Université Paris-Dauphine
M1
Année 2017

Analyse des données

Patrice Bertrand et Denis Pasquignon

Table des matières

Introduction

L'analyse des données (AD), et plus généralement la fouille des données (FD), est constituée d'un ensemble de techniques qui ont pour but de déterminer les structures possédées par l'ensemble des données. Ces structures peuvent être de nature descriptive (partition, hiérarchie, plan factoriel,...) ou explicative (arbre de décision, analyse factorielle discriminante,...). L'analyse de données peut être considérée comme une science expérimentale : propriétés démontrées après avoir été observées, indice empirique pour l'interprétation des résultats, codages établis de façon heuristique.

Par ailleurs, les premiers résultats fournis par une analyse factorielle sont généralement évidents, alors que les résultats suivants ne sont pas triviaux et sont souvent intéressants.

Les données peuvent se présenter sous différentes formes : tableaux individus $\times$ variables (dans un but descriptif, l'interprétation établira des liens entre variables et groupes d'individus qui se ressemblent selon ces variables), tableaux de distances (représentation des individus dans un plan, sur une droite, etc ou partitionement de l'ensemble des individus), tableaux de contingence (ces tableaux croisent les ensembles de modalités de deux caractères qualitatifs), tableaux de présence-absence (0/1), tableaux de notes, tableaux de pourcentage...

Les techniques d'analyse de données se différencient non seulement par les outils mathématiques utilisés (algèbre linéaire dans le cas de l'analyse factorielle, théorie des graphes et combinatoire pour certaines méthodes de classification) mais aussi par les buts poursuivis qui peuvent être un but descriptif ou un but prévisionnel. Le but descriptif consiste à essayer d'obtenir une représentation simplifiée aussi proche que possible des données initiales, le but prévisionnel consiste à expliquer et prévoir une ou plusieurs variables en fonction d'autres variables. Dans ce cours, nous présenterons les techniques suivantes :

— Analyse en composantes principales (ACP) : rechercher des axes d'inertie d'un système de points affectés de poids, ce qui permet d'en déduire des sous-espaces de dimensions réduites sur lesquels la projection des points est la moins déformante.

— Analyse des correspondances (AC) : double ACP ayant un but à la fois descriptif et prévisionnel (étude de liens existants entre lignes et colonnes d'un tableau).

Chapitre 1

Nuages de points

1.1 Tableau de données

On observe p variables quantitatives mesurées sur un échantillon de taille n. Les données sont rassemblées en un tableau ou matrice de n lignes et p colonnes. On note X ce tableau de données, son terme général x_i^j, situé à la $i$ème et $j$ème colonne, désigne la valeur prise par le $i$ème individu pour la variable j.

On note $I = [\![1, n]\!]$ et $J = [\![1, p]\!]$ qui sont les ensembles d'indices désignant respectivement les n individus et les p variables.

$$X = (x_i^j)_{i \in I, j \in J} \in \mathcal{M}_{n,p}(\mathbb{R}).$$

Ainsi les valeurs prises par la variable x^j pour les n individus se lisent sur la $j$ème colonne et les valeurs prise par l'individu i pour les p variables se lisent sur la $i$ème ligne. On note x^j la $j$ème variable et x_i le $i$ème individu :

$$\forall (i,j) \in [\![1, n]\!] \times [\![1, p]\!], \;\; x^j = \begin{pmatrix} x_1^j \\ \vdots \\ x_n^j \end{pmatrix} \in \mathbb{R}^n \text{ et } x_i = \begin{pmatrix} x_i^1 \\ \vdots \\ x_i^p \end{pmatrix} \in \mathbb{R}^p.$$

Ainsi

$$X = [x^1, \cdots, x^p] = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}.$$

1.2 Nuages des individus et nuages des variables

On munit $\mathbb{R}^p$ de la base canonique, O étant l'origine de ce repère, on peut alors associé à chaque individu i le point M_i tel que

$$\forall i \in [\![1, n]\!], \;\; \overrightarrow{OM_i} = x_i.$$

Chaque axe représente une variable. L'ensemble des points $\mathcal{M}_X = \{M_i, \; 1 \leq i \leq n\}$ est appelé le nuage des individus et $\mathbb{R}^p$ est l'espace des individus.

De même, on munit $\mathbb{R}^n$ de la base canonique, on peut alors associé à chaque variable le point N^j tel que

$$\forall j \in [\![1, p]\!], \;\; \overrightarrow{ON^j} = x^j.$$

Chaque axe représente un individu. L'ensemble des points $\mathcal{N}_X = \{N^j,\ 1 \leq j \leq p\}$ est appelé le nuage des variables et $I\!R^n$ est l'espace des variables.

Les ensembles R^n et $I\!R^p$ sont considérés comme des espaces affines. Dans l'annexe A, on rappelle les principales notions à connaitre pour ce cours.

1.3 Centre de gravité du nuage $\mathcal{M}_X$

Chaque individu i est muni d'une masse, appelée aussi poids, notée p_i et telle que

$$\forall i \in I,\ \ p_i > 0 \text{ et } \sum_{i=1}^{n} p_i = 1.$$

On note D_p la matrice diagonale définie par

$$D_p = diag(p_1, \cdots, p_n).$$

En général, les poids sont tous égaux à $1/n$, mais ce n'est pas toujours le cas comme par exemple en Analyse des Correspondances.

Le centre de gravité du nuage des individus M_i affecté du poids p_i est le point G tel que

$$G = \sum_{i=1}^{n} p_i M_i.$$

La $j$ème coordonnée de G est donnée par

$$g_j = \sum_{i=1}^{n} p_i x_i^j = \overline{x^j}.$$

Ainsi g_j est la moyenne de la variable x^j et les coordonnées du point G sont les p moyennes des p variables.

Proposition 1.3.1 *On note 1_n le vecteur de $I\!R^n$ dont toutes les coordonnées sont égales à 1, on a*

$$g = \overrightarrow{OG} = \begin{pmatrix} g_1 \\ \vdots \\ g_p \end{pmatrix} = X' D_p 1_n.$$

preuve : On remarque que

$$\forall j \in [\![1, p]\!],\ \ g_j = {x^j}' D_p 1_n,$$

ce qui donne le résultat.

■

On en déduit que g_j est l'abscisse de la projection orthogonale pour la métrique D_p de x^j sur Vect(1_n).

Il est naturel de centrer le nuage des individus sur le centre de gravité G ce qui revient à construire un nouveau tableau Y tel que

$$\forall (i,j) \in [\![1, n]\!] \times [\![1, p]\!],\ \ y_i^j = x_i^j - \overline{x^j},$$

soit

$$\forall i \in [\![1, n]\!], \quad y_i = M_i - G.$$

Ainsi dans ce nouveau tableau de données, toutes les variables y^j, $1 \leq j \leq p$, sont de moyennes nulles.

Proposition 1.3.2 *On a*

$$Y = X - 1_n g'$$

Par ailleurs

$$y^j = x^j - g_j 1_n = (Id - P_{\text{Vect}(1_n)})(x^j),$$

ce qui signifie que y^j est la projection de x^j sur l'hyperplan orthogonal à 1_n.

1.4 Support des nuages

Définition 1.4.1 *On appelle support d'un nuage le plus petit sous-espace affine contenant les points du nuage. On note*

$$S_X = supp(\mathcal{M}_X) \text{ et } S_Y = supp(\mathcal{M}_Y).$$

Puisque le nuage $\mathcal{M}_Y$ est centré, le support S_Y contient l'origine et est assimilé à un sous-espace vectoriel

$$S_Y = \text{Vect}(y_1, \cdots, y_n) = \text{Im}\, Y'.$$

On en déduit que la dimension de S_Y est égale au rang de Y.

1.5 Métriques sur $I\!R^p$

Pour étudier la proximité entre deux individus d'un même nuage de points, on introduit une distance notée d entre les individus i et i' égale à la norme du vecteur joignant ces deux points :

$$d(i, i') = \|\overrightarrow{M_i M_{i'}}\|.$$

Cette norme provient d'un produit scalaire sur l'espace vectoriel $I\!R^p$. Etant donnée la base canonique $(e_1, \cdots, e_p)$ de $I\!R^p$, le produit scalaire est caractérisée par la donnée d'une matrice carrée d'ordre p dont le terme courant est pour tout $1 \leq i, j \leq p$

$$m_{i,j} = M(e_i, e_j) =< e_i, e_j >_M = e_i' M e_j = e_j' M e_i.$$

Cette matrice M est symétrique, définie et positive. Réciproquement, toute matrice d'ordre p symétrique, définie positive permet de définir un produit scalaire dans $I\!R^p$. Cette matrice définit une métrique de l'espace $I\!R^p$.

Dans la suite, nous noterons M la métrique de l'espace $I\!R^p$ et l'espace des individus $I\!R^n$ est muni de la métrique D_p.

Si l'on suppose que la matrice M est diagonale $M = diag(m_1, \cdots, m_p)$, alors

$$d(y_i, y_{i'}) = \sqrt{\sum_{j=1}^{p} m_j (y_i^j - y_{i'}^j)^2}.$$

De même, la distance entre deux variables y^j et $y^{j'}$ est donnée par

$$d(y^j, y^{j'}) = \sqrt{\sum_{i=1}^{n} p_i (y_i^j - y_i^{j'})^2}.$$

1.6 Matrice Variance

Par définition, la matrice variance, notée V, des p variables pour les n individus est une matrice carré d'ordre p et de terme courant $v_{j,j'}$ donné par

$$\forall (j,j') \in [\![1,p]\!], \;\; v_{j,j'} = \mathrm{Cov}(x^j, x^{j'}) = \sum_{i=1}^{n} p_i (x_i^j - g_j)(x_i^{j'} - g_{j'}) =< y^j, y^{j'} >_{D_p} .$$

Proposition 1.6.1 *En notation matricielle, on a*

$$V = Y' D_p Y = (X - 1_n g')' D_p (X - 1_n g') = X' D_p X - g g'.$$

Remarque 1.6.2 *Si la matrice V est définie positive, elle fournit une métrique sur $I\!R^p$, métrique induite par D_p et Y. Si V n'est pas régulière, on aura seulement une pseudo métrique.*

Proposition 1.6.3 *Soient $u = \begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}$ et $v = \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix}$ de $I\!R^p$, on définit deux nouvelles variables z et t par*

$$z = \sum_{j=1}^{p} u_j x^j \;\; \text{et } t = \sum_{j=1}^{p} v_j x^j.$$

Alors la covariance entre z et t est donnée par

$$\mathrm{Cov}(z,t) = u' V v.$$

Si la matrice V est définie positive, V définit une métrique pour laquelle la covariance entre z et t est le produit scalaire entre les vecteurs z et t et la variance de la variable z est le carré de la norme de z soit

$$\mathrm{Cov}(z,t) =< u, v >_V \;\; \text{et } \; \mathrm{V}(z) = \|u\|_V^2.$$

preuve en TD

1.7 Effet d'une transformation linéaire A du nuage des individus

Soit A une matrice carrée d'ordre p. On note

$$\mathcal{M}_Z = \{z_1, \cdots, z_n\} \text{ avec } \forall i \in [\![1,n]\!], \;\; z_i = A x_i.$$

On obtient ainsi une nouvelle matrice Z dont les lignes sont les $z_1, \cdots, z_n$ soit

$$Z' = A X' \text{ donc } Z = X A'.$$

Proposition 1.7.1 *Le centre de gravité de $\mathcal{M}_Z$ affectés des poids $p_1, \cdots, p_n$ noté g_Z est*

$$g_Z = Ag.$$

La matrice de variance de Z noté $Var(Z)$ est

$$Var(Z) = Var(XA') = AVA'.$$

preuve : On a

$$g_Z = \sum_{i=1}^{n} p_i z_i = \sum_{i=1}^{n} p_i A x_i = Ag.$$

On note Z_c la matrice centrée

$$Z_c = Z - 1_n g_Z' = XA' - 1_n g' A' = YA',$$

donc

$$Var(Z) = Z_c' D_p Z_c = AY' D_p Y A' = AVA'.$$

■

1.8 Inerties

Inertie par rapport à un point

Définition 1.8.1 *Soit A un point, l'inertie du nuage $\mathcal{M} = (x_i)_{1 \le i \le n}$ par rapport au point A est*

$$I_A(\mathcal{M}) = \sum_{i=1}^{n} p_i \|x_i - A\|_M^2.$$

Si $A = G$ le centre de gravité , $I_G(\mathcal{M})$ est appelée inertie totale du nuage :

$$I_T(\mathcal{M}) = I_G(\mathcal{M}).$$

Si l'on suppose que $M = diag(m_1, \cdots, m_n)$ alors

$$I_T(\mathcal{M}) = \sum_{i=1}^{n} p_i \|y_i\|_M^2 = \sum_{i=1}^{n} p_i \sum_{j=1}^{p} m_j (y_i^j)^2 = \sum_{j=1}^{p} m_j \, \mathrm{V}(y^j),$$

où $\mathrm{V}(y^j)$ représente la variance de y^j. L'inertie totale est ainsi la somme pondérée des variances des variables initiales, elle mesure la dispersion du nuage autour du centre de gravité.

Proposition 1.8.2 théorème de Huyghens *On a*

$$I_A(\mathcal{M}) = I_T(\mathcal{M}) + \|A - G\|_M^2.$$

preuve :

$$\begin{aligned}\|x_i - A\|_M^2 &= \|x_i - G + G - A\|_M^2,\\ &= \|x_i - G\|_M^2 + \|G - A\|_M^2 + 2 < x_i - G, G - A >_M .\end{aligned}$$

On en déduit que

$$\begin{aligned}I_A(\mathcal{M}) &= \sum_{i=1}^n p_i \|x_i - A\|_M^2,\\ &= \sum_{i=1}^n p_i \|x_i - G\|_M^2 + \sum_{i=1}^n p_i \|G - A\|_M^2 + 2 < \sum_{i=1}^n p_i (x_i - G), G - A >_M,\\ &= I_T(\mathcal{M}) + \|G - A\|_M^2.\end{aligned}$$

■

Inertie par rapport à un sous-espace affine

Définition 1.8.3 *Soit $\mathcal{E}$ un sous-espace affine de $\mathbb{R}^n$ et E le sous-espace vectoriel associé muni de la métrique M. Soit A un point de $\mathcal{E}$ et B un point de $\mathbb{R}^n$, la distance de B à $\mathcal{E}$ est*

$$d_M(B, \mathcal{E}) = \|(Id - P_E)(\overrightarrow{AB})\|_M,$$

où P_E est la projection orthogonale sur E.
On appelle inertie du nuage $\mathcal{M} = (M_i)_{1\leq i\leq n}$ par rapport au sous-espace affine $\mathcal{E}$

$$I_{\mathcal{E}}(\mathcal{M}) = \sum_{i=1}^n p_i d_M^2(M_i, \mathcal{E}).$$

Remarque 1.8.4 *La définition de $d_M(B, \mathcal{E})$ ne dépend pas du point A de $\mathcal{E}$.*

Proposition 1.8.5 *On a*

$$I_{\mathcal{E}}(\mathcal{M}) = I_{\mathcal{E}_G}(\mathcal{M}) + \|(Id - p_E)(\overrightarrow{AG})\|_M^2,$$

où $\mathcal{E}_G$ est le sous-espace affine passant par G de direction E et A un point de $\mathcal{E}$.

preuve :

$$\begin{aligned}I_{\mathcal{E}}(\mathcal{M}) &= \sum_{i=1}^n p_i d_M^2(M_i, \mathcal{E}),\\ &= \sum_{i=1}^n p_i \|(Id - P_E)(\overrightarrow{AM_i})\|^2,\\ &= \sum_{i=1}^n p_i \|(Id - P_E)(\overrightarrow{AG} + \overrightarrow{GM_i})\|^2,\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{n} p_i \|(Id - P_E)(\overrightarrow{AG})\|^2 + \sum_{i=1}^{n} p_i \|(Id - P_E)(\overrightarrow{GM_i}))\|^2 \\
&\quad +2 < (Id - P_E)(\overrightarrow{AG}), \sum_{i=1}^{n} p_i (Id - P_E)(\overrightarrow{GM_i}) >_M, \\
&= \|(Id - P_E)(\overrightarrow{AG})\|^2 + I_{\mathcal{E}_G}(\mathcal{M}).
\end{aligned}$$

■

Ce résultat montre que parmi tous les sous-espaces affine parallèles à E, celui qui possède une inertie minimale est celui qui passe par le centre de gravité du nuage.

Par la suite, on recherche le ou les sous-espaces affine de dimension k donnée par rapport auquel(s) le nuage a une inertie minimale : c'est l'objectif de l'ACP.

On voit donc que ces sous-espaces optimaux passent nécessairement par G. C'est la raison pour laquelle on supposera, en général, par la suite que le tableau X est centré. Si ce n'est pas le cas, on raisonnera sur Y.

Proposition 1.8.6 *On note $\mathcal{E}^\perp$ le sous espace affine passant par G et de direction $E^\perp$, on a*

$$I_T = I_{\mathcal{E}}(\mathcal{M}) + I_{\mathcal{E}^\perp}(\mathcal{M}).$$

On pose

$$J_{\mathcal{E}}(\mathcal{M}) = I_{\mathcal{E}^\perp}(\mathcal{M}).$$

$J_{\mathcal{E}}(\mathcal{M})$ est l'inertie totale de la projection de $\mathcal{M}$ sur $\mathcal{E}$

preuve : On a la relation $P_E + P_{E^\perp} = Id$, d'où en utilisant Pythagore

$$I_{\mathcal{E}}(\mathcal{M}) + I_{\mathcal{E}^\perp}(\mathcal{M}) = \sum_{i=1}^{n} p_i \|(Id - P_E)(\overrightarrow{GM_i})\|^2 + \sum_{i=1}^{n} p_i \|(Id - P_{E^\perp})(\overrightarrow{GM_i})\|^2 = I_T.$$

Pour le dernier point, il suffit d'appliquer la définition :

$$J_{\mathcal{E}}(\mathcal{M}) = \sum_{i=1}^{n} p_i \|(Id - P_E^\perp)(\overrightarrow{GM_i})\|^2 = \sum_{i=1}^{n} p_i \|(P_E)(\overrightarrow{GM_i})\|^2.$$

■

Ainsi la recherche de $\mathcal{E}$ qui minimise $I_{\mathcal{E}}(\mathcal{M})$ est équivalent à rechercher $\mathcal{E}$ qui maximise $J_{\mathcal{E}}(\mathcal{M})$.

Inertie par rapport à une droite affine passant par G

Soit ϵ_1 un vecteur unitaire pour la métrique M de $I\!R^p$. Soit $\mathcal{E}_1$ la droite affine passant par G associée à Vect(ϵ_1). Puisque pour tout vecteur u de $I\!R^p$, on a

$$P_{\text{Vect}(\epsilon_1)}(u) =< u, \epsilon_1 >_M \epsilon_1.$$

On en déduit que

$$J_{\mathcal{E}_1}(\mathcal{M}) \quad = \quad \sum_{i=1}^{n} p_i \|(P_E)(\overrightarrow{GM_i})\|^2,$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} p_i < y_i, \epsilon_1 >_M^2, \\
&= \sum_{i=1}^{n} p_i \epsilon_1' M y_i y_i' M \epsilon_1, \\
&= \epsilon_1' M V M \epsilon_1.
\end{aligned}
$$

Décomposition de l'inertie

Proposition 1.8.7 *Soit* $(\epsilon_1, \cdots, \epsilon_k)$ *une base orthonormale de* E *pour la métrique* M*, on complète cette base en une base orthonormale de* $\mathbb{R}^p$ *soit* $(\epsilon_1, \cdots, \epsilon_k, \epsilon_{k+1}, \cdots, \epsilon_p)$ *une base orthonormale de* $\mathbb{R}^p$*. On a*

$$I_{\mathcal{E}_G}(\mathcal{M}) = \sum_{l=k+1}^{p} J_{\mathcal{E}_l}(\mathcal{M}),$$

où $\mathcal{E}_l$ *est la droite affine passant par* G *de direction* $\mathrm{Vect}(\epsilon_l)$.

preuve : On a pour tout vecteur u

$$P_E(u) = \sum_{i=1}^{k} < u, \epsilon_i >_M \epsilon_i,$$

on en déduit que pour i fixé

$$\|(Id - P_E)(\overrightarrow{GM_i})\|^2 = \| \sum_{l=k+1}^{p} < \overrightarrow{GM_i}, \epsilon_l >_M \epsilon_l\|^2,$$

ce qui donne

$$\|(Id - P_E)(\overrightarrow{GM_i})\|^2 = \sum_{l=k+1}^{p} < \overrightarrow{GM_i}, \epsilon_l >_M^2,$$

et matriciellement

$$\|(Id - P_E)(\overrightarrow{GM_i})\|^2 = \sum_{l=k+1}^{p} \epsilon_l' M y_i y_i' M \epsilon_l.$$

Par conséquent on obtient par interversion de somme

$$
\begin{aligned}
I_{\mathcal{E}_G}(\mathcal{M}) &= \sum_{i=1}^{n} p_i \sum_{l=k+1}^{p} \epsilon_l' M y_i y_i' M \epsilon_l, \\
&= \sum_{l=k+1}^{p} \epsilon_l' M \sum_{i=1}^{n} p_i y_i y_i' M \epsilon_l, \\
&= \sum_{l=k+1}^{p} \epsilon_l' M V M \epsilon_l, \\
&= \sum_{l=k+1}^{p} J_{\mathcal{E}_l}(\mathcal{M}).
\end{aligned}
$$

∎

Calcul de l'inertie totale

Proposition 1.8.8 *On a*
$$I_T = tr(VM).$$

preuve : On choisit comme base orthonormale une base constitué de vecteurs propres de MV soit $(u_1, \cdots, u_p)$, on a

$$\begin{aligned}
I_T &= I_{(\mathbb{R}^p)^\perp}(\mathcal{M}), \\
&= \sum_{j=1}^{p} u_j' MVMu_j, \\
&= \sum_{j=1}^{p} \lambda_j \|u_j\|_M^2, \\
&= \sum_{j=1}^{p} \lambda_j, \\
&= tr(MV) = tr(VM).
\end{aligned}$$

On peut aussi raisonner directement : puisque la trace de AB est égal à la trace de BA, on en déduit

$$\begin{aligned}
I_T &= \sum_{i=1}^{n} p_i \|y_i\|_M^2, \\
&= \sum_{i=1}^{n} p_i y_i' M y_i, \\
&= tr(\sum_{i=1}^{n} p_i y_i' M y_i), \\
&= \sum_{i=1}^{n} p_i tr(y_i' M y_i), \\
&= \sum_{i=1}^{n} p_i tr(M y_i y_i'), \\
&= tr(M \sum_{i=1}^{n} p_j y_i y_i'), \\
&= tr(MV) = tr(VM).
\end{aligned}$$

■

Chapitre 2

Analyse en Composantes Principales

Soit $\mathcal{N} = \{x_i, i \in I\} \subset I\!R^p$ un nuage de points de l'espace $I\!R^p$ muni de la métrique M. Chaque point x_i est muni de la masse $p_i > 0$ avec $\sum_{i \in I} p_i = 1$.

2.1 Recherche du meilleur sous-espace de dimension k représentant $\mathcal{N}$

L'objectif de l'ACP est de rechercher pour un entier k fixé le ou les sous-espaces affine de dimension k par rapport auquel(s) le nuage a une inertie minimale. D'après ce qui précède, on sait que le meilleur sous-espace $\mathcal{E}_k$ passe par G le centre de gravité de $\mathcal{N}$. On peut donc prendre l'origine en $O = G$ et il est équivalent de rechercher un sous-espace vectoriel E_k de dimension k tel que l'inertie $In(E_k)$ soit minimal. Comme

$$I_T = I_{E_k} + J_{E_k},$$

il est équivalent de rechercher E_k tel que J_{E_k} soit maximale.

Le théorème suivant utilise la remarque suivante : soit E un sous-espace vectoriel de $I\!R^p$ de dimension k, on considère une base orthonormale $(\epsilon_1, \cdots, \epsilon_k)$ de E, on a

$$J_E(\mathcal{N}) = \sum_{j=1}^{k} \epsilon_j' MVM\epsilon_j = \sum_{j=1}^{k} q(\epsilon_k),$$

en notant q la forme quadratique associée à VM, matrice M-symétrique.

Proposition 2.1.1 *La matrice VM est une matrice M-symétrique, positive. On en déduit que VM est diagonalisable, que ses valeurs propres sont des réels et il existe une base M-orthonormale $(u_1, \cdots u_p)$ constituée de vecteurs propres de VM associés aux valeurs propres respectives*

$$\lambda_1 \geq \cdots \geq \lambda_p \geq 0.$$

Enfin on a

$$\forall u \in I\!R^p, \;\; \lambda_p \|u\|_M^2 \leq q(u) \leq \lambda_1 \|u\|_M^2.$$

Preuve : On a pour tout vecteur u et v de $I\!R^p$:

$$\begin{aligned} < u, VMv >_M &= u'MVMv, \\ &= (VMu)'Mv, \\ &= < VMu, v >_M . \end{aligned}$$

donc la matrice VM est une matrice M-symétrique. De plus

$$< VMu, u >_M = u'MVMu = q_V(Mu),$$

or V en tant que matrice variance covariance est positive donc $q_V(Mu)$ est positif. Ainsi la matrice VM est une matrice M-symétrique, positive. Le reste du théorème est une application du théorème sur les matrices symétriques.

∎

Théorème 2.1.2 *Soit $(u_1, \cdots u_p)$ une base orthonormale de $I\!R^p$ constituée de vecteurs propres de VM associés aux valeurs propres respectives*

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0,$$

on pose

$$\forall k \in [\![1, k]\!], \quad E_k = \text{Vect}(u_1, \cdots, u_k).$$

Alors on a

$$\dim E_k = k,$$

$$E_1 \subset E_2 \subset \cdots \subset E_p = I\!R^p,$$

et

$$\forall k \in [\![1, p]\!], \quad J_{E_k} = \sum_{i=1}^{k} \lambda_i = \max_{E e.v. dim E = k} (J_E).$$

Réciproquement si F est un sous-espace vectoriel de dimension k tel que $J_F = \sum_{i=1}^{k} \lambda_i$, alors il existe une base orthonormale $(v_1, \cdots v_p)$ constituée de vecteurs propres de VM associé aux valeurs propres respectives $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, telle que

$$F = \text{Vect}(v_1, \cdots, v_k).$$

Preuve : La première partie se démontre par récurrence sur k :
Le résultat est vrai pour $k = 1$ puisque pour tout vecteur unitaire u, on a

$$J(\text{Vect}(u)) = q(u) \leq \lambda_1 = q(u_1) = J(E_1).$$

On suppose le résultat vrai pour un entier $k \leq p - 1$, Soit E un sous-espace vectoriel de dimension $k + 1$, on pose

$$F = E \cap \text{Vect}(u_1, \cdots, u_k)^{\perp}.$$

Avec la formule de Grassman, on a

$$p \geq \dim(E + \text{Vect}(u_1, \cdots, u_k)^{\perp}) = \dim E + \dim \text{Vect}(u_1, \cdots, u_k)^{\perp} - \dim F,$$

d'où
$$\dim F \geq k+1+p-k-p = 1.$$
Ainsi la dimension de F est supérieure ou égale à 1, donc il existe un vecteur unitaire z dans F, on construit une base orthonormale de E à partir de la famille libre (z) soit $(\epsilon_1, \cdots, \epsilon_k, z)$, on a
$$J(E) = J(\mathrm{Vect}(\epsilon_1, \cdots, \epsilon_k)) + q(z).$$
D'après l'hypothèse de récurrence, on a
$$J(\mathrm{Vect}(\epsilon_1, \cdots, \epsilon_k)) \leq \sum_{i=1}^{k} \lambda_i,$$
de plus l'espace $\mathrm{Vect}(u_1, \cdots, u_k)^\perp$ est $\mathrm{Vect}(u_{k+1}, \cdots, u_p)$, on en déduit que
$$q(z) \leq \lambda_{k+1}.$$
Par conséquent
$$J(E) \leq \sum_{i=1}^{k+1} \lambda_i = J(E_{k+1}).$$
La réciproque se démontre aussi par récurrence :

— Le résultat est vrai pour $k=1$. En effet soit v un vecteur unitaire tel que $q(v) = \lambda_1$. Puisque
$$E_{\lambda_1} \oplus \bigoplus_{\lambda_i \neq \lambda_1} E_{\lambda_i} = I\!R^p,$$
où E_{λ_i} est le sous-espace propre associé à λ_i, on peut décomposer v en
$$v = v_1 + w \text{ où } v_1 \in E_{\lambda_1}, \quad w \in \bigoplus_{\lambda_i \neq \lambda_1} E_{\lambda_i}.$$
On a alors
$$\lambda_1 = q(v) = q(v_1) + q(w) = \lambda_1 \|v_1\|^2 + q(w).$$
Par ailleurs
$$w = \sum_{2 \leq i \leq p,\ \lambda_i \neq \lambda_1} \alpha_i u_i,$$
donc en notant i_0 le plus petit indice i tel que $\lambda_i \neq \lambda_1$
$$q(w) = \sum_{2 \leq i \leq p,\ \lambda_i \neq \lambda_1} \alpha_i^2 \lambda_i \leq \lambda_{i_0} \|w\|^2,$$
d'où l'on déduit que
$$q(w) = \lambda_1 \|w\|^2 \leq \lambda_{i_0} \|w\|^2.$$
Or $\lambda_{i_0} < \lambda_1$, donc $w = 0$ ainsi $v = v_1$. Donc v est un vecteur propre unitaire associé à λ_1.

— On suppose que E est de dimension $k+1$ et vérifie
$$J_E = \sum_{i=1}^{k+1} \lambda_i,$$

alors on reprend le raisonnement précédent : il existe une base orthonormale de E à partir de la famille libre (z) soit $(\epsilon_1, \cdots, \epsilon_k, z)$, on a

$$J(E) = J(\text{Vect}(\epsilon_1, \cdots, \epsilon_k)) + q(z).$$

Or on a

$$J(\text{Vect}(\epsilon_1, \cdots, \epsilon_k)) \leq \sum_{i=1}^{k} \lambda_i \text{ et } q(z) \leq \lambda_{k+1}.$$

Etant donnée l'égalité, les deux inégalités sont des égalités. En utilisant l'hypothèse de récurrence, la première montre que $\text{Vect}(\epsilon_1, \cdots, \epsilon_k)$ corresponds à F_k.
Par ailleurs, le vecteur unitaire z est combinaison linéaire de $u_{k+1}, \cdots, u_p$

$$z = \sum_{i=k+1}^{p} \alpha_i u_i \text{ et } \sum_{i=k+1}^{p} \alpha_i^2 = 1.$$

La deuxième égalité donne donc

$$\sum_{i=k+1}^{p} \alpha_i^2 \lambda_i = \lambda_{k+1}.$$

Soit i_0 le plus petit indice supérieur à $k+1$ tel que $\lambda_i < \lambda_{k+1}$, on a

$$\lambda_{k+1} \sum_{i=k+1}^{i_0-1} \alpha_i^2 + \sum_{i=i_0}^{p} \lambda_i \alpha_i^2 = \lambda_{k+1} \sum_{i=k+1}^{p} \alpha_i^2,$$

d'où

$$\sum_{i=i_0}^{p} \lambda_i \alpha_i^2 = \lambda_{k+1} \sum_{i=i_0}^{p} \alpha_i^2,$$

puisque $\lambda_i < \lambda_{k+1}$, l'égalité n'est possible que si $\displaystyle\sum_{i=i_0}^{p} \alpha_i^2 = 0$. Ainsi z est dans $E_{\lambda_{k+1}}$.

■

On peut introduire les définition suivantes :

Définition 2.1.3 *Soit $(u_1, \cdots u_p)$ une base orthonormale de vecteurs propres de VM associé aux valeurs propres respectives*

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0,$$

pour tout entier $1 \leq \alpha \leq p$,

- *l'axe* $\mathrm{Vect}(u_\alpha)$ *est appelé le αième axe factoriel du nuage de points $\mathcal{N}$.*
- *$\varphi_\alpha = Mu_\alpha$ est appelé le αième facteur,*
- *$\forall i \in [\![1, n]\!]$, $\psi_{i,\alpha} =< y_i, u_\alpha >_M = y_i' M u_\alpha = y_i' \varphi_\alpha$ est l'abscisse de la projection de y_i sur* $\mathrm{Vect}(u_\alpha)$ *:*

$$\psi_\alpha = \begin{pmatrix} \psi_{1,\alpha} \\ \vdots \\ \psi_{n,\alpha} \end{pmatrix} = Y\varphi_\alpha = YMu_\alpha \text{ est appelée } \alpha \text{ ième composante principale.}$$

- *le taux d'inertie expliquée par le αième axe factoriel, noté τ_α, est la quantité*

$$\tau_\alpha = \frac{\lambda_\alpha}{I_T} = \frac{\lambda_\alpha}{\sum_{i=1}^{p} \lambda_i}.$$

- *le taux d'inertie expliquée par E_α, noté $\tau_{1\cdots\alpha}$, est la quantité*

$$\tau_{1\cdots\alpha} = \frac{\lambda_1 + \cdots + \lambda_\alpha}{I_T} = \sum_{i=1}^{\alpha} \tau_i.$$

2.2 Représentations des individus

Proposition 2.2.1 *Si la matrice V est de rang r, alors le nuage $\mathcal{N}$ centré a pour support* $E_r = \mathrm{Vect}(u_1, \cdots, u_r)$.

Preuve : En effet V et VM ont même rang puisque M est inversible. Donc les valeurs propres $\lambda_{r+1}, \cdots, \lambda_p$ sont nulles. Or si un axe factoriel est trivial c'est-à-dire associé à une valeur propre nulle alors le nuage $\mathcal{N}$ est inclus dans l'hyperplan orthogonal à cet axe.

■

Ainsi lorsque V est de rang r, un individu i a $p - r$ coordonnées nulles donc est caractérisé par r valeurs $\psi_{i,1}, \cdots, \psi_{i,r}$ au lieu des p coordonnées initiales dans la base canonique.

Si le taux $\tau_{1,2}$ est proche de 1, on visualise le nuage $\mathcal{N}$ dans le plan $\mathrm{Vect}(u_1, u_2)$, noté plan 1×2. Sinon on complète cette représentation par les projections sur les plans 1×3, 2×3, voire si $\tau_{1,2,3}$ est trop faible, sur les plans 1×4, 2×4, etc.

Définition 2.2.2 Qualité de représentation *La qualité de la représentation de l'individu i sur E_k est*

$$QLT(y_i, E_k) = \cos^2(\theta_{i,E_k}),$$

où θ_{i,E_k} est l'angle entre y_i et E_k.

Proposition 2.2.3 *On a*

$$QLT(y_i, E_k) = \sum_{\alpha=1}^{k} QLT(y_i, \mathrm{Vect}(u_\alpha)) = \sum_{\alpha=1}^{k} (\frac{\psi_{i,\alpha}}{\|y_i\|_M})^2.$$

Preuve : On note Π la projection orthogonale sur E_k, on a

$$\Pi(y_i) = \sum_{\alpha=1}^{k} < y_i, u_\alpha > u_\alpha,$$

donc

$$\cos^2(\theta_{i,E_k}) = \left(< \frac{y_i}{\|y_i\|}, \frac{\Pi(y_i)}{\|\Pi(y_i)\|} >_M\right)^2 = \sum_{\alpha=1}^{k} (\frac{\psi_{i,\alpha}}{\|y_i\|_M})^2.$$

■

Plus ce facteur de qualité se rapproche de 1, mieux est représenté l'individu i. S'il vaut 1, alors y_i est dans E_k.

On note parfois sur les listings, $COR_\alpha(i)$ pour désigner $1000 \times \cos^2(\theta_{i,\mathrm{Vect}(u_\alpha)})$ et aussi $QLT_{E_k}(i) = 1000 \times \cos^2(\theta_{i,E_k})$.

2.3 Représentation des variables

Les variables y^j sont représentés par les vecteurs de l'espace $\mathbb{R}^n$ muni de la métrique D_p. Pour cette métrique, la norme d'un vecteur est l'écart-type de la variable et le produit scalaire entre deux vecteurs est la covariance entre les deux variables. La composante principale ψ_α est un vecteur de $\mathbb{R}^n$.

On suppose que r est le rang de V.

Proposition 2.3.1

$$\forall(\alpha, \beta) \in [\![1, p]\!]^2, \ < \psi_\alpha, \psi_\beta >_{D_p} = \begin{cases} \lambda_\alpha & si\ \alpha = \beta, \\ 0 & si\ \alpha \neq \beta \end{cases}$$

On pose

$$\forall \alpha \in [\![1, r]\!], \ v_\alpha = \frac{\psi_\alpha}{\sqrt{\lambda_\alpha}}.$$

On en déduit que pour tout $1 \leq k \leq r$, $(v_1, \cdots, v_k)$ est une base D_p-orthonormale de $\mathrm{Vect}(v_1, \cdots, v_k) = F_k$.

En particulier $\mathrm{Vect}(v_1, \cdots, v_r)$ *est une base orthonormale de* $\mathrm{Vect}(y^1, \cdots, y^p) = \mathrm{Im}\, Y$*, la αième coordonnée de y^j est donnée par*

$$\eta_j^\alpha = < y^j, \frac{\psi_\alpha}{\sqrt{\lambda_\alpha}} >_{D_p} .$$

On a

$$\eta^\alpha = \begin{pmatrix} \eta_1^\alpha \\ \vdots \\ \eta_r^\alpha \end{pmatrix} = \sqrt{\lambda_\alpha} u_\alpha \ et \ \|\eta_\alpha\|_M^2 = \lambda_\alpha.$$

Preuve On a

$$
\begin{aligned}
< \psi_\alpha, \psi_\beta >_{D_p} &= \psi'_\alpha D_p \psi_\beta, \\
&= u'_\alpha M Y' D_p Y M u_\beta, \\
&= u'_\alpha M V M u_\beta, \\
&= \lambda_\beta < u_\alpha, u_\beta >_M .
\end{aligned}
$$

D'où le premier résultat. Puis on a

$$\eta_j^\alpha = \frac{1}{\sqrt{\lambda_\alpha}} y^{j'} D_p Y' M u_\alpha$$

donc

$$\eta^\alpha = \frac{1}{\sqrt{\lambda_\alpha}} Y' D_p Y M u_\alpha = \sqrt{\lambda_\alpha} u_\alpha.$$

■

Définition 2.3.2 Qualité de représentation *La qualité de la représentation de la variable y^j sur $F_k = \mathrm{Vect}(v_1, \cdots, v_k)$, avec $1 \leq k \leq r$ est*

$$QLT(y^j, F_k) = \cos^2(\theta_{j,F_k}),$$

où θ_{j,F_k} est l'angle entre y^j et F_k.

Proposition 2.3.3 *On a*

$$QLT(y^j, F_k) = \sum_{\alpha=1}^{k} QLT(y^j, \mathrm{Vect}(v_\alpha)) = \sum_{\alpha=1}^{k} < \frac{y^j}{\|y^j\|}, v_\alpha >^2_{D_p} = \sum_{\alpha=1}^{k} r^2_{i,\alpha}.$$

où $r_{j,\alpha}$ désigne la correlation entre y^j et v_α puisque ces deux variables sont centrées.

2.4 Décompositions de l'inertie

2.4.1 Décomposition de l'inertie selon les individus

Puisque l'inertie totale I_T est égale à la somme des valeurs propres et comme chaque valeur propre λ_α est le carré de la norme de la composante principale associée ψ_α pour la métrique D_p, on a

$$I_T = \sum_{\alpha=1}^{r} \lambda_\alpha = \sum_{\alpha=1}^{r} \|\psi_\alpha\|^2_{D_p} = \sum_{\alpha=1}^{r} \sum_{i=1}^{n} p_i (\psi_{i,\alpha})^2.$$

On en déduit la définition suivante :

Définition 2.4.1 Contribution relative
La contribution relative de l'individu y_i à l'inertie de l'axe α est

$$CTR_\alpha(i) = \frac{p_i(< y_i, u_\alpha >_M)^2}{\lambda_\alpha} = \frac{p_i(\psi_{i,\alpha})^2}{\lambda_\alpha},$$

De même la contribution relative de l'axe α à l'inertie de l'individu y_i est

$$COR_\alpha(i) = \frac{(\psi_{i,\alpha})^2}{\sum_{k=1}^{n}(\psi_{k,\alpha})^2} = \cos^2(\theta_{i,\alpha}),$$

où $\theta_{i,\alpha}$ est l'angle entre y_i et u_α.

Sur les listings, $CTR_\alpha(i)$ et $COR_\alpha(i)$ sont souvent multipliés par 1000.

2.4.2 Décomposition de l'inertie selon les variables

On suppose que la matrice M est diagonale :

$$M = diag(m_1, \cdots, m_p) \text{ où les réels } m_j,\ 1 \leq j \leq p, \text{ sont strictement positifs..}$$

Puisque l'inertie totale I_T est égale à la somme des valeurs propres et comme chaque valeur propre λ_α est le carré de la norme de η_α pour la métrique M, on a

$$I_T = \sum_{\alpha=1}^{r} \lambda_\alpha = \sum_{\alpha=1}^{r} \|\eta^\alpha\|_M^2 = \sum_{\alpha=1}^{r}\sum_{j=1}^{p} m_j(\eta_j^\alpha)^2.$$

On en déduit la définition suivante :

Définition 2.4.2 Contribution relative
La contribution relative de la variable y^j à l'inertie de l'axe α est

$$CTR_\alpha(j) = \frac{m_j(< y^j, v_\alpha >_{D_p})^2}{\lambda_\alpha} = m_j u_{j,\alpha}^2,$$

De même la contribution relative de l'axe α à l'inertie de la variable y^j est

$$COR_\alpha(j) = r_{i,\alpha}^2 = \cos^2(\theta_{j,\alpha}),$$

où $\theta_{j,\alpha}$ est l'angle entre y^j et v_α.

Sur les listings, $CTR_\alpha(j)$ et $COR_\alpha(j)$ sont souvent multipliés par 1000.

Pour tout $1 \leq \alpha \leq p$, le vecteur u_α est unitaire pour la métrique M donc $\sum_{j=1}^{p} m_j u_{j,\alpha}^2 = 1$, soit

$$\lambda_\alpha = \sum_{j=1}^{p} \lambda_\alpha m_j u_{j,\alpha}^2.$$

On retrouve les formules de contributions relatives pour une variable y^j.

2.4.3 Eléments supplémentaires

Observation douteuse, élément aberrant, cas nouveau, centre de gravité d'un groupe ("homme", "femme"), éléments de nature différente (opinion/CSP).

Individu supplémentaire : tout individu y_s de $\mathbb{R}^p$ n'ayant pas participé à l'analyse. L'abscisse $\psi_{s,\alpha}$ de sa projection sur Vect(u_α)vérifie

$$\psi_{s,\alpha} = y^{s'} M u^\alpha.$$

Il est clair que $\psi_{s,\alpha}$ s'obtient en effectuant l'analyse factorielle du tableau $X_1 = \begin{pmatrix} X \\ x'_s \end{pmatrix}$ et en donnant un poids nul à s. En effet dans ce cas, les seuls points ayant une inertie non nulle sont les x_i pour $1 \leq i \leq n$.

Exercice : Montrer que l'on a

$$\psi_{s,\alpha} = \frac{1}{\lambda_\alpha} \sum_{i=1}^{n} w_{s,i} p_i \psi_{i,\alpha} \text{ avec } w_{s,i} =< y_s, y_j >_M .$$

variable supplémentaire : toute variable x^s de $\mathbb{R}^n$ n'ayant pas participé à l'analyse pourra être représentée par ses projections sur les nouveaux axes v_α, on note y^s la variable centrée

$$\eta_s^\alpha =< y^s, \frac{\psi_\alpha}{\sqrt{\lambda_\alpha}} >_M .$$

2.5 Analyse en composantes principales

La recherche des axes factoriels, facteurs, composantes principales d'un nuage de points dans $\mathbb{R}^p$ muni de la métrique M s'appelle Analyse en Composantes Principales (ACP).

2.5.1 ACP sur matrice variance

On suppose que

$$M = I_p \text{ et } D_p = \frac{1}{n} I_n.$$

Dans ce cas, la matrice VM est la matrice de variance-covariance.

On effectue souvent la représentation des variables dans le cercle de corrélations c'est-à-dire au lieu de représenter les variables selon leurs covariances avec les facteurs, on les représente par leurs corrélations avec les facteurs.

Si dans un plan, une variable est sur le cercle de corrélations, alors elle parfaitement représentée, donc expliquée, par les deux facteurs associés.

2.5.2 ACP sur matrice de corrélation ou ACP normée

En divisant chaque variable par son écart-type, on obtient un nouveau tableau Z dont les variables sont toutes centrées et réduites. On a

$$Z = Y\Delta \text{ où. } \Delta = diag(\frac{1}{\sqrt{v_{11}}}, \cdots, \frac{1}{\sqrt{v_{pp}}}).$$

Dans ce cas, la matrice $Z' D_p Z$ est la matrice de corrélations. On réalise l'ACP sur Z avec $M = I_p$. L'inertie totale du nuage est alors égal à p, le nombre de variables.

On obtient le même résultat en réalisant une ACP sur Y avec

$$M = \Delta^2 = diag(\frac{1}{v_{11}}, \cdots, \frac{1}{v_{pp}}).$$

Dans ce cas, on diagonalise $VM = Y'D_pYM$. Soit u_α un axe factoriel associé à la valeur propre λ_α, on a

$$Z'D_pZ = \Delta V \Delta.$$

Donc

$$Z'D_pZ\Delta u_\alpha = \Delta VMu_\alpha = \lambda_\alpha \Delta u_\alpha,$$

ainsi Δu_α est un axe factoriel dans l'analyse de Z associé à la valeur propre λ_α. Les composantes principales des deux analyses sont égales car

$$Z\Delta u_\alpha = YMu_\alpha.$$

2.6 Analyse factorielle d'un système de points munis de poids et de distances

On considère un système de points M_i, $i \in I$, munis de poids p_i dans un espace affine. On identifie les points M_i à leurs vecteurs de coordonnées x_i dans un espace euclidien muni de la métrique M.

Théorème 2.6.1 *La représentation du nuage $\mathcal{N}_X$ des x_i affectés des poids p_i dans le système des axes factoriels ne dépend que des poids p_i et des distances entre M_i et $M_{i'}$ où*

$$d^2(M_i, M_{i'}) = \|x_i - x_{i'}\|_M^2.$$

preuve Soit ψ_α la composante principale associée au αième axe factoriel, on a

$$YMY'D_p\psi_\alpha = \lambda_\alpha\psi_\alpha, \text{ où } Y \text{ est le tableau centré associé à } X.$$

Il est clair que

$$\|y_i - y_{i'}\|_M^2 = \|x_i - x_{i'}\|_M^2.$$

Pour démontrer le résultat, il suffit de prouver que la matrice $YMY'D_p$ ne dépend que des poids p_i et des distances $d^2(M_i, M_{i'})$ et même que YMY' ne dépend que des poids p_i et des distances $d^2(M_i, M_{i'}) = d(i, i')$. La matrice YMY' est la matrice de Gram associée à la famille de vecteurs $(y_i)_{1\leq i\leq n}$ qui vérifie

$$\sum_{i=1}^{n} p_i y_i = 0.$$

On pose

$$\forall i \in [\![1, n]\!], \ d(\cdot, i) = \sum_{i'=1}^{n} p_{i'} d(i, i'), \ \text{ et } d(\cdot, \cdot) = \sum_{i=1}^{n} p_i d(\cdot, i).$$

On a la première relation

$$\forall (i, i') \in [\![1, n]\!]^2, \ < y_i, y_{i'} >_M = \frac{-1}{2}(\|y_i - y_{i'}\|_M^2 - \|y_i\|_M^2 - \|y_{i'}\|_M^2).$$

On somme de $i' = 1$ à n en pondérant par $p_{i'}$ d'où

$$0 = \frac{1-}{2}(d(\cdot, i) - \|y_i\|_M^2 - I_T) \Longrightarrow \|y_i\|_M^2 = d(\cdot, i) - I_T,$$

où I_T est l'inertie totale : $I_T = \sum_{i=1}^{n} p_j \|y_i\|_M^2$.

Puis on somme de $i = 1$ à n en pondérant par p_i d'où

$$0 = \frac{1}{2}(d(\cdot, \cdot) - 2I_T).$$

De cette dernière relation, on déduit que

$$I_T = \frac{1}{2} d(\cdot, \cdot).$$

Par conséquent

$$\forall (i, i') \in [\![1, n]\!]^2, \;\; < y_i, y_{i'} >_M = \frac{-1}{2}(d(i, i') - d(\cdot, i) + I_T - d(\cdot, i') + I_T),$$

donc

$$\forall (i, i') \in [\![1, n]\!]^2, \;\; < y_i, y_{i'} >_M = \frac{-1}{2}(d(i, i') - d(\cdot, i) - d(\cdot, i') + d(\cdot, \cdot)).$$

■

2.7 Approche SVD

2.7.1 Introduction

Soit X un tableau $n - p$ et Y le tableau centré. On considère les métriques

$$D_p = diag(p_1, \cdots, p_n) \text{ et } M = diag(m_1, \cdots, m_p).$$

On exprime l'inertie totale à l'aide de la norme de Frobénius sur les matrices

$$\forall A \in \mathcal{M}_{n,p}(I\!R), \;\; \|A\|_F^2 = tr(A'A) = \sum_{i=1}^{n} \sum_{j=1}^{p} a_{ij}^2.$$

On en déduit donc que

$$I_T = \sum_{i=1}^{n} p_i \|y_i\|_M^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} p_i m_j (y_i^j)^2,$$

soit

$$I_T = \|D_p^{1/2} Y M^{1/2}\|_F^2.$$

Soit r le rang de Y, soit $1 \leq k \leq r$, le problème de l'ACP est de déterminer un sous-espace vectoriel E de dimension inférieure ou égale à k tel que l'inertie totale du nuage Y projeté orthogonalement sur E soit maximale. Or ce problème est équivalent à rechercher une matrice B de rang inférieur à k qui approche le mieux $D_p^{1/2} Y M^{1/2}$ au sens de la norme de Frobenius, c'est-à-dire en notant P_E la projection orthogonale sur E

$$\max_{dimE \leq k} I_T(P_E(Y)) = \min_{dimE \leq k} I_T(Y - P_E(Y)) = \min_{rangB \leq k} \|D_p^{1/2} Y M^{1/2} - B\|_F.$$

On a alors $E = \operatorname{Im} B'$.

En effet la matrice $P_E(Y)$ est la matrice dont la $i$ème ligne est la projection de y_i sur E. Ainsi toutes les lignes sont dans E donc le rang de $P_E(Y)$ est inférieur à la dimension de E soit k. De plus les matrices $D_p^{1/2}$ et $M^{1/2}$ sont inversibles, donc $D_p^{1/2} P_E(Y) M^{1/2}$ est de rang inférieur à k. Ainsi

$$\min_{dim E \leq k} I_T(Y - P_E(Y)) = \min_{dim E \leq k} \|D_p^{1/2} Y M^{1/2} - D_p^{1/2} P_E(Y) M^{1/2}\|_F^2 \geq \min_{rang B \leq k} \|D_p^{1/2} Y M^{1/2} - B\|_F.$$

Réciproquement, soit B une matrice de rang inférieur à k, on considère la matrice B_1 telle que $B = D_p^{1/2} B_1 M^{1/2}$. On pose $E = \operatorname{Im} B_1'$. Puisque

$$\forall i \in [\![1, n]\!], \quad \|y_i - b_i\|_M^2 \geq \|y_i - p_E(y_i)\|_M^2,$$

on en déduit que

$$\|D_p^{1/2}(Y - B_1) M^{1/2}\|_F = \|D_p^{1/2} Y M^{1/2} - B\|_F \geq \|D_p^{1/2} Y M^{1/2} - D_p^{1/2} P_E(Y) M^{1/2}\|_F.$$

Ainsi il y a égalité

$$\min_{dim E \leq k} \|D_p^{1/2} Y M^{1/2} - D_p^{1/2} P_E(Y) M^{1/2}\|_F^2 = \min_{rang B \leq k} \|D_p^{1/2} Y M^{1/2} - B\|_F.$$

Ce problème d'approximation d'une matrice par une matrice de rang inférieur est résolu par la décomposition SVD. Nous rappelons cette décomposition démontrée en annexe :

Théorème 2.7.1 SVD *Soit Y une matrice de format $n \times p$ à coefficients réels. On note r le rang de Y, $r \leq s = min(n,p)$. Les espaces $\mathbb{R}^n$ et $\mathbb{R}^p$ sont munis d'une structure euclidienne canonique. Alors il existe*

1. *$(u_1, \cdots, u_n)$ une base orthonormale de $\mathbb{R}^n$,*
2. *$(v_1, \cdots, v_p)$ une base orthonormale de $\mathbb{R}^p$,*
3. *r réels positifs : $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$,*

tel que

$$Y = \sum_{i=1}^{r} \sigma_i u_i v_i'.$$

Les r réels σ_i sont uniques. On les appelle valeurs singulières de Y.
Matriciellement, on pose

$$U = [u_1, \cdots, u_r] \text{ et } V = [v_1, \cdots, v_r], \quad \Sigma = Diag(\sigma_1, \cdots, \sigma_r),$$

L'équation précédente s'écrit

$$Y = U \Sigma V',$$

ou encore on peut poser

$$U = [u_1, \cdots, u_n] \text{ et } V = [v_1, \cdots, v_n],$$

et Σ est une matrice $n \times p$ dont les coefficients diagonaux sont des réels positifs ou nuls et tous les autres sont nuls. Les termes diagonaux de Σ sont rangés par ordre décroissant . Les matrices U et V sont deux matrices orthogonales d'ordre respectif n et p ($U'U = UU' = I_n$ et $VV' = V'V = I_p$)
Dans les deux cas, la matrice Σ est unique.

2.7.2 Approximation par une matrice de rang inférieur

Etant donné une matrice Y de format $n \times p$, de rang r, on cherche à approcher Y par une matrice Y_k de rang inférieur ou égal à k. On utilise dans ce théorème la norme euclidienne canonique.

Proposition 2.7.2 *Soit Y une matrice de format $n - p$, de rang r, on a*

$$Y = U\Sigma V' = \sum_{i=1}^{r} \sigma_i u_i v_i'.$$

Soit k un entier non nul inférieur ou égal à r, on pose

$$Y_k = \sum_{i=1}^{k} \sigma_i u_i v_i'.$$

On a alors

$$\min_{B \in \mathcal{M}_{n,p}(\mathbb{R}), rang(B) \leq k} \|Y - B\|_F = \|Y - \sum_{i=1}^{k} \sigma_i u_i v_i'\|_F = \sum_{i=k+1}^{r} \sigma_i^2.$$

Y_k est ainsi la meilleure approximation de Y par une matrice de rang inférieur ou égal à k.

preuve :

1. soit U et V des matrices orthogonales d'ordre respectifs p et n, on a pour toute matrice A de format $n - p$

$$\|VA\|_F = \|AU\|_F = \|A\|_F.$$

en effet

$$\|VA\|_F^2 = tr(A'V'VA) = tr(A'A) = \|A\|_F^2.$$

et

$$\|AU\|_F^2 = tr(U'A'AU) = tr(A'A) = \|A\|_F^2.$$

2. On en déduit que

$$\|Y - \sum_{i=1}^{k} \sigma_i u_i v_i'\|_F = \sqrt{\sum_{i=k+1}^{r} \sigma_i^2}.$$

en effet en posant $S = diag(0, \cdots, 0, \sigma_{k+1}, \cdots, \sigma_r)$

$$\begin{aligned}
\|Y - \sum_{i=1}^{k} \sigma_i u_i v_i'\|_F^2 &= \|\sum_{i=k+1}^{r} \sigma_i u_i v_i'\|_F^2, \\
&= \|USV'\|_F^2, \\
&= \|S^2\|_F, \\
&= tr(SS'), \\
&= \sum_{i=k+1}^{r} \sigma_i^2.
\end{aligned}$$

3. Soit B une matrice de $\mathcal{M}_{n,p}(I\!R)$ de rang inférieur ou égal à k. On a

$$\|Y - B\|_F = \|\Sigma - V'BU\|_F.$$

En effet

$$\begin{aligned}
\|Y - B\|_F &= \|V\Sigma U' - B\|_F, \\
&= \|V(\Sigma - V'BU)U'\|_F, \\
&= \|\Sigma - V'BU\|_F, \\
&= \sum_{i=1}^{r}(\sigma_i - (V'BU)_{ii})^2 + \sum_{i\neq j}(V'BU)^2_{ij} + \sum_{i=r+1}^{n}(V'BU)^2_{ii}.
\end{aligned}$$

On en déduit que pour réaliser le minimum, $V'BU$ est diagonale de rang k et annule les k plus grandes valeurs de σ_i.

■

Ce résultat peut s'étendre à d'autres normes de matrices.

Proposition 2.7.3 *Soit Y une matrice de format $n - p$, de rang r, on a*

$$Y = U\Sigma V' = \sum_{i=1}^{r}\sigma_i u_i v_i'.$$

Soit k un entier non nul inférieur ou égal à r, on pose

$$Y_k = \sum_{i=1}^{k}\sigma_i u_i v_i'.$$

On a alors

$$\min_{B\in\mathcal{M}_{n,p}(I\!R), rang(B)\leq k}\|Y - B\|_2 = \|Y - \sum_{i=1}^{k}\sigma_i u_i v_i'\|_2 = \sigma_{k+1}.$$

où

$$\|A\|_2 = \sup_{\|X\|=1}(\|AX\|) \text{ et } \|X\| \text{ est la norme euclidienne canonique.}$$

Y_k est ainsi la meilleure approximation de Y par une matrice de rang inférieur ou égal à k.

preuve :

1. On a

$$\|Y - \sum_{i=1}^{k}\sigma_i u_i v_i'\|_2 = \|\sum_{i=k+1}^{r}\sigma_i u_i v_i'\|_2,$$

soit $x \in I\!R^p$ de norme 1, on a

$$\|\sum_{i=k+1}^{r}\sigma_i u_i v_i' x\|_2 = \|\sum_{i=k+1}^{r}\sigma_i v_i' x\, u_i\|_2 = \sqrt{\sum_{i=k+1}^{r}\sigma_i^2(v_i'x)^2} \leq \sigma_{k+1}\|x\|_2 = \sigma_{k+1}.$$

Cette valeur est atteinte pour $x = v_{k+1}$ donc

$$\|Y - \sum_{i=1}^{k} \sigma_i u_i v_i'\|_2 = \sigma_{k+1}.$$

2. Soit B une matrice quelconque de $\mathcal{M}_{n,p}(I\!R)$ de rang inférieur ou égal à k.
   (a) En utilisant le théorème du rang, on a

   $$\dim \operatorname{Ker} B \geq p - k.$$

   (b) Puisque

   $$p \geq \dim(\operatorname{Ker} B + \operatorname{Vect}(v_1, \cdots, v_{k+1})) =$$

   $$\dim \operatorname{Ker} B + \dim \operatorname{Vect}(v_1, \cdots, v_{k+1}) - \dim(\operatorname{Ker} B \bigcap \operatorname{Vect}(v_1, \cdots, v_{k+1})),$$

   donc

   $$\dim(\operatorname{Ker} B \bigcap \operatorname{Vect}(v_1, \cdots, v_{k+1})) \geq \dim \operatorname{Ker} B - p + k + 1 \geq 1,$$

   on en déduit que $\operatorname{Ker} B \bigcap \operatorname{Vect}(v_1, \cdots, v_{k+1})$ n'est pas réduit au vecteur nul.

   (c) Soit w un vecteur unitaire de $\operatorname{Ker} B \cap \operatorname{Vect}(v_1, \cdots, v_{k+1})$, on a

   $$\begin{aligned} \|(Y - B)w\|^2 &= \|Yw\|2, \\ &= \|\sum_{i=1}^{r} \sigma_i u_i v_i' w\|^2, \\ &= \|\sum_{i=1}^{k+1} \sigma_i u_i v_i' w\|^2, \\ &= \sum_{i=1}^{k+1} \sigma_i^2 (v_i' w)^2, \\ &\geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} (v_i' w)^2, \\ &\geq \sigma_{k+1}^2. \end{aligned}$$

   Ainsi

   $$\|Y - B\|_2 \geq \sigma_{k+1}.$$

■

2.7.3 Application à l'ACP

Etant donné un tableau de donnée X de format $n \times p$, on commence par centrer ce tableau comme dans l'ACP, on obtient la matrice Y. On commence par étudier la cas d'une ACP sur matrice variance c'est-à-dire $M = I_p$ et $D_p = \frac{1}{n} I_n$.

D'après la décomposition SVD de Y, si r est le rang de Y, il existe deux matrices orthogonales F et U de formats respectifs n et p et r réels positifs $\sigma_1 > \cdots > \sigma_r > 0$ tels que

$$Y = F \Sigma U',$$

où Σ est une matrice de format $n - p$, vérifiant

$$\forall (i,j) \in [\![1,n]\!] \times [\![1,p]\!], \quad \Sigma_{i,j} = \begin{cases} \sigma_i & \text{si } 1 \leq i = j \leq r \\ 0 & \text{sinon} \end{cases}$$

Les axes factoriels sont obtenus en diagonalisant la matrice $VM = V$, or on a

$$VM = V = Y'D_pY = \frac{1}{n}Y'Y = \frac{1}{n}U\ diag(\sigma_i^2)\, U'$$

On en déduit que les axes factoriels sont les vecteurs colonnes de U et les valeurs propres $\frac{1}{n}\sigma_i^2$. Pour les composantes principales, on diagonalise

$$YMY'D_p = \frac{1}{n}YY' = \frac{1}{n}F\ diag(\sigma_i^2)\, F'.$$

Ainsi la composante principale associé à l'axe α ψ_α est colinéaire à la αème colonne de F notée F_α. De plus la norme de ψ_α pour la métrique D_p est $\frac{\sigma_i}{\sqrt{n}}$, on en déduit que

$$\psi_\alpha = \sigma_\alpha F_\alpha.$$

Dans le cas plus général où M n'est pas l'identité et D_p n'est pas une matrice scalaire, on suppose que M est une matrice diagonale : $M = diag(m_1, \cdots, m_p)$ et $D_p = diag(p_1, \cdots, p_n)$. On se ramène au cas précédent de la manière suivante

— on transforme Y en $Z = D_p^{1/2} Y M^{1/2}$.

— on applique la décomposition SVD à Z d'où

$$Z = F_0 \Sigma U_0'$$

— puis on revient sur Y

$$Y = D_p^{-1/2} F_0 \Sigma (M^{-1/2} U_0)' = F\Sigma U'.$$

On peut conclure comme précédemment que

$$VM = U diag(\sigma_i^2) U'M.$$

Or on a

$$U'MU = U_0' M^{-1/2} M M^{-1/2} U_0 = U_0' U_0 = I_p.$$

Ce qui signifie que les vecteurs colonnes de U : $u_1, \cdots, u_p$ forment une base orthonormale de $I\!R^p$ muni de la métrique M . Ainsi puisque

$$VM = U diag(\sigma_i^2) U^{-1},$$

$u_1, \cdots, u_p$ sont les p axes factoriels associées respectivement aux valeurs propres $\sigma_1^2, \cdots, \sigma_r^2, 0, \cdots, 0$ où r est le rang de Y.

De même les vecteurs colonnes de F : $F_1, \cdots, F_n$ forment une base orthonormale de $I\!R^n$ muni de la métrique D_p constitués de vecteurs propres de $YMY'D_p$. La composante principale associée à l'axe u_i est

$$\psi_i = \sigma_i F_i.$$

2.7.4 Exemple

On considère le tableau de données suivant :

$I \setminus J$	x	y	z
1	1	6	0
2	2	5	1
3	3	4	2
4	4	3	2
5	5	2	1
6	6	1	0

associé aux résultats de trois variables x, y et z mesurées sur un échantillon I de six individus. On suppose que chaque individu i de I $(1 \leq i \leq 6)$ est muni de la masse $1/6$ et $M = I_3$.

On note X le tableau associé. Le tableau centré est

$$Y = \begin{pmatrix} -2.5 & 2.5 & -1 \\ -1.5 & 1.5 & 0 \\ -0.5 & 0.5 & 1 \\ 0.5 & -0.5 & 1 \\ 1.5 & -1.5 & 0 \\ 2.5 & -2.5 & -1 \end{pmatrix}$$

On applique la décomposition SVD à la matrice $D_p^{1/2}Y = \frac{1}{\sqrt{6}}Y$, on ne donne que les trois premières colonnes de U et de V puisque le rang r de Y ne peut dépasser 3 :

$$\frac{1}{\sqrt{6}}Y = \begin{pmatrix} -0.5976143 & -5.000000e-01 & -0.6210748 \\ -0.3585686 & 0.000000e+00 & 0.2975208 \\ -0.1195229 & 5.000000e-01 & -0.3769329 \\ 0.1195229 & 5.000000e-01 & -0.4863153 \\ 0.3585686 & 2.775558e-17 & -0.2975208 \\ 0.5976143 & -5.000000e-01 & -0.2421734 \end{pmatrix} \times \begin{pmatrix} 2.415229 & 0 & 0 \\ 0 & 8.164966e-01 & 0 \\ 0 & 0 & 1.470706e-16 \end{pmatrix}$$

$$\times \begin{pmatrix} 7.071068e-01 & -1.424533e-16 & -7.071068e-01 \\ -7.071068e-01 & -7.744525e-17 & -7.071068e-01 \\ 4.596760e-17 & 1.000000e+00 & -1.554917e-16 \end{pmatrix}'$$

Donc

$$Y = \sqrt{6} \begin{pmatrix} -0.5976143 & -5.000000e-01 & -0.6210748 \\ -0.3585686 & 0.000000e+00 & 0.2975208 \\ -0.1195229 & 5.000000e-01 & -0.3769329 \\ 0.1195229 & 5.000000e-01 & -0.4863153 \\ 0.3585686 & 2.775558e-17 & -0.2975208 \\ 0.5976143 & -5.000000e-01 & -0.2421734 \end{pmatrix} \times \begin{pmatrix} 2.415229 & 0 & 0 \\ 0 & 8.164966e-01 & 0 \\ 0 & 0 & 1.470706e-16 \end{pmatrix}$$

$$\times \begin{pmatrix} 7.071068e-01 & -1.424533e-16 & -7.071068e-01 \\ -7.071068e-01 & -7.744525e-17 & -7.071068e-01 \\ 4.596760e-17 & 1.000000e+00 & -1.554917e-16 \end{pmatrix}'$$

Les valeurs propres sont alors les carrés des termes de la diagonale, les axes factoriels sont donnés par les colonnes de U et les composantes principales se déduisent des colonnes de F. Dans le tableau ci-dessous, on donne les valeurs propres, puis les axes factoriels, puis les composantes principales :

λ_1	λ_2	λ_3
5.833333e+00	6.666667e-01	2.664535e-15
7.071068e-01	-1.424533e-16	-7.071068e-01
-7.071068e-01	-7.744525e-17	-7.071068e-01
4.596760e-17	1.000000e+00	-1.554917e-16
-3.5355339	-1.000000e+00	-2.237410e-16
-2.1213203	0.000000e+00	1.071813e-16
-0.7071068	1.000000e+00	-1.357893e-16
0.7071068	1.000000e+00	-1.751941e-16
2.1213203	5.551115e-17	-1.071813e-16
3.5355339	-1.000000e+00	-8.724248e-17

Ces résultats ont été obtenus en utilisant la commande svd de R

Programme en R.

```
> X=matrix(c(1,2,3,4,5,6,6,5,4,3,2,1,0,1,2,2,1,0),6,3)
> moy=apply(X,MARGIN=2,mean)
> Y=1/sqrt(6)*(X-t(moy\%*\%matrix(c(1,1,1,1,1,1),1,6)))
> s=svd(Y)
```

Chapitre 3

Analyse Factorielle des Correspondances

3.1 Introduction

L'analyse Factorielle des Correspondances (AFC) a été introduite pour analyser les tableaux de contingence. Un tableau de contingence croise les ensembles I et J de deux variables qualitatives X et Y. Un tel tableau peut se noter k_{IJ} et a alors pour terme général le nombre $k(i,j)$ d'individus qui ont pris simultanément la modalité i pour la variable X et la modalité j pour la variable Y.

L'AFC consiste à effectuer deux ACP, l'une sur le tableau des profils lignes, l'autre sur celui des profils colonnes de k_{IJ}.

L'AFC peut être appliquée à des tableaux de nombres positifs de types divers : tableaux de contingence, tableaux homogènes de nombres positifs, tableaux d'échanges, tableau de rangs, tableaux de présence /absence, tableau disjonctifs complets,....

3.2 Définition des nuages étudiés par l'AFC

3.2.1 Notations

On étudie deux variables qualitatives X et Y, X a p modalités et Y q modalités. Le tableau de contingence k_{IJ} est une matrice de format $p \times q$. On pose $I = \{1, \cdots p\} = [\![1, p]\!]$ et $J = \{1, \cdots q\} = [\![1, q]\!]$.

On note

$$k_I = (k(i,\cdot))_{i \in I} \in I\!R^p \text{ avec } k(i,\cdot) = \sum_{j=1}^{q} k(i,j),$$

$$k_J = (k(\cdot,j))_{j \in J} \in I\!R^q \text{ avec } k(\cdot,j) = \sum_{i=1}^{p} k(i,j),$$

$$k = \sum_{j=1}^{q} \sum_{i=1}^{p} k(i,j).$$

On transforme les effectifs en **fréquences** : on obtient un nouveau tableau F_{IJ} ou F de terme courant

$$\forall (i,j) \in I \times J, \ \ f_{i,j} = \frac{k(i,j)}{k}.$$

On a les lois marginales :

$$f_I = (f_{i\cdot})_{i\in I} \in I\!R^p \text{ avec } f_{i\cdot} = \sum_{j=1}^{q} f_{i,j} = \frac{k(i\cdot)}{k},$$

$$f_J = (f_{\cdot j})_{j\in J} \in I\!R^q \text{ avec } f_{\cdot j} = \sum_{i=1}^{p} f_{i,j} = \frac{k(\cdot j)}{k}.$$

f_I est la loi marginale sur I et f_j sur J. Ainsi f_I et f_J sont des distributions de probabilités donc

$$\sum_{i\in I}\sum_{j\in J} f_{i,j} = \sum_{i\in I} f_{i\cdot} = \sum_{j\in J} f_{\cdot j} = 1.$$

On peut aussi introduire la loi conditionnelle sur I sachant j appelé profil de la colonne j :

$$f_I^J = (f_i^j)_{i\in I, j\in J} \text{ avec } f_i^j = \frac{f_{i,j}}{f_{\cdot j}} = \frac{k(i,j)}{k(\cdot,j)},$$

Ainsi f_I^J est une matrice de format $p\times q$ et f_I^j est le $j$ième vecteur colonne de $I\!R^p$.

De même on a la loi conditionnelle sur J sachant i appelé profil de la ligne i :

$$f_J^I = (f_j^i)_{i\in I, j\in J} \text{ avec } f_j^i = \frac{f_{i,j}}{f_{i\cdot}} = \frac{k(i,j)}{k(i,\cdot)},$$

Ainsi f_J^I est une matrice de format $q\times p$ et f_J^i est le $i$ème vecteur de $I\!R^q$.

Puisque f_I^j et f_J^i sont des distributions de probabilités, on a

$$\sum_{i\in I} f_i^j = \sum_{j\in J} f_j^i = 1.$$

S'il n'y a pas d'ambiguité, on note f_i pour $f_{i\cdot}$, f_j pour $f_{\cdot j}$, $k(i)$ pour $k(i,\cdot)$ et $k(j)$ pour $k(\cdot,j)$.

On suppose qu'aucune ligne ou colonne de $K_{I,J}$ n'est nulle. Donc f_i et f_j sont non nulles et f_i^j et f_j^i sont bien définies.

3.2.2 Nuages et métriques

L'AFC consiste à effectuer deux ACP sur deux nuages différents mais présentant une certaine symétrie. On note

$$D_{f_I} = Diag(f_i)_{i\in I} \in \mathcal{M}_p(I\!R) \text{ et } D_{f_J} = Diag(f_j)_{j\in J} \in \mathcal{M}_q(I\!R).$$

On a

$$D_{f_I}^{-1} = Diag(1/f_i)_{i\in I} = D_{1/f_I} \text{ et } D_{f_J}^{-1} = Diag(1/f_j)_{j\in J} = D_{1/f_J}.$$

On dit que D_{1/f_I} (respectivement D_{1/f_J}) est la métrique du chi-deux de centre f_I (respectivement f_J).

On considère les nuages suivants :
— $\mathcal{N}(J) = \{f_I^j, j\in J\}$**, appelé nuage des profils colonnes**, où chaque point f_I^j de $I\!R^p$ est muni du poids f_j et $I\!R^p$ est muni de la métrique D_{1/f_I}.
— $\mathcal{N}(I) = \{f_J^i, i\in I\}$**, appelé nuage des profils lignes**, où chaque point f_J^i de $I\!R^q$ est muni du poids f_i et $I\!R^q$ est muni de la métrique D_{1/f_J}.

On note

$$F_1 = f_I^J = (f_I^1, \cdots, f_I^q) \text{ et } F_2 = f_J^I = (f_J^1, \cdots, f_J^p).$$

F_1 est le tableau des profils colonnes et F_2 des profils lignes. On peut remarquer que F_1' et F_2' sont les matrices correspondantes à X.

Proposition 3.2.1 *On a*

$$F_1 = FD_{1/f_J} \text{ et } F_2 = F'D_{1/f_I}.$$

On en déduit que le rang de F est égal au rang de F_1 et à celui de F_2.

preuve : Les matrices D_{1/f_J} et D_{1/f_I} sont inversibles d'où le résultat.

3.3 Nuage $\mathcal{N}(J)$

3.3.1 Support

Puisque l'on a pour tout $j \in J$

$$\sum_{i \in I} f_i^j = 1,$$

on en déduit que tous les points du nuage $\mathcal{N}(J)$ sont dans l'hyperplan affine de $\mathbb{R}^p$ d'équation

$$\sum_{i \in I} x_i = 1.$$

3.3.2 Centre de gravité

Proposition 3.3.1 *Le centre de gravité du nuage $\mathcal{N}(J)$ est f_I.*
Le support du nuage $\mathcal{N}(J)$ est inclus dans l'hyperplan affine passant par f_I et D_{1/f_I}-orthogonal à f_I.

preuve : En effet soit G_I ce centre de gravité, on a

$$G_I = f_I^J D_{f_J} 1_q = F 1_q = f_I.$$

Par ailleurs

$$< f_I^j - f_I, f_I >_{D_{1/f_I}} = (f_I^j - f_I)' D_{1/f_I} f_I = (f_I^j - f_I)' 1_p = 0.$$

■

3.3.3 Effet du non centrage

On effectue une ACP sur la matrice F_1' avec les métriques $M = D_{1/f_I}$ et $D_p = D_{f_J}$. Le centre de gravité des individus pondérés par D_p est f_I, et la matrice de variance V est

$$V = (F_1 - f_I 1_q') D_{f_J} (F_1 - f_I 1_q')' = F_1 D_{f_J} F_1' - f_I f_I'.$$

Le support de $\mathcal{N}(J)$ est inclus dans l'hyperplan affine d'équation $\sum_{i \in I} x_i = 1$, ce qui se traduit par

$$F_1' 1_p = 1_q.$$

Proposition 3.3.2 *La matrice* $VM = VD_{1/f_I}$ *et la matrice* $F_1 D_{f_J} F_1' D_{1/f_I}$ *ont les mêmes vecteurs propres et*

$$Spectre(VD_{1/f_I}) = \{\lambda_p = 0 \leq \lambda_{p-1} \cdots \leq \lambda_1\} \text{ et } Spectre(VD_{1/f_I}) = \{1, \lambda_{p-1}, \cdots, \lambda_1\}.$$

preuve : En effet on a

$$F_1 D_{f_J} F_1' D_{1/f_I} f_I = F_1 D_{f_J} F_1' 1_p = F_1 D_{f_J} 1_q = f_I, \text{ et } f_I f_I' D_{1/f_I} f_I = f_I f_I' 1_p = f_I.$$

On en déduit que f_I est un vecteur propre associé à la matrice VD_{1/f_I} et à la matrice $F_1 D_{f_J} F_1' D_{1/f_I}$ avec les valeurs propres 0 et 1 respectivement. Comme ces matrices représentent des endomorphismes D_{1/f_I}-symétriques, l'orthogonal de Vect(f_I) est stable pour ces deux matrices. Or soit u un vecteur de Vect$(f_I)^\perp$, on a

$$f_I f_I' D_{1/f_I} u =< f_I, u >_{D_{1/f_I}} f_i = 0.$$

Ainsi

$$\forall u \in \text{Vect}(f_I)^\perp, \;\; VD_{1/f_I} u = F_1 D_{f_J} F_1' D_{1/f_I} u,$$

par conséquent la restriction à Vect$(f_I)^\perp$ des endomorphismes représentés par VD_{1/f_I} et par $F_1 D_{f_J} F_1' D_{1/f_I}$ sont identiques donc les deux matrices ont mêmes valeurs propres et même vecteurs propres.

■

On en déduit que pour obtenir les axes factoriels de l'ACP, le centrage n'est pas nécéssaire. Pour le calcul des composantes principales, il n'est pas nécessaire de centrer non plus :

Soit u_I un axe factoriel orthogonal à f_I, la composante principale ψ^J associée à l'axe u_I est

$$\forall j \in J, \;\; \psi^j =< (f_I^j - f_I), u_I >=< f_I^j, u_I > .$$

A l'axe factoriel trivial f_I, on associe la composante triviale $\psi_o = F_1' D_{1/f_I} f_I = 1_q$.

3.3.4 Axes Factoriels, facteurs et composantes principales

Proposition 3.3.3 *L'ACP du nuage $\mathcal{N}(J)$ consiste à diagonaliser F_1F_2.*
Les axes factoriels sont solutions de

$$\left\{\begin{array}{l} F_1F_2u_I^\alpha = \lambda_\alpha u_I^\alpha, \\ < u_I^\alpha, u_I^\beta >_{D_{1/f_I}} = \delta_{\alpha,\beta}, \\ < u_I^\alpha, f_I >_{D_{1/f_I}} = 0. \end{array}\right.$$

Les facteurs φ_α^I sont solutions de

$$\left\{\begin{array}{l} F_2'F_1'\varphi_\alpha^I = \lambda_\alpha \varphi_\alpha^I, \\ < \varphi_\alpha^I, \varphi_\beta^I >_{D_{f_I}} = \delta_{\alpha,\beta}, \\ < \varphi_\alpha^I, 1_I >_{D_{f_I}} = 0. \end{array}\right.$$

Les composantes principales ψ_α^J sont solutions de

$$\left\{\begin{array}{l} F_1'F_2'\psi_\alpha^J = \lambda_\alpha \psi_\alpha^J, \\ < \psi_\alpha^J, \psi_\beta^J >_{D_{f_J}} = \lambda_\alpha\delta_{\alpha,\beta}, \\ < \psi_\alpha^J, 1_J >_{D_{f_J}} = 0. \end{array}\right.$$

Toutes les valeurs propres λ_α sont positives et inférieures à 1.

preuve : Pour trouver les axes factoriels, on diagonalise

$$F_1D_{f_J}F_1'D_{1/f_I},$$

or $F_1 = FD_{1/f_J}$, on a

$$F_1D_{f_J}F_1'D_{1/f_I} = F_1D_{f_J}D_{1/f_J}F'D_{1/f_I}, = F_1F_2.$$

Les facteurs sont vecteurs propres de MV sans avoir besoin de centrer donc la matrice

$$D_{1/f_I}F_1D_{f_J}F_1' = F_2'F_1' \text{ puisque } F_2 = F'D_{1/f_I}.$$

Enfin les composantes principales sont vecteurs propres de

$$F_1'D_{1/f_I}F_1D_{f_J} = F_1'F_2'.$$

Enfin les valeurs propres sont positives. De plus le terme courant (j,k) de $F_1'F_2'$ est

$$\sum_{i=1}^p f_i^j\, f_k^i$$

donc l'égalité $F_1'F_2'\psi = \lambda\psi$ devient

$$\sum_{k=1}^q \sum_{i=1}^p f_i^j\, f_k^i \psi(k) = \lambda\psi(j),$$

en notant $\psi(j_0)$ la plus grande coordonnée de ψ, on a

$$\lambda\psi(j) \leq \sum_{k=1}^q \sum_{i=1}^p f_i^j\, f_k^i\, \psi(j_0) = \psi(j_0),$$

on en déduit que $0 \leq \lambda \leq 1$.

■

3.4 Le nuage $\mathcal{N}(I)$

L'étude de $\mathcal{N}(I)$ se déduit de celle de $\mathcal{N}(J)$ en intervertissant les rôles de I et de J. Ainsi le centre de gravité de $\mathcal{N}(I)$ est f_J, le support de $\mathcal{N}(I)$ est inclus dans l'hyperplan affine d'équation $\sum_{j\in J} x_j = 1$. On échange F_1 et F_2, donc pour trouver les axes factoriels on diagonalise F_2F_1, les facteurs, on diagonalise $F_1'F_2'$ et les composantes principales, on diagonalise $F_2'F_1'$. On a donc

Proposition 3.4.1 *L'ACP du nuage $\mathcal{N}(I)$ consiste à diagonaliser F_2F_1.*
Les axes factoriels sont solutions de

$$\left\{\begin{array}{l} F_2F_1u_J^\alpha = \lambda_\alpha u_J^\alpha, \\ < u_J^\alpha, u_J^\beta >_{D_{1/f_J}} = \delta_{\alpha,\beta}, \\ < u_J^\alpha, f_J >_{D_{1/f_J}} = 0. \end{array}\right.$$

Les facteurs φ_α^J sont solutions de

$$\left\{\begin{array}{l} F_1'F_2'\varphi_\alpha^J = \lambda_\alpha\varphi_\alpha^J, \\ < \varphi_\alpha^J, \varphi_\beta^J >_{D_{f_J}} = \delta_{\alpha,\beta}, \\ < \varphi_\alpha^J, 1_J >_{D_{f_J}} = 0. \end{array}\right.$$

Les composantes principales ψ_α^I sont solutions de

$$\left\{\begin{array}{l} F_2'F_1'\psi_\alpha^I = \lambda_\alpha\psi_\alpha^I, \\ < \psi_\alpha^I, \psi_\beta^I >_{D_{f_I}} = \lambda_\alpha\delta_{\alpha,\beta}, \\ < \psi_\alpha^I, 1_I >_{D_{f_I}} = 0. \end{array}\right.$$

Toutes les valeurs propres λ_α sont positives et inférieures à 1.

La proposition suivante établit des relations entre les deux ACP :

Proposition 3.4.2 Formules de transition,

On a

$$\psi_\alpha^J = F_1'\varphi_\alpha^I = \sqrt{\lambda_\alpha}\,\varphi_\alpha^J, \ \text{ et } \psi_\alpha^I = F_2'\varphi_\alpha^J = \sqrt{\lambda_\alpha}\varphi_\alpha^I.$$

Ces relations sont appelées relations ou formules de transition.

preuve : Soit λ_α une valeur propre non nulle de $F_2'F_1'$

$$F_2'F_1'\varphi_\alpha^I = \lambda_\alpha\varphi_\alpha^I,$$

en multipliant par F_1', on obtient que $F_1'\varphi_\alpha^I$ est non nul et est donc un vecteur propre de $F_1'F_2'$. On normalise ce vecteur propre, pour cela on calcule sa norme

$$\begin{array}{rcl} \|F_1'\varphi_\alpha^I\|^2 & = & \varphi_\alpha^I F_1 D_{f_J} F_1'\varphi_\alpha^I, \\ & = & \varphi_\alpha^{I\prime} F F_1'\varphi_\alpha^I, \\ & = & \varphi_\alpha^{I\prime} D_{f_I} F_2'F_1'\varphi_\alpha^I, \\ & = & \lambda_\alpha\|\varphi_\alpha^I\|^2, \\ & = & \lambda_\alpha. \end{array}$$

Par conséquent $\dfrac{1}{\sqrt{\lambda_\alpha}} F_1' \varphi_\alpha^I$ est un vecteur propre unitaire de $F_1' F_2'$ associé à la valeur propre λ_α.

De plus soit k et l deux indices distincts, on a

$$\begin{aligned}
< \frac{F_1'\varphi_k^I}{\sqrt{\lambda_k}}, \frac{F_1'\varphi_l^I}{\sqrt{\lambda_l}} > &= \frac{1}{\sqrt{\lambda_k\lambda_l}} \varphi_k^{I,} F_1 D_{f_J} F_1' \varphi_l^I, \\
&= \frac{1}{\sqrt{\lambda_k\lambda_l}} \varphi_k^{I,} D_{f_I} D_{1/f_I} F D_{1/f_J} F' \varphi_l^I, \\
&= \frac{1}{\sqrt{\lambda_k\lambda_l}} \varphi_k^{I,} D_{f_I} F_2' F_1' \varphi_l^I, \\
&= 0.
\end{aligned}$$

On note r le nombre de valeurs propres non nulles de $F_2' F_1'$, c'est-à-dire le rang de $F_2' F_1'$. Ainsi l'image par F_1' de la base orthonormale $(\varphi_1^I, \cdots, \varphi_r^I, \varphi_{r+1}^I, \cdots, \varphi_p^I)$ de $\mathbb{R}^p$ muni de la métrique D_{f_I} donne une famille orthogonale que l'on peut normaliser soit $(\frac{F_1'\varphi_1^I}{\sqrt{\lambda_1}}, \cdots, \frac{F_1'\varphi_r^I}{\sqrt{\lambda_r}})$, ce qui donne une famille orthonormale de $\mathbb{R}^q$ muni de la métrique D_{f_J} constituée de vecteurs propres de $F_1' F_2'$.

On en déduit que le rang de $F_1' F_2'$ est supérieure à r. Par symétrie entre les deux analyses, on en déduit que $F_1' F_2'$ et $F_2' F_1'$ ont même rang et donc les mêmes valeurs propres non nulles.

Par conséquent pour toute valeur propre non nulle, on a

$$F_1' \varphi_\alpha^I = \sqrt{\lambda_\alpha}\, \varphi_\alpha^J,$$

d'où les formules de transition.

Pour une valeur propre nulle, $\lambda_\alpha = 0$, le calcul de la norme de $F_1' \varphi_\alpha^I$ montre que

$$F_1' \varphi_\alpha^I = 0.$$

Les formules de transition sont encore satisfaites.

■

Remarque 3.4.3 *Il existe diverses formulations des relations de transition. Par exemple si la valeur propre λ_α est non nulle, on peut écrire*

$$\psi_\alpha^J = \frac{1}{\sqrt{\lambda_\alpha}} F_1' \psi_\alpha^I.$$

On en déduit

$$\forall j \in J, \;\; \psi_\alpha^j = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i \in I} f_i^j \psi_\alpha^i$$

De même en inversant i et j, on a aussi

$$\forall i \in I, \;\; \psi_\alpha^i = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j \in J} f_j^i \psi_\alpha^j.$$

On a aussi les même relations de transition pour les facteurs.

Représentation simultanée. En AFC, on effectue une représentation simultanée des modalités $i \in I$ et $j \in J$. Plus précisément, sur chaque axe α, on représente $i \in I$ par le point d'abscisse ψ_α^i et $j \in J$ par le point d'abscisse ψ_α^j. Autrement dit, on superpose les représentations des nuages $\mathcal{N}(I)$ et $\mathcal{N}(J)$ dans leurs systèmes d'axes respectifs. D'après les formules de transitions, il en résulte qu'au facteur $\frac{1}{\sqrt{\lambda_\alpha}}$ près, le point j est le barycentre des points i affectés des poids f_i^j. De même le point i est le barycentre des points j affectés des poids f_j^i.

3.5 Inerties

3.5.1 Inertie totale

Proposition 3.5.1 *Les nuages $\mathcal{N}(I)$ et $\mathcal{N}(J)$ ont même inertie totale égale à*

$$I_T = \sum_{i\in I}\sum_{j\in J}\frac{(f_{i,j}-f_if_j)^2}{f_if_j}.$$

preuve : On a

$$\begin{aligned} I_T &= \sum_{j\in J} f_j\|f_I^j - f_I\|^2_{D(1/f_I)},\\ &= \sum_{j\in J} f_j\sum_{i\in I}\frac{1}{f_i}(f_i^j - f_i)^2,\\ &= \sum_{j\in J}\sum_{i\in I}\frac{f_j}{f_i}(\frac{f_{i,j}}{f_j} - f_i)^2,\\ &= \sum_{i\in I}\sum_{j\in J}\frac{(f_{i,j}-f_if_j)^2}{f_if_j}. \end{aligned}$$

∎

Remarque 3.5.2 *En écrivant :*

$$(f_{i,j}-f_if_j)^2 = f_{i,j}^2 - 2f_if_jf_{i,j} + f_i^2f_j^2,$$

et en remarquant que

$$\frac{-2f_if_jf_{i,j}+f_i^2f_j^2}{f_if_j} = -2f_{i,j} + f_if_j,$$

on en déduit que

$$\sum_{i\in I}\sum_{j\in J}\frac{-2f_if_jf_{i,j}+f_i^2f_j^2}{f_if_j} = -2\sum_{i\in I}\sum_{j\in J}f_{i,j} + \sum_{i\in I}f_i\sum_{j\in J}f_j = -2+1=-1.$$

Par conséquent, on a

$$I_T = \sum_{i\in I}\sum_{j\in J}\frac{f_{i,j}^2}{f_if_j} - 1.$$

3.5.2 Interprétation de l'inertie totale dans le cas d'un tableau de contingence

On suppose que K est un tableau de contingence, et plus précisément que I (resp. J) est l'ensemble des modalités d'une variable qualitative X (resp. Y). Ainsi K donne les effectifs de co-occurence des couples de modalités (i,j) sur un échantillon de taille k. Donc F est un estimateur de la mesure de probabilité théorique $p_{I,J}$ (loi jointe de (X,Y)). On sait alors que asymptotiquement, i.e. pour k tendant vers l'infini, on a

$$k\sum_{i\in I}\sum_{j\in J}\frac{(f_{i,j}-p_{ij})^2}{p_{ij}} \to \chi^2_{pq-1},$$

où $p = \mathrm{Card}(I)$ et $q = \mathrm{Card}(J)$.

Lorsque l'on teste l'hypothèse :

$$H_0 : p_{IJ} = p_I p_J,$$

H_0 représente l'hypothèse d'indépendance des variables aléatoires X et Y, on est amené à estimer les lois marginales p_I par f_I et p_J par f_J. Pour p_I, on estime $p-1$ paramètres puisque la somme des p_i vaut 1, de même pour p_J on estime $q-1$ paramètres. Il en résulte que

$$k \sum_{i \in I} \sum_{j \in J} \frac{(f_{i,j} - f_i f_j)^2}{f_i f_j} \to \chi^2_\mu,$$

avec

$$\mu = pq - 1 - (p-1) - (q-1) = (p-1)(q-1).$$

On en conclut que la quantité

$$kI_T = k \sum_{i \in I} \sum_{j \in J} \frac{(f_{i,j} - f_i f_j)^2}{f_i f_j}$$

permet de tester l'hypothèse d'indépendance des variables X et Y.

Dans le cas d'indépendance, kI_T aura tendance à être faible ($kI_T \leq c_\alpha$), et par conséquent, puisque I_T est la somme des valeurs propres, plus les valeurs propres sont faibles moins les facteurs sont interprétables.

Si X et Y ne sont pas indépendants, l'AFC permet de voir comment f_{IJ} s'écarte de l'indépendance, les axes factoriels associés aux plus grandes valeurs propres traduisant les liaisons entre X et Y.

3.5.3 Décomposition de l'inertie, Contributions

On exprime I_T en fonction des composantes principales des deux ACP, on note r le nombre de valeurs propres non nulles, on a

$$I_T = \sum_{\alpha=1}^{r} \sum_{j \in J} f_j (\psi_\alpha^j)^2 = \sum_{\alpha=1}^{r} \sum_{i \in I} f_i (\psi_\alpha^i)^2.$$

De plus la norme d'une composante principale valant la valeur propre, on a

$$\lambda_\alpha = \sum_{j \in J} f_j (\psi_\alpha^j)^2 = \sum_{i \in I} f_i (\psi_\alpha^i)^2.$$

En intervertissant les sommes

$$I_T = \sum_{j \in J} f_j \sum_{\alpha=1}^{r} (\psi_\alpha^j)^2 = \sum_{j \in J} f_j \rho^2(j),$$

où $\rho^2(j)$ est la distance au carré entre f_I^j et f_I.

De même

$$I_T = \sum_{i \in I} f_i \sum_{\alpha=1}^{r} (\psi_\alpha^i)^2 = \sum_{i \in I} f_i \rho^2(i),$$

où $\rho^2(i)$ est la distance au carré entre f_J^i et f_J. On en déduit les définitions suivantes :

Définition 3.5.3 *La contribution de j et i à l'inertie de l'axe α sont respectivement :*

$$CTR_\alpha(j) = \frac{f_j(\psi_\alpha^j)^2}{\lambda_\alpha} \text{ et } CTR_\alpha(i) = \frac{f_i(\psi_\alpha^i)^2}{\lambda_\alpha}.$$

La contribution de l'axe α à l'inertie de j et de i sont

$$COR_\alpha(j) = \frac{(\psi_\alpha^j)^2}{\rho^2(j)} = \cos^2(\theta_{j,\alpha}) \text{ et } COR_\alpha(i) = \frac{(\psi_\alpha^i)^2}{\rho^2(i)} = \cos^2(\theta_{i,\alpha}),$$

où $\theta_{i,\alpha}$ et $\theta_{j,\alpha}$ désignent respectivement les angles formés entre $f_J^i - f_J$ et u_J^α d'une part et entre $f_I^j - f_I$ et u_I^α d'autre part.

Exemple 3.5.4 *Montrer que les relations suivantes sont vérifiées :*

$$\cos^2(\theta_{j,\alpha}) = corr^2(f_j^I, \varphi_\alpha^I) \text{ et } \cos^2(\theta_{i,\alpha}) = corr^2(f_i^J, \varphi_\alpha^J),$$

où $corr^2(f_j^I, \varphi_\alpha^I)$ et $corr^2(f_i^J, \varphi_\alpha^J)$ sont calculées respectivement avec les mesures de probabilté f_i et f_J.

3.6 Principe d'équivalence distributionnelle

Proposition 3.6.1 *Si deux lignes i_1 et i_2 (resp. colonnes j_1 et j_2) du tableau f_{IJ} ou k_{IJ} sont proportionnelles, alors on ne change pas les résultats de l'analyse des correspondances en remplaçant ces deux lignes (resp. colonnes) par leur somme i_0 (resp. j_0) affectée de la somme de leurs poids :*

$$\forall j \in J, \;\; f_{i_0 j} = f_{i_1 j} + f_{i_2 j}.$$

Preuve : On suppose que les deux lignes i_1 et i_2 du tableau f_{IJ} ou k_{IJ} sont proportionnelles, alors il existe un réel a tel que

$$\forall j \in J, \;\; f_{i_1 j} = a \times f_{i_2 j},$$

on en déduit que

$$f_{i_1\cdot} = \sum_{j\in J} f_{i_1 j} = a \times \sum_{j\in J} f_{i_2 j} = a f_{i_2\cdot}.$$

Ainsi dans le nuage $\mathcal{N}(I)$, les deux profils lignes i_1 et i_2 sont confondus :

$$\forall j \in J, \;\; f_j^{i_1} = \frac{f_{i_1 j}}{f_{i_1\cdot}} = \frac{a f_{i_2 j}}{a f_{i_2\cdot}} = f_j^{i_2}.$$

Par conséquent l'ACP du nuage $\mathcal{N}(I)$ n'est pas modifié si l'on réunit les deux individus i_1 et i_2 en un individu i_0 affecté du poids $f_{i_0\cdot} = f_{i_1\cdot} + f_{i_2\cdot}$:

$$f_J^{i_0} = f_J^{i_1} = f_J^{i_2}.$$

Ainsi pour tout $j \in J$,

$$\begin{aligned} f_{i_0 j} &= f_{i_0\cdot} f_j^{i_0}, \\ &= f_{i_1\cdot} f_j^{i_0} + f_{i_2\cdot} f_j^{i_0}, \\ &= f_{i_1\cdot} f_j^{i_1} + f_{i_2\cdot} f_j^{i_2}, \\ &= f_{i_1 j} + f_{i_2 j}. \end{aligned}$$

Pour le nuage $\mathcal{N}(J)$ initial, les distances entre les colonnes j et j' sont

$$d^2(j,j') = \sum_{i\in I} \frac{1}{f_{i\cdot}}(f_i^j - f_i^{j'})^2,$$

et pour le nuage en tenant compte du regroupement des lignes, on a

$$d^2(j,j') = \sum_{i\in I\setminus\{i_1,i_2\}} \frac{1}{f_{i\cdot}}(f_i^j - f_i^{j'})^2 + \frac{1}{f_{i_0\cdot}}(f_{i_0}^j - f_{i_0}^{j'})^2.$$

Or

$$\begin{aligned}
\frac{1}{f_{i_0\cdot}}(f_{i_0}^j - f_{i_0}^{j'})^2 &= f_{i_0\cdot}(\frac{f_{i_0}^j - f_{i_0}^{j'}}{f_{i_0\cdot}})^2,\\
&= (f_{i_1\cdot} + f_{i_2\cdot})(\frac{f_j^{i_0} - f_{j'}^{i_0}}{f_{\cdot j}})^2,\\
&= f_{i_1\cdot}(\frac{f_j^{i_0} - f_{j'}^{i_0}}{f_{\cdot j}})^2 + f_{i_2\cdot}(\frac{f_j^{i_0} - f_{j'}^{i_0}}{f_{\cdot j}})^2,\\
&= f_{i_1\cdot}(\frac{f_j^{i_1} - f_{j'}^{i_1}}{f_{\cdot j}})^2 + f_{i_2\cdot}(\frac{f_j^{i_2} - f_{j'}^{i_2}}{f_{\cdot j}})^2.
\end{aligned}$$

Par conséquent les distances entre les individus j et j' sont les mêmes dans les situations, donc l'ACP de $\mathcal{N}(J)$ est identique dans les deux situations.

■

Remarque 3.6.2 *Cette propriété garantit une invariance des résultats vis à vis du choix de la momenclature pour la construction des modalités d'une variable, sous réserve de regrouper des modalités aux profils similaires.*

3.7 Tableau de Burt

Soit $G = (g_{jj'})_{j,j'\in J}$ le tableau défini par

$$G = F'D_{1/f_I}F.$$

On a

$$\forall (j,j') \in J^2,\ \ g_{jj'} = \sum_{i\in I} \frac{f_{ij}f_{ij'}}{f_{i\cdot}}.$$

On appelle tableau de Burt associé à k_{IJ} le tableau B

$$\forall (j,j') \in J^2,\ \ b_{jj'} = \sum_{i\in I} \frac{k_{ij}k_{ij'}}{k_{i\cdot}} = kg_{jj'}.$$

La matrice G, appelée tableau de Burt, est symétrique donc les deux marges sont égales. On note g_J cette marge commune, on a

$$\forall j \in J,\ \ g_{\cdot j} = g_{j\cdot} = \sum_{i\in I}\sum_{j'\in J} \frac{f_{ij}f_{ij'}}{f_{i\cdot}} = f_{\cdot j}.$$

Donc $g_J = f_J$. Les matrices profil ligne et profil colonne, G_1 et G_2 sont

$$G_1 = GD_{1/f_J} = F'D_{1/f_I}FD_{1/f_J} = F_2F_1 \text{ et } G_2 = G'D_{1/f_J} = G_1.$$

On en déduit que l'AFC de g_{JJ} donne les mêmes axes factoriels u_J^α et les mêmes facteurs de variance 1 φ_α^J que ceux obtenus dans l'AFC de k_{IJ}. Les valeurs propres de l'AFC de G sont les carrés des valeurs propres de L'AFC de k_{IJ}.

Remarque 3.7.1 *Si l'on veut représenter l'ensemble I, il suffit de rajouter f_{IJ} en supplémentaire au tableau $g_{II} = G$. On obtient la même représentation que dans l'AFC de f_{IJ}.*

Chapitre 4

ACM

4.1 Notations-Tableau disjonctif complet-tableau de Burt

4.1.1 Notations et définitions

On note :

Q : ensemble de questions ou de variables qualitatives,
I : ensemble des individus qui ont répondu aux questions, avec $n = |\,I\,|$,
J : ensemble de toutes les modalités de réponse à toutes les questions, avec $p = |\,J\,|$,
J_q : ensemble de toutes les modalités de réponse à la question q,
k_{IJ} : tableau de taille $n \times p$ défini par

$$k(i,j) = \begin{cases} 1 & \text{si l'individu } i \text{ a adopté la modalité } j \text{ de } J, \\ 0 & \text{sinon.} \end{cases}$$

Définition 4.1.1 *Le tableau k_{IJ} est dit* disjonctif *si chaque individu choisit au plus une modalité par question (deux modalités d'une même question s'excluent mutuellement). Le tableau k_{IJ} est dit* complet *si chaque individu choisit au moins une modalité par question.*

Proposition 4.1.2 *Un tableau k_{IJ} est disjonctif complet (TDC) si et seulement si :*

$$\sum_{j \in J_q} k(i,j) = 1 \textit{ pour toute question } q \in Q \textit{ et tout individu } i \in I.$$

4.1.2 Propriétés des tableaux disjonctifs complets

Proposition 4.1.3 *Pour tout individu $i \in I$, toute modalité $j \in J$ et toute question $q \in Q$, on a :*

$$\begin{aligned} k(i) &= \sum_{j\in J} k(i,j) = \sum_{q\in Q}\sum_{j\in J_q} k(i,j) = \operatorname{Card} Q, \\ k(j) &= \sum_{i\in I} k(i,j) = \text{ nombre d'individus ayant choisi la modalité } j, \\ \sum_{j\in J_q} k(j) &= n, \\ k &= \sum_{i\in I}\sum_{j\in J} k(i,j) = \sum_{i\in I} k(i) = n \operatorname{Card} Q. \end{aligned}$$

Exemple 4.1.4

	J_1	J_2	J_3	*total*
1				
$\vdots$				
i	0 1	0 0 1	0 0 1 0	$k(i) = \operatorname{Card} Q$
$\vdots$				
n				
total			$k(j)$	$k = n \operatorname{Card} Q$

4.2 Tableau de Burt

Définition 4.2.1 *Soit k_{IJ} un tableau disjonctif complet, le tableau de Burt associé à k_{IJ}, noté B_{JJ}, est défini pour tout $j, j' \in J$ par :*

$$B(j,j') = \sum_{i\in I} k(i,j)\, k(i,j') = \text{ nombre d'individus qui ont choisi les modalités } j \text{ et } j'.$$

Si $j, j' \in J_q$, alors

$$B(j,j') = \begin{cases} 0 & \text{si } j \neq j' \\ k(j) & \text{sinon.} \end{cases}$$

Proposition 4.2.2 *Pour toute modalité $j \in J$ et toute question $q \in Q$, on a :*

$$\begin{aligned} \sum_{j'\in J_q} B(j,j') &= k(j), \\ B(j) &= \sum_{j'\in J} B(j,j') = k(j) \operatorname{Card} Q, \\ B &= \sum_{j\in J, j'\in J} B(j,j') = \sum_{j\in J} k(j) \operatorname{Card} Q = k \operatorname{Card} Q = n(\operatorname{Card} Q)^2. \end{aligned}$$

Exemple 4.2.3

	J_1	J_2	J_3	*total*
J_1	$\begin{pmatrix} k(1) & 0 & 0 \\ 0 & k(2) & 0 \\ 0 & 0 & k(3) \end{pmatrix}$			
J_2				
J_3				$B(i) = k(j)\,\mathrm{Card}\,Q$
total				$k = n(\mathrm{Card}\,Q)^2$

On va établir que l'AFC de k_{IJ} équivaut à celle de B_{JJ}.

Remarquons tout d'abord que :

$$\frac{B(j)}{B} = \frac{k(j)}{n\ \mathrm{Card}\,Q} = \frac{k(j)}{k}.$$

La marge selon J du tableau k_{IJ} est égale à la marge selon J du tableau B_{JJ}.
Donc les métriques des AFC de k_{IJ} et B_{JJ} dans l'espace $\mathbb{R}^p$ sont identiques.

D'après le chapitre précédent, ces deux AFCs sont en fait équivalentes car on a :

$$B(j,j') = \sum_{i\in I} k(i,j)\,k(i,j') = \mathrm{Card}\,Q \sum_{i\in I} \frac{k(i,j)\,k(i,j')}{k(i)} = k\,(\mathrm{Card}\,Q)\,g_{jj'}.$$

4.3 Propriétés de l'AFC d'un questionnaire

1) Soit $(\varphi^I_\alpha, \varphi^J_\alpha)$ les deux facteurs issus de l'AFC de k_{IJ} de rang α, et donc associés à la même valeur propre λ_α. Alors φ^J_α est le facteur issu de l'AFC de B_{JJ} de rang α, et donc associé à la valeur propre $\mu_\alpha = (\lambda_\alpha)^2$.

2) Soit F^I_α (resp. G^J_α) les projections des profils-lignes (resp. profils-colonnes) sur l'axe de rang α issu de l'AFC de k_{IJ}. Soit $F^J_{B\alpha}$ (resp. $G^J_{B\alpha}$) les projections des profils-lignes (resp. profils-colonnes) sur l'axe de rang α issu de l'AFC de B_{JJ}. On a :

$$F^J_{B\alpha} = G^J_{B\alpha} = \sqrt{\mu_\alpha}\,\varphi^J_\alpha = \lambda_\alpha\,\varphi^J_\alpha = \sqrt{\lambda_\alpha}\,G^J_\alpha.$$

Par ailleurs, les formules de transition entraînent que, pour tout $i \in I$, on a :

$$F_\alpha(i) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j\in J} \frac{k(i,j)}{k(i)}\,G_\alpha(j) = \sum_{j\in J} \frac{k(i,j)}{\mathrm{Card}\,Q}\,\varphi^j_\alpha.$$

En notant $q(i)$ la modalité j de la question $q \in Q$ choisie par l'individu i, on obtient :

$$F_\alpha(i) = \sum_{q\in Q} \frac{\varphi^{q(i)}_\alpha}{\mathrm{Card}\,Q}.$$

Autrement dit, $F_\alpha(i)$ est égal à la moyenne des $\varphi^{q(i)}_\alpha = \frac{G^{q(i)}_\alpha}{\sqrt{\lambda_\alpha}}$, coordonnées "normalisées" des modalités qui ont été choisies par l'individu i. Autrement dit encore, sur chaque axe, la représentation de chaque individu coïncide avec la moyenne des modalités qu'il a choisies à $1/\sqrt{\lambda_\alpha}$ près.

3) Du fait de la structure en blocs des tableaux k_{IJ} et B_{JJ}, on a pour toute question $q \in Q$:

$$\sum_{j \in J_q} f_{\bullet j} \varphi_\alpha^j = 0,$$

où φ_α^J est un facteur non trivial de l'AFC de B_{JJ} ou de k_{IJ}. On en déduit donc :

$$\sum_{j \in J_q} f_{\bullet j} G_\alpha(j) = 0.$$

Autrement dit, lors de l'AFC de k_{IJ}, le centre de gravité des profils f_I^j pour $j \in J_q$ est confondu avec le centre de gravité global. On a le même résultat pour les profils (lignes ou colonnes) du tableau (symétrique) B_{JJ}.

4) En pratique, on effectue l'AFC de B_{JJ} et on met le tableau k_{IJ} en supplémentaire. On a alors :

$$\begin{aligned} G_{B\alpha}(j) &= \sqrt{\lambda_\alpha}\, G_\alpha(j) = \sum_{i \in I} \frac{k(i,j)}{k(j)} F_\alpha(i) \\ G_{B\alpha}(j) &= \sum_{q(i)=j} \frac{F_\alpha(i)}{k(j)}. \end{aligned}$$

Autrement dit, pour tout axe factoriel, chaque modalité $j \in J$ est représentée par le centre de gravité des individus l'ayant choisie.

4.4 Contributions en ACM

On considère un tableau disjonctif complet. L'inertie totale est donnée par

$$I_T = \sum_{j \in J} f_{\bullet j} \rho^2(j) = \sum_{q \in Q} \sum_{j \in J_q} f_{\bullet j} \rho^2(j),$$

On pose

$$CR(j) = f_{\bullet j} \rho^2(j), \text{ et } CR(J_q) = \sum_{j \in J_q} f_{\bullet j} \rho^2(j).$$

On note $p(j)$ la proportion des individus ayant adopté la modalité j, on a

$$p_j = \frac{k(j)}{n}.$$

On démontre alors que

$$\rho^2(j) = \frac{1 - p_j}{p_j}.$$

Comme $f_{\bullet j} = \frac{k(j)}{k} = \frac{p_j}{\mathrm{Card}\, Q}$, on a

$$CR(j) = \frac{1 - p_j}{\mathrm{Card}\, Q},$$

d'où

$$CR(J_q) = \frac{\mathrm{Card}\, J_q - 1}{\mathrm{Card}\, Q},$$

et

$$I_T = \frac{\operatorname{Card} J}{\operatorname{Card} Q} - 1.$$

Décompositions en fonction des axes

On a

$$I_T = \sum_{\alpha} \lambda_\alpha = \sum_{\alpha} \sum_{q \in Q} \sum_{j \in J_q} f_{\bullet j} G_\alpha^2(j),$$

on pose

$$CR_\alpha(j) = f_{\bullet j} G_\alpha^2(j), \;\; CR_\alpha(J_q) = \sum_{j \in J_q} f_{\bullet j} G_\alpha^2(j).$$

On pose

$$CTR_\alpha(q) = \frac{CR_\alpha(J_q)}{\lambda_\alpha}$$

est la contribution relative de J_q à l'inertie de l'axe α. On peut poser

$$COR_\alpha(q) = \frac{CR_\alpha(J_q)}{CR(J_q)}, \;\; QLT(q) = \sum_{\alpha} COR_\alpha(q),$$

$$INR(q) = \frac{CR(J_q)}{CR(J)} = \frac{CR(J_q)}{I_T}.$$

Règles d'interprétation

1. Proximité entre individus : deux individus se essemblent s'ils ont choisi les mêmes modalités.
2. Proximité entre deux modalités de variables différentes : ces modalités correspondent aux points moyens des individus les ayany choisies et sont proches parce qu'elles concernent les mêmes individus ou des individus semblables.
3. Proximité entre deux modalités d'une même variable : par construction, elles s'excluent. Si elles sont proches, c'est que les groupes des individus les ayant choisies se ressemblent.

Annexe A

Espace affine

A.1 Définitions

Définition A.1.1 *Soit E un espace vectoriel, on dit que $\mathcal{E}$ est un espace affine de direction E si il existe une application f de $\mathcal{E} \times \mathcal{E}$ dans E notée*

$$\forall (A, B) \in \mathcal{E} \times \mathcal{E}, \;\; f((A, B)) = \overrightarrow{AB},$$

vérifiant les deux conditions suivantes

— *A1 : Relation de Chasles*

$$\forall (A, B, C) \in \mathcal{E} \times \mathcal{E} \times \mathcal{E}, \;\; \overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC},$$

— *A2 : Pour tout $A \in \mathcal{E}$, l'application f_A définie de $\mathcal{E}$ dans E par*

$$\forall M \in \mathcal{E}, \;\; f_A(M) = \overrightarrow{AM} \text{ est une bijection .}$$

Les éléments de $\mathcal{E}$ sont appelés points et ceux de E vecteurs. On appelle dimension de $\mathcal{E}$ la dimension de E.

Remarque A.1.2 *Pour tout entier n non nul, $\mathbb{R}^n$ est un espace affine de direction $\mathbb{R}^n$ espace vectoriel. Ainsi la notation $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ peut être vu comme un vecteur de $\mathbb{R}^n$ ou un point de l'espace affine $\mathbb{R}^n$.*

Notations : Soit $A \in \mathcal{E}$ et $u \in E$, $A + u$ désigne l'unique point B de $\mathcal{E}$ tel que $\overrightarrow{AB} = u$. Ainsi

$$\forall (A, B) \in \mathcal{E} \times \mathcal{E}, \;\; \forall u \in E, \;\; \overrightarrow{AB} = u \Longleftrightarrow B = A + u \Longleftrightarrow B - A = u.$$

Définition A.1.3 *On considère $\mathcal{E}$ un espace affine de direction E, on dit que $\mathcal{F}$ est un sous-espace affine si il existe un point A de $\mathcal{E}$ et un sous-espace vectoriel F de E tels que*

$$\mathcal{F} = A + F = \{M \in \mathcal{E},\ \exists u \in F, M = A + u\}.$$

La dimension de $\mathcal{F}$ est celle de F.

Exemple A.1.4 *Une droite affine de $\mathcal{E}$ est un sous-espace affine de dimension 1. Dans ce cas $F = \mathrm{Vect}(u)$ où u est non nul, soit A un point de la droite affine, on note $\mathcal{D}_{A,u}$ la droite affine passant par A de direction $\mathrm{Vect}(u)$. On dit encore que u est un vecteur directeur de la droite affine $\mathcal{D}_{A,u}$, on a*

$$\mathcal{D}_{A,u} = \{M \in \mathcal{E},\ \ \exists t \in \mathbb{R},\ \ M = A + t\,u\}.$$

On peut aussi définir une droite affine par deux points distincts A et B, alors la droite affine passant par A et B est $\mathcal{D}_{A,\overrightarrow{AB}}$

Exemple A.1.5 *Un plan affine de $\mathcal{E}$ est un sous-espace affine de dimension 2. Dans ce cas $F = \mathrm{Vect}(u, v)$ où u et v sont des vecteurs non colinéaires, soit A un point du plan affine, on note $\mathcal{P}_{A,(u,v)}$ le plan affine passant par A de direction $\mathrm{Vect}(u, v)$. On a*

$$\mathcal{P}_{A,(u,v)} = \{M \in \mathcal{E},\ \ \exists (t, s) \in \mathbb{R}^2,\ \ M = A + t\,u + s\,v\}.$$

On peut aussi définir un plan affine par trois points non alignés A, B et C, alors le plan affine passant par A, B et C est $\mathcal{P}_{A,(\overrightarrow{AB},\overrightarrow{AC})}$

Exemple A.1.6 *On considère le système linéaire*

$$AX = b \text{ où } A \in \mathcal{M}_{n,p}(\mathbb{R}),\ \ X = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \in \mathbb{R}^p,\ \ b \in \mathbb{R}^n.$$

On suppose qu'il existe une solution particulière X_0, alors l'ensemble des solutions du système linéaire est le sous-espace affine $X_0 + \mathrm{Ker}\,A$ de $\mathbb{R}^p$, de dimension $\dim \mathrm{Ker}\,A = p - rg(A)$, où $rg(A)$ est le rang de A.

Exemple A.1.7 *Un hyperplan affine de $\mathcal{E}$ est un sous-espace affine de dimension $dimE - 1$.*

Remarque A.1.8 *Lorsque l'on fixe un point O dans un espace affine $\mathcal{E}$ de direction E, on vectorialise l'espace affine, c'est-à-dire à l'aide de la fonction f_O on construit une structure d'espace vectoriel sur $\mathcal{E}$, tout point M de $\mathcal{E}$ est assimilé au vecteur $\overrightarrow{OM}$.*

A.2 Barycentre

Définition A.2.1 *On considère $\mathcal{E}$ un espace affine de direction E, soit $M_1, \cdots, M_n$ n points de $\mathcal{E}$, et pour tout $1 \leq i \leq n$, on affecte à chaque point M_i un coefficient ou poids p_i qui est un réel. Soit O une origine,*

— *si* $\sum_{i=1}^{n} p_i = 0$, *alors le vecteur* $\sum_{i=1}^{n} p_i \overrightarrow{OM_i}$ *est indépendant de O.*

— *si* $\sum_{i=1}^{n} p_i = p \neq 0$, *alors le point G défini par*

$$G = \frac{1}{p} \sum_{i=1}^{n} p_i M_i = O + \frac{1}{p} \sum_{i=1}^{n} p_i \overrightarrow{OM_i} \text{ est indépendant de } O.$$

On dit que g est le barycentre des $(M_i, p_i)_{1 \leq i \leq n}$.

Exemple A.2.2 *Le milieu de deux points A et B est le barycentre de A et B affectés des poids $1/2$ et $1/2$.*

Définition A.2.3 *On considère $\mathcal{E}$ un espace affine de direction E, soit $\mathcal{M} = \{M_1, \cdots, M_n\}$ un ensemble de n points de $\mathcal{E}$, on note $< \mathcal{M} >$ l'ensemble des barycentres des points de $\mathcal{M}$ affectés de poids quelconques. Alors $< \mathcal{M} >$ est un sous-espace affine de $\mathcal{E}$. On dit que $< \mathcal{M} >$ est le sous-espace affine engendré par $\mathcal{M}$. C'est le plus petit sous-espace affine contenant $\mathcal{M}$.*

Proposition A.2.4 *Le sous-espace affine engendre par $\mathcal{M} = \{M_1, \cdots, M_n\}$ est associé au sous-espace vectoriel* $\mathrm{Vect}(\overrightarrow{M_1M_2}, \cdots, \overrightarrow{M_1M_n})$. *la dimension de $< \mathcal{M} >$ est au plus $n - 1$.*

A.3 Applications affines

Définition A.3.1 *On considère $\mathcal{E}$ un espace affine de direction E, soit f une application de $\mathcal{E}$ dans $\mathcal{E}$. On dit que f est une application affine si il existe un point O de $\mathcal{E}$ tel que l'application $\vec{f}$ de E dans E qui à tout vecteur u de E associé le vecteur $f(O)f(O+u)$ est linéaire. On appelle $\vec{f}$ l'application linéaire associée à f.*

Remarque A.3.2 *Une application affine f est caractérisée par sa valeur en un point et son application linéaire associée.*

Exemple A.3.3 *Une translation de vecteur u est une application affine telle que*

$$\forall M \in \mathcal{E}, \quad f(M) = M + u.$$

Exemple A.3.4 *Une projection orthogonale affine sur le sous-espace affine $\mathcal{F}$ est une application affine telle qu'il existe un point O de $\mathcal{F}$ vérifiant*

$$\forall M \in \mathcal{E}, \quad f(M) = O + p(OM) \text{ où } p \text{ est la projection orthogonale linéaire sur } F.$$

Proposition A.3.5 *On considère $\mathcal{E}$ un espace affine de direction E, soit f une application de $\mathcal{E}$ dans $\mathcal{E}$. l'application f est une application affine si et seulement si f conserve les barycentres c'est-à-dire pour tout entier n*

$$\forall (x_i, t_i) \in \mathcal{E} \times I\!R, \ \textit{avec} \ \sum_{i=1}^{n} t_i = 1, \ \ f(\sum_{i=1}^{n} t_i x_i) = \sum_{i=1}^{n} t_i f(x_i).$$

Annexe B

Endomorphisme symétrique

Définition B.0.1 *Soit E un espace euclidien muni d'une métrique M, et f un endomorphisme de E, on dit que f est un endomorphisme symétrique si pour tous x et y de E, on a l'égalité*

$$< x, f(y) >_M = < f(x), y >_M .$$

Proposition B.0.2 *Tout endomorphisme symétrique est diagonalisable. De plus il existe une base orthonormale de E constituées de vecteurs propres de f.*

La matrice d'un endomorphisme symétrique dans une base orthomormale quelconque de E est une matrice symétrique à coefficients réels. On en déduit le théorème suivant appelé théorème spectral.

Proposition B.0.3 *Soit A une matrice symétrique à coefficients réels de $\mathcal{M}(\mathbb{R})$, alors A est diagonalisable, il existe donc une matrice diagobale D et P une matrice inversible tels que*

$$A = PDP^{-1}.$$

De plus il est possible de choisir P orthogonale dans ce cas, l'égalité devient

$$A = PDP'.$$

Annexe C

Décomposition SVD

Théorème C.0.1 SVD *Soit Y une matrice de format $n \times p$ à coefficients réels. On note r le rang de Y, $r \leq s = min(n,p)$. Les espaces $I\!R^n$ et $I\!R^p$ sont munis d'une structure euclidienne canonique. Alors il existe*

1. $(u_1, \cdots, u_n)$ *une base orthonormale de* $I\!R^n$,
2. $(v_1, \cdots, v_p)$ *une base orthonormale de* $I\!R^p$,
3. r *réels positifs :* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$,

tel que

$$Y = \sum_{i=1}^{r} \sigma_i u_i v_i'.$$

Les r réels σ_i sont uniques. On les appelle valeurs singulières de Y.
Matriciellement, on pose

$$U = [u_1, \cdots, u_r] \text{ et } V = [v_1, \cdots, v_r], \quad \Sigma = Diag(\sigma_1, \cdots, \sigma_r),$$

L'équation précédente s'écrit

$$Y = U \Sigma V',$$

ou encore on peut poser

$$U = [u_1, \cdots, u_n] \text{ et } V = [v_1, \cdots, v_n],$$

et Σ est une matrice $n \times p$ dont les coefficients diagonaux sont des réels positifs ou nuls et tous les autres sont nuls. Les termes diagonaux de Σ sont rangés par ordre décroissant . Les matrices U et V sont deux matrices orthogonales d'ordre respectif n et p ($U'U = UU' = I_n$ et $VV' = V'V = I_p$)
Dans les deux cas, la matrice Σ est unique.

preuve

1. On montre que la matrice $Y'Y$ est symétrique à coefficients réels et positive.
   En effet $Y'Y$ est symétrique à coefficients réels et de plus

$$\forall u \in I\!R^p, \ \ u'Y'Yu = (Yu)'Yu = \|Yu\| \geq 0.$$

La norme utilisée est la norme euclidienne canonique. Par conséquent $Y'Y$ est symétrique, positive.

2. On montre que le rang de $Y'Y$ est égal au rang de Y, noté r.
En effet, d'après la relation précédente, on montre que
$$\operatorname{Ker} Y = \operatorname{Ker} Y'Y,$$
donc en utilisant le théorème du rang, on obtient le résultat.
3. D'après le théorème spectral, les valeurs propres de $Y'Y$ sont r réels strictement positifs $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ et 0 si $r < p$, et il existe une matrice orthogonale dont les colonnes sont des vecteurs propres $(v_1, \cdots, v_p)$ de $Y'Y$ telle que
$$V'Y'YV = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \text{ où } D = diag(\lambda_1, \cdots, \lambda_r).$$
4. On pose
$$\forall 1 \leq i \leq r, \;\; u_i = \frac{1}{\sqrt{\lambda_i}} Y v_i;$$
On montre que la famille $(u_1, \cdots, u_r)$ est une famille orthonormale de vecteurs propres de YY'.
En effet pour tout $1 \leq i \leq r$, on a
$$YY'u_i = \frac{1}{\sqrt{\lambda_i}} YY'Yv_i = \sqrt{\lambda_i} Y v_i = \lambda_i u_i,$$
et pour tout $1 \leq i, j \leq r$, on a
$$< u_i, u_j >= u_i'v_j = \frac{1}{\sqrt{\lambda_i}\sqrt{\lambda_j}} v_i'Y'Yv_j = \frac{\sqrt{\lambda_j}}{\sqrt{\lambda_i}} v_i'v_j = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon} \end{cases}$$
On complète cette famille orthonormale en une base orthonormale de $I\!R^n$ soit $(u_1, \cdots, u_n)$.
5. On pose $\sigma_i = \sqrt{\lambda_i}$, $1 \leq i \leq r$ et
$$A = \sum_{i=1}^{r} \sigma_i u_i v_i'.$$
On montre que A est égal à Y. Pour cela il suffit de montrer que pour tout $1 \leq j \leq p$, Av_j est égal à Yv_j.
En effet on a
$$Av_j = \sum_{i=1}^{r} \sigma_i u_i (v_i'v_j) = \begin{cases} \sigma_j u_j = Y v_j & \text{si } 1 \leq j \leq r \\ 0 & \text{sinon} \end{cases}$$
Par conséquent
$$Y = \sum_{i=1}^{r} \sigma_i u_i v_i'.$$
6. Ecriture matricielle.
On a
$$[u_1, \cdots, u_r] diag(\sigma_1, \cdots, \sigma_r)[v_1, \cdots, v_r]' = [\sigma_1 u_1, \cdots, \sigma_r u_r] \begin{pmatrix} v_1' \\ \vdots \\ v_r' \end{pmatrix},$$
d'où en effectuant des produit par blocs
$$[u_1, \cdots, u_r] diag(\sigma_1, \cdots, \sigma_r)[v_1, \cdots, v_r]' = \sum_{i=1}^{r} \sigma_i u_i v_i'.$$

7. Les valeurs singulières sont uniques, en effet si $Y = U\Sigma V'$ alors $Y'Y = V\Sigma'\Sigma V'$, la matrice $\Sigma'\Sigma$ est une matrice diagonale $diag(\sigma_1^2, \cdots, \sigma_p^2)$. Ainsi les σ_i^2 sont les valeurs propres de $Y'Y$. On en déduit l'unicité.

$\blacksquare$

Définition C.0.2 *Etant donnée une valeur singulière σ, on dit que le vecteur unitaire u de $\mathbb{R}^n$ et le vecteur unitaire v de $\mathbb{R}^p$ sont respectivement vecteur singulier à gauche et vecteur singulier à droite pour σ si*

$$Yv = \sigma u \text{ et } Y'u = \sigma v.$$