



HAL
open science

Tableaux de Burt et tableaux de Condorcet associés à un tableau ku dont l'un des deux ensembles I ou J est muni d'une partition

P. Cazes, J. Moreaux

► **To cite this version:**

P. Cazes, J. Moreaux. Tableaux de Burt et tableaux de Condorcet associés à un tableau ku dont l'un des deux ensembles I ou J est muni d'une partition. *Annales de l'ISUP*, 1995, XXXIX (1), pp.45-66. hal-03659956

HAL Id: hal-03659956

<https://hal.science/hal-03659956>

Submitted on 5 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tableaux de Burt et tableaux de Condorcet associés à un tableau k_{IJ} dont l'un des deux ensembles I ou J est muni d'une partition

P. CAZES

CEREMADE, Université Paris IX Dauphine,
 Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16

J. MOREAU

Centre Vaudois de Recherches Pédagogiques.
 56, rue Marterey, 1005 Lausanne, Suisse

Key words: analyse relationnelle, analyse des correspondances, analyse interclasses, analyse intraclasses, classification

1 Introduction

On étudie ici les propriétés des tableaux de Burt et de Condorcet associés à un tableau de contingence k_{IJ} muni d'une structure définie par une partition soit de l'ensemble I de ses lignes, soit de l'ensemble J de ses colonnes. Après avoir rappelé (§ 2) la définition et les propriétés de ces tableaux, on étudie (§ 3) les propriétés des tableaux de Burt et de Condorcet associés au tableau dont l'Analyse Factorielle des Correspondances (A.F.C.) réalise l'analyse intraclasses du tableau k_{IJ} , puis on examine le cas particulier fort important dans la pratique où k_{IJ} est un tableau disjonctif complet. Au § 4, on montre, k_{IJ} étant toujours un tableau disjonctif complet, que l'A.F.C. du tableau de Condorcet peut être considérée comme une analyse interclasses, ce résultat subsistant si on remplace k_{IJ} par le tableau dont l'A.F.C. réalise l'analyse intraclasses de k_{IJ} , quand on a une partition sur J. Enfin au § 5, on généralise quelques critères relationnels en classification, critères introduits par Marcotorchino et Michaud, et l'on indique comment modifier ces critères pour s'affranchir de la structure a priori dont est muni le tableau k_{IJ} .

2 Notations. Rappels sur le tableau de Burt et le tableau de Condorcet

On considère un tableau de contingence $k_{IJ} = \{k(i,j) \mid i \in I, j \in J\}$ défini sur le produit de deux ensembles finis I et J , et on désigne par $k_I = \{k(i,.) \mid i \in I\}$ et $k_J = \{k(.,j) \mid j \in J\}$ ses marges:

$$k(i,.) = \sum \{k(i,j) \mid j \in J\}; k(.,j) = \sum \{k(i,j) \mid i \in I\} \quad (0)$$

Dans certaines parties du texte, on supposera que

- soit I est muni d'une partition: $I = \cup \{I_p \mid p \in P\}$

- soit J est muni d'une partition: $J = \cup \{J_r \mid r \in R\}$

- soit I et J sont tous les deux munis d'une partition.

On posera alors:

$$\left. \begin{aligned} k(p,j) &= \sum \{k(i,j) \mid i \in I_p\} \\ k(p,.) &= \sum \{k(i,.) \mid i \in I_p\} \\ k(i,r) &= \sum \{k(i,j) \mid j \in J_r\} \\ k(.,r) &= \sum \{k(.,j) \mid j \in J_r\} \\ k(p,r) &= \sum \{k(i,j) \mid i \in I_p, j \in J_r\} \end{aligned} \right\} \quad (1)$$

On notera aussi $k(i)$, $k(j)$, $k(p)$, $k(r)$ pour $k(i,.)$, $k(.,j)$, $k(p,.)$ et $k(.,r)$ respectivement quand il n'y aura pas de confusion à craindre, et l'on désignera par k le total du tableau k_{IJ} .

Rappelons que le tableau de Burt sur $J \times J$ associé à k_{IJ} est le tableau B_{JJ} de terme général $B(j,j')$ défini par:

$$B(j,j') = \sum \{k(i,j)k(i,j') / k(i,.) \mid i \in I\} \quad (2)$$

tandis que le tableau de Burt sur $I \times I$ associé au tableau k_{IJ} est le tableau C_{II} de terme général $C(i,i')$ donné par:

$$C(i,i') = \sum \{k(i,j)k(i',j) / k(.,j) \mid j \in J\} \quad (3)$$

Ce tableau, qui a été appelé tableau de Condorcet pondéré par Marcotorchino et Michaud dans le cas où k_{IJ} , est un tableau disjonctif complet et que nous appellerons ici tableau de Condorcet, réservant le nom de tableau de Burt au tableau B_{JJ} .

Aux tableaux B_{JJ} et C_{II} , on peut associer les tableaux complémentaires \bar{B}_{JJ} et \bar{C}_{II} dont les termes généraux $\bar{B}(j,j')$ et $\bar{C}(i,i')$ sont respectivement définis par:

$$\left. \begin{aligned} \bar{C}(i,i') &= \frac{C(i,i) + C(i',i')}{2} - C(i,i') \\ \bar{B}(j,j') &= \frac{B(j,j) + B(j',j')}{2} - B(j,j') \end{aligned} \right\} \quad (4)$$

Rappelons quelques propriétés des tableaux B_{JJ} et C_{II} .

Les deux marges de B_{JJ} (resp C_{II}) sont égales à k_J (resp k_I).

Si (φ^I, φ^J) est un couple de facteurs associés issu de l'Analyse Factorielle des Correspondances (A.F.C.) du tableau k_{IJ} et relatif à la valeur propre λ , alors (φ^J, φ^I) (resp (φ^I, φ^J)) est un couple de facteurs associés issu de l'A.F.C. de B_{JJ} (resp C_{II}) et relatif à la valeur propre λ^2 .

Si $I_T(k_{IJ})$ désigne l'inertie totale dans l'A.F.C. du tableau k_{IJ} , on a:

$$I_T(k_{IJ}) = \sum_j \frac{B(j,j)}{k(j)} - 1 = \sum_i \frac{C(i,i)}{k(i)} - 1 \quad (5)$$

Par ailleurs, on déduit immédiatement des relations (2), (3), (4) définissant les tableaux B_{JJ} , C_{II} , \bar{B}_{JJ} et \bar{C}_{II} que:

$$\left. \begin{aligned} \bar{B}(j,j') &= \frac{1}{2} \sum_i \frac{(k(i,j) - k(i,j'))^2}{k(i)} \\ \bar{C}(i,i') &= \frac{1}{2} \sum_j \frac{(k(i,j) - k(i',j))^2}{k(j)} \end{aligned} \right\} \quad (6)$$

Dans le cas où $k(i)$ est constant, et donc égal à $k / \text{Card}(I)$, on a, si $d^2(i,i')$ désigne le carré de la distance du chi-2 entre les profils des lignes i et i' du tableau k_{IJ} :

$$d^2(i,i') = [2 (\text{Card}(I)^2 / k)] \bar{C}(i,i') \quad (7)$$

L'inertie totale $I_T(k_{IJ})$ dans l'A.F.C. du tableau k_{IJ} , qu'on notera aussi I_T , peut alors s'écrire puisque la masse p_i de chaque ligne i est égale à $k(i) / k = 1 / \text{Card}(I)$:

$$I_T = (1/2) \sum_{i, i'} p_i p_{i'} d^2(i, i') = (1/k) \sum_{i, i'} \bar{C}(i, i') = \bar{C}(.,.) / k = \bar{C}(.,.) / C(.,.) \quad (8)$$

$\bar{C}(.,.)$ et $C(.,.)$ désignant les totaux respectifs des tableaux \bar{C}_{II} et C_{II} .

De même, si $k(j)$ est constant, on a avec des notations symétriques des précédentes:

$$d^2(j, j') = [2 (\text{Card}(J)^2 / k)] \bar{B}(j, j') \quad (7 \text{ bis})$$

$$I_T = \bar{B}(.,.) / k = \bar{B}(.,.) / B(.,.) \quad (9)$$

Les relations (6) et (8) ont été données par Marcotorchino (1989) dans le cas particulier où k_{IJ} est un tableau disjonctif complet. Toujours dans ce cas, Marcotorchino montre que la relation (9) reste valable (bien que, en général, $k(.,j)$ ne soit pas constant).

3 Cas d'un tableau structuré par une partition sur l'un des deux ensembles I ou J

3.1 Cas d'une partition sur J

On suppose donc qu'on a la partition de J définie par $J = \cup \{ J_r \mid r \in R \}$. On sait que l'analyse intraclasse (cf. par exemple Benzécri (1983), Cazes & Moreau (1991)) revient à effectuer l'A.F.C. du tableau R_{IJ} de terme général $R(i, j)$ donné par:

$$\forall j \in J_r : R(i, j) = k(i, j) - \frac{k(j)k(i, r)}{k(r)} + \frac{k(i)k(j)}{k} \quad (10)$$

cette analyse permettant de s'affranchir de la structure définie par la partition de J.

On désignera par B_{RJ} et C_{RII} les tableaux de Burt et de Condorcet associés au tableau R_{IJ} , i.e. tableaux obtenus à partir des formules (2) et (3) où le tableau k_{IJ} est remplacé par R_{IJ} .

Compte tenu de ce que les tableaux k_{IJ} et R_{IJ} ont même marge, on obtient après un calcul sans difficultés, les expressions suivantes pour les termes généraux $B_R(j, j')$ et $C_R(i, i')$ des tableaux B_{RJ} et C_{RII} (cf. Cazes & Moreau (1991)):

$$\forall j \in J_r \quad \forall j' \in J_{r'} :$$

$$B_R(j, j') = B(j, j') - \frac{k(j')}{k(r')} B(j, r') - \frac{k(j)}{k(r)} B(j', r) + \frac{k(j)k(j')}{k(r)k(r')} B(r, r') + \frac{k(j)k(j')}{k} \quad (11)$$

avec

$$\left. \begin{aligned} B(j, r') &= \sum_{j' \in J_r} B(j, j') = \sum_i \frac{k(i, j)k(i, r')}{k(i)} = B(r', j) \\ B(r, r') &= \sum_{\substack{j \in J_r \\ j' \in J_r}} B(j, j') = \sum_i \frac{k(i, r)k(i, r')}{k(i)} \end{aligned} \right\} \quad (12)$$

$$\forall i, i' \in I : C_R(i, i') = C(i, i') - \sum_r \frac{k(i, r)k(i', r)}{k(r)} + \frac{k(i)k(i')}{k} \quad (13)$$

Remarque

Compte tenu de ce que les deux marges de B_{JJ} sont égales à k_J , on déduit de (11) que le tableau B_{RJ} est le tableau dont l'A.F.C. réalise l'analyse intraclasses double (ou analyse interne) (cf. Cazes, Chessel & Doledec (1988), Cazes & Moreau (1991)) du tableau B_{JJ} , chacun des deux ensembles J étant muni de la partition définie par les classes J_r ($r \in R$).

Tenant compte de (4) et de (13), on obtient comme terme général $\bar{C}_R(i, i')$ du tableau complémentaire \bar{C}_{RII} l'expression:

$$\bar{C}_R(i, i') = \bar{C}(i, i') - \frac{1}{2} \sum_r \frac{(k(i, r) - k(i', r))^2}{k(r)} + \frac{(k(i) - k(i'))^2}{2k}, \quad (14)$$

expression qui se simplifie si $k(i)$ est constant, le dernier terme de cette expression étant alors nul.

L'inertie interclasses $I_B(k_{IJ})$ du tableau k_{IJ} relativement à la partition R sur J est l'inertie totale $I_T(k_{IR})$ du tableau k_{IR} de terme général $k(i, r)$ ($i \in I, r \in R$), soit:

$$I_B(k_{IJ}) = I_T(k_{IR}) = \sum_i \sum_r \frac{k^2(i, r)}{k(i)k(r)} - 1 \quad (15)$$

$$= \sum_i \sum_r \sum_{j \in J_r} \sum_{j' \in J_r} \frac{k(i, j)k(i, j')}{k(i)k(r)} - 1 \quad (16)$$

$$= \sum_r \sum_{j \in J_r} \sum_{j' \in J_r} \frac{B(j, j')}{k(r)} - 1 \quad (17)$$

$$= \sum_r \frac{B(r, r)}{k(r)} - 1 \quad (18)$$

d'où pour l'inertie intraclasses $I_W(k_{IJ}) = I_T(R_{IJ}) = I_T(k_{IJ}) - I_B(k_{IJ})$, et compte tenu de (5):

$$I_W(k_{IJ}) = \sum_j \frac{B(j, j)}{k(j)} - \sum_r \frac{B(r, r)}{k(r)} \quad (19)$$

Nous allons donner une autre expression de ces inerties interclasses et intraclasses à partir de la relation d'équivalence Y associée à la partition R et définie par:

$$y_{jj'} = 1 \quad \text{si } j \text{ et } j' \text{ appartiennent à la même classe}$$

$$y_{jj'} = 0 \quad \text{sinon} \quad (20)$$

On déduit alors de (17) que:

$$I_B(k_{IJ}) = \sum_{j \in J} \sum_{j' \in J} \frac{B(j, j')}{k(r(j))} y_{jj'} - 1 \quad (21)$$

$r(j)$ désignant la classe de j dans la partition R .

$$\text{comme: } k(r(j)) = \sum \{ k(j'') \mid j'' \in J \} \quad (22)$$

(21) s'écrit encore:

$$I_B(k_{IJ}) = \sum_{j, j'} \frac{B(j, j')}{k(j)} \frac{y_{jj'}}{\sum_{j''} k(j'') y_{jj''}} - 1 \quad (23)$$

L'inertie intraclasses s'écrit alors, compte tenu de (5):

$$I_W(k_{IJ}) = \sum_j \frac{B(j, j)}{k(j)} - \sum_{j, j'} \frac{B(j, j')}{k(j)} \frac{y_{jj'}}{\sum_{j''} k(j'') y_{jj''}} \quad (24)$$

Dans le cas où $k(j)$ est constant et donc égal à $L = k / \text{Card}(J)$, les relations précédentes se simplifient, et on a :

$$I_B(k_{IJ}) = \frac{1}{L} \sum_r \frac{B(r,r)}{\text{Card}(J_r)} - 1 \quad (25)$$

$$= \frac{1}{L} \sum_{j,j'} B(j,j') \frac{y_{jj'}}{y_j} - 1 \quad (26)$$

$$I_W(k_{IJ}) = \frac{1}{L} \left(\sum_j B(j,j) - \sum_r \frac{B(r,r)}{\text{Card}(J_r)} \right) \quad (27)$$

$$I_W(k_{IJ}) = \frac{1}{L} \sum_{j,j'} \bar{B}(j,j') \frac{y_{jj'}}{y_j} \quad (28)$$

$$\text{avec: } y_j = \sum \{y_{jj'} \mid j' \in J\} \quad (29)$$

Notons que la relation (28) peut s'obtenir à partir de (4) et de (24) en tenant compte du fait que :

$$\sum_j B(j,j) = \frac{1}{2} \sum_{j,j'} (B(j,j) + B(j',j')) y_{jj'} / y_j \quad (30)$$

cette dernière relation résultant du fait que si $y_{jj'}$ est non nul (j et j' appartiennent à la même classe J_r) $y_j = y_{j'} (= \text{Card}(J_r))$.

La relation (23) a déjà été donnée par Marcotorchino (1989) dans le cas particulier où k_{IJ} est un tableau disjonctif complet. Toujours dans ce cas (étudié au § 3.3) Marcotorchino donne d'autres expressions relationnelles (i.e. basées sur les $y_{jj'}$) pour l'inertie intraclasses, expressions que l'on peut déduire aisément de (24) en tenant compte de ce que $k^2(i,j) = k(i,j)$ ($= 0$ ou 1) et de ce que $k(i,.)$ est constant.

3.2 Cas d'une partition sur I

Tous les résultats précédents se transposent immédiatement quand on a une partition $I = \cup \{I_p \mid p \in P\}$ sur I. Il suffit d'invertir les rôles de I et J, de remplacer le tableau R_{IJ} par le tableau S_{IJ} de terme général $S(i,j)$ défini par :

$$\forall i \in I_p : S(i,j) = k(i,j) - \frac{k(i)k(p,j)}{k(p)} + \frac{k(i)k(j)}{k} \quad (31)$$

de remplacer R par P, $B_{R_{IJ}}$ par $C_{S_{IJ}}$ (tableau de Condorcet associé à S_{IJ}), $C_{R_{II}}$ par $B_{S_{JJ}}$ (tableau de Burt associé à S_{IJ}) et la relation d'équivalence Y par la relation

d'équivalence X de terme général $x_{ij} = 1$ si i et i' appartiennent à la même classe, et 0 sinon.

L'inertie interclasses $I_B(k_{IJ})$ s'écrit alors, compte tenu de (23):

$$I_B(k_{IJ}) = \sum_{i,i'} C(i,i') \frac{x_{ij}}{\sum_{i''} k(i'')x_{ij''}} - 1 \quad (31 \text{ bis})$$

formule qui nous servira au § 5.

Nous allons maintenant donner quelques formules dans le cas où $k(i)$ est constant et donc égal à $M = k / \text{Card}(I)$. Dans ce cas, on a en désignant par n_p le cardinal de I_p , et avec des notations évidentes:

$$\forall i \in I_p : S(i,j) = k(i,j) - \frac{k(p,j)}{n_p} + \frac{k(j)}{\text{Card}(I)} \quad (32)$$

$$\forall i \in I_p, \forall i' \in I_{p'} : C_S(i,i') = C(i,i') - \frac{C(i,p')}{n_{p'}} - \frac{C(i',p)}{n_p} + \frac{C(p,p')}{n_p n_{p'}} + \frac{M}{\text{Card}(I)} \quad (33)$$

$$I_B(k_{IJ}) = \frac{1}{M} \sum_p \frac{C(p,p)}{n_p} - 1 \quad (34)$$

$$= \frac{1}{M} \sum_{i,i'} C(i,i') \frac{x_{ij}}{x_i} - 1 \quad (34 \text{ bis})$$

$$I_W(k_{IJ}) = I_T(S_{IJ}) = \frac{1}{M} \left(\sum_i C(i,i) - \sum_p \frac{C(p,p)}{n_p} \right) \quad (35)$$

$$= \frac{1}{M} \sum_{i,i'} \bar{C}(i,i') \frac{x_{ij}}{x_i} \quad (35 \text{ bis})$$

On déduit de (8) et de (35), et compte tenu de ce que $k = M \text{Card}(I)$, que le total $\bar{C}_S(.,.)$ des éléments du tableau \bar{C}_{SII} s'écrit:

$$\bar{C}_S(.,.) = \text{Card}(I) \left(\sum_i C(i,i) - \sum_p \frac{C(p,p)}{n_p} \right) \quad (36)$$

Notons que les formules (34 bis) et (35 bis) ont été données par Marcotorchino (1989) dans le cas particulier où k_{IJ} est un tableau disjonctif complet.

3.3 Cas où k_{IJ} est un tableau disjonctif complet

On suppose ici que I est un ensemble d'individus et que $J = \cup \{ J_q \mid q \in Q \}$, J_q étant l'ensemble des modalités de réponses à la question q .

Si $q(i)$ désigne la modalité de J_q adoptée par l'individu i , on a:

$$\forall j \in J_q \subset J : k(i,j) = \delta_j^{q(i)} = 1 \text{ si } q(i) = j \text{ et } 0 \text{ sinon} \quad (37)$$

On a alors $k(i) = \text{Card}(Q)$, $k = \text{Card}(I)\text{Card}(Q)$, $k(j)$ étant le nombre d'individus ayant adopté la modalité j .

Nous supposons de plus que chaque ensemble J_q est muni d'une partition dont les classes sont indicées par R_q ($J_q = \cup \{ J_r \mid r \in R_q \}$). L'ensemble J est alors muni de la partition R définie à partir de l'union des R_q . L'A.F.C. du tableau R_{IJ} donné par (10), i.e. l'analyse intraclasse de k_{IJ} relativement à la partition R de J correspond à l'analyse intraclasse (ou interne) multiple (cf. Cazes & Moreau (1991)).

Gardant les notations définies par (1) et désignant par $r(j)$ la classe à laquelle appartient j , on a:

$$k(i,r) = \delta_{r(q(i))}^r = 1 \text{ si } q(i) \in J_r \text{ et } 0 \text{ sinon} \quad (38)$$

Le terme général $R(i,j)$ du tableau R_{IJ} s'écrit alors, compte tenu de (10):

$$\forall j \in J_r \subset J_q : R(i,j) = \delta_j^{q(i)} - \frac{\delta_{r(q(i))}^r k(j)}{k(r)} + \frac{k(j)}{\text{Card}(I)} \quad (39)$$

tandis que les tableaux C_{RII} et \bar{C}_{RII} ont respectivement pour terme général:

$$C_R(i,i') = C(i,i') - \frac{\delta_{r(q(i))}^r}{k(r(q(i)))} + \frac{\text{Card}(Q)}{\text{Card}(I)} \quad (40)$$

$$\bar{C}_R(i,i') = \bar{C}(i,i') - \frac{1}{2} \sum_r \frac{(k(i,r) - k(i',r))^2}{k(r)} \quad (41)$$

L'inertie totale $I_T(k_{IJ})$ qui est égale à $(\text{Card}(J) - \text{Card}(Q)) / \text{Card}(Q)$ s'écrit aussi, compte tenu de (8):

$$I_T(k_{IJ}) = \frac{\text{Card}(J) - \text{Card}(Q)}{\text{Card}(Q)} = \frac{\bar{C}(\dots)}{\text{Card}(I)\text{Card}(Q)} \quad (42)$$

$$\text{Ce qui montre que: } \bar{C}(\dots) = \text{Card}(I) (\text{Card}(J) - \text{Card}(Q)) \quad (43)$$

Par ailleurs l'inertie interclasses s'écrit d'après (15), et puisque $k^2(i,r) = k(i,r)$ et $k(i) = \text{Card}(Q)$:

$$I_B(k_{IJ}) = \frac{1}{\text{Card}(Q)} \sum_i \sum_r \frac{k(i,r)}{k(r)} - 1 = \frac{\text{Card}(R)}{\text{Card}(Q)} - 1 = \frac{\text{Card}(R) - \text{Card}(Q)}{\text{Card}(Q)} \quad (44)$$

résultat que l'on aurait pu déduire sans calcul, puisque $I_B(k_{IJ})$ est l'inertie totale $I_T(k_{IR})$ issue de l'A.F.C. du tableau k_{IR} obtenu en cumulant toutes les colonnes d'un même ensemble J_r ($r \in R$), tableau qui est un tableau disjonctif complet ayant $\text{Card}(R)$ colonnes.

L'inertie intraclasses $I_W(k_{IJ}) = I_T(k_{IJ}) - I_B(k_{IJ})$ s'écrit alors:

$$I_W(k_{IJ}) = \frac{\text{Card}(J) - \text{Card}(R)}{\text{Card}(Q)} \quad (45)$$

$$\text{Compte tenu de ce que: } I_W(k_{IJ}) = I_T(R_{IJ}) = \frac{\bar{C}_R(\dots)}{\text{Card}(I)\text{Card}(Q)} \quad (46)$$

puisque R_{IJ} a mêmes marges que k_{IJ} et en particulier une marge constante sur I, on a: $\bar{C}_R(\dots) = \text{Card}(I) (\text{Card}(J) - \text{Card}(R))$.

Supposons maintenant que I est muni d'une partition $I = \bigcup \{ I_p \mid p \in P \}$ et désignons toujours par n_p le cardinal de I_p . Tous les résultats du § 3.2 s'appliquent avec $M = k(i) = \text{Card}(Q)$.

Compte tenu de ce que:

$$\sum_i C(i,i) = \sum_i \sum_j \frac{k^2(i,j)}{k(j)} = \sum_i \sum_j \frac{k(i,j)}{k(j)} = \text{Card}(J) \quad (47)$$

l'expression (35) de l'inertie intraclasses se simplifie et on a:

$$I_W(k_{IJ}) = I_T(S_{IJ}) = \left(\text{Card}(J) - \frac{\sum_p C(p,p)}{n_p} \right) / \text{Card}(Q) \quad (48)$$

et donc:

$$\bar{C}_S(\dots) = \text{Card}(I) \left(\text{Card}(J) - \frac{\sum_p C(p,p)}{n_p} \right) \quad (49)$$

Notons que l'on aurait pu obtenir directement l'inertie intraclasse à partir de l'expression (34) de l'inertie interclasses et du fait que

$$I_T(k_{IJ}) = \frac{\text{Card}(J) - \text{Card}(Q)}{\text{Card}(Q)}$$

4 Interprétation de l'A.F.C. du tableau de Condorcet comme une analyse interclasses quand k_{IJ} est un tableau disjonctif complet

On suppose toujours que k_{IJ} est un tableau disjonctif complet. Le tableau k_{IJ} est donc partitionné suivant les blocs $\{k_{IJq} \mid q \in Q\}$ associés à chaque question q de Q .

Nous allons montrer que l'A.F.C. du tableau de Condorcet peut s'interpréter comme l'analyse interclasses d'un tableau $D_{II'}$ (où $I' = I \times Q$) dont l'ensemble I' est convenablement partitionné, tableau dont l'A.F.C. est équivalente à celle de k_{IJ} . L'inertie $I_T(C_{II})$ du tableau de Condorcet (inertie encore égale à $I_T(B_{JJ}) = \sum \{\lambda_\alpha^2 \mid \alpha \in A\}$, $\{\lambda_\alpha \mid \alpha \in A\}$ désignant l'ensemble des valeurs propres non triviales issues de l'A.F.C. de k_{IJ}) peut donc être considérée comme une inertie interclasses.

4.1 Définition et propriétés du tableau $D_{II'}$

Soit C_{II}^q le tableau de Condorcet associé au tableau k_{IJq} . Le tableau $D_{II'}$ est obtenu par superposition des C_{II}^q ($q \in Q$). On a donc:

$$\left. \begin{aligned} \forall i, i' \in I : C^q(i, i') &= \sum_{j \in J_q} \frac{k(i, j)k(i', j)}{k(j)} = \frac{1}{k(q(i))} \delta_{q(i)}^{q(i')} \\ &= \frac{1}{k(j)} \text{ si } q(i) = q(i') = j, \quad 0 \text{ sinon} \end{aligned} \right\} \quad (50)$$

$$\forall i \in I, \forall (i', q) \in I' : D(i, (i', q)) = C^q(i, i') = \frac{1}{k(q(i))} \delta_{q(i)}^{q(i')} \quad (51)$$

$q(i)$ désignant, rappelons-le, la modalité de J_q adoptée par i .

Nous allons considérer deux partitions de I' . La première nous permettra de voir, d'après le principe d'équivalence distributionnelle que l'A.F.C. de $D_{II'}$ est équivalente à l'A.F.C. du tableau k_{IJ} , tandis que la seconde nous permettra de voir

que l'analyse interclasses associée de $D_{II'}$ n'est rien d'autre que l'A.F.C. du tableau de Condorcet $C_{II'}$.

4.1.1 Etude de la première partition de I'

$\forall j \in J_q$, posons: $I_j^q = \{(i,q) \in I' \mid k(i,j) = 1\}$; I_j^q correspond à l'ensemble des individus ayant adopté la modalité j de J_q , et l'on a la partition $I_q = \cup \{I_j^q \mid j \in J_q\}$ de I (où l'on a noté I_q au lieu de I pour traduire le fait que la partition de I est associée à la question q).

$I' = \cup \{I_q \mid q \in Q\} = \cup \{I_j^q \mid j \in J_q, q \in Q\}$ est alors partitionné suivant les I_j^q

Or les colonnes (i,q) et (i',q) du tableau $D_{II'}$, appartenant à I_j^q sont identiques. En effet, on a puisque si i' appartient à I_j^q , $q(i') = j$, et d'après (51):

$$\forall (i',q) \in I_j^q : D(i,(i',q)) = \frac{\delta_{q(i)}^j}{k(j)} \quad (52)$$

valeur indépendante de i' (pourvu que $(i',q) \in I_j^q$).

Toutes les colonnes (i',q) de I_j^q étant identiques, d'après le principe d'équivalence distributionnelle, on ne change pas l'A.F.C. de $D_{II'}$ en remplaçant ces colonnes par leur somme.

L'analyse des correspondances du tableau $D_{II'}$ est donc équivalente à l'A.F.C. du tableau D_{IJ} dont le terme général est donné par:

$$\begin{aligned} \forall j \in J_q : D(i,j) &= \sum \left\{ D(i,(i',q)) \mid (i',q) \in I_j^q \right\} \\ &= \sum \left\{ \frac{\delta_{q(i)}^j}{k(j)} \mid (i',q) \in I_j^q \right\} \\ &= \delta_{q(i)}^j = k(i,j) \end{aligned} \quad (53)$$

puisque $k(j)$ représente le nombre d'individus ayant adopté la modalité j de J_q , i.e. l'effectif de I_j^q .

L'A.F.C. de $D_{II'}$ est donc équivalente à celle de k_{IJ} . En particulier les inerties totales associées à l'A.F.C. de ces deux tableaux sont égales.

4.1.2 Etude de la seconde partition de I'

Cette partition est définie par: $I' = \cup \{ \{i\} \times Q \mid i \in I \} = \cup \{ I_i \mid i \in I \}$

avec $I_i = \{i\} \times Q$ (54)

Ainsi, chaque bloc D_{II} de D_{II}' est associé aux modalités de réponses de l'individu i à l'ensemble Q des $\text{Card}(Q)$ questions. L'analyse interclasses de D_{II}' , associée à la partition précédente revient à faire l'A.F.C. du tableau D_{II} de terme général $D(i,i')$ défini par:

$$\begin{aligned} D(i,i') &= \sum \{ D(i,(i',q)) \mid (i',q) \in I_i \} \\ &= \sum \{ D(i,(i',q)) \mid q \in Q \} = \sum \{ C^q(i,i') \mid q \in Q \} \\ &= \sum \left\{ \frac{k(i,j)k(i',j)}{k(j)} \mid j \in J_q, q \in Q \right\} \\ &= C(i,i') \end{aligned}$$

L'analyse interclasses de D_{II}' est donc l'A.F.C. du tableau C_{II} , ce qui donne une interprétation à la somme des carrés des valeurs propres issues de l'A.F.C. du tableau k_{IJ} (puisque, rappelons-le, les valeurs propres issues de l'A.F.C. de C_{II} sont les carrés des valeurs propres issues de l'A.F.C. de k_{IJ}).

4.2 Cas où chaque sous-ensemble J_q est muni d'une partition R_q

On se place ici dans les conditions définies au début du § 3.3 auquel cas J est muni de la partition R définie par l'union des R_q . Le tableau R_{IJ} défini par (10) et dont l'A.F.C. réalise, rappelons-le, l'analyse intraclasses de k_{IJ} relativement à la partition R , est donc partitionné suivant les blocs $\{ R_{IJ_q} \mid q \in Q \}$. On désigne par C_{RII}^q le tableau de Condorcet associé à R_{IJ_q} et par D_{RII}' la superposition des C_{RII}^q :

$$\begin{aligned} \forall i,i' \in I : C_{RII}^q(i,i') &= \sum_{j \in J_q} \frac{R(i,j)R(i',j)}{R(j)} \\ &= C^q(i,i') - \sum_{r \in R_q} \frac{k(i,r)k(i',r)}{k(r)} + \frac{1}{\text{Card}(I)} \end{aligned} \quad (55)$$

Cette expression est déduite de (13) et l'on a tenu compte de ce que les tableaux R_{IJ_q} et k_{IJ_q} ont même marge k_J sur J , et même marge sur I , à savoir 1, le total de chacun de ces deux tableaux étant égal à $\text{Card}(I)$.

Compte tenu de ce que, pour r appartenant à R_q , $k(i,r)$ est égal à 1 si la modalité $q(i)$ adoptée par i à la question q appartient à J_r , sinon $k(i,r)$ est nul, on a en désignant par $r(j)$ la classe à laquelle appartient la modalité j dans la partition R :

$$\forall i, i' \in I : C_R^q(i, i') = C^q(i, i') - \frac{\delta_{r(q(i))}^{r(q(i'))}}{k(r(q(i)))} + \frac{1}{\text{Card}(I)} \left. \vphantom{\frac{\delta_{r(q(i))}^{r(q(i'))}}{k(r(q(i)))}}} \right\} \quad (56)$$

$$= \frac{\delta_{q(i)}^{q(i')}}{k(q(i))} - \frac{\delta_{r(q(i))}^{r(q(i'))}}{k(r(q(i)))} + \frac{1}{\text{Card}(I)}$$

On a alors:

$$\forall i, i' \in I, \forall q \in Q : D_R(i, (i', q)) = C_R^q(i, i')$$

$$= \frac{\delta_{q(i)}^{q(i')}}{k(q(i))} - \frac{\delta_{r(q(i))}^{r(q(i'))}}{k(r(q(i)))} + \frac{1}{\text{Card}(I)} \quad (57)$$

Considérons la partition de I définie par les I_j^q . Comme précédemment, toutes les colonnes (i', q) de I_j^q dans le tableau $D_{RII'}$ sont identiques. En effet, on a:

$$\forall (i', q) \in I_j^q : D_R(i, (i', q)) = \frac{\delta_{q(i)}^j}{k(j)} - \frac{\delta_{r(q(i))}^{r(j)}}{k(r(j))} + \frac{1}{\text{Card}(I)} \quad (58)$$

quantité indépendante de (i', q) .

Il en résulte que l'A.F.C. de $D_{RII'}$ est équivalente à l'A.F.C. du tableau D_{RII} obtenu en cumulant toutes les colonnes associées à un même sous-ensemble I_j^q . Le terme général du tableau D_{RII} est donc donné, compte tenu de (58) et de ce que l'effectif de I_j^q est égal à $k(j)$, par:

$$D_R(i, j) = \sum \left\{ D_R(i, (i', q)) \mid (i', q) \in I_j^q \right\}$$

$$= k(j) \left[\frac{\delta_{q(i)}^j}{k(j)} - \frac{\delta_{r(q(i))}^{r(j)}}{k(r(j))} + \frac{1}{\text{Card}(I)} \right]$$

$$= k(i, j) - \frac{k(j)k(i, r(j))}{k(r(j))} + \frac{k(i)k(j)}{k} = R(i, j) \quad (59)$$

Pour obtenir le résultat précédent, on a tenu compte de ce que $k(i)$ étant constant dans un tableau disjonctif complet, $k(i) / k$ est égal à $1 / \text{Card}(I)$.

L'A.F.C. de $D_{RII'}$ est donc équivalente à celle de R_{II} .

Si l'on considère maintenant la partition de I définie par les $I_j = \{i\} \times Q$ (cf. formule (54)), l'analyse interclasses associée de D_{RII} est l'A.F.C. du tableau D_{RII} de terme général:

$$D_R(i,i') = \sum \{D_R(i,(i',q)) \mid q \in Q\} = \sum \{C_R^q(i,i') \mid q \in Q\} \quad (60)$$

Compte tenu de (55), de ce que $\sum \{C^q(i,i') \mid q \in Q\} = C(i,i')$ (cf. § 4.1.2) et de ce que $k(i)k(i')/k = \text{Card}(Q)/\text{Card}(I)$, on a:

$$\begin{aligned} D_R(i,i') &= C(i,i') - \sum_r \frac{k(i,r)k(i',r)}{k(r)} + \frac{k(i)k(i')}{k} \\ &= C_R(i,i') \end{aligned} \quad (61)$$

L'analyse interclasses de D_{RII} relative à la partition précédente est donc l'A.F.C. du tableau de Condorcet C_{RII} .

Les résultats relatifs aux tableaux k_{IJ} , C_{II} et D_{II} (cf. §4.1) se transposent donc aux tableaux R_{IJ} , C_{RII} et D_{RII} .

5 Analyse relationnelle de données structurées

Marcotorchino & Michaud (1979) ont proposé un certain nombre de critères pour rechercher une partition d'un ensemble I . Cette partition étant caractérisée par la matrice $X = \{x_{ij'} \mid i \in I, i' \in I\}$ du graphe associé, ils recherchent la matrice booléenne X maximisant une fonction $f(X)$ sous les contraintes:

$$\left. \begin{aligned} \forall i, i' \in I : x_{ij'} &= 0 \text{ ou } 1 && \text{(caractère booléen de } X) \\ x_{ij'} &= x_{i'j} && \text{(symétrie)} \\ \forall i, i', i'' \in I : x_{ij'} + x_{i'j''} - x_{ij''} &\leq 1 && \text{(transitivité)} \end{aligned} \right\} \quad (62)$$

contraintes traduisant que X est la matrice d'un graphe associé à une partition.

Nous nous restreignons ici aux critères basés sur le tableau de Condorcet (qui correspond, rappelons-le, au tableau de Condorcet pondéré dans la terminologie de Marcotorchino (1989)) défini à partir d'un tableau de correspondance quelconque k_{IJ} (alors que Marcotorchino se place uniquement dans le cas où k_{IJ} est un tableau disjonctif complet, ce qui a l'avantage de faciliter l'interprétation des critères qu'il utilise). Nous considérons deux critères donnés par Marcotorchino (1989), le premier découlant directement des recherches de Marcotorchino et de Michaud sur le tableau de Condorcet, tandis que le second est basé sur des critères d'inertie.

Avec le premier critère, on recherche la matrice booléenne X maximisant sous les contraintes (62) la quantité

$$C_1(X) = \sum \{C(i,i')x_{ii'} + \bar{C}(i,i')\bar{x}_{ii'} \mid i \in I, i' \in I\} \quad (63)$$

$$\text{avec} \quad \bar{x}_{ii'} = 1 - x_{ii'} \quad (64)$$

ce qui revient à maximiser le critère

$$C'_1(X) = \sum \{(C(i,i') - \bar{C}(i,i'))x_{ii'} \mid i \in I, i' \in I\} \quad (65)$$

qui s'écrit encore, compte tenu de (4):

$$C'_1(X) = 2 \sum \{C(i,i')x_{ii'} \mid i \in I, i' \in I\} - \sum \{C(i,i)x_i \mid i \in I\} \quad (66)$$

$$\text{avec} \quad x_i = \sum \{x_{ii'} \mid i' \in I\} \quad (67)$$

Avec le second critère, on cherche dans l'A.F.C. de k_{IJ} , la partition de I maximisant l'inertie interclasses $I_B(k_{IJ})$ du nuage des profils des lignes, ce qui est équivalent à minimiser l'inertie intraclasses $I_W(k_{IJ})$ ou encore à maximiser la différence

$$C_2(X) = I_B(k_{IJ}) - I_W(k_{IJ}) \quad (68)$$

C'est ce dernier critère qui est proposé par Marcotorchino (1989). Nous pensons qu'il est plus simple de maximiser directement $I_B(k_{IJ})$, ce qui revient encore, compte tenu de (31 bis) à maximiser la quantité:

$$C'_2(X) = 1 + I_B(k_{IJ}) = \sum_{i,i'} \frac{C(i,i') x_{ii'}}{\sum_{i''} k(i'') x_{ii''}} \quad (69)$$

sous les contraintes (62).

Notons que l'on a, puisque $C_2(X) = I_B(k_{IJ}) - I_W(k_{IJ}) = 2 I_B(k_{IJ}) - I_T(k_{IJ})$, et en tenant compte de (5):

$$C_2(X) = 2C'_2(X) - \sum_i \frac{C(i,i)}{k(i)} - 1 \quad (70)$$

Dans le cas où k_{IJ} est un tableau disjonctif complet, les expressions précédentes se simplifient, et l'on a en particulier, compte tenu de (34 bis) et de (35 bis) (M désignant, rappelons-le, la marge constante sur I et égale ici à $\text{Card}(Q)$):

$$C_2(X) = \frac{1}{M} \sum \left\{ (C(i,i') - \bar{C}(i,i')) \frac{x_{ii'}}{x_i} \mid i \in I, i' \in I \right\} - 1 \quad (71)$$

critère dont l'optimisation n'est pas équivalente à celle de $C_1(X)$ et $C'_1(X)$. En effet, d'après (65), les critères $C'_1(X)$ et $M(C_2(X) + 1)$ sont très voisins, mais différents par le fait que, dans le second, s'introduit la quantité x_i au dénominateur, quantité qui n'intervient pas dans $C'_1(X)$.

Si maintenant l'ensemble I (ou J) est muni d'une partition a priori, on peut, quand on recherche une partition de I , s'affranchir de la structure du tableau k_{IJ} . Il suffit dans les critères précédents de remplacer le tableau de Condorcet C_{II} par le tableau C_{SII} (cas d'une partition a priori sur I) ou par le tableau C_{RII} (cas d'une partition a priori sur J).

Remarques

1) Si l'on désire obtenir une partition de J (et non de I) il suffit, dans ce qui précède, d'intervertir le rôle des ensembles I et J , de remplacer le tableau de Condorcet par le tableau de Burt, et la matrice X que l'on recherche par la matrice Y associée à la partition sur J .

2) On peut, au lieu de chercher une partition de I ou de J , rechercher une partition conjointe de I et de J en optimisant un critère de bloc-sériation. S'inspirant toujours de Marcotorchino (1989), on tentera de trouver une matrice booléenne $Z = \{z_{ij} \mid i \in I, j \in J\}$ maximisant la quantité

$$k(Z) = \sum \{k(i,j)z_{ij} + \bar{k}(i,j)\bar{z}_{ij} \mid i \in I, j \in J\} \quad (72)$$

$$\text{avec } \left. \begin{array}{l} \bar{k}(i,j) = \text{Max}_{i,j} k(i,j) - k(i,j) \\ \bar{z}_{ij} = 1 - z_{ij} \end{array} \right\} \quad (73)$$

sous les contraintes:

$$\left. \begin{array}{l} z_{ij} = 0 \text{ ou } 1 \quad (Z \text{ est Booléen}) \\ \forall i \in I : \sum \{z_{ij} \mid j \in J\} \geq 1 \\ \forall j \in J : \sum \{z_{ij} \mid i \in I\} \geq 1 \\ \forall i, i' \in I, \forall j, j' \in J : z_{ij} + z_{ij'} + z_{i'j} - z_{i'j'} \leq 1 \text{ (triades impossibles)} \end{array} \right\} (74)$$

Dans le cas où k_{IJ} est un tableau disjonctif complet, on a $\text{Max}_{i,j} k(i,j) = 1$ et donc $\bar{k}(i,j) = 1 - k(i,j)$, et l'on retrouve la formulation donnée par Marcotorchino.

Bibliographie

Benzécri, J.-P. (1983). Analyse de l'inertie intraclasse par l'analyse d'un tableau de correspondances. ANA. INTRACLASSE. *Cahiers de l'Analyse des Données*, VIII, 3, 351-358.

Cazes, P., Chessel, D. & Doledec, S. (1988). L'analyse des correspondances internes d'un tableau partitionné. Son usage en hydrobiologie. *Revue de Statistique Appliquée*, XXXVI, 1, 39-54.

Cazes, P. & Moreau, J. (1991). Analysis of a contingency table in which the rows and the columns have a graph structure. Journées Internationales. *Analyse des Données et Apprentissage Symbolique-Numérique*. Versailles, 18-20 septembre 1991.

Marcotorchino, F. & Michaud, P. (1979). *Optimisation en analyse ordinale des données*. Paris: Masson.

Marcotorchino, F. (1989). *Liaisons Analyse factorielle-Analyse relationnelle. 1 Dualité BURT-CONDORCET. 2 Formulations Relationnelles en A.F.C.M.* Etude F 142 du Centre Scientifique IBM France.

Moreau, J. (1991). *Analyse factorielle et relationnelle de données structurées*. Lausanne: Centre Vaudois de Recherches Pédagogiques.