Poster Session on Tuesday May 22\textsuperscript{nd}, 2018

Poster room: CMU, Rangos Auditorium, Ballroom 1 & 2 at the Jared L. Cohon University Center, CMU.

Presenting Time: 6:00 P.M. - 7:30 P.M.
Set Up Time: 2:30 P.M. - 3:00 P.M.
# Table of Content: Presenters, Titles and Poster Numbers

<table>
<thead>
<tr>
<th>Presenter(s)</th>
<th>Title</th>
<th>Poster Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altenburger, Kristen M.</td>
<td>A Principal Stratification Approach to Uncomplicate Causal Inference Complications on Social Networks</td>
<td>32</td>
</tr>
<tr>
<td>Antonello, Richard</td>
<td>Robust Unsupervised Extraction of Macrovariables</td>
<td>61</td>
</tr>
<tr>
<td>Artman, William</td>
<td>Power Analysis in a SMART Design: Sample Size Estimation for Determining the Best Dynamic Treatment Regime</td>
<td>10</td>
</tr>
<tr>
<td>Bahamyirou, Asma</td>
<td>Doubly Robust Adaptive Lasso for Effect Modifier Discovery</td>
<td>13</td>
</tr>
<tr>
<td>Ben-Michael, Eli and Feller, Avi</td>
<td>Matrix constraints for covariate balance</td>
<td>60</td>
</tr>
<tr>
<td>Bonvini, Matteo and Edward H. Kennedy</td>
<td>Efficient Estimation of the Average Treatment Effect on the Treated (ATT) with One-Sided Non-Compliance</td>
<td>11</td>
</tr>
<tr>
<td>Botelho, Anthony</td>
<td>Deep Learning, Auxiliary Data, and Randomization: Analyzing Experiments Run Within a Computerized Math Tutor</td>
<td>35</td>
</tr>
<tr>
<td>Branson, Zach</td>
<td>Is My Matched Dataset As-If Randomized, More, Or Less? Unifying the Design and Analysis of Observational Studies</td>
<td>9</td>
</tr>
<tr>
<td>Buchanan, Ashley</td>
<td>Evaluating the Effects of Attitudes on Health-Seeking Behavior among a Network of People Who Inject Drugs</td>
<td>28</td>
</tr>
<tr>
<td>Canes, Aran</td>
<td>Is the Ignorability Assumption of Propensity Score Analysis being Met in Pharmaceutical Observational Studies?</td>
<td>27</td>
</tr>
<tr>
<td>Carrion, Carlos</td>
<td>Application of Causal Bayesian Networks to Environmental Data</td>
<td>31</td>
</tr>
<tr>
<td>Chen, Ling-Wan</td>
<td>Quantile Association Model for Bivariate Survival Data</td>
<td>41</td>
</tr>
</tbody>
</table>
Chin, Alex .......................................................... 17
Central limit theorems via Stein's method for randomized experiments under interference:
Poster 52 .......................................................... 17

Comment, Leah ................................................... 18
Time-varying survivor average causal effects with semi-competing risks: Poster 16 .......... 18

D’Amour, Alexander ............................................. 19
Extrapolation parameterizations for assessing sensitivity to unmeasured confounding: Poster
46 .......................................................... 19

Fisher, Aaron ....................................................... 20
Model Class Reliance: Variable Importance when all Models are Wrong, but *Many* are
Useful: Poster 5 .................................................. 20

Fulcher, Isabel ...................................................... 21
Nonparametric identification and robust estimation of indirect causal effects in the presence
of exposure-outcome confounding: Poster 14.............................. 21

Furman, Gleb ....................................................... 22
Selecting a generalizable sample in field trials with stratification using balanced sampling: A
simulation study: Poster 62 .......................................... 22

Gellar, Jonathan ................................................. 24
Optimal matching in health policy evaluations under rolling enrollment: Poster 21........... 24

Greifer, Noah ....................................................... 25
A Moderated Nonlinear Factor Analysis Solution to the Problem of Differential Measurement
Error in Propensity Score Analysis: Poster 18.......................... 25

Hasegawa, Raiden ............................................... 26
Extended sensitivity analysis for heterogeneous unmeasured confounding with an
application to sibling studies of returns to education: Poster 34 ......................... 26

Hejazi, Nima ......................................................... 27
Robust Nonparametric Inference for Stochastic Interventions Under Multi-Stage Sampling:
Poster 49 .......................................................... 27

Heng, Siyu ............................................................ 29
Finding the Strength in a Weak Instrument in a Study of Cognitive Outcomes Produced by
Catholic High Schools: Poster 33 ...................................... 29

Hirshberg, David .................................................. 30
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augmented Minimax Linear Estimation: Generalizing Double Robust Balancing Estimators: Poster 8</td>
<td>30</td>
</tr>
<tr>
<td>Howard, Steve</td>
<td>31</td>
</tr>
<tr>
<td>Uniform non-asymptotic confidence sequences, with applications to sequential testing and estimation: Poster 50</td>
<td>31</td>
</tr>
<tr>
<td>Hunt, Lamar</td>
<td>32</td>
</tr>
<tr>
<td>Assessing Therapeutic Equivalence of Brand and Generic Drugs Using Observational Data: Poster 19</td>
<td>32</td>
</tr>
<tr>
<td>Ju, Cheng</td>
<td>33</td>
</tr>
<tr>
<td>On Adaptive Propensity Score Truncation in Causal Inference: Poster 15</td>
<td>33</td>
</tr>
<tr>
<td>Keller, Bryan</td>
<td>34</td>
</tr>
<tr>
<td>Permutation-Based Variable Selection with Random Forests: Poster 65</td>
<td>34</td>
</tr>
<tr>
<td>Kim, Yongnam</td>
<td>36</td>
</tr>
<tr>
<td>An Example where the Fundamental Problem of Causal Inference Does Not Occur: The Causal Effect of Answer Changing: Poster 2</td>
<td>36</td>
</tr>
<tr>
<td>Kim, Kwangho</td>
<td>38</td>
</tr>
<tr>
<td>Effect of low-dose aspirin on pregnancy outcomes based on nonparametric causal effects: Poster 23</td>
<td>38</td>
</tr>
<tr>
<td>Lee, Youjin</td>
<td>39</td>
</tr>
<tr>
<td>Collective problem solving, causal inference, and chain graph: Poster 57</td>
<td>39</td>
</tr>
<tr>
<td>Lei, Lihua</td>
<td>40</td>
</tr>
<tr>
<td>Statistical assessment of strict population overlap in observational studies: Poster 53</td>
<td>40</td>
</tr>
<tr>
<td>Li, Fan (Frank)</td>
<td>41</td>
</tr>
<tr>
<td>Liu, Leqi</td>
<td>42</td>
</tr>
<tr>
<td>Offline Heterogeneous Policy Evaluation for Contextual Bandits: A Causal Approach: Poster 43</td>
<td>42</td>
</tr>
<tr>
<td>Lorenzi, Elizabeth</td>
<td>43</td>
</tr>
<tr>
<td>Heterogeneous Treatment Effects with Subgroups via the Overlap Weights: Poster 45</td>
<td>43</td>
</tr>
<tr>
<td>Lu, Rui</td>
<td>44</td>
</tr>
<tr>
<td>Data Based Covariate Selection for High Dimension Low Sample Size Data: Poster 59</td>
<td>44</td>
</tr>
</tbody>
</table>
Malenica, Ivana .................................................................................................................. 45

Targeted Maximum Likelihood Estimation of Causal Effects Based on Observing a Single
Time Series: Poster 42 ........................................................................................................ 45

Martínez, Sebastián ......................................................................................................... 46

Causal, Inferential, Dynamic Network Analysis for Public Health: Poster 29 ............... 46

Mauro, Jacqueline ........................................................................................................... 48


Mishler, Alan ................................................................................................................... 49

Counterfactual Fairness in Decision Making with Risk Assessment Tools: Poster 4 .......... 49

Mitchell, Shira ................................................................................................................ 51

A retrospective control study of the Millennium Villages Project: Poster 38 .............. 51

Mogensen, Søren Wengel ............................................................................................... 52

Local independence graphs for time series: Poster 58 .................................................. 52

Montero-Hernandez, Samuel .......................................................................................... 53

Intervals of Causal Effects for Learning Causal Graphical Models: Poster 56 .............. 53

Nabi, Razieh & Dan Malinsky (presenter) ...................................................................... 54

Fair Inference On Outcomes: Poster 24 .......................................................................... 54

Nabi, Razieh ...................................................................................................................... 56

Semi-Parametric Causal Sufficient Dimension Reduction Of High Dimensional Treatments:
Poster 25 .......................................................................................................................... 56

Nabi, Razieh & Ilya Shpitser (presenter) ........................................................................ 57

Estimation of Optimal Path-Specific Policies: Poster 26 ............................................... 57

Nattino, Giovanni ............................................................................................................ 58

Model Assisted Sensitivity Analyses for Hidden Bias with Binary Outcomes: Poster 48.... 58

Nevo, Daniel .................................................................................................................... 59

On the difference method for mediation analysis in generalized linear models: Poster 12... 59

Nie, Xinkun ....................................................................................................................... 60

Quasi-Oracle Estimation of Heterogeneous Treatment Effects: Poster 37 ..................... 60

Pashley, Nicole ................................................................................................................ 61

Insights on Variance Estimation for Blocked and Matched Pairs Designs: Poster 55 .... 61

Rose, Reagan ................................................................................................................... 62
Causal Inference for Observational Data with Partially Observed and Time-Dependent Treatment Assignment: Poster 6

Sales, Adam

Fine-Grained Compliance and Treatment Effects in the Cognitive Tutor Effectiveness Trial: Poster 36

Samarasinghe, Savini

Opportunities for Causal Discovery in the Geosciences: Poster 3

Sanchez-Romero, Ruben

Causal Discovery of Feedback Networks with Functional Magnetic Resonance Imaging: Poster 22

Sherman, Eli

Identification of Personalized Path-Specific Effects: Poster 39

Siddique, Arman Alam

Causal Inference for Polypharmacy: Propensity score estimation with multiple concurrent medications: Poster 20

Starling, Jennifer

Functional BART for Causal Inference: Poster 47

Suzuki, Etsuji

Covariate balance for no confounding in the sufficient-cause model: Poster 63

Takeuchi, Yoshinori

Martingale residual-based imputation for unmeasured confounders in marginal structural Aalen additive hazard model using validation sample: Poster 54

Tan, Sarah

Probabilistic Matching: Incorporating Uncertainty to Improve Propensity Score Matching: Poster 17

Walter, Simon

Analyzing the modified outcome method for estimating heterogeneous treatment effects: Poster 40

Wang, Yixin and Blei, David

The Blessings of Multiple Causes: Poster 1

Wang, Ran
Identify Heterogeneous Treatment Effect and Confounder's Effect via L1-Regularized Nonlinear Regression: Poster 44

Wang, Jingshen

Debiased inference on treatment effect in high dimensional model: Poster 64

Wieczorek, Aleksander

Causal Time Series Classification: Poster 7
A Principal Stratification Approach to Uncomplicate Causal Inference
Complications on Social Networks: Poster 32

Abstract: In this work, we extend the principal stratification framework (Frangakis, 2002) to address treatment non-compliance specifically in social network settings. We assume a random subset of nodes are first encouraged to share information at time $t$, then they either share or don’t share, and the real estimand of interest is the spillover effect that has among the node’s friends or followers at a later time $(t+1)$. For concreteness, suppose that users are encouraged to tweet a link to all their followers (Coppock 2016) but not every unit follows their treatment assignment to tweet the link. We outline below the possible stratification of the two sets of nodes we define as: the “broadcasters” are those nodes randomly encouraged to share information, and the “followers” are those nodes that follow at least one broadcaster node, excluding any broadcasters that follow each other. Stratification of Broadcasters: Suppose a seed set of users is randomly encouraged to share information to all their friends. This assignment-to-treatment variable will be encompassed by the treatment vector $Z$ where $Z_i = 1$ if user $i$ is encouraged, and 0 otherwise. We define $S_{i,t}$ as an indicator for whether broadcaster node $i$ shares information. Then, per the usual set-up in encouragement designs, the principal strata are defined as: BB: $S_{i,t}(1)=S_{i,t}(0)=1$, users who would broadcast regardless of treatment assignment [always-takers], BN: $S_{i,t}(1)=1$, $S_{i,t}(0)=0$, users who would broadcast if treated and would not otherwise [compliers] NB: $S_{i,t}(1)=0$, $S_{i,t}(0)=1$, users who would not broadcast if treated and would otherwise [defiers] NN: $S_{i,t}(1)= S_{i,t}(0)=0$, users who would never broadcast regardless of treatment status [never-takers] Stratification of Followers: We define the treatment assignment as $Z^* = 1$ if user $i$ follows either at least one always-taker/complier or only never-takers that were encouraged to share information and 0 otherwise. We define $S^*_{i,t+1}$ as an indicator for whether a follower node $i$ receives broadcasted information. The non-existence of defiers can be enforced by design or is common to assume does not exist in most settings, and we adopt the same assumption on the broadcaster set here as well. The stratification of follower nodes receiving, R, treatment is: RR: $S^*_{i,t+1}(1) = S^*_{i,t+1}(0) = 1$, users who follow at least one
always-taker RN: $S^*_{i,t+1}(1) = 1$, $S^*_{i,t+1}(0)=0$, users who only follow compliers or compliers and never-takers NN: $S^*_{i,t+1}(1) = S^*_{i,t+1}(0)=0$ users who only follow never-takers. If user $i$ follows broadcaster users that belong to multiple distinct strata, then we justify the following stratification of followers. If user $i$ follows an always-taker and a never-taker, then user $i$ is considered an always-taker since user $i$ will always be exposed to the treatment. If user $i$ follows a never-taker and a complier, then we consider them a complier since the complier broadcaster node will determine treatment exposure. A complier is now defined a user who only follows compliers, or follows a combination of compliers and never-takers only. This works main conceptual contribution is that in network settings, the latent principal strata of nodes that are initially treated defines the subsequent principal stratification of their friends in the network.
Robust Unsupervised Extraction of Macrovariables: Poster 61

Abstract: Robust Unsupervised Extraction of Macrovariables  We take as example the El Nino phenomenon, which is known to have significant global consequences, both directly in terms of changed weather patterns, and indirectly, in terms of economic consequences. In both scientific and popular literature, El Nino is commonly described as if it were a distinct causal variable driving effects around the globe. In spite of this significance, its definition remains remarkably crude: it is defined by an average temperature deviation of more than 0.4°C for a period of 180 days in a rectangular geographic region over the equatorial Pacific known as the El Nino 3.4 region. Clearly, this is at best a good indicator measure of the true phenomenon, suggesting more a lack of knowledge about the actual causally relevant (micro-)phenomena, rather than a distinct causal variable. Moreover, the definition has changed over the years and remains controversial. But what then is the actual phenomenon of El Nino? In attempting to address this question, we think that if El Nino is indeed a distinct causal macro-variable, then it should not be the result of an arbitrary definition, but instead be detectable in measurement data as a macro-scale variable. What does that mean in practice? We should be able to discover El Nino by aggregating micro-level measurement data without pre-defining what the relevant macro-features are. That is, if El Nino is not a mere definitional feature, but really exists as a climate macro-phenomenon, it should detectable by an unsupervised method. Additionally, what other climate macro-features does an unsupervised method detect? Causal Feature Learning (CFL) is a algorithm by Chalupka (1) that takes measures of two high-dimensional X and Y variables as input and -- rather than naively clustering each individually -- coarsens them into macro-variables with respect to their relation P(Y|X). We reapply this method to daily-averaged zonal wind maps (X) and surface air temperature maps (Y) over the equatorial Pacific from the NCEP-DOE AMIP-II Reanalysis dataset and show that they were able to automatically recover El Nino as a macro-level climate phenomenon, without informing the algorithm about past periods of El Nino. We additionally report on a new extensive robustness analysis we performed on these results, where we show significant improvements over standard clustering approaches, and second, we apply the same unsupervised method to the entire global NCEP Reanalysis II dataset and note the appearance of several interesting climatological
phenomena. We see these results as a proof of concept that it is possible to automatically construct features of scientific interest that used to be handcrafted by domain experts. We conclude that CFL provides a data-driven and theoretically-grounded approach to the construction of variables and the discovery of relations between them. The poster will primarily focus on the robustness analysis of the algorithm, showing improvements over traditional clustering methods such as k-means and robustness in changes to initial parameters.

Reference:

Power Analysis in a SMART Design: Sample Size Estimation for Determining the Best Dynamic Treatment Regime: Poster 10

Abstract: Sequential, multiple assignment, randomized trial (SMART) designs have become increasingly popular in the field of precision medicine by providing a means for comparing sequences of treatments tailored to the individual patient, i.e., dynamic treatment regime (DTR). The construction of evidence-based DTRs promises a replacement to adhoc one-size-fits-all decisions pervasive in patient care. However, there are substantial statistical challenges in sizing SMART designs due to the complex correlation structure between the DTRs embedded in the design. Since the primary goal of SMARTs is the construction of an optimal DTR, investigators are interested in sizing SMARTs based on the ability to screen out DTRs inferior to the optimal DTR by a given amount which cannot be done using existing methods. We fill this gap by developing a rigorous power analysis framework that leverages multiple comparisons with the best methodology. Our method employs Monte Carlo simulation in order to compute the minimum number of individuals to enroll in an arbitrary SMART. We evaluate our method through extensive simulation studies. We illustrate our method by retrospectively computing the power in the Extending Treatment Effectiveness of Naltrexone SMART study.
Bahamyirou, Asma

Author(s) Email: asma.bahamyirou@umontreal.ca

Institution: Université de Montréal

Poster Number: 13

Presenting Time: 6:00 P.M. - 7:30 P.M.

Doubly Robust Adaptive Lasso for Effect Modifier Discovery: Poster 13

Abstract: Effect modification occurs when the effect of the treatment on an outcome differs according to the level of a third variable (the effect modifier). Knowing which variables are effect modifiers is not a straight-forward task even for a subject matter expert. A natural way to assess effect modification in practice is by subgroup analysis, in which observations are stratified based on the covariates and stratum-specific estimates are calculated. One can also include the interaction terms between the treatment and the covariates in an outcome regression analysis. The latter, however, does not target a parameter of a marginal structural model (MSM) and also relies on a correctly specified outcome model. Our aim is to develop a data-adaptive method to automatically select effect modifying variables in a MSM with a single time point exposure. A two-stage procedure is thus proposed. First, we estimate two nuisance quantities (the conditional outcome expectation and propensity score) and plug these quantities into a doubly robust loss function based on the efficient influence curve (Rubin and van der Laan, 2006). Second, we use the adaptive LASSO (Zou, 2006) to select the effect modifiers and estimate marginal structural model coefficients. Simulations studies are performed in order to verify the performance (selection, estimation, and double robustness) of the proposed methods.
**Ben-Michael, Eli and Feller, Avi**

**Author(s) Email:** ebenmichael@berkeley.edu

**Institution:** University of California, Berkeley

**Poster Number:** 60

**Presenting Time:** 6:00 P.M. - 7:30 P.M.

---

**Matrix constraints for covariate balance: Poster 60**

**Abstract:** Achieving covariate balance is a critical step in the design of observational studies. As an alternative to inverse propensity score weighting, several methods instead propose to directly find weights that globally balance (functions of) the observed covariates, typically via a constrained optimization problem. Interestingly, recent work shows that such weights implicitly estimate a (possibly regularized) propensity score model. We generalize these methods to allow for matrix balance criteria that simultaneously constrain global and within-subgroup covariate imbalance. We show that the resulting optimization problem estimates a separate propensity score model within each subgroup with regularization penalties that globally regularize the separate models together (e.g. by enforcing group-structured sparsity or a low-rank condition on the propensity score parameters). This framework nests current approaches and allows for weights that directly achieve covariate balance in many new settings. First, this is a natural approach for estimating heterogeneous treatment effects in observational studies. More generally, this allows for a single set of weights in settings that have previously required several weights multiplied together, such as observational studies with missing outcomes that require both inverse propensity score and inverse probability of missingness weights. We evaluate the performance of these different estimators with simulation studies and on several canonical applications.
Efficient Estimation of the Average Treatment Effect on the Treated (ATT) with One-Sided Non-Compliance: Poster 11

Abstract: We consider an experimental setting that is plagued with one-sided non-compliance. This is common, for example, in experiments involving human subjects or treatments not accessible by subjects assigned to control. Here it is well-known that the local average treatment effect (LATE) estimand identifies the average treatment effect on the treated (ATT) under complete randomization. More recently, Frölich and Melly (2013) derived an alternative identification formula under conditional randomization (where the classic LATE estimand no longer equals the ATT), and proposed a nonparametric “plug-in” estimator. However, efficiency bounds were unknown under both complete and conditional randomization, as was whether the LATE, plug-in, or some other estimator attained such bounds. Our contributions are threefold: We derive the semiparametric efficiency bound for the ATT under complete and conditional randomization, propose and analyze a novel propensity score-weighted yet doubly robust LATE estimator that attains this bound, and characterize the corresponding efficiency gains relative to previous estimators. We illustrate our results and evaluate performance via simulations.

References:

Deep Learning, Auxiliary Data, and Randomization: Analyzing Experiments Run Within a Computerized Math Tutor: Poster 35

Abstract: Experimental effect estimates typically ignore data from the “remnant” of an experiment—potential or historical subjects who were not randomized, and hence differ statistically from experimental subjects. Traditional causal estimates also ignore high-dimensional or complex covariate data, due to sample size and other modeling limitations. These ignored data could still be useful, particularly if they can explain variation in the outcome. This poster illustrates a technique for incorporating both types of data into a design/randomization-based causal estimator, without sacrificing any of the benefits or guarantees of design based estimation. The procedure is as follows: first, fit a predictive model to data from the remnant, predicting outcomes as a function of covariates. Next, use the fitted model to generate predicted outcomes for the subjects who were randomized. Finally, estimate average treatment effects using prediction residuals in place of outcomes. Since the outcome predictions are based on a model fit to data from outside of the experiment, and pre-treatment covariates, they are invariant to treatment assignment. Therefore, the effect of treatment assignment on residuals is identical to its effect on outcomes. Any unbiased estimator that is linear in outcomes will remain unbiased when outcomes are residualized, and design-based inference procedures remain unchanged. If the predictions are any good, residualization removes ancillary variation in the outcome, and increases precision and power. We illustrate the method by estimating effects in 22 different experiments run within the ASSISTments computerized tutoring system, using a longitudinal deep learning model fit to historical student log-data. The procedure decreased standard errors in all but three of the experiments, in most cases by roughly 30% or more. The extent of the improvement was a function of the predictive accuracy of the deep learning model.
Is My Matched Dataset As-If Randomized, More, Or Less? Unifying the Design and Analysis of Observational Studies: Poster 9

Abstract: Matching is a useful preprocessing tool for mitigating the issue of covariate imbalance in observational studies. Matching involves finding a subset of units—a matched dataset—where units are plausibly “as-if randomized” to treatment or control. This “as-if randomized” assumption is typically supported by demonstrations of covariate balance between the treatment and control groups using rule-of-thumb diagnostics such as tables of covariate mean differences and Love plots. Then, the resulting matched dataset is analyzed as if it were from a randomized experiment. With modern computational tools, matching algorithms are able to produce datasets with strong levels of balance, such as many covariate mean differences being close to zero. Consequently, treatment effect estimators tend to be relatively unbiased. However, standard uncertainty intervals for these estimators do not account for the types of covariate balance that these matching algorithms produce. We find that, as a result, these intervals tend to be unnecessarily conservative. We develop an alternative approach that conditions on the types of covariate balance that matching algorithms instigate by design. Developing this approach involves three contributions. First, we formalize the “as-if randomized” assumption that is commonly alluded to in observational studies. Our formalization is a generalization of many assumptions employed in the literature, and it allows for any assignment mechanism of interest for a particular matched dataset, such as complete randomization, paired randomization, and randomization that is restricted to high levels of covariate balance. Second, we develop a randomization test for this generalized as-if randomized assumption that is more valid than the aforementioned rule-of-thumb diagnostics. This randomization test allows researchers to determine the assignment mechanism that their matched dataset best approximates, and thus this test encapsulates the design stage of an observational study. Third, we provide a treatment effect estimation strategy that uses the same assignment mechanism that was determined during the design stage. Thus, our approach unifies the design and analysis stages of observational studies, similar to how they are unified in the design and analysis of randomized experiments. We find—both theoretically and through simulation—that our approach results in more precise causal inferences than current
approaches by taking advantage of assignment mechanisms that account for the types of covariate balance generated by state-of-the-art matching algorithms.
Evaluating the Effects of Attitudes on Health-Seeking Behavior among a Network of People Who Inject Drugs: Poster 28

Abstract: This study employed causal inference methods in the presence of dissemination (or interference) in an observational study to assess attitudes toward HIV/AIDS risk among people who inject drugs (PWIDs) and their effect on health-seeking behaviors. We analyzed data from the Social Factors and HIV Risk Study (SFHR), a sociometric network study conducted between 1991 and 1993 in Bushwick, Brooklyn, New York that investigated how HIV/AIDS infection spread among PWIDs through shared sexual and injection risk behaviors. We evaluated the effects of PWIDs’ locus of control (internal vs. external) and blame (self vs. others) attitudes separately on both their own health-seeking behavior and that of other members in their subnetworks. Subnetworks were defined to include members that were closely related via risk behavior and had sparser connections with individuals outside of the subnetwork. For the health-seeking behavior outcomes, we evaluated receipt of study-based HIV testing result and attending a medical visit within the past year. First, we applied a modularity-based community detection algorithm to determine subnetworks within the full SFHR network. We then employed a network-based causal inference methodology for clustered observational data. PWIDs who believe uncontrollable factors determine whether or not one gets HIV/AIDS (i.e. with external locus) were 16% less likely to receive their SFHR HIV test result when they were in a 50% coverage (i.e., 50% in subnetwork reporting internal locus) subnetwork (95% confidence interval (CI): -0.27, -0.06). When the coverage of internal locus decreased from 70% to 50%, the likelihood of receiving HIV test result decreased 3% for those with external locus (95% CI: -0.05, -0.01). Furthermore, when the coverage of self-blame was decreased from 99% to 50%, the likelihood of having a recent doctor visit increased 27% for those who blame others (95% CI: 0.07, 0.47). The results obtained from this study provide evidence of indirect effects of attitudes among PWID networks. These results will inform the development of more effective network-based interventions that consider attitudes to prevent HIV/AIDS and improve care among PWIDs.
Is the Ignorability Assumption of Propensity Score Analysis being Met in Pharmaceutical Observational Studies? : Poster 27

Abstract: Objective: To test whether the ignorability from confounding assumption of propensity score analysis holds in the often studied question of the effectiveness of Direct Acting Oral Coagulants (DOACs) versus Warfarin.

Method: A retrospective observational study was performed on Cigna customers who used Eliquis (n=1,236), Xarelto (n=1,237) or Warfarin (1,043) between 1/1/2016 and 8/31/2016. Customers were followed for one year after their first script between 1/1/16 and 8/31/16 and any instances of adverse bleeding or stroke were recorded. Prior to implementing propensity score analysis, confounders were chosen using a Gradient Boosted Random Forest. Over 1,000 potential confounders were tested for predictability of treatment assignment between Eliquis, Xarelto and Warfarin. After calculating propensity scores, the different therapies were evaluated as to safety (fewer adverse bleeding events) and efficacy (fewer strokes). Despite its lack of inclusion in any published clinical trial or observational study comparing Warfarin to Eliquis or Xarelto, the most important predictor for treatment assignment was prescriber specialty. Even though not controlled for in the DOAC experimental literature, prescriber specialty has been shown to be a predictor of which drug is prescribed in several clinical research papers. In our analysis, patients who were prescribed Warfarin or a DOAC from a cardiologist were significantly less likely to have a stroke than patients prescribed Warfarin or a DOAC from an internal medicine specialist or a family practitioner showing the variable’s importance in treatment assignment and outcome.

Results: Including this variable in the propensity score led to similar efficacy between Eliquis, Warfarin and Xarelto in reducing strokes. Using a Cox Proportional Hazards Model led to a hazard ratio for Eliquis of 0.970 (CI=0.721-1.305 p-value 0.8411) and 0.803 for Xarelto (CI=0.580-1.111 p-value 0.1847). The same Cox Proportional Hazards Model led to a significant reduction in adverse bleeding events for Eliquis (Hazard Ratio=0.786 CI=0.620-0.996 p-value=0.0463) and a statistically insignificant difference for Xarelto (Hazard Ratio=0.848 CI=0.675-1.065 p-value=0.1566). Overall, the model had a
misclassification rate of 51.82%, a substantial lift over treatment assignment predicted at random (64.82%).

**Conclusions:** Propensity Score Analysis is dependent on the assumption that all confounders have been properly identified. The lack of inclusion of prescriber specialty in observational studies leads one to question their accuracy regarding DOACs and Warfarin. We further suggest that identifying confounders in an observational study by replicating variables found in a clinical trial may not be enough and that prediction of treatment assignment by machine learning may lead to more accurate findings.

**Limitations:** The causal mechanism of prescriber specialty is not completely understood. Are clinical outcomes different for various prescriber specialties because of pre-treatment qualities of the patient or because of differences in provider care? Is prescriber specialty a cause on its own or is it serving as a proxy for latent variables? I will continue to investigate these questions and hope that other researchers will examine similar questions with regard to atrial fibrillation and other conditions in order to assess the accuracy of propensity score analysis in pharmaceutical settings.
Application of Causal Bayesian Networks to Environmental Data: Poster 31

Abstract: In agriculture, we are often interested in studying how crop yield responds to changes in soil nutrients measured during a growing season with experimental data. For this purpose, it is necessary to identify the confounders in the soil nutrients-yield relationship because crop yield is the result of combined effects of soil nutrients, environment, management and weather. In addition, there are other potential problems with the data such as measurement variability due to different sampling times or sampling resolution. In this paper, we applied multiple constraint-based Causal Bayesian Networks Learning Algorithms to Climate Research Farm data from five farms spanning the U.S. corn belt to infer the structure of the relationships between the variables of interest and identify potential confounders. Moreover, we compared the inferred networks with domain knowledge in order to interpret results. We discovered several scenarios for which data suggested that yield responded to soil nutrients, as well as several scenarios where it did not. We identified potential confounders, which could be useful for correctly modeling the soil nutrients-yield relationship.Lastly, we also identified several aspects of the current Causal Bayesian Networks Learning Algorithms that can be improved.
Chen, Ling-Wan

**Author(s) Email**: lic76@pitt.edu

**Institution**: Department of Statistics, University of Pittsburgh

**Poster Number**: 41

**Presenting Time**: 6:00 P.M. - 7:30 P.M.

**Quantile Association Model for Bivariate Survival Data: Poster 41**

**Abstract**: The association analyses are useful, since understanding how two events are related helps scientists to develop strategies to prevent or promote the occurrence of an event, when they observe an associated event. For bivariate survival data, several global association measures were proposed, such as Kendall’s tau and the correlation between two quantitative variates. However, they cannot capture the dynamic local association pattern over time. Various local association measures have been proposed via a frailty (copula) framework because the copula model allows time-dependent association between two failure times. Furthermore, assessing the potential risk factors in the association analysis is of scientific importance, where the conditional association is adjusted for the confounder effects and the potential predictors. Many studies have been proposed to allow the covariate effects on the marginal distributions only. In practice, risk factors may affect the local association directly, in addition to their effects on the marginal distributions. To handle this challenge, we adopt a novel quantile-specific association measure, which is independent of the marginal distributions, and establish a quantile association model to allow covariate effects on this quantile-based association measure between two failure times. We develop an estimating equation for the quantile association coefficients via the relationship between this quantile-based measure and the copula. The asymptotic properties for the resulting estimators are studied. Furthermore, the covariance estimation is often challenging in the analyses of quantile regression and quantile association due to non-smooth objective functions. To address this issue, we extend the idea of the induced smoothing technique to our quantile association analysis. Simulations were conducted to show that our estimator is unbiased and the covariance estimation procedure is robust. We illustrate our proposed model through an analysis of an age-related macular degeneration data.
Central limit theorems via Stein's method for randomized experiments under interference: Poster 52

Abstract: Controlling for interference through design and analysis can consume both engineering resources and statistical power, so it is of interest to understand the extent to which estimators and confidence intervals constructed under the SUTVA assumption are still valid in the presence of interference. Toward this end, Sävje et al. (2017) provide laws of large numbers for standard estimators of the average treatment effect under a limited form of interference characterized by an interference dependence graph. In this paper, we link that view of interference to the dependency graph version of Stein's method. We prove a central limit theorem for a variant of the difference-in-means estimator if the o(n) restriction on average dependency degree in Sävje et al. (2017) is replaced by an o(n^{1/4}) constraint on the maximal dependency degree. We then provide a central limit theorem that can handle interference that exists between all pairs of units, provided the interference is approximately local. The asymptotic variance admits a decomposition into two terms: (a) the variance that is expected under no-interference and (b) the additional variance contributed by interference. The results arise as an application of two flavors of Stein's method: the dependency graph approach and the generalized perturbative approach.
Time-varying survivor average causal effects with semi-competing risks: Poster 16

Abstract: In semi-competing risks problems, non-terminal time-to-event outcomes such as time to hospital readmission are subject to truncation by death. Such settings are often evaluated with parameters from illness-death models, but evaluating causal treatment effects with hazard models is problematic due to the evolution of incompatible risk sets over time. As an alternative, the survivor average causal effect (SACE) is a principal stratum causal effect of a treatment on the non-terminal event among units that would survive regardless of the assigned treatment. Traditional SACE formulations specify a single time point past which an individual is deemed to have always survived. In contexts with a high and persistent hazard of death -- as in late-stage cancer or other serious illness -- it may be impractical to identify only one time point at which to estimate the causal contrast. We propose a new causal estimand, the time-varying SACE (TV-SACE), for non-terminal events in the semi-competing risks setting. For non-terminal events representing a permanent state change, we propose a second causal estimand, the restricted mean survival average causal effect (RM-SACE). We adopt a Bayesian estimation procedure that is anchored to parameterization of illness-death models for both treatment arms but maintains causal interpretability. We further outline a frailty specification that can accommodate within-person correlation between non-terminal and terminal event times.
Extrapolation parameterizations for assessing sensitivity to unmeasured confounding: Poster 46

Abstract: A fundamental challenge in observational causal inference is that assumptions about unconfoundedness are not testable from data. As such, it is important to assess the sensitivity of all causal estimates to plausible unmeasured confounding. The core idea in sensitivity analysis is that a fixed observed data distribution is compatible with many (potentially contradictory) true causal effects, depending on the value of some unidentifiable sensitivity parameter. We argue that, as a general principle, the implementation of sensitivity analysis should reflect this division: specifically, the observed data distribution implied by the working model in a sensitivity analysis should not depend on the sensitivity parameter. Many commonly used sensitivity analysis approaches violate this invariance principle. This principle is particularly important in the context of Bayesian or other model-based causal inference. In that case, the model specification may constrain the sensitivity parameter even though the observed data provide no information about the parameter. We propose a Bayesian method for assessing sensitivity that has the desired invariance. Our approach is based on a factorization for missing data densities first attributed to John Tukey and more recently termed “the extrapolation factorization.” In this factorization, we explicitly separate likelihood factors of the fully identified observed data distribution from likelihood factors of the fully unidentified treatment assignment mechanism. We demonstrate the utility of this approach by estimating the effects of diuretics on blood pressure using the NHANES data. We show that our approach can be used in conjunction with nonparametric methods for response surface modeling (e.g. Bayesian Adaptive Regression Trees) and flexible models of the potential outcome distributions (e.g. Dirichlet Process Mixtures). Finally, we provide a procedure for interpreting and calibrating the sensitivity parameters.
Fisher, Aaron

**Author(s) Email:**  aafisher@hsph.harvard.edu

**Institution:**  Harvard T.H. Chan School of Public Health

**Poster Number:**  5

**Presenting Time:**  6:00 P.M. - 7:30 P.M.

---

**Model Class Reliance: Variable Importance when all Models are Wrong, but *Many* are Useful: Poster 5**

**Abstract:**  Variable importance (VI) tools are typically used to examine the inner workings of prediction models. However, currently available VI measures tend to not be comparable across model types, can obscure implicit assumptions about the data generating distribution, or can give seemingly incoherent results when multiple prediction models fit the data well. In this paper, we propose a framework of VI measures for describing how much any model class (e.g. all linear models of dimension p), any model-fitting algorithm (e.g. Ridge regression with fixed regularization parameter), or any individual prediction model (e.g. a single linear model with fixed coefficient vector), relies on covariate(s) of interest. The building block of our approach, Model Reliance (MR), compares a prediction model's expected loss with that model's expected loss on a pair of observations in which the value of the covariate of interest has been switched. Expanding on MR, we propose Model Class Reliance (MCR) as the upper and lower bounds on the degree to which any well-performing prediction model within a class may rely on a variable of interest, or set of variables of interest. Thus, MCR describes reliance on a variable while accounting for the fact that many prediction models, possibly of different parametric forms, may fit the data well. We give probabilistic bounds for MR and MCR, leveraging existing results for U-statistics. These bounds can be generalized to create finite-sample confidence regions for the best-performing models from any class, which can then be used in sensitivity analyses. We also illustrate connections between MR and conditional causal effects. Finally, we apply MR & MCR in a public dataset of Broward County criminal records to study the reliance of recidivism prediction models on sex and race.
Nonparametric identification and robust estimation of indirect causal effects in the presence of exposure-outcome confounding: Poster 14

Abstract: Current mediation methods to identify and estimate indirect and direct causal effects require stringent no unmeasured confounding assumptions for the exposure-mediator, mediator-outcome, and exposure-outcome relationships. These assumptions are unlikely to hold in observational studies where mediation methods are often applied. The goal of this paper is to develop methodology to identify and estimate a novel causal effect, the population intervention indirect effect, in the presence of exposure-outcome unmeasured confounding. The population intervention indirect effect (PIIE) is the indirect component of the population intervention effect, which may be of primary interest in observational settings as argued by Hubbard and Van der Laan (2008). Interestingly, the conditions to empirically identify the PIIE are a generalization of Judea Pearl’s front-door criterion, leading to robustness of exposure-outcome unmeasured confounding. In contrast with the well-known natural indirect effect which fails to be nonparametrically identified in the presence of unmeasured confounding of the exposure-outcome relation, the PIIE remains empirically identified whether or not exposure-outcome unmeasured confounding exists making it an invaluable tool in the current causal mediation analysis literature. For estimation of the PIIE, we provide parametric and semiparametric estimators, including a doubly robust semiparametric locally efficient estimator, that perform very well in simulation studies. Finally, these methods are used to measure the effectiveness of monetary saving recommendations among women enrolled in a maternal health program in Tanzania.
Selecting a generalizable sample in field trials with stratification using balanced sampling: A simulation study: Poster 62

Abstract: Recently, causal estimates from multi-site field trials have come under criticism for their limited external validity due to non-random site selection (Bell & Stuart, 2016). Stratification using balanced sampling (SUBS; Tipton, 2013) has been proposed as an approach to improve the generalizability of results from such studies when random sampling is not feasible. SUBS is a structured and data-driven method of selecting a sample representative of a target population in terms of covariates that may explain variation in treatment effects. The population is first divided into strata using cluster-analysis methods. Within strata, units are ranked from most to least “typical,” with higher ranked units prioritized for recruitment. When implemented well, the result is a self-weighting sample that is representative of the target population, making generalizations from the sample to population straightforward and comparatively efficient. In practice, implementing SUBS requires several analytic decisions. Originally, k-means clustering was recommended to generate strata using a general similarity measure to calculate distance and comparing the proportion of between sums of squares to total sums of squares to determine the number of strata. However, as of yet there has been no methodological work evaluating this approach to implementing SUBS, nor comparing SUBS to alternative recruitment methods. Furthermore, in practice SUBS may place unrealistic expectations on recruiters. For instance, in recruiting for an educational field trial, school districts serve as gatekeepers and require dense and unstandardized research proposal submissions. If approved, researchers may then recruit schools, whose participation is not guaranteed by districts. In this context, SUBS can result in recruiters spending time on small districts with schools that are high priority but have low likelihood of participation. The goal of this study is to assess the effectiveness of SUBS for selecting a sample that is representative of a target population. We generate a population frame using real data and design a Monte Carlo simulation process to emulate realistic conditions. We generate samples using SUBS, simple random sampling and convenience sampling. The operating characteristics of these sampling methods are compared using a generalizability index (Tipton, 2014) as well as across covariate standardized mean differences between the sample and the target population.
Recruitment statistics such as proportion of rejections are also compared. The response generation model is manipulated to test the sampling methods under different overall participation rates. The performance of SUBS with an omitted variable is also examined. Preliminary results indicate that SUBS performs well in selecting a generalizable sample, but that recruitment costs are quite high in terms of number of schools and districts that need to be contacted.

**Reference:**

Abstract: In many policy evaluations, propensity score matching is used to construct a matched control group that appears similar to the treatment group to minimize nonrandom selection bias. When subjects enroll in the treatment group on a rolling basis, several complications are introduced. These complications stem from the fact that, while each member of the treatment group is enrolled on a particular date, no analogous date exists for members of the control group. Thus, defining a single baseline period of matching covariates for the control group is more difficult. When treatment eligibility and enrollment is preceded by an acute event, such as a stroke or other type of hospitalization, the importance of properly aligning control subjects is compounded. We discuss several strategies to handle these complications, including a novel approach to optimize matching while disallowing a unique potential control to be matched to more than one treatment subject. This approach is available as the R package groupmatch, which is compatible with other popular matching packages such as optmatch and MatchIt.
A Moderated Nonlinear Factor Analysis Solution to the Problem of Differential Measurement Error in Propensity Score Analysis: Poster 18

Abstract: Adjusting for confounding has been the primary focus of much causal inference of the last several decades. Relatively little attention has been paid to the problem of measurement error in the confounders in the context of propensity score analysis, despite evidence that the presence of measurement error can invalidate inferences by biasing causal effect estimates. One solution is to compute estimates of the latent confounding variable from its observed indicators and use those estimates in place of the true variable to estimate causal effects. Raykov (2012) and Jakubowski (2015) independently proposed such a method. Nguyen et al. (forthcoming) advanced this method by including the treatment variable as an indicator of the latent variable and modeling the covariances between the observed confounders and the latent confounder, which they call the “fully inclusive” method. The question of how these methods function in the face of differential measurement error has been unaddressed, though much research in psychometrics has focused on how to address this problem in the context in scale scoring. I propose the application of a generalization of factor analysis called moderated nonlinear factor analysis (MNLFA) to model the effects of differential measurement error and generate factor scores for use in propensity score weighting to estimate the causal effect of a binary treatment. The fully inclusive method without MNLFA is robust to the differential measurement studied here, but incorporating MNLFA into the scoring algorithm improves precision when differential measurement is present and has no detrimental effects when differential measurement is absent. Though propensity score weighting yields biased effect estimates when using the MNLFA fully inclusive method due to poor balancing performance of the weights, requiring exact balance using entropy balancing (Hainmueller, 2012) eliminated these biases. I recommend using the MNLFA fully inclusive method in the face of uncertainty about differential measurement because the estimated factor scores provide the most information about the true latent confounder and suffer no losses when the applied model is larger than the data-generating model.
Hasegawa, Raiden

Author(s) Email: raiden@wharton.upenn.edu

Institution: The Wharton School, University of Pennsylvania

Poster Number: 34

Presenting Time: 6:00 P.M. - 7:30 P.M.

---

Extended sensitivity analysis for heterogeneous unmeasured confounding with an application to sibling studies of returns to education: Poster 34

Abstract: The conventional model for assessing insensitivity to hidden bias in paired observational studies constructs a worst-case distribution for treatment assignments subject to bounds on the maximal bias to which any given pair is subjected. In studies where rare cases of extreme hidden bias are suspected, the maximal bias may be substantially larger than the typical bias across pairs, such that a correctly specified bound on the maximal bias would yield an unduly pessimistic perception of the study’s robustness to hidden bias. We present an extended sensitivity analysis which allows researchers to simultaneously bound the maximal and typical bias perturbing the pairs under investigation while maintaining the desired Type I error rate. We motivate and illustrate our method with two sibling studies on the impact of schooling on earnings, one containing information of cognitive ability of siblings and the other not. Cognitive ability, clearly influential of both earnings and degree of schooling, is likely similar between members of most sibling pairs yet could, conceivably, vary drastically for some siblings. The method is straightforward to implement, simply requiring the solution to a quadratic program.
Robust Nonparametric Inference for Stochastic Interventions Under Multi-Stage Sampling: Poster 49

Abstract: Perhaps too often, work in statistical causal inference focuses on the effect of deterministic interventions, under which, for each unit, the magnitude of the treatment is set to a fixed value. Under violations of the assumption of positivity, the evaluation of such interventions faces a host of problems, among them non-identification and inefficiency. Prior work has proposed a flexible solution: stochastic shift interventions, under which, in the simplest case, for each unit, the treatment is set to be an additive shift of the observed value of the treatment. What’s more, in real-life applications, data analyses are often further complicated by pragmatic sub-sampling schemes, the effects of which cannot safely be ignored when drawing statistical inferences. Building on much previous work, we present a novel approach for such settings --- an augmented targeted maximum likelihood estimator for interventions that shift observed values of the treatment, with consistency and efficiency guarantees even in the presence of multi-stage sampling, and we show that this estimator enjoys these essential theoretical properties by way of a form of multiple robustness inherited from its constituent parts. After providing a general characterization of shift interventions, we illustrate the utility of employing our proposed nonparametric estimator via simulation studies, showing that it attains fast convergence rates even when incorporating machine learning estimators; moreover, we introduce a recent software implementation (the "txshift" R package) and apply this methodology in an investigation of the effects of immune response biomarkers on HIV vaccine efficacy, contrasting our proposed approach with several classical techniques. Specifically, we show that our proposed method obtains efficient inference on a parameter defined as the overall risk of HIV infection in the vaccine arm of an efficacy trial, under various posited shifts of the distribution of an immune response biomarker away from its observed distribution in the efficacy trial. Our proposed technique provides a highly interpretable variable importance measure -- defined through the formalism of statistical causal inference -- for ranking multiple immune responses based on their utility as immunogenicity study endpoints in future HIV-1 vaccine trials that evaluate putatively improved versions of the vaccine. Time permitting, we discuss extensions of this approach that consider recent and novel ideas.
in stochastic interventions, such as the induction of shifts in the treatment in terms of the propensity score rather than on the observed treatment scale.
Finding the Strength in a Weak Instrument in a Study of Cognitive Outcomes Produced by Catholic High Schools: Poster 33

**Abstract:** The strength of an instrument is incompletely characterized by the proportion of compliers. For a fixed small proportion of compliers, the presence of an equal number of always-takers and never-takers weakens an instrument, whereas the absence of always-takers or, equivalently, the absence of never-takers, strengthens an instrument. Here, the strength of an instrument refers to its ability to recognize and reject a false hypothesis about a structural parameter. This ability is measured by the Bahadur efficiency of a test that assumes the instrument is flawless, or the Bahadur efficiency of a sensitivity analysis that assumes the instrument may be somewhat biased. Studies of the effects of Catholic high schools on academic test performance have used being Catholic as an instrument for attending a Catholic high school, and the application concerns such a comparison using the US National Educational Longitudinal Study. Most Catholics do not attend Catholic school, so there are few compliers, but it is quite rare for non-Catholics to attend Catholic school, so there are very few always-takers.
Augmented Minimax Linear Estimation: Generalizing Double Robust Balancing Estimators: Poster 8

Abstract: For estimation of average treatment effects, balancing weights are an increasingly popular alternative to inverse probability weights. Their use has strong empirical and theoretical justification. In particular, several popular balancing estimators fall into the class of minimax linear estimators, which have been shown to be nearly optimal in fixed design among all estimators. This property holds for minimax linear estimators for a large class of estimands: the class of linear functionals of regression functions. Balancing weights for binary treatments are chosen to ensure the weighted average of an unknown regression function over one (e.g. control) subsample approximates the expectation of that function on another (e.g. treatment) subsample. And for any linear functional, the minimax-optimal weights ensure a generalization of this balance property: the weighted average of an unknown regression function approximates the linear functional evaluated at that function. The existence of such weights follows from the Reisz representation theorem under a generalization of the familiar ‘overlap’ condition. Like balancing weights for more familiar problems, they can be estimated by taking this approximate equality, for a large set of possible regression functions, as estimating equations for the weights. I will describe a general approach to estimating linear functionals satisfying this generalized overlap condition. The proposed estimators incorporate the minimax-optimal weights into a double robust augmented weighting estimator. I will give general conditions under which these estimators are semi-parametrically efficient and demonstrate via simulation the performance of these estimators. As examples, I will discuss estimation of several types of average treatment effect for continuous-valued treatments.
Uniform non-asymptotic confidence sequences, with applications to sequential testing and estimation: Poster 50

**Abstract:** We develop non-asymptotic, nonparametric confidence sequences holding over unbounded time horizon and achieving arbitrary precision. Our technique draws a connection between the Cramèr-Chernoff method, the law of the iterated logarithm (LIL), and the sequential probability ratio test, extending the first to produce time-uniform concentration bounds, characterizing the second in a non-asymptotic fashion, and generalizing the third to nonparametric settings, including sub-Gaussian and Bernstein conditions, self-normalized processes, and matrix martingales. We tighten and substantially generalize existing constructions of non-asymptotic iterated logarithm bounds, illustrating the generality of our proof techniques by deriving a novel upper LIL bound for the maximum eigenvalue of a sum of random matrices. Finally, we demonstrate the utility of our approach with applications to covariance matrix estimation and to estimation of sample average treatment effect under the Neyman-Rubin potential outcomes model. In the latter application, we give a sequence of confidence intervals updating after each subject is randomized and observed. Each interval estimates the average treatment effect for the finite population observed up to that point. The coverage guarantees hold non-asymptotically and uniformly over unbounded time under a randomization inference model in which potential outcomes are considered fixed and known to be uniformly bounded.
Assessing Therapeutic Equivalence of Brand and Generic Drugs Using Observational Data: Poster 19

Abstract: Although generic drugs are required to be bio equivalent to brand, generic producers are not required to establish therapeutic equivalence through clinical trials. We describe a method to assess therapeutic equivalence of brand and generic drugs using insurance claims data with the anti-depressant venlafaxine. We use time to drug failure as an outcome. Generic market entry is typically followed by a large shift among new users towards initiation on generic, with little overlap in initiation times of brand and generic users. This creates temporal confounding when observed survival times are affected by changes over time in unmeasured variables. There is also time varying confounding. Our method addresses both of these concerns by applying Regression Discontinuity to counterfactual survival curves, with a discontinuity in the probability of initiation to generic at the date when generic becomes available. The survival curves themselves are estimated using G-Computation to account for the time-varying confounding.
Ju, Cheng

Author(s) Email: jucheng1992@gmail.com

Institution: University of California, Berkeley

Poster Number: 15

Presenting Time: 6:00 P.M. - 7:30 P.M.

On Adaptive Propensity Score Truncation in Causal Inference: Poster 15

Abstract: The positivity assumption, or the experimental treatment assignment (ETA) assumption, is important for identifiability in causal inference. Even if the positivity assumption holds, practical violations of this assumption may jeopardize the finite sample performance of the causal estimator. One of the consequences of practical violations of the positivity assumption is extreme values in the estimated propensity score (PS). A common practice to address this issue is truncating the PS estimate when constructing PS-based estimators. In this study, we propose a novel adaptive truncation method, Positivity-C-TMLE, based on the collaborative targeted maximum likelihood estimation (C-TMLE) methodology. We demonstrate the outstanding performance of our novel approach in a variety of simulations by comparing it with other commonly studied estimators. Results show that by adaptively truncating the estimated PS with a more targeted objective function, the Positivity-C-TMLE estimator achieves the best performance for both point estimation and confidence interval coverage among all estimators considered.
Permutation-Based Variable Selection with Random Forests: Poster 65

Abstract: To identify an average causal effect with a conditioning strategy it is necessary to assume ignorability (Rosenbaum & Rubin, 1983), which is more likely to be tenable if the set of conditioning variables, X, is large. Although indiscriminate inclusion of conditioning variables will not make an otherwise consistent estimator inconsistent, it will decrease efficiency. We define variable selection as the process by which $X_s \subset X$ is identified with the aim of satisfying ignorability, such that $\{Y_0,Y_1\} \perp T|X \Rightarrow \{Y_0,Y_1\} \perp T|X_s$. de Luna, Waernbaum, & Richardson (2011) and VanderWeele & Shpitser (2011) proposed algorithms for variable selection based on non-parametric statements of conditional independence. To maintain freedom from assumptions, a method for conditional independence testing that relies on minimal assumptions is also desired. Random forests (RFs) are useful because they are based on regression trees (Breiman, Friedman, Olshen, & Stone, 1984), which nonparametrically handle interactions and non-linear relationships. There is no consensus on how to best use RF variable importance to implement variable selection. The use of a global cutoff is problematic because (a) variables on different scales may produce importance values that are not directly comparable and (b) RF variable importance favors correlated predictors (Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008). Furthermore, the accuracy of such methods depends entirely on the arbitrary choice for the cutoff value. In this poster we will describe a new method that avoids the deficits associated with using an arbitrary cutoff. By generating a null distribution for each predictor via permutation, the decision to retain or discard each predictor is made as follows:

Step 1. Run RFs and estimate the initial importance for all predictors.

Step 2. Create R permuted copies of the predictor matrix, $X_1, X_2, ..., X_R$, by randomly shuffling the rows R times.

Step 3. Run RFs on the R replicates to generate R importance estimates for each variable.
Step 4. Compare each variable’s initial importance to the permutation distribution.

For a level \( \alpha \) test, a variable that exceeds the \((1 - \alpha)\) percentile is retained.

We illustrate the RF permutation algorithm with an ECLS-K data set that has been used to examine the causal effect of exposure to special education on later math achievement in school age children. We then present results from simulation studies that demonstrate the performance of the RF permutation algorithm for linear and non-linear data generation processes. We revisit the ECLS-K example and dig more deeply into the selection of a variable that has no marginal linear association with the outcome but, nevertheless, seems important; the variable is the reading IRT pretest score. We find that the reading pretest is involved in a crossing-type interaction with a math pretest variable such that the main effect of reading pretest is zeroed out. The permutation RF method correctly identifies the reading pretest as an important predictor, while algorithms based on assumptions of linear relationships fail to retain the reading variable. We conclude with some discussion and a link to an R package that implements the algorithm.
An Example where the Fundamental Problem of Causal Inference Does Not Occur: The Causal Effect of Answer Changing: Poster 2

Abstract: The fundamental problem of causal inference, as put forward by Holland (1986), has been presumed to exist in all causal inference studies. We present an intriguing real example where the fundamental problem does not occur and all individuals’ counterfactuals are known to researchers. Educational measurement researchers have long debated whether changing initial answers is beneficial or harmful on final scores when taking multiple-choice tests. That is, after an examinee chose an initial response, he or she can self-select to change (i.e., treatment) or retain (i.e., control) it. Depending on the choice, the examinee has a different score (i.e., correct or incorrect) on the item. It turns out that this long-standing puzzle can be easily resolved by applying the potential outcomes framework. When taking true-false tests, researchers observe the potential treatment outcomes of those who change their answers (i.e., treatment group) but also can infer their potential control outcomes because the outcomes must be identical to their initial answers. In the same way, researchers observe the potential control outcomes of those who retain their initial answers (i.e., control group) but also can infer their potential treatment outcomes because the outcomes must be opposite to their initial answers (e.g., if an initial answer was correct, the potential treatment outcome must be incorrect). Since both potential outcomes are known to researchers, the fundamental problem does not occur and researchers can directly compute the individual causal effects without any identification assumptions (e.g., strong ignorability). We extend the approach to general multiple-choice items in which more than two alternatives are present and show that, in this case, the average treatment effect on the treated (ATT) can be exactly computed while the average treatment effect (ATE) and the average treatment effect on the untreated (ATU) can be bounded. Analyzing a real-world student-response dataset, we show that the ATT is positive while the ATE and ATU are negative. In addition, using the same dataset, we demonstrate how matching and difference-in-differences (DiD) methods succeed to identify causal effects. In the context of answer changing, the potential control outcomes must be identical to the initial answers because by definition, the potential control outcomes are the counterfactual final answer correctness if one would have retained the initial response.
Viewing the initial answer variable as a covariate, the potential control outcome is independent of the treatment conditional on the initial response, therefore, the ATT can be identified by matching methods (i.e., matching on the initial response, computing the stratum-specific effects and weights, and computing the weighted average across strata). The equivalence between the potential control outcome and the initial response also establishes the common trend or the additive equi-confounding assumption, required for DiD methods (i.e., the initial and final answers are viewed as repeated measures). We show that both matching and DiD estimates are numerically identical to the ATT estimate we obtained by directly inferring the potential outcomes (without relying on matching and DiD methods). We discuss theoretical and practical implications of our findings.
Kim, Kwangho

Author(s) Email: kwhanghk@stat.cmu.edu

Institution: Carnegie Mellon University

Poster Number: 23

Presenting Time: 6:00 P.M. - 7:30 P.M.

Estimating causal effect as difference in counterfactual outcome distributions:

Poster 23

Abstract: We develop a novel framework for estimating causal effect based on a degree of distributional discrepancy between unobserved counterfactual distributions. In our setting, a causal effect is defined as a mean distribution distance between different counterfactual outcome distributions, not a mean difference in their values. To directly compare counterfactual outcome distributions can provide more valuable information about causality than simply to compare their mean values. We propose estimators to effectively measure the degree of discrepancy between two counterfactual outcome distributions under single- and multi-source randomized studies, and observational study respectively. We analyze error bound and asymptotic properties of the proposed estimators. Also, we propose methods to construct confidence intervals for the unknown mean distribution distance. Finally, we verify their effectiveness by empirical studies.
Collective problem solving, causal inference, and chain graph: Poster 57

Abstract: In social networks, treatments may spill over from the treated individual to his or her social contacts and outcomes may be contagious. Researchers interested in causal inference have developed methods for interference -- when one individual's treatment or exposure affects not only his/her own outcome but also the outcomes of his/her contacts -- and researchers interested in social networks have attempted to model the spread of contagious outcomes across network ties. In both of these settings, causal inference using observational data from a single social network requires observing longitudinal data on treatments and outcomes as they evolve in real-time, so that each spillover or contagious event appears in the data. This results in two roadblocks for researchers. First, in most settings it is impossible to collect the kind of real-time data required. Second, even if the full longitudinal data are available, the resulting model will generally be high-dimensional and often too big to fit to the available data. As a practical matter, most researchers deal with reduced data, comprised of observations collected at one or a small number of time points. But the omission of some time points not only renders some causal effects unidentified, but also renders all outcomes at any observed time marginally and conditionally dependent, resulting in a saturated likelihood with many parameters for each observed time point even before treatment and covariates are included. In our research, we propose and justify a parsimonious parameterization for social network data with interference and contagion. Our parameterization corresponds to a particular family of graphical models known as chain graphs. We demonstrate that chain graph models approximate the projection of the full longitudinal data onto the observed data, which is missing most of the time points from the full data. We illustrate the use of chain graphs for approximate causal inference about contagion, interference, and collective decision making in social networks when the longitudinal evolution of treatments/outcomes is not fully observed. Finally, we apply the chain graph model to data on the collective decisions made by nine supreme court justices.
Statistical assessment of strict population overlap in observational studies: Poster 53

**Abstract:** In causal inference, strict population overlap, also known as positivity or common support, is a central assumption for identification and efficient estimation of causal effects. This assumption asserts that the propensity score is bounded away from 0 and 1. Strict overlap is often treated as axiomatic; however, in this paper, we show that the strict overlap assumption has empirical implications that can be observed in finite samples. In particular, we specify several estimators for the overlap slack, or the constant that bounds the propensity score away from 0 and 1. Our procedures provide upper confidence bounds, which can be interpreted as optimistic estimates of the overlap slack in the study population. Assuming i.i.d. observations, we specify an exact, finite-sample confidence bound, based on the Dvoretzky-Kiefer-Wolfowitz-Massart inequality, and an empirical Bennett inequality. Surprisingly, the confidence bound require no propensity score model and no conditions on the covariates, and can be made more powerful by incorporating modern machine learning or ensemble classification methods. Our estimators can be used as a diagnostic in observational studies, where a small estimated overlap slack would suggest re-defining the causal estimand or carefully choosing the causal estimator. Our result can be applied to any study that requires overlap, particularly useful in studies with high-dimensional covariates, where poor overlap can be difficult to discern by the conventional practice of examining fitted values from a propensity score model, which is likely to be mis-specified. We demonstrate our methodology on simulated and real data.
Li, Fan (Frank)

Author(s) Email: frank.li@duke.edu

Institution: Duke University, Department of Biostatistics and Bioinformatics

Poster Number: 30

Presenting Time: 6:00 P.M. - 7:30 P.M.


Abstract: A central goal in traffic safety research is to evaluate the effectiveness of safety countermeasures. Observational before-after design is common in these studies, where safety outcomes are recorded for each roadway segment both before and after the countermeasures are implemented. Despite its causal nature from a statistical perspective, such an evaluation has rarely utilized causal inference methods. In this paper, motivated from a real application of evaluating the effects of rumble strips on reducing vehicle crashes, we consider a difference-in-differences (DID) framework for causal inference in traffic safety before-after studies. Within this framework, we examine existing estimators based on propensity score weighting and outcome regression. We further propose a new double-robust DID estimator that hybridizes regression and propensity score weighting. We assess the fundamental parallel trend assumption in DID indirectly through analyzing the crash outcomes in the pre-treatment periods. A simulation study is conducted to demonstrate the advantage of the double-robust method over alternative methods. Our empirical results from a Pennsylvania Department of Transportation (PennDOT) data suggest that rumble strips are marginally effective in reducing the run-off-the-road and total crashes.
Liu, Leqi

Author(s) Email: leqil@andrew.cmu.edu

Institution: Machine Learning Department, Carnegie Mellon University

Poster Number: 43

Presenting Time: 6:00 P.M. - 7:30 P.M.

Offline Heterogeneous Policy Evaluation for Contextual Bandits: A Causal Approach: Poster 43

Abstract: Current methods for offline policy evaluations for contextual bandits concern themselves with coarse summaries, namely an average over a distribution of contexts. Recent advances in causal inference have moved beyond estimating the average treatment effect (ATE) to estimating heterogeneous treatment effects (HTE), which considers the treatment effect for a particular individual as described by a set of covariates. We study the similarity of estimating treatment effects and offline policy evaluation, survey their existing methods, and propose a doubly robust estimator using recent advances in estimating HTE to improve offline policy evaluation in the bandit setting. These methods are important in applications ranging from personalized medicine to social policy-making to online advertising.
Lorenzi, Elizabeth  
**Author(s) Email:** ecl21@duke.edu  
**Institution:** Duke University  
**Poster Number:** 45  
**Presenting Time:** 6:00 P.M. - 7:30 P.M.

**Heterogeneous Treatment Effects with Subgroups via the Overlap Weights:**
**Poster 45**

**Abstract:** Learning heterogeneous treatment effects across subgroups is becoming the norm rather than the exception in most comparative effectiveness research. Covariate balance is crucial to valid causal inference. The literature has shown that in the context of linear propensity models, exact mean balance is sufficient for unbiased treatment effect estimation, regardless of whether other aspects of the propensity model are correct. However, in practice, propensity scores are estimated using the full sample, which might balance the covariates in the overall population, but not necessarily in the subgroups. This results in the problem that covariate imbalance in subgroups leads to bias in estimating subgroup causal effects. We propose the use of overlap weights to achieve both overall and subgroup balance and address the bias-variance tradeoff. We prove that a properly specified propensity model including interactions between subgroups and covariates results in exact covariate balance within subgroups. In high dimensional cases, we suggest combining overlap weights with machine-learning methods (e.g. Lasso) to better select the appropriate interaction terms in the propensity score model. We explore these ideas through simulations, showing that overlap weights provide better subgroup balance compared to inverse-probability weighting (IPW), regardless of how the propensity score is estimated. Among linear regression, Lasso, and boosted regression, Lasso performs the best in terms of balancing between bias and variance in estimating the propensity score.
Data Based Covariate Selection for High Dimension Low Sample Size Data:
Poster 59

Abstract: With few exceptions, the propensity score literature has focused on estimating causal effects with moderate to large sample sizes. In the social and medical sciences, however, non-equivalent comparison group designs with small sample sizes are not atypical. Conditioning on many covariates in an attempt to satisfy the ignorability assumption may lead to p > n estimation problems or inefficiency. In such cases, data-driven algorithms for selecting minimum covariate subsets may be useful. The primary aim of this study is to investigate the properties of three data-driven covariate selection techniques when used with small sample sizes under varying conditions. Stepwise logistic regression, Bayesian networks and random forests, are studied in a Monte Carlo simulation. In each scenario, we simulate small samples ranging from 50 to 500 with 90 noise covariates and 10 target covariates that have some association with either the propensity score or the potential outcomes. We generate data from several DAGs and implement de Luna, Waernbaum & Richardson’ (2011) algorithms for covariate selection. Rosenbaum and Rubin’s (1984; 2009) stepwise logistic regression approach is used as benchmark for comparison. The simulation results indicate random forest and Bayesian networks (using mutual information) outperform stepwise logistic regression by successfully reducing the dimension of the data set and including appropriate covariates suggested by the backdoor path criteria. Propensity scores based on selected covariate sets are used to assess bias and mean square error for each method. Results and implications for covariate selection with small samples are discussed.

Reference

Targeted Maximum Likelihood Estimation of Causal Effects Based on Observing a Single Time Series: Poster 42

Abstract: Causal inference from time-series data is a crucial problem in many fields. In particular, it allows tailoring interventions over time to evolving needs of a unit, painting a granular picture of the current status. In medicine, wealth of information available in time-series data hints at an exciting opportunity to explore the very definition of precision medicine-studies that focus on a single person. We present targeted maximum likelihood estimation (TMLE) of data-dependent and marginal causal effects based on observing a single time-series. A key feature of the estimation problem is that the statistical inference is based on asymptotics in time. We focus largely on the data-dependent causal effects that can be estimated in a double robust manner, therefore fully utilizing the sequential randomization. We propose a TMLE of a general class of averages of conditional causal parameters, and establish asymptotic consistency and normality results. Finally, we demonstrate our general framework for the data-adaptive setting with a number of examples and simulations, including a sequentially adaptive design that learns the optimal treatment rule for the unit over time.
Martínez, Sebastián  
Author(s) Email: mail@smartinez.co  
Institution: University of Glasgow  
Poster Number: 29  
Presenting Time: 6:00 P.M. - 7:30 P.M.

Causal, Inferential, Dynamic Network Analysis for Public Health: Poster 29

Abstract: The main problem I am trying to _understand_ in this poster is that of confoundedness between social influence and social selection. Recent research in the social network analysis literature has hinted at how, without very specific information or very strong assumptions, it is impossible to disentangle and identify these two factors. The part of my research relevant for this conference aims at quantifying the effect assumptions have on the conclusions arrived at by researchers. This approach has two main objectives: gathering possible identification strategies that might lead to causal inference, and, therefore, broadening the toolset researchers have when dealing with observational data in social network research. More generally, through this research I want to study the concept of bounded causality in network dynamics. This is, to what extent can we say something about the causal effect a certain variable has on an observed outcome. Social influence (sometimes referred to as peer or spillover effects) is very similar to the idea behind interference, i.e. when an intervening variable *may* have an effect on non-intervened units, something well studied in the causal inference literature. Identifying the source of the interference, and trying to control for it, depends on the way data was gathered. Network data can be divided into four groups: clustered and randomized, clustered but not randomized, randomized but not clustered, and neither clustered nor randomized. Most observational network data available lie in the last category, mostly because causal inference was not the main goal of gathering said data. All other types of observational data allow for some degree of independence allowing for a more traditional causal inference approach (‘normal’ or two-step randomization, for example). Dealing with the problem of confoundedness in social network analysis using the interference framework may shed new light. Some of the possible identification strategies available in the literature include: mendelian randomization as a way of finding independent tools (although the social implications of this research are debatable); a more general instrumental variables approach (an approach that has been hinted at in recent publications); propensity score matching (but some have claimed that this is not possible without strong assumptions); relational event modelling (as social networks, human and otherwise, are inherently relational). It is my goal to understand the implications assumptions made in each one of these cases can
have. The insights developed through the use of this poster will feed into the general research done for my Ph.D. thesis. This deals with extensions and implementations to the temporal exponential random graph model and the temporal network autocorrelation model, and looks to develop a specific set of guidelines on how to carry out a network study in the field of public health.
Mauro, Jacqueline

Author(s) Email: jmauro@andrew.cmu.edu

Institution: Carnegie Mellon University

Poster Number: 51

Presenting Time: 6:00 P.M. - 7:30 P.M.

Nonparametric Estimation of the Causal Effects of Optimal Sorting: Poster 51

Abstract: We propose nonparametric doubly-robust methods to estimate the effects of optimally sorting a population to a discrete set of treatment values. This work is motivated by the question of how best to sort prisoners among 25 Pennsylvania prisons in order to minimize recidivism. In this case, each inmate's assignment depends on the assignment of all other prisoners due to capacity constraints, posing a violation of common causal assumptions about treatment assignment. In order to estimate the effects on recidivism of such an optimal assignment, we propose new resampling and influence-function based estimators.
Counterfactual Fairness in Decision Making with Risk Assessment Tools: Poster 4

Abstract: Risk assessment tools are increasingly used as decision aids in the criminal justice system. These tools aim to predict whether people are likely to commit future crimes, and they are used in various stages of the justice process, including bail setting, pre-trial release decisions, and parole conditions. Proponents claim that these tools can lead to better decisions that reduce incarceration rates and costs without increasing crime. However, there has also been much debate in the past several years about whether these tools are racist or lead to disparate racial outcomes, even when they do not directly take race into account. A highly publicized ProPublica article (Angwin et al., 2016) that investigated one particular risk assessment tool found that false positive rates for black defendants (i.e., a prediction of recidivism when defendants were not known to reoffend) were much higher than for white defendants, while false negative rates (a prediction of no recidivism followed by known re-offense) were much higher for white defendants. False positive and false negative rates are just two of a number of measures that arise in discussions of fairness. Many analyses of algorithmic bias turn instead to predictive parity, under which the probability of a future crime among people who are classified as high risk is the same regardless of their race. Indeed, a counterpoint to the ProPublica analysis found that the risk assessment tool in question satisfied predictive parity (Dietrich et al., 2016). Recent work by Chouldechova (2017) showed that these apparently conflicting results were not accidental but inevitable: predictive parity and error rate balance cannot both be achieved when recidivism rates differ across groups. Kleinberg, et al. (2016) similarly showed that several common measures of fairness cannot be simultaneously satisfied except under constrained and highly improbable circumstances. These results render nontrivial the question of how best to characterize fairness, since different notions of fairness are generally incompatible. Discussions of fairness are generally couched in terms of observed outcomes, such as recidivism rates, but risk assessment tools can only be deemed effective insofar as they change one or more outcomes of interest. Additionally, (un)fairness arises from the use of risk assessment tools in context: a tool with perfect predictive accuracy could nevertheless lead to unfair outcomes depending on how a judge chose to utilize it (Stevenson, 2018).
We suggest that it makes sense to estimate counterfactual versions of fairness metrics in context: how good is a risk assessment tool at predicting whether recidivism would occur if the tool were not used, taking into account the intervention that the tool’s prediction would lead to? We show under several simple models that observed false positive and false negative rates almost necessarily differ from their counterfactual versions, and that the differences depend on the interventions that the use of the tool leads to.
Mitchell, Shira  
Author(s) Email: smitchell@mathematica-mpr.com  
Institution: Mathematica Policy Research  
Poster Number: 38  
Presenting Time: 6:00 P.M. - 7:30 P.M.

A retrospective control study of the Millennium Villages Project: Poster 38  
Abstract: The Millennium Villages Project (MVP) was a ten-year, multi-sector, rural development project implemented in ten sub-Saharan African countries. We estimate the MVP's effect on a variety of development indicators. Causal inference for the MVP context presents many challenges: a nonrandomized design, limited baseline data for candidate control areas, and the assignment of treatment to only ten sites, limiting effective sample sizes. We carry out a matching procedure tailored to small samples and designed to facilitate communication with subject-matter experts. We fit a hierarchical Bayesian model that partially-pools across multiple sites and multiple outcomes to ameliorate the problem of "multiple comparisons", and compare to results from a classical analysis. Our model is fit in Stan, a state-of-the-art platform for statistical modeling.
Local independence graphs for time series: Poster 58

Abstract: Local independence was originally introduced as an asymmetric notion of independence in continuous-time stochastic processes and local independence graphs have been defined as means for graphical modeling of such structures. In a local independence graph each node represents an entire coordinate process of the system, thus time is implicit in the graph. Local independence can also be defined for time series and has very close ties to Granger causality. We consider graphs in which time is explicit, so-called unrolled graphs, and define maps to go back and forth between these graphs and local independence graphs. One can show that the graphical separation models are equivalent in a suitable sense. The graphical translation aids the understanding of local independence in time series models and illustrates that the notion of local independence is relevant for dynamical structural causal models. Starting from a dynamical structural causal model with a certain structural time homogeneity, the framework allows one to find a graphical representation that concisely represents the causal structure of the model. This also holds in the presence of latent processes that act as confounders in the system.
**Montero-Hernandez, Samuel**  
**Author(s) Email**: samuel@inaoep.mx  
**Institution**: Instituto Nacional de Astrofísica, Óptica y Electrónica  
**Poster Number**: 56  
**Presenting Time**: 6:00 P.M. - 7:30 P.M.

**Intervals of Causal Effects for Learning Causal Graphical Models**: Poster 56  
**Abstract**: Structure learning algorithms aim to retrieve the true causal structure. However, under common assumptions, only an equivalence class can be recovered and a unique model cannot be singled out. We hypothesized that unsettled relations could be resolved through the assessment of causal effects intervals among the involved variables. We introduce SLICE (Structural Learning with Intervals of Causal Effects), a new algorithm to decide on unresolved relations, which taps on the computation of the causal effects and the acceptability index; a strategy for intervals comparison. We generated several synthetic datasets varying the number of variables, the density and the number of drawn samples. Comparison against LiNGAM algorithm is made to establish the performance of the algorithm over the simulated scenarios for validation purposes. By using the normalized structural hamming distance (SHD) as assessment metric, the estimated models were contrasted with the true models. The retrieved structures with SLICE showed smaller SHD values compared to LiNGAM, improving the structure of the retrieved causal model regarding correctly found edges. The proposed strategy represents a new complement tool for deciding unraveled causal relations in the presence of observational data only.
Abstract: With the massive expansion of available data and advancements in machine learning algorithms, increasingly important decisions are being automated. Unfortunately, this increases the potential for discriminatory biases to become “baked in” to automated systems that influence people’s lives. Without careful adjustments for these biases during learning and deployment of automated systems, these systems could indeed put certain individuals at risk of discrimination. Here, we consider discrimination with respect to a sensitive feature, such as race or gender, that arises in inference problems involving outcome variables, such as in classification or regression problems. Research on fairness requires mathematical definition of fairness. In discussing the extent to which a particular approach is “fair” we believe the gold standard is human intuition. Inspired by formal and informal definitions which appeared in the legal and causal inference literature (Pearl 2009, Bertrand and Mullainathan 2004), we propose that discrimination ought to be formalized as the presence of certain path-specific effects (PSEs). The specific paths which correspond to discrimination are a domain specific issue. For example, a path from gender to the result of a physical test to hiring may be appropriate for a fire department, but inappropriate for an accounting firm. To represent discrimination formally, we assume the observed data distribution is induced by a causal model, and the PSE is identified as a function of observed data. We fix upper and lower bounds on the PSE, representing the degree of discrimination we are willing to tolerate. Our proposal is to transform the inference problem on the observed data distribution, the “unfair” world, into an inference problem on another distribution, the “fair” world, which is close, in the KL-divergence sense, to the observed data distribution while also having the property that the PSE lies within the specified bounds. Given a finite dataset, we frame the problem as one where we maximize the likelihood subject to constraints that restrict the magnitude of the PSE. Crucial to our proposal, statistical inference on previously unseen instances cannot be carried out until instances are mapped onto the fair world. This is because unseen instances will generally be drawn from unfair world, rather than the fair world. We propose a simple conservative approach for doing so, although others are possible. One of the advantages of our approach is it can be readily extended to concepts like affirmative action and “the wage
gap” in a way that matches human intuition. One methodological difficulty with our approach is the need for a computationally challenging constrained optimization problem. An alternative is to think about re-parameterizing the observed data likelihood in terms of causal parameter for an arbitrary PSE (which is currently an open problem).
Nabi, Razieh

Author(s) Email: rnabiab1@jhu.edu

Institution: Johns Hopkins University

Poster Number: 25

Presenting Time: 6:00 P.M. - 7:30 P.M.

Semi-Parametric Causal Sufficient Dimension Reduction Of High Dimensional Treatments: Poster 25

Abstract: Cause-effect relationships are typically evaluated by comparing the outcome responses to binary treatment values, representing cases and controls. However, in certain applications, treatments of interest are continuous and high dimensional. For example, in oncology, the causal relationship between severity of radiation therapy, represented by a high dimensional vector of radiation exposure values at different parts of the body, and side effects is of clinical interest. In such circumstances, a more appropriate strategy for making interpretable causal inferences is to reduce the dimension of the treatment. If individual elements of a high dimensional feature vector weakly affect the outcome, but the overall relationship between the feature vector and the outcome is strong, careless approaches to dimension reduction may not preserve this relationship. The literature on sufficient dimension reduction considers strategies that avoid this issue. Parametric approaches to sufficient dimension reduction in regression problems (Li 1991) were generalized to semi-parametric models in (Ma and Zhu 2012). Methods developed for regression problems do not transfer in a straightforward way to causal inference due to complications arising from confounding. In this paper, we use semi-parametric inference theory for structural models (Robins 1999) to give a general approach to causal sufficient dimension reduction of a high dimensional treatment.
Estimation of Optimal Path-Specific Policies: Poster 26

**Abstract:** The goal of personalized decision making is to map a unit’s characteristics to an action tailored to maximize the expected outcome for that unit. Obtaining high-quality mappings of this type is the goal of the dynamic regime literature. In healthcare settings, optimizing policies with respect to a particular causal pathway may be of interest as well. For example, we may wish to maximize the chemical effect of a drug given data from an observational study where the chemical effect of the drug on the outcome is entangled with the indirect effect mediated by differential adherence. In such cases, we may wish to optimize the direct effect of a drug, while keeping the indirect effect to that of some reference treatment. Combining mediation analysis and dynamic treatment regime ideas yields path-specific policies and counterfactual responses to these policies. We derive a variety of methods for learning high quality policies of this type from data, in a causal model corresponding to a longitudinal setting of practical importance. We illustrate our methods via a dataset of HIV patients undergoing therapy, gathered in the Nigerian PEPFAR program.
Model Assisted Sensitivity Analyses for Hidden Bias with Binary Outcomes: Poster 48

Abstract: In medical and health sciences, observational studies are a major data source for inferring causal relationships. Unlike randomized experiments, observational studies are vulnerable to the hidden bias introduced by unmeasured confounders. The impact of unmeasured covariates on the causal effect can be assessed by conducting a sensitivity analysis. A comprehensive framework of sensitivity analyses has been developed for matching designs. Sensitivity parameters are introduced to capture the association between the missing covariates and the exposure or the outcome. Fixing sensitivity parameter values, it is possible to compute the bounds of the p-value of a randomization test on causal effects. We propose a model assisted sensitivity analysis with binary outcomes for the general 1:k matching design, which provides results equivalent to the conventional nonparametric approach in large sample. By introducing a conditional logistic outcome model, we substantially simplify the implementation and interpretation of the sensitivity analysis. More importantly, we are able to provide a closed form representation for the set of sensitivity parameters for which the maximum p-values are non-significant. This methodology can be easily extended to matching designs with multilevel treatments. We illustrate our method using a U.S. trauma care database to examine mortality difference between trauma care levels.
On the difference method for mediation analysis in generalized linear models:

**Abstract:** Causal mediation analysis is often carried out to assert whether the effect of an exposure on an outcome of interest is mediated by other covariates. In spite of recent developments, the classical "difference method", that compares estimates from a model with and without the suspected mediator, remains popular. Assumptions for causal interpretation of these estimators were previously provided. However, the validity of fitting separate models was not investigated, and asymptotic variance estimator was previously unavailable. We formulate the problem for generalized linear models. We first consider the issue of whether the same link function holds for the marginal and conditional models with respect to the mediator, a property which we term as "g-linkability". Under g-linkability, we utilize a data duplication algorithm together with a generalized estimation equations approach to estimate direct and indirect effects and to derive a variance estimator that is calculated very fast. We investigate the performance of the estimator and its sensitivity to violation of g-linkability in simulations.
Nie, Xinkun

Author(s) Email: xinkun@stanford.edu

Institution: Stanford University

Poster Number: 37

Presenting Time: 6:00 P.M. - 7:30 P.M.

Quasi-Oracle Estimation of Heterogeneous Treatment Effects: Poster 37

Abstract: We develop a general class of two-step algorithms for heterogeneous treatment effect estimation in observational studies. We first estimate marginal effects and treatment propensities to form an objective function that isolates the heterogeneous treatment effects, and then optimize the learned objective. This approach has several advantages over existing methods. From a practical perspective, our method is very flexible and easy to use: In both steps, we can use any method of our choice, e.g., penalized regression, a deep net, or boosting; moreover, these methods can be fine-tuned by cross-validating on the learned objective. Meanwhile, in the case of penalized kernel regression, we show that our method has a quasi-oracle property, whereby even if our pilot estimates for marginal effects and treatment propensities are not particularly accurate, we achieve the same regret bounds as an oracle who has a-priori knowledge of these nuisance components. We implement variants of our method based on penalized regression in a variety of simulation setups, and find promising performance relative to existing baselines. Joint work with Stefan Wager.
Insights on Variance Estimation for Blocked and Matched Pairs Designs: Poster 55

Abstract: The current literature on blocking has two main branches. The first focuses on larger blocks, with multiple treatment and control units in each block. The second focuses on matched pairs, with a single treatment and control unit in each block. In the former case, variance estimation is straightforward as one can estimate the variance of units within each group and treatment arm. In the latter case, this is not possible so researchers have proposed alternative estimators that look at variance across the blocks. These alternative estimators have been evaluated under different assumptions than found in the large block literature. Neither literature handles cases with blocks of varying size that contain singleton treatment or control units. Differences in the two literatures has also created some confusion regarding the benefits of blocking in general. In this paper, we reconcile the literatures by carefully examining the performance of different estimators of treatment effect and of associated variance estimators under several different frameworks. We provide variance estimators for experiments containing blocks of different sizes, allowing for singleton units in a treatment arm. We finally discuss in which situations blocking is or is not guaranteed to reduce the variance of our estimator.
Causal Inference for Observational Data with Partially Observed and Time-Dependent Treatment Assignment: Poster 6

Abstract: Large health care databases offer a rich data source for analyzing treatment practices and evaluating outcome trajectories across a diverse patient population. It is often of interest to use these sources to evaluate the causal effects of a particular therapeutic treatment such as medication therapy. While initiation of medical therapy (time of treatment assignment) is typically recorded for study participants receiving treatment, defining the initialization time of “control” presents a philosophical challenge for causal inference. We consider estimating the causal effects of the administration versus withholding of a therapeutic treatment in observational studies when the time of treatment assignment is only partially observed. Here, participants that are not observed to receive treatment during the study period may receive the active treatment eventually or may be controls with unknown times of assignment. In studies with a time-to-event outcome, which is defined relative to the time of treatment assignment, these unknown times of assignment need to be inferred in order to make valid comparisons between treatment and control groups. We present a framework for conceptualizing time-to-event outcomes in this setting, which relies on separating the process that governs the time of treatment assignment for each unit from the mechanism that determines the assignment received by that unit. Next, we develop a two-step inferential approach based on: 1) inferring the missing times of treatment assignment, and 2) estimating the causal effects of treatment based on inferential conclusions from Step 1. This approach allows us to incorporate uncertainty about times of treatment assignment, which induces uncertainty in measurement of the outcomes. We apply these methods to study the effects of prescribing contraindicated phosphodiesterase type 5 inhibitors (PDE5Is) for treatment of pulmonary hypertension using administrative data from the Veterans Affairs (VA) health care system.
Fine-Grained Compliance and Treatment Effects in the Cognitive Tutor Effectiveness Trial: Poster 36

Abstract: In a large-scale randomized effectiveness trial of the Cognitive Tutor Algebra I curriculum, whose centerpiece is an educational computer program, researchers estimated indeterminate average effects on a post-test in the first year of implementation, and positive effects in the second year. These estimated effects mask considerable, observable, variation. First, students in the treatment condition varied widely in how, and how much, they used the software. Specifically, the standard version of the curriculum contained 37 units, but the median number of units worked was 10. In essence, each unit defines a type of compliance—a student who would work on problems from that unit if assigned to treatment is a complier for that unit. Second, the 31 post-test items measured different Algebra I skills, and treatment effects varied from problem to problem. This is especially notable since some of the problems were unrelated to the Cognitive Tutor software. In essence, there were 31 separate outcomes measured. This poster will present a Bayesian principal-stratification model to estimate a large set of complier average treatment effects (CATEs)—one CATE for each unit-problem combination. The model first predicts which students would work on which units, if assigned to treatment, and uses those predictions to estimate treatment effects for each post-test item, within an item-response model. Data-driven partial pooling regularizes each stage of the model. The fine-grained notion of compliance and of problem level treatment effects has the potential to enhance and clarify our understanding of the Cognitive Tutor and other complex interventions.
Opportunities for Causal Discovery in the Geosciences: Poster 3

Abstract: MOTIVATION: The earth is an incredibly complex system with countless mechanisms interacting with each other across space and time. While geoscientists have successfully identified many of these interactions, many others are yet unknown. Climate change adds urgency to the search for a deeper understanding of mechanisms, such as the strongest causal pathways that create climate change and the effects to expect from altered atmospheric and ocean flow patterns, including increased occurrence of extreme weather events (floods, droughts, hurricanes, etc.).

CURRENT STATE-OF-THE-ART: By far the most common approach for causality analysis in climate science (and many other geosciences) is to compare uni- and bivariate regression models using Granger causality. A few research groups have applied structure learning algorithms based on probabilistic graphical models, graphical Granger models, or Gaussian models, and the insights generated are gaining the attention of geoscientists. The goal of this poster is to attract more causality experts to work with geoscientists on these types of studies.

NATURE OF GEOSCIENCE APPLICATIONS: Geoscience processes are temporal in nature and cannot be treated as static mechanisms. The temporal models needed to study these mechanisms can be achieved by using the PC algorithm with lagged copies of the original variables and temporal constraints, as introduced by Chu, Danks and Glymour (2005). This allows us to analyze systems with significant feedback loops. For example, we recently analyzed the interactions between Arctic temperature and jet streams (speed and latitude) and found indications of strong interactions between them, including positive (destabilizing) feedback loops and lag times at which they occur. A second type of application deals with spatio-temporal settings, e.g., tracing interaction of processes between (virtual) grid points around the globe. This type of study can provide insights, for example, into the changes of atmospheric flow patterns due to climate change (using the output of climate model predictions as input data) or the mechanisms that lead to stronger Asian monsoon seasons. These problems are high-dimensional, because with 100s to 1000s of grid points, several different fields, and many lagged copies, the number of nodes can reach 100,000. While our optimized implementation can handle that many
nodes, a harder problem arises from spatial autocorrelation, which leads to grid-based distortions in the results whenever anisotropic grids are used. This is because points close to each other are strongly affected by unmodeled environmental factors, masking true interactions between grid points. In some cases, one can apply causal discovery instead in a different domain, e.g., in the space of spherical harmonics, but this is not always feasible.

*QUESTIONS FOR FUTURE RESEARCH:* How can we avoid grid-based distortions? Which methods are most robust in high-dimensional spaces? Which methods are most sensitive to pick up weak signals? For nonlinear processes, how do we best assess the strength of causal interactions (after structure learning is completed)? Which methods can detect hidden common causes in high-dimensional settings? As ground truth is rarely available in the geosciences, how can we create more benchmarks from simulations to evaluate the causal discovery methods?
Causal Discovery of Feedback Networks with Functional Magnetic Resonance Imaging: Poster 22

Abstract: We test the adequacies of several proposed and two new statistical methods for recovering the causal structure of systems with feedback that generate noisy time series closely matching real BOLD time series. We compare: an adaptation for time series of the first correct method for recovering the structure of cyclic linear systems; multivariate Granger causal regression; the GIMME algorithm; the Ramsey et al. non-Gaussian methods; two non-Gaussian methods proposed by Hyvarinen and Smith; a method due to Patel, et al.; and the GlobalMIT algorithm. We introduce and also compare two new methods, the Fast Adjacency Skewness (FASK) and Two-Step, which exploit non-Gaussian features of the BOLD signal in different ways. We give theoretical justifications for the latter two algorithms. Our test models include feedback structures with and without direct feedback (2-cycles), excitatory and inhibitory feedback and models using experimentally determined structural connectivities of macaques. We find that averaged over all of our simulations, including those with 2-cycles, several of these methods have a better than 80% orientation precision (i.e., the probability a directed edge is in the true generating structure given that a procedure estimates it to be so) and the two new methods also have better than 80% recall (probability of recovering an orientation in the data generating model). Recovering inhibitory direct feedback loops between two regions is especially challenging.
Identification of Personalized Path-Specific Effects: Poster 39

Abstract: Unlike classical causal inference, where the goal is to estimate average causal effects within a population, in settings such as personalized medicine, the goal is to map a unit’s characteristics to a treatment tailored to maximize the expected outcome for that unit. Obtaining high quality mappings of this type is the goal of the dynamic treatment regime literature. In healthcare settings, optimizing policies with respect to a particular causal pathway is often of interest as well. For example, one may wish to maximize the chemical effect of a drug given data from an observational study where chemical effect is entangled with differential adherence. In such cases, one may wish to optimize the direct effect of a drug, while keeping the indirect effect mediated by adherence to that of some reference treatment. In the context of average treatment effects, direct and indirect effects are considered in the mediation analysis literature. In this work, we combine mediation analysis and dynamic treatment regime ideas and consider how unit characteristics may be used to tailor a treatment strategy that maximizes a particular path-specific effect. In particular, we define counterfactuals associated with the path-specific effects of a policy, give a general identification algorithm for these counterfactuals, and prove completeness of the algorithm for unrestricted policies. A corollary of our results is that the identification algorithm for responses to policies given by Tian in 2008 is complete for arbitrary policies.
Causal Inference for Polypharmacy: Propensity score estimation with multiple concurrent medications: Poster 20

Abstract: The poster provides a framework for causal estimation under multiple concurrent medications. Our parameter of interest is the marginal mean counterfactual outcome under different combinations of medications. We explore parametric and machine learning methods to estimate the generalized propensity score proposed by Imbens (2000). We then apply three causal estimation approaches (inverse probability of treatment weighting, propensity score adjustment, and targeted maximum likelihood estimation) to estimate the causal parameter of interest. We compare the results obtained using these methods in a simulation study with four potentially concurrent medications. We implement a second simulation study where specific combinations of medications either occur rarely or don’t occur in the dataset. Finally, we apply the methods explored to contrast different antibiotic regimens for patients with multiple drug-resistant tuberculosis.
Functional BART for Causal Inference: Poster 47

Abstract: Bayesian Additive Regression Trees (BART) has been shown to be an effective framework for modeling nonlinear regression functions, with strong predictive performance in a variety of contexts. The BART prior over a regression function is defined by independent prior distributions on tree structure and leaf or end-node parameters. Leaf parameters are typically modeled as independent and identically distributed from a carefully-scaled Gaussian distribution. While highly effective in practice, functions sampled from this prior are not smooth in any of the covariates. In many contexts, it is desirable to enforce or encourage a degree of smoothness in “special” covariates such as a time index. We develop a new version of the BART prior for modeling a function that evolves smoothly over time given other covariates. We demonstrate the utility of this approach by applying our model to a timely women’s health problem, where the outcome of interest is measured at different gestational ages for each subject, conditional on covariates. We discuss the benefits of this approach in a variety of women’s health and obstetrics modeling problems where gestational age is a typical covariate, as well as potential applications in a range of other disciplines.
Suzuki, Etsuji

Author(s) Email: esuzuki@hsph.harvard.edu

Institution: Harvard University

Poster Number: 63

Presenting Time: 6:00 P.M. - 7:30 P.M.

Covariate balance for no confounding in the sufficient-cause model: Poster 63

Abstract: The counterfactual approach to confounding has been widely accessible to epidemiologists, and the concept of confounding is now explained in the counterfactual framework. Much of the literature on this topic has explained that exchangeability between the exposed and the unexposed groups is a core concept to make causal inference. In this context, covariate balance is often addressed as a key feature to control confounding in epidemiology, and many researchers have been concerned about whether covariate balance is achieved between the exposed and the unexposed groups in their analyses. Despite its significance, however, a covariate is broadly defined as a “variable that is possibly predictive of the outcome under study”, and the term “covariate” has been often used interchangeably with the term “confounder”. In this presentation, we aim to show conditions of covariate balance for no confounding in the sufficient-cause model and discuss its relationship with exchangeability conditions. In so doing, we consider the link between the sufficient-cause model and the counterfactual model, emphasizing that the target population plays a key role when discussing these conditions. Furthermore, we incorporate sufficient causes within the directed acyclic graph framework. We propose to use each of the background factors in sufficient causes as representing a set of covariates of interest and discuss the presence of covariate balance by comparing joint distributions of the relevant background factors between the exposed and the unexposed groups. This would fit the concept of covariate balance because one innately focuses on the “factor” or “mechanism” that induces confounding when discussing it. We show conditions for partial covariate balance, covariate balance, and full covariate balance, each of which is stronger than partial exchangeability, exchangeability, and full exchangeability, respectively. This is consistent with the fact that the sufficient-cause model is a “finer” model than the counterfactual model. Although our conceptualization of covariate imbalance is closely related to the counterfactual-based definition of a confounder, the concepts of covariate balance and confounder should be clearly distinguished. In conclusion, we proposed a mapping between covariate balance under the sufficient-cause model and exchangeability conditions in the counterfactual model, highlighting the facts that covariate balance is a stronger condition than no confounding and that the required covariate balance
depends on the target population of interest. Our formalization of the notion of covariate balance will be useful in clarifying the meaning of confounding.
Martingale residual-based imputation for unmeasured confounders in marginal structural Aalen additive hazard model using validation sample: Poster 54

Abstract: The information of additional confounders from smaller-scale validation subsample of the dataset can be used to overcome the drawback of routinely-collected medical records database, which often fails to collect potential confounders’ information. Recently, Burne & Abrahamowicz (2016, 2017) proposed the martingale residual (MR)-based imputation method in Cox proportional hazard model as novel method that uses the validation sample to adjust for unmeasured confounding. In this study, to increase the flexibility of the MR-based imputation method in time-to-event analysis, we will propose the imputation method based on the MRs constructed by Aalen additive hazard model. Aalen model can be a more flexible alternative to Cox model since it doesn’t need proportional hazard assumption and easily incorporates time-varying covariate effects. MRs based on Aalen model are used as the proxy of outcome variable for imputation models in our proposed method instead of those based on Cox model as originally proposed. Specifically, MRs of each observation are calculated through “reduced” Aalen model including only exposure and covariates, which were fully observed throughout the dataset. The covariates which were observed in only validation sample are subsequently imputed by conditional models including MRs. In the simulation study, a setting designed to evaluate the effect of exposure X (e.g., antihyperlipidemic use) on time-to-event Y (e.g., diabetes mellitus) was used. We assumed that the status of exposure and covariates were time-fixed, but their effects were time-varying. We generated fully observed covariates V, partially observed covariates U, X, time-to-censor C and Y under additive hazard setting and subsequently artificially missed U from 90.0–97.5% of patients. Partially missing covariates U were multiple-imputed (10-times) by fully conditional specification method which used conditional models including MRs calculated by reduced Aalen model. Propensity scores for exposure X were calculated by “full” logistic model including V and (partially) imputed U. Finally, cumulative hazard difference for X was estimated by inverse probability of treatment weighted Aalen model for each imputed dataset and was combined according to Rubin’s rule. As a result, the proposed method reduced the bias caused by unmeasured confounding.
under following scenarios: 1) non-null or null exposure effect, 2) various levels of the unmeasured confounding, 3) U were missed according to missing-at-random or missing-not-at-random, 4) C were exponentially or Weibull distributed, and 5) the distributions of covariates were different between validation sample and everyone else. Additional simulation scenarios demonstrated the equivalent performance of the method in the presence of time-varying exposure and unmeasured covariates acting as both confounders and predictors of the subsequent exposure (i.e., exposure-confounder feedback). The imputation method for marginal structural Aalen model is illustrated in Japan Medical Data Center claims database, where the risk of diabetes mellitus among statin initiators compared with other antihyperlipidemic agents’ users. In the case-study, we defined the patients’ sex, age, concurrent medications and past medical histories as fully observed covariates V and defined the clinical and laboratory data as partially observed covariates U. Our poster will further demonstrate the details of the settings and results for both simulation and case studies.
Probabilistic Matching: Incorporating Uncertainty to Improve Propensity Score Matching: Poster 17

Abstract: Matching methods such as propensity score matching are commonly used to construct artificial treatment and control groups from observational data to determine the causal effect of treatment. However, propensity scores, once estimated, are frequently treated as known, and the uncertainty inherent in their estimation is ignored. Two exceptions are: firstly, some variance estimates of treatment effect estimates now include the uncertainty of the first-step propensity score estimation procedure. While reporting accurate estimates of uncertainty is important, we posit that this uncertainty can be further leveraged to improve the second-step estimation of treatment effects, which motivates the proposed procedure, probabilistic matching. Secondly, Bayesian approaches to address uncertainty in the estimation of propensity scores have been proposed. In this paper, we offer a frequentist approach. We introduce probabilistic matching, a proposed improvement on propensity score matching, that incorporates the uncertainty of the estimated propensity score into the subsequent matching process. This is done by weighting matches by the estimated probability of matching. Notably, this is equivalent to averaging the estimated treatment effect over the propensity score distribution, given the data. Probabilistic matching can be used to improve any propensity score estimation model with computable uncertainty. We demonstrate it on propensity scores estimated using logistic regression, random forests, and covariate-balanced propensity scores -- all of which have computable uncertainty. Probabilistic matching achieves comparable bias and lower variance when compared to vanilla propensity score matching. We draw connections and illustrate differences between probabilistic matching and common matching schemes, such as nearest-neighbor matching and matching using a caliper. While we focus on matching in this paper, the idea of incorporating uncertainty can also be brought into other ways of utilizing estimated propensity scores, such as weighing and sub-stratification.
Analyzing the modified outcome method for estimating heterogeneous treatment effects: Poster 40

Abstract: We wish to estimate the expected treatment effect in a randomized experiment as a function of observed covariates; we explore the modified outcome method for this purpose. In the simplest case, where the two equiprobable treatment assignments are indicated by $W_i$ and the observed outcome by $Y_i$, the modified outcome method fits a regression model to $Y_i^* = 2(W_i - 1) Y_i$, a quantity equal, in expectation, to the individual treatment effect. We provide a new motivation for the modified outcome method and introduce notions of optimality for regression adjustment when using the method to estimate the conditional average treatment effect (CATE). We show that all optimal estimators of the CATE can be cast as modified outcome estimators with optimal regression adjustment and, under minimal assumptions, regression adjustment reduces the asymptotic variance of the estimator without increasing its bias, even when the regression adjustment model is not correctly specified. We provide asymptotically correct confidence intervals for parametric models of the treatment effect using Huber-White standard errors. And we show that if the true treatment effect is a linear function of the observed covariates, using the modified outcome method with the lasso converges to the CATE when $\log p << n$. Finally, we demonstrate in simulations and in a real data example that the modified outcome method with optimal regression adjustment is competitive with existing approaches. This is joint work with Jasjeet Sekhon and Bin Yu.
The Blessings of Multiple Causes: Poster 1

**Abstract:** Many scientific questions are causal, while many scientific studies are observational. To make causal claims with observation data, we often assume strong ignorability. It requires all confounders be observed. This strong ignorability assumption is standard yet untestable; it is impossible to determine if it holds or not. When it does not hold, how to perform causal inference is largely unexplored. In this paper, we present multiple causal inference: a class of problems where strong ignorability is no longer necessary. Multiple causal inference works with two or more causes. Multiple causes bring blessings into causal inference: their dependency structure elicits information about unobserved multi-cause confounders; this information allows us to correct for unobserved confounders with latent variable models. We hence propose the deconfounder for multiple causal inference: it first infers a substitute confounder that renders all causes conditionally independent, and then estimates the potential outcome function conditional on the substitute confounder. We demonstrate both theoretically and empirically that the deconfounder protects us from unobserved confounders: it leads to unbiased causal estimates. It also helps researchers identify potential unobserved confounders.
Identify Heterogeneous Treatment Effect and Confounder's Effect via $L_1$-Regularized Nonlinear Regression: Poster 44

Abstract: Recently, in causal inference, two problems are widely discussed among statistic and social science researchers: the identification of heterogeneous treatment effect and the unbiased estimation of structure parameters for treatment effect when many nuisance parameters exist in the functions of confounders. Although researchers have developed some powerful machine learning-based methods that can solve two problems well enough, it is still hard to estimate the "pure" heterogeneous treatment effect, which only affects the outcomes of treatments, since we need to identify whether an effect comes from heterogeneous factors or confounders in a nonlinear setting. In this paper, we develop a $L_1$-regularized nonlinear variable selection method to identify the heterogeneous treatment effect and confounders' effect separately. We also prove that, given several conditions, this method can identify correct subset of variables consistently in a nonlinear setting. In simulation experiment, we construct a $L_1$-regularized soft-decision tree and find it outperforms other variable selection methods substantially. At last, we also apply this method into an empirical study.
Debiased inference on treatment effect in high dimensional model: Poster 64

Abstract: This paper considers the problem of treatment effect estimation and inference on a scalar outcome when a large number of covariates are present in a linear or partially linear model. When inference on the treatment effect is made after variable selection, the risk of bias cannot be ignored. While the estimation bias in an under-fitted model is well understood, we address a lesser known bias that arises from an over-fitted model. We show that the over-fitting bias can be reduced or eliminated through data splitting, and more importantly, smoothing over random data splits is necessary to mitigate the loss of efficiency due to data splitting. To mitigate the risk of under-fitting and over-fitting, we incorporate a new projection based approach. Under appropriate conditions we show that the proposed estimators for the treatment effect are asymptotically normal and their variances can be well estimated. We discuss the merit of these methods and provide comparisons of finite-sample performance in a variety of settings.
Causal Time Series Classification: Poster 7

Abstract: We propose a new method of discovering causal relationships in temporal data based on the notion of causal compression. We adopt the setting of Pearlian graphs and the formalism of Pearl's do calculus. Expressing them in terms of information theoretic concepts capturing the difference between interventional and observational distributions resulted in rich literature. This has led to asymmetrical information theoretic measures being used for modelling causal relationships in graphical models. We first specify conditions under which the information theoretic approach is well defined and can be imbued with a causal interpretation as defined by the interventional calculus. We relate to the points of criticism towards directed information raised in literature and show how they are handled by the information theoretic paradigm we propose. We then unify existing frameworks of directed information (in time series, time series with reduced assumptions regarding time ordering and in general Pearlian graphs) and use the directed information as an information theoretic tool for quantifying causality. We introduce chain rule for directed information and use it to motivate causal sparsity. We prove that modelling of causality in our information theoretic framework only requires estimating the copula density of the data distribution and thus does not depend on its marginals. Subsequently, we show an application of the proposed method: causal time series classification which selects time points capturing the causal flow between time points belonging to different signals. We evaluate the method on EEG data. Unlike cognate approaches to causality modelling in time series, we do not treat whole time series as nodes and do not model relationships between such nodes. We propose to model causal relationships between specific time points of the time series instead. We also do not have to make any assumptions concerning stationarity.