Potential Outcomes

Peng Ding, Email: pengdingpku@berkeley.edu

1. Experimentalists' view of causal inference

Rubin (1975)'s classical quote: "No causation without manipulation."

We will define causal effects using the potential outcomes framework. In this framework, we consider an experiment or at least a thought experiment, which has an intervention, a manipulation or a treatment.

Example 1. We are interested in the effect of taking aspirin or not on head ache. The intervention is taking aspirin.

Example 2. We are interested in the effect of participating in a job training program or not on employment and wage. The intervention is participating in a job training program.

Example 3. We are interested in the effect of studying in a small classroom or a large classroom on standardized test scores. The intervention is studying in a small classroom.

Example 4. Gerber et al. (2008) were interested in the effect of different get-out-to-vote (GOTV) messages on the voting behavior. The intervention is different GOTV messages.

Example 5. Pearl (2018) claimed that we could infer the effect of obesity on life span. The intervention is the body mass index (BMI).

However, there are different levels of ambiguity of the interventions above. The meanings of interventions in Examples 1–4 are relatively clear, but the meaning of intervention on BMI in Example 5 is less clear. In particular, we can imagine different versions of BMI reduction: healthier diet, more physical exercise, bariatric surgery, etc. In this class, we view the intervention in Example 5 as ill-defined.

Another ill-defined intervention is race. Racial discrimination is an important issue in labor market, but it is not easy to imagine an experiment on race. Below is an interesting experiment that partially answers the question.

Example 6. Bertrand and Mullainathan (2004) randomly changed the names on the CVs, and compared the callback rates of CVs with African-American- or White-sounding names. For each CV, the intervention is the binary indicator of African-American- or White-sounding name, and the outcome is the binary indicator of callback. So we can summarize the data in a two-by-two table.

Table 1: Bertrand and Mullainathan (2004)'s data

	no callback	callback
African-American	2278	157
White	2200	235

From Table 1, we can compare the risk difference

$$\frac{157}{2278+157} - \frac{235}{2200+235} = 6.45\% - 9.65\% = -3.20\% < 0$$

with p-value from the Fisher exact test much smaller than 0.001.

2. Formal notation of potential outcomes

Experimental units: $i = 1, \ldots, n$

Treatment levels: 1 for treatment and 0 for control

Outcome of interest Y has two versions for each unit i: $Y_i(1)$ and $Y_i(0)$, which are potential outcomes under the hypothetical interventions 1 and 0. The founding father of the Berkeley Statistics Department, Professor Jerzy Neyman, first used this notation in his master thesis written in Polish (Neyman 1923). This notation seems intuitive, but it has some hidden assumptions. Rubin (1980) made the following clarifications on the hidden assumptions:

(1) Unit i's potential outcomes do not depend on other units' treatments. This is the no-interference assumption. It can be violated in infectious diseases or network experiments.

(2) There are no other versions of the treatment. Equivalently, we require that the treatment level be well defined, or have no ambiguity at least for the outcome of interest. It can be violated if the type of cigarettes matters or the type or major of college education matter.

Rubin (1980) called (1) and (2) above the Stable Unit Treatment Value Assumption (SUTVA). Rubin (2005) further called the $n \times 2$ matrix of potential outcomes the Science Table.

Table 2: Science Table

i	$Y_i(1)$	$Y_i(0)$
1	$Y_1(1)$	$Y_1(0)$
2	$Y_2(1)$	$Y_2(0)$
$\vdots$	$\vdots$	$\vdots$
n	$Y_n(1)$	$Y_n(0)$

Causal effects are functions of the Science Table. We focus on the average causal effect (ACE) in this course:

$$\tau = n^{-1}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\} = n^{-1}\sum_{i=1}^{n} Y_i(1) - n^{-1}\sum_{i=1}^{n} Y_i(0).$$

But we can easily extend our discussion to many other estimands.

3. Treatment assignment mechanism

Treatment vector: $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ where Z_i is the binary treatment indicator for unit i

Observed outcome for unit i:

$$\begin{aligned} Y_i &= \begin{cases} Y_i(1), & \text{if } Z_i = 1 \\ Y_i(0), & \text{if } Z_i = 0 \end{cases} && (1) \\ &= Z_i Y_i(1) + (1 - Z_i) Y_i(0) && (2) \\ &= Y_i(0) + Z_i\{Y_i(1) - Y_i(0)\}. && (3) \end{aligned}$$

Equation(1) is the definition of the observed outcome. Equation (2) is equivalent to (1), but Judea Pearl viewed it as the fundamental bridge between the potential outcomes and the observed outcome. Equation (3) highlights the feature that the individual causal effect $\tau_i = Y_i(1) - Y_i(0)$ can be heterogeneous across units.

The treatment assignment mechanism, i.e., the probability distribution of $\boldsymbol{Z}$, plays an important role in inferring causal effects.

```
> ## potential outcomes and treatment assignment
> n  = 500
> Y0 = rnorm(n)
> tau = - 0.5 + Y0
> Y1 = Y0 + tau
>
> ## perfect doctor 1
> Z = (tau >= 0)
> Y = Z*Y1 + (1 - Z)*Y0
> mean(Y[Z==1]) - mean(Y[Z==0])
[1] 2.379509
>
> ## perfect doctor 2
> Z = (rank(tau) >= n/2)
> Y = Z*Y1 + (1 - Z)*Y0
> mean(Y[Z==1]) - mean(Y[Z==0])
[1] 1.936433
>
> ## nearly perfect doctor 1
> Z = (Y0 >= 0)
> Y = Z*Y1 + (1 - Z)*Y0
> mean(Y[Z==1]) - mean(Y[Z==0])
[1] 1.946643
>
> ## nearly perfect doctor 2
> Z = (rank(Y0) >= n/2)
> Y = Z*Y1 + (1 - Z)*Y0
```

```
> mean(Y[Z==1]) - mean(Y[Z==0])
[1] 1.936433
>
> ## clueless doctor: randomized trials
> Z = rbinom(n, 1, 0.5)
> Y = Z*Y1 + (1 - Z)*Y0
> mean(Y[Z==1]) - mean(Y[Z==0])
[1] -0.3506366
```

References

Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American Economic Review*, 94(4):991–1013.

Gerber, A. S., Green, D. P., and Larimer, C. W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, 102(1):33–48.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles (with discussion). section 9 (translated). reprinted ed. *Statistical Science*, 5:465–472.

Pearl, J. (2018). Does obesity shorten life? or is it the soda? on non-manipulable causes. *Journal of Causal Inference*, 6(2).

Rubin, D. B. (1975). Bayesian inference for causality: The importance of randomization. In *The Proceedings of the social statistics section of the American Statistical Association*, volume 233, page 239. American Statistical Association Alexandria, VA.

Rubin, D. B. (1980). Comment on "Randomization analysis of experimental data: the Fisher randomization test" by D. Basu. *Journal of American Statistical Association*, 75:591–593.

Rubin, D. B. (2005). Causal inference using potential outcomes: Degisn, modeling, decisions. *Journal of American Statistical Association*, 100:322–331.