Lecture 12: Causality and Experiments
Modeling Social Data, Spring 2019
Columbia University

April 26, 2019

Notes from bp2471

1 Observational Estimates

Picking up from last week, we know that oftentimes we are trying to compute causal effect from observed estimates, specified as follows:

$$\Delta_{obs} = [\text{(Sick and went to hospital)} - \text{(Sick if stayed home)}] +$$
$$[\text{(Sick if stayed home)} - \text{(Healthy and stayed home)}]$$

We also briefly reviewed how selection bias comes into play here, and how simply taking the difference between observed outcomes produces a biased estimate of effect.

Observed difference = Causal effect – Selection bias

Selection bias is likely negative here, making the observed difference an underestimate of the causal effect

2 Simpson's Paradox

We more thoroughly covered Simpson's paradox this week. Simpson's Paradox is a phenomenon where a trend across a certain grouping of observations "disappears" or reverses when these groups are aggregated. **It is a warning against casual interpretations of causal effect, and forces us to seriously consider the underlying narrative of a causal system when making decisions on data partitioning.**

More thoroughly detailed on Wikipedia the UC Berkeley gender bias study is a well-known example of Simpson's Paradox. Essentially, at an aggregate level, it appeared as through admissions policy in the 1970's at Berkeley discriminated against women.

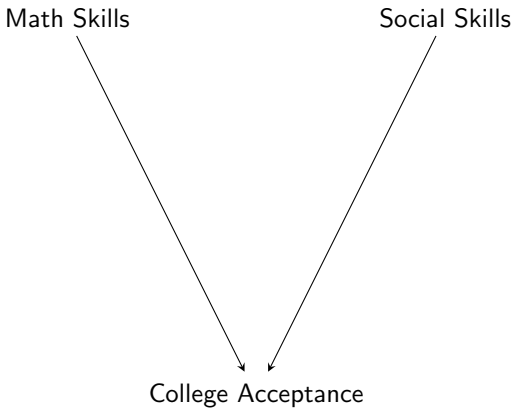
	Men		Women	
	Applicants	Admitted	Applicants	Admitted
Total	8442	**44%**	4321	35%

However, when dis-aggregated by department, the study concluded that women were more likely to apply to competitive departments with low rates of admission even for those qualified, while men tended to apply to less-competitive departments with high rates of admission for those qualified.

3 Selection Bias and Causal Graphs

It is common to represent causal systems by drawing directed graph structures, in which the direction of the edge implies that the origin node of the edge causes the destination node. The figure below shows how we might draw the relationship between college acceptance and a student having math skills and social skills. We would anticipate that having math and social skills (possibly determined via test scores or interviews) would contribute to a student's acceptance to college.

Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
A	***825***	62%	108	**82%**
B	***560***	63%	25	**68%**
C	325	**37%**	***593***	34%
D	417	33%	375	**35%**
E	191	**28%**	***393***	24%
F	373	6%	341	**7%**



Let's think about how selection bias could come into play in a situation like this. Selection bias is introduced by the selection of observations for analysis in a non-random way, resulting in a data sample that is not representative of the population we would like to analyze.

Imagine that we condition our frequency statistics on College Acceptance, ie. we only observe students who have been accepted to college, and we want to investigate the causal effect of math vs. social skills on acceptance. Only observing accepted students, we introduce a kind of selection bias which engenders a relationship between math skills and social skills. This is because in our hypothetical college sample, there are people who are good at math and social skills, people who are bad at math but make up for it in social skills, people who are good at math but bad at social skills, but no people who are bad at both math and social skills because they don't even make it to college. Therefore, in our sample of accepted college students, you will find a negative correlation between math skills and social skills. This is a common problem in causal inference and again highlights how we must take into consideration all the possible relations between events in a system, lest we produce biased estimates of effects.

Note that sometimes selection bias is ok, if were are only interested in the **internal validity** of our estimates—in this example, if we are only interested in accepted students at this specific college our finding of a negative correlation is robust and correct. However, our finding certainly does not meet standards of **external validity**, our findings do not generalize to non-accepted students.

4 Randomization and Experiments

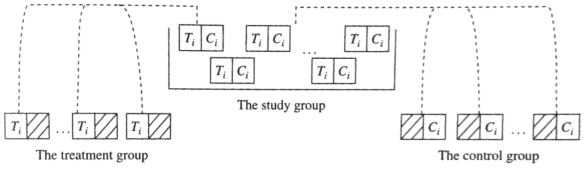
As previously shown, working with observational data is difficult, and requires a lot of hand wringing in order to properly estimate causal effects. In some cases, it is possible for us to actually randomize treatment via experimental design, which allows use to directly compute the effect of treatment. However, there are some caveats to this approach.

1. Randomization often isn't feasible and/or ethical

2. Experiments are costly in terms of time/money
3. It is difficult to create convincing parallel worlds
4. Effects in the lab can differ from the real world
5. **Inevitably people deviate from their random assignments**

There are four possible personas within a treatment assignment framework:

- Compilers: Those individuals who take a treatment when assigned to a treatment groups, and do not take a treatment when assigned to a control group
- Always takers: Those individuals who always take a treatment when assigned to a treatment groups AND when assigned to a control group
- Never takers: Those individuals who never take a treatment when assigned to a treatment groups or when assigned to a control group
- Defiers: Those individuals who will always do the opposite of what will tell them: will take treatment when assigned to control, or will refuse treatment when assigned to treatment (note it is rare to find true defiers in real world experiments)



The Neyman model.
Here, we are drawing at random from a box with N tickets. Each ticket represents one unit in the natural-experimental study group. Here, T_i and C_i are the potential outcomes under treatment and control, respectively. If unit i is sampled into treatment, we observe T_i but not C_i; if unit i is assigned to control, we observe C_i but not T_i. The average of the T_is in the treatment group estimates the average of all the T_is in the box, while the average of the C_is in the control group estimates the average of all the C_is.

Note that the Average Treatment Effect (often our effect of interest) can only be computed from the difference between complying individuals.