Applied Causality Reading Assignment 1

Feihan Lu

January 29, 2017

This reading response summarizes Chapter 2 of [1]. Several important messages conveyed in this chapter include (1) defining causal states and the outcomes caused by these states; (2) defining treatment effects based on the observed outcomes and potential outcomes; (3) estimating the treatment effects defined in (2). I will break down the logic of the entire chapter into the following 3 steps, along with 2 key assumptions.

1 From causal states to observed and potential outcomes

In order to answer the question of how one thing, A, causes another thing, B, using observational data, an intuitive way is to first construct a framework where A can be split into several states, which are mutually exclusive and result in different outcomes of B, so that comparing these outcomes gives rise to the causal effect of A on B. Ideally, each unit in the population or sample can experience all states at the same time, so all outcomes can be observed. However, in reality, only one state is experienced by a certain unit at a certain moment, and so only one outcome can be observed, under the same physical and philosophical condition. The other outcomes are unobserved, and they are called the "potential outcomes". Comparing the observed outcomes with the latent potential outcomes in order to estimate the causal effects is the so-called "counter-factual framework".

2 From observed and potential outcomes to definition of treatment effects

Now that the framework has been constructed, the next step would be to mathematically derive the desired causal effects. Suppose one wants to analyze the causal effect of a random variable D on another random variable Y in the population, where D has two states $D = 1$ and $D = 0$, and Y has two corresponding outcomes Y^1 and Y^0. Under the counter-factual framework, the population-level causal effect is the random variable $\delta = Y^1 - Y^0$ There are three major summary effects of interest: (1). Average Treatment Effects, or ATE = $E[\delta]$; (2). Average Treatment Effects for Treatment group, or ATT = $E[\delta|D = 1]$; (3). Average Treatment Effects for Control group, or ATC = $E[\delta|D = 0]$.
Assumption 1: Note that, these effects are defined under the stable-unit-treatment-value-assumption or SUTVA, meaning that for each unit, i, the outcomes, y_i^1 and y_i^0, are not functions of D for other units, and neither is the treatment effect δ_i. This assumption greatly simplifies the estimation of the treatment effects. A lot of analyses, however, fall outside SUTVA, such as many in the network analysis where people influence each other.

3 From definition of treatment effects to estimation of them

Now that the treatment effects of interest are defined, the final step would be to estimate them using the random sample at hand. The chapter listed several ways to estimate different treatment effects, including (1). Naive estimator of ATE, which is the difference of sample means of the observed treatment and control outcomes; (2). Naive estimator for ATE; (3). Naive estimator for ATT; (4) prediction-based method of over-time treatment effects.
Assumption 2: Note that since the "potential outcome" is unobserved, the naive estimators will not be unbiased and consistent without further assuming "ignorability", i.e., independence of Y^1 and D and \ or Y^2 and D.

References

[1] Winship and Morgan *Counterfactuals and Causal Inference*. Cambridge University Press (2014).