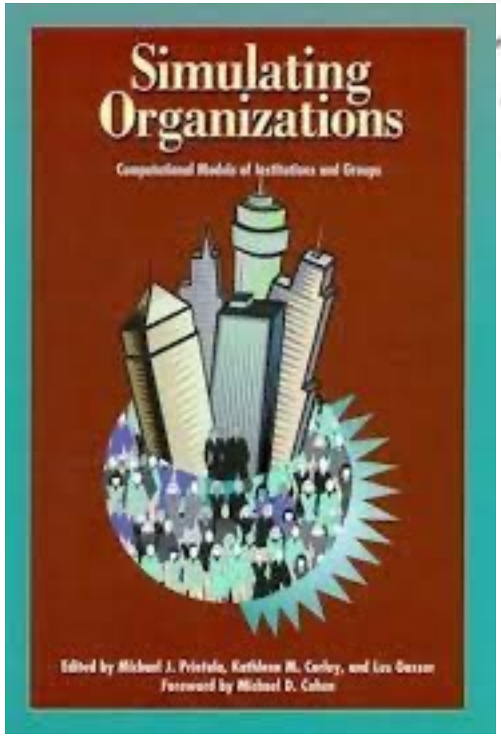
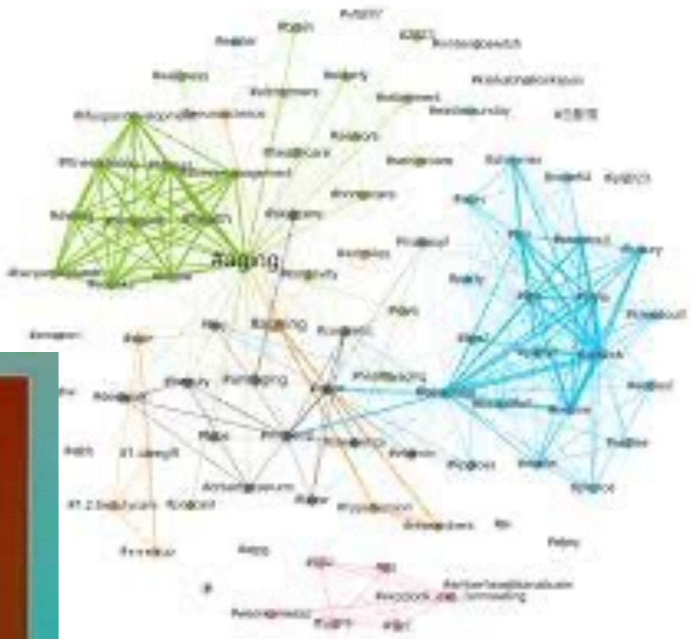
Causal Inference Lecture 3

Kenneth (Kenny) Joseph

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

Passphrase: Kathleen Carley





Overview of our goals for this unit

Three things you need for your project/to do good [computational social] science (IMO)

1. The ability to **understand/reason about** the social world
2. The ability to use that understanding to **design interventions** on the world
3. The ability to **conduct measurements** that assess the [potential] **effect** of that intervention

Tool: causal inference (and causal graphical models)

"Domain": Racial inequality in healthcare

Relation to IP

1 PROBLEM What is the problem you have identified and what contributed to this formation? Of course your high level problem is to end white supremacy but what is the "narrower" problem that you will tackle. Also the domain of the problem that you will choose has to be within the three domains we will cover in our classes: health care, policing and mis-information. You can choose to combine one or more of these. Further, we expect your problem to have the following two parts:

- Tell us about your problem, at a high level (the `zoom out`). How does your problem represent a (hopefully major) impediment in the *CURRENT WORLD* to reaching that vision of the world articulated in *FUTURE WORLD* below?
- Tell us about your problem, at a narrow level (the `zoom in`). What chunk of high level problem from the bullet above did you choose to bite off, and why?

2 CURRENT WORLD Given the problem y'all have identified in the previous question, what is it that has led to and that sustains the problem that you wish to address?

- Also include a discussion of any prior work that has been done on your chosen problem.

3 FUTURE WORLD How will your problem/domain look like in your imagined future world with no white supremacy?

- Specifically, what does **your ideal world with no white supremacy look like?** Be imaginative and bold here. The future does not have to look anything like the current world (in fact if your future world is similar to the current world then you have not gone far enough!)

Relation to IP (cont.)

4 HISTORY+TECH For both *CURRENT WORLD* and *FUTURE WORLD* , your group **must specifically** address **both** of these:

- What are the historical roots?
- How is tech (ideally ML but we can discuss other CSE related tech) involved?

5 PLAN What is your group's proposed plan to go from *CURRENT WORLD* to *FUTURE WORLD* ? Why will it be effective?

- Your plan should start off from the specific `zoomed in` problem above in the *CURRENT WORLD* and end up in the `zoomed out` *FUTURE WORLD* with no white supremacy. What are your first steps in this process? What are the medium-term steps in this plan? What are the long-term steps in this plan? What needs to happen so that each of these steps are successful?

6 TOOLS How do all the Rage tool(s) and (all) the ML+Soc tool(s) that y'all learned in this unit/semester help you address **at least one** of *CURRENT WORLD* , *FUTURE WORLD* or *PLAN* ?

Today

- More details on how to use causal graphical models, causal inference
- Little bit of time to discuss Midpoint Assignment as a class
- Questions?

Reflections on diagram drawing!

- **Two minute** warmup:
  - Get with your project group
  - Gimme **one**
    - Thing that was **hard**
    - Thing drawing it out helped you **think about**
    - Brainless TV show/movie you **all think is really good**

Terminology note

- I'm going to use the following interchangeably:
  - Causal graphical model
  - Directed Acyclic Graph (DAG)
  - Probabilistic graphical model
- They are not the exact same thing. But for our purposes they are close enough (if you want formality, we're mostly talking about causal graphical models)

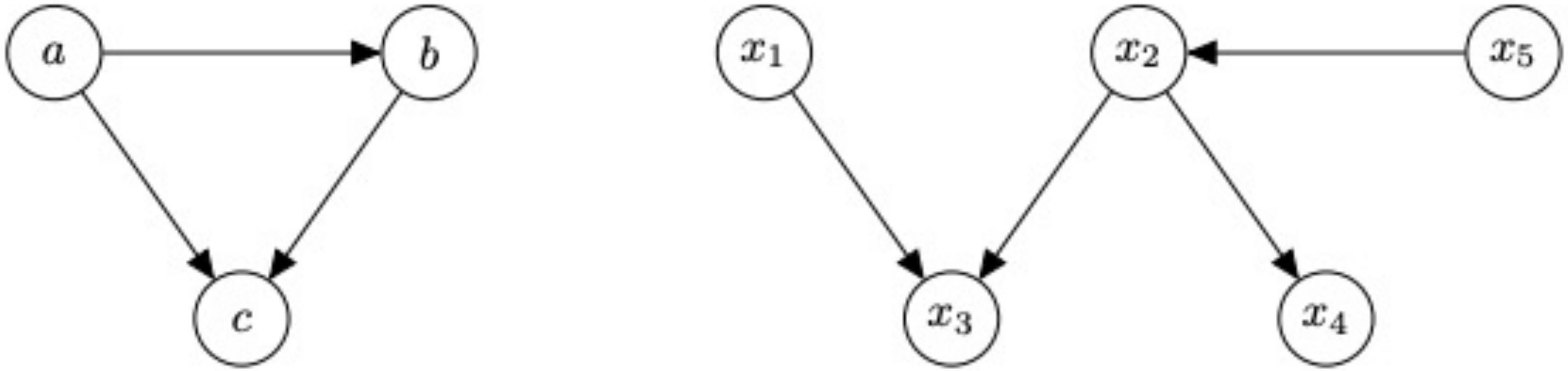
My thoughts/guide to our lecture

- DAGs have limitations!
  - Cycles are real
  - Hard to differentiate certain kinds of probabilistic relationships
- How do you know there's a node? An edge?
  - What counts as evidence?
  - How do we reference evidence?
- Some things are hard to put a number on
  - We have to get suuuper specific to actually run experiments, and **we have to be right**
  - "No causation without intervention"
- We can really use historians...
  - "History" is not an actionable node
- We disagree on both semantics and beliefs about the issues
  - How do we resolve differences between your DAGs?

What can we encode in DAGs

- DAGs can encode valid factorizations of a joint probability distribution
  - ???
- Examples of what DAGs can encode:
  - Causal links
  - Unmeasured confounders
  - “Complex causes”
    - E.g. Moderators
- Examples of what DAGs **can’t** encode
  - When two things cause each other

Practice: Get Distribution from DAG



(a) Fully connected. (b) Not fully connected.

Practice: Draw DAG from Specified Distribution

Draw a Causal Chain, write p()

Draw a moderator, write p()

What does this tell us?

- We can now do a few important things:
  - Formalize what an experiment does for us
  - Think through the **generative story** a particular model tells
  - Think more about why causal models cannot encode cycles (although we can do other things to address this…)

My thoughts/guide to our lecture

- DAGs have limitations!
  - Cycles are real
  - Hard to differentiate certain kinds of probabilistic relationships
- How do you know there's a node? An edge?
  - What counts as evidence?
  - How do we reference evidence?
- Some things are hard to put a number on
  - We have to get suuuper specific to actually run experiments
  - “No causation without intervention”
- We can really use historians...
  - “History” is not an actionable node
- We disagree on both semantics and beliefs about the issues
  - How do we resolve differences between your DAGs?

How do we know?

- **Prior work**

- **Activity:**
  - Try to find an academic article that provides support for the following statement:

[The more severe your illness, the more likely you are to die, regardless of treatment]

How do we know?

- **Prior work**
- Conditional dependencies + Theory
  - ***What is theory?***
  - Sidebar: How do we identify conditional dependencies that are not likely to be random?
  - Sidebar to the sidebar: The "standard" way here is to assume linearity, do linear regression, and look at statistical significance of coefficients. This is *fiiiineeee*.
- Causal Evidence
  - Do an experiment!
  - Do causal inference with causal data

My thoughts/guide to our lecture

- DAGs have limitations!
  - Cycles are real
  - Hard to differentiate certain kinds of probabilistic relationships
- How do you know there's a node? An edge?
  - What counts as evidence?
  - How do we reference evidence?
- Some things are hard to put a number on
  - We have to get suuuper specific to actually run experiments, and we **have to be right**
  - "No causation without intervention"
- We can really use historians...
  - "History" is not an actionable node
- We disagree on both semantics and beliefs about the issues
  - How do we resolve differences between your DAGs?

Putting numbers on things – an example

DIONE CORWIN

8555 PROSACCO SQUARES, CHICAGO, IL

PHONE
+1 (555) 870 1476

EXPERIENCE

BAUMBACH, KOZEY AND NIENOW 04/2018 – present
Philadelphia, PA // *Director of Internal Audit*

- Resides in the Greater Houston Marketplace
- Knowledge of accounting principles, practices, and financial reporting
- Experience working with automated ledger applications
- Strong knowledge of the PCAOB standards, SOX Compliance requirements and COSO's Internal Control Integrated Framework
- Requires strong written and verbal communication skills, analytical and project management skills
- Excellent communication, collaboration and presentation skills
- Requires strong collaboration skills, working closely with management and external audit teams to develop efficient audit plans that address key risks

YUNDT, JACOBI AND SCHINNER 01/2012 – 11/2017
Philadelphia, PA // *Head of Audit*

- Strong experience in the financial services industry, recent experience
- Internal or external audit experience
- To drive the audit framework continuous review and improvements
- Lead the audits in the Digital Bank, focusing on all risks arising from the activities of the Bank
- Plan and perform risk-based assessments, understand and evaluate the business environment, related controls and processes
- Review draft/ final report and participate in the presentation of audit findings

EDUCATION

OTIS COLLEGE OF ART AND DESIGN
Bachelor's in Accounting

SKILLS

- Business acumen including knowledge of the Company's strategy, industry, key risks, operations, and culture
- Outstanding communications skills, including presenting to, advising, and educating executive leadership and Board of Directors
- A quick learner that displays excellent judgment and problem-solving skills which enable timely and appropriately risk balanced advice and guidance
- A team player who displays self-confidence, encourages collaboration, and establishes credibility that earns organizational trust from superiors and peers who can find common ground in solving problems
- Extensive experience with risk-based internal auditing and annual planning

Rachel Corwin

8555 PROSACCO SQUARES, CHICAGO, IL

PHONE
+1 (555) 870 1476

EXPERIENCE

BAUMBACH, KOZEY AND NIENOW 04/2018 – present
Philadelphia, PA // *Director of Internal Audit*

- Resides in the Greater Houston Marketplace
- Knowledge of accounting principles, practices, and financial reporting
- Experience working with automated ledger applications
- Strong knowledge of the PCAOB standards, SOX Compliance requirements and COSO's Internal Control Integrated Framework
- Requires strong written and verbal communication skills, analytical and project management skills
- Excellent communication, collaboration and presentation skills
- Requires strong collaboration skills, working closely with management and external audit teams to develop efficient audit plans that address key risks

YUNDT, JACOBI AND SCHINNER 01/2012 – 11/2017
Philadelphia, PA // *Head of Audit*

- Strong experience in the financial services industry, recent experience
- Internal or external audit experience
- To drive the audit framework continuous review and improvements
- Lead the audits in the Digital Bank, focusing on all risks arising from the activities of the Bank
- Plan and perform risk-based assessments, understand and evaluate the business environment, related controls and processes
- Review draft/ final report and participate in the presentation of audit findings

EDUCATION

OTIS COLLEGE OF ART AND DESIGN
Bachelor's in Accounting

SKILLS

- Business acumen including knowledge of the Company's strategy, industry, key risks, operations, and culture
- Outstanding communications skills, including presenting to, advising, and educating executive leadership and Board of Directors
- A quick learner that displays excellent judgment and problem-solving skills which enable timely and appropriately risk balanced advice and guidance
- A team player who displays self-confidence, encourages collaboration, and establishes credibility that earns organizational trust from superiors and peers who can find common ground in solving problems
- Extensive experience with risk-based internal auditing and annual planning

Draw the causal story here

Sidebar: This isn't enough!

- We have to actually specify our assumptions about the probability distributions as well. We won't get there in this class, though.

Threats to validity

How Black Are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies

S. Michael Gaddis

University of California, Los Angeles

Abstract: Online correspondence audit studies have emerged as the primary method to examine racial discrimination. Although audits use distinctive names to signal race, few studies scientifically examine data regarding the perception of race from names. Different names treated as black or white may be perceived in heterogeneous ways. I conduct a survey experiment that asks respondents to identify the race they associate with a series of names. I alter the first names given to each respondent and inclusion of last names. Names more commonly given by highly educated black mothers (e.g., Jalen and Nia) are less likely to be perceived as black than names given by less educated black mothers (e.g., DaShawn and Tanisha). The results suggest that a large body of social science evidence on racial discrimination operates under a misguided assumption that all black names are alike, and the findings from correspondence audits are likely sensitive to name selection.

Keywords: racial discrimination; inequality; names; audit studies; experiments

What happens when we can't experiment?

- A basic propensity-scoring approach
  - Model
  - Define estimand
  - Estimate
  - Check
- A nod to more complex approaches

Step 1: Model

- We start by building a model of the data generating process
- How?!
- Once we've done that, 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 2: Define an Estimand of interest

estimand

estimate

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.

estimator

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.

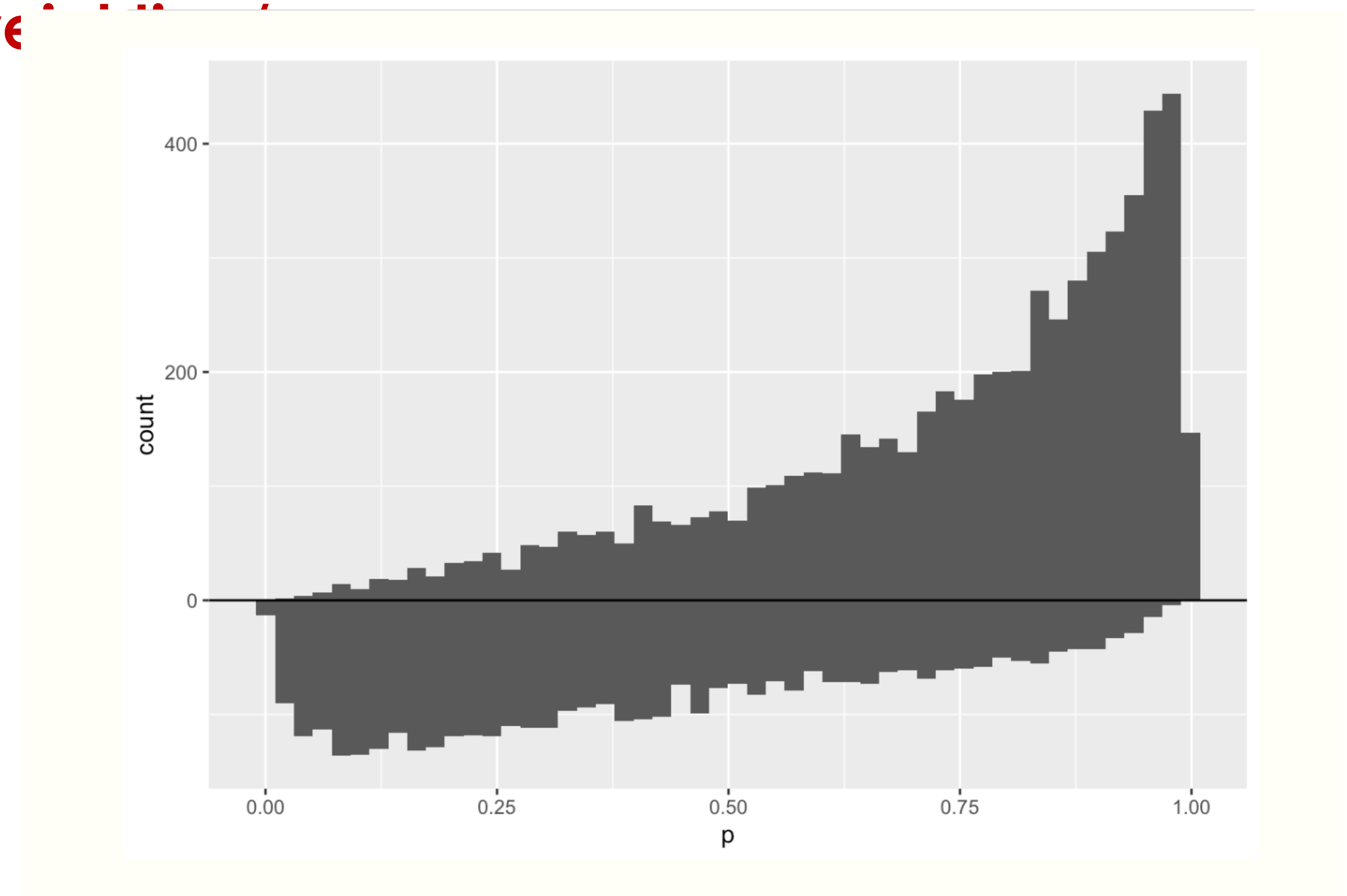
https://livefreeordichotomize.com/2019/01/17/understanding-propensity-score-weighting/

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

Example estimand/estimators using propensity scores

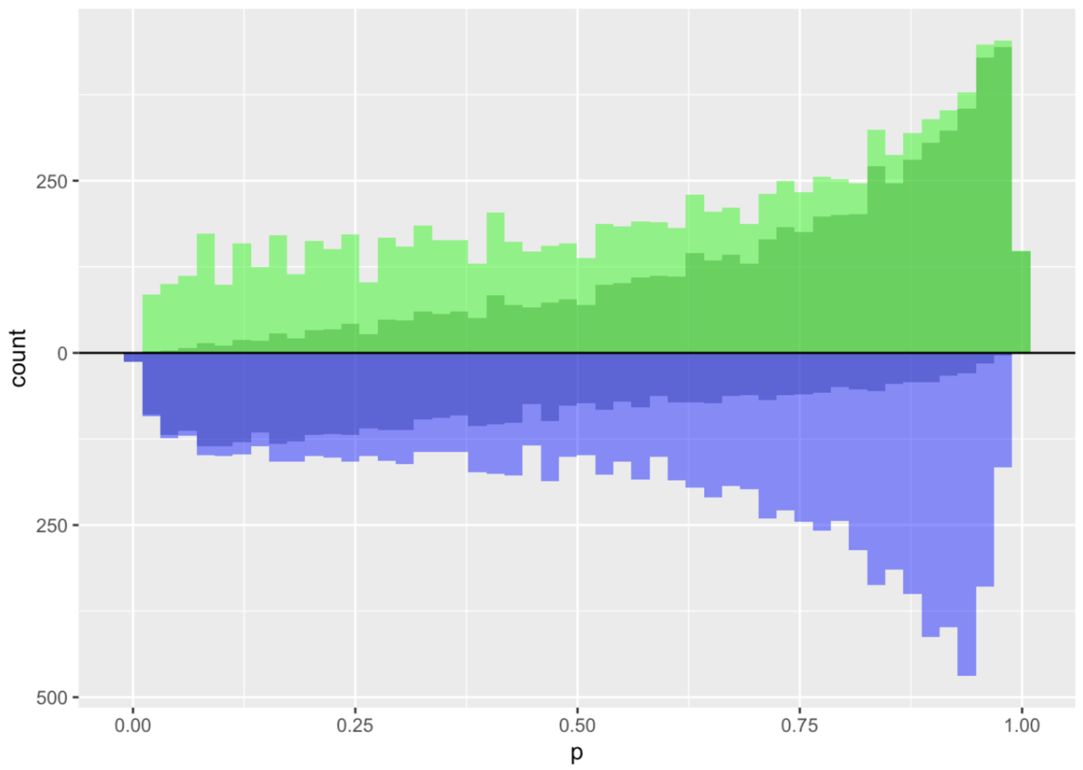
https://livefreeordichotomize.com/posts/2019-01-17-understanding-propensity-score-we



Example estimand/estimators using propensity scores for the ATE

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

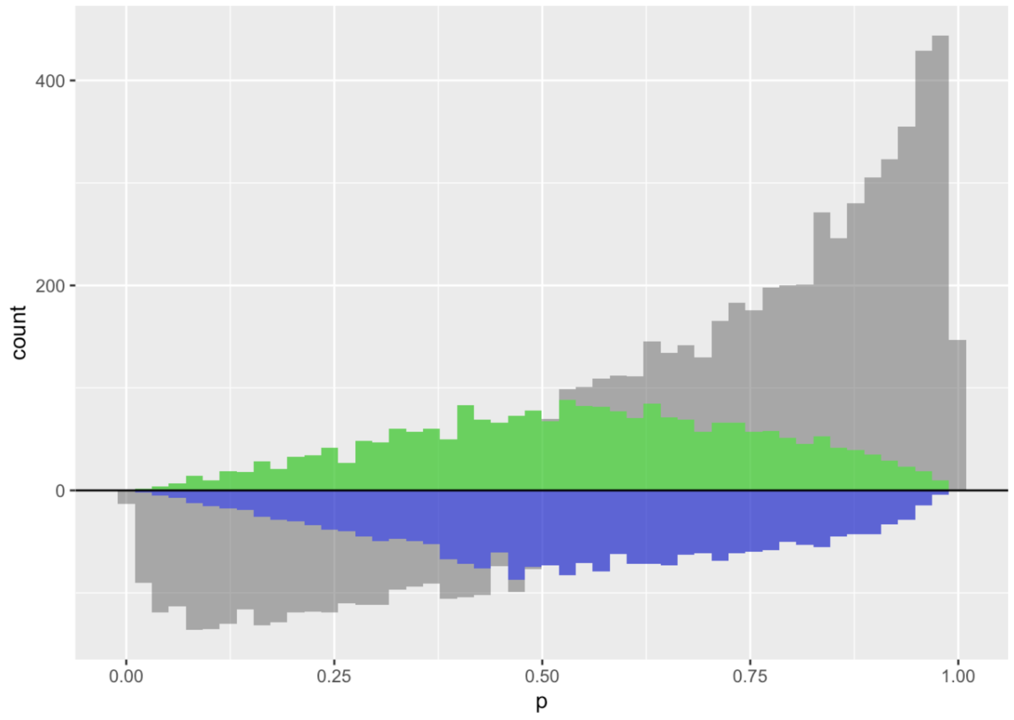
$$w_{ATE} = \frac{Z_i}{e_i} + \frac{1-Z_i}{1-e_i}$$



Example estimand/estimators using propensity scores for the ATO

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

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



@_kenny_joseph

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 (to where?)

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

Where's the ML in this?

Double Machine Learning for Causal Inference from a Partially Linear Model

Double/Debiased machine learning can be used to recover causal effects even if relationships between confounders and the treatment, and between confounders and the outcome are nonlinear.

Zachary Clement · Follow
6 min read · Dec 31, 2022

My thoughts/guide to our lecture

- DAGs have limitations!
  - Cycles are real
  - Hard to differentiate certain kinds of probabilistic relationships
- How do you know there's a node? An edge?
  - What counts as evidence?
  - How do we reference evidence?
- Some things are hard to put a number on
  - We have to get suuuper specific to actually run experiments
  - "No causation without intervention"
- We can really use historians...
  - "History" is not an actionable node
- We disagree on both semantics and beliefs about the issues
  - How do we resolve differences between your DAGs?