
GROUPE DE TRAVAIL STAPH :
STATISTIQUE FONCTIONNELLE ET OPÉRATORIELLE 7

Recueil de résumés 2005-2006

Coordinateurs

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

TABLE DES MATIERES

Introduction.	5
Alain BOUDOU : Proximité de fonctions aléatoires stationnaires et commutativité	7
Christophe CRAMBES : Modèle Linéaire Fonctionnel dans le cas d'une variable explicative bruitée	9
Ali LAKSACI : Contribution aux modèles nonparamétriques conditionnels pour variables fonctionnelles	13
Antonio CUEVAS : Some topics on resampling, clustering and classification with functional data	15
Christine DELEUZE-BREZINS : Utilisation des analyses proches Infra-Rouge (NIR) dans le domaine papetier : quelles perspectives ?	17
Célestin C. KOKONENDJI : Fonction poids de Poisson relative à sur et sousdispersion	19
Magali JAOUL-GRAMMARE : Techniques de cointégration et causalité à la Granger pour un test économétrique en économie de l'éducation	23
Anne-Séverine BOUDOU : Inégalités Fonctionnelles pour des Systèmes de Particules	25
Luboš PRCHAL : Approche statistique de problème de rupture dans des courbes de radioactivité atmosphérique : Modèle Linéaire Fonctionnel dans le cas d'une variable explicative bruitée	27
Michal BENKO : Common functional principal components	29
Marie Josée MARTINEZ : Modèle de mélange pour données répétées de loi exponentielle	31
G.A. KOSHEVOY : Tests for independence based on zonotopes	33
Rappels des activités antérieures de STAPH	35

¹ STAPH : Bilan d'activités 2005-2006

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

Coordinateurs du groupe de travail STAPH
Laboratoire de Statistique et Probabilités

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

L'objectif du groupe de travail STAPH au sein du LSP est d'alimenter et dynamiser les différents aspects fonctionnels et opératoires de la statistique qui recouvrent notamment le lissage de données et la statistique dans les espaces de grandes dimensions. Ces thématiques occupent désormais une place importante dans la recherche en statistique : les thèmes les plus actuels dans ce domaine ont trait à la modélisation statistique pour variables fonctionnelles. En effet, les appareillages modernes de mesure ainsi que les moyens informatiques permettent le recueil de données volumineuses sur des grilles dont la finesse conduit à considérer des objets fonctionnels. Les méthodes classiques de la statistique multidimensionnelle ont montré leur limite dans l'étude de telles données. Il est donc nécessaire de développer des méthodes *ad hoc* au travers de modélisations prenant en compte la richesse de l'information ainsi acquise. Cela engendre une gamme de problématiques englobant les grands thèmes de la statistique et par ailleurs un besoin d'interdisciplinarité. Les aspects théoriques s'articulent principalement autour d'interactions fortes entre statistique, probabilités dans des espaces fonctionnels, optimisation, algèbre, analyse fonctionnelle et informatique. Parallèlement, ces axes de recherche débouchent sur un large champ d'applications (imagerie, télédétection, météorologie, médecine, reconnaissance vocale, spectrométrie, ...).

Les travaux des membres du groupe de travail STAPH s'inscrivent dans ces objectifs. Pour une part, ils couvrent des domaines théoriques et/ou appliqués qui se traduisent par de nombreuses publications. Ils ont pour cadre des collaborations avec d'autres Universités : Bordeaux 2, Montpellier 2, Grenoble (Pierre Mendès-France), Lille 3, les Universités de Béjaia et Sidi-bel-Abbès (Algérie), Novarra/Turin (Italie), Alicante, Bilbao, Madrid, La Corogne, Grenade, Santander, Santiago (Espagne), Berlin, Mayence (Allemagne), Brno, Prague (Tchéquie),

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.

Montréal (Canada), Bergen (Norvège), Une autre part a trait à la création et la diffusion de programmes (logiciels), à l'étude de cas, ... Ce dernier aspect alimente des coopérations avec le CNES, Météo-France, l'Observatoire Midi-Pyrénées,...

Egalement, ce groupe de travail a pour vocation l'encadrement de jeunes chercheurs : 7 thèses et habilitations ont été soutenues depuis 2000 et 7 autres sont en préparation.

Les travaux réalisés ont permis d'avancer dans la compréhension de ces modèles fonctionnels tout en mettant en lumière de nombreux problèmes ouverts, tant théoriques qu'appliqués. C'est dans un souci d'interdisciplinarité que nous poursuivrons et développerons nos activités.

Toulouse, Juillet 2006

Proximité de fonctions aléatoires stationnaires et commutativité

Alain BOUDOU *

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

SÉANCE DU 17 OCTOBRE, 16H, SALLE 214

Résumé

Etant donné deux séries stationnaires $(X_n^1)_{n \in \mathbb{Z}}$ et $(X_n^2)_{n \in \mathbb{Z}}$ telles que l'on ait $\|X_n^1 - X_n^2\| \leq \varepsilon$, pour tout n de \mathbb{Z} , il est légitime de penser que les mesures aléatoires qui leur sont associées (c'est-à-dire les mesures aléatoires dont elles sont transformées de Fourier) possèdent une propriété de proximité analogue. Nous parvenons à établir un tel résultat lorsque les opérateurs de décalage, d'une façon plus précise les opérateurs $X_n^1 \mapsto X_{n+1}^1$, commutent. On parle alors de processus stationnaires qui commutent. Ces résultats peuvent être étendus au cas où l'ensemble d'indiciage est un groupe abélien localement compact G pour lequel :

il existe un élément α de \mathbb{R}_+^ tel que si la fonction aléatoire continue stationnaire $(X_g)_{g \in G}$, de mesure aléatoire associée Z , prend ses valeurs dans la boule centrée en h de rayon ε alors nous avons : $\|Z(\{0\}) - h\| \leq \alpha\varepsilon$. Outre les groupes \mathbb{Z} et \mathbb{R} , pour lesquels le coefficient α est égal à 1, $\mathbb{R}^k, \mathbb{Z}^k, \Pi$ et Π^k possèdent cette propriété.*

Pour réaliser ce travail nous avons utilisé comme outil la mesure spectrale, au sens de la théorie des opérateurs, le produit de convolution de mesures spectrales tel que celui-ci est défini en Boudou-Romain, le groupe des opérateurs unitaires déduit d'une mesure spectrale.

Références

Arveson, W., 2001. A Short Course On Spectral Theory. Springer, New York.

Boudou, A., Romain, Y., 2002. On spectral and random measures associated to discrete and continuous-time processes. *Statist. Probab. Lett.* 59 145-157.

Rozañov, Yu. A., 1982. Markov Random Fields. Springer, New York.

Rudin, W., 1967. Fourier Analysis On Groups. Wiley-Inter science, New York.

Modèle Linéaire Fonctionnel dans le cas d'une variable explicative bruitée

Christophe CRAMBES *

* Adresse pour correspondance :

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

SÉANCE DU 24 OCTOBRE, 16H, SALLE 214

Résumé

Un des problèmes usuels en statistique est d'expliquer les effets d'une variable X déterministe sur une variable aléatoire Y (cf. par exemple Cuevas *et. al.*, 2002, Frank et Friedman, 1993). On considère ici une variable réelle Y , alors que la variable X sera supposée appartenir à un espace fonctionnel, en l'occurrence $L^2(I)$, l'espace des fonctions de carré intégrable sur I où I est un intervalle de \mathbb{R} . On se place ici dans le cadre du *Modèle Linéaire Fonctionnel* (cf. Ramsay et Dalzell, 1991, Ramsay et Silverman, 1997), c'est-à-dire que l'on considère le modèle

$$Y = \int_I \alpha(t)X(t)dt + \epsilon, \quad (1)$$

où $\alpha \in L^2(I)$ est inconnue et ϵ est une variable aléatoire réelle telle que $\mathbb{E}(\epsilon) = 0$ et $\mathbb{E}(\epsilon^2) = \sigma_\epsilon^2$. Disposant d'un échantillon $(X_i, Y_i)_{i=1, \dots, n}$, de nombreux estimateurs de la fonction α peuvent être construits (voir par exemple Cardot *et. al.*, 2003). On propose dans un premier temps un estimateur construit en utilisant une approche provenant des splines de lissage (cf. Wahba, 1990). Considérant une discrétisation $t_1 < \dots < t_p$ de I (équirépartie pour simplifier), on cherche l'estimateur de $\boldsymbol{\alpha} = (\alpha(t_1), \dots, \alpha(t_p))^T \in \mathbb{R}^p$ solution du problème de minimisation suivant

$$\min_{\boldsymbol{a} \in \mathbb{R}^p} \left\{ \frac{1}{n} \left\| \mathbf{Y} - \frac{1}{p} \mathbf{X} \boldsymbol{a} \right\|^2 + \frac{\rho}{p} \boldsymbol{a}^T \mathbf{A}_m \boldsymbol{a} \right\}, \quad (2)$$

où $\|\cdot\|$ désigne la norme usuelle de \mathbb{R}^p , ρ est un paramètre de lissage, $\mathbf{Y} = (Y_1, \dots, Y_n)^\tau$, \mathbf{X} est la matrice $n \times p$ de terme général $X_i(t_j)$ et $\frac{1}{p}\mathbf{a}^\tau \mathbf{A}_m \mathbf{a}$ est la version discrétisée de $\int_I a^{(m)}(t)^2 dt$. La résolution de ce problème de minimisation permet d'obtenir l'écriture explicite de l'estimateur de $\boldsymbol{\alpha}$, donnée par

$$\widehat{\boldsymbol{\alpha}}_{FLS,X} = \frac{1}{np} \left(\frac{1}{np^2} \mathbf{X}^\tau \mathbf{X} + \frac{\rho}{p} \mathbf{A}_m \right)^{-1} \mathbf{X}^\tau \mathbf{Y}. \quad (3)$$

Un résultat de convergence est alors établi pour cet estimateur.

Dans un deuxième temps, revenant au modèle (1), on constate qu'il fait l'hypothèse sous-jacente que la variable X est observée sans erreur, toutes les erreurs étant confinées dans la variable Y par le biais de ϵ . Cette hypothèse peut se révéler peu réaliste en pratique et des erreurs peuvent aussi affecter la variable X . C'est pourquoi on peut naturellement considérer que la variable X n'est pas directement observée, mais que la variable dont on dispose en réalité est $W = X + \delta$. Plus précisément, on a en chaque point de mesure t_1, \dots, t_p ,

$$W(t_j) = X(t_j) + \delta(t_j), \quad (4)$$

où $(\delta(t_j))_{j=1, \dots, p}$ est une suite de variables aléatoires indépendantes identiquement distribuées telle que $\mathbb{E}(\delta(t_j)) = 0$ et $\mathbb{E}(\delta(t_j)^2) = \sigma_\delta^2$. Dans ce contexte, le but est à nouveau de proposer un estimateur de α sur la base d'un échantillon $(W_i, Y_i)_{i=1, \dots, n}$. On adapte pour cela la méthode des *Total Least Squares* (cf. par exemple Golub et Van Loan, 1980, Van Huffel et Vandewalle, 1991) à ce cadre fonctionnel. On cherche l'estimateur de $\boldsymbol{\alpha} \in \mathbb{R}^p$ solution du problème de minimisation

$$\min_{\boldsymbol{\alpha}, \mathbf{X}_i \in \mathbb{R}^p} \left\{ \frac{1}{n} \sum_{i=1}^n \left[\left(Y_i - \frac{1}{p} \mathbf{X}_i^\tau \boldsymbol{\alpha} \right)^2 + \frac{1}{p} \|\mathbf{X}_i - \mathbf{W}_i\|^2 \right] + \frac{\rho}{p} \mathbf{a}^\tau \mathbf{A}_m \mathbf{a} \right\}, \quad (5)$$

où \mathbf{X}_i et \mathbf{W}_i désignent les vecteurs de \mathbb{R}^p de termes généraux respectifs $X_i(t_j)$ et $W_i(t_j)$. On obtient là encore une écriture de l'estimateur de $\boldsymbol{\alpha}$, donnée par

$$\widehat{\boldsymbol{\alpha}}_{FTLS} = \frac{1}{np} \left(\frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} + \frac{\rho}{p} \mathbf{A}_m - \frac{\sigma_\delta^2}{p^2} \mathbf{I}_p \right)^{-1} \mathbf{W}^\tau \mathbf{Y}, \quad (6)$$

où \mathbf{W} désigne la matrice $n \times p$ de terme général $W_i(t_j)$ et \mathbf{I}_p désigne la matrice identité de taille p . Dans le cas où σ_δ n'est pas connu, on l'estime de façon nonparamétrique comme dans Gasser *et. al.* (1986). Un résultat de convergence est ensuite établi pour cet estimateur $\widehat{\boldsymbol{\alpha}}_{FTLS}$.

Dans un dernier temps, cet estimateur *FTLS* est étudié par le biais de simulations. On regarde notamment en quoi l'apport d'une "dérégularisation" $-\frac{\sigma_\delta^2}{p^2} \mathbf{I}_p$ dans (6) permet d'apporter une amélioration à l'estimation.

Références

- Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline Estimators for the Functional Linear Model. *Statistica Sinica*, **13**, 571-591.
- Cuevas, A., Febrero, M. and Fraiman, R. (2002). Linear Functional Regression : the case of a Fixed Design and Functional Response. *Canadian Journal of Statistics*, **30**, 285-300.
- Frank, I.E. and Friedman, J.H. (1993). A Statistical View of some Chemometrics Regression Tools. *Technometrics*, **35**, 109-135.
- Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual Variance and Residual Pattern in Nonlinear regression. *Biometrika*, **3**, 625-633.
- 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.
- 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.
- Ramsay, J.O. and Silverman, B.W. (1997). *Functional Data Analysis*. Springer-Verlag.
- Van Huffel, S. and Vandewalle, J. (1991). *The Total Least Squares Problem : Computational Aspects and Analysis*. SIAM, Philadelphia.
- Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.

Contribution aux modèles nonparamétriques conditionnels pour variables fonctionnelles

Ali LAKSACI

* Adresse pour correspondance :
Université Paul Sabatier
Laboratoire Statistique et Probabilités
Toulouse, France
e-mail : alilak@yahoo.fr ou laksaci@cict.fr

SÉANCE DU 28 NOVEMBRE, 10H, SALLE 15 (Bat1R2)
Soutenance de thèse de Doctorat

Résumé

Dans cette thèse, nous nous proposons d'étudier le problème de la modélisation non paramétrique lorsque les données statistiques sont des courbes. Précisément, nous nous intéressons à des problèmes de prévision à partir d'une variable explicative à valeurs dans un espace de dimension infinie, et nous cherchons à développer des alternatives à la méthode de régression. En effet, on suppose qu'on dispose d'une variable aléatoire réelle (réponse), notée Y , et d'une variable fonctionnelle (explicative), notée X . Le modèle non paramétrique utilisé pour étudier le lien entre X et Y concerne la distribution conditionnelle dont la fonction de répartition (resp. densité), notée F (resp. f), est supposée appartenir à un espace fonctionnel approprié.

Dans un premier temps, on considère une suite d'observations i.i.d. Dans ce contexte, nous construisons des estimateurs par la méthode du noyau pour la fonction de répartition conditionnelle, la densité conditionnelle et ses dérivées. On établit la vitesse de convergence presque complète de ces estimateurs. On déduit des estimateurs précédents ceux pour estimer le mode conditionnel et les quantiles conditionnels, pour lesquels, on donne la vitesse de convergence presque complète.

Dans un second temps, nous supposons que les observations sont fortement

mélangeantes et nous nous fixons comme objectif l'estimation du mode conditionnel. Nous étudions les propriétés asymptotiques de cet estimateur, en donnant l'expression de sa vitesse de convergence. Ce résultat peut être utilisé pour le problème de la prévision en série chronologique.

Notre étude met en évidence le phénomène de concentration de la mesure de probabilité de la variable fonctionnelle sur des petites boules. Plus précisément, des hypothèses portant sur les probabilités de petites boules nous permettent de proposer une solution originale au problème du fléau de la dimension et ainsi de généraliser à la dimension infinie de nombreux résultats asymptotiques existant dans le cas multivarié. De plus, en utilisant les nombreux résultats récents en théorie des probabilités sur les petites boules, on précise nos résultats pour de nombreux processus à temps continu.

L'originalité de cette thèse est qu'elle aborde et développe aussi bien des aspects pratiques que théoriques. Nos méthodes sont appliquées à des données réelles de type spectrométrique ou de pollution.

Some topics on resampling, clustering and classification with functional data

Antonio CUEVAS*, **Manuel FEBRERO** and **Ricardo FRAIMAN**

* Adresse pour correspondance :

Antonio Cuevas, Departamento de Matemáticas, Facultad de Ciencias,
Universidad Autónoma de Madrid, 28049-Madrid, Spain
e-mail : antonio.cuevas@uam.es

SÉANCE DU 28 NOVEMBRE, 16H, SALLE 214

Résumé

The bootstrap methodology for functional data and functional estimation target is considered. A Monte Carlo study analyzing the performance of the bootstrap confidence bands (obtained with different resampling methods) of several functional estimators is presented. Some of these estimators (e.g., the trimmed functional mean) are defined via some depth notions for functional data and do not have received yet much attention in the literature. A real data example in cardiology research is also analyzed. In a more theoretical aspect, a brief discussion is given providing some insights on the asymptotic validity of the bootstrap methodology when functional data, as well as a functional parameter, are involved. These results are included in a recent manuscript by Cuevas Febrero and Fraiman (CFF, 2005).

Some additional ideas, still not completely developed, are also presented in connection with clustering and classification procedures for functional data. They rely, respectively, on the use of infinite-dimensional level sets and on some new concepts of functional depth. See CFF (2000, 2001) and Fraiman and Muniz (2001) for closely related references.

Références

Cuevas, A., Febrero, M. and Fraiman, R. (2000). Estimating the number of clusters. *Canadian Journal of Statistics*, 28, 367-382.

Cuevas, A., Febrero, M. and Fraiman, R. (2001). Cluster analysis : a further approach based on density estimation. *Computational Statistics and Data Analysis*, 36 441-459.

Cuevas, A., Febrero, M. and Fraiman, R. (2005). On the use of the bootstrap for estimating functions with functional data. *Computational Statistics and Data Analysis* (to appear).

Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *Test*, 10, 419-440.

Utilisation des analyses proches Infra-Rouge (NIR) dans le domaine papetier : quelles perspectives ?

**Christine DELEUZE-BREZINS*, Denilson DA SILVA PEREZ,
Alain BOUVET**

* Adresse pour correspondance :
AFOCEL
Station territoriale Nord-Est
Route de Bonnencontre, 21170 Charrey-sur-Saône
FRANCE
e-mail : deleuze@afocel.fr

SÉANCE DU 5 DECEMBRE, 16H, SALLE 214

Résumé

L'AFOCEL est un organisme de recherche de statut privé (Association, loi 1901), dont la mission est de contribuer à la compétitivité durable du système forêt-bois-papier en France, face à une ressource forestière particulièrement riche et complexe. Son laboratoire Bois-Process a en particulier la tâche d'aider les industriels de la pâte et du papier à mieux tirer parti de la diversité des ressources forestières françaises dans leurs process et produits. Beaucoup de projets visent donc à étudier la variabilité des propriétés de la pâte en fonction de la ressource forestière. Ces propriétés sont actuellement étudiées en réalisant des tests souvent longs et coûteux en laboratoire. C'est pourquoi l'AFOCEL s'est tournée vers la spectroscopie NIR pour estimer ces propriétés, aussi bien sur les copeaux de bois, la pâte ou le papier.

L'objectif de cette présentation est de montrer la richesse des données récoltées, de présenter le type d'analyses effectuées et de débattre sur la possibilité d'utiliser des outils plus pointus pour tirer profit de ces données et de cette technologie.

Une perspective à plus long terme existe aussi dans la mise en place d'outils d'aide à l'analyse pour des NIR installés en ligne chez des industriels.

Fonction poids de Poisson relative à sur- et sousdispersion

Célestin C. KOKONENDJI *

*Adresse pour correspondance :
Université de Pau et des Pays de l'Adour
Laboratoire de Mathématiques Appliquées
Département STID - IUT des Pays de l'Adour
Avenue de l'Université - 64000 Pau, France.
e-mail : celestin.kokonendji@univ-pau.fr

Séance du 12 Décembre 2005

Résumé

Dans de nombreux domaines (économétrie, épidémiologie, environnement, agriculture, finance, ...) surviennent des données ou des séries à valeurs entières. La loi de Poisson a longtemps été considérée comme la distribution "normale" pour les données de comptage. Cependant elle est *equidispersée*, à savoir la variance est égale à la moyenne. Pour beaucoup de données de comptage observées, on constate que la variance empirique est supérieure ou inférieure à la moyenne empirique ; ces phénomènes sont respectivement appelés *surdispersion* et *sousdispersion* par rapport à Poisson. Ce sont des phénomènes bien connus et sont dus à une ou plusieurs causes possibles, telles l'hétérogénéité et l'aggrégation pour la surdispersion et la répulsion pour la sousdispersion bien que moins fréquente. Leurs études nécessitent toujours des modifications de la loi de Poisson. On peut consulter [1]-[4], [7], [14], [17]-[21].

Depuis quelques années, nous nous intéressons aussi à des problèmes reliés à l'analyse et la modélisation des données de comptage (surdispersées principalement) en ayant comme objectif le développement d'outils pouvant être utilisés tout autant par un épidémiologiste que par un statisticien-chercheur. Voir [5], [8]-[13].

Dans cet exposé, basé essentiellement sur [13], nous étudierons les deux phénomènes de surdispersion et de sousdispersion pour les données de comptage à travers le modèle de Poisson pondéré (e.g. [3]-[4]). Il s'agit d'un modèle assez "natrel"

dans la situation des données de comptage (e.g. [6], [15]-[16]). Nous présenterons des propriétés générales sur la *fonction poids* de Poisson conduisant à (voire caractérisant) la surdispersion et la sousdispersion. Pour une meilleure compréhension de ces deux phénomènes, nous introduirons une notion de dualité entre deux distributions de Poisson pondérées de manière que le produit (ponctuel) de leurs fonctions poids donnent l'unité. Des exemples pratiques de la littérature seront présentés pour illustrer les résultats.

Enfin, nous discuterons sur une estimation non-paramétrique de la fonction poids de Poisson en cours de développement, laquelle méthode conduirait à une estimation semi-paramétrique de la densité de la loi Poisson pondérée.

Références

- [1] Bosch, R.J., Ryan, L.M., 1998. Generalized Poisson models arising from Markov processes. *Statist. Probab. Letters* 39, 205-212.
- [2] Cameron, A.C., Johansson, P., 1997. Count data regression using series expansions : with applications. *J. Appl. Econometrics* 12, 203-223.
- [3] Castillo, J., Pérez-Casany, M., 1998. Weighted Poisson distributions for overdispersion and underdispersion situations. *Ann. Inst. Statist. Math.* 50, 567-585.
- [4] Castillo, J., Pérez-Casany, M., 2005. Overdispersed and underdispersed Poisson generalizations. *J. Statist. Plann. Inference* 134, 486-500.
- [5] Dossou-Gbété, S., Kokonendji, C.C., Mizère, D., 2005. Quelques tests de la loi de Poisson contre des alternatives générales basés sur l'indice de dispersion de Fisher. *Rev. Statist. Appl.* (accepté pour publication).
- [6] Fisher, R.A., 1934. The effects of methods of ascertainment upon the estimation of frequencies. *Ann. Eugenics* 6, 13-25.
- [7] Gupta, R.C., Ong, S.H., 2005. Analysis of long-tailed count data by Poisson mixtures. *Commun. Statist.- Theory Meth.* 34 (3), 557-573.
- [8] Jourdan, A., Kokonendji, C.C., 2002. Surdispersion et modèle binomial négatif généralisé. *Rev. Statist. Appl.* 50 (3), 73-86.
- [9] Kokonendji, C.C., 1999. Le problème d'Anscombe pour les lois binomiales négatives généralisées. *Canadian J. Statist.* 27 (1), 199-205.
- [10] Kokonendji, C.C., Demétrio, C.G.B., Dossou-Gbété, S., 2004. Some discrete exponential dispersion models : Poisson-Tweedie and characterization of weighted Poisson distribution. Technical Report No. 0523, LMA-University of Pau (submitted for publication).

- [14] Luceño, A., 2005. Recursive characterization of a large family of discrete probability distribution showing extra-Poisson variation. *Statistics* 39, 261-267.
- [15] Patil, G.P., 2002. Weighted distributions. In : El-Shaarawi, A.H., Piegorisch, W.W. (Eds.), *Encyclopedia of Environmetrics*, Wiley, Chichester, vol. 4, 2369-2377.
- [16] Patil, G.P., Rao, C.R., 1978. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics* 34, 179-184.
- [17] Puig, P., Valero, J., 2005. Count data distributions : some characterizations with applications. *J. Amer. Statist. Assoc.* (to appear).
- [18] Rao, C.R., 1965. On discrete distributions arising out of methods of ascertainment. In : Patil, G.P. (Ed.), *Classical and Contagious Discrete Distributions*, Pergamon Press and Statistical Publishing Society, Calcutta, 320-332.
- [19] Ridout, M.S., Besbeas, P., 2004. An empirical model for underdispersed count data. *Statist. Modelling* 4, 77-89.
- [20] Shmueli, G., Minka, T.P., Kadane, J.P., Borle, S., Boatwright, P., 2005. A useful distribution for fitting discrete data : revival of the Conway-Maxwell-Poisson distribution. *J. Roy. Statist. Soc. Ser. C* 54, 127-142.
- [21] Steutel, F.W., 1985. Log-concave and log-convex distributions. In : Kotz, S., Johnson, N.L., Read, C.B. (Eds.), *Encyclopedia of Statistical Sciences*, Wiley, New York, vol. 5, 116-117.

Techniques de cointégration et causalité à la Granger pour un test économétrique en économie de l'éducation

Magali JAOUL-GRAMMARE *

* Adresse pour correspondance :
LAMETA - Association Française de Cliométrie
Université Montpellier I, Faculté des Sciences Economiques
Espace Richter, Avenue de la Mer, CS 79 606
34960 Montpellier Cedex 2, France.
e-mail : m.jaoul@lameta.univ-montp1.fr

SÉANCE DU 16 JANVIER, 16H, SALLE 214

Résumé

Dans de nombreux pays, le système d'enseignement supérieur, comme le système économique, a connu des phases d'expansion suivies de périodes de stagnation ou de baisse. Aussi, la possibilité d'une liaison entre les deux systèmes a souvent été avancée. Pour Diebolt (2001), ces mouvements sont liés aux mouvements du marché du travail. Dans sa théorie de l'engorgement, il considère que le comportement des étudiants dans leurs choix de carrière dépend des revenus espérés et des disponibilités au cœur du secteur professionnel. L'objet de cet article, est de proposer un test économétrique de cette théorie, pour le cas de l'Allemagne (1820-1941), à l'aide de la cointégration et de la causalité. Les relations mises en évidence dans notre analyse semblent aller dans le sens d'une confirmation de cette théorie : nous observons une substituabilité entre filières universitaires et entre les diverses professions d'une part et une complémentarité entre effectifs universitaires et professions correspondantes, par ailleurs. Toutefois, ces relations sont essentiellement des relations de court terme. En effet, nous rejetons la présence d'une relation de long terme stable entre les variables.

Inégalités Fonctionnelles pour des Systèmes de Particules

Anne-Séverine BOUDOU

* Adresse pour correspondance :
LSP, Toulouse
France
e-mail : boudou@math.unipd.it

SÉANCE DU 30 JANVIER, 16H, SALLE 214

Résumé

Étant donné un processus de Markov $(X_t)_{t \geq 0}$ et une mesure invariante μ de ce processus, nous souhaitons examiner l'éventualité de la convergence de la loi conditionnelle $P(X_t \in \cdot | X_0 = \eta)$ vers μ .

L'étude de l'inégalité de Poincaré et l'étude de l'inégalité logarithmique de Sobolev modifiée sont des approches standards. Rappelons que ces inégalités permettent de donner certaines caractéristiques de la vitesse de convergence à l'équilibre.

Une généralisation de l'approche de Bakry-Emery nous permet de minorer le Spectral Gap, constante définie par l'équation de Poincaré. Jusqu'à présent, cette approche a largement été utilisée pour des processus de diffusion, processus dont les trajectoires sont continues. L'étude de cette approche pour des processus dont les trajectoires ne sont pas continues soulève de nouvelles difficultés. Après avoir présenté un modèle général dans le cas du Spectral Gap, nous l'appliquerons à des modèles particuliers.

Approche statistique de problème de rupture dans des courbes de radioactivité atmosphérique : Modèle Linéaire Fonctionnel dans le cas d'une variable explicative bruitée

Luboš PRCHAL*

* Adresse pour correspondance :

Département de Statistique, Université Charles à Prague
Sokolovská 83, 186 00 Prague 8, République tchèque
&

Laboratoire de Statistique et Probabilités, Université Paul Sabatier,
118, route de Narbonne, 31062 Toulouse Cedex, France

e-mail : prchal@cict.fr

SÉANCE DU 6 FEVRIER, 16H, SALLE 214

Abstract

Since 1994, vertical profiles of atmospheric radiation have been measured at the upper air meteorological station of the Czech Hydrometeorological Institute in Prague. Antoch *et al.* (2006) study individual profiles and propose parametric estimators to explain the dependance of radiation intensity on the altitude. Considering estimated profiles as functional variables, the aim of this contribution is to model changes in radiation during the measured period – the last decade.

To model the dependance of the radiation on time we suggest a functional regression in the form

$$m(a, t) = \mathbf{E}[Y(a) | T = t],$$

where $Y(a)$ denotes a radiation profile (a functional variable) and T represents a real valued time. Several estimators of $m(a, t)$ are provided. We apply a kernel

smoother proposed by Cardot (2005) at first. Parametric proportional estimators

$$\widehat{m}(a, t; \boldsymbol{\theta}) = Y_0(a; \boldsymbol{\theta}) \exp \{f(t; \boldsymbol{\theta})\}$$

with a baseline profile $Y_0(a)$ and time dependent proportion $f(t)$ detailed studied by Hlubinka and Prchal (2006) are then discussed. Finally, results of permutation lack-of-fit tests (see Cardot *et al.*, 2006) are shown to prove the adequacy of the parametric estimators.

Acknowledgement. The research was partially supported by the grant No. 201/05/H007 of the Grant Agency of the Czech Republic and by the Research Project No. MSM 0021620839 of the Ministry of Education of the Czech Republic.

Références

1. Antoch, J., Hlubinka, D., and Prchal, L. (2006). Statistical methods for analysis of meteorological measurements, *to appear*.
2. Cardot, H. (2005). Nonparametric regression for functional responses with application to conditional functional principal component analysis. *Technical Report LSP, Université Paul Sabatier, LSP-2005-01*, 18 p.
3. Cardot, H., Prchal, L., and Sarda, P. (2006). No-effect and lack-of-fit permutation tests for functional regression, *to appear*.
4. Hlubinka, D., and Prchal, L. (2006). Changes in atmospheric radiation from the statistical point of view, *to appear*.

Common functional principal components

Michal BENKO*,
Wolfgang HAERDLE and Alois KNEIP

* Adresse pour correspondance :
CASE - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
Spandauerstr 1, D 10178 Berlin, Germany

e-mail : benko@wiwi.hu-berlin.de

SÉANCE DU 20 FEVRIER, 16H, SALLE 214

Abstract

Functional principal component analysis based on the Karhunen-Loève decomposition has been successfully applied in many applications, mainly for one sample problems. Analysis of (functional) principal components appears to be often only practically feasible way of accessing the distribution of the random functions and understanding its structure. In this talk we consider "common functional components" and focus on two sample problems.

First we present a new method for estimation of functional principal components from discrete noisy data and discuss the asymptotic properties of proposed method. It consists in an adaptation of a technique introduced by Kneip and Utkal (2001) for the case of density functions. The key-idea is to represent the components of the Karhunen-Loève expansion in terms of an (L^2) scalar-product matrix of the sample.

The obtained asymptotic results imply conditions under which the error caused by estimation from the discrete noisy data is first-order asymptotically negligible in comparison to the estimation based on a directly observed functional sample and further inference may proceed "as if" the functions were directly observed. Obtained theoretical results also motivate the construction of bootstrap tests for one and two sample inference.

Next we consider two (random) samples of functions. We are interested in comparison of the distributions of these two samples by similarities of the corresponding Karhunen-Loève decompositions. For this reason, we propose bootstrap test for testing the equality of eigenvalues, eigenfunctions, and mean functions of two functional samples. We illustrate the practical usage by a simulation study and application to the implied volatility analysis, improving the (multivariate) method proposed by Fengler, Härdle and Villa (2003) into the functional data framework.

We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft and the Sonderforschungsbereich 649 “Ökonomisches Risiko”.

Références

- Benko, M. and Kneip, A. (2005). Common functional component modelling *Proceedings of 55th Session of the International Statistical Institute*, Sydney 2005
- Fengler, Härdle and Villa (2003). The Dynamics of Implied Volatilities : A Common Principle Components Approach, *Review of Derivative Research* 6 : 179-202.
- Kneip, A. and Utikal, K. (2001). Inference for Density Families Using Functional Principal Components Analysis, *Journal of the American Statistical Association* 96 : 519-531.

Modèle de mélange pour des données répétées de loi exponentielle

Marie-José MARTINEZ

* Adresse pour correspondance :
Institut de Mathématiques et de Modélisation de Montpellier
Université Montpellier II
Place Eugène Bataillon, 34095 Montpellier Cedex 5

e-mail : martinez@math.univ-montp2.fr

SÉANCE DU 6 MARS, 16H, SALLE 214

Résumé

Nous nous intéressons à un modèle de mélange pour des données répétées de loi exponentielle. Les composants du mélange traduisent différents états possibles des individus. Pour chacun de ces composants, on modélise la dépendance et l'extra-variabilité dues à la répétition des données par l'introduction d'effets aléatoires. Un exemple d'application est la modélisation de durées de séjours hospitaliers répétés pour lesquels on cherche à mettre en évidence des classes de patients non contrôlées a priori. En fiabilité, une application est la modélisation de données de défaillance de matériels supposés sans vieillissement et réparables selon l'hypothèse "as good as new". On peut alors associer les composants du mélange à différentes catégories de matériels.

Dans ce modèle, l'estimation des paramètres n'est pas directement envisageable par l'algorithme EM. En effet, l'approche proposée par Celeux, Martin et Lavergne (2005) dans le cadre d'un mélange de modèles linéaires mixtes (L2M) est confrontée au problème de la non accessibilité de la distribution marginale de chaque composant du mélange. Deux méthodes sont alors envisagées. La première méthode proposée combine une linéarisation spécifique du modèle exponentiel mixte associé à chaque composant du mélange et l'utilisation de l'algorithme EM pour un mélange de L2M. La seconde méthode, quant à elle, s'appuie sur une étape de Metropolis-Hastings pour construire un algorithme de type MCEM.

Cette méthode, contrairement à la précédente, est applicable dans le cadre d'un mélange quelconque de modèles à effets aléatoires c'est-à-dire sans spécification particulière de la loi.

Les performances des algorithmes d'estimation qui en découlent sont illustrées sur des simulations. On discutera de la qualité des estimations et de l'intérêt de prendre en compte les répétitions.

Références

Celeux G., Martin O. et Lavergne C. (2005) Mixture of linear mixed models for clustering gene expression profiles from repeated microarray experiments. *Statistical Modelling*, 5, 243-267.

Gaudoin O., Lavergne C. et Soler J.L. (1994) A generalized geometric de-eutrophication software reliability model. *IEEE Transactions on Reliability*, 43, 536-541.

McCulloch C.E. (1997) Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92, 162-170.

McLachlan G. et Peel D. (2000) *Finite mixture models*, Wiley.

Tests for independence based on zonotopes

G.A. KOSHEVOY*,
J. MÖTTÖNEN and H. OJA

* e-mail : koshevoy@ihes.fr

SÉANCE DU 10 AVRIL, 16H, SALLE 214

Abstract

We propose new nonparametric tests based on zonotopes. Specifically, let $(x_i^T, y_i^T)^T, i = 1, \dots, n$ denote a sample of random vectors pairs, where $x - i$ and y_i have dimensions p and q , respectively. We wish to test the null hypothesis.

$$H_0 : x_i \text{ and } y_i \text{ are independent.}$$

Let $Z_p(X)$, $Z_q(Y)$, and $Z_{p+q}((X^T, Y^T)^T)$ to denote the zonoids of the corresponding random vectors. Note that $Z - p(X)$ and $Z_q(Y)$ are just projections of $Z_{p+q}((X^T, Y^T)^T)$ to the first p components (x -space) and to the last q components (y -space), respectively. If the marginal random vectors x and y are linearly dependent then the extreme points of $Z_{p+q}((X^T, Y^T)^T)$ in the x -direction are extreme points in y -directions as well. If we project these points to the y -space we get

$$\{ave\{sign(u^T x_i) y_i\} : u^T u = 1\}.$$

Similar projections on the extreme y -points in the x -space are

$$\{ave\{sign(v^T y_i) x_i\} : v^T v = 1\}.$$

We construct the tests of independence based on the statistics

$$T_1(u) = \{ave\{sign(u^T x_i) y_i\}\},$$

$$T_2(u) = \frac{2}{n-1} \sum_{i < j} \{sign(u^T (x_i - x_j) (y_i - y_j))\},$$

$$S_1(v) = \{ave\{sign(v^T y_i) x_i\}\},$$

and

$$T_2(v) = \frac{2}{n-1} \sum_{i < j} \{ \text{sign}(v^T(y_i - y_j)(x_i - x_j)) \}.$$

We establish the limiting distributions of the proposed statistics under certain alternative sequences, and compare our new test statistics to Wilks' test by using asymptotic relative efficiencies and simulation studies.

Rappels des activités antérieures de STAPH

Les documents ci-dessous, de même que l'ensemble de nos activités,
sont accessibles sur le site :

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

- K. Benhenni, A. Boudou, H. Cardot, F. Ferraty, M. Rachdi, Y. Romain, P. Sarda, S. Viguier et Ph. Vieu (2006). 4èmes Journées de Statistique Fonctionnelle et Opératoireielle. *Pub. Labo. Stat. Proba. Toulouse*, **LSP 2006-04**.
- A. Boudou, H. Cardot, F. Ferraty, Y. Romain, P. Sarda, S. Viguier et Ph. Vieu (2005). Statistique Fonctionnelle 6 : Actes du groupe de travail STAPH 2004-2005. *Pub. Labo. Stat. Proba. Toulouse*, **LSP 2005-10**.
- A. Boudou, H. Cardot, F. Ferraty, Y. Romain, P. Sarda, S. Viguier et Ph. Vieu (2005). 3èmes Journées de Statistique Fonctionnelle et Opératoireielle. *Pub. Labo. Stat. Proba. Toulouse*, **LSP 2005-05**.
- A. Boudou, H. Cardot, F. Ferraty, Y. Romain, P. Sarda, S. Viguier et Ph. Vieu (2004). Statistique Fonctionnelle 5 : Actes du groupe de travail STAPH 2003-2004. *Pub. Labo. Stat. Proba. Toulouse*, **LSP 2004-06**.
- A. Boudou, H. Cardot, F. Ferraty, Y. Romain, P. Sarda, S. Viguier et Ph. Vieu (2003). Statistique Fonctionnelle 4 : Actes du groupe de travail STAPH 2002-2003. *Pub. Labo. Stat. Proba. Toulouse*, **LSP 2003-05**.
- S. Viguier, A. Boudou, H. Cardot, F. Ferraty, Y. Romain, P. Sarda, Ph. Vieu et A. Yousfate (2003). Actes des 2èmes journées de Statistique Fonctionnelle et Opératoireielle, Toulouse Juin 2003. *Pub. Labo. Stat. Proba. Toulouse*, **LSP 2003-06**.
- H. Cardot, F. Ferraty, Y. Romain, P. Sarda et Ph. Vieu (2002). Statistique Fonctionnelle 3 : Actes du groupe de travail STAPH 2001-2002. *Pub. Labo. Stat. Proba. Toulouse*, **LSP 2002-12**.
- H. Cardot, F. Ferraty, A. Mas, Y. Romain, P. Sarda, Ph. Vieu et A. Yousfate (2002). Actes des journées de Statistique Fonctionnelle, Toulouse 10-11 Juin 2002. *Pub. Labo. Stat. Proba. Toulouse*, **LSP 2002-09**.
- H. Cardot, F. Ferraty, Y. Romain, P. Sarda et Ph. Vieu (2001). Statistique Fonctionnelle 2 : Actes du groupe de travail STAPH 2000-2001. *Pub. Labo. Stat. Proba. Toulouse*, **LSP 2001-07**.
- H. Cardot, F. Ferraty, Y. Romain, P. Sarda et Ph. Vieu (2001). Statistique Fonctionnelle 1 : Actes du groupe de travail STAPH 1999-2000. *Pub. Labo. Stat. Proba. Toulouse*, **LSP 2001-05**.