Semester: Winter 2006
Location: Flarsheim Hall, Room 304
Day & Time: Thursday, 1:00-1:45 pm

Organizer: Yong Zeng, 235-5850
Email: zengy@umkc.edu

Dates, Titles, Speakers (with Abstracts as available)

Thursday Jan. 19
Infinite Free Resolutions in Commutative Algebra
Liana Sega, UMKC, Mathematics

A main topic of commutative algebra is the study of systems of polynomial equations. I will consider rings which are quotients of a polynomial ring, modulo an ideal generated by homogeneous polynomials. One can think of a module as being the cokernel of a linear operator given by a matrix with coefficients in the ring. Such a presentation allows us to identify the generators of the module, and also the relations among the generators. A free resolution of a module extends this presentation by choosing generators for the relations, and then generators for the relations between the relations, and so on; the result is an (often) infinite sequence of matrices which is called a free resolution. I will present some recent progress in the study of infinite resolutions, concerning the study of their symmetry and linearity properties in some cases of interest.

Thursday Jan. 26
There is No Place Like Home:
Estimating Intra-Conference Home Field Advantage in College Football
Byron J. Gajewski, Schools of Allied Health and Nursing,
University of Kansas Medical Center

In this talk I present a method to measure the impact of the home field advantage for intra-conference college football. The method models longitudinal data across several years while utilizing a unique home field parameter for each individual team. Additionally, two novel yet intuitive measures of home field advantage are proposed. As a case study of the method and the definitions of home field advantage, teams with the best and worst home field advantages within their respective conferences are determined.

Thursday Feb. 23, 11-11:50 am (The first of two graduate talks today.)
Beyond Intention to Treat:
Efficient Estimation of Causal Effects in Randomized Trials with Noncompliance
Jing Cheng, Division of Biostatistics, University of Pennsylvania

Randomized trials provide a powerful tool for estimating the effect of a treatment. However, noncompliance is a common problem in randomized trials. When there is noncompliance, there is often interest in estimating the causal effect of actually receiving the treatment compared to receiving the control. Three approaches to analysis of randomized trials with noncompliance are intention-to-treat (ITT) analysis, as-treated (AT) analysis and instrumental variables (IV) analysis. ITT analysis, which compares averages outcomes by randomization assignment regardless of the treatment actually received, does not estimate the effect of actually receiving the treatment but instead estimates the effect of assignment to the treatment group. AT analysis, which compares averages outcomes by actual treatment received regardless of randomization assignment, does aim to estimate the effect of actually receiving the treatment but could be biased because it compares self-selected rather than randomized groups. Like AT analysis, IV analysis aims to estimate the effect of actually receiving the treatment but unlike AT analysis, it does so it in a way that compares randomized rather than self-selected groups. In this talk, I discuss the motivation for going beyond ITT analysis to estimate the causal effect of actually receiving the treatment and present a review of the basic assumptions and principles underlying IV analysis. I then describe a new estimator for IV analysis that we have developed that is more efficient than the standard estimator used in IV analysis. Our estimator involves applying empirical likelihood with moment restrictions to mixture outcome distributions. I present simulation studies that demonstrate the gains provided by our new estimator and apply our method to data from a randomized trial of an encouragement intervention to improve adherence to prescribed depression treatments among depressed elderly patients in primary care practices.

Thursday Feb. 23, 1-1:50 pm (The second of two graduate talks today.)
Meta-analysis and Diagnostic Tests
Steve Simon, Research Biostatistician, Children's Mercy Hospital

Meta-analysis is the quantitative combination of results from multiple research studies. Meta-analysis is a relatively new field in Statistics, and standards for the proper data analysis are still evolving. Meta-analysis of studies of diagnostic tests, in particular, is especially controversial, with many conflicting approaches for computing an overall estimate from the individual sensitivity or specificity values from these studies. In the first half of this talk, I will review the general methods for the quantitative combination of results in a meta-analysis, and work out two examples using R and the meta library. In the second half, I will use data from a meta-analysis of 20 studies of endovaginal ultrasonography for detecting endometrial cancer to illustrate and critically evaluate several competing approaches for quantitatively combining results from diagnostic studies. All the data sets used in this presentation come from journal articles where the full free text is available on the web. A

handout for this seminar can be found at
http://www.childrens-mercy.org/stats/training/hand67.asp

Thursday Mar. 16
Efficient Pricing of American Options in a
Double-Exponential Jump-Diffusion Model
Farid AitSahalia, Industrial and Engineering Systems, University of Florida

In this talk I extend the numerical technique developed in AitSahlia and Lai (2001) to the pricing of American options where the underlying asset price follows a double exponential jump-diffusion. In particular, I show that the early exercise boundaries are well approximated by linear splines (in the Brownian scale) with very few knots, leading the way to a fast and accurate method to compute the option prices through a new decomposition formula.

Thursday Mar. 23
Models of Set Theory Where the Axiom of Choice Fails
Eric Hall, UMKC, Mathematics

Certain important classes of sets can be regarded as models of set theory in which the Axiom of Choice (AC) does not hold. This is one reason why there is still interest in what can happen when AC is not true or not assumed. I'll introduce the idea of permutation models, which was one of the first known techniques for creating a model of set theory without AC, and discuss some of my recent work on classification of permutation models and using permutation models to achieve independence results in ZF (the Zermelo-Fraenkel axioms for set theory).

Thursday Mar. 30
An Introduction to Cis-Regulatory Motif Discovery

Chengpeng Bi, Biostatistician, Children's Mercy Hospital

With the complete human genome sequence (3.0 billion base pairs long) now in hand, we face the enormous challenge of interpreting it and learning how to use that information to understand the biology of human health and disease. Only 1.5% human sequences make about 24,000 human genes. The majority of human sequences are non-coding that used to be called "junk" DNA. However some short non-coding sequences are called cis-regulatory motifs that are functioning in gene regulation (i.e.
activate or repress gene expression), but little are known. The goal of motif discovery is to find these

motif sequences and make a complete functional catalog of these so-called "junk" DNA in a genome.

The motif discovery can be formulated as multiple local alignments, word counting statistic, or multiple change-point problems. In this talk we focus on the multiple local alignment method. A brief overview of molecular biology relevant to motif discovery problem will be given. A DNA sequence is treated as a string consisting of four letters: A, C, G and T. The mixture probabilistic models for motif sequences are
introduced: a Markov model for background sequences and a product of multinomial distributions for motif sequences. The objective functions are formulated based on Maximum Likelihood (ML), Bayesian approaches or Shannon's entropy models. Correspondingly, three motif search algorithms are described: (1) Expectation Maximization (EM) method is specially suited for the ML sequence model; (2) Gibbs motif sampler is applied to optimize the Bayesian motif models; (3) A greedy algorithm is used to minimize entropy models. Finally we use a golden standard example to illustrate the above motif discovery methods with comparison and discuss their limitations.

Thursday Apr. 6
Stock Market Liberalization and International Risk Sharing
Shu Wu, Economics, University of Kansas

We empirically examine what macroeconomic risks are shared (or not shared) internationally after stock market liberalization in several developing countries. To address this issue, we incorporate an international asset pricing model into a non-linear structural Vector Autoregression (VAR) system that identifies various sources of macroeconomic risks. We find that most of the risks corresponding to exogenous financial market shocks are surprisingly well shared, although other macroeconomic risks associated with exogenous shocks to output, inflation and monetary policies are not fully shared across countries. Our results suggest that one of the main benefits from stock market liberalization is to allow a country to better hedge against exogenous and idiosyncratic financial market shocks, and stock market liberalization should be accompanied by other macroeconomic reforms in order to achieve the full benefits of international risk sharing.

Thursday Apr. 13
An Asymmetric Information Modeling Framework for Ultra-High Frequency Transaction Data: A Non-Linear Filtering Approach

Yoonjung Lee, Statistics, Harvard University

The paper proposes a new asymmetric information modeling framework that provides a theoretical explanation for some of the observed interactions among the key quantities in financial markets: the price impact of a trade, the duration between trades, and the degree of information asymmetry. In

the model, a private signal is partially revealed through trades, while new public information arrives continuously at the market. The market maker utilizes a non-linear filtering technique to set a competitive price that rationally incorporates these two sources of information. The pricing rule depends on the actual sequence of order arrivals, not just the total number of buy/sell orders. The price impact of a trade tends to decrease when the duration between trades gets longer. The speed at which the information gets incorporated into the price depends on the quality of the private signal and the trading rate of informed traders. I provide a procedure to estimate the parameters and discuss the sampling distribution of the parameter estimates.
The optimal trading strategy of a monopolistic insider is studied in the continuous trading model. If the private information is slowly revealed to the public, the insider trades more aggressively than he would otherwise.

Thursday Apr. 20
What has happened to the term premium? – Evidence from yield spreads between nominal and inflation indexed treasuries

Pu Shen, Senior Economist, Research Department, Federal Reserve Bank of Kansas City

The yield curve of nominal treasuries has become much flatter since the FOMC started raising the federal funds rate in June 2004. Many market analysts suggest that one of the reasons for the flattening is the decline of term premium in the nominal treasury yields. In this article, we provide some direct evidence, based on the yield spread between nominal and inflation indexed treasuries, that indeed part of the recent flattening of the yield curveis due to the decline of the term premium. The yield spread between nominal and inflation indexed treasuries mainly consists of market expectation of inflation, an inflation risk premium, and a liquidity risk premium. We first shows that the liquidity risk premium was large but may have declined considerably in the past few years, and therefore, it is important to account for the liquidity risk premium in the estimation of the inflation risk premium. Joint estimates provide evidence that the inflation risk premium has generally been declining since the FOMC started the current tightening cycle. As the inflation risk premium is an important part of the term premium, the decline of the inflation risk premium explains a large portion of the decline of the term premium, but not all of it.

Semester: Fall 2006
Location: Haag Hall, Room 312 (Unless otherwise noted)
Day & Time: Wednesdays, 2:00-2:45 pm (Unless otherwise noted)
Campus Map for Talks (PDF Format)

Organizer: Dr. Jie Chen, 235-2844
Email: chenj@umkc.edu

GO TO UMKC Department of Mathematics & Statistics

Dates, Titles, Speakers (with Abstracts as available)

Wednesday Sept. 6

Detecting Statistically Significant DNA Copy Number Changes for Array Comparative Genomic Hybridization Data

Jie Chen, Department of Mathematics and Statistics, UMKC

Cancer development is usually associated with DNA copy number changes in the genome. DNA copy number changes correspond to chromosomal aberrations and signify abnormality of a cell. Therefore, identifying statistically significant DNA copy number changes is evidently crucial in cancer research, clinical diagnostic applications, and other related genomic research. The problem can be formulated with a statistical change point theory. We propose to use the mean and variance change point model to study the DNA copy number changes from the microarray comparative genomic hybridization (aCGH) profile. The approximate p-value of identifying a change point is derived from the use of Schwarz information criterion (SIC). The procedure to detect all changes in the data is carried out using binary segmentation procedure (BSP). The proposed method has been validated by Monte-Carlo simulation and applications to aCGH profiles from several cell lines (fibroblast cancer cell line, breast tumor cell line, and breast cancer cell line). The results indicate that the proposed method is effective in identifying DNA copy number changes.

Wednesday Sept. 13

Reconstructing Signaling Pathways from High Throughput Data

Dongxiao Zhu, Bioinformatics Center, Stowers Institute for Medical Research

Many bioinformatics problems can be tackled from a fresh angle offered by the network perspective. Taking into account the network constraints on gene interaction, we propose a series of logically coherent approaches to reconstruct signaling pathways from high throughput data. The approaches proceed in three consecutive steps: co-expression network construction with controlled biological significance and statistical significance, network constrained clustering, and reconstruction of the order of pathway components.

The first step relies on detecting pairwise gene co-expression. We attack the problem from both frequentist statistics and Bayesian statistics perspectives. We designed and implemented a frequentist two-stage co-expression detection algorithm that controls both statistical significance (False Discovery Rate, FDR) and biological significance (Minimum Acceptable Strength, MAS) of the discovered co-expressions. In order to regularize variances of the correlation estimation in small

sample scenarios, we also designed and implemented a Bayesian hierarchical model, in which correlation parameters are assumed to be exchangeable and sampled from a parental Gaussian distribution. Using simulated data and the galactose metabolism data, we demonstrated advantages of our approaches and compared their differences.

The second problem considered is distance based clustering that accounts for ``network constraints" extracted from the Giant Connected Component (GCC) of the network discovered from the data. The clustering is performed using a ``hybrid" distance matrix composed of direct distance between adjacent genes and ``shortest-path" distance between non-adjacent genes in the network. The third problem considered is the reconstruction of the order of pathway components. We applied a first-order Markov model, originally developed and applied to a network tomography problem in telecommunication networks, to reconstructing three well-known signaling pathways. The methods proposed here can also be applied to other high throughput data analysis problems.

Wednesday Sept. 20
Generalized Scheffe's Confidence Intervals
on All Linear Combinations of
Multiple Heteroscedastic Means
Xin Yan, Department of Mathematics and Statistics, UMKC, (formerly Senior Biometrician, Dept. of Clinical Biostatistics, Merck Research Laboratories)

We proposed a method to construct simultaneous confidence intervals on all linear combination of multiple heteroscedastic means by utilizing Satterthwaite's approximation. The proposed simultaneous confidence intervals, to which we refer as the generalized Scheffe's simultaneous confidence intervals, have an explicit format that is similar to their classical counterpart. The proposed generalized Scheffe's intervals have exactly the same format as the well-known Scheffe's intervals when all means have an equal variance. However, by removing equal variance assumption, the proposed simultaneous confidence intervals can be potentially applied to efficacy assessment of multiple clinical trials. Moreover, we demonstrate through full simulations that the proposed generalized Scheffe's intervals preserve familywise error rate (FWER) in various configurations of means, variances, and sample sizes, especially in situations where the classical Scheffe's intervals fail to preserve FWER.

Wednesday Oct. 4
Perturbation Bounds using Relative Gap
Noah Rhee, Department of Mathematics and Statistics, UMKC

Two perturbation bounds for invariant subspaces of complex matrices are discussed, one using absolute gap and the other using relative gap. The goal is to provide intuition as well as an idea for

why the bounds hold and why they look the way they do. Then we present a bound (involving relative gap) for the distance between the original and the perturbed right singular vector subspaces of a general matrix with full column rank.

Wednesday Oct. 11
Integer Factorization in the Context of Cryptography
Venkatesulu Mandadi, Visiting Professor, Department of Mathematics and Statistics, UMKC

By the fundamental theorem of arithmetic, every positive integer has a unique prime factorization. Given an algorithm for integer factorization, one can factor any integer down to its constituent primes by repeated application of this algorithm. But there are NO polynomial time factoring algorithms.The hardness of this problem is at the heart of several important cryptographic systems.

Cryptography is an important building block of e-commerce systems.In particular, public key cryptography can be used for ensuring the confidentiality, authenticity, and integrity of information in an organization.

RSA is one of the most popular public key cryptographic algorithms in used in e-commerce.A fast integer factorization algorithm would mean that the RSA public-key algorithm was insecure.

In contrast, it may turn out that there are attacks on the RSA problem more efficient than integer factorization, though none are currently published.

Monday Oct. 16, 3:00-3:45 pm, in FL 312
Training Health Care Professionals in Data Mining
Morgan C. Wang, Department of Statistics and Actuarial Science, University of Central Florida, Orlando, FL

Wednesday Oct. 25
Similarity measures and similarity searches in genome-wide data sets
Arcady Mushegian, Director, Bioinfomatics Center, Stowers Institute for Medical Research, and Department of Microbiology, Molecular Genetics and Immunology, University of Kansas School of Medicine

Genome era produces large multidimensional datasets, which need to be analyzed in robust,

quantitative ways. Although the general picture of dependencies between genes and their products can be obtained by various clustering methods, in fact many biological questions asked of the genome-wide measurements have little to do with global clustering or with laying out the whole network. Rather, a commonly encountered task is to discover the neighbors of a point, which represents a set of measurements associated with a gene or a protein. Finding such groups does not require the knowledge of all genome-wide correlations – the fundamental task here is to discover and rank similarities that are local with regards to the complete measurement space. I will present an approach to this problem.

Wednesday Nov. 1
A stochastic EM-type algorithm for motif-finding in biopolymer sequences
Chengpeng Bi, Children's Mercy Hospitals

Position weight matrix-based (i.e. product of multinomial distribution) statistical modeling for the identification and characterization of motif sites in a set of unaligned biopolymer sequences is presented. A new algorithm, the Stochastic EM-type Algorithm for Motif finding (SEAM), is described and compared with its deterministic counterpart (i.e. EM-based algorithms) and other popular MCMC-based algorithms such as Gibbs sampling. The gold standard example, cyclic adenosine monophosphate receptor protein (CRP) binding sequences, together with other biological sequences, is used to illustrate the performance of the new algorithm. The convergence of the new algorithm is shown by simulation. Experimental results using simulated and biological examples illustrate the power and robustness of the new algorithm SEAM in de novo motif discovery.

Wednesday Nov. 8
Sampling and eigenvalues of non-self adjoint Sturm-Liouville problems
Amin Boumenir, Mathematics, University of West Georgia

In this talk we introduce a new method for the computation of eigenvalues of Sturm-Liouville operators in the non-selfadjoint case. In this case, the eigenvalues are scattered in the complex plane and classical root finding methods fail. The sampling method, which is based on interpolation, reconstructs the characteristic function over the whole complex plane and then approximates its zeros which are the eigenvalues.

Wednesday Nov. 15
Definitions of Finiteness
Eric Hall, UMKC, Mathematics

There are many characterizations of what it means for a set to be finite which are equivalent assuming the axiom of choice (AC), but not provably equivalent if AC is not assumed. We will consider various definitions of finiteness and their relationships without assuming AC. If one assumes AC, the content may be interpreted as exploring when the existence of certain combinatorial structures on an infinite set make it possible to define certain other structures.