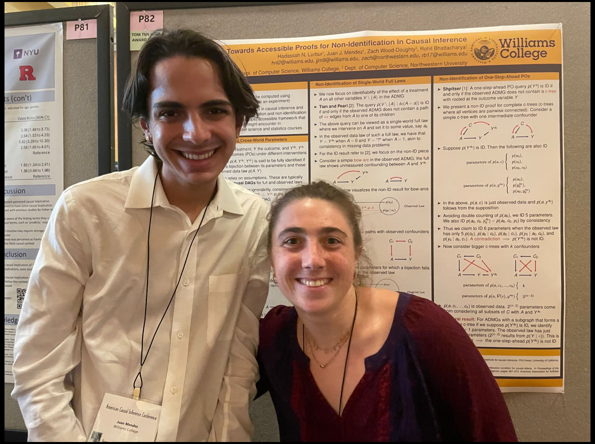
ID, Non-ID, and Parameterizing Causal Graphs with Hidden Variables

Rohit Bhattacharya

Williams College

Collaborators



Juan Mendez + Hadassah Lurbur

Williams College

Zach Wood-Doughty

Northwestern University

RB acknowledges funding from NSF CRII grant 2348287

What I'll cover

- Causal graphs with hidden variables
- Identification and non-identification
- Verma constraints and testing such constraints using observed data

What I’ll cover

non-ID proofs
are underrated

- Causal graphs with hidden variables
- Identification and non-identification
- Verma constraints and testing such constraints using observed data

What I'll cover

non-ID proofs
are underrated

- Causal graphs with hidden variables
- Identification and non-identification
- Verma constraints and testing such constraints using observed data

also understudied, especially in context of estimation

Non-identification proofs are underrated

"D'Amour (2019) raised some concerns about identifiability of the mean potential outcomes using the deconfounder framework in Wang & Blei (2019a) in the static setting and illustrated some pathological examples where identifiability might not hold. In practical settings, the outcome estimates from the Time Series Deconfounder are identifiable …"

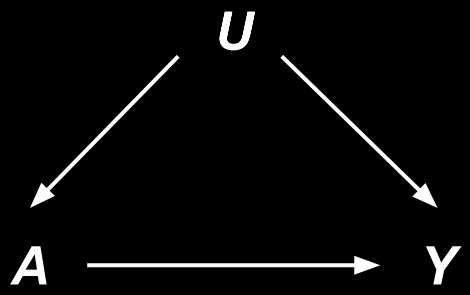
Bica, Alaa, and van der Schaar. *ICML 2020*.

Non-identification proofs are underrated

“D’Amour (2019) raised some concerns about identifiability of the mean potential outcomes using the deconfounder framework in Wang & Blei (2019a) in the static setting and illustrated some pathological examples where identifiability might not hold. In practi om the Time Series Deconfounder are ic

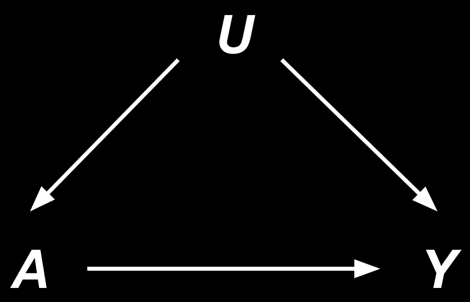
).

True that some examples of non-ID seem "pathological"



- To establish non-ID, sufficient to construct 2 causal models M1 and M2 st
  - $p_1(A, Y) = p_2(A, Y)$
  - $p_1(Y^{a1}) \neq p_2(Y^{a1})$

True that some examples of non-ID seem "pathological"



Structural Equations M1

$U \leftarrow$ fair coin

$A \leftarrow U$

$Y \leftarrow A$ xor U

$$p_1(a, y) = p_2(a, y) \;\forall\; a, y$$

$$p_1(Y^{a1} = 0) = 0.5$$
$$\neq\; p_2(Y^{a1} = 0) = 1$$

Structural Equations M2

$U \leftarrow$ fair coin

$A \leftarrow U$

$Y \leftarrow 0$

Also requires post-hoc smoothing over of positivity issues

Response

- We have modernized identification theory[1] using
    - Modern definitions of causal graphs that mix potential outcomes
- Worth our time to also modernize non-identification theory using
    - Modern definitions above, and
    - Modern parameterizations of causal graphs with hidden variables

1. Richardson and Robins (2013); Malinsky, Shpitser, and Richardson (2019); Shpitser, Richardson, and Robins (2022); Zhao (2025).

Response

- Good news
  - Modernizing is possible; groundwork for doing so already exists
- Bad news
  - Does still require a lot of setup

A ~~simple~~ example without graphs

Observed data law

- A distribution defined over observed random variables

Observed data law

- Two observed random variables A, Y
- Example of an observed data law
- Actually only need to specify 3 params
  - Since $\sum_i p_i = 1$

A	Y	$p(A, Y)$
0	0	p_1
0	1	p_2
1	0	p_3
1	1	p_4

Observed data law

- By chain rule of probability
  - $p(A, Y) = p(A)\, p(Y \mid A)$
- So could also specify the joint in terms

  of a marginal $p(A)$ and conditional $p(Y \mid A)$

Observed data law

- By chain rule of probability
  - $p(A, Y) = p(A)\, p(Y \mid A)$
- So could also specify the joint in terms of a marginal $p(A)$ and conditional $p(Y \mid A)$

A	$p(A)$
0	p_1

Y	A	$p(Y \mid A)$
0	0	p_2
0	1	p_3

Full data law

- Distribution defined over observed variables and potential outcomes

Full data law

- Two observed random variables A, Y
- Two counterfactual random variables aka potential outcomes Y^{a1} and Y^{a0}
- Full data law: $p(Y^{a1}, Y^{a0}, A, Y)$

Identification

- The full data law is said to be identified if there exists a bijection between parameters of $p(Y^{a1}, Y^{a0}, A, Y)$ and the observed data law $p(A, Y)$

Identification

- The full data law is said to be identified if there exists a bijection between parameters of $p(Y^{a1}, Y^{a0}, A, Y)$ and the observed data law $p(A, Y)$

- Without assumptions this is impossible
  - $p(Y^{a1}, Y^{a0}, A, Y)$ has 15 params. $p(A, Y)$ has only 3 params

Identification

- By chain rule of probability
  - $p(Y^{a1}, Y^{a0}, A, Y) = p(Y^{a1}, Y^{a0}) \,.\, p(A \mid Y^{a1}, Y^{a0}) \,.\, p(Y \mid A, Y^{a1}, Y^{a0})$

Identification

$$p(Y^{a1}, Y^{a0}, A, Y) = p(Y^{a1}, Y^{a0}) \,.\, p(A \mid Y^{a1}, Y^{a0}) \,.\, p(Y \mid A, Y^{a1}, Y^{a0})$$

Identification

$$p(Y^{a1}, Y^{a0}, A, Y) = p(Y^{a1}, Y^{a0}) \,.\, p(A \mid Y^{a1}, Y^{a0}) \,.\, p(Y \mid A, Y^{a1}, Y^{a0})$$

Assumption 1: consistency

$$Y = A \,.\, Y^{a1} + (1 - A) \,.\, Y^{a0}$$

Identification

$$p(Y^{a1}, Y^{a0}, A, Y) = p(Y^{a1}, Y^{a0}) \,.\, p(A \,\cancel{|\, Y^{a1}, Y^{a0}}) \,.\, p(Y \,|\, A, Y^{a1}, Y^{a0})$$

Assumption 2: ignorability

$$A \perp Y^{a1}, Y^{a0}$$

Assumption 1: consistency

$$Y = A \,.\, Y^{a1} + (1 - A) \,.\, Y^{a0}$$

Identification

Assumption 3: cross-world independence

$$Y^{a1} \perp Y^{a0}$$

$$p(Y^{a1}, Y^{a0}, A, Y) = p(Y^{a1}) \, . \, p(Y^{a0}) \, . \, p(A \mid \cancel{Y^{a1}, Y^{a0}}) \, . \, p(Y \mid A, Y^{a1}, Y^{a0})$$

Assumption 2: ignorability

$$A \perp Y^{a1}, Y^{a0}$$

Assumption 1: consistency

$$Y = A \, . \, Y^{a1} + (1 - A) \, . \, Y^{a0}$$

Identification

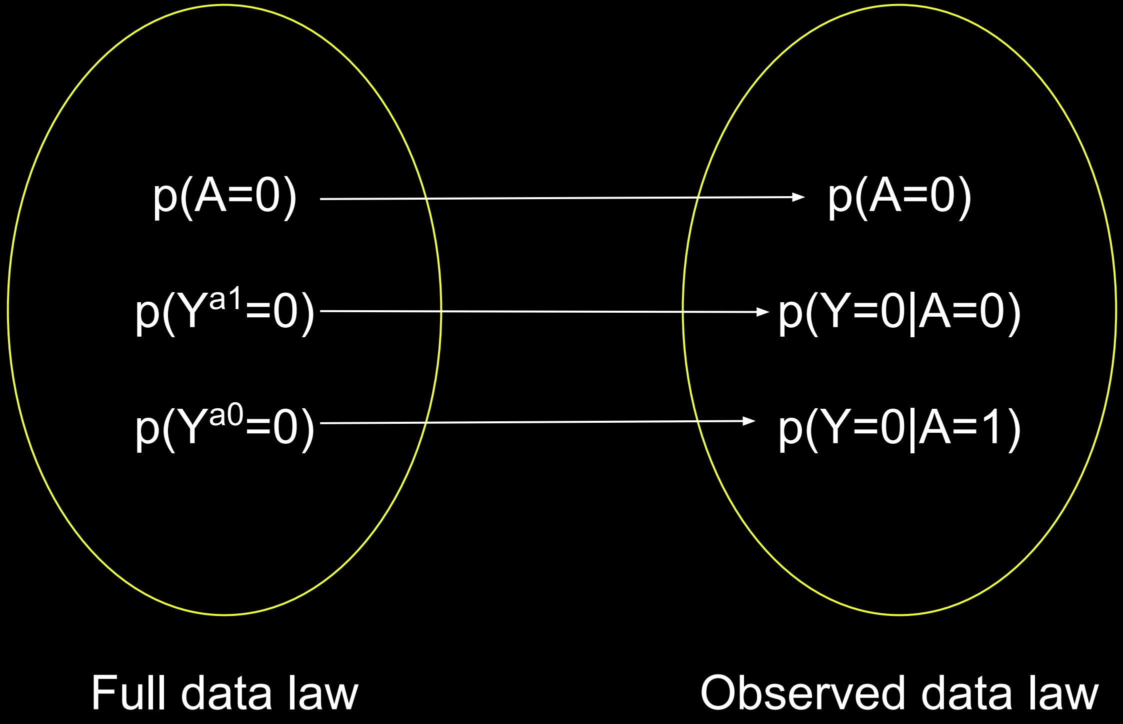
$$p(Y^{a1}, Y^{a0}, A, Y) = p(Y^{a1}) \,.\, p(Y^{a0}) \,.\, p(A) \,.\, p(Y \mid A, Y^{a1}, Y^{a0})$$

Assumption 1: consistency	Assumption 2: ignorability
$Y = A \,.\, Y^{a1} + (1 - A) \,.\, Y^{a0}$	$A \perp Y^{a1}, Y^{a0}$

Assumption 3: cross-world independence

$$Y^{a1} \perp Y^{a0}$$

Identification (bijection) achieved



Re-examining cross-world independence

$$p(Y^{a1}, Y^{a0}, A, Y) = p(Y^{a1}) \,.\, p(Y^{a0}) \,.\, p(A) \,.\, p(Y \mid A, Y^{a1}, Y^{a0})$$

Assumption 1: consistency	Assumption 2: ignorability
$Y = A \,.\, Y^{a1} + (1 - A) \,.\, Y^{a0}$	$A \perp Y^{a1}, Y^{a0}$

Assumption 3: cross-world independence

$$Y^{a1} \perp Y^{a0}$$

Re-examining cross-world independence

$$p(Y^{a1}, Y^{a0}, A, Y) = p(Y^{a1}, Y^{a0}) \, . \, p(A) \, . \, p(Y \mid A, Y^{a1}, Y^{a0})$$

Assumption 1: consistency	Assumption 2: ignorability
$Y = A \, . \, Y^{a1} + (1 - A) \, . \, Y^{a0}$	$A \perp Y^{a1}, Y^{a0}$

Re-examining cross-world independence

$$p(Y^{a1}, Y^{a0}, A, Y) = p(Y^{a1}, Y^{a0}) \,.\, p(A) \,.\, p(Y \mid A, Y^{a1}, Y^{a0})$$

Assumption 1: consistency	Assumption 2: ignorability
$Y = A \,.\, Y^{a1} + (1 - A) \,.\, Y^{a0}$	$A \perp Y^{a1}, Y^{a0}$

- The full data law now requires 4 params, so not ID
- Can use Möbius parameterization to disentangle ID/non-ID pieces

Different ways of parameterizing a distribution *p(X, Y)*

- Directly specify 3 joint probabilities. E.g., $p(x0, y0), p(x0, y1), p(x1, y0)$
- Specify a marginal and 2 conditionals. E.g., $p(x0), p(y1 \mid x0), p(y1 \mid x1)$

- Specify two conditionals $p(x0 \mid y0), p(y0 \mid x0)$, and a "joining term" [1]
- Specify two margins p(x0), p(y0), and a "joining term" [2]

1. Chen (2007). Nabi, Bhattacharya, and Shpitser (2020). Malinsky, Shpitser and Tchetgen Tchetgen (2022).
2. Plackett (1964), Osius (2004), Evans and Richardson (2010), Evans and Didelez (2024).

Binary Möbius parameterization for $p(X, Y)$

- Specify two margins $p(X=0)$ and $p(Y=0)$
- The joining term in this case is simply p(X=0, Y=0)

X	Y	$p(X, Y)$
0	0	$p(X=0, Y=0)$
0	1	$p(X=0) - p(X=0, Y=0)$
1	0	$p(Y=0) - p(X=0, Y=0)$
1	1	$1 - p(X=0) - p(Y=0) + p(X=0, Y=0)$

Binary Möbius parameterization for $p(V_1, \ldots, V_K)$

- Univariate margins $p(V_i = 0)$ for all i in 1, …, K
- Bivariate margins $p(V_i = 0, V_j = 0)$ for all $i \neq j$
- Trivariate margins $p(V_i = 0, V_j = 0, V_k = 0)$ for all $i \neq j \neq k$
- …
- $p(V_1 = 0, \ldots, V_K = 0)$

Binary Möbius parameterization for $p(V_1, \ldots, V_K)$

- Univariate margins $p(V_i = 0)$ for all i in 1, …, K
- Bivariate margins $p(V_i = 0, V_j = 0)$ for all $i \neq j$
- Trivariate margins $p(V_i = 0, V_j = 0, V_k = 0)$ for all $i \neq j \neq k$
- …
- $p(V_1 = 0, \ldots, V_K = 0)$

Some of these can simplify if we have independence constraints. E.g., if $V_i \perp V_j$

Then $p(V_i = 0, V_j = 0) = p(V_i = 0) \,.\, p(V_j = 0)$

Binary Möbius parameterization for $p(V_1, \ldots, V_K)$

- Given Mobius parameters $\mathcal{M}$ for a distribution $p(V)$ with no constraints
- Let $\mathcal{M}[C]$ denote the specific Möbius parameter for the margin $C \subseteq V$.

  Trivially, we will let $\mathcal{M}[\varnothing] = 1$
- Then to get $p(V = v)$ for some specific values v, let $Z \subseteq V$ denote

  variables whose values are 0. Then by Möbius inversion,

$$p(V=v) = \sum_{\{C \,:\, Z \subseteq C \subseteq V\}} (-1)^{|C \setminus Z|} \mathcal{M}[C]$$

Quick sanity check for *p(X, Y)*

$$p(V{=}v) = \sum_{\{ C \,:\, Z \subseteq C \subseteq V \}} (-1)^{|C \setminus Z|} \mathcal{M}[C]$$

$\mathcal{M}[\varnothing] = 1$

$\mathcal{M}[\ \{X\}\ ] = p(X{=}0)$

$\mathcal{M}[\ \{Y\}\ ] = p(Y{=}0)$

$\mathcal{M}[\ \{X,\ Y\}\ ] = p(X{=}0,\ Y{=}0)$

X	*Y*	*p(X, Y)*
0	0	
0	1	
1	0	
1	1	

Quick sanity check for $p(X, Y)$

$$p(V{=}v) = \sum_{\{ C \,:\, Z \subseteq C \subseteq V \}} (-1)^{|C \setminus Z|} \mathcal{M}[C]$$

$$\mathcal{M}[\varnothing] = 1$$

$$\mathcal{M}[\ \{X\}\ ] = p(X{=}0)$$

$$\mathcal{M}[\ \{Y\}\ ] = p(Y{=}0)$$

$$\mathcal{M}[\ \{X, Y\}\ ] = p(X{=}0, Y{=}0)$$

X	Y	$p(X, Y)$
0	0	$p(X{=}0, Y{=}0)$
0	1	$p(X{=}0) - p(X{=}0, Y{=}0)$
1	0	$p(Y{=}0) - p(X{=}0, Y{=}0)$
1	1	$1 - p(X{=}0) - p(Y{=}0) + p(X{=}0, Y{=}0)$

Back to our problem

$$p(Y^{a1}, Y^{a0}, A, Y) = p(Y^{a1}, Y^{a0}) \,.\, p(A) \,.\, p(Y \mid A, Y^{a1}, Y^{a0})$$

Assumption 1: consistency	Assumption 2: ignorability
$Y = A \,.\, Y^{a1} + (1 - A) \,.\, Y^{a0}$	$A \perp Y^{a1}, Y^{a0}$

- The full data law now requires 4 params, so not ID
- Can use Möbius parameterization to disentangle ID/non-ID pieces

Back to our problem

Parameterize this piece using Möbius

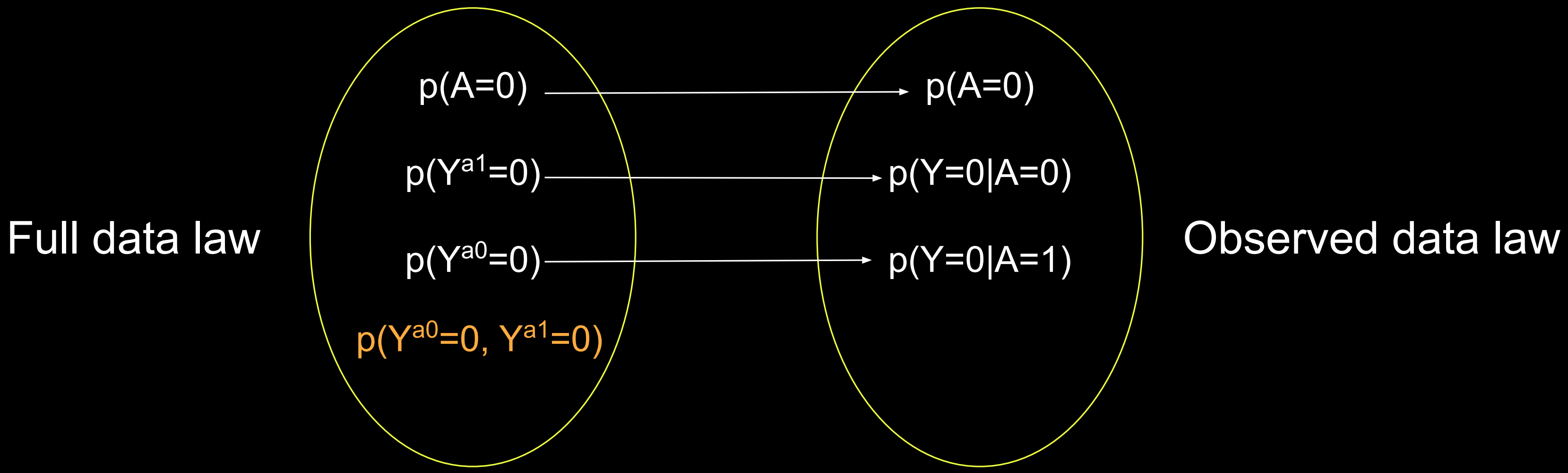
$$p(Y^{a1}, Y^{a0}, A, Y) = p(Y^{a1}, Y^{a0}) \,.\, p(A) \,.\, p(Y \mid A, Y^{a1}, Y^{a0})$$

Assumption 1: consistency	Assumption 2: ignorability
$Y = A \,.\, Y^{a1} + (1 - A) \,.\, Y^{a0}$	$A \perp Y^{a1}, Y^{a0}$

- The full data law now requires 4 params, so not ID
- Can use Möbius parameterization to disentangle ID/non-ID pieces

Non-ID of cross-world parameter

Assumption 1: consistency	Assumption 2: ignorability
$Y = A \,.\, Y^{a1} + (1 - A) \,.\, Y^{a0}$	$A \perp Y^{a1}, Y^{a0}$



Now can construct as many counterexamples as we like

- Specify the identified params $p(Y^{a0}=0)$ and $p(Y^{a1}=0)$. Say 0.3 and 0.4
- Then choose *any* value of $p(Y^{a1}=0, Y^{a0}=0)$ that gives a valid distribution

(all numbers below are [0, 1] and sum to 1. E.g., try with 0.2 and 0.1

Y^{a0}	Y^{a1}	$p(Y^{a0}, Y^{a1})$
0	0	$p(Y^{a1}=0, Y^{a0}=0)$
0	1	$p(Y^{a0}=0) - p(Y^{a1}=0, Y^{a0}=0)$
1	0	$p(Y^{a1}=0) - p(Y^{a1}=0, Y^{a0}=0)$
1	1	$1 - p(Y^{a0}=0) - p(Y^{a1}=0) + p(Y^{a1}=0, Y^{a0}=0)$

Moving onto causal graphs

Will now discuss ID and non-ID of single-world parameters

Statistical DAG models

- DAG: a graph with directed edges (→) s.t. there are no directed cycles
- A statistical model of a DAG G is the set of distributions that factorize as
  - $p(V_1, \ldots, V_K) = \prod_{k=1 \ldots K} p(\, V_k \mid pa_G(V_k))$
- Equivalently, distributions that satisfy the global Markov property wrt G
  - For disjoint subsets X, Y, Z we have $(X \perp Y \mid Z)_{\text{d-sep}} \Rightarrow (X \perp Y \mid Z)_p$

Pearl (1988)

Statistical ADMG models

- ADMG: a graph with directed ($\rightarrow$) and bidirected ($\leftrightarrow$) edges s.t. there are no directed cycles, and at most one directed edge and one bidirected edge between each pair of vertices

- A statistical model of an ADMG G is the set of distributions that satisfy the nested Markov factorization wrt G

Richardson, Evans, Robins, Shpitser (2023)

Statistical ADMG models

- Nested Markov factorization of a distribution *p(V)* wrt to an ADMG G

  implies the following ordinary global Markov property

  - For disjoint subsets X, Y, Z we have $(X \perp Y \mid Z)_{\text{m-sep}} \Rightarrow (X \perp Y \mid Z)_p$
  - Other equality restrictions also exist (covered in part 2)

- m-sep (mixed-separation), generalizes d-separation in a straightforward

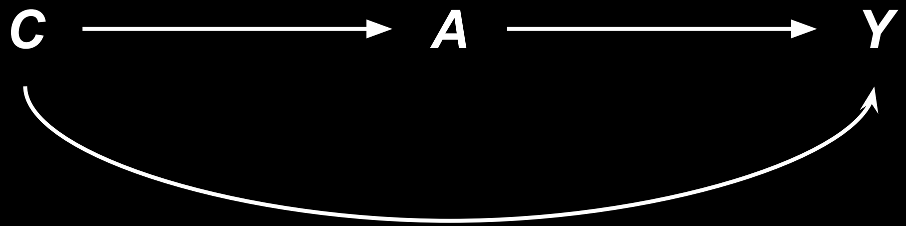
  manner by considering → o ↔, ↔ o ←, and ↔ o ↔ to also be colliders

Richardson (2003); Richardson, Evans, Robins, Shpitser (2023)

Causal models of a DAG

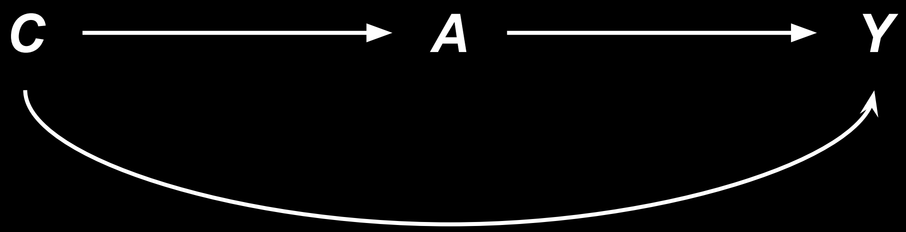
1. Start with a DAG G
2. Define one-step-ahead potential outcomes
3. Pose an independence model for the one-step-aheads
   - Typically NPSEM-IE (Pearl 2009) or FFRCISTG (Robins 1986)
4. Other potential outcomes and links to observed data are defined via recursive substitution + consistency

1. Start with a DAG G



2. Define one-step-ahead potential outcomes

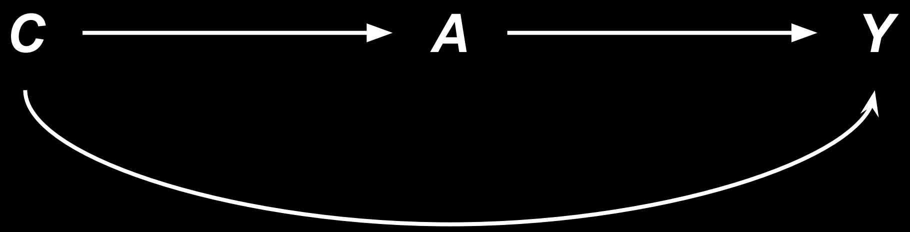
- For each V_i define potential outcomes of the form $V_i^{pa(Vi)}$
- Describes what would happen to V_i if we intervened on its parents



2. Define one-step-ahead potential outcomes

$Y^{c1,a0}$ $Y^{c1,a1}$

A^{c0} A^{c1} $Y^{c0,a0}$ $Y^{c0,a1}$



3. Pose an independence model on one-step-aheads

- NPSEM-IE (Pearl 2009)
  - $V_i^{pa(Vi)} \perp V_j^{pa(Vj)}$ for all i ≠ j and all possible values of $pa(V_i)$, $pa(V_j)$
- FFRCISTG (Robins 1986)
  - $V_i^{pa(Vi)} \perp V_j^{pa(Vj)}$ for all i ≠ j but only when $pa(V_i)$, $pa(V_j)$ agree on value assignments to all variables — no cross-world independences

3. Pose an independence model on one-step-aheads

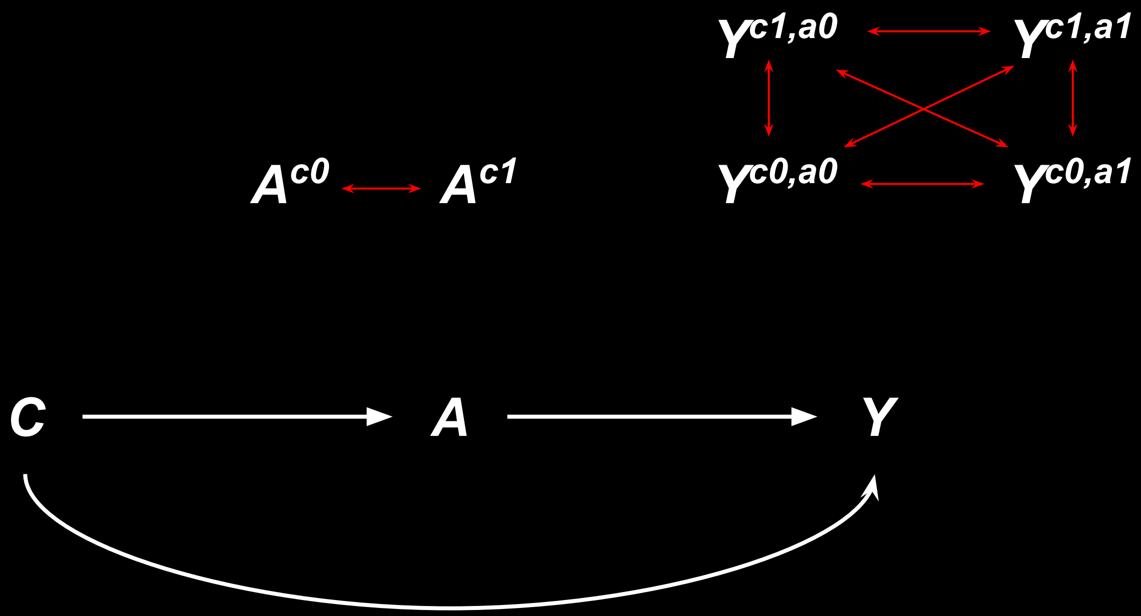
- NPSEM-IE (Pearl 2009)
  - $V_i^{pa(Vi)} \perp V_j^{pa(Vj)}$ for all i ≠ j and all possible values of $pa(V_i)$, $pa(V_j)$

Notice these are all marginal independences.

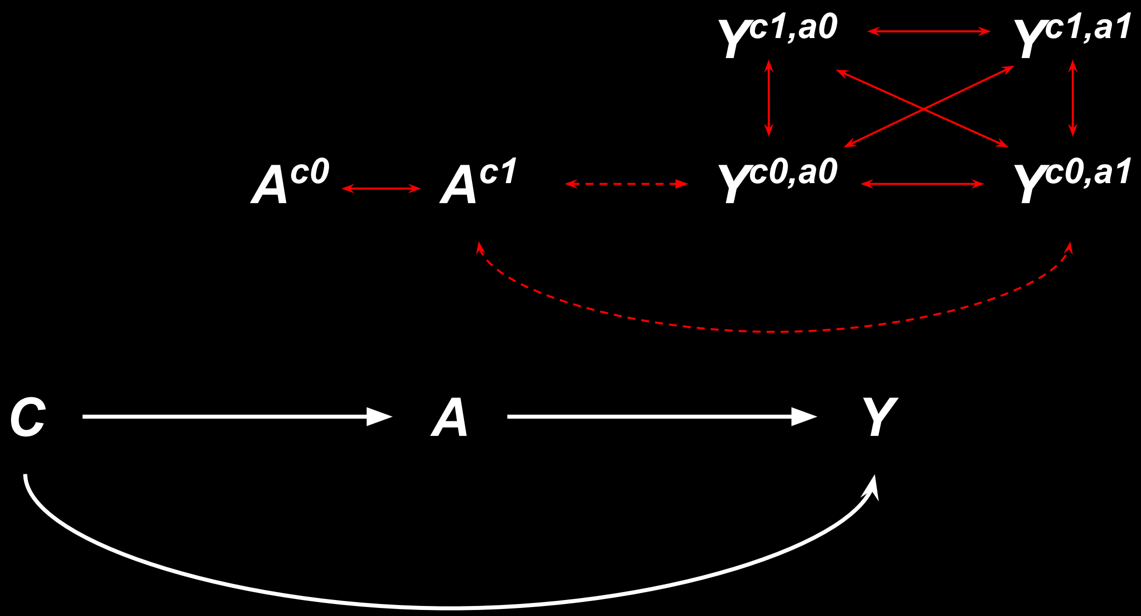
Marginal independence models can be faithfully represented using bidirected graphs

- FFRCISTG (Robins 1986)
  - $V_i^{pa(Vi)} \perp V_j^{pa(Vj)}$ for all i ≠ j but only when $pa(V_i)$, $pa(V_j)$ agree on value assignments to all variables — no cross-world independences

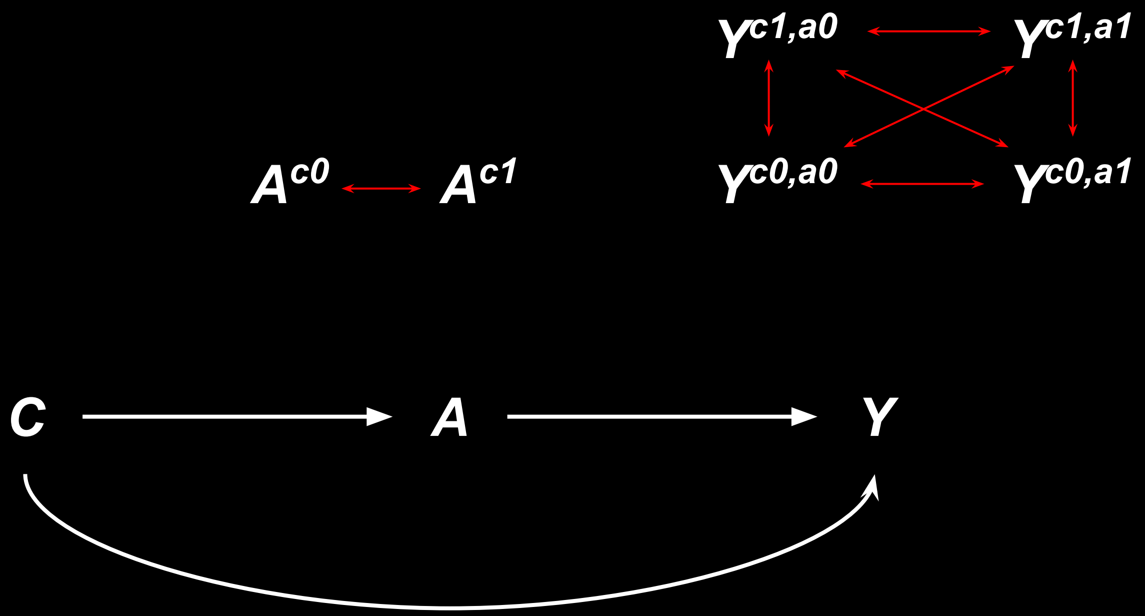
4. NPSEM-IE independences on one-step-aheads



4. Examples of additional edges in FFRCISTG



4. Will use NPSEM-IE for remainder of talk



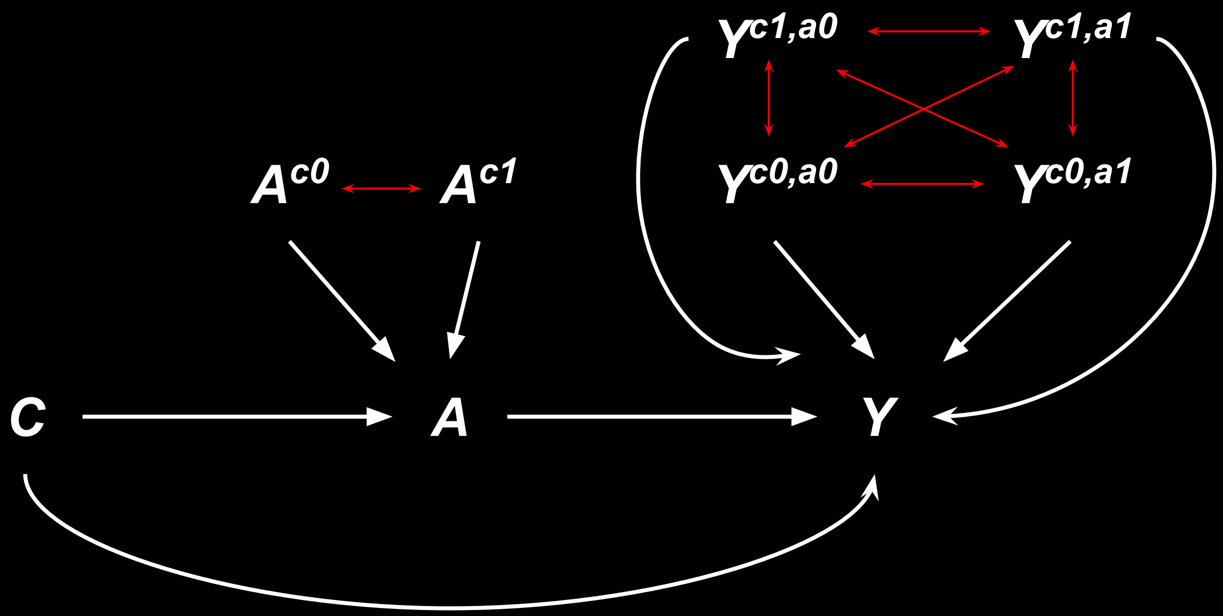
5. Defining other counterfactuals/linking to observed data

- For some subset $A \subset V$, we'd like to know what V_i^a is when $A \neq pa(V_i)$
- Recursive substitution. Code

$$V_i^a \equiv V_i^{\left\{a_j \text{ if } V_j \in A \text{ else } V_j^a \text{ for each } V_j \in \text{Pa}_{\mathcal{G}}(V_i))\right\}}$$

- Implies consistency and causal irrelevance

5. Defining other counterfactuals/linking to observed data



Latent projection

- Suppose we want to study the margin $O \subseteq V$ of an ADMG G
- Rest of the variables $L = V \setminus O$ are considered latent
- The margin $p(O) = \Sigma_L\, p(V)$ also nested Markov factorizes according to an ADMG G with vertices O obtained by the latent projection operator

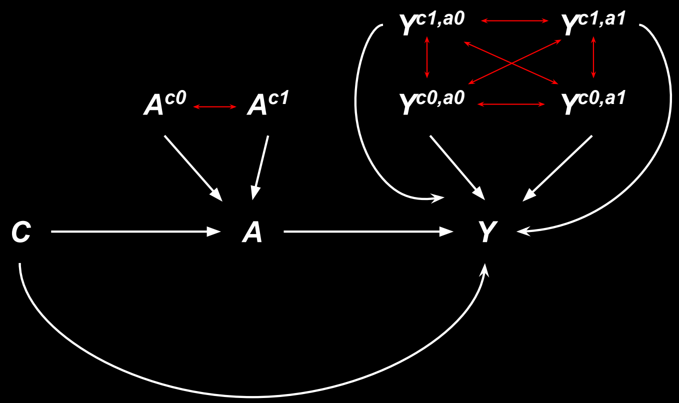
Verma and Pearl (1990); Richardson, Evans, Robins, Shpitser (2023)

Latent projection

- Given ADMG $G(O \cup L)$, the latent projection ADMG G(O) is obtained as,
  - $O_i \rightarrow O_j$ in G(O) iff there exists a directed path $O_i \rightarrow \ldots \rightarrow O_j$ in $G(O \cup L)$ where … are all variables in L
  - $O_i \leftrightarrow O_j$ in G(O) iff there exists a path $O_i \leftarrow o \ldots o \rightarrow O_j$ in G where … are all variables in L and there are no colliders

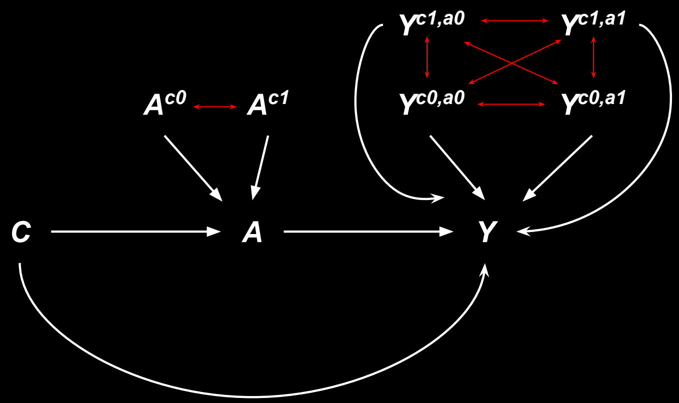
Verma and Pearl (1990); Richardson, Evans, Robins, Shpitser (2023)

Margins of the full data law

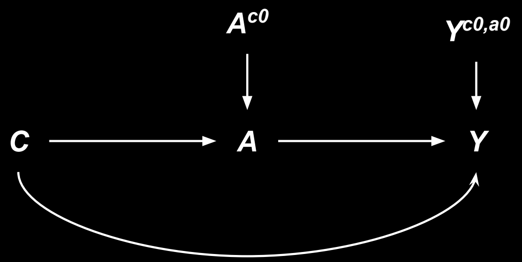


Full data law

Margins of the full data law

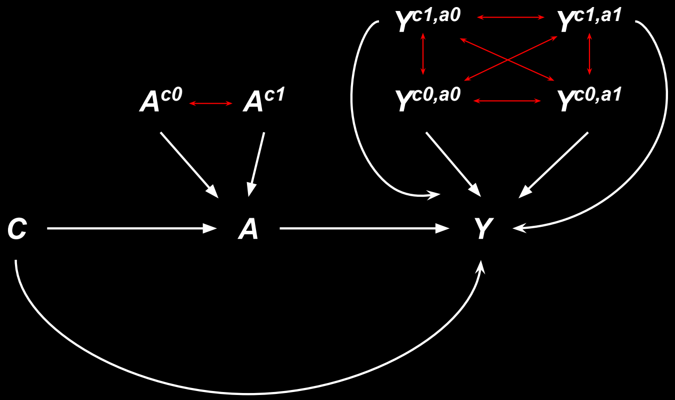


Full data law

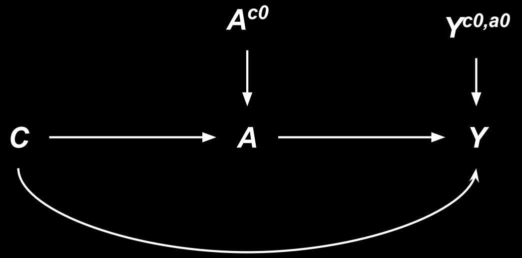


Single-world margin

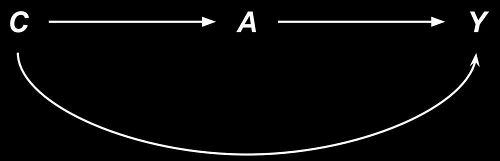
Margins of the full data law



Full data law

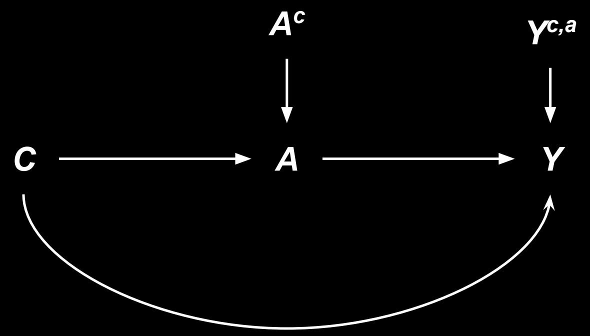


Single-world margin

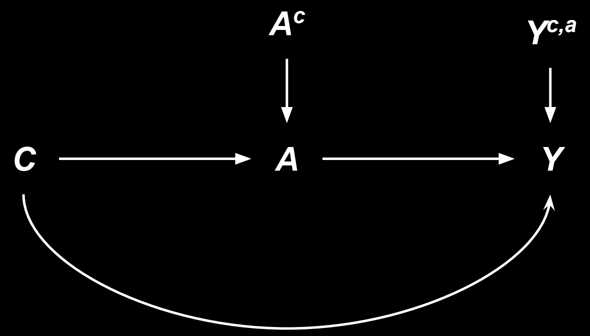


Observed data law

Single-world template for full law identification + g-formula



Single-world template for full law identification + g-formula



$p(C = 0) = p(C = 0)$

$p(A^c = 0) = p(A = 0 \mid C = c)$

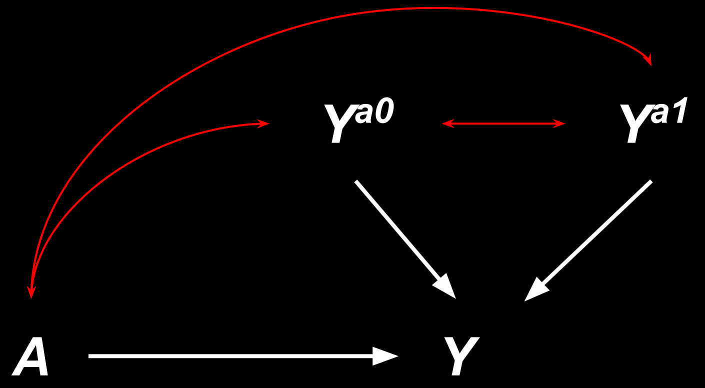
$p(Y^{c,\,a} = 0) = p(Y = 0 \mid A = a, C = c)$

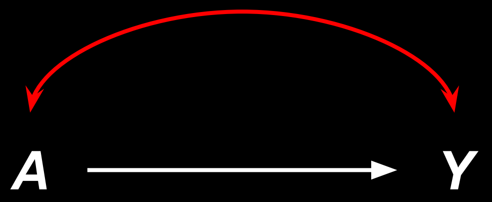
All single-world parameters ID

g-formula

$$p(Y^a) = p(Y^{C,a})$$
$$= \sum_c p(C = c,\ Y^{c,a})$$
$$= \sum_c p(C = c).\ p(Y^{c,\,a})$$
$$= \sum_c p(C = c).\ p(Y \mid A{=}a,\ C{=}c)$$

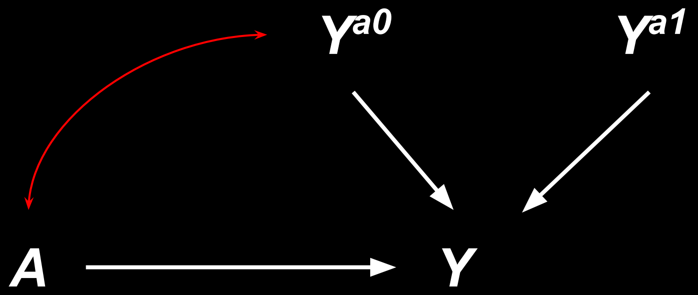
What happens if FFRCISTG does not hold?





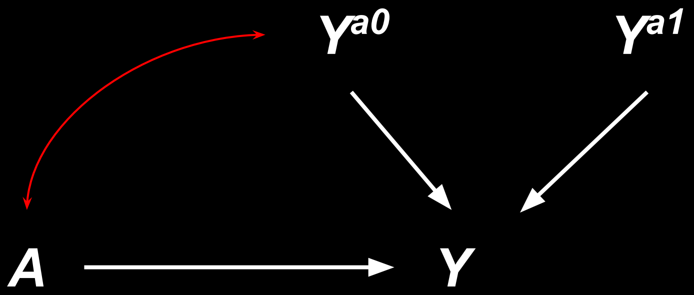
- Observed data law no longer a DAG
- Identification of all single-world parameters *may* not be possible

Non-identification in the “bow-arc” graph

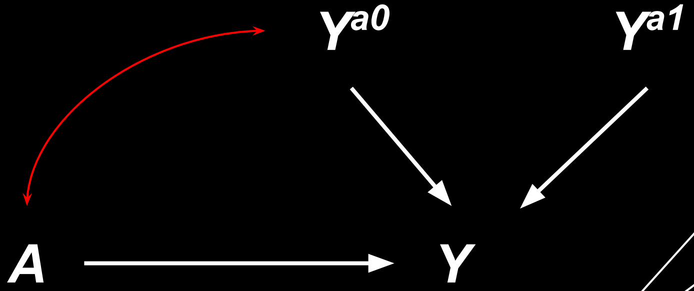


- Sufficient to add just a single edge that violates FFRCISTG assumptions

Non-identification in the "bow-arc" graph



Non-identification in the "bow-arc" graph



$p(Y^{a1} = 0) = p(Y = 0 \mid A = 1)$

$p(A = 0, Y^{a0} = 0) = p(A = 0, Y = 0)$

$p(A = 0) = p(A = 0)$

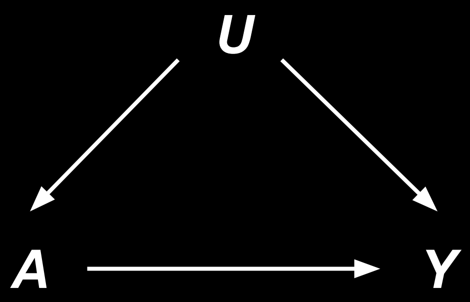
$p(Y^{a0} = 0) = ???$

All single-world parameters not ID

Möbius params of $p(A, Y^{a0})$

Easy to produce counterexamples of non-ID by specifying $p(A = 0)$ and $p(A=0, Y^{a0} = 0)$. Then free to choose any valid value of $p(Y^{a1} = 0)$

Recap: Comparison to older proof



Structural Equations M1

$U \leftarrow$ fair coin

$A \leftarrow U$

$Y \leftarrow A$ xor U

$$p_1(a, y) = p_2(a, y)\ \forall\ a, y$$

$$p_1(Y^{a1} = 0) = 0.5$$
$$\neq\ p_2(Y^{a1} = 0) = 1$$

Structural Equations M2

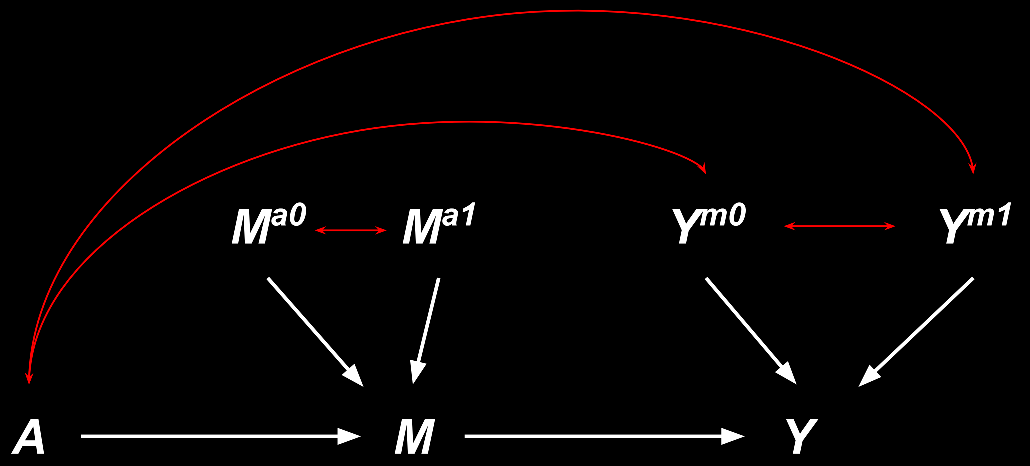
$U \leftarrow$ fair coin

$A \leftarrow U$

$Y \leftarrow 0$

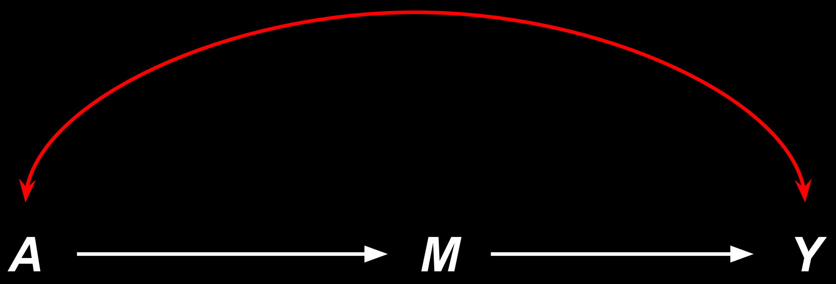
Also requires post-hoc smoothing over of positivity issues

Front-door: Not all unmeasured confounding is non-ID

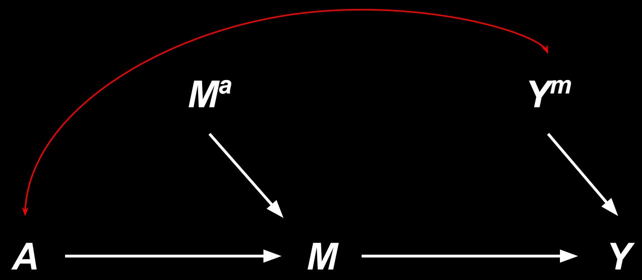


Full data law

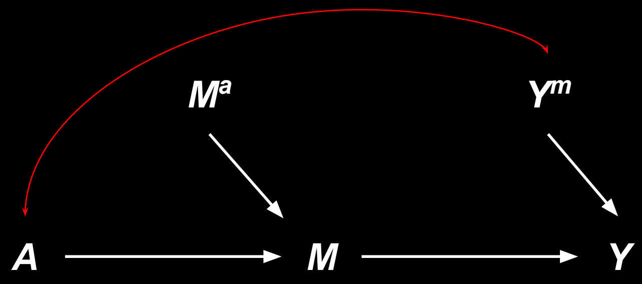
Observed data law



Single-world template and identification of front-door



Single-world template and identification of front-door



$p(M^a = 0) = p(M = 0 \mid A = a)$

$p(A = 0, Y^m = 0) = p(A = 0) \,.\, p(Y \mid A{=}0, M{=}m)$

$p(A = 0) = p(A = 0)$

$p(Y^m = 0) = \sum_a p(A = a) \,.\, p(Y \mid A{=}a, M{=}m)$

All single-world parameters ID

Can combine these to get front-door formula:

$p(Y^a) = \sum_m p(M{=}m \mid A{=}a) \sum_{a'} p(A = a') \,.\, p(Y \mid A{=}a', M{=}m)$

When do we fail to identify single-world parameters?

- Bow-arc graph, but what else?

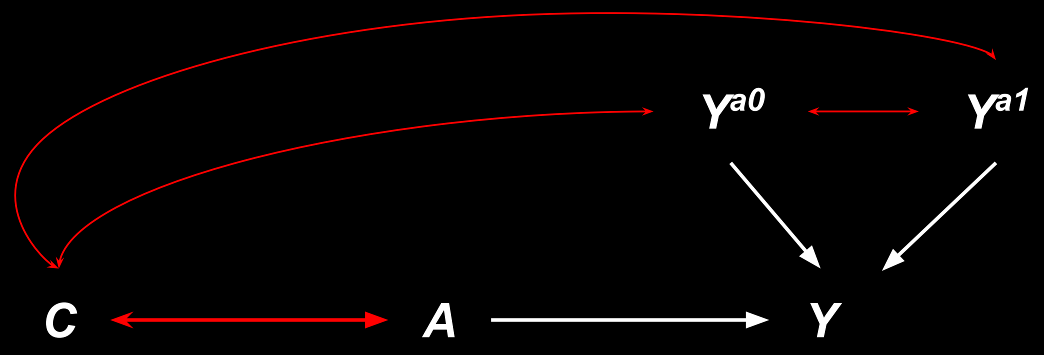
When do we fail to identify single-world parameters?

- Tian and Pearl 2002: The effect of a single treatment *A* on all other variables on the graph $p([\, V \setminus \{A\} \,]^{a})$ is ID if and only if there is no bidirected path from *A* to any of its direct children

- Simple criterion. Covers many popular graphs like the front-door graph
- Estimation under this criterion also well studied[1]

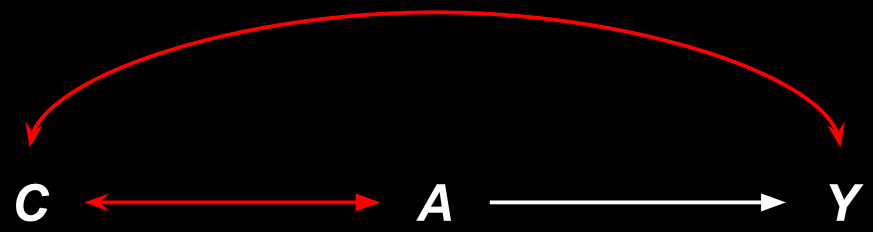
1. Bhattacharya, Nabi, Shpitser (2022); Guo and Nabi (2025); Jung, Tian, Bareinboim (2021)

Non-ID of full law due to a bidirected path to direct child

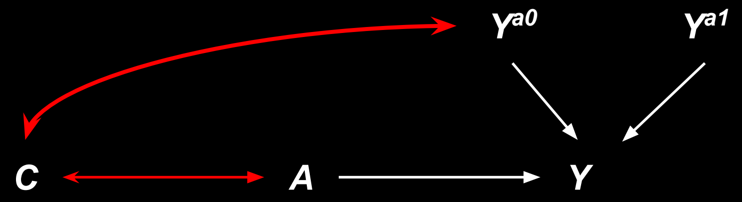


Full data law

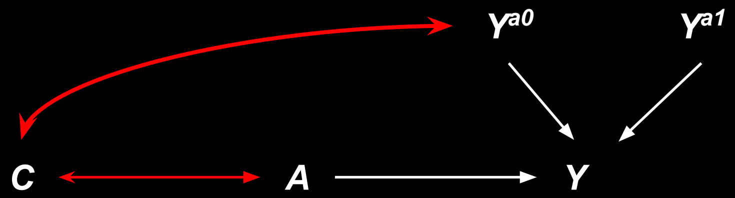
Observed data law



Non-ID of full law due to a bidirected path to direct child



Non-ID of full law due to a bidirected path to direct child

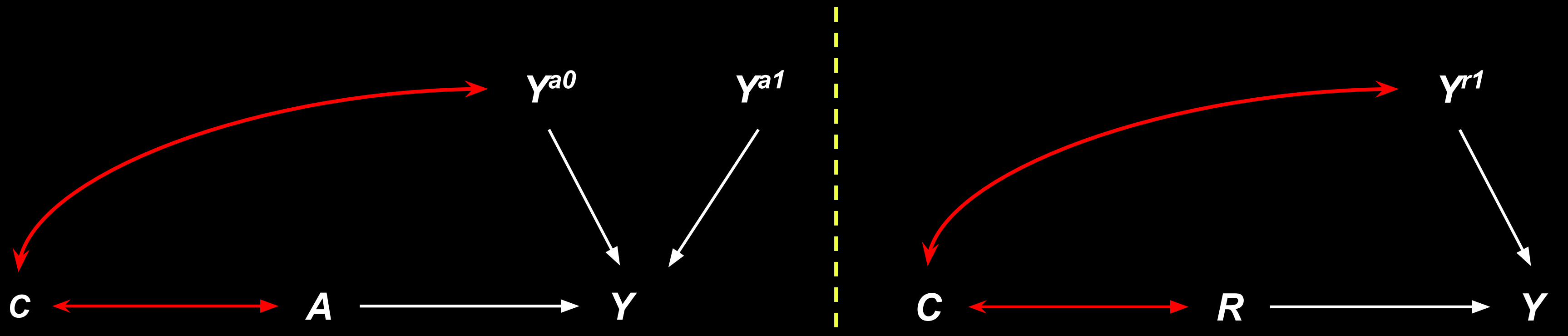


Möbius parameterization of $p(C, A, Y^{a0})$ shows that $p(C=0, Y^{a0}=0)$ is not ID

Caveats to this proof technique

- Must have a parameterization that captures all equality constraints

This unifies a completeness proof in causal and missing data

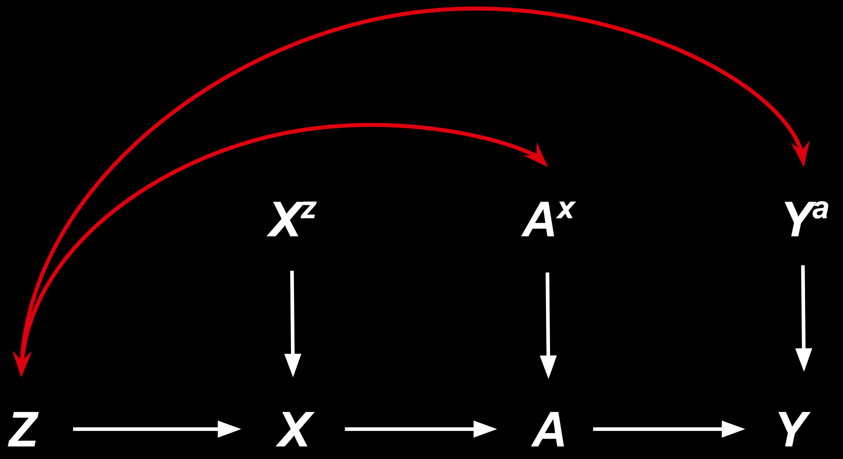


- Same structure leads to non-ID full laws in causal and missing data[1]
- Only difference (in this case) is missing data has only one potential outcome
- However, other problems in missing data seem quite different than causal[2]

1. Bhattacharya, Nabi, Shpitser (2021); Bhattacharya, Srinivasan, Nabi, Ogburn, Shpitser (2024)

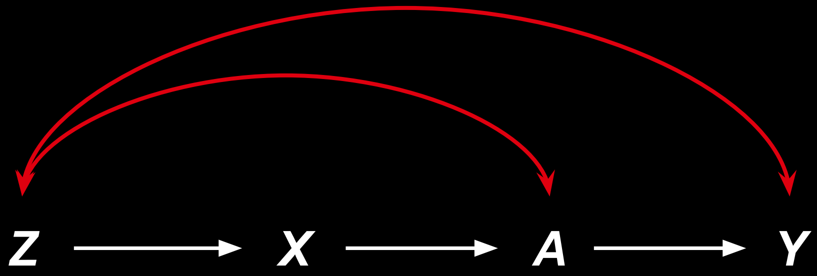
2. Nabi, Bhattacharya, Shpitser, Robins (2025)

Napkin-graph: non-trivial case of full law not ID but p(Y^a) is

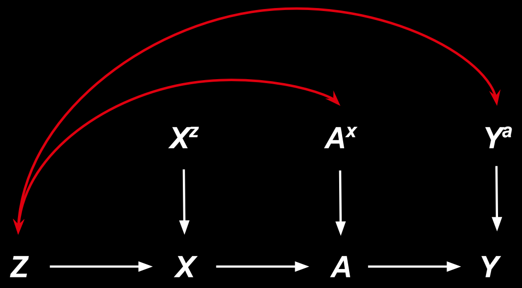


Single-world template

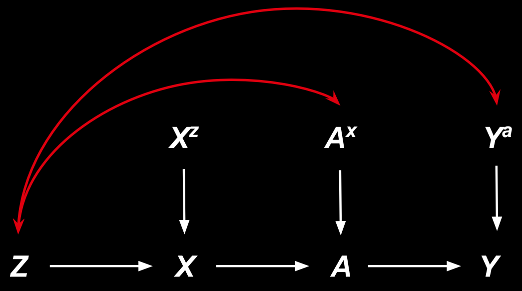
Observed data law



Napkin-graph: non-trivial case of full law not ID but $p(Y^a)$ is



Napkin-graph: non-trivial case of full law not ID but $p(Y^a)$ is



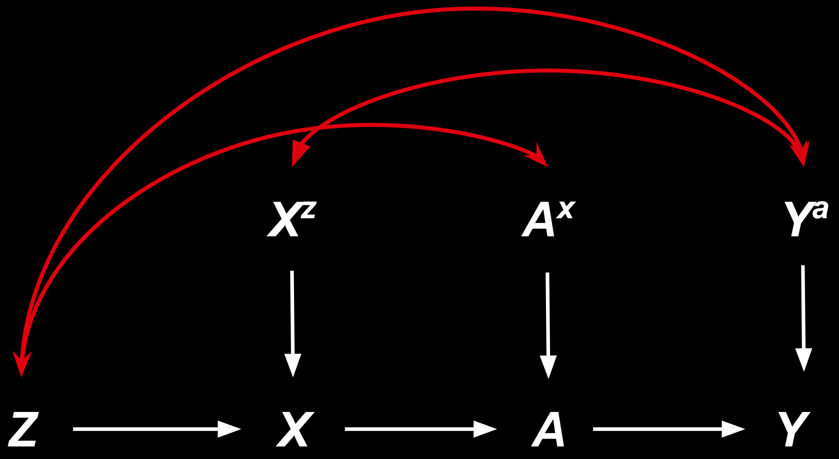
$$p(Y^a) = \left\{ \sum_z p(Z) \,.\, p(A=a, Y \mid Z, X) \right\} \Big/ \left\{ \sum_z p(Z) \,.\, p(A=a \mid Z, X) \right\}$$

When is a one-step-ahead potential outcome not ID?

- Shpitser 2006; Drton, Foygel, Sullivant 2011: A one-step-ahead potential outcome $V_i^{pa(Vi)}$ is ID if and only if there is no c-tree rooted at V_i in G

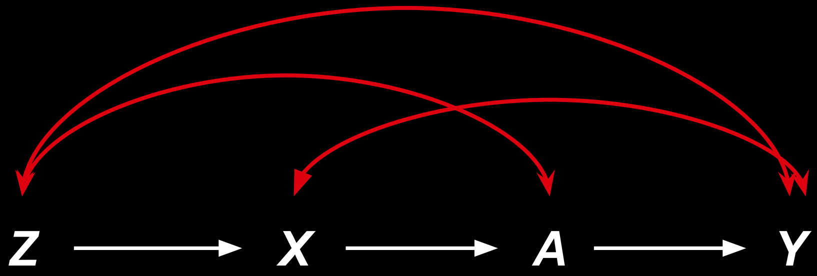
- We have a V_i-rooted c-tree in G, if there exists a subgraph G' of G where
  - G' consists of a single bidirected connected component
  - All vertices in G' have a single directed path to V_i in G'

When is a one-step-ahead potential outcome not ID?

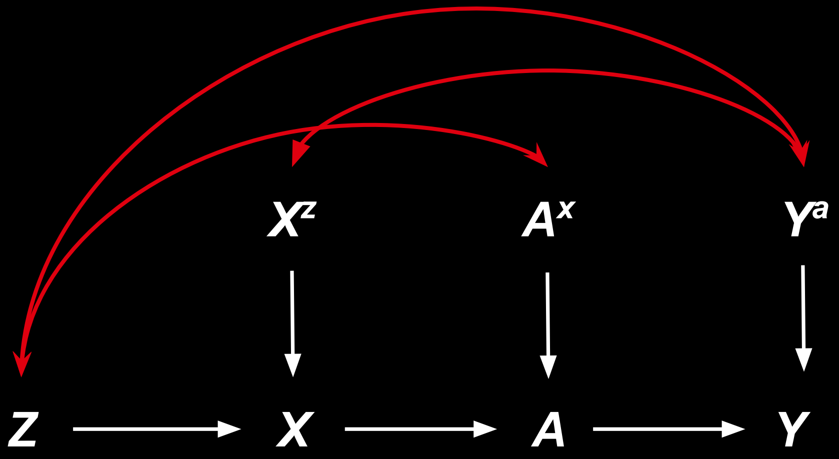


Single-world template

Observed data law



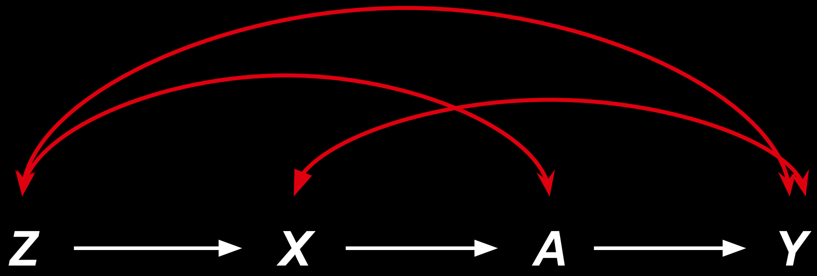
When is a one-step-ahead potential outcome not ID?



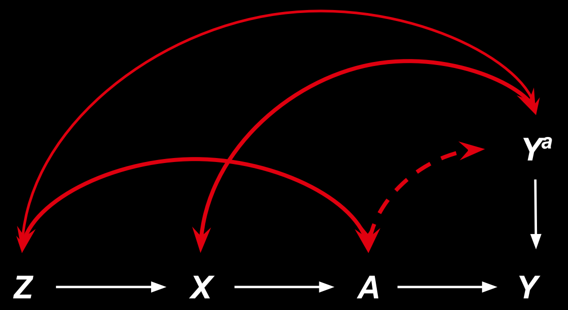
Single-world template

Y^a is no longer ID

Observed data law

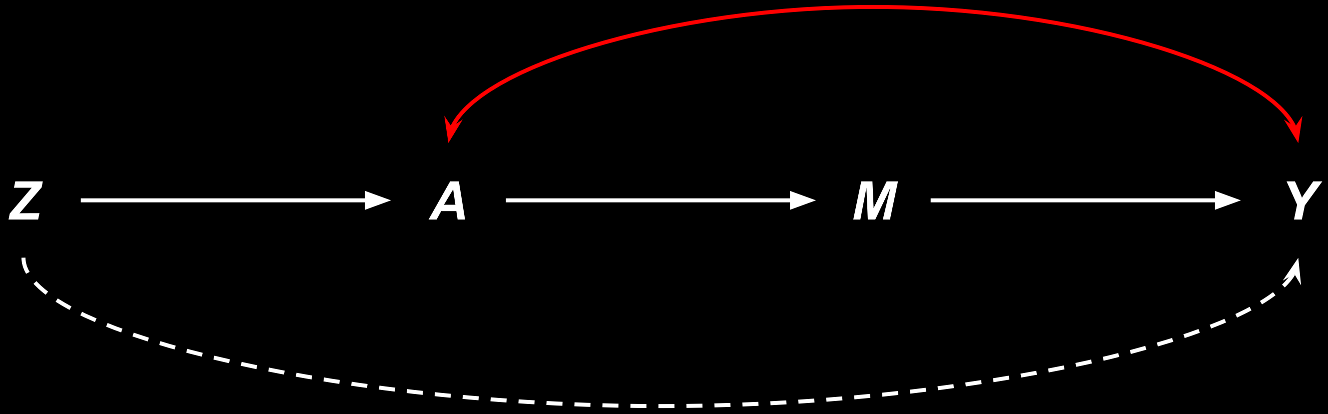


When is a one-step-ahead potential outcome not ID?



Verma constraints

Verma constraints



$p(Y^m) = \sum_a p(A{=}a \mid Z{=}z) \,.\, p(Y \mid Z{=}z, A{=}a, M{=}m)$ is not a function of Z

BIC score for model selection

- Natural parameterizations of p(Z, A, M, Y) lead to the g-null paradox[1]
- If all variables are binary/discrete can use Möbius parameterization of nested Markov model and use the BIC score

1. Robins and Wasserman (1997)

Testing generalized equality constraints

- A more general method is to do an MSM-style weighted conditional independence test[1]
- Code

1. Bhattacharya and Nabi (2021); Robins and Wasserman (1997)

Summary and takeaways

- Mixing potential outcomes and graphs has made identification clearer
- Doing this for non-identification proofs, may make them stronger as well
- There are equality restrictions in hidden variable causal graphs that are not ordinary conditional independences
- Reliable finite-sample testing and using such restrictions in causal effect estimation are open problems

References

- D’Amour 2020. “On Multi-Cause Causal Inference with Unobserved Confounding” *AISTATS.*
- Wang and Blei 2019. “The Blessing of Multiple Causes.” *JASA*.
- Malinsky, Shpitser, and Richardson 2019. “A Potential Outcomes Calculus for Identifying Conditional Path-Specific Effects”. *AISTATS* 2019.
- Richardson and Robins 2013. “Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality.” *arXiv.*
- Shpitser, Richardson, and Robins 2022. “Multivariate Counterfactual Systems And Causal Graphical Models”. *Works of Judea Pearl.*
- Zhao 2025. “On statistical and causal models associated with acyclic directed mixed graphs”. *arXiv*
- Chen 2007. “A semiparametric odds ratio model for measuring association”. *Biometrics*.
- Malinsky, Tchetgen Tchetgen, and Shpitser 2022. “Semiparametric inference for nonmonotone missing-not-at-random data: the no self-censoring model.” *JASA.*
- Nabi, Bhattacharya, and Shpitser 2020. “Full law identification in graphical models of missing data: Completeness results.” *ICML.*
- Plackett 1964. “A class of Bivariate Distributions” *JASA.*

References

- Osius 2004. “The association between two random elements: A complete characterization and odds ratio models.” *Metrika.*
- Evans and Richardson 2010. “Maximum likelihood fitting of acyclic directed mixed graphs to binary data.” *UAI*.
- Evans and Didelez 2024. “Parameterizing and simulating from causal models.” *JRSS-B*.
- Pearl 1988. “*Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*.”
- Pearl 2009. “*Causality*.”
- Richardson, Evans, Robins, and Shpitser 2023. “Nested Markov properties for acyclic directed mixed graphs.” *Annals of Statistics.*
- Richardson 2003. “Markov properties for acyclic directed mixed graphs.” *Scandinavian Journal of Statistics.*
- Verma and Pearl 1990. “Equivalence and synthesis of causal models.” *UAI.*
- Bhattacharya, Nabi, and Shpitser 2022. “Semiparametric Inference For Causal Effects In
- Graphical Models With Hidden Variables”. *JMLR.*

References

- Jung, Tian, and Bareinboim 2021. “Estimating Identifiable Causal Effects through Double Machine Learning.” *AAAI.*
- Guo and Nabi 2025. “Average Causal Effect Estimation in DAGs with Hidden Variables: Beyond Back-Door and Front-Door Criteria”. *arXiv*.
- Bhattacharya, Nabi, and Shpitser 2021. “Full law identification in graphical models of missing data: Completeness results”. *ICML*.
- Nabi, Bhattacharya, Shpitser, and Robins 2025. “Causal and counterfactual views of missing data models.” *Statistica Sinica*.
- Bhattacharya and Nabi 2022. “On testability of the front-door model via Verma constraints.” *UAI*.
- Robins and Wasserman 1997. “Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs.” *UAI*.
- Bhattacharya, Srinivasan, Nabi, Ogburn, and Shpitser 2024. “Graphical models of entangled missingness.” *arXiv*.
- Shpitser 2006. *“Complete identification methods for causal inference.”*
- Drton, Foygel, and Sullivant 2011. “Global identifiability of linear SEMs” *Annals of Statistics*

Questions