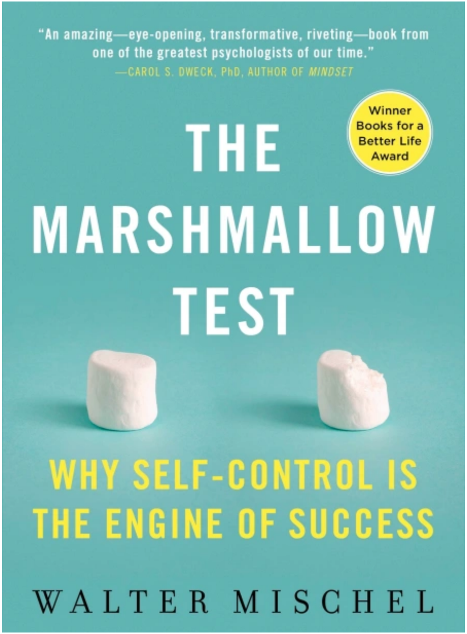
Identifying and Estimating Causal Effects with Incomplete Causal Information

Emilija Perković
University of Washington

some joint work with F. Richard Guo,
Andrea Rotnitzky, Marloes Maathuis, Leonard Henckel

Stanford Marshmallow Experiment



Marshmallow Test

?

Marshmallow Test

THE GLOBE AND MAIL SUBSCRIBE LOG

How having self-control as a kid can affect your health later

BUSINESS INSIDER Subscribe

How Self Control Leads To Success In Life, According To This Legendary Stanford Psychologist

Drake Baer Oct 21, 2014, 6:16 PM

FINANCIAL TIMES myFT

Personal Finance + Add to myFT

Can you resist instant gratification for your finances?

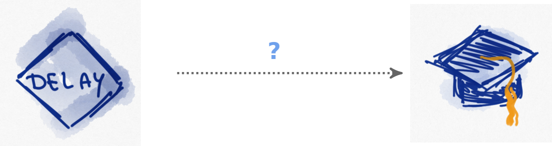
The marshmallow test – and your money

The Guardian

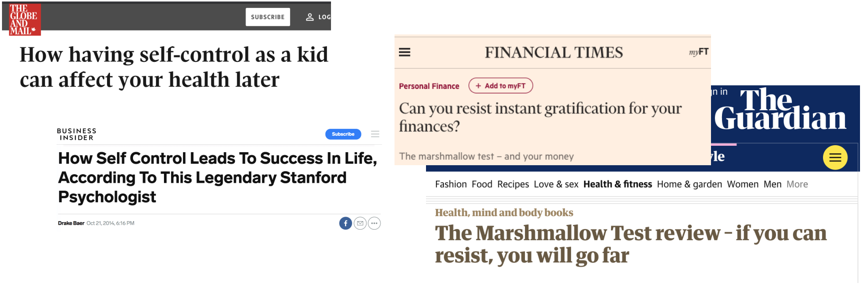
Fashion Food Recipes Love & sex **Health & fitness** Home & garden Women Men More

Health, mind and body books

The Marshmallow Test review – if you can resist, you will go far



Marshmallow Test


THE GLOBE AND MAIL

SUBSCRIBE LOG

How having self-control as a kid can affect your health later

BUSINESS INSIDER

Subscribe

How Self Control Leads To Success In Life, According To This Legendary Stanford Psychologist

Drake Baer Oct 21, 2014, 6:16 PM

FINANCIAL TIMES

myFT

Personal Finance + Add to myFT

Can you resist instant gratification for your finances?

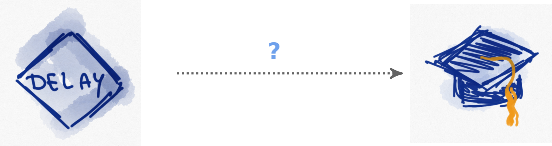
The marshmallow test – and your money

The Guardian

Fashion Food Recipes Love & sex Health & fitness Home & garden Women Men More

Health, mind and body books

The Marshmallow Test review - if you can resist, you will go far



Should we train the delay of gratification?

Marshmallow Test

Assumptions:

- Access to observational data + domain knowledge.
- Do we know all variables that explain or moderate link?
- Do we know all relationships between these variables?

Causal Relationships

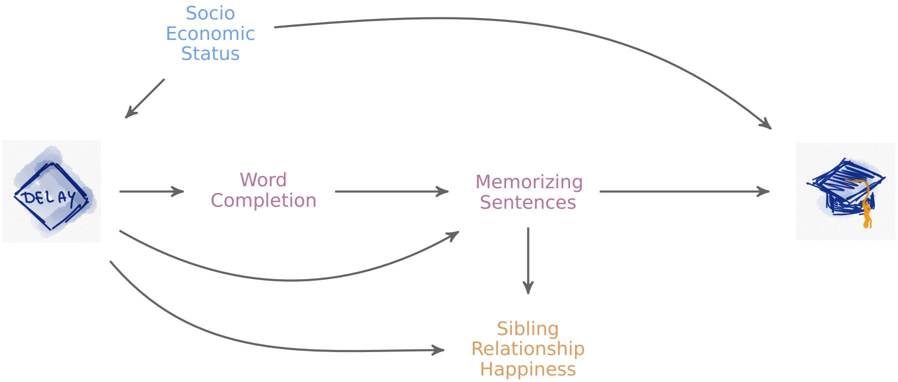
Socio
Economic
Status

Word
Completion

Memorizing
Sentences

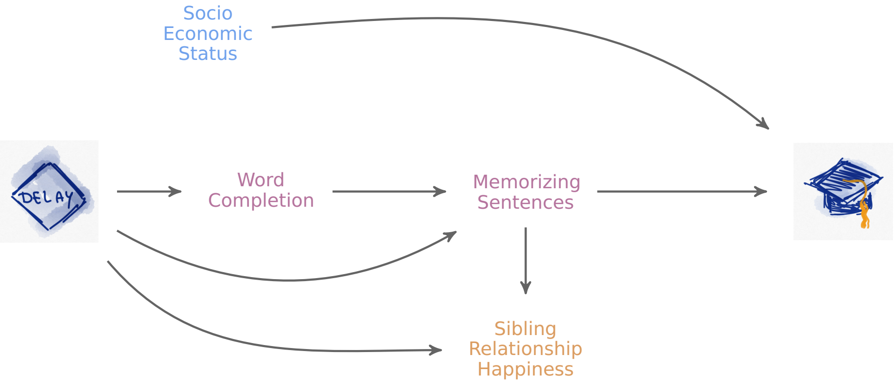
Sibling
Relationship
Happiness

Causal Relationships



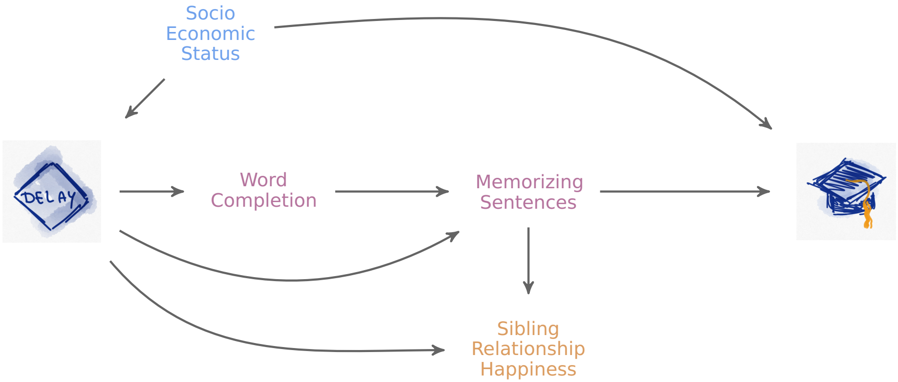
Causal Directed Acyclic Graph (DAG) $\mathcal{D}$.

Interventional DAG



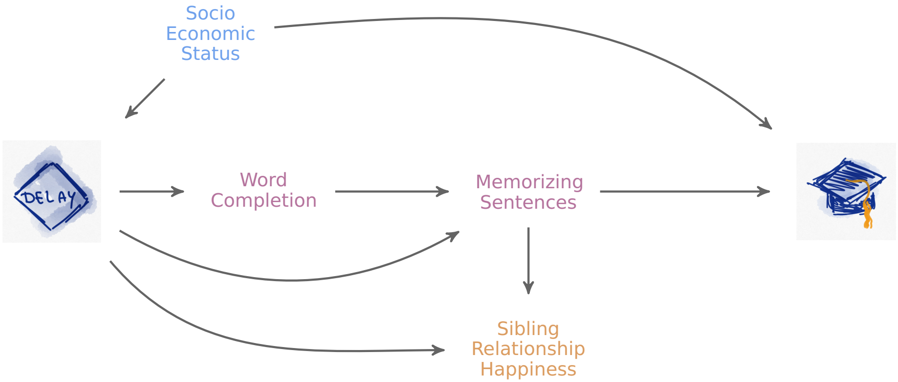
- *Randomized experiment*, e.g: each participant is randomly assigned to treatment or control.
- Any change in response due to a change in treatment goes through *causal paths*.
- $do(x_A)$: an intervention that sets variable X_A to x_A.
- $f(x_Y|do(x_A)) \rightarrow$Causal Effect

Observational Causal DAG



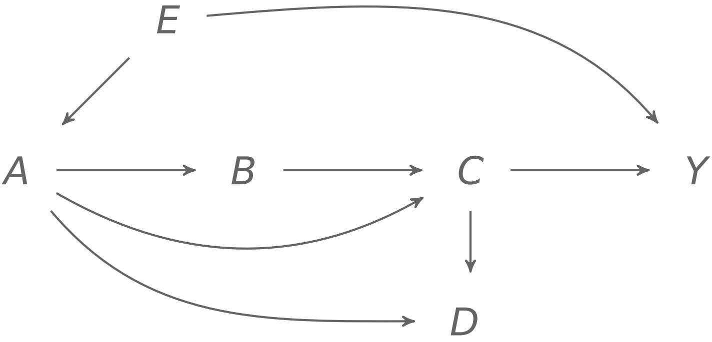
- $f(x_{\mathbf{V}}) \to$ Observational Data
- Access to: $f(x_Y|x_A), f(x_Y), \ldots$
- **Issues**: 1. In general, $f(x_Y|do(x_A)) \neq f(x_Y|x_A)$.

Observational Causal DAG



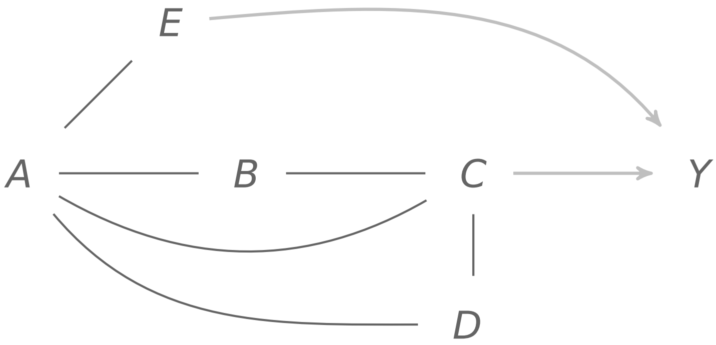
- $f(x_{\mathbf{V}}) \rightarrow$ Observational Data
- Access to: $f(x_Y|x_A), f(x_Y), \ldots$
- **Issues**: 1. In general, $f(x_Y|do(x_A)) \neq f(x_Y|x_A)$. 2. We may not know the full graph.

What if we do not know the DAG?



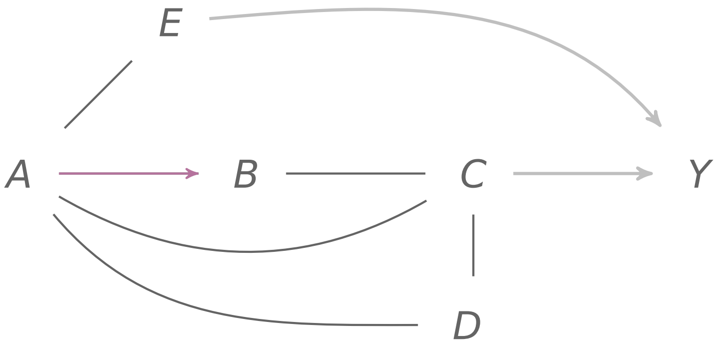
Causal Directed Acyclic Graph (DAG) $\mathcal{D}$.

What if we do not know the DAG?



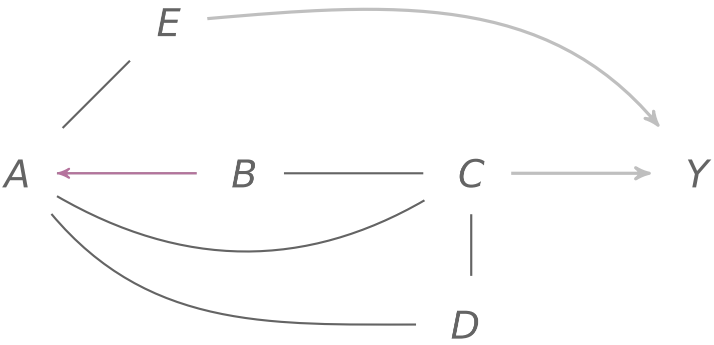
Completed Partially Directed Acyclic Graph (CPDAG).

What if we do not know the DAG?



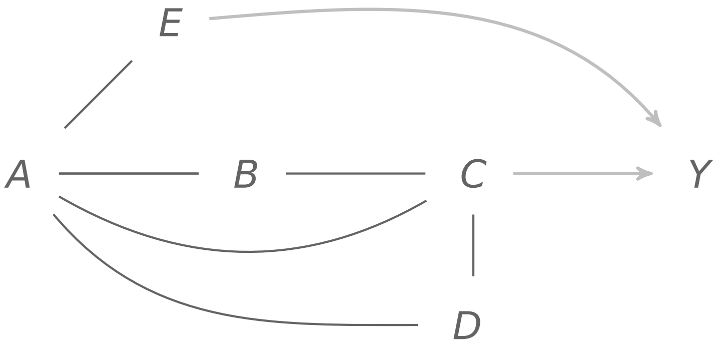
Completed Partially Directed Acyclic Graph (CPDAG).

What if we do not know the DAG?



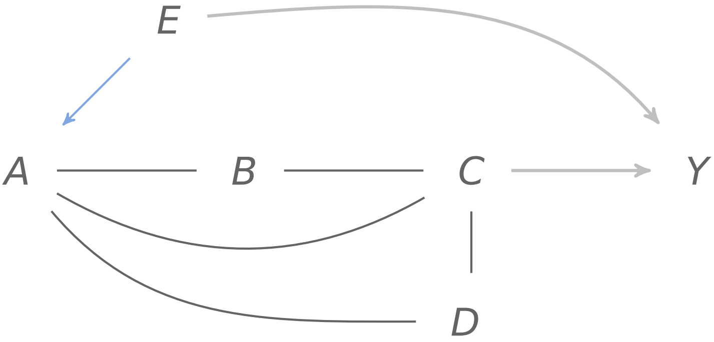
Completed Partially Directed Acyclic Graph (CPDAG).

What if we do not know the DAG?



Completed Partially Directed Acyclic Graph (CPDAG).

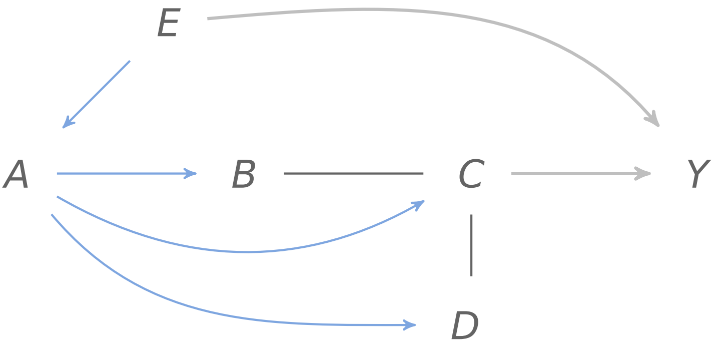
What if we do not know the DAG?



Partially Directed Acyclic Graph (PDAG).

- Expert knowledge of causal relations, previous experiments, model restrictions. . .

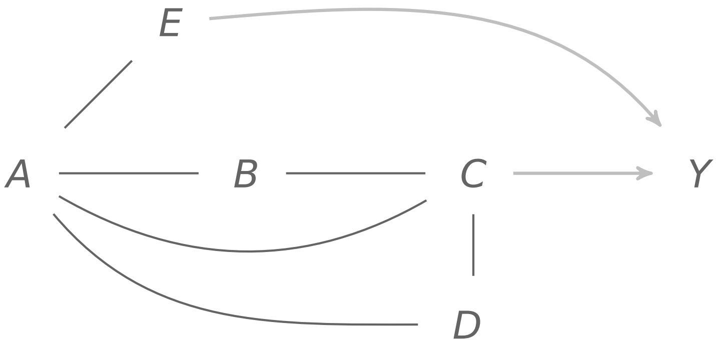
What if we do not know the DAG?



Maximally oriented Partially Directed Acyclic Graph (MPDAG).

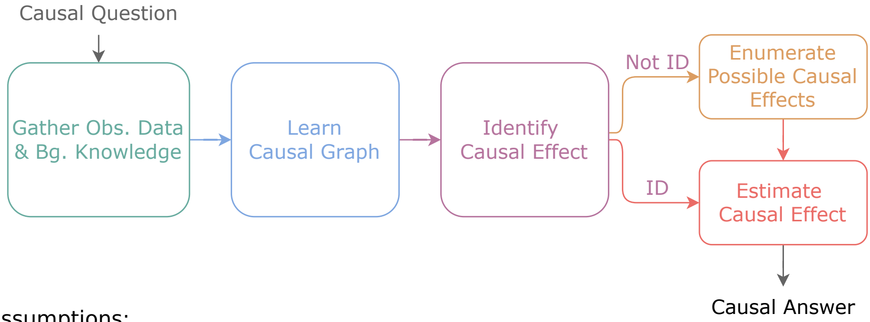
Expert knowledge of causal relations, previous experiments, model restrictions. . .

What if we do not know the DAG?



Completed Partially Directed Acyclic Graph (CPDAG).

Causal Framework



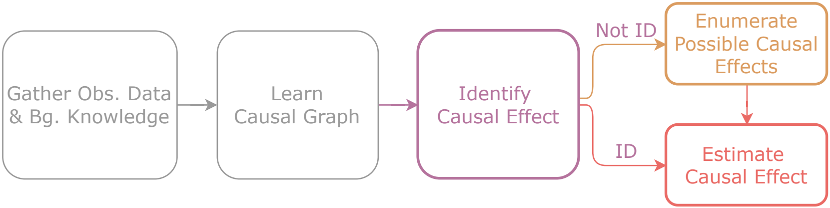
Assumptions:

- Access to observational data + domain knowledge.
- Do we know all variables that explain or moderate link? Yes.
- Do we know all relationships between these variables? No.

1) Can we uniquely identify the causal effect or a set of possible effects?

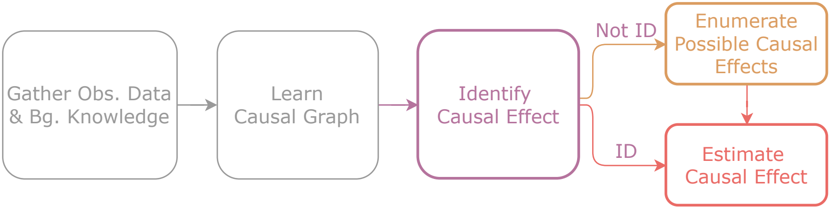
2) How strong is this causal relationship?
   - How to construct an estimator?
   - What estimator is optimal in terms of minimal variance?

My Work



- Perković, Textor, Kalisch and Maathuis (2015). A Complete Generalized Adjustment Criterion. *UAI 2015*.
- Perković, Kalisch and Maathuis (2017). Interpreting and Using CPDAGs with Background Knowledge. *UAI 2017*.
- Perković, Textor, Kalisch and Maathuis (2018). Complete Graphical Characterization and Construction of Adjustment Sets in Markov Equivalence Classes of Ancestral Graphs. *JMLR*.
- Perković (2020). Identifying total causal effects in MPDAGs. *UAI 2020*.
- Guo and Perković (2021). Minimal enumeration of all possible total effects in a Markov equivalence class. *AISTATS 2021*.
- Guo and Perković (2022). Efficient Least Squares for Estimating Total Effects under Linearity and Causal Sufficiency. *JMLR*.
- Henckel, Perković, and Maathuis (2022). Graphical Criteria for Efficient Total Effect Estimation via Adjustment in Causal Linear Structural Equation Models. *JRSS:B*.
- Guo, Perković, and Rotnitzky (2022). Variable elimination, graph reduction, and efficient g-formula. *Biometrika*.

My Work



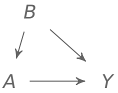
- Perković, Textor, Kalisch and Maathuis (2015). A Complete Generalized Adjustment Criterion. *UAI 2015*.
- Perković, Kalisch and Maathuis (2017). Interpreting and Using CPDAGs with Background Knowledge. *UAI 2017*.
- Perković, Textor, Kalisch and Maathuis (2018). Complete Graphical Characterization and Construction of Adjustment Sets in Markov Equivalence Classes of Ancestral Graphs. *JMLR*.
- Perković (2020). Identifying total causal effects in MPDAGs. *UAI 2020*.
- Guo and Perković (2021). Minimal enumeration of all possible total effects in a Markov equivalence class. *AISTATS 2021*.
- Guo and Perković (2022). Efficient Least Squares for Estimating Total Effects under Linearity and Causal Sufficiency. *JMLR*.
- Henckel, Perković, and Maathuis (2022). Graphical Criteria for Efficient Total Effect Estimation via Adjustment in Causal Linear Structural Equation Models. *JRSS:B*.
- Guo, Perković, and Rotnitzky (2022). Variable elimination, graph reduction, and efficient g-formula. *Biometrika*.

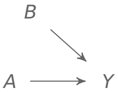
DAGs and Distributions

- Observational density $f(x_{\mathbf{V}})$
- Interventional density $f(x_{\mathbf{V}} | do(x_A))$.

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

$$f(x_{\mathbf{V}}) = \prod_{J \in \mathbf{V}} f(x_J | x_{pa(J)}) \quad \text{and} \quad f(x_{\mathbf{V}} | do(x_A)) = \prod_{J \in \mathbf{V} \setminus \{A\}} f(x_J | x_{pa(J)})$$





$$f(x_B, x_A, x_Y) = f(x_Y | x_B, x_A) f(x_A | x_B) f(x_B)$$

$$f(x_B, x_Y | do(x_A)) = f(x_Y | x_B, x_A) f(x_B)$$

How to define a causal effect?

Total causal effect

- Total causal effect, τ_{AY}, always defined as some function of $f(x_Y|do(X_A = x_A))$, E.g:

$$\tau_{AY} = \mathbb{E}[X_Y|do(X_A = x_A + 1)] - \mathbb{E}[X_Y|do(X_A = x_A)]$$

Identifiability

- A total causal effect is identifiable from observational data and a causal graph if $f(x_Y|do(x_A))$ can be expressed as a function of $f(x_{\mathbf{v}})$.

How to define a causal effect?

Total causal effect

- Total causal effect, τ_{AY}, always defined as some function of $f(x_Y|do(X_A = x_A))$, E.g:

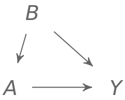
$$\tau_{AY} = \mathbb{E}[X_Y|do(X_A = x_A + 1)] - \mathbb{E}[X_Y|do(X_A = x_A)]$$

Identifiability

- A total causal effect is identifiable from observational data and a causal graph if

$f(x_Y|do(x_A))$ can be expressed as a function of $f(x_{\mathbf{v}})$.

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



$$\begin{aligned} f(x_Y|do(x_A)) &= \int f(x_B, x_Y|do(x_A))dx_B \\ &= \int f(x_Y|x_B, x_A)f(x_B)dx_B. \end{aligned}$$

G-formula (Robins ’86, Pearl ’93)

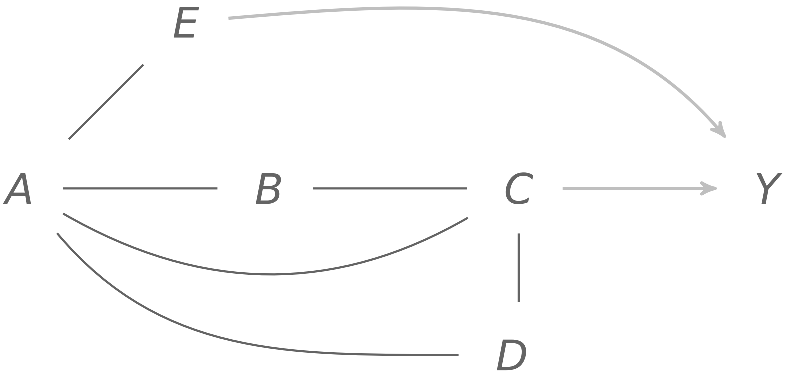
What if we don't know the DAG?

- A causal effect is **not always** identifiable from obs. data and a causal MPDAG.

Graphical criterion	DAG	CPDAG	MPDAG
Adjustment (Pearl '93, Shpitser et al '10)	$\Rightarrow$		
Generalized Adjustment (Perković et al '15, '17, '18)	$\Rightarrow$	$\Rightarrow$	$\Rightarrow$
G-formula, Truncated Factorization (Robins '86, Pearl '93)	$\Leftrightarrow$		
Generalized G-formula (Perković '20)	$\Leftrightarrow$	$\Leftrightarrow$	$\Leftrightarrow$

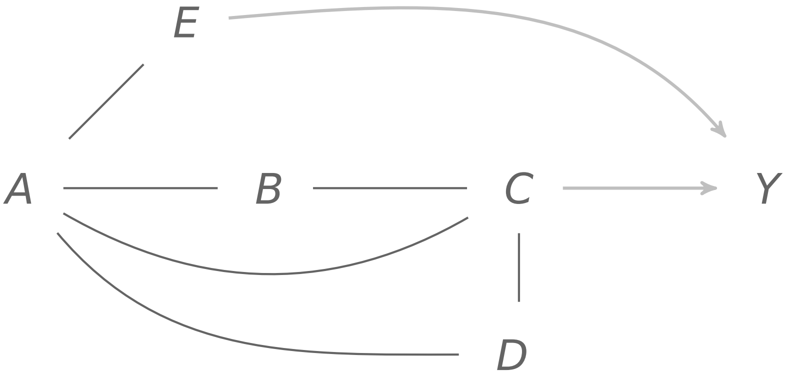
$\Rightarrow$ - sufficient for identification,
$\Leftrightarrow$ - necessary and sufficient for identification

Identifiability Condition



- Can we uniquely identify the effect?

Identifiability Condition

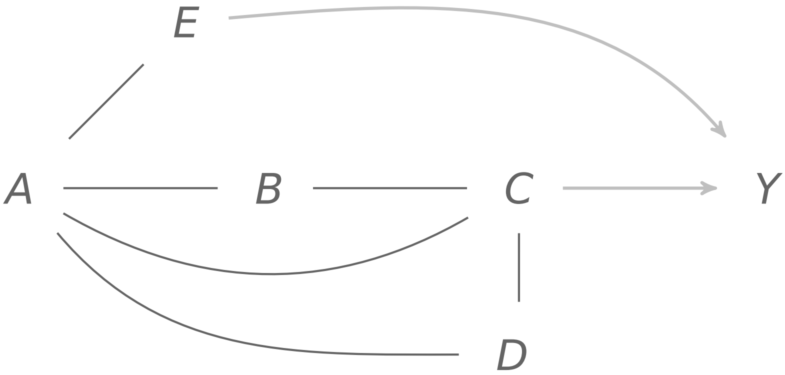


Theorem (Perković, 2020)

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

- Can we uniquely identify the effect?

Identifiability Condition

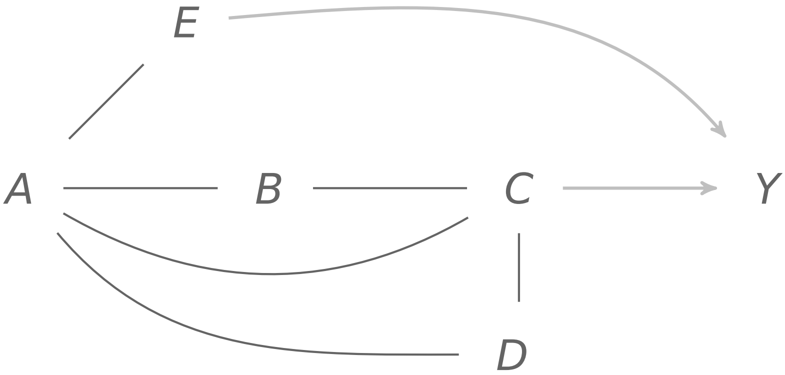


Theorem (Perković, 2020)

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

- Can we uniquely identify the effect? **No**.

Identifiability Condition



Theorem (Perković, 2020)

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

- Can we uniquely identify the effect? **No.**
- Can we identify the set of possible causal effects? **Yes.**

Set Identification

We want to have a list of possible total effects (**set identification**).

Partition of the equivalence class of DAGs such that **set identification** is

1) **complete**: $f(x_Y|\mathrm{do}(x_A))$ is identifiable under each partition

2) **minimal**: $\mathbb{E}[X_Y|\mathrm{do}(x_A)]$ are distinct functionals of x_A between partitions!

Set Identification

We want to have a list of possible total effects (**set identification**).

Partition of the equivalence class of DAGs such that **set identification** is

1) **complete**: $f(x_Y|\mathrm{do}(x_A))$ is identifiable under each partition
   We could enumerate over
   - all DAGs (Maathuis et al, '09)
   - the valid parent sets of A (Maathuis et al, '09, Nandy et al, '17, Perković et al, '17, Witte et al, '20, Fang and He, '20)
   - orientation of $A-$ on possibly causal paths to Y (Liu et al, '20)
2) **minimal**: $\mathbb{E}[X_Y|\mathrm{do}(x_A)]$ are distinct functionals of x_A between partitions!

Theorem (Perković, 2020)

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

Set Identification

We want to have a list of possible total effects (**set identification**).

Partition of the equivalence class of DAGs such that **set identification** is

1) **complete**: $f(x_Y|\text{do}(x_A))$ is identifiable under each partition
   We could enumerate over
   - all DAGs (Maathuis et al, '09)
   - the valid parent sets of A (Maathuis et al, '09, Nandy et al, '17, Perković et al, '17, Witte et al, '20, Fang and He, '20)
   - orientation of $A-$ on possibly causal paths to Y (Liu et al, '20)
2) **minimal**: $\mathbb{E}[X_Y|\text{do}(x_A)]$ are distinct functionals of x_A between partitions!
   - None of the above are minimal. Why is Liu et al, 20 not minimal?

Theorem (Perković, 2020)

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

Optimal enumeration

Theorem (Perković, 2020)

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

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

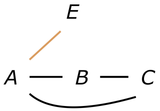
Algorithm FirstTry

1. Pick $A - V_1$ such that there is a possibly causal path $A, V_1, \ldots, Y$.
2. $\mathcal{G}_1 \leftarrow \text{MPDAG}(\mathcal{G}, A \to V_1)$, $\mathcal{G}_2 \leftarrow \text{MPDAG}(\mathcal{G}, A \leftarrow V_1)$
3. Recurse on $\mathcal{G}_1$ and $\mathcal{G}_2$ until $f(x_Y|\text{do}(x_A))$ is identified

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

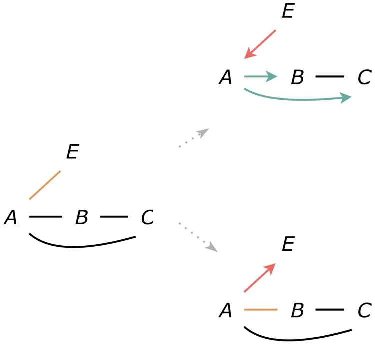
Enumeration

Omitted *D* and *Y* for simplicity.



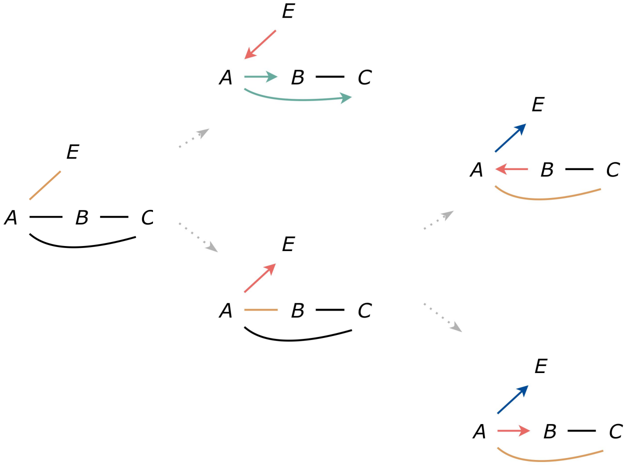
Enumeration

Omitted D and Y for simplicity.



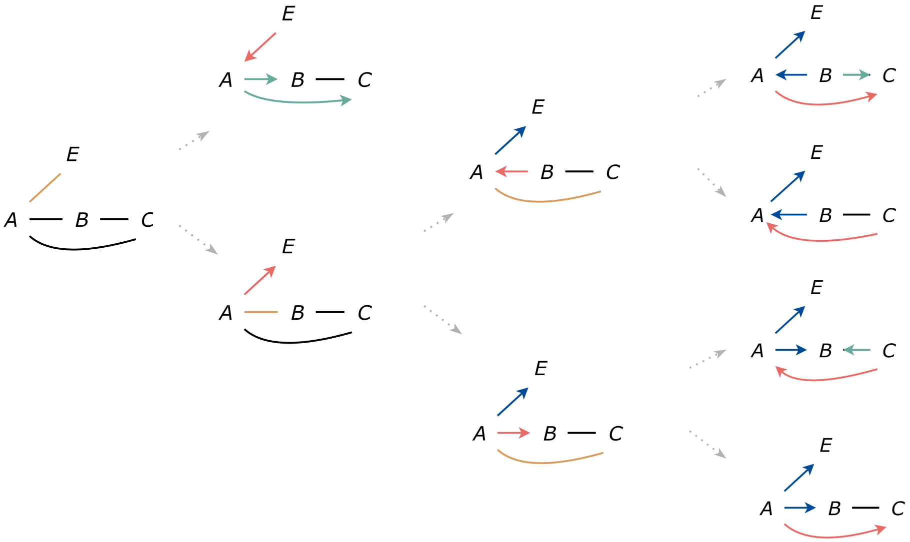
Enumeration

Omitted D and Y for simplicity.



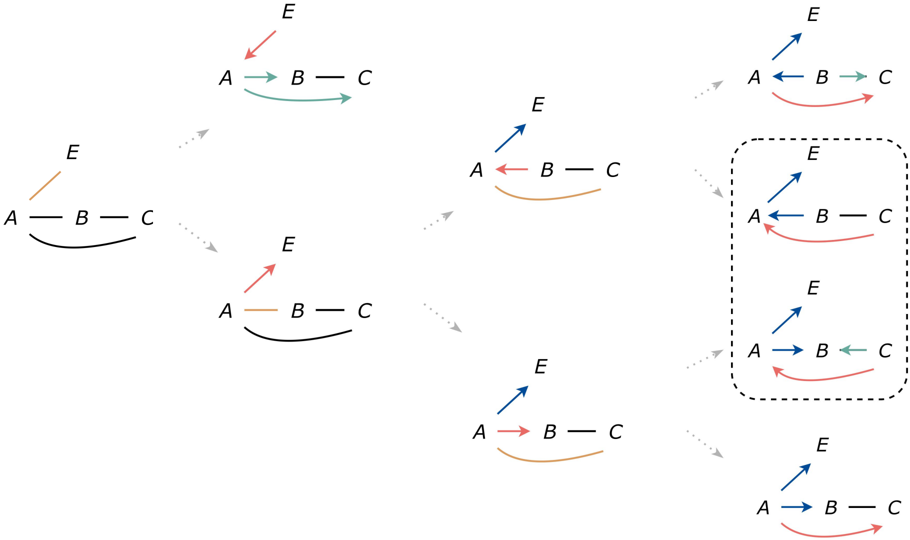
Enumeration

Omitted D and Y for simplicity.



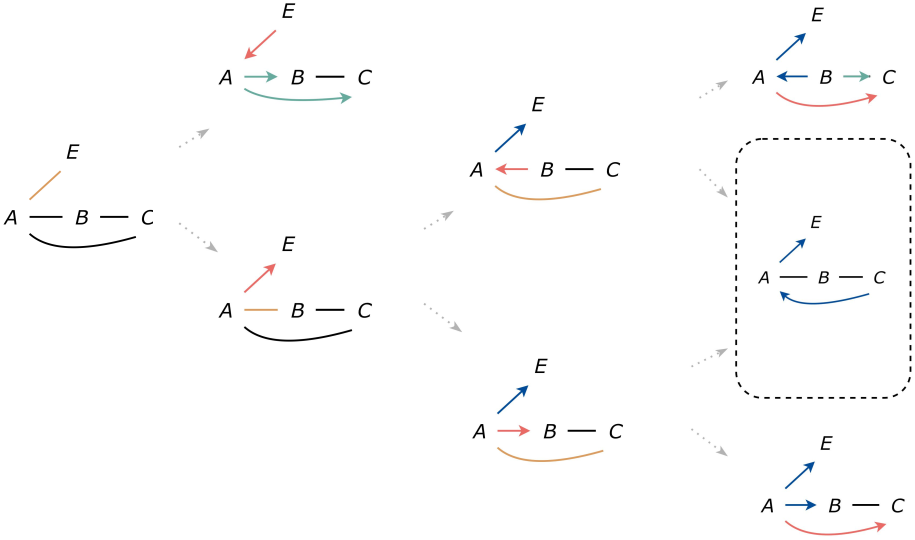
Enumeration

Omitted D and Y for simplicity.



Enumeration

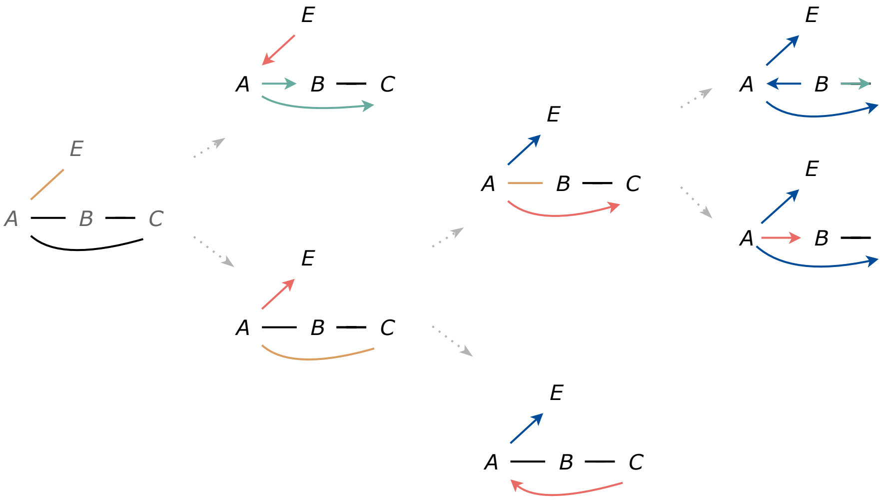
Omitted D and Y for simplicity.



Optimal Enumeration

Orienting $A - E$ then $A - C$ ...

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



Optimal Enumeration

Algorithm IDGraphs, (Guo & Perković, 2021)

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

Theorem (Guo & Perković, 2021)

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

- A small change makes a big difference!
- Have a version for the multiple exposure case as well.
- In R package `eff2`.

Marshmallow Test

Assumptions:

- Access to observational data + domain knowledge.
- Do we know all variables that explain or moderate link? Yes.
- Do we know all relationships between these variables? No.

1) Can we uniquely identify the causal effect or a set of possible effects?

   Yes (Perković 2020, Guo & Perković, 2021).

2) How strong is this causal relationship?

   - How to construct an estimator?
   - What estimator is optimal in terms of minimal variance?

Marshmallow Test

Assumptions:

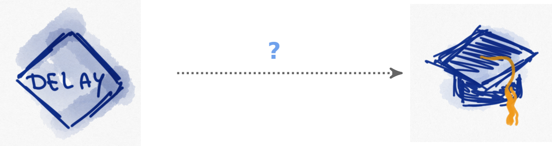
- Access to observational data + domain knowledge.
- Do we know all variables that explain or moderate link? Yes.
- Do we know all relationships between these variables? No.
- Data is generated by a linear structural causal model (SCM).

1) Can we uniquely identify the causal effect or a set of possible effects?

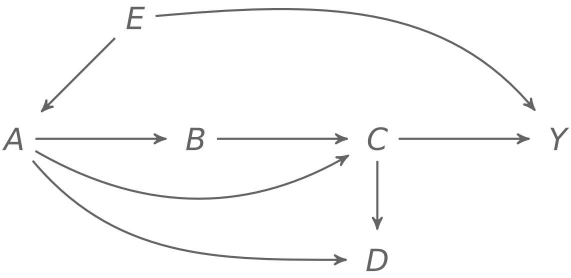
   Yes (Perković 2020, Guo & Perković, 2021).

2) How strong is this causal relationship?

   - How to construct an estimator?
   - What estimator is optimal in terms of minimal variance?



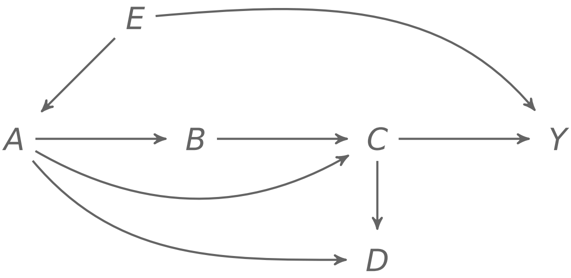
Causal DAG, Linear Structural Causal Model (SCM)



- Data is generated by:

$$
\begin{aligned}
X_E &= \epsilon_E \\
X_A &= \gamma_{EA} X_E + \epsilon_A \\
X_B &= \gamma_{AB} X_A + \epsilon_B \\
X_C &= \gamma_{AC} X_A + \gamma_{BC} X_B + \epsilon_C \\
X_D &= \gamma_{AD} X_A + \gamma_{CD} X_C + \epsilon_D \\
X_Y &= \gamma_{BY} X_B + \gamma_{CY} X_C + \gamma_{EY} X_E + \epsilon_Y \\
\mathbb{E}\,\epsilon &= 0, \quad 0 < \operatorname{var} \epsilon_i < \infty, \quad \epsilon_i \text{ are mutually independent,}
\end{aligned}
$$

Causal DAG, Linear Structural Causal Model (SCM)



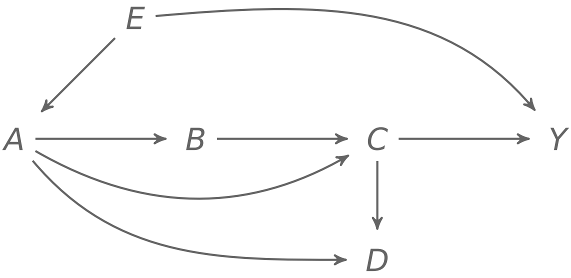
- Data is generated by:

$$X = \Gamma^\intercal X + \epsilon, \qquad \Gamma = (\gamma_{ij}), \quad I \not\to J \Rightarrow \gamma_{ij} = 0,$$

$$\mathbb{E}\,\epsilon = 0, \quad 0 < \operatorname{var}\epsilon_i < \infty, \quad \epsilon_i \text{ are mutually independent,}$$

$$\Gamma \text{ is the weighted adjacency matrix.}$$

Causal DAG, Linear Structural Causal Model (SCM)



- Data is generated by:

$$X = \Gamma^\intercal X + \epsilon, \qquad \Gamma = (\gamma_{ij}), \quad I \not\to J \Rightarrow \gamma_{ij} = 0,$$

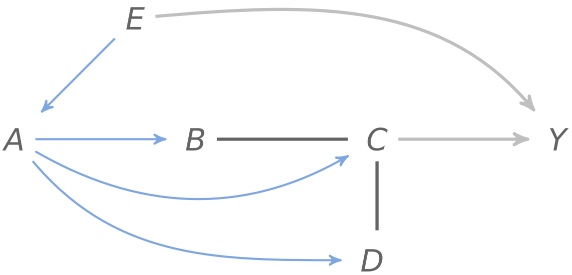
$$\mathbb{E}\,\epsilon = 0, \quad 0 < \operatorname{var} \epsilon_i < \infty, \quad \epsilon_i \text{ are mutually independent,}$$

Γ is the weighted adjacency matrix.

- By the path tracing rules (Wright, 1934) and the G-formula:

$$\tau_{AY} = \cdots = \gamma_{ac}\gamma_{cy} + \gamma_{ab}\gamma_{bc}\gamma_{cy}.$$

Block-recursive Reparametrization

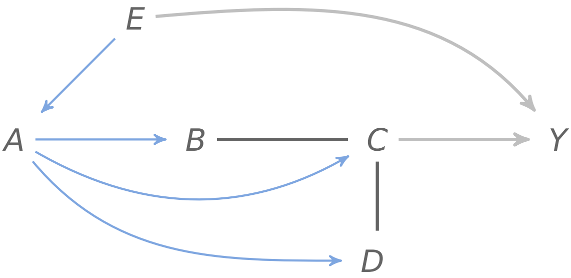


- Data is generated by

$$X = \Gamma^\intercal X + \epsilon, \qquad \Gamma = (\gamma_{ij}), \quad I \not\to J \Rightarrow \gamma_{ij} = 0,$$
$$\mathbb{E}\,\epsilon = 0, \quad 0 < \operatorname{var}\epsilon_I < \infty, \quad \epsilon_I \text{ are mutually independent.}$$

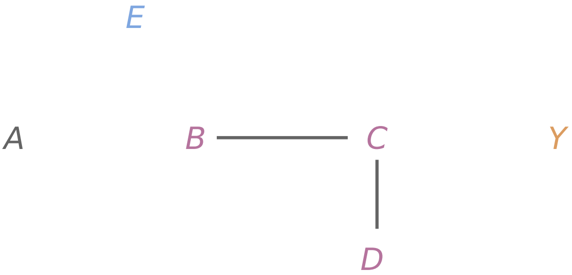
- **Problem:** Γ is **not uniquely identified**.

Block-recursive Reparametrization



- **Idea**: Consider **buckets** (maximal undirected connected components) in $\mathcal{G}$:

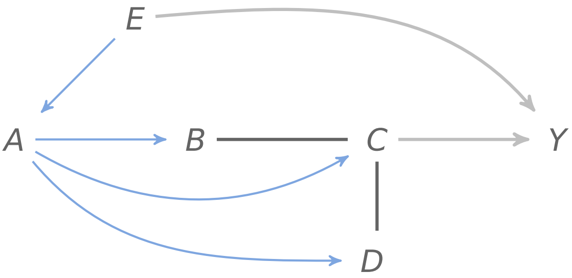
Block-recursive Reparametrization



- **Idea**: Consider **buckets** (maximal undirected connected components) in $\mathcal{G}$:

$$\mathbf{B_1} = \{E\},\ \mathbf{B_2} = \{A\},\ \mathbf{B_3} = \{B, C, D\},\ \mathbf{B_4} = \{Y\}.$$

Block-recursive Reparametrization



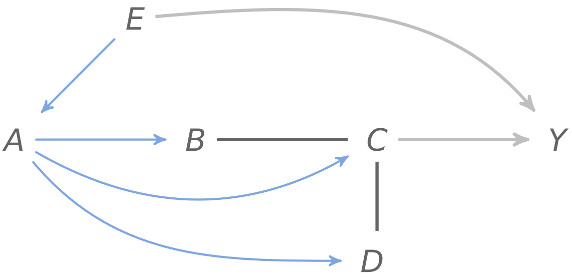
- **Idea**: Consider **buckets** (maximal undirected connected components) in $\mathcal{G}$:

$$\mathbf{B_1} = \{E\},\ \mathbf{B_2} = \{A\},\ \mathbf{B_3} = \{B, C, D\},\ \mathbf{B_4} = \{Y\}.$$

  1. The “between bucket” causal effects are **identifiable**. (Perković 2020).
  2. **Restrictive property:** Each node in a bucket has the same out-of-bucket parents (Guo and Perković, 2022).

- We use this to reparametrize the SCM.

Block-recursive Reparametrization



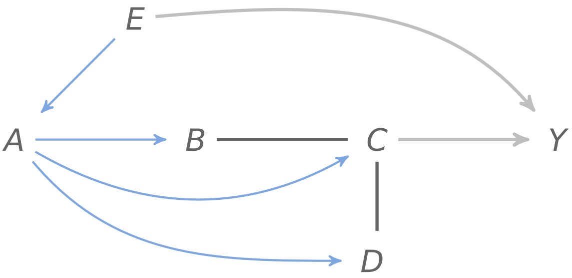
- **Idea**: Consider **buckets** (maximal undirected connected components) in $\mathcal{G}$:

$$\mathbf{B_1} = \{E\},\ \mathbf{B_2} = \{A\},\ \mathbf{B_3} = \{B, C, D\},\ \mathbf{B_4} = \{Y\}.$$

$$X_{\mathbf{B_i}} = \Gamma^{\top}_{\mathrm{pa}(\mathbf{B_i},\mathcal{G}),\mathbf{B_i}} X_{\mathrm{pa}(\mathbf{B_i},\mathcal{G})} + \Gamma^{\top}_{\mathbf{B_i}} X_{\mathbf{B_i}} + \epsilon_{\mathbf{B_i}},$$

,

Block-recursive Reparametrization



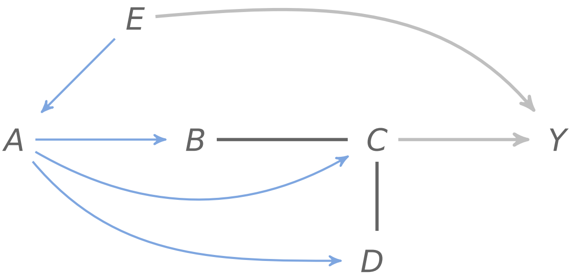
- **Idea**: Consider **buckets** (maximal undirected connected components) in $\mathcal{G}$:

$$\mathbf{B_1} = \{E\},\ \mathbf{B_2} = \{A\},\ \mathbf{B_3} = \{B, C, D\},\ \mathbf{B_4} = \{Y\}.$$

$$X_{\mathbf{B_i}} = \Gamma^{\top}_{\mathrm{pa}(\mathbf{B_i},\mathcal{G}),\mathbf{B_i}} X_{\mathrm{pa}(\mathbf{B_i},\mathcal{G})} + \Gamma^{\top}_{\mathbf{B_i}} X_{\mathbf{B_i}} + \epsilon_{\mathbf{B_i}},$$

$$\begin{aligned} X_{\mathbf{B_i}} &= \left(I - \Gamma_{\mathbf{B_i}}\right)^{-\top} \Gamma^{\top}_{\mathrm{pa}(\mathbf{B_i},\mathcal{G}),\mathbf{B_i}} X_{\mathrm{pa}(\mathbf{B_i},\mathcal{G})} + \left(I - \Gamma_{\mathbf{B_i}}\right)^{-\top} \epsilon_{\mathbf{B_i}} \\ &= \Lambda^{\top}_{\mathrm{pa}(\mathbf{B_i},\mathcal{G}),\mathbf{B_i}} X_{\mathrm{pa}(\mathbf{B_i},\mathcal{G})} + \varepsilon_{\mathbf{B_i}}, \end{aligned}$$

Block-recursive Reparametrization



- **Idea**: Consider **buckets** (maximal undirected connected components) in $\mathcal{G}$:

$$\mathbf{B_1} = \{E\},\ \mathbf{B_2} = \{A\},\ \mathbf{B_3} = \{B, C, D\},\ \mathbf{B_4} = \{Y\}.$$

$$X_{\mathbf{B_i}} = \Gamma^{\top}_{\mathrm{pa}(\mathbf{B_i},\mathcal{G}),\mathbf{B_i}} X_{\mathrm{pa}(\mathbf{B_i},\mathcal{G})} + \Gamma^{\top}_{\mathbf{B_i}} X_{\mathbf{B_i}} + \epsilon_{\mathbf{B_i}},$$

$$X_{\mathbf{B_i}} = \left(I - \Gamma_{\mathbf{B_i}}\right)^{-\top} \Gamma^{\top}_{\mathrm{pa}(\mathbf{B_i},\mathcal{G}),\mathbf{B_i}} X_{\mathrm{pa}(\mathbf{B_i},\mathcal{G})} + \left(I - \Gamma_{\mathbf{B_i}}\right)^{-\top} \epsilon_{\mathbf{B_i}}$$
$$= \Lambda^{\top}_{\mathrm{pa}(\mathbf{B_i},\mathcal{G}),\mathbf{B_i}} X_{\mathrm{pa}(\mathbf{B_i},\mathcal{G})} + \varepsilon_{\mathbf{B_i}},$$

- Suggests re-writing τ_{AY} using elements of Λ and estimating $\Lambda_{\mathrm{pa}(\mathbf{B_i},\mathcal{G}),\mathbf{B_i}}$ using least squares coefficients from $\mathbf{B_i} \sim \mathrm{pa}(\mathbf{B_i}, \mathcal{G})$ $\rightarrow$ $\mathcal{G}$-regression.

Efficiency

Theorem ($\mathcal{G}$-regression, Guo and Perković, 2022)

Suppose τ_{AY} is identifiable given MPDAG $\mathcal{G}$ and let

$\hat{\tau}^{\mathcal{G}}_{AY}$ be the **$\mathcal{G}$-regression estimator**.

Then for **any consistent estimator** $\hat{\tau}_{AY}$ of τ_{AY} such that

$\hat{\tau}_{AY}$ is a **differentiable function of the sample covariance**

it holds that

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

avar - asymptotic variance.

This includes estimators based on:

- covariate adjustment (Henckel et al, 2022, Witte et al, 2020),
- recursive regressions (Nandy et al, 2017, Gupta et al, 2020),
- modified Cholesky decomposition (Nandy et al, 2017).

Marshmallow Test

Assumptions:

- Access to observational data + domain knowledge.
- Do we know all variables that explain or moderate link? Yes.
- Do we know all relationships between these variables? No.
- Data is generated by a linear structural causal model (SCM).

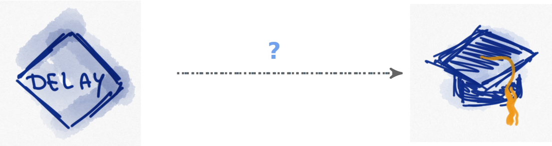
1) Can we uniquely identify the causal effect or a set of possible effects?

   Yes (Perković 2020, Guo & Perković, 2021).

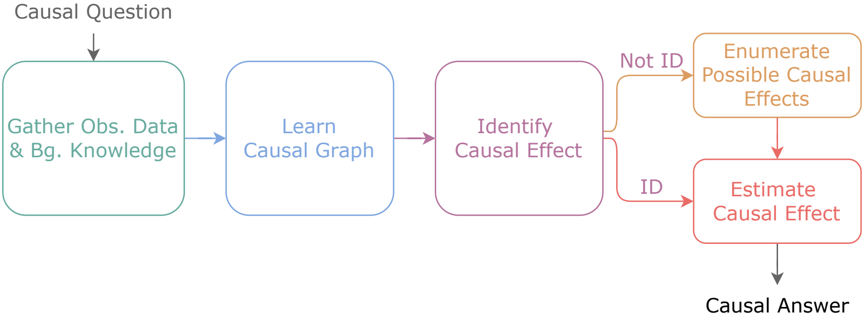
2) How strong is this causal relationship?

   - How to construct an estimator? **Generalized G-Formula** (Perković 2020, Guo & Perković, 2022, Guo, Perković, & Rotnitzky (2022)).

   - What estimator is optimal in terms of minimal variance? **$\mathcal{G}$-regression** (Guo & Perković, 2022).



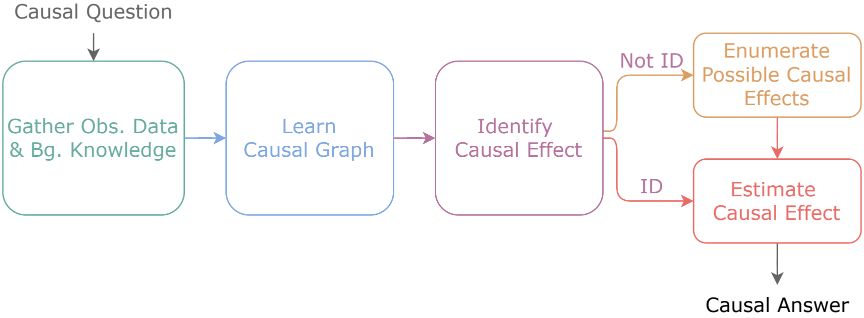
Causal Framework



Assumptions:

- Access to observational data + domain knowledge.
- Do we know all variables that explain or moderate link? Yes.
- Do we know all relationships between these variables? No.

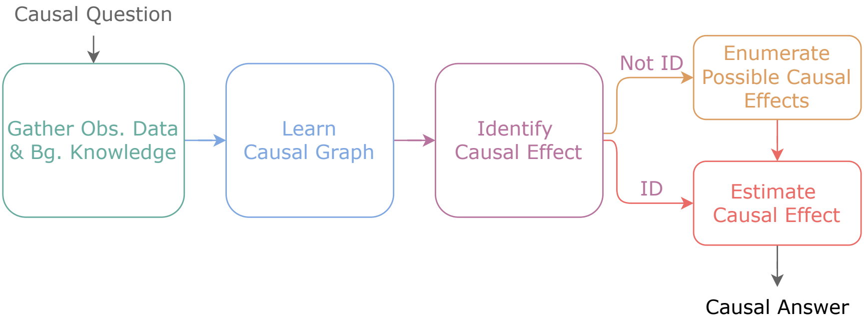
Causal Framework



Assumptions:

- Access to observational data + domain knowledge.
- Do we know all variables that explain or moderate link? No. → **Many open problems**.
- Do we know all relationships between these variables? No.

Causal Framework

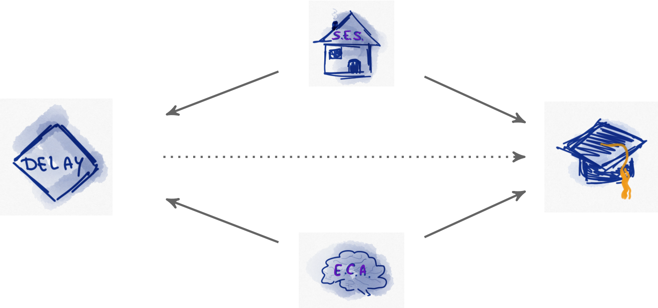


Assumptions:

- Access to observational data + domain knowledge.
- Do we know all variables that explain or moderate link? No. $\rightarrow$ **Many open problems**.
- Do we know all relationships between these variables? No.

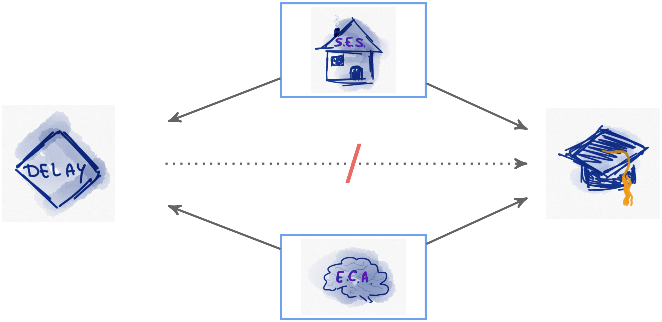
Thanks!

Marshmallow Test Revisited



- Watts, T.W., Duncan, G.J., and Quan, H. (2018) in *Psychological science.*

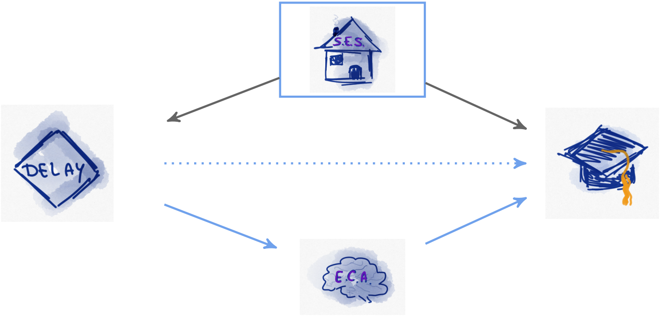
Marshmallow Test Revisited



- Watts, T.W., Duncan, G.J., and Quan, H. (2018) in *Psychological science.*

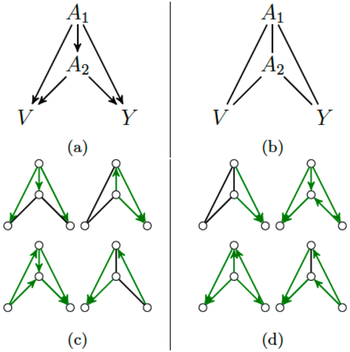
→ "...Associations between delay time and measures of behavioral outcomes at age 15 were much smaller and rarely statistically significant."

Marshmallow Test Re-Revisited



- Doebel, S., Michaelson, L.E., and Munakata, Y. (2019), *Psychological Science.*
- Falk, A., Kosse, F., and Pinger, P. (2019), *Psychological Science.*
- Watts, T.W., and Duncan, G.J. (2019), *Psychological Science.*
- Benjamin, D.J., Laibson D., **Mischel, W.**, Peake, P.K., Shoda, Y., Wellsjo, A.S., and Wilson N.W. (2020), *Journal of Economic Behavior & Organization*

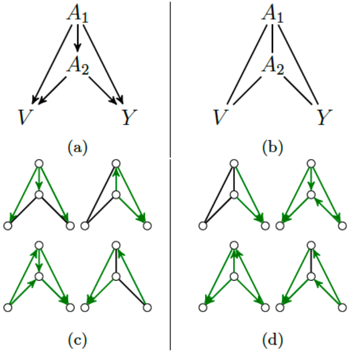
Simulation results



	A_1 on Y (c)
true effect	3
true poss. effects	$\{3, 2, 1.8, 0\}$
our method	$\{2.9, 2.1, 1.9, 0\}$
IDA (optimal)	$\{2.9, (2.1)^2, 1.9, 0\}$
IDA (local, collapsible)	$\{2.9, 2.1, 2.2, 1.9, 0\}$
joint-IDA	—

- Generated with a linear structural causal model with Gaussian errors and $n = 100$.
- $(a)^b$ denotes that a appears with multiplicity b.

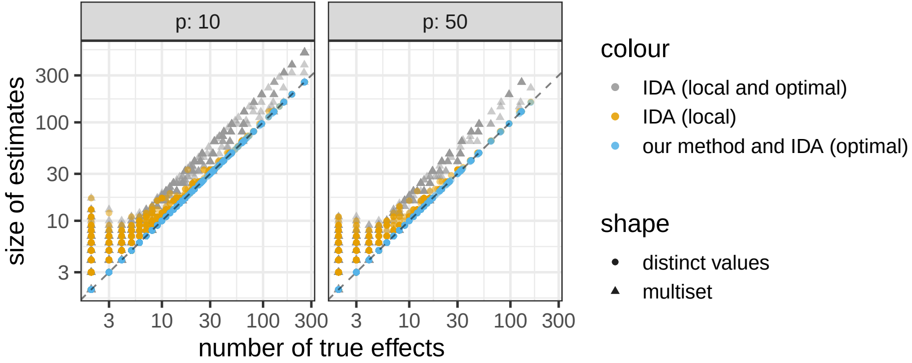
Simulation results



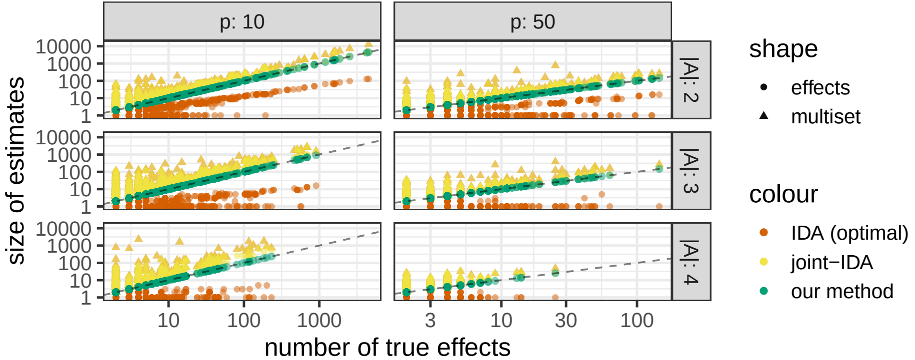
	A_1, A_2 on Y (d)
true effect	(2,1)
true poss. effects	$\{(2,1), (3,0), (0,2), (0,0)\}$
our method	$\{(2.1, 0.9), (2.9, 0), (0, 1.9), (0, 0)\}$
IDA (optimal)	$\{(2.1, 0.9)^6, (0,0)^2, (\texttt{NA}, \texttt{NA})^2\}$
IDA (local, collapsible)	—
joint-IDA	$\{(2.1, 0.9)^2, (2.2, 0.9), (1.9, 1.1),$ $(2.2, 1.1)^2, (0, 1.9), (2.9, 0), (0, 0)^2\}$

- Generated with a linear structural causal model with Gaussian errors and $n = 100$.
- $(a)^b$ denotes that a appears with multiplicity b.

Simulation: size of possible effects



Simulation: size of possible effects



Overview

	Comp. Cost	$\lvert A\rvert = 1$	$\lvert A\rvert > 1$	Duplicates
Naive - Enumerate all DAGs:				
global IDA (Maathuis et al, 2009)	$\mathcal{O}(\lvert V\rvert !)$	✓	-	Yes
global joint IDA (Nandy et al, 2017)	$\mathcal{O}(\lvert V\rvert !)$	✓	✓	Yes
Enumerate valid parent sets of A:				
local IDA (Maathuis et al, 2009, Fang & He, 2020)	$\mathcal{O}(2^{l(\mathcal{G})})$	✓	-	Yes
semi-local IDA, joint IDA (P. et al, 2017,Nandy et al, 2017)	$\mathcal{O}(2^{l(\mathcal{G})} poly(\lvert V\rvert))$	✓	✓	Yes
optimal IDA (Witte et al, 2020)	$\mathcal{O}(2^{l(\mathcal{G})} poly(\lvert V\rvert))$	✓	$\sim$	No
Enum. $A-$ on poss. causal paths to Y:				
collapsible IDA (Liu et. al, 2020)	$\mathcal{O}((\lvert V\rvert + \lvert E\rvert)2^{r(\mathcal{G})})$	✓	-	Yes
Recursively enum. over shortest problem paths				

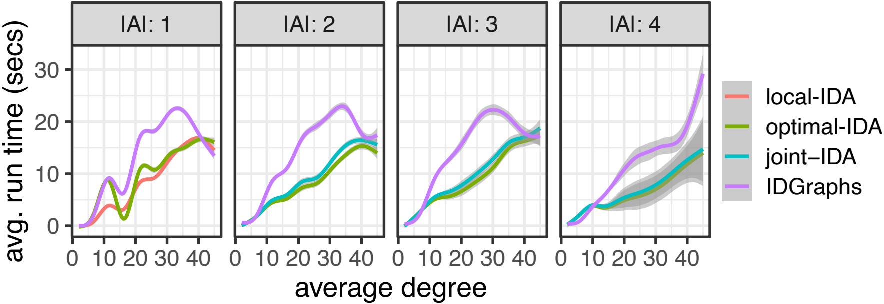
- $l(\mathcal{G})$ - # of undirected edges connected to A
- $r(\mathcal{G})$ - # of edges $A-$ on possibly causal paths to Y, $r(\mathcal{G}) \leq l(\mathcal{G})$

Overview

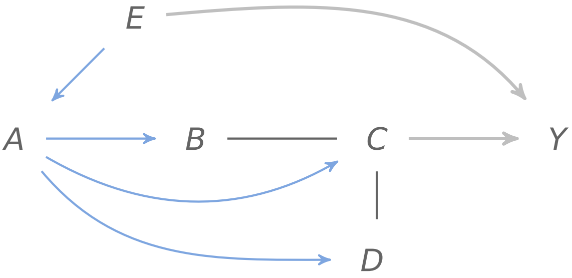
	Comp. Cost	$\lvert A\rvert = 1$	$\lvert A\rvert > 1$	Duplicates
Naive - Enumerate all DAGs:				
global IDA (Maathuis et al, 2009)	$\mathcal{O}(\lvert V\rvert!)$	✓	-	Yes
global joint IDA (Nandy et al, 2017)	$\mathcal{O}(\lvert V\rvert!)$	✓	✓	Yes
Enumerate valid parent sets of A:				
local IDA (Maathuis et al, 2009, Fang & He, 2020)	$\mathcal{O}(2^{l(\mathcal{G})})$	✓	-	Yes
semi-local IDA, joint IDA (P. et al, 2017,Nandy et al, 2017)	$\mathcal{O}(2^{l(\mathcal{G})}poly(\lvert V\rvert))$	✓	✓	Yes
optimal IDA (Witte et al, 2020)	$\mathcal{O}(2^{l(\mathcal{G})}poly(\lvert V\rvert))$	✓	$\sim$	No
Enum. $A-$ on poss. causal paths to Y:				
collapsible IDA (Liu et. al, 2020)	$\mathcal{O}((\lvert V\rvert + \lvert E\rvert)2^{r(\mathcal{G})})$	✓	-	Yes
Recursively enum. over shortest problem paths				
`IDGraphs` (Guo & Perković)	$\mathcal{O}(2^{m(\mathcal{G})}poly(\lvert V\rvert))$	✓	✓	No

- $l(\mathcal{G})$ - # of undirected edges connected to A
- $r(\mathcal{G})$ - # of edges $A-$ on possibly causal paths to Y, $r(\mathcal{G}) \leq l(\mathcal{G})$
- $m(\mathcal{G})$ - # of recursively id. edges $A-$ on proper possibly causal paths to Y, $m(\mathcal{G}) \leq r(\mathcal{G})$

Average runtime simulation comparison



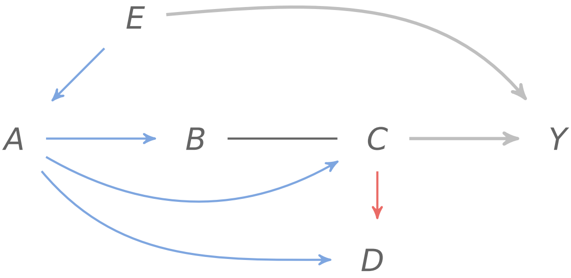
Generalized G-Formula and $\mathcal{G}$-Regression



- Generalized G-Formula and $\mathcal{G}$-regression:

$$\mathbb{E}[X_Y|\mathrm{do}(x_A)] = \int \mathbb{E}[X_Y|x_B, x_C, x_E] f(x_B, x_C|x_A) f(x_E) dx_B dx_C dx_E$$

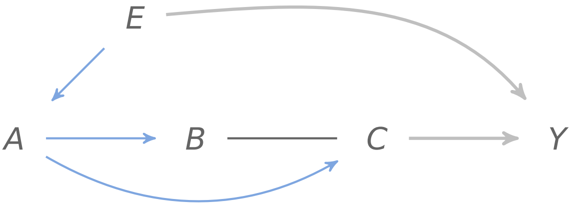
Same Generalized G-Formula and $\mathcal{G}$-Regression



- The generalized G-formula is the same in the above MPDAG.

$$\mathbb{E}[X_Y|\text{do}(x_A)] = \int \mathbb{E}[X_Y|x_B, x_C, x_E] f(x_B, x_C|x_A) f(x_E) dx_B dx_C dx_E$$

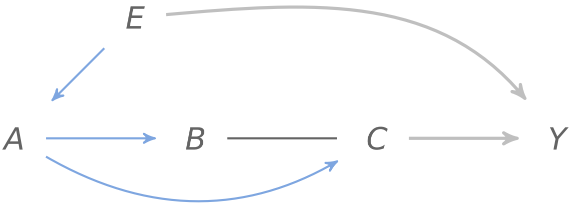
Same Generalized G-Formula and $\mathcal{G}$-Regression



- As well as in the above MPDAG.

$$\mathbb{E}[X_Y|\mathrm{do}(x_A)] = \int \mathbb{E}[X_Y|x_B, x_C, x_E]f(x_B, x_C|x_A)f(x_E)dx_B dx_C dx_E$$

Same Generalized G-Formula and $\mathcal{G}$-Regression



- As well as in the above MPDAG.

$$\mathbb{E}[X_Y|\mathrm{do}(x_A)] = \int \mathbb{E}[X_Y|x_B, x_C, x_E]f(x_B, x_C|x_A)f(x_E)dx_B dx_C dx_E$$

- Indicating that: **measurement** of all variables not needed for efficient causal estimation.
- We explore these implications in Guo, Perković and Rotnitzky (2022). Opportunities for future work.

Block-recursive reparametrization

Proposition (Block-recursive form, Guo and Perković, 2022)

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

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

Two nice things happen under this re-parametrization:

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

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

  The bucket-wise error distribution is a nuisance.
- Under Gaussian errors, the MLE for each $\Lambda_{\mathrm{pa}(\mathbf{B_i}, \mathcal{G}), \mathbf{B_i}}$ corresponds to the least squares coefficients from $\mathbf{B_i} \sim \mathrm{pa}(\mathbf{B_i}, \mathcal{G})$. $\rightarrow$ $\mathcal{G}$-regression.

Efficiency

Theorem ($\mathcal{G}$-regression, Guo and Perković, 2022)

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

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

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

Then for **any consistent estimator** $\hat{\tau}_{AY}$ of τ_{AY} such that $\hat{\tau}_{AY}$ is a **differentiable function of the sample covariance** it holds that

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

avar - asymptotic variance.

This includes estimators based on:

- covariate adjustment (Henckel et al, 2022, Witte et al, 2020),
- recursive regressions (Nandy et al, 2017, Gupta et al, 2020),
- modified Cholesky decomposition (Nandy et al, 2017).

Simulation results

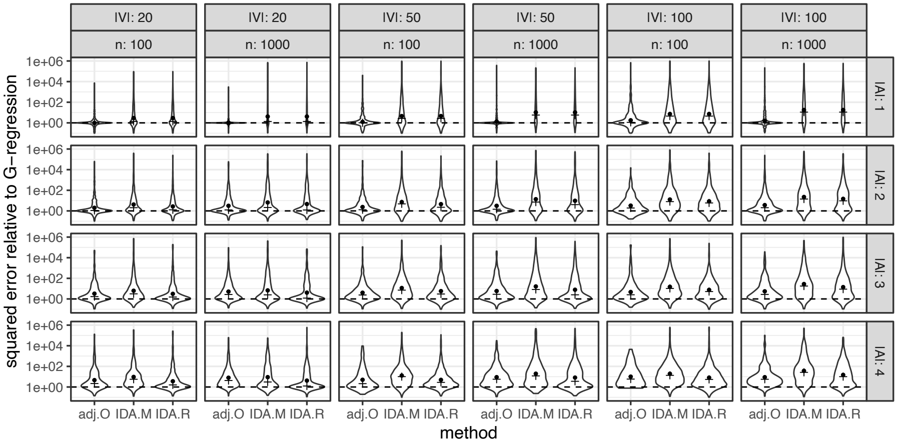
An instance is simulated by the following steps.

1. Draw $\mathcal{D}$ from a random graph ensemble.
2. Take $\mathcal{G} = \text{CPDAG}(\mathcal{D})$.
3. Simulate data from a linear SCM with random error type (normal, t, logistic, uniform).
4. Choose (A, Y) such that τ_{AY} is identified from $\mathcal{G}$.
5. Compute squared error $err = \|\tau_{AY} - \hat{\tau}_{AY}\|^2$.

We compare $\mathcal{G}$-regression to the following estimators:

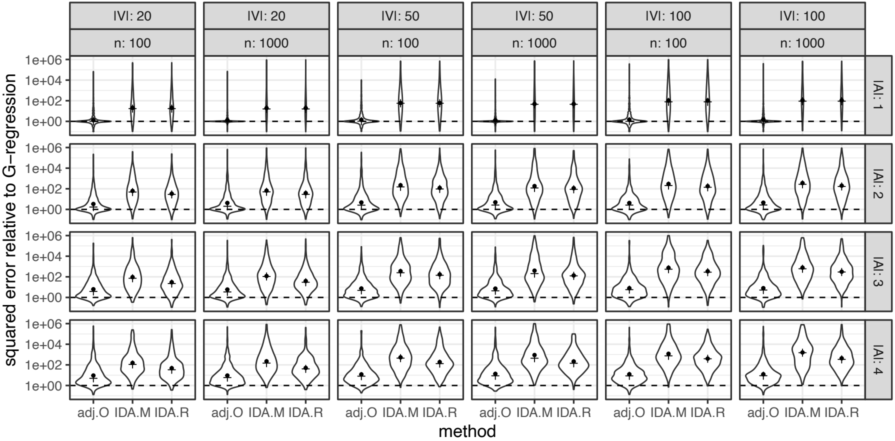
- `adj.O`: optimal adjustment estimator (Henckel et al, 2022), or
- `IDA.M`: joint-IDA estimator based on modifying Cholesky decompositions (Nandy et al, 2017), or
- `IDA.R`: joint-IDA estimator based on recursive regressions (Nandy et al, 2017).

Simulation results



Violin plots displaying relative squared errors $\frac{estimator.err}{\mathcal{G}-reg.err}$ given GES estimated CPDAG.

Simulation results



Violin plots displaying relative squared errors $\frac{estimator.err}{\mathcal{G}-reg.err}$ given the true DAG.

Simulation results

Table: Percentage of identified instances not estimable using contending estimators. All instances are estimable with $\mathcal{G}$-regression.

Estimator	$\|\mathbf{A}\|$	$\|\mathbf{V}\| = 20$	$\|\mathbf{V}\| = 50$	$\|\mathbf{V}\| = 100$
`adj.O`	1	0%	0%	0%
	2	17%	10%	5%
	3	30%	18%	15%
	4	36%	29%	22%
`IDA.M`	1	29%	32%	32%
	2	47%	51%	50%
	3	61%	59%	63%
	4	72%	69%	71%
`IDA.R`	1	29%	32%	32%
	2	47%	51%	50%
	3	61%	59%	63%
	4	72%	69%	71%

Simulation results

Table: Geometric average of squared errors relative to $\mathcal{G}$-regression, computed from estimable instances.

$\lvert\mathbf{A}\rvert$	$\lvert\mathbf{V}\rvert = 20$ $n = 100$	$\lvert\mathbf{V}\rvert = 20$ $n = 1000$	$\lvert\mathbf{V}\rvert = 50$ $n = 100$	$\lvert\mathbf{V}\rvert = 50$ $n = 1000$	$\lvert\mathbf{V}\rvert = 100$ $n = 100$	$\lvert\mathbf{V}\rvert = 100$ $n = 1000$
`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 results

Table: Geometric average of squared errors relative to $\mathcal{G}$-regression, computed from estimable instances given GES estimated CPDAG

	$\lvert\mathbf{V}\rvert = 20$		$\lvert\mathbf{V}\rvert = 50$		$\lvert\mathbf{V}\rvert = 100$	
$\lvert\mathbf{A}\rvert$	$n = 100$	$n = 1000$	$n = 100$	$n = 1000$	$n = 100$	$n = 1000$
`adj.O`						
1	1.0	1.0	1.2	1.3	1.8	1.6
2	2.0	3.1	2.4	3.1	3.2	3.7
3	3.3	5.2	4.0	5.9	4.7	5.5
4	4.6	7.9	5.0	9.0	10	8.9
`IDA.M`						
5	2.9	4.1	4.5	10	7.3	18
6	4.2	6.6	7.3	14	13	22
7	6.2	6.8	12	16	15	28
8	9.5	9.0	13	20	19	37
`IDA.R`						
9	2.9	4.1	4.5	10	7.3	18
10	2.7	4.6	4.5	9.6	8.5	15
11	3.1	4.1	5.8	7.8	7.6	14
12	3.6	4.2	4.9	8.2	8.1	15

Identification of total causal effect

$\mathbf{S_1}, \ldots, \mathbf{S_K}$ is a partition of $\mathbf{S} = An(Y, \mathcal{G}_{\mathbf{V}\setminus\{A\}})$ induced by $\mathbf{B_1}, \ldots, \mathbf{B_K}$.
Let $\mathbf{F_k} = \{A\} \cap \mathrm{pa}(\mathbf{S_k}, \mathcal{G})$, for all $k \in \{1, \ldots, k\}$. Then

$$P(X_\mathbf{S}|\mathrm{do}(x_A)) = \prod_{k=1}^{K} P(X_{\mathbf{S_k}}|X_{\mathrm{pa}(\mathbf{S_k},\mathcal{G})}) = \prod_{k=1}^{K} P(X_{\mathbf{S_k}}|X_{\mathrm{pa}(\mathbf{S_k},\mathcal{G})\setminus\mathbf{F_k}}, X_{\mathbf{F_k}} = x_{\mathbf{F_k}}),$$

where $x_{\mathbf{F_k}}$ is fixed by the $\mathrm{do}(x_A)$ operation.

$$\begin{aligned} X_{\mathbf{S_k}} \mid \left\{X_{\mathrm{pa}(\mathbf{S_k},\mathcal{G})\setminus\mathbf{F_k}}, X_{F_i} = x_{\mathbf{F_k}}\right\} &=_d \Lambda^{\mathsf{T}}_{\mathrm{pa}(\mathbf{S_k},\mathcal{G})\setminus\mathbf{F_k},\mathbf{S_k}} X_{\mathrm{pa}(\mathbf{S_k},\mathcal{G})\setminus\mathbf{F_k}} + \Lambda_{\mathbf{F_k},\mathbf{S_k}} x_{\mathbf{F_k}} + \varepsilon_{\mathbf{S_k}} \\ &= \Lambda^{\mathsf{T}}_{\mathrm{pa}(\mathbf{S_k},\mathcal{G})\cap\mathbf{S},\mathbf{S_k}} X_{\mathrm{pa}(\mathbf{S_k},\mathcal{G})\cap\mathbf{S}} + \Lambda_{\mathrm{pa}(\mathbf{S_k},\mathcal{G})\cap\{A\},\mathbf{S_k}} x_{\mathrm{pa}(\mathbf{S_k},\mathcal{G})\cap\{A\}} + \varepsilon_{\mathbf{S_k}} \end{aligned}$$

The fact that the display above holds for every $k = 1, \ldots, K$ implies that the joint interventional distribution $P(X_\mathbf{S}|\mathrm{do}(x_A))$ satisfies

$$X_\mathbf{S} = \Lambda^{T}_{\mathbf{S},\mathbf{S}} X_\mathbf{S} + \Lambda^{\mathsf{T}}_{A,\mathbf{S}} x_A + \varepsilon_\mathbf{S}.$$

It follows that $X_\mathbf{S} = (I - \Lambda_{\mathbf{S},\mathbf{S}})^{-\mathsf{T}}(\Lambda^{\mathsf{T}}_{A,\mathbf{S}} x_A + \varepsilon_\mathbf{S})$ and since $Y \in \mathbf{S}$, we have

$$\tau_{AY} = \frac{\partial}{\partial x_A}\,\mathbb{E}[X_Y \mid \mathrm{do}(x_A)] = \Lambda_{A,\mathbf{S}}\left[(I - \Lambda_{\mathbf{S},\mathbf{S}})^{-1}\right]_{\mathbf{S},Y}.$$

Efficiency theory

Let Σ_n be the sample covariance. Consider the class of estimators

$$\mathcal{T} = \Big\{ \hat{\tau}(\Sigma_n) : \mathbb{R}_{\text{PD}}^{|\mathbf{V}|\times|\mathbf{V}|} \to \mathbb{R}^{|\mathbf{A}|} : \hat{\tau}(\Sigma_n) \text{ is a differentiable and consistent estimator of } \tau_{AY} \Big\}.$$

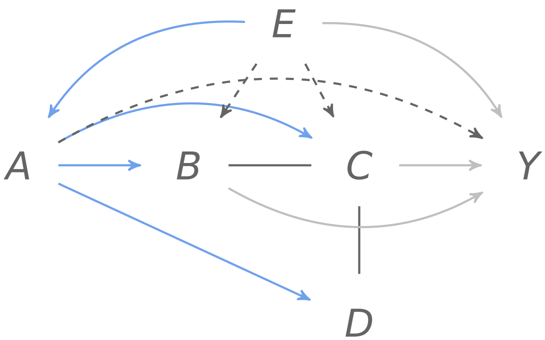
The efficiency theory entails two parts.

- Establish an efficiency bound on $\mathcal{T}$.
  The bound is derived from the gradient condition on $\mathcal{T}$ (as in standard semiparametric efficiency theory) and a **diffeomorphism**

  $$\mathbb{R}_{\text{PD}}^{|\mathbf{V}|\times|\mathbf{V}|} \longleftrightarrow ((\Lambda_{\text{pa}(\mathbf{B_k},\bar{\mathcal{G}}),\mathbf{B_k}}, \Omega_k) : k = 1, \ldots, K) \text{ associated with } \bar{\mathcal{G}},$$

  where $\bar{\mathcal{G}}$ is the saturated version of $\mathcal{G}$.
  This generalizes a result from Drton (2018).
- Verify that $\hat{\tau}_{AY}^{\mathcal{G}}$ achieves this bound.

Efficiency theory



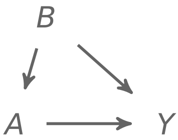
Saturated $\bar{\mathcal{G}}$ according to buckets.

$$\mathbf{B_1} = \{E\},\ \mathbf{B_2} = \{A\},\ \mathbf{B_3} = \{B, C, D\},\ \mathbf{B_4} = \{Y\}.$$

Proof sketch

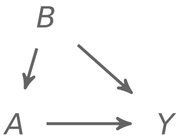
1. Suppose $|\mathbf{A}| = 1$. Rewrite $\hat{\tau} \in \mathcal{T}$ as
$$\hat{\tau}(\Sigma_n) = \hat{\tau}\left((\hat{\Lambda}_k)_{k,\mathcal{G}}, (\hat{\Lambda}_k)_{k,\mathcal{G}^c}, (\hat{\Omega}_k)_k\right),$$
where $(\hat{\Lambda}_k)_{k,\mathcal{G}^c} = (\hat{\Lambda}_k)_{k,\bar{\mathcal{G}}\setminus\mathcal{G}}$ are introduced dashed edges.
2. Consistency of $\hat{\tau}$ implies
$$\frac{\partial \hat{\tau}}{\partial \hat{\Lambda}_{k,\mathcal{G}}} = \frac{\partial \tau_{\mathcal{G}}}{\partial \hat{\Lambda}_{k,\mathcal{G}}} \ (k = 2, \ldots, K), \quad \frac{\partial \hat{\tau}}{\partial \hat{\Omega}_k} = \mathbf{0} \ (k = 1, \ldots, K),$$
but $\frac{\partial \hat{\tau}}{\partial \hat{\Lambda}_{k,\mathcal{G}^c}}$ is free to vary.
3. Compute acov of $\left((\hat{\Lambda}_{k,\mathcal{G}})_k, (\hat{\Lambda}_{k,\mathcal{G}^c})_k\right)$ via asymptotic linear expansions.
4. By the delta method, an upper bound can be derived from quadratic form
$$\begin{aligned}\operatorname{avar}(\hat{\tau}) &= \begin{pmatrix} \frac{\partial \hat{\tau}}{\partial (\hat{\Lambda}_{k,\mathcal{G}})_k} \\ \frac{\partial \hat{\tau}}{\partial (\hat{\Lambda}_{k,\mathcal{G}^c})_k} \end{pmatrix}^{\mathsf{T}} \operatorname{acov}\left((\hat{\Lambda}_{k,\mathcal{G}})_k, (\hat{\Lambda}_{k,\mathcal{G}^c})_k\right) \begin{pmatrix} \frac{\partial \hat{\tau}}{\partial (\hat{\Lambda}_{k,\mathcal{G}})_k} \\ \frac{\partial \hat{\tau}}{\partial (\hat{\Lambda}_{k,\mathcal{G}^c})_k} \end{pmatrix} \\ &\leq \sup_{\partial\hat{\tau}/\partial(\hat{\Lambda}_{k,\mathcal{G}^c})_k} \begin{pmatrix} \frac{\partial \hat{\tau}}{\partial (\hat{\Lambda}_{k,\mathcal{G}})_k} \\ \frac{\partial \hat{\tau}}{\partial (\hat{\Lambda}_{k,\mathcal{G}^c})_k} \end{pmatrix}^{\mathsf{T}} \operatorname{acov}\left((\hat{\Lambda}_{k,\mathcal{G}})_k, (\hat{\Lambda}_{k,\mathcal{G}^c})_k\right) \begin{pmatrix} \frac{\partial \hat{\tau}}{\partial (\hat{\Lambda}_{k,\mathcal{G}})_k} \\ \frac{\partial \hat{\tau}}{\partial (\hat{\Lambda}_{k,\mathcal{G}^c})_k} \end{pmatrix}.\end{aligned}$$

What if we don't know the DAG?

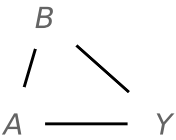


DAG

What if we don’t know the DAG?

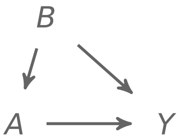


DAG

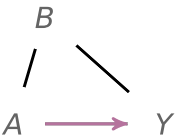


Completed Partially Directed Acyclic Graph (CPDAG)

What if we don't know the DAG?

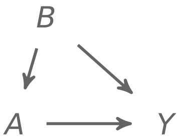


DAG

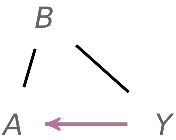


Completed Partially Directed Acyclic Graph (CPDAG)

What if we don't know the DAG?

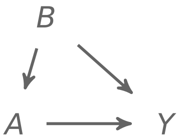


DAG

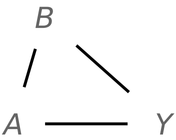


Completed Partially Directed
Acyclic Graph (CPDAG)

What if we don't know the DAG?

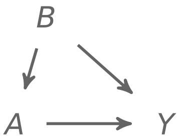


DAG

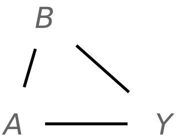


Completed Partially Directed Acyclic Graph (CPDAG)

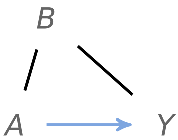
What if we don't know the DAG?



DAG

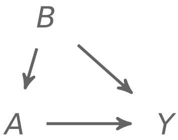


Completed Partially Directed Acyclic Graph (CPDAG)

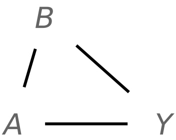


Maximally Oriented PDAG (MPDAG)

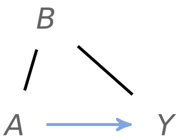
What if we don’t know the DAG?



DAG



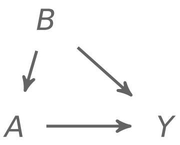
Completed Partially Directed Acyclic Graph (CPDAG)



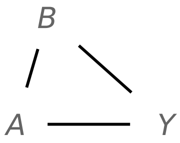
Maximally Oriented PDAG (MPDAG)

- A causal effect is **not always** identifiable from obs. data and a causal MPDAG.

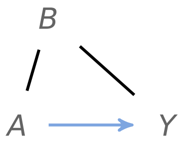
What if we don't know the DAG?



DAG



Completed Partially Directed Acyclic Graph (CPDAG)



Maximally Oriented PDAG (MPDAG)

- A causal effect is **not always** identifiable from obs. data and a causal MPDAG.

Graphical criterion	DAG	CPDAG	MPDAG
Adjustment (Pearl '93, Shpitser et al '10)	$\Rightarrow$		
Generalized Adjustment (Perković et al '15, '17, '18)	$\Rightarrow$	$\Rightarrow$	$\Rightarrow$
G-formula, Truncated Factorization (Robins '86, Pearl '93)	$\Leftrightarrow$		
Generalized G-formula (Perković '20)	$\Leftrightarrow$	$\Leftrightarrow$	$\Leftrightarrow$

$\Rightarrow$ - sufficient for identification,
$\Leftrightarrow$ - necessary and sufficient for identification