Bio^3 Seminar Series

Ying Zhang, Ph.D.

Professor
& Director of Biostatistics Education,
Department of Biostatistics,
Indiana University

Model-Free Causal Inference in Observational Studies

Causal inference is a key component for comparative effectiveness research in observational studies. The inverse-propensity weighting (IPW) technique and augmented inverse-propensity weighting (AIPW) technique, which is known as a double-robust method, are the common methods for making causal inference in observational studies. However, these methods are known not stable, particularly when the models for propensity score and the study outcome are wrongly specified. In this work, we propose a model-free approach for causal inference. While possessing standard asymptotic properties, this method also enjoys excellent finite sample performance and robustness. Simulation studies were conducted to compare with the well-known IPW and AIPW methods for causal inference. A real-life example from an ongoing Juvenile Idiopathic Arthritis Study was applied for the illustration of the proposed method.

FRIDAY, September 14, 2018

10:00am – 11:00am

New Research Auditorium

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio^3 Seminar Series

Yangxin Huang, Ph.D.

Professor,
Department of Epidemiology and Biostatistics
College of Public Health
University of South Florida - Tampa

Bayesian Joint Models for Multivariate Longitudinal and Event Time Data with Multiple Features, with an Application to Diabetes Study

Joint modeling of longitudinal and survival data is an active area of statistics research and becoming increasingly essential in most epidemiological and clinical studies. As a result, a considerable number of statistical models and analysis methods have been suggested for analyzing such longitudinal-survival data. However, the following issues may standout. (i) a common assumption for longitudinal variables is that model errors are normally distributed due to mathematical tractability and computational convenience. This requires the variables to be "symmetrically" distributed. A violation of this assumption could lead to misleading inferences; (ii) in practice, many studies are often to collect multiple longitudinal exposures which may be significantly correlated, ignoring their correlations may lead to bias and reduce efficiency in estimation; (iii) the longitudinal responses may encounter non-ignorable missing; (iv) repeatedly measured observations in time are often interrelated with a time-to-event of interest. Inferential procedures may complicate dramatically when one analyzes data with these features together. Under the umbrella of Bayesian inference, this talk explores a multivariate mixed-effects joint models with skewed distributions for longitudinal measures with an attempt to mediate correlation from multiple responses, adjust departure from normality, and tailor accuracy from non-ignorable missingness as well as overcome shortage of confidence in specifying a time-to-event model. A data set arising from diabetes study is analyzed to demonstrate the methodology. Simulation studies are conducted to assess the performance of the proposed joint models and method under various scenarios.

FRIDAY, September 28, 2018

10:00am – 11:00am

Proctor Harvey Amphitheater, Med-Dent C104

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio^3 Seminar Series

Kosuke Imai, Ph.D.

Professor, Department of Government,
Department of Statistics,
Harvard University

Causal Inference with Interference and Noncompliance in Two-Stage Randomized Experiments

In many social science experiments, subjects often interact with each other and as a result one unit's treatment influences the outcome of another unit. Over the last decade, a significant progress has been made towards causal inference in the presence of such interference between units. Researchers have shown that the two-stage randomization of treatment assignment enables the identification of average direct and spillover effects. However, much of the literature has assumed perfect compliance with treatment assignment. In this paper, we establish the nonparametric identification of the complier average direct and spillover effects in two-stage randomized experiments with interference and noncompliance. In particular, we consider the spillover effect of the treatment assignment on the treatment receipt as well as the spillover effect of the treatment receipt on the outcome. We propose consistent estimators and derive their randomization-based variances under the stratified interference assumption. We also prove the exact relationship between the proposed randomization-based estimators and the popular two-stage least squares estimators. Our methodology is motivated by and applied to the randomized evaluation of the India's National Health Insurance Program (RSBY), where we find some evidence of spillover effects on both treatment receipt and outcome. The proposed methods are implemented via an open-source software package.

FRIDAY, October 12, 2018

10:00am – 11:00am

New Research Auditorium

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio^3 Seminar Series

Vernon M. Chinchilli, Ph.D.

Distinguished Professor & Chair,
Department of Public Health Sciences,
Hershey College of Medicine, Penn State University

N-of-1 Trials: Do They Have a Role in Clinical Research?

Physicians frequently use N-of-1 (single-patient) trial designs in an informal manner to identify an optimal treatment for an individual patient. An N-of-1 clinical trial that focuses exclusively on optimizing the primary outcome for a specific patient clearly may not be very useful for generalizability to a population of patients. A series or collection of N-of-1 clinical trials, however, could be generalizable. We review current literature on the design and analysis of N-of-1 trials, and this includes Bayesian approaches as well. We next describe the "Best African American Response to Asthma Drugs (BARD)" trial, which invokes a four-way crossover design and has the flavor of a series of N-of-1 trials. We propose a nonlinear mixed-effects model with a quadrinomial logistic regression for the analysis of the BARD data that constructs six pairwise comparisons of the four asthma treatments to (1) assess the optimal treatment for each study participant and (2) estimate population-level treatment comparisons.

FRIDAY, October 26, 2018

10:00am – 11:00am

New Research Auditorium

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio^3 Seminar Series

Michael Proschan, Ph.D.

Mathematical Statistician,
Biostatistics Research Branch
National Institute of Allergy & Infectious Diseases (NIAID)

Probability Paradoxes That Actually Happened

Probability books sometimes present "cooked" counterexamples to warn students of the lurking dangers of compromising rigor. Can such anomalies actually occur in practice? This talk is proof that they can! I present actual examples from my clinical trials experience in which I either fell into, or almost fell into, probability traps. These experiences were a major motivation for our book, "Essentials of Probability Theory for Statisticians" (Proschan and Shaw, 2016, CRC Press, Taylor & Francis Group).

Wednesday, November 7, 2018

10:00am – 11:00am

New Research Building - Auditorium

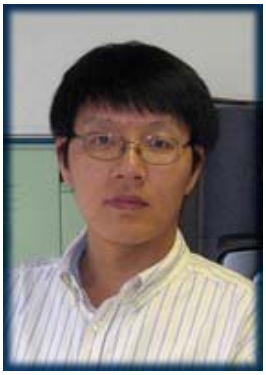
Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio^3 Seminar Series

Faming Liang, Ph.D.

Professor, Department of Statistics,
Purdue University



Extended Stochastic Gradient MCMC Algorithms for Large-Scale Bayesian Computing

The stochastic gradient Markov chain Monte Carlo (SGMCMC) algorithms, such as stochastic gradient Langevin dynamics and stochastic gradient Hamilton Monte Carlo, have recently received much attention in Bayesian computing for large-scale data for which the sample size can be very large, or the dimension can be very high, or both. However, these algorithms can only be applied to a small class of problems for which the parameter space has a fixed dimension and the log-posterior density is differentiable with respect to the parameters. We propose a class of extended SGMCMC algorithms which, by introducing appropriate latent variables and utilizing Fisher's identity, can be applied to more general large-scale Bayesian computing problems, such as those involving dimension jumping and missing data. For a large-scale dataset with sample size N and dimension p, the proposed algorithms can achieve a computational complexity of O(N^{1+\epsilon} p^{1-\epsilon'}) for some small constants \epsilon and \epsilon', which is quite comparable with the computational complexity O(N p^{1-\epsilon'}) achieved in general by the stochastic gradient descent (SGD) algorithm. The proposed algorithms are illustrated using high-dimensional variable selection, sparse deep learning with large-scale data, and a large-scale missing data problem. The numerical results show that the proposed algorithms have a significant computational advantage over traditional MCMC algorithms and can be highly scalable when mini-batch samples are used in simulations. Compared to frequentist methods, they can produce more accurate variable selection and prediction results, while exhibiting similar CPU costs when the dataset contains a large number of samples. The proposed algorithms have much alleviated the pain of Bayesian methods in large-scale computing.

FRIDAY, January 11, 2019

10:00am – 11:00am

Warwick Evans, Building D

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio^3 Seminar Series

Hui Quan, Ph.D.

Associate VP and Global Head of Methodology, Biostatistics and Programming Department, Sanofi

Considerations on trial design and data analysis of multi-regional clinical trials

Extensive research has been conducted in the Multi-Regional Clinical Trial (MRCT) area. To effectively apply an appropriate approach to a MRCT, we need to synthesize and understand the features of different approaches. In this presentation, numerical and real data examples are used to illustrate considerations regarding design, conduct, analysis and interpretation of result of MRCTs. We compare different models as well as their corresponding interpretations of the trial results. We highlight the importance of paying special attention to trial monitoring and conduct to prevent potential issues associated with the final trial results. Besides evaluating the overall treatment effect for the entire MRCT, we also consider other key analyses including quantification of regional treatment effects within a MRCT and the assessment of consistency of these regional treatment effects.

FRIDAY, January 25, 2019

10:00am – 11:00am

Building D, Warwick Evans Conference Room

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio^3 Seminar Series

Yanxun Xu, Ph.D.

Assistant Professor, Department of Applied Mathematics and Statistics, Johns Hopkins University

ASIED: A Bayesian Adaptive Subgroup-Identification Enrichment Design

Developing targeted therapies based on patients' baseline characteristics and genomic profiles such as biomarkers has gained growing interests in recent years. Depending on patients' clinical characteristics, the expression of specific biomarkers or their combinations, different patient subgroups could respond differently to the same treatment. An ideal design, especially at the proof of concept stage, should search for such subgroups and make dynamic adaptation as the trial goes on. When no prior knowledge is available on whether the treatment works on the all-comer population or only works on the subgroup defined by one biomarker or several biomarkers, it's necessary to estimate the subgroup effect adaptively based on the response outcomes and biomarker profiles from all the treated subjects at interim analyses. To address this problem, we propose an Adaptive Subgroup-Identification Enrichment Design, ASIED, to simultaneously search for predictive biomarkers, identify the subgroups with differential treatment effects, and modify study entry criteria at interim analyses when justified. More importantly, we construct robust quantitative decision-making rules for population enrichment when the interim outcomes are heterogeneous. Through extensive simulations, the ASIED is demonstrated to achieve desirable operating characteristics and compare favorably against the alternatives.

FRIDAY, February 8, 2019

10:00am – 11:00am

Building D, Warwick Evans Conference Room

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio^3 Seminar Series

Steven G. Heeringa, Ph.D.

Senior Research Scientist & Associate Director,
Survey Research Center, University of Michigan - Ann Arbor

Neuroscience on a Population Scale: Design, Measurement, Data Integration and Analysis

Neuroscience has a strong research tradition that employs experimental and observational studies in laboratory settings and controlled testing and evaluation in both clinical, educational and volunteer populations. In the past two decades, there has been increasing interest in conducting population-scale epidemiological studies of early age brain development and functioning as well as later age neurological functioning including cognitive impairment, dementias and Alzheimer's disease. The data collected in these population-based studies is not restricted to observations on neurological systems and functioning but is collected in parallel with a wide array of information on participants' life events, medical history, social and environmental exposures, genetics and genomics. This rich array of observational data has the potential to greatly advance our understanding of how complex neurological systems develop, are modified by internal or external factors or otherwise change over the life course. The growing field of epidemiological research also presents many challenging problems in design, measurement, data integration and analysis that those of us trained in biostatistics, bioinformatics and biomathematics will be called on to help to solve.

This presentation will use two cases studies to illustrate the nature of the statistical challenges in conducting population-scale neuroscientific research, describe current best practices and outline opportunities for future research. The first case study will be the Adolescent Brain Cognitive Development project, a 12-year longitudinal investigation of brain morphology and functional development in U.S. adolescents and teens. The second case study will focus on the challenges in design, measurement and analysis faced in special supplemental investigations of dementia and Alzheimer's disease conducted under the auspices of the larger Health and Retirement Study. Each case study review will include a description of the specific study challenges and current solutions. The major aim of this presentation is to increase awareness of these emerging lines of research and to promote interest on the part of the next generation of statisticians and data scientists who will be called upon to advance the various methodologies that will be required to better understand complex neurological systems and how they relate to our individual attributes and the world around us.

FRIDAY, February 22, 2019

10:00am – 11:00am

Building D, Warwick Evans Conference Room

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio^3 Seminar Series

Abera Wouhib, Ph.D.

Program Chief, Statistical Methods in Psychiatry Program, NIH

Estimation of Heterogeneity Parameters in Multivariate Meta-Analysis

Similar to its univariate counterpart, multivariate meta-analysis is a method to synthesize multiple outcome effects by taking in to account the available variance-covariance structure. It can improve efficiency over separate univariate syntheses and enables joint inferences across the outcomes. Multivariate meta-analysis is required to address the complexity of the research questions. Multivariate data can arise in meta-analysis due to several reasons. The primary studies can be multivariate in nature by measuring multiple outcomes for each subject, typically known as multiple-endpoint studies, or it may arise when primary studies involve several comparisons among groups based on a single outcome or measures several parameters. Although it possesses many advantages over the more established univariate counterpart, multivariate meta-analysis has some challenges including modelling and estimating the parameter of interests. Under random-effects model assumption, we discuss the methods of estimating the heterogeneity parameters and effect sizes of the multivariate data and its application by using illustrative example and simulation results.

FRIDAY, April 12, 2019

10:00am – 11:00am

Building D, Warwick Evans Conference Room

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio^3 Seminar Series

Ming Yuan, Ph.D.

Professor, Department of Statistics,
Columbia University

Quantitation in Colocalization Analysis: Beyond "Red + Yellow = Green

"I see yellow; therefore, there is colocalization." Is it really so simple when it comes to colocalization studies? Unfortunately, and fortunately, no. Colocalization is in fact a supremely powerful technique for scientists who want to take full advantage of what optical microscopy has to offer: quantitative, correlative information together with spatial resolution. Yet, methods for colocalization have been put into doubt now that images are no longer considered simple visual representations. Colocalization studies have notoriously been subject to misinterpretation due to difficulties in robust quantification and, more importantly, reproducibility, which results in a constant source of confusion, frustration, and error. In this talk, I will share some of our effort and progress to ease such challenges using novel statistical and computational tools.

FRIDAY, April 26, 2019

10:00am – 11:00am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center