

REVUE DE STATISTIQUE APPLIQUÉE

P. BERTRAND

E. DIDAY

Une généralisation des arbres hiérarchiques : les représentations pyramidales

Revue de statistique appliquée, tome 38, n° 3 (1990), p. 53-78

http://www.numdam.org/item?id=RSA_1990__38_3_53_0

© Société française de statistique, 1990, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

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

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

UNE GÉNÉRALISATION DES ARBRES HIÉRARCHIQUES : LES REPRÉSENTATIONS PYRAMIDALES *

P. BERTRAND et E. DIDAY

Université Paris-9, place du Maréchal de Lattre de Tassigny, 75775 Paris Cédex 16
INRIA, Domaine de Voluceau-Rocquencourt, BP 105, 78153 Le Chesnay Cédex

RÉSUMÉ

Les représentations pyramidales constituent une extension naturelle des arbres hiérarchiques. Comme les hiérarchies, les représentations pyramidales sont des ensembles de parties (appelées *classes* ou *paliers*) de l'ensemble des objets à classer. Cependant la classification pyramidale permet d'obtenir des relations plus complexes entre les classes. En particulier deux classes non disjointes ne sont pas nécessairement emboîtées, comme c'est le cas pour une hiérarchie. Une propriété intéressante des représentations pyramidales est leur capacité à produire un petit nombre d'ordres qui respectent les contraintes de proximité entre les objets à classer. Dans cet article, nous exposons une synthèse des principaux résultats théoriques et pratiques, récemment obtenus sur le sujet.

Mots-clés : Analyse des données, classification hiérarchique, compatibilité entre un ordre et un indice de dissimilarité, classes empiétantes, sériation.

SUMMARY

The pyramidal representations are extensions of the hierarchical trees. Similarly to the hierarchical structure, a pyramid is a collection of subsets of the set of objects to be clustered. However the pyramidal clustering can give rise to complex relations between clusters. In particular, contrary to the hierarchical case, two not disjointed clusters are not necessarily nested. An attractive property of the pyramidal representation is its ability to produce a few number of orders on the set of objects, which are compatible with the proximities between objects. This paper presents the main theoretical and practical advances recently obtained on the subject.

Key-words : Data analysis, hierarchical clustering, compatibility between an order and a dissimilarity, overlapping clusters, seriation.

Introduction

Les arbres hiérarchiques figurent parmi les plus anciens modèles utilisés en Taxonomie (voir Aristote et son arbre de vie). Plusieurs auteurs (par exemple

* Article reçu en avril 1988 et révisé en novembre 1989.

Benzecri (1973)), ont étudié la formalisation mathématique de ce type de modèle. Dans la pratique, la classification hiérarchique est fréquemment utilisée pour compléter ou résumer les résultats obtenus par d'autres méthodes de Classification Automatique : Analyse Factorielle, méthode des Nuées Dynamiques, ... En effet, les arbres hiérarchiques constituent un modèle souvent éloigné des données rencontrées dans les traitements usuels en analyse de données car le modèle hiérarchique impose aux données de vérifier la contrainte ultramétrique (i.e. tous les triplets de points doivent être des triangles isocèles). Aussi, plusieurs extensions des arbres hiérarchiques ont récemment été proposées :

- les B_k clusters de Jardine et Sibson (1971),
- les résultats de Hubert (1974) en théorie des graphes,
- les arbres dits « additifs » qui ont été étudiés par de nombreux auteurs depuis Buneman (1971); voir le livre récent de Barthélémy et Guénoche (1988).
- les pseudo-hiérarchies de Fichet (Durand et Fichet, 1988), notion proche de la représentation pyramidale.

La représentation pyramidale, encore appelée *pyramide*, proposée par Diday (1984), constitue également une généralisation du modèle hiérarchique. Comme les hiérarchies, les pyramides sont des ensembles de parties, appelées classes ou paliers, de l'ensemble des objets à classer. Cependant certaines représentations pyramidales constituent des structures plus « complètes » des données : en effet, deux paliers d'une pyramide peuvent avoir une intersection non vide, sans être inclus l'un dans l'autre. Autrement dit, contrairement au cas hiérarchique, le modèle pyramidal illustre l'idée intuitive que certains objets à classer, peuvent appartenir à deux classes qui ne sont pas emboîtées l'une dans l'autre (classes *empiétantes*). D'autre part, certaines représentations pyramidales sont des sériations, c'est à dire définissent un nombre restreint d'ordres sur l'ensemble des objets étudiés. Par conséquent, les représentations pyramidales forment un ensemble de classifications plus vaste que celui des arbres hiérarchiques, cet élargissement s'effectuant en proposant des classifications plus complexes, pouvant dans certains cas ordonner les objets entre eux. Mais les pyramides ne constituent pas seulement une extension mathématique des hiérarchies : elles présentent aussi l'avantage, du point de vue de l'Analyse des Données, d'introduire une nouvelle représentation graphique synthétique des classes et des objets, plus précise et plus fidèle aux données initiales, que la représentation hiérarchique.

Les pyramides étant une extension des hiérarchies, nous rappelons d'abord les propriétés des hiérarchies, puis nous indiquons la généralisation de ces propriétés au cas pyramidal. Après avoir comparé les représentations pyramidales aux pseudo-hiérarchies, nous terminons en donnant deux exemples de traitements de jeux tests de données.

1. Notations & Rappels

La Classification Automatique étudie un ensemble fini qui est l'ensemble des objets à classer (cet ensemble sera noté ici Ω). La finalité d'une méthode de Classification Automatique est de déterminer, par des algorithmes, un ensemble de parties de Ω caractérisant aussi fidèlement que possible l'information disponible

sur Ω . Les méthodes évoquées ici, utilisent toutes le même type de données : un indice de dissimilarité.

1.1 Indices de dissimilarité

Nous rappelons dans ce paragraphe, les principales notions relatives aux indices de dissimilarité.

Définition d'un indice de dissimilarité

On rappelle qu'un indice de dissimilarité \mathbf{d} défini sur Ω , est une application de Ω^2 dans \mathbf{R}^+ telle que :

$$(i) \text{ pour tout } x \text{ élément de } \Omega, \quad \mathbf{d}(x, x) = 0$$

$$(ii) \text{ pour tout } x, y \text{ élément de } \Omega, \quad \mathbf{d}(x, y) = \mathbf{d}(y, x)$$

Un indice de dissimilarité peut être donné a priori ou calculé à partir des données, par exemple si les données sont des coordonnées de points d'un espace euclidien.

Ultramétrie

Une ultramétrie \mathbf{u} , est un indice de dissimilarité défini sur Ω , qui vérifie l'inégalité suivante, appelée inégalité ultramétrique :

$$\text{pour tout } x, y, z \text{ élément de } \Omega, \quad \mathbf{u}(x, z) \leq \max\{\mathbf{u}(x, y), \mathbf{u}(y, z)\}$$

Indice d'agrégation

Au cours d'une classification hiérarchique ou pyramidale, on utilise souvent un indice d'agrégation afin de mesurer la proximité entre les classes déjà formées. On appelle indice d'agrégation entre groupes d'objets, une application μ de $P(\Omega) \times P(\Omega)$ (où $P(\Omega)$ l'ensemble des parties de Ω dans \mathbf{R}^+ telle que :

$$\text{pour tout } A, B \text{ élément de } P(\Omega), \quad \mu(A, B) = \mu(B, A) \text{ (symétrie)}$$

Dans le cas de la classification hiérarchique, les parties A et B sont par construction disjointes. La définition ci-dessus est donc une extension de la notion d'indice d'agrégation au cas de parties non disjointes. Désignant par d un indice de dissimilarité quelconque sur Ω , et par A et B deux parties quelconques de Ω , donnons la définition de quelques extensions au cadre pyramidal d'indices classiques en classification hiérarchique :

$$\mu(A, B) = \min\{d(x, y) | x \in A - B \text{ et } y \in B - A\} \text{ (extension du saut minimum)}$$

$$\mu(A, B) = \max\{d(x, y) | x \in A \text{ et } y \in B\} \text{ (extension du saut maximum)}$$

$$\mu(A, B) = \Sigma\{d^2(x, y) | x \in A \text{ et } y \in B\} \text{ (extension de l'indice d'inertie)}$$

Matrice d'un indice de dissimilarité

On suppose Ω ordonné par une relation d'ordre notée Δ . Les éléments de Ω peuvent être identifiés à leurs rangs selon l'ordre Δ . Ainsi on notera Ω sous la forme :

$$\Omega = \{1, 2, \dots, n\} \text{ où } n \text{ est le cardinal de } \Omega$$

Soit $m_{ij} = d(i, j)$ la valeur prise par un indice de dissimilarité d pour les deux éléments i et j de Ω . La matrice de terme général m_{ij} obtenue en ordonnant les valeurs des indices i et j (c'est à dire ses lignes et ses colonnes) selon l'ordre Δ , sera notée par la suite $\mathbf{M}(d, \Delta)$. Par construction, cette matrice est symétrique, à termes positifs et tous les termes de sa diagonale sont nuls.

1.2 Rappels sur les hiérarchies

1.2.1. Définitions

Par la suite, le symbole \subset désignera l'inclusion stricte.

Hiérarchie

Une hiérarchie totale \mathbf{H} sur Ω est un ensemble de parties de Ω vérifiant :

- (i) L'ensemble Ω et les singletons de Ω appartiennent à \mathbf{H} ,
- (ii) L'intersection de deux éléments h_1 et h_2 de \mathbf{H} est soit l'ensemble vide, soit l'un des deux éléments h_1 ou h_2 .

Les éléments de H sont appelés «paliers». Par abus de notation, on écrira «hiérarchie» pour «hiérarchie totale».

Prédécesseur, successeur

Etant donné une hiérarchie H , on dit qu'un palier s est successeur du palier p si :

- (i) $s \subset p$ (au sens strict)
- (ii) il n'existe pas de palier h différent de p et de s tel que l'on ait :

$$s \subset h \subset p$$

On dit aussi que p est *prédécesseur* du palier s .

Dans la pratique, on indice les hiérarchies, c'est à dire qu'à chaque palier de la hiérarchie on associe un réel positif qui mesure le degré d'agrégation des objets de ce palier. On dit alors que la hiérarchie est *indicée*. Plus précisément :

Hiérarchie indicée

Une hiérarchie indicée sur Ω est un couple (H, f) où H est une hiérarchie sur Ω et f une application définie de H dans \mathbb{R}^+ telle que :

- (i) pour tout x élément de Ω , $f(\{x\}) = 0$
- (ii) pour tout h_1 et h_2 élément de H , on a : $h_1 \subset h_2 \Rightarrow f(h_1) < f(h_2)$

Par exemple, à tout palier d'une hiérarchie on peut associer, pour mesurer le degré d'agrégation de ses objets, la plus grande des dissimilarités séparant deux objets de ce palier.

Indice de dissimilarité induit

Soit (H, f) une hiérarchie indicée sur Ω . On appelle indice de dissimilarité

induit par H , et on note d_H , l'application de Ω^2 dans R^+ définie par :

$$d_H(x, y) = \min\{f(h) | h \in H, x \in h, y \in h\}$$

Autrement dit $d_H(x, y)$ est la hauteur du plus « bas » palier contenant les objets x et y . On vérifie facilement que d_H est une ultramétrique sur Ω .

Représentation visuelle (« dendrogrammes »)

Remarquons d'abord qu'à toute hiérarchie, on peut associer le graphe dont les sommets sont les paliers et dont les arcs (p, q) présents indiquent que p est successeur de q . En fait ce graphe est un arbre et pour représenter une hiérarchie indiquée, on tient compte de la fonction d'indilage f en symbolisant chaque palier (sommets du graphe) par un trait horizontal dont la hauteur est proportionnelle à la quantité positive $f(h)$.

Donnons un exemple :

$$\Omega = \{1, 2, 3, 4, 5\}$$

$$H = \{\Omega, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, h_1, h_2, h_3\}$$

avec :

$$\begin{array}{lll} h_1 = \{1, 2\} & h_2 = \{1, 2, 3\} & h_3 = \{4, 5\} \\ \text{et } f(h_1) = 0.5 & f(h_2) = 1.5 & f(h_3) = 1 \end{array} \quad f(\Omega) = 2.5$$

(H, f) est visualisée par l'arbre hiérarchique ci dessous où nous remarquons, par exemple, que les successeurs de h_2 sont h_1 et $\{3\}$.

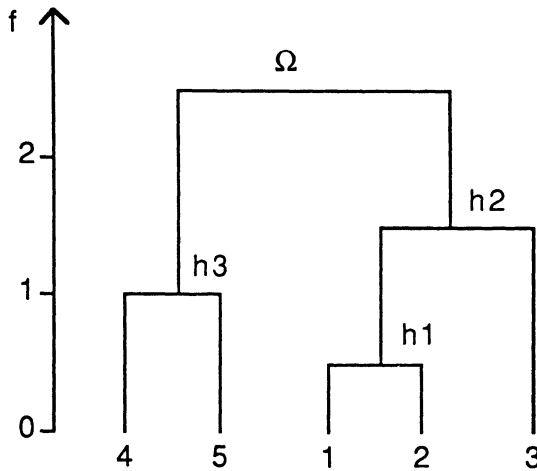


FIGURE 1

Il existe d'autres arbres hiérarchiques visualisant correctement (H, f) : pour les obtenir, il suffit d'afficher les paliers de H en changeant l'ordre d'affichage des objets de Ω (nœuds terminaux de l'arbre). Cependant la modification de l'ordre doit être telle que tout palier de H reste un intervalle vis à vis de l'ordre modifié, ceci afin de garder une bonne lisibilité de l'arbre. Nous reviendrons plus loin sur cette contrainte.

Théorème de Johnson - Benzécri (Johnson (1967), Benzécri (1973))

Il existe une bijection entre l'ensemble des hiérarchies indicées et celui des ultramétriques.

La démonstration de ce résultat tient au fait que l'application qui à chaque hiérarchie indicée (H, f) associe l'indice de dissimilarité induit d_H , est bijective (voir par exemple Diday et al. 1982).

Algorithme de la classification ascendante hiérarchique (CAH)

Il existe de nombreux algorithmes permettant de construire une hiérarchie indicée représentant l'indice de dissimilarité initial d défini a priori sur Ω . Le plus utilisé est celui de la classification ascendante hiérarchique (dit algorithme de la C. A. H.). Notons μ un indice d'agrégation entre parties de Ω , indice choisi au préalable. L'algorithme de la C.A.H. s'énonce alors de la façon suivante :

Etape 1 (initialisation) : les seuls paliers définis sont les singletons de Ω .

Etape 2 (agrégation) : on réunit les deux paliers qui, parmi les paliers n'ayant pas déjà été réunis, sont les plus proches au sens de l'indice d'agrégation μ . On prend pour hauteur du palier créé, la valeur de μ pour le couple de paliers réunis.

Etape 3 (test d'arrêt) : l'algorithme se termine lorsque le dernier palier créé est Ω . Sinon, l'algorithme reprend à l'étape d'agrégation.

Cet algorithme possède la propriété suivante, propriété concernant les données, c'est à dire l'indice de dissimilarité initial (pour une démonstration voir par exemple DIDAY et al. (1982)).

Proposition 1

L'indice d'agrégation du saut minimum ayant été choisi, l'algorithme de la CAH construit une hiérarchie indicée (H, f) pour laquelle d_H est la sous-dominante, c'est à dire l'enveloppe supérieure des ultramétriques inférieures à l'indice de dissimilarité défini a priori sur Ω .

2. Des hiérarchies aux pyramides...

Examinons les caractéristiques des ordres qui permettent d'afficher correctement une hiérarchie donnée, c'est à dire les ordres pour lesquels tout palier de la hiérarchie est un intervalle. Cette propriété sera également exigée dans le cas d'une pyramide.

2.1. Ordres sans croisement, ordres compatibles

Croisement :

On dit qu'un ordre $<$ défini sur Ω donne lieu à un croisement pour une hiérarchie H , s'il existe un triplet d'objets (x, y, z) et un palier h contenant x et z , et tels que :

$$x < y < z \quad \text{et} \quad y \text{ n'est pas élément de } h.$$

Une configuration typique de croisement est donnée ci-dessous :

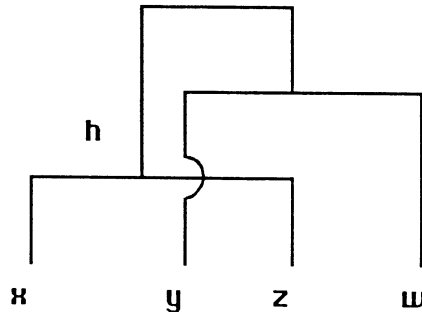


FIGURE 2

Pour éviter ce type de configuration, qui nuit à la lisibilité de l'arbre hiérarchique, on impose à l'ordre d'affichage des objets de Ω d'être sans croisement pour H . Remarquons qu'un ordre est sans croisement pour une hiérarchie, si et seulement si, tout palier de la hiérarchie est un intervalle vis à vis de l'ordre considéré.

Il existe plusieurs autres propriétés équivalentes à l'absence de croisement. Pour les énoncer, nous introduisons la notion d'ordre compatible et la notion de matrice de Robinson.

Ordre compatible

On dit qu'un ordre Δ défini sur Ω , est compatible avec un indice de dissimilarité d si pour tout triplet (x, y, z) d'éléments de Ω vérifiant $x\Delta y\Delta z$, on a :

$$d(x, z) \geq \max\{d(x, y), d(y, z)\}$$

Matrice de Robinson

On dit qu'une matrice réelle symétrique, notée $(m_{i,j})_{1 \leq i, j \leq n}$, est une *matrice de Robinson* si les termes de ses lignes et de ses colonnes sont croissants à partir de chaque terme de sa diagonale, soit formellement :

Pour tout $i < j$	$m_{i,j} \leq m_{i,(j+1)}$	(<i>i</i> ème ligne croissante)
	$m_{i,j} \leq m_{(i-1),j}$	(<i>j</i> ème colonne croissante)

Exemple :

0	1	3	5
	0	2	4
		0	4
			0

FIGURE 3

Caractérisation des ordres sans croisement

Reprenant les notations introduites aux §1.1. et 1.2.1., la proposition ci-dessous est démontrée en annexe.

Proposition 2

Soit Δ un ordre défini sur Ω et (H, f) une hiérarchie indicée sur Ω , les conditions suivantes sont équivalentes :

- (1) l'ordre Δ est sans croisement pour la hiérarchie indicée (H, f)
- (2) d_H et Δ sont compatibles
- (3) $M(d_H, \Delta)$, c'est-à-dire la matrice de terme général $d(i, j)$ où les lignes et les colonnes sont ordonnées suivant Δ , est une matrice de Robinson.

2.2. Introduction aux pyramides

Dans ce paragraphe, nous énonçons une condition générale, notée (1), qui sera vérifiée par les extensions de l'ensemble des hiérarchies indicées que nous considérerons par la suite : à savoir l'ensemble des pyramides indicées au sens large et l'ensemble des pseudo-hiérarchies indicées.

2.2.1. Condition d'extension des hiérarchies

Par souci de simplicité, nous désignerons par «représentation de Ω » tout couple (P, f) où P est un ensemble de parties non vides de Ω , contenant Ω lui-même, et f une application de P dans \mathbf{R}^+ . Nous emploierons également la notion d'indice de dissimilarité induit par une représentation (P, f) , noté d_P , dont la définition est analogue à celle de l'indice de dissimilarité induit par une hiérarchie, c'est à dire :

$$\text{pour tout } x, y \text{ de } \Omega, d_P(x, y) = \min\{f(h) \mid h \in P, x \in h, y \in h\}$$

Nous savons que l'ensemble des hiérarchies indicées est en bijection avec l'ensemble des ultramétriques (théorème de Johnson-Benzécri). Plus précisément, l'application qui à chaque hiérarchie indicée associe l'indice de dissimilarité induit par cette hiérarchie, est bijective.

La condition d'extension, notée (1), que nous allons considérer est l'extension de cette bijection à des ensembles plus vastes que celui des ultramétriques. On trouve aussi l'idée de la condition (1) dans les travaux de Batbedat (1988).

Condition (1)

Soient S un ensemble d'indices de dissimilarité sur Ω contenant les ultramétriques, et E un ensemble de représentation de Ω . Considérons l'application Ψ qui à toute représentation de Ω associe l'indice de dissimilarité induit par cette représentation. On dira que E vérifie la condition (1) si l'application Ψ est bijective de E sur S .

Par la suite (§2.2.2. et §5), nous introduirons deux choix possibles pour E et S .

2.2.2. Définition des pyramides

Dans le cas pyramidal, le choix de S a été guidé par le souci d'obtenir des représentations de Ω pour lesquelles il existe toujours au moins un ordre sans croisement. D'après la proposition 2, il semble raisonnable de choisir pour S l'ensemble des indices de dissimilarité pour lesquels il existe au moins un ordre compatible. Ces indices sont appelés «pyramidaux», plus précisément :

Indice pyramidal

Un indice pyramidal s est un indice de dissimilarité tel que :

- (i) $s(x, y) = 0 \Rightarrow x = y$
- (ii) il existe au moins un ordre compatible avec s , c'est à dire un ordre Δ tel que :

$$x \Delta y \Delta z \Rightarrow s(x, z) \geq \max\{s(x, y), s(y, z)\}$$

Nous allons définir l'ensemble des pyramides indicées au sens large, de sorte que la condition (1) soit vérifiée lorsqu'on prend pour S l'ensemble des indices pyramidaux et pour E l'ensemble des pyramides indicées au sens large.

Pyramide

Soit P un ensemble de parties non vides de Ω . On dit que P est une pyramide sur Ω , si :

- (i) Ω et les singletons de Ω , appartiennent à P ,
- (ii) pour tout h_1, h_2 élément de P , $h_1 \cap h_2$ est soit \emptyset , soit un élément de P ,
- (iii) il existe un ordre sur Ω , tel que tout palier de P soit un intervalle de cet ordre.

Remarques

- 1 – La condition (iii) exprime l'absence de croisement.
- 2 – Les éléments d'une pyramide sont appelés paliers (comme dans le cas hiérarchique)

3 – Les notions de *prédécesseur* et de *successeur* se définissent de la même façon que dans le cas hiérarchique.

Pyramide indicée au sens large

On appelle pyramide indicée au sens large, tout couple (P, f) où P est une pyramide sur Ω , et f une application de P dans R^+ , telle que :

- (1) $f(h) = 0 \iff h$ est un singleton
- (2) pour tout h_1, h_2 élément de P , on a : $h_1 \subset h_2 \Rightarrow f(h_1) \leq f(h_2)$
- (3) $(h_1 \subset h_2 \text{ et } f(h_1) = f(h_2)) \Rightarrow$ il existe deux paliers distincts h et h' , différents de h_1 , tels que : $h \cap h' = h_1$

Remarque

On aurait pu exprimer la condition (3) de la manière suivante :

$(h_1 \subset h_2 \text{ et } f(h_1) = f(h_2)) \Rightarrow$ il existe 2 prédécesseurs distincts de h_1

En effet, nous verrons plus loin (proposition 7) qu'un palier quelconque d'une pyramide admet au plus deux prédécesseurs distincts.

Visualisation d'une pyramide indicée au sens large

Comme dans le cas d'une hiérarchie, on peut associer à toute pyramide un graphe dont les sommets sont les paliers de la pyramide et dont les arcs représentent la relation de succession entre deux paliers. Dans le cas d'une pyramide ce graphe, qui est planaire, n'est pas un arbre car il peut comporter des cycles. Pour représenter une pyramide indicée au sens large (P, f) , chaque palier p non réduit à un singleton est représenté par un trait horizontal dont la hauteur est proportionnelle à la valeur $f(p)$. Contrairement au cas hiérarchique, il est possible que deux branches relient un palier à ses deux prédécesseurs.

Exemple

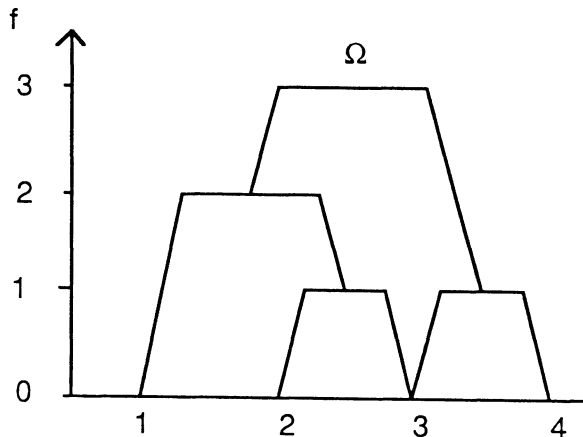


FIGURE 4

$$\Omega = \{1, 2, 3, 4\}$$

$$P = \{\Omega, \{1\}, \{2\}, \{3\}, \{4\}, h_1, h_2, h_3\}$$

avec $h_1 = \{2, 3\}$, $h_2 = \{1, 2, 3\}$, $h_3 = \{3, 4\}$ et pour tout h , $f(h) = \text{card}(h) - 1$

3. Propriétés des pyramides

Les propriétés des pyramides sont, pour la plupart, des généralisations des propriétés des pyramides. Les propositions 3, 4, 5 et 6 montrent que l'ensemble des pyramides indicées au sens large vérifie la condition 1, condition d'extension de l'ensemble des hiérarchies définie au §2.2.1.

Proposition 3

L'indice de dissimilarité induit par une pyramide indicée au sens large est un indice pyramidal.

Démonstration

Soit (P, f) une pyramide indicée au sens large, et d_P l'indice de dissimilarité induit par cette pyramide indicée. Soit Δ un ordre compatible avec P et (x, y, z) un triplet de Ω ordonné selon Δ , c'est à dire :

$$x \Delta y \Delta z$$

Notons h_{xz} (respectivement h_{xy}) le plus bas palier contenant x et z (respectivement x et y). D'après la définition d'une pyramide, h_{xz} est un intervalle pour l'ordre Δ , donc h_{xz} contient y . Par conséquent h_{xy} est inclus dans h_{xz} , et donc :

$$f(h_{xz}) \geq f(h_{xy})$$

d'où $d_P(x, z) \geq d_P(x, y)$. On montre de même que $d_P(x, z) \geq d_P(y, z)$. Donc :

$$d_P(x, z) \geq \max\{d_P(x, y), d_P(y, z)\}$$

Ce qui montre que d_P est bien est un indice pyramidal.

Proposition 4

L'ensemble des hiérarchies indicées est inclus dans l'ensemble des pyramides indicées au sens large.

Démonstration

Soit (H, f) une hiérarchie indicée. Montrons que H est une pyramide. Les conditions (i) et (ii) de la définition d'une pyramide sont trivialement vérifiées. On sait aussi que pour toute hiérarchie indicée, il existe un ordre sans croisement (il s'agit d'un ordre qui créé une chaîne de plus courte longueur : voir Diday 1983).

Ce qui prouve que la condition (iii) est vraie. Donc H est une pyramide, et d'après la définition de f , H est une pyramide indicée au sens large.

Proposition 5

L'ensemble des ultramétriques est inclus dans l'ensemble des indices pyramidaux.

Démonstration

Soit u une ultramétrique. On sait que u est l'indice de dissimilarité induit d'au moins une hiérarchie indicée (voir le théorème de Johnson-Benzécri). Mais d'après la proposition précédente, cette hiérarchie indicée est aussi une pyramide indicée au sens large, dont l'indice de dissimilarité induit est égal à u , donc est un indice pyramidal d'après la proposition 3.

On trouvera une démonstration de la proposition suivante dans Diday (1984).

Proposition 6 (extension du théorème de Johnson-Benzécri)

Notons F l'application qui à toute pyramide indicée au sens large, associe l'indice de dissimilarité qu'elle induit. Cette application F est une bijection de l'ensemble des pyramides indicées au sens large sur l'ensemble des indices pyramidaux.

Les propositions 4, 5, et 6 peuvent être résumées par le diagramme suivant, où i représente l'injection naturelle définie à l'aide de l'inclusion.

$$\begin{array}{ccc}
 \{\text{pyramides indicées} & \xleftarrow{F} & \{\text{indices pyramidaux}\} \\
 \text{au sens large}\} & & \\
 \uparrow i & & \uparrow i \\
 \{\text{hiérarchies indicées}\} & \longleftrightarrow & \{\text{ultramétriques}\}
 \end{array}$$

La proposition suivante montre que les graphes représentant les pyramides sont planaires.

Proposition 7

Tout palier d'une pyramide admet au plus deux prédécesseurs.

Démonstration

Raisonnons par l'absurde. Supposons qu'un palier p de P admette au moins trois prédécesseurs distincts, notés p_1 , p_2 et p_3 .

Comme ces paliers p_i sont prédécesseurs du même palier p , nous avons :

$$(1) \quad p = p_1 \cap p_2 = p_1 \cap p_3 = p_2 \cap p_3$$

Choisissons un ordre noté \leq , qui soit compatible avec P . Les paliers de P sont des intervalles relativement à l'ordre \leq . Donc les paliers p_i peuvent se mettre sous

la forme :

$$p_i = [a_i, b_i]$$

où i prend les valeurs 1, 2 et 3, et où a_i et b_i sont des éléments de Ω . Sans perte de généralité, nous pouvons supposer que les éléments a_i sont rangés de la manière suivante :

$$a_1 \leq a_2 \leq a_3$$

La deuxième égalité de (1) peut alors s'écrire sous la forme :

$$[a_2, \min(b_1, b_2)] = [a_3, \min(b_1, b_3)],$$

Par conséquent $a_2 = a_3$, donc il existe une relation d'inclusion entre p_2 et p_3 , ce qui est contradictoire. Donc le palier p admet au plus deux prédécesseurs.

4. Algorithme de construction

Il existe plusieurs algorithmes de construction de pyramides indicées au sens large. Nous présentons celui qui est le plus proche, dans son principe, de l'algorithme de CAH décrit au §1.2. Il s'agit de l'algorithme de la classification ascendante pyramidale, noté CAP. Avant de l'énoncer, nous donnons les notations et conventions suivantes :

Relations d'ordre sur l'ensemble des parties de Ω

Soient A et B deux parties de Ω , et notons \leq une relation d'ordre sur Ω . Notons également $\min(A)$ le plus petit élément de A , et $\max(A)$ le plus grand élément de A selon l'ordre \leq . Ces deux éléments $\min(A)$ et $\max(A)$ sont aussi appelés bornes de la partie A . Nous dirons que B est *avant* A , si :

$$(\min(B) < \min(A) \text{ et } \max(B) < \max(A)) \text{ ou } A = B$$

Nous dirons que B est à l'intérieur de A , si :

$$\min(A) < \min(B) \leq \max(B) < \max(A)$$

Composante

A toute étape intermédiaire de l'algorithme de CAP, nous pouvons considérer le graphe G dont les noeuds sont les paliers déjà créés par l'algorithme, et les arêtes joignent les couples de paliers prédécesseur-successeur. On appellera composante d'un palier p , et notera $C(p)$, l'union des paliers appartenant à la composante connexe du nœud associé au palier p dans le graphe G .

Exemple

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

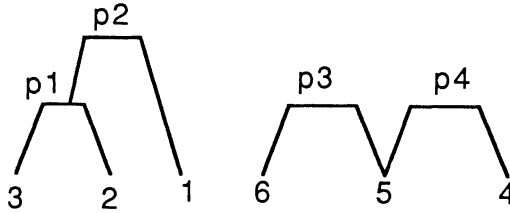


FIGURE 5

Pour cette pyramide «en construction», il y a deux composantes : $\{1, 2, 3\}$ et $\{4, 5, 6\}$.

Algorithme de la Classification Ascendante Pyramidale

Nous choisissons au préalable un indice d'agrégation entre parties de Ω , noté μ .

ETAPE 1 (*initialisation*)

Les seuls paliers existants sont les singletons de Ω . Un ordre arbitraire est choisi entre les éléments de Ω (il est nécessairement compatible avec les singletons de Ω).

ETAPE 2 (*agrégation*)

On réunit deux paliers p^* et q^* , tels que $\mu(p^*, q^*)$ réalise le minimum de μ relativement à l'ensemble des couples de paliers (p, q) vérifiant :

- (i) p est avant q ,
- (ii) il n'existe pas de palier tel que p ou q soient à l'intérieur de ce palier,
- (iii) si p et q appartiennent à la même composante connexe, alors tout palier x de cette composante vérifie :

$$\max(p) < \max(x) \Rightarrow \min(q) \leq \min(x)$$

- (iv) si p et q n'appartiennent pas à la même composante connexe, alors le palier p contient une borne de $C(p)$ et le palier q contient une borne de $C(q)$.

ETAPE 3 (*mise à jour de l'ordre compatible*)

Si avant l'étape 2, p^* et q^* n'appartiennent pas à la même composante connexe, alors les éléments de $C(q^*)$ sont placés consécutivement à ceux de $C(p^*)$. De plus, l'ordre des éléments d'une ou deux de ces composantes est, le cas échéant, inversé de manière à ce que $\min(q^*)$ soit consécutif à $\max(p^*)$.

ETAPE 4 (*test d'arrêt*)

Si le dernier palier créé est Ω , l'algorithme s'arrête, sinon l'algorithme se poursuit à l'étape d'agrégation.

Remarque 1

Dans l'exemple de la figure 5, les couples de paliers vérifiant les quatre conditions de l'étape d'agrégation, sont :

a) pour des composantes connexes identiques :

$$(p_3, p_4)$$

b) pour des composantes connexes différentes :

$$\{\{3\}, p_1, p_2, \{1\}\} \times \{\{6\}, p_3, p_4, \{4\}\}$$

Remarque 2

La complexité de cet algorithme est polynomiale d'ordre compris entre 3 et 4. Il existe un autre algorithme de construction ascendante pyramidale, dont la complexité est polynomiale d'ordre 2. Mais cet algorithme n'est utilisable qu'avec certains indices d'agrégation (pour plus de détails voir Bertrand 1986).

Remarque 3

On montre que l'algorithme de CAP construit, dans tous les cas de figure, une pyramide sur Ω (pour plus de détails voir Bertrand 1986). Notons P cette pyramide et f la fonction définie par :

$$\text{pour tout } h \in P, f(h) = \mu(p^*, q^*),$$

où p^* et q^* sont les deux paliers réunis à l'étape 2 et dont l'union est égale à h . On montre facilement que (P, f) est une hiérarchie indicée au sens large si μ est monotone croissant, c'est à dire si : $A \subset B \Rightarrow \mu(A, C) \leq \mu(B, C)$. Dans le cas contraire, il peut se produire des inversions, c'est à dire l'existence de deux paliers p et q tels que :

$$p \subset q \text{ et } f(p) > f(q).$$

5. Comparaison avec les pseudo-hiérarchies

B. Fichet propose une autre extension de l'ensemble des hiérarchies : les pseudo-hiérarchies indicées (voir Durand et Fichet (1986, 1988)). Afin de simplifier la comparaison avec le cas pyramidal, nous supposerons ici que les pseudo-hiérarchies sur Ω sont totales, c'est à dire contiennent les singletons de Ω . Cette hypothèse étant admise, «pseudo-hiérarchie» devient synonyme de «pyramide». Par contre, la notion de «pseudo-hiérarchie indicée» devient un type particulier de «pyramide indicée au sens large».

Pseudo-hiérarchie (totale) indicée

On appelle pseudo-hiérarchie indicée, tout couple (H, f) où H est une pseudo-hiérarchie (i.e. pyramide) sur Ω , et f une application de H dans \mathbf{R}^+ , telle que :

$$(1) h \text{ est un singleton} \Rightarrow f(h) = 0$$

$$(2) \text{ si deux paliers distincts } h \text{ et } h' \text{ sont tels que } h \subset h', \text{ alors } f(h) < f(h')$$

Remarque

L'implication de (1) est en fait une équivalence, l'implication inverse étant entraînée par (2) et le fait que H soit totale.

D'après cette définition, il est clair que les pseudo-hiérarchies indicées sont des pyramides indicées au sens large. Si l'on reprend les notations du §2.2.1., les résultats de C. Durand et B. Fichet (1988), peuvent s'exprimer comme étant un autre choix d'extension de l'ensemble des hiérarchies indicées, choix qui vérifie la condition (1). Si l'on reprend les notations du §2.2.1., pour l'ensemble \mathbf{E} (extension des hiérarchies indicées), les auteurs choisissent l'ensemble des pseudo-hiérarchies indicées, et pour l'ensemble \mathbf{S} (extension des ultramétriques), l'ensemble des indices de dissimilarité fortement de Robinson.

Indice de dissimilarité fortement de Robinson

On dit qu'un indice de dissimilarité s est fortement de Robinson, si :

(i) il existe un ordre Δ compatible avec s

(ii) pour tout (x, y, z) triplet de Ω ordonné selon Δ , c'est à dire $x\Delta y\Delta z$, on a :

$$a) s(x, z) = s(y, z) \Rightarrow (\text{si } z\Delta w, \text{ alors } s(x, w) = s(y, w))$$

$$b) s(x, z) = s(x, y) \Rightarrow (\text{si } w\Delta x, \text{ alors } s(w, z) = s(w, y))$$

En d'autres termes, un indice de dissimilarité s sur Ω est fortement de Robinson s'il existe un ordre Δ sur Ω tel que la matrice $M(s, \Delta)$ soit une matrice de Robinson et vérifie les deux conditions supplémentaires a) et b).

Il est clair que l'ensemble des ultramétriques est inclus dans l'ensemble des indices fortement de Robinson, qui est lui-même inclus dans celui des indices pyramidaux.

En résumé, les pseudo-hiérarchies indicées constituent une extension des hiérarchies indicées qui vérifie la même condition d'extension que les pyramides indicées au sens large, mais cette extension est incluse dans l'extension pyramidale.

L'avantage des pseudo-hiérarchies indicées est d'éviter le cas de figure où un palier p est inclus strictement dans un palier q avec $f(p) = f(q)$. En effet, ce cas de figure prête à confusion lors de la visualisation d'une pyramide. Donnons un exemple de pyramide indicée au sens large pour lequel il existe deux paliers dans ce cas de figure.

Exemple

Soit $\Omega = \{a, b, c, d, e\}$. Considérons l'indice de dissimilarité initial \mathbf{d} (données) précisé par la matrice relative à l'ordre a, b, c, d, e :

	b	c	d	e	
	1	2	3	5	a
		1	3	4	b
			1	2	c
				1	d

En utilisant l'indice d'agrégation du saut maximum, l'algorithme de CAP construit une pyramide indicée au sens large, notée (P, f) , dont l'indice de dissimilarité induit est ici l'indice de dissimilarité initial \mathbf{d} .

Sa visualisation pour l'ordre compatible (a, b, c, d, e) est donnée ci-dessous :

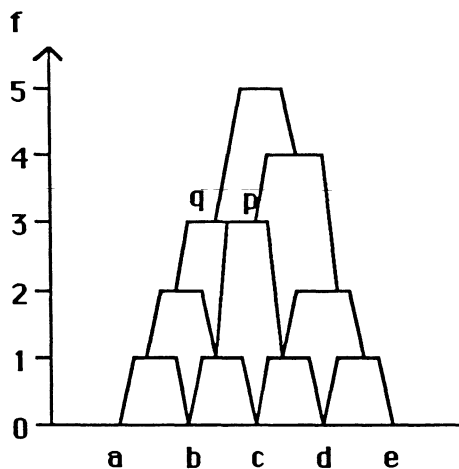


FIGURE 6

Les paliers $p = \{b, c, d\}$ et $q = \{a, b, c, d\}$ réalisent le cas de figure $f(p) = f(q)$ avec p inclus strictement dans q . Par conséquent (P, f) n'est pas une pseudo-hiérarchie indicée. Cependant, on remarquera que l'indice de dissimilarité induit par (P, f) peut être arbitrairement approché (au sens des moindres carrés, par exemple) par les indices de dissimilarité induits par les pseudo-hiérarchies indicées

de la famille $(P, g_e)_e$, où g_e est définie par :

$$\begin{aligned} g_e(x) &= f(x) && \text{si } x \text{ est différent de } q \\ \text{et } g_e(q) &= f(q) + e && \text{où } e \text{ est un réel positif} \\ &&& \text{arbitrairement voisin de zéro} \end{aligned}$$

On remarquera que les visualisations de ces pseudo-hiérarchies admettent pour cas limite la visualisation de la pyramide (P, f) lorsque e tend vers 0. D'ailleurs, dans la pratique lorsque le nombre e est suffisamment petit, les visualisations de (P, g_e) et de (P, f) sont confondues.

6. Exemples de traitement

Nous présentons maintenant deux aspects caractéristiques de la classification pyramidale, en décrivant deux traitements utilisant l'algorithme de CAP sur des jeux de données tests. Cet algorithme a été programmé en FORTRAN 77 par Bertrand (1986), et figure dans une version, actuellement en cours de validation, du logiciel SICLA (Système Interactif de Classification Automatique) développé à l'I.N.R.I.A.

6.1 Sériation par l'approche pyramidale

Dans cet exemple, Ω est un ensemble de douze points du plan euclidien, dont les positions sont indiquées sur la figure 7. L'indice de dissimilarité \mathbf{d} qui constitue avec Ω les données de cet exemple, est la distance euclidienne restreinte à l'ensemble Ω .

Les données se présentent sous la forme du tableau des dissimilarités associées à \mathbf{d} , les éléments de Ω étant rangés selon l'ordre naturel des entiers. Par conséquent, ces données ne contiennent aucune information explicite sur l'ordre entre les points de Ω , que l'on identifie en regardant la figure 7.

Cet ordre, à savoir l'ordre 9, 12, 4, 7, 8, 3, 11, 10, 1, 6, 2, 5 apparaît naturellement comme étant la suite suivant laquelle les points s'ordonnent le long d'une courbe ayant approximativement la forme suivante :



Nous avons effectué une classification ascendante pyramidale en utilisant l'algorithme de CAP (voir figure 8). Cet exemple montre que l'on peut retrouver un ordre implicite entre les objets classés sous la forme d'un ordre compatible avec la pyramide construite par l'algorithme de CAP, c'est à dire ici quasiment l'ordre d'affichage des éléments de Ω (la seule différence concerne le couple d'éléments (1,6) qui apparaît selon cet ordre sur la courbe de la figure 7 et selon l'ordre inverse sur la visualisation de la pyramide de la figure 8).

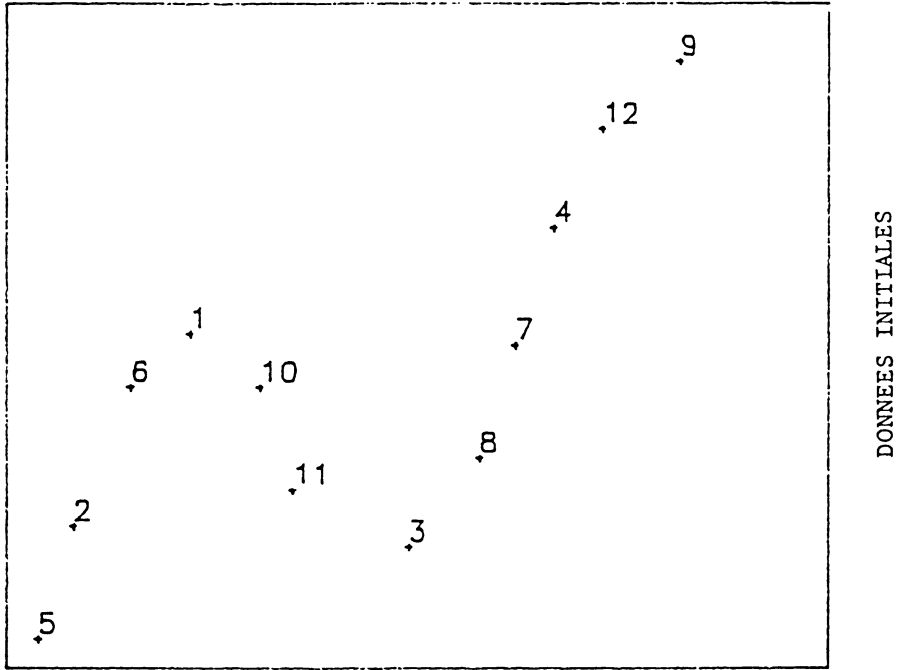


FIGURE 7

Remarquons que l'ordre implicite des données n'aurait pas été retrouvé si la courbe avait comporté une ou plusieurs configurations des types suivants :



6.2. Classification par l'approche pyramidale

Nous présentons un deuxième exemple de classification ascendante pyramidale sur le jeu test classique des données de Ruspini. Ici l'ensemble Ω est constitué de 75 points du plan euclidien répartis en quatre classes distinctes. La figure 9 représente les éléments de Ω par des numéros (1 à 75) les identifiant. L'indice de dissimilarité d qui constitue avec Ω les données de cet exemple, est comme précédemment, la distance euclidienne restreint à l'ensemble Ω .

Ayant choisi l'indice d'agrégation d'inertie, nous avons construit à l'aide de l'algorithme de CAP une pyramide indicée au sens large représentée sur la figure 10. L'ordre d'affichage des éléments de Ω tel qu'il apparaît lors de la visualisation de cette pyramide est aussi indiqué sur la figure 9, par une ligne joignant chacun des points de Ω à son suivant selon cet ordre.

INDICE DU MAXIMUM

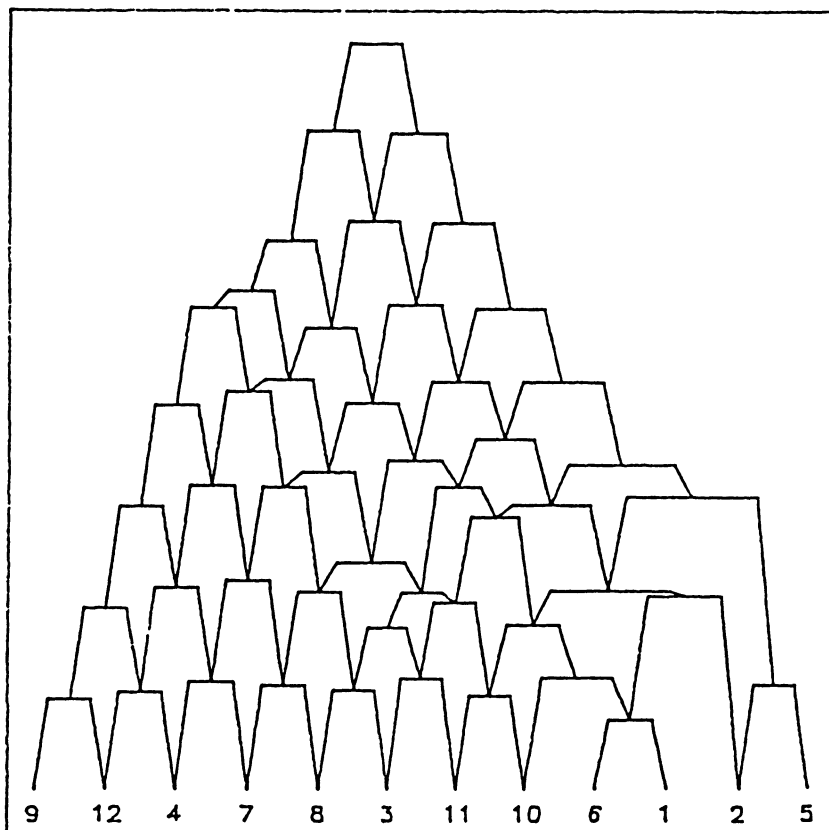


FIGURE 8

La figure 10 est difficilement interprétable car il y a trop de paliers à la même hauteur qui sont dans le bas de la pyramide. Aussi nous avons effectué un grossissement de la base de la pyramide, en ne retenant que les paliers ayant une hauteur inférieure à 4% de la hauteur totale, ce qui donne le graphe de la figure 11. Sur cette dernière figure, les quatre classes des données de Ruspini sont mises en évidence, et apparaissent dans un ordre qui respecte la position intermédiaire de certains points entre deux classes : par exemple les points 46, 47, 48.

RUSPINI

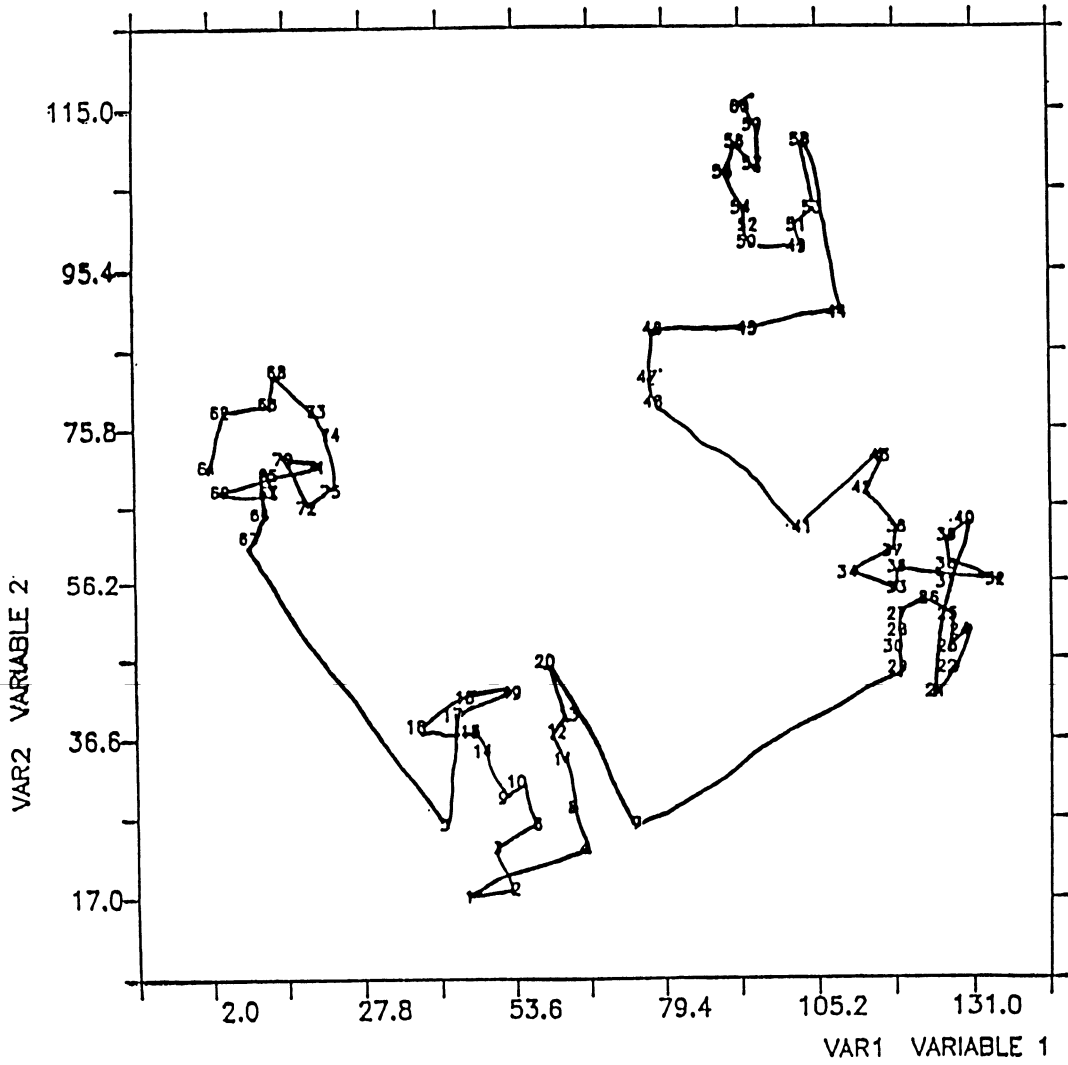
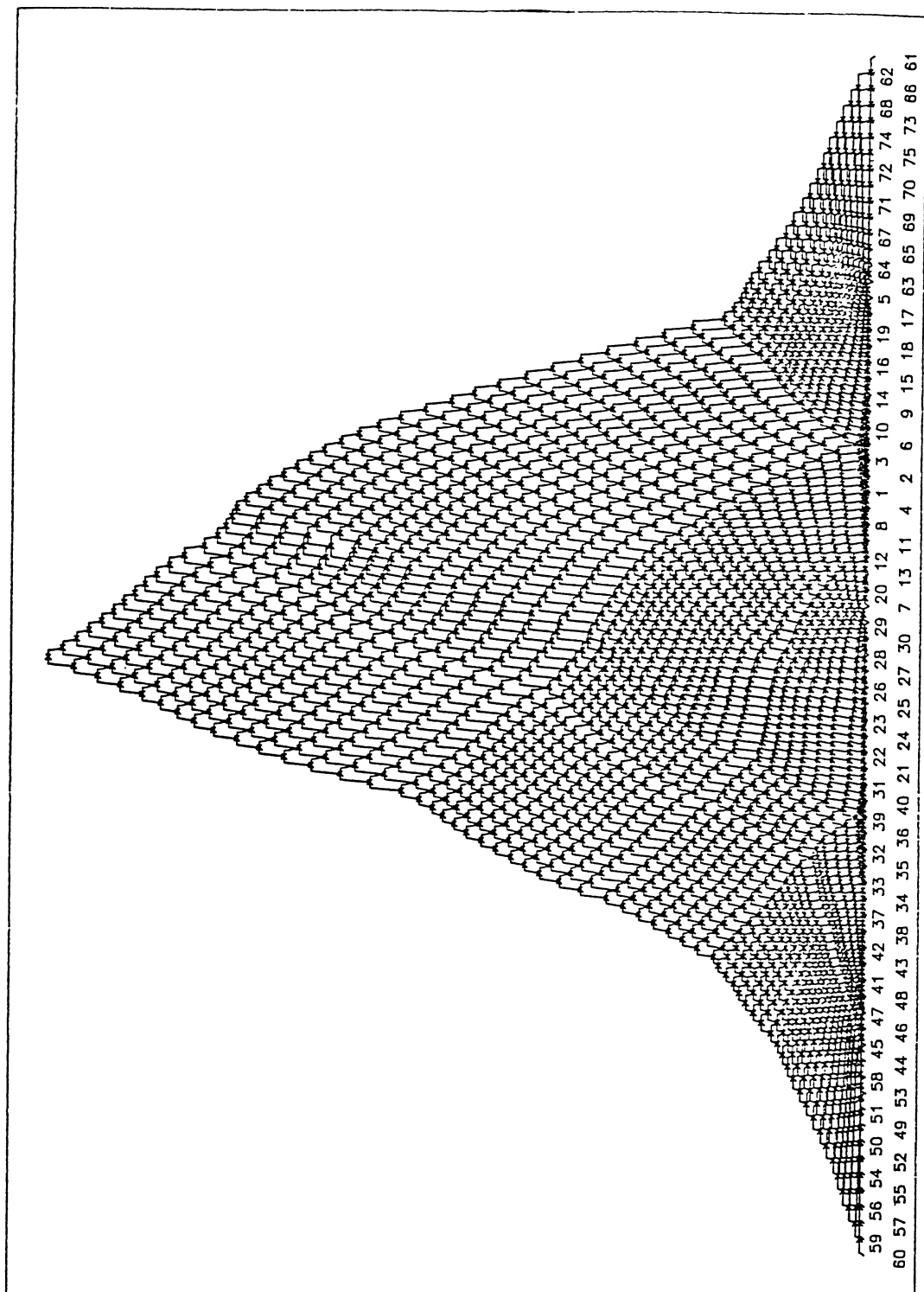


FIGURE 9



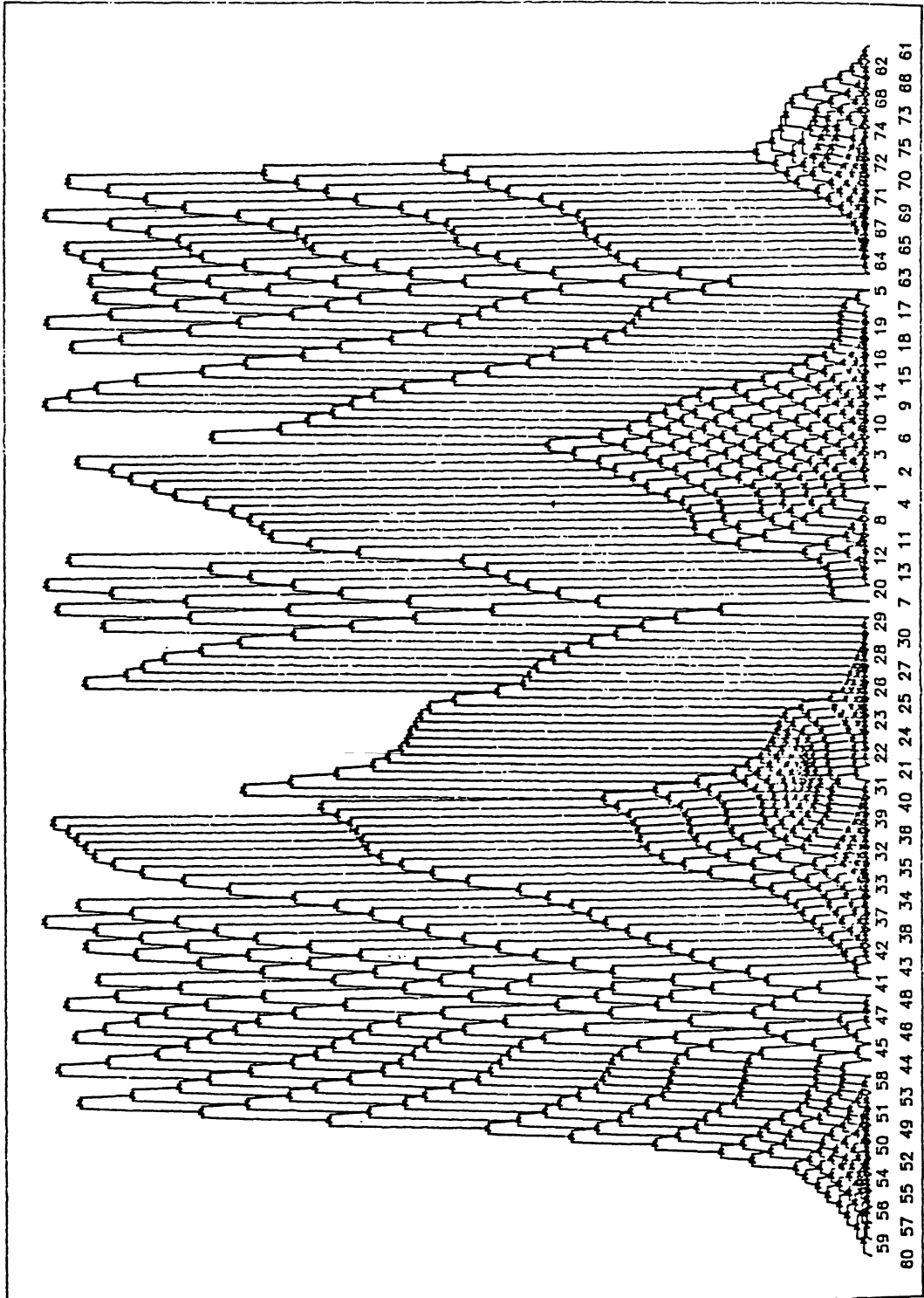


FIGURE 11

Annexe

Démonstration de la proposition 2

(1) \Rightarrow (2)

Soit (x, y, z) un triplet de Ω ordonné par l'ordre Δ , c'est à dire :

$$x\Delta y\Delta z$$

Par définition de d_H , nous savons qu'il existe deux paliers h_1 et h_2 tels que :

$$\begin{aligned} d_H(x, z) &= f(h_1) = \min\{f(h)|h \text{ palier de } H, x \in h, z \in h\} \\ d_H(y, z) &= f(h_2) = \min\{f(h)|h \text{ palier de } H, y \in h, z \in h\} \end{aligned}$$

L'ordre Δ ne donnant lieu à aucun croisement, y appartient à h_1 . Donc $f(h_1) \geq f(h_2)$, et par conséquent $d_H(x, z) \geq d_H(y, z)$.

On montre de la même manière que $d_H(x, z) \geq d_H(x, y)$, ce qui prouve (2).

(2) \Rightarrow (3)

Notons les éléments de Ω par leur rang suivant l'ordre Δ . Comme d_H et Δ sont compatibles, nous avons pour tout $i < j$:

$$\begin{aligned} d_H(i, j+1) &\geq d_H(i, j) \\ d_H(i-1, j) &\geq d_H(i, j) \end{aligned}$$

Ces deux inégalités prouvent que la matrice $M(d_H, \Delta)$ est une matrice de Robinson.

(3) \Rightarrow (1)

Soit (x, y, z) un triplet de Ω ordonné par l'ordre Δ . Montrons que si un palier h contient x et z , alors il contient aussi y : nous aurons ainsi prouvé que l'ordre Δ est sans croisement pour H . Comme la relation $x\Delta y\Delta z$ est vérifiée et que $M(d_H, \Delta)$ est une matrice de Robinson, on a :

$$(i)d_H(x, z) \geq d_H(x, y)$$

D'autre part, il existe deux paliers h_1 et h_2 tels que :

$$\begin{aligned} d_H(x, z) &= f(h_1) = \min\{f(h)|h \text{ palier de } H, x \in h, z \in h\} \\ d_H(x, y) &= f(h_2) = \min\{f(h)|h \text{ palier de } H, x \in h, y \in h\} \end{aligned}$$

Les paliers h_1 et h_2 sont inclus l'un dans l'autre, car ils ont une intersection non vide. Le palier h_1 ne peut pas être strictement inclus dans h_2 , sinon on aurait $f(h_1) < f(h_2)$, ce qui contredit (i). Donc h_2 est inclus dans h_1 . Soit h un palier contenant x et z . On montre de la même manière que h_1 est inclus dans h . Donc h_2 est inclus dans h et par conséquent l'élément y appartient à h .

Bibliographie

- BATBEDAT A. (1988) - L'algorithme Proxel pour les dissimilarités - *Mathématiques Informatique et Sciences Humaines*, N° 102, pp. 31-38.
- BARTHÉLÉMY J.P., GUÉNOCHE A. (1988) - *Les Arbres et les Représentations des proximités* - Masson.
- BENZÉCRI J.P. & coll. (1973) - *L'Analyse des Données, Tome 1 : la Taxonomie* - Dunod.
- BERTRAND P. (1986) - *Etude de la représentation pyramidale* - Thèse de 3° cycle, Université de Paris-Dauphine.
- BERTRAND P., DIDAY E. (1985) - A visual representation of the compatibility between an order and a dissimilarity index : the pyramids - *Computational Statistics quarterly* 2, pp. 31-44.
- BROSSIER G. (1980) - Représentation ordonnée des classifications hiérarchiques - *Statistiques et Analyse des Données*, vol 2, pp. 31-44.
- BUNEMAN P. (1971). - The recovery of trees from measures of dissimilarity - *Mathematics in Archeological and Historical Sciences*, eds : F.R. Hodson, D.G. Kendall, P. Tautu, Edimburgh University Press, pp. 387-395.
- DE LA VEGA F. (1977) - *Raisonnement et méthodes mathématiques* - Editions du C.N.R.S., pp. 157-170.
- DIDAY E. (1983) - Croisements, ordres et ultramétries - *Mathématiques et Sciences humaines*, pp. 31-54.
- DIDAY E. (1984) - Une représentation visuelle des classes empiétantes : les pyramides - Rapport I.N.R.I.A. n° 291, paru en 1986 dans la revue R.A.I.R.O. APII, vol 20 n° 5, pp. 475-526, présenté également en 1984 aux journées de L'A.S.U. de Montpellier.
- DIDAY E., LEMAIRE J., POUGET J., TESTU F. (1982) - *Eléments d'analyse des données* - Dunod.
- DURAND C. (1986) - *Sur la Représentation Pyramidale en Analyse des Données* - Mémoire de DEA, Université de Provence.
- DURAND C., FICHET B. (1988) - One-to-one correspondences in pyramidal representation : a unified approach - in H.H. Bock, ed., *Classification and Related Methods of Data Analysis*, North-Holland, pp. 85-90.
- HARTIGAN J.A. (1975) - *Clustering Algorithm* - Wiley.
- HARTIGAN J.A. (1977) - Clustering as modes - First International Symposium on data analysis and informatics, 7-9 septembre, vol. 2, pp. 433-448, colloques I.R.I.A.
- HUBERT L. (1974) - Some applications of graph theory to clustering - *Psychometrika* 1974, 39, pp. 283-309.
- JARDINE N., SIBSON R. (1971) - *Mathematical Taxonomy* - New York, Wiley.
- JOHNSON S.C. (1967) - Hierarchical clustering schemes - *Psychometrika*, vol. 12, n° 3, pp. 241-254.
- MONJARDET B. (1980) - *Théorie des graphes et Taxonomie mathématique* - in *Regards sur la théorie des graphes*, Presses polytechniques Romandes, Lausanne, pp. 111-125.

- ROHLF F. (1975) - A new approach to the computation of the Jardine-Sibson Bk clusters - *Computer Journal*, 18, pp. 164-168.
- RUSPINI E.H. (1970) - Numerical Methods for Fuzzy Clustering - *Information Sciences*, vol 2, n° 3, pp. 319-350
- SHEPARD R., ARABIE P. (1979) - Additive clustering : Representation of Similarities as combinations of Discrete Overlapping Properties - *Psychological Review*, vol. 86, num 2, pp. 87-123.
- SNEATH P., SOKAL R. (1973) - *Numerical Taxonomy*, Freeman.
- TVERSKY A. (1977) - Features of similarity - *Psychological Review*, 84, pp. 327-352.