Topics in Causal Inference

DRP Final Presentation

Omkar A. Katta

April 30, 2020

Outline

CORRELATION IS NOT CAUSATION.

Causation

- Outcome Y
- Possible Cause X
- $\Delta X \to \Delta Y$?
- *ceteris paribus* - all else equal
- The changes in Y can only be attributed to the differences in X.

Rubin Causal Model

- Treatment Indicator: $D_i = \mathbb{1}(i \text{ is treated})$
- $Y_i(1)$ is outcome if i is treated.
- $Y_i(0)$ is outcome if i is untreated.
- We want: $\mathbb{E}(Y_i(1) - Y_i(0))$ *ceteris paribus*
- Fundamental Problem of Causal Inference

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0) = Y_i(D_i).$$

Outline

What is the causal impact of a treatment on the outcome?

- Randomization
- Differences-in-Differences
  - Find a *counterfactual*.
  - First difference: washes out systematic differences.
  - Second difference: average causal effect.
- Regression Discontinuity Design
  - An *exogeneous* shock
  - E.g., merit-based scholarship, SAT cutoffs
- Instrumental Variables, Structural Models, Propensity Score Matching, ...

What could go wrong?

- We cannot randomize.
- Selection Bias
- Omitted Variable Bias
- Simultaneity
- Violation of Stable Unit Treatment Value Assumption (*SUTVA*)
- Quantification of Uncertainty

Outline

Frontier

- Is randomization all that great?
- Dynamic Potential Outcomes Model
- Non-parametric and Semi-parametric designs
- Machine Learning, Network Theory
  1. unsupervised learning: heterogeneous treatment effects
  2. prediction techniques: synthetic control
  3. Big Data: finite population uncertainty
  4. networks model interference effects: relax SUTVA and adopt *NIA*
  5. model-driven vs. data-driven ; standard errors and statistical properties

Outline

References

Abadie, Alberto and Javier Gardeazabal (2003). "The economic costs of conflict: A case study of the Basque Country". In: *American economic review* 93.1, pp. 113–132.

Abadie, Alberto et al. (2014). *Finite population causal standard errors*. Tech. rep. National Bureau of Economic Research.

Athey, S and GW Imbens (2019). "Machine Learning Methods Economists Should Know About". In: *arXiv preprint arXiv: 1903.10075*.

Athey, Susan (2015). "Machine learning and causal inference for policy evaluation". In: *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 5–6.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). "How much should we trust differences-in-differences estimates?" In: *The Quarterly journal of economics* 119.1, pp. 249–275.

References II

Chalfin, Aaron et al. (2016). "Productivity and selection of human capital with machine learning". In: *American Economic Review* 106.5, pp. 124–27.

Chandler, Dana, Steven D Levitt, and John A List (2011). "Predicting and preventing shootings among at-risk youth". In: *American Economic Review* 101.3, pp. 288–92.

Einav, Liran and Jonathan Levin (2014). "Economics in the age of big data". In: *Science* 346.6210, p. 1243089.

Jagadeesan, Ravi, Natesh Pillai, and Alexander Volfovsky (2017). "Designs for estimating the treatment effect in networks with interference". In: *arXiv preprint arXiv:1705.08524*.

Mullainathan, Sendhil and Jann Spiess (2017). "Machine learning: an applied econometric approach". In: *Journal of Economic Perspectives* 31.2, pp. 87–106.

References III

Sussman, Daniel L and Edoardo M Airoldi (2017). "Elements of estimation theory for causal effects in the presence of network interference". In: *arXiv preprint arXiv:1702.03578*.

Torgovitsky, Alexander (2019). "Nonparametric inference on state dependence in unemployment". In: *Econometrica* 87.5, pp. 1475–1505.