Syllabus for POLI 891: Advanced Topics in Causal Inference

Ye Wang
University of North Carolina at Chapel Hill

Fall 2024

Course description

This graduate-level seminar aims to equip participants with a modern perspective on the progress in causal inference and the skill to apply these frontier methods to social science studies. The course consists of two separate yet interdependent parts. The first part includes six lectures on the foundation of causal inference and causal analysis of cross-sectional data. We start from basic concepts such as the Neyman-Rubin model and counterfactual. Then, we discuss various experimental designs, how to conduct estimation and statistical inference under each of them, as well as the connection between experiments and observational studies. We conclude this part with a review of covariate adjustment techniques in observational studies, including matching, weighting, regression, balancing, and doubly robust estimation.

In the second part, students will give presentations on three selected topics in political methodology: basic algorithms in machine learning (2 weeks), the application of machine learning in causal inference (2 weeks) and panel data analysis (3 weeks). Students are expected to present their own research design in the last week.

Time and location

Classes: 12:20pm-3:10pm on Mondays, Room 0351 in Hamilton Hall.
There will no class on Oct. 16 (Invited talk) or Nov. 27 (Thanksgiving).
Office hours: 12:30pm-3:30pm on Tuesdays, Room 322 in Hamilton Hall.

Texts and software

The course will draw a lot from the following textbooks:

1. Imbens and Rubin (2015)
2. Hernán and Robins (2010)
3. Gareth et al. (2013)

We may also refer to other textbooks and research papers for certain topics. More details can be found in the section on course outline. It is expected that students try to read materials listed as references before class.

We will working with R in this course, which is an open-source computing language that is very widely used in statistics. You can download it for free from www.r-project.org. You are also encouraged to use Rmarkdown for your homework.

Requirements

Students enrolling in this course are expected to have a solid understanding of probability theory, matrix algebra, and calculus, along with practical experience in scripting with R. It is highly recommended that participants have completed POLI-783 and POLI-784 offered by the Department of Political Science or a comparable course.

Grading of the course is based on class participation (20%), presentation (40%), and the final proposal (40%). In the second part of the course, you are expected to read the assigned papers or book chapters carefully prior to class and participate actively in the discussion of these papers. Each student needs to give two to three presentations during the seminar. The presentation should cover the basic ideas of the method, assumptions and contexts for it to work, how to implement it in R, and its potential applications in political science.

You are required to submit a two-page description of a research proposal in the middle of the semester (before October 23) and a complete draft of the proposal before December 1. There is no restriction on the research question in the proposal. But it must include a section in which you justify your research design from the design-based perspective, such as what assumptions you have imposed for causal identification, why they are convincing in your context and what methods are proper in this case. In the last week of the course, each student will give a 10-minute presentation on the proposal that is going to be submitted.

Course outline

Section I: Lectures on the foundation of quantitative social science

Lecture 1: Introduction

Week 1, August 21

Estimand, estimator, and estimate.
Measure the quality of an estimator.
The Neyman-Rubin model.
The fundamental problem of causal inference: two solutions.
Random assignment.

References: Holland (1986), Samii (2016)

Lecture 2: Experimental analysis I

Week 2, August 28

Complete randomization vs. Bernoulli trial.
Horvitz-Thompson estimator vs. Hajek estimator. Neyman variance and its estimation.
From unbiasedness to consistency.
Large sample inference.

References: Ch3, Ch4, and Ch6 of Imbens and Rubin (2015)

Lecture 3: Experimental analysis II

Week 3, September 4

Fisher's randomization test.
Bootstrap and jackknife.
Control for covariates under random assignment.
Lin's regression.
Block randomization.
Clustering experiments.
Clustered standard errors.

References: Ch5, Ch7, and Ch9 of Imbens and Rubin (2015)

Lecture 4: Experimental analysis III

Week 4, September 11

The moderator effect.
Interpret the interactive effect.
Caveats of interactions.
Nonparametric regression.
Kernels.
Estimating the moderator effect with kernel regression.

References: Hainmueller, Mummolo, and Xu (2019)

Lecture 5: From experiments to observational studies

Week 5, September 18

Strong ignorability.
The central role of propensity score. DAG.
Matching.

References: Ch10 and Ch12 of Imbens and Rubin (2015), Ch6 of Hernán and Robins (2010)

Lecture 6: Covariate adjustment

Week 6, September 25

Inverse probability of treatment weighting estimators.
Entropy balancing.
Covariate balancing propensity scores.
Doubly robust estimators.

References: Hainmueller (2012), Imai and Ratkovic (2014), Aronow and Samii (2016)

Section II: Selective topics

Topic 1: Basics of machine learning

Week 7, October 2, Linear methods in machine learning (1 presenter)

References: Ch6 of Gareth et al. (2013)

Week 8, October 9, Tree-based methods in machine learning

References: Ch8 of Gareth et al. (2013), Montgomery and Olivella (2018),

Topic 2: Machine learning and causal inference

Week 9, October 23, Estimating heterogeneous treatment effects

References: Chernozhukov et al. (2017), Athey and Wager (2019)

Week 10, October 30, Partial identification

References: Slough (2023), Knox, Lowe, and Mummolo (2020), Samii, Wang, and Zhou (2022)

Topic 3: Panel data analysis

Week 11, November 6: Caveats of fixed effects models (1 presenter)

References: Strezhnev (2017), Imai and Kim (2019)

Week 12, November 13: Flexible models

References: Liu, Wang, and Xu (2020), Arkhangelsky et al. (2019)

Week 13, November 20: Methods under sequential ignorability

References: Blackwell and Glynn (2018), Ch20 of Hernán and Robins (2010)

Student presentation

Week 14, December 4

References:

Arkhangelsky, Dmitry, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager. 2019. "Synthetic Difference in Differences." National Bureau of Economic Research.

Aronow, Peter M, and Cyrus Samii. 2016. "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science* 60 (1): 250–67.

Athey, Susan, and Stefan Wager. 2019. "Estimating Treatment Effects with Causal Forests: An Application." *Observational Studies* 5 (2): 37–51.

Blackwell, Matthew, and Adam N Glynn. 2018. "How to Make Causal Inferences with Time-Series Cross-Sectional Data Under Selection on Observables." *American Political Science Review* 112 (4): 1067–82.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. 2017. "Double/Debiased/Neyman Machine Learning of Treatment Effects." *American Economic Review* 107 (5): 261–65.

Gareth, James, Witten Daniela, Hastie Trevor, and Tibshirani Robert. 2013. *An Introduction to Statistical Learning: With Applications in r*. Spinger.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis*, 25–46.

Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. 2019. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice." *Political Analysis* 27 (2): 163–92.

Hernán, Miguel A, and James M Robins. 2010. "Causal Inference." CRC Boca Raton, FL.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60.

Imai, Kosuke, and In Song Kim. 2019. "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" *American Journal of Political Science* 63 (2): 467–90.

Imai, Kosuke, and Marc Ratkovic. 2014. "Covariate Balancing Propensity Score." *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 243–63.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Knox, Dean, Will Lowe, and Jonathan Mummolo. 2020. "Administrative Records Mask Racially Biased Policing." *American Political Science Review* 114 (3): 619–37.

Liu, Licheng, Ye Wang, and Yiqing Xu. 2020. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." *Available at SSRN 3555463*.

Montgomery, Jacob M, and Santiago Olivella. 2018. "Tree-Based Models for Political Science Data." *American Journal of Political Science* 62 (3): 729–44.

Samii, Cyrus. 2016. "Causal Empiricism in Quantitative Research." *The Journal of Politics* 78 (3): 941–55.

Samii, Cyrus, Ye Wang, and Junlong Aaron Zhou. 2022. "Generalizing Covariate-Tightened Trimming Bounds for Sample Selection Using Adaptive Kernels."

Slough, Tara. 2023. "Phantom Counterfactuals." *American Journal of Political Science* 67 (1): 137–53.

Strezhnev, Anton. 2017. "Generalized Difference-in-Differences Estimands and Synthetic Controls." *Unpublished Manuscript*.