LPSM Kickoff Conference, Paris, June 18-20, 2018

Abstracts

Laurent Mazliak (LPSM)

The road to unity

The talk will present some aspects of Emile Borel's interest for the mathematics of randomness, leading him in the aftermath of the Great War to give an impulse for the creation of the Statistical Institute of the University of Paris in 1922 and the Henri Poincaré Institute in 1928. The new laboratory LPSM can be seen as a direct heir of both branches of Borel's institutional activity.

Yuval Peres (Microsoft)

Probability and Statistics: Gravitational allocation to uniform points and Trace reconstruction for the deletion channel

Given n uniform points on a two-dimensional sphere, how can we partition the sphere fairly among them ? It turns out that if the n points apply two-dimensional gravity to the rest of the sphere, then the basins of attraction yield such a partition with exactly equal areas.

See http://www.ams.org/publications/journals/notices/201705/rnoti-cvr1.pdf .

With Nina Holden and Alex Zhai, we proved that this partition minimizes, up to a bounded factor, the average distance between points in the same cell. I will also present an application to almost optimal matching of n uniform blue points to n uniform red points on the sphere.

In the second (unrelated) part of the talk, I will discuss the trace reconstruction problem: Suppose that an unknown string x of n bits is observed through the deletion channel; how many independent outputs (traces) of this channel are needed to reconstruct x with high probability? The best lower bound known is linear in n. Currently, the best upper bound is exponential in the cube root of n, proved with F. Nazarov, STOC 2017 (and independently by De, O Donnell and Servedio). If the string x is random, we showed (in joint works with Zhai, Holden and Pemantle) that a subpolynomial number of traces suffices, by comparison to a random walk.

Christina Goldschmidt (Oxford)

Voronoi cells in the Brownian continuum random tree

Take a uniform random tree with n vertices and select k of those vertices independently and uniformly at random; call them sites. (We assume that n is large and k is fixed, so that with high probability the sites are distinct.) Find the associated Voronoi cells: for each vertex in the tree, assign it to the cell of the site (or sites) which is closest in the graph distance. Now consider the vector of the proportions of the vertices lying in each of the k cells. We prove that this vector converges in distribution to the Dirichlet(1,1,. . . ,1) distribution (that is, it is asymptotically uniform on the (k-1)-dimensional simplex). In fact, this is most easily formulated as a result about the scaling limit of the uniform random tree, namely the Brownian continuum random tree: if we pick k independent sites from the mass measure of the tree, their Voronoi cells have masses which are jointly Dirichlet(1,1,. . . ,1) distributed. An analogue of this result also holds for (the scaling limit of) uniform unicellular random maps on surfaces of arbitrary genus. Joint work with Louigi Addario-Berry, Omer Angel, Guillaume Chapuy and Éric Fusy.

Susan Holmes (Stanford)

Statistical challenges from the study of the human microbiome

The human microbiome is a complex assembly of bacteria that are sensitive to many perturbations. We have developed specific tools for studying the vaginal, intestinal and oral microbiomes under different perturbations (pregnancy, hypo-salivation inducing medications and antibiotics are some examples).We will show tools we have developed for analysing longitudinal multi-table data composed of 16s rRNA reads combined with clinical data, transcriptomic and metabolomic profiles. Challenges we have addressed include information leaks, the integration of phylogenetic information,longitudinal dependencies? ?and uncertainty quantification.Our methods enable the detection of ecological gradients and their uncertainty quantificationas well as the integration of tree-aware multivariate representations.

This contains joint work with Joey McMurdie, Lan Huong Nguyen, Pratheepa Jeganathan,Sergio Bacallado, Ben Callahan, Julia Fukuyama, Kris Sankaran, Claire Donnat, and David Relman's Lab members from Stanford.

Anton Thalmaier (Luxemburg)

Characterization of Ricci curvature and Ricci flow by Brownian motion

The observation of Aaron Naber that bounded Ricci curvature on a Riemannian manifold controls the analysis on path space, in a manner analogous to how lower Ricci curvature bounds control the analysis on the manifold, gave new impetus to the field. Certain gradient estimateson path space turn out to be equivalent to bounded Ricci curvature. We present recent work on characterizing Ricci curvature and Ricci flow in terms of functional inequalities for heat semigroups on manifolds. The inequalities are strong enough to characterize in particular Einstein manifolds and Ricci solitons. The talk includes extensions of these methods to geometric flows on manifolds, as well as to the path space of Riemannian manifolds evolving under a geometric flow.

Olivier Bousquet (Google)

Deep Learning and Generalization

Deep Learning is at the heart of the current AI revolution. It is based onsimple principles that date back from the 80s but are still poorly understood. Recently a lot of work has been done on trying to explain some of the surprising phenomena observed when training and using deep neural networks. But a lot of their properties remain mysterious. This talk will present some of these phenomena and what is currently understood about them. It will also highlight the questions that remain open and that can be key to the future developments of the field.

Amir Dembo (Stanford)

Averaging principle and shape theorem for growth with memory

We consider a family of random growth models in n-dimensional space.These models capture certain features expected to manifest at the mesoscopic level for certain self-interacting microscopic dynamics (such as once-reinforced random walk with strong reinforcement and origin-excited random walk). In a joint work with Pablo Groisman, Ruojun Huang and Vladas Sidoravicius, we establish for such models an averaging principle and deduce from it the convergence of the normalized domain boundary, to a limiting shape. The latter is expressed in terms of the invariant measure of an associated Markov chain.

Martin Hairer (Imperial College)

A new universality class for 1+1 dimensional dynamic

Lai-Sang Young (NYU)

Comparing chaotic and random dynamical systems

In this talk I will compare and contrast (deterministic) chaotic dynamical systems and their stochastic counterparts, i.e. when small random perturbations are added to such systems to model uncontrolled fluctuations. Three groups of results, some old and some new, will be discussed. The first has to do with how deterministic systems, when sufficiently chaotic, produce observations resembling those from genuinely random processes. The second compares the ergodic theories of chaotic systems and of random maps (as in stochastic flows of diffeomorphisms generated by SDEs). One will see that results on SRB measures, Lyapunov exponents, entropy, fractal dimension, etc. are all nicer in the random setting. I will finish by suggesting that to improve the applicability of existing theory of chaotic systems, a little bit of random noise can go a long way.

Philip Protter (Columbia)

Semimartingale decompositions under a continuous expansion of the filtration

We are concerned with the role of information plays in various settings, including the financial markets. A few years ago, Younes Kchia and the speaker developed a model for expanding a filtration continuously with new information, for example coming from an external stochastic process. Sufficient conditions were obtained for a semimartingale to remain a semimartingale under the expansion, but the techniques did not supply access to the new resulting decomposition. Within a Brownian paradigm examples showed that the new resultant finite variation terms could or could not remain absolutely continuous with respect to dt.In new work with Léo Neufcourt we show how one can access the new semimartingale decompositions and determine when the finite variation terms retain the absolute continuity of their paths. Such properties are key to a use of Girsanov's theorem and the absence of arbitrage opportunities in financial markets. Various possible applications are discussed. One of interest in Math Finance is how these new techniques give a way to model insider trading and the resulting advantage the insider can gain.This talk is based on joint work with Léo Neufcourt of Michigan State University.

Sara van de Geer (ETH)

Some concentration results for the LASSO

In this talk we aim at presenting tight bounds for the prediction error of the Lasso. We consider the linear model

$$Y = X\beta^0 + \epsilon,$$

with $Y \in \mathbb{R}^p$ a vector of response variables, $X \in \mathbb{R}^{n\times p}$ a design matrix, $\beta^0 \in \mathbb{R}^p$ a vector of unknown regression coefficients, and $\epsilon \sim \mathcal{N}(0, I)$ unobservable noise. The Lasso is

$$\hat{\beta} = \arg\min\bigg\{ \|Y - Xb\|_2^2 + 2\lambda\|b\|_1 \bigg\}.$$

Let $S_0 := \{j : \ \beta_j^0 \neq 0\}$ be the active set of β^0 and $s_0 := |S_0|$ its size.

We first consider random design. Suppose that the rows of X are i.i.d. Gaussian random vectors and let $\Sigma_0 := \mathbb{E}X^T X/n$. Let

$$\beta^* \in \arg\min\Big\{ n\|\Sigma_0^{1/2}(b-\beta^0)\|_2^2 + 2\lambda\|b\|_1 \Big\}.$$

We then show for proper choice of the tuning parameter $\lambda \asymp \sqrt{n\log p}$, with probability at least $1-2\exp[-x]$,

$$\|X(\hat\beta-\beta^*)\|_2 \le \gamma\sqrt{n}\|\Sigma_0^{1/2}(\beta^*-\beta^0)\|_2 + 4\sqrt{\log 2} + \sqrt{2x},$$

where γ depends on the largest eigenvalue $\Lambda_{\max}^2$ of Σ_0 and on the compatibility constant. If $\Lambda_{\max}^2 = o(\log p)$ then $\gamma = o(1)$ in a large number of cases (e.g. in regimes where the compatibility constant stays away from zero and $s_0 \log p/n = o(1)$). Thus, the squared “bias" $n\|\Sigma_0^{1/2}(\beta^*-\beta^0)\|_2^2$ of the Lasso dominates the “variance" $\|X(\hat\beta-\beta^*)\|_2^2$.

We will present the definition of the compatibility constants and discuss their relation with restricted eigenvalues.

For the case of fixed design, we establish upper and lower bounds for the prediction error. As an example, we consider the model $Y = f^0 + \epsilon$ and the least squares estimator with total variation penalty

$$\hat f = \arg\min_{f\in\mathbb{R}^n}\Big\{ \|Y-f\|_2^2 + 2\lambda \mathrm{TV}(f) \Big\},$$

where $\mathrm{TV}(f) := \sum_{i=2}^n |f_i - f_{i-1}|$. We prove in this case that the gap between upper and lower bounds for $\|\hat f - f^0\|_2$ is with probability at least $1-\exp[-x]$ at most

$$\lambda\sqrt{(1+s_0)\log p/n} + 4\sqrt{\log 2} + 2\sqrt{2x}.$$

Here s_0 is the number of jumps of f^0. We assume here that $\lambda \asymp \sqrt{n\log p}$ is suitably chosen and that the jumps of f^0 are sufficiently large.

Sebastian Roch (Wisconsin-Madison)

From genomes to evolutionary trees and beyond

The reconstruction of the Tree of Life is a classical problem in evolutionary biology that has benefited from numerous branches of mathematics, including probability, information theory, combinatorics, and geometry. Modern DNA sequencing technologies are producing a deluge of new genetic data -- transforming how we view the Tree of Life and how it is reconstructed. I will survey recent progress on some statistical questions that arise in this context.

Perla Sousi (Cambridge)

Capacity of random walk and Wiener sausage in 4 dimensions

In four dimensions we prove a non-conventional CLT for the capacity of the range of simple random walk and a strong law of large numbers for the capacity of the Wiener sausage.

This is joint work with Amine Asselah and Bruno Schapira.

Chris Rogers (Cambridge)

Economics: science or sudoku?

When we are ill, most of us would prefer to receive treatment that was supported by scientific evidence, rather than anecdotal tradition or superstition. When a nation's economy is ill, policy-makers turn to economists for advice, but how well is their advice supported by evidence? This talk critiques the value of economic theory in practice, and tries to suggest ways of increasing the practical relevance of the subject.

Bin Yu (UC Berkeley)

Three principles of data science, predictability, computability, and stability

In this talk, I'd like to discuss the importance and connections of three principles of data science in the title and introduce the PCS workflow for the data science life cycle. PCS will be demonstrated in the context of two collaborative projects in neuroscience and genomics, respectively. The first project in neuroscience uses transfer learning to integrate fitted convolutional neural networks (CNNs) on ImageNet with regression methods to provide predictive and stable characterizations of neurons from the challenging primary visual cortex V4. Our DeepTune characterization provides a rich description of the diverse V4 selection patterns. The second project proposes iterative random forests (iRF) as a stablized RF to seek predictable and interpretable high-order interactions among biomolecules. For an enhancer status prediction problem for Drosophila based on high-throughput data, iRF was able to find 20 stable gene-gene interactions, of which 80% had been physically verified in the literature in the past few decades. Last but not least, the data results from both projects provide experimentally testable hypotheses and hence PCS can also serve as a scientific recommendation system for follow-up experiments.

Bernard Derrida (Collège de France)

Renormalization and disorder: a simple toy model

The problem of the depinning transition of a line from a random substrate is one of the simplest problems in the the theory of disordered systems. It has a long history among physicisits and mathematicians. Still there are many open questions about the nature of this transition. After a brief review of our present understanding of the problem, I will discuss a simple tree-like toy model which indicates that, when disorder is relevant, the depinning transition becomes an infinite order transition of the Kosterlitz Thouless type. I will also try to present some recent developments allowing to understand how the precise nature of the singularity at the tansition depends on the distribution of disorder.