Analyse en Composantes Principales (ACP)

Abdelhakim Necir

Département de Mathématiques
Université de Biskra

Master 1, 2025-2026

Matrice des observations

Etant donné un nuage de points $(\mathbf{e}_1, ..., \mathbf{e}_n)$ dans $\mathbb{R}^p$, c'est à dire avec

$$\text{chaque } \mathbf{e}_i \in \mathbb{R}^p.$$

Chaque point correspondant à un individu pour lequel on observe p variables quantitatives

$$(\mathbf{X}_1, ..., \mathbf{X}_p).$$

On note x_{ij} la valeur de la j-ème variable $\mathbf{X}_j$ mesurée sur le i-ème individu $\mathbf{e}_i$, ainsi

$$\mathbf{X}_j = (x_{1j}, ..., x_{nj})^t \in \mathbb{R}^n \text{ avec } 1 \leq i \leq n \text{ et } 1 \leq j \leq p,$$

désigne le vecteur contenant les observations de la variable j pour tous les individus,

Tableau de contingence: variables qualitatives, modalités

et

$$\mathbf{e}_i = (x_{i1}\ , ..., x_{ip})^t \in \mathbb{R}^p$$

désigne le vecteur des observations de l'individu i. Ces observations sont rassemblées dans une matrice notée $\mathbf{X}^*$, dont les colonnes correspondent aux vecteurs $(\mathbf{X}_1, ..., \mathbf{X}_p)$:

$$\mathbf{X}^* = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \in \mathcal{M}\,(n \times p)\,.$$

On obtient donc en ligne, les observations de chaque individu, et en colonne, les observations de chaque variable.

Examples

Les notes, sur 10 de 6 élèves, en trois matières; mathématiques, Français et sciences naturelles:

$$\mathbf{X}^* = \begin{pmatrix} 8 & 1 & 0 \\ 4 & 6 & 5 \\ 6 & 8 & 7 \\ 10 & 4 & 7 \\ 8 & 2 & 5 \\ 0 & 3 & 6 \end{pmatrix} \in \mathcal{M}\left(6 \times 3\right).$$

Choix de l'origine et réduction des variables

Afin de faciliter les représentations graphiques du nuage de points, on choisit en général comme origine le centre de gravité du nuage.

$$\mathbf{g} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1} \\ \frac{1}{n}\sum_{i=1}^{n} x_{ij} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{ip} \end{pmatrix} \in \mathcal{M}\left(p, 1\right).$$

Autrement dit, le centre de gravité $\mathbf{g}$ est le vecteur dont la j-ème coordonnée g_j qui correspond à la valeur moyenne de la variable j sur les n individus. On se retrouve alors avec une version *centrée* de la matrice $\mathbf{X}^*$, que l'on note

$$\mathbf{X} = \mathbf{X}^* - \mathbf{1}_n \mathbf{g}^t.$$

Choix de l'origine et réduction des variables

Pus précisément, on a

$$\mathbf{1}_n\mathbf{g}^t = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} g_1, & g_2, & \cdots & g_p \end{pmatrix} = \begin{pmatrix} g_1 & \cdots & g_p \\ \vdots & \ddots & \vdots \\ g_1 & \cdots & g_p \end{pmatrix} \in \mathcal{M}(n,p),$$

ansi

$$\mathbf{X} = \begin{pmatrix} x_{11} - g_1 & \cdots & x_{1p} - g_p \\ \vdots & \ddots & \vdots \\ x_{n1} - g_1 & \cdots & x_{np} - g_p \end{pmatrix} \in \mathcal{M}(n,p).$$

Choix de l'origine et réduction des variables

De même, il est souvent plus commode de travailler avec des variables réduites. En d'autres termes, on divise les coordonnées de chaque colonne par l'écart-type correspondant. Plus précisément, notons

$$s_j^2 := \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}-g_j\right)^2,\ j=1,...,p,$$

la variance empirique de la variable $\mathbf{X}_j$, et $1/s=\left(1/s_1,...,1/s_p\right)^t$. On définit la matrice poids suivante:

$$\mathbf{D}_{1/s} := \begin{pmatrix} 1/s_1 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & 1/s_{p-1} & 0 \\ 0 & \cdots & 0 & 1/s_p \end{pmatrix} \in \mathcal{M}\left(p\times p\right).$$

Choix de l'origine et réduction des variables

La version centrée et réduite de la matrice des observations est

$$\mathbf{Z} = \mathbf{X}\mathbf{D}_{1/s} = \begin{pmatrix} \frac{x_{11}-g_1}{s_1} & \cdots & \frac{x_{1p}-g_p}{s_p} \\ \vdots & \ddots & \vdots \\ \frac{x_{n1}-g_1}{s_1} & \cdots & \frac{x_{np}-g_p}{s_p} \end{pmatrix} \in \mathcal{M}\left(n, p\right).$$

Choix de l'origine et réduction des variables

On définit, respectivement, les matrices de variance-covariance et de corrélation associées au vecteur $(\mathbf{X}_1, ..., \mathbf{X}_p)$, par

$$\mathbf{V} = \frac{1}{n}\mathbf{X}^t\mathbf{X} \in \mathcal{M}(p \times p) \text{ et } \mathbf{R} = \frac{1}{n}\mathbf{Z}^t\mathbf{Z} \in \mathcal{M}(p \times p).$$

On note que les deux matrices $\mathbf{V}$ et $\mathbf{R}$ sont symétriques:

$$\mathbf{V}^t = \frac{1}{n}\left(\mathbf{X}^t\mathbf{X}\right)^t = \frac{1}{n}\mathbf{X}^t\mathbf{X} = \mathbf{V}$$

et

$$\mathbf{R}^t = \frac{1}{n}\left(\mathbf{Z}^t\mathbf{Z}\right)^t = \frac{1}{n}\mathbf{Z}^t\mathbf{Z} = \mathbf{R}.$$

Choix de l'origine et réduction des variables

Application: Pour notre exemple, le centre de gravité du nuage est

$$\mathbf{g} = \frac{1}{6}\begin{pmatrix} 8+4+6+10+8+0 \\ 1+6+8+4+2+3 \\ 0+5+7+7+5+6 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 36 \\ 24 \\ 30 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \\ 5 \end{pmatrix},$$

ansi la matrice centrée du nuage est

$$\mathbf{X} = \begin{pmatrix} 8-6 & 1-4 & 0-5 \\ 4-6 & 6-4 & 5-5 \\ 6-6 & 8-4 & 7-5 \\ 10-6 & 4-4 & 7-5 \\ 8-6 & 2-4 & 5-5 \\ 0-6 & 3-4 & 6-5 \end{pmatrix} = \begin{pmatrix} 2 & -3 & -5 \\ -2 & 2 & 0 \\ 0 & 4 & 2 \\ 4 & 0 & 2 \\ 2 & -2 & 0 \\ -6 & -1 & 1 \end{pmatrix}.$$

Choix de l'origine et réduction des variables

Calculons la matrice de variance-covariance $\mathbf{V}$. Le transposé de $\mathbf{X}$ est

$$\mathbf{X}^t = \begin{pmatrix} 2 & -2 & 0 & 4 & 2 & -6 \\ -3 & 2 & 4 & 0 & -2 & -1 \\ -5 & 0 & 2 & 2 & 0 & 1 \end{pmatrix},$$

donc

$$\mathbf{V} = \frac{1}{6}\mathbf{X}^t\mathbf{X} = \frac{1}{6}\begin{pmatrix} 2 & -2 & 0 & 4 & 2 & -6 \\ -3 & 2 & 4 & 0 & -2 & -1 \\ -5 & 0 & 2 & 2 & 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & -3 & -5 \\ -2 & 2 & 0 \\ 0 & 4 & 2 \\ 4 & 0 & 2 \\ 2 & -2 & 0 \\ -6 & -1 & 1 \end{pmatrix}.$$

Après le calcul on trouve

$$\mathbf{V} = \begin{pmatrix} 10.667 & -1.333 & -1.333 \\ -1.333 & 5.666 & 3.666 \\ -1.333 & 3.666 & 5.666 \end{pmatrix}.$$

Choix de l'origine et réduction des variables

Calculons la matrice de corrélation **R**. L'écart-type de chaque variable colonne sont:

$$s_1 = 3.\,266\,0, s_2 = 2.\,380\,5, s_3 = 2.\,380\,5.$$

Par conséquent $1/s = (0.306,\ 0.420,\ 0.420)^t$, et la matrice poids est

$$\mathbf{D}_{1/s} = \begin{pmatrix} \frac{1}{s_1} & 0 & 0 \\ 0 & \frac{1}{s_2} & 0 \\ 0 & 0 & \frac{1}{s_3} \end{pmatrix} = \begin{pmatrix} 0.306 & 0 & 0 \\ 0 & 0.420 & 0 \\ 0 & 0 & 0.420 \end{pmatrix}.$$

Ainsi, la matrice de nuage centré-réduit est

$$\mathbf{Z} = \mathbf{XM} = \begin{pmatrix} 2 & -3 & -5 \\ -2 & 2 & 0 \\ 0 & 4 & 2 \\ 4 & 0 & 2 \\ 2 & -2 & 0 \\ -6 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0.306 & 0 & 0 \\ 0 & 0.420 & 0 \\ 0 & 0 & 0.420 \end{pmatrix}.$$

Le calcul donne

$$\mathbf{Z} = \begin{pmatrix} 0.612 & -1.260 & -2.100 \\ -0.612 & 0.840 & 0 \\ 0 & 1.680 & 0.840 \\ 1.224 & 0 & 0.840 \\ 0.612 & -0.840 & 0 \\ -1.837 & -0.420 & 0.420 \end{pmatrix}$$

Espace des individus et espace des variables

Ainsi la matrice de corrélation est

$$
\begin{aligned}
\mathbf{R} &= \frac{1}{6}\mathbf{Z}^t\mathbf{Z} \\
&= \frac{1}{6}\begin{pmatrix} 0.612 & -0.612 & 0 & 1.224 & 0.612 & -1.837 \\ -1.260 & 0.840 & 1.680 & 0 & -0.840 & -0.420 \\ -2.100 & 0 & 0.840 & 0.840 & 0 & 0.420 \end{pmatrix} \\
&\qquad \times \begin{pmatrix} 0.612 & -1.260 & -2.100 \\ -0.612 & 0.840 & 0 \\ 0 & 1.680 & 0.840 \\ 1.224 & 0 & 0.840 \\ 0.612 & -0.840 & 0 \\ -1.837 & -0.420 & 0.420 \end{pmatrix}
\end{aligned}
$$

Espace des individus et espace des variables

Ainsi la matrice de corrélation est Le calcul donne

$$\mathbf{R} = \begin{pmatrix} 0.999 & -0.171 & -0.171 \\ -0.171 & 0.999 & 0.647 \\ -0.171 & 0.647 & 0.999 \end{pmatrix}.$$

Espace des individus et espace des variables

Deux points de vue peuvent être adoptés pour l'étude du nuage de points :
- on peut associer à chaque individu i le vecteur $\mathbf{e}_i$ contenant ses observations sur les p variables (i.e. la i-ème ligne de $\mathbf{X}$). Ce vecteur appartient à un espace vectoriel de dimension p : c'est l'espace des individus.
- on peut associer à chaque variable j le vecteur $\mathbf{X}_j$ contenant les observations de la variable j (i.e. la j-ème colonne de $\mathbf{X}$). Ce vecteur appartient à un espace vectoriel de dimension n : c'est l'espace des variables.

Choix d'une distance

L'étape suivante consiste à introduire une distance entre les points de chacun des deux espaces définis ci-dessus. En ACP, on utilise par convention la distance euclidienne classique. Autrement dit, la distance entre deux individus, représentés par les vecteurs $\mathbf{e}_i$ et $\mathbf{e}_{i'}$, est donnée par :

$$d^2\left(\mathbf{e}_i, \mathbf{e}_{i'}\right) = \sum_{j=1}^{p} \left(x_{ij} - x_{i'j}\right)^2.$$

Le choix de la distance euclidienne permet de donner le même poids à chacune des variables. Cependant, il est possible de considérer une métrique définie par une matrice symétrique définie positive $\mathbf{M} \in \mathcal{M}\left(p \times p\right)$ de la façon suivante:

$$d^2\left(\mathbf{e}_i, \mathbf{e}_{i'}\right) = \left(\mathbf{e}_i - \mathbf{e}_{i'}\right)^t \mathbf{M} \left(\mathbf{e}_i - \mathbf{e}_{i'}\right).$$

La distance euclidienne classique correspond donc au cas $\mathbf{M} = \mathbf{Id}_p$

Inertie

On définit ensuite la notion d'inertie totale:

$$I_T = \frac{1}{n}\sum_{i=1}^{n} d^2\left(\mathbf{e}_i, \mathbf{g}\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(x_{ij} - g_j\right)^2.$$

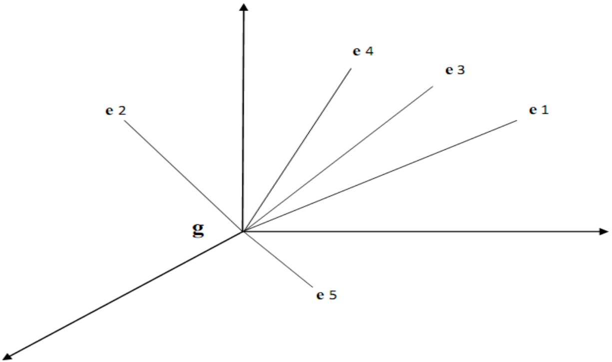

Fig. 3

Inertie

L'inertie *mesure la dispersion des points du nuage* par rapport à son centre de gravité: plus l'inertie est grande, plus le nuage est dispersé, et à l'inverse plus l'inertie est petite, plus le nuage est concentré autour de son centre de gravité. L'inertie du nuage de points I_T s'interprète facilement en termes de variance. En effet, on a :

$$\begin{aligned} I_T &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(x_{ij}-g_j\right)^2 = \sum_{j=1}^{p}\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}-g_j\right)^2\right] \\ &= \sum_{j=1}^{p}\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}-\frac{1}{n}\sum_{i=1}^{n}x_{ij}\right)^2\right], \end{aligned}$$

ainsi on a

$$I_T = \sum_{j=1}^{p} s_j^2 = \text{Trace } \mathbf{V}.$$

Inertie par rapport à un sous-espace vectoriel

Notons $\mathbf{E}$ un sous-espace vectoriel passant par le centre de gravité $\mathbf{g}$. L'inertie du nuage de points par rapport à ce sous-espace vectoriel $\mathbf{E}$ est définie par:

$$I_{\mathbf{E}} = \frac{1}{n} \sum_{i=1}^{n} d^2 \left(\mathbf{e}_i, \mathbf{Proj}_{\mathbf{E},i} \right),$$

où $\mathbf{Proj}_{\mathbf{E},i}$ désigne la projection orthogonale de $\mathbf{e}_i$ sur $\mathbf{E}$.

Décomposition de l'inertie

Considérons une décomposition de $\mathbb{R}^p$ en deux sous-espaces vectoriels supplémentaires, orthogonaux, et passant par et passant par $\mathbf{g}$, c'est à dire

$$\mathbb{R}^p = \mathbf{E} \oplus \mathbf{E}^\perp,$$

où $\mathbf{E}^\perp$ désigne l'espace orthogonal à $\mathbf{E}$. On peut montrer facilement que :

$$I_T = I_E + I_{E^\perp}. \tag{1}$$

En effet, par le théorème de Pythagore, on a

$$d^2\left(\mathbf{e}_i, \mathbf{Proj}_{E,i}\right) + d^2\left(\mathbf{e}_i, \mathbf{Proj}_{E^\perp,i}\right) = d^2\left(\mathbf{e}_i, \mathbf{g}\right).$$

Décomposition de l'inertie

Si on note $\mathbf{u}$ le vecteur directeur unitaire de l'axe $\mathbf{E}$, i.e. $\|\mathbf{u}\| = 1$, on a l'expression matricielle suivante pour l'inertie expliquée par cet axe :

$$I_{E^{\perp}} = \mathbf{u}^{t}\mathbf{V}\mathbf{u} = \langle \mathbf{u}, \mathbf{V}\mathbf{u} \rangle . \tag{2}$$

En effet,

$$\begin{aligned}
I_{E^\perp} &= \frac{1}{n}\sum_{i=1}^{n} d^2\left(\mathbf{e}_i, \mathbf{Proj}_{E^\perp,i}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \left\|\mathbf{e}_i - \mathbf{Proj}_{E^\perp,i}\right\|^2 = \frac{1}{n}\sum_{i=1}^{n} \left\|\mathbf{Proj}_{E,i}\right\|^2 \text{(relation de Chasles)} \\
&= \frac{1}{n}\sum_{i=1}^{n} \left\|\frac{\langle \mathbf{e}_i, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}\mathbf{u}\right\|^2 = \frac{1}{n}\sum_{i=1}^{n} \left\|\langle \mathbf{e}_i, \mathbf{u}\rangle \mathbf{u}\right\|^2 \quad \left(\text{car } \|\mathbf{u}\| = 1\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \langle \mathbf{e}_i, \mathbf{u}\rangle^2 \|\mathbf{u}\|^2 = \frac{1}{n}\sum_{i=1}^{n} \langle \mathbf{e}_i, \mathbf{u}\rangle^2 = \frac{1}{n}\sum_{i=1}^{n} \left(\mathbf{e}_i^t \mathbf{u}\right)^2.
\end{aligned}$$

Décomposition de l'inertie

Comme $\mathbf{e}_i^t\mathbf{u}$ est un scalaire, alors $\mathbf{e}_i^t\mathbf{u} = \left(\mathbf{e}_i^t\mathbf{u}\right)^t = \mathbf{u}^t\mathbf{e}_i$, par conséquent

$$I_{E^\perp} = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{u}^t\mathbf{e}_i\right)\left(\mathbf{e}_i^t\mathbf{u}\right) = \mathbf{u}^t\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{e}_i\mathbf{e}_i^t\right]\mathbf{u}.$$

D'après la notion du produit matricielle:

$$\mathbf{X}^t\mathbf{X} = \sum_{i=1}^{n}\mathbf{e}_i\mathbf{e}_i^t,$$

ainsi

$$I_{E^\perp} = \mathbf{u}^t\left[\frac{1}{n}\mathbf{X}^t\mathbf{X}\right]\mathbf{u} = \mathbf{u}^t\mathbf{V}\mathbf{u}.$$

Décomposition de l'inertie

La décomposition de l'inertie obtenue en (1) est connue sous le nom de **Théorème de Huygens**. On peut le généraliser au cas d'une décomposition de $\mathbb{R}^p$ comme somme de p sous-espaces de dimension 1 et orthogonaux entre eux:

$$\mathbb{R}^p = \mathbf{E}_1 \oplus \mathbf{E}_2 \oplus ... \oplus \mathbf{E}_p.$$

Ainsi

$$I_T = I_{E_1^\perp} + I_{E_2^\perp} + ... + I_{E_p^\perp}.$$

Recherche du premier axe principal

On cherche donc un axe $\mathbf{E}_1$, passant par $\mathbf{g}$, et tel que I_{E_1} soit minimum, ou de façon équivalente tel que $I_{E_1^\perp}$ soit maximum. Notons $\mathbf{u}_1$ le vecteur directeur unitaire de l'axe $\mathbf{E}_1$. On cherche donc $\mathbf{u}_1$ tel que $I_{E_1^\perp}$ soit maximum sous la contrainte $\left\|\mathbf{u}_1\right\|^2 = 1$. On obtient alors le problème d'optimisation sous contrainte suivant :

$$\left\{\begin{array}{l} \max_{\mathbf{u}_1} I_{E_1^\perp} \\ \left\|\mathbf{u}_1\right\|^2 = 1 \end{array}\right. \iff \left\{\begin{array}{l} \max_{u_1} \mathbf{u}_1^t \mathbf{V} \mathbf{u}_1 \\ \mathbf{u}_1^t \mathbf{u}_1 = 1 \end{array}\right. .$$

Recherche du premier axe principal

Pour résoudre ce problème de maximisation sous contraintes, on utilise la méthode des multiplicateurs de Lagrange. On doit alors trouver $\mathbf{u}_1$ vérifiant

$$\frac{\partial}{\partial \mathbf{u}_1}\left(\mathbf{u}_1^t \mathbf{V} \mathbf{u}_1 - \lambda_1\left(\mathbf{u}_1^t \mathbf{u}_1 - 1\right)\right) = 0.$$

En utilisant la dérivée matricielle

$$2\mathbf{V}\mathbf{u}_1 - 2\lambda_1 \mathbf{u}_1 = 0 \iff \mathbf{V}\mathbf{u}_1 = \lambda_1 \mathbf{u}_1.$$

Recherche du premier axe principal

Autrement dit, $\mathbf{u}_1$ est un vecteur propre de la matrice de variance-covariance des données $\mathbf{V}$. Or, d'après l'équation (2) on a:

$$I_{E_1^\perp} = \mathbf{u}_1^t \mathbf{V} \mathbf{u}_1 = \mathbf{u}_1^t (\lambda_1 \mathbf{u}_1) = \lambda_1 \mathbf{u}_1^t \mathbf{u}_1 = \lambda_1 \|\mathbf{u}_1\|^2 = \lambda_1.$$

Comme on cherche à maximiser $I_{E_1^\perp}$ on va également chercher à maximiser λ_1 et on choisira donc la plus grande valeur propre de la matrice $\mathbf{V}$.

Finalement, le vecteur directeur unitaire de l'axe $\mathbf{E}_1$ est le vecteur propre associé à la plus grande valeur propre de la matrice $\mathbf{V}$.

Recherche des axes suivants

Une fois que le premier axe a été identifié, on cherche l'axe $\mathbf{E}_2$, orthogonal à $\mathbf{E}_1$ et tel que l'inertie I_{E_2} soit minimale, ou de façon équivalent tel que $I_{E_2^{\perp}}$ soit maximale. En notant $\mathbf{u}_2$ le vecteur directeur unitaire de $\mathbf{E}_2$, on doit alors résoudre le système suivant :

$$\left\{ \begin{array}{l} \max_{\mathbf{u}_2} \mathbf{u}_2^t \mathbf{V} \mathbf{u}_2 \\ \left\| \mathbf{u}_2 \right\|^2 = \mathbf{u}_2^t \mathbf{u}_2 = 1 \\ \mathbf{u}_2 \perp \mathbf{u}_1 \Longleftrightarrow \langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \mathbf{u}_2^t \mathbf{u}_1 = 0. \end{array} \right. .$$

Recherche des axes suivants

Par rapport au cas précédent, on a donc une contrainte de plus car le nouvel axe doit être orthogonal au premier. On utilise la encore la méthode des multiplicateurs de Lagrange. On cherche donc $\mathbf{u}_2$ qui vérifiant

$$\frac{\partial}{\partial \mathbf{u}_2}\left(\mathbf{u}_2^t \mathbf{V} \mathbf{u}_2 - \lambda_2\left(\mathbf{u}_2^t \mathbf{u}_2 - 1\right) - \mu \mathbf{u}_2^t \mathbf{u}_1\right) = 0,$$

c'est-à-dire :

$$2\mathbf{V}\mathbf{u}_2 - 2\lambda_2 \mathbf{u}_2 - \mu \mathbf{u}_1 = 0. \tag{3}$$

En multipliant à gauche par $\mathbf{u}_1^t$ et en utilisant le fait que $\mathbf{u}_1$ et $\mathbf{u}_2$ sont orthogonaux, et que $\mathbf{u}_1$ est de norme 1, on obtient :

$$\begin{aligned}
2\mathbf{u}_1^t\mathbf{V}\mathbf{u}_2 - 2\lambda_2\mathbf{u}_1^t\mathbf{u}_2 - \mu\mathbf{u}_1^t\mathbf{u}_1 &= 0 \Longrightarrow 2\mathbf{u}_1^t\mathbf{V}\mathbf{u}_2 - \mu\mathbf{u}_1^t\mathbf{u}_1 = 0 \\
&\Longrightarrow 2\left(\mathbf{V}^t\mathbf{u}_1\right)^t\mathbf{u}_2 - \mu\mathbf{u}_1^t\mathbf{u}_1 = 0 \\
&\Longrightarrow 2\left(\mathbf{V}\mathbf{u}_1\right)^t\mathbf{u}_2 - \mu\mathbf{u}_1^t\mathbf{u}_1 = 0 \\
&\Longrightarrow 2\lambda_1\mathbf{u}_1^t\mathbf{u}_2 - \mu\mathbf{u}_1^t\mathbf{u}_1 = 0 \\
&\Longrightarrow \mu = 0.
\end{aligned}$$

Recherche des axes suivants

Finalement, en reprenant l'équation (3), on obtient :

$$\mathbf{V}\mathbf{u}_2 = \lambda_2 \mathbf{u}_2.$$

Le vecteur directeur unitaire de $\mathbf{E}_2$ est donc également un vecteur propre de $\mathbf{V}$. Comme on cherche à maximiser l'inertie $I_{\mathbf{E}_2^\perp}$ on va raisonner de même que pour le premier axe et maximiser λ_2. On choisira donc pour $\mathbf{u}_2$, le vecteur propre associé à la deuxième plus grande valeur propre de $\mathbf{V}$.

On raisonne de même pour trouver les axes suivants, dont les vecteurs directeurs unitaires sont tous des vecteurs propres de la matrice $\mathbf{V}$, associés aux valeurs propres ordonnées par ordre décroissant

Recherche des axes suivants

La matrice $\mathbf{V}$ étant symétrique, elle possède bien p vecteurs propres qui forment une base orthogonale de $\mathbb{R}^p$.

Les axes $\mathbf{E}_1$, $\mathbf{E}_2, ..., \mathbf{E}_p$ sont appelées axes factoriels, ou axes principaux d'inertie.

Contribution des axes à l'inertie totale

Par le théorème de Huygens, on a :

$$I_T = I_{\mathbf{E}_1^\perp} + I_{\mathbf{E}_2^\perp} + ... + I_{\mathbf{E}_p^\perp} = \lambda_1 + \lambda_2 + ... + \lambda_p.$$

L'inertie expliquée par l'axe $\mathbf{E}_j$ est :

$$I_{\mathbf{E}_j^\perp} = \lambda_j,$$

et l**e pourcentage d'inertie expliquée** par cet axe, aussi appelé **contribution relative**, est égale à :

$$\frac{\lambda_j}{\lambda_1 + \lambda_2 + ... + \lambda_p}.$$

Contribution des axes à l'inertie totale

De la même façon, on peut définir le **pourcentage d'inertie expliquée par le sous-espace**

$$\mathbf{F}_h = \mathbf{E}_1 \oplus \mathbf{E}_2 \oplus ... \oplus \mathbf{E}_h$$

par:

$$\frac{\lambda_1 + \lambda_2 + ... + \lambda_h}{\lambda_1 + \lambda_2 + ... + \lambda_p}.$$

Représentation des individus dans les nouveaux axes

Notons y_{ij} la coordonnée de l'individu i selon l'axe $\mathbf{E}_j$. Par définition, on a :

$$y_{ij} = \langle \mathbf{e}_i, \mathbf{u}_j \rangle = \mathbf{e}_i^t \mathbf{u}_j.$$

En notant $\mathbf{y}_i$ le vecteur des coordonnées de l'individu i dans la nouvelle base et $\mathbf{U}$ la matrice des vecteurs propres, on obtient alors :

$$\mathbf{y}_i^t = \mathbf{e}_i^t \mathbf{U}.$$

En notant $\mathbf{Y}$ la matrice contenant en ligne les coordonnées des individus dans la nouvelle base, on a :

$$\mathbf{Y} = \mathbf{X}\mathbf{U}, \tag{4}$$

Qualité de représentation des individus

Le cosinus au carrée de l'angle α_{ij} entre $\mathbf{e}_i$ et $\mathbf{u}_j$ est donné par :

$$\cos^2 \alpha_{ij} = \frac{\langle \mathbf{e}_i, \mathbf{u}_j \rangle^2}{\|\mathbf{e}_i\|^2} = \frac{(\mathbf{e}_i^t \mathbf{u}_j)^2}{\|\mathbf{e}_i\|^2} = \frac{y_{ij}^2}{\|\mathbf{e}_i\|^2}. \tag{5}$$

De même, on peut calculer le cosinus au carrée de l'angle $\alpha_{ijj'}$ entre e_i et le sous-espace engendré par les axes $\mathbf{E}_j$ et $\mathbf{E}_{j'}$ par:

$$\cos^2 \alpha_{ijj'} = \cos^2 \alpha_{ij} + \cos^2 \alpha_{ij'}.$$

Comme $\|\mathbf{e}_i\| = \|\mathbf{y}_i\|$ (il s'agit simplement d'un changement de base, la norme du vecteur étant inchangée), on a

$$\sum_{j=1}^{p} \cos^2 \alpha_{ij} = 1.$$

Interprétation des nouveaux axes en fonction des individus

On rappelle que l'inertie expliquée par un axe $\mathbf{E}_j$ est:

$$I_{\mathbf{E}_j^\perp} = \frac{1}{n}\sum_{i=1}^{n} d^2\left(\mathbf{e}_i, \mathrm{Pr}\,\mathbf{oj}_{\mathbf{E}_j^\perp,i}\right) = \frac{1}{n}\sum_{i=1}^{n} \langle \mathbf{e}_i, \mathbf{u}_j \rangle^2 = \lambda_j,\ \ j = 1, ..., p.$$

Comme tous les individus ont le même poids, ils contribuent chacun à cette inertie à hauteur de

$$\mathrm{ca}\,(i, j) = \frac{1}{n} y_{ij}^2.$$

On parle de *contribution absolue*. On peut aussi s'intéresser à la *contribution relative* d'un individu à un axe $\mathbf{E}_j$. L'inertie portée par l'axe étant λ_j, la contribution relative est donnée par:

$$\mathrm{cr}\,(i, j) = \frac{\mathrm{ca}\,(i, j)}{\lambda_j} = \frac{y_{ij}^2}{n\lambda_j}.$$

Représentation des variables

- La recherche des axes factoriels nous a permis d'identifier une nouvelle base orthonormée de l'espace $\mathbb{R}^p$.
- Les nouveaux axes obtenus sont des combinaisons linéaires des axes initiaux, et on peut donc associer aux variables initiales les combinaisons linéaires correspondantes afin de former de "nouvelles variables", appelées *composantes principales*.
- On peut alors s'intéresser à l'interprétation de ces "nouvelles variables", en cherchant à les relier aux "anciennes variables".

Definition

On note $\mathbf{c}_1, ..., \mathbf{c}_p$ les composantes principales, avec $\mathbf{c}_k$ la composante principale correspond á l'axe $\mathbf{E}_k$. On a

$$\mathbf{c}_k = \sum_{j=1}^{p} u_{kj}\mathbf{X}_j = \mathbf{X}\mathbf{u}_k.$$

Proposition. Les composantes principales vérifient :
- $E[\mathbf{c}_j] = 0$
- $\mathbf{Var}\,[\mathbf{c}_j] = \lambda_j$
- $\mathbf{Cov}\,[\mathbf{c}_j\mathbf{c}_k] = 0,\ j \neq k.$

Les composantes principales sont centrées comme combinaisons linéaires de variables centrées (rappelons que l'on a supposé que le nuage de points était centré). On se place maintenant dans $\mathbb{R}^n$ pour calculer la covariance entre deux composantes principales:

$$\begin{aligned}
\mathbf{Cov}\left[\mathbf{c}_k, \mathbf{c}_j\right] &= \frac{1}{n}\left(\mathbf{X}\mathbf{u}_k\right)^t\left(\mathbf{X}\mathbf{u}_j\right) \\
&= \frac{1}{n}\mathbf{u}_k^t\mathbf{X}^t\mathbf{X}\mathbf{u}_j = \mathbf{u}_k^t\mathbf{V}\mathbf{u}_j \\
&= \lambda_j\mathbf{u}_k^t\mathbf{u}_j = \left\{\begin{array}{cc} 0 & \text{si } k \neq j \\ \lambda_j & \text{si } k = j \end{array}\right\}.
\end{aligned}$$

Qualité de la représentation et cercle des corrélations

Afin d'étudier le lien entre les composantes principales et les variables initiales, on s'intéresse aux corrélations entre les nouvelles et les anciennes variables. On définit d'abord la covariance entre une variable initiale X_j et une composante principale $\mathbf{c}_k$:

$$\begin{aligned}
\mathbf{Cov}\left(\mathbf{X}_j, \mathbf{c}_k\right) &= \frac{1}{n}\mathbf{c}_k^t\mathbf{X}_j = \frac{1}{n}\left(\mathbf{X}\mathbf{u}_k\right)^t\mathbf{X}_j = \frac{1}{n}\mathbf{u}_k^t\mathbf{X}^t\mathbf{X}_j \\
&= \frac{1}{n}\mathbf{u}_k^t\mathbf{X}^t\mathbf{X}\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{u}_k^t\mathbf{V}\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \\
&= \lambda_k\mathbf{u}_k^t\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \end{pmatrix} = \lambda_k u_{kj}.
\end{aligned}$$

Qualité de la représentation et cercle des corrélations

On en déduit la corrélation entre X_j et $\mathbf{c}_k$:

$$\mathbf{COR}\left[\mathbf{X}_j, \mathbf{c}_k\right] = \frac{\mathbf{Cov}\left(\mathbf{X}_j, \mathbf{c}_k\right)}{\sqrt{\mathbf{Var}\left(\mathbf{X}_j\right)\mathbf{Var}\left(\mathbf{c}_k\right)}} = \frac{\lambda_k u_{kj}}{\sqrt{s_j^2 \lambda_k}} = \sqrt{\lambda_k}\frac{u_{kj}}{s_j}.$$

Sous forme matricielle, on a

$$\mathbf{COR} = \mathbf{D}_{\sqrt{\lambda}}\mathbf{U}^t\mathbf{D}_{\mathbf{1}/\sigma} \in \mathcal{M}\left(p \times p\right).$$

Si les variables sont réduites, alors

$$\mathbf{COR}\left[\mathbf{X}_j, \mathbf{c}_k\right] = \sqrt{\lambda_k} u_{kj}.$$

Sous forme matricielle, on a

$$\mathbf{COR} = \mathbf{D}_{\sqrt{\lambda}}\mathbf{U} \in \mathcal{M}\left(p \times p\right).$$

Théorème. Nous avons

$$\sum_{k=1}^{p} \mathbf{COR}^2 \left[\mathbf{X}_j, \mathbf{c}_k \right] = \sum_{k=1}^{p} \cos^2 \alpha_{jk} = 1,\ j = 1, ..., p,$$

où α_{jk} est l'angle entre les deux vecteurs $\mathbf{X}_j$ et $\mathbf{c}_k$.

Preuve. D'après la décomposition spectrale $(\mathbf{??})$, on a $\mathbf{V} = \sum_{k=1}^{p} \lambda_k \mathbf{u}_k \mathbf{u}_k^t$. Ceci implique, en particulier, l'égalité des diagonaux des deux matrices $\mathbf{V}$ et $\sum_{k=1}^{p} \lambda_k \mathbf{u}_k \mathbf{u}_k^t$, ainsi on a

$$\sigma_j^2 = \mathbf{Var} \left(\mathbf{X}_j \right) = \sum_{k=1}^{p} \lambda_k u_{kj}^2,\ j = 1, ..., p.$$

Qualité de la représentation et cercle des corrélations

Par conséquent

$$\sum_{k=1}^{p} \mathbf{COR}^2 \left[\mathbf{X}_j, \mathbf{c}_k \right] = \sum_{k=1}^{p} \lambda_k \frac{u_{kj}^2}{s_j^2} = \frac{\sum_{k=1}^{p} \lambda_k u_{kj}^2}{s_j^2} = \frac{s_j^2}{s_j^2} = 1. \,(\mathrm{CQFD})$$

Remarque. Si

$$\mathbf{COR}^2\left[\mathbf{X}_j, \mathbf{c}_1\right] + \mathbf{COR}^2\left[\mathbf{X}_j, \mathbf{c}_2\right] \simeq 1$$

alors

$$\sum_{k=3}^{p} \mathbf{COR}^2\left[\mathbf{X}_j, \mathbf{c}_k\right] \simeq 0,$$

et par conséquent $\mathbf{COR}\left[\mathbf{X}_j, \mathbf{c}_k\right]$, $j = 3, .., p$ sont aussi proches de zéro. Ceci explique que la variable $\mathbf{X}_j$ a une forte corrélation avec les premières composantes principales et non corrélées avec le reste des composantes.

Qualité de la représentation et cercle des corrélations

- La quantité $\mathbf{COR}\left[\mathbf{X}_j, \mathbf{c}_k\right]$ donne la qualité de représentation de la variable j sur l'axe $\mathbf{E}_k$.
- Plus elle est proche de 1 en valeur absolue, plus la variable est bien représentée par l'axe k.
- Le signe de la corrélation permet de savoir si la variable contribue positivement ou négativement à la définition de l'axe k.
- On représente souvent graphiquement ces corrélations en dimension 2, à l'aide de ce que l'on appelle un cercle des corrélations.
- Plus précisément, on choisit deux axes $\mathbf{E}_j$ et $\mathbf{E}_k$, et l'on trace les points dont les coordonnées sont les corrélations de chacune des variables avec les j-ème et k-ème composantes principales. Un exemple est fourni dans la figure 1.1.

Qualité de la représentation et cercle des corrélations