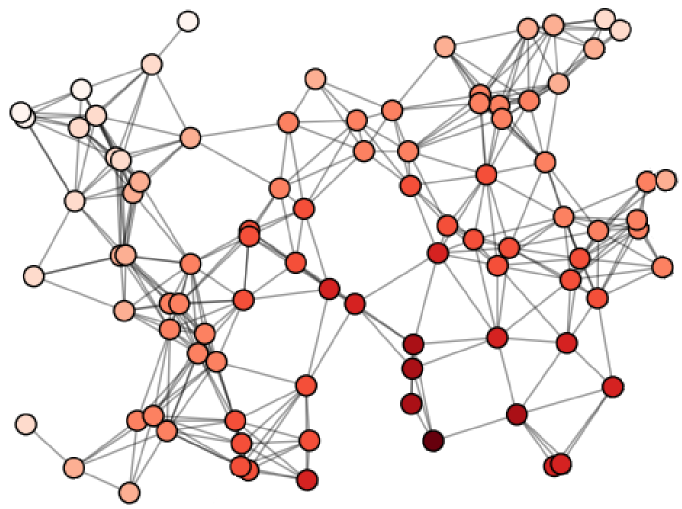

Mathematical Methods of Modern Statistics

Luminy

10th to 14th July 2017

Organizing Committee

Piotr Graczyk (Université d'Angers)
Fabien Panloup (Université d'Angers)
Frédéric Proïa (Université d'Angers)
Étienne Roquain (Université Pierre-et-Marie-Curie)
Jacek Wesołowski (Warsaw Polytechnique)

Scientific Committee

Małgorzata Bogdan (Wrocław University)
Emmanuel Candès (Stanford University)
Hélène Massam (York University)
Pascal Massart (Université Paris-Sud)
Judith Rousseau (Université Paris Dauphine)

Monday, 10th July

M. Bogdan (morning) and P. Graczyk (afternoon) as chairmen

8h50	9h00		*Opening*
9h00	9h45		**Emmanuel Candès** (Stanford University)
9h50	10h25		**Nicolas Verzelen** (INRA Montpellier)
			Coffee break
10h45	11h20		**David Dunson** (Duke University)
11h20	11h55		**Ismael Castillo** (Université Pierre-et-Marie-Curie)
11h55	12h30		**Julie Josse** (École Polytechnique)
			Lunch & break
15h00	15h35		**Hélène Massam** (York University)
15h35	16h10		**Dominique Picard** (Université Paris 7)
16h10	16h45		**Nicolai Meinshausen** (ETH Zürich)
16h45	17h20		**Surya Tokdar** (Duke University)
			Coffee break
17h35	18h10		**Steffen Lauritzen** (University of Copenhagen)
18h10	18h45		**Jacek Wesołowski** (Warsaw Polytechnique)
18h50	19h30		**Posters**
			Dinner

40' + 5' of questions   30' + 5' of questions   10' of lightning talk

Tuesday, 11th July

H. Massam (morning) and F. Panloup (afternoon) as chairmen

Start	End	Session
9h00	9h45	**Judith Rousseau** (Université Paris Dauphine)
9h50	10h25	**Philippe Biane** (Université Paris-Est)
		Coffee break
10h45	11h20	**Jonathan Taylor** (Stanford University)
11h20	11h55	**Oleg Lepski** (Université d'Aix-Marseille)
11h55	12h30	**Daniel Yekutieli** (Tel-Aviv University)
		Lunch & break
15h00	15h35	**Aad Van der Vaart** (Leiden University)
15h35	16h10	**Rafał Latała** (Warsaw University)
16h10	16h45	**Richard Samworth** (University of Cambridge)
16h45	17h15	**Agnieszka Piliszek** (Warsaw Polytechnique) **Julyan Arbel** (INRIA Grenoble Rhône-Alpes) **Guillaume Kon Kam King** (Collegio Carlo Alberto)
		Coffee break
17h35	18h10	**Hideyuki Ishi** (Nagoya University)
18h10	18h45	**Takaaki Nomura** (Kyūshū University)
18h50	19h30	**Geneviève Robin** (École Polytechnique) **Piotr Sobczyk** (Wrocław Polytechnique) **Aymeric Dieuleveut** (ENS Ulm) **Michał Lemańczyk** (Warsaw University)
		Dinner

40' + 5' of questions   30' + 5' of questions   10' of lightning talk

Wednesday, 12th July

E. Candès as chairman

9h00	9h45	**Yoav Benjamini** (Tel-Aviv University)
9h50	10h25	**Étienne Roquain** (Université Pierre-et-Marie-Curie)
		Coffee break
10h45	11h20	**Sara Van de Geer** (ETH Zürich)
11h20	11h55	**Felix Abramovich** (Tel-Aviv University)
11h55	12h30	**Małgorzata Bogdan** (Wrocław University)
		Lunch & break

40' + 5' of questions — 30' + 5' of questions — 10' of lightning talk

Thursday, 13th July

J. Rousseau (morning) and J. Wesołowski (afternoon) as chairmen

9^{h}00	9^{h}45		**Chiara Sabatti** (Stanford University)
9^{h}50	10^{h}25		**Jean-Philippe Vert** (ENS Ulm)
			Coffee break
10^{h}45	11^{h}20		**David Siegmund** (Stanford University)
11^{h}20	11^{h}55		**Ruth Heller** (Tel-Aviv University)
11^{h}55	12^{h}30		**Sylvain Arlot** (Université Paris-Sud)
			Lunch & break
15^{h}35	16^{h}10		**Gérard Letac** (Université de Toulouse)
16^{h}10	16^{h}45		**Włodzimierz Bryc** (University of Cincinnati)
16^{h}45	17^{h}20		**Clément Marteau** (Université Claude Bernard)
			Coffee break
17^{h}35	18^{h}10		**Jean-Marc Bardet** (Université Paris 1)
18^{h}10	18^{h}45		**Yann Ollivier** (Université Paris-Sud)
18^{h}50	19^{h}20		**Svetlana Gribkova** (Université Paris Diderot) **Wojciech Rejchel** (Nicolaus Copernicus University) **Thomas Bonis** (Télécom ParisTech)
			Dinner

40' + 5' of questions — 30' + 5' of questions — 10' of lightning talk

Friday, 14th July

E. Roquain as chairman

9h00	9h45	**Yannick Baraud** (Université Nice-Sophia-Antipolis)
9h50	10h25	**Lucien Birgé** (Université Pierre-et-Marie-Curie)
		Coffee break
10h45	11h20	**Mathias Drton** (University of Washington)
11h20	11h55	**Christophe Giraud** (Université Paris-Sud)
		Lunch & break

40' + 5' of questions   30' + 5' of questions   10' of lightning talk

Plenary talks

Abramovich Felix (Tel-Aviv University, Tel-Aviv, Israel)
felix@post.tau.ac.il

Title. From model selection in GLM to sparse logistic classification.
Abstract. In the first part of the talk we consider model selection in a general high-dimensional GLM by penalized maximum likelihood estimation with a complexity penalty on the model size extending the existing results for Gaussian regression. We derive non-asymptotic upper bounds for the Kullback-Leibler risk of the resulting estimators and the corresponding minimax lower bounds for sparse GLM. For the properly chosen complexity penalty the proposed estimator is shown to be asymptotically minimax and adaptive to the unknown sparsity. We discuss also several alternative model selection procedures (Lasso, Slope) computationally feasible for high-dimensional data. In the second part we apply the obtained general results for model selection in GLM to logistic regression and sparse logistic classification. We derive the misclassification excess risk of the resulting classifier and discuss its optimality among the class of sparse linear classifiers. This is a joint work with Vadim Grinshtein, The Open University of Israel.

Arlot Sylvain (Université Paris-Sud, Paris, France)
sylvain.arlot@math.u-psud.fr

Title. Analysis of some purely random forests.
Abstract. Random forests (Breiman, 2001) are a very effective and commonly used statistical method, but their full theoretical analysis is still an open problem. As a first step, simplified models such as purely random forests have been introduced, in order to shed light on the good performance of Breiman's random forests. In the regression framework, the quadratic risk of a purely random forest can be written as the sum of two terms, which can be understood as an approximation error and an estimation error. Robin Genuer (2010) studied how the estimation error decreases when the number of trees increases for some specific model. In this talk, we study the approximation error (the bias) of some purely random forest models in a regression framework, focusing in particular on the influence of the size of each tree and of the number of trees in the forest. Under some regularity assumptions on the regression function, we show that the bias of an infinite forest decreases at a faster rate (with respect to the size of each tree) than a single tree. As a consequence, infinite forests attain a strictly better risk rate (with respect to the sample size) than single trees. Joint work with Robin Genuer.

[1] Arlot S. and Genuer R. *Analysis of purely random forests bias.* arXiv : 1407.3939. 2014.

[2] Arlot S. and Genuer R. *Comments on : "A Random Forest Guided Tour" by G. Biau and E. Scornet.* arXiv : 1604.01515. 2016.

☞ **Baraud Yannick** (Université Nice-Sophia-Antipolis, Nice, France)
✉ baraud@unice.fr

Title. How to make Bayes estimators robust (common talk with Lucien Birgé).
Abstract. Bayes Estimators, as the MLE, are not robust to a possible misspecification of the model which means that a small departure of the true distribution from the model, i.e. the support of the prior, may cause a huge estimation error. It is therefore impossible to build a general theory on the performance of Bayes estimators unless one puts strong enough assumptions, usually that the KL-divergence between the true distribution and some point within the model is small enough, which implies that these two probabilities are close in total variation distance and in Hellinger distance, the converse being definitely not true since the true distribution may be extremely close to the model in total variation distance or Hellinger distance while its KL-divergence to any point of the model may be large and even infinite. This lack of robustness, which has been known for a long time, is mainly due to the fact that the posterior distribution is based on the log-likelihood, as is the MLE, and is sensitive to a possible outlier among the observations. We have developed, in previous papers, a robust alternative to the MLE by replacing, in the log-likelhood ratios, the log function by another one which is close to it in a vicinity of 1 but bounded. We wondered whether a similar replacement in the definition of the posterior distribution could lead to some robust Bayes-like estimators. The purpose of these two talks is to present an attempt to solve this problem. We shall explain the idea which underlines our construction and what performance is to be expected from this new method, in terms of concentration of the posterior and robustness to deviations (with respect to the Hellinger distance) from the model given by the support of the prior. When the model is exact and under suitable assumptions, we shall also control the Hellinger distance between the posterior distribution derived from our method and the classical Bayes one.

☞ **Bardet Jean-Marc** (Université Paris 1, Paris, France)
✉ bardet@univ-paris1.fr

Title. Statistical analysis of causal affine processes.
Abstract. We present a general class of causal time series including usual ARMA, GARCH, APARCH, ..., processes. The asymptotic behavior of quasi-maximum likehood estimators of parameters of such time series are studied. Model selection is also derived as well as multiple change detection. Simulations exhibit the strong accuracies of the estimators.

☞ **Benjamini Yoav** (Tel-Aviv University, Tel-Aviv, Israel)
✉ ybenja@post.tau.ac.il

Title. A survey of challenges in large scale selective inference.
Abstract. I shall classify current approaches to multiple inferences according to goals, and discuss the basic approaches being used. I shall then highlight a few challenges that await our attention : some are simple inequalities, others arise in particular applications.

☞ **Biane Philippe** (Université Paris-Est, Marne-la-Vallée, France)
✉ biane@univ-mlv.fr

Title. Free probability and random matrices.

Abstract. I will explain how free probability, which is a theory of independence for non-commutative random variables, can be applied to understand the spectra of various models of random matrices.

☞ **Birgé Lucien** (Université Pierre-et-Marie-Curie, Paris, France)
✉ lucien.birge@upmc.fr

Title. How to make Bayes estimators robust (common talk with Yannick Baraud).

Abstract. See Yannick Baraud above.

☞ **Bogdan Małgorzata** (Wrocław University, Wrocław, Poland)
✉ malgorzata.bogdan@pwr.edu.pl

Title. Sorted L-One Penalized Estimation (SLOPE).

Abstract. Sorted L-One Penalized Estimation (SLOPE) is an extension of LASSO, which allows for FDR control in case when regressors can be modeled as independent random variables and the vector of true regression parameters is sparse. Additionally, it has a property of assigning the same regression coefficients to strongly correlated predictors. In this talk we will present the general idea of the method and present its several extensions targeting FDR control (group SLOPE and logistic SLOPE) as well as exploiting its clustering properties (portfolio optimization). This is a joint work with E. van den Berg, D. Brzyski, E. Candes, A. Gossmann, M. Kos, P. Kremmer, S. Lee, S. Paterlini, C. Sabatti and W.Su.

☞ **Bryc Włodzimierz** (University of Cincinnati, Cincinnati, USA)
✉ brycwz@ucmail.uc.edu

Title. Cauchy-Stieltjes families with polynomial variance functions.

Abstract. Cauchy-Stieltjes families of probability measures are defined analogously to the celebrated exponential families of statistics by using the Cauchy-Stieltjes kernel $1/(1-\theta x)$ instead of the exponential kernel $\exp(\theta x)$. In both cases, one can re-parameterize the measures in the family by their mean m, and study how the variance depends on m. One of the questions of interest is which functions of m can serve as the variances. The case of quadratic functions is understood quite completely. Higher order polynomials are more difficult to analyze. In a recent joint work with Raouf Fakhfakh we study polynomial functions and “generalized orthogonality" for associated families of polynomials.

☞ **Candès Emmanuel** (Stanford University, Stanford, USA)
✉ candes@stanford.edu

Title. A new read of the knockoffs framework : new statistical tools for replicable selections.
Abstract. A common problem in modern statistical applications is to select, from a large set of candidates, a subset of variables which are important for determining an outcome of interest. For instance, the outcome may be disease status and the variables may be hundreds of thousands of single nucleotide polymorphisms on the genome. In this talk, we develop an entirely new read of the knockoffs framework of Barber and Candès (2015), which proposes a general solution to perform variable selection under rigorous type-I error control, without relying on strong modeling assumptions. We show how to apply this solution to a rich family of problems where the distribution of the covariates can be described by a hidden Markov model (HMM). In particular, we develop an exact and efficient algorithm to sample knockoff copies of an HMM, and then argue that combined with the knockoffs selective framework, they provide a natural and powerful tool for performing principled inference in genome-wide association studies with guaranteed FDR control. Finally, our methodology is applied to several datasets aimed at studying the Crohn's disease and several continuous phenotypes, e.g. levels of cholesterol. This is joint work with Yingying Fan, Lucas Janson, Jinchi Lv, Chiara Sabatti and Matteo Sesia.

☞ **Castillo Ismael** (Université Pierre-et-Marie-Curie, Paris, France)
✉ ismael.castillo@upmc.fr

Title. Uniform estimation of a class of random graph functionals.
Abstract. We consider estimation of certain functionals of a random graph generated by a stochastic block model (SBM). The number of classes is fixed or grows with the number of vertices. Minimax lower and upper bounds of estimation along specific submodels are derived. The results are nonasymptotic and imply that uniform estimation of a single connectivity parameter is much slower than the expected asymptotic pointwise rate. Specifically, the uniform quadratic rate does not scale as the number of edges, but only as the number of vertices. The lower bounds are local around any possible SBM. An analogous result is derived for functionals of a class of smooth graphons. This is joint work with Peter Orbanz (Columbia).

☞ **Drton Mathias** (University of Washington, Seattle, USA)
✉ md5@uw.edu

Title. Regularized score matching for graphical models : Non-Gaussianity and missing data.
Abstract. Regularizing HyvÃďrinen's score matching loss provides a convenient framework for estimation in high-dimensional graphical models. As it avoids the need to know normalizing constants, the framework greatly simplifies the treatment of non-Gaussian models. I will present results for specific examples of such non-Gaussian models and show how the framework readily accommodates incomplete data.

☞ **Dunson David** (Duke University, Durham, USA)
✉ `dunson@duke.edu`

Title. Bayesian manifold learning.
Abstract. It is common to collect high-dimensional (large p) data that have a potentially much lower intrinsic dimension. For example, although the ambient space is a p-dimensional Euclidean space, the data may be concentrated near a d-dimensional subspace (e.g., a manifold or collection of separated manifolds) with d much less than p. In such a case, one can potentially exploit the true low dimensional structure in the data to overcome the curse of dimensionality and obtain dramatically better performance in density estimation, regression and classification problems. Although many methods are available for manifold or subspace learning, such approaches typically : (i) fail to fully characterize uncertainty in manifold learning in statistical inferences and predictions ; and (ii) rely on locally linear approximations to the manifold which may face practical problems in high curvature settings. In this talk, we consider Bayesian nonparametric approaches to address (i)-(ii). We first show that usual Gaussian process regression can exploit manifold structure to obtain optimal (frequentist) adaptive rates. We then propose a novel "spherelets" method to more efficiently approximate manifolds, illustrating dramatic improvements over state-of-the-art geometric multiresolution analysis.

☞ **Giraud Christophe** (Université Paris-Sud, Paris, France)
✉ `christophe.giraud@polytechnique.edu`

Title. Optimal clustering with convex optimisation.
Abstract. We will explained how a relaxed and corrected version of Kmeans achieves some minimax optimal clustering rates in various settings : mixture model, variable clustering and stochastic block model. Joint work with N. Verzelen, F. Bunea, M. Royer, I. Lemhadri and Y. Emin.

☞ **Heller Ruth** (Tel-Aviv University, Tel-Aviv, Israel)
✉ `ruheller@gmail.com`

Title. Inference following selection by aggregate level hypothesis testing.
Abstract. In many genomic applications, it is common to perform tests using aggregate-level statistics within naturally defined classes for powerful identification of signals. Following aggregate-level testing, it is of interest to infer on the individual units that are within classes that contain signal. Failing to account for class selection will produce biased inference. We develop multiple testing procedures that allow rejection of individual level null hypotheses while controlling for conditional (familywise or false discovery) error rates. We use simulation studies to illustrate validity and power of the proposed procedures in comparison to several possible alternatives.

☞ **Ishi Hideyuki** (Nagoya University, Nagoya, Japan)
✉ hideyuki@math.nagoya-u.ac.jp

Title. Wishart laws for a wide class of regular convex cones.
Abstract. We introduce a new class of regular convex cones consisting of positive definite matrices with a specific block decomposition. The class contains homogeneous cones and cones associated to decomposable graphs. Ideas and Techniques in analysis on homogeneous cones such as Gindikin Gamma integrals and pseudo-inverse maps are generalized to the cones in the class. Based on the new analysis, we define Wishart laws with multi-dimensional shape parameter over the cones. Our Wishart laws cover the ones considered by Letac-Massam and Andersson-Klein as special cases. We give moment formulas and explicit solution of the Maximum Likelihood equations for the Wishart laws.

☞ **Josse Julie** (École Polytechnique, Palaiseau, France)
✉ julie.josse@polytechnique.edu

Title. Multiple imputation for continuous and categorical variables with low-rank methods.
Abstract. Multiple imputation is a powerful method to deal with missing values. In this talk, we will review recent strategies including joint modeling and fully conditional specification. We will detail approaches based on low-rank matrix estimation and highlight their theoretical and practical strengths. We apply the methods to impute missing quantitative and categorical values in a medical data base aiming at modeling decisions for severe trauma patients.

☞ **Latała Rafał** (Warsaw University, Warsaw, Poland)
✉ rlatala@mimuw.edu.pl

Title. Comparison of weak and strong moments for vectors with independent coordinates.
Abstract. We show that for $p \geq 1$, the p–th moment of suprema of linear combinations of independent centered random variables are comparable with the sum of the first moment and the weak p–th moment provided that $2q$–th and q–th integral moments of these variables are comparable for all $q \geq 2$. The latest condition turns out to be necessary in the i.i.d. case. Talk will be based on the joint work with Marta Strzelecka.

☞ **Lauritzen Steffen** (University of Copenhagen, Copenhagen, Denmark)
✉ lauritzen@math.ku.dk

Title. Total positivity and conditional independence.
Abstract. The lecture summarizes results in [1] and [2]. We analyze distributions that are multivariate totally positive of order 2 (MTP_2) and discuss various properties of such distributions. MTP_2 distributions appear in the context of positive dependence, ferromagnetism in the Ising model and various latent models. A multivariate real-valued distribution with density f w.r.t. a product

measure μ is *multivariate totally positive of order 2* (MTP_2) if the density satisfies

$$f(x)f(y) \leq f(x \wedge y)f(x \vee y).$$

Regular multivariate Gaussian distributions are MTP_2 if and only if their inverse covariance matrix (concentration matrix) K is an M-matrix, i.e. iff all off-diagonal elements are non-positive. We show that the maximum likelihood problem in the case of a multivariate Gaussian distribution is a convex optimization problem having a unique solution whenever the number of observations is at least two : the solution $\hat{K} = \hat{\Sigma}^{-1}$ is then determined by the equation system

$$\begin{aligned} \hat{k}_{uv} &\leq 0 \text{ for all } u \neq v, && (1)\\ \hat{\sigma}_{vv} - s_{vv} &= 0 \text{ for all } v \in V, && (2)\\ (\hat{\sigma}_{uv} - s_{uv}) &\geq 0 \text{ for all } u \neq v, && (3)\\ (\hat{\sigma}_{uv} - s_{uv})\hat{k}_{uv} &= 0 \text{ for all } u \neq v,, && (4) \end{aligned}$$

where $S = \{s_{uv}\}$ is the sample covariance matrix. The condition (4) ensures that the MLE $\hat{K}$ is automatically sparse. If we let $\hat{G}$ denote the graph induced by non-zero entries of $\hat{K}$, we also show that the *maximum weight spanning forest MWSF*(R) *of the correlation matrix is a subgraph of* $G(\hat{K})$. Positive associations between random variables can be described in a variety of ways. One of strongest is multivariate total positivity of order two (MTP2) introduced by Karlin and Rinott [2]. The MTP2 property is stable under a number of operations, including marginalization and conditioning. The lecture investigates how the property interacts with conditional independence structures, implying, for example, that *any strictly positive MTP2 distribution becomes faithful to its pairwise independence graph.* In addition we shall study how this property is manifested in Gaussian, discrete, and conditional Gaussian distributions, and give a number of examples of such distributions. The lecture is based on joint work with S. Fallat, K. Sadeghi, C. Uhler, N. Wermuth, and P. Zwiernik ; [1,2].

[1] Fallat S., Lauritzen S., Sadeghi K., Uhler C., Wermuth N., and Zwiernik P. *Total positivity in Markov structures.* arXiv : 1510.01290. To appear in Ann. Statist.

[2] Lauritzen S., Uhler C. and Zwiernik P. *Maximum likelihood estimation in Gaussian models under total positivity.* arXiv : 1702.04031. 2017.

Lepski Oleg (Université d'Aix-Marseille, Marseille, France)
oleg.lepski@univ-amu.fr

Title. Estimation in the convolution structure density model.

Abstract. We study the problem of nonparametric estimation under L_p–loss, $p > 1$, in the framework of the convolution structure density model on $\mathbb{R}^d$. This observation scheme is a generalization of two classical statistical models, namely density estimation under direct and indirect observations. The original pointwise selection rule from a family of "kernel-type" estimators is proposed. For the selected estimator, we prove an L_p–norm oracle inequality and several of its consequences. In particular, the problem of adaptive minimax estimation under L_p–loss over the scale of anisotropic Nikol'skii classes is addressed. We fully characterize the behavior of the minimax risk for different relationships between regularity parameters and norm indexes in the definitions of the functional

class and of the risk. We prove that proposed selection rule leads to the construction of an optimally or nearly optimally (up to logarithmic factor) adaptive estimator.

☞ **Letac Gérard** (Université de Toulouse, Toulouse, France)
✉ gerard.letac@math.univ-toulouse.fr

Title. A generalisation of the Sabot-Tarrès integral and the multivariate normal law with non positive correlations.
Abstract. Christophe Sabot and Pierre Tarrès have explicitely computed in 2015 the integral

$$ST_n(y) = \int \exp(-\langle x, y\rangle)(\det M_x)^{-1/2}dx$$

where M_x is a symmetric matrix of order n with fixed non positive off- diagonal coefficients $-w_{ij}$ and with diagonal $(2x_1, \dots, 2x_n)$. The domain of integration is the part of $\mathbb{R}^n$ for which M_x is positive definite. The result is strickingly simple although the proof is involved. We calculate more generally for $b_1 > 0, \dots b_n > 0$ the integral

$$GST_n(y) = \int \exp\left(-\langle x, y\rangle - \frac{1}{2}b^* M_x^{-1} b\right)(\det M_x)^{-1/2}dx.$$

Here again, the result is simple. We also give several presentations of the result, in particular by using the Laplacian of the undirected graph on $\{1, \dots, n\}$ defined by the set of edges $\{(i, j);\ w_{ij} > 0\}$. Finally when $B \sim N(0, M_x)$ we apply this to the study of

$$x \mapsto \Pr(B_1 > 0, \dots, B_n > 0).$$

☞ **Marteau Clément** (Université Claude Bernard, Lyon, France)
✉ marteau@math.univ-lyon1.fr

Title. Parameter recovery in two-component contamination mixtures.
Abstract. We consider a parametric density contamination model. We work with a sample of i.i.d. data with a common density $f^\star = (1 - \lambda^\star)\phi + \lambda^\star \phi(. - \mu^\star)$, where the shape ϕ is assumed to be known. We establish the optimal rates of convergence for the estimation of the mixture parameters. In particular, we prove that the classical parametric rate cannot be reached when at least one of these parameters is allowed to tend to 0 with the size of the sample.

☞ **Massam Hélène** (York University, Toronto, Canada)
✉ massamh@mathstat.yorku.ca

Title. Existence of the maximum likelihood estimate in discrete graphical models.
Abstract. Discrete graphical models form an important subclass of the class of discrete hierarchical loglinear models. Interest for the maximum likelihood estimate of the parameter in contingency

tables with multinomial or Poisson sampling started in the 1950 and 1960s. The first necessary and sufficient condition for the existence of the mle was given in the 70's. A hierarchical loglinear model can be defined by the domain of its means which is a convex polyhedron called the marginal polytope. Those results can be interpreted in terms of the marginal polytope and are equivalent to saying that the mle exists if, and only if, the data vector belongs to the interior of the polytope. We will look at recent results and methodology developed to identify the smallest face of the polytope containing the data when the mle does not exist. This is joint work with Nanwei Wang and Johannes Rauh.

☞ **Meinshausen Nicolai** (ETH Zürich, Zürich, Switzerland)
✉ `meinshausen@stat.math.ethz.ch`

Title. Causal Dantzig : fast inference in linear structural equation models with hidden variables under additive interventions.
Abstract. Causal inference is known to be very challenging when only observational data are available. Randomized experiments are often costly and impractical and in instrumental variable regression the number of instruments has to exceed the number of causal predictors. It was recently shown in Peters et al. (2016) that causal inference for the full model is possible when data from distinct observational environments are available, exploiting that the conditional distribution of a response variable is invariant under the correct causal model. Two shortcomings of such an approach are the high computational effort for large-scale data and the assumed absence of hidden confounders. Here we show that these two shortcomings can be addressed if one is willing to make a more restrictive assumption on the type of interventions that generate different environments. Thereby, we look at a different notion of invariance, namely inner-product invariance. This leads to a fast algorithm for large-scale causal inference in linear structural equation models.

☞ **Nomura Takaaki** (Kyūshū University, Fukuoka, Japan)
✉ `tnomura@math.kyushu-u.ac.jp`

Title. Homogeneous open convex cones : recent results.
Abstract. This talk presents some of the recent results on homogeneous open convex cones obtained by collaborating with Ishi, Nakashima, Yamasaki as well as of their owns. The topics include matrix realizations, basic relative invariants, various characterizations of symmetric cones among homogenous cones, interesting classes and examples of homogeneous cones, etc.

☞ **Ollivier Yann** (Université Paris-Sud, Paris, France)
✉ `cirm@yann-ollivier.org`

Title. Real-time gradient descents for learning dynamical systems.
Abstract. Learning the parameters of a dynamical system so that its trajectories have desired properties lies at the core of the uses of recurrent networks in machine learning, such as machine

translation. Yet the standard gradient descent approach, backpropagation, goes "backward in time" for a dynamical system, working from the end of the trajectory. This makes real-time training impossible or inconvenient. Time-forward algorithms exist but are computationally prohibitive. Here we present a mathematical analysis of the convergence guarantees of time-forward algorithms, together with a new, computationally cheap algorithm with the same theoretical guarantees.

Picard Dominique (Université Paris 7, Paris, France)
picard@math.univ-paris-diderot.fr

Title. Clustering high dimensional data with sparsity.
Abstract. We consider the problem of clustering high dimensional data. We observe a matrix Y of size $n \times d$. Typically d is much larger than n (but not necessarily) and each column vector represents an individual denoted by Y_i, $i \leq n$, of dimension d. For sake of simplicity, we assume that there is only two classes, i.e. that there exists $A \subset \{1, \dots, n\}$ and two vectors of $\mathbb{R}^d$, θ_- and θ_+, such that $Y_i \sim \mathcal{N}(\theta_-, \sigma^2 I)$ for $i \in A$, $Y_i \sim \mathcal{N}(\theta_+, \sigma^2 I)$ for $i \in A^c$. We assume that θ_- and θ_+ are unknown and sparse in the sense that they belong to the regularity set :

$$\Theta(s, L) := \{\theta \in \mathbb{R}^d, \sup_{K \in \mathbb{N}^*} K^{2s} \sum_{k \geq K} (\theta^k)^2 \leq L^2\}.$$

Again, for sake of simplicity we suppose that A is of the form $\{1, \dots, n\tau\}$, for $\tau \in (0, 1)$ unknown. For technical reason we also assume that in fact $\tau \in]\varepsilon, 1 - \varepsilon[$ (for some known and fixed parameter $0 < \varepsilon < 1/2$) and we put :

$$\Delta^2 = \sum_{l=1}^{d} (\theta_+ - \theta_-)^2.$$

Our problem is to determine whether or not it is efficient to smooth the data i.e. to replace the vectors $Y_i := Y_i(d)$, $i \leq n$ by, for $T < d$, $Y_i(T)$, $i \leq n$, the vectors of $\mathbb{R}^T$, of the T first coordinates of Y_i. Then, if smoothing reveals to be useful, how to choose T ideally in an adaptive way (without knowing the regularity s). In this talk, we prove that for K-means algorithm, the classification rate is improved by smoothing the data, under conditions on the separation Δ of the classes (identifiability conditions). We prove that T can be chosen in an adaptive way using a Lepski-procedure on pseudo data constructed for this purpose. We also propose an adaptive algorithm to estimate θ_+ and θ_- with minimax rates.

Roquain Étienne (Université Pierre-et-Marie-Curie, Paris, France)
etienne.roquain@upmc.fr

Title. Post hoc inference via joint family-wise error rate control.
Abstract. We introduce a general methodology for post hoc inference in a large-scale multiple testing framework. The approach is " user-agnostic " in the sense that the statistical guarantee on the number of correct rejections holds for any set of candidate items selected by the user (possibly after having seen the data). This task is investigated by defining a suitable criterion, named the

joint-family-wise-error rate (JER). We propose several procedures for controlling the JER, with a special focus on incorporating dependencies while adapting to the unknown quantity of signal. We show that our proposed setting incorporates as particular cases a version of the higher criticism as well as the closed testing based approach of Goeman and Solari (2011). Our theoretical statements are supported by numerical experiments. This is a joint work with Gilles Blanchard and Pierre Neuvial.

☞ **Rousseau Judith** (Université Paris Dauphine, Paris, France)
✉ rousseau@ceremade.dauphine.fr

Title. Bayesian nonparametric inference for multivariate Hawkes processes.
Abstract. Les processus de Hawkes forment une classe des processus ponctuels pour lesquels l'intensité s'écrit comme :

$$\lambda(t) = \int_0^{t^-} h(t-s)\,\mathrm{d}N_s + \nu$$

où N représente le processus de Hawkes, et $\nu > 0$. Les processus de Hawkes multivariés ont une intensité similaire sauf que des interractions entre les différentes composantes du processus de Hawkes sont autorisées. Les paramètres de ce modèle sont donc les fonctions d'interractions $h_{k,\ell}$, $k, \ell \leq M$ et les constantes ν_ℓ, $\ell \leq M$. Dans ce travail nous étudions une approche bayésienne nonparamétrique pour estimer les fonctions $h_{k,\ell}$ et les constantes ν_ℓ. Nous présentons un théorème général caractérisant la vitesse de concentration de la loi a posteriori dans de tels modèles. L'intérêt de cette approche est qu'elle permet la caractérisation de la convergence en norme L_1 et demande assez peu d'hypothèses sur la forme de la loi a priori. Une caractérisation de la convergence en norme L_2 est aussi considérée. Nous étudierons un exemple de lois a priori adaptées à l'étude des interractions neuronales. Travail en collaboration avec S. Donnet et V. Rivoirard.

☞ **Sabatti Chiara** (Stanford University, Stanford, USA)
✉ sabatti@stanford.edu

Title. Selective inference in genetics.
Abstract. Geneticists have always been aware that, when looking for signal across the entire genome, one has to be very careful to avoid false discoveries. Contemporary studies often involve a very large number of traits, increasing the challenges of "looking every-where". I will discuss novel approaches that allow an adaptive exploration of the data, while guaranteeing reproducible results.

☞ **Samworth Richard** (University of Cambridge, Cambridge, UK)
✉ rjs57@cam.ac.uk

Title. Efficient multivariate entropy estimation via k-nearest neighbour distances.
Abstract. Many statistical procedures, including goodness-of-fit tests and methods for independent component analysis, rely critically on the estimation of the entropy of a distribution. In this work,

we seek entropy estimators that are efficient in the sense of achieving the local asymptotic minimax lower bound. To this end, we initially study a generalisation of the estimator originally proposed by Kozachenko–Leonenko (1987), based on the k-nearest neighbour distances of a sample of n independent and identically distributed random vectors in $\mathbb{R}^d$. When $d \leq 3$ and provided $k/log^5 n \to \infty$ (as well as other regularity conditions), we show that the estimator is efficient ; on the other hand, when $d \geq 4$, a non-trivial bias precludes its efficiency regardless of the choice of k. This motivates us to consider a new entropy estimator, formed as a weighted average of Kozachenko–Leonenko estimators for different values of k. A careful choice of weights enables us to obtain an efficient estimator in arbitrary dimensions, given sufficient smoothness.

☞ **Siegmund David** (Stanford University, Stanford, USA)
✉ `siegmund@stanford.edu`

Title. Detection and estimation of local signals.

Abstract. I will discuss a general framework for detection of local signals, primarily defined by change-points, in sequences of data. Changes can occur continuously, e.g., a change in the slope of a regression line, or discontinuously, e.g., a jump in the level of a process. A motivating example of jump discontinuities is provided by copy number variation (CNV). I will focus on the simplest version of the problem : segmentation of independent normal observations according to changes in the mean. Results will be illustrated by simulations and applications. Confidence regions for the change-points and some difficulties associated with dependent observations will also be discussed. Aspects of this research involve collaboration with Fang Xiao, Li Jian, Liu Yi, Nancy Zhang, Benjamin Yakir and Li (Charlie) Xia.

☞ **Taylor Jonathan** (Stanford University, Stanford, USA)
✉ `jonathan.taylor@stanford.edu`

Title. Inferactive data analysis.

Abstract. We describe inferactive data analysis, so-named to denote an interactive approach to data analysis with an emphasis on inference after data analysis. Our approach is a compromise between Tukey's exploratory (roughly speaking "model free") and confirmatory data analysis (roughly speaking classical and "model based"), also allowing for Bayesian data analysis. We view this approach as close in spirit to current practice of applied statisticians and data scientists while allowing frequentist guarantees for results to be reported in the scientific literature, or Bayesian results where the data scientist may choose the statistical model (and hence the prior) after some initial exploratory analysis. While this approach to data analysis does not cover every scenario, and every possible algorithm data scientists may use, we see this as a useful step in concrete providing tools (with frequentist statistical guarantees) for current data scientists. The basis of inference we use is selective inference, in particular, its randomized form. The randomized framework, besides providing additional power and shorter confidence intervals, also provides explicit forms for relevant reference distributions (up to normalization) through the selective sampler. The reference distributions are constructed from a particular conditional distribution formed from what we call a DAG-DAG, a

Data Analysis Generative DAG. As sampling conditional distributions in DAGs is generally complex, the selective sampler is crucial to any practical implementation of inferactive data analysis.

☞ **Tokdar Surya** (Duke University, Durham, USA)
✉ st118@stat.duke.edu

Title. Linear quantile regression : joint estimation of quantile planes.
Abstract. In spite of the recent surge of interest in quantile regression, joint estimation of linear quantile planes remains a great challenge in statistics and econometrics. We propose a novel parametrization that characterizes any collection of non-crossing quantile planes over arbitrarily shaped convex predictor domains in any dimension by means of unconstrained scalar, vector and function valued parameters. Statistical models based on this parametrization inherit a fast computation of the likelihood function, enabling penalized likelihood or Bayesian approaches to model fitting. We introduce a complete Bayesian methodology by using Gaussian process prior distributions on the function valued parameters and develop a robust and efficient Markov chain Monte Carlo parameter estimation. The resulting method is shown to offer posterior consistency under mild tail and regularity conditions. We present several illustrative examples where the new method is compared against existing approaches and is found to offer better accuracy, coverage and model fit.

☞ **Van de Geer Sara** (ETH Zürich, Zürich, Switzerland)
✉ geer@stat.math.ethz.ch

Title. Estimating equations and sharp oracle results.
Abstract. We consider the estimation of a high-dimensional parameter $\beta \in \mathcal{B} \subset \mathbb{R}^p$ using estimating equations. Let $\beta \mapsto \hat{\mathbf{r}}_n(\beta) \in \mathbb{R}^p$ a given function depending on data and let $\| \cdot \|_{\mathcal{B}}$ be a norm in $\mathbb{R}^p$. The estimation equations we will look at are

$$\hat{\mathbf{r}}_n(\hat{\beta}) + \hat{z} = 0$$

with $\hat{z} \in \partial \|\hat{\beta}\|_{\mathcal{B}}$, the subdifferential of $\| \cdot \|_{\mathcal{B}}$ at $\hat{\beta}$. Sharp oracle inequalities are derived, which involve decomposability conditions on the norm $\| \cdot \|_{\mathcal{B}}$, margin conditions and compatibility conditions. The theory will be illustrated with some examples including sparse principal component analysis.

☞ **Van der Vaart Aad** (Leiden University, Leiden, Netherlands)
✉ avdvaart@math.leidenuniv.nl

Title. Statistical estimation of a network model.
Abstract. We consider the estimation of the attachment function in the preferential attachment model for generating a random graph. We consider both empirical estimation in the nonparametric situation and maximum likelihood estimation in parametric models. Joint work with F. Gao.

☞ **Vert Jean-Philippe** (ENS Ulm, Paris, France)
✉ jean-philippe.vert@ens.fr

Title. Learning on the symmetric group.

Abstract. Many data can be represented as rankings or permutations, raising the question of developing machine learning models on the symmetric group. When the number of items in the permutations gets large, manipulating permutations can quickly become computationally intractable. I will discuss two computationally efficient embeddings of the symmetric groups in Euclidean spaces leading to fast machine learning algorithms, and illustrate their relevance on biological applications and image classification.

☞ **Verzelen Nicolas** (INRA Montpellier, Montpellier, France)
✉ nicolas.verzelen@inra.fr

Title. On optimal graphon estimation.

Abstract. Consider the nonparametric W-random graph model. Given a realization of the graph, one aims at estimating the corresponding graphon function. In this talk, I will discuss optimal convergence rates and computational-statistical tradeoffs for graphon estimation under various metrics.

☞ **Wesołowski Jacek** (Warsaw Polytechnique, Warsaw, Poland)
✉ j.wesolowski@mini.pw.edu.pl

Title. Morality and immorality for discrete graphical models.

Abstract. Let $\underline{X} = (X_1, \ldots, X_n)$ be a random vector with discrete distribution concentrated on a set $\mathcal{I} = \mathcal{I}_1 \times \ldots \times \mathcal{I}_n$. Denote by p the distribution of $\underline{X}$, i.e. $p(\underline{i}) = \mathbb{P}(\underline{X} = \underline{i})$, $\underline{i} \in \mathcal{I}$. Let $\mathcal{G}$ be a DAG (directed acyclic graph) with vertices $V = \{1, \ldots, n\}$ and the parent function $\mathfrak{pa} : V \to 2^V$. We say that the law of $\underline{X}$ is Markov with respect to $\mathcal{G}$ iff

$$\mathbb{P}(\underline{X} = \underline{i}) = \prod_{v \in V} \mathbb{P}(X_v = i_v | \underline{X}_{\mathfrak{pa}(v)} = \underline{i}_{\mathfrak{pa}(v)}), \quad \underline{i} \in \mathcal{I}.$$

Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two DAGs with a common skeleton. We say that $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent when p is Markov with respect to $\mathcal{G}_1$ iff it is Markov with respect to $\mathcal{G}_2$. A triplet of vertices $(u; v, w)$ is called an immorality in a DAG $\mathcal{G}$ iff $v, w \in \mathfrak{pa}(u)$ and there is no edge between v and w. It is well known that two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ (with the same skeleton) are Markov equivalent iff they have identical immoralities. Such DAGs with common immoralities fall into the same Markov equivalence class which can be identified with a special chain graph called an essential graph. A chordal undirected graph G is an essential graph for the class of moral DAGs (DAGs with no immoralities) having skeleton G. The hyper-Dirichlet distribution, introduced in Dawid and Lauritzen (1993), is a popular Bayesian prior assigned to the vector of (parameters) probabilities $p = (p_{\underline{i}})_{\underline{i} \in \mathcal{I}}$ in such a decomposable graphical model. In the G is a complete graph it reduces to the classical Dirichlet law. We will describe a natural extension of the hyper-Dirichlet distribution, which we call the $\mathcal{P}$-Dirchlet, where $\mathcal{P}$ is an arbitrary family of moral DAGs with skeleton G. If $\mathcal{P}$ is the family of all such moral DAGs,

the $\mathcal{P}$-Dirichlet reduces to the hyper-Dirichlet. We will discuss useful properties of this new family of priors. Geiger and Heckerman (1997) presented a characterization of the classical Dirichlet by so called local and global independence of parameters with respect to two specially chosen DAGs with the common complete skeleton. Using the moments method we will show how this result can be generalized to cover the cases of $\mathcal{P}$-Dirichlet and hyper-Dirichlet laws. More details can be found in Massam and Wesołowski (2016). The approach we developed for moral DAGs can be used to define and analyze analogues of the hyper-Dirichlet (or $\mathcal{P}$-Dirichlet) law for any essential graph. However, in the talk we will rather concentrate on characterizations of essential graphs among chain graphs. Andersson, Madigan and Perlman (1997) gave such a characterization based on no flags and strongly protected arrows. We present a new characterization which seems to have a more systematic nature. In particular, it allows to design a new, simple and efficient alogorithm for construction of the right essential graph from a given DAG. In the process we define a notion of q-essential graphs, which seems to be of its own interest. q-essential graps (with a given skeleton) form a poset with minimal elements being the DAGs of the Markov equivalence class and the unique maximal element being the essential graph of the class. Every q-essential graph appears to be a union, in the sense of Frydenberg (1999), of a certain family of DAGs in the given Markov equivalence class. The subject is still under joint study with Helene Massam (York Univ., Toronto) and John Noble (Warsaw Univ.)

[1] Anderson S.A., Madigan D. and Perlman M.D. *A characterization of Markov equivalence classes for acyclic digraphs.* Ann. Statist. 25, 505–541, 1997.

[2] Dawid A.P. and Lauritzen S.L. *Hyper-Markov laws in the statistical analysis of decomposable graphical models.* Ann. Statist. 21, 1272–1317, 1993.

[3] Frydenberg M. *The chain graph Markov property.* Scand. J. Statist. 333–353, 1990.

[4] Geiger D. and Heckerman D. *A characterization of the Dirichlet distribution through global and local parameter independence.* Ann. Statist. 25, 1344–1369, 1997.

[5] Massam H. and Wesołowski J. *A new prior for discrete DAG models with a restricted set of directions.* Ann. Statist. 44, 1010–1037, 2016.

Yekutieli Daniel (Tel-Aviv University, Tel-Aviv, Israel)

`yekutiel@post.tau.ac.il`

Title. Confidence Intervals for the CDF from "noiy" iid samples.

Abstract. I will present non parametric Bayesian methodology for constructing confidence intervals for the CDF from a sample of iid observations buried in noise.

Lightning talks

Arbel Julyan (INRIA Grenoble Rhône-Alpes, Grenoble, France)
julyan.arbel@inria.fr

Title. Investigating predictive probabilities of Gibbs-type priors.
Abstract. Gibbs-type processes are random probability measures useful as prior distributions in Bayesian nonparametrics, typically employed as extensions to the default choices embodied by the Dirichlet process and the Pitman-Yor process. We recall the Gibbs-type processes predictive distributions and our main result is a second order approximation to the predictive weights for large sample size n. Our finding is that while at the first order approximation, all Gibbs-type processes boil down to the stable process (loss of parameters in the approximation), the second order approximation is enough to comprehend all parameters. Our approximation is thus discriminant, and in the same time it is simple and fast to compute, with an error term vanishing as $o(1/n)$. We derive connexions to recent works on stochastic differential equations arising in population models, as well as on estimation of discovery probabilities. Joint work with Stefano Favaro (Turin University).

Bonis Thomas (Télécom ParisTech, Paris, France)
thomas.bonis@telecom-paristech.fr

Title. On invariant measures of random walks on random geometric graphs.
Abstract. Random geometric graphs are at the center of many data analysis algorithms such as Spectral clustering or Diffusion maps. In this talk, I'll show how one can study the invariant measures of random walks on such graphs using Stein's method.

Dieuleveut Aymeric (ENS Ulm, Paris, France)
aymeric.dieuleveut@gmail.com

Title. Stochastic gradient descent as a Markov chain.
Abstract. In modern statistics, the large amounts of data one has to deal with enforces the use of computationally cheap algorithms : when both the dimension of the space of parameters and the number of observations are large, first order stochastic methods have gained the practitioner recognition and have become the method of choice. We analyze stochastic gradient descent with constant step sizes as a Markov chain. We explain the different behaviors depending on the fact that the optimized function is either quadratic (as in least-squares regression) or smooth and (locally) strongly convex (as in logistic regression). We summarize the various quantities that dictate the convergence speed and propose a very simple modification of the algorithm that allows to improve the convergence rate.

Gribkova Svetlana (Université Paris Diderot, Paris, France)
svetlana.gribkova@math.univ-paris-diderot.fr

Title. Dimension reduction for single cell RNA-Seq data : principal component analysis for zero inflated over-dispersed counts.

Abstract. Single cell RNA-Seq is a recent and revolutionary technics which allows to measure gene expressions in individual cells. Cellular transcriptomic heterogeneity play an important role in numerous biological processes such as malignant transformations or development of tissues. Single cell RNA-Seq data allows to study the structures of heterogeneous populations of individual cells from their transcriptomes. A necessary step for such analysis often consists in reducing the dimension of data by representing individual cells as points in a low dimensional space. Single cell RNA-Seq data are overdispersed counts with zero-inflation, meaning that the proportion of zeroes in such data is very high and may attain up to 50% of the total number of observations. These characteristics of the data make inefficient the standard technics of dimension reduction like principal component analysis. In this talk we will propose a new method of dimension reduction which is adapted to the structure of the data. We suppose that the observed gene expression matrix is a noisy version of a low rank rank matrix, with the zero-inflated negative binomial model for the noise. The low rank matrix is to be inferred from the data. It is represented as a sum of several factorized matrices. These terms correspond to sources of known variation (product of the known matrix of covariates and an unknown matrix of coefficients) and of unknown variation (product of two unknown matrices of factors and their loadings) in the data. Therefore our method allow to do principal component analysis adjusted to the presence of covariates and to the distribution of noise. The inference is made via numerical maximisation of the likelihood of the data using the gradient descent. Numerical experiences show that our method outperforms principal component analysis and is competitive with other methods taking into account the presence of zero inflation in single cell data.

[1] Risso D., Perraudeau F., Gribkova S., Dudoit S. and Vert J.P. *ZINB-WaVE : A general and flexible method for signal extraction from single-cell RNA-seq data*. Preprint on BioRxiv (http://biorxiv.org/content/early/2017/04/06/125112). 2017.

Kon Kam King Guillaume (Collegio Carlo Alberto, Turin, Italy)
guillaume.konkamking.work@gmail.com

Title. Bayesian nonparametric modelling of trended monotonic functional time series.

Abstract. Functional time series naturally appear in contexts where phenomena are measured regularly. Examples include the income distribution over time, the evolution of molecular size distribution during polymerisation, or daily demand/offer curves in an exchange market. Trends are common in these series : higher incomes might tend to increase while lower incomes stagnate or decrease, polymerisation increases molecule sizes globally, and prices commonly show rising or falling trends. The functional nature of the data raises a challenge for the inference and indeed, the likelihood can be intractable in the case of fully observed functions. We present a likelihood-free approach for functional data forecast with a trend phenomenon. We develop a bayesian nonparametric model based on a dependent process. It builds on particle system models, which originate from population genetics. This construction provides a means to flexibly specify the correlation of the dependent process. We take advantage of the expressiveness of interacting particle models to

embed a local and transient trend mechanism. To this aim, we draw inspiration from interaction potentials between physical particle systems in molecular dynamics. We perform the likelihood-free inference by means of Approximate Bayesian Computation (ABC). We discuss the elicitation of informative summary statistics for stochastic processes building on the idea of semi-automatic summaries. Coupled with a population ABC, this results in a very versatile inference method. We show the increased robustness of the trended model and comment on the generality of our approach for building functional forecast models. Joint work with Matteo Ruggiero and Antonio Canale.

☞ **Lemańczyk Michał** (Warsaw University, Warsaw, Poland)
✉ miclem1@wp.pl

Title. Concentration inequalities for Markov chains.
Abstract. We present a method allowing to obtain Berstein-like concentration inequalities for 1-dependent random variables in two particular cases : the two-block-factors and 1-dependent Markov chains. We show also how to use this method (together with so-called splitting of Markov chain) to get concentration inequalities for general Markov chains.

☞ **Piliszek Agnieszka** (Warsaw Polytechnique, Warsaw, Poland)
✉ a.piliszek@mini.pw.edu.pl

Title. Message hidden in the independence of matrix–Kummer and Wishart matrices.
Abstract. If X and Y are independent random variables and $U = Y/(1 + X)$ and $V = X(1 + U)$ are also independent, then distributions of X and Y are determined (Kummer and Gamma, respectively). This has been proved recently in [Piliszek, Wesołowski, *Kummer and gamma laws through independences on trees - another parallel with the Matsumoto-Yor property*, J. Multivar. Anal. **152** (2016)]. During my talk I will formulate an analogous theorem when X and Y are random matrices and give some insight into the crucial points of the proof.

☞ **Rejchel Wojciech** (Nicolaus Copernicus University, Toruń, Poland)
✉ wrejchel@gmail.com

Title. Penalized Monte Carlo methods in high-dimensional Ising model.
Abstract. Finding interactions between binary random variables is an important challenge in biology, genetics, physics or social network analysis. The undirected graph (V, E), where $V = 1, \ldots, p$ is a set of vertices and E is a set of edges, is often used to model this problem. We consider the high-dimensional scenario that p can be (much) larger than the sample size n. Thus, the number of possible edges (interactions) is even larger, but we assume that the model is sparse, i.e. the number of true (existing) edges is relatively small. The popular approach to this problem is applying the Ising model. However, there is an intractable norming constant in this model. In the literature [2, 3, 4] to overcome this difficulty as well as high-dimensionality one uses the pseudolikelihood method [1] with the penalty, for instance Lasso or SCAD. We propose to apply Monte Carlo methods instead of the pseudolikelihood. We prove that our procedure is model selection consistent in the

high-dimensional scenario under mild assumptions that are weaker than those in [2, 3, 4]. We also compare the methods experimentally. The talk is based on the joint work with Blazej Miasojedow (University of Warsaw).

[1] Besag J. *Spatial interaction and the statistical analysis of lattice systems.* J.R. Statist. Soc. B. 36, 192–236, 1974.

[2] Jalali A., Johnson C. and Ravikumar P. *On learning discrete graphical models using greedy methods.* Advances in Neural Information Processing Systems. 24, 1935–1943, 2011.

[3] Ravikumar P., Wainwright M.J. and Lafferty J. *High-dimensional Ising model selection using l1 regularized logistic regression.* Ann. Statist. 38, 1287–1319, 2010.

[4] Xue L., Zou H. and Cai T. *Nonconcave penalized composite conditional likelihood estimation of sparse Ising models.* Ann. Statist. 40, 1403–1429, 2012.

Robin Geneviève (École Polytechnique, Palaiseau, France)
genevieve.robin@polytechnique.edu

Title. Low rank interaction contingency tables.
Abstract. Log-linear models are popular tools to analyze contingency tables, particularly to model row and column effects as well as row-column interactions in two-way tables. In this paper, we introduce a regularized log-linear model designed for denoising and visualizing count data, which can incorporate side information such as row and column features. The estimation is performed through a convex optimization problem where we minimize a negative Poisson log-likelihood penalized by the nuclear norm of the interaction matrix. We derive an upper bound on the Frobenius estimation error, which improves previous rates for Poisson matrix recovery, and an algorithm based on the alternating direction method of multipliers to compute our estimator. To propose a complete methodology to users, we also address automatic selection of the regularization parameter. A Monte Carlo simulation reveals that our estimator is particularly well suited to estimate the rank of the interaction in low signal to noise ratio regimes. We illustrate with two data analyses that the results can be easily interpreted through biplot vizualization. The method is available as an `R` code.

Sobczyk Piotr (Wrocław Polytechnique, Wrocław, Poland)
piotr.sobczyk@pwr.edu.pl

Title. Sparse inverse covariance matrix estimation with graphical SLOPE.
Abstract. Estimation of the sparse inverse covariance (a.k.a. precision) matrix of multivariate Gaussian variables has been studied quite actively in recent years, as it provides a practical tool to understand statistical relations of variables in complex data in forms of a simple undirected graph, which often reveals meaningful interactions of genes, users, news articles, operational parts of a human driver, and so on. We propose a new procedure, we call it Graphical SLOPE, to estimate the inverse covariance matrix of jointly Gaussian random variables under a groupwise sparsity assumption, where groups are unknown and variables in the same group can be highly correlated. Sparsity on parameters is imposed by penalizing the sorted ℓ_1 (SL1)-norm. We made an efficient yet simple learning algorithm suitable for parallel processing. Our contribution includes also theoretical

analysis of estimation error and neighborhood FDR control. Joint work with Sangkyun Lee and Małgorzata Bogdan.

Posters

Baumer Gili (Technion, Haifa, Israel)
gili.baumer@gmail.com

Title. Discovering replicated findings across several studies of high dimension.

Abstract. The aim of replicability analysis is to identify the findings that replicate across independent studies that examine the same features, and quantify the strength of replication. These features can be single-nucleotide polymorphisms (SNPs) examined for associations with disease, genes examined for differential expression, etc. The importance of replicability analysis is well recognized in many fields, where the intention is to show that the result is consistent over different studies and is not unique to a specific study or setting. We introduce a powerful replicability analysis procedure that extends the partial conjunction approach by incorporating an assumed lower bound for the fraction of hypotheses that are null in all the studies. We show the performance of our method in simulations as well as on real data from independent studies of different psychiatric disorders. Joint work with Marina Bogomolov.

Furmańczyk Konrad (Warsaw University, Warsaw, Poland)
konfur@wp.pl

Title. Stepdown procedures with Lasso for high dimensional parametric model selection.

Abstract. Consider the variable selection problem in the high-dimensional linear model

$$Y = X\beta + \varepsilon, \tag{5}$$

where X is a fixed $n \times p$ design matrix, β is a true $p \times 1$ parameter vector $\beta \in \mathbb{R}^p$, where p is much larger than n ($p \gg n$) and ε_1 has a subgaussian distribution subG(σ^2) with the known constant σ. We propose a new selection procedure LSD which combines a stepdown multiple testing procedure with Lasso. In the first step we use the Lasso

$$\hat{\beta}_L = \arg\min_{\beta \in \mathbb{R}^p} \left(\frac{1}{2n} \|Y - X\beta\|_2^2 + \lambda_n \|\beta\|_1 \right), \tag{6}$$

for some parameter $\lambda_n > 0$ and we select a set

$$S := \left\{ 1 \le j \le p : \left| \hat{\beta}_{L,j} \right| > 0 \right\}. \tag{7}$$

In this step, under some assumptions, we select s variables such that, with high probability $s \leq n$. In the next step we apply a stepdown multiple testing procedure based on the OLS estimators in linear model, where p-value for a single hypothesis

$$H_i : \beta_i = 0 \text{ versus } H_i^{'} : \beta_i \neq 0, \text{ for } i \in S,$$

has the form $\pi_i = 2\left(1 - \Phi\left(|t_i|\right)\right)$, where Φ is the cumulative distribution function of the standard normal distribution and $t_i = \frac{\hat{\beta}_{OLS,i}}{se\left(\hat{\beta}_{OLS,i}\right)}$. As a result, we obtain correct selection in linear model with high probability. Our theoretical results are supported by simulation study which confirmed practical applications of our a two step LSD procedure. Joint work with Wojciech Rejchel (Nicolaus Copernicus University).

[1] Furmańczyk K. *On some stepdown procedures with application to consistent variable selection in linear regression.* Statistics. 49, 614–628, 2015.

[2] Furmańczyk K and Rejchel W. *Stepdown procedures with Lasso for high dimensional parametric model selection.* Preprint. 2017.

[3] Tibshirani R. *Regression shrinkage and selection via the lasso.* J. Roy. Statist. Soc. Ser. B. 58, 267–288, 1996.

[4] Ye F. and Zhang C.H. *Rate minimaxity of the Lasso and Dantzig selector for the q loss in r balls.* J. Mach. Learn. Res. 11, 3519–3540, 2010.

Giniewicz Andrzej (Wrocław Polytechnique, Wrocław, Poland)
andrzej.giniewicz@pwr.edu.pl

Title. Risk propagation in large-scale graphs.

Abstract. A problem of risk propagation in large-scale graphs describing resource transfer is considered. The problem is simplified to a weighted, personalized PageRank algorithm, implemented using GraphX Pregel API of Apache Spark Big-Data platform. Propagation properties are numerically evaluated for sample random graphs. Special attention is paid to possible applications in finance, logistics and epidemiology.

Kulishov Sergii (Ukrainian Medical Stomatologic Academy, Poltava, Ukraine)
kulishov@meta.ua

Title. Creative solutions as derivatives of selective multiple testing.

Abstract. Algorithm of creative solutions as derivatives of selective multiple testing : A. Initial selection of multiple testing methods. A1. Selection of independent and dependent variability ; Calculating the of mean, standard error of mean, standard deviation, 95% confidence interval for mean, median, minimum, maximum, range, quartiles ; Determination of the variabilities distribution – parametric or nonparametric by single-factor the Kolmogorov-Smirnov test ; Shapiro-Wilk W test and graphical methods : frequency distribution histograms stem & leaf plots ; scatter plots ; box & whisker plots ; normal probability plots : PP and QQ plots ; graphs with error bars (Graphs : Error Bar). A2. ANOVA (Analysis of Variance) test is used for parametric variabilities distribution. If

deviations are homogeneous by Levene test would used the method of multiple comparison groups by Tukey HSD, Scheffe, Bonferroni, and in the cases without homogeneity we must use the criteria Tamhane's T2, Games-Howell; Kruskal-Wallis test, nonparametric equivalent of the ANOVA, is used for nonparametric variabilities distribution. A3. The selection of variabilities, as criteria for making decisions, with $P = .05$ or less, and / or minimal false discovery rate, q-value (Gyorffy B, Gyorffy A, Tulassay Z :. The problem of multiple testing and its solutions for genom-wide studies. Orv Hetil, 2005;146(12) :559-563) Determination of the sensitivity and specificity of these variabilities. B. Secondary screening the variabilities for multiple test methods. B1. These numerical dependent variabilities with $P = .05$ or less, and / or minimal false discovery rate, with high sensitivity and specificity by diagnostic capabilities must use for formation of new variabilities as descendants of $2, 3, 4, \ldots, n$ numerical dependent variabilities as the derivatives of various mathematical transformations as Cantor, Sierpinski, von Koch sets, etc., anti-fractal sets; Moebius strip like aggregates, oxymoron combinations (Kulishov S.K., Iakovenko O.M. : Fractal and antifractal oxymorons, Moebius strip like transformations of biomedical data as basis for exploratory subgroup analysis. Book of abstract of International Conference on Trends and Perspective in Linear Statistical Inference; LinStat, 2014, Linkoping, Sweden, August 24-28, 2014; 2014, 58); and others mathematical transformations derivatives. C. Check the newly formed variabilities similar to step A to estimate the effectiveness of such changes. D. Comparison of multiple testing of more informative primary and secondary variabilities by accuracy, sensitivity and specificity of diagnostic possibilities. E. If it's necessary, the search of new selection principles of variabilities for multiple testing must be continued.

☞ **Masmoudi Khalil** (University of Sfax, Sfax, Tunisia)
✉ khalil.masmoudi@centraliens.net

Title. An EM algorithm for learning mixtures of Gaussian Markov random fields.
Abstract. In this talk, we introduce a new algorithm, for parameters' estimation and data clustering, suitable for mixtures of Gaussian Markov Random fields. This algorithm called Graphical Expectation Maximization (GEM) extends the classical EM algorithm by taking into account several graphs' structures and by using an original initialization technique. The performances of the proposed algorithm are compared to those of the classical EM algorithm on the ground of a simulation study.

☞ **Nakashima Hideto** (Kyūshū University, Fukuoka, Japan)
✉ h-nakashima@math.kyushu-u.ac.jp

Title. An explicit formula of the basic relative invariants of homogeneous cones.
Abstract. Open convex cones are fundamental objects in many areas such as statistics and analysis, and among them, homogeneous cones form a good class. In fact, their group actions enable us to deal with many things systematically; for example, Laplace transforms of relatively invariant functions can be computed explicitly. Homogeneous cones are described as the positivity sets of the basic relative invariants, which are irreducible relatively invariant polynomials, so that they play

an important role in the study of homogeneous cones. However, we only know so far that they are obtained as irreducible factors of some relative invariant polynomials related to an algebraic structure of the cones (Ishi, 2001). In the poster, I will give an explicit closed formula of them and present some applications.

Sid-Ali Ahmed (Université Laval, Quebec City, Canada)
ahmed.sid-ali.1@ulaval.ca

Title. System of interacting particles for random social networks with community detection.
Abstract. We formalize the problem of modeling social networks into punctual measure and Poisson point measures. We obtain a simple model that represent each member of the network at virtual state as a Dirac measure. We set the exact Monte Carlo scheme of this model and its representation as a stochastic process. By assuming that the spatial dependence of the kernels and rates used to build the model are bounded in some sense, we show that the random network will grow a.s. with bound. By assuming the compactness of the virtual space, we study the extinction and the survival properties of the network. Furthermore, we use a renormalization technique, which have the effect that the density of the network population must grow to infinity, to prove that the rescaled network converges in law towards the solution of a deterministic equation. Joint work with Khader Khadraoui.

Timetable

	Monday	Tuesday	Wednesday	Thursday	Friday
9h00	Candès	Rousseau	Benjamini	Sabatti	Baraud
9h50	Verzelen	Biane	Roquain	Vert	Birgé
10h25			☕		
10h45	Dunson	Taylor	Van de Geer	Siegmund	Drton
11h20	Castillo	Lepski	Abramovich	Heller	Giraud
11h55	Josse	Yekutieli	Bogdan	Arlot	
12h30					
15h00	Massam	Van der Vaart			
15h35	Picard	Latała		Letac	
16h10	Meinshausen	Samworth		Bryc	
16h45	Tokdar	Lightning talks		Marteau	
17h20	☕			☕	
17h35	Lauritzen	Ishi		Bardet	
18h10	Wesołowski	Nomura		Ollivier	
18h50	Posters	Lightning talks		Lightning talks	