Programme

Monday, 9th September

- 9:30am–10.15am
  - Ioannis Kontoyiannis: Bayesian entropy estimation for discrete time series
- 10:20am–11.05am
  - Ismael Castillo: Uniform estimation of a class of random graph functionals
- **Refreshment Break**
- 11.40am–12.15pm
  - Olivier Collier: Estimators of general functionals in the sparse Gaussian mean model
- **Lunch**
- 2pm–2.45pm
  - Mathew Penrose: Limit theory for entropy and other estimators based on multi-dimensional spacings
- 2.50pm–3.35pm
  - Alexandra Carpentier: Some problems of composite and minimax hypothesis testing
- **Refreshment Break**
- 4.10pm–4.55pm
  - Varun Jog: Bridging the inequality gap
- **Reception**, 5pm
- **Conference Dinner** 7.30pm, at St John's College, for invited speakers.

Tuesday, 10th September

- 9.30am–10.15am
  - Po-Ling Loh: Teaching and learning in uncertainty
- 10:20am–11.05am
  - Alon Orlitsky: TBC

- **Refreshment Break**
- 11.40am–12.15pm
  - Chao Gao: Optimal Estimation of Variance in Nonparametric Regression with Random Design
- **Lunch**
- 2pm–2.45pm
  - Tom Berrett: Efficient functional estimation and the super-oracle phenomenon
- 2.50pm–3.35pm
  - Barnabás Póczos: TBC
- **Refreshment Break**

Wednesday, 11th September

- 9:30am–10.15am
  - Judith Rousseau: TBC
- 10:20am–11.05am
  - Yury Polyanskiy: Smoothed Empirical Measures and Entropy Estimation
- **Refreshment Break**
- 11.40am–12.15pm
  - Zhiyi Zhang: Unfolding Entropy

Abstracts

Ioannis Kontoyiannis: Monday, 9.30am–10.15am
It has long been recognised that the main difficulty in obtaining reasonable estimates for the entropy rate of discrete time series, is in accurately capturing the underlying dependence structure. In fact, identifying useful temporal structure in discrete empirical data is in itself an important component of algorithms used for many tasks in statistical inference and machine learning. We will discuss a new methodological approach for Bayesian entropy estimation based on general discrete time series, which offers good results on both small and big data, and which admits strong theoretical justifications. The gist of our approach is the use of a rich class of Bayesian models for variable-memory Markov chains. We will discuss the accompanying theory, and present empirical results on both simulated and real

data from a broad range of applied areas, including neuroscience, animal communications and finance.

This is joint work with I. Papageorgiou (Cambridge).

Ismael Castillo: Monday, 10:20am–11.05am
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).

Olivier Collier: Monday, 11.40am–12.15pm
We observe a sparse mean vector through Gaussian noise and we want to estimate general functionals of this vector in the minimax sense. Adapting the method that has been successfully applied to the linear and quadratic functionals, ie the sums of the coefficients and of the squared coefficients, we obtain new results for the sums of the other powers of the coefficients, then for general additive functionals under very broad assumptions. However, our new estimators crucially depend on the knowledge of the noise level and the noise distribution, which can lead to larger rates of estimation. This is a joint work with Latitia Comminges, Mohamed Ndaoud and Alexandre Tsybakov.

Mathew Penrose: Monday, 2pm–2.45pm
Consider an empirical point process governed by a probability density function in d-space or in a sumbmanifold thereof. We discuss the limit theory (laws of large numbers and central limit theorems) for certain statistics based on nearest-neighbour distances within the sample, which is one way of defining spacings for a multivariate sample. As well as entropy estimators based on these spacings, we intend to discuss estimators of dimension and of divergence. The limit theorems are based on the theory of stabilizing functionals. Most of the results discussed are joint work with Yuliy Barishnikov and Joe Yukich.

Alexandra Carpentier: Monday, 2.50pm–3.35pm
We will consider the problem of composite-composite minimax hypothesis testing setting, which can also be seen as the problem of estimating an indicator function. The objective of this talk will be to highlight that the the difficulty of this problem is related to the shape of the null hypothesis in a complex way. We will consider the Gaussian model in (high) dimension p where the data are of the form $X = \theta + \sigma\epsilon$, where ϵ is a standard Gaussian vector with identity covariance matrix. An important hypothesis testing question consists in deciding whether θ belongs to a given subset Θ_0 of R^p (null hypothesis) or whether the l_2 distance between θ and the set Θ_0 is larger than some quantity ρ (alternative hypothesis). We will investigate how difficult, or easy, this testing problem is, namely how large ρ has to be so that the testing problem has a meaningful solution - i.e. that a non-trivial tests exists. We will see through several examples that the answer to this question depends on the shape

of Θ_0 in an interesting way. Based on joint works with Gilles Blanchard, Sylvain Delattre, Maurilio Gutzeit, Etienne Roquain, Nicolas Verzelen.

Varun Jog: Monday, 4.10pm–4.55pm
Reconstructing probability distributions from projections is a fundamental problem in many scientific applications. Geometric and information theoretic inequalities provide important mathematical tools for understanding the behavior of such projections—in particular, for characterizing extremal distributions with respect to different lower-dimensional properties of interest. This talk will consist of two parts: First, we introduce new methods to bound the size of an unseen geometric object using information derived from its lower-dimensional projections. Second, we present a new information inequality that relates the entropy of a random variable to that of its lower-dimensional marginals. Both parts highlight the advantages of working with information inequalities instead of their equivalent geometric or functional formulations. This is joint work with Jing Hao (UW-Madison), Chandra Nair (CUHK), and Venkat Anantharam (UC Berkeley).

Po-Ling Loh: Tuesday, 9.30am–10.15am
We investigate a simple model for social learning with two characters: a teacher and a student. The teacher's goal is to teach the student the state of the world Theta. However, the teacher herself is not certain about Theta and needs to simultaneously learn it and teach it. We examine several natural strategies the teacher may employ to make the student learn as fast as possible when the state of the world is 0,1 and transmissions occur through a binary channel. Our primary technical contribution is analyzing the exact learning rates for these strategies by studying the large deviation properties of the sign of a transient random walk on the integer grid. We also discuss a Gaussian variant of this problem and contrast the conclusions reached in the binary vs. Gaussian settings. This is joint work with Varun Jog.

Alon Orlitsky: Tuesday 10.20am–11.05am
TBC

Chao Gao: Tuesday, 11.40am–12.15pm
We consider the heteroscedastic nonparametric regression model with random design. We derive the minimax rate of estimating the variance function. The result extends the fixed design rate derived in Wang et al. [2008] in a non-trivial manner, as indicated by the entanglement of the smoothness parameters of both mean function and variance function. An implication is that variance estimation is easier in the random design setting. In the special case of constant variance, we show that the minimax rate is n8/(4+1)n1 for variance estimation, which further implies the same rate for quadratic functional estimation and thus unifies the minimax rate under the nonparametric regression model with those under the density model and the white noise model.

Tom Berrett: Tuesday, 2pm–2.45pm
We consider the estimation of two-sample density functionals

$$T(f,g) = \int f(x)\phi(f(x), g(x), x)dx,$$

based on independent d-dimensional random vectors $X_1, \ldots, X_m$ with density f and $Y_1, \ldots, Y_n$ with density g. The interest in such functionals arises from many applications: for instance, many divergences such as the KL divergence, total variation and Hellinger distances are of this form.

The estimators we consider can be expressed as weighted sums of preliminary estimators based on nearest neighbour distances. We provide conditions under which these estimators are efficient, in the sense of achieving the local asymptotic minimax lower bound, and under which they are asymptotically normal. As well as the significant theoretical contributions in this work, we also show how our results enable the construction of asymptotically valid confidence intervals. Our results also reveal an interesting phenomenon in which the natural 'oracle' estimator, requiring knowledge of f and g, can be outpeformed by our estimators. For some functionals of interest we show that the asymptotic limit of the ratio of the L_2 risks is strictly less than one, uniformly over suitable classes of densities. This is based on joint work with Richard Samworth and Ming Yuan.

Barnabás Póczos: Tuesday, 2.50pm–3.35pm
TBC

Judith Rousseau: Wednesday, 9.30am–10.15am
TBC

Yury Polyanskiy: Wednesday, 10.20am–11.05am
In this talk we discuss behavior of the empirical measure P_n corresponding to iid samples from a distribution P on a d-dimensional space. Let Q_n and Q denote the result of convolving P_n and P, respectively, with an isotropic standard Guassian kernel. We discuss convergence of the p-Wasserstein, KL and other distances between Q_n and Q. Curiously, for some distances (like 1-Wasserstein) we get parametric $1/\sqrt{n}$ speed of convergence regardless of dimension, whereas for some other distances (like 2-Wasserstein) the $1/\sqrt{n}$ rate can change to $\omega(1/\sqrt{n})$. We give an if and only if characterization in the class of subgaussian P for the parametric rate.

As an application, we show that differential entropy of Q_n converges to that of Q at parametric rate regardless of dimension. An estimator of differential entropy of Q, in turn, allows us to estimate the input-output mutual information in noisy neural networks.

Joint work with Ziv Goldfeld, Kristjan Greenewald and Jonathan Weed.

Zhiyi Zhang: Wednesday, 11.40am–12.15pm
This talk includes three sections. In Section 1, an estimator of entropy is described and motivated in a perspective induced from Turings formula. The convergence rates and other asymptotic distributional properties of the estimator under different classes of underlying distributions are summarized. In Section 2, several fundamental questions are considered. What does entropy tell us? What do we really want to know through entropy, that is, what is the real underpinning object of interest behind entropy? Do we really want to estimate entropy, or are we creating a difficult mathematical problem without sufficient merit? The contemplation of these questions leads to the definition of a new notion: the entropic probability distribution, or the entropic distribution in short, which could serve as

one of possibly many anchor points in the foundation of Statistics pertaining to information theory, or Information-Theoretical Statistics (for lack of a better term). Functionals that collectively characterize the entropic distribution are discussed. In Section 3, several results derived from the perspective of the entropic distribution are introduced, which may prove to be of some general interest in probability and statistics. The results discussed include 1) domains of attraction on countable alphabets, 2) diversity indices and their estimation, 3) generalization of mutual information, etc.