Causal Inference in Small-scale Studies

Instructor: Nick Doudchenko

If you want to assess the causal effect of an intervention, it's hard to beat a randomized experiment with 100,000 units. (Actually, perhaps, not that hard—just take a randomized experiment with 1 million units…) Unfortunately, such experiments are not always feasible or practical. Often, you need to estimate causal effects from observational data with the intervention affecting a single large geographic unit such as, for example, a country. Even when you do have the luxury of designing the experiment, you may have to (or wish to) select only a small number of units for treatment.

How do we estimate causal effects in such settings? If we can decide which units receive the treatment, how should we choose those? How should we conduct statistical inference? In this short course we will discuss some modern—as well as not so modern—approaches to these problems.

Prerequisites: Some familiarity with linear algebra and statistics will certainly be helpful and to some degree required, but the content of the course isn't going to be too technical. The lab portion of this course will contain exercises in Python and, possibly, another lower-level language. Being able to implement simple functions performing typical statistical tasks in a language such as Python or R will be very beneficial.

Instructor's bio: I (Nick Doudchenko) am a Software Engineer at Google Research in New York. Most of my current work is focused on experimental design and estimation of causal effects in the context of online marketplaces. I am particularly interested in combining statistical ideas with techniques from machine learning and combinatorial optimization. Before joining Google I worked as an Economist at Facebook which was my first job after completing a PhD in Economics from Stanford GSB in 2018 where I was advised by Guido Imbens and Lanier Benkard. Prior to that I received an MA in Economics from New Economic School and a BS in Mathematics from Moscow State University.

Syllabus

We will have five 4-hour sessions split equally between the "theoretical" part during which we will discuss the methods and the "practical" part in the lab during which we will implement some of those methods.

Session 1. "Observational studies: Difference-in-differences and synthetic controls."
Difference-in-differences and, more recently, synthetic controls are some of the most popular tools in observational causal inference. We will discuss both of these approaches in the common setting.

Literature:
Card, David. *"The impact of the Mariel boatlift on the Miami labor market."* (1990).
Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. *"Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program."* (2010).
Doudchenko, Nick, and Guido W. Imbens. *"Balancing, regression, difference-in-differences and synthetic control methods: A synthesis."* (2016).

Session 2. "Observational studies: Beyond synthetic controls."
There are a great many ways to conduct a synthetic-control-like study, but how exactly should you do that and what if your setting does not perfectly match the traditional synthetic-control setting? We will look at a couple of related models that go a bit beyond the standard setting.

Literature:
Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. *"Synthetic difference in differences."* (2019).
Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. *"Matrix completion methods for causal panel data models."* (2021).

Session 3. "Observational studies: Inference in small scale studies."
Inference in synthetic-control settings is always tricky. We will look at a couple of most common approaches as well as discuss the assumptions that justify these inference procedures.

Literature:
Firpo, Sergio, and Vitor Possebom. *"Synthetic control method: Inference, sensitivity analysis and confidence sets."* (2018).
Bottmer, Lea, Guido Imbens, Jann Spiess, and Merrill Warnick. *"A design-based perspective on synthetic control methods."* (2021).
Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu. *"An exact and robust conformal inference method for counterfactual and synthetic controls."* (2021).

Session 4. "Experiments: Randomized vs. optimized"
Randomized trials are extremely common and for good reasons—they are easy to motivate and design, and they are robust to many specifics of the data-generation process. We will discuss the alternative approaches to designing experiments, their benefits and drawbacks.

Literature:
Kasy, Maximilian. *"Why experimenters might not always want to randomize, and what they could do instead."* (2016).
Banerjee, Abhijit V., Sylvain Chassang, Sergio Montero, and Erik Snowberg. *"A theory of experimenters: Robustness, randomization, and balance."* (2020).

Session 5. “Experiments: Designing Synthetic-control Studies.”

Even when randomization is available, a simple “A/B test” is not always the most practical solution. “Synthetic control experiments” are becoming increasingly common. We will discuss a couple of approaches introduced in this emerging area.

Literature:

Doudchenko, Nick, David Gilinson, Sean Taylor, and Nils Wernerfelt. *"Designing experiments with synthetic controls."* (2019).

Doudchenko, Nick, Khashayar Khosravi, Jean Pouget-Abadie, Sebastien Lahaie, Miles Lubin, Vahab Mirrokni, and Jann Spiess. *"Synthetic Design: An Optimization Approach to Experimental Design with Synthetic Controls."* (2021).

Abadie, Alberto, and Jinglong Zhao. *"Synthetic controls for experimental design."* (2021).