

HOSSAM ALWARD

Un cas modèle pour l'analyse d'un ensemble redondant de variables découpées en classes

Les cahiers de l'analyse des données, tome 15, n° 4 (1990), p. 391-402

http://www.numdam.org/item?id=CAD_1990__15_4_391_0

© Les cahiers de l'analyse des données, Dunod, 1990, 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/>

UN CAS MODÈLE POUR L'ANALYSE D'UN ENSEMBLE REDONDANT DE VARIABLES DÉCOUPÉES EN CLASSES

[MOD. VAR. CLASS.]

ALWARD Hossam*

1 Analyse des questionnaires et redondance

Afin d'analyser les réponses fournies par un ensemble I de sujets à l'ensemble Q des questions d'un questionnaire clos (ou, plus généralement, les valeurs relevées sur un ensemble I d'objets pour un ensemble hétérogène Q de variables descriptives), il est commun de découper les variables en classes. Le procédé le plus simple est le codage disjonctif complet.

En bref, à chaque question q (ou variable q : dans la suite, on utilisera "question" et "variable" comme des synonymes pour soutenir l'intuition), on attribue un ensemble J_q de modalités: pour une question fermée, J_q sera l'ensemble des réponses possibles; pour une variable continue, ce sera une suite de sous-intervalles en lesquels on choisit de découper l'intervalle des valeurs possibles... On crée un tableau dont l'ensemble J des colonnes est la réunion des J_q :

$$J = \cup \{ J_q \mid q \in Q \};$$

et l'on pose, pour tout individu $i \in I$ et toute modalité $j \in J$:

$$k(i, j) = 1, \text{ si } i \text{ rentre dans la modalité } j; \text{ et zéro, sinon ;}$$

ainsi, la ligne afférente à tout individu i comporte, dans chaque bloc J_q de colonnes, une suite de zéros avec un seul chiffre 1.

L'intérêt pratique du codage disjonctif complet est incontestable. Mais, notamment dans le découpage de variables continues, on peut regretter l'imprécision du codage par classes; et, en général, les relations de proximité ou de similitude entre modalités de réponse à une même question ne sont pas directement prises en compte. C'est pourquoi on a conçu le codage flou, suivant

(*) Université Pierre et Marie Curie, 3-ème cycle; et جامعة بغداد.

lequel, dans chaque bloc J_q , un individu i a, généralement, des notes non nulles dans deux modalités successives: ainsi est exprimé mathématiquement le fait que i occupe une position intermédiaire. Selon M. MEZOUGH, il peut même être justifié d'étendre à 3 modalités l'ensemble des notes non nulles représentant la réponse d'un sujet i à une question q .

Cependant, les critiques que l'on vient de formuler se fondent sur la considération des variables prises isolément; or le codage n'est qu'une étape intermédiaire dans la réalisation d'une analyse de correspondance. Si, en général, les résultats de l'analyse sont bien plus satisfaisants qu'on ne l'attendrait *a priori*, c'est parce que le jeu concordant des variables restitue, au travers de l'analyse, une image fidèle des individus. Du fait de la redondance du questionnaire, chaque thème est présenté dans plusieurs questions: il se trouve ainsi recevoir de chaque individu plusieurs modalités de réponse, même si l'on n'a pas fait de codage flou. Et, de par leurs associations avec les modalités d'autres variables, les modalités d'une même variable q montrent entre elles des corrélations dont rien ne se voit si l'on considère isolément le bloc de colonnes J_q afférent à cette variable.

L'objet de la présente note est de montrer, sur des cas modèle très simples, comment joue, mathématiquement, l'effet de la redondance. On supposera qu'il y a un ensemble Q de $\text{card}Q=n$ questions qui, au fond, sont identiques entre elles; et que les ensembles J_q se correspondent deux à deux biunivoquement. Chaque individu i est caractérisé, relativement à cette question de fond, par une variable *latente* unique v ; les réponses de i aux n questions de l'ensemble Q sont comme des tirages aléatoires indépendants faits suivant une même loi ne dépendant que de v .

Pour plus de clarté, on considérera d'abord (§2) un cas précis très simple; avec un ensemble J_q de trois modalités; et deux valeurs possibles seulement pour l'état latent. Le tableau de BURT construit pour ce modèle rentre dans un schéma général dont l'analyse est toujours aisée; ce qui permet de traiter une grande variété de cas modèles; avec un ensemble quelconque d'états latents; et des variables codées suivant un même ensemble J_q ayant un nombre quelconque de modalités (§3); voire par codage flou (§4).

2 Un modèle simple de questionnaire redondant

Nous supposons qu'il y a n questions; et que chaque question q comporte 3 modalités notées, comme on le fait communément, $\{q_-, q_+, q_\approx\}$. L'état latent des sujets peut prendre deux valeurs $\{v_-, v_+\}$; il y a autant de sujets dans chaque état; un sujet dans l'état v_- fournit, équiprobablement, à chaque question q l'une des deux réponses $\{q_-, q_\approx\}$; de même, un sujet dans l'état v_+ fournit équiprobablement les réponses $\{q_+, q_\approx\}$.

Si l'on suppose qu'il y a un nombre très élevé de sujets, le jeu des probabilités produit un tableau de BURT qui diffère arbitrairement peu (quant à

2	0	0	1	1	0
0	4	0	1	2	1
0	0	2	0	1	1
-----			-----		
1	1	0	2	0	0
1	2	1	0	4	2
0	1	1	0	0	2

D	S	S	...
-----	-----	-----	-----
S	D	S	...
-----	-----	-----	-----
S	S	D	...
-----	-----	-----	-----
...

son profil; i.e., à une constante multiplicative près) du modèle donné ici explicitement pour le cas n=2; et schématisé pour n quelconque.

Expliquons ce tableau. On a des blocs diagonaux, notés |D|, et des blocs symétriques |S| extérieurs à la diagonale. On sait que chaque bloc diagonal donne les poids respectifs (tels qu'on les trouve sur la marge) des modalités de la question correspondante. Ici, la modalité moyenne, q≈, de toute question a pour poids relatif (1/2); car, à toute question, tout sujet, quel que soit son état, fournit, avec la probabilité (1/2), cette réponse moyenne. Les modalités extrêmes {q-, q+} se partagent le poids restant. Ceci suffit à expliquer la valeur du bloc |D|; à ceci près qu'on a choisi pour le total des poids la valeur 8, afin d'éviter d'écrire des fractions dans les tableaux |D| et |S|.

Pour expliquer le bloc extra-diagonal |S|, on tiendra compte séparément des sujets qui sont dans les états respectifs {v-, v+}. Considérons deux questions {q, q'}. Les sujets dont l'état latent est v- fournissent avec la même probabilité les quatre couples de réponses:

$$\{(q-, q'-); (q-, q'≈); (q≈, q'-); (q≈, q'≈)\};$$

de même, les sujets en l'état v+ fournissent les quatre couples:

$$\{(q≈, q'≈); (q≈, q'+); (q+, q'≈); (q+, q'+)\};$$

sur le tableau de BURT, cela fait deux carrés de 1; d'où, après addition, les valeurs portées dans |S|.

En vue de l'analyse, nous rappelons d'abord que le tableau de BURT B doit être regardé comme une mesure sur le produit J x J; et que les facteurs sont des fonctions sur J qui sont vecteurs propres de la transition associée à B; transition dont le tableau a pour lignes les profils des lignes de B:

$$\text{tableau de BURT: } B_{JJ} = \{b(j, j') \mid j, j' \in J\};$$

$$\text{transition associée: } B^J_J = \{B^j_j\} = \{b(j, j') / b(j) \mid j, j' \in J\};$$

(où b(j) désigne le total de la ligne j); d'où l'équation des facteurs:

$$\lambda \varphi^J = B^J_J \varphi^J; \quad \lambda \varphi^j = \sum \{B^j_{j'} \varphi^{j'} \mid j' \in J\};$$

dans cette formule, λ désigne la valeur propre pour le tableau disjonctif complet en (0, 1); ou encore la racine carrée de la valeur propre résultant de l'analyse de correspondance du tableau de BURT proprement dit.

Ceci dit, l'analyse du cas modèle considéré ici peut se faire en conjuguant des modèles généraux qui seront rappelés au §3. Mais nous croyons plus instructif de proposer directement les résultats de l'analyse en tenant compte des symétries du modèle. Afin de permettre au lecteur de vérifier les calculs, nous reproduisons les tableaux avec la présente page.

L'identité foncière de toutes les questions impose de chercher d'abord des facteurs prenant les mêmes valeurs pour toutes les modalités de même signe, indépendamment de la question particulière considérée:

$$\varphi = \{ \varphi_-, \varphi_+, \varphi_-, \varphi_+, \varphi_-, \varphi_+, \dots, \varphi_-, \varphi_+, \varphi_+ \} ;$$

et la symétrie entre $\{q_-, q_+\}$ conduit aux valeurs des facteurs (données d'abord sans normalisation):

$$\varphi_1 = \{ -1, 0, +1, -1, 0, +1, \dots, -1, 0, +1 \} ;$$

$$\varphi_2 = \{ +1, -1, +1, +1, -1, +1, \dots, +1, -1, +1 \} .$$

Pour appliquer à ces fonctions φ^J sur J la transition B^J associée au tableau de BURT, il suffit d'en combiner les valeurs avec les coefficients fournis par chaque ligne de ce tableau divisée par son total (qui est $4n$ pour une ligne q_- ou q_+ ; et $8n$ pour une ligne q_+). Il est équivalent de dire que, pour chaque bloc J_q de 3 lignes consécutives $\{q_-, q_+, q_+\}$, on applique, à la fonction $\{\varphi_-, \varphi_+, \varphi_+\}$, une fois la transition associée à la matrice diagonale $|D|$ (transition qui n'est autre que l'identité); et $(n-1)$ fois la transition associée à la matrice $|S|$; et qu'on effectue la moyenne.

Si l'on note $M = \{ -, =, + \}$ l'ensemble des 3 modalités typiques, le calcul ci-dessus peut s'écrire:

$$\mu \varphi^M = S^M_M \varphi^M ; \quad \varphi^M = D^M_M \varphi^M ; \quad \lambda \varphi^J = ((1 + (n-1) \mu) / n) \varphi^J = B^J_J \varphi^J;$$

dans cette formule, μ désigne la valeur propre afférente à la transition associée au petit tableau, 3×3 , $|S|$; et l'on calcule la valeur propre λ pour la transition associée au grand tableau B :

$$\mu_1 = 1/2 ; \lambda_1 = ((1 + ((n-1)/2)) / n) ; \quad \mu_2 = 0 ; \lambda_2 = ((n-1) / (2n)) ;$$

le fait que $\lambda_2 = 0$ n'est qu'une particularité propre au modèle choisi.

2	0	0	1	1	0
0	4	0	1	2	1
0	0	2	0	1	1
-----			-----		
1	1	0	2	0	0
1	2	1	0	4	2
0	1	1	0	0	2

D	S	S	...
-----			---
S	D	S	...
-----			---
S	S	D	...
-----			---
...

Les facteurs ϕ sont normalisés pour avoir variance λ , si on se réfère à l'analyse du tableau en (0,1); ou variance λ^2 si on est censé soumettre le tableau de BURT lui-même à l'analyse des correspondances. Quelle que soit la normalisation choisie, l'ensemble J se présente dans le plan (1,2) (des deux facteurs construits) comme un triangle isocèle dont la base {-, +} est parallèle à l'axe 1; et le sommet {=} est sur le demi-axe (F2<0). Toutes les modalités de même signe q-, q= ou q+ sont respectivement superposées en ces trois points {-, =,+}. Quand le nombre n des questions redondantes tend vers l'infini, les points {-, =, +} tendent à s'aligner sur l'axe 1 (parce que $\mu_2 = 0$).

Quant aux facteurs qui ne prennent pas les mêmes valeurs pour toutes les modalités de même signe, ils se divisent en deux familles. On a des facteurs triviaux, relatifs à la valeur propre nulle, qui sont les fonctions constantes sur chacun des Jq et de moyenne nulle sur J tout entier; ce qu'on peut écrire:

$$\phi = \{\phi(q), \phi(q), \phi(q), \phi(q'), \phi(q'), \phi(q'), \dots, \phi(q^n), \phi(q^n), \phi(q^n)\};$$

avec la condition: $\sum\{\phi(q) \mid q \in Q\} = 0$.

Il y a, d'autre part, des facteurs non triviaux (même s'ils sont sans intérêt pour l'interprétation), qui constituent des sous-espaces propres de dimension n-1 associés à chacun des facteurs issus de l'analyse de la correspondance |S|; et que l'on construit et étudie comme on l'a fait pour ϕ_1 et ϕ_2 . On pose:

$$\phi a = \{a(q)\phi-, a(q)\phi=, a(q)\phi+, \dots, a(q^n)\phi-, a(q^n)\phi=, a(q^n)\phi+\};$$

avec la condition $\sum\{a(q) \mid q \in Q\} = 0$; et, en notant $\phi^M = \{\phi-, \phi=, \phi+\}$, on a:

$$\mu \phi^M = S^M_M \phi^M; \quad \lambda \phi a^J = ((1 - \mu) / n) \phi a^J = B^J_J \phi a^J;$$

pour vérifier cette formule, il suffit de considérer le cas particulier où $a(q')=1$, $a(q^n)=-1$, les autres valeurs $a(q)$ étant nulles: en effet, le cas général d'une fonction $a(q)$ quelconque est engendré par combinaison linéaire de facteurs dont le support est restreint à la réunion de deux sous-ensembles Jq' et Jq''.

3 Modèle général de tableau disjonctif complet pour un questionnaire redondant

Pour engendrer un tableau de réponse en (0,1), nous reprenons, en termes généraux, ce qui a été fait au §2. En fait, l'analyse portera exclusivement sur le tableau de BURT, qui sera seul explicitement défini. Le questionnaire comprend un ensemble redondant, Q , de n questions dont chacune admet le même ensemble typique de modalités M : $Jq \approx M$. Le comportement de chaque sujet est régi par une variable d'état v , élément d'un ensemble, ou espace probabilisable, V . Sur l'ensemble I des sujets, la variable d'état a une loi de probabilité:

$$\pi_v = \{\pi_v \mid v \in V\}, \text{ où } \pi_v \text{ est la probabilité de } v;$$

les réponses d'un sujet i , qui est dans l'état v , aux n questions de l'ensemble Q sont des tirages aléatoires indépendants faits suivant une même loi ne dépendant que de v . De façon précise, on a une probabilité de transition de V vers M :

$p_M^v = \{p_m^v \mid m \in M, v \in V\}$, où p_m^v est la probabilité qu'un sujet qui est dans l'état v adopte la modalité m .

On a pour le tableau de BURT la structure du §2, avec des blocs diagonaux $|D|$ et des blocs extra-diagonaux $|S|$; soit, en normalisant à 1 la masse des blocs:

$$D = \sum \{ \pi_v p_m^v \delta_M^m \otimes \delta_M^m \mid v \in V, m \in M \};$$

$$S = \sum \{ \pi_v p_m^v p_{m'}^v \delta_M^m \otimes \delta_M^{m'} \mid v \in V, m \in M, m' \in M \};$$

formules où on a noté δ_M^m la masse unité en m , i.e. le vecteur dont les composantes sont toutes nulles, exceptée celle de rang m qui vaut 1.

Comme dans l'exemple du §2, l'analyse du tableau de BURT $J \times J$, fournit des facteurs répartis en trois groupes.

I) Facteurs prenant les mêmes valeurs pour toutes les modalités de même nom. Ces facteurs correspondent biunivoquement aux facteurs φ issus du tableau de correspondance $|S|$ défini ci-dessus. En faisant abstraction de la normalisation, on a:

$$\mu \varphi^M = S_M^M \varphi^M; \quad \varphi^J = \{\varphi^M, \varphi^M, \dots, \varphi^M\};$$

$$\lambda \varphi^J = ((1 + (n-1) \mu) / n) \varphi^J = B_J^J \varphi^J;$$

on notera que la valeur propre λ est supérieure à la valeur propre μ correspondante; et, de plus, supérieure ou égale à $(1/n)$. Ceci résulte du fait que

les μ sont ≥ 0 , parce que $|S|$ est combinaison linéaire positive de carrés tensoriels:

$$|S| = \sum \{ \pi_v p_M^v \otimes p_M^v \mid v \in V \}.$$

II) Facteurs non triviaux ne prenant pas les mêmes valeurs pour toutes les modalités de même nom. Ces facteurs engendrent $(\text{card}M-1)$ sous-espaces propres de dimension $(n-1)$, dont chacun correspond à un facteur φ issu de l'analyse de $|S|$. De façon précise on a:

$$\varphi^J = \{ a(q)\varphi^M, \dots, a(q^n)\varphi^M \} \quad ; \quad \text{où} : \sum \{ a(q) \mid q \in Q \} = 0 ;$$

$$\mu \varphi^M = S^M_M \varphi^M \quad ; \quad \lambda \varphi^J = ((1 - \mu) / n) \varphi^J = B^J_J \varphi^J .$$

On notera que la valeur propre λ est comprise entre 0 et $(1/n)$; et, par conséquent, inférieure ou égale aux valeurs propres obtenues avec les facteurs de I; l'égalité n'étant réalisée (avec $1/n$) que si $|S|$ admet la valeur propre 0 (cf. cas de dégénérescence du §2). L'analyse du tableau de Burt fournit donc d'abord les $(\text{card}M-1)$ facteurs du groupe I; lesquels, du point de vue de l'interprétation, sont précisément ceux qui rendent compte de la structure redondante du modèle.

III) Facteurs triviaux, relatifs à la valeur propre nulle, qui sont les fonctions constantes sur chacun des J_q et de moyenne nulle sur J tout entier. Ces facteurs engendrent un sous-espace propre de dimension $(n-1)$; et l'on peut écrire:

$$\varphi^J = \{ \varphi(q)\delta^M, \dots, \varphi(q^n)\delta^M \} \quad ; \quad \text{où} : \sum \{ \varphi(q) \mid q \in Q \} = 0 ; \text{ et } \delta^M \text{ désigne la fonction constante et égale à 1 sur } M.$$

Pour démontrer les résultats que nous venons d'énumérer, on peut procéder par calcul direct, en suivant la même voie qu'au §2. On peut aussi invoquer des modèles généraux déjà considérés dans la littérature: d'une part dans [BLOCS TENS.], in *Pratique de l'Analyse des Données*, Vol 2, pp. 227sqq.; d'autre part dans [Corr. Sym.], in *L'Analyse Des Données*, Tome 2, B n°9 §3, pp. 234 sqq.

En effet, si, dans le schéma du tableau de BURT B, on remplace les blocs diagonaux $|D|$ par zéro, on obtient un tableau T qui, étant produit tensoriel du bloc $|S|$ par le tableau $(n \times n)$ rempli de 1 à l'exclusion de la diagonale qui est nulle, rentre dans le cas de [BLOCS TENS.]. Cependant B apparaît comme la somme du tableau T et d'un tableau diagonal qui (à un coefficient de proportionnalité près) n'est autre que celui obtenu en mettant sur la diagonale du carré $(n \times n)$ la loi marginale de T: on est donc dans le cas de [Corr. Sym.], §3.

On remarquera que la recherche explicite des vecteurs propres de la transition associée à B a montré que tous les facteurs issus de B sont directs.

4 Modèle général redondant avec codage flou

Comme au §3, le questionnaire consiste en un ensemble redondant, Q , de n questions dont chacune admet le même ensemble typique de modalités M : $Jq \approx M$. De plus, chaque question est recodée, par codage flou, sur un même ensemble X de modalités. En termes mathématiques, ce codage, qui répartit chaque modalité m de M suivant une distribution sur X dont la masse totale est 1, peut être regardé comme une transition probabiliste, c_X^M , de M vers X :

$c_X^M = \{ c_x^m \mid x \in X, m \in M \}$; où $c_x^m = \{ c_x^m \mid x \in X \}$ désigne la suite des valeurs c_x^m attribuées à la modalité m recodée sur X ; (ces valeurs étant, dans la pratique, toutes nulles exceptées deux ou trois d'entre elles).

On a, d'une part, un ensemble J de modalités de réponses primaires, qui est une suite de n exemplaires de M ; et, d'autre part, un ensemble, noté Z , de modalités des réponses recodées; Z étant une suite de n exemplaires de X . La loi des réponses primaires n'est pas changée: elle est définie par la loi π_V de la variable d'état, et par la transition p_M^V , de V vers M .

Pour les réponses recodées, on ne peut parler de tableau de BURT au sens strict de tableau de cooccurrence; mais il est commun d'en généraliser la définition en posant:

$$B_{ZZ} = \{ B(z, z') \mid z \in Z, z' \in Z \}; \quad B(z, z') = \sum \{ k(i, z) k(i, z') \mid i \in I \}.$$

Le tableau B , ainsi défini, a une structure de bloc analogue à celle déjà vue; les blocs sont n^2 carrés $X \times X$; il faut distinguer entre blocs diagonaux, D , et blocs extra-diagonaux, S ; mais, présentement les blocs D (situés sur la diagonale de B) ne sont pas réduits à leur diagonale. De façon précise, on a (en normalisant à 1, comme au §3, la masse totale de chaque bloc):

$$D = \sum \{ \pi_v p_m^v c_X^m \otimes c_X^m \mid v \in V, m \in M \};$$

$$S = \sum \{ \pi_v p_m^v p_{m'}^v c_X^m \otimes c_X^{m'} \mid v \in V, m \in M, m' \in M \};$$

formules où les δ_M^m sont remplacés par les profils de codage c_X^m .

En introduisant la transition τ_X^V , composée de c_X^M et p_M^V , on peut écrire sous une autre forme la matrice $|S|$:

$$\tau_X^V = \{ \tau_x^v \mid x \in X, v \in V \}; \quad \text{où: } \tau_x^v = \sum \{ c_x^m p_m^v \mid m \in M \};$$

$$S_{XX} = \sum \{ \pi_v \tau_X^v \otimes \tau_X^v \mid v \in V \}.$$

Comme au §3, nous réduirons complètement l'analyse de B à celles des matrices |S| et |D| ou, plus précisément, d'autres tableaux carrés $X \times X$ qui s'en déduisent par un calcul simple. On retrouvera d'abord aisément les facteurs de classe I (qui seuls comptent pour l'interprétation); ainsi que les facteurs triviaux (III) qui subsistent ici tels quels. La recherche des facteurs de classe II se fera ensuite, en tenant compte des facteurs déjà trouvés.

I) Facteurs prenant les mêmes valeurs pour toutes les modalités de même nom. Ces facteurs correspondent biunivoquement aux facteurs φ^X issus du tableau de correspondance |Δ| défini ci-dessous:

$$\Delta_{XX} = (1/n) (D_{XX} + (n-1) S_{XX}) ;$$

on a, en notant Δ^X_X la transition associée à Δ :

$$\lambda \varphi^X = \Delta^X_X \varphi^X \quad ; \quad \varphi^Z = \{\varphi^X, \varphi^X, \dots, \varphi^X\} \quad ; \quad \lambda \varphi^Z = B^Z_Z \varphi^Z ;$$

Pour démontrer cette formule il suffit de remarquer que le profil d'un bloc de lignes Xq du tableau Δ est constitué d'une suite de n blocs dont (n-1) sont égaux au profil de |S| divisé par n; et un seul, au profil de |D| divisé par n; ce qui, au total, puisque φ^Z est une suite de segments identiques à φ^X , donne $\Delta^X_X \varphi^X$ pour le bloc de lignes Xq de $B^Z_Z \varphi^Z$.

III) Les facteurs triviaux, relatifs à la valeur propre nulle, s'écrivent ici:

$\varphi^Z = \{\varphi(q)\delta^X, \dots, \varphi(q^n)\delta^X\} \quad ; \quad \text{où} : \sum\{\varphi(q) \mid q \in Q\} = 0 \quad ; \quad \text{et } \delta^X$
désigne la fonction constante et égale à 1 sur X.

II) Quant aux facteurs restants, nous savons d'abord qu'ils doivent être orthogonaux sur Z aux facteurs déjà trouvés; c'est-à-dire qu'ils sont de moyenne nulle sur chaque bloc Xq (cf. III), et orthogonaux à toute fonction sur Z *périodique*, i.e. se reproduisant identique à elle-même sur tous les blocs Xq (cf. I).

Cette dernière propriété implique qu'est nulle (pour un tel facteur φ^Z) la fonction μ^X , obtenue en faisant la somme (ou la moyenne) des restrictions de φ^Z aux n segments Xq dont est constitué l'ensemble Z: en effet, si μ^X était non nulle, la fonction μ^Z , coïncidant avec μ^X sur chacun des Xq, serait une fonction périodique non orthogonale à φ^Z .

De la nullité de μ^X , il résulte que l'effet sur φ^Z de la transition B^Z_Z , associée au tableau de BURT B, est le même que celui de la transition associée à la matrice

V formée de n^2 blocs carrés $X \times X$, avec des blocs extra-diagonaux nuls et des blocs diagonaux de la forme $|V| = (n |F| + |D| - |S|)$; où $|F|$ désigne la matrice $X \times X$, carré tensoriel de la loi marginale f_X (de masse totale 1) commune à $|D|$ et $|S|$:

$$F_{XX} = f_X \otimes f_X ;$$

dans cette formule, $|F|$ est introduit afin que, dans la matrice V, chaque ligne ait même masse que dans B; et le terme en $|F|$ est d'autre part sans effet sur le facteur cherché ϕ^Z parce que celui-ci a moyenne nulle sur tout bloc Xq .

Pour montrer que les deux transitions V^Z_Z et B^Z_Z ont même effet sur ϕ^Z , il suffit de remarquer qu'à la correction près introduite par $|F|$ pour que la masse soit conservée, V s'obtient à partir de B en retranchant la matrice S, dont tous

$$\begin{array}{c}
 \mathbf{V} = \begin{array}{|c|c|c|c|} \hline v & 0 & 0 & \dots \\ \hline 0 & v & 0 & \dots \\ \hline 0 & 0 & v & \dots \\ \hline \dots & \dots & \dots & \dots \\ \hline \end{array}
 \end{array}
 \qquad
 \begin{array}{c}
 \mathbf{S} = \begin{array}{|c|c|c|c|} \hline s & s & s & \dots \\ \hline s & s & s & \dots \\ \hline s & s & s & \dots \\ \hline \dots & \dots & \dots & \dots \\ \hline \end{array}
 \end{array}$$

les blocs sont égaux à $|S|$. Or S, appliquée à ϕ^Z donne 0, du fait de la nullité de μ^X , remarquée ci-dessus; car, en bref, combiner les valeurs de ϕ^Z avec les coefficients fournis par une ligne de S, équivaut à faire agir la ligne correspondante de $|S|$ sur la fonction μ^X , somme des segments consécutifs dont se compose ϕ^Z .

On peut maintenant réduire la recherche des facteurs du groupe II à l'analyse de la correspondance $|V|$. À la normalisation près, on procède, avec les facteurs ϕ^X issus de $|V|$, comme on l'a fait au §3 avec les facteurs issus de $|S|$. De façon précise on pose:

$$\begin{aligned}
 \phi^Z &= \{a(q)\phi^X, \dots, a(q^n)\phi^X\} \quad ; \quad \text{où} : \sum \{a(q) \mid q \in Q\} = 0 ; \\
 \lambda \phi^X &= F^X_X \phi^X \quad ; \quad \lambda \phi^Z = B^Z_Z \phi^Z .
 \end{aligned}$$

Ici, la valeur propre λ est la même pour $|V|$ et B, parce qu'on a pris soin d'introduire le terme en $|F|$. On peut vérifier que le présent calcul fournit, pour le modèle du §3, les résultats déjà obtenus. En effet, si $|D|$ est diagonal, les

vecteurs propres sont les mêmes pour $|S|$, $|D|-|S|$ ou $|V|$; on est dans le cas traité dans [Corr. Sym.] §3), cité plus haut: $|V|$ est combinaison linéaire de $|S|$, du carré $|F|$ de sa loi marginale f_X et de $|D|$ qui porte sur sa diagonale la loi f_X ; et la valeur propre λ , afférente à $|V|$, se déduit de μ , afférente à $|S|$ par la formule $\lambda = ((1 - \mu) / n)$.

Dans la cas général du §4, on peut remarquer que les facteurs de $|V|$ ne dépendent pas de la valeur de n . De plus, pour n donné, comme au §3, la valeur propre λ ne peut dépasser $(1/n)$; car, en bref, quelle que soit la fonction φ^X , de moyenne nulle et variance 1 (pour la loi f_X), on a:

$$0 \leq \sum \{S_{xx'} \varphi^x \varphi^{x'} \mid x \in X, x' \in X\} \leq 1 \quad ;$$

$$0 \leq \sum \{D_{xx'} \varphi^x \varphi^{x'} \mid x \in X, x' \in X\} \leq 1 \quad ;$$

d'où il résulte, par différence, que, pour la tableau $|V/n|$ de masse totale 1, on a:

$$(-1/n) \leq \sum \{(1/n) V_{xx'} \varphi^x \varphi^{x'} \mid x \in X, x' \in X\} \leq (1/n) \quad ;$$

plus précisément, cette somme est positive parce qu'elle donne une valeur propre de la transition associée à B ; valeur propre qui n'est autre que celle du tableau de correspondance des notes recodées.

5 Perspectives et conclusions

Nous avons dit que, de par leurs associations avec les modalités d'autres variables, les modalités d'une même variable q montrent entre elles des corrélations dont rien ne se voit si l'on considère isolément le bloc de colonnes J_q afférent à cette variable.

Pour confirmer cette assertion, nous avons, dans le présent travail, analysé un modèle de stricte redondance; où il n'y a véritablement qu'une seule question plusieurs fois présentée mais toujours avec le même découpage en modalités et le même codage; la diversité des réponses fournies par un même sujet ne résultant que d'un comportement aléatoire.

On peut imaginer encore qu'une question de base unique soit plusieurs fois présentée avec des découpages différents: en ce cas, même avec des réponses non aléatoires, on aura des corrélations révélant la structure de l'ensemble des modalités. Le modèle le plus simple serait celui d'une variable continue sous-jacente unique dont chaque question réalise le codage suivant un découpage différent.

Quoi qu'il en soit de ces modèles, c'est, assurément, sous une forme plus souple que la redondance s'introduit dans les données naturelles et en favorise

l'analyse. Mais c'est toujours par référence à la redondance qu'on doit, selon nous, concevoir les formes de codage et en critiquer les effets.

Références bibliographiques

Les notations du calcul tensoriel des transitions, utilisé ici, sont expliquées, notamment, dans le traité sur *L'Analyse Des Données*, Tome II B n°1, [Note Lim.]; (1-ère éd. 1973). Ce même Tome traite, dans [Corr. Sym.], §3, B n°9, du "polygone convexe de lois associé à une loi symétrique". Les produits tensoriels de correspondances sont étudiés dans [BLOCS TENS.], in *Pratique de L'Analyse Des Données*, Vol 2, IV, n°11.

Sur le codage linéaire par morceaux, on trouvera un exposé général avec bibliographie dans [CODAGE LIN.], in *CAD*, Vol XIV, n°2, pp. 203-210; (1989). Nous citerons d'autre part la thèse de M. MEZOUGHİ, *Influence du codage suivant l'équation personnelle et conservation de l'ordre des modalités...*; Paris, (1990).