
GROUPE DE TRAVAIL STAPH :
STATISTIQUE FONCTIONNELLE ET OPÉRATORIELLE

Partie V : Recueil de résumés 2003-2004

Coordinateurs

A. BOUDOU, H. CARDOT, F. FERRATY, Y. ROMAIN,
P. SARDA, P. VIEU et S. VIGUIER-PLA

Résumé

Ce document a pour objectif de présenter les résumés (plus ou moins détaillés selon les souhaits de leurs auteurs) des divers exposés qui ont eu lieu lors des séances du groupe de travail STAPH durant l'année universitaire 2003-2004.

Rappelons que ce groupe de travail en Statistique Fonctionnelle et Opératoire, créé il y a cinq ans au sein du Laboratoire de Statistique et Probabilités de Toulouse, s'inscrit dans la dynamique actuelle autour des divers aspects fonctionnels de la statistique moderne. Les exposés qui sont présentés traitent de divers aspects de la Statistique Fonctionnelle (estimation nonparamétrique, statistique opératoire, modèles de réduction de dimension, modèles pour variables fonctionnelles, ...); ils sont de nature différentes (exposés didactiques ou bibliographiques, exposés de résultats nouveaux en Statistique Appliquée et/ou Théorique, ...); ils témoignent enfin de la perpétuelle ouverture de la démarche par la grande diversité des exposants.

En préambule de ce document, un court texte est présenté afin de tirer le bilan de nos activités et afin surtout de mieux préparer l'avenir en faisant perdurer cette dynamique de recherche.

Pour terminer signalons que l'intégralité des activités de ce groupe de travail est disponible sur notre page web :

[http : //www.lsp.ups - tlse.fr/Fp/Ferraty/staph.html](http://www.lsp.ups-tlse.fr/Fp/Ferraty/staph.html)

Abstract

We present the abstracts (of size more or less important according to the wishes of their authors) of the several different talks given during the sessions of the working group STAPH along the academic year 2003-2004. This group in Functional and Operatorial Statistics is born five years ago at the Laboratoire de Statistique et Probabilités of the Université Paul Sabatier de Toulouse, and its aim was to participate at the actual dynamic existing around the different functional features of modern statistics.

These talks were about different functional topics (nonparametric estimation, statistics of operators, models for functional data, models for dimension reduction, ...). They were of different kinds (didactic, bibliographic, applied, theoretic, ...) and were presented by a large variety of statisticians.

As a foreword, a short text is presented to take the stock of the activities of this group in order to make an efficient preparation of the future.

As a conclusion, note that all the activities of this group are reachable through the following web address :

[http : //www.lsp.ups - tlse.fr/Fp/Ferraty/staph.html](http://www.lsp.ups-tlse.fr/Fp/Ferraty/staph.html).

Sumario

This documento presenta resumens (mas o menos cortos segun los deseos de sus autores) de charlas que han sido presentadas durante las sesiones de trabajo del grupo STAPH durante el ano academico 2003-2004. Este grupo de trabajo en el campo de Estadistica Funcional y Operatorial ha sido creado hace cuatros anos en el Laboratoire de Statistique et Probabilités de l'Université Paul Sabatier de Toulouse, para animar investigaciones en varios aspectos funcionales de la estadística moderna.

Estas conferencias fueran sobre temas variados (estimacion noparametrica, estadística de operadores, modelos para variables funcionales, modelos de reduccion de dimension, ...) y fueran de tipos diferentes (conferencias didacticas o bibliograficas, presentacion de resultados nuevos en estadística teorica o/y applicada, ...).

Al principio del documento, empezamos con un corto texto de presentacion en lo cual hacemos un chequeo de las actividades de este grupo y en lo cual planteamos los fundamentos para el proximo futuro.

Al final, queremos apuntar que todas nuestras actividades pueden ser consultadas a la direccion :

[http : //www.lsp.ups - tlse.fr/Fp/Ferraty/staph.html](http://www.lsp.ups-tlse.fr/Fp/Ferraty/staph.html).

TABLE DES MATIERES

Résumé/Abstract/Summario.	3
Introduction.	7
Christophe CRAMBES : Estimation Spline de quantiles conditionnels pour variables explicatives fonctionnelles.	9
Célestin KOKONENDJI : Problèmes de données surdispersées pour les modèles de comptage.	13
Nour-Eddin EL FAOUZI : Quelques problèmes statistiques en ingénierie du trafic routier.	15
Cornelia ENACHESCU : Une approche par réseau de neurones pour la prédiction du niveau de l'antigène spécifique de prostate.	17
Sylvie VIGUIER-PLA : Récents développements sur la comparaison de plusieurs populations par analyses factorielles.	19
Jean AVEROUS : M- Quantiles et profondeurs : le cas unidimensionnel.	23
Sandrine POIRAUD-CASANOVA : Identification de points non centraux par régression et application à la mesure d'efficacité d'une activité médicale.	25
Abdelaati DAOUIA : Analyse non-paramétrique des frontières de production et des mesures d'efficacité à l'aide de quantiles conditionnels non-standards.	27
Mikhail NIKULIN : Statistical Modeling in Survival Analysis and Reliability : Recent results.	29
Jean AVEROUS : M- Quantiles et profondeurs : le cas multidimensionnel.	31
Ali GANNOUN* , J. SARACCO , W. URFER et G. BONNEY : Nonparametric analysis of replicated microarray experiments.	33
Roger LAFOSSE : 20 ans de recherche : un bilan.	35
Salim LARDJANE : Vitesse optimale du cas i.i.d. pour l'estimation non-paramétrique de la densité invariante d'un système dynamique chaotique.	37
Antoine DE FALGUEROLLES : Modèles de rang réduit pour l'analyse des données.	39
Rima IZEM * , J.S MARRON , J.G KINGSLOVER : Analyse de Variation Nonlinéaire de Données Fonctionnelles, Nouvelles Perspectives en Biologie Evolutive.	41
Dominique MIZERE* , Simplice DOSSOU-GBETE and A. KIYINDOU : The translated Poisson distribution for fitting underdispersed discrete data : application to the discrimination of host plants ...	43
Sommaire des exposés des années précédentes.	45

¹ STAPH : Bilan de l'année 2003-2004 et perspectives

**Alain Boudou, Hervé Cardot, Frédéric Ferraty
Yves Romain, Pascal Sarda, Philippe Vieu et Sylvie Viguier-Pla**

Coordinateurs du groupe de travail STAPH
Laboratoire de Statistique et Probabilités

boudou@cict.fr, cardot@toulouse.inra.fr, ferraty@cict.fr, romain@cict.fr
sarda@cict.fr, vieu@cict.fr, viguier@cict.fr

Avec 5 années d'existence au sein du Laboratoire de Statistique et Probabilités de Toulouse, notre groupe de travail STAPH qui est (faut-il le rappeler ?) axé sur les différents aspects fonctionnels et opératoriels de la Statistique moderne, a continué cette année ses activités sur le rythme désormais habituel de (environ) une séance chaque quinze jours.

Bien sûr, la plupart de nos séances ont porté sur des thèmes qui nous sont chers ici à Toulouse, avec une attention spéciale cette année sur diverses questions liées aux notions de quantiles, mais comme chaque année nous avons aussi consacré quelques séances à des questions *a priori* plus éloignées de nos préoccupations immédiates qui ont été autant d'occasion de tisser ou renforcer des liens scientifiques avec d'autres équipes de statisticiens. Enfin, nous avons essayé de réaliser un équilibre entre des séances centrées sur des applications statistiques et d'autres plus axées sur les aspects théoriques. Ce souhait de ne jamais privilégier les aspects théoriques ou les aspects appliqués des choses, mais plutôt de les faire inter-agir en permanence, est inhérent à notre conception de la recherche en Mathématiques Appliquées et continuera à nous guider dans nos activités futures.

Cet été, notre groupe de travail s'est impliqué dans l'organisation de sessions invitées portant sur nos thèmes de recherche privilégiés lors de deux congrès internationaux : les rencontres franco-canadiennes de mathématiques de Toulouse (en Juillet) et le congrès COMPSTAT de Prague (en Août). La préparation de ces deux événements importants a été l'occasion de nouer/renforcer des collaborations et nul doute que leur déroulement sera l'occasion de diffuser/partager des développements récents. Ainsi, malgré le succès des deux éditions passées de nos Journées de Statistique Fonctionnelle et Opératorielle (Juin 2002 et Juin 2003),

¹Toutes nos activités sont accessibles sur la page web

<http://www.lsp.ups-tlse.fr/Fp/Ferraty/staph.html>

Par ailleurs, les sommaires de nos activités passées sont présentés à la fin de ce document.

il nous a paru scientifiquement plus judicieux de ne pas en organiser cette année et de remettre leurs troisième édition à 2005.

Outre ces aspects directement liés aux activités de notre groupe, nous avons constaté avec plaisir l'engouement de plus en plus affiché par la communauté statistique internationale pour les aspects fonctionnels de notre discipline. C'est donc tout naturellement que nous poursuivrons nos activités l'année prochaine, en nous souhaitant toujours plus ouverts sur des thèmes connexes ou sur de nouvelles collaborations.

En remerciant tous nos intervenants et participants et en leur souhaitant de bonnes vacances,

Toulouse, Juillet 2004

Estimation spline de quantiles conditionnels pour variables explicatives fonctionnelles

Christophe CRAMBES

* Adresse pour correspondance :

bureau 2 bâtiment 1R2, Laboratoire de Statistique et Probabilités, Université
Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex, France
e-mail : crambes@cict.fr

Séance du 13 Octobre 2003

Résumé

Ce travail réalisé sous la direction de Pascal SARDA (GRIMM, Université Toulouse-Mirail, 31058 Toulouse Cedex, France) et Hervé CARDOT (Unité Biométrie et Intelligence Artificielle, INRA, Toulouse, BP 27, 31326 Castanet-Tolosan Cedex, France).

Soit $(X_i, Y_i)_{i=1, \dots, n}$ un échantillon de couples indépendants de variables aléatoires définies sur un même espace de probabilités, de même loi que (X, Y) . La variable explicative X est à valeurs dans un espace fonctionnel H (dans l'exposé, $H = L^2([0, 1])$, l'espace des fonctions de carré intégrable sur l'intervalle $[0, 1]$), et Y est la variable aléatoire réponse à valeurs dans \mathbb{R} .

Soit $\alpha \in [0, 1]$ fixé ; le quantile conditionnel d'ordre α de Y sachant $[X = x]$, $x \in H$, est le réel $g_\alpha(x)$ défini par

$$P^{[X=x]}(Y \leq g_\alpha(X)) = \alpha, \quad (1)$$

où $P^{[X=x]}$ désigne la probabilité conditionnellement à l'événement $[X = x]$.

L'opérateur $g_\alpha : H \rightarrow \mathbb{R}$, appelé *opérateur quantile conditionnel d'ordre α* , définit ainsi le modèle statistique

$$Y_i = g_\alpha(X_i) + \epsilon_i, i = 1, \dots, n, \quad (2)$$

où $\epsilon_1, \dots, \epsilon_n$ sont n variables aléatoires réelles indépendantes identiquement distribuées vérifiant la relation $P^{[X=x]}(\epsilon_i \leq 0) = \alpha$. On suppose de plus que ϵ_i est indépendant de X_i .

Supposons aussi que g_α soit un opérateur linéaire continu de H dans \mathbb{R} ; le théorème de Riesz nous assure l'existence d'une unique fonction Ψ_α appartenant

à H telle que $g_\alpha(X_i) = \langle \Psi_\alpha, X_i \rangle = \int_0^1 \Psi_\alpha(t) X_i(t) dt$, $i = 1, \dots, n$, la notation $\langle \cdot, \cdot \rangle$ désignant le produit scalaire usuel de H .

Notons enfin que $g_\alpha(x)$ peut être défini de manière équivalente comme solution du problème de minimisation

$$\min_{a \in \mathbb{R}} E^{[X=x]}(l_\alpha(Y - a)), \quad (3)$$

l_α étant définie par $l_\alpha(u) = |u| + (2\alpha - 1)u$ (voir Koenker et Bassett, 1978).

Généralisant les travaux de He et Shi (1994) d'une part et de Cardot, Ferraty et Sarda (2003) d'autre part, on se propose alors de construire un estimateur spline de la fonction Ψ_α .

Soit $q \in \mathbb{N}$ et $k = k_n \in \mathbb{N}^*$. On se donne $k - 1$ nœuds (équirépartis dans la suite) définissant une subdivision de $[0, 1]$ en k sous-intervalles. Une fonction spline est une fonction polynômiale de degré q par morceaux sur chaque sous-intervalle de $[0, 1]$, et $(q - 1)$ fois dérivable sur $[0, 1]$. L'espace de ces fonctions splines est un espace vectoriel de dimension $k + q$. Une base de cet espace vectoriel est l'ensemble des fonctions B -splines, que l'on notera $B_{k,q} = {}^t(B_1, \dots, B_{k+q})$ (voir de Boor, 1978).

On estime alors Ψ_α par une combinaison linéaire des B_l , $1 \leq l \leq k + q$: on est ramené à trouver un vecteur $\hat{\theta} = {}^t(\hat{\theta}_1, \dots, \hat{\theta}_{k+q}) \in \mathbb{R}^{k+q}$ tel que $\widehat{\Psi}_\alpha = {}^t B_{k,q} \hat{\theta}$ avec $\hat{\theta}$ solution du problème de minimisation suivant

$$\min_{\theta \in \mathbb{R}^{k+q}} \left\{ \frac{1}{n} \sum_{i=1}^n l_\alpha(Y_i - \langle {}^t B_{k,q} \theta, X_i \rangle) + \rho \| ({}^t B_{k,q} \theta)^{(m)} \|^2 \right\}, \quad (4)$$

où $({}^t B_{k,q} \theta)^{(m)}$ est la dérivée d'ordre m de ${}^t B_{k,q} \theta$, ρ est un paramètre de pénalisation dont le but est de contrôler le degré de "régularité" de l'estimateur cherché et $\|\cdot\|$ est la norme associée au produit scalaire de H .

On s'intéresse enfin à la convergence de cet estimateur. Pour cela, on introduit les notations suivantes. Soit $\Gamma : H \rightarrow H$ l'opérateur de covariance de X tel que pour tout h appartenant à H , $\Gamma h = \mathbb{E}(\langle X, h \rangle X)$. De plus, pour tout h appartenant à H , on définit par $\|h\|_2^2 = \langle \Gamma h, h \rangle$ la norme induite par l'opérateur positif Γ .

Afin d'établir le résultat de convergence pour l'estimateur $\widehat{\Psi}_\alpha$, on suppose que les hypothèses suivantes sont vérifiées.

$$(H.1) \quad \|X_i\| \leq C_0 < +\infty, \quad ps.$$

La fonction Ψ_α est supposée être p' fois dérivable et $\Psi_\alpha^{(p')}$ vérifie

$$(H.2) \quad \left| \Psi_\alpha^{(p')}(t) - \Psi_\alpha^{(p')}(s) \right| \leq C_1 |t - s|^\nu, \quad s, t \in [0, 1],$$

où $C_1 > 0$ et $\nu \in [0, 1]$. Dans la suite, on pose $p = p' + \nu$ et on suppose que $q \geq p \geq m$.

$$(H.3) \quad \text{Les valeurs propres de } \Gamma \text{ sont strictement positives.}$$

$$(H.4) \quad \text{On suppose que les } \epsilon_i \text{ ont une densité } f_\epsilon \text{ telle que } f_\epsilon(0) > 0.$$

Théorème 1 *Sous les hypothèses (H.1) – (H.4) et si on suppose de plus qu'il existe $\beta, \gamma \in]0, 1[$ tels que $k_n \sim n^\beta$ et $\rho \sim n^{(\gamma-1)/2}$, alors*

(i) $\widehat{\Psi}_\alpha$ existe sauf sur un ensemble dont la probabilité tend vers 0 lorsque n tend vers l'infini,

$$(ii) \|\widehat{\Psi}_\alpha - \Psi_\alpha\|_2^2 = O_P\left(\frac{1}{k_n^{2p}} + \frac{k_n}{n\rho} + \rho + \rho k_n^{2(m-p)}\right).$$

Références

- Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline Estimators for the Functional Linear Model. *Statistica Sinica*, **13**, 571-591.
- Dauxois, J., Pousse, A. et Romain, Y. (1982). Asymptotic theory for the principal component analysis of a random vector function : some applications to statistical inference. *J. of Mult. Analysis*, **12**, 136-154.
- de Boor, C. (1978). *A Practical Guide to Splines*. Springer, New-York.
- He, X. et Shi, P. (1994). Convergence Rate of B -Spline Estimators of Nonparametric Conditional Quantile Functions. *Nonparametric Statistics*, **3**, 299-308.
- Koenker et Bassett (1978). Regression Quantiles. *Econometrica*, **46**, 33-50.
- Lejeune, M. et Sarda, P. (1988). Quantile Regression : A Nonparametric Approach. *Computational Statistics and Data Analysis*, **6**, 229-239.
- Poiraud-Casanova, S. et Thomas-Agnan, C. (1998). Quantiles Conditionnels. *Journal de la Société Française de Statistique*, **139**, 31-44.
- Ramsay, J.O. et Silverman, B.W. (1997). *Functional Data Analysis*. Springer-Verlag.
- Stone, C. (1982). Optimal rates of convergence for nonparametric models. *Ann. Statist.*, **10**, 1040-1053.
- Stone, C. (1985). Additive regression and other nonparametric models. *Ann. Statist.*, **13**, 689-705.

Problèmes de données surdispersées pour les modèles de comptage

Célestin C. KOKONENDJI

* Adresse pour correspondance :
Université de Pau et des Pays de l'Adour
Laboratoire de Mathématiques Appliquée
CNRS FRE 2570 - Département STID
celestin.kokonendji@univ-pau.fr

Séance du 27 Octobre 2003

Résumé

Les données de comptage sont souvent surdispersées comparées à la distribution de Poisson. Dans ce cas, il est d'abord et presque toujours fait appel à la loi binomiale négative laquelle est un mélange de Poisson par la loi gamma ([3], [4]). Nous illustrerons sur quelques jeux des données émanant de portefeuille automobile (sur le nombre de sinistres occasionnés par un véhicule pendant une année) qu'aucune loi de probabilité, comparable et surdispersée par rapport à Poisson, ne paraît émerger comme fournissant uniformément le meilleur ajustement ([3], [5]).

Dans le cadre des modèles de régression pour les données de comptage ([7]), la régression de Poisson est cependant remplacée par un modèle comparable et de fonction variance de la forme ([1]) suivante

$$m \mapsto m + \phi m^p, \quad (5)$$

où $\phi > 0$ est un paramètre de dispersion et $p \in \mathbb{R}$ fixé. Pour cet exposé, il sera question de caractériser ([6]) les lois de probabilité des familles exponentielles associées à (5) et de discuter, sur un exemple de [2], le choix de p dans (5).

Références

- [1] Hinde J., Demétrio C.G.B., (1998). *Overdispersion : Models and Estimation*, ABE, Sao Paulo.
- [2] Hougaard P., Lee M-L.T., Whitmore G.A., (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes, *Biometrics*, vol. **53**, 1225-1238.
- [3] Jourdan A., Kokonendji C.C., (2002). Surdispersion et modèle binomial négatif généralisé, *Revue de statistique appliquée*, vol. **L** (3), 73-86.
- [4] Kokonendji C.C., (1999). Le problème d'Anscombe pour les lois binomiales négatives généralisées, *The Canadian Journal of Statistics*, vol. **27** (1), 187-193.
- [5] Kokonendji C.C., Khoudar M., (2003). On strict arcsine distribution, *In revision*.
- [6] Kokonendji C.C., Demétrio C.G.B., Dossou-Gbété S., (2003). Overdispersion and Poisson-Tweedie exponential dispersion models, *To submit*.
- [7] Marque S., Kokonendji C.C., (2003). A strict arcsine regression model, *Submitted*.

Quelques problèmes statistiques en ingénierie du trafic routier

Nour-Eddin EL FAOUZI

* Adresse pour correspondance :
Laboratoire d'ingénierie Circulation Transports
INRETS-ENTPE
25, Ave. F. Mitterrand, 69675 Bron Cedex
e-mail : elfaouzi@inrets.fr

Séance du 17 Novembre 2003

Résumé

La modélisation dynamique et l'exploitation des réseaux de routes fait partie des problèmes relatifs à l'ingénierie du trafic. Il s'agit d'une part, de comprendre et de modéliser la physique du trafic, de la reproduire au travers des outils de simulation et d'autre part, l'élaboration et la mise en œuvre d'outils et d'indicateurs de circulation permettant la caractérisation des situations de trafic et l'activation des actions de gestion et de régulation du trafic.

Dans cette présentation, on se propose d'abord de décrire brièvement les principales méthodes de modélisation des réseaux. En particulier, les modèles d'écoulement et de l'affectation du trafic.

Une seconde partie de l'exposé sera consacrée à la formulation et à la résolution statistique de quelques problèmes liés à la prévision court terme du trafic, à l'estimation de matrices origines - destinations et à l'estimation des temps de parcours dans laquelle les aspects fusion de données et de modèles seront abordés.

Références

Barnett J. A. (1981). Computational Methods for a Mathematical Theory of Evidence., *Proceedings of the 7th Joint Conference on A.I.*, Vancouver, British

- Columbia, Canada, Vol 2, pp. 868-875.
- Bates M. and W. J. Granger (1969) : The combination of Forecasts. *Operational Research Quarterly*, 20, pp. 451-468.
- Bogler P. (1987). Shafer-Dempster Reasoning with Applications to Multisensor Target Identification Systems. *IEEE Transactions on Systems, man and Cybernetics*, 17(6), pp. 968-977.
- Caselton W., W. Luo (1992). Decision Making With Imprecise Probabilities : D-S. Theory and Application., *Water Resources Research*, 28(12), pp. 3071-3083.
- Dempster A.P. (1967). Upper and Lower Probabilities Induced by Multivalued Mapping., *Annals of Mathematical Statistics*, Vol. 38, p. 325-339.
- Dempster A. P. (1968). A Generalization of Bayesian Inference., *Journal of the Royal Statistical Society*, 30B, p. 205-247, 1968.
- El Faouzi N.-E. (2003). Recueil multiforme et fusion de données en circulation routière. Actes journées spécialisée du 24-10-2002, Collection INRETS.
- El Faouzi N.-E. (2002) : Evidential Data Fusion for Short-Term Travel time Prediction : A discrete Case. *PrePrint INRETS-LICIT*, 2002.
- El Faouzi N.-E., J.-B. Lesort (2001). Fusion de données pour l'estimation du temps de parcours : Réflexions théoriques et méthodologiques., Rapport INRETS N° 0101, Novembre 2001.
- El Faouzi N.-E. (1999) Combining predictive schemes in short-term traffic forecasting. *Transportation and Traffic Theory - Proceedings of the 14th International Symposium on Transportation and Traffic Theory*, Ceder, A. ed., Pergamon, Amsterdam, p. 471-487.
- El Faouzi. (1997) Heterogenous data sources fusion for impedance indicators. *The 8th IFAC symposium on transportation systems*, Chania, Greece.
- May A. D. (1990) *Traffic Flow Fundamentals*, Prentice Hall, Englewood Cliffs, N-J.
- Sherali H., A. Narayanan, R. Sivanandan (2003) Estimation of origin-destination trip tables based on partial set of traffic link volumes. *Transportation Research*, B, pp. 815-836.
- Smith B., B. Williams, K. Oswald (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C.*, pp. 303-321.
- Okutani I., Stephanedes Y. (1984) Dynamic Prediction of Traffic Volume through Kalman Filtering Theory. *Transportation Research B.*, pp. 1-11.

Une approche par réseau de neurones pour la prédiction du niveau de l'antigène spécifique de prostate

Cornelia ENACHESCU

* Adresse pour correspondance :

Institut de Statistique Mathématique et Mathématique Appliquée
Casa Academiei, calea 13 Septembrie, 13
Bucarest, Roumanie
e-mail : cenachescu@k.ro

Séance du 24 Novembre 2003

Résumé

Dans cette présentation on se propose de prédire l'antigène spécifique de la prostate à partir de mesures cliniques. Les données viennent d'une étude de Stamey et al (1989).

Dans notre étude nous proposons comme modèle prédictif un réseau de perceptrons à deux couches (MLP) qui est un modèle statistique non linéaire particulier. Le modèle MLP a des paramètres inconnus, appelés poids, et on cherche les valeurs pour ces poids qui ajustent au mieux les données d'apprentissage. On utilise comme mesure d'ajustement la moyenne des carrés des erreurs (MSE) et comme algorithme d'apprentissage celui de Levenberg-Marquard qui est connu pour avoir (pour des réseaux avec une architecture modérée) la convergence la plus rapide.

Il y a tout un art dans l'apprentissage des réseaux de neurones. En général le réseau a mémorisé les exemples, mais il n'a pas appris à généraliser les situations nouvelles. Afin d'améliorer le pouvoir de généralisation du MLP, on utilise la technique de régularisation et la normalisation des données d'entrée. Afin de réduire le nombre de variables d'entrée et par conséquent le nombre de mesures cliniques, il est nécessaire de choisir le meilleur sous-ensemble de variables.

En comparant avec un certain nombre de modèles de régression alternatifs (régression standard, ridge régression, lasso régression, régression en composantes principales et régression PLS), on montre que le réseau MLP produit une erreur de prédiction environ fois inférieure à celle de ces autres modèles.

Références

- DEMUTH, H. AND BEALE, M., (2000) *Neural network toolbox*. The Mathworks Inc., Natick, USA.
- FORESEE, F. AND HAGAN, M., (1997) *gauss-newton approximation to Bayesian regularization*, Proc. of Intern. Joint Conference on Neural networks, 1930-55.
- NGUYEN, D. AND WIDROW, D., (1990) *Improving the learning speed of 2layer neural networks by choosing initial values of the adaptive weights*, Proc. of Intern. Joint Conference on Neural networks, Vol. 3, 21-26.
- HASTIE, T., TIBSHIRANI, R. AND FRIEDMAN, J., (2002) *The elements of statistical learning. data mining, inference and prediction*, Springer Verlag, New-York.
- STAMEY, T., KABALIN, J. MCNEAL, J. JOHNSTONE, I., FREIHA, F., REDWINE, E. AND YANG, N., (1989). Prostate specific antigen in the diagnosis of adenocarcinoma of the prostate II. radical prostatectomy treated patients. *Journal of urology*, **16**, 1076-1083.

Récents développements sur la comparaison de plusieurs populations par analyses factorielles

Sylvie VIGUIER-PLA

* Adresse pour correspondance :
Laboratoire de Statistique et Probabilités,
Université Paul Sabatier, 31062 Toulouse Cedex 4
e-mail : viguiер@cict.fr

Séance du 15 Décembre 2003

Résumé

On appelle analyse factorielle (A.F.) une méthode statistique multivariée qui peut être mise en œuvre par l'analyse spectrale d'un opérateur auto-adjoint, ou plus communément la diagonalisation d'une matrice symétrique. Ainsi, on peut étudier la structure des liens entre variables dans un échantillon issu d'une population homogène.

Si on considère plusieurs échantillons pour lesquels on possède les mêmes mesures, on peut étudier les liens entre variables par une A.F. pour chaque échantillon. Peut alors venir le problème de comparaison de ces populations au moyen de ces A.F. opérées sur des échantillons issus de ces populations.

Lorsque l'A.F. est une analyse en composantes principales (A.C.P.), le problème revient à celui de la comparaison des matrices de covariance ou de corrélation. Dans le cas normal, des tests de comparaison de matrices de covariances ont été proposés par plusieurs auteurs. On peut citer

- le test d'égalité des matrices de covariance (Muirhead, 1982, Anderson, 1984),
- le test de proportionnalité entre matrices de covariance (Jensen and Johansen, 1987, Eriksen, 1987, Flury, 1986),
- les tests de diagonalisation simultanée totale ou partielle, c'est-à-dire d'égalité entre des ensembles de vecteurs propres et le test d'égalité de sous-espaces engendrés par les vecteurs propres (Flury, 1987).

D'autres auteurs ont étudié plus particulièrement le cas de deux échantillons, on peut citer alors

- le test d'égalité de deux sous-espaces associés aux q plus grandes valeurs propres, appelés sous-espaces factoriels (Schott, 1988),
- le test d'existence d'un sous-espace commun à deux sous-espaces factoriels (Chen and Robinson, 1989).

Ce dernier test a été repris par Dauxois, Romain and Viguié (1993), qui l'ont adapté à toute A.F., sous condition d'existence des moments jusqu'à l'ordre 4 des variables aléatoires (v.a.) sous-jacentes, et élargi au cas où ces v.a. ne sont pas forcément de loi normale.

Les tests pré-cités ont été rassemblés, avec d'autres traitements tels que le test de multi-normalité et l'analyse discriminante, dans un logiciel (Romain and Viguié, 1989), et des exemples sont traités dans Viguié-Pla (1994).

On considère dans cet exposé le cas d'un nombre quelconque fini k d'échantillons. Le problème sera alors de tester l'hypothèse :

H_0 : il existe un sous-espace de dimension q commun à k sous-espaces factoriels.

La résolution se fait en utilisant une généralisation de l'analyse canonique de deux à k espaces. On propose une statistique de test, valable quelque soit l'A.F. considérée, pour des v.a. à valeurs dans un espace de Hilbert H de dimension non nécessairement finie.

L'explicitation complète de la loi de la statistique est faite dans le cas où H est de dimension finie, et dans le cas où la covariance asymptotique de l'opérateur dont on fait l'analyse spectrale est connue.

Un exemple tiré de la littérature est repris. Il s'agit de la comparaison de matrices de covariance. Les résultats des différents tests précédemment cités sont analysés. Cette analyse se veut être la proposition d'une démarche cohérente d'utilisation combinée des résultats de tous ces tests, pour la comparaison de plusieurs échantillons.

Les programmes pour réaliser ces tests ont été écrits en langage GAUSS. Les derniers tests sont aussi programmés dans ce logiciel, pour le cas de la comparaison de matrices de covariance ou de corrélation. Il sont, avec quelques fichiers de données, disponibles, sur la page internet :

<http://www.lsp.ups-tlse.fr/Viguié/prepublications.html>

Parmi les très nombreux travaux existant dans la littérature et consacrés au problème de comparaison de populations, on peut citer ceux de Krzanowski (1984), Meredith (1964), très proches de nos préoccupations. Des travaux plus récents (Neuenschwander and Flury, 1995, 2000, Boik, 2003) sont le reflet du fait que les méthodes actuelles ne répondent pas correctement aux problèmes du praticien. C'est le cas lorsqu'il s'agit de comparer d'autres matrices que celles de covariance. C'est aussi le cas lorsque les échantillons ne sont pas issus de v.a. indépendantes. Ces interrogations sont autant de pistes pour des travaux à venir.

Références

- ANDERSON, T.W. (1984) *An Introduction to Multivariate Statistical Analysis*. 2nd edition, Wiley, New York.
- ARCONTE, A. (1980) Etude asymptotique de l'analyse en composantes principales et de l'analyse canonique. *Thèse de 3^e cycle, Mathématiques Appliquées, Université de Pau et des Pays de l'Adour*.
- BOIK, R. J. (2003). Principal component models for correlation matrices. *Biometrika* **90** 679-701
- CHEN, K. H. AND ROBINSON, J. (1989). Comparison of factor spaces of two related populations. *J. Multivariate Anal.* **8** 190-203.
- DAUXOIS, J., ET POUSSE, A. (1975). Une extension de l'analyse canonique. Quelques applications. *Ann. Inst. Henri Poincaré* **XI** 355-379.
- DAUXOIS, J., ROMAIN, Y. AND VIGUIER, S. (1993) Comparison of two factor subspaces. *J. Multivariate Anal.* **44** 160-178.
- DAUXOIS, J., ROMAIN, Y. AND VIGUIER-PLA, S. (1994) *Tensor products and Statistics Linear Alg. and its Appl. - Special Issue on Linear Algebra and Statistics* **210** 59-88.
- ERIKSEN, P.S. (1987) Proportionalty of covariance matrices. *Ann. Stat.* **15** 732-748
- FLURY, B.N. (1986) Proportionality of k covariance matrices. *Stat. and Prob. Letters* **4** 29-33.
- FLURY, B.N. (1987) Two generalizations of the common principal component model. *Biometrika* **74** 59-69.
- FLURY, B. N. (1988). *Common Principal Components and Related Multivariate Models*. Wiley, New York.
- JENSEN, S. T. AND JOHANSEN, S. (1988) Estimation of proportional covariances. *Stat. and Prob. Letters* **6** 83-85.
- W.J. KRZANOWSKI, W.J. (1984) Principal component analysis in the presence of group structure. *Appl. Statist.* **33** 164-168
- MEREDITH, W. (1964) Notes on factorial invariance. *Psychometrika* **29** 177-185
- MUIRHEAD, R.J. (1982) *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- NEUENSCHWANDER, B. E. AND FLURY, B.D. (1995). Common Canonical Variates. *Biometrika* **90** 553-560.
- NEUENSCHWANDER, B. E. AND FLURY, B.D. (2000). Common Principal Components for Dependent Random Vectors. *J. Multivariate Anal.* **75** 163-183.
- ROMAIN, Y. ET VIGUIER, S. (1989) COMPCOV : un outil de comparaison de plusieurs matrices de covariance. *Stat. et Anal. des Données* **14** 37-52.
- SCHOTT, J.R. (1988) Common principal components subspaces in two groups. *Biometrika* **75** 229-236.
- VIGUIER-PLA, S. (1994) Factor-based comparison of several populations using COMPAR software. *Computational Statistics* **9** 135-163.

VIGUIER-PLA, S. (2004) Factor-based comparison of k populations. *Statistics*, to appear.

M- Quantiles et profondeurs : le cas unidimensionnel

Jean AVEROUS

* Adresse pour correspondance :

Laboratoire de Statistique et Probabilités, Université Paul Sabatier
118 Route de Narbonne, 31062 Toulouse Cedex, France
e-mail : averous@cict.fr

Séance du 19 Janvier 2004

Résumé

Ces dix dernières années, la statistique descriptive multidimensionnelle s'est beaucoup développée autour des deux concepts quantiles multidimensionnels et profondeurs qui permettent d'étendre aux vecteurs aléatoires les résultats sur la localisation, la dispersion, la dissymétrie ou le poids des queues, fondés, pour les lois réelles, sur les quantiles usuels.

Dans un premier exposé, on s'intéressera au cas unidimensionnel pour dégager les concepts à la base des diverses définitions de profondeur ou de quantiles généralisés qui ont été proposées dans ce cadre. Les liens entre quantiles et profondeurs y sont plus faciles à mettre en évidence, et ces liens serviront de guide pour l'étude des extensions au cas multidimensionnel qui sera présenté ultérieurement.

Références

- R.Y. Liu, J.M. Parelus and K. Singh. Multivariate analysis by data depth : Descriptive statistics, graphics and inference(with discussion). *Annals of Statistics* 27(1999) 783-858.
- K. Mosler. Multivariate Dispersion, Central Regions and Depth. The Lift Zonoid Approach. *Lecture Notes in Statistics* 165, Springer (2002).
- Y. Zuo and R. Serfling. General notions of statistical depth function. *Annals of Statistics* 28(2000) 461-482.

Identification de points non centraux par régression expectile et application à la mesure d'efficacité d'une activité médicale

Sandrine POIRAUD-CASANOVA

* Adresses pour correspondance :

G.R.E.M.A.Q., Université des Sciences Sociales,
21, allée de Brienne, 31000 TOULOUSE, France

et

L.S.P., Université Paul Sabatier,
118, Route de Narbonne, 31062 TOULOUSE cédex 4, France
e.mail : spoiraud@univ-tlse1.fr

Séance du 26 Janvier 2004

Résumé

En régression, en alternative aux quantiles, les expectiles (Newey et Powell, 1987, Breckling et Chambers, 1988) conditionnels de Y sachant $X = x$ sont une mesure de localisation pertinente lorsqu'on s'intéresse au comportement des individus loin de la moyenne. Etant donné un point (y, x) , son ordre expectile pour la distribution $F(.|X = x)$ peut être utilisé comme mesure de performance de ce point (Kokik *et al.*, 1997).

Nous proposons d'estimer les expectiles conditionnels par les méthodes de la constante locale et des polynômes locaux (degré 1) (Yao et Tong, 1996) pour une grille de valeurs d'ordres et nous en déduisons l'ordre expectile estimé par interpolation. L'estimation par polynôme local ne donnant pas nécessairement un estimateur monotone par rapport à q , nous envisageons différentes techniques de monotonisation. Alternativement, nous proposons une technique d'estimation directe de l'ordre expectile utilisant des estimateurs à noyau de type Nadaraya-Watson calibrés (Hall *et al.*, 1999). Les propriétés à distance finie de nos méthodes sont alors comparées par simulation.

Enfin, nous appliquons nos méthodes à une population de médecins exerçant une activité libérale dans la région Midi-Pyrénées durant l'année 1999. On connaît

pour chacun un certain nombre de caractéristiques, notamment : le montant de prescription pharmaceutique, considéré comme variable d'activité, la répartition de la clientèle par tranches d'âge, le sexe, l'âge du médecin et l'ancienneté de son installation. On veut mesurer l'efficacité de l'activité de chaque médecin par rapport à l'ensemble des médecins considérés. Nous proposons d'utiliser comme mesure, l'ordre de l'expectile de régression de l'activité sur les autres caractéristiques, passant par le point représentant le médecin dans l'espace des variables. Nous introduisons une mesure de qualité d'ajustement inspirée du coefficient de détermination de la régression linéaire. Le nombre élevé de variables rend nécessaire une réduction préalable de dimension. Celle-ci est menée par SIR. La régression expectile est ensuite faite sur les indices obtenus par SIR, d'où l'on déduit la mesure d'efficacité de chaque observation.

Références

- Breckling, J., Chambers, R. (1988) M-quantiles. *Biometrika*, Vol. 75, 761-771.
- Hall, P., Wolff R.C.L., Yao Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, Vol. 94, 154-163.
- Kokic, Ph., Chambers, R., Breckling, J., Beare, S., (1997) A Measure of Production Performance. *Journal of Business Economic Statistics*, Vol. 15, 445-451.
- Newey, W.K., Powell, J.K. (1987) Asymmetric least squares estimation and testing. *Econometrica*, Vol. 46, 33-50.
- Yao, Q., Tong, H. (1996) Asymmetric Least Squares Regression Estimation : A Nonparametric Approach. *Nonparametric Statistics*, Vol. 6, 273-292.

Analyse non-paramétrique des frontières de production et des mesures d'efficacité à l'aide de quantiles conditionnels non-standards

Abdelaati DAOUIA

* Adresse pour correspondance :
UFR MIG - BAT 1R2 - Bureau 147 - 1er étage
118 route de Narbonne - 31062 Toulouse
e-mail : Abdelaati.Daouia@math.ups-tlse.fr

Séance du 23 Février 2004

Résumé

Cet exposé présente une approche non paramétrique pour estimer la frontière du support d'une loi multivariée. Les applications visées concernent l'analyse de l'efficacité de production. La problématique sera expliquée dans la première section. En se basant sur les quantiles conditionnels la seconde section introduit un concept de frontière quantile d'ordre α permettant la mise en œuvre d'estimateurs non paramétriques robustes de la vraie frontière, et étudie les propriétés de cette frontière partielle ainsi que de son estimateur non paramétrique. La section 3 explicite de façon formelle les propriétés de robustesse de cet estimateur, alors que la section 4 propose une analyse plus fine de ses propriétés asymptotiques. La section 5 montre que l'ordre α définit une mesure d'efficacité acceptable et la dernière section propose une version monotone et plus robuste de l'estimateur quantile qui garde les mêmes propriétés de convergence.

Statistical Modeling in Survival Analysis and Reliability : Recent results

Mikhail S. NIKULIN

* Adresse pour correspondance :
EA 2961, Statistique Mathématique
University of Bordeaux-2, 33076, Bordeaux, France
e-mail : M.S.Nikouline@u-bordeaux2.fr

Séance du 1 Mars 2004

Keywords : Accelerated life testing ; GPH model ; Cox model ; Frailty model ; Partial likelihood ; Resource ; Survival analysis ; Transfer functional ; Transformation models.

Abstract

Suppose that $\alpha_{x(\cdot)}(t)$ is the *hazard rate* and $S_{x(\cdot)}(t)$ is the survival function of the time-to-failure $T_{x(\cdot)}$ given the time-varying covariate vector $x(\cdot) = (x_1(\cdot), \dots, x_k(\cdot))^T$, a vector-valued function on $[0, +\infty)$. In reliability theory, $x(\cdot)$ is known as a *stress*. We suppose that $x(\cdot)$ belongs to a set of all possible stresses \mathcal{E} , which will be called the *covariates space*.

The *proportional hazards* (PH) or *Cox* model is given by specifying

$$\alpha_{x(\cdot)}(t) = r\{x(t)\}\alpha_0(t),$$

where $\alpha_0(t)$ is the *baseline hazard rate* which does not depend on $x(\cdot)$, and r is a positive-valued function (e.g. $\exp(\beta^T x(t))$, where β is a vector of parameters). If the PH model is true, then for any two time-varying covariates, $x(\cdot)$ and $x_0(\cdot)$,

$$\frac{\alpha_{x(\cdot)}(t)}{\alpha_{x_0(\cdot)}(t)} = \frac{r\{x(t)\}}{r\{x_0(t)\}}. \quad (1.1)$$

In reliability theory, $x_0(\cdot)$ often denotes the *standard* stress, the normal condition a testing item is in. Let

$$A_{x(\cdot)}(t) = \int_0^t \alpha_{x(\cdot)}(\tau) d\tau = -\log S_{x(\cdot)}(t)$$

denote the *cumulative hazard function*. Then random variable $R = A_{x(\cdot)}(T_{x(\cdot)})$ has the standard exponential distribution with the survival function

$$S(t) = \Pr\{R > t\} = e^{-t}I_{\{t \geq 0\}},$$

which does not depend on $x(\cdot)$. Note that, for all $x(\cdot)$,

$$\Pr\{T_{x(\cdot)} > t\} = \Pr\{R > A_{x(\cdot)}(t)\}.$$

We call $A_{x(\cdot)}(t)$ the resource used up to time t given stress $x(\cdot)$ and R the *exponential resource*.

The PH model (1.1) assumes that the ratio of rates of resource usage for any two individuals at time t depends only on the values of their respective covariates at time t , not on the resources used up to time t . Thus, if x_0 and x_1 are two constant (in time) stresses,

$$x_0(t) \equiv x_0 = \text{const}, \quad x_1(t) \equiv x_1 = \text{const},$$

and

$$x(t) = \begin{cases} x_1, & 0 \leq t < t_1, \\ x_0, & t \geq t_1, \end{cases}$$

then, for all $t \geq t_1$,

$$\alpha_{x(\cdot)}(t) = \alpha_{x_0}(t).$$

A stress $x(\cdot)$ is said to be *accelerated* with respect to another stress $y(\cdot)$, if

$$S_{x(\cdot)}(t) \leq S_{y(\cdot)}(t), t \geq 0.$$

If x_1 is some accelerated stress with respect to the normal stress x_0 , then the resource $A_{x(\cdot)}(t_1)$ used until the time t_1 under the stress $x(\cdot)$ is larger than the resource used until t_1 under the stress x_0 . Nevertheless, the proportional hazards model entails that the rate of resource usage remains the same after time t_1 . This is not very natural when there is an aging effect—hazard rate of an individual who used up more resource before t_1 will be higher than that of one used up less, given that they are subject to the same stress x_0 at t_1 . To accommodate this, we need to generalize the model so that possible dependence of the rate of resource usage at time t on the amount of the resource used up to t can be incorporated.

M- Quantiles et profondeurs : le cas multidimensionnel

Jean AVEROUS

* Adresse pour correspondance :

Laboratoire de Statistique et Probabilités, Université Paul Sabatier

118 Route de Narbonne, 31062 Toulouse Cedex, France

e-mail : averous@cict.fr

Séance du 8 Mars 2004

Résumé

Les quantiles multidimensionnels, profondeurs, et domaines centraux introduits indépendamment ces dernières années sont en fait des concepts étroitement liés. Après avoir présenté une demi-douzaine de ces concepts on montrera comment expliciter ces liens et obtenir ainsi des outils pour une description cohérente des lois multidimensionnelles.

Nonparametric analysis of replicated microarray experiments

Ali GANNOUN (1,2,3)*,
Jérôme SARACCO (1) , Wolfgang URFER (4) et George E. BONNEY (2,3)

(1) Labo de Probabilités et Statistique, Montpellier II, France

(2) Department of Microbiology, Howard University, USA

(3) Statistical Genetics and Bioinformatics Unit, Howard University, USA

(4) Department of Statistics, University of Dortmund, Germany

* Adresse pour correspondance :

Laboratoire de Probabilités et Statistique, Université Montpellier II, France

e-mail : gannoun@darboux.math.univ-montp2.fr

Séance du 15 Mars 2004

Abstract : Microarrays are part of a new class of biotechnologies which allow the monitoring of expression levels of thousands of genes simultaneously. In microarray data analysis, the comparison of gene-expression profiles with respect to different conditions and the selection of biologically interesting genes are crucial tasks. Multivariate statistical methods have been applied to analyze these large data sets. In particular, Dudoit et al. (2002) developed methods using t-statistics with p -values calculated through permutations, and with the Westfall and Young (1993) step-down approach to correct for multiple testing. Thomas et al. (2001) developed a regression modelling approach. Following the idea of Efron et al. (2000) and Tusher et al. (2001), Pan (2002) proposed a Normal mixture modelling approach that relaxes many strong assumptions on the null distributions of the test statistics.

This paper makes two contributions to the analysis of microarray data. The first is the introduction of a new method for the calculation of the cut-off point and the acceptance region, and the second is the replacement of the based Normal mixture density estimators proposed by Pan et al. (2002), with less restrictive kernel nonparametric ones. A useful modification is suggested in order to increase the performance of the kernel estimator on the tail of the distribution. We apply our approach to leukaemia data of Golub et al. (1999) and compare our results

to those of Pan (2002).

Références

- Dudoit, S., Yang, Y.H., Speed, T.P., Callow, M.J. (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica*, **12**, 111-139.
- Efron, B., Tibshirani, R., Goss, V., Chu, G. (2000) Microarrays and their use in a comparative experiment. Technical report, Stanford University.
- Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S. (1999) Molecular classification of cancer : class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531-537.
- Pan, W. (2002) A Comparative Review of Statistical Methods for Discovering Differentially Expressed Genes in Replicated Microarray Experiments. *Bioinformatics*, **12**, 546-554.
- Pan, W., Lin, J., Le, C.T. (2002) A Mixture Model Approach to Detecting Differentially Expressed Genes with Microarray Data. *Genome Biology* (To appear).
- Thomas, J.G., Olson, J.M., Tapscott, S.J., Zhao, L.P. (2001) An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Research*, **11**, 1227-1236.
- Tusher, V.G., Tibshirani, R., Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5116-5121.
- Westfall, P.H., Young, S.S. (1993) *Resampling-based multiple testing : examples and methods for p-value adjustment*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. New York, NY : Wiley.

20 ans de recherche : un bilan

Roger LAFOSSE

* Adresse pour correspondance :
Laboratoire de Statistique et Probabilités,
Université Paul Sabatier, 31062 Toulouse Cedex 4
e-mail : lafosse@cict.fr

Séance du 29 Mars 2004

Résumé

Quand j'ai commencé (Oct. 1981), il s'agissait de définir des mesures de ressemblance et de différence entre deux matrices $n \times p$ et $n \times q$ mesurées sur les mêmes individus, pensant qu'une étude comparative de deux objets était une bonne façon d'enrichir l'étude d'un seul. Cette recherche s'est alimentée d'allers et retours entre l'étude de 3 objets et l'étude de 2 objets. Elle a finalement conduit (avec l'aide d'Hanafi M.)

- à considérer que l'emploi de métriques quelconques, et notamment celle de Mahalanobis, est peu utile en analyse des données, la métrique identité paraissant presque toujours suffisante.

- à définir un concept de composantes monogames entre 2 matrices et de composantes polygames pour plus de 2 matrices, l'endogamie étant réservée à une matrice (ACP).

- à étendre la notion de SVD d'une matrice à celle de SVD entre plusieurs matrices, plus exactement à définir la SVD d'une matrice partitionnée selon les colonnes.

- les applications se trouvent notamment en analyse factorielle, en régression multivariée et en reconnaissance de formes.

- On s'est intéressé aussi à définir des biplots juxtaposés, pouvant servir par exemple à décrire la relation entre présent et passé pour un processus multivarié stationnaire.

Le saut suivant s'est produit en 2003, (grâce à Ten Berge, Cazes, Kissita, et toujours Hanafi).

- La polygamie est étendue à une notion de partouze entre composantes liant K matrices à K' matrices.
- On définit la SVD d'une matrice partitionnée selon les lignes et les colonnes, permettant en particulier une analyse de liens entre cubes de données.
- les applications se trouvent toujours notamment en analyse factorielle, en régression multivariée et en reconnaissance de formes.
- les outils graphiques et autre, comme un covariogramme croisé de 2 processus multivariés, deviennent d'une mise en oeuvre complexe.

Finalement, ce travail se réduit à peu de chose sur le plan numérique : seulement cette histoire de SVD pour matrice partitionnée. Cependant, la SVD étant utilisée dans bien des calculs, on peut espérer que le peu soit souvent utilisé.

Références

- Kissita G., Cazes P., Hanafi M., Lafosse R. (2004) *Deux méthodes d'analyse factorielle du lien entre deux tableaux de variables partitionnés*. à paraître dans Rev. Stat. Appliquée.
- Lafosse R. (1997) *Analyse de concordance de deux tableaux : monogamies, simultanités et découpages*. Rev. Stat. Appliquée, 45, 3, 45-72.
- Lafosse R. et Hanafi M. (1997) *Concordance d'un tableau avec K tableaux : définition de $K+1$ uplés synthétiques*. Rev. Stat. Appliquée, 45, 4, 111-126.

Vitesse optimale du cas i.i.d. pour l'estimation non-paramétrique de la densité invariante d'un système dynamique chaotique

Salim LARDJANE

* Adresse pour correspondance :

Université de Bretagne Sud & CREST-ENSAI
CREST-ENSAI, Laboratoire de Statistique et Modélisation
Rue Blaise Pascal, 35170 Bruz, France
e-mail : lardjane@ensai.fr

Séance du 26 Avril 2004

Résumé

Les systèmes dynamiques déterministes sont souvent utilisés pour modéliser des phénomènes évolutifs en écologie, en physique et en économie [6,7,8].

Je m'intéresse ici à l'estimation non-paramétrique de la densité invariante d'un système dynamique appartenant à une classe de transformations chaotiques. Cette densité décrit le comportement à long terme du système ou encore un état d'équilibre statistique de celui-ci [1,11].

Les systèmes considérés sont unidimensionnels et évoluent en temps discret ; leurs orbites $(x_t)_{t \in \mathbb{N}}$ sont définies par des équations de la forme $x_{t+1} = S(x_t)$ où x_0 est l'état initial du système et où S est sa loi d'évolution, également appelée *transformation itérée* [5].

Je commence par démontrer diverses inégalités de covariance [4]. Je démontre ensuite que le problème est *équivalent* à celui de l'estimation de la densité marginale d'un processus stochastique stationnaire $(X_t)_{t \in \mathbb{N}}$ tel que $X_{t+1} = S(X_t)$ où X_0 est une variable aléatoire fixée [4]. De tels processus sont appelés *processus dynamiques* et sont hautement dépendants. Il ne peuvent, en particulier, être supposés mélangeants [2,3].

Dans un premier temps, j'utilise cette équivalence pour étudier les propriétés de l'estimateur de Parzen-Rosenblatt de la densité marginale d'un processus dynamique $(X_t)_{t \in \mathbb{N}}$ basé sur l'observation de X_0, X_1, \dots, X_{n-1} . Celui-ci est défini,

en tout point x , par

$$f_n(x) = \frac{1}{nh_n} \sum_{t=0}^{n-1} K\left(\frac{x - X_t}{h_n}\right),$$

où K est le noyau uniforme $\chi_{[-1/2,1/2]}$ et où $(h_n)_{n \in \mathbb{N}}$ est une suite de fenêtres décroissant vers 0 [9,10].

Je démontre que l'estimateur peut, *dans certains cas*, converger en moyenne quadratique *avec la vitesse optimale du cas i.i.d.*, pour les même suites de fenêtres que dans le cas i.i.d [4].

Dans un deuxième temps, j'applique ce résultat aux systèmes dynamiques chaotiques considérés, et donne une interprétation pratique de l'erreur quadratique moyenne dans ce contexte [4].

Références

- 1 G. D. Birkhoff, G. D. (1931). Proof of the ergodic theorem. . *Proc. Nat. Acad. Sci. USA*, **17**, 656-660.
- 2 Bosq, D. (1995). Optimal asymptotic quadratic error of density estimators for strong mixing or chaotic data. *Statistics and Probability Letters*, **22**, 339-347.
- 3 Bosq, D. (1998). *Nonparametric statistics for stochastic processes : estimation and Prediction*. Lecture Notes in Statistics, **110**. Springer.
- 4 Lardjane, S. (2002). Optimal speed nonparametric density estimation for one-dimensional dynamical systems. *Série des document de travail du CREST, INSEE*, **16**. Disponible sur www.crest.fr.
- 5 Lasota, A. and Mackey, M. (1994). *Chaos, Fractals and Noise : Stochastic Aspects of Dynamics*. Springer-Verlag.
- 6 May, R. (1976). Simple mathematical models with very complicated dynamics. *Nature*, **261**, 459-467.
- 7 Medio, A. (1992). *Chaotic Dynamics : Theory and applications to Economics*. Cambridge.
- 8 Neimark, Y. and Landa, P. (1992). *Stochastic and Chaotic Oscillations*. Kluwer.
- 9 Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Stat.*, **33**, 1065-1076.
- 10 Rosenblatt, M. (1956). Remarks on some non parametric estimates of a density function. *Ann. Math. Stat.*, **27**, 832-837.
- 11 Ruelle, D. (1989). *Chaotic evolution and strange attractors*. Cambridge.

Modèles de rang réduit pour l'analyse des données

Antoine DE FALGUEROLLES

* Adresse pour correspondance :
LSP, Université Paul Sabatier (Toulouse III)
e-mail : falguero@cict.fr

Séance du 7 Juin 2004

Résumé

L'analyse exploratoire des données multidimensionnelles repose souvent sur une utilisation fine de la décomposition en valeurs singulières, simple ou généralisée, de matrices. Ces dernières sont construites à partir des données avec une grande ingéniosité.

L'objectif est de (tenter de) mettre en évidence certaines formes structurelles présentes dans les données. Les propriétés de la décomposition en valeurs singulière de ces matrices servent un double objectif : obtention d'approximations de rang réduit et construction de représentations graphiques variées, des "biplots" en particulier.

Un exemple simple est fourni par l'analyse des correspondances d'un tableau d'effectifs à deux entrées (obtenu par croisement de deux facteurs de classification) ; la décomposition en valeurs singulières généralisée de diverses matrices construites à partir de ce tableau fournit des approximations de rang réduit du tableau initial et des représentations graphiques variées. L'analyse des correspondances multiples d'un tableau de Burt est plus énigmatique. Les données de base constituent un tableau à plusieurs dimensions dont les cellules sont associés au croisement de tous les facteurs de classification. La matrice construite à partir du tableau de données, le tableau de Burt, est obtenue par concaténation des tableaux d'effectifs conjoints résultant du croisements deux à deux de tous les facteurs de classification. La décomposition en valeurs singulières généralisée de cette matrice assez particulière, fournit des approximations des marges d'ordre deux du tableau de données et des représentations graphiques associées.

Dans certains cas, il est assez facile de présenter ces méthodes exploratoires comme des procédures d'ajustement de modèles linéaires généralisés dont les prédicteurs comprennent un terme d'interaction de rang réduit entre deux fac-

teurs. Ainsi, si l'on suppose que les effectifs conjoints d'un tableau obtenu par croisement de deux facteurs sont des observations de variables aléatoires normales indépendantes, de variances structurées d'une certaine façon, et de moyennes définies, au travers d'une certaine fonction de lien non canonique, par une interaction de rang réduit, on retrouve exactement l'analyse des correspondances. Ainsi dénaturée ou accommodée, l'analyse des correspondances devient une méthode essentiellement univariée. Cela peut paraître réducteur. En contrepartie, il y a de nombreuses ouvertures possibles : autres lois (Poisson, binomiale, binomiale négative, Gamma, inverse-gaussienne ...), autres fonctions liens (identité, logarithme, racine carrée ...), prédicteurs incorporant les effets d'autres variables explicatives, données structurellement absentes...

Mais à ce jour, l'analyse des correspondances multiples semble ne pas se réduire à une telle conception. Plusieurs stratégies sont possibles : modèle classique ou modèles marginaux ? Dans la première stratégie on cherche à construire un modèle dont le prédicteur linéaire comprend toutes les interactions de deux facteurs ; ces interactions sont alors contraintes pour assurer cette réduction du rang des interactions. Dans la seconde, on définit des modèles pour toutes les (respectivement, certaines) marges d'ordre deux du tableau de données ; on cherche alors s'il existe une probabilité conjointe pour le tableau de données qui satisfait aux contraintes ainsi définies ; le cas échéant, elle est estimée. Les deux approches posent des problèmes et des perspectives de généralisation spécifiques.

Dans cet exposé, je détaillerai les résultats obtenus pour les analyses des correspondances. Je présenterai aussi d'autres points de vue récents sur des décompositions en valeurs singulières non standards ou sur des modèles de rang réduit plus généraux.

Analyse de Variation Nonlinéaire de Données Fonctionnelles, Nouvelles Perspectives en Biologie Evolutive

Rima IZEM *, J.S MARRON, J.G KINGSLOVER

* Adresse pour correspondance :
Statistics Department, Science Center
One Oxford Street
Cambridge, MA 02138-2901 USA
e-mail : rizem@stat.harvard.edu (après 07-04)
rizem@email.unc.edu (avant 07-04)

Séance du 14 Juin 2004

Résumé

Un ensemble de courbes ou d'images de forme commune est un jeu de données de plus en plus courant dans les sciences, et en particulier en biologie. L'analyse de variations dans ces données est souvent ce qui importe le plus aux biologistes. Quand les modes de variation sont linéaires, ils sont bien décrits par des méthodes statistiques comme l'Analyse en Composante Principale. Quand les modes de variation sont non-linéaires, ces méthodes usuelles ne donnent pas de résultats interprétables. Je présente ici une nouvelle méthode en Analyse des Données Fonctionnelles pour la décomposition de variations non-linéaires. Cette méthode a été motivée par un problème concret en biologie évolutive : quantifier l'effet de la sélection sur l'évolution de certains organismes. Notre analyse atteint deux buts importants. Le premier est de décomposer la variation dans un ensemble de courbes de forme commune en modes prédéterminés linéaires ou non-linéaires. Le second est de quantifier chaque mode pour comparer les différentes contributions à la variation totale. On définit un nouveau quotient de sommes des carrés qui permet la quantification des modes de variation linéaires ou non-linéaires. Cette décomposition et quantification prend en compte la géométrie de l'espace de variation. La méthode est appliquée à un exemple de normes de réaction continues en biologie évolutive : courbes représentant la croissance de chenilles à différentes températures. Notre décomposition montre que les modes génétiques non-linéaires sont dominants.

Références

- Bhattacharya, R. Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. *Annals of statistics*, **31**, 1-29.
- Chiou, J.M., Muller, H.G., Wang, J.L. (2003). Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **65**, 405-423, Part 2.
- Fréchet M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. (French) *Ann. Inst. H. Poincaré* 10, (1948). 215–310.
- Izem R., Marron J.S, Kingsolver J.G (2004). Analyzing nonlinear variation of thermal performance curves. *Proceedings of the American Statistical Association*.
- Izem R., Marron J.S (2003). Quantifying nonlinear modes of variation of curves. *Proceeding of the International Statistical Institute 54th session, Berlin, Germany*.
- Kingsolver J.G., Gomulkiewicz R. , Carter P.A. (2001). Variation, selection and evolution of function-valued traits. *Genetica*, **112-113**, 87-104.
- Kirkpatrick, M. and Heckman, N. (1989). A quantitative genetic model for growth, shape, reaction norms and other infinite-dimensional characters. *Journal of Mathematical Biology*, **27**, 429-450.
- Ramsay J. O. , and Silverman B. W. , (1997) *Functional data analysis*, Springer : Springer Series in Statistics.

The translated Poisson distribution for fitting
 underdispersed discrete data : application to the
 discrimination of host plants on the growth of the spiralling
 whitefly *Aleurodicus dispersus Russel Homoptera :*
Aleyrodidae

D. MIZERE*, **S. DOSSOU-GBETE** and **A. KIYINDOU**

* Adresse pour correspondance :

Faculté des Sciences, Université Marien NGOUABI, BP.69, Brazzaville
 (République du Congo),

et

Laboratoire de Mathématiques Appliquées, Université de Pau (France)

e-mail : dmizere@yahoo.fr

Séance du 28 Juin 2004

Abstract

In this paper, we have undertaken to deal the probability distribution which fits the underdispersed discrete data. We propose that the underdispersed variable will be decomposed in an additive model of an unknown constant and Poisson latent variable for seeking its probability distribution.

The model will be applied on data collected in Republic of Congo Brazzaville by A. KIYINDOU et al.[3] and related to a sample of spiralling whitefly *Aleurodicus dispersus Russel (Homoptera : Aleyrodidae)* described by the preimaginal span (from the egg to the adult stage) in order to point out the host plants which are favourable to the growth of this insect, among which are fruit species of great consumption such as Safou tree (*dacryodes edulis*), Mango tree (*mangifera indica*), Avocado tree (*Persea americana*) and Citrus tree (*citrus paradisi*) and one shade species the Hura (*hura crepitans*). These plants were entombed separately then covered with a mosquito net for isolating its ones anothers. Let remember that the *Aleurodicus dispersus Russel* is originated from the Caribbean islands and Central America. It colonized Asia and Afrika including the Republic of Congo Brazzaville. This parasite is devastating for the plants that it infests (sucking the sap, decline in photosynthesis activity, fall of leaves,...).

Elementary statistics :

Host plants	\bar{Y}	S^2
Safou tree	30.81132	0.6944848
Mango tree	33.346946	4.856293
Hura	28.88	8.148571
Avocado tree	29.57895	5.223329
Citrus tree	29.53846	10.57085

The asymptotic Poisson underdispersed test (see D. MIZERE and al.[7]) based on

$$U_{sous} = \sqrt{\frac{n-1}{2}} \frac{(\bar{Y} - S^2)}{S^2} \sim N(0, 1) \quad (6)$$

is $U_{sous} = 221.1227$ for Safo tree, $U_{sous} = 28.74108$ for Mango tree, $U_{sous} = 12.59305$ for Hura, $U_{sous} = 20.05569$ for Avocado tree and $U_{sous} = 7.82131$ for Citrus tree, indicating a strong presence of underdispersion.

Références

- 1 BOHNING D. : “ A note on a test for Poisson overdispersion”. *Biometrika*(1994), 81, 2, pp.418-9.
- 2 JORGENSEN B. :”The theory of Dispersion Models”, Chapman & Hall, 1997.
- 3 KINYINDOU A. , ADOUMBAYE I. P., MIZERE D. et MOUSSA J.B. :” Influence de la plante hôte sur le développement et la reproduction de l’aleurode *Aleurodicus dispersus Russel* (Hom. :*Aleyrodidae*) en République du Congo”. *Fruits*, vol.54, pp.115-122, Elsevier, Paris, 1999.
- 4 KIYINDOU A. :” Contribution à l’étude de l’écologie et de la biologie des homoptères ravageurs des arbres fruitiers et autres plantes II.-Influence de la plante hôte et sa phénologie sur l’évolution des infestations de l’*Aleurodicus dispersus Russel*(Hom. ;*Aleyrodidae*) en République du Congo”. *Tropicultura*, 2003,21,1,10-15.
- 5 NORMAN L. J., SAMUEL KOTZ and ADRIENNE W. KEMP. :”Univariate discrete distributions”. Second Edition, Wiley Series in Probability and Mathematical statistics, 1993.
- 6 Mc CULLAG and NELDER T. A. :” Generalized linear model”. Second Edition, Chapman & Hall, 1989,
- 7 MIZERE D. and DOSSOU-BGETE S. :” A test on underdispersed for poisson distribution”. Pré-publication, Laboratoire de Mathématiques Appliquées, Université de Pau, 2003.

SOMMAIRE DES ANNÉES PCÉDENTES

Année 2002-2003

Publications du # LSP 2003-05

- *Un exemple d'analyse de données fonctionnelles en océanologie : le cas de l'Etang de Berre*, David NERINI et Claude MANTÉ.
- *Sur la décomposition en valeurs singulières complexes et applications*, Simplicie DOSSOU-GBETE et Antoine de FALGUEROLLES.
- *ACP dans le domaine des fréquences appliquée à la météorologie*, Olivier CAUMONT.
- *Consistance de l'estimation des paramètres d'un modèle non linéaire sur données fonctionnelles.*, Fabrice ROSSI.
- *Méthodes multivariées pour caractériser la distribution spatiale et saisonnière de la population de merlan de la mer Celtique*, Marion VERDOIT.
- *Contribution à l'étude des modèles de régression pour variables aléatoires fonctionnelles*, Aldo GOIA.
- *Analyse de la dépendance simultanée entre deux ensembles.*, Gabriel KISSITA.
- *Optimal Classification Trees*, Jarumir ANTOCH, Jan KLASCHKA and Petr SAVICKÝ.
- *Kernel estimates of hazard functions and application*, Ivana HOROVÁ.
- *Centrages et disparités : peut-on "négocier" avec les données?* , Jacques VANPOUCKE.
- *Artificial neural network to classify ECG beats.*, M.A. CHIKH, N. BELGACEM et F. BERESKI-REGUIG.
- *Data-driven rate optimal specification testing in non parametric regression models*, Emmanuel GUERRE et Pascal LAVERGNE.
- *Modélisation statistique pour variables aléatoires fonctionnelles : Théorie et applications*, Frédéric FERRATY.

2èmes Journées de Statistique Fonctionnelle :
Toulouse 12-13 Juin 2003
Publications du # LSP 2003-06

- *Using density level sets for nonparametric control charts*, Amparo BAILLO.
- *ACP conditionnelle*, Mohamed E. BAUCHE, Tawfik BENCHIKH, Fatiha RACHEDI et Abderrahmane YOUSFATE.
- *ACP d'un processus stationnaire sous contrainte de B-mesurabilité*, Tawfik BENCHIKH et Abderrahmane YOUSFATE.
- *Recalage de courbes et analyse de variance fonctionnelle par ondelettes*, Jérémie BIGOT.
- *Modèle mixte de régression multiple à coefficients lisses*, Anestis ANTONIADIS et Noëlle BRU.
- *Comparison of parametric and semiparametric estimates in a degradation model*, Vincent COUALLIER.
- *Modèle linéaire généralisé et Directions Révélatrices*, Michel DELECROIX.
- *Etude asymptotique de l'Analyse Canonique : de l'approche matricielle et analytique à l'approche opératoire et tensorielle*, Jeanne FINE.
- *Estimation du potentiel d'interaction de paires d'un processus de Gibbs*, Jean-Michel BILLIOT et Michel GOULARD.
- *Vitesse optimale du cas i.i.d. pour l'estimation non-paramétrique de la densité invariante d'un système dynamique chaotique*, Salim LARDJANE.
- *Estimation de la fonction du taux d'occurrence d'événements ponctuels*, Christophe BONALDI et Nicolas MOLINARI.
- *Estimation fonctionnelle des modes conditionnels*, Abbes RABHI et Abderrahmane YOUSFATE.
- *Propriétés extrémales des valeurs singulières d'un opérateur compact*, Jean Jacques TÉCHENÉ.

Année 2001-2002

Sommaire de la publication # LSP 2002-12

- *Estimation fonctionnelle d'un opérateur de transition d'un processus de Markov à états continus*, Abderrahmane YOUSFATE.
- *PLS regression on a stochastic process*, Christian PREDA et Gilbert SA-PORTA.
- *Réconcilions ridge regression et troncature spectrale en testant la moyenne d'une courbe aléatoire*, André MAS.
- *Estimation fonctionnelle et Ondelettes*, Antoine AYACHE et Jean Michel LOUBES.
- *Analyse de l'activité d'un centre de renseignement téléphonique : étude par modèle additif avec composante d'interaction de dimension réduite*, Simplicie DOSSOU-GBÉTÉ.
- *Quelques principes de déviations modérées et lois du logarithme itéré dans le modèle autorégressif hilbertien*, Ludovic MENNETEAU.
- *Nonparametric estimation applied to sismicity of Galicia*, Alejandro QUINTELA DEL RIO et Graciela ESTEVEZ.
- *A modification of cross-validation procedure in kernel hazard estimation from dependent samples*, Graciela ESTEVEZ et al.
- *Partially linear models with dependent errors : some notes on estimation, bandwidth selection and testing of hypotheses.*, German ANEIROS PEREZ.
- *Sélection des variables en régression linéaire ; lien avec le modèle linéaire fonctionnel*, Guy Martial NKIET.
- *ACP Banachique*, Tawfik BENCHIKH et Abderrahmane YOUSFATE.
- *ACP dans le domaine des fréquences*, Alain BOUDOU et Sylvie VIGUIER-PLA.
- *Une approche unificatrice pour l'estimation non-paramétrique des distributions de valeurs extrêmes multivariées*, Belkacem ABDOUS.

Année 2001-2002
Sommaire de la publication # **LSP 2002-12**

- *Estimation fonctionnelle d'un opérateur de transition d'un processus de Markov à états continus*, Abderrahmane YOUSFATE.
- *PLS regression on a stochastic process*, Christian PREDA et Gilbert SA-PORTA.
- *Réconcilions ridge regression et troncature spectrale en testant la moyenne d'une courbe aléatoire*, André MAS.
- *Estimation fonctionnelle et Ondelettes*, Antoine AYACHE et Jean Michel LOUBES.
- *Analyse de l'activité d'un centre de renseignement téléphonique : étude par modèle additif avec composante d'interaction de dimension réduite*, Simplicie DOSSOU-GBÉTÉ.
- *Quelques principes de déviations modérées et lois du logarithme itéré dans le modèle autorégressif hilbertien*, Ludovic MENNETEAU.
- *Nonparametric estimation applied to sismicity of Galicia*, Alejandro QUIN-TELA DEL RIO et Graciela ESTEVEZ.
- *A modification of cross-validation procedure in kernel hazard estimation from dependent samples*, Graciela ESTEVEZ et al.
- *Partially linear models with dependent errors : some notes on estimation, bandwidth selection and testing of hypotheses.*, German ANEIROS PEREZ.
- *Sélection des variables en régression linéaire ; lien avec le modèle linéaire fonctionnel*, Guy Martial NKIET.
- *ACP Banachique*, Tawfik BENCHIKH et Abderrahmane YOUSFATE.
- *ACP dans le domaine des fréquences*, Alain BOUDOU et Sylvie VIGUIER-PLA.
- *Une approche unificatrice pour l'estimation non-paramétrique des distributions de valeurs extrêmes multivariées*, Belkacem ABDOUS.

Journées de Statistique Fonctionnelle :
Toulouse 10-11 Juin 2002
Sommaire de la publication # LSP 2002-09

- *Méthode hongroise pour les accroissements limites d'un processus de Wiener.*, Abdelkader BAHRAM et Abderrahmane YOUSFATE.
- *ACP conditionnelle*, Mohamed E. BAUCHE, Tawfik BENCHIKH, Fatiha RACHEDI et Abderrahmane YOUSFATE.
- *Estimation localement suroptimale et adaptative de la densité*, Denis BOSQ.
- *ACP dans le domaine des fréquences : applications*, Alain BOUDOU et Sylvie VIGUIER-PLA.
- *Test d'additivité en régression non paramétrique sous des conditions de β -mélange*, Christine CAMLONG-VIOT.
- *On functional linear models and anova tests*, Antonio CUEVAS.
- *Modèles de régression sur variables fonctionnelles*, Frédéric FERRATY.
- *Un modèle semi-paramétrique Hilbertien*, Louis FERRÉ.
- *Estimation fonctionnelle en Ψ -régression*, Ali LAKSACI.
- *Un test d'homoscedasticité conditionnelle dans les modèles*, Djamal LOUANI.
- *Prédiction dans le modèle linéaire fonctionnel*, André MAS.
- *On the (intradaily) seasonality and dynamics of a financial point process : a semi-parametric approach*, Juan M. RODRIGUEZ POO.
- *Le produit tensoriel saurait-il mieux la Statistique que le statisticien ?*, Yves ROMAIN.
- *Une approche semi-paramétrique pour l'estimation de courbes de références*, Jérôme SARACCO.
- *Sur l'estimation fonctionnelle des opérateurs de transition des processus U-markoviens*, Abderrahmane YOUSFATE.
- *Pourquoi les scores de second ordre sont des opérateurs à signe non constant pour les distributions générant une variété à courbure négative*, Abdelghani ALI-ZAZOU et Abderrahmane YOUSFATE.

Année 2000-2001
Sommaire de la publication # **LSP 2001-07**

- *Sur les effets de la dimension en estimation fonctionnelle du réel vers le fonctionnel*, P. Vieu
- *Estimations dans le modèle linéaire fonctionnel*, F. Ferraty, H. Cardot et P. Sarda
- *Differential equation and inverse problems*, A. Vanhems
- *Quelques aspects des grandes déviations en estimation fonctionnelle*, D. Louani
- *Non uniformity of job matching in a transition economy : A nonparametric analysis for the czech republic*, S. Sperlich et S. Profit
- *Modèle additif de régression sous des conditions de mélange*, C. Camlong-Viot
- *Contributions à la Statistique Multidimensionnelle Opératoireielle*, Y. Romain
- *Contributions à l'Estimation Fonctionnelle*, P. Sarda
- *A propos de flux paramétriques*, J. Ramsay
- *Nonlinear alignment of time series with applications to varve chronologies*, D. Tjostheim
- *Boosting wavelets in electrophoresis*, J.Y. Koo
- *The deepest regression method*, P. Rousseuw
- *Estimation de l'occupation des sols à partir de l'évolution temporelles des images du capteur végétation SPOT*, R. Faivre, H. Cardot, M. Goulard et H. Vialard
- *Estimation pour le modèle de Lotka-Volterra*, S. Froda
- *Perturbations d'opérateurs aléatoires et applications*, J. Fine
- *Tests d'hypothèse dans le modèle de régression linéaire fonctionnel*, A. Goia
- *Produits (tensoriels et de convolution) de mesures (aléatoires et spectrales)*, A. Boudou et Y. Romain
- *Analyses factorielles de densités estimées par noyaux gaussiens*, R. Boumaza

Année 1999-2000
Sommaire de la publication # **LSP 2001-05**

- *Estimation fonctionnelle*, P. Sarda et P. Vieu
- *Modélisation pour variables fonctionnelles dans un contexte explicatif*, H. Cardot et F. Ferraty
- *Sur et pour une approche fonctionnelle en statistique*, Y. Romain
- *Produit de convolution de mesures spectrales*, A. Boudou
- *The geometrical theory of estimating functions*, C. Small
- *Inférence statistique pour des estimateurs de discontinuités dans un cadre non paramétrique*, V. Couallier
- *Nonparametric estimation in null recurrent time series* D. Tjøstheim
- *ACP de fonctions de densité. Application aux données climatiques*, T. Antoniadou et al.
- *Modèle non linéaire fonctionnel : une approche par régression inverse*, A.F. Yao et L. Ferré
- *Estimation bayésienne de l'intensité d'un processus de Cox non homogène par une méthode MCMC à saut réversible*, M. Goulard
- *Permutation tests in change point analysis*, J. Antoch et M. Hušková
- *Inférence statistique pour la localisation d'une discontinuité par régression linéaire locale*, G. Grégoire
- *Non causalité et discrétisation fonctionnelle, théorèmes limites pour un processus ARHX(1)*, S. Guillas
- *Data exploration using piecewise polynomial regression trees*, P. Chaudhuri