

REVUE DE STATISTIQUE APPLIQUÉE

R. ROUSSEAU

Analyse des correspondances d'un tableau ayant des cases à poids prépondérants

Revue de statistique appliquée, tome 48, n° 2 (2000), p. 73-83

http://www.numdam.org/item?id=RSA_2000__48_2_73_0

© Société française de statistique, 2000, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

ANALYSE DES CORRESPONDANCES D'UN TABLEAU AYANT DES CASES À POIDS PRÉPONDÉRANTS

R. Rousseau

*Institut de Mathématiques Appliquées
Université Catholique de l'Ouest
3, place André Leroy, 49000 Angers, France*

RÉSUMÉ

Un tableau de correspondance peut présenter, dans la pratique, quelques cases dont les poids sont prépondérants par rapport à ceux des autres cases du tableau. En considérant, sur chacun des deux ensembles (individus et variables) du tableau de correspondance, la partition en deux classes (éléments à marges prépondérantes et éléments à marges négligeables), cet article étudie la limite de l'analyse des correspondances d'un tel tableau quand les poids des cases négligeables tendent vers zéro.

Mots-clés : *Analyse factorielle des correspondances, limite, poids prépondérant, blocs diagonaux*

ABSTRACT

In practice, a correspondence data matrix may have some high weight elements. This article studies the limit of the correspondence analysis on a such table when the weights of the negligible elements converge at zero.

Keywords : *Correspondence analysis, limit, high weight, diagonal blocks.*

1. Position du problème

On considère un tableau de correspondance k_{IJ} de la forme suivante $k_{IJ} = A_{IJ} + \varepsilon B_{IJ}$ avec $\varepsilon > 0$, A_{IJ} et B_{IJ} étant des tableaux de nombres positifs vérifiant, pour des partitions $\{I_0, I_p\}$ et $\{J_0, J_p\}$ en deux classes de I et J respectivement, les propriétés suivantes :

$$\forall (i, j) \notin I_0 \times J_0 \quad A_{ij} = 0 \quad \text{et} \quad \forall (i, j) \in I_0 \times J_0 \quad B_{ij} = 0$$

les marges de A_{IJ} de tout élément de I_0 et de J_0 sont non nulles

les marges de B_{IJ} de tout élément de I_p et de J_p sont non nulles

Autrement dit, k_{IJ} est un tableau de correspondance du type suivant :

	J_0	J_p
I_0		ε
I_p	ε	ε

L'objet de l'article est l'étude de la limite de l'Analyse Factorielle des Correspondances (notée AFC) de k_{IJ} quand ε tend vers zéro. On étudiera, d'abord jusqu'au § 8, le cas $I_p \neq \emptyset$ et $J_p \neq \emptyset$, puis, au § 9, le cas $I_p = \emptyset$ ou $J_p = \emptyset$ et enfin, au § 10, une généralisation au cas où certaines des cases de $A_{I_0J_0} = k_{I_0J_0}$ sont aussi d'ordre ε .

2. Étude du nuage $N(I)$ des individus (pour $J_p \neq \emptyset$), (nuage des profils des lignes du tableau k_{IJ})

On utilisera les notations habituelles de l'AFC, cf. [BEN80], à savoir, pour k_{IJ} tableau de correspondance :

k_I sa marge sur I , de terme général $k(i) = \sum\{k(i, j) | j \in J\}$

k_J sa marge sur J , de terme général $k(j) = \sum\{k(i, j) | i \in I\}$

k son total, $k = \sum\{k(i) | i \in I\} = \sum\{k(j) | j \in J\}$

f_{IJ} le tableau de fréquence associé, $\forall i \in I, \forall j \in J : f_{ij} = k(i, j)/k$

de marges $f_I = \{f_i | i \in I\}$ sur I et $f_J = \{f_j | j \in J\}$ sur J .

Ces notations se transposent aux deux tableaux A_{IJ} et B_{IJ} ainsi qu'au tableau $A_{I_0J_0}$ (sous-tableau de A_{IJ}). On suppose $J_p \neq \emptyset$.

Les marges du tableau k_{IJ} sont donc les suivantes et tendent, quand ε tend vers 0, vers :

$$\begin{aligned} \forall j \in J_0 \quad k(j) &= A(j) + \varepsilon B(j) \quad \longrightarrow A(j) \\ \forall j \in J_p \quad k(j) &= \varepsilon B(j) \quad \longrightarrow 0 \end{aligned}$$

Donc, quand ε tend vers 0, f_J tend vers $(f_{J_0}, 0_{J_p})$ et par symétrie, pour $I_p \neq \emptyset$, f_I tend vers $(f_{I_0}, 0_{I_p})$. Ainsi, les masses des éléments de I_0 et de J_0 tendent vers celles associées au tableau de correspondance $A_{I_0J_0}$ et celles de I_p et J_p tendent vers zéro.

L'écriture des profils des individus, $\forall i \in I, \forall j \in J :$

$$f_j^i = \frac{k(i, j)}{k(i)} = \frac{A(i, j) + \varepsilon B(i, j)}{A(i) + \varepsilon B(i)},$$

montre les limites suivantes :

$$\begin{aligned}\forall i \in I_0, \forall j \in J_0 : f_j^i &\rightarrow A_j^i = \frac{A(i, j)}{A(i)} \\ \forall i \in I_0, \forall j \in J_p : f_j^i &\rightarrow 0 \\ \forall i \in I_p, \forall j \in J : f_j^i &= B_j^i = \frac{B(i, j)}{B(i)}\end{aligned}$$

Le profil f_j^i de l'individu i de I tend donc vers le vecteur $(A_{J_0}^i, 0_{J_p})$ pour $i \in I_0$ et est égal au vecteur constant B_j^i pour $i \in I_p$.

Le calcul de la norme $\|f_j^i\|$, avec la métrique du Chi-2 de centre f_J , montre aisément que celle-ci tend, si $i \in I_0$, vers $\|A_{J_0}^i\|$, relativement à la norme du Chi-2 de centre $f_{A_{J_0}}$ (loi marginale sur J_0 de $A_{I_0 J_0}$), et, si $i \in I_p$, vers $+\infty$ (de l'ordre de $1/\varepsilon^{1/2}$). Ainsi, les profils des individus de I_p , même invariants, ont, pour $J_p \neq \emptyset$, du fait de la modification de la distance du Chi-2 sur J , une norme tendant vers l'infini, quand ε tend vers 0.

Le sous-nuage $N(I_0)$ tend donc vers une forme limite qui est celle correspondant au sous-tableau $A_{I_0 J_0}$, tandis que les profils fixes des points de I_p de masse légère d'ordre ε s'éloignent de l'origine à des distances d'ordre $1/\varepsilon^{1/2}$.

3. Étude de la matrice f_J^J (pour $I_p \neq \emptyset$)

Pour $I_p \neq \emptyset$, en décomposant I selon sa partition $\{I_0, I_p\}$, on calcule aisément la limite du terme général, $f_{j'}^j = \sum \left\{ \frac{k(i, j')k(i, j)}{k(i)k(j)} \mid i \in I \right\}$, de la matrice f_J^J :

$$\forall j \in J_0, \forall j' \in J_0 : f_{j'}^j \rightarrow \sum \left\{ \frac{A(i, j')A(i, j)}{A(i)A(j)} \mid i \in I_0 \right\} = A_{j'}^j,$$

$$\forall j \in J_0, \forall j' \in J_p : f_{j'}^j \rightarrow 0$$

$$\forall j \in J_p, \forall j' \in J_0 :$$

$$f_{j'}^j \rightarrow \sum \left\{ \frac{A(i, j')B(i, j)}{A(i)B(j)} \mid i \in I_0 \right\} + \sum \left\{ \frac{B(i, j')B(i, j)}{B(i)B(j)} \mid i \in I_p \right\}$$

$$\forall j \in J_p, \forall j' \in J_p : f_{j'}^j \rightarrow \sum \left\{ \frac{B(i, j')B(i, j)}{B(i)B(j)} \mid i \in I_p \right\}$$

ce qui donne, comme limite de f_J^J , une matrice de la forme :

$$\lim_{\varepsilon \rightarrow 0} f_J^J = g_J^J = \begin{pmatrix} A_{J_0}^{J_0} & g_{J_0}^{J_p} \\ 0_{J_p}^{J_0} & g_{J_p}^{J_p} \end{pmatrix}.$$

On montre aisément que la matrice g_j^j est une transition probabiliste, i.e., que la somme de toute colonne est égale à 1; en effet, pour tout $j \in J_p$:

$$\begin{aligned}
\sum \{g_{j'}^j | j' \in J\} &= \sum \{g_{j'}^j | j' \in J_0\} + \sum \{g_{j'}^j | j' \in J_p\} \\
&= \sum \{ \sum \{A_{j'}^i, B_i^j | i \in I_0\} + \sum \{B_{j'}^i, B_i^j | i \in I_p\} | j' \in J_0\} \\
&\quad + \sum \{ \sum \{B_{j'}^i, B_i^j | i \in I_p\} | j' \in J_p\} \\
&= \sum \{ \sum \{A_{j'}^i, B_i^j | i \in I_0\} | j' \in J_0\} + \sum \{ \sum \{B_{j'}^i, B_i^j | i \in I_p\} | j' \in J\} \\
&= \sum \{ \sum \{A_{j'}^i, B_i^j | i \in I_0\} | j' \in J_0\} - \sum \{ \sum \{B_{j'}^i, B_i^j | i \in I_0\} | j' \in J\} \\
&\quad + \sum \{ \sum \{B_{j'}^i, B_i^j | i \in I\} | j' \in J\} \\
&= \sum \{B_i^j (\sum \{A_{j'}^i | j' \in J_0\} - \sum \{B_{j'}^i | j' \in J\}) | i \in I_0\} \\
&\quad + \sum \{B_{j'}^j | j' \in J\} \\
&= \sum \{B_i^j (1 - 1) | i \in I_0\} + 1 \\
&= 1
\end{aligned}$$

Il est à noter que, sauf pour $g_{j_0}^{j_0} = 0_{j_0}^{j_0}$, la forme particulière de la matrice g_j^j n'existe jamais pour la matrice f_j^j associée à un tableau de correspondance quelconque. En effet, les égalités suivantes $\langle f_I^j, f_I^{j'} \rangle_{f_I} = \frac{1}{f_j^j} f_j^{j'} = \frac{1}{f_{j'}^{j'}} f_{j'}^j$ impliquent $f_j^{j'} = 0$ si et seulement si $f_{j'}^j = 0$ ($\langle f_I^j, f_I^{j'} \rangle_{f_I}$ désignant le produit scalaire entre f_I^j et $f_I^{j'}$ dans la métrique du Chi-2 de centre f_I).

Le cas particulier $g_{j_0}^{j_0} = 0_{j_0}^{j_0}$ correspond à la situation suivante :

$$\forall j \in J_p, \forall j' \in J_0 :$$

$$g_{j'}^j = \sum \left\{ \frac{A(i, j')B(i, j)}{A(i)B(j)} | i \in I_0 \right\} + \sum \left\{ \frac{B(i, j')B(i, j)}{B(i)B(j)} | i \in I_p \right\} = 0$$

$$\text{i.e. } \forall j \in J_p, \forall j' \in J_0$$

$$(\forall i \in I_0 : A(i, j')B(i, j) = 0) \text{ et } \forall i \in I_p : B(i, j')B(i, j) = 0)$$

$$\text{i.e. } \forall i \in I_0, \forall j \in J_p, \forall j' \in J_0 : A(i, j')B(i, j) = 0$$

$$\text{et } \forall i \in I_p, \forall j \in J_p, \forall j' \in J_0 : B(i, j')B(i, j) = 0$$

$$\text{i.e. } \forall i \in I_0, (\text{ou } \exists j' \in J_0 : A(i, j') \neq 0 \text{ et alors } \forall j \in J_p : B(i, j) = 0, \text{ ou } \forall j' \in J_0 : A(i, j') = 0)$$

$$\text{et } \forall i \in I_p (\text{ou } \exists j' \in J_0 : B(i, j') \neq 0 \text{ et alors } \forall j \in J_p : B(i, j) = 0, \text{ ou } \forall j' \in J_0 : B(i, j') = 0)$$

$$\text{i.e. } \forall i \in I_0, \forall j \in J_p : B(i, j) = 0$$

(du fait que la marge de i sur I_0 pour le tableau A_{IJ} est non nulle par hypothèse)

et $\forall i \in I_p$ (ou $\exists j' \in J_0 : B(i, j') \neq 0$ et alors $\forall j \in J_p : B(i, j) = 0$, ou $\forall j' \in J_0 : B(i, j') = 0$)

ce qui est équivalent à une décomposition du tableau k_{IJ} sous la forme suivante :

	J_0	J_p
I_0		0
I_{p1}	ε	0
I_{p2}	0	ε

Dans ce cas, le tableau k_{IJ} est décomposable en deux blocs diagonaux et son AFC revient donc, à des termes correctifs près dépendant des masses des deux blocs, à l'AFC du premier bloc (cas particulier du problème présent avec $J_p = \emptyset$), à l'AFC invariante du second bloc et à ajouter le facteur opposant les deux blocs, cf. [BEN85].

NB : si de plus $I_{p1} = \emptyset$, alors le tableau k_{IJ} est décomposable en deux blocs diagonaux selon les partitions $\{I_0, I_p\}$ et $\{J_0, J_p\}$ de I et J .

Un autre cas particulier de la matrice g_J^J existe, à savoir $g_{Jp}^{Jp} = 0_{Jp}^{Jp}$, correspondant au cas suivant :

$$\forall j \in J_p, \forall j' \in J_p : g_{j'}^j = \sum \left\{ \frac{B(i, j')B(i, j)}{B(i)B(j)} \mid i \in I_p \right\} = 0$$

$$\text{i.e. } \forall j \in J_p, \forall j' \in J_p, \forall i \in I_p : B(i, j')B(i, j) = 0$$

$$\text{i.e. } \forall i \in I_p, \forall j \in J_p : B(i, j) = 0$$

$$\text{i.e. } B_{I_p J_p} = 0_{I_p J_p}$$

i.e. aucune interaction entre deux éléments quelconques de I_p et J_p

4. Diagonalisation de la matrice g_J^J et de la matrice g_I^I ($I_p \neq \emptyset$ et $J_p \neq \emptyset$)

Si 1 est valeur propre de f_J^J de multiplicité $k \geq 2$, le tableau k_{IJ} est alors décomposable en k blocs diagonaux et son AFC revient à faire les AFC des k blocs auxquelles on rajoute les $(k-1)$ facteurs séparant les k blocs. Dans ce cas, cela revient à étudier k limites d'AFC et à faire les correctifs nécessaires d'après les masses des blocs. Aussi, on supposera dorénavant que le tableau k_{IJ} n'est pas décomposable en blocs diagonaux (ce qui élimine le cas particulier $g_{J_0}^{J_p} = 0_{J_0}^{J_p}$ précité au paragraphe précédent). On suppose $I_p \neq \emptyset$ et $J_p \neq \emptyset$.

Vu la forme particulière de la matrice $g_J^J = \begin{pmatrix} A_{J_0}^{J_0} & g_{J_0}^{J_p} \\ 0_{J_p}^{J_0} & g_{J_p}^{J_p} \end{pmatrix}$, ses valeurs propres sont celles de $A_{J_0}^{J_0}$ et celles de $g_{J_p}^{J_p}$. On notera Λ , Λ_0 et Λ_p les ensembles

indiquant les valeurs propres non nulles de g_J^J , $A_{J_0}^{J_0}$ et $g_{J_p}^{J_p}$ respectivement. On a donc l'égalité $\Lambda = \Lambda_0 \cup \Lambda_p$ avec $\Lambda_p = \emptyset$ si et seulement si $B_{I_p J_p} = 0_{I_p J_p}$. La forme particulière de g_J^J implique de plus que les facteurs sur J , vecteurs propres à gauche de g_J^J , relatifs à $\alpha \in \Lambda_p$ peuvent s'écrire sous la forme $(0^{J_0}, \varphi^{J_p})$ et que les axes factoriels dans R_J , vecteurs propres à droite de g_J^J , relatifs à $\alpha \in \Lambda_0$ peuvent s'écrire sous la forme ${}^t(U_{J_0}, 0_{J_p})$. Par symétrie, la matrice f_I^I tend vers la matrice

$$g_I^I = \lim_{\varepsilon \rightarrow 0} f_I^I = \begin{pmatrix} A_{I_0}^{I_0} & g_{I_0}^{I_p} \\ 0_{I_p}^{I_0} & g_{I_p}^{I_p} \end{pmatrix}. \text{ Les valeurs propres non nulles de } f_I^I \text{ et } g_J^J \text{ étant}$$

identiques, ainsi que celles de $A_{I_0}^{I_0}$ et $A_{J_0}^{J_0}$, celles de $g_{I_p}^{I_p}$ et $g_{J_p}^{J_p}$ sont donc identiques. De plus, les facteurs sur I relatifs à $\alpha \in \Lambda_p$ peuvent s'écrire sous la forme $(0^{I_0}, \varphi^{I_p})$ et les axes factoriels dans R_I relatifs à $\alpha \in \Lambda_0$ peuvent s'écrire sous la forme ${}^t(U_{I_0}, 0_{I_p})$.

On notera également que le facteur trivial δ^J sur J (i.e. le facteur constant et égal à 1 sur J) est bien facteur de g_J^J (car g_J^J est une transition probabiliste), qu'il n'est pas associé à une valeur propre de $g_{J_p}^{J_p}$ vu sa forme et vu le fait que $k_{I,J}$ n'est pas décomposable en blocs diagonaux, et qu'il est donc associé à la valeur propre 1 de $A_{J_0}^{J_0}$. Idem pour le facteur trivial δ^I sur I .

5. Facteurs sur I et sur J associés à Λ_0 ($I_p \neq \emptyset$ et $J_p \neq \emptyset$)

Les axes factoriels du tableau $A_{I_0 J_0}$ engendrant le support de $N(I_0)$ sont donc, à la limite quand ε tend vers zéro, axes factoriels du nuage entier $N(I)$. Les points $i_0 \in I_0$ sont, à la limite quand ε tend vers zéro, représentés sur ces axes comme dans l'AFC de $A_{I_0 J_0}$, tandis que les points $i_p \in I_p$ y sont représentés par des points-limites qui ne sont pas, en général, associés à leurs projections orthogonales sur le support de $N(I_0)$ (pour la métrique du Chi-2 de centre $f_{A_{J_0}}$, loi marginale sur J_0 de $A_{I_0 J_0}$).

En effet, on a : $\forall i_p \in I_p, \forall i_0 \in I_0$:

$$\begin{aligned} \langle f_{J_0}^{i_0}, f_{J_0}^{i_p} \rangle f_J &= \frac{1}{f_{i_0}^{i_0}} f_{i_0}^{i_p} \\ &\longrightarrow \frac{1}{f_{A_{i_0}}} g_{i_0}^{i_p} = \sum \left\{ \frac{A}{A(j)} \frac{A(i_0, j) B(i_p, j)}{A(i_0) B(i_p)} \mid j \in J_0 \right\} \\ &\quad + \sum \left\{ \frac{A}{B(j)} \frac{B(i_0, j) B(i_p, j)}{A(i_0) B(i_p)} \mid j \in J_p \right\} \\ &\longrightarrow \langle A_{J_0}^{i_0}, B_{J_0}^{i_p} \rangle_{f_{A_{J_0}}} + \sum \left\{ \frac{A}{B(j)} \frac{B(i_0, j) B(i_p, j)}{A(i_0) B(i_p)} \mid j \in J_p \right\} \end{aligned}$$

où A représente le total du tableau $A_{I,J}$, i.e., du bloc $A_{I_0 J_0}$.

Ainsi, l'élément $i_p \in I_p$ possède une représentation, à la limite quand ε tend vers zéro, généralement différente de celle que l'on obtiendrait en la mettant en supplémentaire au tableau $A_{I_0 J_0}$, sauf s'il est à termes nuls sur J_p (ce qui est vérifié quand $B_{I_p J_p} = 0_{I_p J_p}$) ou s'il est orthogonal sur J_p à tous les éléments de I_0 (ce qui correspond, si cela est vrai pour tous les éléments de I_p , à une décomposition en deux blocs du sous-tableau k_{I_p} avec $\{I_0, I_p\}$ comme partition sur I).

6. Facteurs associés à Λ_p ($I_p \neq \emptyset$ et $J_p \neq \emptyset$)

Considérons le tableau de correspondance $h_{I',J'}$ défini à partir du tableau k_{IJ} de la façon suivante : on cumule les lignes de I_0 en une seule ligne i_0 , de même on cumule les colonnes de J_0 en une seule colonne j_0 . Autrement dit, on note $I' = \{i_0\} \cup I_p$ et $J' = \{j_0\} \cup J_p$. Cette opération permet au tableau $h_{I',J'}$ d'avoir les mêmes marges sur I_p et sur J_p que dans le tableau k_{IJ} . On obtient ainsi un tableau du type suivant :

	i_0	J_p
i_0	ε	ε
I_p	ε	ε

qui est un cas particulier du tableau k_{IJ} avec $I_0 = \{i_0\}$ et $J_0 = \{j_0\}$. L'étude précédente montre donc les résultats suivants :

$$H_{J'}^{J'} = \lim_{\varepsilon \rightarrow 0} h_{J'}^{J'} = \begin{pmatrix} 1 & H^{J_p} \\ 0_{J_p} & g_{J_p}^{J_p} \end{pmatrix}$$

car

$$\forall j \in J_p, \forall j' \in J_p : h_{j'}^j \longrightarrow \sum \left\{ \frac{B(i, j')B(i, j)}{B(i)B(j)} \mid i \in I_p \right\} = g_{j'}^j$$

- $H_{J'}^{J'}$ est une transition probabiliste;
- le facteur trivial $\delta^{J'}$ est l'unique facteur sur J' associé à l'unique valeur propre 1 de $A_{J_0}^{J_0}$ réduite à l'élément 1;
- $\Lambda_p = \emptyset$ si et seulement si $B_{I_p J_p} = 0_{I_p J_p}$;
- les facteurs sur J' associés aux valeurs propres non nulles de $g_{J_p}^{J_p}$ sont de la forme $(0, \varphi^{J_p})$.

On peut remarquer que $\forall j \in J_p : H^j = \sum \{g_{j_0}^j \mid j_0 \in J_0\}$ du fait que g_j^j et $H_{J'}^{J'}$ sont des transitions probabilistes, i.e., la première ligne de $H_{J'}^{J'}$ est le cumul des lignes de J_0 de la matrice $g_{J'}^J$.

Ainsi, pour $B_{I_p J_p} \neq 0_{I_p J_p}$, les axes factoriels de la projection orthogonale de $N(I_p)$ dans l'orthogonal du support de $N(I_0)$ sont, à la limite quand ε tend vers zéro, ceux de l'AFC de $h_{I',J'}$. On peut remarquer aussi que les valeurs propres de Λ_p sont d'ordre comparable à celles de Λ_0 et donc que le tableau B_{IJ} crée des interférences dans l'AFC de k_{IJ} de même ordre de grandeur que dans l'AFC du tableau $A_{I_0 J_0}$.

On retrouve ici le cas IV n°10 [CASE PRED], analyse d'un tableau dont une case a un poids déterminant, in [BEN85]. Ce cas est un tableau particulier du type général k_{IJ} défini dans cet article, dans lequel I_0 et J_0 sont tous deux réduits à un

seul élément. On généralise donc dans cet article les résultats obtenus dans le cas IV n° 10, à une nuance près concernant l'interprétation de la valeur propre 1 : celle-ci est la limite de la valeur propre triviale (liée au fait que la matrice $H_{J'}^{J'}$ est une transition probabiliste) et non une valeur propre non triviale liée à une décomposition en blocs diagonaux de la limite du tableau $h_{I'J'}$ (cf. exemple suivant § 7).

7. Exemple 1 d'un tableau à une seule case de poids prédominant

	a	b
1	1	0
2	ε	2ε
3	0	ε

On considère le tableau de correspondance ci-dessus où $\varepsilon > 0$. Ce tableau ne possède qu'une seule valeur propre non triviale λ . On montre aisément, après calculs, les résultats suivants :

$$f_I = (1/(1+4\varepsilon), 3\varepsilon/(1+4\varepsilon), \varepsilon/(1+4\varepsilon)) \longrightarrow (1, 0, 0)$$

$$f_J = ((1+\varepsilon)/(1+4\varepsilon), 3\varepsilon/(1+4\varepsilon)) \longrightarrow (1, 0)$$

$$\lambda = (7+\varepsilon)/(9(1+\varepsilon)) \longrightarrow 7/9 \neq 1$$

$$\varphi^J = [1/(3\varepsilon(1+\varepsilon))]^{1/2}(3\varepsilon, -(1+\varepsilon)) \text{ et donc } \Psi^J = \lambda^{1/2}\varphi^J \longrightarrow (0, -\infty)$$

$$\Psi^I = [1/(3\varepsilon(1+\varepsilon))]^{1/2}(3\varepsilon, (\varepsilon-2)/3, -(1+\varepsilon)) \longrightarrow (0, -\infty, -\infty)$$

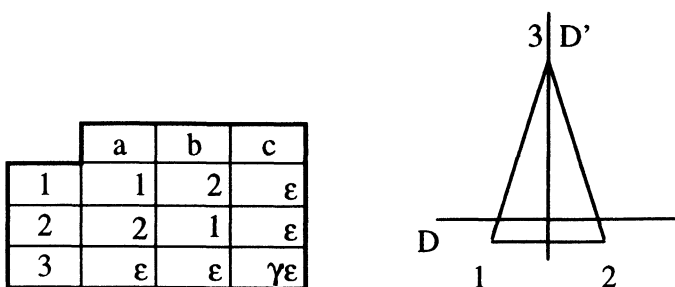
φ^J désignant le facteur non trivial sur J de variance 1, et Ψ^I et Ψ^J les composantes principales sur I et J respectivement.

$$f_J^J \longrightarrow g_J^J = \begin{pmatrix} 1 & 2/9 \\ 0 & 7/9 \end{pmatrix} \text{ ayant comme valeurs propres 1 et } 7/9.$$

On y constate que la limite de la seule valeur propre non triviale est différente de 1 et on y retrouve les résultats énoncés précédemment sur les facteurs.

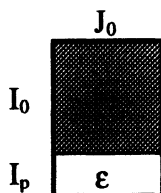
8. Exemple 2 avec card $I_0 = 2$

On considère le tableau de correspondance ci-dessous où $\varepsilon > 0$ et $\gamma > 0$. Le triangle 123 est isocèle de sommet 3. La médiatrice D' passant par 3 est toujours axe factoriel du nuage $N(I)$. L'autre axe factoriel D est donc orthogonal à cette médiatrice et passe par le centre de gravité du nuage. A la limite sur ε tendant vers zéro, l'axe factoriel D tend vers la droite portée par le segment [1,2] et le facteur associé sur I de la forme $(\alpha, -\alpha, 0)$ tend vers l'unique facteur du sous-tableau $\{1, 2\} \times \{a, b\}$. Le second facteur sur I tend vers $(0, 0, +\infty)$ puisque $\gamma > 0$.



9. Cas particuliers $I_p = \emptyset$ ou $J_p = \emptyset$

I et J jouant des rôles symétriques en AFC, le cas $I_p = \emptyset$ est identique au cas $J_p = \emptyset$. On supposera donc $J_p = \emptyset$, i.e., $k_{IJ_0} = A_{IJ_0} + \epsilon B_{IJ_0}$. Le tableau k_{IJ} est donc de la forme :



A la différence du cas $J_p \neq \emptyset$, la norme $\|f_J^{ip}\|$ tend vers une limite finie qui est celle de son profil constant B_J^i pour la norme de la métrique du Chi-2 de centre fA_{J_0} (loi marginale sur J_0 de $A_{I_0J_0}$).

En reprenant la démarche du § 3, on montre que $g_J^J = \lim_{\epsilon \rightarrow 0} f_J^J = A_{J_0}^{J_0}$ et $g_I^I = \lim_{\epsilon \rightarrow 0} f_I^I = \begin{pmatrix} A_{I_0}^{I_0} & g_{I_0}^{I_p} \\ 0_{I_p}^{I_0} & 0_{I_p}^{I_p} \end{pmatrix}$. Les axes factoriels du nuage $N(I)$ dans $R_J = R_{J_0}$ tendent vers ceux de l'AFC de $A_{I_0J_0}$ (ce qui est normal, vu que les éléments de I_p de profils fixes ont des masses tendant vers zéro) et les facteurs sur $J = J_0$ tendent vers ceux de l'AFC de $A_{I_0J_0}$. Les axes factoriels du nuage $N(J)$ dans R_I tendent sur I_0 vers ceux de l'AFC de $A_{I_0J_0}$ et sur I_p vers zéro. Les facteurs sur I tendent sur I_0 vers ceux de l'AFC de $A_{I_0J_0}$ et sur I_p vers ceux associés aux projections orthogonales (pour la métrique du Chi2 de centre fA_{J_0} , loi marginale sur J_0 de $A_{I_0J_0}$) des éléments de I_p sur le support de $N(I_0)$ (qui interviennent de moins en moins du fait de leurs masses tendant vers zéro), ce qui correspond à la situation de mise en éléments supplémentaires au tableau $A_{I_0J_0}$. On retrouve donc les résultats précédents sans avoir ici l'interférence des facteurs associés à Λ_p et donc du bloc $I_p \cdot J_p$.

10. Généralisation et exemple d'un tableau à deux cases de poids prédominants

On a supposé, jusqu'à présent, qu'aucune case du sous-tableau $A_{I_0 J_0}$ n'était d'ordre ε . Les démonstrations précédentes s'étendent aisément au cas où certaines d'entre elles le sont, à condition que toutes les lignes et colonnes de $A_{I_0 J_0}$ aient des marges d'ordre 1.

Par exemple, considérons un tableau à deux cases de poids prépondérants de même ordre. Notons (i_0, j_0) et (i_1, j_1) ces deux cases avec $i_0 \neq i_1$ et $j_0 \neq j_1$. En effet, si $j_0 = j_1$ (le cas $i_0 = i_1$ est analogue), la matrice $A_{J_0}^{J_0}$ se réduit alors à une seule case, ce qui nous ramène au cas d'un tableau ayant une seule case à poids prépondérant.

Supposons de plus $k(i_0, j_0) = 1$ et $k(i_1, j_1) = \gamma$. On a donc $I_0 = \{i_0, i_1\}$ et $J_0 = \{j_0, j_1\}$.

La matrice limite est donc

$$g_J^J = \lim_{\varepsilon \rightarrow 0} f_J^J = \begin{pmatrix} A_{J_0}^{J_0} & g_{J_0}^{J_p} \\ 0_{J_p}^{J_0} & g_{J_p}^{J_p} \end{pmatrix} \quad \text{avec} \quad A_{J_0}^{J_0} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Il existera, à la limite sur ε tendant vers zéro, une valeur propre 1 non triviale qui opposera les éléments i_0 et i_1 , les éléments de I_p se situant d'après leurs projections-limites sur cet axe, comme le montre aussi la forme du facteur associé sur I égale à $(\approx \gamma, \approx -1, \varphi^{I_p})$ à un coefficient de normalisation près. On a un résultat analogue pour les éléments j_0 et j_1 et ceux de J_p .

11. Conclusion

En conclusion, pour I_p et J_p non vides, les ensembles I_0 et J_0 représentent les individus i_0 et les variables j_0 de la 1^{ère} classe ayant entre eux des relations $k(i_0, j_0)$ dont les marges $k(i_0) = \sum \{k(i_0, j_0) | j_0 \in J_0\}$ et $k(j_0) = \sum \{k(i_0, j_0) | i_0 \in I_0\}$ ont un ordre de grandeur égal à 1. Il en est donc nécessairement de même pour l'une des cases $k(i_0, j)$, où $j \in J_0$, et l'une des cases $k(i, j_0)$, où $i \in I_0$. Les ensembles I_p et J_p représentent les individus et les variables de la seconde classe ayant entre eux, et avec les éléments de I_0 et J_0 , des relations dont l'ordre de grandeur est ε .

La structure globale de k_{IJ} donnée par son AFC est, à la limite quand ε tend vers zéro, obtenue notamment par la structure de $k_{I_0 J_0}$ donnée par son AFC : les axes factoriels de $N(I)$ et $N(J)$ étant ceux de $N(I_0)$ et $N(J_0)$ respectivement, les facteurs sur I_0 et J_0 sont ceux de l'AFC de $k_{I_0 J_0}$ auxquels les éléments de I_p et J_p s'ajoutent par leurs projections-limites sur les supports de $N(I_0)$ et de $N(J_0)$ respectivement, projections-limites qui diffèrent en général (sauf notamment quand $B_{I_p J_p} = 0_{I_p J_p}$) de celles que l'on obtiendrait en mettant ces éléments en supplémentaire au tableau $A_{I_0 J_0}$.

De plus, pour $B_{I_p J_p} \neq 0_{I_p J_p}$, la structure globale de k_{IJ} doit tenir compte aussi des seuls individus et variables de la seconde classe qui définissent des axes

de variances comparables à celles des axes de $k_{I_0 J_0}$, des facteurs tendant vers zéro sur I_0 et J_0 et pouvant être d'ordre $1/\varepsilon^{1/2}$ sur I_p et J_p . Il est à noter que les valeurs propres des deux groupes peuvent se mêler et même certaines liées aux individus et variables de la seconde classe être supérieures à celles liées aux individus et variables de la première classe, l'exemple typique étant celui du tableau k_{IJ} décomposé en les deux blocs diagonaux $A_{I_0 J_0}$ et $B_{I_p J_p}$ (les partitions $\{I_0, I_p\}$ et $\{J_0, J_p\}$ de I et de J étant celles de la décomposition en blocs) où les valeurs propres sont invariables.

Dans le cas où I_p (ou J_p) est vide, les résultats précédents restent évidemment valables avec la simplification que les interférences dues aux individus et variables de la seconde classe n'existent pas et ne créent pas de structure particulière dans la structure globale de k_{IJ} et, avec la propriété toujours vérifiée, que les résultats sur I_p (ou J_p) sont, à la limite quand ε tend vers zéro, analogues à ceux obtenus en mettant cet ensemble en supplémentaire au tableau $k_{I_0 J_0}$.

Références

- [BEN80] J.P. et F. BENZECRI, *Pratique de l'analyse des données, tome 1, analyse des correspondances, (exposé élémentaire)*, Bordas, Paris, 1980.
- [BEN85] Ch. BASTIN, J.P. BENZECRI, CH. BOURGARIT, P. CAZES, *Pratique de l'analyse des données, tome 2, (abrégé théorique, étude de cas modèle)*, Dunod, 2^e édition, Paris, 1985.