Mathematisches Forschungsinstitut Oberwolfach

Report No. 26/2015

DOI: 10.4171/OWR/2015/26

Probabilistic Techniques in Modern Statistics

Organised by
Vladimir Koltchinskii, Atlanta
Richard Nickl, Cambridge UK
Markus Reiß, Berlin
Sara van de Geer, Zürich

17 May – 23 May 2015

Abstract. The aim of the workshop was to bring together researchers in modern probability theory and statistics to discuss recent advances and challenging open problems at the intersection of both fields. It focussed on the most promising areas in which fruitful interactions between probability theory and statistics are currently developing.

Mathematics Subject Classification (2010): 60, in particular: 60B12, 60B20, 60D05, 60E07, 60F17, 60Gxx; 62, in particular: 62B15, 62Gxx, 62Jxx, 62Mxx.

Introduction by the Organisers

The workshop **Probabilistic Techniques in Modern Statistics** was organized by Vladimir Koltchinskii (Georgia Tech), Richard Nickl (University of Cambridge), Markus Reiss (Humboldt-Universität, Berlin) and Sara van de Geer (ETH, Zürich) and it took place on May 17–May 23, 2015.

The goal of the workshop was to bring together researchers in modern probability, statistics and related areas and to discuss recent advances and open problems at the intersection of these fields. The main focus was on the areas of the most intense interactions of probability and statistics with a significant impact on the development of novel methods of statistical inference for complex, high- and infinite-dimensional data sets. Among recent advances in these areas are deep understanding of the role of concentration of measure and concentration inequalities in high-dimensional inference, the development of non-asymptotic theory of random matrices and the progress on generic chaining and concentration bounds for empirical and related classes of stochastic processes.

The list of specific topics discussed at the workshop included:

- concentration of measure and its applications in statistical inference;
- probabilistic and geometric methods in high-dimensional statistics;
- Gaussian and empirical processes methods;
- non-asymptotic bounds for random matrices;
- statistics of stochastic processes;
- nonparametric methods, model selection and adaptive estimation;
- Bayesian nonparametrics.

In total, 51 mathematicians and statisticians participated in the workshop, including a number of junior researchers and PhD students. The program included 25 regular talks (their abstracts are given below) and a short evening session with several presentations by PhD students: Nicolay Baldin (Berlin), Claire Boyer (Toulouse), Emilie Devijver (Orsay), Ester Mariucci (Grenoble) and Benjamin Stucky (Zurich).

The workshop has stimulated fruitful discussions, exchanges and potential collaborations between probabilists and statisticians working in cutting edge areas of their fields.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-1049268, “US Junior Oberwolfach Fellows”.

Workshop: Probabilistic Techniques in Modern Statistics

Table of Contents

Abstracts

Nonparametric estimation of service time distribution in the $M/G/\infty$ queue and related estimation problems

Alexander Goldenshluger

The subject of this talk is the problem of estimating service time distribution G of the $M/G/\infty$ queue from incomplete data on the queue. This problem has been studied under different assumptions on the available data. The following three observation schemes have been considered in the literature:

(a) observation of arrival and departure epochs without their matchings;
(b) observation of the queue–length (number–of–busy–servers) process;
(c) observation of the busy–period process.

We note that observation schemes (a) and (b) are equivalent up to initial conditions on the queue length.

In setting (a) Brown [2] proposed an estimator of G which is based on the idea of pairing every departure epoch with the closest arrival epoch to the left. Differences between these epochs constitute an ergodic stationary random sequence whose marginal distribution is related to the service time distribution G by a simple formula. Then estimation of G can be achieved by inverting the formula and substituting the empirical marginal distribution of the differences. Brown [2] proved that the proposed estimator is consistent.

Nonparametric estimation of service time distribution G under observation schemes (b) and (c) was considered in [1]. It is well known that in the steady state the queue–length process $\{X(t)\}$ is stationary with Poisson marginal distribution and correlation function

$$H(t) = 1 - G^*(t), \quad G^*(t) := \Big[\int_0^\infty [1-G(x)]dx\Big]^{-1} \int_0^t [1-G(x)]dx. \tag{1}$$

This fact suggests that function G^* can be reconstructed by estimating correlation function of the queue–length process. Bingham & Pitts [1] discuss this approach and provides standard results from the time series literature for estimators of G^*. The idea of reconstructing the service time distribution from correlation structure of the queue–length process was also exploited by Pickands & Stine [3]. The model considered in that paper assumes that a Poisson number of customers arrives at discrete times $1, 2, \dots, T$, and service times are i.i.d. random variables taking values in the set of non–negative integer numbers. In this discrete setting estimation of the service time distribution is equivalent to estimating a linear form of the correlation function of the queue–length process. For the latter problem standard results from the time series literature are applicable.

Although estimation of G under different observation schemes was considered in the literature, the most interesting and important statistical questions remain to be open. In particular, it is not clear what is the achievable estimation accuracy in such problems, and how to construct optimal estimators.

In this work we adopt minimax approach for measuring estimation accuracy. It is assumed that the estimated distribution G belongs to a given functional class, and accuracy of any estimator is measured by its worst–case mean squared error on the class. The functional class is defined in terms of restrictions on smoothness and tail behavior of G. We concentrate on the observation scheme (b) when the queue–length process is observed on a fixed interval at the points of the regular grid. We want to estimate G at a fixed point using such observations.

We develop an estimator of G which is based on the relationship between distribution G and covariance function of the queue–length process, as discussed in [1] and [3] [cf. (1)]. In particular, estimating G at a fixed point is reduced to estimating derivative of the covariance function of the queue–length process at this point. We analyze accuracy of our estimator over a suitable class of target distributions and derive an upper bound on the maximal risk. The upper bound is expressed in terms of the functional class parameters and the observation horizon. The problem of estimating the arrival rate is discussed as well.

A natural question is: what is the achievable estimation accuracy in the $M/G/\infty$ problem? This question calls for a lower bound on the minimax risk. Since explicit formulas for finite dimensional distributions of the queue–length process in the $M/G/\infty$ model are not available, derivation of lower bounds on the minimax risk seems to be analytically intractable. Therefore, driven by a Gaussian approximation to the queue–length process, we consider a closely related estimation problem for a Gaussian model. Specifically, let $\{X(t), t \in R\}$ be a continuous–time stationary Gaussian process which is observed at the points of a regular grid on a given time interval. Using such discrete observations we want to estimate the derivative of the covariance function of $\{X(t), t \in R\}$. We derive a lower bound on the minimax risk in this problem, and show that under suitable conditions it converges to zero at the same rate as the risk of our estimator in the $M/G/\infty$ estimation problem. This fact strongly suggests that our estimator of the service time distribution is rate–optimal.

References

[1] N. H. Bingham and S. M. Pitts, *Non–parametric estimation for the $M/G/\infty$ queue.* Ann. Inst. Statist. Math. **51** (1999), 71–97.
[2] M. Brown, *An $M/G/\infty$ estimation problem.* Ann. Math. Statist. **41** (1970), 651–654.
[3] J. Pickands and R. A. Stine, *Estimation for an $M/G/\infty$ queue with incomplete information.* *Biometrika* **84** (1997), 295–308.

Gaussian approximation of suprema of empirical processes

Kengo Kato
(joint work with Victor Chernozhukov,Denis Chetverikov)

We develop a new direct approach to approximating suprema of general empirical processes by a sequence of suprema of Gaussian processes, without taking the route of approximating whole empirical processes in the sup-norm. We prove an abstract approximation theorem applicable to a wide variety of statistical problems, such as construction of uniform confidence bands for functions. Notably, the bound in the main approximation theorem is nonasymptotic and the theorem allows for functions that index the empirical process to be unbounded and have entropy divergent with the sample size. The proof of the approximation theorem builds on a new coupling inequality for maxima of sums of random vectors, the proof of which depends on an effective use of Stein's method for normal approximation, and some new empirical process techniques. We study applications of this approximation theorem to local and series empirical processes arising in nonparametric estimation via kernel and series methods, where the classes of functions change with the sample size and are non-Donsker. Importantly, our new technique is able to prove the Gaussian approximation for the supremum type statistics under weak regularity conditions, especially concerning the bandwidth and the number of series functions, in those examples.

References

[1] Chernozhukov, V., Chetverikov, D., and Kato, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564-1597.

[2] Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Process. Appl.*, to appear.

How large is the norm of a random matrix?

Ramon van Handel
(joint work with Afonso S. Bandeira)

Let X be an $n \times n$ symmetric random matrix with independent Gaussian entries $X_{ij} \sim N(0, b_{ij}^2)$. If the variances b_{ij}^2 of the entries are all the same (that is, the entries are i.i.d.) or of the same order, this model is known as a *Wigner matrix* and has been widely studied in the literature. Due to the large amount of symmetry of such models, extremely precise analytic results are available on the limiting behavior and distributions of fine-scale spectral properties of the matrix.

Our interest, however, goes in an orthogonal direction. We consider the case where the variances b_{ij}^2 are given but arbitrary: that is, we consider *structured random matrices* where the structure is given by the variance pattern of the entries. For example, one could consider "sparse Wigner matrices" where a certain sparsity pattern of the matrix is given, and only the nonzero entries are made i.i.d. standard Gaussian. The challenge in investigating such matrices is to understand how the given structure of the matrix variances b_{ij}^2 (for example, the sparsity pattern) is reflected in the spectral properties of the matrix.

In particular, we are interested in the location of the edge of the spectrum, that is, in the expected spectral norm $\mathbf{E}\|X\|$ of the matrix. Understanding the spectral norm of random matrices is a problem of basic interest in several areas of pure mathematics (probability theory, functional analysis, combinatorics) and in applied mathematics, statistics, and computer science. While the spectral norm of classical random matrix models such as Wigner matrices is well understood, existing methods almost always fail to be sharp in the presence of nontrivial structure. For example, the widely used "matrix concentration" method, which gives

$$\mathbf{E}\|X\| \lesssim \sigma\sqrt{\log n}, \qquad \sigma^2 := \max_i \sum_j b_{ij}^2$$

in our setting, is sharp for diagonal matrices with i.i.d. entries ($b_{ij} = \mathbf{1}_{i=j}$) but fails to be sharp even for Wigner matrices ($b_{ij} = 1$). On the other hand, by estimating all the variances b_{ij}^2 from above by the maximal variance, we can estimate the norm of X by that of a Wigner matrix whose entries have variance $\max_{ij} b_{ij}^2$:

$$\mathbf{E}\|X\| \lesssim \sigma_* \sqrt{n}, \qquad \sigma_*^2 := \max_{ij} b_{ij}^2.$$

This bound is sharp for Wigner matrices, but fails to be sharp for diagonal matrices. Another well-known bound due to Latała is similarly sharp essentially only for Wigner matrices. None of these bounds succeeds in capturing precisely how the variance structure is reflected in the spectral norm of the matrix.

In [1] we give a nearly optimal solution to this problem: we show that

$$\mathbf{E}\|X\| \lesssim \sigma + \sigma_* \sqrt{\log n},$$

which could be viewed as a sort of interpolation between the two extreme bounds that are stated above. The beauty of this bound is that it is matched by the almost identical lower bound (that is essentially trivial, see below)

$$\mathbf{E}\|X\| \gtrsim \sigma + \mathbf{E}\max_{ij} |X_{ij}|.$$

It is classical that $\mathbf{E}\max_{ij} |X_{ij}| \sim \sigma_* \sqrt{\log n}$ under mild assumptions. Thus the result of [1] evidently captures precisely how the structure of the variances—in terms of the parameters σ and σ_*—is reflected in the norm of the matrix. This result has already proved to be extremely useful in statistical applications such as community detection and matrix completion, and makes it possible to effortlessly address otherwise nontrivial problems of classical random matrix theory such as identifying the phase transition of the spectral edge of random band matrices.

The lower bound given above is extremely suggestive. It is obtained by averaging two trivial bounds. First, the norm of a matrix is always trivially bounded below by the magnitude of its largest entry:

$$\mathbf{E}\|X\| \geq \mathbf{E}\max_{ij} |X_{ij}|.$$

On the other hand, by the Poincaré inequality and Jensen's inequality,

$$\mathbf{E}\|X\| \gtrsim [\mathbf{E}\|X\|^2]^{1/2} \geq \|\mathbf{E}X^2\|^{1/2} = \sigma.$$

Thus the two terms in the lower bound reflect two distinct mechanisms that control the edge of the spectrum: the spectral norm is large either if the matrix itself is large on average (which is quantified by $\sigma^2 = \|\mathbf{E}X^2\|$; note that the expectation here is *inside* the norm!), or if one of the entries of the matrix has a large fluctuation (which is quantified by $\mathbf{E}\max_{ij}|X_{ij}|$). In view of the nearly matching upper bound, this strongly suggests that these are the *only* two reasons why the spectral norm can be large. If this surprising phenomenon is true, then the lower bound should be sharp, that is, we have the following *conjecture*:

$$\mathbf{E}\|X\| \overset{?}{\asymp} \sigma + \mathbf{E}\max_{ij}|X_{ij}|.$$

Unfortunately, the proof of the upper bound in [1] proceeds by an entirely different route that sheds no light on this conjecture. The idea behind the proof of the upper bound is a "dimensional compression" argument: the norm of the structured $n \times n$ random matrix X is compared with the norm of an (unstructured) Wigner matrix of much smaller dimension of order $\sim (\sigma/\sigma_*)^2 + \log n$, for which classical estimates are available. The (easy) proof of this compression argument is combinatorial in nature, and does not help understand the more delicate probabilistic mechanism conjectured above. Resolving this conjecture will require an entirely different approach to the problem. Some progress in this direction is reported in [2].

References

[1] A. S. Bandeira and R. van Handel, *Sharp nonasymptotic bounds on the norm of random matrices with independent entries* (2015), Ann. Probab., to appear.

[2] R. van Handel, *On the spectral norm of inhomogeneous random matrices* (2015), preprint arXiv:1502.05003.

Statistical estimation in transport-fragmentation models

Marc Hoffmann

(joint work with M. Doumic, N. Krell, A. Olivier, L. Robert)

We consider simple branching processes with deterministic evolution between jump times. Such models appear as toy models for population growth in cellular biology. We wish to statistically estimate the parameters of the model, in order to ultimately discriminate between different hypotheses related to the mechanisms that trigger cell division. We structure the model by state variables for each individual like size, age, growth rate, DNA content and so on. The evolution of the particle system is described by a common mechanism:

(1) Each particle grows by "ingesting a common nutrient" = deterministic evolution.
(2) After some time, depending on a structure variable, each particle gives rise to $k = 2$ offsprings by cell division = branching event.

In the talk we focus on structuring variables that are either age or size. The population evolution is associated with an infinite marked binary tree $\mathcal{U} = \bigcup_{n=0}^{\infty}\{0,1\}^n$ with $\{0,1\}^0 := \emptyset$. To each cell or node $u \in \mathcal{U}$, we associate a cell with size at birth given by ξ_u and lifetime ζ_u, a birth time b_u and a time of death d_u so that $\zeta_u = d_u - b_u$.

Observation scheme I: genealogical data. Set $|u| = n$ if $u = (u_1, \ldots, u_n) \in \mathcal{U}$, $uv = (u_1, \ldots, u_n, v_1, \ldots, v_m)$ if $v = (v_1, \ldots, v_m) \in \mathcal{U}$. For some (large) $n = 2^{k_n}$, define $\mathcal{U}_{[n]} = \{u \in \mathcal{U},\ |u| \leq k_n\}$. We observe $\{\xi_u$ and/or $\zeta_u, u \in \mathcal{U}_{[n]}\}$, *i.e.* the whole (unbiased) tree over the first k_n generations.

Observation scheme II: temporal data. For some (large) $T > 0$, define $\mathcal{U}_T = \{u \in \mathcal{U}, b_u \leq T\}$. We have $\mathcal{U}_T = \mathring{\mathcal{U}}_T \cup \partial\mathcal{U}_T$, with

$$\mathring{\mathcal{U}}_T = \big\{u, d_u \leq T\big\} \text{ and } \partial\mathcal{U}_T = \big\{u, b_u \leq T < d_u\big\}$$

We observe $\{\zeta_u^T$ and/or $\xi_u^T, u \in \mathcal{U}_T\}$ where $\zeta_u^T = \min\{d_u, T\} - b_u$, and $\xi_u^T = \xi_u$ if $d_u \leq T$ and the "size of u at time T" otherwise. This induces a bias: small lifetimes are more often observed than large lifetimes.

We are able to charaterise the optimal (in a min-max sense) rate of convergence for the branching rate $z \rightsquigarrow B(z)$ in two separate cases: **1)** from genealogical data I when $B = B(x)$ depends on the size x of the cell only, or **2)** from the more challenging case of temporal data II but in the mathematically simpler model when $B = B(a)$ depends on the age a of the cell only (recovering $B(a)$ from genealogical data I becomes irrelevant since the resulting statistical model is equivalent to density estimation).

Estimation of the size dependent $B(x)$ from observation scheme I. Each cell grows according to the simple deterministic evolution $dX(t) = \tau X(t)dt$ where τ is the common growth rate of the population. After division according to the branching rate $B(X(t))$, each cell splits into two offsprings with the same size. We thus have $\mathbb{P}(\zeta_u \in [t, t+dt] \,|\zeta_u \geq t, \xi_u = x) = B(xe^{\tau t})dt$, from which we obtain the density of the lifetime of the parent ζ_{u^-} conditional on its size at birth $\xi_{u^-} = x$: $t \rightsquigarrow B(xe^{\tau t}) \exp\Big(-\int_0^t B(xe^{\tau s})ds\Big)$. Using $2\,\xi_u = \xi_{u^-} \exp\big(\tau\zeta_{u^-}\big)$, we further infer

$$\mathbb{P}\big(\xi_u \in dx' \big|\, \xi_{u^-} = x\big) = \frac{B(2x')}{\tau x'} \mathbf{1}_{\{x' \geq x/2\}} \exp\big(-\int_{x/2}^{x'} \tfrac{B(2s)}{\tau s} ds\big)dx'.$$

We obtain a simple an explicit representation for the transition kernel $\mathcal{P}_B\big(x, dx'\big) = \mathbb{P}\big(\xi_u \in dx' \big|\, \xi_{u^-} = x\big)$. Under appropriate conditions on B, the Markov chain associated to $\mathcal{P}_B$ on $(0, \infty)$ is geometrically ergodic; in particular, there exists a unique invariant probability $\nu_B(dx) = \nu_B(x)dx$ on $[0, \infty)$. Expanding the equation $\nu_B\mathcal{P}_B = \nu_B$, one can easily prove that

$$\nu_B(y) = \frac{B(2y)}{\tau y}\mathbb{P}_{\nu_B}\big(\xi_{u^-} \leq 2y,\ \xi_u \geq y\big) = \frac{B(2y)}{\tau y}\int_y^{2y} \nu_B(x)dx$$

which yields a strategy for estimating $B(x)$ by recovering $\nu_B(x/2)$ by kernel estimation and $\int_{x/2}^x \nu_B(z)dz$ by empirical mean. Under appropriate assumptions on the ergodicity of the model that can be specified over an appropriate function class for B, the s-Hölder function $x \rightsquigarrow B(x)$ can be recovered in squared-error loss over nontrivial compact intervals in $(0,\infty)$ with the (normalised) rate $n^{-s/(2s+1)}$ and this rate is optimal in a minimax sense. The precise statements and assumptions are developed in [1].

Estimation of the age dependent $B(a)$ from observation scheme II. In that setting, we have $\mathbb{P}(\zeta_u \in [a, a+da]\,|\,\zeta_u \geq a) = B(a)da$. Let $f_B(a) = B(a)\exp\big(-\int_0^\infty B(s)ds\big)$. We have a law of large numbers

$$\frac{1}{|\mathring{\mathcal{U}}_T|}\sum_{u\in\mathring{\mathcal{U}}_T} g(\zeta_u) \stackrel{\mathbb{P}}{\to} 2\int_0^\infty g(a)e^{\lambda_B a} f_B(a)da,\ \ T\to\infty$$

for an appropriate test function g, where λ_B is the Malthus parameter of the model, that governs (for instance) the size of the population in the following sense: $\mathbb{E}[|\mathring{\mathcal{U}}_T|] \sim \kappa_B e^{\lambda_B T}$ as $T\to\infty$, for some constant $\kappa_B > 0$. The convergence is appended with the expected rate $e^{\lambda_B T/2}$ uniformly in B belonging to an appropriate function class. Using data from $\partial\mathcal{U}_T$, it is then possible to estimate λ_B with the rate $e^{\lambda_B T/2}$ as well. This yields a strategy for recovering $B(a)$ by picking g as a weighted (random) kernel estimator. Under appropriate assumptions that can be specified over an explicit class of functions B, the s-Hölder function $a \rightsquigarrow B(a)$ can be recovered in squared-error loss over nontrivial compact intervals in $(0,\infty)$ with the (normalised) rate $e^{-s\lambda_B T/(2s+1)}$ and this rate is optimal in a minimax sense. The precise statements and assumptions are developed in [4].

Open questions, future research. A formal link can be made rigorous between the above statistical model and the analysis of the transport-fragmentation equation. In the size-dependent model, it is well known (see for instance [5]) that the evolution $n(t,x)$ of the number of cells of size x alive at time t solves the PDE

$$\partial_t n(t,x) + \partial_x\big(\tau x n(t,x)\big) + B(x)n(t,x) = 4B(2x)n(t,2x),$$

with initial condition $n(t, x=0) = 0$, $t > 0$ and an appropriate initial condition $n(0,x)$ for $x \geq 0$. By renormalising the equation and approximating its solution by a steady state $n(t,x) = e^{\lambda_B t}N_B(x)$, where $N_B(x)$ denotes the probability density of a "typical" cell alive at time t and λ_B is the Malthus parameter of the model, it is possible to define a proxy statistical experiment mimicking data from $\partial\mathcal{U}_T$ (cells alive at time t) in which $x \rightsquigarrow B(x)$ can estimated via an inverse problem of order 1, see [3, 2]. The extension of the approach in the simple age-dependent model of [4] to a size-dependent model suggests a strategy for proving in full generality the results of the proxy model, with the hope to establish that the optimal rate from data $\mathring{\mathcal{U}}_T$ is indeed $e^{-s\lambda_B T/(2s+3)}$, as suggested by the studies [3, 2].

References

[1] M. Doumic, M. Hoffmann, N. Krell, L. Robert, *Statistical estimation of a growth-fragmentation model observed on a genealogical tree. Bernoulli*, (2015) **21**, 1760–1799.

[2] M. Doumic, M. Hoffmann, P. Reynaud-Bouret, V. Rivoirard, *Nonparametric estimation of the division rate of a size-structured population.* SIAM Journal on Numerical Analysis. (2012) **50**, 925–950.

[3] M. Doumic, B. Perthame, J. Zubelli, *Numerical Solution of an Inverse Problem in Size-Structured Population Dynamics*. Inverse Problems, **25**, (2009), 045008, 25pp.
[4] M. Hoffmann, A. Olivier, *Nonparametric estimation of the division rate of an age dependent branching process*. arXiv:1412.5936. 32pp. (2014).
[5] B. Perthame, Transport equations in biology. *Birckhauser* (2007).

Likelihood ratios for eigenvalues in spiked multivariate models

Iain M. Johnstone
(joint work with Prathapa Dharmawansa, Alexei Onatski)

It is just over 50 years since James' 1964 paper on the distribution of matrix variates and latent roots, in which he gave a remarkable classification of many of the eigenvalue distribution problems of multivariate statistics. We revisit the classification, now from the viewpoint of high dimensional models and low rank departures from the usual null hypotheses.

We consider spiked models representing each of the five classes of multivariate statistical problems identified by James [1]. For each of the models, we describe the phase transition of the largest eigenvalue, and derive the asymptotic behavior of the likelihood ratios that correspond to null and alternative hypotheses about sub- and super-critical spikes. We find that the statistical experiment of observing the eigenvalues in the super-critical regime, parameterized by local deviations of the spike from its value under the null, converges to a simple Gaussian shift experiment, and therefore, the best test about a single super-critical spike is based on the largest eigenvalue only. Our findings for the sub-critical regime are totally different. In that regime, the experiment of observing the eigenvalues converges to a Gaussian sequence experiment, and no optimal test about a sub- critical spike is available. We derive the asymptotic power envelopes for such tests. The current state of this work in progress is described in manuscripts [2, 3].

References

[1] A. T. James. *Distributions of matrix variates and latent roots derived from normal samples*, Annals of Mathematical Statistics **35**, (1964), 475–501.
[2] Prathapasinghe Dharmawansa, Iain M. Johnstone, Alexei Onatski. *Local Asymptotic Normality of the spectrum of high-dimensional spiked F-ratios*, (2014) ArXiv: 1411.3875.
[3] Iain M. Johnstone, Alexei Onatski. *Testing in high-dimensional spiked models*, Manuscript in preparation.

Change-point analysis of volatility

Mathias Vetter
(joint work with Markus Bibinger, Moritz Jirak)

Suppose one observes a continuous Itô semi-martingale, i.e. a time-continuous stochastic process of the form

$$X_t = X_0 + \int_0^t a_s\, ds + \int_0^t \sigma_s\, dW_s\,,$$

with a standard Brownian motion W and adapted drift and volatility processes a and σ. In the situation of high-frequency asymptotics, in which the process is recorded at discrete regular times $i\Delta_n$ with a mesh $n^{-1} = \Delta_n \to 0$, we are interested in inference on the smoothness of the underlying volatility process. Central for the analysis in this talk is to check whether jumps occur or whether paths become rougher after a certain point in time. Thus, change-point techniques become important.

We focus on volatilities which are almost surely locally bounded and strictly positive adapted processes. For our testing problem we consider classes of squared volatilities

$$\Sigma(\mathfrak{a}, L) = \Big\{ (\sigma_t^2(\omega))_{t\in[0,1]} \Big| \sup_{s,t\in[0,1],|s-t|<\delta} \big|\sigma_t^2(\omega) - \sigma_s^2(\omega)\big| \le L(\omega)\delta^{\mathfrak{a}} \Big\}\,,$$

for an almost surely bounded random variable L. The regularity exponent $\mathfrak{a} > 0$ is the key parameter to describe the null hypothesis H_0.

Our core idea is to estimate spot volatility over small time blocks and to idenfity breaks from too large deviations between two successive local estimators. Precisely, if we set $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$, we use

$$X_{n,i} = \frac{n}{k_n} \sum_{j=1}^{k_n} (\Delta_{ik_n+j}^n X)^2\,,\ i = 0, \ldots, \lfloor n/k_n \rfloor - 1\,,$$

as a spot volatility estimator over $[ik_n\Delta_n, (i+1)k_n\Delta_n]$, where $k_n \to \infty$ is an auxiliary sequence. Then we consider

$$V_n = \max_{i=0,\ldots,\lfloor n/k_n\rfloor-2} |X_{n,i}/X_{n,i+1} - 1|\,,$$

measuring the largest deviation of ratios of spot volatility estimators to one, as well as an analogous statistic V_n^* computed over all overlapping blocks of k_n increments. Under mild assumptions on the processes as well as a certain growth condition on k_n which depends on $\mathfrak{a}$, we are able to prove convergence in distribution of a rescaled version of V_n (and its analogue V_n^*) to a Gumbel distribution, utilizing a result from [3]. Using this theorem, an asymptotic level α test for changes in the volatility is readily obtained. In case of jumps in the price process X and under further conditions, a similar result is obtained for truncated spot volatility estimators following [2].

On the other hand, we are interested in detection bounds on changes in the smoothness. Suppose that the volatility at time θ either has a jump of absolute size larger than b_n or that its smoothness drops to $\mathfrak{a}'$ on an interval at least of length $b_n^{1/\mathfrak{a}}$. In the spirit of [1] we are interested in minimax bounds on b_n, i.e. in finding the smallest order of b_n such that

$$\lim_{n\to\infty} \inf_{\psi} \gamma_\psi\big(\mathfrak{a}, b_n\big) = 0$$

holds, where ψ is a test and $\gamma_\psi\big(\mathfrak{a}, b_n\big)$ denotes the sum of the supremum over the error of the first kind in case σ^2 is $\mathfrak{a}$ smooth and the supremum of the error of the second kind in case the respective alternative holds. In both cases,

$$b_n \sim \big(n/\log(n)\big)^{-\frac{\mathfrak{a}}{2\mathfrak{a}+1}}$$

is proven to be the minimax bound.

References

[1] Y. Ingster, *Asymptotically minimax hypothesis testing for nonparametric alternatives I, II, III*, Mathematical Methods of Statistics **2** (1993), 85–114; 171–189; 249–268.

[2] C. Mancini, *Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps*, Scandinavian Journal of Statistics **36** (2009), 270–296.

[3] W. Wu, Z. Zhou, *Inference of trends in time series*, Journal of the Royal Statistical Society (Series B) **69** (2007), 391–410.

High-dimensional robust regression

Po-Ling Loh

We present results for high-dimensional linear regression using robust M-estimators with a regularization term. We show that when the derivative of the loss function is bounded, our estimators are robust with respect to heavy-tailed noise distributions and outliers in the response variables, with the usual $\mathcal{O}\left(\sqrt{\frac{k \log p}{n}}\right)$ rates for high-dimensional statistical estimation. Our results continue a line of recent work concerning local optima of nonconvex M-estimators with possibly nonconvex penalties, where we adapt the theory to settings where the loss function only satisfies a form of restricted strong convexity within a local neighborhood. We also discuss second-order results concerning the asymptotic normality of our estimators, and provide a two-step M-estimation algorithm for obtaining statistically efficient solutions within the local region.

References

[1] P. Loh, *Statistical consistency and asymptotic normality for high-dimensional robust M-estimators*, arXiv preprint 1501.00312, January 2015.

Limit theorems for stationary increments Lévy driven moving average processes

Mark Podolskij
(joint work with Andreas Basse-O'Connor, Raphaël Lachièze-Rey)

In this talk we present some new limit theorems for power variation of kth order increments of stationary increments Lévy driven moving averages. In this infill sampling setting, the asymptotic theory gives very surprising results, which (partially) have no counterpart in the theory of discrete moving averages.

We consider a *stationary increments Lévy moving average* process

$$X_t = X_0 + \int_{-\infty}^{t} \{g(t-s) - g_0(-s)\} dL_s,$$

where L is a pure jump Lévy process and the function g is assumed to be of the form

$$g(x) = x^\alpha f(x), \qquad \alpha > 0,$$

with $f : \mathbb{R}^+ \to \mathbb{R}$ being a smooth quickly decaying function with $f(0) \neq 0$. Next, we define the *kth order differences of* X via

$$\Delta_{i,k}^n X := \sum_{j=0}^{k} (-1)^j \binom{k}{j} X_{(i-j)/n}.$$

For instance, $\Delta_{i,1}^n X = X_{i/n} - X_{(i-1)/n}$ and $\Delta_{i,2}^n X = X_{i/n} - 2X_{(i-1)/n} + X_{(i-2)/n}$. The power variation of kth order differences of X is given by the statistic

$$V(X,p,k)_n := \sum_{i=k}^{n} |\Delta_{i,k}^n X|^p.$$

The *Blumenthal-Getoor index* of L is playing an important role in the description of the first order asymptotic results for the statistic $V(X,p,k)_n$. We recall that the Blumenthal-Getoor index β of L is defined via

$$\beta := \inf\Big\{ r \geq 0 : \int_{-1}^{1} |x|^r \nu(dx) < \infty \Big\} = \inf\Big\{ r \geq 0 : \sum_{s\in[0,1]} |\Delta L_s|^r < \infty \Big\}.$$

For instance, a ρ-stable Lévy process with $\rho \in (0,2)$ has the Blumenthal-Getoor index ρ. Our first theorem comprise the first order asymptotic theory for power variation $V(X,p,k)_n$.

Theorem: Under certain differentiability and integrability conditions on the kernel function g we obtain the following results:

(i) When $\alpha \in (0, k-1/p)$ and $p > \beta$, we obtain the stable convergence

$$n^{\alpha p} V(X,p,k)_n \xrightarrow{d_{st}} |f(0)|^p \sum_{m:\ T_m\in[0,1]} |\Delta L_{T_m}|^p \left(\sum_{l=k}^{\infty} |h_k(l+U_m)|^p \right),$$

where (T_m) are jump times of L, $(U_m)_{m\geq 1}$ is a sequence of iid $\mathcal{U}([0,1])$-distributed random variables independent of L and the function h_k is defined via

$$h_k(x) := \sum_{j=0}^{k} (-1)^j \binom{k}{j} (x-j)_+^{\alpha}.$$

(ii) Assume that L is a symmetric β-stable process with $\beta \in (0,2)$. When $\alpha \in (0, k-1/\beta)$ and $p < \beta$, we obtain

$$n^{p(\alpha+1/\beta)-1} V(X,p,k)_n \xrightarrow{\mathbb{P}} c_p,$$

where c_p are certain positive constants.

(iii) When $\alpha > k - 1/\max(p,\beta)$ and $p \geq 1$ we deduce

$$n^{kp-1} V(X,p,k)_n \xrightarrow{\mathbb{P}} \int_0^1 |F_s^{(k)}|^p ds, \qquad F_s^{(k)} = \int_{-\infty}^{s} g^{(k)}(s-u) dL_u.$$

We remark that the above limit theorem essentially comprises all possible cases with an exception of critical cases $\alpha = k - 1/p$, $\alpha = k - 1/\beta$ and $p = \beta$. We also notice that our asymptotic results uniquely identify the parameters α and β. We may apply our limit theorem to construct statistical estimates of α and β as follows. Define the statistic $S_{\alpha,\beta}(p,k)_n := -\frac{\log V(X,p,k)_n}{\log n}$. Then it holds that

$$S_{\alpha,\beta}(p,k)_n \xrightarrow{\mathbb{P}} S_{\alpha,\beta}(p,k) := \begin{cases} \alpha p: & \alpha < k-1/p, p > \beta \\ p(\alpha+1/\beta) - 1: & \alpha < k - 1/\beta, p < \beta \\ kp-1: & \alpha > k - 1/\max(p,\beta) \end{cases}$$

Notice that the limit $S_{\alpha,\beta}(p,k)$ is a piecewise linear function in p. For this reason a natural estimator of (α,β) is given via

$$(\hat{\alpha},\hat{\beta}) := \operatorname{argmin}_{(\alpha,\beta)} \sum_{k=1}^{\bar{k}} \int_0^{\bar{p}(k)} \left(S_{\alpha,\beta}(p,k)_n - S_{\alpha,\beta}(p,k)\right)^2 dp,$$

where the integral needs to be discretised for practical applications. Finally, there exists a weak limit theory associated to case (ii) of the above theorem. We refer to [1] for more details.

References

[1] A. Basse-O'Connor, R. Lachièze-Rey and M. Podolskij *Limit theorems for stationary increments Lévy driven moving average processes*, (2015), working paper.

Asymptotic equivalence for regression under dependent noise

Johannes Schmidt-Hieber

Consider the model $\mathbf{Y} = (Y_{1,n}, \ldots, Y_{n,n})$ with

$$Y_{i,n} = f\big(\tfrac{i}{n}\big) + N_i^H, \quad i = 1, \ldots, n, \tag{1}$$

and $(N_i^H)_i$ denotes a fractional Gaussian noise process (fGN) with Hurst index $H \in (0,1)$, that is, a centered Gaussian process with autocovariance function $k \mapsto \frac{1}{2}(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H})$. The different choices for H capture possible behavior of dependent data. For $H > 1/2$, fGN exhibits long-range dependence; for $H < 1/2$, the process has many negative correlations leading to cancellation of the noise with a faster rate than in the Central Limit Theorem; in the special case $H = 1/2$, fGN is just i.i.d. Gaussian white noise.

This model is inconvenient to work with due to the discrete design and the dependence of the errors. Working with the likelihood becomes very challenging as it involves the inverse covariance matrix for which no good approximations are known. Thus, even simple questions such as efficiency for constant f turn out to be very difficult. It has therefore been suggested in [2] to replace the discrete model by a continuous version, where we observe the path $(Y_t)_{t\in[0,1]}$ with

$$Y_t = \int_0^t f(u)du + n^{H-1} B_t^H, \quad t \in [0,1] \tag{2}$$

and $(B_t^H)_t$ a fractional Brownian motion, that is, a Gaussian process with covariance function $K(s,t) := \operatorname{Cov}(B_s^H, B_t^H) = \frac{1}{2}(|s|^{2H} + |t|^{2H} - |t-s|^{2H})$. Using the partial sum process of the observations (1), the closeness of the models can be motivated.

We discuss whether these models are asymptotically equivalent in the sense of Le Cam. To describe the approximation quality, we make heavily use of the geometry induced by the fBM through its uniquely defined reproducing kernel Hilbert space (RKHS), which in the following will be denoted by $\mathbb{H}$. As a first result, we obtain that asymptotic equivalence holds for the approximation of (1) by (2) with parameter space Θ, provided that

$$\sup_{f\in\Theta} \inf_{(\alpha_1,\ldots,\alpha_n)^t\in\mathbb{R}^n} \Big\| \int_0^{\cdot} f(u)du - \sum_{j=1}^n \alpha_j K\big(\cdot, \tfrac{j}{n}\big)\Big\|_{\mathbb{H}} = o(n^{H-1}) \tag{3}$$

and that the functions f can be approximated by step functions in $L^2[0,1]$ with rate $o(n^{-1} \wedge n^{2H-2})$. Thus, the question of asymptotic equivalence can be reduced to two approximation conditions. The key step in the proof is to show that given observations (1), one can construct a path which is asymptotically indistinguishable from (2). Although the partial sum process seems to be natural for interpolating discrete data it leads to intractable problems. Instead, for a well chosen vector $(x_1, \ldots, x_n)$, one has to use the projection property of the interpolation function

$$t \mapsto \mathbb{E}[B_t^H | B_{\ell/n}^H = x_\ell, \ell = 1, \ldots, n], \quad t \in [0,1]$$

in order to arrive at (3).

Unfortunately, it is close to impossible to check with elementary tools whether the approximation condition (3) holds for given Θ. Indeed, the function $t \mapsto K(t, s)$ does not concentrate on a small interval and has low Hölder smoothness due to the irregular behavior at $t = 0$ and $t = s$. Moreover, except for $H = \frac{1}{2}$, we have no clear understanding of the RKHS norm $\|\cdot\|_{\mathbb{H}}$. With the spectral representation of the RKHS, we can, however, rewrite (3). Denote by $\mathcal{F}$ the Fourier transform and write $\|\cdot\|_{\mathrm{hSob}(\gamma)}$ for the norm of the homogeneous Sobolev space with index γ. Suppose that for any $f \in \Theta$ there exists a function g with support on $[0, 1]$ such that

$$f = \mathcal{F}^{-1}\big(|\cdot|^{1-2H}\mathcal{F}(g)\big)\Big|_{[0,1]} \tag{4}$$

and

$$\sup_{f\in\Theta} \inf_{(\beta_1,\dots,\beta_n)^t\in\mathbb{R}^n} \Big\|g - \sum_{j=1}^{n} \beta_j \mathbb{I}_{\left(\frac{j-1}{n},\frac{j}{n}\right]}(\cdot)\Big\|_{\mathrm{hSob}(\frac{1}{2}-H)} = o(n^{H-1}). \tag{5}$$

This condition implies (3). Approximation by step functions in the Sobolev norm is feasible. The difficulty is in (4). There is no obvious way to invert the Fourier multiplier because of the support restrictions on g and f. One way to overcome this is to introduce a space of smooth functions g such that (5) holds and to declare the parameter space Θ as being all functions that are of the form (4). Nevertheless, a much better characterization can be obtained using the following explicit solutions of the Fourier multiplier representation, which are a consequence of the fundamental results in [1]. Denote by $\ldots < \omega_{-1} < \omega_0 := 0 < \omega_1 < \ldots$ the ordered, real roots of the Bessel function of the first kind and index $1 - H$. For any integer k, define

$$s \mapsto g_k(s) := \mathbb{I}_{(0,1)}(s)\partial_s \int_0^s e^{i2\omega_k(s-u)}(u - u^2)^{1/2-H} du,$$

and $f_k = e^{2i\omega_k \cdot}$. Then, (f_k, g_k) are (up to constants) solutions to (4). Using the theory of non-harmonic Fourier series, it follows that $(f_k)_k$ and $(g_k)_k$ are bi-orthogonal Riesz bases of $L^2[0, 1]$. In particular, any $f \in L^2[0, 1]$ has a unique representation as non-harmonic Fourier series $f = \sum_k \theta_k e^{2i\omega_k \cdot}$. It is natural to consider then Sobolev balls

$$\Theta(\alpha, R) = \big\{f = \sum_k \theta_k e^{2i\omega_k \cdot} : \sum_k (1 + |k|)^{2\alpha}|\theta_k|^2 \le R^2\big\}.$$

For integer smoothness, explicit boundary conditions can be derived. Now, we are able to state the main result.

Theorem 1. *If $H \ge 1/2$, then, asymptotic equivalence holds for $\Theta(\alpha, R)$ if and only if $\alpha > 1/2$. If $H \in (\frac{1}{4}, \frac{1}{2})$, then, asymptotic equivalence holds if $\alpha > (1 - H)/(H + 1/2) + H - 1/2$ and it fails to hold if $\alpha < 1 - H$.*

Thus, for long-range dependence we have found sharp conditions. In the case $1/4 < H < 1/2$, we have a small gap between the upper and lower bound and need an additional symmetry condition (cf. [3], Theorem 5). What is interesting is that in this case the boundary is not at 1/2 as typical for asymptotic equivalence. For $H \leq 1/4$, we have some heuristic arguments indicating that asymptotic equivalence cannot hold in this case.

From the theory, we can immediately deduce a sequence space representation of the continuous regression model (2), which is of the form

$$Z_k = \theta_k + \sigma_k n^{H-1}\epsilon_k, \quad \epsilon_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \quad k \in \mathbb{Z},$$

where θ_k are the coefficients in the non-harmonic Fourier representation of f and $\sigma_k \asymp |k|^{1/2-H}$. The growth of σ_k and the noise level n^{H-1} completely characterize the convergence rates that can be obtained for estimation of f or functionals of it. Moreover, in the continuous/sequence space model, construction of estimators are much more straightforward than in the discrete regression model with dependent noise.

References

[1] K. Dzhaparidze and H. van Zanten (2005), *Krein's spectral theory and the Payley- Wiener expansion for fractional Brownian motion.* Ann. Probab. **33**, 620–644.

[2] I. Johnstone and B. Silverman (1997), *Wavelet threshold estimators for data with correlated noise*, J. Royal Statist. Society, Ser. B **59**, 319–351.

[3] J. Schmidt-Hieber (2015), *Asymptotic equivalence for regression under fractional noise*, Ann. Statist., *to appear.*

Regression and the Offset Rademacher Complexity

Alexander Rakhlin

(joint work with T. Liang and K. Sridharan)

We consider regression with square loss and general classes of functions without the boundedness assumption. We introduce a notion of offset Rademacher complexity that provides a transparent way to study localization both in expectation and in high probability. Given $x_1, \ldots, x_n \in \mathcal{X}$, the offset Rademacher process is defined as a stochastic process

$$f \mapsto \frac{1}{n}\sum_{t=1}^{n} \epsilon_t f(x_t) - c f(x_t)^2$$

indexed by $f \in \mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$. Here $c \geq 0$ and ϵ_t's are independent Rademacher random variables. Given a class $\mathcal{F}$ of functions, as well as data $\{(X_i, Y_i)\}_{i=1}^{n}$ consider the two-step estimator

$$\widehat{g} = \operatorname*{argmin}_{f\in\mathcal{F}} \widehat{\mathbb{E}}(f(X) - Y)^2, \ \widehat{f} = \operatorname*{argmin}_{f\in\operatorname{star}(\mathcal{F},\widehat{g})} \widehat{\mathbb{E}}(f(X) - Y)^2$$

where $\widehat{\mathbb{E}}$ is the empirical expectation and $\operatorname{star}(\mathcal{F}, \widehat{g})$ denotes the star hull of $\mathcal{F}$ around $\widehat{g}$. For any (possibly non-convex) class $\mathcal{F}$, the excess loss of this two-step estimator is shown to be upper bounded by the offset complexity through a novel geometric Pythagorean-like inequality

$$\|h - Y\|_n^2 - \|\widehat{f} - Y\|_n^2 \geq c \cdot \|\widehat{f} - h\|_n^2$$

for any $h \in \mathcal{F}$ and $c = 1/18$. Here $\|\cdot\|_n$ is the empirical distance.

In the convex case, the estimator reduces to an empirical risk minimizer. The method recovers the results of (Rakhlin, Sridharan, Tsybakov '15) for the bounded case while also providing guarantees without the boundedness assumption.

Adaptive confidence bands for Markov chains and diffusions: Estimating the invariant measure and the drift

Jakob Söhl
(joint work with Mathias Trabs)

We consider the problem of nonparametric drift estimation for a diffusion process X following the equation $\mathrm{d}X_t = b(X_t)\mathrm{d}t + \mathrm{d}W_t$, where W is a Brownian motion. We base our estimation method on low-frequency observations $X_0, X_\Delta, \dots, X_{(n-1)\Delta}$ for $n \to \infty$ and with $\Delta > 0$ fixed. First the invariant density is estimated by a wavelet projection estimator and subsequently the drift function by a plug-in approach. The low-frequency observations follow a Markov chain obtained from restricting the diffusion to the discrete observation times. So we consider the setting of geometrically ergodic Harris-recurrent Markov chains. The first result is a functional central limit theorem in a multi-scale space for the estimator of the invariant density. This is used to construct confidence bands with L^∞-diameters that shrink at a $(\log n/n)^{s/(2s+1)}$-rate (up to undersmoothing), where s is the Hölder smoothness of the invariant density. We apply our results to the diffusion model, where in addition we prove a functional central limit theorem for the drift estimator and construct confidence bands for the drift. In order to construct confidence bands that adapt to the unknown smoothness of the drift function we assume a self-similarity assumption. Using Lepski's method we estimate the projection level that balances bias term and the stochastic term. This leads to an estimator of the smoothness of the drift function. We conclude with adaptive confidence bands for the drift whose L^∞-diameters shrink at a $(\log n/n)^{s/(2s+3)}$-rate (up to undersmoothing), where s is the Hölder smoothness of the drift function. Our proofs rely on the central limit theorem for Markov chains and on a concentration inequality for Markov chains recently obtained by Adamczak and Bednorz [1].

References

[1] R. Adamczak and W. Bednorz, *Exponential concentration inequalities for additive functionals of Markov chains*, ESAIM Probab. Stat. (2015+), to appear.
[2] J. Söhl and M. Trabs, *Adaptive confidence bands for Markov chains and diffusions: Estimating the invariant measure and the drift*, arXiv preprint 1412.7103 (2014).

Inference for Random Satisfiability Problems

Quentin Berthet

We consider high-dimensional inference problems on various random satisfiability problems, mainly the detection and estimation of planted solutions. We show that the study of these statistical problems is closely related to the properties of random instances. We will also describe the algorithmic aspects of these problems, giving positive and negative results pertaining to the performance of computationally efficient procedures.

Tail index estimation, concentration and adaptivity

Stéphane Boucheron
(joint work with Maud Thomas)

We present an adaptive version of the Hill estimator based on Lespki's model selection method. This simple data-driven index selection method is shown to satisfy an oracle inequality and is checked to achieve the lower bound recently derived by Carpentier and Kim. In order to establish the oracle inequality, we derive non-asymptotic variance bounds and concentration inequalities for Hill estimators. These concentration inequalities are derived from Talagrand's concentration inequality for smooth functions of independent exponentially distributed random variables combined with three tools of Extreme Value Theory: the quantile transform, Karamata's representation of slowly varying functions, and Rényi's characterisation of the order statistics of exponential samples. The performance of this computationally and conceptually simple method is illustrated using Monte-Carlo simulations.

References

[1] S. Boucheron and M. Thomas. Tail index estimation, concentration and adpativity, arXiv preprint 1503.05077.

On Pólya tree posterior densities

Ismaël Castillo

The first purpose of this talk is to investigate convergence of the Bayesian posterior distribution when a Pólya tree is used as a prior distribution on density functions. If posterior consistency has been established in a weak or Hellinger sense for this class of priors [1], [7], no rate-result is available in the literature so far. Another purpose is to show that this class of priors naturally illustrates the multiscale framework for analysis of posterior distributions recently introduced in a series of papers [2], [3], [4]. In particular, we investigate posterior limiting shape results via a nonparametric Bernstein-von Mises-type result.

Let $\mathcal{E} := \cup_{l\geq 0}\{0,1\}^l \cup \{\varnothing\}$ be the set of finite binary sequences. If $\varepsilon \in \{0,1\}^l$, we write $|\varepsilon| = l$. Let us consider the collection of regular dyadic partitions of $[0,1]$.

Set $I_\varnothing = [0,1)$ and, for any $\varepsilon \in \mathcal{E}$ such that $\varepsilon = \varepsilon(l,k)$ is the expression of length l in base 1/2 of $k2^{-l}$, set

$$I_\varepsilon := \left[\frac{k}{2^l}, \frac{k+1}{2^l}\right).$$

A random probability measure P follows a Pólya tree distribution $PT(\mathcal{A})$ with parameters $\mathcal{A} = \{\alpha_\varepsilon,\ \varepsilon \in \mathcal{E}\}$ on the sequence of partitions $\mathcal{I} = \{(I_\varepsilon)_{\varepsilon:\,|\varepsilon|=l},\ l \geq 0\}$ if there exist random variables $0 \leq Y_\varepsilon \leq 1$ such that,

1. the variables $Y_{\varepsilon 0}$ for $\varepsilon \in \mathcal{E}$ are mutually independent and $Y_{\varepsilon 0}$ follows a Beta$(\alpha_{\varepsilon 0}, \alpha_{\varepsilon 1})$ distribution.
2. for any $\varepsilon \in \mathcal{E}$, we have $Y_{\varepsilon 1} = 1 - Y_{\varepsilon 0}$
3. for any $l \geq 0$ and $\varepsilon = \varepsilon_1 \ldots \varepsilon_l \in \{0,1\}^l$, we have

$$P(I_\varepsilon) = \prod_{j=1}^{l} Y_{\varepsilon_1 \ldots \varepsilon_j}.$$

A typical choice of parameters assumes that α_ε only depends on the depth $|\varepsilon|$ of the tree at ε, that is $\alpha_\varepsilon = a_{|\varepsilon|}$, for some sequence $(a_l)_{l\geq 0}$. If $\sum a_l^{-1} < \infty$, it can be shown that the Pólya tree law produces random distributions that are Lebesgue-absolutely continuous, and hence can be used as prior on *density* functions.

Posterior convergence in L^∞-norm. A first main result in [5] is as follows.

Theorem 1. *Let $X = (X_1, \ldots, X_n)$ be i.i.d. from law P_0 with density f_0. Let f_0 belong to $\mathcal{C}^\alpha[0,1]$, for $\alpha \in (0,1]$ and suppose f_0 is bounded away from 0 on $[0,1]$. Let Π be a Pólya tree with parameters $\mathcal{A} = \{\alpha_\varepsilon,\ \varepsilon \in \mathcal{E}\}$ chosen as $\alpha_\varepsilon = a_{|\varepsilon|} \vee 8$ for any $\varepsilon \in \mathcal{E}$, with*

$$a_l = l2^{2l\alpha}, \qquad l \geq 0.$$

Then as $n \to \infty$, for any $M_n \to \infty$, it holds

$$E_{f_0}^n \Pi[f:\ \|f - f_0\|_\infty \leq M_n \varepsilon_{n,\alpha}^* \,|\, X] \to 1.$$

The class of density Pólya tree prior distributions thus provides an example of canonical – not dependent on n – prior, for which convergence of the posterior distribution in the supremum norm occurs at the minimax rate.

Limiting shape of the posterior distribution. Let $w := (w_l)_{l\geq 0}$ be such that $w_l/\sqrt{l} \uparrow \infty$. Let us define as in [4] the multiscale sequence space $\mathcal{M}_0 = \mathcal{M}_0(w)$

$$\mathcal{M}_0 = \Big\{x = \{x_{lk}\}:\ \lim_{l\to\infty} \max_k \frac{|x_{lk}|}{w_l} = 0\Big\}, \tag{1}$$

equipped with the norm $\|x\|_{lk} := \sup_l \max_k |x_{lk}|/w_l$. It is a separable Banach space. A function f is said to belong to $\mathcal{M}_0$ if the sequence of its wavelet coefficients $\langle f, \psi_{lk}\rangle$ over the Haar basis $\{\psi_{lk}\}$ on $[0,1]$ belongs to $\mathcal{M}_0$.

Now we define a limiting process. For P a given probability distribution on $[0,1]$, let $\mathbb{G}_P$ be a *P-white bridge process*, the Gaussian process indexed by the

Hilbert space $L^2(P) \equiv \{f : [0,1] \to \mathbb{R} : \int_0^1 f^2 dP < \infty\}$ with covariance function

$$\mathbb{E}\left[\mathbb{G}_P(g)\mathbb{G}_P(h)\right] = \int_0^1 (g - Pg)(h - Ph)dP.$$

The process $\mathbb{G}_P$ defines a tight Borel Gaussian variable in $\mathcal{M}_0$.

Let P_n denote the empirical measure associated to the observed data X. Let T_n be a smoothed version of P_n defined by, for L_n given in (3) below,

$$\langle T_n, \psi_{lk}\rangle = \langle P_n, \psi_{lk}\rangle 1\!\!1_{l\le L_n}. \tag{2}$$

For a given $\delta > 0$, let L_n be the *largest* integer such that

$$2^{L_n} \le n^{\frac{1}{2\delta+1}}. \tag{3}$$

Let $\beta_{\mathcal{M}_0(w)}$ denote the bounded-Lipschitz metric on $\mathcal{M}_0(w)$.

Theorem 2. *Let $X = (X_1, \ldots, X_n)$ be i.i.d. from law P_0 with density f_0. Let $f_0 \in \mathcal{C}^\alpha[0,1]$, for $\alpha \in (0,1]$ and suppose f_0 is bounded away from 0 on $[0,1]$. Let Π be a Pólya tree $PT(\mathcal{A})$ with parameters $\mathcal{A} = \{a_l,\ l \ge 1\}$, where $a_l = 2^{2l\delta} \vee 8$ for some $\delta > 0$.*

Let $\tau_{T_n} : f \to \sqrt{n}(f - T_n)$ and let $w = (w_l)_l$ be a weighting sequence such that $w_l/\sqrt{l} \uparrow \infty$. For any parameters $\alpha \in (0,1], \delta \le \alpha$, if T_n is given by (2), as $n \to \infty$,

$$\beta_{\mathcal{M}_0(w)}(\Pi(\cdot \mid X) \circ \tau_{T_n}^{-1}, \mathbb{G}_{P_0}) \to^{P_0} 0.$$

This is a nonparametric Bernstein-von Mises result in the space $\mathcal{M}_0(w)$ for density estimation, which is obtained for a class of non-n-dependent priors.

Applications. A natural use of Theorem 2 is the derivation of Bernstein-von Mises theorems for semiparametric functionals via the continuous mapping theorem. A prototypical example is the map $f \to \int_0^{\cdot} f = F(\cdot)$, leading to a Donsker-type result for the distribution function F, see [5] for a statement, paralleling Lo's [6] result for the Dirichlet process. Other smooth functionals can be considered as well, as in [2], [4]. Another important application, see [4], is the construction of confident credible bands.

References

[1] A. Barron, M. J. Schervish, and L. Wasserman. *The consistency of posterior distributions in nonparametric problems*, *Ann. Statist.*, 27(2): 536–561, 1999.

[2] I. Castillo and R. Nickl, *Nonparametric Bernstein-von Mises theorems in Gaussian white noise*, *Ann. Statist.*, 41(4):1999–2028, 2013.

[3] I. Castillo, *On Bayesian supremum norm contraction rates*, *Ann. Statist.*, 42(5):2058–2091, 2014.

[4] I. Castillo and R. Nickl, *On the nonparametric Bernstein-von Mises phenomenon*, *Ann. Statist.*, 42(5):1941–1969, 2014.

[5] I. Castillo, *On Pólya tree posterior densities*, *in preparation*, 2015.

[6] A.Y. Lo, *Weak convergence for Dirichlet processes*, *Sankhyā*, 45(1):105–111, 1983.

[7] S. Walker, *New approaches to Bayesian consistency*, *Ann. Statist.*, 32(5), 2028–2043, 2004.

Approximations of Functions on the Sphere

Friedrich Götze
(joint work with S. Bobkov, G. Chistyakov, A. Naumov, V. Ulyanov)

We investigate optimal Poincaré inequalities for classes of functions on the sphere S^{n-1} with respect to the uniform measure σ_{n-1} which are orthogonal to the class of restrictions of affine functions to S^{n-1}, say $\mathcal{A}$. Let $||f''(\theta)||_{HS}$ denote the Hilbert-Schmidt norm of the 2nd order derivative, say $f''(\theta)$, of f with respect to the tangent space at θ. We prove inequalities with optimal constant of type

$$\int f^2\, d\sigma_{n-1} \leq \frac{1}{2n(n+2)} \int \|f_S''\|_{\mathrm{HS}}^2\, d\sigma_{n-1}, \quad f \in \mathcal{A}^\perp,$$

For $||f_S''(\theta)|| \leq 1$ this leads to optimal exponential bounds of type

$$\int \exp\Big\{\frac{n-1}{2}\,|f|\Big\}\, d\sigma_{n-1} \leq \exp\Big\{\frac{n-1}{n+2} \int \|f_S''\|_{\mathrm{HS}}^2\, d\sigma_{n-1}\Big\}.$$

These results can be extended to Euclidean derivatives 2nd order derivative f'' of functions defined on a neighborhood of S^{n-1} in $\mathbf{R}^n$ with similar constants of asymptotic optimal size. More precisely, for "almost" orthogonal functions to $\mathcal{A}$, i.e. $|| \int \theta f(\theta) d\sigma_{n-1}(\theta)||_2 \leq b_0^{1/2} n^{-1}$ and $\int f d\sigma_{n-1} = 0$, we obtain for $\|f'' - aI_n\| \leq 1$ and $\int \|f'' - aI_n\|_{\mathrm{HS}}^2\, d\sigma_{n-1} \leq b^2$

$$\int \exp\Big\{\frac{n-1}{2(1+b_0^2+4\,b^2)}\,|f|\Big\}\, d\sigma_{n-1} \leq 2.$$

This is joint work with S. Bobkov and A. Chistyakov, [1].

This result applies in particular to the concentration of measure phenomenon for distribution of weighted empirical processes $e_n(\theta, \mathbf{X}) := \sum_{j=1}^n \theta_j(\delta_{X_j} - P)$ based on i.i.d. observations $X_j \sim P$ with weights θ taken from the sphere, like $h_n(\theta) := \mathbf{E}H(e_n(\theta, \mathbf{X}))$, $\quad \theta \in S^{n-1}$, where H denotes a smooth functional of e_n. The class of smooth symmetric functions on $\mathbf{R}^n$ arising in this context can be described by the properties

(1) $$h_n(\theta_{\pi(1)}, ..., \theta_{\pi(n)}) = h_n(\theta_1, ..., \theta_n) \text{ for all permutations } \pi,$$

(2) $$h_{n+1}(\theta_1, ..., \theta_n, 0) = h_n(\theta_1, ..., \theta_n);$$

(3) $$\frac{\partial}{\partial\theta_j} h_n(\theta_1, ..., \theta_j, ..., \theta_n)\Big|_{\theta_j=0} = 0 \quad \text{for all } j = 1, ..., n;$$

Let $\theta^d := \sum_{j=1}^n \theta_j^d$, $\quad |\theta|^d := \sum_{j=1}^n |\theta_j|^d$, $\quad d \geq 1$ denote the d-th power sums. Assuming that $h_n(\cdot), n \geq 1$, satisfies (1)–(3) and

(4) $$|D^\alpha h_n(\theta)| \leq B,$$

for all θ, some $B > 0$, $\quad \alpha = (\alpha_1, ..., \alpha_r), r \leq 3$ and $\alpha_j \geq 2$, $\quad \sum_{j=1}^r (\alpha_j - 2) \leq 1$. there is a function $h_\infty(0; \theta^2)$ such that the following "Berry-Esseen" type bound holds

$$|h_n(\theta) - h_\infty(0; \theta^2)| \leq c \cdot B \cdot \max((\theta^2)^{1/2}, (\theta^2)^{3/2})|\,\theta|^3, \quad c > 0 \text{ absolute}.$$

Moreover, assuming condition (4) the functions

$$h_\infty(\lambda_1, ..., \lambda_s; \lambda^2) := \lim_{k\to\infty} h_{k+s}\left(\lambda_1, ..., \lambda_s, \frac{\lambda}{\sqrt{k}}, ..., \frac{\lambda}{\sqrt{k}}\right).$$

exist. Assume for $h_n(\theta), n \geq 1$, conditions (1)–(3) and (4) with $s \geq 3$, as well as uniform derivatives up to $\sum_{j=1}^r (\alpha_j - 2) \leq s-2,\ \alpha_j \geq 2$. Then we have for $\theta \in S^{n-1}$ the "Edgeworth"-type expansion

$$h_n(\theta) = h_\infty(0;1) + \sum_{l=1}^{s-3} P_l(\theta^*\kappa^*) h_\infty(\lambda_1, ..., \lambda_s; 1)\big|_{\lambda_1=...=\lambda_s=0} + R_s,$$

where $|R_s| \leq c_s{\cdot}B{\cdot}|\theta|^s, \quad c_s > 0$, where "Edgeworth"-polynomials P_l of differential operators (at zero) are used. For example the polynomials P_1 and P_2 are given by

$$P_1(\theta^* D^*) = \frac{1}{6}\theta^3 D^3, \quad D^r := \frac{\partial^r}{\partial\lambda^r}, \ \ D^r D^l := \frac{\partial^r}{\partial\lambda_1^r}\frac{\partial^l}{\partial\lambda_2^l}$$

$$P_2(\theta^* D^*) = \frac{1}{24}\theta^4(D^4 - 3D^2D^2) + \frac{1}{72}(\theta^3)^2 D^3 D^3.$$

This is joint work with A. Naumov and V. Ulyanov, [2]. For further applications of the scheme to expansions in the "CLT" in Free Probability, see [3].
For symmetric functions of such type restricted to the sphere, we have, assuming conditions (1)–(3) and (4) with $s = 4$, that for $\theta \in S^{n-1}$

$$h_n(\theta) = h_\infty(0;1) + \frac{1}{6}\theta^3 D^3 h_\infty(\lambda;1)|_{\lambda=0} + O(\theta^4).$$

Here, a 2nd order concentration of measure phenomenon for $f_n(\theta) := h_n(\theta) - \int h_n d\sigma_{n-1}$ holds. This means that there exists $c > 0$ depending on B such that for $n \geq n_0$

$$\int \exp\left\{\frac{n-1}{c}\,|f_n|\right\} d\sigma_{n-1} \leq 2.$$

.

References

[1] S.G. Bobkov and G.P. Chistyakov and F. Götze, *Second order concentration on the sphere*, http://arxiv.org/pdf/1502.04178v1, (2015).

[2] F. Götze, A. Naumov and V. Ulyanov, *Asymptotic analysis of symmetric functions*, http://arxiv.org/pdf/1502.06267v1, (2015).

[3] F. Götze and A. Reshetenko, *Asymptotic Expansions in Free Limit Theorems*, http://arxiv.org/pdf/1408.1360v2, (2014).

Uncertainty quantification for high dimensional linear problems

Alexandra Carpentier
(joint work with Jens Eisert, David Gross, and Richard Nickl)

We consider the following prototypical high dimensional model. We observe noisy *inner products*

$$Y_i = \langle X^i, \theta \rangle + \varepsilon_i, \ \ i = 1, \dots, n;$$

The noise is, e.g., $\varepsilon_i \sim^{i.i.d.} N(0, \sigma^2), \sigma > 0$ and $\langle ., . \rangle$ is a scalar product in the space where the problem is posed. We consider two settings for this model (and we use unified notations in order to reduce the notational complexity).

- The *"vector" model*, which is the classic linear regression model in dimension $p > 0$. In this setting, the $X^i \in \mathbb{R}^p$ are sensing vectors, $\theta \in \mathbb{R}^p$ is the unknown parameter, and $\langle a, b \rangle = \sum_{j=1}^p a_j b_j$ is the classical scalar product in l_2, with associated norm $\|.\|$ (which is the l_2 norm). In this setting the number n of observations is *small* compared to dimension p but the vector θ is k-sparse and we write that $\theta \in \mathcal{M}(k)$.
- The *"matrix" model*, which is a linear regression model where the parameters and the design elements are matrices of dimension $d > 0$. In this setting, the $X^i \in \mathbb{C}^{d \times d}$ are sensing matrices, $\theta \in \mathbb{C}^{d \times d}$ is the unknown parameter, and $\langle A, B \rangle = \mathrm{tr}(A^* B)$ is the trace scalar product, with associated norm $\|.\|$ (which is the Frobenius norm). In this setting the number n of observations is *small* compared to dimension d^2 but the vector θ is of low rank k and we write that $\theta \in \mathcal{M}(k)$.

We consider two different design assumptions on the design elements $(X^i)_{i \leq n}$.

- The design is Gaussian and i.i.d.
- In the matrix model, we also consider Pauli random design, which is the one considered in *quantum tomography.* For this design, we use the *quantum constraint*, i.e.

$$\theta \in \{u \in \mathbb{C}^{d \times d}, u \succeq 0, \mathrm{tr}(u) = 1\}.$$

For these designs, there exists an estimator $\tilde{\theta}$ of θ that satisfies the following statements uniformly over $\theta \in \mathcal{M}(k)$ and with high probability.

- Vector model : if $k \log(p) \lesssim n$: $\|\tilde{\theta} - \theta\| \lesssim \sqrt{\frac{k \log p}{n}} := r(k)$.
- Matrix model : if $kd (\log d)^\gamma \lesssim n$: $\|\tilde{\theta} - \theta\| \lesssim \sqrt{\frac{kd}{n}} := r(k)$.

Here the unified notation $r(k)$ stands for the minimax optimal rate of estimation in both models [3, 4].

The main problem that we aim at solving is on *quantifying the uncertainty* on θ when the model $\mathcal{M}(k)$ is unknown, i.e. when k is unknown. More precisely, we want that this error *scales correctly with the unknown model k*.

There are works on "local" quantification of the uncertainty [8, 12, 13, 6] in high dimensional models but our objective is different. Our aim will be to construct the following objects for a given $k > 0$.

- *An adaptive and honest confidence set for θ over $\mathcal{M}(k)$*, i.e. a set C_n such that for a given $\alpha > 0$

$$\inf_{\theta\in\mathcal{M}(k)} P_\theta(\theta \in C_n) \geq 1-\alpha \quad \text{and} \quad \forall k_0 \leq k, \sup_{\theta\in\mathcal{M}(k_0)} E_\theta|C_n| \lesssim r(k_0),$$

  where $r(k_0)$ is the k-minimax-optimal rate of estimation of θ in $\|.\|$ norm.
- *An adaptive certificate for θ over $\mathcal{M}(k)$*, i.e. a stopping time and an associated estimator $(\hat{n}, \tilde{\theta}_{\hat{n}})$ that satisfy for a given $\epsilon > 0$

$$\sup_{\theta\in\mathcal{M}(k)} \|\tilde{\theta}_{\hat{n}} - \theta\| \lesssim^{whp} \epsilon \quad \text{and} \quad \forall k_0 \leq k, \sup_{\theta\in\mathcal{M}(k_0)} \hat{n} \lesssim^{whp} \frac{m(k_0)}{\epsilon^2},$$

  where $m(k_0)$ is the model complexity for estimation (respectively $m(k_0) = k_0 \log p$ for the vector model or $m(k_0) = k_0 d(\log d)^\gamma$ for the matrix model).

These two concepts are interesting because they aim at quantifying uncertainty in ways that depend on the unknown model $\mathcal{M}(k_0)$ to which θ belongs. They are also linked : if adaptive certificates exist, then adaptive confidence sets exist. If adaptive confidence sets exist for all n, then adaptive certificates exist.

[10] have proved that there are *no adaptive confidence sets (and therefore no adaptive certificates)* in the vector model. This is not very surprising when one considers the literature on non-parametric confidence sets that mainly consists of negative results [9, 2, 11, 7, 1]. On the other hand [5] have proved that in the matrix model, *confidence sets and adaptive certificates exist.*

This can seem surprising but it can be explained by the following testing argument. Consider, for $k_0 \leq k$, the testing problem

$$H_0 : \theta \in \mathcal{M}(k_0) \quad \text{vs} \quad H_1 : \theta \in \mathcal{M}(k), \quad \|\theta - \mathcal{M}(k_0)\| > \rho,$$

for $\rho > 0$. Let $\rho := \rho(k, k_0)$ be the *minimax-optimal rate of testing.* The following statements hold.

- If for n large enough, and for some $k_0 \leq k$, it holds that $\rho(k, k_0) \gg r(k_0)$, then adaptive and honest confidence sets, and therefore adaptive certificates, do not exist. This is the case for the vector model with large k, since a lower bound on $\rho(k, k_0)$ is

$$\min\Big(\sqrt{\frac{k\log(p)}{n}}, n^{-1/4}, p^{1/4}\sqrt{\frac{1}{n}}\Big) \gg \sqrt{\frac{\log(p)}{n}} = r(1).$$

- If for any n and for any $k_0 \leq k$, it holds that $\rho(k, k_0) \leq r(k_0)$, then adaptive and honest confidence sets and adaptive certificates do exist. This is the case for the matrix model since an upper bound on $\rho(k, k_0)$ is

$$\min\Big(n^{-1/4}, \sqrt{\frac{d}{n}}\Big) + \sqrt{\frac{k_0 d}{n}} \leq \sqrt{\frac{k_0 d}{n}} = r(k_0).$$

This highlights a main difference between the sparsity constraint and the low rank constraint. The sparsity constraint of the vector model induces a too radical dimension reduction for confidence sets (and adaptive certificates) whereas the dimension reduction induced by the low rank model is not too strict - the low rank model's dimension is kd and is always larger than the square root of the full dimension d^2.

In the case of the vector model, we are also interested in constructing adaptive and honest confidence sets in trace norm, since this norm is a meaningful concept for quantum tomography in the context of quantum states distinguishability. Without the quantum constraint, even in the case of Gaussian design, one can prove that if k is large enough, adaptive and honest confidence sets do not exist, which highlights a fundamental difference when comparing with the results in Frobenius norm. However, when one considers that the quantum constraint holds, then it is possible to prove that adaptive and honest confidence sets exist. This implies that for the trace norm that has a less regular geometry than the isotropic Frobenius norm, a shape constraint is necessary for uncertainty quantification.

There are many open problems that remain to be solved in this area, in particular the construction of really practical trace norm confidence intervals, the construction of trace norms certificates, the study of other confidence intervals in other norms, and the study of alternative methods for confidence interval construction like Bayesian methods of the Bootstrap.

References

[1] A. Bul, and R. Nickl, *Adaptive confidence sets in L^2* Probability Theory and Related Fields **156.3-4** (2013), 889–919.

[2] T. Cai, and M. Low, *An adaptation theory for nonparametric confidence intervals*, The Annals of statistics (2004), 1805–1840.

[3] E. Candes, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Communications on pure and applied mathematics **59.8** (2006), 1207–1223.

[4] E. Candes, and Y. Plan, *Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements.* IEEE Transactions on Information Theory, **57.4** (2011), 2342–2359.

[5] A. Carpentier, J. Eisert, D. Gross, and R. Nickl. *Uncertainty Quantification for Matrix Compressed Sensing and Quantum Tomography Problems.* arXiv preprint arXiv:1504.03234 (2015).

[6] A. Carpentier, and A. Kim, *An iterative hard thresholding estimator for low rank matrix recovery with explicit limiting distribution.* arXiv preprint arXiv:1502.04654 (2015).

[7] M. Hoffmann, and R. Nickl, *On adaptive inference and confidence bands*, The Annals of Statistics **39.5** (2011), 2383–2409.

[8] A. Javanmard, and A. Montanari, *Confidence intervals and hypothesis testing for high-dimensional regression*, The Journal of Machine Learning Research **15.1** (2014), 2869–2909.

[9] M. Low, *On nonparametric confidence intervals*, The Annals of Statistics **25.6** (1997), 2547–2554.

[10] R. Nickl, and S. van de Geer, *Confidence sets in sparse regression*, The Annals of Statistics **41.6** (2013), 2852–2876.

[11] J. Robins, and A. Van Der Vaart, *Adaptive non parametric confidence sets*, The Annals of Statistics **34.1** (2006), 229–253.

[12] S. Van de Geer, P. Bühlmann, Y. Ritov, and R. Dezeure, *On asymptotically optimal confidence regions and tests for high-dimensional models*, The Annals of Statistics **42.3** (2014), 1166–1202.

[13] C.H. Zhang, and S. Zhang, *Confidence intervals for low dimensional parameters in high dimensional linear models.* Journal of the Royal Statistical Society: Series B (Statistical Methodology) **76.1** (2014), 217–242.

On Hoeffding's One-Sided Inequality

Emmanuel Rio

In this talk we will give two improvements of Theorem 3 in Hoeffding (1963) for sums of independent random variables bounded on the right. The second improvement, which deals with large values of the deviation, will then be used to get additional results for sums of random variables with values in $[0, 1]$.

Colored microstructure noise, irregular sampling, and estimation of integrated volatility

Jean Jacod

(joint work with Yingying Li and Xinghua Zheng)

Our setting contains three basic ingredients.

1) An underlying one-dimensional Brownian semimartingale X:

$$X_t = X_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s,$$

with both processes b_t and σ_t being themselves Itô semimartingales with locally bounded characteristics (possibly with jumps), on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$.

2) At stage n, that is, for a given frequency of observations, the successive observations occur at times $0 = T(n,0) < T(n,1) < \cdots$ for a sequence $T(n,i)$ of finite stopping times, and we set

$$N_t^n = \sum_{i\geq 1} 1_{\{T(n,i)\leq t\}}, \qquad \Delta(n,i) = T(n,i) - T(n,i-i).$$

We suppose that, at stage n, the time lags $\Delta(n,i)$ are of the same order of magnitude Δ_n for all i, where $\Delta_n \to 0$ is a *non observable* sequence. More specifically,

(i) $\Delta_n N_t^n \stackrel{\mathbb{P}}{\longrightarrow} A_t := \int_0^t \alpha_s\, ds$, where α_t is an Itô semimartingale with locally bounded characteristics and $\alpha_t > 0$ and $\alpha_{t-} > 0$.

(ii) For all $s, t > 0$ the sequence $\Delta_n^{1/2+\rho'}\big(N_t^n - N^n_{(t-s\Delta_n^{\rho'})^+}\big)$ is bounded in probability, where $\rho' \in (0, 1/2)$.

(iii) There are a sequence τ_m of stopping times increasing to ∞, a sequence w_m of reals, and $\rho > 1/4$ and $\kappa \geq 8 \vee \frac{1}{1-2\rho'}$ such that

$$T(n,i) \leq \tau_m \;\Rightarrow\; \begin{cases} \big|\mathbb{E}\big(\alpha_{T(n,i)}\Delta(n,i+1) \mid \mathcal{F}_{T(n,i)}\big) - \Delta_n\big| \leq w_m \Delta_n^{1+\rho} \\ \mathbb{E}\big(|\alpha_{T(n,i)}\Delta(n,i+1)|^\kappa \mid \mathcal{F}_{T(n,i)}\big) \leq w_m \Delta_n^\kappa. \end{cases}$$

This assumption accommodates *regular sampling schemes*, of course, and *Poisson schemes* or *modulated Poisson schemes* and also *modulated random walk schemes.*

3) A *microstructure noise*, meaning that at time $T(n,i)$ the observation is not $X_{T(n,i)}$, but rather $Y_i^n = X_{T(n,i)} + \varepsilon_i^n$, with a noise ε_i^n. We suppose that the noise

can be realized as $\varepsilon_i^n = \gamma_{T(n,i)} \cdot \chi_i$, where γ_t is a nonnegative Itô semimartingale with locally bounded characteristics and $(\chi_i)_{i\in\mathbb{Z}}$ is a stationary process, independent of $\mathcal{F}_\infty = \bigvee_{t>0} \mathcal{F}_t$, centered with variance 1 and finite moments of all orders, and which is v-polynomially ρ-mixing for some $v > 3$. We denote by $r(m)$ the autocovariance of (χ_i) (so $r(0) = 1$), and observe that

$$|r(m)| \leq \frac{K}{(|m|+1)^v}, \qquad \text{so } R = \sum_{m\in\mathbb{Z}} r(m) \text{ is well defined.}$$

Our aim is to estimate $C_T = \int_0^T \sigma_s^2\, ds$, on the basis of the noisy observation within a *fixed* time interval $[0,T]$. When the noise is i.i.d., several methods can be used, see e.g. [5], [3], [1] or [2], but they all introduce a typically strong bias when the noise is colored. We show here how to take out this bias, using the pre-averaging method plus some new estimators for the covariance $r(m)$. This necessitates to average the data over two windows of h_n (for pre-averaging) and k_n (for estimating the noise) successive observations, with $2 \leq k_n < h_n$. For the estimation of the asymptotic variance we need two other window sizes h'_n and k'_n. We need $h_n \to \infty$, and the order of magnitude of k_n, h'_n, k'_n is constrained by the choice of h_n, according to a rule given below.

The choice of these tuning parameter is rather subtle, and is in principle driven by the unobserved "mean" time lag Δ_n if we wish asymptotic rate-efficiency (the efficient rate is $1/\Delta_n^{1/4}$. A proxy for Δ_n is $1/N_t^n$, which is observable. However, in practice the choice of h_n, h'_n, k_n, k'_n has to be determined by simulation studies, and below we pick the following choice (where $a_n \asymp b_n$ means that $1/K \leq a_n/b_n \leq K$ for some constant $K \geq 1$, and $\theta > 0$ is a constant).

$$h_n \sim \frac{\theta}{\sqrt{\Delta_n}}, \qquad h'_n \asymp \frac{1}{\Delta_n^\eta} \text{ with } \frac{1}{2} < \eta < \frac{3}{4}, \qquad k_n \asymp \frac{1}{\Delta_n^{1/5}}, \qquad k'_n \asymp \frac{1}{\Delta_n^{1/8}}$$

(once h_n is chosen, this amounts to having $h'_n \asymp h_n^{2\eta}$ and $k_n \asymp h_n^{2/5}$ and $k'_n \asymp h_n^{1/4}$).

For pre-averaging we need a weight (or, kernel) function g on $\mathbb{R}$, which satisfies

$$\begin{array}{l} g \text{ is continuous, piecewise } C^1 \text{ with a piecewise Lipschitz derivative } g', \\ s \notin (0,1) \ \Rightarrow\ g(s) = 0, \qquad \int_0^1 g(s)^2 ds \ > \ 0. \end{array}$$

We associate with g and the sequence h_n the following numbers (indexed by $n \geq 1$ and $i, j \in \mathbb{Z}$), and functions:

$$\begin{array}{lll} g_i^n = g(i/h_n), & \overline{g}_i^n \ = \ g_{i+1}^n - g_i^n & \\ \phi_j^n = \frac{1}{h_n} \sum_{i\in\mathbb{Z}} g_i^n\, g_{i-j}^n, & \overline{\phi}_j^n = h_n \sum_{i\in\mathbb{Z}} \overline{g}_i^n\, \overline{g}_{i-j}^n & \\ \phi(s) = \int g(u) g(u-s)\, du, & \overline{\phi}(s) = \int g'(u) g'(u-s)\, du & \\ \Phi_{00} = \int_0^1 \phi(s)^2\, ds, & \Phi_{01} = \int_0^1 \phi(s)\overline{\phi}(s)\, ds, & \Phi_{11} = \int_0^1 \overline{\phi}(s)^2\, ds. \end{array}$$

In the simple case when $g(x) = x \bigwedge (1-x)$ for $x \in (0,1)$, we have

$$\phi(0) = \frac{1}{12}, \quad \overline{\phi}(0) = 1, \quad \Phi_{00} = \frac{151}{80640}, \quad \Phi_{01} = \frac{1}{96}, \quad \text{and} \quad \Phi_{11} = \frac{1}{6}.$$

For any process V we write $V_i^n = V_{T(n,i)}$, and also $\Delta_i^n V = V_i^n - V_{i-1}^n$ (so, for example, $\Delta_i^n X$ is the ith return), and $\Delta_i^n Y = Y_i^n - Y_{i-1}^n$ is the ith observed (noisy) return. If V_i^n is any array of variables, we set

$$\widetilde{V}_i^n = \sum_{j=1}^{h_n-1} g_j^n\, \Delta_{i+j}^n V = -\sum_{j=0}^{h_n-1} \overline{g}_j^n\, V_{i+j}^n.$$

Below the function g is kept fixed, but we occasionally replace h_n by h'_n: in this case, $g_i'^n$, $\overline{g}'_i$, $\phi_j'^n$, $\overline{\phi}_j'^n$, $\widetilde{V}_i'^n$. We also use the following simple averages:

$$\overline{V}_i^n = \frac{1}{k_n} \sum_{j=0}^{k_n-1} V_{i+j}^n, \qquad \overline{V}_i'^n = \frac{1}{k'_n} \sum_{j=0}^{k'_n-1} V_{i+j}^n.$$

Then for $m = 0, \ldots, k_n$ we set

$$U(m)_t^n = \sum_{i=0}^{N_t^n - 5k_n} (Y_i^n - \overline{Y}_{i+2k_n}^n)(Y_{i+m}^n - \overline{Y}_{i+4k_n}^n).$$

We are now ready to exhibit our estimators for C_t:

$$\widehat{C}_t^n = \frac{1}{h_n\, \phi_0^n} \sum_{i=0}^{N_t^n - h_n} (\widetilde{Y}_i^n)^2 - \frac{1}{h_n^2\, \phi_0^n} \sum_{m=-k'_n}^{k'_n} \overline{phi}_m^n\, U(|m|)_t^n.$$

These estimators are consistent for estimating C_t, and enjoy a Central Limit Theorem, for any fixed time t. However, to make the CLT feasible, we need consistent estimators for the (conditional) variance. For this, with $m, m' = 0, \ldots, k_n$ we set

$$\begin{aligned}
\overline{U}(m,m')_t^n = &\sum_{i=0}^{N_t^n - 11k_n} (Y_i^n - \overline{Y}_{i+2k_n}^n)(Y_{i+m}^n - \overline{Y}_{i+4k_n}^n) \\
&\qquad (Y_{i+6k_n}^n - \overline{Y}_{i+8k_n}^n)(Y_{i+m'+6k_n}^n - \overline{Y}_{i+10k_n}^n) \\
V(m)_t^n = &\sum_{i=0}^{N_t^n - h'_n - 6k_n} (\widetilde{Y}_i'^n)^2 \big(Y_{i+h'_n+k_n}^n - \overline{Y}_{i+h'_n+3k_n}^n\big)\big(Y_{i+m+h'_n+k_n}^n - \overline{Y}_{i+h'_n+5k_n}^n\big)
\end{aligned}$$

$$V_t^{n,1} = \sum_{i=0}^{N_t^n - h'_n} (\widetilde{Y}_i'^n)^4, \quad V_t^{n,2} = \sum_{m=-k'_n}^{k'_n} V(|m|)_t^n, \quad V_t^{n,3} = \sum_{m,m'=-k'_n}^{k'_n} \overline{U}(|m|,|m'|)_t^n.$$

Finally we set

$$\Sigma_t^n = \frac{4}{\phi(0)^2} \Big(\frac{h_n}{h_n'^2}\, \frac{\Phi_{00}}{3\phi(0)^2}\, V_t^{n,1} + \frac{1}{h_n h'_n}\, \frac{2\Phi_{01}}{\phi(0)}\, V_t^{n,2} + \frac{1}{h_n^3}\, \Phi_{11} V_t^{n,3} \Big).$$

With all this notation, and under all previous assumptions, we obtain that for any $t > 0$ the sequence $\frac{1}{\Delta_n^{1/4}}\, (\widehat{C}_t^n - C_t)$ converges $\mathcal{F}_\infty$-stably in law to a limiting variable defined on an extension of the original space, and which is of the form

$Y_t = \int_0^t \beta_s \, dB_s$, where B is a standard Wiener process independent of $\mathcal{F}$ and β_t is the square-root of

$$\beta_t^2 \;=\; \frac{4}{\phi(0)^2}\Big(\Phi_{00}\,\frac{\theta\sigma_t^4}{\alpha_t} + 2\Phi_{01}\,\frac{\sigma_t^2\gamma_t^2}{\theta}\,R + \Phi_{11}\,\frac{\gamma_t^4\alpha_t}{\theta^3}\,R^2\Big)$$

and also $\frac{1}{\sqrt{\Delta_n}}\,\Sigma_t^n$ converges in probability to $\int_0^t \beta_s^2 \, ds$.

Therefore, for any $t > 0$, the sequence $(\widehat{C}_t^n - C_t)/\sqrt{\Sigma_t^n}$ converges stably in law to an $\mathcal{N}(0,1)$ variable, independent of $\mathcal{F}$.

The last claim is free from the unknown sequence Δ_n and allows us to construct confidence intervals for C_t in a standard way.

References

[1] Barndorff-Nielsen O.E., P. Hansen, A. Lunde and N. Shephard (2088). Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. *Econometrica* **76**, 1481-1536.

[2] Bibinger, M., Hautsch, N., Malec, P., Rei, M. (2014). Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *Annals of Statistics* **42**, 1312–1346.

[3] Jacod J., Y. Li, P. Mykland, M. Podolskij and M. Vetter (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stoch. Proc. Appl.* **119**, 2249–2276.

[4] Jacod J., Y. Li, and X. Zheng (2013). Statistical Properties of Microstructure Noise. *submitted*

[5] Zhang L. (2006). Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach. *Bernoulli* **12**, 1019-1043.

Approximation of spectra of integral operators by spectra of kernel random matrices based on Markov chains.

Radosław Adamczak

(joint work with Witold Bednorz)

Consider a measurable space $(\mathcal{X}, \mathcal{F})$ equipped with a probability measure π and let $h\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a symmetric measurable kernel, square integrable with respect to $\pi \otimes \pi$. Let $\mathbf{H}$ be the associated integral operator

$$\mathbf{H}f(x) = \int_{\mathcal{X}} h(x,y)f(y)\pi(dy). \tag{1}$$

$\mathbf{H}$ is a symmetric Hilbert-Schmidt operator on $L_2(\pi)$ and with a little bit of ambiguity we can identify its spectrum $\lambda(\mathbf{H})$ with an element of ℓ_2, the space of all real square-summable sequences (in this abstract, whenever considering a finite-dimensional operator, we append its spectrum with an infinite sequence of zeros).

In [9] Koltchinskii and Giné considered the problem of approximating $\lambda(\mathbf{H})$ by the spectrum of a random matrix $\mathbf{H}_n$, given by

$$\mathbf{H}_n = \frac{1}{n}((1-\delta_{ij})h(X_i, X_j))_{0 \le i,j \le n-1}, \tag{2}$$

where $X_0, X_1, \ldots$ is a sequence of i.i.d. random variables with law π and δ_{ij} stands for the Kronecker symbol. In particular they proved that

$$\int_{\mathcal{X}\times\mathcal{X}} h^2(x,y)\pi(dx)\pi(dy) < \infty$$

is equivalent to the almost sure convergence

$$d_2(\lambda(\mathbf{H}_n), \lambda(\mathbf{H})) \to 0,$$

where for $x, y \in \ell_2$,

$$\delta_2(x,y) = \inf_\sigma \Big(\sum_{i=0}^{\infty}(x_i - y_{\sigma(i)})^2\Big)^{1/2}$$

(the infimum is taken over all permutations of the set $\mathbb{N}$ of nonnegative integers). They also obtained corresponding limit theorems and rates of convergence under some stronger assumptions on h.

In [4] we investigated a counterpart of the above law of large numbers in the Markov chains setting, more precisely when $X_0, X_1, \ldots$ is a sample from a Harris ergodic Markov chain with invariant measure π. Recall that a Markov chain on $\mathcal{X}$ with transition function $P\colon \mathcal{X}\times\mathcal{F} \to [0,1]$ and invariant measure π is called Harris ergodic if for every initial point $x \in \mathcal{X}$,

$$\|P^n(x,\cdot) - \pi\|_{TV} \to 0,$$

where $\|\cdot\|_{TV}$ denotes the total-variation distance and P^n is the n-step transition function of the chain.

We consider two types of random matrices, the one defined by (2) as well as

$$\tilde{\mathbf{H}}_n = \frac{1}{n}(h(X_i, X_j))_{0\le i,j\le n-1} \tag{3}$$

(we remark that as pointed out in [9], even in the i.i.d. case, in general $\lambda(\tilde{\mathbf{H}}_n)$ may not approximate well the spectrum of $\mathbf{H}$, however it does so under some additional assumptions).

Our main result is

Theorem 1. *Let $\mathbf{X} = (X_n)_{n\ge 0}$ be a Harris ergodic Markov chain on $(\mathcal{X}, \mathcal{F})$ with invariant probability measure π and let $h\colon \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ be a symmetric measurable function. Assume that there exists $F\colon \mathcal{X} \to \mathbb{R}$, such that $\int_{\mathcal{X}} F^2(x)\pi(dx) < \infty$ and $|h(x,y)| \le F(x)F(y)$ for all $x, y \in \mathcal{X}$. Let $\mathbf{H}\colon L^2(\pi) \to L^2(\pi)$ be the linear operator given by (1) and $\tilde{\mathbf{H}}_n$, $\mathbf{H}_n$ be defined by (3),(2) respectively. Then for every initial measure μ of the chain $\mathbf{X}$, with probability one,*

$$\delta_2(\lambda(\tilde{\mathbf{H}}_n), \lambda(\mathbf{H})), \delta_2(\lambda(\mathbf{H}_n), \lambda(\mathbf{H})) \to 0.$$

The main tool in the proof of the above theorem is Nummelin's splitting technique for Markov chains, which allows to decompose the trajectory of a chain into one dependent blocks of random length (see [12, 14, 13, 7]). We use it to adapt the original argument of Koltchinskii and Giné. As a tool, which may be of independent interest, we also obtain a law of large numbers for U-statistics of order two of a Markov chain started at a point. This complements earlier results of many authors [1, 6, 8, 5].

Our main motivation for investigating convergence of spectra of kernel matrices based on Markov chains (besides its intrinsic mathematical interest) is the possibility of obtaining Markov Chain Monte Carlo algorithms for approximation of spectra of integral operators. From this perspective it is desirable to complement the law of large numbers with an appropriate exponential inequality. This goal in general cannot be achieved without imposing some restrictions on the kernel h and the Markov chain. The estimates we are able to obtain work for Mercer's type kernels and geometrically ergodic Markov chains, i.e. chains for which there exists $0 < \rho < 1$ such that for every $x \in \mathcal{X}$ and some constant $M(x)$, we have for every $n \geq 0$,

$$\|P^n(x, \cdot) - \pi\|_{TV} \leq M(x)\rho^n.$$

The tail inequality is given by the following theorem.

Theorem 2. *Let π be a probability measure on $(\mathcal{X}, \mathcal{F})$, where $\mathcal{X}$ is a metric space and $\mathcal{F}$ the Borel σ-field. Let $h\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a bounded function and $\mathbf{H}$ the corresponding kernel operator defined by* (1)*. Assume that there exist continuous functions $\phi_n\colon \mathcal{X} \to \mathbb{R}$, $n \in I$ (where $I = \{0, \ldots, R\}$ or $I = \mathbb{N}$) which form an orthonormal system in $L_2(\pi)$ and a sequence of non-negative numbers $\lambda = (\lambda_n)_{n\in I} \in \ell_2(I)$ such that we have a point-wise equality*

$$h(x, y) = \sum_{n\in I} \lambda_n \phi_n(x)\phi_n(y),$$

with the series converging absolutely and almost uniformly on $\mathcal{X} \times \mathcal{X}$. Assume furthermore that $\mathbf{X} = (X_n)_{n\geq 0}$ is a geometrically ergodic Markov chain with invariant measure π, started at a point z. Then

$$\mathbb{P}(\delta_2(\lambda(\tilde{\mathbf{H}}_n), \lambda(\mathbf{H})) \geq t) \leq 2 \exp\Big(- \frac{1}{L} n \min\Big(\frac{t^2}{\sup_{x\in\mathcal{X}} h(x, x)^2}, \frac{t}{\sup_{x\in\mathcal{X}} h(x, x)}\Big)\Big),$$

where the constant L depends only on the transition function P and the starting point z.

We remark that the constant L can be made explicit in terms of hitting times of certain sets or in terms of drift conditions guaranteeing geometric ergodicity (see [2, 3]). Let us also mention that in the i.i.d. case some inequalities of a similar flavour were obtained in [10, 11, 15].

References

[1] J. Aaronson, R. Burton, H. Dehling, D. Gilat, T. Hill, and B. Weiss. Strong laws for L- and U-statistics. *Trans. Amer. Math. Soc.*, 348(7):2845–2866, 1996.

[2] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13:no. 34, 1000–1034, 2008.

[3] R. Adamczak and W. Bednorz. Exponential concentration inequalities for additive functionals of Markov chains. To appear in ESAIM: Probability and Statistics.

[4] R. Adamczak, W. Bednorz, Some remarks on MCMC estimation of spectra of integral operators. To appear in Bernoulli.

[5] C. Andrieu, A. Jasra, A. Doucet, and P. Del Moral. On nonlinear Markov chain Monte Carlo. *Bernoulli*, 17(3):987–1014, 2011.
[6] M. A. Arcones. The law of large numbers for U-statistics under absolute regularity. *Electron. Comm. Probab.*, 3:13–19 (electronic), 1998.
[7] K. B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.*, 245:493–501, 1978.
[8] P. Bertail and S. Clémençon. A renewal approach to Markovian U-statistics. *Math. Methods Statist.*, 20(2):79–105, 2011.
[9] V. Koltchinskii and E. Giné. Random matrix approximation of spectra of integral operators. *Bernoulli*, 6(1):113–167, 2000.
[10] S. Mendelson and A. Pajor. Ellipsoid approximation using random vectors. In *Learning theory*, volume 3559 of *Lecture Notes in Comput. Sci.*, pages 429–443. Springer, Berlin, 2005.
[11] S. Mendelson and A. Pajor. On singular values of matrices with independent rows. *Bernoulli*, 12(5):761–773, 2006.
[12] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.
[13] E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete*, 43(4):309–318, 1978.
[14] E. Nummelin. *General irreducible Markov chains and nonnegative operators*, volume 83 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 1984.
[15] S. Smale and D.-X. Zhou. Geometry on probability spaces. *Constr. Approx.*, 30(3):311–323, 2009.

Optimalities in Network Analysis

Harrison Zhou
(joint work with Chao Gao, Yu Lu, Zongming Ma, Anderson Zhang)

Network science has become one of the most active research areas over the past few years. It has applications in many disciplines, for example, physics , sociology biology, and Internet. Detecting and identifying communities is fundamentally important to understand the underlying structure of the network. Many models and methodologies have been proposed for community detection from different perspectives, including RatioCut, Ncut and spectral method from computer science, Newman–Girvan Modularity from physics, semi-definite programming from engineering, and maximum likelihood estimation from statistics.

Deep theoretical developments have been actively pursued as well. Recently, celebrated works of Mossel et al. and Massoulie considered balanced two-community sparse network, and discovered the threshold phenomenon for both weak and strong consistency of community detection. Further extensions to slowly growing number of communities have been made. Recently in statistical literature, theoretical properties of various methods had been investigated as well, usually under weaker conditions, but the convergence rates are often sub-optimal.

Despite recent active and significant developments in network analysis, assumptions and conclusions can be very different in different papers. There is not a integrated framework on optimal community detection. In this paper, we attempt to give a fundamental and unified understanding of the community detection problem for Stochastic Block Model (SBM). Our framework is very general, including homogeneous and inhomogeneous SBM, dense and sparse networks, equal and non-equal community sizes, and finite and growing number of communities. For example, the connection probability can be as small as an order of $1/n$, or as large as a constant order, and the total number of communities can be as large as $n/\log n$. Under such framework, a sharp minimax result is obtained with an exponential rate. This result gives a clear and smooth transition from weak consistency (partial recovery) to strong consistency (exact recovery), i.e. clustering error rate from $o(1)$ to $o(n^{-1})$. As a consequence, we obtain phase transitions for non-consistency and strong consistency, under various settings, which recover the tight thresholds for phase transition in literature.

The Stochastic Block Model is possibly the most studied model in network community detection. Consider an undirected network with totally n nodes and K communities, labeled as $\{1, 2 \dots, K\}$. Each node is assigned to one community. Denote σ to be the assignment, and $\sigma(i)$ is the community assignment for the i-th node. Thus $n_k = |\{i : \sigma(i) = k\}|$ is the size for k-th community, for each $k \in \{1, 2, \dots, K\}$. We observe the connectivity of the network, which could be encoded into the adjacency matrix $\{A_{i,j}\}$ taking values in $\{0, 1\}^{n\times n}$. If there exists a connection between two nodes, $A_{i,j}$ equals 1, and 0 otherwise. We assume $A_{i,j}$ for any $i \geq j$ to be an independent Bernoulli random variable with success probability $\theta_{i,j}$. Let $\theta_{i,i} = 0$ (no self-loop) and $A_{i,j} = A_{j,i}$ (symmetry). In SBM, $\{\theta_{i,j}\}$ is assumed to have a block structure, in the sense that $\theta_{i,j} = \theta_{i',j'}$ when $\sigma(i) = \sigma(i')$ and $\sigma(j) = \sigma(j')$. We usually require that the within-community probabilities larger than the between-communities probabilities, as in reality individuals from the same community are more likely to be connected.

We consider a general SBM with parameter space defined as follows,

$$\Theta(n, K, a, b, \beta) \triangleq \Bigg\{ (\sigma, \{\theta_{i,j}\}) : \ \sigma : [n] \to [K]^n, n_k \in \left[\frac{n}{\beta K}, \frac{\beta n}{K}\right], \forall k \in [K],$$

$$\{\theta_{i,j}\} \in [0,1]^{n\times n}, \theta_{i,j} \geq \frac{a}{n} \text{ if } \sigma(i) = \sigma(j) \text{ and } \theta_{i,j} \leq \frac{b}{n} \text{ if } \sigma(i) \neq \sigma(j),\ \theta_{i,j} = \theta_{j,i} \Bigg\},$$

where $\beta \geq 1$ and is bounded. When $\beta = 1 + o(1)$, all communities have almost the same size. The parameters a/n and b/n have straightforward interpretation, with the former one as the smallest within-community probability and the later as the largest between-community probability. Throughout the paper, we assume $\epsilon < b < a$ and $a/n < 1 - \epsilon$ for a small constant $\epsilon > 0$, allowing the network to be very sparse or dense.

We use the mis-match ratio $r(\sigma, \hat{\sigma})$ to measure the performance of community detection. It is as the proportion of nodes mis-clustered by $\hat{\sigma}$ against the truth σ. The minimax rate for the parameter space $\Theta(n, K, a, b, \beta)$ in terms of the mis-match ratio loss is given in the following theorem.

Theorem 1. *Assume* $\frac{nI}{K\log K}\to\infty$,

$$(1)\qquad \inf_{\hat\sigma}\sup_{\Theta(n,K,a,b,\beta)}\mathbb{E}r(\hat\sigma,\sigma)=\begin{cases}\exp\big(-(1+o(1))\frac{nI}{2}\big), k=2,\\ \exp\big(-(1+o(1))\frac{nI}{\beta k}\big), k\geq 3, 1\leq\beta<c,\end{cases}$$

where $c\leq\sqrt{5/3}$. *In addition if* $nI/K=O(1)$, *there are at least a constant proportion of nodes mis-clustered, i.e.* $\inf_{\hat\sigma}\sup_{\Theta(n,K,a,b,\beta)}\mathbb{E}r(\hat\sigma,\sigma)\geq c_1$.

Bayesian Clustering of Functional Data Using Local Features

Subhashis Ghosal
(joint work with Adam Suarez)

Most traditional clustering techniques for functional data apply multivariate clustering methods on a vector of estimated basis coefficients, assuming that the underlying signal functions live in the L2-space. Bayesian methods use models which imply the belief that some observations are realizations from some signal plus noise models with identical underlying signal functions. The method we propose differs in this respect: we employ a model that does not assume that any of the signal functions are truly identical, but possibly share many of their local features, represented by coefficients in a multiresolution wavelet basis expansion. We cluster each wavelet coefficient of the signal functions using conditionally independent Dirichlet process priors. Each possible clustering gives rise to a model, and an uncountable number of such models exist. An appropriate topology on the model space is given by the product of discrete topologies on the partitions of each wavelet coefficients. Under the asymptotic regime that the noise level goes to zero but with a fixed number of subjects, we show that the posterior probability of every neighborhood of the true clustering pattern among subjects tends to one in probability, giving a frequentist justification of the proposed Bayesian clustering procedure. We describe efficient Markov chain Monte Carlo computing techniques for the posterior distribution and a method of identifying the posterior expected cluster. We demonstrate the proposed method using the popular Canadian weather data.

Upper and lower bounds for suprema of canonical processes

Rafał Latała
(joint work with Tomasz Tkocz)

In many problems arising in probability theory and its applications one needs to estimate the supremum of a stochastic process. In particular it is very useful to be able to find two-sided bounds for the mean of the supremum. The modern approach to this challenge is based on the chaining methods (cf. the recent monograph of Michel Talagrand [8]).

In this talk, based on [4], we are going to discuss the class of *canonical processes* (X_t) of the form $X_t=\sum_{i=1}^{\infty}t_iX_i$, where X_i are independent random variables.

If X_i are *standardized*, i.e. have mean zero and variance one, then this series converges a.s. for $t \in \ell^2$ and one may try to estimate $\mathbb{E}\sup_{t\in T} X_t$ for $T \subset \ell^2$. In fact it is more convenient to work with the quantity $\mathbb{E}\sup_{s,t\in T}(X_t - X_s)$. Observe however that if the set T or the variables X_i are symmetric then

$$\mathbb{E}\sup_{s,t\in T}(X_s - X_t) = \mathbb{E}\sup_{s\in T} X_s + \mathbb{E}\sup_{t\in T}(-X_t) = 2\mathbb{E}\sup_{t\in T} X_t.$$

In the case when X_i are i.i.d. $\mathcal{N}(0,1)$ r.v.s, X_t is the canonical Gaussian process. Moreover, any centered separable Gaussian process has the Karhunen-Loève representation of such form. For Gaussian processes the behaviour of $\sup_{t\in T} X_t$ is related to the geometry of the metric space (T, d_2), where d_2 is the ℓ^2-metric $d_2(s,t) = (\mathbb{E}|X_s - X_t|^2)^{1/2}$. The celebrated Fernique-Talagrand [2, 6] majorizing measure bound can be expressed in the form

$$\frac{1}{C}\gamma_2(T) \le \mathbb{E}\sup_{t\in T} X_t \le C\gamma_2(T).$$

Here and in the sequel C denotes a universal constant,

$$\gamma_2(T) := \inf\sup_{t\in T}\sum_{n=0}^{\infty} 2^{n/2}\Delta_2(A_n(t)),$$

the infimum runs over all admissible sequences of partitions $(\mathcal{A}_n)_{n\ge 0}$ of the set T, $A_n(t)$ is the unique set in $\mathcal{A}_n$ which contains t, and Δ_2 denotes the ℓ^2-diameter. An increasing sequence of partitions $(\mathcal{A}_n)_{n\ge 0}$ of T is called *admissible* if $\mathcal{A}_0 = \{T\}$ and $|\mathcal{A}_n| \le N_n := 2^{2^n}$ for $n \ge 1$.

Let us start with a general simple upper bound. For $p \ge 1$ and a finite set T we have

$$\mathbb{E}\sup_{s,t\in T}(X_s - X_t) \le \Big(\mathbb{E}\sup_{s,t\in T}|X_s - X_t|^p\Big)^{1/p} \le \Big(\mathbb{E}\sum_{s,t\in T}|X_s - X_t|^p\Big)^{1/p}$$
$$\le |T|^{2/p}\sup_{s,t\in T}\|X_s - X_t\|_p.$$

Hence

$$|T| \le e^p \quad \Rightarrow \quad \mathbb{E}\sup_{s,t\in T}(X_s - X_t) \le e^2\Delta_p(T),$$

where $\Delta_p(T)$ is the diameter of T with respect to the metric

$$d_p(s,t) = \|X_s - X_t\|_p = (\mathbb{E}|X_t - X_s|^p)^{1/p}.$$

It is natural to ask whether this estimate may be reversed.

Definition. We say that a process $(X_t)_{t\in T}$ satisfies *the Sudakov minoration principle with constant* $\kappa > 0$ if for any $p \ge 1$, $S \subset T$ with $|S| \ge e^p$ such that $\|X_s - X_t\|_p \ge u$ for all $s, t \in S$, $s \ne t$ we have

$$\mathbb{E}\sup_{s,t\in S}(X_s - X_t) \ge \kappa u.$$

In the case of centered Gaussian process $(G_t)_{t\in T}$ we have $\|G_s-G_t\|_p \sim \sqrt{p}\|G_s-G_t\|_2$ and it is not hard to see that the Sudakov minoration principle in the sense above is equivalent to the classical one:

$$\mathbb{E}\sup_{t\in S} G_t \geq \frac{1}{C}u\sqrt{\log|S|} \quad \text{if } (\mathbb{E}|G_t-G_s|^2)^{1/2} \geq u \text{ for all } s,t\in S,\ s\neq t.$$

Results of [7] and [3] imply that canonical processes based on symmetric r.v's with log-concave tails satisfy Sudakov minoration.

It is easy to check that for a symmetric variable Y with a log-concave tail we have $\|Y\|_p \leq C\frac{p}{q}\|Y\|_q$ for $p\geq q\geq 2$. This motivates the following definition.

Definition. For $\alpha \geq 1$ we say that moments of a random variable X *grow α-regularly* if $\|X\|_p \leq \alpha\frac{p}{q}\|X\|_q$ for $p\geq q\geq 2$.

Theorem 1. *Suppose that $X_1, X_2, \ldots$ are independent standardized r.v.s and moments of X_i grow α-regularly for some $\alpha \geq 1$. Then the canonical process $X_t = \sum_{i=1}^{\infty} t_i X_i$, $t\in\ell^2$ satisfies the Sudakov minoration principle with constant $\kappa(\alpha)$, which depends only on α.*

In fact the assumption on regular growth of moments is necessary for the Sudakov minoration principle in the i.i.d. case.

Proposition 2. *Suppose that a canonical process $X_t = \sum_{i=1}^{\infty} t_i X_i$, $t\in\ell^2$ based on i.i.d. standardized random variables X_i satisfies the Sudakov minoration with constant $\kappa > 0$. Then moments of X_i grow C/κ-regularly.*

Methods developed to prove Theorem 1 allow also to show the following result.

Theorem 3. *Let X_t be as in Theorem 1. Then for any $\emptyset \neq T \subset \ell_2$ and $p\geq 1$,*

$$\left(\mathbb{E}\sup_{t,s\in T}|X_t-X_s|^p\right)^{1/p} \leq C(\alpha)\left(\mathbb{E}\sup_{t,s\in T}|X_t-X_s| + \sup_{t,s\in T}(\mathbb{E}|X_t-X_s|^p)^{1/p}\right).$$

Let us try to refine the simple bound leading to the Sudakov minoration employing this time a chaining argument. We follow closely Talagrand's construction of the γ_2 functional. Let $(X_t)_{t\in T}$ be a general process with T finite (for simplicity). We fix an increasing sequence of admissible partitions $(\mathcal{A}_n)_{n\geq 0}$. For each n we construct a set T_n by choosing exactly one point from every set A of the partition $\mathcal{A}_n$. Hence, $|T_n| \leq 2^{2^n}$. We pick $\pi_n(t)\in T_n$ in such a way that t and $\pi_n(t)$ belong to the same set in the partition $\mathcal{A}_n$. We have $X_t - X_{\pi_1(t)} = \sum_{n\geq 1}\left(X_{\pi_{n+1}(t)} - X_{\pi_n(t)}\right)$. Hence for $u\geq 16$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in T}|X_t - X_{\pi_1(t)}| \geq u\sup_{t\in T}\sum_{n\geq 1} d_{2^n}(\pi_{n+1}(t),\pi_n(t))\right)\\
&\leq \mathbb{P}\left(\exists_{n\geq 1}\exists_{t\in T}|X_{\pi_{n+1}(t)} - X_{\pi_n(t)}| \geq u d_{2^n}(\pi_{n+1}(t),\pi_n(t))\right)\\
&\leq \sum_{n\geq 1}\sum_{s\in T_n, s'\in T_{n+1}} \mathbb{P}(|X_s - X_{s'}| \geq u d_{2^n}(s,s')) \leq \sum_{n\geq 1}|T_n||T_{n+1}|u^{-2^n} \leq \frac{128}{u^2}.
\end{aligned}$$

Therefore

$$\mathbb{E}\sup_{t,s\in T}(X_t - X_s) \le 2\mathbb{E}\sup_{t\in T}|X_t - X_{\pi_1(t)}| + \mathbb{E}\sup_{s,t\in T}|X_{\pi_1(t)} - X_{\pi_1(s)}|)$$
$$\le C\sup_{t\in T}\sum_{n\ge 1} d_{2^n}(\pi_{n+1}(t),\pi_n(t)) + |T_1|^2\cdot\Delta_1(T).$$

Since $\pi_n(t)\in A_n(t)$ and $\pi_{n+1}(t)\in A_{n+1}(t)\subset A_n(t)$ we have $d_{2^n}(\pi_{n+1}(t),\pi_n(t))\le \Delta_{2^n}(A_n(t))$, where $\Delta_{2^n}(A_n(t))$ is the d_{2^n}-diameter of the unique set $A_n(t)$ from $\mathcal{A}_n$ containing t.

The bound obtained above motivates the following defininition

$$\gamma_X(T) := \inf\sup_{t\in T}\sum_{n=0}^{\infty}\Delta_{2^n}(A_n(t)),$$

where the infimum runs over all admissible sequences of partitions $(\mathcal{A}_n)$ of T.

We have thus shown the following bound (observed also in [5]).

Theorem 4. *For* any *process* $(X_t)_{t\in T}$*,*

$$\mathbb{E}\sup_{s,t\in T}(X_s - X_t) \le C\gamma_X(T).$$

It is not hard to see that reversing the above γ_X-bound requires the Sudakov minoration, so in the case of canonical processes the regular growth of moments is necessary. Unfortunately we need one more technical assumption.

Definition. For $\beta<\infty$ we say that moments of a random variable X *grow with speed* β if $\|X\|_{\beta p}\ge 2\|X\|_p$ for $p\ge 2$.

Theorem 5. *Let* $X_t=\sum_{i=1}^{\infty} t_iX_i$*,* $t\in\ell^2$ *be the canonical process based on independent standardized r.v.s* X_i *with moments growing* α*-regularly with speed* β *for some* $\alpha\ge 1$ *and* $\beta>1$*. Then for any nonempty* $T\subset\ell^2$*,*

$$\frac{1}{C(\alpha,\beta)}\gamma_X(T)\le\mathbb{E}\sup_{s,t\in T}(X_s-X_t)\le C\gamma_X(T).$$

Corollary 6. *Let* X_t *be as in the main theorem. Then for any nonempty* $T\subset\ell^2$ *and any process* $(Y_t)_{t\in T}$ *such that for all* $p\ge 1$ *and* $s,t\in T$*,* $\|Y_s-Y_t\|_p\le\|X_s-X_t\|_p$ *we have*

$$\mathbb{P}\left(\sup_{s,t\in T}(Y_s-Y_t)\ge u\right)\le C(\alpha,\beta)\mathbb{P}\left(\sup_{s,t\in T}(X_s-X_t)\ge\frac{1}{C(\alpha,\beta)}u\right)\quad \textit{for } u>0.$$

Another consequence of Theorem 5 is the following convex-hull bound.

Corollary 7. *Let* X_t *be as in our main theorem and let nonempty set* $T\subset\ell^2$ *be such that* $\mathbb{E}\sup_{s,t\in T}(X_s-X_t)<\infty$*. Then there exist* $t^1,t^2,\ldots\in\ell^2$ *such that*

$$T-T\subset\overline{\mathrm{conv}}\{\pm t^n:\ n\ge 1\}$$

and

$$\|X_{t^n}\|_{\log(n+2)}\le C(\alpha,\beta)\mathbb{E}\sup_{s,t\in T}(X_s-X_t)\quad\textit{for all } n.$$

Remark. The reverse statement easily follows by the union bound and Chebyshev's inequality. Namely, for any canonical process $(X_t)_{t\in\ell^2}$ and any nonempty set $T\subset\ell^2$ such that $T-T\subset\overline{\mathrm{conv}}\{\pm t^n\colon\ n\geq 1\}$ and $\|X_{t^n}\|_{\log(n+2)}\leq M$ one has $\mathbb{E}\sup_{s,t\in T}(X_s-X_t)\leq CM$. Indeed,

$$\mathbb{P}\Big(\sup_{s\in T-T}X_s\geq uM\Big)\leq\mathbb{P}\Big(\sup_{n\geq 1}X_{\pm t^n}\geq uM\Big)\leq\sum_{n\geq 1}\mathbb{P}(|X_{t^n}|\geq u\|X_{t^n}\|_{\log(n+2)})$$
$$\leq\sum_{n\geq 1}u^{-\log(n+2)}$$

and integration by parts yields $\mathbb{E}\sup_{s,t\in T}(X_s-X_t)=\mathbb{E}\sup_{s\in T-T}X_s\leq CM$.

Let $(\varepsilon_i)_{i\geq 1}$ be i.i.d. symmetric ±1-valued r.v.s, $X_t=\sum_{i=1}^{\infty}t_i\varepsilon_i$, $t\in\ell^2$ and $T=\{e_n\colon\ n\geq 1\}$, where (e_n) is the canonical basis of ℓ^2. Then obviously $\mathbb{E}\sup_{s,t\in T}(X_s-X_t)=2$, moreover for any $A\subset T$ with cardinality at least 2, we have $\Delta_{2^k}(A)\geq\Delta_2(A)=\sqrt{2}$, hence $\gamma_X(T)=\infty$. Therefore one cannot reverse γ_X-bound for Bernoulli processes, so some assumptions on the nontrivial speed of growth of moments are necessary to get two-sided γ_X estimate.

However, the convex hull bound holds for Bernoulli processes [1] and we believe that it holds for canonical processes based on r.v's with regular growth of moments.

References

[1] W. Bednorz and R. Latała, *On the boundedness of Bernoulli processes*, Ann. of Math. (2) **180** (2014), 1167–1203.

[2] X. Fernique, *Regularité des trajectoires des fonctions aléatoires gaussiennes*, École d'Été de Probabilités de Saint-Flour, IV-1974, Lecture Notes in Math. 480, 1–96, Springer 1975.

[3] R. Latała, *Sudakov minoration principle and supremum of some processes*, Geom. Funct. Anal. **7** (1997), 936–953.

[4] R. Latała and T. Tkocz, *A note on suprema of canonical processes based on random variables with regular moments*, Electron. J. Probab. **20** (2015), no. 36, 1–17.

[5] S. Mendelson and G. Paouris, *On generic chaining and the smallest singular values of random matrices with heavy tails*, J. Funct. Anal. **262** (2012), 3775–3811.

[6] M. Talagrand, *Regularity of Gaussian processes*, Acta Math. **159** (1987), 99–149.

[7] M. Talagrand, *The supremum of some canonical processes*, Amer. J. Math. **116** (1994), 283–325.

[8] M. Talagrand, *Upper and lower bounds for stochastic processes. Modern methods and classical problems*, Springer 2014.

Estimation of functionals under sparsity constraints

Alexandre Tsybakov

(joint work with Olivier Collier, Laëtitia Comminges, Nicolas Verzelen)

Consider the model

$$y_j = \theta_j + \sigma\xi_j, \quad j = 1, \dots, d, \tag{1}$$

where $\theta = (\theta_1, \dots, \theta_d) \in \mathbb{R}^d$ is an unknown vector of parameters, ξ_j are i.i.d. standard normal random variables, and $\sigma > 0$ is the noise level. Based on the observations $y_1, \dots, y_d$, we want to estimate the functionals

- $L(\theta) = \sum_{i=1}^d \theta_i$
- $Q(\theta) = \sum_{i=1}^d \theta_i^2$
- $\|\theta\|_2 = \sqrt{Q(\theta)}$,
- $L_\gamma(\theta) = \sum_{i=1}^d |\theta_i|^\gamma, \quad 0 < \gamma \le 1.$

We assume that θ belongs to a given subset Θ of $\mathbb{R}^d$. We consider two possible choices for Θ:

$$B_0(s) = \{\theta \in \mathbb{R}^d : \|\theta\|_0 \le s\}, \quad \text{and} \quad B_q(r) = \{\theta \in \mathbb{R}^d : \|\theta\|_q \le r\}$$

where $\|\theta\|_0$ denotes the number of non-zero components of θ, and

$$\|\theta\|_q = \left(\sum_{i=1}^d |\theta_i|^q\right)^{1/q}$$

for $0 < q < \infty$. The integer s and $r > 0$ are given constants.

Let $T(\theta)$ be one of the functionals defined above. We measure the performance of an estimator $\hat{T}$ of the functional $T(\theta)$ by the maximum squared risk $\sup_{\theta\in\Theta} E_\theta(\hat{T} - T(\theta))^2$ where E_θ is the expectation with respect to the joint distribution of $y_1, \dots, y_n$ satisfying (1). The best possible quality is characterized by the minimax risk

$$R_T^*(\Theta) = \inf_{\hat{T}} \sup_{\theta\in\Theta} E_\theta(\hat{T} - T(\theta))^2,$$

where $\inf_{\hat{T}}$ denotes the infimum over all estimators. We construct minimax optimal estimators of $T(\theta)$, i.e., estimators $\tilde{T}$ such that

$$\sup_{\theta\in\Theta} E_\theta(\tilde{T} - T(\theta))^2 \asymp R_T^*(\Theta).$$

Here, and below the sign $a \asymp b$ means that $c_1 \le a/b \le c_2$ for some absolute constants $c_1, c_2 > 0$. We study non-asymptotic behavior of the minimax risk on the classes $\Theta = B_0(s)$ and $\Theta = B_q(r)$ for all $1 \le s \le d$, $\sigma, r > 0$, $0 < q \le 2$. For the class $B_0(s)$ we obtain the following results:

$$\begin{aligned} R_L^*(B_0(s)) &\asymp \sigma^2 s^2 \log(1 + d/s^2), \\ R_{L_\gamma}^*(B_0(s)) &\asymp \sigma^{2\gamma} s^2 \log^\gamma(1 + d/s^2) \quad (\text{for } 1 \le s \le \sqrt{d}), \end{aligned} \tag{2}$$

$$\begin{aligned} R_Q^*(B_2(r) \cap B_0(s)) &\asymp \min\{r^4, \max\{\sigma^2 r^2, \psi_\sigma(s,d)\}\}, \\ R_{\sqrt{Q}}^*(B_0(s)) &\asymp \sqrt{\psi_\sigma(s,d)}, \end{aligned} \tag{3}$$

where

$$\psi_\sigma(s,d) = \begin{cases} \sigma^4 s^2 \log^2(1+d/s^2) & \text{if} \quad s < \sqrt{d}, \\ \sigma^4 d & \text{if} \quad s \geq \sqrt{d}. \end{cases}$$

For the quadratic functional Q, we consider in (3) a smaller class $B_2(r) \cap B_0(s)$ rather than $B_0(s)$ since $R^*_Q(B_0(s)) = \infty$.

For the classes $B_q(r)$, the behavior of the minimax risks is characterized in terms of the integer $m = m(B)$ defined for any $B \in \mathcal{B} := \{B_q(r) : 0 < q \leq 2, r > 0\}$ as follows:

$$m = \max\{s \in \mathbb{N} : \sigma^2 \log(1+d/s^2) \leq r^2 s^{-2/q}\} \tag{4}$$

if the set in (4) is non-empty, and $m = 0$ otherwise. Then, for all $0 < q \leq 1$, $r > 0$,

$$R^*_L(B_q(r)) \asymp \begin{cases} \sigma^2 d & \text{if} \quad m > \sqrt{d}, \\ \sigma^2 (r/\sigma)^{2q} \log^{1-q}(1+d(\sigma/r)^{2q}) & \text{if} \quad 1 \leq m \leq \sqrt{d}, \\ r^2 & \text{if} \quad m = 0, \end{cases}$$

and for all $0 < q \leq 2$, $r > 0$,

$$R^*_Q(B_q(r)) \asymp \begin{cases} \max\{\sigma^2 r^2, \sigma^4 d\} & \text{if} \quad m > \sqrt{d}, \\ \max\{\sigma^2 r^2, \sigma^4 (\frac{r}{\sigma})^{2q} \log^{2-q}(1+d(\frac{\sigma}{r})^{2q})\} & \text{if} \quad 1 \leq m \leq \sqrt{d}, \\ r^4 & \text{if} \quad m = 0. \end{cases}$$

$$R^*_{\sqrt{Q}}(B_q(r)) \asymp \begin{cases} \sigma^2 \sqrt{d} & \text{if} \quad m > \sqrt{d}, \\ \sigma^2 (\frac{r}{\sigma})^{q} \log^{1-q/2}(1+d(\frac{\sigma}{r})^{2q}) & \text{if} \quad 1 \leq m \leq \sqrt{d}, \\ r^2 & \text{if} \quad m = 0. \end{cases}$$

In all the cases considered above, we explicitly construct estimators that achieve the minimax rates, cf. [1]. For the classes $B_0(s)$, these estimators depend on s, and for the classes $B_q(r)$, they depend on q and r. A natural question is whether such rates can be attained adaptively, i.e., on the estimators independent of these parameters. We show that the answer to this question is negative for the problem of adaptive estimation of the linear functional $L(\cdot)$. Namely, we prove that the adaptive rates are different from the minimax rates given above. The aim is to construct an estimator of $L(\theta)$ that adapts simultaneously to sparsity s when θ belongs to the class $B_0(s)$, $1 \leq s \leq d$, and to parameters q, r when θ belongs to the class $B_q(r)$, $0 < q \leq 1$, $r > 0$. For this purpose, we consider selection from a family of estimators indexed by integer $k \in [0, \sqrt{d \log d}/2] \cup \{d\}$. Estimators in this family are defined as follows:

$$\hat{L}_k = \begin{cases} \sum_{j=1}^d y_j & \text{for} \quad k = d, \\ \sum_{j=1}^d y_j \ \mathbf{1}_{\{|y_j| > \sigma x_k\}} & \text{for} \quad 1 \leq k \leq \sqrt{d \log d}/2, \\ 0 & \text{if} \quad k = 0. \end{cases}$$

Here, $x_k = A\sqrt{\log\left(1 + \frac{d \log d}{k^2}\right)}$ for a large enough absolute constant $A > 0$ (given explicitly), and we assume that $d \geq 3$. We define the adaptive estimator as $\tilde{L} = \hat{L}_{\hat{k}}$ where

$$\hat{k} = \inf\left\{k \in [0, \sqrt{d \log d}/2] \cup \{d\} : |\hat{L}_k - \hat{L}_{k'}| \leq \omega_{k'}, \forall k' > k\right\},$$

$$\omega_k = \begin{cases} A_1\sigma k\sqrt{\log\left(1+\frac{d\log d}{k^2}\right)} & \text{for } k\in[1,\sqrt{d\log d}/2],\\ A_2\sigma\sqrt{d\log d} & \text{for } k=d,\end{cases}$$

and A_1, A_2 are explicitly given absolute positive constants. We show that there exists an absolute constant $C > 0$ such that

$$\sup_{\theta\in B_0(s)} E_\theta(\tilde{L}-L(\theta))^2 \le C\sigma^2 s^2\log\left(1+\frac{d\log d}{s^2}\right), \quad ,\forall\, 1\le s\le d. \tag{5}$$

Thus, the rate of adaptive estimation differs from the minimax rate only in that we replace d by $d\log d$ under the logarithm. This is the price for not knowing the value of s. If $s = d^a$, $a \in (0, 1/2)$, the right hand side of (5) is of the order $\sigma^2 s^2 \log d$, which coincides with the minimax rate for such s. For $\sqrt{d\log d} \le s \le d$ the right hand side of (5) is of the order $\sigma^2 d\log d$, while the minimax rate is of the order $\sigma^2 d$.

For the adaptive estimation on the scale of classes $\mathcal{B}' := \{B_q(r) : 0 < q \le 1, r > 0\}$, we get the following bound. There exists an absolute constant $C > 0$ such that

$$\inf_{\hat{T}}\sup_{B\in\mathcal{B}'}\sup_{\theta\in B}\frac{E_\theta(\hat{T}-L(\theta))^2}{\Psi_\sigma(B)} \le C$$

where $\inf_{\hat{T}}$ is the infimum over all estimators, and

$$\Psi_\sigma(B) = \begin{cases} \sigma^2 m^2(B)\log\left(1+\frac{d\log d}{m^2(B)}\right) & \text{if } \; m(B)\ge 1,\\ r^2 & \text{if } \; m(B)=0.\end{cases}$$

We also establish lower bounds showing that $\Psi_\sigma(B)$ is the adaptive rate of convergence that cannot be improved on the scale of classes $\mathcal{B}'$ in the sense described in [2].

References

[1] O. Collier, L. Comminges and A. B. Tsybakov, *Minimax estimation of linear and quadratic functionals on sparsity classes*, `arxiv:1502.00665`.

[2] A. B. Tsybakov, *Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes*, Annals of Statistics **26** (1998), 2420–2469.

Achieving Optimal Misclassification Proportion in Stochastic Block Model

Zongming Ma

(joint work with Chao Gao, Anderson Y. Zhang, Harrison H. Zhou)

Community detection is a fundamental statistical problem in network data analysis. Many algorithms have been proposed to tackle this problem. Most of these algorithms are not guaranteed to achieve the statistical optimality of the problem, while procedures that achieve information theoretic limits for general parameter spaces are not computationally tractable. In this paper, we present a computationally feasible two-stage method that achieves optimal statistical performance in misclassification proportion for stochastic block model under weak regularity conditions. Our two-stage procedure consists of a generic refinement step that can take a wide range of weakly consistent community detection procedures as initializer, to which the refinement stage applies and outputs a community assignment achieving optimal misclassification proportion with high probability. The practical effectiveness of the new algorithm is demonstrated by competitive numerical results.

References

[1] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou, Achieving optimal misclassification proportion in Stochastic Block Model, ArXiv preprint arXiv:1505.03772.

Reporter: Benjamin Stucky

Participants

Prof. Dr. Radoslaw Adamczak
Institute of Mathematics
University of Warsaw
ul. Banacha 2
02-097 Warszawa
POLAND

Nikolay Baldin
Mathematische Statistik
Humboldt-Universität Berlin
Rudower Chaussee 25
10099 Berlin
GERMANY

Dr. Denis Belomestny
Fakultät für Mathematik
Universität Duisburg-Essen
45117 Essen
GERMANY

Dr. Quentin Berthet
Computing & Mathematical Sciences
CALTECH
Annenberg Center
1200 East California Blvd.
Pasadena CA 91125
UNITED STATES

Markus Bibinger
Fachbereich Mathematik
Humboldt Universität Berlin
Unter den Linden 6
10099 Berlin
GERMANY

Prof. Dr. Gilles Blanchard
Fachbereich Mathematik
Universität Potsdam
Am Neuen Palais 10
14469 Potsdam
GERMANY

Prof. Dr. Stephane Boucheron
UFR de Mathématiques
Université Denis Diderot Paris 7
5 rue Thomas Mann
75205 Paris Cedex 13
FRANCE

Claire Boyer
Institut de Mathématiques de Toulouse
Université Paul Sabatier
118, route de Narbonne
31062 Toulouse Cedex 9
FRANCE

Prof. Dr. Peter Bühlmann
Seminar für Statistik
ETH Zürich
HG G 17
Rämistr. 101
8092 Zürich
SWITZERLAND

Prof. Dr. Florentina Bunea
Department of Statistical Science
Cornell University
Comstock Hall
Ithaca, NY 14853-2601
UNITED STATES

Alexandra Carpentier
Department of Pure Mathematics
and Mathematical Statistics
University of Cambridge
Wilberforce Road
Cambridge CB3 0WB
UNITED KINGDOM

Prof. Dr. Ismael Castillo
Laboratoire de Probabilites et
Modeles Aleatoires
Université Paris VII
175 rue du Chevaleret
75013 Paris Cedex
FRANCE

Dr. Emilie Devijver
Departément de Mathématiques
Université de Paris-Sud
Centre d'Orsay
Bat. 430
91400 Orsay Cedex
FRANCE

Prof. Dr. John Duchi
Statistics and Electrical Engineering
Stanford University
Sequoia Hall
Stanford CA 94305
UNITED STATES

Prof. Dr. Subhashis Ghoshal
Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203
UNITED STATES

Prof. Dr. Alexander Goldenshluger
Department of Statistics
University of Haifa
Haifa 31905
ISRAEL

Prof. Dr. Friedrich Götze
Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld
GERMANY

Prof. Dr. Marc Hoffmann
CEREMADE
Université Paris Dauphine
Place du Marechal de Lattre de Tassigny
75775 Paris Cedex 16
FRANCE

Prof. Dr. Jean Jacod
IMJ
Université P. et M. Curie
4, Place Jussieu
75252 Paris Cedex 05
FRANCE

Prof. Dr. Iain M. Johnstone
Department of Statistics
Stanford University
Sequoia Hall
Stanford, CA 94305-4065
UNITED STATES

Prof. Dr. Kengo Kato
Faculty of Economics
University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-0033
JAPAN

Prof. Dr. Vladimir Koltchinskii
School of Mathematics
Georgia Institute of Technology
686 Cherry Street
Atlanta, GA 30332-0160
UNITED STATES

Prof. Dr. Rafal Latala
Institute of Mathematics
University of Warsaw
ul. Banacha 2
02-097 Warszawa
POLAND

Prof. Dr. Johannes Lederer
Department of Statistical Science
Cornell University
1188 Comstock Hall
Ithaca, NY 14853-2601
UNITED STATES

Dr. Po-Ling Loh
Department of Statistics
The Wharton School
University of Pennsylvania
3730 Walnut Street
Philadelphia, PA 19104
UNITED STATES

Karim Lounici
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332-0160
UNITED STATES

Dr. Zongming Ma
Department of Statistics
The Wharton School
University of Pennsylvania
3730 Walnut Street
Philadelphia, PA 19104-6340
UNITED STATES

Prof. Dr. Enno Mammen
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg
GERMANY

Ester Mariucci
Laboratoire Jean Kuntzmann
Université Joseph Fourier
Tour IRMA
B.P. 53
38041 Grenoble Cedex
FRANCE

Prof. Dr. Richard Nickl
Statistical Laboratory
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WB
UNITED KINGDOM

Prof. Dr. Mark Podolskij
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg
GERMANY

Prof. Dr. Alexander Rakhlin
Department of Statistics
The Wharton School
University of Pennsylvania
3730 Walnut Street
Philadelphia, PA 19104-6340
UNITED STATES

Kolyan M. Ray
Mathematisch Instituut
Universiteit Leiden
Postbus 9512
2300 RA Leiden
NETHERLANDS

Prof. Dr. Markus Reiß
Institut für Mathematik
Humboldt-Universität Berlin
Unter den Linden 6
10117 Berlin
GERMANY

Prof. Dr. Emmanuel Rio
Lab. de Mathématiques de Versailles
Batiment Fermat
Bureau 2310
45 Ave. des Etats Unis
78035 Versailles Cedex
FRANCE

Prof. Dr. Angelika Rohde
Fakultät für Mathematik
Ruhr-Universität Bochum
Lehrstuhl für Stochastik
44780 Bochum
GERMANY

Prof. Dr. Judith Rousseau
CEREMADE
Université Paris Dauphine
Place du Marechal de Lattre de Tassigny
75016 Paris Cedex
FRANCE

Dr. Johannes Schmidt-Hieber
Mathematical Institute
University of Leiden
Niels Bohrweg 1
2300 RA Leiden
NETHERLANDS

Dr. Jakob Söhl
Centre for Mathematical Sciences
University of Cambridge
Wilberforce Road
Cambridge CB3 0WB
UNITED KINGDOM

Prof. Dr. Vladimir G. Spokoiny
Weierstrass-Institute for Applied
Analysis and Stochastics
Mohrenstr. 39
10117 Berlin
GERMANY

Benjamin Stucky
Seminar für Statistik
ETH Zürich
HG G 10.2
Rämistrasse 101
8092 Zürich
SWITZERLAND

Dr. Botond T. Szabó
Korteweg-de Vries Institute for
Mathematics
University of Amsterdam
Postbus 94248
1090 GE Amsterdam
NETHERLANDS

Mathias Trabs
CEREMADE
Université Paris Dauphine
Place du Maréchal de Lattre de Tassigny
75775 Paris Cedex 16
FRANCE

Prof. Dr. Alexandre B. Tsybakov
CREST
Timbre J 340
3, Avenue P. Larousse
92240 Malakoff Cedex
FRANCE

Prof. Dr. Sara van de Geer
Seminar für Statistik
ETH Zürich; HG G 17
Rämistr. 101
8092 Zürich
SWITZERLAND

Prof. Dr. Aad W. van der Vaart
Mathematisch Instituut
Universiteit Leiden
Postbus 9512
2300 RA Leiden
NETHERLANDS

Prof. Dr. Ramon van Handel
ORFE
Princeton University
Sherrerd Hall 227
Princeton, NJ 08544
UNITED STATES

Prof. Dr. Harry van Zanten
Korteweg-de Vries Institute for
Mathematics
University of Amsterdam
Postbus 94248
1090 GE Amsterdam
NETHERLANDS

Prof. Dr. Mathias Vetter
Mathematisches Seminar
Christian-Albrechts-Universität Kiel
Ludewig-Meyn-Str. 4
24118 Kiel
GERMANY

Prof. Dr. Jon A. Wellner
Department of Statistics
University of Washington
Box 35 43 22
Seattle, WA 98195-4322
UNITED STATES

Prof. Dr. Huibin Zhou
Department of Statistics
Yale University
P.O.Box 208290
New Haven, CT 06520-8290
UNITED STATES