# Graduate Student Seminar Day

Dept. Statistics, U. Toronto

Mar 22, 2012

SPEAKER: **Edwin Lei**, PhD candidate, year 2

TITLE: **Dimension reduction in functional regression via cumulative slicing**

ABSTRACT: *This project concerns the most effective dimension reduction (EDR) space in functional regression settings. Our investigation is motivated by finding a smaller set of derived variables that contains just as much predictive power as the original set. The main idea originated from cumulative slicing for multivariate data in Zhu et al. (2010 JASA), which assesses the EDR space by cumulatively pooling the observed data into incremental slices and performing a spectral decomposition on the space spanned by the resulting slices. Compared to sliced inverse regression, it sidesteps the need of partitioning the data into an arbitrary number of slices and thus ensures the maximum use of the data. It is desirable to adopt this idea due to the "curse of dimensionality" encountered in functional regression. Resolving this issue makes possible the estimation of the EDR space for sparsely observed noisy functional data – a challenging setting in FDA where only a few measurements are available for some or even all of the subjects. To this end, we propose a nonparametric method to overcome the sparsity issue by pooling information together from all subjects in each slice. We present some preliminary numerical results.*

SPEAKER: **Jason Ricci**, PhD candidate, year 2

TITLE: **Calibration of a Generalized Hawkes Processes with Latent Point Types to Large Datasets**

ABSTRACT: *It is well known that the classical Hawkes Process has modeling applications in many fields including biology, neuroscience, seismology, and finance. Motivated by high-frequency finance and algorithmic trading, we propose a larger class of marked point processes that may better represent the DGP for real-world, natural systems. In this class, points are classified as those that influence the underlying intensity process and those that do not, where such classification is latent. Moreover, we provide efficient quasi-maximum-likelihood calibration methods that makes calibration of parameters in large datasets possible. Finally, Sequential Monte Carlo estimators are used for real-time estimation of the state of the corresponding intensity process.*

SPEAKER: **Zeynep Baskurt**, PhD candidate, year 4

TITLE: **A composite likelihood approach using the evidential paradigm to analyze genetic association in pedigrees**

ABSTRACT: *Royall (1997) proposed the evidential paradigm, an alternative to Frequentist and Bayesian paradigms for interpreting data as evidence. The evidential paradigm uses the likelihood ratio (LR) for two simple hypotheses as an objective measure of the strength of statistical evidence. In genetic studies, LRs are commonly used to measure evidence. However, evidential association of pedigrees is not straightforward due to complex family structures. We propose to use composite likelihoods to construct LRs for evidential analysis of families. We show how to make these LRs robust from model misspecification; that they have good operational characteristics; and are consistent with competing methods.*

SPEAKER: **Alex Shestopaloff**, PhD candidate, year 4

TITLE: **Some applications of autoencoder neural networks to Bayesian inference problems**

ABSTRACT: *Autoencoder neural networks are neural networks which are used for dimensionality reduction. We consider modifcations of the standard autoencoder model, first, an autoencoder where an input is left out during training, and second, where some of the connections between the visible and hidden layers are removed. We also consider modifications of these, and the standard autoencoder model, in which the encoding is done conditional on an additional set of inputs. Applications to problems of Bayesian inference, in particular to situations where the likelihood cannot be evaluated exactly will be considered.*

SPEAKER: **Avideh Sabeti**, PhD candidate, year 4

TITLE: **Bayesian Inference for Bivariate Conditional Copula Models with Mixed Outcomes**

ABSTRACT: *Conditional copula models are flexible tools for modelling complex dependence structures in regression settings. We construct Bayesian inference for the conditional copula model adapted to regression settings in which the bivariate outcome is continuous or mixed. The dependence between the copula parameter and the covariate is modelled using cubic splines. The proposed joint Bayesian inference is carried out using adaptive Markov chain Monte Carlo sampling. The deviance information criterion (DIC) is used for selecting the copula family that best approximates the data and for choosing the calibration function. The performances of the estimation and model selection methods are investigated using simulations.*

SPEAKER: **Ramya Thinniyam**, PhD candidate, year 4

TITLE: **Sequencing Elements based on Distances**

ABSTRACT: *This project concerns the problem of estimating the temporal order in which a corpus of text documents were written using only the words in them. We propose a two step approach: 1) obtain distance measures between pairs of documents using their word characteristics; 2) estimate an optimal ordering of the documents based on their pairwise distances. This talk will focus on the latter general topic of sequencing elements using distances. We introduce a greedy sequencing algorithm based on the Least Squares Minimization idea. Inferential procedures such as parameter estimation and confidence regions using MCMC are discussed. The performance of the sequencing algorithm and estimation methods are evaluated using simulated examples.*

SPEAKER: **Chunyi Wang**, PhD candidate, year 4

TITLE: **Gaussian Process Regression with Heteroscedastic Residuals**

ABSTRACT: *Standard Gaussian Process (GP) Regression models typically assume that the residuals have the same variance across all observations. However, applications with input-dependent noise (heteroscedastic) frequently arise in practice. In this paper, we propose a GP Regression model with a latent variable which serves as an additional (unobserved) covariate for the regression. This model addresses the heteroscedasticity issue since it allows the function to have a changing partial derivative with respect to the observed covariates. Also, the conditional distribution of the response given the input is non-Gaussian after tranformation by $f$ on the unobserved covariate, therefore this model also doesn't assume Gaussian noise. We compare our model with Goldberg et al's approach,*

*which uses a main GP to model the mean response and a secondary GP to model the residuals. This model does assume Gaussian noise). Experiments on both synthetic and real world data show that both our approach and Goldberg et al's approach are effective, though ours is more efficient when the noise is non-Gaussian, and the other approach works better when the noise is Gaussian.*

SPEAKER: **Ximing Xu**, PhD candidate, year 4

TITLE: **More or less: some paradoxes concerning composite likelihood inference**

ABSTRACT: *Will a composite likelihood lead to more accurate inference when more component likelihoods are incorporated or its component likelihoods are replaced by more informative ones? In the presence of nuisance parameters, the efficiency of the maximum likelihood estimator of the parameter of interest increases when the value nuisance parameters are known. Does this property also hold for composite likelihood estimators? In this talk these questions are addressed in different scenarios with emphasis on illustrative examples.*