Problem set 1

Intro: Causality fundamentals

Please combine all of your answers into a single, beautiful PDF document. Submit the file on Canvas with a link to your notebook in the comment. Make sure to set the "Sharing" settings correctly so the link provides access. The notebook/code is not graded but can help you get partial credit.

I recommend writing the problem set in Dropbox Paper because (1) it is free, (2) it has Latex support, (3) it is easy to copy-paste code and graphics, and (4) it exports to PDF. Google Docs does not have native math support. Jupyter notebooks are ugly and difficult to export to PDF.

1 Notebook problems

Use the corresponding sections of the notebook from lecture 1 to answer these questions: lecture-01-lln-selection-bias.ipynb.

1.1 Law of large numbers (5 points)

1. Where in the code are the parameters μ and σ^2 set?
2. In the code, what objects correspond to (i) $X_1, X_2, ...$, (ii) S_n, and (iii) $\frac{1}{n}S_n$? Why are these objects in math not *exactly* the same as in a computer?
3. Explain intuitively why the line is more stable at the end.

1.2 Selection bias (10 points)

1. Calculate $E[Y^1 - Y^0]$.
2. Calculate the $NATE$.
3. Calculate the selection bias.
4. What line(s) of code fundamentally creates the selection bias? Explain.
5. Calculate $P(D = 1)$

1.3 Data-generating process for a null experiment (6 points)

1. Make a version of this data-generating process that is a true randomized experiment. The new treatment variable should be D_{exp}. Make the overall probability of treatment the same, i.e., $P(D = 1) = P(D_{\text{exp}} = 1)$. Show that the experiment eliminates selection bias. Calculate the $NATE$.

2. Interpret figure 1 "Density of Y0 by treatment group" w.r.t. the *independence assumption*: $(Y^1, Y^0) \perp D$.

3. Make a new version of figure 1 using the experiment data. Compare.

1.4 Data-generating process for an experiment with a treatment effect (4 points)

1. Make a version of the biased data-generating process with an ATE=0.3. Calculate the NATE. (Hint: This only requires changing one line.)

2. Repeat the D_{exp} experiment to show that NATE=ATE=0.3.

2 Causality warm-up (25 points)

Answer all of these questions in your own words. Do *not* re-use any examples from the lecture notes or class. *Do not use the Wikipedia article on selection bias!* You should need a few sentences per per question.

1. What is selection bias? Describe in math and words. Why is it a problem?

2. Describe two reasons selection bias can occur.

3. Describe what it means for random variables to be independent. You can just describe this in words.

4. Why do experiments remove selection bias?

3 In-class study discussion (16 points)

Write up the answers to the class discussion problems. It is OK to reach the same conclusions as your classmates. However, you need to explain the answers in your own words.

4 Critical reading (24 points)

Find a real study about causality that is *not* an experiment. I suggest searching Google News for something like "new study" or 'reddit.com/r/science'. Please don't pick something we already saw in class.

1. Give a link to the article. (It should be accessible. No paywalls please.)

2. What are the units in the study?
3. Briefly describe the study in terms of the potential outcomes Y^1, Y^0, and the treatment D.
4. What is the ATE in the study? If there is no number given just identify how they characterize the treatment effect.
5. How was the treatment (D) assigned?
6. What kind of selection bias do you think might be in the study? How would it affect the results?
7. How would you design an experiment to get an unbiased estimate of the ATE?
8. Why did you choose this study?

5 Selection bias vs. big data (10 points)

In lecture 1 we did in-class exercises on the law of large numbers and selection bias. Now we will combine them. Start with the "charitability" simulation in the notebook. Remember in this scenario we defined the ads as having no effect!

1. Write down the NATE decomposition equation and write the value of each of the four components.
2. Make a function `selection_bias(n)` that calculates the selection bias in just the first n units of data ("sample size of n"). Demonstrate that the law of large numbers applies to selection bias. Make a graph of bias for n ranging from 0 to 10,000. Where is the line converging to? What does is the lesson about selection bias and big data?

Bonus 1: Why is it called identification? (+5 points)

In the lecture we studied the decomposition of the NATE into three parts:

$$\text{NATE} = \text{ATE} + (\text{selection bias}) + (\text{differential treatment effects}).$$

Suppose that we have infinite data on the variables (Y, D) but not (Y^1, Y^0).

The point of this exercise is that our data always tells us (“identifies") the NATE, but not necessarily the ATE. There may not be a one-to-one relationship between the NATE and the ATE. (Hint: We can view the NATE decomposition as an algebra problem to understand when we have a unique solution for the ATE!)

1. Why can we always identify the NATE even if it is not a true randomized experiment?

2. Without observing the potential outcomes (Y^1, Y^0) can we directly estimate the ATE, selection bias, and differential treatment effects?

3. Suppose the independence assumption may not be true. How many unknowns does this equation have? How many solutions does the equation have? *We will consider unknowns to be parameters that we cannot directly estimate from a sample analog.*

4. If the independence assumption holds, which *additional equations* do we get? (Hint: See the course notes for lesson 1 about the "Independence Assumption".)

5. Now how many equations do we have? So with that many equations and the number of unknowns above, how many solutions does the NATE decomposition have?

Bonus 2 (+5 points)

In the lecture notes, we define a version of the average treatment effect called the *average treatment effect on the treated* (ATT). The ATT is the average treatment effect when looking only at the units that were actually treated: $E[Y^1 - Y^0|D = 1]$.

Prove that NATE $=$ ATT $+$ selection bias. This should only take a few lines. You need to use the trick of "adding zero."