Abstracts for PhD Students' presentations

The 35th Finnish Summer School on Probability Theory and Statistics, 3rd–7th June, 2013

Session I

A Bayesian non-parametric approach for mapping dynamic quantitative traits

Li Zitong
University of Helsinki

In genetics, dynamic or longitudinal quantitative traits are those containing the repeated phenotype measurements over time. The multivariate varying-coefficient linear regression model or so called the functional mapping model can be used to detect the association between the dynamic phenotype data (responses) and the biological markers (covariates). The time dependent coefficients for each marker are often modeled as a smoothing function or curve in order to introduce the smoothness into the model. Because of the recent development of modern high-thoughput genotyping and phenotyping techniques, big dimensional data with either a large number of markers or a large number of time points are often produced, which is challenging from the both statistical analysis and computation point of view. Motivated by these challengings, we propose an efficient Bayesian non-parametric functional mapping method on the basis of the above mentioned multivariate regression model. From the modeling point of view, the coefficients of a marker are represented by B-splines, and a random walk penalty prior is used in order to avoid the overfitting problem. In addition, we also assume a first order autoregressive (AR(1)) covariance in the residual error terms in order to model the time dependency of the phenotype data due to non-genetic (i.e., environmental) factors. From the computational perspective, we use a fast deterministic approximation algorithm called Variational Bayes (VB) for parameter estimation. VB does not only provide posterior mean and uncertainty estimation (i.e., standard error) for each parameter involved in the model, but also a lower bound estimate to the marginal likelihood, which can be used to guide variable selection. Based on the lower bound estimates, a pursuit matching like algorithm is then used to perform variable selection - search a best subset of markers which may highly associate with the dynamic trait among large data panels. We demonstrate the efficiency of our methods on both real and simulated data sets.

Optimal stopping of strong Markov processes

Ta Quoc Bao
Åbo Akademi

We characterize the value function and the optimal stopping time for a large class of optimal stopping problems where the underlying process to be stopped is a fairly general

Markov process. The main result is inspired by recent findings for Lévy processes obtained essentially via the Wiener-Hopf factorization. The main ingredient in our approach is the representation of the β-excessive functions as expected suprema. Some examples are presented

Session II

Paleotemperature reconstructions using a spatio-temporal multicore Bayesian model

Liisa Iivonen
University of Oulu

Nowadays climate change provokes intense discussion. In order to understand the future progress of the climate we need to know the past behavior of climate variables such the temperature. However, for example in Finland the longest instrumental temperature records cover only about the last 150 years. This lack of long observed temperature time series leads us to rely on modeling and proxy records. There are many proxy indicators, like ice cores, tree rings, pollen and chironomid fossils, which can be used for climate reconstructions. Such proxies can be used because they have a statistical relationship with the temperature. In our reconstructions, we use pollen taxa abundances as proxy data and our model assumes a unimodal Gaussian shape for how the proxy responds to the environment. Our approach to temperature reconstructions is a hierarchical Bayes model based on an earlier proposition, the so-called Bummer [2]. The novelty is that our model is a multicore model which takes into consideration the spatial and temporal dependence of temperatures along cores and between them. Dependencies are modeled in the prior distribution of past temperatures. This multivariate normal prior has a separable space-time covariance which is exponential in space and smooths in time. Temporal smoothing is controlled by a parameter which defines the overall variability of the temperatures and this parameter is also modeled as a random variable. The parameters of its prior distribution are elicited using a simulated 1150 year long annual mean temperature series for Southern Finland generated from the NCAR Climate System Model. As an example we present annual mean temperature reconstructions for four cores located in Finland, Estonia and Sweden. In addition to the core taxon abun- dances, the data used include surface sediment abundances of 104 pollen taxa and modern temperatures of 173 training lakes [1].

References

[1] Holmström, L., Ilvonen, L., Seppä, H., & Vest, S. A multi-core spatio-temporal model for Holocene temperature reconstruction. Manuscript.

[2] Vasko, K., Toivonen, H. T., & Korhola, A. (2000). A Bayesian multinomial Gaussian response model for organism-based environmental reconstruction. *Journal of Paleolimnology*, 24(3), 243-250.

Parameter estimation for discretely observed fractional Ornstein-Uhlenbeck process of the second kind

Lauri Viitasaari
Aalto University

Let W_t be a standard Brownian motion. Classical Ornstein-Uhlenbeck processes can be obtained via solution of the Langevin equation $\mathrm{d}X_t = -\theta X_t \mathrm{d}t + \mathrm{d}W_t$ or via Lamperti transform $X_t^{(\alpha)} = e^{-\theta t} W_{\alpha e^{2\theta t}}$. With a particular choice $\alpha = \frac{1}{2\theta}$ these two processes are the same in a sense that they have the same finite dimensional distributions. This is not the case however if one replaces Brownian motion W_t with fractional Brownian motion B_t^H. In particular, the process arising from Lamperti transform can be viewed as a solution to Langevin type equation $\mathrm{d}X_t = -\theta X_t \mathrm{d}t + \mathrm{d}Y_t$ with a noise Y given by

$$Y_t = \int_0^t e^{-s} \mathrm{d}B^H_{He^{\frac{t}{H}}}.$$

As a result we obtain two different fractional Ornstein-Uhlenbeck processes depending on the approach. The solution to Langevin equation $\mathrm{d}X_t = -\theta X_t \mathrm{d}t + \mathrm{d}B_t^H$ is referred to fractional Ornstein-Uhlenbeck process of the first kind and the process arising from Lamperti transform is referred to fractional Ornstein-Uhlenbeck process of the second kind.

An interesting problem in mathematical statistics is to estimate the unknown parameter θ. One approach is to consider LSE estimator based on Skorokhod integrals. This is considered by Hu and Nualart [3] for first kind process and by Azmoodeh and Morlanes [1] for second kind process. However, divergence integrals cannot be computed from the path of the process. Another approach is to observe the path of the process and estimate the unknown parameter θ directly from the observations. We consider discretely observed second kind process and find strongly consistent estimator. We also introduce an estimator based on generalised quadratic variations for Hurst parameter H. Moreover, we derive central limit theorems for our estimators. Similar results for first kind process is derived by Brouste and Iacus [2].

References

[1] Azmoodeh, E. and Morlanes, I. (2012). Drift parameter estimation for Fractional Ornstein-Uhlenbeck process of the second kind. *Manuscript.*

[2] Brouste, A. and Iacus, S. M. (2012). Parameter estimation for the discretely observed fractional Ornstein-Uhlenbeck process and the Yuima R package. *Computational Statistics.* DOI 10.1007/s00180-012-0365-6.

[3] Hu, Y. and Nualart, D. (2010). Parameter estimation for fractional Ornstein-Uhlenbeck processes. *Statist. Probab. Lett.* 80, no. 11-12, 1030-1038.

Session III

Binomial random intersection digraph

Mikko Kuronen
University of Jyväskylä

Random intersection digraph is a model for asymmetric interaction between agents. The model is flexible enough to exhibit power-law behavior for the degree distributions and clustering property. In this talk we review the model and describe some conjectures concerning binomial random intersection digraphs.

Rumor spreading on the complete graph with spreaders and listeners

Pekka Aalto
University of Jyväskylä

We consider a generalization of the standard rumor spreading process where only some of the nodes spread the rumor. This process is investigated in the complete graph when the size of the graph grows to infinity. We compute the limiting distributions for the time when the rumor reaches all of the nodes and for the time when the rumor reaches a random node. These times can be viewed also as a maximum distance from a random node to any other node and as a distance between two random nodes in a graph where the edge lengths are varying.

Session IV

On rough asymptotic behaviour of ruin probabilities in a general discrete risk model

Jaakko Lehtomaa
University of Helsinki

The objective is to study the rough asymptotic behaviour of a general economic risk model in a discrete setting. Both financial and insurance risks are taken into account. The loss in the year n is modelled as a random variable $B_1 + A_1 B_2 + \ldots + A_1 \ldots A_{n-1} B_n$, where A_i corresponds to the financial risk of the year i and B_i represents the insurance risk respectively.

The main result is that ruin probabilities exhibit power law decay under general assumptions. The aim of this work is to find the relevant quantities that describe the speed at which the ruin probability vanishes as the amount of initial capital grows. It turns out that these quantities can be expressed as maximal moments, called moment indices, of suitable random variables.