ANALYSE DES DONNEES (Partiel)

Master M1 MMD - MA, 13 mars 2017

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

Barême approximatif : 10 points pour chacun des deux exercices.

Exercice 1

On considère deux nuages de points $\mathcal{N}_1$ et $\mathcal{N}_2$ dans $\mathbb{R}^p$, $p \geq 1$, muni de la métrique identité I_p.

Le nuage $\mathcal{N}_1$ est constitué de n individus $(x_i)_{1\leq i\leq n}$, chaque individu a un poids p_i avec $p_i > 0$ pour tout i et $\sum_{i=1}^{n} p_i = 1$.

De même le nuage $\mathcal{N}_2$ est constitué de $N - n$ individus $(x_i)_{n+1\leq i\leq N}$, chaque individu a un poids q_i avec $q_i > 0$ pour tout i et $\sum_{i=n+1}^{N} q_i = 1$.

Les résultats de l'ACP de $\mathcal{N}_1$, l'espace $\mathbb{R}^p$ étant muni de la métrique identité I_p, sont notés comme suit :

- le α ème vecteur axial factoriel (pour le nuage des individus) est noté u_α,
- la α ème valeur propre est notée λ_α,
- la α ème composante principale du nuage des individus est notée ψ_α,

avec $\alpha \in \{1, \ldots, p\}$.

Les résultats de l'ACP de $\mathcal{N}_2$, l'espace $\mathbb{R}^p$ étant muni de la métrique identité I_p, sont notés comme suit :

- le α ème vecteur axial factoriel (pour le nuage des individus) est noté v_α,
- la α ème valeur propre est notée μ_α,
- la α ème composante principale du nuage des individus est notée ϕ_α,

avec $\alpha \in \{1, \ldots, p\}$.

On note $\mathcal{N}$ l'union des deux nuages, constitué de N individus $(x_i)_{1\leq i\leq N}$, chaque individu a un poids r_i défini par

$$\forall 1 \leq i \leq N, \quad r_i = \begin{cases} p_i/2 & \text{si } 1 \leq i \leq n, \\ q_i/2 & \text{si } n+1 \leq i \leq N. \end{cases}$$

On note G_1 le centre de gravité de $\mathcal{N}_1$, G_2 de $\mathcal{N}_2$ et G de $\mathcal{N}$.

1. Vérifier que $r_i > 0$ et que $\sum_{i=1}^{N} r_i = 1$.
2. Comment se situe G par rapport à G_1 et G_2 ?
3. Exprimer l'inertie totale de $\mathcal{N}$ en fonction de l'inertie totale de $\mathcal{N}_1$ notée $I_T(\mathcal{N}_1)$, de $\mathcal{N}_2$ notée $I_T(\mathcal{N}_2)$ et de $d(G_1, G_2)^2 = \|\overrightarrow{G_1G_2}\|^2$.
4. On suppose que $G_1 \neq G_2$ on pose

$$u = \frac{\overrightarrow{G_1G_2}}{\|\overrightarrow{G_1G_2}\|}.$$

On note V_1, V_2 et V les matrices de variance des nuages $\mathcal{N}_1$, $\mathcal{N}_2$ et $\mathcal{N}$.

(a) Montrer que

$$V = \frac{1}{2}(V_1 + V_2) + \frac{\|\overrightarrow{G_1G_2}\|^2}{4} uu'.$$

On fait l'hypothèse, dans les deux dernières questions, que pour tout $1 \leq \alpha \leq p, \quad u_\alpha = v_\alpha$**.**

(b) On suppose que $u = u_1 = v_1$

i. Montrer que u est un axe factoriel de $\mathcal{N}$. Est ce le premier axe factoriel ?

ii. Exprimer la composante principale H associée à l'axe factoriel u en fonction de ψ_1 et de ϕ_1.

(c) On suppose que $u \in \ker V_1 \cap \ker V_2$.

i. Montrer que u est un axe factoriel de $\mathcal{N}$ et déterminer une condition sur $d(G_1, G_2)$ pour que u soit le premier axe factoriel de $\mathcal{N}$.

ii. Montrer que $P_u(\mathcal{N})$, où P_u est la projection orthogonale sur la droite passant par G de vecteur directeur u, se réduit à deux points. Calculer l'inertie totale du nuage $P_u(\mathcal{N})$,

Exercice 2

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

$$X = \begin{pmatrix} 2 & 1 & -1 & 3 & -3 \\ 0 & 3 & -1 & 2 & -2 \\ 1 & 2 & -1 & 1 & -1 \end{pmatrix}$$

où la $i^{\text{ème}}$ ligne désigne l'individu x_i et la $j^{\text{ème}}$ colonne désigne la variable x^j.
Chaque individu possède un poids égal à 1/3. On considère les résultats de l'ACP du tableau X lorsque $\mathbb{R}^5$ est muni de la métrique identité.

1. Déterminer le tableau centré Y.
2. Sans calculer la matrice variance V du tableau X, combien existe-t-il d'axes factoriels non triviaux ?
3. Calculer V. En déduire l'inertie totale du nuage étudié.
4. On pose $u = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \\ -1 \end{pmatrix}$.
   (a) Vérifier que u est un vecteur propre de V associé à la valeur propre 2.
   (b) En déduire les valeurs propres de V.
   (c) Déterminer, pour chacun des axes factoriels non triviaux, l'unique vecteur axial factoriel qui le dirige et dont la première coordonnée est positive.
5. Calculer les deux premières composantes principales , notée Ψ_1 et Ψ_2.
6. Représenter les trois individus dans le plan factoriel constitué des deux premiers axes.
7. Calculer la contribution relative de chaque individu à l'inertie du premier axe.
8. Représenter les 5 variables dans le plan factoriel constitué des deux premiers axes.