Causal Inference I

Mixtape Session

MIXTAPE SESSIONS

Roadmap

Directed Acyclic Graphs

- Graph notation
- Backdoor criterion
- Collider bias
- Concluding remarks

Graphs

- Now we turn from potential outcomes modeling of causal effects to causal graphs
- Very important area, very common to see it in computer science intersections with data science, particularly tech, and often very advanced
- My focus is very narrow – I am using it mainly to help us carefully reason through design elements around matching and instrumental variables

Judea Pearl, 2011 Turing Award winner, drinking his first IPA

Judea Pearl and DAGs

- Judea Pearl and colleagues in Artificial Intelligence at UCLA developed DAG modeling to create a formalized causal inference methodology
- They make causality concepts extremely clear, they provide a map to the estimation strategy, and maybe best of all, they communicate to others what must be true about the data generating process to recover the causal effect

Further reading

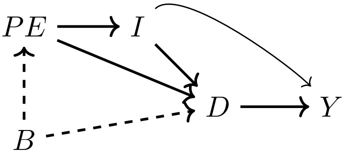
1. Pearl (2018) The Book of Why: The New Science of Cause and Effect, Basic Books (*popular*)
2. Morgan and Winship (2014) Counterfactuals and Causal Inference: Methods and Principles for Social Research, Cambridge University Press, 2nd edition (*excellent*)
3. Pearl, Glymour and Jewell (2016) Causal Inference In Statistics: A Primer, Wiley Books (*accessible*)
4. Pearl (2009) Causality: Models, Reasoning and Inference, Cambridge, 2nd edition (*difficult*)

Design vs. Model

- DAGs tend to be focused more on the theory of treatment assignment in the world
- As such it's compatible with design-based approaches
- DAGs have become extremely common in industry and machine learning, so consider my review very basic comparatively as I will use them mainly to illustrate "good vs bad controls" as well instrumental variables

Causal model

- The causal model is sometimes called the structural model, but for us, I prefer the former as it's less alienating
- It's the system of equations describing the relevant aspects of the world
- It necessarily is filled with causal effects associated with some particular comparative statics
- Consider the following diagram representing the returns to education with simplified confounders



- B is a **parent** of PE and D
- PE and D are **descendants** of B
- There is a **direct (causal) path** from D to Y
- There is a **mediated (causal) path** from B to Y through D
- There are six **paths** from PE to Y but none are direct, but some of them are different in other ways

Where do DAGs come from?

- DAGs are meant to represent “contemporary agreement among experts” – if you aren’t willing to present your DAG before a room of experts, it’s likely you shouldn’t use it at all
- Your DAG should be a reasonable approximation of D and Y parents (confounders) and direct and indirect effects of D on Y
- We get ideas for DAGs from theory, models, observation, experience, prior studies, intuition, as well as conversations with domain experts

Unconfoundedness and the backdoor criterion

- DAGs help us understand the source of problems in our observational (non-experimental) data that make inferring causality hard
- But it also can help us see a way out in some situations
- We will focus today on the unconfoundedness research design, which is best described in causal graphs with the concept of the **backdoor criterion**
- As we will see, the DAG helps you solve the problem of choosing covariates for a model to resolve selection bias, but to do so requires confidence in your DAG

Confounding

- Confounding occurs when when the treatment and the outcomes have a common parent node as that creates spurious correlation between D and Y

$$\begin{tikzcd} D \arrow[rr] & & Y \\ & X \arrow[ul] \arrow[ur] & \end{tikzcd}$$

- The *correlation* between D and Y is a biased measure of the average causal effect of D on Y because of selection bias from the confounder (ignoring for now heterogenous treatment effects bias)

Backdoor Paths

- Confounding creates **backdoor paths** between treatment and outcome ($D \leftarrow X \rightarrow Y$) – i.e., spurious correlations
  - $\rightarrow$ Not the same as a collider path ($D \rightarrow X \leftarrow Y$)
  - $\rightarrow$ and not the same as a mediator path ($D \rightarrow X \rightarrow Y$)
- We can "block" any particular backdoor path by conditioning on variable X so long as it is not a collider
- Once we condition on X, the correlation between D and Y estimates the causal effect of D on Y
- Conditioning means calculating $E[Y|D=1,X] - E[Y|D=0,X]$ for each value of X then combining (e.g., integrating) but we discuss this more later

$$\begin{tikzcd} D \arrow[rr] & & Y \\ & \boxed{X} \arrow[ul] \arrow[ur] & \end{tikzcd}$$

Blocked backdoor paths

A backdoor path is blocked if and only if:

- It contains a noncollider that has been conditioned on
- Or it contains a collider that has not been conditioned on

Note: A path which has a conditioned-on-collider can still be closed, but only with a noncollider-conditioned-on (we will see this later)

Backdoor criterion

Backdoor criterion

Conditioning on X satisfies the backdoor criterion with respect to (D, Y) directed path if:

1. All backdoor paths are blocked by X
2. No element of X is a collider

In words: If X satisfies the backdoor criterion with respect to (D, Y), then controlling for or matching on X identifies the causal effect of D on Y.

And again note that a path which has a conditioned-on-collider can still be closed, but only with a noncollider-conditioned-on

What control strategy meets the backdoor criterion?

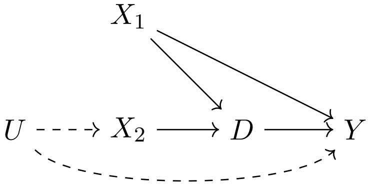
- List all backdoor paths from D to Y. I'll wait.

$$
\begin{tikzcd}
X_1 \arrow[r] \arrow[rrr, bend left] & D \arrow[rr] & & Y \\
& & X_2 \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

- What are the necessary and sufficient set of controls which will satisfy the backdoor criterion?

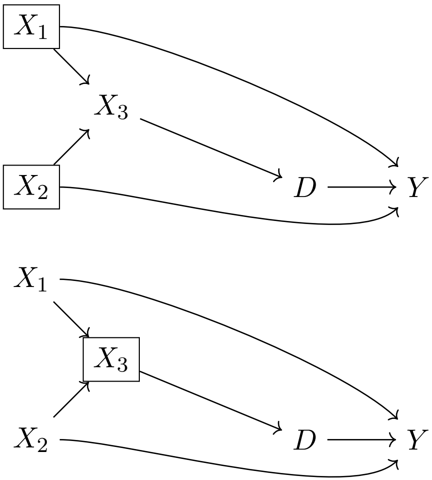
What if you have an unobservable?

- List all the backdoor paths from D to Y.



- What are the necessary and sufficient set of controls which will satisfy the backdoor criterion?
- What about the unobserved variable, U?

Multiple strategies



- Conditioning on the common causes, X_1 and X_2, is sufficient
- …but so is conditioning on X_3

Collider bias

- Backdoor paths can remain open in covariate adjustment strategies through two ways:
  1. You did not close the path because you did not condition on the confounder
  2. Your conditioning variable opened up a previously closed backdoor path because on that path the variable was a **collider**
- Colliders are "bad controls" which when you control for them, *create* new previously non-existent spurious correlations (not commonly discussed, even in economics)
- This is the risk of blindly controlling for variables ("kitchen sink regressions")

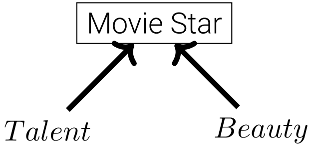
Example 1: Movie stars

Important: Since unconditioned colliders block back-door paths, what exactly does conditioning on a collider do? Let's illustrate with a fun example and some made-up data

- CNN.com headline: Megan Fox voted worst – but sexiest – actress of 2009 (link)
- Are these two things actually negatively correlated in the world?
- Assume talent and beauty are independent, but each causes someone to become a movie star. What's the correlation between talent and beauty for a sample of movie stars compared to the population as a whole (stars and non-stars)?

Movie star DAG

Imagine casting directors pick movie stars based on talent and beauty



Talent and beauty can become correlated even though they are independent

Figure: Top left figure: Non-star sample scatter plot of beauty (vertical axis) and talent (horizontal axis). Top right right figure: Star sample scatter plot of beauty and talent. Bottom left figure: Entire (stars and non-stars combined) sample scatter plot of beauty and talent.

Sample selection?

- Notice that this is clear when we are focused on sample selection
- But even a regression that included “star” would create the issue:

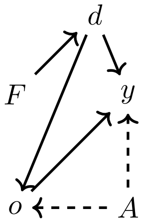
$$talent_i = \alpha + \delta beauty_i + \beta star_i + \varepsilon_i$$

- It’s not just sample selection

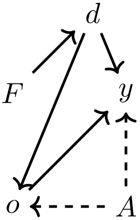
Example 2: Discrimination

- Let's look at another example: very common for think tanks and journalists to say that the gender gap in earnings disappears once you control for occupation.
- But what if occupation is a collider, which it could be in a model with occupational sorting
- Then controlling for occupation in a wage regression searching for discrimination can lead to all kinds of crazy results *even in a simulation where we explicitly design there to be discrimination*

DAG



F is female, d is discrimination, o is occupation, y is earnings and A is ability. Dashed lines mean the variable cannot be observed. Note, by design, being a female has no effect on earnings or occupation, and has no relationship with ability. So earnings is coming through discrimination, occupation, and ability.



Mediation and Backdoor paths

1. $d \rightarrow o \rightarrow y$
2. $d \rightarrow o \leftarrow A \rightarrow y$

Table: Regressions illustrating collider bias with simulated gender disparity

Covariates:	**Unbiased combined effect**	**Biased**	**Unbiased wage effect only**
Female	-3.074***	0.601***	-0.994***
	(0.000)	(0.000)	(0.000)
Occupation		1.793***	0.991***
		(0.000)	(0.000)
Ability			2.017***
			(0.000)
N	10,000	10,000	10,000
Mean of dependent variable	0.45	0.45	0.45

- Recall we designed there to be a discrimination coefficient of -1
- If we do not control for occupation, then we get the combined effect of $d \rightarrow o \rightarrow y$ and $d \rightarrow y$
- Because it seems intuitive to control for occupation, notice column 2 - the sign flips!
- We are only able to isolate the direct causal effect by conditioning on ability and occupation, but ability is unobserved

- **Colliders can be outcomes (and often those are the ones)**
  - → There is only one backdoor path from D to Y

$$\begin{tikzcd}
\boxed{X_1} \arrow[r] \arrow[rrr, bend left] & D \arrow[rr] \arrow[dr] & & Y \arrow[dl] \\
& & X_2 &
\end{tikzcd}$$

  - → Conditioning on X_1 blocks the backdoor path
  - → But what if we also condition on X_2?

$$\begin{tikzcd}
\boxed{X_1} \arrow[r] \arrow[rrr, bend left] & D \arrow[rr] \arrow[dr] & & Y \arrow[dl] \\
& & \boxed{X_2} &
\end{tikzcd}$$

  - → Conditioning on X_2 opens up a new path, creating new spurious correlations between D and Y

- **Colliders could be pre-treatment covariates (called M-bias because it looks like an M)**
  - → Name the backdoor paths. Is it open or closed?

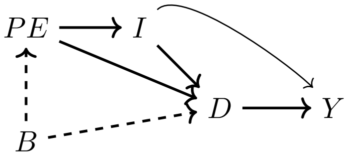
$$\begin{tikzcd}
U_1 \arrow[dr, dashed] \arrow[drrr, dashed, bend left] & & & \\
& X & D \arrow[r] & Y \\
U_2 \arrow[ur, dashed] \arrow[urr, dashed, bend right] & & &
\end{tikzcd}$$

  - → But what if we condition on X?

$$\begin{tikzcd}
U_1 \arrow[dr, dashed] \arrow[drrr, dashed, bend left] & & & \\
& \boxed{X} & D \arrow[r] & Y \\
U_2 \arrow[ur, dashed] \arrow[urr, dashed, bend right] & & &
\end{tikzcd}$$

Testing the Validity of the DAG

- The DAG makes testable predictions
- Conditional on D and I, parental education (PE) should no longer be correlated with Y
- Can be hard to figure this out by hand, but software can help (e.g., Daggity.net is browser based, Causal Fusion is more advanced)
- Causal algorithms tend to be DAG based and are becoming popular in industry



Summarizing all of this

- Your dataset will not come with a codebook flagging some variables as “confounders”, “mechanisms” and “colliders” because those terms are always context specific
- Except for some unique situations that aren’t generally applicable, you also don’t always know statistically you have an omitted variable bias problem; but both of these are fatal for any application
- You only know to do what you’re doing based on *knowledge about data generating process*.
- All identification must be guided by theory, experience, observation, common sense and knowledge of institutions
- DAGs absorb that information and can be then used to write out the explicit identifying model