Causal Inference Workshop (Day 2: Graphical models) Pablo Geraldo (pdgeraldo@ucla.edu) SICSS - UCLA June 25, 2021

Recap from day 1

- We studied the difference between **seeing** (the focus of Stats/ML) and **doing** (the focus of causal inference)
- We studied the importance of using the formal language of potential outcomes to:
 - clarify what do we want to know (estimand)
 - identify *reasons for discrepancies* between what we observe and our target (bias)
 - formalize *what needs to be true* for our estimand to be identified with a given *estimator* (assumptions)
- We studied the role of randomization to identify causal effects by design.

But what do we do when we have *less than perfect* experiments? How to assess our assumptions with *observational data*?

Why graphical models?

Formally, potential outcomes are sufficient to specify our estimand, sources of bias, and assumptions needed for causal identification.

However, assessing the plausibility of identification assumptions rely on the researchers being able to reason about (conditional) independence between *counterfactual* variables.

Any ideas on how to assess the assumption about the *conditional* independence of the **potential outcomes** with respect to the treatment, without randomization?

We can certainly *understand* the statement saying that the treatment is assigned **as-if random** adjusting for covariates. But what about its plausibility?

Why graphical models?

Assessing ignorability

When we say the treatment assignment is strongly ignorable we are stating that $P(Y_x|X=x) = P(Y_x)$, but we never get to observe the full distribution of potential outcomes!

- What type of criteria should we use when discussing others' causal claims?
- What kind of criteria should we use in our own research to judge if we are getting what we are looking for?

Here is where DAGs shine, offering a graphical criteria that is equivalent to the unconfoundedness statement, the *backdoor criterion*.

Structural Causal Model

Structural Causal Model (SCM)

Unifying approach to causal inference, developed by Pearl, Robins, among others:

- (Non-parametric) Structural Equation Models
 - Generalization of the path analysis and SEM you might be familiar with
- Graphical representation using Directed Acyclic Graphs (DAGs)
- Potential Outcomes are **derived** from a SCM
- Transparent representation of qualitative assumptions
- Testable implications of our model of the data generating process

Probabilistic graphical models are mathematical objects that represent relations among variables (probability factorizations)

They are compounded by two ingredients: **nodes** (vertices) and **edges** (links)

Directed Acyclic Graphs (DAGs) are one class of graphical models, with the following characteristics:

- Directed: The edges point *from* one variable *to* another variable
- Acyclic: The paths in the graph flow in certain direction, if you follow the edges you cannot arrive back to the starting point
- Graph: well, you get it!

Important:: under certain conditions, a DAG can be causally interpreted, in which case we talk about "causal DAGs" or causal diagrams

Basically, this happen when we assume that no pair of nodes share a common ancestor that is not included in the DAG

We can go from one variable to another following a *path* along the edges

When you can traverse a path without colliding into an edge in the opposite direction we call it a **connecting path** that transmit information

When you encounter an edge pointing into the opposite direction along a path we call it a **blocking path** that do not transmit information

Faithfulness: d-connection in the graph implies association between variables in reality, while d-separation implies their independence

We can go from one variable to another following a *path* along the edges

When you can traverse a path without colliding into an edge in the opposite direction we call it a **connecting path** that transmit information

When you encounter an edge pointing into the opposite direction along a path we call it a **blocking path** that do not transmit information

• A chain, in which you can travel from X to Y through M, is a d-connected path:

• A fork, in which you can go from a common cause W to both X and Y is a d-connected path:

$$X \leftarrow W
ightarrow Y$$

• A collider, in which you can't go from X to Y due to two edges pointing into a third variable C, is a d-separated path:

$$X \to C \leftarrow Y$$

By adjusting for a variable (represented by a box in the graph), we can turn connecting into blocking paths and vice versa

Therefore, we can have conditional d-separation, and conditional d-connection

- When you adjust for the intermediate variable M in a chain, X and Y become conditionally independent:

• When you adjust for the common cause W in a fork, X and Y become conditionally independent:

$$X \leftarrow W \to Y$$

• When you adjust for a collider variable C, the pair X and Y become conditionally associated:

$$X \cdots \overline{C} \cdots Y$$

Confounding paths

$$egin{aligned} & ext{Confounding} \ & W &= f_w(U_w) \ & X &= f_x(W,U_x) \ & Y &= f_y(W,X,U_y) \end{aligned}$$



Are X and Y marginally independent?

Are they conditionally independent?

The DAG includes the following paths

 $egin{array}{c} X o Y \ X \leftarrow W o Y \end{array}$

Confounding paths





Are X and Y marginally independent?

Are they conditionally independent?

The following path is open

X o Y

But this is now closed

 $X \leftarrow \overline{W}
ightarrow Y$

Confounding paths





Are X and Y marginally independent?

Are they conditionally independent?

Is there a causal effect of X on Y?

The DAG includes the following path

 $X \leftarrow W \to Y$

Mediating paths





Are X and Y marginally independent? Are they conditionally independent?

Is there a causal effect of X on Y?

The DAG includes the following paths

 $egin{array}{c} X o Y \ X o M o Y \end{array}$

Mediating paths





Are X and Y marginally independent? Are they conditionally independent?

Is there a causal effect of X on Y?

The DAG includes the following path

 $X \to Y$

But this path is now closed

X o M o Y

Colliding paths





Are X and Y marginally independent?

Are they conditionally independent?

Is there a causal effect of X on Y?

The DAG includes the following paths

 $egin{array}{c} X o Y \ X o C \leftarrow Y \end{array}$

Colliding paths





Are X and Y marginally independent?

Are they conditionally independent?

Is there a causal effect of X on Y?

The DAG includes the following path

 $X \to C \leftarrow Y$

Colliding paths





Are X and Y marginally independent?

Are they conditionally independent?

Is there a causal effect of X on Y?

The DAG includes the following (open) path

$$X \to \fbox{C} \leftarrow Y$$

Side note: are colliders that important?

One common question (and an area of debate among practitioners) is if colliders are really that important *in applied settings*. This is a hard question to answer, because, you know... we just don't know.

But we know that they are **possible** and that their importance would depend on the structure of our causal graph.

A few compelling examples of collider bias in recent social sciences are discussed by:

- Shalizi and Thomas (2011) in the context of network homophily and contagion
- Richard Breen (2018) in the context of intergenerational mobility
- Knox, Lowe and Mummolo (2020) in the context of police shootings

A great general introduction to the topic is offered by Elwert and Winship (2014)

Do-operator and interventions

Pearl introduced the do-operator to clearly distinguish between passive observations and interventions on the data generating process

In other words, is a form to make explicit the gap between interventional quantities and out more familiar conditional expectations

Causal identification corresponds to removing the do-operator from an expression, following the rules of do-calculus, reducing it to an observational quantity. If there is no equivalece, it means that the quantity of interest is not identified

Given the correspondence between a system of non-parametric structural equations and a given DAG, we can express the operation of **doing** as a *minimal surgery on the structural equation defining the treatment*

Let's start with the following observational data generating process

 $egin{aligned} {
m Structural Causal Model}\ Z_1 &= f_{z1}(U_{z1})\ Z_2 &= f_{z2}(U_{z2})\ W &= f_w(Z_1, Z_2, U_w)\ X &= f_{z1}(Z_1, W, U_x)\ Y &= f_y(Z_2, W, U_y) \end{aligned}$



Intervening in the model to make X=x creates an interventional graph $G_{\bar{X}}$, in which all the incoming arrows into X have been removed

Structural Causal Model $Z_1=f_{z1}(U_{z1})$ $Z_2=f_{z2}(U_{z2})$ $W=f_w(Z_1,Z_2,U_w)$ X=x $Y=f_y(Z_2,W,U_y)$



The purpose of an observational study is to allow only causal paths between the treatment X and the outcome Y, and block all the non causal paths

 $egin{aligned} {
m Structural Causal Model}\ Z_1 &= f_{z1}(U_{z1})\ Z_2 &= f_{z2}(U_{z2})\ W &= f_w(Z_1, Z_2, U_w)\ X &= f_{z1}(Z_1, W, U_x)\ Y &= f_y(Z_2, W, U_y) \end{aligned}$



Adjusting for W blocks a non-causal path, but opens a new one.

P(Y|do(x)) is not identified conditioning on W alone

The purpose of an observational study is to allow only causal paths between the treatment X and the outcome Y, and block all the non causal paths

 $egin{aligned} {
m Structural Causal Model}\ Z_1 &= f_{z1}(U_{z1})\ Z_2 &= f_{z2}(U_{z2})\ W &= f_w(Z_1, Z_2, U_w)\ X &= f_{z1}(Z_1, W, U_x)\ Y &= f_y(Z_2, W, U_y) \end{aligned}$



Adjusting for (Z_1,W) blocks all non-causal paths

P(Y|do(x)) is identified conditioning on (Z_1,W)

The purpose of an observational study is to allow only causal paths between the treatment X and the outcome Y, and block all the non causal paths

 $egin{aligned} {
m Structural Causal Model}\ Z_1 &= f_{z1}(U_{z1})\ Z_2 &= f_{z2}(U_{z2})\ W &= f_w(Z_1, Z_2, U_w)\ X &= f_{z1}(Z_1, W, U_x)\ Y &= f_y(Z_2, W, U_y) \end{aligned}$



Adjusting for (Z_2,W) also blocks all non-causal paths

P(Y|do(x)) is identified conditioning on (Z_2,W)

Back-door Criterion (Pearl)

What we just did can be summarized by the back-door criterion

A set of variables W satisfied the back-door criterion relative to an ordered pair of variables (X,Y) in a DAG G if:

(i) no node in W is a descendant of X; and (ii) W blocks every path between X and Y that contains an arrow into X

Important: satifying the backdoor criterion implies the unconfoundedness assumption

$Y_x \perp\!\!\!\perp X | W$

The back-door adjustment (aka g-formula) indicates that we can recover the effect of X on Y adjusting for any W that satisfy the backdoor criterion

$$P(Y|do(x)) = \sum_w P(Y|X,W)P(W)$$

Pearl, Causality, pp.79-81

SCM: Key insights

- Causal identification is contingent on a given model encoding our assumptions
- Causal identification is finding an observational quantity that is equivalent to an interventional quantity
- Confounding (and bias) is a property of paths in a graph, not variables
- Confounding is relative to the pair (X,Y), not just X
- It is not necessary to adjust for all *parents* of the treatment to block all backdoor paths
- Bias is not monotonically decreasing on the number of variables included
- *do*-calculus can be used to identify the effect of multiple interventions, to recover from missingness data, and to generalize study results.

SCM: limitations

One problem with the Structural Causal Model framework is that you need to assume a certain DAG. Conditional on your model, most identification tasks are rather trivial (algorithmic)

- "Causality is in the model" James Heckman (2005)

However, this is something that we always do! The only matter is how transparent are we about the assumptions we are making anyway

Another problem, more important for issues like mediation analysis and counterfactuals in general, is that we are sometimes making more assumptions that we are willing to

For this cases, other causal (and graphical) models, like the Single World Intervention Graphs (Richardson and Robins, 2013) can help

Being fully non-parametric, certain canonical models are not identified using DAGs (like IVs). However, this only shows that they require parametric assumptions, no matter how weak

Identification Strategies

Identification Strategies: a short tour

Identification strategies are well-known set of assumptions that are sufficient for the causal interpretation of certain estimators:

- Randomization ensures that the potential outcomes are independent from the treatment in experiments. This justifies a causal interpretation of a diff in means estimator.
- Selection on Observables (aka Conditional Ignorability, Unconfoundedness) for multiple regression, (propensity score) matching and (inverse probability) weighting.
 - In FE confounders are assumed to be fully captured by constant characteristics of the individual, group, or time.
- Parallel trends for the difference-in-difference estimator (equivalent to a two-way fixed effects only in the two group, two periods case)
 - Generalizations for multiple groups, multiple adoption periods, and synthetic controls
- Instrumental Variables and quasi-experiments (exogenous variation in the treatment assignment plus exclusion restriction) for the 2SLS and Wald estimator
 - RDD can be interpreted as IVs or under continuity assumptions.

Is there a ladder of identification strategies?

In general, it is assumed that the stronger the assumptions, the less credible an identification strategy would be

We would prefere experiments (the so-called *gold standard*) and, in the lack of experiments, we would prefer strategies in which our assumptions hold by design

Designs in which there is an exogenous source of variation in the treatment status (like IV, RDD, quasi-experiments in general) are considered more plausible

However, is there a natural hierarchy of identification strategies that can tell us, *a priori*, which assumptions are more credible in empirical applications?

Do empirical applications corresponds to the labels that we use to describe a given research design? Any thoughts?

Debates in applied research

Partly as a result of the focus on empirical examples the econometrics literature has developed a small number of canonical settings where researchers view the specific causal models and associated statistical methods as well-established and understood. These causal models correspond to what is nowadays often referred to as identification strategies

Guido Imbens (2019)

About 20 years ago, when asked in a meeting what can be done in observational studies to clarify the step from association to causation, Sir Ronald Fisher replied: 'Make your theories elaborate.' The reply puzzled me at first, since by Occam's razor, the advice usually given is to make theories as simple as is consistent with known data. What Sir Ronald meant, as subsequent discussion showed, was that when constructing a causal hypothesis one should envisage as many different consequences of its truth as possible, and plan observational studies to discover whether each of these is found to hold.

B. G. Cochran (cited in Rosenbaum, 1995)

No one should ever write down a 100 variable DAG and do inference based on that. That would be an insane approach because the analysis would be totally impenetrable. Develop a research design where that 100 variable DAG trivially reduces to a familiar problem (e.g. IV!)

Jason Abaluck (cited in Imbens, 2019)

Where to go now?

- Structural Causal Model
 - The Book of Why (Pearl and MacKenzie)
 - Causal Inference in Statistics. A Primer (Pearl, Glymour and Jewell)
 - Causality (Pearl)
- Potential Outcomes
 - Mostly Harmless Econometrics (Angrist and Pischke)
 - Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction (Imbens and Rubin)
- General and integrative introductions
 - Counterfactuals and Causal Inference (Morgan and Winship)
 - Causal Inference (Hernan and Robins)
 - Causal Inference: The Mixtape (Cunnigham)

We are done!

Activity: Breakout rooms

Open www.dagitty.net and launch it. There, you can create a DAG, obtain a list of testable implications, and check if an effect is identified under the DAG.

- Skim the following paper by Sharkey et al. (see Analytical Approach) and try to reconstruct the underlying DAG for the long-term model:
 - Share your result
 - Is the effect identified?
 - Can you think in possible violations of their assumptions?
- Skim the following paper by Doyle et al. (see Section III. Empirical Strategy) and try to reconstruct the underlying DAG:
 - Share your result
 - Is the effect identified?
 - Can you think in possible violations of their assumptions?