STAT 9220
Advanced Causal Inference
Prof. Eric Tchetgen Tchetgen
Fall 2024

This course will provide an in depth investigation of statistical methods for drawing causal inferences from complex observational studies and imperfect randomized experiments. Formalization will be given for key concepts at the foundation of causal inference, including: confounding, comparability, positivity, interference, intermediate variables, total effects, controlled direct effects, natural direct and indirect effects for mediation analysis, generalizability, transportability, selection bias, etc.... These concepts will be formally defined within the context of a counterfactual causal model. Methods for estimating total causal effects in the context of both point and time-varying exposure will be discussed, including regression-based methods, propensity score techniques and instrumental variable techniques for continuous, discrete, binary and time to event outcomes. Mediation analysis will be discussed from a counterfactual perspective. Causal directed acyclic graphs (DAGs) and associated nonparametric structural equations models (NPSEMs) will be used to formalize identification of causal effects for static and dynamic longitudinal treatment regimes under unconfoundedness and unmeasured confounding settings. This formalization will be used to define, identify and make inferences about the joint effects of time-varying exposures in the presence of (possibly hidden) time-dependent covariates that are simultaneously confounders and intermediate variables. These methods include g-estimation of structural nested models, inverse probability weighted estimators of marginal structural models, and g-computation algorithm estimators. Credible quasi-experimental causal inference methods will be described, leveraging auxiliary variables such as instrumental variables, negative control variables, or more broadly confounding proxy variables. Quasi-experimental methods discussed will include the control outcome calibration approach, proximal causal inference, difference-in-differences and related generalizations of these methods. Semiparametric efficiency and the prospects for doubly robust inference will feature prominently throughout the course, including methods that combine modern semiparametric theory and machine learning techniques.

Instructor: Eric Tchetgen Tchetgen. Office hour: TBD. (email ett@wharton.upenn.edu)

Prerequisite: STAT9210 or equivalent course.

Class notes, homeworks and exam.

There is no required textbook although some parts of the course will sometimes draw from specific books and articles. I will be mentioning the main source for each part as I present it. You will benefit from complementing the class notes with readings of the specific sections from these sources.

Grading system : Homeworks will carry 50%, final project will carry 50% of the grade.

Topics:

<u>PART I</u> Single Treatment Settings

1. Definition of counterfactuals and causal effects for point exposure:

2. Identification of causal effects and estimation of the single occasion g-formula under unconfoundedness: g-computation, ipw and doubly robust estimation.

4. Causal mediation analysis and counterfactual effect decomposition

<u>PART II</u> Time Varying Treatment Settings

5. Generalizations: Nonparametric structural equations model with independent errors (NPSEM-IE), Finest fully randomized causally interpretable structured tree graph (FFRCISTG) for static and dynamic regimes in complex longitudinal studies.

6. Generalizations of g-formula, Marginal structural models and structural nested models for static and dynamic longitudinal treatment regimes.

7. Optimal dynamic treatment regimes.

8. Causal inference in the presence of censoring by death.

<u>PART III</u> Advanced Topics

9. Partial identification of causal effects

10. Instrumental variable Methods for point treatment and time-varying treatments.

11. Causal inference on Networks: interference and homophily

12. Invalid instruments, Bespoke instruments, Negative Controls and Confounding proxies.