NISS Workshop: Causal Inference and Machine Learning/High Dimensional Data

Tuesday May 23, 1pm – 5pm

Machine Learning, Computation, and Causal Inference

Organizer: Cynthia Rudin, Duke University

Causal Inference with High-Dimensional Controls and Parameters of Interest
Alex Belloni, Duke University

Alex will propose and analyze procedures to construct confidence regions for p (infinite dimensional) parameters of interest after model selection for general moment condition models where p is potentially larger than the sample size n. The use of functional parameters allow Alex to investigate various types of causal effects including quantile treatment effects that better characterize heterogeneous effects.

Causal Analysis for Big Data using Techniques from Databases
Sudeepa Roy, Duke University

Causal analysis for large observational datasets using matching techniques has a natural connection to concepts from the area of Database Management in Computer Science like relational databases and SQL queries. In this talk, Sudeepa will describe these connections, present some preliminary results on efficient and scalable matching techniques using SQL queries, and discuss some new research directions related to causal analysis on complex big data.

Estimating Optimal Intervention Strategies Against an Intelligent and Adaptive Adversary with Application to Real-Time Disruption of Human Trafficking in the United States
Eric Laber, North Carolina State University

This talk addresses real-time strategy for causal inference. Eric says that there are spatial spillover effects in the context of a partially observable Markov game.

Hypothesis tests that are robust to subjective choices in matching
Cynthia Rudin, Duke University

Our goal is to create robust matched pairs hypothesis tests for causal inference. These tests implicitly consider all possible reasonably good match assignments and consider the range of possible outcomes for tests on these data. This is a more computationally demanding approach to hypothesis testing than the standard approach where one considers just a single assignment, but the result would be robust to the choice of experimenter.

High-Dimensional Causal Inference
Organizer: Avi Feller, UC Berkeley

Approximate Residual Balancing: De-Biased Inference of Average Treatment Effects in High Dimensions
Stefan Wager, Stanford University

There are many settings where researchers are interested in estimating average treatment effects and are willing to rely on the unconfoundedness assumption, which requires that the treatment assignment be as good as random conditional on pre-treatment variables. The unconfoundedness assumption is often more plausible if a large number of pre-treatment variables are included in the analysis, but this can worsen the performance of standard approaches to treatment effect estimation. In this paper, we develop a method for de-biasing penalized regression adjustments to allow sparse regression methods like the lasso to be used for sqrt{n}-consistent inference of average treatment effects. Our method works under substantially weaker assumptions than other methods considered in the literature: Unlike high- dimensional doubly robust methods recently developed in econometrics, we do not need to assume that the treatment propensities are estimable, and unlike existing de-biasing techniques from the statistics literature, our method is not limited to considering sparse contrasts of the parameter vector. Instead, in addition standard assumptions used to make lasso regression on the outcome model consistent under 1-norm error, we only require overlap, i.e., that the propensity score be uniformly bounded away from 0 and 1. Procedurally, our method combines balancing weights with a regularized regression adjustment.

Overlap and Deconfounding Scores in High Dimensions
Alex D'Amour, UC Berkeley

A key advantage of observational studies with high-dimensional covariates is that the unconfoundedness assumption is often more plausible than in low-dimensional settings. Less discussed is the fact that overlap (i.e., positivity) in the population becomes less plausible with high-dimensional covariates. In this talk, we use results established in the probability literature to draw out some implications of the overlap assumption in high dimensions, which are stronger than most investigators realize. In particular, we show that high-dimensional overlap implies balance between population moments of covariate distributions under some distributional conditions, and more generally implies a variant of sparsity. These results should encourage investigators to be cautious about using methods that require overlap in high-dimensional covariate sets, particularly those that employ propensity score methods in a way that is agnostic to the outcome.

Selecting Subpopulations for Causal Inference in High Dimensional Settings
Alessandra Mattei, University of Florence

Extracting causal information from big data is not always straightforward, especially in regression- discontinuity (RD) studies where the treatment assignment rule depends on some type of cutoff formula. RD designs are often described as local randomized experiments: a RD design can be considered as a randomized experiment for units with a realized value of a so-called forcing variable falling around a pre-fixed threshold. Following Li, Mattei and Mealli (2015), we formally describe RD designs as local randomized experiments under the potential outcome framework. This approach views the forcing variable as a random variable with a probability distribution and reformulates the identifying assumptions – overlap, SUTVA and unconfoundedness – accordingly. The core of this framework is to assume that there exists at least one subpopulation of units for which we can invoke a local version of these assumptions (RD assumptions). Unfortunately we usually do not know the subpopulations for which we can draw valid causal inference. Therefore an important issue in practice concerns the selection of these subpopulations. In this talk, we investigate the use of unsupervised machine learning methods to select suitable subpopulations for causal inference in high dimensional RD designs. Specifically we view the selection of suitable subpopulations around the threshold as a clustering problem. The aim is to classify observations into subpopulations for which RD assumptions hold and we can draw valid causal inference and subpopulations for which we cannot extract any causal information from the observed data. We propose a model-based finite mixture approach to clustering in a Bayesian framework. This approach has important advantages: It explicitly accounts for the uncertainty on subpopulation membership; it does not impose any constraint on the shape of the subpopulation; and it properly works in high-dimensional settings. We illustrate the framework in a high-dimensional regression-discontinuity study concerning the effect of the Borsa Familia program, a social welfare program of the Brazilian government, on leprosy incidence.

Estimating Causal Networks in High-Dimension Observational Data
Organizer: Donglin Zeng, University of North Carolina

Inference in Gaussian DAGs with Known Partial Ordering
Syed Rahman, University of Florida

Directed acyclic graphs (DAGs) are commonly used to represent causal relationships among random variables in graphical models. Applications of these models arise in the study of physical, as well as biological systems, where directed edges between nodes represent the influence of components of the system on each other. There are two important lines of work that have currently dealt with DAG estimation in the Gaussian framework - one where the ordering is known and one where it is unknown. When the nodes exhibit a natural ordering, the problem of estimating directed graphs reduces to the problem of estimating the structure of the network. This leads to a convex problem which can be solved in a fast and efficient manner with the added advantage of convergence guarantees. On the other hand, when the ordering is unknown we get a non-convex problem, which leads to a much slower algorithm. In this paper, we propose a penalized likelihood approach that estimates DAGs when a partial ordering is known. By combining ideas from the known ordering problem, we formulate a more efficient and tractable algorithm than the case of the completely unknown ordering. This is joint work with George Michailidis and Kshitij Khare.

Estimating Latent Causal Network through Sparse Mixed Effects Directed Acyclic Graphs
Yuanjia Wang, Columbia University

Inferring causal relationship between variables from non-experimental data is a highly challenging goal, especially for large-scale data where estimation of directed acyclic graphs is NP-hard. Under the framework of structural equation models, or functional causal models (Pearl 2000), we represent joint distribution of variables in causal directed acyclic graphs (DAGs) as a set of structural equations, where directed edges connecting variables depend on subject-specific covariates and unobserved latent variables. The functional causal model framework allows constructing subject-specific DAGs, where the edges representing strength of network connectivity between variables decompose into a fixed effect term (average network effect given covariates), and a random effect term (unobserved residual network effect). Thus, our framework is a mixed effects DAG model. By pooling information across subjects under this model, we can estimate subject-specific network effects with a much better precision and test heterogeneity of network effects with a better power. We theoretically prove identifiability of our model, assess conditions when the network effects have a causal interpretation, propose a penalized likelihood based approach to handle high-dimensionality of the DAG model space, and a fast computational algorithm to achieve hard-thresholding of the edges. Through extensive simulations, we show substantially improved performance compared to the popular PC-algorithm (Spirtes et al.~2000). Lastly, we apply proposed methods to discover causal relationship among regions of brain atrophy as measured by neuroimaging biomarkers in Huntington's disease patients.

Estimating the Skeleton of High Dimensional Directed Acyclic Graphs
Jichun Xie, Duke University

Estimation of the skeleton of a directed acyclic graph (DAG) is of great importance for understanding the underlying DAG and causal effects can be assessed from the skeleton when the DAG is not identifiable. We propose a novel method named PenPC to estimate the skeleton of a high-dimensional DAG by a two-step approach. We first estimate the non-zero entries of a concentration matrix using penalized regression, and then fix the difference between the concentration matrix and the skeleton by evaluating a set of conditional independence hypotheses. For high dimensional problems where the number of vertices p is in polynomial or exponential scale of sample size n, we study the asymptotic property of PenPC on two types of graphs: traditional random graphs where all the vertices have the same expected number of neighbors, and scale-free graphs where a few vertices may have a large number of neighbors. As illustrated by extensive simulations and applications on gene expression data of cancer patients, PenPC has higher sensitivity and specificity than the state-of-the-art method, the PC-stable algorithm.

Estimation of Sparse Directed Acyclic Graphs through a Lasso Framework and its Applications
Hua Zhong, New York University

Causal networks are conveniently presented by directed acyclic graphs (DAGs). To estimate DAGs from high dimensional data is challenging due to the large number of possible spaces of DAGs, the acyclicity constraint of the structures, the typically nonconvex objective functions, and the problem of equivalent classes from observational data. In this talk, we present an efficient two-stage algorithm to estimate sparse DAGs under adaptive L1-penalized likelihood objective function with the acyclicity constraint. Simulations are presented to demonstrate the efficiency and flexibility of the proposed method. Real data examples are discussed on gene regulatory networks.

Wednesday May 24, 9:30am – 10:45am

Machine Learning and Causal Inference
Organizers: Jennifer Hill and Uri Shalit, New York University

Automated versus Do-It-Yourself Methods for Causal Inference: Lessons Learned from a Data Analysis Competition
Uri Shalit, New York University

Statisticians have made great strides towards assumption-free estimation of causal estimands in the past few decades. However this explosion in research has resulted in a breadth of inferential strategies that both create opportunities for more reliable inference as well as complicate the choices that an applied researcher has to make and defend. Relatedly, researchers advocating for new methods typically compare their method to (at best) 2 or 3 other causal inference strategies and test using simulations that may or may not be designed to equally tease out flaws in all the competing methods. The causal inference data analysis challenge, "Is Your SATT Where It's At?", launched as part of the 2016 Atlantic Causal Inference Conference, sought to make progress with respect to both of these issues. The researchers creating the data testing grounds were distinct from the researchers submitting methods whose efficacy would be evaluated. Results from over 30 competitors in the two parallel versions of the competition (Black Box Algorithms and Do It Yourself Analyses) are presented along with post-hoc analyses that reveal information about the characteristics of causal inference strategies and settings that affect performance. The most consistent conclusion was that the automated (black box) methods performed better overall than the user-controlled methods across scenarios.

Combining Observational and Experimental Data to Find Heterogeneous Treatment Effects
Alex Peysakhovich, Facebook

Every design choice will have different effects on different individuals. Modern research in machine learning tries to estimate these heterogeneous effects. This type of procedure is very data hungry (especially when potential unit-level covariates are high dimensional) and, unfortunately, experimentation is often costly. However, we increasingly have extremely large observational data sets available. Observational data, however, suffers from all sorts of omitted variable biases. We propose a method to combine these observational and experimental data to estimate heterogeneous treatment effects. First, we use observational time series data to estimate a mapping from covariates to unit-level effects. These estimates are likely biased but under some conditions the bias preserves unit-level relative
rank orderings. If these conditions hold, we only need sufficient experimental data to identify a monotonic, one-dimensional transformation from observationally predicted treatment effects to real treatment effects. This reduces power demands greatly and makes the detection of heterogeneous effects much easier. As an application, we show how our method can be used to improve Facebook page recommendations.

Atlantic Causal Inference Conference 2017: Session Abstracts

Bayesian Causal Forests
Richard Hahn, University of Chicago

In this talk I will describe a semi-parametric Bayesian regression model for estimating heterogeneous treatment effects from observational data. Standard nonlinear regression models, which may work quite well for prediction, can yield badly biased estimates of treatment effects when fitted to data with strong confounding. The new Bayesian causal forest model is able to eliminate this adverse bias by jointly modeling the treatment and the response conditional on control variables. Two empirical illustrations are given, analyzing the impact of smoking on medical expenditures and the impact of abortion laws on future crime rates.

Counterfactual Prediction using Deep Instrumental Variables Networks
Greg Lewis, Microsoft

We are in the middle of a remarkable rise in the use and capability of artificial intelligence. Much of this growth has been fueled by the success of deep learning architectures: models that map from observables to outputs via multiple layers of latent representations. These deep learning algorithms are effective tools for unstructured prediction, and they can be combined in AI systems to solve complex automated reasoning problems. This paper provides a recipe for combining ML algorithms to solve for causal effects in the presence of instrumental variables – sources of treatment randomization that are conditionally independent from the response. We show that a flexible IV specification resolves into two prediction tasks that can be solved with deep neural nets: a first-stage network for treatment prediction and a second-stage network whose loss function involves integration over the conditional treatment distribution. This Deep IV framework imposes some specific structure on the stochastic gradient descent routine used for training, but it is general enough that we can take advantage of off-the-shelf ML capabilities and avoid extensive algorithm customization. We outline how to obtain out-of- sample causal validation in order to avoid over-fit. We also introduce schemes for both Bayesian and frequentist inference: the former via a novel adaptation of dropout training, and the latter via a data splitting routine.

Wednesday May 24, 11am – 12:15pm

Causal Inference for Randomized Trials
Organizer: Ashley Naimi, University of Pittsburg
Moderator: Enrique Schisterman, NICHD

Estimating the Effect of Continued Breastfeeding on Infant Hospitalizations in a Cluster Randomized Encouragement Trial
Mireille Schnitzer, University of Montreal

Breastfeeding is considered best practice in early infant feeding, and is recommended by most major health organizations. However, due to the impossibility of directly allocating breastfeeding as a randomized intervention, no direct experimental evidence is available. The PROmotion of Breastfeeding Intervention Trial [Kramer et al 2001, JAMA 285(4):413-20] was a cluster-randomized trial that sought to evaluate the effect of a hospital program that encouraged and supported breastfeeding, thereby producing indirect evidence of its protective effect on infant infections and hospitalizations. We use Longitudinal Targeted Maximum Likelihood Estimation (LTMLE) to estimate the effect of different durations of breastfeeding (a longitudinal exposure) on the number of periods of hospitalization throughout the first year after birth. Because hospitalizations may also affect the continuation of breastfeeding, we consider them a time varying confounder. We adapt a non-targeted regression estimator as well as LTMLE to take into account an outcome that is partially determined by time- varying confounders and the clustering that arises from the nature of the cluster randomized study design.

Generalizing the Adjusted Per-protocol Treatment Effect using Inverse Probability Weights
Haidong Lu, University of North Carolina

Intent-to-treat comparisons of randomized trials provide asymptotically-consistent estimators of the effect of treatment assigned, without regard to compliance. However, decision-makers often wish to know the effect of the treatment protocol, which, under additional assumptions, can be consistently estimated by a per-protocol comparison. Moreover, decision-makers may also wish to know the effect of treatment assignment or treatment protocol in a user-specified target population other than the sample in which the trial was fielded. We consider the use of inverse probability weighting to simultaneously address generalizability and the per-protocol effects. We provide a worked example of estimating the effect of the treatment protocol in the ACTG 5095 trial [Gulick et al 2006, JAMA 296(7): 769-781], generalizing results to the US population diagnosed with HIV as reported by the CDC. We first replicate the intent-to-treat estimate for the effect of assigned treatment (hazard ratio=0.91;
95% confidence interval: 0.69, 1.19). We will present estimates of the per-protocol effect of the treatment plan that adjusts for noncompliance and dropouts by constructing inverse probability weights. Then we will present estimates of the results generalized to the target population for both the intent-to-treat and per-protocol effect estimates using inverse probability-of-sampling weights. The ACTG 5095 study was a 2-arm trial comparing 383 adults provided 4-drug antiretroviral regimen with 382 provided standard 3-drug regimen, and followed for a median of

3 years for virologic failure. The 765 patients had a median baseline age of 37 and 620 were male.

G Computation for Compliance Adjustment in Randomized Trials: An Example using the Effects of Aspirin on Gestation and Reproduction Trial
Ashley Naimi, University of Pittsburgh

First trimester miscarriage is the most common pregnancy complication in the United States. Available treatment/prevention modalities are costly, invasive, and may carry an element of risk to the unborn fetus. Aspirin is a well-tolerated over-the-counter drug that may increase the live birth rate among women at high risk of first trimester miscarriage. The Effects of Aspirin on Gestation and Reproduction [EAGeR; Schisterman et al 2014 Lancet 384(9937): 29-36] trial sought to evaluate the role of aspirin on pregnancy outcomes among 1,228 high-risk women. The ITT parameter of aspirin assignment on the live birth rate was 5.1% (95% CI -0.8 to 11.0). However, there was a non-trivial degree of non-compliance with study protocol. We implement the parametric G computation algorithm to adjust for non-compliance in the EAGeR trial, and quantify the compliance-adjusted effects of aspirin on fetal loss and live birth.

Discussant: Robert Platt, McGill University

Wednesday May 24, 11am – 12:15pm

Survey Sampling and Causal Inference
Organizer: Peng Ding, UC Berkeley

Balancing Covariates via Propensity Score Weighting
Fan Li, Duke University

Covariate balance is crucial for unconfounded descriptive or causal comparisons. However, lack of balance is common in observational studies. This article considers weighting strategies for balancing covariates. We define a general class of weights—the balancing weights—that 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. This class unifies existing weighting methods, including commonly used weights such as inverse-probability weights as special cases. General large-sample results on nonparametric estimation based on these weights are derived. 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 desirable 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. Two applications illustrate these methods and compare them with other approaches.

Using Standard Tools from Finite Population Sampling to Improve Causal Inference forComplex Experiments
Tirthankar Dasgupta, Rutgers University

This article considers causal inference for treatment contrasts from a randomized experiment using potential outcomes in a finite population setting. Adopting a Neymanian repeated sampling approach that integrates such causal inference with finite population survey sampling, an inferential framework is developed for general mechanisms of assigning experimental units to multiple treatments. This framework extends classical methods by allowing the possibility of randomization restrictions and unequal replications. Novel conditions that are "milder" than strict additivity of treatment effects, yet permit unbiased estimation of the finite population sampling variance of any treatment contrast estimator, are derived. The consequences of departures from such conditions are also studied under the criterion of minimax bias, and a new justification for using the Neymanian conservative sampling variance estimator in experiments is provided. The proposed approach can readily be extended to the case of treatments with a general factorial structure.

Atlantic Causal Inference Conference 2017: Session Abstracts

Efficient Estimation of Sample Average Treatment Effects
Yotam Shem-Tov, UC Berkeley

Since Neyman 1923, the question of how to conduct inference for the Sample Average Treatment Effect (SATE) using the difference-in-means estimator has remained open. No consistent variance estimator exists, and various conservative ones have been suggested. We show that when the estimand of interest is the Sample Average Treatment Effect of the Treated (SATT or SATC for controls), a consistent variance estimator exists. Although these estimands are equal to the SATE both in expectation and asymptotically, potentially large difference in both efficiency and coverage can occur by the change of estimand, even asymptotically. We provide analytical results, simulations, and a real data empirical application to illustrate the gains and concerns from a change of estimand. When the estimand of interest is the SATT (or SATC), even a conservative confidence interval for the SATE can provide incorrect coverage. We derive new variance formulas that provide both efficiency gains and correct coverage for the SATT (and SATC).

General forms of finite population central limit theorems with applications to causal inference
Peng Ding, UC Berkeley

Frequentists' inference often delivers point estimators associated with confidence intervals or sets for parameters of interest. Constructing the confidence intervals or sets requires understanding the sampling distributions of the point estimators, which, in many but not all cases, are related to asymptotic Normal distributions ensured by central limit theorems. Although previous literature has established various forms of central limit theorems for statistical inference in super population models, we still need general and convenient forms of central limit theorems for some randomization-based causal analysis of experimental data, where the parameters of interests are functions of a finite population and randomness comes solely from the treatment assignment. We use central limit theorems for sample surveys and rank statistics to establish general forms of the finite population central limit theorems that are particularly useful for proving asymptotic distributions of randomization tests under the sharp null hypothesis of zero individual causal effects, and for obtaining the asymptotic repeated sampling distributions of the causal effect estimators. The new central limit theorems hold for general experimental designs with multiple treatment levels and multiple treatment factors, and are immediately applicable for studying the asymptotic properties of many methods in causal inference, including instrumental variable, regression adjustment, rerandomization, clustered randomized experiments, and so on. Previously, the asymptotic properties of these problems are often based on heuristic arguments, which in fact rely on general forms of finite population central limit theorems that have not been established before. Our new theorems fill this gap by providing more solid theoretical foundation for asymptotic randomization-based causal inference.

Wednesday May 24, 1:30pm – 2:45pm

New Methods for Digital Experimentation
Organizer: Jas Sekhon, UC Berkeley

Adaptive Field Experiments using Bayesian Optimization
Konstantin Kashin, Facebook

Online experiments ("A/B tests") are the workhorse of modern Internet development, yet these experiments are generally limited to evaluating the effects of only one or two variants. In many cases, however, we are interested in evaluating the effects of thousands or a potentially infinite number of possible interventions, such as treatments parametrized by continuous variables, or dynamic contextual policies that map particular states to different actions. I will discuss a new approach to large-scale field experimentation using Gaussian process regression models and Bayesian optimization. Using empirical examples, I will show how we are able to effectively make predictions about yet-to-be- observed treatments, and make substantial improvements to applications ranging from optimizing mobile software for emerging markets to improving machine learning systems.

Trustworthy Analysis of Online A/B Tests: Pitfalls, Challenges and Solutions
Alex Deng, Microsoft

A/B tests (or randomized controlled experiments) play an integral role in the research and development cycles of technology companies. As in classic randomized experiments (e.g., clinical trials), the underlying statistical analysis of A/B tests is based on assuming the randomization unit is independent and identically distributed (i.i.d.). However, the randomization mechanisms utilized in online A/B tests can be quite complex and may render this assumption invalid. Analysis that unjustifiably relies on this assumption can yield untrustworthy results and lead to incorrect conclusions. Motivated by challenging problems arising from actual online experiments, we propose a new method of variance estimation that relies only on practically plausible assumptions, is directly applicable to a wide of range of randomization mechanisms, and can be implemented easily. We examine its performance and illustrate its advantages over two commonly used methods of variance estimation on both simulated and empirical datasets. Our results lead to a deeper understanding of the conditions under which the randomization unit can be treated as i.i.d. In particular, we show that for purposes of variance estimation, the randomization unit can be approximated as i.i.d. when the individual treatment effect variation is small; however, this approximation can lead to variance under-estimation when the individual treatment effect variation is large.

Empirical Bayes Estimators for Online Experiments
Drew Dimmery, Facebook

Online experiments are often thought of as being characterized by a large sample size; in many such cases it can be advantageous to consider a high cardinality of treatments (e.g., in the case of factorial designs in which treatments are parameterized by multiple factors with many levels). Current practice is, generally, to analyze these experiments in the same way as low cardinality treatment regimes. We show that this approach is inadmissible, and we develop an Empirical Bayes Stein-type estimator with various desirable properties. In addition to dominating the traditional approach in terms of mean squared error, our estimator is robust (it is not strongly reliant on a particular distribution of treatment effects). Moreover, we show that modern methods for experimentation--- particularly multi-armed bandit optimization---where treatment allocations adapt to prior responses, can benefit from considering EB estimates, rather than a maximum likelihood estimate that treats arms as independent. We conclude with analyses of several large scale field experiments conducted on Facebook.

Discussant: Erin Hartman, UCLA

Wednesday May 24, 2:50pm – 4:05pm

Generalizing Treatment Effects from One or a Collection of Randomized Trials to a Target Population in the Presence of Treatment Effect Heterogeneity: Identification, Estimation and Sensitivity Analysis
Organizer: Elizabeth Stuart, Johns Hopkins
Chair: Hwanhee Hong, Johns Hopkins

Generalizing Study Results: A Potential Outcomes Perspective
Catherine Lesko, Johns Hopkins

This talk discusses identification conditions for the target population average treatment effect (TATE) using data from a study sample, which include conditions required for both internal and external validity. This talk clarifies when standardization or weighting can be used to estimate TATE.

Sensitivity Analyses for Partially or Fully Unobserved Effect Modifiers when Calibrating the Effect from a Randomized Trial to a Target Population
Trang Quynh Nguyen, Johns Hopkins

Assuming internal validity, we propose methods to assess the sensitivity of TATE estimates based on a randomized trial to a treatment effect modifier observed in the trial but not in the target population, or to one that is observed in neither the trial nor the target population. These methods use and combine effect-modification-outcome-model and probability-weighting approaches.

Transporting the Results of Multiple Randomized Controlled Trials to a Target Population: Towards Causally Interpretable Meta-Analysis
Issa J. Dahabreh, Brown University

This talk tackles the use of multiple trials to infer treatment effects for a target population. This work evaluates the performance of different meta-analysis TATE estimators that are outcome-model-based, probability-of- participation-based and doubly robust.

Discussant: Robert Platt, McGill University

Wednesday May 24, 2:50pm – 4:05pm

Innovations in discovering effect modification
Organizer: Ashkan Ertefaie, University of Rochester

A central question in studying treatments is "what works best for whom?" For many treatments, what works best depends on a person's characteristics. Tailoring treatments based on patients' characteristics can, potentially, improve health outcomes, reduce side effects and reduce treatment costs. In the proposed session, we discuss some novel developments in discovering treatment effect heterogeneities using machine learning approaches that are suitable for applications with Big data. We have four speakers that present their most recent work in this area.

Causal Interaction in Factorial Experiments: Application to Conjoint Analysis
Kosuke Imai, Princeton University

Optimal Policy Learning
Stefan Wager, Stanford University

Post-Selection Inference for the Effect Modifiers Selection
Qingyuan Zhao, University of Pennsylvania

Discussant: Susan Athey, Stanford University

Wednesday May 24, 4:20pm – 5:35pm

Modern Advances in Instrumental Variable Methods
Organizers: Luke Keele, Georgetown University; Edward Kenedy, Carnegie Mellon University

Survivor-Complier Causal Effects in a Study of Prompt ICU Admission with Selection on Treatment
Luke Keele, Georgetown University

Pre-treatment selection or censoring ('selection on treatment') is very common in instrumental vari- able studies: it can occur when two treatment levels are compared ignoring the third option of neither treatment, in 'censoring by death' settings where treatment is only defined for those who survive long enough to receive it, or in general in any study where the treatment is only defined for a subset of the population. Unfortunately, standard analyses are biased in the presence of such selection. In this work we propose a novel estimand for these settings, called the survivor-complier causal effect. Although the effect is not identified under standard assumptions, it is possible to construct bounds. We derive these bounds, and propose a doubly robust and semiparametric efficient approach for estimating them; importantly, our methods allow for high-dimensional confounding adjustment, as well as valid infer- ence even after employing machine learning. We apply the methods in a UK cohort study of critical care patients to examine the mortality effects of prompt admission to the intensive care unit, using ICU bed availability as an instrument. In this illustration the treatment of prompt admission is only defined for those patients who are accepted for admission.

Generalizing local effects with sharp instruments
Edward Kennedy, Carnegie Mellon University

Abstract: It is well-known that, without restricting treatment effect heterogeneity, instrumental variable (IV) methods only identify "local" effects among compliers, i.e., subjects who take treatment only when encouraged by the IV. Local effects are controversial since they apparently only apply to this unidentified subgroup; this has led many to denounce these effects as having no policy relevance. However, we show that such pessimism is not always warranted: complier effects can be closely linked to effects in identifiable subgroups. Specifically, we derive sharp bounds on identifiable subgroup effects, and show that these bounds can collapse to a single point even for arbitrarily weak IVs. Hence strength, the usual measure of instrument quality, is inadequate in this respect. Therefore we present a new measure of how well IVs yield identifiable subgroup effects, called "sharpness". We propose a corresponding nonparametric yet efficient estimator, along with a novel form of sample splitting (called "triple splitting") to accommodate nonregularity. Importantly, using influence functions and triple splitting allows for fast parametric convergence rates, even in the presence of nonparametric nuisance estimation and nonsmooth parameters. Triple splitting has applications in other nonregular problems, include optimal treatment regime estimation and risk estimation for infinite-dimensional causal parameters.

TMLE for Marginal Structural Models Based on an Instrument
Boriska Toth, UC Berkeley

We consider estimation of a causal effect of a possibly continuous treatment when treatment assignment is potentially subject to unmeasured confounding, but an instrumental variable is available. Our focus is on estimating heterogeneous treatment effects, so that the treatment effect can be a function of an arbitrary subset of the observed covariates. One setting where this framework is especially useful is with clinical outcomes. Allowing the causal dose-response curve to depend on a subset of the covariates, we define our parameter of interest to be the projection of the true dose-response curve onto a user-supplied working marginal structural model. We develop a targeted minimum loss-based estimator (TMLE) of this estimand. Our TMLE can be viewed as a generalization of the two-stage regression method in the instrumental variable methodology to a semiparametric model with minimal assumptions. The asymptotic efficiency and robustness of this substitution estimator is outlined. Through detailed simulations, we demonstrate that our estimator's finite-sample performance can beat other semiparametric estimators with similar asymptotic properties. In addition, our estimator can greatly outperform standard approaches. For instance, the use of data-adaptive learning to achieve a good fit can lead to both lower bias and lower variance than for an incorrectly specified parametric estimator. Finally, we apply our estimator to a real dataset to estimate the effect of parents' education on their infant's health.

Testing Endogeneity with Possibly Invalid Instruments and High Dimensional Covariates
Hyunseung Kang, University of Wisconsin, Madison

The Durbin-Wu-Hausman (DWH) test is a commonly used test for endogeneity in econometrics, specifically instrumental variables (IV) regression. Unfortunately, the DWH test depends, among other things, on assuming all the instruments are valid, a rarity in practice. In this paper, we show that the DWH test often has distorted size even if one IV is invalid. Also, the DWH test may have low power when many, possibly high dimensional, covariates are used to make the instruments more plausibly valid. To remedy these shortcomings, we propose a new endogeneity test which has proper size and better power when invalid instruments and high dimemsional covariates are present; in low dimensions, the new test is optimal in that its power is equivalent to the ”oracle” DWH test's power that knows which instruments are valid. The paper concludes with a simulation study of the new test with invalid instruments and high dimensional covariates.

Discussant: Elizabeth Ogburn, Hopkins

Thursday May 25, 9:45am – 11am

Causal Inference in Air Pollution Epidemiology
Organizer: Richard Smith, University of North Carolina

The Effects of Policy-Driven Air Quality Improvements on Children's Respiratory Health
Kiros Berhane, University of Southern California

This presentation will be based on two papers published in NEJM (on lung function) and JAMA (on bronchitic symptoms) that were part of a larger project funded by HEI. The objective with both papers was to document the improved health effects of recent changes in air quality regulations in California. The talk will cover a number of statistical issues associated with such inferences and will provide the background for a more formal development in the framework of causal inference.

Efforts to quantify the causal effect of fine particulate matter on mortality
Zhulin He, Department of Statistics, Iowa State University, and
Zhengyuan Zhu, Department of Statistics, Iowa State University

Outdoor air pollution is a major environmental health problem affecting many countries. Many studies have indicated that higher PM2.5 exposure levels are associated with increases in mortality. However, it remains a challenging task to quantify the causal effect of PM2.5 on mortality. Due to confounding factors, the estimation of the effect of PM2.5 on mortality varies substantially from one study to another. In this talk, we present a counterfactual approach to estimate the causal effect of time-varying PM2.5 exposure on mortality. In particular, by utilizing a directed acyclic graph description of the causal relationship, we propose a structural nested mean model under a spatial-temporal setting to estimate the causal effect of PM2.5 on mortality. This is an ongoing project. Currently we only have summarized mortality information, and our analysis is performed based on a simulated individual-level mortality data combined with PM2.5 data from EPA and meteorological data from NOAA. We plan to apply the methodology to Medicare individual-level mortality data when they become available.

Progress in Automated Inference and Estimation of Causal Concentration-Response Functions in Air Pollution Epidemiology

Tony Cox, Cox Associates and University of Colorado

Current causal graph-learning algorithms, partial dependence plots, and related computational statistical techniques can be used to automatically identify predictive causal relationships (necessary bit not sufficient for manipulative or mechanistic causation) and to estimate or bound the quantitative fraction of health effects caused by controllable conditions, such as current exposure concentrations. This talk discusses different concepts of causation (associational, counterfactual, predictive, computational, manipulative, and mechanistic) and surveys algorithms for estimating and validating predictive causal exposure-response functions from exposure-response data with relevant covariates, and for characterizing remaining uncertainties. We discuss and illustrate the possibility of automating valid causal inference by applying currently available machine-learning, causal discovery, and artificial intelligence technology. Principles and the practical software for fully automated causal analysis and modeling from data are illustrated using epidemiological data on air pollution health effects, analyzed with the free Causal Analytics Toolkit (CAT) software for users of Excel in Windows. (https://regulatorystudies.columbian.gwu.edu/causal-analytics-toolkit-cat-assessing-potential-causal-relations-data).

Discussant: Richard Smith, University of North Carolina

Thursday May 25, 11:15am – 12:30pm

Different Modes of Inference under Interference
Organizer: Laura Forastiere, University of Florence

Randomization Inference in Networks
Dean Eckles, MIT

Social and behavioral scientists are interested in testing of hypotheses about spillovers (i.e. interference, exogenous peer effects) in social networks. However, when there is a single network, this is complicated by lack of independent observations. We explore Fisherian randomization inference as an approach to exact finite-population inference, where the main problem is that the relevant hypotheses are non-sharp null hypotheses. Fisherian randomization inference can be used to test these hypotheses either by (a) making the hypotheses sharp by assuming a model for direct effects or (b) conducting conditional randomization inference such that the hypotheses are sharp. I present both of these approaches, the latter of which is developed in Aronow (2012) and Athey, Eckles & Imbens (2015). I illustrate these methods with application to a large voter turnout experiment on Facebook.

A Folk Theorem on Interference in Experiments
Fredrik Savje, UC Berkeley

It is often assumed that units in experiments do not interference with each other—the stable unit treatment variation assumption (sutva). However, experimenters realize that humans interact in complex ways and that the assumption holds exactly only in special cases. A widely held belief is that a little interference does not hurt our inferences; even if the sutva is violated, our conclusions should remain largely valid as long as the assumption holds approximately. This paper provides rigorous justification for that intuition. We show that unless there is rampant interference, treatment effects can be estimated consistently and without bias even if nothing is known about the interference structure. The estimand—the interference marginalized average treatment effect—is a generalization of the commonly investigated average treatment effect to a setting with interference. Both estimands can be interpreted as the expected effect of administering treatment to the units in the sample under the current experimental setting. Importantly for prac- titioners, we show that if one erroneously assumes sutva in a setting with limited (or even moderate) interference, standard estimators will nevertheless be consistent and retain their interpretation as the expected treatment effect. The corresponding confidence intervals and p-values may, however, be inaccurate.

Estimation and Testing in Two-Stage Randomized Designs with Interference
Avi Feller, UC Berkeley

Two-stage randomization is a powerful design for estimating treatment effects in the presence of social interactions. Our motivating example is a two-stage randomized trial evaluating an intervention to reduce student absenteeism in the School District of Philadelphia. In that experiment, households with multiple students were first assigned to treatment or control; then, in treated households, one student was randomly assigned to treatment. Using this example, we highlight key considerations for estimation and testing in these designs. For estimation, we address additional complications that arise when household sizes vary, draw connections with linear regression, and discuss options for incorporating covariates. For testing, we extend existing methods to allow for more powerful tests in this setting. We apply these methods to the attendance study and find large, substantively meaningful spillover effects.

Exploring encouragement, spillover and attendance effects in a field experiment on museums attendance of high school teens using principal stratification
Fabrizia Mealli, University of Florence

Exploration of causal mechanisms is often important for researchers and policymakers to understand how an intervention or an incentive works and how it can be improved. We explore causal mechanisms characterizing a cluster randomized experiment implemented in some high schools in Florence and designed to evaluate the impact on art museum visits of three different incentives aimed at stimulating attendance to a main museum in Florence, Italy - Palazzo Vecchio. Incentives can be viewed as encouragements and raise the problem of non-compliance, as museum attendance is not enforced and therefore not under experimental control. Encouragements give also rise to a variety of mechanisms, particularly when encouragement is assigned at the cluster level, classrooms in our study. Social interactions among students within the same cluster can result in spillover effects. Understanding the "direct" effect of an incentive and its "indirect" effects through spillover effects and through museum attendance would give a better insight into the different incentives and it could be compelling for planning the scaling-up phase of the program. We use the principal stratification framework to define stratum-specific causal effects, that is, effects for specific latent subpopulations, defined by the joint potential compliance statuses under thed three encouragement conditions. Estimation of causal estimands is performed with Bayesian inferential methods.

Thursday May 25, 11:15am – 12:30pm

Balance at Baseline in Experiments and Observational Studies: New Measures, New Implications
Organizer: Ben Hansen, University of Michigan

The Classification Permutation Test: A Nonparametric Test for Equality of Multivariate Distributions
Johann Gagnon-Bartsch, University of Michigan
* 2016 Tom Ten Have Winner *

The gold standard for identifying causal relationships is a randomized controlled experiment. In many applications in the social sciences and medicine, the researcher does not control the assignment mechanism and instead may rely upon natural experiments, regression discontinuity designs, RCTs with attrition, or matching methods as a substitute to experimental randomization. The standard testable implication of random assignment is covariate balance between the treated and control units. Covariate balance is therefore commonly used to validate the claim of "as-if" random assignment. We develop a new nonparametric test of covariate balance. Our Classification Permutation Test (CPT) is based on a combination of classification methods (e.g. logistic regression or random forests) with Fisherian permutation inference. The CPT is guaranteed to have correct coverage and is consistent under weak assumptions on the chosen classifier. To illustrate the gains of using the CPT, we revisit four real data examples: Lyall (2009); Green and Winik (2010); Eggers and Hainmueller (2009); and Rouse (1995). Monte Carlo power simulations are used to compare the CPT to two existing nonparametric tests of equality of multivariate distributions.

New multivariate tests for assessing covariate balance in matched observational studies
Hao Chen, University of California, Davis

We propose new tests for assessing whether covariates in a treatment group and matched control group are balanced in observational studies. The tests exhibit high power under a wide range of multivariate alternatives, some of which existing tests have little power for. The asymptotic permutation null distributions of the proposed tests are studied and the p-values calculated through the asymptotic results work well in finite samples, facilitating the application of the test to large data sets. The tests are illustrated in a study of the effect of smoking on blood lead levels.

Balancing Basu's elephants: Limiting bias of subgroup effect estimates, particularly in stratified cluster randomized trials
Mark Fredrickson, University of Illinois, Urbana-Champaign

It is generally difficult to state which covariates are important to balance in a randomized trial, because the relationship between background variables and potential outcomes is not known. One important class of background var iables stands out: size variables, variables measuring cluster sizes, sampling weights, or subgroup membership. We show how imbalances of size can have important contributions to first- and second-order biases in design-based estimates of average treatment effects. Using the randomization mechanism, the biases can be very meaningfully controlled, and if post-randomization imbalance is found to fall within certain limits, then a suitable conditional inference may also serve to limit bias and

improve estimation of sampling variability. Our analysis generates guidance on whether it's necessary to collapse small categories of a subgrouping variable before estimating subgroup effects and suggests straightforward mechanisms for improving the statistical properties of simple and widely used effect estimators. These methods constitute a modest elaboration of common approaches to study design and to average treatment effect estimation, and are easily implemented with common software.

Appraising Covariate Balance as Part of Experimental Design
Kari Lock Morgan, Pennsylvania State University

Appraising covariate balance is a cornerstone of causal inference for observational studies, yet it is usually not a salient feature of experimental design. The field of experimental design is devoted primarily to obtaining covariate balance, but appraising covariate balance in the design phase of an experiment is typically bypassed, either because of an assumption that randomization will yield adequate balance or an assumption that nothing can be done in the design phase to correct for observed imbalance. However, neither of these assumptions should be taken for granted. Covariate imbalance can happen despite randomization, and when imbalance is observed before treatment is administered, units can (and should) be rerandomized to treatment groups. This talk will focus on considerations and recommendations regarding the appraisal of covariate balance for this purpose, to decide whether or not to rerandomize units, and will illuminate potential unintended consequences if covariate balance is naively assessed.

Thursday May 25, 1:30pm – 2:45pm

Causal Inference with Unobserved Confounders
Organizer: Eric Tchetgen Tchetgen, Harvard and Wang Miao, Beijing University

Unification of the Instrumental Variable Approach for Causal Inference and Missing Data
Eric Tchetgen Tchetgen, Harvard

Unobserved confounding is a well-known threat to causal inference with observational data. Likewise, selection bias can arise in the presence of missing data if there is an unobserved common cause of the nonresponse process and the potentially unobserved outcome . An instrumental variable for unobserved confounding (IV-C) is a pre-exposure correlate of exposure known to only affect the outcome through its association with exposure. Likewise an instrumental variable for missing data (IV-M) is a predictor of missingness which is otherwise independent of the outcome in the underlying population. We give general necessary and sufficient conditions for nonparametric identification with an IV in settings (IV-C) or (IV-M), thus providing a unification of identification for causal inference and missing data with an IV. The approach equally applies for discrete or continuous IV and outcome. Interestingly, the proposed approach provides an elegant solution to the identification problem of the marginal effect of treatment on the treated with an IV which has been a longstanding problem in causal inference. For statistical inference incorporating high dimensional covariates, we present generalizations of inverse-probability weighting, outcome regression and doubly robust estimation with an instrumental variable that equally apply to IV-C and IV-M. In case identification fails, We describe novel IV bounds for the nonidentified parameter of interest and corresponding methods to account for all sources of uncertainty. We illustrate the approach with simulation studies and several empirical examples.

Bounded, Efficient and Triply Robust Estimation of Average Treatment Effects using Instrumental Variables
Linbo Wang, Harvard

Instrumental variables (IVs) are widely used for estimating causal effects in the presence of unmeasured confounding. Under the standard IV model, however, the average treatment effect (ATE) is only partially identifiable. To address this, we propose novel assumptions that allow for identification of the ATE. Our identification assumptions are clearly separated from model assumptions needed for estimation, so that researchers are not required to commit to a specific observed data model in establishing identification. We then construct multiple estimators that are consistent under three different observed data models, and triply robust estimators that are consistent in the union of these observed data models. We pay special attention to the case of binary outcomes, for which we obtain bounded estimators of the ATE that are guaranteed to lie between -1 and 1. Our approaches are illustrated with simulations and a data analysis evaluating the causal effect of education on earnings.

Optimal Criteria to Exclude the Surrogate Paradox and Sensitivity Analysis
Lan Liu, University of Minnesota

When the primary outcome is hard to collect, surrogate endpoint is typically used as a substitute. However, even when the treatment has a positive average causal effect (ACE) on the surrogate endpoint, which also has a positive ACE on the primary outcome, it is still possible that the treatment has a negative ACE on the primary outcome. Such a phenomenon is called the surrogate paradox and greatly challenges the use of surrogate. In this paper, we provide criteria to exclude the surrogate paradox for both the strong, and non-strong surrogates. Our criteria are optimal in the sense that theyare sufficient and "almost necessary" to exclude the paradox: if the conditions are satisfied, the surrogate paradox is guaranteed to be absent while if the conditions fail, there exists a data generating process with surrogate paradox that can generate the same observed data. That is, our criteria capture all the information in the observed data to exclude the surrogate paradox rather than relying on unverifiable distributional assumptions.

Testing Causative Hypotheses in the Presence of Unmeasured Confounding
Wang Miao, Beijing University Discussant: James Robins, Harvard

Causal effects are often not identified in the presence of unmeasured confounding. In a effort to mitigate this difficulty, I plan to focus on testing the null hypothesis of no causal effect. Rejection of such a null hypothesis indicates causation between the treatment and outcome. When confounders are completely observed, one can test this null hypothesis upon controlling for observed confounders and assessing the conditional independence between the treatment and outcome. However, this approach is clearly not available when a confounder is not observed and we propose a novel measure of the conditional independence which recovers a valid test of the causal null. The proposed approach yields interesting results about Simpson's paradox. I will give several examples in which one has nontrivial power to reject the null hypothesis without external information, which in a sense resolves Simpson's paradox.

Thursday May 25, 2:45pm – 4pm

Interference and Social Networks
Organizer: Lan Liu, University of Minnesota – Twin Cities

Modeling Interference Via Symmetric Treatment Decomposition
Ilya Shpitser, Johns Hopkins

We develop a new approach to decomposing the spillover effect into direct (also known as the contagion effect) and indirect (also known as the infectiousness effect) components that extends the DAG based treatment decomposition approach to mediation found in the literature to causal chain graph models. We show that when these components of the spillover effect are identified in these models, they have an identifying functional, which we call the symmetric mediation formula, that generalizes the mediation formula in DAGs. We further show that, unlike assumptions in classical mediation analysis, an assumption permitting identification in our setting leads to restrictions on the observed data law, making the assumption falsifiable (but not testable). Finally, we discuss statistical inference for the components of the spillover effect in the special case of two interacting outcomes, and discuss a maximum likelihood estimator, and a doubly robust estimator.

Estimation of Monotone Treatment Effects under Interference
David Choi, Carnegie Mellon University

Randomized experiments on social networks pose statistical challenges, due to the possibility of interference between units. We propose new methods for finding confidence intervals on the attributable treatment effect in such settings. The methods do not require partial interference, but instead require an identifying assumption that is similar to requiring "no defiers" or non-negative treatment effects. Network or spatial information can be used to customize the test statistic; in principle, this can increase power without making formal assumptions on the data generating process.

Simulation-Based Sensitivity Analysis for Interference in Observational Studies with Unmeasured Links

Laura Forastiere, University of Florence

Oftentimes estimates on the effect of a treatment are based on the implicit assumption of no interference between units; that is, a subject's value on the response depends only on the treatment to which that subject is assigned, not on the treatment assignments of other subjects. However, interference is common in many social settings (e.g., schools and networks). When data are wrongly analyzed under the "no-interference assumption," very misleading inferences can result. The bias depends on the level of interference but also on the degree of association between a unit's treatment and the treatment received by his neighbors, i.e., units he interacts with. In observational studies, information on links between units is usually unavailable and interference cannot be taken into account. In a way, the neighborhood treatment can be seen as an unmeasured confounder and a proper sensitivity analysis should be of substantive importance. We propose to face this issue by developing a Bayesian simulation-based sensitivity analysis to the violation of the "no-interference assumption".

A Bayesian simulation-based approach to sensitivity analysis repeatedly i) draws a set of sensitivity parameters from a prior distribution, ii) simulates potential confounders, and iii) reestimates the posterior distribution of the effect of interest after adjusting for the simulated confounders.

When the source of potential bias is interference a challenging issue concerns the specification of the model for the confounder: we should posit a model which follows the complex network structure.

We propose a model to generate the unmeasured links, which carries our belief on the level of interference and on the level of association between the individual and the neighborhood treatments. If we assume interference to operate only through a function of the vector of neighbors' treatments, after a network is drawn we can compute such function and estimate the direct effect of the treatment taking interference into account. Different functions can be used. This approach has the additional advantage of adjusting for neighborhood and network covariates.

Discussant: Alex Volfovsky, Duke University