2012 MINGHUI YU MEMORIAL CONFERENCE

Thank you to the Department of Statistics and the Graduate Student Advisory Council for their generous support.

SCHEDULE

Saturday, April 14 — Faculty House, Seminar Rooms

9:30	-	10:00	Breakfast
Morning Session I			Chair: Richard Davis
10:00	-	10:25	Pengfei Zang
10:25	-	10:50	Heng Liu
10:50	-	11:15	Yong Bum Cho
11:15	-	11:30	Break
Morning Session II			Chair: Yang Feng
11:30	-	11:55	Stephanie Zhang
11:55	-	12:20	Gongjun Xu
12:20	-	1:20	Lunch

Keynote Presentation

1:20	-	2:20	Professor Stephen E. Fienberg, CMU

Afternoon Session I			Chair: Jingchen Liu
2:30	-	2:55	Xuan Yang
2:55	-	3:20	Junyi Zhang
3:20	-	3:55	Yi Yu
3:55	-	4:10	Break
Afternoon Session II			Chair: David Madigan
4:10	-	4:35	Subhankar Sadhukhan
4:35	-	5:00	Radka Pickova

Keynote Presentation

Professor Stephen E. Fienberg, CMU

Department of Statistics, Heinz College, Machine Learning Department, and Cylab
Carnegie Mellon University

The Challenge of Privacy Protection for Statistical Network Data

Statisticians and computer scientists have yet to deal with privacy protection for large-scale sparse statistical databases in an adequate and systematic fashion. I will review some of the traditional approaches to disclosure limitation used for more standard rectangular n by p data arrays and discuss them from the perspective of usability (freedom from systematic distortions), transparency (the provision of information bias and variability), and duality (balancing the risk-utility trade-off). Then I will explain why extensions of these approaches to the domain of network data pose even greater challenges and I review progress on the topic to date.

Abstracts

Yong Bum Cho

Asymptotic properties of the spatial empirical extremogram

The extremogram, developed by Davis and Mikosch (2009), is a tool to measure extremal dependence in a time series. Together with its empirical counterpart, the extremogram can be used for estimating extremal dependence and for checking model adequacy as it reflects the dependence in the extremes. We extend the definition of the extremogram to the spatial domain with a view towards applications in the environmental and biological sciences.

Under suitable mixing and anti-clustering conditions, we establish a central limit theorem for the spatial empirical extremogram. The proof is by the well-known big-block small block argument. Since the asymptotic variance of the sample extremogram is intractable, we consider the circular block bootstrap for approximating the sampling distribution of the empirical extremogram. We show that the bootstrap gives asymptotically correct results. The methodology is illustrated with simulated and real spatial data sets. (This is joint work with Richard Davis and Souvik Ghosh.)

Heng Liu

Theory and inference for a class of nonlinear models with application to time series of counts

Richard Davis, Heng Liu

We study theory and inference related to a class of time series models that incorporates nonlinear dynamics. It is assumed that the observations follow a one-parameter exponential family of distributions given an accompanying process that evolves as a function of lagged observations. We employ an iterated random function approach and a special coupling technique to show that, under suitable conditions on the parameter space, the conditional mean process is a geometric moment contracting Markov chain and that the observation process is absolutely regular with geometrically decaying coefficients. Moreover the asymptotic theory of the maximum likelihood estimates of the parameters is established under some mild assumptions. These models are applied to two examples; the first is the number of transactions per minute of Ericsson stock and the second is related to return times of extreme events of Goldman Sachs Group stock.

Radka Pickova

Generalizations of the Volatility-Stabilized Markets

We consider models which generalize the Volatility-stabilized markets introduced in Fernholz and Karatzas (2005). We show how to construct a weak solution of the system of stochastic differential equations under consideration, express the solution in terms of time changed squared-Bessel processes, and argue that this solution is unique in the sense of the probability law. We also discuss strong relative arbitrage opportunities in these models.

Subhankar Sadhukhan

On optimal arbitrage under constraints

We work with a financial market model which, unlike the prevalently discussed models, allows existence of arbitrage opportunities, i.e. the possibilities of making money from nothing by investing in stock markets. We discuss optimal ways of doing so. However, if restrictions are imposed on the investments that can be made, these opportunities might disappear. We study that frontier of restrictions, over which arbitrage opportunities disappear from the market.

Gongjun Xu

Efficient Simulation for Exponential Integrals of Gaussian Random Fields

The integrals of exponential functions of Gaussian random fields play an important role in both applied probability and statistics. In this talk, we focus on the design and analysis of Monte Carlo estimators for computing tail probabilities of such integral random variables. Change of measure based techniques are employed and efficient Monte Carlo estimators are constructed.

Xuan Yang

Bootstrap Variance Estimator for Weighted Samples Quantiles

Quantile estimation is of interest in many areas. When it comes to extreme quantile, the classical sample quantile estimator suffered a problem of large relative variance. Importance sampling could be applied to generate more weighted samples from the neighborhood of the quantile of interest which in turn could help to reduce the estimation error. To evaluate the variance of the weighted sample quantile, we propose a bootstrap variance estimator. We have proved its consistency and derived its convergence rate when sample size goes to infinity. Simulation results are discussed.

Yi Yu

APPLE: Approximate Path for Penalized Likelihood Estimators

In high-dimensional data analysis, penalized likelihood estimators are shown to provide good results in both variable selection and parameter estimation. We propose a new algorithm, APPLE, for calculating the Approximate Path for Penalized Likelihood Estimators. Both the convex penalty (such as LASSO) and the nonconvex penalty (such as MCP) are considered. The APPLE efficiently computes the solution path for the penalized likelihood estimator using a hybrid of the modified predictor-corrector method and the coordinate-descent algorithm. In addition, we define a new rule, modified cross-validation, for selecting the tuning parameter for the penalized likelihood estimation. We compare APPLE with several well-known packages via simulation and two real data studies.

Pengfei Zang

Sparse Vector Autoregressive Modeling

The vector autoregressive (VAR) model has been widely used for modeling temporal dependence in a multivariate time series. For large (and even moderate) dimensions, the number of AR coefficients can be prohibitively large resulting in noisy estimates and difficult-to-interpret temporal dependence. As a remedy, we propose a methodology for fitting sparse VAR models (sVAR) in which most of the AR coefficients are set equal to zero. The first stage in selecting the nonzero coefficients is based on an estimate of the partial spectral coherency (PSC) together with the use of BIC. The PSC is useful for quantifying the conditional relationships between marginal series in a multivariate time series. A refinement stage is then applied to further reduce the number of parameters. The performance of this 2-stage approach is illustrated with both simulated data and several real examples. This is a joint work with Richard Davis and Tian Zheng.

Junyi Zhang

A New Method for Hypothesis Testing in Case of Multicollinearity
- With Application to Evaluating the Model of Resting Energy Expenditure

One of the primary goals of human energy metabolism research is to explore the specific metabolic rate (Ki values) for individual organs/tissues under resting conditions. The specific resting metabolic rates of major organs/tissues in adults were suggested by Elia (1992). A new testing approach is developed to evaluate Elias Ki values on the basis of a mechanistic model that the whole body resting energy expenditure (REE) is the totality of Ki * Ti, where Ti is the mass of individual organs/tissues. Since there is multicollinearity among organ/tissues masses, ridge regression (estimation) or simultaneous confidence interval (testing) or multiple testing procedures (testing) are among the preferred alternatives to a vanilla regression method. In this talk, I will compare the new approach with all alternative methods mentioned above. In addition, the confidence intervals induced by this new method will be discussed.

Stephanie Zhang

Computerized Adaptive Testing for Cognitive Diagnosis

Jingchen Liu, Zhiliang Ying, and Stephanie Zhang

Computerized adaptive testing (CAT) is a sequential experiment design scheme that tailors the selection of experiments to each subject. Such a scheme measures subjects attributes (unknown parameters) more accurately than regular prefixed designs. In this paper, we con- sider CAT for diagnostic classification models, for which the attribute estimation corresponds to a classification problem. We propose to use the asymptotic decaying rate of

the misclassification probability as a criterion for item selection. Based on this criterion, we propose corresponding CAT procedures.

About Minghui Yu

Minghui was born in Shandong, China in 1983. In 2002, he entered the Special Class for the Young and Gifted at the University of Science and Technology of China (USTC), one of the most prestigious universities in China. Minghui was a special person who was not only smart, but also diligent, versatile, modest, easy-going, and 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 with hardly breaking a sweat. Minghui was well liked by his fellows students having 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 and earning a Guo Moruo Scholarship in 2006 from USTC, Minghui entered the PhD program in the Physics Department at Columbia University. After one year, he transferred to the doctorate program in statistics. During his short time at Columbia, Minghui served as the public relations head of Columbia University's Chinese Students and Scholars Association (2007-2008), and was a member of both the Columbia Chinese Basketball Association and the Columbia Graduate Student Consulting Club. His biography on the CUCSSA Web site mentioned his love of movies, photography and delicacies. Minghui described himself in his blog as a boy who wants to combine arts and science together. He was also a huge fan of the Westlife band.

After attending a student organized conference on April 4, 2008 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.

While Minghui was only part of the Statistics Department for a short time, he left an indelible mark that will not be forgotten. May Minghui rest in peace.