Causal effects in MPDAGs: identification and efficient estimation

Emilija Perković and F. Richard Guo

Department of Statistics, University of Washington

Goal

- Estimate the total causal effect of A on Y

Observational data

Randomized control studies

Goal

- Estimate the total causal effect of A on Y
  - the change in Y due to $do(a)$-
  from observational data.

- $do(a)$: an intervention that sets variables A to a.

Observational data	Randomized control studies

Goal

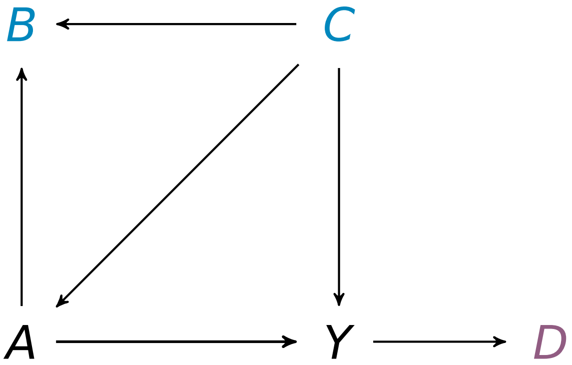
- Estimate the total causal effect of A on Y
  - the change in Y due to $do(a)$-
  from observational data.

- $do(a)$: an intervention that sets variables A to a.
  $f(y|do(a)) \neq f(y|a)$.

Observational data

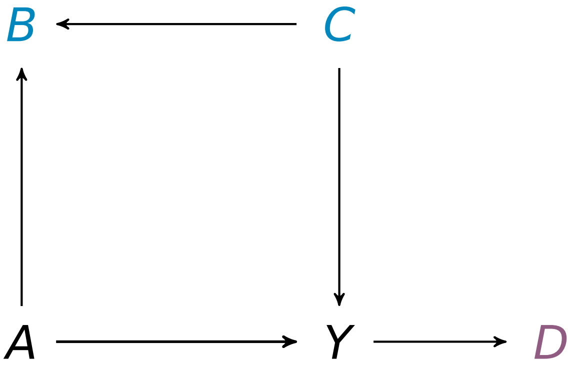
Randomized control studies

Observational Causal DAG



Causal Directed Acyclic Graph (DAG) $\mathcal{D}$.

Interventional Causal DAG



Causal DAG $\mathcal{D}$ **after a "do"-intervention on** A**.**

DAGs and distributions

Interventional density

- $do(a)$: an intervention that sets variables A to a.
- Observational density $f(v)$, Interventional density $f(v|do(a))$.

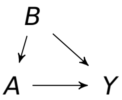
DAGs and distributions

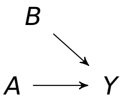
Interventional density

- $do(a)$: an intervention that sets variables A to a.
- Observational density $f(v)$, Interventional density $f(v|do(a))$.

- A DAG $\mathcal{D}$ is causal if for all observational and interventional densities:

$$f(v) = \prod_{V_j \in V} f(v_j|pa(v_j, \mathcal{D})) \quad \text{and} \quad f(v|do(a)) = \prod_{V_j \in V \setminus A} f(v_j|pa(v_j, \mathcal{D})).$$





$$f(b, a, y) = f(y|b, a)f(a|b)f(b)$$

$$f(b, y|do(a)) = f(y|b, a)f(b)$$

$$f(b, y|a) = f(y|b, a)f(b|a) \neq f(b, y|do(a))$$

How to define a causal effect?

Total causal effect

- Total causal effect - τ_{ay} - is *some functional* of $f(y|do(a))$, $P(Y|do(a))$.
- Examples: $E[Y|do(A = a + 1)] - E[Y|do(A = a)]$, $\frac{\partial}{\partial a}E(Y|do(a))$, OR, RR...

Identifiability

- A causal effect is identifiable from observational data if

$$f(y|do(a)) \text{ is computable from } f(v).$$

How to define a causal effect?

Total causal effect

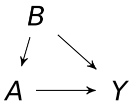
- Total causal effect - τ_{ay} - is *some functional* of $f(y|do(a))$, $P(Y|do(a))$.
- Examples: $E[Y|do(A=a+1)] - E[Y|do(A=a)]$, $\frac{\partial}{\partial a}E(Y|do(a))$, OR, RR...

Identifiability

- A causal effect is identifiable from observational data if

$$f(y|do(a)) \text{ is computable from } f(v).$$

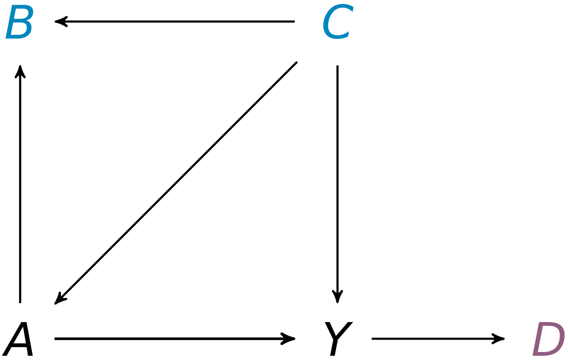
- Given the causal DAG, every total causal effect is identifiable.



$$f(y|do(a)) = \int f(b, y|do(a))db$$
$$= \int f(y|b, a)f(b)db.$$

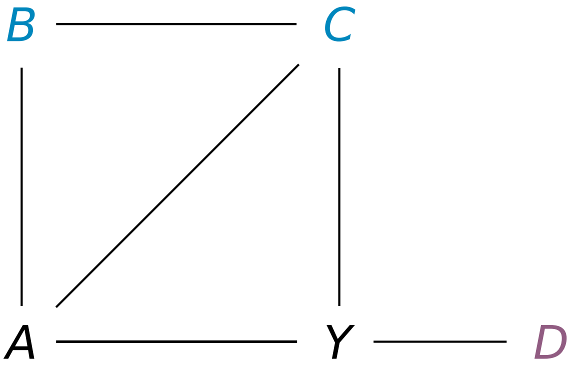
G-formula

Problem solved?



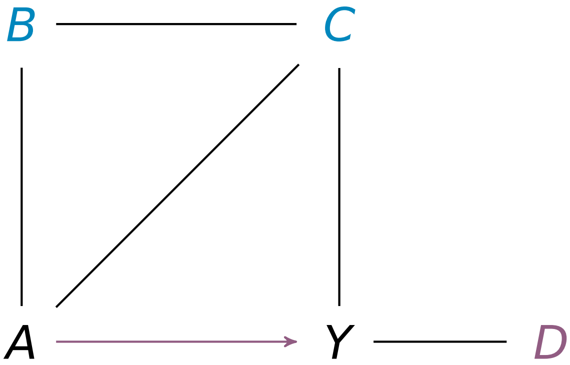
DAG $\mathcal{D}$.

Problem solved?



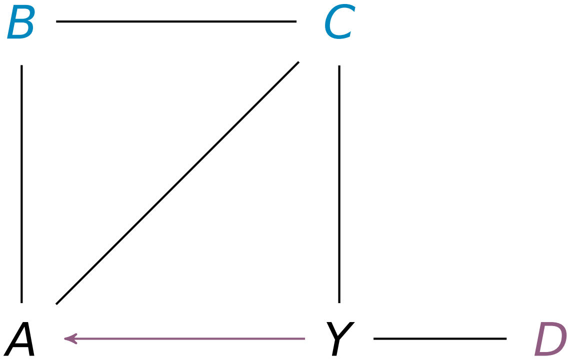
Completed Partially Directed Acyclic Graph (CPDAG) $\mathcal{C}$.

Problem solved?



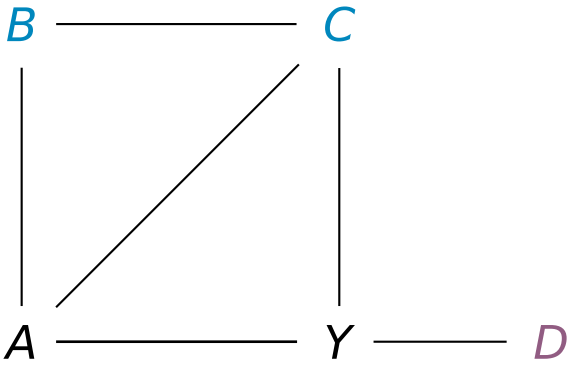
Completed Partially Directed Acyclic Graph (CPDAG) $\mathcal{C}$.

Problem solved?



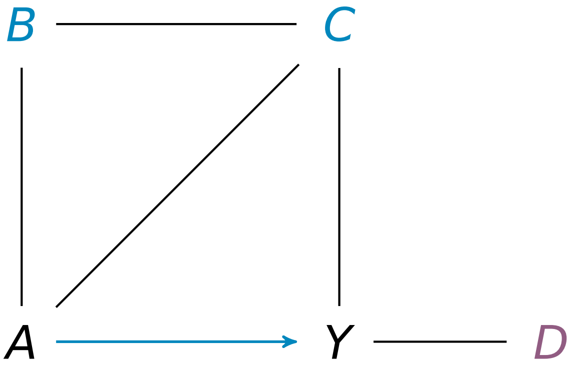
Completed Partially Directed Acyclic Graph (CPDAG) $\mathcal{C}$.

Problem solved?



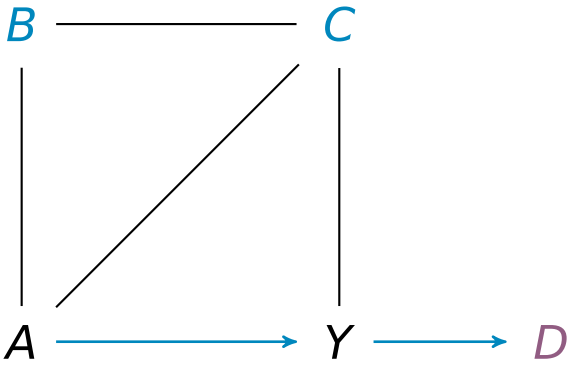
Completed Partially Directed Acyclic Graph (CPDAG) $\mathcal{C}$.

Problem solved?



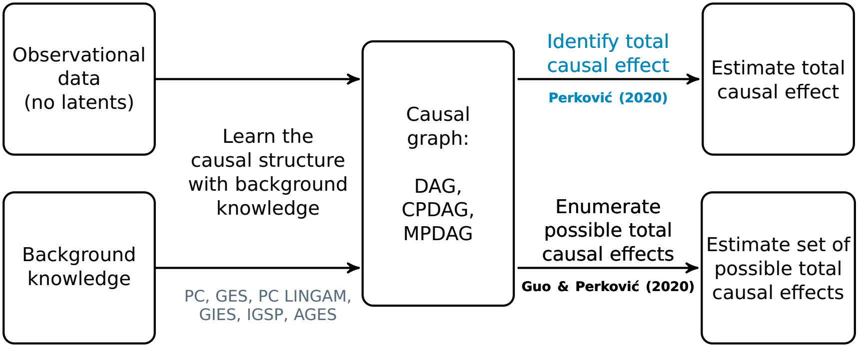
Partially Directed Acyclic Graph (PDAG).

Problem solved?



Maximally oriented Partially Directed Acyclic Graph (MPDAG) $\mathcal{G}$.

Framework

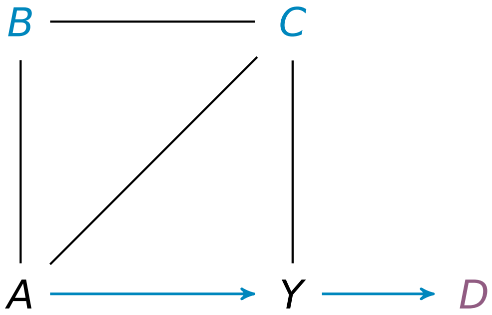


- PC (Spirtes et al, 1993), GES (Chickering, 2002)
- Adding background knowledge (Meek, 1995; TETRAD, Scheines et al., 1998), PC LINGAM (Hoyer et al., 2008), GIES (Hauser and Bühlmann, 2012), IGSP (Wang et al., 2017), etc.

Necessary and sufficient condition

Theorem (Perković, 2020)

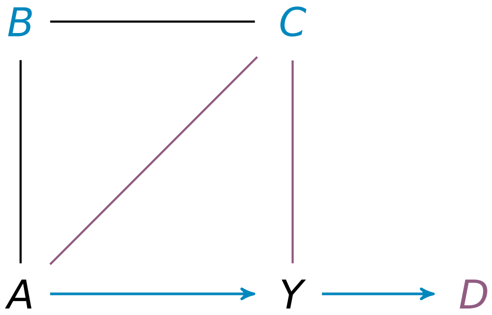
The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.



Necessary and sufficient condition

Theorem (Perković, 2020)

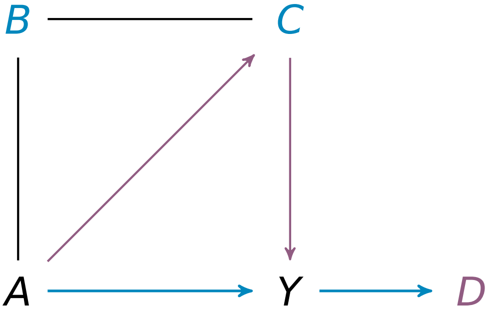
The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.



Necessary and sufficient condition

Theorem (Perković, 2020)

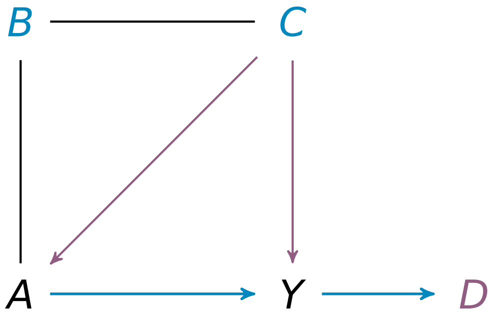
The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.



Necessary and sufficient condition

Theorem (Perković, 2020)

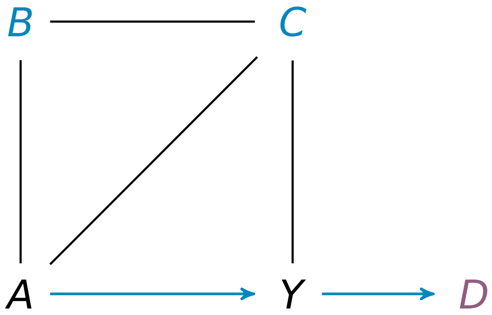
The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.



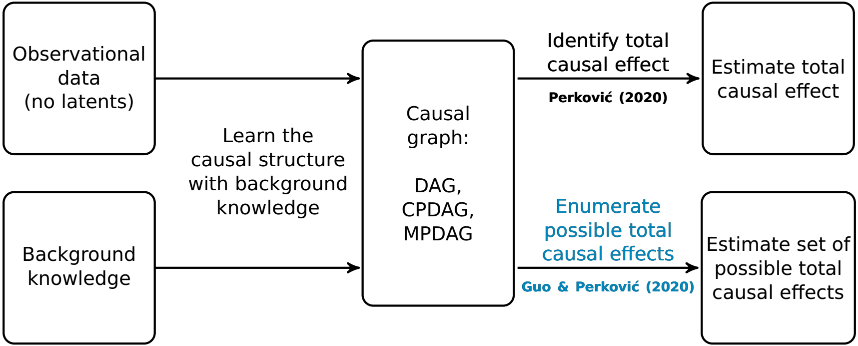
Necessary and sufficient condition

Theorem (Perković, 2020)

The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.



Framework



Observational data (no latents)
Background knowledge
Learn the causal structure with background knowledge
Causal graph:
DAG, CPDAG, MPDAG
Identify total causal effect
Perković (2020)
Enumerate possible total causal effects
Guo & Perković (2020)
Estimate total causal effect
Estimate set of possible total causal effects

Enumeration

Theorem (Perković, 2020)

The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.

Enumerate a set of possible total effects that contain the truth!

Enumeration

Theorem (Perković, 2020)

The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.

Enumerate a set of possible total effects that contain the truth!

- Naive: enumerate all the DAGs in $[\mathcal{G}]$, and identify for each DAG.
  - Computationally prohibitive for large $|V|$ (the complete graph contains $|V|!$ DAGs); see also GIllispie et al (2002), Steinsky et al (2013).

Enumeration

Theorem (Perković, 2020)

The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.

Enumerate a set of possible total effects that contain the truth!

- Naive: enumerate all the DAGs in $[\mathcal{G}]$, and identify for each DAG.
    - Computationally prohibitive for large $|V|$ (the complete graph contains $|V|!$ DAGs); see also GIllispie et al (2002), Steinsky et al (2013).
- Enumerate the valid parent sets of A:
    - $|A| = 1$: IDA algorithm (Maathuis et al 2009).
    - $|A| > 1$: joint-IDA algorithm (Nandy et al 2017).

Enumeration

Theorem (Perković, 2020)

The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.

Enumerate a set of possible total effects that contain the truth!

- Naive: enumerate all the DAGs in $[\mathcal{G}]$, and identify for each DAG.
  - Computationally prohibitive for large $|V|$ (the complete graph contains $|V|!$ DAGs); see also GIllispie et al (2002), Steinsky et al (2013).
- Enumerate the valid parent sets of A:
  - $|A| = 1$: IDA algorithm (Maathuis et al 2009).
  - $|A| > 1$: joint-IDA algorithm (Nandy et al 2017).
  - Yet, the total effect and $f(y|do(a))$ can be the same for two different parent sets!

Optimal enumeration

Theorem (Perković, 2020)

The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.

For an MPDAG $\mathcal{G}$, we look for sub-MPDAGs $\mathcal{G}_1, \ldots, \mathcal{G}_m$ such that

Optimal enumeration

Theorem (Perković, 2020)

The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.

For an MPDAG $\mathcal{G}$, we look for sub-MPDAGs $\mathcal{G}_1, \ldots, \mathcal{G}_m$ such that

1. **complete**: $[\mathcal{G}] = [\mathcal{G}_1] \dot\cup [\mathcal{G}_2] \dot\cup \cdots \dot\cup [\mathcal{G}_m]$
2. $f(y|\mathrm{do}(a))$ is identifiable under each $\mathcal{G}_i$
3. **minimal**: maps $f \mapsto f(y|do(a))$ are distinct under each $\mathcal{G}_i$ (identification formulae are distinct)
   $\Rightarrow$ possible causal effects $f \mapsto \frac{\partial}{\partial a} \mathbb{E}(Y|\mathrm{do}(A) = a)$ are distinct functionals!

Optimal enumeration

Theorem (Perković, 2020)

The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.

For an MPDAG $\mathcal{G}$, we look for sub-MPDAGs $\mathcal{G}_1, \ldots, \mathcal{G}_m$ such that

1. **complete**: $[\mathcal{G}] = [\mathcal{G}_1] \dot\cup [\mathcal{G}_2] \dot\cup \cdots \dot\cup [\mathcal{G}_m]$
2. $f(y|\mathrm{do}(a))$ is identifiable under each $\mathcal{G}_i$
3. **minimal**: maps $f \mapsto f(y|do(a))$ are distinct under each $\mathcal{G}_i$ (identification formulae are distinct)
   $\Rightarrow$ possible causal effects $f \mapsto \frac{\partial}{\partial a} \mathbb{E}(Y|\mathrm{do}(A) = a)$ are distinct functionals!

A natural algorithm is to recursively orient the undirected edges attached to A on **proper possibly causal paths** to Y.

Optimal enumeration

Theorem (Perković, 2020)

The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.

Input: MPDAG $\mathcal{G}$, $Y \in V$ and $A \subset V \setminus \{Y\}$.

Algorithm FirstTry

Optimal enumeration

Theorem (Perković, 2020)

The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.

Input: MPDAG $\mathcal{G}$, $Y \in V$ and $A \subset V \setminus \{Y\}$.

Algorithm FirstTry

1. Pick $A_1 - V_1$ such that $A_1 \in A$ and there is a proper possibly causal path $A, V_1, \ldots, Y$.

Optimal enumeration

Theorem (Perković, 2020)

The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.

Input: MPDAG $\mathcal{G}$, $Y \in V$ and $A \subset V \setminus \{Y\}$.

Algorithm FirstTry

1. Pick $A_1 - V_1$ such that $A_1 \in A$ and there is a proper possibly causal path $A, V_1, \ldots, Y$.
2. $\mathcal{G}_1 \leftarrow \text{MPDAG}(\mathcal{G}, A_1 \to V_1)$, $\mathcal{G}_2 \leftarrow \text{MPDAG}(\mathcal{G}, A_1 \leftarrow V_1)$

Optimal enumeration

Theorem (Perković, 2020)

The total causal effect of A on Y is identifiable in MPDAG $\mathcal{G}$ if and only if **all proper possibly causal paths** from A to Y start with a directed edge in $\mathcal{G}$.

Input: MPDAG $\mathcal{G}$, $Y \in V$ and $A \subset V \setminus \{Y\}$.

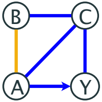
Algorithm FirstTry

1. Pick $A_1 - V_1$ such that $A_1 \in A$ and there is a proper possibly causal path $A, V_1, \dots, Y$.
2. $\mathcal{G}_1 \leftarrow$ MPDAG($\mathcal{G}, A_1 \rightarrow V_1$), $\mathcal{G}_2 \leftarrow$ MPDAG($\mathcal{G}, A_1 \leftarrow V_1$)
3. Recurse on $\mathcal{G}_1$ and $\mathcal{G}_2$ until $f(y|\text{do}(a))$ is identified

MPDAG($\mathcal{G}, R$) adds orientations R to $\mathcal{G}$ and completes Meek orientation rules.

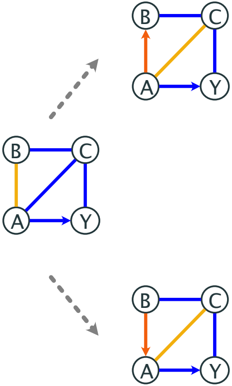
Optimal enumeration

Orienting $A - B$ first ...



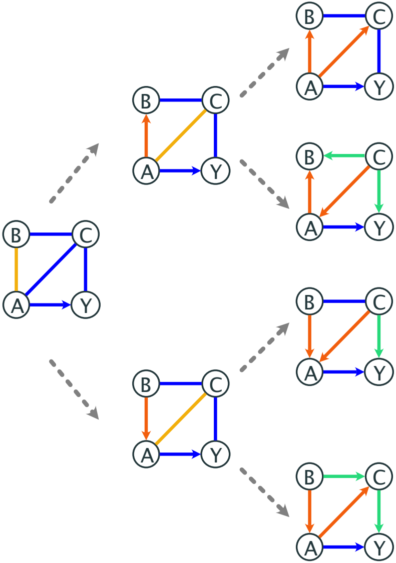
Optimal enumeration

Orienting $A - B$ first ...



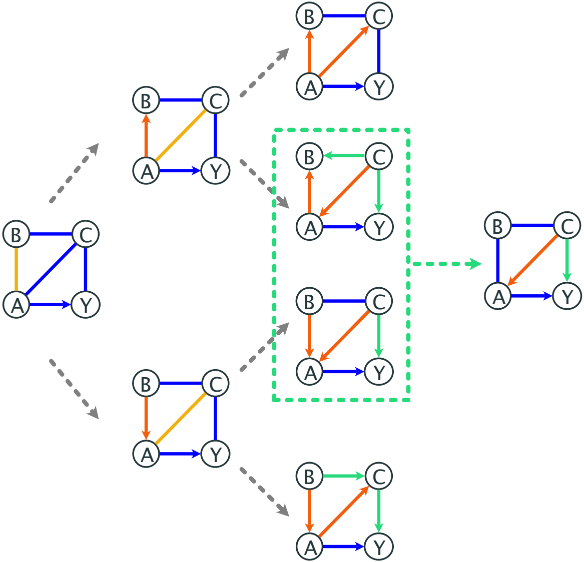
Optimal enumeration

Orienting $A - B$ first ...



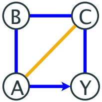
Optimal enumeration

Orienting $A - B$ first ...



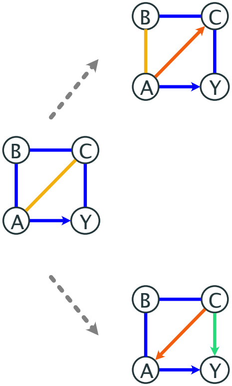
Optimal enumeration

Orienting $A - C$ first ...



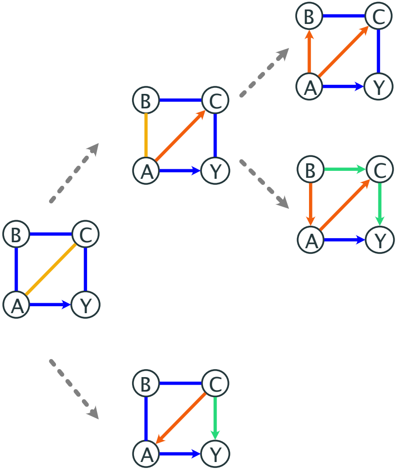
Optimal enumeration

Orienting $A - C$ first ...



Optimal enumeration

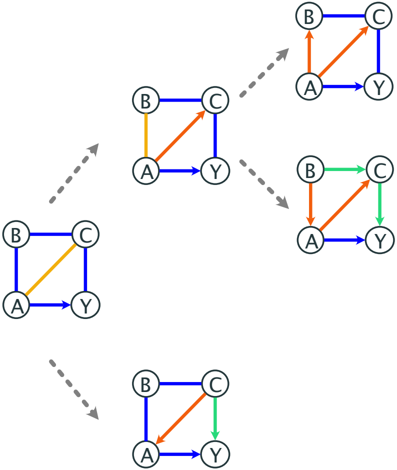
Orienting $A - C$ first ...



Optimal enumeration

Orienting $A - C$ first ...

- $A - C$ should be oriented first because the *status* of $A - B - C - Y$ depends on $A - C - Y$.



Optimal enumeration

Algorithm IDGraphs, (Guo & Perković, 2020)

1. Pick $A_1 - V_1$ such that $A_1 \in A$ and $A_1, V_1, \ldots, Y$ is a shortest proper possibly causal path from A to Y.
2. $\mathcal{G}_1 \leftarrow \text{MPDAG}(\mathcal{G}, A_1 \rightarrow V_1)$, $\mathcal{G}_2 \leftarrow \text{MPDAG}(\mathcal{G}, A_1 \leftarrow V_1)$
3. Recurse on $\mathcal{G}_1$ and $\mathcal{G}_2$ until identified

Optimal enumeration

Algorithm IDGraphs, (Guo & Perković, 2020)

1. Pick $A_1 - V_1$ such that $A_1 \in A$ and $A_1, V_1, \ldots, Y$ is a shortest proper possibly causal path from A to Y.
2. $\mathcal{G}_1 \leftarrow \text{MPDAG}(\mathcal{G}, A_1 \to V_1)$, $\mathcal{G}_2 \leftarrow \text{MPDAG}(\mathcal{G}, A_1 \leftarrow V_1)$
3. Recurse on $\mathcal{G}_1$ and $\mathcal{G}_2$ until identified

Theorem (Guo & Perković, 2020)

$(\mathcal{G}_1, \ldots, \mathcal{G}_m)$ output by the algorithm is **complete** and **minimal**.

Optimal enumeration

Algorithm IDGraphs, (Guo & Perković, 2020)

1. Pick $A_1 - V_1$ such that $A_1 \in A$ and $A_1, V_1, \ldots, Y$ is a shortest proper possibly causal path from A to Y.
2. $\mathcal{G}_1 \leftarrow$ MPDAG($\mathcal{G}, A_1 \rightarrow V_1$), $\mathcal{G}_2 \leftarrow$ MPDAG($\mathcal{G}, A_1 \leftarrow V_1$)
3. Recurse on $\mathcal{G}_1$ and $\mathcal{G}_2$ until identified

Theorem (Guo & Perković, 2020)

$(\mathcal{G}_1, \ldots, \mathcal{G}_m)$ output by the algorithm is **complete** and **minimal**.

Hence, each $\mathcal{G}_i$ represents the minimal set of additional orientations required for a particular interventional distribution/possible effect!

Optimal enumeration

Algorithm IDGraphs, (Guo & Perković, 2020)

1. Pick $A_1 - V_1$ such that $A_1 \in A$ and $A_1, V_1, \ldots, Y$ is a shortest proper possibly causal path from A to Y.
2. $\mathcal{G}_1 \leftarrow \text{MPDAG}(\mathcal{G}, A_1 \rightarrow V_1)$, $\mathcal{G}_2 \leftarrow \text{MPDAG}(\mathcal{G}, A_1 \leftarrow V_1)$
3. Recurse on $\mathcal{G}_1$ and $\mathcal{G}_2$ until identified

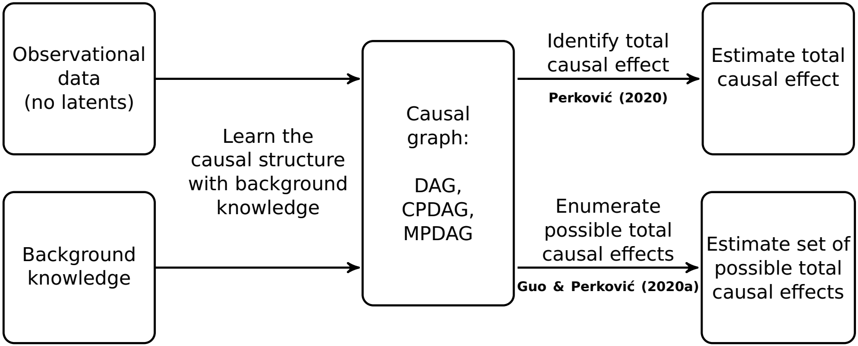
Theorem (Guo & Perković, 2020)

$(\mathcal{G}_1, \ldots, \mathcal{G}_m)$ output by the algorithm is **complete** and **minimal**.

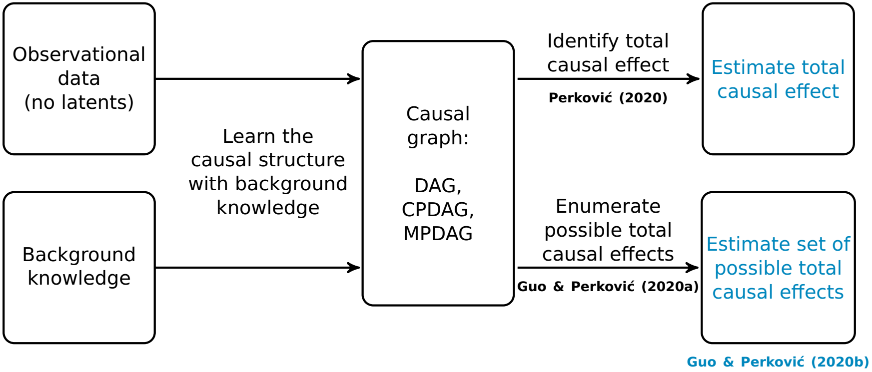
Hence, each $\mathcal{G}_i$ represents the minimal set of additional orientations required for a particular interventional distribution/possible effect!

In contrast, the IDA algorithm will output 4 effects for this example, but two of them are different estimates of the same possible effect!

Framework

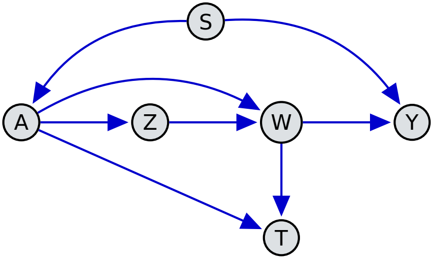

Observational data (no latents)
Background knowledge
Learn the causal structure with background knowledge
Causal graph:
DAG, CPDAG, MPDAG
Identify total causal effect
Perković (2020)
Estimate total causal effect
Enumerate possible total causal effects
Guo & Perković (2020a)
Estimate set of possible total causal effects

Framework

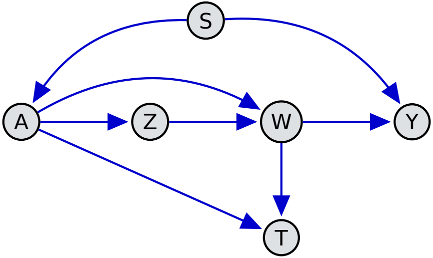


- In the following, we further assume **linearity** in the data generating mechanism.

Causal DAG, linear SEM

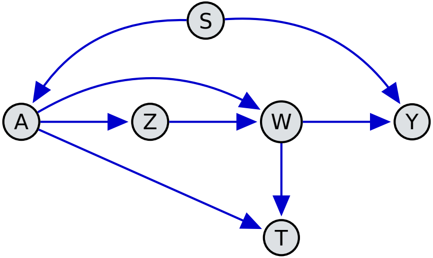


Causal DAG, linear SEM



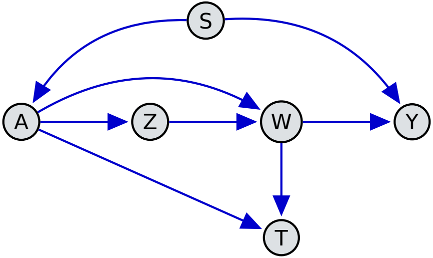
- Suppose $\mathcal{D}$ is the underlying causal DAG.

Causal DAG, linear SEM



- Suppose $\mathcal{D}$ is the underlying causal DAG.
  - $\mathcal{D}$ is **unknown**.

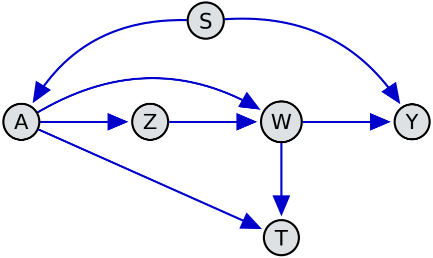
Causal DAG, linear SEM



- Suppose $\mathcal{D}$ is the underlying causal DAG.
  - $\mathcal{D}$ is **unknown**.
- Suppose data is generated by a **linear structural equation model** (SEM)

$$X_v = \sum_{u: u \to v} \gamma_{uv} X_u + \epsilon_u, \quad \mathbb{E}\,\epsilon_u = 0, \quad 0 < \mathrm{var}\,\epsilon_u < \infty.$$

Causal DAG, linear SEM



- Suppose $\mathcal{D}$ is the underlying causal DAG.
  - $\mathcal{D}$ is **unknown**.
- Suppose data is generated by a **linear structural equation model** (SEM)

$$X_v = \sum_{u:u \to v} \gamma_{uv} X_u + \epsilon_u, \quad \mathbb{E}\, \epsilon_u = 0, \quad 0 < \operatorname{var} \epsilon_u < \infty.$$

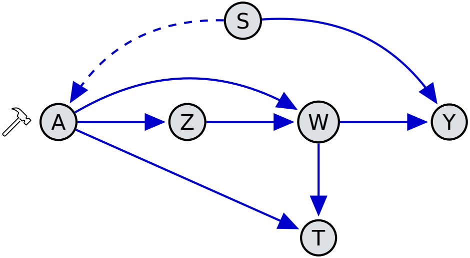
- Under no unobserved confounder, the errors are **mutually independent**.

Total effect

Suppose we want to estimate the **total (causal) effect of A on Y**.

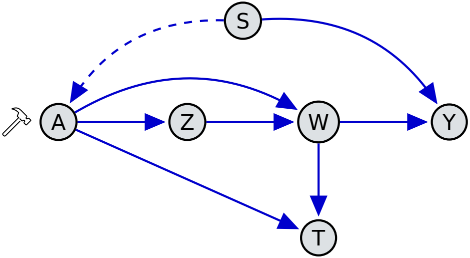
Total effect

Suppose we want to estimate the **total (causal) effect of A on Y**.



Total effect

Suppose we want to estimate the **total (causal) effect of A on Y**.



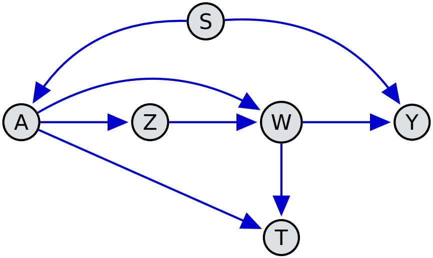
The total effect τ_{AY} is defined as the slope of $x_a \mapsto \mathbb{E}[X_Y | \text{do}(X_A = x_a)]$, given by a sum-product of Wright (1934):

$$\tau_{AY} = \frac{\mathrm{d}}{\mathrm{d}x_a} \mathbb{E}[X_Y | \text{do}(X_A = x_a)] = (\gamma_{AZ}\gamma_{ZW} + \gamma_{AW})\gamma_{WY}.$$

Estimation

Our task is to estimate τ_{AY} from n iid observational sample generated by a linear SEM associated with causal DAG $\mathcal{D}$, given that

$$\mathcal{D} \in [\mathcal{G}] \text{ for MPDAG } \mathcal{G}, \quad \tau_{AY} \text{ is identifiable from } \mathcal{G}.$$

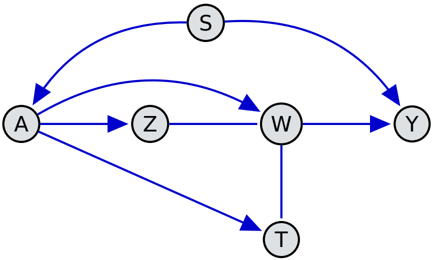


DAG $\mathcal{D}$

Estimation

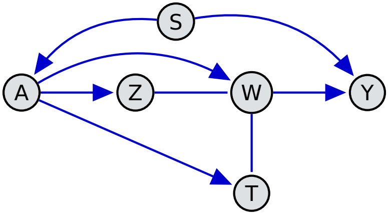
Our task is to estimate τ_{AY} from n iid observational sample generated by a linear SEM associated with causal DAG $\mathcal{D}$, given that

$$\mathcal{D} \in [\mathcal{G}] \text{ for MPDAG } \mathcal{G}, \quad \tau_{AY} \text{ is identifiable from } \mathcal{G}.$$

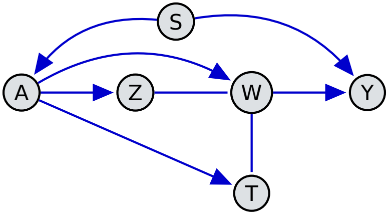


MPDAG $\mathcal{G}$

Buckets

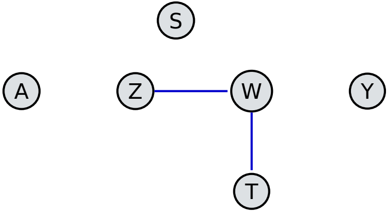


Buckets



- Let “buckets” be the undirected connected components of MPDAG $\mathcal{G}$:

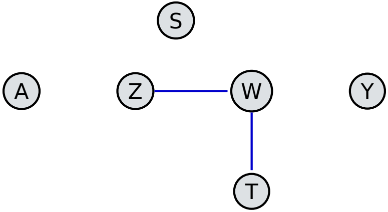
Buckets



- Let "buckets" be the undirected connected components of MPDAG $\mathcal{G}$:

$$B_1 = \{S\},\ B_2 = \{A\},\ B_3 = \{Z, W, T\},\ B_4 = \{Y\}.$$

Buckets

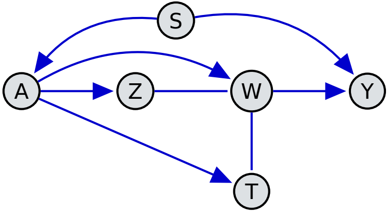


- Let "buckets" be the undirected connected components of MPDAG $\mathcal{G}$:

$$B_1 = \{S\},\ B_2 = \{A\},\ B_3 = \{Z, W, T\},\ B_4 = \{Y\}.$$

1. The "between bucket" causal effects are **identifiable**.

Buckets

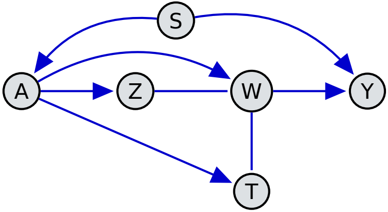


- Let "buckets" be the undirected connected components of MPDAG $\mathcal{G}$:

$$B_1 = \{S\},\ B_2 = \{A\},\ B_3 = \{Z, W, T\},\ B_4 = \{Y\}.$$

  1. The "between bucket" causal effects are **identifiable**.

Buckets

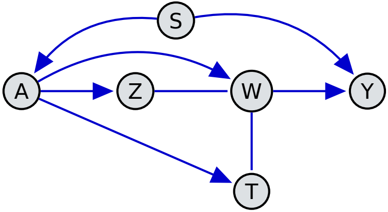


- Let "buckets" be the undirected connected components of MPDAG $\mathcal{G}$:

$$B_1 = \{S\},\ B_2 = \{A\},\ B_3 = \{Z, W, T\},\ B_4 = \{Y\}.$$

1. The "between bucket" causal effects are **identifiable**.
2. **Restrictive property:** Each node in a bucket has the same out-of-bucket parents (Guo and Perković, 2020b).

Buckets



- Let "buckets" be the undirected connected components of MPDAG $\mathcal{G}$:

$$B_1 = \{S\},\ B_2 = \{A\},\ B_3 = \{Z, W, T\},\ B_4 = \{Y\}.$$

  1. The "between bucket" causal effects are **identifiable**.
  2. **Restrictive property:** Each node in a bucket has the same out-of-bucket parents (Guo and Perković, 2020b).

- We use this to reparametrize the SEM.

Block-recursive reparametrization

Proposition (Block-recursive form, Guo and Perković, 2020)

Let $B_1, \ldots, B_K$ be the ordered bucket decomposition of V in MPDAG $\mathcal{G}$. Then

$$X = \Lambda^\intercal X + \varepsilon, \qquad \Lambda = (\lambda_{ij}), j \in B_k,\ i \notin \mathrm{pa}(B_k, \mathcal{G}) \quad \Rightarrow \quad \lambda_{ij} = 0,$$

$$\mathbb{E}\,\varepsilon = 0, \quad \mathbb{E}\,\varepsilon_{B_k}\varepsilon_{B_k}^\intercal \succ \mathbf{0}, \quad \varepsilon_{B_k} \text{ mutually independent,}$$

Block-recursive reparametrization

Proposition (Block-recursive form, Guo and Perković, 2020)

Let $B_1, \ldots, B_K$ be the ordered bucket decomposition of V in MPDAG $\mathcal{G}$. Then

$$X = \Lambda^\intercal X + \varepsilon, \qquad \Lambda = (\lambda_{ij}), j \in B_k, \ i \notin \mathrm{pa}(B_k, \mathcal{G}) \quad \Rightarrow \quad \lambda_{ij} = 0,$$

$$\mathbb{E}\,\varepsilon = 0, \quad \mathbb{E}\,\varepsilon_{B_k}\varepsilon_{B_k}^\intercal \succ \mathbf{0}, \quad \varepsilon_{B_k} \text{ mutually independent},$$

Two nice things happen under this reparametrization:

Block-recursive reparametrization

Proposition (Block-recursive form, Guo and Perković, 2020)

Let $B_1, \dots, B_K$ be the ordered bucket decomposition of V in MPDAG $\mathcal{G}$. Then

$$X = \Lambda^\intercal X + \varepsilon, \qquad \Lambda = (\lambda_{ij}), j \in B_k,\ i \notin \mathrm{pa}(B_k, \mathcal{G}) \quad \Rightarrow \quad \lambda_{ij} = 0,$$
$$\mathbb{E}\,\varepsilon = 0, \quad \mathbb{E}\,\varepsilon_{B_k}\varepsilon_{B_k}^\intercal \succ \mathbf{0}, \quad \varepsilon_{B_k} \text{ mutually independent,}$$

Two nice things happen under this reparametrization:

- For $S = \mathrm{An}(Y, \mathcal{G}_{V \setminus A})$, τ_{AY} can be identified as

$$\tau_{AY} = \Lambda_{A,S} \left[(I - \Lambda_{S,S})^{-1}\right]_{S,Y}.$$

  The bucket-wise **error distribution** is **nuisance**.

Block-recursive reparametrization

Proposition (Block-recursive form, Guo and Perković, 2020)

Let $B_1, \ldots, B_K$ be the ordered bucket decomposition of V in MPDAG $\mathcal{G}$. Then

$$X = \Lambda^{\mathsf{T}} X + \varepsilon, \qquad \Lambda = (\lambda_{ij}), j \in B_k,\ i \notin \mathrm{pa}(B_k, \mathcal{G}) \quad \Rightarrow \quad \lambda_{ij} = 0,$$
$$\mathbb{E}\,\varepsilon = 0, \quad \mathbb{E}\,\varepsilon_{B_k} \varepsilon_{B_k}^{\mathsf{T}} \succ \mathbf{0}, \quad \varepsilon_{B_k} \text{ mutually independent},$$

Two nice things happen under this reparametrization:

- For $S = \mathrm{An}(Y, \mathcal{G}_{V \setminus A})$, τ_{AY} can be identified as

$$\tau_{AY} = \Lambda_{A,S} \left[(I - \Lambda_{S,S})^{-1} \right]_{S,Y}.$$

  The bucket-wise **error distribution** is **nuisance**.

- Under **Gaussian** errors, the MLE for each $\Lambda_{\mathrm{pa}(B_i, \mathcal{G}), B_i}$ is just the **least squares** coefficients of $B_i \sim \mathrm{pa}(B_i, \mathcal{G})$.

Block-recursive reparametrization

Proposition (Block-recursive form, Guo and Perković, 2020)

Let $B_1, \ldots, B_K$ be the ordered bucket decomposition of V in MPDAG $\mathcal{G}$. Then

$$X = \Lambda^{\intercal} X + \varepsilon, \qquad \Lambda = (\lambda_{ij}), j \in B_k,\ i \notin \mathrm{pa}(B_k, \mathcal{G}) \quad \Rightarrow \quad \lambda_{ij} = 0,$$

$$\mathbb{E}\,\varepsilon = 0, \quad \mathbb{E}\,\varepsilon_{B_k} \varepsilon_{B_k}^{\intercal} \succ \mathbf{0}, \quad \varepsilon_{B_k} \text{ mutually independent},$$

Two nice things happen under this reparametrization:

- For $S = \mathrm{An}(Y, \mathcal{G}_{V \setminus A})$, τ_{AY} can be identified as

$$\tau_{AY} = \Lambda_{A,S} \left[(I - \Lambda_{S,S})^{-1} \right]_{S,Y}.$$

  The bucket-wise **error distribution** is **nuisance**.

- Under **Gaussian** errors, the MLE for each $\Lambda_{\mathrm{pa}(B_i, \mathcal{G}), B_i}$ is just the **least squares** coefficients of $B_i \sim \mathrm{pa}(B_i, \mathcal{G})$.
  - This is a special case of "seemingly unrelated regression" under the **restrictive property**.

Efficient estimator

If τ_{AY} is identifiable given MPDAG $\mathcal{G}$, the **$\mathcal{G}$-regression estimator** is defined as:

$$\hat{\tau}_{AY}^{\mathcal{G}} := \hat{\Lambda}_{A,S}^{\mathcal{G}} \left[(I - \hat{\Lambda}_{S,S}^{\mathcal{G}})^{-1} \right]_{S,Y},$$

where $S = \mathrm{An}(Y, \mathcal{G}_{V \setminus A})$, and $\hat{\Lambda}^{\mathcal{G}}$ is matrix consisting of least squares coefficients for each "bucket" regression.

Efficient estimator

If τ_{AY} is identifiable given MPDAG $\mathcal{G}$, the **$\mathcal{G}$-regression estimator** is defined as:

$$\hat{\tau}^{\mathcal{G}}_{AY} := \hat{\Lambda}^{\mathcal{G}}_{A,S}\left[(I - \hat{\Lambda}^{\mathcal{G}}_{S,S})^{-1}\right]_{S,Y},$$

where $S = \mathrm{An}(Y, \mathcal{G}_{V \setminus A})$, and $\hat{\Lambda}^{\mathcal{G}}$ is matrix consisting of least squares coefficients for each “bucket” regression.

Theorem ($\mathcal{G}$-regression, Guo and Perković, 2020)

Then for any regular estimator $\hat{\tau}_{AY}$ that only uses the **first two moments** of the data, it holds that

$$\mathrm{avar}\left(\hat{\tau}_{AY}\right) \geq \mathrm{avar}\left(\hat{\tau}^{\mathcal{G}}_{AY}\right).$$

Efficient estimator

If τ_{AY} is identifiable given MPDAG $\mathcal{G}$, the **$\mathcal{G}$-regression estimator** is defined as:

$$\hat{\tau}_{AY}^{\mathcal{G}} := \hat{\Lambda}_{A,S}^{\mathcal{G}} \left[(I - \hat{\Lambda}_{S,S}^{\mathcal{G}})^{-1} \right]_{S,Y},$$

where $S = \mathrm{An}(Y, \mathcal{G}_{V \setminus A})$, and $\hat{\Lambda}^{\mathcal{G}}$ is matrix consisting of least squares coefficients for each "bucket" regression.

Theorem ($\mathcal{G}$-regression, Guo and Perković, 2020)

Then for any regular estimator $\hat{\tau}_{AY}$ that only uses the **first two moments** of the data, it holds that

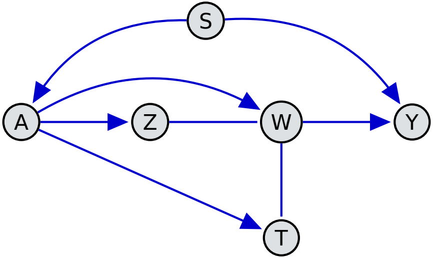
$$\mathrm{avar}\left(\hat{\tau}_{AY}\right) \geq \mathrm{avar}\left(\hat{\tau}_{AY}^{\mathcal{G}}\right).$$

This includes estimators in the literature:

- covariate adjustment (Henckel et al., 2019, Witte et al., 2020),
- recursive regressions (Nandy et al., 2017, Gupta et al., 2020),
- modified Cholesky decomposition (Nandy et al., 2017).

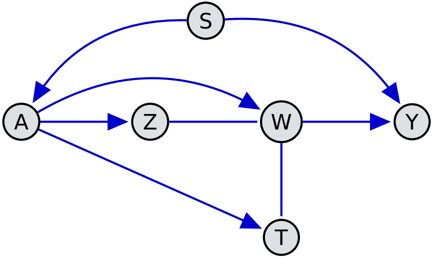
Efficient estimator

The efficiency is achieved by exploiting the conditional independences encoded in $\mathcal{G}$.



Efficient estimator

The efficiency is achieved by exploiting the conditional independences encoded in $\mathcal{G}$.



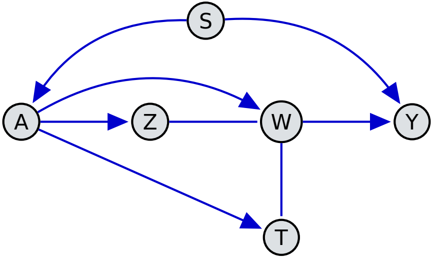
$\mathcal{G}$-regression estimator

$$\hat{\tau}_{AY}^{\mathcal{G}} = \hat{\lambda}_{AW}\hat{\lambda}_{WY},$$

where $\hat{\lambda}_{AW}$, $\hat{\lambda}_{WY}$ are taken from $W \sim A$ and $Y \sim W + S$ respectively.

Efficient estimator

The efficiency is achieved by exploiting the conditional independences encoded in $\mathcal{G}$.



$\mathcal{G}$-regression estimator

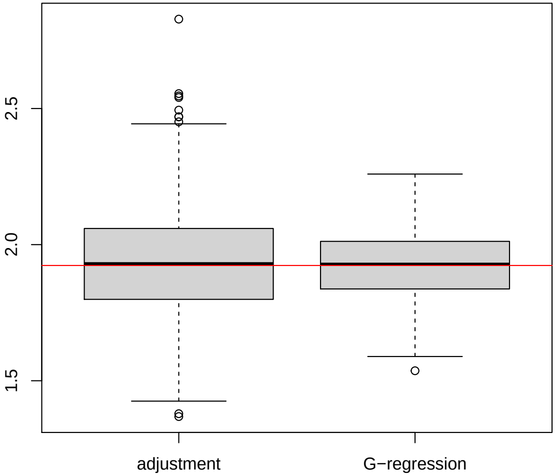
$$\hat{\tau}_{AY}^{\mathcal{G}} = \hat{\lambda}_{AW}\hat{\lambda}_{WY},$$

where $\hat{\lambda}_{AW}$, $\hat{\lambda}_{WY}$ are taken from $W \sim A$ and $Y \sim W + S$ respectively.

adjustment estimator

$$\hat{\tau}_{AY}^{\text{adj}} = \hat{\beta}_{AY} \text{ from } Y \sim A + S.$$

Efficient estimator



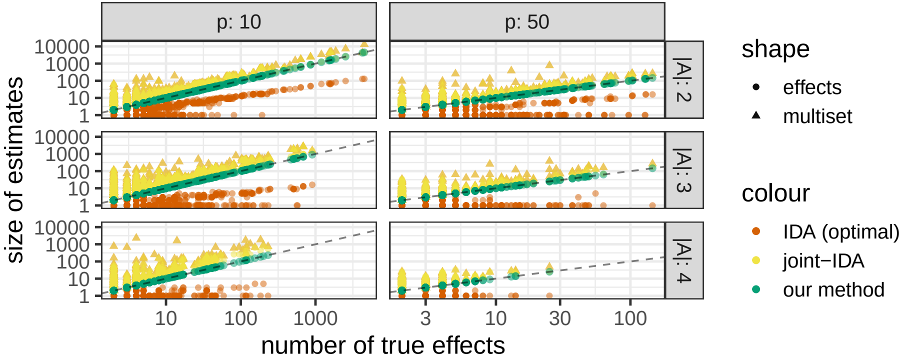
$n = 100$, t_5 errors.

Simulation: relative efficiency

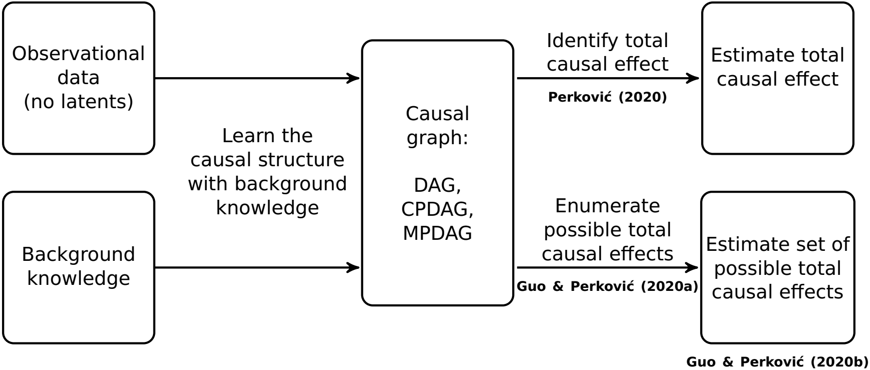
Table: Geometric average of squared errors relative to $\mathcal{G}$-regression:
`adj.O`: optimal adjustment
`IDA.M`: IDA (Cholesky)
`IDA.R`: IDA (recursive regression)

	$\|V\|=20$		$\|V\|=50$		$\|V\|=100$	
$\|A\|$	$n=100$	$n=1000$	$n=100$	$n=1000$	$n=100$	$n=1000$
`adj.O`						
1	1.3	1.3	1.4	1.3	1.5	1.5
2	3.4	4.2	4.7	4.9	4.2	4.5
3	6.3	5.9	7.4	7.2	7.8	8.0
4	9.3	9.3	12	14	12	12
`IDA.M`						
1	20	19	61	48	103	108
2	62	65	220	182	293	356
3	93	119	354	396	749	771
4	154	222	533	895	1188	1604
`IDA.R`						
1	20	19	61	48	103	108
2	33	38	121	113	176	199
3	30	39	171	135	342	312
4	48	50	187	214	405	432

Simulation: size of possible effects



Final remarks



- **R package** `eff`^{2}: github.com/richardkwo/eff2
- **Efficient estimation beyond linear SEMs?** We are working on it!

Thanks!