

# Séminaire de statistique ENSAI

Année 2006-2007

**Responsable Valentin Patilée**

## Vendredi 29 juin 2007

11h00 à 11h50 : **Jean-Marie ROLIN** (Université catholique de Louvain, Belgique)

*Survival analysis with Lévy priors.*

**Résumé.** In a nonparametric Bayesian analysis of survival data, it is usual to use as a prior specification for the baseline distribution function  $F$  a neutral to the right process. Due to the discreteness of such a process, there exist two ways of defining hazard functions, namely:  $\Lambda([0; t]) = -\ln F((t; 1])$  and  $\Lambda^{\{-\}}([0; t]) = \int_{[0; t]} F([s; \infty])^{-1} F(ds)$ . In the presence of an explanatory process, two multiplicative semiparametric models for the hazard function can then be considered: the multiplicative intensity model and the proportional hazards model. This paper compares the posterior distributions of the Euclidean parameter (challenging the Cox's likelihood in the noninformative case) and also the posterior distributions of the baseline distribution function conditional on the Euclidean parameter in these two different models.

## Vendredi 1 juin 2007

11h00 à 11h50 : **Sébastien VAN BELLEGEM** (Université catholique de Louvain, Belgique)

*Local adaptive estimation for nonstationary time series.*

**Résumé.** Increasingly more attention is paid to the analysis on nonstationary time series, that is time series with a time-varying autocovariance function. Several models have been proposed, e.g. models with structural breaks or a smooth evolution of the autocovariance function over time. We present a new method to fit locally a stationary model to a nonstationary data. The method is based on a sequence of local likelihood ratio tests of stationarity against change-point alternative. We establish important results concerning the optimality of the procedure, study its empirical performance for particular structural models and show an application on a real data set.

## Vendredi 4 mai 2007

10h00 à 10h45 : **Harry L. HURD** (University of North Carolina, Chapel Hill)

*Time Series Analysis for Periodically Correlated Sequences of Less than Full Rank*

**Résumé.** In this talk we consider several aspects of time series analysis for periodically correlated (PC) sequences when the observed sequence is of deficient rank. We will begin by discussing the rank of a PC sequence and then give some results on deficient rank PARMA sequences. Finally, we describe some of the considerations of time series analysis under rank deficiency.

11h00 à 11h50 : **Marc HALLIN** (Université Libre de Bruxelles)

*Invariance, distribution-freeness, and semiparametric efficiency -- sixty years of rank-based inference.*

**Résumé.** The modern history of ranks in statistics started in 1945 with Frank Wilcoxon's far-reaching three page paper on rank tests for location. Emphasis in 1945 was on distribution-freeness and ease of applications. Since then, under the impulse of such names as Chernoff, Savage, Hodges, Lehmann, Hajek, and Le Cam, rank-based methods have followed the development of contemporary statistics, and turned into a complete body of modern, flexible and powerful techniques. In this talk, we show how this evolution, from distribution-freeness to group invariance and tangent space projections, eventually may reconcile the enemy brothers of statistics -- efficiency and robustness.

## Vendredi 30 mars 2007

11h00 à 11h50 : **Catherine MATIAS** (CNRS, Laboratoire Statistique et Génome, Evry)

*Estimation conjointe du nombre d'états cachés et de la mémoire dans des chaînes de Markov à régime Markovien.*

**Résumé.** Ce travail porte sur l'identification de l'ordre d'une chaîne de Markov à régime Markovien (MCMR) sur un alphabet fini. L'ordre d'une MCMR est défini comme le couple  $(k, m)$  où  $k$  est le nombre d'états de la chaîne cachée et  $m$ , la mémoire de la chaîne de Markov conditionnelle. La théorie de l'information nous suggère trois estimateurs de l'ordre construits à partir de lois de codage différentes. L'originalité de notre approche réside à la fois dans l'estimation simultanée d'un couple de paramètres de structure du modèle ainsi que dans le fait que les modèles ne sont pas emboîtés. Nous établissons la consistance forte de nos estimateurs, sans hypothèse de borne a priori sur  $k$  et  $m$ . Des simulations illustrent ces résultats. Travail en collaboration avec Antoine Chambaz.

## Vendredi 9 mars 2007

11h00 à 11h50 : **Vladimir SPOKOINY** (WIAS, Berlin)

*Robust Risk Management. Accounting for Nonstationarity and Heavy Tails.*

**Résumé.** In the ideal Black-Scholes world, financial time series are assumed 1) stationary (time homogeneous) and 2) having conditionally normal distribution given the past. These two assumptions have been widely-used in many methods such as the RiskMetrics, one risk

management method considered as industry standard. However these assumptions are unrealistic. The primary aim of the paper is to account for nonstationarity and heavy tails in time series by presenting a local exponential smoothing approach, by which the smoothing parameter is adaptively selected at every time point and the heavy-tailedness of the process is considered. A complete theory addresses both issues. In our study, we demonstrate the implementation of the proposed method in volatility estimation and risk management given simulated and real data. Numerical results show the proposed method delivers accurate and sensitive estimates.

## Vendredi 2 février 2007

11h00 à 11h50 : **Christian P. ROBERT (Ceremade - Université Paris-Dauphine)**  
*K-nearest neighbours: a Bayesian perspective.*

**Résumé.** The  $k$ -nearest-neighbour procedure uses a training dataset  $\left\{ (y_i, x_i) \right\}_{i=1}^n$  to make predictions on new unlabelled data where  $y_i \in \{C_1, \dots, C_G\}$  denotes the class label of the  $i$ -th point and  $x_i$  denotes a vector of  $p$  predictor variables. The prediction for a new point  $\left( (y_{n+1}, x_{n+1}) \right)$  is reported as the most common class found amongst the  $k$ -nearest-neighbour of  $x_{n+1}$  in the set  $\left\{ (x_i) \right\}_{i=1}^n$ . The neighbours of a point are defined via a distance metric  $\rho(x_{n+1}, x_i)$  which is commonly taken to be the Euclidean norm. The  $k$ -nearest-neighbour algorithm is a nonparametric procedure. Traditionally, the value of  $k$  is chosen by minimizing cross-validated misclassification rate. Here we propose some alternative approaches using probability models and Bayesian perspectives. A novel perfect sampling algorithm for resolving difficulties linked with MRFs is introduced on the way. (Joint work with G. Celeux, J.-M. Marin and D.M. Titterton)

## Vendredi 12 janvier 2007

11h00 à 11h50 : **Guillaume CHAUVET (LSE, CREST-ENSAI)**  
*Estimation de précision par Bootstrap dans une enquête en population finie.*

**Résumé.** Lors d'une enquête complexe, on souhaite généralement produire des indicateurs avec leur précision, voire donner un intervalle de confiance. On utilise généralement la technique de linéarisation, mais on capte mal la variance d'une statistique fortement non linéaire telle qu'un fractile. D'autre part, la linéarisation nécessite une connaissance précise du plan de sondage, qui n'est pas forcément accessible à l'utilisateur d'un fichier d'enquête. Nous examinerons comment une méthode de Bootstrap peut être mise en oeuvre pour évaluer la précision sous différents mécanismes d'échantillonnage (tirage poissonien, sondage aléatoire simple, tirage réjectif, échantillonnage équilibré), voire pour un plan de sondage complexe.

## Vendredi 8 décembre 2006

11h00 à 11h50 : **Jean-Pierre FLORENS (Univ. Toulouse I)**  
*Instrumental regression in partially linear regression models.*

**Résumé.** We consider the semiparametric regression  $X'\beta + \phi(Z)$  where  $\beta$  and  $\phi(\cdot)$  are unknown slope coefficient vector and function, and where the variables  $(X, Z)$  are endogeneous. We propose necessary and sufficient conditions for the identification of the parameters in the presence of instrumental variables. We also focus on the estimation of  $\beta$ . An incorrect parametrization of  $\phi$  generally leads to an inconsistent estimator of  $\beta$ , whereas consistent nonparametric estimators for  $\beta$  have a slow rate of convergence. An additional complication is that the solution of the equation necessitates the inversion of a compact operator which can be estimated nonparametrically. In general this inversion is not stable, thus the estimation of  $\beta$  is ill-posed. In this paper, a  $\sqrt{n}$ -consistent estimator for  $\beta$  is derived under mild assumptions. One of these assumptions is given by the so-called *source condition* which we explicit and interpret in the paper. Finally we show that the estimator achieves the semiparametric efficiency bound, even if the model is heteroskedastic.

## Vendredi 17 novembre 2006

11h00 à 11h50 : **Jean-Yves DAUXOIS (Univ. de Franche Comté)**  
*Risques concurrents en Statistique des durées de vie.*

**Résumé.** En Statistique des durées de vie, on parle de risques concurrents quand divers causes peuvent entraîner le décès (resp. la panne) d'un patient (resp. d'un matériel). On observe alors le temps du décès (resp. de la panne) et sa cause. Après une présentation de cette problématique et de la littérature existante, nous donnerons quelques résultats récents que nous avons obtenus dans ce domaine (suite à différentes collaborations). Nous aborderons en particulier les problèmes suivants :

- Risques concurrents et observations sous biais de sélection ;
- Événements récurrents et risques concurrents ;
- Risques concurrents et durées de vie multivariées ;
- Risques concurrents et données manquantes.

## Vendredi 20 octobre 2006

11h00 à 11h50 : **Paul DOUKHAN (Lab. de Statistique du CREST, Paris, et SAMOS-MATISSE, Univ. Paris 1)**  
*Dépendance faible et méthode de Lindeberg.*

**Résumé.** Quelques exemples de séries temporelles stationnaires sont faiblement dépendants et ne possèdent pas les propriétés usuelles de dépendance. Une version simple du théorème de Lindeberg pour des variables dépendante donne lieu aux habituels théorèmes de limite centrale via des techniques de blocs de Bernstein. Des applications de ce lemme donne directement des énoncés en sous échantillonnage comme en statistique non paramétrique.

## Vendredi 29 septembre 2006

11h00 à 11h50 : [Arthur CHARPENTIER](#) (ENSAI)

*Dépendance entre événements extrêmes.*

**Résumé.** L'extention du théorème de Fisher-Tippett en dimension 2, en étudiant les maximas par composantes, a permis d'obtenir les copules extrêmes, satisfaisant une équation de max-stabilité. Le but sera ici de proposer une extention du théorème de Balkema-Pickands-de Haan en dimension 2, en particulier en étudiant les dépendances entre extrêmes dans le cas de copules Archimédiennes. Nous verrons en particulier des résultats pouvant être reliés aux travaux de Ledford et Tawn (1996, 1997), ou plus récemment de Resnick (2004) sur la notion de variation régulière cachée.