

R. ROUSSEAU

**Reconnaissance de la structure de blocs d'un
tableau de correspondance par la classification
ascendante hiérarchique (suite)**

Les cahiers de l'analyse des données, tome 14, n° 3 (1989),
p. 257-266

http://www.numdam.org/item?id=CAD_1989__14_3_257_0

© Les cahiers de l'analyse des données, Dunod, 1989, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

RECONNAISSANCE DE LA STRUCTURE DE BLOCS D'UN TABLEAU DE CORRESPONDANCE PAR LA CLASSIFICATION ASCENDANTE HIÉRARCHIQUE (SUITE)

[REC. BLOC. II]

R. ROUSSEAU*

0 Introduction et rappel du problème

Par structure de blocs, nous entendons ici, comme dans un précédent travail ([REC. BLOC. CAH], in *CAD*, Vol XIII, n°4), la décomposition en blocs diagonaux la plus fine, laquelle, comme on le sait, existe et est unique. Si on effectue sur l'un des deux ensembles en correspondance, (que nous supposons généralement être I), une classification ascendante hiérarchique, (par quoi on entendra toujours, dans la suite, CAH avec pour critère, celui de l'agrégation suivant la variance des partitions, la métrique étant la distance du χ^2), nous dirons que la structure de blocs a été reconnue par la CAH, si chacun des sous ensembles I_b de la partition associée à la décomposition en blocs est une classe de la hiérarchie créée ; auquel cas, toutes les classes créées sont, soit des parties d'un sous ensemble I_b convenable, soit une réunion de tels sous ensembles. On a désigné par (HB) la propriété que toute structure de blocs est reconnue par la CAH: on verra au §6 que cette propriété n'est pas vraie.

Selon [REC. BLOC. CAH], la propriété (HB) équivaut à ce que soit vraie en toute généralité pour un ensemble I en correspondance (non décomposable en blocs) avec un ensemble J , la propriété (G):

$$\forall i \in I, \exists i_1 \in I : \text{crit}(i, i_1) < f_i (1 + d^2(i, I)) \quad (G)$$

dans cette formule, d^2 est la distance du χ^2 ; I désigne le centre de gravité du nuage $N(I)$; et *crit* est le critère de la CAH, calculé comme d'usage.

On a pu démontrer (G) pour $\text{card} I = 2$, mais non au delà.

(*) Université Catholique de l'Ouest, I.M.A., B.P. 808, 49005 ANGERS Cedex.

Dans le présent travail, on envisage plusieurs propriétés équivalentes à celles déjà considérées; on poursuit, au §3, la démonstration de (HB) jusqu'au cas $\text{Card}I = 3$; on considère particulièrement, au §5, le cas d'un tableau sous forme disjonctive complète. On montre au §6, avec l'appui d'un contre-exemple, que la propriété (G) n'est pas vraie en général, à partir de $\text{card}I=4$.

1 Forme équivalente de la propriété (G) pour un tableau de correspondance non décomposable en blocs diagonaux

Par transformations successives des deux membres de (G), on aboutit à une formule (H) où figurent les coefficients de la transition $f_1^I = f_1^J \circ f_J^I$, classiquement calculée pour définir les facteurs. On a:

$$\begin{aligned} \text{crit}(i,i') &= (f_i f_{i'} / (f_i + f_{i'})) d^2(i,i') \\ &= (f_i f_{i'} / (f_i + f_{i'})) \sum \{ ((f_{ij} / f_i) - (f_{ij} / f_{i'}))^2 / f_j \mid j \in J \}. \end{aligned}$$

L'expression \sum de $d^2(i,i')$ peut être réécrite à partir d'une fonction $\text{sym}(i,i')$, symétrique en i et i' , qui n'est autre qu'un produit scalaire entre profils:

$$\begin{aligned} \text{sym}(i,i') &= \sum \{ f_{ij} f_{i'j} / (f_i f_i') \mid j \in J \} = (1/f_i) f_i^i = (1/f_{i'}) f_{i'}^{i'}; \\ d^2(i,i') &= \text{sym}(i,i) + \text{sym}(i',i') - 2 \text{sym}(i,i'); \\ \text{crit}(i,i') &= (f_i f_{i'} / (f_i + f_{i'})) ((1/f_i) f_i^i + (1/f_{i'}) f_{i'}^{i'} - 2 (1/f_i) f_i^i) \\ &= (1 / (f_i + f_{i'})) (f_i f_i^i + f_{i'} f_{i'}^{i'} - 2 f_i f_i^i); \end{aligned}$$

de plus, on a pour le second membre de (G):

$$f_i^i = f_i d^2(i,0) = f_i (1 + d^2(i,I));$$

Ainsi la propriété (G) est équivalente à la propriété (H), définie ci-après, qui concerne la transition composée f_1^I :

$$\begin{aligned} \forall i \in I, \exists i' \in I : (1 / (f_i + f_{i'})) (f_i f_i^i + f_{i'} f_{i'}^{i'} - 2 f_i f_i^i) &< f_i^i \\ \forall i \in I, \exists i' \in I : f_i f_i^i + f_{i'} f_{i'}^{i'} - 2 f_i f_i^i &< f_i f_i^i + f_{i'} f_{i'}^{i'} \\ \forall i \in I, \exists i' \in I : f_i^i &< f_i^i + 2 f_{i'}^i; \end{aligned} \quad (H)$$

Nous dirons donc que la propriété (HB) équivaut à ce que soit vraie en toute généralité pour un ensemble I en correspondance (non décomposable en blocs) avec un ensemble J , la propriété (H).

2 Condition suffisante pour la propriété (H)

Désormais, nous notons $n = \text{card}I$ et $\text{INT} = \text{inertie totale du nuage } N(I)$ par rapport à son centre de gravité. On écrit $H(i, i')$ si est vérifiée l'inégalité (H) pour un couple (i, i') d'éléments distincts de I :

$$H(i, i') : f_i^{i'} < f_i^i + 2 f_i^{i'} .$$

Soit i^o un élément de I ; nous supposons que $H(i^o, i)$ est non vérifiée sauf, peut-être, pour un élément i^o :

$$\forall i \in I - \{i^o, i^o\} : \neg H(i^o, i) ; \text{ (i.e. non H) :}$$

$$\forall i \in I - \{i^o, i^o\} : f_i^i \geq f_{i^o}^{i^o} + 2 f_i^{i^o} ;$$

et cherchons à quelle condition est vérifiée $H(i^o, i^o)$. On a:

$$f_{i^o}^{i^o} = (\text{INT} + 1) - \sum \{f_i^i \mid i \in I - \{i^o\}\} ;$$

or la somme des termes diagonaux et des termes non diagonaux de la matrice f_I^I vaut $n = \text{card}I$; d'où:

$$f_{i^o}^{i^o} = n - (\sum \{f_i^{i'} \mid i \neq i', i \in I, i' \in I\} + \sum \{f_i^i \mid i \in I - \{i^o\}\}) ;$$

$$f_{i^o}^{i^o} \leq n - (\sum \{f_i^{i'} \mid i \neq i', i \in I, i' \in I\} + \sum \{f_{i^o}^{i^o} + 2 f_i^{i^o} \mid i \in I - \{i^o, i^o\}\} + f_{i^o}^{i^o}) ;$$

$$f_{i^o}^{i^o} \leq n - (\sum \{f_i^{i'} \mid i \neq i', i \in I, i' \in I\} + (n-1) f_{i^o}^{i^o} + 2 \sum \{f_i^{i^o} \mid i \in I - \{i^o, i^o\}\}) ;$$

et, compte tenu de ce que la somme des termes $f_i^{i'}$ d'une colonne i' est 1 :

$$f_{i^o}^{i^o} \leq (n-2) + 2 f_{i^o}^{i^o} - (\sum \{f_i^{i'} \mid i \neq i', i \in I, i' \in I\} + (n-3) f_{i^o}^{i^o}) .$$

Ainsi $H(i^o, i^o)$ sera vérifiée si on a la propriété $K(i^o)$ définie ci-après:

$$(n-2) - (\sum \{f_i^{i'} \mid i \neq i', i \in I, i' \in I\} + (n-3) f_{i^o}^{i^o}) < f_{i^o}^{i^o} ;$$

$$(n-2) < (n-2) f_{i^o}^{i^o} + \sum \{f_i^{i'} \mid i \neq i', i \in I, i' \in I\} ; \quad K(i^o)$$

Cette écriture de $K(i^o)$ permet de montrer facilement que (H) et donc (G) sont vraies pour tout tableau de correspondance k_{IJ} non décomposable en blocs si $\text{card}I = n = 2$. Car, en ce cas, $K(i^o)$ équivaut à la positivité de la somme des termes extradiagonaux de f_I^I , donc à l'impossibilité de décomposer en blocs cette matrice; et donc la matrice initiale k_{IJ} .

3 Propriété (K) et résolution du cas $\text{card}I = 3$

La propriété K(i) peut encore s'écrire:

$$(n-3) < (n-3) f_i^i + \sum \{f_i^{i'} \mid i' \neq i, i'' \in I, i' \in I - \{i\}\} .$$

Ainsi la propriété (K) suivante implique les propriétés (H) et (G):

$$\forall i \in I : (n-3) < (n-3) f_i^i + \sum \{f_i^{i'} \mid i' \neq i, i'' \in I, i' \in I - \{i\}\}; \quad (K)$$

Dans le cas d'un tableau de correspondance k_{IJ} non décomposable en blocs, on a les inégalités strictes:

$$\forall i \in I : 0 < \sum \{f_i^{i'} \mid i' \neq i, i'' \in I, i' \in I - \{i\}\}; \quad (Ks)$$

car si l'égalité à 0 est réalisée pour un i , la matrice f_I^I peut se décomposer en deux blocs diagonaux $\{i\} \times \{i\}$ et $\{I - \{i\}\} \times \{I - \{i\}\}$; et de même la matrice initiale k_{IJ} est décomposable.

Dans le cas $\text{card}I = 3$, (K) et (Ks) coïncident; donc (K) et, par conséquent, (H) et (G) sont vérifiées pour $\text{card}I = 3$ si le tableau est non décomposable.

Voici une autre forme de (K) ne mettant en jeu que les termes diagonaux de la matrice f_I^I :

$$\forall i \in I : (n-3) < (n-3) f_i^i + (n-1) - \sum \{f_i^{i'} \mid i' \in I - \{i\}\};$$

$$\forall i \in I : (n-3) < (n-3) f_i^i + (n-1) - (\text{INT} + 1 - f_i^i);$$

$$\forall i \in I : \text{INT} - 1 < (n-2) f_i^i;$$

$$\forall i \in I : f_i^i > (\text{INT} - 1) / (\text{card}I - 2); \quad (K)$$

On peut démontrer directement que, si (G) faux, cette forme de (K) l'est aussi. Car si (G) est faux il existe i tel que:

$$\forall i' \in I - \{i\} : f_i^{i'} \geq f_i^i + 2 f_i^i; \text{ on aurait donc :}$$

$$\text{INT} + 1 - f_i^i = \sum \{f_i^{i'} \mid i' \neq i\}$$

$$\geq (\text{card}I - 1) f_i^i + 2 \sum \{f_i^{i'} \mid i' \neq i\} = (\text{card}I - 1) f_i^i + 2 (1 - f_i^i);$$

$$\text{d'où: } \text{INT} - 1 \geq (\text{card}I - 2) f_i^i .$$

Ceci prouve que la propriété (K) et donc les propriétés (G) et (H) sont vraies pour tout tableau de correspondance d'inertie inférieure ou égale à 1.

On donnera encore deux formes équivalentes de (G) en recourant à l'effet du cumul de deux lignes i et i' en une seule notée $i \cup i'$. On a:

$(f_i + f_{i'}) d^2(i \cup i', I) = f_i d^2(i, I) + f_{i'} d^2(i', I) - \text{crit}(i, i')$;
 d'où pour (G) :

$$\forall i \in I, \exists i' \in I : f_{i'} d^2(i', I) < f_i + (f_i + f_{i'}) d^2(i \cup i', I) ; \quad (G)$$

$$\forall i \in I, \exists i' \in I : f_{i'} (1+d^2(i', I)) < (f_i + f_{i'}) (1+d^2(i \cup i', I)) ; \quad (G1)$$

$$\forall i \in I, \exists i' \in I : f_{i'}^{i'} < f_{i \cup i'}^{i \cup i'} ; \quad (G2)$$

4 Autre forme équivalente, (M), des propriétés (G) et (H)

On considère la propriété (M) suivante:

Pour tout tableau de correspondance k_{IJ} non décomposable en blocs et pour tout couple (i°, i'°) de $(I \times I)$ vérifiant $\text{crit}(i^\circ, i'^\circ) = \inf \{ \text{crit}(i, i') \mid i \in I, i' \in I \}$, on a $\text{crit}(i^\circ, i'^\circ) < \inf (f_{i^\circ}^{i^\circ}, f_{i'^\circ}^{i'^\circ})$; (M)

La propriété (G) implique la propriété (M). En effet, si (M) n'est pas vérifiée, il existe un tableau de correspondance k_{IJ} non décomposable en blocs et une paire d'éléments (i°, i'°) pouvant constituer le 1-er nœud d'une CAH sur I et telle que $\text{crit}(i^\circ, i'^\circ) \geq \inf (f_{i^\circ}^{i^\circ}, f_{i'^\circ}^{i'^\circ})$. Supposons que $f_{i^\circ}^{i^\circ} = \inf (f_{i^\circ}^{i^\circ}, f_{i'^\circ}^{i'^\circ})$. Alors:

$\exists i^\circ \in I, \forall i \in I - \{i^\circ\} : \text{crit}(i, i^\circ) \geq \text{crit}(i^\circ, i'^\circ) \geq f_{i^\circ}^{i^\circ} = f_{i^\circ} (1+d^2(i^\circ, I))$;
 ce qui est la négation de la propriété (G).

Réciproquement, montrons par récurrence sur $n = \text{card}I$ que $(M) \Rightarrow (G)$. Soit k_{IJ} un tableau non décomposable en blocs tel que $\text{card}I = n+1$. Notons:

$$I^\circ = \{ i \mid i \in I ; f_i^i = \inf \{ f_{i'}^{i'} \mid i' \in I \} \} ; \text{ on a pour } i \notin I^\circ \text{ et } i' \in I^\circ :$$

$$f_{i'}^{i'} < f_i^i \leq f_{i^\circ}^{i^\circ} + 2 f_{i'}^{i'} ; \text{ c'est-à-dire } H(i); \text{ reste à montrer } H(i^\circ) \text{ pour } i^\circ \in I^\circ.$$

Deux cas se présentent pour un tel i° selon qu'il peut, ou non, constituer le premier nœud de la CAH sur I. Dans le 1-er cas, notons i' l'autre élément d'un 1-er nœud possible. Alors:

$$\begin{aligned} (f_{i^\circ} + f_{i'}) (1+d^2(i^\circ \cup i', I)) &= (f_{i^\circ} + f_{i'}) + (f_{i^\circ} + f_{i'}) d^2(i^\circ \cup i', I) \\ &= (f_{i^\circ} + f_{i'}) + f_{i^\circ} d^2(i^\circ, I) + f_{i'} d^2(i', I) - \text{crit}(i^\circ, i') \\ &= f_{i^\circ}^{i^\circ} + f_{i'}^{i'} - \text{crit}(i^\circ, i') ; \end{aligned}$$

d'après (M) on a $\text{crit}(i^\circ, i') < \inf (f_{i^\circ}^{i^\circ}, f_{i'}^{i'}) = f_{i^\circ}^{i^\circ}$, car $i^\circ \in I^\circ$. Donc, on a:

$$f_{i'}^{i'} < f_{i^\circ \cup i'}^{i^\circ \cup i'} ; \text{ ce qui montre qu'on a } G2(i^\circ) \text{ et donc } G(i^\circ).$$

Dans le second cas, considérons le tableau k_{IJ} obtenu en cumulant en une seule les deux lignes i' et i'' , où $\{i', i''\}$ est la 1-ère paire agrégée par la CAH sur I, avec $i^\circ \notin \{i', i''\}$, ce qui suppose $\text{card}I = n \geq 2$, où $I = I - \{i', i''\} \cup \{i' \cup i''\}$, et

le tableau k_{IJ} n'est pas décomposable en blocs. D'après l'hypothèse de récurrence il existe $i \in \tilde{I}$ tel que $\text{crit}''(i^\circ, i) < f_i^{i^\circ} = f_i^{i^\circ}$. Or $\text{crit}''(i^\circ, i) = \text{crit}(i^\circ, i)$; donc il existe $i \in \tilde{I}$ tel que $\text{crit}(i^\circ, i) < f_i^{i^\circ}$. Deux cas peuvent alors se présenter selon que i coïncide ou non avec $i'' \cup i'$. Le second cas implique que $G(i)$ est vérifiée dans I . Dans le 1-er cas, on a d'après la formule de la médiane, puisque $i'' \cup i'$ est le 1-er nœud de la CAH sur I : $\text{crit}(i^\circ, i'' \cup i') \geq \inf\{\text{crit}(i^\circ, i'), \text{crit}(i^\circ, i'')\}$. Donc $\text{crit}(i^\circ, i')$ ou $\text{crit}(i^\circ, i'')$ est strictement inférieur à $f_i^{i^\circ}$ et $G(i^\circ)$ est vérifiée; ce qui achève la démonstration car G a été démontrée pour $\text{card}I = 2$ ou $\text{card}I = 3$.

NB Une autre démonstration par récurrence de $(M) \Rightarrow (G)$, basée sur la seule propriété (G) est plus rapide, mais ne montre pas que la difficulté pour obtenir (G) réside dans les cas $i^\circ \in I^\circ$. Avec les mêmes notations que ci-dessus, l'hypothèse de récurrence appliquée au tableau k_{IJ} permet d'affirmer pour $i \in I - \{i'', i'\}$ l'existence de $i \in \tilde{I}$ tel que $\text{crit}(i, i) < f_i^i$. Si i n'est pas $i' \cup i''$, $G(i)$ est vérifiée; et si $i = i' \cup i''$, la formule de la médiane montre que:

$$\inf(\text{crit}(i, i'), \text{crit}(i, i'')) \leq \text{crit}(i, i' \cup i'') < f_i^i;$$

donc $G(i)$ est vérifiée. Quant au cas $i \in \{i'', i'\}$, l'hypothèse (M) suffit à montrer que $G(i)$ est vérifiée puisque $\text{crit}(i', i'') < \inf(f_i^{i'}, f_i^{i''})$, ce qui achève la démonstration, vu que (G) a été démontrée pour $\text{card}I = 2$ ou 3 .

5 Cas des tableaux sous forme disjonctive complète

5.1 L'ensemble J des variables

L'ensemble J des variables (encore appelées 'modalités') d'un tableau sous forme disjonctive complète vérifie la propriété (K) et donc (G) et (H) , pour $\text{card}Q \geq 2$.

Pour une variable j d'un tableau sous forme disjonctive complète on a:

$$\begin{aligned} f_j^j &= f_j (1 + d^2(j, J)) = f_j + f_j d^2(j, J) \\ &= f_j + ((1 - f_j) / \text{card}Q) = (1 + (\text{card}Q - 1) f_j) / \text{card}Q; \end{aligned}$$

d'autre part, on sait que $\text{INT} = (\text{card}J / \text{card}Q) - 1$; d'où :

$$(\text{INT} - 1) / (\text{card}J - 2) = (\text{card}J - 2 \text{card}Q) / ((\text{card}J - 2) \text{card}Q);$$

l'inégalité (K) est donc équivalente à :

$$(1 + (\text{card}Q - 1) f_j) / \text{card}Q > (\text{card}J - 2 \text{card}Q) / ((\text{card}J - 2) \text{card}Q);$$

soit :

$$(\text{card}Q - 1) f_j > 2 (1 - \text{card}Q) / (\text{card}J - 2);$$

Pour $\text{card}Q \geq 2$, on en déduit que $2(1 - \text{card}Q) / (\text{card}J - 2)$ est strictement négatif; et la propriété (K) est donc vraie. Pour $\text{card}Q = 1$, $f_j^j = 1$ et $\text{INT} = \text{card}J - 1$; et alors k_{IJ} est décomposable en blocs.

Ainsi, le tableau de profils B_j^J d'un tableau de Burt vérifie la propriété pour $\text{card}Q \geq 2$:

$$\forall j \in J, \exists j' \in J : B_j^{j'} < B_j^j + 2 B_j^{j'}$$

5.2 L'ensemble I des individus

On démontrera seulement des résultats si $\text{card}J < 3 \text{card}Q$, ou $\text{card}J \leq 3 \text{card}Q$.

Supposons (G) fautive; alors : $\exists i \in I : f_i^i \leq (\text{INT} - 1) / (\text{card}I - 2)$; donc : $\exists i \in I : (\text{card}J - 2\text{card}Q) / (\text{card}Q (\text{card}I - 2)) \geq$

$$(k(i)/k) \sum \{ (1/f_j) (k(i,j)/k(i))^2 \mid j \in J \} = (1/\text{card}Q) \sum \{ k(i,j)/k(i) \mid j \in J \};$$

$$\exists i \in I : (\text{card}J - 2\text{card}Q) / (\text{card}I - 2) \geq \sum \{ 1/k(j(q,i)) \mid q \in Q \},$$

où $j(q,i)$ désigne la modalité de la question q dans laquelle rentre l'individu i . Or la moyenne harmonique est inférieure à la moyenne arithmétique, donc:

$$\exists i \in I : (\text{card}J - 2\text{card}Q) / (\text{card}I - 2) \geq (\text{card}Q)^2 / \sum \{ k(j(q,i)) \mid q \in Q \}; \text{ or:}$$

$$\forall q \in Q : k(j(q,i)) \leq \text{card}I - (\text{card}J_q - 1),$$

car dans toute modalité de J_q rentre au moins un individu i . Donc :

$\sum \{ k(j(q,i)) \mid q \in Q \} \leq \text{card}I \text{card}Q - (\text{card}J - \text{card}Q)$; et, sous l'hypothèse adoptée au départ que (G) est fautive, on aurait :

$$(\text{card}J - 2\text{card}Q) / (\text{card}I - 2) \geq (\text{card}Q)^2 / (((\text{card}I + 1) \text{card}Q) - \text{card}J);$$

ce qui équivaut à l'inéquation du second degré suivante en $(\text{card}J/\text{card}Q)$:

$$(\text{card}J/\text{card}Q)^2 - (\text{card}I + 3) (\text{card}J/\text{card}Q) + 3 \text{card}I \leq 0;$$


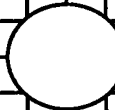
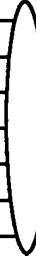
Ainsi $(\text{card}J/\text{card}Q)$ serait compris entre les deux racines de l'équation du second degré associée, qui sont 3 et $\text{card}I$; donc : $3 \leq (\text{card}J/\text{card}Q) \leq \text{card}I$. Or, pour un tableau sous forme disjonctive complète, on a toujours $(\text{card}J/\text{card}Q)$ compris entre 2 et $\text{card}I$.

La propriété (G) est donc vraie pour les individus d'un tableau sous forme disjonctive complète non décomposable en blocs lorsque $\text{card}J < 3 \text{card}Q$ (i.e. lorsque le nombre moyen de modalités par variable est < 3). On montrerait de même que la propriété (G'), définie ci-après, est vraie pour les individus d'un tableau sous forme disjonctive complète quelconque lorsque $\text{card}J \leq 3 \text{card}Q$:

$$\forall i \in I, \exists i' \in I : \text{crit}(i, i') \leq f_i (1 + d^2(i, I)) = f_i^i \quad (G') ;$$

6 Contre-exemples pour le cas cardI ≥ 4

6.1 Contre-exemple aux propriétés (G), (HB), (G')

	j1	j2	j3					jn-1	jn							
i1	1	0	0													
i2	0	1	0													
i3	0	0	1													
				1												
					1											
						1										
							...									
								1								
									1							
i_{n-1}									1							
i_n									0							
i_{n+1}	ε	ε	ε	ε	ε	ε	ε							

On considère le tableau de correspondance k_{IJ} , défini par:

$$\begin{aligned} \text{card}I &= n+1 ; \text{card}J = n \geq 3 ; \\ \forall m \in [1, n] , \forall l \in [1, n] : k(i_m, j_l) &= \delta_m^l ; \\ \forall l \in [1, n] : k(i_{n+1}, j_l) &= \epsilon > 0 ; \end{aligned}$$

ce tableau est non décomposable en blocs. On a:

$$\begin{aligned} f_J &= (1/n) \delta_J ; \quad f_J^{in+1} = f_J ; \\ f_{in+1}^{in+1} &= f_{in+1} (1 + d^2(i_{n+1}, I)) = f_{in+1} = n\epsilon / (n (1+\epsilon)) = \epsilon/(1+\epsilon) ; \\ \forall m \in [1, n] : d^2(i_{n+1}, i_m) &= d^2(i_m, I) = \sum\{(1/(1/n)) (f_j^{im} - (1/n))^2 \mid j \in J\}; \\ &= n ((1-(1/n))^2 + (n-1)(1/n)^2) \\ &= n (n-1) (((n-1)/n^2) + (1/n^2)) = n-1 . \end{aligned}$$

Donc l'écart entre i_m et i_{n+1} pour la CAH est:

$$\text{crit}(i_m, in+1) = (f_{im}f_{in+1}/(f_{im}+f_{in+1})) d^2(i_m, i_{n+1})$$

$$= (1/(n(1+\epsilon))) (n\epsilon/(n(1+\epsilon))) (n-1) ((1/(n(1+\epsilon)))+ (\epsilon/(1+\epsilon)))^{-1}$$

$$= ((n-1)\epsilon / (n(1+\epsilon)^2)) ((1+n\epsilon)/(n(1+\epsilon)))^{-1} = (n-1)\epsilon / ((1+\epsilon)(1+n\epsilon)).$$

Ainsi, sont équivalentes les inégalités suivantes:

$$\text{crit}(i_m, i_{n+1}) \geq f_{i_{n+1}}^{i_{n+1}+1} ;$$

$$(n-1)\epsilon / ((1+\epsilon)(1+n\epsilon)) \geq \epsilon/(1+\epsilon) ;$$

$$n-1 \geq 1+n\epsilon ; \epsilon \leq (n-2)/n .$$

Donc, pour $\epsilon \leq (n-2)/n$ on a:

$$\forall i \in I - \{i_{n+1}\} : \text{crit}(i, i_{n+1}) \geq f_{i_{n+1}} i^{i_{n+1}+1} ; \text{i.e. } G(i_{n+1}) \text{ non vérifiée.}$$

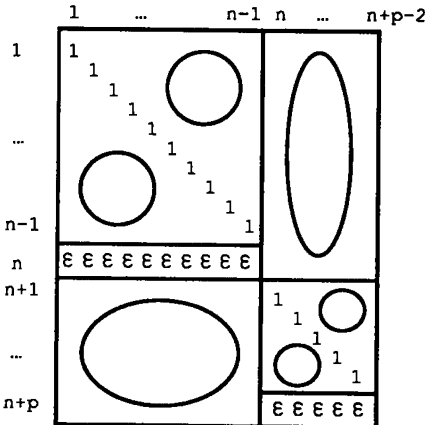
Et de même, pour $\epsilon < (n-2)/n$ on a:

$$\forall i \in I - \{i_{n+1}\} : \text{crit}(i, i_{n+1}) > f_{i_{n+1}} i^{i_{n+1}+1} ; \text{i.e. } G'(i_{n+1}) \text{ non vérifiée.}$$

Ainsi, les propriétés (G), (G') et (HB) ne sont pas vérifiées pour $\text{card } I \geq 4$. On peut également vérifier que la première paire agrégée par la CAH sur I contient nécessairement i_{n+1} ; car pour tout m et m' compris entre 1 et n on a:

$$\text{crit}(i_m, i_{m'}) = 1/(1+\epsilon) ; \text{ or : } \text{crit}(i_m, i_{n+1}) = (n-1)\epsilon / ((1+\epsilon)(1+n\epsilon)) < 1/(1+\epsilon).$$

6.2 Structure de bloc non reconnue par la CAH



On a, dans le tableau proposé ici, une structure de bloc:

$$I = [1, n+p] ; I_1 = [1, n] ; I_2 = [n+1, n+p] ; n \geq p \geq 4;$$

$$J = [1, n+p-2] ; J_1 = [1, n-1] ; J_2 = [n, n+p-2] ;$$

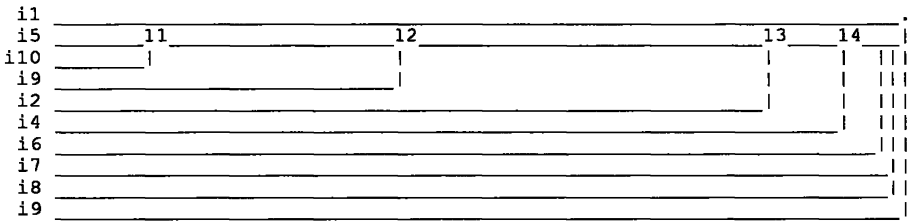
chacun des deux blocs $I_1 \times J_1$ et $I_2 \times J_2$ comprenant une ligne de ϵ sous une matrice carrée dont tous les termes diagonaux valent 1 les autres étant nuls.

Un calcul analogue à ceux du §6.1 montre que l'écart minimum inter-blocs est strictement inférieur à l'écart minimum intrabloc si et seulement si:

$$\epsilon < (p-3) / (p-1) .$$

Dans ce cas, la CAH sur I agrège en premier lieu les éléments n de I₁ et n+p de I₂: la structure de blocs n'est donc pas reconnue.

L'arbre ci-joint a été obtenu pour n = p = 5, ε = 1/10: on voit qu'effectivement les lignes i5 et i10, dernières de chacun des deux blocs (lignes qui sont des suites de nombres égaux à 0.1), s'agrègent d'abord, au niveau le plus bas, pour former le nœud i11. La suite des agrégations ultérieures est contingente, dans la mesure où tout élément du premier ou du deuxième bloc, autre que i5 et i10, est à la même distance de i11.



contre-exemple à RECBLOC
 nombre de colonnes = 8

	j1	j2	j3	j4	j5	j6	j7	j8
i1	1	0	0	0	0	0	0	0
i2	0	1	0	0	0	0	0	0
i3	0	0	1	0	0	0	0	0
i4	0	0	0	1	0	0	0	0
i5	.1	.1	.1	.1	0	0	0	0
i6	0	0	0	0	1	0	0	0
i7	0	0	0	0	0	1	0	0
i8	0	0	0	0	0	0	1	0
i9	0	0	0	0	0	0	0	1
i10	0	0	0	0	.1	.1	.1	.1

7 Conclusion générale

Ainsi, quand l'inertie des partitions calculée suivant la distance du χ^2 est prise pour critère, la reconnaissance de la décomposition en blocs d'un tableau de correspondance par la CAH est assurée si chacun des blocs a une cardinalité inférieure ou égale à 3 ou une inertie inférieure ou égale à 1; elle n'est pas assurée si deux blocs de la partition ont une cardinalité ≥ 4 (comme l'atteste le §6); ou même si c'est le cas pour un seul des blocs (comme le prouvent des calculs complémentaires: cf. [REC. BLOC. Iibis], ce n°, pp. 377-378) .