Groupe de travail STAPH :
Statistique Fonctionnelle et Opératorielle

Recueil de résumés 2007-2008

Coordinateurs
A. Boudou, 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 2007-2008.

Rappelons que ce groupe de travail en Statistique Fonctionnelle et Opératorielle, créé il y a quelques années 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 opératorielle, modèles de réduction de dimension, modèles pour variables fonctionnelles, ...) ; ils sont de natures 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.

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

http://www.lsp.ups-tlse.fr/staph.html

TABLE DES MATIERES

STAPH : Bilan d'activités 2007-2008

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

Coordinateurs du groupe de travail STAPH

boudou@cict.fr, ferraty@cict.fr, romain@cict.fr
sarda@cict.fr, vieu@cict.fr, viguier@cict.fr

Depuis plusieurs années l'objectif du groupe de travail STAPH au sein de l'Equipe de Statistique et Probabilités de l'Institut de Mathématiques de Toulouse est d'alimenter et de dynamiser les différents aspects fonctionnels et opératoriels de la Statistique, thématiques qui occupent désormais une place importante dans la recherche en Statistique. Les activités du groupe STAPH couvrent plusieurs domaines théoriques et/ou appliqués de la Statistique, et elles se traduisent par de nombreuses publications et collaborations nationales et internationales[1]. Le présent recueil de résumés récapitule les divers exposés présentés lors des séances du groupe de travail STAPH durant l'année universitaire 2007-2008.

L'année qui vient de s'écouler a été, pour le groupe STAPH, particulièrement riche en événements scientifiques, avec le 1er Workshop International en Statistique Fonctionnelle et Opératorielle (IWFOS2008) qui s'est tenu à Toulouse du 19 au 21 Juin 2008. Cette conférence, qui a réuni plus de 100 participants venus d'une vingtaine de pays, a vu se succéder des exposés donnés par quelques uns des leaders mondiaux de la recherche en Statistique et des exposés donnés par des jeunes chercheurs[2]. Incontestablement, cette conférence a été un franc succès, qui doit nous inciter à renouveler cette expérience.

Une 2ème édition du Workshop International en Statistique Fonctionnelle et Opératorielle est d'ores et déjà prévue dans un futur assez proche. D'ici là, le groupe STAPH donne rendez-vous pour la 6ème édition des Journées de Statistique Fonctionnelle et Opératorielle qui aura lieu en juin prochain à l'Université de Bourgogne de Dijon.

Enfin, nous tenons à exprimer nos remerciements à tous ceux qui ont accepté de présenter leurs travaux lors de nos séances de travail ou lors du workshop IWFOS'2008. La diversité de ces intervenants (tant thématique que géographique) est une des clés du succès de nos activités.

[1] Nos activités sont accessibles sur $http://www.lsp.ups-tlse.fr/staph/$

[2] S. Dabo Niang and F. Ferraty, 2008. Functional and Operatorial Statistics (Eds.). Contributions to Statistics, Physica Verlag

Combining registration and fitting for functional models

KNEIP Alois

* Adresse pour correspondance :
Institute for Economics and Social Sciences
University of Bonn
Adenauerallee, 24
53113 Bonn, Allemagne
e-mail : akneip@uni-bonn.de

Abstract

A registration method is vaguely defined as a process of aligning features of a sample of curves by monotonically transforming their domain. The aligned curves exhibit only amplitude variation, and the domain transformations, called warping functions, capture the phase amplitude variation, and variation in the original curves. In this talk, we precisely define a new type of registration process, in which the warping functions optimize the fit of a principal components decomposition to the aligned curves. The principal components are effectively the "features" that this process aligns. We discuss the relation of registration to closure of a function space under convex operations and define consistency for registration methods. An explicit decomposition of functional variation into amplitude and phase partitions is defined, and an algorithm for combining registration with principal components analysis is developed and applied to simulated and real data.

.

Regularized posteriors in linear ill-posed inverse problems

FLORENS Jean Pierre et SIMONI* Anna

* Adresse pour corespondance :
GREMAQ, Université des Sciences Sociales,
21 Allées de Brienne, 31000 Toulouse, France
e-mail : simoni.anna@gmail.com

Résumé

The aim of the paper is to obtain a solution for a signal-noise problem, namely we want to make inference on an unobserved infinite dimensional parameter through a noisy indirect observation of this quantity.
Let π and ρ be two measures on $\mathbb{R}$, we consider the following functional equation in Hilbert spaces :

$$Y = Kx, \tag{1}$$

with $x \in L^2_\pi$, $Y \in L^2_\rho$ and $K : L^2_\pi \to L^2_\rho$ a known, injective and Hilbert-Schmidt operator. The functional parameter of interest is x and it appears as the solution of an inverse problem that is ill-posed. Curve $\{Y\}$ represents the observed quantity, but it is typically measured with error, so that $Y \approx Kx$ and, due to the ill-posed nature of the problem, these small errors may cause errors of arbitrary size in the recovered x. We will denote with $\hat{Y}$ the noisy observation.
We place us in a Bayesian framework and we consider the statistical specification of the problem

$$\hat{Y} = Kx + U, \tag{2}$$

with x and U two Hilbert random variables with trajectories in L^2_π and L^2_ρ respectively. Bayesian approach to ill-posed inverse problems is an attempt to remove the ill-posedness by restating the inverse problem as a *well-posed extension* in a larger space of probability distributions. Therefore the solution to the inference problem is the posterior distribution of x.

We assume Gaussian prior and sampling distributions with nuclear covariance operators. Application of Bayes Theorem in functional space poses some difficulty

due to the non invertibility of covariance matrices, that still are compact operators. Thus, the solution to inverse problem (2) is defined to be a *regularized version of the posterior distribution*, namely a gaussian distribution with mean function and covariance operator computed by regularizing the functional equations identifying them.
After having defined this object, we study its asymptotics properties. In particular we analyze frequentist consistency. This means that we think data as generated from a distribution characterized by the true value of the parameter and we check accumulation of the regularized posterior distribution in a neighborhood of this true value. We prove that, for a true value of the parameter of interest satisfying some regularity conditions, posterior consistency is verified. However, we find a result of prior inconsistency, that is the specified prior distribution is not able to generate a value of the parameter x that satisfies the required regularity condition. It perfectly agrees with previous literature and confirms once again the possible prior inconsistency in infinite-dimensional Bayesian experiments already stressed by Diaconis and Freedman (1986)[5]. However, the prior distribution that we specify is able to generate trajectories of the parameter of interest very closed to the true value.
We also extend our analysis to variations of the basic model (2). Firstly, we develop asymptotic theory for the case with unknown operator K that is estimated by a kernel smoothing, without entering the Bayesian experiment. Secondly, we consider the case in which operator K is specific to each observations

$$\hat{Y}_i = K_i x + U_i, \quad i = 1, \ldots, N.$$

This is the typical case of a regression. The more compact way to make inference requires the use of a sufficient statistics. Therefore, we determine and prove sufficiency of a statistics for infinite dimensional parameters.
Finally, a Monte Carlo simulation confirms goods properties of the proposed estimator and we analyze the impact of the prior distribution on the estimated solution.

Bibliographie

[1] Belitser, E., and S., Ghosal (2003), *Adaptive Bayesian Inference on the Mean of an Infinite Dimensional Normal Distribution*, Annals of Statistics, **31**, 536-559.

[2] Carrasco, M., Florens, J.P., and E., Renault (2005), *Estimation based on Spectral Decomposition and Regularization*, forthcoming in Handbook of Econometrics, J.J. Heckman and E. Leamer, eds., **6**, Elsevier, North Holland.

[3] Cavalier, L., Golubev, G.K., Picard, D., and A.B., Tsybakov (2002), *Oracle Inequalities for Inverse Problems*, Annals of Statistics, **30**, 843-874.

[4] Darolles, S., Florens, J.P., and E., Renault (2006), *Nonparametric Instrumental Regression*, Working paper.

[5] Diaconis, F., and D., Freedman (1986), *On the Consistency of Bayes Estimates*, Annals of Statistics, **14**, 1-26.

[6] Florens, J.P., Mouchart, M., and J.M., Rolin (1990), *Elements of Bayesian Statistics*, Dekker, New York.

[7] Florens, J.P., and A., Simoni (2007), *Nonparametric Estimation of Instrumental Regression : a Bayesian Approach Based on Regularized Posterior*, working paper.

[8] Kaipio, J. and E., Somersalo (2005), *Statistical and Computational Inverse Problems*, **160**, Springer-Verlag, New-York.

[9] Kress, R. (1999), *Linear Integral Equation*, Springer.

[10] Van der Vaart, A.W., and J.H., Van Zanten (2000), *Rates of Contraction of Posterior Distributions Based on Gaussian Process Priors*, Working paper.

[11] Vapnik, V.N. (1998), *Statistical Learning Theory*, John Wiley & Sons, Inc.

[12] Zhao, L.H. (2000), *Bayesian Aspects of Some Nonparametric Problems*, Annals of Statistics, **28**, 532-552.

New functional linear regression estimates obtained by presmoothing

GONZÁLEZ-MANTEIGA Wenceslao
and MARTÍNEZ-CALVO* Adela

* Adresse pour correspondance :
Department of Statistics and OR
Faculty of Mathematics
University of Santiago de Compostela
CP 15782, Santiago de Compostela, Spain
e-mail : adelamc@usc.es

Résumé

Let $(E, \langle\cdot,\cdot\rangle)$ be a real separable Hilbert space and let us consider the functional linear model given by

$$Y = m(X) + \epsilon = \langle X, \theta\rangle + \epsilon,$$

where Y is a real random variable, X is a random variable valued in E, θ is a square integrable parameter valued in E and ϵ is a real random variable independent of X with zero mean and constant variance equal to σ^2. In recent years, some authors have studied this model and have proposed techniques for estimating θ (e.g. Ramsay and Silverman (2005) using splines or Cardot *et al.* (1999 and 2003) considering Functional Principal Components Analysis).

We propose two new estimates for the parameter θ obtained by minimizing with respect to θ a distance between the nonparametric estimate

$$\hat{m}_h(\cdot) = \frac{\sum_{i=1}^n Y_i K(h^{-1}d(\cdot, X_i))}{\sum_{i=1}^n K(h^{-1}d(\cdot, X_i))},$$

proposed in Ferraty and Vieu (2006) and studied in Ferraty *et al.* (2007), and the regression operator $m(\cdot) = \langle\theta, \cdot\rangle$. This idea was sucessfully used in the eighties in the real context (Faraldo-Roca and González-Manteiga (1985)) and in the multidimensional case (Cristóbal-Cristóbal *et al.* (1987)). This method allows to obtain a estimate based on the presmoothing of the response variable Y

$$\hat{\theta}_{K_n}^{h,Y} = \sum_{j=1}^{K_n} \frac{\Delta_n^h \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j$$

and a estimate based on the presmoothing of the independent variable X

$$\hat{\theta}_{K_n}^{h,X} = \sum_{j=1}^{K_n} \frac{\Delta_n \hat{v}_j}{\hat{\lambda}_j + h^2 \mu_2(K)} \hat{v}_j,$$

where $\Delta_n = \frac{1}{n}\sum_{i=1}^n X_i \otimes_{E'} Y_i$, $\Gamma_n = \frac{1}{n}\sum_{i=1}^n X_i \otimes_E X_i$ and $\Delta_n^h = \frac{1}{n}\sum_{i=1}^n X_i \otimes_{E'} \hat{m}_h(X_i)$. We can just see both $\hat{\theta}_{K_n}^{h,Y}$ and $\hat{\theta}_{K_n}^{h,X}$ as presmoothing versions of the least-squares estimate

$$\hat{\theta}_{K_n} = \sum_{j=1}^{K_n} \frac{\Delta_n \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j$$

proposed in Cardot *et al.* (1999 and 2003).

For these new estimates, consistency is obtained and the calculations of the mean-square error are used to study their efficiency by comparison with the least-squares estimate.

Références

- H. Cardot, F. Ferraty, and P. Sarda, *Functional linear model*, Statistics and Probability Letters **45** (1999), no. 1, 11–22.
- H. Cardot, F. Ferraty, and P. Sarda, *Spline stimators for the functional linear model*, Statistica Sinica **13** (2003), no. 3, 571–591.
- J. A. Cristóbal-Cristóbal, P. Faraldo-Roca, and W. González-Manteiga, *A class of linear regression parameter estimators constructed by nonparametric estimation*, Annals of Statistics **15** (1987), no. 2, 603–609.
- P. Faraldo-Roca and W. González-Manteiga, *On efficiency of a new class of linear regression estimates obtained by preliminary non-parametric estimation*, New Perspectives in Theoretical and Applied Statistics (New York) (M. Puri, ed.), Wiley, 1985, pp. 229–242.
- F. Ferraty, A. Mas, and P. Vieu, *Nonparametric regression on functional data : inference and practical aspects*, Australian and New Zealand Journal of Statistics **49** (2007), no. 3, 267–286.
- F. Ferraty and P. Vieu, *Nonparametric functional data analysis : theory and practice*, Springer, New York, 2006.
- J.O. Ramsay and B.W. Silverman, *Functional data analysis*, Springer, New York, 2005.

Estimateurs par splines de lissage en régression linéaire fonctionnelle.

CRAMBES Christophe*, KNEIP Alois, SARDA Pascal

* Adresse pour correspondance :
Université Montpellier 2, Case Courrier 051, Place Eugène Bataillon, 34095 MONTPELLIER Cedex
e-mail : ccrambes@math.univ-montp2.fr

Résumé

Dans de nombreuses applications (climatologie, télédétection, linguistique, ...), les données proviennent de l'observation de phénomènes continus du temps ou de l'espace. Ces données sont communément appelées *données fonctionnelles* dans la littérature, et font l'objet de nombreux travaux à l'heure actuelle. Pour une vue d'ensemble des techniques d'analyse de données fonctionnelles, on peut notamment se référer aux monographies [10], [11] et [7].

On s'intéresse ici au *modèle linéaire fonctionnel*, où on souhaite expliquer les effets d'une variable X sur une variable Y, la variable X (variable explicative) étant fonctionnelle et prenant ses valeurs dans l'espace $L^2([0,1])$ des fonctions de carré intégrable sur $[0,1]$ et la variable Y (variable d'intérêt) étant à valeurs réelles. Dans ce contexte, le modèle de régression linéaire fonctionnelle s'écrit alors

$$Y_i = \alpha_0 + \langle \alpha, X_i \rangle + \epsilon_i = \alpha_0 + \int_0^1 \alpha(t) X_i(t) dt + \epsilon_i, \qquad (3)$$

pour $i = 1, \ldots, n$, où la fonction $\alpha \in L^2([0,1])$ est inconnue et ϵ_i est une variable aléatoire d'erreur. Le but est alors d'estimer la fonction α à partir des observations $(X_i, Y_i)_{i=1,\ldots,n}$, de même loi que (X, Y). Ce modèle a été étudié par plusieurs auteurs. Les premiers travaux sur ce modèle peuvent être trouvés dans [9]. Plus récemment dans [3] et [4], deux estimateurs de α ont été proposés, l'un par régression fonctionnelle sur composantes principales, l'autre basé sur les splines de régression.

Dans ce travail, on présente un estimateur basé sur les splines de lissage (pour une vue d'ensemble sur les splines de lissage, voir par exemple [6]). Pour cela, on considère les points de mesure $0 < t_1 < \ldots < t_p < 1$ des courbes X_i. Pour

définir l'estimateur $\widehat{\alpha}$ de α, on considère un problème de minimisation de type moindres carrés auquel on rajoute un terme de pénalité qui permet de contrôler le lissage de l'estimateur. En modifiant légèrement cette pénalité, l'existence de notre estimateur est assurée sous des conditions très générales. On définit ensuite un estimateur de α_0 en posant $\widehat{\alpha}_0 = \overline{Y} + \langle \widehat{\alpha}, \overline{X} \rangle$.

On établit un résultat de convergence de cet estimateur, relativement à la semi-norme induite par l'opérateur de covariance de X,

$$\Gamma := \mathbb{E}\left[\left(\langle X - \mathbb{E}(X) \right), . \rangle \left(X - \mathbb{E}(X) \right) \right].$$

Le fait d'utiliser cette semi-norme permet d'interpréter nos résultats en termes d'erreur de prévision. En effet, pour une nouvelle observation (X_{n+1}, Y_{n+1}) avec X_{n+1} indépendante de $X_1, \ldots, X_n$, on prévoit Y_{n+1} par la quantité $\widehat{\alpha}_0 + \langle \widehat{\alpha}, X_{n+1} \rangle$ et on montre que l'on a

$$\mathbb{E}\left[\left((\widehat{\alpha}_0 + \langle \widehat{\alpha}, X_{n+1} \rangle) - (\alpha_0 + \langle \alpha, X_{n+1} \rangle) \right)^2 | \widehat{\alpha}_0, \widehat{\alpha} \right] = \|\widehat{\alpha} - \alpha\|_\Gamma^2 + O_P\left(n^{-1}\right).$$

Les vitesses obtenues font essentiellement appel à des hypothèses de régularité des courbes X_i et de la fonction α. Plus précisément, si α est m fois dérivable et si sa dérivée d'ordre m est dans $L^2([0,1])$, alors on montre que

$$\|\widehat{\alpha} - \alpha\|_\Gamma^2 = O_P\left(n^{-(2m+2q+1)/(2m+2q+2)}\right),$$

où q est un entier lié à la structure des courbes X_i, plus précisément à la décroissance des valeurs propres de l'opérateur de covariance Γ. On montre alors que la vitesse obtenue est optimale relativement à une certaine classe de fonctions α et de courbes X en ce sens qu'elle constitue une borne inférieure pour la vitesse de convergence dans cette classe. On compare ces vitesses avec celles obtenues dans [1] qui s'intéressent à l'erreur $(\widehat{\alpha}_0 + \langle \widehat{\alpha}, x \rangle) - (\alpha_0 + \langle \alpha, x \rangle)$ pour un x fixé non aléatoire, ce qui constitue une différence importante dans le cadre du modèle linéaire fonctionnel.

Dans le modèle (3), la variable X est supposée être observée sans erreur, ce qui peut se révéler assez peu réaliste en pratique, étant donné que des erreurs (notamment de mesure) peuvent empêcher de connaître X exactement. Ainsi, on considère que les observations disponibles pour la variable explicative sont $W_1, \ldots, W_n$, telles qu'en chaque point de mesure, on a

$$W_i(t_j) = X_i(t_j) + \delta_{ij}, \tag{4}$$

pour tout $i = 1, \ldots, n$ et tout $j = 1, \ldots, p$, où $(\delta_{ij})_{i=1,\ldots,n, j=1,\ldots,p}$ est une suite de variables aléatoires indépendantes et identiquement distribuées représentant les erreurs faites en chaque point $t_1, \ldots, t_p$. Le but est donc d'estimer la fonction α sur la base des observations $(W_i, Y_i)_{i=1,\ldots,n}$. On adapte pour cela l'estimateur spline construit dans le cas où la variable X est réellement observée, en utilisant le principe des moindres carrés orthogonaux (voir par exemple [8] ou [12]). Cette méthode consiste à faire intervenir, dans le problème de minimisation de type moindres carrés (pénalisés) non seulement les carrés des résidus sur la variable d'intérêt mais aussi les carrés des résidus sur la variable explicative. On obtient là aussi un résultat de convergence pour cet estimateur $\widehat{\alpha}_W$ vis-à-vis de la semi-norme induite par Γ. La vitesse obtenue est, sous certaines conditions, la même que celle dans le cas où il n'y a pas de bruit sur la variable explicative. L'ensemble des résultats présentés sont issus de [2].

Références

[1] Cai, T.T. and Hall, P. (2006). Prediction in functional linear regression. *Annals of Statistics*, **34**, 2159-2179.

[2] Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statistics and Probability Letters*, **45**, 11-22.

[3] Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica*, **13**, 571-591.

[4] Crambes, C., Kneip, A. and Sarda, P. (2007). Smoothing splines estimators for functional linear regression. *To appear.*

[5] Eubank, R.L. (1988). *Spline smoothing and nonparametric regression.* Marcel Dekker.

[6] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis : theory and practice.* Springer, New York.

[7] Golub, G.H. and Van Loan, C.F. (1980). An analysis of the total least squares problem. *SIAM, Journal of Numerical Analysis*, **17**, 883-893.

[8] Ramsay, J.O. and Dalzell, C.J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B*, **53**, 539-572.

[9] Ramsay, J.O. and Silverman, B.W. (2002). *Applied functional data analysis.* Springer, New York.

[10] Ramsay, J.O. and Silverman, B.W. (2005). *Functional data analysis* (Second edition). Springer, New York.

[11] Van Huffel, S. and Vandewalle, J. (1991). *The total teast squares problem : computational aspects and analysis.* SIAM, Philadelphia.

.

Some recent results for stationary processes via their associated measures

CABRAL Emmanuel N.

Adresse pour correspondance :
Laboratoire de Statistique et Probabilités, UMR CNRS C55830
118, route de Narbonne
31062 Toulouse cedex.
e-mail : cabral@cict.fr

Abstract

Let $(X_g)_{g\in G} \subset H$ be a complex Hilbertian stationary process where the index space G is a locally compact Abelian group (with dual space $\widehat{G}$ admitting a countable basis). In particular G may be anyone of the sets $\mathbb{Z}$, $\mathbb{Z}^k$, $\mathbb{R}$, $\mathbb{R}^k$, $\Pi = [-\pi, \pi[$, or Π^k and this approach permits considering jointly discrete and continuous-time processes. To any stationary process $(X_g)_{g\in G}$, we know (see, e.g. Azencott and Dacunha-Castelle, 1984), by the Fourier transform, we may associated a *random measure (r.m.)* which is denoted by Z. This measure is unique and we also may associate to this r.m. a unique *spectral measure (s.m.)* denoted by ε. To this s.m. defined on $\mathcal{B}_{\widehat{G}}$, the Borel field of $\widehat{G}$, for the complex Hilbertian space H, we may associate a unique unitary operator denoted by U, called *the unitary operator deduced from* ε, and conversely.

With these tools associated to a stationary process, we propose an example of discrete case where the index set G is $\mathbb{Z}$ and $(X_g)_{g\in G}$ is such that $X_g = \sum_{j\in J} e^{i\lambda_j g} Z_j$, $\|J\| < \infty$, for all $g \in \mathbb{Z}$.

Under the previous assumption, results for the tensor product of discrete and continuous-time processes are presented (see also Birman et al., 1996 ; Boudou et al., 2002). They concern first the convolution product of two spectral measures and secondly the convolution product of random measures. The obtained results are applied to stationary processes statistics with both following examples :

(i) An interpolation problem : given a stationary series $(X_n)_{n\in\mathbb{Z}}$, how can we define all stationary series $(Y_n)_{n\in\mathbb{Z}}$ such as $Y_{nq} = X_n$, $q \in \mathbb{N}^*$?

(ii) An inverse Fourier transform problem : let $(X_n Y_n)_{n\in\mathbb{Z}}$ be a multiplicatively decomposable series, how can we define the random measure whose Fourier transform is the considered series ?

References

Azencott R., Dacunha-Castelle D. (1984). *Séries d'observations irrégulières. Modélisation et prévision.* Techniques Stochastiques, Masson.

Birman M., Solomjak M. (1996). Tensor product of a finite number of spectral measures is always a spectral measure. *Integ. Eq. Oper. Th.*, 24, 179-187.

Boudou A. (2003). Interpolation de processus stationnaires. *C.R. Acad. Sci. Paris*, SérieI, Math., 336, n°12, 1021-1024.

Boudou A. (2007). Groupe d'opérateurs unitaires déduit d'une mesure spectrale - une application. *C.R. Acad. Sci. Paris*, SérieI, Math.,344, n°12, 791-794.

Boudou A., Romain Y. (2002). On spectral and random measures associated to continuous and discrete time processes. *Stat. Proba. Letters*, 59, 145-157.

Carte de Kohonen par noyau et application à la classification de sommets de graphes

VILLA Nathalie* et ROSSI Fabrice

* Adresse pour correspondance :
Institut de Mathématiques de Toulouse, 118 route de Narbonne, 31062 Toulouse cedex 9
e-mail : nathalie.villa@math.univ-toulouse.fr

Résumé

Plusieurs modifications de l'algorithme de cartes auto-organisatrices permettent de s'intéresser à des données non vectorielles. Récemment, nous avons introduit une version par noyau qui s'adaptent à un très grand nombre de types de données. Nous montrerons comment cette version est reliée aux précédentes modifications du SOM et nous illustrerons son utilisation par une application à la classification de sommets de graphes.

Références.

Boulet, R., Jouve, B., Rossi, F. & Villa, N. (2008). Batch kernel SOM and related Laplacian methods for social network analysis. *Neurocomputing*. To appear.

El Golli, A., Rossi, F., Conan-Guez, B. & Lechevallier, Y. (2006). Une adaptation des cartes auto-organisatrices pour des données décrites par un tableau de dissimilarités. *Revue de Statistique Appliquée*, **LIV**(3), 33–64.

Kohohen, T. & Somervuo, P. (1998). Self-Organizing maps of symbol strings. *Neurocomputing*, **21**, 19–30.

Kohonen, T. (2001). *Self-Organizing Maps, 3rd Edition*, volume 30. Springer, Berlin, Heidelberg, New York.

Lau, K., Yin, H. & Hubbard, S. (2006). Kernel self-organising maps for classification. *Neurocomputing*, **69**, 2033–2040.

Villa, N. & Boulet, R. (2007). Clustering a medieval social network by SOM using a kernel based distance measure. In *Proceedings of ESANN 2007* (ed. Verleysen, M.) 31–36, Bruges, Belgium.

Villa, N. & Rossi, F. (2007). A comparison between dissimilarity SOM and kernel SOM for clustering the vertices of a graph. In *Proceedings of the 6th Workshop on Self-Organizing Maps (WSOM 07)*, Bielefield, Germany.

von Luxburg, U. (2007). A tutorial on spectral clustering. Technical Report TR-149, Max Planck Institut für biologische Kybernetik. Avaliable at
http ://www.kyb.mpg.de/publications/attachments/luxburg06_TR_v2_4139%5B% 1%5D.pdf.

Nonlinear Principal Components for a Random Variable

GOIA Aldo* et SALINELLI Ernesto

* Adresse pour correspondance :
Università del Piemonte Orientale, Dipartimento SEMeQ,
Via Perrone, 18 - 28100 NOVARA (ITALY)
e-mail : goia@eco.unipmn.it, salinelli@eco.unipmn.it

Résumé

In the univariate context, only few cases lead to find a set of "simple" variables associated to a given random variable (r.v.) : examples are the works of Lancaster (1969) and its scholars about the canonical variates or, more recently, the ones of Cuadras and its coauthors (see e.g. Cuadras, 2005).

In the above-mentioned spirit, in Goia and Salinelli (2007) a definition of nonlinear principal components (NLPCs) for a r.v. X has been introduced specializing to the univariate case the variance maximizing problem on a set of centered and square integrable transformations used to define NLPCs for a random vector in Salinelli (1998 and 2007).

In particular, let X be an absolutely continuous r.v. with zero mean, finite variance $\mathbb{E}[X^2]$ and density f_X, and $\dot{W}^{1,2}_{f_X}$ the weighted Sobolev space of centered, square integrable functions with first derivative square integrable too. We say that the r.v. $Z_j = \varphi_j(X)$ is the j-th nonlinear principal component of X if φ_j is a solution of the problem

$$\begin{array}{lllr} \max & \mathbb{E}\left[u(X)^2\right] & & \\ \text{sub} & \mathbb{E}\left[u'(X)^2\right] = 1 & u \in \dot{W}^{1,2}_{f_X} & (5) \\ & \mathbb{E}\left[u(X)\varphi_s(X)\right] = 0 \quad s = 1, 2, \ldots, j-1, & j > 1. & \end{array}$$

We show that finding NLPCs is equivalent to find the j-th eigenfunction φ_j with corresponding eigenvalue λ_j of the covariance operator G associated to X, or to determine the j-th solution of a (second order) differential boundary value problem.

Some results on the existence of transformations φ_j are pointed out, also in terms of the moment generating function of X_j. Referring to classical results on self-adjoint, compact operators on Hilbert spaces we show how, under suitable

assumptions, G admits countably many real eigenvalues λ_j that are positive, simple $\lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_j > \cdots$, with 0 as unique limit point; the corresponding eigenfunctions φ_j are mutually orthogonal and form a complete set in $\dot{W}^{1,2}_{f_X}$. Furthermore, we prove that the first transformation φ_1 is strictly monotone, odd if f_X is even, concave in $[0,1]$ if f_X is unimodal symmetric. An important statement is that φ_1 characterizes the distribution of X. We relate our results to the literature concerning the so-called *Chernoff* (or *Poincaré*) *inequality* (see e.g. Chernoff, 1981) and with the characterizations discussed in Purkayastha and Bhandari (1990).

An estimation procedure of NLPC transformations φ_j is proposed. Consider a sample $\{X_i,\ i = 1, 2, \ldots, n\}$ of i.i.d. r.v.s drawn from X. Denoting by $\mathcal{S}_{k,d}$ the $k + d$ dimensional linear space of spline functions of order d defined on D, and by $\widehat{\dot{W}}^{1,2}_k$ the space of spline functions $u \in \mathcal{S}_{k,d}$ such that $\frac{1}{n}\sum_{i=1}^n u(X_i) = 0$, *a.s.*, we define the estimator $\widehat{\lambda}_{j,k,n}$ of λ_j as

$$\widehat{\lambda}_{j,k,n} = \max_{u \in \widehat{\dot{W}}^{1,2}_k} \frac{\sum_{i=1}^n (u(X_i))^2}{\sum_{i=1}^n (u'(X_i))^2} \tag{6}$$

subject to $\frac{1}{n}\sum_{i=1}^{n} u(X_i)\,\widehat{\varphi}_{h,k,n}(X_i) = 0$ *a.s.*, $1 \le h \le j-1$, and the estimator of φ_j as the maximizer $\widehat{\varphi}_{j,k,n} \in \widehat{\dot{W}}^{1,2}_k$ corresponding to $\widehat{\lambda}_{j,k,n}$.

We show how problem (6) can be converted into a generalized eigenvalue problem and then solved by using some available computer packages. Under suitable hypothesis, we obtain some asymptotic results. The practical performances of the estimator are analyzed by a simulation study for the uniform, truncated exponential, normal and centered and scaled beta distributions.

Several possible statistical applications of NLPCs are discussed. The first one concerns the use of characterization properties of the variance of the first NLPC transformation in defining some goodness-of-fit test that use a statistic based on $\widehat{\lambda}_1$. In particular we focus on two examples : testing the hypothesis that a distribution is Uniform and testing that a density is Wigner against an other unimodal symmetric one. For various sample sizes the powers of the tests are analyzed via simulation for many alternatives and compared with those of the Kolmogorv-Smirnov test.

The second application is based on the fact that the set of NLPC transformations represents a "special" basis : this enables us to analyze the dependence structure between two r.v.s by estimating suitable indexes of correlation and dependence. Finally, we show how the recalled property of NLPC transformations allows to define some classes of bivariate distributions with fixed margins : this idea is illustrated by obtaining some densities in the class of the Sarmanov-Lee family, that is a generalization of the popular Farlie-Gumbel-Morgenstern family.

Bibliographie

[1] Chernoff, H. (1981). A note on an inequality involving the normal distribution. *The Annals of Probability*, **9 (3)**, 533-535.

[2] Cuadras, C.M. (2005). First principal component characterization of a continuous random variable. In : *Advances on Models, Characterizations and Applications* (Balakrishnan, N., Bairamov, I. and Gebizlioglu, O., Eds.). Chapman & Hall/CRC-Press, New York, 189-199.

[3] Goia, A. and Salinelli E. (2007). Nonlinear Principal Components III. Univariate Distributions. WP 17/2007, Dipartimento SEMeQ, Università del Piemonte Orientale.

[4] Goia, A. and Salinelli E. (2007). Rate of convergence for spline estimates of Nonlinear Principal Components of Random Variables. WP 18/2007, Dipartimento SEMeQ, Università del Piemonte Orientale.

[5] Lancaster, H.O. (1969). *The chi-squared distribution.* Wiley, New York.

[6] Purkayastha, S. and Bhandari, S.K. (1990). Characterization of uniform distributions by inequality of Chernoff-type. *Sankhya*, **52**, Series A, 376-382.

[7] Salinelli E. (1998). Nonlinear principal components I. Absolutely continuous random variables with positive bounded densities, *The Annals of Statistics*, 26 (2), 596-616.

[8] Salinelli E. (2007). Nonlinear Principal Components II : Characterization of Normal Distributions. WP 16/2007, Dipartimento SEMeQ, Università del Piemonte Orientale.

.

ACP fonctionnelle avec des données collectées par sondage : estimateurs de type Horwitz-Thompson et fonction d'influence

CARDOT Hervé *, CHAOUCH Mohamed, GOGA Camélia et LABRUÈRE Catherine

* Adresse pour correspondance :
Institut de Mathématiques de Bourgogne, UMR CNRS 5584
Université de Bourgogne,
9 Avenue Alain Savary, BP 47870, 21078 DIJON Cedex, FRANCE
e-mail : Herve.Cardot@u-bourgogne.fr

Résumé

L'objectif de l'ACP fonctionnelle est de décrire les grands modes de variation de trajectoires de processus autour de leur trajectoire moyenne. La construction d'un sous-espace permettant de reconstruire au mieux ces variations respose sur l'estimation des vecteurs propres de la covariance associés aux plus grandes valeurs propres (Dauxois *et al.*, 1982, Ramsay & Silverman, 2005).

La manière dont les données fonctionnelles sont observées n'est que rarement prise en compte dans la littérature (cf Dessertaine, 2006 qui tient compte de la procédure de sondage) et les procédures usuelles de la statistique inférentielle peuvent se révéler complètement inadaptées.

Nous montrons comment la trajectoire moyenne et ces vecteurs propres peuvent etre estimés avec des estimateurs de type Horwitz-Thompson qui prennent en compte les probabilités d'inclusion dans l'échantillon lorsque les données sont sélectionnées avec des techniques de sondage. Nous proposons des estimateurs de leur variance basées sur des techniques de linearisation introduites par Deville (1999) dans un cadre général et la theorie des perturbations (Kato, 1966). En adoptant le cadre asymptotique de superpopulation d'Isaki & Fuller (1982) nous montrons la convergence de nos estimateurs. Des simulations confirment le bon comportement de ces estimateurs.

Nous montrons ensuite comment ces approches peuvent etre utilisées simplement pour prendre en compte nonparamétriquement une information auxiliaire et effectuer ainsi des prédictions.

Références

Cardot, H, Chaouch, M, Goga, C. and Labruère, C. (2007). Functional Principal Components Analysis with Survey Data. *Preprint.* Rapport Technique IMB, Université de Bourgogne.

Chiky, R, Hébrail, G. (2007). Generic tool for summarizing distributed data streams. *Preprint.*

Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a random vector function : some applications to statistical inference. *J. Multivariate Anal.*, **12**, 136-154.

Dessertaine A. (2006). Sondage et séries temporelles : une application pour la prévision de la consommation electrique. *38èmes Journées de Statistique*, Clamart, Juin 2006.

Deville, J.C. (1999). Variance estimation for complex statistics and estimators : linearization and residual techniques. *Survey Methodology*, **25**, 193-203.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Am. Statist. Ass.*, **47**, 663-685.

Isaki, C.T. and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *J. Am. Statist. Ass.* **77**, 89-96.

Kato, T. (1966). *Perturbation theory for linear operators.* Springer Verlag, Berlin.

Ramsay, J. O. and Silverman, B.W. (2005). *Functional Data Analysis.* Springer-Verlag, 2nd ed.

Skinner, C.J, Holmes, D.J, Smith, T.M.F (1986). The Effect of Sample Design on Principal Components Analysis. *J. Am. Statist. Ass.* **81**, 789-798.

Shift, Slope and Curvature of yields correlation matrices

SALINELLI Ernesto* and SGARRA Carlo

* Adresse pour correspondance :
Università del Piemonte Orientale, Dipartimento SEMeQ,
Via Perrone, 18 - 28100 NOVARA (ITALY)
e-mail : salinelli@eco.unipmn.it, carlo.sgarra@polimi.it

Résumé

The aim of this talk is to illustrate some recent results about the so-called shift, slope and curvature for yields correlation matrices, and some open questions and current works on this subject.

In the last decade a great deal of attention has been devoted to the development of statistical methods suitable to describe the movements of the yield curve and to empirically justify some important models for interest rate dynamics based on one or more factors (see e.g. [12], [19]). Principal Component Analysis has turned out to be one of the most useful tools to find such factors.

Consider a yield curve completely described (see [10]) by an n-dimensional random vector $\mathbf{X}^T = [X_1, X_2, \ldots, X_n]$ with $\mathbb{E}[\mathbf{X}] = \mathbf{0}$ and $\mathbb{E}[X_i^2] = 1$, whose components represent the forward rates corresponding to maturities $t_1 < t_2 < \ldots < t_n$. This vector can be expressed as a linear transformation $V\mathbf{Y}$ of the n-dimensional random vector $\mathbf{Y}$ of *principal components* (named *factors*) of $\mathbf{X} : Y_j = \mathbf{V}_j^T\mathbf{X}$. It is well known that the $\mathbf{V}_j$**'s**, for $j = 1, 2, \ldots, n$, are the normalized eigenvectors of the correlation matrix $R = \mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right]$ of $\mathbf{X}$ with corresponding eigenvalues λ_j such that : $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.

Given the high degree of correlation among the variables involved, the first three PCs are usually considered to be sufficient to account for the total variability of $\mathbf{X}$, i.e. $\sum_{s=1}^{3} \lambda_s \simeq \mathrm{tr}(R) = n$. It has been empirically observed (see [19]) that the first three eigenvalues of R are simple, and :

- the *first eigenvector* $\mathbf{V}_1$, called **shift** (or **level**), has approximately equal components,
- the *second eigenvector* $\mathbf{V}_2$, called **slope**, has elements of increasing magnitude approximately equal in modulus and with opposite signs at the end points of the maturity range,
- the *third eigenvector* $\mathbf{V}_3$, called **curvature**, has components first decreasing

then increasing, approximately equal at the end points of the maturity range and twice large and of the opposite sign**S** in the middle.

The previous empirically founded terms have been formalized by Salinelli and Sgarra [21] in the following way :

Definition. Given a correlation matrix R having the first three eigenvalues simple, we define :

– its *first eigenvector :* **pure shift** if it is $\mathbf{1} = \begin{bmatrix} 1, & 1, & \cdots, & 1 \end{bmatrix}^T$; **shift** if all its components are positive, strictly increasing and then strictly decreasing ; **weak shift** if all its components are positive ;

– its *second eigenvector :* **slope** if its elements are strictly increasing (strictly decreasing) presenting a unique change of sign ; **weak slope** if its elements present a unique change of sign ;

– its *third eigenvector :* **curvature** if its elements are strictly decreasing (increasing) and then strictly increasing (decreasing) presenting only two changes of sign ; **weak curvature** if its elements present only two changes of sign.

The summentioned authors, and Lord and Pelsser [13] have studied the connections between these spectral properties and some empirical features observed in the correlation matrices, i.e. :

a) interest rates at different maturities are always positively correlated ;

b) the correlation coefficients decrease when the distance between the indices increases ;

c) the previous reduction in the correlation between variables corresponding to the same difference in the indices tends to decrease as the maturities of both the variables increase.

Formalizing these properties and relating them to the property of *total positivity*, ([1], [8]) the authors have proved results on the sign changes of the first three eigenvectors of a correlation matrix for forward rates, i.e. on the presence of shift, slope and curvature in a weak sense.

More recently, Salinelli and Sgarra in [22] have considered the following correlation model

$$\rho_{ij} = \rho^{|i-j|} \qquad i, j = 1, 2, \ldots, n, \quad \rho \in (0, 1)\,. \tag{7}$$

that (as noted in [20, p.196]) is "less ad hoc and arbitrary than one might expect, and can be arrived at in several ways as a financially justifiable first-order building block of more realistic correlation functions." They have shown that for any choice of $\rho \in (0, 1)$ the first eigenvector of $R = [\rho_{ij}]$ is shift and there exists $\rho^* \in (0, 1)$ such that for all $\rho > \rho^*$ the second eigenvector of R is slope.

Bibliographie

[1] Ando T., Totally Positive Matrices, *Linear Algebra and Appl.* 90 (1987), 165-219.

[2] Forzani L., Tolmasky C.F., A Family of Models explaining the Level-Slope-Curvature effect, *International Journal of Theoretical and Applied Finance*, Vol.6, No.3, (2003), 239-255.

[3] Forzani L., Tolmasky C.F., On the spectral decomposition of empirical correlation matrices, *Journal of Knot Theory and Its Ramifications*, 10, n.8, (2001), 1201-1213.

[4] Friedman S., Weisberg H.F., Interpreting the first eigenvalue of a correlation matrix, *Educational and Psychological Measureance* 41 (1981), 11-21.

[5] Gantmacher F.R., Krein M.G., *Oscillation matrices and kernels and small vibrations of mechanical systems*, Dept. of Commerce, Washington, (1961).

[6] Golub B.W., Tilman L.M., Measuring Yield Curve Risk Using Principal Components Analysis, Value at Risk, and Key Rate Durations, *The Journal of Portfolio Management* (1997), 72-94.

[7] Kaiser H.F., A measure of the average intercorrelation, *Educational and Psychological Measurement* 28, (1968), 245-247.

[8] Karlin S., *Total Positivity*, Vol.1, Stanford University Press, Stanford CA, (1968).

[9] Kotz S., Pearn W.L., Wichern D.W., Eigenvalue-eigenvectors analysis for a class of patterned correlation matrices with an application, *Statistics & Probability Letters* 2 (1984), 119-125.

[10] Lardic S., Priaulet P., Priaulet S., PCA of the Yield Curve Dynamics : Questions of Methodologies, *Journal of Bond Trading and Management* 1 (2003), 327-349.

[11] Ledermann W., Bounds for the greatest latent roots of a positive matrix, *Journal of the London Math. Soc.* 35 (1960), 265-268.

[12] Litterman R., Scheinkman J., Common Factors Affecting Bond Returns, *Journal of Fixed Income* 1 (1991), 54-61.

[13] Lord R., Pelsser A., Level, Slope and Curvature : Art or Artifact ?, *Appl. Math. Finance* 14 (2) (2007), 105-130.

[14] Martellini L., Priaulet P., Priaulet S., *Fixed-Income Derivatives and Portfolio Management*, J. Wiley and Sons, (2006).

[15] Meyer E.P., A Measure of the average intercorrelation, *Educational and Psychological Measureament* 35 (1975), 67-72.

[16] Minc H., *Nonnegative Matrices*, Wiley, New York, (1988).

[17] Ostrowski A.M. , Bounds for the greatest latent root of a positive matrix, *J. London Math. Soc.* 27 (1952), 253-256.

[18] Ostrowski A.M., On the eigenvector belonging to the maximal root of a non-negative matrices, *Proc. Edinburgh Math. Soc.* 12 (1961), 107-112.

[19] Rebonato R., *Modern Pricing of Interest-Rate Derivatives*, Princeton University Press, Princeton (2002).

[20] R. Rebonato, Modern Pricing of Interest-rate Derivatives : The LIBOR Market Model and Beyond, Princeton University Press, (2002).

[21] Salinelli E., Sgarra C., Correlation matrices of yields and total positivity, *Linear Algebra Appl.* 418 (1-2) (2006), 682-692.

[22] Salinelli E., Sgarra C., Shift, Slope and Curvature for a Class of Yields Correlation Matrices, *Linear Algebra Appl.* 426 (1-2) (2007), 650-666.

[23] Schoenmakers J.G.M., Coffey B., Systematic Generation of Correlation Structures for the Libor Market Model, *Int. Journal of Theoretical and Applied Finance* 6 (4) (2003), 149-185.

[24] Willner R. , A new Tool for Portfolio Managers : Level, Slope, and Curvature Durations, *The Journal of Fixed Income*, (1996) 48-59.

Impact d'une mauvaise spécification de la variance sur la statistique du test F d'un modèle linéaire : étude de séries temporelles de richesse en sucre de la canne

LAURENT-CHABALIER Sabine

* Adresse pour correspondance :
Laboratoire de Probabilités et Statistique
Institut de mathématiques et de modélisation de Montpellier (I3M)
Université Montpellier II
Place Eugene Bataillon
34095, Montpellier, Cedex 5, France
e-mail : sabinel@wanadoo.fr

Résumé

Ce travail vise à étudier l'impact d'une mauvaise spécification de la variance sur la statistique du test F, pour l'étude de séries temporelles de richesse en sucre de la canne sur l'île de la Réunion. Nous voulons construire un modèle de prédiction des courbes de l'évolution de la richesse de la canne à sucre. Nous considérons le modèle linéaire $Y = X\beta + \epsilon$, où X est une matrice orthogonale supposée de plein rang et ϵ est le vecteur d'erreur de loi normale d'espérance nulle et de matrice de variance-covariance Ω inconnue. Nous proposons d'étudier la statistique du test F, dans le cas d'une mauvaise spécification de cette matrice de variance-covariance en considérant à tort qu'elle est égale à la matrice identité. Nous voulons approcher la loi de cette statistique de test pour un certain type de test d'hypothèse. Pour ceci, nous calculons les quatre premiers moments de la statistique du test F et étudions la densité de sa loi. Nous considérons trois formes de matrices Ω : AR(1), ARMA(1,1) et équi-corrélée. Ceci permet de mesurer l'impact de la mauvaise spécification sur l'espérance, la variance et les coefficients d'asymétrie et d'aplatissement de la loi de cette statistique de test.

Mots-clés : Modèle linéaire ; Test F ; misspecification ; moment ; formes quadratiques ; richesse en sucre ; canne à sucre

.

Recent results in change point analysis

HUSKOVÁ Marie

* Adresse pour correspondance :
Department of Statistics, Charles University,
Sokolovská 83, 186 75 Praha
Czech Republic
e-mail : huskova@karlin.mff.cuni.cz

Abstract

Change point analysis concerns procedures on stability of statistical models. Typically one has a sequence of observations $Y_1, \ldots, Y_n$ obtained at the ordered time points $t_1 < \ldots < t_n$ such that the first m observations follow a certain statistical model and after the m-th observation the model changes and the remaining $n - m$ observations follow another model. The point m is unknown and is called change point. The problem is to detect (test H_0 : no change & H_1 : there is a change) and to identify the location of such a change (estimate m).
The above formulated problem has many variants. It has attracted many researchers both from theoretical and practical point of view. Applications can be found in meteorology, climatology, hydrology or environmental studies, econometric time series, statistical quality control among others.

Recent results on detection and identification of changes in regression models and time series together with their properties will be presented. Particularly,

• approximations to the critical values based on resampling methods,

• ratio type test statistics for detection of changes,

• test procedures for distinguishing between a change in location parameter and random walk alternatives

will be discussed.

Theoretical results will be accompanied by applications on real data sets and by results of simulation study.

.

Asymptotics in regression on functional variable : Estimation and Structural tests

DELSOL Laurent

* Adresse pour correspondance :
Institut de Mathématiques, Université de Toulouse et C.N.R.S. (U.M.R. 5219),
118, route de Narbonne, 31062 Toulouse Cedex 9.
France
e-mail :laurent.delsol@math.univ-toulouse.fr

Summary

Functional data analysis is a typical issue in modern statistics. During the last years, many papers have been devoted to theoretical results or applied studies on models involving functional data. In this talk, we focus on regression models where a real response variable depends on a functional random variable taking its values in an infinite dimensional space.

We firstly focus on the problem of estimating the regression operator nonparametrically. We consider the kernel estimator proposed by [14] as a generalisation of the so-called Nadaraya-Watson estimator to the case of a functional covariate. Many results have been obtained for this estimator (see [15], [19], [13], [7], ...). We present various results linked with the asymptotic properties of this nonparametric kernel estimator. We firstly assume that the dataset under study is composed of strong-mixing variables and focus on a nonparametric regression model. A first result gives asymptotic normality of the kernel estimator with explicit expressions of the asymptotic dominant bias and variance terms (see [8] and [11]). On one hand, we propose from this result a way to construct asymptotic pointwise confidence bands. On the other hand, we decline from both asymptotical normality and uniform integrability results the explicit expressions of the asymptotic dominant terms of centered moments and $\mathbb{L}^p$ errors of the kernel estimator (see [8] and [9]).

During the last years, many articles have been devoted to the study of regression models with functional covariate. Various estimation methods depending on the nature (linear, single index, nonparametric, ...) of the regression operator

have been proposed (see for instance [21], [6],[1], [20], ...). In multivariate statistics, there are plenty of tests (see for instance [18], [4], [17], ...) proposed to check if the regression function have a specific nature (parametric, semi-parametric, ...). However there exist few results focussing on structural tests in the case of a functional covariate. The existing litterature seems reduced to structural tests in linear regression models (see [2] and [3]), no effect tests based on projection (see [16]) and an heuristic goodness-of-fit test proposed by [5]. We construct a test statistic based on the comparison between the nonparametric kernel estimator and a particular one that must be chosen in accordance to the null hypothesis we want to test. We now assume that the dataset is composed of independent random variables and focus on structural testing procedures in regression on functional data. We then state both asymptotic normality of our test statistic under the null hypothesis and its divergence under the alternative (see [10] and [12]). Moreover, we prove our result under general conditions that enable to use our approach to construct innovative structural tests allowing to check if the regression operator is linear, single index, ... Different bootstrap procedures are proposed and compared through various simulation studies.

Bibliographie

[1] Ait-Saïdi, A., Ferraty, F., Kassa, R. et Vieu, P. (2008) Cross-validated estimations in the single functional index model. *submitted*

[2] Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003) Testing Hypothesys in the Functional Linear Model *Scandinavian Journal of Statistics* **30** 241-255.

[3] Cardot, H., Goia, A. et Sarda, P. (2004) Testing for no effect in functional linear regression models, some computational approaches. *Comm. Statist. Simulation Comput.* **33** (1) 179-199.

[4] Chen, S.X. and Van Keilegom I. (2006) A goodness-of-fit test for parametric and semiparametric models in muliresponse regression. *Inst de Statistique, U.C.L., Discussion paper* **0616**

[5] Chiou, J.M. and Müller H.-G. (2007) Diagnostics for functional regression via residual processes *Computational Statistics & Data Analysis* **51**, (10) 4849-4863.

[6] Crambes, C., Kneip, A. and Sarda, P. (2008) Smoothing splines estimators for functional linear regression *Annals of Stat.*.

[7] Dabo-Niang, S. and Rhomari, N. (2003) Estimation non paramétrique de la régression avec variable explicative dans un espace métrique. (French. English, French summary) [Kernel regression estimation when the regressor takes values in metric space] *C. R. Math. Acad. Sci. Paris* **336** (1) 75-80.

[8] Delsol, L. (2007a) CLT and $\mathbb{L}^q$ errors in nonparametric functional regression *C. R. Math. Acad. Sci.* **345** (7) 411-414.

[9] Delsol, L. (2007b) Régression non-paramétrique fonctionnelle : Expressions asymptotiques des moments *Annales de l'I.S.U.P.* **LI** (3) 43-67.

[10] Delsol, L. (2008a) Tests de structure en régression sur variable fonctionnelle. *C. R. Math. Acad. Sci. Paris* **346** (5-6) 343-346.

[11] Delsol, L. (2008b) Advances on asymptotic normality in nonparametric functional Time Series Analysis *Statistics* (à paraître)

[12] Delsol, L., Ferraty, F., Vieu, P. (2008) Structural test in regression on functional variables (submitted)

[13] Ferraty, F., Mas, A. and Vieu, P. (2007) Advances on nonparametric regression for fonctionnal data. *ANZ Journal of Statistics In print.*

[14] Ferraty, F. and Vieu, P. (2000) Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés *Compte Rendus de l'Académie des Sciences* Paris, **330**, 403-406.

[15] Ferraty F. and Vieu P. (2006a) *Nonparametric modelling for functional data.* Springer-Verlag, New York.

[16] Gadiaga, D. and Ignaccolo, R.(2005) Test of no-effect hypothesis by nonparametric regression. *Afr. Stat.* **1** (1) 67-76.

[17] González-Manteiga, W., Quintela-del-Río, A. and Vieu, P. (2002) A note on variable selection in nonparametric regression with dependent data *Statistics and Probability Letters* **57** 259-268.

[18] Härdle, W. and Mammen, E. (1993) Comparing Nonparametric Versus Parametric Regression Fits *The Annals of Statistics* **21**, (4) 1926-1947.

[19] Masry, E. (2005) Nonparametric regression estimation for dependent functional data : asymptotic normality *Stochastic Process. Appl.* **115** (1) 155-177.

[20] Preda, C. (2007) Regression models for functional data by reproducing kernel Hilbert spaces methods. *J. Statist. Plann. Inference* **137** (3) 829-840.

[21] Ramsay, J. and Dalzell, C. (1991) Some tools for functional data analysis *J. R. Statist. Soc. B.* **53** 539-572.

Rank Tests and Regression Rank Scores Tests in Measurement Error Models

PICEK Jan

* Adresse pour correspondance :
Technical University of Liberec
Department of Applied Mathematics
Studentska 2
461 17 Liberec
Czech Republic
e-mail : jan.picek@vslib.cz

Résumé

The rank and regression rank score tests of linear hypothesis in the linear regression model are modified for measurement error models. In some situations the modified tests are still distribution free ; their asymptotic relative efficiencies with respect to tests in model without errors are evaluated. The power of the tests is illustrated on the simulated data.

Some recent results for stationary processes via their associated measures

CABRAL Emmanuel N.

Adresse pour correspondance :
Laboratoire de Statistique et Probabilités, UMR CNRS C55830
118, route de Narbonne
31062 Toulouse cedex.
e-mail : cabral@cict.fr

Abstract

Let $(X_g)_{g \in G} \subset H$ be a complex Hilbertian stationary process where the index space G is a locally compact Abelian group (with dual space $\widehat{G}$ admitting a countable basis). In particular G may be anyone of the sets $\mathbb{Z}$, $\mathbb{Z}^k$, $\mathbb{R}$, $\mathbb{R}^k$, $\Pi = [-\pi, \pi[$, or Π^k and this approach permits considering jointly discrete and continuous-time processes. To any stationary process $(X_g)_{g \in G}$, we know (see, e.g. Azencott and Dacunha-Castelle, 1984), by the Fourier transform, we may associated a *random measure (r.m.)* which is denoted by Z. This measure is unique and we also may associate to this r.m. a unique *spectral measure (s.m.)* denoted by ε. To this s.m. defined on $\mathcal{B}_{\widehat{G}}$, the Borel field of $\widehat{G}$, for the complex Hilbertian space H, we may associate a unique unitary operator denoted by U, called *the unitary operator deduced from* ε, and conversely.

With these tools associated to a stationary process, we propose an example of discrete case where the index set G is $\mathbb{Z}$ and $(X_g)_{g \in G}$ is such that $X_g = \sum_{j \in J} e^{i\lambda_j g} Z_j$, $\|J\| < \infty$, for all $g \in \mathbb{Z}$.

Under the previous assumption, results for the tensor product of discrete and continuous-time processes are presented (see also Birman et al., 1996 ; Boudou et al., 2002). They concern first the convolution product of two spectral measures and secondly the convolution product of random measures. The obtained results are applied to stationary processes statistics with both following examples :

(i) An interpolation problem : given a stationary series $(X_n)_{n \in \mathbb{Z}}$, how can we define all stationary series $(Y_n)_{n \in \mathbb{Z}}$ such as $Y_{nq} = X_n$, $q \in \mathbb{N}^*$?

(ii) An inverse Fourier transform problem : let $(X_n Y_n)_{n \in \mathbb{Z}}$ be a multiplicatively decomposable series, how can we define the random measure whose Fourier transform is the considered series ?

References

Azencott R., Dacunha-Castelle D. (1984). *Séries d'observations irrégulières. Modélisation et prévision.* Techniques Stochastiques, Masson.
Birman M., Solomjak M. (1996). Tensor product of a finite number of spectral measures is always a spectral measure. *Integ. Eq. Oper. Th.*, 24, 179-187.
Boudou A. (2003). Interpolation de processus stationnaires. *C.R. Acad. Sci. Paris*, SérieI, Math., 336, n°12, 1021-1024.
Boudou A. (2007). Groupe d'opérateurs unitaires déduit d'une mesure spectrale - une application. *C.R. Acad. Sci. Paris*, SérieI, Math.,344, n°12, 791-794.
Boudou A., Romain Y. (2002). On spectral and random measures associated to continuous and discrete time processes. *Stat. Proba. Letters*, 59, 145-157.

Selected aspects of functional estimation and testing
(**Soutenance de thèse de Doctorat**)

PRCHAL Luboš

* Adresse pour correspondance :
Université Charles de Prague
Faculté de Mathématiques et de Physique
Département de Statistique
Sokolovská 83
18675 Prague 8, République Tchèque
e-mail : lubos.prchal@mff.cuni.cz

Abstract

Various aspects of statistical modelling, estimating and testing for random samples of curves are the key interest presented in this Thesis. Studied models and newly proposed estimators and methods were motivated by and applied to data from real-life experiments. On one hand, this thesis comprises functional regression models, where a response variable is of a functional nature and a predictor is either scalar or functional. On the other hand, statistical analysis of ROC curves forms the second pillar of the presented work.

First, we investigate the functional linear model with a functional response and a functional predictor as a natural extension of the functional linear model with a scalar response. We consider a common setting for regression and autoregression alternatives of the model and we face the estimation of a bivariate functional parameter. Conditions for existence and uniqueness of the parameter are given and an estimator based on a B-splines expansion is proposed using the penalized least squares method. Some convergence results concerning the error of prediction are derived. This model serves to analyze a real data set describing electricity consumption in Sardegna. The interest lies in predicting either oncoming weekend or oncoming weekdays consumption provided actual weekdays consumption is known. This study verifies good predictive behavior of the model as well as practical advises concerning computational aspects of the estimation.

Our attention is then turned to statistical procedures enabling to check whether a real-valued covariate X has an effect on a functional response $Y(t)$. A nonparametric kernel regression is considered to estimate the influence of X on

$Y(t)$ and two test statistics based on residual sums of squares and smoothing residuals are proposed. Their acceptance levels are determined by means of permutations. The lack-of-fit test for a class of parametric models is then discussed as a consequence of the no effect procedure. Monte Carlo simulations provide an insight into the level and the power of the no effect tests. Modelling of changes in atmospheric radiation within the last decade illustrates the behavior of the proposed methods in practice. Parametric functional multiplicative regression models are provided and its adequacy is tested using the proposed permutation procedure.

A real-data example from computational linguistics motivates our interest in statistical analysis of ROC curves. The first part devoted to ROC curves aims at adapting the standard kernel ROC curve smoother. A data-driven procedure based on automatic monotone data transformation is suggested to improve its accuracy. Box-Cox and general B-splines transformations are discussed. Some insight into the performance of the proposed method is provided by the means of Monte Carlo simulations.

Finally, the problem of testing equivalence of two ROC curves is addressed and illustrated on a real data set from computational linguistics. A transformation of ROC curves is suggested so that it motivates a test statistic as a distance of two empirical quantile processes. Its asymptotic distribution is obtained and a simulation scheme for critical values is proposed. The procedure is applied on several ROC curves measuring quality of automatic collocation extraction. It is shown that obtained p-values can be used as a distance between the curves enabling ROC curves clustering.

Résumé

Cette thèse porte sur plusieurs aspects de la modélisation statistique, de l'estimation et de procédures de test pour des échantillons aléatoires de courbes. Les modèles étudiés et les méthodes proposées ont été motivés en partie par des données réelles. D'un coté, nous nous intéressons à des modèles de régression fonctionnelle où la variable d'intérêt est fonctionnelle et le prédicteur est soit fonctionnel soit réel. D'un autre côté, l'analyse statistique des courbes ROC constitue le second pillier du travail présenté.

Dans la première partie, on considère un modèle de régression lineaire fonctionelle où les deux variables (réponse et prédicteur) sont fonctionnelles. On envisage l'estimation du paramètre fonctionnel bivarié de ce modèle. Tout d'abord, des conditions pour l'identifiabilité du modèle (existence et unicité) sont établies. L'estimateur proposé est de type "splines de régression" basé sur des fonctions B-splines : il s'obtient par minimisation d'un critère de type moindres carrés pénalisés. On obtient pour cet estimateur des résultats de convergence concernant l'erreur de prédiction. Ce modèle nous sert à analyser des courbes hebdomadaires de consommation d'électricité en Sardaigne. Il s'agit de prévoir la consommation pour les jours ouvrables d'une semaine ou celle d'un week-end à partir de la consommation de la semaine précédente. Concernant les aspects pratiques, nous nous sommes attachés à fournir des algorithmes de calcul simples dont on a vérifié l'efficacité sur des simulations et sur les données d'électricité.

On s'intéresse ensuite à des procédures statistiques de test d'influence d'une variable réelle X sur une réponse fonctionelle $Y(t)$. Nous considérons un estimateur à noyau de la régression nonparamétrique de X sur $Y(t)$. Deux statistiques de test basées sur des sommes des résidus au carré et des résidus ajustés sont proposées. Le principe de permutations est adopté pour déterminer le seuil. La procédure est adaptée à un test de type "lack-of-fit" pour une classe de modèles paramétriques. Le comportement du test est analysé sur la base de simulations et de la modélisation de courbes de radioactivité atmosphérique. Des modèles paramétriques de régression fonctionnelle multiplicative sont également proposés et leur adéquation aux données est testé en utilisant la procédure de permutation.

Un problème de discrimination en linguistique motive l'analyse statistique de courbes ROC. Il s'agit de savoir si une paire quelconque de mots consécutifs issus d'un texte forment ou non une collocation en vue de leur traitement par un logiciel de traduction automatique. Les courbes ROC servent, de manière générale, à visualiser les performances du classificateur. On s'intéresse, dans un premier temps, à l'estimateur à noyau d'une corbe ROC. Nous proposons une amélioration de la précision de cet estimateur à l'aide d'une procédure basée sur une transformation monotone automatique des données. Nous discutons deux types de transformations : de type Box-Cox d'une part et B-splines d'autre part.

En pratique, il n'est pas toujours aisé de décider si deux classificateurs sont équivalents à la simple vue de leurs courbes ROC respectives. Il est alors utile

d'avoir une procédure de test. Nous construisons une telle procédure à partir d'une transformation des courbes ROC qui permet d'obtenir une statistique de test sous la forme d'une distance de deux processus quantiles empiriques. Nous obtenons sa loi asymptotique et proposons une procédure de calcul pratique de la valeur critique et discutons de procédures alternatives possibles. Le test est appliqué aux données de linguistique et nous montrons que la p-valeur peut être utilisée dans le but de constituer des "nuages" de courbes ROC.

Membres du jury

Jaromir ANTOCH	Université Charles, Prague	Directeur
Hervé CARDOT	Université de Bourgogne	Examinateur
Frédéric FERRATY	Université Toulouse 2	Examinateur
Daniel HLUBINKA	Université Charles, Prague	Examinateur
Marie HUŠKOVÁ	Université Charles, Prague	Rapporteur
André MAS	Université Montpellier 2	Rapporteur
Jan Picek	Technical University, Liberec	Examinateur
Pascal SARDA	Université Toulouse 2	Directeur

Publications issues de la thèse

– Antoch, J. Hlubinka, D. and Prchal, L. (2006). Statistical methods for analysis of meteorological measurements, *Preprint*.
– Cardot, H., Prchal, L. and Sarda, P. (2007). No effect and lack-of-fit permutation tests for functional regression, *Comp. Statist.*, **22**, 371-390.
– Hlubinka, D. and Prchal, L. (2007). Changes in atmospheric radiation from the statistical point of view. *Comp. Statist. and Data Analysis*, **51**, 4926-4941.
– Antoch, J., Prchal, L. de Rosa, R. and Sarda, P. (2008). Electricity consumption prediction with functional linear regressing using splines estimators, *Preprint*.
– Prchal, L. and Sarda, P. (2008). Spline estimator for the functional linear regression with functional response, *Preprint*.
– Prchal, L. (2008). Data-driven kernel ROC curve estimator for skewed diagnostic variables, *Preprint*.
– Antoch, J., Prchal, L. and Sarda, P. (2008). Nonparametric comparison of ROC curves : Testing equivalence and Clustering, *Preprint*.

Apport de la modélisation fonctionnelle dans le dépistage cytologique du cancer broncho-pulmonaire par imagerie cellulaire

MORLAIS Fabrice *,
CLIN Bénédicte, FAFIN Mélanie, GALATEAU-SALLE Françoise,
HERLIN Paulette, LAUNOY Guy et LETOURNEUX Marc

Adresse pour correspondance :
ERI3 INSERM "Cancers & Populations"
EA 3936 Université de Caen
Unité de Recherche et d'Evaluation en Epidémiologie
Faculté de médecine. Avenue Côte de nacre
14032 Caen cedex
e-mail : fabricemorlais@yahoo.fr

Introduction

En 2005, le cancer du poumon était, d'après les chiffres du réseau français des registres (FRANCIM) et l'Institut national de la santé et de la recherche médicale (INSERM), la première cause de mortalité par cancer en France avec environ 26000 décès par an. Comme tout cancer, l'affection se caractérise par la prolifération tumorale de cellules anormales, initialement localisées mais susceptibles de migrer dans l'organisme sous forme de métastases. Le dépistage actuel du cancer broncho-pulmonaire est basé avant tout sur des techniques radiologiques : la radiographie pulmonaire et le scanner thoracique. Quelques récents travaux (3, 4, 5) ont montré l'intérêt d'une nouvelle technique de dépistage de ces cancers : la cytologie automatisée des expectorations. Il s'agit d'une méthode permettant l'analyse des cellules d'un crachat sur la lame d'un microscope. Une caméra numérique reliée à un ordinateur permet le stockage et l'analyse des images de cette lame par un logiciel d'analyse spécifique. Ainsi pour chaque cellule, 19 paramètres sont mesurés : les principaux étant la taille, la quantité d'ADN et la texture. Une personne peut donc être représentée par la distribution multidimensionnelle des mesures effectuées sur chacune de ses cellules. Chaque axe dimensionnel représente la distribution des cellules de la lame pour un paramètre cellulaire donné (taille, texture, ?).

Problématique

Les travaux réalisés jusqu'à présent pour analyser cette technique se basent sur les arbres de décisions ou sur des taux de cellules suspectes. Le problème de ces analyses se situent dans le fait qu'elles ne prennent pas en compte toute l'information importante contenue dans les distributions cellulaires et quelles sont totalement dépendantes de l'observation de cellules suspectes (ce qui en pratique reste extrêmement rare et sujet à beaucoup d'aléas). La théorie des MACs (1,2) stipule qu'un changement de caractéristiques des cellules normales est observé dans des cas de malignité : les cellules « normales » de personnes ayant un cancer auraient des caractéristiques légèrement différentes des cellules de personnes saines. S'agissant de modifications minimes, elles ne peuvent être identifiées ni par l' ?il de l'anatomopathologiste, ni par les méthodes par arbre de décision. La modélisation fonctionnelle (6, 7, 8) peut parfaitement répondre à cette problématique en permettant une observation (ACP de courbes) et une comparaison (Régression logistique fonctionnelle) des distributions des mesures cellulaires de personnes saines et de personnes ayant un cancer broncho-pulmonaire. Chaque individu étant représenté par la distribution de ses caractéristiques cellulaires. Notre étude se basera sur les prélèvements d'expectorations bronchiques d'au moins 30 personnes ayant un cancer dûment confirmé et d'au moins 30 personnes saines

Résultats préliminaires

Même si nos résultats ne sont pas encore consolidés, il apparaît clairement des différences distributionnelles entre les deux types de populations. En effet les distributions moyennes correspondant à la taille des cellules, à leur contenu en ADN et à leur texture s'avèrent assez différentes entre les cas et les témoins : les distributions cellulaires des personnes ayant un cancer semblent plus étalées que les distributions cellulaires des personnes saines. On observe aussi une plus grande hétérogénéité des distributions chez les personnes ayant un cancer que chez les personnes saines.

Discussion

Si ces résultats venaient à être confirmés nous les appliquerions à une cohorte de personnes à risque de cancer broncho-pulmonaire pour lesquelles ont été conservés des échantillons d'expectoration préparés par cytologie automatisée. Nous pourrons ainsi évaluer cette technique de dépistage, en terme de sensibilité et spécificité, pour pouvoir ensuite la comparer aux techniques de dépistage par radiographie pulmonaire et scanner thoracique.

Références

[1] Nieburgs HE. *Diagnostic Cell Pathology in Tissue Sections and Smears.* New York : Grune & Stratton, 1967.

[2] Nieburgs HE. Recent progress in the interpretation of malignancy associated changes (MAC). *Acta Cytol* 1968 ; 12 :445-453.

[3] Palcic B, Garner DM, Beveridge J, et al. Increase of sensitivity of sputum cytology using high-resolution image cytometry : field study results. *Cytometry* 2002 ; 50 :168-176.

[4] Payne PW, Sbebo TJ, Doudkine A, et al. Sputum screening by quantitative microscopy : a reexamination of a portion of the National Cancer Institute Cooperative Early Lung Cancer Study. *Mayo Clin Proc* 1997 ; 72 :697-704.

[5] Belien JAM, Baak JPA, van Diest PJ, Misere BNLHM, Meijer GA, Bergers L. Prognostic value of image and flow cytometric DNA ploidy assessments in invasive breast cancer. *Electr. J. Pathol.* 1997 ;3 ;972-979.

[6] Ramsay JO, Silverman BW. *Applied Functional Data Analysis Methods and Case Studies* 2002.

[7] Ramsay JO, Silverman BW. *ctional Data Analysis Second Edition* 2005.

[8] Ferraty F, Vieu P. *Non Parametric Functional Data Analysis Theory and Practice* 2006.

.

Régression sur variable fonctionnelle : Estimation, Tests de structure et Applications

(**Soutenance de thèse de Doctorat**)

DELSOL Laurent

* Adresse pour correspondance :
Institut de Mathématiques de Toulouse
Université Paul Sabatier
Toulouse
delsol@cict.fr

Résumé

Au cours des dernières années, la branche de la statistique consacrée à l'étude de variables fonctionnelles a connu un réel essor tant en terme de développements théoriques que de diversification des domaines d'application. Nous nous intéressons plus particulièrement dans ce mémoire à des modèles de régression dans lesquels la variable réponse est réelle tandis que la variable explicative est fonctionnelle, c'est à dire à valeurs dans un espace de dimension infinie.

Les résultats que nous énonçons sont liés aux propriétés asymptotiques de l'estimateur à noyau généralisé au cas d'une variable explicative fonctionnelle. Nous supposons pour commencer que l'échantillon que nous étudions est constitué de variables α-mélangeantes et que le modèle de régression est de nature non-paramétrique. Nous établissons la normalité asymptotique de notre estimateur et donnons l'expression explicite des termes asymptotiquement dominants du biais et de la variance. Une conséquence directe de ce résultat est la construction d'intervalles de confiance asymptotiques ponctuels dont nous étudions les propriétés aux travers de simulations et que nous appliquons sur des données liées à l'étude du courant marin El Niño. On établit également à partir du résultat de normalité asymptotique et d'un résultat d'uniforme intégrabilité l'expression explicite des termes asymptotiquement dominants des moments centrés et des erreurs $\mathbb{L}^p$ de notre estimateur.

Nous considérons ensuite le problème des tests de structure en régression sur variable fonctionnelle et supposons maintenant que l'échantillon est com-

posé de variables indépendantes. Nous construisons une statistique de test basée sur la comparaison de l'estimateur à noyau et d'un estimateur plus particulier dépendant de l'hypothèse nulle à tester. Nous obtenons la normalité asymptotique de notre statistique de test sous l'hypothèse nulle ainsi que sa divergence sous l'alternative. Les conditions générales sous lesquelles notre résultat est établi permettent l'utilisation de notre statistique pour construire des tests de structure innovants permettant de tester si l'opérateur de régression est de forme linéaire, à indice simple, ... Différentes procédures de rééchantillonnage sont proposées et comparées au travers de diverses simulations. Nos méthodes sont enfin appliquées dans le cadre de tests de non effet à deux jeux de données spectrométriques.

Publications issues de la thèse

- Delsol, L. (2007a) CLT and $\mathbb{L}^q$ errors in nonparametric functional regression *C. R. Math. Acad. Sci.* **345** (7) 411-414.
- Delsol, L. (2007b) Régression non-paramétrique fonctionnelle : Expressions asymptotiques des moments (in French) [Nonparametric functional regression : Asymptotic expressions of moments] *Annales de l'I.S.U.P.* **LI** (3) 43-67.
- Delsol, L. (2008a) Tests de structure en régression sur variable fonctionnelle. (in French) [Structural tests in regression on functional variable] *C. R. Math. Acad. Sci. Paris* **346** (5-6) 343-346.
- Delsol, L. (2008b) Advances on asymptotic normality in nonparametric functional Time Series Analysis *Statistics* (In press)
- Delsol, L., Ferraty, F., Vieu, P. (2008) Structural test in regression on functional variables (submitted).