

Notions de mathématiques

Sommaire

| | | |
|------------|---|-----------|
| 2.1 | Les permutations | 15 |
| 2.1.1 | Représentation des permutations | 15 |
| 2.1.2 | Statistiques sur les permutations | 16 |
| 2.1.3 | Relations entre les statistiques | 16 |
| 2.1.4 | Résultats provenant de la combinatoire | 17 |
| 2.2 | Théorie des probabilités | 19 |
| 2.2.1 | Espace de probabilité | 19 |
| 2.2.2 | Variable aléatoire | 21 |
| 2.2.3 | Probabilité conditionnelle et dépendance | 21 |
| 2.2.4 | Espérance et Variance | 22 |
| 2.2.5 | Formules usuelles | 22 |
| | Identité sur les mesures de probabilités | 23 |
| | Identité sur les variables aléatoires | 23 |
| 2.2.6 | Chaîne de Markov discrète | 24 |
| | Propriétés essentielles des chaînes de Markov | 25 |
| | Distribution stationnaire | 27 |
| | Le théorème ergodique | 28 |

Dans ce chapitre, nous introduisons les notions mathématiques qui nous seront nécessaires au cours de cette thèse. La première partie introduit le groupe de symétrie sur un ensemble de taille n dont les éléments sont appelés des permutations. Nous verrons en particulier des définitions de statistiques sur ces permutations ainsi que de bijections qui préservent ces dernières. La deuxième partie concerne la théorie des probabilités. Nous nous intéresserons en particulier aux chaînes de Markov dont nous aurons besoin dans le cadre de l'analyse dans le cas moyen dans le contexte de la prédiction de branchements.

2.1 Les permutations

Soit un ensemble fini X à n éléments, nous pouvons attribuer un numéro à chaque élément de cet ensemble à l'aide d'une bijection $e : X \rightarrow [n]$. Nous nous intéressons ici au groupe de symétrie des ensembles finis, c'est à dire l'ensemble des bijections de ces ensembles sur eux mêmes. Comme il existe une bijection e pour ces ensembles finis, il n'est pas nécessaire de travailler sur l'ensemble X , nous pouvons donc nous intéresser uniquement au groupe de symétrie sur l'ensemble $[n]$. Nous notons \mathfrak{S}_n le groupe de symétrie de l'ensemble $[n]$. Les éléments qui appartiennent à \mathfrak{S}_n sont appelés des permutations de taille n .

2.1.1 Représentation des permutations

Il existe de nombreuses façons de représenter une permutation, et nous allons en montrer deux.

La première est l'utilisation typique issue de l'algèbre sur les permutations. Comme les permutations appartiennent à un groupe fini, il est possible de représenter ces dernières comme un ensemble de cycles. Un *cycle* dans la permutation $\sigma \in \mathfrak{S}_n$ de représentant $i \in [n]$ est défini par la séquence $(i, \sigma(i), \sigma(\sigma(i)), \dots, \sigma^{-1}(i))$. Nous pouvons donc ainsi représenter une permutation comme un ensemble de tous ses cycles. Par exemple si $\sigma = \{(1, 2), (3)\}$, alors nous avons la permutation $\sigma(1) = 2$, $\sigma(2) = 1$ et $\sigma(3) = 3$. Pour ce cas, nous avons un cycle de taille 1 ce qui nous donne un *point fixe*.

La deuxième représentation provient d'un contexte que nous verrons plus souvent dans ce manuscrit. Il s'agit de représenter la permutation sous la forme d'un mot. Nous nous en servons particulièrement pour le cas du problème du tri, où cette représentation est utilisée pour encoder l'ordre dans la séquence d'entrée. Supposons que nous ayons une séquence ordonnée que nous voulons trier $S = (s_1, \dots, s_n)$. Il existe alors une permutation $\sigma \in \mathfrak{S}_n$ telle que $S' = (s_{\sigma^{-1}(1)}, \dots, s_{\sigma^{-1}(n)})$ est la séquence triée des éléments de S . Pour tout i le rang de s_i est ainsi donné par $r_i = \sigma(i)$. La séquence (r_1, \dots, r_n) est alors une autre façon de définir σ . Nous verrons par la suite qu'il existe des algorithmes de tri qui n'utilisent que la comparaison entre les éléments pour trier une telle séquence S . Pour ces algorithmes, il n'y a alors aucune différence, en terme de coût, à trier S ou à trier la séquence (r_1, \dots, r_n) . Nous écrivons de cette façon $\sigma = (r_1, \dots, r_n)$ sa représentation sous forme de mot et qui définit un ordre sur la séquence S . Si nous reprenons, notre exemple précédent, nous avons que la permutation $\sigma = (2, 1, 3) = \{(1, 2)(3)\}$. Par la suite, nous omettrons souvent les parenthèses et les virgules, lorsqu'il n'y a aucune ambiguïté, nous aurons ainsi sur notre exemple $\sigma = 213 = (21)(3)$.

2.1.2 Statistiques sur les permutations

Dans ce qui suit, nous allons introduire des statistiques et des notations sur les permutations qui auront un intérêt particulier au chapitre 6 sur la génération de permutations non-uniformes.

Cycles : Nous l'avons vu précédemment, nous pouvons écrire une permutation sous la forme d'une décomposition en cycles. Il se trouve que le nombre de cycles est une statistique d'intérêt et est même à la base de la distribution d'Ewens qui sera étudiée au chapitre 6. Nous dénotons par $\mathbf{cyc}(\sigma)$ le nombre de cycles d'une permutation $\sigma \in \mathfrak{S}_n$. Nous dénotons de plus par $\mathbf{cyc}_{\geq k}(\sigma)$ le nombre de cycles dont la taille est supérieure ou égale à k . Finalement nous utilisons $\mathbf{cyc}_k(\sigma)$ pour le nombre de cycles dont la taille est égale à k . En particulier $\mathbf{cyc}_1(\sigma)$ compte le nombre de points fixes de σ .

Records : Les records sont de deux types possibles. Un min-record est un élément i d'une permutation σ tel que dans sa représentation en mot $\sigma(i)$ est plus petit que tout élément qui le précède, soit $\forall j \in [i-1]$ on a $\sigma(i) < \sigma(j)$. Respectivement, un max-record est défini de la même façon pour la relation $>$. Sur tout \mathfrak{S}_n , ces statistiques avec le nombre de cycles ont les mêmes cardinalités et nous verrons par la suite des bijections pour le prouver. Dans la permutation $\sigma = 46527381$ nous comptons 4 max-records et 3 min-records.

Descentes et montées : Une permutation σ admet une descente en position i si l'on a la relation $\sigma(i) > \sigma(i+1)$. Respectivement, une montée en position i signifie que l'on a la relation $\sigma(i) < \sigma(i+1)$. Par exemple, dans la permutation $\sigma = 46527381$ nous comptons 3 montées et 4 descentes. Le nombre de montées et le nombre de descentes est une statistique bien connue en combinatoire sur les permutations.

2.1.3 Relations entre les statistiques

Nous l'avons évoqué précédemment, il existe des relations entre certaines de ces statistiques, et nous allons le démontrer à l'aide de bijections.

La bijection fondamentale : La première bijection que nous allons voir est la bijection fondamentale (voir [13, p. 109-110] pour plus de détails et références). Cette bijection permet de montrer qu'à partir d'une permutation de taille n et de k cycles, il est possible de construire une autre permutation de taille n et de k min-records. La bijection fondamentale procède en écrivant la permutation dans sa décomposition en cycles. Pour chacun de ces cycles, nous faisons en sorte de prendre la représentation avec le plus grand élément en première position. Enfin, nous trions les cycles dans l'ordre croissant par les valeurs prises par leurs premiers éléments, puis nous retirons les parenthèses pour obtenir la forme en mot de la permutation image.

Exemple. Prenons la permutation $\sigma = 926857341$. Sa représentation en cycle est $\sigma = (367)(19)(5)(48)(2)$. Nous pouvons la réarranger en faisant en sorte que le plus grand élément de chaque cycle apparaît en premier ce qui donne $\sigma = (736)(91)(5)(84)(2)$. On tri ensuite dans l'ordre croissant de ces premiers

éléments et on obtient $\sigma = (2)(5)(736)(84)(91)$. En retirant les parenthèses nous obtenons l'image de σ par la bijection fondamentale soit $F(\sigma) = 257368491$. On voit bien ici, que chaque élément de début de cycle de σ est alors un record dans $F(\sigma)$ et qu'il y a donc conservation entre ces deux statistiques. De plus, il est facile de voir à l'aide de cette remarque comment revenir en arrière puisqu'il suffit de chercher les min-records de l'image pour retrouver la décomposition en cycles de la permutation initiale et donc de voir que cette application est bien une bijection.

Remarque. Il est facile d'obtenir une bijection fondamentale pour les max-records en mettant le plus grand élément de chaque cycle en première position et en les triant dans l'ordre décroissant.

Bijections entre montées et descentes : Il existe deux involutions qui permettent à partir d'une permutation de transformer toutes les montées en descentes et réciproquement. La première est de représenter une permutation σ sous forme de mot et de prendre le mot miroir $\tilde{\sigma}$ qui pour chaque $i \in [n]$ associe $\tilde{\sigma}(i) = \sigma(n + 1 - i)$. Dans ce cas, une montée en position $i \in [n]$ dans $\sigma \in \mathfrak{S}_n$ devient une descente en position $n - i$ dans $\tilde{\sigma}$. Une autre involution, notons la η , qui possède la même propriété d'inversion des descentes et des montées procède en remplaçant chaque élément $\sigma(i)$ de $\sigma \in \mathfrak{S}_n$, avec $i \in [n]$, par $n + 1 - \sigma(i)$. Cette application laisse invariant la position des montées devenues des descentes et réciproquement.

Exemple. Prenons à nouveau la permutation $\sigma = 926857341$. Nous avons alors $\tilde{\sigma} = 143758629$ et $\eta(\sigma) = 184253769$.

2.1.4 Résultats provenant de la combinatoire

Nombres de Stirling de la première espèce non-signés : Les nombres de Stirling apparaissent dans de nombreux problèmes de combinatoire. En particulier, les nombres de Stirling de la première espèce non-signés comptent les permutations de \mathfrak{S}_n composées de k cycles. Ces nombres sont notés $c(n, k)$. Il n'est alors pas étonnant de les voir apparaître dans les expressions des probabilités faisant intervenir des permutations distribuées en fonction de leur nombre de cycles, comme c'est le cas de la distribution d'Ewens dont nous discuterons au cours du chapitre 6.

Remarque. Comme nous avons pu le voir précédemment, les records sont en bijection avec les cycles, et ainsi $c(n, k)$ comptent également les permutations de taille n avec k records.

Par convention, nous considérons que $c(0, 0) = 1$. Par définition, comme il est impossible d'avoir des permutations de tailles non-nulles et composées de moins de 0 cycle, nous avons pour tout $n > 0$ et tout $k \leq 0$ que $c(n, k) = 0$. De la même façon, il est impossible d'avoir des permutations composées d'un plus grand nombre de cycles que son nombre d'éléments, autrement dit pour tout $n \geq 0$ et pour tout $k > n$, nous avons $c(n, k) = 0$.

Nous allons maintenant présenter des propriétés remarquables de ces nombres.

Lemme 1. Pour tout $n > 0$, pour tout $k > 0$, nous avons la relation de récurrence

$$c(n + 1, k) = nc(n, k) + c(n, k - 1)$$

qui est vérifiée.

Démonstration. La preuve se fait par construction. Pour obtenir une permutation de taille $n + 1$ qui est composée de k cycles, nous pouvons partir d'une permutation de taille n et insérer l'élément $n + 1$ afin d'obtenir une nouvelle permutation de taille $n + 1$. Pour avoir une permutation de taille $n + 1$ composée de k cycle après l'insertion de l'élément $n + 1$, il y a deux cas possibles :

- Soit la permutation initiale possède déjà k cycles, dans ce cas, il suffit de rajouter l'élément $n + 1$ dans un des cycles déjà existants. Il y a n positions possibles dans ce cas. Comme nous avons $c(n, k)$ permutations qui ont exactement k cycles, il y a en tout $nc(n, k)$ possibilités d'obtenir une permutation de taille $n + 1$ à k cycles de cette façon.
- Soit la permutation initiale possède déjà $k - 1$ cycles, dans ce cas il est nécessaire de créer un nouveau cycle constitué uniquement de $n + 1$. Comme il y a en tout $c(n, k - 1)$ permutations qui ont exactement $k - 1$ cycles, il y a en tout $c(n, k - 1)$ possibilités d'obtenir une permutation de taille $n + 1$ à k cycles de cette façon.

En combinant les contributions de ces deux cas, nous obtenons le résultat annoncé. \square

Lemme 2. Pour tout $x \in \mathbb{C}$, l'identité

$$\sum_{k=0}^n c(n, k)x^k = \prod_{k=0}^{n-1} x + k$$

est vérifiée.

Démonstration. Il est possible de démontrer cette relation à l'aide de la méthode symbolique et de l'analyse combinatoire. La preuve peut, par exemple, être trouvée dans le livre de Philippe Flajolet [25]. Nous ne prouverons ici ce résultat qu'avec une preuve par récurrence.

Remarquons que pour $n = 0$ nous avons

$$\prod_{k=0}^{n-1} x + k = 1,$$

car il s'agit d'un produit vide. La convention $c(0, 0) = 1$ a été choisie pour rendre cette relation vraie pour $n = 0$. Supposons que la relation est vraie pour un certain rang n . Développons alors l'expression pour $n + 1$,

$$\sum_{k=0}^{n+1} c(n + 1, k)x^k.$$

En appliquant le Lemme 1, nous avons que cette somme est égale à

$$\sum_{k=0}^{n+1} nc(n, k)x^k + \sum_{k=0}^{n+1} c(n, k - 1)x^k.$$

Après changement de la variable de la somme de droite par $k' = k - 1$ et

comme $c(n, n + 1) = c(n, -1) = 0$, nous avons que cette quantité est égale à

$$\sum_{k=0}^n nc(n, k)x^k + \sum_{k'=0}^n c(n, k')x^{k'+1}.$$

Nous pouvons faire une factorisation et ainsi obtenir que cette quantité est égale à

$$(n + x) \sum_{k=0}^n c(n, k)x^k.$$

En appliquant l'hypothèse de récurrence sur l'expression $\sum_{k=0}^n c(n, k)x^k$, nous obtenons que

$$\sum_{k=0}^{n+1} c(n + 1, k)x^k = (n + x) \prod_{k=0}^{n-1} x + k,$$

L'hypothèse de récurrence pour n implique que la relation est vérifiée pour $n + 1$ et est donc vraie pour tout $n \in \mathbb{N}$. \square

2.2 Théorie des probabilités

Dans le cadre de cette thèse, nous utilisons des éléments de la théorie des probabilités. Nous allons donc dans cette section faire quelques rappels des axiomes de cette théorie ainsi que de propriétés simples qui en découlent.

2.2.1 Espace de probabilité

Nous devons, pour commencer, définir ce qu'est un espace probabilisé, ou espace de probabilité. Soit Ω un *univers* qui est l'ensemble des états qui peuvent être observés au cours d'une expérience aléatoire. Par exemple, dans le cas d'un lancer de dé à six faces, nous pouvons avoir $\Omega = \{1, 2, 3, 4, 5, 6\}$. Pour définir ce qu'est un espace probabilisé, nous devons premièrement introduire la notion de *tribu*.

Définition 1. Soit \mathcal{A} un ensemble de sous-ensembles d'un univers Ω . L'ensemble \mathcal{A} est une tribu si elle vérifie les trois propriétés suivantes :

1. L'ensemble Ω appartient à \mathcal{A} .
2. L'ensemble \mathcal{A} est stable par complémentaire. Si $E \in \mathcal{A}$ alors $E^c = \Omega \setminus E \in \mathcal{A}$.
3. L'ensemble est stable à l'intersection et à l'union dénombrable. Soit une famille $(A_i)_{i \in \mathbb{N}}$ telle que pour tout $i \in \mathbb{N}$ $A_i \in \mathcal{A}$. Nous avons alors

$$\bigcap_{i \in \mathbb{N}} A_i \in \mathcal{A},$$

et

$$\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}.$$

Intuitivement, une mesure de probabilité peut être vue comme la fréquence d'apparition d'un événement. Si nous réalisons un grand nombre N de fois une expérience aléatoire, nous nous attendons à ce que la probabilité d'un événement soit proche du nombre de fois que ce dit événement s'est produit divisé par N . Une conséquence de cette intuition, est qu'une probabilité prend une valeur comprise dans l'intervalle $[0, 1]$.

Définition 2. Une *mesure de probabilité* est une application $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$. De plus, cette application doit vérifier les deux propriétés suivantes :

1. $\mathbb{P}(\Omega) = 1$.
2. Soit une famille $(A_i)_{i \in \mathbb{N}}$ telle que pour tout $i \in \mathbb{N}$ $A_i \in \mathcal{A}$ disjoints deux-à-deux, c'est-à-dire $A_i \cap A_j = \emptyset$ si $i \neq j$. Nous avons

$$\mathbb{P} \left(\bigcup_{i \in \mathbb{N}} A_i \right) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i). \quad (2.1)$$

Nous définissons un *espace probabilisé* par un tel triplet $(\Omega, \mathcal{A}, \mathbb{P})$. L'équation (2.1) est appelée la propriété de σ -additivité.

Définition 3. Un espace de probabilité est défini par un triplet $(\Omega, \mathcal{A}, \mathbb{P})$, où Ω est un univers, \mathcal{A} une tribu dans cet univers, et \mathbb{P} une mesure de probabilité qui associe à chaque événement de \mathcal{A} une valeur dans l'intervalle $[0, 1]$.

Remarque. Il est naturel de choisir $\mathcal{A} = 2^\Omega$, cependant il existe des Ω et des \mathbb{P} , pour lesquels \mathbb{P} ne peut pas être définie sur tous les ensembles de 2^Ω sans créer de contradictions. Dans le cas où Ω est un ensemble fini ou dénombrable, il est possible de créer un espace probabilisé $(\Omega, 2^\Omega, \mathbb{P})$ à la condition que, pour tout $\omega \in \Omega$, il existe $p(\omega) \in [0, 1]$ tel que $\sum_{\omega \in \Omega} p(\omega) = 1$ et en définissant pour tout $A \in 2^\Omega$, $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega)$. Dans de nombreux cas, nous aurons à faire à des espaces probabilisés de ce type.

Des exemples d'espaces probabilisés

Distribution uniforme sur un ensemble fini : En utilisant ce qui a été dit lors de la remarque précédente, nous pouvons en déduire un espace trivial pour le cas où Ω est un ensemble fini. L'ensemble $\mathcal{A} = 2^\Omega$ et nous choisissons d'attribuer le même poids à chaque élément de Ω . Nous définissons ainsi, pour tout $A \in \mathcal{A}$ la mesure de probabilité uniforme $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$.

Distribution uniforme sur l'intervalle $[0, 1]$: Un espace que nous verrons souvent par la suite, est la distribution uniforme pour le cas où $\Omega = [0, 1]$. Nous n'avons pas $\mathcal{A} = 2^\Omega$. \mathcal{A} est construit en générant la tribu minimale à partir des intervalles ouverts $]a, b[$, avec $0 \leq a \leq b \leq 1$. Il s'agit donc de la tribu la plus petite possible qui contient tous ces intervalles. La mesure de probabilité est la *mesure de Lebesgue* v qui pour un intervalle $]a, b[$ associe $v(]a, b[) = b - a$. En particulier, si $a = 0$, $v(]a, b[) = b$.

Nous verrons d'autres distributions par la suite, notamment au cours du Chapitre 6.

2.2.2 Variable aléatoire

Il est courant de transformer l'espace probabilisé à l'aide de variables aléatoires. Concrètement, une variable aléatoire est une application $X : \Omega \rightarrow E$ dans un espace E . Bien souvent, une variable représente une grandeur et nous choisissons $E = \mathbb{R}$. Nous pouvons transformer l'espace probabilisé pour la variable aléatoire en posant une nouvelle mesure de probabilité que nous définissons par

$$\mathbb{P}_X(x) = \mathbb{P}(X^{-1}(x)),$$

où pour tout $x \in E$, $X^{-1}(x) = \{\omega \in \Omega \mid X(\omega) = x\}$. Cette nouvelle mesure est appelée la loi de la variable X . Une autre notation courante est d'écrire $\mathbb{P}_X(x) = \mathbb{P}(X = x)$. Nous pouvons ainsi définir un nouvel espace probabilisé pour l'univers E et la mesure \mathbb{P}_X .

Un exemple de variable aléatoire peut être la somme de deux dés à six faces. Choisissons la distribution uniforme en prenant l'espace probabilisé où $\Omega = \{1, \dots, 6\}^2$, $\mathcal{A} = 2^\Omega$ et pour $\omega = (d_1, d_2) \in \Omega$, nous avons $\mathbb{P}(\omega) = \frac{1}{36}$. Nous posons la variable aléatoire $X(d_1, d_2) = d_1 + d_2$. Nous avons alors, par exemple, $\mathbb{P}_X(2) = \frac{1}{36}$ puisque $(1, 1) \in \Omega$ est le seul lancer qui permet d'obtenir une somme égale à 2, et nous avons aussi $\mathbb{P}_X(3) = \frac{2}{36}$ puisque $(1, 2)$ et $(2, 1)$ sont les seuls lancers possibles pour obtenir une somme égale à 3.

2.2.3 Probabilité conditionnelle et dépendance

Nous allons maintenant définir des notions de dépendances entre les événements. Soit deux événements $A, B \in \mathcal{A}$, nous voulons connaître les chances que l'événement A se réalise si l'on sait que l'événement B s'est réalisé.

Définition 4. Soit $A, B \in \mathcal{A}$ deux événements dans un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. La probabilité conditionnelle que A se réalise sachant que B s'est réalisé est définie par la mesure

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Cette relation définit une nouvelle mesure de probabilité dans cette espace.

Pour mieux comprendre cette définition, nous allons réutiliser l'intuition fréquentiste, qui considère qu'une probabilité est une fréquence limite obtenue après la répétition d'un grand nombre N d'expériences aléatoires. Nous voulons déterminer le nombre de fois que l'événement B a été obtenu sur l'ensemble de ces expériences, si nous notons N_B ce nombre, nous avons alors

$$N_B \sim N\mathbb{P}(B).$$

De la même façon, nous pouvons définir le nombre de fois que l'événement $A \cap B$ s'est réalisé et nous avons

$$N_{A \cap B} \sim N\mathbb{P}(A \cap B).$$

Nous voulons connaître la fréquence d'apparition de l'événement A pour toutes les fois où l'événement B a eu lieu, cette fréquence est alors la fraction

du nombre de fois où nous avons eu à la fois A et B divisé par le nombre de fois où nous avons eu B , ce qui donne

$$\frac{N_{A \cap B}}{N_B} \sim \frac{N\mathbb{P}(A \cap B)}{N\mathbb{P}(B)} \sim \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

ce qui semble correspondre avec la définition de probabilité conditionnelle. Ce raisonnement n'est cependant pas assez rigoureux et il faut bien définir ce que nous voulons dire par convergence d'une probabilité en une fréquence, il s'agit donc plus d'une intuition pour mieux comprendre cette définition.

Il est possible que la réalisation de B ne donne aucune information quant à la réalisation de A . Dans ce cas nous déclarons ces événements comme indépendants.

Définition 5. Soit $A, B \in \mathcal{A}$, A et B sont indépendants si et seulement si nous avons

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Notons que si A et B sont indépendants, nous avons alors $P(A|B) = P(A)$.

2.2.4 Espérance et Variance

En statistique, nous voulons souvent déterminer la valeur moyenne d'un échantillon de données et savoir à quel point ces données sont proches de cette valeur moyenne en calculant la variance. En probabilité, la notion d'espérance et de variance est relativement similaire. Si nous avons une variable aléatoire X , nous aimerions déterminer la valeur moyenne après un échantillon de N observations de X . La définition dans le cas où l'ensemble Ω est dénombrable rappelle l'expression de la moyenne en fonction des fréquences.

Définition 6. Soit un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ avec Ω dénombrable. Soit $X : \Omega \rightarrow \mathbb{R}$, une variable aléatoire. L'espérance de X est définie par la relation

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} \mathbb{P}(\omega)X(\omega) \tag{2.2}$$

La variance représente l'écart entre les observations de X et de $E[X]$, elle est définie comme suit.

Définition 7. Soit un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ avec Ω dénombrable. Soit $X : \Omega \rightarrow \mathbb{R}$, une variable aléatoire. La variance de X est définie par la relation

$$Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Pour des espaces probabilisés plus complexes, il est possible de définir rigoureusement l'espérance à l'aide d'éléments de la théorie de la mesure en topologie. Ces définitions dépassent, cependant, l'utilisation que nous aurons de la théorie des probabilités au cours de cette thèse.

2.2.5 Formules usuelles

Nous allons dans cette section donner quelques formules usuelles en probabilité que nous utiliserons à de nombreuses occasions.

Identité sur les mesures de probabilités

Monotonie de la mesure de probabilité : Soit $A \subset B$ deux événements de \mathcal{A} . Nous avons $B = A \cup (B \setminus A)$. Comme $A \cap (B \setminus A) = \emptyset$, l'union est disjointe et par additivité, nous avons $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$. Comme $\mathbb{P}(B \setminus A) \geq 0$, nous obtenons l'inégalité :

$$\mathbb{P}(A) \leq \mathbb{P}(B). \quad (2.3)$$

Principe d'inclusion-exclusion : Soit deux événements A et B dans \mathcal{A} . D'après la propriété de σ -additivité de l'équation (2.1), nous avons

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \end{aligned}$$

Le raisonnement peut être généralisé à un ensemble fini d'événements A_1, \dots, A_n de \mathcal{A} avec $n \in \mathbb{N}$. Nous obtenons alors dans ce cas l'équation :

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}). \quad (2.4)$$

Sous-additivité : Soit une famille $(A_i)_{i \in \mathbb{N}}$ d'éléments de \mathcal{A} . Il découle, de la propriété de σ -additivité et de la monotonie, l'identité

$$\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) \leq \sum_{i \in \mathbb{N}} \mathbb{P}(A_i). \quad (2.5)$$

Probabilités totales : Soit $(E_i)_{i \in \mathbb{N}}$ une partition dénombrable de Ω , avec pour tout $i \in \mathbb{N}$ on a $E_i \in \mathcal{A}$. Pour tout $A \in \mathcal{A}$, nous avons, en utilisant la définition de probabilité conditionnelle ainsi que la σ -additivité :

$$\mathbb{P}(A) = \sum_{i \in \mathbb{N}} \mathbb{P}(A|E_i)\mathbb{P}(E_i). \quad (2.6)$$

Identité sur les variables aléatoires

Autres écritures de l'espérance : Il est possible de ré-écrire l'équation (2.2) en agglomérant les $\omega \in \Omega$ tels que $X(\omega) = x$ avec $x \in \mathbb{R}$. Nous avons alors

$$\mathbb{E}[X] = \sum_{x \in X^{-1}(\Omega)} x \mathbb{P}(X = x). \quad (2.7)$$

Dans le cas où $X^{-1}(\Omega) = \mathbb{N}$, il est également possible de ré-écrire cette dernière équation par

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{P}(X \geq i). \quad (2.8)$$

Pour se convaincre de cette formule, il suffit de voir que les $\omega \in \Omega$ tels que $X(\omega) = i$ appartiennent à l'ensemble $X \geq j$ pour tout les $j \in \{1, \dots, i\}$. De cette façon, chacun de ces ω sont comptés exactement i fois.

Autre écriture d'une loi associée à une variable X : Soit une variable aléatoire $X : \Omega \rightarrow \mathbb{N}$, nous avons la relation pour $k \in \mathbb{N}$

$$\mathbb{P}(X = k) = \mathbb{P}(X \geq k) - \mathbb{P}(X \geq k + 1). \quad (2.9)$$

Inégalité de Markov : Soit $X : \Omega \rightarrow \mathbb{R}^+$ une variable aléatoire réelle positive, nous avons pour une valeur réelle $\alpha > 0$,

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[X]}{\alpha}. \quad (2.10)$$

Dans le cadre de cette thèse, nous pourrions être amenés à utiliser toutes ces formules. Nous allons maintenant nous intéresser aux processus stochastiques des chaînes de Markov dont nous allons avoir besoin pour des problèmes de prédictions de branchement, que nous étudierons en détail au cours du Chapitre 3 Section 3.5 et du Chapitre 5.

2.2.6 Chaîne de Markov discrète

Nous allons donc nous intéresser aux processus stochastiques des chaînes de Markov. Cette section est un survol de la théorie sur les chaînes de Markov discrètes dont les détails peuvent être retrouvés dans de nombreux livres sur la théorie des probabilités et de la stochastique. Nous avons pris une partie des résultats donnés dans le chapitre 8 du livre de Rosenthal [38], les preuves sont omises mais peuvent être retrouvées dans ce livre. Nous allons dès à présent noter par S un ensemble fini ou dénombrable d'états. C'est sur cet ensemble d'états que nous faisons une marche aléatoire. Nous notons pour tout $t \in \mathbb{N}$, la variable aléatoire $X(t)$ qui prend une valeur sur l'ensemble S . Cette variable X va décrire un mouvement sur cette ensemble S dans le temps. Pour tout couple d'états $(x, y) \in S \times S$, nous connaissons la probabilité de transition p_{xy} . Cette probabilité de transition correspond à la probabilité lors de la marche d'aller de l'état x vers l'état y . Plus concrètement, si à tout instant $t \in \mathbb{N}$, nous avons $X(t) = x$, alors nous avons

$$\mathbb{P}(X(t+1) = y | X(t) = x) = p_{xy}.$$

Bien évidemment, nous devons avoir pour tout $x \in S$,

$$\sum_{y \in S} p_{xy} = 1.$$

Nous devons enfin considérer une distribution initiale qui, pour tout $x \in S$, associe la probabilité de commencer par cet état $\nu_x = \mathbb{P}(X(0) = x)$.

Une chaîne de Markov est la séquence des valeurs prises par $(x_t)_{t \in \mathbb{N}}$, telles que $X(t) = x_t$ pour tout $t \in \mathbb{N}$. Pour les N premières étapes de la marche, nous avons alors la probabilité

$$\mathbb{P}(X(0) = x_0, \dots, X(N) = x_N) = \nu_{x_0} p_{x_0 x_1} \dots p_{x_{N-1} x_N} \quad (2.11)$$

qui est vérifiée.

Nous pouvons écrire tous les paramètres de la chaîne de Markov sous forme matricielle. Nous posons ainsi le vecteur ν , dont les composantes sont les dis-

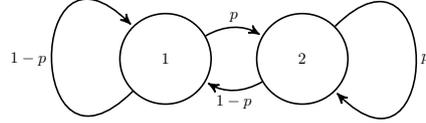


FIGURE 2.1 – Représentation sous forme d’automate d’une chaîne de Markov avec $S = \{1, 2\}$.

tributions initiales ν_x pour tout $x \in S$. Nous notons également la matrice de transition $\mathcal{M} = [p_{xy}]_{(x,y) \in S \times S}$. À chaque instant $t \in \mathbb{N}$, nous avons la distribution d’être à chaque état qui est donnée par les composantes du vecteur $\pi(t)$. Nous avons bien évidemment $\pi(0) = \nu$, et pour tout $n \in \mathbb{N}$, nous avons

$$\pi(n) = (\mathcal{M}^T)^n \nu.$$

Nous allons terminer cette section avec un exemple que nous verrons plus en détail au cours du chapitre 3. Nous allons voir qu’il est également possible de représenter cette chaîne comme un automate (voir Figure 2.1).

Dans cet exemple, l’ensemble des états est $\{1, 2\}$, et nous avons pour un réel $0 < p < 1$ les probabilités de transitions suivantes :

$$p_{11} = 1 - p \qquad p_{12} = p$$

$$p_{21} = 1 - p \qquad p_{22} = p$$

Nous pouvons compresser ces probabilités en une matrice de transition :

$$\begin{bmatrix} 1 - p & p \\ 1 - p & p \end{bmatrix}.$$

Nous continuerons à utiliser cet exemple pour illustrer les propos qui vont suivre.

Propriétés essentielles des chaînes de Markov

Nous allons maintenant définir des propriétés intéressantes que certaines chaînes de Markov peuvent posséder. Nous verrons à la section suivante que ces propriétés sont essentielles pour démontrer la convergence vers une distribution stationnaire unique.

À partir de maintenant, pour tout $x \in S$ et pour tout événement A , nous notons $\mathbb{P}_x(A) = \mathbb{P}(A|X(0) = x)$. Nous définissons les deux probabilités suivantes :

Pour tout états $i, j \in S$,

$$f_{ij}^{(n)} = \mathbb{P}_i(X(n) = j, \forall m \in [n - 1] X(m) \neq j)$$

et

$$f_{ij} = \sum_{n=0}^{\infty} f_{ij}^{(n)},$$

qui sont respectivement la probabilité qu'en commençant par l'état i , nous atteignons pour la première fois j au temps n et la probabilité que l'état j soit atteignable en commençant par l'état i .

Définition 8. Un état $x \in S$ est *récurrent* si et seulement si $f_{xx} = 1$, autrement dit, nous sommes certains de revenir à l'état x en commençant la chaîne par ce dernier. Si un état n'est pas récurrent, on dit alors qu'il est *transient*.

Définition 9. Une chaîne de Markov est *irréductible*, si et seulement si, pour tout $i, j \in S$, nous avons $f_{ij} > 0$. S'il existe $i, j \in S$ tels que $f_{ij} = 0$, alors la chaîne de Markov est *réductible*.

Une chaîne de Markov qui est irréductible est donc une chaîne où il est possible d'atteindre n'importe quel état en commençant par n'importe quel autre état. Il suffit donc de montrer qu'il existe un chemin entre ces deux états qui a une probabilité non-nulle. Si nous notons pour $x, y \in S$ la probabilité $p_{xy}^{(n)} = \mathbb{P}_x(X(n) = y)$ nous avons le théorème suivant qui est vérifié.

Théorème 1. Soit une chaîne de Markov qui est définie sur un ensemble d'états S et par les probabilités de transitions $\{p_{ij}\}_{i,j \in S}$. Si cette chaîne est irréductible, alors les propriétés suivantes sont équivalentes :

1. Il existe un état dans S qui est récurrent.
2. Pour tout $i, j \in S$, nous avons $f_{ij} = 1$ et ainsi en particulier, tous les états sont récurrents.
3. Il existe $k, l \in S$ tels que $\sum_{n=1}^{\infty} p_{kl}^{(n)} = \infty$.
4. Pour tout $i, j \in S$, nous avons $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$.

Une conséquence de ce théorème est que, sur une chaîne de Markov irréductible, les états sont, soit tous récurrents, soit tous transients.

Définition 10. Pour tout état $x \in S$, la *période* de cet état est le PGCD de l'ensemble des n tels que $p_{xx}^{(n)} > 0$. Si tous les états des S ont une période de 1, alors on dit que la chaîne de Markov est *apériodique*.

Cette définition est sensiblement similaire à la définition de périodicité sur les automates. Nous avons de plus le théorème suivant :

Théorème 2. Si une chaîne de Markov est irréductible, alors tous ses états ont la même période.

Autrement dit, si nous parvenons à démontrer qu'il existe un état de période 1 dans une chaîne de Markov irréductible, nous montrons aussi qu'elle est apériodique. En particulier, cela est vraie si nous avons un état $x \in S$ tel que $p_{xx} > 0$. Ce qui est le cas de notre exemple où nous avons $S = \{1, 2\}$.

Distribution stationnaire

Soit une distribution sur les états qui est définie par un vecteur $\pi^* = (\pi_x^*)_{x \in S}$.

Définition 11. La distribution π^* est *stationnaire* si pour toutes ses composantes, nous avons

$$\sum_{x \in S} \pi_x^* P_{xy} = \pi_y^*.$$

Nous pouvons écrire cette relation sous forme matricielle, et nous avons alors

$$\mathcal{M}^T \pi^* = \pi^*. \quad (2.12)$$

Le terme stationnaire vient du fait que si à un temps $t \in \mathbb{N}$ nous sommes sous cette distribution, alors pour tout $k \in \mathbb{N}$, nous restons dans cette distribution au temps $t + k$. Si une chaîne de Markov présente les bonnes propriétés, il est alors possible qu'elle ait une unique distribution stationnaire vers laquelle elle converge quelle que soit la distribution initiale. Autrement dit, nous avons un rang T tel que pour tout $t > T$, nous avons $\pi(t) \approx \pi^*$. Nous avons un théorème que nous pouvons utiliser pour démontrer qu'une chaîne de Markov converge vers une telle distribution.

Théorème 3. *Pour toutes chaînes de Markov sur un ensemble fini d'états. Si la chaîne est irréductible et apériodique, alors il existe une unique distribution stationnaire π^* vers laquelle elle converge indépendamment de la distribution initiale. Nous avons donc pour tout $x \in S$, $\lim_{n \rightarrow \infty} \mathbb{P}(X(n) = x) = \pi_x^*$.*

Démonstration. Il s'agit de la combinaison du corollaire 8.3.11 et de la proposition 8.4.10 que l'on peut trouver dans [38][p.93 et p.98]. \square

Nous allons maintenant montrer une méthode pour déterminer la distribution stationnaire dans le cas où le Théorème 3 est vérifié. Remarquons dans ce cas que par définition, l'Equation (2.12) est vérifiée. Il se trouve que cette équation, signifie que la distribution stationnaire est un vecteur propre de \mathcal{M}^T associé à la valeur propre de 1. Nous savons également, comme il s'agit d'une distribution que la somme de ses composantes est égale à 1. Il nous suffit donc de déterminer ce vecteur propre, en déterminant le noyau de la matrice $\mathcal{M}^T - Id$ où Id est la matrice identité de même format que \mathcal{M}^T . Il nous suffit donc de résoudre le système linéaire

$$(\mathcal{M}^T - Id) \times \pi^* = 0$$

Appliquons ces nouvelles connaissances sur notre exemple. Comme pour tout x dans $S = \{1, 2\}$, nous avons $p_{xx} = p_{xx}^{(1)} > 0$, il est évident d'après tous les théorèmes que nous venons de voir, que la chaîne est irréductible et apériodique, et comme S est fini, nous convergions vers une unique distribution stationnaire. La matrice de transitions et sa transposée sont données par

$$\mathcal{M} = \begin{bmatrix} 1-p & p \\ 1-p & p \end{bmatrix}$$

$$\mathcal{M}^T = \begin{bmatrix} 1-p & 1-p \\ p & p \end{bmatrix}$$

Si nous essayons de résoudre le système linéaire

$$(\mathcal{M}^T - Id)x = \begin{bmatrix} -p & 1-p \\ p & p-1 \end{bmatrix} \times x = 0$$

Nous trouvons alors que le noyau est l'ensemble des vecteurs de la forme suivante

$$\mathcal{N}(\mathcal{M}^T - Id) = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| x_1 = \frac{1-p}{p}x_2 \right\}.$$

Si nous cherchons dans ce noyau le seul vecteur qui a la somme de ses composantes à 1, nous avons alors le vecteur de la distribution stationnaire

$$\pi^* = \begin{bmatrix} 1-p \\ p \end{bmatrix}.$$

Le théorème ergodique

Dans le cadre de l'analyse d'algorithmes dans le cas moyen, nous sommes bien souvent intéressés par l'asymptotique de l'espérance de certaines variables aléatoires. Le théorème ergodique, nous permet de cette façon d'obtenir de tels équivalents asymptotiques. Pour nos besoins, nous allons utiliser un corollaire du théorème ergodique original. Nous donnons directement ce corollaire, que nous appellerons par la suite le théorème ergodique bien qu'il s'agit d'un léger abus.

Théorème 4. *Soit (\mathcal{M}, ν) une chaîne de Markov irréductible et apériodique qui est définie sur un ensemble S . Notons π^* sa distribution stationnaire. Soit E un ensemble de couples $(x, y) \in S^2$ de probabilités de transitions non-nulles $\mathcal{M}(x, y) > 0$. Pour tout entier strictement positif n , soit L_n une variable aléatoire strictement positive telle que $\lim_{n \rightarrow \infty} \mathbb{E}[L_n] = +\infty$. Soit X_n une variable aléatoire qui compte le nombre de fois qu'une transition dans E est prise sur une marche aléatoire de longueur L_n dans \mathcal{M} . Indépendamment de la distribution initiale ν , nous avons l'équivalence asymptotique*

$$\mathbb{E}[X_n] \sim \mathbb{E}[L_n] \sum_{(x,y) \in E} \pi(x)\mathcal{M}(x, y).$$

qui est vérifiée.

Démonstration. D'après le théorème ergodique original [32, p.58], si F est un sous-ensemble de S , et si Y_n est une variable aléatoire qui compte le nombre de fois qu'un élément de F est atteint au cours d'une marche aléatoire de taille n sur \mathcal{M} , alors, quand n est grand, nous avons l'équivalence asymptotique

$$\frac{1}{n} \mathbb{E}[Y_n] \sim \sum_{x \in F} \pi(x).$$

Soit B l'ensemble de tous les couples $(x, y) \in S^2$ tels que $\mathcal{M}(x, y) > 0$. Nous considérons la chaîne de Markov de deuxième ordre \mathcal{M}_2 dont l'ensemble des états est B , et ses transitions sont uniquement définies pour tous les couples $(x, y), (y, z) \in B$ avec $\mathcal{M}_2((x, y), (y, z)) = \mathcal{M}(y, z)$.

\mathcal{M}_2 est une chaîne de Markov irréductible. Pour tous les couples $(x, y), (w, z) \in B$, nous avons alors $f_{(x,y),(w,z)} \geq \mathcal{M}(x, y)f_{yw}\mathcal{M}(y, w)$, et comme \mathcal{M} est irréductible et par définition $\mathcal{M}(x, y) > 0$ et $\mathcal{M}(w, z)$, nous avons ainsi $f_{(x,y),(w,z)} > 0$.

\mathcal{M}_2 est apériodique. Pour tout état $(x, y) \in B$, comme \mathcal{M} est irréductible, il existe une marche de longueur $n - 1$ avec $n > 1$ commençant par l'état y et se terminant par l'état x de probabilité non nulle, ainsi

$$\mathcal{M}^{(n-1)}(y, x) > 0.$$

De plus nous avons que la probabilité d'un cycle de longueur n dans \mathcal{M}_2 commençant par l'état (x, y) vérifie l'expression

$$\mathcal{M}_2^{(n)}((x, y), (x, y)) = \mathcal{M}^{(n-1)}(y, x)\mathcal{M}(x, y),$$

ce qui est non-nulle pour le même n . Comme \mathcal{M} est irréductible, nous avons $f_{xx} > 0$ et donc il existe un ensemble $\Delta = \{\delta_1, \delta_2, \dots\}$ tel que pour tout $\delta \in \Delta$, nous avons

$$\mathcal{M}^{(\delta)}(x, x) > 0.$$

Il existe donc des cycles de probabilités non-nulles dans \mathcal{M}_2 commençant par l'état (x, y) de tailles $n + \delta$ pour tout $\delta \in \Delta \cup \{0\}$. Si nous notons par ℓ_1, ℓ_2, \dots les longueurs d'autres cycles à probabilités non-nulles qui ne s'expriment pas comme une somme $n + \delta$, alors nous avons que la période de l'état (x, y) dans \mathcal{M}_2 est définie par l'expression

$$\text{pgcd}(n, n + \delta_1, n + \delta_2, \dots, \ell_1, \ell_2, \dots).$$

Si nous prouvons que $\text{pgcd}(n, n + \delta_1, n + \delta_2, \dots)$ alors nous démontrons que la période de (x, y) vaut 1 et par irréductibilité de \mathcal{M}_2 , tous les états sont de périodicité 1 et donc \mathcal{M}_2 est irréductible. Une première étape consiste à ré-écrire ce PGCD sous la forme suivante :

$$\text{pgcd}(n, n + \delta_1, n + \delta_2, \dots) = \text{pgcd}(\text{pgcd}(n, n + \delta_1), \text{pgcd}(n, n + \delta_2), \dots).$$

De cette façon nous pouvons utiliser la relation provenant de l'algorithme d'Euclide, qui dit que $\text{pgcd}(a, a + b) = \text{pgcd}(a, b)$. Nous avons ainsi,

$$\text{pgcd}(\text{pgcd}(n, n + \delta_1), \text{pgcd}(n, n + \delta_2), \dots) = \text{pgcd}(\text{pgcd}(n, \delta_1), \text{pgcd}(n, \delta_2), \dots).$$

Nous pouvons ré-écrire cette dernière égalité par

$$\text{pgcd}(\text{pgcd}(n, \delta_1), \text{pgcd}(n, \delta_2), \dots) = \text{pgcd}(n, \delta_1, \delta_2, \dots).$$

Comme \mathcal{M} est apériodique, le PGCD des éléments de Δ définissent la période de l'état x dans \mathcal{M} qui vaut donc 1. Nous avons alors

$$\begin{aligned} \text{pgcd}(n, n + \delta_1, n + \delta_2, \dots) &= \text{pgcd}(n, \text{pgcd}(\delta_1, \delta_2, \dots)) \\ &= \text{pgcd}(n, 1) \\ &= 1, \end{aligned}$$

ce qui nous permet de conclure sur que \mathcal{M}_2 est apériodique.

Par définition, nous avons directement que \mathcal{M}_2 converge vers l'unique dis-

tribution stationnaire π_2^* définie pour tout couple $(x, y) \in B$ par

$$\pi_2^*(x, y) = \pi^*(x)\mathcal{M}(x, y).$$

Nous pouvons maintenant appliquer le théorème ergodique original sur la variable aléatoire Z_ℓ qui compte le nombre de fois qu'un couple dans E est parcouru durant une marche aléatoire de longueur ℓ dans \mathcal{M}_2 . Nous avons ainsi

$$\mathbb{E}[Z_\ell] = \ell \left(\sum_{(x,y) \in E} \pi^*(x)\mathcal{M}(x, y) \right) (1 + \epsilon_\ell),$$

avec $\epsilon_\ell \rightarrow 0$ quand ℓ tend vers l'infini. Intéressons nous à présent au cas où la longueur de la marche aléatoire est la variable aléatoire L_n . Nous avons alors

$$\begin{aligned} \mathbb{E}[X_n] &= \sum_{\ell \geq 0} \mathbb{P}(L_n = \ell) \mathbb{E}[X_n | L_n = \ell] \\ &= \sum_{\ell \geq 0} \mathbb{P}(L_n = \ell) \mathbb{E}[Z_\ell] \\ &= \sum_{\ell \geq 0} \mathbb{P}(L_n = \ell) \ell \left(\sum_{(x,y) \in E} \pi^*(x)\mathcal{M}(x, y) \right) (1 + \epsilon_\ell) \\ &= (\mathbb{E}[L_n] + \mathbb{E}[L_n \epsilon_{L_n}]) \left(\sum_{(x,y) \in E} \pi^*(x)\mathcal{M}(x, y) \right). \end{aligned}$$

Pour terminer la preuve, il nous reste à démontrer que $\mathbb{E}[L_n \epsilon_{L_n}]$ est en $o(\mathbb{E}[L_n])$. Par définition, comme ϵ_ℓ tend vers 0 quand ℓ est grand, pour tout $\alpha > 0$, il existe un rang ℓ_0 tel que pour tout $\ell > \ell_0$, nous avons

$$\epsilon_\ell \leq \frac{\alpha}{2}.$$

Posons $\eta_0 = \max\{\epsilon_\ell : \ell \in \{0, \dots, \ell_0\}\}$, nous avons alors

$$\begin{aligned} \mathbb{E}[L_n \epsilon_{L_n}] &= \sum_{\ell \geq 0} \mathbb{P}(L_n = \ell) \ell \epsilon_\ell \\ &= \sum_{\ell=0}^{\ell_0} \mathbb{P}(L_n = \ell) \ell \epsilon_\ell + \sum_{\ell > \ell_0} \mathbb{P}(L_n = \ell) \ell \epsilon_\ell \\ &< \sum_{\ell=0}^{\ell_0} \mathbb{P}(L_n = \ell) \ell \epsilon_\ell + \frac{\alpha}{2} \mathbb{E}[L_n] \\ &< \eta_0 \ell_0 + \frac{\alpha}{2} \mathbb{E}[L_n]. \end{aligned}$$

Comme $\mathbb{E}[L_n]$ tend vers l'infini quand n est suffisamment grand, il existe un rang n_0 tel que pour tout $n > n_0$ nous avons

$$\frac{\eta_0 \ell_0}{\mathbb{E}[L_n]} \leq \frac{\alpha}{2}.$$

Ainsi pour tout $\alpha > 0$, il existe un rang n_0 tel que pour tout $n > n_0$, nous avons $\mathbb{E}[L_n \epsilon_{L_n}] \leq \alpha \mathbb{E}[L_n]$, ce qui par définition signifie que $\mathbb{E}[L_n \epsilon_{L_n}] = o(\mathbb{E}[L_n])$ ce qui nous permet de conclure cette preuve. \square