

I. C. LERMAN

## **Les présentations factorielles de la classification. II**

*RAIRO. Recherche opérationnelle*, tome 13, n° 3 (1979),  
p. 227-251

[http://www.numdam.org/item?id=RO\\_1979\\_\\_13\\_3\\_227\\_0](http://www.numdam.org/item?id=RO_1979__13_3_227_0)

© AFCET, 1979, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Recherche opérationnelle » 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/>

## LES PRÉSENTATIONS FACTORIELLES DE LA CLASSIFICATION. II (\*) (1)

par I. C. LERMAN (2)

Résumé. — *Différentes tentatives de rapprochement entre l'Analyse Factorielle et la classification ont été proposées. On montre que toutes ces tentatives correspondent en fait à des présentations « factorielles » du problème de la classification. La tentation de telles présentations est en effet grande compte tenu du caractère plus établi de l'histoire de l'Analyse Factorielle en composantes qui fournit la solution optimale pour le critère de l'inertie expliquée.*

*Notre but dans ce travail est d'analyser chacune de ces tentatives pour préciser sa véritable nature, la généraliser et l'adapter à des situations nouvelles. Ce faisant, nous contribuerons par des résultats nouveaux à chacune des approches en les situant, dans une synthèse, les unes par rapport aux autres. Ce qui nous permettra de nous rendre compte de l'intérêt relatif de ces différentes tentatives.*

Abstract. — *Attempts have been made to connect Factor Analysis and classification. We show that all these attempts propose, in fact, to interpret the classification problem in terms of Factor Analysis. This kind of interpretation is rather tempting because of the historically established character of the Principal Component Analysis that has the advantage to provide an optimal solution for the retained variance criterion.*

*The purpose of this paper is to analyse each of these attempts in order to find out their real nature and to adapt them to the new situations. Thus the new results will be contributed to each of these approaches by locating their relative positions. This allows us to realise the relative interests of these attempts.*

### A. 2. Critère du « saut minimal »; recherche d'un système de facteurs

#### 1. INTRODUCTION

Nous nous proposons de reprendre ici l'intéressant travail de M. Gondran (*cf.* [2]) mais avec une présentation formelle dégagée de certaines contingences algorithmiques et qui respecte mieux la nature du rapprochement voulu entre l'analyse factorielle et une même classification d'un ensemble  $I$  muni d'un indice de distance ou de proximité. Cette classification est celle d'un niveau donné de l'arbre binaire des classifications sur  $I$ , bâti pas à pas à partir du critère du « saut minimal », et que nous présenterons rapidement au paragraphe suivant.

Bien qu'il s'agisse de l'interprétation d'une classification, il ne faut pas oublier que cette dernière est celle d'un même niveau de l'arbre binaire des classifications

(\*) Reçu avril 1978.

(1) Voir *R.A.I.R.O.*, V-2, juin 1979, p. 107-128.

(2) Laboratoire de Statistique, I.R.I.S.A., Université de Rennes-I, Rennes.

mentionné ci-dessus. D'un certain point de vue ce paragraphe A.2 constitue donc un passage entre la section A et celle B, où il s'agit de présenter la recherche d'un arbre des classifications. D'ailleurs nous verrons au paragraphe B.2.1 que dans ce dernier cas, ce type d'approche devient plus naturel.

Sans restreindre en rien la généralité, nous supposons que  $I$  est muni d'un indice de proximité «  $s$  » à valeurs dans l'intervalle  $[0, 1]$ . Ceci nous conduira à munir cet ensemble de deux opérations  $\vee$  et  $\wedge$  (i.e. « max » et « min ») qu'on notera par l'« addition »  $\oplus$  et par la « multiplication »  $\star$ . L'élément neutre pour  $\oplus$  (resp.  $\star$ ) est 0 (resp. 1). Le triplet  $([0, 1], \oplus, \star)$  définit dans ces conditions un semi-anneau. Ces notations sont beaucoup plus conformes à l'usage que celles de [2]; elles ont d'autre part, un rôle important dans la reconnaissance de la forme retenue de l'équation factorielle pour l'interprétation de la classification.

En exprimant la vraie nature de cette forme de l'équation factorielle, nous verrons qu'il s'agit de la recherche d'un système de facteurs dont chacun est défini comme une fonction sur  $I$ ; alors que dans [2], on parle de façon un peu abstraite de la recherche d'un système de vecteurs de  $S^n$ , où  $S$  représente  $R^+ \cup \{\infty\}$  et  $n$  est le cardinal de  $I$ .

D'autre part, nous nous dégagerons d'un intermédiaire algorithmique non nécessaire dans la construction de cette présentation factorielle de la classification et qui est l'« arbre de longueur minimale » (cf. [2]). En effet, la notion essentielle est celle de proximité ultramétrique sur  $I$  associée à l'arbre binaire des classifications basé sur le critère du « saut minimal » (cf. le paragraphe ci-dessous) auquel plusieurs algorithmes peuvent conduire. De la sorte, certaines démonstrations deviendront plus directes.

## 2. ARBRE DES CLASSIFICATIONS ET MATRICE ASSOCIÉE DES PROXIMITÉS ULTRAMÉTRIQUES

On considère un ensemble  $I$  d'objets de cardinal  $n$ , qu'on peut noter sous la forme  $I = \{1, 2, \dots, i, \dots, n\}$ , muni d'un indice de proximité «  $s$  ».  $s$  est défini comme une application de  $I \times I$  dans l'intervalle  $[0, 1]$ ; associant à chaque couple  $(i, i')$  de  $I \times I$ , un nombre  $s(i, i')$ , compris entre 0 et 1 ( $0 \leq s(i, i') \leq 1$ ), qui « mesure » la « ressemblance » entre  $i$  et  $i'$ . La fonction  $s(i, i')$  respecte les deux conditions :

- (a)  $(\forall (i, i') \in I \times I), s(i, i') = s(i', i)$ ; condition de symétrie;
- (b)  $(\forall i \in I), s(i, i) = 1$ ; condition d'échelle.

En tenant compte de la condition (a), on peut restreindre la définition de  $s$  à l'ensemble des paires d'objets distincts de  $I$ , c'est-à-dire, à l'ensemble  $P_2(I)$  des parties à deux éléments de  $I$ . On supposera pour simplifier que la fonction  $s$  est

injective sur ce dernier ensemble; c'est-à-dire qu'il n'existe pas deux paires distinctes sur lesquelles l'indice  $s$  prend la même valeur. Dans ce dernier cas en effet, l'arbre que nous allons associer à  $s$  est nécessairement binaire.

La définition algorithmique la plus simple de l'arbre des classifications basé sur le critère du « saut minimal » est sans doute, la suivante :

On démarre avec la partition « discrète » où chaque classe contient un seul élément de l'ensemble  $I$  à organiser. A chaque niveau on réunit les deux classes les plus proches au sens de l'indice de proximité suivant entre deux parties disjointes quelconques  $K$  et  $L$  de  $I$  :

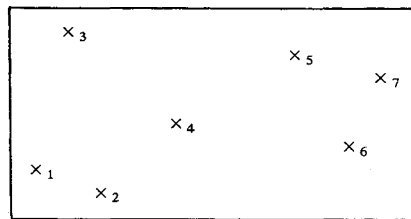
$$p(K, L) = \max \{ s(i, j) / (i, j) \in K \times L \}, \tag{1}$$

De la sorte, dans le cas ci-dessus mentionné où  $s$  est injective sur  $P_2(I)$ ; si le niveau 0 est celui défini par la partition discrète à  $n$  classes, celui 1, définit une partition à  $(n-1)$  classes. Le dernier niveau est celui d'indice  $(n-1)$ , il correspond à la partition « grossière » à une seule classe.

Considérons à titre d'illustration, la configuration géométrique suivante de 7 points et la matrice des proximités associée. Pour établir cette dernière, nous avons posé, en utilisant le centimètre comme unité :

$$(\forall (i, j) \in I^2), \quad s(i, j) = [7 - d(i, j)] / 7,$$

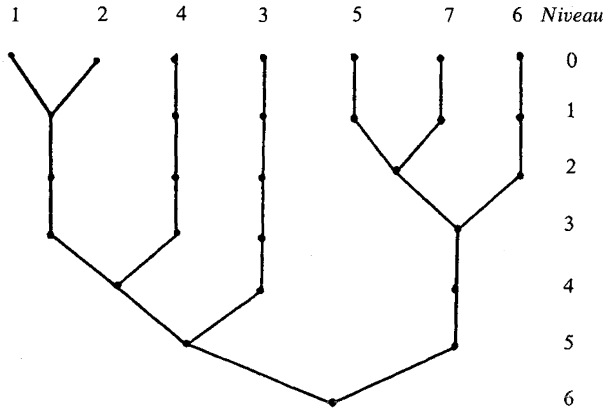
où  $d(i, j)$  est la distance  $i$  et  $j$ . A une homothétie près on obtient la figure suivante :



	1	2	3	4	5	6	7
1	1.						
2	.857	1.					
3	.671	.586	1.				
4	.657	.714	.664	1.			
5	.264	.336	.364	.607	1.		
6	.186	.300	.193	.500	.743	1.	
7	.093	.179	.171	.436	.814	.800	1.

Matrice des proximités.

L'arbre binaire des classifications associé par l'algorithme précédent, est le suivant :



De nombreux algorithmes d'inspirations informatiques différentes conduisent à un tel arbre des classifications. On peut commencer par citer celui de M<sup>me</sup> Nicolau (*cf.* [12]) qui utilise le plus directement une notion de proximité quelconque, en particulier celle (1) ci-dessus pour aboutir à la représentation polonaise de l'arbre des classifications. L'algorithme de Roux (*cf.* [13]) qui fait suite aux travaux de N. Jardine et de R. Sibson (*cf.* [7]) fait appel à la notion de « distance ultramétrique inférieure maximale ». L'algorithme « lexicographique » que nous avons proposé (*cf.* [10]) met en évidence que la seule donnée nécessaire pour la construction d'un tel arbre, a un caractère ordinal : il s'agit de la « préordonnance » sur  $I$ ; c'est-à-dire, le préordre total sur  $P_2(I)$  associé à l'indice de proximité sur  $I$ . On peut également citer l'algorithme de S. C. Johnson (*cf.* [8]). Enfin, l'algorithme de M. Gondran (*cf.* [2]) utilise de façon privilégiée l'arbre de longueur minimale de J. B. Kruskal (*cf.* [9]) de façon conforme à J. C. Gower et G. J. S. Ross (*cf.* [3]).

Nous désignerons par  $I_i$  la classe définie par le nœud formé au niveau  $i$ ; ainsi dans l'exemple précédent  $I_3 = \{5, 6, 7\}$ ,  $I_5 = \{1, 2, 3, 4\}$ .  $a_i$  et  $b_i$  indiqueront les deux classes composantes dont la réunion définit  $I_i$ ; ainsi  $a_3 = \{5, 7\}$  et  $b_3 = \{6\}$ ,  $a_5 = \{1, 2, 4\}$  et  $b_5 = \{3\}$ . On notera par  $\lambda_i$  la valeur de l'indice de proximité entre  $a_i$  et  $b_i$  conformément à la formule (1) :

$$\lambda_i = \max \{ s(i, j) / (i, j) \in a_i \times b_i \}. \quad (2)$$

De par la définition de l'algorithme, on voit sans peine que la suite des valeurs de  $\lambda_i$  est strictement décroissante :

$$\lambda_1 > \lambda_2 > \dots > \lambda_i > \dots > \lambda_{(n-1)}. \quad (3)$$

On distinguera enfin par  $(C_1^1, C_2^1, \dots, C_k^1, \dots, C_{(n-1)}^1)$  la suite des classes de la partition de niveau 1. Pour fixer les idées et sans restreindre en rien la généralité, on suppose cette suite des classes rangée comme suit :

- (a)  $C_1^1 = I_1,$
- (b)  $p(C_1^1, C_k^1)$

est une fonction décroissante de  $k; 1 \leq k \leq (n-1).$

L'ensemble des classes  $\{C_k^1 / 1 \leq k \leq (n-1)\}$  représente, ce que nous appellerons au paragraphe B.1.1 suivant, l'ensemble des « parties maximales » au niveau  $(1+1).$

La proximité ultramétrique entre deux éléments  $i$  et  $j$  de  $I,$  conformément à l'arbre binaire des classifications ci-dessus défini, peut être exprimée de deux façons que nous montrerons bientôt équivalentes. La première est  $s_1^*(i, j) = \lambda_{\iota};$  où  $\iota$  est l'indice du niveau où  $i$  et  $j$  se retrouvent réunis pour la première fois dans  $I_{\iota}.$

La deuxième façon est

$$s_2^*(i, j) = \bigvee_{c \in \mathcal{C}(i, j)} \{s(i_0, i_1) \wedge s(i_1, i_2) \wedge \dots \wedge s(i_r, i_{r+1})\}, \tag{4}$$

où  $c = (i_0 = i, i_1, i_2, \dots, i_r, i_{r+1} = j)$  est une chaîne reliant  $i$  à  $j$  et où  $\mathcal{C}(i, j)$  est l'ensemble de toutes les chaînes reliant  $i$  à  $j.$

LEMME 1 :  $(\forall (i, j) \in I \times I), s_1^*(i, j) = s_2^*(i, j).$

En appelant  $s(i, j)$  valeur de l'arc  $(i, j);$  dans toute chaîne  $c(i, j),$  l'arc le plus « tenu » a une valeur nécessairement inférieure ou égale à  $\lambda_{\iota},$  compte tenu de la définition de  $\lambda_{\iota}$  (cf. [2]). D'où

$$s_2^*(i, j) \leq \lambda_{\iota} = s_1^*(i, j).$$

Soient  $(i_a, i_b)$  le couple de sommets de  $(a_{\iota} \times b_{\iota})$  tel que

$$\lambda_{\iota} = s(i_a, i_b);$$

il existe nécessairement entre  $i$  et  $i_a$  (resp.  $j$  et  $i_b$ ) une chaîne dont les sommets appartiennent à  $a_{\iota}$  (resp.  $b_{\iota}$ ) et dont chaque arc possède une valeur supérieure à  $\lambda_{\iota}.$  Une telle chaîne peut être construite pas à pas en partant des deux sommets à relier et en considérant à chaque fois l'arc de valeur maximale joignant les deux composantes d'un même nœud, lesquelles contenant respectivement deux sommets intermédiaires de la chaîne.

Par conséquent, on a nécessairement

$$s_2^*(i, j) = \lambda_{\iota}.$$

C.Q.F.D.

Nous notons désormais  $s^*(i, j)$  la proximité ultramétrique entre  $i$  et  $j$  et par  $S^*$  la matrice carrée  $I \times I$  des proximités ultramétriques. Dans l'exemple précédent cette matrice est la suivante :

	1	2	3	4	5	6	7
1	1						
2	.857	1					
3	.671	.671	1				
4	.714	.714	.671	1			
5	.607	.607	.607	.607	1		
6	.607	.607	.607	.607	.800	1	
7	.607	.607	.607	.607	.814	.800	1

Matrice des proximités ultramétriques  $S$ .

LEMME 2 : *Les sommets d'un même nœud sont à égale proximité ultramétrique de tout sommet extérieur à ce nœud.*

Soit  $I_i$  un nœud de l'arbre et soit  $j$  un sommet n'appartenant pas à  $I_i$ . Il s'agit de prouver que  $(\forall i, i' \in I_i), s^*(i, j) = s^*(i', j)$ .

Or, on a nécessairement

$$s^*(i, i') > s^*(i, j) \quad \text{et} \quad s^*(i, i') > s^*(i', j),$$

ce qui permet de se rendre compte du résultat en faisant appel à la forme  $s_2^*$  de la proximité ultramétrique [cf. (4)]; ou bien en faisant appel à la propriété caractéristique d'un espace ultramétrique : « Tout triangle est isocèle, la base étant le plus petit côté ».

C.Q.F.D.

Ces lemmes seront utilisés aux paragraphes suivants.

### 3. FORME DE L'ÉQUATION FACTORIELLE

L'intervalle  $[0, 1]$  est l'ensemble des valeurs possibles de l'indice «  $s$  ». Munissons  $[0, 1]$  des deux opérations  $\vee$  et  $\wedge$  (i.e. « max » et « min ») qu'on peut noter par l'« addition » et par la « multiplication »  $\oplus$  et  $\star$ . Dans ces conditions  $([0, 1], \oplus, \star)$  définit une structure de semi-anneau où l'élément neutre pour l'addition est 0 et celui, pour la multiplication, 1.

Étendons la définition de l'addition  $\oplus$  et de la multiplication  $\star$  à l'ensemble  $M_n([0, 1])$  des matrices carrées d'ordre  $n$ , dont les éléments appartiennent à l'intervalle  $[0, 1]$ . Muni de ces deux opérations,  $M_n([0, 1])$  admet alors la même

structure de semi-anneau où l'élément nul (i. e. neutre pour  $\oplus$ ) se trouve défini par la matrice dont tous les éléments sont nuls et où l'élément unité (i. e. neutre pour  $\star$ ) est défini par la matrice dont les seuls éléments non nuls sont ceux de la diagonale principale qui sont tous égaux à 1.

On indiquera par  $0_n$  la matrice nulle et par  $1_n$ , celle, unité. Dans ces conditions,  $S$  (resp.  $S^*$ ) désignant la matrice des proximités (resp. des proximités ultramétriques) sur  $I$  (cf. paragraphe précédent), on a les propriétés suivantes, qu'il est facile de démontrer en utilisant la forme  $s_2^*(i, j)$  [cf. formule (4)] de la proximité ultramétrique  $s^*(i, j)$  entre deux sommets quelconques  $i$  et  $j$  :

$$S^* = \sum_{1 \leq k \leq (n-1)} S^k, \quad (5)$$

$$S^* = S \star S^* = S^* \star S, \quad (6)$$

et

$$S^* = S^* \oplus 1_n = 1_n \oplus S^*, \quad (7)$$

où le signe  $\Sigma$  est, bien entendu, relatif à l'« addition » définie.

La forme de l'équation factorielle nécessaire ici pour interpréter la classification est la même que celle (1) du paragraphe A.1.2 précédent, où l'inconnue est un système de facteurs dont chacun est défini comme une fonction sur  $I$ . Ce système de facteurs, attaché à un même niveau  $\iota$  de l'arbre binaire des classifications, comportera  $(n-\iota)$  éléments dont chacun indiquera une même classe  $C_k^\iota$ ,  $1 \leq k \leq (n-\iota)$ , de ce niveau (cf. paragraphe 2 ci-dessus).

Avec les notations de ce paragraphe, l'équation factorielle devient ici

$$(\forall i \in I), \quad \sum_{i' \in I} s(i', i) \star \varphi(i') = \lambda \star \varphi(i), \quad (8)$$

puisque chaque élément de  $I$  est affecté de la masse unité.

Dans la formule (8), la valeur propre est indiquée par  $\lambda$  qui est un nombre compris entre 0 et 1. La valeur que nous considérons bientôt de  $\lambda$  pour interpréter la classification  $\{C_k^\iota / 1 \leq k \leq (n-\iota)\}$  est précisément définie par  $\lambda_\iota = p(a_\iota, b_\iota)$  [cf. formule (2)].

Commençons par rappeler dans notre cadre les propriétés préliminaires (lemmes 2 et 3 de [2]). Pour cela désignons par  $\Phi_\lambda$  l'ensemble des fonctions sur  $I$  répondant à l'équation (8).

PROPRIÉTÉ 1 : Pour toute fonction  $\varphi$  de  $\Phi_\lambda$ , on a :  $(\forall i \in I), \varphi(i) \leq \lambda$ .

En effet, puisque  $s(i, i) = 1$ ,  $\varphi(i) \leq \lambda \star \varphi(i)$ ; ce qui implique nécessairement  $\varphi(i) \leq \lambda$ , pour tout  $i$  appartenant à  $I$ .





La première, relation donne

$$\varphi(i_1) \leq \varphi(i_2) < \dots \leq \varphi(i_r),$$

alors que la seconde relation donne

$$\varphi(i_r) \leq \varphi(i_{r-1}) \leq \dots \leq \varphi(i_1).$$

Donc

$$\varphi(i_1) = \varphi(i_2) = \dots = \varphi(i_r).$$

C.Q.F.D.

Remarquons que la propriété 2 démontrée reste vraie pour toute valeur propre  $\lambda \leq \lambda_{\iota}$ ; de façon précise, si  $\lambda \leq \lambda_{\iota}$ , toute fonction solution de l'équation factorielle (8) est constante sur chacune des classes de la partition  $\{C_k^{\iota} / 1 \leq k \leq (n - \iota)\}$ .

#### 4. LE SYSTÈME DES FACTEURS

Reprenons la suite des classes  $(C_1^{\iota}, C_2^{\iota}, \dots, C_k^{\iota}, \dots, C_{(n-\iota)}^{\iota})$  de la partition de niveau  $\iota$  de l'arbre binaire des classifications. Cette suite des classes a été définie au paragraphe 2 ci-dessus après l'introduction des paramètres  $\lambda_{\iota}$  [cf. formules (2) et (3)], à partir des conditions (a) et (b). Nous désignerons ici par  $i_k$  un élément courant de la classe  $C_k^{\iota}$ .

Le lemme 2 ci-dessus (cf. paragraphe 2) nous permet d'assurer que la proximité ultramétrique entre deux sommets appartenant respectivement à  $C_k^{\iota}$  et à  $C_h^{\iota}$  ( $k \neq h$ ), est la même quels que soient les deux éléments choisis; l'un dans  $C_k^{\iota}$  et l'autre dans  $C_h^{\iota}$ . Cette proximité ultramétrique pourra donc être notée  $s^*(i_k, i_h)$ .

Considérons dans ces conditions la suite des fonctions  $(\varphi_1, \varphi_2, \dots, \varphi_k, \dots, \varphi_{(n-\iota)})$  où la fonction  $\varphi_k$  sur  $I$  est définie de la manière suivante :

$$\left. \begin{aligned} &(\forall i_k \in C_k^{\iota}), \quad \varphi_k(i_k) = \lambda_{\iota}, \\ &(\forall h \neq k), \quad (\forall i_h \in C_h^{\iota}), \quad \varphi_k(i_h) = s^*(i_k, i_h), \quad 1 \leq k \leq (n - \iota). \end{aligned} \right\} \quad (10)$$

**THÉORÈME :** *La suite des fonctions qu'on vient de définir  $(\varphi_1, \varphi_2, \dots, \varphi_{(n-\iota)})$ , constitue une base unique de l'espace des solutions  $\Phi(\lambda_{\iota})$  de l'équation factorielle (8) relativement à la valeur propre  $\lambda_{\iota}$ .*

Les formules (10) montrent que chacune des fonctions indique une classe de la classification.  $\varphi_k$  est attachée à  $C_k^{\iota}$ ,  $1 \leq k \leq (n - \iota)$ .

Commençons par montrer que  $\varphi_k$  est bien solution de l'équation factorielle (8) qui, nous l'avons vu, se réduit à celle (9). Le premier membre de cette équation

$$\sum_{i' \in I} s(i', i) \star \varphi_k(i'), \tag{11}$$

est, ne l'oublions pas, la transcription de

$$\bigvee_{1 \leq i' \leq n} (s(i', i) \wedge \varphi_k(i')). \tag{11'}$$

Considérons deux cas :

(a)  $i = i_k \in C_k^1;$

(b)  $i = i_h \in C_h^1$  où  $h \neq k$ .

Dans le premier cas, (a), la valeur maximale d'un même argument  $s(i', i_k) \star \varphi_k(i')$ , est obtenue pour  $i' = i_k$ , où elle vaut exactement  $\lambda_1$  qui est la valeur de  $\varphi_k(i_k)$ . On a donc bien

$$(\forall i_k \in C_k^1), \sum_{i' \in I} s(i', i_k) * \varphi_k(i') = \varphi_k(i_k).$$

Dans le deuxième cas (b), l'expression (11) peut se mettre sous la forme

$$\sum_{i' \in C_k^1} s(i', i_h) * \varphi_k(i') \oplus \sum_{j \neq k} \sum_{i' \in C_j^1} s(i', i_h) * \varphi_k(i').$$

Or

$$\varphi_k(i') = \lambda_1 \quad \text{pour } i' \in C_k^1 \tag{12}$$

et

$$\varphi_k(i') = s^*(i_j, i_k) = s^*(i', i_k) \quad \text{pour } i' \in C_j^1$$

et pour un quelconque sommet  $i_k$  de  $C_k^1$ .

D'autre part, par construction, en se donnant un tel sommet  $i_k$ ,

$$\lambda_1 \leq s^*(i', i_k) \quad \text{pour } i' \in C_k^1, \tag{13}$$

de sorte que l'expression (12) peut se mettre sous la forme

$$\sum_{1 \leq i' \leq n} s(i_h, i') \star s^*(i', i_k), \tag{14}$$

où  $i_k$  est un élément quelconque fixé dans  $C_k^1$ .

Mais (14) définit le terme de la  $i_h$ -ième ligne et de la  $i_k$ -ième colonne de la matrice  $S \star S^*$ . Comme  $S \star S^* = S^*$  [cf. formule (6), paragraphe 3 ci-dessus], l'expression (14) est égale à

$$s^*(i_h, i_k) = \varphi_k(i_h), \tag{15}$$

pour  $h \neq k$  [cf. (10)].

On a donc bien aussi dans le deuxième cas (b), le résultat.

Soit à présent  $\varphi$  une fonction quelconque de  $\Phi(\lambda, \nu)$ ; on a

$$(\forall k = 1, \dots, (n-1)), \quad \varphi \geq \varphi(i_k) \star \varphi_k. \tag{16}$$

Cette relation est trivialement vérifiée pour  $i_k \in C_k^1$ . Considérons alors  $i_h \in C_h^1$ , pour  $h \neq k$ . Il s'agit de prouver

$$(\forall i_h \in C_h^1), \quad \varphi(i_h) \geq \varphi(i_k) \star \varphi_k(i_h), \tag{16'}$$

relation qui se met sous la forme

$$(\forall i_h \in C_h^1), \quad \varphi(i_h) \geq s^*(i_h, i_k) \star \varphi(i_k). \tag{17}$$

Or, par définition

$$\varphi(i_h) = \sum_{i' \in I} s(i_h, i') \star \varphi(i'). \tag{18}$$

Soit  $(i'_0 = i_h, i'_1, i'_2, \dots, i'_{(j+1)} = i_k)$  la chaîne reliant  $i_h$  à  $i_k$  qui réalise

$$s^*(i_h, i_k) = s(i'_0, i'_1) \star s(i'_1, i'_2) \star \dots \star s(i'_j, i'_{(j+1)}). \tag{19}$$

La relation (18) entraîne la suite des inégalités

$$\varphi(i'_m) \geq s(i'_m, i'_{(m+1)}) \wedge \varphi(i'_{(m+1)}), \tag{20}$$

pour  $0 \leq m \leq j$ .

Ce qui, par transitivité, implique l'inégalité (17) et donc celle, générale, (16).

Cette relation (16) permet d'écrire

$$\varphi \geq \sum_{1 \leq k \leq (n-1)} \varphi(i_k) \star \varphi_k. \tag{21}$$

Mais on a

$$\varphi(i_r) \leq \varphi(i_r) \star \varphi_r(i_r)$$

puisque,

$$\varphi(i_r) \leq \lambda(r) = \varphi_r(i_r).$$

Par conséquent,

$$\left. \begin{aligned} & (\forall r = 1, \dots, (n-1)), \quad (\forall i_r \in C_r^1), \\ & \varphi(i_r) \leq \sum_{1 \leq k \leq (n-1)} \varphi(i_k) \star \varphi_k(i_r), \end{aligned} \right\} \tag{22}$$

qui est le reflet de l'inégalité contraire de celle (21); donc

$$\varphi = \sum_{1 \leq k \leq (n-1)} \gamma_k \varphi_k, \quad (23)$$

où pour tout  $k = 1, 2, \dots, (n-1)$ ,  $\gamma_k = \varphi(i_k)$ .

Ainsi, toute fonction solution de l'équation factorielle est combinaison linéaire du système de facteurs  $\{\varphi_k / 1 \leq k \leq (n-1)\}$ . Il reste à montrer qu'il s'agit du seul système libre. Supposons en effet que l'on puisse avoir, pour  $k$  fixé,

$$\varphi_k = \sum_{1 \leq h \leq m} \beta_h \varkappa_h, \quad (24)$$

où, pour tout  $h$ ,  $\varkappa_h \in \Phi(\lambda_1)$ .

Comme, pour  $i_k \in C_k^1$ ,

$$\varphi_k(i_k) = \lambda_1 \geq \varkappa_h(i) \quad \text{pour tout } i;$$

il existe nécessairement un indice  $h_0$ , pour lequel

$$\varkappa_{h_0}(i_k) = \lambda_1 \quad \text{avec} \quad \beta_{h_0} > \lambda_1,$$

de sorte que

$$\beta_{h_0} \star \varkappa_{h_0} = \varkappa_{h_0};$$

la relation (24) montre alors que

$$\varkappa_{h_0} \leq \varphi_k. \quad (25)$$

D'autre part, la relation (16) donne

$$\varkappa_{h_0} \geq \lambda_1 \star \varphi_k = \varphi_k, \quad (26)$$

D'où, nécessairement,

$$\varkappa_{h_0} = \varphi_k.$$

C.Q.F.D.

Nous avons ainsi repris de façon plus naturelle la démonstration du théorème de M. Gondran, en nous dégageant de l'accessoire que constitue l'« arbre de longueur minimale » dont l'intérêt est, comme nous l'avons signalé ci-dessus, essentiellement algorithmique.

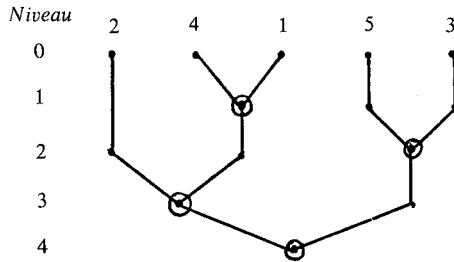
**B. PRÉSENTATION FACTORIELLE DE LA RECHERCHE D'UN ARBRE BINAIRE DES CLASSIFICATIONS**

**B. 1. Critère de l'inertie expliquée**

**B. 1. 1. Recherche d'un système de facteurs et d'axes factoriels associés**

**1. DÉFINITIONS ET PROPRIÉTÉS PRÉLIMINAIRES LIÉES A L'ARBRE DES CLASSIFICATIONS**

$I = \{1, 2, \dots, n\}$  désigne un ensemble de cardinal  $n$ . Un arbre binaire des classifications sur  $I$  dont la suite des niveaux est indexée par  $0, 1, \dots, (n-1)$ ; peut être présenté comme la suite de ses nœuds  $(I_1, I_2, \dots, I_j, \dots, I_{(n-1)})$  respectivement formés aux niveaux  $1, 2, \dots, (n-1)$ . Ainsi, l'arbre binaire suivant sur  $I = \{1, 2, 3, 4, 5\}$  sera représenté par la suite des parties  $(\{1, 4\}, \{3, 5\}, \{1, 2, 4\}, \{1, 2, 3, 4, 5\})$ .



De façon générale, un nœud définira une *partie maximale au niveau  $j$* , dans l'un des deux cas suivants :

- (a) la partie contient un seul élément extérieur à chacune des parties  $I_1, I_2, \dots, I_j$  de  $I$ ;
- (b) il s'agit de l'un des sous-ensembles  $I_h, 1 \leq h \leq j$ , tel qu'aucune autre partie de  $I$ , parmi  $I_1, I_2, \dots, I_{(j-1)}$ , ne l'inclue proprement.

Dans ces conditions, on se rend compte que la suite des nœuds  $(I_1, I_2, \dots, I_{(n-1)} = I)$  d'un arbre binaire des classifications est telle que le nœud  $I_j (1 \leq j \leq (n-1))$  se déduit de la réunion de deux parties maximales au niveau  $j$ .

Réciproquement, soit une suite de parties  $(I_1, I_2, \dots, I_{(n-1)} = I)$  de  $I$  telle que pour tout  $j = 1, 2, \dots, (n-1)$ ,  $I_j$  est formée de la réunion de deux parties maximales au niveau  $j$ . Une telle suite définit nécessairement un arbre binaire des classifications sur  $I$ . Les deux parties maximales dont  $I_j$  est la réunion seront respectivement notées  $a(j)$  et  $b(j)$ .

Nous revenons ici dans le cadre de la classification d'un nuage de points  $\mathcal{N}(I) = \{(M_i, \mu_i) / i \in I\}$  situé dans un espace euclidien muni d'une métrique  $q$ . On a une décomposition du moment total d'inertie du nuage

$$\mathfrak{M} = \sum_{i \in I} \mu_i \|M_i - G\|^2, \quad (1)$$

selon les différents nœuds d'un arbre, en l'occurrence binaire des classifications:

$$\mathfrak{M} = \sum_{1 \leq j \leq (n-1)} \lambda_j, \quad (2)$$

où  $\lambda_j$  est la différence entre le moment d'inertie de la nouvelle classe  $I_j$  et de ses deux sous-classes composantes

$$\lambda_j = \mathfrak{M}(I_j) - [\mathfrak{M}(a(j)) + \mathfrak{M}(b(j))]; \quad (3)$$

il s'agit de la part d'inertie perdue en remplaçant  $I_j$  par ses deux classes composantes  $a(j)$  et  $b(j)$ . On a

$$\lambda_j = \frac{v[a(j)] v[b(j)]}{v(I_j)} \|G[a(j)] - G[b(j)]\|^2, \quad (4)$$

où

$$v[a(j)], \quad v[b(j)] \quad \text{et} \quad v(I_j) = v[a(j)] + v[b(j)]$$

désignent respectivement les masses de  $a(j)$ ,  $b(j)$  et  $I_j$ , où, d'autre part,  $G[a(j)]$  et  $G[b(j)]$  sont les centres de gravité des classes  $a(j)$  et  $b(j)$ .

Nous supposons à présent que l'arbre binaire des classifications sur  $I$  que nous allons devoir bientôt interpréter de façon « factorielle », est bâti de manière ascendante en partant de la partition « discrète » où chaque classe contient exactement un objet, pour aboutir à la partition « grossière » à une classe  $I$ . A chaque pas de la construction de l'arbre on réunit la paire de classes, supposée unique, qui minimise le critère (3) à partir de sa forme (4).

En notant la quantité (3) ou (4) sous la forme  $\Delta[a(j), b(j)]$ , on a

$$\Delta(I_j, I_h) = \frac{1}{v(I_j) + v(I_h)} \{ [v(a(j)) + v(I_h)] \Delta(a(j), I_h) + [v(b(j)) + v(I_h)] \Delta[b(j), I_h] - v(I_h) \Delta[a(j), b(j)] \}, \quad (5)$$

pour toute partie maximale  $I_h$  au niveau  $j$ .

Cette formule permet de montrer que la suite des valeurs  $(\lambda_1, \lambda_2, \dots, \lambda_{(n-1)})$  de (4) attachée à la suite croissante des niveaux de l'arbre, est *strictement croissante*.

## 2. BASE ORTHONORMALE DE FONCTIONS SUR $I$

Commençons par signaler que la forme de l'équation factorielle retenue pour interpréter l'arbre binaire des classifications sur  $I$ , construit en optimisant à chaque pas, la part locale affectée du critère de l'inertie expliquée, est la même que celle du paragraphe A. 1. 2; soit

$$\sum_{i \in I} \mu_i q(i', i) \varphi(i') = \lambda \varphi(i), \quad (6)$$

avec les notations adoptées dans ce paragraphe.

Nous allons, de la même manière, que dans [1], construire dans un cadre ici plus général, une base orthonormale  $(\varphi^1, \varphi^2, \dots, \varphi^r, \dots, \varphi^{(n-1)})$  dans l'espace des fonctions à moyenne nulle sur  $I$ . La fonction  $\varphi^r$  associée au nœud  $I_r$ , de la suite des nœuds  $(I_1, \dots, I_r, \dots, I_{(n-1)})$  représentant l'arbre binaire, est définie comme suit

$$\varphi^r(i) = \begin{cases} \alpha_r & \text{si } i \in a(r), \\ \beta_r & \text{si } i \in b(r), \\ 0 & \text{si } i \notin I_r, \end{cases} \quad (7)$$

où

$$\left. \begin{aligned} \alpha_r v[a_r] + \beta_r v[b_r] &= 0, \\ \alpha_r^2 v[a_r] + \beta_r^2 v[b_r] &= 1. \end{aligned} \right\} \quad (8)$$

En convenant de  $\alpha_r > 0$  et  $\beta_r < 0$ , on obtient :

$$\left. \begin{aligned} \alpha_r &= [v(b_r)/v(a_r) v(I_r)]^{1/2}, \\ \beta_r &= -[v(a_r)/v(b_r) v(I_r)]^{1/2}. \end{aligned} \right\} \quad (9)$$

Il est alors facile de vérifier que le système des fonctions  $\varphi^r$ ,  $1 \leq r \leq (n-1)$ , est bien orthonormal; c'est-à-dire,

$$(\forall 1 \leq r, s \leq (n-1)), \quad \sum_{i \in I} \mu_i \varphi^r(i) \varphi^s(i) = \delta_{rs}, \quad (10)$$

où  $\delta_{rs} = 1$  (resp. 0) selon que  $r = s$  (resp.  $r \neq s$ ).

L'espace des fonctions de moyenne nulle sur  $I$  étant de dimension  $(n-1)$ ;  $(\varphi^1, \varphi^2, \dots, \varphi^{(n-1)})$  constitue une base orthonormale de cet espace.

## 3. PRÉSENTATION FACTORIELLE DE L'ARBRE DES CLASSIFICATIONS

### 3.1. Métrique sur $I$ associée à la base $(\varphi^1, \dots, \varphi^{(n-1)})$

Relativement à la suite des fonctions  $(\varphi^1, \dots, \varphi^{(n-1)})$  que nous venons de définir ci-dessus [formules (7) et (9)], considérons la matrice des proximités



suivantes entre éléments de  $I$  :

$$\{Q(i, i') / (i, i') \in I \times I\}, \quad (11)$$

où

$$(\forall (i, i') \in I^2), \quad Q(i, i') = \sum_{1 \leq r \leq (n-1)} \lambda(r) \varphi^r(i) \varphi^r(i'), \quad (12)$$

l'étendue de la somme précédente peut en fait se réduire à  $n$  ( $i, i' \leq r \leq (n-1)$ ) où  $n(i, i')$  est le plus petit indice du nœud incluant les deux sommets  $i$  et  $i'$ .

A partir de l'extension linéaire de  $\varphi^r$ ,  $1 \leq r \leq (n-1)$ , à l'espace engendré par  $\{(M_i - G) / i \in I\}$ , on peut étendre linéairement  $Q$  par rapport à l'un de ses arguments. Rappelons ici que pour alléger les formules précédentes [(6), (10), (11) et (12)] on a noté  $i$  pour  $(M_i - G)$ .

L'extension linéaire de  $\varphi^r$ ,  $1 \leq r \leq (n-1)$ , sera définie par la formule

$$\varphi^r(\sum_{i \in I} p_i i) = \sum_{i \in I_r} p_i \varphi^r(i) = p(a_r) \alpha_r + p(b_r) \beta_r, \quad (13)$$

où  $\{p_i / i \in I\}$  est une pondération sur  $I$  à coefficients réels, où

$$p(a_r) = \sum_{i \in a_r} p_i, \quad p(b_r) = \sum_{i \in b_r} p_i,$$

où enfin  $\alpha_r$  et  $\beta_r$  ont été définis dans les formules (9).

Avec la formule (13), en considérant une masse totale 1, on vérifie :  $\varphi^r(\sum_{i \in I} \mu_i i) = \varphi^r(G) = 0$ , pour tout  $r = 1, \dots, (n-1)$ .

Avec une telle extension, on a

$$Q(\sum_{i \in I} p_i i, i') = \sum_{1 \leq r \leq (n-1)} \lambda(r) p(a_r) \alpha_r \varphi^r(i') + \sum_{1 \leq r \leq (n-1)} \lambda(r) p(b_r) \beta_r \varphi^r(i') = \sum_{i \in I} p_i Q(i, i'), \quad (14)$$

3.2. THÉORÈME : La suite des fonctions  $(\varphi^1, \varphi^2, \dots, \varphi^{(n-1)})$  définit la suite des facteurs normalisés de moyenne nulle sur  $I$ , de la matrice des proximités sur  $I$  définie par (11). La valeur propre associée à  $\varphi^r$  est  $\lambda(r)$ ,  $r = 1, \dots, (n-1)$ . L'axe factoriel unitaire associé à  $\varphi^r$  au sens de la métrique  $Q$ , est défini par  $[G(a_r) - G(b_r)] / \|G(a_r) - G(b_r)\|$ ,  $r = 1, \dots, (n-1)$ , où la norme  $\|\cdot\|$  reste relative à  $q$ .

L'équation factorielle associée à  $Q$  se met sous la forme

$$\sum_{i' \in I} \mu_{i'} Q(i, i') \varphi(i') = \lambda \varphi(i), \quad (15)$$

pour tout  $i \in I$ .

En remplaçant  $Q(i, i')$  par le second membre de (12) et en considérant pour  $\varphi$  la fonction  $\varphi^s$ , le premier membre de (15) devient

$$\sum_{i' \in I} \mu_r \left[ \sum_{1 \leq r \leq (n-1)} \lambda(r) \varphi^r(i) \varphi^r(i') \right] \varphi^s(i'). \tag{16}$$

En intervertissant à présent les deux signes sommes, l'expression précédente devient

$$\sum_{1 \leq r \leq (n-1)} \lambda(r) \varphi^r(i) \left[ \sum_{i' \in I} \mu_r \varphi^r(i') \varphi^s(i') \right]. \tag{17}$$

Compte tenu à présent du caractère orthonormal de la famille de fonctions  $\{\varphi^r / 1 \leq r \leq (n-1)\}$ , la quantité entre crochets n'est différente de 0 que si  $s=r$  où elle vaut 1. L'expression (17) se réduit par conséquent à  $\lambda(s) \varphi^s(i)$ ; ce qui prouve que la fonction  $\varphi^s$  est facteur, solution de (15), relativement à la valeur propre  $\lambda(s)$ ,  $s = 1, 2, \dots, (n-1)$ .

Reprenons le premier membre de (15) où  $\varphi$  est remplacé par  $\varphi^s$ ; soit

$$\sum_{i' \in I} \mu_r Q(i', i) \varphi^s(i'), \tag{18}$$

compte tenu de la définition de  $\varphi^s$  [cf. (7), (8) et (9)], l'expression précédente se réduit à

$$\alpha_s \sum_{i' \in a_s} \mu_r Q(i', i) + \beta_s \sum_{i' \in b_s} \mu_r Q(i', i), \tag{19}$$

en tenant compte à présent de l'extension linéaire de  $Q$  par rapport à l'un de ses arguments (cf. paragraphe 3.1 précédent) et des valeurs de  $\alpha_s$  et de  $\beta_s$ , on obtient pour (19) :

$$\left[ \frac{v(a_s) v(b_s)}{v(I_s)} \right]^{1/2} Q[G(a_s) - G(b_s), i]. \tag{20}$$

Considérons à présent le second membre de l'équation factorielle (15) avec  $\varphi = \varphi^s$  et  $\lambda = \lambda_s$  donné par la formule (4) ci-dessus. On obtient alors la formule classique en analyse factorielle :

$$\varphi^s = \frac{1}{\sqrt{\lambda_s}} Q(A_s), \tag{21}$$

où

$$A_s = [G(a_s) - G(b_s)] / \|G(a_s) - G(b_s)\|, \tag{22}$$

est l'axe factoriel unitaire au sens de  $q$ , associé au facteur  $\varphi^s$  au sens de  $Q$ .

On a ainsi démontré les relations

$$Q(A_s, i) = \begin{cases} [v(b_s)/v(I_s)] \cdot \|G(a_s) - G(b_s)\| & \text{pour } i \in a_s, \\ -[v(a_s)/v(I_s)] \cdot \|G(a_s) - G(b_s)\| & \text{pour } i \notin b_s, \\ 0 & \text{pour } i \notin I_s. \end{cases} \quad (23)$$

A partir de là, on peut vérifier la  $Q$ -orthonormalité du système d'axes  $\{A_s/1 \leq s \leq (n-1)\}$ .

C.Q.F.D.

#### 4. RECONSTITUTION DU TABLEAU DES DONNÉES

##### 4.1. Cas général

Considérons la situation du paragraphe 3 de A.1.1 où le nuage  $\mathcal{N}(I)$  est associé à la description d'un ensemble fini  $I$  d'individus au moyen d'un ensemble fini  $V$  de variables descriptives numériques. La reconstitution de la suite des composantes  $(x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{im})$  du point  $M_i$  sera définie par la suite des composantes du point  $\hat{M}_i$  défini par

$$\hat{M}_i = G + \sum_{1 \leq r \leq (n-1)} Q(M_i - G, A_r) A_r. \quad (24)$$

Le nuage des points

$$\hat{\mathcal{N}}(I) = \{\hat{M}_i, \mu_i / i \in I\}, \quad (25)$$

a bien pour centre de gravité le point  $G$ , centre de gravité du nuage  $\mathcal{N}(I)$ . D'autre part, on peut vérifier que par rapport à la métrique  $Q$ , le système des facteurs de  $\mathcal{N}(I)$  est bien défini par la suite de fonctions  $(\varphi^1, \varphi^2, \dots, \varphi^{(n-1)})$  précisé ci-dessus [cf. formules (7) et (9)].

On reconstitue la  $j$ -ième composante orthogonale  $x_{ij}$  du point  $M_i$  au moyen de la  $j$ -ième composante orthogonale  $\hat{x}_{ij}$  de  $\hat{M}_i$ , définie par

$$\hat{x}_{ij} = g_j + \sum_{1 \leq r \leq (n-1)} \sqrt{\lambda_r} \varphi^r(i) \sigma(A_r, e_j), \quad (26)$$

où  $g_j$  [resp.  $\sigma(A_r, e_j)$ ] désigne la  $j$ -ième composante du centre de gravité  $G$  (resp. de l'axe factoriel  $A_r$ ).

##### 4.2. Cas d'un tableau de contingence

Dans le cas de l'analyse d'un nuage  $\mathcal{N}(I)$  associé à l'ensemble des lignes d'un tableau de contingence  $I \times J$  (cf. paragraphe 5 de A.1.1), la formule (26)

précédente devient

$$\hat{f}_j^i = p_{.j} + \sum_{1 \leq r \leq (n-1)} \sqrt{\lambda_r} \varphi^r(i) \sigma(A_r, e_j), \tag{27}$$

où  $\hat{f}_j^i$  reconstitue  $f_j^i$  et où

$$A_r = (g_j^a - g_j^b) / \|g_j^a - g_j^b\|$$

avec

$$g_j^a = \frac{1}{p(a_r)} \sum_{i \in a_r} p_i f_j^i, \tag{28}$$

(revoir les notations du paragraphe 5 de A.1.1).

A partir de la formule (27) on obtient

$$\hat{f}_{ij} = p_i p_{.j} \left\{ 1 + \sum_{1 \leq r \leq (n-1)} \sqrt{\lambda_r} \varphi^r(i) q(A_r, e_j) \right\}, \tag{29}$$

où  $q$  est la métrique dont on munit l'espace  $R^J$  [cf. formule (43), paragraphe 5 de A.1.1].

Puisque le point  $g_j = (p_{.1}, \dots, p_{.j}, \dots, p_{.m})$  est centre de gravité du nuage

$$\hat{\mathcal{N}}(I) = \{(\hat{f}_j^i, p_i) / i \in I\},$$

on a bien

$$\sum_{i \in I} \hat{f}_{ij} = p_{.j}, \tag{30}$$

mais on vérifie également à partir de la formule (29) qu'on a

$$\sum_{j \in J} \hat{f}_{ij} = p_i.$$

### 4.3. Cas d'une correspondance symétrique

Une correspondance symétrique sur  $I \times I$  est définie par un système de nombres positifs de somme 1 :

$$\{f_{i' i} / (i, i') \in I \times I\}, \tag{31}$$

tel que

$$(\forall (i, i') \in I \times I), f_{i' i} = f_{i i'}. \tag{32}$$

Il est très rare qu'un tel système (31) de nombres représente celui des fréquences défini par une vraie table de contingence. Le plus souvent, le tableau carré (31) des nombres  $f_{i' i}$ , se trouve défini à partir des formules

$$(\forall (i, i') \in I \times I), f_{i' i} = \sum_{j \in J} (f_{ij} f_{i' j} / p_j), \tag{33}$$

où

$$f_{I \times J} = \{ f_{ij} / (i, j) \in I \times J \}, \quad (34)$$

est la distribution empirique de probabilité sur  $I \times J$  associée à un véritable tableau de contingence  $I \times J$  (cf. paragraphe 5 de A. 1. 1).

Or l'analyse factorielle du tableau (34), orientée vers la définition des facteurs, est équivalente à celle du tableau symétrique (31). Pour reconstituer les nombres de ce dernier conformément aux facteurs  $\varphi^1, \varphi^2, \dots, \varphi^{(n-1)}$ , précédemment définis et associés à l'arbre binaire des classifications sur  $I$  (cf. paragraphe 1 ci-dessus); on « oubliera » la représentation vectorielle de  $I$  à travers  $J$  et on utilisera directement la formule de reconstitution

$$\hat{f}_{ii'} = p_i p_{i'} \left\{ 1 + \sum_{1 \leq r \leq (n-1)} \sqrt{\lambda_r} \varphi^r(i) \varphi^r(i') \right\}, \quad (35)$$

pour tout couple  $(i, i') \in I \times I$ .

On peut alors vérifier que la correspondance (35) admet comme facteurs relativement aux valeurs propres  $\lambda_1, \lambda_2, \dots, \lambda_{(n-1)}$ , la suite des fonctions  $\varphi^1, \varphi^2, \dots, \varphi^{(n-1)}$ ; pour une représentation de  $I$  à travers  $I$  dans  $R^I$ , muni cette fois-ci de la métrique habituelle du  $\kappa^2$ .

C'est cette formule (35), directement proposée dans [1] pour une correspondance symétrique [cf. (31) et (32)] qui a été le point de départ de la recherche des sections A. 1. 2 et B. 1. 1. Par rapport à cette formule, nous avons commencé par nous situer dans le cadre général d'un nuage de points dans un espace euclidien. Nous avons considéré l'arbre binaire des classifications (cf. paragraphe ci-dessus) [resp. la classification (cf. paragraphe 1 de A. 1. 2)] qui s'impose dans cette présentation factorielle; c'est-à-dire, dont la construction est basée sur le critère de l'inertie expliquée. Nous avons en conséquence précisé les valeurs propres [cf. formules (4) du paragraphe 1 ci-dessus (resp. formules (7) du paragraphe 1 de A. 1. 2)] et donné la correspondance entre facteur et axe factoriel associé. Nous avons enfin établi une reconstitution conforme du tableau des données.

## B. 2. Critère du « saut minimal »; recherche d'un système de facteurs

### 1. INTRODUCTION

Le cadre est ici le même que celui du paragraphe A. 2. 1. Dans ce dernier on a interprété factoriellement l'ensemble des classes de la partition définie à un niveau donné de l'arbre binaire des classifications bâti pas à pas sur le critère du « saut minimal ». Cette forme de l'interprétation nous paraît plus

naturellement adaptée pour la présentation factorielle de la suite des nœuds  $(I_1, I_2, \dots, I_1, \dots, I_{(n-1)})$  de l'arbre binaire des classifications (cf. paragraphe 1 de B. 1. 1). On attachera à chaque nœud  $I_\iota$ ,  $1 \leq \iota \leq (n-1)$ , un facteur  $\psi_\iota$  qui l'indiquera, relativement à la valeur propre  $\lambda_\iota$  [cf. formule (2), paragraphe 2 de A. 2).

Commençons par rappeler la forme de l'équation factorielle

$$(\forall i \in I), \quad \sum_{i' \in I} s(i', i) \star \psi(i') = \lambda \psi(i). \quad (1)$$

La propriété 1 du paragraphe 3 de A. 2 reste trivialement vraie; de sorte que l'ensemble  $\Psi(\lambda)$  des solutions de l'équation (1), apparaît comme l'ensemble des fonctions sur  $I$  à valeurs dans l'intervalle  $[0, \lambda]$ , répondant à l'équation

$$(\forall i \in I), \quad \sum_{i' \in I} s(i', i) \star \psi(i') = \psi(i). \quad (2)$$

D'autre part, la propriété 2 (paragraphe 3 de A. 2) est vraie sur  $C_1^1 = I_1$ ; elle est donc vraie sur chaque partie  $I_\iota$  de  $I$ , définie par un nœud de l'arbre;  $1 \leq \iota \leq (n-1)$ .

## 2. LE SYSTÈME DE FACTEURS

A chaque partie  $I_\iota$  de  $I$  définie par un nœud de l'arbre, associons la fonction  $\psi_\iota$  sur  $I$  définie comme suit

$$(\forall i_1 \in I_\iota), \quad \psi_\iota(i_1) = \lambda_\iota$$

[cf. formule (2), paragraphe 2 de A. 2].

$$(\forall i \notin I_\iota), \quad \psi_\iota(i) = s^*(i, i_1), \quad (3)$$

où  $s^*(i, i_1)$  est la proximité ultramétrique entre le point  $i$  et l'un des sommets  $i_1$  de  $I_1$  (voir ici les lemmes 1 et 2, paragraphe 2 de A. 2).

L'ensemble des parties maximales au niveau  $(\iota + 1)$  est précisément l'ensemble des classes  $\{C_1^\iota, C_2^\iota, \dots, C^{(\iota+1)}\}$  de la partition de niveau  $k$  définie dans la section A. 2 (cf. paragraphe 2), où on avait posé  $I_1 = C_1^1$ . Si  $i \in C_j^\iota$ , pour  $j > 1$ ,  $s^*(i, i_1) = s^*(i_j, i_1)$  quel que soit le couple de sommets  $(i_j, i_1)$  de  $C_j^\iota \times C_1^1$ .

**THÉORÈME :** *La suite des fonctions qu'on vient de définir  $(\psi_1, \psi_2, \dots, \psi_1, \dots, \psi_{(n-1)})$  constitue un système unique de facteurs indépendants où le facteur  $\psi_\iota$  est relatif à la valeur propre  $\lambda_\iota$ .*

La démonstration est analogue à celle du théorème correspondant du paragraphe 4 de la section A. 2. On peut remarquer que la fonction  $\psi_\iota$  est

identique à celle  $\varphi_1$  associée au niveau  $\iota$  de l'arbre [cf. formule (10), paragraphe 4 de A.2]. Par conséquent  $\psi_1$  est facteur relativement à la valeur propre  $\lambda_\iota$ ;  $1 \leq \iota \leq (n-1)$ .

D'autre part, pour la raison qu'on vient d'invoquer, si  $\psi$  est facteur relativement à la valeur propre  $\lambda_\iota$ , on a

$$\psi > \gamma_\iota \star \psi_1, \quad (4)$$

où  $\gamma_\iota$  est la valeur commune de  $\psi$  sur  $I_\iota$ . Cette relation est le reflet de la relation (16) (paragraphe 4 de A.2) pour  $k=1$ . D'autre part,

$$(\forall i_1 \in I_\iota), \quad \psi(i_1) \leq \gamma_\iota \star \psi_1(i_1) \quad \text{puisque} \quad \psi_1(i_1) = \lambda_\iota \geq \psi(i_1) = \gamma_\iota,$$

et

$$(\forall i \notin I_\iota), \quad \psi(i) \leq \gamma_\iota \star \psi_1(i) = \psi(i) \star s^*(i, i_1),$$

où  $i_1$  est un point quelconque de  $I_\iota$  et où  $s^*(i, i_1)$  est la proximité ultramétrique de  $i$  à  $i_1$ ; en vertu de la formule (2) ci-dessus. Donc

$$\psi \leq \gamma_\iota \star \psi_1, \quad (5)$$

finalement

$$\psi = \gamma_\iota \star \psi_1, \quad (6)$$

ainsi  $\psi$  est « colinéaire » au facteur  $\psi_1$ .

Les valeurs  $\lambda_\iota$  ( $1 \leq \iota \leq (n-1)$ ) étant fixées; il nous reste à prouver qu'il s'agit du seul système libre de facteurs. Supposons en effet qu'il existe un autre système libre de facteurs respectivement associés aux  $\lambda_\iota$  :

$$\{ \omega_k / 1 \leq k \leq (n-1) \}, \quad (7)$$

et que l'on puisse avoir

$$\psi_1 = \sum_{1 \leq k \leq (n-1)} \alpha_k \omega_k. \quad (8)$$

Puisque  $\psi_1(i) \leq \lambda_\iota$ , pour tout  $i$ ; la relation (8) se réduit nécessairement à

$$\psi_1 = \sum_{\iota \leq k \leq (n-1)} \beta_k \omega_k. \quad (9)$$

Soit  $i_1 \in I_\iota$ ; comme  $\psi_1(i_1) = \lambda_\iota \geq \omega_k(i_1)$ , pour tout  $i$  de  $I$  et tout  $k$  compris entre  $\iota$  et  $(n-1)$ ; il existe nécessairement un indice  $k_0$  tel que

$$\omega_{k_0}(i_1) = \lambda_\iota \quad \text{avec} \quad \beta_{k_0} \geq \lambda_\iota.$$

De sorte que

$$\beta_{k_0} \star \omega_{k_0} = \omega_{k_0}.$$

D'ailleurs,  $k_0$  est nécessairement égal à 1; car pour  $k > 1$ , on a

$$(\forall i \in I), \quad \omega_k(i) < \lambda_1.$$

Ainsi, en vertu de (9),

$$\omega_1 \leq \psi_1. \quad (10)$$

Mais, nous avons montré ci-dessus [cf. formule (4)], la relation

$$\omega_1 \geq \omega_1(i_1) \star \psi_1 = \psi_1, \quad (11)$$

car  $\psi_1(i) \leq \lambda_1 = \omega_1(i_1)$ , pour tout  $i$  de  $I$ .

Donc, finalement,

$$\omega_1 = \psi_1, \quad (12)$$

pour tout  $1 = i_1, \dots, (n-1)$ .

C.Q.F.D.

### C. CONCLUSION

Au terme de notre étude, nous allons chercher à répondre brièvement à la question : « Quel est l'intérêt de ces différents types de rapprochement entre la Classification et l'Analyse factorielle ? ».

Nous avons montré avec insistance que ces rapprochements correspondent tous à des présentations factorielles du résultat de la Classification « non hiérarchique » (section A) ou bien « hiérarchique » (section B).

L'intérêt concret de telles présentations est de « comprendre » la nature de l'interprétation du résultat d'une classification par rapport à celle d'une analyse factorielle et d'avoir ainsi un élément de réponse relativement à la structure la plus adaptée pour représenter synthétiquement la table de données; s'agit-il d'une structure en classes de proximité? ou bien d'une structure par axes en composantes (cf. paragraphe A.1.1, surtout).

Ces présentations ont d'autre part le mérite d'« expliquer » aux praticiens des méthodes des deux formes d'approche de l'analyse des données (i. e. Analyse Factorielle et Classification) pourquoi il arrive qu'une région du plan factoriel soit dominée par une classe (cf. paragraphe A.1.2); pourquoi une opposition entre deux parties du plan factoriel, reflète une opposition entre deux nœuds de



l'arbre des classifications (*cf.* paragraphe B. 1 et [6]). Ces analogies ont d'autant plus de sens que le même critère est en jeu (*cf.* paragraphes A. 1 et B. 1). C'est au paragraphe A. 1. 1 que la liaison est faite de la façon la plus intime puisque les deux côtés du tableau de données sont le plus explicitement présents dans le rapprochement entre les deux types de structure de synthèse.

Un intérêt majeur de ces présentations factorielles de la classification est théorique. Ces dernières permettent des rapprochements formels insoupçonnés (*cf.* paragraphes A. 2 et B. 2, surtout) entre les deux types de techniques en analyse des données. Ces interprétations factorielles séduiront certainement les tenants de l'approche géométrique dans l'analyse des données. Toutefois, ils doivent se garder de croire que de cette façon ils saisiront le propre de la Classification qui reste essentiellement algébrique, combinatoire et statistique.

En effet, alors que pour l'analyse factorielle, l'équation a un rôle *opérationnel* puisqu'elle conduit à la solution de la représentation condensée; ici, l'équation factorielle ne fait qu'« habiller » une classification ou un arbre des classifications obtenus par un algorithme tout à fait indépendant qui optimise un critère de forme donnée. D'ailleurs, deux critères seulement ont été ici envisagés; ce sont les plus classiques. Mais, il en existe d'autres, d'importance. On peut signaler par exemple celui basé sur la quantité d'information mutuelle pour l'analyse d'un tableau de contingence. Indiquons en ce qui nous concerne, que l'indice de proximité entre classes, permettant de construire l'arbre « détaillé » des classifications de notre méthode de classification hiérarchique, ne se réduit à aucun des critères présentés. Il s'agit en effet de la vraisemblance d'un lien qui a la même forme que (1) (paragraphe 2 de A. 2); mais où les similarités  $s(i, j)$  se réfèrent à une échelle  $[0, 1]$  de probabilité où la notion de « ressemblance » est mesurée par une notion de « vraisemblance ».

Pour terminer, nous dirons, compte tenu de tout ce qui précède : « Une équation factorielle ne permettra jamais de construire une classification ».

#### BIBLIOGRAPHIE

1. J. P. BENZÉCRI, *Analyse de correspondance et hiérarchie indicée* dans L'analyse des données, chap. IIB, n° 11, Dunod, Paris, 1973.
2. M. GONDRAU, *Valeurs propres et vecteurs propres en classification hiérarchique*, Revue R.A.I.R.O., série R.
3. J. C. GOWER et G. J. S. ROSS, *Minimum Spanning Trees and Single Linkage Cluster Analysis*, Appl. Stat., vol. 18, n° 1, p. 54-64.
4. F. R. HELMERT, *Die Genauigkeit der formel von Peters zur berechnung des wahrscheinlichen beobachtungs fehlers director beobachtungen gleicher genauigkeit*, Astron. Nachr., vol. 88, p. 112-131.

5. N. HOWARD, *Least Squares Classification and Principal Component Analysis: A Comparison in Quantitative Ecological Analysis in the Social Sciences*, M. DOGAN et S. ROKKAN, éd., Cambridge, M.I.T., Press, 1969.
6. M. JAMBU, *Programme de calcul des contributions mutuelles entre classes d'une hiérarchie et facteurs d'une correspondance*, Les cahiers de l'analyse des données, vol. 1, n° 1, 1976, Paris.
7. N. JARDINE et R. SIBSON, *Mathematical Taxonomy*, Wiley, New York, 1971.
8. S. C. JONHSON, *Hierarchical Clustering Schemes*, Psychometrika, vol. 32, n° 3, 1967.
9. J. B. KRUSKAL, *On the Shortest Spanning Subtree of a Graph and the Travelling Salesman Problem*, Proc. Amer. Math. Soc., n° 7, 1956.
10. I. C. LERMAN, *Les bases de la classification automatique*, Gauthier-Villars, Paris, 1970.
11. I. C. LERMAN, *Reconnaissance et classification des structures finies en Analyse des données*, rapport n° 70, I.R.I.S.A., Rennes, 1977.
12. M. H. NICOLAÛ, *Analyse d'un algorithme de classification*, Thèse de 3<sup>e</sup> cycle, Université Paris-VI, I.S.U.P., novembre 1972.
13. M. ROUX, *Un algorithme pour construire une hiérarchie particulière*, Thèse de 3<sup>e</sup> cycle, Université Paris-VI, L.S.M., 1968.