Field Experiments, Machine Learning, and Causality
Spring 2024
David Blei and Donald Green

Draft of January 22, 2024

Description

This course explores the challenges of extracting inferences about cause and effect in policy-relevant domains. Specifically, we will study modern field experiments and how machine learning methods can help design and analyze them. Focusing on real-world designs and real-world data, we will evaluate the strengths and weaknesses of modeling choices and methods, and we will study how to use ML-based insights to suggest more informative design choices.

Prerequisites and workload

The course is designed for doctoral students in social science, computer science, and statistics. It is also open to masters students and undergraduates with sufficient preparation.

We expect students to have basic knowledge of probability and statistics, and to be comfortable with computer programming for data analysis.

Workload: a small mid-semester project; short weekly reading responses; a final paper.

Syllabus (tentative)

1. What are field experiments? What is machine learning? (Green and Blei)
   Gerber et al. (2004); Blei (2014)
2. Causality : Potential outcomes (Green)
   Imbens and Rubin (2015), Chapter 1; Ding (2023)
3. Causality : Graphical models (Blei)
   Pearl et al. (2016); Freedman (2004)
4. Covariate section (Green)
   Bloniarz et al. (2016); Lin (2013); Lu et al. (2023); Su et al. (2023)
5. Synthetic controls (Blei)
   Abadie et al. (2010); Abadie (2021); Athey et al. (2021); Ben-Michael et al. (2021); Bottmer et al. (2023); Shi et al. (2022)
6. Benchmarking (Green)
   Arceneaux et al. (2006); DellaVigna and Pope (2018); Gentzel et al. (2021); Gordon et al. (2019); Keith et al. (2023)
7. Causality and text data (Blei)
   Egami et al. (2022); Fong and Grimmer (2023); Sridhar and Blei (2022); Veitch et al. (2020)
8. Heterogenous treatment effects (Green)
   Chernozhukov et al. (2018); Ding et al. (2016)

9. Bayesian optimization and sequential experimental design (Blei)
   Garnett (2023)
10. Factorial design (Green)
    Egami and Imai (2018); Goplerud et al. (2022); Zhao and Ding (2022)
11. Multiple environments and invariance (Blei)
    Peters et al. (2016); Arjovsky et al. (2019)
12. Multiple outcomes (Green)
    Arnold and Ercumen (2016); Shi et al. (2020); Vickerstaff et al. (2021)
13. Instrumental variables (Green)
    Danieli et al. (2022); Hartford et al. (2017)

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425.

Abadie, A., Diamond, A., and Hainmueller, J. (2010). 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.

Arceneaux, K., Gerber, A. S., and Green, D. P. (2006). Comparing experimental and matching methods using a large-scale voter mobilization experiment. *Political Analysis*, 14:37–62.

Arjovsky, M., Bottou, L., Gulrajani, I., and Lopez-Paz, D. (2019). Invariant risk minimization. *arXiv:1907.02893*.

Arnold, B. F. and Ercumen, A. (2016). Negative control outcomes: A tool to detect bias in randomized trials. *JAMA*, 316(24):2597–2598.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). Matrix completion methods for causal panel data models. *arXiv:1710.10251*.

Ben-Michael, E., Feller, A., and Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, 116(536):1789–1803.

Blei, D. (2014). Build, compute, critique, repeat: Data analysis with latent variable models. *Annual Review of Statistics and Its Application*, 1:203–232.

Bloniarz, A., Liu, H., Zhang, C.-H., Sekhon, J. S., and Yu, B. (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proceedings of the National Academy of Sciences*, 113(27):7383–7390.

Bottmer, L., Imbens, G. W., Spiess, J., and Warnick, M. (2023). A design-based perspective on synthetic control methods. *Journal of Business & Economic Statistics*, pages 1–12.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernandez-Val, I. (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. Technical report, National Bureau of Economic Research.

Danieli, O., Nevo, D., Walk, I., Weinstein, B., and Zeltzer, D. (2022). Negative controls for instrumental variable designs.

DellaVigna, S. and Pope, D. (2018). Predicting experimental results: Who knows what? *Journal of Political Economy*, 126(6):2410–2456.

Ding, P. (2023). *A First Course in Causal Inference*.

Ding, P., Feller, A., and Miratrix, L. (2016). Randomization inference for treatment effect variation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(3):655–671.

Egami, N., Fong, C. J., Grimmer, J., Roberts, M. E., and Stewart, B. M. (2022). How to make causal inferences using texts. *Science Advances*, 8.

Egami, N. and Imai, K. (2018). Causal interaction in factorial experiments: Application to conjoint analysis. *Journal of the American Statistical Association*, 113(522):222–233.

Fong, C. and Grimmer, J. (2023). Causal inference with latent treatments. *American Journal of Political Science*, 67:374–389.

Freedman, D. A. (2004). Graphical models for causation, and the identification problem. *Evaluation Review*, 28(4):267–293.

Garnett, R. (2023). *Bayesian Optimization*. Cambridge University Press.

Gentzel, A., Pruthi, P., and Jensen, D. (2021). How and why to use experimental data to evaluate methods for observational causal inference. *arXiv:2010.03051*.

Gerber, A., Green, D., and Kaplan, E. (2004). The illusion of learning from observational research. In Shapiro, I., Smith, R., and Massoud, T., editors, *Problem and Methods in the Study of Politics*. Cambridge University Press.

Goplerud, M., Imai, K., and Pashley, N. E. (2022). Estimating heterogeneous causal effects of high-dimensional treatments: Application to conjoint analysis. *arXiv:2201.01357*.

Gordon, B. R., Zettelmeyer, F., Bhargava, N., and Chapsky, D. (2019). A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. *Marketing Science*, 38(2):193–225.

Hartford, J., Lewis, G., Leyton-Brown, K., and Taddy, M. (2017). Deep IV: A flexible approach for counterfactual prediction. In *International Conference on Machine Learning*, volume 70, pages 1414–1423.

Imbens, G. and Rubin, D. (2015). *Causal Inference in Statistics, Social and Biomedical Sciences: An Introduction*. Cambridge University Press.

Keith, K. A., Feldman, S., Jurgens, D., Bragg, J., and Bhattacharya, R. (2023). RCT rejection sampling for causal estimation evaluation. *arXiv:2307.15176*.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. pages 295–318.

Lu, X., Yang, F., and Wang, Y. (2023). Debiased regression adjustment in completely randomized experiments with moderately high-dimensional covariates.

Pearl, J., Glymour, M., and Jewell, N. (2016). *Causal Inference in Statistics: A Primer*. John Wiley & Sons.

Peters, J., Bühlmann, P., and Meinshausen, N. (2016). Causal inference by using invariant prediction: Identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):947–1012.

Shi, C., Sridhar, D., Misra, V., and Blei, D. (2022). On the assumptions of synthetic control methods. In *Artificial Intelligence and Statistics*.

Shi, X., Miao, W., and Tchetgen Tchetgen, E. (2020). A selective review of negative control methods in epidemiology. *Current Epidemiology Reports*, 7:190–202.

Sridhar, D. and Blei, D. (2022). Causal inference from text: A commentary. *Science Advances*, 8(42).

Su, F., Mou, W., Ding, P., and Wainwright, M. J. (2023). A decorrelation method for general regression adjustment in randomized experiments.

Veitch, V., Sridhar, D., and Blei, D. (2020). Adapting text embeddings for causal inference. In *Uncertainty in Artificial Intelligence*.

Vickerstaff, V., Ambler, G., and Omar, R. Z. (2021). A comparison of methods for analysing multiple outcome measures in randomised controlled trials using a simulation study. *Biometrical Journal*, 63(3):599–615.

Zhao, A. and Ding, P. (2022). Regression-based causal inference with factorial experiments: estimands, model specifications and design-based properties. *Biometrika*, 109(3):799–815.