

P. CAZES

**L'analyse de certains tableaux rectangulaires
décomposés en blocs : généralisation des
propriétés rencontrés dans l'étude des
correspondances multiples**

Les cahiers de l'analyse des données, tome 6, n° 1 (1981),
p. 9-18

<http://www.numdam.org/item?id=CAD_1981__6_1_9_0>

© Les cahiers de l'analyse des données, Dunod, 1981, 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/>

L'ANALYSE DE CERTAINS TABLEAUX RECTANGULAIRES

DÉCOMPOSÉS EN BLOCS :

GÉNÉRALISATION DES PROPRIÉTÉS RENCONTRÉS DANS L'ÉTUDE DES CORRESPONDANCES MULTIPLES *

III. - CODAGE SIMULTANÉ

DE VARIABLES QUALITATIVES ET QUANTITATIVES

[ANA. BLOCS III] (à suivre)

par P. Cazes (¹)

5 Codage simultané de variables qualitatives et quantitatives - Relations entre analyse des correspondances et analyse en composantes principales

5.1 Introduction : On généralise ici l'idée de B. Escofier (cf. [qualitatives et quantitatives], *Cahiers*, Vol IV n° 2, pp 137-146) qui traite du mélange de variables qualitatives codées de façon disjonctive complète, et de variables quantitatives x_q codées (i.e. remplacées) par les deux variables $(1 + x_q)/2$ et $(1 - x_q)/2$, x_q ayant été préalablement centrée.

Après avoir formulé ce codage et donné le tableau de Burt associé (§5.2), on compare les méthodes déduites de ce codage et de l'analyse des correspondances multiples (analyse du tableau de Burt ou de sous-tableaux de ce tableau) avec des méthodes plus classiques, d'abord quand on n'a que des variables quantitatives (§5.3) puis quand on étudie les liaisons entre un groupe de variables qualitatives et un groupe de variables quantitatives (§5.4).

Dans le premier cas, si l'on désire décrire les liaisons entre les variables quantitatives, on obtient l'analyse en composantes principales usuelle (§ 5.3.1) ; si par contre on recherche les liaisons entre deux groupes de variables quantitatives, on obtient une technique qui diffère légèrement de l'analyse canonique usuelle (§ 5.3.2).

Dans le deuxième cas (§ 5.4) on compare en particulier la méthode obtenue lorsque l'on n'a qu'une seule variable qualitative avec l'analyse factorielle discriminante.

On conclut enfin sur le type de méthode présentée ici (§ 5.5).

5.2 Traitement simultané de variables qualitatives et quantitatives :

On suppose ici que sur un échantillon I de n individus, on a mesuré un ensemble Q de variables, Q étant divisé en deux groupes K et K', les variables du groupe K étant qualitatives et celles du groupe K' quantitatives. On suppose de plus que les variables quantitatives sont centrées :

* suite des articles parus sous le même titre dans les Cahiers: [ANA BLOCS I], (§§ 1, 2 et 3, formules 1 à 55) Vol V n° 2, pp 145-161 ; [ANA. BLOCS II], (§ 4 ; formules 56 à 66) Vol V n° 4 pp 387-403 ; articles auxquels nous faisons ici de fréquentes références.

(1) Maître-Assistant, Laboratoire de statistique. Université Pierre et Marie Curie.

$$\forall x_q \in K' : \sum \{x_q(i) | i \in I\} = 0$$

$x_q(i)$ étant la valeur pour l'individu i de la variable quantitative x_q .

A toute variable x_q de K' , on associe les deux variables $x_q^+ = (1 + x_q)/2$ et $x_q^- = (1 - x_q)/2$, et l'on pose :

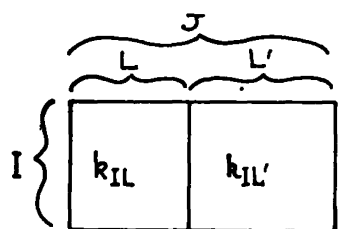
$$J_q = \{x_q^+, x_q^-\}, L' = \cup \{J_q | q \in K'\}$$

(les x_q^+ , x_q^- pouvant, cf. *infra*, prendre des valeurs négatives).

A l'ensemble K' des variables quantitatives est donc associé un tableau $k_{IL'}$ à $2C_{K'}$ colonnes.

A l'ensemble K des variables qualitatives est associé de façon classique le tableau disjonctif complet k_{IL} où $L = \cup \{J_q | q \in K\}$, J_q étant l'ensemble des modalités de la variable q de K .

Posons $J = L \cup L' = \cup \{J_q | q \in K \cup K'\}$ et considérons le tableau $k_{IJ} = (k_{IL}; k_{IL'})$ obtenu par juxtaposition de k_{IL} et $k_{IL'}$. Ce tableau (on l'a dit) peut comporter des éléments négatifs; par contre ses



Juxtaposition de k_{IL} et de $k_{IL'}$

marges sur I et J ne comportent (à supposer qu'on ait au préalable retiré les colonnes nulles, correspondant aux modalités des variables qualitatives qui ne sont prises par aucun individu), comme on le vérifie aisément, que des éléments strictement positifs (on a en particulier : $\forall i \in I, k(i) = C_Q$; $\forall j \in L' : k(j) = C_I/2$).

On peut donc définir l'analyse des correspondances du tableau k_{IJ} , comme on le fait pour un tableau de nombres positifs, et comme le réalisent les programmes d'analyse des correspondances, à partir de la double analyse du nuage des profils des lignes et des colonnes de ce tableau, les poids et les métriques étant définis classiquement à partir des marges. C'est cette analyse qu'on effectuera pour étudier conjointement l'ensemble Q de toutes les variables, suivant la voie qui a été proposée par B. Escofier (cf. *Cahiers*, Vol IV n° 2, pp 137-146) dans le cas où les variables quantitatives sont réduites.

L'analyse des correspondances du tableau k_{IJ} ne peut plus par contre s'interpréter comme l'analyse canonique dans R^{IJ} des sous-espaces R^I et R^J , (sous-espaces des fonctions ne dépendant respectivement que de I et de J). En effet la forme quadratique (dont la valeur pour une fonction ψ^{IJ} s'écrit : $\sum \{\psi^{ij} k(i,j)/k | (i,j) \in I \times J\}$, k désignant toujours le total du tableau k_{IJ} égal ici à $C_I C_Q$) qui permet de définir la métrique

dans le cas d'un tableau de nombres positifs, n'est plus définie si k_{IJ} comporte des éléments négatifs ; il en résulte que l'on peut avoir des valeurs propres supérieures à 1 (de même qu'en géométrie hyperbolique avec une forme quadratique de signature mixte, on a des cosinus hyperboliques...).

Donnons quelques caractéristiques associées à l'analyse des correspondances de k_{IJ} :

On peut d'abord noter que le tableau k_{IJ} vérifie la relation (4) avec $b_q = 1/C_Q$:

$$\forall i \in I : \sum \{k(i, j) \mid j \in J_q\} = b_q, k(i) \quad (4)$$

le centre de gravité de chaque sous-ensemble J_q de J est donc confondu avec le centre de gravité total. Il en résulte en particulier que pour toute variable quantitative x_q , les points x_q^+ et x_q^- sont sur tout axe factoriel symétriques par rapport à l'origine puisqu'ils ont même poids; cette propriété subsiste dans l'analyse du tableau de Burt B_{JJ} associé à k_{IJ} , ou de sous-tableaux de B_{JJ} . Il en résulte en particulier que si l'on n'a que des variables quantitatives ($K = \emptyset$, $L = \emptyset$, $K' = Q$, $L' = J$, $C_J = 2C_Q$) on peut se contenter de ne représenter que les x_q^+ , et de diagonaliser une matrice de dimension deux fois plus petite que la matrice que l'on diagonalise classiquement (i.e. on diagonalise une matrice d'ordre $C_J/2 = C_Q$ au lieu de C_J , dans le cas de k_{IJ} ou B_{JJ}).

Par ailleurs le produit scalaire entre les profils de x_q^ε et $x_q^{\varepsilon'}$ (avec $\varepsilon = \pm 1$, $\varepsilon' = \pm 1$, et où on a noté x_q^ε pour x_q^+ si $\varepsilon = 1$ et x_q^- si $\varepsilon = -1$, et de même pour $x_q^{\varepsilon'}$) ramenés au centre de gravité est égal à la covariance entre les variables εx_q et $\varepsilon' x_q$, tandis que les contributions de x_q^+ et x_q^- à l'inertie totale $CR(J)$ de k_{IJ} sont égales, leur valeur commune étant au facteur $(1/(2C_Q))$ près la variance de x_q notée $Var(x_q)$. La contribution de $J_q = \{x_q^+, x_q^-\}$ à $CR(J)$ est donc égale à $Var(x_q)/C_Q$.

Les contributions $CR(j)$, $CR(J_q)$, d'une modalité j de L , et d'un sous-ensemble J_q de L sont toujours données par les deux premières formules (57):

$$CR(j) = (1 - p_j)/C_Q$$

$$CR(J_q) = (\text{Card } J_q - 1)/C_Q$$

et l'inertie totale s'écrit :

$$CR(J) = (C_L - C_K + \sum \{Var(x_q) \mid q \in K'\})/C_Q \quad (67)$$

On peut ensuite remarquer que le sous-tableau k_{IJ_q} ($q \in K'$) ne comporte que deux colonnes ; il lui est donc associé une seule valeur propre non triviale qui est égale à son inertie, inertie qui est égale à la variance de x_q , et qui vaut donc 1 si x_q est réduite. Si toutes les variables quantitatives sont réduites, on déduit de (67) que $CR(J) = (C_J - C_Q)/C_Q$; on obtient ainsi la même formule que dans le cas d'un questionnaire.

Le tableau de Burt B_{JJ} , dont l'analyse est rappelons-le équivalente à celle de k_{IJ} , se décompose en quatre blocs: $B_{LL}, B_{LL'}, B_{L'L}, B_{L'L'}$, B_{LL} étant (au facteur C_Q/C_K près avec la définition correspondant à la formule (31)) le tableau de Burt usuel associé au sous-tableau k_{IL} ; les termes généraux des tableaux $B_{LL'}, B_{L'L}$ et $B_{L'L'}$, s'écrivent respectivement :

$$\forall j \in L, \forall j' = x_q^E \in L' : B(j, j') = B(j', j) = k(j) (1 + \bar{x}_q^j) / (2C_Q) \quad (68)$$

$$\forall (q, q') \in K'^2, \forall j = x_q^E, j' = x_{q'}^E : B(j, j') = B(j', j) = C_I (1 + \varepsilon \varepsilon' \text{Cov}(x_q, x_{q'})) / (4C_Q) \quad (69)$$

$\bar{x}_q^j$ désignant la moyenne de x_q pour les individus ayant pris la modalité j de L , individus en nombre égal à $k(j)$, tandis que $\text{Cov}(x_q, x_{q'})$ désigne la covariance entre x_q et $x_{q'}$.

Le sous-tableau $B_{JqJq'}$ ($q, q' \in Q$) ayant pour total C_I/C_Q , le sous-tableau de total 1, $p_{JqJq'}$, qui lui est proportionnel est donc égal à $(C_Q/C_I) B_{JqJq'}$. L'inertie du tableau $B_{JqJq'}$ ($q, q' \in K'$) étant égale à $\text{Cov}^2(x_q, x_{q'})$, l'inertie totale de B_{JJ} peut s'écrire d'après (13) (où $a_q = b_q = 1/C_Q$, $\text{IN}(q, q') = \phi_{qq}^2$, si $q, q' \in K$, avec $\phi_{qq}^2 = \text{Card } J_q - 1$, si $q \in K$), et compte-tenu de ce que l'inertie du bloc $q \times q'$ ($q \in K, q' \in K'$) est égale à la variance interclasse $\text{Var}_q(x_{q'})$ de $x_{q'}$, associée à la partition induite par q :

$$\begin{aligned} \text{CR}(J) = & (C_L - C_K + \sum \{ \phi_{qq}^2 \mid q \in K, q' \in K, q' \neq q \} + \sum \{ \text{Var}^2(x_q) \mid q \in K' \} + \sum \{ \text{Cov}^2(x_q, x_{q'}) \\ & \mid q \in K', q' \in K', q \neq q' \} + 2 \sum \{ \text{Var}_q(x_{q'}) \mid (q, q') \in K \times K' \}) / (C_Q)^2 \quad (70) \end{aligned}$$

Si les variables quantitatives x_q ($q \in K'$) sont de variance 1, les blocs $B_{JqJq'}$ ($q \in K'$) sont diagonaux (*) (x_q^+ et x_q^- sont non corrélés) comme le sont les blocs B_{JqJq} ($q \in K$). B_{JJ} a donc la structure d'un tableau de Burt usuel où tous les blocs diagonaux sont diagonaux. Dans l'expression (70) de l'inertie, on peut remplacer $C_L - C_K + \sum \{ \text{Var}^2(x_q) \mid q \in K' \}$ par $C_J - C_Q$, et si l'on veut annuler l'influence des termes diagonaux qui apportent une information triviale, on pourra comme dans le cas d'un questionnaire, annuler ces termes et considérer le tableau de Burt modifié (cf. § 4.4 in [ANA. BLOCS II] Cahiers Vol V n° 4).

Remarques : 1) Tous les tableaux k_{IJq} (resp. $B_{JqJq'}$) ont même poids, la somme de leurs éléments valant C_I (resp. C_I/C_Q) ; on peut donner une pondération à J_q , soit pour égaliser les contributions de chaque sous-tableau, soit pour tenir compte de la variance originale de x_q ($q \in K'$) si l'on a réduit les variables pour avoir des blocs $J_q \times J_q$ diagonaux, ou pour que tous les résultats des §§ 2.4 et 3.2.2 restent valables (cf. remarque suivante).

2) Les propriétés extrémales données au § 3.2.1 (cf. [ANA. BLOCS I])

(*) De façon précise ces blocs sont égaux, au coefficient $C_I/(2C_Q)$ près, à la matrice unité d'ordre 2.

Cahiers Vol V n° 2) pour les tableaux construits à partir du tableau k_{IJ} considéré ici, et croisant deux séries Q_1 et Q_2 de variables restent valables. Par contre les limitations données aux §§ 2.4, 3.2.2 (Th 1, 2, 3), et dans l'addendum (Th. 3', cf. *Cahiers* Vol V n° 4, pp 404-406) ne sont plus valables puisque les sous-tableaux k_{IJq} ($q \in K'$), $p_{JqJq'}$ ($q \in K \cup K'$, $q' \in K'$) pouvant comporter des termes négatifs, on ne peut plus appliquer l'inégalité de Schwarz qui est à la base de ces résultats. Néanmoins, on peut montrer que si toutes les variables quantitatives sont de variance inférieure ou égale à 1, l'inégalité de Schwarz reste valable. De façon précise, si l'on considère par exemple le sous-tableau p_{IJq} (tableau de total 1 et égal à k_{IJq}/C_I) avec $q \in K'$, on a pour toute fonction φ^I centrée, et pour toute fonction φ^{Jq} :

$$\begin{aligned} |\Sigma p_{IJq}(i,j) \varphi^i \varphi^j| & \leq \|\varphi^I\|_{p_I} \|\varphi^{Jq}\|_{p_{Jq}} (\text{Var}(x_q))^{1/2} \\ & \leq \|\varphi^I\|_{p_I} \|\varphi^{Jq}\|_{p_{Jq}} \quad \text{si } \text{Var}(x_q) \leq 1 \end{aligned}$$

Si donc toutes les variables quantitatives sont de variance inférieure ou égale à 1, les calculs effectués aux §§ 2.4, 3.2.2 et dans l'addendum restent valables, et donc les théorèmes 1, 2 (où il faut remplacer la covariance entre deux fonctions par la valeur pour ces deux fonctions de la forme bilinéaire associée) 3 et 3' le sont également. Dans ce cas, toutes les valeurs propres issues de k_{IJ} , B_{JJ} ou de sous-tableaux quelconques de k_{IJ} et B_{JJ} (dont en particulier k_{IJq} pour $q \in K'$, et $B_{JqJq'}$ pour $q \in K \cup K'$, $q' \in K'$) sont inférieures ou égales à 1.

5.3 Cas où l'on n'a que des variables quantitatives

Dans ce cas, on a $K' = Q$, $L' = J$, et toutes les métriques associées à l'analyse des correspondances de k_{IJ} , de B_{JJ} , ou d'un sous-tableau de B_{JJ} (cf. § 5.3.2) sont proportionnelles (si du moins on ne donne pas de pondération aux variables) à la métrique unité, ce qui compte-tenu de la symétrie des deux représentations de Q par rapport à l'origine, permet d'interpréter facilement ces analyses.

5.3.1 Analyse du tableau k_{IJ}

L'analyse du tableau k_{IJ} défini au § 5.2, i.e. l'analyse des correspondances du tableau des $\{(1+x_q)/2, (1-x_q)/2 | q \in Q\}$ est alors équivalente à l'analyse en composantes principales (A.C.P.) du tableau X des $\{x_q | q \in Q\}$, cette dernière analyse revenant à diagonaliser la matrice variance V des $\{x_q | q \in Q\}$. On peut noter que les valeurs propres issues de l'analyse des correspondances de k_{IJ} (*) sont égales aux valeurs propres de l'A.C.P. de X divisées par C_Q , et que la représentation des individus dans l'analyse de k_{IJ} est identique au coefficient $1/\sqrt{C_Q}$ près à celle obtenue dans l'A.C.P. de X ; quant à l'ensemble Q des variables, du fait du dédoublement, on obtient, comme on l'a déjà noté au § 5.2, deux représentations de Q , symétriques par rapport à l'origine, l'abscisse de la projection de x_q^+ sur l'axe α étant égale à la covariance

(*) De façon précise, l'analyse des correspondances de k_{IJ} revient à diagonaliser la matrice $W = (1/(2C_Q)) \begin{pmatrix} V & -V \\ -V & V \end{pmatrix}$.

(et donc la corrélation, si x_q est réduite) entre la variable x_q et le α -ème facteur sur I de variance 1.

Si les variables x_q sont réduites, l'analyse de k_{IJ} revient donc au facteur $1/C_Q$ près à diagonaliser la matrice de corrélation des x_q dont la trace est C_Q . On retrouve ainsi que la trace de k_{IJ} vaut 1 (résultat que l'on aurait pu déduire directement de (67)) comme dans le cas d'un tableau de 0-1 dédoublé, et que dans ce cas toutes les valeurs propres sont inférieures ou égales à 1.

Remarques : 1) Soit $\{y_q | q \in Q\}$ une suite de variables binaires ne prenant que les valeurs 0 ou 1. On sait (cf. [BIN. MULT.], *Cahiers*, Vol II n° 1 § 3.5 pp 67 sqq) que l'analyse des correspondances du tableau des y_q dédoublé revient à diagonaliser au facteur $1/C_Q$ près la matrice de corrélation des y_q .

Dans l'optique adoptée ici, cette analyse peut aussi être regardée comme l'analyse des correspondances du tableau des $(1+x_q)/2$ dédoublé (par rapport à 1), x_q étant la variable centrée réduite associée à y_q .

2) Au lieu d'associer à une variable x_q de variance σ_q^2 les variables $(1+x_q)/2$ et $(1-x_q)/2$ qui sont corrélées (le sous-tableau B_{JqJq} du tableau de Burt B_{JJ} n'étant pas diagonal), on peut associer les variables $y_q^+ = \sigma_q (1+x_q/\sigma_q)/2 = (\sigma_q + x_q)/2$ et $y_q^- = (\sigma_q - x_q)/2$, qui sont non corrélées, ce qui permet de garder l'effet de variance de x_q , tout en ayant des blocs B_{JqJq} diagonaux. Dans ce cas tous les éléments du tableau B_{JJ} sont positifs ou nuls, comme on le vérifie aisément ; toutes les propriétés données au § 3.2.2 et dans son addendum sont donc valables, comme on l'a déjà signalé, tandis que les valeurs propres issues de l'analyse des correspondances du tableau associé aux y_q^+ et y_q^- sont inférieures ou égales à 1. Notons que cette analyse revient à faire l'A.C.P. du tableau des $x_q/\sqrt{\sigma_q}$ (et non x_q/σ_q , comme dans le cas de l'A.C.P. sur matrice de corrélation).

3) Faire l'A.C.P. par l'analyse des correspondances du tableau k_{IJ} ne présente d'intérêt pratique que si l'utilisateur ne possédant pas de programme d'A.C.P., a un programme d'analyse des correspondances à sa disposition avec de bonnes sorties (contributions, éléments supplémentaires, graphiques, etc.), puisque le passage par l'analyse des correspondances revient à diagonaliser une matrice de dimension double de celle que l'on diagonalise en A.C.P..

4) Dans la thèse de F. Nakhlé (cf. *Cahiers* Vol I n° 3 pp 243-257 et n° 4 pp 367-379(*)), on considère un tableau de notes dédoublé, construit à partir d'un tableau de notes $k_0(i,q)$ suivant les formules :

$$k(i,q^+) = k_0(i,q) ; k(i,q^-) = a(q) - k_0(i,q),$$

où la constante $a(q)$ définit l'intervalle de variation $(0, a(q))$ de l'épreuve q ; F. Nakhlé démontre que (p désignant le nombre des épreuves) l'analyse du tableau dédoublé se ramène à la diagonalisation d'une matrice $p \times p$. Il est clair qu'aux notations près les résultats de F. Nakhlé ne diffèrent pas de ceux exposés dans ce §.

(*) cf. aussi *Prat. de l'A. des Données*, Vol 2 § V n°s 1 et 2, où ces deux articles sont repris.

5.3.2 Etude de la liaison entre deux groupes de variables quantitatives

Si l'on veut étudier les liaisons entre deux groupes Q_1 et Q_2 de variables (on supposera sans nuire à la généralité que $Q = Q_1 \cup Q_2$), et si l'on fait l'analyse des correspondances du sous-tableau de Burt construit à partir de k_{IJ} et croisant ces deux groupes, on est amené à diagonaliser (*) $V_{12} V_{21} / (C_{Q_1} C_{Q_2})$ et $V_{21} V_{12} / (C_{Q_1} C_{Q_2})$, V_{12} étant la matrice de covariance des variables $\{x_q | q \in Q_1\}$ avec les variables $\{x_q | q \in Q_2\}$ et V_{21} sa transposée.

Dans le cas où les variables x_q ($q \in Q$) sont réduites, et où les variables $\{x_q | q \in Q_1\}$ sont non corrélées entre elles, ainsi que les variables $\{x_q | q \in Q_2\}$, les hypothèses 1) et 2) du § 3.2.1 sont réalisées, à savoir :

$$1) \quad \forall q \in Q, P_{J_q J_q} \text{ est diagonal}$$

$$2) \quad \forall (q, q') \in Q_1^2 \cup Q_2^2 : P_{J_q J_{q'}} = P_{J_q} \otimes P_{J_{q'}}$$

comme on le vérifie immédiatement puisque :

$$\forall j = x_q^e, \forall j' = x_{q'}^e : P_{J_q J_{q'}}(j, j') = (1 + \varepsilon \varepsilon' \text{Cov}(x_q, x_{q'})) / 4 = 1/4$$

$$P_{J_q}^{\cdot}(j) = P_{J_{q'}}^{\cdot}(j') = 1/2$$

Comme de plus les variables ont même pondération, (le tableau analysé vérifie les équations (3) et (4) avec $\forall q \in Q_1 : a_q = 1/C_{Q_1}$, $\forall q \in Q_2, b_q = 1/C_{Q_2}$) on déduit des résultats du § 3.2.1 que l'analyse des correspondances du tableau croisant Q_1 et Q_2 est équivalente à l'analyse des correspondances du tableau k_{IJ} ou du tableau de Burt B_{JJ} , toutes ces analyses étant équivalentes à l'analyse canonique de ces deux groupes de variables. Notons que l'équivalence avec l'analyse canonique était évidente, sans recourir aux résultats du § 3.2.1, puisque dans cette analyse on diagonalise $V_{11}^{-1} V_{12} V_{22}^{-1} V_{21}$ et $V_{22}^{-1} V_{21} V_{11}^{-1} V_{12}$ (où V_{ii} ($i = 1, 2$) est la matrice variance des $\{x_q | q \in Q_i\}$), et qu'avec les hypothèses effectuées V_{ii} ($i = 1, 2$) se réduit à la matrice unité.

L'avantage (par rapport à l'analyse canonique), de la technique proposée ici pour étudier les liaisons entre deux groupes de caractères quantitatifs réside dans sa simplicité, la facilité d'interprétation des résultats, et le fait qu'il existe des programmes d'analyse des correspondances performants, alors que l'analyse canonique si séduisante d'un point de vue théorique, l'est beaucoup moins d'un point de vue pratique. Par contre, contrairement à l'analyse canonique, la diagonalisation de $V_{12} V_{21}$ ne permet pas d'éliminer l'effet variance de chaque variable et l'effet des liaisons à l'intérieur de chaque groupe, effets qui ne sont pas intéressants ici puisqu'on étudie les

(*) De façon précise, si l'on pose $A_{12} = V_{12} V_{21} / (2 C_{Q_1} C_{Q_2})$, l'analyse des correspondances du sous-tableau de Burt croisant Q_1 et Q_2 revient à diagonaliser la matrice $B_{12} = \begin{pmatrix} A_{12} & -A_{12} \\ -A_{12} & A_{12} \end{pmatrix}$ et la matrice B_{21} obtenue en échangeant les indices 1 et 2.

liaisons entre les deux groupes de variables. Pour éliminer le premier effet, il suffit de supposer les variables x_q réduites, mais le second effet subsiste.

Remarque : Si l'on veut étudier les liaisons entre Q_1 et Q_2 en faisant jouer un rôle dissymétrique à Q_1 et Q_2 , Q_1 étant considéré par exemple comme le groupe des variables explicatives, on est amené à diagonaliser $V_{11}^{-1} V_{12} V_{21}$ (cf. Rao, Sankya, séries A, Vol 26, part 4, Déc. 64, pp 329-358, et Obadia, R.S.A., Vol 26 n° 4, 1978, pp 5-28) ce qui est équivalent à la méthode proposée ici si V_{11} se réduit à la matrice unité.

5.4 Etude du cas où l'on croise l'ensemble K des variables qualitatives avec l'ensemble K' des variables quantitatives

Le tableau C_{LL} , dont on fait ici l'analyse des correspondances n'est autre que le sous-tableau B_{LL} , du tableau de Burt B_{JJ} , dont le terme général est donné par (68). L'analyse de ce tableau revient à diagonaliser la matrice $\Sigma\{B_q | q \in K\} / (C_K C_K)$, B_q étant la matrice variance interclasse des variables quantitatives associée à la partition définie par la $q^{\text{ème}}$ variable qualitative. Si les variables quantitatives sont non corrélées et de variance 1, et si les variables qualitatives sont deux à deux indépendantes, compte-tenu de ce que toutes les variables ont même pondération, (C_{LL} vérifie les équations (3) et (4) avec $a_q = 1/C_K$ si $q \in K$, $b_q = 1/C_K$ si $q \in K'$), l'on déduit encore du § 3.2.1 que l'analyse de C_{LL} est équivalente à celle de k_{IJ} et de B_{JJ} , toutes ces analyses étant équivalentes à l'analyse canonique dans R^I des sous-espaces respectivement engendrés par les variables quantitatives, et par les indicatrices des modalités de toutes les variables qualitatives.

Dans le cas important en pratique où l'on n'a qu'une seule variable qualitative ($C_K = 1$) la technique proposée ici revient à diagonaliser la matrice variance interclasse A des variables quantitatives, alors que dans l'analyse factorielle discriminante usuelle, on diagonalise $V^{-1} A$, V étant la matrice variance des variables quantitatives. Ici encore, on a intérêt à supposer les variables quantitatives réduites; et la différence entre l'analyse des correspondances (diagonalisation de A) et l'analyse discriminante, réside dans le fait que dans la seconde on élimine (par le choix de la métrique d'inertie V^{-1}) l'effet des liaisons entre les variables quantitatives, effet qui subsiste dans l'analyse des correspondances. En fait d'un point de vue pratique, on aura intérêt à effectuer une analyse factorielle (A.C.P. ou analyse des correspondances) sur le tableau des variables quantitatives et à effectuer une analyse discriminante usuelle sur les premiers facteurs issus de cette analyse.

5.5 Conclusion : portée des méthodes étudiées ici

Les méthodes proposées ici permettent de considérer l'A.C.P. comme une analyse de correspondance particulière, et fournissent d'autres techniques pour étudier la liaison entre deux groupes de variables quelconques, en effectuant l'analyse des correspondances du sous-tableau de

Burt associé, ce qui précise notre conception théorique d'ensemble de l'analyse des données. Néanmoins, dans la pratique, il nous semble préférable de traiter un mélange de variables qualitatives et quantitatives en découpant les variables quantitatives en classes de façon à homogénéiser l'étude, pour appliquer l'analyse des correspondances usuelle : ceci permet d'une part de tenir compte des liaisons non linéaires entre variables quantitatives, et d'autre part d'éviter l'hétérogénéité qui subsiste avec le codage proposé au § 5.2, puisque dans le tableau k_{IJ} défini au § 5.2, une colonne associée à une modalité d'une variable qualitative ne peut comporter que deux valeurs 0 ou 1, alors que les colonnes x_q^+ ou x_q^- associées à une variable quantitative peuvent comporter autant de valeurs qu'il y a d'individus (avec même des nombres négatifs).

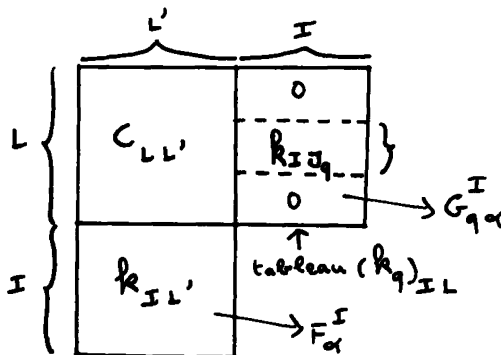
(à suivre)

Addendum au § 4.5.1 de [ANA. BLOCS II] sur les contributions d'une question à un axe factoriel.

Nous donnons ci-dessous une remarque relative aux contributions d'une question sur un axe factoriel α issu d'une analyse de correspondances multiples, remarque qui est à insérer à la fin du § 4.5.1 de [ANA. BLOCS II] (cf. *Cahiers*, Vol V n° 4, p. 399).

Remarque : On peut donner une autre interprétation aux contributions $CR_\alpha(q)$, $CR'_\alpha(q)$, $CR''_\alpha(q)$.

Prenons par exemple le cas de la contribution $CR''_\alpha(q)$ de q ($q \in K$) à la variance λ''_α du α -ème facteur issu de C_{LL} . Si l'on adjoint en élément supplémentaire de C_{LL} , le tableau $(k_q)_{IL}$ dont les blocs $I \times J_q$, ($q' \in K$) sont nuls sauf le bloc $I \times J_q$ qui est égal à k_{IJ_q} et si l'on



désigne par $G_{q\alpha}^I$ le facteur associé, on a :

$$CR''_\alpha(q) = \text{Cov}(F_\alpha^I, G_{q\alpha}^I) \sqrt{\lambda''_\alpha} / C_K$$

$$\sqrt{CR''_\alpha(q)} = \text{Corr}(F_\alpha^I, G_{q\alpha}^I) \sqrt{\text{Var}(F_\alpha^I) / C_K}$$

Faisant $L = L' = J$, $K = K' = Q$, on en déduit, dans le cas du tableau de Burt, et du tableau disjonctif complet que :

$$CR'_\alpha(q) = \text{Cov}(F_\alpha^I, G_{q\alpha}^I) \lambda_\alpha / C_Q = \lambda_\alpha CR_\alpha(q)$$

$$\sqrt{CR'_\alpha(q)} = \text{Corr}(F_\alpha^I, G_{q\alpha}^I) \sqrt{\lambda_\alpha / C_Q} = \sqrt{\lambda_\alpha CR_\alpha(q)}$$

λ_α désignant, rappelons-le, la valeur propre associée au α -ème facteur issu du tableau disjonctif complet k_{IJ} .

N.B. a) La formule reliant $CR_\alpha(q)$ et $\text{Corr}(F_\alpha^I, G_{q\alpha}^I)$ est donnée sous une forme légèrement différente par Tenenhaus (cf. L'analyse en composantes principales de variables qualitatives, à paraître dans Les Publications du Laboratoire de Statistique de l'Université Paul Sabatier, Toulouse).

b) On peut noter que le facteur $G_{q\alpha}^I$ obtenu en mettant $(k_q)_{IL}$ en élément supplémentaire de C_{LL} , n'est rien d'autre que la fonction $\varphi_{q\alpha}^I = \varphi_\alpha^{Jq} \circ p_{Jq}^I$ donnée par (39) où φ_α^{Jq} est la restriction à J_q du facteur de variance 1 φ_α^L issu de C_{LL} .

c) Les formules précédentes (où il faut remplacer $1/C_K$ (resp. $1/C_Q$) par a_q (resp. c_q) restent valables si l'on considère les tableaux plus généraux définis au § 3, à condition de supposer les tableaux p_{JqJq} ($q \in K$) diagonaux, auquel cas on a $\text{Var}(G_{q\alpha}^I) = \text{Var}(\varphi_\alpha^{Jq})$, cette dernière variance étant égale d'après (62) (où $1/C_K$ est remplacé par a_q) à $CR''_\alpha(q) / (a_q \lambda''_\alpha)$.

Si ce n'est pas le cas, les formules relatives aux $CR''_\alpha(q)$, $CR'_\alpha(q)$, $CR_\alpha(q)$ restent exactes, mais les formules relatives aux racines carrées de ces quantités ne sont plus valables, car $\text{Var}(G_{q\alpha}^I)$ est alors en général différent de $\text{Var}(\varphi_\alpha^{Jq})$.

d) La formule reliant $\text{Cov}(F_\alpha^I, G_{q\alpha}^I)$ et $CR''_\alpha(q)$ résulte de la relation $\text{Cov}(F_\alpha^I, G_{q\alpha}^I) = \sqrt{\lambda''_\alpha} \text{Var} \varphi_\alpha^{Jq}$, relation qui se démontre de façon analogue à la relation (43) (cf. [ANA. BLOCS I], Cahiers Vol V n° 2 p. 157).