Statistical Sciences
UNIVERSITY OF TORONTO

Graduate Student Seminars

November 20, 2014 at 3:30pm
Sidney Smith Hall, Room SS 1083
Refreshments will be served at 3:15pm

Lingling Fan

On Bias Adjustments for Web Surveys

Web surveys are becoming an attractive data collection mode over the last decades, but by design they exclude the entire non-internet population. And also they do not have high response rates, thus non-coverage and non-response biases are more worrisome in web surveys. Imputation is a commonly used method to deal with item non-response, by which a complete data set can be created by filling in missing values. In this study, we will use imputation methods including hot deck imputation and recursive partitioning tree-based imputation methods to address non-coverage bias in web surveys. We present simulation results to illustrate the performance of the methods under different scenarios depending on the availability of additional information for the reference population, which look promising in some cases. Possible extensions of these approaches and directions for future work will also be discussed.,

Keywords: Web surveys, non-coverage, non-response, imputation

Tadeu Ferreira

Discriminative Gaussian-Bernoulli RBM with applications on LOBs.

Most electronic markets nowadays are quote driven, where quotes (prices and volumes that individuals are willing to buy and sell) are collected in the Limit Order Book (LOB), and are price and time prioritized. In this talk, I will present an extension and adaptation of the Restricted Boltzmann Machine (RBM) named Discriminative Gaussian-Bernoulli RBM (DGB-RBM). RBMs in particular are superior to many methods at capturing hidden structures in various kinds of data. For algorithmic trading, however, we are interested in conditional predictions for bid/ask movements. The DGB-RBM is tailor made for optimized learning on such conditional predictions and have better performance, are faster to train and easier to implement than the contrastive divergence used in RBMs. We show the efficacy of this method by applying it to some high frequency NASDAQ data.

Zhen Qin

Ambiguity Aversion in Commodity Markets

In this talk I will address the question of how agents who trade commodities exposed to jump risks account for uncertainty in their model. To achieve this goal, the agent's reference model P is driven by a Poisson random measure and a Brownian driver and she considers alternate models Q by applying a class of measure changes which preserve the polar sets. Under the alternate models, the agent considers models with stochastic intensity, alternate jump sizes and stochastic drift factors. Alternate models are penalized using a convex measure of distance between measures H(Q|P) which reduces to relative entropy under some restrictions. The agent then values contracts according to a robust form of certainty equivalent and I explore the impact that model uncertain has. A verification theorem is proven and in some cases, closed form results are attainable, while in others, perturbation methods provide approximate results. Some numerical experiments with finite difference are also carried out.

Dameng Tang

Modeling Correlated Frequencies with Application in Operational Risk Management

In this presentation, we propose a copula-free approach for modeling correlated frequency distributions using an Erlang-based multivariate mixed Poisson distribution. We investigate some of the properties possessed by this class of distributions and derive a tailor-made expectation-maximization (EM) algorithm for fitting purposes. The applicability of the proposed distribution is illustrated in an operational risk management context, where this class is used to model the operational loss frequencies and their complex dependence structure in a high dimensional setting. Furthermore, by assuming that operational loss severities follow the
mixture of Erlang distributions, our approach leads to a closed-form expression for the total aggregate loss distribution and its Value at Risk (VaR) can be calculated easily by any numerical method. The efficiency and accuracy of the proposed approach are analyzed using a
modified real operational loss dataset.

Jinyoung (Jennifer) Yang

Automatically Tuned General-Purpose MCMC via New Adaptive Diagnostics

Adaptive MCMC is an attempt to modify a Markov chain `on the fly’ so the chain can converge quicker. As many adaptive techniques involve using information from the past iterations of the Markov chain, the chain loses its Markovian property as well as the guaranteed convergence. The algorithm introduced here employs a few adaptive rules to tune a symmetric random walk Metropolis algorithm. The adaption stops once the algorithm diagnoses that further adaption will not significantly improve the convergence speed. This choice avoids the theoretical difficulties which arise in proving the convergence of chains which adapt infinitely. Once the chain stops adapting, the algorithm runs a standard non-adaptive MCMC to converge to a target distribution and estimate means of functionals.

Jialin Zou

Probability Enhanced Effective Dimension Reduction for Classifying Sparse Functional Data

One of the challenging problems in functional data analysis (FDA) is how to discriminate different curve into corresponding class. This field is under rapid development and has been thoroughly studied recently for densely observed functional data. However, the literature on classification for sparse functional data is relatively scarce. In this article, inspired from probability estimation by weight support vector machine (WSVM) for multivariate data, we proposed a two-step classification procedure called probability enhanced functional cumulative slicing (PEFCS), which integrates the technique from WSVM into functional cumulative slicing (FCS), to estimate EDR directions under sparsely observed functional data setting. As a consequence, PEFCS inherits the advantage of FCS about maximizing the utility of the data but not suering from the problem about homogeneity in slice in FCS.