

J.-L. CHANDON

F.-F. BOCTOR

**Approximation d'une préordonnance
par une partition**

RAIRO. Recherche opérationnelle, tome 19, n° 2 (1985),
p. 159-184

http://www.numdam.org/item?id=RO_1985__19_2_159_0

© AFCET, 1985, 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/>

APPROXIMATION D'UNE PRÉORDONNANCE PAR UNE PARTITION (*)

par J.-L. CHANDON ⁽¹⁾ et F.-F. BOCTOR ⁽¹⁾

Résumé. — *Nous nous intéressons aux problèmes d'analyse typologique où les mesures de similarité entre objets à classer sont de type ordinal. Plusieurs chercheurs ont proposé de construire une partition de l'ensemble à classer, de façon à maximiser un critère ordinal d'adéquation de la partition à la préordonnance. Une présentation unifiée des divers critères classiques est proposée.*

Après un exemple formel de ces critères et de la manière dont ils prennent en compte les paires équivalentes, une nouvelle approche, appelée « partitions efficientes » est développée. Un algorithme permettant d'énumérer toutes les partitions efficientes de rang zéro est présenté.

Mots clés : Analyse des données; Typologie; Préordonnance.

Abstract. — *In this paper we consider the problems of constructing a partition of a set of N objects which maximizes some ordinal criterion of goodness of the fit between the partition and the objects' similarity matrix.*

A unified review of various ordinal criteria of the goodness of the fit is given. After an evaluation of the classical criteria, focussing on their treatment of tied pairs, a new approach called « efficient partitioning » is proposed.

Mots clés : Data Analysis; Clustering; Set partitioning.

I. INTRODUCTION

Nous nous intéressons aux problèmes d'analyse typologique présentant les deux caractéristiques suivantes :

(a) Les mesures de proximité utilisées pour définir les dissimilarités (res : similarités) entre objets sont de type ordinal. L'information disponible se présente sous la forme d'une préordonnance ⁽²⁾ qui permet seulement de décider si les objets de la paire p sont plus semblables entre eux que les objets

(*) Reçu en juin 1982.

⁽¹⁾ Institut d'Administration des Entreprises, Clos Guiot, chemin des Camus, 13540 Puyricard.

⁽²⁾ Terminologie introduite par Benzecri (1965).

de la paire q . Elle ne permet pas de quantifier avec précision le degré de similitude.

(b) L'information disponible ne permet ni d'avoir une idée *a priori* de la structure de l'ensemble des données, ni de formuler des hypothèses précises concernant le nombre de classes d'objets.

On peut diviser les méthodes de classification en deux catégories : celles qui recherchent une partition et celles qui recherchent une hiérarchie de partitions. On se limitera ici à la recherche d'une partition de l'ensemble des objets. Cette structure a le mérite de la simplicité. Elle a cependant l'inconvénient de requérir, pour la plupart des méthodes de classification disponibles actuellement, une hypothèse concernant le nombre de classes.

Nous présentons ici une nouvelle approche, appelée partition efficiente, qui ne nécessite pas de choisir *a priori* le nombre de classes.

Nous montrerons également que le critère ordinal d'adéquation d'une partition à une préordonnance que nous proposons, présente plusieurs avantages vis-à-vis des critères précédemment utilisés dans la littérature typologique. Ce critère peut également servir à déterminer, parmi la suite des partitions emboîtées générées par une méthode de classification hiérarchique, celles qui sont les plus proches de la préordonnance.

Les méthodes de partition peuvent être classées en trois catégories :

1. Les méthodes nécessitant un choix préalable du nombre de classes

Parmi celles-ci, la méthode des « k -means » de Mac Queen (1967), et la méthode des « nuées dynamiques » de Diday (1971, 1979), sont les plus connues. Cette dernière permet de varier les centres de classes choisis pour initialiser un algorithme de réallocation des objets aux classes. En comparant ensuite les diverses partitions obtenues, on décide de ne retenir que les classes stables ou « formes fortes ». On peut également faire varier le nombre de classes et observer que certaines classes restent vides. Une telle approche s'est avérée utile en pratique, mais il n'est pas certain qu'elle converge vers le nombre de classe optimal. De plus, le critère d'adéquation utilisé est en général de type métrique, quoique certaines applications au traitement des données de préférences, telles que Chandon, Lemaire (1977) et Lemaire (1977), aient utilisé un critère ordinal.

Lorsque les objets à classer sont décrits sur des variables discrètes, la méthode des « partitions centrales » de Régnier (1965), permet d'approcher, à l'aide d'un algorithme de transferts, la partition la plus proche des partitions induites

par chacune des variables discrètes. Malheureusement, ici aussi, la convergence vers la (ou les) partition(s) centrales n'est pas certaine et le problème du choix d'une partition initiale demeure ⁽³⁾.

2. Les méthodes qui remplacent le choix préalable du nombre de classes par le choix de seuils d'homogénéité ou de séparation des classes

La méthode la plus simple est celle du « leader » décrite par Tou et Gonzalez (1974) et par Hartigan (1975). Un seuil d'agrégation d_A est choisi et les objets sont considérés séquentiellement. Le premier objet est le premier « leader ». L'objet suivant est affecté au premier groupe rencontré pour lequel sa distance au leader est inférieure au seuil d'agrégation. S'il n'existe pas de tel groupe, on en crée un nouveau dont il devient le leader et ainsi de suite jusqu'à épuisement des objets. Le résultat obtenu dépend du seuil d_A et de l'ordre dans lequel on considère les objets.

Une seconde approche due à Sébestyen (1962) consiste à se donner un seuil d'agrégation d_A et un seuil de séparation d_B , avec $d_A < d_B$. L'algorithme est une généralisation de la méthode du leader où les objets leader sont remplacés par le centre de gravité de la classe.

Il est clair que le choix des seuils est tout aussi arbitraire que celui du nombre de classes. Diverses variantes de cette approche sont mises en œuvre par le programme Reloc de Wishart (1969) et Isodata de Ball et Hall (1965, *a* et *b*).

La méthode des « boules optimisées » de Benzecri (1979) propose une variante itérative dans laquelle on fixe le nombre maximum de classes et le rayon R initial des classes. Chaque objet est affecté à la classe la plus proche si sa distance au centre de classe est inférieure à R , sinon il devient centre d'une nouvelle classe. Si le rayon initial est insuffisant pour affecter tous les objets à une classe, il est doublé. Les centres de classe sont les centres de gravité et ils sont recalculés après que tous les objets aient été affectés.

L'algorithme converge rapidement, mais le choix du nombre maximum de classes et du rayon demeure arbitraire ⁽⁴⁾.

3. Les méthodes qui déterminent simultanément le nombre de classes et l'affectation des objets aux classes

Le principe de base de ces méthodes est de choisir un critère d'adéquation d'une partition à la préordonnance puis de mettre en œuvre un algorithme maximisant le critère choisi.

⁽³⁾ Le problème des partitions centrales a été résolu récemment par Marcotorchino (1981) à l'aide d'une autre approche ne nécessitant pas de choisir le nombre de classes.

⁽⁴⁾ Un grand rayon conduira automatiquement à un petit nombre de classes.

Les différences entre les méthodes résident dans le critère d'adéquation choisi et dans l'algorithme (exact ou heuristique), utilisé pour atteindre ou approcher la (ou les) partition(s) optimale(s).

Parmi les critères d'adéquation les plus connus, nous examinerons le critère de Benzecri (1973), le critère de la différence symétrique de De la Vega (1967) et le critère de Lerman (1969).

Nous adoptons ici cette troisième approche. Nous donnons d'abord une présentation unifiée des différents critères classiques d'adéquation, puis nous développons ensuite un nouveau critère ordinal de l'adéquation d'une partition à une préordonnance. Enfin, un algorithme d'énumération des partitions efficaces de rang zéro est présenté.

Il existe d'autres méthodes qui déterminent simultanément le nombre de classes et l'affectation des objets aux classes, telle que la méthode des « pôles d'attraction » de Lerman et Leredde (1977), Lerman (1981). Ces méthodes ne seront pas envisagées ici car elles n'optimisent pas un critère basé sur la préordonnance.

II. APPROXIMATION D'UNE PRÉORDONNANCE PAR UNE PARTITION

Nous allons transformer ce problème en celui, plus facile à résoudre, de la comparaison de deux préordres totaux définis sur un même ensemble.

a) Définitions

Soient : N l'ensemble à classer et n le nombre d'éléments de N .

M l'ensemble des sous-ensembles à 2 éléments (paires) de N .

m le nombre des éléments de M ; $m = n(n-1)/2$.

d une application de M dans $\mathbb{N}^+$ indicatrice du rang de dissimilarité entre paires d'objets.

P le préordre total (préordonnance) défini sur M par la relation :

$$pPq \Leftrightarrow dp \leq dq ; p, q \in M$$

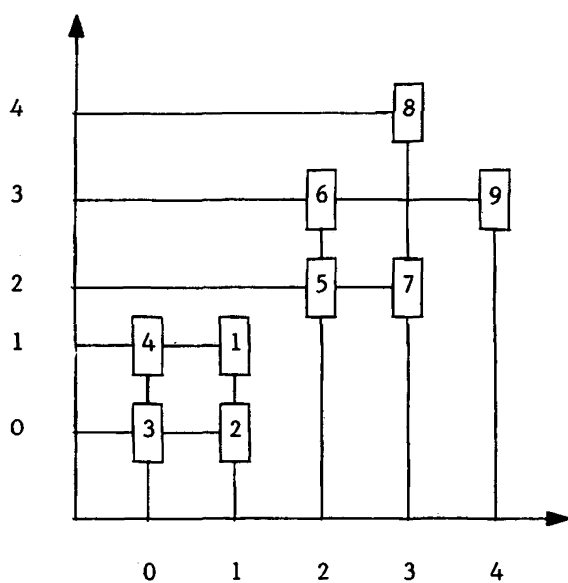
On dira que la paire p précède la paire q si les deux objets composant p se ressemblent plus que ceux composant q . L'ensemble ordonné des classes d'équivalence du préordre P sera noté

$$P_1 < P_2 \dots < P_t$$

Le préordre P sera représenté par son graphe W_p dans $M \times M$

$$W_p = \{ (p, q) \in M \times M / pPq \}$$

W_p est l'ensemble des couples de paire (p, q) tels que les objets de la paire p



a) Positions géographiques des 9 villes

	1	2	3	4	5	6	7	8	9
1	-	1	2	1	2	4	4	7	7
2		-	1	2	4	6	5	9	8
3			-	1	5	7	7	10	10
4				-	4	5	6	8	9
5					-	1	1	4	4
6						-	2	2	3
7							-	3	2
8								-	2
9									-

b) Rang des distances entre les 9 villes

Figure 1. — L'exemple de 9 villes.

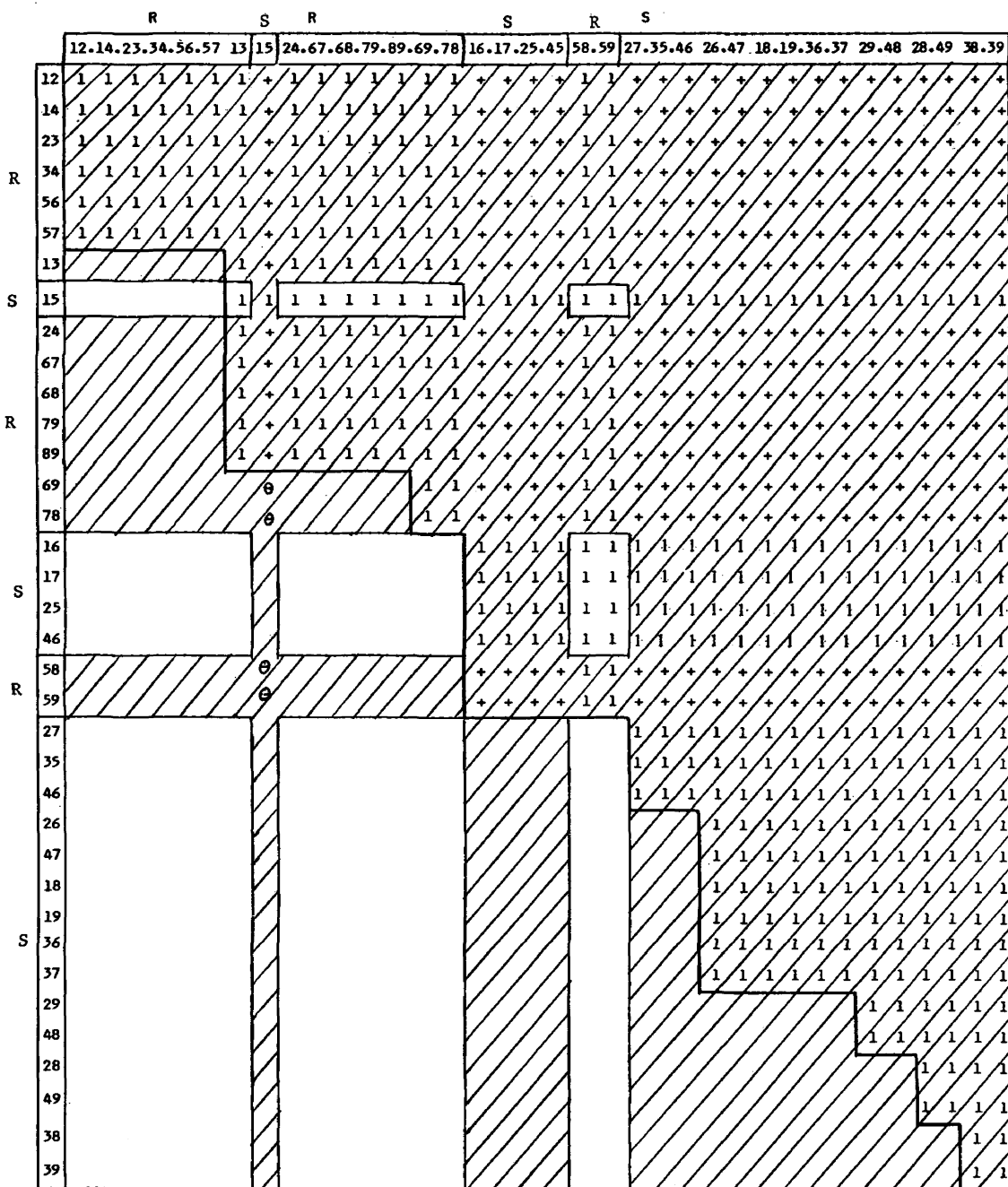


Figure 2. — L'exemple de 9 villes : représentation simultanée de l'ensemble W_p et de l'ensemble W_π pour la partition : $\pi = (1, 2, 3, 4), (5, 6, 7, 8, 9)$.

Légende : 1 indique que $(p, q) \in W_p - R \times S$ θ indique que $(p, q) \in R \times S - W_p$
 + indique que $(p, q) \in W_p \cap R \times S$ // indique que $(p, q) \in W_\pi$
 1 ou + indique que $(p, q) \in W_p$

sont séparés par une distance inférieure ou égale à celle séparant les objets de la paire q .

Soit maintenant : π une partition quelconque de N ,

R l'ensemble des paires réunies selon la partition π ,

S l'ensemble des paires séparées selon π .

On notera r , et s , le nombre d'éléments des ensembles R et S .

Soit Π le préordre total à 2 classes d'équivalence, $R < S$, induit sur M par la partition π en considérant que deux objets quelconques d'une paire réunie sont éloignés d'une distance δ_R inférieure à la distance δ_S de deux objets quelconques d'une paire séparée.

Le préordre Π sera représenté par son graphe W_π dans $M \times M$.

$$W_\pi = \{ (p, q) \in M \times M / p \Pi q \}$$

La comparaison d'une partition π à une dissimilarité d se ramène donc à la comparaison de deux préordres P et Π tous deux définis sur M .

Pour illustrer ces différentes définitions, nous utiliserons l'exemple de la figure 1 ⁽⁵⁾. Sur la figure 2, nous présentons l'ensemble W_P et l'ensemble W_π liés à une partition π .

b) Croisement de deux préordres

Afin d'établir une base de comparaison entre les diverses mesures d'associations entre 2 préordres, formons la table croisant les classes d'équivalence de P avec les classes d'équivalence de Π .

		Π : Préordre induit par la partition		
		$\overbrace{\quad R \quad S \quad}$		
P : Préordre induit par la matrice des dissimilarités	P_1	m_{11}	m_{12}	$m_{1.}$
	$\vdots$			
	P_2	m_{21}	m_{22}	$m_{2.}$
	$\vdots$			
	P_i	m_{i1}	m_{i2}	$m_{i.}$
	$\vdots$			
	P_l	m_{l1}	m_{l2}	$m_{l.}$
		$m_{.1}$	$m_{.2}$	m

⁽⁵⁾ Exemple tiré de Chandon, Pinson (1981).

Il en découle les relations suivantes :

$$m_{i1} = |P_i \cap R|$$

$$m_{i2} = |P_i \cap S|$$

$$m_{i.} = |P_i|$$

$$m_{.1} = r$$

$$m_{.2} = s$$

En utilisant la décomposition canonique du graphe W_P d'un préordre P

en sa partie symétrique, $I_P = \{(p, q) \in M \times M / d_p = d_q\}$

et sa partie asymétrique, $V_P = \{(p, q) \in M \times M / d_p < d_q\}$

et en considérant de plus la partie $\bar{V}_P$, opposée à V_P .

$$\bar{V}_P = \{(p, q) \in M \times M / d_p > d_q\}$$

On observe que l'ensemble V_P^c complémentaire de V_P par rapport à $M \times M$ est l'union des ensembles I_P et $\bar{V}_P$.

On note également les 2 relations suivantes :

$$W_P = V_P \cup I_P$$

$$W_P^c = \bar{V}_P$$

il est clair que tout couple de paires est affecté à l'une des trois catégories suivantes :

Paires équivalentes : $(p, q) \in I_P$

Paires concordantes : $(p, q) \in V_P$

Paires discordantes : $(p, q) \in \bar{V}_P$

En croisant ces 3 catégories pour les deux préordres P et Π on obtient une table qui précise les 9 modalités de relation entre les 2 préordres.

		Préordre Π			
		$\delta_p < \delta_q$	$\delta_p = \delta_q$	$\delta_p > \delta_q$	
Préordre P	$dp < dq$	$V_P \cap V_\pi$	$V_P \cap I_\pi$	$V_P \cap \bar{V}_\pi$	$\left. \begin{array}{l} V_P \\ I_P \\ \bar{V}_P \end{array} \right\} W_P$
	$dp = dq$	$I_P \cap V_\pi$	$I_P \cap I_\pi$	$I_P \cap \bar{V}_\pi$	
	$dp > dq$	$\bar{V}_P \cap V_\pi$	$\bar{V}_P \cap I_\pi$	$\bar{V}_P \cap \bar{V}_\pi$	
		$\underbrace{V_\pi \quad I_\pi}_{W_\pi}$		$\bar{V}_\pi$	$M \times M$

Le cardinal des divers sous-ensembles de $M \times M$ peut être calculé comme suit à partir de la table des m_{ij}

$$\begin{aligned}
 |M \times M| &= m^2 \\
 |I_P| &= \sum_{i=1}^l m_{i1}^2 & |I_\pi| &= r^2 + s^2 \\
 |V_P| = |\bar{V}_P| &= \left(m^2 - \sum_{i=1}^l m_{i1}^2 \right) / 2 & |V_\pi| = |\bar{V}_\pi| &= (m^2 - r^2 - s^2) / 2 = r \times s \\
 |W_P| &= \left(m^2 + \sum_{i=1}^l m_{i1}^2 \right) / 2 & |W_\pi| &= (m^2 + r^2 + s^2) / 2 \\
 |I_P \cap I_\pi| &= \sum_{i=1}^l (m_{i1}^2 + m_{i2}^2) & |I_P \cap V_\pi| = |I_P \cap \bar{V}_\pi| &= \left(\sum_{i=1}^l m_{i1}^2 - \sum_{i=1}^l (m_{i1}^2 + m_{i2}^2) \right) / 2 \\
 |V_P \cap I_\pi| = |\bar{V}_P \cap I_\pi| &= \left(r^2 + s^2 - \sum_{i=1}^l (m_{i1}^2 + m_{i2}^2) \right) / 2 \\
 |V_P \cap V_\pi| = |\bar{V}_P \cap \bar{V}_\pi| &= \sum_{i=1}^l m_{i1} \left(\sum_{i' > 1} m_{i'2} \right) \\
 |V_P \cap \bar{V}_\pi| = |\bar{V}_P \cap V_\pi| &= \sum_{i=1}^l m_{i1} \left(\sum_{i' < i} m_{i'2} \right)
 \end{aligned}$$

Croisement d'une préordonnance avec 2 partitions.

$\pi_1 = (1, 2, 3, 4) (5, 6, 7, 8, 9)$

$\pi_2 = (1, 2, 3, 4) (5, 6, 7) (8, 9)$

Paires	Réunies	Séparées
12, 14, 23, 34 56, 57	6	0
13, 15, 24, 67 68, 79, 89	6	1
69, 78	2	0
16, 17, 25, 45 58, 59	2	4
27, 35, 46	0	3
26, 47	0	2
18, 19, 36, 37	0	4
29, 48	0	2
28, 49	0	2
38, 39	0	2

16

20

6

7

2

6

3

2

4

2

2

2

Réunies	Séparées
6	0
4	3
0	2
0	6
0	3
0	2
0	4
0	2
0	2
0	2

10

26

Décomposition canonique.

		π_1						π_2			
		$\delta p < \delta q$	$\delta p = \delta q$	$\delta p > \delta q$				$\delta p < \delta q$	$\delta p = \delta q$	$\delta p > \delta q$	
P	$dp < dq$	302	259	4	565	$dp < dq$	248	317	0	565	
	$dp = dq$	14	138	14	166	$dp = dq$	12	142	12	166	
	$dp > dq$	4	259	302	565	$dp > dq$	0	317	248	565	
		320	656	320	1 296			260	776	260	1 296

Les deux tableaux précédents et la figure 3 illustrent l'application de ces calculs aux données de la figure 1 et 2.

Les effectifs des cases du tableau de la décomposition canonique du croisement de 2 préordres peuvent également être obtenus en appliquant le codage de Kendall :

Le codage a_{pq} de la paire (p, q) , notée pq , selon le préordre P , est donné par :

$$a_{pq} = \text{sign}(d_p - d_q) = \begin{cases} 1 & \text{si } d_p < d_q \\ 0 & \text{si } d_p = d_q \\ -1 & \text{si } d_p > d_q \end{cases}$$

De même, le codage b_{pq} de la paire (p, q) , selon le préordre Π , est donné par :

$$b_{pq} = \text{sign}(\delta_p - \delta_q) = \begin{cases} 1 & \text{si } \delta_p < \delta_q \\ 0 & \text{si } \delta_p = \delta_q \\ -1 & \text{si } \delta_p > \delta_q \end{cases}$$

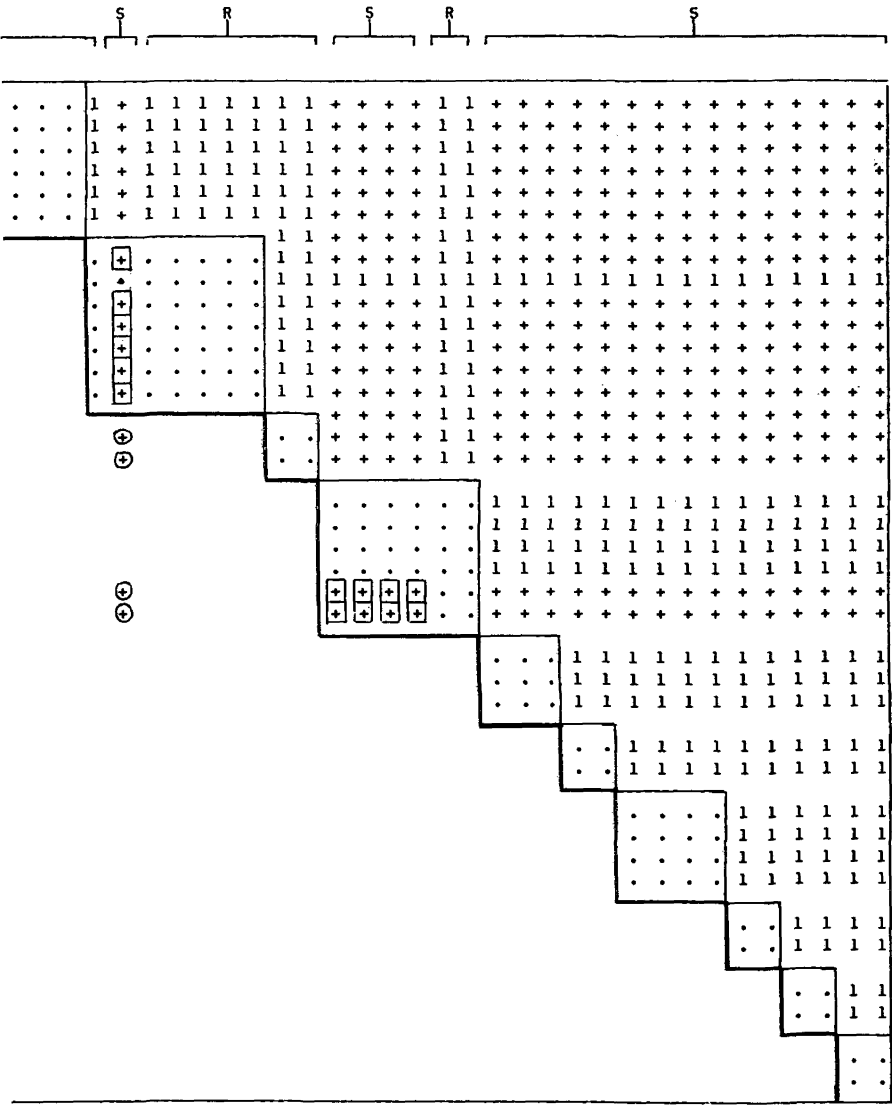
On obtient alors :

$$\begin{aligned} |V_P| &= \sum_{p < q} a_{pq}^2 & |V_\Pi| &= \sum_{p < q} b_{pq}^2 \\ |I_P| &= m^2 - \sum_{p < q} a_{pq}^2 & |I_\Pi| &= m^2 - \sum_{p < q} b_{pq}^2 \\ |V_P \cap V_\Pi| - |V_P \cap \bar{V}_\Pi| &= \sum_{p < q} a_{pq} b_{pq} \end{aligned}$$

c) Critères d'adéquation d'une partition π à un préordre P

Il est clair que tout critère d'adéquation doit croître avec $|V_P \cap V_\pi|$ qui représente le nombre de couples de paires concordantes dans les deux préordres et décroître avec $|V_P \cap \bar{V}_\pi|$ qui représente le nombre de paires discor-

$$\pi = (1, 2, 3, 4)(5, 6, 7, 8, 9)$$



$e : \cdot$ ou $\boxplus I_p$ $++ V_p \cap V_n$ $\boxplus \boxplus I_p \cap V_n$
1 ou $+ V_p$ $\oplus \oplus \bar{V}_p \cap V_n$

Figure 3. — L'exemple de 9 villes
Représentation des ensembles $I_p, V_p, V_p \cap V_n, V_p \cap V_n, I_p \cap V_n$.

dantes c'est-à-dire les paires dont l'ordre est inversé d'un préordre à l'autre.

La pondération des paires équivalentes comporte une bonne part d'arbitraire, selon qu'elles seront incluses ou exclues, on obtiendra des critères larges ou des critères stricts.

On distinguera également les critères purement ensemblistes des critères bornés qui sont indépendants du nombre de paires et donc du nombre d'objets à classer.

CRITÈRES ENSEMBLISTES

1. Nombre de paires concordantes

(1a) au sens large : $|W_P \cap W_\pi|$

(1b) au sens strict : $|V_P \cap V_\pi|$

Ces critères proposés par Benzecri (1965) présentent l'inconvénient de ne pas tenir compte du nombre de paires incompatibles. Ainsi, ils classeront *ex aequo* deux partitions π_1 et π_2 ayant le même nombre de paires compatibles même si leur nombre de paires incompatibles est différent.

2. Différence symétrique

L'adéquation sera d'autant meilleure que la distance entre les deux préordres, mesurée par leur différence symétrique sera plus faible. Le complément, dans l'ensemble du couple des paires, de la différence symétrique sera donc une mesure d'adéquation.

(2a) au sens large : $|M \times M - W_P \Delta W_\pi|$

(2b) au sens strict : $|M \times M - V_P \Delta V_\pi|$

Les deux critères ci-dessus peuvent se mettre sous la forme :

$$\text{— au sens large : } |M|^2 - |W_P| + 2 \left(|W_P \cap W_\pi| - \frac{1}{2} |W_\pi| \right)$$

$$\text{— au sens strict : } |M|^2 - |V_P| + 2 \left(|V_P \cap V_\pi| - \frac{1}{2} |V_\pi| \right)$$

On démontre dans l'annexe 1 que ces deux critères sont identiques et qu'ils sont de plus égaux à la somme des critères 1a et 1b.

Le nombre total de paires et le préordre P ne variant pas, on peut se restreindre à la partie entre parenthèses, et l'on obtient ainsi les deux critères de De la Vega.

$$\text{— au sens large : } |W_P \cap W_\pi| - \frac{1}{2} |W_\pi| = |W_P \cap W_\pi| - (m^2 + r^2 + s^2)/4$$

— au sens strict : $|W_P \cap V_\pi| - \frac{1}{2} |V_\pi| = |V_P \cap V_\pi| - r \times s/2$

S. Chah (1983) a récemment donné une linéarisation des critères de De la Vega permettant l'application des techniques de programmation linéaire à la recherche des partitions optimales selon ces critères.

Lerman (1969, 1981) propose de normaliser les critères de De la Vega :

— au sens large et à une constante multiplicative près,

$$\left(|W_P \cap W_\pi| - \frac{1}{2} |W_\pi| \right) \sqrt{|W_\pi|}$$

— au sens strict et à une constante multiplicative près,

$$\left(|W_P \cap V_\pi| - \frac{1}{2} |V_\pi| \right) \sqrt{|V_\pi|}$$

Sous l'hypothèse d'équiprobabilité de l'ensemble des partitions de même type, les critères de Lerman tendent vers une distribution normale centrée réduite.

3. Nombre de paires concordantes moins nombre de paires discordantes

(3a) Au sens large : $|(W_P \cap W_\pi)| + |(W_P^c \cap W_\pi^c)| - |W_P \cap W_\pi^c| - |W_\pi \cap W_P^c|$. Ceci peut s'écrire également :

$$|W_P \cap W_\pi| + |\bar{V}_P \cap \bar{V}_\pi| - |W_P \Delta W_\pi|$$

(3b) au sens strict $|V_P \cap V_\pi| - |V_P \cap \bar{V}_\pi|$.

La version stricte du critère présente l'inconvénient suivant : elle classe ex aequo la partition π_1 , possédant $|V_{\pi_1} \cap V_P|$ paires concordantes et aucune paire discordante avec la partition π_2 telle que :

$$|V_{\pi_2} \cap V_P| - |\bar{V}_{\pi_2} \cap V_P| = |V_{\pi_1} \cap V_P|$$

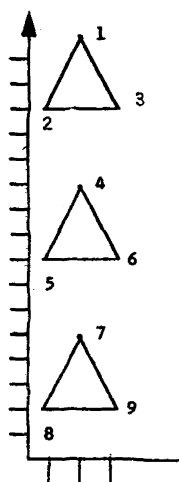
On démontre dans l'annexe 1 que le critère au sens large est égal au double du critère au sens strict plus $|I_P \cap I_\pi|$.

Par ailleurs, I. C. Lerman ⁽⁶⁾, a établi la relation :

$$(W_P \Delta W_\pi) \Delta (V_P \Delta V_\pi) = I_P \Delta I_\pi.$$

On notera que la plupart des critères précédents peuvent s'écrire sous la forme $\sum_{p < q} a_{pq} b_{pq}$ par le choix d'un codage adéquat des a_{pq} et b_{pq} , comme le montre la figure 4.

⁽⁶⁾ Communication personnelle. Nous tenons à remercier le professeur Lerman pour ses nombreuses remarques et suggestions qui ont beaucoup contribué à la version finale de cet article.



PRÉORDONNANCE

$$P : \begin{cases} 23 = 56 = 89 < 12 = 13 = 45 = 46 = 78 = 79 < 24 = 34 = 57 = 67 \\ < 14 = 25 = 36 = 47 = 58 = 69 < 35 = 26 = 68 = 59 < 15 = 16 = 48 = 49 \\ < 27 = 37 < 17 = 28 = 39 < 38 = 29 < 18 = 19 \end{cases}$$

PARTITION NATURELLE (i. e. parfaitement compatible)

$$1 : \begin{cases} \pi_1 = (1, 2, 3)(4, 5, 6)(7, 8, 9) \\ 12 = 13 = 23 = 45 = 46 = 56 = 78 = 79 = 89 \\ < 14 = 15 = 16 = 17 = 18 = 19 = 24 = 25 = 26 = 27 = 28 = 29 = 34 = 35 \\ = 36 = 37 = 38 = 39 = 47 = 48 = 49 = 57 = 58 = 59 = 67 = 68 = 69 \end{cases}$$

PARTITION NON NATURELLE

$$2 : \begin{cases} \pi_2 = (1, 2, 3, 4, 5, 6)(7, 8, 9) \\ 12 = 13 = 14 = 15 = 16 = 23 = 24 = 25 = 26 = 34 = 35 = 36 = 45 = 46 \\ = 56 = 78 = 79 = 89 \\ < 17 = 18 = 19 = 27 = 28 = 29 = 37 = 38 = 39 = 47 = 48 = 49 = 57 = 58 \\ = 59 = 67 = 68 = 69 \end{cases}$$

Décompositions canoniques

$P \backslash \pi_1$	$\delta p < \delta q$	$\delta p = \delta q$	$\delta p > \delta q$	
$dp < dq$	243	330	0	573
$dp = dq$	0	150	0	150
$dp > dq$	0	330	243	573
	243	810	243	1 296

Croisement de P avec π_1

$P \backslash \pi_2$	$\delta p < \delta q$	$\delta p = \delta q$	$\delta p > \delta q$	
$dp < dq$	273	270	30	573
$dp = dq$	21	108	21	150
$dp > dq$	30	270	273	573
	324	648	324	1 296

Croisement de P avec π_2

1) Nombre de paires concordantes

a-strictes

	1	0	0	
1				
0				
0				
243				273

b-larges

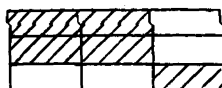
	1	1	0	
1				
1				
0				
723				672

Figure 4. — Partition naturelle et critères d'adéquation.

II) Complément de la différence symétrique

strictes et larges

966



945

121.5
196.5
7.79
6.06

DE LA VEGA stricte
large
LERMAN stricte
large

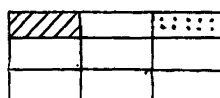
111
186
6.17
5.97

(à une constante multiplicative près)

III) Paires concordantes moins paires discordantes

a-strictes

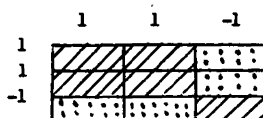
243



243

b-larges

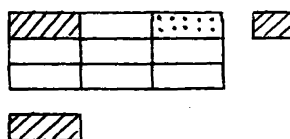
636



594

IV) Tau b de Kendall

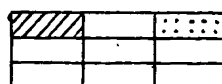
$$\frac{243 - 0}{\sqrt{243 \times 573}} = 0.651$$



$$\frac{273 - 30}{\sqrt{324 \times 573}} = 0.564$$

V) Gamma de Goodman, Kruskal

$$\frac{243 - 0}{243 + 0} = 1.000$$



$$\frac{273 - 30}{273 + 30} = 0.802$$

Tous les critères détectent la partition naturelle π_1 .

Légende  Partie de $M \times M$ contribuant positivement au critère.
 Partie de $M \times M$ contribuant négativement au critère.

Figure 4. — Partition naturelle et critères d'adéquation (suite).

CRITÈRES BORNÉS

1. Pourcentage de paires concordantes au sens strict

$$2 |V_P \cap V_\pi| / |M|^2$$

2. Tau de Kendall ajusté pour les paires équivalentes

$$\tau_b = \frac{|V_P \cap V_\pi| - |V_P \cap \bar{V}_\pi|}{[|V_P| \times |V_\pi|]^{1/2}} = \frac{\sum_{p < q} a_{pq} b_{pq}}{\left[\sum_{p < q} a_{pq}^2 \sum_{p < q} b_{pq}^2 \right]^{1/2}}$$

3. Gamma de Goodman, Kruskal.

$$\gamma = \frac{|V_P \cap V_\pi| - |V_P \cap \bar{V}_\pi|}{|V_P \cap V_\pi| + |V_P \cap \bar{V}_\pi|} = \frac{2 \sum_{p < q} a_{pq} b_{pq}}{m^2 - \sum_i m_i^2 - \sum_j m_j^2 + \sum_i \sum_j m_{ij}^2}$$

Nous ne retiendrons pas ici le critère de corrélation des rangs de Spearman que l'on peut considérer comme une approximation linéaire du τ de Kendall (cf. Guilbaud 1972).

On peut donner une interprétation statistique à ces 3 critères.

. Le premier donne une estimation de la probabilité d'obtenir deux paires concordantes en échantillonnant au hasard l'ensemble des paires. Ce critère peut être considéré comme étant une extension du critère de Rand (1971), qui est fréquemment utilisé pour mesurer l'adéquation de deux partitions. Il présente cependant le grave inconvénient de ne pas tenir compte des paires incompatibles.

. Le tau b de Kendall (1948) est une mesure couramment employée de la corrélation des rangs en présence d'ex aequo, cependant Lerman (1981), pages 142 et 143, démontre qu'il devient biaisé lorsque l'on passe de la comparaison d'ordre totaux et stricts à la comparaison de préordre totaux.

. Le gamma de Goodman, Kruskal (1954, 1959, 1963) donne la probabilité que deux paires, tirées au hasard parmi les paires non équivalentes, soient concordantes moins la probabilité qu'elles soient discordantes. Son erreur type est inférieure à celle de τ_b et on vérifie toujours que $|\gamma| > |\tau_b|$. De plus, $\gamma = 1$ lorsqu'il n'y a pas de paires discordantes, ce qui n'est pas vérifié pour τ_b . Dans ce cas particulier, la partition sera dite parfaitement compatible avec le préordre P puisqu'il n'existe pas d'objets réunis séparés par une distance supérieure à la plus petite distance entre deux objets séparés :

$$\pi \text{ est parfaitement compatible avec } P \text{ si } \max_{p \in R} dp \leq \min_{q \in S} dq$$

Cette propriété, conforme à la définition de « partition naturelle », adoptée par Cormack (1971), Everitt (1974), Rao (1971) et Vinod (1969), ainsi que les résultats de plusieurs études de simulation conduites par Baker et Hubert (1975) et Milligan (1981, 1983) militent en faveur de l'utilisation de γ comme critère d'adéquation.

En effet, comme le montre le tableau de la figure 5, seul γ croît avec le nombre de couples de paires concordantes, décroît avec le nombre de couples de paires discordantes, est indépendant du nombre de paires équivalentes et atteint sa valeur maximum uniquement pour les « partitions naturelles ». Cependant, γ donne un poids égal aux couples de paires concordantes et discordantes ce qui est relativement arbitraire.

	$ V_P \cap V_\pi $	$ W_P \cap W_\pi $	$ M \times M - W_P \Delta W_\pi $	$ V_P \cap V_\pi - V_P \cap \bar{V}_\pi $	τ_b	γ
1. Croît avec le nombre de couples de paires concordantes	OUI	OUI	OUI	OUI	OUI	OUI
2. Décroît avec le nombre de couples de paires discordantes	NON	OUI	OUI	OUI	OUI	OUI
3. Est indépendant du nombre de paires équivalentes dans la partition	OUI	NON	NON	OUI	NON	OUI
4. Atteint sa valeur maximum uniquement pour les « partitions naturelles » (parfaitement compatibles)	NON	NON	NON	NON	NON	OUI

Figure 5. — Evaluation de 6 critères d'adéquation.

Nous proposons d'évaluer l'adéquation d'une partition par une approche bicritère, dans l'esprit des travaux de Hansen, Delattre (1980), consistant à maximiser le nombre de couples de paires concordantes sous contrainte d'un nombre maximum de couples de paires discordantes à ne pas dépasser.

d) Partitions efficaces

Posons $a = |V_P \cap V_\pi|$ et $b = |V_P \cap \bar{V}_\pi|$.

On appelle partition efficace de niveau b , les partitions π_b^* qui maximisent a sans présenter plus de b couples de paires discordantes.

$$\pi_b^* : a(\pi_b) = \max_{\pi_d} \{ a(\pi_d) / d \leq b \}$$

L'application de ce concept aux données de la figure 1 donne les résultats suivants.

Partition efficace Critère	π_0^* (1, 2, 3, 4) (5, 6, 7) (8, 9)	π_4^* (1, 2, 3, 4) (5, 6, 7, 8, 9)
$ W_P \cap W_\pi $	719	713
$a = V_P \cap V_\pi $	248	302
$b = \bar{V}_\pi \cap V_P $	0	4
$a - b$	248	296
De la Vega		
stricte	118	142
large	201	225
Lerman		
stricte (à une constante multipli- cative près)	7.32	7.94
large	6.24	7.20
$ M \times M - W_P \Delta W_\pi $ et $ M \times M - V_P \Delta V_\pi $	967	1 015
τ_b	0.647	0.701
$\gamma = \frac{a-b}{a+b}$	1.000	0.974

Figure 6. — Comparaison des critères d'adéquation pour deux partitions efficaces sur les données des 9 villes.

L'utilisateur est le seul à pouvoir décider du meilleur compromis entre une partition efficace de niveau zéro, c'est-à-dire parfaitement compatible avec la préordonnance mais pouvant comporter un nombre de classes très élevé, et une partition efficace comportant un petit nombre d'incompatibilités en échange d'un nombre de classes restreint.

Lorsque les partitions efficaces forment une hiérarchie, ainsi que pour l'exemple de la figure 1, c'est une information supplémentaire importante qui serait perdue si l'on optait pour l'un quelconque des critères d'adéquation.

III. RECHERCHE DES PARTITIONS EFFICIENTES DE NIVEAU ZÉRO

Nous nous intéressons maintenant au problème de la détermination des partitions efficaces du niveau zéro ⁽⁷⁾. Le problème général de la détermination d'une partition efficace d'un niveau quelconque b , est un problème combinatoire non-polynomial. Deux approches de résolution peuvent être envisagées : une approche basée sur le principe d'énumération par séparation et évaluation et une approche par l'énumération implicite ⁽⁸⁾.

Algorithme pour déterminer la partition π_0^ :*

Dans l'exposé qui suit, nous supposons que les paires d'objets sont numérotées tel que :

$$d_1 \leq d_2 \dots \leq d_m$$

Nous présentons d'abord la notation que nous utiliserons, ensuite nous exposerons les définitions relatives à notre algorithme et, finalement, nous donnerons les détails de fonctionnement de celui-ci.

Notations : π_0 une partition parfaitement compatible ;
 E_0 l'ensemble de toutes les partitions parfaitement compatibles ;
 $a(\pi_0)$ le nombre de paires compatible de π_0 ;
 π_0^* une partition efficace de niveau zéro.

DÉFINITION. — (1) Soit Y_p^π la variable indicatrice, selon la partition π , de la réunion des objets i et j de la paire p . Une partition π de X est alors définie par le vecteur Y^π de m éléments tels que :

$$Y_p^\pi = \begin{cases} 1 & \text{si } p \in R \\ 0 & \text{si } p \in S \end{cases}$$

Dans ce cas, la propriété de transitivité suivante est vérifiée :

soit i, j les objets de la paire p ;
 soit j, k les objets de la paire q ; et
 soit i, k les objets de la paire z

alors $Y_p^\pi = 1$ et $Y_q^\pi = 1$ impliquent $Y_z^\pi = 1$.

(2) Soit Z_p le rang de la paire p dans la préordonnance P :

$$Z_p = |\{ q \in M / d_q < d_p \}|$$

⁽⁷⁾ La partition π_0^* est unique si la préordonnance P est une ordonnance.

⁽⁸⁾ Les algorithmes d'énumération implicite testés par les auteurs s'avèrent inefficients.

A chaque rang Z correspond une partition G de M à deux classes R_g et S_g telles que

$$R_g = (p/Z_p < Z) \quad \text{et} \quad S_g = (p/Z_p > Z)$$

La variable
$$Y_P^G = \begin{cases} 1 & \text{si } p \in R_g \\ 0 & \text{si } p \in S_g \end{cases}$$

(3) Soit G_i et G_{i+1} deux partitions de rangs successifs de M , e un sous-ensemble quelconque de $R_{g_{i+1}} - R_{g_i}$.

Chaque partition dont $RgUe$ est la classe des réunis sera appelée partition intermédiaire entre G_i et G_{i+1} .

ALGORITHME

Phase 1 : On classe les paires par ordre de distance croissante.

Phase 1 : On construit une suite $G_1, G_2, \dots, G_i, \dots, G_t$ de partitions de M de rang croissant. Pour chaque partition G_i on détermine si G_i est également partition de X . Si G_i n'est pas une partition de X on cherche la partition $\pi(G_i)$ de X déduite de G_i par « fermeture transitive ». C'est-à-dire, on ajoute pour tout couple p et $q \in R_g$ les paires $z \in S_g$ nécessaires à la vérification de la propriété de transitivité.

- On détermine si la partition $\pi(G_i)$ est parfaitement compatible avec la préordonnance P .
- On retient la partition $\pi(G_i)$ qui maximise a le nombre de couples de paires de $\pi(G_i)$ compatible avec P .

Phase 2 : On cherche à améliorer la partition obtenue à l'issue de la Phase 1 en explorant systématiquement toutes les partitions intermédiaires entre G_i et G_{i+1} , en vue de déterminer s'il existe des paires pouvant être réunies sans créer de contradiction avec P .

PROPOSITION : L'algorithme ci-dessus énumère toutes les partitions efficaces π_0^* .

Démonstration : Il s'agit de montrer que la partition $\pi(G_i)$ qui maximise a est telle qu'il n'existe pas de paires p compatibles avec P qui ne soient pas contenues dans les partitions intermédiaires entre G_i et G_{i+1} .

Dans la phase 1 on obtient la partition $\pi(G_i)$ telle que :

$$a[\pi(G_i)] = \max \{ a[\pi(G_1)], \dots, a[\pi(G_m)] \}$$

(1) Une réunion $Y_p = 1$ ne peut se produire que de 2 façons différentes :

A. Soit comme conséquence transitive de réunions effectuées à l'étape $i-1$.

B. Soit comme formation d'un groupe isolé.

(2) Seul l'ensemble $(Rg_{i+1} - Rg_i)$ doit être exploré car (étant donné que Rg_{i+1} contient au moins un zéro) une réunion dans Rg_{i+2} entraîne nécessairement au moins une contradiction.

(3) L'ensemble $Rg_{i+1} - Rg_i$ contient nécessairement au moins un zéro car sinon il est possible d'augmenter A dans la phase 2. CQFD

L'annexe 2 présente un code en langage BASIC de l'algorithme ci-dessus.

ANNEXE 1

PROPOSITION 1. — *Démonstration que $|M|^2 - |V_P \Delta V_\pi| = |M|^2 - |W_P \Delta W_\pi|$.
Il suffit de montrer que $|W_P \Delta W_\pi| = |V_P \Delta V_\pi|$.*

Démonstration. — $|V_P \Delta V_\pi| = |V_P| + |V_\pi| - 2|V_P \cap V_\pi|$.

On ajoute et on retranche $|I_P|$ et $|I_\pi|$.

$$\begin{aligned} |V_P \Delta V_\pi| &= |V_P| + |V_\pi| - 2|V_P \cap V_\pi| + |I_P| - |I_P| + |I_\pi| - |I_\pi| \\ |V_P \Delta V_\pi| &= |W_P| + |W_\pi| - 2|V_P \cap V_\pi| - |I_P| - |I_\pi| \end{aligned}$$

et comme $I_P \cap (M \times M) = I_P$ on peut écrire

$$|V_P \Delta V_\pi| = |W_P| + |W_\pi| - 2|V_P \cap V_\pi| - |I_P \cap M \times M| - |I_\pi \cap M \times M|$$

et comme $(M \times M) = (I_\pi \cup \bar{V}_\pi \cup V_\pi)$ on a

$$\begin{aligned} |V_P \Delta V_\pi| &= |W_P| + |W_\pi| - 2|V_P \cap V_\pi| - |I_P \cap (I_\pi \cup V_\pi \cup \bar{V}_\pi)| - |I_\pi \cap (I_P \cup V_P \cup \bar{V}_P)| \\ |V_P \Delta V_\pi| &= |W_P| + |W_\pi| - 2|V_P \cap V_\pi| - |I_P \cap I_\pi| - |I_P \cap V_\pi| - |I_P \cap \bar{V}_\pi| - |I_\pi \cap I_P| \\ &\quad - |I_\pi \cap V_P| - |I_\pi \cap \bar{V}_P| \end{aligned}$$

mais $|I_P \cap V_\pi| = |I_P \cap \bar{V}_\pi|$

et $|I_\pi \cap V_P| = |I_\pi \cap \bar{V}_P|$

donc

$$\begin{aligned} |V_P \Delta V_\pi| &= |W_P| + |W_\pi| - 2|V_P \cap V_\pi| - 2|I_P \cap I_\pi| - 2|I_P \cap V_\pi| - 2|I_\pi \cap V_P| \\ |V_P \Delta V_\pi| &= |W_P| + |W_\pi| - 2|W_P \cap W_\pi| \end{aligned}$$

et donc $|V_P \Delta V_\pi| = |W_P \Delta W_\pi|$

CQFD

PROPOSITION 2. — *Démontrons que le critère 2 est égal à la somme des critères 1a et 1b*

$$|M \times M| - |V_P \Delta V_\pi| = |W_P \cap W_\pi| + |V_P \cap V_\pi|$$

Démonstration :

$$|M \times M| - |V_P \Delta V_\pi| = |M \times M| - |V_P| - |V_\pi| + 2|V_P \cap V_\pi|$$

et comme $M \times M = W_P \cup \bar{V}_P = W_\pi \cup \bar{V}_\pi$ et que, de plus $(M \times M) \cap (M \times M) = M \times M$, il vient

$$= |(W_P \cup \bar{V}_P) \cap (W_\pi \cup \bar{V}_\pi)| - |V_P| - |V_\pi| + 2|V_P \cap V_\pi|$$

En développant le premier cardinal, nous obtenons :

$$= |W_P \cap W_\pi| + |\bar{V}_P \cap W_\pi| + |(W_P \cup \bar{V}_P) \cap V_\pi| - |V_P| - |V_\pi| + 2|V_P \cap V_\pi|$$

Le 3^e cardinal vaut $|\bar{V}_\pi| = |V_\pi|$ et s'annule avec l'avant-dernier cardinal

$$\begin{aligned} &= |W_P \cap W_\pi| + |\bar{V}_P \cap W_\pi| - |V_P| + |V_P \cap V_\pi| + |V_P \cap V_\pi| \\ &= |W_P \cap W_\pi| + |\bar{V}_P \cap W_\pi| - |V_P| + |\bar{V}_P \cap \bar{V}_\pi| + |V_P \cap V_\pi| \end{aligned}$$

En regroupant le 2^e cardinal et l'avant-dernier, on obtient :

$$= |W_P \cap W_\pi| + |\bar{V}_P \cap (W_\pi \cup \bar{V}_\pi)| - |V_P| + |V_P \cap V_\pi|$$

Le deuxième cardinal vaut $|\bar{V}_P| = |V_P|$ et s'annule avec le 3^e cardinal.

$$= |W_P \cap W_\pi| + |V_P \cap V_\pi| \quad \text{CQFD}$$

PROPOSITION 3. — *Démontrons que :*

$$\begin{aligned} &|W_P \cap W_\pi| + |W_P^c \cap W_\pi^c| - |W_P \cap W_\pi^c| - |W_\pi \cap W_P^c| \\ &= 2[|V_P \cap V_\pi| - |V_P \cap \bar{V}_\pi|] + |I_P \cap I_\pi| \end{aligned}$$

Démonstration. — Décomposons le premier cardinal utilisant $W_P = V_P \cup I_P$ et $W_\pi = V_\pi \cup I_\pi$. Remplaçons également W_P^c par $\bar{V}_P$

$$|(V_P \cup I_P) \cap (V_\pi \cup I_\pi)| + |\bar{V}_P \cap \bar{V}_\pi| - |(V_P \cup I_P) \cap \bar{V}_\pi| - |(V_\pi \cup I_P) \cap \bar{V}_P|$$

En développant, il vient :

$$\begin{aligned} &|V_P \cap V_\pi| + |V_P \cap I_\pi| + |I_P \cap V_\pi| + |I_P \cap I_\pi| + |\bar{V}_P \cap \bar{V}_\pi| \\ &\quad - |V_P \cap \bar{V}_\pi| - |I_P \cap \bar{V}_\pi| - |V_\pi \cap \bar{V}_P| - |I_\pi \cap \bar{V}_P| \end{aligned}$$

En utilisant les relations suivantes :

$$|I_P \cap V_\pi| = |I_P \cap \bar{V}_\pi| \quad \text{et} \quad |I_\pi \cap V_P| = |I_\pi \cap \bar{V}_P|$$

ainsi que

$$|V_P \cap V_\pi| = |\bar{V}_P \cap \bar{V}_\pi| \quad \text{et} \quad |V_P \cap \bar{V}_\pi| = |\bar{V}_P \cap V_\pi|$$

il vient

$$2[|V_P \cap V_\pi| - |V_P \cap \bar{V}_\pi|] + |I_P \cap I_\pi| \quad \text{CQFD}$$

PROPOSITION 4, due à I. C. Lerman.

$$(W_P \Delta W_\pi) \Delta (V_P \Delta V_\pi) = I_P \Delta I_\pi$$

Démonstration. — Développons la 1^{re} parenthèse.

$$W_P \Delta W_\pi = (W_P \cap W_\pi^c) \cup (W_P^c \cap W_\pi)$$

en utilisant $W_\pi^c = \bar{V}_\pi$ et $W_P^c = \bar{V}_P$, il vient :

$$W_P \Delta W_\pi = (W_P \cap \bar{V}_\pi) \cup (\bar{V}_P \cap W_\pi)$$

en utilisant $W_P = V_P \cup I_P$ et $W_\pi = V_\pi \cup I_\pi$, il vient :

$$W_P \Delta W_\pi = (V_P \cap \bar{V}_\pi) \cup (I_P \cap \bar{V}_\pi) \cup (\bar{V}_P \cap V_\pi) \cup (\bar{V}_P \cap I_\pi)$$

Développons la 2^e parenthèse :

$$V_P \Delta V_\pi = (V_P \cap V_\pi^c) \cup (V_P^c \cap V_\pi)$$

en utilisant $V_\pi^c = \bar{V}_P \cap I_\pi$ et $V_P^c = V_P \cap I_P$, il vient :

$$V_P \Delta V_\pi = [V_P \cup (\bar{V}_\pi \cap I_\pi)] \cup [(\bar{V}_P \cap I_P) \cap V_\pi]$$

$$V_P \Delta V_\pi = (V_P \cap \bar{V}_\pi) \cup (V_P \cap I_\pi) \cup (\bar{V}_P \cap V_\pi) \cup (I_P \cap V_\pi)$$

Nous pouvons maintenant calculer la différence symétrique entre les deux premières parenthèses.

$$(W_P \Delta W_\pi) \Delta (V_P \Delta V_\pi) = (I_P \cap \bar{V}_\pi) \cup (V_P \cap I_\pi) \cup (\bar{V}_P \cap I_\pi) \cup (I_P \cap V_\pi)$$

Les parenthèses 1 et 4 du membre de gauche peuvent être factorisées comme suit :

$$(I_P \cap \bar{V}_\pi) \cup (I_P \cap V_\pi) = I_P \cap (\bar{V}_\pi \cup V_\pi) = I_P \cap I_\pi^c$$

De même, les parenthèses 2 et 3 du membre de gauche deviennent :

$$(V_P \cap I_\pi) \cup (\bar{V}_P \cap I_\pi) = (V_P \cup \bar{V}_P) \cap I_\pi = I_P^c \cap I_\pi$$

Donc :

$$(W_P \Delta W_\pi) \Delta (V_P \Delta V_\pi) = (I_P \cap I_\pi^c) \cup (I_P^c \cap I_\pi) = I_P \Delta I_\pi \quad \text{CQFD}$$

ANNEXE 2

```

10 READN:M=N*(N-1)/2:DIMI(M+1),J(M+1),D(M+1),Y(M),X(M),U(N):P=INT(M/2):D=P:F=P
20 FORS=0TOP-1:D(S)=-1E30:NEXTS:FORS=PTOM+1:D(S)=1E30:NEXTS
60 READY:IFY=1THEN80
70 READC:DIMC(N,C):FORI=1TON:FORS=1TOC:READC(I,J):NEXTJ,I
80 FORJ=2TON:FORS=1TOJ-1:IFY=1READX:GOTO90
95 X=0:FORS=1TOC:X=X+(C(I,S)-C(J,S))*(C(I,S)-C(J,S)):NEXTS
90 K=D-1:IFX>D(P)LETK=P
100 IFX<D(K+1)THEN200ELSEK=K+1:GOTO100

```



```

200 IFD=1GOTO220
210 IFF=MOK-D<F-KGOTO300
220 IFK=FGOTO240
230 FORS=FTOK+1STEP-1:D(S+1)=D(S):I(S+1)=I(S):J(S+1)=J(S):NEXTS
240 IFF<MLETF=F+1
250 K=K+1:GOTO400
300 IFK<DGOTO350
310 FORS=DTOK:D(S-1)=D(S):I(S-1)=I(S):J(S-1)=J(S):NEXTS
350 IFD>1LETD=D-1
400 P=INT((D+F)/2):D(K)=X:I(K)=I:J(K)=J:NEXTI,J
490 LPRINT" I  J", "DIST", "CONTRIBUTION":FORS=1T048:LPRINT"-":NEXTS:LPRINT
495 FORS=1TOM:Y(S)=M-S+1:IFD(S)=D(S-1)LETY(S)=Y(S-1)
500 LPRINTI(S):J(S):D(S):Y(S):NEXTS:FORS=1T048:LPRINT"-":NEXTS:LPRINT
510 FORI=1TOM:U(I)=I:NEXTI:FORS=1TOM:D(S)=Y(S):Y(S)=0:NEXTS
600 LPRINT:LPRINT"PARTITIONS EFFICIENTES DE RANG 0":LPRINTSTRING$(32,61)
610 D=1:K=1:S=K:Z=0
620 S=S+1:IFD(S)=D(D)GOTO620
640 K=S-1:FORS=DTOK:Y(S)=1:NEXTS:IFK*(M-K)<K1*(M-K1)THEN800
650 IFZ=0THENK1=K:D1=D
660 GOSUB1700:GOSUB1740:IFZ=2GOTO800
680 D=K+1:S=K+1:GOTO620
800 X=1:K=K1:D=D1:FORS=K+1TOM:Y(S)=0:NEXTS
805 IFK>INT((M+1)/2)THENX=K-INT((M+1)/2)
810 Y=X-1:P=K-D+1:T=P
820 Y=Y+1:IFY>PSTOP
830 FORJ=DTOK:Y(J)=1:NEXTJ
840 FORJ=1TOY:X(J)=J:Y(D-1+X(J))=0:NEXTJ:GOSUB1700:GOSUB1840:IFZ=0GOSUB1500:P=Y
850 FORI=YTO1STEP-1:IFX(I)<T-Y+IGOTO870
860 NEXTI:GOTO820
870 FORJ=DTOK:Y(J)=1:NEXTJ
880 FORJ=YTO1STEP-1:X(J)=X(I)+J-I+1:Y(D-1+X(J))=0:NEXTJ
890 FORJ=1TOI:Y(D-1+X(J))=0:NEXTJ:GOSUB1700:GOSUB1840:IFZ=0GOSUB1500:P=Y
900 GOTO850
1500 A=0:FORS=1TOK:IFY(S)>0THENA=A+1
1501 :NEXTS
1503 LPRINT:A=A*(M-A):LPRINT"A=":A:LPRINT:W=0:I=1:LPRINT"GROUPES:"
1510 IFU(I)>0:LETU=U(I):U(I)=0:GOTO1530
1520 I=I+1:IFI>NLPRINT:RETURNELSEGOTO1510
1530 W=W+1:LPRINT"GROUPE":W:" COMPOSE DES INDIVIDUS ":I:IFI=NLPRINT:RETURN
1540 FORJ=I+1TOM:IFU(J)=ULPRINT","J:U(J)=0
1550 NEXTJ:LPRINT:GOTO1520
1700 FORI=1TOM:U(I)=I:NEXTI

```

```

1705 Z=0:FORS=1TOK:IFY(S)=0ORU(I(S))=U(J(S))GOTO1720
1710 Z=Z+1:IFU(I(S))<U(J(S))THENU(J(S))=U(I(S))ELSEU(I(S))=U(J(S))
1720 NEXTS:IFZ>0GOTO1705ELSEReturn
1740 S=K
1750 S=S+1:IFS>MTHENZ=2:Return
1760 IFU(I(S))=U(J(S))GOTO1750
1770 T=S
1780 S=S+1:IFS>MReturn
1790 IFU(I(S))=U(J(S))ANDD(S)<D(T)GOTO1800ELSEGOTO1780
1800 Z=1:Return
1840 FORS=K+1TOM:IFU(I(S))=U(J(S))GOTO1860
1850 NEXTS:Return
1860 Z=1:Return

```

BIBLIOGRAPHIE

- F. B. BAKER, L. J. HUBERT (1972), Measuring the power of hierarchical cluster analysis, *Journal of the American Statistical Association*, vol. 70, 1972, p. 31-38.
- G. H. BALL, D. J. HALL (1965 a), Isodata, an Iterative Method of Multivariate Analysis and Pattern Classification. *Proceedings of the IFIPS Congress*, 1965.
- G. H. BALL, D. J. HALL (1965 b), Isodata a Novel Method of Data Analysis and Pattern Classification. *NATIS Report AD 699616*, 1965.
- J. P. BENZECRI (1965), Problèmes et Méthodes de la Taxinomie, *Cours ISUP*, Paris, 1965.
- J. P. BENZECRI (1973), *L'analyse des Données*, Tome 1 : La Taxinomie. Dunod, Paris, 1973.
- J. P. BENZECRI, G. FLAMEMBAUM, J. THIERY (1979), Agrégation en Boules de Rayons Fixes et Centres Optimisés, *Cahiers de l'Analyse des données*, vol. 14, n° 3, 1979, p. 357-364.
- S. CHAH (1983), *Optimisation linéaire en classification automatique*, Thèse de 3^e cycle, Université Pierre et Marie Curie, Paris.
- J. L. CHANDON, J. LEMAIRE (1977), Agrégation typologique de quasi-ordres, un nouvel algorithme. *1^{re} Journée Internationale, Analyse des données et Informatique*, IRIA, Versailles.
- J. L. CHANDON, S. PINSON (1981). *Analyse typologique : Théories et Applications*. Masson, Paris, 1981.
- R. M. CORMACK (1971), A Review of Classification *Journal of the Royal Statistical Society*, vol. 134, Series A, Part. 3, 1971, p. 321-327.
- W. F. DE LA VEGA (1967), Techniques de classification automatique utilisant un indice de ressemblance. *Revue Française de sociologie*, 1967.
- W. F. DE LA VEGA (1970) : Quelques aspects du problème de la détermination automatique des classifications. *Quality and Quantity*, vol. 4, n° 1, 1970, p. 117-152.
- W. F. DE LA VEGA (1975) : 3 Algorithmes de Classification Automatique. *Informatique et Sciences Humaines*, n° 25, 1975, p. 61-74.
- E. DIDAY (1971). Une nouvelle méthode en classification automatique et reconnaissance des formes. *Revue de Statistique Appliquée*, vol. 19, n° 2, 1971.

- E. DIDAY et collaborateurs (1979), *Optimisation en Classification Automatique*, INRIA, le Chesnay, 1979.
- B. EVERITT (1974), *Cluster Analysis*, Social Sciences Research Council, Heineman Educational Books, 1974.
- L. A. GOODMAN, W. H. KRUSKAL (1954, 1959, 1963), Measures of association for cross classifications, Part. I, II et III, *Journal of American Statistical Association*, vol. 49, p. 732, vol. 54, p. 123, vol. 58, p. 310.
- G. Th. GUILBAUD (1980), Relation entre les deux coefficients de corrélation des rangs, *Mathématique et Sciences Humaines*, vol. 18, Hiver 1980, p. 45-59.
- P. HANSEN, M. DELATTRE (1980), Bi-criterion Cluster Analysis, IEEE, *Transaction on Pattern Analysis and Machine Intelligence*, vol. PAMI 2, n° 4, July 1980.
- J. A. HARTIGNAN (1975), *Clustering Algorithms*, Wiley 1975.
- M. J. KENDALL (1948), *Rank correlation Methods*, Griffin London, 4^e édition 1971.
- J. LEMAIRE (1977), Agrégation typologique des préférences, *Mathématiques et Sciences Humaines*, vol. 58, p. 31-50.
- I. C. LERMAN (1969), On two criteria of classification, in *Numerical Taxonomy*, Cole A. J. Ed. p. 114-127, Academic Press, London, 1969.
- J. C. LERMAN, H. LEREDDE (1977), La méthode des pôles d'attraction, *Actes du Colloque « Analyse des Données et Informatique »*, I.R.I.A., Versailles, 1977.
- I. C. LERMAN (1981), *Classification et Analyse Ordinale des données*, Dunod, Paris, 1981.
- J. MACQUEEN (1967), Some Methods for Classification and Analysis of Multivariate Data. *Proceeding 5th Berkeley Symposium*, vol. 1, 1967, p. 281-297.
- G. W. MILLIGAN (1981), A monte carlo study of thirty internal criteria measures for cluster analysis, *Psychometrika*, vol. 46, p. 187-199.
- G. W. MILLIGAN, M. C. COOPER (1983), An examination of procedures for determining the number of clusters, *Working Paper series*, p. 83-51, College of Administrative Science, Ohio State University.
- R. MOJENA (1977), Hierarchical Grouping Methods and Stopping Rules. *Computer Journal*, vol. 20, 1977, p. 359-363.
- W. M. RAND (1971), Objective criteria for the evaluation of clustering methods, *Journal of the American Statistical Association*, vol. 66, 1971, p. 846-850.
- M. R. RAO (1971), Cluster Analysis and Mathematical Programming, *Journal of the American Statistical Association*, vol. 66, n° 335, p. 622-626.
- S. RÉGNIER, Sur quelques aspects mathématiques des problèmes de la classification automatique, *J. C. C. Bull.*, vol. 4, 1965, p. 175-191.
- G. S. SEBESTYEN (1962), *Décision Making Process in Pattern Recognition*. MacMillan, 1982.
- J. T. TOU, R. C. GONZALEZ (1974), *Pattern Recognition Principles*, Addison-Wesley, 1974.
- H. D. VINOD (1969), Integer Programming and Grouping. *Journal of the American Statistical Association*, p. 506-519, June 1969.
- D. WISHART (1969), Mode Analysis: a Generalization of Nearest Neighbour which Reduces Chaining Effects. In *Numerical Taxonomy*, Cole A. J., ed., p. 282-311, Academic Press, 1969.