

# STATISTIQUE ET ANALYSE DES DONNÉES

SOCIÉTÉ FRANÇAISE DE CLASSIFICATION

**Résumés - Journées de Statistique, Paris, mai 1979**

*Statistique et analyse des données*, tome 4, n° 3 (1979), p. 79-84

[http://www.numdam.org/item?id=SAD\\_1979\\_\\_4\\_3\\_79\\_0](http://www.numdam.org/item?id=SAD_1979__4_3_79_0)

© Association pour la statistique et ses utilisations, 1979, tous droits réservés.

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

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

SOCIETE FRANÇAISE  
DE CLASSIFICATION

(Résumés - Journées de Statistique, Paris, MAI 1979)



ETUDE DISTRIBUTIONNELLE DE QUELQUES INDICES DE  
SIMILARITE POUR DES DONNEES BINAIRES

Auteur : H. BACELAR NICOLAU  
Centre de Statistique et Applications (C.E.A.- I.N.I.C.)  
Faculté des Sciences de Lisbonne  
58, rua da Escola Politécnica  
Lisboa - PORTUGAL

Dans ce travail nous faisons une étude comparative du point de vue distributionnel des coefficients le plus couramment utilisés pour les données binaires.

Nous démontrons que, sous une certaine hypothèse que l'on précise, ils se partagent en deux classes : l'une contenant ceux qui ont la même loi exacte (hypergéométrique), l'autre rassemblant les coefficients qui ne sont équivalents qu'en loi limite (normale).

Nous proposons alors comme coefficient général de comparaison pour ce type de données un coefficient probabiliste basé sur la valeur de la fonction de répartition normale limite commune à tous les coefficients. Ce coefficient généralise la notion de similarité selon la "vraisemblance du Lien" introduite par I.C. Lerman.

UN MODELE DE MORPHOGENESE FONDE SUR L' ALGORITHME  
DES NUEES DYNAMIQUES

Auteur : B. CRAIGNOU  
IRIA-Laboria  
Rocquencourt  
B.P. 105  
78150 LE CHESNAV  
Tél 9549020 poste 668

Modélisation d'une forme naturelle. Un algorithme de morphogénèse décrit l'évolution de la forme des feuilles dès son initialisation. Les cellules sont les unités morphologiques du modèle.

La première partie de l'algorithme décrit la croissance de chacune des cellules de la feuille (alimentation, grossissement, division, état stationnaire).

La seconde partie décrit la croissance globale (rangement, lissage).

Le mouvement de l'ensemble des cellules (conséquence de leur grossissement) pose des problèmes dans le déroulement de l'algorithme. Une application dans le cas continu de l'algorithme des Nuées Dynamiques permet l'organisation de ce mouvement et la détermination postérieure de la forme des cellules.

UNE NOUVELLE CLASSE DE METHODES D'AGREGATION EN CLASSIFICATION  
AUTOMATIQUE BASEE SUR LA VRAISSEMBLANCE DU LIEN.

Auteur : F. COSTA NICOLAU

Centre de Statistique et Applications (C.E.A.-I.N.I.C.)  
Faculté des Sciences de Lisbonne  
58, Rua da Escola Politécnica  
Lisboa - Portugal

Nous proposons l'exploitation systématique de la notion de Vraisemblance du Lien (I.C.Lerman - méthode A.V.L., basée sur le lien maximal entre deux classes), et dont des résultats récents montrent l'aspect global, unificateur et sans biais pour le cas des données binaires (H. Bacelar Nicolau).

Nous y traitons surtout la méthode AVD (agrégation par la Vraisemblance du Lien minimal entre deux classes), les méthodes AMVLD (agrégation par les moyennes - arithmétique, géométrique, harmonique, et cas limites- des Vraisemblances des Liens maximal et minimal ), et la méthode AVM ( agrégation par la Vraisemblance du Lien moyen entre deux classes ) .

CLASSIFICATION HIERARCHIQUE ET CONNEXITE

Auteur : M. GONDRAN M.  
EDF, Direction des Etudes et Recherches  
1, Avenue du Général de Gaulle  
92141 CLAMART CEDEX

On présente ici une nouvelle classification hiérarchique basée sur des données de similarité.

Cette classification hiérarchique est robuste (pas d'effet de chaîne) et optimale. Elle s'appuie sur des résultats de la théorie des graphes : la liaison entre les différents flots d'un graphe symétrique donnée par le théorème de Gomory-Hu.

Dans le cas où les données de similarité sont du type "tout ou rien", les classes de niveau k de la classification hiérarchique correspondent aux composantes k-arêtes-connexes du graphe des similarités.

## VALEURS PROPRES ET VECTEURS PROPRES

### EN ANALYSE DES PREFERENCES

Auteur : M. GONDRAN M.  
EDF, Direction des Etudes et Recherches  
1, Avenue du Général de Gaulle  
92141 CLAMART CEDEX

Trois méthodes d'analyse des préférences sont présentées. Elles consistent à déterminer les valeurs propres et les vecteurs propres de la matrice des préférences par couple dans une structure algébrique bien choisie : les dioïdes. Si le dioïde est  $(\mathbb{R}^+, +, \times)$  on trouve un ordre moyen ; si le dioïde est  $(\mathbb{R}^+, \max, \times)$  on trouve la classe des éléments les plus liés entre eux ; si le dioïde est  $(\mathbb{R}, \max, \min)$  on trouve une classification hiérarchique de préordres.

## CROISEMENT DE CLASSIFICATIONS FLOUES

LERMAN I.C.  
Laboratoire de Statistique  
I.R.I.S.A. - Université de Rennes  
35042 - RENNES Cédex

On montre que la donnée d'une classification sur un ensemble  $A$  d'attributs descriptifs permet, en introduisant un degré d'appartenance relative d'un individu donné à une même classe d'attributs, la définition et la mise en évidence d'une classification "floue" sur l'ensemble  $E$  des individus décrits. On développe dans ces conditions l'analyse du croisement de deux classifications floues par la construction d'indices affectés aux cases du croisement conformément à la statistique du  $\chi^2$  et par la généralisation adéquate de la théorie classique dans cette nouvelle situation. Ce qui permet de situer, l'une par rapport à l'autre, deux typologies sur le même ensemble  $E$  d'individus, respectivement définies par une classification monothétique et une classification polythétique ou bien par deux classifications polythétiques à partir de deux ensembles disjoints de traits descriptifs. Une telle approche est particulièrement pertinente dans l'analyse des questionnaires pouvant être scindés en deux parties ; la première comprenant les attributs de description de la structure du sujet et la seconde, les attributs d'attitude ou de comportement.

UN PROBLEME DE LA CLASSIFICATION AUTOMATIQUE POSE PAR  
L'INTERROGATION DES RASSEMBLEMENTS DOCUMENTAIRES

Auteur: F. MURTAGH,  
Laboratoire de Statistiques Mathématiques,  
Université de Paris VI.

Nous proposons un traitement rigoureux de la classification automatique appliquée à l'interrogation des rassemblements documentaires. En particulier on expose une nouvelle approche à l'analyse probabiliste des résultats d'une classification ascendante hiérarchique. Quelques résultats nécessaires à la mise au point de cette théorie sont fournis. Bien que cette étude s'applique à - et s'explique par - l'exemple de l'interrogation des systèmes documentaires, on en suggère la généralisation aux autres applications de la classification automatique.

APPLICATION DES METHODES D'ANALYSE STATISTIQUE MULTIDIMENSIONNELLE  
A L'ETUDE DE LA CHARGE D'UN ORDINATEUR

Auteur : M. RALAMBONDRAINY H.,  
LABORIA - BP 105 -78150 LE CHESNAY

RESUME : Alors qu'on envisage à l'IRIA une extension et des changements de configuration, il est nécessaire d'avoir une vue d'ensemble de la charge de l'IRIS 80 du centre de calcul. C'est le but de cette étude.

LES DONNEES : Le fichier de comptabilité de SIRIS 8 recense les différentes consommations machines des divers travaux soumis à l'ordinateur. Pour donner un ordre de grandeur du volume d'informations mis en jeu, signalons qu'il a été traité au mois de mars 1977 à peu près 13000 travaux batch. Les données portent sur une période de dix mois : janvier à octobre 1977. A partir du fichier comptable ont été extraits divers fichiers sur lesquels ont porté nos analyses : fichier utilisateurs, fichier travaux, fichier sessions temps partage, etc,...

LES ANALYSES EFFECTUEES :

- . Etude des consommations mensuelles globales : définition de mois types.
- . Typologie des utilisateurs pendant les mois types.
- . Suivi des utilisateurs sur la période de mesure.
- . Etude de la charge journalière : définition de jours types.
- . Typologie des travaux batch dans les différents jours types.
- . Typologie des sessions temps partagé.

METHODES MISES EN OEUVRE :

- . Analyse factorielle des correspondances.
- . Analyse en composantes principales.
- . Classification hiérarchique ascendant.
- . Nuées dynamiques.