

# Notes de cours

22 janvier 2024

## 1 Inertie : changement de point de repère

On cherche à comparer  $I_{\mathbf{g}}$  et  $I_{\mathbf{a}}$ , pour un vecteur  $\mathbf{a}$  quelconque.

$$\begin{aligned} I_{\mathbf{a}} &= \sum_{i=1}^n p_i \|\mathbf{e}_i - \mathbf{a}\|_{\mathbf{M}}^2 = \sum_{i=1}^n p_i \|\mathbf{e}_i - \mathbf{g} + \mathbf{g} - \mathbf{a}\|_{\mathbf{M}}^2 \\ &= \sum_{i=1}^n p_i \|\mathbf{e}_i - \mathbf{g}\|_{\mathbf{M}}^2 + \sum_{i=1}^n p_i \|\mathbf{g} - \mathbf{a}\|_{\mathbf{M}}^2 + 2 \times \left\langle \sum_{i=1}^n p_i (\mathbf{e}_i - \mathbf{g}), \mathbf{g} - \mathbf{a} \right\rangle_{\mathbf{M}} \\ &= I_{\mathbf{g}} + \|\mathbf{g} - \mathbf{a}\|_{\mathbf{M}}^2 + 2 \times \langle 0, \mathbf{g} - \mathbf{a} \rangle_{\mathbf{M}} = I_{\mathbf{g}} + \|\mathbf{g} - \mathbf{a}\|_{\mathbf{M}}^2. \end{aligned}$$

On voit donc que  $I_{\mathbf{a}}$  n'est différent de  $I_{\mathbf{g}}$  que par une valeur fixe. Il ne présente pas d'intérêt.

## 2 Seconde forme de l'inertie

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n p_i p_j \|\mathbf{e}_i - \mathbf{e}_j\|_{\mathbf{M}}^2 &= \sum_{i=1}^n \sum_{j=1}^n p_i p_j \left[ \|\mathbf{e}_i - \mathbf{g}\|_{\mathbf{M}}^2 + \|\mathbf{g} - \mathbf{e}_j\|_{\mathbf{M}}^2 - 2 \langle \mathbf{e}_i - \mathbf{g}, \mathbf{e}_j - \mathbf{g} \rangle_{\mathbf{M}} \right] \\ &= 2 \sum_{i=1}^n p_i \|\mathbf{e}_i - \mathbf{g}\|_{\mathbf{M}}^2 - 2 \left\langle \sum_{i=1}^n p_i (\mathbf{e}_i - \mathbf{g}), \sum_{j=1}^n p_j (\mathbf{e}_j - \mathbf{g}) \right\rangle_{\mathbf{M}} \\ &= 2I_{\mathbf{g}} - 2 \times \langle 0, 0 \rangle_{\mathbf{M}}. \end{aligned}$$

## 3 Inertie et trace

$$I_{\mathbf{g}} = \sum_{i=1}^n p_i \|\mathbf{e}_i - \mathbf{g}\|_{\mathbf{M}}^2 = \sum_{i=1}^n p_i (\mathbf{e}_i - \mathbf{g})' \mathbf{M} (\mathbf{e}_i - \mathbf{g}).$$

D'autre part, comme  $I_{\mathbf{g}}$  est un nombre,

$$\begin{aligned} I_{\mathbf{g}} &= \text{Tr}(I_{\mathbf{g}}) = \sum_{i=1}^n p_i \text{Tr}[(\mathbf{e}_i - \mathbf{g})' \mathbf{M} (\mathbf{e}_i - \mathbf{g})] = \sum_{i=1}^n p_i \text{Tr}[(\mathbf{e}_i - \mathbf{g})(\mathbf{e}_i - \mathbf{g})' \mathbf{M}] \\ &= \text{Tr} \left\{ \left[ \sum_{i=1}^n p_i (\mathbf{e}_i - \mathbf{g})(\mathbf{e}_i - \mathbf{g})' \right] \mathbf{M} \right\} = \text{Tr}(\mathbf{V}\mathbf{M}). \end{aligned}$$

## 4 Propriétés spectrales de $\mathbf{V}\mathbf{M}$

Soit  $\mathbf{T} = \text{diag}(\sqrt{m_j})$  la matrice diagonale qui vérifie  $\mathbf{T}^2 = \mathbf{M}$ .  $\mathbf{T}\mathbf{V}\mathbf{T}$  est une matrice symétrique et possède donc  $p$  valeurs propres réelles que l'on classe arbitrairement par ordre décroissant

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_p,$$

et  $p$  vecteurs propres réels orthonormés  $\mathbf{v}_1, \dots, \mathbf{v}_p$  associés. De plus,  $\mathbf{T}\mathbf{V}\mathbf{T}$  est une matrice semi-définie puisque, pour tout vecteur  $\mathbf{v}$

$$\begin{aligned} \mathbf{v}' \mathbf{T}\mathbf{V}\mathbf{T}\mathbf{v} &= \sum_{j=1}^p \sum_{\ell=1}^p v_j v_{\ell} \sqrt{m_j} \sqrt{m_{\ell}} \text{cov}(\mathbf{x}^j, \mathbf{x}^{\ell}). \\ &= \text{cov} \left( \sum_{j=1}^p v_j \sqrt{m_j} \mathbf{x}^j, \sum_{\ell=1}^p v_{\ell} \sqrt{m_{\ell}} \mathbf{x}^{\ell} \right) = \text{var} \left( \sum_{j=1}^p v_j \sqrt{m_j} \mathbf{x}^j \right) \geq 0 \end{aligned}$$

et ses valeurs propres sont donc positives ou nulles. Si on pose  $\mathbf{a}_k = \mathbf{T}^{-1} \mathbf{v}_k$ , on a alors

$$\begin{aligned} \mathbf{T}\mathbf{V}\mathbf{T}\mathbf{v}_k &= \lambda_k \mathbf{v}_k, \\ \mathbf{V}\mathbf{T}\mathbf{v}_k &= \lambda_k \mathbf{T}^{-1} \mathbf{v}_k, \\ \mathbf{V}\mathbf{M}\mathbf{T}^{-1} \mathbf{v}_k &= \lambda_k \mathbf{T}^{-1} \mathbf{v}_k, \\ \mathbf{V}\mathbf{M}\mathbf{a}_k &= \lambda_k \mathbf{a}_k. \end{aligned}$$

Les  $\mathbf{a}_k$  sont  $\mathbf{M}$ -orthonormés car

$$\langle \mathbf{a}_k, \mathbf{a}_{\ell} \rangle_{\mathbf{M}} = \mathbf{a}'_k \mathbf{M} \mathbf{a}_{\ell} = \mathbf{v}'_k \mathbf{T}^{-1} \mathbf{M} \mathbf{T}^{-1} \mathbf{v}_{\ell} = \mathbf{v}'_k \mathbf{v}_{\ell} = \langle \mathbf{v}_k, \mathbf{v}_{\ell} \rangle.$$

## 5 Preuve partielle du théorème sur l'ACP

On va se contenter de regarder le résultat en dimension 1, c'est-à-dire en projetant sur un axe.

**Expression de la projection** Tout d'abord, on cherche la définition de la projection  $\mathbf{M}$ -orthogonale sur l'axe défini par le vecteur  $\mathbf{a}$ . L'opérateur  $\mathbf{P}$  de projection satisfait pour tout vecteur  $\mathbf{v}$

$$\mathbf{P}\mathbf{v} = \mu\mathbf{a}, \quad (\mathbf{v} - \mathbf{P}\mathbf{v})'\mathbf{M}\mathbf{a} = 0.$$

On en déduit que  $\mathbf{v}'\mathbf{M}\mathbf{a} - \mu\mathbf{a}'\mathbf{M}\mathbf{a} = 0$  et donc  $\mu = \mathbf{a}'\mathbf{M}\mathbf{v}/\mathbf{a}'\mathbf{M}\mathbf{a}$ . Finalement,

$$\mathbf{P} = \frac{\mathbf{a}\mathbf{a}'\mathbf{M}}{\mathbf{a}'\mathbf{M}\mathbf{a}}$$

On vérifie que  $\mathbf{P}^2 = \mathbf{P}$  et  $\mathbf{P}'\mathbf{M} = \mathbf{M}\mathbf{P}$ , ce qui est vrai pour toutes les projections  $\mathbf{M}$ -orthogonales.

**L'inertie du nuage projeté** En reprenant le calcul « Inertie et trace », on trouve pour le nuage projeté  $\mathbf{X}\mathbf{P}$

$$\begin{aligned} I_{\mathbf{g}}(\mathbf{X}\mathbf{P}) &= \sum_{i=1}^n p_i \|\mathbf{P}(\mathbf{e}_i - \mathbf{g})\|_{\mathbf{M}}^2 = \sum_{i=1}^n p_i (\mathbf{e}_i - \mathbf{g})'\mathbf{P}'\mathbf{M}\mathbf{P}(\mathbf{e}_i - \mathbf{g}) \\ &= \sum_{i=1}^n p_i (\mathbf{e}_i - \mathbf{g})'\mathbf{M}\mathbf{P}(\mathbf{e}_i - \mathbf{g}) = \dots = \text{Tr}(\mathbf{V}\mathbf{M}\mathbf{P}) \\ &= \frac{\text{Tr}(\mathbf{V}\mathbf{M}\mathbf{a}\mathbf{a}'\mathbf{M})}{\mathbf{a}'\mathbf{M}\mathbf{a}} = \frac{\text{Tr}(\mathbf{a}'\mathbf{M}\mathbf{V}\mathbf{M}\mathbf{a})}{\mathbf{a}'\mathbf{M}\mathbf{a}} = \frac{\mathbf{a}'\mathbf{M}\mathbf{V}\mathbf{M}\mathbf{a}}{\mathbf{a}'\mathbf{M}\mathbf{a}}, \end{aligned}$$

où on a utilisé la relation  $\mathbf{P}'\mathbf{M}\mathbf{P} = \mathbf{M}\mathbf{P}\mathbf{P} = \mathbf{M}\mathbf{P}$  à la seconde ligne.

**Maximisation de l'inertie** On cherche le  $\mathbf{a}$  qui maximise l'inertie. Pour cela on doit dériver  $I_{\mathbf{g}}(\mathbf{X}\mathbf{P})$  par rapport à  $\mathbf{a}$ , dans le sens

$$\frac{\partial I_{\mathbf{g}}(\mathbf{X}\mathbf{P})}{\partial \mathbf{a}} = \begin{pmatrix} \frac{\partial I_{\mathbf{g}}(\mathbf{X}\mathbf{P})}{\partial a_1} \\ \vdots \\ \frac{\partial I_{\mathbf{g}}(\mathbf{X}\mathbf{P})}{\partial a_p} \end{pmatrix}$$

Il est facile de se persuader que pour toute matrice  $\mathbf{G}$ ,

$$\frac{\partial}{\partial \mathbf{a}} (\mathbf{a}'\mathbf{G}\mathbf{a}) = 2\mathbf{G}\mathbf{a},$$

et donc

$$\frac{\partial I_{\mathbf{g}}(\mathbf{X}\mathbf{P})}{\partial \mathbf{a}} = \frac{\mathbf{a}'\mathbf{M}\mathbf{a} \cdot 2\mathbf{M}\mathbf{V}\mathbf{M}\mathbf{a} - \mathbf{a}'\mathbf{M}\mathbf{V}\mathbf{M}\mathbf{a} \cdot 2\mathbf{M}\mathbf{a}}{(\mathbf{a}'\mathbf{M}\mathbf{a})^2} = 0,$$

ce qui implique

$$\mathbf{M}\mathbf{V}\mathbf{M}\mathbf{a} = \frac{\mathbf{a}'\mathbf{M}\mathbf{V}\mathbf{M}\mathbf{a}}{\mathbf{a}'\mathbf{M}\mathbf{a}} \mathbf{M}\mathbf{a},$$

et comme  $\mathbf{M}$  est inversible,  $\mathbf{a}$  est vecteur propre de  $\mathbf{V}\mathbf{M}$  associé à la valeur propre  $\lambda = I_{\mathbf{g}}(\mathbf{X}\mathbf{P})$  :

$$\mathbf{V}\mathbf{M}\mathbf{a} = \lambda\mathbf{a}.$$

Finalement, l'axe qui maximise l'inertie projetée est le vecteur  $\mathbf{a}_1$  défini en section 4.

## 6 Variables liées et valeurs propres

On cherche à montrer qu'à chaque fois que des variables sont liées par une relation linéaire, l'analyse en composante principale des données correspondantes produit un vecteur  $\mathbf{u}$  associé une valeur propre nulle. On considère pour cela une table  $\mathbf{X} = (x_i^j)$  de données avec  $n$  individus et  $p$  variables, ainsi que sa version centrée réduite  $\mathbf{Z} = (z_i^j)$ , où comme d'habitude  $z_i^j = (x_i^j - \bar{x}^j)/\sigma_j$ ,  $\bar{x}^j$  est la moyenne arithmétique de la variable  $j$  et  $\sigma_j$  son écart type. Pour simplifier, on supposera que toutes les variables ont le même poids  $1/n$ . Dans ce cas, on rappelle que la matrice de corrélation de  $\mathbf{X}$  s'écrit  $\mathbf{R} = \frac{1}{n}\mathbf{Z}'\mathbf{Z}$ .

Dans la suite, on suppose qu'il existe des coefficients non tous nuls  $v_1, \dots, v_p$  et  $v_0$  tels que, pour tout  $i$ ,

$$v_1 x_i^1 + v_2 x_i^2 + \dots + v_p x_i^p = v_0.$$

Les moyennes arithmétiques des variables vérifient alors

$$\begin{aligned} v_1 \bar{x}^1 + v_2 \bar{x}^2 + \dots + v_p \bar{x}^p &= v_1 \frac{1}{n} \sum_{i=1}^n x_i^1 + \dots + v_p \frac{1}{n} \sum_{i=1}^n x_i^p \\ &= \frac{1}{n} \sum_{i=1}^n (v_1 x_i^1 + \dots + v_p x_i^p) \\ &= \frac{1}{n} \sum_{i=1}^n v_0 = v_0. \end{aligned}$$

Les variables centrées vérifient pour chaque individu  $i$

$$v_1 (x_i^1 - \bar{x}^1) + v_2 (\bar{x}^2 - x_i^2) + \dots + v_p (\bar{x}^p - x_i^p) = v_0 - v_0 = 0,$$

et comme  $x_i^j - \bar{x}^j = \sigma_j z_i^j$ , pour les variables centrées réduites on obtient

$$v_1 s_1 z_i^1 + v_2 s_2 z_i^2 + \dots + v_p s_p z_i^p = 0.$$

Un facteur propre non nul  $\mathbf{u}$  associé à la valeur propre nulle est tel que  $\mathbf{R}\mathbf{u} = \mathbf{0}$ , puisque les facteurs propres sont des vecteurs propres de la matrice de corrélation. On considère le vecteur  $\mathbf{u} = (v_1 \sigma_1, \dots, v_p \sigma_p)'$ . Le vecteur est évidemment non nul et on peut réécrire matriciellement la relation trouvée précédemment comme  $\mathbf{Z}\mathbf{u} = \mathbf{0}$ .

On a donc  $\mathbf{R}\mathbf{u} = \frac{1}{n}\mathbf{Z}'\mathbf{Z}\mathbf{u} = \frac{1}{n}\mathbf{Z}'\mathbf{0} = \mathbf{0}$ , ce qui répond à la question posée.

## 7 Propriété de maximisation des corrélations

On suppose la variable  $\mathbf{v}$  centrée. Alors

$$\text{cor}(\mathbf{v}, \mathbf{x}^j) = \text{cor}(\mathbf{v}, \mathbf{z}^j) = \frac{\text{cov}(\mathbf{v}, \mathbf{z}^j)}{\sqrt{\text{var } \mathbf{v}}} = \frac{\mathbf{v}'\mathbf{D}_p \mathbf{z}^j}{\sqrt{\mathbf{v}'\mathbf{D}_p \mathbf{v}}}.$$

Le vecteur des corrélations avec  $\mathbf{v}$  est donc

$$\frac{1}{\sqrt{\mathbf{v}'\mathbf{D}_p\mathbf{v}}}\mathbf{v}'\mathbf{D}_p\mathbf{Z}$$

et la somme des corrélations au carré est

$$\frac{\mathbf{v}'\mathbf{D}_p\mathbf{Z}\mathbf{Z}'\mathbf{D}_p\mathbf{v}}{\mathbf{v}'\mathbf{D}_p\mathbf{v}}.$$

La forme est exactement la même que l'optimisation faite précédemment, avec  $\mathbf{M}$  remplacée par  $\mathbf{D}_p$  et  $\mathbf{VM}$  remplacée par  $\mathbf{ZZ}'\mathbf{D}_p$ . Le maximum est donc atteint pour  $\mathbf{v}$  vérifiant

$$\mathbf{ZZ}'\mathbf{D}_p\mathbf{v} = \lambda\mathbf{v}.$$

Or, on sait que les vecteurs propres de ces matrices sont les  $\mathbf{c}_k$ . La somme des corrélations au carré est donc maximale pour  $\mathbf{v} = \mathbf{c}_1$ .