Cours d'analyse des données et apprentissage : L'analyse en composantes principales

Jean-Marie Monnez

COURS D'ANALYSE DES DONNEES ET APPRENTISSAGE
L'analyse en composantes principales

Jean-Marie MONNEZ
Université de Lorraine, CNRS, Inria, Institut Elie Cartan de Lorraine, France

April 29, 2021

1 Données et buts

1.1 Données

Soit :

- $I = \{1, 2, \cdots, n\}$ un ensemble d'individus ;
- $P = \{p_1, p_2, \cdots, p_n\}$ un ensemble de poids attribués aux individus, avec $0 < p_i < 1$, $\sum_{i=1}^{n} p_i = 1$;
- $x^1, x^2, \cdots, x^p$ des caractères quantitatifs mesurés sur les individus.

On centre les caractères x^j (on calcule les moyennes $\bar{x}^j = \sum_{i=1}^{n} p_i x_i^j$ et on centre les données) ; désormais, on note x_i^j la valeur centrée du caractère x^j pour l'individu i.

On a le tableau de données suivant dans lequel les individus sont en ligne et les caractères en colonne :

$$X = \begin{array}{c} 1 \\ \vdots \\ i \\ \vdots \\ n \end{array} \begin{pmatrix} & & \vdots & & \\ \cdots & \cdots & x_i^j & \cdots & \cdots \\ & & \vdots & & \end{pmatrix}.$$

On note $\underline{x}_i = \begin{pmatrix} x_i^1 \\ \vdots \\ x_i^p \end{pmatrix}$, $\underline{x}^j = \begin{pmatrix} x_1^j \\ \vdots \\ x_n^j \end{pmatrix}$. $\underline{x}_i$ représente l'individu i et $\underline{x}^j$ le caractère x^j.

Alors, $X = \begin{pmatrix} \underline{x}_1' \\ \vdots \\ \underline{x}_i' \\ \vdots \\ \underline{x}_n' \end{pmatrix} = (\underline{x}^1 \cdots \underline{x}^j \cdots \underline{x}^p)$.

On note D la matrice diagonale des poids des individus :

$$D = \begin{pmatrix} p_1 & & & \\ & \ddots & & 0 \\ & & \ddots & \\ 0 & & & p_n \end{pmatrix}.$$

1.2 Buts de l'étude

Les buts de cette étude sont de *décrire* (mettre en évidence et analyser les ressemblances et dissemblances entre individus, les corrélations entre caractères) et *synthétiser* (déterminer des facteurs combinaisons linéaires des caractères) le tableau de données X et d'en effectuer une *représentation graphique*. Elle est divisée en trois parties:

1) une étude dans l'espace des individus $\mathbb{R}^p$, dans laquelle on détermine les *axes principaux* ;

2) une étude dans le dual $\mathbb{R}^{p*}$ de $\mathbb{R}^p$, dans laquelle on détermine les *facteurs principaux* ;

3) une étude dans l'espace des caractères $\mathbb{R}^n$, dans laquelle on détermine les *composantes principales*.

Cette analyse conduit à remplacer un nombre important de caractères par un nombre réduit de facteurs non corrélés. En ce sens, elle peut être mise en oeuvre préalablement à une autre analyse statistique, par exemple une régression multiple.

2 Etude géométrique dans l'espace des individus $\mathbb{R}^p$

Le but de cette étude est de déterminer une représentation graphique des individus par des points de telle manière que deux individus ayant des valeurs voisines (respectivement éloignées) pour les caractères soient représentés par des points voisins (respectivement éloignés).

2.1 Représentation dans un espace euclidien

2.1.1 Représentation d'un individu par un point

On représente l'individu i par le point A_i de $\mathbb{R}^p$ de coordonnées $x_i^1, \cdots, x_i^p$ qui constituent le vecteur $\underline{x}_i$:

$$i \mapsto A_i(\underline{x}_i).$$

On affecte le poids p_i au point A_i.

Le barycentre G des points (A_i, p_i) est à l'origine car les caractères sont centrés : $\overrightarrow{OG} = \sum_{i=1}^{n} p_i \overrightarrow{OA_i}$ a pour j$^{\text{ième}}$ composante $g^j = \sum_{i=1}^{n} p_i \, x_i^j = 0$.

2.1.2 Définition d'une distance entre les points A_i

On convient de représenter la différence entre deux individus vis-à-vis des caractères par un nombre qui est une distance euclidienne entre les points représentatifs de ces individus.

On choisit une métrique M dans $\mathbb{R}^p$ à laquelle est associée la distance euclidienne d. Pour deux points $A_i(\underline{x}_i)$ et $A_{i'}(\underline{x}_{i'})$, on a :

$$d^2(A_i, A_{i'}) = ||\underline{x}_i - \underline{x}_{i'}||^2 = (\underline{x}_i - \underline{x}_{i'})' M (\underline{x}_i - \underline{x}_{i'}).$$

Différents choix de métriques sont possibles.

1) Distance euclidienne usuelle

On prend $M = I$.

$$\boxed{d(A_i, A_{i'}) = \sqrt{\sum_{j=1}^{p} (x_i^j - x_{i'}^j)^2}}$$

Cette distance présente un inconvénient lorsque les caractères mesurés sont hétérogènes, c'est-à-dire ne s'expriment pas dans la même unité, comme le montre l'exemple suivant.

Exemple On a mesuré dans deux pays les exportations, en millions de dollars, et le taux d'escompte, en pourcentage:

	Exportations	Taux d'escompte
Pays 1	28 054	7,00
Pays 2	2 306	5,50
Différence	25 748	1,50

On constate que le taux d'escompte n'interviendra pratiquement pas dans la distance entre les deux pays. De plus, les différences calculées dépendent des unités choisies.■

2) Distance de l'ACP normée

On note s^j l'écart-type du caractère x^j : $s^j = \sqrt{\sum_{i=1}^{n} p_i \, (x_i^j)^2}$. On définit la distance:

$$\boxed{d(A_i, A_{i'}) = \sqrt{\sum_{j=1}^{p} \left(\frac{x_i^j - x_{i'}^j}{s^j} \right)^2} = \sqrt{\sum_{j=1}^{p} \frac{1}{(s^j)^2} (x_i^j - x_{i'}^j)^2}}$$

Les différences entre les valeurs de deux caractères s'expriment alors en écarts-types.

a) Elles ne dépendent plus de l'unité de mesure choisie, car l'unité de s^j est la même que celle de x_i^j.

b) Elles sont du même ordre de grandeur, comprises en général entre 0 et 6.

En effet, le caractère x^j étant centré, d'après l'inégalité de Bienaymé-Tchébychev, au moins $\frac{8}{9}$ de ses valeurs x_i^j sont comprises entre $-3s^j$ et $+3s^j$.

Choisir cette distance revient à prendre dans $\mathbb{R}^p$ la métrique diagonale suivante, appelée métrique de l'inverse des variances :

$$M = \begin{pmatrix} \frac{1}{(s^1)^2} & & & & \\ & \ddots & & & \\ & & \frac{1}{(s^j)^2} & & \\ & & & \ddots & \\ & & & & \frac{1}{(s^p)^2} \end{pmatrix}.$$

On peut aussi considérer que ceci revient à prendre les données centrées réduites $\frac{x_i^j}{s^j}$ et à utiliser la métrique-identité.

3) Autres métriques possibles

On peut utiliser une métrique diagonale de poids attribués aux caractères :

$$M = \begin{pmatrix} q^1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & q^p \end{pmatrix}, \quad d(A_i, A_{i'}) = \sqrt{\sum_{j=1}^{p} q^j \, (x_i^j - x_{i'}^j)^2}.$$

C'est ce que l'on fait par exemple dans l'analyse factorielle multiple, qui est étudiée dans le chapitre sur l'analyse factorielle de tableaux multiples.

En analyse canonique généralisée, également étudiée dans le chapitre cité précédemment, on utilise une métrique M diagonale par blocs :

$$M = \begin{pmatrix} M^1 & & & & \\ & \ddots & & & 0 \\ & & M^2 & & \\ & & & \ddots & \\ 0 & & & & M^q \end{pmatrix}.$$

2.1.3 Visualisation

Si l'on pouvait voir dans $\mathbb{R}^p$, le but de l'étude serait atteint. *On va alors projeter le nuage de points A_i sur un sous-espace de faible dimension, choisi de façon optimale.*

2.2 Détermination du sous-espace de projection

2.2.1 Critére de détermination du sous-espace de projection

On détermine un sous-espace de projection F_r, de dimension fixée r, tel que la projection du nuage des points A_i sur ce sous-espace soit l'image la plus fidèle possible du nuage initial.

1) Première expression

On note P_i la projection de A_i sur F_r. On a :

$$d(P_i, P_{i'}) = ||\overrightarrow{P_i P_{i'}}|| \leq ||\overrightarrow{A_i A_{i'}}|| = d(A_i, A_{i'}).$$

Donc :

$$\underbrace{\sum_{i=1}^{n}\sum_{i'=1}^{n} p_i p_{i'} \, ||\overrightarrow{P_i P_{i'}}||^2}_{\text{dépend du sous-espace } F_r} \leq \underbrace{\sum_{i=1}^{n}\sum_{i'=1}^{n} p_i p_{i'} ||\overrightarrow{A_i A_{i'}}||^2}_{\text{nombre fixe}}$$

Pour que, dans leur ensemble, les distances entre les points P_i soient les plus proches possible des distances entre les points A_i, on détermine F_r de telle manière que la somme $\sum_{i=1}^{n}\sum_{i'=1}^{n} p_i p_{i'} \, ||\overrightarrow{P_i P_{i'}}||^2$ soit maximale.

Critère *F_r est un sous-espace de dimension r qui rend maximale la somme* $\sum_{i=1}^{n}\sum_{i=1}^{n} p_i p_{i'} \, ||\overrightarrow{P_i P_{i'}}||^2$.

2) Deuxième expression

On note G' la projection du barycentre G sur F_r : G' est le barycentre des points (P_i, p_i).

Lemme $\sum_{i=1}^{n}\sum_{i'=1}^{n} p_i p_{i'} ||\overrightarrow{P_iP_{i'}}||^2 = 2\sum_{i=1}^{n} p_i ||\overrightarrow{G'P_i}||^2$.

Démonstration
$||\overrightarrow{P_iP_{i'}}||^2 = ||\overrightarrow{P_iG'} + \overrightarrow{G'P_{i'}}||^2 = ||\overrightarrow{P_iG'}||^2 + ||\overrightarrow{G'P_{i'}}||^2 + 2 < \overrightarrow{P_iG'}, \overrightarrow{G'P_{i'}} >$

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{i'=1}^{n} p_i p_{i'} ||\overrightarrow{P_iP_{i'}}||^2 &= \sum_{i=1}^{n}\sum_{i'=1}^{n} p_i p_{i'} \big(||\overrightarrow{P_iG'}||^2 + ||\overrightarrow{G'P_{i'}}||^2 + 2 < \overrightarrow{P_iG'}, \overrightarrow{G'P_{i'}} > \big) \\
&= 2\sum_{i=1}^{n} p_i ||\overrightarrow{P_iG'}||^2 + 2 < \sum_{i=1}^{n} p_i \overrightarrow{P_iG'}, \sum_{i'=1}^{n} p_{i'} \overrightarrow{G'P_{i'}} > \\
&= 2\sum_{i=1}^{n} p_i ||\overrightarrow{G'P_i}||^2 \ \text{ car } \ \sum_{i'=1}^{n} p_{i'} \overrightarrow{P_iG'} = \overrightarrow{0}.\blacksquare
\end{aligned}$$

Dans le triangle $G'A_iP_i$ rectangle en P_i, on a :

$$||\overrightarrow{G'A_i}||^2 = ||\overrightarrow{G'P_i}||^2 + ||\overrightarrow{A_iP_i}||^2.$$

$$\sum_{i=1}^{n} p_i \, ||\overrightarrow{A_iP_i}||^2 = \sum_{i=1}^{n} p_i ||\overrightarrow{G'A_i}||^2 - \underbrace{\sum_{i=1}^{n} p_i ||\overrightarrow{G'P_i}||^2}_{\max}$$

D'après le lemme, le critère précédent revient à maximiser $\sum_{i=1}^{n} p_i ||\overrightarrow{G'P_i}||^2$.

$$\begin{aligned}
\sum_{i=1}^{n} p_i ||\overrightarrow{G'A_i}||^2 &= \sum_{i=1}^{n} p_i \, ||\overrightarrow{G'G} + \overrightarrow{GA_i}||^2 \\
&= \sum_{i=1}^{n} p_i \big(||\overrightarrow{G'G}||^2 + 2 < \overrightarrow{G'G}, \overrightarrow{GA_i} > + ||\overrightarrow{GA_i}||^2\big) \\
&= ||\overrightarrow{G'G}||^2 + 2 < \overrightarrow{G'G}, \ \sum_{i=1}^{n} p_i \ \overrightarrow{GA_i} > + \sum_{i=1}^{n} p_i \ ||\overrightarrow{GA_i}||^2
\end{aligned}$$

Or : $\sum_{i=1}^{n} p_i \ \overrightarrow{GA_i} = \overrightarrow{0}$. Donc :

$$\sum_{i=1}^{n} p_i \, ||A_iP_i||^2 = ||\overrightarrow{G'G}||^2 + \sum_{i=1}^{n} p_i \, ||\overrightarrow{GA_i}||^2 - \sum_{i=1}^{n} p_i \, ||\overrightarrow{G'P_i}||^2.$$

$$\sum_{i=1}^{n} p_i \, ||\overrightarrow{G'P_i}||^2 \ \max \quad et \quad G' = G \Longleftrightarrow \sum_{i=1}^{n} p_i \, ||A_iP_i||^2 \ \min$$

C'est ce critère des moindres carrés que l'on utilise pour définir le sous-espace affine F_r, qui contient alors le barycentre G.

Critère *F_r est un sous-espace affine de dimension r qui rend minimale la somme $\sum_{i=1}^{n} p_i \, ||\overrightarrow{A_iP_i}||^2$: il ajuste au mieux le nuage des points (A_i, p_i) au sens des moindres carrés, les distances des points A_i à F_r étant prises orthogonalement.*

On appelle :
- inertie des points A_i par rapport à G : $\sum_{i=1}^{n} p_i \, ||\overrightarrow{GA_i}||^2$;
- inertie des points P_i par rapport à G : $\sum_{i=1}^{n} p_i \, ||\overrightarrow{GP_i}||^2$;
- inertie des points A_i par rapport à F_r : $\sum_{i=1}^{n} p_i \, ||\overrightarrow{A_iP_i}||^2$.

2.2.2 Détermination des axes principaux

On note $(\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_r)$ une base M-orthonormée de F_r.

Lemme $\sum_{i=1}^{n} p_i ||\overrightarrow{GP_i}||^2 = \sum_{k=1}^{r} \underline{u}'_k \, MX'DXM \, \underline{u}_k$.

Démonstration

1) Notons P_i^k la projection de P_i sur l'axe $(G, \vec{u}_k)$. D'après le théorème des trois perpendiculaires, P_i^k est la projection de A_i sur l'axe $(G, \vec{u}_k)$.

En effet, comme $\vec{u}_k$ est unitaire et $\overrightarrow{A_iP_i}$ orthogonal à $\overrightarrow{u}_k$:

$$\overline{GP_i^k} =< \overrightarrow{GP_i}, \vec{u}_k >=< \overrightarrow{GA_i} + \overrightarrow{A_iP_i}, \vec{u}_k >=< \overrightarrow{GA_i}, \vec{u}_k > .$$

$\overline{GP_i^k} =< \overrightarrow{GA_i}, \vec{u}_k >$; donc, P_i^k est la projection orthogonale de A_i sur $(G, \vec{u}_k)$.

2) Comme la base $(\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_r)$ est orthonormée, on a :

$$\left\| \overrightarrow{GP_i} \right\|^2 = \sum_{k=1}^{r} ||\overrightarrow{GP_i^k}||^2 \Longrightarrow \sum_{i=1}^{n} p_i \, ||\overrightarrow{GP_i}||^2 = \sum_{k=1}^{r} \sum_{i=1}^{n} p_i \, ||\overrightarrow{GP_i^k}||^2.$$

3) $\overline{GP_i^k} = \overrightarrow{GA_i} \cdot \vec{u}_k = \underline{x}'_i M \underline{u}_k = \underline{u}'_k M \underline{x}_i$

$$\sum_{i=1}^{n} p_i \, ||\overrightarrow{GP_i^k}||^2 = \sum_{i=1}^{n} p_i \, \underline{u}'_k M \underline{x}_i \underline{x}'_i M \underline{u}_k = \underline{u}'_k M \Big(\sum_{i=1}^{n} p_i \, \underline{x}_i \underline{x}'_i \Big) M \underline{u}_k$$

$$\sum_{i=1}^{n} p_i \, \underline{x}_i \underline{x}'_i = (\underline{x}_1 \cdots \underline{x}_i \cdots \underline{x}_n) \begin{pmatrix} p_1 & & & \\ & \ddots & & 0 \\ & & \ddots & \\ 0 & & & p_n \end{pmatrix} \begin{pmatrix} \underline{x}'_1 \\ \vdots \\ \underline{x}'_i \\ \vdots \\ \underline{x}'_n \end{pmatrix} = X'DX$$

L'élément (l, m) de $X'DX$ est égal à $\sum_{i=1}^{n} p_i \, x_i^l x_i^m$. On a :

$$\begin{aligned} \sum_{i=1}^{n} p_i \, ||\overrightarrow{GP_i^k}||^2 &= \underline{u}'_k MX'DXM\underline{u}_k \\ \sum_{i=1}^{n} p_i \, ||\overrightarrow{GP_i}||^2 &= \sum_{k=1}^{r} \underline{u}'_k MX'DXM\underline{u}_k . \blacksquare \end{aligned}$$

Le problème est donc maintenant de déterminer une base $(\underline{u}_1, \cdots, \underline{u}_r)$ M-orthonormée telle que $\sum_{k=1}^{r} \underline{u}'_k MX'DXM\underline{u}_k$ soit maximale. On montre, en utilisant le théorème suivant, que l'on peut procéder pas à pas en déterminant d'abord $\underline{u}_1$, puis $\underline{u}_2$, ..., puis $\underline{u}_r$.

Théorème *A partir d'un sous-espace F_{r-1} vérifiant le critère, on peut construire un sous-espace F_r le vérifiant en lui adjoignant une droite M-orthogonale $\Delta\underline{u}_r$ telle que la forme quadratique $\underline{u}'_r MX'DXM\underline{u}_r$ soit maximale sous la contrainte $\underline{u}'_r M\underline{u}_r = 1$.*

Ceci revient à dire : pour maximiser $\sum_{k=1}^{r} \underline{u}_k' MX'DXM\underline{u}_k$ sous la contrainte $(\vec{u}_1, \cdots, \vec{u}_{r-1}, \vec{u}_r)$ M-orthonormée, on maximise $\sum_{k=1}^{r-1} \underline{u}_k' MX'DXM\underline{u}_k$ sous la contrainte $(\vec{u}_1, \cdots, \vec{u}_{r-1})$ M-orthonormée et $\underline{u}_r' MX'DXM\underline{u}_r$ sous les contraintes $\vec{u}_r$ unitaire et $\vec{u}_r$ orthogonal à $\vec{u}_j$ pour $j = 1, \cdots, r-1$.

En pratique, on peut donc procéder pas à pas pour déterminer les vecteurs $\vec{u}_k$:

$$\begin{cases} F_1 = (G, \vec{u}_1) & : \underline{u}_1' MX'DXM\underline{u}_1 \text{ max}, \quad \underline{u}_1' M\underline{u}_1 = 1. \\ F_2 = (G, \vec{u}_1, \vec{u}_2) & : \underline{u}_2' MX'DXM\underline{u}_2 \text{ max}, \quad \underline{u}_2' M\underline{u}_2 = 1, \underline{u}_2' M\underline{u}_1 = 0\ . \\ F_3 = (G, \vec{u}_1, \vec{u}_2, \vec{u}_3) & : \underline{u}_3' MX'DXM\underline{u}_3 \text{ max}, \quad \underline{u}_3' M\underline{u}_3 = 1, \underline{u}_3' M\underline{u}_1 = 0, \underline{u}_3' M\underline{u}_2 = 0. \\ etc... \end{cases}$$

D'après le théorème MSC, on a le résultat suivant :

Théorème *$\vec{u}_k$ est vecteur propre de $X'DXM$ associé à sa $k^{ième}$ plus grande valeur propre λ_k. On a : $\sum_{i=1}^{n} p_i \, ||\overrightarrow{GP_i^k}||^2 = \lambda_k$.*

Démonstration

$\vec{u}_k$ rend maximale la forme quadratique $\underline{u}' MX'DXM\underline{u}$

sous les contraintes : $\begin{cases} \underline{u}' M\underline{u} = 1, \\ \underline{u}' M\underline{u}_j = 0, \quad j = 1, ..., k-1. \end{cases}$

La matrice $A = MX'DXM$ est symétrique. D'après le théorème de maximisation sous contraintes d'une forme quadratique étudié dans le chapitre précédent, $\vec{u}_k$ est vecteur propre de $M^{-1}A = M^{-1}MX'DXM = X'DXM$ associé à la $k^{ième}$ plus grande valeur propre. D'après la démonstration du lemme, on a :

$$\sum_{i=1}^{n} p_i \, ||\overrightarrow{GP_i^k}||^2 = \underline{u}_k' \, MX'DXM \, \underline{u}_k = \lambda_k \, \underline{u}_k' \, M \, \underline{u}_k = \lambda_k. \blacksquare$$

Définition *L'axe $(G, \vec{u}_k)$ est appelé le $k^{\text{ième}}$ axe principal.*

Remarque $V = X'DX$ est la matrice des covariances des caractères $x^1, x^2, \cdots, x^p$. Si les caractères sont réduits, $X'DX$ est la matrice de corrélation (matrice des coefficients de corrélation linéaire) dont l'élément général (l, m) est :

$$\sum_{i=1}^{n} p_i \, \frac{x_i^l}{s^l} \, \frac{x_i^m}{s^m} = \frac{1}{s^l s^m} \, \sum_{i=1}^{n} p_i x_i^l x_i^m = r(x^l, x^m). \blacksquare$$

2.2.3 Qualité globale de représentation du nuage par projection

Définition *On appelle qualité globale de représentation du nuage des points A_i par projection sur F_r le rapport*

$$\frac{\sum_{i=1}^{n} \sum_{i'=1}^{n} p_i p_{i'} \, ||\overrightarrow{P_i P_{i'}}||^2}{\sum_{i=1}^{n} \sum_{i=1}^{n} p_i p_{i'} ||\overrightarrow{A_i A_{i'}}||^2} = \frac{\sum_{i=1}^{n} p_i ||\overrightarrow{GP_i}||^2}{\sum_{i=1}^{n} p_i \, ||\overrightarrow{GA_i}||^2} = \frac{\text{inertie des projections } P_i \text{ par rapport à } G}{\text{inertie des points } A_i \text{ par rapport à G}}.$$

On l'appelle aussi pourcentage d'inertie expliquée par le sous-espace F_r.

Proposition *La qualité globale de représentation par projection sur F_r est égale à $\dfrac{\lambda_1 + \lambda_2 + \cdots + \lambda_r}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} = \dfrac{\lambda_1 + \lambda_2 + \cdots + \lambda_r}{\text{Trace}\,(X'DXM)}$. C'est la somme des qualités globales de représentation par projection sur les axes principaux qui engendrent F_r.*

Démonstration

$$\sum_{i=1}^{n} p_i ||\overrightarrow{GP_i}||^2 = \sum_{k=1}^{r}\sum_{i=1}^{n} p_i ||\overrightarrow{GP_i^k}||^2 = \sum_{k=1}^{r} \lambda_k$$

Comme $(\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_p)$ est une base M-orthonormée de $\mathbb{R}^p$ et comme P_i^k est la projection de A_i sur $(G, \overrightarrow{u}_k)$, on a :

$$\begin{aligned} ||\overrightarrow{GA_i}||^2 &= \sum_{k=1}^{p} ||\overrightarrow{GP_i^k}||^2. \\ \sum_{i=1}^{n} p_i ||\overrightarrow{GA_i}||^2 &= \sum_{k=1}^{p}\sum_{i=1}^{n} p_i ||\overrightarrow{GP_i^k}||^2 = \sum_{k=1}^{p} \lambda_k. \end{aligned}$$

La qualité globale de représentation par projection sur l'axe principal $(G, \vec{u}_k)$ est égale à

$$\frac{\sum_{i=1}^{n} p_i ||\overrightarrow{GP_i^k}||^2}{\sum_{i=1}^{n} p_i ||\overrightarrow{GA_i}||^2} = \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_p}.\blacksquare$$

Remarque Dans le cas de l'ACP normée, la matrice $X'DXM$ a des 1 sur la diagonale. Alors Trace $(X'DXM) = p.\blacksquare$

2.3 Projection du nuage

2.3.1 Projection des points $\mathbf{A}_i$

1) Abscisse de la projection d'un point sur un axe principal

Proposition *L'abscisse de la projection du point A_i sur l'axe principal $(G, \vec{u}_k)$, $\overline{GP_i^k}$, est égale à $\underline{x}_i' M \underline{u}_k$.*

2) Qualité de la représentation d'un point par projection

Définition *On appelle qualité individuelle de représentation du point A_i par projection sur l'axe principal $(G, \vec{u}_k)$ le rapport $\frac{||\overrightarrow{GP_i^k}||^2}{||\overrightarrow{GA_i}||^2} = \cos^2(\overrightarrow{GA_i}, \overrightarrow{GP_i^k})$. On appelle qualité individuelle de représentation de A_i par projection sur le sous-espace F_r le rapport $\frac{||\overrightarrow{GP_i}||^2}{||\overrightarrow{GA_i}||^2} = \cos^2(\overrightarrow{GA_i}, \overrightarrow{GP_i})$.*

On a :

$$\begin{aligned} ||\overrightarrow{GP_i}||^2 &= \sum_{k=1}^{r} ||\overrightarrow{GP_i^k}||^2; \frac{||\overrightarrow{GP_i}||^2}{||\overrightarrow{GA_i}||^2} = \sum_{k=1}^{r} \frac{||\overrightarrow{GP_i^k}||^2}{||\overrightarrow{GA_i}||^2}; \\ ||\overrightarrow{GP_i^k}||^2 &= (\underline{x}_i' M u_k)^2 \ ; \ ||\overrightarrow{GA_i}||^2 = \underline{x}_i' M \underline{x}_i. \end{aligned}$$

Proposition *La qualité individuelle de représentation de A_i par projection sur F_r est égale à $\sum_{k=1}^{r} \frac{(\underline{x}_i' M \underline{u}_k)^2}{\underline{x}_i M \underline{x}_i}$; c'est la somme des qualités individuelles de représentation par projection sur les axes principaux qui engendrent F_r.*

Interprétation On doit interpréter avec prudence la position relative par rapport aux autres de la projection d'un point ayant une faible qualité de représentation. La proximité entre deux projections de points ayant une faible qualité de représentation n'implique pas que ces individus soient semblables vis-à-vis des caractères.

3) Contribution d'un point à l'inertie expliquée par un axe

Définition *On appelle contribution de la projection P_i^k du point A_i sur l'axe principal $(G, \vec{u}_k)$ à l'inertie expliquée par cet axe le rapport* $\dfrac{p_i||\overrightarrow{GP_i^k}||^2}{\sum_{l=1}^{n} p_l||\overrightarrow{GP_l^k}||^2}$.

On a : $\sum_{l=1}^{n} p_l||\overrightarrow{GP_l^k}||^2 = \lambda_k$.

Proposition *Cette contribution est égale à* $\frac{p_i(\underline{x}_i' M \underline{u}_k)^2}{\lambda_k}$.

Interprétation L'examen des contributions permet de repérer d'éventuels individus aberrants ayant une contribution à l'inertie expliquée par un axe principal très forte par rapport à celle des autres, et qui peuvent parfois expliquer à eux seuls la construction de cet axe. Dans certains cas, on peut être amené à éliminer ces individus et à les faire figurer en individus supplémentaires ou passifs.

2.3.2 Projection des axes de la base canonique de $\mathbb{R}^p$

Soit $(\vec{e}_1, \cdots, \vec{e}_p)$ la base canonique de $\mathbb{R}^p$. L'axe $(G, \vec{e}_j)$ représente le j^{ième} caractère x^j.

On note u_k^l la $l^{ième}$ composante de $\underline{u}_k$.

Proposition *La mesure algébrique de la projection de $\vec{e}_j$ sur l'axe principal $(G, \vec{u}_k)$ est égale à* $< \vec{e}_j, \vec{u}_k >= \underline{e}_j' M \underline{u}_k = \sum_{l=1}^{p} M_{jl} u_k^l$. *Pour M diagonale, elle est égale à $M_{jj} u_k^j$.*

3 Etude statistique dans le dual $\mathbb{R}^{p*}$ de $\mathbb{R}^p$

Dans l'étude précédente, on a déterminé des axes principaux $(G, \vec{u}_k)$ dans $\mathbb{R}^p$.

Dans $\mathbb{R}^p$, un axe de la base canonique représente un caractère ; un axe principal représente un "nouveau caractère", combinaison linéaire des caractères $x^1, ..., x^p$ centrés, que l'on appelle *facteur principal*. L'introduction de la notion de facteur va permettre d'interpréter statistiquement l'étude géométrique effectuée.

3.1 Définition des facteurs principaux

Définition *Le $k^{ième}$ facteur principal est une combinaison linéaire des caractères $x^1, x^2, \cdots, x^p$ centrés telle que sa valeur pour l'individu i soit l'abscisse de la projection P_i^k du point A_i sur l'axe principal $(G, \vec{u}_k)$, $\overline{GP_i^k} = \underline{x}_i' M \underline{u}_k$.*

Notons $\underline{a}_k = M\underline{u}_k = \begin{pmatrix} a_k^1 \\ \vdots \\ a_k^p \end{pmatrix}$.

La valeur du k^{ième} facteur par l'individu i est par définition :

$$\underline{x}_i' M \underline{u}_k = \underline{x}_i' \underline{a}_k = (x_i^1 ..... x_i^p) \begin{pmatrix} a_k^1 \\ \vdots \\ a_k^p \end{pmatrix} = a_k^1 x_i^1 + \cdots + a_k^p x_i^p.$$

Le k^{ième} facteur est la combinaison linéaire $a_k^1 x^1 + a_k^2 x^2 + \cdots + a_k^p x^p$ des caractères $x^1, ..., x^p$ centrés.

Définition *Le vecteur $\underline{a}_k = M\underline{u}_k$ des coefficients de la combinaison linéaire qui définit le k^{ième} facteur est un élément du dual $\mathbb{R}^{p*}$ de $\mathbb{R}^p$ que l'on appelle aussi $k^{ième}$ facteur principal.*

On définit l'application M de $\mathbb{R}^p$ dans $\mathbb{R}^{p*}$:

$$\begin{aligned} M &: \mathbb{R}^p \longrightarrow \mathbb{R}^{p*} \\ M &: \underline{u}_k \longmapsto \underline{a}_k = M\underline{u}_k \end{aligned}$$

Théorème *$\underline{a}_k$ est vecteur propre M^{-1}-unitaire de la matrice $MX'DX$ associé à sa $k^{ième}$ plus grande valeur propre λ_k.*

Démonstration

$\underline{u}_k$ est vecteur propre M-unitaire de $X'DXM$ associé à sa k$^{\text{ième}}$ plus grande valeur propre λ_k : $X'DXM\,\underline{u}_k = \lambda_k\,\underline{u}_k$.

$$\begin{aligned} X'DXM\underline{u}_k &= \lambda_k\underline{u}_k \Longleftrightarrow MX'DXM\,\underline{u}_k = \lambda_k M\underline{u}_k \Longleftrightarrow MX'DX\underline{a}_k = \lambda_k\,\underline{a}_k \\ \underline{u}_k'M\underline{u}_k &= 1 \Longleftrightarrow \underline{u}_k'MM^{-1}M\underline{u}_k = 1 \Longleftrightarrow \underline{a}_k'M^{-1}\underline{a}_k = 1.\blacksquare \end{aligned}$$

3.2 Interprétation statistique de l'ACP

Théorème *Pour $\lambda_{k-1} \neq 0$, $\underline{a}_k$ est solution du problème de la recherche de $\underline{a} \in \mathbb{R}^{p*}$ tel que $\underline{a}'X'DX\,\underline{a}$ soit maximal sous les contraintes $\underline{a}'X'DX\underline{a}_j = 0$, $j = 1, 2, \cdots, k-1$, $\underline{a}'M^{-1}\underline{a} = 1$.*

Démonstration

$\underline{u}_k$ est solution du problème de la recherche de $\underline{u} \in \mathbb{R}^p$ tel que

$$\underline{u}'MX'DXM\,\underline{u} \quad maximal,$$

sous les contraintes

$$\begin{aligned} \underline{u}'M\underline{u}_j &= 0, \quad j = 1, \cdots, k-1, \\ \underline{u}'M\underline{u} &= 1. \end{aligned}$$

. $\underline{u}'MX'DXM\underline{u}$ maximal $\Longleftrightarrow \underline{a}'X'DX\underline{a}$ maximal ;
. $\underline{u}'M\underline{u} = 1 \Longleftrightarrow \underline{u}'MM^{-1}M\underline{u} = 1 \Longleftrightarrow \underline{a}'M^{-1}\underline{a} = 1$;
. Pour $j = 1, \cdots, k-1$, $\underline{u}_j$ est vecteur propre de $X'DXM$: $X'DXM\underline{u}_j = \lambda_j\underline{u}_j \Longleftrightarrow \underline{u}_j = \frac{1}{\lambda_j}X'DXM\underline{u}_j$, pour $\lambda_j \neq 0$; donc :
$\underline{u}'M\underline{u}_j = 0 \Longleftrightarrow \underline{u}'MX'DXM\underline{u}_j = 0 \Longleftrightarrow \underline{a}'X'DX\underline{a}_j = 0.\blacksquare$

Interprétation

$\underline{a}_k$ est le vecteur des coefficients d'une combinaison linéaire des caractères $x^1, x^2, \cdots, x^p$.

$$X\underline{a}_k = \begin{pmatrix} & x_i^1 \ldots\ldots x_i^p & \end{pmatrix} \begin{pmatrix} a_k^1 \\ \vdots \\ a_k^p \end{pmatrix} = \begin{pmatrix} \vdots \\ \sum_{j=1}^{p} a_k^j x_i^j \\ \vdots \end{pmatrix} \leftarrow \text{valeur du k}^{\text{ième}} \text{ facteur pour l'individu } i.$$

$\underline{a}_k'X'DX\underline{a}_k = (X\underline{a}_k)'D(X\underline{a}_k) =$

$$\Big(\cdots \sum_{j=1}^{p} a_k^j x_i^j \cdots\Big) \begin{pmatrix} p_1 & & & \\ & \ddots & & 0 \\ & & \ddots & \\ & 0 & & p_n \end{pmatrix} \begin{pmatrix} \vdots \\ \sum_{j=1}^{p} a_k^j x_i^j \\ \vdots \end{pmatrix} = \sum_{i=1}^{n} p_i \Big(\sum_{j=1}^{p} a_k^j x_i^j\Big)^2 \text{ : c'est la variance des valeurs}$$

du k$^{\text{ième}}$ facteur.

$\underline{a}_k'X'DX\underline{a}_j$ est la covariance du k$^{\text{ième}}$ et du j$^{\text{ième}}$ facteur.

On a donc le résultat suivant :

Proposition *Le k*$^{\text{ième}}$ *facteur de l'ACP est une combinaison linéaire des caractères* $x^1, x^2, \cdots, x^p$ *centrés de variance maximale sous les contraintes qu'elle soit non corrélée aux facteurs précédents et la contrainte de normalisation* $\underline{a}_k' M^{-1} \underline{a}_k = 1$.

On peut présenter l'ACP comme *la recherche*
. d'une combinaison linéaire des caractères centrés de variance maximale,
. puis d'une deuxième combinaison non corrélée à la première et de variance maximale,
. puis d'une troisième combinaison non corrélée aux deux premières et de variance maximale et ainsi de suite.

L'ACP conduit à remplacer p caractères par un nombre réduit de facteurs non corrélés qui peuvent alors être utilisés dans une autre analyse.

Remarque La variance du k$^{\text{ième}}$ facteur est $\underline{a}_k' X'DX\underline{a}_k = \underline{u}_k' MX'DXM\underline{u}_k = \sum_{i=1}^n p_i \left\| \overrightarrow{GP_i^k} \right\|^2 = \lambda_k$; la qualité globale de représentation par projection sur le k$^{\text{ième}}$ axe principal, $\dfrac{\lambda_k}{\lambda_1 + \cdots + \lambda_p}$, est aussi appelée *pourcentage de variance expliquée par le* $k^{\text{ième}}$ *facteur.*■

3.3 Deuxième interprétation statistique de l'ACP

On note $V = X'DX$ la matrice des covariances de $x^1, \cdots, x^p$. On la suppose définie positive (ceci est réalisé si et seulement si il n'existe pas de relation linéaire entre les caractères $x^1, \cdots, x^p$ centrés).

Théorème *Le vecteur* $\underline{b}_k = \dfrac{\underline{a}_k}{\sqrt{\lambda_k}}$ *de* $\mathbb{R}^{p*}$ *est solution du problème de la recherche de* $\underline{b} \in \mathbb{R}^{p*}$ *tel que* $\underline{b}'VMV\underline{b}$ *soit maximal sous les contraintes* $\underline{b}'V\underline{b}_j = 0, j = 1, \cdots, k-1$, $\underline{b}'V\underline{b} = 1$.

Démonstration
Soit $A = VMV$. On cherche $\underline{b}$ tel que $\underline{b}'A\underline{b}$ soit maximal sous les contraintes $\underline{b}'V\underline{b}_j = 0$, $j = 1, \cdots, k-1$ et $\underline{b}'V\underline{b} = 1$.
Toute solution de ce problème est vecteur propre de $V^{-1}A = MV$ associé à la k$^{\text{ième}}$ plus grande valeur propre, comme $\underline{a}_k$.
Mais, $\underline{a}_k' V\underline{a}_k = \underline{a}_k' X'DX\underline{a}_k = \lambda_k$ (variance du k$^{\text{ième}}$ facteur) ; on peut prendre $\underline{b}_k = \dfrac{\underline{a}_k}{\sqrt{\lambda_k}}$.■

Cas où la métrique M est diagonale
On note M_{jj} le j^{ième} élément diagonal de M et $\underline{e}_j$ le j^{ième} vecteur de la base canonique de $\mathbb{R}^p$.

$$\underline{e}_j \underline{e}_j' = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} (0 \cdots 1 \cdots 0) = \begin{pmatrix} 0 & & & & \\ & \ddots & & & 0 \\ & & 1 & & \\ 0 & & & \ddots & \\ & & & & 0 \end{pmatrix} \longleftarrow j$$

$$M = \begin{pmatrix} M_{11} & & & & \\ & \ddots & & & 0 \\ & & \ddots & & \\ 0 & & & \ddots & \\ & & & & M_{pp} \end{pmatrix} = \sum_{j=1}^{p} M_{jj} \underline{e}_j \underline{e}_j'$$

$$\begin{aligned}
\underline{b}_k' V M V \underline{b}_k &= \underline{b}_k' V \sum_{j=1}^{p} M_{jj}\, \underline{e}_j \underline{e}_j' V \underline{b}_k \\
&= \sum_{j=1}^{p} M_{jj}\, (\underline{b}_k' V \underline{e}_j)^2 \\
\underline{b}_k' V \underline{e}_j &= \frac{1}{\sqrt{\lambda_k}}\, \underline{a}_k' V \underline{e}_j = \frac{1}{\sqrt{\lambda_k}}\, \underline{a}_k' X' D X \underline{e}_j = \frac{1}{\sqrt{\lambda_k}} (X \underline{a}_k)' D X \underline{e}_j
\end{aligned}$$

$X\underline{a}_k$ est le vecteur des valeurs du k$^{\text{ième}}$ facteur pour les n individus;

$$X\underline{e}_j = (\underline{x}^1 \cdots \underline{x}^j \cdots \underline{x}^p) \begin{pmatrix} 0 \\ 1 \\ \\ 0 \end{pmatrix} = \underline{x}^j.$$

$(X\underline{a}_k)' D (X\underline{e}_j)$ est la covariance du k$^{\text{ième}}$ facteur et du caractère x^j ; $\frac{1}{\sqrt{\lambda_k}} (X\underline{a}_k)' D (X\underline{e}_j)$ est la covariance du k$^{\text{ième}}$ facteur réduit, noté v^k, et du caractère x^j, $s_{v^k x^j}$.

Donc, $\underline{b}_k' V M V \underline{b}_k = \sum_{j=1}^{p} M_{jj}\, s^2_{v^k x^j}$.

On en déduit la proposition suivante :

Proposition *Dans le cas où la métrique M est diagonale, le k$^{\text{ième}}$ facteur réduit est une combinaison linéaire v^k des caractères $x^1, x^2, \cdots, x^p$ centrés qui rend maximale $\sum_{j=1}^{p} M_{jj} s^2_{vx^j}$ sous les contraintes d'être non corrélée aux facteurs précédents.*

Cas de l'*ACP* normée

Dans ce cas, on a :

$M_{jj} = \frac{1}{(s^j)^2}$; $\sum_{j=1}^{p} M_{jj}\, s^2_{v^k x^j} = \sum_{j=1}^{p} \frac{1}{(s^j)^2}\, s^2_{v^k x^j} = \sum_{j=1}^{p} r^2_{v^k x^j}$.

On en déduit la proposition :

Proposition *Le k$^{\text{ième}}$ facteur réduit de l'ACP normée est une combinaison linéaire v^k des caractères $x^1, \cdots, x^p$ centrés qui rend maximale $\sum_{j=1}^{p} r^2_{vx^j}$ sous les contraintes d'être non corrélée aux facteurs précédents.*

4 Etude dans l'espace des caractères $\mathbb{R}^n$

4.1 Structure euclidienne de $\mathbb{R}^n$

Dans $\mathbb{R}^n$, on représente chaque caractère x^j par le point B^j de coordonnées $(x^j_1, \cdots, x^j_n)$: $B^j(\underline{x}^j)$.

On munit $\mathbb{R}^n$ de la métrique des poids des individus :

$$D = \begin{pmatrix} p_1 & & & \\ & \ddots & & 0 \\ & & \ddots & \\ 0 & & & p_n \end{pmatrix}.$$

Alors :

a) $< \underline{x}^j, \underline{x}^k >= (\underline{x}^j)'\, D\, \underline{x}^k = (x_1^j .....x_n^j) \begin{pmatrix} p_1 & & & \\ & \ddots & & 0 \\ & & \ddots & \\ 0 & & & p_n \end{pmatrix} \begin{pmatrix} x_1^k \\ \vdots \\ \vdots \\ x_n^k \end{pmatrix}$

$= \sum_{i=1}^n p_i x_i^j x_i^k$ est la covariance de x^j et de x^k ;

b) $||\underline{x}^j||$ est l'écart-type de $\underline{x}^j$;

c) $\cos\,(\underline{x}^j, \underline{x}^k) = \dfrac{< \underline{x}^j, \underline{x}^k >}{||\underline{x}^j||\;||\underline{x}^k||}$ est le coefficient de corrélation linéaire de x^j et x^k.

Remarque Le caractère x^j est centré :

$$\sum_{i=1}^n p_i x_i^j = 0 =< \underline{x}^j, \underline{u} > \quad \text{avec } \underline{u} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Le caractère x^j est centré si et seulement si le vecteur $\underline{x}^j$ appartient à l'orthogonal à la droite $\Delta\underline{u}$ engendrée par $\underline{u}$.■

4.2 Les composantes principales

La valeur du $k^{ième}$ facteur pour l'individu i est, par définition, $\underline{x}_i' M \underline{u}_k = \underline{x}_i' \underline{a}_k$.

Définition *Le vecteur $\underline{c}^k$ des valeurs du $k^{ième}$ facteur pour les n individus est appelé la $k^{ième}$ composante principale.*

Expression

$$\underline{c}^k = \underset{(n,1)}{\begin{pmatrix} \underline{x}_1' \underline{a}_k \\ \underline{x}_2' \underline{a}_k \\ \vdots \\ \underline{x}_n' \underline{a}_k \end{pmatrix}} = \underset{(n,p)}{\begin{pmatrix} \underline{x}_1' \\ \underline{x}_2' \\ \vdots \\ \underline{x}_n' \end{pmatrix}} \underline{a}_k = X\underline{a}_k = XM\underline{u}_k.$$

On définit l'application :

$$\begin{aligned} X \quad &: \quad \mathbb{R}^{p*} \longrightarrow \mathbb{R}^n \\ X \quad &: \quad \underline{a}_k \longmapsto X\underline{a}_k \end{aligned}$$

Proposition *$\underline{c}^k$ est un vecteur propre de la matrice $XMX'D$ associé à la $k^{ième}$ plus grande valeur propre λ_k ; on a $||\underline{c}^k|| = \sqrt{\lambda_k}$.*

Démonstration

· $||\underline{c}^k||$ est l'écart-type du $k^{ième}$ facteur, égal à $\sqrt{\lambda_k}$.

· Montrons que les matrices $MX'DX$ et $XMX'D$ ont même ensemble de valeurs propres non nulles. Supposons $\lambda_k \neq 0$:

$$\begin{aligned} &\text{a)} \quad MX'DX\,\underline{a}_k = \lambda_k \underline{a}_k \Longrightarrow XMX'D(X\underline{a}_k) = \lambda_k (X\underline{a}^k), \quad X\underline{a}^k \neq \underline{0} \\ \Longleftrightarrow \quad &XMX'D\,\underline{c}^k = \lambda_k \underline{c}^k \\ &\text{b)} \quad XMX'D\,\underline{c}^k = \lambda_k\,\underline{c}^k \Longrightarrow MX'DX(MX'D\,\underline{c}^k) = \lambda_k\,(MX'D\,\underline{c}^k), \quad MX'D\,\underline{c}^k \neq \underline{0}. \end{aligned}$$

Toute valeur propre non nulle de l'une est valeur propre de l'autre ; ces deux matrices ont donc même ensemble de valeurs propres non nulles. En outre, $\underline{c}^k$ est vecteur propre de $XMX'D$ associé à la valeur propre λ_k.■

4.3 ACP non centrée des points B^j

Proposition $\underline{v}^k = \frac{1}{\sqrt{\lambda_k}}\,\underline{c}^k = \frac{1}{\sqrt{\lambda_k}}\,X\,\underline{a}_k$ *est solution du problème de la recherche de* $\underline{v} \in R^n$ *tel que* $\underline{v}'DXMX'D\,\underline{v}$ *soit maximal sous les contraintes* $\underline{v}'D\underline{v}^j = 0, \quad j = 1, \cdots, k-1$ *et* $\underline{v}'D\underline{v} = 1$.

Démonstration
Notons $A = DXMX'D$. La matrice A est symétrique.
On cherche $\underline{v}$ tel que $\underline{v}'A\underline{v}$ soit maximal sous les contraintes $\underline{v}'D\underline{v}^j = 0, \quad j = 1, \cdots, k-1$ et $\underline{v}'D\underline{v} = 1$.
On sait que toute solution de ce problème est vecteur propre de

$$D^{-1}A = D^{-1}(DXMX'D) = XMX'D$$

associé à la $k^{ième}$ plus grande valeur propre λ_k, comme $\underline{c}^k$.
Mais, $||\underline{c}^k||^2 = \lambda_k$ (variance du $k^{ième}$ facteur). Donc, $\frac{\underline{c}^k}{\sqrt{\lambda_k}}$ est solution de ce problème.■

Proposition *Lorsque* M *est diagonale, d'éléments diagonaux* M_{jj}, $\frac{1}{\sqrt{\lambda_k}}\,\underline{c}^k = \underline{v}^k$ *est un vecteur unitaire du* $k^{ième}$ *axe principal de l'ACP non centrée des points* $B^j(\underline{x}^j)$ *affectés des poids* M_{jj} *dans* (R^n, D).

Démonstration
On a établi que $\underline{v}^k = \frac{1}{\sqrt{\lambda_k}}\,\underline{c}^k$ est solution du problème de la recherche de $\underline{v} \in \mathbb{R}^n$ qui rend maximal

$$\underline{v}'DXMX'D\underline{v}$$

sous les contraintes

$$\begin{aligned} \underline{v}'D\underline{v}^j &= 0, \quad j = 1, \cdots, k-1, \\ \underline{v}'D\underline{v} &= 1. \end{aligned}$$

Or, dans l'ACP du tableau X dans $\mathbb{R}^p$ muni de la métrique M, la matrice des poids étant D, notée ACP (X, M, D), on recherche pour $k = 1, \cdots, r$, $\underline{u}_k \in \mathbb{R}^p$ qui rend maximal

$$\underline{u}'MX'DXM\underline{u}$$

sous les contraintes

$$\begin{aligned} \underline{u}'M\underline{u}_j &= 0, \quad j = 1, \cdots, k-1, \\ \underline{u}'M\underline{u} &= 1. \end{aligned}$$

Donc, on peut interpréter le premier problème comme étant l'ACP du tableau X' dans $\mathbb{R}^n$ muni de la métrique D, la matrice des poids étant M, notée ACP (X', D, M) ; $\underline{v}^k$ est alors un vecteur unitaire du $k^{ième}$ axe principal de cette ACP. On a :

$$X' = \begin{pmatrix} \underline{x}^{1'} & B^1 \\ \vdots & \vdots \\ \underline{x}^{p'} & B^p \end{pmatrix}$$

L'origine de $\mathbb{R}^n$ n'est pas le barycentre des points (B^j, M_{jj}). On dit alors que l'on fait une *ACP non centrée.*■

Pour l'ACP (X', D, M), on définit les applications :

$$\begin{aligned} D &: \mathbb{R}^n \longrightarrow \mathbb{R}^{n*} \\ D &: \underline{v}^k \longmapsto D\underline{v}^k \ (k^{ième} \text{ facteur de cette ACP}) \\ X' &: \mathbb{R}^{n*} \longrightarrow \mathbb{R}^p \\ X' &: D\underline{v}^k \longmapsto X'D\underline{v}^k \ (k^{ième} \text{ composante principale de cette ACP}) \end{aligned}$$

4.4 Le schéma de dualité

En rassemblant les définitions des applications, on construit le schéma de dualité de l'ACP :

$$\begin{array}{ccccc} (\underline{u}_k) & \mathbb{R}^p & \xleftarrow{X'} & \mathbb{R}^{n*} & \\ & \downarrow \mathrm{M} & & & \uparrow \mathrm{D} \\ (\underline{a}_k) & \mathbb{R}^{p*} & \xrightarrow{X} & \mathbb{R}^n & (\underline{c}^k) \end{array}$$

On lit sur ce schéma les données et les résultats principaux de l'ACP :
$\underline{u}_k$ est vecteur propre de $X'DXM$;
$\underline{a}_k = M\underline{u}_k$ est vecteur propre de $MX'DX$;
$\underline{c}^k = XM\underline{u}_k$ est vecteur propre de $XMX'D$.
On peut aussi retrouver facilement les critères de détermination de $\underline{u}_k, \underline{a}_k, \underline{c}^k$.

Remarque $X'D\underline{v}^k = \dfrac{1}{\sqrt{\lambda_k}}\, X'D\underline{c}^k = \dfrac{1}{\sqrt{\lambda_k}}\, X'DXM\underline{u}_k = \dfrac{1}{\sqrt{\lambda_k}}\, \lambda_k \underline{u}_k = \sqrt{\lambda_k}\, \underline{u}_k$.■

4.5 Analyse factorielle d'un tableau de distances

$\underline{c}^k$ est vecteur propre de $\underbrace{XMX'}_{=W} D$.

$$XMX' = \begin{pmatrix} \underline{x}'_1 \\ \vdots \\ \underline{x}'_n \end{pmatrix} \quad M\ (\underline{x}_1 \cdots \underline{x}_n)$$

L'élément (i,j) de W est $w_{ij} = \underline{x}'_i M \underline{x}_j =< \underline{x}_i, \underline{x}_j >$ dans $(\mathbb{R}^p, M)$.
Notons d_{ij} la distance entre les points A_i et A_j représentatifs des individus i et j dans $\mathbb{R}^p$:

$$\begin{aligned} d^2_{ij} &= ||\underline{x}_i - \underline{x}_j||^2 = ||\underline{x}_i||^2 - 2 < \underline{x}_i, \underline{x}_j > + ||x_j||^2 \\ &= w_{ii} - 2w_{ij} + w_{jj}. \end{aligned}$$

Notons $\overline{d^2_i} = \sum_{l=1}^{n} p_l\, d^2_{il}, \quad \overline{d^2} = \sum_{l=1}^{n}\sum_{m=1}^{n} p_l p_m d^2_{lm}$.

Proposition *On a* $w_{ij} = \dfrac{1}{2}(\overline{d^2_i} + \overline{d^2_j} - d^2_{ij} - \overline{d^2})$.

Démonstration
Soit $\mathcal{J}$ l'inertie de l'ensemble des points (A_i, p_i) par rapport au barycentre.

$$\begin{aligned} \mathcal{J} &= \sum_{i=1}^{n} p_i ||\underline{x}_i||^2 = \sum_{i=1}^{n} p_i\, w_{ii}. \\ \overline{d^2_i} &= \sum_{j=1}^{n} p_j d^2_{ij} = w_{ii} + \mathcal{J}, \text{ car } \sum_{1}^{n} p_j \underline{x}_j = \underline{0}; \\ \overline{d^2_j} &= w_{jj} + \mathcal{J} \quad ; \quad \overline{d^2} = \sum_{i=1}^{n} p_i \overline{d^2_i} = 2\mathcal{J}; \\ \overline{d^2_i} &+ \overline{d^2_j} - d^2_{ij} - \overline{d^2} = 2w_{ij}.\blacksquare. \end{aligned}$$

Donc, pour calculer la matrice $W = (w_{ij})$, il suffit de connaître les distances entre les points A_i.

Par conséquent, *si on donne le tableau des distances (euclidiennes) entre les points A_i, on peut construire la matrice W, donc WD, et on peut calculer les composantes principales. On peut alors faire les représentations graphiques des projections des points A_i sur les axes principaux.*

4.6 Le cercle des corrélations

Notons y^j le caractère réduit associé au caractère x^j : $y^j = \frac{x^j}{s^j}$.

Dans $\mathbb{R}^n$: $\underline{y}^j = \dfrac{\underline{x}^j}{||\underline{x}^j||}$; $||\underline{y}^j|| = 1$.

Notons $C^j(\underline{y}^j)$ le point de $\mathbb{R}^n$ représentatif du caractère y^j.

1) Dans $\mathbb{R}^n$, les points C^j sont situés sur la sphère de centre l'origine et de rayon 1. $\cos(\underline{x}^l, \underline{x}^m) = \cos(\underline{y}^l, \underline{y}^m)$ représente le coefficient de corrélation linéaire entre les caractères x^l et x^m.

Si ces caractères sont fortement corrélés positivement (cosinus proche de 1 $\Longleftrightarrow$ angle proche de 0), alors les points C^l et C^m sont proches.

Si ces caractères sont fortement corrélés négativement (cosinus proche de - 1 $\Longleftrightarrow$ angle proche de π), alors les points C^l et C^m sont presque diamétralement opposés.

2) Soit un plan passant par l'origine et engendré par les vecteurs $\underline{v}^{k_1} = \frac{\underline{c}^{k_1}}{\|\underline{c}^{k_1}\|}$ et $\underline{v}^{k_2} = \frac{\underline{c}^{k_2}}{\|\underline{c}^{k_2}\|}$ (dans le cas où M est diagonale, c'est un plan principal de l'ACP des points B^j).

L'intersection de la sphère de rayon 1 et de ce plan est un cercle de rayon 1 appelé *cercle des corrélations.*

Si, sur ce plan, les projections de C^l et de C^m sont proches du cercle, ces points sont bien représentés par projection.

Si, en outre, ces projections sont proches l'une de l'autre (respectivement presque diamétralement opposées), ceci traduit que les points C^l et C^m sont proches (respectivement presque diamétralement opposés), donc que les caractères x^l et x^m sont fortement corrélés positivement (respectivement négativement).

3) La projection de C^j sur l'axe $(0, \underline{v}^k)$ (on note O l'origine de $\mathbb{R}^n$) a pour abscisse $< \underline{y}^j, \underline{v}^k >$: c'est la covariance de y^j et de v^k, donc le coefficient de corrélation linéaire du caractère x^j et du $k^{ième}$ facteur.

Donc, dans le plan de projection $(0, \underline{v}^1, \underline{v}^2)$, on représente le caractère x^j par un point qui a pour coordonnées les coefficients de corrélation linéaire de ce caractère avec les deux premiers facteurs. On fait de même pour tout autre plan de projection $(0, \underline{v}^{k_1}, \underline{v}^{k_2})$.

On calcule ces coefficients de corrélation linéaire :

$$< \underline{y}^j, \underline{v}^k >= (\underline{y}^j)' D \underline{v}^k = \frac{1}{s^j} \frac{1}{\sqrt{\lambda_k}} (\underline{x}^j)' D \underline{c}^k = \frac{1}{s^j} \frac{1}{\sqrt{\lambda_k}} (\underline{x}^j)' DXM \underline{u}_k.$$

Or: $X' = \begin{pmatrix} \underline{x}^{1'} \\ \vdots \\ \underline{x}^{j'} \\ \vdots \\ \underline{x}^{p'} \end{pmatrix}$; $(\underline{x}^j)' DXM \underline{u}_k$ est la $j^{ième}$ ligne de $X'DXM\underline{u}_k = \lambda_k \underline{u}_k$.

Donc : $< \underline{y}^j, \underline{v}^k >= \sqrt{\lambda_k}\, \dfrac{u_k^j}{s^j}$.

Proposition *Le coefficient de corrélation linéaire du caractère x^j et du $k^{ième}$ facteur est égal à $\sqrt{\lambda_k}\, \dfrac{u_k^j}{s^j}$.*

Remarque Dans $\mathbb{R}^p$, la mesure algébrique de la projection de $\vec{e}_j$ sur le $k^{ième}$ axe principal $(G, \vec{u}_k)$ est égale, dans le cas où M est diagonale, à $M_{jj}\, u_k^j$.

Le rapport du coefficient de corrélation linéaire du caractère x^j et du $k^{ième}$ facteur et de cette mesure algébrique est égal à $\frac{\sqrt{\lambda_k}}{M_{jj}\, s^j}$.

Dans le cas de l'ACP normée, lorsque l'on prend des données centrées réduites et $M = I$, ce rapport est égal à $\sqrt{\lambda_k}$ et ne dépend pas de j. C'est pourquoi on superpose parfois dans ce cas la projection des points C^j et la projection des points A_i.■

5 Interprétation des résultats d'une ACP

On représente la projection des points-individus A_i et des axes-caractères $(G, \vec{e}_j)$ sur le plan principal $(G, \vec{u}_1, \vec{u}_2)$ et éventuellement sur d'autres plans principaux. On représente les cercles de corrélation correspondants.

On peut diviser le plan d'interprétation en quatre parties.

5.1 Axes principaux à conserver

On note la qualité globale de représentation sur chaque axe (appelée aussi pourcentage d'inertie expliquée par l'axe ou pourcentage de variance expliquée par le facteur correspondant). Pour décider quels axes conserver, on peut s'appuyer sur les trois principes suivants, qui peuvent être contradictoires :

1. On conserve tout axe que l'on peut interpréter.
2. On arrête l'introduction des axes lorsque la variance d'un facteur correspondant à un axe est inférieure à celle d'un caractère.

   Par exemple, en ACP normée, lorsqu'on utilise des caractères réduits, on arrête l'introduction lorsque la variance est plus petite que 1.
3. On arrête l'introduction lorsque la qualité de représentation sur un axe est beaucoup plus faible que celle du précédent.

   Exemple Axe 1 : 30 % - Axe 2 : 20 % - Axe 3 : 18 % ; on peut conserver l'axe 3.

   Axe 1 : 30 % - Axe 2 : 20 % - Axe 3 : 6 % ; on peut ne pas conserver l'axe 3.

5.2 Etude des corrélations entre les caractères initiaux

On étudie le cercle des corrélations et la matrice des corrélations. On détermine les caractères fortement corrélés positivement et ceux fortement corrélés négativement. On rappelle que :

- Lorsque deux caractères sont représentés par des *points proches du cercle et proches l'un de l'autre*, il existe une forte corrélation linéaire positive entre eux ;
- lorsque deux caractères sont représentés par des *points proches du cercle et presque diamétrialement opposés*, il existe une forte corrélation linéaire négative entre eux.

On lit les coefficients dans la matrice des corrélations.

Remarque Des caractères fortement corrélés positivement varient dans le même sens et peuvent représenter le même phénomène. Si l'on conserve trop de caractères fortement corrélés positivement, la différence entre deux individus, qui seraient semblables vis-à-vis de ces caractères, mais dissemblables vis-à-vis des autres, peut apparaître fortement atténuée. Pour remédier à ceci, on peut éliminer certains de ces caractères ou leur donner un poids qui diminue leur importance dans l'étude, ce qui revient à changer de métrique M. On peut aussi effectuer une ACP sur le tableau des mesures de ces caractères et remplacer l'ensemble de ces caractères par le premier facteur de cette ACP.■

5.3 Interprétation des facteurs

Que représentent les combinaisons linéaires des caractères initiaux centrés que l'on a déterminées ? On peut faire cette interprétation en s'appuyant sur l'une ou l'autre des deux analyses suivantes.

5.3.1 Analyse des corrélations des facteurs avec les caractères

On détermine les caractères fortement corrélés positivement d'une part, négativement d'autre part, avec les facteurs. Ceci peut permettre de donner une signification à un facteur, en ayant déterminé les caractères qui varient en moyenne dans le même sens que ce facteur et ceux qui varient en moyenne en sens contraire.

Exemple Dans une analyse où l'on a mesuré des variables économiques sur les pays du monde, le premier facteur peut être corrélé positivement avec des caractères tels que le PNB par habitant, les importations, les exportations, le pourcentage d'actifs dans le secteur tertiaire et corrélé négativement avec des variables concernant l'importance relative du secteur primaire (agriculture, sylviculture, pêche) ; l'axe principal correspondant oppose alors des pays très développés économiquement (qui ont des valeurs fortes pour les caractères du premier groupe et des valeurs faibles pour les caractères du deuxième) à des pays moins développés (qui sont dans le cas contraire) : on peut dire que le premier facteur représente le développement économique.■

Remarque *Cas d'un facteur de taille* : c'est un facteur tel que tous les caractères sont fortement corrélés positivement (ou tous négativement) avec lui. L'axe principal correspondant oppose les individus ayant de fortes valeurs pour tous les caractères à ceux ayant de faibles valeurs. Par exemple : on fait une étude sur les notes d'étudiants dans différentes matières ; il peut apparaître que certains ont de bonnes notes partout, alors que d'autres en ont de mauvaises partout ; le premier facteur peut alors être un facteur de taille, l'axe principal opposant les étudiants ayant de bonnes notes à ceux ayant de mauvaises notes.

On peut parfois remédier au phénomène de taille en *transformant des caractères*, en remplaçant dans ce cas des caractères de taille par des caractères de structure. Par exemple : on fait une étude sur les départements français ; le caractère "nombre de médecins" dépend de la population du département : on peut l'interpréter comme un caractère de taille ; on peut alors le remplacer par le caractère "nombre de médecins pour 1 000 habitants" (ou "densité médicale"), qui est un caractère de structure.■

5.3.2 Analyse de la position d'invidus ayant une contribution importante

Si certains individus ont une contribution très importante à l'inertie expliquée par un axe principal et un type bien défini, l'étude de leur position par rapport à cet axe peut permettre de donner une signification au facteur correspondant. Certains individus peuvent parfois expliquer à eux seuls la construction d'un axe principal.

Remarque Certains individus ayant une contribution très importante peuvent être considérés, après examen des valeurs des caractères, comme aberrants. On peut alors les éliminer de la détermination des axes principaux, et cependant faire figurer leur projection sur les plans (*individus supplémentaires ou non actifs*).■

Exemple On fait une étude sur les départements français. Il peut apparaître que les départements d'outre-mer ont des valeurs de certains caractères très différentes de celles des autres. Après avoir étudié leurs caractéristiques, on peut dans ce cas limiter éventuellement l'analyse aux départements de la métropole.■

Une analyse de données se fera donc généralement en plusieurs étapes : après un premier passage, on interprète les résultats et on élimine éventuellement des caractères ou des individus ; on effectue alors un deuxième passage et on reprend le même procédé.

5.4 Interprétation de la position des projections des points-individus

On interprète la position par rapport aux axes principaux de la projection des points représentatifs des individus.

Si un individu a une abscisse positive (respectivement négative) élevée en valeur absolue sur un axe principal, c'est qu'il a une forte (respectivement faible) valeur du facteur correspondant et, en général, une forte (respectivement faible) valeur des caractères fortement corrélés positivement avec ce facteur et une faible (respectivement forte) valeur des caractères fortement corrélés négativement avec ce facteur. On confirme ceci en consultant le tableau des données.

La proximité entre deux individus sur un plan principal ne traduit pas nécessairement une ressemblance entre ces individus vis-à-vis de tous les caractères, en particulier si la qualité de représentation de ces individus sur le plan considéré est faible, mais traduit une ressemblance vis-à-vis des facteurs correspondants.

Si le nombre d'individus est trop important, ou si les individus constituent une population ou un échantillon représentatif de cette population pour laquelle l'étude des résultats individuels ne présente pas d'intérêt, alors on peut regrouper les individus en *classes*

- empiriquement, à partir des graphiques de l'ACP ;
- automatiquement, par une méthode de *classification*.

On interprète alors la position des classes par rapport aux axes principaux.

Exemple Dans une banque, on peut étudier la population des clients et vouloir segmenter cet ensemble en classes dont on étudie les caractèristiques globales sans étudier celles de chaque client.■

Remarque Dans le logiciel SPAD, la classification est faite non pas à partir des valeurs des caractères, mais à partir des valeurs des facteurs. L'utilisateur déclare le nombre de facteurs qu'il souhaite introduire. Après la classification, on peut représenter automatiquement les classes sur les plans d'axes principaux, le code d'un individu étant remplacé sur la figure par le numéro de la classe.■

Remarque *Représentation d'un caractère qualitatif* : on peut représenter un caractère qualitatif sur un plan principal. On représente chaque modalité du caractère par un point qui est le barycentre des projections sur ce plan des points représentatifs des individus ayant cette modalité. Si ce caractère est ordinal, on peut joindre ces points par des flèches orientées en respectant l'ordre des modalités.■