

P. CAZES

Note sur les éléments supplémentaires en analyse des correspondances. II. Tableaux multiples

Les cahiers de l'analyse des données, tome 7, n° 2 (1982), p. 133-154

http://www.numdam.org/item?id=CAD_1982__7_2_133_0

© Les cahiers de l'analyse des données, Dunod, 1982, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » 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/>

NOTE SUR LES ÉLÉMENTS SUPPLÉMENTAIRES EN ANALYSE DES CORRESPONDANCES

II. TABLEAUX MULTIPLES [EL. SUPP. II]

par P. Cazes (1)

4 Cas d'un tableau ternaire

Nous envisageons ici une série de tableaux $\{(k_t)_{I_t \times J_t} \mid t \in T\}$ indicés par t , t décrivant un ensemble fini T : ce peut être par exemple une série de tableaux relevés à différentes époques. Nous désignerons par IT (resp. JT) l'union disjointe des I_t (resp. J_t) :

$$IT = \cup \{I_t \mid t \in T\} \quad ; \quad JT = \cup \{J_t \mid t \in T\}$$

Nous supposons qu'au moins un des deux ensembles en correspondance est indépendant de t , ce qui donne les trois situations possibles :

- a) $\forall t \in T : J_t = J$, I_t quelconque
- b) $\forall t \in T : I_t = I$, J_t quelconque
- c) $\forall t \in T : I_t = I$, $J_t = J$

Nous n'étudierons pas le cas b) qui se déduit du cas a) par interversion des rôles de I et J , ainsi que de ceux de IT et JT . Par contre, bien que le cas c) puisse être considéré comme un cas particulier de a) (ou de b)), du fait du format des tableaux, il se prête à des analyses intéressantes non possibles en général dans les cas a) et b). Nous examinerons donc d'abord le cas a) (§ 4.1), avant de développer le cas c) (§ 4.2).

Nous supposons pour fixer les idées que $T = \{1, 2, \dots, r\} =]r]$, et sans perte de généralité (i.e. en ne tenant pas compte de l'ordre naturel qui lui est associé) qu'il correspond au temps.

4.1 Etude du cas où $J_t = J$

Dans ce cas, on peut placer les tableaux $(k_t)_{I_t \times J}$ les uns en dessous des autres, ce qui revient à considérer le tableau $k_{IT \times J}$ croisant IT avec J :

$$k_{IT \times J} = ((k_1)_{I_1 \times J}, (k_2)_{I_2 \times J}, \dots, (k_r)_{I_r \times J})$$

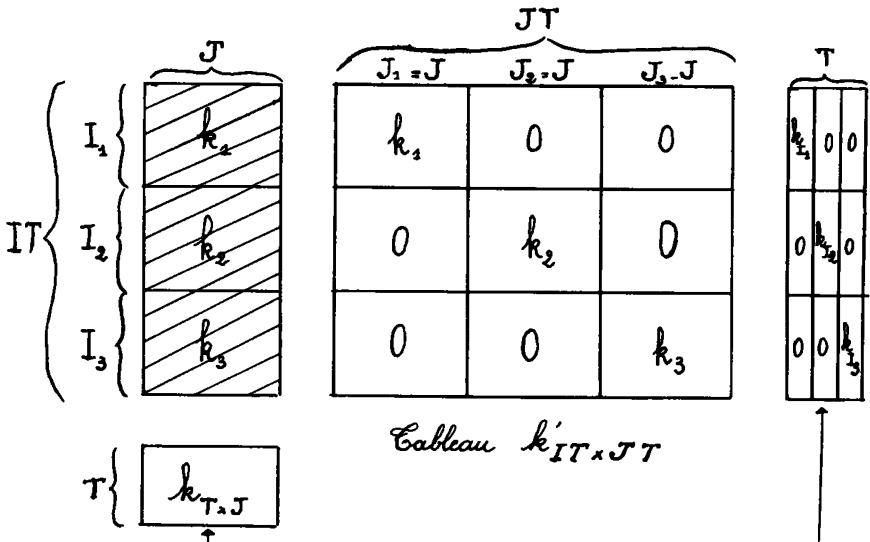
* Suite de l'article paru sous le même titre, dans les Cahiers, Vol VII n° 1, pp 9-23.

(1) Professeur de statistique à l'Université de Paris Dauphine.

et classiquement on effectue l'analyse des correspondances du tableau $k_{IT \times J}$, ce qui permet de représenter IT et J . On peut représenter chaque élément t de T , en représentant le sous-ensemble I_t de IT à l'aide de la 2-ème méthode (colonne supplémentaire) ou de la 3-ème méthode (ligne supplémentaire) décrite au § 3 ; dans ce dernier cas, chaque t de T est représenté par le centre de gravité des éléments i de I_t , ce qui revient à adjoindre en lignes supplémentaires au tableau $k_{IT \times J}$ le tableau $k_{T \times J}$ défini par :

$$\forall (t, j) \in T \times J : k(t, j) = \Sigma \{k_t(i, j) \mid i \in I_t\} \tag{8}$$

Si l'on veut représenter chaque élément j de J à chaque instant t de T , ce qui revient à représenter l'ensemble JT (égal ici à $J \times T$) on adjoint en élément supplémentaire du tableau $k_{IT \times J}$, le tableau $k'_{IT \times JT}$ ainsi défini : tous les blocs non diagonaux $I_t \times J_t$, ($= I_t \times J$) du tableau k' sont nuls ($t \neq t'$) tandis que tous les blocs diagonaux $I_t \times J_t$ ($= I_t \times J$) sont identiques au tableau $(k_t)_{I_t \times J_t}$ (cf. figure 2).



Représentation de T (méthode du §3.5) :
 La ligne t_j du tableau $k_{T \times J}$ correspond à la marge sur J (k_{t_j}) du tableau k_t .
 Le tableau $k_{T \times J}$ permet de calculer la variance interclasse (inter t) dans R_J ou sur un axe α

Représentation de T (méthode du §3.2) :
 k_T est la marge sur I_t du tableau k_t .
 La colonne t_{IT} associée à t est aussi la marge sur IT du bloc $IT \times J_t$ du tableau $k'_{IT \times JT}$

Analyse d'une suite de tableaux $\{(k_t)_{I_t \times J_t} \mid t \in T\}$.
 (On a hachuré le tableau principal analysé $k_{IT \times J}$).

Figure 2

Soit j un élément de J , $jt (1 \leq t \leq r)$ cet élément à l'instant t , on a :

$$\forall i \in IT : k(i, j) = \Sigma \{k'(i, jt) | t \in T\} \quad (9)$$

puisque, si $i \in I_t$, tous les $k'(i, jt)$ sont nuls par construction, sauf $k'(i, jt')$ qui est égal à $k_t(i, j)$ i.e. à $k(i, j)$.

On déduit de la relation (9) que dans l'analyse de $k_{IT \times J}$, chaque point j de J est au centre de gravité des $jt (t \in T)$, jt étant affecté de la masse $k_t(j) = \Sigma \{k'(i, jt) | i \in IT\} = \Sigma \{k_t(i, j) | i \in I_t\}$ ce qui permet de voir l'évolution (si du moins T est muni d'un ordre naturel, comme c'est le cas si c'est effectivement le temps) et la dispersion des jt autour de j .

Soit $(F_\alpha^{IT}, G_\alpha^J)$ un couple de facteurs associés de variance la valeur propre λ_α , issus du tableau $k_{IT \times J}$. La valeur $G_\alpha(jt)$ du facteur α pour l'élément supplémentaire jt s'écrit :

$$G_\alpha(jt) = \Sigma \{k_t(i, j) F_\alpha(i) | i \in I_t\} / (\sqrt{\lambda_\alpha} k_t(j)) \quad (10)$$

Le principe barycentrique relativement à $I_t \times J_t$ reste donc valable : sur un axe factoriel α , jt est au facteur $(1/\sqrt{\lambda_\alpha})$ près au centre de gravité des éléments i de I_t , i étant affecté de la masse $k_t(i, j)$ (de même que it est (toujours au facteur $1/\sqrt{\lambda_\alpha}$ près) au centre de gravité des j de J affectés des masses $k_t(i, j)$).

La représentation de JT permet donc d'interpréter plus facilement les résultats de l'analyse de $k_{IT \times J}$. Cette représentation a été en particulier utilisée par M. Jabbour dans sa thèse, comme on l'a signalé au § 1.3.

Remarques

1) On déduit de (10) que sur un axe factoriel α , l'abscisse $G_\alpha(J_t)$ du centre de gravité des jt pour jt décrivant J_t , jt étant affecté de la masse $k_t(j)$, est au facteur $(1/\sqrt{\lambda_\alpha})$ près au centre de gravité des éléments i de I_t , i étant affecté de la masse $k_t(i) = \Sigma \{k_t(i, j) | j \in J_t\} (= \Sigma \{k(i, j) | j \in J\})$.

Il en résulte que si on a :

$$\forall j \in J : k_t(j) = c_t k(j) \quad (11)$$

où $k(j)$ désigne la somme de la colonne j du tableau $k_{IT \times J}$, et où c_t est une constante indépendante de j , on a $G_\alpha(J_t) = 0$ puisque dans ce dernier cas, le centre de gravité du sous-ensemble I_t de I coïncide avec le centre de gravité global (i.e. le centre de gravité de IT).

Si donc la relation (11) est vérifiée, le centre de gravité des j_t de J_t est confondu avec celui de J (et d'après (9) avec celui de JT). Ce cas est en particulier vérifié pour tout t de T si chaque tableau $k_{I_t J}$ est un tableau disjonctif complet (dont J est l'ensemble des individus et I_t l'ensemble des modalités) ou si c'est un tableau dédoublé (l'ensemble dédoublé étant l'ensemble I_t).

2) Si on garde tous les facteurs non triviaux issus du tableau $k_{IT \times J}$, on reconstitue exactement le tableau supplémentaire $k_{T \times J}$, puisque chaque ligne t de ce tableau est le centre de gravité des i de I_t qui sont parfaitement reconstitués.

3) Le tableau $k_{IT \times J}$ peut être considéré comme un tableau du type de celui étudié au § 3.4, où I est remplacé par IT , et C par T . On peut donc lui appliquer toutes les considérations développées dans ce §, dont certaines sont reprises dans les remarques suivantes.

4) Pour voir l'influence de I_t sur le facteur α issu de $k_{IT \times J}$, on peut calculer la contribution $CR_\alpha(I_t)$ de I_t (somme des contributions des éléments de I_t) à ce facteur, et reporter cette contribution à la valeur propre λ_α associée à α , ce qui donne la part relative de I_t dans la détermination du facteur α , part que l'on pourra noter $CTR_\alpha(I_t)$. Si l'on rapporte $CR_\alpha(I_t)$ à la contribution totale (i. e. à l'inertie) de I_t , on obtient la part relative, qu'on notera $COR_\alpha(I_t)$, de α à la reconstitution des éléments de I_t . Les notations CTR et COR adoptées pour les contributions relatives de I_t à α et de α à I_t respectivement, correspondent aux notations usuelles employées quand I_t se réduit à un élément.

On a également intérêt, en reprenant les considérations d'inerties interclasse et intraclasse développées au § 3.4 (dont on conservera les notations, à ceci près, comme on l'a déjà dit, qu'il faut remplacer I par IT et C par T) à considérer les 3 quantités suivantes :

$$\begin{aligned} \Sigma\{CTR_\alpha(i) \mid i \in I_t\} - CTR_\alpha(t_J) &= CTR_\alpha(I_t) - CTR_\alpha(t_J) \\ &= In_\alpha(I_t) / \lambda_\alpha \end{aligned}$$

Cette quantité représente l'inertie de I_t sur l'axe α rapportée à l'inertie totale λ_α de cet axe.

$$\Sigma\{CTR_\alpha(t_J) \mid t \in T\} = In_\alpha(T) / \lambda_\alpha = CTR_\alpha(T)$$

Cette quantité représente la variance interclasse (intertemps) du facteur α , rapportée à λ_α .

$$\begin{aligned} \Sigma\{(CTR_\alpha(I_t) - CTR_\alpha(t_J)) \mid t \in T\} &= CTR_\alpha(IT - T) \\ &= 1 - CTR_\alpha(T) \end{aligned}$$

Cette quantité représente la variance intraclasse (intratemp) du facteur α , rapportée à λ_α .

Ces trois quantités, facilement calculables à partir des résultats de l'analyse factorielle, permettent de voir si un axe est déterminé par le temps, ou par la dispersion d'un ou plusieurs ensembles I_t , ou par les deux.

On peut également faire des calculs analogues aux précédents en considérant l'inertie dans l'espace R_J tout entier.

Les calculs de contributions et d'inerties considérés ici sont explicités au § 5 dans le cas particulier, fort important en pratique, des correspondances multiples.

5) On peut, au lieu d'analyser $k_{IT \times J}$ (avec $k_{T \times J}$ en élément supplémentaire) analyser le tableau $k_{T \times J}$ avec $k_{IT \times J}$ en supplémentaire. Pour comparer les résultats de ces deux façons de procéder, on peut comme on l'a déjà dit (cf. remarque 3 du § 3.4) calculer les corrélations entre facteurs issus de ces deux analyses, corrélations qui peuvent s'exprimer à l'aide des éléments supplémentaires (cf. § 6). L'intérêt de ces calculs qui complètent les calculs de contributions donnés dans la remarque précédente, est de permettre d'interpréter plus facilement les facteurs issus de $k_{IT \times J}$, et en particulier de se rendre compte si un facteur issu de $k_{IT \times J}$ peut être considéré comme un facteur général temporel, ou comme un facteur exprimant les fluctuations individuelles au cours du temps.

4.2 Etude du cas où $I_t = I$, $J_t = J$

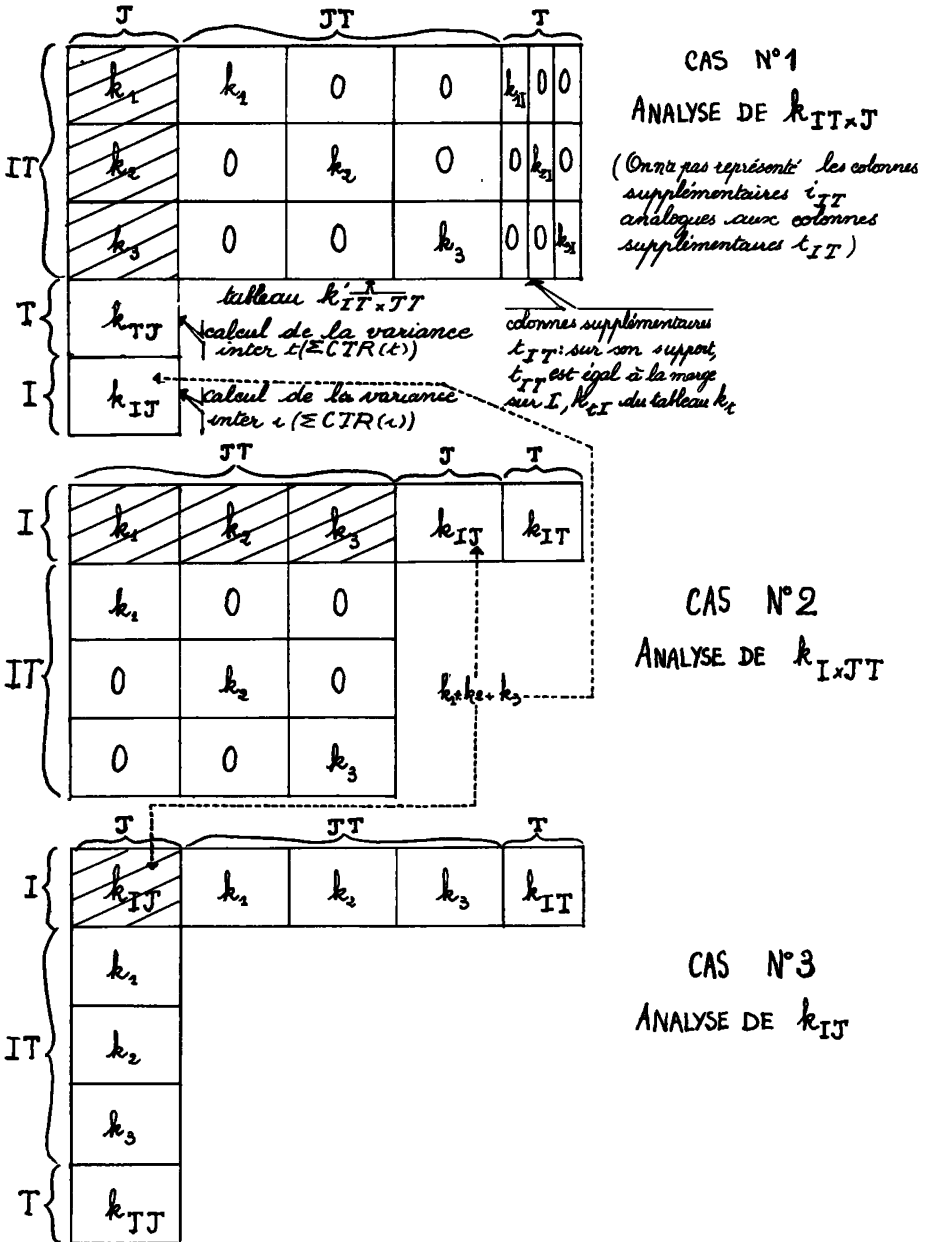
4.2.1 Principe des analyses

Dans le cas où $I_t = I$, $J_t = J$, les tableaux de données se présentent en fait sous la forme d'un tableau ternaire k_{IJT} (où $k(i, j, t) = k_t(i, j)$). On alors $IT = I \times T$, $JT = J \times T$, et l'on désignera par k_{IJ} , k_{IT} , k_{TJ} les tableaux marginaux (d'ordre 2) associés à k_{IJT} :

$$\left. \begin{aligned} k_{IJ}(i, j) &= \Sigma \{k_{IJT}(i, j, t) \mid t \in T\} \\ k_{IT}(i, t) &= \Sigma \{k_{IJT}(i, j, t) \mid j \in J\} \\ k_{TJ}(t, j) &= \Sigma \{k_{IJT}(i, j, t) \mid i \in I\} \end{aligned} \right\} \quad (12)$$

Notons que ce dernier tableau correspond au tableau défini par la relation (8) du § 4.1.

On peut alors appliquer la technique décrite au § 4.1 en analysant le tableau $k_{IT \times J}$ (avec les tableaux k_{IJ} et $k'_{IT \times JT}$ en éléments supplémentaires, ainsi que les colonnes supplémentaires t_{IT} ($t \in T$) de support I_t et égales sur I_t à la marge restreinte associée de $k_{IT \times J}$). On a intérêt à adjoindre aussi en élément supplémentaire de $k_{IT \times J}$ le tableau k_{IJ} : de cette façon, on représente tout élément i de I par le centre de gravité des it ($= (i, t)$) pour t décrivant T : ainsi tout i en tant qu'élément supplémentaire est au centre de gravité des éléments principaux it pour t décrivant T , tandis que tout j en tant qu'élément principal est au centre de gravité des éléments supplémentaires jt pour t décrivant T .



Analyses de $k_{I \times JT}$: on a hachuré les tableaux analysés, les autres étant en éléments supplémentaires. Les cas n° 1 et n° 2 correspondent à la procédure du § 4.1 (figure 2).

Dans le cas n° 2, on n'a pas représenté les lignes supplémentaires correspondant à la procédure du § 3.2.

Figure 3

Notons que quand on analyse le tableau $k_{IT \times J}$, les ensembles I et T jouent des rôles parfaitement symétriques ; on peut donc représenter tout couple (i, j) de la même façon que l'on représente tout couple (t, j) , et l'on peut aussi représenter tout i à l'aide d'une colonne supplémentaire i_{IT} , de la même façon qu'à t est associé t_{IT} .

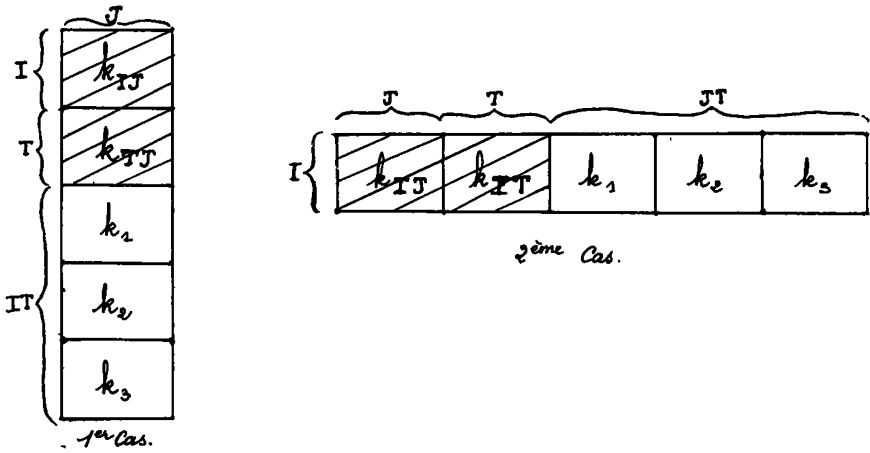
Remarque :

L'adjonction des tableaux supplémentaires k_{IJ} et k_{TJ} à $k_{IT \times J}$ permet de calculer les variances inter I et intertemps sur chaque axe factoriel, ou dans l'espace R_J .

Au lieu d'analyser le tableau $k_{IT \times J}$, on aurait pu de façon symétrique analyser le tableau $k_{I \times JT}$, ce qui revient à permuter les rôles de I et de J.

Une autre possibilité que celles que l'on vient de décrire est d'analyser le tableau k_{IJ} , en lui adjoignant les tableaux $k_{IT \times J}$ et $k_{I \times JT}$ en éléments supplémentaires, ainsi que les tableaux marges k_{IT} et k_{TJ} . L'avantage de cette façon de procéder réside dans les rôles symétriques qu'elle fait jouer aux ensembles I et J : en particulier tout élément principal i (resp. j) est au centre de gravité des éléments supplémentaires it (resp. jt) pour t décrivant T. Notons qu'ici les représentations de t en tant que ligne supplémentaire ou colonne supplémentaire ne sont plus reliées comme au § 4.1 par la relation (7) du § 3.3 : toute ligne supplémentaire t est le centre de gravité des it pour i décrivant I, tandis que toute colonne supplémentaire t est le centre de gravité des jt pour j décrivant J.

Indiquons une dernière procédure pour étudier le tableau k_{IJT} : on analyse le tableau (k_{IJ}, k_{TJ}) superposition de k_{IJ} et k_{TJ} , en lui adjoignant le tableau supplémentaire $k_{IT \times J}$; ainsi tout élément principal i est au centre de gravité des éléments supplémentaires it pour t décrivant T, tandis que tout élément principal t est au centre de gravité des éléments supplémentaires it pour i décrivant I. On peut de façon symétrique analyser le tableau (k_{IJ}, k_{IT}) en lui adjoignant le tableau supplémentaire $k_{I \times JT}$. Un exemple de cette façon de procéder se trouve dans la thèse de A. Bener *op. cit.*.



Etude de k_{IJT} par analyses de tableaux superposition de 2 tableaux de marge d'ordre 2.

On a hachuré les tableaux analysés, les autres étant en éléments supplémentaires.

Figure 4

Remarque

Dans l'exposé précédent, on a fait jouer un rôle particulier à T, en considérant que I et J jouaient des rôles symétriques. En fait les trois ensembles I, J, T, jouent des rôles parfaitement symétriques. On peut donc, pour chacune des analyses préconisées, en déduire deux autres en permutant les ensembles I, J, T. En particulier, dans le cas numéro 3 de la figure 3, on peut au lieu d'analyser k_{IJ} (avec les quatre tableaux $k_{IT \times J}$, k_{TJ} , $k_{I \times JT}$ et k_{IT} en supplémentaires) analyser k_{TJ} ou k_{IT} , avec dans chaque cas les quatre tableaux possibles associés en éléments supplémentaires.

4.2.2 Calculs d'inertie

Plaçons nous dans l'espace R_J pour fixer les idées. On a dans cet espace trois nuages: $N(IT_J)$, $N(I_J)$ et $N(T_J)$, d'où, comme on l'a dit, trois analyses factorielles possibles, suivant le nuage choisi pour déterminer les axes factoriels ($N(IT_J)$ dans le cas 1 de la figure 3, $N(I_J)$ dans le cas 3 de cette même figure et $N(T_J)$) et même quatre, si l'on fait l'analyse factorielle du nuage $N(I_J) \cup N(T_J)$ (cf. figure 4, 1-er cas).

Quel que soit le nuage sur lequel porte l'analyse (les autres étant en supplémentaires) l'espace R_J est muni de la même métrique, à savoir la métrique du χ^2 de centre p_J , p_J étant la loi marginale sur J du tableau $P_{IJT} = k_{IJT}/k$ (avec k total des éléments du tableau k_{IJT}). On peut donc, par des considérations analogues à celles développées au § 3.4, calculer d'une part l'inertie des nuages $N(I_J)$, $N(T_J)$, $N(IT_J)$ dans R_J (inerties indépendantes de l'analyse effectuée, puisqu'on a toujours la même métrique), et d'autre part, l'inertie de ces mêmes nuages sur les axes factoriels issus des différentes analyses réalisées.

Au lieu de se placer dans R_J , on peut se placer dans R_I ou dans R_T , d'où des calculs d'inerties analogues aux précédents. Tous ces calculs que l'on peut résumer dans un tableau, permettent de voir l'importance de chaque interaction binaire (interaction totale, ou sur un axe) et peuvent ainsi aider à mieux interpréter les analyses. En général T correspond au temps, et les interactions $T \times I$ ou $T \times J$ (interactions mesurées par l'inertie du tableau marge correspondant) sont beaucoup plus faibles que l'interaction $I \times J$, tandis que cette dernière interaction (égale à l'inertie de $N(I_J)$ ou $N(J_T)$) est souvent peu différente des inerties des nuages $N(IT_J)$ et $N(JT_I)$. On peut trouver un exemple d'application de ces calculs dans [AGRI. SYRIE], *Cahiers*, Vol VII n° 1, pp 67-91 où I est un ensemble de produits agricoles, J un ensemble de régions, et T un ensemble d'années, $\kappa(i, j, t)$ désignant la production du produit i dans la région j pendant l'année t .

5 Cas des correspondances multiples

On suppose ici qu'on a un tableau disjonctif complet k_{IJ} où I représente suivant l'usage l'ensemble des individus et $J = \cup \{J_q | q \in Q\}$ l'ensemble des modalités de toutes les questions, J_q étant l'ensemble des modalités de la question q . On désignera par B_{JJ} le tableau de Burt associé à k_{IJ} , et par C_{LL} , le sous-tableau de B_{JJ} croisant deux séries de questions K et K' de Q , L (resp. L') désignant l'union disjointe des J_q pour $q \in K$ (resp. $q \in K'$).

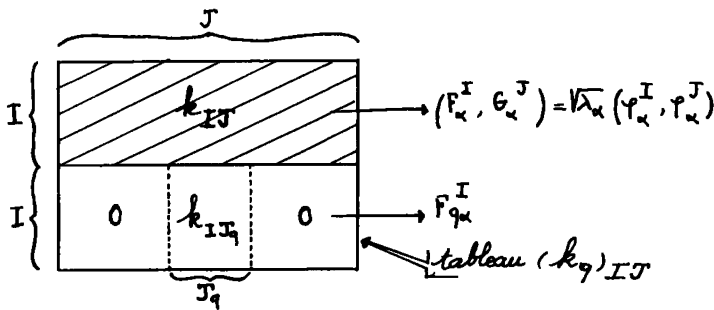
Aux §§ 5.1 et 5.2, on examine l'intérêt des éléments supplémentaires dans le cas des tableaux k_{IJ} et B_{JJ} , tandis qu'au § 5.3 on se place dans le cadre du tableau C_{LL} ; ; enfin au § 5.4, on rappelle quelques résultats sur la représentation de variables quantitatives (non découpées en classes) sur les axes factoriels issus d'une analyse de correspondances multiples.

5.1 Cas du tableau k_{IJ}

Le tableau k_{IJ} rentre (avec des notations légèrement différentes, l'ensemble J correspondant à l'ensemble JT du § 4) dans le cadre (cas b) du début du § 4) des tableaux examinés au § précédent. On pourra donc s'inspirer des idées développées au § 4.1 pour représenter les ensembles J_q , ainsi que les éléments de l'ensemble I caractérisés par chaque question q de Q (on notera dans ce cas I_q pour I).

La représentation des sous-ensembles J_q ($q \in Q$) offre peu d'intérêt puisque chacun de ces sous-ensembles est représenté par l'origine (que l'on adopte les méthodes 1, 2, ou 3 du § 3), résultat classique en analyse des correspondances multiples, puisque le centre de gravité de chaque sous-ensemble J_q ($q \in Q$) de J est confondu avec le centre de gravité de J .

En ce qui concerne la représentation des éléments de I_q , on adjoind en tableau supplémentaire de k_{IJ} le tableau $(k_q)_{IJ}$ (cf. fig. 5)



Le tableau disjonctif complet k_{IJ} avec le tableau $(k_q)_{IJ}$ en élément supplémentaire (chaque ligne du tableau $(k_q)_{IJ}$ ne comporte que des 0, sauf le $q(i)^{ème}$ élément qui vaut 1).

Figure 5

dont tous les blocs $I \times J_q$, sont nuls, sauf le bloc $I \times J_q$ qui est égal au sous-tableau k_{IJq} de k_{IJ} (la superposition des tableaux $(k_q)_{IJ}$ pour $q \in Q$ donne un tableau $k'_{IQ \times J}$ où $IQ = I \times Q$, qui correspond au tableau $k'_{IT \times JT}$ du § 4.1, après interversion des rôles des lignes et des colonnes de ce dernier tableau).

Soit $(\varphi_\alpha^I, \varphi_\alpha^J)$ un couple de facteurs associés, de variance 1, issu de k_{IJ} , λ_α la valeur propre correspondante, $(F_\alpha^I, G_\alpha^J) = \sqrt{\lambda_\alpha} (\varphi_\alpha^I, \varphi_\alpha^J)$ le couple de facteurs de variance λ_α . Il est immédiat de voir que si $F_{q\alpha}^I$ désigne le facteur sur I associé au tableau $(k_q)_{IJ}$, on a :

$$F_{q\alpha}^I(i) = \varphi_\alpha^J \tag{13}$$

j désignant la modalité de la question q prise par l'individu i modalité que l'on notera encore $q(i)$.

Représenter i caractérisé par la question q (on notera iq le point correspondant) revient finalement à placer iq au point $q(i)$ de J_q , si l'on adopte sur chaque axe factoriel la représentation barycentrique $(F_\alpha^I, \varphi_\alpha^J)$ associée au tableau k_{IJ} : dans cette représentation, tout i est au centre de gravité des j de J affectés des masses $k(i, j)$. Comme $k(i, j)$ pour $j \in J_q$ est nul, sauf pour $j = q(i)$, auquel cas il vaut 1, i est au centre de gravité des iq pour $q \in Q$ (iq étant affecté de la masse $1/\text{Card}Q$), résultat classique. Notons que le centre de gravité des iq pour i décrivant I , iq étant affecté de la masse unité, n'est autre que l'origine, i.e. le centre de gravité de I (calculé à partir du tableau k_{IJ}) : ceci résulte de la relation (11) où il faut remplacer j par i , et où le coefficient c_t est égal à $1/\text{Card}Q$; on peut aussi le vérifier directement à partir de (13), en se souvenant que φ_α^J est centré sur chaque J_q .

Tous les résultats précédents sont relativement classiques. L'intérêt du facteur $F_{q\alpha}^I$ qui peut être considéré comme un codage de la variable q (cf. Tenenhaus : L'analyse en composantes principales de variables qualitatives, à paraître dans les Publications du Laboratoire de Statistique de l'Université Paul Sabatier, Toulouse) réside dans le fait qu'il permet d'interpréter la contribution $CR_\alpha(q)$ d'une question q à l'inertie λ_α du facteur α .

On a les résultats suivants (cf. [ANA. BLOCS II], § 4.5.1 *Cahiers*, Vol V n° 4, p. 398, et *addendum*, *Cahiers*, Vol VI n° 1, pp 17-18 (où le facteur $G_{q\alpha}^I$ qui y est défini en mettant (k_q^J) en élément supplémentaire de B_{JJ} est égal à $F_{q\alpha}^I$) :

$$\left. \begin{aligned} CR_\alpha(q) &= \Sigma\{k(j) G_\alpha^2(j) \mid j \in J_q\} / (\text{Card } I \text{ Card } Q) \\ &= \text{Cov}(F_\alpha^I, F_{q\alpha}^I) / \text{Card } Q \\ \sqrt{CR_\alpha(q)} &= \text{Corr}(F_\alpha^I, F_{q\alpha}^I) / \sqrt{\text{Card } Q} \end{aligned} \right\} \quad (14)$$

formules où Cov désigne la covariance, Corr la corrélation, et $k(j)$ le total de la colonne j de k_{IJ} (i.e. le nombre d'individus ayant adopté la modalité j).

Si $CTR_\alpha(q)$ désigne la contribution de q à l'axe α , rapportée à λ_α , on a donc :

$$\left. \begin{aligned} CTR_\alpha(q) &= CR_\alpha(q) / \lambda_\alpha = \text{Cov}(F_\alpha^I, F_{q\alpha}^I) / (\lambda_\alpha \text{ Card } Q) \\ \sqrt{CTR_\alpha(q)} &= \text{Corr}(F_\alpha^I, F_{q\alpha}^I) / \sqrt{\lambda_\alpha \text{ Card } Q} \end{aligned} \right\} \quad (15)$$

L'intérêt de la contribution $CR_\alpha(q)$ est de permettre une représentation non triviale de la question q sur l'axe α (puisque la représentation classique de q comme centre de gravité des j de J_q est l'origine comme on l'a rappelé ci-dessus). Cette représentation est, par analogie avec celle des modalités, basée sur la double décomposition de l'inertie en fonction des α et des q par l'intermédiaire des $CR_\alpha(q)$. De façon précise, si f_q est le poids de la question q (poids égal à la somme des poids des modalités j de J_q), égal ici à $1/\text{Card } Q$, nous proposons de prendre comme abscisse $G_\alpha(q)$ de q sur l'axe α la quantité :

$$G_\alpha(q) = \sqrt{CR_\alpha(q)/f_q} \quad (16)$$

soit, en tenant compte de (14) :

$$G_\alpha(q) = (CR_\alpha(q) \text{ Card } Q)^{1/2} = \text{Corr}(F_\alpha^I, F_{q\alpha}^I) \quad (17)$$

$G_\alpha(q)$ est donc égal à la corrélation entre F_α^I et $F_{q\alpha}^I$.

On a alors comme décomposition de l'inertie totale In (égale ici à $(\text{Card } J - \text{Card } Q)/\text{Card } Q$) de k_{IJ} la formule :

$$In = \sum \{f_q (G_\alpha(q))^2 \mid \alpha \in A, q \in Q\} \quad (18)$$

où A désigne l'ensemble des facteurs non triviaux issus de k_{IJ} .

Cette décomposition est analogue à la décomposition usuelle de l'inertie en fonction des facteurs et des modalités :

$$In = \sum \{f_j (G_\alpha(j))^2 \mid \alpha \in A, j \in J\}$$

f_j (égal ici à $k(j)/(\text{Card } I \text{ Card } Q)$) désignant la masse de j .

Remarques

a) Avec la représentation précédente des questions, si on se place dans l'espace E_k des k premiers facteurs issus de k_{IJ} , toutes les questions sont situées dans le cône positif de E_k .

b) Les f_q étant constants, on peut aussi prendre pour abscisse de q sur l'axe α la quantité $\sqrt{CR_\alpha(q)}$.

c) Les considérations d'inerties interclasse et intraclasse développées au § 3.4 (où il faut intervertir les rôles de I et J , et où Q joue le rôle de C) se simplifient ici, puisque le centre de gravité de chaque J_q étant l'origine, la variance interclasse est nulle, et donc l'inertie de la classe q^* est égale à sa contribution $CR_\alpha(q)$, tandis que l'inertie intraclasse $In(J-Q)$ est égale à l'inertie totale. Ces propriétés restent vraies pour tous les tableaux intervenant en analyse des correspondances multiples, et en particulier dans le cas des tableaux B_{JJ} et C_{LL} .

* sur l'axe α .

5.2 Cas du tableau de Burt B_{JJ}

Les résultats associés au tableau de Burt B_{JJ} se déduisent immédiatement de ceux associés à k_{IJ} , à condition de mettre les tableaux k_{IJ} et $(k_q)_{IJ}$ en éléments supplémentaires de B_{JJ} , les facteurs correspondants sur I étant les mêmes que précédemment, i.e. F_α^I pour k_{IJ} , $F_{q\alpha}^I$ pour $(k_q)_{IJ}$. Par ailleurs, on sait que B_{JJ} et k_{IJ} ont mêmes facteurs φ_α^J de variance 1 sur J , les valeurs propres λ'_α issues de B_{JJ} étant égales aux carrés des valeurs propres λ_α issues de k_{IJ} .

On a donc, en désignant par un prime les caractéristiques associées à B_{JJ} , et avec des notations évidentes :

$$\left. \begin{aligned} G'_\alpha{}^J &= \sqrt{\lambda'_\alpha} \varphi_\alpha^J = \sqrt{\lambda_\alpha} G_\alpha^J \\ CR'_\alpha(q) &= \lambda_\alpha CR_\alpha(q) \\ CTR'_\alpha(q) &= CTR_\alpha(q) \\ G'_\alpha(q) &= \sqrt{\lambda_\alpha} G_\alpha(q) = \sqrt{\lambda_\alpha} \text{Corr}(F_\alpha^I, F_{q\alpha}^I) \end{aligned} \right\} (19)$$

On a intérêt à raisonner sur le tableau de Burt plutôt que sur le tableau disjonctif complet. En effet d'une part les pourcentages d'inertie calculés sur le tableau de Burt ont plus de sens que ceux calculés sur le tableau disjonctif complet, puisqu'on a un tableau de comptage, et que la trace associée reflète les liaisons entre les variables, contrairement au cas du tableau disjonctif complet où la trace ne reflète que la structure en blocs du tableau, et ne dépend que du nombre de questions et du nombre total de modalités.

D'autre part, si une série de questions Q_1 ($Q_1 \subset Q$) ressort sur les axes factoriels, il peut être intéressant de rapporter l'ensemble J tout entier aux axes factoriels déterminés uniquement par l'ensemble JQ_1 des modalités des questions de Q_1 (cf. § 1.5, 2-ème exemple). Si l'on pose $JQ_2 = J - JQ_1$, cela revient à analyser dans le cas du tableau disjonctif complet le sous-tableau $k_{I \times JQ_1}$ avec $k_{I \times JQ_2}$ en supplémentaire, tandis que dans le cas du tableau de Burt, on analysera la bande $B_{J \times JQ_1}$ avec $B_{J \times JQ_2}$ en supplémentaire. Dans le premier cas, si Q_1 se réduit à une question q , on obtient une analyse triviale, puisque toute fonction de moyenne nulle sur J_q est facteur de k_{IJ_q} relatif à la valeur propre 1, et la position de J sur les premiers axes factoriels (qui seront choisis quelconques dans le support du nuage $N(J_q)$) issu de k_{IJ_q} ne présentera pas d'intérêt. Dans le second cas (Q_1 se réduisant toujours à q), l'analyse de la bande B_{JJ_q} est équivalente à celle du tableau $B_{JQ_2 \times J_q}$ croisant $JQ_2 (= J - J_q)$ avec J_q , et les facteurs ainsi déterminés traduisent bien les liaisons de JQ_2 avec J_q , contrairement au premier cas.

Nous allons donc supposer (dans le cas général où Q_1 ne se réduit pas forcément à une question q) qu'on analyse la bande $B_{J \times JQ_1}$ (avec $B_{J \times JQ_2}$ en supplémentaire). Il est alors facile, à partir du listing de trouver l'inertie totale $\text{Trace}(J)$ (ou l'inertie $\text{In}_\alpha(J)$ sur l'axe α issu de $B_{J \times JQ_1}$) du nuage des colonnes de B_{JJ} .

En effet, si l'on considère la partition C suivante de J :

$$J = (\cup \{j\} | j \in JQ_1) \cup (\cup \{J_q | q \in Q_2\})$$

et si l'on désigne par B_{JC} le tableau déduit de B_{JJ} en cumulant toutes les colonnes d'une même classe, compte tenu de ce que chaque bloc $J \times J_q$ a même marge sur J , l'analyse de B_{JC} est équivalente (à un coefficient de proportionnalité près) à l'analyse de $B_{J \times JQ_1}$. On peut donc appliquer (en les adaptant au cas étudié ici) les résultats du § 3.4 et en particulier ceux de la remarque 2 située à la fin de ce §. Nous nous contenterons de donner deux formules, renvoyant pour plus de détails à [BANDES BURT] déjà cité :

$$\begin{aligned} \text{Trace}(J) &= (\text{Card } Q_1 / \text{Card } Q) \text{Trace}(JQ_1) (\Sigma \{ \text{INR}(j) | j \in J \}) \\ &= (\text{Card } Q_1 / \text{Card } Q) \text{Trace}(JQ_1) (1 + \Sigma \{ \text{INR}(j) | j \in JQ_2 \}) \end{aligned}$$

$$\begin{aligned} \text{In}_\alpha(J) &= (\text{Card } Q_1 / \text{Card } Q) \lambda_\alpha (\Sigma \{ \text{CTR}_\alpha(j) | j \in J \}) \\ &= (\text{Card } Q_1 / \text{Card } Q) \lambda_\alpha (1 + \Sigma \{ \text{CTR}_\alpha(j) | j \in JQ_2 \}) \end{aligned}$$

$\text{Trace}(JQ_1)$ désignant l'inertie de JQ_1 dans l'analyse de $B_{J \times JQ_1}$, λ_α la α -ème valeur propre issue de cette analyse, tandis que dans les sommations des $\text{INR}(j)$ ou $\text{CTR}_\alpha(j)$, les j sont considérés comme colonnes de $B_{J \times JQ_1} \cup B_{J \times JQ_2}$ (et non comme lignes de $B_{J \times JQ_1}$).

On peut calculer le pourcentage d'inertie $\tau_\alpha(J) = 100 \text{In}_\alpha(J) / \text{Trace}(J)$, mais il faut bien noter que l'analyse ayant été faite sur la bande $J \times JQ_1$ (et non sur $J \times J$), $\text{In}_\alpha(J)$ et $\tau_\alpha(J)$ ne sont pas en général des fonctions décroissantes de α .

Dans l'enquête sur l'enfant au cycle primaire et dans sa famille (cf. [ENQUETE ECOLE], Vol VII n° 1, pp 45-65) les 89 questions retenues dans l'analyse correspondant à plus de 300 modalités, on a analysé ces questions en effectuant des analyses par bandes ; on a successivement pris pour bandes les différentes parties du questionnaire : loisirs et vacances ; vie professionnelle des parents ; l'enfant à l'école et dans sa famille ; habitation et conditions économiques.

Dans chacune de ces analyses, on a donné, outre les pourcentages d'inertie usuels, les valeurs $\text{In}_\alpha(J)$ et $\tau_\alpha(J)$, ces deux dernières quantités décroissant avec α (au moins pour $\alpha \leq 5$) quand on étudie les bandes relatives aux deux premières parties du questionnaire ; par contre elles sont croissantes quand l'on passe du premier au troisième axe factoriel, si l'on analyse les bandes relatives aux deux dernières parties du questionnaire.

Notons que dans cette étude, certaines modalités de faible poids, risquant de perturber les analyses, ont été regroupées. C'est ainsi que dans l'analyse de la bande relative aux loisirs et aux vacances, on a remplacé toutes les colonnes (principales) associées aux non-réponses par une colonne (qui reste principale), somme de ces colonnes. D'un point de vue théorique, si JQ_1 désigne la bande initiale analysée (ici loisirs-vacances), J_{nr} l'ensemble des modalités de non-réponse ($J_{nr} \subset JQ_1$), et si $J_1 = JQ_1 - J_{nr}$, on est amené à considérer la partition C suivante de J :

$$J = (\cup\{j\} \mid j \in J_1) \cup J_{nr} \cup JQ_2$$

JQ_2 désignant toujours le complémentaire de JQ_1 dans J . Comme précédemment, l'analyse du tableau $J \times C$ obtenu en cumulant toutes les colonnes d'une même classe se réduit si l'on désigne par j_{nr} la colonne somme des colonnes de J_{nr} , et si l'on pose : $J'_1 = J_1 \cup \{j_{nr}\}$, à l'analyse du tableau $J \times J'_1$ composé du bloc $J \times J_1$ auquel on adjoint la colonne j_{nr} . Cette façon de procéder qui constitue un cas particulier du § 3.4 (remarque 2) est développée dans [BANDES BURT] *Cahiers*, Vol VII n° 1, pp 33-43.

5.3 Cas du tableau C_{LL}

Rappelons que C_{LL} croise l'ensemble des modalités des questions q de K , avec l'ensemble des modalités des questions q' de K' . Ce tableau intervient fréquemment en régression quand on désire expliquer un ensemble K' de variables en fonction d'un ensemble K de variables explicatives, toutes les variables ayant été rendues qualitatives par découpage en classes des variables quantitatives. L'adjonction en élément supplémentaire de C_{LL} , du tableau disjonctif complet (sous tableau de k_{IJ}) k_{IL} associé aux variables explicatives, permet alors d'avoir des facteurs sur I à partir desquels on peut pour toute variable q' de K' appliquer les techniques usuelles de régression, régression par boule (si q' était quantitative avant découpage en classes), ou de discrimination (si q' était initialement qualitative) (cf. § 1.4).

Outre k_{IL} , on peut adjoindre en tableau supplémentaire de C_{LL} , d'une part le tableau disjonctif complet k_{IL} , d'autre part pour $q \in K$ (resp. $q' \in K'$) le tableau $(k_q)_{IL}$ (resp. $(k_{q'})_{IL}$) dont tous les blocs $I \times J_q$ sont nuls sauf le bloc $I \times J_q$ (resp. $I \times J_{q'}$) qui est égal au sous tableau k_{IJ_q} (resp. $k_{IJ_{q'}}$) de k_{IJ} (cf. figure 6).

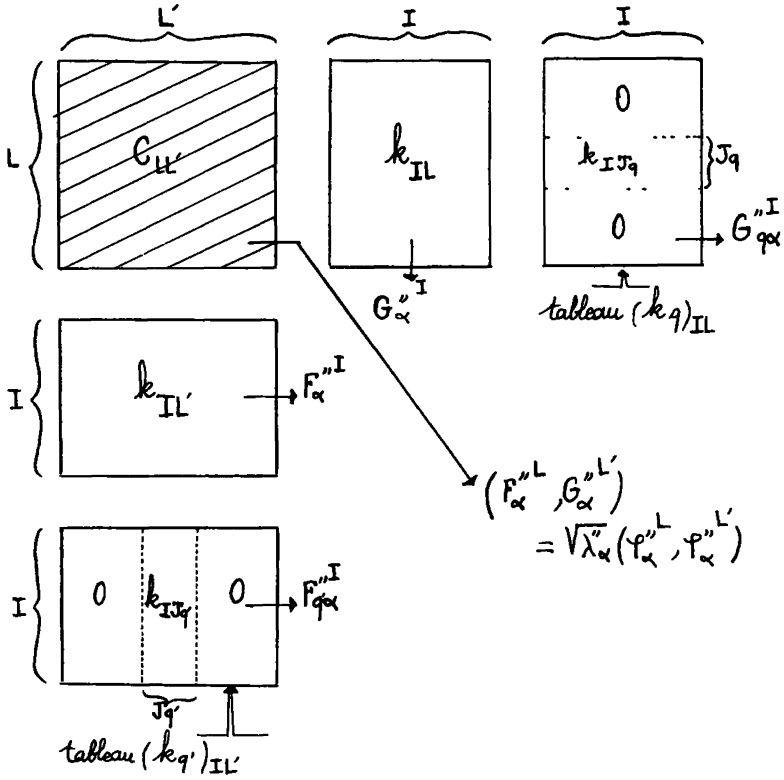


Tableau C_{LL} , et tableaux adjoints en éléments supplémentaires de C_{LL} .

Figure 6

Soit $(\varphi_{\alpha}^{II}, \varphi_{\alpha}^{II'})$ un couple de facteurs associés de variance 1 issu de C_{LL} , λ_{α}^{II} la valeur propre correspondante et $(F_{\alpha}^{II}, G_{\alpha}^{II}) = \sqrt{\lambda_{\alpha}^{II}} (\varphi_{\alpha}^{II}, \varphi_{\alpha}^{II'})$ le couple de facteurs de variance λ_{α}^{II} (nous utilisons l'indice seconde pour les caractéristiques de C_{LL} , de la même façon qu'on a utilisé l'indice prime pour les caractéristiques de B_{JJ}).

Nous désignerons encore par F_{α}^{II} , G_{α}^{II} , $G_{q\alpha}^{II}$, $F_{q\alpha}^{II}$ les valeurs du facteur α pour k_{IL} , $k_{IL'}$, $(k_q)_{IL}$, $(k_q')_{IL'}$ respectivement.

Rappelons que tous ces facteurs sont centrés et que la covarian-
ce entre F_{α}^{II} et G_{α}^{II} est égale à $\sqrt{\lambda_{\alpha}^{II}}$ * tandis qu'on peut établir une
minoration de la corrélation entre F_{α}^{II} et G_{α}^{II} (cf. [ANA. BLOCS I], *Cahiers*,
Vol V n° 2, § 3.2.2, et *addendum* Vol V n° 4, pp 404-406).

Afin de ne pas rallonger inutilement le texte, nous allons dans
la suite considérer I caractérisé par les variables q de K, tous les
résultats donnés se transposant immédiatement si I est caractérisé
par les variables q' de K', en intervertissant les rôles de K et K'
(et donc ceux de L et L').

q(i) désignant toujours la modalité de q prise par i, on a :

$$\left. \begin{aligned} \forall q \in K : G_{q\alpha}^{II}(1) &= \varphi_{\alpha}^{L}(q(i)) \\ G_{\alpha}^{II} &= (1/\text{Card } K) \Sigma \{G_{q\alpha}^{II} | q \in K\} \end{aligned} \right\} (20)$$

relation qui s'écrit encore :

$$G_{\alpha}^{II}(1) = (1/\text{Card } K) \Sigma \{\varphi_{\alpha}^{L}(q(i)) | q \in K\} \quad (21)$$

et qui traduit que tout i (caractérisé par L) est au centre de gra-
vité des iq (i.e. i caractérisé par q) pour q décrivant K, i.e. au
centre de gravité des modalités q(i) de K adoptées par i, si on re-
présente, sur chaque axe non trivial α , L par le facteur de varian-
ce $1 \varphi_{\alpha}^{L}$.

Par ailleurs si $CR_{\alpha}^{II}(q)$ désigne la contribution de la question q
de K à l'axe α (cf. § 4.5.1 de [ANA. BLOCS II] et *addendum*, *Cahiers*,
Vol VI n° 1, pp 17-18), et $CTR_{\alpha}^{II}(q)$ cette contribution rapportée à λ_{α}^{II} ,
on a :

$$\left. \begin{aligned} CR_{\alpha}^{II}(q) &= \text{Cov}(F_{\alpha}^{II}, G_{q\alpha}^{II}) \sqrt{\lambda_{\alpha}^{II}} / \text{Card } K \\ \sqrt{CR_{\alpha}^{II}(q)} &= \text{Corr}(F_{\alpha}^{II}, G_{q\alpha}^{II}) \sqrt{\text{Var}(F_{\alpha}^{II}) / \text{Card } K} \\ CTR_{\alpha}^{II}(q) &= CR_{\alpha}^{II}(q) / \lambda_{\alpha}^{II} = \text{Cov}(F_{\alpha}^{II}, G_{q\alpha}^{II}) / (\sqrt{\lambda_{\alpha}^{II}} \text{Card } K) \end{aligned} \right\} (22)$$

ce qui permet d'interpréter ces contributions.

Comme dans le cas de k_{IJ} et B_{JJ} , pour obtenir une représentation
non triviale de q sur l'axe α , on peut considérer la quantité $G_{\alpha}^{II}(q)$
 $= \sqrt{CR_{\alpha}^{II}(q) / f_{\alpha}^{II}(q)}$ où $f_{\alpha}^{II}(q)$ est la masse de q (égale ici à $1/\text{Card } K$, puisque
 $q \in K$).

On a alors d'après (22) :

$$G_{\alpha}^{II}(q) = \sqrt{CR_{\alpha}^{II}(q) \text{Card } K} = \text{Corr}(F_{\alpha}^{II}, G_{q\alpha}^{II}) \sqrt{\text{Var } F_{\alpha}^{II}} \quad (23)$$

* Si on considère deux facteurs α et β différents, la covariance
entre F_{α}^{II} et G_{β}^{II} est nulle.

Remarques

1) Tous les résultats (et en particulier les formules (20) à (23)) données dans ce § restent valables si $K = K' = Q$ (et donc $L = L' = J$) et quel cas C_{LL} , est égal au tableau de Burt B_{JJ} . En particulier les formules (22) et (23) se simplifient, compte tenu de ce que $\lambda''_{\alpha} = \lambda'_{\alpha}$, λ_{α}^2 , $F''_{\alpha} = F'_{\alpha}$, $G''_{\alpha} = G'_{\alpha}$, $\text{Var } F'_{\alpha} = \lambda_{\alpha}$, et permettent de retrouver les formules (19) (où il faut remplacer $CR_{\alpha}(q)$, $CTR_{\alpha}(q)$ par leur valeurs fournies par (14) et (15)).

2) Si l'on avait appliqué les considérations du § 4.1 au tableau C_{LL} , en le considérant par exemple comme une suite de tableaux $\{C_{J_q L}, | q \in K\}$, on aurait été amené à mettre en colonnes supplémentaires de C_{LL} , les tableaux $(C_q)_{LL}$, pour q décrivant K , tableaux dont tous les blocs $J_{q''} \times L'$ ($q'' \in K$) sont nuls, sauf le bloc $J_q \times L'$ qui est égal à $C_{J_q L}$. Cette procédure n'apporte *a priori* ici rien d'intéressant.

5.4 Représentation d'une variable quantitative sur les axes factoriels issus d'une analyse de correspondances multiples.

Nous reprenons ici quelques résultats donnés au § 4.6 de [ANA. BLOCS II] (cf. *Cahiers*, Vol V, n° 4, pp 401-403) auquel nous renvoyons le lecteur pour plus de détails.

Soit x une variable quantitative mesurée sur l'ensemble I des individus, x_I l'ensemble des valeurs prises par x sur I . Nous supposons cette variable centrée ($\Sigma\{x_i | i \in I\} = 0$) et posons $x_I^+ = (1 + x_I)/2$. Bien que x_i^+ puisse être négatif, le profil (sur I) de x_I^+ est bien défini, puisque la somme des x_i^+ est égale à $\text{Card } I/2$. On pourra donc adjoindre x_I^+ en élément supplémentaire du tableau disjonctif complet k_{IJ} .

Si on désigne par \bar{x}^j la moyenne de x pour les individus ayant adopté la modalité j ($\bar{x}^j = \Sigma\{k(i, j) x_i | i \in I\}/k(j)$) et si $\varphi_{\alpha}^J = \{\varphi_{\alpha}^{Jq} | q \in Q\}$ désigne toujours le α -ème facteur de variance issu de k_{IJ} et λ_{α} , la valeur propre associée, l'abscisse $G_{\alpha}(x_I^+)$ de x_I^+ sur l'axe α s'écrit (cf. [ANA. BLOCS II]) :

$$G_{\alpha}(x_I^+) = (1/\sqrt{\lambda_{\alpha}}) \Sigma\{\text{Cov}(\varphi_{\alpha}^{Jq}, \bar{x}^{Jq}) | q \in Q\} / \text{Card } Q \quad (24)$$

En ce qui concerne le tableau de Burt B_{JJ} , on adjoindra en élément supplémentaire de ce tableau le vecteur y_J de composantes $y_j = k(j)(1 + \bar{x}^j)/2$. La coordonnée $G'_{\alpha}(y_J)$ de y sur l'axe α est alors égale à $\sqrt{\lambda_{\alpha}} G_{\alpha}(x_I^+)$.

Enfin, dans le cas du tableau C_{LL} , on pourra adjoindre en élément supplémentaire de ce tableau soit le vecteur $y_L = \{y_j | j \in L\}$, soit le vecteur $y'_L = \{y_j | j \in L'\}$, (ou les deux). Si l'on considère par exemple y_L , la coordonnée $G_\alpha^n(y_L)$ de cet élément sur le α -ème axe factoriel issu de C_{LL} , s'écrit :

$$G_\alpha^n(y_L) = \Sigma \{ \text{Cov}(\varphi_\alpha^{Jq}, \bar{x}^{Jq}) | q \in K \} / \text{Card } K \quad (25)$$

On a bien sûr une formule analogue pour y'_L .

Remarque

$G_\alpha^n(y_L)$ peut aussi s'interpréter comme la covariance entre la variable x , et le facteur sur I obtenu en mettant le sous tableau k_{II} de k_{IJ} en élément supplémentaire de C_{LL} , facteur noté G_α^I au § 5.3.

6 Utilisation des éléments supplémentaires pour comparer deux analyses différentes, et calculer les corrélations entre facteurs issus de ces deux analyses (cf. [CORREL. FAC.] Prat. de l'an. des données Vol 2, § 3 n° 11)

Soit k_{IJ} et k'_{IK} deux tableaux de données, p_{IJ} et p'_{IK} les tableaux de correspondance associés (i.e. tableaux proportionnels aux tableaux initiaux et dont la somme des éléments vaut 1). Nous supposons que p_{IJ} et p'_{IK} ont même loi marginale sur I : $p_I = p'_I$, et nous désignerons, suivant l'usage par p_J et p'_K les lois marginales sur J et sur K respectivement associées aux tableaux p_{IJ} et p'_{IK} . I peut être par exemple un ensemble d'individus, et J et K deux ensembles de modalités de réponses à un questionnaire (par exemple J correspond à des variables socio-économiques, et K à des variables d'opinion), ensembles que l'on désire comparer.

Soit (F_α^I, G_α^J) (resp. F_β^I, G_β^K) un couple de facteurs associés non triviaux issu de k_{IJ} (resp. k'_{IK}) de variance la valeur propre λ_α (resp. λ'_β) correspondante. Nous désignerons par G_α^K l'extension du facteur G_α^J à l'ensemble K, extension obtenue en adjoignant au tableau k_{IJ} le tableau k'_{IK} en élément supplémentaire.

Pour comparer les facteurs issus de k_{IJ} et k'_{IK} , il est d'usage de calculer la corrélation $r_{\alpha\beta}$ entre les facteurs F_α^I et F_β^I , I étant muni de la mesure p_I . On peut montrer (cf. Prat. de l'an. des données, Vol 2, § 3 n° 11) que cette corrélation s'exprime uniquement en fonction de $G_\alpha^K, G_\beta^K, \lambda'_\beta$ et p'_K . De façon précise on a :

$$\left. \begin{aligned} r_{\alpha\beta} = \text{Corr}(F_\alpha^I, F_\beta^I) &= \Sigma \{ p'_k G_\alpha(k) G_\beta(k) | k \in K \} / \lambda'_\beta \\ &= \text{Cov}(G_\alpha^K, G_\beta^K) / \lambda'_\beta \end{aligned} \right\} (26)$$

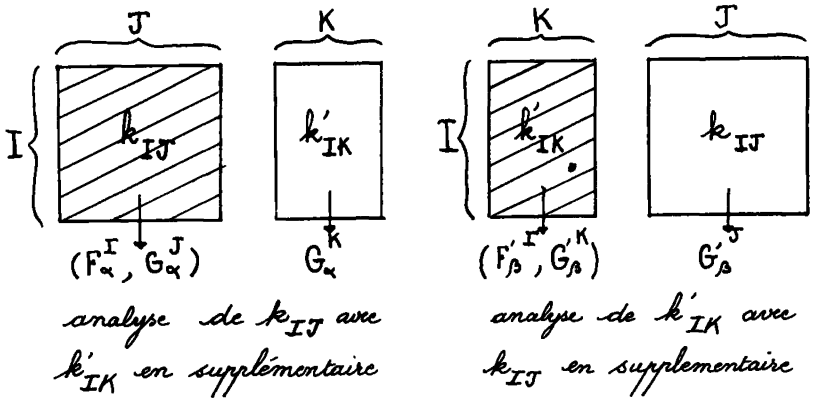


Figure 7.

Dans le cas où le cardinal de I est élevé, et le cardinal de K relativement faible, on a ainsi grâce à la formule précédente un moyen rapide de calculer, sans nouveau passage sur l'ordinateur, la corrélation entre F_{α}^I et F_{β}^I connaissant les résultats des analyses de k_{IJ} (avec k'_{IK} en supplémentaire) et de k'_{IK} .

Remarques

1) Soit G_{β}^J le facteur obtenu en adjoignant en élément supplémentaire de k'_{IK} le tableau k_{IJ} ; $r_{\alpha\beta}$ est aussi égal à $Cov(G_{\alpha}^J, G_{\beta}^J) / \lambda_{\alpha}$, J étant muni de la mesure p_J . Dans le cas où Card J est inférieur à Card K, on a intérêt à utiliser cette dernière formule, au lieu de (26) pour calculer $r_{\alpha\beta}$.

2) Si p'_I est différent de p_I , et si l'on calcule la corrélation entre F_{α}^I et F_{β}^I , I étant muni de la mesure p'_I , on a :

$$r_{\alpha\beta} = (\lambda_{\alpha}^{1/2} / (s_{\alpha} \lambda'_{\beta})) Cov(G_{\alpha}^K, G_{\beta}^K)$$

s_{α}^2 désignant la variance (pour p'_I) de F_{α}^I .

3) Les calculs de corrélation précédents sont très utiles pour comparer les facteurs issus de diverses analyses en bandes d'un tableau de Burt $J \times J$ (cf. § 5.2 *in fine*). Supposons qu'on analyse les bandes $J \times JQ_1$ et $J \times JQ'_1$: les marges sur le premier ensemble J sont identiques dans les deux analyses. Appliquant les résultats précédents où il faut remplacer I, J, K par J, JQ_1 , JQ'_1 respectivement, on obtient avec des notations évidentes :

$$\left. \begin{aligned} r_{\alpha\beta} &= Corr(F_{\alpha}^J, F_{\beta}^J) = Cov(G_{\alpha}^{JQ'_1}, G_{\beta}^{JQ'_1}) / \lambda'_{\beta} \\ &= \{ \sum Cov(G_{\alpha}^{Jq}, G_{\beta}^{Jq}) \mid q \in Q'_1 \} / (Card Q'_1 \lambda'_{\beta}) \end{aligned} \right\} (27)$$

Remplaçant dans (27) Q'_1 par Q_1 et λ'_β par λ_α , on obtient d'après la remarque 1), une autre expression pour $r_{\alpha\beta}$.

Notons qu'au lieu de calculer la corrélation entre F_α^J et F_β^J , facteurs associés aux lignes des bandes analysées, on aurait pu aussi calculer la corrélation entre les facteurs G_α^J et G_β^J associés à l'ensemble J des colonnes (principales + supplémentaires) dans les deux analyses.

Revenons maintenant au cas général considéré ici (cf. figure 7). Pour voir si le facteur F_α^I peut être expliqué par l'ensemble B des facteurs non triviaux issus de k'_{IK} , on considère la quantité :

$$C_\alpha = \Sigma \{r_{\alpha\beta}^2 | \beta \in B\}$$

quantité qui n'est autre, puisque les F_β^I sont non corrélés, que le carré du coefficient de corrélation multiple de F_α^I par rapport aux $\{F_\beta^I | \beta \in B\}$, i.e. le carré de la corrélation (ou du cosinus) de F_α^I avec sa projection sur le sous espace E' de R^I engendré par les $\{F_\beta^I | \beta \in B\}$.

Si C_α est égal à 1, F_α^I peut s'exprimer en combinaison linéaire des F_β^I , ou encore en combinaison linéaire des $p'_k{}^I$ ($k \in K$) (avec $p'_k{}^i = p'_{ik}/p'_i = p'_{ik}/p_i$) puisque les F_β^I sont combinaisons linéaires des $p'_k{}^I$. Si C_α est nul, F_α^I est orthogonal au sous espace E' , et on peut montrer (cf. [CORREL. FAC.] *op. cit.*) que pour tout k de K , $G_\alpha(k) = 0$.

Si pour tout α de A , (où A désigne l'ensemble des facteurs non triviaux issus de k_{IJ}), C_α vaut 1, alors compte tenu de ce que $p_I = p'_I$, le sous espace vectoriel des profils p_I^j des colonnes de p_{IJ} (ou de k_{IJ}) est contenu dans le sous espace vectoriel engendré par les profils p_I^k des colonnes de p'_{IK} (ou de k'_{IK}). Dans ce cas toute colonne k_{Ij} de k_{IJ} s'exprime en combinaison linéaire des colonnes k'_{Ik} de k'_{IK} .

De même qu'on a introduit C_α , on pourra définir la quantité :

$$C'_\beta = \Sigma \{r_{\alpha\beta}^2 | \alpha \in A\}$$

qui est le carré de la corrélation multiple de F_β^I en fonction des $\{F_\alpha^I | \alpha \in A\}$ et voir dans quelle mesure on peut expliquer F_β^I en fonction des F_α^I .

Les considérations précédentes sont en particulier fort utiles quand le tableau k'_{IK} est obtenu par sommation de certaines colonnes

du tableau k_{IJ} . On se trouve en particulier dans cette situation déjà examinée de façon générale au § 3.4 quand on analyse des tableaux ternaires (cf. § 4, et remarques nos 4 et 5 du § 4.1), ou des tableaux de régression (§ 1.4.2, 1.4.3) ou des tableaux intervenant en correspondances multiples (§ 5).

Supposons que : $J = \cup \{J_k \mid k \in K\}$

et que : $\forall k \in K : k'(i,k) = \Sigma \{k(i,j) \mid j \in J_k\}$ (28)

(auquel cas, k'_{IJ} et k'_{IK} ont bien même loi marginale sur I).

Le tableau k'_{IK} peut correspondre par exemple à un effet moyen ou à un effet général (qui peut être un effet temporel si K représente le temps), et il peut être intéressant de voir si certains facteurs issus de k'_{IJ} peuvent être considérés comme des facteurs traduisant des propriétés générales, ou si ce sont des facteurs exprimant les différences à l'intérieur de chaque J_k (ou les fluctuations temporelles individuelles si K représente le temps et si J_k est indépendant de k). Si un facteur F_α^I issu de k'_{IJ} est tel que $C_\alpha = 1$, c'est un facteur général. Dans tous les cas, on pourra considérer que C_α est la contribution relative des variables générales au facteur α .

On peut en outre noter que du fait de (28), $\forall \beta \in B$, C_β^I vaut 1, puisque le support des profils des colonnes de k'_{IK} est contenu dans le support des profils des colonnes de k'_{IJ} .

Un exemple d'application rentrant dans le cadre précédent est donné dans [*LIENS ENFANTS*] (cf. *Cahiers*, Vol I n° 2, pp 197-216) où k'_{IJ} est un tableau disjonctif complet tel que $J = I \times K$, et où K correspond à un ensemble de 6 attitudes.

De façon précise, I est un ensemble d'élèves ayant émis des opinions sur leurs camarades ($\forall (i,i') \in I \times I$: soit i accepte i' , soit i rejette i' , soit i est indifférent à i'), et $K = K^+ \cup K^-$ avec :

$K^+ = \{\text{accepte, rejette, est indifférent}\}$

$K^- = \{\text{est accepté, est rejeté, est considéré comme indifférent}\}$

A tout élève i de I correspondent deux variables i^+ et i^- à 3 modalités : i^+ caractérisé par K^+ correspond à l'élève i actif (i.e. jugeant ses camarades), tandis que i^- caractérisé par K^- correspond à l'élève i passif (i.e. jugé par ses camarades).

Le tableau k'_{IK} donne alors pour chaque élève i le nombre total de ses acceptations, de ses rejets et de ses indifférences, ainsi que le nombre total de fois qu'il est accepté, rejeté, ou considéré comme indifférent par un élève de sa classe.

L'analyse factorielle des tableaux k'_{IJ} et k'_{IK} , complétée par les calculs de corrélation précédents, permet de voir que les contributions des qualités générales aux facteurs issus de k'_{IJ} se font essentiellement sur les trois premiers facteurs issus de cette analyse, bien que ces facteurs ne soient pas exclusivement des facteurs généraux. On pourra pour plus de détails se reporter à [*LIENS ENFANTS*].