Causal Inference: DAGs, reasoning and discovery

Charles Assaad - Lecture
Timothée Loranchet - First graded lab
Simon Ferreira - Second graded lab

February 2026

SANTÉ SORBONNE UNIVERSITÉ

Sommaire

SANTÉ SORBONNE UNIVERSITÉ

1

Causal hierarchy

Some epidemiological questions

Some epidemiological questions

- Do people who exercise more tend to have lower rates of heart disease?

Some epidemiological questions

- Do people who exercise more tend to have lower rates of heart disease?
- Can we predict who is at higher risk of developing diabetes using BIG DATA?

Some epidemiological questions

- Do people who exercise more tend to have lower rates of heart disease?
- Can we predict who is at higher risk of developing diabetes using BIG DATA?
- Does exercising reduce cholesterol?

Some epidemiological questions

- Do people who exercise more tend to have lower rates of heart disease?
- Can we predict who is at higher risk of developing diabetes using BIG DATA?
- Does exercising reduce cholesterol?
- What happens to blood pressure if we prescribe a new hypertension drug?

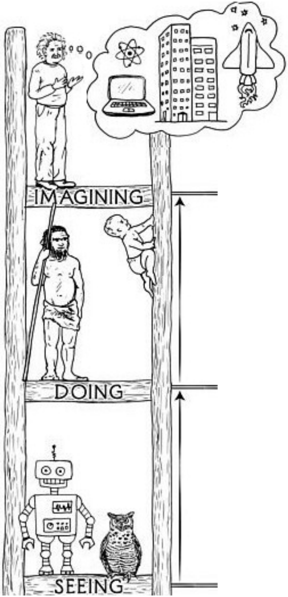
Some epidemiological questions

- Do people who exercise more tend to have lower rates of heart disease?
- Can we predict who is at higher risk of developing diabetes using BIG DATA?
- Does exercising reduce cholesterol?
- What happens to blood pressure if we prescribe a new hypertension drug?
- Would COVID-19 outcomes in France have changed with a different policy?

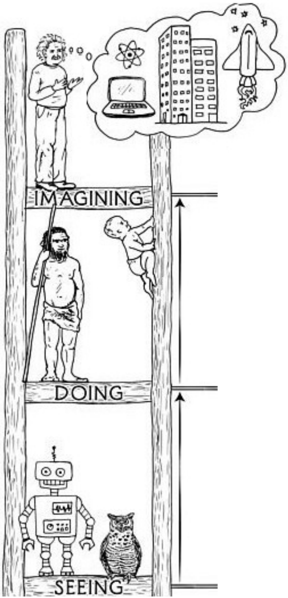
Some epidemiological questions

- Do people who exercise more tend to have lower rates of heart disease?
- Can we predict who is at higher risk of developing diabetes using BIG DATA?
- Does exercising reduce cholesterol?
- What happens to blood pressure if we prescribe a new hypertension drug?
- Would COVID-19 outcomes in France have changed with a different policy?
- Would air pollution affect mortality if it didn't affect respiratory diseases?

Pearl causal hierarchy

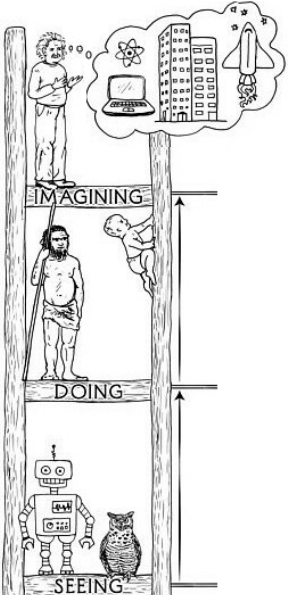


Pearl causal hierarchy



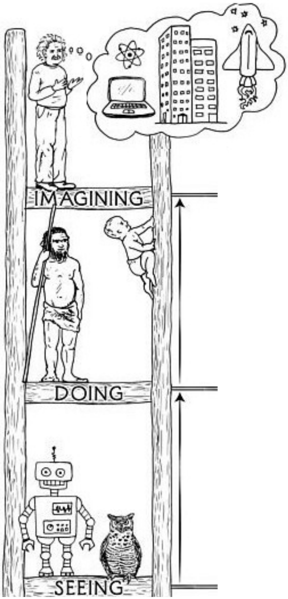
- Associations
  - Questions : What if I see ...?
  - Do people who exercise more tend to have lower rates of heart disease?
  - Can we predict who is at higher risk of developing diabetes using BIG DATA?

Pearl causal hierarchy



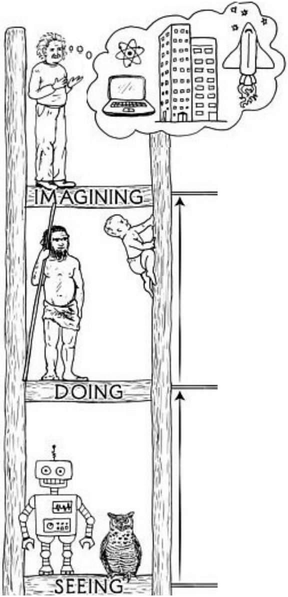
- Interventions
  - Questions: What if I do ...? How?
  - Does exercising reduce cholesterol?
  - What happens to blood pressure if we prescribe a new hypertension drug?
- Associations
  - Questions : What if I see ...?
  - Do people who exercise more tend to have lower rates of heart disease?
  - Can we predict who is at higher risk of developing diabetes using BIG DATA?

Pearl causal hierarchy



- Counterfactuals
  - Questions: What if I had done ...? Why?
  - Would COVID-19 outcomes in France have changed with a different policy?
  - Would air pollution affect mortality if it didn't affect respiratory diseases?
- Interventions
  - Questions: What if I do ...? How?
  - Does exercising reduce cholesterol?
  - What happens to blood pressure if we prescribe a new hypertension drug?
- Associations
  - Questions : What if I see ...?
  - Do people who exercise more tend to have lower rates of heart disease?
  - Can we predict who is at higher risk of developing diabetes using BIG DATA?

Pearl causal hierarchy



- Interventions
  - Questions: What if I do ...? How?
  - Does exercising reduce cholesterol?
  - What happens to blood pressure if we prescribe a new hypertension drug?

Causal effect

The causal effect of A on Y

$$= \mathbb{E}(Y \mid do(A = a)) - \mathbb{E}(Y \mid do(A = a'))$$

The operator $do()$ represents interventions.

Causal effect

$$\text{The causal effect of } A \text{ on } Y$$
$$= \mathbb{E}(Y \mid do(A = a)) - \mathbb{E}(Y \mid do(A = a'))$$

The operator $do()$ represents interventions.

Population

Causal effect

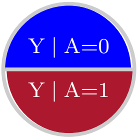
The causal effect of A on Y

$$= \mathbb{E}(Y \mid do(A = a)) - \mathbb{E}(Y \mid do(A = a'))$$

The operator $do()$ represents interventions.

Population

Sub-populations



Causal effect

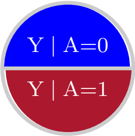
The causal effect of A on Y

$$= \mathbb{E}(Y \mid do(A = a)) - \mathbb{E}(Y \mid do(A = a'))$$

The operator $do()$ represents interventions.

Population

Sub-populations

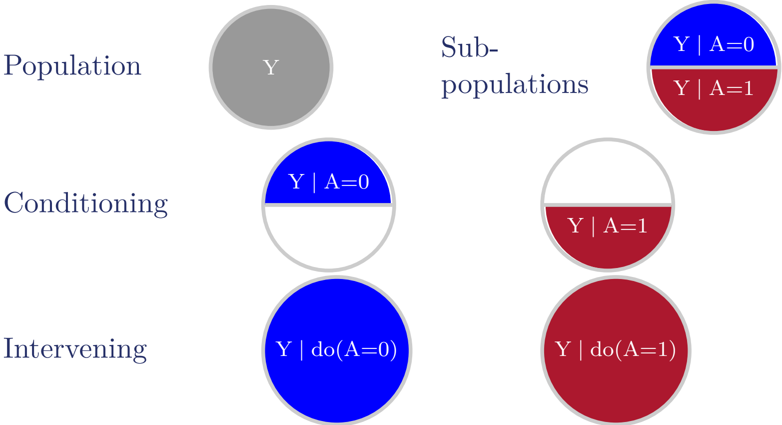


Conditioning

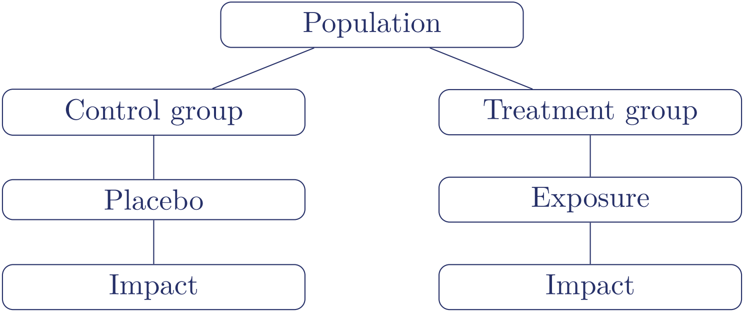
The causal effect of A on Y

$$= \mathbb{E}(Y \mid do(A = a)) - \mathbb{E}(Y \mid do(A = a'))$$

The operator $do()$ represents interventions.



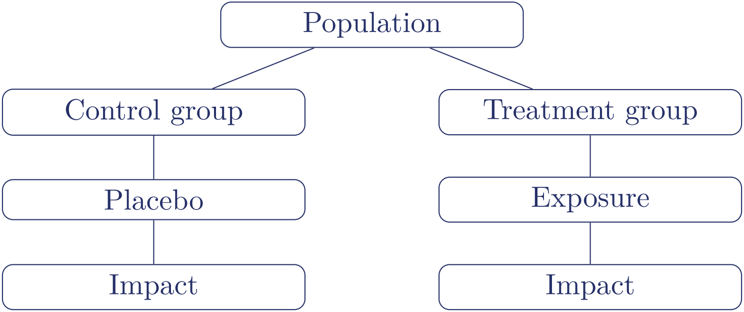
Gold standard: Randomized controlled trials



Control:

- Eliminate temporal bias

Gold standard: Randomized controlled trials



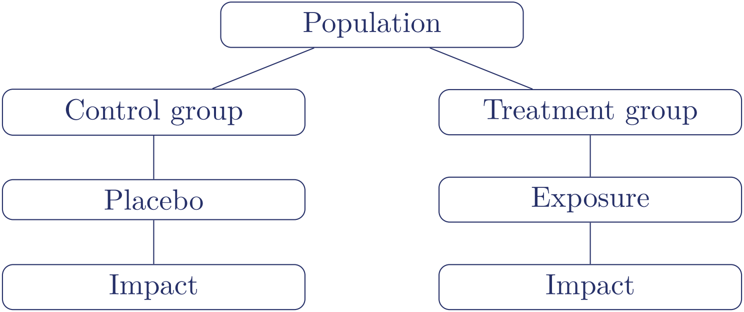
Control:

- Eliminate temporal bias

Random selection:

- Eliminate confounding bias

Gold standard: Randomized controlled trials



Control:

- Eliminate temporal bias

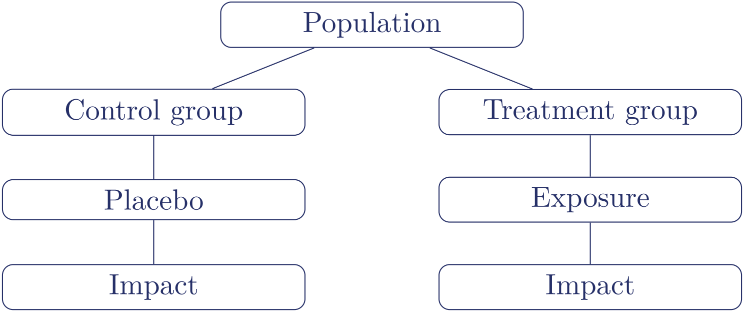
Random selection:

- Eliminate confounding bias

Limitations:

- Costly
- Unethical
- infeasible

Gold standard: Randomized controlled trials



Control:

- Eliminate temporal bias

Random selection:

- Eliminate confounding bias

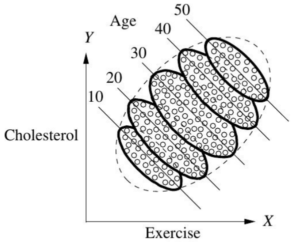
Limitations:

- Costly
- Unethical
- infeasible

$\Longrightarrow$ Sometimes we have to rely on observational studies

Simpson paradox [3]

In a study, we measure weekly exercise and cholesterol levels for various age groups.



SANTÉ SORBONNE UNIVERSITÉ

2

Bayesian networks

Notations

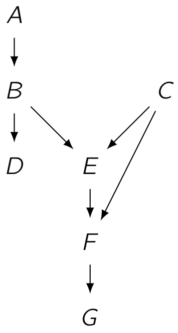
- Sets: $\mathbb{A} = \{X, Y, Z\}$
- Statistical independence: $\perp\!\!\!\perp_P$
- Statistical dependence: $\not\perp\!\!\!\perp_P$
- $P(Y = y \mid X = x) \equiv P(y \mid x)$

Directed graphs

A graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ is said to be a directed graph iff

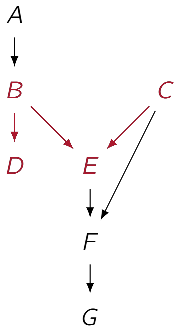
- $\mathbb{V}$ is the set of vertices (usually each vertex corresponds to a random variable),
- $\mathbb{E}$ is the set of edges,
- $\forall (X, Y) \in \mathbb{E},$ there is an arrow pointing from X to Y.

Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



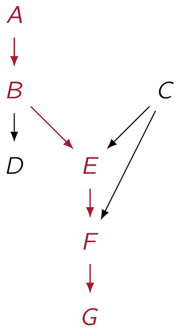
Directed graphs: basic concepts

Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



Path: $D \leftarrow B \rightarrow E \leftarrow C$

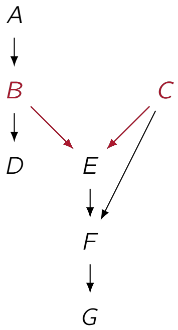
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$

Directed graphs: basic concepts

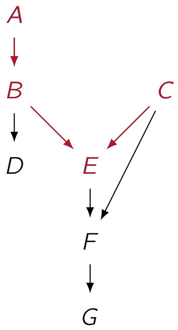
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$

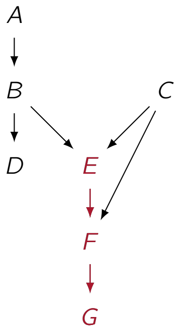
Directed graphs: basic concepts

Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



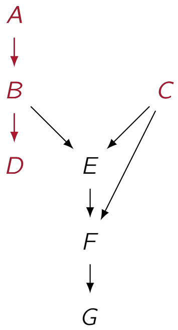
Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancestors: $An(E) = \{A, B, C, E\}$

Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



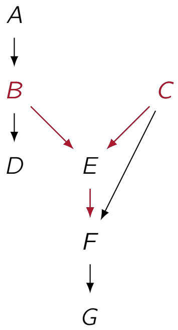
Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancestors: $An(E) = \{A, B, C, E\}$
Descendants: $De(E) = \{E, F, G\}$

Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancestors: $An(E) = \{A, B, C, E\}$
Descendants: $De(E) = \{E, F, G\}$
Non-descendants: $Nd(E) = \{A, B, C, D\}$

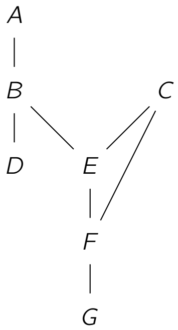
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancestors: $An(E) = \{A, B, C, E\}$
Descendants: $De(E) = \{E, F, G\}$
Non-descendants: $Nd(E) = \{A, B, C, D\}$
Neighbors: $Ne(E) = \{B, C, E\}$

Directed graphs: basic concepts

Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancestors: $An(E) = \{A, B, C, E\}$
Descendants: $De(E) = \{E, F, G\}$
Non-descendants: $Nd(E) = \{A, B, C, D\}$
Neighbors: $Ne(E) = \{B, C, E\}$
Skeleton of $\mathcal{G}$

A directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ is said to be a directed acyclic graph (DAG) iff

$$\forall X \in \mathbb{V}, An(X) \cap De(X) = \{X\},$$

i.e., there are no cycle in $\mathcal{G}$.

DAGs

A directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ is said to be a directed acyclic graph (DAG) iff

$$\forall X \in \mathbb{V}, An(X) \cap De(X) = \{X\},$$

i.e., there are no cycle in $\mathcal{G}$.

From now on, we will only deal with DAGs.

Python code: graphs using networkx package

```
#Load the networkx package
import networkx as nx
#Define the Directed graph
g = nx.DiGraph()
g.add_nodes_from(["A","B","C"])
g.add_edge("B", "A")
g.add_edge("B", "C")
g.add_edge("A", "C")
g.remove_edge("A", "C")
#Print nodes and edges
print(g.nodes)
print(g.edges)
#Print parents and children
print(g.predecessors("A"))
print(g.successors("B"))
```

Definition

A distribution $P(\mathbb{V})$ is compatible with a DAG $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ if

$$P(\mathbb{V}) = \prod_{X \in \mathbb{V}} P(X \mid Pa(X))$$

Definition

A distribution $P(\mathbb{V})$ is compatible with a DAG $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ if

$$P(\mathbb{V}) = \prod_{X \in \mathbb{V}} P(X \mid Pa(X))$$

$\Leftrightarrow \forall X \in \mathbb{V},\ X \perp\!\!\!\perp_P Nd(X) \mid Pa(X)$ (Markov condition)

Definition

A distribution $P(\mathbb{V})$ is compatible with a DAG $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ if

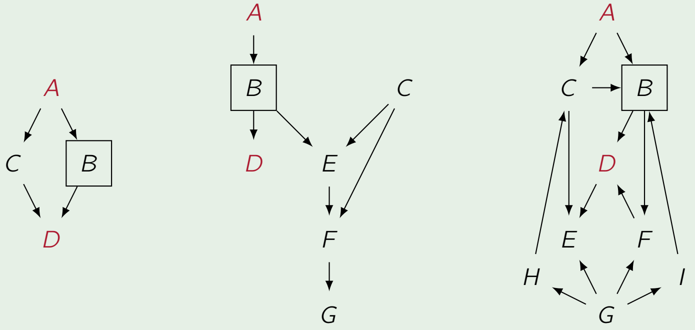
$$P(\mathbb{V}) = \prod_{X \in \mathbb{V}} P(X \mid Pa(X))$$

$\Leftrightarrow \forall X \in \mathbb{V}, X \perp\!\!\!\perp_P Nd(X) \mid Pa(X)$ (Markov condition)

A DAG $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ is a Bayesian network iff there exists a joint distribution $P(\mathbb{V})$ that is compatible with $\mathcal{G}$.

Example

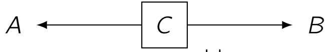
$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



Basic structures

Fork: contains a confounder

$A \overset{?}{\perp\!\!\!\perp}_P B$

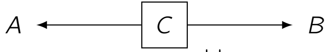


$A \perp\!\!\!\perp_P B \mid C$

Basic structures

Fork: contains a confounder

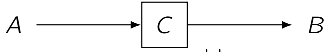
$A \overset{?}{\perp\!\!\!\perp}_P B$



$A \perp\!\!\!\perp_P B \mid C$

Chain: contains a mediator or an intermediate cause

$A \overset{?}{\perp\!\!\!\perp}_P B$



$A \perp\!\!\!\perp_P B \mid C$

SANTÉ SORBONNE UNIVERSITÉ

Basic structures

Fork: contains a confounder

$$A \longleftarrow C \longrightarrow B \qquad\qquad A \longleftarrow \boxed{C} \longrightarrow B$$

$$A \overset{?}{\perp\!\!\!\perp}_P B \qquad\qquad A \perp\!\!\!\perp_P B \mid C$$

Chain: contains a mediator or an intermediate cause

$$A \longrightarrow C \longrightarrow B \qquad\qquad A \longrightarrow \boxed{C} \longrightarrow B$$

$$A \overset{?}{\perp\!\!\!\perp}_P B \qquad\qquad A \perp\!\!\!\perp_P B \mid C$$

Collider : contains a common effect

$$A \longrightarrow C \longleftarrow B \qquad\qquad A \longrightarrow \boxed{C} \longleftarrow B$$

$$A \perp\!\!\!\perp_P B \qquad\qquad A \overset{?}{\perp\!\!\!\perp}_P B \mid C$$

Basic structures

Fork: contains a confounder

$A \leftarrow C \rightarrow B$

$A \overset{?}{\perp\!\!\!\perp}_P B$

$A \leftarrow \boxed{C} \rightarrow B$

$A \perp\!\!\!\perp_P B \mid C$

Chain: contains a mediator or an intermediate cause

$A \rightarrow C \rightarrow B$

$A \overset{?}{\perp\!\!\!\perp}_P B$

$A \rightarrow \boxed{C} \rightarrow B$

$A \perp\!\!\!\perp_P B \mid C$

Collider : contains a common effect

$A \rightarrow C \leftarrow B$

$A \perp\!\!\!\perp_P B$

$A \rightarrow C \leftarrow B$, $C \rightarrow \boxed{D}$

$A \overset{?}{\perp\!\!\!\perp}_P B \mid D$

Basic structures

Fork: contains a confounder

$A \leftarrow C \rightarrow B$ $\qquad$ $A \overset{?}{\perp\!\!\!\perp}_P B$

$A \leftarrow \boxed{C} \rightarrow B$ $\qquad$ $A \perp\!\!\!\perp_P B \mid C$

Chain: contains a mediator or an intermediate cause

$A \rightarrow C \rightarrow B$ $\qquad$ $A \overset{?}{\perp\!\!\!\perp}_P B$

$A \rightarrow \boxed{C} \rightarrow B$ $\qquad$ $A \perp\!\!\!\perp_P B \mid C$

Collider : contains a common effect

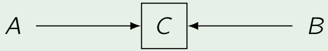
$A \rightarrow C \leftarrow B$ $\qquad$ $A \perp\!\!\!\perp_P B$

$A \rightarrow C \leftarrow B$, $C \rightarrow \boxed{D}$ $\qquad$ $A \overset{?}{\perp\!\!\!\perp}_P B \mid D$

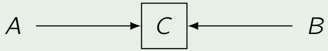
A collider is unshielded if its extremities are not adjacent.

Artificial correlation in Colliders: example 1

Example



Example

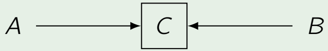


$$A = \begin{cases} \text{Mother carrier} \\ \text{Mother not carrier} \end{cases} \qquad B = \begin{cases} \text{Father carrier} \\ \text{Father not carrier} \end{cases}$$

$$C = (A \text{ or } B) = \begin{cases} \text{Child carrier} \\ \text{Child not carrier} \end{cases}$$

Artificial correlation in Colliders: example 1

Example



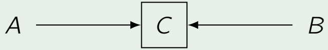
$$A = \begin{cases} \text{Mother carrier} \\ \text{Mother not carrier} \end{cases} \qquad B = \begin{cases} \text{Father carrier} \\ \text{Father not carrier} \end{cases}$$

$$C = (A \text{ or } B) = \begin{cases} \text{Child carrier} \\ \text{Child not carrier} \end{cases}$$

If $C = \text{Child carrier} \implies$

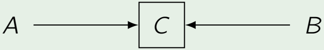
$$\begin{cases} \text{If } A = \text{Mother not carrier then } B = \text{Father carrier} \\ \text{If } B = \text{Father not carrier then } A = \text{Mother carrier} \end{cases}$$

Example



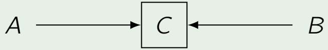
Artificial correlation in Colliders: example 2

Example

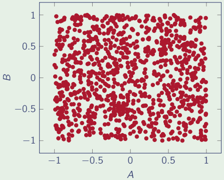


$A, B \sim U(-1, 1)$ $\xi_c \sim N(0, \frac{1}{2})$ $C = 2AB + \xi_c$

Example



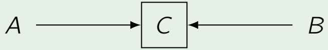
$A, B \sim U(-1, 1)$ $\xi_c \sim N(0, \frac{1}{2})$ $C = 2AB + \xi_c$



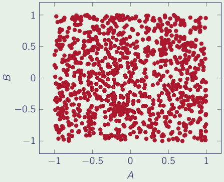
$Corr(A; B) = 0.002$

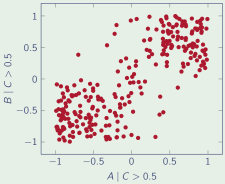
Artificial correlation in Colliders: example 2

Example



$A, B \sim U(-1, 1)$ $\xi_c \sim N(0, \frac{1}{2})$ $C = 2AB + \xi_c$





$Corr(A; B) = 0.002$ $Corr(A; B \mid C > 0.5) = 0.8$

Blocked paths

A path is said to be blocked by a set of vertices $\mathbb{Z} \in \mathbb{V}$ if:

- it contains a chain $A \to B \to C$ or a fork $A \leftarrow B \to C$ and $B \in \mathbb{Z}$; or
- it contains a collider $A \to B \leftarrow C$ such that no descendant of B is in $\mathbb{Z}$.

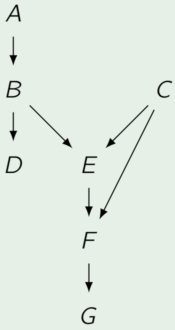
Blocked paths

A path is said to be blocked by a set of vertices $\mathbb{Z} \in \mathbb{V}$ if:

- it contains a chain $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ and $B \in \mathbb{Z}$; or
- it contains a collider $A \rightarrow B \leftarrow C$ such that no descendant of B is in $\mathbb{Z}$.

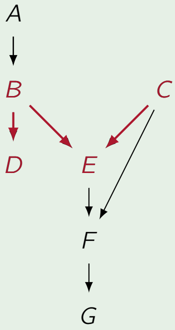
A path that is not blocked is active.

Example



Blocked path: examples

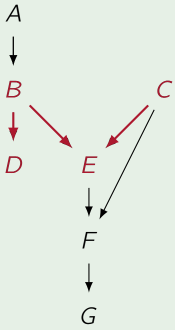
Example



Is the path $< D, B, E, C >$ blocked?

Blocked path: examples

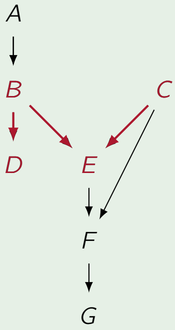
Example



Is the path $< D, B, E, C >$ blocked? Yes

Blocked path: examples

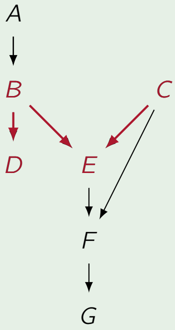
Example



Is the path $< D, B, E, C >$ blocked? Yes

Is the path $< D, B, E, C >$ blocked by E?

Example

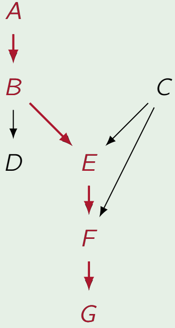


Is the path $< D, B, E, C >$ blocked? Yes

Is the path $< D, B, E, C >$ blocked by E? No

Blocked path: examples

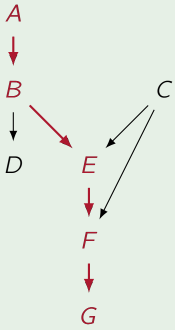
Example



Is the path $< D, B, E, C >$ blocked? Yes
Is the path $< D, B, E, C >$ blocked by E? No
Is the path $< A, B, E, F, G >$ blocked?

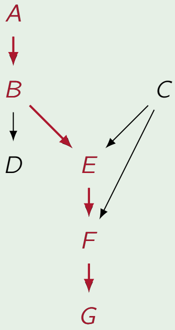
Blocked path: examples

Example



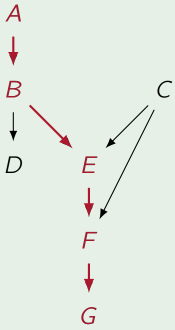
Is the path $< D, B, E, C >$ blocked? Yes
Is the path $< D, B, E, C >$ blocked by E? No
Is the path $< A, B, E, F, G >$ blocked? No

Example



Is the path $< D, B, E, C >$ blocked? Yes
Is the path $< D, B, E, C >$ blocked by E? No
Is the path $< A, B, E, F, G >$ blocked? No
Is the path $< A, B, E, F, G >$ blocked by E?

Example



Is the path $< D, B, E, C >$ blocked? Yes

Is the path $< D, B, E, C >$ blocked by E? No

Is the path $< A, B, E, F, G >$ blocked? No

Is the path $< A, B, E, F, G >$ blocked by E? Yes

d-separation

Given disjoint sets $\mathbb{X}, \mathbb{Y}, \mathbb{Z} \subseteq \mathbb{V}$, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-separated by $\mathbb{Z}$ if every path between a vertex in $\mathbb{X}$ and a vertex in $\mathbb{Y}$ is blocked by $\mathbb{Z}$ and we write $\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

d-separation

Given disjoint sets $\mathbb{X}, \mathbb{Y}, \mathbb{Z} \subseteq \mathbb{V}$, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-separated by $\mathbb{Z}$ if every path between a vertex in $\mathbb{X}$ and a vertex in $\mathbb{Y}$ is blocked by $\mathbb{Z}$ and we write $\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

If one path is not blocked, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-connected given $\mathbb{Z}$ and we write $\mathbb{X} \not\!\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

Given disjoint sets $\mathbb{X}, \mathbb{Y}, \mathbb{Z} \subseteq \mathbb{V}$, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-separated by $\mathbb{Z}$ if every path between a vertex in $\mathbb{X}$ and a vertex in $\mathbb{Y}$ is blocked by $\mathbb{Z}$ and we write $\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

If one path is not blocked, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-connected given $\mathbb{Z}$ and we write $\mathbb{X} \not\!\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

Theorem

$\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \Rightarrow \forall P$ compatible with $\mathcal{G}, \mathbb{X} \perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$

d-separation

Given disjoint sets $\mathbb{X}, \mathbb{Y}, \mathbb{Z} \subseteq \mathbb{V}$, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-separated by $\mathbb{Z}$ if every path between a vertex in $\mathbb{X}$ and a vertex in $\mathbb{Y}$ is blocked by $\mathbb{Z}$ and we write $\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

If one path is not blocked, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-connected given $\mathbb{Z}$ and we write $\mathbb{X} \not\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

Theorem

$\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \Rightarrow \forall P$ compatible with $\mathcal{G}, \mathbb{X} \perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$
but $\mathbb{X} \not\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \not\Rightarrow \forall P$ compatible with $\mathcal{G}, \mathbb{X} \not\perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$

Given disjoint sets $\mathbb{X}, \mathbb{Y}, \mathbb{Z} \subseteq \mathbb{V}$, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-separated by $\mathbb{Z}$ if every path between a vertex in $\mathbb{X}$ and a vertex in $\mathbb{Y}$ is blocked by $\mathbb{Z}$ and we write $\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

If one path is not blocked, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-connected given $\mathbb{Z}$ and we write $\mathbb{X} \not\!\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

Theorem

$\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \Rightarrow \forall P$ compatible with $\mathcal{G}, \mathbb{X} \perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$
but $\mathbb{X} \not\!\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \not\Rightarrow \forall P$ compatible with $\mathcal{G}, \mathbb{X} \not\!\perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$ instead
$\mathbb{X} \not\!\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \Rightarrow \exists P$ compatible with $\mathcal{G}, \mathbb{X} \not\!\perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$

d-separation

Given disjoint sets $\mathbb{X}, \mathbb{Y}, \mathbb{Z} \subseteq \mathbb{V}$, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-separated by $\mathbb{Z}$ if every path between a vertex in $\mathbb{X}$ and a vertex in $\mathbb{Y}$ is blocked by $\mathbb{Z}$ and we write $\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

If one path is not blocked, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-connected given $\mathbb{Z}$ and we write $\mathbb{X} \not\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

Theorem

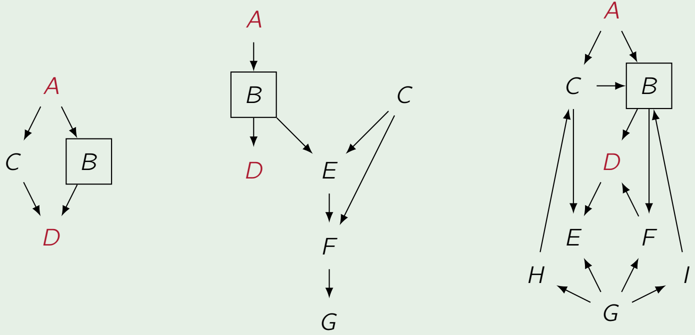
$\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \Rightarrow \forall P$ compatible with $\mathcal{G}, \mathbb{X} \perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$
but $\mathbb{X} \not\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \not\Rightarrow \forall P$ compatible with $\mathcal{G}, \mathbb{X} \not\perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$ instead
$\mathbb{X} \not\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \Rightarrow \exists P$ compatible with $\mathcal{G}, \mathbb{X} \not\perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$

$$\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \Rightarrow \mathbb{X} \perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$$

$$\mathbb{X} \not\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \Rightarrow \mathbb{X} \overset{?}{\perp\!\!\!\perp}_{P} \mathbb{Y} \mid \mathbb{Z}$$

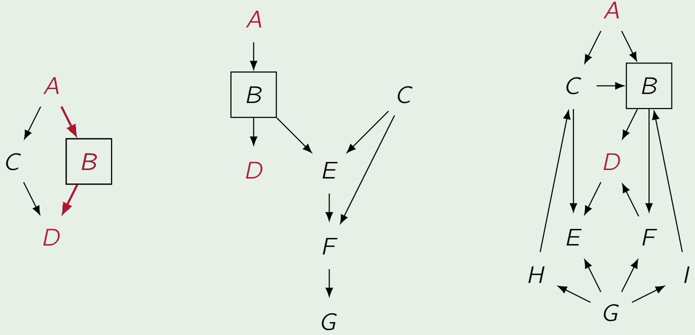
Example

Is $A \perp\!\!\!\perp_P D \mid B$?



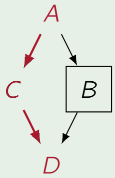
Example

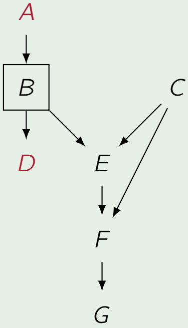
Is $A \perp\!\!\!\perp_P D \mid B$?

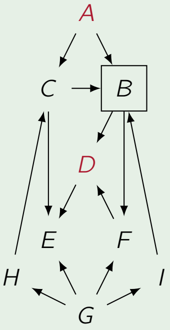


Example

Is $A \perp\!\!\!\perp_P D \mid B$?

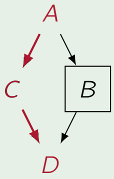






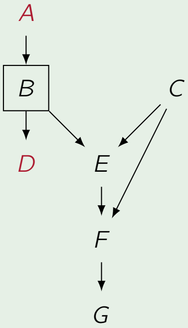
Example

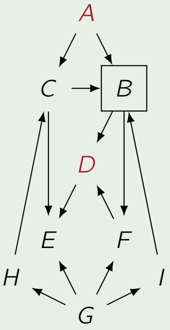
Is $A \perp\!\!\!\perp_P D \mid B$?



$< A, C, D >$ is not blocked

$\overset{?}{\Longrightarrow} A \perp\!\!\!\perp_P D \mid B$

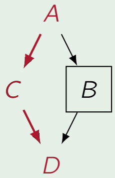




Reading conditional independencies using d-separation

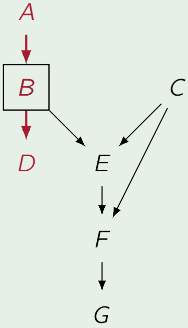
Example

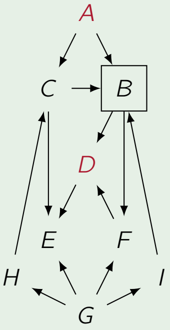
Is $A \perp\!\!\!\perp_P D \mid B$?



$< A, C, D >$ is not blocked

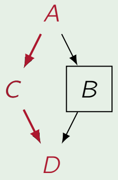
$\overset{?}{\implies} A \perp\!\!\!\perp_P D \mid B$





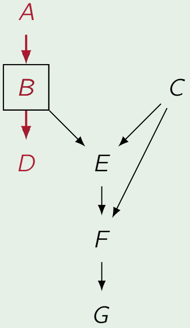
Example

Is $A \perp\!\!\!\perp_P D \mid B$?



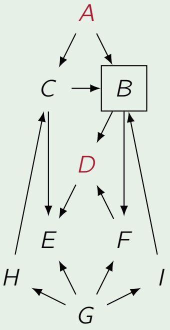
$< A, C, D >$ is not blocked

$\implies A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



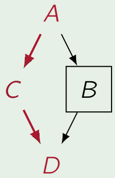
All paths are blocked

$\implies A \perp\!\!\!\perp_P D \mid B$



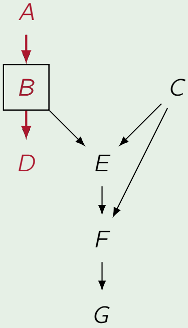
Example

Is $A \perp\!\!\!\perp_P D \mid B$?



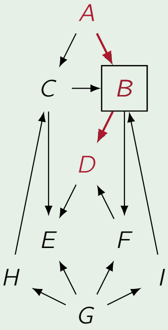
$< A, C, D >$ is not blocked

$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



All paths are blocked

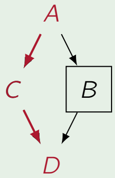
$\Longrightarrow A \perp\!\!\!\perp_P D \mid B$



Reading conditional independencies using d-separation

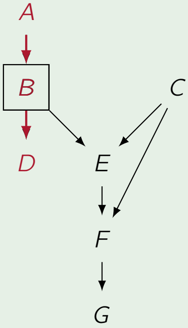
Example

Is $A \perp\!\!\!\perp_P D \mid B$?



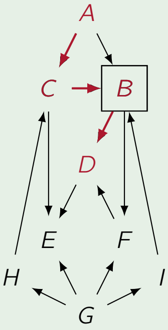
$< A, C, D >$ is not blocked

$\implies A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



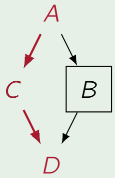
All paths are blocked

$\implies A \perp\!\!\!\perp_P D \mid B$



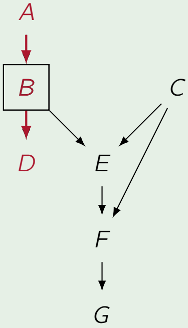
Example

Is $A \perp\!\!\!\perp_P D \mid B$?



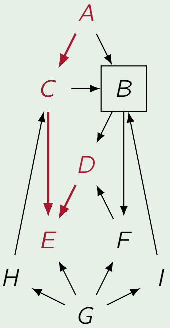
$< A, C, D >$ is not blocked

$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



All paths are blocked

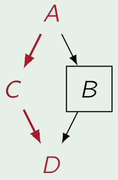
$\Longrightarrow A \perp\!\!\!\perp_P D \mid B$



Reading conditional independencies using d-separation

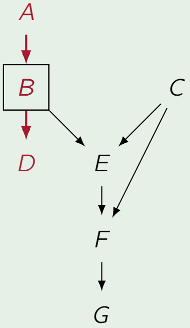
Example

Is $A \perp\!\!\!\perp_P D \mid B$?



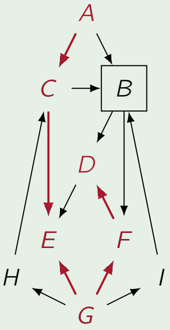
$< A, C, D >$ is not blocked

$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



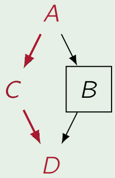
All paths are blocked

$\Longrightarrow A \perp\!\!\!\perp_P D \mid B$



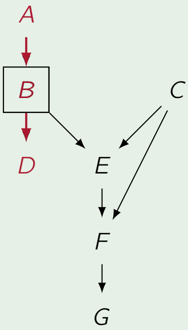
Example

Is $A \perp\!\!\!\perp_P D \mid B$?



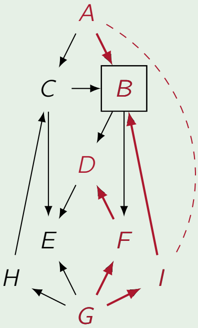
$< A, C, D >$ is not blocked

$\implies A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



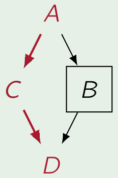
All paths are blocked

$\implies A \perp\!\!\!\perp_P D \mid B$



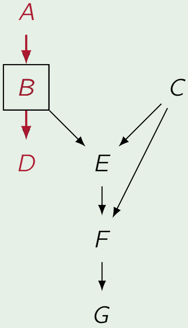
Example

Is $A \perp\!\!\!\perp_P D \mid B$?



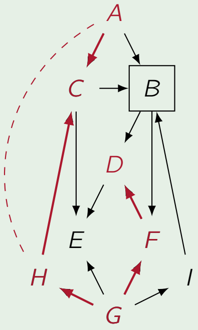
$< A, C, D >$ is not blocked

$\implies A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



All paths are blocked

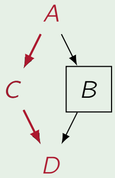
$\implies A \perp\!\!\!\perp_P D \mid B$



Reading conditional independencies using d-separation

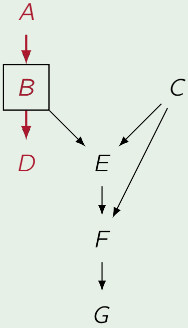
Example

Is $A \perp\!\!\!\perp_P D \mid B$?



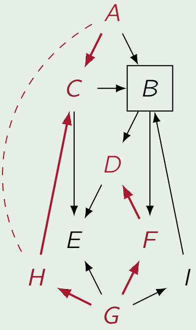
$< A, C, D >$ is not blocked

$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



All paths are blocked

$\Longrightarrow A \perp\!\!\!\perp_P D \mid B$



$< A, B, I, G, F, D >$ is not blocked

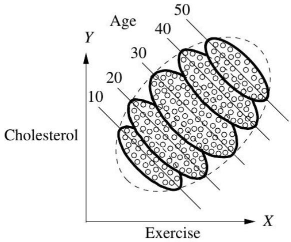
$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_P D \mid B$

Python code: d-separation using networkx package

```
#Load the networkx package
import networkx as nx
#Define the Directed graph
g = nx.DiGraph()
g.add_nodes_from(["A","B","C"])
g.add_edge("B", "A")
g.add_edge("B", "C")^^I^^I
#Check if B d-separates A and C
dsep = nx.is_d_separator(g, "A", "B", "C")
print(dsep)
```

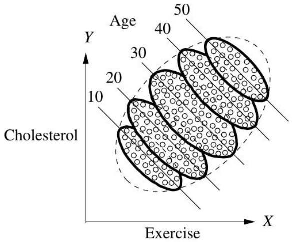
Simpson paradox [3]

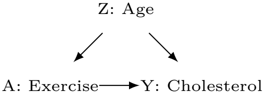
In a study, we measure weekly exercise and cholesterol levels for various age groups.



Simpson paradox [3]

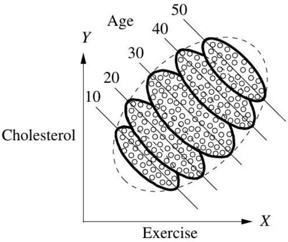
In a study, we measure weekly exercise and cholesterol levels for various age groups.

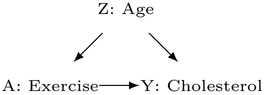


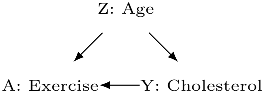


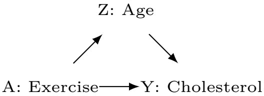
Simpson paradox [3]

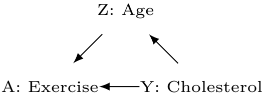
In a study, we measure weekly exercise and cholesterol levels for various age groups.









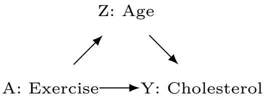


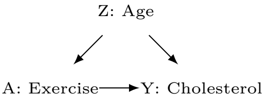
SANTÉ SORBONNE UNIVERSITÉ

3

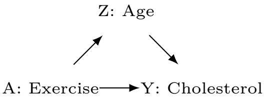
Causal DAGs

Bayesian networks vs causal DAGs

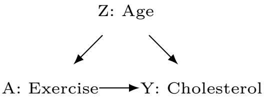




Bayesian networks vs causal DAGs

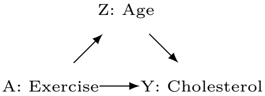


Not a causal DAG

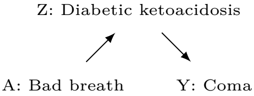


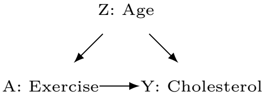
Causal DAG

Bayesian networks vs causal DAGs

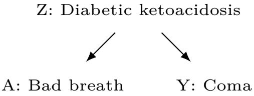


Not a causal DAG

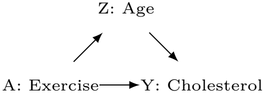




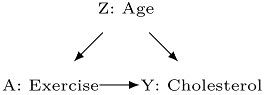
Causal DAG



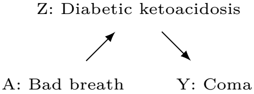
Bayesian networks vs causal DAGs



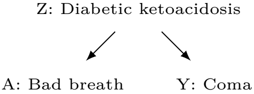
Not a causal DAG



Causal DAG

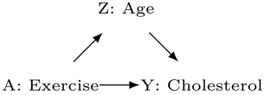


Not a causal DAG

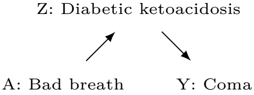


Causal DAG

Bayesian networks vs causal DAGs

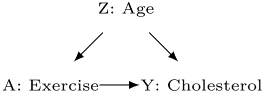


Not a causal DAG

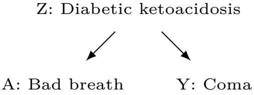


Not a causal DAG

Oracle for conditional independence



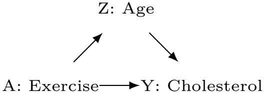
Causal DAG



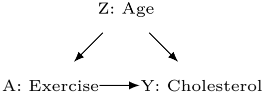
Causal DAG

Oracle for intervention

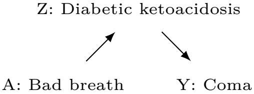
Bayesian networks vs causal DAGs



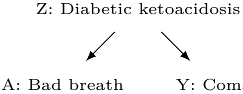
Not a causal DAG



Causal DAG



Not a causal DAG



Causal DAG

Oracle for conditional independence

Oracle for intervention

It is impossible to determine, without additional assumptions, which of the two DAGs is causal based solely on the observed distribution!

Truncated factorization [2]: If we intervene on a subset $\mathbb{S} \subset \mathbb{V}$, then

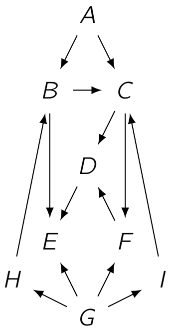
$$P(v_1, \cdots, v_d \mid do(s)) = \prod_{V_i \notin \mathbb{S}} P(v_i \mid Pa(v_i))$$

Truncated factorization [2]: If we intervene on a subset $\mathbb{S} \subset \mathbb{V}$, then

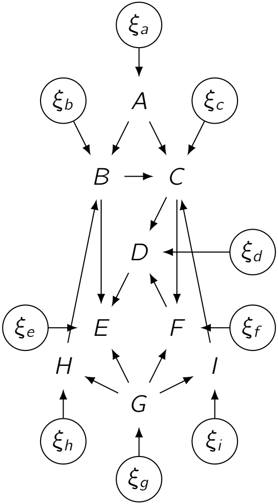
$$P(v_1, \cdots, v_d \mid do(s)) = \prod_{V_i \notin \mathbb{S}} P(v_i \mid Pa(v_i))$$

Let $P(\mathbb{V})$ be a probability distribution and let P_* denote the set of all interventional distributions $P(\mathbb{V} \mid do(s))$. A bayesian network $\mathcal{G}$ is said to be a causal DAG compatible with P_* iff $\mathcal{G}$ and P_* satisfy the truncated factorization.

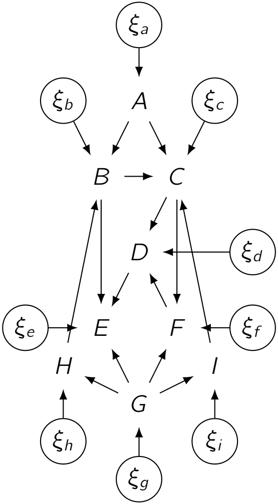
Structural causal model



Structural causal model

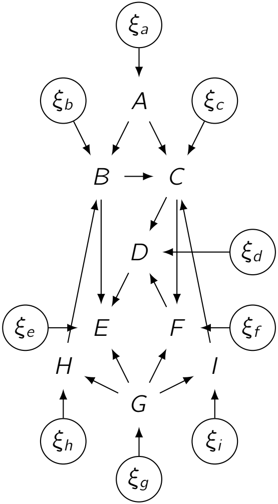


Structural causal model



$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

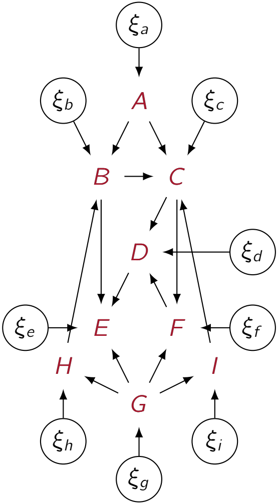
Structural causal model



$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

A structural causal model (SCM) is a tuple that contains:

Structural causal model

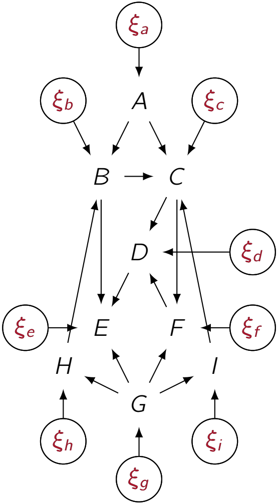


$$
M: \begin{cases}
A := f_a(\xi_a) \\
B := f_b(A, H, \xi_b) \\
C := f_c(A, B, I, \xi_c) \\
D := f_d(C, F, \xi_d) \\
E := f_e(B, G, \xi_e) \\
F := f_f(C, G, \xi_f) \\
G := f_g(\xi_g) \\
H := f_h(G, \xi_h) \\
I := f_i(G, \xi_i)
\end{cases}
$$

A structural causal model (SCM) is a tuple that contains:

- Endogenous variables

Structural causal model

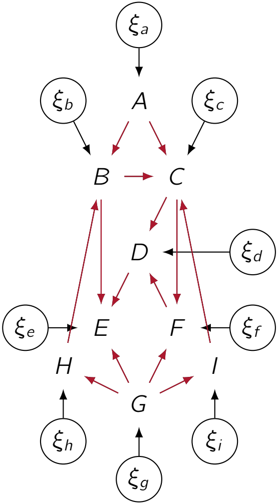


$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

A structural causal model (SCM) is a tuple that contains:

- Endogenous variables
- Exogenous variables

Structural causal model



$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

A structural causal model (SCM) is a tuple that contains:

- Endogenous variables
- Exogenous variables
- Causal mechanisms for generating endogenous variables

Guidlines for constructing causal DAGs

In causal DAGs, the information lies in the absence of edges and the direction of arrows, rather than merely in the presence of edges.

Guidlines for constructing causal DAGs

In causal DAGs, the information lies in the absence of edges and the direction of arrows, rather than merely in the presence of edges.

- Think carrefully about the orientation

Guidlines for constructing causal DAGs

In causal DAGs, the information lies in the absence of edges and the direction of arrows, rather than merely in the presence of edges.

- Think carrefully about the orientation
- When you are not sure if you need to add or not an edge (for example $Z \to A$) to the graph, ADD IT! (as long as you keep the graph acyclic)

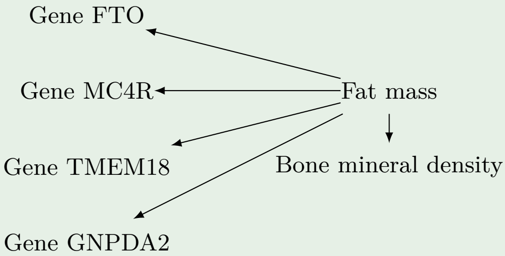
Guidlines for constructing causal DAGs

In causal DAGs, the information lies in the absence of edges and the direction of arrows, rather than merely in the presence of edges.

- Think carrefully about the orientation
- When you are not sure if you need to add or not an edge (for example $Z \rightarrow A$) to the graph, ADD IT! (as long as you keep the graph acyclic)

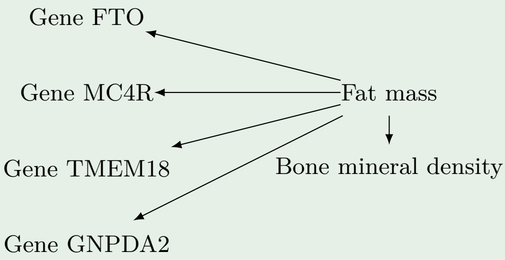
If you cannot keep the graph acyclic, do not worry, there exists new tools for cyclic graphs.
But in this lecture, we will focus only on acyclic graphs.

Example



Is this DAG causal?

Example

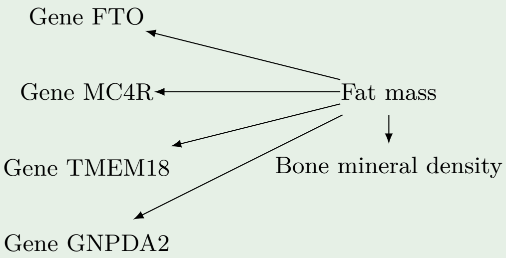


Is this DAG causal?

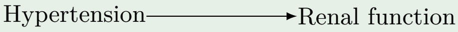
No

DAG Vs DAG causal: exemples

Example

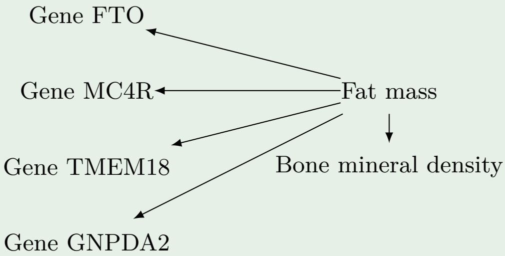


Is this DAG causal?
No

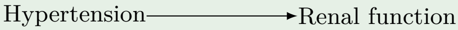


Is this DAG causal?

Example



Is this DAG causal?
No



Is this DAG causal?
Yes

SANTÉ SORBONNE UNIVERSITÉ

4

Causal reasoning

Identifiability and positivity

A causal effect is said to be identifiable if it is uniquely computable from an observational distribution $P(\mathbb{V})$.

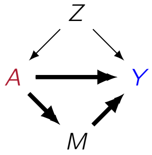
The causal effect of A on Y
$= \mathbb{E}(\mathbb{Y} \mid do(A = a)) - \mathbb{E}(Y \mid do(A = a'))$

Identifiability and positivity

A causal effect is said to be identifiable if it is uniquely computable from an observational distribution $P(\mathbb{V})$.

The causal effect of A on Y

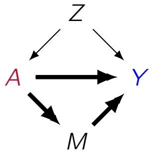
$= \mathbb{E}(\mathbb{Y} \mid do(A = a)) - \mathbb{E}(Y \mid do(A = a'))$



Identifiability and positivity

A causal effect is said to be identifiable if it is uniquely computable from an observational distribution $P(\mathbb{V})$.

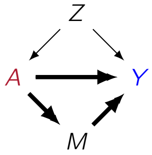
The causal effect of A on Y
$= \mathbb{E}(\mathbb{Y} \mid do(A = a)) - \mathbb{E}(Y \mid do(A = a'))$



The causal effect is identifiable iff $P(y \mid do(a))$ is identifiable.

A causal effect is said to be identifiable if it is uniquely computable from an observational distribution $P(\mathbb{V})$.

The causal effect of A on Y
$$= \mathbb{E}(\mathbb{Y} \mid do(A = a)) - \mathbb{E}(Y \mid do(A = a'))$$

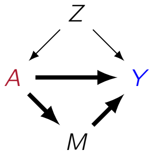


The causal effect is identifiable iff $P(y \mid do(a))$ is identifiable.

Causal reasoning involves utilizing a causal DAG to determine whether $P(y \mid do(a))$ is identifiable.

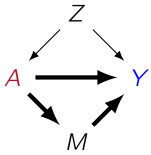
Key challenges of not using a causal DAG (1/3)

Confounding bias (e.g., Simpson paradox):



Key challenges of not using a causal DAG (1/3)

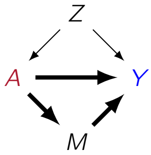
Confounding bias (e.g., Simpson paradox):



We must eliminate all confounding bias by adjusting for confounders.

Key challenges of not using a causal DAG (1/3)

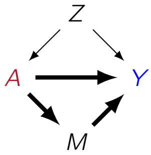
Confounding bias (e.g., Simpson paradox):



We must eliminate all confounding bias by adjusting for confounders.
Should we always adjust for all available variables?

Key challenges of not using a causal DAG (1/3)

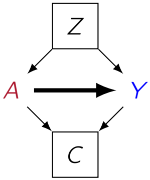
Confounding bias (e.g., Simpson paradox):



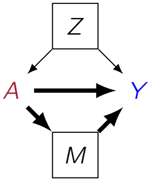
We must eliminate all confounding bias by adjusting for confounders.
Should we always adjust for all available variables? No!

Key challenges of not using a causal DAG (2/3)

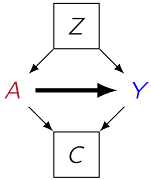
Bias due to adjusting for colliders:



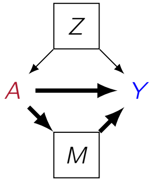
Bias due to incorrect adjustment for mediators:



Bias due to adjusting for colliders:



Bias due to incorrect adjustment for mediators:



Should we adjust on everything that is temporally prior to the exposure?

Key challenges of not using a causal DAG (3/3)

Let $\mathbb{Z}$ be the set of observed variables in a problem that are not affected by A. The set $\mathbb{Z}$ satisfies Association Criterion if each element $Z \in \mathbb{Z}$ meets the following conditions:

- Z is associated with A; and
- Z is associated with Y, conditionally on A.

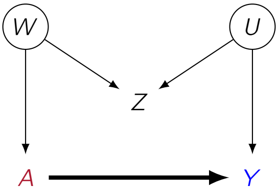
.

Key challenges of not using a causal DAG (3/3)

Let $\mathbb{Z}$ be the set of observed variables in a problem that are not affected by A. The set $\mathbb{Z}$ satisfies Association Criterion if each element $Z \in \mathbb{Z}$ meets the following conditions:

- Z is associated with A; and
- Z is associated with Y, conditionally on A.

Counterexample



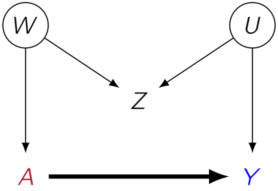
.

Key challenges of not using a causal DAG (3/3)

Let $\mathbb{Z}$ be the set of observed variables in a problem that are not affected by A. The set $\mathbb{Z}$ satisfies Association Criterion if each element $Z \in \mathbb{Z}$ meets the following conditions:

- Z is associated with A; and
- Z is associated with Y, conditionally on A.

Counterexample

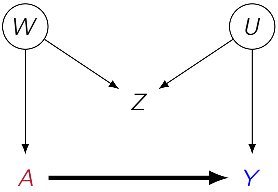


Z is associated with A and Z is associated with Y, conditionally on A.

Let $\mathbb{Z}$ be the set of observed variables in a problem that are not affected by A. The set $\mathbb{Z}$ satisfies Association Criterion if each element $Z \in \mathbb{Z}$ meets the following conditions:

- Z is associated with A; and
- Z is associated with Y, conditionally on A.

Counterexample



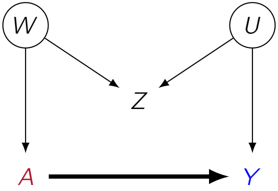
Z is associated with A and Z is associated with Y, conditionally on A. Should we adjust on Z?

Key challenges of not using a causal DAG (3/3)

Let $\mathbb{Z}$ be the set of observed variables in a problem that are not affected by A. The set $\mathbb{Z}$ satisfies Association Criterion if each element $Z \in \mathbb{Z}$ meets the following conditions:

- Z is associated with A; and
- Z is associated with Y, conditionally on A.

Counterexample

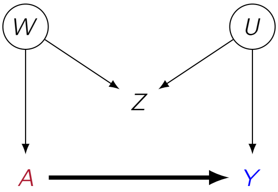


Z is associated with A and Z is associated with Y, conditionally on A. Should we adjust on Z? No!

Let $\mathbb{Z}$ be the set of observed variables in a problem that are not affected by A. The set $\mathbb{Z}$ satisfies Association Criterion if each element $Z \in \mathbb{Z}$ meets the following conditions:

- Z is associated with A; and
- Z is associated with Y, conditionally on A.

Counterexample

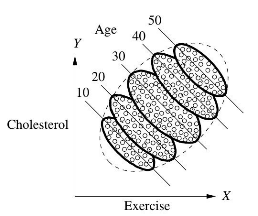


Z is associated with A and Z is associated with Y, conditionally on A. Should we adjust on Z? No!
This criterion is incorrect!

Simpson paradox and a simple solution

Example

In a study, we measure weekly exercise and cholesterol levels for various age groups.

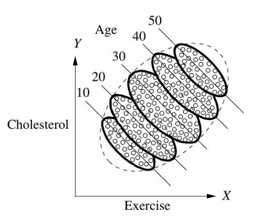


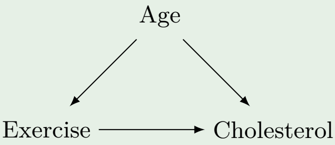
What is the effect of exercise on cholesterol $P(c \mid do(e))$?

Simpson paradox and a simple solution

Example

In a study, we measure weekly exercise and cholesterol levels for various age groups.



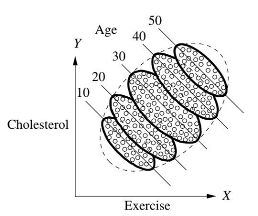


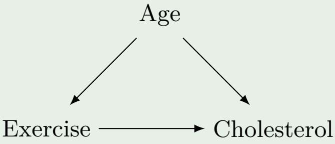
What is the effect of exercise on cholesterol $P(c \mid do(e))$?

Simpson paradox and a simple solution

Example

In a study, we measure weekly exercise and cholesterol levels for various age groups.





What is the effect of exercise on cholesterol $P(c \mid do(e))$?

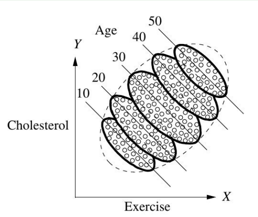
$$P(a, e, c) = P(a)P(e \mid a)P(c \mid a, e) \qquad \text{(Compatibility)}$$

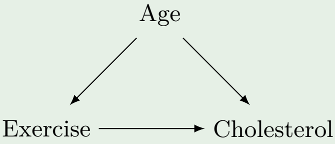
$$P(a, c \mid do(e)) = P(a)P(c \mid a, e) \qquad \text{(Truncated factorization)}$$

Simpson paradox and a simple solution

Example

In a study, we measure weekly exercise and cholesterol levels for various age groups.





What is the effect of exercise on cholesterol $P(c \mid do(e))$?

$$P(a, e, c) = P(a)P(e \mid a)P(c \mid a, e) \quad \text{(Compatibility)}$$

$$P(a, c \mid do(e)) = P(a)P(c \mid a, e) \quad \text{(Truncated factorization)}$$

$$P(c \mid do(e)) = \sum_a P(a)P(c \mid a, e) \quad \text{(Marginalizing)}$$

A simple solution: direct causes adjustment [2]

Theorem

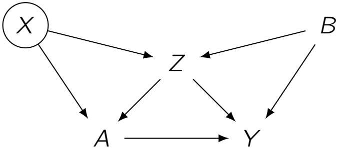
Given a causal DAG $\mathcal{G}$ in which a subset $\mathbb{V}$ of variables are measured, the causal effect $P(y \mid do(a))$ is identifiable whenever $\{A \cup Y \cup Parents(A)\} \subseteq \mathbb{V}$, and is given by:

$$P(y \mid do(a)) = \sum_{Z \in Pa(A)} P(y \mid a, z) P(z)$$

Limitations of the direct causes adjustment

- Sometimes the set of parents is too large. Is it possible to find a smaller set?

- Sometimes the set of observed parents is not sufficient for adjustment. Is it possible to find another set?



The back-door criterion

A set of variables $\mathbb{Z}$ satisfies the back-door criterion relative to an ordered pair of variables (A, Y) in causal DAG $\mathcal{G}$ if:

- No node in $\mathbb{Z}$ is a descendant of A; and
- $\mathbb{Z}$ blocks all paths between A and Y that contain an arrow pointing toward A.

A set of variables $\mathbb{Z}$ satisfies the back-door criterion relative to an ordered pair of variables (A, Y) in causal DAG $\mathcal{G}$ if:

- No node in $\mathbb{Z}$ is a descendant of A; and
- $\mathbb{Z}$ blocks all paths between A and Y that contain an arrow pointing toward A.

Theorem

If $\mathbb{Z}$ satisfies the back-door criterion with respect to (A, Y) and if $P(a, z) > 0$, then $P(y \mid do(a))$ is identifiable and is given by:

$$P(y \mid do(a)) = \sum_{z} P(y \mid a, z) P(z).$$

The back-door criterion

A set of variables $\mathbb{Z}$ satisfies the back-door criterion relative to an ordered pair of variables (A, Y) in causal DAG $\mathcal{G}$ if:

- No node in $\mathbb{Z}$ is a descendant of A; and
- $\mathbb{Z}$ blocks all paths between A and Y that contain an arrow pointing toward A.

Theorem

If $\mathbb{Z}$ satisfies the back-door criterion with respect to (A, Y) and if $P(a, z) > 0$, then $P(y \mid do(a))$ is identifiable and is given by:

$$P(y \mid do(a)) = \sum_z P(y \mid a, z) P(z).$$

This criterion can be extended to a set $\mathbb{A}$ and a set $\mathbb{Y}$.

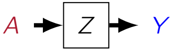
The back-door criterion: intuition

- Why ”no node in $\mathbb{Z}$ is a descendant of A”?

- Why ”$\mathbb{Z}$ blocks all paths between A and Y that contain an arrow pointing toward A”?

The back-door criterion: intuition

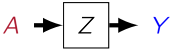
- Why "no node in $\mathbb{Z}$ is a descendant of A"?



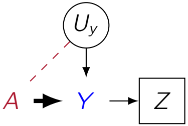
To avoid blocking on intermediate causes.

- Why "$\mathbb{Z}$ blocks all paths between A and Y that contain an arrow pointing toward A"?

- Why "no node in $\mathbb{Z}$ is a descendant of A"?



To avoid blocking on intermediate causes.

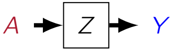


To avoid introducing artificial confounding bias resulting from conditioning on a collider.

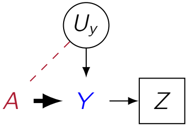
- Why "$\mathbb{Z}$ blocks all paths between A and Y that contain an arrow pointing toward A"?

The back-door criterion: intuition

- Why "no node in $\mathbb{Z}$ is a descendant of A"?



To avoid blocking on intermediate causes.

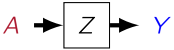


To avoid introducing artificial confounding bias resulting from conditioning on a collider.
To avoid selection bias!

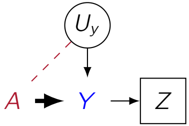
- Why "$\mathbb{Z}$ blocks all paths between A and Y that contain an arrow pointing toward A"?

The back-door criterion: intuition

- Why "no node in $\mathbb{Z}$ is a descendant of A"?



To avoid blocking on intermediate causes.



To avoid introducing artificial confounding bias resulting from conditioning on a collider.
To avoid selection bias!

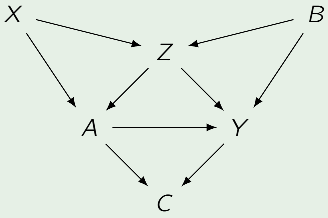
- Why "$\mathbb{Z}$ blocks all paths between A and Y that contain an arrow pointing toward A"?

To eliminate confounding bias.

Back-door criterion in action (1/2)

Example

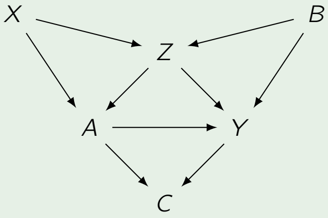
$P(y \mid do(a))$?



Back-door criterion in action (1/2)

Example

$P(y \mid do(a))$?



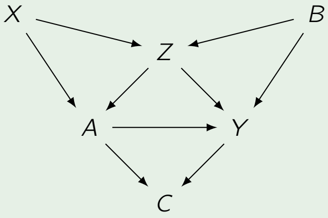
Z ?

Back-door criterion in action (1/2)

Example

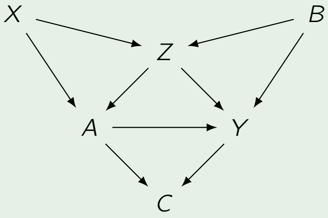
$P(y \mid do(a))$?

Z ? No



Example

$P(y \mid do(a))$?



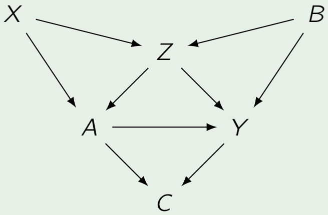
Z ? No

X ?

Back-door criterion in action (1/2)

Example

$P(y \mid do(a))$?



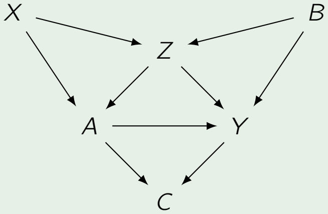
Z ? No

X ? No

Back-door criterion in action (1/2)

Example

$P(y \mid do(a))$?



Z ? No

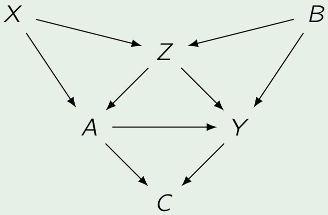
X ? No

B ?

Back-door criterion in action (1/2)

Example

$P(y \mid do(a))$?



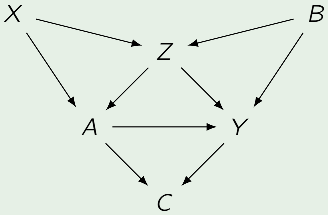
Z ? No

X ? No

B ? No

Example

$P(y \mid do(a))$?



Z ? No

X ? No

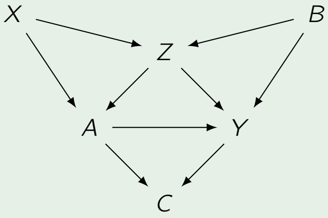
B ? No

C ?

Back-door criterion in action (1/2)

Example

$P(y \mid do(a))$?

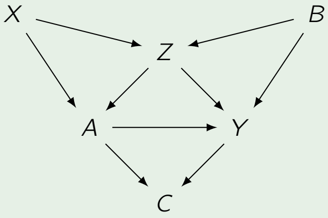


Z ? No
X ? No
B ? No
C ? No

Back-door criterion in action (1/2)

Example

$P(y \mid do(a))$?



Z ? No

X ? No

B ? No

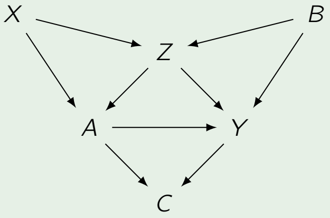
C ? No

X, B?

Back-door criterion in action (1/2)

Example

$P(y \mid do(a))$?



Z ? No

X ? No

B ? No

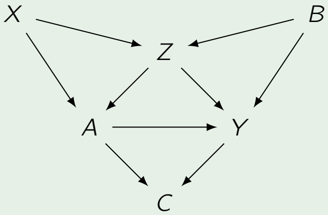
C ? No

X, B? No

Back-door criterion in action (1/2)

Example

$P(y \mid do(a))$?



Z ? No

X ? No

B ? No

C ? No

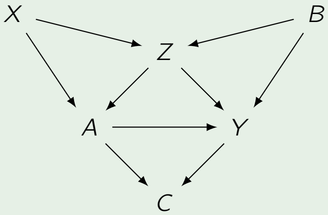
X, B? No

Z, X ?

Back-door criterion in action (1/2)

Example

$P(y \mid do(a))$?



Z ? No

X ? No

B ? No

C ? No

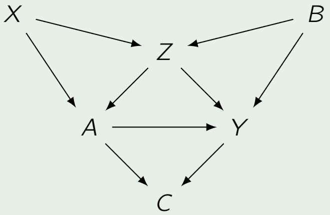
X, B? No

Z, X ? Yes

Back-door criterion in action (1/2)

Example

$P(y \mid do(a))$?



Z ? No

X ? No

B ? No

C ? No

X, B? No

Z, X ? Yes

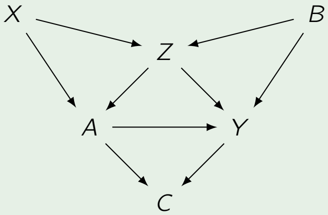
Z, B ?

SANTÉ SORBONNE UNIVERSITÉ

Back-door criterion in action (1/2)

Example

$P(y \mid do(a))$?



Z ? No

X ? No

B ? No

C ? No

X, B? No

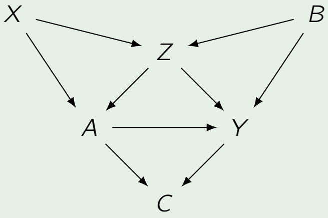
Z, X ? Yes

Z, B ? Yes

Back-door criterion in action (1/2)

Example

$P(y \mid do(a))$?



Z ? No

X ? No

B ? No

C ? No

X, B? No

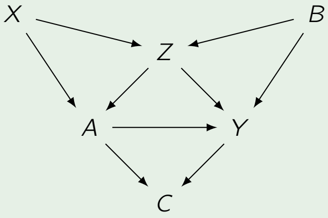
Z, X ? Yes

Z, B ? Yes

Z, X, B ?

Example

$P(y \mid do(a))$?



Z ? No

X ? No

B ? No

C ? No

X, B? No

Z, X ? Yes

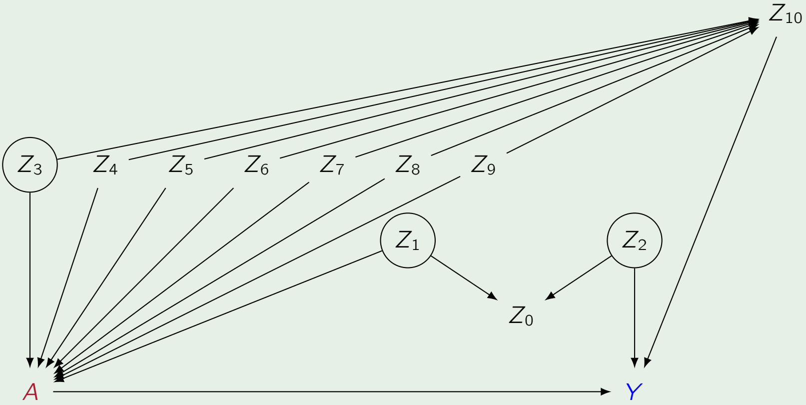
Z, B ? Yes

Z, X, B ? Yes

Back-door criterion in action (2/2)

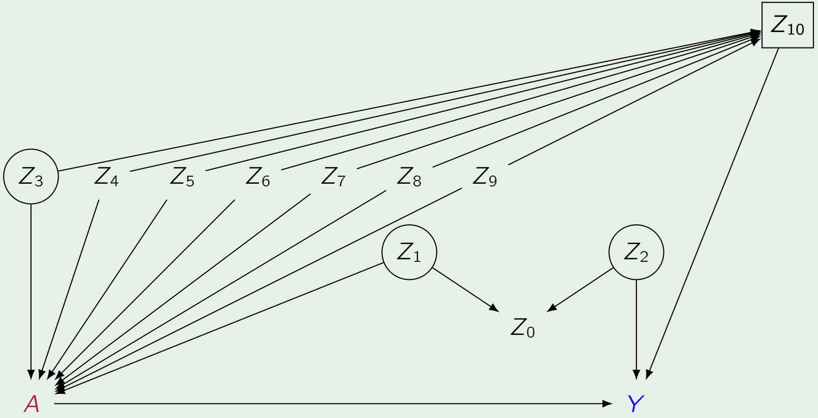
Example

Suppose that all variable are temporally prior to A and Y.
$P(y \mid do(a))$?



Example

Suppose that all variable are temporally prior to A and Y.
$P(y \mid do(a))$?



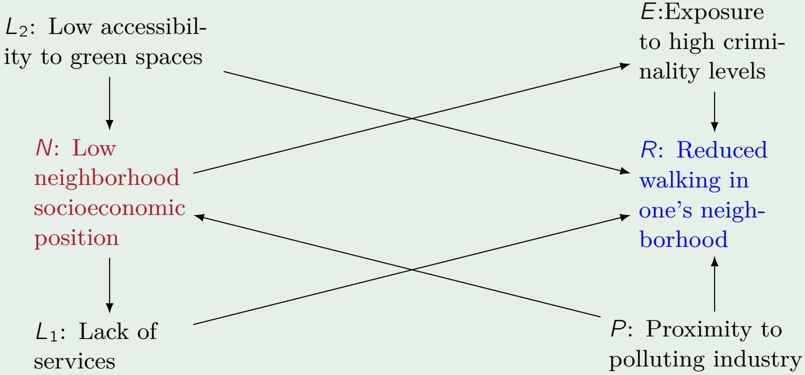
Back-door criterion in a real application

Example

In this study, we aim to estimate the effect of the neighborhood's socioeconomic status (N) on the reduction of walking within the neighborhood (R), $P(r \mid do(n))$?

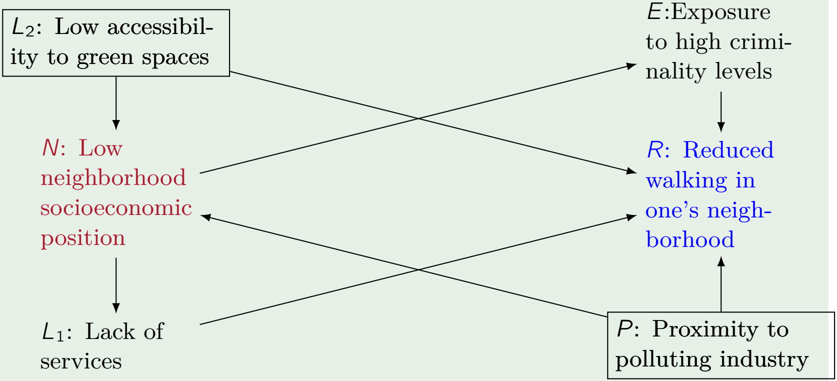
Example

In this study, we aim to estimate the effect of the neighborhood's socioeconomic status (N) on the reduction of walking within the neighborhood (R), $P(r \mid do(n))$?



Example

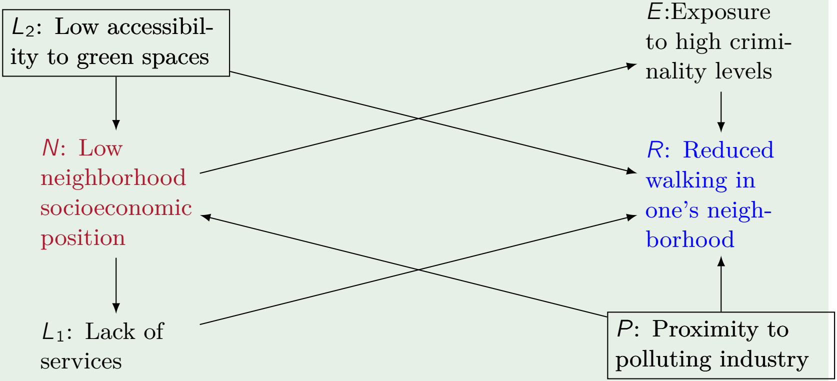
In this study, we aim to estimate the effect of the neighborhood's socioeconomic status (N) on the reduction of walking within the neighborhood (R), $P(r \mid do(n))$?



Back-door criterion in a real application

Example

In this study, we aim to estimate the effect of the neighborhood's socioeconomic status (N) on the reduction of walking within the neighborhood (R), $P(r \mid do(n))$?



$$P(r \mid do(n)) = \sum P(r|n, l_2, p)P(l_2, p)$$

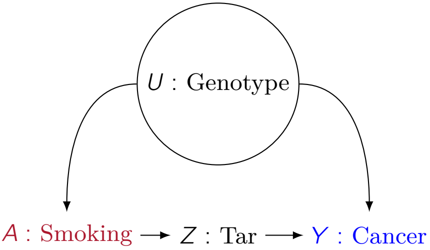
Incompleteness of the back-door criterion

- The back-door criterion is sound but not complete:

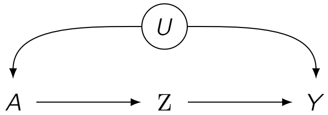
Incompleteness of the back-door criterion

- The back-door criterion is sound but not complete:
  - If there exists a set that satisfy the back-door criterion for $P(y \mid do(a))$, then $P(y \mid do(a))$ is identifiable;
  - If there exists a no set that satisfy the back-door criterion for $P(y \mid do(a))$, then $P(y \mid do(a))$ is not necesarly not identifiable.

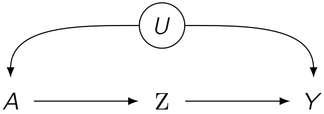
Going beyond the back-door



Going beyond the back-door

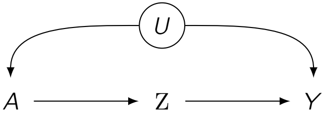


Going beyond the back-door



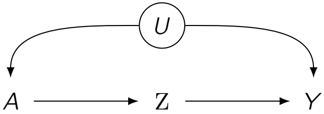
- $P(z \mid do(a)) = P(z \mid a)$ (No back-door path)

Going beyond the back-door



- $P(z \mid do(a)) = P(z \mid a)$ (No back-door path)
- $P(y \mid do(z)) = \sum_a P(y \mid z, a) P(a)$ (A blocks the back-door)

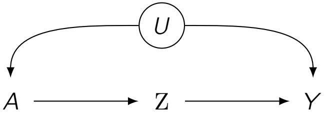
Going beyond the back-door



- $P(z \mid do(a)) = P(z \mid a)$ (No back-door path)
- $P(y \mid do(z)) = \sum_a P(y \mid z, a) P(a)$ (A blocks the back-door)

$$P(y \mid do(a)) = \sum_z P(y \mid do(z)) P(z \mid do(a))$$

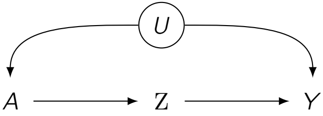
Going beyond the back-door



- $P(z \mid do(a)) = P(z \mid a)$ (No back-door path)
- $P(y \mid do(z)) = \sum_a P(y \mid z, a)P(a)$ (A blocks the back-door)

$$
\begin{aligned}
P(y \mid do(a)) &= \sum_z P(y \mid do(z))P(z \mid do(a)) \\
&= \sum_z P(z \mid a) \sum_{a'} P(y \mid z, a')P(a')
\end{aligned}
$$

Front-door criterion



A set of variables $\mathbb{Z}$ satisfies the front-door criterion relative to an ordered pair of variables (A, Y) in causal DAG $\mathcal{G}$ if:

- $\mathbb{Z}$ intercepts all directed paths from A to Y;
- There is no back-door path from A to $\mathbb{Z}$;
- All back-door paths from $\mathbb{Z}$ to Y are blocked by A.

Theorem ([1])

If $\mathbb{Z}$ satisfies the front-door criterion relative to (A, Y) and if $P(a, z) > 0$, then the causal effect of A on Y is identifiable and is given by

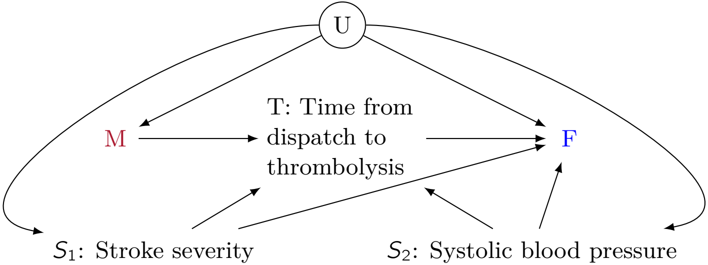
$$P(y \mid do(a)) = \sum_{z} P(z \mid a) \sum_{a'} P(y \mid a', z) P(a').$$

Back-door and Front-door criteria in a real application

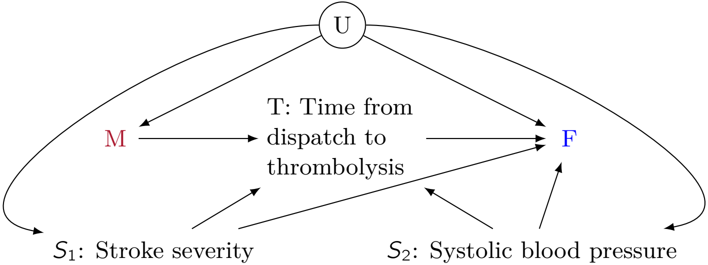
In this study, Piccininni et al. were interested in estimating the effect of Mobile Stroke Unit dispatch (M) on functional outcomes (F). In other words, $P(f \mid do(m))$?

Back-door and Front-door criteria in a real application

In this study, Piccininni et al. were interested in estimating the effect of Mobile Stroke Unit dispatch (M) on functional outcomes (F). In other words, $P(f \mid do(m))$?

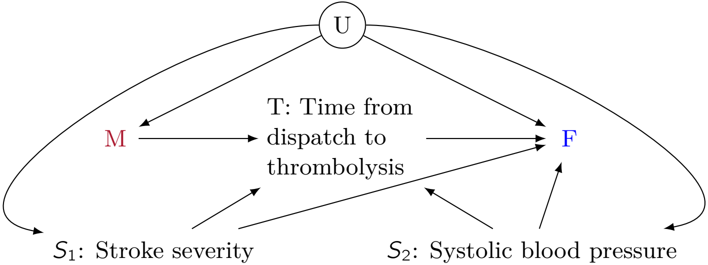


In this study, Piccininni et al. were interested in estimating the effect of Mobile Stroke Unit dispatch (M) on functional outcomes (F). In other words, $P(f \mid do(m))$?



The back-door criterion is not satisfied for any set of variables. However, it is possible to identify $P(f \mid do(m))$ by leveraging a combination of the back-door and front-door criteria!

In this study, Piccininni et al. were interested in estimating the effect of Mobile Stroke Unit dispatch (M) on functional outcomes (F). In other words, $P(f \mid do(m))$?



The back-door criterion is not satisfied for any set of variables. However, it is possible to identify $P(f \mid do(m))$ by leveraging a combination of the back-door and front-door criteria!

Best paper in Epidemiology in 2024!

- The front-door criterion is sound but not complete:

Incompleteness of the front-door criterion

- The front-door criterion is sound but not complete:
  - If there exists a set that satisfy the front-door criterion for $P(y \mid do(x))$, then $P(y \mid do(x))$ is identifiable;
  - If there exists a no set that satisfy the front-door criterion for $P(y \mid do(x))$, then $P(y \mid do(x))$ is not necesarly not identifiable.

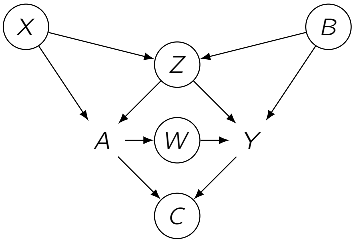
Incompleteness of the front-door criterion

- The front-door criterion is sound but not complete:
  - If there exists a set that satisfy the front-door criterion for $P(y \mid do(x))$, then $P(y \mid do(x))$ is identifiable;
  - If there exists a no set that satisfy the front-door criterion for $P(y \mid do(x))$, then $P(y \mid do(x))$ is not necesarly not identifiable.

The combination of the back-door and front-door criterions is also not complete.

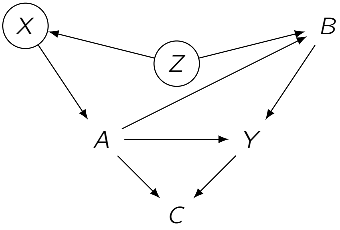
Exercise 1

Consider that in the following causal DAG, only A and Y, and one additional variable can be measured. Which variable would allow the identification of $P(y \mid do(a))$?



Exercise 2

- Consider the following causal DAG. List all sets of variables that satisfy the back-door criterion for $P(y \mid do(a))$;

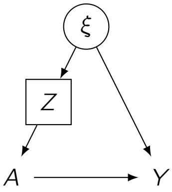


Exercise 3

Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?

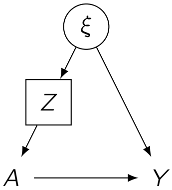
Exercise 3

Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



Exercise 3

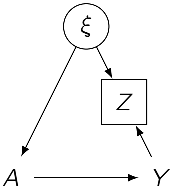
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



- Z blocks all back-door paths and it is not a descendant of A $\implies$ $\{Z\}$ is a good adjustment set.

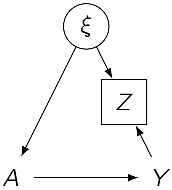
Exercise 3

Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



Exercise 3

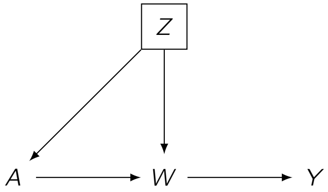
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



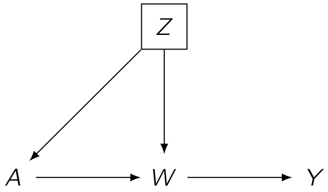
- Z activates a back-door path
  $\implies \{Z\}$ is a bad adjustment set.

Exercise 3

Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



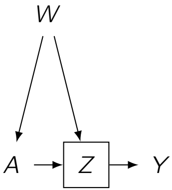
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



- Z blocks all back-door paths and it is not a descendant of A $\implies \{Z\}$ is a good adjustment set.

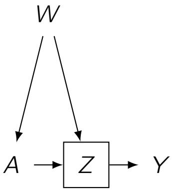
Exercise 3

Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



Exercise 3

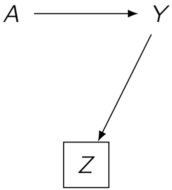
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



- Z d-separates A from Y
  $\implies \{Z\}$ is a bad adjustment set.

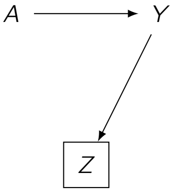
Exercise 3

Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



Exercise 3

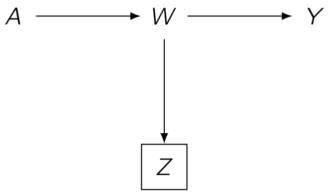
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



- Selection bias
  $\implies$ $\{Z\}$ is a bad control.

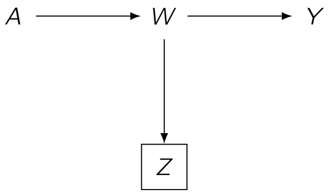
Exercise 3

Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



Exercise 3

Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



- Z is a descendant of A
  $\implies \{Z\}$ is a bad adjustment set.

SANTÉ SORBONNE UNIVERSITÉ

5

Estimation

Linear regression and back-door criterion

Goal: estimate the causal effect

$$\mathrm{Q} := \mathbb{E}[Y \mid do(A=1)] - \mathbb{E}[Y \mid do(A=0)] \quad \text{or} \quad \mathrm{Q} := \frac{d}{da}\mathbb{E}[Y \mid do(A=a)]$$

Assume the back-door criterion holds with covariates $\mathbb{Z}$.

$$\begin{aligned}
\mathbb{E}[Y \mid do(A=a)] &= \int_{\mathbb{z}} \mathbb{E}[Y \mid A=a, \mathbb{Z}=\mathbb{z}] P(\mathbb{z}) d\mathbb{z} \\
&= \int_{\mathbb{z}} (\alpha_1 A + \alpha_2 Z + \alpha_0) P(\mathbb{z}) d\mathbb{z} \qquad \text{(linear model)} \\
&= \alpha_1 A + \alpha_2 \mathbb{E}[Z] + \alpha_0
\end{aligned}$$

Causal interpretation (under correct specification):

$$\widehat{\mathrm{Q}} = \hat{\alpha}_1$$

Linear regression and back-door criterion

Goal: estimate the causal effect

$$\mathrm{Q} := \mathbb{E}[Y \mid do(A=1)] - \mathbb{E}[Y \mid do(A=0)] \quad \text{or} \quad \mathrm{Q} := \frac{d}{da}\mathbb{E}[Y \mid do(A=a)]$$

Assume the back-door criterion holds with covariates $\mathbb{Z}$.

$$\begin{aligned}
\mathbb{E}[Y \mid do(A=a)] &= \int_{\mathbb{z}} \mathbb{E}[Y \mid A=a, \mathbb{Z}=\mathbb{z}] P(\mathbb{z}) d\mathbb{z} \\
&= \int_{\mathbb{z}} (\alpha_1 A + \alpha_2 Z + \alpha_0) P(\mathbb{z}) d\mathbb{z} \qquad \text{(linear model)} \\
&= \alpha_1 A + \alpha_2 \mathbb{E}[Z] + \alpha_0
\end{aligned}$$

Causal interpretation (under correct specification):

$$\widehat{\mathrm{Q}} = \hat{\alpha}_1$$

$\mathbb{E}[Y \mid A=a, \mathbb{Z}=\mathbb{z}]$ can be estimated by OLS.

Linear regression and front-door criterion

Goal: estimate the causal effect

$$\mathrm{Q} := \mathbb{E}[Y \mid do(A=1)] - \mathbb{E}[Y \mid do(A=0)] \quad \text{or} \quad \mathrm{Q} := \frac{d}{da}\mathbb{E}[Y \mid do(A=a)]$$

Assume the front-door criterion holds with variable Z.

Linear model:

$$\mathbb{E}[Z \mid do(A=a)] = \alpha_0 + \alpha_1 a \qquad \text{(can be estimated by OLS)}$$

$$\mathbb{E}[Y \mid do(Z=z)] = \beta_0 + \beta_1 z + \beta_2 \mathbb{E}[A] \qquad \text{(can be estimated by OLS)}$$

$$\begin{aligned} \mathbb{E}[Y \mid do(A=a)] &= \beta_0 + \beta_1\, \mathbb{E}[Z \mid do(A=a)] + \beta_2 \mathbb{E}[A] \\ &= \beta_0 + \beta_1(\alpha_0 + \alpha_1 a) + \beta_2 \mathbb{E}[A] \\ &= \beta_0 + \beta_1 \alpha_0 + \beta_1 \alpha_1 a + \beta_2 \mathbb{E}[A] \end{aligned}$$

Causal interpretation (under correct specification):

$$\widehat{\mathrm{Q}} = \hat{\beta}_1 \times \hat{\alpha}_1$$

Limitations of linear regression

- Model misspecification: incorrect functional form leads to biased causal effects.
- Poor extrapolation: limited overlap in covariates can make linear predictions unreliable.

Machine learning (ML)?

- ML methods can automatically capture nonlinear effects and interactions without explicit specification.
- But naive use of ML breaks standard inference guarantees

Limitations of linear regression

- Model misspecification: incorrect functional form leads to biased causal effects.
- Poor extrapolation: limited overlap in covariates can make linear predictions unreliable.

Machine learning (ML)?

- ML methods can automatically capture nonlinear effects and interactions without explicit specification.
- But naive use of ML breaks standard inference guarantees

ML minimize prediction error: $min_f \mathbb{E}[(Y - f(A, Z))^2] \neq \mathbb{E}[Y \mid do(A = a)]$

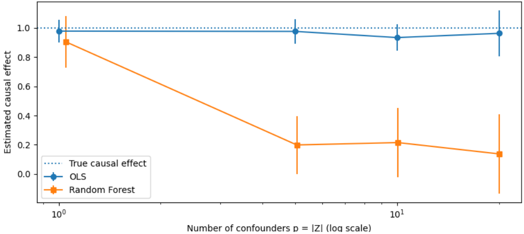
Limitations of linear regression

- Model misspecification: incorrect functional form leads to biased causal effects.
- Poor extrapolation: limited overlap in covariates can make linear predictions unreliable.

Machine learning (ML)?

- ML methods can automatically capture nonlinear effects and interactions without explicit specification.
- But naive use of ML breaks standard inference guarantees

ML minimize prediction error: $min_f \mathbb{E}[(Y - f(A, Z))^2] \neq \mathbb{E}[Y \mid do(A = a)]$



(simulated using a linear model)

Targeted Maximum Likelihood Estimation

TMLE relies on two nuisance models:

$$
\begin{aligned}
O(A, \mathbb{Z}) &= \mathbb{E}[Y \mid A, \mathbb{Z}] \quad \text{(outcome model)}, \\
E(\mathbb{Z}) &= P(A = 1 \mid \mathbb{Z}) \quad \text{(exposure model)}.
\end{aligned}
$$

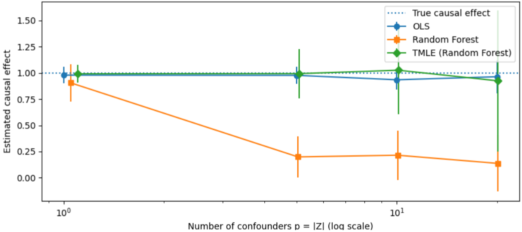
A targeting step updates $\hat{O}$ using information from $\hat{E}$ to focus estimation on the causal effect.
These models can be estimated using ML algorithms.

TMLE relies on two nuisance models:

$$O(A, \mathbb{Z}) = \mathbb{E}[Y \mid A, \mathbb{Z}] \quad \text{(outcome model)},$$
$$E(\mathbb{Z}) = P(A = 1 \mid \mathbb{Z}) \quad \text{(exposure model)}.$$

A targeting step updates $\hat{O}$ using information from $\hat{E}$ to focus estimation on the causal effect.

These models can be estimated using ML algorithms.



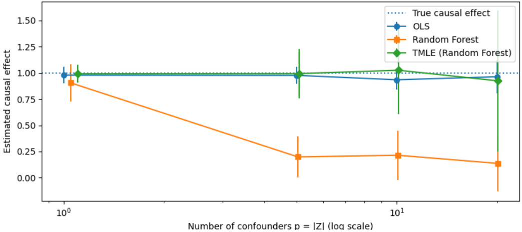
(simulated using a linear model)

Targeted Maximum Likelihood Estimation

TMLE relies on two nuisance models:

$$O(A, \mathbb{Z}) = \mathbb{E}[Y \mid A, \mathbb{Z}] \quad \text{(outcome model)},$$
$$E(\mathbb{Z}) = P(A = 1 \mid \mathbb{Z}) \quad \text{(exposure model)}.$$

A targeting step updates $\hat{O}$ using information from $\hat{E}$ to focus estimation on the causal effect.
These models can be estimated using ML algorithms.



(simulated using a linear model)

TMLE is doubly robust: Consistent $\hat{Q}$ if either $\hat{O}$ or $\hat{E}$ is correctly specified.

Cross-fitting

- Flexible ML models can overfit.
- Overfitting leads to biased estimates and invalid confidence intervals.

Cross-fitting idea:

- Split data into folds.
- Estimate nuisance models on one fold.
- Predict on another fold.
- Combine results across folds.

Take-home: TMLE + ML + cross-fitting = flexible estimation with valid confidence intervals.

Python code: causal estimation using zepid package

```
#Load data
import pandas as pd
data = pd.read_csv("./data.csv")
#Load zepid and sklearn
from zepid.causal.doublyrobust import TMLE
from sklearn.ensemble import RandomForestRegressor as rf
#Estimate causal effect
tmle = TMLE( df=data, exposure='A', outcome='Y' )
tmle.exposure_model('Z')
tmle.outcome_model('A+Z', custom_model=rf())
tmle.fit()
#Show results
tmle.summary()
```

Remark: cross-fitting is not implemented!

SANTÉ SORBONNE UNIVERSITÉ

6

Causal discovery

Causal discovery using conditional independencies

Given observational data, is it possible to infer a causal DAG using conditional independencies?

Causal discovery using conditional independencies

Given observational data, is it possible to infer a causal DAG using conditional independencies? In general no!

Given observational data, is it possible to infer a causal DAG using conditional independencies? In general no!

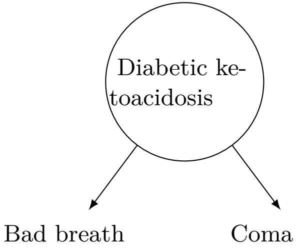
We cannot even construct the skeleton of the graph because

- $\not\perp\!\!\!\perp_P \not\Longrightarrow \not\perp\!\!\!\perp_{\mathcal{G}}$
- $\perp\!\!\!\perp_P \not\Longrightarrow \perp\!\!\!\perp_{\mathcal{G}}$

Given observational data, is it possible to infer a causal DAG using conditional independencies? In general no!

We cannot even construct the skeleton of the graph because

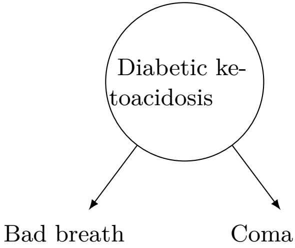
- $\not\perp\!\!\!\perp_P \not\Longrightarrow \not\perp\!\!\!\perp_{\mathcal{G}}$
- $\perp\!\!\!\perp_P \not\Longrightarrow \perp\!\!\!\perp_{\mathcal{G}}$

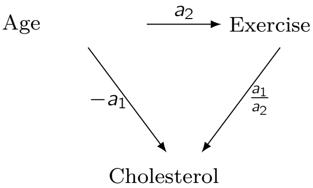


Given observational data, is it possible to infer a causal DAG using conditional independencies? In general no!

We cannot even construct the skeleton of the graph because

- $\not\perp\!\!\!\perp_P \not\Longrightarrow \not\perp\!\!\!\perp_{\mathcal{G}}$
- $\perp\!\!\!\perp_P \not\Longrightarrow \perp\!\!\!\perp_{\mathcal{G}}$





Additional assumptions

Causal sufficiency: No unmeasured confounding.

Additional assumptions

Causal sufficiency: No unmeasured confounding.

Faithfulness: All conditional independence relations true in P are entailed by the d-separation applied to $\mathcal{G}$:

Additional assumptions

Causal sufficiency: No unmeasured confounding.

Faithfulness: All conditional independence relations true in P are entailed by the d-separation applied to $\mathcal{G}$:

$$A \perp\!\!\!\perp_{\mathcal{G}} B \mid \mathrm{S} \iff A \perp\!\!\!\perp_{P} B \mid \mathrm{S}$$

Additional assumptions

Causal sufficiency: No unmeasured confounding.

Faithfulness: All conditional independence relations true in P are entailed by the d-separation applied to $\mathcal{G}$:

$$A \perp\!\!\!\perp_{\mathcal{G}} B \mid \mathrm{S} \iff A \perp\!\!\!\perp_{P} B \mid \mathrm{S}$$

Given observational data, is it possible to infer a causal DAG using conditional independencies under the assumptions of faithfulness and causal sufficiency?

Additional assumptions

Causal sufficiency: No unmeasured confounding.

Faithfulness: All conditional independence relations true in P are entailed by the d-separation applied to $\mathcal{G}$:

$$A \perp\!\!\!\perp_{\mathcal{G}} B \mid \mathrm{S} \iff A \perp\!\!\!\perp_{P} B \mid \mathrm{S}$$

Given observational data, is it possible to infer a causal DAG using conditional independencies under the assumptions of faithfulness and causal sufficiency? In general no!

Markov equivalence class [7]

Equivalence in terms of conditional independence

$X \rightarrow Y$	$X \leftarrow Y$	$X \rightarrow Z \leftarrow Y$
$X \leftarrow Z \rightarrow Y$	$X \leftarrow Z \leftarrow Y$	$X \rightarrow Z \rightarrow Y$

Markov equivalence class [7]

Equivalence in terms of conditional independence

$X \rightarrow Y$	$X \leftarrow Y$	$X \rightarrow Z \leftarrow Y$
$X \leftarrow Z \rightarrow Y$	$X \leftarrow Z \leftarrow Y$	$X \rightarrow Z \rightarrow Y$

Theorem

Two causal DAGs are Markov equivalent iff they have the same skeleton and the same unshielded colliders.

Markov equivalence class [7]

Equivalence in terms of conditional independence

$X \rightarrow Y$	$X \leftarrow Y$	$X \rightarrow Z \leftarrow Y$
$X \leftarrow Z \rightarrow Y$	$X \leftarrow Z \leftarrow Y$	$X \rightarrow Z \rightarrow Y$

Theorem

Two causal DAGs are Markov equivalent iff they have the same skeleton and the same unshielded colliders.

- All equivalent graphs can be represented by a completed partially DAG (CPDAG)
- This CPDAG is called the representative of the Markov equivalence class

Markov equivalence class [7]

Equivalence in terms of conditional independence

$X \rightarrow Y$	$X \leftarrow Y$	$X \rightarrow Z \leftarrow Y$
$X \leftarrow Z \rightarrow Y$	$X \leftarrow Z \leftarrow Y$	$X \rightarrow Z \rightarrow Y$

Theorem

Two causal DAGs are Markov equivalent iff they have the same skeleton and the same unshielded colliders.

- All equivalent graphs can be represented by a completed partially DAG (CPDAG)
- This CPDAG is called the representative of the Markov equivalence class

Given observational data, is it possible to infer a CPDAG using conditional independencies under the assumptions of faithfulness and causal sufficiency?

Markov equivalence class [7]

Equivalence in terms of conditional independence

$X \rightarrow Y$	$X \leftarrow Y$	$X \rightarrow Z \leftarrow Y$
$X \leftarrow Z \rightarrow Y$	$X \leftarrow Z \leftarrow Y$	$X \rightarrow Z \rightarrow Y$

Theorem

Two causal DAGs are Markov equivalent iff they have the same skeleton and the same unshielded colliders.

- All equivalent graphs can be represented by a completed partially DAG (CPDAG)
- This CPDAG is called the representative of the Markov equivalence class

Given observational data, is it possible to infer a CPDAG using conditional independencies under the assumptions of faithfulness and causal sufficiency? Yes!

Finding skeleton and unshielded colliders

Theorem

If $P(\mathbb{V})$ is faithful to some causal DAG $\mathcal{G}$ with vertex $\mathbb{V}$ then:

- For $X, Y \in \mathbb{V}$, X and Y are adjacent iff $\forall \mathbb{S} \subseteq \mathbb{V} \backslash \{X, Y\}$, $X \not\perp\!\!\!\perp_P Y \mid \mathbb{S}$;

Theorem

If $P(\mathbb{V})$ is faithful to some causal DAG $\mathcal{G}$ with vertex $\mathbb{V}$ then:

- For $X, Y \in \mathbb{V}$, X and Y are adjacent iff $\forall \mathbb{S} \subseteq \mathbb{V} \backslash \{X, Y\}$, $X \not\perp\!\!\!\perp_P Y \mid \mathbb{S}$;
- For $X, Y, Z \in \mathbb{V}$ such that X is adjacent to Z and Z is adjacent to Y and X and Y are not adjacent, $X \rightarrow Z \leftarrow Y$ in $\mathcal{G}$ iff $\forall \mathbb{S} \subseteq \mathbb{V} \backslash \{X, Y\}$ such that $Z \in \mathbb{S}$, $X \not\perp\!\!\!\perp_P Y \mid \mathbb{S}$.

Finding skeleton and unshielded colliders

Theorem

If $P(\mathbb{V})$ is faithful to some causal DAG $\mathcal{G}$ with vertex $\mathbb{V}$ then:

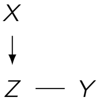
- For $X, Y \in \mathbb{V}$, X and Y are adjacent iff $\forall \mathbb{S} \subseteq \mathbb{V} \backslash \{X, Y\}$, $X \not\perp\!\!\!\perp_P Y \mid \mathbb{S}$;
- For $X, Y, Z \in \mathbb{V}$ such that X is adjacent to Z and Z is adjacent to Y and X and Y are not adjacent, $X \to Z \leftarrow Y$ in $\mathcal{G}$ iff $\forall \mathbb{S} \subseteq \mathbb{V} \backslash \{X, Y\}$ such that $Z \in \mathbb{S}$, $X \not\perp\!\!\!\perp_P Y \mid \mathbb{S}$.

- Point 1 can be used to discover the skeleton of $\mathcal{G}$ from $P(\mathbb{V})$;
- Given the skeleton of $\mathcal{G}$, point 2 can be used to find all unshielded colliders.

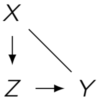
Orientation rules

Suppose we already found the skeleton and all unshielded colliders:

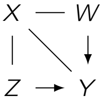
Meek-Rule 1:



Meek-Rule 2:



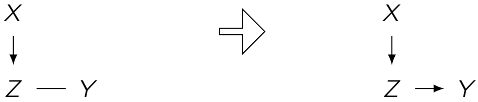
Meek-Rule 3:



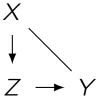
Orientation rules

Suppose we already found the skeleton and all unshielded colliders:

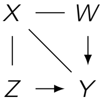
Meek-Rule 1:



Meek-Rule 2:



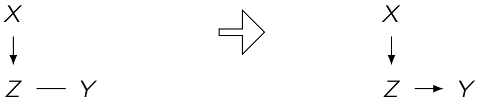
Meek-Rule 3:



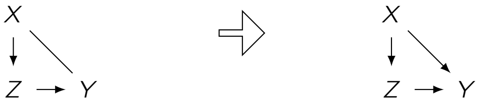
Orientation rules

Suppose we already found the skeleton and all unshielded colliders:

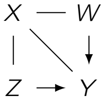
Meek-Rule 1:



Meek-Rule 2:



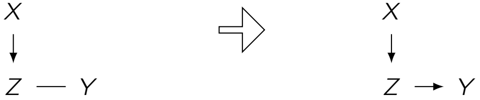
Meek-Rule 3:



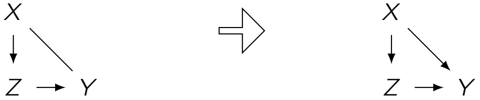
Orientation rules

Suppose we already found the skeleton and all unshielded colliders:

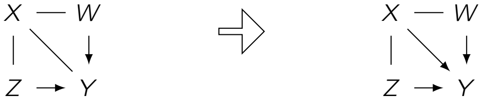
Meek-Rule 1:



Meek-Rule 2:



Meek-Rule 3:



The PC algorithm [6]

- Step 1: skeleton construction:
  - Construct a complete non oriented graph
  - Prune unnecessary edges (optimal) from the skeleton using $\perp\!\!\!\perp_P$

The PC algorithm [6]

- Step 1: skeleton construction:
  - Construct a complete non oriented graph
  - Prune unnecessary edges (optimal) from the skeleton using $\perp\!\!\!\perp_P$
- Step 2: orientation
  - Find unshielded colliders (optimal)
  - Meek-Rules 1, 2, 3

The PC algorithm [6]

- Step 1: skeleton construction:
  - Construct a complete non oriented graph
  - Prune unnecessary edges (optimal) from the skeleton using $\perp\!\!\!\perp_P$
- Step 2: orientation
  - Find unshielded colliders (optimal)
  - Meek-Rules 1, 2, 3

Theorem ([6])

Assume the distribution P is compatible and faithful to some causal DAG $\mathcal{G}$ and assume that we are given perfect conditional independence information about all pairs of variables. The PC algorithm returns the CPDAG of $\mathcal{G}$.

Pseudocode of the PC algorithm

Algorithm 1 PC

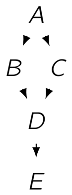
Input: $P(\mathbb{V})$

Output: CPDAG $\mathcal{G}^*$

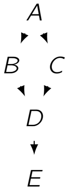
1: Initialize a complete undirected graph $\mathcal{G}^*$ on vertex set $\mathbb{V}$
2: Let $n = 0$
3: **repeat**
4:     **for all** $X - Y$ in $\mathcal{G}^*$ such that $|Ne(X, \mathcal{G}^*) \backslash \{Y\}| \geq n$
    and subsets $\mathbb{S} \subseteq Ne(X, \mathcal{G}^*) \backslash \{Y\}$ such that $|\mathbb{S}| = n$ **do**
5:         **if** $X \perp\!\!\!\perp_P Y \mid \mathcal{S}$ **then**
6:             Delete edge $X - Y$ from $\mathcal{G}^*$
7:             Let $sepset(X, Y) = sepset(Y, X) = \mathbb{S}$
8:         **end if**
9:     **end for**
10:     Let $n = n + 1$
11: **until** for each pair of adjacent vertices (X, Y), $|Ne(X, \mathcal{G}^*) \backslash \{Y\}| < n$
12: For each triple $X - Z - Y$ in $\mathcal{G}^*$, if $X \notin Ne(Y, \mathcal{G}^*)$ and $Z \notin sepsep(X, Y)$
then orient the triple as $X \rightarrow Z \leftarrow Y$
13: Recursively apply Meek-Rules until no more edges can be oriented
14: Return $\mathcal{G}^*$

PC in action

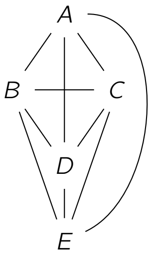
- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness



- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness



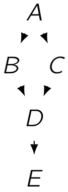
Skeleton construction:



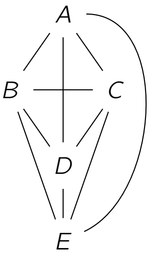
$card = 0$

PC in action

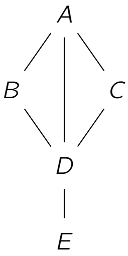
- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness



Skeleton construction:



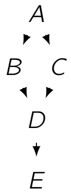
$card = 0$



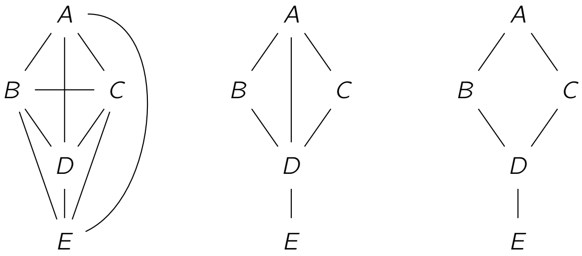
$card = 1$

PC in action

- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness



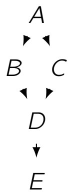
Skeleton construction:



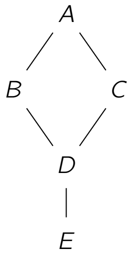
$card = 0$ $card = 1$ $card = 2$

PC in action

- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness

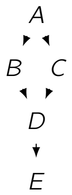


Orientation:

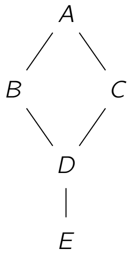


PC in action

- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness



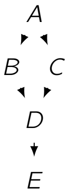
Orientation:



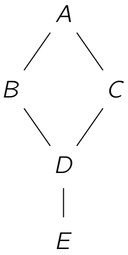
- Find unshielded colliders

PC in action

- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness

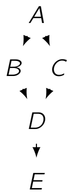


Orientation:

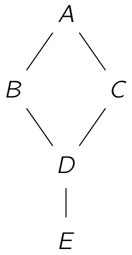


- Find unshielded colliders
  - $B \perp\!\!\!\perp_P C \mid A \implies B \rightarrow D \leftarrow C$

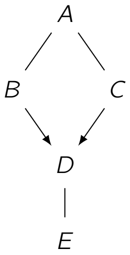
- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness



Orientation:

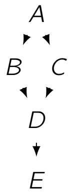


- Find unshielded colliders
  - $B \perp\!\!\!\perp_P C \mid A \implies B \rightarrow D \leftarrow C$

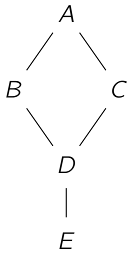


PC in action

- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness

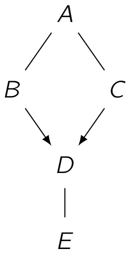


Orientation:

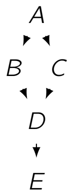


- Find unshielded colliders
  - $B \perp\!\!\!\perp_P C \mid A \implies B \to D \leftarrow C$

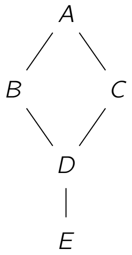
- Meek-Rule 1



- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness

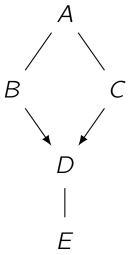


Orientation:

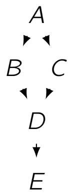


- Find unshielded colliders
  - $B \perp\!\!\!\perp_P C \mid A \implies B \to D \leftarrow C$

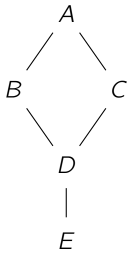
- Meek-Rule 1
  - $B \to D$ & $D - E \implies D \to E$



- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness

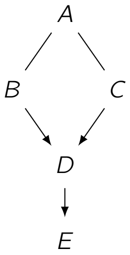


Orientation:



- Find unshielded colliders
  - $B \perp\!\!\!\perp_P C \mid A \implies B \rightarrow D \leftarrow C$

- Meek-Rule 1
  - $B \rightarrow D \;\&\; D - E \implies D \rightarrow E$



Incorporating background knowledge, extensions, and more

The PC algorithm can effectively incorporate background knowledge in the form of:

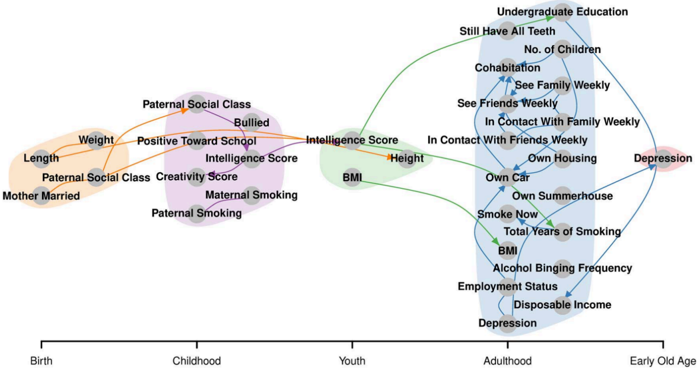
- Forbidden edges
- Required edges
- Forbidden orientations
- Required orientations

The PC algorithm has been extended to settings with unmeasured confounding: FCI algorithm [6]

Many other causal discovery algorithms have been proposed; some are able to recover the underlying DAG under additional assumptions, such as semiparametric modeling assumptions: LiNGAM, ANM, ... [4]

Real application of PC [5]

- Danish men born in 1953, followed from birth to age 65
- Data sources: surveys at ages 12 and 51 + administrative registers
- 33 variables measured across 5 life-course periods

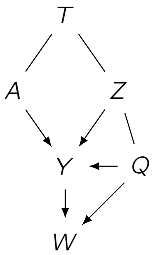


Python code: causal discovery using causallearn package

```
#Load data
import pandas as pd
data = pd.read_csv("./data.csv")
#Load the causallearn package
from causallearn.search.ConstraintBased.PC import pc
#Learn CPDAG
cg = pc(data, alpha=0.05, indep_test='fisherz')
#Plot CPDAG
cg.draw_pydot_graph()
```

Applicability of the back-door criterion to CPDAGs

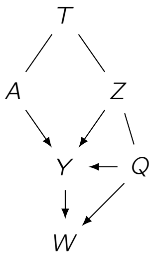
CPDAG



- Is it possible to identify the causal effect of A on Y using the CPDAG?
- What about causal effect of Y on W?

Applicability of the back-door criterion to CPDAGs

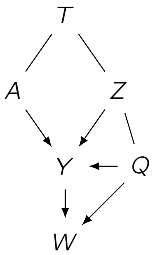
CPDAG



- Is it possible to identify the causal effect of A on Y using the CPDAG? No
- What about causal effect of Y on W? Yes

Applicability of the back-door criterion to CPDAGs

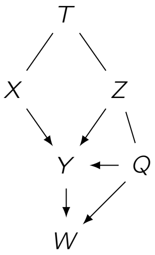
CPDAG



- Is it possible to identify the causal effect of A on Y using the CPDAG? No
- What about causal effect of Y on W? Yes

A set of variables $\mathbb{Z}$ satisfies the back-door criterion relative to an ordered pair of variables (A, Y) in a CPDAG if $\mathbb{Z}$ does not contain a possible descendant of A; and $\mathbb{Z}$ blocks all possible paths between A and Y that contain an arrow pointing toward A.

Exercise 4



Which orientation rules did PC use to orient this CPDAG?

7

References

References I

[1] Judea Pearl. "Causal diagrams for empirical research". In: Biometrika 82.4 (Dec. 1995), pp. 669–688. issn: 0006-3444. doi: 10.1093/biomet/82.4.669.

[2] Judea Pearl. Causality: Models, Reasoning, and Inference. Cambridge University Press, 2000.

[3] Judea Pearl and Dana Mackenzie. The Book of Why: The New Science of Cause and Effect. Basic Books, 2018.

[4] Jonas Peters, D. Janzing, and B. Schölkopf. Elements of Causal Inference: Foundations and Learning Algorithms. Cambridge, MA, USA: MIT Press, 2017.

References II

[5] Anne H. Petersen, Merete Osler, and Claus Ekstrøm. "Data-Driven Model Building for Life Course Epidemiology". In: American Journal of Epidemiology 190 (Mar. 2021). doi: 10.1093/aje/kwab087.

[6] Peter Spirtes, Clark Glymour, and Richard Scheines. Causation, Prediction, and Search. 2nd. MIT press, 2000.

[7] Thomas Verma and Judea Pearl. "Equivalence and synthesis of causal models". In: UAI '90: Proceedings of the Sixth Annual Conference on Uncertainty in Ed. by Piero P. Bonissone et al. Elsevier, 1990, pp. 255–270.

MERCI POUR VOTRE ATTENTION

SANTÉ
SORBONNE
UNIVERSITÉ