

# STATISTIQUE ET ANALYSE DES DONNÉES

ASSOCIATION DES STATISTICIENS UNIVERSITAIRES

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

*Statistique et analyse des données*, tome 4, n° 3 (1979), p. 59-78

[http://www.numdam.org/item?id=SAD\\_1979\\_\\_4\\_3\\_59\\_0](http://www.numdam.org/item?id=SAD_1979__4_3_59_0)

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

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

NUMDAM

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

ASSOCIATION DES STATISTICIENS  
UNIVERSITAIRES

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



IDENTIFICATION D'UN PROCESSUS AUTO-REGRESSIF

MOYENNE MOBILE : LA METHODE DU COIN

Auteurs : MM. BEGUIN J.M., GOURIEROUX C.,  
MONFORT A.

E.N.S.A.E. : 3, avenue P. Larousse  
92240 MALAKOFF - FRANCE

Dans l'approche de Box et Jenkins des séries temporelles, un point important est la détermination des degrés  $p$  et  $q$  des polynômes auto-régressifs et moyenne mobile.

On propose ici un critère simple qui apparaît comme une généralisation des méthodes proposées par Box et Jenkins dans le cas de processus auto-régressifs purs ou moyennes mobiles pures. Ce critère ne nécessite pas d'avoir préalablement estimé les coefficients des polynômes définissant le processus puisqu'il ne repose que sur les valeurs des autocorrélations. Il permet également d'effectuer des tests statistiques, comme dans les cas "purs".

ANALYSE EN COMPOSANTES PRINCIPALES (A.C.P.)

D'UNE VARIABLE ALEATOIRE HILBERTIENNE,  
A.C.P. FILTRANTE ; DISCRETISATION

AUTEUR : BESSE Philippe

Laboratoire de Statistique et Probabilités  
UNIVERSITE PAUL SABATIER  
31077 TOULOUSE CEDEX.

On définit l'A.C.P. d'une v.a.  $X$ , centrée, du second ordre, à valeurs dans un espace hilbertien  $\phi$  (dans le cas où  $\phi$  est  $L^2_H(T, \xi, \mu)$ , c'est l'A.C.P. de la fonction aléatoire vectorielle  $X$ ).

Si  $F$  est une isométrie (un filtre) de  $\phi$  dans un Hilbert  $H$ , l'A.C.P. de  $FX$ , c'est à dire l'A.C.P. de la v.a. filtrée de  $X$  par  $F$ , est, à une isométrie près, l'A.C.P. de  $X$  dans  $\phi$ , que nous appelons A.C.P. filtrante de  $X$ . Ainsi, par exemple, pour rechercher l'A.C.P. des "dérivées" (resp. des "intégrales") des trajectoires d'une fonction aléatoire  $X$ , il suffit de calculer l'A.C.P. de  $X$  dans un espace de Sobolev (resp. dans le dual d'un espace de Sobolev) muni de la norme appropriée.

En pratique, lorsqu'on considère une f.a.  $(X_t)_{t \in T}$ ,  $T$  est discrétisé et se pose alors un problème de convergence d'A.C.P. On montre que l'A.C.P. sur la f.a. discrétisée converge uniformément à la condition que la suite emboîtée  $\{\phi_p\}_{p \in \mathbb{N}}$  ( $\dim \phi_p = p$ ) des sous-espaces d'approximation converge fortement vers  $\phi$ . Dans le cas particulier où  $\phi$  est un sous-espace de Sobolev (ou son dual), on trouve pour espace d'approximation "naturel", l'espace des fonctions "pline" d'interpolation aux instants  $t_1, \dots, t_p$  de discrétisation.

CONVERGENCE FAIBLE DES SUITES MELANGEANTES DE PROCESSUS EMPIRIQUES

NORMALISES TRONQUES MULTIDIMENSIONNELS

ANALYSE, A L'AIDE DE DISTANCES, DE DONNEES PREORDINALES

Hervé ABDI<sup>\*</sup>, Jean-Pierre BARTHELEMY<sup>\*\*</sup>, Xuan LUONG<sup>\*\*\*</sup>

Des résultats récents (sur la combinatoire des ensembles ordonnés) offrent la possibilité de calculer des distances entre des pré-ordres totaux, ces distances s'interprétant comme un cheminement minimal dans un graphe valué. Par ailleurs, la théorie -généralisée- de l'information permet de préciser la signification de certaines de ces métriques.

L'application -à des données de psychologie sociale : étude sur la motivation des injures- de ces derniers résultats, menée parallèlement à d'autres méthodes connues (i.e A.F.C.), aboutit à des résultats significatifs et originaux pour l'étude concernée.

En outre, la comparaison de ces différentes approches offre la possibilité de dégager la spécificité - mais aussi les limites- de chacune d'elles quant au problème étudié.

Auteurs : S. BALACHEFF  
G. DUPONT  
Université de ROUEN  
Faculté des Sciences et des Techniques  
76130 MONT SAINT AIGNAN

La convergence faible des suites mélangeantes de processus empiriques normalisés tronqués multidimensionnels (p.e.n.t.m.) peut être utilisée comme outil dans la démonstration de théorèmes centraux limites relatifs à des suites de statistiques de rang intervenant dans l'étude d'échantillons multidimensionnels (par exemple pour le test de l'indépendance des deux composantes dans un échantillon d'observations bi-dimensionnelles). Cette approche de la convergence des statistiques de rang a été proposée, en particulier, dans les travaux de L. Rüschendorf.

Il apparaît que les suites de p.e.n.t.m. ne satisfont pas les hypothèses des théorèmes classiques de convergence de processus à valeurs dans l'ensemble  $D_k$  des fonctions continues à droite et à limites à gauche de  $[0,1]^k$  dans  ${}^k R$  (ces théorèmes, qui font usage des modules  $w'$  et  $w''$  sur  $[0,1]^k$ , sont dus, en particulier, à Billingsley si  $k = 1$ , et à Bickel et Wichura dans le cas général). C'est pourquoi, l'approche proposée par Rüschendorf était restée jusqu'ici, à notre connaissance, incomplète.

Nous établissons une méthode de démonstration de ce théorème de convergence qui est fondée sur l'emploi du "module de continuité"  $w, (w(f, \delta) = \sup\{|f(t) - f(s)|; s \in [0,1]^k, t \in [0,1]^k \text{ et } |t - s| \leq \delta\} \text{ où } f \in D_k)$ ; cette méthode, qui opère grâce au fait que la probabilité limite est concentrée sur l'ensemble des trajectoires continues, nécessite en effet des hypothèses suffisamment peu restrictives pour être satisfaites par les p.e.n.t.m.

---

\* Centre de Recherche Socio-Economique, Faculté des Lettres,  
30, rue Mégevand, 25030 BESANCON CEDEX.

\*\* E.N.S.M.M., 25030 BESANCON CEDEX.

\*\*\* Laboratoire de Mathématiques et Statistiques, Faculté des Lettres,  
30 rue Mégevand, 25030 BESANCON CEDEX.

CALCUL FONCTIONNEL ET SERIES CHRONOLOGIQUES

Bernard CHARLES, Université de Montpellier  
et Roch ROY, Université de Montréal

Roch ROY, prof.  
Département d'informatique et de recherche opérationnelle  
Université de Montréal  
C.P. 6128, Succursale "A"  
Montréal, P.Q.  
H3C, Canada

L'objectif principal de cet article est de préciser le contexte mathématique le plus général possible dans lequel les manipulations d'opérateurs telles que faites dans le livre de Box et Jenkins (1976) demeurent valides. Ceci nous amène à considérer une famille d'opérateurs engendrée par l'opérateur décalage à gauche  $B$ , famille opérant sur l'espace des processus commençant à 0 ainsi que sur celui des processus commençant à  $-\infty$ . Pour chacun de ces deux espaces, en utilisant uniquement des notions générales d'algèbre linéaire ainsi que les principes de base du calcul fonctionnel, nous déduisons des conditions nécessaires et suffisantes pour l'inversibilité d'un opérateur polynomial en  $B$  ainsi que pour la simplification d'un facteur commun dans l'équation de définition d'un modèle ARMA. Finalement, nous discutons les implications de ces résultats sur différentes questions traitées dans Box et Jenkins.

Auteur : M. COLLOMBIER D.  
Laboratoire de Statistique et Probabilités  
UNIVERSITE Paul SABATIER  
118 Route de Narbonne  
31077 TOULOUSE Cédex

Nombreuses sont les tables de contingence où les caractères observés - s'ils sont considérés comme ensembles de leurs modalités - peuvent être naturellement dotés de structures d'ordre. Ces situations correspondent souvent au cas où les paramètres que l'on est alors amené à introduire dans les modèles d'analyse doivent respecter des contraintes (sous forme d'inéquations) pour être interprétables. A ce jour les spécialistes du modèle log-linéaire ne se sont guère intéressés à l'introduction de tels types de contraintes. Seules ont été envisagées indirectement, quoique dans un cadre plus général, les contraintes d'isotonie.

Cette communication a pour objet de faire le bilan de quelques travaux effectués sur les modèles log-linéaires à contraintes appliqués au traitement des tables de contingence dans deux domaines : la formulation des hypothèses et l'estimation par maximum de vraisemblance. A partir d'un exemple nous indiquons comment un modèle peut être paramétré pour tenir compte de structures d'ordre partiel ou total sur tout ou partie des caractères observés, et comment certaines hypothèses se traduisent alors en termes d'inéquations. Nous donnons ensuite une définition générale des modèles log-linéaires à contraintes et nous en indiquons quelques propriétés. Nous présentons enfin un algorithme d'estimation des paramètres sous de telles contraintes. Il s'agit d'un algorithme de programmation convexe à contraintes linéaires par sous-optimisations sans contrainte. Nous en détaillons le fonctionnement sur quelques exemples et nous indiquons les liens qui existent entre cet algorithme et ceux qui ont déjà été proposés dans le cas de contraintes d'isotonie.

ANALYSE EN COMPOSANTES PRINCIPALES SOUS CONTRAINTES LINEAIRES -  
INTRODUCTION DE LA METRIQUE RELATIONNELLE ET APPLICATIONS.

SUR UNE CLASSE DE TESTS QUI CONTIENT LE TEST DU  $\chi^2$ .

Auteur : M. BOSQ D.  
Université des Sciences et Techniques de Lille I  
U.E.R. de Mathématiques Pures et Appliquées  
59655 - VILLENEUVE D'ASCQ - CEDEX

Auteurs : CARLIER, A., CROQUETTE, A., GRAU, D., HAÏT, J.R., PASTOR, J.,  
SCHEKTMAN, Y.  
E.R.A. - C.N.R.S. n° 591  
Université Paul Sabatier  
118, route de Narbonne  
31077 TOULOUSE Cedex

On présente une classe de tests d'ajustement qui contient les tests du type  $\chi^2$  et notamment le "smooth test" de J. NEYMAN. Le résultat principal est une inégalité qui précise le comportement asymptotique des statistiques utilisées. On donne aussi une condition nécessaire et suffisante pour qu'un test de la classe soit convergent.

Enfin on montre comment le choix d'un test dans la classe considérée peut se faire en fonction de l'alternative envisagée.

Les auteurs définissent quelques types d'analyses en composantes principales sous contraintes linéaires et en donnent quelques propriétés.

Par l'introduction de la métrique "Relationnelle" les auteurs effectuent une synthèse de tous les modèles classiques d'analyse des données multidimensionnelles et formalisent en termes de projections, dans l'espace des unités statistiques, leurs propriétés conjointes. Ils proposent, en outre, des extensions et de nouvelles façons de définir ou d'aborder les problèmes suivants,

- analyse de plusieurs ensembles de variables,
- régression, régression protégée,
- mesure des effets dans les expériences planifiées,
- analyse des données évolutives,

ce qui leur permet d'apporter de nouveaux résultats.

ECHANTILLONNAGE PREFERENTIEL POUR LES SIMULATIONS DE  
CERTAINS PROCESSUS DE SAUTS.

Auteurs : Marie COTTRELL, Jean-Claude FORT, Gérard MALGOUYRES

Université Paris-Sud  
Département Mathématique  
E.R.A. 532 "Statistique Appliquée"  
91405 ORSAY

Auteur: M. DEHEUVELS P.  
ISUP, t 45-55, E3,  
Université Paris VI  
4 Place Jussieu  
75230 PARIS CEDEX 05

La méthode de simulation avec échantillonnage préférentiel ("Importance-Sampling") consiste à effectuer un changement de probabilité recentrant autour des événements entrant pour une grande part dans la valeur à estimer, la difficulté étant que la détermination du changement optimal de probabilité est aussi difficile que le problème initial.

Pour les simulations de certains processus de sauts de taille modérée, les techniques et résultats des grandes déviations suggèrent un changement de probabilité permettant de réduire considérablement la variance des estimateurs. Si les sauts sont petits, elles suggèrent une voie d'approche analytique.

Cette méthode permet d'obtenir des résultats fiables dans le cas où des approximations par des diffusions donnent des résultats erronés, et où les calculs analytiques ne sont pas possibles.

Elle sera développée sur quelques exemples de modèles écologiques, de dynamique des populations et de télécommunications.

Etant donné une suite  $X_n = (X_n(1), \dots, X_n(p)) \in \mathbb{R}^p$ ,  $n=1,2,\dots$  de vecteurs aléatoires indépendantes de même loi, si  $Y_n = (\sup_{1 \leq k \leq n} X_k(1), \dots, \sup_{1 \leq k \leq n} X_k(p))$

désigne son extrême supérieur d'ordre  $n$ , en utilisant des résultats nouveaux concernant la caractérisation des lois limites de  $Y_n$ , en faisant l'hypothèse que la loi de  $Y_n$  appartient au domaine d'attraction d'une loi limite extrême multivariée, on donne des méthodes d'estimation non paramétrique efficaces de cette loi.



## TESTS DE RUPTURE DANS UN MODELE

Auteurs : M. DESHAYES J., Melle PICARD D.  
Université Paris-Sud  
Département Mathématique  
E.R.A. 532 "Statistique Appliquée"  
91405 ORSAY

Nous étudions le comportement asymptotique de différents tests utilisés pour déceler une rupture dans un modèle de régression. On suppose observer les variables :

$$y_k = x'_k \beta_k + u_k, \quad k = 1, 2, \dots, n.$$

- où
- $x_k$  désigne le vecteur sur lequel on régresse :  $x_k \in \mathbb{R}^d$
  - $\beta_k$  est le vecteur des paramètres de la régression au temps  $k$
  - $u_k$  est l'erreur : les  $u_k$  sont indépendantes et  $N(0, \sigma_k^2)$ .

On teste l'hypothèse  $H_0 : \{\beta_1 = \beta_2 = \dots = \beta_n \text{ et } \sigma_1 = \sigma_2 = \dots = \sigma_n\}$  contre l'hypothèse qu'il y a rupture à un instant  $k$  inconnu.

Pour le test du rapport de vraisemblance et pour les tests classiques basés sur les sommes de résidus récursifs ou leurs carrés, nous calculons la vitesse exponentielle (le seuil étant fixé) de l'erreur de première espèce. Nous montrons l'optimalité, de ce point de vue, du test du rapport de vraisemblance (en comparant les vitesses des erreurs de seconde espèce). Nous introduisons un nouveau test basé sur une randomisation sur les instants d'observation qui, du point de vue asymptotique donne des résultats intéressants.

## QUELQUES REMARQUES LIEES À LA DEFINITION D'INDICES DE LIAISON

### PARTIELLE ENTRE VARIABLES VECTORIELLES

Auteurs : Mr. DROUET d'AULIGNY & Mlle. EXCOFFIER C.  
I.M.S.S. B<sup>L</sup> Sciences Humaines & Mathématiques  
Université des Sciences Sociales de GRENoble  
38040 GRENoble cedex

La définition & l'utilisation d'indices de liaison entre variables vectorielles est nécessaire à l'évaluation de la qualité des résumés construits en analyse des données, ainsi qu'à l'interprétation des résultats.

On présente un rapport de synthèse étudiant les différentes méthodes de construction d'indices du point de vue de leur utilisation possible en analyse des données. Le recours à des codages à priori ou à postériori liés à l'échelle de mesure des variables étudiées est analysé.

La distinction entre une évaluation de l'écart à la situation d'indépendance & celle de la force d'une relation, associée à l'opposition entre les caractères symétrique ou non d'une liaison, permet une classification simple des divers indices présentés.

L'extension des résultats au cas vectoriel permet une définition commune d'indices d'association partielle & d'indices d'écart à la situation d'indépendance conditionnelle.

Le rapport est rédigé dans une optique d'analyse des données & l'application des résultats présentés est discutée dans cette direction.

Le rapport sera disponible sur simple demande aux auteurs au début de l'année 1980.

ANALYSE FACTORIELLE SPHERIQUE

Auteurs : Melle Dominique DOMENGES  
Allocataire DG RST  
ISUP, Faculté des Sciences, 4 Place Jussieu 75230 PARIS CEDEX 05

Michel VOLLE  
INSEE  
Division des Comptes Trimestriels  
18, Boulevard A. Pinard  
75675 PARIS CEDEX 14  
tél : 540-18-08

On part de la remarque suivante, due à J.-P. BENZECRI : les distributions sur un ensemble fini  $I$  peuvent être représentées aussi bien sur un orthant de sphère que sur un simplexe. Il est en outre possible de définir sur la sphère une distance entre distributions qui corresponde à la métrique du  $\chi^2$  utilisée sur le simplexe : si l'on considère la métrique  $\chi^2$  centrée sur une distribution  $r$  comme une métrique locale dans le simplexe, on dote celui-ci d'une structure d'espace de Riemann ; et la distance géodésique sur le simplexe se relie très simplement à la distance usuelle sur la sphère.

La représentation des distributions sur la sphère, associée à la métrique de Hellinger :

$$d^2(p, q) = \sum_i (\sqrt{p_i} - \sqrt{q_i})^2,$$

permet de construire une nouvelle méthode d'analyse factorielle. Les applications possibles sont nombreuses : tableaux de fréquences comportant des valeurs négatives ; comparaison de deux tableaux de fréquences ; analyse d'un nuage "vu d'un point" ; traitement de tableaux de transition ou d'échanges.

MONTAGE AUDIO-VISUEL "STATISTIQUE MULTIDIMENSIONNEL"

J.P. FENELON  
CNRS - CREDOC  
140, rue du chevaleret  
75013 PARIS

Après une vingtaine de programmes audio-visuels consacrés aux statistiques classiques (histogrammes, tests, prévision, contrôle de la qualité, etc...) les éditions PRISMATHEQUE publient trois nouveaux tomes, de chacun 80 diapositives.

Le premier est une introduction intuitive à l'Analyse des Données. Le second présente l'Analyse factorielle et sa pratique (choix des métriques, choix des codages). Le troisième porte sur la classification Automatique, la régression, l'analyse de la variance et la segmentation.

Le produit a été testé sur un public de non-mathématiciens (essentiellement des juristes, psychologues, sociologues, archéologues, médecins), désireux de ne pas passer plus de trois journées d'introduction à des techniques qu'ils veulent mettre en oeuvre sur leurs données.

On consacrera la séance à présenter le produit à la fois à ceux d'entre nous qui ont des problèmes de formateurs, et à ceux de nos collègues qui voudraient bien nous donner leur avis sur ce travail de pédagogie pour adultes.

ETUDE DE LA STABILITE PAR ECHANTILLONNAGE  
DANS LA SEGMENTATION

AUTEUR : J. FINE-FONTAN

Laboratoire de Statistique et Probabilités  
Université Paul Sabatier  
118, route de Narbonne  
31077 TOULOUSE CEDEX

Dans un premier temps, on propose un modèle général permettant de faire le lien entre l'échantillonnage en Statistique Inférentielle et le point de vue descriptif de l'Analyse des Données.

On utilise ensuite ce modèle pour étudier la stabilité de la méthode de segmentation : on définit de façon naturelle ce que l'on peut appeler la "segmentation exacte" et on démontre, à chaque étape, la convergence de la méthode lorsque la taille de l'échantillon augmente indéfiniment. Les différents cas suivant la nature du critère sont examinés. On peut ajouter que, pour chaque cas, on obtient, presque sûrement, à partir d'un certain rang, non une approximation, mais précisément les résultats de la segmentation exacte.

UNE AUTRE PRESENTATION DE LA COMPARAISON DES ANALYSES  
EN ANALYSE DES CORRESPONDANCES

B.FICHET

Laboratoire de Physique. Faculté de Médecine.  
Université d'Aix-Marseille II  
27, Bd Jean-Moulin  
13385 MARSEILLE CEDEX 4

Nous proposons une autre méthode pour la comparaison des analyses, dans une analyse des correspondances entre deux ensembles A et B. La démonstration est très courte ; au prix d'une transformation analytique simple de la métrique sur B, liée aux résultats de l'analyse sur A, nous plaçons directement B dans l'espace de la solution, c'est-à-dire dans un espace euclidien ayant pour origine le centre de gravité et pour axes, les axes factoriels du nuage associé à B. Les résultats classiques, comme la propriété barycentrique..., en découlent très simplement.

AUTEUR : J.P. FLORENS  
CERAMMISE  
FACULTE DES SCIENCES ECONOMIQUES  
UNIVERSITE D'AIX-MARSEILLE II  
38, cours Mirabeau  
  
13100 AIX-EN-PROVENCE

Une expérience bayésienne étant définie par une probabilité sur le produit de l'espace des échantillons et de l'espace paramétrique, une sous tribu de l'espace des échantillons est exogène si l'expérience conditionnelle à cette tribu contient toute l'information sur la sous tribu des paramètres d'intérêt.

Dans un modèle séquentiel une suite  $(\mathcal{C}_n)_n$  de sous tribus de l'espace des échantillons est transitive (en échantillonnage ou en prédictive) si la probabilité sur  $\mathcal{C}_{n+1}$  conditionnelle à  $\mathcal{C}_n$  est identique à la probabilité conditionnelle à tout l'échantillon jusqu'à l'instant  $n$ . Cette définition généralise les définitions de non causalité utilisées en économétrie.

Les propriétés de ces deux concepts et leurs relations sont étudiées ainsi que leur application à deux exemples.

Auteurs : Dominique FOURDRINIER  
Gérard GRANCHER  
Université de Rouen  
Faculté des Sciences et des Techniques  
76130 MONT SAINT AIGNAN

On s'intéresse à la comparaison de  $T$  échantillons (à valeurs dans un ensemble totalement ordonné) dont on sait qu'ils sont régis par des probabilités images, par des applications strictement croissantes inconnues  $(h_1, \dots, h_T)$ , d'une même probabilité discrète  $P_0$ . Il s'agit de tester l'identité des probabilités régissant ces échantillons contre une "tendance" se traduisant par un "léger décalage" des applications  $h_t$  (où  $1 \leq t \leq T$ ) entre elles. Cette notion de léger décalage est précisée à la fois à l'aide d'une relation de préordre sur l'ensemble des familles  $(h_t, 1 \leq t \leq T)$  et d'une topologie (relativement à laquelle les familles "légèrement décalées" constituent un voisinage de l'hypothèse à tester).

L'étude est effectuée à l'aide de tests de rang ; on donne une formule générale pour le test de rang de plus forte puissance, et pour le test de plus forte puissance parmi les tests uniformément randomisés (c'est-à-dire tels que, en cas d'ex-aequo, on tire au hasard un rangement avec la probabilité uniforme sur l'ensemble de tous les rangements possibles).

Des calculs de mise en oeuvre des tests, et de puissance dans le cas où la probabilité  $P_0$  est uniforme et où on compare 2 échantillons, ont fait apparaître :

- 1°) l'absence de l'influence de la taille du support de  $P_0$  sur la mise en oeuvre du test uniformément randomisé ;
- 2°) la puissance particulièrement faible (et ce d'autant que la taille du support est élevée) du test uniformément randomisé, alors que la puissance du test de plus forte puissance parmi les tests de rang, localement meilleure, se détériore très vite quand on sort de l'hypothèse locale ;
- 3°) le rôle prépondérant joué par la considération des seules provenances (du premier ou du second échantillon) des valeurs extrêmes observées (la plus faible et la plus forte).

Cette dernière remarque a conduit à proposer également des tests fondés sur la provenance de ces valeurs extrêmes (où aussi des 2 plus faible et des deux plus fortes valeurs), et à en calculer la puissance.

TESTS D'HYPOTHESES SEPARÉES POUR DES

PROCESSUS STATIONNAIRES DU SECOND ORDRE

STATISTIQUE DES PROCESSUS A INDICES DANS  $\mathbb{R}^d$

Auteur : D. GUEGAN

UNIVERSITE PARIS XIII

E.R.A. 532 STATISTIQUES APPLIQUEES

Bât. 425

91405 ORSAY

Xavier GUYON

Université Paris-Sud

Département Mathématique

E.R.A. 532 "Statistique Appliquée"

91405 ORSAY

Nous étudions la statistique limite du test du maximum de vraisemblance pour des processus gaussiens du second ordre. Etant donné  $(X_t)_{t \in \mathbb{R}}$  un tel processus, soit à tester  $\theta_0$  contre  $\theta_1$  hypothèses composées. En un point  $\alpha \in \theta_0$ , la densité spectrale du processus est notée  $f(\lambda, \alpha)$ , et  $g(\lambda, \beta)$  pour  $\beta \in \theta_1$ . On considère des tests basés sur la statistique

$$Z_N = \frac{\sup_{\alpha} L_N(X, \alpha)}{\sup_{\beta} L_N^*(X, \beta)} = \frac{L_N(X, \hat{\alpha}_N)}{L_N^*(X, \hat{\beta}_N)}$$

où  $L_N(X, \cdot)$  représente la log vraisemblance du processus sous l'hypothèse  $H_0$ , et  $\hat{\alpha}_N$  et  $\hat{\beta}_N$  les estimateurs du maximum de vraisemblance sous  $H_0$  et  $H_1$ . Pour cela nous considérons l'approximation de Whittle de la log vraisemblance et nous nous plaçons dans le cas où les hypothèses sont asymptotiquement séparées, au sens de Kullback, i.e.

$$J(\theta_0, \theta_1) = \inf_{\alpha \in \theta_0, \beta \in \theta_1} K(\alpha, \beta) > 0 \quad \text{avec}$$

$$K(\alpha, \beta) = (2\pi)^{-1} \left| \int \log \frac{f(\alpha, \beta)}{g(\alpha, \beta)} d\lambda + 1 - \int \frac{f(\lambda, \alpha)}{g(\lambda, \beta)} d\lambda \right|$$

Nous montrons en fait que la statistique  $T_N = \frac{\hat{L} - E_{\hat{\alpha}} \hat{L}}{(\text{Var}_{\hat{\alpha}} \hat{L})^{1/2}}$  (où

$\hat{L} = L_N(X, \hat{\alpha}_N) - L_N^*(X, \hat{\beta}_N)$ ) convenablement renormée est libre en loi. Nous considérons alors un test basé sur cette statistique, nous montrons qu'il est consistant et que l'on peut construire une région de rejet indépendante des paramètres.

Soit  $X_t, t \in \mathbb{R}^d$ , un processus stationnaire du second ordre de densité spectrale  $f(\lambda, \theta)$ . Les approximations de la vraisemblance de Whittle, Box-Jenkins sont étudiées lorsque les observations sont faites sur un "rectangle"  $N = n_1 \times n_2 \times \dots \times n_d$  : ces approximations sont en  $O_p(N^{1-\frac{1}{d}})$ , du fait d'effet de bord du même ordre. Les estimateurs  $\hat{\theta}_N$  associés sont étudiés : dans le cas gaussien, ils seront efficaces une fois débiaisés d'une quantité  $b_N(\hat{\theta}_N)$ , nulle si  $d = 1$ , en  $N^{-1/2}$  si  $d = 2 \dots$ , ceci d'une part du fait des effets de bord, d'autre part de la plus grande dépendance probabiliste en moyenne des observations lorsque  $d$  augmente.

INVERSIBILITE ET INVERSIBILITE GENERALISEE DES MODELES DE SERIES CHRONOLOGIQUES

M. HALLIN  
Université Libre de Bruxelles CP210  
B1050 Bruxelles Belgique

Un nouveau concept d'inversibilité a été introduit récemment par Granger et Andersen dans le but de généraliser le concept classique. Nous montrons comment leur définition, satisfaisante dans le cas des modèles à coefficients indépendants du temps, ne constitue plus une généralisation correcte lorsque les coefficients du modèle varient dans le temps. Nous donnons également une définition modifiée, qui fournit dans les deux cas une généralisation correcte, tout en jouissant des mêmes propriétés que celle de Granger et Andersen.

SYSTEME INTERACTIF D'ACQUISITION,  
DE REDUCTION ET DE VISUALISATION DE DONNEES VOCALES.

Auteur : Mme HATON M.C.  
Centre de Recherche en Informatique de Nancy  
C.O. 140  
54037 NANCY-CEDEX

Dans le cadre du système SIRENE concernant l'analyse de la parole et son application à la réalisation d'aides visuelles pour la rééducation vocale des enfants non-entendants, nous présentons un ensemble interactif de programmes fonctionnant en trois phases :

- acquisition en temps réel de données vocales prononcées par un ou plusieurs locuteurs et appartenant à des classes bien déterminées (classes de phonèmes du français) pour la phase d'apprentissage ; les données sont constituées des valeurs de l'intensité du signal dans les quinze bandes de fréquence d'un analyseur spectral,
- calcul en ligne des facteurs des deux premiers axes obtenus par analyse factorielle discriminante sur ces données pour la phase de traitement,
- visualisation en temps réel sur écran cathodique des nuages de points correspondant à l'émission de sons prononcés par les enfants à rééduquer.

On montre sur un exemple de plan optimal au sens de la meilleure discrimination des sons l'importance de l'information apportée à l'enfant sur la qualité de son élocution.

CARACTERISATION DES DIRECTIONS ET DES COMPOSANTES PRINCIPALES EN  
ANALYSE EN COMPOSANTES PRINCIPALES - ANALYSE EN COMPOSANTES  
SEMBLABLES.

Auteurs : M. HAÏT J.R. et SCHEKTMAN Y.  
E.R.A. - C.N.R.S. n° 591  
Université Paul Sabatier  
118, route de Narbonne  
31077 TOULOUSE Cedex

Les auteurs présentent plusieurs familles de fonctions numériques, définies sur  $\mathbb{R}^D$ , dont les optima sont, sous certaines conditions, les  $p$  vecteurs axiaux principaux (resp. les  $p$  composantes principales).

Ces propriétés extrémales permettent de définir l'Analyse en Composantes Principales sous des contraintes très générales.

En inversant les extrema des fonctions précédentes, les auteurs proposent le modèle d'Analyse en composantes semblables et en étudient les propriétés. Ils montrent, en particulier, que ce modèle possède d'excellentes caractéristiques (stockage de l'information, reconstitution d'informations perdues) pour la télétransmission d'images multispectrales comprimées.

ADAPTABILITE ET STABILITE DES MODELES ARIMA

Auteur : M. INDJEHAGOPIAN J.P.  
Professeur de Statistique  
Ecole Supérieure des Sciences  
Economiques et Commerciales  
BP 105  
95021 CERGY PONTOISE CEDEX

La modélisation d'une série temporelle à partir des données du passé, permet d'effectuer des prédictions si l'on suppose que la structure du modèle et/ou ses paramètres sont invariants au cours du temps. Si tel n'est pas le cas, les prédictions seront sérieusement altérées.

Dans cette communication, on s'intéresse à la modélisation des séries temporelles à partir des modèles ARIMA et on étudie le changement des paramètres dans de tels modèles. En particulier, on regarde comment les effets d'une intervention déterminée (taxe, promotion, crise, etc ...) sur une série économique, peuvent être modélisés en présence d'un bruit. Ces modèles se présentent sous la forme :

$$z_t = f(\delta, \omega, x, t) + n_t$$

où  $n_t$  suit un ARIMA  $\{p, d, q ; P, D, Q, s\}$  et  $f$  représente l'effet dynamique de la variable exogène  $x_t$  déterministe avec  $\delta, \omega$  une suite de paramètres.

On montre à partir d'un théorème que le biais de prédiction, sur l'horizon  $l$  à partir de l'époque  $n$ , lorsque l'on néglige les effets d'une intervention, se calcule par

$$B_n(l) = h(B) [x_{n+l} - \sum_{j \geq 0} \pi_{j+1}^{(l)} x_{n-j}]$$

où  $h(B)$  est la fonction d'intervention et où les coefficients  $\pi_k^{(l)}$  se calculent par récurrence à l'aide des coefficients du modèle sur le bruit  $n_t$ .

Le calcul de ce biais permet de discuter de l'adaptabilité des prédictions sur les séries modélisées à partir de processus ARIMA lorsque ces séries sont soumises à des interventions déterminées (telles que des impulsions ou des sauts à des périodes données).

Enfin on utilise le critère de Box et Tiao pour tester la stabilité des paramètres de processus ARIMA qui modélisent des séries macroéconomiques françaises soumises à des interventions (récession 1974, etc ...). Dans le cas de rejet de la stabilité, on propose une modélisation avec fonctions d'intervention.

NORMALITE ASYMPTOTIQUE DE CERTAINES STATISTIQUES DE RANGS SERIELLES  
DANS LES PROCESSUS  $\Psi$ -MELANGES

J.-F. INGENBLEEK  
Université Libre de Bruxelles CP 210  
B1050 Bruxelles Belgique

Une classe assez étendue de statistiques utilisées dans les tests d'indépendance peuvent s'écrire sous une forme générale commune, appelée ici statistique de rangs sérielle. Nous étudions la normalité asymptotique de telles statistiques sous l'hypothèse que le processus sous-jacent est  $\Psi$ -mêlé.

OPTIMALITE DES DECOMPOSITIONS DE FORMES QUADRATIQUES  
EN ANALYSE DES DONNEES

Auteur : M. KOBILINSKY A.  
INRA, Laboratoire de BIOMETRIE du CNRA  
Route de Saint-Cyr  
78000 VERSAILLES  
Tél. : 950 75 22 poste 570

Les sous-espaces définis en Analyse en Composantes Principales (ACP) sont optimaux pour un critère beaucoup plus fort que celui de l'inertie. On étudie ce type d'optimalité, en montrant les relations d'ordre qui existent entre critères sur les individus d'une part, sur les variables de l'autre.

A cette occasion, on rappelle certaines propriétés des décompositions de formes quadratiques en régression multivariée, ACP, Analyse Factorielle classique.

On montre enfin sur un exemple que l'analyse en composantes explicatives n'a pas la propriété d'optimalité mentionnée ci-dessus pour l'ACP classique.



ETUDE STATISTIQUE DES PHENOMENES DE  
PERCEPTION EXTRA-SENSORIELLE

Auteur : M. LIGNON Yves  
Laboratoire de Parapsychologie  
UER Mathématiques, Informatique,  
Statistiques et Sciences Expérimentales  
Université Toulouse-Le Mirail  
31061 Toulouse-Cedex

Après un rappel du rôle joué par la statistique dans les expériences classiques de perception extra-sensorielle (méthode de Rhine) on montre que la solution des problèmes posés par l'existence du risque d'erreur de 1ère espèce nécessite la connaissance de la distribution de la v. a.  $X$  qui rend compte de la performance du sujet.

L'auteur expose comment il a étudié ces questions sur 1992 cas grâce à une procédure expérimentale inédite. Il conclut que la distribution observée de  $X$  diffère significativement de celle du hasard. ( $P < .001$ ). De plus on présente les "napiers" peu ou pas connus que Fisher a consacré à ce sujet. Ces napiers présentent d'autre part un intérêt certain en ce qui concerne l'histoire de la théorie des tests statistiques.

REPRESENTATION FACTORIELLE ADAPTEE AU CRITERE QUADRATIQUE  
EN ANALYSE DISCRIMINANTE

Auteur : D.E. LAFAYE de MICHEAUX  
LASSY (Laboratoire de Signaux et Systèmes)  
Equipe de Recherche Associée au CNRS R35  
Université de Nice  
41, boulevard Napoléon III  
06041 NICE CEDEX

L'Analyse Factorielle Discriminante Linéaire (A.F.D.L.) propose un ensemble réduit de variables explicatives (variables factorielles) qui contient toute l'information utile à l'Analyse Discriminante Linéaire.

On étudie, dans cet exposé, le choix de variables factorielles qu'il faut ajouter à celles issues de l'A.F.D.L. afin de résumer complètement l'information utile à une discrimination quadratique : on traite l'aspect géométrique de ce problème ainsi que le cas où les données suivent un modèle probabiliste gaussien.

On illustre la méthode proposée par un exemple basé sur des données médicales concrètes.

PROPRIETE DE MARKOV DE CERTAINS PROCESSUS A INDICE DANS LE PLAN

Auteur : M. PRUM B.  
Université Paris-Sud  
Mathématiques, Bât. 425, ERA 532 "Statistique Appliquée"  
91405 ORSAY

Nous donnons un ensemble ordonné filtrant  $E$  (en pratique  $\mathbb{R}$  ou  $\mathbb{R}^2$ ) et une famille croissante de tribus  $\mathcal{F}_t$ ,  $t \in E$ , nous étudions les processus centrés de carré intégrable  $Y_t$  tels que pour tout couple  $(s, s')$  d'éléments de  $E$  vérifiant  $s \leq s'$ , l'espace vectoriel fermé engendré par  $\{E(Y_t | \mathcal{F}_s)\}$  pour  $t \geq s'$  est indépendant de  $s'$  et de dimension  $d_s$  bornée par un certain  $N$  : c'est la propriété de Markov au sens de Hida : les espérances de  $Y$  futur sont résumées en la connaissance de  $N$  valeurs  $E(Y_{t_i} | \mathcal{F}_s)$ .

Nous liions cette propriété au fait pour  $Y_t$  de s'écrire  $Y_t = \sum f_{\alpha}(t) M_{\alpha}(t)$  où les  $M_{\alpha}(t)$  sont  $N$  martingales linéairement indépendantes.

Nous généralisons cette étude à la prédiction de  $Y_t$ ,  $t \in \mathbb{R}^2$ , connaissant le processus sur un quart de plan, trois quarts de plan, l'extérieur d'un rectangle, etc..., fournissant une classe de modèles de processus markoviens dans le plan.

THEORIE ASYMPTOTIQUE DE L'ANALYSE EN  
COMPOSANTES PRINCIPALES LINEAIRE (A.C.P.L.) D'UNE  
FONCTION ALEATOIRE (F.A.) VECTORIELLE.

AUTEUR : ROMAIN Yves  
Laboratoire de Statistique et Probabilités  
UNIVERSITE PAUL SABATIER  
31077 TOULOUSE CEDEX.

On établit, à partir de propriétés antérieures sur l'échantillonnage statistique, la distribution asymptotique conjointe des valeurs principales et la normalité asymptotique des vecteurs principaux de l'A.C.P.L. d'une f.a. vectorielle.

Ces résultats généralisent donc en dimension infinie et pour une f.a. vectorielle la théorie asymptotique donnée par T.W. ANDERSON (dans le cas gaussien) et par A.DAVIS (pour le cas non-gaussien) pour des v.a. réelles et en dimension finie.

Une attention particulière est portée au cas d'une f.a. gaussienne où les résultats sont alors complètement explicités.

UNE METHODE DE SELECTION D'UN ALGORITHME DE PREVISION A COURT TERME

Auteurs : Guy MELARD et Olivier ROULAND

Institut de Statistique C.P. 210

Université Libre de Bruxelles - Campus Plaine

Boulevard du Triomphe

B-1050 Bruxelles, BELGIQUE

On envisage l'estimation des paramètres de plusieurs algorithmes de prévision à court terme ainsi que la spécification des performances de ceux-ci sur les données du passé. Les algorithmes considérés sont ceux des procédures de lissage exponentiel (de Brown, de Holt et de Winters), de modèles à deux aléas et de modèles de croissance polynomiale de Harrison. La méthode proposée exploite les liens qui existent entre ces algorithmes et les modèles ARIMA. Certaines généralisations récentes des modèles ARIMA sont également appliquées, notamment l'inclusion des transformations de Box et Cox, la possibilité de faire dépendre la variance du temps et le traitement déterministe du mouvement saisonnier. La phase de validation des algorithmes utilise le test de Ljung et Box et plusieurs tests d'homogénéité. Les algorithmes retenus sont comparés au moyen du critère d'Akaike.

Les justifications théoriques de la méthode sont exposées et l'intérêt pratique de celle-ci est discutée.

APPROXIMATION STOCHASTIQUE : RESULTATS THEORIQUES

ET APPLICATIONS

Auteur : J.M. MONNEZ

I.U.T. A, Département Informatique

2 bis, boulevard Charlemagne

54000 NANCY

On donne une généralisation d'un théorème de Blum sur la convergence presque sûre du processus de Robbins-Monro multidimensionnel, applicable à une généralisation de ce processus introduite par Burkholder. Un résultat de rapidité de convergence est également donné.

On applique le résultat de convergence presque sûre à un processus de Robbins-Monro tel que, à chaque étape, on fasse plusieurs observations, en nombre non nécessairement constant, dont on prend la moyenne. En outre, un résultat de normalité asymptotique est donné pour un tel processus, par application d'un théorème de Fabian.

Ces deux derniers résultats sont appliqués dans l'estimation récursive d'une espérance conditionnelle et dans l'estimation récursive d'un paramètre d'une loi de probabilité en utilisant une fonction d'estimation.

UNE DISTANCE ENTRE VARIABLES  
VECTORIELLES DE MEME DIMENSION.

Auteurs : S. SELLAM, A. FORCIOLI, J. AZENCOT  
Laboratoire informatique  
Université Claude Bernard Lyon 1  
43, boulevard du 11 Novembre 1918  
69621 VILLEURBANNE

Un certain nombre de techniques du type analyse de données (en particulier l'analyse des correspondances) ou du type classification automatique (en particulier l'analyse hiérarchique) permettent de comparer entre elles des variables aléatoires unidimensionnelles. Dans cet article, nous définissons une distance entre variations vectorielles qui peut permettre une généralisation de ces méthodes.

Nous verrons dans une première partie, la définition d'un "indice d'écart" entre sous-espaces vectoriels d'un espace vectoriel normé, ainsi que quelques propriétés de cet "indice d'écart". Dans une seconde partie, nous verrons que cet "indice d'écart" est une distance si on l'applique à des sous-espaces vectoriels de même dimension et nous verrons son application à des vecteurs aléatoires (ou variables vectorielles).

MODELES D'ANALYSE DES  
TABLEAUX DE TRANSITION CARRÉS

Auteur : M. THELOT Cl.  
I.N.S.E.E.  
Direction Régionale de NANTES  
76, rue Desaix  
44035 - NANTES CEDEX

Pour analyser des tableaux de transition carrés  $(p_{ij})$ , on propose trois modèles définis de la façon suivante : on se donne a priori une fonction  $f$  de deux variables  $(x,y) \in \mathbb{R}^+ \times \mathbb{R}^+$ , symétrique :  $f(x,y) = f(y,x)$ . Et l'on définit :

$$(P) \quad \forall i, \forall j, \quad f(p_{ij}, p_{ji}) = 0$$

$$(PF) \quad \forall i, \exists p_i > 0, \forall j, \exists r_{ij} > 0, ; p_{ij} = p_i r_{ij} \text{ avec } f(r_{ij}, r_{ji}) = 0$$

$$(PQF) \quad \forall i, \exists p_i > 0, q_i > 0, \forall j, \exists r_{ij} > 0 : p_{ij} = p_i q_j r_{ij} \text{ avec } f(r_{ij}, r_{ji}) = 0$$

Il est facile de voir qu'un certain nombre de modèles classiques (symétrie, quasisymétrie ...) sont des cas particuliers de cette formulation générale.

On étudie quelques propriétés de ces trois modèles (liens avec le modèle loglinéaire, relations avec des modèles marginaux). On examine ensuite quelle signification revêt la forme de l'estimateur du maximum de vraisemblance.

Quelques modèles particuliers (notamment celui qui dérive de la fonction  $f(x,y) = xy - k$ ) donnent lieu à un examen approfondi. Enfin différentes applications (mobilité sociale, migrations géographiques) sont proposées.

DECOMPOSITION HIERARCHIQUE DU KHI-DEUX D'INDEPENDANCE

ASSOCIE A UN TABLEAU DE CONTINGENCE

Auteur : Mme WEISS M.C.

U.E.R. de Mathématiques, Logique formelle et Informatique  
Université René Descartes, 12 rue Cujas, 75005 PARIS

La décomposition algébrique d'un espace de mesures de fréquences associées à des tableaux est utilisée par H. ROUANET dans le domaine des plans d'expérience. L'application qui est faite dans mon article "Décomposition hiérarchique du khi-deux associé à un tableau de contingence à plusieurs entrées", (Revue de Stat. Appl., 1978, XXVI n°1, p.23-33) n'est qu'un exemple de son utilisation.

S'il est connu depuis longtemps que le "khi-deux d'indépendance" peut se décomposer en somme de "khi-deux d'interaction", la méthode et les formules mises au point par CAPERAA permettent de décomposer chacun de ceux-ci selon un produit de partitions sur chacune des variables. Les partitions forment une décomposition hiérarchique. Cette méthode est présentée en utilisant la métrique du khi-deux par rapport à la mesure-produit des mesures marginales simples, et des vecteurs associés aux dichotomies employées. Dans les domaines où ces statistiques sont utilisées, on peut étudier avec précision l'effet du regroupement des modalités de certaines variables, ce qui permet de réduire avec discernement de grands tableaux à multiples entrées.