Towards graphical rules for efficient estimation in causal graphical models

Andrea Rotnitzky .

Universidad Di Tella and Harvard T.H. Chan School of Public Health

Based on

Rotnitzky and Smucler, 2020, Journal of Machine Learning Research, 21 188: 1-86,

Smucler, Sapienza and Rotnitzky, 2021, Biometrika, 109, 1, 49–65.

Guo, Perkovic and Rotnitzky, 2022, https://arxiv.org/abs/2202.11994

BIRS, Kelowna, May 23, 2022

Causality in the 21st century

- $1/2$ a century ago different disciplines had their own opinions about causal inference.

Causality in the 21st century

- $1/2$ a century ago different disciplines had their own opinions about causal inference.
- Today there is nearly unanimous acceptance.

Causality in the 21st century

- 1/2 a century ago different disciplines had their own opinions about causal inference.
- Today there is nearly unanimous acceptance.
- "Causal revolution" in great part due to the emergence and adoption of two formalisms:
  - Counterfactual Models
  - Graphical Models

Graphical Models

- **In epidemiology and medical research**: graphical models are responsible for the acceptance and adoption of modern causal analytic techniques because they facilitate encoding complex causal assumptions and reasoning in an intuitive way

Graphical Models

- **In epidemiology and medical research**: graphical models are responsible for the acceptance and adoption of modern causal analytic techniques because they facilitate encoding complex causal assumptions and reasoning in an intuitive way

- S*imple graphical rules* exist to explain the potential biases of one's preferred estimation procedure and the possible remedial approaches.

Graphical Models

- **In epidemiology and medical research**: graphical models are responsible for the acceptance and adoption of modern causal analytic techniques because they facilitate encoding complex causal assumptions and reasoning in an intuitive way

- S*imple graphical rules* exist to explain the potential biases of one's preferred estimation procedure and the possible remedial approaches.

- No graphical rules existed to explain efficiency (variance) in estimation

Graphical Models

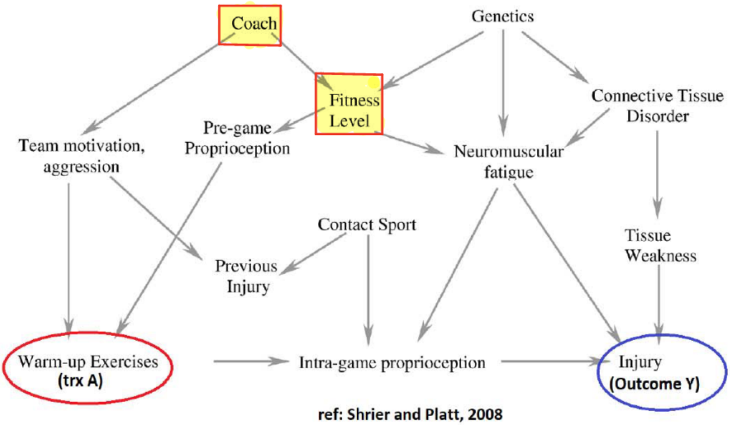
- **In epidemiology and medical research**: graphical models are responsible for the acceptance and adoption of modern causal analytic techniques because they facilitate encoding complex causal assumptions and reasoning in an intuitive way

- *Simple graphical rules* exist to explain the potential biases of one's preferred estimation procedure and the possible remedial approaches.

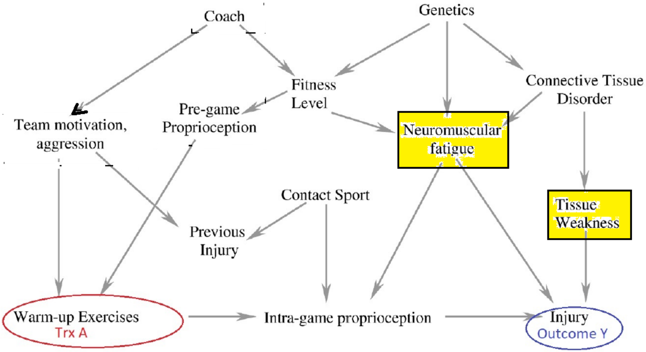
- No graphical rules existed to explain efficiency (variance) in estimation

- In this talk: some work towards filling this gap

An adjustment set



Another adjustment set

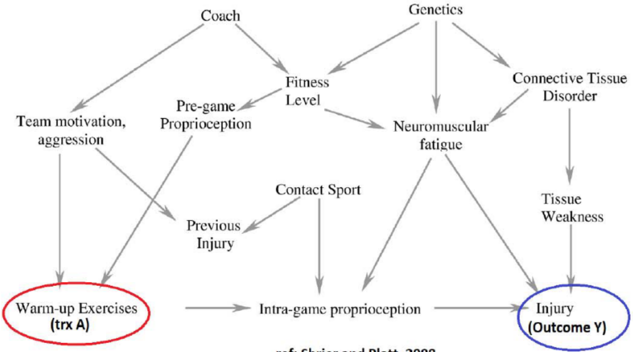


Graph taken from Shrier and Platt, 2008.

Road map of the talk

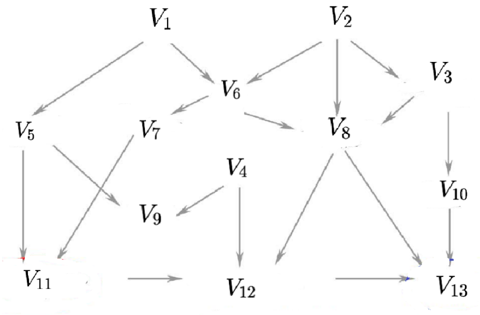
- **Gentle introduction to causal graphical models.**
- Some results with Smucler and Sapienza on optimal adjustment sets
  - Rules for comparing adjustment sets for point exposure studies
  - Time dependent adjustment sets for time dependent exposures
- Some results with Guo and Perkovic on uninformative variables and graph reduction
- Final remarks

Causal Graphical Models in a nutshell



ref: Shrier and Platt, 2008

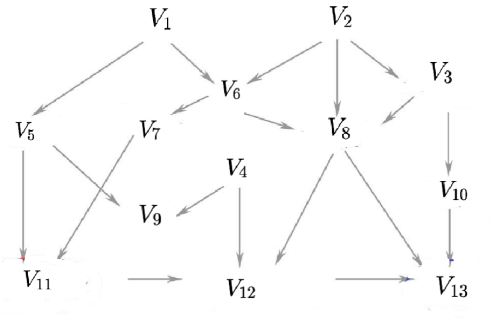
Causal Graphical Models in a nutshell



$$V_1 = f_1(\varepsilon_1)$$
$$V_2 = f_2(\varepsilon_2)$$
$$V_3 = f_3(\varepsilon_3)$$
$$V_4 = f_4(\varepsilon_4)$$
$$V_5 = f_5(V_1, \varepsilon_5)$$
$$\vdots$$
$$V_{11} = f_{11}(V_5, V_7, \varepsilon_{11})$$
$$V_{12} = f_{12}(V_{11}, V_4, \varepsilon_{12})$$
$$V_{13} = f_{13}(V_8, V_{10}, V_{12}, \varepsilon_{13})$$

$\varepsilon_1, \ldots, \varepsilon_{13}$ omitted
non- common causes

Causal Graphical Models in a nutshell

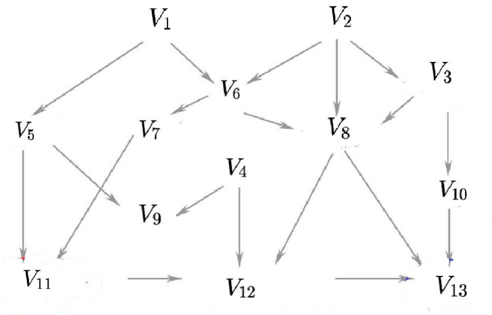


$$V_1 = f_1(\varepsilon_1)$$
$$V_2 = f_2(\varepsilon_2)$$
$$V_3 = f_3(\varepsilon_3)$$
$$V_4 = f_4(\varepsilon_4)$$
$$V_5 = f_5(V_1, \varepsilon_5)$$
$$\vdots$$
$$V_{11} = f_{11}(V_5, V_7, \varepsilon_{11})$$
$$V_{12} = f_{12}(V_{11}, V_4, \varepsilon_{12})$$
$$V_{13} = f_{13}(V_8, V_{10}, V_{12}, \varepsilon_{13})$$

$\varepsilon_1, \ldots, \varepsilon_{13}$ omitted
non- common causes

No omitted common cause assumption formalized as: the ε_j's are mutually independent (Pearl, 1995)

Causal Graphical Models in a nutshell

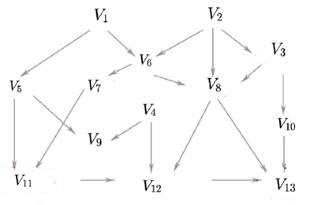


$$V_1 = f_1(\varepsilon_1)$$
$$V_2 = f_2(\varepsilon_2)$$
$$V_3 = f_3(\varepsilon_3)$$
$$V_4 = f_4(\varepsilon_4)$$
$$V_5 = f_5(V_1, \varepsilon_5)$$
$$\vdots$$
$$V_{11} = f_{11}(V_5, V_7, \varepsilon_{11})$$
$$V_{12} = f_{12}(V_{11}, V_4, \varepsilon_{12})$$
$$V_{13} = f_{13}(V_8, V_{10}, V_{12}, \varepsilon_{13})$$

$\varepsilon_1, \ldots, \varepsilon_{13}$ omitted
non- common causes

No omitted common cause assumption formalized as: the $\varepsilon_j's$ are mutually independent (Pearl, 1995)

Causal Graphical Models in a nutshell



$$
\begin{aligned}
V_1 &= f_1(\varepsilon_1)\\
V_2 &= f_2(\varepsilon_2)\\
V_3 &= f_3(\varepsilon_3)\\
V_4 &= f_4(\varepsilon_4)\\
V_5 &= f_5(V_1, \varepsilon_5)\\
&\vdots\\
V_{11} &= f_{11}(V_5, V_7, \varepsilon_{11})\\
V_{12} &= f_{12}(V_{11}, V_4, \varepsilon_{12})\\
V_{13} &= f_{13}(V_8, V_{10}, V_{12}, \varepsilon_{13})
\end{aligned}
$$

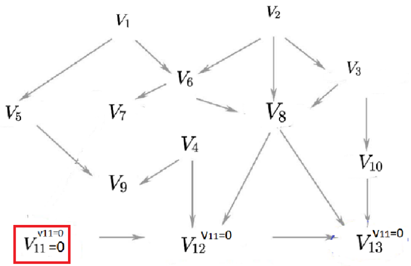
$\varepsilon_1, \ldots, \varepsilon_{13}$ omitted
non- common causes

- **Graphical model with independent ε_j's is tantamount to:**

$$p(\mathbf{v}) = \prod_j p\left(v_j \mid pa_{\mathcal{G}}(v_j)\right)$$

- The collection of laws for V that factor like this is called a Bayesian Network $\mathcal{B}(\mathcal{G})$.

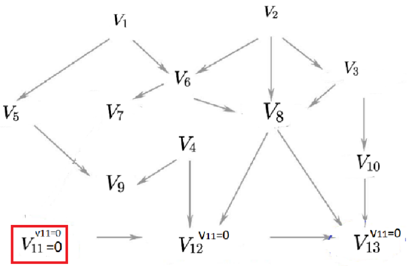
Causal Graphical Models in a nutshell: counterfactual world static intervention



$$V_1 = f_1(\varepsilon_1)$$
$$V_2 = f_2(\varepsilon_2)$$
$$V_3 = f_3(\varepsilon_3)$$
$$V_4 = f_4(\varepsilon_4)$$
$$V_5 = f_5(V_1, \varepsilon_5)$$
$$\vdots$$
$$V_{11}^{v_{11}=0} = 0$$
$$V_{12}^{v_{11}=0} = f_{12}\left(V_{11}^{v_{11}=0}, V_4, \varepsilon_{12}\right)$$
$$V_{13}^{v_{11}=0} = f_{13}\left(V_8, V_{10}, V_{12}^{v_{11}=0}, \varepsilon_{13}\right)$$

$\varepsilon_1, \ldots, \varepsilon_{13}$ omitted
non- common causes

Causal Graphical Models in a nutshell: counterfactual world static intervention



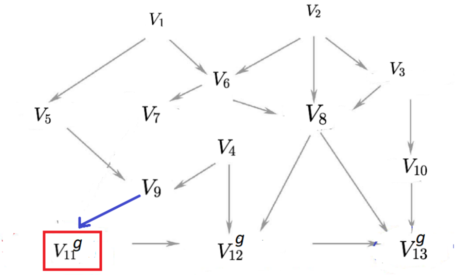
$$V_1 = f_1(\varepsilon_1)$$
$$V_2 = f_2(\varepsilon_2)$$
$$V_3 = f_3(\varepsilon_3)$$
$$V_4 = f_4(\varepsilon_4)$$
$$V_5 = f_5(V_1, \varepsilon_5)$$
$$\vdots$$
$$\boxed{V_{11}^{v_{11}=0} = 0}$$
$$V_{12}^{v_{11}=0} = f_{12}\left(V_{11}^{v_{11}=0}, V_4, \varepsilon_{12}\right)$$
$$V_{13}^{v_{11}=0} = f_{13}\left(V_8, V_{10}, V_{12}^{v_{11}=0}, \varepsilon_{13}\right)$$

$\varepsilon_1, \ldots, \varepsilon_{13}$ omitted
non- common causes

Corollary: counterfactual law is **identified** and given by

$$p_{(v_{11}=0)}(\mathbf{v}) = \prod_{j \neq 11} p\left(v_j | pa_{\mathcal{G}}(v_j)\right) \times I_{\{0\}}(v_{11})$$

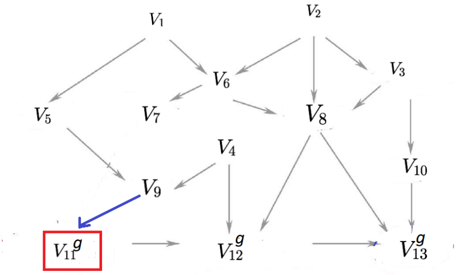
Causal Graphical Models in a nutshell: counterfactual world, deterministic dynamic intervention



$$V_1 = f_1(\varepsilon_1)$$
$$V_2 = f_2(\varepsilon_2)$$
$$V_3 = f_3(\varepsilon_3)$$
$$V_4 = f_4(\varepsilon_4)$$
$$V_5 = f_5(V_1, \varepsilon_5)$$
$$\vdots$$
$$V_{11}^g = g(V_9)$$
$$V_{12}^g = f_{12}(V_{11}^g, V_4, \varepsilon_{12})$$
$$V_{13}^g = f_{13}(V_8, V_{10}, V_{12}^g, \varepsilon_{13})$$

$\varepsilon_1, \ldots, \varepsilon_{13}$ omitted non- common causes

Causal Graphical Models in a nutshell: counterfactual world, deterministic dynamic intervention



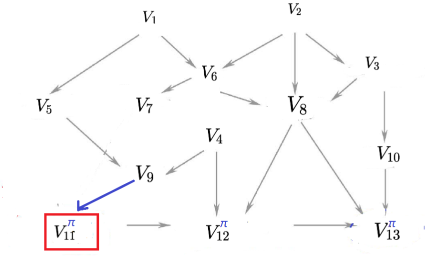
$$V_1 = f_1(\varepsilon_1)$$
$$V_2 = f_2(\varepsilon_2)$$
$$V_3 = f_3(\varepsilon_3)$$
$$V_4 = f_4(\varepsilon_4)$$
$$V_5 = f_5(V_1, \varepsilon_5)$$
$$\vdots$$
$$V_{11}^g = g(V_9)$$
$$V_{12}^g = f_{12}(V_{11}^g, V_4, \varepsilon_{12})$$
$$V_{13}^g = f_{13}(V_8, V_{10}, V_{12}^g, \varepsilon_{13})$$

$\varepsilon_1, \ldots, \varepsilon_{13}$ omitted
non- common causes

Corollary: counterfactual law is **identified** and given by

$$p_g(\mathbf{v}) = \prod_{j \neq 11} p(v_j | pa_{\mathcal{G}}(v_j)) \times I_{\{g(v_9)\}}(v_{11})$$

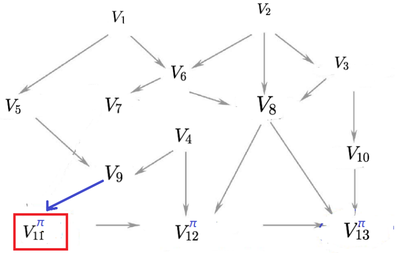
Causal Graphical Models in a nutshell: counterfactual world, random dynamic intervention



$$V_1 = f_1(\varepsilon_1)$$
$$V_2 = f_2(\varepsilon_2)$$
$$V_3 = f_3(\varepsilon_3)$$
$$V_4 = f_4(\varepsilon_4)$$
$$V_5 = f_5(V_1, \varepsilon_5)$$
$$\vdots$$
$$V_{11}^{\pi} = g(V_9, U_{11})$$
$$V_{12}^{\pi} = f_{12}(V_{11}^{\pi}, V_4, \varepsilon_{12})$$
$$V_{13}^{\pi} = f_{13}(V_8, V_{10}, V_{12}^{\pi}, \varepsilon_{13})$$

$\varepsilon_1, \ldots, \varepsilon_{13}$ omitted
non- common causes

Causal Graphical Models in a nutshell: counterfactual world, random dynamic intervention



$$V_1 = f_1(\varepsilon_1)$$
$$V_2 = f_2(\varepsilon_2)$$
$$V_3 = f_3(\varepsilon_3)$$
$$V_4 = f_4(\varepsilon_4)$$
$$V_5 = f_5(V_1, \varepsilon_5)$$
$$\vdots$$
$$V_{11}^{\pi} = g(V_9, U_{11})$$
$$V_{12}^{\pi} = f_{12}(V_{11}^{\pi}, V_4, \varepsilon_{12})$$
$$V_{13}^{\pi} = f_{13}(V_8, V_{10}, V_{12}^{\pi}, \varepsilon_{13})$$

$\varepsilon_1, \ldots, \varepsilon_{13}$ omitted
non- common causes

Corollary: counterfactual law is **identified** and given by

$$p_{\pi}(\mathbf{v}) = \prod_{j \neq 11} p\left(v_j | pa_{\mathcal{G}}(v_j)\right) \times \pi(v_{11} | v_9)$$

Causal graphical models

Causal graphical models

a. **Factual world**. The law p of $\mathbf{V} = (V_1, ..., V_J)$ belongs to *Bayesian Network* $\mathcal{B}(\mathcal{G})$, i.e. it factorizes as

$$p(\mathbf{v}) = \prod_{j=1}^{J} p\left(v_j | pa_{\mathcal{G}}(v_j)\right)$$

where $pa_{\mathcal{G}}(V_j)$ are the parents of V_j in $\mathcal{G}$.

Causal graphical models

a. **Factual world**. The law p of $\mathbf{V} = (V_1, ..., V_J)$ belongs to *Bayesian Network* $\mathcal{B}(\mathcal{G})$, i.e. it factorizes as

$$p(\mathbf{v}) = \prod_{j=1}^{J} p\left(v_j | pa_{\mathcal{G}}(v_j)\right)$$

where $pa_{\mathcal{G}}(V_j)$ are the parents of V_j in $\mathcal{G}$.

b. **Counterfactual world.** For any $\mathbf{A} = (A_1, ..., A_s) \subset \mathbf{V}$, the distrib. of the data when a regime that assigns a_t to A_t with prob. $\pi_t(a_t | \mathbf{Z}_t)$ is implemented in the population (where $\mathbf{Z}_t$ are non-descendants of A_t), is

$$p_\pi(\mathbf{v}) = \prod_{V_j \in \mathbf{V} \setminus \mathbf{A}} p\left(v_j | pa_{\mathcal{G}}(v_j)\right) \times \prod_{t=1}^{s} \pi_t(a_t | \mathbf{z}_t)$$

So, p_π is **identified** from p

Bayesian Network

- *Bayesian Network* $\mathcal{B}(\mathcal{G})$: collection of laws p for V that factorize as

$$p(v) = \prod_{j=1}^{J} p\left(v_j | pa_{\mathcal{G}}(v_j)\right)$$

where $pa_{\mathcal{G}}(V_j)$ are the parents of V_j in DAG $\mathcal{G}$.

Bayesian Network

- *Bayesian Network* $\mathcal{B}(\mathcal{G})$: collection of laws p for V that factorize as

$$p(v) = \prod_{j=1}^{J} p\left(v_j | pa_{\mathcal{G}}(v_j)\right)$$

where $pa_{\mathcal{G}}(V_j)$ are the parents of V_j in DAG $\mathcal{G}$.

- **d-separation:** a sound and complete graphical rule for determining whether a conditional independence holds **under any** $p \in \mathcal{B}(\mathcal{G})$.

Bayesian Network

- *Bayesian Network* $\mathcal{B}(\mathcal{G})$: collection of laws p for V that factorize as
$$p(v) = \prod_{j=1}^{J} p\left(v_j | pa_{\mathcal{G}}(v_j)\right)$$
where $pa_{\mathcal{G}}(V_j)$ are the parents of V_j in DAG $\mathcal{G}$.

- **d-separation:** a sound and complete graphical rule for determining whether a conditional independence holds **under any** $p \in \mathcal{B}(\mathcal{G})$.

- $$A \perp\!\!\!\perp_{\mathcal{G}} B \mid C \quad (A \text{ and } B \text{ are d-separated by } C \text{ in } \mathcal{G})$$

Bayesian Network

- *Bayesian Network* $\mathcal{B}(\mathcal{G})$: collection of laws p for V that factorize as

$$p(v) = \prod_{j=1}^{J} p\left(v_j | pa_{\mathcal{G}}(v_j)\right)$$

where $pa_{\mathcal{G}}(V_j)$ are the parents of V_j in DAG $\mathcal{G}$.

- **d-separation:** a sound and complete graphical rule for determining whether a conditional independence holds **under any** $p \in \mathcal{B}(\mathcal{G})$.
-

$$A \perp\!\!\!\perp_{\mathcal{G}} B \mid C \quad (A \text{ and } B \text{ are d-separated by } C \text{ in } \mathcal{G})$$

- **Theorem (**Geiger, Verma & Pearl, 1990) **:**

$A \perp\!\!\!\perp_{\mathcal{G}} B \mid C \Leftrightarrow$
A is cond. indep. of B given C under any $p \in \mathcal{B}(\mathcal{G})$

d-separation

- A, B single vertices, $C \subset V \setminus \{A, B\}$
- a path from A to B is blocked by C if either

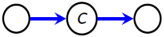
(1) at least one non-collider is in C

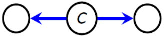
(2) $\exists$ at least one collider, such that neither itself nor its descendants is in C

d-separation

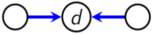
- A, B single vertices, $C \subset V \setminus \{A, B\}$
- a path from A to B is blocked by C if either

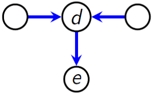
(1) at least one non-collider is in C





(2) $\exists$ at least one collider, such that neither itself nor its descendants is in C



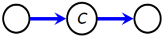


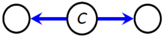
- A and B are d-separated by C if all paths bw A and B are blocked by C

d-separation

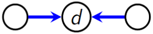
- A, B single vertices, $C \subset V \setminus \{A, B\}$
- a path from A to B is blocked by C if either

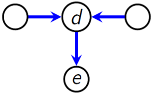
(1) at least one non-collider is in C





(2) $\exists$ at least one collider, such that neither itself nor its descendants is in C





- A and B are d-separated by C if all paths bw A and B are blocked by C
- A set A is d-separated from another set B by $C \subset V \setminus \{A, B\}$ if all $A_j \in A$ and $B_k \in B$ are d-separated by C, in which case we write

$$A \perp\!\!\!\perp_{\mathcal{G}} B \mid C$$

Counterfactual law under a point exposure intervention

- **Counterfactual law.**

$$p_{\pi}\left(\mathbf{v}\right)=\prod_{j:V_j\in\mathbf{V}\setminus A}p\left(v_j|pa_{\mathcal{G}}\left(v_j\right)\right)\times\pi\left(a|\mathbf{z}\right)$$

Counterfactual law under a point exposure intervention

- **Counterfactual law.**

$$p_{\pi}(\mathbf{v}) = \prod_{j:V_j \in \mathbf{V} \setminus A} p\left(v_j | pa_{\mathcal{G}}(v_j)\right) \times \pi(a|\mathbf{z})$$

- Then for $Y = V_J$,

$$E_{\pi}[Y] = \int y \prod_{j:V_j \in \mathbf{V} \setminus A} p\left(v_j | pa_{\mathcal{G}}(v_j)\right) \times \pi(a|\mathbf{z})\, dv$$

Counterfactual law under a point exposure intervention

- **Counterfactual law.**

$$p_{\pi}(\mathbf{v}) = \prod_{j:V_j \in \mathbf{V} \setminus A} p\left(v_j | pa_{\mathcal{G}}\left(v_j\right)\right) \times \pi\left(a | \mathbf{z}\right)$$

- Then for $Y = V_J$,

$$E_{\pi}[Y] = \int y \prod_{j:V_j \in \mathbf{V} \setminus A} p\left(v_j | pa_{\mathcal{G}}\left(v_j\right)\right) \times \pi\left(a | \mathbf{z}\right) dv$$

- But under the Bayesian Network $E_{\pi}(Y)$ is equal to many other functionals

Counterfactual law under a point exposure intervention

- **Counterfactual law.**

$$p_{\pi}(\mathbf{v}) = \prod_{j:V_j \in \mathbf{V} \setminus A} p\left(v_j | pa_{\mathcal{G}}\left(v_j\right)\right) \times \pi\left(a|\mathbf{z}\right)$$

- Then for $Y = V_J$,

$$E_{\pi}[Y] = \int y \prod_{j:V_j \in \mathbf{V} \setminus A} p\left(v_j | pa_{\mathcal{G}}\left(v_j\right)\right) \times \pi\left(a|\mathbf{z}\right) dv$$

- But under the Bayesian Network $E_{\pi}(Y)$ is equal to many other functionals

Adjustment formula and adjustment sets

- **Adjustment formula**:

$$\underbrace{E_\pi[Y]}_{\text{intervention mean}} = \underbrace{\sum_{a=0}^{1}\int E[Y|A=a,\mathbf{L}=\mathbf{l}]\,\pi(a|\mathbf{z})\;p_{\mathbf{L}}(\mathbf{l})\;d\mathbf{l}}_{\text{g-functional}}$$

$$= E_p\left[\frac{\pi(A|\mathbf{Z})}{p(A|\mathbf{L})}Y\right]$$

where $\mathbf{Z} \subset \mathbf{L} \subset \mathbf{V}$

Adjustment formula and adjustment sets

- **Adjustment formula**:

$$\underbrace{E_{\pi}\left[Y\right]}_{\text{intervention mean}} = \underbrace{\sum_{a=0}^{1}\int E\left[Y|A=a,\mathbf{L}=\mathbf{l}\right]\pi\left(a|\mathbf{z}\right)\ p_{\mathbf{L}}\left(\mathbf{l}\right)\ d\mathbf{l}}_{\text{g-functional}}$$

$$= E_{p}\left[\frac{\pi\left(A|\mathbf{Z}\right)}{p\left(A|\mathbf{L}\right)}Y\right]$$

  where $\mathbf{Z} \subset \mathbf{L} \subset \mathbf{V}$
- **Definition:** A $\mathbf{Z}-$ adjustment set for a single trx A and outcome Y is any $\mathbf{L}$ disjoint with A and Y such that
  - $\mathbf{Z} \subset \mathbf{L}$ and,
  - Under the causal graphical model, *for any regime* $\pi\left(A|\mathbf{Z}\right)$, $E_{\pi}\left[Y\right]$ is equal to the corresponding adjustment formula.

Adjustment formula and adjustment sets

- **Adjustment formula**:

$$\underbrace{E_{\pi}[Y]}_{\text{intervention mean}} = \underbrace{\sum_{a=0}^{1} \int E[Y|A=a, \mathbf{L}=\mathbf{l}]\, \pi(a|\mathbf{z})\; p_{\mathbf{L}}(\mathbf{l})\; d\mathbf{l}}_{\text{g-functional}}$$

$$= E_p\left[\frac{\pi(A|\mathbf{Z})}{p(A|\mathbf{L})} Y\right]$$

where $\mathbf{Z} \subset \mathbf{L} \subset \mathbf{V}$

- **Definition:** A $\mathbf{Z}-$ adjustment set for a single trx A and outcome Y is any $\mathbf{L}$ disjoint with A and Y such that
  - $\mathbf{Z} \subset \mathbf{L}$ and,
  - Under the causal graphical model, *for any regime* $\pi(A|\mathbf{Z})$, $E_{\pi}[Y]$ is equal to the corresponding adjustment formula.

- If $\mathbf{Z} = \emptyset$, then we say $\mathbf{L}$ is a *static adjustment set* .

Characterization of Z-adjustment sets

- **Generalized adj. criterion for static (i.e. $\mathbf{Z} = \emptyset$) treatments** (Shpitzer. et. al., 2010, Perkovic et. al., 2015, 2018): $\mathbf{L}$ is static adj. set iff

Characterization of Z-adjustment sets

- **Generalized adj. criterion for static (i.e. $\mathbf{Z} = \emptyset$) treatments** (Shpitzer. et. al., 2010, Perkovic et. al., 2015, 2018): $\mathbf{L}$ is static adj. set iff
  - $\mathbf{L}$ is neither a mediator, nor descendant of Y or of a mediator

Characterization of Z-adjustment sets

- **Generalized adj. criterion for static (i.e. $\mathbf{Z} = \emptyset$) treatments** (Shpitzer. et. al., 2010, Perkovic et. al., 2015, 2018): $\mathbf{L}$ is static adj. set iff
  - $\mathbf{L}$ is neither a mediator, nor descendant of Y or of a mediator
  - $\mathbf{L}$ blocks all non-causal paths between A and Y.

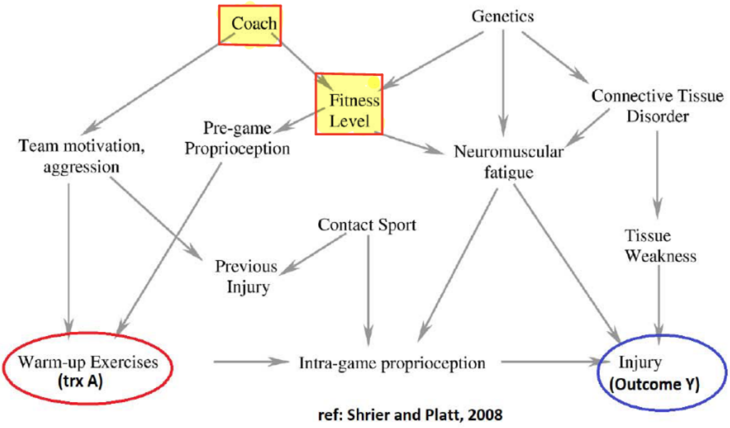
Characterization of Z-adjustment sets

- **Generalized adj. criterion for static (i.e. $\mathbf{Z} = \emptyset$) treatments** (Shpitzer. et. al., 2010, Perkovic et. al., 2015, 2018): $\mathbf{L}$ is static adj. set iff
  - $\mathbf{L}$ is neither a mediator, nor descendant of Y or of a mediator
  - $\mathbf{L}$ blocks all non-causal paths between A and Y.

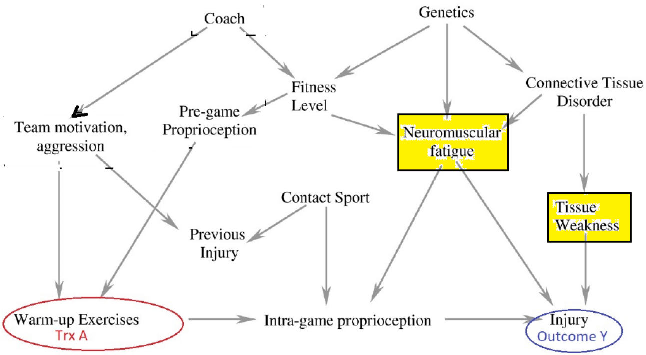
- **Result** (Smucler and Rotnitzky, 2020)**:**

$$\textbf{Class of all } \mathbf{Z}-\textbf{adj sets} = \{\mathbf{L} : \mathbf{L} \text{ is a static adj. set and } \mathbf{Z} \subset \mathbf{L}\}$$

Static adjustment set

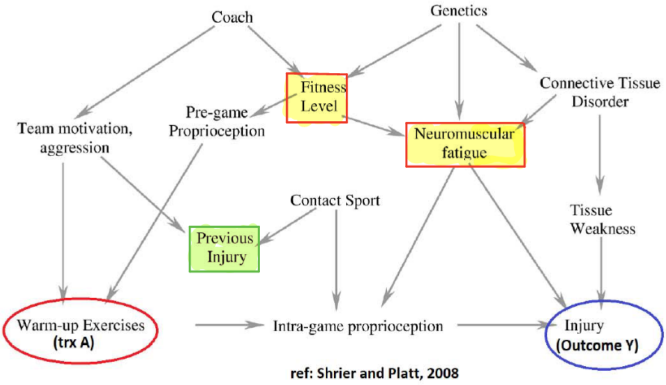


Another static adjustment set

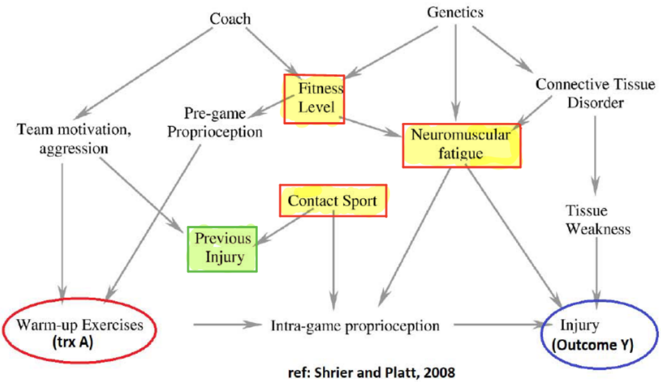


Graph taken from Shrier and Platt, 2008.

An invalid Z-adjustment , Z= previous injury



A valid Z-adjustment set, Z= previous injury



Road map of the talk

- Gentle introduction to causal graphical models.
- **Some results with Smucler and Sapienza on optimal adjustment sets**
  - **Rules for comparing adjustment sets for point exposure studies**
  - Time dependent adjustment sets for time dependent exposures
- Some results with Guo and Perkovic on uninformative variables and graph reduction
- Final remarks

L-NPA estimators of a counterfactual mean

- **Recall:** a $\mathbf{Z}-$ adj. set $\mathbf{L}$ satisfies that for any regime $\pi\left(A|\mathbf{Z}\right)$, the counterfactual mean $E_{\pi}\left(Y\right)$ is equal to

$$\psi_{\pi,\mathbf{L}}\left(P\right) \equiv E_{p}\left[\frac{\pi\left(A|\mathbf{Z}\right)}{p\left(A|\mathbf{L}\right)}Y\right] = \text{g-functional that adjusts for } \mathbf{L}$$

L-NPA estimators of a counterfactual mean

- **Recall:** a $\mathbf{Z}-$ adj. set $\mathbf{L}$ satisfies that for any regime $\pi(A|\mathbf{Z})$, the counterfactual mean $E_\pi(Y)$ is equal to

$$\psi_{\pi,\mathbf{L}}(P) \equiv E_p\left[\frac{\pi(A|\mathbf{Z})}{p(A|\mathbf{L})}Y\right] = \text{g-functional that adjusts for } \mathbf{L}$$

- $\mathbf{L}-$NPA estimators of $\psi_{\pi,\mathbf{L}}(P)$ are those which estimate the prop. score and/or the outcome regression non-parametrically

L-NPA estimators of a counterfactual mean

- **Recall:** a $\mathbf{Z}-$ adj. set $\mathbf{L}$ satisfies that for any regime $\pi(A|\mathbf{Z})$, the counterfactual mean $E_\pi(Y)$ is equal to

$$\psi_{\pi,\mathbf{L}}(P) \equiv E_p\left[\frac{\pi(A|\mathbf{Z})}{p(A|\mathbf{L})}Y\right] = \text{g-functional that adjusts for } \mathbf{L}$$

- $\mathbf{L}-$NPA estimators of $\psi_{\pi,\mathbf{L}}(P)$ are those which estimate the prop. score and/or the outcome regression non-parametrically
- **Key point:** All regular asymptotically linear **L**-NPA estimators of $\psi_{\pi,\mathbf{L}}(P)$ have the **same limiting mean zero normal distribution** with variance denoted, say, as $\sigma^2_{\pi,\mathbf{L}}(p)$

L-NPA estimators of a counterfactual mean

- **Recall:** a $\mathbf{Z}-$ adj. set $\mathbf{L}$ satisfies that for any regime $\pi(A|\mathbf{Z})$, the counterfactual mean $E_\pi(Y)$ is equal to

$$\psi_{\pi,\mathbf{L}}(P) \equiv E_p\left[\frac{\pi(A|\mathbf{Z})}{p(A|\mathbf{L})} Y\right] = \text{g-functional that adjusts for } \mathbf{L}$$

- $\mathbf{L}-$NPA estimators of $\psi_{\pi,\mathbf{L}}(P)$ are those which estimate the prop. score and/or the outcome regression non-parametrically

- **Key point:** All regular asymptotically linear **L**-NPA estimators of $\psi_{\pi,\mathbf{L}}(P)$ have the **same limiting mean zero normal distribution** with variance denoted, say, as $\sigma^2_{\pi,\mathbf{L}}(p)$

- $\sigma^2_{\pi,\mathbf{L}}(p)$ is the variance of the *unique influence function* of the functional $\psi_{\pi,\mathbf{L}}(P)$ under a non-parametric model for P.

L-NPA estimators of a counterfactual mean

- **Recall:** a $\mathbf{Z}-$ adj. set $\mathbf{L}$ satisfies that for any regime $\pi(A|\mathbf{Z})$, the counterfactual mean $E_\pi(Y)$ is equal to

$$\psi_{\pi,\mathbf{L}}(P) \equiv E_p\left[\frac{\pi(A|\mathbf{Z})}{p(A|\mathbf{L})}Y\right] = \text{g-functional that adjusts for } \mathbf{L}$$

- $\mathbf{L}-$NPA estimators of $\psi_{\pi,\mathbf{L}}(P)$ are those which estimate the prop. score and/or the outcome regression non-parametrically
- **Key point:** All regular asymptotically linear **L**-NPA estimators of $\psi_{\pi,\mathbf{L}}(P)$ have the **same limiting mean zero normal distribution** with variance denoted, say, as $\sigma^2_{\pi,\mathbf{L}}(p)$
- $\sigma^2_{\pi,\mathbf{L}}(p)$ is the variance of the *unique influence function* of the functional $\psi_{\pi,\mathbf{L}}(P)$ under a non-parametric model for P.
- Questions that we addressed:.

L-NPA estimators of a counterfactual mean

- **Recall:** a $\mathbf{Z}-$ adj. set $\mathbf{L}$ satisfies that for any regime $\pi\left(A|\mathbf{Z}\right)$, the counterfactual mean $E_{\pi}\left(Y\right)$ is equal to

$$\psi_{\pi,\mathbf{L}}\left(P\right) \equiv E_{p}\left[\frac{\pi\left(A|\mathbf{Z}\right)}{p\left(A|\mathbf{L}\right)}Y\right] = \text{g-functional that adjusts for } \mathbf{L}$$

- $\mathbf{L}-$NPA estimators of $\psi_{\pi,\mathbf{L}}\left(P\right)$ are those which estimate the prop. score and/or the outcome regression non-parametrically

- **Key point:** All regular asymptotically linear **L**-NPA estimators of $\psi_{\pi,\mathbf{L}}\left(P\right)$ have the **same limiting mean zero normal distribution** with variance denoted, say, as $\sigma^{2}_{\pi,\mathbf{L}}\left(p\right)$

- $\sigma^{2}_{\pi,\mathbf{L}}\left(p\right)$ is the variance of the *unique influence function* of the functional $\psi_{\pi,\mathbf{L}}\left(P\right)$ under a non-parametric model for P.

- Questions that we addressed:.
  - Given two adjustment sets, are there graphical rules to determine which one yields an estimator with smaller variance?

L-NPA estimators of a counterfactual mean

- **Recall:** a $\mathbf{Z}-$ adj. set $\mathbf{L}$ satisfies that for any regime $\pi(A|\mathbf{Z})$, the counterfactual mean $E_\pi(Y)$ is equal to

$$\psi_{\pi,\mathbf{L}}(P) \equiv E_p\left[\frac{\pi(A|\mathbf{Z})}{p(A|\mathbf{L})}Y\right] = \text{g-functional that adjusts for } \mathbf{L}$$

- $\mathbf{L}-$NPA estimators of $\psi_{\pi,\mathbf{L}}(P)$ are those which estimate the prop. score and/or the outcome regression non-parametrically
- **Key point:** All regular asymptotically linear $\mathbf{L}$-NPA estimators of $\psi_{\pi,\mathbf{L}}(P)$ have the **same limiting mean zero normal distribution** with variance denoted, say, as $\sigma^2_{\pi,\mathbf{L}}(p)$
- $\sigma^2_{\pi,\mathbf{L}}(p)$ is the variance of the *unique influence function* of the functional $\psi_{\pi,\mathbf{L}}(P)$ under a non-parametric model for P.
- Questions that we addressed:.
  - Given two adjustment sets, are there graphical rules to determine which one yields an estimator with smaller variance?
  - Is there a universally optimal adjustment set and, if so, what graphical rules determine it?

Related literature

- Henckel, Perkovic and Maathuis (2019) provided graphical rules
  - for comparing certain pairs of static adjustment sets
  - for determining the globally optimal static adjustment set
- Also, Kuroki and Miyakawa, 2003 and Kuroki and Cai 2004.
- These works assume:
  - causal graphical **linear** model, i.e. $V_j = \beta_j^T pa_{\mathcal{G}}(V_j) + \varepsilon_j, \{\varepsilon_j : j\}$ indep.
  - treatment effect estimated via **OLS**

Related literature

- Henckel, Perkovic and Maathuis (2019) provided graphical rules
  - for comparing certain pairs of static adjustment sets
  - for determining the globally optimal static adjustment set
- Also, Kuroki and Miyakawa, 2003 and Kuroki and Cai 2004.
- These works assume:
  - causal graphical **linear** model, i.e. $V_j = \beta_j^T pa_{\mathcal{G}}(V_j) + \varepsilon_j, \{\varepsilon_j : j\}$ indep.
  - treatment effect estimated via **OLS**
- Works connected with efficiency implications of inclusion of overadjustment and precision variables in regression and in semip. estimation of ATE:
  - **Linear regression**: Cochran (1968)
  - **Non-linear regression:** Mantel and Haenszel (1959), Breslow (1982), Gail (1988), Robinson and Jewell (1991), Neuhaseuser and Becher (1997) and De Stavola and Cox, (2008).
  - **Semiparametric estimation of a counterfactual mean and of ATE:** Robins and Rotnitzky (1992), Hahn (1998), White and Lu (2011).

Related literature

- Henckel, Perkovic and Maathuis (2019) provided graphical rules
  - for comparing certain pairs of static adjustment sets
  - for determining the globally optimal static adjustment set
- Also, Kuroki and Miyakawa, 2003 and Kuroki and Cai 2004.
- These works assume:
  - causal graphical **linear** model, i.e. $V_j = \beta_j^T pa_{\mathcal{G}}(V_j) + \varepsilon_j, \{\varepsilon_j : j\}$ indep.
  - treatment effect estimated via **OLS**
- Works connected with efficiency implications of inclusion of overadjustment and precision variables in regression and in semip. estimation of ATE:
  - **Linear regression**: Cochran (1968)
  - **Non-linear regression:** Mantel and Haenszel (1959), Breslow (1982), Gail (1988), Robinson and Jewell (1991), Neuhaseuser and Becher (1997) and De Stavola and Cox, (2008).
  - **Semiparametric estimation of a counterfactual mean and of ATE:** Robins and Rotnitzky (1992), Hahn (1998), White and Lu (2011).

Our work with Smucler and Sapienza on adjustment sets

- **Proved** that Henckel et. al. rules also apply when causal graphical model is agnostic and trx effect estimated via non-parametric $\mathbf{L}-$covariate adjustment .
- **Derived graphical rules and efficient algorithms for finding**:
  - globally optimal adj. sets for personalized $\mathbf{Z}-$ dependent regimes
  - optimal static and personalized adj. sets among observable adj. sets
- **Extended** rules for comparing adjustment sets to time dependent treatments and confounding
- **Proved** that optimal time dependent adj. sets do not always exist
- **Characterized** graphs under which the semip. efficient estimator of the counterfactual mean is asym. equivalent to the optimally adjusted estimator

Supplementing adjustment sets with precision variables.

- **Lemma 1.** Suppose $\mathbf{B}$ is a $\mathbf{Z}-$adj. set and $\mathbf{G}$, disjoint with $\mathbf{B}$, satisfies

$$A \perp\!\!\!\perp_{\mathcal{G}} \mathbf{G} \mid \mathbf{B}$$

then, $\mathbf{G} \cup \mathbf{B}$ is *also a* $\mathbf{Z}-$*adj. set* and for all $p \in \mathcal{B}(\mathcal{G})$ and all regimes $\pi(A|\mathbf{Z})$

$$\sigma^2_{\pi,\mathbf{G}\cup\mathbf{B}}(p) \leq \sigma^2_{\pi,\mathbf{B}}(p)$$

Supplementing adjustment sets with precision variables.

- **Lemma 1.** Suppose $\mathbf{B}$ is a $\mathbf{Z}-$adj. set and $\mathbf{G}$, disjoint with $\mathbf{B}$, satisfies

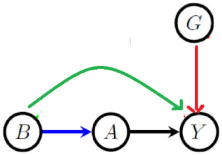
$$A \perp\!\!\!\perp_{\mathcal{G}} \mathbf{G} \mid \mathbf{B}$$

then, $\mathbf{G} \cup \mathbf{B}$ is *also a* $\mathbf{Z}-$*adj. set* and for all $p \in \mathcal{B}(\mathcal{G})$ and all regimes $\pi(A|\mathbf{Z})$

$$\sigma^2_{\pi,\mathbf{G}\cup\mathbf{B}}(p) \leq \sigma^2_{\pi,\mathbf{B}}(p)$$

- In particular, for the *static regime* π that sets A to a,

$$\sigma^2_{\pi,\mathbf{B}}(p) - \sigma^2_{\pi,\mathbf{G}\cup\mathbf{B}}(p) = E\left[\left\{\frac{1}{P(A=a|\mathbf{B})} - 1\right\} var\left\{E(Y|A=a,\mathbf{G},\mathbf{B})\middle| A=a,\mathbf{B}\right\}\right]$$



Deleting overadjustment variables

- **Lemma 2.** Suppose $\mathbf{G}\dot{\cup}\mathbf{B}$ is a $\mathbf{Z}-$adj. set and $\mathbf{B}$ satisfies

$$Y \perp\!\!\!\perp_{\mathcal{G}} \mathbf{B} \mid \mathbf{G}, A$$

  If $\mathbf{Z} \subset \mathbf{G}$, then $\mathbf{G}$ is *also a* $\mathbf{Z}-$*adj. set* and for all $p \in \mathcal{B}(\mathcal{G})$ and all regimes $\pi(A|\mathbf{Z})$

$$\sigma^2_{\pi,\mathbf{G}}(p) \leq \sigma^2_{\pi,\mathbf{G}\cup\mathbf{B}}(p)$$

Deleting overadjustment variables

- **Lemma 2.** Suppose $\mathbf{G}\dot{\cup}\mathbf{B}$ is a $\mathbf{Z}-$adj. set and $\mathbf{B}$ satisfies

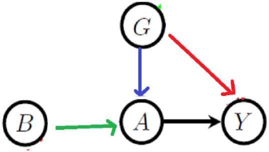
$$Y \perp\!\!\!\perp_{\mathcal{G}} \mathbf{B} \mid \mathbf{G}, A$$

  If $\mathbf{Z} \subset \mathbf{G}$, then $\mathbf{G}$ is *also a* $\mathbf{Z}-$*adj. set* and for all $p \in \mathcal{B}(\mathcal{G})$ and all regimes $\pi(A|\mathbf{Z})$

$$\sigma^2_{\pi,\mathbf{G}}(p) \leq \sigma^2_{\pi,\mathbf{G}\cup\mathbf{B}}(p)$$

- In particular, for the *static regime* π that sets A to a,

$$\sigma^2_{\pi,\mathbf{G}\cup\mathbf{B}} - \sigma^2_{\pi,\mathbf{G}} = E\left[var(Y|A=a,\mathbf{G})\left\{\frac{1}{P(A=a|\mathbf{B},\mathbf{G})} - \frac{1}{P(A=a|\mathbf{G})}\right\}\right]$$



Comparing two arbitrary adjustment sets

- **Corollary:** Suppose that $\mathbf{G}$ and $\mathbf{B}$ are two $\mathbf{Z}-$adj. sets such that

$$A \perp\!\!\!\perp_{\mathcal{G}} (\mathbf{G}\backslash\mathbf{B}) \mid \mathbf{B}$$

  and

$$Y \perp\!\!\!\perp_{\mathcal{G}} (\mathbf{B}\backslash\mathbf{G}) \mid \mathbf{G}, A$$

  Then, for all $p \in \mathcal{B}(\mathcal{G})$ and all regimes $\pi(A|\mathbf{Z})$

$$\sigma^2_{\pi,\mathbf{G}}(p) \leq \sigma^2_{\pi,\mathbf{B}}(p)$$

Comparing two arbitrary adjustment sets

- **Corollary:** Suppose that $\mathbf{G}$ and $\mathbf{B}$ are two $\mathbf{Z}-$adj. sets such that

$$A \perp\!\!\!\perp_{\mathcal{G}} (\mathbf{G}\backslash\mathbf{B}) \mid \mathbf{B}$$

and

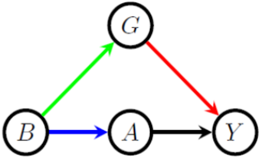
$$Y \perp\!\!\!\perp_{\mathcal{G}} (\mathbf{B}\backslash\mathbf{G}) \mid \mathbf{G}, A$$

Then, for all $p \in \mathcal{B}(\mathcal{G})$ and all regimes $\pi(A|\mathbf{Z})$

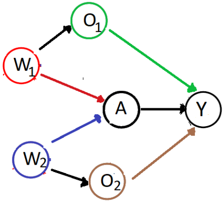
$$\sigma^2_{\pi,\mathbf{G}}(p) \leq \sigma^2_{\pi,\mathbf{B}}(p)$$

- **Proof:**

$$\sigma^2_{\pi,\mathbf{B}} - \sigma^2_{\pi,\mathbf{G}} = \underbrace{\sigma^2_{\pi,\mathbf{B}} - \sigma^2_{\pi,\mathbf{B}\cup(\mathbf{G}\backslash\mathbf{B})}}_{\substack{\text{gain due to supplementation} \\ \text{with precision component } \mathbf{G}\backslash\mathbf{B}}} + \underbrace{\sigma^2_{\pi,\mathbf{G}\cup(\mathbf{B}\backslash\mathbf{G})} - \sigma^2_{\pi,\mathbf{G}}}_{\substack{\text{gain due to deletion} \\ \text{of noisy component } \mathbf{B}\backslash\mathbf{G}}}$$



Not all adjustment sets are comparable



- (O_1, W_2) is preferable to (O_2, W_1) if green association stronger than brown, and blue association weaker than red
- (O_2, W_1) is preferable to (O_1, W_2) if brown association stronger than green, and red association weaker than blue
- but... (O_1, O_2) is more efficient than both

Optimal adjustment set

- **Theorem:** (Henckel, et. al. (2019)). The set

$$\mathbf{O} \quad = \quad \text{non-descendants of } A \text{ that are parents of } Y \text{ or}$$
$$\text{of vertices in the causal path bw } A \text{ and } Y$$

is a *static* adjustment set. Furthermore, for any other static adjustment set **L**,

$$A \perp\!\!\!\perp_{\mathcal{G}} (\mathbf{O} \backslash \mathbf{L}) \mid \mathbf{L}$$

and

$$Y \perp\!\!\!\perp_{\mathcal{G}} (\mathbf{L} \backslash \mathbf{O}) \mid \mathbf{O}, A$$

Optimal adjustment set

- **Theorem:** (Henckel, et. al. (2019)). The set

$$\mathbf{O} \quad = \quad \text{non-descendants of } A \text{ that are parents of } Y \text{ or} \\ \text{of vertices in the causal path bw } A \text{ and } Y$$

  is a *static* adjustment set. Furthermore, for any other static adjustment set $\mathbf{L}$,

$$A \perp\!\!\!\perp_{\mathcal{G}} (\mathbf{O} \backslash \mathbf{L}) \mid \mathbf{L}$$

  and

$$Y \perp\!\!\!\perp_{\mathcal{G}} (\mathbf{L} \backslash \mathbf{O}) \mid \mathbf{O}, A$$

- **Corollary (**Rotnitzky and Smucler, 2020): **O** is the globally optimal static adjustment set.

Optimal adjustment set

- **Theorem:** (Henckel, et. al. (2019)). The set

$$\mathbf{O} = \text{non-descendants of } A \text{ that are parents of } Y \text{ or of vertices in the causal path bw } A \text{ and } Y$$

is a *static* adjustment set. Furthermore, for any other static adjustment set $\mathbf{L}$,

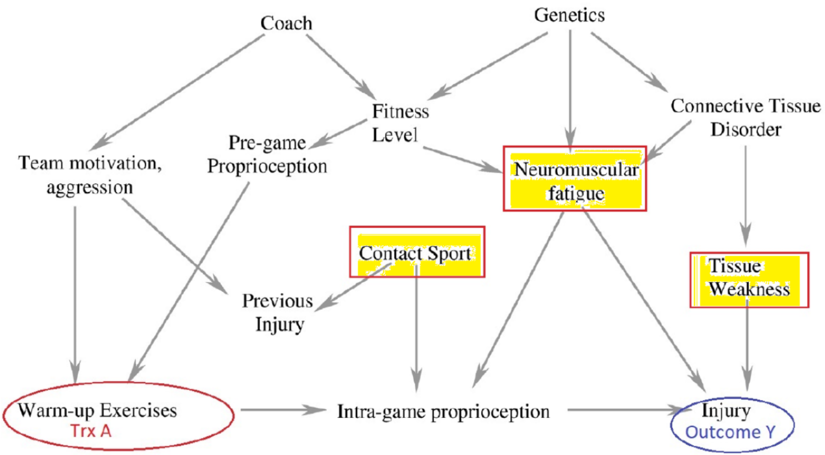
$$A \perp\!\!\!\perp_{\mathcal{G}} (\mathbf{O} \backslash \mathbf{L}) \mid \mathbf{L}$$

and

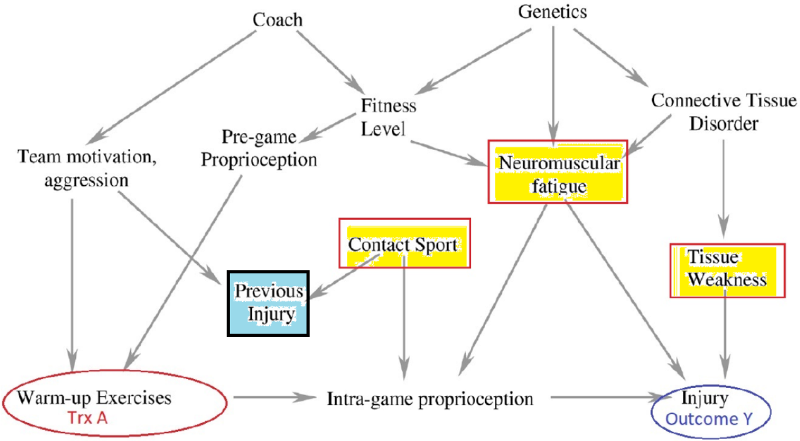
$$Y \perp\!\!\!\perp_{\mathcal{G}} (\mathbf{L} \backslash \mathbf{O}) \mid \mathbf{O}, A$$

- **Corollary (**Rotnitzky and Smucler, 2020): $\mathbf{O}$ is the globally optimal static adjustment set.
- **Lemma** (Smucler, Sapienza and Rotnitzky, 2021): $\mathbf{O} \cup \mathbf{Z}$ is the globally optimal $\mathbf{Z}$ - adjustment set

Globally optimal static adjustment set



Optimal personalized adjustment set



DAGs with hidden variables

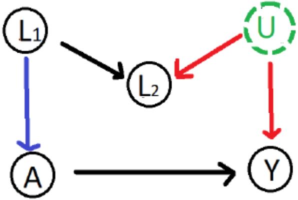
- Suppose that some variables in the DAG are impossible to measure.

DAGs with hidden variables

- Suppose that some variables in the DAG are impossible to measure.
- Then, even if an observable adjustment set exists, a globally optimal adj. set among the **observable** adjustment sets may not exist.

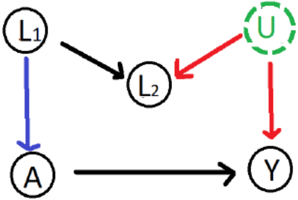
DAGs with hidden variables

- Suppose that some variables in the DAG are impossible to measure.
- Then, even if an observable adjustment set exists, a globally optimal adj. set among the **observable** adjustment sets may not exist.
- **Example:**



DAGs with hidden variables

- Suppose that some variables in the DAG are impossible to measure.
- Then, even if an observable adjustment set exists, a globally optimal adj. set among the **observable** adjustment sets may not exist.
- **Example:**

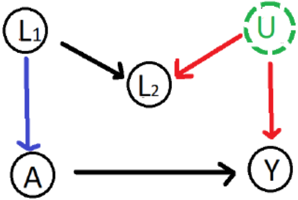


- If U is unobserved, then $\mathbf{L} = \{L_1, L_2\}$ and $\mathbf{L} = \emptyset$ are two valid static adjustment sets which do not dominate each other

DAGs with hidden variables

- Suppose that some variables in the DAG are impossible to measure.
- Then, even if an observable adjustment set exists, a globally optimal adj. set among the **observable** adjustment sets may not exist.

- **Example:**



- If U is unobserved, then $\mathbf{L} = \{L_1, L_2\}$ and $\mathbf{L} = \emptyset$ are two valid static adjustment sets which do not dominate each other
  - $\mathbf{L} = \{L_1\}$ is another adj. set but is dominated by $\mathbf{L} = \emptyset$

Optimal adjustment sets in DAGs with hidden variables

- $\text{An}_{\mathcal{G}}(A, Y, \mathbf{Z})$ = set of nodes that are ancestors of at least one of A, Y or a component of $\mathbf{Z}$
- **Result:** (van der Zander, Liskiewicz and Textor, 2019): if an observable $\mathbf{Z}$−adj. set exists then

$$\mathcal{S} = \{\mathbf{L} : \mathbf{L} \text{ is observable } \mathbf{Z} - \text{adj.set and } \mathbf{L} \subset \text{An}_{\mathcal{G}}(A, Y, \mathbf{Z})\}$$

is not empty.

Optimal adjustment sets in DAGs with hidden variables

- $\text{An}_{\mathcal{G}}(A, Y, \mathbf{Z})$ = set of nodes that are ancestors of at least one of A, Y or a component of $\mathbf{Z}$
- **Result:** (van der Zander, Liskiewicz and Textor, 2019): if an observable $\mathbf{Z}-$adj. set exists then

$$\mathcal{S} = \{\mathbf{L} : \mathbf{L} \text{ is observable } \mathbf{Z} - \text{adj.set and } \mathbf{L} \subset \text{An}_{\mathcal{G}}(A, Y, \mathbf{Z})\}$$

  is not empty.

- **Result (**Smucler et al, 2021)**:** If $\mathcal{S} \neq \emptyset$ then an optimal $\mathbf{Z}-$adj. set exists in the class $\mathcal{S}$.

Optimal adjustment sets in DAGs with hidden variables

- $\text{An}_{\mathcal{G}}(A, Y, \mathbf{Z})$ = set of nodes that are ancestors of at least one of A, Y or a component of $\mathbf{Z}$
- **Result:** (van der Zander, Liskiewicz and Textor, 2019): if an observable $\mathbf{Z}-$adj. set exists then

$$\mathcal{S} = \{\mathbf{L} : \mathbf{L} \text{ is observable } \mathbf{Z} - \text{adj.set and } \mathbf{L} \subset \text{An}_{\mathcal{G}}(A, Y, \mathbf{Z})\}$$

  is not empty.

- **Result (**Smucler et al, 2021**):** If $\mathcal{S} \neq \emptyset$ then an optimal $\mathbf{Z}-$adj. set exists in the class $\mathcal{S}$.
- In Smucler et al, 2021, we derived a graphical algorithm, *based on a particular latent projected undirected moralized graph*, that finds the optimal $\mathbf{Z}$-adj. set in $\mathcal{S}$.

Road map of the talk

- Gentle introduction to causal graphical models.
- **Some results with Smucler and Sapienza on optimal adjustment sets**
  - Rules for comparing adjustment sets for point exposure studies
  - **Time dependent adjustment sets for time dependent exposures**
- Some results with Guo and Perkovic on uninformative variables and graph reduction
- Final remarks

Time dependent treatments

Suppose A_1 and A_2 are two treatments, $A_1 \in \mathrm{nd}_{\mathcal{G}}(A_2)$. Under a causal graphical model represented by DAG G, the mean of Y_{a_0,a_1} when the static regime that sets A_0 to a_0 and A_1 to a_1 is

$$\begin{aligned} E\left(Y_{a_0,a_1}\right) &= E\left\{\frac{I_{a_0}\left(A_0\right)}{p\left(a_0|pa_{\mathcal{G}}\left(A_0\right)\right)}\frac{I_{a_1}\left(A_1\right)}{p\left(a_1|pa_{\mathcal{G}}\left(A_1\right)\right)}Y\right\} \\ &= E\left\{E\left[E\left[Y|a_0,a_1,pa_{\mathcal{G}}\left(A_0\right),pa_{\mathcal{G}}\left(A_1\right)\right]|a_0,pa_{\mathcal{G}}\left(A_0\right)\right]\right\} \end{aligned}$$

Definition: $\mathbf{L} = \left(\mathbf{L}_0, \mathbf{L}_1\right) \subset \mathbf{V}$ is a static time dependent adjustment set relative to trxs $\left(A_0, A_1\right)$ and outcome Y in G iff for all $P \in \mathcal{B}\left(\mathcal{G}\right)$,

$$E\left(Y_{a_0,a_1}\right) = E\left\{E\left[E\left[Y|a_0,a_1,\mathbf{L}_0,\mathbf{L}_1\right]|a_0,\mathbf{L}_0\right]\right\}$$

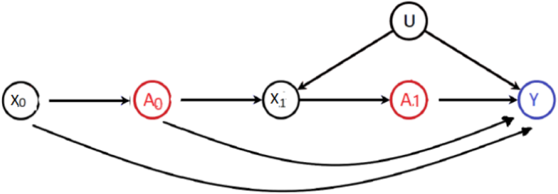
The right hand side is the so-called the g-functional with respect to $\left(\mathbf{L}_0, \mathbf{L}_1\right)$.

Time dependent treatments

Lemma (Robins, 1986) $(\mathbf{L}_0, \mathbf{L}_1)$ is a time-dependent adjustment set if :
(i) $\mathbf{L}_j$ non-descendant of $A_j, j = 0, 1$, and
(ii) Sequential randomization:

$$Y_{a_0,a_1} \amalg A_1 | (A_0, \mathbf{L}_0, \mathbf{L}_1) \text{ and } Y_{a_0,a_1} \amalg A_0 | \mathbf{L}_0$$

Example:



- X_0 is a time 0 adjustment set $(= \mathbf{L}_0)$
- X_1, U and (X_1, U) are time 1 adjustment sets $(= \mathbf{L}_1)$

Time dependent treatments

Lemma: Suppose that $(\mathbf{B}_0, \mathbf{B}_1)$ and $(\mathbf{G}_0, \mathbf{G}_1)$ are time dependent adjustment sets. If

(1)

$$A_0 \perp\!\!\!\perp_{\mathcal{G}} [\mathbf{G}_0 \backslash \mathbf{B}_0] \mid \mathbf{B}_0$$
$$A_1 \perp\!\!\!\perp_{\mathcal{G}} [(\mathbf{G}_0, \mathbf{G}_1) \backslash (\mathbf{B}_0, \mathbf{B}_1)] \mid (\mathbf{B}_0, \mathbf{B}_1, A_0)$$

(2)

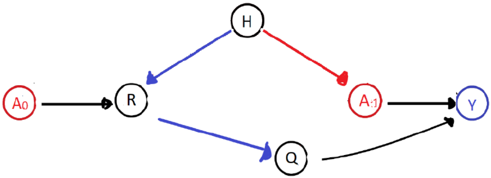
$$\mathbf{G}_1 \perp\!\!\!\perp_{\mathcal{G}} [\mathbf{B}_0 \backslash \mathbf{G}_0] \mid (\mathbf{G}_0, A_0)$$
$$Y \perp\!\!\!\perp_{\mathcal{G}} [(\mathbf{B}_0, \mathbf{B}_1) \backslash (\mathbf{G}_0, \mathbf{G}_1)] \mid (\mathbf{G}_0, \mathbf{G}_1, A_0, A_1)$$

then, for all $P \in \mathcal{B}(\mathcal{G})$

$$\sigma^2_{\mathbf{G}_0, \mathbf{G}_1} \leq \sigma^2_{\mathbf{B}_0, \mathbf{B}_1}$$

where for any adj. set $(\mathbf{L}_0, \mathbf{L}_1)$, $\sigma^2_{\mathbf{L}}$ is the variance of the NP inf. fcn of the g-functional adjusted for $(\mathbf{L}_0, \mathbf{L}_1)$.

Time dependent treatments

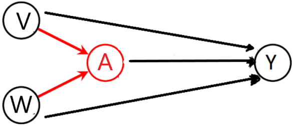


The following adjustment sets dominate all other adjustment sets but they don't dominate each other

Time 0 adj. set $(= \mathbf{L}_0)$	**Time 1 adj. set** $(= \mathbf{L}_1)$	**Better when**
$\emptyset$	Q	red assoc. strong, blue assoc weak
H	Q	red assoc. weak, blue assoc strong

In Rotnitzky and Smucler we exhibited two laws P_1 and P_2 in $\mathcal{B}(\mathcal{G})$ for binary data such that:
(i) under P_1, (H, Q) is 8% more efficient than $(\emptyset, Q)$, and
(ii) under P_2, $(\emptyset, Q)$ is 47% more efficient than (H, Q)

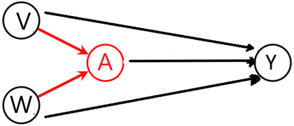
Semip. efficient estimation vs optimal non-parametric adjusted estimation



- The interventional mean $E(Y^a)$ is

$$E\left[E\left(Y|A=a,V,W\right)\right]=\int E\left(Y|A=a,V=v,W=w\right)\underbrace{p\left(v\right)p\left(w\right)}_{=p(v,w)}dvdw$$

Semip. efficient estimation vs optimal non-parametric adjusted estimation

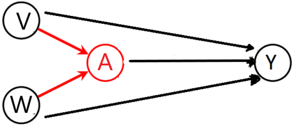


- The interventional mean $E(Y^a)$ is

$$E[E(Y|A=a,V,W)] = \int E(Y|A=a,V=v,W=w)\underbrace{p(v)p(w)}_{=p(v,w)}dvdw$$

- Optimal non-parametric adjusted estimator *ignores* restrictions on the marginal law of covariates, i.e. that V and W are marginally independent.

Semip. efficient estimation vs optimal non-parametric adjusted estimation



- The interventional mean $E\left(Y^{a}\right)$ is

$$E\left[E\left(Y|A=a,V,W\right)\right]=\int E\left(Y|A=a,V=v,W=w\right)\underbrace{p\left(v\right)p\left(w\right)}_{=p\left(v,w\right)}dvdw$$

- Optimal non-parametric adjusted estimator *ignores* restrictions on the marginal law of covariates, i.e. that V and W are marginally independent.

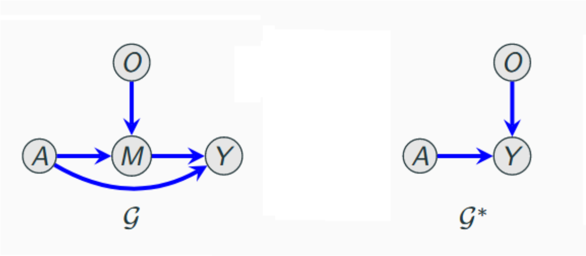
- Semiparametric efficient (SE) *exploits* these restrictions and can be much much more efficient than optimally adjusted NP estimator.

There is also information in the mediators structure

$$
\begin{aligned}
E\left(Y^{a}\right) &= E\left(Y \middle| A=a\right) \\
&= \int\int y \underbrace{p\left(y \middle| m\right) p\left(m \middle| a\right)}_{=p(y,m|a)} dm dy
\end{aligned}
$$

- Markov chain structure carries information about $E\left(Y \middle| A=a\right)$.

However ... in some graphs the optimally adjusted estimator is efficient



- With discrete data the MLE of $p_a(y)$ under $\mathcal{G}$ is

$$\widehat{p}_{a,MLE}(y) = \sum_{m,o} \mathbb{P}_n(y|m,a)\,\mathbb{P}_n(m|a,o)\,\mathbb{P}_n(o)$$

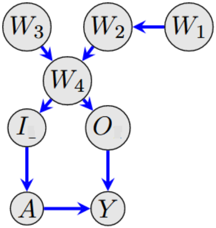
- Surprisingly, $\widehat{p}_{a,MLE}(y)$ is asym. equivalent to the MLE of $p_a(y)$ under $\mathcal{G}^*$ is

$$\widetilde{p}_{a,MLE}(y) = \sum_{o} \mathbb{P}_n(y|o,a)\,\mathbb{P}_n(o)$$

Road map of the talk

- Gentle introduction to causal graphical models.
- **Some results with Smucler and Sapienza on optimal adjustment sets**
  - Rules for comparing adjustment sets for point exposure studies
  - Time dependent adjustment sets for time dependent exposures
- **Some results with Guo and Perkovic on uninformative variables and graph reduction**
- Final remarks

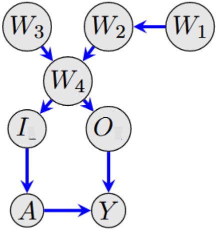
Graph reduction for semiparametric efficient estimation (joint work with Richard Guo and Ema Perkovic)



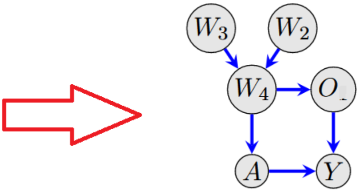
(a) $\mathcal{G}$

$$p_a(y) = \sum_{i, w_1, w_2, w_3, w_4, o} p(y|o, a)\, p(i|w_4)\, p(o|w_4)\, p(w_4|w_2, w_3)\, p(w_3)\, p(w_2|w_1)\, p(w_1)$$

Graph reduction for semiparametric efficient estimation (joint work with Richard Guo and Ema Perkovic)



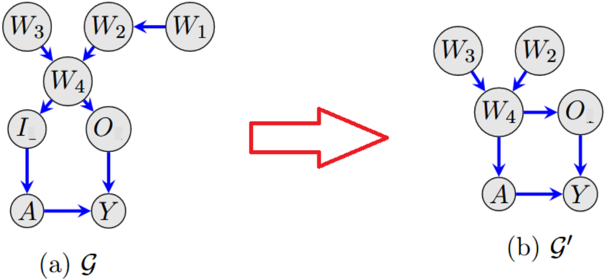
(a) $\mathcal{G}$



(b) $\mathcal{G}'$

$$p_a(y) = \underbrace{\sum_{w_2, w_3, w_4, o} p(y|o, a)\, p(o|w_4)\, p(w_4|w_2, w_3)\, p(w_3)\, p(w_2)}_{\text{g-formula in } \mathcal{G}'}$$

Graph reduction for semiparametric efficient estimation (joint work with Richard Guo and Ema Perkovic)

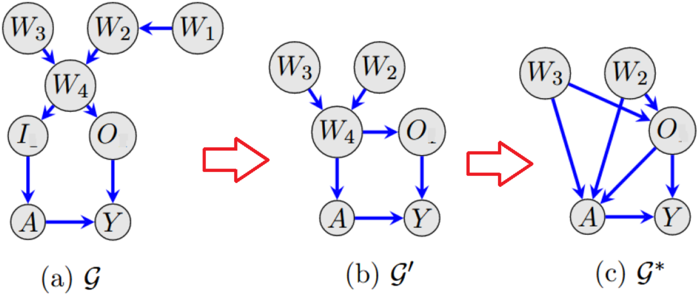


$$p_a(y) = \underbrace{\sum_{w_2, w_3, w_4, o} p(y|o,a)\, p(o|w_4)\, p(w_4|w_2,w_3)\, p(w_3)\, p(w_2)}_{\text{g-formula in } \mathcal{G}'}$$

- With discrete data, MLE under $\mathcal{G}'$ is

$$\widehat{p}_{a,MLE}(y) = \sum_{w_2, w_3, w_4, o} \mathbb{P}_n(y|o,a)\, \mathbb{P}_n(o|w_4)\, \mathbb{P}_n(w_4|w_2,w_3)\, \mathbb{P}_n(w_3)\, \mathbb{P}_n(w_2)$$

Graph reduction for semiparametric efficient estimation (joint work with Richard Guo and Ema Perkovic)



- Surprisingly, MLE under $\mathcal{G}^*$ is asymptotically equivalent to MLE under $\mathcal{G}'$

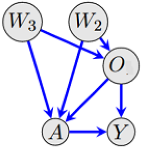
$$\widetilde{p}_{a,MLE}\left(y\right) = \sum_{w_2, w_3, o} \mathbb{P}_n\left(y \middle| o, a\right) \mathbb{P}_n\left(o \middle| w_2, w_3\right) \mathbb{P}_n\left(w_3\right) \mathbb{P}_n\left(w_2\right)$$

Graph reduction for semiparametric efficient estimation of a counterfactual mean

- Given a graph $\mathcal{G}$ we derived an algorithm that outputs another graph $\mathcal{G}^*$over a subset of the variables in $\mathcal{G}$ such that
  - the $g-$formula in $\mathcal{G}^*$ is an identifying formula in $\mathcal{G}$,
  - the semiparametric variance bound for estimation of $E\left(Y^a\right)$ in model $\mathcal{B}\left(\mathcal{G}\right)$ and in model $\mathcal{B}\left(\mathcal{G}^*\right)$ agree
  - $\mathcal{G}^*$ is the smallest such possible graph in the sense that all variables in $\mathcal{G}^*$ are informative. More precisely, the efficient influence function for $E\left(Y^a\right)$ is a function of every variable in $\mathcal{G}^*$ for at least one P in $\mathcal{B}\left(\mathcal{G}^*\right)$

Final remarks

- **Estimation via adjustment vs semip. efficient estimation**:
  - Usual variance/bias trade-off: adjustment relies on less model assumptions
  - Equally or perhaps even more importantly: efficient estimation requires estimation of each cond. density $p\left(V_j | pa_{\mathcal{G}}\left(V_j\right)\right)$. Even debiased, influence-function based, i.e. one-step estimation or TMLE, will hardly control the estimation bias of these densities.

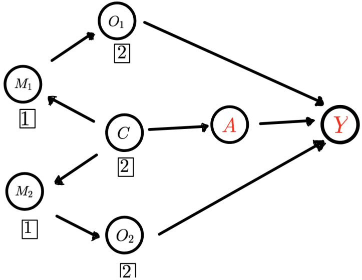


$$\widehat{E}\left(Y^a\right)_{MLE} = \sum_{w_2, w_3, o} \mathbb{E}_n\left(Y | o, a\right) \mathbb{P}_n\left(o | w_2, w_3\right) \mathbb{P}_n\left(w_3\right) \mathbb{P}_n\left(w_2\right)$$

$$\widehat{E}\left(Y^a\right)_{adj} = \sum_{o} \mathbb{E}_n\left(Y | o, a\right) \mathbb{P}_n\left(o\right)$$

Final remarks

- **Study design:** assign cost to each graph variable and find the adjustment set leading to smallest estimation variance:
  - subject to a cost constraint $\longrightarrow$ a universal solution does not exist



  - among adjustment sets of minimum cost $\longrightarrow$ for point exposure we provide the universal solution in Smucler and Rotnitzky, 2022, and

    graphical rules for finding it

Open problems

- Inference about the functional returned by the ID algorithm when no observable adj. set exists
  - Some special cases have been studied, e.g. the generalized front door formula, (Fulcher, et. al. 2019). General theory for an arbitrary functional not yet available.
- Optimal adj. sets and efficient estimation for other parameters e.g., trx effect on the treated, and natural direct and indirect effects

THANKS!

Cuts and moralized graphs.

- **Separation and cuts in undirected graphs**: In an undirected graph $\mathcal{H}$, **A** is separated from **B** by **C**, denoted as

$$\mathbf{A} \perp_{\mathcal{H}} \mathbf{B} \mid \mathbf{C}$$

iff all paths between **A** and **B** have a vertex in **C**. In such case **C** is called a cut between **A** and **B**.

Cuts and moralized graphs.

- **Separation and cuts in undirected graphs**: In an undirected graph $\mathcal{H}$, **A** is separated from **B** by **C**, denoted as

$$\mathbf{A} \perp_{\mathcal{H}} \mathbf{B} | \mathbf{C}$$

  iff all paths between **A** and **B** have a vertex in **C**. In such case **C** is called a cut between **A** and **B**.

- **Moralized graph of a DAG** $\mathcal{G}$ is an undirected graph $\mathcal{G}^m$ with same vertices as $\mathcal{G}$, constructed by keeping the edges of $\mathcal{G}$ but removing their direction and additionally "marrying" the unshielded colliders.

Cuts and moralized graphs.

- **Separation and cuts in undirected graphs**: In an undirected graph $\mathcal{H}$, **A** is separated from **B** by **C**, denoted as

$$\mathbf{A} \perp_{\mathcal{H}} \mathbf{B} | \mathbf{C}$$

  iff all paths between **A** and **B** have a vertex in **C**. In such case **C** is called a cut between **A** and **B**.

- **Moralized graph of a DAG $\mathcal{G}$** is an undirected graph $\mathcal{G}^m$ with same vertices as $\mathcal{G}$, constructed by keeping the edges of $\mathcal{G}$ but removing their direction and additionally "marrying" the unshielded colliders.

- Neighborhood of Y : set of vertices adjacent to Y, denoted with $\partial_{\mathcal{H}}(Y)$

Cuts and moralized graphs.

- **Separation and cuts in undirected graphs**: In an undirected graph $\mathcal{H}$, **A** is separated from **B** by **C**, denoted as

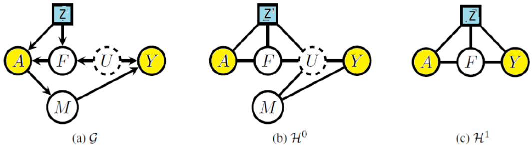
$$\mathbf{A} \perp_{\mathcal{H}} \mathbf{B} | \mathbf{C}$$

  iff all paths between **A** and **B** have a vertex in **C**. In such case **C** is called a cut between **A** and **B**.

- **Moralized graph of a DAG $\mathcal{G}$** is an undirected graph $\mathcal{G}^m$ with same vertices as $\mathcal{G}$, constructed by keeping the edges of $\mathcal{G}$ but removing their direction and additionally "marrying" the unshielded colliders.

- Neighborhood of Y : set of vertices adjacent to Y, denoted with $\partial_{\mathcal{H}}(Y)$

Construction of the latent projected moralized graph



1. $\mathcal{H}^0 \leftarrow \left(\mathcal{G}_{\underline{A}}\left[\mathrm{An}_{\mathcal{G}}\left(A, Y, \mathbf{Z}\right)\right]\right)^m$ (Textor and Liskiewicz, 2011 and van der Zander et al, 2019)
   1.1 compute ancestral subgraph $\mathcal{G}\left[\mathrm{An}_{\mathcal{G}}\left(A, Y, \mathbf{Z}\right)\right]$
   1.2 delete edges pointing out of A
   1.3 moralize the resulting subgraph
2. $\mathcal{H}^1$ constructed from $\mathcal{H}^0$ by
   2.1 Latent project out the hidden nodes and the nodes in $\texttt{forb}(A, Y, \mathcal{G})$
   2.2 Add to latent projected graph edges bw $\mathbf{Z}$ and A and bw $\mathbf{Z}$ and Y