Abstracts

*** Jérémie Bigot**
Moyenne de Fréchet pour l'analyse statistique de formes et des images

Dans cet exposé nous proposons de présenter le problème de l'estimation d'une forme moyenne à partir d'un ensemble de courbes ou d'images similaires. Lorsque la variabilité des données est due à la présence de déformations aléatoires et de bruit additif, ce problème nécessite de définir des distances appropriées entre des courbes ou des images. Ceci peut être fait en modélisant la variabilité de forme des signaux naturels à l'aide de groupes de transformations. Un axe de recherche récent est l'étude des aspects probabilistes et statistiques des modèles de déformations, et le développement de procédures consistentes d'estimation d'une forme commune à partir d'un ensemble de courbes ou d'images. L'exposé sera centré sur cette question. Nous proposons d'étudier en particulier quelques modèles statistiques de déformations, et de montrer les liens existants avec la théorie minimax en statistique non-paramétrique et les problèmes inverses. Les méthodes d'estimation seront également illustrées à l'aide d'exemples numériques.

*** Félix Abramovich**
Model Selection in Regression: some new (?) thoughts on the old (?) problem

We discuss the model selection problem in linear regression, where the number of potential predictors might be even much larger than the number of observations. The proposed Bayesian approach implies the penalized least squares estimation with a complexity penalty associated with a prior on the model size. We present optimality properties of the resulting Bayesian model selector and compare it with several well-known existing counterparts. Computational problems and possible solutions will be also discussed.
This is joint work with Vadim Grinshtein from the Open University of Israel.

*** Umberto Amato**
Anestis Antoniadis: a life for Statistics

*** René Carmona**
Market Mechanisms for GHG Emissions Control: Mathematical Challenges

We review recent attempts to provide rigorous mathematical analyses of the cap-and-trade schemes used as market mechanisms to control greenhouse gas emissions. We include equilibrium models as well as reduced form models proposed to price options. We conclude with a set of new mathematical results showing that some of the singular BSDEs occurring in the pricing of emission allowance can have pathological behaviors.

*** Jianquing Fan**
Journey to ultrahigh-dimensional space

Ultrahigh-dimensionality characterizes many contemporary statistical problems from genomics and engineering to finance and economics. To visit such a vast dimension, large-scale and moderate-scale vehicles are needed for such a venture.
We outline a unified framework to ultrahigh dimensional variable selection problems: Iterative applications of vast-scale screening followed by moderate-scale variable selection, called ISIS.
The framework is widely applicable to many statistical contexts: from multiple regression, generalized linear models, survival analysis to machine learning and compress sensing. The fundamental building blocks are marginal variable screening and penalized likelihood methods. The focus will be on the sure screening property, false selection size, the model selection consistency and oracle properties. The issue of risk approximation in classification will also be discussed. The methods will be convincingly illustrated by carefully designed simulation studies and the empirical studies on disease classifications using microarray data and forecast home price indexes at zip level.

*** Irène Gijbels**
Smoothing and variable selection using P-splines

In this talk we focus on penalized estimation in additive models and varying coefficient models. A main interest is also on variable selection for such models. Particular attention goes to recent variable selection procedures such as grouped Lasso grouped SCAD, grouped Bridge and grouped COSSO, but also to the nonnegative garrote method introduced originally for variable selection in a multiple linear regression model. We show how the latter method combined with P-splines estimation leads to an estimation and variable consistent method in both the settings of additive models and varying coefficient models. The performances of this and other related selection procedures are investigated in a simulation study and illustrations on real data examples are provided.
In the additive varying coefficient model a uniform framework is presented and the various grouped regularization procedures are studied. A discussion on implementation
issues and computational algorithms is provided, and a comparative simulation study is given.
This talk is based on joints work with Anestis Antoniadis, Sophie Lambert-Lacroix and Anneleen Verhasselt.

*** Stéphane Mallat**
Classification by Invariant Scattering

Signal classes are often invariant to groups of operators such as translations or rotations. Classification then requires finding informative invariants which are stable to local deformations. A scattering transform defines a signal representation which is invariant to translations and to other groups of operators. It provides a new representation of stationary processes and linearizes locally the signal deformations. Classifications of complex signals and textures are implemented with low-dimensional affine space models in the scattering domain. Applications to image and texture classification will be shown.

*** Mariana Pensky**

On convergence rates equivalency and sampling strategies in functional deconvolution models.

Using the asymptotical minimax framework, we examine convergence rates equivalency between a continuous functional deconvolution model and its real-life discrete counterpart, over a wide range of Besov balls and for the L^2-risk. For this purpose, all possible models are divided into three groups. For the models in the first group, which we call {\it uniform}, the convergence rates in the discrete and the continuous models coincide no matter what the sampling scheme is chosen and, hence, the replacement of the discrete model by its continuous counterpart is legitimate. For the models in the second group, to which we refer as {\it regular}, one can point out the best sampling strategy in the discrete model, but not every sampling scheme leads to the same convergence rates: there are at least two sampling schemes which deliver different convergence rates in the discrete model (i.e., at least one of the discrete models leads to convergence rates that are different from the convergence rates in the continuous model). The third group consists of models for which, in general, it is impossible to devise the best sampling strategy; we call these models {\it irregular}.

We formulate the conditions when each of these situations takes place.

In the regular case, not only we point out the number and the selection of sampling points which deliver the fastest convergence rates in the discrete model but also investigate when, in the case of an arbitrary sampling scheme, the convergence rates in the continuous model coincide or do not coincide with the convergence rates in the discrete model.

We also study what happens if one chooses a uniform, or a more general pseudo-uniform, sampling scheme which can be viewed as an intuitive replacement of the continuous model.

Finally, as a representative of the irregular case, we study functional deconvolution with a box-car like blurring function since this model has a number of important applications.

All theoretical results presented in the paper are illustrated by numerous examples many of which are motivated directly by a multitude of inverse problems in mathematical physics where one needs to recover initial or boundary conditions on the basis of observations from a noisy solution of a partial differential equation. The theoretical performance of the suggested estimator in the multichannel deconvolution model with a box-car like blurring function is also supplemented by a limited simulation study and compared to an estimator available in the current literature. The paper concludes that in both regular and irregular cases one should be extremely careful when replacing a discrete functional deconvolution model by its continuous counterpart.

*** Jean-Michel Poggi**

Detecting influential observations using tree-based methods

The talk will present two tree-based procedures for the detection of influential observations. The first one dealing with boosting and regression is briefly recalled while the second define, thanks to jackknife trees, influence functions for CART to highlight influential observation in classification problems.

A procedure for detecting outliers in regression problems based on information provided by boosting trees has been proposed (in a joint work with Nathalie Cheze in 2006). The key idea was to select the most frequently resampled observation along the boosting iterations and reiterate boosting after removing it. The selection criterion is based on the maximum over the boosting iterations of the average number of appearances in bootstrap samples. Then the procedure is noise distribution free and fully nonparametric leading to a very general definition of outlier.

More recently, a joint work with Avner Bar-Hen and Servane Gey, focusing on classification problems, deals with measuring the influence of observations on the results obtained with CART. To define the influence of individuals on the analysis, we use the notion of influence and we propose some general criterions to measure the sensitivity of the CART analysis and its robustness. The proposals are based on jackknife trees, are organized around two lines: influence on predictions and influence on partitions. In addition, the analysis is extended to the pruned sequences of CART trees to produce a CART specific notion of influence.

*** Vladimir Spokoiny**

Non-Gaussian Component Analysis using semi-definite relaxation

Non Gaussian Component Analysis is an unsupervised method of extracting a linear structure from a high dimensional data based on estimating a low-dimensional non-Gaussian data component.

This paper offers a new approach to direct estimation of the projector on the target space based on semidefinite programming.

The new procedure differs significantly from the earlier proposals in Blanchard et al (2006) and Diederichs et al (2010) and it improves the method efficiency and sensitivity to a broad variety of deviations from nor mality and decreases the computational effort.

It particularly enables to proceed with much higher dimensions without loss of accuracy.