Mathematical Statistics and Inverse Problems

February 8 – 12 (2016), CIRM, France

1. Butucea Cristina (Université de Paris Est-Marne la Vallée, France)

   Title : Quantum statistical models and inference.

2. Buzun Nazar (WIAS Berlin, Germany)

   Title : Multiplier bootstrap for change point detection.

   Abstract : In Change point detection task Likelihood Ratio Test (LRT) is sequentially applied in a sliding window procedure. Its high values indicate changes in parametric distribution of the data sequence. Correspondingly LRT values require predefined bound for their maximum. The maximum value has unknown distribution and may be calibrated with multiplier bootstrap. Bootstrap procedure convolves independent components of the Likelihood function with random weights, that enables to estimate empirically LRT distribution. For this empirical distribution of the LRT we show its convergence to the real maximum value distribution.

3. Dalalyan Arnak (ENSAE, Paristech, France)

   Title : Convex programming approach to robust estimation of a multivariate Gaussian model.

   Abstract : Multivariate Gaussian distribution is often used as a first approximation to the distribution of high-dimensional data. Determining the parameters of this distribution under various constraints is a widely studied problem in statistics, and is often considered as a prototype for testing new algorithms or theoretical frameworks. In this paper, we develop a non asymptotic approach to the problem of estimating the parameters of a multivariate Gaussian distribution when data are corrupted by outliers. We propose an estimator-efficiently computable by solving

a convex program-that robustly estimates the population mean and the population covariance matrix even when the sample contains a significant proportion of outliers. In the case where the dimension p of the data points is of smaller order than the sample size, our estimator of the corruption matrix is provably rate optimal simultaneously for the entry-wise l_1-norm, the Frobenius norm and the mixed l_2/l_1 norm. Furthermore, this optimality is achieved by a penalized square-root-of-least-squares method with a universal tuning parameter (calibrating the strength of the penalization). These results are partly extended to the case where p is potentially larger than n, under the additional condition that the inverse covariance matrix is sparse.

4. Enikeeva Farida (Université de Poitiers, France)

   Title : Bump detection in heterogeneous Gaussian regression.

   Abstract : We consider the problem of the bump detection problem of a signal in a heterogeneous Gaussian regression model. We allow for a simultaneous change in mean and in variance of the signal and specify its impact on the difficulty to detect the null signal against a single bump. We derive lower and upper bounds of testing that lead to explicit characterizations of the detection boundary in several sub-regimes depending on the asymptotic behavior of the bump heights in mean and variance. In particular, we explicitly identify those regimes, where the additional information about a simultaneous bump in variance eases the detection problem. This effect is made explicit in the constant and / or the rate, appearing in the detection boundary. We also discuss the case of an unknown bump height and provide an adaptive test in that case. This is a joint work with Axel Munk and Frank Werner.

5. Ermakov Mikhail (Institute of Mechanical Engineering Problems, Russie)

   Title : On consistent hypothesis testing.

   Abstract : In this talk we make a survey of results on consistent hypothesis testing and point out necessary and sufficient conditions for existence of consistent tests in the problem of signal detection, linear inverse problem, problem of hypothesis testing on covariance operator of Gaussian random vector in Hilbert space,....

6. Florens Jean-Pierre (Université Toulouse 1, France)

Title : Inverse problems in econometrics: examples and specific theoretical problems.

7. Freyermuth Jean-Marc (Cambridge University, Angleterre)

   Title : Minimax optimal detection of structure for multivariate data.

   Abstract : We overview our results on the performance of some wavelet-based methods for nonparametric function estimation. The methods we are concerned with pool information within geometric structures in the coefficient domain (horizontal/vertical blocks). They have attracted a lot of attention being often motivated by practical reasons. We will discuss how the maxiset approach provides a very convenient way to assess the optimality of such methods for function estimation. The maxiset approach consists in determining the largest functional space over which the risk of an estimator converges at a chosen rate. In the univariate setting, information pooling allows to get large maxisets. Nevertheless, in the multivariate setting, the situation is less straightforward since some of these estimators are much more exposed to the curse of dimensionality. However, thanks to the maxiset approach, we are able to identify cases where information pooling still has a clear benefit. These cases correspond to global structural properties of the estimand that can be related to compound models and to a 'minimal' level of anisotropy. Finally, we will also discuss how the geometry of the hyperbolic wavelet basis allows to construct 'optimal' testing procedures of these structural characteristics.

8. Goldenshluger Alexandre (University of Haifa, Israel)

   Title : The M/G/∞ estimation problem revisited.

   Abstract : The subject of this talk is nonparametric estimation of the service time distribution in the M/G/∞ queue from incomplete data. Two different approaches to constructing estimators will be presented, and properties of the corresponding estimators will be discussed.

9. Hengartner Nicolas (Los Alamos National Laboratory, Etats-Unis)

   Title : : TBA.

10. Hohage Thorsten (Gottingen University, Allemagne)

    Title : Variational Regularization of Nonlinear Statistical Inverse Problems.

11. Huckemann Stephan (Gottingen University, Allemagne)

    Title : Some Statistics for Live Biological Cells. Drift estimation in sparse sequential dynamic imaging and circular scale space theory.

    Abstract : In vivo observation of biological cellular structures below the resolution of visible light has recently been made possible by single marker switching (SMS) microscopy yielding sparse images. This comes at a price, however, of extended exposure time, which can induce blur due to drifts of the probe. We develop a simple semiparametric model for drift correction in sequences of sparse images and propose an M-estimator for the drift and show its asymptotic normality. This is used to correct the final image and it is shown that this purely statistical method is competitive with state of the art calibration techniques which require to incorporate fiducial markers into the specimen. Time permitting, we show how the study of early diversification of stem cells links to circular scale space theory of which we develop some central aspects.

12. Ibragimov Ildar (Steklov Mathematical Institute, Russie)

    Title : Estimation of infinite-dimensional parameter in l_p spaces.

    Abstract : 1. U. Grenander in his book "Abstract Inferences" (J. Wiley, 1981) has considered the following non-parametric estimation problem.

    We are observing a sample

$$X_1, X_2, \ldots X_n$$

    where the random variables X_j take integer values $1, 2, \ldots$ with the probabilities $\theta(k) = \mathbf{P}\{X_j = k\}$. The problem is to estimate the (infinite-dimensional) parameter $\theta = (\theta(1), \theta(2), \ldots$.

    In particular U. Grenander proved the following result :

Theorem : If $\sum_1^\infty \sqrt{\theta(j)} < \infty$, then the maximum likelihood estimates (MLE) satisfy the following limit relation

$$\lim_{n\to\infty} \sqrt{n}||\hat{\theta}_n - \theta||_1 = \lim \sum_1^\infty \sqrt{n}|\hat{\theta}_n(j) - \theta(j)| = \sqrt{\frac{2}{\pi}} \sum_1^\infty \sqrt{\theta(j)(1-\theta(j))}. \quad (1)$$

Here we continue the investigation of U. Grenander.

2. The MLE $\hat{\theta}_n = (\hat{\theta}_n(1), \dots)$ where $\hat{\theta}_n(k) = \frac{\sharp\{j, 1\le j\le n: X_j = k\}}{n}$. It follows that with probability one $\sqrt{n}(\hat{\theta}_n - \theta) \in l_p$, $1 \le p \le \infty$. Denote L_p, 0¡p¡1 the metric space of $x = (x_1, \dots)$, $\sum_1^\infty |x_j|^p < \infty$ with the metric $\rho(x,y) = \sum_1^\infty |x_j - y_j|^p$; if $\sum_1^\infty (\theta(j))^p < \infty$, then $\sqrt{n}(\hat{\theta}_n - \theta) \in l_p$ with probability one.

Denote $Q_n(p,\theta)$ the distribution of the normed MLE $\sqrt{n}(\hat{\theta}_n - \theta)$ in the space l_p/ Denote $Q(p,\theta)$ the distribution in l_p of the Gaussian random sequence

$$\eta = (\eta_1, \dots) = (\xi_1\sqrt{\theta(1)(1-\theta(1))}, \dots \xi_k\sqrt{\theta(k)(1-\theta(k)}, \dots)$$

where $\xi_1, \xi_2, \dots$ are Gaussian random variables such that

$$\mathbf{E}\xi_k = 0,\ \mathbf{E}\xi_k^2 = 1,\ \mathbf{E}\xi_k\xi_l = -\sqrt{\frac{\theta(k)\theta(l)}{(1-\theta(k))(1-\theta(l)}}.$$

Theorem : Let $0 < p \le \infty$. If $\sum_1^\infty (\theta(k))^{p/2} < \infty$, then the distributions $Q_n(p,\theta)$ converge to the distribution $Q(p,\theta)$, $n \to \infty$. In other words for any real valued bounded continuous in l_p function $\varphi(x)$

$$\mathbf{E}_\theta \varphi(\sqrt{n}(\hat{\theta}_n - \theta)) = \mathbf{E}\varphi(\eta).$$

Theorem : Let $l(x) \uparrow, x \in [0,\infty)$. There exists a constant $a > 0$ such that if $l(x) \le const \cdot e^{ax^2}$, then under the conditions of the previous theorem

$$\lim_n \mathbf{E}_\theta l(\sqrt{n}(\hat{\theta}_n - \theta)||_p) = \mathbf{E}l(||\eta||_p), 1 \le p \le \infty. \quad (2)$$

In particular ((cf (1))

$$\lim \mathbf{E}_\theta ||\sqrt{n}(\hat{\theta}_n - \theta)||_p^p = \frac{1}{\sqrt{2\pi}} 2^{\frac{p+1}{2}} \Gamma(\frac{p+1}{2}) \sum_1^\infty (\theta(j)(1-\theta(j))^{p/2}. \quad (3)$$

The last equality is valid for all $p > 0$.

The MLE estimates are asymptotically efficient in the sense of I.Ibragimov, R. Khasminski , Statistical estimation: Asymptotic Theory, Springer, 1981.

13. Jongbloed Geurt (Amsterdam University, Pays-Bas)

    Title : TBA.

14. Kerkyacharian Gérard (Université de Paris Pierre et Marie Curie, France)

    Title : Geometry and inverse problems. Example tomography and astrophysics.

15. Lepski Oleg (Aix-Marseille Université, France)

    Title : Adaptive Estimation in the Convolution Structure Density Model.

16. Marteau Clément (INSA Toulouse, France)

    Title : Minimax goodness-of-fit testing in ill-posed inverse problems with partially unknown operators.

    Abstract : We consider a Gaussian sequence model that contains ill-posed inverse problems as special cases. We assume that the associated operator is partially unknown in the sense that its singular functions are known and the corresponding singular values are unknown but observed with Gaussian noise. For the considered model, we study the minimax goodness-of-fit testing problem. Working with certain ellipsoids in the space of squared-summable sequences of real numbers, with a ball of positive radius removed, we obtain lower and upper bounds for the minimax separation radius in the non-asymptotic framework, i.e., for fixed values of the involved noise levels. Examples of mildly and severely ill-posed inverse problems with ellipsoids of ordinary-smooth and super-smooth sequences are examined in detail and minimax rates of goodness-of-fit testing are obtained for illustrative purposes.

17. Mathé Peter (WIAS, Allemagne)

    Title : Discrepancy based model selection in statistical inverse problems.

    Abstract : Model selection/parameter choice is one important aspect in regularization of statistical ill-posed problems. Various strategies are studied within statistics.

However, within the classical theory of ill-posed problems the discrepancy (data misfit) is the overwhelming method of choice. For some regularization schemes, model selection requires specific principles. This holds for instance for conjugate gradient regularization. Then the question is how to use these restrictions for statistical problems. In a series of papers we have analyzed the discrepancy principle for statistical problems. The key aspects are pre-smoothing the data to make the discrepancy available, then to show why the discrepancy, and even the related Raus-Gfrerer rule can be substantiated. We will highlight, that the applicability is restricted to 'well-behaved' solutions, but that this constraint holds true for an everywhere dense set.

18. Munk Axel (Gottingen University, Allemagne)

    Title : Statistical Blind Source Separation.

    Abstract : We provide a new methodology for statistical recovery of single linear mixtures of piecewise constant signals (sources) with unknown mixing weights and change points in a multi-scale fashion. We discuss identifiability issues and show under suitable conditions exact recovery within an ϵ-neighborhood of the mixture when the sources take only values in a known finite alphabet. Based on this we provide the SESAME (SEparateS-finite-Alphabet- MixturEs) estimators for the mixing weights and sources for gaussian error. We obtain uniform confidence sets and optimal rates (up to log-factors) for all quantities. SESAME is efficiently computed as a nonconvex optimization problem by a dynamic program tailored to the finite alphabet assumption. Its performance is investigated in a simulation study. Finally, it is applied to assign copy-number aberrations (CNAs) from genetic sequencing data to different tumor clones and to estimate their proportions.

19. Pensky Marianna (University of Central Florida, Etats-Unis))

    Title : Laplace deconvolution and its application to the analysis of dynamic enhanced medical imaging data.

    Abstract : In the present paper we consider the problem of Laplace deconvolution with noisy discrete non-equally spaced observations on a finite time interval whose length may increase with a sample size. The study is motivated by perfusion imaging using a short injection of contrast agent, a procedure which is applied for medical assessment of micro-circulation within tissues such as cancerous tumors.

Unlike Fourier deconvolution which was extensively studied in the last 20 years, Laplace deconvolution received very little attention in statistical literature. In the present talk we consider several recent approaches to the problem: the kernel-based estimation, the Laguerre basis approach and application of overcomplete dictionaries. This is a joint work with Felix Abramovich, Tel-Aviv University, Israel, and Fabienne Comte, Charles-Andre Cuenod and Yves Rozenholc.

20. Pereverzev Sergei (Johann Radon Institute, Autriche)

    Title : Aggregation of regularized rankers by means of a linear functional strategy.

    Abstract : Regularization schemes are frequently used for performing ranking tasks. This topic has been intensively studied in recent years. However, to be effective a regularization scheme should be equipped with a suitable strategy for choosing a regularization parameter. In the presentation we are going to discuss an approach, which is based on the idea of a linear combination of regularized rankers corresponding to different values of the regularization parameter.The coefficients of the linear combination are estimated by means of the so-called linear functional strategy. We provide a theoretical justification of the proposed approach and illustrate it by application in ranking the risk of nocturnal hypoglycemia of diabetes patients.

21. Reiss Markus (Humboldt University, Allemagne)

    Title : From prediction error to estimation error bounds.

    Abstract : In statistical inverse problems of the form $Y = Af + \epsilon$ the error for the prediction error $\|A(\hat{f} - f)\|$ can be estimated via the residual and adaptive schemes for early stopping rely on that. It is not clear, however, whether estimators with close to optimal prediction error also have small estimation error $\|\|hatf - f\|$. We shall discuss oracle inequality results for general linear regularization methods. A clear picture is obtained to which extent this transfer of error bounds is feasible. This clarifies the power of statistical adaptation methods.

22. Tsybakov Alexandre (CREST-ENSAE, France)

Title : Sharp minimax and adaptive variable selection.

Abstract : We derive non-asymptotic bounds for the minimax risk of variable selection under expected Hamming loss in the Gaussian mean model in R^d for classes of s-sparse vectors separated from 0 by a constant $a > 0$. In some cases, we get exact expressions for the nonasymptotic minimax risk as a function of (d, s, a) and find explicitly the minimax selectors. Analogous results are obtained for the probability of wrong recovery of the sparsity pattern. As corollaries, we derive necessary and sufficient conditions for such asymptotic properties as almost full recovery and exact recovery. Moreover, we propose data-driven selectors that provide almost full and exact recovery adaptive to the parameters of the classes. This is a joint work with C. Butucea and N. Stepanova.