

STATISTIQUE ET ANALYSE DES DONNÉES

Résumés des Journées de Toulouse (Mai 1980)

Statistique et analyse des données, tome 5, n° 2 (1980), p. 65-114

http://www.numdam.org/item?id=SAD_1980__5_2_65_0

© Association pour la statistique et ses utilisations, 1980, 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/>

RESUMES
DES
JOURNEES DE TOULOUSE
(Mai 1980)

SUR LA DETECTION DES VALEURS ABERRANTES DANS LE MODELE LINEAIRE -
EFFETS FIXES OU ALEATOIRES

ESTIMATION D'UNE PARTITION DU CUBE UNITAIRE A n-DIMENSIONS
A L'AIDE D'UNE PROCEDURE D'APPRENTISSAGE

Auteur : ABBADY Fouad
Laboratoire de Statistique et Probabilités
Université Paul Sabatier
31077 TOULOUSE Cedex

J. AGUILAR MARTIN

M. BALSSA

Considérons d'abord le modèle linéaire à effets fixes :

$$X = A\beta + \varepsilon$$

où A est une matrice donnée ($n \times p$), β le vecteur ($p \times 1$) des paramètres et ε un vecteur aléatoire ($n \times 1$) gaussien de coordonnées indépendantes de même variance, et *entrées sous l'hypothèse nulle.*

Les hypothèses alternatives assignent des moyennes différentes de zéro à certaines coordonnées de ε , dites aberrantes (outliers).

Ce problème a suscité un intérêt considérable ces dernières années : voir entre autres [2].

Pour la détection d'une seule valeur aberrante, une procédure bayésienne optimale dans la classe des tests invariants a été présentée par FERGUSON [1]. Nous généralisons cette étude au cas de plusieurs valeurs aberrantes.

Nous étudions d'autre part des problèmes analogues concernant les modèles à effets aléatoires ou mixtes :

$$X = A\beta + \sum_{i=1}^m A_i \xi_i + \varepsilon$$

(détection de valeurs aberrantes pour l'un quelconque des vecteurs aléatoires ξ_i ou ε).

Références : [1] FERGUSON T.S. (1961) - On the rejection of outliers, Proc. Fourth Berkeley Symp. Math. Stat. Prob. 1, pp.253-287, Univ. of California.

[2] JOHN J.A. (1979) - Outliers in factorial experiments, Applied Stat. 27 n°2, p.111-119.

Laboratoire d'Automatique et d'Analyse des Systèmes du CNRS
7, Avenue du Colonel Roche - 31400 TOULOUSE

Dans cet article nous analysons en détail le fonctionnement d'un algorithme de classification avec auto-apprentissage. Cet algorithme accepte des données multidimensionnelles dont les n composantes sont positives et inférieures à un. La moyenne des éléments attribués à une classe suffit à la représenter grâce à l'utilisation d'une loi de probabilité exponentielle dont le support est le cube unitaire. Le calcul de l'appartenance d'un élément à une classe à l'aide d'un critère de décision bayésien peut être simplifié en adoptant une approximation dont la justification est décrite dans l'article. Le nombre de classes étant inconnu a priori, nous soulignons l'importance de la période initiale de l'algorithme et du mécanisme de création et initialisation des classes. Plusieurs exemples de son fonctionnement illustrent les points précédents.

67

Publications ayant une certaine parenté avec ce travail :

- E. DIDAY : "Apprentissage et Mesures de Ressemblances Adaptatives"
Computer Oriented Learning Processes - NATO, 1976 (J.C. SIMON)
- ... et d'autres publications sur "Nuées Dynamiques"
- ANDERBERG M.R. : "Cluster Analysis for applications". Academic Press, 1973.
- ... et d'autres publications sur "Clustering".

PREMIERE VERSION EXPERIMENTALE DU SYSTEME LAPUMS

Auteur : J.P. BANQUET, LENA, La Salpêtrière
47, bd. de l'Hôpital, 75651 - PARIS CEDEX 13.

Auteurs : J. BADIA, O. CADET, A. CROQUETTE,
D. GRAU, J. JOCKIN, J. PASTOR,
Y. SCHEKIMAN, J. VANPOUCKE, D. VIELLE

Les caractéristiques principales du système LAPUMS (Logiciel d'Aide à la Programmation et à l'Utilisation des Modèles Statistiques) et de son langage de programmation le LOM (Langage d'Organisation de Modules) sont présentées.

Les objectifs assignés au système LAPUMS sont :

- d'assister le statisticien dans la conception des logiciels et la réalisation d'interfaces entre logiciels. Dans cette dernière mission, les programmes intégrés dans le LAPUMS sont plongés dans un environnement plus riche. Ils deviennent plus puissants et généralement d'une utilisation plus souple.

- d'assister l'utilisateur en rendant les logiciels statistiques directement et facilement accessibles aux non-initiés en Informatique.

Les auteurs présentent entre autres deux applications :

- conception d'un système intégré d'analyse de la variance multivariées à partir de la programmation de l'Institut National de la Recherche Agronomique ;

- création, autour du logiciel GLIM (Generalised Linear Interactive Modelling) d'une bibliothèque de programmes d'analyse de la variance.

Mots clés : Système d'Aide à la Programmation, Langage Statistique, Meta-Traducteur, Production Assistée de Logiciels.

L'hypothèse d'asymétrie hémisphérique dans l'activité de l'EEG du sommeil et son implication sur le nombre et l'indépendance des générateurs du rythme delta de sommeil, ont été testées par l'étude des relations inter- et intrahémisphériques. Les paramètres de puissance (PS) et de cohérence (CH) spectrales ont été calculées pour des périodes de 25 sec. de sommeil, sur six dérivations EEG enregistrées sur six sujets et préalablement classées manuellement. Les analyses factorielles pratiquées ont consisté en : analyse en composantes principales (ACP) sur les données brutes, analyse des correspondances (ACR) sur les classes (Benzécri, 1973 ; Lebart, 1977), analyse discriminante (AD) et analyse sur opérateurs (Escoufier, 1979), classification automatique (Diday, 1970).

Les résultats concernent :

- La mise en évidence d'une oscillation des potentiels cérébraux de longue période (1 à 5 min.).
- La décroissance de la CH globale dans le passage de la veille au sommeil profond.
- L'existence d'une asymétrie interhémisphérique stable des activités EEG au cours du sommeil, variable selon les sujets.

Dans la mise en évidence de ces résultats, ont surtout été "opératoires" :

- 1° - L'étude comparée des ACR et ACP qui montre une concordance d'ensemble des résultats et une plus grande quantité d'informations fournie par l'ACR en particulier la mise en évidence des relations non linéaires.
- 2° - L'analyse de sous-ensembles des individus ou des variables, survenant après l'analyse de l'ensemble des données, de façon à mettre en évidence des sources particulières de variance.
- 3° - La robustesse des méthodes d'analyse factorielle en même temps que leur aptitude à mettre en valeur de faibles sources de variance sur les plans factoriels de rang élevé.
- 4° - Les méthodes d'analyse discriminante (AD) dont les résultats se rapprochent de ceux de l'ACR, alors que l'analyse des opérateurs fournit des résultats similaires à ceux de l'ACP.
- 5° - Enfin des tests statistiques classiques (en particulier T^2 de Hotelling, appliqué sur les distances interclasses de l'AD) ont permis une "séparation" des différents stades de sommeil individualisés.

METHODES DE CLASSIFICATION AUTOMATIQUE
DANS L'ANALYSE DE L'ELECTROENCEPHALOGRAMME DE SOMMEIL

Auteur : J.P.BANQUET, LENA, La Salpêtrière
47, bd. de l'Hôpital, 75651 - PARIS CEDEX 13.

Les méthodes d'agrégation autour de centres mobiles et de classification hiérarchique ascendante ont été utilisées sur des données électroencéphalographiques (EEG) de sommeil.

Les données proviennent de l'enregistrement sur papier et bande magnétique du sommeil de six sujets sur six dérivations EEG. Le calcul des paramètres de puissance (PS) et cohérence (CH) spectrales a été réalisé sur des époques de 512 points (5.12 sec.).

L'agrégation autour des centres mobiles ou nuées dynamiques, utilisant la procédure de recherche des formes fortes ou encore groupements stables (E. Diday, 1972) rend les partitions indépendantes des premiers centres choisis. Les classes sont construites en utilisant la distance du χ^2 à partir de l'ensemble des coordonnées du tableau de données.

Les résultats se présentent sous 2 formes :

- 1° - Tableau des correspondances entre les 15 classes C_i - retenues sur les 216 classes de la partition croisée et les différents stades P_i du continuum veille-sommeil. Ces tableaux de correspondance permettent de définir des index d'homogénéité et de dispersion.
- 2° - Une analyse des correspondances (ACR) à partir des classes prises comme individus afin de permettre une comparaison avec l'ACR pratiquée sur les données primitives (Lebart, 1977). Les corrélations élevées entre les facteurs des deux analyses confirment la qualité des classes obtenues.

Les classes de la classification hiérarchique ascendante (Benzécri, 1976) ont été constituées à partir des 6 premières coordonnées factorielles d'une analyse de correspondance (ACR), (Jambu, 1979), successivement sur les ensembles de sommeil et de veille active. La distance du χ^2 et le critère d'agrégation des classes par la maximisation du moment centré d'ordre deux d'une partition, ont été retenus.

La CAH sur les individus permet ici également de constituer un tableau de correspondance entre classes C_i et stades P_i , avec des résultats similaires à ceux des nuées dynamiques. La CAH présente l'avantage de donner la généalogie des C_i générées, et de distinguer un polymorphisme réel des C_i (niveau d'agrégation à indice élevé) d'un polymorphisme apparent (convergence des C_i à un niveau de bas indice).

Une comparaison et une correspondance est établie entre les deux méthodes de classification et l'ACR.

ETUDE DES PROPRIETES STATISTIQUES
DES ESTIMATIONS DES PARAMETRES D'UNE COURBE
DE REPONSE OBTENUES PAR LA METHODE DE L'ESCALIER.

Auteurs : BASTENAIRE François - REGNIER Lionel
Service Statistique
Institut de Recherches de la Sidérurgie (IRSID)
78105 SAINT-GERMAIN-EN-LAYE CEDEX

La méthode de l'escalier [1] permet l'estimation de la moyenne μ et de l'écart type σ de la courbe de probabilité de réponse à l'intensité d'une action lorsque cette courbe est supposée sigmoïde normale.

La précision de ces estimations dépend fortement du nombre d'essais réalisés, les auteurs de la méthode suggérant un nombre supérieur à 40.

Or, en expérimentation, les questions de coût et parfois de durée d'essais limitent bien souvent l'exécution de cette méthode à un nombre d'essais inférieur à 20.

C'est pourquoi il nous a semblé intéressant de calculer les propriétés statistiques (biais, dispersion, loi de distribution) des estimations μ et σ obtenues par cette méthode pour un nombre d'essais allant de 5 à 20 et en fonction des diverses conditions d'application qui peuvent, en pratique, se présenter.

Référence :

- [1] W.J. DIXON - A.M. MOOD : "A method for obtaining and analyzing sensitivity data" - Journal of the American Statistical Association, 43 (1948), 109.

TESTS DE L'AUTOCORRELATION DES ERREURS DANS UN MODELE
LINEAIRE GAUSSIEN.

Auteur : Mohammed BENHELLI
Laboratoire de Statistique et Probabilités
Université Paul Sabatier
31062 TOULOUSE Cedex

On considère un modèle linéaire gaussien classique $Y = X\beta + u$ et on teste l'hypothèse de l'indépendance des erreurs contre l'hypothèse selon laquelle les erreurs sont engendrées par un processus de Markov stationnaire du premier ordre (ou d'un ordre supérieur).

Techniques utilisées :

Parmi les techniques existantes, on considère celles qui utilisent les résidus BLUS [1] BLUF [2] et BAUS [3] ; on introduit aussi des techniques nouvelles [4].

On évalue la puissance des tests correspondants par une méthode numérique, puis par simulation.

Références :

- [1] THEIL, H. (1965). "The analysis of disturbances in regression analysis". J. Amer. Statist. Ass. 60, 1067-1079.
- [2] ABRAHMSE A.P.J. and J. KOERTS (1971). "New Estimators of disturbances in regression analysis". J. Amer. Statist. Ass. 66,71.
- [3] CLIFFORD, H. (1971). "A note on approximate regression disturbances" Discussion paper n°2, january 1971. Depart. of economics University of Minnesota.
- [4] BENHELLI M. Thèse de 3ème cycle, Université Paul Sabatier, Toulouse (prévue pour juin 1980).

PRESENTATION DU LANGAGE GEOL

Auteurs : BOUCHET P., CUNIN Y., GRIFFITHS M., MALLET J.-L.,
ROYER J.-J.
Ecole Nationale Supérieure de Géologie
Centre de Recherche en Informatique de Nancy
54001 NANCY Cedex B.P. n° 452

A partir des problèmes spécifiques posés par la cartographie automatique et l'analyse des données géologiques et minières, les auteurs ont conçu et développent actuellement un langage de programmation adapté. Ce langage présente entre autre particularité d'introduire dans sa syntaxe la notion de tableau de données sous une forme plus naturelle que celle induite par les langages classiques.

Références :

- P. BOUCHET, Y. CUNIN, M. GRIFFITHS, J.-L. MALLET, J.J.-ROYER :
"Langage d'analyse de données géologiques" (Congrès de la "Convencion Informatica Latina", Barcelone, 1979).

ANALYSES EN COMPOSANTES PRINCIPALES D'UNE FONCTION
ALEATOIRE HILBERTIENNE

Auteur : BOUDOU Alain
Laboratoire de Statistique et Probabilités
Université Paul Sabatier
31077 TOULOUSE Cedex

On étudie différents types d'analyses en "composantes principales" d'une fonction aléatoire hilbertienne, chacun d'eux pouvant être présenté comme une extension "naturelle" de l'analyse en composantes principales de variables aléatoires réelles. Lorsqu'on considère "le cas fini", on obtient alors des méthodes d'analyses de tableaux ternaires. On donne un exemple d'application qui appartient au domaine électoral.

Références :

- J.M. BOUROCHE : "Analyse des données ternaires : la double analyse en composantes principales" - COREF - conseil, recherche et formation.
- J. DAUXOIS, A. POUSSE : "Les Analyses factorielles en Calcul des Probabilités et en Statistiques : essai d'étude synthétique".
Thèse - Université Paul Sabatier, Toulouse, 1976.

MODELE MATHEMATIQUE POUR L'ETUDE DE LA FERTILITE
ET DE LA VIABILITE DES CROISEMENTS CONSANGUINS.

Auteurs : BOUFFETTE Armide-René
BOUFFETTE Jacqueline
Département de Mathématiques
Université Claude Bernard - LYON 1
43, boulevard du 11 Novembre 1918
69622 VILLEURBANNE Cedex

A partir des hypothèses génétiques proposées par C. BIEMONT [1] [2] on se propose de calculer et de comparer la fertilité et la viabilité dans les cas de croisements frère-soeur, demi-frère soeur, cousins et double cousins.

Puis les variations de la fertilité et de la viabilité en fonction des caractéristiques de la population initiale sont étudiées. Enfin, on compare la notion de viabilité et celle de fardeau génétique telle que l'utilisent N.E. MORTON, J.F. CROW et H.J. MULLER [3].

Références :

- [1] C. BIEMONT : "Inbreeding effects : evidence for a genetic system which regulates viability in Drosophila Melanogaster Populations" - Mechanisms of Ageing and Development - Vol. 8 - 1978.
- [2] C. BIEMONT : "An inbreeding sensitivity gene in Drosophila Melanogaster" - Experientia - Vol. 36 - 1980.
- [3] N.E. MORTON, J.F. CROW, H.J. MULLER : "An estimate of the mutational damage in man from data on consanguineous marriages" - P.N.A.S. - Vol. 42, p. 855-863 - 1956.

ANALYSE EN COMPOSANTES PRINCIPALES
D'UNE FONCTION ALEATOIRE QUALITATIVE

Auteur : BOUMAZA Rachid
Laboratoire de Statistique et Probabilités
Université Paul Sabatier
31077 Toulouse Cedex

A partir de l'A.C.P. non linéaire d'une f.a. introduite en [2], on développe une méthode d'A.C.P. d'une f.a. $(X_t)_{t \in T}$ où, pour tout t de T , X_t est qualitative à k modalités. On montre que cette A.C.P. est équivalente à l'A.C.P. linéaire de la f.a. $(\tilde{X}_t)_{t \in T}$, où \tilde{X}_t est la v.a. à valeurs dans \mathbb{R}^k ayant pour composantes les indicatrices des modalités de X_t , et on rejoint l'analyse harmonique qualitative développée parallèlement en [3].

On envisage diverses approximations de ces analyses. On étudie également les représentations des individus, variables ou modalités qu'on en déduit.

- Références : [1] R. BOUMAZA - Contribution à l'étude descriptive d'une fonction aléatoire qualitative. Thèse de 3ème Cycle. Université Paul Sabatier, Toulouse, 1980.
- [2] J. DAUXOIS et A. POUSSE - Les analyses factorielles en Calcul des Probabilités et en Statistique : essai d'étude synthétique. Thèse, Université Paul Sabatier, Toulouse, 1976.
- [3] J.C. DEVILLE et G. SAPORTA - Analyse harmonique qualitative, Colloque IRIA, Octobre 1979.

REPRESENTATION ORDONNÉE DES
CLASSIFICATIONS HIÉRARCHIQUES

Auteur : Gildas BROSSIER
U.E.R. des Sciences et Techniques
Université de Haute-Bretagne
35043 - RENNES

Dans la représentation graphique d'une hiérarchie, la distance entre deux éléments est représentée par la hauteur du noeud commun qui les réunit. Ceci entraîne que seul l'axe vertical du graphique usuel est porteur d'informations.

On propose d'ordonner les noeuds terminaux de l'arbre hiérarchique de façon à pouvoir interpréter, en plus, la proximité des points sur l'axe horizontal.

On décrit une procédure simple permettant de choisir parmi les 2^{n-1} représentations possibles d'un arbre celle qui va nous donner une ordonnance interprétable au regard des données initiales.

Cette procédure est indépendante de l'algorithme de classification utilisé et n'intervient qu'au moment du tracé de l'arbre.

L'axe vertical inchangé, et l'axe horizontal étant maintenant porteur d'informations, l'interprétation globale s'en trouve facilitée surtout pour le non-spécialiste.

RECHERCHE DE BI-CLASSES
PAR LA METHODE DES NUEES DYNAMIQUES

Auteurs : G. BROSSIER

Université de Haute Bretagne, Rennes II
6, avenue Gaston Berger, 35043 RENNES CEDEX

D. JACQUET

INRIA, Domaine de Voluceau
78150 LE CHESNAY

Une bi-classe B_i est composée de deux classes B_i^+ et B_i^- . On veut que les objets composant une classe (B_i^+ ou B_i^-) se ressemblent le plus possible et donc que leur distance au noyau de cette classe soit aussi faible que possible. D'autre part on veut que les deux classes composant la bi-classe s'opposent, ce qui entraîne que la distance d'un objet au noyau de l'autre classe soit aussi grande que possible.

On utilise un critère exprimant l'adéquation d'une bi-classe à ces deux propriétés.

On montre que l'algorithme améliore ce critère par itérations successives jusqu'à convergence.

On présente les résultats obtenus sur un exemple pédagogique.

APPROXIMATION OF REGRESSION FUNCTIONS AND MODEL CHOICE.

Auteur : BUNKE O.

Humboldt University
BERLIN - RDA

The approximation of regression functions of unknown form by linear or non linear models is treated in the basis on statistical goodness criteria like the mean square approximation or prediction error (M.S.E). First we discuss transformations to get approximatively homogeneous variances, normality and less non-linearity. Then we present the essential asymptotic results for inadequate models and for testing model hypotheses. We discuss weighted least squares and (in the linear case) useful minimax estimators. As tools for model choice we compare different estimators for the MSE.

Gilles CELEUX et Yves LECHEVALIER

INRIA
Domaine de Voluceau
78150 LE CHESNAY

ANALYSE MULTIDIMENSIONNELLE DE L'OPINION D'UN JURY.

Auteur : CARAUX Gilles
Chaire de Mathématiques Appliquées.
Ecole Nationale Supérieure Agronomique
34 000 - MONTPELLIER.

Après avoir introduit une représentation matricielle de l'opinion d'un jury sur un ensemble d'objets, nous suivrons la démarche des analyses descriptives multidimensionnelles qui consiste à rechercher une image simplifiée, mais aussi pertinente que possible, de la réalité.

Cette approche aboutira à dégager une représentation spatiale des objets comparés respectant le caractère antisymétrique de la relation de préférence portée sur eux.

Des exemples très variés seront présentés à chaque étape du raisonnement pour illustrer les résultats trouvés.

Références :

G. CARAUX : "Contribution à l'analyse multidimensionnelle des préférences" - Thèse - Université PARIS VI - 1979.

On présente un algorithme de J. Friedman de segmentation non paramétrique pour discriminer 2 classes. Cet algorithme utilise la distance de Kolmogorov-Smirnov entre 2 distributions de probabilité et conduit à la construction d'un arbre binaire de décision.

Deux extensions sont proposées

- On généralise cet algorithme au cas multi-classes et on aboutit à la construction d'un seul arbre de décision binaire pour discriminer les différentes classes.
- On propose une variante qui permet de construire des arbres de décision à 3 branches au lieu de 2, l'une des branches correspondant à une non décision.

Références :

- [1] C. CELEUX - "Contribution statistique à la reconnaissance automatique de spectres de machines tournantes" - Thèse de 3ème cycle - Université Paris VI - Novembre 1978.
- [2] J.H. FRIEDMAN - "A recursive partitioning decision rule for non parametric classification" - IEEE-Transaction on Computers - Avril 77.
- [3] L. GORDON, R.A. OHLSEN - "Asymptotically efficient solution to the classification problem" - Annals of statistics. Vol.6 n°3 - 78.

DETECTION D'UNE RUPTURE DE NIVEAU DANS UN PROCESSUS "ARMA".

Auteur : CHALMOND Bernard
 Département d'Informatique Médicale
 Centre Hospitalier Régional Universitaire
 2 Bd du Maréchal de Lattre de Tassigny
 21034 DIJON Cedex.

A la vue d'une réalisation (z_1, \dots, z_n) d'un processus ARMA(p,q), on cherche à savoir si à partir d'une date t_0 inconnue ($t_0 = 1, n$), le phénomène observé a subi une altération, par rupture de la stationnarité de son niveau moyen.

Nous voulons décider de l'une des hypothèses suivantes :

$$H_0 : z_t = \beta + u_t, \quad H_1 : z_t = \beta + \delta_t \mathbb{1}_{t < t_0} + u_t$$

où u_t est le processus ARMA(p,q) : $\Phi(B)u_t = \Theta(B)a_t$;

Seuls les paramètres p et q sont supposés connus.

Pour ce faire, un indicateur de rupture basé sur la statistique classique de la somme des déviations cumulées :

$$S_r = \sum_{t=k+1}^r d_t \quad \text{où} \quad d_t = z_t - \frac{1}{t-1} \sum_{j=1}^{t-1} z_j \quad \text{est considéré.}$$

La propriété de suite mélangeante du processus z_t permet alors de déterminer la loi limite de cet indicateur sous H_0 , et par là d'établir une règle de décision. Sa puissance est considérée sur un processus AR(1) et pour une rupture par saut : $\delta_t = \delta$, et ce, par simulation.

Par ailleurs le test du rapport des vraisemblances maximales est abordé pour un processus AR(1).

Références : BROWN-DURBIN-EVANS, JRSS, B, vol. 37, n°2, (1975).
 HAWKINS, JASA, vol. 72, n°357, (1977)

CALCUL DES COEFFICIENTS D'IDENTITE
 INEGALITES, ET DISTANCES GENETIQUES.

Auteur : CHEVALET Claude
 Laboratoire de Génétique cellulaire
 I.N.R.A. Centre de Recherches de Toulouse
 31320-CASTANET-TOLOSAN

Dans une population, les relations de parenté définissent une relation d'ordre entre zygotes, qui induit une autre relation d'ordre, R , dans l'ensemble G des gènes qu'ils portent. On identifie tout événement du processus de ségrégation mendélienne à un locus, à une partition de G en sous-ensemble qui sont, selon R , des arbres. Cette analyse donne des méthodes de calcul des coefficients d'identité, et un système de notation qui permet l'établissement d'inégalités entre coefficients d'identité moyens. Enfin on construit une famille de distances génétiques fondées sur les coefficients d'identité.

ULTRAMETRIQUES ADAPTATIVES ET OPTIMISATION
D'UNE HIERARCHIE PAR LA METHODE DES NULLS
DYNAMIQUES

1. Ultramétriques adaptatives : R. CHIFFLET, CAP-SOGETI.

Il s'agit de munir un ensemble E de k structure hiérarchiques localement optimales au sens d'un critère bien défini minimiser ce critère revient à minimiser la somme des distances (prises d'une famille d'ultramétriques admissibles) de tous les points d'une classe à l'élément représentatif de cette classe. On ne cherche pas à définir une hiérarchie optimale sur E tout entier mais on vise à trouver une partition de E en k classes et à associer simultanément à chacune de ces classes une hiérarchie localement optimale.

2. Hiérarchies localement optimales : H. RAJAMBONDRAJNY, I.N.R.I.A.,
Domaine de Voluceau, 78150 Le CHESNAY

Il s'agit de trouver une hiérarchie sur E optimale au sens d'un critère global exprimant la somme des inerties de chacun des paliers de cette hiérarchie.

L'algorithme se généralise au cas des k -structure hiérarchiques c.a.d. à l'ensemble des k -uplets (H_1, \dots, H_k) , de hiérarchies. On cherche alors simultanément k hiérarchies locales minimisant le critère d'inertie.

ESTIMATION LOCALE DE LA SECONDE FORME QUADRATIQUE
FONDAMENTALE D'UNE VARIÉTÉ SUPPORT D'UNE MESURE DE
PROBABILITÉ.

Auteur : CIRIER Guy
I.U.T. Avenue de Versailles
142, Avenue de Versailles
75016 - Paris 1

On construit un estimateur de la seconde forme quadratique fondamentale d'une sous-variété M de R^d , de dimension q , de classe $p \geq 2$ support d'une mesure de probabilité P , en un point O fixé de M .

La méthode d'estimation proposée s'appuie sur la connaissance - ou l'estimation, lorsque M est inconnu - d'un repère canonique, [1], solution de l'opérateur de contiguïté de Geary [2] à partir d'un échantillon de points tirés au hasard dans la loi P . On étudie la convergence de l'estimateur lorsque la taille de l'échantillon tend vers l'infini sous différentes hypothèses : on connaît le plan tangent et le sous-espace fondamental du second ordre, M est une hypersurface, et enfin, le cas où M est une sous-variété inconnue de R^d .

Références :

- [1] G. CIRIER. "Application des méthodes statistiques à l'étude locale d'une variété". Thèse, 1979, Paris.
- [2] GEARY. "The contiguity ratio and statistical mappings". Incomp. Stat. 1954, 5, p.115.

PROPRIETES ASYMPTOTIQUES DE QUELQUES METHODES
NON PARAMETRIQUES UTILISEES EN ANALYSE DISCRIMINANTE.

Auteur : COLLOMB Gérard
Laboratoire de Statistique et Probabilités
Université Paul Sabatier
31077 TOULOUSE Cedex

Soit (X_i, Z_i) $i=1, n$, une suite de n observations indépendantes du couple (X, Z) , où X est un vecteur aléatoire de \mathbb{R}^p et Z une variable aléatoire à valeurs dans un ensemble C de cardinal fini (ensemble de classes ou modalités). La loi du couple (X, Z) n'est assujettie à aucune hypothèse permettant d'utiliser des méthodes paramétriques pour le problème de discrimination suivant : (x, z) étant une observation de (X, Z) , x connu et z inconnu, "prévoir" z ou "affecter" x à un élément de C . Nous considérons les deux méthodes non paramétriques suivantes

- *méthode de la boule* : on affecte x à la classe la plus fréquente parmi les Z_i , $i=1, n$, tels que X_i appartienne à la boule de centre x et rayon h_n ;
- *méthode des k points les plus proches* : on affecte x à la classe la plus fréquente parmi les Z_i , $i=1, n$, tels que X_i appartienne à l'ensemble des k_n points X_j , $j=1, n$ les plus proches de x .

On étudie quelques propriétés asymptotiques de ces discriminateurs : ces propriétés sont liées au comportement asymptotique des suites $(h_n)_{\mathbb{N}}$ ou $(k_n)_{\mathbb{N}}$.

SEPARABILITE DES HYPOTHESES DANS L'ANALYSE DES TABLES
DE CONTINGENCE A 2 ET 3 DIMENSIONS.

Auteurs : D. COLLOMBIER - J.R. MATHIEU
Université Paul Sabatier
Laboratoire de Statistique et Probabilités
31072 TOULOUSE Cedex

L'analyse des tables de contingence observées par sondage aléatoire se heurte à des difficultés pratiques dues au fait que - sauf exceptions - les estimations sous-contraintes et les valeurs des statistiques de test usuelles ne sont obtenues que par recours à des procédures itératives. Ainsi en est-il, par exemple, de l'étude de l'identité des marges dans un tableau carré à deux dimensions ou de l'étude de la non interaction d'ordre deux dans une table complète à trois dimensions.

Il peut advenir cependant dans de telles situations qu'on soit en mesure d'introduire une ou plusieurs hypothèses auxiliaires séparables de H_0 : l'hypothèse d'analyse, et séparables entre elles au sens envisagé par J.R. MATHIEU [2] suite aux travaux de J. AITCHISON [1]. Il suffit alors que les statistiques du chi-deux de K. PEARSON et les estimations de maximum de vraisemblance sous chacune des hypothèses auxiliaires soient connues sous une forme explicite pour que l'on puisse proposer une approximation, explicite elle aussi, de la statistique de K. PEARSON et de l'estimation maximum de vraisemblance sous H_0 . Une telle éventualité est fréquente dans la pratique des tables de contingence complètes.

Nous nous intéressons ici à l'analyse des tables complètes à deux et trois dimensions sous certaines contraintes linéaires et log-linéaires. Les deux premières parties de notre communication sont consacrées à la présentation de la structure statistique dans laquelle nous nous plaçons puis au rappel de la définition et de quelques propriétés de la séparabilité des hypothèses. Les deux parties suivantes concernent l'analyse des tables à deux dimensions, les deux dernières l'analyse des tables à trois dimensions sous hypothèse de non-interaction d'ordre 2.

Références :

- [1] J. AITCHISON (1962) : "large sample restricted parametric tests". J. Roy. Stat. Soc. A. 24, p.234-250.
- [2] J.R. MATHIEU (1978) : "Contribution à l'étude de la séparabilité des hypothèses, au sens du test du χ^2 , dans la théorie asymptotique". Thèse, Université Paul Sabatier, Toulouse.

LE RECENSEMENT GENERAL DE L'AGRICULTURE 1979-1980
EN FRANCE

Auteur : CORDIER Paul
Service Régional de Statistique Agricole Midi-Pyrénées
31074- TOULOUSE Cédex

Organisé conjointement par les Services Statistiques du Ministère de l'Agriculture (Service Central des Enquêtes et Etudes Statistiques, Services Régionaux et Services Départementaux) et l'Institut National de la Statistique et des Etudes Economiques, le Recensement Général de l'Agriculture 1979-1980 s'achèvera sur le terrain à la fin du mois de Mai.

1 300 000 exploitations agricoles auront alors reçu la visite d'un enquêteur venu remplir un questionnaire.

Faisant suite à ceux réalisés en 1929, 1955 et 1970, il s'inscrit dans un programme de recensement mondial de l'agriculture centré autour de l'année 1980, établi par la FAO et confirmé par une directive de la Communauté Economique Européenne.

Cette importante opération statistique répond à plusieurs objectifs :

- la connaissance actuelle détaillée des structures de production de l'agriculture française, à un niveau géographique fin,
- la mesure des mutations qu'elles ont subi depuis une dizaine d'années,
- l'établissement d'une base de sondage pour le tirage des échantillons d'enquêtes pour une période assez prolongée.

Avant d'étudier l'organisation des différentes phases, la nature des données recueillies, les résultats qui en sont attendus, on examinera les diverses solutions qui pouvaient être envisagées, quant au champ du recensement, aux techniques utilisables aux différents types de questionnaires, aux informations à collecter, et on donnera les raisons qui ont abouti à celles finalement retenues.

On se demandera en conclusion, quel est l'avenir des recensements généraux de l'agriculture en France, au moins dans leur forme actuelle, compte tenu essentiellement du développement des grands fichiers informatisés et de la notion qui se fait jour d'exploitation agricole professionnelle.

ESTIMATION LINEAIRE DES SAISONNALITES MULTIPLES ET EVOLUTIVES

Auteur : COSTANTINI Marcel
CESMAP
67, avenue Lénine
94112 ARCUEIL

L'introduction de saisonnalités multiples entraîne une dégénérescence qui est habituellement résolue par l'introduction de contraintes faisant partie intégrante du modèle. Les estimations et formules de test sont donc établies à partir d'une g -inverse G adaptée au problème résolu. La prise en compte d'une évolution linéaire de la saisonnalité introduit de nouvelles contraintes non liées à la dégénérescence et conduit à de nouvelles expressions. En parallèle à l'établissement des estimations et des tests on portera une attention particulière à la notion d'estimabilité dans ce type de modèle.

Références :

- [1] Searle : Linear Models - Wiley
- [2] Rao : Linear Statistical Inference - Wiley
- [3] Rao-Mitra : Generalized Inverses and its applications - Wiley

L'ANALYSE SUBJECTIVE

UNE GENERALISATION DE L'ANALYSE DES DONNEES

Auteur : René-Charles CROS

Unité de Recherches de Pathologie Digestive
INSERM U 31
46, boulevard de la Gaye
13009 MARSEILLE

En raison des difficultés qui s'opposent, en tout domaine, à la constitution de données à la fois précises et exhaustives, il semble qu'en général les outils d'analyse mathématiques soient inadaptés à l'étude des problèmes tels qu'ils se posent concrètement.

L'analyse subjective propose d'exprimer toute variable par 2 limites ce qui fournit un cadre permettant de décrire les données imprécises et manquantes. Cette formalisation donne en outre la possibilité de neutraliser intentionnellement toute variable et de définir ainsi des objets ne présentant qu'un caractère déterminé (ex : un sujet âgé de plus de 65 ans, sans autre précision) ou des combinaisons de caractère jugés intéressantes (ex : sujets de plus de 65 ans et de sexe masculin). L'étude des relations entre caractères des objets peut donc être remplacée par l'étude des relations entre objets "épurés" -qui ne comportent pas de caractères "parasites"- ces derniers étant plongés dans la même espèce de représentation que les objets réels qui apportent le substratum informationnel nécessaire à l'étude des relations.

Cette méthode clarifie certains problèmes de reconnaissance des formes -par exemple l'aide au diagnostic médical- en élucidant les propriétés d'entités telles que "sujet présentant un des symptômes S sans être nécessairement atteints par la maladie M" "sujets certainement atteints par M sans précision sur les symptômes S qu'ils éprouvent" etc.

INCIDENCE DES VARIATIONS DE LA MORTALITE

POUR UNE CAUSE DONNEE SUR LA MORTALITE GENERALE - ASPECT METHODOLOGIQUE.

Auteurs : DAMIANI Paul*, MASSÉ Hélène**, AUBENQUE Maurice*
* INSEE - 18, Bd. Adolphe Pinard, 75675, PARIS Cedex 14
** INSERM - 101, rue de Tolbiac, 75645, PARIS Cedex 13

On a exposé une méthode permettant de mesurer l'incidence des variations de la mortalité pour une cause donnée sur la mortalité générale en supposant la non indépendance des causes de décès.

On a appliqué cette méthode aux données régionales françaises de la période 1967-1969 pour huit causes principales de décès.

Pour une cause k, on a admis le modèle suivant, pour le groupe d'âge i et un sexe donnés :

$$\text{Log} (q_i - h_{ki}) = a_k + b_k \text{Log} q_{ki}$$

où : q_i , q_{ki} , quotients de mortalité générale et de la cause k,

h_{ki} , quotient limite en l'absence de la cause k.

Le modèle suppose que les droites de régression de $\text{Log} (q_i - h_{ki})$ en fonction de $\text{Log} q_{ki}$ sont confondues quel que soit l'âge pour une cause et un sexe donnés.

On expose la méthode de calcul de h_{ki} et des coefficients a_k et b_k . Cette méthode débute par la détermination graphique d'une valeur approchée h'_{ki} de h_{ki} .

On détermine ensuite les gains en espérance de vie dus à la disparition d'une cause k. On compare ces gains à ceux calculés avec l'hypothèse d'indépendance des causes de décès.

UNE MANIFESTATION DE LA LOI D'ARC-SINUS EN METEOROLOGIE

Auteurs : DANVEL Corinne, DER-MEGREDITCHIAN Guy

Les lois d'arc-sinus révèlent un comportement surprenant de la mémoire des phénomènes aléatoires, illustré par la mise en défaut du "bon sens gaussien" ; la valeur moyenne étant la moins probable, les valeurs extrêmes étant les plus probables, et la densité de probabilité correspondante étant une courbe en "U" ($f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$) au lieu d'une courbe en cloche.

Nous avons recherché si ces lois s'appliquaient au domaine météorologique, d'abord sur des fichiers de température, puis de géopotential. L'insuffisance de la taille des fichiers de température n'a permis de déceler qu'une image tronquée de la première loi. Par contre, en prenant les échantillons-année pour un nombre suffisant ($N \geq 30$) de composantes principales du champ de géopotential 500 mb, la première loi se manifeste avec une grande netteté.

Ce phénomène a été confirmé avec un degré élevé de signification par les tests d'adéquation correspondants paramétriques (Student, Fisher) et non-paramétriques (Kolmogorov, Khi-deux, informationnel).

Références :

- [1] SPITZER F. - "Principles of Random Walk", Princeton, 1964.
- [2] FELLER W. - "An Introduction to Probability Theory and its applications". Tome 1, Wiley.

ESTIMATION AUTOMATIQUE DE LA DENSITE

Auteur : Paul DEHEUVELS

Université Paris VI

& Ecole Pratique des Hautes Etudes (1ère section)

Etant donné un échantillon X_1, \dots, X_n de v.a. indépendantes réelles de même loi admettant une densité f sur \mathbb{R} , l'estimation de f par la méthode du noyau de Parzen-Rosenblatt est donnée par $f_n(x) = (n \delta_n)^{-1} \sum_{i=1}^n K((X_i - x)/\delta_n)$, où K est une densité fixée, et δ_n un nombre positif dépendant de n . On constate que le choix de K parmi les noyaux habituellement utilisés est peu important par rapport au choix de la "fenêtre" δ_n . On se propose de présenter une procédure permettant un choix automatique de δ_n en fonction des seules valeurs de l'échantillon, pour K fixé.

Ce procédé permet de réaliser une estimation automatique de la densité sans choix extérieur. Les justifications de cette méthode sont liées, d'une part au fait qu'elle fournit une estimation convergent p.s. uniformément, avec un ordre de consistance asymptotiquement optimal selon le critère du M.I.S.E., d'autre part du fait qu'elle réalise pour des lois usuelles (normales, gamma, ..) des estimations très voisines des estimations optimales pour de petits échantillons ($n = 10, 20$) comme on a pu le constater empiriquement par simulations.

Il est nécessaire de supposer que f a une dérivée seconde de carré intégrable. Ces méthodes sont à comparer aux résultats de Silverman pour la distance uniforme.

Références :

- DEHEUVELS P., HOMINAL P., 1980, Estimation automatique de la densité, Revue de Statistique appliquée (en cours publ.)
- SILVERMAN B., 1978, Choosing the window when estimating a density, Biometrika 65, 1, p. 1-11.

ITERATIONS DANS UN PROCESSUS DE GALTON-WATSON

Auteur : DEMONGEOT Jacques
IRMA
Université Scientifique et Médicale de Grenoble
BP 53X 38041 Grenoble Cedex

Nous faisons d'abord un bilan des propriétés mathématiques concernant, dans le cas sur-critique, la taille de la $n^{\text{ème}}$ génération d'une population régie par un processus de Galton-Watson ; ces propriétés sont liées à la considération d'itérations de la fonction génératrice g du processus. En considérant ensuite la fonction $(1-g)$, nous montrons quels renseignements supplémentaires, concernant la taille de la population, on peut tirer de l'étude de ses points fixes et de ses cycles éventuels.

Nous appliquons enfin ces résultats à des cas particulièrement simples (processus de Galton-Watson à loi de reproduction homogène binômiale, poissonnienne, etc...).

METHODES D'ETUDE DE LA STRUCTURE DE L'INTERACTION "GENOTYPE * MILIEU"
ET DE RECHERCHE D'UN MODELE EXPLICATIF A EFFETS FIXES : APPLICATION
A L'ANALYSE DES RESULTATS D'UN RESEAU D'ESSAIS DE VARIETES DE BLE TENDRE.

Auteurs : J.B. DENIS*
J.P. GOUET**
J. TRANCHEFORT***

Les informations obtenues sur les rendements de variétés de céréales dans les conditions particulières d'un essai isolé, ne sont pas immédiatement transposables à l'ensemble d'une région agricole. En revanche, l'analyse d'un réseau d'essais constituant un échantillon de diverses situations possibles peut permettre de trouver des conclusions plus générales, donc largement vulgarisables. Mais on a bien souvent constaté que lorsque les lieux d'essais sont trop dissemblables, l'interaction "génotypes * milieux" devient importante et enlève tout intérêt aux conseils qui pourraient être déduits de l'examen des effets principaux. ☞

Pour constituer des groupes d'essais, homogènes quant au classement relatif des génotypes expérimentés, on a donc appliqué et comparé différentes méthodes d'étude de la structure de cette interaction (arrangement visuel de matrices, généralisation du test de TUKEY, vecteurs de MANDEL, classification automatique).

Ensuite, pour rechercher parmi les caractères des variétés et les conditions de milieu, des covariables explicatives de cette structure, on a utilisé la régression factorielle (ou analyse de covariance factorielle) et l'analyse discriminante lorsque des groupes se discernaient clairement.

Ces études ont été conduites en admettant l'hypothèse d'un modèle à effets fixes.

* Laboratoire de Biométrie, I.N.R.A.-C.N.R.A., route de St-Cyr
78000 VERSAILLES.

** Service des Etudes Statistiques et Informatiques, Institut Technique des Céréales et des Fourrages (I.T.C.F.), Station Expérimentale - Boigneville - 91720 MAISSE.

DIFFICULTES DE L'UTILISATION DE LA STATISTIQUE POUR DES RECHERCHES EMPIRIQUES

Auteurs : J. De Prins et W. Malbecq

Laboratoire des Etalons de Fréquence

Université Libre de Bruxelles,

Stanley 7 - B 1980 TERVUREN - Belgique

La Science empirique consiste dans l'acquisition et la systématisation des connaissances concernant des phénomènes observés. Ainsi, en chronobiologie, l'expérimentateur va détecter des rythmes qu'il voudra caractériser de manière valable, et ce malgré l'absence de tout modèle explicatif du rythme. Le chercheur utilisera pour ce faire la statistique. Cette dernière a la structure de toute théorie mathématique, c.à.d. que sur la base d'axiomes, des propriétés sont déduites. Pour concilier l'empirisme et la statistique, l'expérimentateur sera conduit à utiliser une méthode hypothéco-déductive.

Nous devons dès lors constater que :

- 1) Le problème sera généralement un "problème mal posé" [1]
- 2) Certains concepts (tels les "zones de confiance") ne donneront des résultats acceptables que si les hypothèses choisies sont quasi-exactes [2].
- 3) Il n'y a aucune raison de croire que certains concepts mathématiques (erreurs indépendantes, processus autorégressifs) introduits pour la simplicité ou la facilité de la théorie reflètent une quelconque "réalité" du phénomène observé [3].

Références

- [1] A. TIKHONOV et V. ARSENINE - Méthodes de résolution de problèmes mal posés - Edit. MIR - 1976
- [2] J. DE PRINS et G. CORNELISSEN - Methods Workshop - Cours OTAN Chronobiology Hanovre 1979 (Sous presse).
- [3] P.R. GUTOWSKI, E.R. ROBINSON et S. TREITEL : Spectral estimation : factor fiction, IEEE Trans. Geosci. Electron. 16, pp. 80-84, 1978.

UNE NOUVELLE DEMONSTRATION DE LA LOI DU TEST DE FISZ ET CHANG

Auteur : DER-MEGREDITCHIAN Guy
Météorologie Nationale
73-77, rue de Sèvres
92100 BOULOGNE

Les tests non paramétriques d'homogénéité de K échantillons présentent un grand intérêt pour les applications pratiques. S'ils sont nombreux pour le cas classique $k=2$, ils le sont beaucoup moins pour $k=3$ et encore moins pour le cas général $k>3$.

Le test de Fisz et Chang présente sur certains autres l'avantage de ne pas introduire un nouveau problème de tabulation des résultats, puisque les lois limites de la statistique utilisée s'expriment en fonction de la loi de Kolmogorov.

Nous donnons ici une démonstration nouvelle des résultats de Fisz et Chang basée sur une approche combinatoire, et dont l'idée a été avancée pour la première fois par Gnedenko dans le cadre du test de Kolmogorov-Smirnov.

Références :

- FISZ - "Probability theory and mathematical statistics". Wiley, 1963.
GNEDENKO - "Cours de la théorie des probabilités". Nauka, 1961.

TRAITEMENT STATISTIQUE DES DONNEES
RECUEILLIES DANS LES ELEVAGES.

Auteurs : DEROO M., STRIZYK S., BONY D.

ACTA (Association de Coordination Technique Agricole)
149, rue de Bercy
75595 PARIS

L'estimation de la valeur génétique des reproducteurs conduit à effectuer grand nombre de mesures "à l'élevage".

Le traitement statistique de telles données pose des problèmes multiples.

Créations de fichier volumineux .

Prise en compte de l'effet milieu tant au niveau des conditions d'élevage que de leurs variations dans le temps.

Structure "démographique" complexe des élevages : insémination artificielle et monte naturelle, périodicité d'utilisation des reproducteurs.

Nature des variables mesurées et erreur de mesure.

Confrontés à ces problèmes les praticiens ont développé des solutions "approchées" intégrant leurs contraintes de rapidité et de capacité de traitement. On se propose d'étudier sur quelques exemples l'incidence de ces approximations en montrant le sens et l'importance quantitative des biais qui peuvent en résulter.

N.B. - Cet exposé "bilan" sera une synthèse des travaux menés par un groupe de travail réunissant généticiens et statisticiens des différents Instituts d'Élevage et de l'ACTA..

ANALYSE HARMONIQUE QUALITATIVE
et ANALYSES CANONIQUES DE DEUX TRIBUS

Auteur : DEVILLE Jean-Claude

Institut National de la Statistique et
des Etudes Economiques (INSEE)
18 Boulevard Adolphe Pinard
75675 PARIS CEDEX 14

On montre qu'on peut associer à un processus qualitatif une mesure de probabilité sur un espace du type $\Omega \times T$. Celui-ci se trouve ainsi ipso facto de 3 tribus, permettant de définir 9 analyses canoniques.

Certaines d'entre elles sont triviales, les autres présentent un certain intérêt.

L'une d'entre elle, se ramène à l'analyse harmonique qualitative.

Il se trouve qu'elle est la seule à :

- respecter l'information de base
- être praticable sur des données individuelles nombreuses.

Cette approche permet en outre d'étudier très simplement les problèmes de discrétisation et d'approximation des données ainsi que la convergence de ces procédures.

ANALYSE EN COMPOSANTE PRINCIPALE
DES PROCESSUS ("ANALYSE HARMONIQUE") et
ANALYSE SPECTRALE DES PROCESSUS STATIONNAIRES

Auteur : DEVILLE Jean-Claude

Institut National de la Statistique et
des Etudes Economiques (INSEE)
18 Boulevard Adolphe Pinard
75675 PARIS CEDEX 14

On rappelle ce qu'est l'A.C.P. d'un processus du second ordre sur un intervalle compact.

On montre en quoi un processus stationnaire sur un groupe abélien localement compact est représentable sous forme d'intégrale stochastique.

Dans le cas où le groupe est fini on montre que la première notion est une généralisation de la seconde.

Pour le processus en temps discret, purement non déterministe, on montre l'existence d'une généralisation commune à ces deux analyses.

On signale aussi quelques problèmes non résolus. On termine par quelques idées pour représenter un ensemble fini de variables en les indiquant par des éléments d'un groupe fini.

APPROXIMATION DE L'A.C.P. SEMI-LINEAIRE D'UNE FONCTION
ALEATOIRE QUELCONQUE PAR L'A.C.P. SEMI-LINEAIRE D'UNE
FONCTION ALEATOIRE QUALITATIVE.

Auteur : DOSSOU-GBETE Simplicie

Université de PAU et des PAYS de l'ADOUR
Faculté des Sciences Exactes
Département de Mathématiques et Informatique
Avenue Philippon
64000 PAU

Ce travail concerne l'étude descriptive des fonctions aléatoires à valeurs dans un ensemble quelconque E . Nous avons résolu le problème de l'approximation de l'Analyse en Composantes Principales (A.C.P.) d'une telle fonction aléatoire par discrétisation de E à partir d'une partition de cet ensemble. Nous avons aussi explicité la forme générale des opérateurs qui définissent cette A.C.P., et précisé les conditions de leur compacité.

Références :

- [1] BOUMAZA, R. - "Contribution à l'étude d'une fonction aléatoire qualitative". Thèse de 3ème cycle, Toulouse 1980.
- [2] DEVILLE, J.C. et SAPORTA, G. - "Analyse harmonique qualitative". Communication aux Journées Internationales "Analyse des données et Informatique" de l'I.R.I.A., octobre 1979.
- [3] DAUXOIS, J. et POUSSE, A. - "Les analyses factorielles en calcul des Probabilités et en Statistique : essai d'étude synthétique". Thèse, Toulouse, 1976.
- [4] DOSSOU-GBETE, S. - "Approximation de l'analyse en Composantes Principales (A.C.P.) semi-linéaire d'une fonction aléatoire par l'A.C.P. semi-linéaire d'une fonction aléatoire qualitative". Publication du Laboratoire de Statistique et Probabilités. n°02-80, janvier 1980.

PRESENTATION COMPAREE DE METHODES D'ESTIMATION
DE COMPOSANTES DE LA VARIANCE SUR MODELE DE
GENETIQUE ANIMALE.

Auteur : DOUAIRE M. - MALLARD J.

ENSAR, 65, rue de Saint Brieuc
35042 RENNES

L'héritabilité, paramètre utilisé en sélection animale, peut se calculer à partir des composantes de la variance. Par ailleurs, la pratique du sélectionneur et la théorie de la génétique quantitative permettent d'avoir des idées a priori sur ce paramètre.

Nous utilisons sur une population animale quelques unes des méthodes d'estimation des composantes de la variance (Henderson, Rao, MMV restreint) pour comparer les estimations obtenues, et, lorsque c'est possible, les variances des estimateurs.

Aucune de ces méthodes ne garantit une estimation de l'héritabilité compatible avec ce qu'on connaît du paramètre, peut-être parce qu'elles n'intègrent pas vraiment la connaissance a priori qu'on en a, et que le critère de non-biais n'est pas forcément bien adapté.

QUELQUES REMARQUES SUR LA MODELISATION
DES SERIES CHRONOLOGIQUES SAISONNIERES

Auteur : DROESBEKE F.

Université Libre de Bruxelles
Institut de Statistique
Campus de la Plaine, C.P. 210
B - 1050 BRUXELLES

L'analyse des séries chronologiques saisonnières et leur désaisonnalisation constituent des problèmes importants, notamment en Economie.

L'élaboration d'un modèle adéquat est une étape utile pour une telle analyse.

Les méthodes de désaisonnalisation, par contre, ne reposent pas toujours sur une spécification précise d'un modèle générateur et ont parfois été élaborées en recourant à des considérations empiriques.

Nous examinons le problème de la modélisation d'un processus générateur pour certaines de ces méthodes et en étudions les caractéristiques essentielles.

COMPARAISON DE TABLEAUX DE CONTINGENCE

Auteur : ESCOPIER Brigitte
Laboratoire de Probabilités et Statistique
I. N. S. A.
20, avenue des Buttes de Coësmes
35043 RENNES CEDEX

Pour comparer des tableaux de contingence définis par le même couple de variables qualitatives sur des populations différentes, nous proposons de projeter les centres de gravité et les premiers axes d'inertie des différents nuages de points représentant ces populations, sur les premiers axes d'inertie du nuage de la population toute entière.

Ces axes sont, à la métrique près, les facteurs de l'analyse des correspondances des différents tableaux et du tableau moyen. On pourra ainsi repérer et interpréter, par exemple, des facteurs d'opposition entre tableaux, des facteurs communs à tous les tableaux, ou spécifiques de certains d'entre eux.

SUR DES APPROXIMATIONS D'INDICES DE DISSEMBLANCE
VIA LES REPRÉSENTATIONS EUCLIDIENNE ET HIERARCHIQUE

Auteur : FICHET Bernard
Laboratoire de Physique, Faculté de Médecine
Université d'Aix-Marseille II
13385 MARSEILLE CEDEX 4

Le problème de l'approximation d'un indice de dissemblance pour les représentations euclidienne et hiérarchique est étudié en termes de projection sur des cônes convexes.

Pour la représentation euclidienne, on considère une configuration de points dans un certain espace et une forme quadratique, non nécessairement positive, respectant l'indice de dissemblance donné. L'approximation est obtenue par projection de cette forme, au sens d'une certaine norme, dans le cône des formes quadratiques positives.

Pour la représentation hiérarchique, l'indice de dissemblance est approché par l'indice ultramétrique le plus proche au sens des moindres carrés, dans une certaine préordonnance ultramétrique ; cette dernière est la moins fine des préordonnances ultramétriques plus fines que la préordonnance associée à l'indice donné. La solution est obtenue analytiquement et un algorithme simple, reposant essentiellement sur un classement de nombres, permet de la construire.

REDUCTIONS INITIALE ET SEQUENTIELLE DES EXPERIENCES BAYESIENNES

Auteurs : FLORENS Jean-Pierre
C) RAMMISI
Faculté des Sciences Economiques
Université d'Aix-Marseille
5 avenue R. Schuman
13100 Aix-en-Provence

MOUCHART Michel
C.O.R.E.
Université Catholique de Louvain
34 voie du Roman Pays
B-1348 Louvain-la-Neuve
Belgique

Les réductions admissibles sont analysées dans le cas d'une expérience bayésienne séquentielle. De telles réductions sont obtenues à partir de tribus suffisantes et ancillaires. Elles sont considérées aussi bien dans le modèle global (c-à-d qui engendre n observations) que dans les modèles séquentiels (c-à-d qui engendrent l'observation i conditionnellement aux $(i-1)$ premières). Dans le cadre de problèmes d'inférence et de prédiction, on montre pourquoi il peut être utile de construire des réductions qui soient admissibles à la fois initialement et séquentiellement. On montre que le concept de transitivité apparaît naturellement lorsqu'on analyse sous quelles conditions une réduction séquentiellement admissible est aussi initialement admissible. On montre que cette condition de transitivité est également une condition nécessaire : elle est en effet impliquée par une réduction qui serait admissible à la fois initialement et séquentiellement. Enfin, on montre la relation entre le concept de transitivité et celui de "non-causalité" tel qu'introduit en économétrie.

PREVISION D'UNE SUITE DE TABLEAUX DE PROBABILITES SOUS CONTRAINTE D'UN SOUS-TABLEAU

Auteur : FOUcart Thierry
Institut Universitaire de Technologie
56000 VANNES

On utilise les propriétés des opérateurs que l'on peut associer à un tableau de données (1), en particulier le produit scalaire défini par la trace du produit de composition (2), et la méthode STATIS (3) pour analyser une suite de tableaux de probabilités et effectuer une prévision sous contrainte d'un sous-tableau.

Un exemple numérique montre quel genre d'applications on peut envisager.

Références :

- (1) F. CAILLEZ - J.P. PAGES : "Introduction à l'analyse des données" SMASH 1976.
- (2) Y. ESCOUFIER : "Echantillonnage dans une population de variables aléatoires réelles".
Thèse d'Etat, publication de l'ISUP Vol. XIX
fasc. 4 - p. 1-47, PARIS 1970.
- (3) H. L'HERMIER DES PLANTES : "Structuration des tableaux à trois indices de la statistique : théorie et application d'une méthode d'analyse conjointe".
Thèse de 3e cycle, Université des Sciences et Techniques du Languedoc, MONTPELLIER 1976.

ESTIMATION DES VALEURS PRISES PAR DES VARIABLES ALEATOIRES
DANS UN MODELE MIXTE : TECHNIQUES DE RESOLUTION DANS LE CAS
MULTIVARIABLE

Auteur : GOFFINET Bruno
I.N.R.A. - Laboratoire de Biométrie
Chemin de Borde Rouge
31320 CASTANET TOLOSAN

On présente deux techniques complémentaires pour estimer les valeurs prises par les variables aléatoires dans un modèle mixte multivariable.

Une technique classique réduit le problème aux nombres de niveaux des facteurs aléatoires.

Une autre méthode est proposée qui nécessite certaines conditions d'équilibre mais qui permet de traiter des lots de données plus importants.

Référence :

- (1) HENDERSON C.R. : "Sire Evaluation and Genetic Trends".
In Pro. Anim. Breed. Genet. Symp. Blacksburg - Virginia
p. 10-41. American Society of Animal Science
Champaign Illinois.

PRISE EN COMPTE DE L'INTERACTION GENOTYPE-MILIEU
EN SELECTION VEGETALE

Auteurs :

GOFFINET Bruno
I.N.R.A.-Laboratoire Biométrie
Chemin de Borde Rouge
31320 CASTANET TOLOSAN

VINCOURT Patrik
I.N.R.A.-Station d'Amélioration
des Plantes Fourragères
86600 LUSIGNAN

Pour prendre en compte l'interaction génotype x milieu en sélection végétale on a étudié le modèle :

$$Y_{ij} = \mu + G_i + \beta_j + C_j D_i + E_{ij}$$

où G_i et D_i suivent des lois normales non indépendantes.

On propose deux méthodes d'estimation des valeurs prises par G_i et D_i que l'on applique à un essai multilocal de maïs.

∞

Références :

- (1) FREEMAN G.H. : "Heredity" - Vol 31 - 339.354 - 1973.
(2) MANDEL J. : "The partitioning of interaction in analysis of variance".
Journal of Research of the National Bureau of Standards
B.73.B N°4 - 309.328.

ELABORATION ET EVALUATION D'UN GRAPHE D'IMPLICATION

CLASSIFICATION DANS LE TEMPS

Auteurs : GRAS R. et ROSTAM H.
Laboratoire de Statistique - IRISA -
Université de RENNES I
Campus de Beaulieu
35042 - RENNES Cédex

Auteur : G. GOVAERT
Université Paris Val de Marne
INRIA
Domaine de Voluceau
78150 Le CHESNAY

Disposant d'un ensemble de stimuli et d'un ensemble de réponses associées, le didacticien se pose à tout moment la question : tel comportement de réponse au stimulus i entraîne-t-il tel comportement de réponse au stimulus j ? Son objectif est alors de connaître les subordinations et de prévoir les enchaînements possibles.

Ayant une population E répartie en sous populations correspondant à T valeurs régulièrement réparties d'une variable continue (le temps, la profondeur,...), on cherche à faire T partitions (une pour chaque sous-populations) telle que l'évolution des centres soit polynomiale.

On ne cherche pas à classer des séries chronologiques mais on cherche des tendances ainsi que l'affectation de chaque individu (à chaque époque) à une de ces tendances.

Un nouvel indice d'implication permet de répondre à cette attente et de construire des graphes d'implication suivant les formes de cet indice. Par exemple, pour un test donné à des élèves, le graphe est l'image de l'ordre partiel des comportements de réussite. On analyse ce graphe en termes de contenu, de procédures et de complexité taxinomique. En particulier, la "proximité" entre le préordre à priori défini par la taxinomie sur l'ensemble des items et le graphe d'implication est révélée par un indice de comparaison entre deux relations pondérées sur le même ensemble.

Références :

- [1] R. GRAS : "Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques". Thèse d'Etat - Université de Rennes I - 1979.
- [2] I.C. LERMAN : "Etude formelle et statistique de la notion de ressemblance" - I.R.I.S.A. Publication interne n° 107 - 1978.

MODELES k-INVERSIBLES DE PROCESSUS MOYENNE-MOBILE

Auteur : HALLIN Marc

Université Libre de Bruxelles - Institut de Statistique
Campus Plaine - CP 210 - Boulevard du Triomphe
B-1050 BRUXELLES

Soit $b(z)$ la fonction génératrice associée à un modèle moyenne-mobile d'ordre q . Si l'équation $b(z) = 0$ possède k ($0 \leq k \leq q$) racines à l'intérieur du cercle unité (les q racines étant supposées distinctes et de module différent de 1), l'inversion de ce modèle fait intervenir k paramètres arbitraires ; nous dirons qu'il est k -inversible (l'inversibilité classique correspondant à la q -inversibilité). Il est alors évident que le modèle moyenne-mobile défini par la fonction génératrice $b(\frac{1}{z})$ est $(q-k)$ -inversible.

Cette notion et cette propriété sont fondamentales dans le problème de la construction d'un modèle pour une fonction d'auto-corrélation donnée. Nous les étendons au cas des processus à coefficients dépendant du temps. L'outil commode des fonctions génératrices doit cependant être remplacé par celui de système fondamental de solutions de l'équation aux différences homogène associée aux opérateurs moyenne-mobile.

MODELISATION DES PROFILS DE DENSITE OPTIQUE DE CHROMOSOMES HUMAINS EN VUE D'UNE RECONNAISSANCE AUTOMATIQUE DES ANOMALIES DE STRUCTURE DU CARYOTYPE.

Auteur : HAZOUT S.

Université de Paris 7.

Une étude précédente (1) avait porté sur la reconnaissance semi-automatique des anomalies chromosomiques humaines basée sur un décodage visuel des bandes obtenues par une technique de marquage (bandes R). Le programme sélectionnait les chromosomes normaux susceptibles d'intervenir dans les "solutions possibles" du remaniement, simulait toutes les transformations possibles (neuf types d'aberrations peuvent être reconnus) et comparait les images simulées à l'image analysée. De plus le mécanisme globale d'échange dans un groupe de chromosomes anormaux apparaissant dans un caryotype humain peut être recherché.

L'étude relative à cet exposé correspond à l'étape technique suivante qui concerne la modélisation des profils de densité optique des chromosomes humains dans le contexte précédemment évoqué. Nous comparons diverses techniques de représentation des profils (notamment les transformées de Fourier ; décomposition en somme de lois normales). Les problèmes de la reproductibilité des représentations et de la quantification des variabilités (inter chromatides, entre chromosomes homologues et entre cellules d'un même individu) sont discutés.

Référence :

- [1] S. HAZOUT, A.M. VENUAT, A.J. VALLERON et C. ROSENFELD :
Computer-Aided Analysis of Chromosomal Aberrations occurring
in an abnormal human karyotype. Hum. Genet. 49, 133-145 (1979)

INTERVALLES DE CONFIANCE SIMULTANES DE

TUKEY et SCHEFFE

Auteur : HUPET Georges

Unité de Calcul des Probabilités et d'Analyse Statistique
Université Catholique de Louvain
2, chemin du Cyclotron
1348 - LOUVAIN LA NEUVE - Belgique

Plusieurs auteurs ont analysé le problème des intervalles de confiance de Tukey et Scheffé dans le modèle de l'analyse de la variance en blocs aléatoires complets mais avec un choix particulier de la matrice de la covariance.

On propose ici d'aborder cette question de manière générale en recherchant toutes les matrices de covariance pour lesquelles on peut trouver une solution au problème posé.

On étudie aussi des procédures séquentielles pour des intervalles du type Scheffé de longueur fixée.

LE MODELE GAUSSIEN POUR LA SERIATION
DES DONNEES HISTORIQUES

Auteur: IHM Peter
Institut de Statistique Médicale
Philipps-Universität
D-3550 Marburg

La sériation est appliquée en archéologie et écologie pour ordonner des unités selon un gradient qui est - en archéologie - généralement le temps. Nous considérons une matrice H dont l'élément h_{ik} est égal à la fréquence du k -ième type dans le i -ième site. On suppose dans le cadre du modèle gaussien que l'espérance mathématique de h_{ik} soit égale à

$$\lambda_{ik} = E h_{ik} = R_i S_k \exp\left[-\frac{1}{2\sigma^2} (t_i - \mu_k)^2\right].$$

On suppose, en plus, que les h_{ik} ont des distributions de Poisson indépendantes avec paramètre λ_{ik} .

Le but est l'estimation des paramètres t_i , $\log S_k$ et μ_k qui sont définis sur une échelle d'intervalle.

Nous avons appliqué la méthode du maximum de vraisemblance. Le système d'équations non-linéaires est résolu par la méthode d'itération de Newton. Des valeurs initiales sont obtenues par l'analyse des correspondances. Différents cas particuliers du modèle seront discutés.

TABLES DE CONTINGENCE AVEC PARAMETRES ALEATOIRES

Auteur : IM SOTAN

Laboratoire de Biométrie - I.N.R.A.
Chemin de Borde Rouge
31320 CASTANET TOLOSAN

Laboratoire de Statistiques et Probabilités
Université Paul Sabatier
31077 TOULOUSE CEDEX

Lorsqu'une table de contingence est observée avec répétition, les probabilités des cellules peuvent varier d'une répétition à l'autre. Pour tenir compte de cette hétérogénéité et tenter de la mesurer, par analogie avec l'analyse de variance à effets aléatoires, on suppose que les probabilités des cellules sont des variables aléatoires ayant une distribution conjointe de Dirichlet dont on cherche à estimer les paramètres.

Techniquement le problème d'estimation débouche sur des méthodes itératives qui ont été programmées.

On a appliqué la méthode sur des données en génétique animale où divers élevages constituent les répétitions.

Références :

- (1) BISHOP Y.M.M. - FIENBERG S.E. - HOLLAND P.W. :
"Discrete Multivariate Analysis" MIT Press 1975.
- (2) CROWDER M.J. : "Inference about the intraclass correlation coefficient in the Beta-Binomial ANOVA for proportions"
J.R.S.S.B. Vol 41 N°2 - 1979.
- (3) GOOD I.J. : "The Estimation of Probabilities"
Cambridge, Mass., The MIT Press - 1965.

REPRESENTATION SOUS FORME DE VARIABLE D'ETAT D'UN MODELE DE REGRESSION A COEFFICIENTS VARIANT SUIVANT DES PROCESSUS STOCHASTIQUES - ESTIMATION ET PREVISION.

Auteur : INDJEHAGOPIAN Jean-Pierre
Ecole Supérieure des Sciences
Economiques et Commerciales
B.P. 105
95021 PONTOISE Cedex

Dans cette communication on se propose d'étudier le modèle de régression linéaire dans lequel le vecteur des coefficients varie suivant un processus stochastique linéaire non stationnaire. On montre que le modèle de régression et le modèle d'évolution du vecteur aléatoire des coefficients peuvent être représentés sous forme d'un modèle d'évolution markovien d'un vecteur d'état. On en déduit une procédure d'estimation des coefficients de régression et une méthode de prévision. On analyse ensuite les propriétés statistiques du prédicteur proposé. Enfin, on étudie le problème de l'estimation des coefficients lorsque les paramètres du processus stochastique caractérisant l'évolution du vecteur des coefficients de régression sont inconnus.

Références :

- AKAIKE H., "Markovian Representation of Stochastic Processes by canonical variables", *Siam J. of Control*, Vol. 13, 1975.
- COOLEY T. and E. PRESCOTT, "Estimation in the Presence of Stochastic Parameter Variation", *Econometrica*, Vol. 44, n° 1, Janv. 1976.
- RUCKEBUSCH G., "Représentation Markovienne de processus Gaussiens stationnaires et applications statistiques", Rapport interne n° 18, 1978, Centre de Mathématiques Appliquées, Ecole Polytechnique.

Auteur : INGENBLEK Jean-François
Université Libre de Bruxelles
Institut de Statistique
Campus Plaine - C.P. 210
Boulevard du Triomphe
B-1050 BRUXELLES

Il existe bon nombre de tests non paramétriques de bruit blanc (univarié); ces tests sont d'ordre p ($p = 1, 2, \dots$), en ce sens qu'ils établissent des bilans entre p rangs successifs. On présente une extension de ces tests permettant de traiter des variables multivariées.

L'accent est mis sur l'étude de tests simultanés d'ordres différents.

A. JACQUARD
Université de Paris VI
INED
4, place Jussieu
75005 PARIS

Notre esprit domine aisément les mécanismes additifs ; est-ce une particularité "innée" de notre structure mentale ou le résultat de l'entraînement que nous avons subi lorsqu'à l'école notre premier contact avec la "science" a été l'addition ? Avant même de savoir définir un nombre, nous savions que "deux et deux font quatre". Plus tard, notre apprentissage de la physique a surtout concerné les forces ; même si nous ne comprenions pas bien de quoi il s'agissait, nous savions que, si deux forces s'exercent simultanément en un même point d'un objet, elles "s'ajoutent" ; autrement dit, tout se passe comme si une seule force, égale à la somme vectorielle des deux premières, était en action. Conséquence de cette merveilleuse propriété, on peut, sans changer les données du problème, remplacer une force par deux ou plusieurs autres, pourvu que celles-ci aient une somme vectorielle égale à la première ; le choix astucieux de la direction ou de l'intensité de ces forces partielles, les "composantes", permet bien souvent de résoudre simplement un problème apparemment complexe, d'où les "solutions élégantes" dont nous félicitions nos professeurs du secondaire.

Le monde réel qui nous attend à la sortie de l'Université n'a malheureusement pas la merveilleuse structure simple de l'arithmétique à l'école communale, ou de la physique des forces au baccalauréat.

Méfions-nous de nos réflexes. La formulation même des questions que nous posons, à propos des phénomènes ou des processus que nous observons, risque d'impliquer une conception totalement irréaliste de ceux-ci ; les réponses à ces questions ne pourront alors être que des non-sens.

PROPOSITIONS D'UN PROFANE (SOCIOLOGUE NON STATISTICIEN) POUR L'ANALYSE DESCRIPTIVE DE "TABLEAUX STATISTIQUES (dits de contingence) MULTIDIMENSIONNELS" : méthode d'Analyse Structurelle des Interférences (A.S.I.), selon la technique de calcul booléen du "TREILLIS DISTRIBUTIF DES COEFFICIENTS DE RARETE", baptisés coefficients "upsilon".

Auteur : JENNY Jacques, alias "upsilon",
Centre d'Ethnologie Sociale et de Psychosociologie (C.E.S.P.),
1 rue du Onze Novembre, 92120 MONTRouGE

La technique de calcul proposée consiste à construire, pour chacun des sous-ensembles (classes, cases, ou cubes) d'un parallélépipède de contingence, à trois dimensions [I_xJ_xK] (voire à 4 ou 5 dimensions, dans les limites de ce qui est praticable et compte tenu de la taille de l'effectif N de l'ensemble), les treillis booléens distributifs à x générateurs (x=1, 2 ou 3, voire 4 ou 5, selon l'ordre de complexité où opère l'analyse), en prenant pour générateurs le log. de l'inverse des proportions, qu'on propose d'appeler les coefficients de rareté "upsilon": (i₋)= log N/i₋, etc..., et en prenant pour plus grand majorant: (i₋j₋k₋)=log N/i₋j₋k₋.

cf. par ex., ci-joint, le treillis distributif à trois générateurs (i₋), (j₋) et (k₋), avec les formules de calcul correspondantes, extrait de la référ.[1].

La méthode d'analyse, baptisée A.S.I., consiste à utiliser telle ou telle décomposition (partielle, le plus souvent) de tel ou tel coefficient de rareté des treillis, selon les exigences du programme de traitement des données élaboré "pas à pas" par le chercheur, et -si possible (grâce à l'assistance de micro-ordinateurs à écran de visualisation graphique)- en mode conversationnel... On a déjà écrit toute une gamme de programmes de calcul pour la T.I. 59, en attendant de pouvoir disposer d'un matériel plus performant...

On comparera cette méthode et cette technique à d'autres, plus ou moins usitées, aux objectifs analogues, et notamment à l'"analyse de l'incertitude" inspirée par la "théorie de l'information", qui opère "globalement" (c'est-à-dire sur les répartitions numériques en tant que telles) à l'aide de "coefficients d'entropie" qu'on appellera "UPSILON" (ils correspondent aux sommes pondérées des coefficients de rareté "upsilon"...) et non "localement", comme la méthode A.S.I.

Dans cette comparaison, on sera amené à constater certaines contradictions du "formulaire" de l'analyse de l'incertitude, qui correspondent précisément aux ambiguïtés signalées par ex. par QUASTLER [2], de même qu'elles correspondent par analogie aux ambiguïtés signalées par ex. par ROY, MITRA et KASTENBAUM [3] dans leurs travaux sur la décomposition du CHI-DEUX multiple.

Références :

- [1] J.JENNY : "La naissance et/ou le diplôme. Analyse Structurelle des Interférences observables dans une table de mobilité sociale intergénérationnelle à trois dimensions....." - Informatique et Sciences humaines - N°39 - Décembre 1978.
- [2] cité notamment par F. ATTNEAVE : "Applications of Information Theory to Psychology" - New York, Henry Holt and Co, 1959.
- [3] cités notamment par KENDALL et STUART : "The Advanced Theory of Statistics" - Vol. 2, pages 580 à 584 : "Multi way tables"

PROCESSUS ARMA A COEFFICIENTS DEPENDANT DU TEMPS -
UN ALGORITHME RAPIDE DE CALCUL DE LA FONCTION DE
VRAISEMBLANCE EXACTE

Auteurs : J.L. KIEHM
G. MELARD

Le calcul efficient de la fonction de vraisemblance exacte associée à une série chronologique générée par un processus ARMA (p,q) a fait l'objet d'articles récents ([1],[2]).

On présente ici un algorithme équivalent à celui de Ansley si $p \leq q$ mais qui lui est supérieur si $p > q$. de plus l'algorithme est court, ne nécessite pratiquement pas de place en mémoire et ne fait pas appel à d'autres algorithmes.

On compare l'efficience théorique par rapport à des méthodes approchées courantes et l'efficience effective avec les méthodes décrites dans [1] et [2].

[1] ANSLEY, Craig F. - "An algorithm for the exact likelihood of a mixed autoregressive-moving average process".
Biometrika (1979)66, pp.59-65.

[2] LJUNG, Greta M. and BOX, George E.P. - "The likelihood function of stationary autoregressive-moving average models",
Biometrika (1979)66, pp.265-270.

L'ESTIMATION DANS LES SERIES CHRONOLOGIQUES
EN VUE D'UN CHOIX DANS LA PREVISION.

Auteur : KHABIE-ZEITOUNE E.
Département de Mathématiques
North East London Polytechnic
Romford Road, Londres E. 15
Angleterre.

On établit, de façon exacte, des méthodes de calcul de diverses fonctions économiques à partir des paramètres de modèles ARMA (p,q) de série chronologique gaussienne à plusieurs dimensions.

Cette étude, prolongement d'un exposé de E. KHABIE-ZEITOUNE [2], permet une très grande souplesse dans le choix des fonctions économiques à utiliser dans l'approche des séries chronologiques introduite par G.E.P. BOX et G.M. JENKINS [1] (les fonctions économiques de la prévision n'étant, bien sûr, que des objectifs partiels dans les modèles plus étendus).

On considère, comme exemple d'application dans le domaine de gestion des stocks, le cas de demandes gaussiennes obéissant à un modèle ARMA (p,q) avec durée d'attente aléatoire. Ceci généralise, de façon exacte, certains résultats établis, en partie par simulation, par W.D. RAY dans le cas d'un modèle AR(1).

Références :

- [1] G.E.P. BOX et G.M. JENKINS : "Time Series Analysis" - Holden-Day, 1970.
- [2] E. KHABIE-ZEITOUNE : "Closed form Likelihood in Multivariate Weakly Stationary Time Series", dans "TIME-SERIES", Ed. O.D. ANDERSON - North-Holland, 1980.
- [3] W.D. RAY : "The significance of Correlated Demands and Variable Lead Time for Stock Control Policies" - The Journal of the Operational Research Society, Vol. 31, N°2, 1980.

UN LOGICIEL D'ANALYSE DES SERIES CHRONOLOGIQUES
UNIDIMENSIONNELLES.

Auteur : KHABIE-ZEITOUNE E.
Mathematics Department
North East London Polytechnic
Romford Road, LONDON E 15

Un logiciel d'analyse des séries chronologiques à une seule dimension est présenté. Ce logiciel est conçu pour le micro-ordinateur CBM(PET) de 32K de mémoire.

Ce logiciel facilite les phases interactives d'identification, d'estimation, de vérification et de prévision. L'identification des séries est basée sur l'étude des autocorrélations, autocorrélations partielles, autocorrélations inverses et sur la méthode "du coin" (corner method), ainsi que sur diverses analyses spectrales des autocorrélations. Les transformations instantanées sont permises, ainsi que les différences finies et saisonnières.

La méthode d'estimation utilisée est basée sur l'optimisation numérique de Fletcher-Reeves. Divers modèles et fonctions de vraisemblance peuvent être utilisés.

"MICRO CHADOC"
LOGICIEL D'ANALYSE DE DONNEES POUR
MICRO ORDINATEURS

Auteurs : D. LAFAYE DE MICHEAUX, J. LEMAIRE, J. POUGET
Laboratoire de Signaux et Systèmes (ERA 835)
Université de Nice - 41 Boulevard Napoléon III
06041 NICE CEDEX

S. NAZHAOUI, M. SOUKAL
DEA d'Informatique de l'Université de Nice

"MICRO CHADOC" (Chaîne d'Analyse de Données à Organisation Cohérente sur MICRO-ordinateur peut intéresser les chercheurs et les enseignants ne disposant que d'un micro-ordinateur).

Ecrite en BASIC sur un TEKTRONIX 4052, elle n'utilise de ce langage que les instructions non spécifiques. Sa structure modulaire et une normalisation des entrées/sorties facilitent la modification des programmes existants et l'inclusion de nouveaux programmes.

Les données et les commandes des différentes analyses sont enregistrées en mode conversationnel par un pilote qui sert d'interface entre l'utilisateur et les programmes de la chaîne.

0116

UNE APPROCHE ECONOMETRIQUE INDIVIDUALISEE
DES FINANCES LOCALES.

Auteur : LANCRY Pierre-Jean
Groupe de Recherches en Economie
de la Répartition
Université de Paris-Val-de-Marne
94210 LA VARENNE ST-HILAIRE

La collecte de données temporelles individualisées (panel data) permet d'étudier de façon intéressante les phénomènes de diversité du comportement individuel [1]. En effet grâce à ce type de données et aux nouveaux traitements résultants, on peut englober dans un même modèle l'information aux deux niveaux individuel et temporel tout en rendant compte des effets propres à chaque niveau [2].

Nous avons appliqué cette méthodologie à l'analyse du comportement financier et fiscal des trente-huit communes de plus de dix mille habitants qui constituent le département du Val-de-Marne [3]. Au delà d'une logique macroéconomique traditionnelle il nous apparaissait important d'aborder de manière plus individualisée les problèmes des finances locales.

Références :

- [1] C.N.R.S.-I.N.S.E.E. "The econometrics of panel data".
Annales de l'INSEE - n°30/31 - 1978.
- [2] MAZODIER P. & TROGNON A. "Données en coupes répétées et modèles à double indice" - miméo INSEE - 1975.
- [3] LANCRY P.J. "Une analyse individualisée des finances communales : le cas du Val-de-Marne (1966-1976)"
Université Paris XII - 1980

TREILLIS MESURES ET EFFET
D'INTERACTION EN SOCIOLOGIE

auteurs R. Langevin
Univ. de Paris XI-Orsay
J.C. Sifre
Univ. de Paris VII

Ce texte a été suscité par l'article de J. Jenny la naissance et/ou le diplôme. Nous reprenons ici le formalisme de Jenny pour en dégager la simplicité mathématique.

Pour chaque entier n nous pouvons définir un treillis distributif \mathcal{L}_n . Soit maintenant A_1, \dots, A_n n parties mesurables d'un ensemble \mathcal{E} muni d'une mesure de probabilité p . Nous montrons que la donnée des nombres $p(A_1)$, $p(A_1 \cap A_2)$, \dots , $p(A_1 \cap A_2 \cap \dots \cap A_n)$ permet de définir de façon unique une mesure sur \mathcal{L}_n .

Les mesures de certains éléments du treillis s'interprètent comme coefficients d'interaction multiple.

Nous proposons enfin d'utiliser la valeur absolue de la mesure ainsi définie comme "mesure d'oscillation".

Bibliographie.

- [1] J. Jenny. La naissance et/ou le diplôme. Informatique et sciences humaines. N° 39 dec. 1978 publication de l'institut des sciences humaines appliquées.
- [2] G. Grätzer. Lattice theory. W.H. Freeman and company San Francisco.

PRESENTATION DE PROGRAMMES DE STATISTIQUE
SUR MATERIEL HP. 9825.

Auteur : LANNUZEL Bernard
Laboratoire de micro-info-graphique
UER des Lettres et Sc. Humaines de Rouen
76130 MONT-SAINT-AIGNAN

Présentation de programmes d'Analyse des Données (A.C.P., Analyse des Correspondances, Classification Hiérarchique etc...) sur HP. 9825 A (+ table traçante 9872 A.)
Ces programmes sont actuellement utilisés sur des données géographiques.

ANALYSE D'INDICES D'IMPLICATION POUR DES DONNEES BINAIRES

INFERENCE FIDUCIO-BAYESIENNE

POUR LES COMPARAISONS A PLUSIEURS DEGRES DE LIBERTE

Auteurs : LECOUTRE Bruno et ROUANET Henry
Groupe de Recherche Mathématiques et Psychologie
Université René Descartes
Bureau 519, 12, rue Cujas
75005 PARIS

Les tests de signification, qui sont la procédure usuellement utilisée dans l'analyse de la variance, permettent de se prononcer sur l'existence d'un effet parent mais pas sur l'importance de cet effet.

Dans le cas d'une comparaison à plusieurs degrés de liberté, on pourra prendre comme indicateur de l'importance de l'effet observé une quantité proportionnelle à l'inertie associée à cette comparaison. La distribution fiducio-bayésienne sur le paramètre correspondant est obtenue pour les plans usuels et sous différents modèles (1). Ce résultat fait intervenir une distribution, appelée distribution du ψ , qui généralise les distributions du t de Student - qui intervient dans le cas de l'inférence fiducio-bayésienne sur un contraste entre moyennes (2) - et du χ^2 non centré.

Références :

- (1) B. LECOUTRE : "Procédures fiducio-bayésiennes pour l'analyse des données expérimentales". Thèse en cours de rédaction - Université René Descartes - Paris, 1980.
- (2) D. LEPINE, H. ROUANET : "L'inférence fiduciaire sur un contraste entre moyennes" - Cahiers de Psychologie - Vol. 18, p. 193-218 1975.

Auteur : LERMAN I.C.
Laboratoire de Statistique - I.R.I.S.A.
Université de RENNES I
Campus de Beaulieu
35042 - RENNES Cédex

Par rapport au problème du didacticien de la recherche d'une relation d'"implication" entre comportements de réponses à un ensemble de stimuli, nous analysons par rapport à différentes formes de l'hypothèse d'absence de lien, les différentes formes que peut revêtir un indice d'implication et nous étudions de façon comparative les propriétés de ces indices.

Un tel indice permet la définition d'un graphe orienté valué et sans cycles qui donne lieu à un graphe d'"implication" obtenu par la définition d'un seuil. On propose un critère d'évaluation de chacun des deux graphes (pondéré et discret) par rapport à un préordre total sur l'ensemble des sommets du graphe que définit une "taxinomie d'objectifs cognitifs".

La méthode mise au point a un caractère suffisamment général pour s'appliquer à d'autres champs que celui de la Didactique ; par exemple, recherche de sériations en Archéologie.

Références :

- [1] R. GRAS : "Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques". Thèse d'Etat - Université de Rennes I - 1979.
- [2] G. LECALVE : "Problèmes d'analyse des données" - Thèse d'Etat (2ème partie) - Université de Rennes I - 1976.
- [3] I.C. LERMAN : "Formal analysis of a general notion of proximity between variables" - Proceedings of European Congress of Statisticians - 1976.
- [4] I.C. LERMAN : "Etude formelle et statistique de la notion de ressemblance" - I.R.I.S.A. Publication interne n° 107 - 1978.

SUR UNE METHODE DE FISHER POUR
L'ETUDE DES "ACCORDS" DANS LES
TESTS DE PERCEPTION EXTRA-SENSORIELLE
(ESP) AVEC CARTES A JOUER.

Auteur* : LIGNON Yves
UER Mathématiques
Université de Toulouse-Le Mirail
31081 TOULOUSE Cedex

Après un rappel des résultats antérieurs [4] on présente la méthode inédite de R.A. Fischer [1] basée sur la loi $N(0;10)$ et qui permet de construire des tests ESP avec jeux de cartes ; en application on donne les résultats de Fisher et ceux, plus récents de Kanthamani et Kelly [2],[3].

En seconde partie on expose les recherches récentes effectuées à l'Université de Toulouse-Le Mirail à partir de la méthode de Fisher, on insiste en particulier sur les distributions d'échantillonnage (moyenne et écart-type) déjà obtenues.

Références :

- [1] R.A. FISHER : "A method of scoring coincidences in tests with playing cards". Proc.Soc.Psy.Res. Vol.34,181-5-1924.
[2] J.H. KANTHAMANI et E.F. KELLY : "Cards experiments with a special subject-Journal of Parapsychology". Vol.38 N°1, 16-26,1974.
[3] J.H. KANTHAMANI et E. KELLY : "On the relation between visual and ESP confusion structures". Jour.Am.Soc.Psy.Res. Vol.69 N°1,1-31,1975.
[4] Y. LIGNON : "Etude statistique des phénomènes de perception extra-sensorielle". Communication aux Journées de Statistique. Paris, ENSAE, 30 mai 1979.

* Avec la collaboration de Pierre LAGIER (informatique).

MOUJIK : UN LOGICIEL DE TRAITEMENT DES
SERIES TEMPORELLES SUR MICROORDINATEUR

Auteur : MALAMOUD Georges
Service d'Analyse Economique - Ministère des Transports
55, rue Brillat Savarin - 75013 PARIS
Equipe de Statistiques - Université Paris Sud
Bât. 425 - 91405 ORSAY CEDEX.

Ce logiciel permet d'effectuer les traitements suivants : créations, modifications, transformations, graphes, listes de séries ainsi que leur stockage dans des fichiers sur disquettes. Il permet en outre de lisser et désaisonnaliser les séries, d'ajuster des modèles de régression variés - multiple, avec autocorrélation des résidus, non linéaire, pondérée - d'utiliser le lissage triple exponentiel ainsi que les procédures de Box et Jenkins univariates complètes.

Son utilisation complètement conversationnelle et sa structure segmentée permettent de rajouter des modules nouveaux à tout moment et d'enchaîner sans avoir à recharger le logiciel.

Actuellement implanté sur matériel TEKTRONIX (série 4050), MOUJIK peut être facilement transporté sur d'autres microordinateurs acceptant le BASIC et possédant quelques caractéristiques minimales.

Ce logiciel est particulièrement adapté aux études économiques et à l'utilisation des méthodes de Box et Jenkins sur un nombre raisonnable de séries. Pour des modèles comportant beaucoup d'équations ou beaucoup de données, il vaut mieux utiliser les procédures traditionnelles de traitement par lots sur un ordinateur classique.

REGRESSION SOUS CONTRAINTES
ET CODAGES MONOTONES

Auteur : MALLET Jean-Laurent
Centre de Recherches Pétrographiques et Géochimiques
15, rue Notre-Dame des Pauvres
Case officielle n° 1
54500 VANDOEUVRE

On propose une technique de régression sous contraintes linéaires *quelconques* (1) permettant entre autre chose de construire facilement des codages optimaux *monotones* inspirés de (2).

A quelques facteurs correctifs près, d'expression très simple, la présentation adoptée conduit à un formalisme dont la particularité est d'être quasi identique à celui de la régression sans contraintes.

Références :

- (1) J.L. MALLET : "Régression sous contraintes linéaires : Application au codage des V.A." - Revue de Statistique Appliquée - à paraître - 1980.
- (2) G. SAPORTA : "Liaison entre plusieurs ensembles de variables et codage de données qualitatives". Thèse 3ème cycle - Paris - 1975.

QUELQUES REMARQUES SUR LES CODAGES FLOUS

Auteur : J.F.MARTIN
Laboratoire de Statistique et Probabilités
Université Paul Sabatier
31077 TOULOUSE Cedex

On est souvent amené, en statistique, à discrétiser des variables aléatoires réelles.

Le découpage en classes, suivi de l'attribution de chaque observation à une classe précise (codage disjonctif) présente, dans certains cas, les inconvénients bien connus : perte d'information, découpage arbitraire, introduction artificielle de discontinuités.

Plusieurs auteurs utilisent des méthodes empiriques pour pallier ces inconvénients, consistant à attribuer à une observation des poids d'appartenance aux classes de la partition (codage flou).

On tente de synthétiser ces méthodes, de juger de leur utilité dans diverses situations, d'étudier leur stabilité et leurs limites.

Référence : J.F. MARTIN - Une approche des codages flous ; quelques propriétés - Publications du Laboratoire de Statistique et Probabilités, 1980, Toulouse.

ESTIMATION DES COMPOSANTES DE LA VARIANCE
SUR MODELE MIXTE : REVUE ET BILAN

Auteur : MASSON J.P.

ENSAR, 65, rue de Saint Briec
35042 RENNES

On procèdera à une revue des différentes méthodes d'Estimation des Composantes de la variance sur Modèle mixte :

- a) Méthode d'HENDERSON
- b) Méthodes de C.R. RAO
- c) Méthodes liées au maximum de vraisemblance
 - C.1 Maximum de vraisemblance
 - C.2 Maximum de vraisemblance restreint

On indiquera les logiciels disponibles ; ceux notamment écrits à l'INRA et permettant d'utiliser les méthodes : HENDERSON III (a/), b, c-2
En conclusion après avoir comparé succinctement les différentes approches on fera une rapide revue sur l'estimation quadratique (des composantes de la variance) biaisée et assurant la positivité.

SUR LA CONVERGENCE DE PROCESSUS DE VRAISEMBLANCE
EN VARIABLES MARKOVIENNES.

Auteurs : MILHAUD X.

Laboratoire de Statistique et Probabilités
Université Paul Sabatier
31077 TOULOUSE Cedex

OPPENHEIM G. - VIANO M.C.
U.E.R. de Mathématiques, Logique Formelle
et Informatique - Université PARIS V
75005 Paris

On établit la convergence faible de la suite de processus de vraisemblances $(dP(X_1, \dots, X_n, \theta + t/\sqrt{n}) / dP(X_1, \dots, X_n, \theta), t \in \mathbb{R})$, $n \in \mathbb{N}$ vers un processus Z , dans l'espace $(C_0(\mathbb{R}), \text{Sup})$. Les résultats sont établis pour une classe de Markov homogènes récurrents au sens de HARRIS, lorsque la loi initiale du processus est une loi de DIRAC. Des résultats semblables ont été obtenus antérieurement pour des Markov stationnaires.

Références :

IBRAGUIMOV I.A., KHASMINSKI R.Z. : "Asymptotic behaviour of statistical estimators in the smooth case. I - Study of the likelihood Ratio - Theory of Probability and its applications, 1972, XVII n°3, p.445-462.

ANALYSE DES ESTIMATEURS LINEAIRES DE BAYES
DANS QUELQUES CAS D'APPLICATION.

par Michel MOUCHART *
et Léopold SIMAR **

Cet article présente quelques applications de résultats obtenus dans MOUCHART - SIMAR (1979) sur l'approximation, au sens des moindres carrés, d'espérances a posteriori.

L'utilisation de statistiques judicieusement choisies pour construire ces approximations permet d'obtenir des formes simples où le rôle de certaines hypothèses apparaît clairement. Les cas concrets analysés sont l'estimation d'une moyenne, le problème de la prédiction, le modèle linéaire et l'estimation de fonctions de distribution.

Référence : MOUCHART M. and L. SIMAR (1979), Least squares Approximations in Bayesian Analysis. Cahiers du S.M.A.S.H. n° 7905, Facultés Universitaires Saint-Louis, Bruxelles (à paraître dans Trabajos de Estadística).

* C.O.R.E., Université Catholique de Louvain, Louvain-la-Neuve.

** Facultés Universitaires Saint-Louis, Bruxelles.

DESIRABLE QUALITIES IN STATISTICAL SYSTEMS,
WITH SPECIAL REFERENCE TO GENSTAT.

Auteur : J.A. NELDER

Rothamsted Experimental Station
Harpenden, Herts, AL5 2JQ
ENGLAND

The following aspects will be discussed : data structures supported ; algorithms ; flexibility of interpretive language ; requirements for multi-stage analysis ; ease of communication with other programs. Examples will be drawn mainly from Genstat and Glim.

CLASSIFICATION SOUS CONTRAINTE DE CONTIGUITE CONTINUE

Christophe PERRUCHET
CISI - 35 Bd Brune - 75015 PARIS

Les techniques de Classification Automatique permettent de répartir un ensemble de données en groupes homogènes. Leur champ d'application est donc vaste.

Il peut être nécessaire de tenir compte des proximités induites par une relation de contigüité dans l'ensemble à classifier, ce que permet la "Classification sous Contrainte de Contigüité Continue".

En pratique, la classification obtenue dépend du choix d'un seuil de contigüité représentant la distance au delà de laquelle deux points ne peuvent être agrégés. L'estimation statistique de ce seuil s'est révélée extrêmement satisfaisante et opérationnelle.

D'autre part, des règles relatives à la sélection de partitions et à la recherche d'anomalies facilitent la tâche du praticien dans l'interprétation d'une classification.

Enfin, une application de ces techniques à des données géochimiques fournies par le Bureau de Recherches Géologiques et Minières est présentée.

LONGEVITE ET PROFIT : UN MODELE STOCHASTIQUE
D'ACTIVITE COMMERCIALE.

Auteur : PINHAS Max
Université des Sciences Sociales
Place Anatole France
31070 TOULOUSE Cedex

Le modèle stochastique d'activité commerciale exposé dans la présente note tire son caractère opérationnel de certains résultats de la Théorie du Risque Collectif.

Parmi les utilisations variées qu'il offre nous avons choisi de mettre l'accent sur celle qui a trait à l'emploi. Plus précisément nous pensons qu'à travers cette maquette une information quantitative claire peut être dégagée de nombreuses situations concrètes, qui aide à résoudre le délicat arbitrage profit-longévité de l'entreprise.

TESTS SUR LA LOI D'UN PROCESSUS PONCTUEL.

Auteur : PONS Odile
I.N.R.A. - Biométrie
78350 Jouy-en-Josas
Equipe CNRS 532 - Statistique Appliquée

Observons au cours du temps les réalisations d'un processus ponctuel réel de dimension k et paramétré, (N, P_θ) . Au vu de ses réalisations nous proposons des tests sur la loi P_θ quand elle est absolument continue par rapport à celle d'un processus de Poisson généralisé P .

La théorie des processus stochastiques et des martingales nous permet d'introduire la notion de compensateur prévisible \tilde{N} de N ; nous supposons que $\tilde{N}^{\theta^*} = \int X \theta d\tilde{N}$ où X est un processus prévisible et θ une fonction càg lād. Quand deux composantes de θ sont proportionnelles nous donnons un test optimum de leur égalité, et plus généralement nous éprouvons la valeur de θ sur un intervalle donné, tout en calculant la puissance as. de ces tests.

C'est ainsi que l'on peut comparer des processus markoviens ou semi-markoviens, des processus censurés etc; il apparait un grand nombre d'applications.

Références :

- [1] O. AALEN : An. Stat., 6, 1978, p. 701.
- [2] J. JACOD : Z. Wahrsh., 31, 1975, p. 235.
- [3] R. REBOLLEDO : Lect. Notes Math., 636, Springer verlag, 1977.

ECHANTILLONNAGE REPRESENTATIF
ET DIMENSION INTRINSEQUE

Auteur : QUINQUETON Joël
I.N.R.I.A.
B.P. 105
78150 ROCQUENCOURT

De nombreuses méthodes pour obtenir une estimation de la dimension intrinsèque ont déjà été proposées dans la littérature.

Nous essayons d'utiliser certaines de ces méthodes pour évaluer le caractère représentatif d'un sous ensemble de points de départ.

Les méthodes obtenues, quoique incertaines, nous semblent présenter un certain intérêt pour la simplification de problèmes d'analyse de données sur de grands ensembles de points.

ESTIMATION EFFICACE POUR UN PROCESSUS AUTOREGRESSIF
EXPONENTIEL A DENSITE DISCONTINUE.

Auteur : RAFTERY Adrian E.
Laboratoire de Statistique Mathématique
Université de Paris VI
75230 PARIS cedex 05

On considère un modèle linéaire markovien pour des suites de variables aléatoires exponentielles corrélées, qui est analytiquement traitable et facile à simuler. Il peut servir comme modèle aussi bien pour les séries temporelles non-gaussiennes que pour les processus ponctuels aux intervalles dépendants. Il a des applications entre autres en climatologie statistique (débits des rivières, vitesses des vents) et en recherche opérationnelle (fiabilité, files d'attente avec corrélation, systèmes informatiques).

La densité conditionnelle des observations est discontinue et de ce fait, ainsi que de la dépendance des observations, la théorie classique de l'estimation efficace ne s'applique pas. On propose un estimateur des paramètres dont on démontre la consistance forte et l'efficacité asymptotique. On donne également sa relation avec l'estimateur de maximum de vraisemblance. Pour un des paramètres l'estimateur est "super-efficace" et pour les autres il est asymptotiquement normal.

BASES GENETIQUES ET STATISTIQUES DES INDEX
DE SELECTION

Auteur : RAZUNGLES J.

I.N.R.A. Station d'Amélioration Génétique
des Animaux
B.P. 12 - 31320 CASTANET-TOLOSAN

Le potentiel héréditaire d'un individu n'est pas observable. Un index de sélection attribué à un individu permet de prédire ce potentiel. Il est actuellement défini comme une forme linéaire de variables aléatoires observables dont les coefficients sont déterminés par différentes méthodes. Cependant cette forme ne correspond pas à une relation de parenté quelconque entre l'individu du "prédit" et les individus "prédicteurs".

L'index sera défini comme l'espérance conditionnelle de la valeur zygotique d'un individu aux valeurs phénotypiques de l'individu lui-même et de ses apparentés.

Nous établissons la forme explicite de l'index pour un prédicteur consanguin dans le cas général d'une loi quelconque du milieu pour des caractères régis par une ou plusieurs lois indépendantes. Nous particularisons cette forme à des milieux où l'influence peut se traduire par une loi normale ou une loi Gamma.

L'établissement de la loi conjointe a priori des génotypes de trois individus non consanguins nous permet ensuite par récurrence de déterminer la forme générale d'un index pour des prédicteurs non consanguins lorsque l'influence individuelle des gènes sur la variabilité du caractère et le milieu quelconque.

PARTITIONS CENTRALES ET RELATIONS CENTRALES

Auteur: REGNIER Simon

Service de Mathématiques Appliquées et de Calcul
Maison des Sciences de l'Homme
54, boulevard Raspail
75270 PARIS Cedex 06

L'auteur donne une revue des résultats que d'autres ont développé autour du concept de partition centrale:

- 1° Sont-elles effectivement calculables par une programmation linéaire en nombres entiers.
- 2° Comment étendre leur définition et leur calcul lorsque l'ensemble à classer est infini.
- 3° Comment associer un centre de même nature à une "préordonnance" sur l'ensemble des paires d'objets.
- 4° Dans quelle mesure cette notion s'étend-elle à d'autres problèmes d'agrégation d'opinions.

Références:

- (1) S. REGNIER: "Sur quelques aspects mathématiques des problèmes de classification automatique" - I.C.C. Bulletin, 4, Rome - 1965.
- (2) A. de FALGUEROLLES: "Classification automatique. Emploi généralisé de l'algorithme de S. Régnier" - mai 1975.
- (3) J.-P. BARTHELEMY: "A propos des partitions centrales sur un ensemble non nécessairement fini" - mars 1977.
- (4) J.-F. MARCOTORCHINO - P. MICHAUD: "Optimisation en analyse ordinale des données" - Masson - 1979.
- (5) J.-P. BARTHELEMY - B. MONTJARDET: "Ajustement et résumé de données relationnelles: les relations centrales" - février 1979.

Auteur : REY Jean-François
INSTITUT SUPERIEUR AGRICOLE
Rue Pierre Wagué
60026. BEAUVAIS CEDEX

Le logiciel présenté a été conçu en cherchant à exploiter les facilités offertes par le micro-ordinateur : souplesse et simplicité d'exploitation, choix conversationnel des traitements et résultats progressifs. Il réalise de manière totalement intégrée :

- GESTION DE TABLEAUX DE DONNEES : entrées, corrections, extractions sur critères logiques, regroupements, combinaisons de transformations diverses, construction de tableaux de contingence, éditions et reports graphiques des résultats.
- DEPOUILLEMENT DE TABLEAUX.
n questions croisées et caractérisation des sous-groupes obtenus.
- CORRELATIONS : simples, multiples, résiduelles, canoniques.
- ANALYSES DE DONNEES : Composantes principales, Factorielles des correspondances, Données centrées, Analyse discriminante.

Ecrit en BASIC pour un WANG 2200 T ou un WANG VP (clavier-écran, imprimante de 132 c, 2 disquettes de 256 K chacune), ce logiciel est construit pour traiter et interpréter des tableaux de données relativement importants :

- . avec 12 K utiles de mémoire centrale, les matrices de dispersion ou de corrélation d'origine admettent une quarantaine de variables ;
- . interruptions et reprises des traitements prévues.

Cet ensemble de programme, conçu avec la collaboration de Mr J. LEFEBVRE (Laboratoire de Génétique Factorielle - C.N.R.Z. - Jouy-en-Josas), est utilisé en self-service par les étudiants de l'I.S.A.B. depuis 3 ans.

Référence : J. LEFEBVRE - "Introduction aux Analyses Statistiques Multidimensionnelles" - Masson - 1980.

LA VARIANCE DU COEFFICIENT DE CORRELATION ENTRE DEUX
PROCESSUS SPATIAUX AUTOCORRELÉS.

Auteurs : RICHARDSON Sylvia, HEMON Denis
INSERM - U170
16 bis, Avenue Paul Vaillant Couturier
94800 VILLEJUIF

Une approximation de la variance du coefficient de corrélation empirique entre deux processus spatiaux définis sur un réseau régulier est calculée sous l'hypothèse d'indépendance mutuelle des deux processus. La variance est évaluée dans trois cas particuliers : processus markoviens, autorégressifs et doublement géométriques. Nos résultats montrent que la variance du coefficient de corrélation empirique est très sensible à l'existence d'auto-corrélations. Ils suggèrent par ailleurs un facteur de correction de la taille de l'échantillon.

ANCILLARITE ET SUFFISANCE EN ANALYSE DISCRIMINANTE

Auteur : ROLIN Jean-Marie
Unité de Calcul des Probabilités et d'Analyse Statistique
Université Catholique de Louvain
2, Chemin du Cyclotron
1348 - LOUVAIN-LA-NEUVE - Belgique

Le problème de sélection globale et de sélection pas à pas de variables en analyse discriminante est formellement analogue au problème de la recherche de statistiques ancillaires et de statistiques suffisantes pour le paramètre indexant les sous-populations, respectivement dans le modèle général et dans le modèle conditionnel.

Dans le cadre de l'analyse de la variance multivariée, le rapport de vraisemblance généralisé permet d'obtenir les statistiques pour éprouver l'hypothèse de suffisance et l'hypothèse d'ancillarité d'une transformation linéaire arbitraire des variables. Ces statistiques sont ensuite utilisées pour la détermination des variables les plus discriminantes et fournissent des règles d'arrêt dans la sélection de celles-ci.

QUELQUES PROBLEMES INFERENTIELS LIES A
L'ANALYSE EN COMPOSANTES PRINCIPALES (A.C.P.)
D'UNE FONCTION ALEATOIRE (f.a.)

Auteur : ROMAIN Yves
Laboratoire de Statistique et Probabilités
Université Paul Sabatier
31077 TOULOUSE Cedex

A la suite des résultats obtenus dans l'étude asymptotique de l'A.C.P. d'une f.a., on propose ici quelques applications inférentielles dans le cas d'une f.a. gaussienne.

On s'intéresse notamment à l'estimation (ponctuelle et par intervalle) d'éléments propres de l'A.C.P. et à des tests d'hypothèses : par exemple, test de sphéricité (égalité des valeurs propres), test d'identification d'un facteur principal (ou "harmonique")...

Références : [T.W. ANDERSON] - "Asymptotic theory for Principal Component Analysis" Ann. Maths. Statist. Vol.34, n°1, 1963.

[Y. ROMAIN] - "Etude asymptotique des approximations par échantillonnage de l'analyse en composantes principales d'une fonction aléatoire. Quelques applications".
Thèse de Spécialité, Toulouse 1979.

UNE CHAÎNE D'ANALYSE DE QUESTIONNAIRES
SUR MICRO-ORDINATEUR HEWLETT PACKARD 9825

Auteur : ROSENZVEIG Christian
Laboratoire de Statistique Mathématiques
Université Pierre et Marie CURIE (PARIS VI)

Une chaîne complète de traitement effectuant l'analyse des correspondances d'un questionnaire est présentée. Cette chaîne qui comporte 21 modules d'encombrement minimum, utilise en particulier, dans l'espace de dimension r des facteurs une base orthonormée pouvant être conservée dans un tableau à deux lignes et r colonnes.

Références :

- (1) "EISPACK guide" (1974) Matrix Eigensystem Routines. Lecture Notes in Computer Science, n° 6, Springer Verlag
- (2) WILKINSON J.H. (1965) The Algebraic Eigenvalue Problem Clarendon Press, Oxford
- (3) WILKINSON J.H., REINSCH C. (1971) Handbook for Automatic Computation. Vol. 2, Linear Algebra, Springer Verlag

MODELE D'ESTIMATION DE LA VALEUR GENETIQUE
DES ANIMAUX REPRODUCTEURS.

Auteur : ROUVIER Roger
Station d'Amélioration Génétique des Animaux
Centre de TOULOUSE - INRA
31320 CASTANET TOLOSAN - B.P. 12

Il s'agit d'un exposé de réflexion sur les méthodes d'estimation de la valeur génétique additive des reproducteurs, pour un ou plusieurs caractères, et compte tenu des objectifs du sélectionneur. L'on s'est efforcé de montrer dans la théorie maintenant classique (ROUVIER, 1969, 1977) et dans le cas de loi de distribution multinormale des aléatoires considérées, l'utilisation conjointe des concepts de la génétique quantitative issus des travaux de R. FISHER et des méthodes de la statistique mathématique: partition de la variance génotypique, expression de la covariance entre phénotypes et génotypes d'apparentés, régression linéaire multiple et analyse de variance. Les objectifs de la sélection sur plusieurs caractères et la prise en compte d'interaction génotype×milieu dans la sélection sur les valeurs additives sont également discutés. Plusieurs références permettant d'analyser l'évolution de la pensée ou indiquant des applications à l'amélioration génétique animale sont données.

Références

- ROUVIER R., 1969. Contribution à l'étude des index de sélection sur plusieurs caractères. Thèse 3è cycle, Fac. Sci. Paris.
- ROUVIER R., 1977. Mise au point sur le modèle classique d'estimation de la valeur génétique. Ann. Génét. Sél. anim. 9 (1), 17-26.

SUR L'UTILISATION DE METRIQUES
EN ANALYSE HARMONIQUE
APPLICATION A L'ETUDE D'UN
ENSEMBLE DE COURBES

Auteur : SAPORTA Gilbert
Institut Universitaire de Technologie
143, Avenue de Versailles
75016 PARIS

L'analyse harmonique introduite en [3] pour l'étude d'un processus X_t est la généralisation de l'ACP usuelle, avec $M = I$, à une famille continue de variables aléatoires. L'introduction d'une métrique quelconque sous forme d'un opérateur intégral, pour mesurer la distance entre deux trajectoires permet :

- 1) d'étendre à un processus les résultats classiques de l'ACP. On explicitera les notions d'axes, de facteurs, d'inertie et les équations à résoudre ; la généralisation du schéma de dualité ayant été faite en [2].
- 2) de prendre en compte et généraliser des transformations linéaires usuelles en analyse des séries chronologiques (différences premières, moyennes mobiles ...)
- 3) de répondre aux nécessités d'approximation et d'interpolation voir [1].

Diverses propositions sont alors faites concernant l'étude d'un ensemble de n courbes sur $[0 ; T]$ selon qu'on les considère comme n trajectoires (courbes "individus") ou n mesures sur une population d'instantants (courbes "variables").

- [1] P. BESSE "Etude descriptive d'un processus". Thèse de 3e cycle - Université Paul SABATIER - Toulouse - 1979.
- [2] J. DAUXOIS - A. POUSSE "Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique". Thèse - Université Paul SABATIER - Toulouse - 1976.
- [3] J.C. DEVILLE "Méthodes statistiques et numériques de l'analyse harmonique" Annales de l'INSEE n° 15 3-101 - 1974.

OPERATEURS LOGIQUES EN CLASSIFICATION

Auteur : SIDI Jacqueline

INRIA
Domaine de Voluceau
78150 Le CHESNAY

Il s'agit de faire une classification d'un ensemble d'objets caractérisés par des variables qualitatives. Contrairement à l'habitude chaque objet peut prendre simultanément plusieurs modalités de la même variable ; cela nous amène à utiliser des formules logiques.

Nous appelons formule logique une conjonction de prédicats élémentaires du type $[X \# A]$, dans lequel X est le nom de la variable, A une valeur ou un ensemble de valeurs, # un opérateur logique (par exemple : >, <, =). Le prédicat $[X \# A]$ est vrai si et seulement si la valeur prise par la variable X est en relation # avec l'ensemble A.

Par exemple : une formule s'écrit sous la forme

$$[X^1 = A^1] \wedge [X^2 \neq A^2] \wedge [X^3 > A^3] \wedge [X^4 < A^4]$$

Nous utilisons l'algorithme des Nuées Dynamiques avec Distances Adaptatives sur un ensemble de données, caractérisées par un nombre fini de variables qualitatives, en utilisant des formules logiques comme noyau.

Utilisant les résultats ainsi obtenus, nous appliquons un algorithme de type agglomératif afin d'obtenir une description de chaque classe sous forme de formules logiques.

L'utilisation des formules logiques nous permet de mettre en valeur les relations existantes entre les différentes variables afin de pouvoir classer de nouveaux éléments.

ETUDES DU GRAPHE DES CHEMINS POUR DES RESEAUX DE CONTROLE LOGIQUE

Auteur : SNOUSSI E.H. et GOLES CHACC E.

Laboratoire IMAG Mathématiques Appliquées
Université Scientifique et Médicale
B.P. 53 X
38041 GRENOBLE Cedex - FRANCE

Un système de contrôle génétique peut être considéré comme un réseau d'éléments fonctionnellement connectés par des interactions positives ou négatives. Un tel système est décrit par une structure logique de la manière suivante : à chaque gène S on associe trois variables booléennes R, Γ et ρ . R exprime la présence ($R=1$) ou l'absence ($R=0$) du gène S ; Γ décrit l'expression du gène quand celui-ci est présent ; l'expression du gène S conduit à l'apparition du produit ρ (protéine). Ainsi on peut décrire un système de contrôle génétique soit par le graphe orienté décrivant les interactions soit par un système d'équations logiques du type :

$$\begin{aligned} \Gamma_i &= R_i f_i(\rho_1, \dots, \rho_n) \quad i=1,2,\dots,n \\ \rho_i &= D(\Gamma_i) \end{aligned}$$

Cette modélisation booléenne a été étudiée par R. Thomas qui a décrit ainsi, le comportement dynamique d'une boucle et d'une connexion de deux boucles. Notre travail consiste à formaliser ce modèle pour pouvoir le généraliser au cas de connexion de plusieurs boucles avec des fonctions logiques f_i du type "ET", "OU", "OU-exclusif" et fonctions à seuil.

Un tel formalisme prend tout son intérêt dans l'étude de l'expression conjointe de nombreux gènes en interaction : il permet en particulier de prévoir la nature (état stationnaire, ou cycle, stable ou instable) des attracteurs de la dynamique du processus d'expression génétique.

PROPRIETES D'UNE REGLE DE SUBSTITUTION DE L'ANALYSE DISCRIMINANTE
MULTIVARIEE.

Auteur: STREIT Franz
Section de Mathématiques
Université de Genève
2, Rue du Lièvre
CH - 1211 GENEVE 24

En appliquant les méthodes de l'analyse discriminante paramétrique il arrive souvent que les paramètres des distributions à comparer ne sont que partiellement connus. Ce fait complique considérablement la recherche d'une règle d'identification optimale et l'étude de ses propriétés.

Le but de ce travail est d'examiner le comportement d'une règle de substitution pour le problème d'identification suivant: Il s'agit d'attribuer une unité à l'une de deux populations G_1 et G_2 , qui suivent une distribution Gaussienne de dimension k de vecteur moyen connu μ_1 resp. μ_2 [$\mu_1 \neq \mu_2$] et de matrice de covariance inconnue $\Sigma_1 = \Sigma_2 = \Sigma$. Des expressions analytiques explicites pour les probabilités de classification incorrecte à G_1 resp. G_2 sont indiquées. La détermination des valeurs numériques de ces quantités pour différents choix de μ_1 , μ_2 , Σ et k (taille de l'échantillon qui sert à estimer Σ) donnent des indications sur les risques entraînés par l'utilisation de ce procédé.

CODAGE ET DISTANCES ADAPTATIVES
EN CLASSIFICATION AUTOMATIQUE

Auteur : F. TALENG
INRIA
Domaine de Voluceau
78150 Le CHESNAY

En classification automatique on considère généralement que la mesure de ressemblance choisie ainsi par le codage des variables sont généralement fixés, les algorithmes sont alors utilisés sans remettre en question ces choix.

Nous présentons une famille de critères à optimiser dépendant de la partition cherchée, du codage et de la mesure de ressemblance. Les algorithmes proposés font évoluer le codage et la mesure de ressemblance en améliorant leur adéquation à la partition cherchée au sens du critère à optimiser, par amélioration itérative jusqu'à obtention d'une solution stable.

L'intérêt pratique de cette approche est double, d'une part elle permet de traiter des tableaux hétérogènes (mélange de variables qualitatives et quantitatives) d'autre part elle donne à l'utilisateur une information complémentaire sur ses données par l'interprétation que l'on peut faire du nouveau codage obtenu.

ALGORITHME DE NUAGE DYNAMIQUE SUR UN TABLEAU
DE DISTANCES NON EUCLIDIENNES.

Auteur : THAURONT Gérard

Institut de Recherche des Transports
2, avenue du Cal Malleret-Joinville
B.P. 34
94 114 - ARCUEIL-CEDEX

Lorsque l'on ne dispose, sur un ensemble à classifier, que d'un tableau de distances, et en particulier lorsque la notion de coordonnées n'existe pas, l'écriture d'un algorithme de Nuée Dynamique nécessite une notion de centre et une notion de dispersion calculable uniquement à partir des distances.

Nous proposons plusieurs solutions et leur application au cas de distances entre paire de points.

Ce travail a été effectué à l'occasion d'un rapport de thèse de 3ème cycle à l'université Pierre et Marie-Curie - Paris 6è

ANALYSE DE DONNEES SUR MICRO-ORDINATEUR

Auteur : TIXIER Jean-Michel

Département de Mathématiques Appliquées
Complexe Scientifique des Cégeaux
Boite Postale N°45
63170 AUBIERE

L'Analyse des Données offre un large éventail d'outils descriptifs dont les qualités sont, cependant, conditionnées par le mode de représentation choisi. Pour tenter d'améliorer celle-ci, on peut utiliser des affichages couleurs sur écran, par exemple, avec un équipement aussi modeste que le micro ordinateur APPLE II connecté avec un poste T.V. couleur.

Ainsi en Analyse en Composantes Principales, on remplace les identificateurs utilisés par une imprimante, pour représenter les projections sur les plans principaux, par des petits rectangles de couleurs. De plus si on introduit une classification sur les individus, en affichant plusieurs de ceux-ci dans une même couleur, il est possible de tester visuellement cette classification, en remarquant s'il apparaît des "tâches" de couleur. D'autre part, grâce à une structure conversationnelle du programme on obtient plusieurs affichages qui sont une aide à l'interprétation du phénomène étudié.

En classification hiérarchique, outre un tracé très net des dendrogrammes, on peut visualiser les noeuds significatifs (cf. Lermann, rapport Irisa N° 70) de la hiérarchie. Ceci est difficilement réalisable avec une imprimante qui ne donne généralement que la valeur du critère considéré.

On peut donc dire que si l'utilisation de la couleur ne change pas les méthodes d'Analyse de Données, elle permet par contre à l'utilisateur une meilleure visualisation du problème qui le préoccupe.

CATASTROPHE IN TIME SERIES ANALYSIS?

Author: Dr Howell TONG
Department of Mathematics
University of Manchester Institute of Science & Technology
Manchester
M60 1QD
England

In this paper, we make the empirical observation that the much analysed Canadian lynx data has an estimated bivariate probability density function for chronologically adjacent observations in the shape of a crater. Based on this seemingly overlooked evidence, we argue that the recently developed threshold autoregression [1] is a natural quantitative expression of René Thom's catastrophe theory in the empirical science of time series modelling.

References

- [1] H TONG & K S LIM: "Threshold autoregression, limit cycles and cyclical data". Paper read to the Royal Statistical Society on 19 March 1980. Paper with discussions to appear in J Roy Statist Soc (B) 42 (1980).

UNE ETUDE DES PROCESSUS ARIMA DU POINT DE VUE OPERATORIEL

Auteur : TRUONG-VAN-MANH Benoît
Laboratoire de Statistiques et de Probabilités
Université Paul Sabatier
31062 TOULOUSE Cédex

On donne une étude sur la nature de l'opérateur retard B_X associé à des processus ARIMA. Ces processus ainsi appelés par [1] sont des solutions non stationnaires d'un type important d'équations aux différences (voir [1] et [2]).

En particulier d'une part on montre que ces processus ne sont pas linéarisables, contrairement à ce qui est affirmé dans [1] et que leur opérateur retard B_X ne peut pas être linéaire et de norme inférieure à l'unité, d'autre part on attire l'attention sur la propriété que la nature de B_X (linéarité ou/et continuité) donc des processus ARIMA, eux-mêmes, dépend fortement du choix de leurs conditions initiales.

Références :

- [1] G. BOX - G. JENKINS : "Time Series Analysis, forecasting and control" - Holden Day - San Francisco-1976.
[2] G.J. HANNAN : Multiple Time Series - Wiley 1970.

LES PROBABILITES CONDITIONNELLES D'AFFECTATION D'UNE
OBSERVATION A UNE CLASSE EN DISCRIMINATION PARAMETRIQUE

Auteur : VEYSSIERE Jean-Michel
Météorologie nationale
73-77 rue de Sèvres
92100 Boulogne

L'Analyse Discriminante Paramétrique permet, lorsque l'on dispose d'un prédictand ne pouvant prendre qu'un nombre fini de valeurs et de plusieurs prédicteurs continus, de donner une prévision de la classe à laquelle appartient le prédictand à partir des valeurs prises par les prédicteurs, en faisant certaines hypothèses sur la loi de probabilité suivie par ces prédicteurs. Dans le cas où le vecteur n-dimensionnel des prédicteurs est gaussien, il est possible de calculer en outre des probabilités d'apparition de telle ou telle classe du prédictand lorsque les valeurs prises par les prédicteurs sont connues. Le but de cette étude est d'explicitier la façon dont sont calculées ces probabilités d'une manière générale, avec des exemples d'application au cas fréquemment rencontré dans la pratique où le prédictand ne peut appartenir qu'à l'une de deux classes seulement : occurrence ou non-occurrence de précipitations, brouillard, avalanches, dépassement d'une valeur critique d'un paramètre, apparition d'un phénomène dangereux. La façon de calculer ces probabilités est étudiée pour l'Analyse Discriminante Paramétrique habituelle, avec et sans l'hypothèse d'égalité des matrices des variances-covariances, et pour l'analyse discriminante linéaire avec inégalité des matrices des variances-covariances (méthode Anderson-Bahadur). Enfin la signification à accorder à ces probabilités dans une utilisation opérationnelle est éclaircie, et leur caractère de probabilités conditionnelles mis en valeur, par opposition aux probabilités globales de réussite qui donnent quant à elles une idée de la qualité des résultats obtenus avec l'équation discriminante choisie.