Université de Poitiers
Département de Mathématiques

Statistique descriptive, 1er semestre, année univ. 2009-2010 **Fiche 9**

Analyse factorielle des correspondances

Exercice 1

Table de contingence, lois marginales, lois conditionnelles

Soit $N = (n_{ij})_{i\in I, j\in J}$ une table de contingence $p \times q$ croisant deux variables qualitatives X et Y respectivement à p et q modalités. Le terme n_{ij} désigne le nombre d'individus ayant les modalités $X = i$ et $Y = j$. On note $n = \sum_{ij} n_{ij}$ le nombre total d'individus, ainsi que $n_{i\cdot} := \sum_{j\in J} n_{ij}$ le nombre d'individu ayant la modalité $X = i$ et $n_{\cdot j} = \sum_{i\in I} n_{ij}$ le nombre d'individu ayant la modalité $Y = j$. On note enfin $F := (f_{ij})_{i\in I, j\in J}$ le tableau des fréquences donné par $F = \frac{1}{n}N$, $f_{i\cdot} := \frac{1}{n}n_{i\cdot}$ et $f_{\cdot j} := \frac{1}{n}n_{\cdot j}$. On note $D_I := \text{diag}(f_{i\cdot})$ et $D_J := \text{diag}(f_{\cdot j})$ les matrices diagonales $p \times p$ et $q \times q$ respectivement.

1) Montrer que F défini une loi de probabilité sur $I \times J$ dont les marginales sur I et J sont données respectivement par $(f_{i\cdot})_{i\in I}$ et $(f_{\cdot j})_{j\in J}$.

2) Montrer que le i-ème profil-ligne $\text{PrL}(i) \in \mathbb{R}^q$ donné comme la i-ème ligne de la matrice $Z_I = D_I^{-1}F$ s'interprète comme la loi conditionnelle de Y sachant $X = i$.

3) De même, montrer que le j-ème profil-colonne $\text{PrC}(j) \in \mathbb{R}^p$ donné comme la j-ème colonne de la matrice $Z_J = FD_J^{-1}$ s'interprète comme la loi conditionnelle de X sachant $Y = i$.

4) Que deviennent ces quantités si la loi de probabilité F sur $I \times J$ est une loi produit *i.e.* lorsque les échantillons X et Y sont indépendants ?

Exercice 2

Espaces, poids, métriques et inertie en AFC

À la matrice des fréquences F on associe le nuage des p profils-lignes dans $\mathbb{R}^q$ donné comme les lignes de la matrice $Z_I = D_I^{-1}F$. La i-ème ligne est affectée du poids $f_{i\cdot}$ et D_I est la matirce des poids. On utilise le produit scalaire sur $\mathbb{R}^q$ associé à la matrice D_J^{-1} et la distance d_{χ^2} associée *i.e.*

$$\langle u; v\rangle_{\chi^2} = u'D_J^{-1}v. \tag{1}$$

1) Quel est le centre de gravité G_I du nuage des profils-lignes ?

2) Que vaut l'inertie totale du nuage des profils-lignes calculée par rapport à G_I ?

3) Montrer que OG_I est orthogonal à l'hyperplan $\sum_j u_j = 1$ contenant $\mathcal{N}_I$.

4) Soit u un vecteur de $\mathbb{R}^q$ normé. Montrer que le vecteur ϕ_u des projections orthogonales des points de $\mathcal{N}_I$ sur la droite passant par O et dirigée par u est donné par $\phi_u = Z_I D_J^{-1}u \in \mathbb{R}^p$.

5) Montrer que l'inertie du nuage projeté sur u calculé en O est donnée par

$$I^{\text{vect}(u)}(\mathcal{N}) = u'D_J^{-1}Z_I'D_I Z_I D_J^{-1}u.$$

6) Soit M_I la matrice définie par $M_I := Z_I'D_I Z_I D_J^{-1}$. Montrer que M_I s'écrit également

$$M_I = F'D_I^{-1}FD_J^{-1} = Z_I'Z_J$$

puis que M_I est symétrique et positive pour le produit scalaire donné par (1). Ainsi M_I se diagonalise en base orthonormée (pour le produit scalaire défini en (1) et a des valeurs propres réelles positives.

7) Montrer que la droite d'inertie maximale est dirigée par un vecteur propre de M_I associée à la plus grande valeur propre.

8) Montrer que 1 est la plus grande valeur propre, associée au vecteur propre $\mathbf{1}_J$ (cela ne dépend pas du tableau de données et donc n'importe pas dans l'AFC).

Exercice 3

Propriété d'équivalence distributionnelle pour la distance du χ^2

1) Vérifier que si deux profils-lignes $\mathrm{PrL}(i_1) = (\frac{f_{i_1 j}}{f_{i_1 \cdot}})_{j\in J}$ et $\mathrm{PrL}(i_2) = (\frac{f_{i_2 j}}{f_{i_2 \cdot}})_{j\in J}$ de poids respectifs $f_{i_1\cdot}$ et $f_{i_2\cdot}$ sont égaux *i.e.*

$$\mathrm{PrL}(i_1) = \mathrm{PrL}(i_2)$$

dans une analyse de type AFC, ils peuvent être remplacés par une nouvelle modalité ■ $i_1 + i_2$ ■ et un unique profil-ligne

$$\mathrm{PrL}(i_1 + i_2) = \mathrm{PrL}(i_1) = \mathrm{PrL}(i_2).$$

de poids $f_{i_1\cdot} + f_{i_2\cdot}$.

2) Vérifier que les distances entre profils-colonnes $d_{\chi^2}(\mathrm{PrC}(j_1), \mathrm{PrC}(j_2))$ ne sont pas modifiées.

Exercice 4

Dualité en AFC

On considère le nuage des q profils-colonnes dans $\mathbb{R}^p$ donné par les colonnes de la matrice $Z_J = FD_J^{-1}$. La j-ème colonne est affectée du poids $f_{\cdot j}$ et D_J est la matice des poids. On utilise le produit scalaire sur $\mathbb{R}^p$ associé à la matrice D_I^{-1} et la distance d_{χ^2} associée.

1) Comment s'adaptent les résultats obtenus sur les profils-lignes pour les profils-colonnes ?

2) Soit M_J la matrice définie par $M_J = Z_J Z_I'$: montrer que M_I et M_J ont les mêmes valeurs propres non nulles, avec même multiplicité. Quelles relations existent entre les vecteurs propres ?

3) En déduire les relations quasi-barycentriques.