Abstracts (authors from alphabetical order)

Emmanuel Candes

Title: Conformal Prediction with Conditional Validity

Abstract: We consider the problem of constructing distribution-free prediction sets with finite-sample conditional guarantees. Prior work has shown that it is impossible to provide exact conditional coverage universally in finite samples. Thus, most popular methods only provide marginal coverage over the covariates. This paper bridges this gap by defining a spectrum of problems that interpolate between marginal and conditional validity. We motivate these problems by reformulating conditional coverage as coverage over a class of covariate shifts. When the target class of shifts is finite dimensional, we show how to simultaneously obtain exact finite sample coverage over all possible shifts. For example, given a collection of protected subgroups, our algorithm outputs intervals with exact coverage over each group. For more flexible, infinite dimensional classes where exact coverage is impossible, we provide a simple procedure for quantifying the gap between the coverage of our algorithm and the target level. Moreover, by tuning a single hyperparameter, we allow the practitioner to control the size of this gap across shifts of interest. Our methods can be easily incorporated into existing split conformal inference pipelines, and thus can be used to quantify the uncertainty of modern black-box algorithms without distributional assumptions.

Alexandra Carpentier

Title: Optimal ranking in crowd-sourcing

Abstract: Consider a crowd sourcing problem where we have n experts and d tasks. The average ability of each expert for each task is stored in an unknown matrix M, from which we have incomplete and noise observations. We make no (semi) parametric assumptions, but assume that both experts and tasks can be perfectly ordered: so that if an expert A is better than an expert B, the ability of A is higher than that of B for all tasks - and that the same holds for the tasks. This implies that if the matrix M, up to permutations of its rows and columns, is bi-isotonic.

We focus on the problem of recovering the optimal ranking of the experts in l2 norm, when the ordering of the tasks is known to the statistician. In other words, we aim at estimating the suitable permutation of the rows of M while the permutation of the columns is known. We provide a minimax-optimal and computationally feasible method for this problem, based on hierarchical clustering, PCA, change-point detection, and exchange of informations among the clusters. We prove in particular - in the case where d > n - that the problem of estimating the expert ranking is significantly easier than the problem of estimating the matrix M.

This talk is based on a joint ongoing work with Alexandra Carpentier and Emmanuel Pilliat.

Fabrice Gamboa

Title: Free deconvolution.

Abstract: One of Elisabeth's favorite scientific topics is deconvolution. We will first take a quick look at Elisabeth's results on this subject. We will then discuss some preliminary results on multiplicative free deconvolution. This joint work with R. Chhaibi (Toulouse), S. Kammoun (Toulouse) and M. Velasco (Bogota) deals with the estimation of a discrete measure

observed through a noisy realization of a spectral measure. We propose a statistical model, study its asymptotic properties and provide inversion tools based on complex analysis to recover the unknown measure.

Alice Guionnet
Title: About non-Bayesian optimal inference
Abstract: We discuss the problem of estimating a vector from measurements of its noisy observation. We consider Bayes and non-Bayes optimal inference for such rank 1 statistical estimation problem and establish a general replica symmetry breaking formula for these models based on large deviations for the associated overlaps. This is based on a joint work with Justin Ko, Florent Krzakala and Lenka Zdeborova.

Thi Thu Hoang
Title: Statistics applied to energy sector
Abstract: Electricity cannot be stored on a large scale. At each moment, the balance between consumption and production must be ensured. The management of the electrical system must be carried out at all time horizons: short-term for the production plan, medium-term for the security of supply and long-term for the investment decisions. We give here some examples of system management at different horizons. The first one concerns an hydrological tool where the goal is to simulate streamflows for the estimation of mid-term hydraulic supply. We use here the results of a thesis co-supervised with Elisabeth Gassiat on bivariate stochastic model of temperature and rainfall based on hidden Markov chain. The second one is about short-term consumption forecast where we show different challenges and diffents methods used to adapt to the fluctuation of electricity consumption.

Éric Moulines
Title: Conformal Prediction for Federated Uncertainty Quantification Under Label Shift
Abstract: Federated Learning (FL) is a machine learning framework where many clients collaboratively train models while keeping the training data decentralized. Despite recent advances in FL, the uncertainty quantification topic (UQ) remains partially addressed. Among UQ methods, conformal prediction (CP) approaches provide distribution-free guarantees under minimal assumptions.
We develop a new federated conformal prediction method based on quantile regression and take into account privacy constraints. This method takes advantage of importance weighting to effectively address the label shift between agents and provides theoretical guarantees for both valid coverage of the prediction sets and differential privacy. Extensive experimental studies demonstrate that this method outperforms current competitors.

Richard Nickl
Title: On the computational complexity of MCMC in high-dimensional non-linear regression models.
Abstract: We discuss recent results, both positive and negative, about the run-time of Markov chain Monte Carlo (MCMC) algorithms that target posterior distributions arising from high-dimensional non-linear statistical regression models with Gaussian process priors. Prototypical applications include inverse problems with partial differential equations (PDEs). We show that cold-start local MCMC may not work (have `exponential in dimension' runtime) even for target measures that are radially strictly decreasing away from their

unique mode, but that warm-start Langevin type MCMC can achieve polynomial runtime for posterior computation under certain `gradient stability' conditions that can be verified in a large class of relevant PDE models.

Judith Rousseau
Title: Scalable variational Bayes inference for multivariate Hawkes processes
Abstract: Multivariate nonlinear Hawkes processes are powerful models for multi- variate point processes with excitation and inhibation phenomenon. Bayesian nonparametric methods have been proposed and studied theoretically, show- ing good properties. However their implementation remain a challenge due to the complexity of the likelihood and the potentially high dimensional space. In this work we propose a two step variational Bayes approach to estimate both the graph of interaction and the functions of interactions. We give theoretical guarantees to the procedure and show that it scales well for moderately high dimensionnal Hawkes processes.
This is a joint work with Deborah Sulem and Vincent Rivoirard.

Adeline Samson
Title: Some statistical models to quantify the effect of climate change on whales in Greenland
Abstract: Human activities have a profound impact on marine ecology in Greenland. In this presentation, I will focus on a study of the impact of these activities on narwhals. I will present different stochastic models to analyze the data from this study: point process with memory, stochastic Langevin diffusions. These diffusion processes can be multidimensional, hypoelliptic (with a degenerate noise) and partially observed. I will discuss the question of parameter estimation when only discrete observations are available.

Jean-Christophe Thalabard
Title: STAFAV : une formation aux statistiques appliquées en Afrique francophone
Abstract: En 2003, le Pr. Maurice Tchuente, alors ministre de l'enseignement supérieur du Cameroun, a souhaité développer une offre moderne de niveau master de formation en mathématiques appliquées adaptées aux enjeux africains, avec en premier lieu les applications pour la santé humaine et son environnement, afin de fédérer les capacités existantes africaines et de former les nouvelles générations en partenariat/co-tutelle avec des formations existantes dans les départements universitaires français partenaires. Sous l'impulsion du Pr. Didier Dacunha-Castelle, et en étroite interaction avec des collègues d'Afrique francophone, le département de mathématique de l'université Paris-Sud a coordonné la mise en place du projet STAFAV de formation aux statistiques, d'abord au Cameroun puis au Sénégal et au Bénin, à laquelle ont contribué de nombreux collègues statisticiens français enseignants chercheurs et/ou chercheurs (Paris 4, Paris Dauphine, Paris 5, Telecom, INRAE, CIRAD, IRD, SPF, ...). Elisabeth Gassiat et Jean-Marc Bardet ont été des acteurs majeurs dans cette aventure qui, au-delà de la découverte des arcanes toujours complexes des relations franco-africaines marquées par le poids de l'histoire, a permis d'établir des collaborations originales entre disciplines sur des sujets d'importance majeure (paludisme, HIV, arboviroses, pathologies végétales, hydro-géologie pour n'en citer que quelques-uns), aboutissant à des travaux statistiques originaux répondant aux attentes du terrain et s'inscrivant pleinement dans les objectifs de développements durables tels que définis par l'OMS, tout en contribuant à former une jeune génération d'acteurs africains. La

présentation essaiera brièvement de retracer les grandes étapes de ce projet à travers les contributions notamment d'Elisabeth Gassiat. Elle nous donne l'occasion de saluer la mémoire de notre collègue et ami camerounais, le Pr. Henri Gwet, profondément associé au projet depuis son début.

Ramon van Handel
Title: Is there a nonlinear empirical process theory?
Abstract: The goal of empirical process theory, a classical workhorse of mathematical statistics and machine learning, is to understand the suprema of sums of independent random variables. One might wonder whether there is an analogous theory that enables one to understand the suprema of arbitrary nonlinear functions of independent random variables. The aim of my (somewhat speculative) talk is to suggest that recent advances in the metric theory of Banach spaces could shed new light on this question.

Aad van der Vaart
Title: Linear methods for nonlinear inverse problems.
Abstract: We consider the recovery of an unknown function f from a noisy observation u_f of the solution to a partial differential equation that can be written in the form Lu_f = c(f, u_f), for a differential operator L that is rich enough to recover f from Lu_f . We transform this problem into the linear inverse problem of recovering Lu_f, and show that Bayesian methods (with priors placed either on u_f or Lu_f) for this problem yield optimal recovery rates not only for u_f , but also for f . Joint work with Geerten Koers and Botond Szabó.