STATS 571/671
Causal Inference
Winter 2026

Richard Guo

Guest Lecture: Time-varying confounding and treatments

April 9, 2026

Overview

Graphical models

Statistical graphs

A statistical graph is a graph over a vertex set V with one or more types of edges.
It is a compact (and intuitive!) form for encoding and reasoning about dependencies.

1. **Probabilistic graphical models**.

absence of edge between $u, v \implies$ some form of independence between X_u and X_v

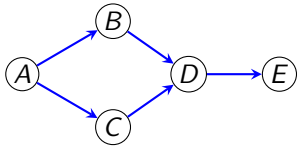
Statistical graphs

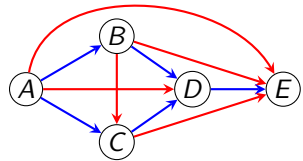
A statistical graph is a graph over a vertex set V with one or more types of edges.
It is a compact (and intuitive!) form for encoding and reasoning about dependencies.

1. **Probabilistic graphical models**.

absence of edge between $u, v \implies$ some form of independence between X_u and X_v

☞ For example, in this Directed Acyclic Graph (DAG)





The missing '$B \to C$' posits

$$P(C \mid A, B) = P(C \mid A) \iff \boxed{B \perp\!\!\!\perp C \mid A}$$

Statistical graphs

A statistical graph is a graph over a vertex set V with one or more types of edges.
It is a compact (and intuitive!) form for encoding and reasoning about dependencies.

1. **Probabilistic graphical models**.

absence of edge between $u, v \implies$ some form of independence between X_u and X_v

- Undirected Graph (UG)
- Directed Acyclic Graph / Bayesian Network (DAG)
- Completed Partially Directed Acyclic Graph / Essential Graph (CPDAG)
- Partial Ancestral Graph (PAG)
- Local Independence Graph

☞ They represent a set of distributions over $(X_v : v \in \mathrm{V})$.

Statistical graphs

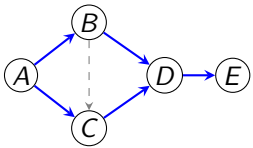
❷ **Causal graphical models**.

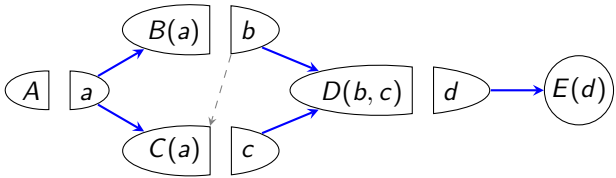
absence of $u \to v \implies$ intervening on X_u has no direct effect on X_v

Statistical graphs

❷ **Causal graphical models**.

absence of $u \to v \implies$ intervening on X_u has no direct effect on X_v





SWIG/FFRCISTG model: The absence of '$B \to C$' encodes that $C(a, b) = C(a)$, i.e., C ignores any input from b.

Statistical graphs

❷ **Causal graphical models**.

absence of $u \to v \implies$ intervening on X_u has no direct effect on X_v

- Directed Acyclic Graph / Bayesian Network (DAG)
- Acyclic Directed Mixed Graph (ADMG)

☞ They represent a set of distributions over all the counterfactuals $X(\cdot)$.

$\hookrightarrow X(\cdot) = (X_v(x_I) : v \in \mathrm{V}, I \subseteq \mathrm{V}, x_I \in \mathcal{X}_I)$

$X(\cdot)$ contains $X(\emptyset)$, the naturally occurring random variables.

Semantics

A DAG can be either a probabilistic graphical model or a causal graphical model, depending on the semantics you use to interpret it.

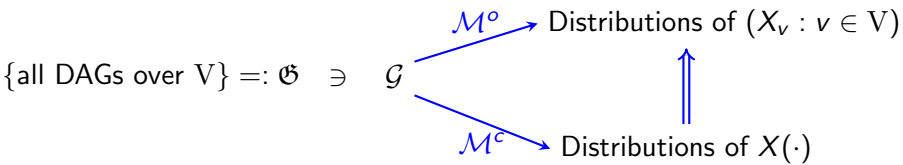
☞ A statistical graph has no intrinsic meaning until you interpret it with certain semantics!

Semantics

A DAG can be either a probabilistic graphical model or a causal graphical model, depending on the semantics you use to interpret it.

☞ A statistical graph has no intrinsic meaning until you interpret it with certain semantics!

- Consider a DAG $\mathcal{G}$ over a finite vertex set V.



- $\mathcal{M}^o(\mathcal{G})$ is a probabilistic model (aka Markov properties)
- $\mathcal{M}^c(\mathcal{G})$ is a causal model
- $\mathcal{M}^c(\mathcal{G}) \implies P(\{X_v(\emptyset) : v \in \mathrm{V}\}) \in \mathcal{M}^o$

☞ causal discovery

DAG as a probabilistic model

Let the probabilistic semantics of a DAG be

$$\mathcal{M}^o(\mathcal{G}) := \{P : p(X) \text{ factorizes according to } \mathcal{G}\}$$
$$= \left\{ P : p(X_v : v \in \mathrm{V}) = \prod_{v \in \mathrm{V}} p(X_v \mid X_{\mathrm{Pa}(v)}) \right\}.$$

Theorem Factorization $\iff$ Global Markov, where

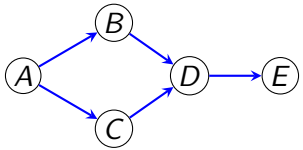
- Global Markov property: For disjoint $A, B, C \subset \mathrm{V}$,

$$A \perp\!\!\!\perp B \mid C \implies X_A \perp\!\!\!\perp X_B \mid X_C \ [P], \quad P \in \mathcal{M}^o(\mathcal{G}).$$

☞ '$A \perp\!\!\!\perp B \mid C$' means C blocks all the d-connecting paths between A and B.

DAG as a probabilistic model: Generating data

We can easily generate data by sequentially drawing from $p(X_v \mid X_{\mathrm{Pa}(v)})$.



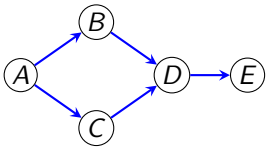
Following the topological ordering $A \prec B \prec C \prec D \prec E$,

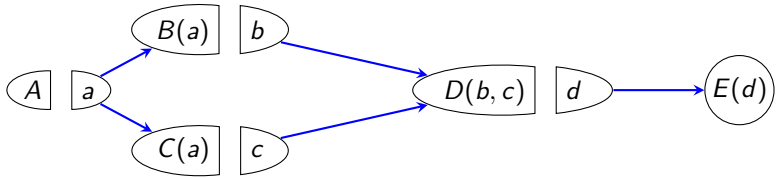
1. Draw $A \sim P(A)$
2. Draw $B \sim P(B \mid A)$, $C \sim P(C \mid A)$
3. Draw $D \sim P(D \mid B, C)$
4. Draw $E \sim P(E \mid D)$

☞ Causal model: an alternative way of generating data.

DAG as a causal model (I): One-step-ahead counterfactuals

- Let $\mathcal{M}^c$ be the FFRCISTG/SWIG causal model (J. Robins, 1986; Richardson & Robins, 2013).





1. Draw $A \sim P(A)$
2. Independent of A, draw ☞ $A \perp\!\!\!\perp B(a), C(a)$ for every a

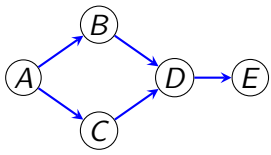
$$B(a) \sim P(B \mid A = a), \quad C(a) \sim P(A \mid A = a), \quad \text{for every } a.$$

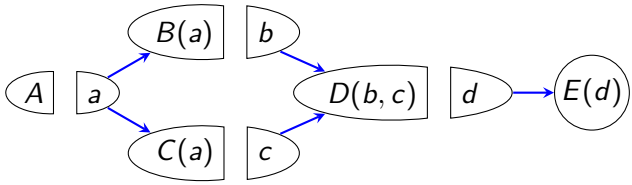
3. Independent of previously drawn, draw

$$D(b, c) \sim P(D \mid B = b, C = c), \quad \text{for every } b, c.$$

4. Independent of previously drawn, draw

DAG as a causal model (II): Recursive substitution





Under no intervention, the observed distribution follows from

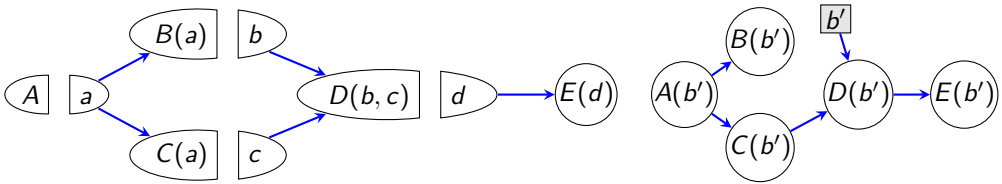
1. $A = A$,
2. $B = B(A)$, $C = C(A)$,
3. $D = D(B, C)$,
4. $E = E(D)$.

- Clearly, $(A, B, C, D, E) \sim P$

DAG as a causal model (III): Intervention

- Consider an intervention that imposes value b' to B. We want to generate counterfactuals

$$A(b'), B(b'), C(b'), D(b'), E(b').$$



1. $A(b') = A$,
2. $B(b') = B(A(b')) = B(A)$ and $C(b') = C(A(b')) = C(A)$
3. $D(b') = D(b', C(b'))$,
4. $E(b') = E(D(b'))$.

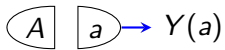
☞ $B(b')$ is the naturally occurring value of B

☞ $P((A, B, C, D, E)(b')) \equiv P(A, B, C, D, E \mid \text{do}(B = b'))$.

Single treatment

Single treatment, randomized

- $A = 1$: treated; $A = 0$: control.



☞ From the SWIG, we can read off

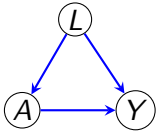
$$A \perp\!\!\!\perp Y(a),$$

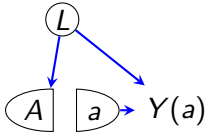
so

$$\boxed{\mathbb{E}\, Y(a) = \mathbb{E}[Y \mid A = a], \quad a = 0, 1.}$$

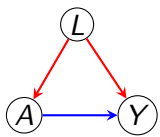
☞ association = causation

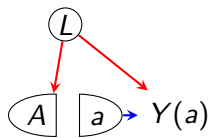
Single treatment, conditionally randomized





Single treatment, conditionally randomized





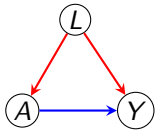
☞ L is an observed confounder between A and Y, so $\mathbb{E}\, Y(a) \neq \mathbb{E}[Y \mid A = a]$.

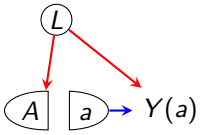
- association $\neq$ causation

There is no unobserved confounding.

☞ We can use L to identify $\mathbb{E}\, Y(a)$ through standardization/adjustment or IPW (inverse probability weighting).

Single treatment, conditionally randomized: Standardization





☞ **Standardization**: From the SWIG, we see that

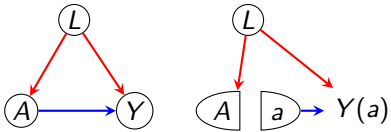
$$A \perp\!\!\!\perp Y(a) \mid L,$$

i.e., A is randomized within every stratum of L.

1. Within stratum $L = l$, we have $\mathbb{E}[Y(a) \mid L = l] = \mathbb{E}[Y \mid A = a, L = l]$.
2. Averaging over l to get the whole population:

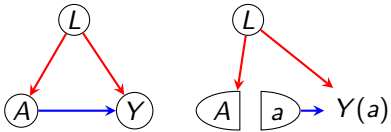
$$\mathbb{E}\, Y(a) = \sum_l \mathbb{E}[Y(a) \mid L = l] P(L = l) = \boxed{\sum_l \mathbb{E}[Y \mid A = a, L = l] P(L = l).}$$

Single treatment, conditionally randomized: IPW

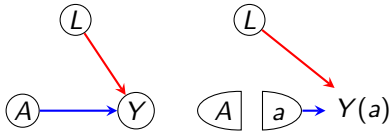


$$p(L = l)p(A = a \mid L = l)p(Y \mid A = a, L = l)$$

Single treatment, conditionally randomized: IPW



$p(L = l)p(A = a \mid L = l)p(Y \mid A = a, L = l)$



$p(L = l)\cancel{p(A = a \mid L = l)}p(Y \mid A = a, L = l)$
$\times q(A = a)/\cancel{p(A = a \mid L = l)}$,
where q is an arbitrary distribution of A.

☞ **IPW**: Weighting by

$1/p(A \mid L)$ ▶ $q(A = 0) = q(A = 1) = 1/2$

or

$p(A)/p(A \mid L)$ ▶ $q(A) = p(A)$, stabilized weight

gives a trial where A is randomized (does not depend on L). ▶ association = causation

Single treatment, conditionally randomized

IPW:

$$\mathbb{E}[Y(a)] = \mathbb{E}\left\{\frac{\mathbb{I}_{A=a}\,Y}{P(A=a\mid L)}\right\} / \mathbb{E}\left\{\frac{\mathbb{I}_{A=a}}{P(A=a\mid L)}\right\} = \mathbb{E}\left\{\frac{\mathbb{I}_{A=a}\,Y}{P(A=a\mid L)}\right\}.$$

Stabilized IPW has the same target

$$\mathbb{E}\left\{\frac{\mathbb{I}_{A=a}\,\cancel{P(A=a)}\,Y}{P(A=a\mid L)}\right\} / \mathbb{E}\left\{\frac{\mathbb{I}_{A=a}\,\cancel{P(A=a)}}{P(A=a\mid L)}\right\} = \mathbb{E}\left\{\frac{\mathbb{I}_{A=a}\,Y}{P(A=a\mid L)}\right\}.$$

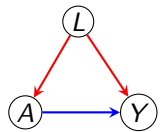
☞ But it makes a difference when fitting marginal structural models with estimated weights...

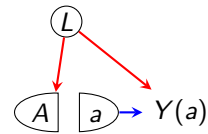
Single treatment, conditionally randomized

- **Standardization** and **IPW** target the same population quantity:

$$\underbrace{\mathbb{E}\left\{\frac{\mathbb{I}_{A=a}\,Y}{P(A=a\mid L)}\right\}}_{\text{IPW}} = \mathbb{E}\left\{\frac{\mathbb{E}[\mathbb{I}_{A=a}\,Y\mid L]}{P(A=a\mid L)}\right\} = \mathbb{E}\left\{\frac{P(A=a\mid L)\,\mathbb{E}[Y\mid A=a,L]}{P(A=a\mid L)}\right\} = \underbrace{\mathbb{E}\{\mathbb{E}[Y\mid A=a,L]\}}_{\text{standardization}}$$

☞ Both standardization and IPW are **adjusting for** L: We use L to block all the non-causal paths (paths not in the shape $A \to \cdots \to Y$).

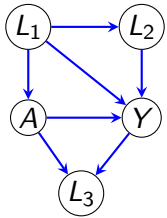




☞ Non-causal ('backdoor') path $A \leftarrow L \rightarrow Y$ is blocked by L (i.e., not d-connected given L).

Quiz

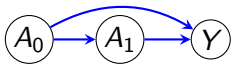
In the setting below, which strategies can correctly identify $\mathbb{E}\, Y(a)$?

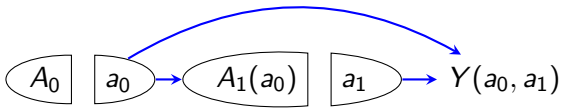


Time-varying treatments

Two treatments, randomized

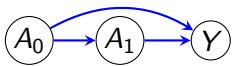
1. Time 0: randomly assign A_0 (1: treated; 0: control)
2. Time 1: randomly assign A_1 (1: treated; 0: control) depending on A_0.
3. Time 2: measure outcome Y

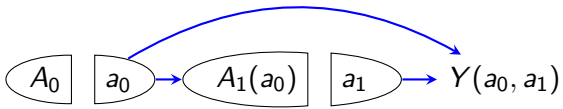




Two treatments, randomized

1. Time 0: randomly assign A_0 (1: treated; 0: control)
2. Time 1: randomly assign A_1 (1: treated; 0: control) depending on A_0.
3. Time 2: measure outcome Y





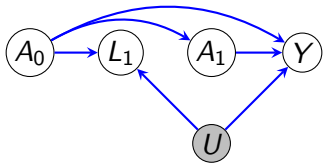
☞ (A_0, A_1) as a whole is randomized, so

$$\mathbb{E}\, Y(a_0, a_1) = \mathbb{E}[Y \mid A_0 = a_0, A_1 = a_1].$$

☞ association = causation

Two treatments: Example on HIV

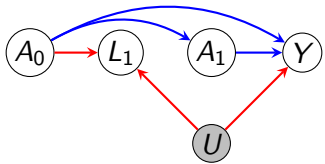
1. Month 0: Assign therapy ($A_0 = 1$: treated; $A_0 = 0$: control) at the start of the follow-up. ☞ Suppose A_0 is randomly assigned.
2. Month 6: Measure blood CD4 counts L_1 and assign therapy ($A_1 = 1$: treated; $A_1 = 0$: control). ☞ Suppose A_1's assignment depends only on A_0 but not L_1.
3. Month 12: Measure the final health score Y.



☞ U represents the unobserved health status that affects both L_1 and Y.
Can we identify $\mathbb{E}\, Y(0,0)$?

Two treatments: Example on HIV

1. Month 0: Assign therapy ($A_0 = 1$: treated; $A_0 = 0$: control) at the start of the follow-up. ☞ Suppose A_0 is randomly assigned.
2. Month 6: Measure blood CD4 counts L_1 and assign therapy ($A_1 = 1$: treated; $A_1 = 0$: control). ☞ Suppose A_1's assignment depends only on A_0 but not L_1.
3. Month 12: Measure the final health score Y.



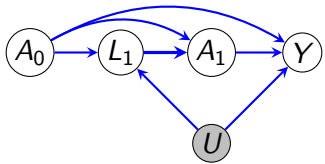
☞ U represents the unobserved health status that affects both L_1 and Y.

Can we identify $\mathbb{E}\, Y(0,0)$?

Yes, non-causal path is not d-connected, unless we condition on L.

Two treatments, with time-varying confounder

1. Month 0: Assign therapy ($A_0 = 1$: treated; $A_0 = 0$: control) at the start of the follow-up. ☞ Suppose A_0 is randomly assigned.
2. Month 6: Measure blood CD4 counts L_1 and assign therapy ($A_1 = 1$: treated; $A_1 = 0$: control). ☞ Suppose A_1's assignment depends on **both A_0 and L_1**.
3. Month 12: Measure the final health score Y.

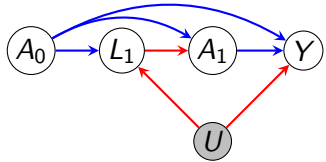


- Can we identify $\mathbb{E}\, Y(a_0, a_1)$? ☞ CD4 count is called a time-varying confounder

Dilemma

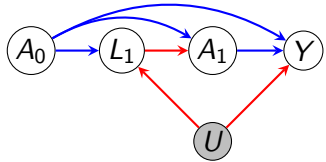
Dilemma

1. Not adjusting for L_1.

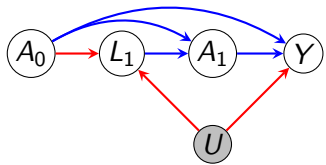


Dilemma

1. Not adjusting for L_1.

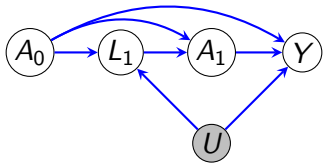


2. Adjusting for L_1.

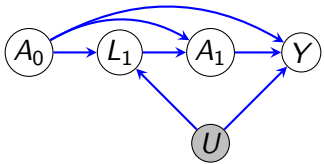


☞ Need something more sophisticated.

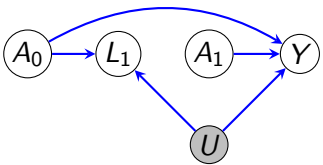
IPW: Removing A_1's dependency on L_1



IPW: Removing A_1's dependency on L_1

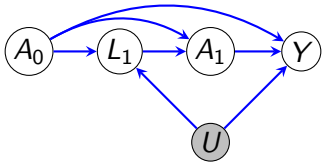


☞ **IPW**

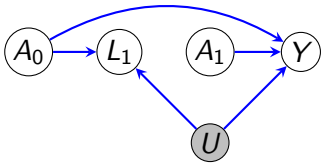


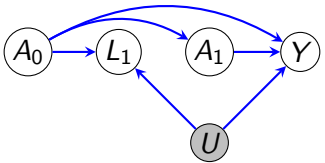
$$\text{weight} = 1/p(A_1 \mid A_0, L_1)$$

IPW: Removing A_1's dependency on L_1



☞ **IPW**





weight $= 1/p(A_1 \mid A_0, L_1)$

stabilized weight $= p(A_1 \mid A_0)/p(A_1 \mid A_0, L_1)$

☞ After reweighting, $\mathbb{E}\, Y(a_0, a_1) = \mathbb{E}[Y \mid A_0 = a_0, A_1 = a_1]$.

IPW: Identification

$$\mathbb{E}\, Y(a_0, a_1) = \mathbb{E}\left\{ \frac{Y\, \mathbb{I}_{A_0=a_0, A_1=a_1}}{P(A_1 = a_1 \mid A_0 = a_0, L_1)} \right\} / \mathbb{E}\left\{ \frac{\mathbb{I}_{A_0=a_0, A_1=a_1}}{P(A_1 = a_1 \mid A_0 = a_0, L_1)} \right\}.$$

☞ Does it make a difference to use the stabilized weight $P(A_1 = a_1 \mid A_0 = a_0)/P(A_1 = a_1 \mid A_0 = a_0, L_1)$?

Quiz

Suppose A_0, L_1, A_1 are all binary. What is the estimate of $\mathbb{E}\, Y(0,0)$ based on IPW?

n	A_0	L_1	A_1	$\mathbb{E}[Y \mid A_0, L_1, A_1]$
50	0	0	0	4
100	0	0	1	-1
100	0	1	0	2
50	0	1	1	-3
50	1	0	0	4
50	1	0	1	-2
100	1	1	0	7
200	1	1	1	11

☞ The first row means there are 50 subjects with $A_0 = 0, L_1 = 0, A_1 = 0$ and their average outcome is 4.

g-formula

The identification formula is called a g-formula

$$\mathbb{E}\, Y(a_0, a_1) = \mathbb{E}\left\{ \frac{Y\, \mathbb{I}_{A_0=a_0, A_1=a_1}}{P(A_1 = a_1 \mid A_0 = a_0, L_1)} \right\} / \mathbb{E}\left\{ \frac{\mathbb{I}_{A_0=a_0, A_1=a_1}}{P(A_1 = a_1 \mid A_0 = a_0, L_1)} \right\},$$

where we first observe

$$\begin{aligned} \mathbb{E}\left\{ \frac{\mathbb{I}_{A_0=a_0, A_1=a_1}}{P(A_1 = a_1 \mid A_0 = a_0, L_1)} \right\} &= \mathbb{E}\left\{ \frac{P(A_1 = a_1, A_0 = a_0 \mid L_1)}{P(A_1 = a_1 \mid A_0 = a_0, L_1)} \right\} \\ &= \mathbb{E}\, P(A_0 = a_0 \mid L_1) = P(A_0 = a_0). \end{aligned}$$

g-formula

Then, the formula can be rewritten as

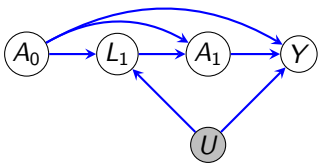
$$
\begin{aligned}
\mathbb{E}\, Y(a_0, a_1) &= \mathbb{E}\left\{ \frac{Y\, \mathbb{I}_{A_0=a_0, A_1=a_1}}{P(A_1 = a_1 \mid A_0 = a_0, L_1)} \right\} / \mathbb{E}\left\{ \frac{\mathbb{I}_{A_0=a_0, A_1=a_1}}{P(A_1 = a_1 \mid A_0 = a_0, L_1)} \right\} \\
&= \mathbb{E}\left\{ \frac{Y\, \mathbb{I}_{A_0=a_0, A_1=a_1}}{P(A_0 = a_0)\, P(A_1 = a_1 \mid A_0 = a_0, L_1)} \right\} \\
&= \mathbb{E}\left\{ \frac{\mathbb{E}[Y\, \mathbb{I}_{A_0=a_0, A_1=a_1} \mid L_1]}{P(A_0 = a_0)\, P(A_1 = a_1 \mid A_0 = a_0, L_1)} \right\} \\
&= \mathbb{E}\left\{ \frac{\mathbb{E}[Y \mid A_0 = a_0, A_1 = a_1, L_1] P(A_1 = a_1, A_0 = a_0 \mid L_1)}{P(A_0 = a_0)\, P(A_1 = a_1 \mid A_0 = a_0, L_1)} \right\} \\
&= \mathbb{E}\left\{ \frac{\mathbb{E}[Y \mid A_0 = a_0, A_1 = a_1, L_1] P(A_0 = a_0 \mid L_1)}{P(A_0 = a_0)} \right\} \\
&= \boxed{\sum_{l_1} \mathbb{E}[Y \mid A_0 = a_0, A_1 = a_1, L_1 = l_1] P(L_1 = l_1 \mid A_0 = a_0),}
\end{aligned}
$$

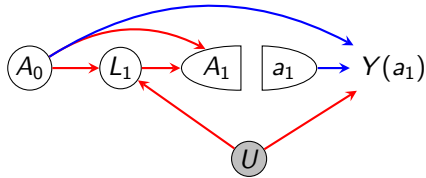
in the form of standardization/adjustment.

g-formula: Intuition

g-formula: Intuition

1. Consider $Y(a_1) := Y(A_0, a_1)$.



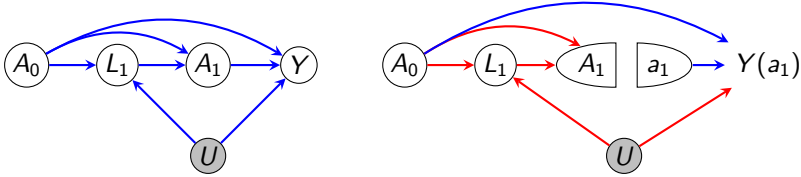


Within the stratum of (A_0, L_1), A_1 is independent of $Y(a_1)$, so (why?)

$$\mathbb{E}[Y(a_1) \mid A_0 = a_0, L_1 = l_1] = \mathbb{E}[Y \mid A_0 = a_0, A_1 = a_1, L_1 = l_1].$$

g-formula: Intuition

1. Consider $Y(a_1) := Y(A_0, a_1)$.



Within the stratum of (A_0, L_1), A_1 is independent of $Y(a_1)$, so (why?)

$$\mathbb{E}[Y(a_1) \mid A_0 = a_0, L_1 = l_1] = \mathbb{E}[Y \mid A_0 = a_0, A_1 = a_1, L_1 = l_1].$$

2. Because A_0 is randomly assigned,

$$\mathbb{E}[Y(a_0, a_1)] = \mathbb{E}[Y(a_1) \mid A_0 = a_0] = \sum_{l_1} \mathbb{E}[Y(a_1) \mid A_0 = a_0, L_1 = l_1] P(L_1 = l_1 \mid A_0 = a_0).$$

Positivity

From the standardization / g-formula

$$\boxed{\mathbb{E}\, Y(a_0, a_1) = \sum_l \mathbb{E}[Y \mid A_1 = a_1, A_0 = a_0, L_1 = l_1] P(L_1 = l_1 \mid A_0 = a_0),}$$

to identify $\mathbb{E}\, Y(a_0, a_1)$, we must have

$$\forall l_1 : P(L_1 = l_1 \mid A_0 = a_0) > 0 \implies \text{data within } (a_0, a_1, l_1),$$

i.e.,

$$\boxed{\forall l_1 : P(L_1 = l_1 \mid A_0 = a_0) > 0 \implies P(A_1 = a_1 \mid A_0 = a_0, L_1 = l_1) > 0.}$$

Quiz

Suppose A_0, L_1, A_1 are all binary. What is the estimate of $\mathbb{E}\, Y(0,0)$ based on standardization / g-formula?

n	A_0	L_1	A_1	$\mathbb{E}[Y \mid A_0, L_1, A_1]$
50	0	0	0	4
100	0	0	1	-1
100	0	1	0	2
50	0	1	1	-3
50	1	0	0	4
50	1	0	1	-2
100	1	1	0	7
200	1	1	1	11

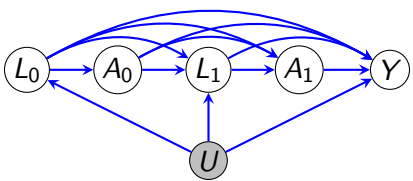
☞ The first row means there are 50 subjects with $A_0 = 0, L_1 = 0, A_1 = 0$ and their average outcome is 4.

Generalization

Let L_0, L_1 be CD4 counts at baseline and time 1. ▶ time-varying confounder

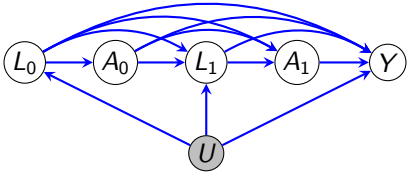
Let A_0, A_1 be antiretroviral therapy at baseline and time 1. ▶ time-varying treatment

1. Month 0: Assign therapy ($A_0 = 1$: treated; $A_0 = 0$: control) at the start of the follow-up. ☞ Suppose $Y(a_0, a_1) \perp\!\!\!\perp A_0 \mid L_0$ for baseline covariates L_0.
2. Month 6: Measure blood CD4 counts L_1 and assign therapy ($A_1 = 1$: treated; $A_1 = 0$: control). ☞ Suppose $Y(a_0, a_1) \perp\!\!\!\perp A_1 \mid L_0, A_0, L_1$.
3. Month 12: Measure the final health score Y.



☞ These conditions are known as sequential ignorability.

Generalization



Under positivity and sequential ignorability

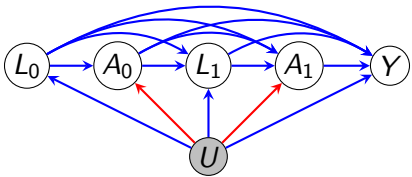
$$Y(a_0, a_1) \perp\!\!\!\perp A_0 \mid L_0,$$
$$Y(a_0, a_1) \perp\!\!\!\perp A_1 \mid L_0, A_0, L_1,$$

$$\mathbb{E}\, Y(a_0, a_1) = \sum_{l_0} \sum_{l_1} \mathbb{E}[Y \mid A_1 = a_1, A_0 = a_0, L_1 = l_1, L_0 = l_0]$$
$$\times P(L_1 = l_1 \mid A_0 = a_0, L_0 = l_0) P(L_0 = l_0).$$

☞ Extends to more time points.

Out of luck

If either red edge is present, then $\mathbb{E}\, Y(a_0, a_1)$ cannot be identified.

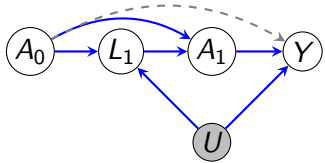


Verma constraint

When there can be latent confounding, a sharp null of no direct effect may not correspond to any independence or conditional independence in the observed distribution.

Verma constraint

When there can be latent confounding, a sharp null of no direct effect may not correspond to any independence or conditional independence in the observed distribution.

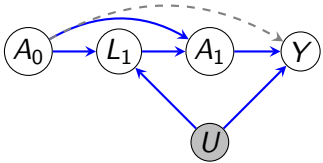


$$H_0 : Y(a_0, a_1) = Y(a_0', a_1) \; \forall a_0, a_0'$$

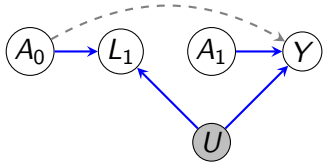
☞ H_0 does not correspond to any marginal or conditional independence on $P(A_0, L_1, A_1, Y)$.

Verma constraint

When there can be latent confounding, a sharp null of no direct effect may not correspond to any independence or conditional independence in the observed distribution.



$H_0 : Y(a_0, a_1) = Y(a_0', a_1)\ \forall a_0, a_0'$



$H_0 : A_0 \perp\!\!\!\perp Y$ under
weight $= 1/p(A_1 \mid A_0, L_1)$

☞ H_0 does not correspond to any marginal or conditional independence on $P(A_0, L_1, A_1, Y)$.

☞ Such a constraint on $P(A_0, L_1, A_1, Y)$ is called a Verma constraint or a generalized conditional independence. Described by the nested Markov model associated with ADMGs.

Marginal structural (mean) model

Consider two treatments A_0, A_1.

☞ A marginal structural (mean) model (J. M. Robins et al., 2000) is to postulate

$$\mathbb{E}[Y(a_0, a_1)] = f(a_0, a_1; \theta).$$

Marginal structural (mean) model

Consider two treatments A_0, A_1.

☞ A marginal structural (mean) model (J. M. Robins et al., 2000) is to postulate

$$\mathbb{E}[Y(a_0, a_1)] = f(a_0, a_1; \theta).$$

For example, when A_0, A_1 are both binary and Y is continous:

- Saturated model

$$\mathbb{E}[Y(a_0, a_1)] = \alpha + \beta_0 a_0 + \beta_1 a_1 + \gamma a_0 a_1$$

- Main effect only

$$\mathbb{E}[Y(a_0, a_1)] = \alpha + \beta_0 a_0 + \beta_1 a_1$$

Fitting model with IPW

If (A_0, A_1) is randomized, we have $\mathbb{E}[Y(a_0, a_1)] = \mathbb{E}[Y \mid A_0 = a_0, A_1 = a_1]$, so the model can be simply fitted with least squares.

☞ Now under time-varying confounding, we can use IPW to reweigh data such that we can treat the data as if A_0, A_1 are randomized.

Fitting model with IPW

If (A_0, A_1) is randomized, we have $\mathbb{E}[Y(a_0, a_1)] = \mathbb{E}[Y \mid A_0 = a_0, A_1 = a_1]$, so the model can be simply fitted with least squares.

☞ Now under time-varying confounding, we can use IPW to reweigh data such that we can treat the data as if A_0, A_1 are randomized.

To fit marginal structural mean model,

1. Estimate the propensity score $\widehat{P}(a_1 \mid a_0, l_1)$ (e.g., with logistic regression)
2. Compute weights $\widehat{w} = 1/\widehat{P}(A_1 \mid A_0, L_1)$ or the stabilized weights

$$\widehat{w}_s = \left(\sum_{l_1} \widehat{P}(A_1 \mid A_0, l_1) \widehat{P}(l_1 \mid A_0) \right) / \widehat{P}(A_1 \mid A_0, l_1).$$

3. Fit least squares using $\widehat{w}$ or $\widehat{w}_s$ as weights.

☞ It makes a difference here.

Further reading

See Chapters 19, 20 and 21 of Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.

References I

- Richardson, T., & Robins, J. M. (2013).Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, *128*(30), 2013.
- Robins, J. (1986).A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, *7*(9-12), 1393–1512.
- Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology.