# Conference on Statistical Estimation

12-14 june 2023
Faculté des Sciences et Techniques
23 Rue du docteur Paul Michelon,
Saint-Étienne, France.

**Schedule**:

- 12 june 2023
  13h30-14h30: welcome coffee
  14h30-15h30: Fabienne Comte
  15h30-16h30: Vincent Rivoirard
  16h30-17h: coffee break
  17h00-18h: Elias Ould Said
- 13 june 2023:
  9h-9h30: welcome coffee
  9h30-10h30: Oleg Lepski
  10h30-11h30: Nicolas Klutchnikoff
  11h30-13h30: lunch
  13h30-14h30: Claire Lacour
  14h30-15h30: Jan Johannes
  15h30-16h: coffee break
  16h00-17h00: Anatoli Juditsky
  Conference dinner
- 14 june 2023
  8h30-9h00: welcome coffee
  9h00-10h00: Flore Sentenac
  10h00-11h00: Angelina Roche
  11h00-12h00: Timothée Mathieu
  12h00: lunch

**Programme**:

- **Fabienne Comte:** Should we estimate a product of density functions by a product of estimators?
  **Abstract:** In this talk, we consider the inverse problem of estimating the product of two densities, given a $d$-dimensional $n$-sample of i.i.d. observations drawn from each distribution. We propose a general method of estimation encompassing both projection estimators with model selection device and kernel estimators with bandwidth selection strategies. The procedures do not consist in making the product of two density estimators, but in plugging an overfitted estimator of one of the two densities, in an estimator based on the second sample. Our findings are a first step toward a better understanding of the good performances of overfitting in regression Nadaraya-Watson estimator.

- **Jan Johannes:** Statistical ill-posed inverse problems: Minimax-optimal and data-driven estimation.
  **Abstract:** *Statistical ill-posed inverse problems* are becoming increasingly important in a diverse range of disciplines, including geophysics, astronomy, medicine and economics. Roughly speaking, in all of these applications the observable signal $g = T\theta$ is a transformation of the functional parameter of interest $\theta$ under a linear operator $T$ between separable Hilbert spaces. Statistical inference on $\theta$ based on an estimation of $g$ is thus called an inverse problem as it usually necessitates an inversion of $T$. Moreover, by ill-posed we mean that the transformation $T$ is not stable, i.e. $T$ does not have a continuous inverse. In this presentation we assume that the transformation $T$ is unitarily equivalent to a multiplication by a Borel-measurable function $\mathfrak{s}_\bullet$ defined on a real measure space $(\mathcal{J}, \mathscr{J}, \nu)$, i.e. $\mathcal{J} \subset \mathbb{R}$, where the unitary maps and the function space over $(\mathcal{J}, \mathscr{J}, \nu)$ are known in advance. In other words the transformation $T$ is known if and only if the multiplicative function $\mathfrak{s}_\bullet$ is. Typical examples are circular convolution, i.e. $\mathcal{J} = \mathbb{Z}$, additive convolution on the real line, i.e. $\mathcal{J} = \mathbb{Z}$, or multiplicative convolution on the positive real line, i.e. $\mathcal{J} = \mathbb{Z}$. In most applications, however, both the signal $g$ and the inherent transformation $T$ are not known in practice, although they can be estimated from the data. More precisely, in this presentation the multiplicative function $\mathfrak{s}_\bullet$ and the unitary image $g_\bullet$ of $g$ both defined on $(\mathcal{J}, \mathscr{J}, \nu)$ are not known. However, given sample size $k, n \in \mathbb{N}$ the observations allow to form empirical processes $\widehat{\mathfrak{s}}_\bullet = \widehat{\mathbb{P}}_k(\psi_\bullet^{\mathcal{S}})$ and $\widehat{g}_\bullet = \psi^{\mathcal{S}}$ an function $\mathfrak{s}_\bullet$ and $g_\bullet$, respectively. Consequently, a statistical inference has to take into account that a random noise is present in both the estimated signal and the estimated operator.
  *Minimax-optimal estimation and adaptation.* Typical questions in this context are the nonparametric estimation of the functional parameter $\theta$ as a whole or the value of a linear functional evaluated at $\theta$, referred to as global or local estimation, respectively. It is well-known that in terms of its risk the attainable accuracy of an estimation procedure is essentially determined in the framework of this presentation by the conditions imposed on the unitary image $\theta_\bullet$ of $\theta$ and the multiplicative function $\mathfrak{s}_\bullet$ , which are, for example, expressed in the form $\theta_\bullet \in \Theta$ and $\mathfrak{s}_\bullet \in \mathcal{S}$ for suitable chosen classes $\Theta$ and $\mathcal{S}$. Minimax-optimality of an estimator is then usually shown by establishing both an upper and a lower bound of the maximal risk over given classes $\Theta$ and $\mathcal{S}_\bullet$ The proposed estimation procedure relies on a multiplication of a thresholded inversion of the empirical process $\widehat{\mathfrak{s}}_\bullet$ and a projection of the empirical process $\widehat{g}_\bullet$ on the interval $[-m, m] \cap \mathcal{J}$ where the dimension parameter $m$ has to be chosen suitably. The upper bounds (both globally and locally) consist (up to the constants) of sums of two terms depending each on one of the sample sizes $n$ and $k$ only. The first summand depending on $n$ matches the lower bound in case the multiplicative function $\mathfrak{s}_\bullet$ is known in advance. The second summand represents the prize to pay for estimating $\mathfrak{s}_\bullet$ . Assuming $\mathcal{J}$ is countable we present a matching lower bound for global estimation, and a lower bound for local estimation featuring a gap. The proposed estimation procedures (as many others) necessitates the choice of a tuning parameter, which in turn, crucially influences the attainable accuracy of the constructed estimator. Its optimal choice, however, follows from a classical squared-bias-variance trade-off and relies on an a-priori knowledge about the classes $\Theta$ and $\mathcal{S}$, which is usually inaccessible in practice. We propose a fully data-driven choice of the dimension parameter by model selection and Goldenshluger-Lepski method for global and local estimation, respectively. We derive global and local oracle inequalities and discuss attainable rates of convergence as $k, n \to +\infty$ considering usual behaviours for $\theta_\bullet$ and $\mathfrak{s}_\bullet$.

- **Anatoli Juditsky:** On robust counterpart of linear inverse problems
  **Abstract:** We consider an *uncertain linear inverse problem* as follows. Given observation $\omega = Ax + \xi$ where $A \in \mathbf{R}^{m\times n}$ and $\xi \in \mathbf{R}^m$ is observation noise, we want to recover unknown signal $x$, known to belong to a convex set $\mathcal{X} \in \mathbf{R}^n$. As opposed to the "usual" setting of such problem, we suppose that sensing matrix $A$, feasible set $\mathcal{X}$, or noise $\xi$ may be *uncertain*. For instance, observation matrix may satisfy $A = A_0 + \delta A$ where the nominal matrix $A_0$ is known and $\delta A$ is unknown perturbation which may be random or belong to a given set $\mathcal{A}$, or observation noise may contain a deterministic or singular component, etc. In a series of problem settings, under various assumptions on the nature of problem uncertainty, we discuss the properties of two types of parameter estimates—linear estimates and *polyhedral estimates* (A particular class of nonlinear estimates as introduced in Juditsky, A., & Nemirovski, A. (2020). On polyhedral estimation of signals via indirect observations. *Electronic Journal of Statistics,* 14(1), 458-502.) We show that in the situation where the signal set is an uncertain *ellitope* (essentially, a symmetric convex set delimited by quadratic surfaces), nearly minimax optimal (up to a moderate suboptimality factor) estimates can be constructed by means of efficient convex optimization routine.
  Joint work with Y. Bekry, and A. Nemirovski.

- **Nicolas Klutchnikoff:** Adaptive estimation of the regression function with Brownian path covariates
  **Abstract:** In this paper, we are interested in estimating a regression function in the presence of functional covariates. More precisely, we wish to estimate the conditional expectation of a real response variable Y with respect to a standard Wiener coprocess $W$. Using the Wiener-Itô chaotic decomposition of $E(Y|W)$, we construct natural estimators and obtain minimax rates of convergence over specific regularity classes. We also define a selection procedure for some hyper-parameters, based on the Goldenshluger-Lepski method and obtain an oracle-like inequality and adaptive results.

- **Claire Lacour:** Semiparametric inference for mixtures of circular data
  **Abstract:** We consider a sample of data on the circle $S^1$, whose distribution is a two-components mixture. The density of the sample is assumed to be $g(x) = pf(x-a) + (1-p)f(x-b)$ where $p$ is the mixing parameter, f a density on the circle, and a and b two angles. The objective is to estimate both the parametric part $(p, a, b)$ and the nonparametric part $f$. We shall study the specific identifiability problems on the circle, which do not appear for real data. Next we shall present our adaptive estimation procedure, its theoretical performances and some numerical simulations
  Joint work with with T-M Pham Ngoc

- **Oleg Lepski:** Minimax estimation of nonlinear functionals
  **Abstract:** We deal with the problem of nonparametric estimating the $L_p$-norm, $p \in (1, \infty)$, of a probability density on $R^d, d \geq 1$, from independent observations. The unknown density is assumed to belong to a ball in the anisotropic Nikolskii's space. We adopt the minimax approach and demonstrate in particular that accuracy of estimation procedures essentially depends on whether $p$ is integer or not. Moreover, we develop a general technique for derivation of lower bounds on the minimax risk in the problems of estimating nonlinear

functionals. The proposed technique is applicable for a broad class of nonlinear functionals, and it is used for derivation of the lower bounds in the $L_p$-norm estimation.

- **Timothée Mathieu:** Robust Multivariate Mean estimation with M-estimators
**Abstract:** Mean estimation is a fundamental problem in statistics, as it is a tool on which a lot of the statistical procedures are based. In the well-controlled case of Gaussian random variables (or sub-gaussian random variables), it is known that the empirical mean perform fairly well. On the other hand, as soon as the distribution becomes either heavy-tailed or corrupted, things get complicated. This can be a major difficulty because in practice a lot of datasets contains outliers (typically in life sciences there are outliers in most datasets). Estimating the mean optimally for corrupted datasets is still unsolved, and most estimators are either theoretically optimal or computationally efficient but (for now) never both at the same time. In this presentation, I will present how to partially solve the problem with M-estimators that are computable and optimal for Heavy-tail distributions and I will explain the challenges to which we are confronted to design optimal and computable estimators.

- **Elias Ould Said:** Strong uniform consistency of the local linear error regression estimator under left truncation
**Abstract:** This paper is concerned with a nonparametric estimator of the regression function based on the local linear method when the loss function is the mean squared relative error and the data left truncated. The proposed method avoids the problem of boundary effects and is robust against the presence of outliers. Under suitable assumptions, we establish the uniform almost sure strong consistency with a rate over a compact set. A simulation study is conducted to comfort our theoretical result. This is made according to different cases, sample sizes, rates of truncation, in presence of outliers and a comparison study is made with respect to classical, local linear and relative error estimators. Finally, an experimental prediction is discussed.

- **Vincent Rivoirard:** Bayesian nonparametric inference for nonlinear Hawkes processes
**Abstract:** Hawkes processes are a specific class of point processes modeling the probability of occurrences of an event depending on past occurrences. Hawkes processes are therefore naturally used when one is interested in graphs for which the temporal dimension is essential. In the linear framework, the statistical inference of Hawkes processes is now well known. We will therefore focus more specifically on the class of nonlinear multivariate Hawkes processes that allow to model both excitation and inhibition phenomena between nodes of a graph. We will present the Bayesian nonparametric estimation of the parameters of the Hawkes model and the posterior contraction rates obtained on Hölder classes. From the practical point of view, since simulating posterior distributions is often out of reach in reasonable time, especially in the mutlivariate framework, we will more specifically use the variational Bayesian approach which provides a direct and fast computation of an approximation of the posterior distributions allowing the analysis in reasonable time of graphs containing several tens of neurons.
Joint work with Déborah Sulem and Judith Rousseau.

- **Angelina Roche:** Minimax rates in regression models for functional data
  **Abstract:** In recent decades, significant research efforts have focused on regression models that involve functional data, which are data that can be modeled as samples of random functions. The minimax rates for the functional linear model and the fully nonparametric model are now well understood, although some aspects of these models still requires further exploration. However, for other models, like the single index model or the models with sparsity, the minimax rates are still unknown. The objective of this presentation is to provide a brief overview of the current state of knowledge regarding these models, as well as ongoning research on them.

- **Flore Sentenac:** Robust Estimation of Discrete Distributions under Local Differential Privacy
  **Abstract:** Although robust learning and local differential privacy are both widely studied fields of research, combining the two settings is just starting to be explored. We consider the problem of estimating a discrete distribution in total variation from $n$ contaminated data batches under a local differential privacy constraint. A fraction $1-\alpha$ of the batches contain $k$ i.i.d. samples drawn from a discrete distribution $p$ over $d$ elements. To protect the users' privacy, each of the samples is privatized using an $\epsilon$-locally differentially private mechanism. The remaining $\alpha n$ batches are an adversarial contamination. The minimax rate of estimation under contamination alone, with no privacy, is known to be $\alpha/\sqrt{k}+\sqrt{d/kn}$. Under the privacy constraint alone, the minimax rate of estimation is $\sqrt{d^2/\epsilon^2 kn}$. We show, up to a $\sqrt{\log(1/\alpha)}$ factor, that combining the two constraints leads to a minimax estimation rate of $\alpha\sqrt{d/\epsilon^2 k}+\sqrt{d^2/\epsilon^2 kn}$, larger than the sum of the two separate rates. We provide a polynomial-time algorithm achieving this bound, as well as a matching information theoretic lower bound.