ANALYSE DES DONNEES (Examen et correction)

Master M1 MMD - MA, 23 mai 2017

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

Barême approximatif : 10 points pour chacun des deux exercices.

Exercice 1

On considère le tableau K suivant où a est un entier non nul :

I/J	j_1	j_2	j_3	j_4	j_5
i_1	a	a	a	0	0
i_2	a	a	0	a	0
i_3	0	a	0	a	a

On pose

$$I = \{i_1, i_2, i_3\} \text{ et } J = \{j_1, j_2, j_3, j_4, j_5\}.$$

On effectue l'analyse factorielle des corresponsdances (AFC) de K.

1. Déterminer les centres de gravité des nuages $\mathcal{N}(I)$ et $\mathcal{N}(J)$.

   On obtient
   $$f_I = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} \text{ et } f_J = \frac{1}{9}\begin{pmatrix} 2 \\ 3 \\ 1 \\ 2 \\ 1 \end{pmatrix}.$$

2. Déterminer la matrice des profils colonnes F_1 ainsi que la matrice des profils lignes F_2 de K.

   On a donc
   $$F_1 = \frac{1}{6}\begin{pmatrix} 3 & 2 & 6 & 0 & 0 \\ 3 & 2 & 0 & 3 & 0 \\ 0 & 2 & 0 & 3 & 6 \end{pmatrix} \text{ et } F_2 = \frac{1}{3}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

3. Vérifier que le produit
   $$F_1F_2 = \frac{1}{18}\begin{pmatrix} 11 & 5 & 2 \\ 5 & 8 & 5 \\ 2 & 5 & 11 \end{pmatrix}.$$

   Le produit donne le bon résultat.

4. Quel est l'influence du réel a sur l'AFC de ce tableau.

   Aucune puisque les individus des nuages ne dépendent pas de a ainsi que les poids et les métriques.

5. Quel est l'axe factoriel trivial, à quelle valeur propre est-il associé ?

   f_I est le vecteur propre de F_1F_2 associé à la valeur propre triviale 1.

6. Quelle est l'inertie du nuage $\mathcal{N}(J)$.

   L'inertie du nuage $\mathcal{N}(J)$ est la trace de F_1F_2 moins 1 donc $30/18 - 1 = 2/3$.

7. On pose
$$w_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \text{ et } w_2 = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$
Montrer que w_1 et w_2 sont des vecteurs propres de F_1F_2, en déduire les axes factoriels non triviaux u_1 et u_2 ainsi que les valeurs propres associés. On choisira u_1 de manière que la première coordonnée soit positive, de même pour u_2. On vérifie que
$$F_1F_2w_1 = \frac{1}{2}w_1 \text{ et } F_1F_2w_2 = \frac{1}{6}w_2.$$
Or la norme de w_1 pour la métrique $D_{1/f_I} = 3I_3$ est $\sqrt{\P}$ donc le premier axe factoriel u_1 associé à la valeur propre $\lambda_1 = 1/2$ est $u_1 = \frac{1}{\sqrt{6}}w_1$ et le deuxième axe factoriel u_2 associé à la valeur propre $\lambda_2 = 1/6$ est $u_2 = \frac{1}{3\sqrt{2}}w_2$

8. On note $\varphi_\alpha(i)$ l'abscisse de la projection du profil de la ligne i sur le α ème axe factoriel. Remplir le tableau suivant avec la contrainte $\varphi_\alpha(i_1) \geq 0$

I/J	φ_1	φ_2	φ_3
i_1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{6}$	0
i_2	0	$\frac{-2\sqrt{3}}{6}$	0
i_3	$\frac{-\sqrt{3}}{2}$	$\frac{\sqrt{3}}{6}$	0

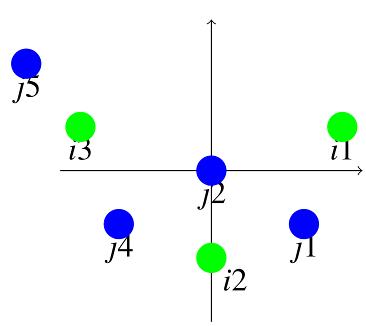
φ_α est une composante principale du nuage $\mathcal{N}(I)$ donc φ_α est un vecteur propre de $F_2'F_1'$. Or on remarque que la matrice F_1F_2 est symétrique donc $F_2'F_1' = F_1F_2$, on a les même vecteurs propres w_1 et w_2 que l'on normalise avec la métrique $1/3I_3$. Donc
$$\varphi_1 = \frac{\sqrt{3}}{2}w_1 \text{ et } \varphi_2 = \frac{\sqrt{3}}{6}w_2.$$

9. On note ψ_α^j l'abscisse de la projection du profil de la colonne j sur le α ème axe factoriel. En utilisant les formules de transition, compléter le tableau suivant

I/J	j_1	j_2	j_3	j_4	j_5
ψ_1	$\frac{\sqrt{6}}{4}$	0	$\frac{2\sqrt{6}}{4}$	$\frac{-\sqrt{6}}{4}$	$\frac{-2\sqrt{6}}{4}$
ψ_2	$\frac{-\sqrt{2}}{4}$	0	$\frac{2\sqrt{2}}{4}$	$\frac{-\sqrt{2}}{4}$	$\frac{2\sqrt{2}}{4}$

Avec les formules de transition on a
$$\psi_1 = \sqrt{2}F_1'\varphi_1 \text{ et } \psi_2 = \sqrt{6}F_1'\varphi_2.$$

10. Représenter les deux nuages $\mathcal{N}(I)$ et $\mathcal{N}(J)$ simultanément dans le plan factoriel 1-2. Les tableaux précédents donnent les coordonnées des points des deux nuages.



11. Calculer la contribution de i_1 à chacun des axes factoriels non triviaux ainsi que la qualité de représentation de i_1 dans le plan factoriel 1-2 c'est-à-dire $COR_1(i_1) + COR_2(i_1)$.

    On a
$$CTR_1(i_1) = \frac{1}{3} \times \frac{3/4}{1/2} = 1/2 \text{ et } CTR_2(i_1) = \frac{1}{3} \times \frac{3/36}{1/6} = 1/6.$$
    Comme il n'y a que deux axes non triviaux , la qualité de représentation de i_1 dans le plan factoriel 1-2 est 1.

Exercice 2

On considère un tableau de contingence, noté $K = (k_{ij})_{1 \le i \le p, 1 \le j \le q}$, qui croise deux variables qualitatives X et Y dont le nombre de modalités est p pour X et q pour Y. On pose
$$I = \{1, \cdots, p\} \text{ et } J = \{1, \cdots, q\}.$$

On effectue une AFC sur ce tableau. On note F_α la composante principale associée à l'axe factoriel α pour le nuage $\mathcal{N}(J)$ et G_α la composante principale associée à l'axe factoriel α pour le nuage $\mathcal{N}(I)$.

1. Rappeler les formules de transitions entre F_α et G_α.

   On rappelle que
$$\forall i \in I, \quad F_\alpha(i) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^{q} f_j^i G_\alpha(j) \text{ et } \forall j \in J, \quad G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^{p} f_i^j F_\alpha(i).$$

2. On suppose qu'il existe une valeur propre non triviale égale à 1 : $\lambda_\alpha = 1$.

   (a) On suppose qu'il existe des indices i_0, j_0 tels que
$$\forall i \in I \setminus \{i_0\}, \quad F_\alpha(i) < F_\alpha(i_0) \text{ et } \forall j \in J \setminus \{j_0\}, \quad G_\alpha(j) < G_\alpha(j_0).$$

   i. Montrer que
$$F_\alpha(i_0) = G_\alpha(j_0).$$
   On a
$$F_\alpha(i) = \sum_{j=1}^{q} f_j^i G_\alpha(j) \le \sum_{j=1}^{q} f_j^i G_\alpha(j_0) = G_\alpha(j_0)$$
   et de même
$$G_\alpha(j_0) \le F_\alpha(i_0).$$
   donc on a l'égalité.

   ii. En déduire que
$$f_j^{i_0} = \begin{cases} 1 & \text{si } j = j_0, \\ 0 & \text{sinon} \end{cases}$$

On suppose que $f_j^{i_0} > 0$ pour un indice $j \neq j_0$, alors

$$\begin{aligned} G_\alpha(j_0) &= \sum_{j=1}^{q} f_j^{i_0} G_\alpha(j), \\ &= f_{j_0}^{i_0} G_\alpha(j_0) + \sum_{j \neq j_0} f_j^{i_0} G_\alpha(j), \\ &< f_{j_0}^{i_0} G_\alpha(j_0) + \left(\sum_{j \neq j_0} f_j^{i_0} \right) G_\alpha(j_0), \\ &< f_{j_0}^{i_0} G_\alpha(j_0) + (1 - f_{j_0}^{i_0}) G_\alpha(j_0), \\ &< G_\alpha(j_0). \end{aligned}$$

L'inégalité est impossible donc pour tout $j \neq j_0$, $f_j^{i_0} = 0$. Comme la somme des composantes de $f_J^{i_0}$ vaut 1, on a le résultat.

iii. Que cela signifie t-il pour les modalités i_0 et j_0 ?

De même on a

$$f_i^{j_0} = \begin{cases} 1 & \text{si } i = i_0, \\ 0 & \text{sinon} \end{cases}$$

Donc tout individu qui possède la modalité i_0 possède la modalité j_0 et réciproquement.

(b) On note I_0 l'ensemble des indices pour lesquels $F_\alpha(i)$ atteint son maximum et J_0 l'ensemble des indices pour lesquels $G_\alpha(j)$ atteint son maximum. Pour simplifier on note $I_0 = \{1, \cdots, p_0\}$ et $J_0 = \{1, \cdots, q_0\}$, montrer que le tableau d'effectif K s'écrit

I/J	J_0	$J \setminus J_0$
I_0	K_0	0
$I \setminus I_0$	0	K_1

où K_0 et K_1 sont des tableaux d'effectifs.

On raisonne comme dans le cas précédent, on a pour $j_0 \in J_0$ et $i_0 \in I_0$

$$\begin{aligned} G_\alpha(j_0) &= \sum_{i \in I_0} f_i^j F_\alpha(i_0) + \sum_{i \in I \setminus I_0} f_i^j F_\alpha(i), \\ &\leq F_\alpha(i_0). \end{aligned}$$

D'où par symétrie

$$\forall i_0 \in I_0, \ \ \forall j \in J_0, \ \ G_\alpha(j_0) = F_\alpha(i_0).$$

et aussi

$$\forall i_0 \in I_0, \ \ \forall j \in J \setminus J_0, \ \ f_j^{i_0} = 0 \text{ et } \forall j_0 \in J_0, \ \ \forall i \in I \setminus I_0, \ \ f_i^{j_0} = 0.$$

En termes d'effectifs, cela donne

$$\forall i_0 \in I_0, \ \ \forall j \in J \setminus J_0, \ \ k_{i_0,j} = 0 \text{ et } \forall j_0 \in J_0, \ \ \forall i \in I \setminus I_0, \ \ f_{ij_0} = 0.$$

on en déduit le résultat.

3. On suppose que $p = q$ et que 1 est la seule valeur propre de F_1F_2.

(a) Montrer que $F_1F_2 = I_p$ où I_p est la matrice identité.

La matrice F_1F_2 est diagonalisable et n'admet que 1 comme valeur propre donc c'est l'identité.

(b) Que peut-on en déduire pour les matrices F_1 et F_2 ?

On en déduit que

$$\forall (i,i') \in I^2, \quad i \neq i' \Longrightarrow \sum_{j=1}^{p} f_i^j f_j^{i'} = 0,$$

donc les termes étant positifs,

$$\forall (i,i') \in I^2, \quad i \neq i' \Longrightarrow \forall j \in I, \quad f_{ij} f_{i'j} = 0,$$

Pour i_0 fixé, la ligne i_0 n'étant pas nulle par hypothèse, il existe j_0 tel que $f_{i_0 j_0} \neq 0$ alors d'après ce qui précède

$$\forall i \neq i_0, \quad f_{i j_0} = 0.$$

Par conséquent la ligne i_0 ne comporte que des zéros sauf pour j_0 donc $f_J^{i_0} = e_{j_0}$ où e_j est le $j$ème vecteur de la base canonique de $\mathbb{R}^p$. La matrice F_2 est donc une matrice de permutation, de même pour F_1, inverse de F_2.

(c) Que conclure sur les variables X et Y ?

Les variables X et Y partitionnent l'ensemble des individus de la même manière.