Fundamentals of causal inference: part 1

Justin Sheen

July 25, 2024

1 Introduction

I became interested in causal inference after noticing differences between causality in the field of causal inference, and causality in pathogen transmission mechanistic models. I find the definitions and formalities in causal inference, and statistics more generally, satisfying.

One of the most fun things about the history of statistics is how rich it is with interesting anecdotes and case studies. I enjoyed listening to Prof. Susan Ellenberg's recounting of its history.[1] It's interesting to think about modern clinical trials only gaining traction in the mid-twentieth century, and the human complications that arise when putting these designs into practice. For instance, is it ethical to have placebo groups if the treatment has the potential to save the life of the participant? At what points should trials stop if we think the treatment is having a positive effect? When vaccines are used widely, and therefore a rare bad health event coincidentally occurs shortly after a baby is vaccinated, what is the best way to address, with scientific evidence, the concerns of a parent? What was also interesting was hearing about the difference between an active control vs. a placebo control, and how even studies where the treatment has a known effect may fail. By it's very nature of applying abstract theory to real-world situations, and encountering all sorts of complications, the statistical literature is full of interesting human stories.

The following is adapted from module one of Prof. Jason Roy's fantastic online Coursera course on causal inference https://www.coursera.org/learn/crash-course-in-causality.[2] I wanted to review the basics of causal inference for myself. I recall a professor in undergrad saying that one should never underestimate the fundamentals. This is part one of five.

2 Preliminary definitions and assumptions

Suppose we are interested in the causal effect of some treatment Z on some outcome Y. For example, $z = 1$ indicates an individual taking a vaccine against influenza and $Z = 0$ otherwise; Y is the time until the individual is infected. We can notate $Y(Z = 1)$ and $Y(Z = 0)$ as *potential outcomes* given treatment.

A *counterfactual outcome* is the outcome that would have been observed with the treatment had been different. For example, if the treatment of an individual was $Z = 0$, then the counterfactual outcome would be $Y(Z = 1)$ and vice-versa. Before treatment assignment,

any outcome is a potential outcome, but after the study there is an observed outcome and a counterfactual outcome. It is normally assumed that counterfactual outcomes are the same as potential outcomes.

It is important that there is clearly one treatment, and that there are no hidden treatments in the mix. For example, we may be interested in the causal effect of BMI on some health outcome. The problem is that there are many potential ways to achieve a particular BMI value, which can lead to different outcomes. Similarly, some variables are simply immutable, such as race. Because race is not manipulable, we cannot identify a causal effect for it. These causal effects certainly exist, but are not easily identified within the causal inference framework.

A causal effect is identified if there is a difference in $Y(Z = 1)$ and $Y(Z = 0)$. However, for any individual we can only ever observe one of these potential outcomes. This is what is known as *the fundamental problem of causal inference*. We get around the problem by considering population-level causal effects rather than individual-level causal effects, where $Y(Z = 1)$ would be the averaged outcome of all individuals if all individuals were assigned treatment $Z = 1$.

We should note that $E[Y(Z = 1) - Y(Z = 0)] \neq E[Y|Z = 1] - E[Y|Z = 0]$. This is because in reality there may be differences in the population that had $z = 1$ vs. $z = 0$; for example, individuals at a higher risk for flu may be more likely to get a flu shot, so this may differ from what would happen if *all* individuals were either assigned to $Z = 1$ or $Z = 0$. Generally, if we are comparing two different populations, then we are not identifying a causal effect, whereas we would be identifying the causal effect if we compare the same population where the only difference is treatment.

3 Causal inference assumptions

These are:

- Stable Unit Treatment Value Assumption (SUTVA), which allows us to write the potential outcomes for individual i in terms solely of the treatment individual i receives
  - No interference: units do not interfere with each other, where the treatment of one individual does not affect the outcome of another
  - One version of treatment
- Consistency
  - The potential outcome $Y(Z)$ is equal to the observed outcome if the actual treatment received is $Z = z$. In other words $Y = Y(Z)$ if $Z = z$ for all z.
- Ignorability (sometimes referred to as the no unmeasured confounders assumption)
  - Given pre-treatment covariates X, who is treated (treatment assignment) is independent from the potential outcomes (formally: $Y(0), Y(1) \perp\!\!\!\perp Z|X$). For instance, maybe sicker individuals are assigned treatment more often than less sick individuals, and they are also more likely to have a bad outcome. In this case, who

is treated Z is not independent of the potential outcomes (note that if treatment assignment is randomized no matter the sickness of the individual we wouldn't have a problem). We are looking for: within each stratum of X (in this case age) treatment Z is randomly assigned, so that potential outcomes within each stratum are independent of treatment assignment.

- Positivity
  - $P(Z = z|X = x) > 0; \forall z$ and x. In other words, the probability of treatment assignment is positive no matter your covariate values of X.

We can get from the observed outcome to the potential outcome through the following proof: $E[Y|Z = z, X = x]$ involves only observed data. $E[Y|Z = z, X = x] = E[Y(z)|Z = z, X = x]$ by consistency. $= E[Y(z)|X = x]$ by ignorability.

If we want a marginal causal effect we can average over the distribution of X. Suppose X is a single categorical variable, then $E[Y(z)] = \sum_x E[Y|Z = z, X = x]Pr(X = x)$. This process is known as *standardization*, which involves stratifying then averaging. In other words we obtain a treatment effect within each stratum then pool across stratum, weighting by the probability (or size) of each stratum. We can follow this procedure with the real data. As long as there is randomization of treatment assignment within each stratum, we can identify a causal effect.

Standardization can be difficult if there are many X variables needed to achieve ignorability, and if there are many empty cells of combinations of X variables. Because of this difficulty, much work has been done on remedies such as matching, inverse probability of treatment weighting, propensity scores, and instrumental variables.

4 Incident user and active comparator designs

Suppose we were interested in the effect of yoga on blood pressure. If we take a cross-section of five people, we might find that two of them are currently practicing yoga and three are not. However, there may be variable individual histories of whether they ever did yoga in the past, which creates complications if we want to compare a group that does currently practice yoga vs. doesn't.

One way to get around this would be an *incident user* study design, where we restrict the treated population to those newly initiating treatment. In this way, we would ask the slightly different causal question of: for the population of individuals who have not practiced yoga in the past, what is the causal effect of practicing yoga.

One last challenge is that although start of follow-up time of treated individuals is clear, i.e., when they initiate yoga, it is less clear what the start of follow-up time is for the control. In this case, we might want to have an *active comparator* control, such as Zumba, which would also have a clear start of follow-up time. There is normally less confounding when using an active comparator, but the causal question becomes more narrow. Finally, active comparators may not be possible, or even what we are interested, so they don't always have to be used.

References

[1] S. Ellenberg. People & perspectives: Susan ellenberg. https://www.youtube.com/watch?v=aMkO5vezUcs.

[2] J. Roy. A crash course in causality: Inferring causal effects from observational data. https://www.coursera.org/learn/crash-course-in-causality.