STATISTICS IN THE DATA SCIENCE ERA:

A Symposium to Celebrate 50 Years of Statistics at the University of Michigan

September 20-21, 2019
Rackham Graduate School
915 E. Washington Street
Ann Arbor, MI 48109

PROGRAM

Friday, September 20, 2019

9:30-10:30 ..... **Coffee and Registration**

10:30-10:50 ..... **Opening Remarks**
Xuming He and Dean Anne Curzan

Keynote Session 1 **Chair: Ambuj Tewari,** *University of Michigan*
10:50-11:50 ..... **Susan Murphy,** *Harvard University*
Online Experimentation with Learning Algorithms in a Clinical Trial

11:50-12:00 ..... **Group Photo**

12:00-1:30 ..... **Lunch and Poster Session**

Invited Session 1 **Chair: Ziwei Zhu,** *University of Michigan*
1:30-2:00 ..... **Sumanta Basu,** *Cornell University*
Large Spectral Density Matrix Estimation by Thresholding

2:00-2:30 ..... **Anindya Bhadra,** *Purdue University*
Horseshoe Regularization for Machine Learning in Complex and Deep Models

2:30-3:00 ..... **Coffee Break**

3:00-4:00 ..... **Academic Panel Discussion**

Keynote Session 2 **Chair: Long Nguyen,** *University of Michigan*
4:00-5:00 ..... **Michael Jordan,** *University of California, Berkeley*
Decisions and Contexts: On Gradient-Based Methods for Finding Game-Theoretic Equilibria

All sessions and panel discussions will be held in the Amphitheater on the 4th floor of the Rackham building. Coffee breaks and lunches will be served in the Assembly Hall across from the Amphitheater. Posters will be displayed in the East and West Conference Rooms on the 4th floor.

Saturday, September 21, 2019

8:30-9:00 ..... **Coffee**

Invited Session 2 **Chair: Snigdha Panigrahi,** *University of Michigan*
9:00-9:30 ..... **Adam Rothman,** *University of Minnesota*
Shrinking Characteristics of Precision Matrix Estimators

9:30-10:00 ..... **Bodhisattva Sen,** *Columbia University*
Multivariate Rank-based Distribution-free Nonparametric Testing using Measure Transportation

10:00-10:30 ..... **Min Qian,** *Columbia University*
Personalized Policy Learning using Longitudinal Mobile Health Data

10:30-11:00 ..... **Coffee Break**

Invited Session 3 **Chair: Kean Ming Tan,** *University of Michigan*
11:00-11:30 ..... **Ali Shojaie,** *University of Washington*
Incorporating Auxiliary Information into Learning Directed Acyclic Graphs

11:30-12:00 ..... **Jing Ma,** *Texas A&M University*
Graphical Models and Differential Networks for Microbiome Data

12:00-12:30 ..... **Eric Laber,** *NC State University*
Sample Size Calculations for SMARTs

12:30-1:30 ..... **Lunch**

1:30-2:30 ..... **Industry Panel Discussion**

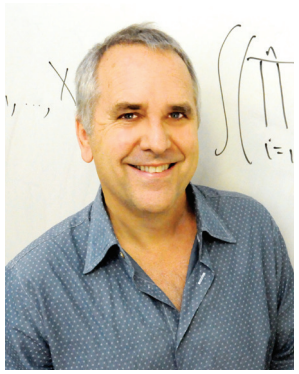
Keynote Session 3 **Chair: Kerby Shedden,** *University of Michigan*
2:30-3:30 ..... **Jeff Wu,** *Georgia Institute of Technology*
Navier-Stokes, Spatial-Temporal Kriging and Combustion Stability: A Prominent Example of Physics-based Analytics

3:30-4:00 ..... **Poster Awards and Concluding Remarks**

PROGRAM

KEYNOTE SPEAKERS

Michael I. Jordan
University of California, Berkeley



Michael I. Jordan is the Pehong Chen Distinguished Professor in the Department of Electrical Engineering and Computer Science and the Department of Statistics at the University of California, Berkeley. His research interests bridge the computational, statistical, cognitive and biological sciences. Prof. Jordan is a member of the National Academy of Sciences and a member of the National Academy of Engineering. He has been named a Neyman Lecturer and a Medallion Lecturer by the Institute of Mathematical Statistics, and he has given a Plenary Lecture at the International Congress of Mathematicians. He received the IJCAI Research Excellence Award in 2016, the David E. Rumelhart Prize in 2015 and the ACM/AAAI Allen Newell Award in 2009.

Decisions and Contexts: On Gradient-Based Methods for Finding Game-Theoretic Equilibria

Statistical decisions are often given meaning in the context of other decisions, particularly when there are scarce resources to be shared. We aim to blend gradient-based methodology with game-theoretic goals, as part of a larger "microeconomics meets statistics" programme. I'll discuss several recent results: (1) a gradient-based algorithm that finds Nash equilibria, and only Nash equilibria; (2) how to define local optimality in nonconvex-nonconcave minimax optimization, and how such a definition relates to stochastic gradient methods; and (3) exploration-exploitation tradeoffs for bandits that compete over a scarce resource.

Susan Murphy
Harvard University

Susan Murphy is a Professor of Statistics, Computer Science and Radcliffe Alumnae Professor at Harvard University. Her lab develops data analysis methods and experimental designs to improve real time sequential decision-making in mobile health. In particular, her lab develops algorithms, deployed on wearable devices, to deliver and continually optimize individually tailored treatments. Dr. Murphy is a member of the National Academy of Sciences and of the National Academy of Medicine, both of the US National Academies. In 2013 she was awarded a MacArthur Fellowship for her work on experimental designs to inform sequential decision making.

Online Experimentation with Learning Algorithms in a Clinical Trial

In sequential decision making we aim to learn when and in which context it is best to deliver treatments. Consider mobile health in which behavioral treatments are given to individuals as they go about their daily life. Operationally designing the sequential treatments involves the construction of decision rules that input current context of an individual and output a recommended treatment. There is much interest in personalizing the decision rules, particularly in real time as the individual experiences sequences of treatment. Here we discuss our work in designing and implementing variants of online "bandit" learning algorithms for use in personalizing mobile health interventions.

KEYNOTE SPEAKERS

Jeff Wu

Georgia Institute of Technology

C. F. Jeff Wu is Professor and Coca Cola Chair in Engineering Statistics at the School of Industrial and Systems Engineering, Georgia Institute of Technology. He was the first academic statistician elected to the National Academy of Engineering (2004); also a Member (Academician) of Academia Sinica (2000). A Fellow of the American Society for Quality, Institute of Mathematical Statistics, of INFORMS, and the American Statistical Association. He received the COPSS (Committee of Presidents of Statistical Societies) Presidents' Award in 1987, the COPSS Fisher Lecture Award in 2011, the Deming Lecture Award in 2012, the inaugural Akaike Memorial Lecture Award in 2016, the George Box Medal from ENBIS in 2017, and numerous other awards and honors. He has published more than 180 research articles and supervised 49 Ph.D.s. Among his students, there are 21 Fellows of ASA, IMS, ASQ, IAQ and IIE, and three editors of Technometrics. He has published two books "Experiments: Planning, Analysis, and Parameter Design Optimization" (with Hamada) and "A Modern Theory of Factorial Designs" (with Mukerjee). He coined the term "data science" in 1998.

Navier-Stokes, Spatial-Temporal Kriging and Combustion Stability: A Prominent Example of Physics-based Analytics

Most "learning" in big data is driven by the data alone. Some people may believe this is sufficient because of the sheer data size. If the physical world is involved, this approach is often insufficient. In this talk I will give a recent study to illustrate how physics and data are used jointly to learn about the "truth" of the physical world. It also serves as an example of physics-based analytics, which in itself has many forms and meanings. In an attempt to understand the turbulence behavior of an injector, a new design methodology is needed which combines engineering physics, computer simulations and statistical modeling. There are two key challenges: the simulation of high-fidelity spatial-temporal flows (using the Navier-Stokes equations) is computationally expensive, and the analysis and modeling of this data requires physical insights and statistical tools. A surrogate model is presented for efficient flow prediction in injectors with varying geometries, devices commonly used in many engineering applications. The novelty lies in incorporating properties of the fluid flow as simplifying model assumptions, which allows for quick emulation in practical turnaround times, and also reveals interesting flow physics which can guide further investigations.

Sumanta Basu, *Cornell University*

Large Spectral Density Matrix Estimation by Thresholding

Spectral density matrix estimation of multivariate time series is a classical problem in time series and signal processing. In modern neuroscience, spectral density based metrics are commonly used for analyzing functional connectivity among brain regions. In this paper, we develop a non-asymptotic theory for regularized estimation of high-dimensional spectral density matrices of Gaussian and linear processes using thresholded versions of averaged periodograms. Our theoretical analysis ensures that consistent estimation of spectral density matrix of a p-dimensional time series using n samples is possible under high-dimensional regime log(p)=o(n) as long as the true spectral density is approximately sparse. A key technical component of our analysis is a new concentration inequality of average periodogram around its expectation, which is of independent interest. Our estimation consistency results complement existing results for shrinkage based estimators of multivariate spectral density, which require no assumption on sparsity but only ensure consistent estimation in a regime p^2=o(n). In addition, our proposed thresholding based estimators perform consistent and automatic edge selection when learning coherence networks among the components of a multivariate time series. We demonstrate the advantage of our estimators using simulation studies and a real data application on functional connectivity analysis with fMRI data.

Anindya Bhadra, *Purdue University*

Horseshoe Regularization for Machine Learning in Complex and Deep Models

Since the advent of the horseshoe priors for regularization, global-local shrinkage methods have proved to be a fertile ground for the development of Bayesian theory and methodology in machine learning. They have achieved remarkable success in computation, and enjoy strong theoretical support. Much of the existing literature has focused on the linear Gaussian case. The purpose of the current talk is to demonstrate that the horseshoe priors are useful more broadly, by reviewing both methodological and computational developments in complex models that are more relevant to machine learning applications. Specifically, we focus on methodological challenges in horseshoe regularization in nonlinear and non-Gaussian models; multivariate models; and deep neural networks. We also outline the recent computational developments in horseshoe shrinkage for complex models along with a list of available software implementations that allows one to venture out beyond the comfort zone of the canonical linear regression problems.

Eric Laber, *NC State University*

Sample Size Calculations for SMARTs

Sequential Multiple Assignment Randomized Trials (SMARTs) are considered the gold standard for estimation and evaluation of treatment regimes. SMARTs are typically sized to ensure sufficient power for a simple comparison, e.g., the comparison of two fixed and non-overlapping treatment sequences. Estimation of an optimal treatment regime is conducted as part of a secondary and hypothesis-generating analysis with formal evaluation of the estimated optimal regime deferred to a follow-up trial. However, running a follow-up trial to evaluate an estimated optimal treatment regime is costly and time-consuming; furthermore, the estimated optimal regime that is to be evaluated in such a follow-up trial may be far from optimal if the original trial was underpowered for estimation of an optimal regime. We derive sample size procedures for a SMART that ensure: (i) sufficient power for comparing the optimal treatment regime with standard of care; and (ii) the estimated optimal regime is within a given tolerance of the true optimal regime with high-probability. We establish asymptotic validity of the proposed procedures and demonstrate their finite sample performance in a series of simulation experiments.

Jing Ma, *Texas A&M University*

Graphical Models and Differential Networks for Microbiome Data

Microorganisms such as bacteria form complex ecological community networks with various interactions. Diet and other environmental factors can greatly impact the composition and structure of these microbial communities. Differential analysis of microbial community networks aims to elucidate such systematic changes during an adaptive response to changes in environment. In this talk, I will present a flexible Markov random field model for microbial network structure and introduce a hypothesis testing framework for detecting the differences between networks, also known as differential network biology. The proposed global test for differential networks is particularly powerful against sparse alternatives. In addition, I will discuss a multiple testing procedure with false discovery rate control to identify the structure of the differential network. The proposed method is applied to a gut microbiome study on UK twins to evaluate how age affects the microbial community network.

Min Qian, *Columbia University*

Personalized Policy Learning using Longitudinal Mobile Health Data

We address the personalized policy learning problem using longitudinal mobile health application usage data. Personalized policy represents a paradigm shift from developing a single policy that may prescribe personalized decisions by tailoring. Specifically, we aim to develop the best policy, one per user, based on estimating random effects under generalized linear mixed model. With many random effects, we consider new estimation method and penalized objective to circumvent high-dimension integrals for marginal likelihood approximation. We establish consistency and optimality of our method with endogenous app usage, and apply the method to develop personalized push ("prompt") schedules in a mobile health application. The proposed method compares favorably to existing estimation methods including using the R function "glmer" in a simulation study.

Adam Rothman, *University of Minnesota*

Shrinking Characteristics of Precision Matrix Estimators

We propose a framework to shrink a user-specified characteristic of a precision matrix estimator that is needed to fit a predictive model. Estimators in our framework minimize the Gaussian negative loglikelihood plus an L1 penalty on a linear function evaluated at the optimization variable corresponding to the precision matrix. We establish convergence rate bounds for these estimators and propose an alternating direction method of multipliers algorithm for their computation. Our simulation studies show that our estimators can perform better than competitors when they are used to fit predictive models. In particular, we illustrate cases where our precision matrix estimators perform worse at estimating the population precision matrix but better at prediction.

INVITED SPEAKERS

INVITED SPEAKERS

Bodhisattva Sen, *Columbia University*

Multivariate Rank-based Distribution-free Nonparametric Testing using Measure Transportation

In this talk we propose a general framework for distribution-free nonparametric testing in multi-dimensions, based on a notion of multivariate ranks which are defined using some recent advances in the theory of measure transportation. Unlike other existing proposals in the literature, these multivariate ranks share a number of useful properties with the usual notion of one-dimensional ranks; most importantly, these ranks are distribution-free. This crucial observation allows us to design nonparametric tests which are based on statistics that are exactly distribution-free under the null hypothesis. We illustrate the applicability of this approach by constructing exact distribution-free tests for two classical nonparametric problems: (i) testing for mutual independence between random vectors, and, (ii) testing for the equality of multivariate distributions. In both these problems we derive the asymptotic null distribution of the proposed statistic. We further show that our tests are consistent against very general alternatives. Moreover, the proposed tests are tuning-free, computationally feasible and are well-defined under minimal assumptions on the underlying distributions (e.g., they do not need any moment assumptions). We also demonstrate the efficacy of these procedure using extensive simulations. In the process of analyzing the theoretical properties of our procedures, we end up proving some new results in the theory of measure transportation and limit theory of permutation statistics using Stein's method for exchangeable pairs, which may be of independent interest.

Ali Shojaie, *University of Washington*

Incorporating Auxiliary Information into Learning Directed Acyclic Graphs

Directed acyclic graphs (DAGs) are commonly used to represent causal relationships in complex social and biological systems. As such, learning DAGs from observational data is a fundamental problem in statistics and machine learning. While learning DAGs with no external information is an NP-hard problem, the problem becomes somewhat trivial if a causal ordering of nodes is available. However, complete causal orderings are rarely available in practice and background scientific knowledge may only provide partially informative constraints on the model parameters. In this talk we examine the extent to which partially informative, and hence more realistic, constraints can lead to improved computational and sample complexity when learning high-dimensional DAGs.

PANEL MEMBERS

Academic Panel

David Hunter, *Pennsylvania State University*

George Michailidis, *University of Florida*

Susan Murphy, *Harvard University*

Naveen Narisetty, *University of Illinois at Urbana-Champaign*

Min Qian, *Columbia University*

Jeff Wu, *Georgia Institute of Technology*

Liza Levina, *University of Michigan* (Moderator)

Industry Panel

Debjit Biswas, *Pfizer*

William Brenneman, *Proctor & Gamble*

David Mease, *Google*

Harsh Singhal, *Wells Fargo*

Yang Yang, *LinkedIn*

Bobby Yuen, *Liberty Mutual*

Ji Zhu, *University of Michigan* (Moderator)

POSTER SESSION

Joseph Dickens, *University of Michigan*
Calibrating Tests of the Indirect Meditation Effect

Robyn Ferg, *University of Michigan*
Tracking Presidential Approval with Twitter

Mohamad Kazem Shirani Faradonbeh, *University of Florida*
Learning and Stabilizing Linear Time Series with Exogenous Inputs

Zheng Gao, *University of Michigan*
Laws of Large Dimensions

Jack Goetz, *University of Michigan*
Active Learning for Nonparametric Regression using Purely Random Trees

Yuqi Gu, *University of Michigan*
Learning Attribute Patterns in High-Dimensional Structured Latent Attribute Models

Aritra Guha, *University of Michigan*
On Posterior Contraction of Parameters in Bayesian Mixture Modeling

Yinqiu He, *University of Michigan*
Asymptotically Independent U-Statistics in High-Dimensional Testing

Young Jung, *University of Michigan*
Regret Bounds for Thompson Sampling in Restless Bandit Problems

Natallia V. Katenka, *University of Rhode Island*
Evaluating the Effects of Attitudes on Health-Seeking Behavior among a Network of People who Inject Drugs

Daniel Kessler, *University of Michigan*
Prediction from Networks with Node Features with Application to Neuroimaging

Baekjin Kim, *University of Michigan*
On the Optimality of Perturbation in both Stochastic and Adversarial Multi-armed Bandit

Michael Law, *University of Michigan*
Linear Mixed Models in High Dimensions

Rayleigh Lei, *University of Michigan*
Modeling Simplex Transformations

Keith Levin, *University of Michigan*
Bootstrapping Networks with Latent Space Structure

Tianxi Li, *University of Virginia*
Network Regression with Inference

Asad Lodhia, *University of Michigan*
Harmonic Means of Wishart Random Matrices

Brook Luers, *University of Michigan*
Mixed Effects Models for Sequential, Multiple Assignment Randomized Trials (SMARTs)

Chenchen Ma, *University of Michigan*
Learning Latent Hierarchical Structures in Latent Attribute Models

Laura Niss, *University of Michigan*
Debiasing Representations by Removing Unwanted Variation due to Protected Attributes

Weijing Tang, *University of Michigan*
SODEN: A Scalable Continuous-time Survival Model through Ordinary Differential Equation Networks

Robert Trangucci, *University of Michigan*
Modeling the Spatial Risk of Diarrheal Illness in Mezquital Valley, Mexico

Edward Wu, *University of Michigan*
The P-LOOP Estimator: Covariate Adjustment for Paired Experiments

Ziping Xu, *University of Michigan*
Perturbation Algorithm on Reinforcement Learning

Drew Yarger, *University of Michigan*
A Functional Data Approach to the Argo Project

Xuefei Zhang, *University of Michigan*
Estimating Joint Latent Space Models for Network Data with High-Dimensional Multivariate Node Variables

Yunpeng Zhao, *Arizona State University*
Network Inference from Temporal-Dependent Grouped Observations

POSTER SESSION

PRESENTED BY...

LSA STATISTICS
UNIVERSITY OF MICHIGAN

The 50th Anniversary Committee

Tailen Hsing
Michael Woodroofe Collegiate Professor of Statistics
Chair of the 50th Anniversary Committee

Nicholle Cardinal
Communications Coordinator

Xuming He
Department Chair
H.C. Carver Professor of Statistics

Liza Levina
Vijay Nair Collegiate Professor of Statistics

Judy McDonald
Executive Assistant

Ambuj Tewari
Associate Professor of Statistics

Ji Zhu
Professor of Statistics

Bebe Zuniga-Valentino
Department Administrator

The Department of Statistics would like to thank the generous sponsors who have made our 50th Anniversary celebration possible!

WELLS FARGO

SCHOOL OF PUBLIC HEALTH
BIOSTATISTICS
UNIVERSITY OF MICHIGAN

LSA PSYCHOLOGY
UNIVERSITY OF MICHIGAN

INSTITUTE FOR
SOCIAL RESEARCH
UNIVERSITY OF MICHIGAN

PRESENTED BY...

THANK YOU!

DID YOU KNOW?

The Department of Statistics is on social media! Follow us on Facebook, Twitter, and Instagram to stay up-to-date on the latest from the department.

#UMStats50

Stat-UM@umich.edu
http://www.lsa.umich.edu/stats/