LANGENHOP LECTURE
& SIU PROB AND STATS
CONFERENCE

Probability Participants:

Peter Baxendale, University of Southern California

Denis Bell, University of North Florida

Wesley Calvert, Southern Illinois University Carbondale

Elton Hsu, Northwestern University

Isabelle Kemajou-Brown, Morgan State University

Kay Kirkpatrick, University of Illinois at Urbana-Champaign

Arash Komaee, Southern Illinois University Carbondale

Nicholas LaRacuente, University of Illinois at Urbana-Champaign

David Nualart, University of Kansas

Nian Yao, Shenzhen University

Flavia Sancier, Antioch College

René Schott, University of Lorraine, Nancy, France

Henri Schurz, Southern Illinois University Carbondale

Lochana Siriwardena, University of Indianapolis

Heinrich von Weizsacker, Technische Universitt Kaiserslautern

Probability Schedule:

Monday, May 14, 2018

Time	**Probability Neckers 240**
9:00-9:50	Elton Hsu
10:00-10:50	Peter Baxendale
11:00-11:30	Henri Schurz
11:30-1:00	Lunch
1:00-1:50	Heinrich von Weizsacker
2:00-2:50	Denis Bell
3:00-3:30	Arash Komaee

Tuesday, May 15, 2018

Time	**Probability Neckers 240**
8:30-9:20	René Schott
9:30-10:20	David Nualart
10:30-11:00	Flavia Sancier
11:00-11:30	Isabelle Kemajou-Brown
11:30-1:00	Lunch
1:00-1:50	Kay Kirkpatrick
2:00-2:30	Nicholas LaRacuente
2:40-3:20	Wesley Calvert
3:30-4:10	Nian Yao

(1) Registration on Monday begins at 8:00 at the atrium of Neckers, followed by a welcome session 8:30-8:45 in Neckers Room 240.
(2) The Langenhop Lecture is in Guyon Auditorium of Morris Library.

Probability Abstract

Random sources and sinks for stochastic dynamical systems

Peter Baxendale
Department of Mathematics
University of Southern California

Phase portraits have proved to be an effective tool for the description of the long term behavior of deterministic dynamical systems. Here we describe some attempts to develop similar techniques for stochastic dynamical systems. We illustrate some of the essential differences between the deterministic and stochastic theories by presenting examples of stochastic dynamical systems (generated by stochastic differential equations) on the circle and the two-dimensional torus.

Smooth Densities for a Class of Degenerate Stochastic Delay-Hereditary Equations

Denis Bell
Department of Mathematics & Statistics
University of North Florida

In this talk I discuss a prior joint work with Salah Mohammed in which we establish the existence of smooth densities for stochastic delay-hereditary equations of the form

$$dx_t = A(x_{t-r})dw_t + H(t,x)dt.$$

Here, r is a fixed (strictly) positive time delay and the diffusion coefficient H is a non-anticipating functional defined on the space of paths. We establish our result under hypotheses that allow for degeneracy of the noise coefficient A on a smooth co-dimension 1 hypersurface in the ambient space. The method of proof is a standard Malliavin calculus type argument but requires new techniques to handle degneracy in this non-Markovian situation. The main mechanism in the proof is a propagation phenomenon induced by the time-delay r, which appears to have no analogue in the classical diffusion setting.

Recent Trends at the Interface of Mathematical Logic and Probability

Wesley Calvert
Department of Mathematics
Southern Illinois University Carbondale

In the late 19th and early 20th centuries, logic and probability were frequently treated as closely related disciplines. Each has, in an important sense, gone its own way, so that neither, in its modern form, is in any proper sense a systematization of the "Laws of Thought," as Boole called them. However, the last four decades have seen a remarkable rapproachment.

On the most obvious level, the various probability logics have developed as formal systems of reasoning in the modern sense of logic. At a deeper level, though, attempts have been made to formulate logics in which model theory of random variables, stochastic processes, and randomized structures can be explored from the perspective of model theory. Continuous first-order logic as a context for stability theory on metric structures is perhaps the most conspicuous example, but others exist.

At the same time, algorithmic randomness in its various forms has come to play a core role in computability theory, while probabilistic computation of various kinds (randomized computation, interactive proofs, and others) has come to dominate major parts of computational complexity. The older recursion-theoretic program of machine learning, initiated by Gold in the 1960s, has become much more important thanks to Valiant's reformulation in probabilistic terms to allow for reasonable errors.

The model theory of random objects, Fraïssé limits, and pseudofinite structures, each of which embodies some important aspect of 0-1 laws, has been important for longer, but advances in stability, simplicity, and the transition from finite to infinite model theory have enriched this subject.

In set theory, too, the study of dynamics that respect probability measures has played a central role in the study of equivalence relations. Probability is frequently at the center of modern descriptive set theory.

In the present talk, we will survey some of these developments, well-known to logicians, but potentially of interest to probabilists and statisticians.

Stochastic De Giorgi Iteration and Regularity of Stochastic Partial Differential Equations

Elton P. Hsu
Department of Mathematics
Northwestern University

De Giorgi iteration is a standard method in proving regularity results in partial differential equations. We will explain a stochastic version of its method which can be used effectively to deal with the regularity problem of stochastic partial differential equations (SPDE). In particular, it can be used to prove a strong Holder continuity of the solution of a class of divergence form semilinear SPDE with multiplicative additive noise. This is a joint work with Yu Wang and Zhenan Wang.

Optimal Control Risk-sensitive Benchmarked Asset Management under Regime Switching

Isabelle Kemajou-Brown, & Olivier Menoukeu-Pamen
Department of Mathematics
Morgan State University
Zhongyang Sun
School of Mathematics
Sun Yat-sen University
People's Republic of China
Olivier Menoukeu-Pamen
African Institute for Mathematical Sciences, Ghana
University of Ghana, Ghana
Institute for Financial and Actuarial Mathematics
Department of Mathematics
University of Liverpool, United Kingdom

We assume the stock is modeled by a Markov regime-switching diffusion process and that, the benchmark depends on the economic factor. Then, we solve a risk-sensitive benchmarked asset management problem of a firm. Our method consists of finding the portfolio strategy that minimizes the risk sensitivity of an investor in such environment, using the general maximum principle.

BIOLOGIC: Biological Computation

Kay Kirkpatrick
Departments of Mathematics and Physics
University of Illinois
O. Osuagwu
Department of Computer Science
CUNY Medgar Evers College

We will discuss newly defined machines that out-perform Turing machines. In his unpublished 1948 paper, Intelligent Machinery, Alan Turing identified several types of machines, with one dichotomy that is false, between active and controlling machines. Ill introduce a new kind of machine and define a subtype, an automatic biochemical machine, that is equivalent to a deterministic Turing machine with two oracle machines as adjuncts. Joint work with O. Osuagwu.

Topics in Estimation of Discrete Events

Arash Komaee
Department of Electrical and Computer Engineering
Southern Illinois University Carbondale

Two problems in estimation of discrete events are considered. The first problem defines a Poisson-Gauss process as the sum of a filtered Poisson process (a Poisson process passed through a linear filter) and a white Gaussian noise. With the observations of a Poisson-Gauss process, three estimation problems are considered: minimum mean squared error (MMSE) estimation of the Poisson process at every fixed but arbitrary time, MMSE estimation of the Poisson intensity, and the maximum likelihood estimation of the intensity. The second problem adopts a Bayesian framework for quickest detection of a single random pulse in the presence of white Gaussian noise.

Asymmetry, Specificity and Repeatability of Stochastic Systems

Nicholas LaRacuente & James O'Dwyer
Departments of Physics and Plant Biology
University of Illinois at Urbana-Champaign

The essence of modeling is to distill replicable features of a system from per-instance specifics, whether the latter are noise or just irrelevant details. Classically, we often assume that averaging over many samples yields cancellation between the idiosyncrasies of particular trials. This cancellation allows an underlying common structure to show through. The state we ultimately seek to measure is not the most precise characterization of the instance under study, but a sort of symmetrized analog over the multitude of possible outcomes. Many of the modern systems we would like to study, such as cities, markets, ecosystems, or many-body states of matter, are far too large or complex to effectively sample configuration space via laboratory replicates. In these cases, there is a fundamental tradeoff between the specificity of a model and the amount of data available to build it.

We relate inherent tradeoffs in modeling to quantification of asymmetry in stochastic systems. Hidden symmetries between systems not a priori given as replicates can improve forecasts by compensating for lack of repetition. We look to maximize the entropy of irrelevant degrees of freedom conditioned on repeatable underlying dynamics, correspondingly minimizing the entropy in a hypothetical symmetrized basis. We demonstrate these principles using a simple algorithm for nonlinear timeseries analysis similar to the existing Empirical Dynamic Modeling method. We show how our algorithm overcomes common problems in timeseries analysis and discuss connections to broader questions in modeling.

Central limit theorems for functionals of Gaussian processes

David Nualart
The University of Kansas

In this talk we will first present a version of the Breuer-Major Theorem for a class of self-similar Gaussian processes, that includes processes without stationary increments like the bifractional Brownian motion. The proof

is based on chaos expansions and the Fourth Moment Theorem. We will also discuss the rate of convergence of the total variation distance in the framework of the Breuer-Major Theorem and its generalizations.

Portfolio and reinsurance for an insurer with default risk

Nian Yao
College of Mathematics and Statistics
Shenzhen University

Abstract: We study the optimal excess-of-loss reinsurance and portfolio for an insurer in a defaultable market by a general stochastic volatility model. The insurer is assumed to buy reinsurance and to invest in the following securities: a bank account, a risky asset with stochastic volatility, and a defaultable corporate bond. We discuss the optimal investment strategy through two subproblems: a pre-default case and a post-default case, respectively. We show the existence of a classical solution to a pre-default case via super-sub solution techniques and give an explicit characterization of the optimal reinsurance and investment policy that maximizes a common used utility associated with the terminal wealth. Verification theorem is established to show the uniqueness of the corresponding solution of HJB equation. Moreover, the portfolio of n defaultable corporate bonds and a bank account with reinsurance are also considered.

Testing a Generalized Black-Scholes Model with Hereditary Structure

Flavia Sancier
Sciences Division
Antioch College

In this talk, we discuss the testing of a continuous option pricing model that has stochastic volatility with hereditary structure. The stock dynamics follows a generalized geometric Brownian Motion described by a nonlinear stochastic functional differential equation. The option pricing formula is the result of an equivalent (local) martingale measure and therefore are written as a conditional expectation that can be simulated via Monte Carlo methods. We assess the model's performance using real market data from the S&P500 index from 2008 to 2010.

On Stochastic Calculus with Respect to q-Brownian Motion

René SCHOTT
Institut Elie Cartan and Loria
University of Lorraine, Nancy, France

We pursue the investigations initiated by Donati-Martin regarding stochastic calculus with respect to q-Brownian motion, and essentialy extend the previous results along two directions:
(i) We develop a robust L^∞-integration theory based on rough-paths principles and apply it to the study of q-Bm driven differential equations;
(ii) We provide a comprehensive description of the multiplication properties in the q-Wiener chaos.
Our presentation follows a probabilistic pattern, in the sense that it only leans on the law of the process and not on its particular construction. Besides, our formulation puts the stress on the rich combinatorics behind non-commutative processes, in the spirit of the machinery developed by Nica and Speicher.
(based on joint work with A. Deya)

L^p-Theory of SFDEs with Infinite Memory

Henri Schurz
Department of Mathematics
Southern Illinois University Carbondale

Consider Itô SFDEs with continuous coefficients and infinite memory

$$dX(t) \;=\; -a(t)X(t)dt + b(t,X(t))dW(t) + [\int_{-\infty}^{t} D(t,s)f(X(t))ds]dt$$

with L^1-integrable kernel D, driven by standard Wiener process W and started at adapted initial values

$$X(s) \;=\; \psi(s) \qquad \text{for } s \leq 0$$

at time $t = 0$, where $\psi \in C^0_{ub}((-\infty, 0])$. We discuss existence and uniqueness of L^p-solutions, L^p-regularity (Hölder-continuity), asymptotic L^p-stability and, if time permits, also convergence of Euler-Riemann-Maruyama-type numerical methods. The major idea is to use Banachs contraction mapping principle (CMP) in order to derive the main conclusions (which was

already exploited by T. Burton in ordinary FDEs). These are results of a joint work with Saeed A. Althubiti and motiviated by the brilliant works of former colleagues Ted Burton and Salah Mohammed in related fields. This talk is especially devoted to my esteemed friend Salah in memory of numerous academic research discussions with him.

Looking back to three decades at Kaiserslautern

Heinrich von Weizsäcker
Fachbereich Mathematik
Technische Universität Kaiserslautern, Germany

Salah Mohammed first visited our department in the early eighties. He returned every few years at least for a couple of weeks, usually bringing part of his family, the last visit being in August 2016 to Berlin. This talk wants to provide an brief overview of what our probability group did during those years, starting from Michael Scheutzow's approach to stochastic delay equation, passing over a few measure theoretic results and ending with unpublished work on diffusions conditioned on survival, by Martin Anders, my last PhD student.

LANGENHOP LECTURE
& SIU PROB AND STATS
CONFERENCE

Statistics Participants:

Ali Arab, Georgetown University

David Banks, Dake University

Sanjib Basu, University of Illinois at Chicago

Emily Berg, Iowa State University

Lynne Billard, University of Georgia

Sounak Chakraborty, University of Missouri-Columbia

Xinyu Chen, Worcester Polytechnic Institute

Priyan DeAlwis, Southern Illinois University Carbondale

Ashkan Ertefaie, University of Rochester

Mohammad Kazem Shirani Faradonbeh, University of Florida

Joyee Ghosh, University of Iowa

Naama Lewis, Southern Illinois University Carbondale

Mohammad Reza Meshkani, Shahid Beheshti University

Nitis Mukhopadhyay, University of Connecticut

Balgobin Nandram, Worcester Polytechnique Institute

Chathurangi Pathiravasan, Southern Illinois University Carbondale

Buddika Peiris, Worcester Polytechnique Institute

Lasanthi Pelawa Watagoda, Appalachian Sate University

Hasthika Rupasinghe, Appalachian State University

Hadi Safari, Southern Illinois University Carbondale

Cao Quy, University of Montana

Mehdi Soleymani, University of Auckland, New Zealand

T.N. Sriram, University of Georgia

Christopher K. Wikle, University of Missouri

Ping Ye, University of North Georgia

Huijun Yi, Troy State University

Yuan Yu, Worcester Polytechnic Institute

Samira Zaroudi, Azad University

Yaser Samadi, Southern Illinois University Carbondale

Econometrics

Nazmul Ahsan, St. Louis University

Anil Bera, University of Illinois at Urbana-Champaign

Sajal Lahiri, Southern Illinois University Carbondale

Biostatistics

John Reeve, Southern Illinois University Carbondale

Lihui Zhao, Northwestern University

Data Science

Mary Frances Dorn, Los Alamos National Laboratory

Statistics Schedule

Monday, May 14, 2018

Time	Keynote Session-Statistics Neckers 440	Time	Biostatistics Neckers 218
9:00-9:50	Nitis Mukopadhyay	1:00-1:35	Lihui Zhao
9:55-10:45	T.N. Sriram	1:40-2:15	Cao Quy
10:50-11:40	Christopher Wikle	2:20-2.55	John Reeve
11:30-1:00	Lunch		

Time	Bayesian Analysis Neckers 440	Time	Statistics Neckers 156
1:00-1:35	Joyee Ghosh	1:00-1:25	Lasanthi Pelawa-Watagoda
1:40-2:15	Mohommad Meshkani	1:25-1:50	Hasthika Rupasinghe
2:20-2.55	Sounak Chakraborty	1:50-2:15	Chathurangi Pathirawasan
		2:15-2:30	Break
		2:30-2:55	Ping Ye
		2:55-3:20	Huijun Yi
		3:20-3:45	Buddika Peiris

(1) Registration on Monday begins at 8:00 at the atrium of Neckers, followed by a welcome session 8:30-8:45 in Neckers Room 240.
(2) The Langenhop Lecture is in Guyon Auditorium of Morris Library.

Tuesday, May 15, 2018

Time	Keynote Session-Statistics Neckers 440	Time	Statistics Neckers 218
8:30-9:20	Balgobin Nandram	9:00-9:35	Nazmul Ahsan
9:20-10:10	Sanjib Basu	9:40-10:15	Anil Bera
10:10-10:20	Break	10:15-10:25	Break
10:20-11:10	David Bank	10:25–11:00	Ashkan Ertefaie
11:10-12:00	Lynne Billard	11:05-11:50	Sajal Lahiri
11:30-1:00	Lunch		

Time	Statistics Neckers 440	Time	Statistics Neckers 218
1:00-1:35	Ali Arab	1:00-1:25	Mary Frances Dron
1:40-2:15	Yaser Samadi	1:25-1:50	Naama Lewis
2:20-2.55	Mohommad Faradonbeh	1:50-2:15	Yuan Yu
2:55-3:10	Break	2:15-2:30	Break
3:10-3:45	Emily Berg	2:30-2:55	Xinyu Chen
3:50:4:25	Mehdi Soleymani	2:55-3:20	Priyan De Aliws
		3:20-3:45	Hadi Safari
		3:45-4:10	Samira Zaroudi

Statistics Abstract

Spatio-Temporal Models for Modeling Rare Events Count Data

Ali Arab
Department of Mathematics & Statistics
Georgetown University

Data on frequency of rare events often include excess zeroes, and potentially, extreme values. This is particularly the case for data on events, conditions, or diseases that are not common in specific areas or specific time periods, or those that are hard to detect or on the rise. The complex nature of the distribution of these data (i.e., excess zeroes, often with heavy tails) makes the modeling of the data challenging as the typical data models based on common count probability distributions (e.g., Poisson or negative binomial) are not capable of providing a reasonable fit to the data and fall short of predicting rare outcomes. Examples include number of cases of Lyme disease in the Midwest (e.g., Illinois) where the disease is not common or is on the rise, number of casualties related to mass shootings over the past several months in a particular state, or number of observed endangered species in an ecological monitoring program. Moreover, in many applications the dynamics of evolution or spread of rare events over space and time, although very difficult to understand, is of great interest for surveillance and preparedness management. Critically, this requires the development of dynamical spatio-temporal models which can effectively address the challenges of rare events data distributions with excess zeroes and heavy tails. In this work, we provide a review of modeling approaches for rare events count data and propose strategies for modeling spatial, temporal, and spatio-temporal rare events count data. Finally, we provide a case study on counts of confirmed cases of Lyme disease in the United States.

Adversarial Risk Analysis

David Banks
Department of Statistical Science
Duke University

Adversarial Risk Analysis (ARA) is a Bayesian alternative to classical game theory. Rooted in decision theory, one builds a model for the decision-making of one's opponent, placing subjective distributions over all unknown quantities. Then one chooses the action that maximizes expected utility. This approach aligns with some perspectives in modern behavioral economics, and enables principled analysis of novel problems, such as a multiparty auction in which there is no common knowledge and different bidders have different opinions about each other.

Bayesian Variable Selection in Linear and Nonlinear Models

Sanjib Basu
University of Illinois at Chicago

We consider the question of variable selection in complex models. This is often a difficult problem due to the inherent nonlinearity of the models and the resulting non-conjugacy in their Bayesian analysis. We consider a comparative review of Bayesian criterion based variable selection. We also propose an efficient variable selection method and illustrate its performance in simulation studies and real example.

Small Area Prediction for County Level Erosion Rates based on a Mixed Effects Quantile Regression Model

Emily Berg & Danhyang Lee
Department of Statistics
Iowa State University

Prediction of small area quantiles has relevance for poverty analysis, monitoring of water quality, and forestry. In the motivating application, estimates of small area quantiles of several measures of erosion are of interest. The data are from a national survey called the Conservation Effects Assessment Project (CEAP). Quantile regression is appealing for CEAP because finding a single family of parametric models that adequately describes the distribution of all variables is difficult and the quantile function is a parameter of

interest. We consider small area prediction based on a mixed effects quantile regression model. The estimation procedure exploits the linearly interpolated generalized Pareto distribution (LIGPD). Empirical Bayes predictors and bootstrap mean squared error estimators are constructed. Through simulation, we compare predictors of small area quantiles based on the LIGPD to alternative predictors and evaluate the quality of bootstrap mean squared error estimators. We apply the LIGPD approach to obtain predictors of county level quantiles for several measures of soil and nutrient loss.

Clustering Histograms

L. Billard
Department of Statistics
University of Georgia, Athens GA
Jaejik Kim
Department of Statistics
Sungkyunkwan University, Seoul

One of the common issues in large dataset analyses is to detect homogeneous groups of objects. We present a divisive hierarchical clustering method for histogram data. Unlike classical data points, a histogram has internal variation of itself as well as location information. However, to find the optimal bipartition, existing divisive monothetic clustering methods for histogram data consider only location information as a monothetic characteristic and they cannot distinguish histograms with the same location but different internal variations. Thus, a divisive clustering method considering both location and internal variation of histograms is described.

Bayesian Kernel Based Models and Analysis of High-dimensional Multiplatform Genomics Data

Sounak Chakraborty
Department of Statistics
University of Missouri-Columbia

In recent years we have seen rapid development of new technologies for genome-wide assays. With increasing reliability and affordability of microarray and next-generation sequencing, patient care decisions are now customized based on the diverse genetic and epigenetic alterations of a disease

for a specific individual. Moreover, nowadays the scale of omcis studies has expanded to measure and include multiple genomic features on a single patient, like gene expression, DNA methylation, gene mutation, copy number variation, promoter binding and protein expression. Combining and modeling multiple genome features coming from different data platforms is a big conceptual challenge and practical hurdle.

In this paper we propose to develop statistical nonlinear models to integrate genomic data from multiple platforms. Our models can incorporate the fundamental biological relationships that exist among the data obtained from different platforms and produce more accurate understanding of the functional responses. The proposed models are developed on the basis of Bayesian trees and Bayesian kernel machine models. Our methodologies are highly flexible in exploring, extracting, and analyzing complex biological systems and datasets from heterogeneous platforms. Combining all available genetic, pathological, and demographic information can dramatically improve the nature of clinical diagnosis and treatment of several human diseases.

Inference in Constrained Linear Regression

Xinyu Chen
Department of Mathematical Science
Worcester Polytechnic Institute

Regression analyses constitutes an important part of the statistical inference and has great applications in many areas. In some applications, we strongly believe that the regression function changes monotonically with some or all of the predictor variables in a region of interest. Deriving analyses under such constraints will be an enormous task. In our work, the restricted prediction interval for a new observation is constructed when two predictors are present. We use a modified likelihood ratio test(LRT) to construct prediction intervals. The formulas developed are applied on a real data set to predict the numbder of days that a new patient will stay in hosptial using patient's age and infection risk as predictors, and compared with the unrestircted model.

Fourier Methods for Estimating the Central Subspaces in Time Series

Priyan De Alwis & S. Yaser Samadi
Department of Mathematics
Southern Illinois University Carbondale, IL

The main objective of time series analysis is to make inference about the conditional mean and the conditional variance functions. Using the Fourier transformation, we have developed a new statistical method to estimate the conditional mean and variance functions of a nonlinear time series. To this end, we have derived the candidate matrices that their column spaces span the central subspaces efficiently. The estimated subspaces are used to estimate the conditional mean and variance functions of the given model. Simulation results for different types of time series models are presented to evaluate the performance of the proposed method and compare it with other existing methods.

Sensitivity Analysis and Power in the Presence of Many Weak Instruments

Ashkan Ertefaie
Department of Biostatistics
University of Rochester

This article discusses a sensitivity analysis for an instrumental variable (IV) estimate in the presence of many instruments that are weakly associated with the endogenous variable. We study the effect of imprisonment on earnings using data on all individuals sentenced for felony in Michigan in the years 2003-2006. Motivated by the random assignment of judges within a county to felony cases, we construct a vector of putative instruments based on judges′ ID. Our data have two important features that cannot be handled using standard IV approaches. First, while some judges exhibit strong tendencies towards a prison or non-prison sentence, many judges that do not have strong tendencies toward a particular sentence type. Second, our data includes only cases that result in conviction and sentencing (e.g., it exclude cases settled by a plea bargain), and thus the standard analyses are subject to selection bias. We propose an estimation procedure that is robust to the presence of many weak instruments and develop a sensitivity analysis that quantifies the effect of the selection bias on the causal effect of interest. A

power formula for the sensitivity analysis is also provided. Analyses show that being sentenced to prison significantly reduces the offenders earnings and the largest effect is found among younger white female offenders. Our simulation studies highlight the value of the proposed method in terms of statistical power and also confirm the validity of our power formula.

Finite Time Identification in Unstable Linear Systems

Mohamad Kazem Shirani Faradonbeh
UF Informatics Institute
University of Florida

Ambuj Tewari
Department of Statistics
University of Michigan

George Michailidis
Department of Statistics
University of Florida

Identification of the parameters of stable linear dynamical systems is a well-studied problem in the literature, both in the low and high-dimensional settings. However, there are hardly any results for the unstable case, especially regarding *finite time bounds*. For this setting, classical results on least-squares estimation of the dynamics parameters are not applicable and therefore new concepts and technical approaches need to be developed to address the issue. Unstable linear systems reflect key real applications in control theory, econometrics, and finance.

This study establishes finite time bounds for the identification error of the least-squares estimates for a fairly large class of heavy-tailed noise distributions, and transition matrices of such systems. The results relate the time length required as a function of the problem dimension and key characteristics of the true underlying transition matrix and the noise distribution. To obtain them, appropriate concentration inequalities for random matrices and for sequences of martingale differences are leveraged.

Robust Bayesian Model Averaging

Joyee Ghosh

Department of Statistics and Actuarial Science

University of Iowa

The majority of Bayesian variable selection methods/algorithms for linear regression have focused on normal errors, which is a venerable problem in its own right, when the number of variables exceeds the sample size. Since estimates obtained under the normality assumption can be sensitive to outliers, robustifying the error distribution may be of interest, especially in high dimensions, when standard model diagnostics do not work well. The Bayesian variable selection approach can handle an unknown degree of sparsity by placing a prior on the inclusion probability of variables. In this work, we develop Bayesian models which allow additional flexibility, by incorporating an unknown degree of tail heaviness in the likelihood. We compare and contrast the results with those obtained from models with normal errors.

Convex Optimization and I-projections for Item Response Theory Models

Naama Lewis
Department of Mathematics
Southern Illinois University Carbondale

The area of convex optimization has been a focus of much recent research due to the wide range of problems that may be cast into the convex optimization framework and the efficiency of algorithms for solving them. Of particular importance in this domain is the notion of Fenchel duality, as many optimization problems are easier to solve in their dual space. (Borwien, Hamilton, 2006). Item Response Models, like the one parameter, two parameters, or normal Ogive, have been discussed for many years. Here we propose a new way of looking at the IRT models using I-projections and duality. We propose a projection algorithm for solving these types of problems.

New Developments in Determining Objective Priors

M. Reza Meshkani
Department of Statistics
Shahid Beheshti University, Tehran, Iran

Prior distributions are the essential part of Bayesian statistical analyses. During the past tow and half centuries since Thomas Bayes work, some researchers have suggested intuitive, heuristic, and rigorous rules for choosing priors. Great efforts have been devoted to refine the procedure of determining a type of prior which minimum effect on Bayesian inference.

In this talk we review the historical development of determining objective priors. Having explained the background ideas of subjective probability, we examine the advantages and disadvantages of proposed methods. The talk culminates in the recent developments of reference priors.

A General Sequential Fixed-Accuracy Confidence Interval Estimation Methodology for a Positive Parameter

Nitis Mukhopadhyay
Department of Statistics
University of Connecticut, Storrs

Estimation of positive parameters is important in areas including ecology, biology, medicine, nuclear power, and study of cell membranes. Mukhopadhyay and Banerjee (2014, Sequential Analysis 33: 251-285) developed a fixed-accuracy sequential confidence interval methodology for the mean of a negative binomial (NB) distribution having its thatch parameter unknown with applications in statistical ecology. In this presentation, we will outline a broad structure for fixed-accuracy sequential confidence interval estimation methodology for a positive parameter of an arbitrary distribution which may be discrete or continuous. We construct a confidence interval of the form: [T/d, dT] with d ¿ 1, based on a maximum likelihood (ML) estimator T. We will point out that the methodology enjoys attractive properties such as asymptotic (as d?1) consistency and asymptotic first-order efficiency. Specific illustrations will be included. Data analyses from large-scale simulations will be briefly incorporated which would substantiate encouraging performances

of the proposed estimation methodology. We will emphasize illustrations corresponding to the Bernoulli distribution (odds-ratio of poisonous mushrooms), Poisson distribution (radioactive decay of isotopes), and a normal distribution with the same mean and variance (real-time 911 calls dispatch). This presentation is based on joint research with Professor Swarnali Banerjee (Ph.D., UConn-Storrs, July 2014) who is at Loyola University, Chicago.

Bayesian Projective Inference of Finite Population Proportions for Sub-areas

Balgobin Nandram
Department of Mathematical Sciences
Worcester Polytechnic Institute

A standard problem in official statistics is to predict the finite population proportion of a small area when individual-level data are available from a survey and more extensive data (covariates but not responses) are available from a census. The census and the survey consist of the same strata and primary sampling units that are matched, but the households are not matched and the covariates in the sample and the census are different. The Nepal Standards of Living Survey and the recent census provide an example in which a PPS sample of the wards (PSUs) is selected and a systematic sample of the households within the wards is selected. In the largest stratum less than one percent of the wards and 12 households within the sampled wards are sampled.We are interested in the health portion of the survey in which each individual in a household is categorized into one of four health classes. Using a two-stage procedure, we study the counts in the households within the wards and a projection method to infer about the nonsampled households and wards. This is accommodated by a four- stage hierarchical Bayesian model for multinomial counts as it is necessary to accommodate heterogeneity. To fit the model, we compare two computational methods, an approximate method and an exact method, that are used to obtain the distribution of the proportions in each health class, and then we use this distribution to do projective inference for the finite population proportions. In addition, we compare the heterogeneous model, with household effects, and a homogeneous model, without household effects, and two projection procedures (nonparametric and parametric). Key Words: Dirichlet distribution, Hierarchical Bayesian model, Iterative reweighted

A new semi-parametric approach to one-way ANOVA

Chathurangi Pathiravasan & Bhaskar Battacharya
Department of Mathematics
Southern Illinois University Carbondale

The one-way analysis of variance (ANOVA) is mainly based on several assumptions and can be used to compare the means of two or more independent groups of a factor. To relax the normality assumption in one-way ANOVA, recent studies have considered exponential distortion or tilt of a reference distribution. The reason for the exponential distortion was not investigated before; thus the main objective of the study is to closely examine the reason behind it. As a result of that, a new generalized semi-parametric approach for one-way ANOVA is introduced. This generalized approach relaxes all the assumptions in one-way ANOVA and it can be applied to any type of distribution. The performance of our method is demonstrated on simulated data examples, and compared with existing techniques for one-way ANOVA.

Bayesian Analysis of a ROC Curve for Categorical Data

Buddika Peiris
Department of Mathematical Sciences
Worcester Polytechnic Institute

In a taste-testing experiment, foods are withdrawn from storage at various times and a panel of tasters are asked to rate the foods, which are rated on a nine-point hedonic scale. We provide a statistical procedure that can assess the difference between fresh foods and foods withdrawn a few months later. Thus, we have two sets of ordinal data, one for the fresh foods and the other for the foods which are withdrawn. A natural and popular way to compare two withdrawals is to use the receiver operating characteristic (ROC) curve and the area under the curve (AUC). We perform Bayesian methods, which incorporate a stochastic ordering, to obtain the AUC, but these methods are well known. However, our first method, which robustifies the binormal model, is novel. Our second method is more innovative because it uses a skew binormal model with additional robustification like the binormal method. We use the Gibbs sampler to fit both models in order to estimate the ROC curves and the AUCs. These AUCs demonstrate that there is not much difference between fresh foods and those withdrawn later using both methods. However

we have shown, using marginal likelihoods, that skew binormal model is better.

Comparing Shrinkage Estimators With Asymptotically Optimal Prediction Intervals

Lasanthi Pelawa Watagoda
Department of Mathematical Sciences
Appalachian State University

Consider the multiple linear regression model $Y = \beta_1 x_1 + \cdots + \beta_p x_p + e$. If n is the sample size, prediction intervals are developed that may be useful even if $p > n$. The length and the coverage of the prediction intervals can be used to compare shrinkage estimators such as forward selection, ridge and lasso.

Bayesian Tensor Regression for Neuroimaging Data

Hossein Moradi Rekabdarkolaee
Department of Statistical Sciences and Operations Research
Virginia Commonwealth University

Data in the form of a multidimensional array often referred to as tensor, are used in neuroimaging and other big data applications. We propose a Tensor linear model to a neuroimaging problem with brain image as response and a vector of predictors. Our method provides estimates for the parameters of interest by using a generalized sparsity principle. This procedure is in fully Bayesian setting to characterize different sources of uncertainty and inference is performed using MCMC. We show the posterior consistency and develop a computational efficient Markov Chain Monte Carlo algorithm using a block Gibbs sampler for posterior computation. The effectiveness of our approach is illustrated through simulation studies and analysis of Cocaine addiction's effect on brain connectivity.

A new regularization and variable selection technique - HRLR

Hasthika S. Rupasinghe Arachchige Don
Department of Mathematical Sciences
Appalachian State University

This work propose a new variable selection and parameter estimation method for the multiple linear regression model $Y = \beta_1 x_1 + \cdots + \beta_p x_p + e$. This new method is a hybrid of ridge regression and relaxed lasso regularization. Theoretical and simulated results demonstrate that the new method produces sparser models with equal or lower prediction loss than the regular Lasso and Relaxed Lasso estimators for high dimensional data.

Modeling Count Data via Copulas: Comparison of Kendall's tau and Spearman's rho

Hadi Safari .K & S. Yaser Samadi
Department of Mathematics
Southern Illinois University Carbondale

Copula models have been widely used to model dependence between continuous random variables, but modelling count data via copulas has recently become popular in the statistics literature. Spearman's rho is a widely used measure for the strength of association between two random variables. In this talk, we propose the population version of Spearman's rho correlation via copulas when both random variables are discrete. The explicit form of the Spearman correlation are obtained for some copulas of simple structure such as Archimedean copula family. Then, the Spearman's rho correlations are compared with their corresponding Kendall's tau values. Finally, the results are applied to model the count data in our simulation study and a real data analysis.

Statistical Methods for Integrative Analysis of Multiple Data Types

Sandra Safo
Division of Biostatistics
University of Minnesota

Over the past years, a host of researchers have become committed to developing statistical methods for integrating multiple data types, recognizing that the mechanisms that underlie complex diseases may not be unraveled by single-type data analyses alone but by the examination and integration of these multifaceted data types. However, integrating these different data types is statistically challenging partly because the data are complex, heterogeneous, and of high dimension. Dimension reduction methods capable of feature selection, and can account for the complex relationships among variables, or heterogeneity in each data type may prove useful.

First, we present a dimension reduction method that exploits the relationship between multivariate analysis methods and the generalized eigen value problem to produce sparse and interpretable solution vectors for both individual and integrative analysis of high dimensional data types. Second, given that many biomedical data are complex, with the variables functionally structured in pathways, we extend the sparse methods to incorporate known biological information via undirected graphical networks to help unravel complex mechanisms in complex diseases. We demonstrate the utility of the methods via unsupervised statistical learning methods specifically for assessing association between transcriptomic and metabolomic data from a Predictive Health Institute study that includes healthy adults at a high risk of developing cardiovascular diseases. With time permitting, I will talk about our ongoing supervised statistical learning method for simultaneous data integration and classification.

Sequential bagging for classification and regression

Mehdi Soleymani
Department of Statistics
University of Auckland
Stephen Lee
Department of Statistics and Actuarial Science
University of Hong Kong

The resampling step of conventional bagged classifiers requires the predictors and the labels to be bonded together implicitly, i.e. each resampled point has the same class label as that originally observed in the data set. The sequential algorithm reflects the chance mechanism underlying the generation of class labels conditional on the observed sample by assigning a random distribution to the class labels in each draw. This assigned random distribution is estimated by a prediction step carried on the original data set.

Online Sequential Leveraging Sampling Method for Streaming Time Series Data

Rui Xie, T. N. Sriram & Ping Ma
Department of Statistics
University Georgia
Wei Biao Wu
Department of Statistics
University of Chicago

Advances in data acquisition technology pose challenges in analyzing large volumes of streaming data. Sampling is a natural yet powerful tool for analyzing such data sets due to their competent estimation accuracy and low computational cost. Unfortunately, sampling methods and their statistical properties for streaming data, especially streaming time series data, are not well studied in the literature. In this article, we propose an online leverage-based sequential sampling algorithm for streaming time series data, which is assumed to come from an autoregressive model of order $p1$ (AR(p)). The proposed sequential leveraging sampling method samples only one consecutively recorded block from the data stream for inference. While the starting point of the sequential sampling scheme is chosen using a random mechanism based on leverage scores of the data, the subsample size is decided by

a sequential sampling threshold. We show that an appropriately normalized sequential least squares estimator of the AR parameter vector is uniformly asymptotically normally distributed for non-explosive AR(p) model. Simulation studies and real data examples are presented to evaluate the empirical performance of the proposed sequential leveraging sampling method.

Multi Time-Scale Spatio-Temporal Dynamic Models Motivated by Machine Learning

Christopher K. Wikle
Department of Statistics
University of Missouri

Spatio-temporal data are ubiquitous in engineering and the sciences, and their study is important for understanding and predicting a wide variety of processes. One of the chief difficulties in modeling spatial processes that change with time is the complexity of the dependence structures that must describe how such a process varies, and the presence of high-dimensional complex datasets and large prediction domains. It is particularly challenging to specify parameterizations for nonlinear dynamical spatio-temporal models that are simultaneously useful scientifically and efficient computationally. Statisticians have developed some ?deep? mechanistically-motivated models that can accommodate process complexity as well as the uncertainties in the predictions and inference. However, these models can be expensive and are typically application specific. On the other hand, the science, engineering, and machine learning communities have developed alternative approaches for nonlinear spatio-temporal modeling, in some cases with fairly parsimonious parameterizations. These approaches can be quite flexible and sometimes can be implemented quite efficiently, but typically without formal uncertainty quantification. Here, we present a multi time-scale spatio-temporal dynamical model that places a special parsimonious class of recurrent neural networks in a statistical framework that can account for uncertainty. This is illustrated on a multi-scale process related to long lead-time forecasting of atmospheric events given ocean conditions.

Bring Undergraduate Research In Data Science Into The Classroom

Ping Ye
Department of Mathematics
University of North Georgia

Stepping out of the own comfort zone is not easy for both the students and faculty advisor. PIC Math program provides a great opportunity for the faculty advisor to integrate research into the undergraduate classroom and for students to collaborate with the world outside of the University. The experience to go through the whole process turns out to be extremely valuable for everybody. The undergraduate research project of the University of North Georgia students will be introduced in this paper.

A Bayesian Analysis of Contingency Tables with Constrained Sample Data

Huijun Yi
Department of Mathematics
Troy University

In the analysis of contingency tables, often we face the situation: the reported data fail to match the true population. That is, sampled and target populations are not identical. Thus the resulting estimates of cell probabilities lead to the relative loss of information due to nontruthful-reporting. It is of interest to study Bayesian method by imposing constraints on priors, for example, the independent Dirichlets on marginal probabilities which carry the information of unknown parameters. In the comparison study, we also consider the unconstrained or noninformative Dirichlet priors. In the simulation study we compute posterior predictive expectations to measure missing information. An example is then used to illustrate methods.

Bayesian Analysis of Unrelated Question Design for Correlated Sensitive Questions from Small Areas

Yuan Yu & Balgobin Nandram
Department of Mathematics
Worcester Polytechnic Institute

In sample surveys with sensitive questions, random response techniques, like the unrelated question methodology, have a huge advantage in estimating population proportions by adjusting for non-response or untruthful response. Given binary response data from combined sensitive questions of many areas, the counts for each combination follows a multinomial distribution within each area. We assume the probability parameters have a Dirichlet prior. Based on this, we can construct a hierarchical Bayesian model with latent variables. Markov chain Monte Carlo methods are applied to predict the finite population proportions. We explore how the correlation between the two sensitive questions will affect the estimation. We use a simulation study and an application on body mass index data from the Third National Health and Nutrition Examination Survey to study our procedures.

Application of Copula for Modeling Reserves in Life Insurance Market

Samira Zaroudi
Department of Statistics
Science and Research, Azad University, Tehran, Iran

Copula models are very appropriate tools for assessing the relationship between two lifetimes. In this talk, life insurance reserves have been calculated for couples lifetime by copula. These calculations have been carried out under some Archimedean copulas for "survival of both individuals" and "death of one individual". The proposed model has been applied to Iran insurance industry using its lifetime table. Our estimation results indicate that insurer reserves with using Archimedean copula are greater than independence condition. So, according to this result, considering the association between two lifetimes for calculating the optimal reserves by the insurer is highly recommended.

Time Series Analysis for Symbolic interval-valued Data

S. Yaser Samadi & Lynne Billard
Department of Mathematics
Southern Illinois University, University of Georgia

While many series record a single value for each time point, many other series record the observations as intervals. This is particularly so with financial data, where, e.g., assets have two prices (bid and ask prices) and the interval between them represents all possible prices at which the asset can be traded. There are countless examples. Therefore, in comparison with standard classical data, they are more complex and can have structures (especially internal structures) that impose complications that are not evident in classical data. As a result of dependency in time series observations, it is difficult to deal with symbolic interval-valued time series data and take into account their complex structure and internal variability. In the literature, the proposed procedures for analyzing interval-valued time series data used either midpoint or radius that are inappropriate surrogates for symbolic interval variables. All previously available methods in the literature fail in some way to use all the variations inherent in the interval-valued data; there is a loss of information. We develop a methodology using the information contained in the complete intervals (and not just on the two point values represented by the end points and/or the center-range values) to analyze interval time series data.

Econometrics Abstract

Social Promotion and Learning Outcomes: Evidence from the Right to Education Act in India

Nazmul Ahsan
Department of Economics
Saint Louis University
Ashna Arora
Department of Economics
Columbia University
Rakesh Banerjee
Baker Institute for Public Policy
Rice University
Siddharth Hari
Department of Economics
Virginia Tech

While primary school enrollment rates in several developing countries have grown exponentially over the past few decades, learning levels often continue to remain extremely low, and drop out rate high. It is therefore crucial to understand what policies improve learning outcomes. In this paper, we study the effect of social promotion policies - under which students in primary school cannot be asked to repeat a grade. Ex-ante, it is not clear what effect such a policy will have. On the one hand, it might lower the incentives to invest effort in learning, both by students and parents. On the other, it might encourage students to learn in a fear-free environment, leading to better outcomes. Additionally, in developing countries such a policy might help in reducing drop out rates. We analyze this question in the context of a policy introduced by the Government of India in 2009, which prohibited the failing of students till grade 8. Prior to the introduction of this policy, different states followed similar policies till different grades. Using this variation, we setup a difference-in-differences strategy exploiting the variation in treatment at the state-grade level. Using a large scale survey data on learning outcomes for the period 2007-2015, we find that the policy improved average reading score by 2 percent and math score by 4 percent. In addition, we find children of educated mothers benefit the most, indicating compensatory investment by parents. The different mechanisms through which these effects operate are also discussed.

Asymptotic Variance of Test Statistics in ML and QML Frameworks

Anil K. Bera & Osman Dogan
Economics Program
University of Illinois
Suleyman Taspinar
Economics Program
Queens College, The City University of New York

In this study, we consider test statistics that can be written as the sample averages of data and derive their limiting distribution under the maximum likelihood (ML) and the quasi-maximum likelihood (QML) frameworks. We first generalize the asymptotic variance formula suggested in Pierce (1982) in the ML framework and illustrate its applications through some well-known test statistics: (i) the skewness statistic, (ii) the kurtosis statistic, (iii) the Coxs statistic,(iv) the information matrix test statistic, and (v) the Durbins h-statistic. We next provide a similar result in the QML setting and illustrate its applications by providing two examples. Illustrations show the simplicity and the effectiveness of our results for the asymptotic variance of test statistics, and therefore, they are recommended for practical applications.

Is the Internet Bringing Down Language-based Barriers to International Trade?

Erick Kitenge
College of Business
Central State University, Wilberforce, Ohio
and
Sajal Lahiri
Department of Economics
Southern Illinois University Carbondale

This paper studies the effect of Internet access on language-based barriers to trade, using bilateral manufacturing export data for eight sectors from 210 countries. Using recent developments in gravity analysis, we find that Internet access is breaking down language-based barriers to trade. The result is strong and robust. We find, *inter alia*, that a 1% increase in Internet access offsets, on average, the impact of native languages on trade by about 1.8%, the impact of spoken languages by about 2.9%, and the impact of official

languages by about 0.7%. Therefore, international institutions such as the World Trade Organization should put more emphasis on the expansion of Information Technologies (IT) access among the population in their member countries.

Data Science

Predicting the impact of storms on electric power outages

Mary Frances Dorn & Kimberly Kaufeld
Statistical Sciences Group
Los Alamos National Laboratory

Tropical and winter storms can cause widespread damage to electric distribution networks. These distribution networks are mostly above ground and are exposed to direct damage from severe weather conditions associated with these storms. During winter storms, the combined stress of the weight of ice, the increased wind resistance of the conductors, and broken tree limbs can damage lines, poles, and support structures. Patterns in weather forecasts and records were analyzed to identify winter storm events, which were validated against historical storm records from the National Weather Service. Information from electric power companies were used to develop a model to predict outages when a storm is forecasted. This talk will include a discussion of the many challenges presented by this problem.

Biostatistics Abstracts

Using acceleromets to improve mortality risks estimation in NHANES

Ekaterina Smirnova & Quy Cao
Department of Mathematics
University of Montana

Technological and scientific developments over the past decade have brought new opportunities to collect, store and study heterogeneous sources of data, which can vary in size from very small (a few observations) to very big (Terabytes of data). One example of such data is the National Health Examination (NHANES) study, which contains objectively measured physical activity data collected using hip-worn accelerometers from multiple cohorts. However, using the accelerometry data has proven daunting because: 1) sampling weights need to be carefully adjusted and accounted for in individual analyses; 2) there is a lack of reproducible software that transforms the data from its published format into analytic form; and 3) the high dimensional nature of accelerometry data complicates analyses. The recently developed NHANES data package in R, helps disseminate high quality, processed activity data combined with mortality and demographic information. We illustrate using this package to evaluate the mortality risks via accelerometry measured total physical activity and all-cause mortality risks in models adjusted for other predictors, such as age, gender, race/ethnicity, education, body mass index, and comorbidities. We rank health biomarkers by their ability to predict both all-cause and cardiac mortality, and identify directions for improving currently existing cardiovascular risks estimation algorithms.

Using *Dictyostelium* amoebas to test traveling wave theory

John D. Reeve
Department of Zoology
Southern Illinois University Carbondale

There are a number of ecological phenomena that display traveling waves, including the spread of invasive species, natural enemies, and diseases. An extensive theory has been developed to explain these phenomena, mostly based on reaction-diffusion equations. Quantitative tests of the theory are uncommon, however, because dispersal behavior is difficult to observe in the

field. *Dictyostelium* amoebas and their bacterial prey provide a model system for testing this theory. This system displays traveling waves as the amoebas advance through a bacterial colony, and the dispersal of individual amoebas can often be observed. It is also feasible to quantify population growth rates and other quantities needed to predict wave speed, and manipulate prey density by varying media strength. Preliminary results will be presented for dispersal behavior, growth rates, and wave speed for five amoeba species. The dispersal of individual amoebas shows strong chemotaxis in response to prey, but in many cases is not diffusive. Some alternative dispersal kernels are also fitted to these observations. Dispersal distances and growth rates vary greatly among species, and appear correlated with wave speed.

Dynamic Association Based on Time-Dependent Risk Factors in Longitudinal Studies

Lihui Zhao
Division of Biostatistics/Department of Preventive Medicine
Feinberg School of Medicine/Northwestern University

In multiple large cohort studies, it has been observed that the risk factors levels at remote past could be more predictive to the current cardiovascular risk than the more recent risk factor levels. In general, the predictiveness of historical risk factor levels varies with their measurement time. It has been hypothesized that the early cumulative exposure to suboptimal risk factor levels such as higher blood pressure may cause irreversible organ damage and permanently elevates the long term cardiovascular risk regardless of the subsequent change in the levels of the same risk factors due to various intervention at later stage of the life. Therefore it is important to understand the prospective association between the underlying trajectory of the past risk factor levels and a response variable reflecting one's recent cardiovascular health status. In this paper, we propose a set of statistical models allowing estimating patterns of the trajectory of the risk factor levels for predicting the outcome of interest. In our proposal, the underlying trajectory for a patient is characterized by a function of fixed effects as well as individualized random effects, while the targeted pattern is modeled via a linear functional of the trajectory within a given time window. For the well parametrized functional of the trajectory, we propose asymptotically valid inference procedure. We apply the proposed methods to the data from a cardiovascular disease cohort study.