

## Statistics Seminar for Fall 2018

### Schedule and Abstract

Date	Speaker	Talk Title
September 21	Yinghao Pan, UNC Charlotte	<b>Improved doubly robust estimation in learning individualized treatment rules</b>
September 28	Yu Shen, The University of Texas MD Anderson Cancer Center	<b>Density Ratio Model for Analyzing Length-biased Data, Division of Quantitative Sciences</b>
October 5	Ram Tiwari, FDA	<b>(Part 1) Job/career opportunities at FDA, (Part 2) Bayesian approaches for benefit-risk assessment with examples</b>
October 12	Mei-Ling Ting Lee, University of Maryland	<b>First-hitting-time Based Threshold Regressions with Shocks for Degradation Processes</b>
October 26	Yang Feng, Columbia University	<b>Neyman-Pearson classification algorithms and NP receiver operating characteristics</b>
November 9	Hongyuan Cao, Florida State University	<b>Regression analysis of longitudinal data with omitted asynchronous longitudinal covariate</b>
November 16	Yifan Cui, University of Pennsylvania	<b>Tree-based Survival Models and Precision Medicine</b>
November 30	Fei Gao, University of Washington	<b>Non-iterative Estimation Update for Parametric and Semiparametric Models with Population-based Auxiliary Information</b>
December 7	Final Exams	

**Speaker #1:****Date:** September 21, 2018**Time and location:** 11:00am-12:00noon, Fretwell 315**Speaker:** Prof. Yinghao Pan, UNC Charlotte**Title:** Improved doubly robust estimation in learning individualized treatment rules

**Abstract:** Due to patient's heterogeneous response to treatment, there is a growing interest in developing novel and efficient statistical methods in estimating individualized treatment rules (ITRs). The central idea is to recommend treatment according to patient characteristics, and the optimal ITR is the one that maximizes the expected clinical outcome if followed by the patient population. We propose an improved estimator of the optimal ITR that enjoys two key properties. First, it is doubly robust, meaning that the proposed estimator is consistent if either the propensity score or the outcome model is correct. Second, it achieves the smallest variance among its class of doubly robust estimators when the propensity score model is correctly specified, regardless of the specification of the outcome model. Simulation studies show that the estimated optimal ITR obtained from our method yields better clinical outcome than its main competitors. Data from Sequenced Treatment Alternatives to Relieve Depression (STAR\*D) study is analyzed as an illustrative example.

(Hosted by Dr. Yinghao Pan, UNC Charlotte)

**Speaker #2****Date:** September 28<sup>th</sup>, 2018**Time and location:** 11:00am-12:00noon, Fretwell 315**Speaker:** Professor Yu Shen, Division of Quantitative Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX**Title:** Density Ratio Model for Analyzing Length-biased Data**Abstract:**

For analyzing complex length-biased failure time data, computational algorithms and theoretical properties are not readily available, especially when a likelihood function involves infinite-dimensional parameters. Relying on the invariance property of length-biased failure time data under the semiparametric density ratio model, we present two likelihood approaches for the estimation and assessment of the difference between two survival distributions. The most efficient maximum likelihood estimators are obtained by the EM algorithm and profile likelihood. We also provide a simple numerical method for estimation and inference based on conditional likelihood, which can be generalized to k-arm settings. Unlike conventional survival data, the mean of the population failure times can be consistently estimated given right-censored length-biased data under mild regularity conditions. To check the semiparametric density ratio model assumption, we use a test statistic based on the area between two survival distributions.

(Hosted by Dr. Yanqing Sun, UNC Charlotte)

**Speaker #3:**

**Date:** October 5, 2018

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

**Speaker:** Ram Tiwari, Director, Division of Biostatistics, Center for Devices and Radiological Health, Office Surveillance and Biometrics, U.S. Food & Drug Administration

**Title:** (Part 1) Job/career opportunities at FDA, (Part 2) Bayesian approaches for benefit-risk assessment with examples

**Abstract:** An important aspect of the drug/device evaluation process is to have an integrated benefit-risk assessment to determine, using some quantitative measures, whether the benefit outweighs the risk for the target population. The subject-level benefit-risk response is a five-category random variable with cell counts following a multinomial distribution. Assuming that the cell probabilities follow a Dirichlet distribution, we develop a Bayesian approach for the longitudinal assessment of benefit-risk using the global measures proposed by Chuang-Stein et al. In a more generalized approach, a power prior is used through the likelihood function to discount the information from previous visits. For the subject-level benefit-risk assessment, the cell-probability of the subject, with respect to a reference category, is modeled, on the logarithmic scale, as a generalized linear model using a Dirichlet process as a prior. The model is applied to drug/device clinical trial datasets.

(Hosted by Dr. Yanqing Sun, UNC Charlotte)

**Speaker #4:**

**Date:** October 12, 2018

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

**Speaker:** Prof. Mei-Ling Ting Lee, University of Maryland

**Title:** First-hitting-time Based Threshold Regressions with Shocks for Degradation Processes

**Abstract:** People's health or engineering systems experience gradual degradation while simultaneously being exposed to a stream of random shocks of varying magnitudes that eventually cause death or failure when a shock exceeds the residual strength. I'll present theory and statistical properties of this model. Applications including a study of osteoporotic hip fractures in the elderly.

(Hosted by Dr. Yang Li, UNC Charlotte)

**Speaker #5:****Date:** October 26, 2018**Time and location:** 11:00am-12:00noon, (Fretwell 106)**Speaker:** Dr. Yang Feng, Columbia University**Title:** Neyman-Pearson classification algorithms and NP receiver operating characteristics

**Abstract:** In many binary classification applications, such as disease diagnosis and spam detection, practitioners commonly face the need to limit type I error (that is, the conditional probability of misclassifying a class 0 observation as class 1) so that it remains below the desired threshold. To address this need, the Neyman-Pearson (NP) classification paradigm is a natural choice; it minimizes type II error (that is, the conditional probability of misclassifying a class 1 observation as class 0) while enforcing an upper bound,  $\alpha$ , on the type I error. Despite its century-long history in hypothesis testing, the NP paradigm has not been well recognized and implemented in classification schemes. Common practices that directly limit the empirical type I error to no more than  $\alpha$  do not satisfy the type I error control objective because the resulting classifiers are likely to have type I errors much larger than  $\alpha$ , and the NP paradigm has not been properly implemented in practice. We develop the first umbrella algorithm that implements the NP paradigm for all scoring-type classification methods, such as logistic regression, support vector machines, and random forests. Powered by this algorithm, we propose a novel graphical tool for NP classification methods: NP receiver operating characteristic (NP-ROC) bands motivated by the popular ROC curves. NP-ROC bands will help choose  $\alpha$  in a data-adaptive way and compare different NP classifiers. We demonstrate the use and properties of the NP umbrella algorithm and NP-ROC bands, available in the R package nproc, through simulation and real data studies.

(Hosted by Dr. Jiancheng Jiang, UNC Charlotte)

**Speaker #6:****Date:** November 9, 2018**Time and location:** 11:00am-12:00noon, Fretwell 315**Speaker:** Dr. Hongyuan Cao, Florida State University**Title:** Regression analysis of longitudinal data with omitted asynchronous longitudinal covariate

**Abstract:** Long term follow-up with longitudinal data is common in many medical investigations. In such studies, some longitudinal covariate can be omitted for various reasons. Naïve approach that simply ignores the omitted longitudinal covariate can lead to biased estimators. In this article, we propose new unbiased estimation methods to accommodate omitted longitudinal covariate. In addition, if the omitted longitudinal covariate is asynchronous with the longitudinal response, a two stage approach is proposed for valid statistical inference. Asymptotic properties of the proposed estimators are established. Extensive simulation studies provide numerical support for the theoretical findings. We illustrate the performance of our method on dataset from an HIV study.

(Hosted by Dr. Shaoyu Li, UNC Charlotte)

**Speaker #7:****Date:** November 16, 2018**Time and location:** 11:00am-12:00noon, Fretwell 315**Speaker:** Dr. Yifan Cui, University of Pennsylvania**Title:** Tree-based Survival Models and Precision Medicine

**Abstract:** In the first part, we develop a theoretical framework for survival tree and forest models. We first investigate the method from the aspect of splitting rules. We show that existing approaches lead to a potentially biased estimation of the within-node survival and cause non-optimal selection of the splitting rules. Based on this observation, we develop an adaptive concentration bound result which quantifies the variance component for survival forest models. Furthermore, we show with two specific examples how these concentration bounds, combined with properly designed splitting rules, yield consistency results. In the second part, we focus on one application of survival trees in precision medicine which estimates individualized treatment rules nonparametrically under right censoring. We extend the outcome weighted learning to right censored data without requiring either inverse probability of censoring weighting or semi-parametric modeling of the censoring and failure times. To accomplish this, we take advantage of the tree-based approach to nonparametrically impute the survival time in two different ways. In simulation studies, our estimators demonstrate improved performance compared to existing methods. We also illustrate the proposed method on a phase III clinical trial of non-small cell lung cancer.

(Hosted by Dr. Yinghao Pan, UNC Charlotte)

**Speaker #8:****Date:** November 30, 2018**Time and location:** 11:00am-12:00noon, Fretwell 315**Speaker:** Dr. Fei Gao, University of Washington**Title:** Non-iterative Estimation Update for Parametric and Semiparametric Models with Population-based Auxiliary Information

**Abstract:** With the advancement in disease registries and surveillance data, population-based information on disease incidence, survival or other important biological variables become increasingly available. Such information can be leveraged in studies that collect detailed measurements but from smaller samples. In contrast to recent proposals that formulate the additional information as constraints in optimization problems, we develop a general framework to construct simple estimators that update usual regression estimators with some functionals of data and the initial estimator based on the additional information. We consider general settings which include nuisance parameters in the auxiliary information, non-i.i.d. data such as case-control sampling, and semi-parametric models with infinite dimensional parameters. Detailed examples of several important data and sampling settings are provided.

(Hosted by Dr. Qinging Zhou, UNC Charlotte)