Lecture06

April 20, 2023

1 POLI 273

1.1 Causal Inference

1.1.1 Lecture 06 - DAGs for Causal Identification and Regression

1.2 Announcements

- This class GitHub: https://github.com/umbertomig/POLI273
- PS01: How is it going?
- Qualtrics Videos: It is still missing the Conjoint Video and the Case-Control Video.
  - I'll post them this week

1.3 Agenda for Today

- DAGs:
  - d-Separation
  - Backdoor criterion
  - Why they are good / connections with PO.
- Regression Analysis:
  - Linear Regression
  - Consistency
  - Regression for Causal Inference

1.4 DAGs

- Three relations in the paths:
  - Chains:

$$T \longrightarrow C \longrightarrow Y$$

  - Forks:

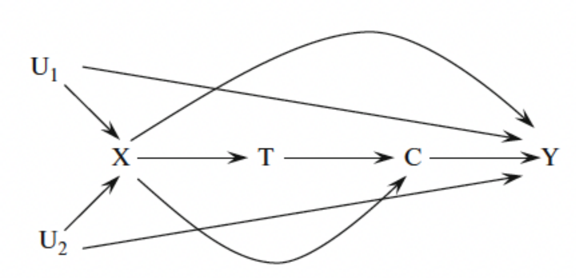
$$T \longleftarrow C \longrightarrow Y$$

  - Inverted Forks (Colliders):

$$T \longrightarrow C \longleftarrow Y$$

1.5 DAGs

- Consider this DAG (Elwert, 2013):



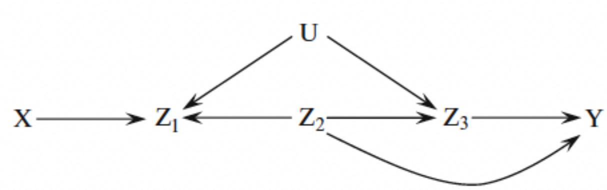
- Let us see what happens in the data.

1.6 DAGs

- [Pearl 1988] **d-Separation**: A path between two variables A and B is said to be d-separated (blocked; closed) if:
  1. The path contains a **non-collider that has been conditioned on**, OR
  2. The path contains a **collider that has *not* been conditioned on**.
- 2. If two variables A and B are d-separated conditional on a third variable (or sets of variables) C, then $A \perp B | C$.
  3. If your DAG is faithful and the variables A and B are d-connected (or not d-separated), then they are dependent (faithfulness: assumes your DAG is certain; weak faithfulness: assume your DAG is a conjecture).

1.7 DAGs

- Are X and Y d-separated?



1. Draw the paths.
2. Study them.
3. What happens when you start conditioning?

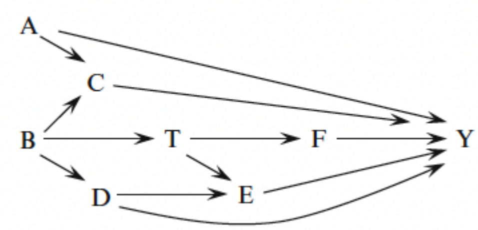
1.8 DAGs

- [Shpister et al. 2010]: **Adjustment criterion**: A set of observed variables Z (that may be empty) satisfies the adjustment criterion relative to the total causal effect of a treatment T on an outcome Y if:

1. Z blocks all non-causal paths from Z to Y, AND
2. No variable in Z lies on or decents from a causal path from T to Y.

1.9 DAGs

Consider the following DAG, where we are interested in the effect of a treatment T on Y:



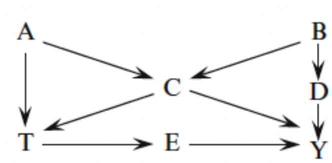
1. Draw the paths.
2. Study them.
3. Can we use the adjustment criterion to identify the effects causally? (Hint: Nine possible adjustments. Find two).

1.10 DAGs

-
- [Pearl 1993]: **Backdoor criterion**: A set of observed variables Z (that may be empty) satisfies the backdoor criterion relative to the total causal effect of a treatment T on an outcome Y if:
  1. No element of Z is a descendant of T, AND
  2. Z blocks all backdoor paths from T to Y.

1.11 DAGs

Consider the following DAG, where we are interested in the effect of a treatment T on Y:



1. Draw the paths.
2. Study them.
3. Can we use the *backdoor criterion* to identify the effect causally? (Met by seven adjustments. Find two.)

1.12 DAGs

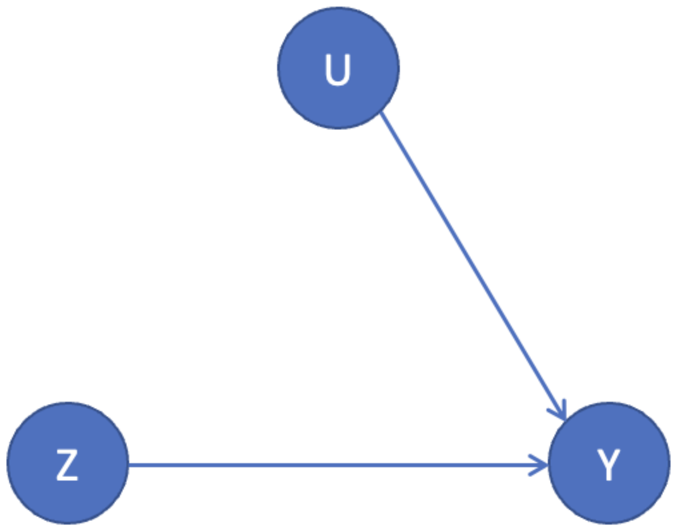
- More on DAGs later. Now, let us see the implication for Experimental Political Science.
- Suppose we have a random treatment Z and an outcome of interest Y. And let us assume that:

$$(Y_1, Y_0) \perp Z \text{ and positivity}$$

This means that:

- $Y_1 \perp Z$ and $Y_0 \perp Z$. From what we know from the adjustment criterion:

$$\tau = \mathbb{E}[Y_1 - Y_0] = \mathbb{E}[Y_1] - \mathbb{E}[Y_0]$$



1.13 DAGs

- Suppose we have a treatment Z, a variable X, and an outcome of interest Y. And let us assume that:

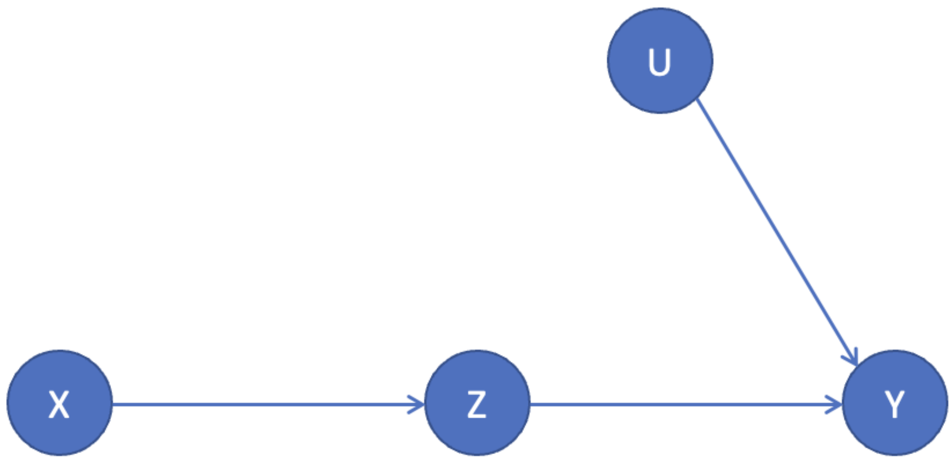
$$(Y_1, Y_0) \perp Z|X \text{ and positivity}$$

- This means that:

1. $\mathbb{E}[Y_1] = \sum_{\text{Supp}(X)} \mathbb{E}[Y|Z = 1, X = x]\mathbb{P}(X = x)$
2. $\mathbb{E}[Y_0] = \sum_{\text{Supp}(X)} \mathbb{E}[Y|Z = 0, X = x]\mathbb{P}(X = x)$

And:

$$\tau = \mathbb{E}[Y_1 - Y_0] = \mathbb{E}[Y_1] - \mathbb{E}[Y_0]$$



2 Regression for Causal Identification

2.1 Regression for Causal Identification

- A linear regression is a *special* projection on a space.
- Suppose we have two sets of variables (X, y) (X can be more than one var).
- The conditional expectation of y given X can be decompose as:

$$\mathbf{y} = \mathbb{E}\big[\mathbf{y} \mid \mathbf{X}\big] + \varepsilon$$

2.2 Regression for Causal Identification

$$\mathbf{y} = \mathbb{E}\big[\mathbf{y} \mid \mathbf{X}\big] + \varepsilon$$

And if this is the case, then:

Theorem: $\mathbb{E}\big[\varepsilon \mid \mathbf{X}\big] = \mathbf{0}$. Moreover:

1. For any function ϕ, $\mathbb{E}\big[\phi(\mathbf{X})\varepsilon\big] = \mathbf{0}$.
2. In particular: $\$\mathrm{E}\big[\mathbf{X'}\boldsymbol{\varepsilon}\ \big] = \backslash\mathbf{bf\{0\}}\ \$$

2.3 Regression for Causal Identification

- Now, pick the result:

$$\mathbb{E}(\mathbf{X}'\varepsilon) = \mathbf{0}$$

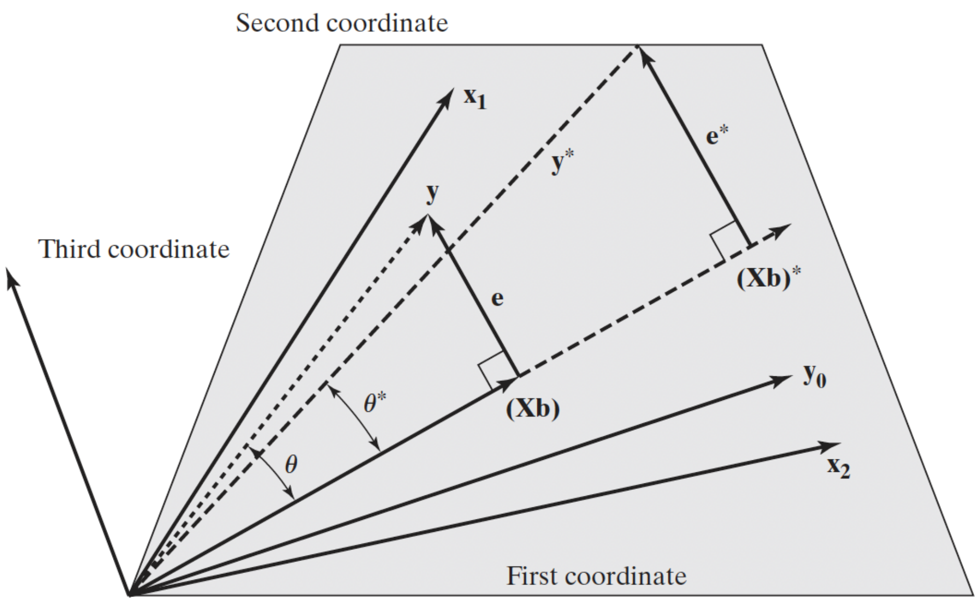
- What does it mean?
  - In linear-algebra-terms, it means that if you project the variables on the residuals, the relationship is *orthogonal*.

– Remember that, by the cosine law: the angle between two vectors is $cos(\theta) = \dfrac{\mathbf{a}'\mathbf{b}}{|\mathbf{a}|\ |\mathbf{b}|}$.

2.4 Regression for Causal Identification

- Or:

FIGURE A.3 Least Squares Projections.



2.5 Regression for Causal Identification

Let $f(\mathbf{X}) = \mathbf{X}'$ and that the conditional expectation of y given X is linear (i.e., $E(**y**|**X**) = **X\bf{β} $). Then:**

$$
\begin{aligned}
\mathbf{0} &= \mathbb{E}(f(\mathbf{X})\varepsilon) \\
&= \mathbb{E}(\mathbf{X}'\varepsilon) \\
&= \mathbb{E}(\mathbf{X}'(\mathbf{y} - \mathbb{E}(\mathbf{y}|\mathbf{X}))) \\
&= \mathbb{E}(\mathbf{X}'(\mathbf{y} - \mathbf{X}\beta)) \\
&= \mathbb{E}(\mathbf{X}'\mathbf{y}) - \mathbb{E}(\mathbf{X}'\mathbf{X}\beta)
\end{aligned}
$$

Thus:

$$\beta \ = \ \mathbb{E}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbb{E}\left(\mathbf{X}'\mathbf{y}\right)$$

2.6 Regression for Causal Identification

- **Let us now look at each individual value.**

Theorem: Assuming that for each i:

$$Y_i \;=\; \mathbb{E}(Y_i|X_i) + \varepsilon_i$$

Then, $\mathbb{E}\big[\varepsilon_i|X_i\big] = 0$.

2.7 Regression for Causal Identification

What is the `minimum mean squared error predictor` (MHE, Chapter 3)?

Theorem: Suppose that there exists a function $m(X_i)$ that minimizes:

$$S(m(X_i)) \;=\; \mathbb{E}\Big[\big(Y_i - m(X_i)\big)^2\Big]$$

Then, $m(X_i) = \mathbb{E}\big(Y_i|X_i\big)$.

2.8 Regression for Causal Identification

- **Since we operate in a sample, can we recover the *population* parameters using the *sample* parameters?**

Theorem: Let \$β $\;=\; \mathbf{E}\big[\mathbf{X'X}\big]^{-1}\mathbb{E}\big[\mathbf{X}'\mathbf{y}\big]$ and let the MSE estimator in the sample $\widehat{\beta} \;=\; \big[\mathbf{X}'\mathbf{X}\big]^{-\mathbf{1}}\mathbf{X}'$\bf{y}\$.Then :

$$\mathbb{E}\big[\widehat{\beta}|\mathbf{X}\big] \;=\; \beta + \mathbb{E}\Big[\big[\mathbf{X}'\mathbf{X}\big]^{-\mathbf{1}}\mathbf{X}'\varepsilon \mid \mathbf{X}\Big] \;=\; \beta$$

But how does that help us to find the ATE?

2.9 Regression for Causal Identification

Let the ATE:

$$\tau \;=\; \mathbb{E}\big[Y_{1i} - Y_{0i}\big]$$

And assume all assumptions hold: positivity + CIA. This means that $(Y_1, Y_0) \perp Z$ (and remember that $Y_{1i} = (Y_i|Z_i = 1)$).

2.10 Regression for Causal Identification

And let us consider that Y_i is equal to:

$$Y_i \;=\; Z_i\beta + \varepsilon_i$$

It is easy to see that:

$$\mathbb{E}\big[Y_i|Z_i\big] \;=\; \mathbb{E}\big[Z_i|Z_i\big]\beta + \mathbb{E}\big[\varepsilon_i|Z_i\big]$$

And since these are exogenous: $\mathbb{E}\big[\varepsilon_i|Z_i\big] = 0$, what makes:

$$\beta \;=\; \mathbb{E}\big[Y_i|Z_i\big]$$

2.11 Regression for Causal Identification

This is still not that helpful, but remember that $Y_i = Z_i Y_{1i} + (1 - Z_i)Y_{0i}$. A little algebra gets:

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})Z_i$$

And plugging this in:

$$\beta = \mathbb{E}\big[Y_{0i} + (Y_{1i} - Y_{0i})Z_i|Z_i\big]$$

And now what? Complete the proof to find that: $\beta = \tau$!

2.12 Regression for Causal Identification

2.12.1 The linear regression coefficient is equivalent to the differences-in-means estimator.

2.13 Questions?

2.14 See you in the next class!