L'ANALYSE EN COMPOSANTES PRINCIPALES (ACP) " CLASSIQUE "

1.1. Domaine d'application

L'ACP permet d'analyser tout tableau de données statistiques $X(n,p)$ (n lignes, p colonnes) représentant n individus décrits par p variables quantitatives. Son domaine d'application est donc très vaste. Ainsi si l'ensembe des individus doit être homogène (ensemble d'entreprises ou ensemble de personnes par exemple), l'ensemble des variables peut être hétérogène (chiffre d'affaire, nombre d'employés pour une entreprise ou taille, poids d'un individu par exemple).

1.2. Cadre de l'ACP

1.2.a) Nuages de points associés au tableau des données

Soit $X = \{x_i^j, i = 1, ..., n; j = 1, ..., p\}$ le tableau des données, l'individu i est décrit par le vecteur de $\mathbb{R}^p$ $\;X_i = (x_i^j, j = 1, ..., p)$. De plus, chaque individu i est muni d'un poids p_i : $\forall i = 1, ..., n \;\; p_i > 0$; $\displaystyle\sum_{i=1}^{n} p_i = 1$ (en général, on a $p_i = \frac{1}{n}$, pour $i = 1, ..., n$).

La variable j est décrite par le vecteur de $\mathbb{R}^n$: $X^j = (x_i^j, i = 1, ..., n)$.

au tableau des données X sont donc associés deux nuages de points

- Le nuage de points pesants $N(I) = \{(X_i, p_i), i = 1, ..., n\} \subset \mathbb{R}^p$ dit nuage des individus.
- Le nuage $N(J) = \{X^j, j = 1, ..., p\} \subset \mathbb{R}^n$ dit nuage des variables.

Avant de préciser le but de l'ACP, nous présentons maintenant les notions qu'elle utilise et quelques résultats préliminaires.

1.2. b) Centre de gravité du nuage $N(I)$
C'est un vecteur de $\mathbb{R}^p$ que l'on notera g qui s'écrit :

$$g = \left(g^j = \sum_{i=1}^{n} p_i \; x_i^j, j = 1, ..., p \right)$$

Dans la suite, on supposera pour simplifier les calculs que $g = 0$, on peut toujours se ramener à ce cas en centrant les p variables. Matriciellement, l'égalité $g = 0$ s'écrit : ${}^tXD_p\mathbb{I} = 0$. où ${}^tX$ est la matrice transposée de X.
et avec $\mathbb{I} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$, $Dp = \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix}$ matrice des poids des individus.
Afin de définir des distances entre individus et des distances entre variables, on munit les espace $\mathbb{R}^p$ et $\mathbb{R}^n$ de métriques euclidiennes (c'est à dire associées à des matrices symétriques définies positives).

1.2.c) Métrique D_p sur l'espace des variables $\mathbb{R}^n$
La métrique dont on munit $\mathbb{R}^n$ est la métrique Dp dite métrique des poids. Le choix de cette métrique est naturel. En effet :

$$\langle X^j, X^{j'} \rangle_{D_p} = \; {}^tX^j D_p X^{j'} = \sum_{i=1}^{n} p_i \; x_i^j \; x_i^{j'} = COV(X^j, X^{j'})$$

car les variables sont centrées. De manière analogue :

$$||X^j||_{D_p}^2 = \; {}^tX^j \; D_p \; X^j = \sum_{i=1}^{n} p_i (x_i^j)^2 = Var(X^j)$$

1.2. d) Métrique M sur l'espace des individus $\mathbb{R}^p$
Le choix de la matrice M dépend des caractéristiques des données. Rappelons que les plus usuelles sont la matrice identité et la matrice diagonale D_{1/σ^2}, dont le terme général de la diagonale est l'inverse de la variance des variables. Nous reviendrons sur ce choix important au paragraphe 1) 11)

1.2. e) Matrice variance - covariance du nuage $N(I)$

Le terme général de la matrice variance $V_{(p,p)}$ du nuage $N(I)$ s'écrit

$$\forall j, j' = 1, ..., p \quad COV(X^j, X^{j'}) = \sum_{i=1}^{n} p_i \; x_i^j \; x_i^{j'} \; = \; {}^tX^j D_p X^{j'}$$

Matriciellement, on a donc : $V = \sum_{i=1}^{n} p_i \; X_i \; {}^tX_i$, d'où $V = \; {}^tXD_p \; X.$

1.3. Inerties

1.3.a) Inertie par rapport à un point

On rappelle que l'inertie du nuage $N(I) \subset \mathbb{R}^p$ muni de la métrique M, par rapport à un point a de $\mathbb{R}^p$ s'écrit :

$$I_a = \sum_{i=1}^{n} p_i \; d_M^2(X_i, a) = \sum_{i=1}^{n} p_i \; {}^t(X_i - a) \; M(X_i - a)$$

L'inertie par rapport au centr de gravité O de $N(I)$ s'écrit :

$$I_O = \sum_{i=1}^{n} p_i \; ||X_i||_M^2 = \sum_{i=1}^{n} p_i \; {}^tX_i \; MX_i.$$

D'après le théorème de Huygens, l'inertie de $N(I)$ par rapport à son centre de gravité est minimum. Cette quantité est désignée comme l'inertie totale du nuage $N(I)$, on notera désormais $I_O = I_T$.

Théorème de Huygens [1]

Si $g = \sum_{i=1}^{n} p_i \; X_i$ est le centre de gravité du nuage $N(I)$ on a :

$$\forall a \in \mathbb{R}^p, \; I_a = I_g + d_M^2(a, g)$$

Démonstration

$\forall a \in \mathbb{R}^p, \; I_a = \sum_{i=1}^{n} p_i \; {}^t(X_i - a) \; M(X_i - a)$

$I_a = \sum_{i=1}^{n} p_i \; {}^t(X_i - g + g - a) \; M(X_i - g + g - a)$

$I_a = \sum_{i=1}^{n} p_i \; {}^t(X_i - g) \; M(X_i - g) + 2 \sum_{i=1}^{n} p_i \; {}^t(g - a) \; M(X_i - g)$

$+\sum_{i=1}^{n} p_i \; {}^t(g-a)\; M(g-a) \quad$ (car M est symétrique)

$I_a = I_g + d_M^2(g,a) + 2^t(g-a)\; M \sum_{i=1}^{n} p_i \; (X_i - g)$ (car $\sum_{i=1}^{n} p_i = 1$)

Or $\sum_{i=1}^{n} p_i \; (X_i - g) = 0$ par définition de g, d'où le résultat annoncé.

Remarques

- Le centre de gravité g est le point par rapport auquel l'inertie du nuage est minimum.
- I_g est l'inertie totale du nuage $N(\Omega)$ et sera souvent notée I ou T.

1.3.b) Inertie par rapport à un sous - espace affine

Soit E_1 un sous - espace vectoriel de $E = \mathbb{R}^p$. Considérons la décomposition en somme directe de E : $E = E_1 \oplus E_1^{\perp}$, $E_1^{\perp}$ étant le sous - espace vectoriel orthogonal à E_1 pour la métrique M ; $\forall\; X_i \in E$, on a $X_i = \alpha_i + \beta_i$, $\alpha_i \in E_1$, $\beta_i \in E_1^{\perp}$; cette décompositioni étant unique.

Soient F_1 et $F_1^{\perp}$ les sous - espaces affines associés à E_1 et $E_1^{\perp}$ passant par un point a de E, dont la décomposition suivant E_1 et $E_1^{\perp}$ s'écrit : $a = a_1 + a_2$. On appelle inertie du nuage $N(I)$ par rapport à F_1 la quantité :

$$I_{F_1} = \sum_{i=1}^{n} p_i \; d_M^2(\beta_i, a_2) = \sum_{i=1}^{n} p_i \; d_M^2(X_i, F_1)$$

De même on définit l'inertie $N(I)$ par rappport à $F_1^{\perp}$:

$$I_{F_1^{\perp}} = \sum_{i=1}^{n} p_i \; d_M^2(\alpha_i, a_1) = \sum_{i=1}^{n} p_i \; d_M^2(X_i, F_1^{\perp})$$

Par le théorème de Pythagore, on a :

$$d_M^2(X_i, a) = d_M^2(\alpha_i, a_1) + d_M^2(\beta_i, a_2).$$

Puisque $(\alpha_i - a_1) \in E_1$ est orthogonal à $(\beta_i - a_2) \in E_1^{\perp}$, on peut donc en déduire $I_a = I_{F_1} + I_{F_1^{\perp}}$. En particulier, si $a = 0$, on a $I_T = I_{E_1} + I_{E_1^{\perp}}$

Par le théorème de Huygens, on a :

$$\forall a \in \mathbb{R}^p,\; I_{F_1} = I_{E_1} + d_M^2(a_2, 0).$$

Autrement dit, le sous - espace affine associé à E_1 minimisant I_{F_1} est celui qui contient le centre de gravité de $N(I)$.

Remarque
Les relations précédentes permettent de voir que $I_{E_1^\perp}$ peut aussi s'interpréter comme l'inertie de la projection du nuage $N(I)$ sur E_1. On désignera $I_{E_1^\perp}$ comme l'inertie portée par E_1.

1.4 But de l'ACP

Le but de l'analyse en composantes principales est d'obtenir une réprésentation du nuage $N(I)$ de $\mathbb{R}^p$ dans un espace de dimension réduirte de telle manière que l'inertie portée par cet espace soit la plus grande possible.
La principale opération de l'ACP est de déterminer les axes principaux d'inertie du nuage autour de son centre de gravité. Ce sont les axes qui prennent le mieux en compte la dispersion du nuage au sens de la distance d_M définie sur $\mathbb{R}^p$. Ces axes principaux d'inertie appelés axes factoriels permettent de représenter les points du nuage sur des espaces de dimension réduite. Par exemple, on obtiendra une représentation plane du nuage en projetant orthogonalement au sens de la métrique M tous les points sur le plan principal d'inertie, c'est-à-dire sur l'espace de dimension 2 qui porte le plus d'inertie.

1.5. Formulation du problème de l'ACP

Mathématiquement, le problème s'énonce ainsi : Trouver le sous - espace affine E_k de dimension k $(k < p)$ tel que l'inertie du nuage $N(I)$ par rapport à E_k : $I_{E_k} = \sum_{i=1}^{n} p_i \; d_M^2(X_i, E_k)$ soit minimum.
E_k est l'espace tel que l'inertie $I_{E_k^\perp}$ du nuage projeté sur E_k soit maximum.
D'après le théorème de Huygens, E_k contient nécessairement le centre de gravité O du nuage.
Nous donnons maintenant deux théorèmes qui vont permettre de traiter le problème en plusieurs étapes.
Théorème 1 d'inclusion
Si E_{k-1} est un sous - espace vectoriel optimal de dimension $k-1$, alors la recherche d'un sous - espace vectoriel optimal de dimension k peut se faire parmi l'ensemble des sous - espaces vectoriels de dimension k contenant E_{k-1}.

Démonstration
Soit F_k un sous - espace vectoriel de dimension k et $H = F_k + E_{k-1}^\perp$.

$F_k \cap E_{k-1}^{\perp}$ ne peut être réduit au vecteur nul. Sinon on aurait

$$H = F_k \oplus E_{k-1}^{\perp} \text{ et } dim(H) = k + (p - (k-1)) = p + 1.$$

Ce qui est absurde puisque $H \subset \mathbb{R}^p$. Il existe donc $v \neq 0 \in F_k \cap E_{k-1}^{\perp}$. Soit Δv l'axe engendré par v.

Soit G l'espace supplémentaire M- orthogonal à Δv dans F_k : $F_k = G \oplus \Delta V$ et soit $E_k = E_{k-1} \oplus \Delta v$. On a $I_{F_k} = I_G + I_{\Delta v}$ car G est orthogonal à Δv, mais par hypothèse E_{k-1} est optimal, donc $I_{k-1} \leq I_G$ d'où $I_{E_k} \leq I_{F_k}$.
On peut donc restreindre la recherche d'un sous - espace optimal aux sous - espaces contenant E_{k-1}.
Théorème 2
La recherche d'un sous - espace vectoriel E de dimension k contenant un espace vectoriel F de dimension $k-1$ minimisant I_E est équivalente à la recherche d'un axe Δv , M - orthogonal à F et mnimisant $I_{\Delta v}$

Démonstration
Quel que soit l'espace E contenant F, on a une décomposition $E = F \oplus \Delta v$ avec $\Delta v \perp F$ donc $I_E = I_F + I_{\Delta v}$: I_F étant constant, minimiser I_E revient à minimiser $I_{\Delta v}$.
A partir de ces théorèmes, on ramène donc le problème de l'ACP au problème suivant :

1. Rechercher un axe $E_1 = \Delta u_1$ à inertie minimum, u_1 étant le vecteur unitaire engendrant E_1.

2. Rechercher un axe Δu_2, M - orthogonal à Δu_1 et à inertie minimum. Soit $E_2 = \Delta u_1 \oplus \Delta u_2$; E_2 est un sous - espace optimal de dimension 2.

3. Rechercher un axe Δu_k, M - orthogonal à E_{k-1} et à inertie minimum. Soit $E_k = E_{k-1} \oplus \Delta u_k$, E_k est alors une solution du problème.
   On a $E_k = \Delta u_1 \oplus \Delta u_2 \oplus ... \oplus \Delta u_k$. Les axes $\Delta u_1, \Delta u_2, ..., \Delta u_k$ sont appelés les axes factoriels.

Remarque
On a du même coup obtenu toutes les solutions pour $h < k$.

1.6. Résolution du problème

1.6. a) Résultats Préliminaires

• **Expression de l'Inertie totale I_T**

$$I_T = \sum_{i=1}^{n} p_i \ {}^tX_i \ M \ X_i.$$

Notons $tr(A)$ la trace d'une matrice A. Comme l'expression ${}^tX_i \ M \ X_i$ est un réel, on a : ${}^tX_i \ MX_i = tr({}^tX_i \ MX_i) = tr(X_i \ {}^tX_i \ M)$ car
$tr(AB) = tr(BA)$. d'où :

$$I_T = tr\left(\left(\sum_{i=1}^{n} p_i \ X_i \ {}^tX_i\right)M\right)$$

On a donc : $I_T = tr \ VM$

• **Expression de l'inertie portée par un axe**
Soit Δu l'axe engendré par le vecteur unitaire u : on a $\forall X_i \in E, \ X_i = \alpha_i + \beta_i$, avec $\alpha_i \in \Delta u$ et $\beta_i \in \Delta u^{\perp}$. Or α_i s'écrit $a_i = \alpha_i u, a_i \in \mathbb{R}$.
Donc $I_{\Delta u^{\perp}} = \displaystyle\sum_{i=1}^{n} p_i \ d^2_M(\alpha_i, 0)$.
α_i est la projection orthogonale de X_i sur Δu, donc $a_i = \langle u, X_i\rangle = \ {}^tu \ M \ X_i$ et $d^2(\alpha_i, 0) = ||a_i \ u||^2 = a_i^2 = \ {}^tu \ M \ X_i \ {}^tX_i \ Mu$ d'où

$$I_{\Delta u^{\perp}} = \ {}^tuM \left[\sum_{i=1}^{n} p_i \ X_i \ {}^tX_i\right] Mu$$

Soit $I_{\Delta u^{\perp}} = \ {}^tuMVMu$.

On en déduit que $I_{\Delta u} = \ tr(VM) - \ {}^tuMVMu$

• **Etude de** VM :
Rappelons tout d'abord que M est symétrique définie positive et que V est symétrique positive. Par ailleurs, VM est M symétrique : ${}^t(VM)M = M(VM)$.
On en déduit les propriétés suivantes :
- Les valeurs propres de VM sont réelles, positives ou nulles.
Il existe une base M - orthonormée de $E = \mathbb{R}^p$ constituée de vecteurs propres de VM.

1.6. b) Détermination des axes factoriels

On commence par chercher le premier axe factoriel Δu_1.
Le problème s'écrit maintenant :

Problème 1 :
Maximiser ${}^t uMVMu$ sous la contrainte ${}^t uMu = 1$.
Munissons E de la base M - orthonormée constituée des vecteurs propres $e_1, ..., e_p$ de VM associés aux valeurs propres $\lambda_1, ..., \lambda_p$ les valeurs propres étant rangées par ordre décroissant $(\lambda_1 > \lambda_2 > ... > \lambda_p \geq 0)$.
Dans cette base, le vecteur u_1 cherché s'écrit :

$$u_1 = \sum_{i=1}^{p} \alpha_j \ e_j \ \text{ avec } \sum_{i=j}^{p} \alpha_j^2 = 1$$

et on a :

$${}^t u_1 \ MVMu_1 = \Big\langle VM \left(\sum_{j=1}^{p} \alpha_j \ e_j \right), \ \sum_{j'=1}^{p} \alpha_{j'} e_{j'} \Big\rangle = \langle \sum_{j=1}^{p} \lambda_j \alpha_j \ e_j, \ \sum_{j'=1}^{p} \alpha_{j'} e_{j'} \rangle$$

d'où ${}^t u_1 MVMu_1 = \displaystyle\sum_{j=1}^{p} \lambda_j \ \alpha_j^2$.

On doit donc maximiser $\displaystyle\sum_{i=j}^{p} \lambda_j \ \alpha_j^2$ sous la contrainte $\displaystyle\sum_{j=1}^{p} \alpha_j^2 = 1$.

Or

$$\sum_{j=1}^{p} \lambda_j \ \alpha_j^2 \leq \lambda_1 \sum_{j=1}^{p} \alpha_j^2 = \lambda_1.$$

Il suffit donc de prendre $\alpha_1 = 1$ et $\alpha_j = 0$, pour $j > 1$.
Finalement :
u_1 est le vecteur propre de VM associé à la plus grande valeur propre λ_1.
$\lambda_1 = I_{\Delta u_1^{\perp}}$ représente l'inertie du nuage $N(I)$ projeté sur le premier axe factoriel Δu_1.
$\dfrac{\lambda_1}{tr(VM)}$ est la part d'inertie du nuage porté par le premier axe factoriel.
Le deuxième axe factoriel Δu_2 est engendré par le vecteur u_2 , qui est solution du problème.

Problème 2 :
Maximiser ${}^t u_2 MVMu_2$ sous les contraintes : ${}^t u_2 Mu_2 = 1$ et ${}^t u_2 Mu_1 = 0$.

Partant de l'écriture $u_2 = \sum_{j=1}^{p} \alpha_j\ e_j$ avec $\sum_{j=1}^{p} \alpha_j^2 = 1$ dans la base des vecteurs propres de VM. On montre de manière analogue à l'étape précédente que u_2 est le vecteur propre de VM associé à la deuxième plus grande valeur propre λ_2 qui est l'inertie portée par l'axe Δu_2.

$\dfrac{\lambda_1 + \lambda_2}{tr(VM)}$ est la part d'inertie du nuage $N(I)$ porté par le premier plan factoriel engendré par (u_1, u_2).
La recherche du k^{ime} axe factoriel Δu_k engendré par u_k se mène de manière analogue.
u_k est le vecteur propre unitaire de VM associé à la k^{ime} plus grande valeur propre λ_k qui est l'inertie portée par l'axe u_k.
$\dfrac{\lambda_1 + \lambda_2 + ... + \lambda_k}{tr(VM)}$ est la part d'inertie du nuage $N(I)$ portée par l'espace factoriel E_k de dimension k, avec :

$$E_k = \Delta u_1 \oplus \Delta u_2 \oplus \Delta u_3 \oplus ... \oplus \Delta u_k$$

1.7. Facteurs associés aux axes factoriels

A tout vecteur unitaire u de $E = \mathbb{R}^p$ est canoniquement associé la forme linéaire b sur $\mathbb{R}^p$ définie par l'opérateur de la projection sur l'axe Δu. On a donc $b(X) = {}^tXMu$ que l'on notera ${}^tX.b$ en identifiant le vecteur Mu à la forme linéaire b,

$${}^tXMu = {}^tXb = b(X)$$

Ainsi aux axes factoriels de vecteurs unitaires $u_1, u_2, ..., u_r$ (r étant le rang de X) sont associées les formes linéaires $b_1, b_2, ..., b_r$ appelées facteurs de l'analyse en composantes principales.
Il est facile de voir que le premier facteur b_1 est vecteur propre de MV associé à la valeur propre λ_1, que le deuxième facteur b_2 est vecteur propre de MV associé à la valeur propre λ_2. etc...
Les facteurs caractérisent les axes factoriels aussi bien que les valeurs $u_1, ..., u_r$. Ainsi, on montre de mnaière immédiate que la recherche du premier axe factoriel qui est de maximiser ${}^tuMVMu$ sous la contrainte ${}^tuMu = 1$ revient à la recherche de la forme linéaire $b = Mu$ qui maximise ${}^tbVb$ sous la contrainte ${}^tbM^{-1}b = 1$; $b_1 = Mu_1$ est la solution de ce problème. Plus généralement la recherche du k^{ime} axe factoriel Δu_k revient à rechercher la forme linéaire $b = Mu$ qui maximise ${}^tbVb$ sous les contraintes ${}^tbM^{-1}b = 1$ et ${}^tb_\ell M^{-1}b = 0$ pour $\ell = 1, ..., k-1; b_k = Mu_k$ est solution de ce problème.

1.8. Composantes principales

1.8.a) Définition

Pour tout $i = 1, .., n$ la projection de X_i sur le premier axe factoriel Δu_1 s'écrit : $C_1^i u_1$ avec $C_1^i = \langle X_i, u_1 \rangle \ = \ {}^tX_i M u_1 = \ {}^tX_i b_1$.
Le vecteur $C_1 = (C_1^i, i = 1, ..., n)$ de $\mathbb{R}^n$ s'appelle la première composante principale et s'écrit : $C_1 = XMu_1 = Xb_1$.
On définit de manière analogue les autres composantes principales. On notera C_k la k^{ime} composante principale.

1.8.b) Propriétés des composantes principales

Proposition 1
Les composantes principales C_k sont centrées, de variance λ_k et non corrélées deux à deux.

Démonstration
La moyenne de la k^{ime} composante C_k s'écrit :

$$\sum_{i=1}^{n} p_i \ C_k^i = \sum_{i=1}^{n} p_i \ {}^tu_k \ MX_i = \ {}^tu_k \ M \left(\sum_{i=j}^{n} p_i \ X_i \right) = 0$$

Calculons la covariance de C_k et $C_{k'}$.

$$COV(C_k, C_{k'}) \ = \ {}^tC_k \ D_p \ C_{k'} = \ {}^tb_k \ {}^tX \ D_p \ Xb_{k'} = \ {}^tb_k V b_{k'} = \ {}^tu_k MVM u_{k'}$$

$$COV(C_k, C_{k'}) \ = \lambda_k \langle u_k, u_{k'} \rangle$$

D'où $COV(C_k, C_{k'} = 0$, Si $k \neq k'$ et on obtient $Var(C_k) = \lambda_k$.
proposition 2
Les composantes principales C_k sont vecteurs propres de XM^tXD_p associées aux valeurs propres λ_k.

Démonstration
De $MVb_k = \lambda_k \, b_k$ on tire $XMVb_k = \lambda_k X b_k$ Ce qui implique $XM^tXD_pXb_k = \lambda_k X b_k$ d'où le résultat puisque $C_k = Xb_k$.

1.9. Représentation des individus

La projection du nuage $N(I)$ dans les sous - espace vectoriel, de dimension

k, E_k en donne une image approximative. La qualité globale de cette représentation est mesurée par le pourcentage d'inertie pris en compte par E_k :

$$\frac{\lambda_1 + ... + \lambda_k}{tr(VM)} \times 100.$$

Il est important de pouvoir juger de la qualité de représentation de chaque point X_i sur les axes factoriels. Les vecteurs unitaires $u_1, ..., u_p$ des axes factoriels constituent une base M - orthonormée de $\mathbb{R}^p$ et on a :

$$X_i = \sum_{k=1}^{p} c_k^i \; u_k \; , \; c_k = (c_k^i, \; i = 1, ..., n)$$

étant la k^{ime} composante principale. D'où

$$||X_i||_M^2 = \sum_{k=1}^{p} (C_k^i)^2 \text{ et } I_T = \sum_{i=1}^{n} p_i ||X_i||_M^2 = \sum_{k=1}^{p} \sum_{i=1}^{n} p_i (c_k^i)^2 = \sum_{k=1}^{p} \lambda_k$$

On en déduit les indices de qualité de représentation d'un point X_i.

• $(C_k^i)^2 \Big/ ||X_i||_M^2$ est la contribution relative du k^{ime} axe factoriel à l'inertie de X_i : c'est la part d'inertie de X_i prise en compte par cet axe. Cette quantité est le cosinus carré de l'angle θ_k^i formé par X_i et u_k.

• De plus, $\sum_{\ell=1}^{k} (C_\ell^i)^2 \Big/ ||X_i||_M^2 = \sum_{\ell=1}^{k} cos^2 \; \theta_\ell^i$ est la contribution relative de l'espace factoriel E_k engendré par les k premiers axes factoriels à l'inertie de X_i.

• Par ailleurs, $\dfrac{p_i \; (C_k^i)^2}{\lambda_k}$ est la contribution relative de X_i à l'inertie du kième axe.
C'est la part d'inertie de cet axe prise en compte ("expliquée") par le point X_i.

1.10. Représentation des variables

A chaque vecteur propre unitaire u_k de VM correspond une composante principale $C_k \in \mathbb{R}^n$: $C_k = (C_k^i = {}^t u_k M X_i, \; i = 1, ...n)$.
$\mathbb{R}^n$ étant muni de la métrique D_p, on a vu au paragraphe 1) 8) que les p

composantes principales $\left(\dfrac{C_k}{\sqrt{\lambda_k}}\ ,\ k = 1, ..., p\right)$ forment un système D_p -orthonormé de $\mathbb{R}^n$. (Si X est de rang $r < p$, on se restreint aux r composantes principales associées aux valeurs propres non nulles, on suppose ici que $r = p$).
On peut donc représenter les variables X^j dans la base constituée par les vecteurs $\left(\dfrac{C_k}{\sqrt{\lambda_k}}\ , k = 1, ..., p\right)$ du nuage formé par les p points $(X^1, X^2, ..., X^p)$ de $\mathbb{R}^n$.
La k^{ime} coordonnée de X^j dans cette base s'écrit :

$$d_j^k = \frac{{}^tC_k}{\sqrt{\lambda_k}}\ D_p\ X^j \quad \text{et on a } ||X^j||^2_{D_p} = \sum_{k=1}^{p} (d_j^k)^2,$$

d_j^k est la covariance entre la variable X^j et la composante principale normée $\dfrac{C_k}{\sqrt{\lambda_k}}$.
La qualité de représentation d'une variable X^j sur l'axe engendré par $\dfrac{C_k}{\sqrt{\lambda_k}}$ se mesure, comme pour les individus, par le Cosinus carré de l'angle formé par X^j et C_k :
$(d_j^k)^2 \Big/ ||X^j||^2_{D_p}$ est le carré du coefficient de corrélation entre X^j et C_k.
De même, la qualité de représentation d'une variable X^j sur l'espace engendré par les k premières composantes principales normées $\left(\dfrac{C_1}{\sqrt{\lambda_1}}, ..., \dfrac{C_k}{\sqrt{\lambda_k}}\right)$ sera mesurée par :

$$\sum_{\ell=1}^{k} \frac{(d_j^\ell)^2}{||X^j||^2_{D_p}}$$

qui est le carré du coefficient de corrélation multiple de X^j avec $(C_1, ..., C_k)$: une variable sera d'autant mieux représentée que cette quantité sera proche de un.
Par ailleurs, dans le cas très général où la métrique M est diagonale et si on note M_j le terme diagonal générique de M, on montre [Cazes, 1985] que l'inertie totale peut s'écrire :

$$I_T = \sum_{k=1}^{p} \sum_{j=1}^{p} M_j\ (d_j^k)^2 \quad \text{avec} \quad \sum_{j=1}^{p} M_j\ (d_j^k)^2 = \lambda_k.$$

Le rapport $\dfrac{M_j\ (d_j^k)^2}{\lambda_k}$ est alors la contribution relative de la variable X^j à l'inertie portée par le k^{ime} axe factoriel.

1.11. Choix de la Métrique M

L'analyse en composantes principales impose au départ la définition d'une métrique euclidienne M sur $\mathbb{R}^p$. Le choix de cette métrique est très important car il influe beaucoup sur les résultats de l'analyse. On a vu que les deux métriques les plus usitées sont les suivantes :

$M = I_p$: Cette métrique s'emploie lorsque les variables sont mesurées dans des unités identiques et ont des variances de même ordre.

La matrice à diagonaliser $VM = VI_p = V$ est alors la matrice de variance - covariance des variables.

$M = D_{1/\sigma^2}$: Cette matrice s'emploie lorsque les variables sont mesurées dans des unités différentes et plus généralement lorsqu'elles ont des variance notablement différentes.

L'emploie de cette métrique revient à analyser le nuage centré réduit avec la métrique I_p. Le choix de cette métrique rend les résultats de l'analyse indépendants des unités de mesure choisies pour les variables. Par ailleurs, il permet de rédurie considérablement l'effet taille de l'analyse en composantes principales sur les variables.

1.12. Les éléments illustratifs

Dans toute analyse factorielle, il est possible de projeter dans le système d'axes trouvé des individus ou des variables n'ayant pas participé en analyse. On parlera d'éléments illustratifs ou supplémentaires. Les formules de projection de ces points sur un axe factoriel sont les suivantes :

• Pour un individu illustratif X_s, l'abscisse de sa projection sur le k^{ime} axe Δu_k est ${}^tX_sMu_k = {}^tX_sb_k$.

• Pour une variable illustrative X^s, l'abscisse de sa projection sur le k^{ime} axe Δ_{C_k} est ${}^t(X^s)\ D_p\ \dfrac{C_k}{\sqrt{\lambda_k}}$.