REVUE DE STATISTIQUE APPLIQUÉE

R. ABDESSELAM
Y. SCHEKTMAN

Une analyse factorielle de l'association dissymétrique entre deux variables qualitatives

Revue de statistique appliquée, tome 44, n° 2 (1996), p. 5-34
<http://www.numdam.org/item?id=RSA_1996__44_2_5_0>

Rev. Statistique Appliquée, 1996, **XLIV** (2), 5-34

UNE ANALYSE FACTORIELLE DE L'ASSOCIATION DISSYMÉTRIQUE ENTRE DEUX VARIABLES QUALITATIVES

* R. Abdesselam, ** Y. Schektman

* *Laboratoire Lemme DIEM – Université Paul Sabatier Toulouse (France)*
** *Maison de la Recherche – Université Toulouse le Mirail (France)*

RÉSUMÉ

Les «coefficients d'association relationnels» dissymétriques sont présentés. Les mesures de ces coefficients sont exprimées en termes d'inerties dans l'espace des individus muni d'un «produit scalaire relationnel». Cette vision géométrique et mécanique des associations permet de synthétiser, d'étendre, les méthodes classiques d'analyse des données, basées sur la recherche des moments principaux du nuage des individus, et d'en proposer de nouvelles. Pour analyser l'association dissymétrique entre deux variables qualitatives, détermination des moments principaux et représentations graphiques, nous proposons une analyse factorielle applicable à une famille de coefficients d'association dissymétriques, incluant le tau de Goodman-Kruskal et ses extensions : tau pondéré ou équipondéré. Cette analyse affine les résultats de l'analyse proposée par D'Ambra et Lauro et offre un champ d'application plus vaste. De plus, il est intéressant de noter que l'on retrouve l'Analyse Factorielle des Correspondances de J.P. Benzecri, si on applique cette analyse au carré moyen de contingence de Pearson. Un exemple sur données simulées est présenté.

Mots-clés : *Analyse factorielle, distance relationnelle, coefficient d'association symétrique et dissymétrique, carré moyen de contingence de Pearson, rapport de corrélation, corrélations canoniques, tau de Goodman- Kruskal, coefficient de Stewart-Love.*

ABSTRACT

Dissymmetrical «relational association coefficients» are described. Measurements of these coefficients are expressed as inertia in the individual-space with a «relational inner product». This geometrical and mechanical point of view on associations analysis leads to a synthesis, an extension, of classical data analysis methods, based on the research of principal axes of a configuration of points, and to new methods. We propose a factorial analysis fitted to a family of dissymmetrical associations coefficients between two qualitative variables, including the Goodman-Kruskal tau and its weighted or equally weighted extensions. This analysis improves results proposed by D'Ambra and Lauro and gives to you more possibilities. Besides, it is interesting to note that Factorial Correspondence Analysis is obtained by applying

proposed analysis to the symmetrical mean square contingency Pearson association coefficient. One example is described.

Keywords : *Factorial analysis, relational distance, symmetrical and dissymmetrical association coefficients, mean square contingency Pearson coefficient, correlation ratio, canonical correlations, Goodman-Kruskal tau, Stewart-Love coefficient.*

1. Introduction – Notations

On se propose d'analyser l'association dissymétrique «x explique y», où x et y sont deux variables qualitatives ayant respectivement p et q modalités. Pour cela, on définit une analyse factorielle qui décrit la «forme de l'association», mesurée par un coefficient choisi dans une famille infinie de coefficients d'association dissymétriques contenant notamment le tau de Goodman-Kruskal [12], noté τ. Cette analyse affine les résultats de l'analyse non symétrique des correspondances proposée par Lauro et D'Ambra [10,14], et son domaine d'application est plus vaste. De plus, elle admet, comme cas particulier, l'analyse de l'association symétrique mesurée par le carré moyen de contingence de Pearson, noté ϕ^2 : elle est alors équivalente à l'Analyse Factorielle des Correspondances (AFC) [4].

$E_y = \mathbb{R}^q$ étant le sous-espace des individus, associé par dualité aux variables indicatrices centrées des modalités de y, notées $\{y^k; k = 1, q\}$, l'ensemble des coefficients d'association dissymétriques étudiés est en correspondance biunivoque avec l'ensemble des produits scalaires de $\mathbb{R}^q$, c'est-à-dire l'ensemble des matrices symétriques définies positives d'ordre q. L'expression générale de ces coefficients est donnée dans le paragraphe 2.3 : elle mesure l'inertie d'un nuage de l'espace des individus $E = E_x \oplus E_y$ muni d'un produit scalaire relationnel [16], où $E_x = \mathbb{R}^p$ est le sous-espace des individus associé par dualité aux variables indicatrices centrées $\{x^j; j = 1, p\}$ des modalités de x. Nous présentons, dans le paragraphe 2, trois coefficients : le tau de Goodman-Kruskal et ses dérivés.

On rappelle qu'un produit scalaire relationnel dans E fournit une traduction, en termes géométriques, de la structure des associations observées entre les variables $\{x^j\}$ et $\{y^k\}$. Les produits scalaires relationnels sont présentés succinctement dans le paragraphe 2.3, ils sont à l'origine des méthodes proposées notamment dans [1,2,5,8,11,13,20].

La définition et les propriétés de l'analyse factorielle proposée sont présentées dans le paragraphe 3 et une application sur données simulées est commentée dans le paragraphe 4.

On utilise les notations suivantes :

- $X(n,p)$ et $Y(n,q)$, où n est le nombre d'individus, étant respectivement les matrices des valeurs des variables $\{x^j\}$ et $\{y^k\}$, on pose $V_x = {}^tXDX$, $V_y = {}^tYDY$ et $V_{xy} = {}^tXDY$, où $D = (1/n)I_n$ est la matrice diagonale des poids des n individus et I_n la matrice unité d'ordre n,
- M est la matrice du produit scalaire de référence dans E et M_y[resp. M_x] est celle du produit scalaire de l'espace, noté E_y[resp. E_x], isomorphe au sous-espace de même nom, via l'injection canonique notée Iny[resp. Inx],

- $N_y = \{y_i \in E_y; i = 1,n\}$[resp. $N_x = \{x_i \in E_x; i = 1,n\}$] est le nuage des individus associé à Y [resp. X], il est constitué d'au plus q[resp. p] points distincts qui représentent aussi les modalités de y[resp. x],
- $g(y/x^j)$ est le q-uple $[(njk/nj.) - (n.k/n);\ k = 1,q]$, où $nj. = \sum_{k=1}^{q} njk$ et $n.k = \sum_{j=1}^{p} njk$, njk étant le nombre d'individus possédant les modalités j de x et k de y,
- $N_g(y/x) = \{g(y/x^j); j = 1,p\}$ est le nuage des p q-uples $\{g(y/x^j)\}$ dans $\mathbb{R}^q$,
- D_y[resp. D_x] est la matrice diagonale (des poids) définie par $[D_y]_{kk} = n.k/n$ pour tout k [resp. $[D_x]_{jj} = nj./n$ pour tout j], où de façon générale $[A]_{kl}$ désigne l'élément de la $k^{\text{ème}}$ ligne et de la $l^{\text{ème}}$ colonne de A,
- D_{1/σ_y^2} est la matrice diagonale définie par $[D_{1/\sigma_y^2}]_{kk} = 1/\text{var}\,(y^k)$ pour tout k, que l'on écrira encore $D_{1/\sigma_y^2} = \text{Diag}\,[1/\text{var}\,(y^k)]$, où $\text{var}\,(y^k)$ est la variance de y^k,
- $\chi_y^2 = D_y^{-1}$[resp. $\chi_x^2 = D_x^{-1}$] est la matrice de la distance du khi-deux dans $\mathbb{R}^q$[resp. $\mathbb{R}^p$].

De plus, pour ne pas multiplier les notations, on utilisera les mêmes symboles pour les matrices et les applications qui leur correspondent dans le schéma de dualité [6] : ainsi le symbole χ_y^2 servira aussi à nommer le produit scalaire ou l'isomorphisme associé. De même, le même symbole désignera un vecteur et la matrice unicolonne de ses coordonnées dans une base donnée. Ces différents langages seront utilisés indifféremment selon leur commodité.

Pour simplifier les écritures on identifiera les vecteurs des espaces E_x et E_y avec leur image par injection canonique.

On notera que $N_g(y/x)$ est le nuage des p lois conditionnelles centrées de y sachant $x^j (j = 1,p)$, et que son centre de gravité, relativement à la matrice des poids D_x, est confondu avec l'origine de $\mathbb{R}^q$.

2. Expressions de quelques coefficients d'association dissymétriques

Trois approches sont utilisées pour donner les définitions du τ et de ses dérivés : l'approche géométrique permet d'introduire naturellement les dérivés du τ, les approches algébrique et surtout mécanique apportant les bases de la méthode que nous proposons.

2.1. Expressions géométriques dans l'espace des variables – Interprétation statistique

$Q_x(y^k)$ étant la projection orthogonale de $y^k \in \mathbb{R}^n$ sur le sous-espace engendré par les $\{x^j; j = 1,p\}$, vu que

$$||Q_x(y^k)||^2 = \sum_{j=1}^{p} (nj./n)[(njk/nj.) - (n.k/n)]^2 \qquad (1)$$

et $||y^k||^2 = n.k(n - n.k)/n^2$,

le tau de Goodman-Kruskal peut s'écrire [19] :

$$\bullet\ \tau(x;y) = \sum_{k=1}^{q} ||Q_x(y^k)||^2 / \sum_{l=1}^{q} ||y^l||^2. \tag{2}$$

Il est alors raisonnable de proposer les définitions suivantes.

Définition 1 :

• tau pondéré :

$$\tau_p(x;y) = \sum_{k=1}^{q} n.k||Q_x(y^k)||^2 / \sum_{l=1}^{q} n.l||y^l||^2 \tag{3}$$

• tau équipondéré :

$$\tau_{ep}(x;y) = (1/q) \sum_{k=1}^{q} [||Q_x(y^k)||^2 / ||y^k||^2]. \tag{4}$$

Les appellations utilisées, notamment pour le τ_{ep}, sont justifiées par les expressions suivantes qui se déduisent immédiatement des expressions (2), (3) et (4),

$$\bullet\ \tau(x;y) = \sum_{k=1}^{q} [\text{var}\,(y^k) / \sum_{l=1}^{q} \text{var}\,(y^l)]\ \ r^2[y^k;\ \{x^j\}]$$

$$\bullet\ \tau_p(x;y) = \sum_{k=1}^{q} [n.k\ \text{var}\,(y^k) / \sum_{l=1}^{q} n.l\ \text{var}\,(y^l)] r^2[y^k; \{x^j\}]$$

$$\bullet\ \tau_{ep}(x;y) = (1/q) \sum_{k=1}^{q} r^2[y^k; \{x^j\}].$$

où $r[y^k;\ \{x^j\}]$ est le coefficient de corrélation multiple entre y^k et les $\{x^j; j = 1, p\}$.

Remarque 1 : Si on désire utiliser un coefficient d'association dissymétrique indépendant des variances des $\{y^k\}$, alors on pourra prendre le τ_{ep}; dans le cas contraire on pourra prendre le τ ou le τ_p. L'utilité du τ_p est mise en évidence dans le paragraphe 4.

2.2. Expressions algébriques

Pour mesurer l'association dissymétrique entre x et y, il est raisonnable de proposer la formule suivante :

$$\bullet\ F(x;y) = \sum_{j=1}^{p} (nj./n) \sum_{k=1}^{q} m_k(y)[(njk/nj.) - (n.k/n)]^2 \tag{5}$$

avec $m_k(y) \geqslant 0$.

De l'expression (1) de $||Q_x(y^k)||^2$, on déduit que

$$F(x;y) = \sum_{k=1}^{q} m_k(y)||Q_x(y^k)||^2;$$

pour établir la propriété suivante, il suffira donc d'utiliser (2), (3), (4) et l'expression (1) de $||y^k||^2$.

Propriété 1 :

- $F(x;y) = \tau(x;y)$ si $\forall k \;\; m_k(y) = n^2 / \sum_{l=1}^{q} n.l(n - n.l)$
- $F(x;y) = \tau_p(x;y)$ si $\forall k \;\; m_k(y) = n^2 \; n.k / \sum_{l=1}^{q} n_{.l}^2(n - n.l)$
- $F(x;y) = \tau_{ep}(x;y)$ si $\forall k \;\; m_k(y) = n^2/q \; n.k(n - n.k)$.

Remarques 2 :

a) Dans $\mathbb{R}^q$ muni de la distance $M_y = \text{Diag}\,[m_k(y)]$, relativement à D_x, les trois coefficients précédents ont pour valeur l'inertie, notée I, de $N_g(y/x)$ par rapport à l'origine, ou encore par rapport à son centre de gravité. En effet (5) s'écrit :

- $F(x;y) = \sum_{j=1}^{p} (nj./n)||g(y/x^j)||^2 = I[N_g(y/x)]$.

b) Si $\forall k \;\; m_k(y) = n/n.k$, c'est-à-dire $M_y = \chi_y^2$,

alors $F(x;y) = \phi^2(x,y) = [F(y;x) \text{ si } M_x = \chi_x^2]$.

c) Si $\forall k \;\; n.k = n/q$, c'est-à-dire si les variances des variables $\{y^k; k = 1, q\}$ sont égales, alors $\tau(x;y) = \tau_{ep}(x;y) = \tau_p(x;y) = \phi^2(x,y)/(q-1)$.

2.3. Expressions mécaniques dans l'espace des individus muni d'une géométrie relationnelle

La définition des «coefficients d'association relationnels» [17,1] est basée sur la notion de «produit scalaire relationnel».

On pose $M_{xy} = {}^t\text{Inx}\; M \;\text{Iny}$ où ${}^t\text{Inx}$ est la transposée de Inx.

On rappelle que dans l'espace des individus $E = E_x \oplus E_y$, le produit scalaire M est dit relationnel relativement aux variables $\{x^j\}$ et $\{y^k\}$, et est noté $R[M_x, M_y]$, si et seulement si :

$$M_{xy} = M_x[(V_x M_x)^{1/2}]^+ V_{xy} M_y[(V_y M_y)^{1/2}]^+ \qquad (6)$$

où $[(V_x M_x)^{1/2}]^+$ est l'inverse généralisée de Moore-Penrose, pondérée par M_x, de $(V_x M_x)^{1/2}$.

L'introduction d'inverses généralisées est une conséquence de la singularité des matrices V_x et V_y , puisque rang $[V_x] = p - 1$ et rang $[V_y] = q - 1$. On montre [22] que les propriétés de $R[M_x, M_y]$ pour des matrices V_x et/ou V_y singulières sont presque toujours identiques à celles établies pour des matrices régulières. On effectue, en annexe, une présentation succincte des inverses généralisées pondérées de Moore-Penrose et on montre quelques propriétés utiles en Analyse de Données.

Remarques 3 :

a) Il existe de nombreuses propriétés caractéristiques des produits scalaires relationnels [8, 22]. Donnons à titre illustratif une condition nécessaire et suffisante de type géométrique [16] : soient $\{c_j(x); j = 1, p\}$[resp. $\{c_k(y); k = 1, q\}$] les vecteurs axiaux principaux du triplet (N_x, M_x, D)[resp. (N_y, M_y, D)] et $\{C^j(x); j = 1, p\}$[resp. $\{C^k(y); k = 1, q\}$] les composantes principales correspondantes, on a $M = R[M_x, M_y]$ est équivalent à

$$\cos[c_j(x), c_k(y)] = \begin{cases} \cos[C^j(x), C^k(y)] & \text{si } ||C^j(x)|| \neq 0 \text{ et } ||C^k(y)|| \neq 0 \\ 0 & \text{sinon.} \end{cases}$$

b) Il est parfois utile [poids de certains individus nuls, couples de variables (x^j, y^k) très corrélées] de pouvoir définir et utiliser des semi-produits scalaires relationnels : sous certaines conditions [21] M_x, M_y ou/et D peuvent être des semi-produits scalaires.

On note $I_x[N_y]$ l'inertie (par rapport à son centre de gravité) de la projection orthogonale de N_y sur E_x, V_x^+ l'inverse généralisée de Moore-Penrose de V_x (pondérée par M_x^{-1} et M_x) et P_x l'opérateur de projection orthogonale sur E_x.

Lemme 1 : Si $M = R[M_x, M_y]$ alors $I_x[N_y] = \sum_{k=1}^{q} \sum_{k'=1}^{q} [M_y]_{kk'} D[Q_x(y^k), Q_x(y^{k'})]$.

Preuve :

sachant que $Q_x = XV_x^+ {}^tXD$ (cf. annexe) et $P_x = \text{Inx}\, M_x^{-1}\, {}^t\text{Inx}\, M$, (7)

partant de $I_x[N_y] = \text{trace}\,[P_x \text{Iny}\, V_y\, {}^t\text{Iny}\, M]$,

en utilisant (6), les décompositions spectrales de $[(V_xM_x)^{1/2}]^+$ et $[(V_yM_y)^{1/2}]^+$, et la relation $(V_xM_x)^+ = M_x^{-1}V_x^+$ données en annexe, on a :

$$\begin{aligned} I_x[N_y] &= \text{trace}[M_x^{-1}M_{xy}V_yM_{yx}] \\ &= \text{trace}[[(V_xM_x)^{1/2}]^+V_{xy}M_y[(V_yM_y)^{1/2}]^+V_yM_{yx}] \\ &= \text{trace}[V_{xy}M_yV_{yx}V_x^+] = \text{trace}[M_y\, {}^tYDQ_xY] \\ &= \sum_{k=1}^{q} \sum_{k'=1}^{q} [M_y]_{kk'} D[Q_x(y^k), Q_x(y^{k'})]. \end{aligned}$$

Remarque 4 : Il est naturel d'observer, dans le lemme 1, que $I_x[N_y]$ ne dépend pas du choix de M_x ; il sera souvent utile de choisir $M_x = \chi^2_x$, ce sera le cas dans la définition 3 du paragraphe 3.2.

Définition 2 : Sachant que $M = R[M_x, M_y]$, on appelle Coefficient d'Association Relationnel Dissymétrique entre les variables $\{x^j\}$ et $\{y^k\}$, relativement à M_y, la quantité :

$$\text{CARD}[x; y] = I_x[N_y]/I[N_y].$$

Propriété 2 :

$$\begin{aligned} \text{CARD}[x; y] &= \tau(x; y) && \text{si } M_y = I_q \\ &= \tau_p(x; y) && \text{si } M_y = D_y \\ &= \tau_{ep}(x; y) && \text{si } M_y = D_{1/\sigma^2_y}. \end{aligned}$$

Preuve :

en utilisant le lemme 1 et les expressions (2), (3) et (4), la propriété découle des calculs suivants :

– pour $M_y = I_q$ $\quad I_x[N_y] = \sum_{k=1}^{q} D[Q_x(y^k), Q_x(y^k)] = \sum_{k=1}^{q} ||Q_x(y^k)||^2$

de plus $\quad I[N_y] = \text{trace}[V_y M_y] = \sum_{k=1}^{q} \text{var}\,(y^k) = \sum_{k=1}^{q} ||y^k||^2.$

– pour $M_y = D_y$ $\quad I_x[N_y] = \sum_{k=1}^{q} (n.k/n)||Q_x(y^k)||^2$ de plus

$$I[N_y] = \sum_{k=1}^{q} (n.k/n)||y^k||^2.$$

– pour $M_y = D_{1/\sigma^2_y}$ $\quad I_x[N_y] = \sum_{k=1}^{q} ||Q_x(y^k)||^2/||yk||^2$ de plus $I[N_y] = q.$

Remarques 5 :

a) Si les variables $\{x^j\}$ et $\{y^k\}$ sont quantitatives alors, pour $M_y = I_q$, la valeur du coefficient CARD$[x; y]$ est égale [17] à celle du coefficient d'association dissymétrique de Stewart-Love [23].

b) Selon le type et le nombre de variables, dans les cas particuliers où $M_y = V_y^{-1}$ ou $M_y = \chi^2_y$, on notera que le numérateur $I_x[N_y]$ du coefficient CARD$[x; y]$ est égal ([17], ou voir preuve du lemme 1) à la somme des carrés des coefficients de corrélation canoniques, soit encore, au carré du coefficient de Bravais-Pearson, au ϕ^2 de Pearson ou au rapport de corrélation généralisé.

On déduit de la démonstration de la propriété précédente et de la remarque 4,

Corollaire 1 : M_x étant un produit scalaire quelconque, si $M = R[M_x, M_y]$ et

- si $M_y = (1/\sum_{l=1}^{q} ||y^l||^2) I_q$ alors $\tau(x;y) = I_x[N_y]$.
- si $M_y = (1/\sum_{l=1}^{q} (n.l/n)||y^l||^2) D_y$ alors $\tau_p(x;y) = I_x[N_y]$.
- si $M_y = (1/q) D_{1/\sigma_y^2}$ alors $\tau_{ep}(x;y) = I_x[N_y]$.

Diag$[a_l]$ étant une matrice diagonale dont le $l^{\text{ème}}$ élément diagonal est a_l, on déduit du corollaire 1, des égalités (1), de la propriété 1 et de la remarque 2a),

Propriété 3 : M_x étant un produit scalaire quelconque, si $M = R[M_x, M_y]$ et

- si $M_y = \text{Diag}[n^2/\sum_{l=1}^{q} n.l(n-n.l)]$ alors $\tau(x;y) = I_x[N_y] = I[N_g(y/x)]$.
- si $M_y = \text{Diag}[n^2 n.k/\sum_{l=1}^{q} n_{.l}^2(n-n.l)]$ alors $\tau_p(x;y) = I_x[N_y] = I[N_g(y/x)]$.
- si $M_y = \text{Diag}[n^2/qn.l(n-n.l)]$ alors $\tau_{ep}(x;y) = I_x[N_y] = I[N_g(y/x)]$.

3. Analyses factorielles de l'association dissymétrique entre deux variables qualitatives

On utilisera les notations suivantes :

- $[\{e_j(x) \in E_x; j = 1,p\} \cup \{e_k(y) \in E_y; k = 1,q\}]$ est la base canonique de $E = E_x \oplus E_y$ et $[\{e_j^*(x); j = 1,p\} \cup \{e_k^*(y); k = 1,q\}]$ sa base duale,
- Pry est le projecteur cartésien de E sur l'espace E_y,
- $\underline{y}_i^k$[resp. $\underline{x}_i^j$] étant la valeur de l'indicatrice non centrée $\underline{y}^k$[resp. $\underline{x}^j$] pour l'individu i, on pose $\underline{y}_i = \sum_{k=1}^{q} \underline{y}_i^k e_k(y)$ et $\underline{x}_i = \sum_{j=1}^{p} \underline{x}_i^j e_j(x)$,
- y_i^k[resp. x_i^j] étant la valeur de l'indicatrice centrée y^k[resp. x^j] pour l'individu i, on pose $y_i = \sum_{k=1}^{q} y_i^k e_k(y)$ et $x_i = \sum_{j=1}^{p} x_i^j e_j(x)$,
- $\underline{X}(n,p)$ étant la matrice d'éléments $[\underline{X}]_{ij} = \underline{x}_i^j$ pour tout (i,j), on pose $V_{y\underline{x}} = {}^tY D \underline{X}$,
- $g(\underline{x}) = (1/n)\sum_{i=1}^{n} \underline{x}_i = \sum_{j=1}^{p} (n_{j.}/n) e_j(x)$, $g(\underline{y}) = (1/n)\sum_{i=1}^{n} \underline{y}_i = \sum_{k=1}^{q} (n_{.k}/n) e_k(y)$,

- $g(x/y^k) = \sum_{j=1}^{p}[(n_{jk}/n_{.k}) - (n_{j.}/n)]e_j(x)$, le centre de gravité des p points $\{e_j(x) - g(\underline{x}); j = 1,p\}$, relativement au système de poids $\{n_{jk}/n_{.k}; j = 1,p\}$,
- $N_g(x/y) = \{g(x/y^k); k = 1,q\}$,
- $g(y/x^j) = \sum_{k=1}^{q}[(n_{jk}/n_{j.}) - (n_{.k}/n)]e_k(y)$, le centre de gravité des q points $\{e_k(y) - g(\underline{y}); k = 1,q\}$, relativement au système de poids $\{n_{jk}/n_{j.}; k = 1,q\}$,
- $N_g(y/x) = \{g(y/x^j); j = 1,p\}$.

On notera que $\forall i = 1,n$ si $\underline{y}_i^k = 1$ alors on a $\underline{y}_{\,i} = e_k(y)$ et $y_i = e_k(y) - g(\underline{y})$; de même $\forall i = 1,n$ $\underline{x}_i^j = 1$ implique $\underline{x}_{\,i} = e_j(x)$ et $x_i = e_j(x) - g(\underline{x})$.

Remarques 6 : Il est immédiat de montrer que :

- $({}^t\underline{X}D\underline{X})^{-1} = \chi_x^2$ et $V_{y\underline{x}} = V_{yx}$.
- $g(\underline{x}) = \text{Inx } {}^t\underline{X}D1_n$, où 1_n est le vecteur de coordonnées 1 dans $\mathbb{R}^n$. (8)
- $g(y/x^j) = \sum_{k=1}^{q}[V_{y\underline{x}}\chi_x^2]_{kj}e_k(y) = \sum_{k=1}^{q}[V_{yx}\chi_x^2]_{kj}e_k(y)$. (9)

3.1. Traduction de l'Analyse Factorielle des Correspondances dans le modèle euclidien relationnel

On sait que l'AFC est équivalente aux deux Analyses en Composantes Principales (ACP) suivantes :

$$\text{ACP du triplet } [N_g(y/x); \chi_y^2; D_x] \tag{10}$$

et

$$\text{ACP du triplet } [N_g(x/y); \chi_x^2; D_y]. \tag{11}$$

Dans (10) [resp. (11)] les vecteurs de l'espace E_y[resp. E_x] ont été identifiés avec leur image dans le sous-espace $E_y \subset E$[resp. $E_x \subset E$], via l'injection canonique Iny[resp. Inx].

Faire une AFC revient donc à effectuer une décomposition en moments principaux de la valeur de l'association symétrique, entre les variables x et y, mesurée par le ϕ^2 de Pearson : la valeur de ce dernier est égale [cf. remarques 2 a) et b)] à $I[N_g(y/x)] = I[N_g(x/y)]$.

Dans $E = E_x \oplus E_y$ muni du produit scalaire relationnel $R[\chi_x^2, \chi_y^2]$, c'est-à-dire dans un Modèle Euclidien Relationnel (MER), rappelons ([9] et lemmes 2 et 2^{bis} ci-après) que les moments principaux non nuls et les représentations des individus (chaque individu étant nommé par le nom de la modalité correspondant à la qualité qu'il possède) des deux ACP (10) et (11) sont respectivement identiques (en identifiant

les vecteurs des espaces E_x, E_y avec leur image par injection canonique) à ceux des deux ACP suivantes :

$$\text{ACP du triplet } [\{P_y(x_i); i = 1, n\}; R[\chi^2_x, \chi^2_y]; D] \tag{10'}$$

et

$$\text{ACP du triplet } [\{P_x(y_i); i = 1, n\}; R[\chi^2_x, \chi^2_y]; D]. \tag{11'}$$

On dira, par extension, que l'ACP $(10')$ [resp. $(11')$] est équivalente à l'ACP (10) [resp. (11)].

Dans (10) et (11), les nuages de points $N_g(y/x)$ et $N_g(x/y)$ sont respectivement dans des espaces $\mathbb{R}^q$ et $\mathbb{R}^p$ distincts, alors que ceux de $(10')$ et $(11')$ sont respectivement dans les sous-espaces $E_y = \mathbb{R}^q$ et $E_x = \mathbb{R}^p$ d'un même espace euclidien $E = E_x \oplus E_y$, relationnel pour les ensembles de variables $\{x^j\}$ et $\{y^k\}$.

Etablissons les lemmes suivants dans le cadre plus général où les résultats sont indépendants du type des variables $\{y^k\}$. Pour cela, on note V_y^- [resp. V_x^-] une inverse généralisée interne de plein rang de V_y[resp. V_x] et on rappelle (cf. annexe) que χ^2_x est une inverse généralisée interne de plein rang de V_x.

Lemme 2 : Si $M = R[\chi^2_x, V_y^-]$ alors

$$P_y[g(\underline{x})] = 0 \text{ et } [\underline{x}_i^j = 1 \Rightarrow P_y(x_i) = P_y[e_j(x)] = g(y/x^j)].$$

Preuve :

Vu que $M = R[\chi^2_x, V_y^-]$ on a $M_{yx} = V_y^- V_{yx} \chi^2_x$ (cf. annexe), en utilisant (7) mais pour P_y, il vient,

$$P_y \text{Inx} = \text{Iny}\, V_{yx}\chi^2_x = \text{Iny}\, V_{y\underline{x}}\chi^2_x. \tag{12}$$

Sachant que $Q_{\underline{x}} = \underline{X}\chi^2_x\ {}^t\underline{X}D$ est une expression de l'opérateur de projection orthogonale sur $\text{Im}\,\underline{X} \subset \mathbb{R}^n$, on déduit de (8) et (12),

$P_y[g(\underline{x})] = \text{Iny}^t Y D Q_{\underline{x}} 1_n = \text{Iny}^t Y D 1_n = 0$ puisque $1_n \in \text{Im}\,\underline{X}$ et que les $\{y^k\}$ sont centrées; ce qui implique $P_y(x_i) = P_y[\underline{x}_i - g(\underline{x})] = P_y(\underline{x}_i)$.

Finalement en utilisant (12) puis (9), et en supposant $\underline{x}_i^j = 1$, on a

$$P_y(\underline{x}_i) = P_y[e_j(x)] = \sum_{k=1}^{q} < \text{Iny}\, V_{yx}\chi^2_x\, Prx\ \ e_j(x), e_k^*(y) > e_k(y)$$

$$= \sum_{k=1}^{q} [V_{yx}\chi^2_x]_{kj}\, e_k(y) = g(y/x^j).$$

On démontrerait de même, quand y est qualitative, et quel que soit le type des $\{x^j\}$

Lemme 2bis : Si $M = R[V_x^-, \chi_y^2]$ alors

$$P_x[g(\underline{y})] = 0, \quad \text{et} \quad [\underline{y}_i^k = 1 \Rightarrow P_x(y_i) = P_x[e_k(y)] = g(x/y^k)].$$

Remarque 7 : Pour évaluer les choix effectués, on notera (démonstration semblable à celle proposée dans [22]) que (sous des conditions de normalité sur M, non restrictives en pratique) si $M_x = V_x^-$ et $M_y = V_y^-$, alors

$$\sum_{i=1}^{n}(1/n)||y_i - P_y(x_i)||_M^2 \quad [\text{resp.} \sum_{i=1}^{n}(1/n)||x_i - P_x(y_i)||_M^2]$$

est minimum si et seulement si $M = R[V_x^-, V_y^-]$.

Les représentations graphiques simultanées et barycentriques [4] des points modalités $\{x^j\} \cup \{y^k\}$, où les $\{x^j\}$ sont les centres de gravité des $\{y^k\}$, sont donc obtenues, dans le MER, avec $R[\chi_x^2, \chi_y^2]$, en projetant orthogonalement sur les plans principaux de l'ACP (11′) le nuage :

$$\{P_x[e_k(y)]; k = 1, q\} \cup \{P_x P_y[e_j(x)]; j = 1, p\}. \tag{13}$$

Ces représentations sont justifiées par l'égalité des moments principaux non nuls des ACP (10) et (11), mais aussi par les formules cohérentes de transition [4] entre les composantes principales. Ces représentations trouvent dans le MER des justifications supplémentaires dans (13) et dans la propriété suivante.

Propriété 4 : $\{(\lambda_r, a_r)\}$ et $\{(\lambda_r, b_r)\}$ étant les moments principaux non nuls et les vecteurs axiaux principaux respectivement des ACP (10′) et (11′) on a $P_x(a_r) = \sqrt{\lambda_r} b_r$.

Preuve :

En effet, on sait (formules de transition) que pour tout moment principal non nul λ_r, on a :

$$b_r = (1/\sqrt{\lambda_r})\text{Inx}\, V_{xy}\chi_y^2 \ \text{Pry}\, a_r$$

or, en utilisant la formule (12) mais pour P_x, on a :

$$P_x(a_r) = \text{Inx}\, V_{xy}\chi_y^2 \ \text{Pry}\, a_r = \sqrt{\lambda_r} b_r.$$

Remarques 8 :

a) En AFC on peut proposer d'autres représentations, simultanées et barycentriques, en permutant les rôles de x et y, c'est-à-dire en projetant le nuage $\{P_y[e_j(x)]; j = 1, p\} \cup \{P_y P_x[e_k(y)]; k = 1, q\}$ sur les plans principaux de l'ACP (10′).

b) La propriété 4 est une conséquence d'un cas particulier ($M_x = \chi_x^2$ et $M_y = \chi_y^2$) d'une propriété caractéristique des produits scalaires relationnels. En effet,

- on sait [7,6] que : $a_r = \text{Iny}^t Y D A^r$ et $b_r = \text{Inx}^t X D B^r$
  où (A^r, B^r) sont les couples de variables canoniques de (F_y, F_x, D), F_x[resp. F_y] étant le sous-espace de l'espace des variables engendré par les vecteurs

$$\{x^j; j = 1, p\} \text{ [resp. } \{y^k; k = 1, q\}];$$

- les deux conditions utiles [22] qui caractérisent $M = R[M_x, M_y]$ sont :
  (i) les coefficients de corrélation canonique non nuls, de (E_x, E_y, M) et (F_x, F_y, D), sont égaux,
  (ii) les couples de variables canoniques de même rang r $(a'_r \in E_y, b'_r \in E_x)$ et $(A^r \in F_y, B^r \in F_x)$ se correspondent par les relations suivantes :

$$a'_r = \text{Iny}[(V_y M_y)^{1/2}]^+ {}^t Y D A^r = \text{Iny } {}^t Y D A^r \quad \text{pour} \quad M_y = \chi_y^2$$
$$b'_r = \text{Inx}[(V_x M_x)^{1/2}]^+ {}^t X D B^r = \text{Inx } {}^t X D B^r \quad \text{pour} \quad M_x = \chi_x^2.$$

On en déduit $a'_r = a_r$ et $b'_r = b_r$, or $P_x(a'_r) = \sqrt{\lambda_r} b'_r$, d'où le résultat.

3.2. Définition

La propriété 3 et les rappels précédents nous amènent à proposer la définition suivante.

Définition 3 : On dit que l'on fait l'analyse factorielle de l'association dissymétrique entre la variable qualitative x et la variable qualitative y, mesurée par un coefficient Q, si et seulement si :

a) on effectue les deux ACP suivantes :

$$\text{ACP du triplet } [N_g(y/x); M_y; D_x] \tag{14}$$

et

$$\text{ACP du triplet } [\{P_x(y_i); i = 1, n\}; R[\chi_x^2, M_y]; D], \tag{15}$$

b) le produit scalaire M_y est tel que les valeurs des inerties des nuages d'individus de (14) et (15) soient égales à la valeur du coefficient d'association dissymétrique Q.

A priori M_x pourrait être quelconque, le choix $M_x = \chi_x^2$ simplifie les calculs et permet de retrouver des résultats fondamentaux dans le cas particulier où $M_y = \chi_y^2$. Nous supposerons donc dans la suite que $M_x = \chi_x^2$.

Existence : D'après la propriété 3, il existe des produits scalaires M_y pour les coefficients τ, τ_p et τ_{ep}. Plus généralement, il découle du 1) de la propriété 7, ci-après, que si Q est un CARD alors il existe toujours un produit scalaire M_y : en

effet, si le coefficient CARD a été défini relativement à $M = R[M_x, M'_y]$, alors $M_y = M'_y/I[N_y]$ où $I[N_y]$ est calculée relativement à M'_y .

Remarque fondamentale : Pour $M_y = \chi^2_y$, l'ACP (15) est identique à l'ACP(11′) et donc équivalente à l'ACP (11); dans ce cas très particulier, l'analyse factorielle proposée est donc équivalente à l'AFC : elle analyse l'association *symétrique*, entre les variables x et y, mesurée par le ϕ^2.

On propose donc d'appeler l'analyse correspondant à la définition 3, «l'Analyse Factorielle des Correspondances Dissymétriques» (AFCD).

On peut en donner une définition équivalente en remplaçant (15) par l'ACP équivalente suivante :

$$\text{ACP du triplet}[\{P_x[e_k(y) - g(\underline{y})]; k = 1, q\}; R[\chi^2_x, M_y]; D_y]. \qquad (15')$$

3.3. Propriétés

On déduit la propriété suivante de (9) .

Propriété 5 : L'ACP (14) est identique à

$$\text{l'ACP du triplet } [\chi^2_x V_{xy}; M_y; D_x] \qquad (16)$$

Propriété 6 : L'ACP (15) est équivalente à

$$\text{l'ACP du triplet } [Y M_y[(V_y M_y)^{1/2}]^+ V_{yx}; \chi^2_x; D] \qquad (17)$$

Preuve :

Etant donné que $M_x = \chi^2_x$, en utilisant (6) et (7), sachant (cf. annexe) que $[(V_x \chi^2_x)^{1/2}]^+ V_{xy} = V_{xy}$, et $y_i = \text{Iny } {}^tY f^*_i$ [6], où $\{f^*_i; i = 1, n\}$ est la base duale de la base canonique de l'espace des variables ($\mathbb{R}^n$) de l'ACP (15), on a

$$P_x(y_i) = \text{Inx } M_x^{-1} M_{xy}\ {}^tY f^*_i = \text{Inx } V_{xy} M_y[(V_y M_y)^{1/2}]^+\ {}^tY f^*_i .$$

On en déduit

$$< P_x(y_i), e^*_j(x) >= [V_{xy} M_y[(V_y M_y)^{1/2}]^+\ {}^tY]_{ji}.$$

Les coordonnées des points $\{P_x(y_i); i = 1, n\}$, relativement à $\{e_j(x); j = 1, p\}$, sont donc les éléments des lignes de la matrice $Y M_y[(V_y M_y)^{1/2}]^+ V_{yx}$, vu que $[V_y M_y)^{1/2}]^+$ est M_y-symétrique (cf. annexe).

Soient,

– $\{\gamma_r(x)\}$ les moments principaux non nuls de l'ACP (16) ou (14) et $\{u_r\}$ les vecteurs axiaux principaux correspondants,

– $\{\gamma_r(y)\}$ les moments principaux non nuls de l'ACP (17) [resp. (15)] et $\{v_r\}$[resp. $\{\mathrm{Inx}\, v_r\}$] les vecteurs axiaux principaux correspondants.

Propriété 7 : M_y étant un produit scalaire quelconque, $\forall r$ on a :

$$1)\ \gamma_r(x) = \gamma_r(y) = \gamma_r$$

$$2)\quad u_r = (1/\sqrt{\gamma_r})Av_r \qquad \text{où}\quad A = V_{yx}\chi_x^2 \tag{18}$$

$$v_r = (1/\sqrt{\gamma_r})Bu_r \qquad \text{où}\quad B = V_{xy}M_y \tag{18'}$$

Preuve :

Vu que $\chi_x^2 D_x = I_p$, les moments et vecteurs axiaux principaux de l'ACP (16) sont les éléments propres (normés) de l'opérateur : ${}^t(\chi_x^2 V_{xy})D_x\chi_x^2 V_{xy}M_y = AB$.

En utilisant la M-symétrie de $[(V_yM_y)^{1/2}]^+$, les moments et vecteurs axiaux principaux de l'ACP (17) sont les éléments propres (normés) de l'opérateur :

$$V_{xy}\,{}^t([(V_yM_y)^{1/2}]^+)M_yV_yM_y[(V_yM_y)^{1/2}]^+V_{yx}\chi_x^2 = V_{xy}M_yV_{yx}\chi_x^2 = BA.$$

AB et BA ayant mêmes valeurs propres non nulles, on déduit aisément les assertions 1) et 2), compte tenu de ce que u_r [resp. v_r] est normé pour M_y[resp. χ_x^2].

Remarque 9 : En examinant les démonstrations des propriétés 6 et 7, on constate que les résultats établis sont valables quel que soit le type des variables $\{y^k\}$.

La propriété suivante est utilisée pour mettre en évidence dans la propriété 8, des formules de transition, entre les composantes principales de (14) et (15′), identiques, pour $M_y = \chi_y^2$, à celles établies en AFC.

Propriété 6$^{\text{bis}}$: L'ACP (15′) est équivalente à

$$\text{l'ACP du triplet } [\chi_y^2(V_yM_y)^{1/2}V_{yx};\chi_x^2;D_y] \tag{17'}$$

Preuve :

En posant $\widetilde{y}_k = e_k(y) - g(\underline{y}) = \mathrm{Iny}\,{}^t\widetilde{Y}f_k^*$, où $\{f_k^*; k = 1,q\}$ est la base duale de la base canonique de l'espace des variables $(\mathbb{R}^q)$ de l'ACP (15′), on montre, comme dans la propriété 6, que les coordonnées des points $\{P_x(\widetilde{y}_k); k = 1,q\}$, relativement à $\{e_j(x); j = 1,p\}$, sont les éléments des lignes de la matrice $\widetilde{Y}M_y[(V_yM_y)^{1/2}]^+V_{yx}$. Or, vu que

$-\ \underline{y}_i^k = 1 \Rightarrow y_i = \widetilde{y}_k$, on a $V_y = {}^tYDY = {}^t\widetilde{Y}D_y\widetilde{Y}$,

$-\ < e_k(y) - g(\underline{y}), e_l^*(y) >= [I_q - 1_q\,{}^t1_qD_y]_{kl}$ on a $\widetilde{Y} = I_q - 1_q\,{}^t1_qD_y$,

où 1_q est la matrice unicolonne dont les q éléments sont égaux à 1 et I_q la matrice unité d'ordre q.

On en déduit

$$V_y = [I_q - D_y1_q\,{}^t1_q]D_y\widetilde{Y} = D_y\widetilde{Y}$$

car les colonnes de $\widetilde{Y}$ sont centrées relativement à D_y

D'où $\widetilde{Y} = \chi^2_y V_y$ et finalement

$$\widetilde{Y} M_y[(V_y M_y)^{1/2}]^+ V_{yx} = \chi^2_y (V_y M_y)^{1/2} V_{yx}.$$

On note $\{U^r\}$[resp. $\{V^r\}$] les composantes principales de l'ACP (16) [resp. (17′)], correspondant aux vecteurs axiaux principaux $\{u_r\}$[resp. $\{v_r\}$]. On rappelle que l'ACP (17′) est équivalente à l'ACP (15′), elle même équivalente à l'ACP (15).

Propriété 8 : M_y étant un produit scalaire quelconque, $\forall r$ on a :

$$Ur = (1/\sqrt{\gamma_r})\, \chi^2_x V_{xy}\, {}^t[(V_y M_y)] V^r \tag{19}$$
$$V^r = (1/\sqrt{\gamma_r})\, \chi^2_y (V_y M_y)^{1/2} V_{yx} U^r \tag{19'}$$

Preuve : On a [6]

$$U^r = \chi^2_x V_{xy} M_y u_r \tag{20}$$
$$V^r = \chi^2_y (V_y M_y)^{1/2} V_{yx} \chi^2_x v_r. \tag{21}$$

En remplaçant v_r par son expression (18′) dans (21) et en utilisant (20) on obtient (19′).

De même, en remplaçant u_r par son expression (18) dans (20) et en utilisant (21) on a,

$$\begin{aligned} U^r &= (1/\sqrt{\gamma_r})\, \chi^2_x V_{xy} M_y V_{yx} \chi^2_x v_r \\ &= (1/\sqrt{\gamma_r})\, \chi^2_x V_{xy} M_y [(V_y M_y)^{1/2}]^+ D_y V^r \\ &= (1/\sqrt{\gamma_r})\, \chi^2_x V_{xy}\, {}^t[(V_y M_y)^{1/2}] M_y (V_y M_y)^+ D_y V^r \\ &= (1/\sqrt{\gamma_r})\, \chi^2_x V_{xy}\, {}^t[(V_y M_y)^{1/2}] V_y^+ D_y V^r \quad \text{car} \quad (V_y M_y)^+ = M_y^{-1} V_y^+ \\ &= (1/\sqrt{\gamma_r})\, \chi^2_x V_{xy}\, {}^t[(V_y M_y)^{1/2}] V_y^+ V_y V^r \end{aligned}$$

car $V_y = {}^t Y D Y = {}^t \underline{Y} D Y = D_y - {}^t \underline{Y} D 1_n\, {}^t 1_q D_y$, et ${}^t 1_q D_y V^r = 0$ puisque V^r est centrée, et finalement

$$U^r = (1/\sqrt{\gamma_r})\, \chi^2_x V_{xy}\, {}^t[(V_y M_y)^{1/2}] V^r$$

en effet

$${}^t[(V_y M_y)^{1/2}] V_y^+ V_y = {}^t[V_y V_y^+ (V_y M_y)^{1/2}] = {}^t[(V_y M_y)^{1/2}]$$

puisque (cf. annexe) $V_y V_y^+$ est un opérateur de projection sur $\mathrm{Im}\, V_y = \mathrm{Im}^t Y = \mathrm{Im}(V_y M_y)^{1/2}$ et ${}^t(V_y^+) = V_y^+$.

Remarque 10 : Si $M_y = \chi_y^2$, alors (18), (18′), (19) et (19′) sont les formules de transition de l'AFC.

En AFCD, les représentations simultanées et barycentriques, où les $\{x^j\}$ sont les centres de gravité des $\{y^k\}$ relativement au système de poids $\{njk/nj.; k = 1, q\}$, sont obtenues en projetant orthogonalement le nuage $\{P_x[e_k(y) - g(y)]; k = 1, q\} \cup \{P_x[g(y/x^j)]; j = 1, p\}$ sur les plans principaux de l'ACP (15). Contrairement à l'AFC, la dissymétrie des rôles joués par x et y ne permet pas ici de proposer, pour le coefficient $Q(x; y)$, les représentations obtenues en permutant les rôles de x et y (cf. remarque 8) : en effet, ces représentations seraient celles de l'AFCD pour le coefficient $Q(y; x)$. En AFCD, les barycentres sont toujours les représentants des modalités explicatives.

3.4. Lien avec l'Analyse Non Symétrique des Correspondances (ANSC)

Dans [10,14], N. Lauro et L. D'Ambra définissent l'ANSC entre x et y par les deux ACP suivantes :

$$\text{ACP du triplet } [\{((njk/nj.) - (n.k/n); k = 1, q); j = 1, p\}; M_y = I_q; D_x] \quad (22)$$

et

$$\text{ACP du triplet } [\{((njk/nj.) - (n.k/n); j = 1, p); k = 1, q\}; M_x = D_x; I_q]. \quad (23)$$

Etant donné que (propriété 1)

$$[\sum_{l=1}^{q} n.l(n - n.l)/n^2]\tau(x; y) = \sum_{j=1}^{p}(nj./n)\sum_{k=1}^{q}[(njk/nj.) - (n.k/n)]^2 \quad (24)$$

$$= \sum_{k=1}^{q}\sum_{j=1}^{p}(nj./n)[(njk/nj.) - (n.k/n)]^2. \quad (25)$$

On observe que :

a) (25) est obtenu à partir de (24) en permutant les signes de sommation.

b) (24) est l'expression de l'inertie du nuage du triplet (22), et (25) celle du nuage du triplet (23), ce qui, compte tenu du problème posé, est cohérent avec les choix des ACP (22) et (23).

On constate que, dans le cas particulier où $M_y = I_q$,

a) l'ACP (22) est identique à l'ACP (14) de la définition 3, pour (cf. propriété 3)

$$Q(x; y) = [\sum_{l=1}^{q} n.l(n - n.l)/n^2]\ \tau(x; y).$$

b) l'ACP (23) est non équivalente à l'ACP (15′) ou (15) : en effet, on déduit de (9) que les moments principaux et les vecteurs axiaux principaux de l'ACP (23) sont les éléments propres normés de l'opérateur $\chi_x^2 V_{xy} I_q V_{yx} \chi_x^2 D_x = \chi_x^2 V_{xy} V_{yx}$, alors que l'opérateur correspondant de l'ACP (17′) ou (17) est, pour $M_y = I_q$, $V_{xy} V_{yx} \chi_x^2$ (cf. preuve propriété 7).

En notant (γ'_r, v'_r, V'^r) les moments principaux non nuls, les vecteurs axiaux principaux et les composantes principales correspondants de l'ACP (23), on a :

$$\forall r \ \gamma'_r = \gamma_r, \quad v'_r = \chi_x^2 v_r$$

de plus, compte tenu de (21), on a

$$V'_r = V_{yx} v'_r = V_{yx} \chi_x^2 v_r = [(V_y M_y)^{1/2}]^+ D_y V^r = [V_y^{1/2}]^+ D_y V^r \text{ pour } M_y = I_q.$$

En conclusion :

a) Pour l'analyse de l'association dissymétrique entre x et y, mesurée par le τ de Goodman-Kruskal, bien que l'ANSC et l'AFCD paraissent être voisines, notamment dans le cas très particulier où $\chi_x^2 = pI_p$, en fait les résultats graphiques de l'ACP (23) peuvent être très différents de ceux de l'ACP (17′), c'est-à-dire ceux de l'ACP (15′) ou (15) pour $M_y = I_q$ (cf. paragraphe 4).

b) Vu que, dans la définition 3, le produit scalaire M_y est quelconque, l'AFCD offre un champ d'application plus vaste que celui de l'ANSC : ainsi, pour $M_y = \chi_y^2$ (cas symétrique) elle s'identifie à l'AFC, mais surtout, dans le cas d'une association dissymétrique, elle permet d'effectuer une analyse en prenant un coefficient CARD «aussi bien adapté que possible» à la structure de données observée (cf. paragraphe 4).

4. Un exemple sur des données simulées

4.1. Les données

On a $p = 5$, $q = 4$ et $n = 6391$; les données sont présentées dans le tableau 1, où $\{x_j; j = 1,5\}$[resp. $\{y_k; k = 1,4\}$] représentent les modalités de la variable explicative x [resp. à expliquer y].

TABLEAU 1 :
Effectifs (njk)

	x_1	x_2	x_3	x_4	x_5	(n.k)
y_1	126	124	85	12	13	360
y_2	130	11	87	124	17	369
y_3	9	818	672	13	1352	2864
y_4	8	10	657	1065	1058	2798
(nj.)	273	963	1501	1214	2440	6391

Dans les tableaux 2, 3 et 4 on donne des quantités qui entrent dans les formules des coefficients d'association (cf. paragraphes 2.1 et 2.2), et qui jouent donc des rôles importants dans les analyses.

TABLEAU 2 :
Profils colonnes (njk /nj.).

	x_1	x_2	x_3	x_4	x_5	(n.k/n)
y_1	0.4615	0.1288	0.0566	0.0099	0.0053	0.0563
y_2	0.4762	0.0114	0.0580	0.1021	0.0070	0.0577
y_3	0.0330	0.8494	0.4477	0.0107	0.5541	0.4481
y_4	0.0293	0.0104	0.4377	0.8773	0.4336	0.4378

TABLEAU 3 :
Profils colonnes centrés $((n_{jk}/n_{j.}) - (n_{.k}/n))$

	x_1	x_2	x_3	x_4	x_5
y_1	0.4052	0.0725	0.0003	−0.0464	−0.0510
y_2	0.4185	−0.0463	0.0003	0.0444	−0.0507
y_3	−0.4151	0.4013	−0.0004	−0.4374	0.1060
y_4	−0.4085	−0.4274	−0.0001	0.4395	−0.0042

TABLEAU 4 :
Variances des $\{y^k\}$.

y_1	y^2	y^3	y^4
0.0531	0.0544	0.2473	0.2461

4.2. Quelques résultats numériques

On a effectué cinq analyses :

– trois AFCD, respectivement pour les coefficients τp, τ et τep, notées AFCD (τp), AFCD (τ) et AFCD (τep), c'est-à-dire respectivement pour

$$M_y = \text{Diag}\,[n^2\, n.k/\sum_{l=1}^{q} n_{.l}^2(n - n.l)],\quad M_y = \text{Diag}\,[n^2/\sum_{l=1}^{q} n.l(n - n.l)]$$

et
$$M_y = \text{Diag}\,[n^2/q\; n.l(n - n.l)].$$

– l'AFC [ou AFCD pour le ϕ^2] et l'ANSC avec $M_y = I_q/[\sum_{l=1}^{q} n.l(n - n.l)/n^2]$ (somme des moments principaux égale au τ) afin de faciliter les comparaisons, notamment au paragraphe 4.3.2.

Dans le tableau 5, on donne quelques résultats numériques obtenus avec les AFCD et l'AFC : ces résultats correspondent aux graphiques des figures 1 et 2.

TABLEAU 5 :
Analyses Factorielles des Correspondances Dissymétriques $(\tau_p, \tau, \tau_{ep})$
et Symétriques (AFC : ϕ^2*).*

Analyse du tau pondéré $\tau_p(x;y) = 0,288$				Analyse du tau de Goodman-Kruskal $\tau(x;y) = 0,269$			
	1	2	Total		1	2	Total
Valeur propre	0,251	0,037	0,288	Valeur propre	0,214	0,055	0,269
%	87,22	12,77	99,99	%	79,55	20,39	99,94
Contributions (%)				Contributions (%)			
x_1	0,01	87,72	11,21	x_1	0,00	87,71	17,89
x_2	40,69	0,44	35,55	x_2	40,93	0,41	32,66
x_3	0,00	0,00	0,00	x_3	0,00	0,00	0,00
x_4	57,42	0,03	50,09	x_4	57,27	0,02	45,58
x_5	1,88	11,81	3,15	x_5	1,80	11,86	3,87
y_1	0,05	44,14	5,68	y_1	0,59	44,59	9,59
y_2	0,14	44,44	5,80	y_2	0,79	44,00	9,63
y_3	49,74	5,36	44,08	y_3	49,07	5,47	40,15
y_4	50,07	6,06	44,44	y_4	49,55	5,94	40,63
Analyse du tau équipondéré $\tau_{ep}(x;y) = 0,231$				Analyse des correspondances $\phi^2(x,y) = 0,645$			
	1	2	Total		1	2	Total
Valeur propre	0,137	0,094	0,231	Valeur propre	0,327	0,318	0,645
%	59,33	40,51	99,84	%	50,59	49,20	99,79
Contributions (%)				Contributions (%)			
x_1	0,02	87,65	35,53	x_1	87,64	0,03	44,37
x_2	41,69	0,46	24,98	x_2	0,19	42,61	21,14
x_3	0,00	0,00	0,00	x_3	0,00	0,00	0,00
x_4	56,60	0,04	33,63	x_4	0,01	56,09	27,64
x_5	1,69	11,85	5,86	x_5	12,16	1,27	6,85
y_1	2,68	45,46	20,08	y_1	44,40	5,54	25,29
y_2	3,36	43,12	19,54	y_2	44,14	4,33	24,55
y_3	46,97	5,36	30,04	y_3	6,57	43,95	24,95
y_4	46,99	6,06	30,34	y_4	4,89	46,18	25,21

Les valeurs des contributions des modalités, aux deux premiers moments principaux, sont cohérentes avec les statistiques des tableaux 2, 3 et 4 : on notera notamment que

a) les deux groupes de couples de modalités bien associées $[(x_2;y_3),(x_4;y_4)]$ et $[(x_1;y_1),(x_1;y_2)]$ jouent des rôles différents dans les AFCD (τ_{ep}, τ et τ_p), et dans l'AFC (ϕ^2).

b) les valeurs des contributions (Total) des modalités (y_3, y_4) aux coefficients d'association croissent en partant d'un minimum observé dans l'AFC, en passant par l'AFCD (τ_{ep}) et l'AFCD (τ), pour atteindre un maximum observé dans l'AFCD (τ_p) ; de même, les valeurs des contributions de ces mêmes modalités, au premier moment principal, dans l'AFCD (τ_{ep}) sont inférieures à celles observées dans l'AFCD (τ), qui sont elles mêmes inférieures à celles observées dans l'AFCD (τ_p).

Avec ce jeu de données simulées, il est très clair que les résultats des analyses des associations dissymétriques mesurées par les coefficients $\{\tau_{ep}, \tau \text{ et } \tau_p\}$ sont différents de ceux de l'AFC.

Dans le tableau 6, on donne les mêmes résultats numériques, mais obtenus avec l'ANSC : si on compare ces résultats avec ceux de l'AFCD (τ) (Tableau 5), on observe (i) que les valeurs propres et les contributions des modalités $\{x_j\}$ sont bien identiques pour les deux analyses, et (ii) que, par contre, les contributions des modalités $\{y_k\}$ sont différentes, essentiellement celles concernant le deuxième moment principal. Ces résultats correspondent au graphique de la figure 3.

TABLEAU 6 :
Analyse Non Symétrique des Correspondances

Analyse du tau de Goodman-Kruskal $\tau(x;y) = 0,269$			
	1	2	Total
Valeur propre	0,214	0,055	0,269
%	79,55	20,39	99,94
Contributions (%)			
x_1	0,00	87,71	17,89
x_2	40,93	0,41	32,66
x_3	0,00	0,00	0,00
x_4	57,27	0,02	45,58
x_5	1,80	11,86	3,87
y_1	0,63	25,28	5,68
y_2	0,74	24,72	5,66
y_3	49,78	24,56	44,61
y_4	48,85	25,44	44,05

4.3. Quelques résultats graphiques

4.3.1. Graphiques

Un point X_j[resp. Y_k] de la figure 1(a) [resp. 1(b)] représente la projection orthogonale du point $g(y/x^j)$ [resp. $P_x[e_k(y) - g(y)]$] sur le premier plan principal de l'ACP (14) [resp. (15) ou (15')] pour $Q = \tau$.

Nous observons immédiatement, comme nous l'avons constaté sur les résultats numériques, que ces résultats graphiques sont très différents de ceux produits par l'AFC (cf. figure 2).

Sur la figure 1(a) [resp. 1(b)]

- plus un point modalité X_j[resp. Y_k] est éloigné de l'origine, plus sa contribution au coefficient d'association dissymétrique (ici $Q = \tau$) est importante, donc plus la modalité «x_j de x explique» [resp. «y_k de y est expliquée par»] la variable y[resp. x].
- plus deux points sont voisins, plus les modalités correspondantes de x [resp. y] expliquent y [resp. sont expliquées par x] de façon semblable.

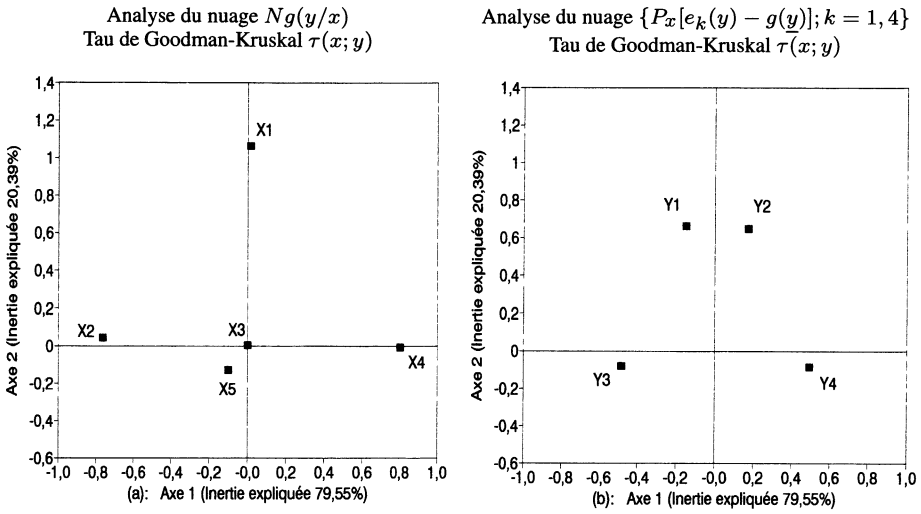

FIGURE 1
Analyse Factorielle des Correspondances Dissymétriques pour le τ de Goodman-Kruskal

Le graphique de la figure 1(a) [resp. 1(b)] est une description des «potentiels à expliquer» [resp. «à être expliquées»] des modalités de x [resp. y] .

Ainsi, le pourcentage d'inertie expliquée par le premier plan principal étant voisin de 100%,

– le potentiel de prévision de X_3 est nul, celui de X_1 est le plus élevé, et ceux de X_2 et X_4 sont élevés et très différents.

– les potentiels à être expliquées de Y_1 et Y_2 sont plus ressemblants que ceux de Y_3 et Y_4.

Ces affirmations sont cohérentes avec la structure des profils colonnes centrés du tableau 3.

Dans une AFCD, pour représenter graphiquement l'association «x explique y», les représentations simultanées et barycentriques (figure 2), définies et justifiées à la fin du paragraphe 3.3, semblent être, comme en AFC, des outils utiles.

Sur la figure 3, on a représenté les graphiques de l'ANSC, c'est-à-dire ceux produits par les ACP (22) et (23). On notera que le nuage des $\{Y_k\}$ a une «petite

taille» relativement à celle du nuage des $\{X_j\}$: cela est dû au fait que, par définition, (i) ces deux nuages ont même inertie et que (ii) les points $\{Y_k\}$ ont tous une masse égale à 1, alors que les poids des $\{X_j\}$, plus nombreux, sont évidemment plus petits, respectivement égaux à $\{0.04, 0.15, 0.24, 0.19, 0.38\}$.

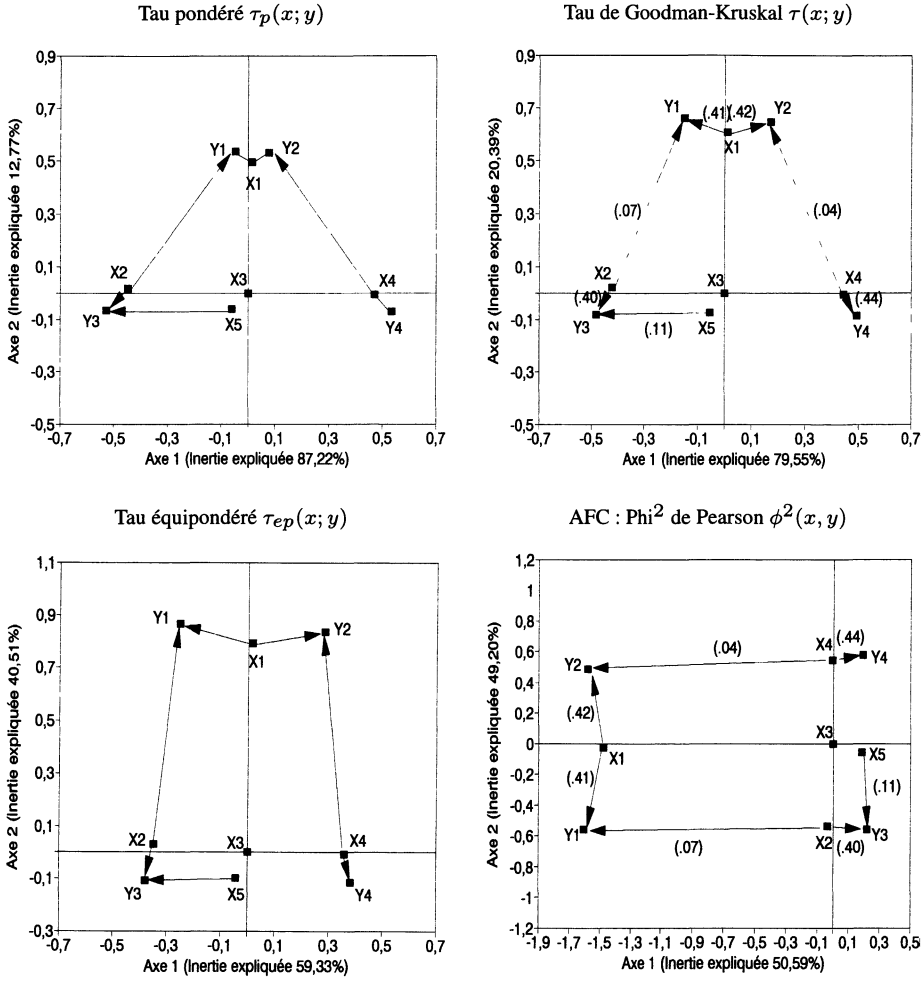

FIGURE 2
Analyses Factorielles des Correspondances Dissymétriques
(τ_p, τ, τ_{ep}) et Symétriques (AFC : ϕ^2)
(représentations simultanées et barycentriques)

Pour faciliter la lecture, dans certains graphiques, nous avons reporté sur les flèches, matérialisant la direction des associations positives, les valeurs des profils colonnes centrés (tableau 3) correspondants. De même, pour faciliter les comparaisons

proposées dans le paragraphe 4.3.2, les échelles sur les axes des abscisses et des ordonnées sont les mêmes.

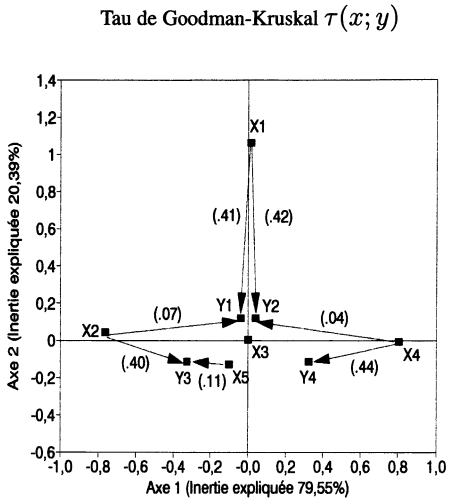

FIGURE 3
Analyse Non Symétrique des Correspondances (représentation simultanée)

Remarques 11 : Pour représenter graphiquement les associations, il serait intéressant d'étudier les propriétés des ACP suivantes :

1) ACP du triplet $[\{\{P_x[e_k(y) - g(\underline{y})]; k = 1, q\} \cup N_g(y/x)\}; R[\chi^2_x, M_y]; D_{y \cup x}]$
avec $[D_{y \cup x}]_l = n.l/2n$ pour $l \leqslant q$
$= nl./2n$ pour $q+1 \leqslant l \leqslant p+q$.

2) ACP du triplet $[\{\{P_x[e_k(y) - g(\underline{y})]; k = 1, q\} \times N_g(y/x)\}; R[\chi^2_x, M_y]; D_{q \times p}]$
où $D_{q \times p}$ est une matrice diagonale de poids de dimension $q \times p$.

4.3.2. Quelques éléments pour évaluer les analyses

Ces évaluations sont basées sur la comparaison des valeurs des distances, mesurées sur les représentations graphiques (figures 2 et 3) entre les points modalités explicatives $\{X_j\}$ et les points modalités à expliquer $\{Y_k\}$ pour une analyse donnée, avec les valeurs se trouvant dans le tableau 3.

Tout d'abord, vu que toutes les valeurs de la troisième colonne du tableau 3, correspondant à x_3, sont égales (nulles), on est amené à penser que le graphique de l'AFCD (τ_p) peut représenter correctement l'association «x explique y», puisque pour cette analyse, les distances de X_3 avec les $\{Y_k\}$, notées $\mathrm{d}(X_3, Y_k)$, sont approximativement égales (tableau 7).

TABLEAU 7 :

Distances du point modalité explicatif X_3 aux points modalités à expliquer $\{Y_k\}$.

	AFCD			AFC	ANSC
	τ_p	τ	τ_{ep}		
$d(X_3, Y_1)$	0,54	0,67	0,90	1,70	0,12
$d(X_3, Y_2)$	0,54	0,67	0,88	1,65	0,12
$d(X_3, Y_3)$	0,53	0,49	0,39	0,60	0,35
$d(X_3, Y_4)$	0,54	0,50	0,40	0,61	0,34

Une deuxième observation conduit à la même conclusion.

Soient $\varepsilon = \{\varepsilon_{jk} = (njk/nj.) - (n.k/n); j = 1,p \text{ et } k = 1,q\}, D = \{d(X_j, Y_k); j = 1,p \text{ et } k = 1,q\}$ l'ensemble des distances [mesurées sur le plan principal (1,2)] entre les couples (X_j, Y_k) pour une analyse donnée et $C(\varepsilon, D)$ l'indice de «concordance» de rangs de Kendall entre les deux séries de mesures ε et D : ici la valeur de $C(\varepsilon, D)$ est égale au nombre de produits négatifs parmi les $pq(pq-1)/2$ produits de la forme $(\varepsilon_{jk} - \varepsilon_{lm})[d(X_j, Y_k) - d(X_l, Y_m)]$, puisqu' en situation «idéale» on devrait avoir : $\varepsilon_{jk} > \varepsilon_{lm} \Rightarrow d(X_j, Y_k) < d(X_l, Y_m)$.

La lecture du tableau 8 montre que parmi les valeurs de l'indice de concordance $C(\varepsilon, D)$, calculé à partir des graphiques des figures 2 et 3, celle de l'AFCD (τ_p) est la plus grande; on notera que le maximum de $C(\varepsilon, D)$ vaut 190, puisque $pq = 20$.

TABLEAU 8 :

Valeurs de l'indice $C(\varepsilon, D)$ de concordance de rangs de Kendall

AFCD			AFC	ANSC
τ_p	τ	τ_{ep}		
164	161	140	140	114

Pour évaluer visuellement, à partir des valeurs de $C(\varepsilon, D)$, les résultats de l'AFCD (τ_p), de l'AFC et de l'ANSC, on a représenté, sur la figure 4, pour chacune de ces trois analyses, les graphes de ε et de D en fonction des couples (X_j, Y_k) rangés dans l'ordre des valeurs décroissantes des $\{\varepsilon_{jk}\}$ (cf. tableau 3).

Remarques 12 :

1) On suggère ici une technique, justifiée et utilisée par ailleurs [1,5,11,18], qui a permis d'améliorer, au sens du critère de concordance $[C(\varepsilon, D) = 165]$, les résultats obtenus avec l'AFCD (τ_p) : on a choisi «le meilleur» M_y [obtenu pour $\alpha_2 = 14/15$] dans une famille de distances définie par :

$$M_y = (1-\alpha_1)\,\chi^2_y + \alpha_1 I_q \quad \text{et} \quad M_y = (1-\alpha_2) I_q + \alpha_2 D_y$$

$$\text{où} \quad \alpha_1, \alpha_2 \in \{0,\ 1/15,\ 2/15, ..., 1\}.$$

2) Nous avons analysé avec l'AFCD (τ) les données étudiées par F. Benzecri avec l'AFC dans [3], et reprises par Lauro et D'Ambra [10] avec l'ANSC. Nous ne publions pas ici ces résultats par manque de place; de plus, ces derniers ne présentent qu'un intérêt très moyen : en effet, bien que les moments principaux soient différents, les graphiques sur le premier plan principal sont très ressemblants.

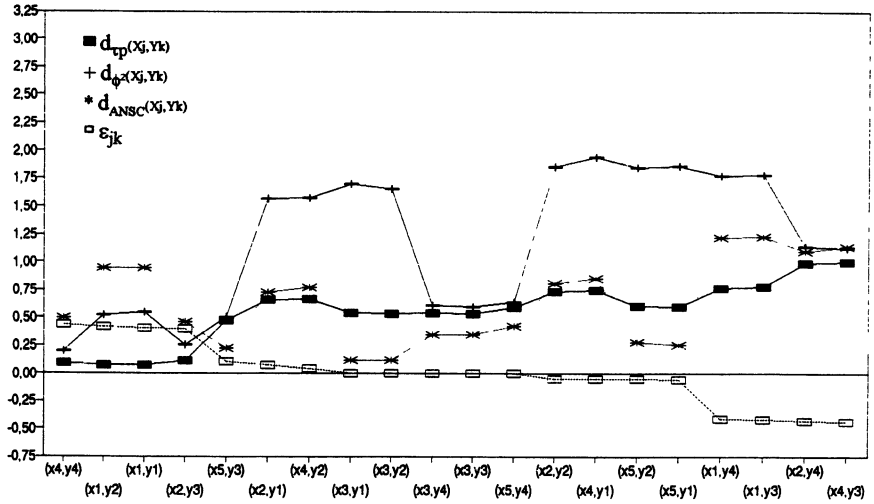

FIGURE 4

Graphes des $\varepsilon_{jk} = (n_{jk}/nj.) - (n.k/n)$, et des distances $d(X_j, Y_k)$ pour l'AFCD (τ_p), l'AFC (ϕ^2) et l'ANSC

5. Conclusion - Une ouverture

La définition 3 et les résultats énoncés dans le lemme 2, les propriétés 6 et 7, restent valables si les variables $\{y^k\}$ ne sont pas les indicatrices d'une variable qualitative. Les ACP (14) et (15) peuvent donc être utilisées pour analyser l'association dissymétrique, mesurée par un coefficient CARD, entre une variable qualitative explicative x et des variables quantitatives à expliquer $\{y^k\}$. Cette analyse englobe, dans le cas particulier où $M_y = V_y^{-1}$ [cf. ACP (14)], l'Analyse Factorielle Discriminante (AFD) [6] : le coefficient d'association analysé est alors symétrique puisque le numérateur du coefficient CARD est égal au rapport de corrélation généralisé [cf. remarque 5b)].

On notera que les représentations simultanées et barycentriques sont les projections orthogonales, sur les plans principaux de l'ACP (15), du nuage $\{P_x(y_i); i = 1, n\} \cup \{P_x[g(y/x^j)]; j = 1, p\}$, alors que celles de l'AFD sont les projections orthogonales, sur les plans principaux de l'ACP (14), du nuage $\{g(y/x^j) = P_y[e_j(x)]; j = 1, p\} \cup \{y_i; i = 1, n\}$. La démarche pour analyser les résultats serait semblable à celle présentée dans les paragraphes 4.2 et 4.3.

On notera, enfin, que le lemme 2 et les propriétés 6 et 7 seraient des outils utiles pour aborder la «MANOVA» [5], ou encore pour écrire certains algorithmes de classification [13].

Annexe
Quelques éléments sur les inverses généralisées de Moore-Penrose pondérées [15,8,9]

1. Généralités

H et G sont deux espaces vectoriels munis respectivement des produits scalaires S et T, et A est une application de H dans G.

Définition : On dit que A^+ est l'Inverse Généralisée de Moore-Penrose (IGMP), pondérée[1] par le couple (S,T), de A si et seulement si :

$AA^+A = A \quad \Leftrightarrow A^+$ est une inverse généralisée interne (IGI) de A (1)

$A^+AA^+ = A^+ \quad \Leftrightarrow A^+$ est une inverse généralisée externe de A (2)

AA^+[resp. A^+A] est T-symétrique [resp. S-symétrique].

On déduit de (1) ou (2) l'idempotence des opérateurs AA^+ et A^+A .

De plus on a, $\mathrm{Im}AA^+ = \mathrm{Im}A$ car $\mathrm{Im}A \supset \mathrm{Im}AA^+ \supset \mathrm{Im}AA^+A = \mathrm{Im}A$, et de même $\mathrm{Im}A^+A = \mathrm{Im}A^+$.

On en déduit les propriétés suivantes.

Propriété 1 : Les deux assertions suivantes sont équivalentes :

a1) A^+ est l'IGMP, pondérée par (S,T), de A

a2) AA^+ est l'opérateur de projection T-orthogonale sur $\mathrm{Im}A$

A^+A est l'opérateur de projection S-orthogonale sur $\mathrm{Im}A^+$.

Propriété 2 : Si A^- est une inverse généralisée interne de A, alors AA^- est un projecteur sur $\mathrm{Im}A$ parallèlement à $\mathrm{Ker}\, AA^-$.

2. Quelques propriétés utiles en Analyse des Données

Soient $X_{(n,p)}$[resp. $\underline{X}_{(n,p)}$] la matrice des valeurs des indicatrices centrées [resp. non centrées] des p modalités d'une variable qualitative observée sur n individus et $D = 1/n\, I_n$ la matrice des poids des individus, où I_n est la matrice unité d'ordre n. On note $\{(\lambda_j, u_j); j = 1,p\}$ les valeurs propres et les vecteurs propres normés de l'opérateur VM de l'espace des individus $\mathbb{R}^p$, où M est l'isomorphisme associé au produit scalaire dans $\mathbb{R}^p$ et $V = {}^tXDX$. On a rang $[V] = p - 1$, VM n'est donc pas inversible. Il est cependant utile d'avoir une expression de son IGMP,

[1] On dit encorte inverse généralisée orthogonale.

pondérée par M^2, notée $(VM)^+$. Soit P_j l'opérateur de projection M-orthogonale sur $\{\alpha u_j / \alpha \in \mathbb{R}\}$, vu que $\mathrm{Im}[\sum_{\{j/\lambda_j \neq 0)} P_j] = \mathrm{Im}^t X$, il découle immédiatement de la propriété 1 et de $VM = \sum_{\{j/\lambda_j \neq 0\}} \lambda_j P_j$ que $(VM)^+ = \sum_{\{j/\lambda_j \neq 0\}} (1/\lambda_j) P_j$.

On en déduit que :

– P_j étant M-symétrique, $(VM)^+$ l'est aussi et $\mathrm{Im}(VM)^+ = \mathrm{Im}^t X$.

– $[(VM)^{1/2}]^+ = \sum_{\{j/\lambda_j \neq 0\}} (1/\sqrt{\lambda_j}) P_j$ est M-symétrique et $\mathrm{Im}[(VM)^{1/2}]^+ = \mathrm{Im}^t X$.

– $[(VM)^{1/2}]^+ (VM)^{1/2}$ est une expression de l'opérateur de projection M-orthogonale sur $\mathrm{Im}^t X$.

Sachant que ${}^t P_j$ est l'opérateur de projection M^{-1}-orthogonale sur $\{\alpha M u_j / \alpha \in \mathbb{R}\}$, vu que $\mathrm{Im}[M \sum_{\{j/\lambda_j \neq 0\}} (1/\lambda_j) P_j] = \mathrm{Im}[\sum_{\{j/\lambda_j \neq 0\}} (1/\lambda_j)\, {}^t P_j M] = \mathrm{Im}[\sum_{\{j/\lambda_j \neq 0\}} {}^t P_j]$ et en notant que $V = \sum_{\{j/\lambda_j \neq 0\}} \lambda_j P_j M^{-1}$, on montre en utilisant la propriété 1 que :

$V^+ = M \sum_{\{j/\lambda_j \neq 0\}} (1/\lambda_j) P_j$ est l'IGMP, pondérée par (M^{-1}, M), de V, et $\mathrm{Im}\, VV^+ = \mathrm{Im}^t X$.

On en déduit immédiatement,

a) ${}^t(V^+) = \sum_{\{j/\lambda_j \neq 0\}} (1/\lambda_j)\, {}^t P_j M = \sum_{\{j/\lambda_j \neq 0\}} (1/\lambda_j) M P_j = V^+$.

b) $(VM)^+ = M^{-1} V^+$.

c) $Q_x = X V^+ {}^t X D$ est une expression de l'opérateur de projection D-orthogonale sur $\mathrm{Im}\, X$: en effet l'idempotence de Q_x découle du fait que V^+ est une inverse généralisée externe, et la D-symétrie découle de la symétrie de V^+; de plus $\mathrm{Im}\, Q_x = \mathrm{Im}\, X$ car d'une part $\mathrm{Im}\, Q_x \subset \mathrm{Im}\, X$ et d'autre part $\mathrm{trace}\,[Q_x] = \mathrm{trace}[VV^+] = \dim[\mathrm{Im}\, VV^+] = \dim[\mathrm{Im}^t X]$.

Remarque 1 : Une inverse généralisée interne particulière

Partant de $\chi^2 = ({}^t \underline{X} D \underline{X})^{-1}$, en notant Q_{1n} l'opérateur de projection D-orthogonale sur $\Delta 1n$, on a,

---

[2] (M, M).

$$
\begin{aligned}
V\chi^2 V &= V\chi^2\, {}^t\underline{X}\, D(I_n - Q_{1n})\underline{X} \\
&= V - {}^t\underline{X}\, D(I_n - Q_{1n})Q_{\underline{x}}Q_{1n}\underline{X} \\
&\quad \text{où } Q_{\underline{x}} \text{ est l'opérateur de projection } D\text{-orthogonale sur Im } \underline{X} \\
&= V \quad \text{car } (I_n - Q_{1n})Q_{\underline{x}}Q_{1n} = (I_n - Q_{1n})Q_{1n} = 0, \text{ puisque } 1_n \in \operatorname{Im}\underline{X}.
\end{aligned}
$$

χ^2 est donc une IGI de V.

On en déduit que $V\chi^2$ est (cf. propriété 2) le projecteur sur $\mathrm{Im}V = \mathrm{Im}^t X$ parallèlement à $\mathrm{Ker}V\chi^2$.

Remarque 2 : Une simplification utile dans $E = E_x \oplus E_y$.

Notons V_x^-[resp. V_y^-] une IGI de plein rang de V_x[resp. V_y]. Vu que $V_x V_x^-$ est un projecteur sur $\mathrm{Im}^t X$,

si $M = R[V_x^-, M_y]$ alors $M_{xy} = V_x^- V_{xy} M_y [(V_y M_y)^{1/2}]^+$.

On a évidemment $M_{xy} = V_x^- V_{xy} V_y^-$ si $M = R[V_x^-, V_y^-]$.

Remerciements

Nous remercions P. CAZES pour les remarques et les conseils qu'il nous a très aimablement prodigués.

Références

[1] ABDESSELAM R. (1988), *Contribution à l'analyse des associations dissymétriques*. Thèse 3e cycle, Université Paul Sabatier (UPS), Toulouse.

[2] ABDESSELAM R., SCHEKTMAN Y. (1989), *Dissymmetrical association analysis between two qualitatives variables*. Data Analysis, Learning Symbolic and Numeric Knowledge, Diday E. (Eds.), INRIA – Nova Science Publishers, New-York. Budapest, 39-46.

[3] BENZECRI F. (1980), *Introduction à l'analyse des correspondances d'après un exemple de données médicales*. [INT. CORR. MED.], les Cahiers de l'Analyse des Données, Vol. 5, N° 3, 283-310.

[4] BENZECRI J-P. (1982), *L'Analyse des Données : l'analyse des correspondances*, (Tome 2), Dunod, 4e Edition.

[5] CADET O, SCHEKTMAN Y. (1989), *A method for analysing multidimensional experimental data.* Data Analysis, Learning Symbolic and Numeric Knowledge, Diday E. (Eds.), INRIA – Nova Science Publishers, New-York. Budapest, 87-94.

[6] CAILLIEZ F., PAGÈS J.-P. (1976), *Introduction à l'Analyse des Données*, Smash, ASU, Buro, Paris.

[7] CAZES P. (1970), *Application de l'Analyse des Données au traitement de problèmes géologiques*. Thèse 3ème cycle, Université Paris VI.

[8] CROQUETTE A. (1980) *Quelques résultats synthétiques en Analyse des Données Multidimensionnelles : Optimalité et Métriques à effets relationnels.* Thèse 3e cycle, UPS, Toulouse

[9] CROQUETTE A. (1983) *Cours d'Analyse des Données*. UPS, Toulouse

[10] D'AMBRA L, LAURO N. (1989), *Non symmetrical analysis of three-way contingency tables*. Multiway Data Analysis, Coppi R. & Bolasco S. Eds., North-Holland, Amsterdam, 301-315.

[11] FABRE C. (1986), *Contribution à la protection des méthodes relationnelles.* Thèse 3e cycle, UPS, Toulouse.

[12] GOODMAN L.A., KRUSKAL W.H. (1954), *Measures of association for cross classifications*. J.A.S.A., 49, 732-764.

[13] LABRÈCHE S., SCHEKTMAN Y., TRÉJOS J., TROUPÉ M. (1992), *Les distances relationnelles : deux applications récentes*. Congrès International sur analyse en distance, Distancia'92, Joly S & Le Calve G, (Eds), Rennes, 369-372.

[14] LAURO N., D'AMBRA L. (1983), *L'Analyse non symétrique des correspondances*. Data Analysis and Informatics. E. Diday and al. Eds, North-Holland, Amsterdam, 433-446.

[15] NASHED M.Z. (1976), *Generalized Inverses and Applications*. Proceedings of an Advanced Seminar – Mathematics Research Center – The University of Winconsin – Madison, Academic Press, New York – San Francisco – London.

[16] SCHEKTMAN Y. (1978), *Contribution à la mesure en facteurs dans les sciences expérimentales et à la mise en œuvre des calculs statistiques*. Thèse de Doctorat d'Etat, UPS, Toulouse.

[17] SCHEKTMAN Y. (1988), *A general euclidean approach for measuring and describing associations between several sets of variables*. Recent Developments in Clustering and Data Analysis, C. Hayashi and al. Eds, Academic Press, Inc., Tokyo, 37-48.

[18] SCHEKTMAN Y., FABRE C. (1984), *Un point de vue métrique sur la protection des modèles relationnels*. Résumé des communications des Journées de Statistique, Montpellier, 72.

[19] SCHEKTMAN Y. (1989), *Cours de DEA. Analyse et traitement informatique de données*. UPS, Toulouse

[20] SCHEKTMAN Y. (1989), *Inner products and association indices useful for analysing some multiway tables*. Multiway Data Analysis, Coppi R. & Bolasco S. Eds., North-Holland, Amsterdam, 203- 212.

[21] SCHEKTMAN Y. (1991), *Eléments mathématiques de base pour la définition de semi-produits scalaires utiles en Analyse de données*. Quelques applications. Note interne, Laboratoire MLAD, UPS, Toulouse.

[22] SCHEKTMAN Y. (1994), *Propriétés des produits scalaires relationnels*. Note interne, Laboratoire Lemme-DIEM, UPS, Toulouse (47 pages).

[23] STEWART D., LOVE W. (1968), *A general canonical correlation index*. Psychological Bull, vol. 70, 160-163.