

REVUE DE STATISTIQUE APPLIQUÉE

F. MARCOTORCHINO

N. EL AYOUBI

Paradigme logique des écritures relationnelles de quelques critères fondamentaux d'association

Revue de statistique appliquée, tome 39, n° 2 (1991), p. 25-46

http://www.numdam.org/item?id=RSA_1991__39_2_25_0

© Société française de statistique, 1991, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

PARADIGME LOGIQUE DES ÉCRITURES RELATIONNELLES DE QUELQUES CRITÈRES FONDAMENTAUX D'ASSOCIATION

F. MARCOTORCHINO

*CEMAP IBM France, 68/76 quai de la Rapée, Rives de Seine
75592 Paris Cedex 12*

N. EL AYOUBI

*Université Pierre et Marie Curie, Paris VI
Laboratoire de Statistique Théorique et Appliquée*

I. Introduction

L'Analyse Relationnelle¹ développée par F. Marcotorchino et P. Michaud depuis 1978 trouve des applications dans de nombreux domaines tels que l'agrégation des similarités ou des préférences, la linéarisation des critères d'association sur tableaux de contingence, la sériation, la quadri-décomposition, etc.

Plus récemment, nous avons montré dans : [5] et [9] qu'il existe un pont, mathématiquement validé, entre l'Analyse Relationnelle et l'Analyse Factorielle.

L'objectif de cet article, faisant suite aux récents travaux de P. Michaud, F. Marcotorchino, C. Bédécarrax, H. Messatfa, N. EL Ayoubi, etc, est de montrer comment les écritures relationnelles des critères fondamentaux d'association établissent un compromis, ou encore éliminent le paradoxe, entre deux écritures contingentiels de ces critères.

Nous nous intéresserons au critère de Rand et à ses dérivés ainsi qu'au critère du Chi-deux. Nous décomposerons leurs écritures contingentiels et relationnelles sur la base du critère de Condorcet; c'est-à-dire en une somme de deux associations, l'une positive et l'autre négative.

Dans [13], en utilisant le concept de comparaisons par paires, nous avons prouvé que le critère de Condorcet était égal au numérateur du critère de Rand. Ceci montre que le critère de Condorcet, qui est un critère logique, possède une écriture contingentielle et que le critère de Rand admet une écriture linéaire en

¹ Il s'agit d'une méthode d'Analyse des Données fondée sur le principe de présentation relationnelle ou logique des données (qui doivent être par nature qualitatives), de critères d'association ou d'agrégation dont l'optimisation par rapport à une relation cherchée (équivalence ou ordre total) s'effectue par programmation linéaire. De nombreuses références à cette méthodologie sont données en bibliographie.

fonction du terme général du tableau relationnel ² représentant la partition supposée inconnue.

Partant de l'écriture modifiée du critère de Rand (voir [12]), nous allons donner une décomposition contingentielle de ce critère. Cette décomposition nous permettra de retrouver le critère du Chi-deux par pondération en lignes et en colonnes.

Afin de montrer ce résultat, il est nécessaire d'étudier le problème à la base et de s'interroger sur les variables sur lesquelles nous travaillons et sur les tableaux associés.

Les variables considérées sont qualitatives nominales. Une variable est dite qualitative lorsqu'elle prend des modalités numériques; suivant les modalités on a affaire à :

1. une variable nominale si les modalités ne sont pas fixées a priori; par exemple les catégories socio-professionnelles
2. une variable ordinale si les modalités sont ordonnées; par exemple des classements, des notes
3. une variable binaire si elle prend seulement les valeurs 0 ou 1; c'est le cas du codage disjonctif.

On ne peut pas travailler directement sur de telles variables; il faut les transformer au préalable en d'autres variables, autrement dit les coder.

Il existe plusieurs représentations suivant l'outil mathématique utilisé. Dans cette étude nous nous intéresserons aux représentations logiques : les tableaux relationnels (tableaux de Condorcet) et les tableaux disjonctifs.

Nous venons de parler des tableaux de Condorcet et des tableaux disjonctifs sans les définir. Le propos du paragraphe suivant est de présenter : le tableau de Condorcet (noté C), le tableau de Condorcet pondéré (noté \hat{C}), et le tableau disjonctif (noté K).

Si l'on se situe dans le cadre du croisement de deux variables nominales, c'est le tableau de contingence que nous étudierons. Nous reprenons les notations et les formules de passage contingences-paires présentées dans [13]. Nous définirons et démontrerons également d'autres formules qui sont importantes pour la suite.

Les quatre tableaux dont nous venons de parler sont appelés : les tableaux basiques.

2. Les tableaux basiques

Supposons que nous disposions de m variables nominales, notées $C^1, \dots, C^k, \dots, C^m$, observées sur le même ensemble d'individus I , ayant respectivement

² Que l'on appelle également tableau de Condorcet ou tableau des comparaisons par paires (voir [15] et [16]).

$p_1, \dots, p_k, \dots, p_m$ modalités dont la somme est donnée par :

$$(1) \quad \sum_{k=1}^m p_k = p$$

Nous poserons pour tout le texte :

N = nombre d'individus

m = nombre de variables

p = nombre total de modalités

A partir de ces quantités nous allons rappeler les propriétés des tableaux K , C et \hat{C} .

2.1 Le tableau disjonctif

Le tableau K est associé aux m variables. Il est de dimension $N \times p$. Son terme général, noté k_{ij} , est défini par :

$$k_{ij} = \begin{cases} 1 & \text{si l'individu } i \text{ possède la modalité } j \\ 0 & \text{sinon} \end{cases}$$

Il possède les propriétés suivantes :

- L'effectif marginal d'un individu est égal au nombre de variables, soit :

$$(2) \quad k_{i\bullet} = \sum_{j=1}^p k_{ij} = m$$

Chaque individu possède en effet une seule modalité de chaque variable.

- L'effectif marginal d'une modalité est donné par :

$$(3) \quad k_{\bullet j} = \sum_{i=1}^N k_{ij} = \text{nombre d'individus qui possèdent la modalité } j$$

L'effectif total du tableau K est égal au produit du nombre de variables et du nombre d'individus.

$$(4) \quad k_{\bullet\bullet} = \sum_{i=1}^N \sum_{j=1}^p k_{ij} = \sum_{i=1}^N k_{i\bullet} = \sum_{j=1}^p k_{\bullet j} = mN$$

2.2 Le tableau de Condorcet

Chacune des variables considérées sera représentée par un tableau relationnel³ dans l'espace des individus, de dimension $N \times N$, dont le terme général est défini⁴ comme suit :

$$c_{ii'}^k = \begin{cases} 1 & \text{si les individus } i \text{ et } i' \text{ possèdent la même modalité de la variable } C^k \\ 0 & \text{sinon} \end{cases}$$

La somme de ces tableaux est appelée tableau collectif. Son terme général, noté $c_{ii'}$, est la somme des termes généraux des tableaux relationnels $C^1, \dots, C^k, \dots, C^m$, c'est-à-dire :

$$(5) \quad c_{ii'} = \sum_{k=1}^m c_{ii'}^k$$

En d'autres termes :

$c_{ii'}$ = nombre de variables pour lesquelles les individus i et i' possèdent la même modalité .

Les tableaux C et K sont liés de la façon suivante (lorsque C est la somme de variables⁵ qualitatives) :

$$(6) \quad c_{ii'} = \sum_{j=1}^p k_{ij} k_{i'j}$$

que l'on peut écrire matriciellement sous la forme :

$$C = KK'$$

où K' est la matrice transposée de K .

Les tableaux collectifs et individuels possèdent des propriétés structurelles et relationnelles. Nous reprenons ici quelques propriétés essentielles démontrées par J-F. Marcotorchino dans [13].

1. Le terme général de la diagonale est égal au nombre de variables :

$$(7) \quad c_{ii} = m \quad (\forall i \in I)$$

2. Le tableau C est symétrique, c'est-à-dire :

$$(8) \quad c_{ii'} = c_{i'i} \quad (\forall i, i' \in I)$$

³ $C^1, \dots, C^k, \dots, C^m$ désigneront à la fois les variables et les tableaux relationnels qui leur sont associés (ces tableaux sont également appelés tableaux de Condorcet ou tableaux de comparaisons par paires).

⁴ Voir [15] et [16].

⁵ C peut représenter une somme de tableaux relationnels non équivalents à des variables à modalités donc plus généraux. C'est ce qui rend l'approche relationnelle plus générale que l'approche par tableaux disjonctifs.

3. Le tableau C vérifie la propriété de transitivité globale :

$$(9) \quad c_{ii'} + c_{i'i''} - c_{ii''} \leq m \quad (\forall i, i', i'' \in I)$$

4. la somme totale des éléments de C est donnée par :

$$(10) \quad c_{..} = \sum_{i=1}^N \sum_{i'=1}^N c_{ii'} = \sum_{j=1}^p k_{\bullet j}^2$$

Les variables sont toutes des variables à modalités, c'est-à-dire des partitions; on peut donc, selon [13], associer au tableau C son tableau complémentaire, noté \bar{C} , dont le terme général est défini par :

$$(11) \quad \bar{c}_{ii'} = \sum_{k=1}^m \bar{c}_{ii'}^k = m - c_{ii'}$$

où

$$\bar{c}_{ii'}^k = 1 - c_{ii'}^k = \begin{cases} 0 & \text{si les individus } i \text{ et } i' \text{ possèdent la même modalité de la} \\ & \text{variable } C^k \\ 1 & \text{sinon} \end{cases}$$

$\bar{c}_{ii'}$ peut également être défini à partir du terme général du tableau K , de la façon suivante :

$$\bar{c}_{ii'} = \frac{1}{2} \sum_{j=1}^p (k_{ij} - k_{i'j})^2$$

On en déduit que le tableau \bar{C} est un tableau de distances entre les individus.

2.3 Le tableau de Condorcet pondéré

Le tableau de Condorcet pondéré, noté \hat{C} , peut se définir de deux façons ⁶

1. soit à partir du tableau relationnel individuel par pondération de son terme général par la moyenne arithmétique des effectifs des classes correspondant à ses indices,
2. soit à partir du tableau disjonctif complet par pondération de la contribution par colonne d'une modalité.

En d'autres termes :

$$(12) \quad \hat{c}_{ii'} = \sum_{k=1}^m \frac{2c_{ii'}^k}{c_{i\bullet}^k + c_{\bullet i'}^k} = \sum_{j=1}^p \frac{k_{ij}k_{i'j}}{k_{\bullet j}}$$

⁶ Voir [5] et [9].

Soit encore, si nous notons $\hat{c}_{ii'}^k = \frac{2c_{ii'}^k}{c_{i\bullet}^k + c_{\bullet i'}^k}$:

$$(13) \quad \hat{c}_{ii'} = \sum_{k=1}^m \hat{c}_{ii'}^k$$

A l'instar des tableaux K et C , le tableau \hat{C} possède des propriétés structurelles et relationnelles démontrées dans [5] et [9]. Nous allons en citer quelques-unes :

1. Le terme général de la diagonale est égal à la somme des inverses des effectifs des classes correspondant à ses indices :

$$(14) \quad \hat{c}_{ii} = \sum_{k=1}^m \frac{1}{c_{i\bullet}^k} \quad \forall i \in I$$

2. La valeur marginale pour un individu est égale au nombre de variables, soit :

$$(15) \quad \hat{c}_{i\bullet} = m \quad \forall i \in I$$

$$\text{Car : } \hat{c}_{i\bullet} = \sum_{i'=1}^N \hat{c}_{ii'} = \sum_{i'=1}^N \sum_{j=1}^p \frac{k_{ij}k_{i'j}}{k_{\bullet j}} = \sum_{j=1}^p k_{ij} = k_{i\bullet} = m$$

Ceci est dû aux formules (12), (3) et (2).

3. La somme totale du tableau \hat{C} est égale à la somme totale du tableau K .

$$(16) \quad \hat{c}_{\bullet\bullet} = mN$$

En effet :

$$\hat{c}_{\bullet\bullet} = \sum_{i=1}^N \sum_{i'=1}^N \hat{c}_{ii'} = \sum_{i=1}^N \hat{c}_{i\bullet} = \sum_{i=1}^N m = mN$$

Ceci résulte de la double somme sur i et i' de la formule (15).

4. Le tableau \hat{C} possède des propriétés relationnelles que l'on peut qualifier de floues, à savoir :

$$(17-1) \quad 0 \leq \hat{c}_{ii} \leq 1 \quad (\forall i \in I)$$

La symétrie floue :

$$(17-2) \quad \hat{c}_{ii'} = \hat{c}_{i'i} \quad (\forall i, i' \in I)$$

La transitivité floue :

$$(17-3) \quad \hat{c}_{ii'} \geq \min(\hat{c}_{ii''}, \hat{c}_{i''i'}) \quad (\forall i, i', i'' \in I)$$

Au tableau \widehat{C} est associé le **tableau complémentaire**, noté $\overline{\widehat{C}}$. Son terme général est défini comme suit :

$$(18) \quad \overline{\widehat{c}}_{ii'} = \frac{\widehat{c}_{ii} + \widehat{c}_{i'i'}}{2} - \widehat{c}_{ii'}$$

que nous pouvons écrire également, à partir du terme général du tableau K , de la façon suivante :

$$(19) \quad \overline{\widehat{c}}_{ii'} = \sum_{j=1}^p \frac{\left(\frac{k_{ij} + k_{i'j}}{2} - k_{ij}k_{i'j} \right)}{k_{\bullet j}}$$

A partir de cette formule et du fait que $k_{ij} = k_{ij}^2 \forall i, j$ (car $k_{ij} = 1$ ou 0), nous déduisons que :

$$(20) \quad \overline{\widehat{c}}_{ii'} = \frac{1}{2} \sum_{j=1}^p \frac{1}{k_{\bullet j}} (k_{ij} - k_{i'j})^2$$

Il en résulte donc que ce terme est égal, à la constante $\frac{m}{2N}$ près, à la distance du Chi-deux entre les profils des individus i et i' (voir [1]).

La somme totale de $\overline{\widehat{C}}$ est donnée par :

$$(21) \quad \overline{\widehat{c}}_{\bullet\bullet} = \sum_{i=1}^N \sum_{i'=1}^N \overline{\widehat{c}}_{ii'} = pN - mN$$

2.4 Le tableau de contingence et les formules de linéarisation

2.4.1 Le tableau de contingence

Supposons ici que C représente une variable à p modalités et que X soit une partition à q classes de l'ensemble I .

Le tableau de contingence croisant C et X est alors de dimension $p \times q$. Il est caractérisé par :

- Le terme général :

n_{uv} = l'effectif de la case (u, v)

Il est lié aux termes généraux des tableaux disjonctifs associés à C et X par la formule suivante :

$$(22) \quad n_{uv} = \sum_{i=1}^N k_{iu}k_{iv}$$

- Le terme général marginal associé aux colonnes :
 $n_{u\bullet}$ = nombre d'individus ayant la modalité u de C .
 Soit encore :

$$(23) \quad n_{u\bullet} = \sum_{v=1}^p n_{uv} = \sum_{i=1}^N k_{iu} = k_{\bullet u}$$

- Le terme général marginal associé aux lignes :
 $n_{\bullet v}$ = nombre d'individus ayant la modalité v de X

$$(24) \quad n_{\bullet v} = \sum_{u=1}^p n_{uv} = \sum_{i=1}^N k_{iv} = k_{\bullet v}$$

- La somme totale :

$$(25) \quad n_{\bullet\bullet} = \sum_{u=1}^q \sum_{v=1}^q n_{uv} = \sum_{u=1}^p n_{u\bullet} = \sum_{v=1}^q n_{\bullet v} = N$$

2.4.2 Les formules de linéarisation

Selon [13], le terme **linéarisation** doit être compris par rapport au terme général du tableau de comparaisons par paires représentant la variable X supposée inconnue.

Le tableau C représente une seule relation d'équivalence c'est-à-dire $C = C^k$, pour un k fixé. Ce tableau C n'est pas le tableau de Condorcet défini précédemment, mais le tableau relationnel représentant une variable V qualitative donnée.

Si l'on superpose les tableaux relationnels C et X , on a la même information que dans le tableau de contingence croisant les variables C et X .

Des formules de passage contingences-paires ont été proposées et démontrées par **Kendall** ([7] pour la formule (26)) et par **Marcotorchino** ([13] pour (27), (28), (29), (30), (31) et (32)). Rappelons celles qui sont utiles dans le cadre de cette étude :

$$(26) \quad \sum_{u=1}^p \sum_{v=1}^q n_{uv}^2 = \sum_{i=1}^N \sum_{i'=1}^N c_{ii'} x_{ii'}$$

$$(27) \quad \sum_{u=1}^p n_{u\bullet}^2 = \sum_{i=1}^N \sum_{i'=1}^N c_{ii'}$$

$$(28) \quad \sum_{v=1}^q n_{\bullet v}^2 = \sum_{i=1}^N \sum_{i'=1}^N x_{ii'}$$

$$(29) \quad \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}^2}{n_{u\bullet}} = \sum_{i=1}^N \sum_{i'=1}^N \frac{2c_{ii'}}{c_{i\bullet} + c_{\bullet i'}} x_{ii'}$$

$$(30) \quad \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}^2}{n_{u\bullet} n_{\bullet v}} = \sum_{i=1}^N \sum_{i'=1}^N \frac{c_{ii'}}{c_{i\bullet} x_{\bullet i'}}$$

$$(31) \quad p = \sum_{i=1}^N \frac{1}{c_{i\bullet}}$$

$$(32) \quad q = \sum_{i=1}^N \frac{1}{x_{i\bullet}}$$

A partir de la définition du terme général du tableau \widehat{C} donnée par la formule (12) (avec $m = 1$ où \widehat{C} représente une seule variable), les formules (29), (30), (31) et (32) peuvent s'écrire :

$$(29') \quad \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}^2}{n_{u\bullet}} = \sum_{i=1}^N \sum_{i'=1}^N \widehat{c}_{ii'}, x_{ii'}$$

$$(30') \quad \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}^2}{n_{u\bullet} n_{\bullet v}} = \sum_{i=1}^N \sum_{i'=1}^N \widehat{c}_{ii'}, \widehat{x}_{ii'}$$

$$(31') \quad p = \sum_{i=1}^N \widehat{c}_{ii} \quad (\forall i \in I)$$

$$(32') \quad q = \sum_{i=1}^N \widehat{x}_{ii} \quad (\forall i \in I)$$

En tenant compte de la formule (15) ($m = 1$), les deux dernières formules s'expriment comme suit :

$$(31'') \quad p = \sum_{i=1}^N \sum_{i'=1}^N \frac{\widehat{c}_{ii} + \widehat{c}_{i'i'}}{2} \widehat{x}_{ii'}$$

$$(32'') \quad q = \sum_{i=1}^N \sum_{i'=1}^N \frac{\widehat{x}_{ii} + \widehat{x}_{i'i'}}{2} \widehat{c}_{ii'}$$

En effet, le membre de droite de la formule (31'') s'écrit :

$$\sum_{i=1}^N \sum_{i'=1}^N \frac{\widehat{c}_{ii}}{2} \widehat{x}_{ii'} + \sum_{i=1}^N \sum_{i'=1}^N \frac{\widehat{c}_{i'i'}}{2} \widehat{x}_{ii'}$$

Il se simplifie, en effectuant respectivement les sommes sur i' et i de $\widehat{x}_{ii'}$

$$\sum_{i=1}^N \frac{\widehat{c}_{ii}}{2} \widehat{x}_{i\bullet} + \sum_{i'=1}^N \frac{\widehat{c}_{i'i'}}{2} \widehat{x}_{\bullet i'}$$

Ce qui donne, en remplaçant $\widehat{x}_{i\bullet}$ et $\widehat{x}_{\bullet i'}$ par leurs valeurs données par la formule (15) ($m = 1$) et selon la formule (31') :

$$\sum_{i=1}^N \frac{\widehat{c}_{ii}}{2} + \sum_{i'=1}^N \frac{\widehat{c}_{i'i'}}{2} = \frac{p}{2} + \frac{p}{2} = p \quad C.Q.F.D.$$

Même démonstration pour la formule (32'').

A partir de ces formules nous proposons et démontrons la formule suivante :

$$(33) \quad \sum_{u=1}^q \sum_{v=1}^q \frac{n_{uv} n_{\bullet v}}{n_{u\bullet}} = \sum_{i=1}^N \sum_{i'=1}^N \frac{\widehat{c}_{ii} + \widehat{c}_{i'i'}}{2} x_{ii'}$$

En effet, en utilisant les formules (22), (23) et (24), le membre de gauche de la formule peut s'écrire en notations disjonctives de la façon suivante :

$$\sum_{u=1}^p \sum_{v=1}^q \frac{\sum_i k_{iu} k_{iv} \sum_i k_{i'v}}{\sum_{i''} k_{i''u}}$$

Soit après permutation des sommations et changement de position des termes et en vertu de la formule (23) :

$$\sum_{i=1}^N \sum_{i'=1}^N \sum_{u=1}^p \frac{k_{iu}}{k_{\bullet u}} \sum_{v=1}^q k_{iv} k_{i'v}$$

quantité que nous pouvons écrire en vertu de (6) et (12) :

$$\sum_{i=1}^N \sum_{i'=1}^N \widehat{c}_{ii} x_{ii'}$$

ou encore, du fait de la symétrie du tableau X :

$$\sum_{i=1}^N \sum_{i'=1}^N \frac{\widehat{c}_{ii} + \widehat{c}_{i'i'}}{2} x_{ii'} \quad C.Q.F.D.$$

De la même façon, nous pouvons montrer que :

$$(34) \quad \sum_{u=1}^p \sum_{v=1}^q \frac{N n_{uv}}{n_{u\bullet} n_{\bullet v}} = \sum_{i=1}^N \sum_{i'=1}^N \hat{c}_{ii} \hat{x}_{ii}$$

3. Paradoxe de l'écriture contingentielle

De nombreux critères d'association sont calculés sur tableaux de contingence (voir [7] et [12]). Nous ne prétendons pas donner ici un inventaire correspondant à l'intégralité de la production scientifique dans le domaine des critères d'association statistiques. Nous ne nous intéresserons qu'aux critères de **Rand** et du **Chi-deux**, puisqu'ils sont au centre de cet article.

Nous allons démontrer tout d'abord que le critère de Rand admet deux décompositions. Puis en effectuant les pondérations de ces deux nouvelles écritures contingentielles, nous montrerons qu'elles conduisent à un paradoxe. Ensuite, nous établirons le critère du Chi-deux à partir du critère de Rand pondéré.

3.1 Décomposition contingentielle du critère de Rand

En 1971, Rand ([21]) a introduit un critère d'association entre deux variables nominales, par exemple C et X , sous la forme contingentielle suivante :

$$R'(C, X) = \frac{1}{N(N-1)} \left[2 \sum_{u=1}^p \sum_{v=1}^q n_{uv}^2 - \sum_{u=1}^p n_{u\bullet}^2 - \sum_{v=1}^q n_{\bullet v}^2 + N(N-1) \right]$$

En 1984, pour établir l'équivalence avec le critère de **Condorcet** ([12]) nous avons modifié ce critère en :

$$R(C, X) = \frac{1}{N^2} \left[2 \sum_{u=1}^p \sum_{v=1}^q n_{uv}^2 - \sum_{u=1}^p n_{u\bullet}^2 - \sum_{v=1}^q n_{\bullet v}^2 + N^2 \right]$$

Ce critère est toujours inférieur à 1 et égal à 1 lorsque $C \equiv X$. A partir de cette formule nous allons établir les décompositions du critère de Rand mentionnées plus haut.

Première décomposition

En utilisant les formules (23), (24) et (25) et en mettant en facteur la double somme, nous pouvons reformuler l'égalité précédente de la façon suivante :

$$R(C, X) = \frac{1}{N^2} \sum_{u=1}^p \sum_{v=1}^q (2n_{uv}^2 - n_{u\bullet} n_{uv} - n_{\bullet v} n_{uv} + n_{u\bullet} n_{\bullet v})$$

Puis en effectuant la factorisation des termes entre crochets, il découle que :

$$(35) \quad R(C, X) = \frac{1}{N^2} \sum_{u=1}^p \sum_{v=1}^q [n_{uv}^2 + (n_{u\bullet} - n_{uv})(n_{\bullet v} - n_{uv})]$$

Ce qui donne, en effectuant les sommes sur u et v des termes entre crochets :

$$(36) \quad R(C, X) = \frac{1}{N^2} \sum_{u=1}^p \sum_{v=1}^q n_{uv}^2 + \frac{1}{N^2} \sum_{u=1}^p \sum_{v=1}^q (n_{u\bullet} - n_{uv})(n_{\bullet v} - n_{uv})$$

Si nous notons :

$N_{uv} = n_{uv}$ = nombre d'individus qui possèdent simultanément les modalités u de C et v de X ,

$N_{u\bar{v}} = n_{u\bullet} - n_{uv}$ = nombre d'individus possédant la modalité u de C et ne possédant pas la modalité v de X et

$N_{\bar{u}v} = n_{\bullet v} - n_{uv}$ = nombre d'individus ayant la modalité v et n'ayant pas la modalité u ,

Nous obtenons :

$$R(C, X) = \frac{1}{N^2} \sum_{u=1}^p \sum_{v=1}^q N_{uv}^2 + \frac{1}{N^2} \sum_{u=1}^p \sum_{v=1}^q N_{u\bar{v}} N_{\bar{u}v}$$

Cette formulation du critère de Rand nous permet d'énoncer le théorème suivant :

Théorème 1

Le critère de Rand se décompose en une mesure d'association entre les individus qui ont simultanément des modalités des deux variables plus une mesure d'association entre les individus qui possèdent des modalités d'une variable et ne possèdent pas de modalités de l'autre.

D'où, nous déduisons que :

le critère de Rand peut être décomposé en une somme des carrés des éléments du tableau de contingence (croisant C et X) divisé par N^2 , plus une somme des produits des complémentaires des éléments n_{uv} aux effectifs marginaux correspondant à ces indices divisée par N^2 .

Nous avons donc une décomposition analogue à la décomposition du critère de Rand en notations relationnelles, démontrée dans [5], qui se présente sous la forme :

$$R(C, X) = \frac{1}{N^2} \sum_{i=1}^N \sum_{i'=1}^N c_{ii'} x_{ii'} + \frac{1}{N^2} \sum_{i=1}^N \sum_{i'=1}^N \bar{c}_{ii'} \bar{x}_{ii'} = \frac{1}{N^2} C(X)$$

qui n'est autre, à la constante $\frac{1}{N^2}$, que le critère de Condorcet lorsque nous ne considérons qu'une variable C^k .

D'où **le premier résultat**

Le critère de Rand décomposable en deux parties contingentielles est équivalent, dans son écriture relationnelle, au critère de Condorcet $C(X)$ lorsque nous croisons une variable qualitative C et une variable qualitative X .

Deuxième décomposition

Par ailleurs, il est facile de voir que :

$$(37) \quad \sum_{u=1}^p \sum_{v=1}^q (n_{u\bullet} - n_{uv})(n_{\bullet v} - n_{uv}) = \sum_{u=1}^p \sum_{v=1}^q n_{uv}(N - n_{u\bullet} - n_{\bullet v} + n_{uv})$$

Ainsi, à partir de la décomposition donnée par la formule (36) et en utilisant la formule (37), le critère de Rand s'écrit de la façon suivante :

$$R(C, X) = \frac{1}{N^2} \sum_{u=1}^p \sum_{v=1}^q n_{uv}^2 + \frac{1}{N^2} \sum_{u=1}^p \sum_{v=1}^q n_{uv}(N - n_{u\bullet} - n_{\bullet v} + n_{uv})$$

Ce qui donne, en mettant en facteur $\frac{1}{N^2}$ et la double somme :

$$(38) \quad R(C, X) = \frac{1}{N^2} \sum_{u=1}^p \sum_{v=1}^q [n_{uv}^2 + n_{uv}(N - n_{u\bullet} - n_{\bullet v} + n_{uv})]$$

3.2 Les critères pondérés dérivés du critère de Rand

A partir du critère de Rand calculé sur le tableau de contingence croisant les modalités de deux variables nominales, nous considérons deux types de pondérations possibles :

1. soit par la contribution d'une modalité de l'une des deux variables,
2. soit par les contributions des deux modalités des deux variables.

Nous allons donc expliciter le critère de Rand pondéré, à partir des contributions des modalités de la variable C dans le premier cas, et à partir des contributions des modalités de C et X dans le deuxième cas.

3.2.1 Le critère de Rand pondéré en C

En effectuant la pondération des termes entre crochets de la formule (35) ou (38) par la contribution de la modalité u , nous obtenons un critère de Rand pondéré⁷ que nous notons $R(\hat{C}, X)$.

$$(39) \quad R(\hat{C}, X) = \frac{1}{Np} \left[\sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}^2}{n_{u\bullet}} + \sum_{u=1}^p \sum_{v=1}^q \frac{1}{n_{u\bullet}} (n_{u\bullet} - n_{uv})(n_{\bullet v} - n_{uv}) \right]$$

⁷ Nous avons divisé le numérateur du critère de Rand par Np au lieu de N^2 pour avoir un critère normé par 1

qui s'écrit également de la façon suivante :

$$R(\widehat{C}, X) = \frac{1}{Np} \left[\sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}^2}{n_{u\bullet}} + \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}}{n_{u\bullet}} (N - n_{u\bullet} - n_{\bullet v} + n_{uv}) \right]$$

Soit, après développement et selon la formule (25) :

$$(40) \quad R(\widehat{C}, X) = \frac{1}{Np} \left[\sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}^2}{n_{u\bullet}} + \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}^2}{n_{u\bullet}} - N - \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv} n_{\bullet v}}{n_{u\bullet}} + pN \right]$$

3.2.2 Ecriture relationnelle des critères de Rand et de Condorcet pondérés

Partant de la formule (40) du critère de Rand pondéré, en exploitant les formules (29'), (16), (33) et (31') et en mettant en facteur la double somme, nous obtenons :

$$R(\widehat{C}, X) = \frac{1}{Np} \sum_{i=1}^N \sum_{i'=1}^N \left[\widehat{c}_{ii'} x_{ii'} + \widehat{c}_{i'i'} x_{i'i'} - \widehat{c}_{ii'} - \frac{\widehat{c}_{ii} + \widehat{c}_{i'i'}}{2} x_{ii'} + \frac{\widehat{c}_{ii} + \widehat{c}_{i'i'}}{2} \right]$$

D'où il découle, après factorisation du second membre et en vertu des formules (11) et (18), que :

$$R(\widehat{C}, X) = \frac{1}{Np} \sum_{i=1}^N \sum_{i'=1}^N (\widehat{c}_{ii'} x_{ii'} + \widehat{c}_{i'i'} \bar{x}_{i'i'}) = \frac{1}{Np} C_p(X)$$

Et l'on déduit **Le deuxième résultat** :

Le critère de Rand pondéré est égal, à la constante $\frac{1}{Np}$ près, au critère de Condorcet pondéré $C_p(X)$ présenté dans [5] et [9].

3.2.3 Le critère de Rand pondéré en C et X

Dans la formule (35) ou (38), si nous effectuons la pondération des termes entre crochets par l'inverse du produit des effectifs des modalités u de C et v de X . Nous obtenons un critère doublement pondéré de Rand⁸ noté $R(\widehat{C}, \widehat{X})$.

Ce qui donne, en effectuant les sommes sur u et v .

$$1) R_1(\widehat{C}, \widehat{X}) = \frac{1}{p^2} \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}^2}{n_{u\bullet} n_{\bullet v}} + \frac{1}{p^2} \sum_{u=1}^p \sum_{v=1}^q \frac{(n_{u\bullet} - n_{uv})(n_{\bullet v} - n_{uv})}{n_{u\bullet} n_{\bullet v}}$$

⁸ Pour avoir un critère normé par 1, il faut diviser le numérateur du critère de Rand par p^2 , car :

1) $R_1(\widehat{C}, \widehat{X}) = 1$ si $C \equiv X$

2) $R_2(\widehat{C}, \widehat{X}) = 1$ si $C \equiv X$ et dans le cas optimal.

$$2) R_1(\widehat{C}, \widehat{X}) = \frac{1}{p^2} \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}^2}{n_{u\bullet} n_{\bullet v}} + \frac{1}{p^2} \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}(N - n_{u\bullet} - n_{\bullet v} + n_{uv})}{n_{u\bullet} n_{\bullet v}}$$

Après développement et compte tenu des formules (23) et (24), le critère $R_1(\widehat{C}, \widehat{X})$ se présente comme suit :

$$(41) \quad R_1(\widehat{C}, \widehat{X}) = \frac{1}{p^2} \left[\sum_{u=1}^p \sum_{v=1}^q 2 \frac{n_{uv}^2}{n_{u\bullet} n_{\bullet v}} + pq - p - q \right]$$

Mais en effectuant la pondération dans la formule (38), le critère doublement pondéré de Rand $R_2(\widehat{C}, \widehat{X})$ devient :

$$(42) \quad R_2(\widehat{C}, \widehat{X}) = \frac{1}{p^2} \left[\sum_{u=1}^p \sum_{v=1}^q 2 \frac{n_{uv}^2}{n_{u\bullet} n_{\bullet v}} + N \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}}{n_{u\bullet} n_{\bullet v}} - p - q \right]$$

Les deux expressions du critère doublement pondéré de Rand, données par (41) et (42) ne sont pas égales. Ceci est dû à la différence entre les termes pq et $N \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}}{n_{u\bullet} n_{\bullet v}}$.

D'où le **PARADOXE** de la double pondération.

Théorème 2 :

Bien que les deux expressions non pondérées soient égales sans pondération et égales avec une seule pondération, elles sont inégales lorsque l'on applique la double pondération.

3.3 Résolution du paradoxe par une écriture logique en notations relationnelles

Pour résoudre le paradoxe précédent, nous proposons une écriture logique en notations relationnelles applicable au critère doublement pondéré de Rand dont nous allons montrer qu'elle constitue un compromis entre les deux écritures contingentiellles données par (41) et (42).

Le critère relationnel doublement pondéré de Rand

Nous avons vu précédemment que les écritures relationnelles des critères de Rand et de Rand pondéré sont équivalentes à des constantes près, respectivement au critère de Condorcet et au critère de Condorcet pondéré.

Selon le même principe de décomposition, nous pouvons écrire le critère doublement pondéré de Rand en notations relationnelles sous la forme suivante :

$$(43) \quad R(\widehat{C}, \widehat{X}) = \frac{1}{p^2} \sum_{i=1}^N \sum_{i'=1}^N (\widehat{c}_{ii'} \widehat{x}_{ii'} + \bar{\widehat{c}}_{ii'} \bar{\widehat{x}}_{ii'}) = \frac{1}{p^2} C_{pp}(X)$$

où $C_{pp}(X)$ est le critère de Condorcet doublement pondéré que nous pouvons considérer, par analogie avec les critères de Condorcet et de Condorcet pondéré,

comme suit :

$$(44) \quad C_{pp}(X) = \sum_{i=1}^N \sum_{i'=1}^N (\hat{c}_{ii'} \hat{x}_{ii'} + \bar{\hat{c}}_{ii'} \bar{\hat{x}}_{ii'})$$

En notations relationnelles, que vaut :

$$\frac{1}{p^2} \sum_{i=1}^N \sum_{i'=1}^N \hat{c}_{ii'} \hat{x}_{ii'} + \frac{1}{p^2} \sum_{i=1}^N \sum_{i'=1}^N \bar{\hat{c}}_{ii'} \bar{\hat{x}}_{ii'}$$

Si l'on développe cette expression, en utilisant la formule (18) et les formules (31'), (31''), (32') et (32''), nous obtenons :

$$\begin{aligned} \frac{1}{p^2} \left[\sum_{i=1}^N \sum_{i'=1}^N \hat{c}_{ii'} \hat{x}_{ii'} + \frac{N}{4} \sum_{i=1}^N \hat{c}_{ii} \hat{x}_{ii} + \frac{N}{4} \sum_{i'=1}^N \hat{c}_{i'i'} \hat{x}_{i'i'} \right. \\ \left. + \frac{pq}{2} - p - q + \sum_{i=1}^N \sum_{i'=1}^N \bar{\hat{c}}_{ii'} \bar{\hat{x}}_{ii'} \right] \end{aligned}$$

Soit en notations contingentielles, selon les formules (30') et (34) :

$$(45) \quad \frac{1}{p^2} \left[2 \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}^2}{n_{u\bullet} n_{\bullet v}} + \frac{N}{2} \sum_u \sum_v \frac{n_{uv}}{n_{u\bullet} n_{\bullet v}} + \frac{pq}{2} - p - q \right]$$

Que nous pouvons écrire :

$$\begin{aligned} \frac{1}{2p^2} \left\{ \left[2 \sum_u \sum_v \frac{n_{uv}^2}{n_{u\bullet} n_{\bullet v}} + pq - p - q \right] \right. \\ \left. + \left[2 \sum_u \sum_v \frac{n_{uv}^2}{n_{u\bullet} n_{\bullet v}} - p - q + N \sum_u \sum_v \frac{n_{uv}}{n_{u\bullet} n_{\bullet v}} \right] \right\} \end{aligned}$$

D'où, nous obtenons :

$$\frac{1}{2} [R_1(\hat{C}, \hat{X}) + R_2(\hat{C}, \hat{X})]$$

Théorème 3

Le critère pondéré relationnel est la moyenne des deux expressions contingentielles du critère de Rand doublement pondéré

4. Formulation relationnelle à partir du critère doublement pondéré de Rand

4.1 Du critère du Chi-deux

D'après la formule (149) de [13], le critère du Phi-deux s'écrit :

$$\phi^2 = \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}^2}{n_{u\bullet} \cdot n_{\bullet v}} - 1$$

En utilisant la formule (30'),

$$\phi^2 = \sum_{i=1}^N \sum_{i'=1}^N \hat{c}_{ii'} \hat{x}_{ii'} - 1$$

Mais en vertu de la formule (45), nous avons :

$$\phi^2 = \frac{1}{2} \left[p^2 R(\hat{C}, \hat{X}) + p + q - \frac{1}{2} \left(pq + N \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}}{n_{u\bullet} \cdot n_{\bullet v}} \right) \right] - 1$$

Soit en notations relationnelles selon (44) et les formules (31''), (32'') et (34) :

$$\phi^2 = \frac{1}{2} \left[p + q - 2 - \sum_{i=1}^N \sum_{i'=1}^N (\hat{c}_{ii'} \bar{\bar{x}}_{ii'} + \bar{\bar{c}}_{ii'} \hat{x}_{ii'}) \right]$$

D'autre part, à partir de la formule (5) de [12], le critère du Chi-deux s'écrit :

$$\chi^2 = N\phi^2$$

D'où finalement, nous obtenons :

$$\chi^2 = \frac{N}{2} \left[p + q - 2 - \sum_{i=1}^N \sum_{i'=1}^N (\hat{c}_{ii'} \bar{\bar{x}}_{ii'} + \bar{\bar{c}}_{ii'} \hat{x}_{ii'}) \right]$$

Donc maximiser le χ_2 , qui implique une forte dépendance des variables C et X , revient à un nombre de classes fixé, à minimiser le dual du critère relationnel⁹.

⁹ où $C_{pp}(\bar{X}) = \sum_{i=1}^N \sum_{i'=1}^N (\hat{c}_{ii'} \bar{\bar{x}}_{ii'} + \bar{\bar{c}}_{ii'} \hat{x}_{ii'})$ est le dual du critère relationnel.

4.2 Distance entre deux variables en notations relationnelles

La distance entre les tableaux relationnels C et X pondérés en lignes ou en colonnes est donnée par¹⁰

$$D^2(C, X) = \sum_{i=1}^N \sum_{i'=1}^N \left(\frac{c_{ii'}}{c_{i\bullet}} - \frac{x_{ii'}}{x_{\bullet i'}} \right)^2 = p + q - 2(\phi^2 + 1)$$

En remplaçant ϕ^2 par sa valeur dans cette formule nous obtenons :

$$D^2(C, X) = \sum_{i=1}^N \sum_{i'=1}^N (\hat{c}_{ii'} \bar{x}_{ii'} + \bar{c}_{ii'} \hat{x}_{ii'})$$

5. Le dual des critères relationnels

5.1 Le critère de Condorcet

En se référant à [18], il se présente comme suit :

$$C(X) = \sum_{i=1}^N \sum_{i'=1}^N (c_{ii'} x_{ii'} + \bar{c}_{ii'} \bar{x}_{ii'})$$

D'autre part selon [5], le dual de ce critère est donné par :

$$C(\bar{X}) = \sum_{i=1}^N \sum_{i'=1}^N (c_{ii'} \bar{x}_{ii'} + \bar{c}_{ii'} x_{ii'})$$

Ce qui nous donne, en effectuant la somme de ces deux critères et du fait que $c_{ii'} + \bar{c}_{ii'} = 1$ et $x_{ii'} + \bar{x}_{ii'} = 1$:

$$C(X) + C(\bar{X}) = N^2$$

5.2 Le critère de Condorcet pondéré

Nous avons démontré dans [5] et [9] que le critère de Condorcet pondéré s'écrit :

$$C_P(X) = \sum_{i=1}^N \sum_{i'=1}^N (\hat{c}_{ii'} x_{ii'} + \bar{c}_{ii'} \bar{x}_{ii'})$$

¹⁰ Voir la formule (151) dans [13]

Par analogie avec le critère de Condorcet et son dual, le dual du critère de Condorcet pondéré peut s'écrire de la façon suivante :

$$C_P(\bar{X}) = \sum_{i=1}^N \sum_{i'=1}^N (\hat{c}_{ii'} \bar{x}_{ii'} + \bar{c}_{ii'} x_{ii'})$$

A partir de ces deux formules et des formules¹¹ (11), (18) et (31'), la somme du critère de Condorcet et de son dual est égale au produit du nombre d'individus et du nombre de modalités de la variable C , soit :

$$C_P(X) + C_P(\bar{X}) = Np$$

5.3 Le critère doublement pondéré de Condorcet

Nous avons vu (c.f. la formule 44) que :

$$C_{PP}(X) = \sum_{i=1}^N \sum_{i'=1}^N (\hat{c}_{ii'} \hat{x}_{ii'} + \bar{c}_{ii'} \bar{x}_{ii'})$$

et

$$C_{PP}(\bar{X}) = \sum_{i=1}^N \sum_{i'=1}^N (\hat{c}_{ii'} \bar{x}_{ii'} + \bar{c}_{ii'} \hat{x}_{ii'})$$

Il en résulte que :

$$1) C_{PP}(X) + C_{PP}(\bar{X}) = \frac{pq}{2} + \frac{N}{2} \sum_{u=1}^p \sum_{v=1}^q \frac{n_{uv}}{n_{u\bullet} n_{\bullet v}}$$

Ceci est dû aux formules (18) (où $m = 1$), (31'), (32') et (34).

$$2) C_{PP}(X) + C_{PP}(\bar{X}) = p^2 \text{ dans le cas optimal}$$

Notons que l'on a :

$$C_{PP}(\bar{X}) = D^2(C, X) = p + q - 2(\phi^2 + 1)$$

5.4 Conclusion

Nous avons vu que dans l'approche logique des données et du résultat, nous travaillons dans un espace relationnel à N^2 degrés de liberté (le tableau d'une relation), tandis que dans l'approche probabiliste des données et logique du résultat, l'espace a Np degrés de liberté (le tableau disjonctif complet). Enfin dans l'approche probabiliste des données et du résultat, l'espace considéré a p^2 degrés de liberté (le tableau contingenciel croisant C et X dans le cas optimal).

D'où, nous déduisons **le troisième résultat** :

¹¹ Nous sommes en présence d'une seule variable; en tenant compte de ces formules avec $m = 1$.

Le critère de Condorcet pondéré joue un rôle intermédiaire entre le critère de Condorcet et le critère doublement pondéré de Condorcet. D'autre part, dans le cas dual, le dual du critère de Condorcet pondéré joue un rôle intermédiaire entre le dual du critère de Condorcet et le critère du Chi-deux

6. Comparaison relationnelle-contingentielle des critères et Conclusion

Dressons le tableau récapitulatif des écritures relationnelles et contingentielles des critères et des écritures logiques et probabilistes des données et des résultats.

RELATIONNEL		CONTINGENTIEL	
Logique - Logique		1 ^{er} CAS	$N^2 R(C, X) =$
$C(C, X) = \sum_{i=1}^N \sum_{i'=1}^N C_{ii'} X_{ii'} + \bar{C}_{ii'} \bar{X}_{ii'}$	\equiv		$\left\{ \begin{array}{l} \sum_u \sum_v n_{uv}^2 + (n_{u.} - n_{uv})(n_{.v} - n_{uv}) \\ \sum_u \sum_v n_{uv}^2 + n_{uv}(N - n_{u.} - n_{.v} + n_{uv}) \end{array} \right.$
Logique - Probabi.		2 ^{ème} CAS	$Np R(\hat{C}, X) =$
$C(\hat{C}, X) = \sum_{i=1}^N \sum_{i'=1}^N \hat{C}_{ii'} X_{ii'} + \bar{\hat{C}}_{ii'} \bar{X}_{ii'}$	\equiv		$\left\{ \begin{array}{l} \sum_u \sum_v \frac{n_{uv}^2}{n_{u.}} + \frac{(n_{u.} - n_{uv})(n_{.v} - n_{uv})}{n_{u.}} \\ \sum_u \sum_v \frac{n_{uv}^2}{n_{u.}} + \frac{n_{uv}(N - n_{u.} - n_{.v} + n_{uv})}{n_{u.}} \end{array} \right.$
Probabi. - Probabi.		3 ^{ème} CAS	
$C(\hat{C}, \hat{X}) = \sum_{i=1}^N \sum_{i'=1}^N \hat{C}_{ii'} \hat{X}_{ii'} + \bar{\hat{C}}_{ii'} \bar{\hat{X}}_{ii'}$	\equiv		$p^2 R_1(\hat{C}, \hat{X}) = \sum_u \sum_v \frac{n_{uv}^2}{n_{u.} n_{.v}} + \frac{(n_{u.} - n_{uv})(n_{.v} - n_{uv})}{n_{u.} n_{.v}}$ $\frac{1}{2} (p^2 R_1(\hat{C}, \hat{X}) + p^2 R_2(\hat{C}, \hat{X}))$ $p^2 R_2(\hat{C}, \hat{X}) = \sum_u \sum_v \frac{n_{uv}^2}{n_{u.} n_{.v}} + \frac{n_{uv}(N - n_{u.} - n_{.v} + n_{uv})}{n_{u.} n_{.v}}$

L'examen de ce tableau nous amène à la conclusion suivante :

Le paradigme relationnel ne génère aucune ambiguïté alors que celui de type contingentiel implique dans le cas 3, une double écriture du critère contingentiel.

7. Références

- [1] BENZÉCRI J-P. et al. *Analyse des données, Tome 1 : L'analyse des correspondances*, Dunod, (1973).
- [2] BÉDÉCARRAX C. *Classification automatique en Analyse Relationnelle : la Quadri-décomposition et ses applications*, Etude du Centre Scientifique IBM de Paris No F-0132, (1988).

- [3] BÉDÉCARRAX C. *Classification automatique en Analyse Relationnelle : la Quadri-décomposition et Applications*, Thèse de l'Université Paris VI, (1989).
- [4] BÉDÉCARRAX C. et WARNESSON I. *Relational Analysis and Dictionaries*, Proceedings of the Fourth International Symposium on Applied Stochastic Processus and Data Analysis, INRIA, Nancy,(1988).
- [5] EL AYOUBI N. *Liaison Analyse Relationnelle-Analyse Factorielle : Extensions*, Thèse de l'Université Paris VI, (1990).
- [6] EL AYOUBI N. *Analyse Relationnelle et Analyse Non Symétrique des Correspondances sur tableau disjonctif*, Colloque International sur les méthodes de blocs sériation et applications, Strasbourg, France, (Avril 1990).
- [7] KENDALL M-G et STUART A. *The Advanced Theory of Statistics*, Vol 2, Griffin, Londres, (1961).
- [8] JORDAN C. *Les coefficients d'Intensité Relative de Körösy*, Revue de la Société Hongroise de Statistique, Vol 5, Budapest, (1927).
- [9] MARCOTORCHINO F. *Liaison Analyse Factorielle-Analyse Relationnelle (I) : "Dualité Burt-Condorcet"*, Etude du Centre Scientifique IBM France, No F142, (1989).
- [10] MARCOTORCHINO F. *Les techniques de sériation : Etat de l'art*, Proceedings du colloque International sur les méthodes de blocs sériation et applications, Strasbourg, France, (Avril 1990).
- [11] MARCOTORCHINO F. *Inerties Relationnelles (II)* , Etude du Centre Européen de mathématiques Appliquées – IBM France, No MA03 (à paraître fin 1990).
- [12] MARCOTORCHINO F. *Utilisation des Comparaisons par Paires en Statistique des Contingences (Partie I)*, Etude du Centre Scientifique IBM France, No F069, (1984).
- [13] MARCOTORCHINO F. *Utilisation des Comparaisons par Paires en Statistique des Contingences (Partie II)*, Etude du Centre Scientifique IBM France, No F071, (1984).
- [14] MARCOTORCHINO F. *Maximal Association Theory as a Tool for Classification*, in Classification as a Tool for Research, (Gaul W. and Schader M. eds), North Holland, Amsterdam, (1986).
- [15] MARCOTORCHINO F. et MICHAUD P. *Optimisation en Analyse Ordinale des Données*, Masson, Paris, (1979).
- [16] MARCOTORCHINO F. et MICHAUD P. *Agrégation des Similarités en Classification Automatique*, Revue de Statistique Appliquée, Vol 30, No2, (1981).
- [17] MICHAUD P. et MARCOTORCHINO F. *Modèles d'optimisation en Analyse des Données Relationnelles*, Math. et Sciences Humaines, No 67, Gauthier Villars, Paris, (1979).
- [18] MICHAUD P. *Agrégation à la Majorité I : Hommage à Condorcet*, Etude du Centre Scientifique IBM France, No F051, (1982).
- [19] MICHAUD P. *Agrégation à la Majorité II : Analyse du Résultat d'un vote*, Etude du Centre Scientifique IBM France, No F052, (1985).

- [20] MESSATFA H. *Unification Relationnelle des critères et Structures de Contingence*, Thèse de l'Université Paris VI, (1990).
- [21] RAND W-H. *Objective Criteria for the Evaluation of Clustering Methods*, Journal of the American Statistical Association, Vol 66, (1971).