# Statistics Seminar for Fall 2017

## Schedule and Abstract

| Date | Speaker | Talk Title |
|---|---|---|
| September 8 | Qingning Zhou | **Outcome-dependent sampling with interval-censored failure time data,** UNC Charlotte |
| September 22 | Jun Song | **Linear and Nonlinear Sufficient Dimension Reduction for Functional Data,** UNC Charlotte |
| September 29 | Yanqing Sun | **Analysis of Stratified Mark-Specific Proportional Hazards Models under Two-Phase Sampling with Application to HIV Vaccine Efficacy Trials,** UNC Charlotte |
| October 13 | Akim Adekpedjou | **Semiparametric Estimation with Spatially Correlated Recurrent Events,** University of Missouri- Science & Technology, Hosted by Yanqing Sun |
| October 20 | Weizhen Wang | **On Exact Inferences Using Binary Data in Two or Multi-stage Designs**, Wright State University, Hosted by Jiancheng Jiang |
| October 27 | Ling Ma | **Joint modeling of longitudinal functional feature and discrete time-to-event**, Clemson University, Hosted by Qingning Zhou |
| November 3 | Yanyuan Ma | **On Estimation of General Index Model for Survival Data,** Penn State University, Hosted by Yanqing Sun |
| November 10 | Won Chang | **Calibrating an ice sheet model using high-dimensional binary spatial data,** University of Cincinnati, Hosted by Jun Song |
| December 1 | Ram Tiwari | Division of Biostatistics, Center for Devices and Radiological Health, Office Surveillance and Biometrics, U.S. Food & Drug Administration, Hosted by Yanqing Sun |

**Speaker #1**
**Date:** September 8th, 2017
**Time and location:** 11:00am-12:00noon, Fretwell 315
**Speaker:** Qingning Zhou, Department of Mathematics and Statistics, UNCCharlotte

**Title: Outcome-dependent sampling with interval-censored failure time data**

**Abstract:** Epidemiologic studies and disease prevention trials often seek to relate an exposure variable to a failure time that suffers from interval-censoring. When the failure rate is low and the time intervals are wide, a large cohort is often required so as to yield reliable precision on the exposure-failure-time relationship. However, large cohort studies with simple random sampling could be prohibitive for investigators with a limited budget, especially when the exposure variables are expensive to obtain. Alternative cost-effective sampling designs and inference procedures are therefore desirable. We propose an outcome-dependent sampling (ODS) design with interval-censored failure time data, where we enrich the observed sample by selectively including certain more informative failure subjects like the case-cohort design. We develop a novel sieve semiparametric maximum empirical likelihood approach for fitting the proportional hazards model to data from the proposed interval-censoring ODS design. This approach employs the empirical likelihood and sieve methods to deal with the infinite-dimensional nuisance parameters, which greatly reduces the dimensionality of the estimation problem and eases the computation difficulty. The consistency and asymptotic normality of the resulting regression parameter estimator are established. The results from our extensive simulation study show that the proposed design and method works well for practical situations and is more efficient than the alternative designs and competing approaches. An example from the Atherosclerosis Risk in Communities (ARIC) study is provided for illustration.

**Speaker #2:**
**Date:** September 22, 2017
**Time and location:** 11:00am-12:00noon, Fretwell 315
**Speaker: Jun Song,** Department of Mathematics and Statistics, UNCCharlotte

**Title: Linear and Nonlinear Sufficient Dimension Reduction for Functional Data**

**Abstract:** When you sign your name on a touch-pad device, your handwriting is recorded as a parameterized curve along with the time. GPS sensors can record a person's movement in three-dimensional space as a function of time. These types of data can be considered as a realization of a random multivariate function. They are increasingly prevalent in modern applications and their analysis intrinsically require probability model over functional space of infinite dimensions.
In this talk, I will discuss sufficient dimension reduction (SDR) for functional data in which both predictor and response are multivariate functional data. First, I will introduce nested reproducing kernel Hilbert space (nested RKHS) which provides a general mechanism for nonlinear functional data analysis. Two layers of function spaces are constructed in a nested fashion so that the first space represents the observed functional data, and the second space characterizes nonlinearity of the random functions. As well as nonlinear functional data analysis, the nested

RKHS enables us to develop linear SDR for multivariate functional data without using a slice of the response. Then I will introduce a method of nonlinear SDR for functional data, the detailed procedure for estimation, and its consistency and convergence rate. The applications to speech recognition and handwriting symbol association show that the reduced predictors can be used for classification with great effectiveness.

**Speaker #3:**
**Date:** September 29, 2017
**Time and location:** 11:00am-12:00noon, Fretwell 315
**Speaker: Yanqing Sun,** Department of Mathematics and Statistics, UNCCharlotte

**Title: Analysis of Stratified Mark-Specific Proportional Hazards Models under Two-Phase Sampling with Application to HIV Vaccine Efficacy Trials**

**Abstract:** An objective of preventive HIV vaccine efficacy trials is to understand how immune responses to specific protein or sub-protein sequences of HIV associate with the level of vaccine efficacy to prevent infection with sequences of HIV targeted by the immune responses. The vaccine-induced immune response biomarkers are often measured via two-phase sampling for efficiency. Motivated by this objective, we investigate the stratified mark-specific proportional hazards model under two-phase sampling, where the mark is the genetic distance of an infecting HIV sequence to an HIV sequence represented inside the vaccine. The estimation and inference procedures based on inverse probability weighting of complete-cases and the augmented inverse probability weighted complete-case are developed. The asymptotic properties are derived, and their finite-sample performances are examined in a comprehensive simulation study. The methods are shown to have satisfactory performance, and are applied to the RV144 vaccine trial to assess whether immune response correlates of HIV infection are stronger for HIV infecting sequences similar to the vaccine than for sequences distant from the vaccine.

**Speaker #4:**
**Date:** October 13, 2017
**Time and location:** 11:00am-12:00noon, Fretwell 315
**Speaker:** Akim Adekpedjou, Department of Mathematics and Statistics, University of Missouri-Science & Technology

**Title: Semiparametric estimation with spatially correlated recurrent events**

**Abstract:** This talk pertains to the analysis of recurrent event data in the presence of spatial correlation. Consider units that are located in different geographical areas described by their longitude and latitude in a two dimensional surface. Existing recurrent events models, to the best of our knowledge will fail to capture any potential spatial patterns that may exist among these regions. In this talk, we propose a new class of semiparametric models for recurrent events that

can be used to identify spatial risk factors at onset and future recurrence of events while accounting for spatial correlation. Since the parameters involved in the models are not directly estimable because of the high dimensional of the likelihood, we borrow the approach of composite likelihood. The approach leads to estimates with population interpretation where their large samples properties are easily obtained under a reasonable set of regularity conditions. Simulation studies suggest that the resulting estimators have a very good finite sampling properties. The methods are illustrated using spatial data on recurrent oesophageal cancer. (Hosted by Dr. Yanqing Sun, UNC Charlotte)

**Speaker #5:**
**Date:** October 20, 2017
**Time and location:** 11:00am-12:00noon, Fretwell 315
**Speaker:** Weizhen Wang, Department of Mathematics and Statistics, Wright State University

Title: On Exact Inferences Using Binary Data in Two or Multi-stage Designs
Abstract: When establish an effective treatment using binary data from a two-stage design, one-sided tests for a proportion p are employed. Researchers typically use the parameter configuration at the boundary of the null hypothesis space to determine a rejection region. However, it is unclear whether the resultant test is of level-$\alpha$, especially in adaptive designs. In this paper, we prove that this is true for a large family of tests in both nonadaptive and adaptive two-stage designs by showing that the power function of any test in the family is a nondecreasing function in p, then establish similar results for nonadaptive multi-stage designs with m (> 2) stages. In addition, we derive optimal lower one-sided 1-$\alpha$ confidence intervals for p.

About the speaker: Weizhen Wang received his B.S. and M.S. at Peking University in 1987 and 1990, respectively, and completed his Ph.D. in Statistics at Cornell University in 1995. After one-year visit at Purdue University, he joined Wright State University, and has been a Professor of Statistics since 2007. His research includes bioequivalence, exact parametric and nonparametric inference, saturated and adaptive designs, categorical data analysis, foundation of statistics, statistical computation, dose-response study and causal inference. His current primary interest is exact statistical inference and its implementation in R.
(Hosted by Dr. Jiancheng Jiang, UNC Charlotte)

**Speaker #6:**
**Date:** October 27, 2017
**Time and location:** 11:00am-12:00noon, Fretwell 315
**Speaker:** Ling Ma, Department of Statistics, Clemson University

**Title: Joint modeling of longitudinal functional feature and discrete time-to-event**

**Abstract:** In longitudinal studies, it is often of interest to investigate how the functional feature of a marker's measurement process is associated with the event time of interest. We make use of B-splines to smoothly approximate the infinite dimensional functional data and propose a joint model of the longitudinal functional feature and the time to event. The proposed approach also allows for prediction of survival probabilities for future subjects based on their available

longitudinal measurements and a fitted joint model. We illustrate our proposals on a prospective pregnancy study, namely Oxford Conception Study, where hormonal measurements of luteinizing hormone which is an important biomarker of ovulation is available. A joint modeling approach using functional analytic approach and discrete survival modeling was used to assess whether the functional feature of hormonal measurements, such as the curvature of the hormonal profile is associated with time to pregnancy. (Hosted by Dr. Qingning Zhou, UNC Charlotte)

**Speaker #7:**
**Date:** November 3, 2017
**Time and location:** 11:00am-12:00noon, Fretwell 315
**Speaker:** Yanyuan Ma, Penn State University

**Title: On Estimation of General Index Model for Survival Data**

**Abstract:** We propose a general index model for survival data, which generalizes many commonly used semiparametric survival models and belongs to the framework of dimension reduction. Using a combination of geometric approach in semiparametrics and martingale treatment in survival data analysis, we devise estimation procedures that are feasible and do not require covariate-independent censoring as assumed in many dimension reduction methods for censored survival data. We establish the root-n consistency and asymptotic normality of the proposed estimators and derive the most efficient estimator in this class for the general index model. Numerical experiments are carried out to demonstrate the empirical performance of the proposed estimators and an application to an AIDS data further illustrates the usefulness of the work. (Hosted by Dr. Yanqing Sun, UNC Charlotte)

**Speaker #8:**
**Date:** November 10, 2017
**Time and location:** 11:00am-12:00noon, Fretwell 315
**Speaker:** Won Chang, University of Cincinnati

**Title: Calibrating an ice sheet model using high-dimensional binary spatial data**

**Abstract:** Rapid retreat of ice in the Amundsen Sea sector of West Antarctica may cause drastic sea level rise, posing significant risks to populations in low-lying coastal regions. Calibration of computer models representing the behavior of the West Antarctic Ice Sheet is key for informative projections of future sea level rise. However, both the relevant observations and the model output are high-dimensional binary spatial data; existing computer model calibration methods are unable to handle such data. Here we present a novel calibration method for computer models whose output is in the form of binary spatial data. To mitigate the computational and inferential challenges posed by our approach, we apply a generalized principal component based dimension reduction method. To demonstrate the utility of our method, we calibrate the PSU3D-ICE model by comparing the output from a 499-member perturbed-parameter ensemble with observations from the Amundsen Sea sector of the ice sheet.

Our methods help rigorously characterize the parameter uncertainty even in the presence of systematic data-model discrepancies and dependence in the errors. Our method also helps inform environmental risk analyses by contributing to improved projections of sea level rise from the ice sheets. A published article based on this work can be found here: http://www.tandfonline.com/doi/full/10.1080/01621459.2015.1108199

(Hosted by Dr. Jun Song, UNC Charlotte)

**Speaker #9:**
**Date:** December 1, 2017
**Time and location:** 11:00am-12:00noon, Fretwell 315
**Speaker:** Ram Tiwari, Center for Devices and Radiological Health, Office Surveillance and Biometrics, U.S. Food & Drug Administration

(Hosted by Dr. Yanqing Sun, UNC Charlotte)