High Dimensional Probability VI

Richard M. Dudley (Massachusetts Institute of Technology),
Christian Houdré (Georgia Institute of Technology),
David Mason (University of Delaware),
Jan Rosinski (University of Tennessee),
Jon A. Wellner (University of Washington).

October 9 – 14, 2011

1 Overview of the Field

Problems in probability theory increasingly involve high dimensions either in the basic sample spaces or in the dimensionality of the classes of functions or sets involved. In statistical problems, this occurs via the vast and fast data collection made possible by new generations of instrumentation in areas as diverse as microarray data in genetics, cosmic background microwave radiation measurements in astronomy, and new imaging methods in medicine and biophysics. Similarly, real world problems of interest in combinatorial optimization are high-dimensional in nature.

High Dimensional Probability is an area of mathematical research with deep roots in the classical limit theorems of probability. The problems of major focus to researchers in this area initially arose from the study of limit theorems in infinite-dimensional spaces such as Hilbert spaces, Banach spaces and normed linear spaces. Interest and motivation for such results goes back at least to the work of Glivenko-Cantelli, Kolmogorov-Smirnov and Donsker on the large sample behavior of empirical distribution functions.

The desire to provide a rigorous framework for the derivation and understanding of results of this type was the primary impetus for the development of a general theory of limit theorems in infinite dimensions. This goal has in large part been achieved.

But the most remarkable feature of this program is that it has resulted in the creation of powerful new tools and perspectives, whose range of application has extended far beyond its original purpose, and with an increasing number of interactions with other areas of mathematics, statistics, and computer science. These include additive combinatorics, random matrix theory, nonparametric statistics, empirical process theory, statistical learning theory, strong and weak approximations, distribution function estimation in high dimensions, combinatorial optimization, and random graph theory. For example, lately we have seen surprising connections between high dimensional probability and dependent sequences in additive combinatorics, and between concentration of measure inequalities and statistics.

High dimensional probability theory continues to develop innovative new tools, methods, techniques and perspectives to analyze the increasingly manifold phenomena happening around us.

2 Outcome of the Meeting

The primary objectives of this workshop were:

(1) To bring together experts in high dimensional probability and those in a number of the "areas of strong interaction" to discuss some of the major problems in this area and report on progress towards their solution.

(2) To facilitate interactions and communications between the experts actively involved in the development of new theory in high dimensional probability, and leading researchers in statistics, machine learning, and computer science. Our intention is to deepen contacts between several different communities with common research interests focusing on probability inequalities, empirical processes, strong approximations, Gaussian and related chaos processes of higher order, Markov processes, and applications of these methods to a wide range of problems in other areas of mathematics or to applications in statistics, optimization theory, and machine learning.

(3) To foster and develop interest in this area of research. There are many interesting and exciting problems, which can be formulated in a way that can be understood by graduate students, postdoctoral students, and new researchers.

Our goal was to focus on the following areas:

A. Applications of concentration of measure results and methods to random matrices.

B. Concentration of measure inequalities.

C. Interactions between small ball probabilities, approximation theory, prior distributions for nonparametric Bayes procedures, entropy bounds for high-dimensional function classes.

D. Applications of high dimensional probability methods to analysis and stochastic analysis.

E. Applications of modern empirical process and strong approximation methods to treat problems in machine learning, nonparametric estimation and inference, with a particular focus on high- and infinite-dimensional statistical models.

F. Identification of major problems and areas of potentially high impact for applications and use in other areas of mathematics, statistics, and computer science.

Based on the presented results and discussion at the meeting, we believe that we more met our goal. The followup proceedings volume of the meeting is planned to be completed in 2012.

3 Presentation Highlights

Our participants presented talks on a wide variety of subjects, which stimulated a number of lively discussions that will likely lead to new research and collaborations. We have attached the their abstracts, as well as a division of them by subject.

Division of Abstracts by Subject

Random Matrices

Adamczak: *The Circular Law for random matrices with independent log-concave rows*
Eichelsbacher: *Universal moderate deviations for the eigenvalue counting function of Wigner matrices*
Mark Meckes: *Concentration and convergence rates for spectral measures of random matrices*

Small Deviations

Aurzada: *Small deviation probabilities of Gaussian processes and path regularity*
Li: *Recent Developments on Small Value Probabilities*

Infinitely Divisible and Related Processes

Basse-O'Connor and Rosinski: *On the uniform convergence of random series in Skorohod space and representations of cadlag infinitely divisible processes*
Figueroa-Lopez: *Small-time expansions for local jump-diffusions models with infinite jump activity*

Partial Sums, Self-Normalized Sums and Processes

Kevei and Mason: *Self-Normalized Sums and Self-Normalized Levy Processes*

Central Limit Theorems and Density Function Estimation

Random Structures with Gaussian Components

Tail Estimates

Invariance Principles and Approximations

Convexity and Its Applications

Nonparametric Estimation and Inference

Problems motivated by Mathematics Physics

Special Processes

Abstracts

The Circular Law for random matrices with independent log-concave rows

Radosław Adamczak (University of Warsaw)

I will show how the replacement principle by Tao and Vu together with Klartag's thin shell inequality and simple bounds for the smallest singular value allow for a relatively easy proof of the circular law for the class of matrices mentioned in the title.

Small deviation probabilities of Gaussian processes and path regularity

Frank Aurzada (Technische Universität Berlin)

We study the sample path regularity of Gaussian processes and relate it directly to the theory of small deviations. For example, we show that if the path of a centered Gaussian process is n-times differentiable then the exponential rate of decay of its L_∞-small deviations is at most $\varepsilon^{-1/n}$. We also show a similar result if n is not an integer. Further generalizations are given, which parallel the entropy method – which is recalled in this talk – to determine the small deviations.

On the uniform convergence of random series in Skorohod space and representations of càdlàg infinitely divisible processes

Andreas Basse-O'Connor (Aarhus University)

Let X_n be independent random elements in the Skorohod space $D([0,1];E)$ of càdlàg functions taking values in a separable Banach space E. Let $S_n = \sum_{i=1}^n X_i$. We show that if S_n converges in finite dimensional distributions to a càdlàg process, then $S_n + c_n$ converges a.s. uniformly over $[0,1]$ for some $c_n \in D([0,1];E)$. This result extends the Itô-Nisio Theorem to the space $D([0,1];E)$, which is surprisingly lacking in the literature even for $E = \mathbf{R}$. The main difficulties of dealing with $D([0,1];E)$ in this context are its non-separability under the supremum norm and the discontinuity of the addition under Skorohod's J_1-topology.

We use this result to prove the uniform convergence of various series representations of càdlàg infinitely divisible processes. As a consequence, we obtain explicit representations of the jump process, and of related path functionals, in a general non-Markovian setting. Finally, we illustrate our results on an example of stable processes. To this aim we obtain criteria for such processes to have càdlàg modifications, which may be of independent interest.

Renormalization in the model of Brownian motions in Poissonian potentials

Xia Chen (University of Tennessee)

The model of Brownian motion in Poissonian potential describes a typical trajectory of a Brownian particle surviving from being attracted by the obstacles randomly located in the space (think about the stars in the universe). In the existing literature, the random potential is defined as the convolution between a Poissonian field and a bounded and locally supported function. According to the Newton's law of universal attraction and some other related laws in physics, the most natural way of constructing the random potential is to define it as the Riesz potential of the Poissonian field. On the other hand, the Riesz potential of the Poissonian field blows up. In this talk, this problem will be fixed by the way of renormalization. In addition, some asymptotic patterns of our models will be established and more problems will be asked. Part of the talk is based on some collaborative works with Kulik and Rosinski and Xiong.

The almost sure invariance principle for unbounded functions

Jérôme Dedecker (Université Paris Descartes)

We consider two classes of piecewise expanding maps T of $[0,1]$: a class of uniformly expanding maps for which the Perron-Frobenius operator has a spectral gap in the space of bounded variation functions, and a class of expanding maps with a neutral fixed point at zero. In both cases, we give a large class of unbounded functions f for which the partial sums of $f \circ T^i$ satisfy an almost sure invariance principle. This class contains piecewise monotonic functions (with a finite number of branches) such that:

- For uniformly expanding maps, they are square integrable with respect to the absolutely continuous invariant probability measure.
- For maps having a neutral fixed point at zero, they satisfy an (optimal) tail condition with respect to the absolutely continuous invariant probability measure.

This is a joint work with Sébastien Gouëzel (université Rennes 1) and Florence Merlevède (université Paris Est-Marne la vallée).

Uniform-in-Bandwidth Functional Limit Laws for the Empirical Process and Applications

Paul Deheuvels (LSTA - Université Pierre et Marie Curie)

The functional limit laws for increments of the uniform empirical process, due to Deheuvels and Mason (1992) and Deheuvels (1992), allow to describe the limiting behavior of maximal deviations of nonparametric functional estimators. Deheuvels and Ouadah (2011) have established the following uniform-in-bandwidth in probability version of these results. Denoting by $\{\alpha_n(t) : 0 \le t \le 1\}$ the usual uniform empirical process, set $\xi_n(h;t;u) := \alpha_n(t+hu) - \alpha_n(t)$, for $0 < h < 1$, $0 \le u \le 1$ and $0 \le t \le 1-h$. Set $\log_+ v := \log(v \vee e)$, and let, for each $n \ge 1$, $\mathcal{F}_n(h) := \{(2h\log_+(1/h))^{-1/2}\xi_n(h;t;\cdot) : 0 \le t \le 1-h\}$. Denote by $\mathcal{S}$ the unit ball of the reproducing kernel Hilbert space pertaining to the Wiener process on $[0,1]$, and let $\Delta(\cdot,\cdot)$ denote the Hausdorff set-distance induced by the sup-norm of bounded functions on $[0,1]$. Set $\mathcal{J}_n := [a_n, b_n]$, where $a_n \le b_n$ are such that

$$na_n/\log n \to \infty \quad \text{and} \quad b_n \to 0.$$

Deheuvels and Ouadah (2011) showed that, under these assumptions, as $n \to \infty$,

$$\sup_{h \in \mathcal{J}_n} \Delta\left(\mathcal{F}_n(h), \mathcal{S}\right) \xrightarrow{\mathbb{P}} 0.$$

We extend this result to multivariate empirical processes. As an example of application, we consider an i.i.d. sequence $X_1, X_2, \ldots$ of $\mathbb{R}^d$-valued random vectors with continuous density $f(\cdot)$ on a neighborhood of $\mathcal{I} := \prod_{j=1}^d [c_j, d_j]$, with $c_j < d_j$ for $j = 1, \ldots, d$. Letting $K(\cdot)$ denote a function of bounded variation, and bounded support, integrating to 1 on $\mathbb{R}^d$, we consider the kernel estimator of $f(\cdot)$ defined by

$$f_{n,h}(x) := \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x - X_i}{h^{1/d}}\right) \quad \text{for} \quad x \in \mathbb{R}^d.$$

We will show that our limit law implies that, under the assumptions above, as $n \to \infty$,

$$\sup_{h \in \mathcal{J}_n}\left|\sup_{x \in \mathcal{I}}\left\{\frac{nh}{2\log_+(1/h)}\right\}^{1/2} \pm \left\{f_{n,h}(x) - E(f_{n,h}(x))\right\} - \left\{\sup_{x\in\mathcal{I}} f(x)\int_{\mathbb{R}^d} K^2(t)dt\right\}^{1/2}\right| \xrightarrow{\mathbb{P}} 0.$$

How Long will it Take?

Victor de la Pena (Columbia University)

In this talk we introduce an approach to estimate the first hitting time of random processes. The approach provides a natural extension of the concept of boundary crossingby non-random functions to the case of stochastic processes. Two examples are provided: one motivated by cancer research, and the other on dealing with drought predictions for the Southwest US and the Mediterranian region. This is Joint work with Mark Brown, Yochanan Kushnir and Tony Sit.

Universal moderate deviations for the eigenvalue counting function of Wigner matrices

Peter Eichelsbacher (Ruhr-University Bochum)

We establish a moderate deviation principle (MDP) for the number of eigenvalues of a Wigner matrix in an interval. The proof relies on fine asymptotics of the variance of the eigenvalue counting function of GUE matrices due to Gustavsson. The extension to large families of Wigner matrices is based on the Tao and Vu Four Moment Theorem and applies localization results by Erdös, Yau and Yin. Moreover we investigate families of covariance matrices as well. This is joint work with Hanna Döring.

Small-time expansions for local jump-diffusions models with infinite jump activity

José E. Figueroa-López (Purdue University)

We consider a Markov process X with deterministic initial condition x, which is the solution of a stochastic differential equation driven by a Lévy process Z and an independent Wiener process W. Under some regularity conditions, including non-degeneracy of the diffusive and jump components of the process as well as smoothness of the Lévy density of Z, we obtain a small-time second-order polynomial expansion in time for the tail distribution and the transition density of the process X. The method of proof combines a recent approach by Figueroa-López and Houdré (2009) for regularizing the tail distribution of a Lévy process with classical results of Malliavin calculus for purely-jump processes, which have to be extended here to deal with the mixture model X. As an application, the leading term for out-of-the-money option prices in short maturity under a local jump-diffusion model is also derived. This is a joint work with Cheng Ouyang.

Rates of Contraction for Posterior Distributions in L^r-metrics, $1 \le r \le \infty$

Evarist Giné and Richard Nickl (University of Connecticut and University of Cambridge)

The frequentist behavior of nonparametric Bayes estimates, more specifically, rates of contraction of posterior distributions to shrinking L^r-norm neighborhoods of the unknown parameter, $1 \leq r \leq \infty$, are studied. A theorem for nonparametric density estimation is proved under general assumptions on the prior. The result is applied to a variety of common examples, including Gaussian processes and wavelet series. The rates of contraction are minimax-optimal for $1 \leq r \leq 2$, but deteriorate as r increases beyond 2. In the case of Gaussian nonparametric regression, a Gaussian prior is devised for which the posterior contracts at the optimal rate in all L^r-norms, $1 \leq r \leq \infty$. Modern theory of Gaussian processes, including lower bounds for small ball probabilities, is used.

Asymptotic Approximations in the CLT in Free and Classical Probability and Applications

Friedrich Götze (Bielefeld University)

In the classical CLT expansions of the entropy distance to the class of normal distribution are shown assuming the existence of moments and entropy of sums only. This extends Barron's (1986) CLT in the entropic distance to higher order approximations assuming minimal conditions. Explicit bounds for the entropic distance of order $O(n^{-1})$ (assuming four moments) are shown. Extensions to the multidimensional and stable case and relations to results e.g. by Rio using Wasserstein distances are discussed as well.

For comparison we discuss asymptotic approximations of first and second order in the CLT of free probability assuming four moments and a corresponding expansion of Voiculescu's free Entropy. This is joint work with G. Chistyakov and S. Bobkov.

Concentration of measure and optimal transport

Nathael Gozlan (Université Paris Est - Marne la Vallée)

This talk is devoted to Talagrand's transport entropy inequality $\mathbf{T}_2$. First we will recall its connection to Gaussian dimension free concentration of measure. Then we will present a new result showing that Talagrand's inequality is equivalent to a modified Log-Sobolev inequality. This latter result improves a celebrated result by Otto and Villani. Joint work with C. Roberto and P-M Samson.

Tail Bounds and Extremal Behavior of Light-tailed Perpetuities

Paweł Hitczenko (Drexel University)

We study the tail and extremal behavior of a sequence of random variables (R_n) defined by the recurrence $R_n = M_n R_{n-1} + Q_n$, $n \geq 1$, where R_0 is arbitrary, (M_n, Q_n) are iid copies of a non–degenerate random vector (M, Q) satisfying $0 \leq M \leq 1$, and $0 \leq Q \leq q$. The imposed conditions guarantee that (R_n) converge in distribution to a random variable R, usually referred to as perpetuity. We provide an upper bound for the tail of a limiting random variable R. Our bound is similar in nature to a lower bound obtained under the additional assumption that $Q \equiv q > 0$ by Goldie and Grübel (1996). Furthermore, we apply our result to obtain some information on the extremal behavior of the sequence (R_n). Specifically, we show that when $Q \equiv q > 0$ then under mild and natural conditions on M the suitably normalized extremes of (R_n) converge in distribution to a double exponential random variable. This partially complements a result of de Haan, Resnick, Rootzén, and de Vries (1989) who considered extremes of the sequence (R_n) under the assumption that $P(M > 1) > 0$.

Slepian's Inequality and Stochastic Orderings

Jorgen Hoffmann-Jorgensen (University of Aarhus)

It is well-known that Slepian's inequality is of great importance in the theory of Gaussian processes. There exists many forms of Slepian's inequality in the literature. One of the most general versions can be found in Ledoux and Talagrand's book (Probability in Banach spaces), but theorem is wrong as it stands (this can be shown by easy counterexamples). It is well-known that Slepian's inequality is closely related to various stochastic orderings; for instance the supermodular ordering. Recall that $f : R^k \to R$ is supermodular if $f(x \vee y) + f(x \wedge y) \leq f(x) + f(y)$ and if $X = (X_1, \ldots, X_k)$ and if $Y = (Y_1, \ldots Y_k)$ are random vectors, we write $X \leq_{\mathrm{sm}} Y$ if $Ef(X) \leq Ef(Y)$ for all supermodular functions f for which the expectations exist. In the modern literature it is generally stated that it suffices to verify this inequality for bounded supermodular functions. In the talk, I shall show that this is correct if $k \leq 2$ but wrong if $k \geq 3$. One commonly stated consequence of Slepian's inequality says that if X and Y are Gaussian vectors such that var $X_i = \mathrm{var} Y_i \ \forall i$ and cov $(X_i, X_j) \leq \mathrm{cov}(Y_i, Y_j) \ \forall i \neq j$, then $X \leq_{\mathrm{sm}} Y$. However., this result relies on the wrong statement above. This means that the statement is correct if $k \leq 2$ but I suspect that it fails if $k \geq 3$. The correct versions of Slepian's inequality involves 3 types of conditions (a): Smoothness (twice or more times differentiability); (b): Growth conditions (polynomial or exponential); (c): Regularity of the covariance matrix. In the talk I shall present a version of Slepian's inequality which covers all the correct versions I know, including Fernique's version (where covariance are replaced by intrinsic metric). The minimal conditions, under which I have been able to establish Slepian's inequality allows $f(x)$ to grow as e^{ax^2} for some $a > 0$, it do not require differentiability (it allows f to be an indicator function) but it do requires a weak type of approximate, directional continuity together with a condition on the range of the covariance matrices (which becomes redundant if f is sufficiently directional continuous).

Self-Normalized Sums and Self-Normalized Lévy Processes

Péter Kevei (University of Szeged)

Let $X, X_1, X_2, \ldots$, and $Y, Y_1, Y_2, \ldots$ be iid random variables and non-negative iid random variables, respectively, independent of each other. Define the self-normalized sum $R_n = \sum_{i=1}^n X_i Y_i / \sum_{i=1}^n Y_i$. We investigate the possible limiting distributions of R_n along the subsequences of the natural numbers. In particular, we show that if Y is in the centered Feller-class and $E|X| < \infty$, then all the possible subsequential limits of R_n has a C^∞ density function. We also consider the process version of the problem. This is an ongoing joint work with David Mason.

A hyperplane inequality for measures of convex bodies

Alexander Koldobsky (University of Missouri-Columbia)

The hyperplane problem asks whether there exists an absolute constant C so that for any origin-symmetric convex body K in R^n

$$|K|^{\frac{n-1}{n}} \leq C \max_{\xi \in S^{n-1}} |K \cap \xi^\perp|, \tag{1}$$

where $\xi^\perp$ is the central hyperplane in R^n perpendicular to ξ, and $|K|$ stands for volume of proper dimension. The problem is still open, with the best-to-date estimate $C \sim n^{1/4}$ established by Klartag, who slightly improved the previous estimate of Bourgain. In the case where the dimension $n \leq 4$, the inequality (1) was proved with the best possible constant $C = \left|B_2^n\right|^{\frac{n-1}{n}} / \left|B_2^{n-1}\right|$ (this constant is less than 1), where B_2^n is the unit Euclidean ball in R^n. In this talk we show that the latter result can be extended to arbitrary measures in place of volume. Namely, if $n \leq 4$, K is an origin-symmetric convex body in R^n, and μ is a measure in R^n with non-negative continuous density f (so that $\mu(K) = \int_K f$) then

$$\mu(K) \leq \frac{n}{n-1} \frac{\left|B_2^n\right|^{\frac{n-1}{n}}}{\left|B_2^{n-1}\right|} \max_{\xi \in S^{n-1}} \mu(K \cap \xi^\perp) \, |K|^{1/n}.$$

The constant is sharp, and the inequality holds true in higher dimensions under an additional assumption that K is an intersection body.

Complexity Penalization in Low Rank Matrix Recovery

Vladimir Koltchinskii (Georgia Institute of Technology)

We will discuss a problem of estimation of a large Hermitian $m \times m$ matrix A based on a finite number of measurements of randomly picked linear functionals of this matrix. In the noiseless case, the goal is to recover the target matrix exactly with a high probability based on the number of measurements of the order $m\mathrm{rank}(A)$ (up to constants and log-factors). Among the examples of this problem are the matrix completion (when the functionals are just the entries of the matrix) and the estimation of a density matrix in quantum state tomography. Recent results of Candes and Tao (2010) and Gross (2011) show that, in the case of matrix completion, such a recovery is possible provided that the target matrix satisfies certain "low coherence" conditions. The method is based on minimizing the nuclear norm over the affine space of matrices that agree with the data. We will be more interested in the case of noisy recovery (matrix or trace regression) where various versions of penalized least squares method can be used with complexity penalties based either on the nuclear norm, or on von Neumann entropy (in the case of quantum state tomography). We will discuss recent results of Koltchinskii (2010, 2011), Koltchinskii, Lounici and Tsybakov (2011) on oracle inequalities showing how the error of such estimators depends on the rank of the target matrix and other parameters of the problem. The proofs of these results are based on a variety of tools, including matrix versions of Bernstein inequalities and generic chaining bounds.

A CLT for Empirical Processes and Empirical Quantile Processes Involving Time Dependent Data

James Kuelbs (University of Wisconsin-Madison)

We establish empirical quantile process CLT's based on n independent copies of a stochastic process $\{X_t : t \in E\}$ that are uniform in $t \in E$ and quantile levels $\alpha \in I$, where I is a closed sub-interval of $(0, 1)$. Typically $E = [0, T]$, or a finite product of such intervals. Also included are CLT's for the empirical process based on $\{I_{X_t \leq y} - \mathrm{Pr(X_t \leq y) : t \in E, y \in R}\}$ that are uniform in $t \in E, y \in R$. The process $\{X_t : t \in E\}$ may be chosen from a broad collection of Gaussian processes, compound Poisson processes, stationary independent increment stable processes, and martingales.

Tail estimates for sums of independent log-concave random vectors

Rafał Latała (University of Warsaw)

We will present new tail estimates for order statistics of sums of independent log-concave vectors and show how they may be applied to derive deviation inequalities for l_r norms and norms of projections of such vectors. Part of the talk is based on the joint work with Radosław Adamczak, Alexander Litvak, Alain Pajor and Nicole Tomczak-Jaegermann.

Recent Developments on Small Value Probabilities

Wenbo Li (University of Delaware)

Small value probabilities or small deviations study the decay probability that positive random variables behave near zero. In particular, small ball probabilities provide the asymptotic behavior of the probability measure inside a ball as the radius of the ball tends to zero. In this talk, we will provide an overview on some recent developments, including symmetrization inequalities in high dimension, smooth Gaussian processes, and branching related processes.

Gaussian summation processes and weighted summation operators on trees

M.A. Lifshits, W.Linde (St.Petresburg State University, F. Schiller University Jena)

We study Gaussian summation processes on trees

$$X(t) := \sigma(t) \sum_{s \preceq t} \alpha(s)\xi(s) \, , \quad t \in T \ ,$$

where T is a tree, α and σ are given weights on T, and $\xi(s)$ are independent standard Gaussian rvs. In some important cases, we provide necessary and sufficient conditions for boundedness of X.

In parallel, we investigate compactness properties of weighted summation operators $V_{\alpha,\sigma}$ as mapping from $\ell_1(T)$ into $\ell_q(T)$ for some $q \in (1, \infty)$, defined by

$$(V_{\alpha,\sigma}x)(t) := \alpha(t) \sum_{s \succeq t} \sigma(s)x(s) \, , \quad t \in T \ .$$

These operators are natural discrete analogues of Volterra operators. We introduce a metric d on T such that compactness properties of (T, d) imply two–sided estimates for the (dyadic) entropy numbers of $V_{\alpha,\sigma}$ (recall that behavior of entropy numbers is directly related to small deviation probabilities of X). The results are applied to a large variety of trees and weights.

Permanental Processes

Michael B. Marcus (CUNY)

An α-permanental process $\theta := \{\theta_x, x \in S\}$, is a real valued positive stochastic process that is determined by a real valued kernel $\Gamma = \{\Gamma(x, y), x, y \in T\}$, in the sense that its finite joint distributions are given by

$$E\left(\exp\left(-\sum_{i=1}^{n} \lambda_i \theta_{x_i}\right)\right) = |I + \Lambda\widetilde{\Gamma}|^{-\alpha}, \tag{2}$$

where I is the $n \times n$ identity matrix, Λ is the $n \times n$ diagonal matrix with entries $(\lambda_1, \dots, \lambda_n)$, $\widetilde{\Gamma} = \{\Gamma(x_i, x_j)\}_{i,j=1}^{n}$ is an $n \times n$ matrix and $\alpha > 0$. When Γ is symmetric and positive definite and $\alpha = 1/2$, $\theta = \{G_x^2, x \in S\}$, where $\{G_x, x \in S\}$ is a mean zero Gaussian process with covariance Γ. However, the right-hand side of (2) can define a stochastic process for kernels Γ that are not symmetric. For example, when Γ is the zero potential density of a transient Markov process, that is not symmetric. Therefore, permanental processes are positive stochastic processes that generalize processes that are the squares of Gaussian processes.

In certain cases the squares of Gaussian processes are related by the Dynkin Isomorphism Theorem to the local times of symmetric Markov processes. The more general permanental processes are similarly related to the local times of Markov processes that need not be symmetric.

In this paper we discuss recent results about sample path properties of permanental processes obtained with Hana Kogan and Jay Rosen. In particular we identify permanental processes with kernels that are the potential densities of transient Markov processes as the loop soup local times of the Markov processes.

A More General Maximal Bernstein-Type Inequality

Péter Kevei and David M. Mason (University of Szeged and University of Delaware)

We describe a new and unexpected general maximal Bernstein-type inequality, along with a number of interesting applications.

Projections of Probability Distributions: A Measure-theoretic Dvoretzky Theorem

Elizabeth Meckes (Case Western Reserve University)

There is a widely studied phenomenon which can be described by saying that high-dimensional random vectors typically have Gaussian marginals. For example, for a probability measure on R^d, under mild conditions, most one-dimensional marginals are approximately Gaussian if d is large. This fact has important implications in statistics, in particular for the procedure known as graphical projection pursuit. Natural questions are then, for example, how many projections typically have to be tried before finding something non-Gaussian, or in another direction one could ask how large k can be as a function of the ambient dimension for k-dimensional marginals to be approximately Gaussian. I will discuss a quantitative approach to this phenomenon which sheds light on both of these questions. In particular, I will give the main ideas of the proof that, under mild conditions, a probability measure on R^d has mostly Gaussian marginals if $k < 2log(d)/log(log(d))$, and that this estimate is best possible in the metric considered.

Concentration and convergence rates for spectral measures of random matrices

Mark W. Meckes (Case Western Reserve University)

I will discuss how a combination of concentration of measure and metric entropy techniques can be used to estimate the Wasserstein distance between the empirical spectral distribution of a random matrix and its mean. The techniques apply to several different models of random matrices, and allow us in particular to improve on previous results of Diaconis–Shahshahani, Hiai–Petz, Guionnet–Zeitouni, Chatterjee, and Kargin.

On strong approximation for the empirical process of stationary sequences

Florence Merlevède (University of Paris Est)

In this talk, I shall present a strong approximation result with rate for the empirical process associated to a stationary sequence of real-valued random variables, under dependence conditions involving only indicators of half lines. This strong approximation result also holds for the empirical process associated to iterates of expanding maps with a neutral fixed point at zero as soon as the correlations decay more rapidly than $n^{-1-\delta}$ for some positive δ, which shows that our conditions are in some sense optimal. (Joint work with D. Dedecker and E. Rio).

Structure of the Gibbs measure in the Sherrington-Kirkpatrick model

Dmitry Panchenko (Texas A&M University)

In the Sherrington-Kirkpatrick spin glass model, given i.i.d. standard normal r.v.s $(g_{i,j})$, one considers a Gaussian process

$$H_N(\vec{\sigma}) = \sum_{i,j=1}^{N} g_{i,j}\sigma_i\sigma_j$$

indexed by vectors $\vec{\sigma} = (\sigma_1, \ldots, \sigma_N) \in \{-1, +1\}^N/\sqrt{N}$ of length one and defines a random probability measure, called the Gibbs measure,

$$G_N(\{\vec{\sigma}\}) = \frac{\exp(\beta\sqrt{N} H_N(\vec{\sigma}))}{Z_N}$$

for some inverse temperature parameter $\beta > 0$. As the size of the system N goes to infinity, one would like to understand the geometry of the set on which the measure G_N concentrates or, in one possible interpretation, given an i.i.d. sample $(\vec{\sigma}^l)$ from G_N, one would like to find the asymptotic distribution of the Gram matrix $(\vec{\sigma}^l \cdot \vec{\sigma}^{l'})_{l,l' \geq 1}$ under all the randomness involved, $\mathrm{E}G_N^{\otimes\infty}$. This distribution was predicted by the Italian physicist Giorgio Parisi and the main feature of his theory is that the measure G_N in the limit must concentrate on an ultrametric set. I will review several results from recent years that partially confirm Parisi's predictions.

Exact asymptotics for linear processes

Magda Peligrad (University of Cincinnati)

In this talk we present several new asymptotic results for linear processes. We shall address first the following question: giving a sequence of identically distributed random variables in the domain of attraction of a normal law, does the associated linear process satisfy the central limit theorem? We study this question for several classes of stationary processes. For independent and identically distributed random variables we show that the central limit theorem for the linear process is equivalent to the fact that the variables are in the domain of attraction of a normal law, answering in this way an open problem in the literature. Further, we present the convergence to the fractional Brownian motion and we also discuss the self-normalized version of this theorem. Then, we introduce the exact moderate and large deviation asymptotics in non-logarithmic form. We give an asymptotic representation for probability of the tail of the normalized sums and specify the zones in which it can be approximated either by a standard normal distribution or by the marginal distribution of the innovation process. Applications to fractionally integrated processes, regression models and computation of value at risk and expected shortfall are pointed out. Finally, we shall mention some open questions. The results presented are based on recent joint papers with Hailin Sang, Wei Biao Wu and Yunda Zhong.

Direct Bootstrapping Technique and its Applications to Novel Goodness of Fit Test

Dragan Radulovic (Florida Atlantic University)

We prove a general theorem that characterizes situations in which we could have asymptotic closeness between the original statistics H_n and its Bootstrap version H_n^* , without stipulating the existence of weak limits. As one possible application we introduce a novel goodness of fit test based on modification of Total Variation metric. This new statistic is more sensitive than Kolmogorov Smirnov statistics, it applies to higher dimensions and it does not converge weakly; but we show that it can be Bootstrapped.

Some Lasso procedure for multivariate counting processes and its particular link with some exponential inequalities for martingales

Patricia Reynaud-Bouret (CNRS and University of Nice Sophia-Antipolis)

Genomics and neurosciences produce a huge amount of high-dimensional data, where complex structures of dependence are involved. If one can infer those structures, the result may hint for particular synergy between neurons or transcription regulatory elements for instance. One of the natural model of such data is the multivariate Hawkes process, which is a particular case of counting process. The inference of the dependency structure in this set-up is a sparse non-parametric problem for which Lasso procedure is particularly adequate. However to obtain a full data driven procedure, some parameters of the procedure need to be calibrated in a full adaptive way. At this stage, sharp and particular exponential bounds for martingales are really useful. In those inequalities, the bracket of the martingale is estimated and plugged in the deviation bound. A discussion on the best way to perform this may follow. This is a joint work, still in progress, with N. R. Hansen (Copenhagen) and V. Rivoirard (Dauphine).

Stein's Method and Applications

Qi-Man Shao (Hong Kong University of Science and Technology)

Stein's method is a powerful tool in estimating accuracy of various probability approximations. It works for both independent and dependent random variables. It works for normal approximation and also for non-normal approximation. The method has been successfully applied to study the absolute error of approximations and the relative error as well. In this talk, we shall review the latest developments on Stein's method and its applications. The focus will be on Cramér type moderate deviation theorems under a general Stein's identity, for Studentized non-linear statistics as well as applications to Curie-Weiss models in statistical physics.

Probabilistic Analysis of Large Geometric Structures

Joseph E. Yukich (Lehigh University)

Fundamental questions pertaining to large random geometric structures often involve sums of spatially correlated terms having short range interactions but complicated long range dependence. This phenomenon arises in a range of fields, including (i) the statistical estimation of entropy and dimension via graphical methods, (ii) the study of functionals of the convex hull of an i.i.d. sample, (iii) models for random sequential packing, and (iv) random geometric graphs and networks. This talk will survey general methods for establishing limit theorems for geometric functionals of large random structures. It is shown that if the geometric functionals satisfy a spatial dependency condition known as stabilization then laws of large numbers, variance asymptotics, and central limit theorems follow, subject to appropriate scaling. The general theory is shown to answer questions in the areas (i)-(iv).