Causal Inference: Presenting Tools to Ascend the Ladder of Causation

Charles Assaad charles.assaad@inserm.fr

L'Institut Pierre Louis d'Epidémiologie et de Santé Publique, Inserm, Sorbonne Université

June 2025

SANTÉ SORBONNE UNIVERSITÉ

Sommaire

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?

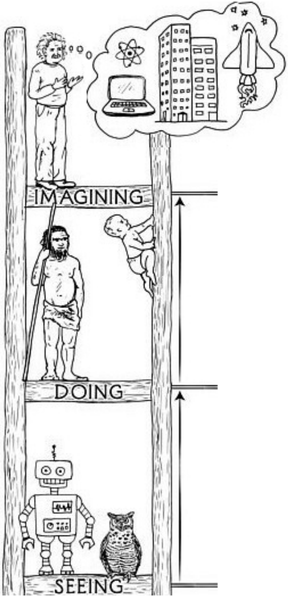
Some epidemiological questions

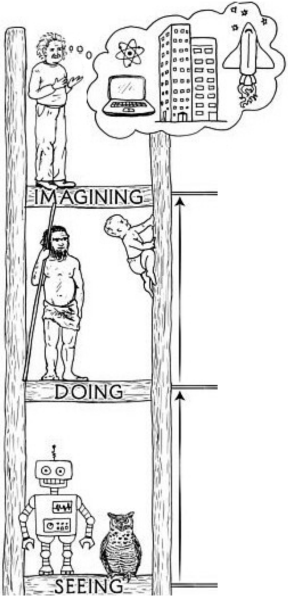
- Do people who exercise more tend to have lower rates of heart disease?
- Can we predict who is at higher risk of developing diabetes using BIG DATA?
- Does exercising reduce cholesterol?
- What happens to blood pressure if we prescribe a new hypertension drug?
- Would COVID-19 outcomes in France have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

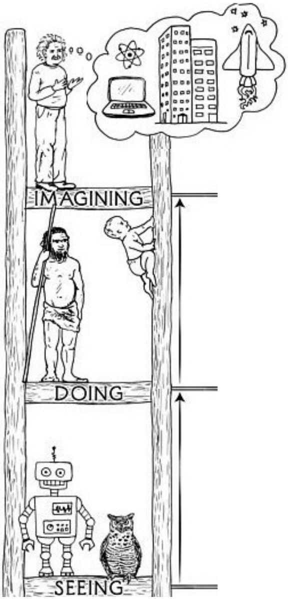
Pearl causal hierarchy





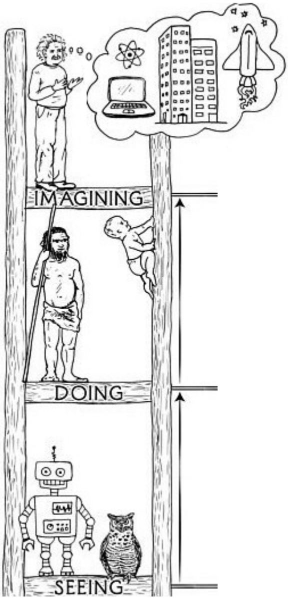
- Associations
  - Questions : What if I see ...?
  - Do people who exercise more tend to have lower rates of heart disease?
  - Can we predict who is at higher risk of developing diabetes using BIG DATA?

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?

2

Association

- Bayesian networks
- d-separation
- Prediction
- Do we need more than associations?

2

Association

Bayesian networks

d-separation

Prediction

Do we need more than associations?

Notations

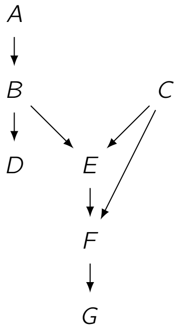
- Sets: $\mathbb{A} = \{X, Y, Z\}$
- Statistical independence: $\perp\!\!\!\perp_P$
- Statistical dependence: $\not\perp\!\!\!\perp_P$
- $P(Y = y \mid X = x) \equiv P(y \mid x)$

Directed graphs

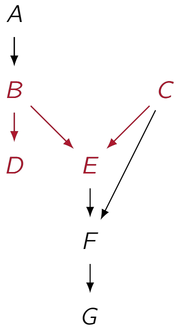
A graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ is said to be a directed graph iff

- $\mathbb{V}$ 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 $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



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

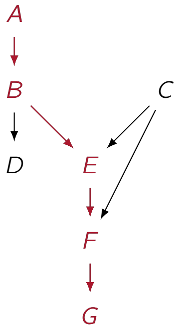


Path: $D \leftarrow B \rightarrow E \leftarrow C$

Directed graphs: basic concepts

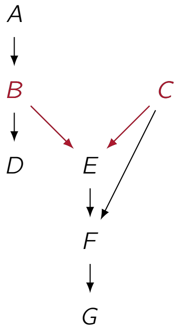
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:

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



Directed graphs: basic concepts

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

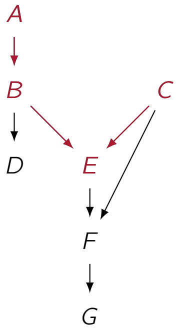


Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$

Directed graphs: basic concepts

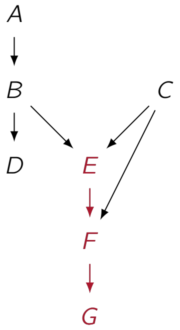
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:

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: $An(E) = \{A, B, C, E\}$



Directed graphs: basic concepts

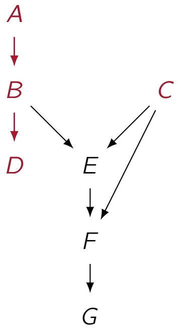
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



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: $De(E) = \{E, F, G\}$

Directed graphs: basic concepts

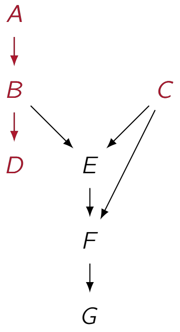
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



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: $Nd(E) = \{A, B, C, D\}$

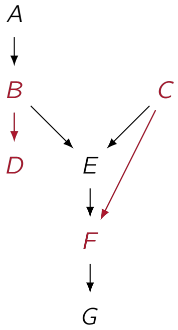
Directed graphs: basic concepts

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: $Nd(E) = \{A, B, C, D\}$
Ancestral sets: a subset of vertices $\mathbb{S}$ is ancestral if $\forall X \in \mathbb{S}, An(X) \subseteq \mathbb{S}$

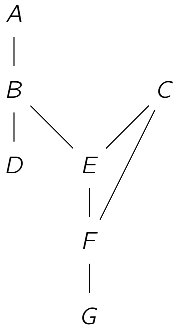
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



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: $Nd(E) = \{A, B, C, D\}$
Ancestral sets: a subset of vertices $\mathbb{S}$ is ancestral if $\forall X \in \mathbb{S}, An(X) \subseteq \mathbb{S}$
Subgraph $\mathcal{G}[\mathbb{S}]$: $\mathcal{G}[\{B, C, D, F\}]$

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: $Nd(E) = \{A, B, C, D\}$
Ancestral sets: a subset of vertices $\mathbb{S}$ is ancestral if $\forall X \in \mathbb{S}, An(X) \subseteq \mathbb{S}$
Subgraph $\mathcal{G}[\mathbb{S}]$: $\mathcal{G}[\{B, C, D, F\}]$
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.

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 $\mathcal{G}$.

Decomposing with respect to ancestral sets

Theorem

If P is compatible with $\mathcal{G}$ and $\mathbb{S} \subseteq \mathbb{V}$ is an ancestral set, then $P(\mathbb{S})$ is compatible with $\mathcal{G}[\mathbb{S}]$.

Decomposing with respect to ancestral sets

Theorem

If P is compatible with $\mathcal{G}$ and $\mathbb{S} \subseteq \mathbb{V}$ is an ancestral set, then $P(\mathbb{S})$ is compatible with $\mathcal{G}[\mathbb{S}]$.

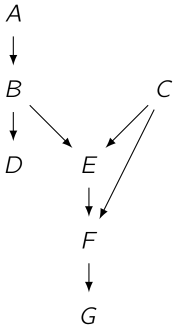
$$\begin{array}{ccc} A & & \\ \downarrow & & \\ B & & C \\ \downarrow & \searrow \quad \swarrow & \downarrow \\ D & E & \\ & \downarrow & \swarrow \\ & F & \\ & \downarrow & \\ & G & \end{array}$$

Bayesian network compatible with P

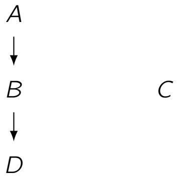
Decomposing with respect to ancestral sets

Theorem

If P is compatible with $\mathcal{G}$ and $\mathbb{S} \subseteq \mathbb{V}$ is an ancestral set, then $P(\mathbb{S})$ is compatible with $\mathcal{G}[\mathbb{S}]$.



Bayesian network compatible with P



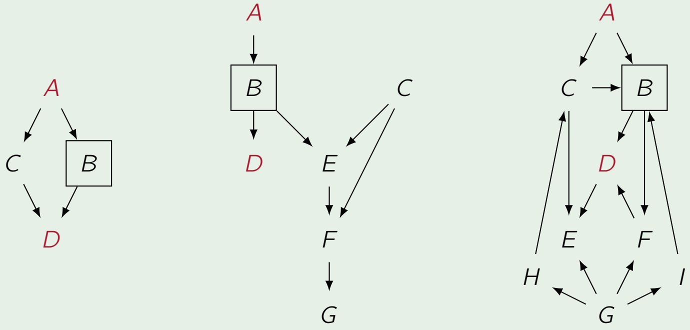
Bayesian network compatible with $P(A, B, C, D)$

2

Association

Example

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



Basic structures

Fork: contains a confounder

$$A \longleftarrow C \longrightarrow B$$

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

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

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

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

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

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

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

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

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

V-structure: contains a collider

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

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

Fork: contains a confounder

$$\begin{tikzcd} A & C \arrow[l] \arrow[r] & B \end{tikzcd} \qquad \begin{tikzcd} A & \boxed{C} \arrow[l] \arrow[r] & B \end{tikzcd}$$

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

Chain: contains a mediator or an intermediate cause

$$\begin{tikzcd} A \arrow[r] & C \arrow[r] & B \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r] & \boxed{C} \arrow[r] & B \end{tikzcd}$$

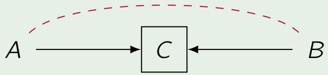
$$A \overset{?}{\perp\!\!\!\perp}_P B \qquad A \perp\!\!\!\perp_P B \mid C$$

V-structure: contains a collider

$$\begin{tikzcd} A \arrow[r] & C & B \arrow[l] \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r] \arrow[rr, dashed, bend left, no head] & C \arrow[d] & B \arrow[l] \\ & \boxed{D} & \end{tikzcd}$$

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

Example



Artificial correlation in V-structures: example 1

Example

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

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

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

Artificial correlation in V-structures: example 1

Example

$$\begin{tikzcd} A \arrow[r] \arrow[rr, bend left, dashed, no head] & \boxed{C} & B \arrow[l] \end{tikzcd}$$

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

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

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

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

Artificial correlation in V-structures: example 2

Example

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

Artificial correlation in V-structures: example 2

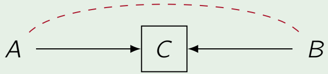
Example

$$\begin{tikzcd} A \arrow[r] \arrow[rr, dash, dashed, bend left] & \boxed{C} & B \arrow[l] \end{tikzcd}$$

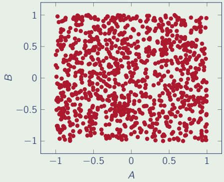
$A, B \sim U(-1, 1)$ $\qquad$ $\xi_c \sim N(0, \frac{1}{2})$ $\qquad$ $C = 2AB + \xi_c$

Artificial correlation in V-structures: example 2

Example



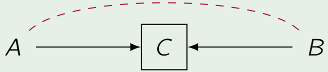
$A, B \sim U(-1, 1)$ $\quad\quad$ $\xi_c \sim N(0, \frac{1}{2})$ $\quad\quad$ $C = 2AB + \xi_c$



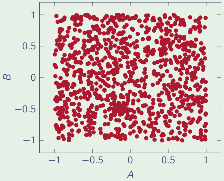
$Corr(A; B) = 0.002$

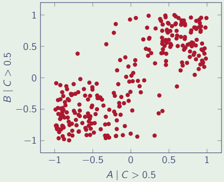
Artificial correlation in V-structures: example 2

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:

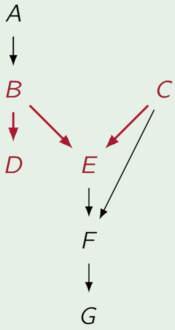
- it contains a chain $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ and $B \in \mathbb{Z}$; or
- it contains a collider $A \rightarrow B \leftarrow C$ such that no descendant of B is in $\mathbb{Z}$.

A path that is not blocked is active.

Example

$$
\begin{tikzcd}
A \arrow[d] & & \\
B \arrow[d] \arrow[dr] & & C \arrow[dl] \arrow[ddl] \\
D & E \arrow[d] & \\
& F \arrow[d] & \\
& G &
\end{tikzcd}
$$

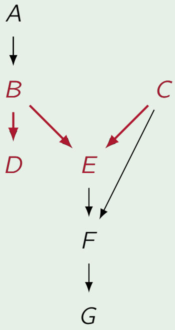
Example



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

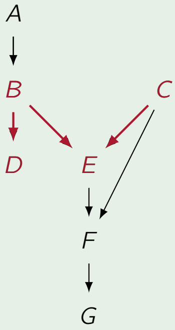
Blocked path: examples

Example



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

Example

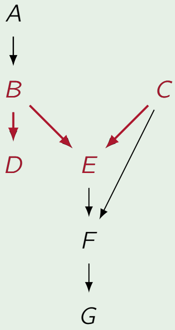


Is the path $< D, B, E, C >$ blocked? Yes

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

Blocked path: examples

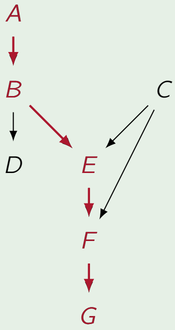
Example



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

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

Example



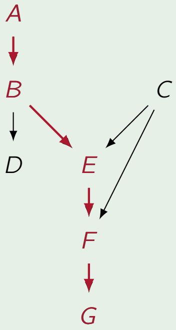
Is the path $< D, B, E, C >$ blocked? Yes

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

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

Blocked path: examples

Example



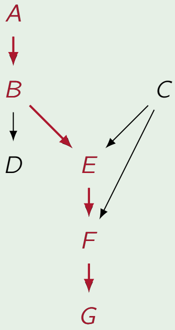
Is the path $< D, B, E, C >$ blocked? Yes

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

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

Blocked path: examples

Example



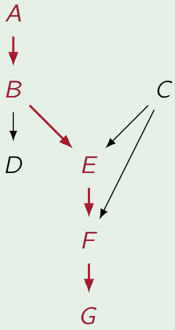
Is the path $< D, B, E, C >$ blocked? Yes

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

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

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

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

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

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

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 $\mathbb{X} \not\!\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

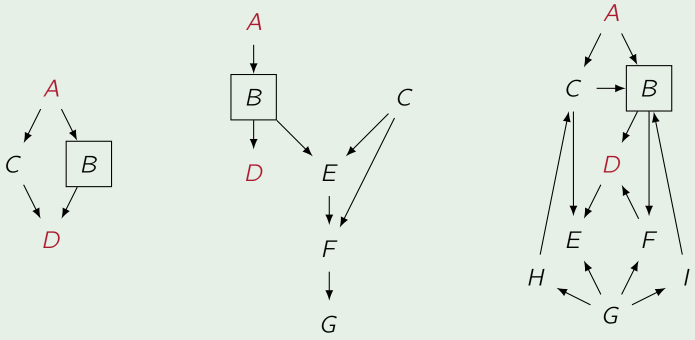
Theorem

$\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \Rightarrow \mathbb{X} \perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$
but $\mathbb{X} \not\!\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \not\Rightarrow \mathbb{X} \not\!\perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$ instead

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

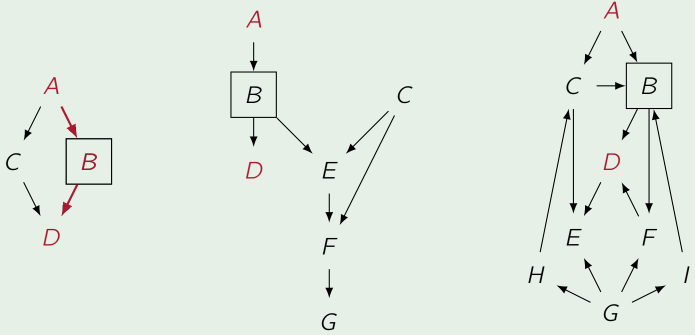
Example

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



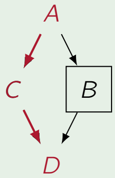
Example

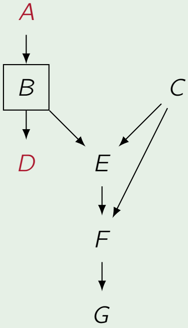
Is $A \perp\!\!\!\perp_P D \mid B$?

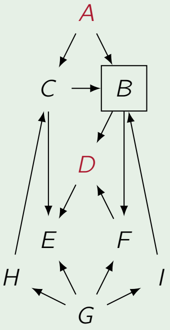


Example

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



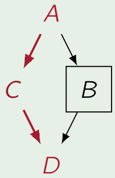


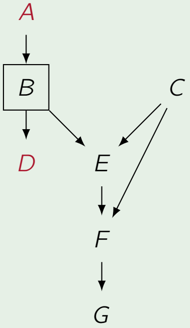


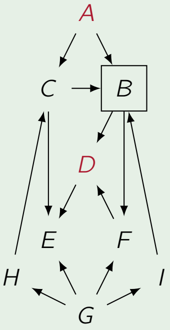
Reading conditional independencies using d-separation

Example

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







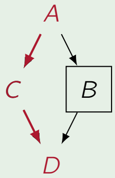
$< A, C, D >$ is not blocked

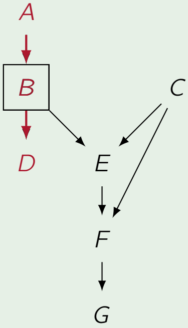
$\overset{?}{\implies} A \perp\!\!\!\perp_P D \mid B$

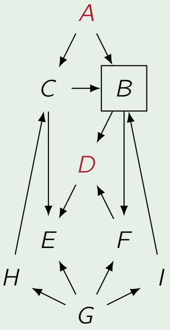
Reading conditional independencies using d-separation

Example

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







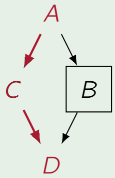
$< A, C, D >$ is not blocked

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

Reading conditional independencies using d-separation

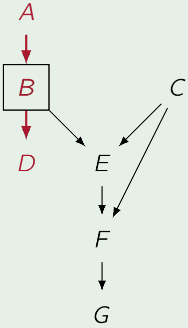
Example

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



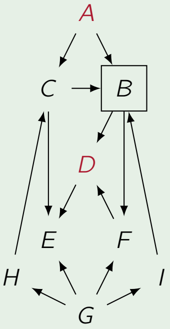
$< A, C, D >$ is not blocked

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



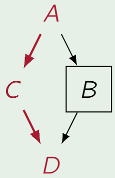
All paths are blocked

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



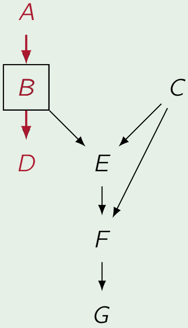
Example

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



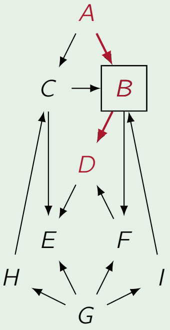
$< A, C, D >$ is not blocked

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



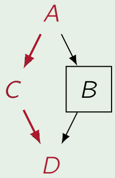
All paths are blocked

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



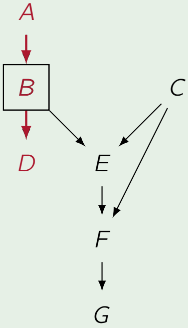
Example

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



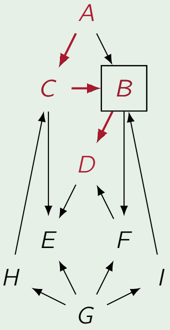
$< A, C, D >$ is not blocked

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



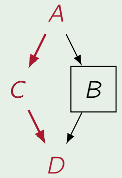
All paths are blocked

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

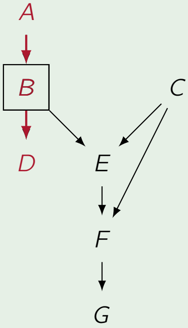


Example

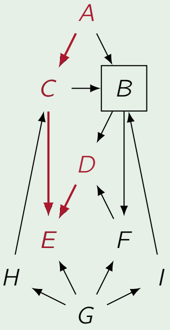
Is $A \perp\!\!\!\perp_P D \mid B$?



$< A, C, D >$ is not blocked $\overset{?}{\Longrightarrow} A \perp\!\!\!\perp_P D \mid B$

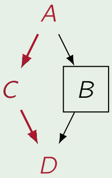


All paths are blocked $\Longrightarrow A \perp\!\!\!\perp_P D \mid B$



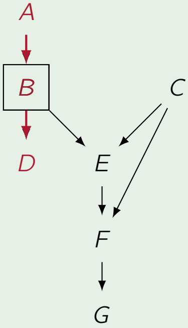
Example

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



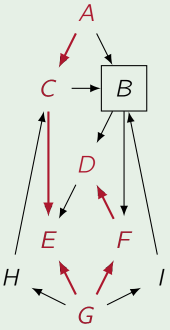
$< A, C, D >$ is not blocked

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



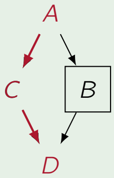
All paths are blocked

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



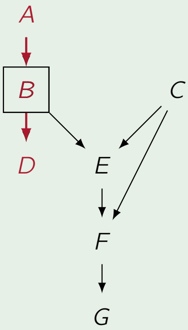
Example

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



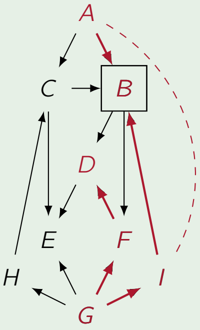
$< A, C, D >$ is not blocked

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



All paths are blocked

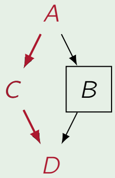
$\implies A \perp\!\!\!\perp_P D \mid B$



Reading conditional independencies using d-separation

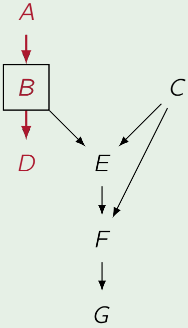
Example

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



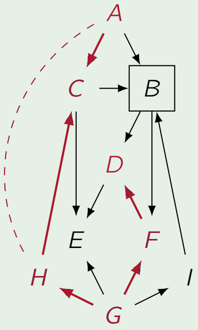
$< A, C, D >$ is not blocked

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



All paths are blocked

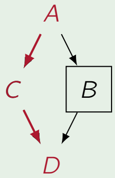
$\implies A \perp\!\!\!\perp_P D \mid B$



Reading conditional independencies using d-separation

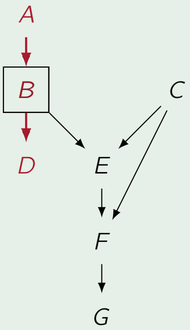
Example

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



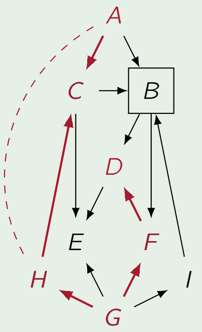
$< A, C, D >$ is not blocked

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



All paths are blocked

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



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

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

2

Association

- Bayesian networks
- d-separation
- Prediction
- Do we need more than associations?

Bayesian networks for prediction

Bayesian networks can be used to detect relevant variables that contains all the useful information for prediciton!

Bayesian networks can be used to detect relevant variables that contains all the useful information for prediciton! (Assuming the distribution does not change)

Bayesian networks for prediction

Bayesian networks can be used to detect relevant variables that contains all the useful information for prediciton! (Assuming the distribution does not change)

Relevant variables for prediciting B:

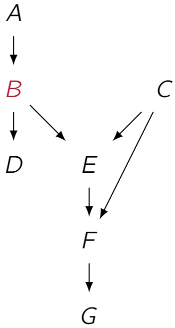
- Parents of B;
- Children of B;
- Other parents of the children of B.

Bayesian networks for prediction

Bayesian networks can be used to detect relevant variables that contains all the useful information for prediciton! (Assuming the distribution does not change)

Relevant variables for prediciting B:

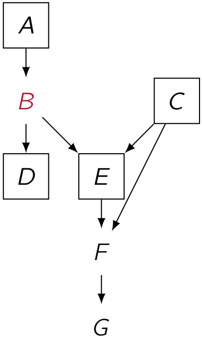
- Parents of B;
- Children of B;
- Other parents of the children of B.



Bayesian networks can be used to detect relevant variables that contains all the useful information for prediciton! (Assuming the distribution does not change)

Relevant variables for prediciting B:

- Parents of B;
- Children of B;
- Other parents of the children of B.

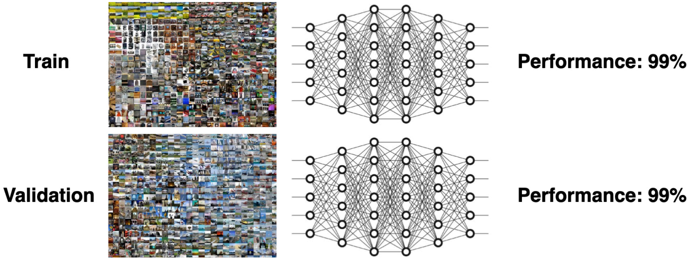


2

Association

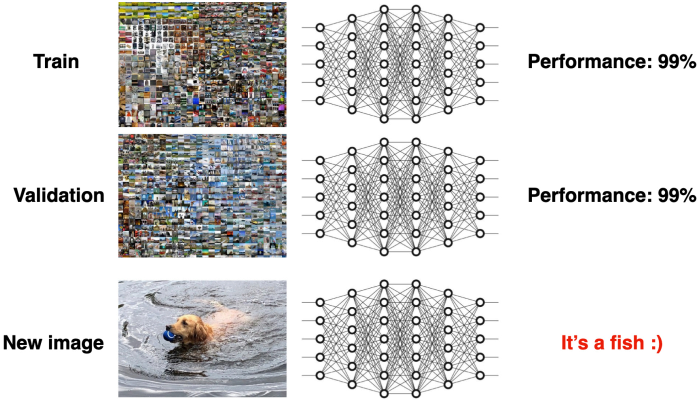
Bayesian networks
d-separation
Prediction
Do we need more than associations?

Distribution shifts



P. Cui and T. Zhang. Causal Inference and Stable Learning. Tutoriel at ICML, 2019.

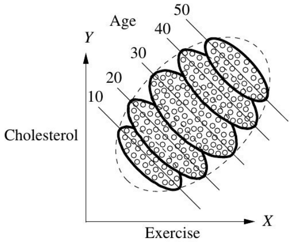
Distribution shifts



P. Cui and T. Zhang. Causal Inference and Stable Learning. Tutoriel at ICML, 2019.

Simpson paradox [6]

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



Exercise 1

Suppose this DAG is a bayesian network compatible with P:

$$\begin{tikzcd} A \arrow[r] & B \arrow[dl] \arrow[d] & C \arrow[l] \arrow[dl] \arrow[ddl] \\ D & E \arrow[d] & \\ G \arrow[r] & F & \end{tikzcd}$$

Do the following subgraphs represent Bayesian networks that are compatible with their respective distributions?

$$\begin{tikzcd} A \arrow[r] & B & C \arrow[l] \end{tikzcd}$$

$P(A, B, C)$

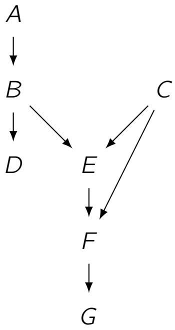
$$\begin{tikzcd} B \arrow[d] \\ E \arrow[d] \\ F \end{tikzcd}$$

$P(B, E, F)$

$$\begin{tikzcd} A \arrow[r] & B \arrow[d] & C \arrow[l] \arrow[dl] \\ & E & \end{tikzcd}$$

$P(A, B, C, E)$

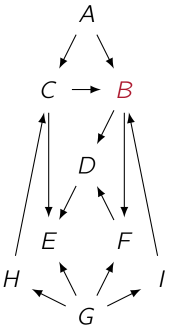
Exercise 2



- $B \perp\!\!\!\perp_P G \mid F$?
- $A \perp\!\!\!\perp_P F \mid E$?
- $B \perp\!\!\!\perp_P E \mid A, C, F$?

Exercise 3

What are the relevant variables for predicting B?



3

Intervention

Causal DAGs
Causal reasoning

3

Intervention

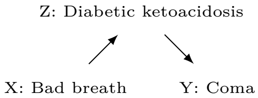
Causal DAGs

Causal reasoning

Bayesian networks vs causal DAGs

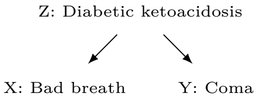
$$\begin{tikzcd}
& \text{Z: Diabetic ketoacidosis} \arrow[dr] & \\
\text{X: Bad breath} \arrow[ur] & & \text{Y: Coma}
\end{tikzcd}
\qquad
\begin{tikzcd}
& \text{Z: Diabetic ketoacidosis} \arrow[dl] \arrow[dr] & \\
\text{X: Bad breath} & & \text{Y: Coma}
\end{tikzcd}$$

Bayesian networks vs causal DAGs



Bad breath $\perp\!\!\!\perp_P$ Coma $|$ Diabetic ketoacidosis

Bayesian network



Bad breath $\perp\!\!\!\perp_P$ Coma $|$ Diabetic ketoacidosis

Bayesian network

Bayesian networks vs causal DAGs

$$\begin{tikzcd} & \text{Z: Diabetic ketoacidosis} \arrow[dr] & \\ \text{X: Bad breath} \arrow[ur] & & \text{Y: Coma} \end{tikzcd}$$

Bad breath $\perp\!\!\!\perp_P$ Coma | Diabetic ketoacidosis

Bayesian network

Not a causal DAG

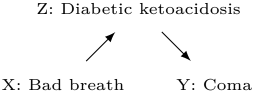
$$\begin{tikzcd} & \text{Z: Diabetic ketoacidosis} \arrow[dl] \arrow[dr] & \\ \text{X: Bad breath} & & \text{Y: Coma} \end{tikzcd}$$

Bad breath $\perp\!\!\!\perp_P$ Coma | Diabetic ketoacidosis

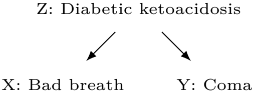
Bayesian network

Causal DAG

Bayesian networks vs causal DAGs



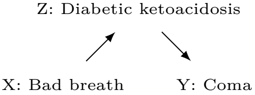
Bad breath $\perp\!\!\!\perp_P$ Coma |
Diabetic ketoacidosis
Bayesian network
Not a causal DAG



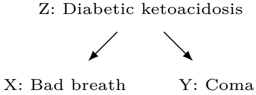
Bad breath $\perp\!\!\!\perp_P$ Coma |
Diabetic ketoacidosis
Bayesian network
Causal DAG

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

Bayesian networks vs causal DAGs



Bad breath $\perp\!\!\!\perp_P$ Coma |
Diabetic ketoacidosis
Bayesian network
Not a causal DAG



Bad breath $\perp\!\!\!\perp_P$ Coma |
Diabetic ketoacidosis
Bayesian network
Causal DAG

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

Oracle for conditional independence

Oracle for intervention

Bayesian networks vs causal DAGs

$$\begin{tikzcd}[column sep=small] & \text{Z: Diabetic ketoacidosis} \arrow[dr] & \\ \text{X: Bad breath} \arrow[ur] & & \text{Y: Coma} \end{tikzcd} \qquad \begin{tikzcd}[column sep=small] & \text{Z: Diabetic ketoacidosis} \arrow[dl] \arrow[dr] & \\ \text{X: Bad breath} & & \text{Y: Coma} \end{tikzcd}$$

Bad breath $\perp\!\!\!\perp_P$ Coma $|$ Diabetic ketoacidosis
Bayesian network
Not a causal DAG

Bad breath $\perp\!\!\!\perp_P$ Coma $|$ Diabetic ketoacidosis
Bayesian network
Causal DAG

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

Oracle for conditional independence

Oracle for intervention

The $do()$ operator encodes (structural) interventions.

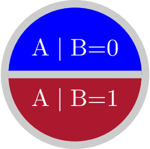
Conditioning vs Intervening

Population

Conditioning vs Intervening

Population

Sub-populations

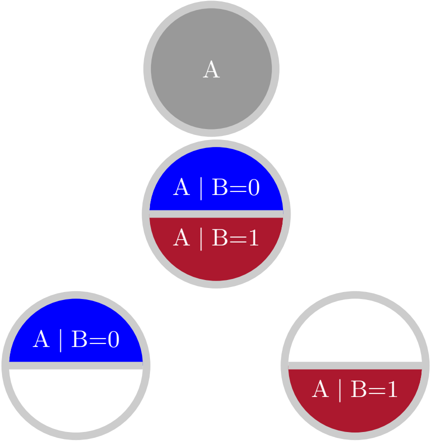


Conditioning vs Intervening

Population

Sub-populations

Conditioning



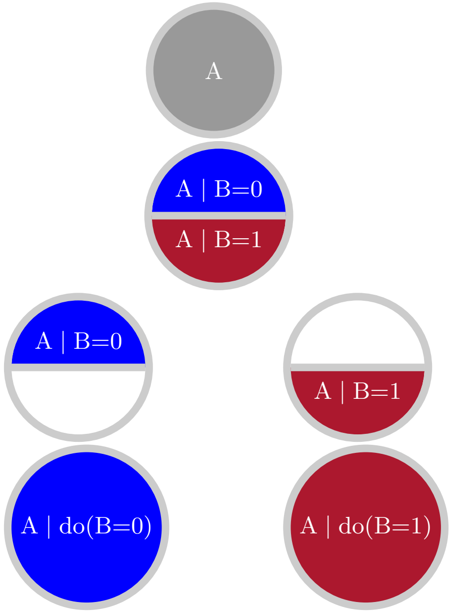
Conditioning vs Intervening

Population

Sub-populations

Conditioning

Intervening



Causal Bayesian networks - Causal DAG

Truncated factorization [4]: If we intervene on a subset $\mathbb{S} \subset \mathbb{V}$, then

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

Causal Bayesian networks - Causal DAG

Truncated factorization [4]: If we intervene on a subset $\mathbb{S} \subset \mathbb{V}$, then

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

Also known as G-computation formula [2] or Manipulation theorem [10].

Truncated factorization [4]: If we intervene on a subset $\mathbb{S} \subset \mathbb{V}$, then

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

Also known as G-computation formula [2] or Manipulation theorem [10].

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.

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 $Z \rightarrow A$) to the graph, ADD IT! (as long as you keep the graph acyclic)

Guidlines for constructing causal DAGs

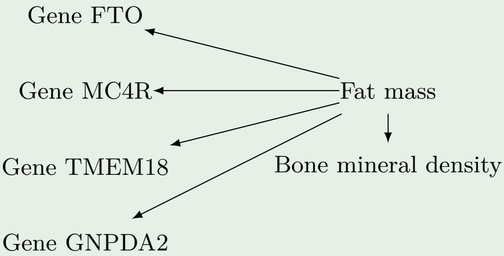
In causal DAGs, the information lies in the absence of edges and the direction of arrows, rather than merely in the presence of edges.

- Think carrefully about the orientation
- When you are not sure if you need to add or not an edge (for example $Z \to A$) 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 presentation we will focus only on acyclic graphs.

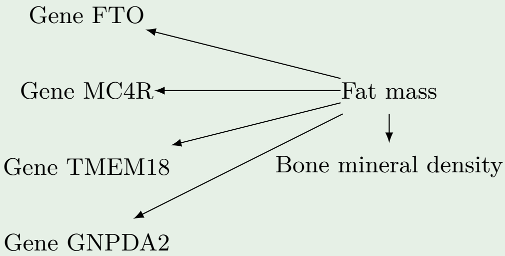
DAG Vs DAG causal: exemples

Example



Is this DAG causal?

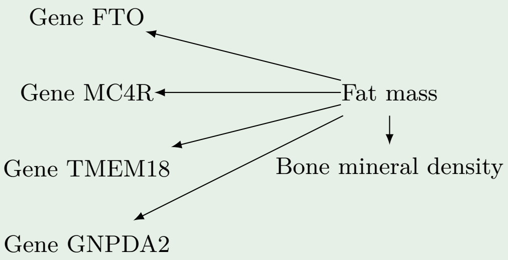
Example



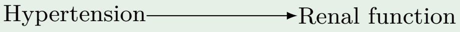
Is this DAG causal?
No

DAG Vs DAG causal: exemples

Example



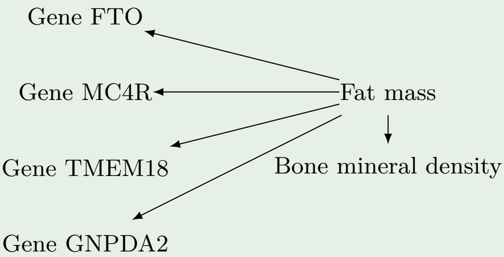
Is this DAG causal?
No



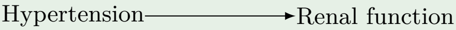
Is this DAG causal?

DAG Vs DAG causal: exemples

Example



Is this DAG causal?
No



Is this DAG causal?
Yes

3

Intervention

Causal DAGs

Causal reasoning

Identifiability and positivity

A quantity Q is said to be identifiable if it is uniquely computable from a positive distribution $P(\mathbb{V})$.

Identifiability and positivity

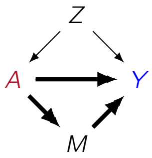
A quantity Q is said to be identifiable if it is uniquely computable from a positive distribution $P(\mathbb{V})$.

In the following, we will assume that all distributions are positive.

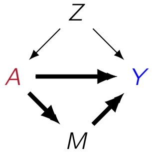
The total effect of $\mathbb{A}$ on $\mathbb{Y}$
$= \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a})) - \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a}'))$

The total effect of $\mathbb{A}$ on $\mathbb{Y}$

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

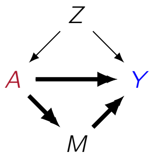


The total effect of $\mathbb{A}$ on $\mathbb{Y}$
$= \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a})) - \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a}'))$



The total effect is identifiable iff $P(\mathbb{Y} \mid do(\mathbb{A}))$ is identifiable.

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

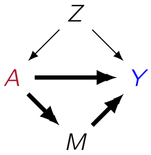


The total effect is identifiable iff $P(\mathbb{Y} \mid do(\mathbb{A}))$ is identifiable.

Causal reasoning involves utilizing a causal DAG to determine whether $P(\mathbb{Y} \mid do(\mathbb{A}))$ is identifiable.

Causal reasoning (1/2)

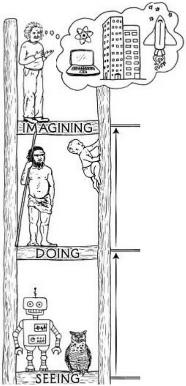
The total effect of $\mathbb{A}$ on $\mathbb{Y}$
$$= \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a})) - \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a}'))$$



The total effect is identifiable iff $P(\mathbb{Y} \mid do(\mathbb{A}))$ is identifiable.

Causal reasoning involves utilizing a causal DAG to determine whether $P(\mathbb{Y} \mid do(\mathbb{A}))$ is identifiable.

In what follows, we will use the terms total effect and causal effect interchangeably.



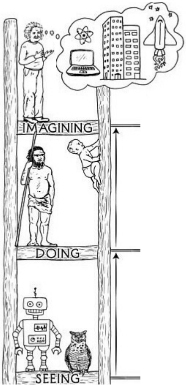
Observational data

Causal reasoning

Causal reasoning (2/2)

Causal DAG

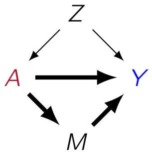
Observational data



Causal reasoning

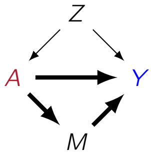
Key challenges of not using a causal DAG (1/3)

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



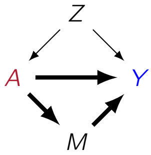
Key challenges of not using a causal DAG (1/3)

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



We must eliminate all confounding bias by adjusting for confounders.

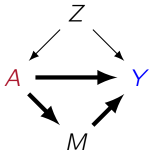
Confounding Bias (e.g., Simpson paradox):



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

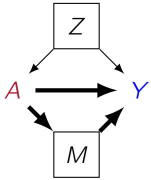
Key challenges of not using a causal DAG (1/3)

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



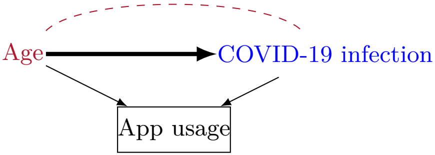
We must eliminate all confounding bias by adjusting for confounders.
Should we always adjust for all available variables? No!

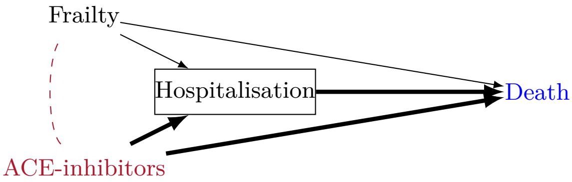
Bias due to incorrect adjustment for mediators:



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

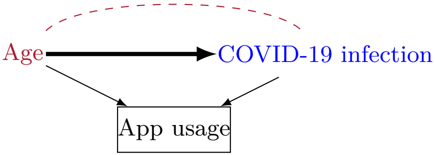
Bias due to adjusting for colliders:

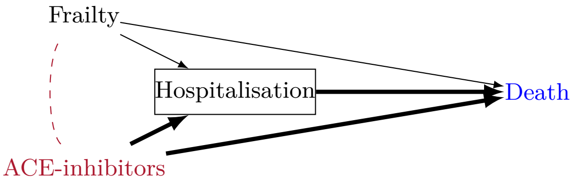




Key challenges of not using a causal DAG (3/3)

Bias due to adjusting for colliders:

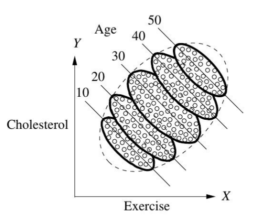




We cannot adjust for everything that is temporally prior to the treatment!

Example

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

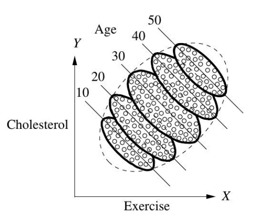


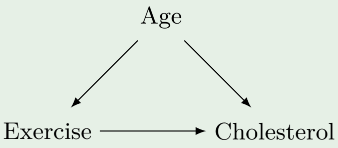
What is the effect of exercise on cholesterol $P(c \mid do(e))$?

Simpson paradox and a simple solution

Example

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



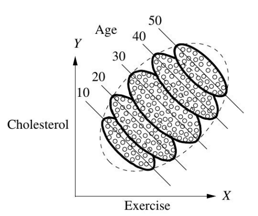


What is the effect of exercise on cholesterol $P(c \mid do(e))$?

Simpson paradox and a simple solution

Example

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



$$\begin{tikzcd} & \text{Age} \arrow[dl] \arrow[dr] & \\ \text{Exercise} \arrow[rr] & & \text{Cholesterol} \end{tikzcd}$$

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

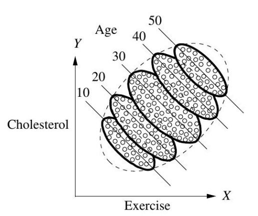
$$P(a, e, c) = P(a)P(e \mid a)P(c \mid a, e) \qquad \text{(Compatibility)}$$

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

Simpson paradox and a simple solution

Example

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



$$
\begin{tikzcd}
 & \text{Age} \arrow[dl] \arrow[dr] & \\
\text{Exercise} \arrow[rr] & & \text{Cholesterol}
\end{tikzcd}
$$

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

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

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

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

A simple solution: direct causes adjustment [4]

Theorem

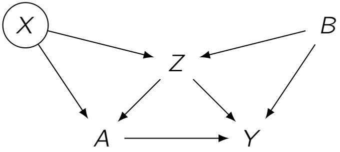
Given a causal DAG $\mathcal{G}$ in which a subset $\mathbb{V}$ of variables are measured, the causal effect $P(y \mid do(a))$ is identifiable whenever $\{A \cup Y \cup Parents(A)\} \subseteq \mathbb{V}$, and is given by:

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

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.

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:

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

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"?

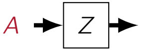
$$A \rightarrow \boxed{Z} \rightarrow Y$$

To avoid blocking on intermediate causes.

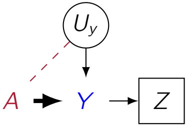
- Why "$\mathbb{Z}$ blocks all paths between A and Y that contain an arrow pointing toward A"?

The back-door criterion: intuition

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



To avoid blocking on intermediate causes.

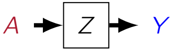


To avoid introducing artificial confounding bias resulting from conditioning on a collider.

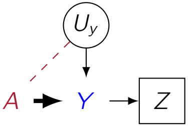
- Why ”$\mathbb{Z}$ blocks all paths between A and Y that contain an arrow pointing toward A”?

The back-door criterion: intuition

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



To avoid blocking on intermediate causes.

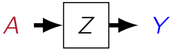


To avoid introducing artificial confounding bias resulting from conditioning on a collider.
To avoid selection bias!

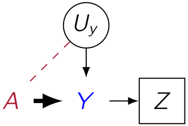
- Why "$\mathbb{Z}$ blocks all paths between A and Y that contain an arrow pointing toward A"?

The back-door criterion: intuition

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



To avoid blocking on intermediate causes.



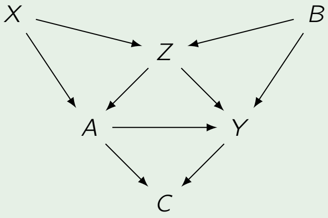
To avoid introducing artificial confounding bias resulting from conditioning on a collider.
To avoid selection bias!

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

To eliminate confounding bias.

Example

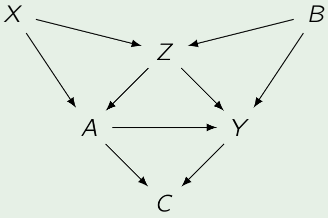
$P(y \mid do(a))$?



Example

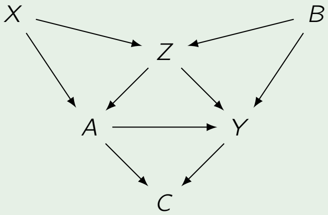
$P(y \mid do(a))$?

Z ?



Example

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

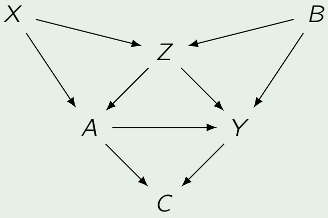


Z ? No

Back-door criterion in action

Example

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



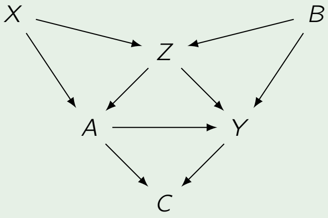
Z ? No

X ?

Back-door criterion in action

Example

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



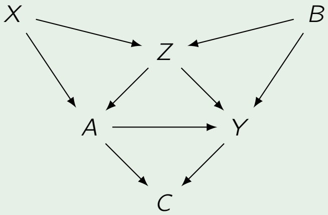
Z ? No

X ? No

Back-door criterion in action

Example

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



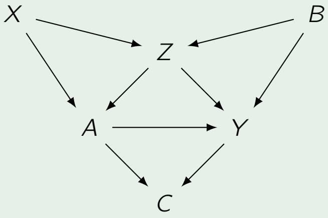
Z ? No

X ? No

B ?

Example

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



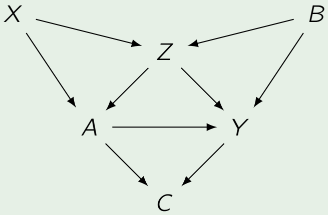
Z ? No

X ? No

B ? No

Example

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



Z ? No

X ? No

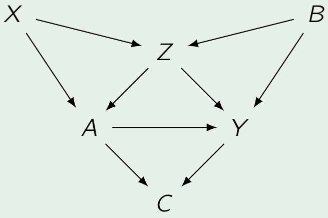
B ? No

C ?

Back-door criterion in action

Example

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



Z ? No

X ? No

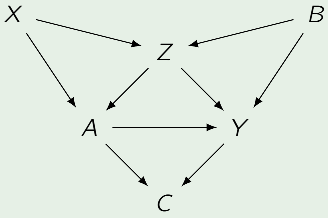
B ? No

C ? No

Back-door criterion in action

Example

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



Z ? No

X ? No

B ? No

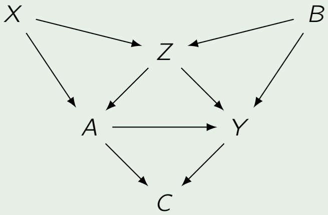
C ? No

X, B?

Back-door criterion in action

Example

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



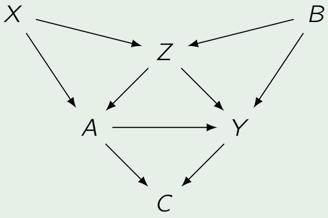
Z ? No
X ? No
B ? No
C ? No
X, B? No

SANTÉ SORBONNE UNIVERSITÉ

Back-door criterion in action

Example

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



Z ? No

X ? No

B ? No

C ? No

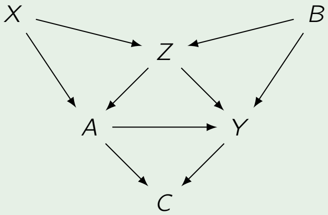
X, B? No

Z, X ?

Back-door criterion in action

Example

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



Z ? No

X ? No

B ? No

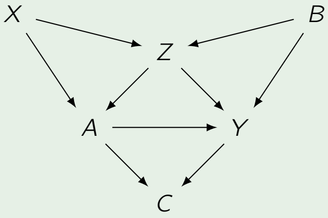
C ? No

X, B? No

Z, X ? Yes

Example

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



Z ? No

X ? No

B ? No

C ? No

X, B? No

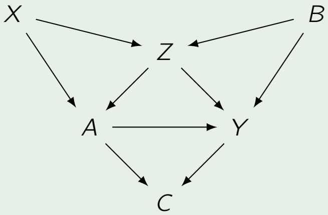
Z, X ? Yes

Z, B ?

Back-door criterion in action

Example

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



Z ? No

X ? No

B ? No

C ? No

X, B? No

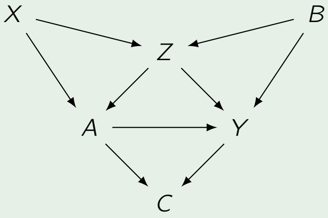
Z, X ? Yes

Z, B ? Yes

Back-door criterion in action

Example

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



Z ? No

X ? No

B ? No

C ? No

X, B? No

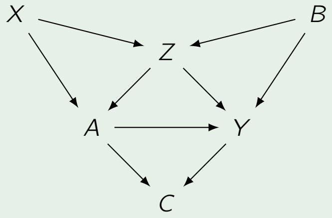
Z, X ? Yes

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: example 1

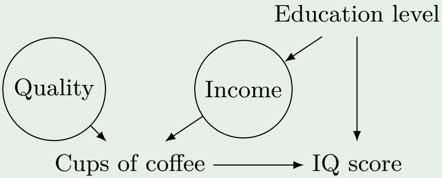
Example

In a study, we measure the number of coffee intake and IQ score for a sample of a population with various education level.

What is the effect of the number cups of coffee on IQ score $P(i \mid do(c))$?

Example

In a study, we measure the number of coffee intake and IQ score for a sample of a population with various education level.

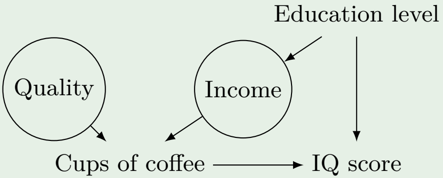


What is the effect of the number cups of coffee on IQ score $P(i \mid do(c))$?

Back-door criterion: example 1

Example

In a study, we measure the number of coffee intake and IQ score for a sample of a population with various education level.



What is the effect of the number cups of coffee on IQ score $P(i \mid do(c))$?

$$P(i \mid do(c)) = \sum_e P(i \mid c, e) P(e)$$

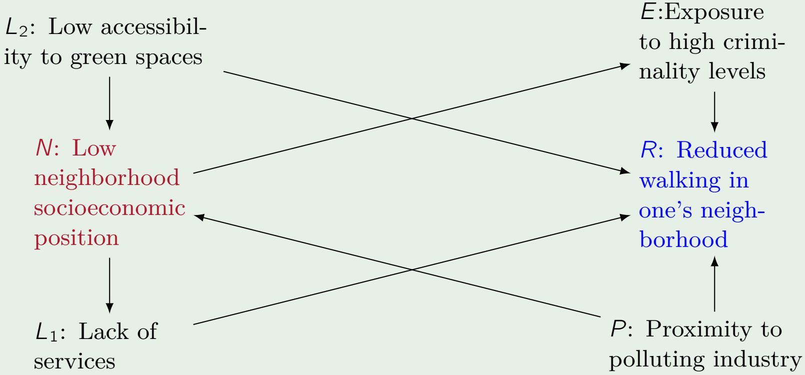
Back-door criterion: example 2

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), $P(r \mid do(n))$?

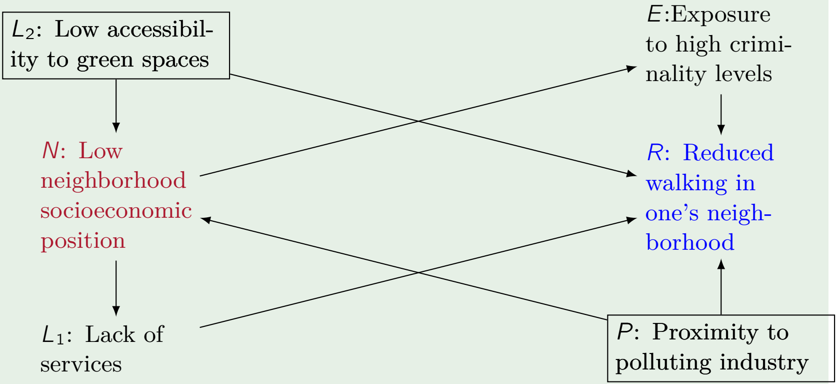
Example

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

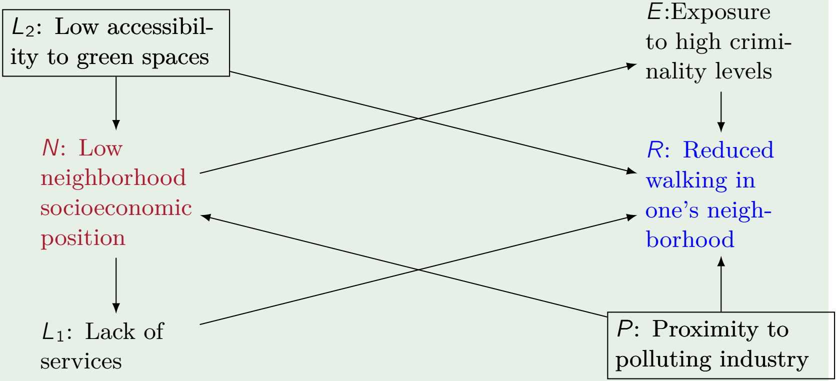
In this study, we aim to estimate the effect of the neighborhood's socioeconomic status (N) on the reduction of walking within the neighborhood (R), $P(r \mid do(n))$?



Back-door criterion: example 2

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), $P(r \mid do(n))$?

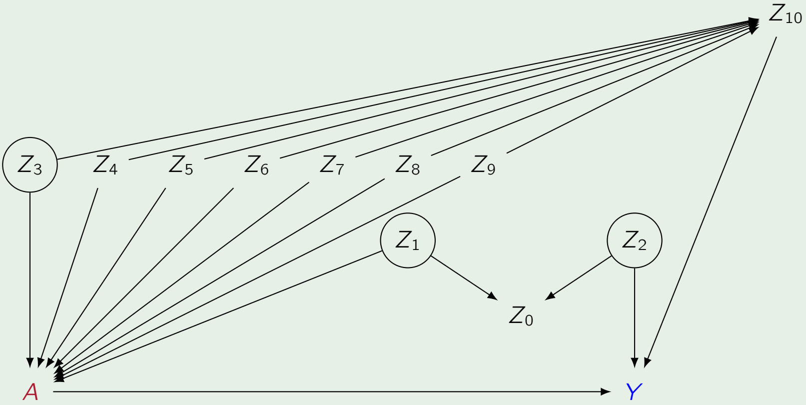


$$P(r \mid do(n)) = \sum P(r|n, l_2, p)P(l_2, p)$$

Back-door criterion: example 3

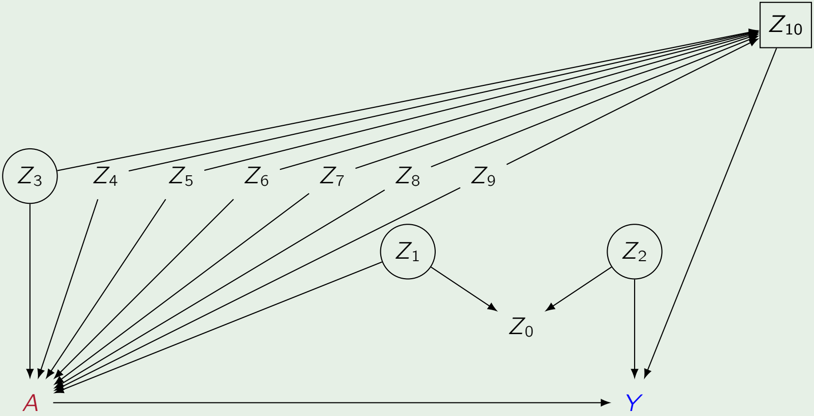
Example

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



Finding the optimal set satisfying the back-door criterion

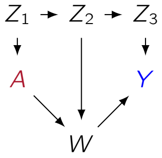
Many sets of variables can satisfy the back-door criterion:

- All these sets successfully remove bias.
- However, in practice, they do not all yield the same variance.

Finding the optimal set satisfying the back-door criterion

Many sets of variables can satisfy the back-door criterion:

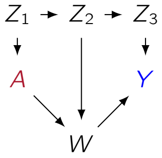
- All these sets successfully remove bias.
- However, in practice, they do not all yield the same variance.

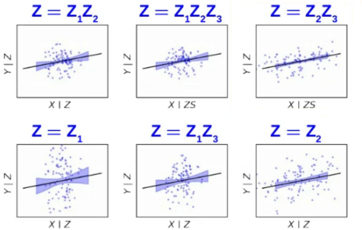


Finding the optimal set satisfying the back-door criterion

Many sets of variables can satisfy the back-door criterion:

- All these sets successfully remove bias.
- However, in practice, they do not all yield the same variance.

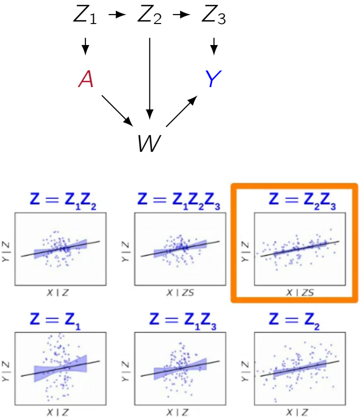




Finding the optimal set satisfying the back-door criterion

Many sets of variables can satisfy the back-door criterion:

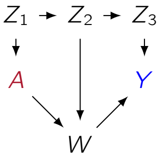
- All these sets successfully remove bias.
- However, in practice, they do not all yield the same variance.

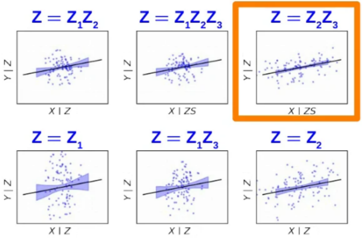


Finding the optimal set satisfying the back-door criterion

Many sets of variables can satisfy the back-door criterion:

- All these sets successfully remove bias.
- However, in practice, they do not all yield the same variance.
- Is it possible to find the optimal set that yields the smallest variance?

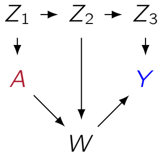




Finding the optimal set satisfying the back-door criterion

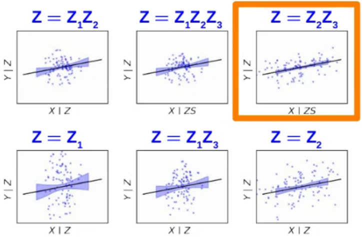
Many sets of variables can satisfy the back-door criterion:

- All these sets successfully remove bias.
- However, in practice, they do not all yield the same variance.
- Is it possible to find the optimal set that yields the smallest variance?



A set $\mathbb{Z}$ satisfying the back-door criterion relative to (A, Y) optimal if:

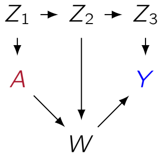
- $\mathbb{Z}$ includes all parents of Y that are not mediators between A and Y, and



Finding the optimal set satisfying the back-door criterion

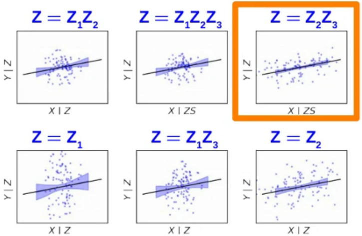
Many sets of variables can satisfy the back-door criterion:

- All these sets successfully remove bias.
- However, in practice, they do not all yield the same variance.
- Is it possible to find the optimal set that yields the smallest variance?



A set $\mathbb{Z}$ satisfying the back-door criterion relative to (A, Y) optimal if:

- $\mathbb{Z}$ includes all parents of Y that are not mediators between A and Y, and
- $\mathbb{Z}$ includes all parents of any mediator between A and Y, that are not themselves mediators between A and Y.



Optimal set in action

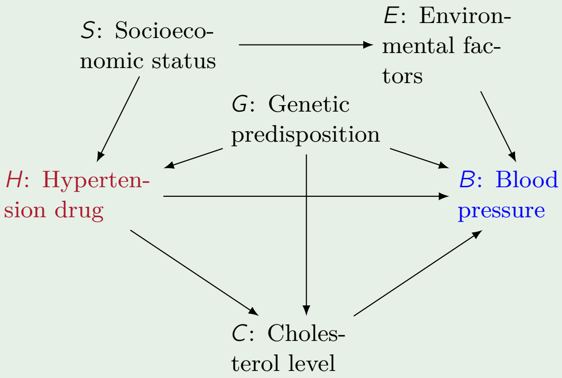
Example

How to optimally estimate the effect of new hypertension drug (H) on blood pressure (B) $P(b \mid do(h))$?

Optimal set in action

Example

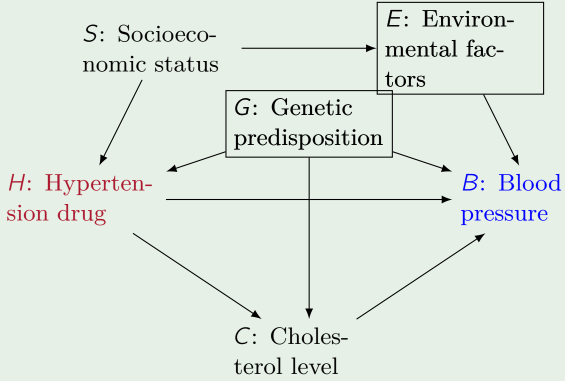
How to optimally estimate the effect of new hypertension drug (H) on blood pressure (B) $P(b \mid do(h))$?



Optimal set in action

Example

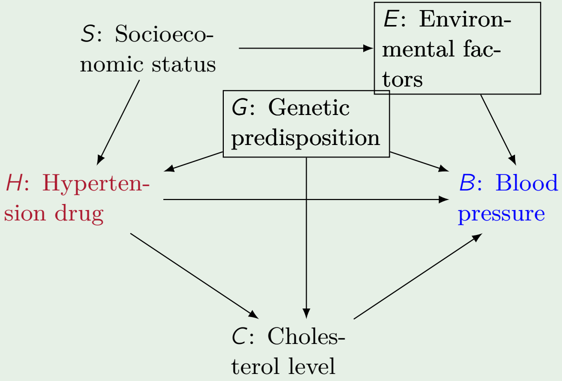
How to optimally estimate the effect of new hypertension drug (H) on blood pressure (B) $P(b \mid do(h))$?



Optimal set in action

Example

How to optimally estimate the effect of new hypertension drug (H) on blood pressure (B) $P(b \mid do(h))$?



$$P(b \mid do(h)) = \sum_{g,e} P(b|h,g,e)P(g,e)$$

Association Criterion 1 for identifying confounding factors (incorrect)

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 1 if each element $Z \in \mathbb{Z}$ meets the following conditions:

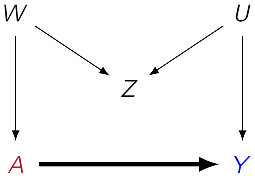
- Z is associated with A; and
- Z is associated with Y, conditionally on A.

Association Criterion 1 for identifying confounding factors (incorrect)

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 1 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

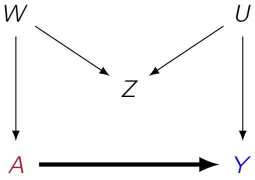


Association Criterion 1 for identifying confounding factors (incorrect)

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 1 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



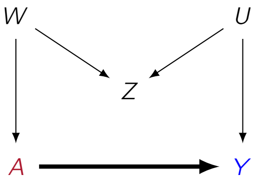
Z is associated with A and Z is associated with Y, conditionally on A

Association Criterion 1 for identifying confounding factors (incorrect)

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 1 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

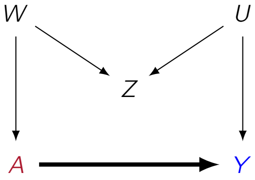
Should we adjust on Z?

Association Criterion 1 for identifying confounding factors (incorrect)

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 1 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

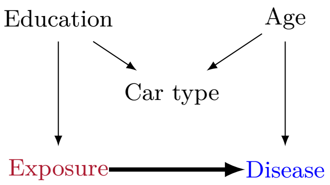
Should we adjust on Z? No!

Association Criterion 1 for identifying confounding factors (incorrect)

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 1 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!

Association Criterion 2 for identifying confounding factors (incorrect)

Let $\mathbb{Z}$ be the set of variables in a problem that are not affected by X. The set $\mathbb{Z}$ satisfies Association Criterion 2 if it can be divided into two subsets, $\mathbb{Z}_1$ and $\mathbb{Z}_2$, which meet the following conditions:

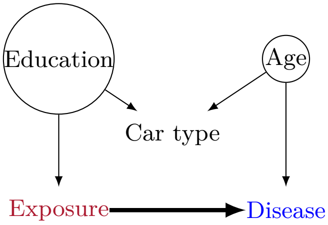
- $\mathbb{Z}_1$ is associated with A; and
- $\mathbb{Z}_2$ is associated with Y, conditionally on A and $\mathbb{Z}_1$.

Association Criterion 2 for identifying confounding factors (incorrect)

Let $\mathbb{Z}$ be the set of variables in a problem that are not affected by X. The set $\mathbb{Z}$ satisfies Association Criterion 2 if it can be divided into two subsets, $\mathbb{Z}_1$ and $\mathbb{Z}_2$, which meet the following conditions:

- $\mathbb{Z}_1$ is associated with A; and
- $\mathbb{Z}_2$ is associated with Y, conditionally on A and $\mathbb{Z}_1$.

Counterexample

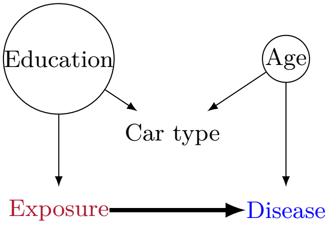


Association Criterion 2 for identifying confounding factors (incorrect)

Let $\mathbb{Z}$ be the set of variables in a problem that are not affected by X. The set $\mathbb{Z}$ satisfies Association Criterion 2 if it can be divided into two subsets, $\mathbb{Z}_1$ and $\mathbb{Z}_2$, which meet the following conditions:

- $\mathbb{Z}_1$ is associated with A; and
- $\mathbb{Z}_2$ is associated with Y, conditionally on A and $\mathbb{Z}_1$.

Counterexample



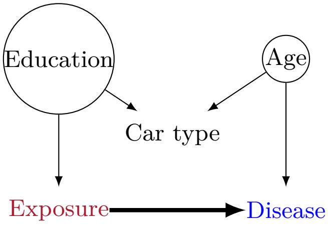
Car type is associated with Exposure

Association Criterion 2 for identifying confounding factors (incorrect)

Let $\mathbb{Z}$ be the set of variables in a problem that are not affected by X. The set $\mathbb{Z}$ satisfies Association Criterion 2 if it can be divided into two subsets, $\mathbb{Z}_1$ and $\mathbb{Z}_2$, which meet the following conditions:

- $\mathbb{Z}_1$ is associated with A; and
- $\mathbb{Z}_2$ is associated with Y, conditionally on A and $\mathbb{Z}_1$.

Counterexample



Car type is associated with Exposure

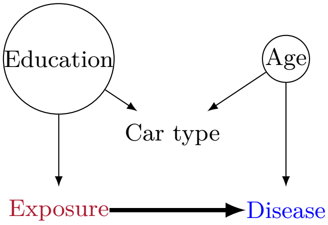
Should we adjust on Car type?

Association Criterion 2 for identifying confounding factors (incorrect)

Let $\mathbb{Z}$ be the set of variables in a problem that are not affected by X. The set $\mathbb{Z}$ satisfies Association Criterion 2 if it can be divided into two subsets, $\mathbb{Z}_1$ and $\mathbb{Z}_2$, which meet the following conditions:

- $\mathbb{Z}_1$ is associated with A; and
- $\mathbb{Z}_2$ is associated with Y, conditionally on A and $\mathbb{Z}_1$.

Counterexample



Car type is associated with Exposure

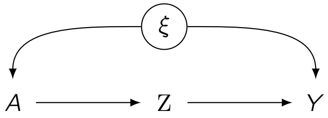
Should we adjust on Car type? No!

Incompleteness of the back-door criterion

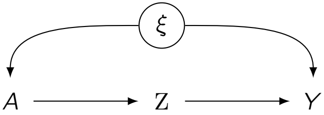
- The back-door criterion is sound but not complete:

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.


ξ
A
Z
Y

Going beyond the back-door



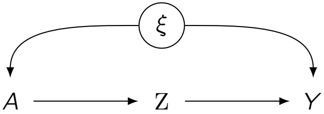
- $P(z \mid do(a)) = P(z \mid a)$ (No back-door path)

Going beyond the back-door

$$
\begin{tikzcd}
 & \xi \arrow[dl, bend right] \arrow[dr, bend left] & \\
A \arrow[r] & Z \arrow[r] & Y
\end{tikzcd}
$$

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

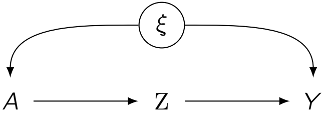
Going beyond the back-door



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

$$P(y \mid \color{red}{do(a)}\color{black}) = \sum_z P(y \mid \color{red}{do(z)}\color{black})P(z \mid \color{red}{do(a)}\color{black})$$

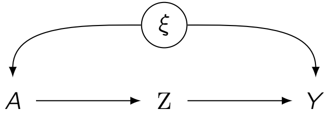
Going beyond the back-door



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

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

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.

Front-door adjustment

Theorem ([3])

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

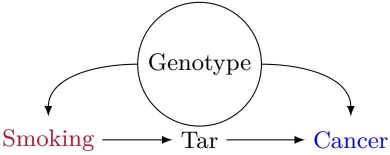
$$P(y \mid do(a)) = \sum_{z} P(z \mid a) \sum_{a'} P(y \mid a', z) P(a').$$

Front-door criterion: example

In a study, we measure the tar and the % of cancer among smokers and non smokers in a randomly selected sample of the population.
What is the effect of smoking on cancer $P(c \mid do(s))$?

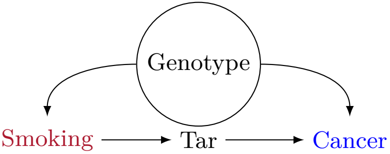
Front-door criterion: example

In a study, we measure the tar and the % of cancer among smokers and non smokers in a randomly selected sample of the population.
What is the effect of smoking on cancer $P(c \mid do(s))$?



Front-door criterion: example

In a study, we measure the tar and the % of cancer among smokers and non smokers in a randomly selected sample of the population.
What is the effect of smoking on cancer $P(c \mid do(s))$?

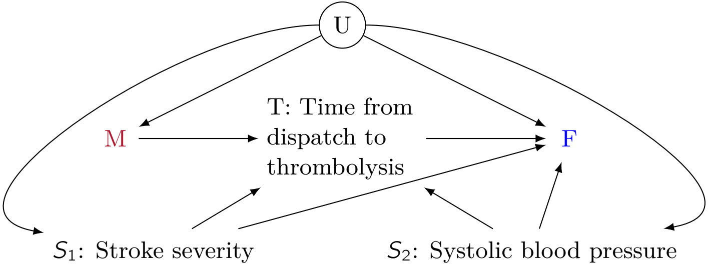


$$P(c \mid do(s)) = \sum_t P(t|s) \sum_{s'} P(c \mid t, s')P(s')$$

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

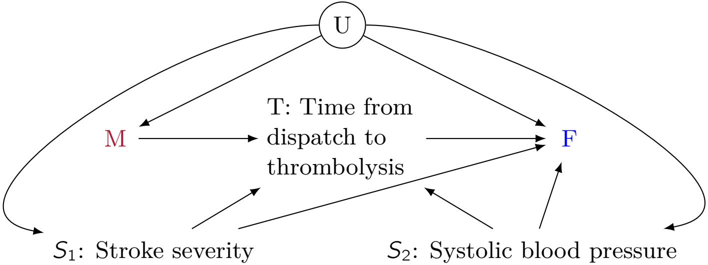
In this study, Piccininni et al. were interested in estimating the effect of Mobile Stroke Unit dispatch (M) on functional outcomes (F). In other words, $P(f \mid do(m))$?

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, $P(f \mid do(m))$?



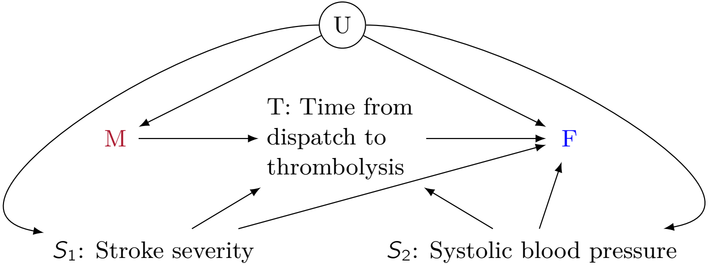
Back-door and Front-door criteria in a real application

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, $P(f \mid do(m))$?



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!

Incompleteness of the front-door criterion

- 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.

A glimpse of the do-calculus [3]

The do calculus consists of three rules :

- Rule 1: generalization of the d-seperation
- Rule 2: generalization of back-door criterion
- Rule 3: complicated

A glimpse of the do-calculus [3]

The do calculus consists of three rules :

- Rule 1: generalization of the d-seperation
- Rule 2: generalization of back-door criterion
- Rule 3: complicated

$P(y \mid do(x))$ is identifiable iff by repeatedly applying the rules of do-calculus, we obtain a do-free expression.

A glimpse of the do-calculus [3]

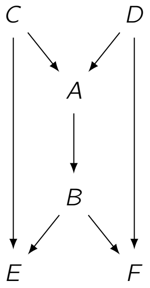
The do calculus consists of three rules :

- Rule 1: generalization of the d-seperation
- Rule 2: generalization of back-door criterion
- Rule 3: complicated

$P(y \mid do(x))$ is identifiable iff by repeatedly applying the rules of do-calculus, we obtain a do-free expression.

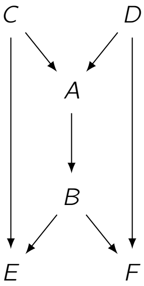
The do calculus is sound and complete!

Mutilated Graphs

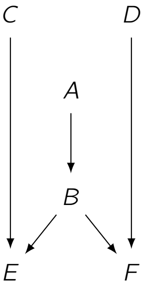


$\mathcal{G}$

Mutilated Graphs

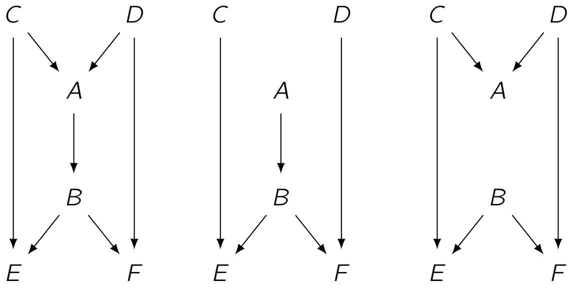


$\mathcal{G}$



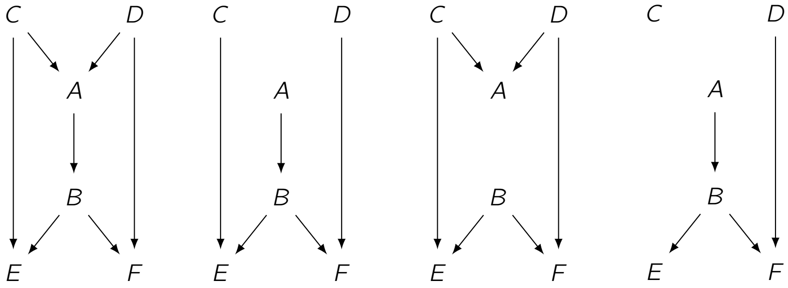
$\mathcal{G}_{\overline{A}}$

Mutilated Graphs



$\mathcal{G}$ $\mathcal{G}_{\overline{A}}$ $\mathcal{G}_{\underline{A}}$

Mutilated Graphs



$\mathcal{G}$ $\quad$ $\mathcal{G}_{\overline{A}}$ $\quad$ $\mathcal{G}_{\underline{A}}$ $\quad$ $\mathcal{G}_{\overline{A}\underline{C}}$

The do-calculus

Theorem

Let $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ be a causal graph. Let $\mathbb{X}, \mathbb{Y}, \mathbb{Z}, \mathbb{W} \subseteq \mathbb{V}$ be disjoint. We have:

$$P(y|do(x), z, w) = P(y|do(x), w) \quad \text{if} \quad \mathbb{Y} \perp\!\!\!\perp_{\mathcal{G}_{\overline{\mathbb{X}}}} \mathbb{Z}|\mathbb{X}, \mathbb{W}$$

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \quad \text{if} \quad \mathbb{Y} \perp\!\!\!\perp_{\mathcal{G}_{\overline{\mathbb{X}}\underline{\mathbb{Z}}}} \mathbb{Z}|\mathbb{X}, \mathbb{W}$$

$$P(y|do(x), do(z), w) = P(y|do(x), w) \quad \text{if} \quad \mathbb{Y} \perp\!\!\!\perp_{\mathcal{G}_{\overline{\mathbb{X}}\overline{\mathbb{Z}(\mathbb{W})}}} \mathbb{Z}|\mathbb{X}, \mathbb{W}$$

where $\mathbb{Z}(\mathbb{W})$ is the set of $\mathbb{Z}$-vertices that are not ancestors of any $\mathbb{W}$-vertex in $\mathcal{G}_{\overline{\mathbb{X}}}$

$$P(y|\cancel{do(x)}, do(z), w) = P(y|\cancel{do(x)}, w) \quad \text{if} \quad (\mathbb{Y} \perp\!\!\!\perp \mathbb{Z}|\cancel{\mathbb{X}}, \mathbb{W})_{\mathcal{G}_{\overline{\cancel{\mathbb{X}}\mathbb{Z}(\mathbb{W})}}}$$

where $\mathbb{Z}(\mathbb{W})$ is the set of $\mathbb{Z}$-vertices that are not ancestors of any $\mathbb{W}$-vertex in $\mathcal{G}_{\overline{\cancel{\mathbb{X}}}}$

Intuition for Rule 3

$$P(y|do(z), w) = P(y|w) \quad \text{if} \quad (\mathbb{Y} \perp\!\!\!\perp \mathbb{Z}|\mathbb{W})_{\mathcal{G}_{\overline{\mathbb{Z}(\mathbb{W})}}}$$

where $\mathbb{Z}(\mathbb{W})$ is the set of $\mathbb{Z}$-vertices that are not ancestors of any $\mathbb{W}$-vertex in $\mathcal{G}$

Intuition for Rule 3

$$P(y|do(z), w) = P(y|w) \quad \text{if} \quad (\mathbb{Y} \perp\!\!\!\perp \mathbb{Z}|\mathbb{W})_{\mathcal{G}_{\overline{\mathbb{Z}(\mathbb{W})}}}$$

where $\mathbb{Z}(\mathbb{W})$ is the set of $\mathbb{Z}$-vertices that are not ancestors of any $\mathbb{W}$-vertex in $\mathcal{G}$

Suppose

$P(y \mid do(z), w_1, w_2)$
$= P(y \mid w_1, w_2)$
if $(Y \perp\!\!\!\perp Z|W_1, W_2)_{\mathcal{G}_{\overline{Z}}}$

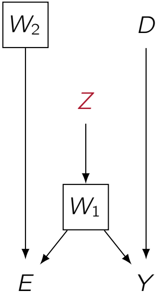
Intuition for Rule 3

$$P(y|do(z), w) = P(y|w) \quad \text{if} \quad (\mathbb{Y} \perp\!\!\!\perp \mathbb{Z}|\mathbb{W})_{\mathcal{G}_{\overline{\mathbb{Z}(\mathbb{W})}}}$$

where $\mathbb{Z}(\mathbb{W})$ is the set of $\mathbb{Z}$-vertices that are not ancestors of any $\mathbb{W}$-vertex in $\mathcal{G}$

Suppose

$P(y \mid do(z), w_1, w_2)$
$= P(y \mid w_1, w_2)$
if $(Y \perp\!\!\!\perp Z|W_1, W_2)_{\mathcal{G}_{\overline{Z}}}$



$\mathcal{G}_{\overline{Z}}$

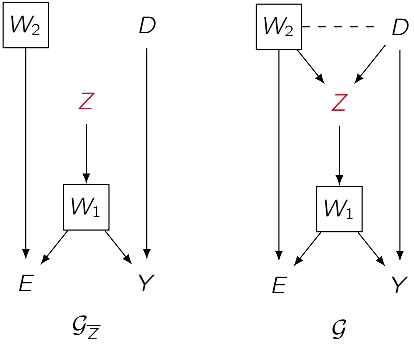
Intuition for Rule 3

$$P(y|do(z), w) = P(y|w) \quad \text{if} \quad (\mathbb{Y} \perp\!\!\!\perp \mathbb{Z}|\mathbb{W})_{\mathcal{G}_{\overline{\mathbb{Z}(\mathbb{W})}}}$$

where $\mathbb{Z}(\mathbb{W})$ is the set of $\mathbb{Z}$-vertices that are not ancestors of any $\mathbb{W}$-vertex in $\mathcal{G}$

Suppose

$P(y \mid do(z), w_1, w_2)$
$= P(y \mid w_1, w_2)$
if $(Y \perp\!\!\!\perp Z|W_1, W_2)_{\mathcal{G}_{\overline{Z}}}$



Intuition for Rule 3

$$P(y|do(z),w) = P(y|w) \quad \text{if} \quad (\mathbb{Y} \perp\!\!\!\perp \mathbb{Z}|\mathbb{W})_{\mathcal{G}_{\overline{\mathbb{Z}(\mathbb{W})}}}$$

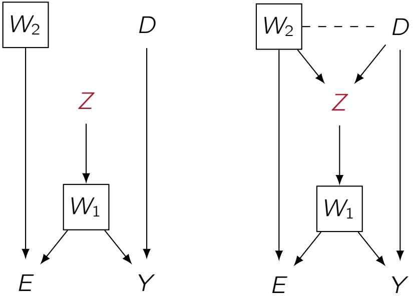
where $\mathbb{Z}(\mathbb{W})$ is the set of $\mathbb{Z}$-vertices that are not ancestors of any $\mathbb{W}$-vertex in $\mathcal{G}$

Suppose

$P(y \mid do(z), w_1, w_2)$

$= P(y \mid w_1, w_2)$

if $(Y \perp\!\!\!\perp Z|W_1, W_2)_{\mathcal{G}_{\overline{Z}}}$



$\mathcal{G}_{\overline{Z}}$ $\mathcal{G}$

$Y \perp\!\!\!\perp_{\mathcal{G}_{\overline{Z}}} W_2 \mid W_1$

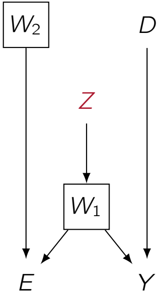
Intuition for Rule 3

$$P(y|\color{red}{do(z)}, w) = P(y|w) \quad \text{if} \quad (\mathbb{Y} \perp\!\!\!\perp \mathbb{Z}|\mathbb{W})_{\mathcal{G}_{\overline{\mathbb{Z}(\mathbb{W})}}}$$

where $\mathbb{Z}(\mathbb{W})$ is the set of $\mathbb{Z}$-vertices that are not ancestors of any $\mathbb{W}$-vertex in $\mathcal{G}$

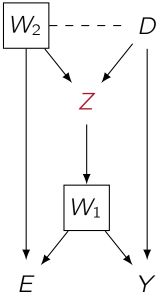
Suppose

$P(y \mid \color{red}{do(z)}, w_1, w_2)$
$= P(y \mid w_1, w_2)$
if $(Y \perp\!\!\!\perp Z|W_1, W_2)_{\mathcal{G}_{\overline{Z}}}$



$\mathcal{G}_{\overline{Z}}$

$Y \perp\!\!\!\perp_{\mathcal{G}_{\overline{Z}}} W_2 \mid W_1$



$\mathcal{G}$

$Y \not\!\perp\!\!\!\perp_{\mathcal{G}_{\overline{Z}}} W_2 \mid W_1$

Intuition for Rule 3

$$P(y|do(z), w) = P(y|w) \quad \text{if} \quad (\mathbb{Y} \perp\!\!\!\perp \mathbb{Z}|\mathbb{W})_{\mathcal{G}_{\overline{\mathbb{Z}(\mathbb{W})}}}$$

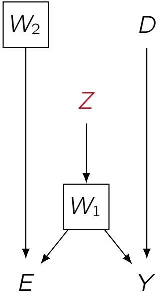
where $\mathbb{Z}(\mathbb{W})$ is the set of $\mathbb{Z}$-vertices that are not ancestors of any $\mathbb{W}$-vertex in $\mathcal{G}$

~~Suppose~~

$\cancel{P(y \mid do(z), w_1, w_2)}$

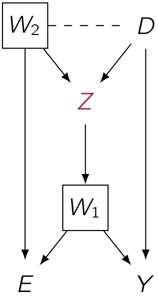
$\cancel{= P(y \mid w_1, w_2)}$

$\cancel{\text{if } (Y \perp\!\!\!\perp Z|W_1, W_2)_{\mathcal{G}_{\overline{Z}}}}$



$\mathcal{G}_{\overline{Z}}$

$Y \perp\!\!\!\perp_{\mathcal{G}_{\overline{Z}}} W_2 \mid W_1$



$\mathcal{G}$

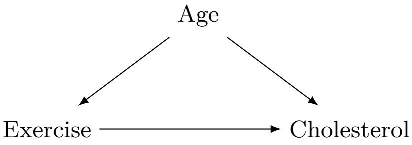
$Y \not\perp\!\!\!\perp_{\mathcal{G}_{\overline{Z}}} W_2 \mid W_1$

Completness of the do-calculus

Theorem ([3, 9])

A causal effect $P(y \mid do(x))$ is identifiable in a model characterized by a graph $\mathcal{G}$ if and only if there exists a finite sequence of transformations, each conforming to one the Rules 1-3, that reduces $P(y \mid do(x))$ into a standard (i.e., "do"-free) probability expression involving observed quantities.

From the do-calculus to the back-door adjustment



What's the effect of exercice on cholesterol?

From the do-calculus to the back-door adjustment

$$\begin{array}{ccc} & \text{Age} & \\ \swarrow & & \searrow \\ \text{Exercise} & \longrightarrow & \text{Cholesterol} \end{array}$$

What's the effect of exercice on cholesterol?

$$P(c \mid do(e)) = \sum_a P(c \mid do(e), a) P(a \mid do(e)) \qquad \text{(Probability Axioms)}$$

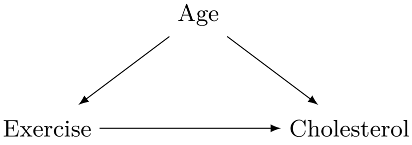
From the do-calculus to the back-door adjustment

$$
\begin{tikzcd}
 & \text{Age} \arrow[dl] \arrow[dr] & \\
\text{Exercise} \arrow[rr] & & \text{Cholesterol}
\end{tikzcd}
$$

What's the effect of exercice on cholesterol?

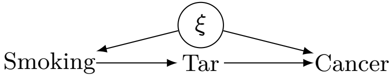
$$
\begin{aligned}
P(c \mid do(e)) &= \sum_a P(c \mid do(e), a) P(a \mid do(e)) && \text{(Probability Axioms)} \\
&= \sum_a P(c \mid e, a) P(a \mid do(e)) && \text{(Rule 2)}
\end{aligned}
$$

From the do-calculus to the back-door adjustment



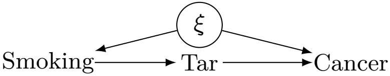
What's the effect of exercice on cholesterol?

$$\begin{aligned} P(c \mid do(e)) &= \sum_a P(c \mid do(e), a) P(a \mid do(e)) && \text{(Probability Axioms)} \\ &= \sum_a P(c \mid e, a) P(a \mid do(e)) && \text{(Rule 2)} \\ &= \sum_a P(c \mid e, a) P(a) && \text{(Rule 3)} \end{aligned}$$



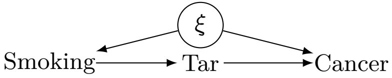
What's the effect of smoking on cancer?

From the do-calculus to the front-door adjustment



What's the effect of smoking on cancer?

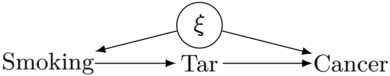
$$P(c \mid do(s)) = \sum_t P(c \mid do(s), t) P(t \mid do(s)) \qquad \text{(Probability Axioms)}$$



What's the effect of smoking on cancer?

$$
\begin{aligned}
P(c \mid do(s)) &= \sum_t P(c \mid do(s), t) P(t \mid do(s)) && \text{(Probability Axioms)} \\
&= \sum_t P(c \mid do(s), do(t)) P(t \mid do(s)) && \text{(Rule 2)}
\end{aligned}
$$

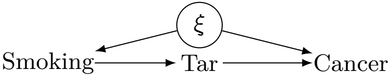
From the do-calculus to the front-door adjustment



What's the effect of smoking on cancer?

$$
\begin{aligned}
P(c \mid do(s)) &= \sum_t P(c \mid do(s), t) P(t \mid do(s)) && \text{(Probability Axioms)} \\
&= \sum_t P(c \mid do(s), do(t)) P(t \mid do(s)) && \text{(Rule 2)} \\
&= \sum_t P(c \mid do(s), do(t)) P(t \mid s) && \text{(Rule 2)}
\end{aligned}
$$

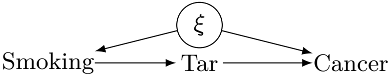
From the do-calculus to the front-door adjustment



What's the effect of smoking on cancer?

$$
\begin{aligned}
P(c \mid do(s)) &= \sum_t P(c \mid do(s), t) P(t \mid do(s)) && \text{(Probability Axioms)} \\
&= \sum_t P(c \mid do(s), do(t)) P(t \mid do(s)) && \text{(Rule 2)} \\
&= \sum_t P(c \mid do(s), do(t)) P(t \mid s) && \text{(Rule 2)} \\
&= \sum_t P(c \mid do(t)) P(t \mid s) && \text{(Rule 3)}
\end{aligned}
$$

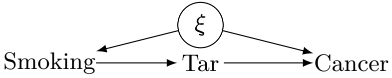
From the do-calculus to the front-door adjustment



What's the effect of smoking on cancer?

$$
\begin{aligned}
P(c \mid do(s)) &= \sum_t P(c \mid do(s), t) P(t \mid do(s)) && \text{(Probability Axioms)} \\
&= \sum_t P(c \mid do(s), do(t)) P(t \mid do(s)) && \text{(Rule 2)} \\
&= \sum_t P(c \mid do(s), do(t)) P(t \mid s) && \text{(Rule 2)} \\
&= \sum_t P(c \mid do(t)) P(t \mid s) && \text{(Rule 3)} \\
&= \sum_{s'} \sum_t P(c \mid do(t), s') P(s' \mid do(t)) P(t \mid s) && \text{(Probability Axioms)}
\end{aligned}
$$

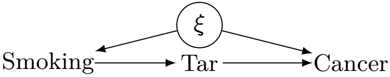
From the do-calculus to the front-door adjustment



What's the effect of smoking on cancer?

$$
\begin{aligned}
P(c \mid do(s)) &= \sum_t P(c \mid do(s), t) P(t \mid do(s)) && \text{(Probability Axioms)} \\
&= \sum_t P(c \mid do(s), do(t)) P(t \mid do(s)) && \text{(Rule 2)} \\
&= \sum_t P(c \mid do(s), do(t)) P(t \mid s) && \text{(Rule 2)} \\
&= \sum_t P(c \mid do(t)) P(t \mid s) && \text{(Rule 3)} \\
&= \sum_{s'} \sum_t P(c \mid do(t), s') P(s' \mid do(t)) P(t \mid s) && \text{(Probability Axioms)} \\
&= \sum_{s'} \sum_t P(c \mid t, s') P(s' \mid do(t)) P(t \mid s) && \text{(Rule 2)}
\end{aligned}
$$

From the do-calculus to the front-door adjustment



What's the effect of smoking on cancer?

$$
\begin{aligned}
P(c \mid do(s)) &= \sum_t P(c \mid do(s), t) P(t \mid do(s)) && \text{(Probability Axioms)} \\
&= \sum_t P(c \mid do(s), do(t)) P(t \mid do(s)) && \text{(Rule 2)} \\
&= \sum_t P(c \mid do(s), do(t)) P(t \mid s) && \text{(Rule 2)} \\
&= \sum_t P(c \mid do(t)) P(t \mid s) && \text{(Rule 3)} \\
&= \sum_{s'} \sum_t P(c \mid do(t), s') P(s' \mid do(t)) P(t \mid s) && \text{(Probability Axioms)} \\
&= \sum_{s'} \sum_t P(c \mid t, s') P(s' \mid do(t)) P(t \mid s) && \text{(Rule 2)} \\
&= \sum_{s'} \sum_t P(c \mid t, s') P(s') P(t \mid s) && \text{(Rule 3)}
\end{aligned}
$$

From a calculus toward an automated algorithm

Limitations of the do-calculus:

- The do-calculus demands a lot of manual labor
- Non-identifiability is complicated

From a calculus toward an automated algorithm

Limitations of the do-calculus:

- The do-calculus demands a lot of manual labor
- Non-identifiability is complicated

Is it possible automatize it?

From a calculus toward an automated algorithm

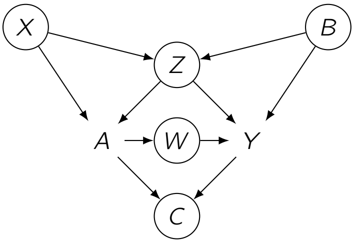
Limitations of the do-calculus:

- The do-calculus demands a lot of manual labor
- Non-identifiability is complicated

Is it possible automatize it? Yes! There exists many algorithms, e.g., the ID algorithm [9].

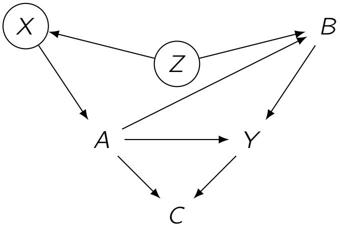
Exercise 4

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 5

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

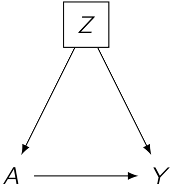


Exercise 6

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

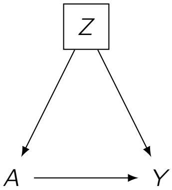
Exercise 6

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



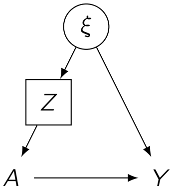
Exercise 6

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



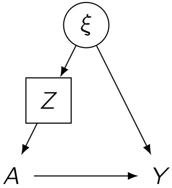
- Z blocks all back-door paths and it is not a descendant of A $\implies$ $\{Z\}$ is a good adjustment set.

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



Exercise 6

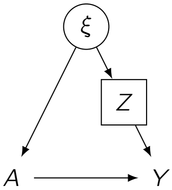
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



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

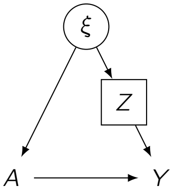
Exercise 6

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



Exercise 6

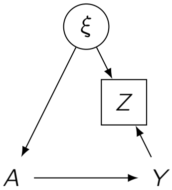
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



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

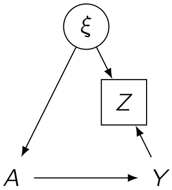
Exercise 6

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



Exercise 6

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.

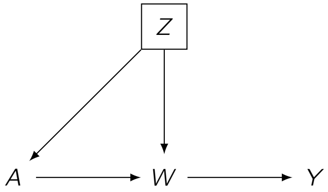
Exercise 6

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

$$\begin{tikzcd}
 & \boxed{Z} \arrow[dl] \arrow[d] & \\
A \arrow[r] & W \arrow[r] & Y
\end{tikzcd}$$

Exercise 6

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 6

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

$$A \rightarrow \boxed{Z} \rightarrow Y$$

Exercise 6

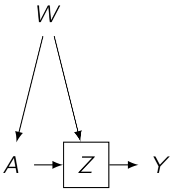
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?

$$A \rightarrow \boxed{Z} \rightarrow Y$$

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

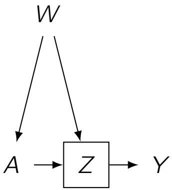
Exercise 6

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



Exercise 6

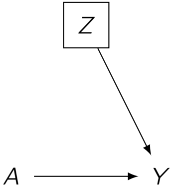
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



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

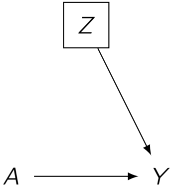
Exercise 6

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



Exercise 6

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



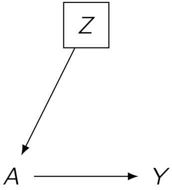
- Z does not open any back-door paths from A to Y $\implies$ $\{Z\}$ is a neutral adjustment set;

- Adjusting on Z can reduces the variation of Y, and helps improve the precision of the estimate in finite samples.

SANTÉ SORBONNE UNIVERSITÉ

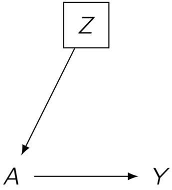
Exercise 6

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



Exercise 6

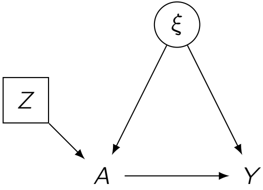
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



- Z does not open any back-door paths from A to Y
  $\implies \{Z\}$ is a neutral adjustment set;

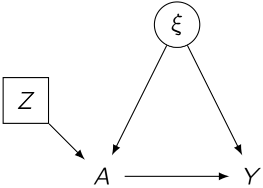
- Adjusting on Z can reduces the variation of A and so may hurt the precision of the estimate in finite samples.

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



Exercise 6

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



- Z does not block existing back-door path from A to Y
  $\implies \{Z\}$ is a bad adjustment set;

- In linear models, adjusting on Z amplify any existing bias.

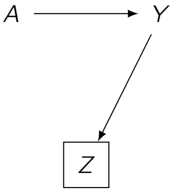
Exercise 6

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

$$\begin{tikzcd} A \arrow[r] & Y \arrow[dl] \\ \boxed{Z} & \end{tikzcd}$$

Exercise 6

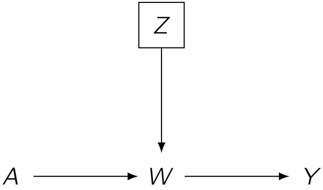
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(a))$?



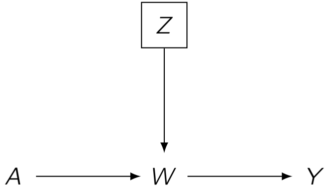
- Selection bias
  $\implies$ $\{Z\}$ is a bad control.

Exercise 6

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

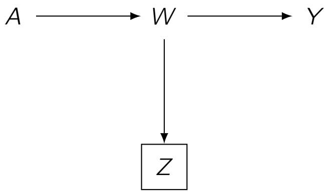


- Z does not open any back-door paths from A to Y $\implies$ $\{Z\}$ is a neutral adjustment set;

- Adjusting on Z can reduces the variation of W, and helps improve the precision of the estimate in finite samples.

Exercise 6

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



Exercise 6

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

$$\begin{tikzcd} A \arrow[r] & W \arrow[r] \arrow[d] & Y \\ & \boxed{Z} & \end{tikzcd}$$

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

SANTÉ SORBONNE UNIVERSITÉ

4

Counterfactuals

- Structural causal models
- Counterfactual reasoning
- Mediation analysis
- Difference with the potential outcome framework

4

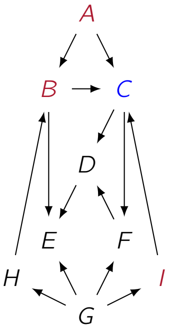
Counterfactuals

- Structural causal models
- Counterfactual reasoning
- Mediation analysis
- Difference with the potential outcome framework

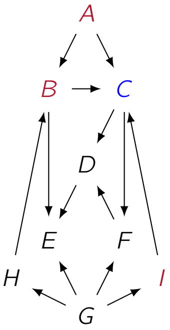
Pure imagination vs constructive imagination

https://www.youtube.com/embed/0lpY0Kt4bn8?rel=0&autoplay=1

Causal mechanism

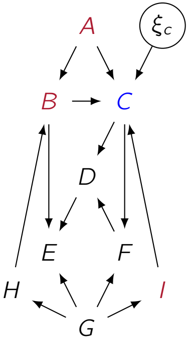


Causal mechanism



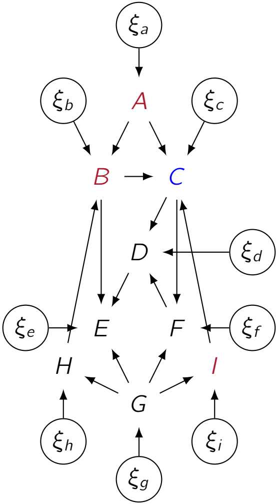
$$C = f(A, B, I)$$

Causal mechanism



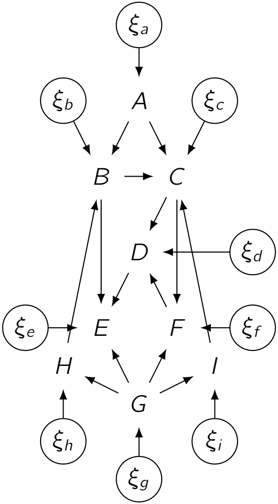
$$C := f_c(A, B, I, \xi_c)$$

Causal mechanism

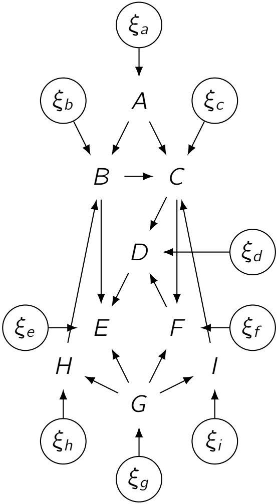


$$C := f_c(A, B, I, \xi_c)$$

Structural causal model

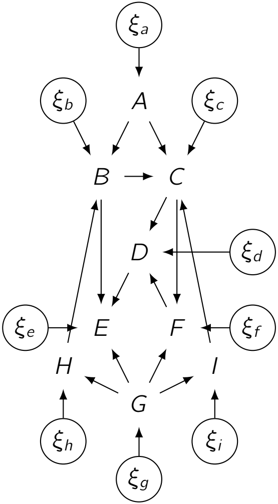


Structural causal model



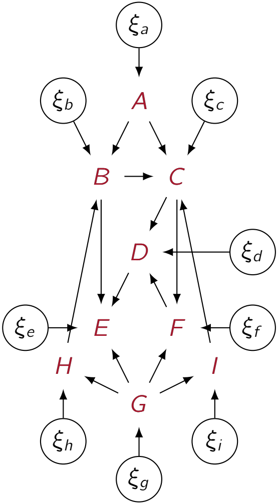
$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

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:

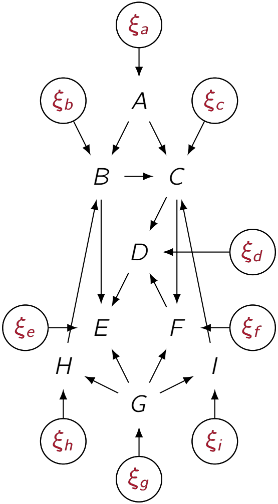


$$
M : \begin{cases}
A := f_a(\xi_a) \\
B := f_b(A, H, \xi_b) \\
C := f_c(A, B, I, \xi_c) \\
D := f_d(C, F, \xi_d) \\
E := f_e(B, G, \xi_e) \\
F := f_f(C, G, \xi_f) \\
G := f_g(\xi_g) \\
H := f_h(G, \xi_h) \\
I := f_i(G, \xi_i)
\end{cases}
$$

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

- Endogenous variables

Structural causal model

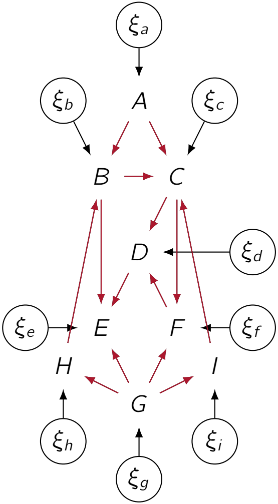


$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

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

Interventions in a structural causal model

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

Interventions in a structural causal model

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

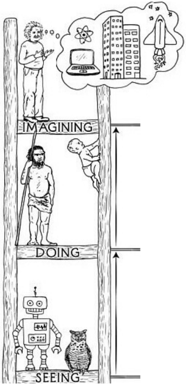
Interventions in a structural causal model

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

4

Counterfactuals

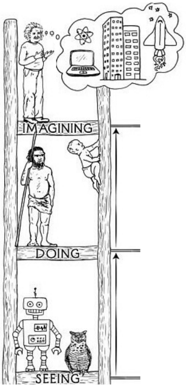
SCM



Counterfactual reasoning

Observational data

SCM



Counterfactual reasoning

Observational data

SCMs naturally helps us in computing counterfactual queries:

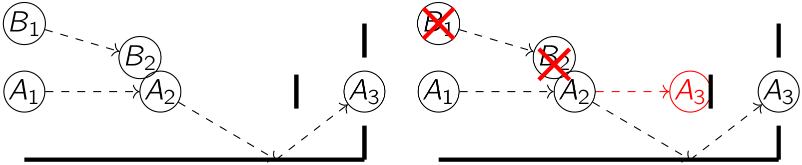
$$P(Y_{A=a} \mid A = a_0, Y = y_0)$$

1. Abduction: use the observations to determine the value of ξ_y
2. Action: modify the model M by removing the structural equations for the variable A and replacing it with the appropriate functions $A = a$, to obtain the interventional SCM M_a
3. Prediction: use M_a and the value of ξ_y to compute the value of $Y_{A=a}$

If $Y_{A=a} \neq y_0$ then a_0 is an actual cause of y_0

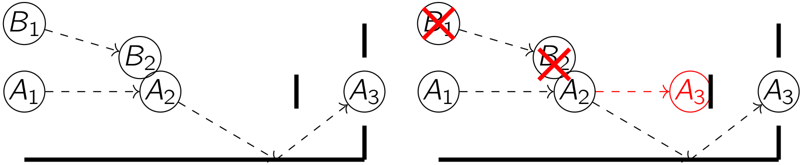
Factual world

Conterfactual world



Factual world

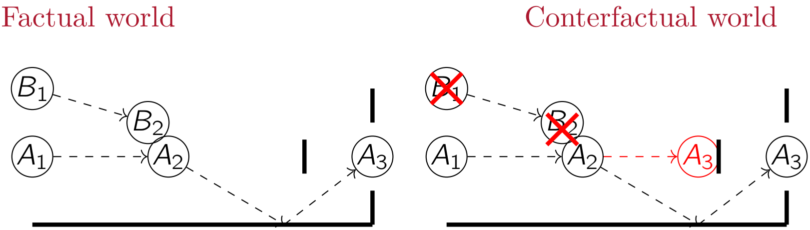
Conterfactual world



Toy example (without noise on C, W, and G):

$$M : \begin{cases} H \quad (B \text{ hits A}) := \xi_h \\ C \quad (A \text{ changes tragectory}) := H \\ W \quad (A \text{ hits the wall}) := 1 - C \\ G \quad (A \text{ reaches the goal}) := 1 - W \end{cases}$$

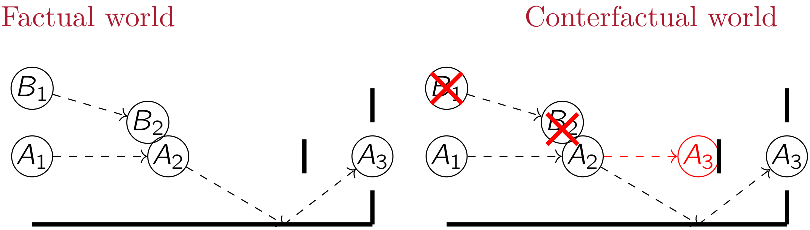
Computing counterfactuals using SCMs: example



Toy example (without noise on C, W, and G):

$$M : \begin{cases} H \quad (B \text{ hits A}) := \xi_h \\ C \quad (A \text{ changes tragectory}) := H \\ W \quad (A \text{ hits the wall}) := 1 - C \\ G \quad (A \text{ reaches the goal}) := 1 - W \end{cases} \begin{cases} H := 1 \\ C := 1 \\ W := 0 \\ G := 1 \end{cases}$$

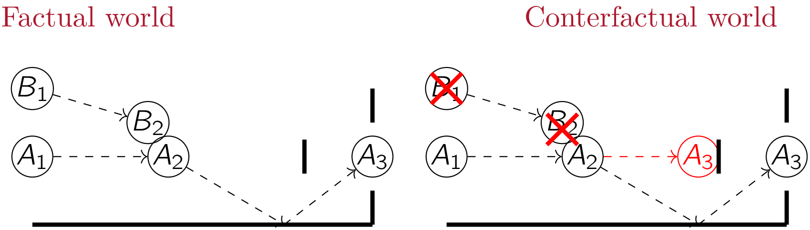
Computing counterfactuals using SCMs: example



Toy example (without noise on C, W, and G):

$$M : \begin{cases} H \quad (B \text{ hits A}) := \xi_h \\ C \quad (A \text{ changes tragectory}) := H \\ W \quad (A \text{ hits the wall}) := 1 - C \\ G \quad (A \text{ reaches the goal}) := 1 - W \end{cases} \begin{cases} H := 1 \\ C := 1 \\ W := 0 \\ G := 1 \end{cases} \qquad M_{H=0} : \begin{cases} H := 0 \\ C := 0 \\ W := 1 \\ G := 0 \end{cases}$$

Computing counterfactuals using SCMs: example



Toy example (without noise on C, W, and G):

$$M : \begin{cases} H & (B \text{ hits A}) := \xi_h \\ C & (A \text{ changes tragectory}) := H \\ W & (A \text{ hits the wall}) := 1 - C \\ G & (A \text{ reaches the goal}) := 1 - W \end{cases} \begin{cases} H := 1 \\ C := 1 \\ W := 0 \\ G := 1 \end{cases} \qquad M_{H=0} : \begin{cases} H := 0 \\ C := 0 \\ W := 1 \\ G := 0 \end{cases}$$

$G_{H=0} \neq 1 \implies B$ causes A to reach the goal

Computing counterfactuals using causal DAGs

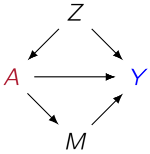
There are methods for computing counterfactual queries directly from a causal DAG [9]; however, these are beyond the scope of this tutorial.

We will focus on a specific class of counterfactual queries that are particularly relevant for mediation analysis.

4

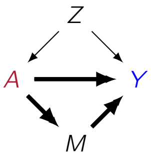
Counterfactuals

There exists three main types of causal effects:



There exists three main types of causal effects:

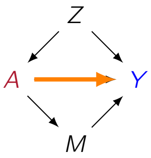
- Total effect



Different types of causal effects

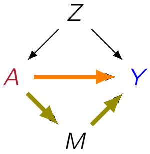
There exists three main types of causal effects:

- Total effect
- Direct effect



There exists three main types of causal effects:

- Total effect
- Direct effect
- Indirect effect



Natural direct effect (NDE)

Natural direct effect (NDE):

$$NDE(x, x'; Y) = \mathbb{E}(y \mid do(x, pa^{x'}_{Y \setminus X})) - \mathbb{E}(y \mid do(x', pa^{x'}_{Y \setminus X}))$$

where $pa^{x'}_{Y \setminus X}$ is the value of $Pa(Y) \setminus \{X\}$ when $X = x'$.

Natural direct effect (NDE):

$$NDE(x, x'; Y) = \mathbb{E}(y \mid do(x, pa^{x'}_{Y\setminus X})) - \mathbb{E}(y \mid do(x', pa^{x'}_{Y\setminus X}))$$

where $pa^{x'}_{Y\setminus X}$ is the value of $Pa(Y)\setminus\{X\}$ when $X = x'$.

Advantage:

- It allows for the complete decomposition of the total effect into the NDE and NIE

Natural direct effect (NDE):

$$NDE(x, x'; Y) = \mathbb{E}(y \mid do(x, pa^{x'}_{Y \setminus X})) - \mathbb{E}(y \mid do(x', pa^{x'}_{Y \setminus X}))$$

where $pa^{x'}_{Y \setminus X}$ is the value of $Pa(Y) \setminus \{X\}$ when $X = x'$.

Advantage:

- It allows for the complete decomposition of the total effect into the NDE and NIE

Limitation:

- NDE cannot always be measured via experiments since it is a conterfactual expression

Natural indirect effect (NIE)

$$TE = NDE - NIE_r$$

where NIE_r stands for the reverse transition, from $X = 1$ to $X = 0$; it becomes additive in linear systems, where reversal of transitions entails sign reversal.

Natural indirect effect (NIE)

$$TE = NDE - NIE_r$$

where NIE_r stands for the reverse transition, from $X = 1$ to $X = 0$; it becomes additive in linear systems, where reversal of transitions entails sign reversal.

This implies that NIE is identifiable whenever NDE and the total effect are identifiable

$$TE = NDE - NIE_r$$

where NIE_r stands for the reverse transition, from $X = 1$ to $X = 0$; it becomes additive in linear systems, where reversal of transitions entails sign reversal.

This implies that NIE is identifiable whenever NDE and the total effect are identifiable

Natural indirect effect (NIE):

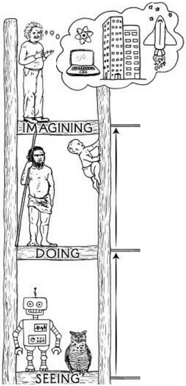
$$NIE(x, x'; Y) = \mathbb{E}(y \mid do(x, pa^{x}_{Y\setminus X})) - \mathbb{E}(y \mid do(x, pa^{x'}_{Y\setminus X}))$$

where $pa^{x}_{Y\setminus X}$ is the value of $Pa(Y)\setminus\{X\}$ when $X = x$.

Linearity

Causal DAG

Observational data



Mediation analysis

Linear structural causal models

A linear structural causal model consists of a set of structural equations of the form:

$$y := \sum_{x \in Pa(y)} \alpha_{xy} x + \xi_y$$

where $Pa(y)$ are direct causes of y, ξ_y represent errors due to ommitted factors and α_{xy} which are known as a structural coefficient represents the strength of the causal relation.

Linear structural causal models

A linear structural causal model consists of a set of structural equations of the form:

$$y := \sum_{x \in Pa(y)} \alpha_{xy} x + \xi_y$$

where $Pa(y)$ are direct causes of y, ξ_y represent errors due to ommitted factors and α_{xy} which are known as a structural coefficient represents the strength of the causal relation.

α_{xy} is also known as the direct effect of X on Y.

Linear structural causal models

A linear structural causal model consists of a set of structural equations of the form:

$$y := \sum_{x \in Pa(y)} \alpha_{xy} x + \xi_y$$

where $Pa(y)$ are direct causes of y, ξ_y represent errors due to ommitted factors and α_{xy} which are known as a structural coefficient represents the strength of the causal relation.

α_{xy} is also known as the direct effect of X on Y.

The direct effect is not necessarily equivalent to a regression coefficient.

Consider a causal DAG $\mathcal{G}$ and a direct effect α_{xy} associated with link $X \to Y$. Let $\mathcal{G}_{\alpha_{xy}}$ denote the graph that results when $X \to Y$ is deleted from $\mathcal{G}$. A set of variables $\mathbb{Z}$ satisfies the single-door criterion iff:

- $\mathbb{Z}$ contains no descendant of Y;
- $\mathbb{Z}$ d-separates X from Y in $\mathcal{G}_{\alpha_{xy}}$.

Single-door adjustment [11]

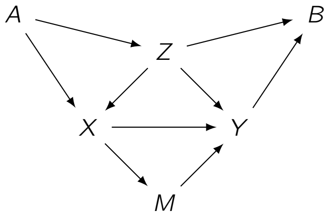
Theorem

If $\mathbb{Z}$ satisfies the single-door criterion relative to (X, Y), then the direct effect α_{xy} is identifiable and is given by

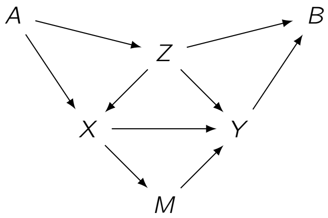
$$\alpha_{xy} = r_{XY.Z}$$

where $r_{XY.\mathbb{Z}}$ represents the residual correlation between Y and X after $\mathbb{Z}$ is "partialled out".

Single-door in action

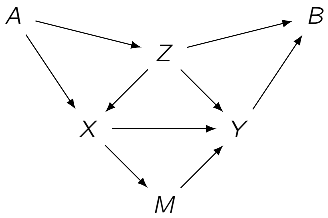


Does the following set satisfy the single-door criterion?



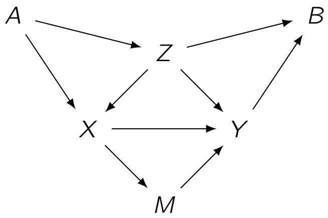
Does the following set satisfy the single-door criterion?

$\{Z\}$?



Does the following set satisfy the single-door criterion?

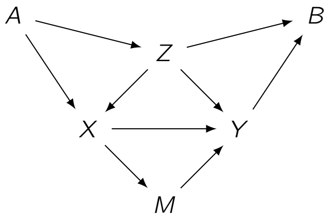
$\{Z\}$? No



Does the following set satisfy the single-door criterion?

$\{Z\}$? No
$\{A\}$?

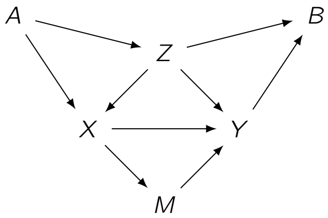
Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No

$\{A\}$? No



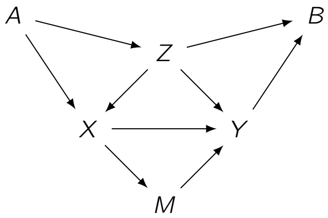
Does the following set satisfy the single-door criterion?

$\{Z\}$? No

$\{A\}$? No

$\{B\}$?

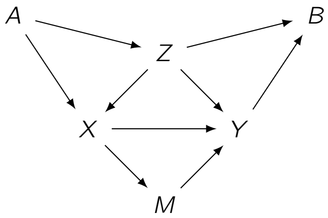
Single-door in action



Does the following set satisfy the single-door criterion?

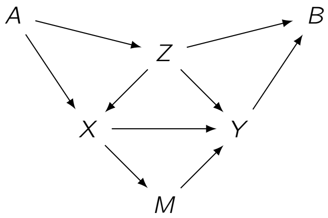
$\{Z\}$? No
$\{A\}$? No
$\{B\}$? No

Single-door in action



Does the following set satisfy the single-door criterion?

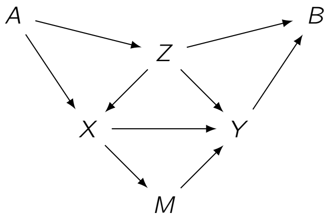
$\{Z\}$? No
$\{A\}$? No
$\{B\}$? No
$\{M\}$?



Does the following set satisfy the single-door criterion?

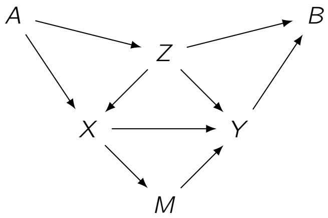
$\{Z\}$? No
$\{A\}$? No
$\{B\}$? No
$\{M\}$? No

Single-door in action



Does the following set satisfy the single-door criterion?

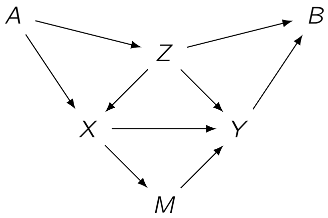
$\{Z\}$? No
$\{A\}$? No
$\{B\}$? No
$\{M\}$? No
$\{A, B\}$?



Does the following set satisfy the single-door criterion?

$\{Z\}$? No
$\{A\}$? No
$\{B\}$? No
$\{M\}$? No
$\{A, B\}$? No

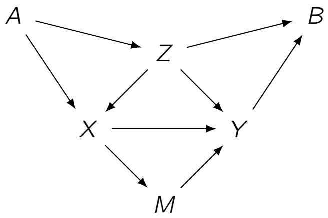
Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No
$\{A\}$? No
$\{B\}$? No
$\{M\}$? No
$\{A, B\}$? No
$\{Z, M\}$?

Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No

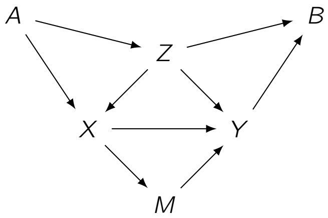
$\{A\}$? No

$\{B\}$? No

$\{M\}$? No

$\{A, B\}$? No

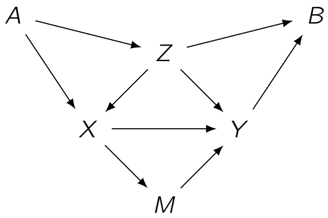
$\{Z, M\}$? Yes



Does the following set satisfy the single-door criterion?

$\{Z\}$? No
$\{A\}$? No
$\{B\}$? No
$\{M\}$? No
$\{A, B\}$? No
$\{Z, M\}$? Yes
$\{Z, M, B\}$?

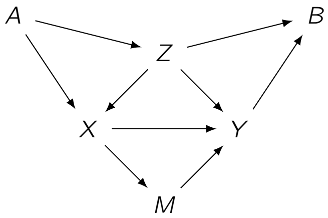
Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No
$\{A\}$? No
$\{B\}$? No
$\{M\}$? No
$\{A, B\}$? No
$\{Z, M\}$? Yes
$\{Z, M, B\}$? No

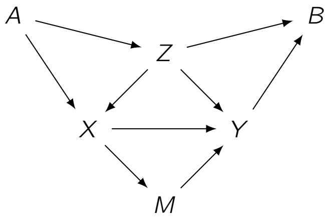
Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No
$\{A\}$? No
$\{B\}$? No
$\{M\}$? No
$\{A, B\}$? No
$\{Z, M\}$? Yes
$\{Z, M, B\}$? No
$\{Z, M, A\}$?

Single-door in action



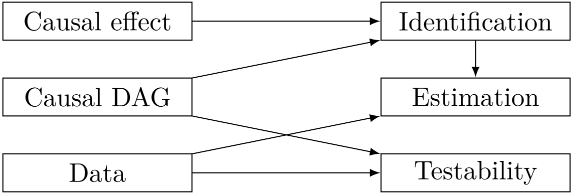
Does the following set satisfy the single-door criterion?

$\{Z\}$? No
$\{A\}$? No
$\{B\}$? No
$\{M\}$? No
$\{A, B\}$? No
$\{Z, M\}$? Yes
$\{Z, M, B\}$? No
$\{Z, M, A\}$? Yes

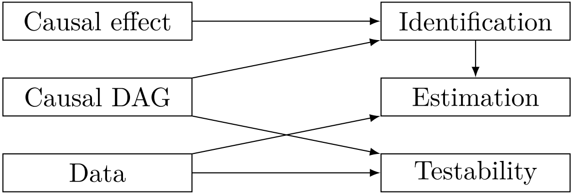
4

Counterfactuals

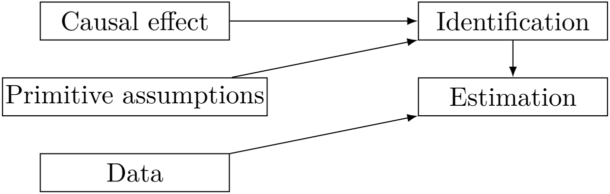
Structural causal models



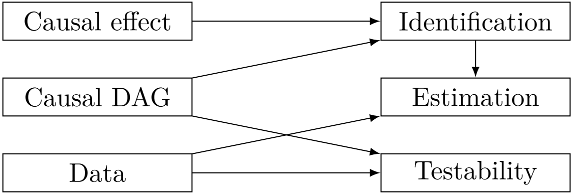
Structural causal models



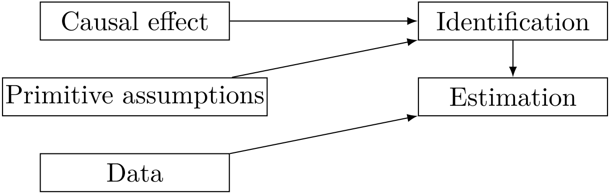
Potential outcome



Structural causal models



Potential outcome



In the structural causal models framework, the primitive assumptions becomes theorems!

SANTÉ SORBONNE UNIVERSITÉ

5

Conclusion

Conclusion

Counterfactual data

Experimental data

Observational data

SCMs
counterfactual reasoning
mediation analysis

Causal DAG
causal reasoning

Bayesian networks
prediction

6

References

References I

[1] Harri Helajärvi et al. "Exploring Causality between TV Viewing and Weight Change in Young and Middle-Aged Adults. The Cardiovascular Risk in Young Finns Study". In: PloS one 9 (July 2014), e101860. doi: 10.1371/journal.pone.0101860.

[2] M.A. Hernan and J.M. Robins. Causal Inference: What If. Chapman & Hall/CRC Monographs on Statistics & Applied Probab. CRC Press, 2023. isbn: 9781420076165.

[3] 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.

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

[5] Judea Pearl. "Direct and Indirect Effects". In: Probabilistic and Causal Inference (2001).

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

[7] Richard Scheines. “A Woefully Incomplete History of Empirical Applications of Algorithmic Causal Discovery”. The History and Development of Search Methods for Causal Structure Workshop at the 39th Conference on Uncertainty in Artificial Intelligence. 2023.

[8] Shohei Shimizu et al. “A Linear Non-Gaussian Acyclic Model for Causal Discovery”. In: Journal of Machine Learning Research 7.72 (2006), pp. 2003–2030.

References III

[9] Ilya Shpitser and Judea Pearl. “Complete Identification Methods for the Causal Hierarchy”. In: Journal of Machine Learning Research 9.64 (2008), pp. 1941–1979. url: http://jmlr.org/papers/v9/shpitser08a.html.

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

[11] PETER SPIRTES et al. “Using Path Diagrams as a Structural Equation Modeling Tool”. In: Sociological Methods & Research 27.2 (1998), pp. 182–225. doi: 10.1177/0049124198027002003.

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

MERCI POUR VOTRE ATTENTION