Variable elimination, graph reduction and efficient g-formula

F. Richard Guo

May 4th 2022, Fellows Seminar, Simons Institute

Statistical Laboratory, University of Cambridge

Ema Perković
Washington

Andrea Rotnitzky
Torcuato Di Tella

Motivating example

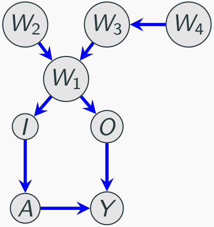
Suppose data comes from a causal DAG $\mathcal{G}$ over vetices $\mathbf{V}$.

Treatment A (discrete), outcome Y. **No unobserved confounders.**

Motivating example

Suppose data comes from a causal DAG $\mathcal{G}$ over vetices $\mathbf{V}$.

Treatment A (discrete), outcome Y. **No unobserved confounders.**

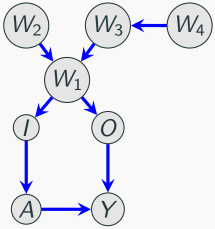


$$
\begin{aligned}
p &= \prod_v p(v \mid \mathrm{Pa}(v)) \\
&= p(a \mid i)p(y \mid a, o)p(i \mid w_1)p(o \mid w_1) \\
&\quad \times p(w_1 \mid w_2, w_3)p(w_3 \mid w_4)p(w_2)p(w_4).
\end{aligned}
$$

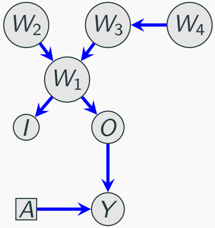
Motivating example

Suppose data comes from a causal DAG $\mathcal{G}$ over vetices $\mathbf{V}$.

Treatment A (discrete), outcome Y. **No unobserved confounders.**



$$
\begin{aligned}
p &= \prod_v p(v \mid \mathrm{Pa}(v)) \\
&= p(a \mid i)p(y \mid a, o)p(i \mid w_1)p(o \mid w_1) \\
&\quad \times p(w_1 \mid w_2, w_3)p(w_3 \mid w_4)p(w_2)p(w_4).
\end{aligned}
$$



$$
\begin{aligned}
p(\mathrm{do}(A = a)) &= \prod_{v \neq a} p(v \mid \mathrm{Pa}(v)) \\
&= p(y \mid a, o)p(i \mid w_1)p(o \mid w_1) \\
&\quad \times p(w_1 \mid w_2, w_3)p(w_3 \mid w_4)p(w_2)p(w_4).
\end{aligned}
$$

g-formula

Suppose we are interested in the counterfactual mean $\mathbb{E}\, Y(a)$.

g-formula

Suppose we are interested in the counterfactual mean $\mathbb{E}\, Y(a)$.

g-formula (Robins, 1986)

aka *manipulated distribution formula* (Spirtes, Glymour, and Scheines, 2000), *truncated factorization formula* (Pearl, 2000)

$$\mathbb{E}\, Y(a) = \Psi_a(P; \mathcal{G}) \equiv \sum_{\mathbf{V} \setminus \{A\}} y \left\{ \prod_{v \neq a} P(v \mid \mathrm{Pa}(v, \mathcal{G})) \right\} \Bigg|_{A=a}.$$

g-formula

Suppose we are interested in the counterfactual mean $\mathbb{E}\, Y(a)$.

g-formula (Robins, 1986)

aka *manipulated distribution formula* (Spirtes, Glymour, and Scheines, 2000), *truncated factorization formula* (Pearl, 2000)

$$\mathbb{E}\, Y(a) = \Psi_a(P;\mathcal{G}) \equiv \sum_{\mathbf{V}\setminus\{A\}} y \left\{ \prod_{v \neq a} P(v \mid \mathrm{Pa}(v,\mathcal{G})) \right\} \Bigg|_{A=a}.$$

$$\Psi_a(P;\mathcal{G}) = \sum_{y,o,i,w_1,w_2,w_3,w_4} yp(y \mid A=a, o)p(i \mid w_1)p(o \mid w_1) \\ \times p(w_1 \mid w_2, w_3)p(w_3 \mid w_4)p(w_2)p(w_4).$$

g-formula

Suppose we are interested in the counterfactual mean $\mathbb{E}\, Y(a)$.

g-formula (Robins, 1986)

aka *manipulated distribution formula* (Spirtes, Glymour, and Scheines, 2000), *truncated factorization formula* (Pearl, 2000)

$$\mathbb{E}\, Y(a) = \Psi_a(P; \mathcal{G}) \equiv \sum_{\mathbf{V} \setminus \{A\}} y \left\{ \prod_{v \neq a} P(v \mid \mathrm{Pa}(v, \mathcal{G})) \right\} \Bigg|_{A=a}.$$

$$\Psi_a(P; \mathcal{G}) = \sum_{y, o, i, w_1, w_2, w_3, w_4} y p(y \mid A = a, o) p(i \mid w_1) p(o \mid w_1) \\ \times p(w_1 \mid w_2, w_3) p(w_3 \mid w_4) p(w_2) p(w_4).$$

☞ Due to the factorization of P, in the model, $\mathbb{E}\, Y(a)$ can be expressed **in different forms**.

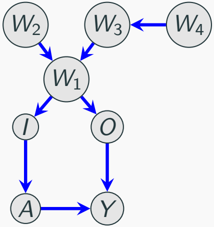
Other identifying formulae

We have back-door formulae

$$\Psi_a^{\mathrm{ADJ}}(P; \mathcal{G}) = \mathbb{E}[\mathbb{E}[Y \mid A = a, \mathbf{L}]],$$

where **adjustment set** $\mathbf{L}$ can take

$$\{O\}, \{I, O\}, \{I, W_1, O\}, \{I, O, W_1, W_2\}, \ldots.$$



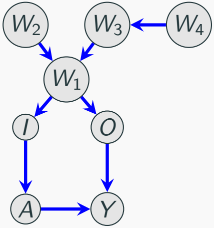
Other identifying formulae

We have back-door formulae

$$\Psi_a^{\text{ADJ}}(P;\mathcal{G}) = \mathbb{E}[\mathbb{E}[Y \mid A = a, \mathbf{L}]],$$

where **adjustment set** $\mathbf{L}$ can take

$$\{O\},\ \{I, O\},\ \{I, W_1, O\},\ \{I, O, W_1, W_2\}, \ldots .$$



☞ Which formula should I use?

Plugin estimates

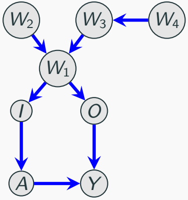
We can compare the **large-sample performance** of plugin estimators of formulae. (☞ Suppose all variables take only finitely many levels.)

Plugin estimates

We can compare the **large-sample performance** of plugin estimators of formulae. (☞ Suppose all variables take only finitely many levels.)

☞ g-formula is efficient (MLE):

$$\Psi_a(\mathbb{P}_n;\mathcal{G}) = \sum_{y,o,i,w_1,w_2,w_3,w_4} y\,\mathbb{P}_n(y \mid A=a,o)\mathbb{P}_n(i \mid w_1)$$
$$\times\mathbb{P}_n(o \mid w_1)\mathbb{P}_n(w_1 \mid w_2, w_3)\mathbb{P}_n(w_3 \mid w_4)\mathbb{P}_n(w_2)\mathbb{P}_n(w_4).$$



$\mathcal{G}$

Plugin estimates

We can compare the **large-sample performance** of plugin estimators of formulae. (☞ Suppose all variables take only finitely many levels.)

☞ g-formula is efficient (MLE):

$$\Psi_a(\mathbb{P}_n;\mathcal{G}) = \sum_{y,o,i,w_1,w_2,w_3,w_4} y\,\mathbb{P}_n(y \mid A=a,o)\mathbb{P}_n(i \mid w_1)$$

$$\times\mathbb{P}_n(o \mid w_1)\mathbb{P}_n(w_1 \mid w_2,w_3)\mathbb{P}_n(w_3 \mid w_4)\mathbb{P}_n(w_2)\mathbb{P}_n(w_4).$$

☞ adjustment formulae are **not** efficient:

$$\Psi_a^{\mathrm{ADJ}}(\mathbb{P}_n;\mathcal{G}) = \sum_{\mathbf{L}} \mathbb{P}_n(\mathbf{L}) \sum_y y\mathbb{P}_n(y \mid A=a,\mathbf{L})$$

$$\begin{tikzcd}
W_2 \arrow[dr] & & W_3 \arrow[dl] & W_4 \arrow[l] \\
& W_1 \arrow[dl] \arrow[dr] & & \\
I \arrow[d] & & O \arrow[d] & \\
A \arrow[rr] & & Y &
\end{tikzcd}$$

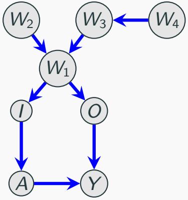
$\mathcal{G}$

Plugin estimates

We can compare the **large-sample performance** of plugin estimators of formulae. (☞ Suppose all variables take only finitely many levels.)

☞ g-formula is efficient (MLE):

$$\Psi_a(\mathbb{P}_n;\mathcal{G}) = \sum_{y,o,i,w_1,w_2,w_3,w_4} y\,\mathbb{P}_n(y \mid A=a,o)\mathbb{P}_n(i \mid w_1)$$
$$\times\mathbb{P}_n(o \mid w_1)\mathbb{P}_n(w_1 \mid w_2,w_3)\mathbb{P}_n(w_3 \mid w_4)\mathbb{P}_n(w_2)\mathbb{P}_n(w_4).$$



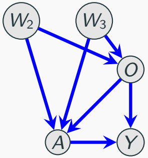
$\mathcal{G}$

☞ adjustment formulae are **not** efficient:

$$\Psi_a^{\mathrm{ADJ}}(\mathbb{P}_n;\mathcal{G}) = \sum_{\mathbf{L}} \mathbb{P}_n(\mathbf{L}) \sum_y y\mathbb{P}_n(y \mid A=a,\mathbf{L})$$

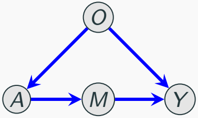
☞ Yet, a simpler g-formula is **also** efficient!

$$\Psi_a(\mathbb{P}_n;\mathcal{G}^*) = \sum_{y,o,w_2,w_3} y\mathbb{P}_n[y \mid A=a,o]$$
$$\times\,\mathbb{P}_n(o \mid w_2,w_3)\mathbb{P}_n(w_2)\mathbb{P}_n(w_3).$$

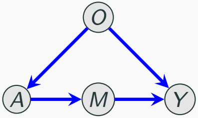


$\mathcal{G}^*$

Another example



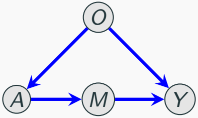
Another example



1. "back-door" (Pearl, 1993)

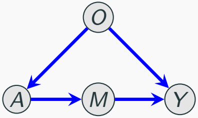
$$\Psi_a^{\mathrm{ADJ}}(P; \mathcal{G}) = \sum_o \mathbb{E}[Y \mid A = a, o] p(o),$$

Another example



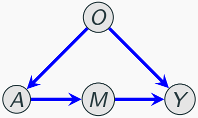
1. "back-door" (Pearl, 1993)
$$\Psi_a^{\mathrm{ADJ}}(P;\mathcal{G}) = \sum_o \mathbb{E}[Y \mid A=a, o]p(o),$$
2. "front-door" (Pearl, 1995)
$$\Psi_a^{\mathrm{FRONT}}(P;\mathcal{G}) = \sum_m \left\{ \sum_a \mathbb{E}[Y \mid m, A=a']p(A=a') \right\} p(m \mid A=a),$$

Another example



1. "back-door" (Pearl, 1993)
$$\Psi_a^{\mathrm{ADJ}}(P;\mathcal{G}) = \sum_o \mathbb{E}[Y \mid A = a, o]p(o),$$
2. "front-door" (Pearl, 1995)
$$\Psi_a^{\mathrm{FRONT}}(P;\mathcal{G}) = \sum_m \left\{ \sum_a \mathbb{E}[Y \mid m, A = a']p(A = a') \right\} p(m \mid A = a),$$
3. or the g-formula (Robins, 1986)
$$\Psi_a(P;\mathcal{G}) = \sum_{m,o} \mathbb{E}[Y \mid m, o]p(o)p(m \mid A = a).$$

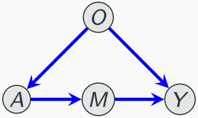
Another example



1. "back-door" (Pearl, 1993)
$$\Psi_a^{\text{ADJ}}(P;\mathcal{G}) = \sum_o \mathbb{E}[Y \mid A=a, o]p(o),$$
2. "front-door" (Pearl, 1995)
$$\Psi_a^{\text{FRONT}}(P;\mathcal{G}) = \sum_m \left\{ \sum_a \mathbb{E}[Y \mid m, A=a']p(A=a') \right\} p(m \mid A=a),$$
3. or the g-formula (Robins, 1986)
$$\Psi_a(P;\mathcal{G}) = \sum_{m,o} \mathbb{E}[Y \mid m, o]p(o)p(m \mid A=a).$$

☞ Which one should be preferred?

Another example



1. "back-door" (Pearl, 1993)
$$\Psi_a^{\mathrm{ADJ}}(P;\mathcal{G}) = \sum_o \mathbb{E}[Y \mid A = a, o]p(o),$$
2. "front-door" (Pearl, 1995)
$$\Psi_a^{\mathrm{FRONT}}(P;\mathcal{G}) = \sum_m \left\{ \sum_a \mathbb{E}[Y \mid m, A = a']p(A = a') \right\} p(m \mid A = a),$$
3. or the g-formula (Robins, 1986)
$$\Psi_a(P;\mathcal{G}) = \sum_{m,o} \mathbb{E}[Y \mid m, o]p(o)p(m \mid A = a).$$

☞ Which one should be preferred? *"Another area which is neglected in my opinion ... given an estimand find the best way of decomposing to estimate it"* — Judea Pearl (OCIS, Nov 17, 2020).

Semiparametric efficiency

Semiparametric efficiency

The **semiparametric** efficiency bound is defined with respect to

$$\mathcal{M}(\mathcal{G}, \mathbf{V}) := \{P(\mathbf{V}) : P \text{ factorizes according to } \mathcal{G}\}.$$

Semiparametric efficiency

The **semiparametric** efficiency bound is defined with respect to

$$\mathcal{M}(\mathcal{G}, \mathbf{V}) := \{P(\mathbf{V}) : P \text{ factorizes according to } \mathcal{G}\}.$$

☞ The efficiency bound for estimating $\mathbb{E}\, Y(a)$ at P (assuming positivity) with respect to $\mathcal{M}(\mathcal{G}, \mathbf{V})$ is characterized by the **efficient influence function** $\chi_{\text{eff},P}(\mathbf{V}; \mathcal{G})$.

Semiparametric efficiency

The **semiparametric** efficiency bound is defined with respect to

$$\mathcal{M}(\mathcal{G}, \mathbf{V}) := \{P(\mathbf{V}) : P \text{ factorizes according to } \mathcal{G}\}.$$

☞ The efficiency bound for estimating $\mathbb{E}\, Y(a)$ at P (assuming positivity) with respect to $\mathcal{M}(\mathcal{G}, \mathbf{V})$ is characterized by the **efficient influence function** $\chi_{\mathsf{eff},P}(\mathbf{V}; \mathcal{G})$.

☞ Observation: for some graphs, certain variables **"cancels out"** from $\chi_{\mathsf{eff},P}(\mathbf{V}; \mathcal{G})$.

Informative variables

Lemma (Rotnitzky and Smucler, 2020) Let $\mathcal{M}$ be a semiparametric model on vector $\mathbf{V}$. Suppose $\mathbf{V}'$ is a subvector of $\mathbf{V}$, such that

1. $\Psi(P)$ depends on P only through margin $P(\mathbf{V}')$
2. and $\chi_{\text{eff},P}(\mathbf{V};\mathcal{M})$ only depends on $\mathbf{V}$ through $\mathbf{V}'$ for every $P \in \mathcal{M}(\mathbf{V})$,

then

$$\chi_{\text{eff},P}(\mathbf{V};\mathcal{M}) = \chi_{\text{eff},P}(\mathbf{V}';\mathcal{M}') \quad P\text{-a.s. for every } P \in \mathcal{M},$$

where $\mathcal{M}' \equiv \{P(\mathbf{V}') : P(\mathbf{V}) \in \mathcal{M}\}$ is the induced marginal model over $\mathbf{V}'$.

Informative variables

Lemma (Rotnitzky and Smucler, 2020) Let $\mathcal{M}$ be a semiparametric model on vector $\mathbf{V}$. Suppose $\mathbf{V}'$ is a subvector of $\mathbf{V}$, such that

1. $\Psi(P)$ depends on P only through margin $P(\mathbf{V}')$
2. and $\chi_{\text{eff},P}(\mathbf{V};\mathcal{M})$ only depends on $\mathbf{V}$ through $\mathbf{V}'$ for every $P \in \mathcal{M}(\mathbf{V})$,

then

$$\chi_{\text{eff},P}(\mathbf{V};\mathcal{M}) = \chi_{\text{eff},P}(\mathbf{V}';\mathcal{M}') \quad P\text{-a.s. for every } P \in \mathcal{M},$$

where $\mathcal{M}' \equiv \{P(\mathbf{V}') : P(\mathbf{V}) \in \mathcal{M}\}$ is the induced marginal model over $\mathbf{V}'$.

Definition Given graph $\mathcal{G}$ over $\mathbf{V}$, we say subset $\mathbf{U} \subset \mathbf{V}$ is **uninformative** for estimating $\mathbb{E}\, Y(a)$ if

1. $\mathbb{E}\, Y(a)$ is identified from $P(\mathbf{V} \setminus \mathbf{U})$,
2. and $\chi_{\text{eff},P}(\mathbf{V};\mathcal{G})$ does not depend on $\mathbf{U}$ P-a.e. for all $P \in \mathcal{M}(\mathcal{G},\mathbf{V})$.

Informative variables

Lemma (Rotnitzky and Smucler, 2020) Let $\mathcal{M}$ be a semiparametric model on vector $\mathbf{V}$. Suppose $\mathbf{V}'$ is a subvector of $\mathbf{V}$, such that

1. $\Psi(P)$ depends on P only through margin $P(\mathbf{V}')$
2. and $\chi_{\text{eff},P}(\mathbf{V};\mathcal{M})$ only depends on $\mathbf{V}$ through $\mathbf{V}'$ for every $P \in \mathcal{M}(\mathbf{V})$,

then

$$\chi_{\text{eff},P}(\mathbf{V};\mathcal{M}) = \chi_{\text{eff},P}(\mathbf{V}';\mathcal{M}') \quad P\text{-a.s. for every } P \in \mathcal{M},$$

where $\mathcal{M}' \equiv \{P(\mathbf{V}') : P(\mathbf{V}) \in \mathcal{M}\}$ is the induced marginal model over $\mathbf{V}'$.

Definition Given graph $\mathcal{G}$ over $\mathbf{V}$, we say subset $\mathbf{U} \subset \mathbf{V}$ is **uninformative** for estimating $\mathbb{E}\, Y(a)$ if

1. $\mathbb{E}\, Y(a)$ is identified from $P(\mathbf{V} \setminus \mathbf{U})$,
2. and $\chi_{\text{eff},P}(\mathbf{V};\mathcal{G})$ does not depend on $\mathbf{U}$ P-a.e. for all $P \in \mathcal{M}(\mathcal{G},\mathbf{V})$.

Irreducible informative set $\mathbf{V}^*(\mathcal{G}) \equiv \{\text{smallest } \mathbf{V}' : \mathbf{V} \setminus \mathbf{V}' \text{ is uninformative}\}$.

Agenda

1. **Variable elimination**: identify informative variables $\mathbf{V}^*(\mathcal{G})$.
2. **Graph reduction**: characterize the marginal model over $\mathbf{V}^*(\mathcal{G})$.
3. Derive a **simpler g-formula**.

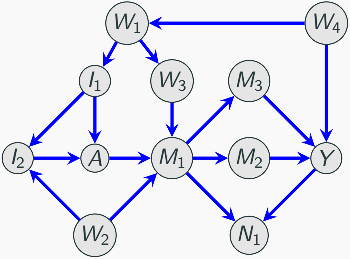
Taxonomy of vertices

Taxonomy of vertices

We write $u \mapsto v$ if u is an ancestor of v. (We always suppose $A \mapsto Y$.)

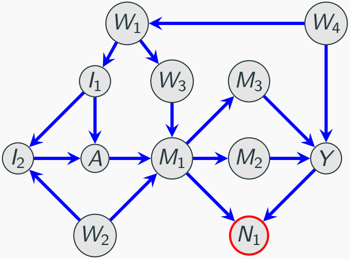
Taxonomy of vertices

We write $u \mapsto v$ if u is an ancestor of v. (We always suppose $A \mapsto Y$.)



Taxonomy of vertices

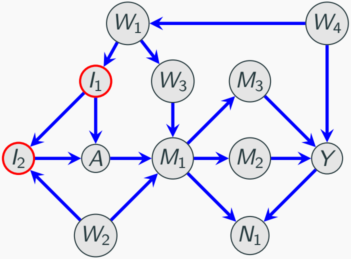
We write $u \mapsto v$ if u is an ancestor of v. (We always suppose $A \mapsto Y$.)



- $\mathbf{N}(\mathcal{G}) = \{v : v \not\mapsto Y\}$
  ☞ non-ancestors of Y

Taxonomy of vertices

We write $u \mapsto v$ if u is an ancestor of v. (We always suppose $A \mapsto Y$.)



- $\mathbf{N}(\mathcal{G}) = \{v : v \not\mapsto Y\}$

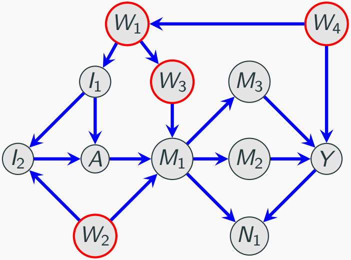
  ☞ non-ancestors of Y

- $\mathbf{I}(\mathcal{G}) = \{v \neq A : v \mapsto Y \text{ only through } A\}$

  ☞ conditional instruments

Taxonomy of vertices

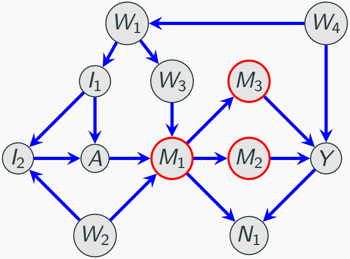
We write $u \mapsto v$ if u is an ancestor of v. (We always suppose $A \mapsto Y$.)



- $\mathbf{N}(\mathcal{G}) = \{v : v \not\mapsto Y\}$
  ☞ non-ancestors of Y
- $\mathbf{I}(\mathcal{G}) = \{v \neq A : v \mapsto Y \text{ only through } A\}$
  ☞ conditional instruments
- $\mathbf{W}(\mathcal{G}) = \{v \notin \mathbf{I}(\mathcal{G}) : A \not\mapsto v, v \mapsto Y\}$
  ☞ baseline covariates

Taxonomy of vertices

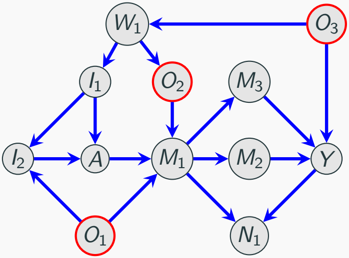
We write $u \mapsto v$ if u is an ancestor of v. (We always suppose $A \mapsto Y$.)



- $\mathbf{N}(\mathcal{G}) = \{v : v \not\mapsto Y\}$
  ☞ non-ancestors of Y
- $\mathbf{I}(\mathcal{G}) = \{v \neq A : v \mapsto Y \text{ only through } A\}$
  ☞ conditional instruments
- $\mathbf{W}(\mathcal{G}) = \{v \notin \mathbf{I}(\mathcal{G}) : A \not\mapsto v, v \mapsto Y\}$
  ☞ baseline covariates
- $\mathbf{M}(\mathcal{G}) = \{v : A \mapsto v \mapsto Y\}$
  ☞ mediators

Taxonomy of vertices

We write $u \mapsto v$ if u is an ancestor of v. (We always suppose $A \mapsto Y$.)



- $\mathbf{N}(\mathcal{G}) = \{v : v \not\mapsto Y\}$
  ☞ non-ancestors of Y
- $\mathbf{I}(\mathcal{G}) = \{v \neq A : v \mapsto Y \text{ only through } A\}$
  ☞ conditional instruments
- $\mathbf{W}(\mathcal{G}) = \{v \notin \mathbf{I}(\mathcal{G}) : A \not\mapsto v, v \mapsto Y\}$
  ☞ baseline covariates
- $\mathbf{M}(\mathcal{G}) = \{v : A \mapsto v \mapsto Y\}$
  ☞ mediators

☞ A special subset of **W** plays an important role:

$$\mathbf{O}(\mathcal{G}) := \mathrm{Pa}(\mathbf{M} \cup \{Y\}) \setminus (\mathrm{De}(\mathbf{M} \cup \{Y\}) \cup \{A\})$$

is the **optimal adjustment set** (Henckel, Perković, and Maathuis, 2022; Rotnitzky and Smucler, 2020).
(**O** consists of direct parents of **M** or Y that do not block causal paths.)

N and I are uninformative

Using conditional independences on $\mathcal{G}$, Rotnitzky and Smucler (2020) showed

$$\chi_{\text{eff},P}(\mathbf{V},\mathcal{G}) = \sum_{j=1}^{J} \left(\mathbb{E}[b(\mathbf{O}) \mid W_j, \text{Pa}(W_j)] - \mathbb{E}[b(\mathbf{O}) \mid \text{Pa}(W_j)]\right)$$
$$+ \sum_{k=1}^{K+1} \left(\mathbb{E}[AY/\pi(\mathbf{O}) \mid M_k, \text{Pa}(M_k)] - \mathbb{E}[AY/\pi(\mathbf{O}) \mid \text{Pa}(M_k)]\right),$$

where $M_{K+1} \equiv Y$, $b(\mathbf{O}) = \mathbb{E}[Y \mid A = 1, \mathbf{O}]$, $\pi(\mathbf{O}) = P(A = 1 \mid \mathbf{O})$.

N and I are uninformative

Using conditional independences on $\mathcal{G}$, Rotnitzky and Smucler (2020) showed

$$\chi_{\mathrm{eff},P}(\mathbf{V},\mathcal{G}) = \sum_{j=1}^{J}\left(\mathbb{E}[b(\mathbf{O}) \mid W_j, \mathrm{Pa}(W_j)] - \mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_j)]\right)$$
$$+ \sum_{k=1}^{K+1}\left(\mathbb{E}[AY/\pi(\mathbf{O}) \mid M_k, \mathrm{Pa}(M_k)] - \mathbb{E}[AY/\pi(\mathbf{O}) \mid \mathrm{Pa}(M_k)]\right),$$

where $M_{K+1} \equiv Y$, $b(\mathbf{O}) = \mathbb{E}[Y \mid A = 1, \mathbf{O}]$, $\pi(\mathbf{O}) = P(A = 1 \mid \mathbf{O})$.

Corollary This implies that

$$\mathbf{N}(\mathcal{G}) \equiv \{\text{non-ancestors of } Y\} \quad \text{and} \quad \mathbf{I}(\mathcal{G}) \equiv \{\text{indirect ancestors of } Y\}$$

are **uninformative**.

N and I are uninformative

Using conditional independences on $\mathcal{G}$, Rotnitzky and Smucler (2020) showed

$$\chi_{\text{eff},P}(\mathbf{V},\mathcal{G}) = \sum_{j=1}^{J} \left(\mathbb{E}[b(\mathbf{O}) \mid W_j, \mathrm{Pa}(W_j)] - \mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_j)]\right) + \sum_{k=1}^{K+1} \left(\mathbb{E}[AY/\pi(\mathbf{O}) \mid M_k, \mathrm{Pa}(M_k)] - \mathbb{E}[AY/\pi(\mathbf{O}) \mid \mathrm{Pa}(M_k)]\right),$$

where $M_{K+1} \equiv Y$, $b(\mathbf{O}) = \mathbb{E}[Y \mid A = 1, \mathbf{O}]$, $\pi(\mathbf{O}) = P(A = 1 \mid \mathbf{O})$.

Corollary This implies that

$$\mathbf{N}(\mathcal{G}) \equiv \{\text{non-ancestors of } Y\} \quad \text{and} \quad \mathbf{I}(\mathcal{G}) \equiv \{\text{indirect ancestors of } Y\}$$

are **uninformative**.

☞ Yet, $\mathbf{N}(\mathcal{G})$ and $\mathbf{I}(\mathcal{G})$ are defined with respect to $\mathcal{G}$ rather than $\mathcal{M}(\mathcal{G}, \mathbf{V})$.

Causal Markov equivalence

Causal Markov equivalence

Two DAGs $\mathcal{G}$ and $\mathcal{G}'$ are **Markov equivalent** if they define the same set of models

$$\mathcal{G} \simeq \mathcal{G}' \iff \mathcal{M}(\mathcal{G}, \mathbf{V}) = \mathcal{M}(\mathcal{G}', \mathbf{V}).$$

Causal Markov equivalence

Two DAGs $\mathcal{G}$ and $\mathcal{G}'$ are **Markov equivalent** if they define the same set of models

$$\mathcal{G} \simeq \mathcal{G}' \iff \mathcal{M}(\mathcal{G}, \mathbf{V}) = \mathcal{M}(\mathcal{G}', \mathbf{V}).$$

Markov equivalence class (MEC) (Andersson, Madigan, and Perlman, 1997; Verma and Pearl, 1991)

$\mathcal{G} \simeq \mathcal{G}' \iff \mathcal{G}$ and $\mathcal{G}'$ share the same **adjacencies** and **unshielded colliders**

(unshielded collider: $a \rightarrow \circ \leftarrow b$ with a and b non-adjacent.)

Causal Markov equivalence

Two DAGs $\mathcal{G}$ and $\mathcal{G}'$ are **Markov equivalent** if they define the same set of models

$$\mathcal{G} \simeq \mathcal{G}' \iff \mathcal{M}(\mathcal{G}, \mathbf{V}) = \mathcal{M}(\mathcal{G}', \mathbf{V}).$$

Markov equivalence class (MEC) (Andersson, Madigan, and Perlman, 1997; Verma and Pearl, 1991)

$\mathcal{G} \simeq \mathcal{G}' \iff \mathcal{G}$ and $\mathcal{G}'$ share the same **adjacencies** and **unshielded colliders**

(unshielded collider: $a \to \circ \leftarrow b$ with a and b non-adjacent.)

☞ But Markov equivalence does not preserve the causal interpretation:

Causal Markov equivalence

Causal Markov equivalence (with respect to the effect of A on Y):

$$\mathcal{G} \overset{c}{\sim} \mathcal{G}' \iff \mathcal{G} \simeq \mathcal{G}' \text{ and } \Psi(P, \mathcal{G}) = \Psi(P, \mathcal{G}') \text{ for all } P \in \mathcal{M}_{\mathcal{G}}.$$

Causal Markov equivalence

Causal Markov equivalence (with respect to the effect of A on Y):

$$\mathcal{G} \overset{c}{\sim} \mathcal{G}' \iff \mathcal{G} \simeq \mathcal{G}' \text{ and } \Psi(P, \mathcal{G}) = \Psi(P, \mathcal{G}') \text{ for all } P \in \mathcal{M}_{\mathcal{G}}.$$

☞ The causal Markov equivalence class **(c-MEC)** is characterized by Guo and Perković (2021) (with MPDAGs).

Causal Markov equivalence

Causal Markov equivalence (with respect to the effect of A on Y):

$$\mathcal{G} \overset{c}{\sim} \mathcal{G}' \iff \mathcal{G} \simeq \mathcal{G}' \text{ and } \Psi(P, \mathcal{G}) = \Psi(P, \mathcal{G}') \text{ for all } P \in \mathcal{M}_{\mathcal{G}}.$$

☞ The causal Markov equivalence class **(c-MEC)** is characterized by Guo and Perković (2021) (with MPDAGs).

Theorem $\bigcup_{\mathcal{G}' \overset{c}{\sim} \mathcal{G}} \mathbf{I}(\mathcal{G}') \cup \mathbf{N}(\mathcal{G}')$ is uninformative.

Causal Markov equivalence

Causal Markov equivalence (with respect to the effect of A on Y):

$$\mathcal{G} \overset{\mathrm{c}}{\sim} \mathcal{G}' \iff \mathcal{G} \simeq \mathcal{G}' \text{ and } \Psi(P, \mathcal{G}) = \Psi(P, \mathcal{G}') \text{ for all } P \in \mathcal{M}_{\mathcal{G}}.$$

☞ The causal Markov equivalence class **(c-MEC)** is characterized by Guo and Perković (2021) (with MPDAGs).

Theorem $\bigcup_{\mathcal{G}' \overset{\mathrm{c}}{\sim} \mathcal{G}} \mathbf{I}(\mathcal{G}') \cup \mathbf{N}(\mathcal{G}')$ is uninformative.

It can be further shown that there exists $\check{\mathcal{G}} \overset{\mathrm{c}}{\sim} \mathcal{G}$ such that

$$\mathbf{I}(\check{\mathcal{G}}) = \bigcup_{\mathcal{G}' \overset{\mathrm{c}}{\sim} \mathcal{G}} \mathbf{I}(\mathcal{G}'), \quad \mathbf{N}(\check{\mathcal{G}}) = \bigcup_{\mathcal{G}' \overset{\mathrm{c}}{\sim} \mathcal{G}} \mathbf{N}(\mathcal{G}').$$

Causal Markov equivalence

Causal Markov equivalence (with respect to the effect of A on Y):

$$\mathcal{G} \overset{c}{\sim} \mathcal{G}' \iff \mathcal{G} \simeq \mathcal{G}' \text{ and } \Psi(P, \mathcal{G}) = \Psi(P, \mathcal{G}') \text{ for all } P \in \mathcal{M}_{\mathcal{G}}.$$

☞ The causal Markov equivalence class **(c-MEC)** is characterized by Guo and Perković (2021) (with MPDAGs).

Theorem $\bigcup_{\mathcal{G}' \overset{c}{\sim} \mathcal{G}} \mathbf{I}(\mathcal{G}') \cup \mathbf{N}(\mathcal{G}')$ is uninformative.

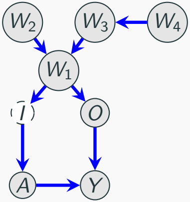
It can be further shown that there exists $\check{\mathcal{G}} \overset{c}{\sim} \mathcal{G}$ such that

$$\mathbf{I}(\check{\mathcal{G}}) = \bigcup_{\mathcal{G}' \overset{c}{\sim} \mathcal{G}} \mathbf{I}(\mathcal{G}'), \quad \mathbf{N}(\check{\mathcal{G}}) = \bigcup_{\mathcal{G}' \overset{c}{\sim} \mathcal{G}} \mathbf{N}(\mathcal{G}').$$

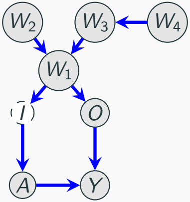
☞ Vertices can be determined uninformative by moving **within** the causal Markov equivalence class (flipping edges).

I and W_4 are uninformative

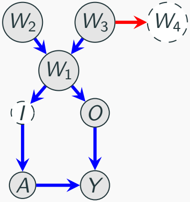
I and W_4 are uninformative



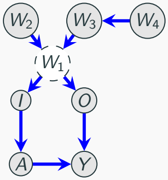
I and W_4 are uninformative



$\overset{c}{\sim}$



Not complete ...



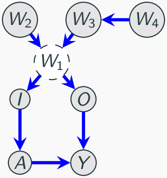
$\mathcal{G}$

$\chi_{\text{eff},P}(\mathbf{V}, \mathcal{G})$ depends on W_1 thought the terms

$$\cancel{\mathbb{E}\{b_a(O) \mid W_1, W_2, W_3\}} \quad + \quad b_a(O) \quad - \quad \cancel{\mathbb{E}\{b_a(O) \mid W_1\}}$$

due to $O \perp\!\!\!\perp_{\mathcal{G}} W_2, W_3 \mid W_1$, where $b_a(O) \equiv \mathbb{E}[Y \mid A = a, O]$.

Not complete ...



$\mathcal{G}$

$\chi_{\text{eff},P}(\mathbf{V}, \mathcal{G})$ depends on W_1 thought the terms

$$\cancel{\mathbb{E}\{b_a(O) \mid W_1, W_2, W_3\}} \quad + \quad b_a(O) \quad - \quad \cancel{\mathbb{E}\{b_a(O) \mid W_1\}}$$

due to $O \perp\!\!\!\perp_{\mathcal{G}} W_2, W_3 \mid W_1$, where $b_a(O) \equiv \mathbb{E}[Y \mid A = a, O]$.

☞ **Uninformative** W_1 cannot be detected this way!

Conspired cancellation

Conspired cancellation

Fix $W_j \equiv W_{j_0} \in \mathbf{W} \setminus \mathbf{O}$. Let $\mathbf{W} \cap \mathrm{Ch}(W_j)$ be topo-sorted as $\{W_{j_1}, \ldots, W_{j_r}\}$.

Conspired cancellation

Fix $W_j \equiv W_{j_0} \in \mathbf{W} \setminus \mathbf{O}$. Let $\mathbf{W} \cap \mathrm{Ch}(W_j)$ be topo-sorted as $\{W_{j_1}, \ldots, W_{j_r}\}$.

☞ Then the EIF only depends on W_j through the terms:

$$+\mathbb{E}[b(\mathbf{O}) \mid W_j, \mathrm{Pa}(W_j)] + \mathbb{E}[b(\mathbf{O}) \mid W_{j_1}, \mathrm{Pa}(W_{j_1})] + \mathbb{E}[b(\mathbf{O}) \mid W_{j_2}, \mathrm{Pa}(W_{j_2})] \quad \cdots \quad + \mathbb{E}[b(\mathbf{O}) \mid W_{j_r}, \mathrm{Pa}(W_{j_r})]$$

$$-\mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_{j_1})] \qquad -\mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_{j_2})] \qquad -\mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_{j_3})] \quad \cdots$$

Conspired cancellation

Fix $W_j \equiv W_{j_0} \in \mathbf{W} \setminus \mathbf{O}$. Let $\mathbf{W} \cap \mathrm{Ch}(W_j)$ be topo-sorted as $\{W_{j_1}, \ldots, W_{j_r}\}$.

☞ Then the EIF only depends on W_j through the terms:

$$+\mathbb{E}[b(\mathbf{O}) \mid W_j, \mathrm{Pa}(W_j)] + \mathbb{E}[b(\mathbf{O}) \mid W_{j_1}, \mathrm{Pa}(W_{j_1})] + \mathbb{E}[b(\mathbf{O}) \mid W_{j_2}, \mathrm{Pa}(W_{j_2})] \quad \cdots \quad + \mathbb{E}[b(\mathbf{O}) \mid W_{j_r}, \mathrm{Pa}(W_{j_r})]$$

$$-\mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_{j_1})] \qquad -\mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_{j_2})] \qquad -\mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_{j_3})] \quad \cdots$$

Conspired cancellation

Fix $W_j \equiv W_{j_0} \in \mathbf{W} \setminus \mathbf{O}$. Let $\mathbf{W} \cap \mathrm{Ch}(W_j)$ be topo-sorted as $\{W_{j_1}, \dots, W_{j_r}\}$.

☞ Then the EIF only depends on W_j through the terms:

$$+\mathbb{E}[b(\mathbf{O}) \mid W_j, \mathrm{Pa}(W_j)] + \mathbb{E}[b(\mathbf{O}) \mid W_{j_1}, \mathrm{Pa}(W_{j_1})] + \mathbb{E}[b(\mathbf{O}) \mid W_{j_2}, \mathrm{Pa}(W_{j_2})] \quad \cdots \quad + \mathbb{E}[b(\mathbf{O}) \mid W_{j_r}, \mathrm{Pa}(W_{j_r})]$$

$$-\mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_{j_1})] \qquad -\mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_{j_2})] \qquad -\mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_{j_3})] \quad \cdots$$

To *make this happen*, we can posit the following graphical **W-criterion**:

1. $W_j \perp\!\!\!\perp_{\mathcal{G}} \mathbf{O} \mid W_{j_r}, \mathrm{Pa}(W_{j_r}) \setminus \{W_j\}$,
2. and for $m = 1, \dots, r$:
   - (i) $W_{j_{m-1}} \to W_{j_m}$ (children are chained)
   - (ii) $\mathrm{Pa}(W_{j_m}) \subseteq \mathrm{Pa}(W_{j_{m-1}}) \cup \{W_{j_{m-1}}\}$ (parent sets are decreasing)
   - (iii) $\mathrm{Pa}(W_{j_{m-1}}) \setminus \mathrm{Pa}(W_{j_m}) \perp\!\!\!\perp_{\mathcal{G}} \mathbf{O} \mid \mathrm{Pa}(W_{j_m})$ (left-over piece is separated from $\mathbf{O}$)

Conspired cancellation

Fix $W_j \equiv W_{j_0} \in \mathbf{W} \setminus \mathbf{O}$. Let $\mathbf{W} \cap \mathrm{Ch}(W_j)$ be topo-sorted as $\{W_{j_1}, \ldots, W_{j_r}\}$.

☞ Then the EIF only depends on W_j through the terms:

$$+\mathbb{E}[b(\mathbf{O}) \mid W_j, \mathrm{Pa}(W_j)] + \mathbb{E}[b(\mathbf{O}) \mid W_{j_1}, \mathrm{Pa}(W_{j_1})] + \mathbb{E}[b(\mathbf{O}) \mid W_{j_2}, \mathrm{Pa}(W_{j_2})] \quad \cdots \quad + \mathbb{E}[b(\mathbf{O}) \mid W_{j_r}, \mathrm{Pa}(W_{j_r})]$$

$$-\mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_{j_1})] \qquad -\mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_{j_2})] \qquad -\mathbb{E}[b(\mathbf{O}) \mid \mathrm{Pa}(W_{j_3})] \quad \cdots$$

To *make this happen*, we can posit the following graphical **W-criterion**:

1. $W_j \perp\!\!\!\perp_{\mathcal{G}} \mathbf{O} \mid W_{j_r}, \mathrm{Pa}(W_{j_r}) \setminus \{W_j\}$,
2. and for $m = 1, \ldots, r$:

   (i) $W_{j_{m-1}} \to W_{j_m}$ (children are chained)

   (ii) $\mathrm{Pa}(W_{j_m}) \subseteq \mathrm{Pa}(W_{j_{m-1}}) \cup \{W_{j_{m-1}}\}$ (parent sets are decreasing)

   (iii) $\mathrm{Pa}(W_{j_{m-1}}) \setminus \mathrm{Pa}(W_{j_m}) \perp\!\!\!\perp_{\mathcal{G}} \mathbf{O} \mid \mathrm{Pa}(W_{j_m})$ (left-over piece is separated from $\mathbf{O}$)

☞ A similar graphical **M-criterion** applies to mediator $M_i \in \mathbf{M}$.

Complete graphical criteria

Theorem The set of informative variables is given by

$$\begin{aligned}\mathbf{V}^*(\mathcal{G}) = \{A, Y\} \cup \mathbf{O} \\ \cup \{W_j \in \mathbf{W} \setminus \mathbf{O} : W_j \text{ fails the W-criterion}\} \\ \cup \{M_i \in \mathbf{M} : M_i \text{ fails the M-criterion}\}.\end{aligned}$$

Complete graphical criteria

Theorem The set of informative variables is given by

$$\begin{aligned}\mathbf{V}^*(\mathcal{G}) = \{A, Y\} &\cup \mathbf{O} \\ &\cup \{W_j \in \mathbf{W} \setminus \mathbf{O} : W_j \text{ fails the W-criterion}\} \\ &\cup \{M_i \in \mathbf{M} : M_i \text{ fails the M-criterion}\}.\end{aligned}$$

Proof sketch:

1. A, Y, $\mathbf{O}$ are informative.

Complete graphical criteria

Theorem The set of informative variables is given by

$$\begin{aligned}\mathbf{V}^*(\mathcal{G}) = \{A, Y\} \cup \mathbf{O} \\ \cup \{W_j \in \mathbf{W} \setminus \mathbf{O} : W_j \text{ fails the W-criterion}\} \\ \cup \{M_i \in \mathbf{M} : M_i \text{ fails the M-criterion}\}.\end{aligned}$$

Proof sketch:

1. A, Y, $\mathbf{O}$ are informative.
2. W_j/M_i satisfies the W/M-criterion $\implies$ W_j/M_i is uninformative
   ☞ by conditional independence.

Complete graphical criteria

Theorem The set of informative variables is given by

$$\begin{aligned}\mathbf{V}^*(\mathcal{G}) = \{A, Y\} \cup \mathbf{O} \\ \cup \{W_j \in \mathbf{W} \setminus \mathbf{O} : W_j \text{ fails the W-criterion}\} \\ \cup \{M_i \in \mathbf{M} : M_i \text{ fails the M-criterion}\}.\end{aligned}$$

Proof sketch:

1. A, Y, $\mathbf{O}$ are informative.
2. W_j/M_i satisfies the W/M-criterion $\implies$ W_j/M_i is uninformative
   ☞ by conditional independence.
3. W_j/M_i fails the W/M-criterion $\implies$ W_j/M_i is informative
   ☞ by constructing certain $P \in \mathcal{M}(\mathcal{G}, \mathbf{V})$ such that $\chi_{\text{eff},P}(\mathbf{V}, \mathcal{G})$ depends on W_j/M_i.

Graph reduction

☞ How do we represent the following **marginal model** of a DAG?

$$\mathcal{M}(\mathcal{G}, \mathbf{V}^*) \equiv \{P(\mathbf{V}^*) : P \in \mathcal{M}(\mathcal{G}, \mathbf{V})\},$$

where $\mathbf{V}^* \equiv \mathbf{V}^*(\mathcal{G})$.

☞ Marginal models of a DAG can be complicated.

Latent projection

☞ One popular approach is the **latent projection** (Verma and Pearl, 1991).

Latent projection

☞ One popular approach is the **latent projection** (Verma and Pearl, 1991).

Suppose $\mathcal{G} = (\mathbf{V} \dot\cup \mathbf{U}, \mathbf{E})$ for observed $\mathbf{V}$ and latent $\mathbf{U}$.

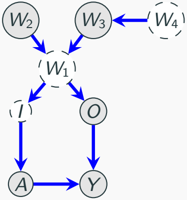
1. Whenever there is a path of the form $w \rightarrow u_1 \rightarrow \cdots \rightarrow u_2 \rightarrow v$ add $w \rightarrow v$ (if not already present).
2. Whenever there is a path of the form $w \leftarrow u_1 \leftarrow \cdots \rightarrow u_2 \rightarrow v$ add $w \leftrightarrow v$.

Latent projection

☞ One popular approach is the **latent projection** (Verma and Pearl, 1991).

Suppose $\mathcal{G} = (\mathbf{V} \dot\cup \mathbf{U}, \mathbf{E})$ for observed $\mathbf{V}$ and latent $\mathbf{U}$.

1. Whenever there is a path of the form $w \rightarrow u_1 \rightarrow \cdots \rightarrow u_2 \rightarrow v$ add $w \rightarrow v$ (if not already present).
2. Whenever there is a path of the form $w \leftarrow u_1 \leftarrow \cdots \rightarrow u_2 \rightarrow v$ add $w \leftrightarrow v$.



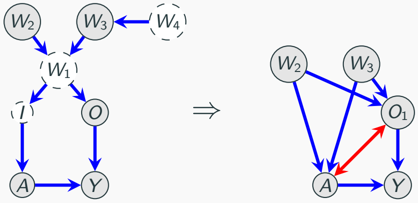
$\Rightarrow$

Latent projection

☞ One popular approach is the **latent projection** (Verma and Pearl, 1991).

Suppose $\mathcal{G} = (\mathbf{V} \dot{\cup} \mathbf{U}, \mathbf{E})$ for observed $\mathbf{V}$ and latent $\mathbf{U}$.

1. Whenever there is a path of the form $w \rightarrow u_1 \rightarrow \cdots \rightarrow u_2 \rightarrow v$ add $w \rightarrow v$ (if not already present).
2. Whenever there is a path of the form $w \leftarrow u_1 \leftarrow \cdots \rightarrow u_2 \rightarrow v$ add $w \leftrightarrow v$.



☞ Results in an ADMG (not always a DAG). Conceptually strange.

Graph reduction algorithm

$\mathbf{V}^* \leftarrow \{A, Y\} \cup \mathbf{W} \cup \mathbf{M}$

$\mathcal{G}^* \leftarrow \mathcal{G}(\mathbf{V}^*)$ by projecting out $\mathbf{N}$ and $\mathbf{I}$ with latent projection

for $v \in \mathbf{V}^*$ **do**

  if ($v \in \mathbf{W} \setminus \mathbf{O}$ and v satisfies the W-criterion) **or** ($v \in \mathbf{M}$ and v satisfies the M-criterion) **then**

    $\mathbf{V}^* \leftarrow \mathbf{V}^* \setminus \{v\}$

    $\mathcal{G}^* \leftarrow \mathcal{G}^*_{-v}$

  end if

end for

return $\mathcal{G}^*$

Graph reduction algorithm

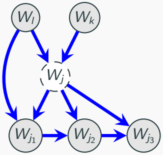
$\mathbf{V}^* \leftarrow \{A, Y\} \cup \mathbf{W} \cup \mathbf{M}$
$\mathcal{G}^* \leftarrow \mathcal{G}(\mathbf{V}^*)$ by projecting out $\mathbf{N}$ and $\mathbf{I}$ with latent projection
for $v \in \mathbf{V}^*$ **do**
    if ($v \in \mathbf{W} \setminus \mathbf{O}$ and v satisfies the W-criterion) **or** ($v \in \mathbf{M}$ and v satisfies the M-criterion) **then**
        $\mathbf{V}^* \leftarrow \mathbf{V}^* \setminus \{v\}$
        $\mathcal{G}^* \leftarrow \mathcal{G}^*_{-v}$
    end if
end for
return $\mathcal{G}^*$

☞ $\mathcal{G}^*_{-v}$ saturates edges from $\mathrm{Pa}(v)$ to $\mathrm{Ch}(v)$ and those within $\mathrm{Ch}(v)$, before removing v.

Graph reduction algorithm

$\mathbf{V}^* \leftarrow \{A, Y\} \cup \mathbf{W} \cup \mathbf{M}$
$\mathcal{G}^* \leftarrow \mathcal{G}(\mathbf{V}^*)$ by projecting out $\mathbf{N}$ and $\mathbf{I}$ with latent projection
for $v \in \mathbf{V}^*$ **do**
  if ($v \in \mathbf{W} \setminus \mathbf{O}$ and v satisfies the W-criterion) **or** ($v \in \mathbf{M}$ and v satisfies the M-criterion) **then**
    $\mathbf{V}^* \leftarrow \mathbf{V}^* \setminus \{v\}$
    $\mathcal{G}^* \leftarrow \mathcal{G}^*_{-v}$
  end if
end for
return $\mathcal{G}^*$

☞ $\mathcal{G}^*_{-v}$ saturates edges from $\mathrm{Pa}(v)$ to $\mathrm{Ch}(v)$ and those within $\mathrm{Ch}(v)$, before removing v.



Graph reduction algorithm

$\mathbf{V}^* \leftarrow \{A, Y\} \cup \mathbf{W} \cup \mathbf{M}$

$\mathcal{G}^* \leftarrow \mathcal{G}(\mathbf{V}^*)$ by projecting out $\mathbf{N}$ and $\mathbf{I}$ with latent projection

for $v \in \mathbf{V}^*$ **do**

  if ($v \in \mathbf{W} \setminus \mathbf{O}$ and v satisfies the W-criterion) **or** ($v \in \mathbf{M}$ and v satisfies the M-criterion) **then**

    $\mathbf{V}^* \leftarrow \mathbf{V}^* \setminus \{v\}$

    $\mathcal{G}^* \leftarrow \mathcal{G}^*_{-v}$

  end if

end for

return $\mathcal{G}^*$

☞ $\mathcal{G}^*_{-v}$ saturates edges from $\mathrm{Pa}(v)$ to $\mathrm{Ch}(v)$ and those within $\mathrm{Ch}(v)$, before removing v.

Graph reduction algorithm

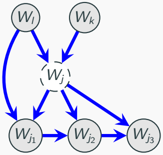
$\mathbf{V}^* \leftarrow \{A, Y\} \cup \mathbf{W} \cup \mathbf{M}$
$\mathcal{G}^* \leftarrow \mathcal{G}(\mathbf{V}^*)$ by projecting out $\mathbf{N}$ and $\mathbf{I}$ with latent projection
for $v \in \mathbf{V}^*$ **do**
    if ($v \in \mathbf{W} \setminus \mathbf{O}$ and v satisfies the W-criterion) **or** ($v \in \mathbf{M}$ and v satisfies the M-criterion) **then**
        $\mathbf{V}^* \leftarrow \mathbf{V}^* \setminus \{v\}$
        $\mathcal{G}^* \leftarrow \mathcal{G}^*_{-v}$
    end if
end for
return $\mathcal{G}^*$

☞ $\mathcal{G}^*_{-v}$ saturates edges from $\mathrm{Pa}(v)$ to $\mathrm{Ch}(v)$ and those within $\mathrm{Ch}(v)$, before removing v.

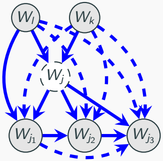
Graph reduction algorithm

$\mathbf{V}^* \leftarrow \{A, Y\} \cup \mathbf{W} \cup \mathbf{M}$
$\mathcal{G}^* \leftarrow \mathcal{G}(\mathbf{V}^*)$ by projecting out $\mathbf{N}$ and $\mathbf{I}$ with latent projection
for $v \in \mathbf{V}^*$ **do**
  if ($v \in \mathbf{W} \setminus \mathbf{O}$ and v satisfies the W-criterion) **or** ($v \in \mathbf{M}$ and v satisfies the M-criterion) **then**
    $\mathbf{V}^* \leftarrow \mathbf{V}^* \setminus \{v\}$
    $\mathcal{G}^* \leftarrow \mathcal{G}^*_{-v}$
  end if
end for
return $\mathcal{G}^*$

☞ $\mathcal{G}^*_{-v}$ saturates edges from $\mathrm{Pa}(v)$ to $\mathrm{Ch}(v)$ and those within $\mathrm{Ch}(v)$, before removing v.



$\Rightarrow$



Graph reduction algorithm

$\mathbf{V}^* \leftarrow \{A, Y\} \cup \mathbf{W} \cup \mathbf{M}$

$\mathcal{G}^* \leftarrow \mathcal{G}(\mathbf{V}^*)$ by projecting out $\mathbf{N}$ and $\mathbf{I}$ with latent projection

for $v \in \mathbf{V}^*$ **do**

  if ($v \in \mathbf{W} \setminus \mathbf{O}$ and v satisfies the W-criterion) **or** ($v \in \mathbf{M}$ and v satisfies the M-criterion) **then**

    $\mathbf{V}^* \leftarrow \mathbf{V}^* \setminus \{v\}$

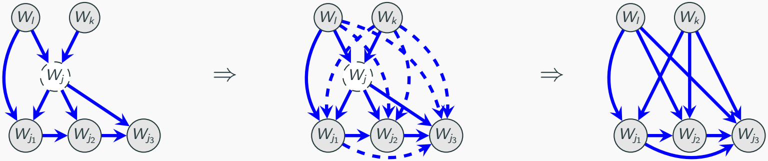
    $\mathcal{G}^* \leftarrow \mathcal{G}^*_{-v}$

  end if

end for

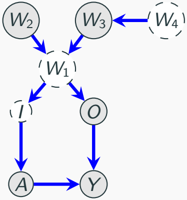
return $\mathcal{G}^*$

☞ $\mathcal{G}^*_{-v}$ saturates edges from $\mathrm{Pa}(v)$ to $\mathrm{Ch}(v)$ and those within $\mathrm{Ch}(v)$, before removing v.



Graph reduction algorithm

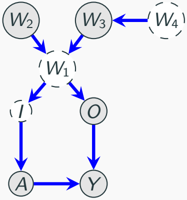
Graph reduction algorithm



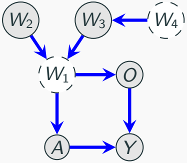
$\mathcal{G}$

$\Rightarrow$

Graph reduction algorithm

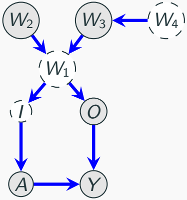


$\Rightarrow$



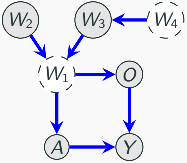
$\Rightarrow$

Graph reduction algorithm

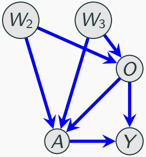


$\mathcal{G}$

$\Rightarrow$

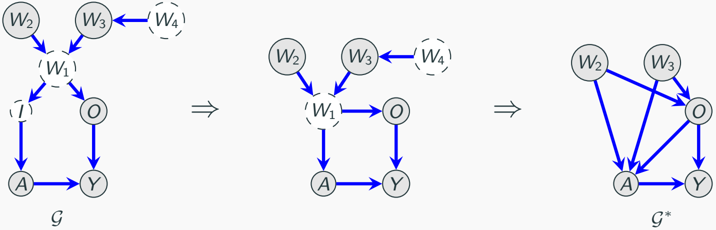


$\Rightarrow$



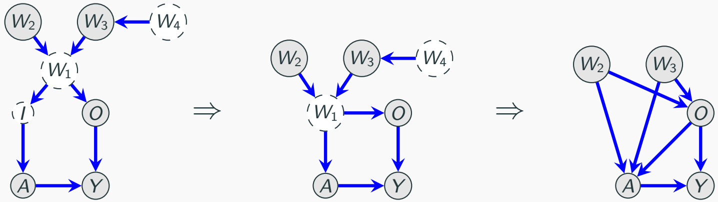
$\mathcal{G}^*$

Graph reduction algorithm



Theorem The reduced graph $\mathcal{G}^*$ is a DAG on vertices $\mathbf{V}^* \equiv \mathbf{V}^*(\mathcal{G})$ with the following properties.

Graph reduction algorithm

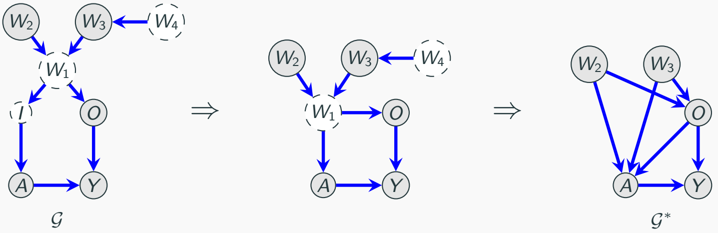


$\mathcal{G}$ $\qquad$ $\mathcal{G}^*$

Theorem The reduced graph $\mathcal{G}^*$ is a DAG on vertices $\mathbf{V}^* \equiv \mathbf{V}^*(\mathcal{G})$ with the following properties.

1. $\mathcal{G}^*$ does not depend on the order that vertices are visited in the Algorithm.

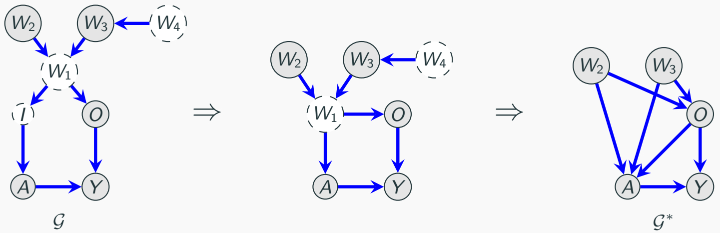
Graph reduction algorithm



Theorem The reduced graph $\mathcal{G}^*$ is a DAG on vertices $\mathbf{V}^* \equiv \mathbf{V}^*(\mathcal{G})$ with the following properties.

1. $\mathcal{G}^*$ does not depend on the order that vertices are visited in the Algorithm.
2. $\mathcal{M}(\mathcal{G}, \mathbf{V}^*) = \mathcal{M}(\mathcal{G}^*, \mathbf{V}^*)$. (☞ when uninformative vars are continuous) (☞ based on mDAGs of Evans, 2016)

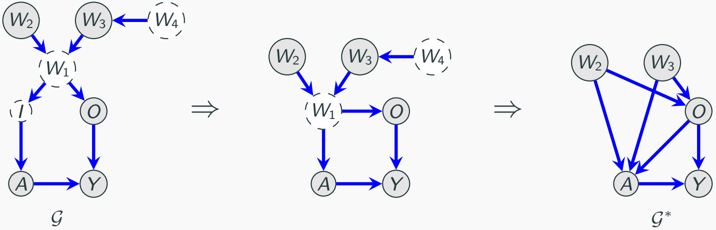
Graph reduction algorithm



Theorem The reduced graph $\mathcal{G}^*$ is a DAG on vertices $\mathbf{V}^* \equiv \mathbf{V}^*(\mathcal{G})$ with the following properties.

1. $\mathcal{G}^*$ does not depend on the order that vertices are visited in the Algorithm.
2. $\mathcal{M}(\mathcal{G}, \mathbf{V}^*) = \mathcal{M}(\mathcal{G}^*, \mathbf{V}^*)$. (☞ when uninformative vars are continuous) (☞ based on mDAGs of Evans, 2016)
3. $\Psi_a(P; \mathcal{G}) = \Psi_a(P; \mathcal{G}^*)$ for every $P \in \mathcal{M}(\mathcal{G}, \mathbf{V})$.

Graph reduction algorithm



Theorem The reduced graph $\mathcal{G}^*$ is a DAG on vertices $\mathbf{V}^* \equiv \mathbf{V}^*(\mathcal{G})$ with the following properties.

1. $\mathcal{G}^*$ does not depend on the order that vertices are visited in the Algorithm.
2. $\mathcal{M}(\mathcal{G}, \mathbf{V}^*) = \mathcal{M}(\mathcal{G}^*, \mathbf{V}^*)$. (☞ when uninformative vars are continuous) (☞ based on mDAGs of Evans, 2016)
3. $\Psi_a(P; \mathcal{G}) = \Psi_a(P; \mathcal{G}^*)$ for every $P \in \mathcal{M}(\mathcal{G}, \mathbf{V})$.
4. $\chi_{\text{eff},P}(\mathbf{V}, \mathcal{G}) = \chi_{\text{eff},P(\mathbf{V}^*)}(\mathbf{V}^*, \mathcal{G}^*)$ P-a.e. for every $P \in \mathcal{M}(\mathcal{G}, \mathbf{V})$.

Simplest efficient g-formula

Simplest efficient g-formula

Corollary $\Psi_a(P; \mathcal{G}^*)$ is the irreducible "efficient" g-formula in the sense that

$$|\Psi_a(\mathbb{P}_n; \mathcal{G}^*) - \Psi_a(\mathbb{P}_n; \mathcal{G})| = o_p(n^{-1/2}) \quad \text{as } n \to \infty$$

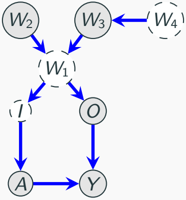
under iid sampling of $P \in \mathcal{M}(\mathcal{G}, \mathbf{V})$ with a finite sample space.

Simplest efficient g-formula

Corollary $\Psi_a(P;\mathcal{G}^*)$ is the irreducible "efficient" g-formula in the sense that

$$|\Psi_a(\mathbb{P}_n;\mathcal{G}^*) - \Psi_a(\mathbb{P}_n;\mathcal{G})| = o_p(n^{-1/2}) \quad \text{as } n \to \infty$$

under iid sampling of $P \in \mathcal{M}(\mathcal{G}, \mathbf{V})$ with a finite sample space.



$\mathcal{G}$

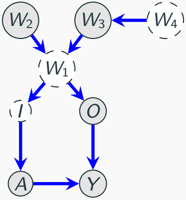
$$\Psi_a(P;\mathcal{G}) = \sum_{y,o,i,w_1,w_2,w_3,w_4} yp(y \mid A=a, o)p(i \mid w_1)p(o \mid w_1) \times p(w_1 \mid w_2, w_3)p(w_3 \mid w_4)p(w_2)p(w_4).$$

Simplest efficient g-formula

Corollary $\Psi_a(P;\mathcal{G}^*)$ is the irreducible "efficient" g-formula in the sense that

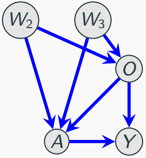
$$|\Psi_a(\mathbb{P}_n;\mathcal{G}^*) - \Psi_a(\mathbb{P}_n;\mathcal{G})| = o_p(n^{-1/2}) \quad \text{as } n \to \infty$$

under iid sampling of $P \in \mathcal{M}(\mathcal{G}, \mathbf{V})$ with a finite sample space.



$\mathcal{G}$

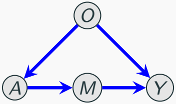
$$\Psi_a(P;\mathcal{G}) = \sum_{y,o,i,w_1,w_2,w_3,w_4} yp(y \mid A = a, o)p(i \mid w_1)p(o \mid w_1) \times p(w_1 \mid w_2, w_3)p(w_3 \mid w_4)p(w_2)p(w_4).$$



$\mathcal{G}^*$

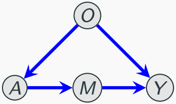
$$\Psi_a(P;\mathcal{G}^*) = \sum_{y,o,w_2,w_3} yP[y \mid A = a, o] \times P(o \mid w_2, w_3)P(w_2)P(w_3).$$

Examples (i)



$\mathcal{G} = \mathcal{G}^*$

Examples (i)

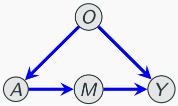


$\mathcal{G} = \mathcal{G}^*$

The g-formula

$$\Psi_a(P;\mathcal{G}) = \sum_{m,o} \mathbb{E}[Y \mid m, o] p(o) p(m \mid a) \text{ is efficient.}$$

Examples (i)



$\mathcal{G} = \mathcal{G}^*$

The g-formula

$$\Psi_a(P;\mathcal{G}) = \sum_{m,o} \mathbb{E}[Y \mid m, o]p(o)p(m \mid a) \text{ is efficient.}$$

☞ Neither the back-door

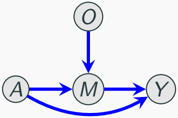
$$\Psi_a^{\text{ADJ}}(P;\mathcal{G}) = \sum_{o} \mathbb{E}[Y \mid a, o]p(o)$$

nor the front-door

$$\Psi_a^{\text{FRONT}}(P;\mathcal{G}) = \sum_{m} \left\{ \sum_{a} \mathbb{E}[Y \mid m, a']p(a') \right\} p(m \mid a)$$

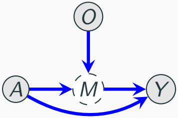
is efficient.

Examples (ii.a)



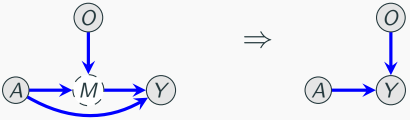
$\mathcal{G}_1$

Examples (ii.a)



$\mathcal{G}_1$

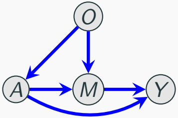
Examples (ii.a)



$\mathcal{G}_1$ $\Rightarrow$ $\mathcal{G}_1^*$

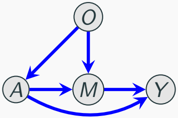
$$\Psi_a(P; \mathcal{G}_1^*) = \sum_o \mathbb{E}[Y \mid A = a, O = o] P(o).$$

Examples (ii.b)



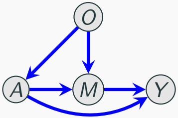
$\mathcal{G}_2$

Examples (ii.b)



$\mathcal{G}_2$

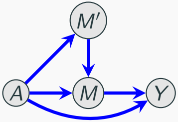
$\Rightarrow$



$\mathcal{G}_2^* = \mathcal{G}_2$

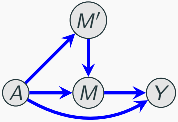
$$\boxed{\Psi_a(P;\mathcal{G}_2^*) = \sum_M \mathbb{E}[Y \mid A = a, M] \sum_O P(M \mid O, a)P(O).}$$

Examples (ii.c)



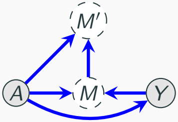
$\mathcal{G}_3$

Examples (ii.c)



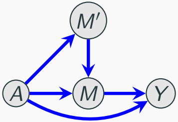
$\mathcal{G}_3$

$\overset{\mathrm{c}}{\sim}$



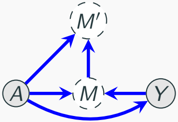
$\check{\mathcal{G}}_3$

Examples (ii.c)



$\mathcal{G}_3$

$\overset{c}{\sim}$



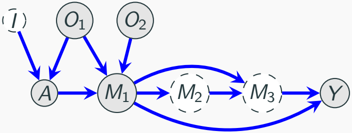
$\check{\mathcal{G}}_3$

$\Rightarrow$

$\mathcal{G}_3^*$

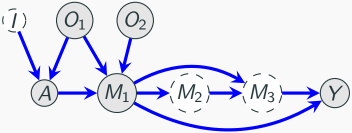
$$\Psi_a(P; \mathcal{G}_3^*) = \mathbb{E}[Y \mid A = a].$$

Examples (iii)



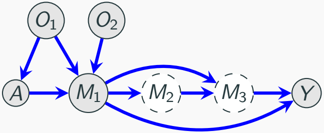
$\mathcal{G}$

Examples (iii)

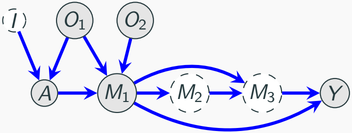


$\mathcal{G}$

$\Rightarrow$

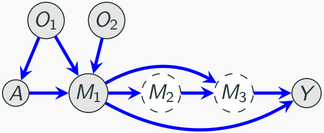


Examples (iii)

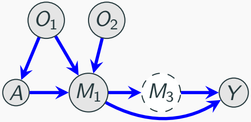


$\mathcal{G}$

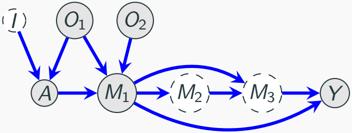
$\Rightarrow$



$\Rightarrow$

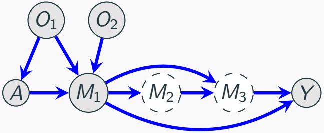


Examples (iii)

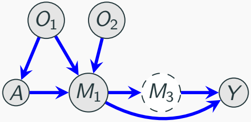


$\mathcal{G}$

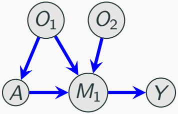
$\Rightarrow$



$\Rightarrow$



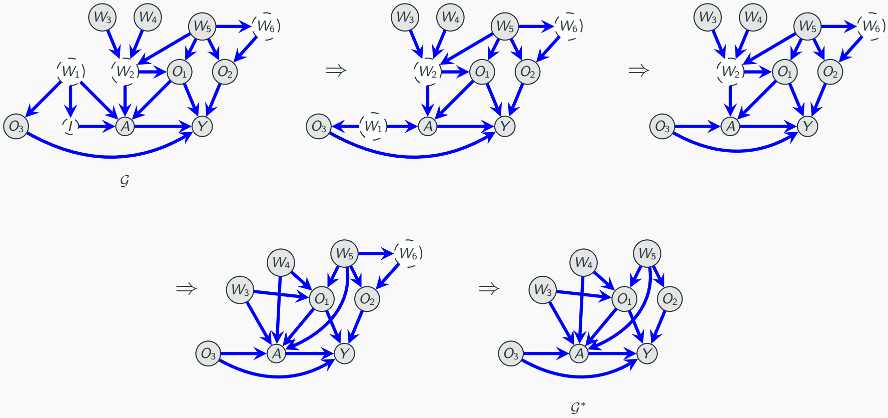
$\Rightarrow$



$\mathcal{G}^*$

$$\Psi_a(P;\mathcal{G}^*) = \sum_{M_1} \mathbb{E}[Y \mid M_1] \sum_{O_1,O_2} P(M_1 \mid O_1, O_2, A = a)P(O_1)P(O_2).$$

Examples (iv)


W_3 W_4 W_5 W_6
W_1 W_2 O_1 O_2
O_3 I A Y
$\mathcal{G}$
$\Rightarrow$
W_3 W_4 W_5 W_6
W_2 O_1 O_2
O_3 W_1 A Y
$\Rightarrow$
W_3 W_4 W_5 W_6
W_2 O_1 O_2
O_3 A Y
$\Rightarrow$
W_4 W_5 W_6
W_3 O_1 O_2
O_3 A Y
$\Rightarrow$
W_4 W_5
W_3 O_1 O_2
O_3 A Y
$\mathcal{G}^*$

reduceDAG

Try simplifying your causal DAG with R package **reduceDAG** available from https://unbiased.co.in

```
library(dagitty)
library(reduceDAG)
g <- dagitty('dag {
    A [pos="0,2", exposure]
    M [pos="1,1"]
    Y [pos="2,2", outcome]
    O [pos="1,0"]
    A -> M -> Y
    A -> Y
    O -> M
}')
cat(gFormula(g))
# sum_{M,Y} Y P[Y | A=a,M] sum_{O} P(M | A=a,O) P(O)
h <- reduceDAG(g, verbose=TRUE)
# Uninformative variables {M} are eliminated.
# Reduced g-formula:
# sum_{O,Y} Y P[Y | A=a,O] P(O)
```

Conclusion

We have studied estimating the counterfactual mean (or the average treatment effect) of a point intervention given a causal DAG.

Conclusion

We have studied estimating the counterfactual mean (or the average treatment effect) of a point intervention given a causal DAG.

- For some graphs, certain variables are uninformative for optimal estimation in large samples.

Conclusion

We have studied estimating the counterfactual mean (or the average treatment effect) of a point intervention given a causal DAG.

- For some graphs, certain variables are uninformative for optimal estimation in large samples.
- We graphically characterized the set of irreducible informative variables $\mathbf{V}^*$.

Conclusion

We have studied estimating the counterfactual mean (or the average treatment effect) of a point intervention given a causal DAG.

- For some graphs, certain variables are uninformative for optimal estimation in large samples.
- We graphically characterized the set of irreducible informative variables $\mathbf{V}^*$.
- The marginal model over $\mathbf{V}^*$ is represented by a DAG $\mathcal{G}^*$.
  ☞ A polynomial time algorithm for constructing $\mathcal{G}^*$.

Conclusion

We have studied estimating the counterfactual mean (or the average treatment effect) of a point intervention given a causal DAG.

- For some graphs, certain variables are uninformative for optimal estimation in large samples.
- We graphically characterized the set of irreducible informative variables $\mathbf{V}^*$.
- The marginal model over $\mathbf{V}^*$ is represented by a DAG $\mathcal{G}^*$.
  ☞ A polynomial time algorithm for constructing $\mathcal{G}^*$.
- For optimal estimation, $\mathcal{G}^*$ is all you need.
  ☞ $\mathcal{G}^*$ prescribes the simplest g-formula that is efficient.
  ☞ $\mathcal{G}^*$ could inform data collection and estimation strategies.

Conclusion

We have studied estimating the counterfactual mean (or the average treatment effect) of a point intervention given a causal DAG.

- For some graphs, certain variables are uninformative for optimal estimation in large samples.
- We graphically characterized the set of irreducible informative variables $\mathbf{V}^*$.
- The marginal model over $\mathbf{V}^*$ is represented by a DAG $\mathcal{G}^*$.
  ☞ A polynomial time algorithm for constructing $\mathcal{G}^*$.
- For optimal estimation, $\mathcal{G}^*$ is all you need.
  ☞ $\mathcal{G}^*$ prescribes the simplest g-formula that is efficient.
  ☞ $\mathcal{G}^*$ could inform data collection and estimation strategies.
- R package `reduceDAG`.

Thanks!

arXiv: 2202.11994

R package: `reduceDAG`

References

- Andersson, Steen A., David Madigan, and Michael D. Perlman (1997). "A characterization of Markov equivalence classes for acyclic digraphs". In: *The Annals of Statistics* 25, pp. 505–541.
- Chickering, David Maxwell (1995). "A transformational characterization of equivalent Bayesian network structures". In: *Proceedings of UAI 1995*.
- Evans, Robin J. (2016). "Graphs for margins of Bayesian networks". In: *Scandinavian Journal of Statistics* 43.3, pp. 625–648.
- Guo, F. Richard and Emilija Perković (2021). "Minimal enumeration of all possible total effects in a Markov equivalence class". In: *Proceedings of the International Conference on Artificial Intelligence and Statistics*.

References ii

Henckel, Leonard, Emilija Perković, and Marloes H. Maathuis (2022). "Graphical criteria for efficient total effect estimation via adjustment in causal linear models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* published online.

Pearl, Judea (1993). "Comment: graphical models, causality and intervention". In: *Statistical Science* 8.3, pp. 266–269.

— (1995). "Causal diagrams for empirical research". In: *Biometrika* 82.4, pp. 669–688.

— (2000). *Causality: Models, Reasoning, and Inference*. 1st. Cambridge University Press.

Robins, James M. (1986). "A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect". In: *Mathematical Modelling* 7, pp. 1393–1512.

Rotnitzky, Andrea and Ezequiel Smucler (2020). "Efficient Adjustment Sets for Population Average Causal Treatment Effect Estimation in Graphical Models". In: *Journal of Machine Learning Research* 21.188, pp. 1–86.

Spirtes, Peter, Clark Glymour, and Richard Scheines (2000). *Causation, Prediction, and Search*. New York: Springer-Verlag.

Verma, Thomas and Judea Pearl (1991). *Equivalence and synthesis of causal models*. Tech. rep. R-150. Department of Computer Science, University of California, Los Angeles.

W-criterion Suppose $\{W_j\} \cup \mathrm{Ch}(W_j) \cap \mathbf{W}$ is topologically sorted as $\{W_{j_0} \equiv W_j, W_{j_1}, \ldots, W_{j_r}\}$. Then $W_j \in \mathbf{W} \setminus \mathbf{O}$ is uninformative if and only if

1. $W_j \perp\!\!\!\perp_{\mathcal{G}} \mathbf{O} \mid W_{j_r}, \mathrm{Pa}(W_{j_r}) \setminus \{W_j\}$,
2. and for $m = 1, \ldots, r$:
   - (i) $W_{j_{m-1}} \to W_{j_m}$ (children are chained)
   - (ii) $\mathrm{Pa}(W_{j_m}) \subseteq \mathrm{Pa}(W_{j_{m-1}}) \cup \{W_{j_{m-1}}\}$ (parent sets are decreasing)
   - (iii) $\mathrm{Pa}(W_{j_{m-1}}) \setminus \mathrm{Pa}(W_{j_m}) \perp\!\!\!\perp_{\mathcal{G}} \mathbf{O} \mid \mathrm{Pa}(W_{j_m})$ (left-over piece is separated from $\mathbf{O}$)

M-criterion Suppose $\{M_i\} \cup \mathrm{Ch}(M_i) \cap \mathbf{M}$ is topologically sorted as $\{M_{i_0} \equiv M_i, M_{i_1}, \ldots, M_{i_k}\}$. Then $M_i \in \mathbf{M}$ is uninformative if and only if

1. $M_i \perp\!\!\!\perp_{\mathcal{G}} \{A, Y\} \cup \mathbf{O}_{\min} \mid M_{i_k}, \mathrm{Pa}(M_{i_k}) \setminus \{M_i\}$,
2. and for $l = 1, \ldots, k$:
   - (i) $M_{i_{l-1}} \to M_{i_l}$ (children are chained)
   - (ii) $\mathrm{Pa}(M_{i_l}) \subseteq \mathrm{Pa}(M_{i_{l-1}}) \cup \{M_{i_{l-1}}\}$ (parent sets are decreasing)
   - (iii) $\mathrm{Pa}(M_{i_{l-1}}) \setminus \mathrm{Pa}(M_{i_l}) \perp\!\!\!\perp_{\mathcal{G}} \{A, Y\} \cup \mathbf{O}_{\min} \mid \mathrm{Pa}(M_{i_l})$ (left-over piece is separated from $A, Y, \mathbf{O}_{\min}$)

Nonparametric model $\mathcal{M}_0(\mathbf{V}) \equiv \{\text{all laws over vector } \mathbf{V}\}$.

Identifying formula Fix a model $\mathcal{M}(\mathbf{V}) \subseteq \mathcal{M}_0(\mathbf{V})$ and a functional $\gamma(P) : \mathcal{M}(\mathbf{V}) \to \mathbb{R}$.

Functional $\chi(P) : \mathcal{M}_0(\mathbf{V}) \to \mathbb{R}$ is an identifying formula for $\gamma(P)$ if $\chi(P) = \gamma(P)$ for every $P \in \mathcal{M}(\mathbf{V})$.

Efficient identifying formula Consider a semiparametric model $\mathcal{M}(\mathbf{V}) \subseteq \mathcal{M}_0(\mathbf{V})$ and a regular functional $\gamma : \mathcal{M}(\mathbf{V}) \to \mathbb{R}$. Let $\gamma^1_{P,\text{eff}}(\mathbf{V})$ be its efficient influence function with respect to $\mathcal{M}(\mathbf{V})$.

An identifying formula $\chi : \mathcal{M}_0(\mathbf{V}) \to \mathbb{R}$ for functional γ is called efficient if $\chi^1_{P,\text{NP}}(\mathbf{V}) = \gamma^1_{P,\text{eff}}(\mathbf{V})$ P-almost-everywhere for every $P \in \mathcal{M}(\mathbf{V})$.