Topics in causal inference

Dmitry Arkhangelsky
d.i.arkhangelsky@gmail.com

March, 2018

Course Description

Causal questions are at heart of any economic decision problem. How can we evaluate a marketing campaign? How efficient was a job training program? What is the economic effect of a new law? What is the best hospital admission policy? Answers to these and many other questions require careful use of experimental and observational data.

Current theoretical literature uses familiar concepts from causal inference and insights from econometrics, statistics, computer science, and operations research. This combination produced a new set of techniques that can be applied to familiar empirical problems. The goal of this course is to introduce some of these methods in a user-friendly way. The principal emphasis will be on algorithms themselves and the reasons why (and when) we should use them instead of other, more standard techniques.

I want to stress that this is not an introductory applied course, there will be no time to discuss basic econometric/statistic/probability concepts. Instead we will focus on current (or recent) **theoretical** research. At the same time, some of these methods are already widely used in companies like Google, Netflix, Facebook, and Amazon, so if you are focused on a career in a data-driven company, this knowledge can be particularly useful. If you are thinking about an academic career then knowing how to apply these techniques can be quite advantageous (and, at a certain point in the future, even essential).

Prerequisites

You should read Imbens and Wooldridge [2009] and understand it (this requires certain maturity in econometrics and statistics). The methods that we will discuss are built on the results described in the paper. Home assignments will require modest programming skills (writing shorts scripts in R).

Course Objectives

By the end of the course, I expect students to have a good understanding of the current theoretical literature (small subset), and the ways they can use the new methods in their own applied work (both academic or industry related). A large part of the reading list consists of the current working papers and recently published articles. The technical side of these papers can be quite demanding, and it is not expected that the students will be able to understand all the details of the papers fully. Instead, the focus will be on motivation and practical implementation.

Course Structure

There will be ten lectures, three home assignments, and one final presentation. Home assignments will mainly ask you to replicate some of the analysis done in the paper (e.g., a Monte-Carlo simulation). The final presentation will be focused either on theoretical papers that were not discussed during the lectures or on the empirical applications. All assignments can (and should) be done in groups. I will decide on the exact rules for the groups during the first week of the class.

Preliminary

- New ways to deal with covariates: machine learning and double robustness (Chernozhukov et al. [2016], Li et al. [2014]).
- Modern approach to a regression discontinuity analysis (Imbens and Wager [2017]).
- Alternative methods for panel data analysis: synthetic control, matrix completion (Abadie et al. [2010], Brodersen et al. [2015], Athey et al. [2017]);
- From treatment effects to decisions: efficient policy learning (Athey and Wager [2017]);
- Combining different data sources: role of the surrogates (Athey et al. [2016])

References

A. Abadie, A. Diamond, and J. Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505, 2010.

S. Athey and S. Wager. Efficient policy learning. *arXiv preprint arXiv:1702.02896*, 2017.

S. Athey, R. Chetty, G. Imbens, and H. Kang. Estimating treatment effects using multiple surrogates: The role of the surrogate score and the surrogate index. *arXiv preprint arXiv:1603.09326*, 2016.

S. Athey, M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi. Matrix completion methods for causal panel data models. *arXiv preprint arXiv:1710.10251*, 2017.

K. H. Brodersen, F. Gallusser, J. Koehler, N. Remy, S. L. Scott, et al. Inferring causal impact using bayesian structural time-series models. *The Annals of Applied Statistics*, 9 (1):247–274, 2015.

V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, et al. Double machine learning for treatment and causal parameters. *arXiv preprint arXiv:1608.00060*, 2016.

G. Imbens and S. Wager. Optimized regression discontinuity designs. *arXiv preprint arXiv:1705.01677*, 2017.

G. Imbens and J. Wooldridge. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86, 2009.

F. Li, K. L. Morgan, and A. M. Zaslavsky. Balancing covariates via propensity score weighting. *arXiv preprint arXiv:1404.1785*, 2014.