2014 MINGHUI YU MEMORIAL CONFERENCE

PhD Student Body
Department of Statistics
Columbia University
April 26, 2014

The organizers would like to thank the Department of Statistics and the Graduate Student Advisory Council for their generous support.

2014 MINGHUI YU MEMORIAL CONFERENCE SCHEDULE

Saturday, April 26 — Davis Auditorium, Schapiro CESPR

9:20	-	9:50	Breakfast
9:50	-	10:00	Opening remarks by Professor Richard Davis, Columbia University
Morning Session I			Chair: Prof. Jingchen Liu
10:00	-	10:25	Yunxiao Chen
10:25	-	10:50	Xiaoou Li
10:50	-	11:15	Roseline Bilina Falafala
11:15	-	11:30	Break
Morning Session II			Chair: Prof. Liam Paninski
11:30	-	11:55	Gonzalo Mena
11:55	-	12:20	Kristen L. Gore
12:20	-	1:20	Lunch

Keynote Presentation

1:30	-	2:30	Professor Peter McCullagh, University of Chicago

Afternoon Session I			Chair: Prof. Bodhisattva Sen
2:35	-	3:00	Benjamin Reddy
3:00	-	3:25	Xuan Yang
3:25	-	3:45	Break
Afternoon Session II			Chair: Prof. Peter Orbanz
3:45	-	4:10	Yuting Ma
4:10	-	4:35	Gustavo Lacerda
4:35	-	5:30	Reception (Statistics Department – 10th floor of School of Social Work at 1255 Amsterdam Avenue)

Keynote Presentation

Prof. Peter McCullagh

John D. MacArthur Distinguished Service Professor
Department of Statistics and the College
University of Chicago

Survival models and health sequences

Medical investigations focusing on patient survival often generate not only a failure time for each patient but also a sequence of measurements on patient health at annual or semi-annual check-ups while the patient remains alive. Such a sequence of random length accompanied by a survival time is called a survival process. Ordinarily robust health is associated with longer survival, so the two parts of a survival process cannot be assumed independent. This talk is concerned with a general technique—temporal realignment—for constructing statistical models for survival processes. A revival model is a regression model in the sense that it incorporates covariate and treatment effects into both the distribution of survival times and the joint distribution of health outcomes. It also allows the sequence of health outcomes to be used clinically for predicting the subsequent trajectory, including the residual survival time. Details available at http://www.stat.uchicago.edu/~pmcc/reports/revival.pdf.

Abstracts

Roseline Bilina Falafala

Mathematical modeling of insider trading

Insider trading is the study of market agents having asymmetrical information and investing in the same financial market. It deals with two agents: a regular trader and an informed trader who possesses additional information. The modeling technique used for insider trading is the theory of the expansion of filtrations. There are two ways to expand a filtration: an initial expansion and a progressive enlargement. In the initial expansion framework, the underlying filtration is enlarged by the information about some random variable. In the context of insider trading, initial expansion models situations where the insider gets some information at the beginning of the trading interval. Progressive expansion models situations where the extra information the insider trader has comes from a continuous flow of knowledge. Having additional information, it would seem that insiders are always able to have free lunches with vanishing risk but we find that under certain conditions, the insider does not necessarily have arbitrage or free lunches. We consider continuous time financial models, a finite trading horizon and model insider trading by initially expanding the filtration of the regular trader at a random time. We think this is a better representation of reality than what a progressive expansion model implies. In a progressive enlargement model, the insider knows a random time which is related to the path of the risky asset. Such knowledge by an insider is extremely unlikely. The kinds of additional knowledge we consider are exogenous or endogenous information that might affect the risky asset's price process but do not explicitly depend on its path properties.

Yunxiao Chen

A regularized approach for the estimation of Q-matrix

Diagnostic classification models have recently gained prominence in educational assessment, psychiatric evaluation, and many other disciplines. Central to the model specification is the so-called Q-matrix that provides a qualitative specification of the item-attribute relationship. We formulate the Q-matrix estimation as a latent variable selection problem and then construct a computationally feasible estimator via the regularized maximum likelihood. The estimator can be applied to basically all diagnostic classification models in use.

Kristen L. Gore

Biased Sampling in Genetic Epidemiology

Understanding penetrance, the probability that a carrier of a particular genetic mutation will exhibit the phenotype of interest, is an important facet of genetic epidemiology. Penetrance estimates are typically calculated via maximum likelihood methods. However, penetrance estimates can be biased and inefficient if all factors pertaining to the familial ascertainment are not taken into account in the likelihood expression. Conditioning on the ascertainment scheme is difficult in practice. Lack of knowledge of the complete pedigree structure also makes expressing the likelihood intractable. Unless the underlying pedigree structure is known for each family brought into a penetrance study, the problem of constructing a likelihood for the observed genotypic and phenotypic information becomes intractable. This talk details how one may designate a set of potential underlying pedigree structures and incorporate knowledge of the proband class, the familial sampling frame, and other factors which can determine what portion of the complete data for each observed family in order to generate efficient and unbiased penetrance estimates.

Gustavo Lacerda

Beta Process Subspace Analysis

Beta Process Factor Analysis (BPFA) is a flexible model for exploratory factor analysis that models sparse weights and an unbounded number of factors via the Indian Buffet Process. We extend BPFA to discover groups of factors that are active together, which leads to a model where instead of factors, we have subspaces that can be on or off.

Xiaoou Li

Level crossing problem and hypothesis testing

TBD

Yuting Ma

Adaptive Sparse Distance Metric Learning

By treating multivariate observations as points in a high-dimensional space, distance measures have been used as natural measures of (dis)similarity and served as foundation of various learning methods. The efficiency of these learning methods heavily depends on the chosen distance measure. Much effort had been contributed to improving the performance of classifiers by learning an appropriate distance metric, particularly for the k-Nearest Neighbor classifier. With dimension of the data increasing, however, traditional metric learning methods face the challenge that many input variables bring in noises that mask the true signal hidden in a low-dimensional subspace, as well as resulting in a formidable computational cost. In this talk, we address these issues by adaptively learning a sparse distance metric in high-dimensional space, with simultaneous feature selection. More specifically, we construct a basis classifier based on a Mahalanobis-type distance metric which unifies the ideas of nearest neighbor and large margin classification. Using this basis learner as a building block, a gradient boosting algorithm is adopted to learn one sparse rank-one matrix at each step. The sparsity is controlled by both a stepwise feature selection mechanism and a total complexity penalty. Moreover, we further extend our method to nonlinear metric learning via a hierarchical expansion with interactions. Close connections to kernel methods is drawn via Representer theorem and Taylor expansion, which illustrates the rudiments of our approach. Simulations and experiments with several real datasets show that our approach compares favorably with the state-of-the-art methods in the current metric learning literature.

Gonzalo Mena

On Quadrature Methods for Refractory Point Process Likelihoods

Parametric models of the conditional intensity of a point process (e.g., generalized linear models) are popular in statistical neuroscience, as they allow us to characterize the variability in neural responses in terms of stimuli and spiking history. Parameter estimation in these models relies heavily on accurate evaluations of the log-likelihood and its derivatives. Classical approaches use a discretized time version of the spiking process, and recent work has exploited the existence of a refractory period (during which the conditional intensity is zero following a spike) to obtain more accurate estimates of the likelihood. In this work we demonstrate that this method can be improved significantly by applying classical quadrature methods directly to the resulting continuous time integral.

Benjamin Reddy

Approximate MLE on random walk growth models of sparse networks

Statistical inference in network models is impaired by the fact that edges in a random network are in general highly dependent random variables. One solution increasingly popular in the current statistics literature is to model the network by a "graphon," conditional on which edges completely decouple. In my talk, I will explain in what sense such conditional independence contradicts network structure. Addressing this problem naturally leads to models of random graphs traced out by random walks on the vertex set, and I will present a class of very simple random walk models in which maximum likelihood estimation is tractable.

Xuan Yang

Composite likelihood inference in a stochastic volatility model in space

The stochastic heteroscedastic process (SHP) generalizes the stochastic volatility time series model which incorporate a volatility component to capture the heterogeneity of variation at different observation sites. Since the full likelihood based inference for the SHP model involves high computational cost, we consider a composite likelihood method for its parameter estimation, where the composite likelihood is formed by a weighted sum of the pairwise likelihood functions. In addition, we give a treatment for the often seen irregularity in observation sites which the sites are assumed to be realization of a spatial point process. Under weak dependence conditions, we provide sufficient conditions for maximum composite likelihood estimator to be consistent and asymptotically normal.

This research project is conducted under supervision of Prof. Richard A. Davis and Prof. Jingchen Liu.

About Minghui Yu

Minghui was born in Shandong, China in 1983. In 2002, he entered the Special Class for the Gifted Young at the University of Science and Technology of China (USTC), one of the most prestigious universities in China. Minghui possessed the rare quality of being not only smart, but also diligent, versatile, modest and easy-going. He was the type of friend who would stand by you no matter the situation. Minghui breezed through the challenging undergraduate program at USTC, ranking at the top of his class. Minghui was well liked by his fellows students and served as the class president from his sophomore year. Although under enormous academic pressure, he still found time to organize a series of student activities, such as hiking, art performances, and athletic contests for his fellow students.

After graduating summa cum laude in 2006 from USTC, Minghui entered the PhD program of the Physics Department at Columbia University. One year later, he transferred to the doctorate program in statistics. During his time at Columbia, Minghui served as the public relations head of the Columbia University's Chinese Students and Scholars Association (2007-2008), and was a member of the Columbia Chinese Basketball Association and the Columbia Graduate Student Consulting Club. His biography on the CUCSSA website mentioned his love of movies, photography and delicacies. Minghui described himself in his blog as a boy who wants to combine art and science together.

On April 4, 2008, after attending a student-organized conference, Minghui escorted his girlfriend home on the west side of campus. On his return, he was accosted by juveniles as he was crossing 122nd and Broadway and in his attempt to flee, he was struck by an automobile on Broadway. Minghui was taken to St. Luke's Hospital where he passed away a short time later.