# Graphical Models For Complex Health Data (P8124)

Daniel Malinsky

Columbia University
dsm2128@cumc.columbia.edu

Discussion of DAGs for causal inference in epi

# Background

A graph $\mathcal{G} = (V, E)$ consists of a set of vertices $V$ and edges $E$

# Background

A graph $\mathcal{G} = (V, E)$ consists of a set of vertices $V$ and edges $E$

A graphical model is a pair $(\mathcal{G}, \mathcal{P})$ where $\mathcal{P}$ is a set of distributions that factorizes according to $\mathcal{G}$ (e.g. Bayesian network model)

# Background

A graph $\mathcal{G} = (V, E)$ consists of a set of vertices $V$ and edges $E$

A graphical model is a pair $(\mathcal{G}, \mathcal{P})$ where $\mathcal{P}$ is a set of distributions that factorizes according to $\mathcal{G}$ (e.g. Bayesian network model)

A causal graphical model is $(\mathcal{G}, \mathcal{P})$ + some *causal interpretation*

$\Rightarrow$ there are multiple causal interpretations out there, can be more or less formally defined

# What is causality about?

There is a very long history in philosophy and in science of debates over the meaning and analysis of causality.

In statistics + CS + social sci, there is not really a single *definition* of causality, but there is some agreement that causality is primarily about (some combo of):

- the consequences of interventions (manipulations)
- counterfactuals ("what if I had taken option B instead of option A?")
- the asymmetric "flow" of "information"

# What is causality about?

Often want to know what to expect from certain actions or policies (what if we prescribed this drug? what if we enacted this policy rule change?).

Causal graphical models can help, especially in settings where we cannot directly experiment (try different actions). Causal models for observational studies are especially common in epidemiology.

# Causal DAG

A causal DAG is a DAG with a causal interpretation. One common interpretation:[1]

- the lack of an arrow from $X$ to $Y$ can be interpreted as the absence of a direct causal effect of $X$ on $Y$ relative to the other variables on the graph
- all common causes, even if unmeasured, of any pair of variables on the graph are themselves on the graph

[1] This is from Hernan and Robins (2020). There are alternative definitions in the literature. See Pearl (2009) and Spirtes et al. (2000).

# Causal DAG

A causal DAG is a DAG with a causal interpretation. One common interpretation:[1]

- the lack of an arrow from $X$ to $Y$ can be interpreted as the absence of a direct causal effect of $X$ on $Y$ relative to the other variables on the graph
- all common causes, even if unmeasured, of any pair of variables on the graph are themselves on the graph

(but what exactly is a “direct causal effect”?)

[1] This is from Hernan and Robins (2020). There are alternative definitions in the literature. See Pearl (2009) and Spirtes et al. (2000).

# Nonparametric Structural Equation Models (NPSEMs)

It can be useful to associate with a DAG a system of equations:

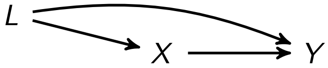
$$X_i = f_i(\mathrm{Pa}(X_i, \mathcal{G}), \epsilon_i)$$

for every $i \in V$

Typically all the error terms are assumed independent:
$p(\epsilon_1, ..., \epsilon_p) = \prod_{i=1}^{p} p(\epsilon_i)$

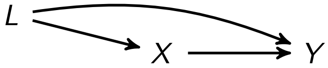
These are called **structural** equations if they are assumed to satisfy certain properties wrt interventions.

# Interventions break arrows, replace some equations

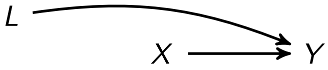


$$\begin{aligned} L &= \epsilon_L \\ X &= f_X(L, \epsilon_A) \\ Y &= f_Y(X, L, \epsilon_Y) \end{aligned}$$

# Interventions break arrows, replace some equations



$$
\begin{aligned}
L &= \epsilon_L \\
X &= f_X(L, \epsilon_A) \\
Y &= f_Y(X, L, \epsilon_Y)
\end{aligned}
$$



$$
\begin{aligned}
L &= \epsilon_L \\
X &= x \\
Y &= f_Y(x, L, \epsilon_Y)
\end{aligned}
$$

# Direct causal effects

In the context of an NPSEM, “no direct causal effect” of $X$ on $Y$ means that $X$ does not appear in the structural equation for $Y$.
$\Rightarrow$ so really the causal interpretation derives from the assumed NPSEM

There are alternative definitions of “direct causal effect” (or causal DAG) based on “potential outcomes” – we won’t get into it here.

# So what's the point of causal DAGs?

Can use graphs to:

- Represent & communicate causal relationships in complex designs or studies
- Determine sources of confounding, decide what to "adjust" for
- Decide using formal rules whether some causal parameter is identified
- Understand and formalize related issues: missing data, drop out, selection into sample
- Sometimes suggest estimation stategies when adjustment won't work
- More...

# Paper by Greenland, Pearl, and Robins

"Causal diagrams can provide a starting point for identifying variables that must be measured and controlled to obtain unconfounded effect estimates."

Key is really about representing assumptions and understanding confounding adjustment.

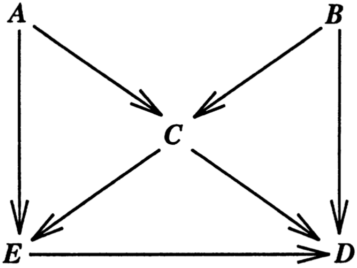
# What is confounding?

Roughly: "... confounding will be present if and only if exposure would remain associated with disease even if all exposure effects were removed, prevented, or blocked."

- Delete all single-headed arrows emanating from the exposure (remove all exposure effects)
- In this new graph, see whether there is any unblocked[2] path from exposure to outcome (see if they remain associated)

This does not tell us which variables are confounders, exactly.

[2] a path is blocked if it has one or more colliders; otherwise it is unblocked



$A$ = air pollution level, $B$ = sex, $C$ = bronchial reactivity, $E$ = antihistamine treatment, $D$ = asthma

Interested in the effect of $E$ on $D$

# Backdoor path

A path from $X$ to $Y$ is called a backdoor path if it has an arrowhead pointing to $X$.

# Backdoor criterion for confounding adjustment

Def: The set $S \subseteq V$ is sufficient for confounding adjustment of the effect of $E$ on $D$ if there is no confounding of the $E - D$ relationship in any strata of $S$.

Criterion: Given a DAG and a set $S$ containing only non-decendents of $E$ or $D$, $S$ is sufficient for adjustment if:

- Every unblocked backdoor path from $E$ to $D$ is intercepted by a variable in $S$
- If every collider on a backdoor path from $E$ to $D$ is either in $S$ or has a descendent in $S$, then $S$ must also contain a noncollider along that path.

($S$ blocks every backdoor path from $E$ to $D$.)

# d-sep formulation

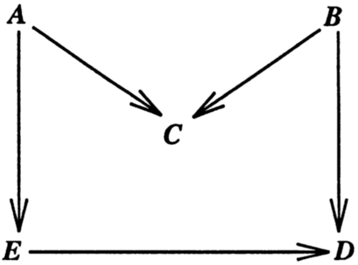
$S$ is sufficient for adjustment for effect of $E$ on $D$ if:

- $S$ contains no descendants of $E$ or $D$ and
- $S$ d-separates $E$ from $D$ in the graph obtained by deleting all arrows emanating from $E$

# Some examples

[on board]

# Do not simply adjust for everything!



“M-bias”

# Minimal sufficient sets

There may be multiple sufficient adjustment sets for an effect in a given graph. A set $S$ is called minimally sufficient for adjustment if $S$ sufficient but no proper subset of $S$ is sufficient.

To find minimal set: may begin with sufficient set $S$ and sequentially delete vars until new set fails backdoor test. However, more recently there have been developed much more efficient algorithms for finding minimal sets, for example the ones implemented in `dagitty` in R.

# Satisfying backdoor

If $D$ is not a parent of $E$, then $\mathrm{Pa}(E, \mathcal{G})$ always satisfies the backdoor criterion for the effect of $E$ on $D$. Why?

# Formalizing interventions

One way to distinguish interventional distributions from observational distributions (and formalize causal effects) is using the do($x$) notation:

$$p(y|\,\mathrm{do}(x))$$

“the distribution of $Y$ under an intervention that forces $X$ to value $x$”

$$p(y|x)$$

“the distribution of $Y$ given that $X$ is observed to take value $x$”

$$p(y|\,\mathrm{do}(x)) \neq p(y|x)$$

# Formalizing interventions

Let $\mathcal{M}$ be a NPSEM, i.e., $V_i = f_i(\text{Pa}(V_i, \mathcal{G}), \epsilon_i)\ \forall i$.

We say $p(x, y, z, ...)$ is the distribution *induced* by the SEM.

Let $\mathcal{M}_x$ denote the NPSEM where $X$ is set to $x$ by an intervention, i.e., replace the eq for $X$ in $\mathcal{M}$ by $X = x$ (and leave all else the same).

The post-intervention distribution $p(y, z, ...|\,\text{do}(x))$ is the distribution induced by this "manipulated SEM" $\mathcal{M}_x$.

# Identification

The causal effect of $X = x$ on $Y$ is defined as some functional of the post-intervention distribution $p(y|\,\mathrm{do}(x))$.

Often we focus on estimating a low-dimensional contrast e.g. $\mathbb{E}[Y|\,\mathrm{do}(x)] - \mathbb{E}[Y|\,\mathrm{do}(x')]$ (called "average causal effect") or similar, but more generally may be interested in $p(y|\,\mathrm{do}(x))$.

We say the effect is *identified* if $p(y|\,\mathrm{do}(x))$ can be expressed as some function of the observational distribution $p(y, x, z, ...)$.

# Backdoor criterion

Theorem. If $Z$ satisfies the backdoor criterion, then

$p(y|\,\mathrm{do}(x)) = \sum_z p(y|x,z)p(z)$.

This is sometimes called the "adjustment formula."[3]

[3] Note: replace $\sum_z$ with integral as appropriate for continuous variables, e.g. $\int_z p(y|x,z)p(z)dz$

# Backdoor criterion

To prove this, it helps to formalize interventions with special "regime indicator" nodes. Consider a graph $\mathcal{G} = (V, E)$ and an augmented graph $\mathcal{G}' = (V', E')$ with:
$V' = V \cup F_i$
$E' = E \cup \{F_i \to X_i\}$.

The factorization property for this augmented model is defined to be the same as for $\mathcal{G}$ except:

$$p(x_i | \, \mathrm{Pa}(X_i, \mathcal{G}')) = \begin{cases} p(x_i | \, \mathrm{Pa}(X_i, \mathcal{G})) & \text{if } F_i = \text{“idle”} \\ 1 & \text{if } F_i = \mathrm{do}(x_i') \text{ and } x_i = x_i' \\ 0 & \text{if } F_i = \mathrm{do}(x_i') \text{ and } x_i \neq x_i' \end{cases}$$

# Backdoor criterion

With this intervention-node representation:

$$p(y|\operatorname{do}(x)) = p(y|F_x = \operatorname{do}(x))$$

$$\begin{aligned}
p(y|F_x = \operatorname{do}(x)) &= \sum_z p(y, z|F_x = \operatorname{do}(x)) \\
&= \sum_z p(y|z, F_x = \operatorname{do}(x)) p(z|F_x = \operatorname{do}(x)) \\
&=^1 \sum_z p(y|z, x, F_x = \operatorname{do}(x)) p(z|F_x = \operatorname{do}(x)) \\
&=^2 \sum_z p(y|z, x, F_x = \operatorname{do}(x)) p(z) \\
&=^3 \sum_z p(y|z, x) p(z)
\end{aligned}$$

(1) $F_x = \operatorname{do}(x) \implies X = x$

(2) $F_x \perp\!\!\!\perp \operatorname{Nd}(F_x, \mathcal{G}')|\operatorname{Pa}(F_x, \mathcal{G}')$ implies $F_x \perp\!\!\!\perp Z$

(3) Backdoor condition implies $F_x \perp\!\!\!\perp Y|Z, X$