P. CAZES

L'analyse de certains tableaux rectangulaires décomposés en blocs : généralisation des propriétés rencontrées dans l'analyse des correspondances multiples. II. Questionnaires : variantes de codages et nouveaux calculs de contributions

Les cahiers de l'analyse des données, tome 5, nº 4 (1980), p. 387-403

<http://www.numdam.org/item?id=CAD_1980__5_4_387_0>

Les Cahiers de l'Analyse des Données
Vol. V 1980 *n°* 4 *p* 387 403

L'ANALYSE DE CERTAINS TABLEAUX RECTANGULAIRES DÉCOMPOSÉS EN BLOCS : GÉNÉRALISATION DES PROPRIÉTÉS RENCONTRÉES DANS L'ANALYSE DES CORRESPONDANCES MULTIPLES II QUESTIONNAIRES : VARIANTES DE CODAGES ET NOUVEAUX CALCULS DE CONTRIBUTIONS

[ANA. BLOCS II] (*) (*à suivre*)

par P Cazes (1)

4 Cas particulier des questionnaires. Généralisations diverses des correspondances multiples.

4.1 Introduction

On suppose ici que I est un ensemble d'individus, et $J = \cup\{J_q | q \in Q\}$ l'ensemble des modalités de toutes les questions d'un questionnaire, J_q désignant l'ensemble des modalités associé à la $q^{ème}$ question.

Nous allons proposer plusieurs codages pour formuler les réponses d'un individu i au questionnaire, ce qui nous donnera des généralisations des correspondances multiples dans le cas du codage pondéré (§ 4.2.2) et du codage flou (§ 4.2.3), puis nous donnerons une autre généralisation des correspondances multiples permettant de traiter de façon différente le codage flou (§ 4.2.4). A chaque codage correspondra un tableau sur $I \times J$ et les tableaux associés sur $J \times J$ et $L \times L'$ (tableaux définis au § 3.1, dont on conservera les notations, à partir du tableau initial sur $I \times J$), tous ces tableaux possédant les caractéristiques des tableaux étudiés aux §§ 2 et 3.1, et vérifiant donc les propriétés données aux §§ 2 et 3.2.

Après avoir traité de ces codages au § 4.2, nous examinons au § 4.3 les contributions d'une question à l'inertie totale des différents tableaux étudiés, et rappelons l'intérêt du tableau de Burt modifié ; nous donnons ensuite au § 4.5 les contributions d'une question à l'inertie d'un facteur ce qui nous permet de suggérer des aides à l'interprétation basées sur les variables, et qui complètent les aides usuelles associées aux modalités ; nous terminons ce § 4 en donnant au § 4.6 des indications sur le moyen de représenter une variable quantitative quelconque sur les axes factoriels issus d'une analyse des correspondances multiples.

4.2 Codages du questionnaire. Généralisations diverses des correspondances multiples.

4.2.1 Codage disjonctif complet (cf. [BIN. MULT.], Cahiers Vol II n° 1, 1977, pp 55 s.q.q.).

C'est le codage usuel :

$$\forall j \in J_q : k_{IJ}(i,j) = 1 \text{ si } i \text{ a pris la modalité } j \text{ de } J_q$$
$$= 0 \text{ sinon.}$$

(*) *Suite de l'article paru sous le même titre, dans les* Cahiers, *Vol V n°2 pp 145-161.*

(1) *Maître Assistant, Laboratoire de Statistique, Université P. & M. Curie.*

Dans ce cas, la relation (30) est vérifiée avec $c_q = 1/C_Q$, C_Q désignant le cardinal de Q.

4.2.2 *Codage disjonctif complet pondéré* (cf. L. Benyamina, rapport de stage de D.E.A., juin 1980, et *Cahiers*, Vol V n° 2 p. 192).

Ici, on code par une quantité α_q (au lieu de 1) la présence à une modalité de J_q.

$$\forall\ j \in J_q\ :\ k_{IJ}(i,j) = \alpha_q \text{ si } i \text{ a pris la modalité } j \text{ de } J_q$$
$$= 0 \text{ sinon.}$$

Dans ce cas, la relation (30) est vérifiée avec :

$$c_q = \alpha_q / \Sigma\{\alpha_{q'} \mid q' \in Q\} \qquad (56)$$

Les coefficients α_q peuvent être choisis de façon à optimiser certains critères (cf. Cazes et alt., Statistique et Analyse des Données, n° 2, 1976, pp 48-62, et Saporta, Statistique et Analyse des Données, n° 3, 1979, pp 19-31), critères liés à l'inertie ou à la somme des k (k≥1) plus grandes valeurs propres issues de k_{IJ} (ou du tableau de Burt associé) ; on impose alors en général une contrainte de normalisation sur les α_q, par exemple que $\Sigma\{\alpha_q^2 \mid q \in Q\} = 1$. On peut aussi choisir les coefficients α_q de façon à rendre les contributions de chaque question proportionnelles à des nombres fixés à l'avance (cf. § 2.3, remarque 1 *in fine*, et § 4.3).

En fait ce codage n'est intéressant que pour équilibrer les contributions des questions dans l'analyse du tableau de Burt B_{JJ} associé au tableau disjonctif pondéré, ou d'un sous-tableau $C_{LL'}$ de B_{JJ} (cf. §§ 4.3.2 & 4.3.3), les contributions d'une question, l'inertie totale, ainsi que les pourcentages d'inertie ne présentant pas d'intérêt dans le cas d'un tableau disjonctif complet (cf. formules (57) et (58), ainsi que la remarque n° 3 du § 4.3.1).

4.2.3 *Codage flou*

Ce codage s'impose quand certaines des variables du questionnaire résultent d'un découpage en classes de variables quantitatives, et que l'effectif Card I du nombre d'individus analysés est relativement faible. Il revient pour chaque question q, à affecter à tout individu i un système de probabilités $\alpha_q(i,j)$ de choisir chaque modalité j de J_q. On aura donc :

$$\forall\ j \in J_q\ :\ k_{IJ}(i,j) = \alpha_q(i,j) \qquad (56\text{ bis})$$

$\alpha_q(i,j)$ appartenant à l'intervalle (0,1), avec :

$$\Sigma\{\alpha_q(i,j) \mid j \in J_q\} = 1$$

La relation (30) est toujours vérifiée avec $c_q = 1/\text{Card}\,Q$.

Il est clair que rien ne permet en toute rigueur de déterminer de véritables probabilités : mais il s'agit seulement d'assouplir le codage usuel en (0,1).

Avec le codage flou, si un individu se trouve à la frontière entre deux classes j et j' de J_q, au lieu de l'affecter soit à j, soit à j', on lui donnera une probabilité 0,5 d'être dans chacune des deux classes précédentes, ce qui fera perdre moins d'information , et devrait donner des résultats d'analyse plus stables.

On peut noter que le tableau de Burt associé à k_{IJ} n'a pas la structure diagonale du tableau de Burt usuel, ou de celui déduit du codage pondéré (cf. § 4.2.2), car les sous-tableaux diagonaux $J_q \times J_q$ comportent des termes non-nuls à l'extérieur de leur diagonale. C'est la raison pour laquelle on introduit au § suivant une généralisation des correspondances multiples, qui dans le cas du codage flou permet de conserver au tableau de Burt sa structure diagonale.

Remarques

1) Dans les trois codages précédents, on peut donner une pondération p_i à chaque individu i, en multipliant le codage de i par p_i. Cette transformation ne modifie pas les propriétés du tableau construit sur $I \times J$ et des tableaux associés construits à partir de la procédure du § 3.1. Tous les résultats donnés dans ce § 4 (ainsi du reste que ceux donnés avec le codage proposé au § 5, ou avec les cas modèles étudiés au § 6) restent valables. En particulier les différentes contributions calculées aux §§ suivants sont inchangées par cette transformation. Notons qu'il faut alors munir I de la loi p_I et non plus de la loi uniforme.

2) On peut combiner le codage flou et le codage pondéré en écrivant que la somme des $\alpha_q(i,j)$ pour j décrivant J_q est égal à α_q (et non à 1 comme pour un système de probabilités).

3) Le codage flou a en particulier été utilisé par Guitonneau et Roux (cf. [Le Genre Erodium], *Cahiers*, Vol II n° 1, 1977, pp 97-113), Lefol (thèse 3-ème cycle, Paris, 1979) et par Gallégo (thèse 3-ème cycle, Paris, 1980).

4) D'un point de vue pratique, le codage d'un individu i pour une variable x_q découpée en classes se fait en général suivant deux classes contiguës (i.e. les $\alpha_q(i,j)$ pour $j \in J_q$ sont nuls sauf pour deux d'entre-eux dont la somme vaut 1).

Le codage flou revient de façon générale à remplacer la valeur $x_q(i)$ par les $\{\alpha_q(i,j) \mid j \in J_q\}$ où $\alpha_q(i,j) = z_j(x_q(i))$, les z_j étant des fonctions de codage. On choisit en général pour les z_j des fonctions linéaires par morceaux . Gallégo (op. cit.) donne dans sa thèse un programme de codage flou (DCFLOU) qui utilise ce type de codage linéaire avec des fonctions exponentielles pour les classes extrêmes (cf. figure 1) ce qui permet à partir des $\{\alpha_q(i,j) \mid j \in J_q\}$ de reconstituer la valeur exacte $x_q(i)$ de x_q.

4.2.4 Une autre présentation des correspondances multiples. Application au codage flou

Nous reprenons ici la présentation proposée dans La Pratique de l'Analyse des Données, Vol II, ch VI n° 0 § 4.1.

On pose : $S = \Pi\{J_q \mid q \in Q\}$, ensemble de tous les systèmes de réponses possibles et l'on suppose que l'on a une loi de probabilité p_S sur S . J désignant toujours l'union disjointe des J_q , on considère

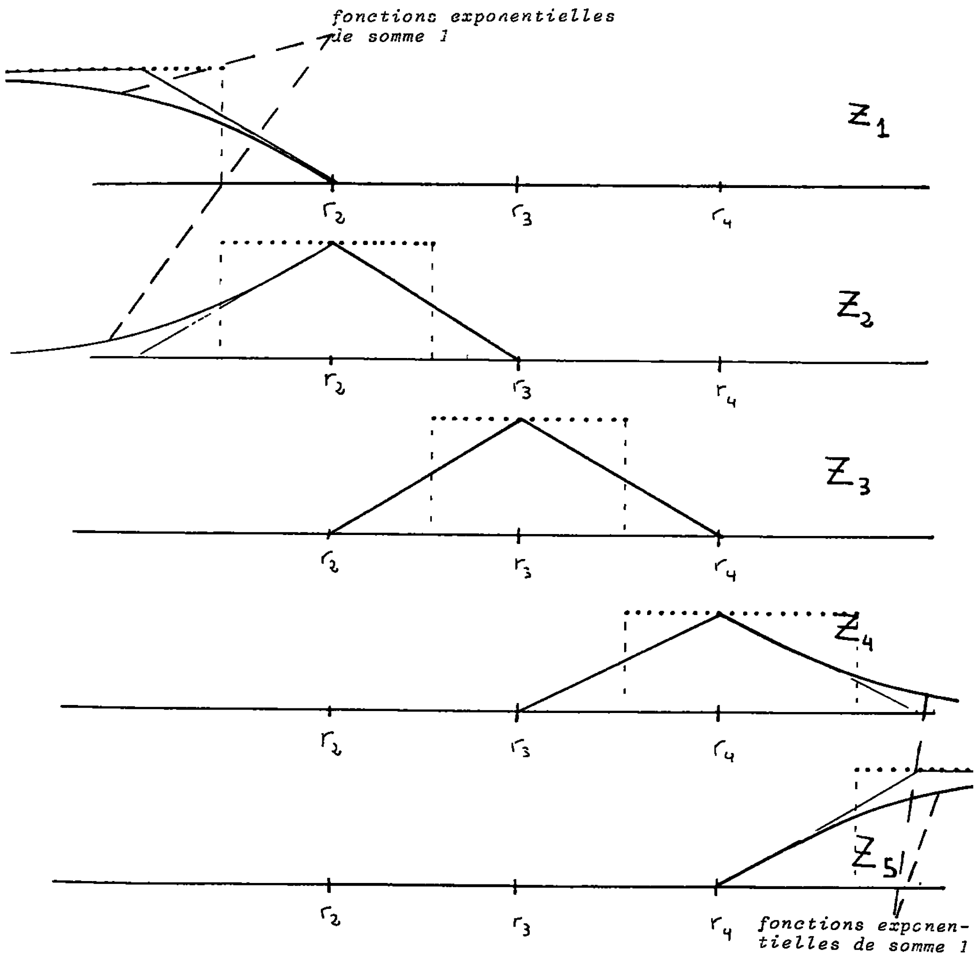

Figure 1 : Fonctions donnant l'éclatement flou d'une variable quantitative en cinq modalités. En pointillé, un découpage disjonctif (extrait de la thèse de F. Gallégo).

le tableau disjonctif complet λ_{SJ} associé à un questionnaire dont tout C_Q - uple de réponses possibles a été pris une fois et une seule :

$$\forall\, j \in J_q\ ,\ \forall\, s = \{j_1\ , j_2\ , \ldots\ j_q\ , \ldots , j_{C_Q}\} \in S\ :$$

$$\lambda_{SJ}(s,j) = 1 \text{ si } j = j_q$$
$$= 0 \text{ sinon}$$

Notons que les facteurs non triviaux issus de λ_{SJ} ne sont autres que les fonctions sur J, centrées sur chaque J_q, ces fonctions engendrant un sous-espace propre de dimension $C_J - C_Q$ associé à la valeur propre $1/C_Q$.

A la loi p_S on associe le tableau k_{SJ} défini par :

$$\forall\ (s,j) \in S \times J\ :\ k_{SJ}(s,j) = p_s\ \lambda_{SJ}(s,j)$$

Dans le cas où un ensemble I d'individus a repondu au questionnaire dont J est l'ensemble des modalités de réponses, p_s n'est rien d'autre que la proportion des individus i de I ayant fourni le système de réponses s. D'après le principe d'équivalence distributionnelle, l'analyse des correspondances du tableau disjonctif complet k_{IJ} associé à I est équivalente à celle du tableau k_{SJ}.

A partir du tableau k_{SJ}, on peut construire à l'aide des formules (31) et (33) où i est remplacé par s, le tableau de Burt B_{JJ}, ou des sous-tableaux de ce tableau. Ces tableaux s'obtiennent, comme il est immédiat de le vérifier, par superposition des lois marginales binaires $p_{J_qJ_{q'}}$ de p_S, la loi $p_{J_qJ_q}$ étant la loi marginale sur J_q, portée par la diagonale de $p_{J_qJ_q}$.

En analysant le tableau B_{JJ}, ou des sous-tableaux de B_{JJ}, on peut ainsi étudier la correspondance multiple associée à toute loi p_S définie sur S.

Application au codage flou.

Nous reprenons ici les notations du § 4.2.3.

Soit i un individu, $\alpha_q(i,j)$ la probabilité que i prenne la modalité j_q de J_q. On supposera que les réponses de i aux différentes questions sont indépendantes. Cette hypothèse semble peu réaliste, mais l'analyse du tableau k_{SJ} définie ci-dessous, fournit dans le cas où le codage flou se réduit au codage disjonctif complet usuel ($\alpha_q(i,j) = 1$ pour un seul j de J_q, et 0 ailleurs) les mêmes résultats que l'analyse du tableau disjonctif complet classique, ce qui légitime dans une certaine mesure, l'hypothèse effectuée.

La probabilité $p_s(i)$ que i prenne le système de réponses $s = \{j_1\ , \ldots , j_q\ , \ldots , j_{C_Q}\}$ s'écrit alors :

$$p_s(i) = \Pi\{\alpha_q(i,j_q) \mid q \in Q\}$$

et la probabilité d'avoir sur l'échantillon I la réponse s s'écrit, C_I désignant le cardinal de I :

$$p_s = \Sigma\{p_s(i) \mid i \in I\}/C_I$$

Pour étudier le questionnaire défini par le codage flou du § 4.2.3, on considère le tableau p_S construit ci-dessus, et on analyse les tableaux k_{SJ}, B_{JJ} (ou des sous-tableaux $C_{LL'}$ de B_{JJ}) déduits de p_S ; on peut alors représenter l'ensemble I des individus en mettant en supplémentaire des analyses précédentes le tableau k_{IJ} des codages flous (ou dans le cas de $C_{LL'}$, les sous-tableaux k_{IL} et $k_{IL'}$ de k_{IJ}). En somme le codage flou est interprété comme superposant au codage usuel un effet de dispersion qui à chaque individu réel associe un nuage d'individus fictifs offrant tous les systèmes de réponses avoisinant (avec pour ces individus fictifs des poids convenablement calculés).

Remarques

1) Soit B'_{JJ} le tableau de Burt associé au tableau k_{IJ} du codage flou défini par la relation (56 bis) du § 4.2.3. B_{JJ} désignant toujours le tableau de Burt associé à p_S, il est immédiat de vérifier que les tableaux B'_{JJ} et $C_I\ B_{JJ}$ sont, hormis leurs termes diagonaux, identiques : pour passer du tableau B'_{JJ} au tableau B_{JJ}, il suffit, au coefficient $1/C_I$ près, d'annuler les termes extradiagonaux de chaque bloc diagonal de B'_{JJ}, après avoir reporté leur masse sur la diagonale, ce qui revient à prendre pour termes diagonaux la marge du bloc diagonal considéré (avant annulation des termes non diagonaux).

En pratique, on aura intérêt à raisonner sur le tableau de Burt B_{JJ}, puisqu'il possède la même structure diagonale que le tableau de Burt usuel. En fait, on verra au § 4.4 que l'on a intérêt à raisonner non sur le tableau de Burt lui-même, mais sur le tableau de Burt modifié obtenu en annulant les blocs diagonaux du tableau de Burt. En opérant ainsi, les deux approches du codage flou que nous avons proposées sont équivalentes.

2) Utilisant la remarque précédente D. Maïti a programmé plusieurs procédures de codage flou basées sur l'approche considéré ici (i.e. analyse des correspondances du tableau de Burt B_{JJ} associé à la loi p_S, avec mise en supplémentaire du tableau k_{IJ}) et il a incorporé ces procédures à la version STECKMA-3 de son programme STECKMA (cf. *Cahiers*, Vol IV n° 4, 1979, pp 465-sqq).

4.3 *Contributions d'une question à l'inertie totale dans le cas d'un tableau disjonctif complet, du tableau de Burt ou de sous-tableaux associés.*

Dans tout ce §, on suppose, sauf spécification contraire, qu'on a le codage disjonctif complet rappelé au § 4.2.1.

4.3.1 *Contributions dans le cas d'un tableau disjonctif complet*

Les contributions CR(j), CR(J_q) (qu'on notera aussi CR(q)) à la trace CR(J) du tableau disjonctif complet k_{IJ} s'écrivent respectivement (cf. La Pratique de l'A. des Données, Vol II, ch VI n° 0 § 2.2), si p_j désigne la proportion des individus ayant adopté la modalité j :

$$\left.\begin{aligned} CR(j) &= (1 - p_j)/C_Q \\ CR(J_q) &= (\mathrm{Card} J_q - 1)/C_Q \\ CR(J) &= (C_J - C_Q)/C_Q \end{aligned}\right\} \quad (57)$$

On voit que la contribution d'une question q à CR(J) ne dépend que du nombre Card J_q de modalités de cette question. Considérons par exemple deux questions, la première ayant deux modalités et la seconde neuf modalités ; cette dernière question aura une contribution huit fois plus forte que la première. On aura donc intérêt, pour équilibrer l'influence des questions dans un questionnaire, à choisir , chaque fois que c'est possible, le même nombre de modalités pour chaque question. Si ce n'est pas possible, on pourra, pour égaliser l'influence de chaque question q , ou pour donner à q une contribution proportionnelle à une quantité fixée d_q , adopter le codage pondéré défini au § 4.2.2, en posant $\alpha_q = d_q/(\mathrm{Card}\, J_q - 1)$ (cf. § 2.3, *in fine*, remarque 1). On a alors si $A_T = \Sigma\{\alpha_q | q \in Q\}$:

$$\left.\begin{aligned} CR(j) &= (\alpha_q/A_T)(1 - p_j) \\ CR(J_q) &= (\alpha_q/A_T)(\mathrm{Card}\, J_q - 1) = d_q/A_T \\ CR(J) &= \Sigma\{d_q | q \in Q\}/A_T \end{aligned}\right\} \quad (58)$$

On retrouve bien les formules (57) si $\alpha_q = 1$ (auquel cas $d_q = \mathrm{Card}\, J_q - 1$, $A_T = \mathrm{Card}\, Q$) , tandis que si $\alpha_q = 1/(\mathrm{Card}\, J_q - 1)$ ($d_q = 1$) les contributions $CR(J_q)$ sont bien toutes égales.

Remarques

1) Si l'on pondère chaque individu par un poids p_i ($\sum_i p_i = 1$), i.e. si l'on pose $k_{IJ}(i,j) = \alpha_q p_i$ si i a pris la modalité j de J_q, et 0 sinon (avec $\alpha_q = 1$ pour le codage disjonctif complet) les formules (58) (et donc (57) quand $\alpha_q = 1$) (où p_j désigne la masse des i ayant adopté la modalité j) , restent valables comme on l'a déjà dit (cf. remarque 1 du § 4.2.3), comme on peut le vérifier aisément, et comme cela résulte immédiatement du principe d'équivalence distributionnelle dans le cas où les p_i sont proportionnels à des nombres entiers, puisque ces formules sont indépendantes de I.

2) Les formules (57) et (58) supposent essentiellement que le tableau k_{IJ} ne comporte pas de colonnes nulles, i.e. qu'on a éliminé les modalités prises par aucun individu ; ceci implique en particulier que le nombre d'individus est supérieur ou égal au nombre total de modalités.

3) Les pondérations pour équilibrer les $CR(J_q)$ présentent peu d'intérêt dans l'analyse d'un tableau disjonctif complet puisque ces contributions ne dépendent que du nombre de modalités de q , et non des liaisons de q avec les autres questions, et que de plus inertie totale CR(J) et pourcentages d'inertie ont peu de signification dans cette analyse, comme on l'a déjà signalé, et comme on le développe au § 4.4. Le cas binaire ($C_Q = 2$, $J = J_{q1} \cup J_{q2}$) permet également de

voir le peu d'intérêt à raisonner sur un tableau disjonctif pondéré. Dans ce cas, l'analyse du tableau pondéré est (comme dans le cas du tableau non pondéré, cas détaillé à la fin du § 4.4) équivalente à l'analyse du tableau croisant J_{q1} et J_{q2}. De façon précise, à tout facteur $\varphi^J = (\varphi^{Jq1}, \varphi^{Jq2})$ de variance 1 issu du tableau non pondéré (facteur qui est tel que φ^{Jq1} et φ^{Jq2} ont même variance) est associé un facteur de même rang de variance 1 $\psi^J = (\psi^{Jq1}, \psi^{Jq2})$ issu du tableau pondéré, avec $\psi^{Jq1} = a\varphi^{Jq1}$, $\psi^{Jq2} = b\varphi^{Jq2}$, a et b étant deux coefficients de proportionnalité positifs, fonctions des pondérations et de la valeur propre du tableau non pondéré. Le seul effet de la pondération est donc de donner des variances différentes aux composantes ψ^{Jq1} et ψ^{Jq2} de ψ^J sur J_{q1} et J_{q2}. On n'emploiera donc d'un point de vue pratique un système de pondérations que pour équilibrer les contributions des questions dans l'analyse d'un tableau de Burt ou d'un sous-tableau de ce tableau.

4) Il est facile de voir que si l'on analyse le tableau k_{IJ} associé au codage flou et défini par la relation (56 bis) du § 4.2.3 , l'inertie CR(q) d'une question q , et l'inertie totale CR(J) sont plus petites que les valeurs données par (57) et associées au tableau disjonctif complet.

4.3.2 *Cas d'un sous-tableau $C_{LL'}$ du tableau de Burt associé à 2 parties K et K' de Q.*

Soit $q \in K$, $q' \in K'$; les contributions respectives CR(q) et CR(q') de q et q' à l'inertie CR(L) = CR(L') de $C_{LL'}$, inertie donnée par (18) s'écrivent :

$$\left.\begin{aligned} CR(q) &= \Sigma\{\phi^2_{qq''} \mid q'' \in K'\}/(C_K C_{K'}) \\ CR(q') &= \Sigma\{\phi^2_{q''q'} \mid q'' \in K\}/(C_K C_{K'}) \end{aligned}\right\} \quad (59)$$

$\phi^2_{qq''}$ (resp. $\phi^2_{q''q'}$) étant le ϕ^2 relatif aux deux variables q et q" (resp. q" et q'), i.e. l'inertie du tableau $C_{JqJq''}$ (resp. $C_{Jq''Jq'}$) croisant ces deux variables. On peut noter d'après (18) que la contribution CR(q,q") du tableau $C_{JqJq''}$ ($q \in K$, $q'' \in K'$) à l'inertie de $C_{LL'}$ est $\phi^2_{qq''}/(C_K C_{K'})$; si on fait l'hypothèse que ce tableau, construit à partir du C_I échantillon étudié ne s'écarte de l'indépendance (i.e. de la loi produit) que par les fluctuations d'échantillonnage, $C_I\, \phi^2_{qq''}$ suit une loi du χ^2 à (Card J_q - 1)(Card $J_{q''}$ - 1) degrés de liberté. Sous cette hypothèse, l'espérance mathématique de CR(q, q") est égale à (Card J_q - 1)(Card $J_{q''}$ - 1)/$(C_I C_K C_{K'})$. Si l'on suppose que l'hypothèse précédente est vérifiée pour tout couple $(q,q') \in K \times K'$ (ce qui exige en particulier que $K \cap K' = \emptyset$) CR(q) , CR(q') et CR(L) suivent (au facteur $1/(C_I C_K C_{K'})$ près) respectivement des lois du χ^2 à (Card J_q - 1)$(C_{L'} - C_{K'})$, (Card $J_{q'}$ - 1)$(C_L - C_K)$ et $(C_L - C_K)(C_{L'} - C_{K'})$ degrés de liberté, ce qui permet en particulier de tester à partir de CR(L) cette hypothèse d'indépendance, et l'on a en désignant par E le symbole espérance mathématique :

$$\forall\ q \in K : E(CR(q\ )) = (\text{Card}\ J_q - 1)(C_{L'} - C_{K'})/(C_I C_K C_{K'})$$

$$\forall\ q' \in K' : E(CR(q')) = (\text{Card}\ J_{q'} - 1)(C_L\ - C_K\ )/(C_I C_K C_{K'})$$

$$E(CR(L\ )) = E(CR(L')) = (C_L - C_K)(C_{L'} - C_{K'})/(C_I C_K C_{K'})$$

On voit là encore que si on veut équilibrer les contributions q de K (resp. q' de K') on a intérêt, quand cela est possible à choisir le même nombre de modalités pour les questions de K (resp. K').

Si l'on veut que les $E(CR(q\ ))$ soient proportionnels à des quantités d_q ($q \in K$) fixées, et que de même les $E(CR(q'))$ soient proportionnels à des quantités $e_{q'}$ ($q' \in K'$) données, il suffit d'adopter le codage pondéré du § 4.2.2, avec $\alpha_q = d_q/(\text{Card}\ J_q - 1)$ si $q \in K$, et $\alpha_{q'} = e_{q'}/(\text{Card}\ J_{q'} - 1)$ si $q' \in K'$, ce qui revient à multiplier le sous-tableau $C_{JqJq'}$ de $C_{LL'}$ par $\alpha_q\ \alpha_{q'}$; la contribution de $C_{JqJq'}$ à $E(CR(L))$ est alors égale, si l'on pose $A_T = \Sigma\{\alpha_q | q \in K\}$, $A'_T = \Sigma\{\alpha_{q'} | q' \in K'\}$, à $d_q e_{q'}/(A_T A'_T C_I)$.

On peut également (cf. § 2.3 remarque 1), ce qui nous semble plus intéressant, rendre les $CR(q)$ ($q \in K$) proportionnels à des quantités fixées d_q en adoptant le codage pondéré défini par $\alpha_q = d_q/CR(q)$ si $q \in K$, $\alpha_{q'} = 1$ si $q' \in K'$, ce qui revient à multiplier $C_{JqJq'}$ par $d_q/CR(q)$. On peut de même rendre les $CR(q')$ ($q' \in K'$) proportionnels à des quantités $e_{q'}$ données, mais on ne peut pas, par le procédé précédent rendre directement les $CR(q)$ proportionnels aux d_q et les $CR(q')$ proportionnels aux $e_{q'}$, comme on l'a déjà noté à la fin de la remarque citée.

4.3.3 Cas du tableau de Burt B_{JJ}.

Ce cas est un cas particulier du cas précédent, celui où $K = K' = Q$. Il suffit donc dans les formules (19) et (59) de faire $K = K' = Q$ pour obtenir respectivement l'inertie totale $CR(J)$ et la contribution $CR(q)$ de q à cette inertie. Compte-tenu de ce que $\phi^2_{qq} = \text{Card}\ J_q - 1$, on a :

$$CR(q) = (\text{Card}\ J_q - 1 + \Sigma\{\phi^2_{qq'} | q' \in Q,\ q' \neq q\})/(C_Q)^2$$

Sous l'hypothèse que tout sous-tableau de B_{JJ} croisant deux questions différentes q et q', ne s'écarte de l'indépendance que par des fluctuations d'échantillonnage on a :

$$E(CR(q\ )) = (\text{Card}\ J_q - 1)(1 + ((C_J - C_Q - \text{Card}\ J_q + 1)/C_I))/(C_Q)^2$$

expression plus compliquée que celle donnée au § 4.3.2 du fait du sous-tableau diagonal croisant q avec lui-même.

Si l'on veut rendre les contributions $CR(q)$ proportionnelles à des nombres fixés d_q , il suffit de multiplier chaque sous-tableau $B_{JqJq'}$ de B_{JJ} par $d_q/CR(q)$. On obtient alors un tableau non symétrique où les contributions des questions (considérées comme lignes) sont proportionnelles aux $d_{q'}$.

On peut également, si l'on veut conserver un tableau de Burt pondéré symétrique, adopter le codage pondéré du § 4.2.2, avec par exemple

$\alpha_q = 1/(\text{Card}\, J_q - 1)$, ce qui revient à multiplier $B_{JqJq'}$ par $\alpha_q \alpha_{q'}$. On a alors, en posant $A_T = \Sigma\{\alpha_q | q \in Q\}$:

$$CR(q) = (\alpha_q^2 (\text{Card}\, J_q - 1) + \Sigma\{\alpha_q \alpha_{q'} \phi^2_{qq'} | q' \in Q,\ q' \neq q\})/A_T)^2$$

d'où l'on déduit par sommation sur q l'inertie totale.

Ce codage nous semble moins intéressant que le précédent, puisque dans ce cas, l'analyse du tableau de Burt pondéré est équivalente à celle du tableau disjonctif pondéré qui présente peu d'intérêt, comme on l'a souligné plus haut (cf. remarque 3 du § 4.3.1).

On peut noter que dans le cas où toute question q de Q est indépendante de toute autre question q' de Q ($\Phi^2_{qq'} = 0$; le tableau de fréquence $p_{JqJq'}$ associé à $B_{JqJq'}$ étant égal au produit $p_{Jq} \otimes p_{Jq'}$ de ses marges), du fait des blocs diagonaux, la trace CR(J) du tableau de Burt usuel (i.e. non pondéré) n'est pas nulle, mais égale d'après (19) (où l'on fait Q = K) à $(C_J - C_Q)/(C_Q)^2$. Dans ce cas toute fonction sur J centrée sur chaque J_q est facteur de B_{JJ} associé à la valeur propre $\lambda = (1/C_Q)^2$; cette valeur propre non triviale de B_{JJ} est de multiplicité $C_J - C_Q$; $1/C_Q$ est alors la seule valeur propre non triviale issue du tableau disjonctif complet k_{IJ} associé à B_{JJ}.

Pour supprimer les facteurs précédents qui sont sans intérêt puisque les questions sont indépendantes, et qu'ils ne traduisent qu'un effet de structure (à savoir les blocs diagonaux qui ne comportent des termes non-nuls que sur leur diagonale) il suffit d'éliminer la contribution $(C_J - C_Q)/(C_Q)^2$ des blocs diagonaux à CR(J) en considérant le tableau de Burt modifié B^M_{JJ} (cf. *Cahiers* Vol IV n° 3 p. 377) tableau qui est identique au tableau de Burt usuel, à ceci près que les blocs diagonaux ont été annulés, et dont les propriétés sont données au § suivant.

Remarque

Dans le cas d'indépendance précédent, si l'on emploie le codage pondéré α_q pour J_q, toute fonction φ^J, centrée et de support J_q est facteur du tableau de Burt correspondant associé à la valeur propre $(\alpha_q/A_T)^2$ qui est donc de multiplicité $\text{Card}\, J_q - 1$.

4.4 *Taux d'inertie dans le cas d'un tableau disjonctif complet et d'un tableau de Burt. Tableau de Burt modifié (cf. Cahiers Vol IV n° 3 pp 377-378)*

On sait que les taux d'inertie dans un tableau disjonctif complet sont faibles, et plus petits (au moins en ce qui concerne le premier facteur) que ceux associés au tableau de Burt correspondant B_{JJ}. Ce dernier tableau étant un tableau de comptage, il semble plus logique de calculer les pourcentages d'inertie à partir de ce tableau plutôt qu'à partir du tableau disjonctif complet. Mais ce tableau B_{JJ} comporte des blocs diagonaux, qui ne fournissent aucune information supplémentaire tout en apportant à la trace une contribution égale à $(C_J - C_Q)/(C_Q)^2$, comme on l'a signalé à la fin du § précédent. Il en résulte en particulier que dans le cas limite d'une correspondance continue (cas où $\forall q \in Q$, Card J_q tend vers l'infini, C_Q restant fini) la trace est infinie, et les taux d'inertie associés aux premiers facteurs sont nuls, même si ces facteurs sont significatifs. Pour éliminer cet effet, et de plus n'avoir que des facteurs triviaux quand toute question q est indépendante de toute question q' (q' $\neq$ q) (cf. § 4.3.3 *in fine*), il suffit d'analyser le tableau de Burt modifié B^M_{JJ} qui se déduit de B_{JJ} en annulant les blocs diagonaux, comme on l'a déjà dit. Ces deux tableaux ont mêmes facteurs, les valeurs propres λ^M_B (de parité ε) et λ_B respectivement issues de l'analyse de B^M_{JJ} et de B_{JJ} étant liées par la relation :

$$\varepsilon\sqrt{\lambda^M_B} = (\sqrt{\lambda_B} - 1/C_Q)(C_Q/(C_Q - 1)) \tag{60}$$

avec $\varepsilon = 1$ si $\sqrt{\lambda_B} > 1/C_Q$, auquel cas on a un facteur direct issu de B^M_{JJ}, sinon $\varepsilon = -1$, auquel cas on a un facteur inverse. La trace $CR_M(J)$ associée à l'analyse de B^M_{JJ} vaut alors d'après (13) où les coefficients a_q et b_q, relatifs au tableau B^M_{JJ} valent $1/(C_Q - 1)$, comme il est immédiat de le vérifier :

$$CR_M(J) = \Sigma\{\phi^2_{qq'} \mid q \in Q,\ q' \in Q,\ q \neq q'\}/(C_Q - 1)^2$$

trace qui reste finie dans le cas limite d'une correspondance continue, si on suppose que les $\phi^2_{qq'}$ $(q \neq q')$ ont une limite finie, hypothèse que l'on fait en général quand on étudie des correspondances continues.

De plus à la valeur propre $(1/C_Q)^2$ issue de B_{JJ} (correspondant à la valeur propre $1/C_Q$ dans le cas du tableau disjonctif complet) est associée la valeur propre nulle, ce qui semble *a priori* satisfaisant, cette valeur propre correspondant à des facteurs de structure en général sans intérêt, comme on peut le constater sur les cas modèles étudiés au § 6 (où l'on fait K = K' = Q). Par contre, aux valeurs propres nulles issues de B_{JJ} (valeurs propres associées à un sous-espace propre de dimension supérieure ou égale à $C_Q - 1$) correspondent des valeurs propres égales à $1/(C_Q - 1)^2$ pour B^M_{JJ} ; leur contribution à la trace issue de B^M_{JJ} est donc supérieure ou égale à $1/(C_Q - 1)$. En fait, ces valeurs propres correspondent à des facteurs inverses, et on va voir, par référence au cas binaire ($C_Q = 2$) qu'il ne faut pas tenir compte des facteurs inverses (i.e. des valeurs propres de B_{JJ} telles que $\sqrt{\lambda_B} < 1/C_Q$, ce qui exclut en particulier les valeurs propres nulles).

En effet, considérons une correspondance binaire ($C_Q = 2$, $J = J_{q1} \cup J_{q2}$). A tout couple de facteurs non triviaux $(\varphi^{Jq1}, \varphi^{Jq2})$ issu du tableau de contingence $J_{q1} \times J_{q2}$ et relatif à la valeur propre μ, sont associés deux facteurs $\varphi^J_\varepsilon = (\varphi^{Jq1}, \varepsilon\varphi^{Jq2})$ $(\varepsilon = \pm 1)$ issus du tableau de Burt B_{JJ} ou du tableau de Burt modifié B^M_{JJ}. En ce qui concerne B_{JJ}, φ^J_ε est associé à la valeur propre $(1 + \varepsilon\sqrt{\mu})^2/4$; pour B^M_{JJ}, φ^J_ε est associé à la valeur propre μ, φ^J_+ $(\varepsilon = 1)$ étant direct et φ^J_- $(\varepsilon = -1)$ étant inverse ; ces deux facteurs apportant la même information, il suffit de ne considérer que le facteur direct φ^J_+ . En se restreignant ainsi aux facteurs directs issus de B^M_{JJ} , on obtient les mêmes valeurs propres et les mêmes pourcentages d'inertie que pour le tableau $J_{q1} \times J_{q2}$. Notons qu'à tout facteur associé à la valeur propre nulle issu de ce dernier tableau correspond un facteur associé à la valeur propre nulle pour B^M_{JJ} et à $1/4 = (1/C_Q)^2$ pour B_{JJ} , d'où l'intérêt encore de raisonner sur B^M_{JJ}. Ces résultats se généralisent au cas modèle étudié au § 6.1.1 où au lieu du tableau croisant q_1 et q_2 , on considère le tableau $C_{LL'}$ croisant deux groupes K et K' de variables, les variables étant deux à deux indépendantes à l'intérieur de chaque groupe.

Dans le cas général ($C_Q > 2$), on ne considérera donc que les facteurs directs issus de B^M_{JJ} , et l'on calculera les taux d'inertie d'après ces seuls facteurs.

4.5 *Contributions d'une question sur un axe factoriel et aides à l'interprétation basées sur les questions*

4.5.1 *Contributions d'une question sur un axe factoriel*

Soit $(F^I_\alpha , G^J_\alpha)$ un couple de facteurs associés, de variance la valeur propre λ_α , issu du tableau disjonctif complet k_{IJ}. La contribution à λ_α de l'ensemble J_q des modalités de la q-ème question s'écrit :

$$CR_\alpha(q) = \Sigma\{k(j)G^2_\alpha(j)/k \mid j \in J_q\}$$

$$= \Sigma\{k(j)G^2_\alpha(j)/(C_I C_Q) \mid j \in J_q\}$$

puisque le total k de k_{IJ} vaut $C_I C_Q$.

Pour interpréter $CR_\alpha(q)$, on va considérer la partition $P_q = \{D_j \mid j \in J_q\}$ de I engendrée par la question q : D_j est l'ensemble des individus ayant adopté la modalité j de J_q , et l'on a $\text{Card } D_j = k(j)$.

Appliquant la formule de transition, on a :

$$G_\alpha(j) = \Sigma\{k(i,j)F_\alpha(i) \mid i \in I\}/(k(j)\sqrt{\lambda_\alpha})$$

$$= \Sigma\{F_\alpha(i) \mid i \in D_j\}/(\text{Card } D_j \sqrt{\lambda_\alpha})$$

Au coefficient $(1/\sqrt{\lambda_\alpha})$ près, $G_\alpha(j)$ est la moyenne des $F_\alpha(i)$ pour i appartenant à D_j.

F^I_α ayant pour variance λ_α , on en déduit que :

$$CR_\alpha(q) = \text{Variance interclasse de } F_\alpha/(C_Q \text{ Variance totale de } F_\alpha) \quad (61)$$

La valeur propre λ_α , qui est la somme des $CR_\alpha(q)$ s'interprète donc comme la moyenne des rapports variance interclasse/variance totale associés aux C_Q partitions P_q.

Si on considère maintenant le tableau de Burt B_{JJ}, le facteur G'^{J}_{α} associé au α-ème axe factoriel est égal à $\sqrt{\lambda_\alpha} G^J_\alpha$, tandis que F^I_α s'obtient en adjoignant en élément supplémentaire de B_{JJ} le tableau k_{IJ}. Il en résulte que la contribution $CR'_\alpha(q)$ de q à l'inertie $\lambda'_\alpha = \lambda^2_\alpha$ du α-ème facteur issu de B_{JJ}, contribution égale à $\lambda_\alpha CR(q)$, s'interprète comme la variance interclasse de F^I_α, associée à la partition P_q définie par q, divisée par C_Q. λ'_α s'interprète donc comme la moyenne des variances interclasses de F^I_α associées aux C_Q partitions P_q.

Si l'on considère enfin un sous-tableau $C_{LL'}$ de B_{JJ}, associé à deux parties K et K' de Q, on peut interpréter la contribution d'une question q à l'inertie d'un axe factoriel par des considérations analogues à celles développées dans le cas du tableau de Burt. Supposons pour fixer les idées que q appartienne à K. Si F^L_α est le α-ème facteur issu de $C_{LL'}$, et relatif à la valeur propre λ''_α, la contribution de q à λ''_α s'écrit :

$$CR''_\alpha(q) = \Sigma\{k(j)F^2_\alpha(j) \mid j \in J_q\}/(C_I C_K) \qquad (62)$$

Si F^I_α désigne le facteur obtenu en adjoignant le sous-tableau $k_{IL'}$ de k_{IJ} en élément supplémentaire de $C_{LL'}$, $F_\alpha(j)$ est la moyenne des $F_\alpha(i)$ pour i ayant adopté la modalité j. $CR''_\alpha(q)$ est donc au facteur $(1/C_K)$ près la variance interclasse de F^I_α, associée à la partition P_q, tandis que λ''_α est la moyenne de ces variances interclasses pour q décrivant K. On peut noter qu'ici, la variance de F^I_α ne s'exprime pas aisément en fonction de la valeur propre, comme c'était le cas pour le tableau disjonctif complet ou le tableau de Burt.

4.5.2 *Utilisation des contributions pour représenter une variable sur un axe factoriel et construire des aides à l'interprétation basées sur les variables*

Pour représenter chaque question q sur un axe factoriel α, on peut utiliser la contribution de q à l'inertie de cet axe, ou mieux la racine carrée de cette contribution, si l'on veut que dans le plan, ou dans l'espace E_k des k premiers facteurs, le carré de la distance de q à l'origine soit égale à la contribution de q sur ces k axes. Notons qu'avec cette représentation, si l'on se place dans l'espace E_k, toutes les questions sont situées dans le cône positif de cet espace. En particulier, sur tout plan factoriel, les questions se trouvent dans le premier quadrant.

On peut aussi rapporter cette contribution que nous noterons $C_\alpha(q)$ (suivant le tableau analysé, $C_\alpha(q)$ est égal à $CR_\alpha(q)$ ou $CR'_\alpha(q)$ ou $CR''_\alpha(q)$) à la valeur propre λ_α associée à l'axe α ; on obtient ainsi la contribution relative de q à l'inertie de l'axe α, ce qui est l'équivalent du "CTR" dans les sorties du programme BENTAB d'analyse des correspondances usuel. On peut également rapporter la contribution $C_\alpha(q)$ à la contribution totale CR(q) de q (CR(q) qui est la contribution de q à l'inertie totale, est la somme des $CR_\alpha(q)$, et sa valeur est donnée par les formules du § 4.3) ; on obtient alors la contribution relative de l'axe α à q, ce qui est l'équivalent du "COR" dans les sorties du programme BENTAB, mais cette quantité ne s'interprète pas *a priori* comme un cosinus carré.

Si l'on considère maintenant les k premiers axes factoriels, on peut calculer :

1) la qualité (QUAL) de la reconstitution de l'inertie de q par ces k axes, ce qui revient à faire la somme des contributions relatives (COR) de ces k axes à q.

2) l'inertie (INR) de q sur ces k axes, i.e. la somme des contributions de q à ces k axes, cette inertie étant rapportée à l'inertie totale en projection dans l'espace de ces k axes i.e. à la somme des k premières valeurs propres.

On peut donc pour chaque question q avoir une sortie analogue à celle que l'on trouve dans le BENTAB pour chaque modalité, à savoir : le poids de q, la qualité (QUAL) et l'inertie (INR) de q sur les axes factoriels gardés, puis pour chacun de ces axes la racine carrée de la contribution de q à l'axe α considéré et les contributions relatives COR et CTR de l'axe α à q et de q à cet axe. Ces sorties, simples à réaliser, devraient faciliter l'interprétation, en particulier dans le cas de questionnaires comportant un nombre élevé de questions.

Remarques

1) Dans le cas d'un tableau disjonctif complet, B. Escofier (R.S.A., Vol 27 n° 4 pp 37-47, 1979) propose de représenter une variable q sur l'axe α à l'aide de la quantité $D_\alpha(q) = CR_\alpha(q)C_Q/\sqrt{\text{Card}\, J_q - 1}$ (*) qui d'après (61) n'est rien d'autre au coefficient $1/\sqrt{\text{Card}\, J_q - 1}$ près que le rapport de la variance interclasse de F^I_α, associée à P_q, à la variance totale de F^I_α ; $D_\alpha(q)$ qui peut s'interpréter comme un cosinus diffère de la contribution relative de α à q qui vaut d'après (57) $CR_\alpha(q)C_Q$ / (Card J_q - 1).

2) Les formules donnant les contributions $CR'_\alpha(q)$, $CR''_\alpha(q)$ peuvent être utilisées dans le cas d'une question q dont les modalités ont été mises en éléments supplémentaires, car elles s'interprètent toujours à un facteur près, comme la variance interclasse du facteur associé à q : q ne contribue plus à la formation de l'axe, par contre l'axe permet de reconstituer une partie des liaisons de q avec les variables qui caractérisent q en tant qu'élément supplémentaire. Supposons pour fixer les idées que l'on analyse $C_{LL'}$, et que l'on adjoigne à $C_{LL'}$ les modalités de q en lignes supplémentaires (i.e. q est caractérisé par ses liaisons avec les questions de K') ; si l'on donne *(a posteriori)* à chaque modalité j de q un poids $f_j = k(j)/(C_I C_K) = p_j/C_K$ (où $k(j)$ est le nombre d'individus ayant adopté la modalité j, et p_j la proportion associée : f_j est le total de la ligne supplémentaire $\{C(j,j') | j' \in L'\}$ divisé par le total C du tableau $C_{LL'}$; dans

(*) *B. Escofier se place dans l'espace R^I, où elle considère le sous-espace W_q^- associé à la variable q (cf. § 3.2.1), et le sous-espace ΔF^I_α engendré par F^I_α. $D_\alpha(q)$ est alors le cosinus entre les opérateurs de projection sur W_q^- et ΔF^I_α, le produit scalaire entre deux opérateurs O_1 et O_2 étant défini par trace $(O_1 O_2)$.*

ce dernier total les $\{C(j,j')|j \in J_q\}$ n'interviennent évidemment pas) l'inertie $CR''(q)$ de l'ensemble des modalités de q est donnée par la première formule (59), bien que cette formule ait été établie en supposant que q était un élément actif ; la contribution $CR''_\alpha(q)$ de l'axe α à $CR''(q)$ est alors donnée par la formule (62) (où $F_\alpha(j)$ désigne la valeur du α-ème facteur de $C_{LL'}$, pour la modalité supplémentaire j de J_q) et s'interprète encore, comme on l'a déjà dit, comme la variance interclasse du facteur F^I_α (obtenu en adjoignant $k_{IL'}$, en élément supplémentaire de $C_{LL'}$) divisée par C_K. Soit A l'ensemble des facteurs non triviaux issus de $C_{LL'}$. Si les facteurs $\{F^{L'}_\alpha|\alpha \in A\}$ forment une base du sous-espace de $R^{L'}$ engendré par les fonctions de moyenne nulle sur chaque J_q, $(q' \in K')$, la somme des $CR''_\alpha(q)$ (pour $\alpha \in A$) est égale à $CR''(q)$. En général cette somme est plus petite que $CR''(q)$. On pourra juger de la qualité de la reconstitution de $CR''(q)$ par l'axe α à l'aide du rapport $COR_\alpha(q) = CR''_\alpha(q)/CR''(q)$.

4.6 *Représentation de variables quantitatives quelconques dans une analyse de correspondances multiples*

La procédure proposée au § 4.5.2 pour représenter une variable sur les axes factoriels issus d'une analyse de correspondances multiples n'est valable que si l'on a une variable qualitative, ou une variable quantitative découpée en classes. De plus on obtient dans chaque plan factoriel une représentation des variables située dans le premier quadrant, comme on l'a remarqué. C'est la raison pour laquelle nous présentons ici une méthode pour représenter une variable quantitative quelconque x_I.

Nous raisonnerons dans l'espace R_I muni de la métrique du χ^2 définie à partir du tableau disjonctif complet k_{IJ}, i.e. de la métrique usuelle au coefficient C_I près. Les espaces R_I et R^I étant isomorphes, nous les identifierons et poserons $x^I = \{x^i|i \in I\}$ avec $x^i = x_i$.

Soit $(\varphi^I_\alpha, \varphi^J_\alpha, \lambda_\alpha)$ un couple de facteurs associés de variance 1, issu de k_{IJ} et relatif à la valeur propre λ_α. Si x_I est une variable positive, on peut la représenter sur l'axe α, en adjoignant x_I en élément supplémentaire de k_{IJ}. La coordonnée de x_I sur cet axe est alors proportionnelle à la covariance entre φ^I_α et x^I.

Si x_I est une variable quelconque, on pourra la représenter sur l'axe α par la covariance entre φ^I_α et x^I, ce qui revient à centrer x^I $(\Sigma x_i = 0)$ et à placer la variable x^+_I, où $x^+_i = (1 + x_i)/2$ (cf. § 5 pour la justification de l'introduction de x^+_I) en élément supplémentaire de k_{IJ}. Si $G_\alpha(x^+_I)$ désigne la coordonnée de x^+_I sur l'axe α, on a, puisque φ^I_α est centrée et en appliquant la formule de transition :

$$G_\alpha(x^+_I) = \mathrm{Cov}(\varphi^I_\alpha, x^I) = \Sigma\{\varphi^i_\alpha x_i|i \in I\}/C_I$$
$$= (1/(C_I C_Q \sqrt{\lambda_\alpha}))\, \Sigma\{\varphi^j_\alpha k(i,j) x_i|i \in I,\ j \in J\} \qquad (63)$$

formule où l'on a tenu compte de ce que $k(i) = C_Q$.

Posons :

$$x_j = \Sigma\{k(i,j)\ x_i | i \in I\} = k(j)\bar{x}^j \quad ; \tag{64}$$

x_j est la somme des x_i tels que i a pris la modalité j et $\bar{x}^j$ la moyenne associée. Tenant compte de ce que $\varphi_\alpha^J = \{\varphi_\alpha^{Jq} | q \in Q\}$ est centré sur chaque J_q, on a :

$$G_\alpha(x^+_I) = (1/\sqrt{\lambda_\alpha})(\Sigma\{\mathrm{Cov}(\varphi_\alpha^{Jq}, \bar{x}^{Jq}) | q \in Q\}/C_Q) \tag{65}$$

Dans le cas du tableau de Burt B_{JJ}, on peut, si x_I est une variable positive mettre $x_J = \{x_j | j \in J\}$ en élément supplémentaire, auquel cas on obtient (au facteur $\sqrt{\lambda_\alpha}$ près) la même coordonnée qu'en mettant x_I en élément supplémentaire de k_{IJ}. Si x_I est une variable quelconque, on peut calculer la covariance entre $F_\alpha^I = \sqrt{\lambda_\alpha}\ \varphi_\alpha^I$ (qui est ici obtenu en mettant k_{IJ} en élément supplémentaire de B_{JJ}), et x^I, ce qui revient à mettre (x^I ayant été centré) le vecteur y_J défini par $y_j = k(j)(1 + \bar{x}^j)/2$ en élément supplémentaire de B_{JJ}. La coordonnée $G'_\alpha(y_J)$ ainsi obtenue est égale à $\sqrt{\lambda_\alpha}G_\alpha(x^+_I)$.

Dans le cas du sous-tableau de Burt $C_{LL'}$, croisant deux séries de questions K et K', on peut représenter x_I dans R_L ou $R_{L'}$. Supposons pour fixer les idées qu'on représente x_I dans R_L. On peut, si x_I est positive ajouter la colonne $x_L = \{x_j | j \in L\}$ en élément supplémentaire de $C_{LL'}$; on obtient alors à une constante près la covariance entre x^I et le facteur G_α^I calculé en mettant le sous-tableau k_{IL} en élément supplémentaire de $C_{LL'}$. On peut dans tous les cas, calculer la covariance entre x^I et G_α^I, ce qui revient à mettre $y_L = \{y_j | j \in L\}$ en élément supplémentaire de $C_{LL'}$. La coordonnée $G_\alpha(y_L)$ ainsi obtenue s'écrit, si $\varphi_\alpha^L = \{\varphi_\alpha^{Jq} | q \in K\}$ est le α-ème facteur non trivial sur L, de variance 1, issu de $C_{LL'}$:

$$G_\alpha(y_L) = \Sigma\{\mathrm{Cov}(\varphi_\alpha^{Jq}, \bar{x}^{Jq}) | q \in K\}/C_K \tag{66}$$

Remarques

1) Si on adopte un codage pondéré, la présence d'une modalité de J_q étant codée α_q, au lieu de 1, les résultats précédents sont encore valables, à condition de prendre dans les formules (65) et (66) la moyenne pondérée des covariances, avec les α_q comme coefficients de pondération.

2) Si I est muni d'une pondération p_I de masse totale 1, les résultats précédents restent valables, à condition de poser $x^i = x_i/p_i$, et $x^+_i = p_i(1 + x^i) = p_i + x_i$.

De façon plus générale, les résultats précédents (où $p_i = k_i/k$) s'étendent immédiatement au cas des tableaux k_{IJ}, B_{JJ}, $C_{LL'}$ considérés au § 3, à condition de prendre des moyennes de covariances pondérées par les coefficients c_q (définis par (30)).

3) On a intérêt dans les calculs précédents à supposer x^I de variance 1.

4) Au lieu de calculer des covariances pour représenter x_I , on peut aussi calculer les corrélations de x^I avec les facteurs considérés.

5) Avec le codage du § 5, une variable quantitative est représentée sur un axe α par la covariance entre cette variable et le facteur associé, ce qui correspond à la démarche adoptée dans ce § . Si l'on adopte la démarche du § 4.5.2, on obtient, à un coefficient multiplicatif près la valeur absolue de cette covariance.