Identifying causal effects in MPDAGs

Emilija Perković
University of Washington

Goal

- Identify the total causal effect of **X** on **Y**
  - the change in **Y** due to $do(\mathbf{x})$-
  from observational data.

- $do(\mathbf{x})$: an intervention that sets variables **X** to **x**.
  $f(\mathbf{y}|do(\mathbf{x})) \neq f(\mathbf{y}|\mathbf{x})$.

Observational data

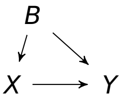
Randomized control studies

DAGs and distributions

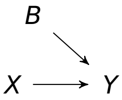
- $do(\mathbf{x})$: an intervention that sets variables $\mathbf{X}$ to $\mathbf{x}$.
- Observational density $f(\mathbf{v})$, Interventional density $f(\mathbf{v}|do(\mathbf{x}))$.

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

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



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



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

How to define a causal effect?

Total causal effect

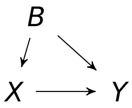
- Total causal effect - $\tau_{\mathbf{yx}}$ - is *some functional* of $f(\mathbf{y}|do(\mathbf{x}))$, $P(\mathbf{Y}|do(\mathbf{x}))$.
- Examples: $E[Y|do(X = x+1)] - E[Y|do(X = x)]$, $\frac{\partial}{\partial x}E(Y|do(x))$, OR, RR...

Identifiability

- A causal effect is identifiable from observational data if

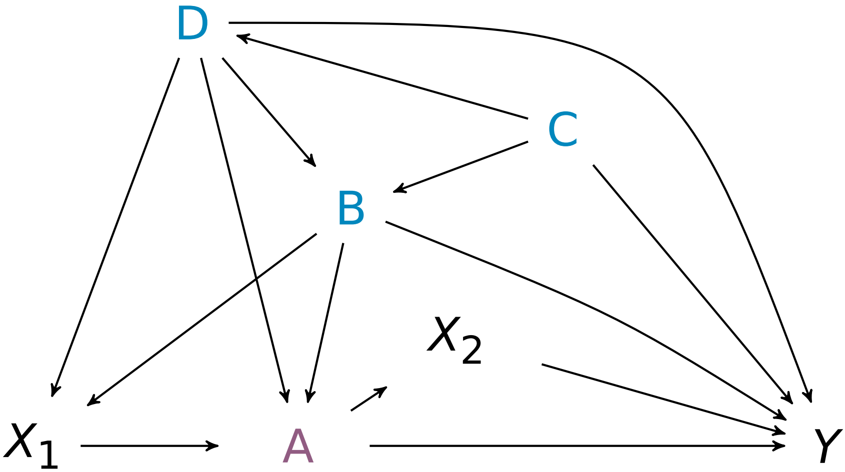
$$f(\mathbf{y}|do(\mathbf{x})) \text{ is computable from } f(\mathbf{v}).$$

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



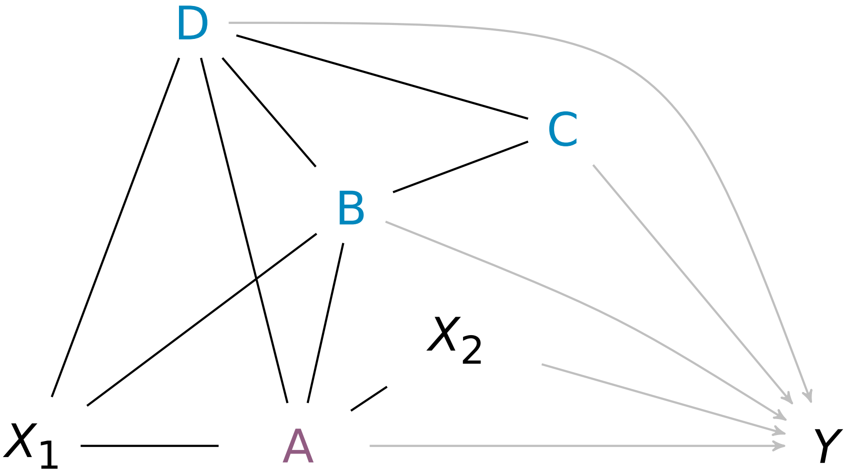
$$f(y|do(x)) = \int f(b, y|do(x))db$$
$$= \int f(y|b, x)f(b)db.$$

Problem solved?



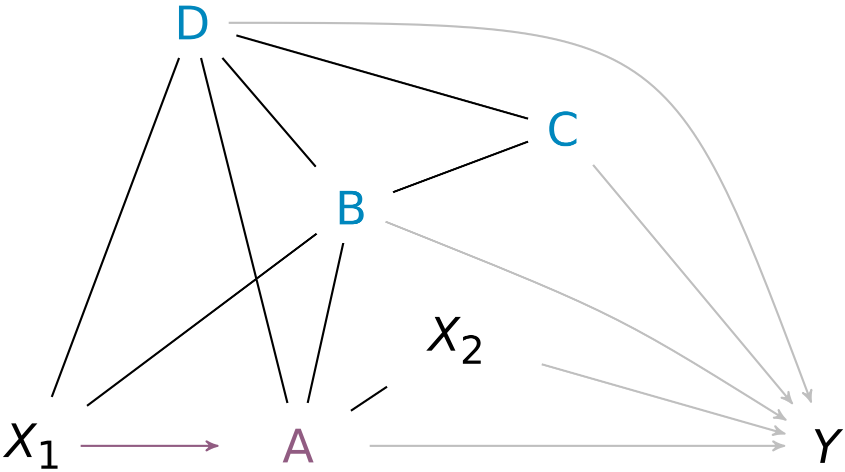
DAG $\mathcal{D}$.

Problem solved?



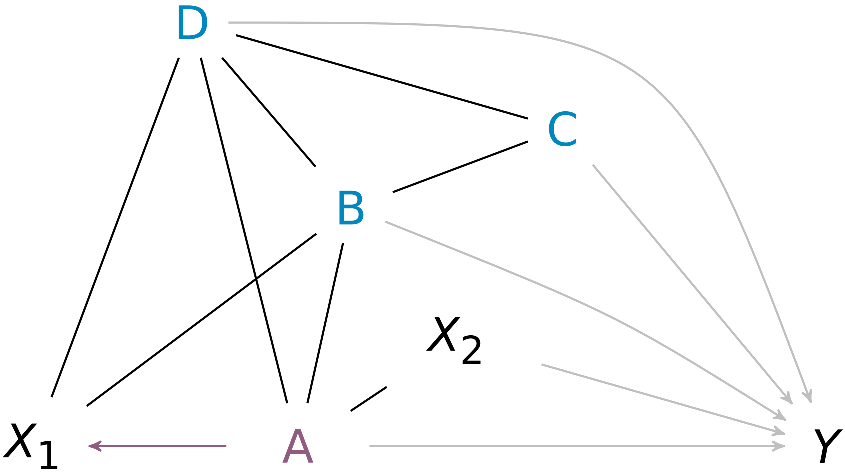
Completed Partially Directed Acyclic Graph (CPDAG) $\mathcal{C}$.

Problem solved?



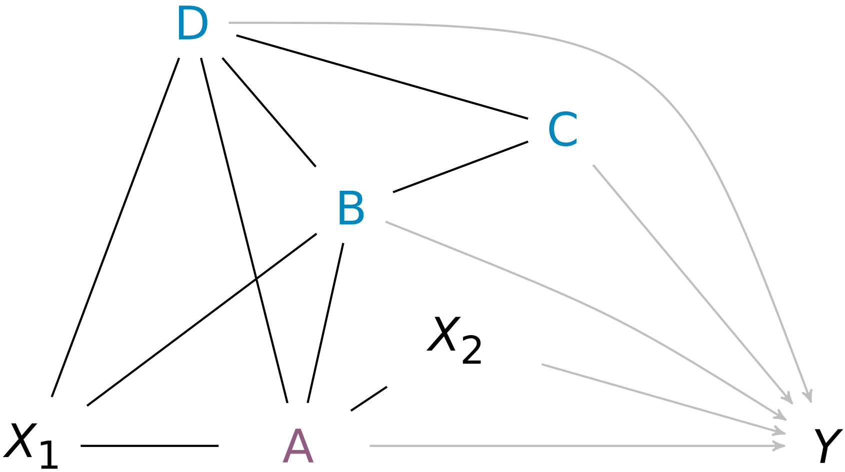
Completed Partially Directed Acyclic Graph (CPDAG) $\mathcal{C}$.

Problem solved?



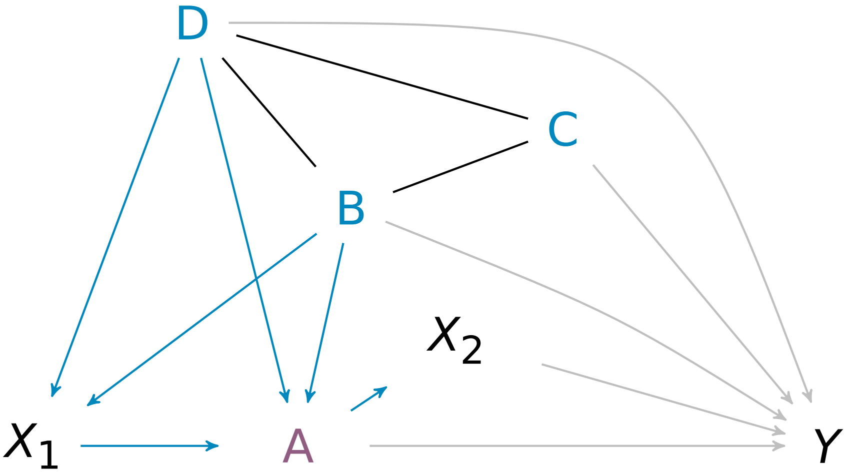
Completed Partially Directed Acyclic Graph (CPDAG) $\mathcal{C}$.

Problem solved?



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

Problem solved?



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

Overview of graphical criteria for identification

Graphical criterion	DAG	CPDAG	MPDAG
Generalized adjustment (Shpitser et al '10, Perković et al '15, '17, '18)	$\Rightarrow$	$\Rightarrow$	$\Rightarrow$
G-formula (Robins '86)	$\Leftrightarrow$		
Causal identification formula (Perković '20)	$\Leftrightarrow$	$\Leftrightarrow$	$\Leftrightarrow$

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

G-formula: Let $\mathbf{V}' = \mathbf{V} \setminus \{\mathbf{X} \cup \mathbf{Y}\}$, then

$$f(\mathbf{y}|do(\mathbf{x})) = \int \prod_{V_i \in \mathbf{V} \setminus \mathbf{X}} f(v_i|pa(v_i, \mathcal{D}))d\mathbf{v}'.$$

Causal identification formula

Theorem (Perković, 2020)

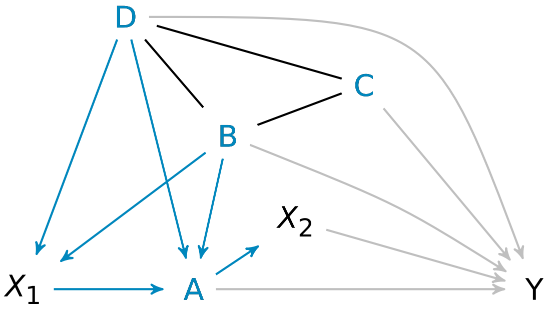
If **all proper possibly causal paths** from **X** to **Y** start with a directed edge in $\mathcal{G}$, then

$$f(\mathbf{y}|do(\mathbf{x})) = \int \prod_{i=1}^{k} f(\mathbf{s_i}|pa(\mathbf{s_i}, \mathcal{G}))d\mathbf{s},$$

where $\mathbf{S} = an(\mathbf{Y}, \mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}) \setminus \mathbf{Y}$,
and $(\mathbf{S_1}, \dots, \mathbf{S_k})$ is a partition of $\mathbf{S} \cup \mathbf{Y}$ into undirected connected sets in $\mathcal{G}$.

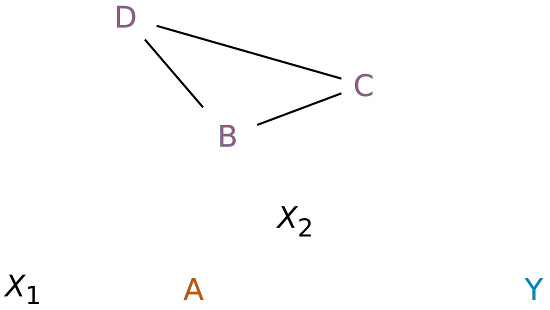
- $\mathbf{S} \cup \mathbf{Y} = an(\mathbf{Y}, \mathcal{G}_{\mathbf{V}\setminus\mathbf{X}})$ - nodes that have a causal path to **Y** that is not through **X**.
- $(\mathbf{S_1}, \dots, \mathbf{S_k})$ - maximal connected components of $\mathbf{S} \cup \mathbf{Y}$ in the induced undirected subgraph of $\mathcal{G}$.

$f(y|do(x_1, x_2)) = ?$



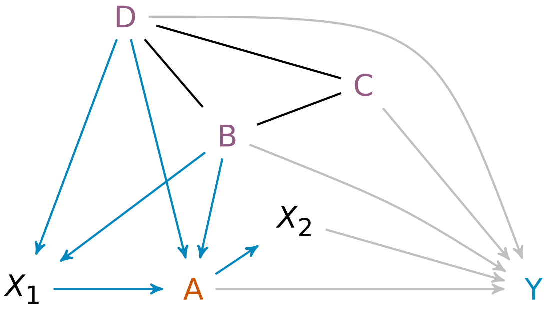
- $\mathbf{S} = an(Y, \mathcal{G}_{\mathbf{V} \setminus \mathbf{X}}) \setminus \{Y\} = \{A, B, C, D\}$, Partition of $\mathbf{S} \cup \{Y\} =$

$f(y|do(x_1, x_2)) = ?$



- $\mathbf{S} = an(Y, \mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}) \setminus \{Y\} = \{A, B, C, D\}$, Partition of $\mathbf{S} \cup \{Y\} = (\{B, C, D\}, \{A\}, \{Y\})$.

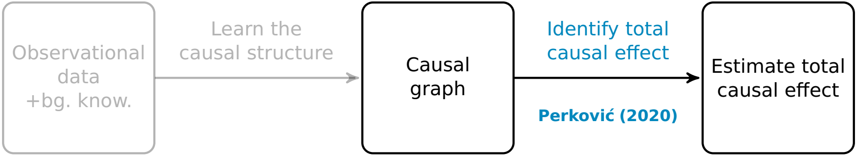
$f(y|do(x_1, x_2)) = ?$



- $\mathbf{S} = an(Y, \mathcal{G}_{\mathbf{V} \setminus \mathbf{X}}) \setminus \{Y\} = \{A, B, C, D\}$, Partition of $\mathbf{S} \cup \{Y\} = (\{B, C, D\}, \{A\}, \{Y\})$.

$$f(y|do(x_1, x_2)) = \int f(y|a, b, c, d, x_2) f(a|b, d, x_1) f(b, c, d) d\mathbf{s}.$$

Summary



- A necessary condition for identification of $f(\mathbf{y}|do(\mathbf{x}))$.
- Necessary and sufficient graphical criterion for identification of causal effects.
- Proposition on the “necessity” of adjustment.
- Slides, both long and short version, on my webpage: https://emilijaperkovic.com/projects/

Thanks!