(x, \gamma) \geq d(\gamma, \Omega).$$

La fonction $\gamma \mapsto d(\gamma, \Omega)$ est continue d'après la question 2, donc elle est bornée et atteint ses bornes sur Ω' . Si son minimum est atteint en γ_0 , on a, pour tout $(x, \gamma) \in \Omega \times \Omega'$,

$$d(x, \gamma) \geq d(\gamma, \Omega) \geq d(\gamma_0, \Omega).$$

On en déduit que, pour tout $(x, y) \in \Omega \times \Omega'$,

$$\inf_{(x,y) \in \Omega \times \Omega'} d(x, y) \geq d(y_0, \Omega).$$

Mais il résulte de la question 3 que, Ω étant fermé, il existe $x_0 \in \Omega$ tel que $d(y_0, \Omega) = d(y_0, x_0)$. Comme Ω et Ω' sont disjoints, x_0 et y_0 sont distincts donc $d(y_0, \Omega) = d(y_0, x_0) > 0$. Finalement, on a

$$\inf_{(x,y) \in \Omega \times \Omega'} d(x, y) > 0.$$

13 1. Cette question a déjà été traitée dans l'exercice 12, question 3.

2. Pour x, y et y' dans \mathbb{R}^n , on a

$$\begin{aligned} 2 \left\| x - \frac{1}{2}(y + y') \right\|^2 + \frac{1}{2} \|y - y'\|^2 &= 2 \|x\|^2 - 2\langle x, y + y' \rangle + \frac{1}{2} (\|y + y'\|^2 + \|y - y'\|^2) \\ &= 2 \|x\|^2 - 2\langle x, y' \rangle - 2\langle x, y \rangle + \|y\|^2 + \|y'\|^2 \\ &= \|x - y\|^2 + \|x - y'\|^2. \end{aligned}$$

Supposons que $d(x, C) = d(x, y) = d(x, y')$. On a alors

$$\left\| x - \frac{1}{2}(y + y') \right\|^2 = (d(x, C))^2 - \frac{1}{4} \|y - y'\|^2.$$

Si $y \neq y'$, on en déduit

$$\left\| x - \frac{1}{2}(y + y') \right\| < d(x, C).$$

Comme C est convexe et contient y et y' , on a $\frac{1}{2}(y + y') \in C$. L'inégalité qui précède est donc contradictoire avec la définition de $d(x, C)$. On a donc $y = y'$ et il existe un unique $y \in C$ tel que

$$d(x, C) = d(x, y).$$

3. a) Soit C un sous-espace vectoriel de \mathbb{R}^n . Si $C = \mathbb{R}^n$, C est fermé. Sinon, C peut s'écrire comme intersection d'un nombre fini d'hyperplans. Comme tout hyperplan est un fermé, on en déduit que C est fermé. Par ailleurs, on sait que C est convexe et non vide (car il contient 0).

b) Soit $x \in \mathbb{R}^n$, y le projeté orthogonal de x sur C . Il existe donc $z \in C^\perp$ tel que $x = y + z$. On a alors, pour tout $u \in C$, comme $y - u$ et z sont orthogonaux,

$$\|x - u\|^2 = \|y - u + z\|^2 = \|y - u\|^2 + \|z\|^2 \geq \|z\|^2 \geq \|x - y\|^2.$$

On a donc $y = \text{proj}_C(x)$.

4. a) Comme y et u sont dans C qui est convexe, $(1 - t)y + tu$ appartient à C . Par définition de y , on a donc

$$\|y - x\| \leq \|(1 - t)y + tu - x\|.$$

Si on écrit

$$\|(1 - t)y + tu - x\|^2 = \|t(u - y) + y - x\|^2 = t^2 \|u - y\|^2 + 2t \langle u - y, y - x \rangle + \|y - x\|^2,$$

l'inégalité précédente devient

$$t^2 \|u - \gamma\|^2 + 2t \langle u - \gamma, \gamma - x \rangle \geq 0$$

et pour $t \in]0, 1]$,

$$t \|u - \gamma\|^2 + 2 \langle u - \gamma, \gamma - x \rangle \geq 0.$$

En faisant tendre t vers 0, on en déduit

$$\langle u - \gamma, \gamma - x \rangle \geq 0.$$

b) On a, pour tout $u \in C$,

$$\|u - x\|^2 = \|u - \gamma + \gamma - x\|^2 = \|u - \gamma\|^2 + \|\gamma - x\|^2 + 2 \langle u - \gamma, \gamma - x \rangle \geq \|\gamma - x\|^2,$$

car $\langle u - \gamma, \gamma - x \rangle \geq 0$. L'inégalité obtenue montre que $\gamma = \text{proj}_C(x)$.

Chapitre 6

1 La fonction f est de classe \mathcal{C}^2 sur $\mathbb{R} \times \mathbb{R}_+^*$ et

$$\frac{\partial f}{\partial x}(x, \gamma) = 2x\sqrt{\gamma}, \quad \frac{\partial f}{\partial \gamma}(x, \gamma) = \frac{1}{2}x^2\gamma^{-\frac{1}{2}},$$

$$\frac{\partial^2 f}{\partial x^2}(x, \gamma) = 2\sqrt{\gamma}, \quad \frac{\partial^2 f}{\partial x \partial \gamma}(x, \gamma) = x\gamma^{-\frac{1}{2}}, \quad \frac{\partial^2 f}{\partial \gamma^2}(x, \gamma) = -\frac{1}{4}x^2\gamma^{-\frac{3}{2}}.$$

2. La fonction f est définie sur l'ouvert

$$D = \{(x, \gamma) \in \mathbb{R}^2, x + \sqrt{x^2 + \gamma^2} > 0\} = \mathbb{R}^2 \setminus \mathbb{R}_- \times \{0\}$$

(l'ensemble D est un ouvert car c 'est l'image réciproque de \mathbb{R}_+^* par l'application continue $(x, \gamma) \mapsto x + \sqrt{x^2 + \gamma^2}$) et

$$\frac{\partial f}{\partial x}(x, \gamma) = \frac{1 + \frac{2x}{2\sqrt{x^2 + \gamma^2}}}{x + \sqrt{x^2 + \gamma^2}} = \frac{1}{\sqrt{x^2 + \gamma^2}}, \quad \frac{\partial f}{\partial \gamma}(x, \gamma) = \frac{\gamma}{x\sqrt{x^2 + \gamma^2} + (x^2 + \gamma^2)},$$

$$\frac{\partial^2 f}{\partial x^2}(x, \gamma) = \frac{-x}{(x^2 + \gamma^2)^{\frac{3}{2}}}, \quad \frac{\partial^2 f}{\partial x \partial \gamma}(x, \gamma) = \frac{-\gamma}{(x^2 + \gamma^2)^{\frac{3}{2}}},$$

$$\frac{\partial^2 f}{\partial \gamma^2}(x, \gamma) = \frac{\frac{x^3}{(x^2 + \gamma^2)^{\frac{3}{2}}} + x^2 - \gamma^2}{(x\sqrt{x^2 + \gamma^2} + x^2 + \gamma^2)^2}.$$

3. La fonction $f : (x, \gamma) \mapsto e^{\gamma \ln(x+\gamma)}$ est définie et de classe \mathcal{C}^2 sur l'ensemble

$$D = \{(x, \gamma) \in \mathbb{R}^2, x + \gamma > 0\}$$

et on obtient

$$\frac{\partial f}{\partial x}(x, \gamma) = \frac{\gamma}{x + \gamma}(x + \gamma)^\gamma = \gamma(x + \gamma)^{\gamma-1}, \quad \frac{\partial f}{\partial \gamma}(x, \gamma) = \left(\ln(x + \gamma) + \frac{\gamma}{x + \gamma} \right) (x + \gamma)^\gamma,$$

$$\frac{\partial f^2}{\partial x^2}(x, y) = (y^2 - y)(x + y)^{y-2}, \quad \frac{\partial f^2}{\partial x \partial y}(x, y) = ((y^2 + xy) \ln(x + y) + y^2 + x)(x + y)^{y-2},$$

$$\frac{\partial f^2}{\partial y^2}(x, y) = (((x + y) \ln(x + y))^2 + 2y(x + y) \ln(x + y) + y^2 + 2x + y)(x + y)^{y-2}.$$

2 1. On obtient, pour tout $(x, y, z) \in \mathbb{R}^3$,

$$\frac{\partial f}{\partial x}(x, y, z) = 2xf(x, y, z), \quad \frac{\partial^2 f}{\partial x^2}(x, y, z) = (2 + 4x^2)f(x, y, z),$$

$$\frac{\partial^2 f}{\partial x \partial y}(x, y, z) = 4xyf(x, y, z),$$

et des formules du même type pour les autres dérivées partielles.

2. On obtient, pour tout $(x, y, z) \in \mathbb{R}^3$,

$$\frac{\partial f}{\partial x}(x, y, z) = y + z, \quad \frac{\partial^2 f}{\partial x^2}(x, y, z) = 0, \quad \frac{\partial^2 f}{\partial x \partial y}(x, y, z) = 1,$$

et des formules du même type pour les autres dérivées partielles.

3. On obtient, pour tout $(x, y, z) \in \mathbb{R}^3$,

$$\frac{\partial f}{\partial x}(x, y, z) = \frac{yz(1 - x^2 + y^2)}{(1 + x^2 + y^2)^2}, \quad \frac{\partial f}{\partial y}(x, y, z) = \frac{xz(1 + x^2 - y^2)}{(1 + x^2 + y^2)^2},$$

$$\frac{\partial f}{\partial z}(x, y, z) = \frac{xy}{1 + x^2 + y^2},$$

$$\frac{\partial^2 f}{\partial x^2}(x, y, z) = \frac{2xyz(-3 + x^2 - 3y^2)}{(1 + x^2 + y^2)^3}, \quad \frac{\partial^2 f}{\partial y^2}(x, y, z) = \frac{2xyz(-3 - 3x^2 + y^2)}{(1 + x^2 + y^2)^3},$$

$$\frac{\partial^2 f}{\partial z^2}(x, y, z) = 0, \quad \frac{\partial^2 f}{\partial x \partial z}(x, y, z) = \frac{y(1 - x^2 + y^2)}{(1 + x^2 + y^2)^2}, \quad \frac{\partial^2 f}{\partial y \partial z}(x, y, z) = \frac{x(1 - x^2 + y^2)}{(1 + x^2 + y^2)^2},$$

$$\frac{\partial^2 f}{\partial x \partial y}(x, y, z) = \frac{z(1 - x^4 + 6x^2y^2 - y^4)}{(1 + x^2 + y^2)^2}.$$

3 1. La fonction f est de classe C^1 donc continue sur $\mathbb{R}^2 \setminus \{(0, 0)\}$.

Pour $(x, y) \neq (0, 0)$,

$$|f(x, y)| \leq \frac{x^2 + y^2}{\sqrt{x^2 + y^2}} \leq \|(x, y)\|.$$

On en déduit que $\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 0$: f est continue en $(0, 0)$.

Pour $x \neq 0$, on a

$$\frac{f(x, 0) - f(0, 0)}{x} = \frac{\sin x^2}{|x|x}.$$

Comme $\frac{f(x, 0) - f(0, 0)}{x} \underset{x \rightarrow 0}{\sim} \frac{x}{|x|}$ n'a pas de limite en 0 : $\frac{\partial f}{\partial x}$ n'est pas définie en $(0, 0)$. On

montre de même que $\frac{\partial f}{\partial y}$ n'est pas définie en $(0, 0)$.

2. La fonction g est de continue sur $\mathbb{R}^2 \setminus \{(0, 0)\}$ et de classe \mathcal{C}^1 sur $(\mathbb{R}^*)^2$.
On a, pour $(x, y) \neq (0, 0)$,

$$|g(x, y)| \leq \|(x, y)\|^{2\alpha-2}.$$

On en déduit que si $\alpha > 1$, on a $\lim_{(x,y) \rightarrow (0,0)} g(x, y) = 0$.

Pour tout réel t , on a $g(t, t) = \frac{1}{2}|t|^{2\alpha-2}$. Si $\alpha \leq 1$, $g(t, t)$ ne tend pas vers 0 quand t tend vers 0, donc g n'a pas pour limite 0 en $(0, 0)$: la fonction g n'est pas continue en $(0, 0)$.
Les applications $x \mapsto g(x, 0)$ et $y \mapsto g(0, y)$ sont nulles donc on a

$$\frac{\partial g}{\partial x}(0, 0) = \frac{\partial g}{\partial y}(0, 0) = 0.$$

Pour $y \neq 0$, on a, pour tout $x \neq 0$, $\frac{g(x, y) - g(0, y)}{x} = \frac{\varepsilon|y|^\alpha|x|^{\alpha-1}}{x^2 + y^2}$, où $\varepsilon = \pm 1$ selon le signe de x .

Si $\alpha > 1$, on a $\lim_{x \rightarrow 0} \frac{g(x, y) - g(0, y)}{x} = 0$ et $\frac{\partial g}{\partial x}(0, y) = 0$.

Si $\alpha < 1$, le rapport tend vers $\pm\infty$ et $\frac{\partial g}{\partial x}$ n'est pas défini en $(0, y)$.

Si $\alpha = 1$, le rapport n'a pas de limite, car les limites à droite et à gauche ne sont pas les mêmes.

Pour $\frac{\partial g}{\partial y}$ le résultat est le même.

En un point (x, y) tel que $xy \neq 0$, on trouve

$$\frac{\partial f}{\partial x}(x, y) = \frac{\varepsilon\alpha y|xy|^{\alpha-1}}{x^2 + y^2} - \frac{2x|xy|^\alpha}{(x^2 + y^2)^2}.$$

4 1. Ce développement limité s'écrit

$$\begin{aligned} f(h, 1+k) &= f(1, 0) + h \frac{\partial f}{\partial x}(1, 0) + k \frac{\partial f}{\partial y}(1, 0) + \|(h, k)\| \varepsilon(h, k) \\ &= h + 2k + \|(h, k)\| \varepsilon(h, k), \end{aligned}$$

avec $\lim_{(h,k) \rightarrow (0,0)} \varepsilon(h, k) = 0$.

2. On en déduit que lorsque h tend vers 0,

$$\begin{aligned} f(-2t, e^t) &= f(-2t, 1 + t + o(t)) = -2t + 2t + \|(t, o(t))\| \varepsilon(-2t, o(t)) = o(t) \\ f\left(t, \frac{e^t + e^{-t}}{2}\right) &= f(t, 1 + o(t)) = t + \|(t, o(t))\| \varepsilon(t, o(t)) = t + o(t). \end{aligned}$$

On obtient enfin

$$\lim_{t \rightarrow 0} \frac{f(-2t, e^t)}{f\left(t, \frac{e^t + e^{-t}}{2}\right)} = \lim_{t \rightarrow 0} \frac{o(t)}{t + o(t)} = 0.$$

5 La fonction $f : X \mapsto \frac{1}{\|X\|^2}$ est définie sur $\mathbb{R}^n \setminus \{(0, \dots, 0)\}$ par

$$f(x_1, \dots, x_n) = \frac{1}{\sum_{i=1}^n x_i^2}.$$

Elle est de classe \mathcal{C}^1 , car c'est l'inverse d'une fonction polynomiale. On obtient, pour $1 \leq i \leq n$,

$$\frac{\partial f}{\partial x_i}(x_1, \dots, x_n) = -\frac{2x_i}{\|(x_1, \dots, x_n)\|^4}.$$

On a donc $\nabla f_{X_0} = -\frac{2}{\|X_0\|^4} X_0$ et le développement limité d'ordre 1 de f en X_0 est

$$f(X_0 + H) = \frac{1}{\|X_0\|^2} - 2\frac{\langle X_0, H \rangle}{\|X_0\|^4} + \|H\| \varepsilon(H).$$

6 1. La fonction g est définie et continue sur l'ouvert $\Omega = \mathbb{R}^* \times \mathbb{R}$ (Ω est la réunion des deux demi-plans ouverts $x > 0$ et $x < 0$). Si F est une primitive de f sur \mathbb{R} , on a, pour $x \neq 0$,

$$g(x, y) = \frac{F(xy) - F(x)}{x}$$

et par la formule des accroissements finis, il existe c entre x et xy (*i.e.* appartenant à $[x, xy]$ ou $[xy, x]$) tel que

$$g(x, y) = \frac{(xy - x)f(c)}{x} = (y - 1)f(c).$$

Supposons que l'on fixe y et que l'on fasse tendre x vers 0, alors c tend vers 0 et $g(x, y)$ tend vers $(y - 1)f(0)$. Si l'on veut prolonger g en une fonction continue sur \mathbb{R}^2 , il faut nécessairement poser, pour tout $y \in \mathbb{R}$,

$$g(0, y) = (y - 1)f(0).$$

Montrons que la fonction ainsi prolongée est continue en tout point $(0, y_0)$.

D'après ce qui précède, pour tout $(x, y) \in \mathbb{R}^* \times \mathbb{R}$, il existe c entre xy et x tel que $g(x, y) = (y - 1)f(c)$. Remarquons que cela reste vrai quand $x = 0$, car alors $x = xy = 0$ et $c = 0$. Quand (x, y) tend vers $(0, y_0)$, c tend vers 0 et par continuité de f , $f(c)$ tend vers $f(0)$. D'autre part, y tend vers y_0 . On obtient

$$\lim_{(x, y) \rightarrow (0, y_0)} g(x, y) = (y_0 - 1)f(0) = g(0, y_0).$$

La fonction g est donc continue en $(0, y_0)$ pour tout réel y_0 donc continue sur \mathbb{R}^2 .

2. La fonction F étant dérivable sur \mathbb{R} , la fonction g possède des dérivées partielles d'ordre 1 sur Ω . Pour tout $x \neq 0$,

$$\begin{aligned} \frac{\partial g}{\partial x}(x, y) &= \frac{yf(xy) - f(x)}{x} - \frac{F(xy) - F(x)}{x^2} = \frac{1}{x} (yf(xy) - f(x) - g(x, y)), \\ \frac{\partial g}{\partial y}(x, y) &= f(xy). \end{aligned}$$

On remarque que ces dérivées partielles sont continues : la fonction g est donc de classe \mathcal{C}^1 sur Ω .

Étudions l'existence de dérivées partielles en $(0, y)$ pour $y \in \mathbb{R}$.

La fonction partielle $\gamma \mapsto g(0, \gamma) = (\gamma - 1)f(0)$ est clairement dérivable, donc

$$\frac{\partial g}{\partial \gamma}(0, \gamma) = f(0).$$

Pour la dérivée partielle par rapport à x , il faut calculer la limite lorsque x tend vers 0 de

$$\frac{g(x, y) - g(0, y)}{x} = \frac{F(xy) - F(x) - x(y-1)f(0)}{x^2}.$$

Si $y = 1$, cette fonction est identiquement nulle donc $\frac{\partial g}{\partial x}(0, 1) = 0$.

Montrons que l'existence de $f'(0)$ est une condition suffisante d'existence d'une dérivée partielle par rapport à x en tout point $(0, y)$.

Si $f'(0)$ existe la fonction $h : t \mapsto \frac{f(t) - f(0)}{t}$ peut être prolongée, en posant $h(0) = f'(0)$ en une fonction continue sur \mathbb{R} qui vérifie, pour tout $t \in \mathbb{R}$, $f(t) - f(0) = th(t)$. En remarquant que $(y-1)f(0) = \frac{1}{x} \int_x^{xy} f(0) dt$, on obtient, pour $x \neq 0$,

$$\begin{aligned} \frac{g(x, y) - g(0, y)}{x} &= \frac{1}{x^2} \left(\int_x^{xy} f(t) dt - \int_x^{xy} f(0) dt \right) = \frac{1}{x^2} \int_x^{xy} th(t) dt \\ &= \frac{1}{x^2} \int_x^{xy} th(0) dt + \frac{1}{x^2} \int_x^{xy} t(h(t) - h(0)) dt \\ &= \frac{y^2 - 1}{2} h(0) + \frac{1}{x^2} \int_x^{xy} t(h(t) - h(0)) dt. \end{aligned}$$

Montrons que le deuxième terme tend vers 0 quand x tend vers 0. Soit $\varepsilon > 0$. Comme h est continue en 0, il existe $\eta > 0$ tel que, pour tout réel t , $|t| \leq \eta$ implique $|h(t) - h(0)| \leq \varepsilon$.

Soit x et que $|x| \leq \frac{\eta}{|y| + 1}$. On a alors, pour tout t entre x et xy , $|t| \leq \eta$ et donc

$$\frac{1}{x^2} \left| \int_x^{xy} t(h(t) - h(0)) dt \right| \leq \frac{\varepsilon}{x^2} \left| \int_x^{xy} t dt \right| \leq \frac{\varepsilon}{x^2} \frac{|x^2 y^2 - x^2|}{2} \leq \frac{|y^2 - 1| \varepsilon}{2}.$$

Comme ε est un réel strictement positif quelconque, on

$$\lim_{x \rightarrow 0} \frac{1}{x^2} \int_x^{xy} t(h(t) - h(0)) dt = 0$$

et

$$\frac{\partial g}{\partial x}(0, y) = \lim_{x \rightarrow 0} \frac{g(x, y) - g(0, y)}{x} = \frac{y^2 - 1}{2} h(0) = \frac{y^2 - 1}{2} f'(0).$$

3. Si la fonction g est indépendante de x , on a, pour tout $(x, y) \in \mathbb{R}^* \times \mathbb{R}$, $g(x, y) = g(0, y)$, soit $F(xy) - F(x) = x(y-1)f(0)$. En dérivant cette égalité par rapport à y , on obtient, pour $(x, y) \in \mathbb{R}^* \times \mathbb{R}$, $xf(xy) = xf(0)$ et donc $f(xy) = f(0)$. Cela implique que f est constante.

Réciproquement, si f est constante, on a pour $x \neq 0$,

$$g(x, y) = \frac{1}{x} \int_x^{xy} f(0) dt = \frac{1}{x}(xy - x)f(0) = (y - 1)f(0).$$

Ceci étant vrai également si $x = 0$, la fonction g est indépendante de x .

7 1. Soit $(x_1, \dots, x_n) \in \mathbb{R}^n$, $i \in \llbracket 1, n \rrbracket$ et $t > 0$. Par définition des dérivées partielles, la fonction $x \mapsto f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)$ est dérivable sur \mathbb{R} et son nombre dérivé en x_i est $\frac{\partial f}{\partial x_i}(x_1, \dots, x_n)$. Le théorème de dérivation des fonctions composées d'une variable montre que la fonction

$$h : x \mapsto f(tx_1, \dots, tx_{i-1}, tx, tx_{i+1}, \dots, tx_n)$$

est dérivable et son nombre dérivé en x_i est

$$h'(x_i) = t \frac{\partial f}{\partial x_i}(tx_1, \dots, tx_n).$$

Mais on sait par ailleurs que, pour tout $x \in \mathbb{R}$,

$$h(x) = t^\alpha f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n).$$

On en déduit que

$$h'(x_i) = t^\alpha \frac{\partial f}{\partial x_i}(x_1, \dots, x_n).$$

De l'égalité des deux expressions de h' , on déduit, pour tout $x \in \mathbb{R}$,

$$\begin{aligned} t \frac{\partial f}{\partial x_i}(tx_1, \dots, tx_n) &= t^\alpha \frac{\partial f}{\partial x_i}(x_1, \dots, x_n) \text{ et donc} \\ \frac{\partial f}{\partial x_i}(tx_1, \dots, tx_n) &= t^{\alpha-1} \frac{\partial f}{\partial x_i}(x_1, \dots, x_n). \end{aligned}$$

Comme ceci est vrai pour tous $t > 0$ et $(x_1, \dots, x_n) \in \mathbb{R}^n$, la fonction $\frac{\partial f}{\partial x_i}$ est positivement homogène de degré $\alpha - 1$.

2. On fixe $(x_1, \dots, x_n) \in \mathbb{R}^n$. La fonction f est de classe \mathcal{C}^1 sur \mathbb{R}^n ; les fonctions $u_i : t \mapsto tx_i$ ($1 \leq i \leq n$) sont de classe \mathcal{C}^1 sur $]0, +\infty[$. On en déduit, par le théorème de dérivation des fonctions composées que la fonction $\varphi : t \mapsto f(tx_1, \dots, tx_n)$ est dérivable sur $]0, +\infty[$ et que

$$\varphi'(t) = \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}(tx_1, \dots, tx_n)$$

Mais on sait par ailleurs que, pour $t > 0$, $\varphi(t) = t^\alpha f(x_1, \dots, x_n)$ et donc

$$\varphi'(t) = \alpha t^{\alpha-1} f(x_1, \dots, x_n).$$

En écrivant l'égalité de ces deux expressions de $\varphi'(t)$ pour $t = 1$, on obtient

$$\sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}(x_1, \dots, x_n) = \alpha f(x_1, \dots, x_n).$$

3. On fixe $(x_1, \dots, x_n) \in \mathbb{R}^n$ et on considère $\varphi : t \mapsto f(tx_1, \dots, tx_n)$. La fonction φ est dérivable sur $]0, +\infty[$ et

$$\varphi'(t) = \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}(tx_1, \dots, tx_n).$$

Comme l'égalité (*) est vérifiée pour tout couple de réels, on peut l'appliquer à (tx_1, \dots, tx_n) . On obtient

$$\sum_{i=1}^n tx_i \frac{\partial f}{\partial x_i}(tx_1, \dots, tx_n) = \alpha f(tx_1, \dots, tx_n),$$

c'est-à-dire $t\varphi'(t) = \alpha\varphi(t)$. Ainsi φ vérifie l'équation différentielle

$$\forall t > 0 \quad \varphi'(t) = \frac{\alpha}{t}\varphi(t).$$

Une primitive de $t \mapsto \frac{\alpha}{t}$ étant $t \mapsto \alpha \ln t$, on sait qu'il existe $C \in \mathbb{R}$ tel que, pour tout $t > 0$, $\varphi(t) = Ce^{\alpha \ln t} = Ct^\alpha$. Comme $C = \varphi(1) = f(x_1, \dots, x_n)$, on obtient, pour tout $t > 0$,

$$f(tx_1, \dots, tx_n) = \varphi(t) = Ct^\alpha = f(x_1, \dots, x_n)t^\alpha$$

et la fonction f est positivement homogène de degré α .

- 8** 1. La fonction f étant de classe \mathcal{C}^1 sur I^2 , ensemble convexe, la formule des accroissements finis s'applique. Si (x, y) et (x', y') sont deux éléments de I^2 , il existe $(c, d) \in I^2$ tel que

$$f(x, y) - f(x', y') = (x - x') \frac{\partial f}{\partial x}(c, d) + (y - y') \frac{\partial f}{\partial y}(c, d).$$

On en déduit

$$\begin{aligned} |f(x, y) - f(x', y')| &\leq |x - x'| \left| \frac{\partial f}{\partial x}(c, d) \right| + |y - y'| \left| \frac{\partial f}{\partial y}(c, d) \right| \\ &\leq \left(\left| \frac{\partial f}{\partial x}(c, d) \right| + \left| \frac{\partial f}{\partial y}(c, d) \right| \right) \max(|x - x'|, |y - y'|) \\ &\leq k \max(|x - x'|, |y - y'|). \end{aligned}$$

2. a) Pour tout $n \in \mathbb{N}$, on a $a_{n+2} = \max(|u_{n+4} - u_{n+3}|, |u_{n+3} - u_{n+2}|)$. D'après la question 1, on a

$$\begin{aligned} |u_{n+3} - u_{n+2}| &= |f(u_{n+2}, u_{n+1}) - f(u_{n+1}, u_n)| \\ &\leq k \max(|u_{n+2} - u_{n+1}|, |u_{n+1} - u_n|) \leq ka_n. \end{aligned}$$

On obtient de même

$$\begin{aligned} |u_{n+4} - u_{n+3}| &= |f(u_{n+3}, u_{n+2}) - f(u_{n+2}, u_{n+1})| \\ &\leq k \max(|u_{n+3} - u_{n+2}|, |u_{n+2} - u_{n+1}|). \end{aligned}$$

D'après ce qui précède, on a $|u_{n+3} - u_{n+2}| \leq ka_n \leq a_n$ et par définition, $|u_{n+2} - u_{n+1}| \leq a_n$, donc $|u_{n+4} - u_{n+3}| \leq ka_n$ et finalement

$$a_{n+2} \leq ka_n.$$

b) On a donc, pour tout entier naturel n ,

$$a_{2n} \leq k^n a_0 \quad \text{et} \quad a_{2n+1} \leq a_1 k^n.$$

On en déduit que, pour tout entier naturel N ,

$$\sum_{n=0}^{2N+1} a_n \leq \sum_{n=0}^N a_0 k^n + \sum_{n=0}^N a_1 k^n \leq (a_0 + a_1) \frac{1 - k^{N+1}}{1 - k} \leq \frac{a_0 + a_1}{1 - k}$$

et

$$\sum_{n=0}^{2N} a_n \leq \sum_{n=0}^{2N+1} a_n \leq \frac{a_0 + a_1}{1 - k}.$$

Les sommes partielles de la série de terme général a_n positif sont majorées donc elle converge.

c) On a par définition, pour tout entier naturel n ,

$$|u_{n+1} - u_n| \leq a_n.$$

Donc d'après le théorème de comparaison des séries à terme positif, la série $\sum |u_{n+1} - u_n|$ converge, c'est-à-dire que la série $\sum (u_{n+1} - u_n)$ converge absolument. Mais on a, pour tout entier $N \in \mathbb{N}^*$,

$$\sum_{n=0}^{N-1} (u_{n+1} - u_n) = u_N - u_0.$$

Comme les sommes partielles de la série $\sum (u_{n+1} - u_n)$ convergent, on en déduit que la

suite $(u_n)_{n \in \mathbb{N}}$ converge vers $u_0 + \sum_{n=0}^{+\infty} (u_{n+1} - u_n)$.

3. Soit (α', β') un autre élément de I^2 , $(v_n)_{n \in \mathbb{N}}$ la suite définie par

$$v_0 = \alpha', \quad v_1 = \beta', \quad \text{et} \quad \forall n \in \mathbb{N}, \quad v_{n+2} = f(v_{n+1}, v_n).$$

On note ℓ la limite de la suite $(u_n)_{n \in \mathbb{N}}$ et ℓ' la limite de la suite $(v_n)_{n \in \mathbb{N}}$. En utilisant l'inégalité de la question 1, on obtient, pour tout entier naturel n ,

$$|u_{n+2} - v_{n+2}| = |f(u_{n+1}, u_n) - f(v_{n+1}, v_n)| \leq k \max(|u_{n+1} - v_{n+1}|, |u_n - v_n|).$$

En passant à la limite dans cette inégalité, on obtient

$$|\ell - \ell'| \leq k \max(|\ell - \ell'|, |\ell - \ell'|) \leq k|\ell - \ell'|.$$

Comme $k \in [0, 1]$, on en déduit $|\ell - \ell'| = 0$ et donc $\ell = \ell'$. La limite de la suite (u_n) est donc indépendante du choix du couple (α, β) .

9 1. La fonction f est de classe \mathcal{C}^1 sur $\mathbb{R}^2 \setminus \{(0, 0)\}$ et on trouve, pour $(x, y) \neq (0, 0)$,

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= \frac{y(x^4 + 4x^2y^2 - y^4)}{(x^2 + y^2)^2} \\ \frac{\partial f}{\partial y}(x, y) &= \frac{x(x^4 - 4x^2y^2 - y^4)}{(x^2 + y^2)^2}. \end{aligned}$$

Les applications partielles en $(0, 0)$ étant nulles, on a

$$\frac{\partial f}{\partial x}(0, 0) = 0 \quad \text{et} \quad \frac{\partial f}{\partial y}(0, 0) = 0.$$

Pour $(x, y) \neq (0, 0)$, on obtient

$$\left| \frac{\partial f}{\partial x}(x, y) \right| \leq \frac{2|y|(x^4 + 2x^2y^2 + y^4)}{(x^2 + y^2)^2} \leq 2|y| \leq 2\|(x, y)\|$$

et de même

$$\left| \frac{\partial f}{\partial y}(x, y) \right| \leq 2|x| \leq 2\|(x, y)\|.$$

On en déduit $\lim_{(x,y) \rightarrow (0,0)} \frac{\partial f}{\partial x}(x, y) = \lim_{(x,y) \rightarrow (0,0)} \frac{\partial f}{\partial y}(x, y) = 0$ et la continuité des dérivées partielles en $(0, 0)$. La fonction f est donc de classe \mathcal{C}^1 sur \mathbb{R}^2 .

2. On a, par définition

$$\frac{\partial^2 f}{\partial y \partial x}(0, 0) = \lim_{y \rightarrow 0} \frac{\frac{\partial f}{\partial x}(0, y) - \frac{\partial f}{\partial x}(0, 0)}{y} = \lim_{y \rightarrow 0} \frac{-y}{y} = -1$$

et

$$\frac{\partial^2 f}{\partial x \partial y}(0, 0) = \lim_{x \rightarrow 0} \frac{\frac{\partial f}{\partial y}(x, 0) - \frac{\partial f}{\partial y}(0, 0)}{x} = \lim_{x \rightarrow 0} \frac{x}{x} = 1.$$

On constate que

$$\frac{\partial^2 f}{\partial x \partial y}(0, 0) \neq \frac{\partial^2 f}{\partial y \partial x}(0, 0).$$

Le théorème de Schwarz ne s'applique pas. Comme manifestement, $\frac{\partial f}{\partial x \partial y}$ et $\frac{\partial^2 f}{\partial y \partial x}$ sont définies sur \mathbb{R}^2 , elles ne sont pas toutes les deux continues en $(0, 0)$. La fonction f n'est pas de classe \mathcal{C}^2 .

10 1. Soit $(x, y) \in \mathbb{R}^2$. La fonction f étant de classe \mathcal{C}^2 possède un développement limité d'ordre 2 en (x, y) qui s'écrit, en posant $a = \frac{\partial f}{\partial x}(x, y)$ et $b = \frac{\partial f}{\partial y}(x, y)$ et en utilisant les notations de Monge,

$$f(x + h, y + k) = f(x, y) + ah + bk + \frac{1}{2}(rh^2 + shk + tk^2) + (h^2 + k^2)\varepsilon(h, k),$$

avec $\lim_{(h,k) \rightarrow (0,0)} \varepsilon(h, k) = 0$. On en déduit que, quand h tend vers 0,

$$\begin{aligned} f(x + h, y + h) &= f(x, y) + (a + b)h + \frac{1}{2}(r + s + t)h^2 + 2h^2\varepsilon(h, h) \\ &= f(x, y) + (a + b)h + \frac{1}{2}(r + s + t)h^2 + o(h^2). \end{aligned}$$

On obtient de même

$$\begin{aligned} f(x - h, y - h) &= f(x, y) - (a + b)h + \frac{1}{2}(r + s + t)h^2 + o(h^2) \\ f(x + h, y - h) &= f(x, y) + (a - b)h + \frac{1}{2}(r - s + t)h^2 + o(h^2) \\ f(x - h, y + h) &= f(x, y) + (-a + b)h + \frac{1}{2}(r - s + t)h^2 + o(h^2). \end{aligned}$$

On a ainsi, quand h tend vers 0,

$$\Delta(x, y, h) = 2sh^2 + o(h^2).$$

La fonction $h \mapsto \Delta(x, y, h)$ est la fonction nulle, donc les coefficients de son développement limité d'ordre 2 sont nuls. On en déduit $\frac{\partial^2 f}{\partial x \partial y}(x, y) = s = 0$. Comme (x, y) est quelconque, la fonction $\frac{\partial^2 f}{\partial x \partial y}$ est nulle.

Pour tout y , la fonction $x \mapsto \frac{\partial f}{\partial y}(x, y)$ a une dérivée nulle. Elle est donc constante, *i.e.* indépendante de x . Il existe donc $h : \mathbb{R} \rightarrow \mathbb{R}$ telle que, pour tout $(x, y) \in \mathbb{R}^2$,

$$\frac{\partial f}{\partial y}(x, y) = h(y).$$

Comme f est de classe \mathcal{C}^2 sur \mathbb{R}^2 , $\frac{\partial f}{\partial y}$ est de classe \mathcal{C}^1 sur \mathbb{R}^2 donc h est de classe \mathcal{C}^1 sur \mathbb{R} .

Soit ψ une primitive sur \mathbb{R} de h . La fonction $g : (x, y) \mapsto f(x, y) - \psi(y)$ vérifie $\frac{\partial g}{\partial y} = 0$, donc elle est indépendante de y . Il existe $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ telle que, pour tout $(x, y) \in \mathbb{R}^2$, $g(x, y) = \varphi(x)$ et donc

$$f(x, y) = \varphi(x) + \psi(y).$$

Comme f est de classe \mathcal{C}^2 , φ et ψ sont de classe \mathcal{C}^2 sur \mathbb{R} .

2. Supposons réciproquement qu'il existe deux fonctions φ et ψ de classe \mathcal{C}^2 sur \mathbb{R} telles que, pour tout $(x, y) \in \mathbb{R}^2$, $f(x, y) = \varphi(x) + \psi(y)$.

On a alors, pour tout $(x, y, h) \in \mathbb{R}^3$,

$$\begin{aligned} \Delta(x, y, h) &= (\varphi(x+h) + \psi(y+h)) + (\varphi(x-h) + \psi(y-h)) \\ &\quad - (\varphi(x+h) + \psi(y-h)) - (\varphi(x-h) + \psi(y+h)) = 0. \end{aligned}$$

- 11** 1. S'il existe une telle application définie sur un ouvert D inclus dans $(\mathbb{R}^*)^2$, ses dérivées partielles sont de classe \mathcal{C}^1 , donc f est de classe \mathcal{C}^2 . On a alors, pour $(x, y) \in D$,

$$\frac{\partial^2 f}{\partial y \partial x} = -\frac{2+x}{y^2} \quad \text{et} \quad \frac{\partial^2 f}{\partial x \partial y} = -\frac{2+y}{x^2}.$$

D'après le théorème de Schwarz, on a l'égalité de ces dérivées partielles, ce qui équivaut à $x^3 + 2x^2 = y^3 + 2y^2$, égalité qui n'est vérifiée sur aucun ouvert. Donc il n'existe pas une telle application f .

2. On cherche une application f définie sur l'ouvert $D = \{(x, y) \in \mathbb{R}^2, x + y \neq -1\}$. En intégrant l'égalité $\frac{\partial f}{\partial x}(x, y) = \frac{1-y}{(x+y+1)^2}$, on obtient l'existence d'une application φ de classe \mathcal{C}^2 sur \mathbb{R} telle que, pour tout $(x, y) \in D$,

$$f(x, y) = \frac{y-1}{x+y+1} + \varphi(y).$$

En dérivant par rapport à y , on obtient, pour tout $(x, y) \in D$,

$$\frac{\partial f}{\partial y}(x, y) = \frac{x+2}{(x+y+1)^2} + \varphi'(y).$$

Il faut donc que φ' soit nulle et φ constante. Les solutions sont les fonctions définies sur D par

$$f(x, y) = \frac{y-1}{x+y+1} + c,$$

où c est une constante.

3. On cherche une application f définie sur l'ouvert $D = \{(x, y) \in \mathbb{R}^2, x + y \neq 0\}$. En intégrant l'égalité $\frac{\partial f}{\partial x}(x, y) = \frac{y^2}{(x+y)^2}$, on obtient l'existence d'une application φ de classe \mathcal{C}^2 sur \mathbb{R} telle que, pour tout $(x, y) \in D$,

$$f(x, y) = \frac{-y^2}{x+y} + \varphi(y).$$

En dérivant par rapport à y , on obtient, pour tout $(x, y) \in D$,

$$\frac{\partial f}{\partial y}(x, y) = \frac{-y^2 - 2xy}{(x+y)^2} + \varphi'(y) = \frac{x^2}{(x+y)^2}.$$

On en déduit $\varphi'(y) = \frac{y^2 + 2xy + x^2}{(x + y)^2} = 1$. Il existe une constante c telle que $\varphi(y) = y + c$.

Les solutions sont les fonctions f définies sur D par

$$f(x, y) = \frac{-y^2}{x + y} + y + c = \frac{xy}{x + y} + c,$$

où c est une constante.

4. On cherche une application f définie sur l'ouvert $D = \mathbb{R} \times \mathbb{R}^*$. En intégrant l'égalité $\frac{\partial f}{\partial x}(x, y) = 2x + \frac{1}{y}$, on obtient l'existence d'une application φ de classe \mathcal{C}^2 sur \mathbb{R} telle que, pour tout $(x, y) \in D$,

$$f(x, y) = x^2 + \frac{x}{y} + \varphi(y).$$

En dérivant par rapport à y , on obtient, pour tout $(x, y) \in D$,

$$\frac{\partial f}{\partial y}(x, y) = -\frac{x}{y^2} + \varphi'(y) = 2y - \frac{x}{y^2}$$

et donc $\varphi'(y) = 2y$. Les solutions sont les fonctions f définies sur D par

$$f(x, y) = x^2 + \frac{x}{y} + y^2 + c,$$

où c est une constante.

- 12** 1. La fonction g est définie sur $\mathbb{R}_+^* \times \mathbb{R}$. Comme f est de classe \mathcal{C}^2 et que $u \mapsto uv$ est de classe \mathcal{C}^2 , pour tout réel v , la fonction $u \mapsto g(u, v)$ est de classe \mathcal{C}^2 sur \mathbb{R}_+^* . Ainsi $\frac{\partial^2 g}{\partial u^2}$ est définie sur $\mathbb{R}_+^* \times \mathbb{R}$. On obtient, pour $(x, y) \in \mathbb{R}_+^* \times \mathbb{R}$,

$$\begin{aligned} \frac{\partial g}{\partial u}(u, v) &= \frac{\partial f}{\partial x}(u, uv) + v \frac{\partial f}{\partial y}(u, uv), \quad \text{puis} \\ \frac{\partial^2 g}{\partial u^2}(u, v) &= \frac{\partial^2 f}{\partial x^2}(u, uv) + 2v \frac{\partial^2 f}{\partial x \partial y}(u, uv) + v^2 \frac{\partial^2 f}{\partial y^2}(u, uv). \end{aligned}$$

On peut démontrer plus généralement que g est de classe \mathcal{C}^2 .

2. L'application $(u, v) \mapsto (u, uv)$ est une bijection de $\mathbb{R}_+^* \times \mathbb{R}$ sur $\mathbb{R}_+^* \times \mathbb{R}$ (dont l'application réciproque est $(x, y) \mapsto \left(x, \frac{y}{x}\right)$). On en déduit que f est solution de (E) si, et seulement si, pour tout $(u, v) \in \mathbb{R}_+^* \times \mathbb{R}$,

$$u^2 \frac{\partial^2 f}{\partial x^2}(u, uv) + 2u^2 v \frac{\partial^2 f}{\partial x \partial y}(u, uv) + u^2 v^2 \frac{\partial^2 f}{\partial y^2}(u, uv) = 0,$$

c'est-à-dire si, et seulement si, pour tout $(u, v) \in \mathbb{R}_+^* \times \mathbb{R}$, $u^2 \frac{\partial^2 g}{\partial u^2}(u, v) = 0$, i.e.

$\frac{\partial^2 g}{\partial u^2}(u, v) = 0$. La fonction f est solution de (E) si, et seulement si, $\frac{\partial^2 g}{\partial u^2} = 0$.

La fonction $\frac{\partial^2 g}{\partial u^2}$ est nulle s'il existe une fonction k définie sur \mathbb{R} telle que, pour tout $(u, v) \in \mathbb{R}_+^* \times \mathbb{R}$,

$$\frac{\partial g}{\partial u}(u, v) = k(v),$$

puis une fonction ℓ définie sur \mathbb{R} telle que, pour tout $(u, v) \in \mathbb{R}_+^* \times \mathbb{R}$,

$$g(u, v) = k(v)u + \ell(v).$$

Comme g est de classe \mathcal{C}^2 sur $\mathbb{R}_+^* \times \mathbb{R}$, les fonctions k et ℓ sont de classe \mathcal{C}^2 . En effet, pour u et u' réels strictement positifs distincts, on a

$$k(v) = \frac{g(u, v) - g(u', v)}{u - u'} \quad \text{et} \quad \ell(v) = \frac{u'g(u, v) - ug(u', v)}{u' - u}.$$

3. Une solution de (E) est donc définie sur $\mathbb{R}_+^* \times \mathbb{R}$, par

$$f(x, y) = g\left(x, \frac{y}{x}\right) = xk\left(\frac{y}{x}\right) + \ell\left(\frac{y}{x}\right),$$

où k et ℓ sont des fonctions de classe \mathcal{C}^2 sur \mathbb{R} .

13 1. En posant $X = (x_1, \dots, x_n)$ et $Y = (y_1, \dots, y_n)$, on obtient

$$\varphi_{X,Y}(t) = f(tx_1 + (1-t)y_1, \dots, tx_n + (1-t)y_n).$$

Comme Ω est convexe, $tX + (1-t)Y$ appartient à Ω pour tout $t \in [0, 1]$ donc $\varphi_{X,Y}$ est définie sur $[0, 1]$. Les fonctions $t \mapsto tx_i + (1-t)y_i$ sont de classe \mathcal{C}^1 ainsi que f . On en déduit, par le théorème de composition, que $\varphi_{X,Y}$ est de classe \mathcal{C}^1 sur $[0, 1]$ et

$$\varphi'_{X,Y}(t) = \sum_{i=1}^n (x_i - y_i) \frac{\partial f}{\partial x_i}(tX + (1-t)Y) = \langle \nabla f_{tX+(1-t)Y}, X - Y \rangle.$$

2. Si f est convexe, il en est de même de $\varphi_{X,Y}$ pour tout couple $(X, Y) \in \Omega$. On sait qu'une fonction g de classe \mathcal{C}^1 sur un intervalle I de \mathbb{R} est convexe si, et seulement si, elle vérifie

$$\forall (u, t) \in I^2, \quad g(u) \geq g(t) + (u-t)g'(t).$$

On a donc ici en particulier $\varphi_{X,Y}(1) \geq \varphi_{X,Y}(0) + \varphi'_{X,Y}(0)$, c'est-à-dire

$$f(X) \geq f(Y) + \langle \nabla f_Y, X - Y \rangle.$$

3. On a par définition $\varphi_{X,Y}(u) = f(uX + (1-u)Y)$, $\varphi_{X,Y}(t) = f(tX + (1-t)Y)$. Comme Ω est convexe, $uX + (1-u)Y$ et $tX + (1-t)Y$ appartiennent à Ω et on a donc par hypothèse

$$\begin{aligned} \varphi_{X,Y}(u) &\geq \varphi_{X,Y}(t) + \langle \nabla f_{tX+(1-t)Y}, [uX + (1-u)Y] - [tX + (1-t)Y] \rangle \\ &= \varphi_{X,Y}(t) + (u-t) \langle \nabla f_{tX+(1-t)Y}, X - Y \rangle. \end{aligned}$$

Mais il résulte de la question 1 que $\varphi'_{X,Y}(t) = \langle \nabla f_{tX+(1-t)Y}, X - Y \rangle$. On a donc, pour tout $(u, t) \in [0, 1]^2$,

$$\varphi_{X,Y}(u) \geq \varphi_{X,Y}(t) + (u-t)\varphi'_{X,Y}(t).$$

Ceci est une condition nécessaire et suffisante pour que $\varphi_{X,Y}$ soit convexe sur $[0, 1]$. La fonction $\varphi_{X,Y}$ est donc convexe pour tout couple $(X, Y) \in \Omega^2$. Cela implique que f est convexe sur $[0, 1]$.

4. En reprenant les notations de la question 1 et la même méthode, on trouve que si f est de classe \mathcal{C}^2 , pour tout $(X, Y) \in \Omega^2$, la fonction $\varphi_{X,Y}$ est de classe \mathcal{C}^2 et

$$\begin{aligned} \varphi''_{X,Y}(t) &= \sum_{i=1}^n (x_i - y_i) \left(\sum_{j=1}^n (x_j - y_j) \frac{\partial^2 f}{\partial x_j \partial x_i} (tX + (1-t)Y) \right) \\ &= q_{tX+(1-t)Y}(X - Y). \end{aligned}$$

La fonction f est convexe si, et seulement si, pour tout $(X, Y) \in \Omega^2$, $\varphi_{X,Y}$ est convexe, c'est-à-dire si, pour tout $(X, Y) \in \Omega^2$ et pour tout $t \in [0, 1]$,

$$\varphi''_{X,Y}(t) = q_{tX+(1-t)Y}(X - Y) \geq 0.$$

- Si, pour tout $A \in \Omega$ et tout $H \in \mathbb{R}^n$, on a $q_A(H) \geq 0$, alors cette condition est réalisée car pour tout $(X, Y) \in \Omega^2$ et tout $t \in [0, 1]$, $tX + (1-t)Y$ appartient à Ω .
- Supposons réciproquement que, pour tout $(X, Y) \in \Omega^2$ et pour tout $t \in [0, 1]$, $q_{tX+(1-t)Y}(X - Y) \geq 0$. Soit $A \in \Omega$ et $H \in \mathbb{R}^n$. Comme Ω est ouvert, il existe $r > 0$ tel que $B(A, r) \subset \Omega$. On choisit $\lambda \in]0, 1]$ tel que $\|\lambda H\| < r$. On a alors $B = A + \lambda H \in \Omega$. En appliquant l'hypothèse avec $X = B$, $Y = A$ et $t = 0$, on obtient $q_A(\lambda H) \geq 0$. Mais q_A étant une forme quadratique, on a $q_A(\lambda H) = \lambda^2 q_A(H) \geq 0$ et comme $\lambda \neq 0$, on en déduit $q_A(H) \geq 0$.

Chapitre 7

1. Soit $\theta \in \mathbb{R}$. On a, pour tout $t \in \mathbb{R}$,

$$\begin{aligned} f_\theta(t) &= f(t \cos \theta, t \sin \theta) = (t^2 \cos^2 \theta - t \sin \theta)(2t^2 \cos^2 \theta - t \sin \theta) \\ &= t^2(t \cos^2 \theta - \sin \theta)(2t \cos^2 \theta - \sin \theta). \end{aligned}$$

- Si $\sin \theta \neq 0$, on a $\lim_{t \rightarrow 0} (t \cos^2 \theta - \sin \theta)(2t \cos^2 \theta - \sin \theta) = \sin^2 \theta > 0$, donc pour tout assez petit, on a $(t \cos^2 \theta - \sin \theta)(2t \cos^2 \theta - \sin \theta) > 0$ et donc $f_\theta(t) > 0$. Comme $f_\theta(0) = 0$, f_θ possède un minimum local strict en 0.
 - Si $\sin \theta = 0$, on a pour tout $t \in \mathbb{R}$, $f_\theta(t) = 2t^4$ et f_θ possède encore un minimum local strict en 0.
2. On a, pour tout $x \in \mathbb{R}^*$,

$$f\left(x, \frac{3}{2}x^2\right) = -\frac{1}{4}x^4 < 0.$$

On trouve dans toute boule de centre $(0, 0)$ des points de la forme $\left(x, \frac{3}{2}x^2\right)$ donc f ne possède pas de minimum local en $(0, 0)$.

- 2 1. On cherche les points critiques. On obtient le système

$$\begin{cases} \frac{\partial f}{\partial x}(x, y) = 2x + y + 2 = 0 \\ \frac{\partial f}{\partial y}(x, y) = x + 2y - 2 = 0, \end{cases}$$

qui donne $(x, y) = (-2, 2)$. On calcule alors les dérivées partielles secondes. On trouve $r = 2$, $s = 1$ et $t = 2$. Comme $rt - s^2 = 3 > 0$ et $r > 0$, on a un minimum local strict en $(-2, 2)$.

2. On cherche les points critiques. On obtient le système

$$\begin{cases} \frac{\partial f}{\partial x}(x, y) = x^2 y^2 (-4x - 3y + 3) = 0 \\ \frac{\partial f}{\partial y}(x, y) = x^3 y (-2x - 3y + 2) = 0. \end{cases}$$

Les points critiques sont les points $(0, y_0)$, y_0 réel quelconque, les points $(x_0, 0)$, x_0 quelconque, et la solution du système $\begin{cases} -4x - 3y + 3 = 0 \\ -2x - 3y + 2 = 0, \end{cases}$ qui est $\left(\frac{1}{2}, \frac{1}{3}\right)$.

- On a $f(0, y_0) = 0$ et comme la fonction $x \mapsto f(x, y)$ est impaire, dans toute boule de centre $(0, y_0)$, f prend des valeurs positives et négatives. Il n'y a pas d'extremum en $(0, y_0)$.
- On a $f(x_0, 0) = 0$. Si $x_0 \neq 0$ et $x_0 \neq 1$, on a

$$\lim_{(x, y) \rightarrow (x_0, 0)} x^3(1 - x - y) = x_0^3(1 - x_0) \neq 0.$$

Il existe une boule de centre $(x_0, 0)$ sur laquelle $x^3(1 - x - y)$ et donc $f(x, y)$ garde un signe constant. La fonction présente donc un extremum qui est un minimum si $x_0^3(1 - x_0) > 0$ et un maximum si $x_0^3(1 - x_0) < 0$.

- En $(1, 0)$, la fonction ne possède pas d'extremum, pour $h \in]0, 1[$ $f(1, h) = -h^3 < 0$ et $f(1, -h) = h^3 > 0$.

- En $\left(\frac{1}{2}, \frac{1}{3}\right)$, on calcule les dérivées partielles secondes. On trouve : $r = -\frac{1}{9}$, $s = -\frac{1}{12}$ et $t = -\frac{1}{8}$. Comme $rt - s^2 = \frac{1}{144} > 0$ et $r < 0$, la fonction présente en $\left(\frac{1}{2}, \frac{1}{3}\right)$ un maximum strict.

3. On a, pour tout $(x, y) \in \mathbb{R}^2$,

$$\frac{\partial f}{\partial x}(x, y) = 3x^2 + 3y^2 + 15 > 0.$$

La fonction f n'a pas de point critique sur \mathbb{R}^2 , donc pas de d'extremum.

4. On cherche les points critiques, en écrivant

$$\begin{cases} \frac{\partial f}{\partial x}(x, y) = 4x^3 - 4x + 4y = 0 \\ \frac{\partial f}{\partial y}(x, y) = 4y^3 + 4x - 4y = 0. \end{cases}$$

On en déduit $y^3 = -x^3$ et donc $y = -x$ puis $4x^3 - 8x = 0$. On a donc trois points critiques $(0, 0)$, $(\sqrt{2}, -\sqrt{2})$ et $(-\sqrt{2}, \sqrt{2})$.

On cherche les dérivées partielles secondes aux points critiques.

- En $(\sqrt{2}, -\sqrt{2})$ et $(-\sqrt{2}, \sqrt{2})$, on trouve $r = t = 20$, $s = 4$. Comme $rt - s^2 > 0$ et $r > 0$, on a un minimum relatif strict.

- En $(0, 0)$, on trouve $r = t = -4$ et $s = -4$. Comme $rt - s^2 = 0$, on ne peut pas conclure. Mais on remarque que

$$f(x, y) = x^4 + y^4 - 2(x - y)^2.$$

On en déduit que, pour tout réel $x \in \mathbb{R}^*$, $f(x, x) = 2x^4 > 0$, alors que

$$f(x, 0) = x^4 - 2x^2 = x^2(x^2 - 2) < 0$$

si $0 \leq |x| < \sqrt{2}$. Dans toute boule de centre $(0, 0)$, il y a des points où f est positive, d'autres où elle est négative : f n'a pas d'extremum en $(0, 0)$.

3 Les points critiques vérifient le système

$$\begin{cases} \frac{\partial f}{\partial x}(x, y, z) = e^{x(y^2+z^2+1)} (1 + (x+z^2)(y^2+z^2+1)) = 0 \\ \frac{\partial f}{\partial y}(x, y, z) = e^{x(y^2+z^2+1)} (x+z^2)(2xy) = 0 \\ \frac{\partial f}{\partial z}(x, y, z) = e^{x(y^2+z^2+1)} (2z + (x+z^2)2zx) = 0 \end{cases}$$

Notons que $x + z^2 \neq 0$ et $x \neq 0$ sinon la première équation n'est pas vérifiée. On a donc $y = 0$ d'après la deuxième équation. Les deux autres équations deviennent

$$\begin{cases} 1 + (x + z^2)(z^2 + 1) = 0 \\ z(1 + (x + z^2)x) = 0. \end{cases}$$

Si $z \neq 0$, on tire des deux équations $x = z^2 + 1$, puis $1 + (2z^2 + 1)(z^2 + 1) = 0$, ce qui est impossible. On a donc $z = 0$ et $x = -1$. Il n'y a qu'un point critique $(-1, 0, 0)$.

On a $f(-1, 0, 0) = -e^{-1}$ et, pour tout $(x, y, z) \in \mathbb{R}^3$,

$$f(x, y, z) \geq xe^{x(y^2+z^2+1)}.$$

Si $x \geq 0$, alors $f(x, y, z) \geq 0$, sinon on peut écrire $y^2 + z^2 + 1 \geq 1$, puis $e^{x(y^2+z^2+1)} \leq e^x$ et enfin

$$f(x, y, z) \geq xe^x.$$

Une rapide étude de fonction montre que, pour $x \leq 0$, on a $xe^x \geq -e^{-1}$, avec égalité si $x = -1$. Finalement, on obtient, pour tout $(x, y, z) \in \mathbb{R}^3$,

$$f(x, y, z) \geq f(-1, 0, 0).$$

La fonction f possède en $(-1, 0, 0)$ un minimum absolu.

4 Les fonctions considérées dans cet exercice sont de classe \mathcal{C}^2 sur leur ensemble de définition. Si (x_0, y_0) est un point critique de f , le plan tangent a pour équation $z = f(x_0, y_0)$.

1. On a, pour $(x, y) \in \mathbb{R}_+^* \times \mathbb{R}$,

$$\frac{\partial f}{\partial x}(x, y) = 2 \ln x + (\ln x)^2 + y^2 \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) = 2xy.$$

Les points critiques vérifient $2 \ln x + (\ln x)^2 + y^2 = xy = 0$. On obtient $y = 0$ et $\ln x(2 + \ln x) = 0$, soit $x = 1$ ou $x = e^{-2}$. Il y a deux points critiques : $(1, 0)$ et $(e^{-2}, 0)$.

Comme $f(1, 0) = 0$ et $f(x, y) \geq 0$, pour tout $(x, y) \in \mathbb{R}_+^* \times \mathbb{R}$, le graphe de f est au dessus de son plan tangent au voisinage du point $(1, 0)$, car ce plan tangent a pour équation $z = f(1, 0) = 0$.

Pour étudier le cas du point $(e^{-2}, 0)$, on calcule les dérivées secondes

$$\frac{\partial^2 f}{\partial x^2}(x, y) = \frac{2}{x}(1 + \ln x), \quad \frac{\partial^2 f}{\partial x \partial y}(x, y) = 2y, \quad \frac{\partial^2 f}{\partial y^2}(x, y) = 2x.$$

On en déduit $r = -2e^2$, $s = 0$, $t = e^{-2}$ et $rt - s^2 = -2$. On a un point selle en $(e^{-2}, 0)$.

2. On a, pour tout $(x, y) \in \mathbb{R}^2$,

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= y^3((1 + 2y)2x + 9x^2) = xy^3(9x + 4y + 2), \\ \frac{\partial f}{\partial y}(x, y) &= x^2((1 + 3x)3y^2 + 8y^3) = x^2y^2(9x + 8y + 3). \end{aligned}$$

Les points critiques vérifient $x = 0$ ou $y = 0$ ou le système $\begin{cases} 9x + 4y + 2 = 0 \\ 9x + 8y + 3 = 0 \end{cases}$, ce qui

donne le point $\left(-\frac{1}{9}, -\frac{1}{4}\right)$.

On calcule les dérivées secondes

$$\frac{\partial^2 f}{\partial x^2}(x, y) = y^3(18x + 4y + 2), \quad \frac{\partial^2 f}{\partial x \partial y}(x, y) = xy^2(27x + 16y + 6)$$

$$\frac{\partial^2 f}{\partial y^2}(x, y) = x^2y(18x + 24y + 6).$$

Au point $A = \left(-\frac{1}{9}, -\frac{1}{4}\right)$, on trouve $r = \frac{1}{64}$, $s = \frac{1}{144}$ et $t = \frac{1}{162}$. Comme $rt - s^2 = \frac{1}{20736} > 0$ et $r > 0$, le graphe est au dessus de son plan tangent pour (x, y) proche de A .

Soit $x_0 \neq -\frac{1}{3}$. On a $f(x_0, 0) = 0$ et pour (x, y) proche de $(x_0, 0)$, $f(x, y)$ a le signe de $y^3(1 + 3x_0)$, i.e. le signe de $y(1 + 3x_0)$: la position du graphe par rapport au plan tangent dépend du signe de y . Si $x_0 = -\frac{1}{3}$, la position du graphe par rapport au plan tangent au point $\left(-\frac{1}{3}, 0\right)$ dépend du signe de $y(1 + 3x + 2y)$, c'est-à-dire de la position de (x, y) par rapport à deux droites sécantes en $(x_0, 0)$.

Soit y_0 différent de 0 et $-\frac{1}{2}$. On a $f(0, y_0) = 0$ et pour (x, y) proche de $(0, y_0)$, $f(x, y)$ a

le signe de $x^2 y_0^3 (1 + 2y_0)$, *i.e.* le signe de $y_0 (1 + 2y_0)$. Selon le signe de cette expression, le graphe est au-dessus ou au-dessous du plan tangent. Si $y_0 = -\frac{1}{2}$, $f(x, y)$ a le signe de $-\frac{1}{8}(1 + 3x + 2y)$: la position du graphe par rapport au plan tangent dépend de la position de (x, y) par rapport à la droite d'équation $3x + 2y + 1 = 0$.

3. On a, pour $(x, y) \in \left] -\frac{\pi}{2}, \frac{\pi}{2} \right[$,

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= \sin y (\cos x \sin(x + y) + \sin x \cos(x + y)) = \sin y \sin(2x + y), \\ \frac{\partial f}{\partial y}(x, y) &= \sin x \sin(x + 2y). \end{aligned}$$

Les coordonnées d'un point critique vérifient soit $x = y = 0$, soit $\sin(x + 2y) = 0$ et $\sin(2x + y) = 0$. Alors, $x + 2y$ et $2x + y$ sont des multiples de π , ainsi que leur différence $x - y$. On a donc $x = y$ et $\sin(3x) = 0$, soit $x = \pm \frac{\pi}{3}$. Il y a trois points critiques $(0, 0)$, $\left(-\frac{\pi}{3}, -\frac{\pi}{3}\right)$ et $\left(\frac{\pi}{3}, \frac{\pi}{3}\right)$.

On a $f(0, 0) = 0$ et au voisinage de $(0, 0)$, $f(x, y)$ a le signe de $xy(x + y)$: la position d'un point du graphe par rapport au plan tangent dépend du signe de $xy(x + y)$, c'est-à-dire de la position de (x, y) par rapport à trois droites.

Pour les deux autres points, on calcule les dérivées secondes

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2}(x, y) &= 2 \sin y \cos(2x + y), \quad \frac{\partial^2 f}{\partial y^2}(x, y) = 2 \sin x \cos(x + 2y), \\ \frac{\partial^2 f}{\partial x \partial y}(x, y) &= \cos y \sin(2x + y) + \sin y \cos(2x + y) = \sin(2x + 2y). \end{aligned}$$

Au point $\left(-\frac{\pi}{3}, -\frac{\pi}{3}\right)$, on obtient $r = \sqrt{3} > 0$, $s = \frac{\sqrt{3}}{2}$, $t = \sqrt{3}$, $rt - s^2 = \frac{9}{4} > 0$. Le graphe de f est au-dessus du plan tangent pour (x, y) proche de $\left(-\frac{\pi}{3}, -\frac{\pi}{3}\right)$.

Au point $\left(\frac{\pi}{3}, \frac{\pi}{3}\right)$, on obtient $r = -\sqrt{3} < 0$, $s = -\frac{\sqrt{3}}{2}$, $t = -\sqrt{3}$, $rt - s^2 = \frac{9}{4} > 0$. Le graphe de f est au-dessous du plan tangent pour (x, y) proche de $\left(-\frac{\pi}{3}, -\frac{\pi}{3}\right)$.

- 5 1. La fonction f est de classe \mathcal{C}^2 sur \mathbb{R}^2 et on a, pour tout couple (x, y) ,

$$\frac{\partial f}{\partial x}(x, y) = 3(x^2 - 1 - y^2) \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) = -6xy.$$

Les points critiques vérifient $x^2 - 1 - y^2 = xy = 0$ et donc $(x, y) = (-1, 0)$ ou $(1, 0)$. On calcule les dérivées secondes aux points critiques. On obtient

$$\frac{\partial^2 f}{\partial x^2}(x, y) = 6x, \quad \frac{\partial^2 f}{\partial x \partial y}(x, y) = -6y, \quad \frac{\partial^2 f}{\partial y^2}(x, y) = -6x.$$

On obtient, au point $(-1, 0)$, $r = -6$, $s = 0$, $t = 6$. On a donc $rt - s^2 < 0$. Il n'y a pas d'extremum en $(-1, 0)$. De même, en $(1, 0)$, $r = 6$, $s = 0$, $t = -6$ et $rt - s^2 < 0$ et il n'y a pas d'extremum en $(1, 0)$.

La fonction f ne possède pas d'extremum local sur \mathbb{R}^2 .

2. a) L'ensemble D est fermé et borné. La fonction continue f possède donc un minimum m et un maximum M sur D . Si m ou M est atteint en un point de la boule ouverte $\{(x, y) \in \mathbb{R}^2, x^2 + y^2 < 1\}$, ce point est un point critique de f . Cela est impossible car f ne possède pas de point critique sur cette boule ouverte d'après la question 1. Donc m et M sont atteints en un point de $S = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 = 1\}$.
- b) Tout point de S ayant des coordonnées de la forme $(\cos t, \sin t)$, avec $t \in [-\pi, \pi]$, pour étudier f sur S , il faut étudier $g : t \mapsto f(\cos t, \sin t)$. On obtient

$$g(t) = \cos^3 t - 3 \cos t(1 + \sin^2 t) = 4 \cos^3 t - 6 \cos t.$$

La fonction g est paire, et on a, pour tout $t \in [0, \pi]$,

$$g'(t) = -12 \cos^2 t \sin t + 6 \sin t = 6 \sin t(1 - 2 \cos^2 t).$$

La dérivée s'annule et change de signe si $\cos t = \pm \frac{1}{\sqrt{2}}$ soit $t = \frac{\pi}{4}$ ou $\frac{3\pi}{4}$; de plus $g'(t) \geq 0$ si $|\cos t| \leq \frac{1}{\sqrt{2}}$, c'est-à-dire si $t \in \left[\frac{\pi}{4}, \frac{3\pi}{4}\right]$. On obtient le tableau de variation suivant

t	0	$\frac{\pi}{4}$	$\frac{3\pi}{4}$	π			
$g'(t)$	-	0	+	0	-		
$g(t)$	-2	\searrow	$-2\sqrt{2}$	\nearrow	$2\sqrt{2}$	\searrow	2

On a donc $m = -2\sqrt{2}$ et $M = 2\sqrt{2}$.

- 6 1. a) On a, pour $(x, y) \neq (0, 0)$, $f(x, y) = e^{x \ln(x^2 + y^2)}$ et f est continue sur $\mathbb{R}^2 \setminus \{(0, 0)\}$ comme composée de fonction continue. De l'inégalité

$$|x \ln(x^2 + y^2)| \leq \|(x, y)\| \ln \|(x, y)\|^2,$$

valable pour $(x, y) \neq (0, 0)$ et de $\lim_{t \rightarrow 0} t \ln t^2 = 0$, on déduit $\lim_{(x,y) \rightarrow (0,0)} x \ln(x^2 + y^2) = 0$, puis $\lim_{(x,y) \rightarrow (0,0)} f(x, y) = e^0 = 1$ et la continuité de f en $(0, 0)$. La fonction f est donc continue sur \mathbb{R}^2 .

- b) La fonction f possède des dérivées partielles d'ordre 1 sur $\mathbb{R}^2 \setminus \{(0, 0)\}$ définies par

$$\frac{\partial f}{\partial x}(x, y) = \left(\ln(x^2 + y^2) + \frac{2x^2}{x^2 + y^2} \right) f(x, y) \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) = \frac{2xy}{x^2 + y^2} f(x, y).$$

La fonction partielle $y \mapsto f(0, y)$ est constante, égale à 1, donc $\frac{\partial f}{\partial y}(0, 0) = 0$. Pour $x \neq 0$, on a

$$\frac{f(x, 0) - f(0, 0)}{x} = \frac{e^{x \ln x^2} - 1}{x} \underset{x \rightarrow 0}{\sim} \frac{x \ln x^2}{x} \underset{x \rightarrow 0}{\sim} \ln x^2$$

et donc $\lim_{x \rightarrow 0} \frac{f(x, 0) - f(0, 0)}{x} = -\infty$. La fonction $\frac{\partial f}{\partial x}$ n'est pas définie en $(0, 0)$.

2. a) On a $h(0) = 0$ et pour $x \neq 0$, $h(x) = e^{x \ln x^2} - 1$. Pour $x \in \mathbb{R}^*$, on obtient

$$h'(x) = (\ln x^2 + 2)e^{x \ln x^2}.$$

On en déduit que $h'(x) \leq 0$ si $\ln x^2 \leq -2$, c'est-à-dire $|x| \leq e^{-1}$ et $h'(x) > 0$ si $|x| > e^{-1}$. Comme de plus h est continue en 0, h décroît sur $[-e^{-1}, e^{-1}]$ et croît sur $] -\infty, -e^{-1}]$ et $[e^{-1}, +\infty[$.

- b) La fonction f est monotone sur $[-e^{-1}, e^{-1}]$ et $h(0) = 0$, donc h change de signe en 0. On en déduit

$$f(x, 0) > f(0, 0) \text{ si } x \in [-e^{-1}, 0[\text{ et } f(x, 0) < f(0, 0) \text{ si } x \in]0, e^{-1}],$$

donc f n'admet pas d'extremum en $(0, 0)$.

3. D'après le calcul des dérivées partielles, un point critique de $\mathbb{R}^2 \setminus \{(0, 0)\}$ vérifie

$$\ln(x^2 + y^2) + \frac{2x^2}{x^2 + y^2} = \frac{2xy}{x^2 + y^2} = 0.$$

On a donc soit $x = 0$ et $\ln y^2 = 0$, ce qui donne les points $(0, -1)$ et $(0, 1)$, soit $y = 0$ ce qui donne $\ln x^2 + 2 = 0$ et les points $(-e^{-1}, 0)$ et $(e^{-1}, 0)$.

4. a) Si $x > 0$, on a pour tout réel y , $\ln(x^2 + y^2) \geq \ln x^2$, $x \ln(x^2 + y^2) \geq x \ln x^2$ et $f(x, y) \geq f(x, 0)$, inégalité qui reste vérifiée si $x = 0$. Comme $h(x) + 1 = f(x, 0)$, on a, pour tout $x \geq 0$,

$$f(x, y) \geq h(x) + 1.$$

- b) Il résulte de l'étude des variations de h que h possède sur \mathbb{R}_+ un minimum en e^{-1} . Comme $h(e^{-1}) + 1 = f(e^{-1}, 0)$, on en déduit que, pour $x \geq 0$ et $y \in \mathbb{R}$,

$$f(x, y) \geq h(x) + 1 \geq f(e^{-1}, 0).$$

Sur la boule ouverte de centre $(e^{-1}, 0)$ et de rayon e^{-1} , on a $f(x, y) \geq f(e^{-1}, 0)$, donc f admet en $(e^{-1}, 0)$ un minimum local.

5. a) Pour tout $x \in \mathbb{R}^*$, on a

$$f(x, 1) - f(0, 1) = e^{x \ln(x^2+1)} - 1.$$

Cette expression a le signe de $x \ln(1 + x^2)$ et donc le signe de x .

- b) Ce qui précède montre que f ne possède pas d'extremum en $(0, 1)$ car toute boule de centre $(0, 1)$ contient des points de la forme $(x, 1)$ avec $x > 0$ et des points $(x, 1)$ avec $x < 0$.

► **Remarque**

On montrerait de même que f possède un maximum local en $(-e^{-1}, 0)$ et n'a pas d'extremum local en $(0, -1)$.

- 7 1. Si $A = (a_{ij})_{1 \leq i, j \leq n}$, on a pour $x = (x_1, \dots, x_n) \in \mathbb{R}^n \setminus \{0\}$,

$$\varphi(x) = \frac{\sum_{1 \leq i, j \leq n} a_{ij} x_i x_j}{\sum_{i=1}^n x_i^2}$$

et φ est continue car quotient de fonctions polynomiales. L'ensemble S est fermé et borné, donc φ est bornée sur S et atteint ses bornes. Il existe en particulier $x_0 \in S$ où φ atteint son maximum.

2. Si x est un élément quelconque de $\mathbb{R}^n \setminus \{0\}$, on pose $z = \frac{1}{\|x\|}x$. Alors $z \in S$ et le vecteur-colonne des coordonnées de z est $Z = \frac{1}{\|x\|}X$. On en déduit que

$$\varphi(x) = \frac{{}^t X A X}{\|x\|^2} = \frac{\|x\|^2 {}^t Z A Z}{\|x\|^2} = {}^t Z A Z = \varphi(z) \leq \varphi(x_0).$$

Cela montre que

$$\varphi(x_0) = \max_{x \in \mathbb{R}^n \setminus \{0\}} \varphi(x).$$

3. La fonction φ est de classe \mathcal{C}^1 , car quotient de fonctions polynomiales qui sont de classe \mathcal{C}^1 . Comme la fonction φ , qui est définie sur l'ouvert $\mathbb{R}^n \setminus \{0\}$, possède un extremum en x_0 , x_0 est un point critique de φ .

Pour $1 \leq i \leq n$, on a

$$\frac{\partial \varphi}{\partial x_i}(x) = \frac{1}{\|x\|^2} 2 \sum_{j=1}^n a_{ij} x_j - \frac{2x_i}{\|x\|^4} \sum_{1 \leq i, j \leq n} a_{ij} x_i x_j.$$

En notant X le vecteur-colonne des coordonnées de x et $Y = AX = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, on obtient

pour un point de S ,

$$\frac{\partial \varphi}{\partial x_i}(x) = 2y_i - 2x_i \varphi(x).$$

Si $x \in S$ est un point critique de φ , on a donc, pour $1 \leq i \leq n$, $y_i = \varphi(x)x_i$, c'est-à-dire $Y = \varphi(x)X$: le vecteur x est vecteur propre de A pour la valeur propre $\varphi(x)$. En particulier x_0 est vecteur propre de A . Cela montre que toute matrice symétrique réelle possède au moins une valeur propre.

- 8 1. L'ensemble U est l'intersection des deux demi-espaces ouverts $x_1 - x_2 < 0$ et $x_2 - x_3 < 0$ qui sont des ensembles ouverts et convexes. Leur intersection U est donc un ouvert convexe.

2. a) On a, pour $x \in \mathbb{R} \setminus \{a_1, a_2, a_3\}$,

$$\ln |P(x)| = \sum_{j=1}^3 \ln |x - a_j| \text{ et donc } \frac{P'(x)}{P(x)} = \sum_{j=1}^3 \frac{1}{x - a_j}.$$

- b) Comme la fonction P est de classe \mathcal{C}^2 , on peut appliquer la formule de Taylor-Young à l'ordre 2 en a_i :

$$\begin{aligned} P(x) &= P(a_i) + (x - a_i)P'(a_i) + \frac{1}{2}(x - a_i)^2P''(a_i) + o((x - a_i)^2) \\ P'(x) &= P'(a_i) + (x - a_i)P''(a_i) + o(x - a_i), \\ (x - a_i)P'(a_i) &= (x - a_i)P'(a_i) + (x - a_i)^2P''(a_i) + o((x - a_i)^2). \end{aligned}$$

On a donc

$$\begin{aligned} (x - a_i)P'(x) - P(x) &= \frac{1}{2}(x - a_i)^2P''(a_i) + o((x - a_i)^2) \\ (x - a_i)P(x) &= (x - a_i)^2P'(a_i) + o((x - a_i)^2). \end{aligned}$$

On en déduit que

$$\begin{aligned} \frac{P'(x)}{P(x)} - \frac{1}{x - a_i} &= \frac{(x - a_i)P'(x) - P(x)}{(x - a_i)P(x)} \\ &= \frac{(x - a_i)^2P''(a_i) + o((x - a_i)^2)}{2(x - a_i)^2P'(a_i) + o((x - a_i)^2)} \\ &= \frac{P''(a_i) + o(1)}{2P'(a_i) + o(1)}, \end{aligned}$$

puis que

$$\lim_{x \rightarrow a_i} \left(\frac{P'(x)}{P(x)} - \frac{1}{x - a_i} \right) = \frac{P''(a_i)}{2P'(a_i)}.$$

Comme pour $x \in \mathbb{R} \setminus \{a_1, a_2, a_3\}$,

$$\frac{P'(x)}{P(x)} - \frac{1}{x - a_i} = \sum_{\substack{j=1 \\ j \neq i}}^3 \frac{1}{x - a_j},$$

on obtient, en faisant tendre x vers a_i ,

$$\sum_{\substack{j=1 \\ j \neq i}}^3 \frac{1}{a_i - a_j} = \frac{P''(a_i)}{2P'(a_i)}.$$

- c) La fonction F est de classe \mathcal{C}^2 sur U comme somme de fonctions de classe \mathcal{C}^2 . On obtient, pour $1 \leq i \leq 3$,

$$\frac{\partial F}{\partial x_i}(x) = 2x_i - 2 \sum_{\substack{j=1 \\ j \neq i}}^3 \frac{1}{x_i - x_j}.$$

Le point a est un point critique de F si, et seulement si, pour $1 \leq i \leq 3$, on a

$$a_i = \sum_{\substack{j=1 \\ j \neq i}}^3 \frac{1}{a_i - a_j} = \frac{P''(a_i)}{2P'(a_i)},$$

c'est-à-dire

$$2a_iP'(a_i) - P''(a_i) = 0.$$

La condition est donc que a_1, a_2, a_3 soient racines de $2XP' - P''$.

- d) Comme ce polynôme est de degré 3, il doit être proportionnel à P . Comme P est unitaire et que le coefficient dominant de $2XP' - P''$ est 6, on obtient finalement que a est point critique de F si, et seulement si,

$$2XP' - P'' = 6P.$$

- e) On cherche P sous la forme $P = X^3 + bX^2 + cX + d$. On trouve

$$\begin{aligned} 2XP' - P'' - 6P &= 2(3X^3 + 2bX^2 + cX) - (6X + 2b) - 6(X^3 + bX^2 + cX + d) \\ &= -2bX^2 - (6 + 4c)X - 2b - 6d. \end{aligned}$$

Il faut donc $b = 0$, $c = -\frac{3}{2}$, $d = 0$. On trouve donc

$$P = X^3 - \frac{3}{2}X,$$

polynôme dont les racines sont 0, $-\sqrt{\frac{3}{2}}$ et $\sqrt{\frac{3}{2}}$. Ainsi F possède un seul point critique

$$\left(-\sqrt{\frac{3}{2}}, 0, \sqrt{\frac{3}{2}}\right).$$

3. a) Comme U est convexe, $tx + (1-t)a$ appartient à U si $t \in [0, 1]$ et Ψ est définie sur $[0, 1]$. On écrit pour tout $t \in [0, 1]$,

$$\Psi(t) = F(tx_1 + (1-t)a_1, tx_2 + (1-t)a_2, tx_3 + (1-t)a_3).$$

Puisque F est de classe \mathcal{C}^1 , ainsi que les fonctions $t \mapsto tx_i + (1-t)a_i$, il en est de même de Ψ et par le théorème de composition,

$$\Psi'(t) = \sum_{i=1}^3 (x_i - a_i) \frac{\partial F}{\partial x_i}(tx + (1-t)a).$$

Puisque a est un point critique de F , on en déduit que

$$\Psi'(0) = \sum_{i=1}^3 (x_i - a_i) \frac{\partial F}{\partial x_i}(a) = 0.$$

Montrons que la fonction F est convexe. On déduira alors de l'exercice 13 du chapitre 4 que l'application Ψ est convexe. La fonction $t \mapsto t^2$ est convexe sur \mathbb{R} car sa dérivée seconde est positive. On en déduit que pour $1 \leq i \leq 3$, la fonction $g_i : x = (x_1, x_2, x_3) \mapsto x_i^2$ est convexe sur \mathbb{R}^3 . On a en effet, pour $(x, y) \in (\mathbb{R}^3)^2$ et $t \in [0, 1]$,

$$\begin{aligned} g_i(tx + (1-t)y) &= (tx_i + (1-t)y_i)^2 \leq tx_i^2 + (1-t)y_i^2 \\ &\leq tg_i(x) + (1-t)g_i(y). \end{aligned}$$

De même, la convexité de la fonction $-\ln$ sur \mathbb{R}_+^* (sa dérivée seconde est positive) entraîne, pour $1 \leq i < j \leq 3$, la convexité sur U de $h_{ij} : x \mapsto -\ln(x_j - x_i)$. En effet,

pour $(x, x') \in U^2$, on a

$$\begin{aligned} h_{ij}(tx + (1-t)x') &= -\ln(t(x_j - x_i) + (1-t)(x'_j - x'_i)) \\ &\leq -t \ln(x_j - x_i) - (1-t) \ln(x'_j - x'_i) \\ &\leq th_{ij}(x) + (1-t)h_{ij}(x'). \end{aligned}$$

Enfin, la fonction F est convexe sur U comme somme de fonctions convexes. On en déduit que Ψ est convexe sur $[0, 1]$.

- b)** D'après les propriétés des fonctions convexes d'une variable réelle, on a, pour tout $t \in [0, 1]$, $\Psi(t) \leq \Psi(0) + t\Psi'(0)$ et en particulier $\Psi(1) \leq \Psi(0) + \Psi'(0)$. Comme $\Psi(0) = F(a)$, $\Psi(1) = F(x)$ et $\Psi'(0) = 0$, on obtient, $F(x) \leq F(a)$.

Comme cela est vrai pour tout $x \in U$, la fonction F admet un minimum absolu en a .

- c)** Comme $a = \left(-\sqrt{\frac{3}{2}}, 0, \sqrt{\frac{3}{2}}\right)$, on obtient

$$F(a) = 2 \times \frac{3}{2} - 2 \left(2 \ln \sqrt{\frac{3}{2}} + \ln 2 \sqrt{\frac{3}{2}}\right) = 3 - \ln \frac{27}{2}.$$

9

- 1.** La fonction f est de classe \mathcal{C}^2 sur \mathbb{R}^3 . On a, pour $(x, y, z) \in \mathbb{R}^3$,

$$\frac{\partial f}{\partial x}(x, y, z) = 2x - 2y, \quad \frac{\partial f}{\partial y}(x, y, z) = -2x + z + 1, \quad \frac{\partial f}{\partial z}(x, y, z) = y - 1.$$

On trouve un seul point critique $(1, 1, 1)$.

Pour tout $(h, k, \ell) \in \mathbb{R}^3$, on a

$$f(1+h, 1+k, 1+\ell) = h^2 - 2hk + k\ell$$

et en particulier $f(1+h, 1, 1) = h^2 > 0$ et $f(1+h, 1+h, 1) = -h^2 < 0$, si $h \neq 0$. Dans toute boule de centre $(0, 0, 0)$, il existe donc des points où f est positive et d'autre où elle est négative : $(1, 1, 1)$ est un point selle. La fonction f ne possède pas d'extremum.

- 2.** Le point $A = (x, y, z) \in \mathcal{C}$ est un point critique sous la contrainte

$$\mathcal{C} \begin{cases} g_1(x, y, z) = 2x - y - 1 = 0 \\ g_2(x, y, z) = x + z - 1 = 0 \end{cases}$$

si, et seulement si, $\nabla f_A \in \text{Vect}(\nabla g_1, \nabla g_2)$. On vérifie facilement que

$$\text{Vect}(\nabla g_1, \nabla g_2) = \text{Vect}((2, -1, 0), (1, 0, 1))$$

a pour équation $x + 2y - z = 0$. Alors ∇f_A appartient à ce plan si et seulement si

$$(2x - 2y) + 2(-2x + z + 1) - (y - 1) = -2x - 3y + 2z + 3 = 0.$$

Comme de plus $A \in \mathcal{C}$, on obtient le système

$$2x - y = 1, \quad x + z = 1, \quad -2x - 3y + 2z = -3.$$

En résolvant le système, on obtient $A = \left(\frac{4}{5}, \frac{3}{5}, \frac{1}{5}\right)$.

On note \mathcal{H} le sous-espace vectoriel de \mathbb{R}^3 défini par $2x - y = 0$ et $x + z = 0$. Un point quelconque de \mathcal{C} s'écrit $X = A + H$, avec $H \in \mathcal{H}$. Il existe h tel que $H = (h, 2h, -h)$ et $X = \left(\frac{4}{5} + h, \frac{3}{5} + 2h, \frac{1}{5} - h\right)$. On obtient en développant

$$f(X) = \frac{1}{5} - 5h^2.$$

On a donc pour tout $X \in \mathcal{C}$,

$$f(X) - f(A) = -5h^2 \leq 0.$$

On a donc au point A un maximum absolu pour f sous la contrainte \mathcal{C} .

10 **1. a)** L'ensemble Δ est limité par les quatre droites d'équation $x = 0$, $y = 0$, $-2x + y + 1 = 0$ et $x - 2y + 1 = 0$.

b) L'ensemble Δ est fermé car c'est une intersection de demi-plans fermés. On a, pour $(x, y) \in \Delta$, $y \leq \frac{x+1}{2}$ et $x \leq \frac{y+1}{2}$. On en déduit que

$$x \leq \frac{y+1}{2} \leq \frac{\frac{x+1}{2} + 1}{2} \leq \frac{x+3}{4}$$

et donc $0 \leq x \leq 1$. On obtient de même $0 \leq y \leq 1$. On en déduit que

$$\|(x, y)\| \leq \sqrt{2}.$$

Ainsi Δ est borné.

c) La fonction g étant continue sur le fermé borné Δ y possède un maximum.

d) L'ensemble Δ' qui est une intersection de demi-plans ouverts est un ouvert. Si g atteignait son maximum en un point de Δ' , celui-ci serait un point critique de g . Mais on a, pour tout $(x, y) \in \mathbb{R}^2$,

$$\frac{\partial g}{\partial x}(x, y) = 3 \neq 0 \quad \text{et} \quad \frac{\partial g}{\partial y}(x, y) = -1 \neq 0.$$

La fonction g ne possède pas de point critique donc le maximum de g sur Δ ne peut pas être atteint en un point de Δ' .

e) Le maximum est donc atteint en un point de $\Delta \setminus \Delta'$, c'est-à-dire en un point de Δ qui appartient à l'une des quatre droites qui limitent Δ . L'ensemble $\Delta \setminus \Delta'$ est la réunion des ensembles

$$\left\{ (x, 0), x \in \left[0, \frac{1}{2}\right] \right\}, \left\{ (0, y), y \in \left[0, \frac{1}{2}\right] \right\}, \left\{ (x, 2x - 1), x \in \left[\frac{1}{2}, 1\right] \right\}$$

$$\text{et} \quad \left\{ (2y - 1, y), y \in \left[\frac{1}{2}, 1\right] \right\}.$$

On a

$$g(x, 0) = 3x + 4 \leq \frac{11}{2} \quad \text{si} \quad x \in \left[0, \frac{1}{2}\right],$$

$$g(0, y) = 3y + 4 \leq \frac{11}{2} \quad \text{si} \quad y \in \left[0, \frac{1}{2}\right],$$

$$g(x, 2x - 1) = x + 5 \leq 6 \quad \text{si} \quad x \in \left[\frac{1}{2}, 1\right],$$

$$g(2y - 1, y) = y + 5 \leq 6 \quad \text{si} \quad y \in \left[\frac{1}{2}, 1\right].$$

Le maximum de g sur Δ est donc 6. Il est atteint en $(1, 1)$.

2. a) Comme $AX = \begin{pmatrix} 2x_1 - x_2 + x_3 \\ -x_1 + 2x_2 + x_4 \end{pmatrix}$, la condition $AX = B$ équivaut à

$$x_3 = -2x_1 + x_2 + 1 \quad \text{et} \quad x_4 = x_1 - 2x_2 + 1.$$

Ainsi $X \in \mathcal{C}$ si, et seulement si,

$$x_1 \geq 0, \quad x_2 \geq 0, \quad -2x_1 + x_2 + 1 \geq 0, \quad x_1 - 2x_2 + 1 \geq 0, \quad x_3 = -2x_1 + x_2 + 1$$

$$\text{et} \quad x_4 = x_1 - 2x_2 + 1,$$

c'est-à-dire si, et seulement si

$$(x_1, x_2) \in \Delta, \quad x_3 = -2x_1 + x_2 + 1 \quad \text{et} \quad x_4 = x_1 - 2x_2 + 1.$$

b) Pour tout $X \in \mathcal{C}$,

$$\begin{aligned} f(X) &= \langle X, W \rangle = 2x_1 + 4x_2 + x_3 + 3x_4 \\ &= 2x_1 + 4x_2 + (-2x_1 + x_2 + 1) + 3(x_1 - 2x_2 + 1) \\ &= 3x_1 - x_2 + 4 = g(x_1, x_2). \end{aligned}$$

Le maximum de f sur \mathcal{C} est donc égal au maximum de g sur Δ . Le maximum de g sur Δ est atteint en $(1, 1)$. Le maximum de f sur \mathcal{C} est donc atteint en $(1, 1, 0, 0)$.

11 1. L'ensemble Γ est l'intersection d'un hyperplan d'équation $g(x_1, \dots, x_n) = 1$ et des demi-espaces fermés $x_i \geq 0$ ($1 \leq i \leq n$). Il est donc fermé car intersection de fermés. D'autre part, pour tout $1 \leq i \leq n$, on a $0 \leq x_i \leq 1$ et donc $\|(x_1, \dots, x_n)\| \leq \sqrt{n}$. La fonction f est continue donc elle possède un maximum M sur le fermé borné Γ . Comme f prend des valeurs positives en certains points de Γ , la maximum n'est pas nul donc il est atteint en un point de $\Gamma \cap (\mathbb{R}_+^*)^n$.

2. On cherche les points critiques de f sur $(\mathbb{R}_+^*)^n$ sous la contrainte $g(x_1, \dots, x_n) = 1$. On a, pour $1 \leq i \leq n$ et $(x_1, \dots, x_n) \in (\mathbb{R}_+^*)^n$,

$$\frac{\partial f}{\partial x_i}(x_1, \dots, x_n) = \prod_{j \neq i} x_j^{\alpha_j} \alpha_i x_i^{\alpha_i - 1} = \frac{\alpha_i}{x_i} f(x_1, \dots, x_n) \quad \text{et} \quad \frac{\partial g}{\partial x_i}(x_1, \dots, x_n) = \alpha_i.$$

En un point critique x de f sous la contrainte $g(x_1, \dots, x_n) = 1$, ∇f_x est proportionnel à ∇g_x et donc $\left(\frac{\alpha_1}{x_1}, \dots, \frac{\alpha_n}{x_n}\right)$ proportionnel à $(\alpha_1, \dots, \alpha_n)$, c'est-à-dire $x_1 = \dots = x_n$.

Comme de plus $x \in \Gamma$, on a $x_1 = \dots = x_n = 1$ (car $\sum_{i=1}^n \alpha_i = 1$).

D'après la question 1, M est atteint en un point de l'ouvert $(\mathbb{R}_+^*)^n$. Ce point est donc un point critique de f sur $(\mathbb{R}_+^*)^n$ sous la contrainte $g(x_1, \dots, x_n) = 1$. C'est donc $(1, \dots, 1)$. Ainsi, $M = f(1, \dots, 1) = 1$.

3. L'inégalité est claire si $x_1 = \dots = x_n = 0$. Pour tout $(x_1, \dots, x_n) \in \mathbb{R}_+^n$, non nul, on pose $s = g(x_1, \dots, x_n)$ et pour $1 \leq i \leq n$, $y_i = \frac{x_i}{s}$. Alors $(y_1, \dots, y_n) \in \Gamma$ et d'après les questions précédentes on a $f(y_1, \dots, y_n) \leq 1$. Mais d'autre part,

$$\begin{aligned} f(y_1, \dots, y_n) &= \prod_{i=1}^n \left(\frac{x_i}{s} \right)^{\alpha_i} = \frac{f(x_1, \dots, x_n)}{s^{\alpha_1 + \dots + \alpha_n}} = \frac{f(x_1, \dots, x_n)}{s} \\ &= \frac{f(x_1, \dots, x_n)}{g(x_1, \dots, x_n)}. \end{aligned}$$

On a donc, pour tout $(x_1, \dots, x_n) \in \mathbb{R}_+^n$,

$$f(x_1, \dots, x_n) \leq g(x_1, \dots, x_n).$$

- 12** On cherche les points fixes de f sur $(\mathbb{R}_+^*)^n$ sous la contrainte $x_1 + \dots + x_n = n$. En un tel point le gradient de f est colinéaire à $(1, \dots, 1)$. Comme, pour tout $i \in \llbracket 1, n \rrbracket$,

$$\frac{\partial f}{\partial x_i}(x_1, \dots, x_n) = 4x_i^3,$$

il faut $4x_1^3 = \dots = 4x_n^3$, c'est-à-dire, puisque la fonction $x \mapsto 4x^3$ est injective sur \mathbb{R} , $x_1 = \dots = x_n$. Comme de plus $x_1 + \dots + x_n = n$, le seul point critique de f sous la contrainte $x_1 + \dots + x_n = n$ est $A = (1, \dots, 1)$.

La fonction f est de classe \mathcal{C}^2 et en tout point de $(\mathbb{R}_+^*)^n$, on a

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \dots, x_n) = \begin{cases} 0 & \text{si } i \neq j \\ 12x_i^2 & \text{si } i = j. \end{cases}$$

La hessienne en tout point de $(\mathbb{R}_+^*)^n$ est diagonale à n valeurs propres positives. On note \mathcal{H} l'hyperplan d'équation $x_1 + \dots + x_n = 0$. Pour tout point $X = (x_1, \dots, x_n)$ tel que $x_1 + \dots + x_n = n$, $X - A = H \in \mathcal{H}$. En appliquant l'égalité de Taylor-Lagrange à l'ordre 1, on obtient l'existence de $\theta \in]0, 1[$ tel que

$$f(A + H) = f(A) + \langle \nabla f_A, H \rangle + \frac{1}{2} q_{A+\theta H}(H) = f(A) + \frac{1}{2} q_{A+\theta H}(H) \geq f(A),$$

car $\nabla f_A \in \mathcal{H}^\perp$. Ainsi, f possède en A un minimum global sous la contrainte $x_1 + \dots + x_n$. Pour démontrer ce résultat, on aurait pu aussi appliquer deux fois l'inégalité de Cauchy-Schwarz. On obtient, pour $(x_1, \dots, x_n) \in (\mathbb{R}_+^*)^n$ tel que $x_1 + \dots + x_n = n$,

$$n^2 = \left(\sum_{i=1}^n x_i \right)^2 = \left(\sum_{i=1}^n 1 \cdot x_i \right)^2 \leq n \sum_{i=1}^n x_i^2 \leq n \sum_{i=1}^n 1 \cdot x_i^2 \leq n \sqrt{n \sum_{i=1}^n x_i^4}$$

et donc

$$f(x_1, \dots, x_n) = \sum_{i=1}^n x_i^4 \geq n \geq f(1, \dots, 1).$$

- 13** 1. Le point M étant intérieur au triangle, l'aire du triangle est égal à la somme des aires des triangles MBC , MCA et MAB , c'est-à-dire $\frac{ax}{2} + \frac{by}{2} + \frac{cz}{2}$. On a donc

$$S = ax + by + cz.$$

Réciproquement, si x , y et z sont trois réels positifs tels que $ax + by + cz = S$, on a $ax \leq S$, donc $x \leq \frac{S}{a}$, où $\frac{S}{a}$ est égal à la hauteur du triangle ABC relative au sommet A . On démontre le même résultat pour les autres côtés. On en déduit qu'il existe un point M , intérieur au triangle dont les distances aux côtés (BC) et (CA) respectivement sont x et y . La distance de M au côté (AB) est alors donnée par $\frac{S - ax - by}{c}$; c'est donc z .

2. L'ensemble $F = \{(x, y, z) \in (\mathbb{R}_+)^3, ax + by + cz = S\}$ est un fermé car intersection d'un plan et des demi-espaces fermés $x \geq 0, y \geq 0, z \geq 0$. Il est borné, car si $(x, y, z) \in F$, alors

$$x \leq \frac{S}{a}, y \leq \frac{S}{b}, z \leq \frac{S}{c} \quad \text{et donc} \quad \|(x, y, z)\| \leq \sqrt{\frac{S^2}{a^2} + \frac{S^2}{b^2} + \frac{S^2}{c^2}}.$$

La fonction g étant continue possède un maximum sur F .

3. La fonction g n'est pas identiquement nulle sur F , donc si le maximum de g sur F est atteint en (x_0, y_0, z_0) , $g(x_0, y_0, z_0) \neq 0$, ce qui équivaut à $x_0 y_0 z_0 \neq 0$. Il est donc obtenu sur l'ouvert $\Omega = (\mathbb{R}_+^*)^3$.
4. On cherche les points critiques de Ω sous la contrainte $ax + by + cz = S$. Le gradient de f en est tel point est colinéaire à (a, b, c) . Comme, pour $(x, y, z) \in \Omega$,

$$\frac{\partial g}{\partial x}(x, y, z) = \alpha x^{\alpha-1} y^\beta z^\gamma = \frac{\alpha g(x, y, z)}{x}, \quad \frac{\partial g}{\partial y}(x, y, z) = \frac{\beta g(x, y, z)}{y}$$

$$\text{et} \quad \frac{\partial g}{\partial z}(x, y, z) = \frac{\gamma g(x, y, z)}{z},$$

il faut $\left(\frac{\alpha}{x}, \frac{\beta}{y}, \frac{\gamma}{z}\right)$ colinéaire à (a, b, c) et donc (x, y, z) colinéaire à $\left(\frac{\alpha}{a}, \frac{\beta}{b}, \frac{\gamma}{c}\right)$. Il existe donc $\lambda > 0$ tel que

$$(x, y, z) = \lambda \left(\frac{\alpha}{a}, \frac{\beta}{b}, \frac{\gamma}{c}\right).$$

On a, de plus

$$ax + by + cz = \lambda(\alpha + \beta + \gamma) = S,$$

ce qui donne $\lambda = \frac{S}{\alpha + \beta + \gamma}$. Finalement il n'y a qu'un point critique

$$\frac{S}{\alpha + \beta + \gamma} \left(\frac{\alpha}{a}, \frac{\beta}{b}, \frac{\gamma}{c}\right).$$

Le maximum de f est atteint nécessairement en ce point critique et il vaut

$$\boxed{\left(\frac{S}{\alpha + \beta + \gamma}\right)^{\alpha + \beta + \gamma} \frac{\alpha^\alpha \beta^\beta \gamma^\gamma}{a^\alpha b^\beta c^\gamma}}.$$

Chapitre 8

1 a) On a, pour tout $k \in \mathbb{N}^*$,

$$P(X = k) = P(X > k - 1) - P(X > k).$$

On en déduit, par un changement d'indice,

$$\begin{aligned} \sum_{k=0}^n kP(X = k) &= \sum_{k=1}^n kP(X = k) = \sum_{k=1}^n k(P(X > k - 1) - P(X > k)) \\ &= \sum_{k=1}^n kP(X > k - 1) - \sum_{k=1}^n kP(X > k) \\ &= \sum_{k=0}^{n-1} (k+1)P(X > k) - \sum_{k=1}^n kP(X > k) \\ &= \sum_{k=0}^{n-1} P(X > k) - nP(X > n). \end{aligned}$$

b) La série $\sum kP(X = k)$ converge et, pour tout $n \in \mathbb{N}$,

$$P(X > n) = \sum_{k=n+1}^{+\infty} P(X = k).$$

On en déduit que

$$0 \leq nP(X > n) \leq \sum_{k=n+1}^{+\infty} nP(X = k) \leq \sum_{k=n+1}^{+\infty} kP(X = k).$$

Comme $\sum_{k=n+1}^{+\infty} kP(X = k)$ est le reste d'une série convergente, on a

$$\lim_{n \rightarrow +\infty} \sum_{k=n+1}^{+\infty} kP(X = k) = 0$$

et, par encadrement,

$$\lim_{n \rightarrow +\infty} nP(X > n) = 0.$$

On en déduit

$$\lim_{n \rightarrow +\infty} \sum_{k=0}^{n-1} P(X > k) = \lim_{n \rightarrow +\infty} \left(\sum_{k=0}^{n-1} kP(X = k) + P(X > n) \right) = \sum_{k=0}^{+\infty} kP(X = k) = E(X).$$

Autrement dit la série de terme général $P(X > n)$ converge et sa somme est $E(X)$.

c) D'après la question a, on a, pour tout $n \in \mathbb{N}^*$,

$$\sum_{k=0}^n kP(X = k) \leq \sum_{k=0}^{n-1} P(X > k) \leq \sum_{k=0}^{+\infty} P(X > k).$$

La série de terme général positif $kP(X = k)$ a ses sommes partielles majorées donc elle converge. Ainsi X admet une espérance et d'après la question précédente,

$$E(X) = \sum_{k=0}^{+\infty} P(X > k).$$

d) Il résulte des questions précédentes que X possède une espérance si et seulement si la série de terme général $P(X > n)$ converge et qu'alors

$$E(X) = \sum_{k=0}^{+\infty} P(X > k).$$

2. On montre comme dans la première question que

$$\begin{aligned} \sum_{k=0}^n k^2 P(X = k) &= \sum_{k=0}^n k^2 P(X = k) = \sum_{k=1}^n k^2 (P(X > k-1) - P(X > k)) \\ &= \sum_{k=0}^{n-1} (k+1)^2 P(X > k) - \sum_{k=1}^n k^2 P(X > k) \\ &= \sum_{k=0}^{n-1} (2k+1)P(X > k) - n^2 P(X > n). \end{aligned}$$

Si X possède une variance, elle possède un moment d'ordre 2, donc la série de terme général $n^2 P(X = n)$ converge et son reste $\sum_{k=n+1}^{+\infty} k^2 P(X = k)$ tend vers 0, quand n tend vers $+\infty$. Des inégalités

$$0 \leq n^2 P(X > n) \leq \sum_{k=n+1}^{+\infty} n^2 P(X = k) \leq \sum_{k=n+1}^{+\infty} k^2 P(X = k)$$

on déduit $\lim_{n \rightarrow +\infty} n^2 P(X > n) = 0$ et donc

$$\lim_{n \rightarrow +\infty} \sum_{k=0}^{n-1} (2k+1)P(X > k) = \lim_{n \rightarrow +\infty} \sum_{k=0}^n k^2 P(X = k) = E(X^2).$$

Ainsi la série de terme général $(2n+1)P(X > n)$ converge et sa somme est $E(X^2)$.

On peut montrer réciproquement que si la série de terme général $(2k+1)P(X > k)$ converge, alors les sommes partielles de la série de terme général $k^2 P(X = k)$ sont majorées donc elle converge. Cela montre que X^2 possède une espérance et X une variance.

3. a) Par définition, on a

$$X = \max(X_1, \dots, X_n).$$

La probabilité pour un oignon de fleurir une année donnée est p . Comme le fait que l'oignon ne fleurisse pas n'influe pas sur le résultat des années suivantes, la variable X_i qui est le temps d'attente de la floraison du i -ième oignon suit une loi géométrique de paramètre p . Les variables X_i et la variable X prennent leurs valeurs dans \mathbb{N}^* . Pour tout $k \in \mathbb{N}$, on a

$$\begin{aligned} P(X \leq k) &= P(\max(X_1, \dots, X_n) \leq k) \\ &= P(X_1 \leq k, \dots, X_n \leq k) = \prod_{i=1}^n P(X_i \leq k), \end{aligned}$$

car les variables X_1, \dots, X_n sont indépendantes.

Comme $P(X_i > k) = (1-p)^k$, car c'est la probabilité que l'oignon n'ait pas fleuri lors des k premières années, on a $P(X_i \leq k) = 1 - (1-p)^k$ et

$$P(X > k) = 1 - P(X \leq k) = 1 - (1 - (1-p)^k)^n.$$

- b) Comme $1-p \in]0, 1[$, on $\lim_{k \rightarrow +\infty} (1-p)^k = 0$ et

$$1 - (1 - (1-p)^k)^n \underset{k \rightarrow +\infty}{\sim} n(1-p)^k.$$

La série $\sum n(1-p)^k$ est une série géométrique convergente, donc la série de terme général $P(X > k)$ converge également. D'après la question 1, X possède une espérance et

$$\begin{aligned} E(X) &= \sum_{k=0}^{+\infty} P(X > k) = \sum_{k=0}^{+\infty} \left(1 - \sum_{i=0}^n \binom{n}{i} (-1)^i (1-p)^{ki} \right) \\ &= \sum_{k=0}^{+\infty} \left(\sum_{i=1}^n \binom{n}{i} (-1)^{i+1} (1-p)^{ki} \right). \end{aligned}$$

Par linéarité de la somme pour les séries convergentes, on obtient

$$E(X) = \sum_{i=1}^n \binom{n}{i} (-1)^{i+1} \left(\sum_{k=0}^{+\infty} (1-p)^{ki} \right) = \sum_{i=1}^n \binom{n}{i} (-1)^{i+1} \frac{1}{1 - (1-p)^i},$$

car on reconnaît des sommes de séries géométriques.

- c) On a

$$(2k+1)P(X > k) \underset{k \rightarrow +\infty}{\sim} (2k+1)n(1-p)^k \underset{k \rightarrow +\infty}{\sim} 2nk(1-p)^k.$$

On reconnaît à une constante près la dérivée de la série géométrique donc une série convergente. Par le théorème de comparaison, on obtient la convergence de la série de terme général $(2k+1)P(X > k)$ et donc l'existence de la variance de X .

2

1. a) On a, pour $n \in \mathbb{N}$ et $t \in [-1, 1]$, $|P(X = k)t^k| \leq P(X = k)$. Comme la série de terme général $P(X = k)$ est convergente, de somme 1, la série de terme général $P(X = k)t^k$ est absolument convergente pour $|t| \leq 1$, donc $\mathcal{G}_X(t)$ est défini pour $t \in [-1, 1]$.

- b) L'ensemble de définition de \mathcal{G}_X contient $[-1, 1]$ et est inclus dans \mathbb{R} (il est égal à \mathbb{R} si la variable X est finie). Si on exige que la série définissant $\mathcal{G}_X(t)$ soit absolument convergente, l'ensemble de définition de \mathcal{G}_X est symétrique par rapport à 0.
2. a) Si la série définissant $\mathcal{G}_X(t)$ est absolument convergente, on a d'après la formule de transfert,

$$\mathcal{G}_X(t) = E(t^X).$$

- b) Pour tout $t \in [-1, 1]$, on a

$$\mathcal{G}_{X+Y}(t) = E(t^{X+Y}) = E(t^X t^Y).$$

Les variables X et Y étant indépendantes, il en est de même des variables t^X et t^Y et

$$\mathcal{G}_{X+Y}(t) = E(t^X t^Y) = E(t^X)E(t^Y) = \mathcal{G}_X(t)\mathcal{G}_Y(t).$$

3. a) On a, pour tout $t \in [0, 1]$,

$$\mathcal{T}(t) = \frac{\mathcal{G}_X(t) - \mathcal{G}_X(1)}{t - 1} = \sum_{k=1}^{+\infty} p_k \frac{t^k - 1}{t - 1} = \sum_{k=1}^{+\infty} p_k (1 + t + \dots + t^{k-1}).$$

Soit t et t' deux réels tels que $0 \leq t \leq t' < 1$. On a alors, pour tout $k \in \mathbb{N}^*$,

$$1 + t + \dots + t^k \leq 1 + t' + \dots + t'^k$$

et donc $\mathcal{G}_X(t) \leq \mathcal{G}_X(t')$: la fonction \mathcal{T} est croissante sur $[0, 1[$.

- b) On a, pour tout $k \in \mathbb{N}^*$,

$$p_k (1 + t + \dots + t^{k-1}) \leq k p_k.$$

Comme X possède une espérance, la série de terme général $k p_k$ converge. Par croissance de la somme d'une série, on obtient, pour tout $t \in [0, 1[$,

$$\sum_{k=1}^{+\infty} p_k (1 + t + \dots + t^{k-1}) \leq \sum_{k=1}^{+\infty} k p_k, \text{ c'est-à-dire } \mathcal{T}(t) \leq E(X).$$

- c) La fonction \mathcal{T} est croissante sur $[0, 1[$ et majorée par $E(X)$. Elle possède donc une limite finie en 1 à gauche. Autrement dit, la fonction \mathcal{G}_X est dérivable en 1 à gauche et par passage à la limite dans l'inégalité précédente, on a

$$(\mathcal{G}_X)'_g(1) = \lim_{t \rightarrow 1^-} \mathcal{T}(t) \leq E(X).$$

- d) D'après la question a, on a, pour $n \in \mathbb{N}^*$ et $t \in [0, 1[$,

$$\sum_{k=1}^n p_k (1 + t + \dots + t^{k-1}) \leq \sum_{k=1}^{+\infty} p_k (1 + t + \dots + t^{k-1}) \leq \mathcal{T}(t).$$

Comme \mathcal{G}_X est dérivable à gauche en 1, on obtient en faisant tendre t vers 1 par valeurs inférieures dans l'inégalité précédente

$$\sum_{k=1}^n k p_k \leq (\mathcal{G}_X)'_g(1).$$

La série à termes positifs $\sum kp_k$ a ses sommes partielles majorées, donc elle converge et X possède une espérance. En faisant tendre n vers $+\infty$ dans l'inégalité précédente, on obtient

$$E(X) \leq (\mathcal{G}_X)'_g(1).$$

- e) Il résulte des questions précédentes qu'il y a équivalence entre X admet une espérance et \mathcal{G}_X est dérivable à gauche en 1. Des inégalités démontrées dans les questions b et c, on déduit que dans le cas où l'espérance existe

$$E(X) = (\mathcal{G}_X)'_g(1).$$

- f) Si on admet que la dérivée seconde de \mathcal{G}_X en 1 s'obtient en dérivant deux fois terme à terme la série définissant \mathcal{G}_X (ce qui est évident si la variable est finie), on obtient

$$(\mathcal{G}_X)''_g(1) = \sum_{k=2}^{+\infty} k(k-1)p_k = E(X(X-1)) = E(X^2) - E(X).$$

Ces espérances existent, car par hypothèse, X possède une variance, donc une espérance et un moment d'ordre 2. On en déduit

$$V(X) = E(X^2) - (E(X))^2 = (\mathcal{G}_X)''_g(1) + (\mathcal{G}_X)'_g(1) - ((\mathcal{G}_X)'_g(1))^2.$$

4. a) Pour la loi binomiale, on trouve, pour tout réel t ,

$$\begin{aligned} \mathcal{G}_X(t) &= \sum_{k=0}^n P(X=k)t^k = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} t^k \\ &= \sum_{k=0}^n \binom{n}{k} (pt)^k (1-p)^{n-k} = (pt + 1 - p)^n. \end{aligned}$$

On en déduit que, pour tout réel t ,

$$\mathcal{G}'_X(t) = np(pt + 1 - p)^{n-1} \quad \text{et} \quad \mathcal{G}''_X(t) = n(n-1)p^2(pt + 1 - p)^{n-2}.$$

On obtient

$$\begin{aligned} E(X) &= \mathcal{G}'_X(1) = np, \\ V(X) &= \mathcal{G}''_X(1) + \mathcal{G}'_X(1) - (\mathcal{G}'_X(1))^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p). \end{aligned}$$

- b) Si X suit la loi uniforme sur $\llbracket 1, n \rrbracket$ on a, pour $k \in \llbracket 1, n \rrbracket$, $P(X=k) = \frac{1}{n}$. On en déduit, pour tout réel t ,

$$\mathcal{G}_X(t) = \sum_{k=0}^n \frac{1}{n} t^k, \quad \mathcal{G}'_X(t) = \sum_{k=1}^n \frac{1}{n} k t^{k-1}, \quad \mathcal{G}''_X(t) = \sum_{k=2}^n \frac{1}{n} k(k-1) t^{k-2}.$$

On obtient

$$\begin{aligned} \mathcal{G}'_X(1) &= \sum_{k=1}^n \frac{1}{n} k = \frac{n+1}{2} \\ \mathcal{G}''_X(1) &= \frac{1}{n} \sum_{k=2}^n k(k-1) = \frac{1}{n} \sum_{k=1}^n (k^2 - k) \\ &= \frac{1}{n} \left(\frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)}{2} \right) \\ &= \frac{(n^2 - 1)}{3} \end{aligned}$$

et donc

$$E(X) = \frac{n+1}{2} \quad \text{et} \quad V(X) = \frac{n^2 - 1}{3} + \frac{n+1}{2} - \frac{(n+1)^2}{4} = \frac{n^2 - 1}{12}.$$

c) Si X suit une loi géométrique de paramètre p , on a pour tout réel t et tout entier $n \in \mathbb{N}^*$

$$P(X = n)t^n = (1-p)^{n-1}pt^n = pt(1-p)^{n-1}.$$

La série $\sum P(X = n)t^n$ converge absolument pour $|(1-p)t| < 1$ soit $|t| < \frac{1}{1-p}$. On obtient alors, pour $|t| < \frac{1}{1-p}$,

$$\begin{aligned} \mathcal{G}_X(t) &= pt \sum_{k=1}^{+\infty} ((1-p)t)^{k-1} = \frac{pt}{1 - (1-p)t}, \\ \mathcal{G}'_X(t) &= \frac{p}{(1 - (1-p)t)^2}, \quad \mathcal{G}''_X(t) = \frac{2p(1-p)}{(1 - (1-p)t)^3}. \end{aligned}$$

Comme $1 < \frac{1}{1-p}$, \mathcal{G}_X possède une dérivée et une dérivée seconde en 1 et

$$\mathcal{G}'_X(1) = \frac{1}{p}, \quad \mathcal{G}''_X(1) = \frac{2(1-p)}{p^2}.$$

On en déduit

$$E(X) = \frac{1}{p} \quad \text{et} \quad V(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

d) Si X suit une loi de Poisson de paramètre, la fonction \mathcal{G}_X est définie sur \mathbb{R} , et pour tout réel t ,

$$\mathcal{G}_X(t) = \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} t^k = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda t)^k}{k!} = e^{-\lambda} e^{\lambda t}.$$

La fonction \mathcal{G}_X est deux fois dérivable sur \mathbb{R} et pour tout réel t ,

$$\mathcal{G}'_X(t) = \lambda e^{-\lambda} e^{\lambda t} \quad \text{et} \quad \mathcal{G}''_X(t) = \lambda^2 e^{-\lambda} e^{\lambda t}.$$

On en déduit $\mathcal{G}'_X(1) = \lambda$, $\mathcal{G}''_X(1) = \lambda^2$ et donc

$$E(X) = \lambda, \quad V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

3

- 1. a)** Par définition, la variable X_1 suit la loi géométrique de paramètre p .
b) L'ensemble des valeurs prises par X_r est $[r, +\infty[\cap \mathbb{N}$.
c) Si k est un entier supérieur à r , l'événement $[X_r = k]$ est réalisé si la k -ième épreuve est un succès et si parmi les $k - 1$ épreuves précédentes, il y a exactement $r - 1$ succès et $k - r$ échecs. Comme il y a $\binom{k-1}{r-1}$ façons de répartir ces succès parmi les $k - 1$ épreuves et que les épreuves sont indépendantes, on obtient

$$P(X_r = k) = \binom{k-1}{r-1} q^{k-r} p^r.$$

- 2. a)** On a, d'après la formule donnée dans l'énoncé, en faisant un changement d'indice

$$\sum_{k=r}^{+\infty} P(X_r = k) = \sum_{k=r}^{+\infty} \binom{k-1}{r-1} q^{k-r} p^r = p^r \sum_{k=r-1}^{+\infty} \binom{k}{r-1} q^{k-r+1} = \frac{p^r}{(1-q)^r} = 1.$$

Cette probabilité est égale à la probabilité de l'événement A : « on obtient au moins un succès dans la suite des épreuves ». Comme X_r est définie sur A , cela montre que X_r est définie sur un ensemble de probabilité nulle. Donc X_r est une variable aléatoire.

- b)** La convergence de la série de terme général $kP(X_r = k)$ découle du résultat donné dans l'énoncé appliqué à l'entier $r + 1$; X_r possède donc une espérance. En utilisant l'égalité $k\binom{k-1}{r-1} = r\binom{k}{r}$, on obtient

$$E(X_r) = \sum_{k=r}^{+\infty} k \binom{k-1}{r-1} q^{k-r} p^r = p^r r \sum_{k=r}^{+\infty} \binom{k}{r} q^{k-r} = \frac{p^r r}{(1-q)^{r+1}} = \frac{r}{p}.$$

- c)** On obtient, de même, en utilisant $k(k+1)\binom{k-1}{r-1} = r(r+1)\binom{k+1}{r+1}$,

$$\begin{aligned} E(X_r(X_r + 1)) &= \sum_{k=r}^{+\infty} k(k+1) \binom{k-1}{r-1} q^{k-r} p^r \\ &= p^r r(r+1) \sum_{k=r}^{+\infty} \binom{k+1}{r+1} q^{k+1-(r+1)} \\ &= \frac{p^r r(r+1)}{(1-q)^{r+2}} = \frac{r(r+1)}{p^2}. \end{aligned}$$

On en déduit que

$$V(X_r^2) = E(X_r(X_r + 1)) - E(X_r) + (E(X_r))^2 = \frac{r(r+1)}{p^2} - \frac{r}{p} - \frac{r^2}{p^2} = \frac{rq}{p^2}.$$

- d) Toujours d'après la même formule, la série donnant $\mathcal{G}_{X_r}(t)$ converge absolument pour $|qt| < 1$ et l'on obtient

$$\begin{aligned} \mathcal{G}_{X_r}(t) &= \sum_{k=r}^{+\infty} \binom{k-1}{r-1} q^{k-r} p^r t^k = (pt)^r \sum_{k=r}^{+\infty} \binom{k-1}{r-1} (qt)^{k-r} \\ &= (pt)^r \sum_{k=r-1}^{+\infty} \binom{k}{r-1} (qt)^{k-r+1} \\ &= \frac{(pt)^r}{(1-qt)^r} = \left(\frac{pt}{1-qt} \right)^r. \end{aligned}$$

La fonction \mathcal{G}_{X_r} est deux fois dérivable sur $\left] -\frac{1}{q}, \frac{1}{q} \right[$ et

$$\begin{aligned} \mathcal{G}'_{X_r}(t) &= \frac{rp}{(1-qt)^2} \left(\frac{pt}{1-qt} \right)^{r-1} \\ \mathcal{G}''_{X_r}(t) &= \frac{2rpq}{(1-qt)^3} \left(\frac{pt}{1-qt} \right)^{r-1} + \frac{r(r-1)p^2}{(1-qt)^4} \left(\frac{pt}{1-qt} \right)^{r-2} \end{aligned}$$

On en déduit

$$\mathcal{G}'_{X_r}(1) = \frac{rp}{p^2} = \frac{r}{p}, \quad \mathcal{G}''_{X_r}(1) = \frac{2rpq}{p^3} + \frac{r(r-1)p^2}{p^4} = \frac{2rq + r(r-1)}{p^2}.$$

On obtient

$$E(X_r) = \frac{r}{p}, \quad V(X_r) = \frac{2rq + r(r-1)}{p^2} + \frac{r}{p} - \frac{r^2}{p^2} = \frac{2rq - r + rp}{p^2} = \frac{rq}{p^2}.$$

3. a) On a $Y_r = X_r - r$.
 b) La variable Y_r est à valeurs dans \mathbb{N} et pour tout $n \in \mathbb{N}$,

$$P(Y_r = k) = P(X_r = k+r) = \binom{k+r-1}{r-1} q^k p^r.$$

De plus, on a

$$E(Y_r) = E(X_r - r) = E(X_r) - r = \frac{r}{p} - r = \frac{rq}{p} \quad \text{et} \quad V(Y_r) = V(X_r) = \frac{rq}{p^2}.$$

- c) On peut écrire, par définition, pour $k \in \mathbb{N}$,

$$\begin{aligned} \binom{k+r-1}{k} &= \frac{(k+r-1)(k+r-2)\dots(r+1)r}{k!} \\ &= \frac{(-1)^k}{k!} (-r)(-r-1)\dots(-r-k+1) \\ &= (-1)^k \binom{-r}{k}. \end{aligned}$$

On en déduit

$$P(Y_r = k) = \binom{k+r-1}{r-1} q^k p^r = (-1)^k \binom{-r}{k} p^r q^k = \binom{-r}{k} (-q)^k p^r.$$

d) Sous réserve de convergence, on a

$$\mathcal{G}_{Y_r}(t) = \sum_{k=0}^{+\infty} P(Y_r = k) t^k = \sum_{k=0}^{+\infty} P(X_r = k+r) t^k.$$

La fonction \mathcal{G}_{Y_r} est donc définie comme \mathcal{G}_{X_r} sur $\left] -\frac{1}{q}, \frac{1}{q} \right[$. On obtient, pour $t \neq 0$,

$$\mathcal{G}_{Y_r}(t) = \frac{1}{t^r} \sum_{k=0}^{+\infty} P(X_r = k+r) t^{k+r} = \frac{1}{t^r} \mathcal{G}_{X_r}(t) = \frac{p^r}{(1-qt)^r}.$$

Cette formule reste vraie pour $t = 0$, car $\mathcal{G}_{Y_r}(0) = P(Y_r = 0) = P(X_r = r) = p^r$.

On en déduit que, pour $t \in \left] -\frac{1}{q}, \frac{1}{q} \right[$,

$$\mathcal{G}'_{Y_r}(t) = \frac{p^r q r}{(1-qt)^{r+1}} \quad \text{et} \quad \mathcal{G}''_{Y_r}(t) = \frac{p^r q^2 r(r+1)}{(1-qt)^{r+2}}.$$

On en déduit

$$\mathcal{G}'_{Y_r}(1) = \frac{qr}{p} \quad \text{et} \quad \mathcal{G}''_{Y_r}(1) = \frac{q^2 r(r+1)}{p^2}.$$

On obtient

$$E(Y_r) = \frac{qr}{p} \quad \text{et} \quad V(Y_r) = \frac{q^2 r(r+1)}{p^2} + \frac{qr}{p} - \frac{q^2 r^2}{p^2} = \frac{qr}{p^2}.$$

4 **1. a)** Pour vider l'urne, il faut effectuer n tirages de rang impair et donc en tout $N = 2n - 1$ tirages.

b) Avant le $(2j)$ -ième tirage, *i.e.* à l'issue du $(2j-1)$ -ième tirage, on a retiré j boules. Il en reste $n-j$.

Avant le $(2j+1)$ -ième tirage, *i.e.* à l'issue du $(2j)$ -ième tirage, on a retiré j boules. Il en reste $n-j$.

2. a) On a $P(X_1 = 1) = \frac{1}{n}$ et

$$\begin{aligned} P(X_2 = 1) &= P(X_1 = 0, X_2 = 1) = P(X_1 = 0)P(X_2 = 1 | X_1 = 0) \\ &= \frac{n-1}{n} \frac{1}{n-1} = \frac{1}{n}, \end{aligned}$$

car pour tirer la boule noire au second tirage, il ne faut pas qu'elle ait été retirée de l'urne au premier tirage.

- b) Plus généralement, pour tirer la boule noire aux $(2j + 1)$ -ième tirage, il faut qu'elle n'ait pas été tirée à l'un des tirages impairs précédents, ce qui donne, en appliquant la formule des probabilités composées,

$$\begin{aligned} P(X_{2j+1} = 1) &= P(X_1 = 0, \dots, X_{2j-1} = 0, X_{2j+1} = 1) \\ &= \frac{n-1}{n} \frac{n-2}{n-1} \dots \frac{n-j}{n-j+1} \frac{1}{n-j} = \frac{1}{n}. \end{aligned}$$

On obtient, de même,

$$\begin{aligned} P(X_{2j} = 1) &= P(X_1 = 0, \dots, X_{2j-1} = 0, X_{2j} = 1) \\ &= \frac{n-1}{n} \frac{n-2}{n-1} \dots \frac{n-j}{n-j+1} \frac{1}{n-j} = \frac{1}{n}. \end{aligned}$$

- c) Toutes les variables X_k suivent la loi de Bernoulli de paramètre $\frac{1}{n}$.
3. a) Avant le $(2n - 2)$ -ième tirage, il reste une boule dans l'urne. Si on n'a pas encore tiré la boule noire, on la tire au $(2n - 2)$ -ième tirage. On en déduit que $P(U_n) = 0$. Pour $j \in \llbracket 1, n \rrbracket$, on a, en appliquant la formule des probabilités composées,

$$\begin{aligned} P(U_j) &= P(X_1 = X_2 = \dots = X_{2j-2} = 0, X_{2j} = 1) \\ &= \frac{n-1}{n} \left(\frac{n-2}{n-1} \right)^2 \dots \left(\frac{n-j+1}{n-j+2} \right)^2 \frac{n-j}{(n-j+1)^2} = \frac{n-j}{n(n-1)}, \end{aligned}$$

car la probabilité de ne pas tirer la boule noire au $(2k)$ -ième et au $(2k + 1)$ -ième tirage, si elle n'a pas encore été tirée est $\frac{n-k-1}{n-k}$.

- b) Si la première boule noire est obtenue lors d'un tirage impair, elle est retirée et la boule noire n'apparaîtra qu'une fois en tout. Si elle est obtenue lors d'un tirage pair, elle est remise et sera tirée de nouveau. On a donc $[X = 1] = U_1 \cup U_2 \dots \cup U_n$ et

$$P(X = 1) = \sum_{j=1}^n \frac{n-j}{n(n-1)} = \frac{1}{n-1} \sum_{k=0}^{n-1} k = \frac{1}{2}.$$

- c) Si l'événement $X_{2j-1} = 1$ est réalisé alors $X \leq j$, car la boule noire ne peut être tirée qu'aux tirages de rang $2, 4, \dots, 2j-2, 2j-1$. On en déduit que

$$[X = n] = [X_1 = X_3 = \dots = X_{2n-3} = 0, X_2 = \dots = X_{2n-2} = 1, X_{2n-1} = 1],$$

puis en appliquant la formule des probabilités composées,

$$P([X = n]) = \frac{n-1}{n} \frac{1}{n-1} \frac{n-2}{n-1} \frac{1}{n-2} \dots \frac{1}{2} \cdot 1 \cdot 1 = \frac{1}{n!}.$$

4. La somme est égale au nombre de tirages où l'on obtient une boule noire, c'est-à-dire à X . Par linéarité de l'espérance, on en déduit

$$E(X) = \sum_{k=1}^{2n-1} E(X_k) = \sum_{k=1}^{2n-1} \frac{1}{n} = \frac{2n-1}{n}.$$

5. a) Pour $j \in \llbracket 1, 2n - 2i - 2 \rrbracket$, on a

$$P(X_{2i+j+1} = 1 | X_{2i+1} = 1) = 0,$$

car si la boule noire est tirée au $(2i+1)$ -ième tirage, elle n'est pas remise et donc ne peut pas être tirée au $(2i+j+1)$ -ième tirage. Les variables X_{2i+1} et X_{2i+j+1} étant de Bernoulli, on a

$$\begin{aligned} E(X_{2i+1}X_{2i+j+1}) &= P(X_{2i+1} = 1, X_{2i+j+1} = 1) \\ &= P(X_{2i+j+1} = 1 | X_{2i+1} = 1)P(X_{2i+1}) = 0. \end{aligned}$$

- b) On en déduit

$$\text{Cov}(X_{2i+1}, X_{2i+j+1}) = -E(X_{2i+1}, X_{2i+j+1}) = -\frac{1}{n^2}.$$

6. a) Pour $k \in \llbracket 1, n - i - 1 \rrbracket$, on obtient

$$P(X_{2i+2k} = 1 | X_{2i} = 1) = \frac{n-i-1}{n-i} \frac{n-i-2}{n-i-1} \cdots \frac{n-i-k}{n-i-k+1} \frac{1}{n-i-k} = \frac{1}{n-i},$$

car il faut obtenir des boules blanches aux tirages de rang impair entre le $(2i+1)$ -ième et le $(2i+2k-1)$ -ième et une boule noire au $(2i+2k)$ -ième.

- b) On obtient de même

$$P(X_{2i+2k+1} = 1 | X_{2i} = 1) = \frac{n-i-1}{n-i} \frac{n-i-2}{n-i-1} \cdots \frac{n-i-k}{n-i-k+1} \frac{1}{n-i-k} = \frac{1}{n-i}.$$

- c) Des questions a et b, il résulte que, pour $j \in \llbracket 1, 2n - 2i - 1 \rrbracket$,

$$P(X_{2i+j} = 1 | X_{2i} = 1) = \frac{1}{n-i}.$$

On en déduit que

$$E(X_{2i+j}X_{2i}) = P(X_{2i+j} = 1, X_{2i} = 1) = P(X_{2i+j} = 1 | X_{2i} = 1)P(X_{2i} = 1) = \frac{1}{n(n-i)},$$

$$\text{Cov}(X_{2i+j}, X_{2i}) = E(X_{2i+j}X_{2i}) - E(X_{2i+j})E(X_{2i}) = \frac{1}{n(n-i)} - \frac{1}{n^2} = \frac{i}{n^2(n-i)}.$$

7. En appliquant la formule donnant la variance d'une somme, on obtient

$$V(X) = \sum_{k=1}^{2n-1} V(X_k) + 2 \sum_{1 \leq k < l \leq 2n-1} \text{Cov}(X_k, X_l).$$

Les variables X_k suivent la loi de Bernoulli de paramètre $\frac{1}{n}$ donc

$$V(X_k) = \frac{1}{n} \left(1 - \frac{1}{n} \right) = \frac{n-1}{n^2}.$$

D'après la question 5, $\text{Cov}(X_{2i+1}, X_l) = \frac{-1}{n^2}$ si $2i + 1 < l$; il y a alors $2n - 2i - 2$ telles valeurs de l et

$$\begin{aligned} 2 \sum_{2i+1 < l \leq n} \text{Cov}(X_{2i+1}, X_l) &= -\frac{2}{n^2} \sum_{i=0}^{n-2} ((2n-2) - 2i) \\ &= -\frac{2}{n^2} (2(n-1)^2 - (n-1)(n-2)) \\ &= -\frac{2(n-1)}{n}. \end{aligned}$$

D'après la question 6, $\text{Cov}(X_{2i}, X_l) = \frac{i}{n^2(n-i)}$ si et $2i < l$ et il y a $(2n - 2i - 1)$ telles valeurs de l et

$$\begin{aligned} 2 \sum_{2i+1 < l \leq n} \text{Cov}(X_{2i}, X_l) &= 2 \sum_{i=1}^{n-1} (2n - 2i - 1) \frac{i}{n^2(n-i)} = \frac{2}{n^2} \sum_{j=1}^{n-1} \frac{(n-j)(2j-1)}{j} \\ &= \frac{2}{n^2} \sum_{j=1}^{n-1} \frac{-n + (2n+1)j - 2j^2}{j} \\ &= -\frac{2}{n} \sum_{j=1}^{n-1} \frac{1}{j} + \frac{2(2n+1)(n-1)}{n^2} - \frac{2(n-1)}{n} \\ &= -\frac{2}{n} \sum_{j=1}^n \frac{1}{j} + \frac{2(n-1)(n+1)}{n^2}. \end{aligned}$$

On en déduit

$$\begin{aligned} V(X) &= \frac{(2n-1)(n-1)}{n^2} - \frac{2(n-1)}{n} - \frac{2}{n} \sum_{j=1}^n \frac{1}{j} + \frac{2(n-1)(n+1)}{n^2} \\ &= \frac{(2n+1)(n-1)}{n^2} - \frac{2}{n} \sum_{j=1}^n \frac{1}{j}. \end{aligned}$$

5 **1. a)** Soit $k \geq 2$. Si P_{k-1} et P_k sont réalisés, alors S_k est réalisé si S_{k-1} ne l'est pas, donc $P_{k-1} \cap P_k \subset S_{k-1} \cup S_k$. On en déduit

$$P_{k-1} \cap P_k = (P_{k-1} \cap P_k \cap S_{k-1}) \cup (P_{k-1} \cap P_k \cap S_k)$$

et la réunion précédente est disjointe, car par définition, S_{k-1} et S_k le sont.

- On a $S_k \subset P_{k-1} \cap P_k$ et donc $P(P_{k-1} \cap P_k \cap S_k) = P(S_k) = a_k$.
- D'autre part, on a $S_{k-1} \subset P_{k-1}$ et P_k est indépendant de S_{k-1} , car S_{k-1} ne dépend que de ce qui s'est passé avant le k -ième lancer. On en déduit que

$$P(P_{k-1} \cap P_k \cap S_{k-1}) = P(P_k \cap S_{k-1}) = P(P_k)P(S_{k-1}) = \alpha a_{k-1}.$$

Finalement, on obtient, pour tout $k \geq 2$,

$$\alpha^2 = P(P_{k-1} \cap P_k) = a_k + \alpha a_{k-1}$$

- b) La suite (a_k) est donc arithmético-géométrique, de point fixe $\frac{\alpha^2}{1+\alpha}$ et de premier terme $a_1 = 0$. On obtient, pour tout $k \geq 2$,

$$a_k - \frac{\alpha^2}{\alpha+1} = -\alpha \left(a_{k-1} - \frac{\alpha^2}{\alpha+1} \right)$$

et donc, pour $k \geq 1$,

$$a_k = \frac{\alpha^2}{1+\alpha} (1 + (-1)^k \alpha^{k-1}).$$

2. On a $S_2 = P_1 \cap P_2 = G_2$ et donc $a_2 = b_2 = a_1 b_1 + b_2$, car $a_1 = 0$; $S_3 = \overline{P_1} \cap P_2 \cap P_3 = G_3$, donc $a_3 = b_3 = a_1 b_2 + a_2 b_1 + b_3$, car $a_1 = b_1 = 0$.

De manière générale, si S_k est réalisé S_{k-1} n'est pas réalisé et soit aucun des événements S_j pour $j < k$ n'est réalisé, soit il y en a moins un qui est réalisé et l'on note i le plus grand $j < k$ tel que S_j soit réalisé. Ainsi S_k est la réunion des événements incompatibles $\overline{S_1} \cap \overline{S_2} \cap \dots \cap \overline{S_{k-1}} \cap S_k = G_k$ et $S_i \cap \overline{S_{i+1}} \cap \dots \cap \overline{S_{k-1}} \cap S_k$, pour $i \in \llbracket 1, k-2 \rrbracket$. On observe que

$$\begin{aligned} P(S_i \cap \overline{S_{i+1}} \cap \dots \cap \overline{S_{k-1}} \cap S_k) &= P(S_i)P(\overline{S_{i+1}} \cap \dots \cap \overline{S_{k-1}} \cap S_k | S_i) \\ &= P(S_i)P(G_{k-i}) = a_i b_{k-i}. \end{aligned}$$

On en déduit

$$a_k = P(S_k) = b_k + \sum_{i=1}^{k-2} a_i b_{k-i} = \sum_{i=1}^{k-1} a_i b_{k-i} + b_k,$$

car $a_1 = 0$.

3. a) La série de terme général $a_k t^k$ est somme de deux séries géométriques convergentes car α et t appartiennent à $[0, 1[$, donc est elle-même convergente et

$$\begin{aligned} F(t) &= \sum_{k=2}^{+\infty} \frac{\alpha^2}{1+\alpha} (1 + (-1)^k \alpha^{k-1}) t^k = \frac{\alpha^2}{1+\alpha} \sum_{k=2}^{+\infty} t^k + \frac{\alpha}{1+\alpha} \sum_{k=2}^{+\infty} (-\alpha t)^k \\ &= \frac{\alpha^2}{1+\alpha} \cdot \frac{t^2}{1-t} + \frac{\alpha}{1+\alpha} \cdot \frac{\alpha^2 t^2}{1+\alpha t} = \frac{\alpha^2 t^2}{(1-t)(1+\alpha t)} \end{aligned}$$

- b) On a $0 \leq b_k t^k \leq t^k$ car b_k est une probabilité et, pour $t \in [0, 1[$, la convergence de la série géométrique de raison t donne la convergence de la série de terme général $b_k t^k$. De l'égalité

$$a_k t^k = b_k t^k + \sum_{i=1}^{k-1} a_i b_{k-i} t^k = b_k t^k + \sum_{i=1}^{k-1} (a_i t^i) (b_{k-i} t^{k-i}),$$

on tire, grâce au résultat admis, pour tout $t \in [0, 1[$,

$$F(t) = H(t) + F(t)H(t).$$

- c) On obtient, pour $t \in [0, 1[$,

$$H(t) = \frac{F(t)}{1+F(t)} = \frac{\alpha^2 t^2}{(1-t)(1+\alpha t) + \alpha^2 t^2}.$$

4. On a $\lim_{t \rightarrow 1^-} H(t) = 1$.

5. Les résultats admis permettent de sommer les séries de termes généraux b_k , $k b_k$ et $k(k-1)b_k$.

a) On obtient

$$\sum_{k=1}^{\infty} b_k = H(1) = \lim_{t \rightarrow 1^-} H(t) = 1,$$

ce qui prouve que l'on est bien en présence d'une loi de probabilité.

b) On calcule, pour $t \in [0, 1[$,

$$H'(t) = \frac{\alpha^2 t(2 + (\alpha - 1)t)}{((1-t)(1+\alpha t) + \alpha^2 t^2)^2}, \quad H''(t) = 2\alpha^2 \frac{-\alpha(\alpha - 1)^2 t^3 + 3(\alpha - \alpha^2)t^2 + 1}{((1-t)(1+\alpha t) + \alpha^2 t^2)^3},$$

d'où l'on tire

$$\lim_{t \rightarrow 1^-} H'(t) = \frac{1 + \alpha}{\alpha^2} \quad \text{et} \quad \lim_{t \rightarrow 1^-} H''(t) = \frac{2(-\alpha^3 - \alpha^2 + 2\alpha + 1)}{\alpha^4}.$$

On obtient alors

$$E(X) = \sum_{k=1}^{\infty} k b_k = H'(1) = \frac{1 + \alpha}{\alpha^2}$$

et de même,

$$E(X(X-1)) = \sum_{k=1}^{\infty} k(k-1)b_k = H''(1) = \frac{2(-\alpha^3 - \alpha^2 + 2\alpha + 1)}{\alpha^4},$$

ce qui donne $V(X) = E(X(X-1)) + E(X) - [E(X)]^2$ et, tous calculs effectués,

$$V(X) = \frac{(1 - \alpha)(1 + 3\alpha + \alpha^2)}{\alpha^4}.$$

6 Soit $n \in \mathbb{N}^*$, fixé. Pour tout $p \in \mathbb{N}$, $p \neq n$,

$$a_{n,p} = \frac{1}{2n} \left(\frac{1}{n-p} + \frac{1}{n+p} \right) = \frac{1}{2n} \left(\frac{1}{p+n} - \frac{1}{p-n} \right).$$

On en déduit que, pour $P > n$, on a

$$\begin{aligned} \sum_{p=0}^P a_{n,p} &= \frac{1}{2n} \left(\sum_{\substack{0 \leq p \leq P \\ p \neq n}} \frac{1}{p+n} - \sum_{\substack{0 \leq p \leq P \\ p \neq n}} \frac{1}{p-n} \right) \\ &= \frac{1}{2n} \left(\sum_{k=n}^{n+P} \frac{1}{k} - \frac{1}{2n} + \sum_{k=1}^n \frac{1}{k} - \sum_{k=1}^{P-n} \frac{1}{k} \right) \\ &= \frac{1}{2n} \left(\sum_{k=1}^{n+P} \frac{1}{k} + \frac{1}{2n} - \sum_{k=1}^{P-n} \frac{1}{k} \right) \\ &= \frac{1}{2n} \left(\sum_{k=P-n+1}^{P+n} \frac{1}{k} + \frac{1}{2n} \right). \end{aligned}$$

L'entier n étant fixé, on trouve $\lim_{p \rightarrow +\infty} \sum_{k=p-n+1}^{p+n} \frac{1}{k} = 0$ et donc

$$\sum_{p=0}^{+\infty} a_{n,p} = \frac{1}{4n^2}.$$

Pour $n = 0$, on obtient $\sum_{p=0}^{+\infty} a_{n,p} = -\sum_{p=1}^{+\infty} \frac{1}{p^2}$. On en déduit que

$$\sum_{n=0}^{+\infty} \sum_{p=0}^{+\infty} a_{n,p} = -\sum_{p=1}^{+\infty} \frac{1}{p^2} + \sum_{n=1}^{+\infty} \frac{1}{4n^2} = -\frac{3}{4} \sum_{p=1}^{+\infty} \frac{1}{p^2}.$$

On remarque que, pour tout $(n, p) \in \mathbb{N}^2$, $a_{p,n} = -a_{n,p}$. On en déduit, d'après ce qui précède que, pour tout $p \in \mathbb{N}$,

$$\sum_{n=0}^{+\infty} a_{n,p} = -\sum_{n=0}^{+\infty} a_{p,n} = \begin{cases} -\frac{1}{4p^2} & \text{si } p \neq 0 \\ \sum_{n=1}^{+\infty} \frac{1}{n^2} & \text{si } p = 0, \end{cases}$$

puis

$$\sum_{p=0}^{+\infty} \sum_{n=0}^{+\infty} a_{n,p} = \frac{3}{4} \sum_{p=1}^{+\infty} \frac{1}{p^2}.$$

On obtient donc

$$\sum_{n=0}^{+\infty} \sum_{p=0}^{+\infty} a_{n,p} = -\sum_{p=0}^{+\infty} \sum_{n=0}^{+\infty} a_{n,p}.$$

On en déduit que la série double de terme général $a_{n,p}$ n'est pas absolument convergente.

7 L'entier $n \in \mathbb{N}^*$ étant fixé, la série $\sum_p a_{n,p}$ est une combinaison linéaire de deux séries géométriques convergentes. On obtient

$$\sum_{p=1}^{+\infty} a_{n,p} = \frac{1}{n+1} \cdot \frac{\frac{n}{n+1}}{1 - \frac{n}{n+1}} - \frac{1}{n+2} \cdot \frac{\frac{n+1}{n+2}}{1 - \frac{n+1}{n+2}} = \frac{n}{n+1} - \frac{n+1}{n+2}.$$

Comme pour $N \in \mathbb{N}^*$, on a

$$\sum_{n=1}^N \left(\frac{n}{n+1} - \frac{n+1}{n+2} \right) = \frac{1}{2} - \frac{N+1}{N+2},$$

car les termes s'éliminent deux à deux, on obtient

$$\sum_{n=1}^{+\infty} \left(\frac{n}{n+1} - \frac{n+1}{n+2} \right) = \frac{1}{2} - 1 = -\frac{1}{2},$$

c'est-à-dire

$$\sum_{n=1}^{+\infty} \sum_{p=1}^{+\infty} a_{n,p} = -\frac{1}{2}.$$

D'autre part, pour tout $p \in \mathbb{N}^*$, on a pour $N \in \mathbb{N}^*$,

$$\begin{aligned} \sum_{n=1}^N a_{n,p} &= \sum_{n=1}^N \left(\frac{1}{n+1} \left(\frac{n}{n+1} \right)^p - \frac{1}{n+2} \left(\frac{n+1}{n+2} \right)^p \right) \\ &= \left(\frac{1}{2} \right)^{p+1} - \frac{1}{N+2} \left(\frac{N+1}{N+2} \right)^p, \end{aligned}$$

car de nouveau les termes s'éliminent deux à deux. On en déduit que

$$\sum_{n=1}^{+\infty} a_{n,p} = \left(\frac{1}{2} \right)^{p+1}.$$

On obtient ensuite

$$\sum_{p=1}^{+\infty} \sum_{n=1}^{+\infty} a_{n,p} = \sum_{p=1}^{+\infty} \left(\frac{1}{2} \right)^{p+1} = \left(\frac{1}{2} \right)^2 \frac{1}{1 - \frac{1}{2}} = \frac{1}{2}.$$

On conclut comme dans la question précédente que la série double n'est pas absolument convergente.

8 Pour tout $n \in \mathbb{N}$, il existe deux termes $a_{n,p}$ non nuls qui sont $a_{n,n} = 1$ et soit $a_{n,n+1} = -1$ soit $a_{n,n-1} = -1$ selon que n est pair ou impair. Dans tous les cas on trouve

donc $\sum_{p=0}^{+\infty} a_{n,p} = 0$ et donc

$$\sum_{n=0}^{+\infty} \sum_{p=0}^{+\infty} a_{n,p} = 0.$$

De même pour tout $p \in \mathbb{N}$, il y a deux termes $a_{n,p}$ non nuls qui sont égaux à 1 et -1 . Ainsi

$$\sum_{p=0}^{+\infty} \sum_{n=0}^{+\infty} a_{n,p} = \sum_{p=0}^{+\infty} 0 = 0.$$

Pourtant la série double $\sum a_{n,p}$ n'est pas absolument convergente car, pour tout $n \in \mathbb{N}$,

$$\sum_{p=0}^{+\infty} |a_{n,p}| = 2$$

et la série de terme général 2 diverge.

9 Pour tout $(i, j) \in \mathbb{N}^2$,

$$a_{i,j} = \frac{i}{i!j!2^{i+j}} + \frac{j}{i!j!2^{i+j}}.$$

Posons $b_{i,j} = \frac{i}{i!j!2^{i+j}}$ et montrons que la série double à terme positifs $\sum b_{i,j}$ est convergente.

Pour tout $j \in \mathbb{N}$, on a

$$\sum_{i=0}^{+\infty} b_{i,j} = \frac{1}{j!2^{j+1}} \sum_{i=1}^{+\infty} \frac{\left(\frac{1}{2}\right)^{i-1}}{(i-1)!} = \frac{1}{j!2^{j+1}} e^{\frac{1}{2}}.$$

On en déduit que

$$\sum_{j=0}^{+\infty} \sum_{i=0}^{+\infty} b_{i,j} = \sum_{j=0}^{+\infty} \frac{1}{j!2^{j+1}} e^{\frac{1}{2}} = \frac{1}{2} e^{\frac{1}{2}} \sum_{j=0}^{+\infty} \frac{\left(\frac{1}{2}\right)^j}{j!} = \frac{1}{2} e^{\frac{1}{2}} e^{\frac{1}{2}} = \frac{e}{2}.$$

On montre de même que $\sum \frac{j}{i!j!2^{i+j}}$ converge et, pour des raisons de symétrie, a aussi pour somme $\frac{e}{2}$. On en déduit que la série double $\sum a_{i,j}$ converge, de somme e .

10 Soit $j \in \mathbb{N}$. On montre que $\sum_i |a_{i,j}|$ converge. On calcule

$$b_j = \sum_{i=0}^{+\infty} \frac{(i+j)!}{i!j!} |x|^{i+j} = \frac{|x|^j}{j!} \sum_{i=0}^{+\infty} (i+j)(i+j-1) \dots (i+1) |x|^i.$$

On reconnaît la dérivée j -ième d'une série géométrique convergente. On obtient

$$b_j = \frac{|x|^j}{j!} \frac{j!}{(1-|x|)^{j+1}} = \frac{1}{1-|x|} \left(\frac{|x|}{1-|x|} \right)^j.$$

Comme $|x| < \frac{1}{2}$, on a $0 \leq \frac{|x|}{1-|x|} < 1$ et b_j est le terme général d'une série géométrique convergente. On en déduit que la série double de terme général $a_{i,j}$ converge absolument. Le même calcul que précédemment montre que

$$\sum_{i=0}^{+\infty} a_{i,j} = \sum_{i=0}^{+\infty} \frac{(i+j)!}{i!j!} x^{i+j} = \frac{1}{1-x} \left(\frac{x}{1-x} \right)^j.$$

On en déduit que

$$\sum_{j=0}^{+\infty} \sum_{i=0}^{+\infty} a_{i,j} = \sum_{j=0}^{+\infty} \frac{1}{1-x} \left(\frac{x}{1-x} \right)^j = \frac{1}{1-x} \cdot \frac{1}{1-\frac{x}{1-x}} = \frac{1}{1-2x}.$$

La série double $\sum a_{i,j}$ a pour somme $\frac{1}{1-2x}$.

11 Considérons la série double définie pour $(p, q) \in (\mathbb{N}^*)^2$ par

$$a_{p,q} = \begin{cases} \frac{(-1)^p}{q^3} & \text{si } q \geq p \\ 0 & \text{sinon.} \end{cases}$$

Montrons que cette série est absolument convergente. Pour tout $q \in \mathbb{N}^*$, on a

$$\sum_{p=1}^{+\infty} |a_{p,q}| = \sum_{p=1}^q \frac{1}{q^3} = \frac{1}{q^2}.$$

Comme la série de terme général $\frac{1}{q^2}$ converge, la série double $\sum a_{p,q}$ est absolument convergente. On en déduit que

$$\sum_{p=1}^{+\infty} \sum_{q=p}^{+\infty} \frac{(-1)^p}{q^3} = \sum_{p=1}^{+\infty} \sum_{q=1}^{+\infty} a_{p,q} = \sum_{q=1}^{+\infty} \sum_{p=1}^{+\infty} a_{p,q} = \sum_{q=1}^{+\infty} \sum_{p=1}^q \frac{(-1)^p}{q^3} = \sum_{q=1}^{+\infty} \frac{1}{q^3} \sum_{p=1}^q (-1)^p.$$

Comme

$$\sum_{p=1}^q (-1)^p = \begin{cases} 0 & \text{si } q \text{ est pair} \\ -1 & \text{si } q \text{ est impair,} \end{cases}$$

on obtient

$$\sum_{p=1}^{+\infty} \sum_{q=p}^{+\infty} \frac{(-1)^p}{q^3} = \sum_{k=0}^{+\infty} \frac{-1}{(2k+1)^3} = -\sum_{k=1}^{+\infty} \frac{1}{k^3} + \sum_{k=1}^{+\infty} \frac{1}{(2k)^3} = -\frac{7}{8} \sum_{k=1}^{+\infty} \frac{1}{k^3}.$$

12

Comme $|x| < 1$, on a, pour tout entier naturel n , $|x^{2n+1}| < 1$, donc

$$\frac{x^{2n+1}}{1 - x^{2n+1}} = \sum_{p=0}^{+\infty} x^{2n+1} (x^{2n+1})^p = \sum_{p=0}^{+\infty} (x^{2n+1})^{p+1}.$$

Cela nous amène à considérer la série double définie par

$$a_{n,p} = (x^{2n+1})^{p+1}.$$

Montrons que cette série double est absolument convergente. Pour n fixé dans \mathbb{N} , on a $|x^{2n+1}| < 1$. On en déduit que

$$\sum_{p=0}^{+\infty} |a_{n,p}| = \sum_{p=0}^{+\infty} |x^{2n+1}|^{p+1} = \frac{|x^{2n+1}|}{1 - |x^{2n+1}|}.$$

Comme $|x| < 1$, on a $\lim_{n \rightarrow +\infty} |x^{2n+1}|$ et

$$\frac{|x^{2n+1}|}{1 - |x^{2n+1}|} \underset{n \rightarrow +\infty}{\sim} |x^{2n+1}| \underset{n \rightarrow +\infty}{\sim} |x|(x^2)^n.$$

La série de terme général $|x|(x^2)^n$ est une série géométrique convergente, donc la série $\sum a_{n,p}$ est absolument convergente. On en déduit que

$$\sum_{n=0}^{+\infty} \frac{x^{2n+1}}{1 - x^{2n+1}} = \sum_{n=0}^{+\infty} \sum_{p=0}^{+\infty} a_{n,p} = \sum_{p=0}^{+\infty} \sum_{n=0}^{+\infty} a_{n,p}.$$

Pour tout $p \in \mathbb{N}$, on a

$$\sum_{n=0}^{+\infty} a_{n,p} = \sum_{n=0}^{+\infty} x^{p+1} (x^{2(p+1)})^n = \frac{x^{p+1}}{1 - x^{2(p+1)}}.$$

On obtient

$$\sum_{p=0}^{+\infty} \sum_{n=0}^{+\infty} a_{n,p} = \sum_{p=0}^{+\infty} \frac{x^{p+1}}{1 - x^{2(p+1)}} = \sum_{n=1}^{+\infty} \frac{x^n}{1 - x^{2n}}$$

et finalement

$$\sum_{n=0}^{+\infty} \frac{x^{2n+1}}{1 - x^{2n+1}} = \sum_{n=1}^{+\infty} \frac{x^n}{1 - x^{2n}}.$$

13 On procède comme dans l'exercice précédent. Comme $|x| < 1$, on a $|x^{an+c}| \leq |x^c| < 1$, donc

$$\frac{x^{bn}}{1 - x^{an+c}} = \sum_{p=0}^{+\infty} x^{bn} (x^{an+c})^p = x^{bn+cp+ap}.$$

on considère la série double définie, pour $(n, p) \in \mathbb{N}^2$ par

$$u_{n,p} = x^{bn+cp+ap}.$$

On montre que cette série double est absolument convergente. Pour tout $n \in \mathbb{N}$, on a

$$|u_{n,p}| = |x^{bn}| (|x^{an+c}|)^p,$$

avec $|x^{an+c}| < 1$, donc

$$\sum_{p=0}^{+\infty} |u_{n,p}| = |x^{bn}| \frac{1}{1 - |x^{an+c}|}.$$

Comme $|x| < 1$, on a

$$|x^{bn}| \frac{1}{1 - |x^{an+c}|} \underset{n \rightarrow +\infty}{\sim} |x^b|^n$$

terme général d'une série géométrique convergente, donc la série de terme général $|x^{bn}| \frac{1}{1 - |x^{an+c}|}$ converge et la série double $\sum u_{n,p}$ est absolument convergente. On en déduit que

$$\sum_{n=0}^{+\infty} \frac{x^{bn}}{1 - x^{an+c}} = \sum_{n=0}^{+\infty} \sum_{p=0}^{+\infty} u_{n,p} = \sum_{p=0}^{+\infty} \sum_{n=0}^{+\infty} u_{n,p}.$$

Comme

$$\sum_{n=0}^{+\infty} u_{n,p} = \sum_{n=0}^{+\infty} x^{cp} (x^{b+ap})^n = \frac{x^{cp}}{1 - x^{b+ap}},$$

on obtient finalement

$$\sum_{n=0}^{+\infty} \frac{x^{bn}}{1 - x^{an+c}} = \sum_{p=0}^{+\infty} \frac{x^{cp}}{1 - x^{b+ap}}.$$

Pour démontrer la deuxième égalité, on considère la série double définie pour $(n, p) \in \mathbb{N}^2$ par

$$v_{n,p} = (-1)^p x^{bn+cp+anp}.$$

Comme $|v_{n,p}| = |u_{n,p}|$, la série $\sum v_{n,p}$ est absolument convergente ; On en déduit que

$$\sum_{n=0}^{+\infty} \frac{x^{bn}}{1 + x^{an+c}} = \sum_{n=0}^{+\infty} \sum_{p=0}^{+\infty} v_{n,p} = \sum_{p=0}^{+\infty} \sum_{n=0}^{+\infty} v_{n,p}.$$

Comme

$$\sum_{n=0}^{+\infty} v_{n,p} = \sum_{n=0}^{+\infty} (-1)^p x^{cp} (x^{b+ap})^n = \frac{(-1)^p x^{cp}}{1 - x^{b+ap}},$$

on obtient finalement

$$\sum_{n=0}^{+\infty} \frac{x^{bn}}{1 + x^{an+c}} = \sum_{p=0}^{+\infty} \frac{(-1)^p x^{cp}}{1 - x^{b+ap}}.$$

14

1. Pour $m \in \mathbb{N}^*$, la série de terme général $\frac{\sqrt{m}}{(m+n)\sqrt{n}}$ converge car

$$\frac{\sqrt{m}}{(m+n)\sqrt{n}} \underset{n \rightarrow +\infty}{\sim} \frac{\sqrt{m}}{n^{\frac{3}{2}}},$$

terme général d'une série de Riemann convergente.

On considère la fonction continue sur $]0, +\infty[$, positive, $f : x \mapsto \frac{1}{(x+1)\sqrt{x}}$. L'intégrale

$\int_0^1 f(x) dx$ converge, par comparaison avec une intégrale de Riemann, car $f(x) \underset{x \rightarrow 0}{\sim} \frac{1}{x^{\frac{1}{2}}}$; il

en est de même de $\int_1^{+\infty} f(x) dx$ car $f(x) \underset{x \rightarrow +\infty}{\sim} \frac{1}{x^{\frac{3}{2}}}$. Finalement $\int_0^{+\infty} f(x) dx$ converge.

La fonction f est décroissante sur $]0, +\infty[$. On a donc, pour tout $n \in \mathbb{N}^*$,

$$\int_{\frac{n-1}{m}}^{\frac{n}{m}} f(x) dx \geq \int_{\frac{n-1}{m}}^{\frac{n}{m}} f\left(\frac{n}{m}\right) \geq \frac{1}{m} f\left(\frac{n}{m}\right) \geq \frac{1}{m} \frac{1}{\left(\frac{n}{m} + 1\right) \sqrt{\frac{n}{m}}} \geq \frac{\sqrt{m}}{(n+m)\sqrt{n}}.$$

Soit $N \in \mathbb{N}^*$. En sommant les inégalités obtenues pour n variant de 1 à N , on obtient

$$\int_0^{\frac{N}{m}} f(x) dx \geq \sum_{n=1}^N \frac{\sqrt{m}}{(n+m)\sqrt{n}}.$$

En faisant tendre N vers $+\infty$, on obtient

$$\int_0^{+\infty} f(x) dx \geq \sum_{n=1}^{+\infty} \frac{\sqrt{m}}{(n+m)\sqrt{n}}.$$

Dans la suite de l'exercice, on note $I = \int_0^{+\infty} f(x) dx$.

2. a) Pour $n \in \mathbb{N}^*$, on écrit

$$\sum_{1 \leq i, j \leq n} \frac{a_i b_j}{i+j} = \sum_{1 \leq i, j \leq n} \frac{a_i i^{\frac{1}{4}}}{j^{\frac{1}{4}} \sqrt{i+j}} \cdot \frac{b_j j^{\frac{1}{4}}}{i^{\frac{1}{4}} \sqrt{i+j}}$$

et l'on applique l'inégalité de Cauchy-Schwarz. On obtient

$$\sum_{1 \leq i, j \leq n} \frac{a_i b_j}{i+j} \leq \left(\sum_{1 \leq i, j \leq n} \frac{a_i^2 \sqrt{i}}{\sqrt{j}(i+j)} \right)^{\frac{1}{2}} \left(\sum_{1 \leq i, j \leq n} \frac{b_j^2 \sqrt{j}}{\sqrt{i}(i+j)} \right)^{\frac{1}{2}}.$$

b) Pour $i \in \llbracket 1, n \rrbracket$, on obtient grâce à la question 1,

$$\sum_{j=1}^n \frac{a_i^2 \sqrt{i}}{\sqrt{j}(i+j)} \leq a_i^2 \sum_{j=1}^{+\infty} \frac{\sqrt{i}}{\sqrt{j}(i+j)} \leq a_i^2 I,$$

puis

$$\sum_{1 \leq i, j \leq n} \frac{a_i^2 \sqrt{i}}{\sqrt{j}(i+j)} \leq \sum_{i=1}^n a_i^2 I \leq I \sum_{i=1}^{+\infty} a_i^2.$$

Le même raisonnement peut être tenu pour l'autre somme et l'on obtient finalement, pour tout $n \in \mathbb{N}^*$,

$$\sum_{1 \leq i, j \leq n} \frac{a_i b_j}{i+j} \leq I \left(\sum_{i=1}^{+\infty} a_i^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^{+\infty} b_j^2 \right)^{\frac{1}{2}}.$$

On en déduit que, pour toutes parties finies A et B de \mathbb{N}^* , on a

$$\sum_{(i,j) \in A \times B} \frac{a_i b_j}{i+j} \leq I \left(\sum_{i=1}^{+\infty} a_i^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^{+\infty} b_j^2 \right)^{\frac{1}{2}},$$

car il existe $n \in \mathbb{N}^*$ tel que $A \times B \subset \llbracket 1, n \rrbracket^2$. Cela montre que la série double à termes positifs $\sum \frac{a_i b_j}{i+j}$ converge et que

$$\sum_{(i,j) \in (\mathbb{N}^*)^2} \frac{a_i b_j}{i+j} \leq I \left(\sum_{i=1}^{+\infty} a_i^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^{+\infty} b_j^2 \right)^{\frac{1}{2}}.$$

On calcule I , en faisant le changement de variable $t = \sqrt{x}$. On obtient

$$I = \int_0^{+\infty} \frac{2t dt}{(t^2+1)t} = 2 \int_0^{+\infty} \frac{dt}{1+t^2} = \pi,$$

ce qui donne l'inégalité attendue.

15 1. On a, par définition $S = \sum_{(i,j) \in \mathbb{N}^2} a_{i,j}$. On pose, pour tout $(i,j) \in \mathbb{N}^2$, $f(i,j) = i + j$

puis, pour $n \in \mathbb{N}$, $u_n = \sum_{f(i,j)=n} a_{i,j} = \sum_{(i,j) \in I_n} a_{i,j}$. Le théorème de sommation par paquets, compte tenu de la bijection entre \mathbb{N}^2 et \mathbb{N} , donne la convergence de la série de terme général u_n et

$$S = \sum_{n=0}^{+\infty} u_n = \sum_{n=0}^{+\infty} \sum_{(i,j) \in I_n} a_{i,j}.$$

2. Pour tout $(i,j) \in \mathbb{N}^2$, on a

$$|a_{i,j}| = \frac{|x|^i |y|^j}{\binom{i+j}{i} i! j!} \leq \frac{|x|^i}{i!} \cdot \frac{|y|^j}{j!}.$$

Pour toutes parties finies A et B de \mathbb{N} , on a

$$\begin{aligned} \sum_{(i,j) \in A \times B} |a_{i,j}| &\leq \sum_{(i,j) \in A \times B} \frac{|x|^i}{i!} \cdot \frac{|y|^j}{j!} \\ &\leq \left(\sum_{i \in A} \frac{|x|^i}{i!} \right) \left(\sum_{j \in B} \frac{|y|^j}{j!} \right) \\ &\leq \left(\sum_{i=0}^{+\infty} \frac{|x|^i}{i!} \right) \left(\sum_{j=1}^{+\infty} \frac{|y|^j}{j!} \right) \\ &\leq e^{|x|} e^{|y|}, \end{aligned}$$

donc la série double $\sum a_{i,j}$ est absolument convergente.

Pour calculer la somme de cette série, on utilise la question 1. On obtient

$$\sum_{(i,j) \in I_n} a_{i,j} = \sum_{(i,j) \in I_n} \frac{x^i y^j}{n!} = \frac{1}{n!} \sum_{i=0}^n x^i y^{n-i}.$$

On en déduit

$$\sum_{(i,j) \in I_n} a_{i,j} = \begin{cases} \frac{1}{n!} \cdot \frac{x^{n+1} - y^{n+1}}{x - y} & \text{si } x \neq y \\ \frac{(n+1)x^n}{n!} & \text{si } x = y \end{cases}$$

La somme de la série est donc

$$\begin{aligned} S &= \sum_{n=0}^{+\infty} \frac{x^{n+1} - y^{n+1}}{n!(x - y)} = \frac{xe^x - ye^y}{x - y} && \text{si } x \neq y, \\ S &= \sum_{n=0}^{+\infty} \frac{(n+1)x^n}{n!} = \sum_{n=1}^{+\infty} \frac{x^n}{(n-1)!} + \sum_{n=0}^{+\infty} \frac{x^n}{n!} \\ &= xe^x + e^x = (x+1)e^x, && \text{si } x = y. \end{aligned}$$

16 1. Pour tout $n \in \mathbb{N}^*$, on a $|x^{n+1}| < 1$, donc on peut écrire

$$\frac{x^n}{1-x^n} = \sum_{p=0}^{+\infty} x^n (x^n)^p = \sum_{p=1}^{+\infty} (x^n)^p.$$

Considérons la série double définie pour $(n, p) \in \mathbb{N}^2$ par

$$a_{n,p} = x^{np}.$$

On obtient, pour tout $n \in \mathbb{N}^*$,

$$\sum_{p=1}^{+\infty} |x|^{np} = \sum_{p=1}^{+\infty} (|x|^n)^p = \frac{|x|^n}{1-|x|^n}.$$

Comme $|x| < 1$, on a $\frac{|x|^n}{1-|x|^n} \underset{n \rightarrow +\infty}{\sim} |x|^n$ et la série de terme général $\frac{|x|^n}{1-|x|^n}$ est convergente. Ainsi, la série $\sum a_{n,p}$ est absolument convergente.

La somme de cette série est égale à

$$S = \sum_{n=1}^{+\infty} \sum_{p=1}^{+\infty} x^{np} = \sum_{n=1}^{+\infty} \frac{x^n}{1-x^n}.$$

On considère la fonction f définie sur $(\mathbb{N}^*)^2$ par $f(n, p) = np$. En posant $v_k = \sum_{f(n,p)=k} a_{n,p}$,

on obtient par le théorème de sommation par paquets, compte tenu de la bijection entre \mathbb{N}^2 et \mathbb{N} ,

$$S = \sum_{k=1}^{+\infty} v_k.$$

On calcule

$$\begin{aligned} v_k &= \sum_{np=k} x^{np} = x^k \text{Card}\{(n, p) \in (\mathbb{N}^*)^2, np = k\} \\ &= x^k \text{Card}\{n \in \mathbb{N}^*, n \text{ divise } k\} = x^k d(k), \end{aligned}$$

d'où l'on tire finalement

$$\sum_{n=1}^{+\infty} \frac{x^n}{1-x^n} = \sum_{k=1}^{+\infty} d(k)x^k.$$

2. On note que, pour $p \geq 2$, la série de terme général $\frac{1}{j^p}$ converge. On a, pour tout $i \in \mathbb{N}^*$,

$$\sum_{j=1}^{+\infty} \frac{1}{i^p j^p} = \frac{1}{i^p} \sum_{j=1}^{+\infty} \frac{1}{j^p}.$$

On en déduit que

$$\sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} \frac{1}{i^p j^p} = \sum_{i=1}^{+\infty} \left(\frac{1}{i^p} \sum_{j=1}^{+\infty} \frac{1}{j^p} \right) = \left(\sum_{i=1}^{+\infty} \frac{1}{i^p} \right) \left(\sum_{j=1}^{+\infty} \frac{1}{j^p} \right) = \left(\sum_{i=1}^{+\infty} \frac{1}{i^p} \right)^2.$$

Ainsi la série double $\sum \frac{1}{i^p j^p}$ converge et sa somme est $\left(\sum_{i=1}^{+\infty} \frac{1}{i^p}\right)^2$. En utilisant le théorème de sommation par paquets comme dans la question 1, on obtient

$$\left(\sum_{i=1}^{+\infty} \frac{1}{i^p}\right)^2 = \sum_{n=1}^{+\infty} v_n,$$

où

$$\begin{aligned} v_n &= \sum_{(i,j) \in \mathbb{N}^2, ij=n} \frac{1}{i^p j^p} = \frac{1}{n^p} \text{Card}\{(i,j) \in \mathbb{N}^2, ij=n\} \\ &= \frac{1}{n^p} \text{Card}\{i \in \mathbb{N}, i \text{ divise } n\} = \frac{d(n)}{n^p}. \end{aligned}$$

On a finalement

$$\left(\sum_{i=1}^{+\infty} \frac{1}{i^p}\right)^2 = \sum_{n=1}^{+\infty} \frac{d(n)}{n}.$$

17 1. On note $I_k = \llbracket 1, k \rrbracket^2$, pour tout $k \in \mathbb{N}^*$. On a alors $u_k = \sum_{(n,p) \in I_k} a_{n,p}$ et donc

$$u_{k+1} - u_k = \sum_{(n,p) \in I_{k+1} \setminus I_k} a_{n,p}.$$

L'ensemble $I_{k+1} \setminus I_k$ possède $(k+1)^2 - k^2 = (2k+1)$ éléments.

Si $(n,p) \in I_{k+1} \setminus I_k$, on a $1 \leq n \leq (k+1)$ et $1 \leq p \leq (k+1)$. On en déduit que

$$(n^2 + p^2)^\alpha \leq (2(k+1)^2)^\alpha \leq 2^\alpha (k+1)^{2\alpha}.$$

Mais d'autre n ou p est égal à $k+1$, sinon (n,p) appartient à I_k donc

$$(n^2 + p^2)^\alpha \geq (k+1)^{2\alpha}.$$

On en déduit que, pour tout $(n,p) \in I_{k+1} \setminus I_k$,

$$\frac{1}{2^\alpha (k+1)^{2\alpha}} \leq a_{n,p} \leq \frac{1}{(k+1)^{2\alpha}}$$

et puisque $I_{k+1} \setminus I_k$ comporte $2k+1$ éléments,

$$\frac{2k+1}{2^\alpha (k+1)^{2\alpha}} \leq u_{k+1} - u_k \leq \frac{2k+1}{(k+1)^{2\alpha}}.$$

2. On a $\frac{2k+1}{2^\alpha (k+1)^{2\alpha}} \underset{k \rightarrow +\infty}{\sim} \frac{2^{1-\alpha}}{k^{2\alpha-1}}$. Si $\alpha \leq 1$, c'est-à-dire si $2\alpha - 1 \leq 1$, la série de terme général $\frac{2k+1}{2^\alpha (k+1)^{2\alpha}}$ diverge et, d'après le théorème de comparaison des séries à termes positifs, la série de terme général $u_{k+1} - u_k$ diverge. On peut écrire pour tout $k > 1$,

$$u_k = u_1 + \sum_{i=1}^{k-1} (u_{i+1} - u_i).$$

Comme la série de terme général positif $u_{k+1} - u_k$ diverge, ses sommes partielles tendent vers $+\infty$ et

$$\lim_{k \rightarrow +\infty} u_k = \lim_{k \rightarrow +\infty} u_1 + \sum_{k=1}^{k-1} (u_{i+1} - u_i) = +\infty.$$

Les sommes $\sum_{(i,j) \in I_k} a_{n,p}$ ne sont pas bornées donc la série double $\sum a_{n,p}$ diverge.

3. On a $\frac{2k+1}{(k+1)^{2\alpha}} \underset{k \rightarrow +\infty}{\sim} \frac{2^{1-\alpha}}{k^{2\alpha-1}}$. Si $\alpha > 1$, c'est-à-dire $2\alpha - 1 > 1$, la série de terme général $\frac{2k+1}{2^\alpha(k+1)^{2\alpha}}$ converge et, d'après le théorème de comparaison des séries à termes positifs, la

série de terme général $u_{k+1} - u_k$ converge. En reprenant la relation $u_k = u_1 + \sum_{k=1}^{k-1} (u_{i+1} - u_i)$,

on en déduit que la suite $(u_k)_{k \in \mathbb{N}^*}$ converge. Soit ℓ sa limite, qui est un majorant de la suite car elle est croissante. Pour toutes parties A et B finies de \mathbb{N}^* , il existe $k \in \mathbb{N}^*$ tel que $A \times B \subset I_k$. On alors

$$\sum_{(n,p) \in A \times B} a_{n,p} \leq \sum_{(n,p) \in I_k} a_{n,p} \leq \ell.$$

Les sommes partielles $\sum_{(n,p) \in A \times B} a_{n,p}$ sont bornées donc la série double $\sum a_{n,p}$ converge.

- 18** 1. Les séries $\sum a_n$ et $\sum na_n$ sont à termes positifs. Comme $a_n \underset{n \rightarrow +\infty}{\sim} \frac{1}{n^2}$ et $na_n \underset{n \rightarrow +\infty}{\sim} \frac{1}{n}$, la première série converge et la seconde diverge.

2. Pour qu'une telle variable existe, il faut que les nombres $\frac{1}{2}a_{|n|}$ soient positifs, ce qui est clair, et que $\sum_{n \in \mathbb{Z}^*} \frac{1}{2}a_{|n|} = 1$. On observe que $\sum_{n \in \mathbb{Z}_+^*} \frac{1}{2}a_{|n|} = \sum_{n \in \mathbb{N}^*} \frac{1}{2}a_n$. Ainsi, on a

$$\sum_{n \in \mathbb{Z}^*} \frac{1}{2}a_{|n|} = \sum_{n \in \mathbb{N}^*} a_n = \sum_{n=1}^{+\infty} \left(\frac{1}{n} - \frac{1}{n+1} \right).$$

Comme $\sum_{n=1}^N \left(\frac{1}{n} - \frac{1}{n+1} \right) = 1 - \frac{1}{N+1}$, on obtient

$$\sum_{n \in \mathbb{Z}^*} \frac{1}{2}a_{|n|} = \lim_{N \rightarrow +\infty} \left(1 - \frac{1}{N+1} \right) = 1.$$

3. On remarque que A_n est l'événement $[|X| = n]$. Comme $|X|$ est à valeurs dans \mathbb{N}^* , la famille $(A_n)_{n \in \mathbb{N}^*}$ forme un système complet d'événements.
4. Pour tout $n \in \mathbb{N}^*$, on a sous réserve d'absolue convergence,

$$E(X|A_n) = \sum_{k \in \mathbb{Z}} kP(X = k|A_n) = \sum_{k \in \mathbb{Z}} kP(X = k | |X| = n).$$

Il est clair que si $|k| \neq n$, la probabilité $P(X = k | |X| = n)$ est nulle. Il en résulte que $E(X|A_n)$ existe et

$$\begin{aligned} E(X|A_n) &= nP(X = n | |X| = n) - nP(X = -n | |X| = n) \\ &= n \frac{P(X = n)}{P(|X| = n)} - n \frac{P(X = -n)}{P(|X| = n)} = 0, \end{aligned}$$

car $P(X = n) = P(X = -n) = \frac{1}{2}a_{|n|}$.

5. La série de terme général $E(X|A_n)P(A_n)$ converge, car c'est la série nulle. Sa somme est nulle.
6. La variable X n'a pas d'espérance puisque la série $\sum_{n \in \mathbb{N}^*} nP(X = n)$, i.e $\sum na_n$ diverge.

Ainsi, pour un système complet d'événements (A_n) , l'existence de l'espérance conditionnelle $E(X|A_n)$ pour tout n et la convergence absolue de la série $\sum E(X|A_n)P(A_n)$ ne suffit pas à assurer l'existence de $E(X)$.

Chapitre 9

- 1** 1. Remarquons que c est c qui doit être non nul. Nous avons,

$$\rho_{aX+b, cY+d} = \frac{\text{Cov}(aX + b, cY + d)}{\sigma(aX + b)\sigma(cY + d)} = \frac{ac\text{Cov}(X, Y)}{|a| |c| \sigma(X)\sigma(Y)} = \frac{ac}{|ac|} \rho_{X, Y}.$$

Si a et c sont de même signe, alors $\rho_{aX+b, cY+d} = \rho_{X, Y}$.

2. a) On a $\rho_{X, aX+b} = \frac{\text{Cov}(X, aX + b)}{\sigma(X)\sigma(aX + b)} = \frac{a\text{Cov}(X, X)}{|a| V(X)} = \pm 1$.
- b) Nous venons de voir que ce coefficient est égal à 1 en valeur absolue.
3. a) On a

$$\begin{aligned} V(Z) &= \frac{1}{\sigma_Y^2} V(Y) + \frac{\rho_{X, Y}^2}{\sigma_X^2} V(X) - 2 \frac{\rho_{X, Y}}{\sigma_X \sigma_Y} \text{Cov}(X, Y) \\ &= 1 + 1 - 2\rho_{X, Y}^2 = 2(1 - \rho_{X, Y}^2). \end{aligned}$$

- b) Si $|\rho_{X, Y}| = 1$ alors $V(Z) = 0$ et donc Z est une variable égale presque sûrement à $E(Z)$, autrement dit si $|\rho_{X, Y}| = 1$ alors il existe deux réels a, b tels que $Y = aX + b$.

- 2** 1. a) Par le calcul, on trouve

X \ Y	1	2	3	4	loi de X
1	0,08	0,04	0,16	0,12	0,4
2	0,04	0,02	0,08	0,06	0,2
3	0,08	0,04	0,16	0,12	0,4
loi de Y	0,2	0,1	0,4	0,3	1

b) On constate que $\forall(x, y) \in \llbracket 1, 3 \rrbracket \times \llbracket 1, 4 \rrbracket$, on a

$$P([X = x] \cap [Y = y]) = P(X = x) \times P(Y = y).$$

On en déduit que les variables X et Y sont indépendantes.

c) Il résulte de l'indépendance que $\text{Cov}(X, Y) = 0$.

2. a) Les variables étant indépendantes, la loi de X sachant $Y = 2$ est la loi marginale de X et la loi de Y sachant $X \in \{1, 3\}$ est la loi marginale de Y .

b) Étant donné que X et Y sont indépendantes, on a

$$E(Y|X) = E(Y) = 1 \times 0,2 + 2 \times 0,1 + 3 \times 0,4 + 4 \times 0,3 = 2,8.$$

(en particulier la variable $E(Y|X)$ est constante).

c) Il résulte de la question précédente que $E(E(Y|X)) = E(Y) = 2,8$.

3. Pour $(i, i) \in \llbracket 1, 3 \rrbracket^2$, nous avons

$$P((\inf(X, Y), \sup(X, Y)) = (i, i)) = P(X = i)P(Y = i)$$

et pour $(i, j) \in \llbracket 1, 3 \rrbracket \times \llbracket 1, 4 \rrbracket$ avec $i < j$, on a

$$P((\inf(X, Y), \sup(X, Y)) = (i, j)) = P(X = i)P(Y = j) + P(X = j)P(Y = i).$$

D'où le tableau suivant pour la loi de $(\inf(X, Y), \sup(X, Y))$:

$\inf(X, Y) \backslash \sup(X, Y)$	1	2	3	4
1	0,08	0,08	0,24	0,12
2	0	0,02	0,12	0,06
3	0	0	0,16	0,12

3 1. a) Nous avons

$X \backslash Y$	1	2	3	4	loi de X
1	0	0	0	0,3	0,3
2	0,2	0	0	0	0,2
3	0	0	0,1	0	0,1
4	0,3	0,1	0	0	0,4
loi de Y	0,5	0,1	0,1	0,3	1

b) On a $P([X = 1] \cap [Y = 1]) = 0 \neq P(X = 1)P(Y = 1)$. Les variables X et Y ne sont pas indépendantes.

c) On a $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.

D'après les lois marginales de X et Y , on a

$$E(X) = 0,3 + 0,4 + 0,3 + 1,6 = 2,6$$

$$\text{et } E(Y) = 0,5 + 0,2 + 0,3 + 1,2 = 2,2.$$

Et en utilisant le tableau de la loi du couple (X, Y) on obtient

$$E(XY) = 4 \times 0,3 + 2 \times 0,2 + 9 \times 0,1 + 4 \times 0,3 + 8 \times 0,1 = 4,5.$$

d'où $\text{Cov}(X, Y) = 4,5 - 2,6 \times 2,2 = -1,22$.

2. a) Pour la loi de X sachant $Y = 2$, on a $P_{Y=2}(X = 4) = 1$ et $P_{Y=2}(X = k) = 0$ si $k \in \llbracket 1, 3 \rrbracket$.

En remarquant que

$$P_{X \in \{1,4\}}(Y = k) = \frac{1}{P(X \in \{1,4\})} (P([Y = k] \cap [X = 1]) + P([Y = k] \cap [X = 3])),$$

on obtient la loi de Y sachant $X \in \{1, 4\}$:

y	1	2	3	4
$P_{X \in \{1,4\}}(Y = y)$	3/7	1/7	0	3/7

- b) Nous avons

$$E_{Y=1}(X) = 2 \times 0,4 + 4 \times 0,6 = 3,2, \quad E_{Y=2}(X) = 4, \quad E_{Y=3}(X) = 3 \quad \text{et} \quad E_{Y=4}(X) = 1.$$

D'après la définition de $E(X|Y)$, il en résulte que

x	1	3	3,2	4
$E(X Y) = x$	0,3	0,1	0,5	0,1

- c) On a

$$E(E(X|Y)) = 0,3 + 0,3 + 1,6 + 0,4 = 2,6 = E(X).$$

Cela confirme le théorème de l'espérance totale.

3. En raisonnant comme dans l'exercice précédent, on trouve

	$\sup(X, Y)$	1	2	3	4
$\inf(X, Y)$					
1		0	0,2	0	0,6
2		0	0	0	0,1
3		0	0	0,1	0
4		0	0	0	0

4

1. On sait que l'on doit avoir $\sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} P([X = j] \cap [Y = k]) = 1$, or

$$\begin{aligned} \sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} P([X = j] \cap [Y = k]) &= \sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} \frac{(j+k)\lambda^{j+k}}{e^j k!} \\ &= \sum_{j=0}^{+\infty} \frac{\lambda^j}{e^j} \left(j \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} + \sum_{k=0}^{+\infty} \frac{k\lambda^k}{k!} \right) \\ &= \sum_{j=0}^{+\infty} \frac{\lambda^j}{e^j} (j + \lambda) e^\lambda = e^{\lambda-1} \left(\sum_{j=1}^{+\infty} \frac{\lambda^j}{(j-1)!} + \lambda \sum_{j=0}^{+\infty} \frac{\lambda^j}{j!} \right) \\ &= e^{\lambda-1} e^\lambda \times 2 \times \lambda = 2\lambda e^{2\lambda-1}. \end{aligned}$$

La fonction $f : \lambda \mapsto 2\lambda e^{2\lambda-1}$ est strictement croissante sur \mathbb{R}^+ , or on remarque que $f(1/2) = 1$.

Nous avons donc $\lambda = \frac{1}{2}$.

2. D'après les probabilités totales, pour $j \in \mathbb{N}$,

$$\begin{aligned} P(X = j) &= \sum_{k=0}^{+\infty} \frac{(j+k)\lambda^{j+k}}{e^j k!} \\ &= \frac{\lambda^j}{e^j} \left(j \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} + \sum_{k=0}^{+\infty} \frac{k\lambda^k}{k!} \right) = \frac{\lambda^j e^{\lambda-1}}{j!} (j + \lambda) \\ &= \left(\frac{1}{2}\right)^j \frac{e^{-\frac{1}{2}}}{j!} \left(j + \frac{1}{2}\right). \end{aligned}$$

3. Un calcul tout à fait analogue aurait prouvé que $P(Y = k) = \left(\frac{1}{2}\right)^k \frac{e^{-\frac{1}{2}}}{k!} \left(k + \frac{1}{2}\right)$.

En particulier $P(X = 0) = P(Y = 0) = \frac{1}{2}e^{-\frac{1}{2}}$ et $P([X = 0] \cap [Y = 0]) = \frac{1}{e}$.

Or $P(X = 0)P(Y = 0) = \frac{1}{4}e^{-1} \neq e^{-1}$. Les variables X et Y ne sont pas indépendantes.

4. D'après le théorème de transfert 2^{X+Y} admet une espérance si la série

double $\sum_{(j,k) \in \mathbb{N}^2} 2^{j+k} \frac{(j+k)\lambda^{j+k}}{e^j k!}$ converge absolument. Elle est à termes positifs et par un calcul analogue à celui de la première question (où λ est remplacé par 2λ) on trouve que

$$\sum_{(j,k) \in \mathbb{N}^2} 2^{j+k} \frac{(j+k)\lambda^{j+k}}{e^j k!} = 4\lambda e^{4\lambda-1} = 2e.$$

D'où $E(2^{X+Y}) = 2e$.

5

1. Nous avons

$$\sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} \frac{1}{ek!2^{j+1}} = \sum_{j=0}^{+\infty} \frac{1}{2^{j+1}} = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1.$$

On définit bien ainsi une loi conjointe.

2. Par les probabilités totales, pour $j \in \mathbb{N}$,

$$P(X = j) = \sum_{k=0}^{+\infty} \frac{1}{ek!2^{j+1}} = \frac{1}{2^{j+1}}.$$

et pour $k \in \mathbb{N}$,

$$P(Y = k) = \sum_{j=0}^{+\infty} \frac{1}{ek!2^{j+1}} = \frac{1}{ek!}.$$

3. On a pour $(j, k) \in \mathbb{N}^2$,

$$P(X = j)P(Y = k) = \frac{1}{2^{j+1}} \frac{1}{ek!} = \frac{1}{ek!2^{j+1}} = P([X = j] \cap [Y = k]).$$

Les variables X et Y sont indépendantes.

4. On a pour $n \in \mathbb{N}$,

$$\sum_{j=0}^n jP(X = j) = \sum_{j=0}^n \frac{j}{2^{j+1}} \rightarrow \frac{1}{2} \frac{\frac{1}{2}}{\left(1 - \frac{1}{2}\right)^2} = 1.$$

La variable X admet une espérance et $E(X) = 1$.

De même

$$\sum_{j=0}^n j^2 P(X = j) = \sum_{j=0}^n \frac{j^2}{2^{j+1}} \rightarrow \frac{1}{2} \frac{\frac{1}{2} \times \frac{3}{2}}{\left(1 - \frac{1}{2}\right)^3} = 3,$$

donc $E(X^2) = 3$ et $V(X) = 3 - 1 = 2$.

On remarque que Y suit une loi de Poisson de paramètre 1, d'où $E(Y) = V(Y) = 1$.

6 Manifestement, on doit raisonner pour $i + j \leq n$ où $n \in \mathbb{N}$.

1. Nous avons

$$\begin{aligned} \sum_{j=0}^n \sum_{k=0}^{n-j} \frac{n! p_1^j p_2^k p_3^{n-j-k}}{j! k! (n-j-k)!} &= \sum_{j=0}^n \sum_{k=0}^{n-j} \binom{n}{j} \binom{n-j}{k} p_1^j p_2^k p_3^{n-j-k} \\ &= \sum_{j=0}^n \binom{n}{j} p_1^j \sum_{k=0}^{n-j} \binom{n-j}{k} p_2^k p_3^{n-j-k} \\ &= \sum_{j=0}^n \binom{n}{j} p_1^j (p_2 + p_3)^{n-j} = (p_1 + p_2 + p_3)^n = 1. \end{aligned}$$

On a bien défini la loi conjointe de deux variables aléatoires.

2. En reprenant le calcul précédent à j fixé, on trouve que pour tout $j \in \llbracket 0, n \rrbracket$,

$$P(X = j) = \binom{n}{j} p_1^j (1 - p_1)^{n-j}, \text{ autrement dit } X \hookrightarrow \mathcal{B}(n, p_1). \text{ En particulier } E(X) = np_1 \text{ et } V(X) = np_1(1 - p_1).$$

3. Pour $k \in \llbracket 0, j \rrbracket$, nous avons

$$\begin{aligned} P_{[X=j]}(Y = k) &= \frac{P([X = j] \cap [Y = k])}{P(X = j)} \\ &= \frac{\frac{n! p_1^j p_2^k p_3^{n-j-k}}{j! k! (n-j-k)!}}{\frac{n!}{j!(n-j)!} p_1^j (1-p_1)^{n-j}} \\ &= \frac{(n-j)! p_2^k p_3^{n-j-k}}{k!(n-j-k)! (1-p_1)^{n-j}} = \binom{n-j}{k} \frac{p_2^k p_3^{n-j-k}}{(1-p_1)^{n-j}}. \end{aligned}$$

On reconnaît une loi binomiale de paramètres $n-j$, $\frac{p_2}{1-p_1}$.

7 **1.** Soit $(i, j) \in \llbracket 1, n \rrbracket^2$ tel que $i < j$. Il y a $\binom{n}{2}$ façons de choisir les places des boules blanches, toutes équiprobables. Une seule telle que $([X = i] \cap [Y = j])$, d'où

$$P([X = i] \cap [Y = j]) = \frac{2}{n(n-1)}.$$

Naturellement, pour $i \geq j$, $P([X = i] \cap [Y = j]) = 0$.

2. On obtient alors

$$P(X = i) = \frac{2(n-i)}{n(n-1)} \text{ et } P(Y = j) = \frac{2(j-1)}{n(n-1)}.$$

3. On trouve

$$E(X) = \frac{n+1}{3}$$

$$V(X) = \frac{1}{18}(n+1)(n-2)$$

$$E(Y) = \frac{2(n+1)}{3}$$

$$V(Y) = \frac{1}{18}(n+1)(n-2)$$

$$E(XY) = \frac{1}{24}(n+1)(3n+2)$$

$$\rho_{X,Y} = -\frac{7n+10}{4(n-2)}.$$

4. La droite de régression de x en y a pour équation

$$y = -\frac{7n+10}{4(n-2)} \left(x - \frac{n+1}{3} \right) + \frac{2(n+1)}{3}$$

et celle de x en y

$$x = -\frac{7n+10}{4(n-2)} \left(y - \frac{2(n+1)}{3} \right) + \frac{n+1}{3}.$$

8 Tout d'abord $Y(\Omega) = \llbracket 0, m \rrbracket$.

Pour $i \in \llbracket 0, m \rrbracket$, on a d'après la formule des probabilités totales et l'hypothèse de l'énoncé,

$$\begin{aligned}
 P(Y = i) &= \sum_{k=0}^m P_{X=k}(Y = i)P(X = k) = \sum_{k=i}^m \binom{k}{i} \varpi^i (1 - \varpi)^{k-i} \binom{m}{k} p^k (1 - p)^{m-k} \\
 &= \sum_{k=i}^m \frac{k!}{i!(k-i)!} \times \frac{n!}{k!(m-k)!} \varpi^i (1 - \varpi)^{k-i} p^k (1 - p)^{m-k} \\
 &= \sum_{k=i}^m \frac{m!}{i!(m-i)!} \times \frac{(m-i)!}{(k-i)!(m-k)!} \varpi^i (1 - \varpi)^{k-i} p^k (1 - p)^{m-k} \\
 &= \binom{m}{i} \varpi^i p^i \sum_{k=i}^m \binom{m-i}{k-i} (1 - \varpi)^{k-i} p^{k-i} (1 - p)^{m-i-(k-i)} \\
 &= \binom{m}{i} \varpi^i p^i \sum_{k=0}^{m-i} \binom{m-i}{k} (p(1 - \varpi))^k (1 - p)^{m-i-k} \\
 &= \binom{m}{i} \varpi^i p^i (p(1 - \varpi) + (1 - p))^{m-i} \\
 &= \binom{m}{i} (p\varpi)^i (1 - p\varpi)^{m-i},
 \end{aligned}$$

et l'on constate que $Y \hookrightarrow \mathcal{B}(m, p\varpi)$.

9 1. Soit $(i, j) \in \mathbb{N}^*$. Si $j \in \llbracket 0, i \rrbracket$, alors

$$P([X = i] \cap [Y = j]) = P_{X=i}(Y = j)P(X = i) = \binom{i}{j} p^j (1 - p)^{i-j} \frac{\lambda^i}{i!} e^{-\lambda},$$

sinon $P([X = i] \cap [Y = j]) = 0$.

2. Par la formule des probabilités totales il vient pour $j \in \mathbb{N}$,

$$\begin{aligned}
 P(Y = j) &= \sum_{i=j}^{+\infty} P([X = i] \cap [Y = j]) = \sum_{i=j}^{+\infty} \binom{i}{j} p^j (1 - p)^{i-j} \frac{\lambda^i}{i!} e^{-\lambda} \\
 &= \sum_{i=j}^{+\infty} p^j (1 - p)^{i-j} \frac{\lambda^i}{j!(i-j)!} e^{-\lambda} \\
 &= \frac{p^j \lambda^j e^{-\lambda}}{j!} \sum_{i=j}^{+\infty} \frac{(\lambda(1 - p))^{i-j}}{(i-j)!} = \frac{p^j \lambda^j e^{-\lambda}}{j!} e^{\lambda(1-p)} \\
 &= \frac{(\lambda p)^j}{j!} e^{-\lambda p},
 \end{aligned}$$

et l'on constate que $Y \hookrightarrow \mathcal{P}(\lambda p)$.

10 1. On a $V(\Omega) = \mathbb{N}$ et $W(\Omega) = \mathbb{Z}$.

Pour $(i, j) \in \mathbb{N} \times \mathbb{Z}$, si $j \geq 0$, on a

$$\begin{aligned} P([V = i] \cap [W = j]) &= P([\inf(X, Y) = i] \cap [X - Y = j]) \\ &= P([Y = i] \cap [X = i + j]) = P(X = i + j)P(Y = i) \\ &= pq^{i+j}pq^i = p^2q^{2i+j}. \end{aligned}$$

Pour $j < 0$, on a

$$\begin{aligned} P([V = i] \cap [W = j]) &= P([\inf(X, Y) = i] \cap [X - Y = j]) \\ &= P([X = i] \cap [Y = i - j]) = P(X = i)P(Y = i - j) \\ &= pq^i pq^{i-j} = p^2q^{2i-j}. \end{aligned}$$

Or pour $i \in \mathbb{N}$,

$$\begin{aligned} P(V = i) &= P([X = i] \cap [Y = i] \cup [X > i] \cap [Y = i] \cup [X = i] \cap [Y > i]) \\ &= P(X = i)P(Y = i) + P(X > i)P(Y = i) + P(X = i)P(Y > i) \\ &= p^2q^{2i} + 2 \left(pq^i \sum_{k=i+1}^{+\infty} pq^k \right) \\ &= p^2q^{2i} + 2pq^i \times pq^{i+1} \times \frac{1}{1-q} \\ &= p^2q^{2i} + 2pq^{2i+1} = p^2q^{2i}(p+2q) \\ &= pq^{2i}(1+q). \end{aligned}$$

Pour $j \geq 0$,

$$\begin{aligned} P(W = j) &= P(X - Y = j) = P\left(\bigcup_{k \in \mathbb{N}} [X = j + k] \cap [Y = k]\right) \\ &= \sum_{k=0}^{+\infty} P([X = j + k] \cap [Y = k]) \\ &= \sum_{k=0}^{+\infty} pq^{j+k}pq^k \\ &= \sum_{k=0}^{+\infty} p^2q^{j+2k} \\ &= p^2q^j \times \frac{1}{1-q^2} \\ &= \frac{pq^j}{1+q}. \end{aligned}$$

Et l'on vérifie alors que $P([V = i] \cap [W = j]) = P(V = i)P(W = j)$.

Si $j < 0$, par symétrie et en reprenant le calcul précédent, on trouve

$$P(W = j) = P(Y - X = -j) = \frac{pq^{-j}}{1+q},$$

et là encore $P([V = i] \cap [W = j]) = P(V = i)P(W = j)$.

Les variables V et W sont indépendantes.

2. On a

$$\begin{aligned} \frac{P([X = n + 1] \cap [Y = n])}{P([X = n] \cap [Y = n])} &= \frac{P([V = n] \cap [W = 1])}{P([V = n] \cap [W = 0])} \\ &= \frac{P([V = n])P([W = 1])}{P([V = n])P([W = 0])} = k. \end{aligned}$$

Par ailleurs,

$$\frac{P([X = n + 1] \cap [Y = n])}{P([X = n] \cap [Y = n])} = \frac{P([X = n + 1])P([Y = n])}{P([X = n])P([Y = n])} = \frac{P(X = n + 1)}{P(X = n)}.$$

On en déduit que la suite $P(X = n)$ est géométrique de raison k , d'où $P(X = n) = P(X = 0)k^n$ et comme Y a la même loi, $P(Y = n) = P(Y = 0)k^n$. Autrement dit, il existe deux réels p et q tels que $P(X = n) = P(Y = n) = pq^n$ (et l'on vérifie que

$$q = 1 - p \text{ à l'aide de } \sum_{k=0}^{+\infty} P(X = k) = 1).$$

11 1. On considère que les variables X et Y sont indépendantes.

Si $S = s$, alors X prend ses valeurs dans $\llbracket 0, s \rrbracket$. Pour $k \in \llbracket 0, s \rrbracket$, on a

$$\begin{aligned} P(X = k | S = s) &= \frac{P([X = k] \cap [S = s])}{P(S = s)} = \frac{P([X = k] \cap [Y = s - k])}{\frac{(\lambda + \mu)^s}{s!} e^{-\lambda - \mu}} \\ &= \frac{s! P([X = k]) P([Y = s - k])}{(\lambda + \mu)^s e^{-\lambda - \mu}} \\ &= \frac{s! \frac{\lambda^k}{k!} e^{-\lambda} \frac{\mu^{s-k}}{(s-k)!} e^{-\mu}}{(\lambda + \mu)^s e^{-\lambda - \mu}} \\ &= \frac{1}{(\lambda + \mu)^s} \binom{s}{k} \lambda^k \mu^{s-k}. \end{aligned}$$

On voit que $X \hookrightarrow \mathcal{B}\left(s, \frac{\lambda}{\lambda + \mu}\right)$.

2. D'après ce qui précède, pour $s \in \mathbb{N}$, on a $P\left(E(X|S) = \frac{s\lambda}{\lambda + \mu}\right) = \frac{(\lambda + \mu)^s}{s!} e^{-\lambda - \mu}$.

3. En reprenant la question précédente, pour $n \in \mathbb{N}$,

$$\begin{aligned} \sum_{s=0}^n \frac{s\lambda}{\lambda + \mu} P\left(E(X|S) = \frac{s\lambda}{\lambda + \mu}\right) &= \sum_{s=1}^n \frac{s\lambda}{\lambda + \mu} \frac{(\lambda + \mu)^s}{s!} e^{-\lambda - \mu} \\ &= \sum_{s=1}^n \lambda \frac{(\lambda + \mu)^{s-1}}{(s-1)!} e^{-\lambda - \mu} \\ &= \lambda e^{-\lambda - \mu} \sum_{s=0}^{n-1} \frac{(\lambda + \mu)^s}{s!} \xrightarrow{+\infty} \lambda. \end{aligned}$$

On en déduit que $E(E(X|S)) = \lambda = E(X)$, ce qui confirme le théorème de l'espérance totale.

12 1. Chaque conducteur a une chance sur n de choisir le guichet numéro i . Les choix des automobilistes étant indépendants, $X_i \hookrightarrow \mathcal{B}(na, 1/n)$. On en déduit que $E(X) = a$ et $V(X) = \frac{(n-1)a}{n}$.

2. a) Chaque voiture passant par un et un seul guichet, on a $X_1 + \dots + X_n = na$, d'où $V(X_1 + \dots + X_n) = 0$.
b) On en déduit que

$$\sum_{k=1}^n V(X_k) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) = 0.$$

Les guichets jouant des rôles identiques, $\text{Cov}(X_i, X_j)$ ne dépend pas de i et j (pour $i \neq j$). Il s'ensuit que $(n-1)a + n(n-1)\text{Cov}(X_1, X_2) = 0$, d'où $\text{Cov}(X_1, X_2) = -\frac{a}{n}$.

3. a) D'après la question précédente, pour $i \neq j$,

$$\rho_{X_i, X_j} = \frac{-\frac{a}{n}}{\frac{(n-1)a}{n}} = -\frac{1}{n-1}.$$

- b) Si $n = 2$, on trouve $\rho_{X_1, X_2} = -1$. On ne s'en étonne pas puisque $X_2 = na - X_1$, autrement dit X_2 est une fonction affine de X_1 .

13 1. C'est la loi binomiale de paramètres j, p .

2. Pour $j \in \mathbb{N}$ et $i \in \llbracket 0, j \rrbracket$, on a

$$P([X = i] \cap [N = j]) = P(X = i | N = j)P(N = j) = \binom{j}{i} p^i q^{j-i} \times \frac{\lambda^j}{j!} e^{-\lambda}.$$

3. a) Soit $i \in \mathbb{N}$, on a

$$\begin{aligned} P(X = i) &= \sum_{j=i}^{+\infty} P([X = i] \cap [N = j]) = \sum_{j=i}^{+\infty} \binom{j}{i} p^i q^{j-i} \times \frac{\lambda^j}{j!} e^{-\lambda} \\ &= \sum_{j=i}^{+\infty} p^i q^{j-i} \times \frac{\lambda^j}{i!(j-i)!} e^{-\lambda} = \frac{\lambda^i p^i e^{-\lambda}}{i!} \sum_{j=i}^{+\infty} \frac{\lambda^{j-i} q^{j-i}}{(j-i)!} \\ &= \frac{\lambda^i p^i e^{-\lambda}}{i!} e^{\lambda q} = \frac{(\lambda p)^i}{i!} e^{-\lambda p}. \end{aligned}$$

On reconnaît une loi de Poisson de paramètre λp .

- b) On a $E(X) = V(X) = \lambda p$.
4. Pour $(i, j) \in \mathbb{N}^2$, on a

$$P([X = i] \cap [Y = j]) = P([X = i] \cap [N = i+j]) = \frac{p^i q^j}{i!j!} \lambda^{i+j} e^{-\lambda}.$$

D'autre part, on montrerait par un raisonnement analogue à celui de la question 3 que Y suit une loi de Poisson de paramètres λq . On en déduit que

$$P(X = i)P(Y = j) = \frac{(\lambda p)^i}{i!} e^{-\lambda p} \frac{(\lambda q)^j}{j!} e^{-\lambda q} = \frac{p^i q^j}{i!j!} \lambda^{i+j} e^{-\lambda}.$$

Et l'on constate que $P([X = i] \cap [Y = j]) = P(X = i)P(Y = j)$. Les variables X et Y sont indépendantes.

5. On a $\text{Cov}(X, N) = \text{Cov}(X, X + Y) = \text{Cov}(X, X) + \text{Cov}(X, Y) = V(X) = \lambda p$.

6. On en déduit que

$$\rho_{X,N} = \frac{\lambda p}{\sqrt{\lambda p} \sqrt{\lambda}} = \sqrt{p}.$$

14 1. Notons les boîtes A, B et C . Notons A_n l'événement « la boîte A est choisie au n^e placement », de même B_n, C_n .
Nous avons pour $k \geq 2$,

$$\begin{aligned} P(Y = k) &= P_{A_1}(Y = k)P(A_1) + P_{B_1}(Y = k)P(B_1) + P_{C_1}(Y = k)P(C_1) \\ &= 3 \times 2 \left(\frac{1}{3}\right)^{k-1} \times \frac{1}{3} = 2 \left(\frac{1}{3}\right)^{k-1}. \end{aligned}$$

2. a) Naturellement pour $\ell \leq k$, $P(Z = \ell | Y = k) = 0$.

Supposons maintenant que $\ell > k$. Supposons que $Y = k$. Il reste une seule boîte vide. Pour que $Z = \ell$, il faut ne pas avoir choisi cette boîte aux étapes $k + 1, k + 2, \dots, \ell - 1$ et l'avoir choisie à l'étape ℓ , d'où

$$P(Z = \ell | Y = k) = \left(\frac{2}{3}\right)^{\ell-k-1} \times \frac{1}{3}.$$

b) On en déduit par les probabilités totales que pour $\ell \geq 3$,

$$\begin{aligned} P(Z = \ell) &= \sum_{k=2}^{\ell-1} P(Z = \ell | Y = k)P(Y = k) = \sum_{k=2}^{\ell-1} \left(\frac{2}{3}\right)^{\ell-k-1} \times \frac{1}{3} \times 2 \left(\frac{1}{3}\right)^{k-1} \\ &= \left(\frac{1}{3}\right)^{\ell-1} 2^{\ell-2} \frac{1 - 2^{2-\ell}}{1 - \frac{1}{2}} = \left(\frac{2}{3}\right)^{\ell-1} (1 - 2^{2-\ell}). \end{aligned}$$

3. a) Pour $k \geq 2$,

$$\begin{aligned} E(Z | Y = k) &= \sum_{\ell=k+1}^{+\infty} \ell P(Z = \ell | Y = k) = \sum_{\ell=k+1}^{+\infty} \ell \left(\frac{2}{3}\right)^{\ell-k-1} \times \frac{1}{3} \\ &= \frac{1}{3} \sum_{\ell=0}^{+\infty} (\ell + k + 1) \left(\frac{2}{3}\right)^{\ell} = \frac{1}{3} \left(\frac{\frac{2}{3}}{\left(1 - \frac{2}{3}\right)^2} + (k + 1) \frac{1}{1 - \frac{2}{3}} \right) = k + 3. \end{aligned}$$

La variable aléatoire $E(Z | Y)$ prend donc comme valeurs les entiers naturels supérieurs ou égal à 5, et pour $\ell \geq 5$,

$$P(E(Z | Y) = \ell) = P(Y = \ell - 3) = 2 \left(\frac{1}{3}\right)^{\ell-4}.$$

b) Par un calcul direct

$$\begin{aligned}
 E(Z) &= \sum_{\ell=3}^{+\infty} \ell P(Z = \ell) = \sum_{\ell=3}^{+\infty} \ell \left(\frac{2}{3}\right)^{\ell-1} (1 - 2^{2-\ell}) \\
 &= \sum_{\ell=0}^{+\infty} (\ell + 3) \left(\left(\frac{2}{3}\right)^{\ell+2} - 2 \left(\frac{1}{3}\right)^{\ell+2} \right) \\
 &= \frac{4}{9} \sum_{\ell=0}^{+\infty} \ell \left(\frac{2}{3}\right)^{\ell} + \frac{4}{3} \sum_{\ell=0}^{+\infty} \left(\frac{2}{3}\right)^{\ell} - \frac{2}{9} \sum_{\ell=0}^{+\infty} \ell \left(\frac{1}{3}\right)^{\ell} - \frac{2}{3} \sum_{\ell=0}^{+\infty} \left(\frac{1}{3}\right)^{\ell} \\
 &= \frac{4}{9} \frac{\frac{2}{3}}{\left(1 - \frac{2}{3}\right)^2} + \frac{4}{3} \frac{1}{1 - \frac{2}{3}} - \frac{2}{9} \frac{\frac{1}{3}}{\left(1 - \frac{1}{3}\right)^2} - \frac{2}{3} \frac{1}{1 - \frac{1}{3}} \\
 &= \frac{8}{3} + 4 - \frac{1}{6} - 1 = \frac{11}{2}.
 \end{aligned}$$

On a également d'après le théorème de l'espérance totale

$$\begin{aligned}
 E(Z) &= \sum_{\ell=5}^{+\infty} \ell \times 2 \left(\frac{1}{3}\right)^{\ell-4} = \sum_{\ell=0}^{+\infty} 2(\ell + 5) \left(\frac{1}{3}\right)^{\ell+1} \\
 &= \frac{2}{3} \sum_{\ell=0}^{+\infty} \ell \left(\frac{1}{3}\right)^{\ell} + \frac{10}{3} \sum_{\ell=0}^{+\infty} \left(\frac{1}{3}\right)^{\ell} \\
 &= \frac{2}{3} \frac{\frac{1}{3}}{\left(1 - \frac{1}{3}\right)^2} + \frac{10}{3} \frac{1}{1 - \frac{1}{3}} = \frac{1}{2} + 5 = \frac{11}{2}.
 \end{aligned}$$

15 1. On a naturellement $R_n + V_n = n$, d'où $D_n = 2R_n - n$.

On en déduit que $D_n(\Omega) = \{2k - n \mid k \in \llbracket 0, n \rrbracket\}$.

Pour $k \in \llbracket 0, n \rrbracket$,

$$P(D_n = 2k - n) = P(R_n = k) = \binom{n}{k} p^k q^{n-k}.$$

On a $E(D_n) = 2E(R_n) - n = 2np - n = n(2p - 1)$ et $V(D_n) = 4V(R_n) = 4np(1 - p)$.

2. Les variables sont clairement indépendantes, et elle suivent une loi de Bernoulli de paramètre p . La matrice recherchée est donc $p(1 - p)I_n$.

Chapitre 10

1 En enlevant la valeur absolue, on obtient

$$\begin{aligned}
 \varphi(a) &= \int_{-\infty}^a (a - t)f(t) dt + \int_a^{+\infty} (t - a)f(t) dt \\
 &= a \int_{-\infty}^a f(t) dt - \int_{-\infty}^a tf(t) dt + \int_a^{+\infty} tf(t) dt - a \int_a^{+\infty} f(t) dt.
 \end{aligned}$$

Compte tenu de $\int_{-\infty}^{+\infty} f(t) dt = 1$ et $\int_{-\infty}^{+\infty} tf(t) dt = E(X)$, on en déduit

$$\begin{aligned} \varphi(a) &= a \int_{-\infty}^a f(t) dt - \int_{-\infty}^a tf(t) dt + E(X) - \int_{-\infty}^a tf(t) dt - a + a \int_{-\infty}^a f(t) dt \\ &= 2aF(a) - 2 \int_{-\infty}^a tf(t) dt - a + E(X), \end{aligned}$$

où F est la fonction de répartition de X . Puisque f est continue sur \mathbb{R} , F et $t \mapsto \int_{-\infty}^t tf(t) dt$ sont dérivables sur \mathbb{R} et pour tout réel a ,

$$\varphi'(a) = 2F(a) + 2af(a) - 2af(a) - 1 = 2F(a) - 1.$$

La fonction F est croissante car $F' = f$.

• Si la densité f ne s'annule qu'en des points isolés, F est strictement croissante et comme $\lim_{-\infty} F = 0$ et $\lim_{+\infty} F = 1$, elle réalise une bijection de \mathbb{R} sur $]0, 1[$ et il existe un unique réel a_0

tel que $F(a_0) = \frac{1}{2}$. Comme F croît, la dérivée est négative avant a_0 et positive après donc φ atteint son minimum en a_0 .

• Si F n'est pas strictement croissante, l'ensemble des $a \in \mathbb{R}$ tels que $F(a) = \frac{1}{2}$ est un intervalle car si $F(a) = F(b) = \frac{1}{2}$, alors $F(x) = \frac{1}{2}$ pour tout $x \in [a, b]$. Cet intervalle est borné car $\lim_{-\infty} F = 0$ et $\lim_{+\infty} F = 1$ et fermé car F est continue. Il est de la forme $[\alpha, \beta]$. La fonction φ atteint son minimum en tout point de l'intervalle $[\alpha, \beta]$.

2 La variable X admet une variance donc une espérance et un moment d'ordre 2 et par linéarité de l'intégrale

$$\begin{aligned} \psi(a) &= \int_{-\infty}^{+\infty} (t^2 - 2at + a^2)f(t) dt \\ &= \int_{-\infty}^{+\infty} t^2f(t) dt - 2a \int_{-\infty}^{+\infty} tf(t) dt + a^2 \int_{-\infty}^{+\infty} f(t) dt \\ &= E(X^2) - 2aE(X) + a^2 \\ &= (a - E(X))^2 + E(X^2) - (E(X))^2 \\ &= (a - E(X))^2 + V(X). \end{aligned}$$

La fonction ψ atteint donc son minimum en $a = E(X)$ et ce minimum vaut $V(X)$.

3 1. La variable X est à valeurs dans \mathbb{R}_+ donc, en appelant f une densité,

$$E(X) = \int_0^{+\infty} tf(t) dt. \text{ On en déduit, pour } x > 0,$$

$$0 \leq x(1 - F_X(x)) = x \int_x^{+\infty} f(t) dt \leq \int_x^{+\infty} tf(t) dt.$$

Comme $\int_0^{+\infty} tf(t) dt$ converge, le reste $\int_x^{+\infty} tf(t) dt$ tend vers 0 quand x tend vers $+\infty$ et, par encadrement

$$\lim_{x \rightarrow +\infty} x(1 - F_X(x)) = 0.$$

2. Si f est continue sur $]0, \infty[$, on obtient, en intégrant par parties, pour tout $x > 0$,

$$\int_0^x (1 - F_X(t)) dt = x - [tF_X(t)]_0^x + \int_0^x tf(t) dt = x(1 - F_X(x)) + \int_0^x tf(t) dt.$$

L'égalité $\int_0^x (1 - F_X(t)) dt = x(1 - F_X(x)) + \int_0^x tf(t) dt$ reste vraie si f est discontinue en un nombre fini de points. En effet les deux membres de l'égalité sont continus sur $]0, +\infty[$, dérivable sauf aux points de discontinuité de f et ont même dérivée $x \mapsto 1 - F_X(x)$. Ces deux fonctions diffèrent d'une constante et comme elles ont même limite 0 en 0 elles sont égales.

- Supposons que $E(X)$ existe. En utilisant la question 1, on obtient

$$\begin{aligned} \lim_{x \rightarrow +\infty} \int_0^{+\infty} (1 - F_X(t)) dt &= \lim_{x \rightarrow +\infty} \left(x(1 - F_X(x)) + \int_0^x tf(t) dt \right) \\ &= \int_0^{+\infty} tf(t) dt = E(X). \end{aligned}$$

L'intégrale $\int_0^{+\infty} (1 - F_X(t)) dt$ converge et vaut $E(X)$.

- Supposons réciproquement que $\int_0^{+\infty} (1 - F_X(t)) dt$ converge. On a alors comme les fonctions $t \mapsto tf(t)$ et $t \mapsto (1 - F_X(t))$ sont positives,

$$\begin{aligned} \int_0^x tf(t) dt &= \int_0^x (1 - F_X(t)) dt - x(1 - F_X(x)) \\ &\leq \int_0^x (1 - F_X(t)) dt \leq \int_0^{+\infty} (1 - F_X(t)) dt. \end{aligned}$$

La fonction $x \mapsto \int_0^x tf(t) dt$ est donc majorée sur $]0, +\infty[$. On en déduit que $\int_0^{+\infty} tf(t) dt$ converge et donc que $E(X)$ existe. D'après la première partie de la démonstration, on a

$$E(X) = \int_0^{+\infty} (1 - F_X(t)) dt.$$

3. a) On applique ce qui précède à la variable X^2 . On sait que X^2 est encore une variable à densité, qui est à valeurs dans \mathbb{R}_+ . On a, pour $x > 0$,

$$P(|X| \geq x) = P(X^2 \geq x^2) = 1 - F_{X^2}(x^2).$$

On en déduit $\lim_{x \rightarrow +\infty} P(|X| \geq x) = 0$, car la limite de F_{X^2} en $+\infty$ est 1.

b) D'après la question 2 et l'expression de $1 - F_{X^2}$ vue dans a, on a

$$E(X^2) = \int_0^{+\infty} (1 - F_{X^2}(x)) dx = \int_0^{+\infty} P(|X| \geq \sqrt{x}) dx.$$

En faisant le changement de variable $t = \sqrt{x}$ dans l'intégrale, on obtient

$$E(X^2) = \int_0^{+\infty} P(|X| \geq t)(2tdt) = 2 \int_0^{+\infty} tP(|X| \geq t) dt.$$

4

1. a) La fonction f_X est continue sur \mathbb{R} sauf en 0. Il faut montrer que

$$\int_{-\infty}^{+\infty} f_X(x) dx = \int_0^{+\infty} \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}(\ln x-1)^2} dx$$

converge. On fait le changement de variable défini par $\ln x = y + 1$ ou $x = e^{y+1}$ qui définit une bijection croissante de \mathbb{R}_+^* sur \mathbb{R} . Le théorème de changement de variable dans les intégrales impropres dit que $\int_0^{+\infty} \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}(\ln x-1)^2} dx$ a même nature et en cas de convergence même valeur que

$$\int_{-\infty}^{+\infty} \frac{1}{e^{y+1}\sqrt{2\pi}} e^{-\frac{1}{2}y^2} e^{y+1} dy = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy.$$

Cette dernière intégrale converge et vaut 1 car on reconnaît l'intégrale sur \mathbb{R} de la densité de la normale centrée réduite. La fonction f_X est bien une densité de probabilité.

b) Le même changement de variable montre que

$$\begin{aligned} E(X) &= \int_0^{+\infty} x \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}(\ln x-1)^2} dx = \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\ln x-1)^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2} e^{y+1} dy \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(y-1)^2} e^{\frac{3}{2}} dy = e^{\frac{3}{2}}. \end{aligned}$$

On obtient, de même,

$$\begin{aligned} E(X^2) &= \int_0^{+\infty} x^2 \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}(\ln x-1)^2} dx = \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} x e^{-\frac{1}{2}(\ln x-1)^2} dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{y+1} e^{-\frac{1}{2}y^2} e^{y+1} dy = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2+2y+2} dy \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-2)^2} e^4 dy = e^4. \end{aligned}$$

On en déduit que

$$V(X) = E(X^2) - (E(X))^2 = e^4 - e^3.$$

2. a) La fonction \exp est de classe \mathcal{C}^1 sur \mathbb{R} , possède en tout point une dérivée strictement positive et réalise une bijection de \mathbb{R} sur $]0, +\infty[$. D'après le premier théorème de transfert, X est une variable aléatoire qui admet une densité, nulle sur \mathbb{R}_- et définie sur \mathbb{R}_+^* par $g(x) = f_Y(\ln(x)) \ln'(x) = \frac{1}{x} f_Y(\ln(x))$, où f_Y est la densité de la variable Y , c'est-à-dire

$$g(x) = \frac{1}{x\sqrt{2\pi\sigma}} e^{-\frac{(\ln x - m)^2}{2\sigma^2}}.$$

- b) D'après le second théorème de transfert, on a, si l'intégrale converge,

$$E(X) = E(e^Y) = \int_{-\infty}^{+\infty} e^t \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(t-m)^2}{2\sigma^2}} dt.$$

On transforme l'exposant de e en écrivant

$$\begin{aligned} t - \frac{(t-m)^2}{2\sigma^2} &= -\frac{1}{2\sigma^2} (t^2 - 2(m+\sigma^2)t + m^2) \\ &= -\frac{1}{2\sigma^2} \left((t - (m + \sigma^2))^2 - (\sigma^4 + 2m\sigma^2) \right) \\ &= -\frac{1}{2\sigma^2} (t - (m + \sigma^2))^2 + \frac{\sigma^2}{2} + m. \end{aligned}$$

On en déduit la convergence de l'intégrale et l'égalité

$$\begin{aligned} E(X) &= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{+\infty} e^{-\frac{(t-(m+\sigma^2))^2}{2\sigma^2}} e^{\frac{\sigma^2}{2} + m} dt \\ &= e^{\frac{\sigma^2}{2} + m}. \end{aligned}$$

De même, on a, sous réserve de convergence

$$E(X^2) = E(e^{2Y}) = \int_{-\infty}^{+\infty} e^{2t} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(t-m)^2}{2\sigma^2}} dt.$$

On écrit de même,

$$\begin{aligned} 2t - \frac{(t-m)^2}{2\sigma^2} &= -\frac{1}{2\sigma^2} (t^2 - 2(m+2\sigma^2)t + m^2) \\ &= -\frac{1}{2\sigma^2} \left((t - (m + 2\sigma^2))^2 - (4\sigma^4 + 4m\sigma^2) \right) \\ &= -\frac{1}{2\sigma^2} (t - (m + 2\sigma^2))^2 + 2\sigma^2 + 2m. \end{aligned}$$

On en déduit la convergence de l'intégrale et l'égalité

$$E(X^2) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{+\infty} e^{-\frac{(t-(m+2\sigma^2))^2}{2\sigma^2}} e^{2\sigma^2+2m} dt = e^{2\sigma^2+2m}.$$

On obtient enfin

$$V(X) = E(X^2) - (E(X))^2 = e^{2\sigma^2+2m} - e^{\sigma^2+2m} = (e^{\sigma^2} - 1)e^{\sigma^2+2m}.$$

3. a) • Si $a > 0$, on écrit

$$Z = aX^b = e^{\ln a} e^{bY} = e^{\ln a + bY}.$$

La variable $\ln a + bY$ qui est une fonction affine de Y suit une loi normale. Comme on a

$$E(\ln a + bY) = \ln a + bE(Y) = \ln a + bm \quad \text{et} \quad V(\ln a + bY) = b^2 V(Y) = (b\sigma)^2,$$

$\ln a + bY$ suit une loi normale de paramètre $(\ln a + bm, (b\sigma)^2)$. Par définition, la variable Z suit une Log-normale de paramètre $(\ln a + bm, (b\sigma)^2)$.

• Si $a < 0$, $Z = -|a|X^b$. La variable $-Z$ suit une loi Log-normale de paramètre $(\ln |a| + bm, (b\sigma)^2)$. La densité de Z est $x \mapsto -g(-x)$ où g est la densité de la variable $-Z$.

• Si $a = 0$, Z est la variable nulle.

b) Dans tous les cas, on a

$$E(Z) = aE(X^b) = aE(e^{bY}) \quad \text{et} \quad V(Z) = a^2 V(e^{bY}).$$

La variable bY suit une loi normale de paramètre $(bm, (b\sigma)^2)$, donc e^{bY} suit une loi Log-normale de paramètre $(bm, (b\sigma)^2)$. On a donc

$$E(Z) = ae^{\frac{(b\sigma)^2}{2} + bm} \quad \text{et} \quad V(Z) = a^2 V(e^{bY}) = a^2 \left(e^{(b\sigma)^2} - 1 \right) e^{(b\sigma)^2 + 2bm}.$$

5 La fonction f_X est définie sur \mathbb{R} , positive si $\lambda \geq 0$ et continue sauf au point 1. Il reste à déterminer λ pour que son intégrale sur \mathbb{R} soit égale à 1. Pour $x \geq 1$, on a

$$f_X(x) = \frac{\lambda}{x^2 - \frac{1}{9}} = \frac{\lambda}{\left(x - \frac{1}{3}\right)\left(x + \frac{1}{3}\right)} = \frac{3\lambda}{2} \left(\frac{1}{x - \frac{1}{3}} - \frac{1}{x + \frac{1}{3}} \right).$$

On en déduit que pour $A > 1$,

$$\int_1^A \frac{\lambda}{x^2 - \frac{1}{9}} dx = \left[\frac{3\lambda}{2} \ln \frac{x - \frac{1}{3}}{x + \frac{1}{3}} \right]_1^A = \frac{3\lambda}{2} \left(\ln \frac{A - \frac{1}{3}}{A + \frac{1}{3}} + \ln 2 \right).$$

On obtient

$$\int_{-\infty}^{+\infty} f_X(x) dx = \int_1^{+\infty} \frac{\lambda}{x^2 - \frac{1}{9}} dx = \lim_{A \rightarrow +\infty} \frac{3\lambda}{2} \left(\ln \frac{A - \frac{1}{3}}{A + \frac{1}{3}} + \ln 2 \right) = \frac{3\lambda \ln 2}{2}.$$

La fonction f_X est une densité de probabilité si $\lambda = \frac{2}{3 \ln 2}$.

L'espérance X est, en cas de convergence de l'intégrale,

$$\int_{-\infty}^{+\infty} x f_X(x) dx = \int_1^{+\infty} \frac{\lambda x}{x^2 - \frac{1}{9}} dx.$$

Comme $\frac{\lambda x}{x^2 - \frac{1}{9}} \underset{x \rightarrow +\infty}{\sim} \frac{\lambda}{x}$, cette intégrale diverge : X ne possède pas d'espérance.

6 Énonçons un résultat utile pour la suite. La fonction sin est une fonction strictement croissante et continue sur $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Elle réalise une bijection de $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ sur $[-1, 1]$. On note Arcsin sa bijection réciproque. La fonction sin est dérivable et $\sin' = \cos$ ne s'annule qu'en $-\frac{\pi}{2}$ et $\frac{\pi}{2}$. Donc Arcsin est dérivable sur $] -1, 1[$ et pour tout $x \in] -1, 1[$,

$$\text{Arcsin}'(x) = \frac{1}{\cos(\text{Arcsin}(x))} = \frac{1}{\sqrt{1-x^2}}.$$

Si $x \in [0, 1]$, alors $\sin(\pi x) \in [0, 1]$. La variable Y est à valeurs dans $[0, 1]$. On a donc pour $x < 0$, $F_Y(x) = 0$ et pour $x \geq 1$, $F_Y(x) = 1$. Si $x \in [0, 1]$, on a

$$\begin{aligned} P(Y \leq x) &= P(\sin(\pi X) \leq x) \\ &= P(0 \leq \pi X \leq \text{Arcsin } x) + P(\pi - \text{Arcsin } x \leq \pi X \leq \pi) \\ &= P\left(0 \leq X \leq \frac{1}{\pi} \text{Arcsin } x\right) + P\left(1 - \frac{1}{\pi} \text{Arcsin } x \leq X \leq 1\right) \\ &= \frac{2 \text{Arcsin } x}{\pi}, \end{aligned}$$

car X suit une loi uniforme sur $[0, 1]$. La fonction F_Y est continue sur \mathbb{R} est dérivable sauf en 0 et 1. Elle possède donc une densité f définie par

$$f(x) = \begin{cases} 0 & \text{si } x \notin]0, 1[\\ \frac{2}{\pi\sqrt{1-x^2}} & \text{si } x \in]0, 1[. \end{cases}$$

On calcule

$$E(X) = \int_0^1 \frac{2x}{\pi\sqrt{1-x^2}} dx = \left[-\frac{2}{\pi} \sqrt{1-x^2}\right]_0^1 = \frac{2}{\pi}.$$

Pour calculer $E(X^2) = \int_0^1 \frac{2x^2}{\pi\sqrt{1-x^2}} dx$, on fait le changement de variable $x = \sin t$ et l'on obtient

$$E(X^2) = \int_0^{\frac{\pi}{2}} \frac{2 \sin^2 t}{\pi \cos t} \cos t dt = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} 2 \sin^2 t dt = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} (1 - \cos(2t)) dt = \frac{1}{2}.$$

On en déduit que

$$V(X) = E(X^2) - (E(X))^2 = \frac{1}{2} - \frac{4}{\pi^2}.$$

7 La fonction tangente réalise est bijection strictement croissante de $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$ dont la fonction réciproque est Arctan. On a, pour tout réel x

$$F_Y(x) = P(Y \leq x) = P(\tan X \leq x) = P(X \leq \text{Arctan } x) = \frac{\text{Arctan } x + \frac{\pi}{2}}{\pi},$$

car X suit une loi uniforme sur $\left] -\frac{\pi}{2}, \frac{\pi}{2} \right]$. La fonction F_Y est de classe \mathcal{C}^1 sur \mathbb{R} donc Y admet une densité définie par

$$f_Y(x) = F'_Y(x) = \frac{1}{\pi(1+x^2)}.$$

La variable Y suit donc une loi de Cauchy. On sait qu'elle ne possède ni espérance, ni variance.

- 8** 1. La fonction F est définie sur \mathbb{R} , croissante au sens large : en effet, elle est nulle sur \mathbb{R}_- et elle croît sur \mathbb{R}_+ , car $a > 0$. Les limites de F en $-\infty$ et $+\infty$ sont 0 et 1. Enfin, F est continue sur \mathbb{R} car $\lim_{t \rightarrow 0^+} F(t) = 1 - e^0 = 0$. Donc F est une fonction de répartition.
2. La fonction F est continue sur \mathbb{R} et \mathcal{C}^1 sauf en 0, donc X est une variable à densité. Une densité f s'obtient en dérivant F . On obtient

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ \left(\frac{a}{b}\right) \left(\frac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^a} & \text{si } x \geq 0. \end{cases}$$

- 3.
4. Traitons en même temps les questions 3 et 4.

Par croissances comparées, on a, pour tout $n \in \mathbb{N}^*$, $x^n f(x) \underset{x \rightarrow +\infty}{=} o\left(\frac{1}{x^2}\right)$ donc $\int_0^{+\infty} x^n f(x) dx$ converge. Ainsi X possède des moments de tout ordre. On obtient, pour $n \in \mathbb{N}^*$,

$$\begin{aligned} E(X^n) &= \int_0^{+\infty} x^n f(x) dx = \int_0^{+\infty} \left(\frac{a}{b}\right) \left(\frac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^a} x^n dx \\ &= ab^{n-1} \int_0^{+\infty} \left(\frac{x}{b}\right)^{a+n-1} e^{-\left(\frac{x}{b}\right)^a} dx. \end{aligned}$$

On fait le changement de variable $t = \left(\frac{x}{b}\right)^a$ ou $x = bt^{\frac{1}{a}}$. On obtient

$$E(X^n) = ab^n \int_0^{+\infty} t^{\frac{a+n-1}{a}} e^{-t} \left(\frac{1}{a} t^{\frac{1}{a}-1} dt\right) = b^n \int_0^{+\infty} t^{\frac{n}{a}} e^{-t} dt = b^n \Gamma\left(\frac{n}{a} + 1\right).$$

On a en particulier

$$E(X) = b\Gamma\left(\frac{1}{a} + 1\right) \quad \text{et} \quad E(X^2) = b^2\Gamma\left(\frac{2}{a} + 1\right).$$

- 9** 1. La fonction f est positive, continue sur \mathbb{R} et pour tout $x \in \mathbb{R}_+$, on a

$$\int_{-\infty}^x f(t) dt = \int_0^x te^{-\frac{t}{2}} = \left[-e^{-\frac{t}{2}}\right]_0^x = 1 - e^{-\frac{x}{2}}.$$

On en déduit

$$\lim_{x \rightarrow +\infty} \int_{-\infty}^x f(t) dt = 1.$$

Donc $\int_{-\infty}^{+\infty} f(t) dt$ converge et vaut 1. La fonction f est une densité de probabilité.

2. La fonction de répartition est nulle sur \mathbb{R}_- et, d'après ce qui précède, pour $x \geq 0$,

$$F(x) = 1 - e^{-\frac{x^2}{2}}.$$

3. Comme au voisinage de $+\infty$, $x^2 f(x) = o\left(\frac{1}{x^2}\right)$, l'intégrale $\int_0^{+\infty} x^2 f(x) dx$ converge. La variable X possède un moment d'ordre 2 et donc une espérance et une variance. On obtient,

$$E(X) = \int_0^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{\sqrt{2\pi}}{2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2}} dx = \frac{\sqrt{2\pi}}{2} E(Z^2),$$

où Z suit la loi normale centrée réduite. De $E(Z^2) = 1$, on déduit

$$E(X) = \sqrt{\frac{\pi}{2}}.$$

Pour calculer le moment d'ordre 2, on intègre par parties. On obtient

$$\begin{aligned} E(X^2) &= \int_0^{+\infty} x^3 e^{-\frac{x^2}{2}} dx = \left[-x^2 e^{-\frac{x^2}{2}}\right]_0^{+\infty} + \int_0^{+\infty} 2xe^{-\frac{x^2}{2}} dx \\ &= 2 \int_0^{+\infty} f(x) dx = 2. \end{aligned}$$

On en déduit

$$V(X) = 2 - \frac{\pi}{2}.$$

4. La variable Y est à valeurs dans \mathbb{R}_+ . On a donc $F_Y(x) = 0$ si $x < 0$. Pour $x \geq 0$, on obtient, puisque X est à valeurs dans \mathbb{R}_+ ,

$$F_Y(x) = P(Y \leq x) = P(X \leq \sqrt{x}) = F_X(\sqrt{x}) = 1 - e^{-\frac{x}{2}}.$$

Ainsi Y suit la loi exponentielle de paramètre $\frac{1}{2}$. On en déduit

$$E(Y) = 2 \quad \text{et} \quad V(Y) = 4.$$

- 10** 1. a) Pour $\lambda \geq 0$, la fonction f est positive sur \mathbb{R} . Elle est continue sauf en x_0 . Il faut déterminer λ pour que

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{x_0+a}^{+\infty} f(x) dx = 1.$$

Cette intégrale converge car f est continue sur $[x_0 + a, +\infty[$ et au voisinage de $+\infty$, $f(x) \sim \frac{\lambda a^{\alpha+1}}{x^{\alpha+1}}$, avec $\alpha + 1 > 1$. On obtient

$$\int_{-\infty}^{+\infty} f(x) dx = \left[-\frac{\lambda a^{\alpha+1}}{\alpha} (x - x_0)^{-\alpha} \right]_{x_0+a}^{+\infty} = \frac{\lambda a^{\alpha+1}}{\alpha} a^{-\alpha} = \frac{\lambda a}{\alpha}.$$

La fonction f est une densité de probabilité si $\lambda = \frac{\alpha}{a}$.

- b)** Notons F la fonction de répartition de X . Comme f est nulle sur $] -\infty, x_0 + a]$, on a $F(x) = 0$ si $x \leq x_0 + a$.
Si $x > x_0 + a$, on obtient

$$\begin{aligned} F(x) &= \int_{x_0+a}^x \frac{\alpha}{a} \cdot \left(\frac{a}{t - x_0} \right)^{\alpha+1} dt = \left[-a^\alpha (t - x_0)^{-\alpha} \right]_{x_0+a}^x \\ &= -a^\alpha (x - x_0)^{-\alpha} + 1 = 1 - \left(\frac{a}{x - x_0} \right)^\alpha. \end{aligned}$$

La fonction F est donc définie par

$$F(x) = \begin{cases} 0 & \text{si } x \leq x_0 + a \\ 1 - \left(\frac{a}{x - x_0} \right)^\alpha & \text{si } x > x_0 + a. \end{cases}$$

Pour alléger les calculs qui suivent, on pose pour $\alpha > 0$, $I_\alpha = \int_{x_0+a}^{+\infty} \left(\frac{a}{x - x_0} \right)^{\alpha+1} dx$.

D'après la première question, cette intégrale vaut $\frac{a}{\alpha}$.

- c)** L'espérance $E(X)$ existe si $\int_{-\infty}^{+\infty} xf(x) dx = \int_{x_0+a}^{+\infty} \frac{\alpha x}{a} \cdot \left(\frac{a}{x - x_0} \right)^{\alpha+1} dx$ converge. Le

seul problème de convergence est en $+\infty$ et $xf(x) \underset{x \rightarrow +\infty}{\sim} \frac{\alpha a^\alpha}{x^\alpha}$. L'espérance existe si, et seulement si, $\alpha > 1$.

Il est plus simple de calculer

$$\begin{aligned} E(X - x_0) &= \int_{x_0+a}^{+\infty} \frac{\alpha}{a} (x - x_0) \left(\frac{a}{x - x_0} \right)^{\alpha+1} dx \\ &= \alpha \int_{x_0+a}^{+\infty} \left(\frac{a}{x - x_0} \right)^\alpha dx \\ &= \alpha I_{\alpha-1} = \frac{\alpha a}{\alpha - 1}. \end{aligned}$$

On en déduit

$$E(X) = x_0 + E(X - x_0) = x_0 + \frac{\alpha a}{\alpha - 1}.$$

d) De même, X admet un moment d'ordre 2 et donc une variance si

l'intégrale $\int_{x_0+a}^{+\infty} \frac{\alpha x^2}{a} \cdot \left(\frac{a}{x-x_0}\right)^{\alpha+1} dx$ converge. Comme $x^2 f(x) \underset{x \rightarrow +\infty}{\sim} \frac{\alpha a^\alpha}{x^{\alpha-1}}$, cela est réalisé si, et seulement si, $\alpha > 2$. On procède comme pour l'espérance. On calcule

$$\begin{aligned} E((X-x_0)^2) &= \int_{x_0+a}^{+\infty} \frac{\alpha}{a} (x-x_0)^2 \left(\frac{a}{x-x_0}\right)^{\alpha+1} dx \\ &= \alpha a \int_{x_0+a}^{+\infty} \left(\frac{a}{x-x_0}\right)^{\alpha-1} dx \\ &= \alpha a I_{\alpha-2} = \frac{\alpha a^2}{\alpha-2}. \end{aligned}$$

Par linéarité de l'espérance, on en déduit

$$E(X^2) = E((X-x_0)^2) + 2x_0 E(X) - x_0^2 = \frac{\alpha a^2}{\alpha-2} + 2 \frac{\alpha a x_0}{\alpha-1} + x_0^2$$

et

$$V(X) = E(X^2) - (E(X))^2 = \frac{\alpha a^2}{\alpha-2} - \frac{\alpha^2 a^2}{(\alpha-1)^2} = \frac{a^2 \alpha}{(\alpha-1)^2 (\alpha-2)}.$$

2. La variable Y est à valeurs dans $]rx_0 + a) + s, +\infty[=]rx_0 + s) + ra, +\infty[$. On a donc

$$F_Y(x) = 0 \quad \text{si } x \leq rx_0 + s + ra.$$

Pour $x > rx_0 + a$, on obtient

$$\begin{aligned} F_Y(x) &= P(rX + s \leq x) = P\left(X \leq \frac{x-s}{r}\right) \\ &= 1 - \left(\frac{a}{\frac{x-s}{r} - x_0}\right)^\alpha \\ &= 1 - \left(\frac{ra}{x - (rx_0 + s)}\right)^\alpha, \end{aligned}$$

car $\frac{x-s}{r} > x_0 + a$. D'après la question 2, la variable Y suit donc la loi de Pareto de paramètre $(\alpha, ra, rx_0 + s)$.

3. a) La variable X est à valeurs dans $]0, +\infty[$. Comme la fonction $x \mapsto \beta \gamma^x$ réalise une bijection de $]0, +\infty[$ sur $] \beta, +\infty[$ la variable Y est à valeurs dans $] \beta, +\infty[$. On a donc $F_Y(x) = 0$ si $x \leq \beta$. Pour $x > \beta$, on obtient

$$\begin{aligned} F_Y(x) &= P(Y \leq x) = P(\beta \gamma^X \leq x) = P\left(\gamma^X \leq \frac{x}{\beta}\right) \\ &= P(X \ln \gamma \leq \ln \frac{x}{\beta}) = F_X\left(\frac{\ln \frac{x}{\beta}}{\ln \gamma}\right). \end{aligned}$$

Comme $x > \beta$ et $\ln \gamma > 0$, on a $\frac{\ln \frac{x}{\beta}}{\ln \gamma} > 0$ et, par définition de la loi exponentielle,

$$\begin{aligned} F_Y(x) &= 1 - \exp\left(-\mu \frac{\ln \frac{x}{\beta}}{\ln \gamma}\right) \\ &= 1 - \exp\left(\frac{-\mu}{\ln \gamma} \ln \frac{x}{\beta}\right) \\ &= 1 - \left(\frac{x}{\beta}\right)^{\frac{-\mu}{\ln \gamma}} = 1 - \left(\frac{\beta}{x}\right)^{\frac{\mu}{\ln \gamma}}. \end{aligned}$$

On note que $\frac{\mu}{\ln \gamma} > 0$. D'après la question 2, la variable Y suit la loi de Pareto de paramètre $\left(\frac{\mu}{\ln \gamma}, \beta, 0\right)$.

- b)** Supposons réciproquement que la variable Y suit une loi de Pareto de paramètre $(\alpha, \beta, 0)$. Considérons $\gamma > 1$ et la variable X définie par $Y = \beta \gamma^X$, c'est-à-dire $X = \frac{\ln Y - \ln \beta}{\ln \gamma}$. Cette définition a un sens car Y est à valeurs dans $]\beta, +\infty[\subset]0, +\infty[$.

De plus, la fonction $y \mapsto \frac{\ln y - \ln \beta}{\ln \gamma}$ réalise une bijection de $]\beta, +\infty[$ sur $]0, +\infty[$ donc X est à valeurs dans $]0, +\infty[$. On a donc $F_X(x) = 0$ si $x \leq 0$ et pour $x > 0$,

$$\begin{aligned} F_X(x) &= P(X \leq x) = P\left(\frac{\ln Y - \ln \beta}{\ln \gamma} \leq x\right) \\ &= P\left(\ln \frac{Y}{\beta} \leq \ln \gamma^x\right) = P(Y \leq \beta \gamma^x). \end{aligned}$$

On obtient, d'après la question 2,

$$F_X(x) = 1 - \left(\frac{1}{\gamma^x}\right)^\alpha = 1 - \gamma^{-\alpha x} = 1 - e^{-\alpha(\ln \gamma)x}.$$

On note que $\mu = \alpha \ln \gamma > 0$. La variable X suit donc une loi exponentielle de paramètre μ . Ainsi une variable Y suit une loi de Pareto de paramètre $(\alpha, \beta, 0)$ si, et seulement si, elle s'écrit $Y = \beta \gamma^X$, où $\gamma > 1$ et X suit une loi exponentielle.

- c)** D'après les deux questions précédentes, on peut écrire $Z = a \gamma^X$, où $\gamma > 1$ et X suit une loi exponentielle de paramètre $\mu = \alpha \ln \gamma$. On a alors

$$Z^c = a^c \gamma^{cX} = a^c (\gamma^c)^X.$$

De la question a, on déduit que Z suit une loi de Pareto de paramètre $\left(\frac{\mu}{\ln \gamma^c}, a^c, 0\right)$.

Comme $\frac{\mu}{\ln \gamma^c} = \frac{\mu}{c \ln \gamma} = \frac{\alpha}{c}$, la variable Z^c suit une loi de Pareto de paramètre $\left(\frac{\alpha}{c}, a^c, 0\right)$.

11 1. a) Pour $(m, n) \in (\mathbb{N}^*)^2$, le changement de variable $u = 1 - t$, donne

$$\beta(n, m) = \int_0^1 u^{n-1}(1-u)^{m-1} du = \int_0^1 (1-t)^{n-1} t^{m-1} dt = \beta(m, n).$$

Pour $m \geq 2$, on obtient en intégrant par parties

$$\begin{aligned} \beta(n, m) &= \int_0^1 u^{n-1}(1-u)^{m-1} du \\ &= \left[\frac{1}{n} u^n (1-u)^{m-1} \right]_0^1 - \int_0^1 \frac{1}{n} u^n (-(m-1))(1-u)^{m-2} du \\ &= \frac{m-1}{n} \beta(n+1, m-1). \end{aligned}$$

b) On calcule

$$\beta(n, 1) = \int_0^1 u^{n-1} du = \frac{1}{n}.$$

On en déduit, en appliquant la relation démontrée dans la question a de manière réitérée que, pour $m \geq 1$,

$$\begin{aligned} \beta(n, m) &= \frac{m-1}{n} \cdot \frac{m-2}{n+1} \cdots \frac{1}{n+m-2} \beta(n+m-1, 1) \\ &= \frac{(m-1)(m-2) \cdots 1}{n(n+1) \cdots (n+m-1)} \\ &= \frac{(m-1)!(n-1)!}{(n+m-1)!}. \end{aligned}$$

2. a) La fonction $f_{m,n}$ est continue sur \mathbb{R} sauf en 0, positive et

$$\int_{-\infty}^{+\infty} f_{m,n}(x) dx = \frac{1}{\beta(n, m)} \int_0^1 x^{n-1}(1-x)^{m-1} dx = \frac{\beta(n, m)}{\beta(n, m)} = 1.$$

Donc $f_{m,n}$ est une densité de probabilité.

b) Sous réserve de convergence, on a

$$E(X) = \int_{-\infty}^{+\infty} x f_{m,n}(x) dx = \frac{1}{\beta(n, m)} \int_0^1 x^n (1-x)^{m-1} dx = \frac{\beta(n+1, m)}{\beta(n, m)}.$$

Cela montre que l'espérance existe et d'après la question 1,

$$E(X) = \frac{(m-1)!n!}{(n+m)!} \cdot \frac{(n+m-1)!}{(m-1)!(n-1)!} = \frac{n}{n+m}.$$

On trouve de même

$$\begin{aligned} E(X^2) &= \frac{\beta(n+2, m)}{\beta(n, m)} = \frac{(m-1)!(n+1)!}{(n+m+1)!} \cdot \frac{(n+m-1)!}{(m-1)!(n-1)!} \\ &= \frac{n(n+1)}{(n+m+1)(n+m)}. \end{aligned}$$

On en déduit

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 = \frac{n(n+1)}{(n+m+1)(n+m)} - \frac{n^2}{(n+m)^2} \\ &= \frac{nm}{(n+m)^2(n+m+1)}. \end{aligned}$$

12 1. Notons f_1 la densité de la loi uniforme sur $[0, 1]$. La variable $X_1 + X_2$ est à valeurs dans $[0, 2]$. Les variables X_1 et X_2 étant indépendantes, $X_1 + X_2$ possède une densité f_2 donnée par le produit de convolution $f_1 * f_1$. On a, pour tout $x \in \mathbb{R}$,

$$f_2(x) = \int_0^1 f_1(t)f_1(x-t) dt = \int_0^1 f_1(x-t) dt = \int_{x-1}^x f_1(u) du.$$

On obtient

- si $x \notin [0, 2]$, $f_2(x) = 0$;
- si $x \in [0, 1]$,

$$f_2(x) = \int_0^x du = x;$$

- si $x \in [1, 2]$,

$$f_2(x) = \int_{x-1}^1 du = 2 - x.$$

2. La variable $X_1 + X_2 + X_3$ est à valeurs dans $[0, 3]$. Les variables $X_1 + X_2$ et X_3 étant indépendantes, $X_1 + X_2 + X_3$ possède une densité $f_3 f_2 * f_1 = f_1 * f_2$. On a, pour tout réel x ,

$$f_3(x) = \int_0^1 f_1(t)f_2(x-t) dt = \int_0^1 f_2(x-t) dt = \int_{x-1}^x f_2(u) du.$$

On obtient donc

- si $x \notin [0, 3]$, $f_3(x) = 0$;
- si $x \in [0, 1]$,

$$f_3(x) = \int_0^x f_2(u) dt = \int_0^x du = \frac{1}{2}x^2;$$

- si $x \in [1, 2]$,

$$\begin{aligned} f_3(x) &= \int_{x-1}^1 u du + \int_1^x (2-u) du = \frac{1}{2} (1 - (x-1)^2 - (2-x)^2 + 1) \\ &= -x^2 + 3x - \frac{3}{2}; \end{aligned}$$

- si $x \in [2, 3]$,

$$f_3(x) = \int_{x-1}^2 (2-u) du = \int_0^{3-x} v dv = \frac{1}{2}(3-x)^2.$$

3. De nouveau X_4 est indépendante de $X_1 + X_2 + X_3$ et $X_1 + X_2 + X_3 + X_4$ possède une densité $f_4 = f_3 * f_1 = f_1 * f_3$, donnée pour tout x par $f_4(x) = \int_{x-1}^x f_3(t) dt$.

On obtient • si $x \notin [0, 4]$, $f_4(x) = 0$.

- si $x \in [0, 1]$,

$$f_4(x) = \int_0^x \frac{1}{2} t^2 dt = \frac{1}{6} x^3;$$

- si $x \in [1, 2]$,

$$\begin{aligned} f_4(x) &= \int_{x-1}^1 \frac{1}{2} t^2 dt + \int_1^x \left(t^2 + 3t - \frac{3}{2} \right) dt \\ &= -\frac{1}{2} x^3 + 2x^2 - 2x + \frac{2}{3}; \end{aligned}$$

- si $x \in [2, 3]$,

$$\begin{aligned} f_4(x) &= \int_{x-1}^2 \left(t^2 + 3t - \frac{3}{2} \right) dt + \int_2^x \frac{1}{2} (3-t)^2 dt \\ &= \frac{1}{2} x^3 - 4x^2 + 10x - \frac{22}{3}; \end{aligned}$$

- si $x \in [3, 4]$,

$$f_4(x) = \int_{x-1}^3 \frac{1}{2} (3-x)^2 = \int_0^{4-x} \frac{1}{2} v^2 dv = \frac{1}{6} (4-x)^3.$$

4. Partant d'une fonction f_1 continue par morceaux, on obtient des fonctions de plus en plus régulières. La fonction f_2 est continue, la fonction f_3 de classe \mathcal{C}^1 et la fonction f_4 de classe \mathcal{C}^2 sur \mathbb{R} . En notant f_n une densité de $X_1 + \dots + X_n$, on obtient, pour tout réel x ,

$$f_n(x) = \int_{x-1}^x f_{n-1}(t) dt,$$

ce qui permet de montrer, par récurrence sur $n \geq 2$ que f_n est de classe \mathcal{C}^{n-2} . La densité de f_n est nulle en dehors du segment $[0, n]$ et sa restriction à chaque intervalle $[k, k+1]$ (pour $0 \leq k \leq n-1$) est un polynôme de degré $n-1$. Son maximum est atteint en $\frac{n}{2}$. Ce maximum décroît, car si on le note M_n , on a pour tout réel x ,

$$f_n(x) \leq \int_{x-1}^x M_{n-1} dt \leq M_{n-1}.$$

La fonction f_n est donc de plus en plus plate. On pourrait démontrer que (M_n) converge vers 0.

13 1. On utilise la fonction Φ de répartition de la loi normale centrée réduite.

La technique est toujours la même : mettre l'exposant sous la forme canonique, puis effectuer le changement de variable indiqué. On donnera chaque fois la valeur exacte de l'intégrale, puis une valeur approchée déterminée par exemple grâce à une table.

- a) On obtient avec le changement de variable $x - 1 = \frac{u}{\sqrt{2}}$,

$$I = \int_0^1 e^{-(x-1)^2} dx = \int_{-\sqrt{2}}^0 e^{-u^2/2} du = \sqrt{\pi} [\Phi(0) - \Phi(-\sqrt{2})] = \frac{\sqrt{\pi}}{2} [2\Phi(\sqrt{2}) - 1]$$

puisque $\Phi(0) = \frac{1}{2}$ et $\Phi(-z) = 1 - \Phi(z)$.

Un calcul approché par interpolation linéaire dans la table du livre, où l'on lit $\Phi(1,41) \approx 0,9207$ et $\Phi(1,42) \approx 0,9222$, donne environ $\Phi(\sqrt{2}) \approx \Phi(1,414) \approx 0,4213$. En résulte la valeur approchée $I \approx 0,7467$, alors qu'un calcul effectué grâce à des logiciels appropriés, donne $I \approx 0,746823$.

- b) On obtient de façon analogue, avec le même changement de variable,

$$I = \int_0^1 e^{-(x-1)^2+1} dx = e \frac{\sqrt{\pi}}{2} [2\Phi(\sqrt{2}) - 1].$$

C'est d'ailleurs clairement le produit par e de l'intégrale précédente. La table donne, comme valeur approchée, 2,0298 au lieu d'environ 2,03007.

- c) On obtient de façon analogue, avec le changement de variable $2x - 1 = \frac{u}{\sqrt{2}}$,

$$I = \int_0^1 e^{-(2x-1)^2} dx = \frac{\sqrt{\pi}}{2} [\Phi(\sqrt{2}) - \Phi(-\sqrt{2})] = \frac{\sqrt{\pi}}{2} [2\Phi(\sqrt{2}) - 1]$$

c'est-à-dire la valeur de la première intégrale.

Cette égalité résulte d'ailleurs de changements de variables très simples (prenant $1 - x$ et $1 - 2x$ comme nouvelles variables) conduisant aussitôt à une autre expression de la valeur commune : $2 \int_0^1 e^{-s^2} ds$.

- d) Il y a ici une faute d'écriture dans l'énoncé, l'intégrale à calculer étant $\int_0^1 e^{-4x^2+4x} dx$.

D'après la question précédente, on obtient aussitôt le produit par e de l'intégrale précédente du c), c'est-à-dire encore la valeur de l'intégrale calculée au b).

2. On utilise la valeur de l'intégrale de Gauss $\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}$.

- a) On obtient avec le changement de variable $x - 1 = t$,

$$I = \int_{-\infty}^{+\infty} e^{-(x-1)^2} dx = \int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}.$$

- b) On obtient de façon analogue, avec le même changement de variable,

$$I = \int_{-\infty}^{+\infty} e^{-(x-1)^2+1} dx = e \int_{-\infty}^{+\infty} e^{-t^2} dt = e \sqrt{\pi}.$$

C'est d'ailleurs clairement le produit par e de l'intégrale précédente.

- c) On obtient avec le changement de variable $2x - 1 = t$,

$$I = \int_{-\infty}^{+\infty} e^{-(2x-1)^2} dx = \frac{1}{2} \int_{-\infty}^{+\infty} e^{-t^2} dt = \frac{1}{2} \sqrt{\pi}.$$

d) Il y a ici une faute d'écriture dans l'énoncé, l'intégrale à calculer étant $\int_{-\infty}^{+\infty} e^{-4x^2+4x} dx$.

D'après la question précédente, on obtient aussitôt $\frac{e}{2}\sqrt{\pi}$.

e) On met le trinôme sous forme canonique. On obtient

$$\begin{aligned} -ax^2 + bx + c &= -a\left(x^2 - \frac{bx}{a}\right) + c = -a\left(x - \frac{b}{2a}\right)^2 + \frac{b^2}{4a} + c \\ &= -\left(x\sqrt{a} - \frac{b}{2\sqrt{a}}\right)^2 + \frac{b^2 + 4ac}{4a}, \end{aligned}$$

puis en faisant le changement de variable $t = x\sqrt{a} - \frac{b}{2\sqrt{a}}$,

$$\begin{aligned} I &= \int_{-\infty}^{+\infty} e^{-ax^2+bx+c} dx \\ &= \int_{-\infty}^{+\infty} e^{-\left(x\sqrt{a} - \frac{b}{2\sqrt{a}}\right)^2 + \frac{b^2+4ac}{4a}} dx \\ &= \frac{e^{\frac{b^2+4ac}{4a}}}{\sqrt{a}} \int_{-\infty}^{+\infty} e^{-t^2} dt \\ &= e^{\frac{b^2+4ac}{4a}} \sqrt{\frac{\pi}{a}}. \end{aligned}$$

On peut ainsi vérifier les résultats des sous-questions précédentes (une même étude aurait pu également être menée dans la question précédente, conduisant alors à l'égalité

$$\int_0^1 e^{-ax^2+bx+c} dx = e^{\frac{b^2+4ac}{4a}} \sqrt{\frac{\pi}{a}} \left[\Phi\left(\frac{2a-b}{\sqrt{2a}}\right) - \Phi\left(-\frac{b}{\sqrt{2a}}\right) \right]$$

avec $a > 0$).

14 1. On note encore Φ la fonction de répartition de la loi normale centrée réduite. On trouve

- a) $P(X \leq 1,36) = \Phi(1,36) \approx 0,9131$, $P(X > 1,36) = 0$,
 $P(X > 1,36) = 1 - P(X \leq 1,36) \approx 0,0869$,
b) $P(X \leq -0,86) = \Phi(-0,86) = 1 - \Phi(0,86) \approx 1 - 0,8051 \approx 0,1949$,
 $P(X > -0,86) = \Phi(0,86) \approx 0,8051$,
c) $P(-0,86 \leq X \leq 1,36) = \Phi(1,36) - \Phi(-0,86) \approx 0,7172$,
 $P(0,86 \leq X \leq 1,36) = \Phi(1,36) - \Phi(0,86) \approx 0,108$,
 $P(-1,36 \leq X \leq -0,86) = P(0,86 \leq X \leq 1,36) \approx 0,108$.
2. a) On trouve $\Phi(1,62) \approx 0,9474$. On veut $\Phi(x) \approx 0,0537$ et donc $\Phi(-x) \approx 0,9463$. On trouve $-x \approx 1,61$ et donc $x \approx -1,61$.
- b) On constate que $\Phi(1,44) \approx 0,9251$ et $\Phi(1,45) \approx 0,9265$. Par interpolation linéaire, on prend

$$x \approx 1,44 + \frac{0,01(0,9259 - 0,9251)}{0,9265 - 0,9251} \approx 1,4457.$$

On veut $\Phi(x) \approx 0,1778$ et donc $\Phi(-x) \approx 0,8222$. On trouve $\Phi(0,92) \approx 0,8212$ et $\Phi(0,93) \approx 0,8238$. Par interpolation linéaire, on obtient

$$-x \approx 0,92 + \frac{0,01 \cdot 10}{26} \approx 0,9238 \quad \text{et} \quad x \approx -0,9238$$

3. La variable $Y = \frac{X-5}{3}$ suit une loi normale centrée réduite.

a) On a

$$P(X \leq 1,36) = P\left(Y \leq \frac{1,36-5}{3}\right) \approx \Phi(-1,213) = 1 - \Phi(1,213).$$

Par interpolation linéaire, on trouve

$$\Phi(1,213) \approx 0,8869 + \frac{3 \cdot 0,0019}{10} \approx 0,8875$$

et donc

$$P(X \leq 1,36) \approx 0,1125.$$

On en déduit $P(X = 1,36) = 0$ et $P(X > 1,36) = 1 - P(X \leq 1,36) \approx 0,8875$.

b) On obtient de même

$$P(X \leq -0,86) = P(Y \leq -1,953) = 1 - \Phi(1,953).$$

Par interpolation linéaire, on obtient

$$\Phi(1,953) \approx 0,9744 + \frac{3 \cdot 0,0006}{10} \approx 0,9746.$$

On en déduit

$$P(X \leq -0,86) \approx 0,0254 \quad \text{et} \quad P(X > -0,86) \approx 0,9746.$$

c) De ce qui précède on déduit

$$P(-0,86 \leq X \leq 1,36) = P(X \leq 1,36) - P(X \leq -0,86) \approx 0,0869.$$

On a par ailleurs

$$P(X \leq 0,86) = P(Y \leq -1,38) = 1 - \Phi(1,38) \approx 0,0838$$

et

$$P(0,86 \leq X \leq 1,36) = P(-1,36 \leq X \leq -0,86) = \Phi(1,36) - \Phi(0,86) \approx 0,0287.$$

4. a) On a $P(X \leq x) = P\left(Y \leq \frac{x-5}{3}\right) = \Phi\left(\frac{x-5}{3}\right)$.

Comme $\Phi(1,62) \approx 0,9474$ et $\Phi(-1,61) = 1 - \Phi(1,61) \approx 0,0537$, on obtient dans le premier cas $\frac{x-5}{3} \approx 1,62$ soit $x \approx 9,86$ et dans le second $\frac{x-5}{3} \approx -1,61$ soit $x \approx 0,17$.

- b) Dans la question 2.b, on a trouvé $\Phi(1, 4457) \approx 0, 9258$ et $\Phi(-0, 9238) \approx 0, 1778$. On obtient donc dans le premier cas $\frac{x-5}{3} \approx 1, 4457$ soit $x \approx 9, 3372$ et dans le second $\frac{x-5}{3} \approx -0, 9238$ soit $x \approx 2, 2286$.

15 1. On cherche $(A, B, C, D) \in \mathbb{R}^4$ tel que, pour tout $x \in \mathbb{R}$,

$$\frac{1}{(u^2 + x^2)(v^2 + (z - x)^2)} = \frac{Ax + B}{u^2 + x^2} + \frac{C(z - x) + D}{v^2 + (z - x)^2},$$

c'est-à-dire

$$1 = (Ax + B)(v^2 + (z - x)^2) + (C(z - x) + D)(u^2 + x^2).$$

On veut que le polynôme

$$P = (AX + B)(v^2 + (z - X)^2) + (C(z - X) + D)(u^2 + X^2) - 1 \in \mathbb{R}_3[X]$$

soit le polynôme nul. Si cela est vrai tout nombre complexe est racine de P et l'on a en particulier

$$\begin{aligned} P(ui) &= (Aui + B)(v^2 + (z - ui)^2) - 1 = 0, \\ P(z + vi) &= (-Cvi + D)(u^2 + (z + vi)^2) - 1. \end{aligned}$$

Si réciproquement ces conditions sont réalisées, ui et $z + vi$ sont racines de P , et comme P est à coefficients réels, leurs conjugués $-ui$ et $z - vi$ sont racines de P . Si $z \in \mathbb{R}^*$, ces quatre racines sont distinctes donc P est le polynôme nul. On suppose désormais $z \in \mathbb{R}^*$. Les réels A, B, C et D vérifient l'égalité demandée si, et seulement si,

$$Aui + B = \frac{1}{v^2 + (z - ui)^2} \quad \text{et} \quad -Cvi + D = \frac{1}{u^2 + (z + vi)^2}.$$

On trouve un quadruplet (A, B, C, D) unique. En effet il est définie par

$$\begin{aligned} B &= \Re \left(\frac{1}{v^2 + (z - ui)^2} \right), & A &= \frac{1}{u} \Im \left(\frac{1}{v^2 + (z - ui)^2} \right), \\ D &= \Re \left(\frac{1}{u^2 + (z + vi)^2} \right), & C &= -\frac{1}{v} \Im \left(\frac{1}{u^2 + (z + vi)^2} \right). \end{aligned}$$

Les deux variables étant indépendantes, elles possèdent une densité obtenue par un produit de convolution. On note f cette densité. On a donc, pour tout $z \in \mathbb{R}^*$,

$$\begin{aligned} f(z) &= \int_{-\infty}^{+\infty} \frac{uv}{\pi^2(u^2 + x^2)(v^2 + (z - x)^2)} dx \\ &= \frac{uv}{\pi^2} \int_{-\infty}^{+\infty} \left(\frac{Ax + B}{u^2 + x^2} + \frac{C(z - x) + D}{v^2 + (z - x)^2} \right) dx. \end{aligned}$$

On calcule d'abord

$$\begin{aligned} \int_{-\infty}^{+\infty} \left(\frac{B}{u^2 + x^2} + \frac{D}{v^2 + (z-x)^2} \right) dx &= \left[\frac{B}{u} \operatorname{Arctan} \left(\frac{x}{u} \right) + \frac{D}{v} \operatorname{Arctan} \left(\frac{x-z}{v} \right) \right]_{-\infty}^{+\infty} \\ &= \pi \left(\frac{B}{u} + \frac{D}{v} \right) = \frac{\pi(Bv + Du)}{uv}. \end{aligned}$$

Reste alors à calculer $\int_{-\infty}^{+\infty} \left(\frac{Ax}{u^2 + x^2} + \frac{C(z-x)}{v^2 + (z-x)^2} \right) dx$. Cette intégrale converge comme différence de deux intégrales convergentes. On note g la fonction à intégrer. Comme $\frac{Ax}{u^2 + x^2} \underset{x \rightarrow +\infty}{\sim} \frac{A}{x}$ et $\frac{C(z-x)}{v^2 + (z-x)^2} \underset{x \rightarrow +\infty}{\sim} \frac{-C}{x}$, on a donc $A = C$, sinon l'intégrale diverge. Une primitive G de g est définie, compte tenu de $A = C$, par

$$G(x) = \frac{A}{2} \ln(u^2 + x^2) - \frac{A}{2} \ln(v^2 + (z-x)^2) = \frac{A}{2} \ln \frac{(u^2 + x^2)}{v^2 + (z-x)^2}.$$

La fonction G a donc pour limite en $\pm\infty$ $\ln 1 = 0$. On obtient donc finalement

$$f(z) = \frac{uv}{\pi^2} \frac{\pi(Bv + Du)}{uv} = \frac{1}{\pi}(Bv + Du).$$

Reste à calculer $Bv + Du = \Re(Z)$, où $Z = \left(\frac{v}{v^2 + (z-uv)^2} + \frac{u}{u^2 + (z+vi)^2} \right)$. On transforme l'expression de Z . On obtient, en écrivant que, pour tout $(a, b) \in \mathbb{C}^2$, $a^2 + b^2 = (a+ib)(a-ib)$,

$$\begin{aligned} Z &= \frac{v}{(v+zi+u)(v-zi-u)} + \frac{u}{u+zi-v)(u-zi+v)} \\ &= \frac{v(u+v-z\hat{i})-u(v+zi+u)}{(v+zi+u)(v-zi-u)(u+v-z\hat{i})} = \frac{(u+v)(v-u-z\hat{i})}{(v+zi+u)(v-zi-u)(u+v-z\hat{i})} \\ &= \frac{u+v}{(u+v+z\hat{i})(u+v-z\hat{i})} = \frac{u+v}{(u+v)^2 + z^2}. \end{aligned}$$

On a finalement

$$Bv + Du = \frac{u+v}{(u+v)^2 + z^2} \quad \text{et} \quad f(z) = \frac{u+v}{\pi((u+v)^2 + z^2)}.$$

La fonction f coïncide avec la densité de la loi $C(u+v)$ sur \mathbb{R}^* . Cela suffit pour conclure que le loi de la somme est $C(u+v)$, la densité d'une variable aléatoire pouvant être modifiée sur un nombre fini de points.

2. Leur somme X suit la loi de Cauchy $C(2t)$. On cherche la loi de $Y = \frac{1}{2}X$. Sa fonction de répartition vérifie, pour tout $x \in \mathbb{R}$,

$$F_Y(x) = P(Y \leq x) = P(X \leq 2x) = F_X(2x).$$

On en déduit sa densité

$$f_Y(x) = 2f_X(2x) = 2 \frac{2t}{\pi((2t)^2 + (2x)^2)} = \frac{t}{\pi(t^2 + x^2)}.$$

La variable Y suit une loi $C(t)$. La moyenne arithmétique de deux variables suivant la loi $C(t)$ suit la loi $C(t)$.

Chapitre 11

1 Il va de soi que le V_n qui figure par erreur dans cet énoncé doit être remplacé par X_n .

Par définition, $X_n \xrightarrow{\mathcal{P}} X$ signifie que $\forall \varepsilon \in \mathbb{R}_+^*$, $\lim_{n \rightarrow +\infty} P(|X_n - X| > \varepsilon) = 0$, c'est-à-dire que

$$\forall \varepsilon \in \mathbb{R}_+^* \forall \delta \in \mathbb{R}_+^*, \exists N(\varepsilon, \delta) \in \mathbb{N}, \quad n \geq N(\varepsilon, \delta) \Rightarrow P(|V_n - X| > \varepsilon) \leq \delta.$$

Il est clair (en prenant $\delta = \varepsilon$) que cela implique

$$\forall \varepsilon \in \mathbb{R}_+^*, \exists N(\varepsilon) \in \mathbb{N}, \quad n \geq N(\varepsilon) \Rightarrow P(|V_n - X| > \varepsilon) \leq \varepsilon.$$

Réciproquement, supposons que $\forall \varepsilon \in \mathbb{R}_+^*, \exists N(\varepsilon) \in \mathbb{N}, \quad n \geq N(\varepsilon) \Rightarrow P(|V_n - X| > \varepsilon) \leq \varepsilon$.
Considérons alors deux réels ε et δ strictement positifs.

Si $\delta \geq \varepsilon$, alors $\exists N(\varepsilon) \in \mathbb{N}, \quad n \geq N(\varepsilon) \Rightarrow P(|V_n - X| > \varepsilon) \leq \varepsilon \leq \delta$.

La conclusion en résulte en posant $N(\varepsilon, \delta) = N(\varepsilon)$.

Si $\delta < \varepsilon$, alors $\{|X_n - X| > \varepsilon\} \subset \{|X_n - X| > \delta\}$ et donc $P(|X_n - X| > \varepsilon) \leq P(|X_n - X| > \delta)$.
En appliquant la propriété au réel strictement positif δ , on obtient alors

$$\exists N(\delta) \in \mathbb{N}, \quad n \geq N(\delta) \Rightarrow P(|V_n - X| > \delta) \leq \delta \leq P(|X_n - X| > \varepsilon)$$

d'où le résultat, en posant $N(\varepsilon, \delta) = N(\delta)$.

2 **1.** Il est clair que $X_n \xrightarrow{\mathcal{P}} X$ si et seulement si $|X_n - X| \xrightarrow{\mathcal{P}} 0$. Donc, pour démontrer

le résultat annoncé, il suffit de démontrer que $X_n \xrightarrow{\mathcal{P}} 0 \iff \lim_{n \rightarrow +\infty} E\left(\frac{|X_n|}{1 + |X_n|}\right) = 0$.

Supposons que $X_n \xrightarrow{\mathcal{P}} 0$. Alors, $\forall \varepsilon \in \mathbb{R}_+^*$, $\lim_{n \rightarrow +\infty} P(|X_n - X| > \varepsilon) = 0$.

D'autre part, si $|X_n| \leq \varepsilon$, alors $\frac{|X_n|}{1 + |X_n|} \leq \varepsilon$, et, pour tout réel x , $0 \leq \frac{|x|}{1 + |x|} \leq 1$.

Il en résulte, en désignant par $\mathbf{1}_A$ la fonction caractéristique de l'ensemble A , et en rappelant que $P(A) = E(\mathbf{1}_A)$, que

$$0 \leq \frac{|X_n|}{1 + |X_n|} \leq \frac{|X_n|}{1 + |X_n|} \mathbf{1}_{(|X_n| > \varepsilon)} + \varepsilon \mathbf{1}_{(|X_n| \leq \varepsilon)} \leq \mathbf{1}_{(|X_n| > \varepsilon)} + \varepsilon.$$

En prenant l'espérance, il en résulte que $0 \leq E\left(\frac{|X_n|}{1 + |X_n|}\right) \leq E(\mathbf{1}_{(|X_n| > \varepsilon)}) + \varepsilon$ c'est-à-dire

$$\text{que } 0 \leq E\left(\frac{|X_n|}{1 + |X_n|}\right) \leq P(|X_n| > \varepsilon) + \varepsilon.$$

Il existe donc n_0 tel que, pour tout $n > n_0$, $0 \leq E\left(\frac{|X_n|}{1+|X_n|}\right) \leq 2\varepsilon$.

Comme ε est un réel strictement positif quelconque, cela prouve que $\lim_{n \rightarrow +\infty} E\left(\frac{|X_n|}{1+|X_n|}\right) = 0$.

Réciproquement, supposons que $\lim_{n \rightarrow +\infty} E\left(\frac{|X_n|}{1+|X_n|}\right) = 0$.

La fonction f définie par $f(x) = \frac{x}{1+x}$ est positive et strictement croissante sur \mathbb{R}_+ . Donc,

pour tout réel strictement positif ε , $0 \leq \frac{\varepsilon}{1+\varepsilon} \mathbf{1}_{(|X_n| > \varepsilon)} \leq \frac{|X_n|}{1+|X_n|} \mathbf{1}_{(|X_n| > \varepsilon)} \leq \frac{|X_n|}{1+|X_n|}$.

En prenant l'espérance et en passant à la limite, on en déduit que

$$\frac{\varepsilon}{1+\varepsilon} \lim_{n \rightarrow +\infty} P(|X_n| > \varepsilon) \leq \lim_{n \rightarrow +\infty} E\left(\frac{|X_n|}{1+|X_n|}\right).$$

Or, par hypothèse, cette dernière limite est nulle. Comme ε est donné, strictement positif, il en résulte que $\lim_{n \rightarrow +\infty} P(|X_n| > \varepsilon) = 0$, c'est-à-dire le résultat annoncé.

2. Dans les démonstrations ci dessus, la forme exacte de la fonction f n'intervient pas. Le lecteur constatera que les mêmes démonstrations s'appliquent dans cette deuxième question, à un détail près. Ce détail consiste en le fait que, dans la première question, la fonction f est majorée par 1, alors que dans la deuxième question, on sait qu'elle est majorée, mais l'on ne connaît pas de majorant. On appellera k un majorant de $f(x)$ sur \mathbb{R}_+ . Nous laissons au lecteur le soin d'apporter les quelques modifications qui en découlent.

3 1. Pour tout entier naturel non nul n , X_n prend trois valeurs ($X_n(\Omega) = \{-n, 0, n\}$) et la somme des probabilités est bien égale à 1.

Pour tout entier naturel n , on a clairement $E(X_n) = 0$ et $E(X_n^2) = 1$, donc $V(X_n) = 1$.

2. Pour tout réel strictement positif ε , et pour tout entier n supérieur à ε , on a $P(|X_n| > \varepsilon) = 0$. Il est donc clair que $\lim_{n \rightarrow +\infty} P(|X_n| > \varepsilon) = 0$, et donc que la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers 0.

Pour tout entier naturel non nul n , $E(|X_n|) = \frac{1}{n}$, et donc $\lim_{n \rightarrow +\infty} E(|X_n|) = 0$, c'est-à-dire que la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en moyenne vers 0 (ce qui d'ailleurs implique qu'elle converge en probabilité vers 0).

Pour tout entier naturel non nul n , $E(X_n^2) = 1$, et donc la suite $(X_n)_{n \in \mathbb{N}^*}$ ne converge pas en moyenne quadratique.

4 Pour tout n , une densité de X_n est la fonction f définie par $f(x) = 0$ si $x < \theta$, et $f(x) = e^{-(x-\theta)}$ pour $x \geq \theta$. La variable aléatoire θ prend ses valeurs dans $[\theta, +\infty[$.

En appliquant l'inégalité de Markov à la variable aléatoire positive $m_n - \theta$ ($= |m_n - \theta|$), on obtient, pour ε réel strictement positif, $P(|m_n - \theta| \geq \varepsilon) = P(m_n - \theta \geq \varepsilon) \leq \frac{E(m_n - \theta)}{\varepsilon}$.

Pour tout réel x , on a $P(m_n > x) = P\left(\bigcap_{k=1}^n (X_k > x)\right) = \prod_{k=0}^n P(X_k > x) = [1 - F(x)]^n$

où $F(x) = \int_{\theta}^x e^{-(t-\theta)} dt = 1 - e^{-(x-\theta)}$.

La fonction de répartition G de m_n est donnée par $G(x) = 1 - [1 - F(x)]^n$ et l'une de ses densités g par $g(x) = nf(x)[1 - F(x)]^{n-1} = ne^{-n(x-\theta)}$.

On en déduit l'espérance de m_n

$$\begin{aligned} E(m_n) &= n \int_{\theta}^{+\infty} te^{-n(t-\theta)} dt = n \int_{\theta}^{+\infty} (t - \theta + \theta) e^{-n(t-\theta)} dt \\ &= \frac{1}{n} \int_0^{+\infty} ue^{-u} du + \theta \int_{\theta}^{+\infty} ne^{-n(t-\theta)} dt. \end{aligned}$$

Tous calculs faits, on obtient $E(m_n) = \frac{1}{n} + \theta$.

De tout cela, il résulte que $0 \leq P(|m_n - \theta| \geq \varepsilon) \leq \frac{1}{n\varepsilon}$, et donc que, pour tout réel strictement positif ε , $\lim_{n \rightarrow +\infty} P(|m_n - \theta| \geq \varepsilon) = 0$, ce qui se traduit par $m_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} \theta$.

5

La démarche est la même que dans l'exercice précédent. On obtient

$$1. P(Y_k = 0) = (1 - p)^2; \quad P(Y_k = 1) = 2p(1 - p), \quad P(X_K = 2) = p^2, \\ E(Y_k) = 2p, \quad V(Y_k) = 2p.$$

2. On a clairement $E(Z_n) = 2p$.

$$\text{D'autre part, } Z_n = \frac{1}{n} \left(X_1 + \sum_{k=2}^{n-1} 2X_k + X_n \right)$$

et, en utilisant l'indépendance de $X_1, 2X_2, \dots, 2X_{n-1}, X_n$

$$V(Z_n) = \frac{1}{n^2} \left(V(X_1) + 4 \sum_{k=2}^{n-1} V(X_k) + V(X_n) \right) = \left(\frac{4n-6}{n^2} \right) p(1-p).$$

3. L'inégalité de Bienaymé-Tchebychev appliquée à Z_n montre que, pour tout réel strictement positif ε , $P(|Z_n - 2p| \geq \varepsilon) \leq \frac{(4n-6)p(1-p)}{n^2\varepsilon^2}$.

Il en résulte que $\lim_{n \rightarrow +\infty} P(|Z_n - 2p| \geq \varepsilon) = 0$, et donc que la suite $(Z_n)_{n \in \mathbb{N}}$ converge en probabilité vers la constante $2p$ (on remarquera à ce sujet que $P(|Z_n - 2p| > \varepsilon) \leq P(|Z_n - 2p| \geq \varepsilon)$).

Comme la convergence en probabilité implique la convergence en loi, on peut conclure que la suite $(Z_n)_{n \in \mathbb{N}}$ converge en loi vers la constante $2p$.

► Remarques

- on pouvait bien sûr déterminer la fonction de répartition F_{Z_n} de Z_n et chercher la limite de $F_{Z_n}(x)$ lorsque n tend vers l'infini, mais les premières questions et la question relative à la loi des grands nombres guidaient vers l'utilisation de la convergence en probabilité.
- La loi des grands nombres ne s'applique pas ici car les Y_k ne sont pas deux à deux non corrélées.

6

1. Soit $F_Y(x) = P(Y \leq x) = P(e^{-X} \leq Y)$.

X prend ses valeurs dans \mathbb{R}_+ , $Y = e^{-X}$ prend ses valeurs dans $[0, 1]$, donc, pour $x \leq 0$, $F_Y(x) = 0$; pour $x \geq 1$, $F_Y(x) = 1$.

$$\text{Pour } x \in]0, 1[, F_Y(x) = P(-X \leq \ln x) = P(X \geq -\ln x) = \int_{-\ln x}^{+\infty} e^{-t} dt = x.$$

On conclut donc que Y est une variable aléatoire qui suit la loi uniforme sur l'intervalle $[0, 1]$.

2. On a $F_{Y_n}(x) = P(Y_n \leq x) = P(e^{\frac{1}{n}} e^{-X} \leq x) = P(e^{-X} \leq x e^{-\frac{1}{n}}) = F_Y(x e^{-\frac{1}{n}})$.
 Pour tout x , la fonction F_Y est continue, donc, quand n tend vers l'infini, $F_{Y_n}(x)$ tend vers $F_Y(x)$, c'est-à-dire que la suite $(Y_n)_{n \in \mathbb{N}^*}$ converge en loi vers la variable aléatoire Y .

► **Remarque**

on peut aussi démontrer, et le lecteur pourra s'y exercer, que la suite $(Y_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers la variable aléatoire Y .

- 7 1. Pour tout réel strictement positif ε , si $n \leq \varepsilon$, $P(|X_n| > \varepsilon) = 0$, et si $n > \varepsilon$, alors $P(|X_n| > \varepsilon) = \frac{1}{n}$. Il en résulte clairement que la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers la variable aléatoire constante nulle, ce qui s'écrit $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} 0$ ou plus simplement $X_n \xrightarrow{\mathcal{P}} 0$.

Pour tout entier non nul k , $E(X_n^k) = n^{k-1}$. Donc, en particulier, $E(X_n^2) = n$, qui tend vers l'infini quand n tend vers l'infini. Cela prouve que la suite $(X_n)_{n \in \mathbb{N}^*}$ ne converge pas en moyenne quadratique.

► **Remarque**

nous avons ici un exemple d'une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ qui converge en probabilité vers une variable aléatoire X , telle qu'aucun moment de X_n ne converge vers le moment correspondant de X .

2. Pour tout réel x

$$\begin{aligned} [Y_n + X_n \leq x] &= ([Y_n \leq x] \cap [X_n = 0]) \cup ([Y_n \leq x - n] \cap [X_n = n]) \\ &= \left(1 - \frac{1}{n}\right) F_{Y_n}(x) + \frac{1}{n} F_{Y_n}(x - n). \end{aligned}$$

Or, pour tout n , $F_{Y_n} = \Phi$ (fonction de répartition de la loi normale centrée réduite). Cette fonction étant bornée, on en déduit clairement que $\lim_{n \rightarrow \infty} F_{X_n + Y_n}(x) = \Phi(x)$, et l'on conclut que la suite de variables aléatoires $(X_n + Y - n)_{n \in \mathbb{N}^*}$ converge en loi vers une variable normale centrée réduite.

3. On a $V(Z_n) = V(X_n) + V(Y_n) = n - 1 + 1 = n$. Donc $\lim_{n \rightarrow \infty} V(Z_n) = +\infty$, alors que Z_n converge en loi vers une variable aléatoire de variance 1.

8 Cet énoncé est entaché d'une malencontreuse erreur, dont nous prions nos lecteurs

de bien vouloir nous pardonner. Il faut lire $\frac{E(X^2) - a^2}{M^2} \leq P(X \geq a) \leq \frac{E(|X|)}{a}$.

1. Si X est une variable aléatoire discrète ($X(\Omega) = \{x_1, x_2, \dots\}$), quel que soit l'entier naturel n , la série de terme général $x_i^n P(X = x_i)$ est absolument convergente, car $|x_i^n P(X = x_i)|$ est majoré par $M^n P(X = x_i)$ qui est clairement le terme général d'une série positive convergente.

De même, si X est une variable admettant la densité f , quel que soit l'entier naturel n , l'intégrale $\int_{-\infty}^{+\infty} t^n f(t) dt$ est absolument convergente, car $\int_{-\infty}^{+\infty} |t^n f(t)| dt$ est majorée par $\int_{-\infty}^{+\infty} M^n f(t) dt$, qui est clairement convergente.

2. Notons $\mathbf{1}_{(X \geq a)}$ la variable indicatrice de $(X \geq a)$.

Si $\mathbf{1}_{(X \geq a)} = 0$, alors $X < a$, donc $X^2 - a^2 < 0$. On a alors $\frac{X^2 - a^2}{M^2} \leq \mathbf{1}_{(X \geq a)} \leq \frac{|X|}{a}$.

Si $\mathbf{1}_{(X \geq a)} = 1$, alors $X \geq a$, $\frac{|X|}{a} \geq 1$, et $\frac{X^2 - a^2}{M^2}$ étant inférieur à $\frac{X^2}{M^2} (\leq 1)$, on aura $\frac{X^2 - a^2}{M^2} \leq 1$. On en déduit donc aussi dans ce cas que $\frac{X^2 - a^2}{M^2} \leq \mathbf{1}_{(X \geq a)} \leq \frac{|X|}{a}$.

On conclut d'après la croissance de l'espérance, sa linéarité, et le fait que $E(\mathbf{1}_{(X \geq a)}) = P(X \geq a)$.

9 1. Pour tout entier naturel non nul n , la fonction f_n est continue et positive sur \mathbb{R} .

D'autre part, $\int_{-\infty}^x f_n(t) dt = \int_0^x \frac{n^2 t}{\sigma^2} e^{-\frac{n^2 t^2}{2\sigma^2}} dt = 1 - e^{-\frac{n^2 x^2}{2\sigma^2}}$.

Donc $\lim_{x \rightarrow +\infty} \int_{-\infty}^x f_n(t) dt = \lim_{x \rightarrow +\infty} 1 - e^{-\frac{n^2 x^2}{2\sigma^2}} = 1$.

Tout cela prouve que, pour tout entier naturel non nul n , f_n est une densité d'une variable aléatoire (noter qu'il s'agit d'une densité et non pas de la densité...).

2. Soit ε un réel strictement positif. On a

$P(|X_n| \geq \varepsilon) = P(X_n \geq \varepsilon) = \int_{\varepsilon}^{+\infty} \frac{n^2 t}{\sigma^2} e^{-\frac{n^2 t^2}{2\sigma^2}} dt = e^{-\frac{n^2 \varepsilon^2}{2\sigma^2}}$.

Donc $\lim_{n \rightarrow +\infty} P(|X_n| \geq \varepsilon) = 0$, c'est-à-dire que la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers la variable aléatoire certaine nulle ($X = 0$).

10 1. Il s'agit d'une simple application du théorème de Bernoulli (qu'il faut savoir démontrer, voir p. 397)

2. On sait que $E(Y_n) = \frac{1}{3}$. L'inégalité de Bienaymé-Tchebychev, appliquée à $n=100000$, et

$\varepsilon = 0,34 - \frac{1}{3} = \frac{1}{150}$ donne $P(|Y_n - E(Y_n)|) < \frac{9 \times 100000}{\left(\frac{1}{150}\right)^2}$, et le résultat

$$P(|Y_n - E(Y_n)|) < 0,05.$$

► **Remarque**

On sait que l'inégalité de Bienaymé-Tchebychev donne des majorations grossières. On peut sans doute (avec d'autres méthodes) trouver une majoration bien plus fine (voir chapitre 12).

11 1. Avec $E(X) = \frac{1}{\lambda}$ et $V(X) = \frac{1}{\lambda^2}$, l'inégalité de Bienaymé-Tchebychev s'écrit

$$P\left(\left|X - \frac{1}{\lambda}\right| \geq \varepsilon\right) \leq \frac{1}{\lambda^2 \varepsilon^2}.$$

Or, l'événement $\left[X \geq \frac{1}{\lambda} + \varepsilon\right]$ implique l'événement $\left[\left|X - \frac{1}{\lambda}\right| \geq \varepsilon\right]$.

Donc $P\left(X \geq \frac{1}{\lambda} + \varepsilon\right) \leq P\left(\left|X - \frac{1}{\lambda}\right| \geq \varepsilon\right) \leq \frac{1}{\lambda^2 \varepsilon^2}$ c'est-à-dire $\int_{\frac{1}{\lambda} + \varepsilon}^{+\infty} \lambda e^{-\lambda t} dt \leq \frac{1}{\lambda^2 \varepsilon^2}$,
soit $e^{-\lambda(\frac{1}{\lambda} + \varepsilon)} \leq \frac{1}{\lambda^2 \varepsilon^2}$.

Ce résultat peut s'écrire $(\lambda \varepsilon)^2 e^{-\lambda \varepsilon} \leq e$, où ε désigne un réel strictement positif quelconque.
En prenant, pour t positif, $\varepsilon = \frac{t}{\lambda}$, on obtient $t^2 e^{-t} \leq e$.

2. En étudiant la fonction, on trouve que f admet un maximum atteint pour $x = 2$ et qui vaut $4e^{-2}$. Le calcul numérique permet d'écrire, dans le cadre de la question précédente, que $t^2 e^{-t} \leq 2,71829$ (par exemple), alors que, dans le deuxième cas, on peut obtenir $t^2 e^{-t} \leq 0,54135$.

On retrouve ici le fait que les majorations obtenues par l'inégalité de Bienaymé-Tchebychev sont bien ... médiocres.

12 Notons X_n le nombre d'apparition de l'as au cours des $6n$ premiers lancers. En supposant les lancers successifs indépendants, X_n suit une loi binomiale de paramètres $6n$ et $\frac{1}{6}$. On a alors $E(X_n) = n$ et $V(X_n) = \frac{5n}{6}$.

Pour tout réel strictement positif ε , l'inégalité de Bienaymé-Tchebychev donne

$$P(|X_n - n| \geq \varepsilon) \leq \frac{5n}{6\varepsilon^2}.$$

Cette inégalité peut encore s'écrire $P\left(\left|\frac{X_n}{6n} - \frac{1}{6}\right| \geq \frac{\varepsilon}{6n}\right) \leq \frac{5n}{6\varepsilon^2}$.

$\frac{X_n}{6n}$ est la fréquence d'apparition de l'as au cours des $6n$ lancers. On veut que cette fréquence s'écarte de $\frac{1}{6}$ de moins de 10^{-2} . Posons donc $\frac{\varepsilon}{6n} = 10^{-2}$, c'est à dire $\varepsilon = 6n10^{-2}$.

Tout calcul fait, l'inégalité devient alors $P\left(\left|\frac{X_n}{6n} - \frac{1}{6}\right| \geq \frac{\varepsilon}{6n}\right) \leq \frac{5 \times 10^4}{216n}$.

Or, on veut que cette probabilité soit inférieure à $\frac{1}{2}$. Il suffira alors que $\frac{5 \times 10^4}{216n} \geq \frac{1}{2}$, c'est à dire $n \geq \frac{100000}{216}$, soit, puisque n est entier, $n \geq 463$.

► **Remarque**

On sait que les majorations données par l'inégalité de Bienaymé-Tchebychev sont grossières. Il est très vraisemblable qu'une autre méthode permette de trouver une majoration plus fine de n (voir chapitre 12).

13 La fonction de répartition de Y_n est définie pour x strictement positif par

$$F_{Y_n}(x) = P(Y_n \leq x) = P(X_n \leq nx) = \sum_{k=1}^{\lfloor nx \rfloor} p_n (1 - p_n)^{k-1} = p_n \frac{1 - (1 - p_n)^{\lfloor nx \rfloor}}{1 - (1 - p_n)}$$

$$= 1 - (1 - p_n)^{\lfloor nx \rfloor}$$

où $\lfloor z \rfloor$ désigne la partie entière de z .

► **Remarque**

Si $\lfloor nx \rfloor < 1$, il n'y a aucun terme dans la somme, on a $F_{Y_n}(x) = 0$.

Soit $\ln(1 - F_{Y_n}(x)) = \lfloor nx \rfloor \ln \left(1 - \frac{\theta}{n} \right)$.

On sait que $nx - 1 < \lfloor nx \rfloor \leq nx$. Donc

$$nx \ln \left(1 - \frac{\theta}{n} \right) < \lfloor nx \rfloor \ln \left(1 - \frac{\theta}{n} \right) \leq (nx - 1) \ln \left(1 - \frac{\theta}{n} \right).$$

Or, quand n tend vers l'infini, $\ln \left(1 - \frac{\theta}{n} \right) \sim -\frac{\theta}{n}$ et, x étant strictement positif, $nx - 1 \sim nx$.

Il en résulte que $\lim_{n \rightarrow \infty} \ln(1 - F_{Y_n}(x)) = -\theta x$, et par suite que $\lim_{n \rightarrow \infty} F_{Y_n}(x) = 1 - e^{-\theta x}$ ce qui permet de conclure que la suite $(Y_n)_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire suivant la loi exponentielle de paramètre θ .

14 1. La constante a est définie par la condition $\int_{-\infty}^{+\infty} \frac{an}{\pi(1+n^2t^2)} dt = 1$

Dans l'intégrale $\int_x^y \frac{n}{\pi(1+n^2t^2)} dt$, posons $u = nt$, et donc $dt = \frac{du}{n}$.

$$\text{Donc } \int_x^y \frac{n}{\pi(1+n^2t^2)} dt = \frac{1}{n} \int_{nx}^{ny} \frac{n}{\pi(1+u^2)} du = \frac{1}{\pi} (\text{Arctan}(ny) - \text{Arctan}(nx)).$$

Il en résulte que, en faisant tendre x vers $-\infty$ et y vers $+\infty$, $\int_{-\infty}^{+\infty} \frac{n}{\pi(1+n^2t^2)} dt = 1$, et donc que $a = 1$.

2. On constate que, au voisinage de l'infini, $xf_n(x) \sim \frac{1}{\pi nx}$ dont l'intégrale est divergente. La variable aléatoire X_n n'admet donc pas d'espérance, et par conséquent, n'admet aucun moment.
3. La fonction de répartition de X_n est donnée par

$$F_{X_n}(x) = \int_{-\infty}^x \frac{an}{\pi(1+n^2t^2)} dt = \frac{1}{\pi} \left(\text{Arctan}(nx) + \frac{\pi}{2} \right).$$

Pour $x < 0$, on a alors $\lim_{n \rightarrow \infty} F_{X_n}(x) = \frac{1}{\pi} \left(-\frac{\pi}{2} + \frac{\pi}{2} \right) = 0$.

Pour $x > 0$, on a alors $\lim_{n \rightarrow \infty} F_{X_n}(x) = \frac{1}{\pi} \left(\frac{\pi}{2} + \frac{\pi}{2} \right) = 1$.

On obtient ainsi, à la limite, la fonction de répartition de la variable aléatoire constante nulle. On peut donc conclure que la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers la variable aléatoire constante nulle ($X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} 0$).

15 1. On reconnaît un exercice traité en première année (voir exercice n°28 p 222 du Tome 1 du présent ouvrage).

Rappelons simplement qu'on utilise la formule de Poincaré, et que l'on trouve

$$p_n = \sum_{k=0}^n n \frac{(-1)^k}{k!}.$$

2. Soit k un entier naturel compris entre 0 et n . On a $\binom{n}{k}$ façons de choisir k éléments distincts deux à deux dans l'ensemble $\llbracket 1, n \rrbracket$, et la probabilité pour que chacun de ces k éléments soit un

point fixe de la permutation et qu'il n'y en ait pas d'autre est

$$\frac{1}{n} \times \frac{1}{n-1} \times \cdots \times \frac{1}{n-k+1} \times p_{n-k} = \frac{(n-k)!}{n!} \times p_{n-k}.$$

La loi de X_n est donc donnée par $P(X_n = k) = \binom{n}{k} \frac{(n-k)!}{n!} \times p_{n-k} = \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!}$.

3. En passant à la limite, on a clairement $\lim_{n \rightarrow \infty} P(X_n = k) = \frac{e^{-1}}{k!}$, ce qui permet de conclure que X_n converge en loi vers la loi de Poisson de paramètre 1.

16 Soient n le nombre de lancers du dé, X la variable aléatoire égale au nombre de parties où le résultat est pair, et G le gain du joueur au bout des n parties.

La variable X suit la loi binomiale $\mathcal{B}\left(n, \frac{1}{2}\right)$, avec $E(X) = \frac{n}{2}$ et $V(X) = \frac{n}{4}$.

$G = 1 \times X + (-1) \times (n - X) = 2X - n$, avec $E(G) = 0$ et $V(G) = n$.

On cherche à majorer n pour que $P(G > -21) \geq 0,95$.

- On suppose que l'on est dans les conditions de l'approximation normale de la loi binomiale, et, avec correction de continuité,

$$[G > -21] = [G > -20,5] = [2X - n > -20,5] = \left[\frac{X - \frac{n}{2}}{\frac{\sqrt{n}}{2}} > \frac{-20,5}{\sqrt{n}} \right].$$

La condition $P(G > -21) \geq 0,95$ équivaut donc environ à $P\left(\frac{X - \frac{n}{2}}{\frac{\sqrt{n}}{2}} > \frac{-20,5}{\sqrt{n}}\right) \geq 0,95$,

c'est-à-dire à la condition $P\left(\frac{X - \frac{n}{2}}{\frac{\sqrt{n}}{2}} \leq \frac{-20,5}{\sqrt{n}}\right) < 0,05$, soit $\Phi\left(\frac{-20,5}{\sqrt{n}}\right) < 0,05$, où Φ

désigne la fonction de répartition de la loi normale centrée réduite.

La condition $\Phi\left(\frac{-20,5}{\sqrt{n}}\right) < 0,05$ équivaut à $\Phi\left(\frac{20,5}{\sqrt{n}}\right) > 0,95$.

Dans la table numérique donnant les valeurs de Φ , on obtient $\Phi(1,645) \approx 0,95$, et $\Phi(x) > 0,95$ si, et seulement si, $x > 1,645$. On prendra $\frac{20,5}{\sqrt{n}} > 1,645$, ce qui donne

$n > 155,3$, soit, puisque n est un entier, $n \geq 156$.

Sans la correction de continuité, on trouve $n \geq 163$.

- L'inégalité de Bienaymé-Tchebychev appliquée à G donne $P(|G| \geq 21) \leq 1 - \frac{n}{441}$.

Or, $[|G| \geq 21] = [G \geq 21] \cup [G \leq -21]$. Compte tenu de la symétrie du jeu, on a $P(|G| \geq 21) = 2P(G \leq -21)$.

La condition de l'énoncé ($P(G \leq -21) \leq 0,05$) sera remplie dès que $1 - \frac{n}{441} \leq 0,1$, ce qui donne $n \geq 396,9$, soit $n \geq 397$.

En utilisant la correction de continuité, on trouverait $n \geq 379$.

Comme toujours, on constate que l'inégalité de Bienaymé-Tchebychev donne des résultats très médiocres.

17 On sait que $E(X) = \frac{n}{3}$, $V(X) = \frac{2n}{9}$, et que $E(F) = m_F = \frac{1}{3}$, $V(X) = \frac{2}{9n}$.

• On suppose que l'on est dans les conditions d'approximation de la loi binomiale par la loi normale.

Soit alors $Y = \frac{X - \frac{n}{3}}{\frac{\sqrt{2n}}{3}}$. La fonction de répartition Φ de la loi normale centrée réduite est une

bonne approximation de la fonction de répartition F_Y de Y . Alors

$$\begin{aligned} \left(\left| F - \frac{1}{3} \right| < 10^{-2} \right) &= \left(\left| X - \frac{n}{3} \right| < n10^{-2} \right) = \left(\left| \frac{X - \frac{n}{3}}{\frac{\sqrt{2n}}{3}} \right| < \frac{n10^{-2}}{\frac{\sqrt{2n}}{3}} \right) \\ &= \left(|Y| < 3\sqrt{\frac{n}{2}} 10^{-2} \right). \end{aligned}$$

Donc

$$P\left(\left|F - \frac{1}{3}\right| < 10^{-2}\right) = P\left(|Y| < 3\sqrt{\frac{n}{2}} 10^{-2}\right) = P\left(-3\sqrt{\frac{n}{2}} 10^{-2} < Y < 3\sqrt{\frac{n}{2}} 10^{-2}\right)$$

admet pour valeur approchée, par cette méthode,

$$\Phi\left(3\sqrt{\frac{n}{2}} 10^{-2}\right) - \Phi\left(-3\sqrt{\frac{n}{2}} 10^{-2}\right) = 2\Phi\left(3\sqrt{\frac{n}{2}} 10^{-2}\right) - 1.$$

Pour que $P\left(\left|F - \frac{1}{3}\right| < 10^{-2}\right) \geq 0,98$, il suffit alors que $2\Phi\left(3\sqrt{\frac{n}{2}} 10^{-2}\right) - 1 \geq 0,98$, c'est-à-dire que $\Phi\left(3\sqrt{\frac{n}{2}} 10^{-2}\right) \geq 0,99$.

Après consultation de la table numérique donnant les valeurs de la fonction Φ (p553), et compte tenu de la monotonie de la fonction Φ , on constate qu'il suffit alors de prendre $3\sqrt{\frac{n}{2}} 10^{-2} \geq 2,33$, ce qui devrait donner $n \geq 12064,2$. Comme n est un entier, on prendra comme résultat, $n \geq 12065$.

► **Remarque**

Il va de soi que les résultats obtenus peuvent varier légèrement en fonction de la table numérique utilisée, du fait que l'on réalise, ou non, une interpolation linéaire,...

• L'inégalité de Bienaymé-Tchebychev donne $P\left(\left|F - \frac{1}{3}\right| < 10^{-2}\right) \geq 1 - \frac{2}{10^{-4}}$. Pour que la condition de l'énoncé soit remplie, il suffit donc que $1 - \frac{2}{10^{-4}} \geq 0,98$, soit $n \geq \frac{10^6}{9}$. Comme n est entier, on prendra $n \geq 111112$.

Comme toujours, on constate que l'inégalité de Bienaymé-Tchebychev donne des résultats très médiocres.

Chapitre 12

1 1. La fonction de répartition F_X est donnée par

$$\begin{aligned} F_X(x) &= P\left(-\frac{1}{c} \ln U \leq x\right) = P(-\ln U \leq cx) = P(\ln U \geq -cx) \\ &= P(U \geq e^{-cx}) = 1 - F_U(e^{-cx}) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-cx} & \text{si } x \geq 0 \end{cases} \end{aligned}$$

La variable aléatoire X suit donc la loi exponentielle de paramètre c .

2. Voir 4.2 p. 531.

2 1. a) On sait que, pour tout réel x , $F_X(x) \in [0, 1]$. Il en résulte que

- si $t < 0$, $\{x \mid F_X(x) \geq t\} = \mathbb{R}$, et par suite, $F_X^*(t) = -\infty$,
- si $t > 1$, $\{x \mid F_X(x) \geq t\} = \emptyset$, et par suite, $F_X^*(t) = +\infty$.

Il en résulte que la fonction F_X^* est définie sur $[0, 1]$.

b) Si X est une variable uniforme discrète sur $\{0, 1, 2, 3\}$, alors

$$\begin{cases} F_X(x) = 0 & \text{si } x < 0 \\ F_X(x) = \frac{1}{4} & \text{si } 0 \leq x < 1 \\ F_X(x) = \frac{1}{2} & \text{si } 1 \leq x < 2 \\ F_X(x) = \frac{3}{4} & \text{si } 2 \leq x < 3 \\ F_X(x) = 1 & \text{si } 3 \leq x \end{cases}$$

On en déduit que

$$\begin{cases} F_X^*(x) = 0 & \text{si } 0 \leq x \leq \frac{1}{4} \\ F_X^*(t) = 1 & \text{si } \frac{1}{4} < x \leq \frac{1}{2} \\ F_X^*(x) = 2 & \text{si } \frac{1}{2} < x \leq \frac{3}{4} \\ F_X^*(x) = 3 & \text{si } \frac{3}{4} < x \leq 1 \end{cases}$$

Les représentations graphiques ne posent pas de problème. Supposons que les courbes soient représentées dans un repère orthonormal. Que remarque-t-on en faisant subir à l'une des deux courbes la symétrie orthogonale par rapport à la première bissectrice ?

2. a) • Il est clair que $x \in \{y \mid F(y) \geq F(x)\}$, donc $x \geq \inf\{y \mid F(y) \geq F(x)\}$, c'est-à-dire $F_X^*(F_X(x)) \leq x$.

• La fonction F_X étant croissante, il est clair que $]F_X^*(t), +\infty[\subset \{x \mid F_X(x) \geq t\}$. Considérons alors une suite décroissante (u_k) d'éléments de $\{x \mid F_X(x) \geq t\}$ convergeant vers $F_X^*(t)$. La fonction F_X étant continue à droite en tout point (comme toute fonction de répartition), en passant à la limite dans l'inégalité $F_X(u_k) \geq t$, on en déduit que $F_X(F_X^*(t)) \geq t$.

• La propriété $\{x \mid F_X(x) \geq t\} = [F_X^*(t), +\infty[$ découle naturellement de ce qui a été prouvé ci-dessus.

- La propriété $F_X(x) \geq t \iff x \geq F_X^*(t)$ (et non $X \geq F_X^*(t)$ comme il est écrit par erreur) est une autre façon d'exprimer la propriété précédente.
- b)** On a vu que, pour tout réel y , la condition $y \in [F_X^*(F_X(x)), +\infty[$ équivaut à la condition $y \in \{y \mid F(y) \geq F(x)\}$.
Soit y un réel satisfaisant à ces conditions. Si $y < x$, alors, puisque F_X est strictement croissante, $F(y) < F(x)$, ce qui serait contradictoire. Donc, $\forall y \in [F_X^*(F_X(x)), +\infty[$, $y \geq x$. Cela prouve que $x \leq F_X^*(F_X(x))$. Compte tenu des résultats de la sous-question précédente, on en déduit que, si F_X est strictement croissante, pour tout réel x , on a $F_X^*(F_X(x)) = x$.
Si F_X n'est pas strictement croissante, il existe deux réels distincts x et x' tels que $F_X(x) = F_X(x')$ et donc que $F_X^*(F_X(x)) = F_X^*(F_X(x'))$. Cela suffit à prouver qu'il existe au moins un réel t tel que $F_X^*(F_X(t)) = t$. Pour construire effectivement un contre-exemple, utiliser le 1.b).
- c)** Par définition de $F_X^*(t)$, pour tout réel y , si $y < F_X^*(t)$, alors $F_X(y) < t$. Considérons alors une suite croissante (y_k) de réels convergeant vers $F_X^*(t)$. Puisque F_X est continue à gauche, en passant à la limite dans $F_X(y_k) < t$, on obtient $F_X(F_X^*(t)) \leq t$, ce qui, joint à un résultat du 1.a) permet de conclure que, si F_X est continue, pour tout $t \in]0, 1[$, on a $F_X(F_X^*(t)) = t$.
Pour construire un contre-exemple, utilisons encore le 1.b). Dans cet exemple, on a $F_X^*(1/4) = 0$ et $F_X(0) = 0$, donc $F_X(F_X^*(1/4)) = \frac{1}{4}$.
- 3.** On sait que $F_X(x) \geq t \iff x \geq F_X^*(t)$, et que F_U est continue (et croissante). Donc, pour tout réel x ,
 $P(F_X^*(U) \leq x) = P(F_X(F_X^*(U)) \leq F_X(x)) = P(U \leq F_X(x)) = F_X(x)$ (car $F_X(x) \in [0, 1]$).
Cela prouve bien que $F_X^*(U)$ admet F_X comme fonction de répartition.
- 4. a)** La fonction F_X étant continue, strictement croissante de \mathbb{R} dans $[0, 1]$, c'est une bijection \mathbb{R} sur son image $F_X(\mathbb{R}) \subset [0, 1]$. S'il existait un réel x tel que $F_X(x) = 0$, alors, pour tout x' inférieur à x , on aurait aussi, $F_X(x') = 0$, et F_X ne serait pas strictement croissante sur \mathbb{R} . On prouve de même que 1 n'a pas d'antécédent par F_X , et donc F_X est une bijection de \mathbb{R} sur $]0, 1[$.
Dans ces conditions, et d'après les questions précédentes, on a alors $F_X \circ F_X^* = F_X^* \circ F_X = id$, c'est-à-dire $F_X^* = F_X^{-1}$.
- b)** $F_X(X)$ prend ses valeurs dans $[0, 1]$. Il en résulte que, si $x < 0$, alors $P(F_X(X) \leq x) = 0$ et que, si $x > 1$, alors $P(F_X(X) \leq x) = 1$.
La bijection F_X étant strictement croissante, il en est de même de sa bijection réciproque F_X^{-1} . On en déduit que, pour tout x de $[0, 1]$,

$$P(F_X(X) \leq x) = P(X \leq F_X^{-1}(x)) = F(F_X^{-1}(x)) = x.$$

Cela prouve bien que $F_X(X)$ suit la loi uniforme sur $[0, 1]$.

- 5.** Si l'on sait calculer les valeurs prises par F_X^* , on sait simuler la variable X . En effet, d'après la question 3., si U suit la loi uniforme sur $[0, 1]$, alors $F_X^*(U)$ admet F_X comme fonction de répartition. Alors, si x est une réalisation de la loi uniforme sur $[0, 1]$, $y = F_X^*(x)$ est une réalisation de X .
Dans le cas d'une variable exponentielle X de paramètre λ , la fonction F_X^* est définie sur $[0, 1[$ par $F_X^*(x) = -\frac{1}{\lambda} \ln(1-x)$. Le programme Pascal correspondant fait l'objet du 4.2 p. 531.

3 Cette question est traitée au 3.2 p. 446.

Pour ce qui concerne la réciproque, elle est malheureusement fautive ! Il suffit de constater qu'il existe des suites (X_n) de variables aléatoires qui convergent en loi vers une variable aléatoire X , sans que la suite $E(X_n)$ n'admette la limite $E(X)$, ou même sans que cette suite n'admette de limite (voir p. 394).

4 **1. a)** Soit V l'événement « la personne répond selon ses convictions » dont la probabilité est α . La formule des probabilités totales appliquée au système complet d'événements $\{V, \bar{V}\}$ donne $q = \alpha p + (1 - \alpha)(1 - p)$.

b) D'après le cours, la fréquence empirique S_n , c'est-à-dire la proportion de réponses « oui » parmi les n personnes interrogées, est un estimateur sans biais (p. 442) et convergent (p. 448) de q .

c) Si $\alpha = \frac{1}{2}$, $q = \frac{1}{2}$ quel que soit p . Ce cas ne présente pas d'intérêt.

Lorsque $\alpha \neq \frac{1}{2}$, on a $p = \frac{q - 1 + \alpha}{2\alpha - 1}$. On peut donc considérer l'estimateur $T_n = \frac{S_n - 1 + \alpha}{2\alpha - 1}$.

Comme T_n est l'image de l'estimateur convergent S_n par une fonction continue, il est lui-même convergent (voir le théorème 12 p. 449).

2. a) Par linéarité de l'espérance, et du fait que $E(S_n) = q$, on obtient $E(T_n) = p$, ce qui n'est pas une surprise, l'estimateur T_n ayant été construit dans ce but. Cet estimateur est donc un estimateur sans biais de p .

b) On sait que la variance de S_n est $V(S_n) = \frac{q(1-q)}{n}$. On obtient alors

$$V(T_n) = \frac{q(1-q)}{n(2\alpha-1)^2}, \text{ c'est-à-dire, en remplaçant } q \text{ par sa valeur, et après calcul,}$$

$$V(T_n) = \frac{p(1-p)}{n} + \frac{\alpha(1-\alpha)}{n(2\alpha-1)^2}.$$

Comme T_n est sans biais et que $\lim_{n \rightarrow +\infty} V(T_n) = 0$, alors T_n est convergent.

c) Pour n fixé, $V(T_n)$ tend vers l'infini quand α tend vers $\frac{1}{2}$. Cette variance est minimum lorsque α est soit nul, soit égal à 1, mais dans ces deux cas, la procédure perd son intérêt. Le problème est alors de choisir un α assez éloigné de $\frac{1}{2}$ pour que la variance de l'estimateur ne soit pas trop grande, mais aussi assez éloigné de 0 et de 1 pour que la confidentialité reste crédible.

3. a) Il faut bien entendu lire qu'une réalisation de S_n est 0, 425. La réalisation correspondante de T_n est alors $\bar{T}_n = 0, 6125$.

En appliquant les formules du cours, avec $t_\alpha = 1, 96$,

- avec $I = \left[\bar{T}_n - t_\alpha \frac{1}{2\sqrt{n}}, \bar{T}_n + t_\alpha \frac{1}{2\sqrt{n}} \right]$, on peut obtenir l'intervalle de confiance $]0, 5815; 0, 6435[$

- avec $I = \left[\bar{T}_n - t_\alpha \sqrt{\frac{\bar{T}_n(1-\bar{T}_n)}{n}}, \bar{T}_n + t_\alpha \sqrt{\frac{\bar{T}_n(1-\bar{T}_n)}{n}} \right]$, on peut obtenir l'intervalle de confiance $]0, 58186; 0, 64314[$, ce qui n'est pas significativement meilleur qu'avec la première formule.

b) Voici une réponse possible.

```

BEGIN
  RANDOMIZE ;
  FOR i := 1 TO 50 DO
    BEGIN
      FOR j := 1 TO 1000 DO
        BEGIN
          a := RANDOM ;
          IF x < 0.425 THEN R := R+1
        END ;
        S := R/1000 ;
        T := (S-1/6)/(1/3-1) ;
        a := T-1.96*1/2 SQRT(1000) ;
        b := T+1.96*1/2 SQRT(1000) ;
        WRITELN (a, ' ', b)
      END
    END
  END ;

```

5 1. a) X suit la loi géométrique de paramètre $\frac{m}{N}$, d'espérance $E(X) = \frac{N}{m}$ et de variance $V(X) = \frac{N(N-m)}{m^2}$.

On sait que la fréquence empirique $F_n = \frac{1}{n} \sum_{k=1}^n X_k$ est un estimateur sans biais de $\frac{N}{m}$.

Par linéarité, mF_n est un estimateur sans biais de N .

La variance de mF_n est égale à $m^2 \frac{V(X)}{n} = \frac{N(N-m)}{n}$. Elle tend donc vers 0 quand n tend vers l'infini. Comme c'est un estimateur sans biais, il est convergent.

L'inconvénient majeur de cet estimateur est qu'il peut, à chaque réalisation de la variable X , nécessiter un très grand nombre de tirages.

b) En posant $p = \frac{m}{N}$, on a $E(Y) = \sum_{k=1}^b kp(1-p)^{k-1} = p \sum_{k=1}^b k(1-p)^{k-1}$

Posons $1-p = x$. On a alors $\sum_{k=1}^b k(1-p)^{k-1} = \sum_{k=1}^b kx^{k-1}$.

Posons $f(x) = \sum_{k=1}^b x^k = x \sum_{k=0}^{b-1} x^k = x \frac{1-x^b}{1-x} = \frac{x}{1-x}(1-x^b)$.

On a alors $\sum_{k=1}^b kx^{k-1} = f'(x) = \frac{1}{(1-x)^2}(1-x^b) - \frac{x}{1-x}bx^{b-1}$.

Avec $x = 1-p$, on obtient alors

$$E(Y) = p \left[\frac{1}{p^2}(1 - (1-p)^b) - \frac{1-p}{p}b(1-p)^{b-1} \right] = \frac{1}{p} - (1-p)^b \left(\frac{1}{p} + b \right).$$

C'est-à-dire $E(Y) = \frac{N}{m} - \left(1 - \frac{m}{N}\right)^b \left(\frac{N}{m} + b\right)$.

On constate que $E(mY) = N - \left(1 - \frac{m}{N}\right)^b (N + mb)$. Le terme $1 - \frac{m}{N}$ est inférieur à 1.

On peut choisir b de façon que $\left(1 - \frac{m}{N}\right)^b (N + mb)$ soit petit devant N . La variable aléatoire mY est un estimateur biaisé de N .

Soient Y_1, Y_2, \dots, Y_p p variables aléatoires suivant la même loi que Y .

On considère la variable aléatoire $Z_b = \frac{\sum_{k=0}^b mY_k}{b}$. On a clairement

$$E(Z_b) = E(mY) = N - \left(1 - \frac{m}{N}\right)^b (N + mb).$$

La variable aléatoire Z_b est un estimateur biaisé de N .

D'autre part, $1 - \frac{m}{N}$ étant strictement compris entre 0 et 1, $\lim_{b \rightarrow \infty} \left(1 - \frac{m}{N}\right)^b (N + mb) = 0$, la variable aléatoire Z_b est un estimateur asymptotiquement sans biais de N .

Enfin, on a $V(Z_b) = \frac{m^2}{b} V(Y)$. Or, il est clair que $V(Y) < V(X)$, donc $0 < V(Z_b) < \frac{N(N-m)}{b}$, et $\lim_{b \rightarrow \infty} V(Z_b) = 0$. Comme Z_b est un estimateur asymptotiquement sans biais, il est convergent.

► **Remarque**

On peut douter de la pertinence d'introduire un tel estimateur, qui n'apporte que peu d'amélioration par rapport à l'estimateur rencontré à la première question.

2. a) La variable aléatoire Z suit la loi de Bernoulli de paramètre $\frac{m}{N}$, avec $E(Z) = \frac{m}{N}$ et $V(Z) = \frac{m(N-m)}{N^2}$.
- b) La variable aléatoire S_n suit la loi binomiale de paramètre $\left(n, \frac{m}{N}\right)$, avec $E(S_n) = n\frac{m}{N}$ (on peut justifier par la linéarité de l'espérance) et $V(S_n) = n\frac{m(N-m)}{N^2}$.
- c) L'expression $T_n = \frac{nm}{S_n}$ ne peut pas définir une variable aléatoire, car elle n'est pas définie pour $S_n = 0$, événement dont la probabilité n'est pas nulle. En d'autres termes, la fonction $g = x \mapsto \frac{nm}{x}$ n'est pas définie sur $S_n(\Omega)$ (voir le théorème 16 p. 769 du Tome 1).
- d) La fonction $h = x \mapsto \frac{1}{x+1}$ est définie sur $S_n(\Omega)$, donc $U_n = \frac{1}{S_n+1}$ est une variable aléatoire (voir le théorème 16 p. 769 du Tome 1).
Pour calculer l'espérance de U_n , on utilise le théorème de transfert

$$E(U_n) = \sum_{k=0}^n \frac{1}{k+1} \binom{n}{k} \left(\frac{m}{N}\right)^k \left(1 - \frac{m}{N}\right)^{n-k}.$$

Or, $\binom{n+1}{k+1} = \frac{n+1}{k+1} \binom{n}{k}$ donc

$$\begin{aligned} E(U_n) &= \sum_{k=0}^n \frac{1}{n+1} \binom{n+1}{k+1} \left(\frac{m}{N}\right)^k \left(1 - \frac{m}{N}\right)^{n-k} \\ &= \frac{N}{m(n+1)} \sum_{k=0}^n \binom{n+1}{k+1} \left(\frac{m}{N}\right)^{k+1} \left(1 - \frac{m}{N}\right)^{n+1-k-1}. \end{aligned}$$

C'est-à-dire

$$\begin{aligned} E(U_n) &= \frac{N}{m(n+1)} \sum_{h=1}^{n+1} \binom{n+1}{h} \left(\frac{m}{N}\right)^h \left(1 - \frac{m}{N}\right)^{n+1-h} \\ &= \frac{N}{m(n+1)} \left(1 - \left(1 - \frac{m}{N}\right)^{n+1}\right). \end{aligned}$$

e) L'estimateur $V_n = \frac{nm}{S_n + 1}$ est donc tel que $E(V_n) = \frac{nN}{n+1} \left(1 - \left(1 - \frac{m}{N}\right)^{n+1}\right)$.

Comme $1 - \frac{m}{N}$ est strictement compris entre 0 et 1, on a $\lim_{n \rightarrow +\infty} E(V_n) = N$. Donc, V_n est un estimateur asymptotiquement sans biais de N .

► **Remarque**

On aurait pu choisir par exemple $V'_n = \frac{(n+1)m}{S_n + 1}$, ce qui simplifie un peu l'expression de l'espérance, voir l'exercice 8.

Pour étudier la convergence de l'estimateur V_n de N , envisageons la variable aléatoire $I_n = \frac{1}{V_n} = \frac{S_n + 1}{nm}$. On a $E(I_n) = \frac{1}{nm}(E(S_n) + 1) = \frac{1}{nm} \left(n \frac{m}{N} + 1\right) = \frac{1}{N} + \frac{1}{nm}$. C'est donc un estimateur biaisé, mais asymptotiquement sans biais, de $\frac{1}{N}$.

D'autre part, $V(I_n) = \frac{1}{n^2 m^2} V(S_n) = \frac{1}{n^2 m^2} n \frac{m(N-m)}{N^2} = \frac{(N-m)}{nmN^2}$. On constate sans peine que cette variance tend vers 0 quand n tend vers l'infini. Comme I_n est un estimateur asymptotiquement sans biais de $\frac{1}{N}$, c'est un estimateur convergent.

La fonction $g = \left[x \mapsto \frac{1}{x} \right]$ est définie et continue sur \mathbb{R}_+^* , donc $g(I_n)$ est un estimateur convergent de $g\left(\frac{1}{N}\right)$ (théorème 12 p. 449 qu'il faut savoir démontrer), c'est-à-dire que V_n est un estimateur convergent de N .

6 1. a) $L(x_1, x_2, \dots, x_n, \theta)$ représente la probabilité pour que la réalisation (x_1, x_2, \dots, x_n) provienne bien d'un échantillon (X_1, X_2, \dots, X_n) correspondant à la valeur θ du paramètre de la loi de probabilité envisagée.

b) L'échantillon étant *iid*, on a $L(x_1, x_2, \dots, x_n, \theta) = \prod_{k=1}^n P_\theta(x_k)$, où $P_\theta(x_k)$ désigne la probabilité de x_k pour la loi correspondant à la valeur θ du paramètre.

2. a) Pour la loi de Bernoulli de paramètre p , on a, pour tout $x_k \in \{0, 1\}$,

$$P_p(x_k) = p^{x_k}(1-p)^{1-x_k}.$$

On aura donc, pour $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$,

$$L(x_1, x_2, \dots, x_n, p) = p^{\left(\sum_{k=1}^n x_k\right)} (1-p)^{\left(n - \sum_{k=1}^n x_k\right)}.$$

- b) Pour la loi binomiale $\mathcal{B}(k, p)$, k étant connu, on a, pour tout $x_j \in \llbracket 0, k \rrbracket$,

$$P_p(x_j) = \binom{k}{x_j} p^{x_j} (1-p)^{(k-x_j)}.$$

On peut donc mettre, pour $(x_1, x_2, \dots, x_n) \in \llbracket 0, k \rrbracket^n$, la fonction de vraisemblance sous la forme

$$L(x_1, x_2, \dots, x_n, p) = \prod_{j=1}^n \binom{k}{x_j} p^{x_j} \prod_{j=1}^n (1-p)^{(k-x_j)}.$$

- c) Pour la loi géométrique de paramètre p , on a, pour tout $x_i \in \mathbb{N}^*$, $P_p(x_i) = p(1-p)^{x_i-1}$.

On aura donc, pour $(x_1, x_2, \dots, x_n) \in \mathbb{N}^{*n}$, $L(x_1, x_2, \dots, x_n, p) = p^n (1-p)^{\left(\sum_{i=1}^n x_i - n\right)}$.

3. a) Pour que la fonction partielle associée à θ (qui est une fonction d'une variable, de classe C^2 dépendant de n paramètres) admette un maximum, il faut, et il suffit, que la dérivée partielle $\frac{\partial L}{\partial \theta}(x_1, x_2, \dots, x_n, \theta)$ s'annule et que la dérivée partielle seconde $\frac{\partial^2 L}{\partial \theta^2}(x_1, x_2, \dots, x_n, \theta)$ soit strictement négative.

- b) Par définition, la fonction L prend des valeurs positives ou nulles. Si elle n'est pas identiquement nulle, et qu'elle admet un maximum, il sera strictement positif. Comme la fonction \ln est croissante, elle respecte l'ordre, et le maximum de $L(x_1, x_2, \dots, x_n, \theta)$ sera atteint au même point que le maximum de $\ln L(x_1, x_2, \dots, x_n, \theta)$, c'est-à-dire en un point où la dérivée partielle $\frac{\partial \ln L}{\partial \theta}(x_1, x_2, \dots, x_n, \theta)$ s'annule et où la dérivée partielle seconde $\frac{\partial^2 \ln L}{\partial \theta^2}(x_1, x_2, \dots, x_n, \theta)$ est strictement négative.

4. a) Pour la loi de Bernoulli de paramètre p , on a

$$L(x_1, x_2, \dots, x_n, p) = p^{\left(\sum_{k=1}^n x_k\right)} (1-p)^{\left(n - \sum_{k=1}^n x_k\right)},$$

donc $\ln L(x_1, x_2, \dots, x_n, p) = \left(\sum_{k=1}^n x_k\right) \ln p + \left(n - \sum_{k=1}^n x_k\right) \ln(1-p)$.

$$\text{Alors, } \frac{\partial \ln L}{\partial p}(x_1, x_2, \dots, x_n, p) = \left(\sum_{k=1}^n x_k\right) \frac{1}{p} - \left(n - \sum_{k=1}^n x_k\right) \frac{1}{1-p}.$$

La condition $\frac{\partial \ln L}{\partial p}(x_1, x_2, \dots, x_n, p) = 0$ conduit à une seule solution $\hat{p} = \frac{\sum_{k=1}^n x_k}{n}$.

D'autre part, on a $\frac{\partial^2 \ln L}{\partial p^2}(x_1, x_2, \dots, x_n, \theta) = -\left(\sum_{k=1}^n x_k\right) \frac{1}{p^2} - (n - \sum_{k=1}^n x_k) \frac{1}{(1-p)^2}$, et la deuxième condition est remplie pour tout p , donc en particulier pour \hat{p} .

Pour tout (x_1, x_2, \dots, x_n) la fonction de vraisemblance admet donc un maximum en \hat{p} . On en déduit que si (X_1, X_2, \dots, X_n) est un échantillon *iid* d'une variable de Bernoulli de paramètre p , l'estimateur du maximum de vraisemblance de p est la fréquence empirique $\bar{X} = \frac{\sum_{k=1}^n X_k}{n}$.

b) Pour la loi binomiale $\mathcal{B}(k, p)$, où k est connu, on a

$$L(x_1, x_2, \dots, x_n, p) = \prod_{j=1}^n \binom{k}{x_j} p^{x_j} (1-p)^{(k-x_j)},$$

et donc $\ln L(x_1, x_2, \dots, x_n, p) = \sum_{j=1}^n \ln \binom{k}{x_j} + \sum_{j=1}^n x_j \ln p + \sum_{j=1}^n (k - x_j) \ln(1-p)$.

Alors, $\frac{\partial \ln L}{\partial p}(x_1, x_2, \dots, x_n, p) = \frac{1}{p} \sum_{j=1}^n x_j - \frac{1}{1-p} \sum_{j=1}^n (k - x_j)$.

La condition $\frac{\partial \ln L}{\partial p}(x_1, x_2, \dots, x_n, p) = 0$ équivaut à $(1-p) \sum_{j=1}^n x_j - pnk + p \sum_{j=1}^n x_j = 0$

qui admet pour unique solution $\hat{p} = \frac{\sum_{j=1}^n x_j}{nk}$.

D'autre part, on a $\frac{\partial^2 \ln L}{\partial p^2}(x_1, x_2, \dots, x_n, p) = -\frac{1}{p^2} \sum_{j=1}^n x_j - \frac{1}{(1-p)^2} \sum_{j=1}^n (k - x_j)$. Comme pour tout j , $x_j \in \llbracket 0, k \rrbracket$, cette expression est négative pour tout p , et la deuxième condition est remplie en particulier pour \hat{p} .

Pour tout (x_1, x_2, \dots, x_n) la fonction de vraisemblance admet donc un maximum en \hat{p} . On en déduit que si (X_1, X_2, \dots, X_n) est un échantillon *iid* d'une variable binomiale

$\mathcal{B}(k, p)$, où k est connu, l'estimateur du maximum de vraisemblance de p est $\hat{X} = \frac{\sum_{j=1}^n X_j}{nk}$.

c) Pour la loi géométrique de paramètre p , on a $L(x_1, x_2, \dots, x_n, p) = p^n (1-p)^{\left(\sum_{i=1}^n x_i - n\right)}$, donc $\ln L(x_1, x_2, \dots, x_n, p) = n \ln p - (n - \sum_{i=1}^n x_i) \ln(1-p)$.

Alors, $\frac{\partial \ln L}{\partial p}(x_1, x_2, \dots, x_n, p) = \frac{n}{p} + (n - \sum_{i=1}^n x_i) \frac{1}{1-p}$.

La condition $\frac{\partial \ln L}{\partial p}(x_1, x_2, \dots, x_n, p) = 0$ conduit à une seule solution $\hat{p} = \frac{n}{\sum_{i=1}^n x_i}$.

D'autre part, on a $\frac{\partial^2 \ln L}{\partial p^2}(x_1, x_2, \dots, x_n, p) = -\frac{n}{p^2} + (n - \sum_{i=1}^n x_i) \frac{1}{(1-p)^2}$. Comme tous

les x_i sont supérieurs ou égaux à 1, $(n - \sum_{i=1}^n x_i)$ est négatif, et la deuxième condition est remplie pour tout p , donc en particulier pour \hat{p} .

Pour tout (x_1, x_2, \dots, x_n) la fonction de vraisemblance admet donc un maximum en \hat{p} . On en déduit que si (X_1, X_2, \dots, X_n) est un échantillon *iid* d'une variable géométrique de paramètre p , l'estimateur du maximum de vraisemblance de p est $\hat{X} = \frac{n}{\sum_{i=1}^n X_i}$.

7 1. Les variables aléatoires X_k suivent toutes la même loi $\mathcal{N}(150; 0, 21^2)$. Comme elles sont indépendantes, la moyenne empirique $\bar{\mu} = \frac{1}{400} \sum_{k=1}^{400}$ suit la loi normale

$$\mathcal{N}\left(150; \frac{0, 21^2}{400}\right).$$

Soit $T = \frac{\bar{\mu} - 150}{0, 0105}$ la variable aléatoire centrée réduite associée à $\bar{\mu}$. Elle suit la loi normale $\mathcal{N}(0; 1)$, dont la fonction de répartition est notée Φ .

La condition $P(|\mu - \bar{\mu}| \leq h) = P(150 - h \leq \bar{\mu} \leq 150 + h) = 0, 95$ équivaut à $P\left(-\frac{h}{0, 0105} \leq T \leq \frac{h}{0, 0105}\right) = 0, 95$, c'est-à-dire à $2\Phi\left(\frac{h}{0, 0105}\right) - 1 = 0, 95$, soit $\Phi\left(\frac{h}{0, 0105}\right) = 0, 975$.

On obtient, à l'aide de la table numérique, $\frac{h}{0, 0105} \approx 1, 96$, d'où $h \approx 0, 0206$.

2. Les variables aléatoires X_k suivent toutes la même loi $\mathcal{N}(28, 2; 0, 027^2)$. Comme elles sont indépendantes, la moyenne empirique $M = \frac{1}{n} \sum_{k=1}^n$ suit la loi normale $\mathcal{N}\left(28, 2; \frac{0, 027^2}{n}\right)$.

Soit $T = \frac{M - 28, 2}{\frac{0, 027}{\sqrt{n}}}$ la variable aléatoire centrée réduite associée à M . Elle suit la loi normale $\mathcal{N}(0; 1)$, dont la fonction de répartition est notée Φ .

La condition $P(28, 195 \leq M \leq 28, 205) = 0, 95$ équivaut alors à

$P\left(-\frac{0, 005}{0, 027} \sqrt{n} \leq T \leq \frac{0, 005}{0, 027} \sqrt{n}\right) = 0, 95$, c'est-à-dire à $2\Phi\left(\frac{5}{27} \sqrt{n}\right) - 1 = 0, 95$, soit $\Phi\left(\frac{5}{27} \sqrt{n}\right) = 0, 975$.

On obtient, à l'aide de la table numérique, $\frac{5}{27} \sqrt{n} \approx 1, 96$, d'où $n \approx 112, 021$.

Comme n est entier, on gardera $n \approx 113$.

8 1. Soit \bar{m} la moyenne observée sur l'échantillon. Il s'agit simplement d'appliquer les formules de la page 458.

À la précision 90%, on obtient l'intervalle de confiance

$$\left[\bar{m} - 1, 645 \frac{200}{6}; \bar{m} + 1, 645 \frac{200}{6}\right] = [\bar{m} - 54, 86; \bar{m} + 54, 86].$$

À la précision 95%, on obtient l'intervalle de confiance

$$\left[\bar{m} - 1, 96 \frac{200}{6}; \bar{m} + 1, 96 \frac{200}{6}\right] = [\bar{m} - 65, 34; \bar{m} + 65, 34].$$

À la précision 99%, on obtient l'intervalle de confiance

$$\left[\bar{m} - 2,575 \frac{200}{6}; \bar{m} + 2,575 \frac{200}{6} \right] = [\bar{m} - 85,84; \bar{m} + 85,84].$$

On remarque que, plus on veut de certitude, moins on a de précision (plus l'intervalle de confiance est grand), ce qui s'explique aisément.

2. De la même façon :

À la précision 90%, on obtient l'intervalle de confiance

$$\left[\bar{m} - 1,645 \frac{200}{20}; \bar{m} + 1,645 \frac{200}{20} \right] = [\bar{m} - 16,45; \bar{m} + 16,45].$$

À la précision 95%, on obtient l'intervalle de confiance

$$\left[\bar{m} - 1,96 \frac{200}{20}; \bar{m} + 1,96 \frac{200}{20} \right] = [\bar{m} - 19,6; \bar{m} + 19,6].$$

À la précision 99%, on obtient l'intervalle de confiance

$$\left[\bar{m} - 2,575 \frac{200}{20}; \bar{m} + 2,575 \frac{200}{20} \right] = [\bar{m} - 25,75; \bar{m} + 25,75].$$

On remarque (bien sûr) la même chose qu'à la question 1. On observe de plus l'influence de la taille de l'échantillon sur la précision des estimations.

- 9 1. Des calculs sans histoire, éventuellement exécutés sur un tableur, conduisent à $\bar{x} = 92,2$ et à $\sigma \approx 7,17$

2. La moyenne empirique est un estimateur sans biais de la moyenne de la loi parente. Une estimation ponctuelle de μ est donc 92,2. Par contre, la variance empirique S_n^2 est un estimateur biaisé de la variance de la loi parente. On sait que $E(S_n^2) = \sigma^2 = \frac{n-1}{n}s^2$.

Une estimation ponctuelle de la variance s^2 de la loi parente est donc $\frac{n}{n-1}E(S_n^2)$, et une

estimation ponctuelle de s est $\sqrt{\frac{n}{n-1}E(S_n^2)}$. Avec les données de l'énoncé, on obtient pour s une estimation de 7,206.

3. En prenant pour écart type de la loi parente la valeur $s = 7,206$ trouvée, on obtient un intervalle de confiance au coefficient 0,99 sous la forme

$$\left[\frac{92,2 - 2,575 \times 7,206}{10}; \frac{92,2 + 2,575 \times 7,206}{10} \right] = [90,34; 94,06].$$

4. Au coefficient 0,95, l'amplitude de l'intervalle de confiance obtenu par la même méthode que ci-dessus est

$$2 \frac{1,96s}{\sqrt{n}} = 2 \frac{1,96 \times 7,206}{\sqrt{n}} = 2 \frac{14,12376}{\sqrt{n}}.$$

On veut que cette amplitude soit inférieure ou égale à 0,5, c'est-à-dire que $2 \frac{14,12376}{\sqrt{n}} \leq 0,5$. Tout calcul fait, on obtient $n \geq 3192$.

10 1. Il s'agit d'estimer le paramètre d'une loi de Bernoulli.

Un intervalle de confiance pour p au niveau 95% est donné par

$$\left[0,51 - 1,96 \frac{1}{2\sqrt{100}}; 0,51 + 1,96 \frac{1}{2\sqrt{100}} \right]$$

(voir p. 460), c'est-à-dire $[0,412; 0,608]$. Compte tenu des approximations de la méthode et des calculs, on gardera $[0,41; 0,61]$.

2. On obtient ici l'intervalle de confiance $\left[0,51 - 1,96 \frac{1}{2\sqrt{1000}}; 0,51 + 1,96 \frac{1}{2\sqrt{1000}} \right]$, soit $[0,479; 0,541]$. En toute rigueur, on devrait conclure à $[0,47; 0,55]$ (pour être sûr d'une probabilité supérieure à 0,95). Mais on ne court pas un grand risque en gardant $[0,48; 0,54]$.
3. La longueur de l'intervalle de confiance obtenu avec la formule utilisée est inférieure à $\frac{1,96}{\sqrt{n}}$. On cherche n pour que $\frac{1,96}{\sqrt{n}} \leq 0,04$, et l'on obtient $n \geq 2401$.

► **Remarque**

Dans cet exercice, on aurait pu utiliser la deuxième formule de la page 460. On reprendra les calculs avec cette formule, et l'on obtiendra des résultats très voisins (et même égaux après arrondi) de ceux que l'on a obtenus.

11 1. Il suffit d'appliquer les formules. La moyenne observée est $\frac{30,25}{25} = 1,21$.

On trouve un intervalle de confiance à 90 % sous la forme

$$\left[1,21 - 1,65 \frac{1}{\sqrt{25}}; 1,21 + 1,65 \frac{1}{\sqrt{25}} \right] = [0,88; 1,54].$$

Cet encadrement est médiocre compte tenu de la petitesse de l'échantillon.

2. La moyenne observée est $\frac{484}{400} = 1,21$.

On trouve un intervalle de confiance à 90% sous la forme

$$\left[1,21 - 1,65 \frac{1}{\sqrt{400}}; 1,21 + 1,65 \frac{1}{\sqrt{400}} \right] = [1,1275; 1,2925].$$

Compte tenu du caractère approché de la méthode et de certains calculs, il serait déraisonnable de garder toutes les décimales. On devrait garder $[1,12; 1,30]$, mais on ne prendrait pas un gros risque en gardant $[1,13; 1,29]$.

- 12 **1. a)** En utilisant le théorème de transfert, et avec le changement de variable (autorisé pour x positif) $u = \frac{t^2}{2\sigma^2}$, on obtient pour tout i , puisque l'échantillon est *iid*,

$$\begin{aligned} E(|X_i|) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} |t| e^{-\frac{t^2}{2\sigma^2}} dt = 2 \frac{1}{\sigma\sqrt{2\pi}} \int_0^{+\infty} t e^{-\frac{t^2}{2\sigma^2}} dt \\ &= 2 \frac{\sigma}{\sqrt{2\pi}} \int_0^{+\infty} e^{-u} du = \sigma \sqrt{\frac{2}{\pi}}. \end{aligned}$$

- b)** Par linéarité de l'espérance, il en résulte que $E\left(\sqrt{\frac{\pi}{2}}|X_i|\right)$. Il en résulte, compte tenu des propriétés de la moyenne empirique, que $T_n = \sqrt{\frac{\pi}{2}} \left(\frac{1}{n} \sum_{i=1}^n |X_i|\right) = \sqrt{\frac{\pi}{2}} D_n$ (avec $D_n = \frac{1}{n} \sum_{i=1}^n |X_i|$) est un estimateur sans biais de σ .

- 2. a)** Pour tout i , $E(|X_i|^2) = E(X^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2\sigma^2}} dt = \sigma^2$ (c'est la variance de la loi $\mathcal{N}(0, \sigma^2)$). Il en résulte que $V(|X_i|) = \left(1 - \sqrt{\frac{2}{\pi}}\right) \sigma^2$.

- b)** La variance de T_n est alors donnée par $V(T_n) = \frac{\pi}{2n} V(|X_i|) = \left(\frac{\pi}{2} - 1\right) \frac{\sigma^2}{n}$. Cette variance est de limite nulle quand n tend vers l'infini. On en déduit clairement que T_n , qui est sans biais, est convergent.

- 3.** D'après le théorème de la limite centrée, la variable centrée réduite $\frac{T_n - \sigma}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{\pi}{2} - 1}}$ converge en loi vers la loi normale centrée réduite. On peut donc approcher cette variable centrée réduite par une variable normale centrée réduite. Il en résulte que $P\left[-1,96 \leq \frac{T_n - \sigma}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{\pi}{2} - 1}} \leq 1,96\right] \approx 0,95$ (on lit dans la table, ou l'on se souvient de cette valeur, souvent rencontrée) c'est-à-dire que, approximativement,

$$\sigma \left[1 - 1,96 \sqrt{\frac{\pi - 2}{2n}}\right] \leq T_n \leq \left[1 + 1,96 \sqrt{\frac{\pi - 2}{2n}}\right],$$

d'où l'encadrement (aux erreurs près dues à la méthode et aux arrondis des calculs)

$$\frac{T_n}{1 + 1,96 \sqrt{\frac{\pi - 2}{2n}}} \leq \sigma \leq \frac{T_n}{1 - 1,96 \sqrt{\frac{\pi - 2}{2n}}};$$

c'est-à-dire $\sqrt{\frac{\pi}{2}} \frac{D_n}{1 + 1,96\sqrt{\frac{\pi-2}{2n}}} \leq \sigma \leq \sqrt{\frac{\pi}{2}} \frac{D_n}{1 - 1,96\sqrt{\frac{\pi-2}{2n}}}$ ce qui donne, avec

les données de l'énoncé un intervalle de confiance de coefficient 0,95 sous la forme [9,25 ; 25,55]. On constate que cet encadrement est de très mauvaise qualité. Cela est dû à ce que l'observation d'une moyenne à 10,84 pour un échantillon des valeurs absolues d'une variable normale de moyenne nulle suppose un grand écart type... Il serait plus cohérent de modifier l'énoncé en supposant que la moyenne obtenue a été de 1,084. On trouve alors un intervalle de confiance à 90% pour σ sous la forme [0,925 ; 2,555].

13 1. a) La fonction f_θ est positive, continue sur $\mathbb{R} \setminus \{\theta\}$.

On calcule $\int_{-\infty}^{+\infty} f_\theta(t) dt = \int_0^\theta \frac{2t}{\theta^2} dt = \frac{1}{\theta^2} [t^2]_0^\theta = 1$.

La fonction f_θ est donc une densité de probabilité.

b) La densité f_θ est nulle en dehors d'un intervalle fermé borné sur lequel elle est bornée. La variable aléatoire X_θ admet donc une espérance, donnée par $E(X_\theta) = \frac{2}{\theta^2} \int_0^\theta t^2 dt = \frac{2}{3}\theta$.

D'autre part, $E(X_\theta^2) = \frac{2}{\theta^2} \int_0^\theta t^3 dt = \frac{1}{2}\theta^2$, d'où $V(X_\theta) = \frac{1}{2}\theta^2 - \frac{4}{9}\theta^2 = \frac{1}{18}\theta^2$.

2. a) On sait que la moyenne empirique $M_n = \frac{1}{n} \sum_{k=1}^n X_k$ est un estimateur sans biais de $E(X_\theta) = \frac{2}{3}\theta$. Compte tenu de la linéarité de l'espérance, on en déduit que $T_n = \frac{3}{2}M_n$ est un estimateur sans biais de θ .

D'autre part, $V(T_n) = \frac{9}{4}V(M_n) = \frac{\theta^2}{8n}$. Donc $\lim_{n \rightarrow +\infty} V(T_n) = 0$. Il en résulte que T_n est un estimateur sans biais et convergent de θ .

b) La fonction de vraisemblance est une fonction de θ qui dépend de n paramètres. Elle est nulle si l'un des x_k est extérieur à l'intervalle $[0, \theta]$, et vaut

$$\varphi(\theta) = L(x_1, x_2, \dots, x_n, \theta) = \left(\frac{2}{\theta}\right)^n \prod_{k=1}^n x_k$$

lorsque $0 \leq \min(x_1, x_2, \dots, x_n) \leq \max(x_1, x_2, \dots, x_n) \leq \theta$.

$\varphi(\theta)$ est donc nulle pour $\theta < \max(x_1, x_2, \dots, x_n)$, et vaut $\left(\frac{2}{\theta}\right)^n \prod_{k=1}^n x_k$ pour $\theta \in [\max(x_1, x_2, \dots, x_n), +\infty[$. Elle est donc décroissante sur cet intervalle, et admet un maximum pour $\theta = \max(x_1, x_2, \dots, x_n)$.

c) L'estimateur du maximum de vraisemblance est donc $M_n = \max(X_1, X_2, \dots, X_n)$.

3. a) La fonction de répartition de M_n est donnée par

$$F_{M_n}(x) = P(M_n \leq x) = P\left[\bigcap_{k=1}^n (X_k \leq x)\right] = \prod_{k=1}^n (P(X_k) \leq x) = [(F_{X_\theta}(x))^n].$$

C'est-à-dire $F_{M_n}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \left[\frac{x^2}{\theta^2}\right]^n & \text{si } 0 \leq x \leq \theta \\ 1 & \text{si } \theta \leq x \end{cases}$ Par dérivation, on obtient une

densité f_{M_n} de M_n sous la forme $f_{M_n}(x) = \begin{cases} 2n \frac{x^{2n-1}}{\theta^{2n}} & \text{si } 0 \leq x \leq \theta \\ 0 & \text{sinon} \end{cases}$

b) Le calcul de l'espérance de M_n donne $E(M_n) = \frac{2n}{\theta^{2n}} \int_0^\theta t^{2n} dt = \frac{2n}{2n+1} \theta$.

c) M_n est donc un estimateur biaisé de θ .

d) Comme estimateur sans biais de θ , on peut, bien entendu prendre $\widehat{M}_n = \frac{2n+1}{2n} M_n$, dont l'espérance est bien égale à θ , par linéarité de l'espérance.

e) On sait que $V(\widehat{M}_n) = \left(\frac{2n+1}{2n}\right)^2 V(M_n)$. Pour calculer $V(M_n)$, calculons d'abord

$$E(M_n^2) = \frac{2n}{\theta^{2n}} \int_0^\theta t^{2n+1} dt = \frac{n}{n+1} \theta^2. \text{ On en déduit que}$$

$$V(M_n) = \frac{n}{n+1} \theta^2 - \frac{4n^2}{(2n+1)^2} \theta^2 = \frac{n(2n+1)^2 - 4n^2(n+1)}{(n+1)(2n+1)^2} \theta^2 = \frac{n}{(n+1)(2n+1)^2} \theta^2$$

$$\text{et que } V(\widehat{M}_n) = \frac{n(2n+1)^2}{4n^2(n+1)(2n+1)^2} \theta^2 = \frac{\theta^2}{4n(n+1)}.$$

Comme $\lim_{n \rightarrow \infty} V(\widehat{M}_n) = 0$, et comme \widehat{M}_n est sans biais, on déduit qu'il est convergent.

Comme d'autre part, $V(T_n) = \frac{\theta^2}{8n}$, on constate que $V(\widehat{M}_n) = 0$ tend plus rapidement

vers 0 que $V(T_n)$. Pour préciser ce point, on calcule $\frac{V(\widehat{M}_n)}{V(T_n)} = \frac{2}{n+1}$. Comme ce

quotient tend vers 0 lorsque n tend vers l'infini, on conclut que la variance de \widehat{M}_n tend vers 0 infiniment plus rapidement que celle de T_n , ce qui prouve que \widehat{M}_n est bien meilleur que T_n .

4. Pour utiliser la loi de M_n , cherchons l'intervalle de confiance sous la forme $[aM_n; bM_n]$. On cherche donc deux réels a et b tels que $P(\theta \in [aM_n; bM_n]) = 0,95$, c'est-à-dire

$$P\left(\frac{\theta}{b} \leq M_n \leq \frac{\theta}{a}\right) = F_{M_n}\left(\frac{\theta}{a}\right) - F_{M_n}\left(\frac{\theta}{b}\right) = 0,95.$$

On ne dispose que d'une condition pour déterminer deux paramètres. Dans ce genre de situation, il est d'usage de procéder de la manière suivante (voir p. 452) :

Posons $0,95 = 1 - \alpha$ (avec $\alpha = 0,05$), et soient deux réels positifs α_1 et α_2

tels que $\alpha_1 + \alpha_2 = \alpha$, $F_{M_n}\left(\frac{\theta}{b}\right) = \alpha_1$ et $F_{M_n}\left(\frac{\theta}{a}\right) = 1 - \alpha_2$. La condition

$F_{M_n}\left(\frac{\theta}{a}\right) - F_{M_n}\left(\frac{\theta}{b}\right) = 1 - \alpha$ est bien remplie, et l'on peut disposer d'une condition

supplémentaire portant sur α_1 et α_2 . En l'absence d'indication supplémentaire, on cherchera un intervalle « symétrique » en posant $\alpha_1 = \alpha_2$ (ici, égaux à 0,025). Avec les

données de l'énoncé, en remplaçant F_{M_n} par sa valeur, et en simplifiant, on cherche a et b tels que $\frac{1}{b^{2n}} = 0,025$, et $\frac{1}{a^{2n}} = 0,975$.

On obtient alors l'intervalle de confiance cherché sous la forme

$$\left[5 \left(\frac{1}{0,975} \right)^{\frac{1}{2n}} ; 5 \left(\frac{1}{0,025} \right)^{\frac{1}{2n}} \right].$$

► **Remarque**

Il manque une application numérique. Pour $n = 20$, on trouve un intervalle de confiance au niveau 95% sous la forme $[5; 5,484]$, et avec $n = 100$, on trouve $[5; 5,188]$.

14 La représentation graphique s'effectue sans problème, et les calculs ne présentent pas de difficulté. Le point moyen est le point $(\bar{X}, \bar{Y}) = (7; 5)$.

Les droites de régression sont les droites d'équations respectives $y = \frac{7}{11}x + \frac{6}{11}$ et $x = -0,5 + 1,5y$, c'est-à-dire $y = \frac{2}{3}x + \frac{1}{3}$.

15 Cet énoncé, comme le suivant, relève davantage des travaux dirigés que des exercices. Ils sont surtout axés sur l'utilisation raisonnée d'un tableur. C'est par erreur que l'énoncé cite le nom d'un tableur. Il ne saurait être question, ici, de privilégier telle ou telle marque, aussi ne donnerons nous pas de corrigé détaillé. Simplement quelques indications. Pour simuler une variable aléatoire, la plupart des tableurs disposent d'un générateur de nombres au hasard, qui renvoie un nombre compris entre 0 et 1. Cette simulation fournit une bonne approximation d'une loi uniforme à densité sur $[0, 1]$. Pour ce premier exercice, il suffit de simuler X et Y dans les deux premières colonnes de la feuille de calcul, de manière indépendante (et sur 50, 100 ou 1000 lignes suivant le cas). On pourra être conduit, dans ce dernier cas à utiliser plusieurs colonnes pour chaque variable aléatoire. Si l'on ne veut pas utiliser la variable uniforme sur $[0, 1]$, on pourra utiliser, pour X , ou pour Y , ou les deux, des fonctions de cette variable aléatoire uniforme.

On choisira ensuite des cellules où l'on programmera les calculs demandés, en recopiant seulement les formules, selon la syntaxe du tableur que l'on utilise.

Enfin, on utilisera l'outil graphique pour représenter automatiquement le nuage de points, et éventuellement le point moyen.

La fonction de recalcul permet de faire fonctionner cette feuille de calculs plusieurs fois. On remarquera la grande variabilité des résultats obtenus, surtout pour les valeurs de n relativement faibles. On se convaincra en particulier que ce n'est pas parce que les variables X et Y sont indépendantes que les droites de régression du nuage de points obtenu sont systématiquement perpendiculaires.

On pourra, dans une même classe faire imprimer par chaque élève le nuage de points correspondant à son travail, et tracer les droites de régression, pour comparer ensuite les résultats.

16 1. Les calculs demandés ne présentent pas de difficulté. On les exécutera avantageusement dans un tableur. Les résultats sont des valeurs approchées, données avec trois décimales.

On trouve $Y = -1,404X + 9,679$ et $X = -0,710Y + 6,886$, ce qui donne $Y = -1,408X + 9,696$ et le coefficient de corrélation vaut $-0,998$.

2. La relation $PV^\gamma = C$ se traduit par $\gamma \ln V + \ln P = \ln C$, soit $\gamma X + Y = \ln C$. On peut donner des estimations assez précises, par $\ln C \approx 9,689$, soit $C \approx 16139$, et $\gamma \approx 1,406$.
3. Pour $V = 100$, on a alors $\ln P \approx -1,406 \ln 100 + 9,68$, c'est-à-dire $\ln P \approx 3,205$, et alors $P \approx 24,66$ (en kg/cm^3).

Pour $P = 15$, on a lors $\ln V = \frac{\ln C - \ln P}{\gamma}$, c'est-à-dire $\ln V \approx \frac{9,689 - 2,708}{1,406} \approx 143,3$.

4. Nous sommes ici dans un cas très particulier. On sait qu'il existe une relation affine entre X et Y . Cela explique la très forte corrélation observée. Le calcul demandé ne vise qu'à donner une estimation des valeurs numériques. Les estimations obtenues sont sans doute de très bonne qualité, malgré le faible nombre d'observations. Mais ces valeurs ne sont crédibles que si P et V prennent des valeurs comprises entre les valeurs extrêmes figurant dans le tableau.

17 1. Dans cet exercice, nous sommes confrontés à un système de 11 équations à 3

$$\text{inconnues, défini matriciellement par } M = \begin{pmatrix} 1 & -5 & 25 \\ 1 & -4 & 16 \\ 1 & -3 & 9 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{pmatrix} \begin{pmatrix} c \\ b \\ a \end{pmatrix} - \begin{pmatrix} 23,2 \\ 31,4 \\ 39,8 \\ 50,2 \\ 62,9 \\ 76 \\ 92 \\ 105,7 \\ 122,8 \\ 131,7 \\ 151,1 \end{pmatrix} = 0$$

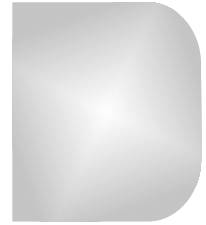
Faute de pouvoir annuler le vecteur M , on cherche à minimiser sa norme (p. 467-468). Pour cela, on multiplie par la matrice transposée de la matrice du système, et l'on obtient (tous calculs faits) un système à trois équations, trois inconnues :

$$\begin{cases} 11c + 110a & = & 886,8 \\ 110b & = & 1929,8 \\ 110c + 1958a & = & 9209 \end{cases}$$

dont on trouve pour solution sans difficulté $(a, b, c) \approx (0,4; 13; 76,64)$. La parabole des moindres carrés est donc (aux erreurs près d'arrondi) la parabole d'équation $y = 0,40x^2 + 13,00x + 76,64$.

2. Les représentations graphiques ne présentent pas de difficulté. Il est quand même nécessaire de bien choisir les unités, et de les effectuer avec le plus grand soin.
3. En calculant les valeurs de γ pour les 11 valeurs des données, on constate que l'approximation est d'assez bonne qualité. Et pour $x = 4,5$, on trouve $\gamma = 143,24$.
4. La détermination de la droite de régression de Y en X conduit à $\gamma = 13x + 80,62$ et à une estimation de 139,12 pour la $x = 4,5$.

Index



A

aiguille de Buffon, 296
anamorphose, 326
application
 bilinéaire, 24
 définie, 24
 positive, 24
 symétrique, 24
approximation
 affine de f , 129
 asymptotiquement sans biais, 333
 normale, 311, 312
 poissonnienne, 310

B

biais d'un estimateur, 329
boule, 93
 fermée, 93
 ouverte, 93

C

changement de variable, 69
Chasles, 66
coalitions, 225
col, 155
comparaison série-intégrale, 75
composition, 113
continue en M_0 , 108

continue sur D , 108
convergence
 absolue, 77
 en distribution, 298
 en loi, 298
 en probabilité, 286
 étroite, 298
 forte, 285
 presque sûre, 285
 stochastique, 286
coordonnée, 222
correction de continuité, 312
corrélation, 219
couple discret, 202
covariance, 217
critère de comparaison, 71

D

densité, 238
dérivée(s)
 d'une fonction composée, 132
 directionnelles en A , 127
 partielle, 125
 partielles d'ordre 2, 135
 seconde directionnelle en A , 143
développement limité
 d'ordre 1, 128, 131
 d'ordre 2, 141

distance euclidienne, 92
 domination et convergence, 72
 droite

affine, 88
 de Mayer, 356
 de régression, 221, 351

E

écart-type, 216
 échantillon, 318
 aléatoire, 319
 observé, 319
 échantillonnage aléatoire simple, 327
 égalité de Taylor-Lagrange à l'ordre 1,
 140

éléments extrémaux, 381

endomorphismes

antisymétriques, 55, 56
 diagonalisables, 10
 orthogonaux, 54
 symétriques, 43

équivalence et convergence, 72

erreur quadratique moyenne, 331

espérance

conditionnelle, 208
 totale, 191, 209

estimateur, 327

consistant, 333
 convergent, 333
 sans biais, 330

estimation

d'un intervalle, 413
 paramétrique, 318, 328

estimer, 318

exponentiation rapide, 369

extremum

absolu, 148
 global, 157
 global sous la contrainte \mathcal{C} , 162
 local, 148
 local sous la contrainte \mathcal{C} , 162

F

factorielle, 366

famille d'événements, 191

fermé de \mathbb{R}^n , 96

Fibonacci, 372

fonction(s)

affine, 104
 beta, 278
 convexe, 147
 de classe \mathcal{C}^1 , 130
 de classe \mathcal{C}^2 , 135
 de répartition, 174
 de répartition empirique, 320
 Γ , 73
 génératrice, 194
 partielle, 125
 positives (intégrale des), 71
 quantile, 338
 vraisemblance, 359

formatage, 402

forme quadratique, 47

formule

de Kœnig-Huygens, 216, 251
 des accroissements finis, 134

G

gestion de listes, 380

gradient, 127

graphe, 104

H

hessienne, 139

hyperplan, 89

affine, 89
 tangent, 130

I

iid, 318

image réciproque par f

d'un intervalle fermé, 116
 d'un intervalle ouvert, 116

impropreté, 59

inégalité

de Bienaymé-Tchebychev, 254, 292
 de Cauchy-Schwarz, 27
 de Markov, 254, 291
 triangulaire, 28

Index

intégrale(s)

- convergente, 59
- de Bertrand, 81
- de Gauss, 74
- de Riemann, 64
- divergente, 59

intégration par parties, 68

intervalle

- de confiance, 337, 417
- de dispersion, 338

K

Khi-deux, 264

L

lemme de Slutsky, 302

lignes de niveau, 106

linéarité, 64

loi

- beta de première espèce, 281
- binomiale négative, 195
- conditionnelle, 207
- conjointe, 204
- d'Erlang, 267
- de Cauchy, 271
- de Laplace, 268
- de Laplace-Gauss, 273
- de Pareto, 281
- de Pascal, 195
- de Rayleigh, 281
- de Student, 345
- de Weibull, 281
- du couple, 204
- exponentielle, 258
- faible des grands nombres, 295
- forte des grands nombres, 295
- Gamma, 263
- hypergéométrique, 309
- Log-normale, 280
- marginale, 205
- parente, 318
- sans mémoire, 260
- uniforme, 255

M

matrice(s)

- des variances covariances, 219
- diagonalisable, 17
- orthogonale, 37
- semblables, 11

maximum

- absolu, 148
- de vraisemblance, 359
- local (ou relatif), 148

minimum

- absolu, 148
- local (ou relatif), 148

moindres carrés, 41, 351

moment, 246

moyenne empirique, 320, 321

N

niveau de confiance, 337

norme euclidienne, 26

notations de Monge, 140, 155

nuage de points, 349

O

opérations sur les fonctions continues,
111

orthogonal, 34

orthonormalisation de Schmidt, 31

ouvert de \mathbb{R}^n , 94

P

paramètre d'échelle, 263

partie

- bornée, 93
- convexe, 98
- fermée et bornée, 117

point moyen, 349

point selle, 155

point(s) critique(s), 149

- de f dans l'optimisation sous
contrainte \mathcal{C} , 162
- sous contrainte, 162

polynôme(s)

- annulateur, 9, 16
- d'un endomorphisme, 8

d'une matrice, 15
 de Tchebychev, 57
 position du graphe par rapport à l'hyper-
 plan tangent, 159
 positivité, 66
 processus binomial, 195
 produit de convolution, 245
 produit scalaire, 24
 projecteur, 3
 projection, 2
 orthogonale, 40
 pseudo-inverse, 357

Q

Quicksort, 386

R

RANDOM, 323, 390
 RANDOMIZE, 323, 390
 récursive, 365
 récursivité, 364
 réduction d'un endomorphisme nil-
 potent, 21
 résolution dichotomique, 367
 risque, 337
 quadratique, 331

S

segment, 89
 semi-convergence, 78
 série
 absolument convergente, 182
 double, 179
 simulation(s)
 de lois réelles, 390
 par convergence en loi, 327
 par fonctions indicatrices, 325
 par inversion de répartition, 326
 sommation par paquets, 180
 somme directe, 1
 de deux sous-espaces, 1
 de n sous-espaces, 4
 sous-espace(s)

propre, 7
 stables, 5
 supplémentaires, 1
 spectre, 7, 12
 statistique, 319
 empirique, 320
 pivotale, 344
 suite double, 179
 supplémentaire orthogonal, 39
 support, 338
 système complet, 191
 d'événements, 203

T

théorème
 central limite, 309
 de Bernoulli, 296
 de la limite centrée, 308
 de Lindeberg-Lévy, 309
 de Pythagore, 29
 de Schwarz, 137
 de transfert, 186, 213, 242, 249
 tri, 384

V

valeur propre, 7, 12
 variable(s)
 aléatoire réelle, 171
 aléatoire réelle à densité, 238
 centrée réduite, 255
 indépendantes, 210
 indépendantes deux à deux, 225
 mutuellement indépendantes, 172
 variance, 216
 empirique, 322
 vecteur(s)
 aléatoire, 221
 espérance, 206
 multinomial, 232
 orthogonaux, 29
 propre, 7, 12
 vraisemblance, 359

Sous la direction de Christian Gautier et André Warusfel
Bruno Caminade • Gonzague de Monicault • Serge Nicolas

MATHÉMATIQUES

TOUT-EN-UN • ECS 2^e ANNÉE

Cours et exercices corrigés

Cet ouvrage couvre en un seul volume la totalité des programmes de mathématiques de la 2^e année des classes préparatoires économiques et commerciales. Il concerne principalement la filière scientifique, mais pourra être également utilisé avec profit par les élèves de la filière économique, ainsi que par les élèves des classes préparatoires BCPST et B/L.

Dans cette 2^e édition revue et corrigée, tous les corrigés des exercices sont présents en fin d'ouvrage.

Conçu comme ouvrage de référence, ce livre propose à son lecteur une vision globale du cours dans le strict respect des programmes. Il se compose de 13 chapitres s'articulant autour de quatre grands thèmes :

- Compléments d'algèbre linéaire, algèbre bilinéaire
- Analyse
- Probabilités, Statistiques
- Informatique

De nombreux exercices et problèmes complètent le cours, ce qui permet au futur candidat de s'entraîner efficacement dans l'optique des concours.

CHRISTIAN GAUTIER
ancien élève de l'École normale supérieure de Saint-Cloud, est professeur au lycée La Bruyère à Versailles.

ANDRÉ WARUSFEL
ancien élève de l'École normale supérieure de la rue d'Ulm, a été professeur de Mathématiques Spéciales au lycée Louis-le-Grand à Paris et Inspecteur général de mathématiques.

MATHÉMATIQUES

PHYSIQUE

CHIMIE

SCIENCES DE L'INGÉNIEUR

INFORMATIQUE