ANALYSE DES DONNEES (Partiel)

Master M1 MMD - MA, 25 mars 2015

Calculatrice autorisée, documents autorisés : 2 feuilles recto-verso.

Barême approximatif : 10 points pour chacun des deux exercices.

Exercice 1

Soit $X = \left(x_i^j\right)_{\substack{1\leq i\leq p\\ 1\leq j\leq n}}$ un tableau de données, comportant p lignes et n colonnes, dont le terme général x_i^j indique la valeur de la $i^{\text{ème}}$ variable pour le $j^{\text{ème}}$ individu. Le $j^{\text{ème}}$ individu est muni d'un poids p_j (avec $p_j > 0$ pour tout j et $\sum_{j=1}^{n} p_j = 1$). On note $I = \{1, \ldots, p\}$, $J = \{1, \ldots, n\}$ et r le rang de la matrice variance V_X de X. Les résultats de l'ACP de X, l'espace $\mathbb{R}^p$ étant muni d'une métrique M, sont notés comme suit :

- le $\alpha^{\text{ème}}$ vecteur axial factoriel (pour le nuage des individus) est noté u^α,
- la $\alpha^{\text{ème}}$ valeur propre est notée λ_α,
- la $\alpha^{\text{ème}}$ composante principale du nuage des individus est notée ψ_α,

avec $\alpha \in \{1, \ldots, r\}$. On considère aussi un tableau de données Y déduit de X par la formule :

$$Y = A + BX,$$

où A est matrice à p lignes et n colonnes, et B une matrice carrée d'ordre p. Par la suite, les résultats de l'ACP de $Y = A + BX$, l'espace $\mathbb{R}^p$ étant muni d'une métrique notée M_1, sont notés comme suit : v^α désigne le $\alpha^{\text{ème}}$ vecteur axial factoriel (pour le nuage des individus), μ_α la $\alpha^{\text{ème}}$ valeur propre et ξ_α la $\alpha^{\text{ème}}$ composante principale du nuage des individus, avec $\alpha \in \{1, \ldots, r_1\}$ où r_1 est le rang de la matrice variance V_Y de Y. De plus, on adopte les notations suivantes : $\mathbb{1}$ est le vecteur de $\mathbb{R}^n$ dont les coordonnées sont toutes égales à 1, $0_{p\times n}$ est la matrice à p lignes et n colonnes dont tous les termes sont égaux à 0, $\mathbf{I}_p$ est la matrice identité d'ordre p et D_b est matrice diagonale dont le $i^{\text{ème}}$ terme sur la diagonale est la $i^{\text{ème}}$ coordonnée b_i d'un vecteur $b \in \mathbb{R}^p$. Enfin, on note $\widetilde{X}$ (resp. $\widetilde{Y}$) le tableau centré déduit de X (resp. Y). On suppose par la suite que, pour tout $\alpha \in \{1, \ldots, r\}$, les valeurs propres (non nulles) λ_α sont simples.

Dans les questions 1. *et* 2.*, on suppose que* $M = M_1 = \mathbf{I}_p$.

1. On suppose que $A = 0_{p\times n}$ et $B = k\,\mathbf{I}_p$ où k est un réel non nul.

1.a. Montrer que pour tout $i \in I$ et tout $j \in J$, on a $\widetilde{y}_i^j = k\,\widetilde{x}_i^j$.

1.b. Montrer que pour tout $\alpha \in \{1, \ldots, r\}$, on a :

$$v^\alpha = \epsilon_\alpha\, u^\alpha,\ \mu_\alpha = k^2\lambda_\alpha \ \text{ et }\ \xi_\alpha = \epsilon_\alpha\, k\,\psi_\alpha \qquad (\text{avec } \epsilon_\alpha = \pm 1).$$

2. On suppose que $A = (a_1\mathbb{1}, \ldots, a_p\mathbb{1})'$ et $B = \mathbf{I}_p$, où les a_i ($i \in I$) sont des réels.

2.a. Déterminer l'expression de y_i^j en fonction de x_i^j, pour tout $i \in I$ et tout $j \in J$.

2.b. Comparer les tableaux centrés $\widetilde{X}$ et $\widetilde{Y}$.

2.c. Pour chaque $\alpha \in \{1, \ldots, r\}$, exprimer v^α en fonction de u^α, et μ_α en fonction de λ_α.

3. On suppose que $A = 0_{p\times n}$ et $B = D_b$ où $b \in \mathbb{R}^p$ est tel que $b_i \neq 0$ pour tout i.
Dans cette question, $M = D_{1/s^2}$, où D_{1/s^2} désigne la matrice diagonale des inverses des variances des p variables du tableau X, et $M_1 = \mathbf{I}_p$.

3.a. Donner l'expression de y^j en fonction de x^j, pour tout $j \in J$.
3.b. Montrer que, pour tout $j, j' \in J$, la quantité $\|y^j - y^{j'}\|^2_{M_1}$ peut s'exprimer sous la forme $\|x^j - x^{j'}\|^2_N$ où N est une métrique dont la valeur ne dépend que de b. En déduire une condition portant sur la valeur absolue de b_i ($i \in I$) qui, si elle est vérifiée, entraîne que $\xi_\alpha = \pm\psi_\alpha$, pour chaque $\alpha \in \{1, \ldots, r_1\}$. Par la suite, on notera (C) cette condition.
3.c. On suppose que (C) est vérifiée. Pour α quelconque, exprimer v^α en fonction de u^α.

4. On suppose que $A = (a_1\mathbb{1}, \ldots, a_p\mathbb{1})'$ et $B = D_b$ où $b \in \mathbb{R}^p$ est tel que $b_i > 0$ pour tout i.
Par ailleurs, la métrique M (resp. M_1) est égale à la matrice diagonale des inverses des variances des p variables du tableau X (resp. Y). Exprimer ξ_α en fonction de ψ_α, pour chaque valeur possible de α.

Exercice 2

On considère le tableau de données, noté X, qui est défini par :

$X =$

	j_1	j_2	j_3	j_4	j_5
i_1	1	2	3	4	5
i_2	6	7	8	9	10
i_3	11	12	13	14	15
i_4	16	17	18	19	20
i_5	21	22	23	24	25

où la $i^{\text{ème}}$ ligne désigne la variable x_i et la $j^{\text{ème}}$ colonne désigne l'individu x^j.
Par la suite, on considère les résultats de l'ACP sur matrice variance du tableau X (i.e., ACP de X lorsque $\mathbb{R}^5$ est muni de la métrique identité et chaque individu possède un poids égal à 1/5).

1. Soit g le centre de gravité du nuage des cinq individus associés au tableau X. Montrer que $g = x^{j_3}$, c'est-à-dire coïncide avec l'individu j_3.

2. Calculer le tableau centré, noté Y, qui est associé à X.

3. Combien existe-t'il d'axes factoriels non triviaux ? (justifier votre réponse)

4. Pour chacun des axes factoriels non triviaux, calculer l'unique vecteur axial factoriel qui le dirige et dont la première coordonnée est positive.

5. Déterminer l'inertie totale du nuage étudié.

6. Calculer la première composante principale de l'individu j_5, notée $\Psi_1^{j_5}$.

7. Calculer la contribution de l'individu j_5 à l'inertie du premier axe, notée $CTR_1(j_5)$.

8. Déterminer la représentation des variables dans le nouveau système d'axes factoriels de l'espace des variables.

9. Dans cette question, on considère les résultats de l'ACP normée du tableau X (i.e., ACP de X lorsque $\mathbb{R}^5$ est muni de la métrique $M = D_{1/s^2}$, où D_{1/s^2} désigne la matrice diagonale des inverses des variances des variables du tableau X, et lorsque chaque individu possède un poids égal à 1/5). Répondre à nouveau aux questions 3), 4), 5) et 6).