

Principes et calculs de la méthode implantée dans le programme CHAVL (Classification Hiérarchique par Analyse de la Vraisemblance des Liens).

- PREMIÈRE PARTIE -

I.C. LERMAN* PH. PETER† et H. LEREDDE‡

Table des matières

- 1 Introduction ; les spécificités de l'approche.
- 2 Vision "classique" de la CAH et justification de l'AVL.
- 3 La méthode de classification AVL dans le cas le plus simple (autour d'un exemple traité à la main).
 - 3.1 Association entre deux attributs.
 - 3.2 Famille de critères d'agrégation de la "vraisemblance du lien maximal".
 - 3.3 Nœuds et niveaux "significatifs" d'un arbre des classifications.
 - 3.4 Aide à l'interprétation : degré de "neutralité" des éléments de l'ensemble à classifier.
 - 3.5 Comment atteindre une échelle discriminante de probabilité, lorsque l'ensemble O des objets devient "gros".
 - 3.6 Indice de similarité probabiliste sur l'ensemble O des objets, décrits par des attributs booléens (cas de l'exemple du tableau 1).

* IRISA, UNIVERSITÉ DE RENNES 1, campus de Beaulieu, avenue du Général Leclerc, 35042 Rennes cedex, tél (33)99.84.71.00, E-mail : lerman@irisa.fr, fax (33)99.38.38.32.

† IRESTE, UNIVERSITÉ DE NANTES, La Chantrerie, CP 3003, 44087 Nantes cedex 03, tél. (33)40 68 30 00, E-mail : ppeter@ireste.fr.

‡ UNIVERSITÉ PARIS NORD, avenue Jean Baptiste Clément, 93430 Villetaneuse, tél. (33)1 49 40 30 00.

Important

Cet article possède un complément qui paraîtra dans le prochain numéro.

Les références bibliographiques présentées à la fin du texte concernent *également* la partie complémentaire.

1 Introduction ; les spécificités de l'approche.

La méthode est centrée sur la notion primordiale de "SIMILARITÉ", au sens logico-statistique du terme, entre entités qui admettent une représentation sous forme de structures mathématico-logiques. L'originalité fondamentale de la méthode concerne la manière dont la notion de similarité est élaborée.

Cette approche utilise bien et développe l'algorithme de la Construction Ascendante Hiérarchique. Mais, elle a bien plus que des aspects algorithmiques et permet en fait une vision très générale des "données" (d'observation, expérimentales ou de connaissance), de leur représentation et de leur synthèse automatique. C'est ainsi que les variables descriptives sont interprétées en termes de relations qui sont représentées au niveau de l'ensemble décrit de façon ensembliste et combinatoire (nous serons plus clairs sur ce point ci-dessous).

D'autre part, cette méthode introduit de façon opérationnelle le concept le plus original de la statistique qui est celui de "vraisemblance". Nous mettons ainsi en évidence la notion de "vraisemblance" dans celle de "ressemblance"; d'où le nom de la méthode. C'est ainsi que nos coefficients d'association entre variables de description (resp. nos indices de similarité entre objets élémentaires ou catégories) seront "probabilistes"; c'est à dire se référeront à une échelle de probabilité pour l'évaluation du lien. Cette démarche est également conforme avec celle sous-jacente au formalisme de la théorie de l'information où un évènement est d'autant plus quantifié qu'*a priori* il était invraisemblable. Dans notre cas les évènements correspondent aux liens observés à partir d'indices "bruts" ou construits d'association. L'expression de ces derniers peut s'imposer de façon "naturelle" ou (non exclusif) se justifier, compte tenu de la représentation mathématico-logique des variables de description; et, celle qui en découle, des unités de données. Le degré d'invraisemblance est calculé par rapport à une hypothèse d'absence de liaison (h.a.l.) ou d'indépendance mutuelle entre les variables. Cette hypothèse est conditionnelle et tient compte de la nature des échelles de description ainsi que des caractéristiques statistiques marginales observées de chacune des variables prises une à une [Daudé 1992, Lerman 1981, 1986, 1987, 1991, 1992_a, 1992_b, Lerman & Peter 1989, Ouali-Allah 1991, Peter 1987].

La démarche précédente est d'une extrême généralité par rapport à la structure logico-statistique du tableau T des données dont l'ensemble des lignes peut représenter soit un ensemble O "d'objets élémentaires" qu'on dit également "individus", soit un ensemble C de "catégories" qu'on dit également "classes", "concepts" ou "modalités". Alors que l'ensemble des colonnes du tableau est étiqueté par un ensemble V de variables descriptives qu'on suppose toutes d'un même type (e.g. quantitatives, booléennes, nominales, ordinales, ...). De sorte, qu'il pourra nous arriver dans notre expression ici, de dépasser le cadre strict des types de données traitées dans CHAVL. Cependant, ce cadre reste très large.

La méthode permet aussi bien et avec la même rigueur conceptuelle, de classier l'ensemble représenté en lignes de T , que celui représenté en colonnes. D'ailleurs, une étape ultime -mais non encore rendue systématique et qui n'existe pas au niveau de CHAVL- peut conduire au croisement de deux arbres condensés de classification, sur l'ensemble des lignes et celui des colonnes de T .

En effet, un point d'importance, distinctif de la méthode consiste en la détection de nœuds ou de niveaux "significatifs" de l'arbre des classifications. En s'exprimant de façon intuitive et abusive, un nœud "significatif" correspond à la complétion d'une classe pour un degré de synthèse donné; alors qu'un niveau "significatif" détermine une partition qui exprime un état d'équilibre dans la synthèse classificatoire.

Au paragraphe 2 ci-dessous, nous rappellerons de façon résumée la vision "classique" de la Classification Ascendante Hiéarchique (CAH). Nous montrerons alors, comment peut s'inscrire la justification de la méthode AVL.

Au paragraphe 3 on explicite la suite des calculs dans le cas d'un tableau de données de petite taille : 10 objets \times 8 attributs. Il s'agit de la structure de la donnée la plus simple pour la méthode, où tous les attributs sont booléens. Nous y exprimerons les principales idées, qu'il s'agit de rendre opérationnelles dans les situations les plus générales, quant à la structure logico-statistique des données.

Les différentes structures du tableau des données pouvant être traitées par CHAVL sont exprimées au paragraphe 1 de la seconde partie. Nous y mentionnerons également les extensions et développements qui ont donné lieu à des programmes informatiques et à des applications d'envergure. Nous terminerons en signalant les structures les plus générales des données pouvant être embrassées par AVL.

Dans ce même paragraphe 1 de la seconde partie nous expliciterons les expressions calcul implantées dans CHAVL. Soulignons que pour chaque cas de figure, un seul indice est proposé. Néanmoins, nous mentionnerons les extensions qui ont été réalisées dans notre environnement ; ainsi que celles potentielles.

Dans le dernier paragraphe 2 de la seconde partie, nous indiquerons diverses extensions méthodologiques dont certaines sont déjà ponctuées par des réalisations importantes. Ce qu'elles ont de commun c'est de partir de la même notion de similarité probabiliste et d'utiliser des modes complémentaires ou d'autres modes de représentation qu'un arbre de classification. Dans les extensions mentionnées, certains travaux sont en cours.

2 Vision "classique" de la CAH et justification de l'AVL.

Pour une présentation synthétique de la Classification Ascendante Hiéarchique (CAH), on peut imaginer que -classiquement- la donnée de base se réduit à un triplet.

$$(O, \mu_O, d) \tag{1}$$

où O est un ensemble d'objets élémentaires sur lequel il y a lieu de découvrir un système de classes et sous-classes d'association, où μ_O est une mesure positive sur O affectant à chacun de ses éléments x un poids μ_x et où d est un indice de distance ou de dissimilarité sur O .

Dans le précédent système, une sous-classe présente une plus grande cohésion que la classe qui l'englobe. D'autre part, le poids μ_x indique "l'importance" avec laquelle doit être considéré l'objet $x, x \in O$. Enfin, on peut signaler que très souvent -dans la vision

classique- relativement à la description de O par un ensemble de variables descriptives, on cherche -de façon plus ou moins justifiée- à représenter (O, μ_O) dans un espace géométrique par un "nuage" de points et à munir cet espace d'une métrique, de telle sorte que la distance entre points reflète la dissemblance entre objets.

En supposant la symétrie de d sur $O \times O$ et une valeur nulle de $d(x, x)$ pour tout x appartenant à O , la donnée (1) est équivalente à celle

$$\{d(x, y), \mu_x, \mu_y / \{x, y\} \in P_2(O)\} \quad (2)$$

où $P_2(O)$ est l'ensemble des paires d'éléments (distincts) de O .

Le propre de la CAH est de passer du triplet (1) à celui

$$(P, \mu_p, \delta) \quad (3)$$

où P est l'ensemble des parties de O , où μ_p est une mesure positive sur P se déduisant de μ_O et où δ est un indice de dissimilarité entre parties de O :

$$\delta : (P \times P, \mu_p) \rightarrow R+ \quad (4)$$

où $R+$ est l'ensemble des nombres positifs.

Il est important bien sûr d'induire δ à partir de d façon cohérente: et, il n'y a pas une seule procédure pour cette cruciale induction. Cependant, de façon formelle, on a toujours:

$$\begin{aligned} \forall (X, Y) \in P \times P, \\ \delta(X, Y) &= f[\{d(x, y) / (x, y) \in (X \cup Y) \times (X \cup Y)\}, \\ &\quad \mu_{X \cup Y} \end{aligned} \quad (5)$$

où nous avons noté μ_Z , le système des poids $\{\mu_z / z \in Z\}$ et où la procédure se trouve définie par la fonction f des distances mutuelles entre éléments pesants de $Z = X \cup Y$.

En vérité, on n'a besoin de la définition de δ qu'entre parties disjointes de O . En effet, à partir de là, le principe mathématique de l'algorithme de "Construction Ascendante Hiérarchique" de l'arbre de classification sur O est "trivial":

A chaque pas, réunir les paires de classes qui rendent minimale la valeur de δ .

Pour l'implantation de l'algorithme, on peut voir la précédente construction ascendante comme l'évolution d'un système dont le k -ème état qui concerne le k -ème niveau de l'arbre, se trouve matérialisé par un couple (T_k, μ^k) ; où T_k est la table des indices δ

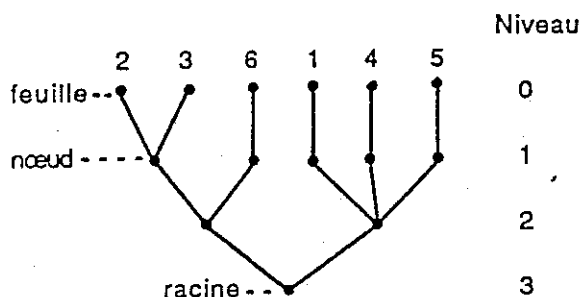


FIG. 1 - Arbre de classification $A_6(O)$ où $O = \{1, 2, 3, 4, 5, 6\}$.

entre classes formées au niveau k et où μ^k est la suite des poids de ces classes. L'état initial (T_0, μ^0) se trouve défini par la matrice T_0 des indices δ entre classes singletons (dont chacune contient exactement un seul élément) et par μ^0 , où le poids d'une classe singleton est celui de l'élément unique qu'elle contient. Si l est le numéro du dernier niveau de l'arbre correspondant à la racine ($l = 3$ dans l'exemple de la figure 1 ci-dessus), on doit considérer la suite des états du système pour $0 \leq k \leq l-1$. Dans ces conditions, on comprend aisément toute l'importance de la formule dite de "réactualisation" qui se présente comme suit :

$$(T_{K+1}, \mu^{k+1}) = \varphi(T_K, \mu^k) \quad (6)$$

état du système état du système
au niveau $k+1$ au niveau k

$0 \leq k \leq l-2$, où φ est une fonction à déterminer compte tenu de δ .

Si c_k est le nombre de classes formées au niveau k , la dimension nécessaire de T_k est $c_k(c_k - 1)/2$. En effet, δ est symétrique et il n'y a pas pertinence à définir δ entre une classe et elle-même, pour la construction ci-dessus.

Relativement à la présentation ci-dessus, on peut déjà signaler -comme d'ailleurs on le verra au paragraphe 3 ci-dessous- que AVL travaille avec des indices de similarité.

D'autre part et surtout, l'algorithme implanté est celui dit des "voisins réciproques", maintenant couramment utilisé pour ce qu'on appelle la classification de "gros ensembles" [Bruynooghe 1989, Peter 1987]. Rappelons rapidement l'idée de cet algorithme en commençant par substituer au triplet (1) ci-dessus, celui correspondant à (T_0, μ^0) ci-dessus :

$$\{P_0, \mu_{P_0}, \delta\} \quad (7)$$

où $P_0 = \{\{x\}/x \in O\}$ est la partition discrète dont chaque classe contient exactement un élément et où μ_{P_0} est la mesure sur P_0 telle que :

$$(\forall x \in O), \mu_{\{x\}} = \mu_x \quad (8)$$

Ici encore, l'algorithme procède par agrégations successives ; mais où, à chaque pas, relativement aux classes en présence on fusionne les paires de classes $\{X, Y\}$, plus proches voisines réciproques l'une de l'autre ; c'est à dire telles que X (resp. Y) est la classe la plus proche au sens de δ de Y (resp. X). Partant de (7), on a, ici encore, besoin de la formule de réactualisation (6), compte tenu des nouvelles classes formées.

On démontre que dans le cas -vérifié par les critères les plus intéressants- où δ possède une propriété dite de "contractance", la famille de classes obtenue dans l'algorithme des voisins réciproques est la même que celle, obtenue dans l'algorithme classique ci-dessus : il s'agit de la même hiérarchie de parties [voir par exemple Bruynooghe 1989]. Signalons déjà que tout élément de la famille de critères de la "vraisemblance du lien maximal" que nous présentons au paragraphe 3 ci-dessous, est bien "contractant" [voir dans Lerman 1981].

Cependant, à partir de la hiérarchie de partie construite par l'algorithme des voisins réciproques, il faut bien retrouver l'arbre des classifications tel qu'il est dessiné dans la figure 1. Il est en effet notamment essentiel d'attacher à chacun des niveaux de l'arbre définissant une partition, la valeur d'un critère appelé "statistique globale des niveaux" qui va permettre de détecter les niveaux et nœuds les plus "significatifs" de l'arbre des classifications. Cette reconstitution est bien effectuée dans l'étape AVLVR (Algorithme de la Vraisemblance du Lien utilisant les Voisins Réciproques) de CHAVL. D'autre part, l'algorithme de AVLVR prévoit la gestion de la transitivité dans le cas où plusieurs paires de classes "voisines réciproques" doivent être agrégées en même temps.

Si on veut faire jouer à la CAH le rôle d'un outil général pour l'analyse d'un tableau T des données, on commencera par s'interroger sur la structure de base de ce dernier, lorsqu'il correspond à une description. Or, il y a deux cas fondamentaux pour cette structure :

- (i) $O \times A$
- (ii) $C \times A$

où O (resp. C) indexe l'ensemble des lignes du tableau et où A indexe celui des colonnes.

O est un ensemble d'objets (on dit encore individus) élémentaires (c'est à dire indivisibles) et C est un ensemble de classes (on dit encore catégories, modalités ou concepts). Dans les deux cas de figure (i) et (ii), A représente un ensemble de variables (on dit encore, attributs) de description :

A l'intersection de la i -ème ligne du tableau des données -étiquetée par l'objet o_i (resp. le concept c_i)- et de la j -ème colonne -étiquetée par l'attribut a^j - (cf. figure 2 ci-dessus), se trouve la "valeur" de a^j sur o_i (resp. c_i). Mais, on se gardera d'interpréter cette "valeur" uniquement par référence à la notion classique d'(attribut, valeur) telle qu'elle est comprise en intelligence artificielle. Déjà en statistiques, en imaginant que la description concerne C (cf. (ii) ci-dessus), $a^j(c_i)$ représente la distribution statistique de la

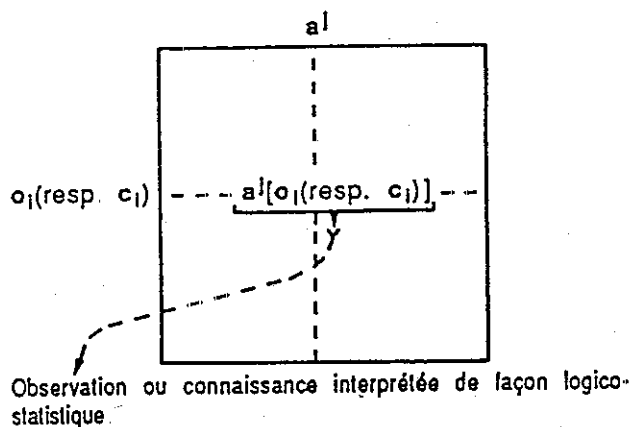


FIG. 2 - Tableau de données.

variable a^j sur la classe c_i , $1 \leq i \leq n, 1 \leq j \leq p$. Le tableau de données correspond en fait à un support de l'information d'une extrême généralité et cela, même dans le cas de données très structurées par la connaissance telles qu'elles se présentent en intelligence artificielle. Ainsi, $a^j[o_i(\text{resp. } c_i)]$ peut par exemple correspondre à une formule logique sur l'ensemble des "valeurs" de la variable; l'ensemble de ces dernières étant muni d'une relation que suppose l'échelle sous jacente à la variable a^j .

Nous venons ci-dessus d'exprimer le contexte classique de présentation générale de la CAH. Ce dernier oublie les deux points fondamentaux suivants :

- (i) A est formé de variables qualitatives, où l'ensemble des "valeurs" ("modalités", ou "catégories") d'une même variable, est (très) structuré par la connaissance du domaine (ou de l'expert);
- (ii) Intérêt spécifique de la classification hiérarchique de l'ensemble A des variables descriptives, en classes et sous-classes "significatives d'association". Une telle décomposition conduit à un système de "facteurs" et "sous-facteurs", où les "sous-facteurs" sont relativement indépendants à l'intérieur du facteur général qui les englobe.

Tenant compte de ce deuxième point, l'ensemble E à classifier peut aussi bien être -ou bien être interprété comme- l'ensemble O des objets (resp. C des concepts) que l'ensemble A des attributs de description :

$$E : \text{ensemble à classifier} \begin{cases} E = A \\ E = O(\text{resp. } C) \end{cases}$$

FIG. 3 - Les deux termes de l'alternative de la classification.

Ce qui rend possible cette double classification en respectant (i) ci-dessus, est précisément la simplicité du principe de l'algorithme de la CAH.

En effet, il y a :

- Avantage scientifique considérable, car la souplesse de l'usage permet d'embrasser les situations de description des données les plus multiples.

Mais aussi, il y a :

- Danger car la "signification" de la synthèse automatique dépend de deux aspects cruciaux :
 - (i) Codage adapté de l'information, résultant de l'observation expérimentale ou de la connaissance.
 - (ii) Notion pertinente -respectant le codage- de proximité sur l'ensemble $P(E)$ des parties de l'ensemble E à classifier.

C'est pour ne pas avoir pris en considération ces deux aspects que des résultats aberrants ont pu être obtenus.

3 La méthode de classification AVL dans le cas le plus simple (autour d'un exemple traité à la main).

Le cas le plus "simple" pour une méthode, concerne la structure de la donnée dans le cadre de laquelle la conception de la méthode est la plus simple à exprimer. Il s'agit ici du cas où A est formé d'attributs booléens de présence-absence et où la description concerne un ensemble O d'objets élémentaires (cf. §1 ci-dessus). Nous allons dans ces conditions procéder à l'illustration la plus simple du schéma de traitement ci-dessus, dans le cas du tableau d'incidence ci-dessous comprenant 10 lignes et 8 colonnes (cf. tableau 1), déjà annoncé dans l'introduction.

3.1 Association entre deux attributs.

Un même attribut a de présence-absence est représenté par le sous-ensemble $O(a)$ des objets où il est présent (on dit encore : où l'attribut booléen est à vrai).

A	a^1	a^2	a^3	a^4	a^5	a^6	a^7	a^8
O_1	0	0	1	1	0	1	0	1
O_2	0	0	0	1	0	0	1	1
O_3	1	1	0	1	1	0	0	0
O_4	0	0	1	0	0	0	1	1
O_5	1	0	1	0	1	0	0	0
O_6	1	0	0	1	1	0	0	0
O_7	1	0	0	0	1	1	1	0
O_8	0	0	1	1	0	0	1	1
O_9	1	0	0	1	1	1	0	0
O_{10}	1	1	0	1	0	0	1	1

Tableau 1 - Table des données.

Relativement à un couple (a, b) d'attributs booléens, introduisons les paramètres suivants qui sont des cardinaux, où $\sim a$ (resp. $\sim b$) indique l'attribut opposé de a (resp. b) :

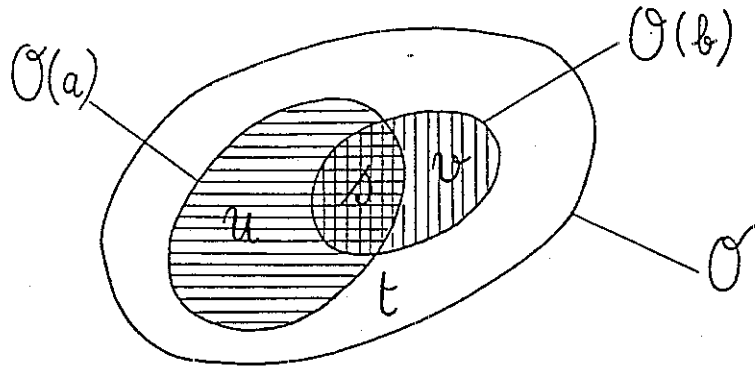
$$\left. \begin{aligned} s &= \text{card}[O(a) \cap O(b)] \\ u &= \text{card}[O(a) \cap O(\sim b)] \\ v &= \text{card}[O(\sim a) \cap O(b)] \\ \text{et } t &= \text{card}[O(\sim a) \cap O(\sim b)] \end{aligned} \right\} \quad (9)$$

s, u, v , et t sont, respectivement, les cardinaux des sous-ensembles de l'ensemble O des objets où sont à vrai les attributs booléens résultants chacun de la conjonction de deux attributs $a \wedge b, a \wedge \sim b, \sim a \wedge b$, et $\sim a \wedge \sim b$. Ces cardinaux prennent place dans le diagramme suivant :

Il est naturel de considérer que les deux attributs a et b sont d'autant plus "ressemblants" que le nombre s d'objets où ils sont tous les deux présents, est relativement grand. On peut donc initialiser la construction de la similarité en considérant l'indice :

$$s(a, b) = s = \text{card}[O(a) \cap O(b)] \quad (10)$$

que nous appelons indice brut de similarité. Il suffit en effet que les deux attributs a et b soient fréquents (resp. rares) pour s'attendre à une valeur plutôt grande (resp. petite) de s . Il y a donc lieu de neutraliser les effets de taille de $\text{card}[O(a)]$ et de $\text{card}[O(b)]$. C'est exactement et de façon implicite ce qu'ont tenté de faire les différents concepteurs d'indices de similarité dans le cas de la comparaison de lignes ou colonnes

FIG. 4 - Comparaison de deux parties de O .

d'un tableau d'incidence formé de zéros et de uns. Chacun de ces indices correspond en fait à une fonction particulière $l(s, u, v)$ du triplet (s, u, v) dans son domaine de définition, croissante par rapport à s , symétrique en u et v et décroissante par rapport à u ; la croissance par rapport à s ou la décroissance par rapport à u étant stricte.

Chacune des fonctions proposées $l(s, u, v)$ se présente en général sous la forme d'un rapport dont le numérateur et le dénominateur sont des fonctions assez simples utilisant les opérations élémentaires (addition, soustraction, multiplication, division et racine carrée) où les paramètres $s, s + u = n(a) = \text{card}[O(a)]$ et $s + v = n(b) = \text{card}[O(b)]$, interviennent le plus directement. Ainsi en est-il d'un indice très utilisé et le plus ancien parmi ceux connus, qui est l'indice de [Jaccard 1908], que nous noterons $J(a, b)$:

$$\begin{aligned}
 J(a, b) &= \frac{s}{s + u + v} = \frac{s}{(s + u) + (s + v) - s} \\
 &= \frac{\text{card}[O(a) \cap O(b)]}{\text{card}[O(a) \cup O(b)]}
 \end{aligned} \tag{11}$$

Notre propre démarche dans l'élaboration d'un indice probabiliste, au niveau de la seule comparaison de deux attributs a et b , consiste à associer au couple $[O(a), O(b)]$ de parties observées de l'ensemble O des objets, un couple (X, Y) de parties aléatoires indépendantes d'un ensemble Ω ; de telle sorte que la suite des cardinaux $[n(a), n(b), n]$ soit "respectée de façon probabiliste" par la suite $[\text{card}(X), \text{card}(Y), \text{card}(\Omega)]$ des trois variables aléatoires entières, respectivement associées (nous serons un peu plus précis sur ce point ci-dessous). La correspondance :

$$[O(a), O(b), O] \longrightarrow [X, Y, \Omega] \tag{12}$$

est, ce que nous appelons "hypothèse d'absence de liaison (h.a.l.)". Dans ce cadre, on considère l'association :

$$s = \text{card}[O(a) \cap O(b)] \longrightarrow S = \text{card}(X \cap Y) \tag{13}$$

faisant correspondre à l'indice brut observé s , l'indice brut aléatoire S . Dans ces conditions, les deux attributs a et b seront considérés d'autant plus ressemblants que s est invraisemblablement grand relativement à la distribution de la variable aléatoire S ; c'est à dire que la probabilité :

$$Pr\{S > s/h.a.l.\} \quad (14)$$

est petite. Dans ces conditions, on prendra comme coefficient d'association (ou de similarité) probabiliste entre les deux attributs a et b , la probabilité complémentaire de (14); soit :

$$P(a, b) = Pr\{S \leq s/h.a.l.\} \quad (15)$$

On se rend compte que de la sorte, on met en évidence le concept de vraisemblance dans celui de ressemblance : encore une fois, les deux attributs seront considérés d'autant plus ressemblants, que le nombre d'objets où ils sont présents ensemble, est invraisemblablement grand, par référence au "hasard parfait" que schématise la correspondance (12). Néanmoins, nous avons pu dégager trois formes fondamentales pour l'h.a.l. [Lerman 1981]. Elles concernent le choix de Ω et le modèle aléatoire de détermination de X (resp. Y) dans l'ensemble $P(\Omega)$ des parties de Ω . Quel que soit l'un des trois modèles aléatoires, on a, en désignant par E l'espérance mathématique :

$$\left. \begin{aligned} E[\text{card}(\Omega)] &= n \\ E[\text{card}(X)] &= n(a) = \text{card}[O(a)] \\ \text{et } E[\text{card}(Y)] &= n(b) = \text{card}[O(b)] \end{aligned} \right\} \quad (16)$$

ce qui correspond, au respect probabiliste ci-dessus mentionné de la cardinalité. En fait, pour chacun des deux premiers modèles $\Omega = O$. Nous nous proposons d'exprimer quelques éléments du second, puisque c'est ce dernier qui conduit aux calculs qui suivront.

Dans le cadre de ce modèle, X (resp. Y) est un élément aléatoire de l'ensemble des parties $P(O)$ de O , muni d'une mesure de probabilité P_a (resp. P_b), telle que, la probabilité d'une partie fixée de cardinal l (resp. m) est donnée par :

$$p(a)^l p(\bar{a})^{n-l} [\text{resp. } p(b)^m p(\bar{b})^{n-m}] \quad (17)$$

où $p(a)$, $p(\bar{a})$, $p(b)$ et $p(\bar{b})$ représentent respectivement les proportions $n(a)/n$, $n(\bar{a})/n$, $n(b)/n$ et $n(\bar{b})/n$, avec des notations que l'on comprend. Dans ces conditions, nous démontrons que la loi de probabilité de la variable aléatoire S (cf. (13)) est binomiale de paramètres n et $p = p(a) \times p(b)$. Il s'ensuit alors, que l'indice (15) ci-dessus se met sous la forme :

$$P(a, b) = \sum_{0 \leq i \leq s} C_n^i p^i (1-p)^{n-i} \quad (18)$$

Nous aurons à l'appliquer dans le cadre de l'exemple avec $n = 10$; et, bien qu'un calcul direct peut bien évidemment être considéré, nous avons préféré présenter ci-dessous la table de la fonction de répartition de la loi binomiale $B(n = 10, p)$ pour différentes valeurs de p .

Par consultation de cette table, on obtient $P(a^3, a^7) = 0.678$ et $P(a^4, a^7) = 0.514$. Plus complètement, on se reportera au tableau 4 qui utilise le tableau 3, pour établir les valeurs des indices :

$$\{P(a^j, a^k)/1 \leq j < k \leq 8\} \quad (19)$$

qui se trouvent au-dessus de la diagonale principale.

A côté du tableau des indices probabilistes, considérons la table des indices "centrés et réduits" (on dit encore "normalisés") par rapport à "l'hypothèse d'absence de liaison". Il s'agit de :

$$\{Q(a^j, a^k)/1 \leq j < k \leq 8\} \quad (20)$$

où

$$Q(a^j, a^k) = \frac{s(a^j, a^k) - E[S(a^j, a^k)]}{\sigma[S(a^j, a^k)]} \quad (21)$$

où E et σ^2 désignent respectivement l'espérance mathématique et la variance, avec $1 \leq j < k \leq 8$.

n	k	p=.05	.10	.15	.20	.25	.30	.35	.40	.45
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025
	1	.9139	.7361	.5443	.3758	.2440	.1493	.0860	.0464	.0233
	2	.9885	.9298	.8202	.6778	.5256	.3828	.2616	.1673	.0996
	3	.9990	.9872	.9500	.8791	.7759	.6496	.5138	.3823	.2660
	4	.9999	.9984	.9901	.9672	.9219	.8497	.7515	.6331	.5044
	5	1.000	.9999	.9986	.9936	.9803	.9527	.9051	.8338	.7384
	6	1.000	1.000	.9999	.9991	.9965	.9894	.9740	.9452	.8980
	7	1.000	1.000	1.000	.9999	.9996	.9984	.9952	.9877	.9726
	8	1.000	1.000	1.000	1.000	1.000	.9999	.9995	.9983	.9955
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.9999	.9997
	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Tableau 2 - Table de la fonction de répartition de la loi binomiale $B(n=10, p): F(k) = \Pr(K \leq k)$.

	a^1	a^2	a^3	a^4	a^5	a^6	a^7	a^8
a^1	6	0.12	0.24	0.42	0.30	0.18	0.30	0.30
a^2	2	2	0.08	0.14	0.10	0.06	0.10	0.10
a^3	1	0	4	0.28	0.20	0.12	0.20	0.20
a^4	4	2	2	7	0.35	0.21	0.35	0.35
a^5	5	1	1	3	5	0.15	0.25	0.25
a^6	2	0	1	2	2	3	0.15	0.15
a^7	2	1	2	3	1	1	5	0.25
a^8	1	1	3	4	0	1	4	5

Tableau 3 - Table des indices bruts entre attributs descriptifs (en-dessous au sens large de la diagonale). Table des paramètres des lois binomiales à consulter (au-dessus strictement de la diagonale).

	a^1	a^2	a^3	a^4	a^5	a^6	a^7	a^8
a^1		0.886	0.270	0.582	0.953	0.735	0.389	0.149
a^2	2.112		0.449	0.842	0.736	0.549	0.736	0.736
a^3	-0.270	0.222		0.440	0.376	0.659	0.678	0.879
a^4	0.614	1.760	0.197		0.514	0.647	0.514	0.752
a^5	3.034	1.182	0.022	0.407		0.820	0.244	0.056
a^6	1.178	0.511	0.875	0.831	1.617		0.544	0.544
a^7	0.041	1.182	0.945	0.407	-0.344	0.496		0.922
a^8	-0.644	1.182	2.048	1.255	-1.059	0.496	2.511	

Tableau 4 - (1ère agrégation): au-dessus de la diagonale; table des indices probabilistes de la "vraisemblance du lien", directement basés sur le modèle binomial. En-dessous de la diagonale: table des $-\log(-\log)$ (à base e) des valeurs des indices probabilistes.

	a^1	a^2	a^3	a^4	a^5	a^6	a^7	a^8
a^1								
a^2	0.778							
a^3	-1.037	-0.933						
a^4	-0.128	0.547	-0.563					
a^5	1.380	0.000	-0.791	-0.331				
a^6	0.165	-0.799	-0.195	-0.078	0.443			
a^7	-0.690	0.000	0.000	-0.331	-1.095	-0.443		
a^8	-1.380	0.000	0.791	0.331	-1.826	-0.443	1.095	

Tableau 5 - Table des indices "centrés et réduits" par rapport au modèle binomial de l'hypothèse d'absence de lien (ou d'indépendance) $\{Q(a^j, a^k) | 1 \leq j < k \leq 8\}$.

L'élaboration de cette table (cf. Tableau 5) s'avère nécessaire dès lors que n n'est pas assez petit; mais où on a:

$$P(a^j, a^k) \cong \phi[Q(a^j, a^k)] \quad (22)$$

$1 \leq j < k \leq 8$, où ϕ est la fonction de répartition de la loi normale centrée et réduite.

On peut signaler que, si on considère le premier (resp. le troisième) modèle aléatoire de "l'hypothèse d'absence de liaison" (cf. ci-dessus), le coefficient $Q(a, b)$ est -au facteur multiplicatif $\sqrt{n-1}$ (resp. \sqrt{n}) près- le coefficient bien connu de K. Pearson (resp. la contribution "orientée" de (a, b) à l'indice χ^2/n du tableau de contingence 2×2 , croisant $\{a, \bar{a}\}$ avec $\{b, \bar{b}\}$).

Comparons à présent les deux associations ; entre a^1 et a^6 d'une part et entre a^4 et a^7 d'autre part. Dans la figure 5, nous avons représenté l'ensemble total O par un rectangle horizontal. Un même attribut a^j sera figuré par un sous-rectangle de même largeur et délimité par deux segments verticaux. La longueur du sous-rectangle représentant a^j est proportionnelle au nombre d'objets où l'attribut a^j prend la valeur 1 ; c'est à dire :

$$\text{card}[(a^j)^{-1}(1)] = \text{card}[O(a^j)] = n(a^j)$$

où les deux dernières expressions ont déjà été introduites. Nous avons hachuré verticalement (resp. horizontalement) les rectangles représentant a^1 et a^4 (resp. a^6 et a^7). D'autre part, la taille du rectangle résultant de l'intersection des deux rectangles qui représentent a^1 et a^6 (resp. a^4 et a^7), est conforme à la cardinalité de $O(a^1) \cap O(a^6)$ [(resp. $O(a^4) \cap O(a^7)$)]. Il n'est pas aisé, en s'appuyant de façon intuitive sur la perception visuelle, de discriminer laquelle des deux paires $\{a^1, a^6\}$ et $\{a^4, a^7\}$, est formée des composantes les plus proches. L'indice de Jaccard donne :

$$J(a^1, a^6) = \frac{2}{7} = 0.286 < J(a^4, a^7) = \frac{3}{9} = 0.333$$

L'inégalité est inversée pour notre indice de "vraisemblance du lien". On a en effet :

$$P(a^1, a^6) = 0.737 > P(a^4, a^7) = 0.514$$

Ainsi, pour l'indice P , une intersection de taille supérieure à 2, lorsqu'il s'agit de comparer deux parties de tailles respectives 3 et 6, est plus invraisemblable qu'une intersection de taille à 3, lorsqu'il s'agit de comparer deux parties de tailles respectives 5 et 7.

Pour être clairs, signalons que les nombres du tableau 4, au-dessus de la diagonale principale, ont été exclusivement obtenus sur la base du tableau 2, en opérant par interpolation linéaire lorsque la valeur du paramètre p n'est pas incluse.

Les différents cas de figure peuvent concerner des aspects statistiques (le nombre n de lignes du tableau des données est plus ou moins grand) ou bien des aspects formels liés à la structure sous-jacente à l'ensemble des modalités d'une même variable relationnelle. Exprimons -avant d'aborder la famille de critères de la vraisemblance du lien maximal (cf. § 3.2 ci-dessous)- que dans tous cas et quel que soit l'ensemble E à classifier [$E = A$ ou $E = O$ (resp. C) (cf. figure 2)], nous aboutissons toujours à une table d'indices probabilistes, comme le suggère le schéma suivant :

$$P = \{P(x, y) / \{x, y\} \in P_2(E)\} \quad (23)$$

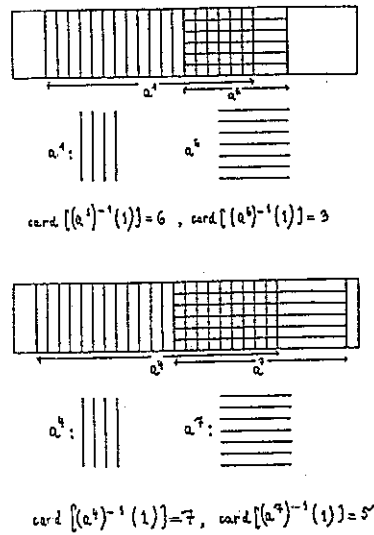


FIG. 5 - (voir tableau 3) Comparaison entre deux associations.

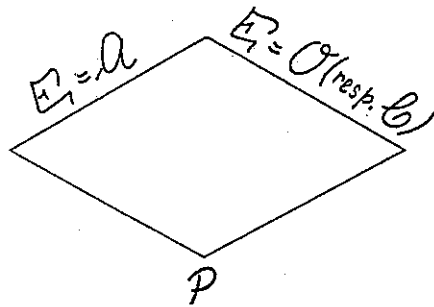


FIG. 6 - Structure probabiliste commune de la similarité.

La table (23) est telle que si $\{x^*, y^*\}$ est la paire d'éléments aléatoires et indépendants associée à $\{x, y\}$ dans l'hypothèse d'absence de liaison, alors $P(x^*, y^*)$ est une variable aléatoire uniformément répartie sur l'intervalle $[0, 1]$, pour toute paire $\{x, y\}$ de $P_2(E)$. Une telle hypothèse est d'indépendance en probabilité entre les différents attributs. Elle se trouve définie par la correspondance :

$$A \longrightarrow A^* \quad (24)$$

où A^* est un ensemble de variables ou attributs aléatoires indépendants, respectivement associés aux attributs de A , en respectant les structures formelles et statistiques de ces derniers pris un par un.

3.2 Famille de critères d'agrégation de la "vraisemblance du lien maximal".

Pour comprendre de façon intuitive ce type de critère de proximité entre classes, nous allons considérer la figure suivante qui se situe dans un contexte simple et classique d'un nuage planaire. Nous suggérons la présence de deux paires de classes $\{C_1, C_2\}$ d'une part et de $\{D_1, D_2\}$ d'autre part. La densité des hachures traduisant la densité des points, C_1 et C_2 sont fortement denses ; alors que D_1 et D_2 sont plus faiblement denses. Commençons par considérer un indice δ -que nous appellerons "brut"- de "distance" entre parties d'un espace métrique et qui est d'ailleurs fourni par la topologie :

$$\delta(X, Y) = \min\{d(x, y) / (x, y) \in X \times Y\} \quad (25)$$

On a :

$$\delta(C_1, C_2) < \delta(D_1, D_2) \quad (26)$$

Mais y-a-t-il pour autant lieu de réunir C_1 et C_2 avant D_1 et D_2 ?

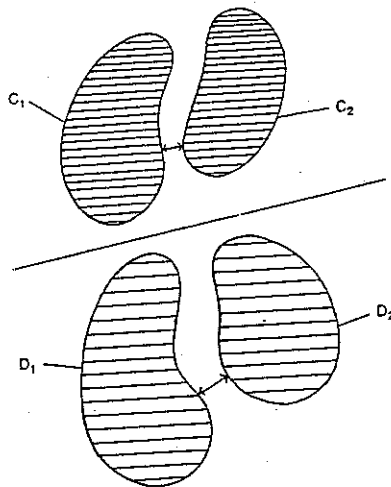


FIG. 7 - La densité des hachures traduit la densité des points.

Le critère de la "vraisemblance du lien maximal" fait l'inverse et commence par agréger D_1 et D_2 avant C_1 et C_2 , car la petitesse de $\delta(D_1, D_2)$ est plus exceptionnelle que celle de $\delta(C_1, C_2)$ eu égard aux densités respectives des deux classes à comparer.

Reprenons la table (23) ci-dessus :

$$\{P(x, y) / \{x, y\} \in P_2(E)\} \quad (27)$$

Le point de départ de la construction du critère de la vraisemblance du lien maximal d'association entre deux classes C et D (parties disjointes de E) est :

$$p(C, D) = \max\{P(c, d)/(c, d) \in C \times D\} \quad (28)$$

Au couple (C, D) de classes, nous associons un couple (C^*, D^*) de classes aléatoires indépendantes, où C^* (resp. D^*) respecte la structure statistique de C (resp. D). Cette association est celle qui définit l'hypothèse H d'absence de liaison (cf. §3.1).

L'indice $p(C^*, D^*)$ devient aléatoire. L'indice définitif d'association entre les deux classes C et D prend la forme :

$$P(C, D) = \frac{Pr\{p(C^*, D^*) \leq p(C, D)/H\}}{[p(C, D)]^{lzm}} \quad (29)$$

où $l = \text{card}(C)$ et $m = \text{card}(D)$

Il s'agit de la forme pure du critère de la vraisemblance du lien maximal qu'on peut noter :

$$VL_1(C, D) = [p(C, D)]^{lzm} \quad (30)$$

alors que $p(C, D)$ correspond -avec l'expression de nos indices- au lien maximal [Lerman 1970 in Lerman 1981].

A la suite des travaux de F. Nicolau [Nicolau 1980], nous avons pu suggérer la famille de critères :

$$VL_\varepsilon(C, D) = [p(C, D)]^{(lzm)^\varepsilon} \quad (31)$$

où ε est un paramètre réel compris entre 0 et 1. Cette famille va du lien maximal ($\varepsilon = 0$), à la forme pure du critère de la vraisemblance du lien maximal ($\varepsilon = 1$).

Des raisons de précision calcul nous conduisent à considérer la fonction strictement croissante :

$$S_\varepsilon(C, D) = -\text{Log}\{-\text{Log}[VL_\varepsilon(C, D)]\} \quad (32)$$

qui conduit exactement au même arbre de classification.

Relativement à cette dernière expression du critère, nous disposons bien d'une formule de réactualisation en cas d'agrégations multiples :

$$S_\varepsilon(C, D) = -\varepsilon \log \left(\sum_{1 \leq i \leq k} c_i \right) - \varepsilon \log \left(\sum_{1 \leq j \leq l} d_j \right) \\ + \max\{S_\varepsilon(C_i, D_j) + \varepsilon \log(c_i) + \varepsilon \log(d_j) \\ / 1 \leq i \leq k, 1 \leq j \leq l\} \quad (33)$$

où nous avons noté $c_i = \text{card}(C_i), 1 \leq i \leq k$ (resp. $d_j = \text{card}(D_j), 1 \leq j \leq l$) et où $C = C_1 \cup C_2 \cup \dots \cup C_k$ (resp. $D = D_1 \cup D_2 \dots \cup D_l$) [Lerman 1989].

Le critère d'agrégation implanté dans CHAVL, correspond à $\varepsilon = 1$. La forme (33) s'avère précieuse dans la pratique car il arrive souvent qu'on ait à gérer des agrégations multiples. Le cas typique est celui de la classification d'un ensemble d'objets de taille suffisamment importante et décrit par des variables qualitatives. Toutefois, la formule que nous utiliserons ci-dessous a la forme classique suivante :

$$S_1(C, D_1 \cup D_2) = -\log(d_1 + d_2) + \max\{S_1(C, D_1) + \log(d_1), S_1(C, D_2) + \log(d_2)\} \quad (34)$$

qui, pour $\varepsilon = 1$, réactualise par rapport à la fusion de deux classes seulement.

Les tableaux 4,6,7,8,9,10 et 11 donnent la suite des états du couple que nous avons désigné au §1 par (T_k, μ^k) pour $0 \leq k \leq 6$. Le tableau 4 correspond à (T_0, μ^0) et le 11, à (T_6, μ^6) . L'arbre comprend par conséquent 8 niveaux si on comprend le niveau 0 (cf. figure 8). La mesure μ^k correspond ici au nombre d'éléments par classe ; elle se trouve indiquée dans la dernière colonne à droite.

Dans notre cas la matrice T_k est de similarité entre classes, $0 \leq k \leq 6$. Nous avons à chaque fois souligné la case indiquant la seule paire de classes qui réalise cette valeur maximale. C'est là une circonstance liée à l'exemple et non un fait général. Ainsi l'arbre est binaire (cf. figure 5).

Lorsque k augmente, $0 \leq k \leq 6$, le nombre de classes formées diminue. Nous avons dans nos tableaux (6,7,8,9,10 et 11) indiqué les différents états de ce qui représente une classe d'étiquette donnée, occupant une colonne (resp. ligne) de la matrice $T_k, 0 \leq k \leq 6$. De cette façon nous retrouvons le contenu d'une classe donnée, concernée dans une table $T_k, 0 \leq k \leq 6$.

1 ← 1 V 5	2 ← 2	3 ← 3	4 ← 4	5 ← 6	6 ← 7	7 ← 8	
							2
1.419							1
-0.617	0.222						1
-0.079	1.760	0.197					1
0.924	0.511	0.875	0.831				1
-0.652	1.182	0.945	0.407	0.496			1
-1.337	1.182	2.048	1.255	0.496	<u>2.511</u>		1

Tableau 6 - (2ème agrégation). Tableau des proximités entre classes. La dernière colonne représente les cardinaux des classes.

1 ← 1	2 ← 2	3 ← 3	4 ← 4	5 ← 5	6 ← 6 V 7	
						2
1.419						1
-0.671	0.222					1
-0.079	1.760	0.197				1
0.924	0.511	0.875	0.831			1
-1.345	0.489	1.355	0.562	-0.197		2

Tableau 7 - (3ème agrégation). Tableau des proximités entre classes. La dernière colonne représente les cardinaux des classes.

1 ← 1	2 ← 2 V 4	3 ← 3	4 ← 5	5 ← 6	
					2
0.726					2
-0.671	-0.471				1
0.924	0.138	0.875			1
-1.345	0.131	1.355	-0.197		2

Tableau 8 - (4ème agrégation). Tableau des proximités entre classes. La dernière colonne représente les cardinaux des classes.

1 ← 1	2 ← 2	3 ← 3 V 5	4 ← 4	
				2
0.726				2
-1.751	-0.537			3
0.924	0.138	-0.224		1

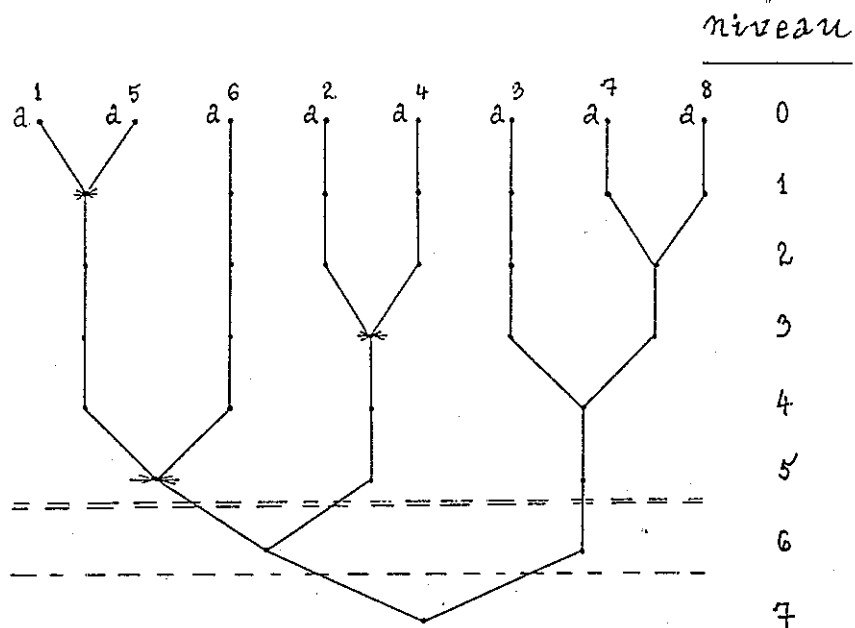
Tableau 9 - (5ème agrégation). Tableau des proximités entre classes. La dernière colonne représente les cardinaux des classes.

1 ← 1 V 4	2 ← 2	3 ← 3	
			3
0.32			2
-1.323	-0.537		3

Tableau 10 - (6ème agrégation). Tableau des proximités entre classes. La dernière colonne représente les cardinaux des classes.

1 ← 1 V 2	2 ← 3	
		5
-1.453		3

Tableau 11 - (7ème agrégation). Tableau des proximités entre classes. La dernière colonne représente les cardinaux des classes.

FIG. 8 - Arbre des classifications sur A .

La suite des partitions ordonnées par finesse décroissante est :

$$\begin{aligned}
 \pi_0 &= \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}\} \\
 \pi_1 &= \{\{1, 5\}, \{2\}, \{3\}, \{4\}, \{6\}, \{7\}, \{8\}\} \\
 \pi_2 &= \{\{1, 5\}, \{2\}, \{3\}, \{4\}, \{6\}, \{7, 8\}\} \\
 \pi_3 &= \{\{1, 5\}, \{2, 4\}, \{3\}, \{6\}, \{7, 8\}\} \\
 \pi_4 &= \{\{1, 5\}, \{2, 4\}, \{3, 7, 8\}, \{6\}\} \\
 \pi_5 &= \{\{1, 5, 6\}, \{2, 4\}, \{3, 7, 8\}\} \\
 \pi_6 &= \{\{1, 2, 4, 5, 6\}, \{3, 7, 8\}\} \\
 \pi_7 &= \{\{1, 2, 3, 4, 5, 6, 7, 8\}\}
 \end{aligned} \tag{35}$$

Introduisons déjà ici, ce qui nous servira directement au paragraphe suivant, les ensembles $R(\pi_i)$ (resp. $S(\pi_i)$), $0 \leq i \leq 7$, où $R(\pi_i)$ (resp. $S(\pi_i)$) est l'ensemble des paires réunies (resp. séparées) par la partition π_i . On a, en notant par xy la paire $\{a^x, a^y\}$, où $x < y$:

$$\begin{aligned}
 R(\pi_0) &= \phi \\
 R(\pi_1) &= \{15\} \\
 R(\pi_2) &= \{15, 78\} \\
 R(\pi_3) &= \{15, 24, 78\} \\
 R(\pi_4) &= \{15, 24, 37, 38, 78\} \\
 R(\pi_5) &= \{15, 16, 56, 24, 37, 38, 78\} \\
 R(\pi_6) &= \{12, 14, 15, 16, 24, 25, 26, 45, 46, 56, \\
 &\quad 37, 38, 78\} \\
 R(\pi_7) &= \{12, 13, 14, 15, 16, 17, 18, 23, 24, 25, \\
 &\quad 26, 27, 28, 34, 35, 36, 37, 38, 45, 46, \\
 &\quad 47, 48, 56, 57, 58, 67, 68, 78\}
 \end{aligned} \tag{36}$$

D'autre part, si $F = P_2(E)$ désigne l'ensemble de toutes les paires de a , on a :

$$S(\pi_i) = F - R(\pi_i), 0 \leq i \leq 7 \quad (37)$$

où la différence est ensembliste.

3.3 Nœuds et niveaux "significatifs" d'un arbre des classifications.

La détermination des niveaux et nœuds les plus "significatifs" d'un arbre des classifications est fondée sur l'élaboration d'un critère d'association -statistiquement normalisé- entre une partition donnée π et une information quant aux ressemblances entre les éléments de l'ensemble E à classifier. La forme la plus pure de cette information est de nature ordinale. Il s'agit de la "pré ordonnance" $w(E)$ sur E ; c'est à dire, du préordre total sur l'ensemble $P_2(E)$ des paires d'éléments de E , associé à la similarité définie. Ce préordre total est tel que le rang d'une paire est d'autant plus grand, que la ressemblance entre ses composantes -évaluée au moyen de l'indice de similarité- est plus forte. Ainsi, la préordonnance $w(A)$, associée à la similarité établie dans le tableau 4, est la suivante, où $r(p)$ désigne une fonction "rang" :

$ \begin{aligned} p : & 58 < 18 < 57 < 13 < 35 < 17 < 34 < 23 < 45 \approx 47 < 67 \approx 68 \\ & < 26 < 14 < 46 < 36 < 37 < 16 < 25 \approx 27 \approx 28 \\ & < 48 < 56 < 24 < 38 < 12 < 78 < 15 \end{aligned} $ <hr style="border-top: 1px dotted black;"/> $ r(p) : 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9,5 \ 11,5 \ 13 \ 14 \ 15 \ 16 \ 17 \ 18 \ 20 \ 22 \ 23 \ 24 \ 25 \ 26 \ 27 \ 28 $ <div style="text-align: right;">(38)</div>
--

La fonction "rang" que nous adoptons en général :

$$\{r(p)/p \in F = P_2(A)\} \quad (39)$$

est celle dite du "rang moyen", où le rang d'un même élément dans une classe d'ex-aequo, pour un préordre total, est égal à la moyenne des rangs des éléments de la classe, pour tout ordre total et strict, compatible. De la sorte, la somme des rangs est toujours constante.

L'usage de cette fonction "rang" permet la généralisation naturelle du critère initialement conçu et calculé dans le cas où l'information de base est un ordre total et strict sur l'ensemble F des paires de l'ensemble E à classifier. L'élaboration du critère suppose qu'on interprète une partition π sur E comme un préordre total à deux classes sur F , $S(\pi)$ et $R(\pi)$, où $S(\pi)$ (resp. $R(\pi)$) indique l'ensemble des paires séparées (resp. réunies)

par la partition π . Relativement à π , une paire séparée (resp. réunie) est formée de composantes "éloignées" (resp. "proches"). De sorte que la partition π est représentée au niveau de $F \times F$ par le produit cartésien :

$$S(\pi) \times R(\pi) \quad (40)$$

alors que l'ordonnance totale $w(E)$ se trouve de la même façon représentée par son graphe strict :

$$gr[w(E)] = \{(p, q) / (p, q) \in F \times F \text{ et } p < q(\text{strictement})\} \quad (41)$$

Dans ces conditions, le critère d'association que nous notons $C(w, \pi)$, entre l'ordonnance $w(E)$ et la partition $\pi(E)$, se met sous la forme ensembliste suivante :

$$C(w, \pi) = \frac{s(w, \pi) - [\tau(\pi) \times s(\pi) / 2]}{\sqrt{\tau(\pi) \times s(\pi)}(f + 1)} \quad (42)$$

où

$$s(w, \pi) = \text{card}\{gr(w) \cap [S(\pi) \times R(\pi)]\} \quad (43)$$

où $\tau(\pi) = \text{card}[R(\pi)]$ (resp. $s(\pi) = \text{card}[S(\pi)]$), et où $f = \text{card}(F) = \tau(\pi) + s(\pi)$ ($f = 28$, dans notre exemple).

$s(w, \pi)$ peut -dans tous les cas- se mettre sous la forme suivante :

$$s(w, \pi) = \Sigma\{v(q) / q \in R(\pi)\} \quad (44)$$

où $v(q)$ est le nombre de paires séparées, strictement à gauche de q pour la préordonnance w .

Nous allons directement utiliser la forme (42) du critère pour confronter notre préordonnance $w(A)$ (cf. (38)), de la suite des niveaux de l'arbre des classifications (cf. figure 8). L'usage de $C(w, \pi)$ suppose implicitement une certaine définition de la fonction "rang" et une procédure de normalisation statistique qui sont parfaitement admissibles dans le cas où la préordonnance est assez fine (i.e. faible nombre d'ex-aequos); ce qui est le cas de la préordonnance (38). La suite des valeurs de (44) :

$$\{s(w, \pi_i) / 0 \leq i \leq 7\} \quad (45)$$

est :

$$\begin{aligned}
s(w, \pi_0) &= \underline{0} ; \\
s(w, \pi_1) &= \underline{27} ; \\
s(w, \pi_2) &= 26 + 26 = \underline{52} ; \\
s(w, \pi_3) &= 25 + 23 + 25 = \underline{73} ; \\
s(w, \pi_4) &= 23 + 22 + 16 + 22 + 23 = \underline{106} ; \\
s(w, \pi_5) &= 21 + 16 + 20 + 20 + 16 + 20 + 21 = \underline{134} ; \\
s(w, \pi_6) &= 15 + 11 + 15 + 12 + 15 + 12 + 11 + 8 \\
&\quad + 11 + 15 + 12 + 15 + 15 = \underline{167} ; \\
s(w, \pi_7) &= \underline{0}.
\end{aligned} \tag{46}$$

Pour $s(w, \pi) = 0$, on a :

$$C(w, \pi) = -\{3r(\pi)s(\pi)/[r(\pi) + s(\pi) + 1]\}^{1/2} \tag{47}$$

Par convention, on posera $C(w, \pi) = 0$ si le produit $r(\pi) \times s(\pi)$ est nul.

En introduisant le taux d'accroissement de $C(w, \pi)$, entre deux niveaux consécutifs de l'arbre des classifications :

$$\tau(w, \pi_i) = C(w, \pi_i) - C(w, \pi_{i-1}) \tag{48}$$

$1 \leq i \leq 7$, on obtient finalement et après calcul, le tableau 12 :

i	$C(w, \pi_i)$	$\tau(w, \pi_i)$
0	0.000	
1	1.671	1.671
2	2.139	0.468
3	2.637	0.498
4	2.909	0.272
5	3.210	0.301
6	3.202	-0.008
7	0.000	-3.202

Tableau 12 - Suite des valeurs des statistiques "globale" et "locale" des niveaux fondées sur la préordonnance.

Un niveau "significatif" correspond à un maximum local de la distribution

$$\{C(w, \pi_i)/0 \leq i \leq 7\} \tag{49}$$

des valeurs du critère C le long de la suite des niveaux de l'arbre des classifications. Il définit une partition qui représente un état d'équilibre dans la synthèse automatique. Dans l'exemple ci-dessus, il y a un seul niveau "significatif" ; c'est le 5ème. Toutefois, en général, il y a différents niveaux "significatifs", conduisant à des partitions plus ou moins fines. Le plus intéressant consiste en la détection d'un nœud "significatif". Ce dernier correspond à un maximum local de la distribution :

$$\{\tau(w, \pi_i) / 1 \leq i \leq 7\} \quad (50)$$

(cf. (48)) du taux d'accroissement de $C(w, \pi)$, le long de la suite des niveaux de l'arbre des classifications. Un nœud "significatif" correspond à l'achèvement d'une classe, à un niveau de synthèse donné. Dans l'exemple ci-dessus, il y a trois nœuds "significatifs" qui sont marqués dans la figure 8 par une étoile \star . Une classe comprenant suffisamment de nœuds "significatifs" est bien structurée pour la compréhension. Dans nos recherches les plus récentes [Lerman et Ghazzali 1991, Ghazzali 1992], nous avons -pour construire une partition intéressante- tendance à oublier la notion de niveau de l'arbre des classifications et retenir un ensemble de nœuds qui sont, soit "significatifs", soit liés par filiation directe à ces derniers; conformément à la relation de parenté induite par l'arbre.

Nous avons ci-dessus présenté la forme ordinale du critère d'adéquation d'une partition aux similarités entre éléments de l'ensemble E à classifier. Nous comparons ainsi, deux structures combinatoires de même type: préordres totaux sur $F = P_2(E)$. Cette forme du critère, qui nécessite un tri sur l'ensemble F des paires, est bien implantée dans certains des programmes élaborés dans notre environnement. Son comportement est comparé au critère dit CCC (Cubic Clustering Criterion) et conçu dans le contexte de la classification d'un nuage de points dans un espace euclidien [Mollière 1986].

Cependant, il existe une forme numérique du critère d'adéquation. Pour cette dernière, l'indice brut qui remplace $s(w, \pi)$ (cf. (44)), s'écrit:

$$S(Q, \pi) = \Sigma\{Q(q)/q \in R(\pi)\} \quad (51)$$

où $Q(q)$ est le coefficient de similarité statistiquement normalisé (indice brut centré et réduit) entre les deux composantes qu'on notera x et y de la paire $q = \{x, y\}$. Ainsi, le tableau 5 ci-dessus donne un exemple de la matrice Q . Très précisément dans CHAVL, pour le cas de la comparaison d'attributs booléens, la normalisation statistique se fait par rapport au modèle Poissonien (nous y reviendrons). Très précisément, la formule (21) s'écrit dans ce cas, avec des notations que l'on comprend:

$$Q(a^j, a^k) = \frac{s(a^j, a^k) - np(a^j)p(a^k)}{\sqrt{np(a^j)p(a^k)}} \quad (52)$$

$$1 \leq j < k \leq 8.$$

Signalons déjà que lorsqu'il s'agira d'établir un indice de similarité sur l'ensemble O des objets ou C des concepts (cf. §III.6), le rôle joué par Q pour l'élaboration du critère d'adéquation, sera joué par l'indice que nous noterons S et que nous appellerons "somme des contributions normalisées".

L'indice brut (51) entre la structure numérique de la similarité

$$\{Q(u)/u \in P_j(E)\} \quad (53)$$

et la partition π , est centré et réduit par rapport à l'hypothèse d'absence de liaison où à π , on associe une partition aléatoire π^* . Cette dernière est prise dans l'ensemble des partitions de même type que π ; c'est à dire, dont la suite des cardinaux des classes est fixée de la même manière que pour π [Lerman 1981, 1983, 1992_a, 1992_b, Ouali-Allah 1991].

Introduisons la moyenne α et la variance λ^2 de la table des valeurs (53). D'autre part, codons E au moyen de l'ensemble $J = \{1, 2, \dots, i, \dots, m\}$ [$m = \text{card}(E)$]. Maintenant, substituons à la table (53), la table des indices "globalement normalisés":

$$\{c(i, j)/1 \leq i < j \leq m\}, \quad (54)$$

où

$$[V(i, j), 1 \leq i < j \leq m], c(i, j) = [Q(i, j) - \alpha]/\lambda \quad (55)$$

Dans ces conditions, le critère d'adéquation se met sous la forme

$$\frac{1}{\sqrt{V}} \Sigma \{c(i, j)/\{i, j\} \in R(\pi)\} \quad (56)$$

où V indique la variance du critère aléatoire:

$$\Sigma \{c(i, j)/\{i, j\} \in R(\pi^*)\} \quad (57)$$

Très précisément, en désignant par $(m_g; 1 \leq g \leq h)$ et par $(\pi_g = m_g/m; 1 \leq g \leq h)$ la suite des proportions, la forme calcul de V est la suivante:

$$\begin{aligned} V &= \binom{m}{2} \left(\sum_{1 \leq g \leq h} \pi_g^2 \right) \left[1 + \left(\sum_{1 \leq g \leq h} \pi_g^2 \right) \right] \\ &+ \left\{ \sum_{1 \leq i \leq m} \left[\sum_{1 \leq j \leq m} c(i, j) \right]^2 - m(m-1) \right\} \\ &\times \left[\left(\sum_{1 \leq g \leq h} \pi_g^3 \right) + \left(\sum_{1 \leq g \leq h} \pi_g^2 \right)^2 \right] \quad (58) \end{aligned}$$

où on a posé $c(j, j) = 0$ pour tout j de J et où $c(i, j) = c(j, i)$ pour tout (i, j) de $J \times J$. $f = \binom{m}{2} = m(m-1)/2$.

Cette forme de la variance est plus précise que dans le cas ordinal. Elle peut être utilisée dans ce dernier cas, à partir de (56) à condition de remplacer la table (53) par celle, de la même forme que (39).

C'est l'expression (56) du critère, obtenue à partir de la table (53), qui se trouve implantée dans CHAVL. Le tableau (14) ci-dessous donne la suite des valeurs du critère

(statistique globale) et de son taux d'accroissement (statistique locale) sur la suite croissante des niveaux. L'arbre concerné est ici directement obtenu à partir des expressions (52), (60), (61) et (63), il diffère très légèrement de celui de la figure 8 ci-dessus. Seul l'ordre des nœuds des niveaux 5 et 6 se trouvent intervertis. Le niveau 6 définit la même partition dans les deux cas.

niveau	statistique globale	statistique locale
1	2.0737	2.0737
2	2.6454	0.5717
3	2.7014	0.0560
4	2.9828	0.2815
5	3.1609	0.1780
6	3.1493	-0.0116
7	-0.0811	-3.2304

Tableau 13 - Suite des valeurs des statistiques "globale" et "locale" des niveaux fondées sur la similarité Q .

Ici, la partition la plus "significative" est $\pi'_5 = \{\{1, 2, 4, 5\}, \{6\}, \{3, 7, 8\}\}$. Mais le calcul précédent est plus précis parce que, exact. Ici (cf. (63)) on passe par l'approximation normale de la loi dans l'h.a.l. de l'indice aléatoire; alors que le nombre d'individus est faible ($n = 10$). Néanmoins, la quasi-identité des résultats est remarquable.

3.4 Aide à l'interprétation: degré de "neutralité" des éléments de l'ensemble à classifier.

Un aspect d'une validation intrinsèque des résultats de la classification concerne la détermination des éléments plus ou moins "moteurs" (ou à l'opposé "neutres") dans l'entraînement d'une classe. A cet égard, nous proposons de mesurer le degré de "neutralité" d'un même élément α de l'ensemble (ici A) à classifier par la petitesse de la variance $V(\alpha)$ de ses proximités aux autres éléments de A . p désignant le cardinal de A , on a :

$$V(\alpha) = \Sigma\{[Q(\alpha, b) - Q(\alpha)]^2/b \in A - \{\alpha\}\}$$

où

$$Q(\alpha) = \frac{1}{p-1} \Sigma\{Q(\alpha, b)/b \in A - \{\alpha\}\} \quad (59)$$

D'où le tableau 14, obtenu dans le cadre de notre exemple, sur la base du tableau 5. La dernière colonne indique les rangs des différents éléments de A , conformément à la

suite décroissante des valeurs de V .

j	$V(a^j)$	$r(a^j)$
1	0.8415	3
2	0.3414	7
3	0.3556	6
4	0.6037	4
5	0.9522	2
6	0.1483	8
7	0.4105	5
8	1.0109	1

Tableau 14 - Valeurs de la variance des proximités $V(\alpha)$ et rangs respectivement associés.

Signalons que l'analyse de la variance des proximités (de chacun des éléments aux autres éléments de l'ensemble à organiser) a conduit à une riche famille de méthodes de "sériation" et de "classification" [Lerman 1972, 1981, chap. 8; Leredde 1979]. Le programme MPATD de la bibliothèque MODULAD constitue une des retombées de cette recherche.

3.5 Comment atteindre une échelle discriminante de probabilité, lorsque l'ensemble O des objets devient "gros".

Dans ce cas en effet [pour $n = \text{card}(O)$ tendant vers l'infini], on peut montrer que la formule (22), conduit soit à la valeur 0 (resp. 1), en cas d'un lien négatif (resp. positif) et ce, quelle que soit l'intensité de ce lien. Une analyse statistique fine [Lerman 1984_a, Daudé 1992], permet de valider -dans l'hypothèse d'absence de liaison schématisée dans (24)- la substitution des indices dits "locaux"

$$\{Q_l(j, k)/1 \leq j < k \leq p = \text{card}(A)\} \quad (60)$$

correspondants à (20) ci-dessus, par ceux dits "globaux":

$$\{Q_g(j, k)/1 \leq j < k \leq p\} \quad (61)$$

avec

$$Q_g(j, k) = \frac{Q_l(j, k) - \text{moy}_e(Q_l)}{\sqrt{\text{var}_e(Q_l)}} \quad (62)$$

où $\text{moy}_e(Q_l)$ et $\text{var}_e(Q_l)$ sont la moyenne et la variance empiriques de table (60) des valeurs de Q_l .

Dans ces conditions, l'indice probabilité devient :

$$P(a^j, a^k) = \phi[Q_g(a^j, a^k)] \quad (63)$$

$1 \leq j < k \leq p$, où ϕ désigne toujours la fonction de répartition de la loi normale centrée et réduite.

De cette façon, l'échelle de probabilité est assez fine pour l'évaluation relative des liaisons observées entre attributs descriptifs de l'ensemble O des objets. Soulignons que de patients travaux ont permis d'étendre et de valider l'approche pour les types les plus généraux de données que nous présenterons au § 1 de la seconde partie.

Nous allons maintenant, dans la situation la plus simple et en nous basant sur l'exemple, donner la démarche générale pour l'élaboration d'un indice de similarité probabiliste sur l'ensemble O (resp. C) des objets (resp. concepts).

3.6 Indice de similarité probabiliste sur l'ensemble O des objets, décrits par des attributs booléens (cas de l'exemple du tableau 1).

(i) Pour tout $j, 1 \leq j \leq p$ ($p = 8$ dans notre exemple), définir un indice "brut" de similarité $s_j(o_i, o_{i'})$ entre les deux objets o_i et $o_{i'}$ à associer.

Proposition d'un indice s_j

Remplacer le tableau d'incidence 1 :

$$\{\varepsilon_i^j / 1 \leq i \leq n, 1 \leq j \leq p\}$$

par

$$\{\eta_i^j / 1 \leq i \leq n, 1 \leq j \leq p\}$$

où

$$\eta_i^j = \varepsilon_i^j / \left(\sum_{1 \leq k \leq p} \varepsilon_i^k \right)^{1/2}$$

On pose

$$s_j(o_i, o_{i'}) = \frac{1}{p} - \frac{1}{2} (\eta_i^j - \eta_{i'}^j)^2 \quad (64)$$

Ainsi,

$$s_1(o_1, o_2) = \frac{1}{8} - \frac{1}{2} \left(\frac{0}{\sqrt{4}} - \frac{0}{\sqrt{3}} \right)^2 = 0.125$$

$$s_4(o_1, o_2) = \frac{1}{8} - \frac{1}{2} \left(\frac{1}{\sqrt{4}} - \frac{1}{\sqrt{3}} \right)^2 = 0.122$$

$$s_7(o_1, o_2) = \frac{1}{8} - \frac{1}{2} \left(\frac{0}{\sqrt{4}} - \frac{1}{\sqrt{3}} \right)^2 = -0.042$$

$$s_3(o_1, o_2) = \frac{1}{8} - \frac{1}{2} \left(\frac{1}{\sqrt{4}} - \frac{0}{\sqrt{3}} \right)^2 = 0.000$$

C'est au niveau de la définition de l'indice brut $s_j(o_i, o_{i'})$ [resp. $s_j(c_i, c_{i'})$] entre deux objets (resp. concepts), qu'il y a lieu de tenir le plus intimement compte de la structure mathématico-logique du tableau des données (cf. figure 2).

(ii) Normaliser statistiquement $s_j(o_i, o_{i'})$ par rapport à la distribution empirique

$$\{s_j(o_i, o_{i'})/1 \leq i, i' \leq n\} \quad (65)$$

pour obtenir :

$$s_j(o_i, o_{i'}) = \frac{s_j(o_i, o_{i'}) - m_e(s_j)}{\sqrt{\text{var}_e(s_j)}} \quad (66)$$

où $m_e(s_j)$ et $\text{var}_e(s_j)$ sont la moyenne et la variance de la précédente distribution (65).

La moyenne et la variance empiriques $m_e(s_j)$ et $\text{var}_e(s_j)$ portent sur n^2 valeurs de $s_j(o_i, o_{i'})$, qui, dans notre exemple, sont équipondérées. Cela ne sera pas le cas, si le poids d'un même objet (resp. concept) n'est pas invariable, quel que soit l'objet (resp. le concept). $S_j(o_i, o_{i'})$ est ce que nous appelons la "contribution normalisée de la jème variable à la comparaison des deux objets".

(iii)

$$S(o_i, o_{i'}) = \sum_{1 \leq j \leq p} S_j(o_i, o_{i'}) \quad (67)$$

est la somme des contributions normalisées.

(iv) Nouvelle normalisation globale par rapport à la distribution empirique

$$\{S(x, y)/\{x, y\} \in P_2(O) [\text{resp. } P_2(C)]\} \quad (68)$$

où $P_2(O)$ [resp. $P_2(C)$] désigne l'ensemble des paires d'éléments de O (resp. C), pour obtenir :

$$S(x, y) \rightarrow Q_g(x, y) = \frac{S(x, y) - m_e(S)}{[\text{var}_e(S)]^{1/2}} \quad (69)$$

où $m_e(S)$ et $\text{var}_e(S)$ sont la moyenne et la variance de la précédente distribution.

On peut signaler que l'indice $Q_g(x, y)$ est analogue à celui $c(i, j)$ défini en (55).

(v) Calcul de l'indice probaliliste

$$P(o_i, o_{i'}) = \text{Prob}\{Q_g(x^*, y^*) \leq Q_g(o_i, o_{i'})\} \quad (70)$$

Cette probabilité est calculée dans l'hypothèse d'absence de liaison schématisée dans (24). Pourvu que le nombre de variables : p , soit assez grand, l'approximation suivante est excellente :

$$P(o_i, o_{i'}) \cong \phi[Q_g(o_i, o_{i'})] \quad (71)$$

$1 \leq i \leq i' \leq n$ où ϕ est la fonction de répartition de la loi normale $N(0, 1)$.

En tenant compte de l'extension de cet indice pour tous les types de données (cf. § 1, seconde partie), on se retrouve devant le schéma de la figure 6.

Dans ces conditions, on obtient les sorties suivantes du programme CHAVL, pour la classification de l'ensemble O défini dans la table d'incidence des données du tableau 1.

Programme de classification hiérarchique CHAVL

Titre de l'étude : Essai incidence

Étape SIMOB

Rappel des paramètres

nombre d'individus	10
nombre de variables	8
attributs de description	
format des données	(3 × 8 i 2)

Fin de l'étape SIMOB

Étape AVLVR

Représentation polonaire de l'arbre

-9 -7 1 -6 -3 -1 2 8 4 10 -8 -5 -4 3 -2 6
9 7 5 0

Fin de l'étape AVLVR

Étape INTRP

Rangement des éléments par valeurs de dispersions croissantes

élément 10:	o10	dispersion : 0.36271
élément 7:	o7	dispersion : 0.43063
élément 1:	o1	dispersion : 0.45478
élément 5:	o5	dispersion : 0.62854
élément 9:	o9	dispersion : 1.04031
élément 6:	o6	dispersion : 1.06048
élément 2:	o2	dispersion : 1.09864
élément 3:	o3	dispersion : 1.10182
élément 8:	o8	dispersion : 1.11017
élément 4:	o4	dispersion : 1.42690

Statistiques des niveaux.

	niveau	statistique globale	statistique locale
	1	1.8040	1.8040
	2	2.4660	0.6619
1 maximum	3	3.2918	0.8258
	4	3.8923	0.6005
	5	4.02541	0.1317
2 maximum	6	4.3176	0.2935
	7	4.5721	0.2546
3 maximum	8	4.8333	0.2612
	9	-0.1124	-4.9457

Fin de l'étape INTRP

Étape DESAB

Paramètres de l'étape

nombre d'éléments	10
nombre de niveaux réels de l'arbre	9
nombre de niveaux représentés	9
un niveau sur 1 est représenté	
longueur d'une ligne du dessin	80

Fin de l'étape DESAB
Fin du programme CHAVL

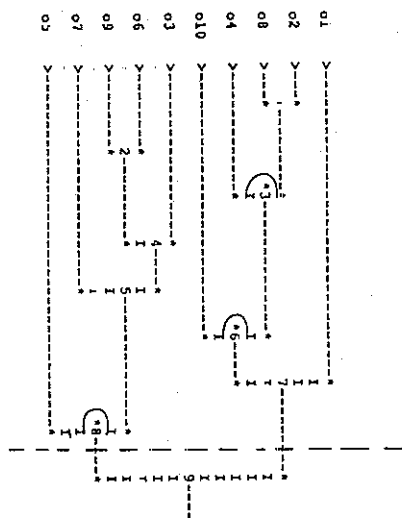


FIG. 9 - Arbre des classifications sur 0.

Références

- [Arabie et Hubert 92] Arabie (P.) et Hubert (L.J.). – Combinatorial data analysis. *Annu. Rev. Psychol.*, n° 43, 1992, pp. 169–203.
- [Barthélémy et Guénoche 88] Barthélémy (J.P.) et Guénoche (A.). – *Les arbres et les représentations des proximités*. – Paris, Masson, 1988.
- [Bruynooghe 89] Bruynooghe (M.). – *Nouveaux algorithmes en classification automatique applicables aux très gros ensembles de données, rencontrés en traitement d'images et en reconnaissance des formes*. – Thèse, Université de Paris VI, 1989.
- [Caillez 83] Caillez (F.). – The analytical solution of the additive constant problem. *Psychometrika*, vol. 48, n° 2, 1983, pp. 305–308.
- [Daude 92] Daude (F.). – *Analyse et justification de la notion de ressemblance entre variables qualitatives dans l'optique de la classification hiérarchique par AVL*. – Thèse, université de Rennes 1, 1992.
- [Ghazzali 92] Ghazzali (N.). – *Comparaison et réduction d'arbres de classification, en relation avec des problèmes de quantification en imagerie numérique*. – Thèse, Université de Rennes I, 1992.
- [Giakoumakis et Monjardet 87] Giakoumakis (V.) et Monjardet (B.). – Coefficients d'accord entre deux préordres totaux. *Rev. Statistique et Analyse des Données*, n° 12, 1987, pp. 46–99.
- [Gras et Larher 92] Gras (R.) et Larher (A.). – L'implication statistique, nouvelle méthode d'analyse des données. *Rev. Math., Inf. Sci. Hum.*, n° 120, 1992.
- [Guenoche 87] Guénoche (A.). – Méthodes combinatoires de sériation à partir d'une dissimilarité. Inria, Cinquièmes Journées Internationales "Analyse des Données Informatique", 29 sept-2 oct. 1987 – Versailles, Inria, North-Holland, 1987.
- [Jaccard 08] Jaccard (P.). – Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.*, n° 44, 1908, pp. 223–2790.
- [Leredde 79] Leredde (H.). – *La méthode des pôles d'attraction; la méthode des pôles d'agrégation: deux nouvelles familles d'algorithmes en classification automatique et sériation*. – Thèse, Université Paris VI, 1979.
- [Lerman 72] Lerman (I.C.). – Analyse du phénomène de la "sériation". *Rev. Math. Sci. Hum.*, n° 38, 1972, pp. 39–57.
- [Lerman 81] Lerman (I.C.). – *Classification et analyse ordinale des données*. – Paris, Dunod, 1981.
- [Lerman 83a] Lerman (I.C.). – Sur la signification des classes issues d'une classification automatique. *Numerical Taxonomy, NATO ASI Series vo. G1*, éd. par Felsenstein (J.), pp. 179–198. – Springer Verlag, 1983.

- [Lerman 83b] Lerman (I.C.). – Association entre variables qualitatives ordinales nettes ou flowes. *Rev. Statistique et Analyse des Données*, vol. 8, n° 7, 1983, pp. 41-73.
- [Lerman 83c] Lerman (I.C.). – Indices d'association partielle entre variables qualitatives nominales. *Rairo/Oper. Res.*, n° 17, 1983, pp. 213-259.
- [Lerman 83d] Lerman (I.C.). – Indices d'association partielle entre variables qualitatives ordinales. *Publ. Inst. Stat. univ. Paris*, n° 28, 1983, pp. 7-46.
- [Lerman 84a] Lerman (I.C.). – Justification et validité statistique d'une échelle $[0,1]$ de fréquence mathématique pour une structure de proximité sur un ensemble de variables observées. *Publ. Inst. Stat. univ. Paris*, n° 29, 1984, pp. 27-57.
- [Lerman 84b] Lerman (I.C.). – Analyse classificatoire d'une correspondance multiple, typologie et régression. *Data Analysis and Informatics, III*, E. Diday. North Holland, 1984, pp. 193-221. –
- [Lerman 86] Lerman (I.C.). – Rôle de l'inférence statistique dans une approche de l'analyse classificatoire des données. *Journal de la société de statistique de Paris*, n° 4, 1986, pp. 238-252.
- [Lerman 87a] Lerman (I.C.). – Construction d'un indice de similarité entre objets décrits par des variables d'un type quelconque. application au problème du consensus en classification. *Rev. Stat. Appl.*, n° 35, 1987, pp. 39-60.
- [Lerman 87b] Lerman (I.C.). – Maximisation de l'association entre deux variables qualitatives ordinales. *Rev. Math. Sci. Hum.*, n° 100, 1987, pp. 49-56.
- [Lerman 89] Lerman (I. C.). – Formules de réactualisation en cas d'agrégations multiples. *Rairo série R.O.*, vol. 23, n° 2, 1989, pp. 151-163.
- [Lerman 91] Lerman (I.C.). – Foundations of the likelihood linkage analysis (lla) classification method. *Applied Stochastic Models and Data Analysis John Wiley*, vol. 7, 1991, pp. 69-76.
- [Lerman 92a] Lerman (I.C.). – Conception et analyse de la forme limite d'une famille de coefficients statistiques d'association entre variables relationnelles I. *Rev. Math. Infor. & Sci. Hum.*, 30e année, Paris, n° 118, 1992, pp. 35-52.
- [Lerman 92b] Lerman (I.C.). – Conception et analyse de la forme limite d'une famille de coefficients statistiques d'association entre variables relationnelles II. *Rev. Math. Infor. & Sci. Hum.*, 30e année, Paris, n° 119, 1992, pp. 75-100.
- [Lerman 93] Lerman (I.C.). – *Likelihood linkage analysis (LLA) classification method: An example treated by hand.*, Biochimie, Elsevier editions, 1993, volume 75, pp. 379-397. –
- [Lerman, Gras et Rostam 81] Lerman (I.C.), Gras (R.) et Rostam (H.). – Elaboration et évaluation d'un indice d'implication pour des données binaires, I et II. *Rev. Mat. Sci. Hum.*, n° 74 et 75, 1981, pp. 5-35 et 5-47.

- [Lerman, Hardouin et Chantrel 80a] Lerman (I.C.), Hardouin (M.) et Chantrel (T.). – Analyse de la situation relative entre deux classifications floues. *Data Analysis and Informatics*. Secondes Journées Internationales Analyse des Données et Informatique, Versailles 17-19 octobre 1979, North Holland, 1980, pp. 523–552. –
- [Lerman et Tallur 80b] Lerman (I.C.) et Tallur (B.). – Classification des éléments constitutifs d'une juxtaposition de tableaux de contingence. *Rev. de Statistique Appliquée*, n° 28, 3, Paris 1980, pp. 5–28.
- [Lerman et Peter 84] Lerman (I.C.) et Peter (P.). – *Analyse d'un algorithme de classification hiérarchique "en parallèle" pour le traitement de gros ensembles*. – rapport de recherche n° 339, Inria, Le Chesnay 1984.
- [Lerman et Peter 85] Lerman (I.C.) et Peter (P.). – *Elaboration et logiciel d'un indice de similarité entre objets d'un type quelconque. Application au problème du consensus en classification*. – publication interne, 72 pages n° 262, IRISA, Rennes, juillet 1985.
- [Lerman et Peter 88] Lerman (I.C.) et Peter (P.). – Structure maximale pour la somme des carrés d'une contingence aux marges fixées ; une solution algorithmique programmée. *Rairo/Oper. Res.* 22, 1988, pp. 83–136.
- [Lerman et Peter 89] Lerman (I.C.) et Peter (P.). – Classification of concepts described by taxonomic preordination variables with multiple choice. application to the structuration of a species set of phlebotomine. *Data Analysis, Learning Symbolic and Numeric Knowledge*, Diday Ed. Inria. Proceed. of the Conf. Antibes, sept. 11-14 1989, pp. 73–86. – Nova Science Publishers Inc., New-York, 1989.
- [Lerman et Ghazzali 91] Lerman (I.C.) et Ghazzali (N.). – What do we retain from a classification tree? an experiment in image coding. *Symbolic-Numeric Data Analysis and Learning*, E. Diday et Y. Lechevallier, Inria, Versailles, sept. 18-20, 1991, pp. 27–42. –
- [Marcotorchino 91] Marcotorchino (F.). – Seriation problems: an overview. *Applied Stochastic Models and Data Analysis*, John Wiley, n° 7, 1991, pp. 139–151.
- [Matusita 67] Matusita (K.). – On the notion of affinity of several distributions and some of its applications. *Ann. Math. Stat.*, vol. 19, n° 2, 1967, pp. 181–192.
- [Messatfa 90] Messatfa (H.). – *Unification relationnelle des critères et structures optimales des tables de contingence*. – Thèse, Université de Paris VI, 1990.
- [Mollière 86] Mollière (J.L.). – What's the real number of clusters? *Classification as a Tool of Research*, éd. par Gaul (W.) et Schader (M.). – North Holland, 1986.
- [Moreau 85] Moreau (A.). – *Elaboration et calcul d'indices d'association entre variables qualitatives "nettes" ou "floues"*. Application à une forme d'interprétation d'une classification de paramètres épidémiologiques – Thèse, Université de Rennes I, 1985.

- [Nicolau 80] Nicolau (F.). – *Crítérios de análise classificatoria hierárquica baseados na furgao de distribuigao*. – Thèse, Faculté des Sciences de Lisbonne, 1980.
- [Nicolau 88] Nicolau (H. Bacelar). – Two probabilistic models for classification of variables in frequency tables. The First Conference of the IFCS, Technical Univ. of Aachen, FRG 29 June-1 July 1987, 1988, pp. 181-186. –
- [Ouali-Allah 91] Ouali-Allah (M.). – *Analyse en préordonnances des données qualitatives. Applications aux données numériques et symboliques*. – Thèse, Université de Rennes I, 1991.
- [Peter 87] Peter (P.). – *Méthodes de classification hiérarchique et problèmes de structuration et de recherche d'informations assistées par ordinateur*. – Thèse, Université de Rennes I, 1987.
- [Tallur 88] Tallur (B.). – *Contribution l'analyse exploratoire de tableaux de contingence par la classification*. – Thèse, Université de Rennes I, 1988.
- [Tarski 54] Tarski (A.). – Contribution to the theory of models, I et II. *Indagationes Mathematicae*, n° 16, 1954, pp. 572-588.