Fiche 3 - AD - M1 MAS

Tran Viet Chi, chi.tran@math.univ-lille1.fr, bureau 316 (bâtiment M3).

Exercice 1 (Analyse d'un tableau de distances)
On considère une population de n individus, caractérisés par p variables. Les données sont rangées dans un tableau $X = (x_{ij}, i \in \{1,\dots n\}, j \in \{1,\dots p\}) \in \mathcal{M}_{n,p}$, où x_{ij} est la valeur de la variable j pour l'individu i. On note p_i, $i \in \{1,\dots n\}$, les poids des individus, avec $\sum_{i=1}^n p_i = 1$. On suppose que $\mathbb{R}^p$ est muni de la métrique M, c'est-à-dire que pour $x \in \mathbb{R}^p$, $\|x\|^2 = {}^t\!xMx$. On suppose aussi que chacune des variables est centrée : pour tout $j \in \{1,\dots p\}$, $\sum_{i=1}^n p_i x_{ij} = 0$.

1. Quel est le centre de gravité du nuage $N = \{(x_i, p_i),\ i \in \{1,\dots n\}\}$? Comment s'écrit dans ce cadre la matrice de variances-covariances V ?

On considère la matrice $D = (d_{ii'}, i, i' \in \{1,\dots n\}) \in \mathcal{M}_{n,n}$ des distances carrées entre individus, où

$$d_{ii'}^2 = {}^t(x_i - x_{i'})M(x_i - x_{i'}).$$

On pose également :

$$d_i^2 = \sum_{i'=1}^n p_{i'} d_{ii'}^2, \qquad d^2 = \sum_{i=1}^n p_i d_i^2.$$

Le but de l'exercice est de montrer que l'analyse du nuage peut être mené à partir de la matrice des distances carrées D.
2. Montrer que pour tout $i \in \{1,\dots n\}$, $d_i^2 = \|x_i\|^2 + I(N)$ où $I(N)$ est l'inertie du nuage N.

3. En déduire que $d^2 = 2\ I(N)$.

4. En développant $\langle x_i - x_{i'}, x_i - x_{i'}\rangle = \|x_i - x_{i'}\|^2$, exprimer le produit scalaire $\langle x_i, x_{i'}\rangle$ en fonction de d_i^2, $d_{i'}^2$, d^2 et $d_{ii'}^2$.

5. Supposons que l'analyse factorielle du nuage génère r axes principaux dirigés par $(u_k)_{k\in\{1,\dots r\}}$, qu'on pourra supposer normés. On note $(C_k)_{k\in\{1\dots r\}}$ les facteurs principaux associés. Ecrire C_k en fonction de X, M et u_k.

6. Montrer que $XMVMu_k = \lambda_k C_k$, où λ_k est la valeur propre de VM associée à u_k.

7. En déduire que C_k est également vecteur propre d'une matrice dont le terme général s'écrit en fonction des distances carrées $d_{ii'}^2$, d_i^2, $d_{i'}^2$ et d^2.

8. En définissant le vecteur $F_k \in \mathbb{R}^n$ dont la composante $i \in \{1,\dots n\}$ est $F_{ik} = \sqrt{p_i}C_{ik}$, déterminer la matrice S dont F_k est le vecteur propre associé à la valeur propre λ_k. Montrer que le vecteur $(\sqrt{p_i})_{i\in\{1,\dots n\}}$ est vecteur propre associé à la valeur propre 0.

9. Montrer que $\sum_{i=1}^n F_{ik}^2 = \lambda_k$ et que pour $k \neq \ell$, $\sum_{i=1}^n F_{ik}F_{i\ell} = 0$.

10. On considère un nuage de 3 individus, $\{1, 2, 3\}$, tel que :

$$d_{12}^2 = d_{23}^2 = 1, \quad d_{13}^2 = 2, \quad p_1 = p_2 = p_3 = \frac{1}{3}.$$

Déterminer S, les valeurs propres λ_k et les vecteurs propres H_k. Représenter les trois individus dans le plan principal.

Exercice 2 (ACP (non normée))
Reprendre les données de l'exercice 2 de la fiche 2.
1. Calculer les axes principaux (donner les vecteurs u_k les engendrant) et les inerties associées λ_k. Donner les parts d'inertie expliquées par chaque axe.

2. Calculer les coordonnées des individus sur l'axe factoriel 1.

3. Calculer les contributions des individus à l'inertie de l'axe 1, $CONT_1(i)$, ainsi que les indicateurs de qualité de représentation $C02_1(i)$.

4. Quelles sont les coordonnées des variables dans le nuage des variables ?