

Statistics Seminar for Fall 2019 and Spring 2020

Schedule and Abstract

Date	Speaker	Talk Title
October 4, 11:00am- 12:00noon, Fretwell 315	Zhihua Su, University of Florida (Hosted by Dr. Yanqing Sun)	Envelope Models and Methods
October 18, 11:00am- 12:00noon, Fretwell 315	Shanshan Zhao, Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences (Hosted by Yang Li)	Accommodating Limit-of-Detection in Environmental Mixture Analysis
October 25, 11:00am- 12:00noon, Fretwell 315	Yanqing Wang, Georgia State University (Hosted by Yinghao Pan)	Learning-Based Biomarker-Assisted Rules for Optimized Clinical Benefit under A Risk-Constraint
November 8, 11:00am- 12:00noon, Fretwell 315	Zhezhen Jin, Columbia University (Hosted by Jiancheng Jiang)	Variance Estimation in Semiparametric Regression Models
November 22, 11:00am- 12:00noon, Fretwell 315	Kevin Lee, Western Michigan University (Hosted by Jun Song)	Clustering Time-Evolving Networks through Temporal ERGMs
February 7, 11:10am- 12:10pm, Fretwell 305	Eliana Christou, UNC Charlotte (Faculty)	Central Quantile Subspace
February 21, 11:10am- 12:10pm, Fretwell 305	Mingyao Li, University of Pennsylvania (Hosted by Shaoyu Li)	Translation of single-cell genomics into human health: methods and applications
TBA 11:10am- 12:10pm, Fretwell 305	Shaoyu Li	TBA

Speaker #1:**Date:** October 4, 2019**Time and location:** 11:00am-12:00noon, Fretwell 315**Speaker:** Prof. Zhihua Su, University of Florida**Title:** Envelope Models and Methods

Abstract: This talk presents a new statistical concept called an envelope. An envelope has the potential to achieve substantial efficiency gains in multivariate analysis by identifying and cleaning up immaterial information in the data. The efficiency gains will be demonstrated both by theory and example. Some recent developments in this area, including partial envelopes and heteroscedastic envelope models, will also be discussed. They refine and extend the enveloping idea, adapting it to more data types and increasing the potential to achieve efficiency gains. Applications of envelopes and their connection to other fields will also be mentioned. (Hosted by Dr. Yanqing Sun, UNC Charlotte)

Speaker #2:**Date:** October 18, 2019**Time and location:** 11:00am-12:00noon, Fretwell 315**Speaker:** Shanshan Zhao, Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences (Hosted by Yang Li)**Title:** Accommodating limit-of-detection in environmental mixture analysis

Abstract: Humans are exposed to a multitude of environmental toxicants daily, and there is a great interest in developing statistical methods for assessing the effects of environmental mixtures on various health outcomes. One difficulty is that multiple chemicals in the mixture can be subject to left-censoring due to varying limits of detection (LOD). Conventional approaches either ignore these measures, dichotomize them at the limits, or replace them with arbitrary values such as LOD/Ö2. Methods have been proposed to handle a single biomarker with limit of detection in such setting, by joint modeling the left-censored biomarker measure with an AFT model and the disease outcome with a generalized linear model. We extend this method to handle multiple correlated biomarkers subject to LOD, through a newly proposed nonparametric estimator of the multivariate survival function and innovative computational approaches. We apply the proposed method to the LIFECODES birth cohort to elucidate the relationship between maternal urinary trace metals and oxidative stress markers.

Shanshan Zhao's website:

<https://www.niehs.nih.gov/research/atniehs/labs/bb/staff/zhao/index.cfm>**Speaker #3:****Date:** October 25, 2019**Time and location:** 11:00am-12:00noon, Fretwell 315**Speaker:** Prof. Yanqing Wang, Georgia State University (Hosted by Yinghao Pan)

Title: Learning-Based Biomarker-Assisted Rules for Optimized Clinical Benefit under A Risk-Constraint

Abstract: Novel biomarkers, in combination with currently available clinical information, have been sought to improve clinical decision making in many branches of medicine, including screening, surveillance, and prognosis. Statistical methods are needed to integrate such diverse information to develop targeted interventions that balance benefit and harm. In the specific setting of disease detection, we propose novel approaches to construct a multiple-marker-based decision rule by directly optimizing a benefit function, while controlling harm at a maximally tolerable level. The new approaches include plug-in and direct-optimization-based algorithms, and they allow for the construction of both nonparametric and parametric rules. A study of asymptotic properties of the proposed estimators is provided. Simulation results demonstrate good clinical utilities for the resulting decision rules under various scenarios. The methods are applied to a biomarker study in prostate cancer surveillance.

Speaker #4:

Date: November 8, 2019

Time and location: 11:00am-12:00noon, Fretwell 315

Speaker: Prof. Zhezhen Jin, Columbia University (Hosted by Jiancheng Jiang)

Title: Variance estimation in semiparametric regression models

Abstract: In semiparametric regression analysis, objective functions and estimating functions for regression parameters are often nonsmooth and non-monotone, which results in difficulty in the corresponding variance estimation. I will discuss the issues and present available and newly developed methods with general theory, implementation and demonstrate the methods with examples.

Speaker #5:

Date: November 22, 2019

Time and location: 11:00am-12:00noon, Fretwell 315

Speaker: Prof. Kevin Lee, Western Michigan University (Hosted by Jun Song)

Title: Clustering Time-Evolving Networks through Temporal ERGMs

Abstract: Model-based clustering of dynamic networks has emerged as one of the increasingly important research topics in statistical network analysis. We present a statistical clustering framework through the temporal exponential-family random graph models (ERGMs). The temporal ERGMs allow the specification of interesting network features (e.g., stability), and the hidden Markov structure allows the inference of the dynamic latent block structure. In this talk, we will introduce two models one with static latent block structure and the other with dynamic latent block structure. Furthermore, we will discuss a variational expectation-maximization algorithm to solve the approximate maximum likelihood estimation. The performance of our proposed methods is demonstrated through an empirical application to the international trade networks.

Speaker #6:**Date:** February 7, 2020**Time and location:** 11:10am-12:10pm, Fretwell 305**Speaker:** Prof. Eliana Christou, UNC Charlotte (Faculty)**Title:** Central Quantile Subspace

Abstract: Quantile regression (QR) is becoming increasingly popular due to its relevance in many scientific investigations. There is a great amount of work about linear and nonlinear QR models. Specifically, nonparametric estimation of the conditional quantiles received particular attention, due to its model flexibility. However, nonparametric QR techniques are limited in the number of covariates. Dimension reduction offers a solution to this problem by considering low-dimensional smoothing without specifying any parametric or nonparametric regression relation. Existing dimension reduction techniques focus on the entire conditional distribution. We, on the other hand, turn our attention to dimension reduction techniques for conditional quantiles and introduce a new method for reducing the dimension of the predictor X . The novelty is threefold. We start by considering a single index quantile regression model, which assumes that the conditional quantile depends on X through a single linear combination of the predictors, then extend to a multi index quantile regression model, and finally, generalize the proposed methodology to any statistical functional of the conditional distribution. The results suggest that this method has a good finite sample performance and often outperforms existing methods.

Speaker #7:**Date:** February 21, 2020**Time and location:** 11:10am-12:10pm, Fretwell 305**Speaker:** Mingyao Li, Ph.D, Department of Biostatistics, Epidemiology & Informatics, University of Pennsylvania Perelman School of Medicine**Title:** Translation of single-cell genomics into human health: methods and applications

Abstract: Recent technological breakthroughs have made it possible to measure gene expression at the single-cell level, thus allowing biologists and clinicians to better understand cellular heterogeneity and modify cell behavior through targeted molecular therapies. However, single-cell RNA sequencing protocols are complex. Even with the most sensitive platforms, the data are often noisy owing to a high frequency of dropout events, and the phenomenon of transcriptional bursting in which pulses of transcriptional activity are followed by inactive refractory periods. In this talk, I will present several statistical and machine learning methods that aim to tackle these challenges for a better understanding of cellular heterogeneity. I will illustrate our methods by showing results from ongoing collaborations on age-related macular degeneration and Alzheimer's disease. With the growing interest in utilizing single-cell RNA sequencing in biomedical research, our methods will aid biomedical researchers to answer medically related questions and make exciting discoveries.