E196: Probabilistic models of protein structure: From theory to applications
Presenter: **Thomas Hamelryck**, University of Copenhagen, Denmark

The so-called protein folding problem is the loose designation for an amalgam of closely related, unsolved problems that include protein structure prediction, protein design and the simulation of the protein folding process. We adopt a unique probabilistic approach to modeling bio-molecular structures, based on graphical models and directional statistics. Notably, we develop a generative probabilistic model of protein structure in full atomic detail. We will give an overview of how rigorous probabilistic models of something as complicated as a protein's atomic structure can be formulated, focusing on the use of graphical models and directional statistics to model angular degrees of freedom. We will also discuss the reference ratio method, a novel statistical method of general interest that can be used to 'glue' different probabilistic models of protein structure together in a consistent way. This method also sheds new light on the potential of mean force, widely used in protein structure prediction, and which was up to now poorly understood and justified. Finally, we will describe some applications, including probabilistic protein structure prediction and the inference of protein structure from experimental data.

E094: Optimal tests for circular reflective symmetry about the median direction
Presenter: **Thomas Verdebout**, University Lille Nord de France, France
Co-authors: Christophe Ley

Optimal tests for reflective circular symmetry about a fixed median direction are proposed. The distributions against which optimality is achieved are the so-called k-sine-skewed distributions. We first show that sequences of k-sine-skewed models are locally and asymptotically normal in the vicinity of reflective symmetry. Following the Le Cam methodology, we then construct optimal (in the maximin sense) parametric tests for reflective symmetry, which we render semi-parametric by a studentization argument. These asymptotically distribution-free tests happen to be uniformly optimal, under any reference density, and are moreover of a very simple and intuitive form. They furthermore exhibit nice small sample properties, as we show through a Monte Carlo simulation study.

E524: Inference on circular orders with application to cell cycle gene expression data
Presenter: **Sandra Barragan**, University of Valladolid, Spain
Co-authors: Cristina Rueda, Miguel A. Fernandez, Shyamal D. Peddada

In recent years there has been considerable interest in drawing inferences regarding order relationships among angular parameters. In particular, in biology the interest is to understand genes participating in cell cycle across multiple species and whether they are functionally conserved. The time to peak expression, known as phase angle, of such a gene can be mapped onto a unit circle. Biologists are not only interested in estimating the phase angles but in determining the relative order of expression of various genes. The final aim is to know whether the order of peak expression among cell cycle genes is conserved evolutionarily across species. These questions are challenging due to large variability among studies and to the circular nature of the data. A methodology to find the underlying circular order in a population is presented. We also propose a solution for the problem of testing equality of circular orders among two or more populations. Unbalanced samples and differences in distributions are taken into consideration. The proposed methodology is illustrated by analyzing data sets from three species: Schizosaccharomyces Pombe, Schizosaccharomyces Cerevisiae and Humans. As a result a set of genes is presented where the circular order is conserved across these three species.

ES123 Room B33 BAYESIAN SEMI- AND NONPARAMETRIC MODELLING II	Chair: Li Ma

E785: An optimal transport based theory of hierarchical Bayesian inference
Presenter: **Long Nguyen**, University of Michigan, United States

Hierarchical Bayesian models present a powerful tool in statistics and machine learning. Statistical dependence can be easily expressed via latent variables, which may also be objects of inference. In a hierarchical model, the latent hierarchy enables the "borrowing of strength" between different data sets through shared parameters in the latent hierarchy. We will discuss an asymptotic theory for hierarchical model-based inference, taking a view that places latent variables at the center of the inference. By analyzing posterior concentration behaviors of the latent hierarchy that arise in a number of Bayesian nonparametric models, including the Dirichlet process mixture and the hierarchical Dirichlet processes, we show how to quantify in a precise sense the benefits of borrowing strength in a hierarchical model. We will also discuss the roles of transportation distances in the theory.

E790: Bayesian semi-parametric estimation of the long-memory parameter under FEXP-priors
Presenter: **Willem Kruijer**, Wageningen University, Netherlands
Co-authors: Judith Rousseau

For a Gaussian time series with long-memory behavior, we use the FEXP-model for semi-parametric estimation of the long-memory parameter d. The true spectral density f_o is assumed to have long-memory parameter d_o and a FEXP-expansion of Sobolev-regularity $\beta > 1$. We prove that when k follows a Poisson or geometric prior, or a sieve prior increasing at the rate $n^{1/(1+2\beta)}$, d converges to d_o at a suboptimal rate. When the sieve prior increases at the rate $n^{1/(2\beta)}$ however, the minimax rate is almost obtained. Our results can be seen as a Bayesian equivalent of the result which Moulines and Soulier obtained for some frequentist estimators.

E234: Confidence in nonparametric Bayesian credible sets
Presenter: **Botond Szabo**, Eindhoven University of Technology, Netherlands
Co-authors: Aad van der Vaart, Harry van Zanten

Adaptive techniques for nonparametric estimation have been widely studied and many rate-adaptive results have been provided for a variety of statistical problems. However, an adaptive estimator without any knowledge of its uncertainty is rather uninformative. In the Bayesian framework credible sets are intended to visualize the remaining uncertainty of the estimator. The coverage of the credible sets describes to what extent credible sets can be viewed as a quantification of uncertainty of the estimator from a frequentist point of view. Consider the problem of constructing Bayesian based confidence sets that are adaptive in L^2-loss over a continuous scale of Sobolev classes ($\beta \in [B, 2B]$, for some fixed $B > 0$) in the Gaussian White noise model. It is shown that both the hierarchical Bayes and marginal likelihood empirical Bayes approaches lead to credible sets with asymptotic coverage zero for certain oddly behaving functions. Then a new empirical Bayes method based on risk estimation is introduced, which provides uniform and adaptive confidence sets over the collection of Sobolev classes.

E184: Posterior consistency of nonparametric location-scale mixtures for multivariate density estimation
Presenter: **Pierpaolo de Blasi**, University of Torino and Collegio Carlo Alberto, Italy
Co-authors: Antonio Canale

Multivariate density estimation represents one of the most successful applications of Bayesian Nonparametrics. In particular, Dirichlet process mixtures of normal kernels are the golden standard for density estimation and their asymptotic properties have been studied extensively, especially in the univariate case. However, a gap between practitioners and the current theoretical literature is present. So far, posterior asymptotic results in the multivariate case are available only for location mixtures of Gaussian kernels with independent prior on the common covariance matrix, while in practice as well as from a conceptual point of view a location-scale mixture is often preferable. We address posterior consistency for such general mixture models by adapting a convergence rate result which combines the usual low-entropy, high-mass sieve approach with a suitable