Solution du partiel d'AD

M1 MMD - MA – 25/03/2015
P. Bertrand

Exercice 1

1a) $y_i^j = kx_i^j$ donc $\overline{y}_i = k\,\overline{x}_i$, et par conséquent : $\widetilde{y}_i^j = y_i^j - \overline{y}_i = k\,x_i^j - k\,\overline{x}_i = k\,\widetilde{x}_i^j$.

1b) Les résultats u^α et λ_α de l'ACP de X sont solutions de $V_X u^\alpha = \lambda_\alpha\, u^\alpha$. Par ailleurs, v^α et μ_α vérifient : $V_Y v^\alpha = \widetilde{Y} D_p \widetilde{Y}' v^\alpha = k^2\,\widetilde{X}\,D_p\,\widetilde{X}'\,v^\alpha = k^2 V_X v^\alpha = \mu_\alpha\, v^\alpha$. Donc $\lambda_\alpha\, u^\alpha = (\mu_\alpha/k^2)\, v^\alpha$, ce qui prouve que $\mu_\alpha = k^2\lambda_\alpha$ et $u^\alpha = \epsilon_\alpha v^\alpha$ car u^α et v^α sont de normes 1.
De plus $\xi_\alpha = \widetilde{Y}' v^\alpha = k\,\widetilde{X}'\epsilon_\alpha\, u^\alpha = \epsilon_\alpha\, k\,\psi_\alpha$.

2a) Avec les hypothèses de cette question, il résulte que (1) équivaut à $y_i = a_i \mathbb{1}_n + x_i$, pour tout $i \in I$. Cette condition s'écrit encore sous la forme $y_i^j = a_i + x_i^j$, pour tout $i \in I$ et tout $j \in J$.

2b) Pour tout $i \in I$ et $j \in J$, on a $\widetilde{y}_i^j = y_i^j - \overline{y}_i = a_i + x_i^j - \overline{y}_i$.
Or, $\overline{y}_i = \sum_{j\in J} p_j\,(a_i + x_i^j) = a_i + \overline{x}_i$. D'où $\widetilde{y}_i^j = x_i^j - \overline{x}_i = \widetilde{x}_i^j$, d'où $\widetilde{Y} = \widetilde{X}$.

2c) Comme les tableaux centrés $\widetilde{Y}$ et $\widetilde{X}$ sont les mêmes, et que $M_1 = M$, nous avons $v^\alpha = \epsilon_\alpha\, u^\alpha$ et $\mu_\alpha = \lambda_\alpha$, pour tout α.

3a) On a $y_i^j = b_i x_i^j$ pour tout $i \in I$ et tout $j \in J$, donc $y^j = D_b x^j$, pour tout $j \in J$.

3b) Pour tout $j, j' \in J$, on a :

$$\|y^j - y^{j'}\|^2_{M_1} = \|D_b x^j - D_b x^{j'}\|^2_{I_p} = (x^j - x^{j'})' D_b D_b (x^j - x^{j'}) = \|x^j - x^{j'}\|^2_{D_{b^2}}.$$

Les composantes principales de l'ACP de X (resp. Y), notées ψ_α (resp. ξ_α) pour α variant de 1 à r, ne dépendent que des distances $\|x^j - x^{j'}\|^2_{D_{1/s^2}}$ (resp. $\|y^j - y^{j'}\|^2_{M_1} = \|x^j - x^{j'}\|^2_{D_{b^2}}$) entre tous les individus $j, j' \in J$. Il en résulte que si $|\,b_i\,| = \dfrac{1}{s_i}$, pour tout $i \in I$, alors $\xi_\alpha = \pm\,\psi_\alpha$ pour toutes les valeurs de α (l'orientation d'un axe étant toujours arbitraire).

3c) Du fait que $Y = D_b\, X$, on sait que $V_Y = D_b V_X D_b$ car D_b est symétrique. On en déduit que $D_b V_X D_b v^\alpha = \mu_\alpha v^\alpha$. En mulpliant à gauche par $D_{1/b}$, on obtient :

$$V_X D_b v^\alpha = V_X D_{b^2} D_{1/b} v^\alpha = \mu_\alpha D_{1/b} v^\alpha.$$

Comme (C) est vérifiée, $D_{b^2} = D_{1/s^2}$, d'où $V_X\, D_{1/s^2}(D_{1/b} v^\alpha) = \mu_\alpha (D_{1/b} v^\alpha)$. On en conclut que $(D_{1/b}\, v^\alpha)$ dirige le $\alpha^{\text{ème}}$ axe factoriel de l'ACP de X, et $\mu_\alpha = \lambda_\alpha$. Montrons plus précisément que $u_\alpha = \pm D_{1/b} v^\alpha$: il suffit pour cela de montrer que $(D_{1/b} v^\alpha)$ est de norme 1 pour $M = D_{1/s^2}$. En effet, on a :

$$\|D_{1/b} v^\alpha\|^2_M = (v^\alpha)' D_{1/b} D_{1/s^2} D_{1/b} v^\alpha = (v^\alpha)' D_{1/b^2} D_{1/s^2} v^\alpha = (v^\alpha)' v^\alpha = 1.$$

4) Pour tout $i \in I$ et $j \in I$, on a $y_i^j = a_i + b_i x_i^j$. Donc $\overline{y}_i = a_i + b_i \overline{x}_i$ et par conséquent $y_i^j - \overline{y}_i = b_i(x_i^j - \overline{x}_i)$. On en déduit que $\mathrm{var}(y_i) = b_i^2\,\mathrm{var}(x_i)$. Il en résulte :

$$\frac{y_i^j - \overline{y}_i}{\sqrt{\mathrm{var}(y_i)}} = \frac{a_i + b_i x_i^j - a_i - b_i \overline{x}_i}{|\,b_i\,|\sqrt{\mathrm{var}(x_i)}} = \frac{x_i^j - \overline{x}_i}{\sqrt{\mathrm{var}(x_i)}},$$

car $b_i > 0$. On en déduit que les tableaux centrés réduits associés à X et Y sont identiques, donc les composantes principales ξ_α et ψ_α des ACPs normées des tableaux X et Y, respectivement, sont identiques au signe près.

Exercice 2

1) $g = \frac{1}{5}\begin{pmatrix}15 & 40 & 65 & 90 & 115\end{pmatrix}' = \begin{pmatrix}3 & 8 & 13 & 18 & 23\end{pmatrix}' = x^{j_3}$

2) $Y = \begin{pmatrix}\dots & x^j - g & \dots\end{pmatrix} = \begin{pmatrix}-2 & -1 & 0 & 1 & 2\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ -2 & -1 & 0 & 1 & 2\end{pmatrix}.$

3) Toutes les lignes de Y étant égales, le rang de Y est égal à 1, et par conséquent celui de V_Y est aussi 1. Il en résulte qu'il existe un et un seul axe factoriel non trivial. Donc toutes les données sont situées sur cet unique axe factoriel.

4) Il est facile de constater que chaque individu centré (colonnes de Y) est colinéaire au vecteur $\mathbb{1}_5 \in \mathbb{R}^5$ qui dirige donc l'unique axe factoriel non trivial. Donc $u = \frac{1}{\sqrt{5}}\mathbb{1}_5$ le premier (et unique) vecteur axial factoriel.

5) Soit I_T l'inertie totale. On a $V_X = V_Y = \frac{1}{5}Y\,Y' = \frac{10}{5}\mathbb{1}_5\,\mathbb{1}_5' = 2\,\mathbb{1}_5\,\mathbb{1}_5'$. D'où $I_T = \mathrm{tr}(V_Y) = 10$.
On a aussi : $I_T = \sum_{j\in J} p_j\, d^2(x^j, g) = \frac{1}{5}\sum_{j\in J} p_j\, \|x^j - g\|^2 = \frac{2}{5}(5\times(-2)^2 + 5\times(-1)^2 + 5\times 0^2) = 10.$

6) $\Psi_1^{j_5} = (y^{j_5})'u = \frac{1}{\sqrt{5}}(2+2+2+2+2) = \frac{10}{\sqrt{5}} \approx 4.472$

7) $\mathrm{CTR}_1(j_5) = \frac{1}{5}\frac{(\Psi_1^{j_5})^2}{\lambda} = \frac{1}{5}\,\frac{100}{5}\,\frac{1}{10} = \frac{2}{5} = 0.4$

8) Il n'existe qu'un seul "axe factoriel" pour représenter les variables. Les coordonnées des variables sur cet axe sont celles du vecteur $\eta = \sqrt{\lambda}\,u$. Donc, sur cet unique axe, chaque variable a la même coordonnée qui vaut $\frac{\sqrt{10}}{\sqrt{5}} = \sqrt{2} = 1.414$

9) Chaque variable ayant une variance égale à 2, on a $D_{1/s^2} = \frac{1}{2}\mathbb{I}_5$. Donc les vecteurs axiaux factoriels v^α et les valeurs propres μ_α sont solutions de $\frac{1}{2}V_Y v^\alpha = \mu_\alpha v^\alpha$ au lieu de l'équation précedente $V_Y u^\alpha = \lambda_\alpha u^\alpha$ (en ACP sur matrice variance) pour u_α et λ_α.

Question 3. La matrice à diagonaliser ayant été divisée par 2, il existe toujours un seul axe factoriel non trivial car la matrice à diagonaliser est toujours de rang 1.

Question 4. L'unique vecteur axial factoriel v est colinéaire à u mais normé pour $M = \frac{1}{2}\mathbb{I}_5$, donc $v = \frac{\mathbb{1}}{\|\mathbb{1}\|_M}$. Par ailleurs, on a $\|\mathbb{1}\|_M^2 = \frac{1}{2}\mathbb{1}'\mathbb{1} = \frac{5}{2}$. D'où $v = \sqrt{\frac{2}{5}}\,\mathbb{1} \approx 0.632\,\mathbb{1}$.

Question 5. Ici, l'inertie totale est égale à la valeur propre μ, d'où $I_T = \mu = \lambda/2 = 10/2 = 5$.

Question 6. Si on note ξ_1 l'unique composante principale, on a :

$$\xi_1^5 = (y^5)'\frac{1}{2}\mathbb{I}_5\, v = \frac{1}{2}\times 2\times \mathbb{1}'\Big(\sqrt{\frac{2}{5}}\,\mathbb{1}\Big) = 5\times\sqrt{\frac{2}{5}} \approx 3.162$$