Edoardo Airoldi

Professor of Statistics and Data Science
Temple University

Title: Model-assisted design of experiments on social and information networks

Abstract: Classical approaches to causal inference largely rely on the assumption of "lack of interference", according to which the outcome of an individual does not depend on the treatment assigned to others, as well as on many other simplifying assumptions, including the absence of strategic behavior. In many applications, however, such as evaluating the effectiveness of health-related interventions that leverage social structure, assessing the impact of product innovations and ad campaigns on social media platforms, or experimentation at scale in large IT companies, assuming lack of interference and other simplifying assumptions is untenable. Moreover, the effect of interference itself is often an inferential target of interest, rather than a nuisance. In this talk, we will formalize technical issues that arise in estimating causal effects when interference can be attributed to a network among the units of analysis, within the potential outcomes framework. We will introduce and discuss several strategies for experimental design in this context centered around a judicious use statistical models, which we refer to as "model-assisted" design of experiments. In particular, we wish for certain finite-sample properties of the estimator to hold even if the model catastrophically fails, while we would like to gain efficiency if certain aspects of the model are correct. We will then contrast design-based, model-based and model-assisted approaches to experimental design from a decision theoretic perspective.

Bio: Edoardo M. Airoldi is the Millard E. Gladfelter Professor of Statistics and Data Science in the Fox School of Business, at Temple University. He also serves as Director of the Fox School's Data Science Center. Prior to joining Fox, Airoldi was on the faculty in the Department of Statistics at Harvard University, where he founded and directed the Harvard Laboratory for Applied Statistics & Data Science, until 2018. Additionally, he held visiting positions at MIT and Yale. His current research focuses on statistical theory and methods for designing and analyzing experiments on large networks and, more generally, modeling and inferential issues that arise in analyses that leverage network data. His work has appeared in journals across statistics, computer science, and general science, including Annals of Statistics, Journal of the American Statistical Association, Journal of Machine Learning Research, Proceedings of the National Academy of Sciences, and Nature. Airoldi has received a Sloan Fellowship, the Shutzer Fellowship from the Radcliffe Institute of Advanced Studies, an NSF CAREER Award, and an ONR Young Investigator Program Award, among other honors. He has delivered a plenary talk at the National Academy of Sciences Sackler Colloquium on "Causal Inference and Big Data," in 2015, and he has given an IMS Medallion Lecture at the Joint Statistical Meetings, in 2017.

Jennifer Hill

Professor of Applied Statistics and Data Science
New York University

Title: Identifying Heterogeneous Treatment Effects with Bayesian Nonparametrics

Abstract: After decades of focus in the causal inference world on average treatment effects, there is now increasing interest in estimation of more targeted effects including even individual-level treatment effects. We explore the tradeoffs between two Bayesian non-parametric approaches to this estimation goal that enforce more and less smoothing. Implications in terms of precision, uncertainty quantification, and ability to detect violations of assumptions are presented. Finally, a compromise approaches capitalizing on the strengths of both BART and Gaussian Processes is described.

Bio: Jennifer Hill develops and evaluates methods that help us answer the causal questions that are vital to policy research and scientific development. In particular, she focuses on situations in which it is difficult or impossible to perform traditional randomized experiments, or when even seemingly pristine study designs are complicated by missing data or hierarchically structured data. Most recently Hill has been pursuing two intersecting strands of research. The first focuses on Bayesian nonparametric methods that allow for flexible estimation of causal models and are less time-consuming and more precise than competing methods (e.g. propensity score approaches). The second line of work pursues strategies for exploring the impact of violations of typical causal inference assumptions such as ignorability (all confounders measured) and common support (overlap). Hill has published in a variety of leading journals including Journal of the American Statistical Association, Statistical Science, American Political Science Review, American Journal of Public Health, and Developmental Psychology. Hill earned her PhD in Statistics at Harvard University in 2000 and completed a post-doctoral fellowship in Child and Family Policy at Columbia University's School of Social Work in 2002.

Luke Keele

Associate Professor of Statistics in Surgery
University of Pennsylvania

Title: Estimation Methods for Cluster Randomized Trials with Noncompliance: A Study of A Biometric Smartcard Payment System in India

Abstract: Many policy evaluations occur in settings with randomized assignment at the cluster level and treatment noncompliance at the unit level. For example, villagers or towns might be assigned to treatment and control, but residents may choose to not comply with their assigned treatment status. For example, in the state of Andhra Pradesh, India, the state government sought to evaluate the use of biometric smartcards to deliver payments from antipoverty programs. Smartcard payments were randomized at the village level, but residents could choose to comply or not. In some villages, more than

90\% of residents complied with the treatment, while in other locations fewer than 15\% of the residents complied. When noncompliance is present, investigators may choose to focus attention on either intention to treat effects or the treatment effect among the units that comply. When analysts focus on effects among compliers, the instrumental variables framework can be used to evaluate identify causal effects. We first review extant methods for instrumental variable estimators in clustered designs which depend on assumptions that are often unrealistic in applied settings. In response, we present a method that allows for possible treatment effect heterogeneity that is correlated with cluster size and uses finite sample variance estimator. We evaluate these methods using a series of simulations and apply them to data from an evaluation of welfare transfers via smartcard payments in India.

Bio: **Luke Keele** (Ph.D., University of North Carolina, Chapel Hill, 2003) Professor Keele specializes in research on applied statistics. His research in focuses on causal inference, design-based methods, matching, and instrumental variables. He also conducts research on topics in educational program evaluation, election administration, and health services research. He has published articles in the *Journal of the American Statistical Association, Annals of Applied Statistics, Journal of the Royal Statistical Society, Series A, The American Statistician, American Political Science Review*, *Political Analysis*, and *Psychological Methods*.

Fan Li

Associate Professor of Statistical Science
Duke University

Title: Introducing the overlap weights in causal inference

Abstract: Covariate balance is crucial for unconfounded descriptive or causal comparisons in observational studies. We propose a unified framework ---the balancing weights---to balance the weighted distributions of the covariates between treatment groups. These weights incorporate the propensity score to weight each group to an analyst-selected target population, and include several commonly used weighting schemes such as inverse-probability weight and trimming as special cases.   We derive the large-sample results on nonparametric estimation based on these weights. We further propose a new weighting scheme, the overlap weights, in which each unit's weight is proportional to the probability of that unit being assigned to the opposite group. The overlap weights are bounded, and minimize the asymptotic variance of the weighted average treatment effect among the class of balancing weights. The overlap weights also possess a small-sample exact balance property, based on which we propose a new method that achieves exact balance for means of any selected set of covariates. We apply the method the Framingham Heart Study to evaluate the effect of statins on health outcomes. Extensions to subgroup analysis and multi-arm treatments will also be discussed.

Bio: Fan Li is an associate professor of Statistical Science, and Biostatistics and Bioinformatics (secondary) at Duke University. She is also an affiliated member of the Duke Clinical Research Institute. She received her Ph.D. in biostatistics from Johns Hopkins University Bloomberg School of Public Health, and postdoctoral training at Harvard Medical School Department of Health Care Policy. Her primary research interest is statistical methods for causal inference, particularly with applications to comparative effectiveness research in health studies. She also works extensively on methods for missing data, Bayesian analysis and high dimensional data analysis. She is the associate editor of Journal of the American Statistical Association (Application and Case Studies) and Bayesian Analysis.

Hongzhe Li

Professor of Biostatistics and Statistics
University of Pennsylvania

Title: Inference for Individual Mediation Effects and Interventional Effects in Sparse High- Dimensional Causal Graphical Models

Abstract: We consider the problem of identifying intermediate variables (or mediators) that regulate the effect of a treatment on a response variable. While there has been significant research on this topic, little work has been done when the set of potential mediators is high-dimensional. A further complication arises when the potential mediators are interrelated. In particular, we assume that the causal structure of the treatment, the potential mediators and the response is a directed acyclic graph (DAG). High-dimensional DAG models have previously been used for the estimation of causal effects from observational data. In particular, methods called IDA and joint-IDA have been developed for estimating the effect of single interventions and the effect of multiple simultaneous interventions respectively. In this paper, we propose an IDA-type method, called MIDA, for estimating mediation effects from high-dimensional observational data. Although IDA and joint-IDA estimators have been shown to be consistent in certain sparse high-dimensional settings, their asymptotic properties such as convergence in distribution and inferential tools in such settings remained unknown. In this paper, we prove high-dimensional consistency of MIDA for linear structural equation models with sub-Gaussian errors. More importantly, we derive distributional convergence results for MIDA in similar high-dimensional settings, which are applicable to IDA and joint-IDA estimators as well. To the best of our knowledge, these are the first distributional convergence results facilitating inference for IDA-type estimators. These results have been built on our novel theoretical results for linear regressions over varying subsets of high-dimensional covariates, which may be of independent interest. Finally, we empirically demonstrate the usefulness of our asymptotic theory in the identification of large mediation effects and we illustrate a practical application of MIDA in genomics with a real dataset.

Bio: Dr. Hongzhe Li is a Professor of Biostatistics and Statistics at the Perelman School of Medicine at the University of Pennsylvania (Penn). He is Vice Chair of Integrative Research in the Department of Biostatistics, Epidemiology and Informatics, Chair of the Graduate Program in Biostatistics and Director of Center of Statistics in Big Data at Penn. Dr. Li has been elected as a Fellow of the American Statistical Association (ASA), a Fellow of the Institute of Mathematical Statistics (IMS) and a Fellow of AAAS. Dr. Li severed on the Board of Scientific Counselors of the National Cancer Institute of NIH and regularly serves on various NIH study sections. He is currently an Associate Editor of Biometrics, Statistica Sinica and also Chief co-Editor of Statistics in Biosciences. He served as Chair of the Section on Statistics in Genomics and Genetics of the ASA. Dr. Li's research has been focused on developing powerful statistical and computational methods for analysis of large-scale genetic, genomics and metagenomics data and high dimensional statistics with applications in genomics. He has published papers in Science, Nature, Nature Genetics, Nature Methods, Science Translational Medicine, JASA, JRSS, Biometrika, Annals of Statistics etc.

Jasjeet Sekhon

Professor of Political Science and Statistics
University of California, Berkeley

Title: Transfer Learning for Estimating Causal Effects using Neural Networks

Abstract: We develop new algorithms for estimating heterogeneous treatment effects, combining recent developments in transfer learning for neural networks with insights from the causal inference literature. By taking advantage of transfer learning, we are able to efficiently use different data sources that are related to the same underlying causal mechanisms. We compare our algorithms with those in the extant literature using extensive simulation studies based on large-scale voter persuasion experiments and the MNIST database. Our methods can perform an order of magnitude better than existing benchmarks while using a fraction of the data.

Bio: Jasjeet Sekhon is a professor of political science and statistics and a senior fellow at the University of California, Berkeley. He received his Ph.D. from Cornell Unversity.

Michael Sobel

Professor of Statistics
Columbia University

Title: Between Causation and Association: Inference for the Role of Judge Attributes in EEOC Litigation Outcomes

Abstract: A longstanding question in the American literature on judicial decision concerns the manner in which case outcomes depend on judge attributes such as race, sex and ideology. The literature is inconclusive at best. Typically, using case covariates and judge attributes as predictors, statistical models of various outcomes are estimated, and the coefficients associated with the attributes interpreted as effects. This interpretation is incorrect: researchers do not regard the attributes as treatments, but as proxies for unmeasured variables that presumably vary more between than within attributes. Thus, it is necessary to back up, articulate the question of interest, and develop a suitable framework to address it. The primary concern in the literature is that judges with different attributes will handle cases differently. Studying how judges with a given attribute handle cases to which they are assigned may not be indicative of how judges with different attributes would handle these cases. Ideally, one wants to compare judges with different attributes on a common set of cases, taking also into consideration within attribute heterogeneity in outcomes. Using judge's potential outcomes, we propose several estimands and discuss their identification. An estimation strategy based on a Bayesian hierarchical model of award outcomes in cases filed by the Equal Employment Opportunity Commission between October 1,1996 and September 30,2006 is used to estimate these quantities. We find little support for the notion that non-white judges favor plaintiffs more than white judges; similar comments apply with respect to male and female judges, and to judges appointed by Republican and Democratic presidents.

Bio: Michael Sobel is a professor in the statistics department at Columbia University. His work is primarily in the area of causal inference, with applications in functional magnetic resonance imaging (fMRI) and social statistics.

Elizabeth Stuart

Associate Dean for Education, Professor of Mental Health, Biostatistics, and Health Policy and Management, John Hopkins Bloomberg School of Public Health

Title: Confounders? Assumptions? Huh? Mediation analysis in applied research

Abstract: Mediation analysis is an appealing method in health research, with the idea that it helps identify mechanisms of action for interventions. This has led some agencies, including the National Institute of Mental Health, to require mediation analyses be conducted in clinical trials. In addition, in recent years there has been an explosion of causal mediation methods, which aim to define mediation effects using potential outcomes and better articulate the underlying assumptions. However, it is unclear how much those causal mediation approaches have been adopted by applied researchers, nor is it clear how well applied researchers understand the assumptions underlying mediation analyses, and how to conduct analyses that best meet those assumptions. To answer these questions we conducted a literature review of mediation analysis in mental health research. All articles considered came from academic journals in the PsycInfo database, published in the English language from 2013-2018. Because we wanted a representation of how mediation analyses were being performed in the most respected top tier journals, we further restricted our search to those journals that were listed among the top ten google scholar publications for each of Psychiatry and Psychology, for a total of 20 journals. We selected all articles with the phrase "mediation analys*" in the title, abstract, or keywords. This resulted in 209 articles whose abstracts we further screened for inclusion into our study. Using AbstrackR software, each abstract was assigned to two independent reviewers and screened for suitable content, resulting in 199 articles carried forward to the extraction process. Two unique reviewers were randomly assigned to extract information on each article. The information extracted related to study design, sample size, covariates adjusted for in the analysis, temporal ordering of variables, and the specific method used to perform the mediation analysis. In preliminary findings, almost none of the articles use causal mediation methods, control for confounding, or discuss the underlying assumptions of their approaches (such as linear relationships between outcome and mediator). These results provide insights for how we as statisticians should aim to communicate our methods, and motivation for more outreach to the applied research community on best practices for mediation analysis.

Bio: Elizabeth A. Stuart (estuart@jhu.edu) is Professor in the Departments of Mental Health, Biostatistics, and Health Policy and Management at the Johns Hopkins Bloomberg School of Public Health, and Associate Dean for Education at JHSPH. She is a Fellow of the American Statistical Association and has received the mid-career award from the Health Policy Statistics Section of the American Statistical Association, the Gertrude Cox Award for applied statistics, and the Myrto Lefkopoulou award from the Harvard University Department of Biostatistics. Dr. Stuart has published influential papers on propensity scores and methods to assess the generalizability of randomized trials and has received funding for her work from the National Institutes of Health, the National Science Foundation, and the US Department of Education Institute of Education Sciences.

Eric Tchetgen Tchetgen

Luddy Family President's Distinguished Professor and Professor of Statistics, The Wharton School University of Pennsylvania

Title: Longitudinal Marginal Structural Models Estimation with Instrumental Variables: Identification and Multiple Robustness

Abstract: In a seminal paper, Robins (1998) introduced marginal structural models (MSMs), a general class of counterfactual models for the joint effects of time-varying treatment regimes in complex longitudinal studies subject to time-varying confounding. He established identification of MSM parameters under a sequential randomization assumption (SRA), which rules out unmeasured confounding of treatment assignment over time. We extend Robins' MSM theory by considering identification of MSM parameters with the aid of a time-varying instrumental variable, when sequential randomization fails to hold due to unmeasured confounding. Our identification conditions essentially require that no unobserved confounder predicts compliance type at each follow-up time. Under this assumption, we obtain a large class of semiparametric estimators that extends standard inverse-probability weighting (IPW) and includes multiply robust estimators, including a locally semiparametric efficient estimator.

Bio: Prof Tchetgen Tchetgen is the Luddy Family President's Distinguished Professor and Professor of Statistics at the Wharton School at The University of Pennsylvania. He is interested in statistical and epidemiological methods to improve inference in presence of confounding bias, selection bias or missing data . His collaborations span several public health areas including HIV/AIDS, environmental Health, genetic epidemiology and Social Epidemiology. He is founding co-editor of the journal Epidemiologic Methods, Associate editor of the American Journal of epidemiology and Statistical Science.

Stijn Vansteelandt

Professor of Statistics, Ghent University (Belgium) and the London School of Hygiene and Tropical Medicine (UK)

Title: Time-to-event mediation analysis of randomised trials with repeatedly measured mediators: a re-analysis of the LEADER trial

Abstract: The LEADER trial found protective cardiovascular effects of liraglutide compared to placebo in patients with Type II diabetes and high cardiovascular risk. Effects were also found on glycated haemoglobin, body weight, blood pressure and heart rate, thereby raising the question to what extent these potential pathways my explain liraglutide's protective effect. We will explain how we addressed this question by expanding modern techniques from causal mediation analysis. In particular, we will show how to identify and infer the path-specific effect of liraglutide on the time to major cardiovascular events via the repeatedly measured glycated haemoglobin levels. The considered proposal addresses complications due to patients dying before the mediator is assessed, due to the mediator being repeatedly measured, and due to post-treatment confounding of the effect of glycated haemoglobin by other mediators, which makes mediation analysis a challenging enterprise.

Bio: Stijn Vansteelandt is Professor of Statistics in the Department of Applied Mathematics, Computer Science and Statistics, and Professor of Statistical Methodology in the Department of Medical Statistics at the London School of Hygiene and Tropical Medicine. He has authored over 150 peer-reviewed publications in international journals on a variety of topics in biostatistics, epidemiology and medicine, such as the analysis of longitudinal and clustered data, missing data, mediation and moderation/interaction, instrumental variables, time-dependent confounding, family-based genetic association studies, analysis of outcome-dependent samples and phylogenetic inference. He is Co-Editor of Biometrics, the leading flagship journal of the International Biometrics Society, and has previously served as Associate Editor for the journals Biometrics, Biostatistics, Epidemiology, Epidemiologic Methods and the Journal of Causal Inference.