Barcelona GSE Summer Forum

Casa Convalescència – Sant Antoni Maria Claret, 171 - Barcelona

HIGH-DIMENSIONAL STATISTIC AND RANDOM STRUCTURES

JUNE 18-19, 2018 Room 07 (Ground Floor)

PROGRAM FOR MONDAY, JUNE 18

09:00	*Registration*
09:20	*Welcome*
Session 1	
09:30-10:15	RICHARD SAMWORTH (University of Cambridge) “Data Perturbation for High-dimensional Statistical Inference”
10:15-11:00	SOMAYEH SOJOUDI (University of California, Berkeley) “Learning Large-Scale Graphical Models: Theory, Algorithm, and Applications”
11:00	*Coffee Break**
Session 2	
11:30-12:15	MOHSEN POURAHMADI (Texas A&M University) “Aspects of Modeling Structured Correlation Matrices” (with Ruey S. Tsay)
12:15-13:00	GERGELY NEU (Universitat Pompeu Fabra) “Iterate Averaging as Regularization for Stochastic Gradient Descent” (with Lorenzo Rosasco)
13:00	*Lunch**
Session 3	
14:15-15:00	MICHAEL WOLF (University of Zurich) “Direct Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices” (with Olivier Ledoit)
15:00-15:45	MARCELO MEDEIROS (Pontifical Catholic University of Rio de Janeiro) “Machine Learning Estimates of Heterogeneous Treatment Effects in Aggregate Data and High Dimensions” (with Ricardo P. Masini)
15:45-16:30	ANDERS KOCK (University of Oxford) “Power in High-Dimensional Testing Problems” (with David Preinerstorfer)
16:30	*Coffee Break**
Session 4	
17:00-17:45	MARLOES MAATHUIS (ETH Zurich) “Learning DAGs with some Latent Variables” (with Benjamin Frot and Preetam Nandy)
20:00	*Workshop Dinner**

PROGRAM FOR TUESDAY, JUNE 19

Session 5	
09:30-10:15	ERIC GAUTIER (University of Toulouse Capitole) “Inference on Social Effects when Networks is Sparse and Unknown” (with Christiern Rose)
10:15-11:00	SOH DE WEN (Institute of High Performance Computing) “Separability and Identifiability in Gaussian Graphical Models”
11:00	*Coffee Break**
Session 6	
11:30-12:15	PRADEEP RAVIKUMAR (Carnegie Mellon University) “MPMLE: Multistaged Piecewise Estimation of Latent Gaussian Models” (with Arun Sai Suggala, Eunho Yang)
12:15-13:00	DAVID PREINERSTORFER (Université Libre de Bruxelles) “Uniformly Valid Confidence Intervals Post-model-selection” (with Francois Bachoc and Lukas Steinberger)
13:00	*Lunch**
Session 7	
14:15-15:00	RUI PIRES DA SILVA CASTRO (TU Eindhoven) “Are There Needles in a (moving) Haystack? Adaptive Sensing for Detection and Estimation of Static and Dynamically Evolving Signals” (with Ervin Tánczos)
15:00-15:45	KARL ROHE (University of Wisconsin - Madison) “A Critical Threshold in Snowball Sampling”
15:45-16:30	STEFFEN LAURITZEN (University of Copenhagen) “Random Networks, Graphical Models, and Exchangeability” (with Alessandro Rinaldo and Kayvan Sadeghi)
16:30	*Coffee Break**

Workshop Organizers:

- CHRISTIAN BROWNLEES (UPF and Barcelona GSE)
- CAROLINE UHLER (Massachusetts Institute of Technology)
- PIOTR ZWIERNIK (UPF and Barcelona GSE)

The organizers gratefully acknowledge financial support from the Spanish Ministry of Economy and Competitiveness, through the “Severo Ochoa” Programme for Centres of Excellence in R&D” (SEV-2015-0563).

* Meals are provided by the organization

ABSTRACTS (alphabetical order):

Rui Pires da Silva Castro (TU Eindhoven): Are There Needles in a (moving) Haystack? Adaptive Sensing for Detection and Estimation of Static and Dynamically Evolving Signals

Abstract: In many practical settings one can sequentially and adaptively guide the collection of future data, based on information extracted from data already collected, in what is known as sequential experimental design, active learning, or adaptive sensing/sampling (depending on the context). The intricate relation between data analysis and acquisition in adaptive sensing paradigms is extremely powerful, and allows for reliable inference in situations where non-adaptive sensing would fail dramatically. In this talk I consider estimation and detection of high-dimensional sparse signals in noise, and will focus in particular on structured signals (where the signal support has some structural properties) and dynamically evolving signals. In all these settings we characterize the difficulty of the detection and/or estimation problem, and illustrate the potential gains adaptive sensing can have over non-adaptive sensing.

Soh De Wen (Institute of High Performance Computing): Separability and Identifiability in Gaussian Graphical Models

Abstract: In learning high-dimensional graphical models, often many sparse structural assumptions are imposed on the model so that the underlying graph can be recovered with low sample complexity. In the case of p-dimensional Gaussian graphical models, assumptions of either k-degree boundedness or k-separability allows us to use techniques to learn the topology with sample complexity of only k lop p, which is efficient that k is small. However, in many real world applications, sparsity is not necessarily promised and there is no efficient method to check whether these techniques are accurate without prior sparsity assumption. In this talk, I will talk about the use of conditional independence relations to identify whether a Gaussian graphical model is sparse, both in the degree sense and in the separability sense, and extend this to learning the graph as well, without knowledge of its graph sparsity.

Eric Gautier (University of Toulouse Capitole): Inference on Social Effects when the Network is Sparse and Unknown

Abstract: This paper considers models of social interaction when the underlying networks are unobserved but sparse and there are endogenous, contextual, and correlated effects. The data consists of a time series of interactions between the N individuals and we assume that the network formation is exogenous and that the parameters are stable over time. We accommodate prior knowledge on the sparsity pattern (group sparsity, known existing or non-existing links) and restrictions on the parameters. We first present identification robust confidence sets on linear functionals of the parameters. When the model is identified we obtain estimation, model selection, and a two-stage inference on vectors of functionals.

Anders Kock (University of Oxford): Power in High-Dimensional Testing Problems

Abstract: Fan et al. (2015) recently introduced a remarkable method for increasing asymptotic power of tests in high-dimensional testing problems. If applicable to a given test, their power enhancement principle leads to an improved test that has the same asymptotic size, uniformly non-inferior asymptotic power, and is consistent against a strictly broader range of alternatives than the initially given test. We study under which conditions this method can be applied and show the following: In asymptotic regimes where the dimensionality of the parameter space is fixed as sample size increases, there often exist tests that can not be further improved with the power enhancement principle. When the dimensionality of the parameter space can increase with sample size, however, there typically is a range of “slowly” diverging rates for which every test with asymptotic size smaller than one can be improved with the power enhancement principle. While the latter statement in general does not extend to all rates at which the dimensionality increases with sample size, we give sufficient conditions under which this is the case.

Steffen Lauritzen (University of Copenhagen): Random Networks, Graphical Models, and Exchangeability

Abstract: We study conditional independence relationships for random networks and their interplay with exchangeability. We show that, for finitely exchangeable network models, the empirical subgraph densities are maximum likelihood estimates of their theoretical counterparts. We then characterize all possible Markov structures for finitely exchangeable random graphs, thereby identifying a new class of Markov network models corresponding to bidirected Kneser graphs. In particular, we demonstrate that the fundamental property of dissociatedness corresponds to a Markov property for exchangeable networks described by bidirected line graphs. Finally we study those exchangeable models that are also summarized in the sense that the probability of a network only depends on the degree distribution, and identify a class of models that is dual to the Markov graphs of Frank and Strauss (1986). Particular emphasis is placed on studying consistency properties of network models

under the process of forming subnetworks and we show that the only consistent systems of Markov properties correspond to the empty graph, the bidirected line graph of the complete graph, and the complete graph.

Marloes Maathuis (ETH Zurich): Learning DAGs with some Latent Variables

Abstract: We introduce a new method to estimate the Markov equivalence class of a directed acyclic graph (DAG) in the presence of latent variables, in settings where the underlying DAG among the observed variables is sparse, and there are a few latent variables that have a direct effect on many of the observed ones. Such settings occur frequently in biological applications, where for example technical factors or unmeasured environmental variables can affect many of the observed variables. Building on the so-called low rank plus sparse framework for latent Gaussian graphical model selection, we suggest a two-stage approach which first removes the effect of the latent variables and then estimates the Markov equivalence class of the DAG among the observed variables. This approach is consistent in certain high-dimensional regimes and performs favorably when compared to the state of the art, both in terms of graphical structure recovery and total causal effect estimation.

Marcelo Medeiros (Pontifical Catholic University of Rio de Janeiro): Machine Learning Estimates of Heterogeneous Treatment Effects in Aggregate Data and High Dimensions

Abstract: Furthermore, the intervention might not be exogenous. We propose a two-step methodology where in the first stage, a counterfactual is estimated on the basis of a large-dimensional set of variables from a pool of untreated units by means of machine learning methods such as random forests(RF), least absolute shrinkage and selection operator (LASSO) or complete subset regression (CSR). In the second stage, we estimate the average intervention effects on a vector of variables, which are consistent and asymptotically normal. As an application, we estimate the price elasticity of demand for a group of products from a large retail chain in Brazil. Our estimates are segregated by different regions and we show that there is a lot of heterogeneity in consumer behavior among different regions of the country.

Gergely Neu (Universitat Pompeu Fabra): Iterate Averaging as Regularization for Stochastic Gradient Descent

Abstract: We propose and analyze a variant of the classic Polyak-Ruppert averaging scheme, broadly used in stochastic gradient methods. Rather than a uniform average of the iterates, we consider a weighted average, with weights decaying in a geometric fashion. In the context of linear least squares regression, we show that this averaging scheme has the same regularizing effect, and indeed is asymptotically equivalent, to ridge regression. In particular, we derive finite-sample bounds for the proposed approach that match the best-known results for regularized stochastic gradient methods.

Mohsen Pourahmadi (Texas A&M University): Aspects of Modeling Structured Correlation Matrices

Abstract: There has been a flurry of activity in the last two decades in reparametrizing the Cholesky factors of correlation matrices using hyperspherical coordinates where the ensuing angles are meaningful geometrically, but hard to interpret statistically. In spite of the lack of broadly accepted statistical interpretation, we demonstrate that these angles are quite flexible and effective for parsimonious modeling of large nearly block-structured correlation matrices commonly encountered in finance, environmental and biological sciences. Asymptotic normality of the maximum likelihood estimates of these angles as new parameters is established. Real examples will be used to demonstrate the flexibility and applicability of the methodology.

David Preinerstorfer (Université libre de Bruxelles): Uniformly Valid Confidence Intervals Post-model-selection

Abstract: We suggest general methods to construct asymptotically uniformly valid confidence intervals post-model-selection. The constructions are based on principles recently proposed by Berk et al. (2013). In particular the candidate models used can be misspecified, the target of inference is model-specific, and coverage is guaranteed for any data-driven model selection procedure. After developing a general theory we apply our methods to practically important situations where the candidate set of models, from which a working model is selected, consists of fixed design homoskedastic or heteroskedastic linear models, or of binary regression models with general link functions. In an extensive simulation study, we find that the proposed confidence intervals perform remarkably well, even when compared to existing methods that are tailored only for specific model selection procedures.

Pradeep Ravikumar (Carnegie Mellon University): MPMLE: Multistaged Piecewise Estimation of Latent Gaussian Models

Abstract: We consider the estimation of Latent Gaussian Models, where the latent variables follow a multivariate Gaussian distribution. While state of the art estimators ranging over Expectation Maximization (EM), variational methods, and spectral methods, come with caveats that are statistical or computational or both, we present the class of MPMLE estimators that are able to finesse these caveats by a multi-staged piecewise estimation of carefully selected sub-problems. We not only show

that our algorithm is computationally efficient, even improving upon local approaches such as EM, but also provide strong statistical guarantees similar to that of the exact but intractable Maximum Likelihood Estimator (MLE).

Karl Rohe (University of Wisconsin, Madison): A Critical Threshold in Snowball Sampling

Abstract: In Snowball sampling and Respondent-Driven Sampling, researchers ask participants to refer their friends into the sample. This talk models snowball sampling as a Markov process on a social graph that is indexed by a Galton-Watson tree. Markov dependence decays exponentially in the number of steps (i.e. referrals). However, the Galton-Watson tree (which indexes the dependence) grows exponentially. The first part of the talk discusses the competition between these exponential rates and how they determine a critical threshold. If m is the expected number of referrals provided by each sample and \lambda_2 is the second eigenvalue of the Markov transition matrix, then the rate is determined by whether or not m < 1/\lambda_2^2. The rest of the talk will discuss ways of overcoming that dependence.
https://arxiv.org/abs/1505.05461
https://arxiv.org/abs/1606.00387
https://arxiv.org/abs/1708.04999

Richard Samworth (University of Cambridge): Data Perturbation for High-dimensional Statistical Inference

Abstract: Traditional statistical inference has tended to rely on postulating a generative model for the data under study, and then applying well-established methodologies such as maximum likelihood for inference. Big Data may be highly heterogeneous, however, and it may be infeasible to propose a realistic model that would facilitate these techniques. In such contexts, running an algorithm once on the entire data set may be unreliable. Instead, I will argue that an attractive and increasingly popular approach is to apply the algorithm to many different perturbations of the original data, and to aggregate the results appropriately. Such perturbations include subsampling, bootstrap sampling, generating random projections or knockoffs and adding artificial noise. I will illustrate this paradigm with examples from my own work.

Somayeh Sojoudi (University of California, Berkeley): Learning Large-Scale Sparse Graphical Models: Theory, Algorithm, and Applications

Abstract: Learning models from data has a significant impact on many disciplines, including computer vision, medical imaging, social networks, neuroscience and signal processing. In the network inference problem, one may model the relationships between the network components through an underlying inverse covariance matrix. The sparse inverse covariance estimation problem is commonly solved using an ℓ_1-regularized Gaussian maximum likelihood estimator, known as "graphical lasso". Despite the popularity of graphical lasso, its computational cost becomes prohibitive for large data sets. In this talk, we will develop new notions of sign-consistent matrices and inverse-consistent matrices to obtain key properties of graphical lasso and prove that although the complexity of solving graphical lasso is high, the sparsity pattern of its solution has a simple formula if a sparse graphical model is sought. We will prove — under mild assumptions— that the graphical lasso estimator can be retrieved by soft-thresholding the sample covariance matrix and solving a maximum determinant matrix completion (MDMC) problem, and describe a Newton-CG algorithm to efficiently solve the MDMC problem. Assuming that the thresholded sample covariance matrix is sparse with a sparse Cholesky factorization, we will show that the algorithm converges to an ϵ-accurate solution in $O(n\log(1/\epsilon))$ time and $O(n)$ memory. We will illustrate our results in different case studies.

Michael Wolf (University of Zurich): Direct Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices

Abstract: This paper introduces a nonlinear shrinkage estimator of the covariance matrix that does not require recovering the population eigenvalues first. We estimate the sample spectral density and its Hilbert transform directly by smoothing the sample eigen values with a variable-bandwidth kernel. Relative to numerically inverting the so-called QuEST function, the main advantages of direct kernel estimation are: (1) it is much easier to comprehend because it is analogous to kernel density estimation; (2) it is only twenty lines of code in Matlab — as opposed to thousands— which makes it more verifiable and customizable; (3) it is 200 times faster without significant loss of accuracy; and (4) it can handle matrices of a dimension larger by a factor of ten. Even for dimension 10,000, the code runs in less than two minutes on a desktop computer; this makes the power of nonlinear shrinkage as accessible to applied statisticians as the one of linear shrinkage.