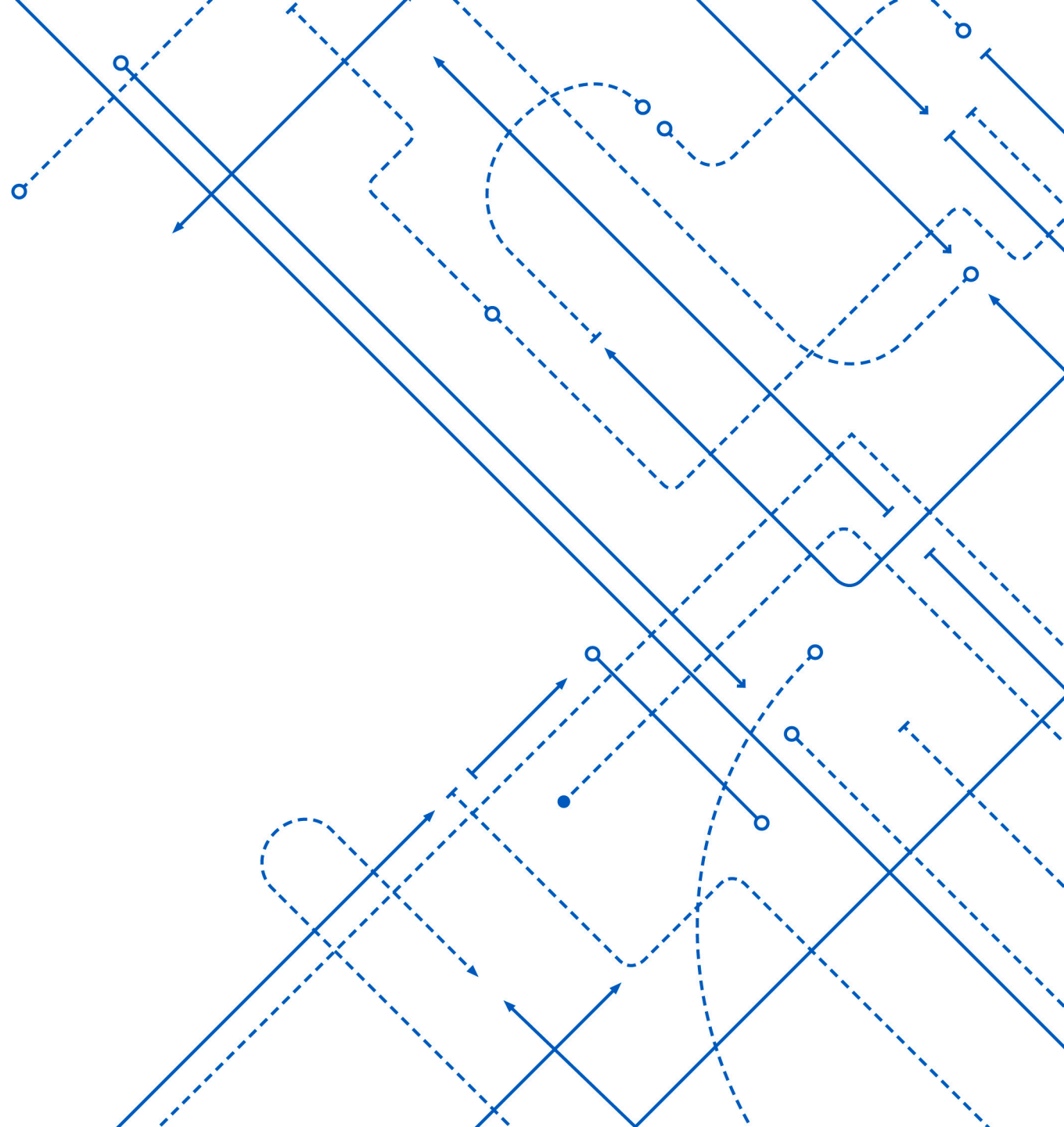


Causal Inference (kind of)

Kenneth (Kenny) Joseph



Pass phrase: **Daphne Koller**

Daphne Koller

🗨️ 27 languages ▾

Article [Talk](#)

[Read](#) [Edit](#) [View history](#) [Tools](#) ▾

From Wikipedia, the free encyclopedia

Daphne Koller (**Hebrew**: דפנה קולר; born August 27, 1968) is an **Israeli-American** computer scientist. She was a professor in the department of **computer science** at **Stanford University**^[4] and a **MacArthur Foundation** fellowship recipient.^[1] She is one of the founders of **Coursera**, an online education platform. Her general research area is **artificial intelligence**^{[5][6]} and its applications in the **biomedical sciences**.^[7] Koller was featured in a 2004 article by *MIT Technology Review* titled "10 Emerging Technologies That Will Change Your World"^[8] concerning the topic of **Bayesian machine learning**.^{[9][10]}

Education [\[edit \]](#)

Koller received a bachelor's degree from the **Hebrew University of Jerusalem** in 1985, at the age of 17, and a master's degree from the same institution in 1986, at the age of 18.^[11] She completed her **PhD** at Stanford in 1993 under the supervision of **Joseph Halpern**.^[2]

Career and research [\[edit \]](#)

After her PhD, Koller did **postdoctoral research** at **University of California, Berkeley** from 1993 to 1995 under **Stuart J. Russell**,^[12] and joined the faculty of the Stanford University computer science department in 1995. She was named a **MacArthur Fellow** in 2004. She was elected a member of the **National Academy of Engineering** in 2011 for contributions to representation, inference, and learning in probabilistic models with applications to robotics, vision, and biology. She was also elected a fellow of the **American Academy of Arts and Sciences** in 2014.

In April 2008, Koller was awarded the first ever \$150,000 **ACM-Infosys** Foundation Award in Computing Sciences.^[13]

She and **Andrew Ng**, a fellow Stanford computer science professor in the AI lab, founded **Coursera** in 2012. She served as the co-CEO with Ng, and then as president of Coursera. She was recognized for her contributions to online education by being named one of *Newsweek's* 10 Most Important People in 2010, *Time* magazine's 100 Most Influential People in 2012, and *Fast Company's* Most Creative People in 2014.^[14]

She left Coursera in 2016 to become chief computing officer at **Calico**.^[14] In 2018, she left Calico to start and lead **Insitro**, a **drug discovery** startup.^[15]

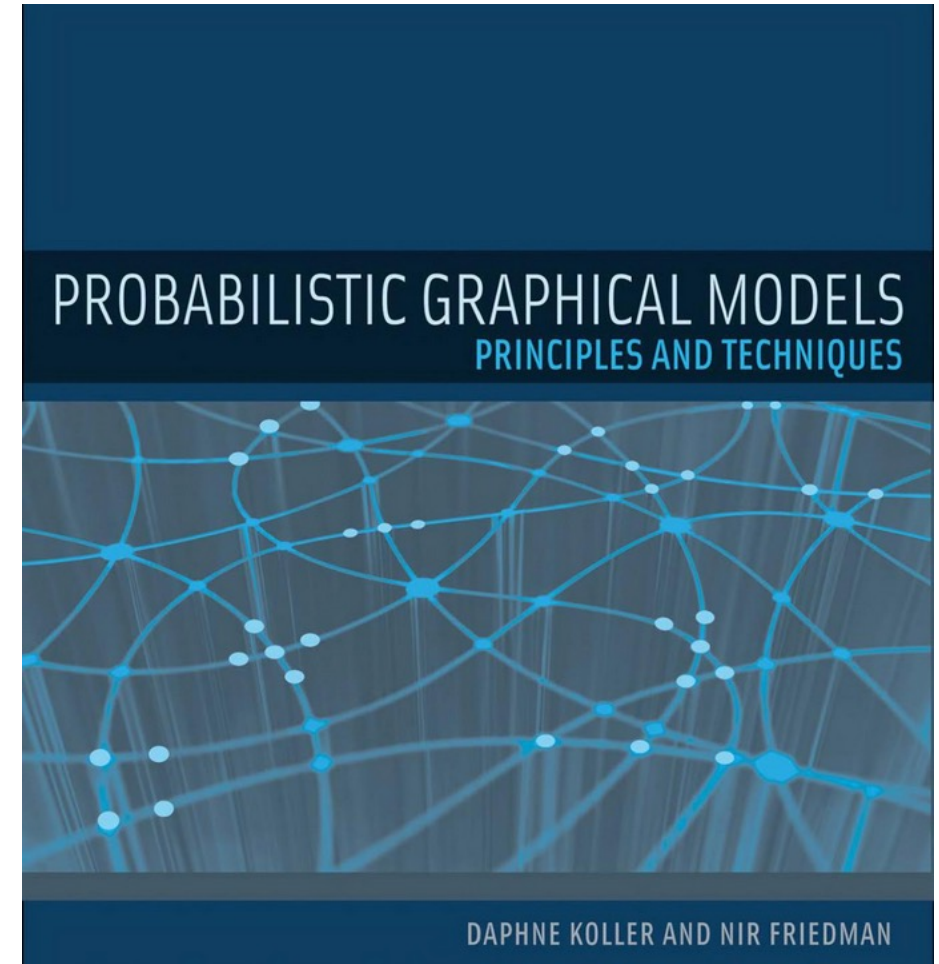
Koller is primarily interested in representation, inference, learning, and decision making, with a focus on applications to **computer vision** and **computational biology**.^[16] Along with **Suchi Saria** and Anna Penn of Stanford University, Koller developed PhysiScore, which uses various data elements to predict whether premature babies are likely to have health issues.^[17]

Daphne Koller



Koller in 2019

Born	August 27, 1968 (age 54) Israel ^{[<i>citation needed</i>]}
Education	Hebrew University of Jerusalem (BSc, MSc) Stanford University (PhD)
Known for	Machine learning Graphical models MOOCs Coursera
Awards	ISCB Fellow (2017) IJCAI Computers and Thought Award (2001) MacArthur Fellow (2004) PECASE (1999) ACM Prize in Computing (2007)
Fields	Scientific career Machine learning Computational biology Computer vision



Announcements

- Putting in the effort in your group
- Wednesday is now a combined class
- No progress summary due, but...
 - Still meet with your groups
 - Leverage office hours
 - Prep for 2nd presentation

What is the goal of machine learning?

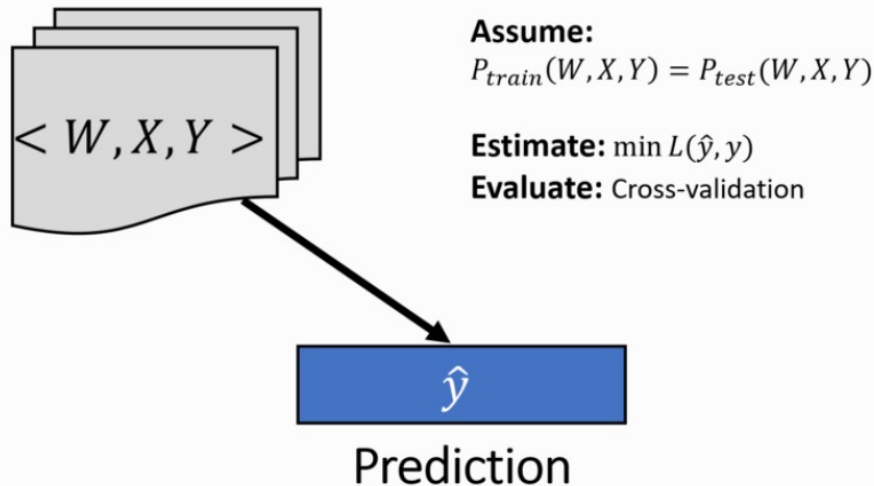
What is the goal of machine learning?

- Can ML answer these questions?
 - 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?
- Why?

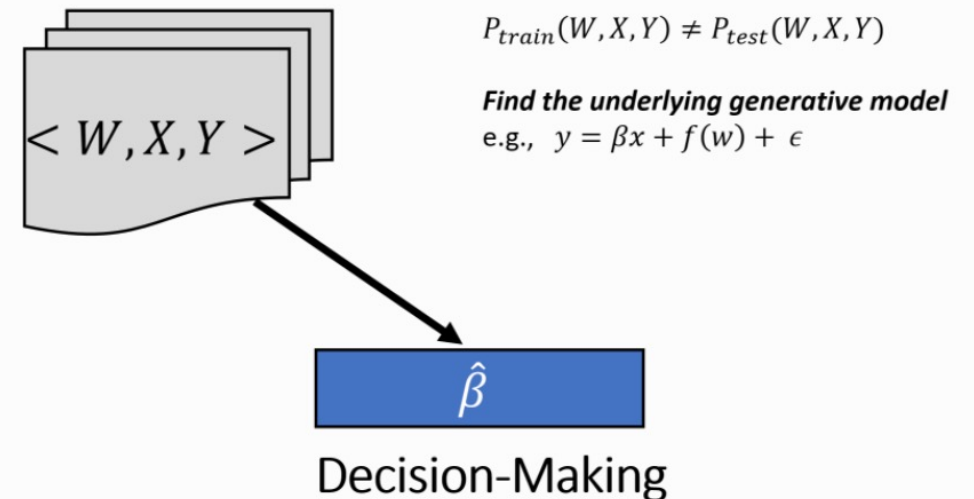
Causal Inference

- Many, many interesting questions are causal inference questions

Supervised Machine Learning



Causal Inference



https://microsoft.github.io/dowhy/example_notebooks/tutorial-causal-inference-machine-learning-using-dowhy-econml.html

Why causal inference is hard

I want to know if telling Vedang he is going to get an A no matter what will make his project better or worse



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



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

I can't both give Vedang an A AND not give it to him!
This is the **Fundamental Problem of Causal Inference**

https://microsoft.github.io/dowhy/example_notebooks/tutorial-causal-inference-machine-learning-using-dowhy-econml.html

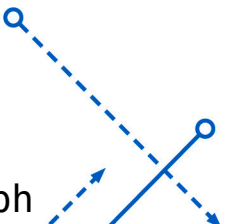
What can we do, then?

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.

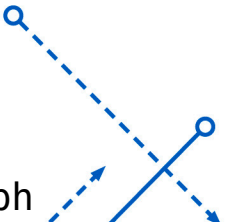


Problems with the simple story

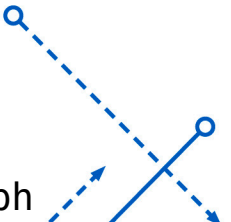
- Sometimes, experimentation is unethical
 - Give half of you As and the other half we don't
- Other times,
 - we might have wanted to experiment but simply couldn't, and are left with a bunch of observational data
 - Even when we experiment, we mess it up
- Since we do not always have access to experimental data, we rely on observational data for estimating causal effect.

How could you mess up an experiment?

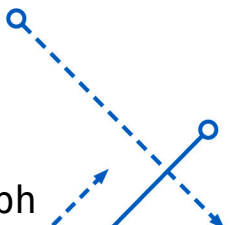
A digression... probabilistic graphical models



Practice: Get Distribution from BayesNet

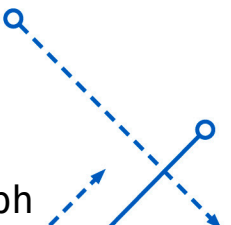


Practice: Draw Bayes Net from Specified Distribution



Summary: PGMs

- PGMs are drawings that represent probability distributions
- These probability distributions represent **generative models, i.e. stories about the (social) processes you are trying to model**
- E.g. ChatGPT's generative story is “the probability of the next word is defined by the previous words” ... although it is much more complex than that
- OK, we'll come back to this...



Problems with the simple story

- Sometimes, experimentation is unethical
 - Give half of you As and the other half we don't
- Other times,
 - we might have wanted to experiment but simply couldn't, and are left with a bunch of observational data
 - Even when we experiment, we mess it up
- Since we do not always have access to experimental data, we rely on observational data for estimating causal effect.

Now what?

- What might we do in these settings?

Now what?

- What might we do in these settings?
 - Create a fake world where we can intervene (simulation)
 - Try to estimate causal effects in observational data by controlling for confounders
 - Many ways to do this, two high level approaches:
 - **Propensity Scoring:** Try to make the people who didn't get the treatment look like the people who did
 - **Conditional Outcome Modeling:** Try to build a predictive model of the outcome and "ask" it what happens if you change the causal feature from 1 to 0

Today

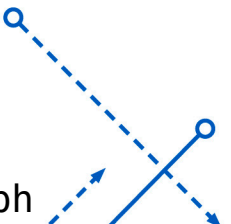
- A basic propensity-scoring approach
 - Model
 - Define estimand
 - Estimate
 - Check
- A nod to more complex approaches

Step 1: Model

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

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 **probabilistic graphical model (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

- What is the effect of being Black on getting a job interview once your resume is submitted?

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.

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

estimator

Average Treatment Effect

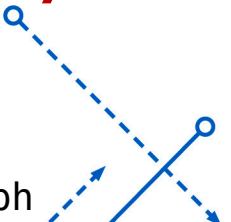
The Average Treatment Effect (ATE) is generally the quantity estimated when running a *randomized* study. The target population is the whole population, both treated and controlled. While this is often declared as the population of interest, it is not always the medically or scientifically appropriate 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/>

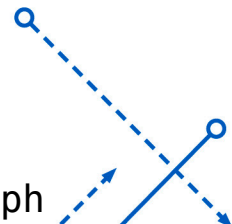
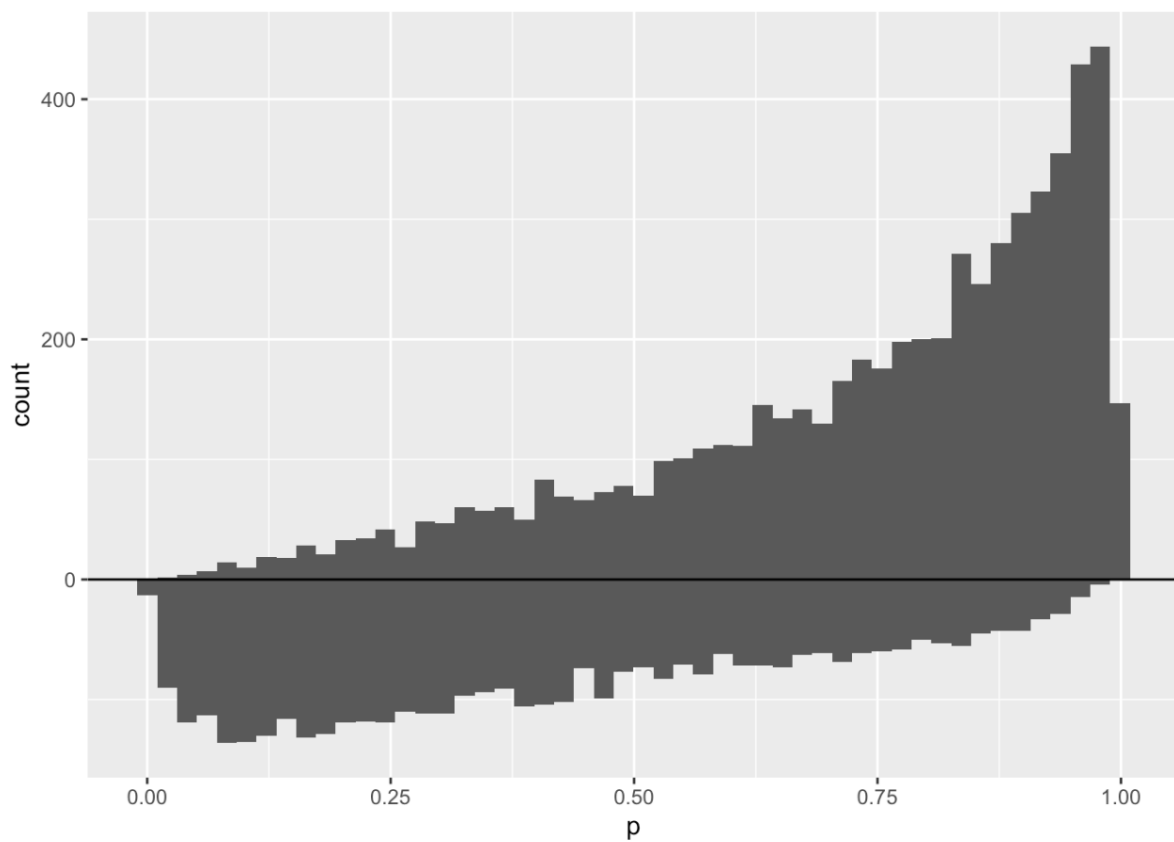


Step 3: Select and estimator & compute an estimate!

- 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

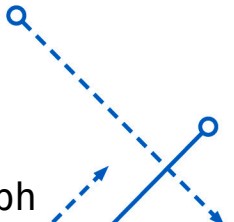
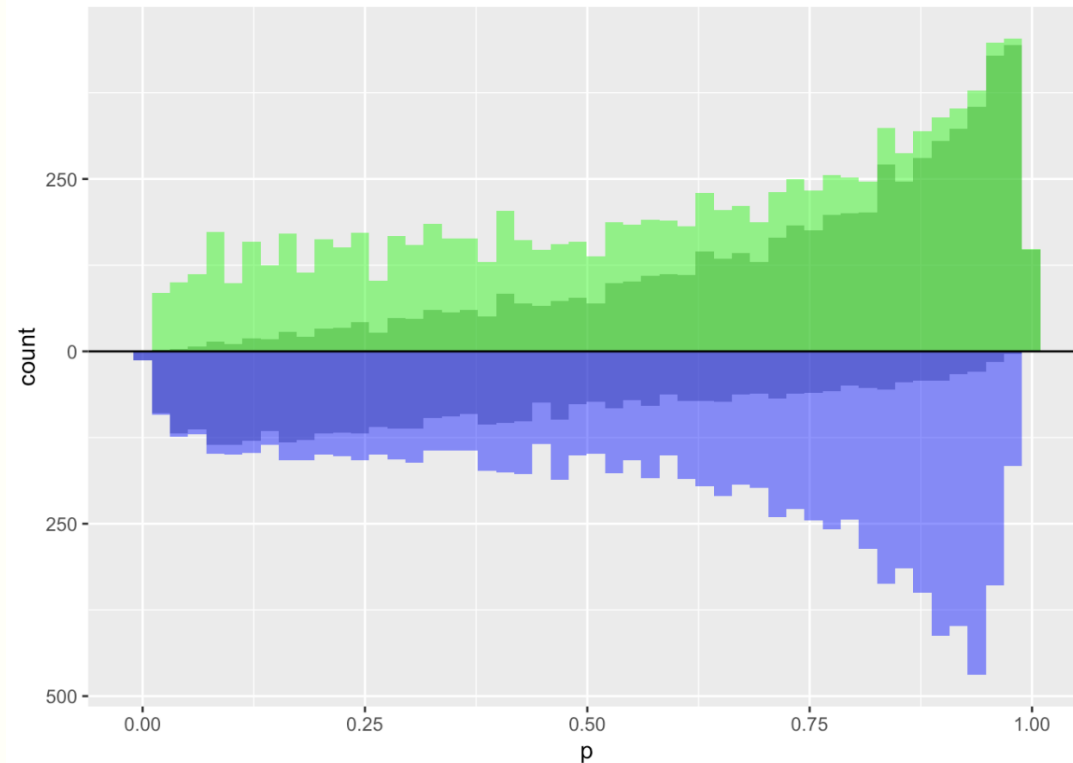
<https://livefreeordichotomize.com/posts/2019-01-17-understanding-propensity-score-weighting/>



Example estimand/estimators using propensity scores for the ATE

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

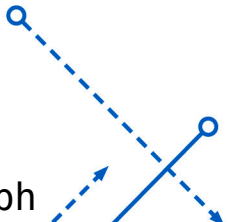
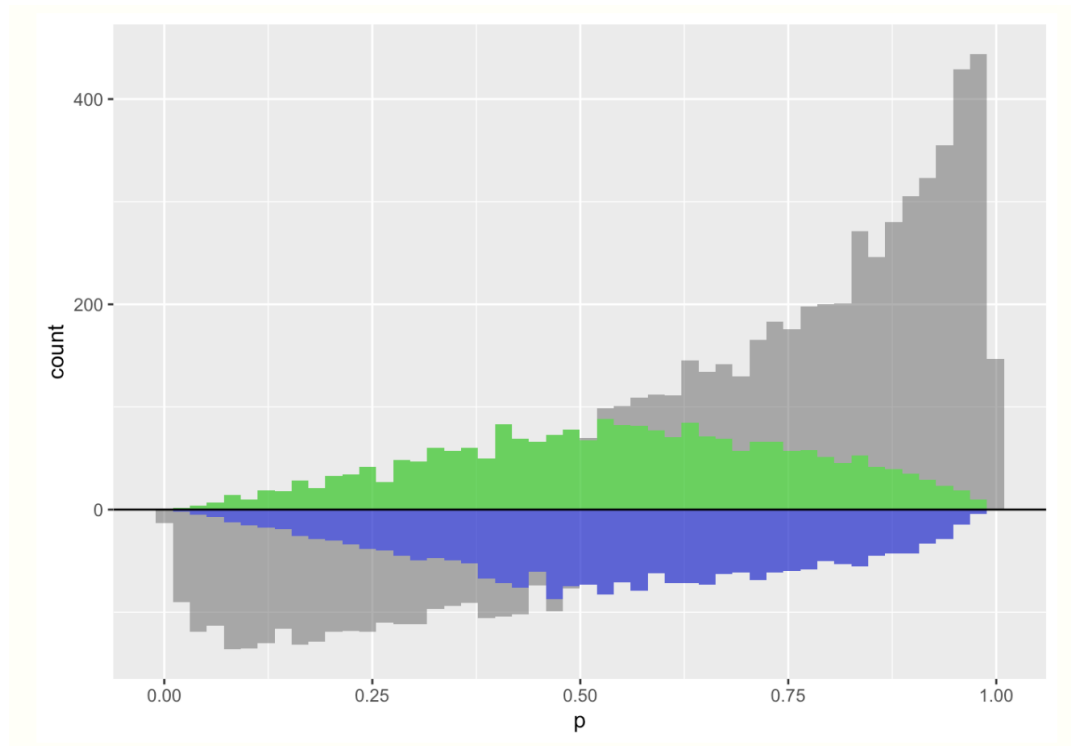
$$W_{ATE} = \frac{Z_i}{e_i} + \frac{1-Z_i}{1-e_i}$$



Example estimand/estimators using propensity scores for the ATO

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

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



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



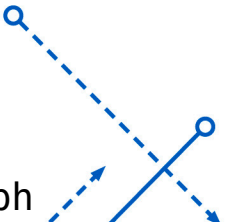
Will Lowe

@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



An example

- https://colab.research.google.com/drive/106f_x1OXA6qeMFxd0EYhSQGP6dkSLk79?usp=sharing

Where's the *ML* in this?
