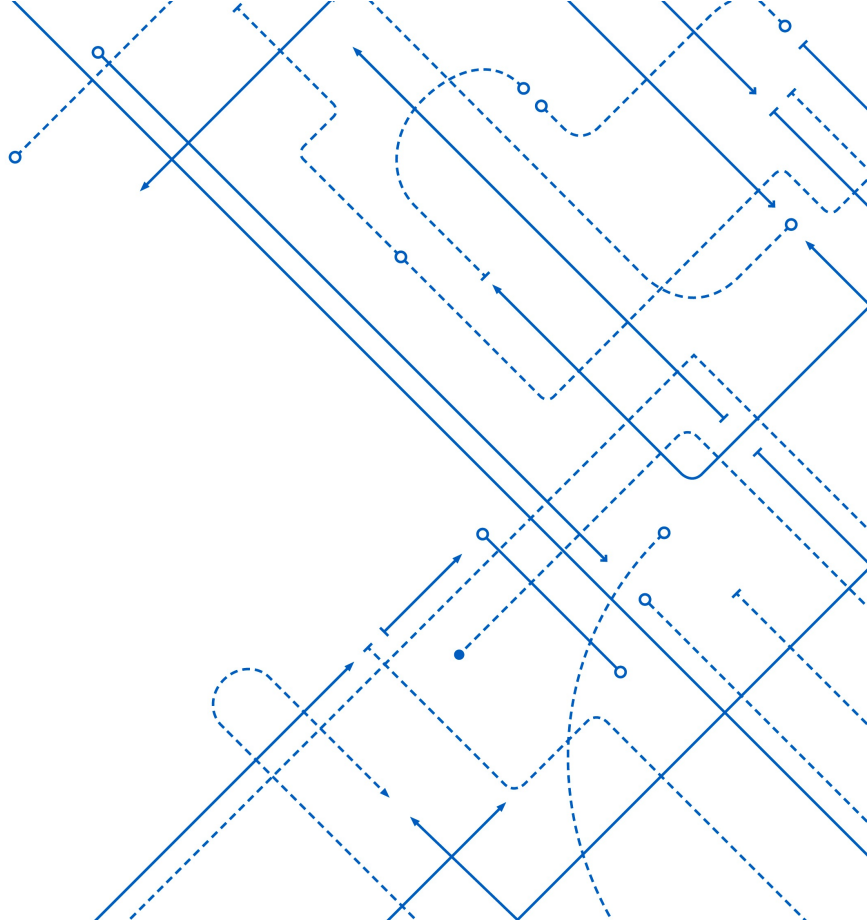
Logistics, Quiz Review, Causal Inference, AMA/Extended OH

Kenneth (Kenny) Joseph

University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences



Announcements

- Quiz 12 due Wednesday night
- PA5 due Friday
- **Any issues grading on anything except PA5 or the Final must be completed before this Friday**
- There's a study guide on Piazza. Don't go beyond it.
- **Thursday – Review Jeopardy!**
  - Same as last time, except more bonus points and fewer questions ☺

Change in Peer Review Req.

- If you also have a group issue, please email me some time before Friday for a discussion
- I will otherwise assume that I already am aware of all group issues

Finals Format

- **Next Tuesday, 7:15PM** (expect the exam should take roughly 1.5 hours, max)
- Exam will be same length, similar format as midterm, although less hand-calculations.. Ish:
  - ~15 MC:
    - Lose points for guessing
  - ~4 Short Answer (think the midterm)
  - Grade will be roughly out of 12 MC, 3 short answer (with max amount of bonus points)
- **Bring nothing but yourself and pen(cil)s**
- **There will be assigned seating, you will learn your assigned seat on 5/16**
- If you talk, I will simply pick up your paper and ask you to leave the room
- Must show your work

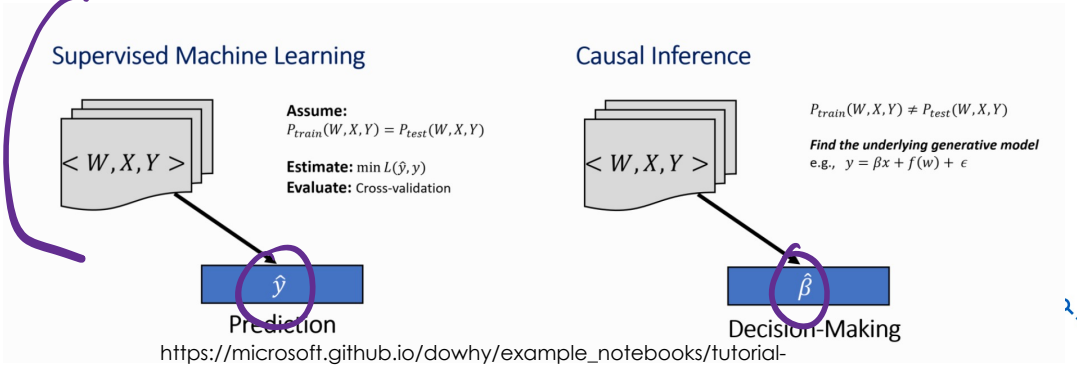
Plan for today

- Review Quiz 10 and Quiz 11
- Causal inference - a whirlwind tour
- Five Things I hope you remember from this course
- Extended OH

Review: Quizzes 10 and 11

Causal Inference

- Why do we want to do causal inference?
  - Did this ad *cause* this person to buy this product?
  - Did this tweet *cause* this person to not get the vaccine?
  - Did this drug *cause* someone to not get cancer?
- Many, many interesting questions are causal inference questions



https://microsoft.github.io/dowhy/example_notebooks/tutorial-causalinference-machinelearning-using-dowhy-econml.html

Why causal inference is hard

I want to know if giving Naveen a study guide will help him ace the final.

Real World: **do(T=1)**

Counterfactual World: **do(T=0)**

I can't both give Naveen the study guide AND not give it to him!
This is the **Fundamental Problem of Causal Inference**

https://microsoft.github.io/dowhy/example_notebooks/tutorial-causalinference-machinelearning-using-dowhy-econml.html

What can we do, then?

1. Target averages/expectations instead of an individualized effect

$$E[Y|do(A=1)] - E[Y|do(A=0)]$$

2. Experiment! (Why does that work?)

Note: do operator indicates an intervention, in an RCT we intervene by randomly assigning treatment and control to comparable groups.

Problems with the simple story

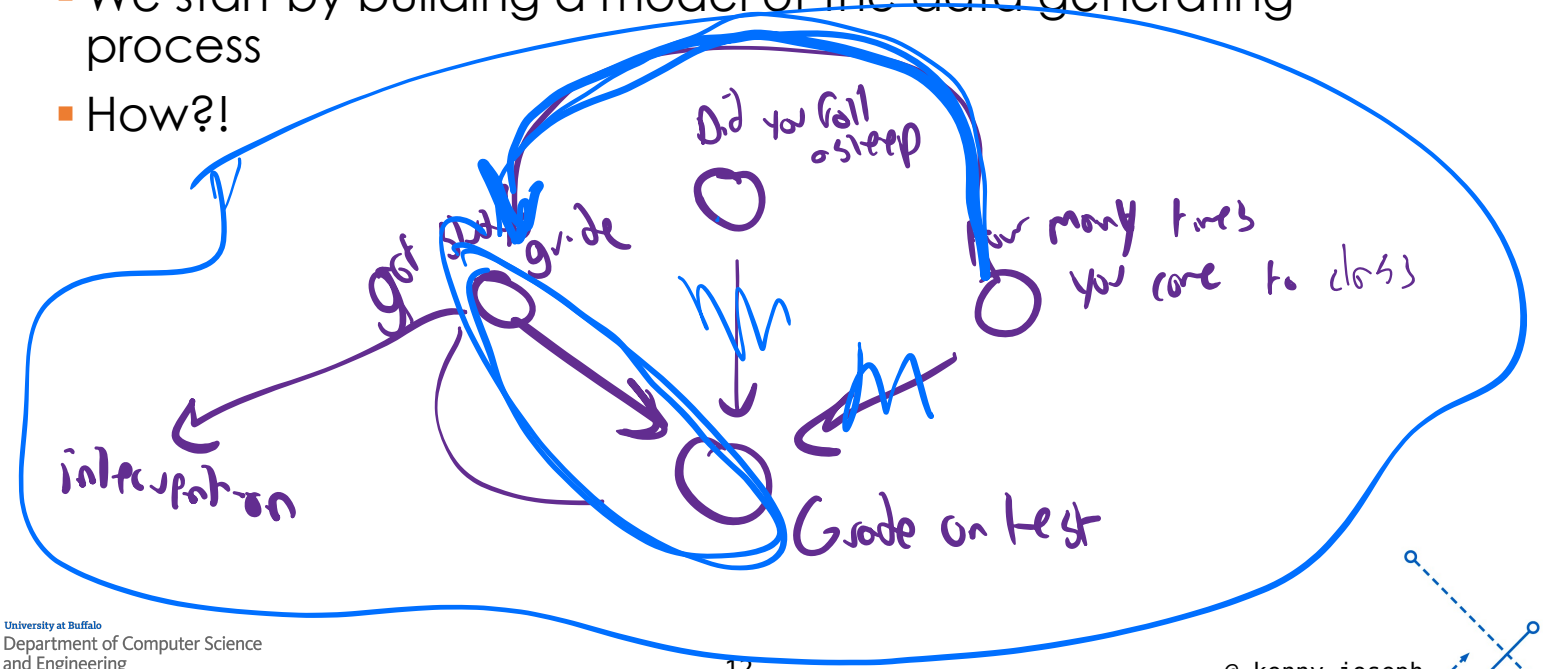
- Sometimes, experimentation is unethical
  - I suspect half of you would be very angry if I gave you a placebo study guide ☺
- Other times, we might have wanted to experiment but simply couldn't, and are left with a bunch of observational data
- Since we do not always have access to experimental data, we rely on observational data for estimating causal effect. Wherein, E[Y | do(A=1)] != E[Y | A=1], as treatment assignment and outcome might be effected by confounding elements, mediator, etc.

Now what?

- Causal inference is a large and growing research field devoted to trying to measure causal effects using observational data
- I am going to show you (at a very high level) **one way to do this**.
- Additional resources on the website!

Step 1: Model

- We start by building a model of the data generating process
- How?!



Step 1: Model

- We start by building a model of the data generating process
- To do so, we can hypothesize the **causal structure of the data using a PGM!**
- Using this PGM tells us what factors, aside from our **treatment**, impact our outcome
- The task of causal inference is then to find a way to estimate, for a given estimand the effect of the treatment on the outcome

Step 1: Model

- Simple model for our example …

What to do?

Find bins!

Step 2: Define an Estimand of interest

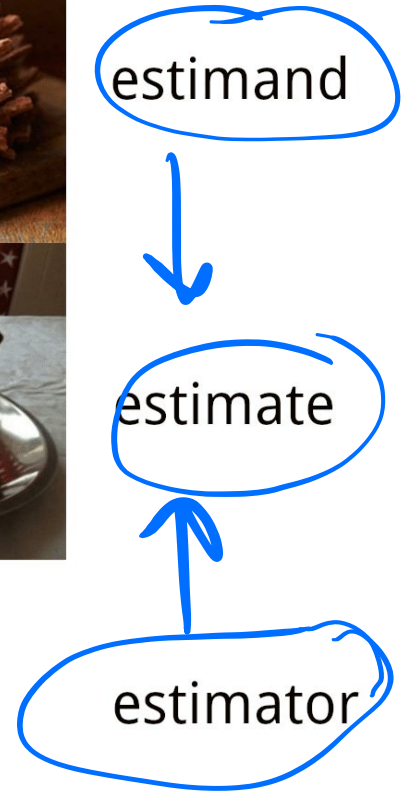
Ingredients

- 150g unsalted butter, plus extra for greasing
- 150g plain chocolate, broken into pieces
- 150g plain flour
- ½ tsp baking powder
- ½ tsp bicarbonate of soda
- 200g light muscovado sugar
- 2 large eggs

Method

1. Heat the oven to 160C/140C fan/gas 3. Grease and base line a 1 litre heatproof glass pudding basin and a 450g loaf tin with baking parchment.

2. Put the butter and chocolate into a saucepan and melt over a low heat, stirring. When the chocolate has all melted remove from the heat.



https://livefreeordichotomize.com/2019/01/17/understanding-propensity-score-weighting/

Average Treatment Effect

The Average Treatment Effect (ATE) is generally the quantity estimated when running a *randomized* study. The target population is the whole population, both treated and controlled. While this is often declared as the population of interest, it is not always the medically or scientifically appropriate population. This is because estimating the ATE assumes that every participant can be switched from their current treatment to the opposite, which doesn't always make sense. For example, it may not be medically appropriate for every participant who didn't receive a treatment to receive it.

https://livefreeordichotomize.com/2019/01/17/understanding-propensity-score-weighting/

Average Treatment Effect Among the Overlap Population

The Average Treatment Effect Among the Overlap Population (ATO) estimates the treatment effect very similar to the ATM, with some improved variance properties. Basically, if you estimated the probability of receiving treatment, the "overlap" population would consist of participants who fall in the middle - you're estimating the treatment effect among those likely to have received either treatment or control. I'll include some graphs in the following sections to help better understand this causal quantity.

https://livefreeordichotomize.com/2019/01/17/understanding-propensity-score-weighting/

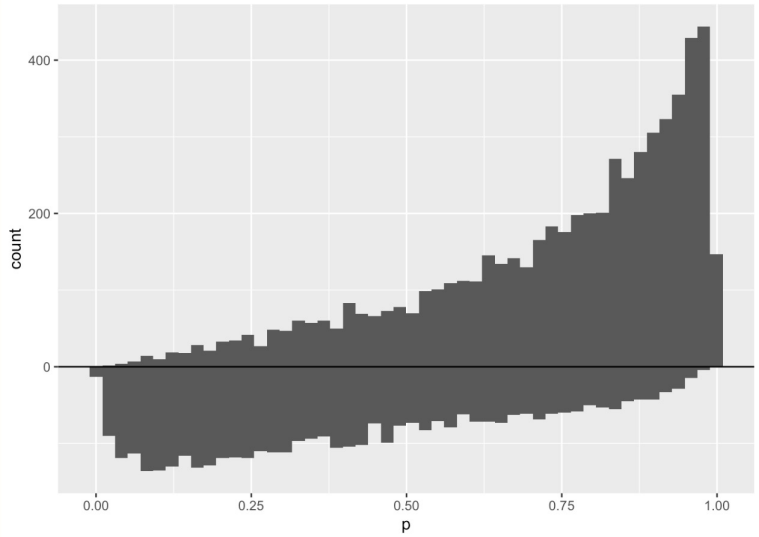
Step 3: Select and estimator & compute an estimate!

- Several common estimators use **propensity scoring**
- Approach to compute propensity scores (very high level):
  - Identify your treatment
  - Identify all confounding variables (informed by structure of your PGM)
    - Build a model to predict whether or not someone was treated from the confounders
- Then you might:
  1. Match on propensity scores
  2. Weight based on propensity scores

Predict who got treatment

Example estimand/estimators using propensity scores

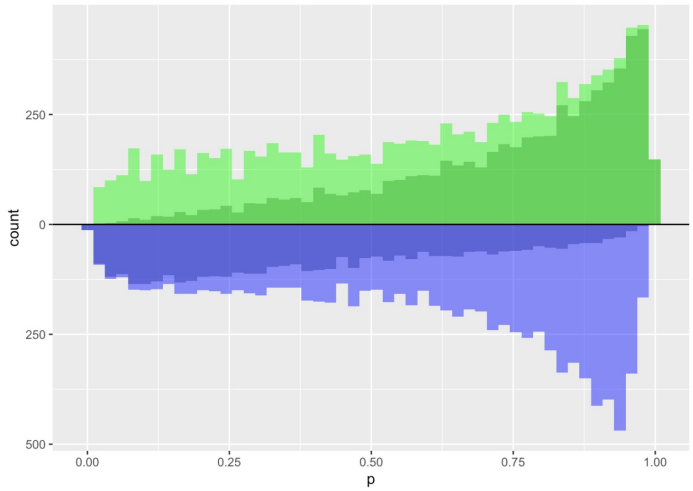
https://livefreeordichotomize.com/2019/01/17/understanding-propensity-score-weighting/



Example estimand/estimators using propensity scores for the ATE

https://livefreeordichotomize.com/2019/01/17/understanding-propensity-score-weighting/

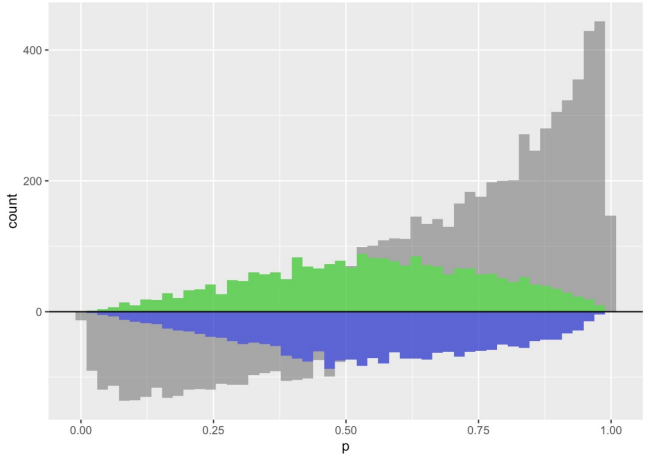
$$w_{ATE} = \frac{Z_i}{e_i} + \frac{1-Z_i}{1-e_i}$$



Example estimand/estimators using propensity scores for the ATO

https://livefreeordichotomize.com/2019/01/17/understanding-propensity-score-weighting/

$$w_{ATO} = (1 - e_i)Z_i + e_i(1 - Z_i)$$



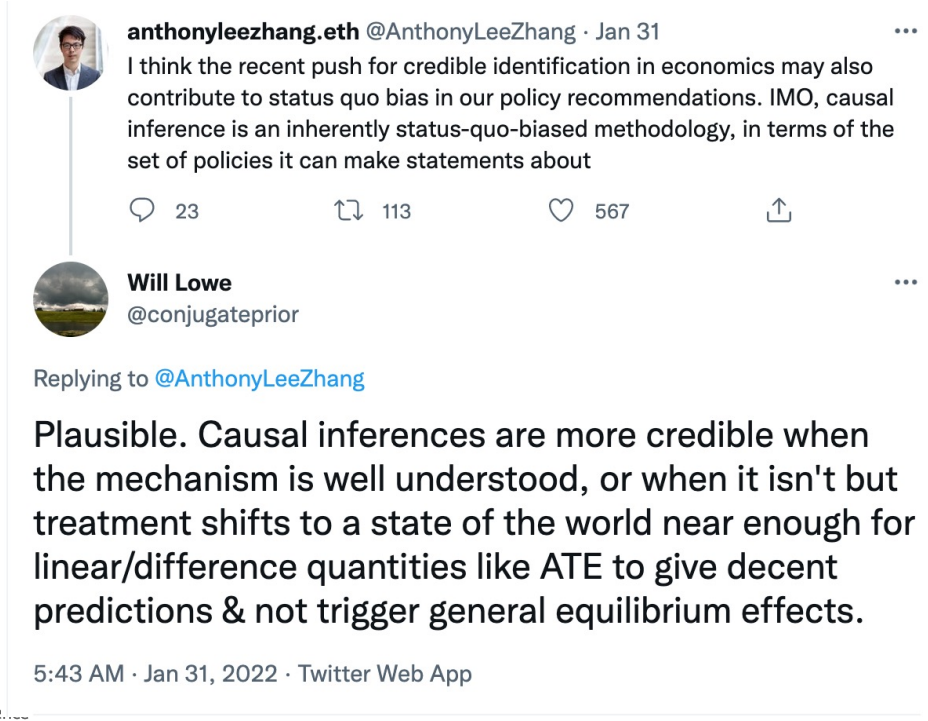
Estimate the treatment effect

$$\frac{\sum Y_i Z_i w_i}{\sum Z_i w_i} - \frac{\sum Y_i (1-Z_i) w_i}{\sum (1-Z_i) w_i}$$

Misc. CI stuff

- There are many, many assumptions baked in here!
- Other areas of work
  - Inferring causal structure (e.g. the DAG)
  - Using ML in various ways (e.g. to do the propensity scoring)
  - …

Beyond Causal Inference



anthonyleezhang.eth @AnthonyLeeZhang · Jan 31

I think the recent push for credible identification in economics may also contribute to status quo bias in our policy recommendations. IMO, causal inference is an inherently status-quo-biased methodology, in terms of the set of policies it can make statements about

23 113 567

Will Lowe
@conjugateprior

Replying to @AnthonyLeeZhang

Plausible. Causal inferences are more credible when the mechanism is well understood, or when it isn't but treatment shifts to a state of the world near enough for linear/difference quantities like ATE to give decent predictions & not trigger general equilibrium effects.

5:43 AM · Jan 31, 2022 · Twitter Web App

Kenny's 5 things he hopes you remember

1. ML is not magic, but it is useful
2. Every ML project has a similar pipeline that it follows
3. ML is just as likely to cause harm as it is to do good
4. Simple methods will generally get you 95% of the way to your goal
5. Strong fundamentals will get you 95% of the way to understanding modern ML

AMA / Open OH