Statistique et analyse des données

Association des statisticiens universitaires

Résumés - Journées de Statistique, Nice 22-26 mai 1978

Statistique et analyse des données, tome 3, nº 2 (1978), p. 15-29

<http://www.numdam.org/item?id=SAD_1978__3_2_15_0>

ASSOCIATION DES STATISTICIENS
UNIVERSITAIRES

(Résumés - Journées de Statistique, Nice 22-26 MAI 1978)

ANALYSE DES CORRESPONDANCES SUR LES GRAPHES DE MOBILITE

A. BELLACICCO

UNIVERSITE DE ROME
Institut de Calcul de Probabilité
ITALIE

L'objet d'étude de ce papier est d'analyser la mobilité spatiale en ITALIE avec la méthode de l'analyse factorielle des correspondances. On considère le graphe de mobilité spatiale en ITALIE, c'est à dire les fréquences de changement d'adresse d'une région à l'autre pendant les dernières années. On obtient une matrice de mobilité qui est appelée matrice de mobilité effective, M et une matrice de mobilité maximum qui peut être obtenue avec une méthode particulière appelée de "graduation controvariante". De la même manière, on peut obtenir une matrice de mobilité nulle M°.

La comparaison entre la matrice de mobilité effective et la matrice de mobilité nulle est amenée par l'analyse factorielle des correspondances sur les deux matrices, soit en analysant les statistiques qu'on peut obtenir, c'est à dire les valeurs propres, les valeurs factorielles, etc..., soit par les graphiques des premiers axes factoriels.

Dans ce papier, est analysé soit l'algorithme de construction de la matrice de mobilité nulle, soit les résultats de l'analyse des données.

AUTOREGRESSION : UNE APPLICATION

D. BOMASSI*
A. de FALGUEROLLES**

* Ingénieur au S.I.M.A., Gaz de France, 93210 LA PLAINE-ST.DENIS
** assistant à l'I.U.T.(B), 31081 TOULOUSE Cédex

Soit $C = \{x_o, x_1, \ldots, x_n\}$ une série chronologique scalaire, réelle, de moyenne empirique $\bar{x}$, de fonction de covariance empirique $\{s_o, s_1, \ldots, s_n\}$ et de corrélogramme empirique $\{r_1, r_2, \ldots, r_n\}$:

$$\bar{x} = \frac{1}{n+1} \sum_{t=0}^{n} x_t, \quad s_j = \frac{1}{n+1} \sum_{t=0}^{n-j} (x_t - \bar{x})(x_{t+j} - \bar{x}) \text{ et } r_j = \frac{s_j}{s_o}.$$

Considérons la série réduite $C^* = \{y_o, y_1, \ldots, y_n\}$, $y_t = (x_t - \bar{x})/\sqrt{s_o}$, et dans l'espace euclidien R^{2n+1} les $n + 1$ vecteurs : $V_n = [y_o, y_1, \ldots, y_{n-1}, y_n, 0, \ldots, 0]$,

$V_{n-1} = [0, y_o, \ldots, y_{n-2}, y_{n-1}, y_n, \ldots, 0]$, ... , $V_o = [0, 0, \ldots, 0, y_o, y_1, \ldots, y_n]$ (d'où $\|V_n\| = \ldots = \|V_o\| = 1$ et $\langle V_{n-i}, V_{n-j} \rangle = r_{|i-j|}$). On note $\hat{V}_n(p) = \sum_{j=1}^{p} \theta_j^p V_{n-j}$ la projection orthogonale de V_n sur le sous-espace de R^{2n+1} engendré par $V_{n-1}, \ldots, V_{n-p}$. Les θ_j^p sont obtenus par résolution des équations de Yule-Walker d'ordre p. La série chronologique C étant considérée comme une partie d'une réalisation d'un processus AR, l'identification de l'ordre de ce processus repose sur l'examen i) du corrélogramme empirique (calculé dans la pratique jusqu'à l'ordre $k \leq \frac{n+1}{4}$) et ii) des coefficients d'autocorrélation partielle, ces derniers n'étant autres que les θ_p^p (p variant de 1 à k). Le cas échéant, la colinéarité des V_{n-i} rend instable les θ_j^p, j fixé, lorsque p varie pas à pas. En remplaçant dans les équations de Yule-Walker les coefficients de corrélation r_j par des coefficients lissés $\tilde{r}_j = \lambda_j r_j$, où les λ_j sont les termes d'une suite de type strictement positif, on considère l'analogue d'une "ridge-régression" susceptible d'atténuer ces fluctuations. A noter que cette méthode de lissage du corrélogramme empirique est utilisée pour estimer la densité spectrale des processus stationnaires.

A titre d'exemple on considère une série des écarts des moyennes journalières de relevés trihoraires de température à la "normale" du jour. Le corrélogramme lissé est obtenu en considérant une "fenêtre" de Bartlett : $\lambda_j = \max\{0, 1 - \frac{j}{K}\}$, K étant un entier fixé. On étudie les effets de ce lissage sur les estimations des coefficients d'autorégression.

RECONNAISSANCE DES FORMES EN PEDOLOGIE

Ph. BONNERIC

UNIVERSITE DE MONTPELLIER II
Cellule Informatique
Place E. Bataillon
34060 - MONTPELLIER Cédex

Le pédologue ayant découpé une carte géographique en un certain nombre de cellules de surfaces égales, il enregistre des caractères (altitude moyenne, pente moyenne etc...) pour chacune d'elles.

Il définit ensuite plusieurs formes en fonction de l'étude à réaliser (exemple : bonne aptitude au pommier, aptitude moyenne etc...), chacune étant matérialisée par des cellules de la carte, dont on connait l'aptitude, ou des cellules théoriques définies par les mêmes caractères.

Il s'agit alors d'affecter chaque cellule de la carte à la forme qui lui est la plus proche. Pour ceci, on calcule un indice de similarité (J.C. GOWER) s entre chaque cellule de la carte et chaque forme, puis on affecte la cellule à la forme avec qui elle a le plus grand s (sous réserve que s soit dans une certaine fourchette).

Le calcul s permet de pondérer les caractères. Il s'agit de déterminer les poids de telle sorte que les formes soient distinctes.

La méthode utilisée résulte des travaux de Y. ESCOUFIER et P. ROBERT qui disent que les poids optimum sont ceux qui donnentle coefficient RV maximum.

Sachant que :

$$RV(R,S) = Tr(R.S) / \left[Tr(R^2).\ Tr(S^2)\right]^{1/2}$$

S étant la matrice de similarité globale entre les cellules composant les formes.
R étant la matrice des similarités entre les formes.

Les poids ainsi déterminés sont alors injectés dans le calcul de similarité celluleforme.

Pour chaque forme on visualise graphiquement la répartition de cellules en fonction de leur indice de similarité et on fixe un seuil en dessous duquel la cellule ne sera pas considérée pour cette forme.

L'affectation cellule-forme est finalement représentée par un schéma localisant les différentes formes sur la carte de départ.

SUR UNE CLASSE D'ESTIMATEURS CENTRES A VARIANCE MINIMALE

M. CARTON

UNIVERSITE DE L'ETAT
17, place Warocqué
MONS (BELGIQUE)

Ce papier est un compte rendu succint d'une étude d'une cinquantaine de pages non encore publiée, relative à la recherche d'estimateurs centrés à variance minimale d'une fonction quelconque (algébrique ou transcendante) de paramètres d'une distribution d'un schéma probabiliste. Cette étude se base sur la borne inférieure de l'inégalité de Bhattacharyya pour laquelle l'ordre r tend vers l'infini. Cette extension permet la recherche d'estimateurs centrés de polynômes algébriques d'ordre infini (développements en série) des paramètres de la distribution. Le procédé d'obtention de ces estimateurs est pratique et relativement simple (pas d'intégration).

Ainsi, pour ne donner qu'un exemple élémentaire, la méthode donne pour l'estimateur centré à variance minimale de la fonction $\exp[\lambda\theta]$ du paramètre θ de la distribution de Poisson :

$$f(x;\theta) = \theta^x e^{-\theta}/x!$$

où λ est une constante, l'expression

$$T = \left(1 + \frac{\lambda}{n}\right)^{n\bar{x}}$$

où n est la taille de l'échantillon et $\bar{x}$ sa moyenne. La méthode donne également la variance correspondante :

$$V(T) = e^{2\lambda\theta}\left[e^{\lambda^2 \frac{\theta}{n}} - 1\right]$$

On voit immédiatement que pour de petits échantillons, l'estimateur

$$T_1 = e^{\lambda\bar{x}}$$

est loin d'être satisfaisant et que c'est la forme exacte T qu'il convient d'utiliser.

SUR UNE METHODE STATISTIQUE RELATIVE A L'EMPLOI DE L'ORDINATEUR ELECTRONIQUE DANS L'ANALYSE DES ELECTROCARDIOGRAMMES

L. V. DE CAROLIS

FACULTE DES SCIENCES STATISTIQUES
Institut de Calcul de Probabilité
UNIVERSITE DE ROME

Ce travail constitue un essai pour chercher à satisfaire aux exigences actuelles concernant l'emploi de l'ordinateur électronique, pour effectuer l'analyse d'un grand nombre d'électrocardiogrammes (E.C.G.). On a commencé par considérer un E.C.G. comme une ligne polygonale. Le fait qu'il y ait des machines qui donnent directement un E.C.G. sous forme d'une ligne brisée, peut justifier ce procédé. Par conséquent, on a utilisé l'équation de la ligne polygonale décrite par l'E.C.G. pour tirer une liaison éventuelle entre une fonction des coefficients de l'équation et le type d'E.C.G. (pathologique ou normal). On a conduit la recherche en prenant un échantillon au hasard de 60 individus (35 pathologiques et 25 normaux) dont on a considéré les douze dérivations d'un E.C.G. On a mis en évidence, en employant des tests d'hypothèse non paramètriques, que certaines fonctions des coefficients des lignes polygonales peuvent donner des valeurs propres à discriminer la plupart des E.C.G. relatifs aux individus pathologiques.

APPROXIMATION D'ANALYSES CANONIQUES NON LINEAIRES DE VARIABLES ALEATOIRES ET APPLICATIONS A LA REGRESSION ET A L'ANALYSE DISCRIMINANTE NON LINEAIRES

D. DE MICHEAUX

I M A N

UNIVERSITE de NICE Parc Valrose
06034 NICE CEDEX

Lorsqu'une grandeur est à prévoir à partir d'autres grandeurs, on utilise régression, an. de la variance, an. de la covariance ou an. discriminante selon le type (quantitatif ou qualitatif) des variables traitées. Ces diverses méthodes font toutes appel à un modèle de relations linéaires et, pour justifier les tests d'hypothèses qui les accompagnent, il faut supposer que les variables quantitatives suivent des lois normales.

Pour se libérer du modèle linéaire trop restrictif, on utilise actuellement de nouvelles démarches plus ou moins empiriques, basées sur le "découpage en classes" des variables quantitatives et sur l'an. factorielle des correspondances.

Dans cet exposé, nous présentons la notion d'approximation d'analyses canoniques non linéaires de variables aléatoires, notion qui justifie sur le plan théorique certaines de ces démarches, et donne naissance à d'autres modes de prévision : ceux-ci permettent, en particulier, d'éviter la perte d'information due au découpage en classes (codage disjonctif) des variables quantitatives.

UTILISATION DES PROCESSUS ARMA DANS L'ETUDE DES SERIES CHRONOLOGIQUES

S. DOSSOU-GBETE

H. DANG-DUC

Laboratoire de Statistique et Probabilités
E.R.A. C.N.R.S. N° 591
UNIVERSITE de TOULOUSE III,
31077 TOULOUSE Cédex

L'objet de cet exposé est de présenter certaines propriétés des processus stationnaires du second ordre.

Partant des notions de bruit blanc, de processus des innovations, de processus régulier, de processus inversible et d'opérateur retard (B), on distingue équation ARMA et processus ARMA (p,q). Soient P' et Q' deux polynômes à coefficients réels de degrés p' et q' respectivement et un bruit blanc $\{\alpha(t)\}_{t\in Z}$, on appelle équation ARMA l'équation

$$P'(B)\ X(t) = Q'(B)\ \alpha(t) \qquad (1)\ .$$

On étudie les conditions sur P' et Q' pour que (1) admette un processus régulier $(X_t)_{t\in Z}$ comme solution. On rappelle qu'une telle solution est unique si elle existe et que l'on peut alors définir deux polynômes P et Q, à coefficients réels et de degrés minimum $p\leq p'$ et $q\leq q'$, tels que

$$P(B)\ \ X(t) = Q(B)\ \varepsilon(t)$$

où $\varepsilon(t)$ est l'innovation au temps t de la solution régulière de (1).
Cette solution est appelée processus ARMA (p,q).

On développe quelques aspects de l'agrégation des processus ARMA (p,q) : agrégation diachronique, agrégation synchronique.

On étudie les méthodes d'estimations préliminaires de G.E.P. BOX et G.M. JENKINS ; ces méthodes définissent-elles toujours des processus ARMA (p,q) réguliers et inversibles ?

QUELQUES ASPECTS DE L'ANALYSE SPECTRALE DE PROCESSUS STOCHASTIQUES NON STATIONNAIRES

F. DROESBEKE

INSTITUT DE STATISTIQUE - U.L.B.
C.P. 210
Bd du Triomphe
B - 1050 - BRUXELLES

Plusieurs tentatives de généraliser les concepts spectraux introduits dans le cas stationnaire ont été envisagées, en particulier au cours de cette décennie. Notre propos est de présenter certains aspects de ces généralisations concernant :

1 - les travaux récents effectués dans l'étude de processus stochastiques non stationnaires univariés et multivariés.

2 - les problèmes posés par l'estimation des concepts spectraux introduits.

3 - le recours à l'analyse spectrale dans l'utilisation de modèles probabilistes.

LES COLLECTIONS STRUCTUREES : LE MODELE STATISTIQUE

J.C. FARGET
IREP - CRISS
Université des Sciences Sociales
B.P. 47
38040 GRENOBLE-CEDEX

Le but de cette étude est d'arriver à une bonne formalisation de l'ensemble des traitements nécessaires à la production des statistiques de base (tables de statistiques). Les données sont décrites et organisées en termes de "collections structurées" qui étend à l'ensemble des données les concepts de la tabulation. L'information est vue alors au travers d'une structuration hiérarchique dont chaque niveau correspond à une ventilation (ensemble de classes). Les ventilations apparaissent comme des objets du système, définis par les utilisateurs. Cette hiérarchie de classes sert de chemin d'accès lors des traitements des données. Ces traitements sont répertoriés en deux grandes catégories : d'une part les opérations internes à une collection (extraction, réductions et linéarisation) et d'autre part les opérations externes (concaténation, fusion, reformattage et finalement opérations arithmétiques).

En conclusion on peut considérer ces outils comme une extension d'APL aux problèmes de gestion d'une base de données orientée vers les statistiques.

SUR LES RESULTATS DE BLUM EN APPROXIMATION ET OPTIMISATION STOCHASTIQUES MULTIDIMENSIONNELLES

B. FICHET

Laboratoire de Physique
FACULTE DE MEDECINE
27, bd Jean Moulin
13385 - MARSEILLE Cédex 4

L'extension au cas multidimensionnel des processus d'approximation stochastique de Robbins-Monro et d'optimisation stochastique de Kiefer-Wolfowitz, fut l'oeuvre de Blum. On lui doit trois importants théorèmes, deux pour l'approximation et un pour l'optimisation. Depuis, de nombreux auteurs, tels Dvoresky, Derman et Sacks, Fabian, Venter, ont établi de nombreux théorèmes de convergence, tant pour ces processus que pour des processus généralisés.

Nous montrons, par cette note, que les résultats de Blum restent d'une grande actualité. Pour différentes régressions, les convergences des processus de Robbins-Monro ou de Kiefer-Wolfowitz peuvent être déduites tant des théorèmes de Blum que de théorèmes postérieurs. Ainsi, à l'aide des théorèmes de Blum, nous démontrons la convergence pour l'approximation d'une régression linéaire donnée par une matrice définie négative, sous l'hypothèse de la variance uniformément bornée ; et même sans cette hypothèse à l'aide d'un exemple. Et nous établissons de même la convergence pour l'optimisation d'une régression de type paraboloïde.

ANALYSE COMPLEXE DE DONNEES

A. GRORUD

Rce les Asters B-I
Chemin des Infirmeries
13100 - AIX en PROVENCE

L'objet du travail présenté ici est l'extension aux données complexes de quelques méthodes classiques d'analyse factorielle.

L'exposé se décompose en deux parties :

1 - Nous montrons d'abord la possibilité théorique de traiter les variables complexes par les méthodes factorielles en utilisant un produit hermitien qui remplace le produit scalaire des analyses réelles. Nous parlons en particulier :

- du traitement d'un ensemble mixte de variables (variables réelles + variables complexes) ;
- des vecteurs aléatoires complexes, des opérateurs linéaires associés et du coefficient RV de deux opérateurs ;
- du schéma de dualité complexe ;
- de la comparaison par le RV du traitement réél et du traitement complexe des données.

2 - Nous appliquons ensuite les résultats de la première partie en une étude pratique de l'Analyse en Composantes Principales Complexe sur des données de vent en Camargue.

PROCESSUS MULTIVARIES AUTOREGRESSIFS MOYENNE MOBILE A COEFFICIENTS DEPENDANT DU TEMPS : CONDITIONS D'INDETERMINABILITE PURE ET D'INVERSIBILITE

M. HALLIN

UNIVERSITE LIBRE DE BRUXELLES
Campus-Plaine - C.P. 210
Bd du Triomphe
1050 - BRUXELLES

Des conditions d'indéterminabilité pure et d'inversibilité ont été obtenues (HALLIN et MELARD - 1977) pour les processus autorégressifs - moyenne mobile univariée et à coefficients dépendant du temps. Ces conditions s'étendent assez simplement aux processus multivariés de rang constant. Les résultats classiques de la théorie des équations aux différences linéaires ne permettent pas de traiter directement le cas des processus de rang variable ; une généralisation de la notion de fonction de GREEN est introduite à cet effet. Cette généralisation permet en outre un calcul récursif, beaucoup plus simple que celui auquel conduit la définition classique.

Références :

M. HALLIN et G. MELARD : indéterminabilité pure et inversibilité des processus autoregressifs - moyenne mobile à coefficients dépendant du temps. Colloque Séries Chronologiques - Approches Fréquentielle et Temporelle - Bruxelles, mai 1977. Cahiers du CERO, 19, 1977, pp. 385-392.

APPLICATION DE LA METHODE DU COL AU DEVELOPPEMENT ASYMPTOTIQUE DE LA LOI DE PROBABILITE D'UN ESTIMATEUR DE LA CLASSE K DANS UN MODELE A EQUATIONS SIMULTANEES

A. HOLLY

UNIVERSITE PARIS IX (Dauphine)
U.E.R. Mathématiques de la Décision
Place du Mal de Lattre de Tassigny
75016 - PARIS
et
ECOLE POLYTECHNIQUE
Laboratoire d'Econométrie
17, rue Descartes
75230 - PARIS Cédex 05

De nombreux auteurs ont obtenu récemment des approximations des fonctions de répartitions d'estimateurs de paramètres, dans le cadre des modèles à équations simultanées. Ces approximations sont essentiellement du type Edgeworth ou Gram-Charlier. Dans ce travail, on étudie l'approximation de la densité de probabilité des estimateurs dits "de classe k" des paramètres d'un système à deux équations à l'aide d'une nouvelle méthode. Celle-ci, connue en analyse classique sous le nom de méthode du col, consiste à trouver un développement asymptotique d'une intégrale dépendant d'un paramètre réel, à l'aide d'une intégration dans le plan complexe.

Dans ce travail, la méthode du col est appliquée à la transformée de Laplace de la loi de l'estimateur de classe k. Par un théorème d'inversion de la transformée de Laplace, on obtient une approximation de la densité exacte.

On compare ensuite cette approximation avec la densité exacte, des estimateurs des moindres carrés ordinaires et des doubles moindres carrés qui sont les cas particuliers les plus importants d'estimateurs de la classe k.

Des comparaisons numériques entre l'approximation ainsi obtenue et les densités exactes, d'une part, et les approximations du type d'Edgeworth, d'autre part, ont été effectuées. Elles montrent que l'approximation par la méthode du col conduit à des résultats très satisfaisants.

UNE PROCEDURE DE CONSTRUCITON DE MODELES ARMA MULTIVARIES EXEMPLE D'APPLICATION EN UTILISANT LE MODELE ARMA ET LE FILTRE DE KALMAN

J.P. INDJEHAGOPIAN *

L'objet de la communication est d'aborder une classe particulière de processus stochastiques vectoriels stationnaires : les modèles ARMA multivariés : $\Phi(B)\underset{\sim}{z}_t = \theta(B)\underset{\sim}{a}_t$.

Une procédure pour modéliser les séries temporelles multidimensionnelles à l'aide de modèles ARMA est développée. Cette procédure consiste à identifier, estimer et contrôler le modèle ARMA (p,q) multivarié.

Les différentes phases, pour construire le modèle, s'appuient respectivement sur l'estimation de la matrice des autocorrélations au décalage k, une généralisation de la méthode d'estimation des paramètres d'un modèle ARMA univarié et sur l'estimation des coefficients de corrélation croisés du vecteur bruit.

A titre d'exemple, on considère les modèles bivariates autorégressifs AR(1), moyenne mobile MA(1) et autorégressif - moyenne mobile ARMA(1,1). A cette occasion, on montre que le modèle de Box et Jenkins avec fonction de transfert pour une variable exogène est un cas particulier de modèle bivariate.

Enfin, cette étude présente une utilisation du filtre de Kalman lorsque les paramètres d'un modèle de régression multiple évoluent au cours du temps selon un processus ARMA multivarié i.e.

$$\Phi(B)\underset{\sim}{b}_{t+1} = \theta(B)\underset{\sim}{a}_{t+1}$$

$$y_t = x_t' \underset{\sim}{b}_t + \varepsilon_t$$

* Professeur de Statistiques à l'Ecole Supérieure des Sciences Economiques et Commerciales - BP 105 - 95001 CERGY.

TESTS DE RANG ET PROCESSUS AUTOREGRESSIFS D'ORDRE UN

J.F. INGENBLEEK

Institut de Statistique
UNIVERSITE LIBRE DE BRUXELLES
Campus Plaine - C.P.210 - Bd du Triomphe
1050 - BRUXELLES

Soit $x_1, x_2, \ldots, x_n$ une réalisation de longueur finie d'un processus aléatoire. On connaît [1] des statistiques linéaires de rang pour tester l'hypothèse que le processus générateur est un bruit blanc stationnaire, par rapport à l'alternative que ce processus comporte une tendance linéaire. On considère ici comme hypothèse alternative un processus autorégressif d'ordre un. On examine grâce à la méthode développée par Hajek et Sidak des conditions suffisantes à imposer aux scores pour que la statistique de rang associée soit asymptotiquement normale. On analyse également la normalité dans le cas d'une alternative contigüe.

Référence : [1] Hajek-Sidak : Theory of rank test. Academic Press, New York (1967).

INTERPOLATION DES SUITES STATIONNAIRES OCCULTEES

B. LACAZE

UNIVERSITE PAUL SABATIER
Laboratoire de Statistique et Probabilité
U.E.R. Mathématique Informatique et Gestion
31000 - TOULOUSE

Soit $\{X_n\}$, $n \in Z$, une suite stationnaire au sens large, et $\{A_n\}$, une suite de v.a indépendante de la précédente, et prenant les valeurs 0 et 1 seulement. On désire estimer X_0 à l'aide de réalisations de la suite $\{A_n X_n\}$, $n \in Z$.

La procédure habituelle, dite de Wiener, consiste à projeter X_0, élément d'un espace de Hilbert de type L^2, sur le sous-espace généré par la suite $\{A_n X_n\}$, $n \in Z$. L'estimateur $\tilde{X}_0$ de X_n est alors de la forme

$$\tilde{X}_0 = \sum_k a_k A_k X_k$$

la suite des a_k étant solution du système d'équations :

$$E\left((X_0 - \tilde{X}_0) A_k X_k \right) = 0 \quad \forall k \in Z.$$

L'estimateur ainsi défini a le grave défaut suivant : la suite $\{a_k\}$ ne dépend pas des observations des A_k. Ainsi, si on observe que $A_0 = 1$, l'observation de $\tilde{X}_0$ sera différente de celle de X_0.

D'où l'idée de l'estimateur "conditionnel" suivant qui évite cet inconvénient. Posons :

$$\begin{cases} \tilde{X}_0 = X_0 \text{ si } A_0 = 1 \\ \tilde{X}_0 = \sum_k a_k A_k X_k \quad \text{si } A_0 = 0 \end{cases}$$

la suite des a_k étant définie par le système d'équations :

$$E\left((X_0 - \tilde{X}_0) A_k X_K \mid A_0 = 0 \right) = 0 \quad \forall k \in Z^*$$

On étudie cet estimateur dans les deux cas suivants :

1 - les v.a A_k sont mutuellement indépendantes

2 - la suite des v.a A_k est stationnaire au sens large.

ANALYSE DES CORRESPONDANCES SUR JUXTAPOSITION DE SOUS TABLEAUX HETEROGENES

J.P. LAMARCHE et M. SUEUR

UNIVERSITE D'ORLEANS
Département de Mathématiques
Faculté des Sciences
45045 - ORLEANS Cédex

Nous disposions de données consistant en résultats à des exercices posés à des classes de 6èmes, visant à trouver des liens entre la formulation des problèmes additifs et soustractifs et la compréhension qu'en ont les élèves, plusieurs séries d'exercices étant posés chacunes à un groupe de classe.

Nous avions donc un "pseudo" tableau de contingence, constitué par la superposition de plusieurs sous tableaux, correspondant chacun à une série d'exercices et des classes différentes. Pour l'étudier, nous avons été amenés à utiliser une méthode d'"analyse des correspondances partielles" pour comparer entre elles les différentes séries d'exercices, en juxtaposant à l'analyse globale de tout le tableau celle de chaque sous tableau, ou groupe, avec une métrique de χ^2 globale, compatible entre toutes les analyses. Nous calculons les inerties correspondantes : inertie des groupes, et inertie inter-groupes.

Sur deux séries de données, ceci nous a permis de voir qu'un des deux tableaux (dont l'inertie inter-groupe était plus forte) était plus concluant pour l'étude pédagogique. La projection simultanée dans le "plan principal" de chaque groupe des exercices et des classes permet de comparer l'influence de la formulation des différentes séries d'exercices. Enfin, par un tracé Benson, nous pouvons projeter l'ensemble des classes et des exercices sur le plan principal d'un des groupes (plus particulièrement intéressant, par exemple), ainsi que les centres de gravité des groupes et leurs axes principaux d'inertie.

ANALYSES STATISTIQUES DE DONNEES DE POLLUTION ATMOSPHERIQUE DE LA REGION DE MONTREAL

U.R. MAAG

Centre de Recherche en Informatique et Gestion
UNIVERSITE des SCIENCES et TECHNIQUES du LANGUEDOC
34075 - MONTPELLIER Cédex

Les analyses portent sur les polluants suivants : l'anhydride sulfureux (SO_2), l'indice de souillure (COH), l'ozone (O_3) et le monoxide de carbone (CO). Certaines de ces mesures sont disponibles depuis 1968, le nombre de stations considérées est de 14. Deux approches statistiques ont été faites, l'une descriptive, l'autre à l'aide de modèles de régression et de séries chronologiques.

En addition des méthodes descriptives unidimensionnelles classiques, des techniques de représentations de données multidimensionnelles ont été employées telles que les faces de Chernoff, les courbes d'Andrews et STATIS. Les conclusions suivantes se dégagent : la pollution atmosphérique est nettement supérieure en hiver qu'en été ; les polluants SO_2, COH, et CO sont très liés et peuvent servir à définir un indice de pollution ; on observe un aspect local qui permet de diviser les zones en trois grandes catégories : la proximité des raffineries, la zone urbaine proprement dite (le secteur commercial du centre-ville, les secteurs de densité et d'industrialisation moyenne ou faible) et les espaces verts.

Des modèles de séries chronologiques servent à décrire la saisonnalité et l'évolution à moyen terme. Ceci permet d'observer des effets d'intervention comme par exemple un règlement municipal fixant le taux de souffre permis dans l'huile de chauffage. Des modèles de régression montrent comment les mesures de pollution dépendent de certaines variables météorologiques.

MODELES ARIMA POUR DES SERIES CHRONOLOGIQUES NON HOMOGENES

G. MELARD

UNIVERSITE LIBRE DE BRUXELLES
Institut de Statistique
Campus Plaine - C.P. 210
Bd du Triomphe
B-1050 - BRUXELLES

Cet exposé présente une ébauche d'étude comparée de deux approches pour la représentation de séries chronologiques comportant une tendance en dispersion, qui dérivent toutes deux de la méthode de BOX et JENKINS.

La première approche consiste à utiliser la méthode de BOX et COX pour déterminer la transformation à appliquer à la série afin de la rendre homogène.

La seconde approche, due à l'auteur, se base sur une classe particulière de modèles ARIMA à coefficients dépendant du temps. Les spécifications sont validées au moyen d'une batterie de tests qui ont pour but de faire apparaître l'incapacité éventuelle du modèle à représenter la série à certaines époques de chaque année, lorsque certains niveaux de la variable sont atteints ou à certaines périodes de l'intervalle d'observation. L'étude comparée porte sur quelques séries présentées dans la littérature ainsi que sur des séries macroéconomiques. Les deux approches conviennent différemment selon les cas, bien que la transformation de BOX et COX semble préférable pour les séries macroéconomiques.

TEST ET ESTIMATION POUR UNE CLASSE DE PROCESSUS DE NAISSANCE BIDIMENSIONNELS

P. PICARD

UNIVERSITE CLAUDE BERNARD (LYON 1)
Département de Mathématiques
43, bd du 11 Novembre 1918
69621 - VILLEURBANNE

Les processus linéaires N.D. (lire "de naissance et de décès") rentrent dans la classe $\mathcal{C}$ des processus qui sont N.D. et pour lesquels

$$P(N_{t+\Delta t} = j / N_t = i) = \begin{cases} \lambda f(t,i)\,\Delta t + o(\Delta t) & \text{si } j=i+1 \\ \mu f(t,i)\,\Delta t + o(\Delta t) & \text{si } j=i-1 \\ o(\Delta t) & \text{si } |j-i|>1 \end{cases}$$

avec $f : \mathbb{R}_+ \times \mathbb{N} \to \mathbb{R}_+$ telle que $\forall t > o, f(t,o)=o$ et λ et μ constants. Si $i \to f(t,i)$ est linéaire, on retrouve un processus N.D. linéaire qui peut être rendu homogène par simple changement de l'échelle des temps. Pour de tels processus, on sait formuler $P(N_t=n/N_o=n_o)$ et résoudre divers problèmes statistiques (test et estimation). On se propose ici d'étendre certains de ces résultats à la classe $\mathcal{C}$ toute entière. Bien qu'on ne puisse pas en général expliciter $P(N_t=n/N_o=n_o)$, la théorie des martingales permet, pour certains temps d'arrêt, de formuler la distribution conjointe de N_T et $\int_o^T f(u,N_u)\,du$. Ceci suggère les schémas de recueil des données statistiques à utiliser, et permet ensuite de construire pour λ et μ des estimateurs du maximum de vraisemblance, ainsi que leur distribution conjointe, aussi bien que de résoudre des problèmes de test du type NEYMAN-PEARSON.

Comme un processus N.D. se ramène à un processus de naissance bidimensionnel (on pose $N_t = X_t - Y_t + n_o$ avec X_t (resp. Y_t) nombre des naissances (resp. des décès) enregistrés entre les dates 0 et t) la théorie peut se généraliser aisément aux processus de naissance bidimensionnel (ou multidimensionnels).

APPROCHE MACRO-ECONOMIQUE DE L'ACTIVITE ACTUARIELLE FONDEE SUR LA THEORIE DES EQUATIONS DIFFERENTIELLES STOCHASTIQUES

M. PINHAS

UNIVERSITE DES SCIENCES SOCIALES
Place Anatole France
31070 - TOULOUSE Cédex

Le modèle que nous proposons est une approche macro-économique de l'activité actuarielle fondée sur la théorie des équations différentielles stochastiques. Quatre applications en ont dérivé.

1 - ACTUALISATION DU FONDS DE RESERVE ET CHARGEMENT DE SECURITE :
les méthodes de calcul du chargement de sécurité, lorsqu'elles ont pour fondement la probabilité de ruine relative à un exercice de durée infinie, ne tiennent pas compte en général de l'actualisation du fond de réserve. L'effet correspondant jugé à priori négligeable ne justifie pas l'accroissement considérable des difficultés techniques nécessaires à sa détermination. En revanche, le présent modèle de diffusion conduit à une évaluation simple de l'effet d'actualisation à partir des paramètres essentiels de l'activité actuarielle considérée.

2 - UN DESCRIPTEUR D'ACTIVITE ACTUARIELLE : L'ESPERANCE ARRETEE :
il s'agit ici de construire un indice simple pour classer des activités actuarielles en tenant compte de l'évolution moyenne du fond de réserve et de la probabilité de ruine. Est aussi envisagée une solution d'entente compagnie-assurés susceptible d'améliorer conjointement les situations des deux partenaires.

3 - ESPERANCE ARRETEE ET EQUILIBRE CONCURRENTIEL :
une autre application de l'indice "espérance arrêtée" réside en la définition d'un équilibre concurrentiel sur le marché des assurances, en lequel apparaît clairement l'influence des taux d'intérêt à court et long terme, des fonds de réserve initiaux, de la dispersion du risque et de l'horizon économique pris en considération.

4 - VALEUR DE SHAPLEY ET ESPERANCE ARRETEE : DEFINITION D'UNE ENTENTE ACTUARIELLE :
la méthode de l'espérance arrêtée voit son champ d'application considérablement élargi par la théorie des jeux. Associée au concept "valeur de Shapley" elle permet d'établir les règles d'une entente entre compagnies d'assurances.

ESTIMATION DE FORMES CONVEXES DU PLAN

Dr. J.P. RASSON

FACULTES UNIVERSITAIRES de NAMUR - BELGIQUE

Nous suggérons une solution au problème suivant proposé par le Prof. D.G. KENDALL : "supposons qu'une réalisation d'un processus de Poisson dans le plan ait été mutilée de telle façon que nous ne puissions observer que les points intérieurs à un domaine convexe compact D, inconnu. Trouvez D en utilisant des méthodes d'inférence statistique".

Ceci est présenté comme l'analogue, dans le plan, du célèbre Problème du Taxi dont la généralisation unidimensionnelle est l'estimation des deux bornes inconnues d'un intervalle alors que nous observons un nombre fixé de points à l'intérieur de celui-ci. Nous justifions le choix des estimations classiques pour ce problème par les principes de l'estimation équivariante appliqués à la recherche d'estimateurs optimaux (U.M.V.).

Nous attirons alors l'attention sur la différence essentielle entre ce problème et son analogue dans le plan ; ici, il y a une nouvelle inconnue appelée la forme. Ce paramètre, qui n'est pas de dimension finie, apparaît comme un paramètre de nuisance pour l'estimation de la position et de la surface. Nous procédons à son élimination en le remplaçant par la forme qui maximise la vraisemblance marginale de ce paramètre et de la surface. Nous donnons alors les estimateurs de l'aire et de la position du domaine D qui correspondent au choix que nous avions fait dans le cas unidimensionnel, correspondance que nous détaillons. Notre solution finale sera une homothétie de l'enveloppe convexe de l'échantillon à partir du centre de gravité de celle-ci.

Finalement, nous signalons comment cette solution se généralise naturellement à R^n et nous donnons les possibilités d'application dans différents domaines.

SUR LE COMPORTEMENT ASYMPTOTIQUE DES AUTOCOVARIANCES ET DES AUTOCORRELATIONS D'UNE SERIE NON-STATIONNAIRE

R. ROY

UNIVERSITE DE MONTREAL
Département d'Informatique et
de Recherche Opérationnelle
Univ. de MONTREAL

CP 6128 SUCCURSALE A MONTREAL PQ CANADA

Soit $Z_1, \ldots, Z_N$ une série chronologique de longueur N et soient

$$c_k = (1/N) \sum_{t=1}^{N-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z}), k = 0, 1, \ldots, K$$

$$r_k = c_k / c_0 , \; k = 1, \ldots, K$$

les fonctions d'autocovariance et d'autocorrélation correspondantes. Lorsque le processus aléatoire décrivant la série est stationnaire, les propriétés asymptotiques $(N \to \infty)$ des c_k et des r_k sont bien connues et en particulier, les fonctions d'autocovariance et d'autocorrélation échantillonnales convergent respectivement vers les fonctions d'autocovariance et d'autocorrélation du processus. Cependant, dans le cas non-stationnaire, relativement peu de choses sont connues sur le comportement des c_k et des r_k. Dans cet exposé, nous présentons quelques propriétés des c_k et des r_k lorsque le processus aléatoire $\{Z_t : t = 0, \pm 1, \ldots\}$ décrivant la série est un processus intégré, c'est à dire dont la première différence $\{w_t = Z_t - Z_{t-1}, \; t = 0 \pm 1, \ldots\}$ est un processus stationnaire. Nous explicitons les deux premiers moments asymptotiques des c_k et nous montrons que pour chaque entier k, $k = 1, \ldots, K$, l'autocorrélation r_k converge vers un en probabilité. L'implication de ces résultats pour l'identification d'un modèle intégré est aussi discutée.

LE PROCESSUS STOCHASTIQUE ET MULTIPARAMETRIQUE DEFINI PAR UNE ESTIMATION DE LA DENSITE DE PROBABILITE

W. SCHLEE

INST. F. STATISTIK U. UNTERNF
Techn. Universität
Arcisstrasse 21
D-8000 MUNCHEN ALLEMAGNE FEDERALE

L'objet de l'étude est une estimation de la densité de probabilité bien adaptée au besoin de calcul automatique.

Cette méthode d'estimation est stimulée par un article de L.I. BONEVA, D. KENDALL et I. STEFANOV : "Spline Transformations : Three New Diagnostic Aids for the Statistical Data-Analyst", Journal of the Royal Statistical Soc. Ser. B 33(1971)1-70.

Dans le texte, ici, il est recherché une généralisation multidimensionnelle, la loi asymptotique de probabilité et la convergence en moyenne quadratique pour tous les éléments particuliers de l'espace R^q. En outre, la convergence en probabilité du processus stochastique, défini par cette estimation de la densité, est démontrée.

Pour être bref, seulement, l'estimation de la densité est notée ici dans le cas de dimension un :

$$f_n(x) = h^{-1} \sum_{i=-\infty}^{\infty} k\left(\frac{x - i\beta}{h}\right) \left(F_n(i\beta) - F_n((i-1)\beta)\right)$$

n dénote les nombres des épreuves, F_n la répartition empirique, h un paramètre dépendant de n, β un paramètre dépendant de h.

Il est supposé que $\lim_{n \to \infty} h(n) = 0$, $\lim_{h \to 0} \beta(h)/h = 0$

k est une propre fonction.

La généralisation multidimensionnelle est obtenue en remplaçant k par un produit de q facteurs et en remplaçant $F_n(i\beta) - F_n((i-1)\beta)$ par la propre différence à plusieurs dimensions.

ENSEIGNEMENT DU CALCUL DES PROBABILITES ET DE LA STATISTIQUE DANS LE SECONDAIRE

J. TEGHEM

UNIVERSITE LIBRE DE BRUXELLES
Département de Mathématiques
C.P. Campus Plaine - bd du Triomphe
B - 1050 - BRUXELLES

Difficultés liées à un enseignement rigoureux des fondements du calcul des probabilités.

Choix du niveau (espaces finis ou dénombrables ou continus) et de la méthode.

Utilité d'un enseignement d'initiation à l'inférence statistique.

Choix du niveau de cet enseignement et du type d'applications. Adaptation à la catégorie d'auditeurs.

Examen critique de programmes et d'ouvrages et avis exprimés aux tables rondes de l'Institut International de Statistique sur le sujet.