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 2024

SANTÉ SORBONNE UNIVERSITÉ

Sommaire

1

Association

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

SANTÉ SORBONNE UNIVERSITÉ

1 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: $\not\!\perp\!\!\!\perp_P$
- $P(Y = y \mid X = x) \equiv P(y \mid x)$

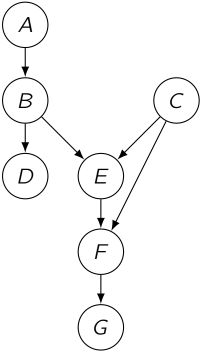
Directed graphs

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

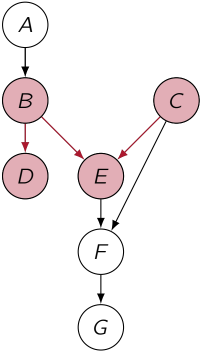
- $\mathbb{V}$ is the set of vertices (usually each vertex corresponds to a random variable),
- $\mathbb{E}$ is the set of edges,
- $\forall (X, Y) \in \mathbb{E},$ there is an arrow pointing from X to Y.

Directed graphs: basic concepts

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

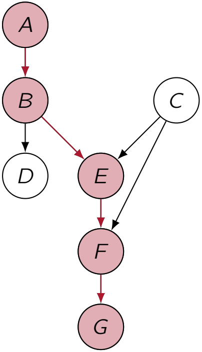


Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



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

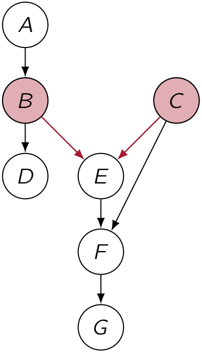
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



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

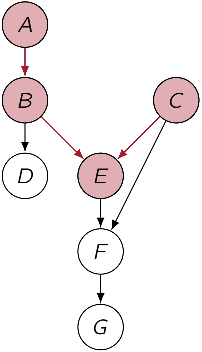
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:

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



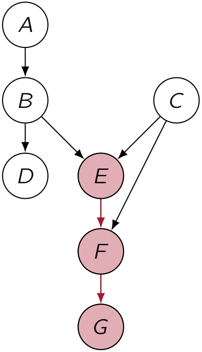
Directed graphs: basic concepts

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



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

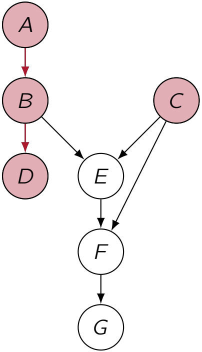
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



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

Directed graphs: basic concepts

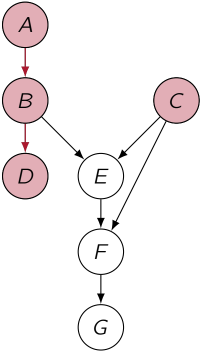
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



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: $Desc(E) = \{E, F, G\}$
Non-descendants: $Nd(E) = \{A, B, C, D\}$

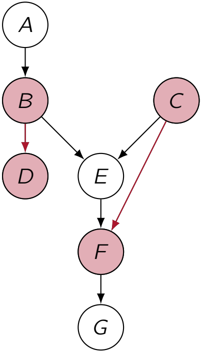
Directed graphs: basic concepts

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

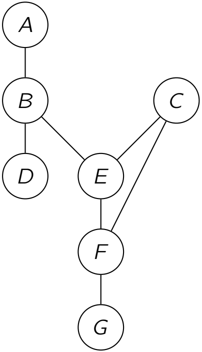
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



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

DAGs

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

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

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

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

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)
generalization of the common cause principle

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)
generalization of the common cause principle

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

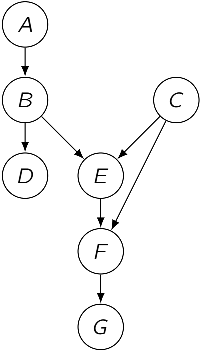
Theorem

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

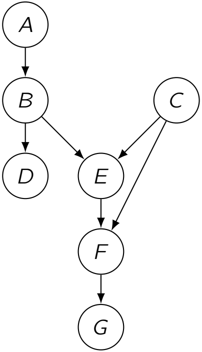
If P is compatible with $\mathcal{G}$ and $\mathbb{S} \subseteq \mathbb{V}$ is an ancestral set, then $P(\mathbb{S})$ is compatible with $\mathcal{G}[\mathbb{S}]$.



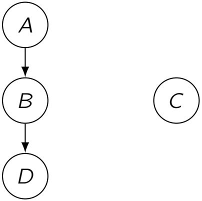
Bayesian network compatible with P

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

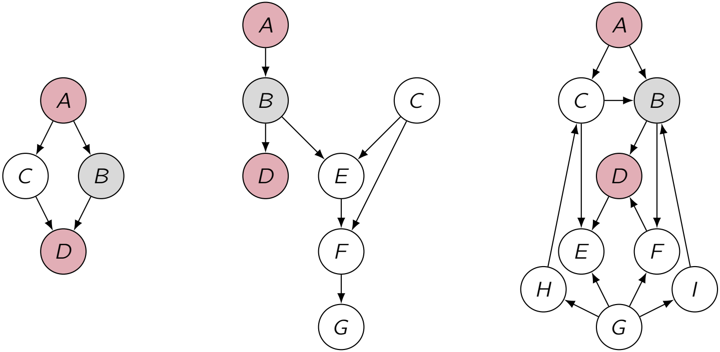
SANTÉ SORBONNE UNIVERSITÉ

1

Association

Reading conditional independencies in graphs

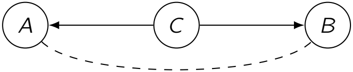
$$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$

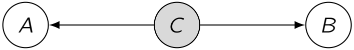


Basic structures

Basic structures

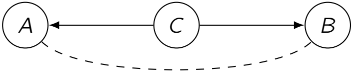
Fork (common cause principle)

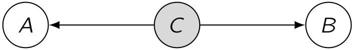




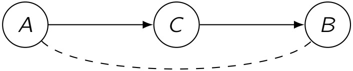
Basic structures

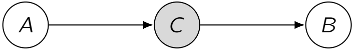
Fork (common cause principle)





Chain





Basic structures

Fork (common cause principle)

Chain

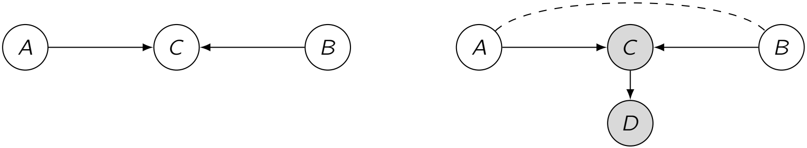
V-structure

Basic structures

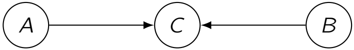
Fork (common cause principle)

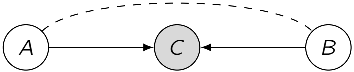
Chain

V-structure

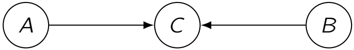


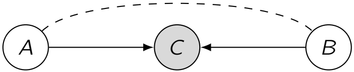
Artificial correlation in V-structures





Artificial correlation in V-structures





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

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

Artificial correlation in V-structures

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

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

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

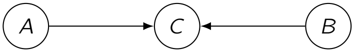
$$\begin{cases} \text{If } A = \text{Mother not carrier then } B = \text{Father carrier} \\ \text{If } B = \text{Father not carrier then } A = \text{Mother carrier} \end{cases}$$

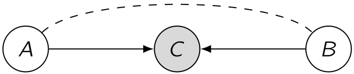
Artificial correlation in V-structures

$$A = \begin{cases} \text{Asthmatic} \\ \text{Not asthmatic} \end{cases} \qquad B = \begin{cases} \text{Sleep apnea} \\ \text{No sleep apnea} \end{cases}$$

$$C = (A \text{ or } B) = \begin{cases} \text{Visit pulmonologist} \\ \text{No visit pulmonologist} \end{cases}$$

Artificial correlation in V-structures





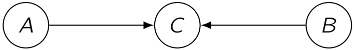
$$A = \begin{cases} \text{Asthmatic} \\ \text{Not asthmatic} \end{cases} \qquad B = \begin{cases} \text{Sleep apnea} \\ \text{No sleep apnea} \end{cases}$$

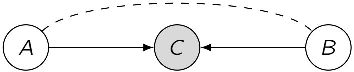
$$C = (A \text{ or } B) = \begin{cases} \text{Visit pulmonologist} \\ \text{No visit pulmonologist} \end{cases}$$

If $C = \text{Visit pulmonologist} \implies$

$$\begin{cases} \text{If } A = \text{Not asthmatic then } B = \text{Sleep apnea} \\ \text{If } B = \text{No sleep apnea then } A = \text{Asthmatic} \end{cases}$$

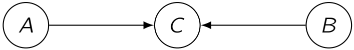
Artificial correlation in V-structures

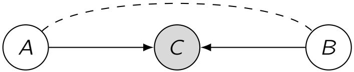




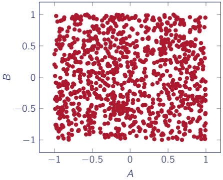
$A, B \sim U(-1, 1)$ $\qquad$ $\xi_c \sim N(0, \frac{1}{2})$ $\qquad$ $C = 2AB + \xi_c$

Artificial correlation in V-structures



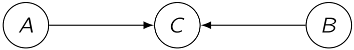


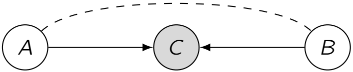
$A, B \sim U(-1, 1)$ $\xi_c \sim N(0, \frac{1}{2})$ $C = 2AB + \xi_c$



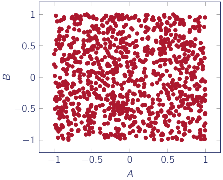
$Corr(A; B) = 0.002$

Artificial correlation in V-structures

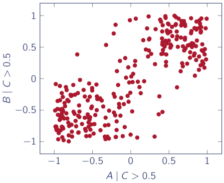




$A, B \sim U(-1, 1)$ $\xi_c \sim N(0, \frac{1}{2})$ $C = 2AB + \xi_c$



$Corr(A; B) = 0.002$



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

Blocked paths

A triple such that $X \to Z \leftarrow Y$ is called a collider[1] . If the two parent vertices are not adjacent, the collider is a v-structure.

[1] We also refer to Z as the collider

Blocked paths

A triple such that $X \rightarrow Z \leftarrow Y$ is called a collider[1] . If the two parent vertices are not adjacent, the collider is a v-structure.

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 $\mathbb{Z}$.

[1] We also refer to Z as the collider

Blocked paths

A triple such that $X \rightarrow Z \leftarrow Y$ is called a collider[1] . If the two parent vertices are not adjacent, the collider is a v-structure.

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.

[1] We also refer to Z as the collider

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

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

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

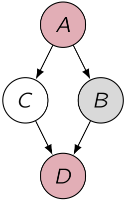
$\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z} \Rightarrow \mathbb{X} \perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$

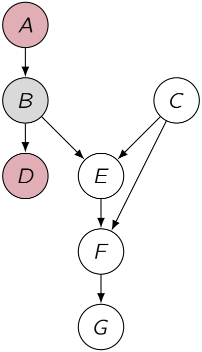
Given disjoint sets $\mathbb{X}, \mathbb{Y}, \mathbb{Z} \subseteq \mathbb{V}$, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-separated by $\mathbb{Z}$ if every path between a vertex in $\mathbb{X}$ and a vertex in $\mathbb{Y}$ is blocked by $\mathbb{Z}$ and we write $\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}}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

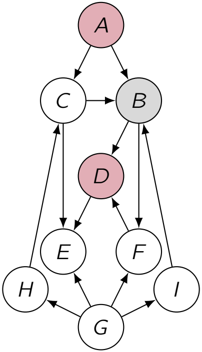
If one path is not blocked, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-connected given $\mathbb{Z}$ and we write $\mathbb{X} \not\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$.

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

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

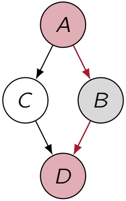


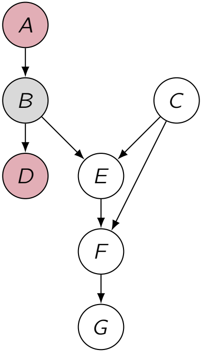


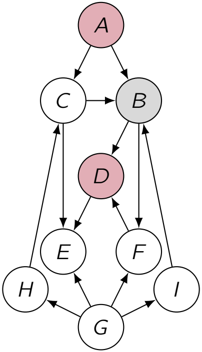


Reading conditional independencies using d-separation

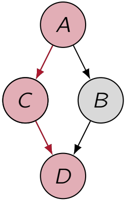
$$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$

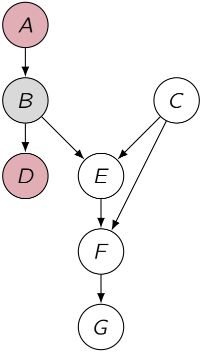


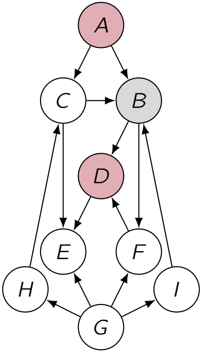




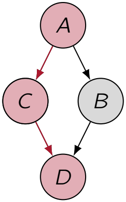
$$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$





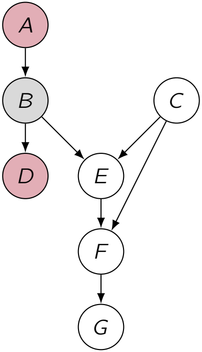


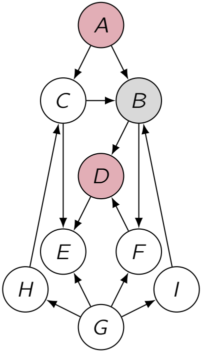
$$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$



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

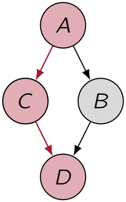
$$\implies A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$

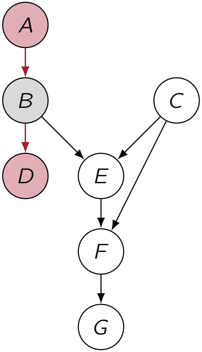


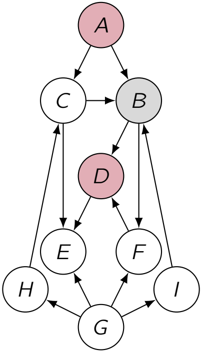


Reading conditional independencies using d-separation

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



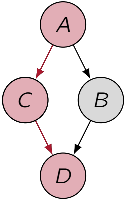




$< A, B, D >$ is not blocked

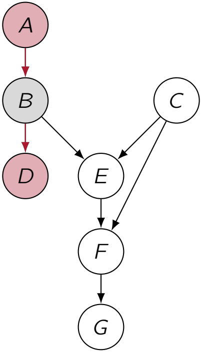
$$\implies A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$

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



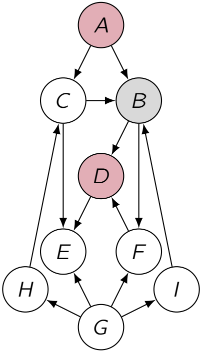
$< A, B, D >$ is not blocked

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



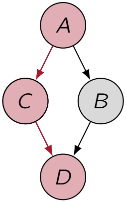
All paths are blocked

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



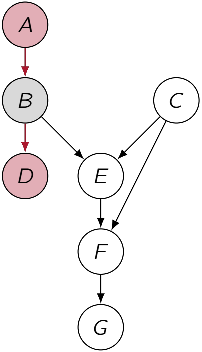
Reading conditional independencies using d-separation

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



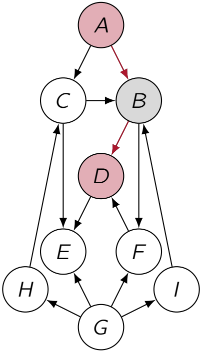
$< A, B, D >$ is not blocked

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



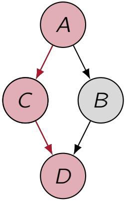
All paths are blocked

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

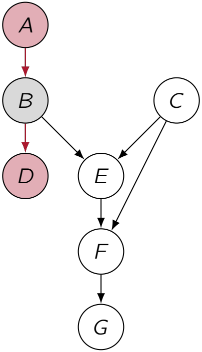


Reading conditional independencies using d-separation

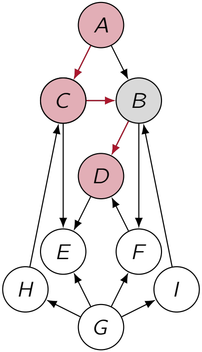
$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



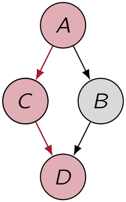
$< A, B, D >$ is not blocked
$\implies A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



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

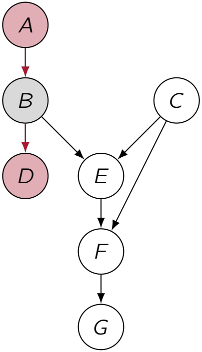


$$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$



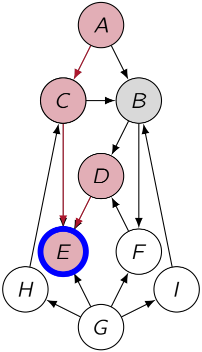
$< A, B, D >$ is not blocked

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

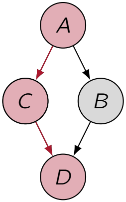


All paths are blocked

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

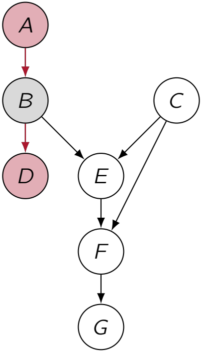


$$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$



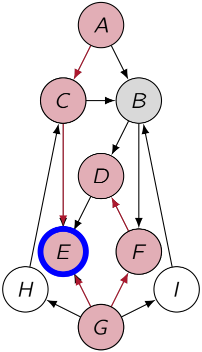
$< A, B, D >$ is not blocked

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

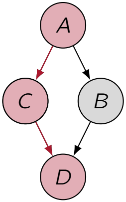


All paths are blocked

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

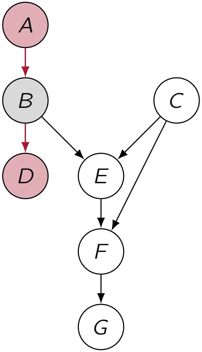


$$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$



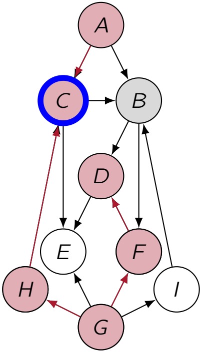
$< A, B, D >$ is not blocked

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

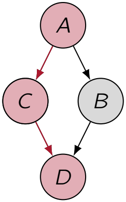


All paths are blocked

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

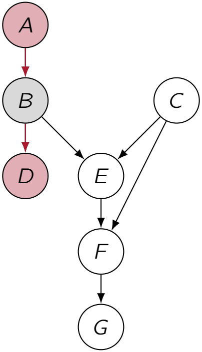


$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



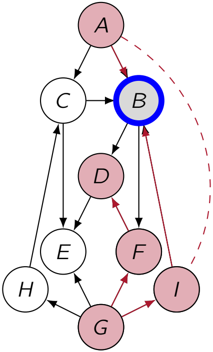
$< A, B, D >$ is not blocked

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

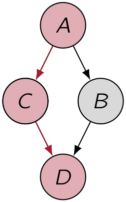


All paths are blocked

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

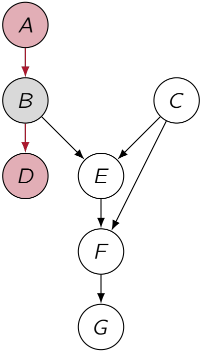


$$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$



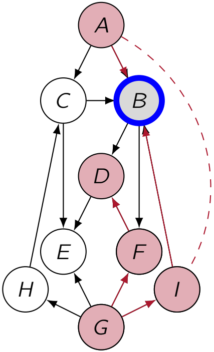
$< A, B, D >$ is not blocked

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



All paths are blocked

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



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

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

1

Association

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 Y:

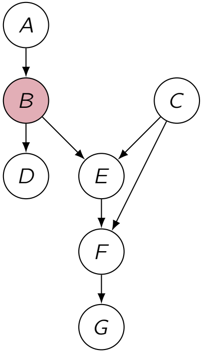
- Parents of Y;
- Children of Y;
- Other parents of the children of Y.

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 Y:

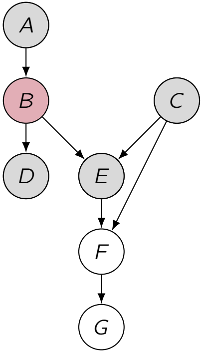
- Parents of Y;
- Children of Y;
- Other parents of the children of Y.



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

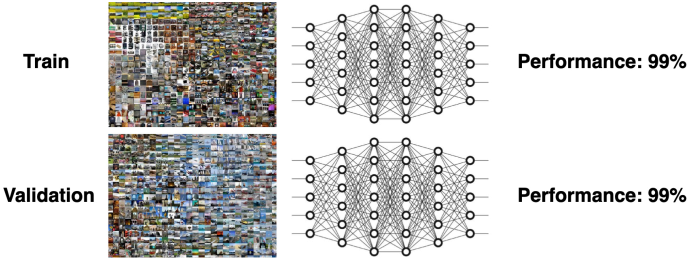
Relevant variables for prediciting Y:

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



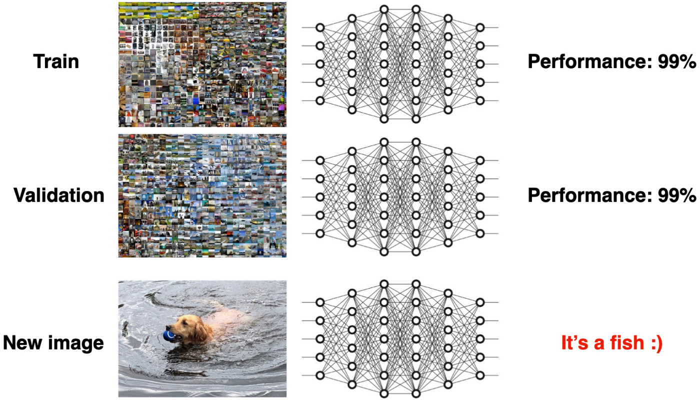
1
Association

Distribution shifts



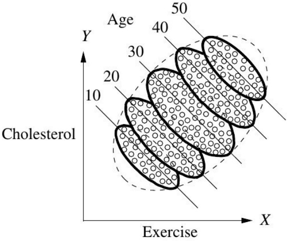
P. Cui and T. Zhang. Causal Inference and Stable Learning. Tutoriel at ICML, 2019.

Distribution shifts

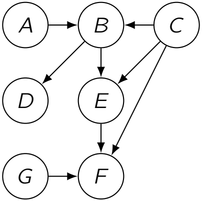


P. Cui and T. Zhang. Causal Inference and Stable Learning. Tutoriel at ICML, 2019.

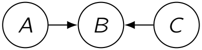
In a study, we measure weekly exercise and cholesterol levels for various age groups.



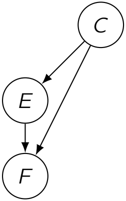
Suppose this DAG is a bayesian network compatible with P:



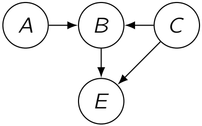
Do the following subgraphs represent Bayesian networks that are compatible with their respective distributions?



$P(A, B, C)$

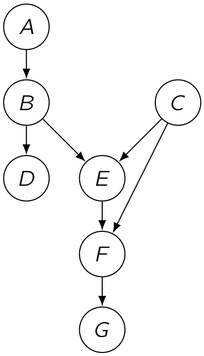


$P(C, E, F)$



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

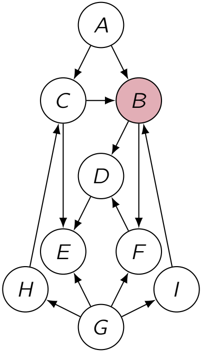
Exercise 2



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

Exercise 3

What are the relevant variables for predicting B?



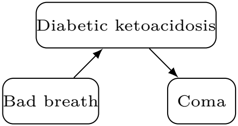
SANTÉ SORBONNE UNIVERSITÉ

2 Intervention

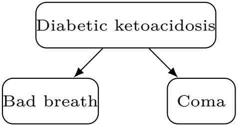
Causal graphs
Causal reasoning
Causal discovery

2 Intervention

Bayesian networks vs causal graph

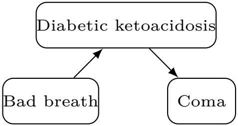


Bad breath $\perp\!\!\!\perp_P$ Coma | Diabetic ketoacidosis
Bayesian network



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

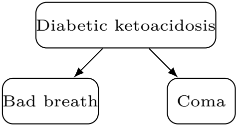
Bayesian networks vs causal graph



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

Bayesian network

Not a causal graph

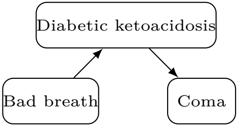


Bad breath $\perp\!\!\!\perp_P$ Coma $|$ Diabetic ketoacidosis

Bayesian network

Causal graph

Bayesian networks vs causal graph

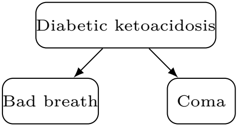


Bad breath $\perp\!\!\!\perp_P$ Coma | Diabetic ketoacidosis

Bayesian network

Not a causal graph

Oracle for conditional independence



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

Bayesian network

Causal graph

Oracle for intervention

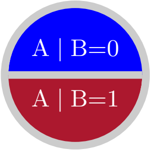
Conditioning vs Intervening (1/2)

Population

Conditioning vs Intervening (1/2)

Population

Sub-populations

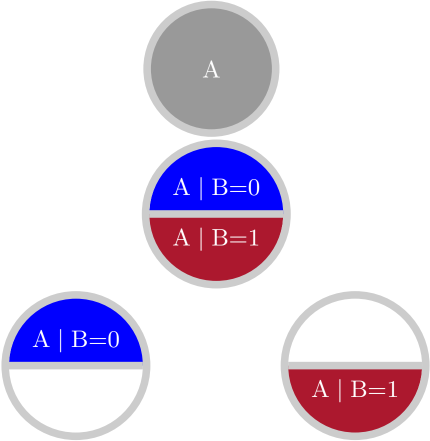


Conditioning vs Intervening (1/2)

Population

Sub-populations

Conditioning



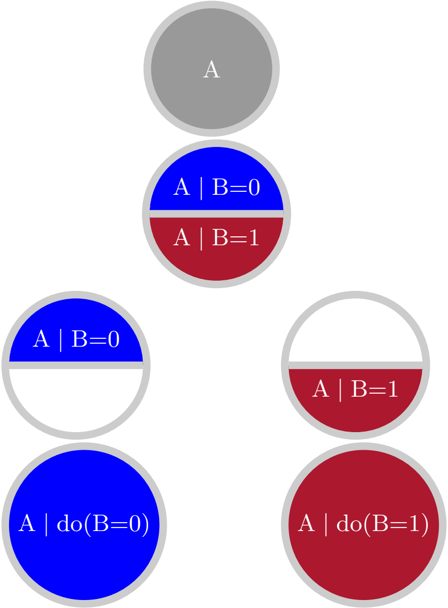
Conditioning vs Intervening (1/2)

Population

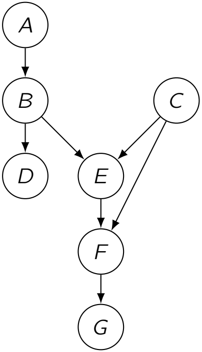
Sub-populations

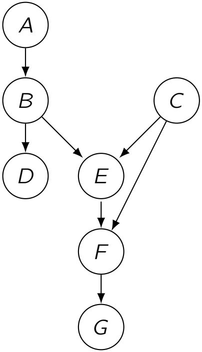
Conditioning

Intervening

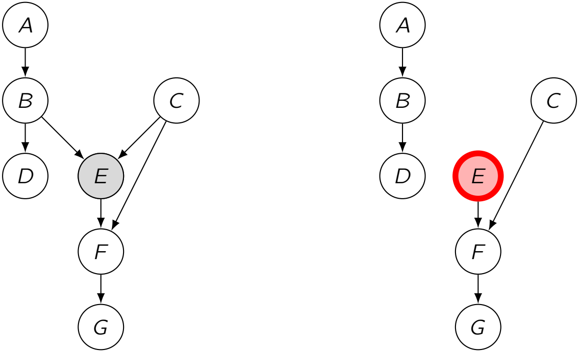


Conditioning vs Intervening (2/2)

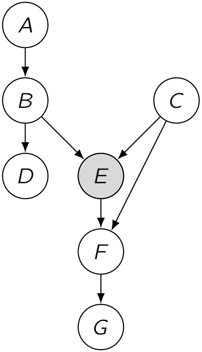


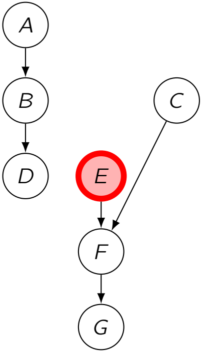


Conditioning vs Intervening (2/2)



Conditioning vs Intervening (2/2)

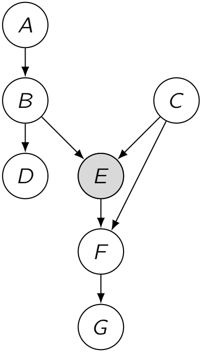


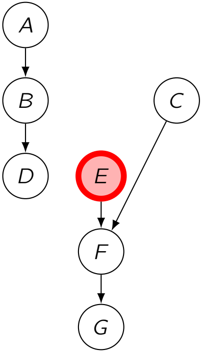


Note that there are two types of interventions:

- Structural (or hard) intervention
- Parametric (or soft) intervention

Conditioning vs Intervening (2/2)

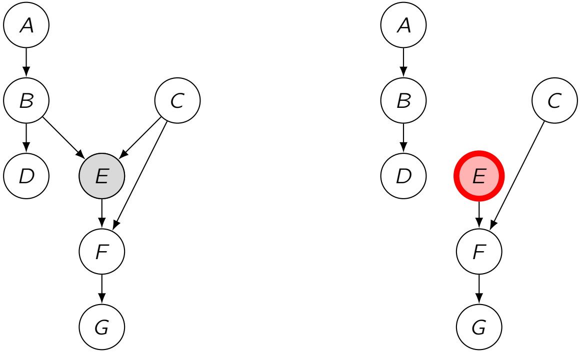




Note that there are two types of interventions:

- Structural (or hard) intervention (we will focus on this)
- Parametric (or soft) intervention

Conditioning vs Intervening (2/2)

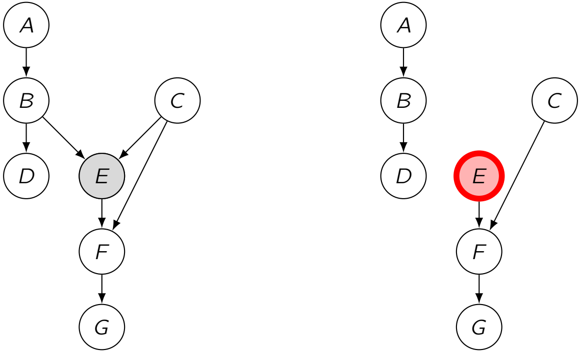


Note that there are two types of interventions:

- Structural (or hard) intervention (we will focus on this)
- Parametric (or soft) intervention

The operator $do()$ is a way to denote (hard) interventions

Conditioning vs Intervening (2/2)



Note that there are two types of interventions:

- Structural (or hard) intervention (we will focus on this)
- Parametric (or soft) intervention

The operator $do()$ is a way to denote (hard) interventions
For example $P(a, b, c, d, f, g \mid do(e))$ or $P_{E=e}(a, b, c, d, f, g)$

Truncated factorization [4]

If we intervene on a subset $\mathbb{S} \subset \mathbb{V}$, then

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

Truncated factorization [4]

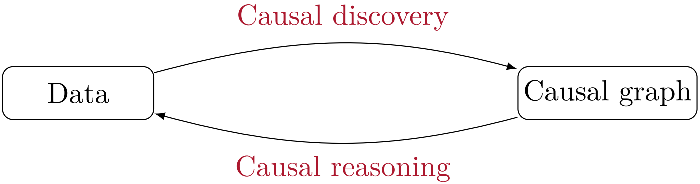
If we intervene on a subset $\mathbb{S} \subset \mathbb{V}$, then

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

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 graph compatible with P_* iff $\mathcal{G}$ and P_* satisfy the truncated factorization.

Working with causal graphs



2

Intervention

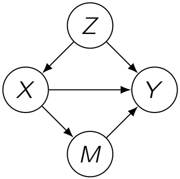
Causal graphs

Causal reasoning

Causal discovery

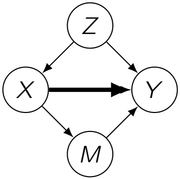
Different types of causal effects

There exists three main types of causal effects:



There exists three main types of causal effects:

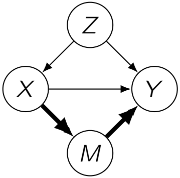
- Direct effect



Different types of causal effects

There exists three main types of causal effects:

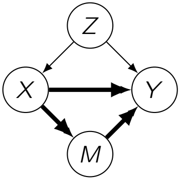
- Direct effect
- Indirect effect



Different types of causal effects

There exists three main types of causal effects:

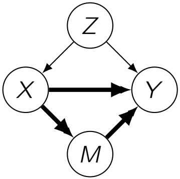
- Direct effect
- Indirect effect
- Total effect



Different types of causal effects

There exists three main types of causal effects:

- Direct effect
- Indirect effect
- Total effect



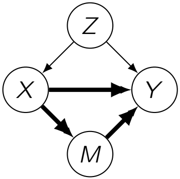
$$\text{Total effect of } X \text{ on } Y = P(y \mid do(x)) - P(y \mid do(x'))$$

where the $do()$ operator represents an intervention.

Different types of causal effects

There exists three main types of causal effects:

- Direct effect
- Indirect effect
- Total effect



$$\text{Total effect of } X \text{ on } Y = P(y \mid do(x)) - P(y \mid do(x'))$$

where the $do()$ operator represents an intervention.

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

Identifiability

The causal effect of an intervention $do(x)$ on a set of variables Y such that $Y \cap X = \emptyset$ is said to be identifiable if $P(y \mid do(x))$ is uniquely computable from $P(\mathbb{V})$.

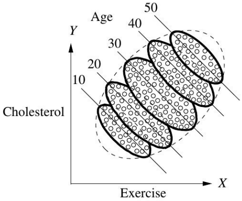
Identifiability

The causal effect of an intervention $do(x)$ on a set of variables Y such that $Y \cap X = \emptyset$ is said to be identifiable if $P(y \mid do(x))$ is uniquely computable from $P(\mathbb{V})$.

in other words, $P(Y \mid do(x))$ is identifiable if we can re-write it using a do-free expression.

Simpson paradox 1 and a simple solution

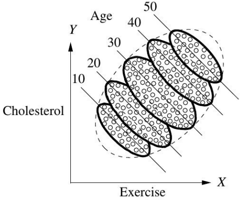
In a study, we measure weekly exercise and cholesterol levels for various age groups.

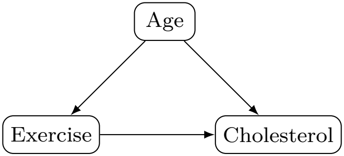


What is the effect of exercise on cholesterol $P(c \mid do(e))$?

Simpson paradox 1 and a simple solution

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

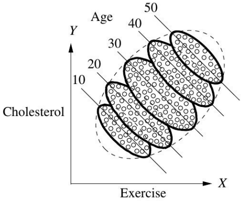


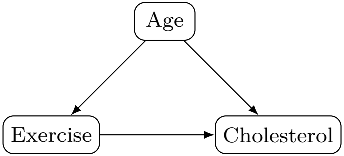


What is the effect of exercise on cholesterol $P(c \mid do(e))$?

Simpson paradox 1 and a simple solution

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





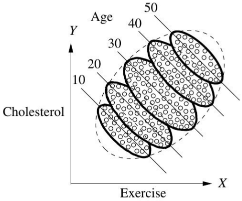
What is the effect of exercise on cholesterol $P(c \mid \color{red}{do(e)})$?

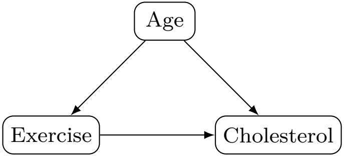
$$P(a, e, c) = P(a)P(e \mid a)P(c \mid a, e) \qquad \text{(Compatibility)}$$

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

Simpson paradox 1 and a simple solution

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





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

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

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

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

A simple solution: direct causes adjustment [4]

Theorem

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

$$P(y \mid do(x)) = \sum_{z \in Parents(x)} P(y \mid x, z) \Pr(z)$$

Limitations of the direct causes adjustment

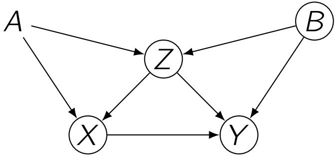
- Sometimes the set of parents is too large. Is it possible to find a smaller set?

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

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?



Back-door criterion

Consider a causal graph $\mathcal{G}$ and a causal effect $P(y \mid do(x))$. A set of variables $\mathbb{Z}$ satisfies the back-door criterion iff:

- no vertex in $\mathbb{Z}$ is a descendant of X;
- $\mathbb{Z}$ blocks every path between X and Y that contains an arrow into X.

Back-door criterion: using d-separation

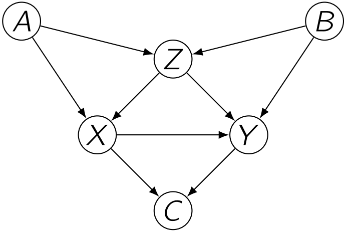
Consider a causal graph $\mathcal{G}$. The mutilated graph $\mathcal{G}_{\underline{X}}$ is a modified version of $\mathcal{G}$ that reflects the intervention on the variable X. This modification involves removing all edges going out of X.

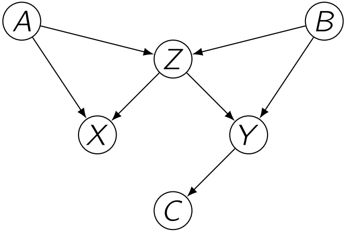
Consider a causal graph $\mathcal{G}$ and a causal effect $P(y \mid do(x))$. A set of variables $\mathbb{Z}$ satisfies the back-door criterion iff:

- $X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid \mathbb{Z}$

Back-door criterion: example

Causal graph $\mathcal{G}$ $P(y \mid do(x))$ Mutilated graph $\mathcal{G}_{\underline{X}}$



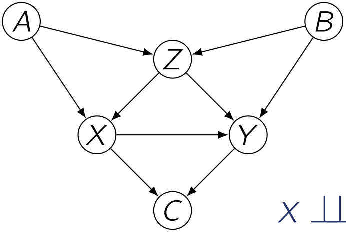


Back-door criterion: example

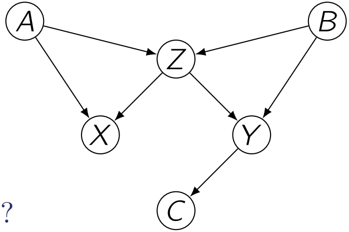
Causal graph $\mathcal{G}$

$P(y \mid do(x))$

Mutilated graph $\mathcal{G}_{\underline{X}}$

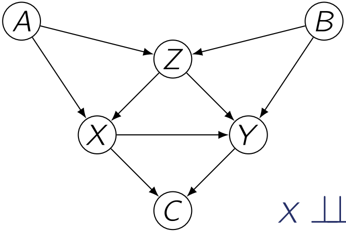


$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$?

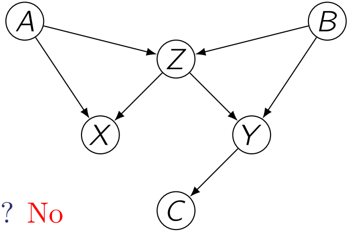


Back-door criterion: example

Causal graph $\mathcal{G}$ $\quad$ $P(y \mid do(x))$ $\quad$ Mutilated graph $\mathcal{G}_{\underline{X}}$



$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

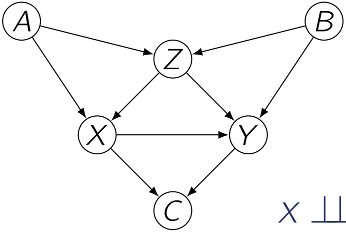


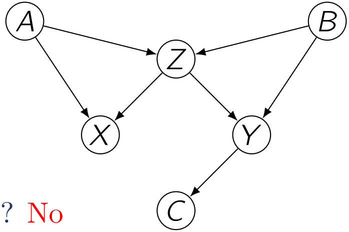
Back-door criterion: example

Causal graph $\mathcal{G}$

$P(y \mid do(x))$

Mutilated graph $\mathcal{G}_{\underline{X}}$



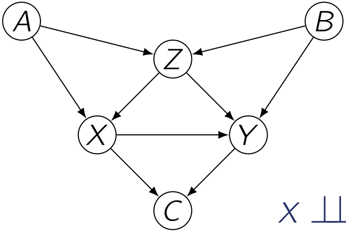


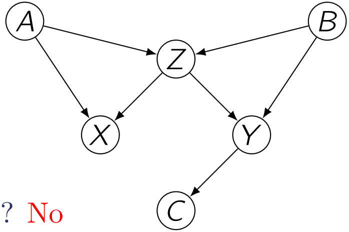
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$?

Back-door criterion: example

Causal graph $\mathcal{G}$ $\quad$ $P(y \mid do(x))$ $\quad$ Mutilated graph $\mathcal{G}_{\underline{X}}$

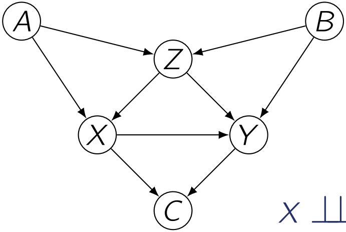




$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

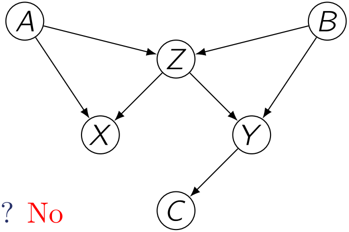
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

Causal graph $\mathcal{G}$



$P(y \mid do(x))$

Mutilated graph $\mathcal{G}_{\underline{X}}$



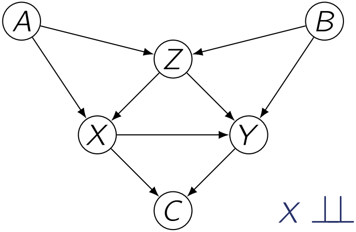
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid B$?

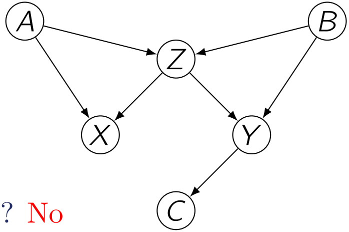
Back-door criterion: example

Causal graph $\mathcal{G}$



$P(y \mid do(x))$

Mutilated graph $\mathcal{G}_{\underline{X}}$



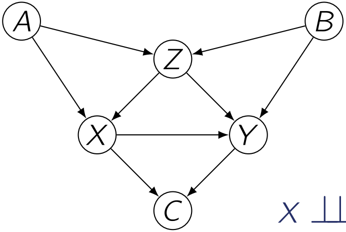
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid B$? No

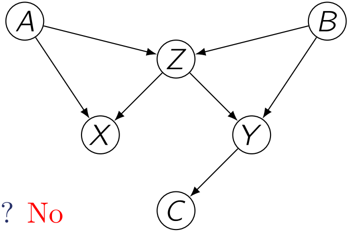
Back-door criterion: example

Causal graph $\mathcal{G}$



$P(y \mid do(x))$

Mutilated graph $\mathcal{G}_{\underline{X}}$



$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

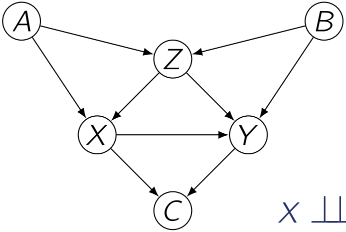
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid B$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid C$?

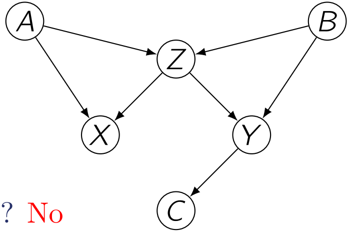
Back-door criterion: example

Causal graph $\mathcal{G}$



$P(y \mid do(x))$

Mutilated graph $\mathcal{G}_{\underline{X}}$



$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

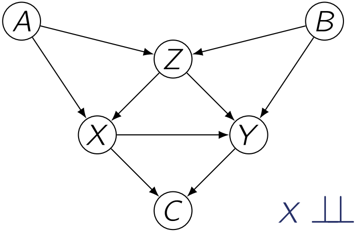
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

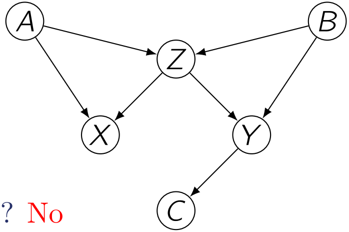
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid B$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid C$? No

Back-door criterion: example

Causal graph $\mathcal{G}$ $\quad$ $P(y \mid do(x))$ $\quad$ Mutilated graph $\mathcal{G}_{\underline{X}}$





$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

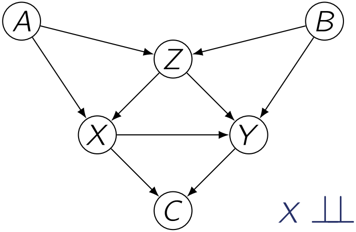
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid B$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid C$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A, B$?

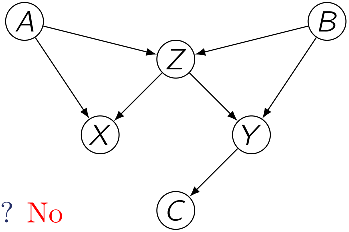
Back-door criterion: example

Causal graph $\mathcal{G}$



$P(y \mid do(x))$

Mutilated graph $\mathcal{G}_{\underline{X}}$



$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

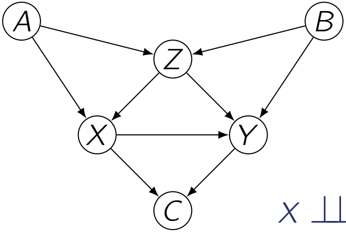
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid B$? No

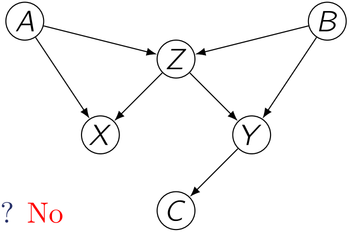
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid C$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A, B$? No

Back-door criterion: example

Causal graph $\mathcal{G}$ $\quad P(y \mid do(x))$ $\quad$ Mutilated graph $\mathcal{G}_{\underline{X}}$





$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid B$? No

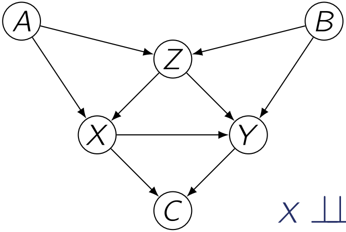
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid C$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A, B$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z, A$?

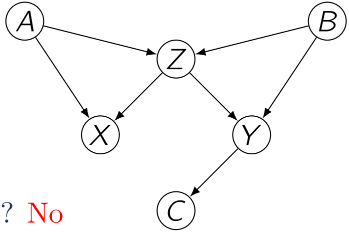
Back-door criterion: example

Causal graph $\mathcal{G}$



$P(y \mid do(x))$

Mutilated graph $\mathcal{G}_{\underline{X}}$



$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid B$? No

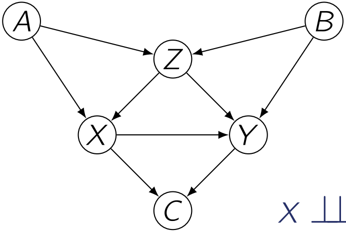
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid C$? No

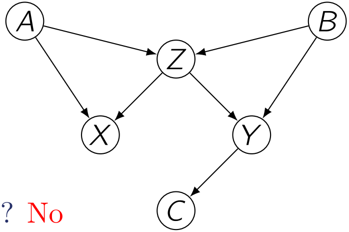
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A, B$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z, A$? Yes

Back-door criterion: example

Causal graph $\mathcal{G}$ $\qquad$ $P(y \mid do(x))$ $\qquad$ Mutilated graph $\mathcal{G}_{\underline{X}}$





$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid B$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid C$? No

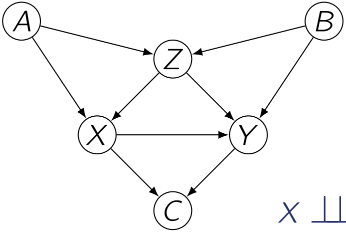
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A, B$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z, A$? Yes

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z, B$?

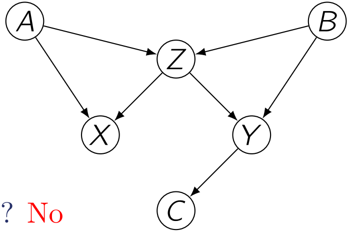
Back-door criterion: example

Causal graph $\mathcal{G}$



$P(y \mid do(x))$

Mutilated graph $\mathcal{G}_{\underline{X}}$



$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid B$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid C$? No

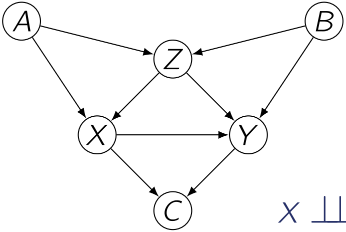
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A, B$? No

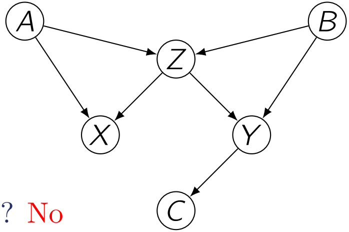
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z, A$? Yes

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z, B$? Yes

Back-door criterion: example

Causal graph $\mathcal{G}$ $\quad P(y \mid do(x))$ $\quad$ Mutilated graph $\mathcal{G}_{\underline{X}}$





$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid B$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid C$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A, B$? No

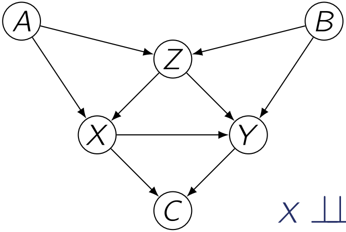
$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z, A$? Yes

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z, B$? Yes

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z, A, B$?

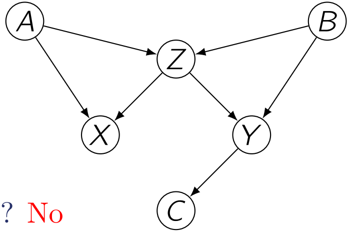
Back-door criterion: example

Causal graph $\mathcal{G}$



$P(y \mid do(x))$

Mutilated graph $\mathcal{G}_{\underline{X}}$



$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid B$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid C$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid A, B$? No

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z, A$? Yes

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z, B$? Yes

$X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z, A, B$? Yes

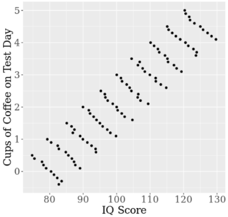
Back-door adjustment [3]

Theorem

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

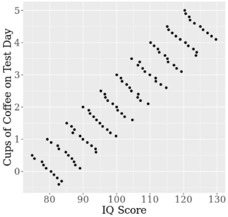
$$P(y \mid do(x)) = \sum_{z} P(y \mid x, z)P(z).$$

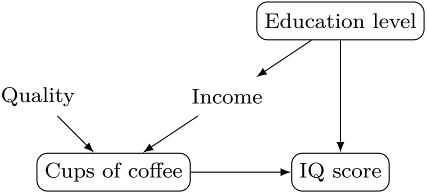
In a study, we measure the number of coffee intake and IQ score for a sample of a population with various education level.



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

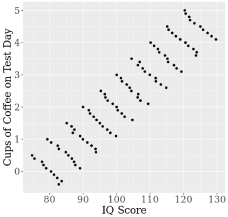
In a study, we measure the number of coffee intake and IQ score for a sample of a population with various education level.

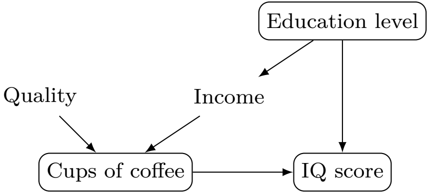




What is the effect of the number cups of coffee on IQ score $P(i \mid do(c))$?

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 $P(i \mid do(c))$?

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

Associational criterion for finding non-confounding

Let $\mathbb{Z}$ be the set of observed variables in a problem that are not affected by X. $\mathbb{Z}$ satisfies the associational criterion if each member Z of $\mathbb{Z}$ satisfied at least one of the following conditions:

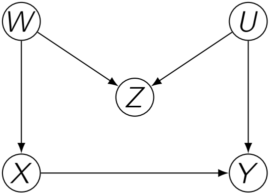
- Z is not associated with X ;
- Z is not associated with Y, conditional on X.

Associational criterion for finding non-confounding

Let $\mathbb{Z}$ be the set of observed variables in a problem that are not affected by X. $\mathbb{Z}$ satisfies the associational criterion if each member Z of $\mathbb{Z}$ satisfied at least one of the following conditions:

- Z is not associated with X ;
- Z is not associated with Y, conditional on X.

Counter-example

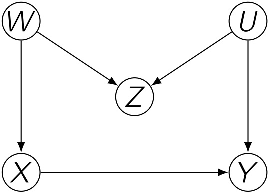


Associational criterion for finding non-confounding

Let $\mathbb{Z}$ be the set of observed variables in a problem that are not affected by X. $\mathbb{Z}$ satisfies the associational criterion if each member Z of $\mathbb{Z}$ satisfied at least one of the following conditions:

- Z is not associated with X ;
- Z is not associated with Y, conditional on X.

Counter-example



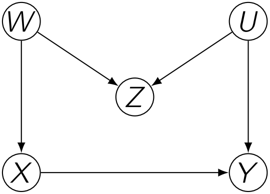
Z is associated with X and Z is associated with Y, conditional on X

Associational criterion for finding non-confounding

Let $\mathbb{Z}$ be the set of observed variables in a problem that are not affected by X. $\mathbb{Z}$ satisfies the associational criterion if each member Z of $\mathbb{Z}$ satisfied at least one of the following conditions:

- Z is not associated with X ;
- Z is not associated with Y, conditional on X.

Counter-example



Z is associated with X and Z is associated with Y, conditional on X

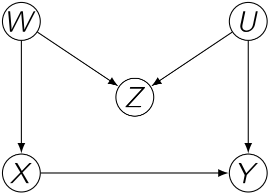
$\implies$ Z should be adjusted on

Associational criterion for finding non-confounding

Let $\mathbb{Z}$ be the set of observed variables in a problem that are not affected by X. $\mathbb{Z}$ satisfies the associational criterion if each member Z of $\mathbb{Z}$ satisfied at least one of the following conditions:

- Z is not associated with X ;
- Z is not associated with Y, conditional on X.

Counter-example



Z is associated with X and Z is associated with Y, conditional on X

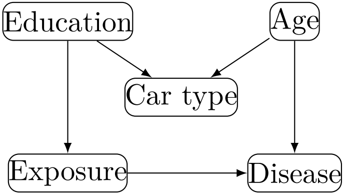
~~$\Longrightarrow$ Z should be adjusted on~~

Associational criterion for finding non-confounding

Let $\mathbb{Z}$ be the set of observed variables in a problem that are not affected by X. $\mathbb{Z}$ satisfies the associational criterion if each member Z of $\mathbb{Z}$ satisfied at least one of the following conditions:

- Z is not associated with X ;
- Z is not associated with Y, conditional on X.

Counter-example



Car type is associated with Exposure and Car type is associated with Disease, conditional on Exposure

~~$\Longrightarrow$ Car type should be adjusted on~~

Modified associational criterion for finding non-confounding

Let $\mathbb{Z}$ be the set of variables in a problem that are not affected by X. $\mathbb{Z}$ satisfies the associational criterion if it can be devided into two sets $\mathbb{Z}_1$ and $\mathbb{Z}_2$ that satisfy the following conditions:

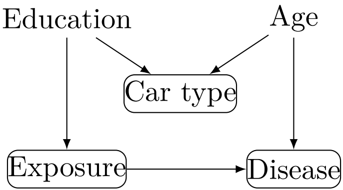
- $\mathbb{Z}_1$ is not associated with X ; and
- $\mathbb{Z}_2$ is not associated with Y, conditional on X and $\mathbb{Z}_1$.

Modified associational criterion for finding non-confounding

Let $\mathbb{Z}$ be the set of variables in a problem that are not affected by X. $\mathbb{Z}$ satisfies the associational criterion if it can be devided into two sets $\mathbb{Z}_1$ and $\mathbb{Z}_2$ that satisfy the following conditions:

- $\mathbb{Z}_1$ is not associated with X ; and
- $\mathbb{Z}_2$ is not associated with Y, conditional on X and $\mathbb{Z}_1$.

Counter-example

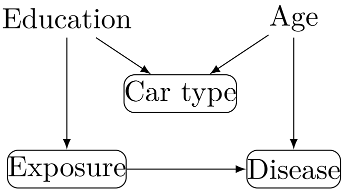


Modified associational criterion for finding non-confounding

Let $\mathbb{Z}$ be the set of variables in a problem that are not affected by X. $\mathbb{Z}$ satisfies the associational criterion if it can be devided into two sets $\mathbb{Z}_1$ and $\mathbb{Z}_2$ that satisfy the following conditions:

- $\mathbb{Z}_1$ is not associated with X ; and
- $\mathbb{Z}_2$ is not associated with Y, conditional on X and $\mathbb{Z}_1$.

Counter-example



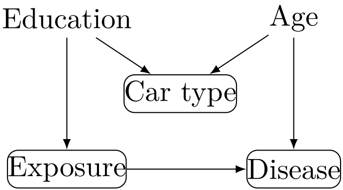
Car type is associated with Exposure

Modified associational criterion for finding non-confounding

Let $\mathbb{Z}$ be the set of variables in a problem that are not affected by X. $\mathbb{Z}$ satisfies the associational criterion if it can be devided into two sets $\mathbb{Z}_1$ and $\mathbb{Z}_2$ that satisfy the following conditions:

- $\mathbb{Z}_1$ is not associated with X ; and
- $\mathbb{Z}_2$ is not associated with Y, conditional on X and $\mathbb{Z}_1$.

Counter-example



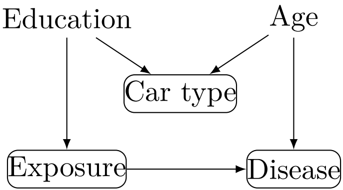
Car type is associated with Exposure
$\implies$ Car type should be adjusted on

Modified associational criterion for finding non-confounding

Let $\mathbb{Z}$ be the set of variables in a problem that are not affected by X. $\mathbb{Z}$ satisfies the associational criterion if it can be devided into two sets $\mathbb{Z}_1$ and $\mathbb{Z}_2$ that satisfy the following conditions:

- $\mathbb{Z}_1$ is not associated with X ; and
- $\mathbb{Z}_2$ is not associated with Y, conditional on X and $\mathbb{Z}_1$.

Counter-example



Car type is associated with Exposure

~~$\implies$ Car type should be adjusted on~~

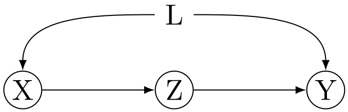
Incompleteness of the back-door criterion

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

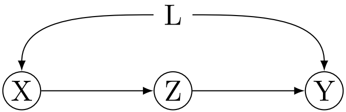
Incompleteness of the back-door criterion

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

Going beyond the back-door

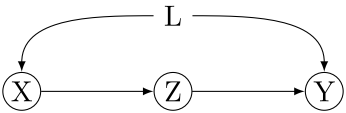


Going beyond the back-door



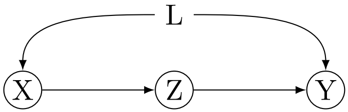
- $P(z \mid do(x)) = P(z \mid x)$ (No back-door path)

Going beyond the back-door



- $P(z \mid \color{red}{do(x)}) = P(z \mid x)$ (No back-door path)
- $P(y \mid \color{red}{do(z)}) = \sum_x P(y \mid z, x) P(x)$ (X blocks the back-door)

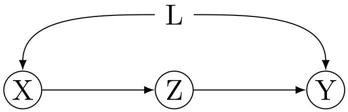
Going beyond the back-door



- $P(z \mid do(x)) = P(z \mid x)$ (No back-door path)
- $P(y \mid do(z)) = \sum_x P(y \mid z, x) P(x)$ (X blocks the back-door)

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

Going beyond the back-door



- $P(z \mid do(x)) = P(z \mid x)$ (No back-door path)
- $P(y \mid do(z)) = \sum_x P(y \mid z, x) P(x)$ (X blocks the back-door)

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

Front-door criterion

Consider a causal graph $\mathcal{G}$ and a causal effect $P(y \mid do(x))$. A set of variables $\mathbb{Z}$ satisfies the front-door criterion iff:

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

Front-door adjustment

Theorem ([3])

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

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

In a study, we measure the tar and the % of cancer among smokers and non smokers in a randomly selected sample of the population.

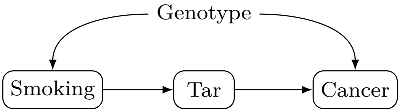
Smokers	Tar	% of cancer
No	No	10
No	Yes	5
Yes	No	90
Yes	Yes	85

What is the effect of smoking on cancer $P(c \mid do(s))$?

In a study, we measure the tar and the % of cancer among smokers and non smokers in a randomly selected sample of the population.

Smokers	Tar	% of cancer
No	No	10
No	Yes	5
Yes	No	90
Yes	Yes	85

What is the effect of smoking on cancer $P(c \mid do(s))$?

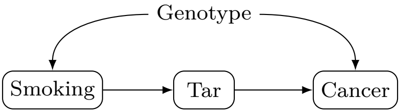


Simpson paradox 3 and the front-door in action

In a study, we measure the tar and the % of cancer among smokers and non smokers in a randomly selected sample of the population.

Smokers	Tar	% of cancer
No	No	10
No	Yes	5
Yes	No	90
Yes	Yes	85

What is the effect of smoking on cancer $P(c \mid do(s))$?



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

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!

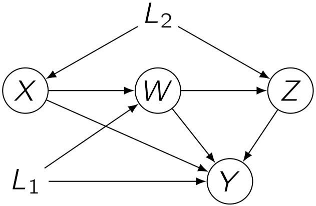
A glimpse of the do-calculus [3]

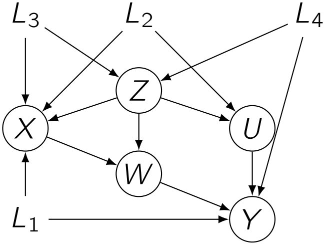
The do calculus consists of three rules :

- 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!





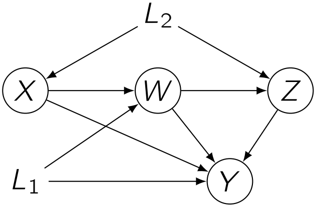
A glimpse of the do-calculus [3]

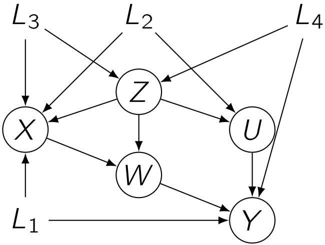
The do calculus consists of three rules :

- 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!

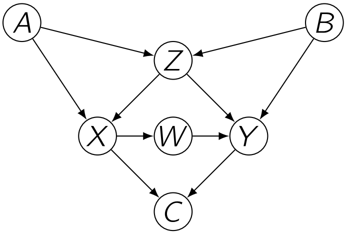




The ID algorithm [9] uses the do-calculus to automatically uncover the do-free expression.

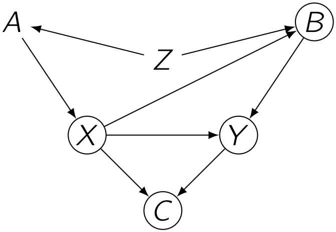
Exercise 4

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



Exercise 5

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

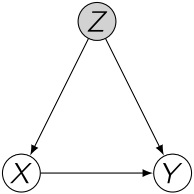


Exercise 6

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

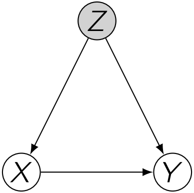
Exercise 6

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



Exercise 6

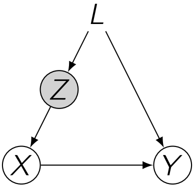
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(x))$?



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

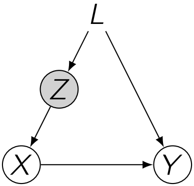
Exercise 6

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



Exercise 6

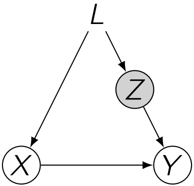
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(x))$?



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

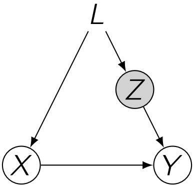
Exercise 6

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



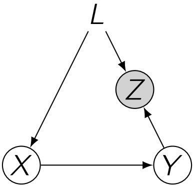
Exercise 6

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



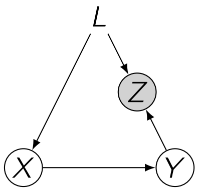
- Z blocks all back-door paths and it is not a descendant of X $\implies$ $\{Z\}$ is a good adjustment set.

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



Exercise 6

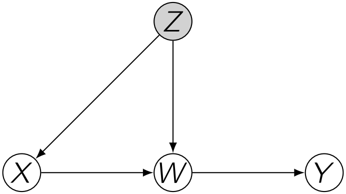
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(x))$?



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

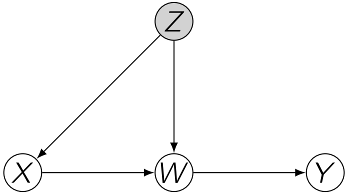
Exercise 6

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



Exercise 6

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



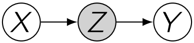
- Z blocks all back-door paths and it is not a descendant of X $\implies$ $\{Z\}$ is a good adjustment set.

Exercise 6

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

Exercise 6

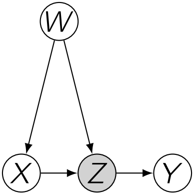
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(x))$?



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

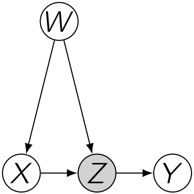
Exercise 6

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



Exercise 6

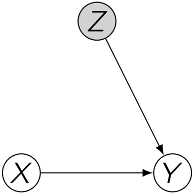
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(x))$?



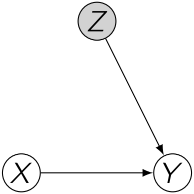
- Z d-separates X from Y
  $\implies$ $\{Z\}$ is a bad adjustment set.

Exercise 6

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



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

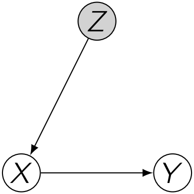


- Z does not open any back-door paths from X to Y $\implies \{Z\}$ is a neutral adjustment set;

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

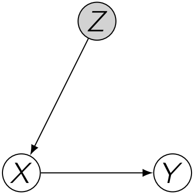
Exercise 6

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



Exercise 6

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

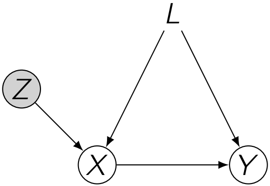


- Z does not open any back-door paths from X to Y $\implies \{Z\}$ is a neutral adjustment set;

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

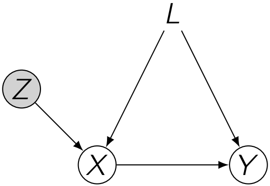
Exercise 6

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



Exercise 6

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

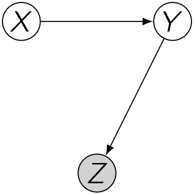


- Z does not block existing back-door path from X to Y $\implies$ $\{Z\}$ is a bad adjustment set;

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

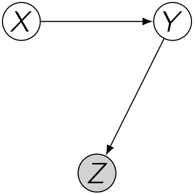
Exercise 6

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



Exercise 6

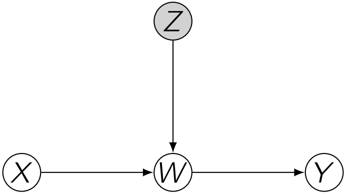
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(x))$?



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

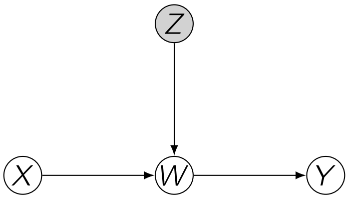
Exercise 6

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



Exercise 6

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

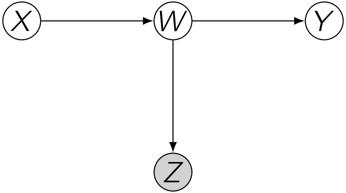


- Z does not open any back-door paths from X to Y
  $\implies \{Z\}$ is a neutral adjustment set;

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

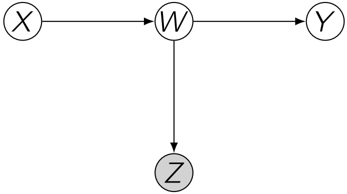
Exercise 6

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



Exercise 6

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

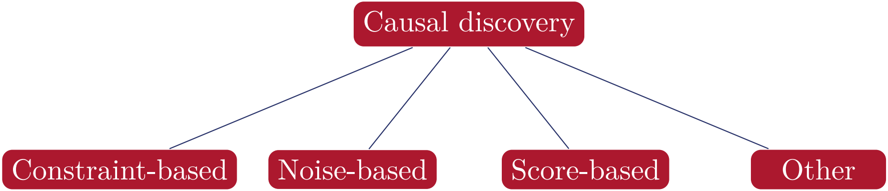


- Z is a descendant of X
  $\implies$ $\{Z\}$ is a bad adjustment set.

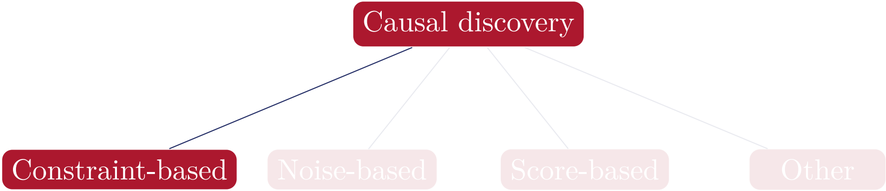
2

Intervention

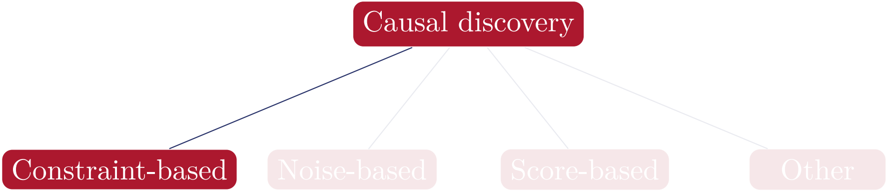
Families of causal discovery methods



Families of causal discovery methods



Families of causal discovery methods



Constraint-based: run local tests of independence to create constraints on space of possible graphs.

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

Causal discovery using conditional independencies

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

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

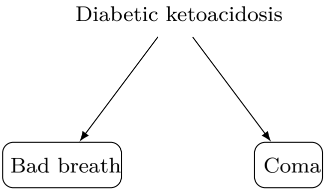
We cannot even construct the skeleton of the graph because

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

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

We cannot even construct the skeleton of the graph because

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

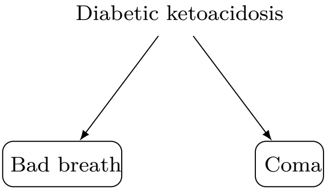


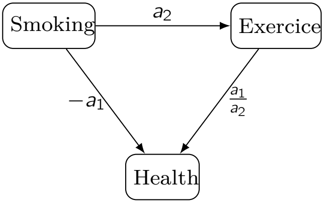
Causal discovery using conditional independencies

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

We cannot even construct the skeleton of the graph because

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





Additional assumptions

Causal sufficiency: No hidden confounding.

Additional assumptions

Causal sufficiency: No hidden confounding.

Faithfulness: All conditional independence relations true in P are entailed by the Markov condition applied to $\mathcal{G}$.

Additional assumptions

Causal sufficiency: No hidden confounding.

Faithfulness: All conditional independence relations true in P are entailed by the Markov condition applied to $\mathcal{G}$.

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

Additional assumptions

Causal sufficiency: No hidden confounding.

Faithfulness: All conditional independence relations true in P are entailed by the Markov condition applied to $\mathcal{G}$.

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

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

Additional assumptions

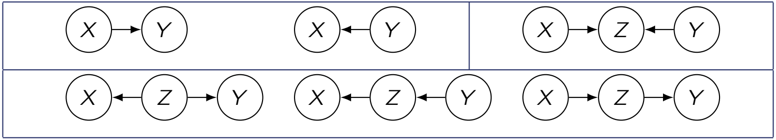
Causal sufficiency: No hidden confounding.

Faithfulness: All conditional independence relations true in P are entailed by the Markov condition applied to $\mathcal{G}$.

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

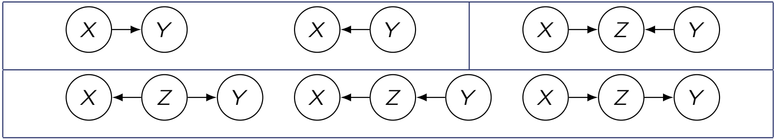
Given observational data, is it possible to infer a causal graph using conditional independencies under the assumptions of faithfulness and causal sufficiency? In general no!

Equivalence in terms of conditional independence



Markov equivalence class [12]

Equivalence in terms of conditional independence

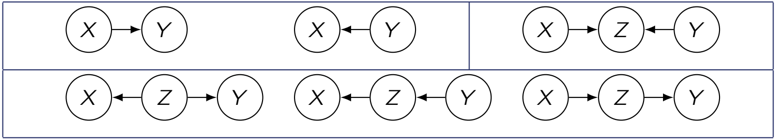


Theorem

Two causal graphs are Markov equivalent iff they have the same skeleton and the same V-structures.

Markov equivalence class [12]

Equivalence in terms of conditional independence



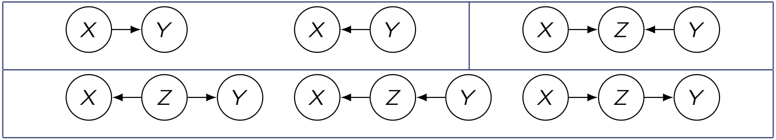
Theorem

Two causal graphs are Markov equivalent iff they have the same skeleton and the same V-structures.

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

Markov equivalence class [12]

Equivalence in terms of conditional independence



Theorem

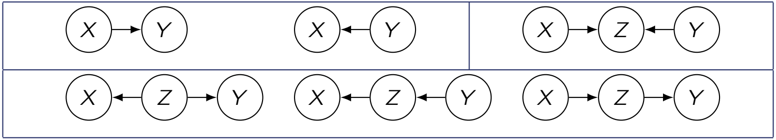
Two causal graphs are Markov equivalent iff they have the same skeleton and the same V-structures.

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

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

Markov equivalence class [12]

Equivalence in terms of conditional independence



Theorem

Two causal graphs are Markov equivalent iff they have the same skeleton and the same V-structures.

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

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

Theorem

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

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

Finding skeleton and v-structures

Theorem

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

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

Finding skeleton and v-structures

Theorem

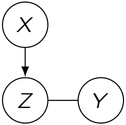
If $P(\mathbb{V})$ is faithful to some causal graph $\mathcal{G}$ with vertex $\mathbb{V}$ then:

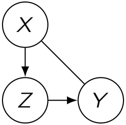
- For $X, Y \in \mathbb{V}$, X and Y are adjacent iff $\forall \mathbb{S} \subseteq \mathbb{V} \backslash \{X, Y\}$, $X \not\perp\!\!\!\perp_P Y \mid \mathbb{S}$;
- For $X, Y, Z \in \mathbb{V}$ such that X is adjacent to Z and Z is adjacent to Y and X and Y are not adjacent, $X \rightarrow Z \leftarrow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

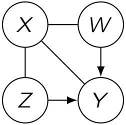
- Point 1 can be used to discover the skeleton of $\mathcal{G}$ from $P(\mathcal{V})$;
- Given the skeleton of $\mathcal{G}$, point 2 can be used to find all v-structures.

Orientation rules

Suppose we already found the skeleton and all v-structures:

PC-Rule 1: 

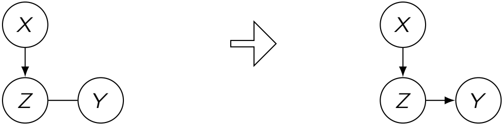
PC-Rule 2: 

PC-Rule 3: 

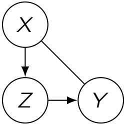
Orientation rules

Suppose we already found the skeleton and all v-structures:

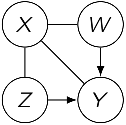
PC-Rule 1:



PC-Rule 2:



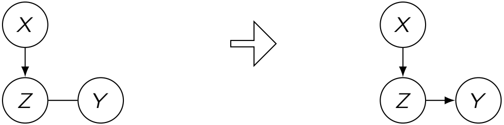
PC-Rule 3:



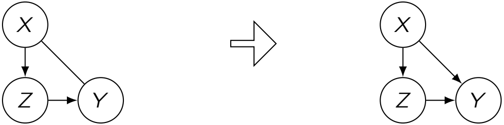
Orientation rules

Suppose we already found the skeleton and all v-structures:

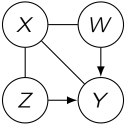
PC-Rule 1:



PC-Rule 2:

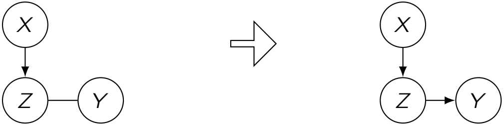


PC-Rule 3:

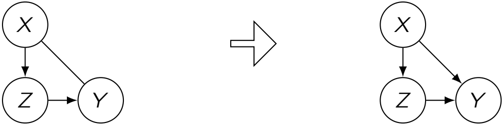


Suppose we already found the skeleton and all v-structures:

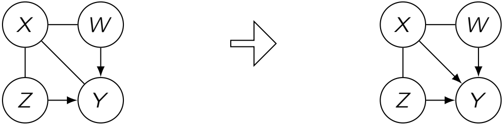
PC-Rule 1:



PC-Rule 2:



PC-Rule 3:



The PC algorithm [10]

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

The PC algorithm [10]

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

- Step 2: orientation
  - PC-Rule 0: find V-structures
  - PC-Rule 1, 2, 3: deduce other orientations by contradiction to V-structures and acyclicity

The PC algorithm [10]

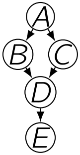
- Step 1: skeleton construction:
  - Construct a complete non oriented graph
  - Prune unnecessary edges (optimal) from the skeleton using $\perp\!\!\!\perp_P$
- Step 2: orientation
  - PC-Rule 0: find V-structures
  - PC-Rule 1, 2, 3: deduce other orientations by contradiction to V-structures and acyclicity

Theorem ([10])

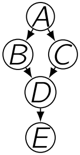
Assume the distribution P is compatible and faithful to some causal graph $\mathcal{G}$ and assume that we are given perfect conditional independence information about all pairs of variables. The PC algorithm returns the CPDAG of $\mathcal{G}$.

PC in action

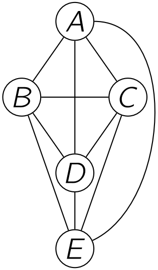
- Suppose the causal graph on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: Markov condition, causal sufficiency, faithfulness



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



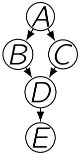
Skeleton construction:



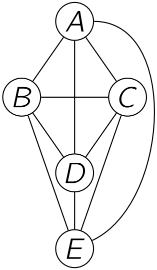
$\boxed{card = 0}$

PC in action

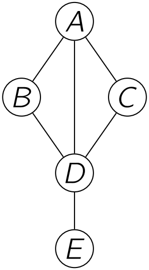
- Suppose the causal graph on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: Markov condition, causal sufficiency, faithfulness



Skeleton construction:



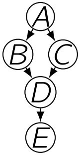
$\boxed{card = 0}$



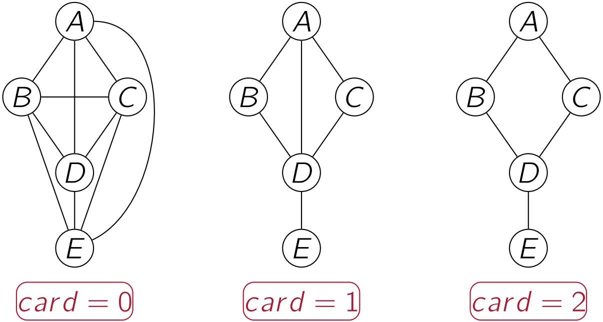
$\boxed{card = 1}$

PC in action

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

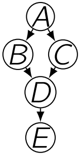


Skeleton construction:

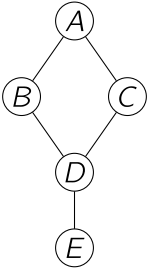


PC in action

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

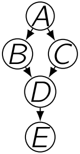


Orientation:

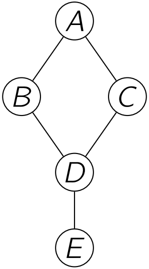


PC in action

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



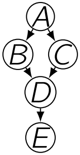
Orientation:



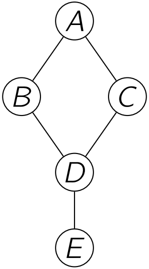
- PC-Rule 0

PC in action

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

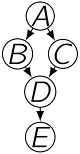


Orientation:



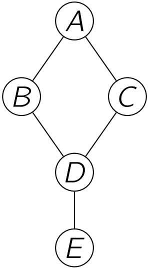
- PC-Rule 0
  - $B \perp\!\!\!\perp_P C \mid A \implies B \to D \leftarrow C$

PC in action

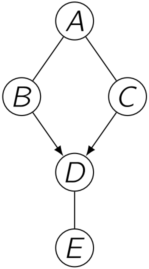


- Suppose the causal graph on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: Markov condition, causal sufficiency, faithfulness

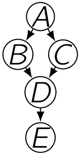
Orientation:



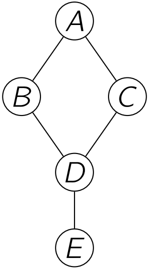
- PC-Rule 0
  - $B \perp\!\!\!\perp_P C \mid A \implies B \to D \leftarrow C$



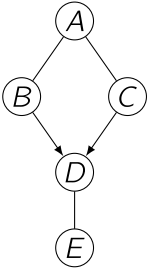
- Suppose the causal graph on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: Markov condition, causal sufficiency, faithfulness



Orientation:

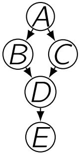


- PC-Rule 0
  - $B \perp\!\!\!\perp_P C \mid A \implies B \to D \leftarrow C$
- PC-Rule 1

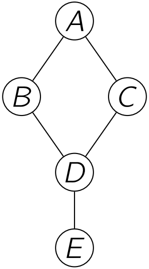


PC in action

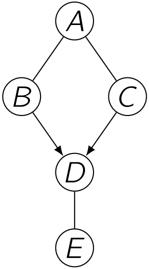
- Suppose the causal graph on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: Markov condition, causal sufficiency, faithfulness



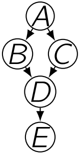
Orientation:



- PC-Rule 0
  - $B \perp\!\!\!\perp_P C \mid A \implies B \to D \leftarrow C$
- PC-Rule 1
  - $B \to D \;\&\; D - E \implies D \to E$

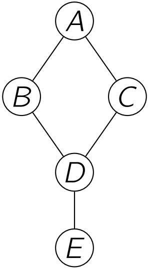


PC in action

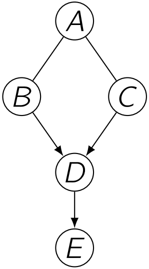


- Suppose the causal graph on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: Markov condition, causal sufficiency, faithfulness

Orientation:



- PC-Rule 0
  - $B \perp\!\!\!\perp_P C \mid A \implies B \rightarrow D \leftarrow C$
- PC-Rule 1
  - $B \rightarrow D \;\&\; D - E \implies D \rightarrow E$



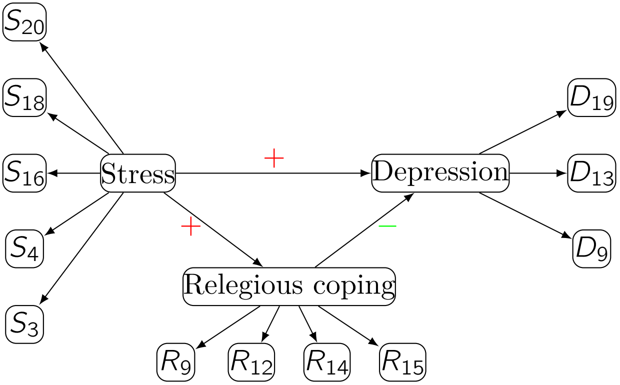
Incorporating background knowledge

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

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

MSW students ($N = 127$); 61 item survey (Likert Scale)

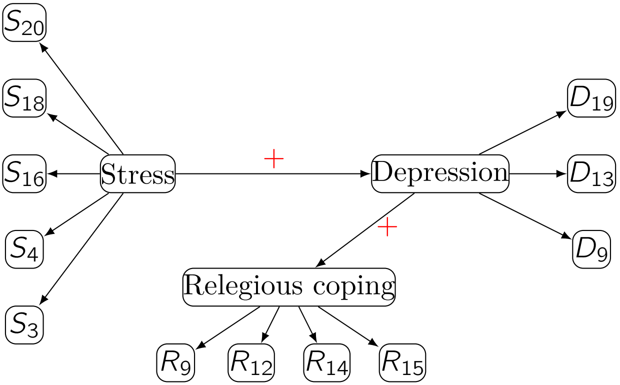
Specified graph



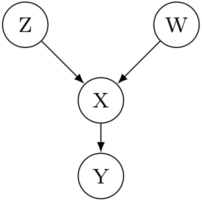
Real application of PC [7]

MSW students ($N = 127$); 61 item survey (Likert Scale)

Inferred graph (assuming stress is temporally prior)



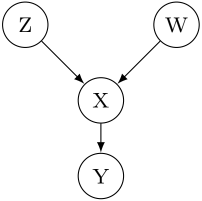
Beyond causal sufficiency (1/2)



Y-structure

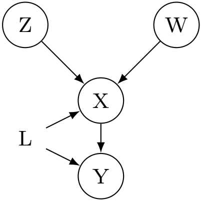
$$Z \perp\!\!\!\perp_P W$$
$$Z \not\perp\!\!\!\perp_P W \mid X$$
$$Y \not\perp\!\!\!\perp_P Z$$
$$Y \perp\!\!\!\perp_P Z \mid X$$
$$Y \not\perp\!\!\!\perp_P W$$
$$Y \perp\!\!\!\perp_P W \mid X$$

Beyond causal sufficiency (1/2)



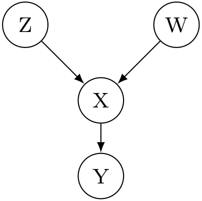
Y-structure

$$Z \perp\!\!\!\perp_P W$$
$$Z \not\!\perp\!\!\!\perp_P W \mid X$$
$$Y \not\!\perp\!\!\!\perp_P Z$$
$$Y \perp\!\!\!\perp_P Z \mid X$$
$$Y \not\!\perp\!\!\!\perp_P W$$
$$Y \perp\!\!\!\perp_P W \mid X$$



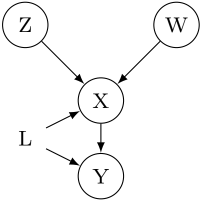
$$Z \perp\!\!\!\perp_P W$$
$$Z \not\!\perp\!\!\!\perp_P W \mid X$$
$$Y \not\!\perp\!\!\!\perp_P Z$$
$$Y \not\!\perp\!\!\!\perp_P Z \mid X$$
$$Y \not\!\perp\!\!\!\perp_P W$$
$$Y \not\!\perp\!\!\!\perp_P W \mid X$$

Beyond causal sufficiency (1/2)



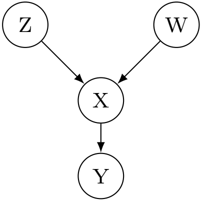
Y-structure

$$Z \perp\!\!\!\perp_P W$$
$$Z \not\perp\!\!\!\perp_P W \mid X$$
$$Y \not\perp\!\!\!\perp_P Z$$
$$Y \perp\!\!\!\perp_P Z \mid X$$
$$Y \not\perp\!\!\!\perp_P W$$
$$Y \perp\!\!\!\perp_P W \mid X$$



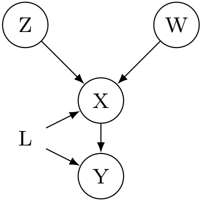
$$Z \perp\!\!\!\perp_P W$$
$$Z \not\perp\!\!\!\perp_P W \mid X$$
$$Y \not\perp\!\!\!\perp_P Z$$
$$Y \not\perp\!\!\!\perp_P Z \mid X$$
$$Y \not\perp\!\!\!\perp_P W$$
$$Y \not\perp\!\!\!\perp_P W \mid X$$

Beyond causal sufficiency (1/2)



Y-structure

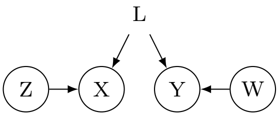
$Z \perp\!\!\!\perp_P W$
$Z \not\!\perp\!\!\!\perp_P W \mid X$
$Y \not\!\perp\!\!\!\perp_P Z$
$Y \perp\!\!\!\perp_P Z \mid X$
$Y \not\!\perp\!\!\!\perp_P W$
$Y \perp\!\!\!\perp_P W \mid X$



$Z \perp\!\!\!\perp_P W$
$Z \not\!\perp\!\!\!\perp_P W \mid X$
$Y \not\!\perp\!\!\!\perp_P Z$
$Y \not\!\perp\!\!\!\perp_P Z \mid X$
$Y \not\!\perp\!\!\!\perp_P W$
$Y \not\!\perp\!\!\!\perp_P W \mid X$

Pattern of independence can rule out hidden confounding.

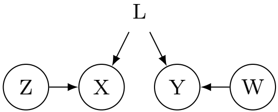
Beyond causal sufficiency (2/2)



W-structure

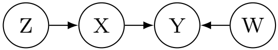
$$
\begin{gathered}
Z \not\perp\!\!\!\perp_P X \\
X \not\perp\!\!\!\perp_P Y \\
Y \not\perp\!\!\!\perp_P W \\
Z \perp\!\!\!\perp_P W \\
Z \perp\!\!\!\perp_P Y \\
X \perp\!\!\!\perp_P W \\
Z \not\perp\!\!\!\perp_P Y \mid X \\
X \not\perp\!\!\!\perp_P W \mid Y
\end{gathered}
$$

Beyond causal sufficiency (2/2)

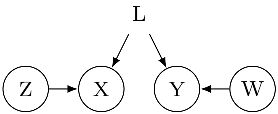


W-structure

$$Z \not\perp\!\!\!\perp_P X$$
$$X \not\perp\!\!\!\perp_P Y$$
$$Y \not\perp\!\!\!\perp_P W$$
$$Z \perp\!\!\!\perp_P W$$
$$Z \perp\!\!\!\perp_P Y$$
$$X \perp\!\!\!\perp_P W$$
$$Z \not\perp\!\!\!\perp_P Y \mid X$$
$$X \not\perp\!\!\!\perp_P W \mid Y$$

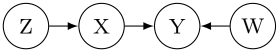


$$Z \not\perp\!\!\!\perp_P X$$
$$X \not\perp\!\!\!\perp_P Y$$
$$Y \not\perp\!\!\!\perp_P W$$
$$Z \perp\!\!\!\perp_P W$$
$$Z \not\perp\!\!\!\perp_P Y$$
$$X \perp\!\!\!\perp_P W$$
$$Z \perp\!\!\!\perp_P Y \mid X$$
$$X \not\perp\!\!\!\perp_P W \mid Y$$



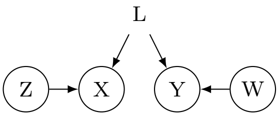
W-structure

$$Z \not\perp\!\!\!\perp_P X$$
$$X \not\perp\!\!\!\perp_P Y$$
$$Y \not\perp\!\!\!\perp_P W$$
$$Z \perp\!\!\!\perp_P W$$
$$Z \perp\!\!\!\perp_P Y$$
$$X \perp\!\!\!\perp_P W$$
$$Z \not\perp\!\!\!\perp_P Y \mid X$$
$$X \not\perp\!\!\!\perp_P W \mid Y$$



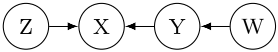
$$Z \not\perp\!\!\!\perp_P X$$
$$X \not\perp\!\!\!\perp_P Y$$
$$Y \not\perp\!\!\!\perp_P W$$
$$Z \perp\!\!\!\perp_P W$$
$$Z \not\perp\!\!\!\perp_P Y$$
$$X \perp\!\!\!\perp_P W$$
$$Z \perp\!\!\!\perp_P Y \mid X$$
$$X \not\perp\!\!\!\perp_P W \mid Y$$

Beyond causal sufficiency (2/2)

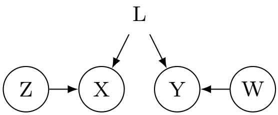


W-structure

$Z \not\perp\!\!\!\perp_P X$
$X \not\perp\!\!\!\perp_P Y$
$Y \not\perp\!\!\!\perp_P W$
$Z \perp\!\!\!\perp_P W$
$Z \perp\!\!\!\perp_P Y$
$X \perp\!\!\!\perp_P W$
$Z \not\perp\!\!\!\perp_P Y \mid X$
$X \not\perp\!\!\!\perp_P W \mid Y$

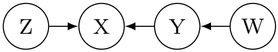


$Z \not\perp\!\!\!\perp_P X$
$X \not\perp\!\!\!\perp_P Y$
$Y \not\perp\!\!\!\perp_P W$
$Z \perp\!\!\!\perp_P W$
$Z \perp\!\!\!\perp_P Y$
$X \not\perp\!\!\!\perp_P W$
$Z \not\perp\!\!\!\perp_P Y \mid X$
$X \not\perp\!\!\!\perp_P W \mid Y$



W-structure

$$Z \not\perp\!\!\!\perp_P X$$
$$X \not\perp\!\!\!\perp_P Y$$
$$Y \not\perp\!\!\!\perp_P W$$
$$Z \perp\!\!\!\perp_P W$$
$$Z \perp\!\!\!\perp_P Y$$
$$X \perp\!\!\!\perp_P W$$
$$Z \not\perp\!\!\!\perp_P Y \mid X$$
$$X \not\perp\!\!\!\perp_P W \mid Y$$



$$Z \not\perp\!\!\!\perp_P X$$
$$X \not\perp\!\!\!\perp_P Y$$
$$Y \not\perp\!\!\!\perp_P W$$
$$Z \perp\!\!\!\perp_P W$$
$$Z \perp\!\!\!\perp_P Y$$
$$X \not\perp\!\!\!\perp_P W$$
$$Z \not\perp\!\!\!\perp_P Y \mid X$$
$$X \not\perp\!\!\!\perp_P W \mid Y$$

Pattern of independence can suggest hidden confounding.

A glimpse of the FCI algorithm

The FCI algorithm extends the PC algorithm to accommodate hidden confounders:

A glimpse of the FCI algorithm

The FCI algorithm extends the PC algorithm to accommodate hidden confounders:

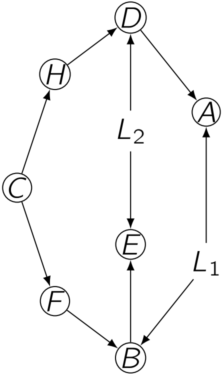
- Skeleton construction is much more complicated;
- Orientation is done using 10 different rules.

A glimpse of the FCI algorithm

The FCI algorithm extends the PC algorithm to accommodate hidden confounders:

- Skeleton construction is much more complicated;
- Orientation is done using 10 different rules.

The FCI algorithm in action:



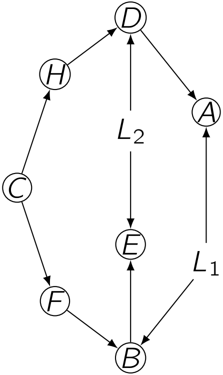
True graph

A glimpse of the FCI algorithm

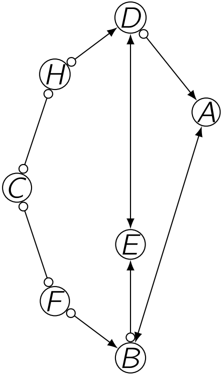
The FCI algorithm extends the PC algorithm to accommodate hidden confounders:

- Skeleton construction is much more complicated;
- Orientation is done using 10 different rules.

The FCI algorithm in action:



True graph



Inferred graph

Pros and cons of constraint-based methods

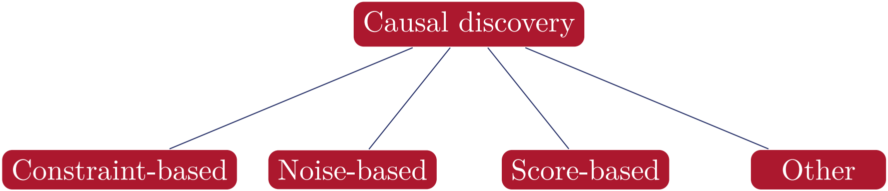
Pros:

- Non-parametric (in practice, it depends on the selected independence test)
- Intuitive

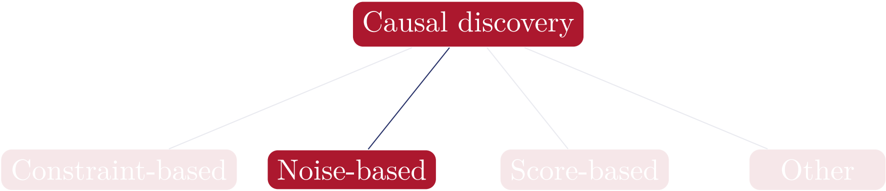
Cons:

- Recover a partially oriented graph
- The faithfulness assumption is not always accepted

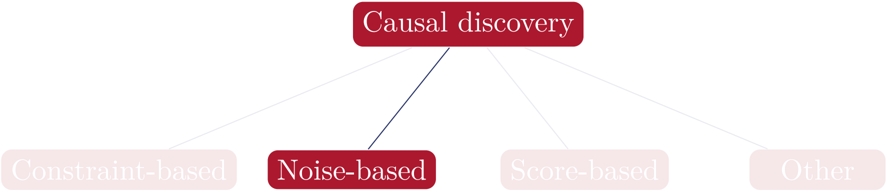
Noise based


Causal discovery
Constraint-based
Noise-based
Score-based
Other

Noise based

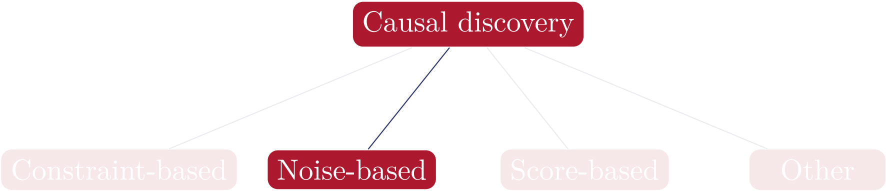

Causal discovery
Constraint-based
Noise-based
Score-based
Other

Noise based



Noise-based: find footprints in the noise that imply causal asymmetry.

Noise based



Noise-based: find footprints in the noise that imply causal asymmetry.
Also known as semi-parametric-based or functional-based.

The intuition behind the noise (1/2)

$$Suppose \begin{cases} X := \xi_x \\ Y := 2X + \xi_y \end{cases}$$

$$Suppose \begin{cases} X := \xi_x \\ Y := 2X + \xi_y \end{cases}$$

Given $P(X, Y)$, one can detect $X - Y$ but what about orientation?

The intuition behind the noise (1/2)

$$Suppose \begin{cases} X := \xi_x \\ Y := 2X + \xi_y \end{cases}$$

Given $P(X, Y)$, one can detect $X - Y$ but what about orientation?

$Y := 2X + \xi_y$?
or
$X := \frac{Y}{2} + \hat{\xi}_x$?

The intuition behind the noise (1/2)

$$Suppose \begin{cases} X := \xi_x \\ Y := 2X + \xi_y \end{cases}$$

Given $P(X, Y)$, one can detect $X - Y$ but what about orientation?

$Y := 2X + \xi_y$?
or
$X := \frac{Y}{2} + \hat{\xi}_x$?

Wihout further assumption we cannot know.

The intuition behind the noise (1/2)

$$Suppose \begin{cases} X := \xi_x \\ Y := 2X + \xi_y \end{cases}$$

Given $P(X, Y)$, one can detect $X - Y$ but what about orientation?

$Y := 2X + \xi_y$?
or
$X := \frac{Y}{2} + \hat{\xi}_x$?

Wihout further assumption we cannot know.

Assume that the noise follow a uniform distribution on $\{-1, 0, 1\}$

The intuition behind the noise (1/2)

$$Suppose \begin{cases} X := \xi_x \\ Y := 2X + \xi_y \end{cases}$$

Given $P(X, Y)$, one can detect $X - Y$ but what about orientation?

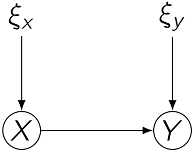
$Y := 2X + \xi_y$?
or
$X := \frac{Y}{2} + \hat{\xi}_x$?

Wihout further assumption we cannot know.

Assume that the noise follow a uniform distribution on $\{-1, 0, 1\}$

X	Y	$\xi_y = Y - 2X$	$\hat{\xi}_x = X - Y/2$
1	2	$0 \in \{-1, 0, 1\}$	$0 \in \{-1, 0, 1\}$
3	6	$0 \in \{-1, 0, 1\}$	$0 \in \{-1, 0, 1\}$
4	9	$1 \in \{-1, 0, 1\}$	$-0.5 \notin \{-1, 0, 1\}$

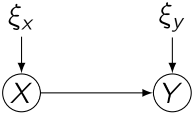
The intuition behind the noise (2/2)



$$M_1 : \begin{cases} X := f_x(\xi_x) \\ Y := f_y(X, \xi_y) \end{cases}$$

- $X \perp\!\!\!\perp_{\mathcal{G}} \xi_y$
- $Y \not\!\perp\!\!\!\perp_{\mathcal{G}} \xi_x$

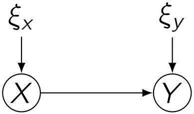
Suppose

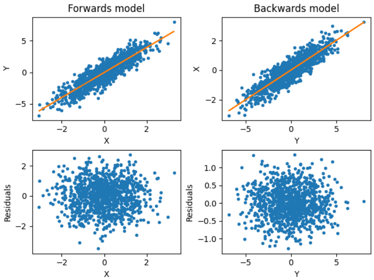


$$X \sim N(0,1)$$
$$\xi_y \sim N(0,1)$$
$$Y := 2X + \xi_y$$

The experiment that changed everything

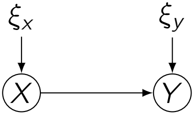
Suppose

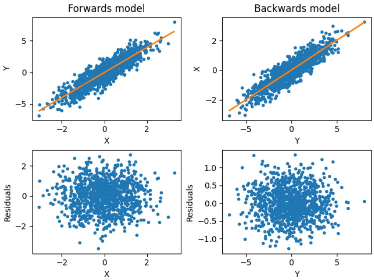




$$X \sim N(0,1)$$
$$\xi_y \sim N(0,1)$$
$$Y := 2X + \xi_y$$

Suppose



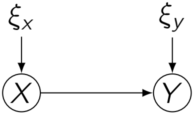


$$X \sim N(0,1)$$
$$\xi_y \sim N(0,1)$$
$$Y := 2X + \xi_y$$

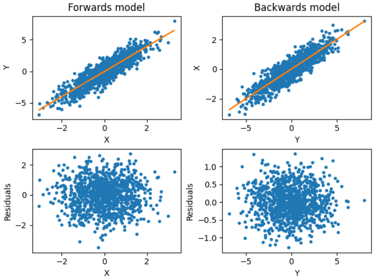
$$X \sim U(0,1)$$
$$\xi_y \sim U(0,1)$$
$$Y := 2X + \xi_y$$

The experiment that changed everything

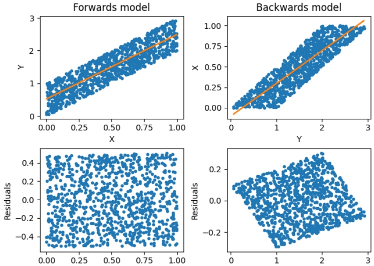
Suppose



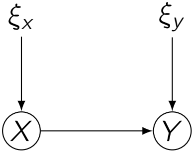
$$X \sim N(0,1)$$
$$\xi_y \sim N(0,1)$$
$$Y := 2X + \xi_y$$



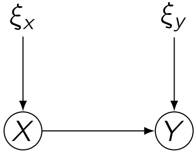
$$X \sim U(0,1)$$
$$\xi_y \sim U(0,1)$$
$$Y := 2X + \xi_y$$



The linear case (1/2)



The linear case (1/2)

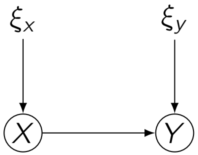


True model:

$$M_1 : \begin{cases} X := \xi_x \\ Y := aX + \xi_y \end{cases}$$

- $X \perp\!\!\!\perp_{\mathcal{G}} \xi_y$

The linear case (1/2)



True model:

$$M_1 : \begin{cases} X := \xi_x \\ Y := aX + \xi_y \end{cases}$$

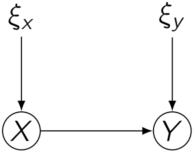
- $X \perp\!\!\!\perp_{\mathcal{G}} \xi_y$

Backwards model:

$$M_2 : \begin{cases} Y := \hat{\xi}_y \\ X := bY + \hat{\xi}_x \end{cases}$$

- $Y \perp\!\!\!\perp_{\mathcal{G}} \hat{\xi}_x$?

The linear case (1/2)



True model:

$$M_1 : \begin{cases} X := \xi_x \\ Y := aX + \xi_y \end{cases}$$

- $X \perp\!\!\!\perp_{\mathcal{G}} \xi_y$

Backwards model:

$$M_2 : \begin{cases} Y := \hat{\xi}_y \\ X := bY + \hat{\xi}_x \end{cases}$$

- $Y \perp\!\!\!\perp_{\mathcal{G}} \hat{\xi}_x$?

$$\begin{aligned} \hat{\xi}_x &= X - bY \\ &= X - b(aX + \xi_y) \\ &= (1 - ba)X - b\xi_y \end{aligned}$$

The linear case (2/2)

$$Y = aX + \xi_y$$
$$\hat{\xi}_x = (1 - ba)X - b\xi_y$$

When $Y \perp\!\!\!\perp_P \hat{\xi}_x$?

The linear case (2/2)

$$Y = aX + \xi_y$$
$$\hat{\xi}_x = (1 - ba)X - b\xi_y$$

When $Y \perp\!\!\!\perp_P \hat{\xi}_x$?

Theorem (Darmois-Skitovich)

Let $X_1, \cdots, X_n$ be independent, non degenerate random variables. If for two linear combinations:

$$l_1 = a_1X_1 + \cdots + a_nX_n$$
$$l_2 = b_1X_1 + \cdots + b_nX_n$$

are independent, then each X_i is normally distributed.

Linear non-Gaussian models (LiNGAM) [8]

Theorem

Assume that $P(X, Y)$ admits the linear model

$$Y := aX + \xi_y, \qquad X \perp\!\!\!\perp_P \xi_y,$$

with continuous random variables X, ξ_y, and Y. Then there exists $b \in \mathbb{R}$ and a random variable $\hat{\xi}_x$ such that

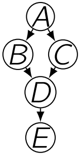
$$X := bY + \hat{\xi}_x, \qquad Y \perp\!\!\!\perp_P \hat{\xi}_x,$$

iff ξ_y and X are Gaussian.

Similar result for the multivariate case

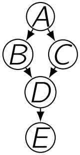
A glimpse of LiNGAM in action

- Suppose the causal graph on the right
- Input: Observational data
- Output: Causal graph
- Assumptions: Markov condition, causal sufficiency, minimality, linearity, non-gaussianity

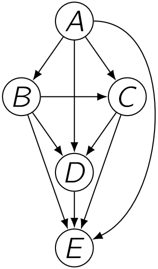


A glimpse of LiNGAM in action

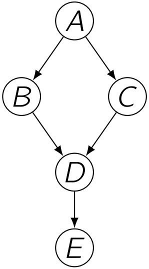
- Suppose the causal graph on the right
- Input: Observational data
- Output: Causal graph
- Assumptions: Markov condition, causal sufficiency, minimality, linearity, non-gaussianity



Causal order:



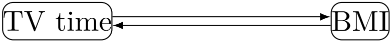
Pruning:

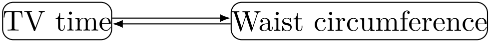


Real application of LiNGAM [1]

Young and middle-aged adults ($N = 2,060$); self-administered questionnaire for TV time

Specified graphs





Real application of LiNGAM [1]

Young and middle-aged adults ($N = 2,060$); self-administered questionnaire for TV time

Inferred graphs

$$\boxed{\text{TV time}} \longrightarrow \boxed{\text{BMI}}$$

$$\boxed{\text{TV time}} \longrightarrow \boxed{\text{Waist circumference}}$$

Extensions

- LiNGAM with hidden confounding
- Non-linear additive noise models
- Post non-linear additive noise models
- ...

Pros and cons of noise-based methods

Pros:

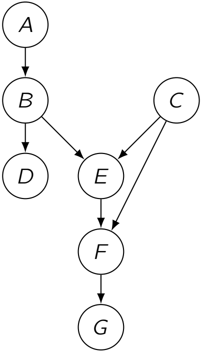
- Recover the complete causal graph

Cons:

- Semi-parametric

Exercise 7

What is the CPDAG of the following causal graph?



3

Counterfactuals

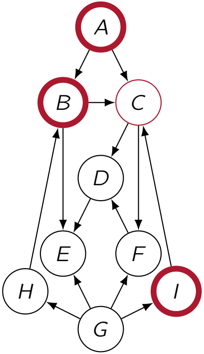
- Structural causal models
- Counterfactual reasoning
- Mediation analysis
- Difference with the potential outcome framework

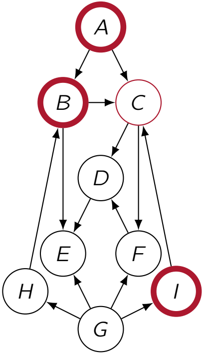
3

Counterfactuals

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

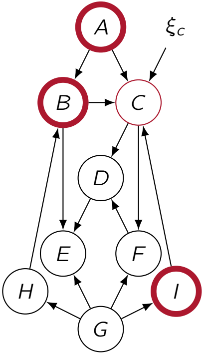
Causal mechanism





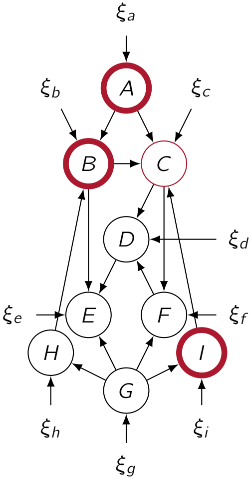
$$C = f(A, B, I)$$

Causal mechanism



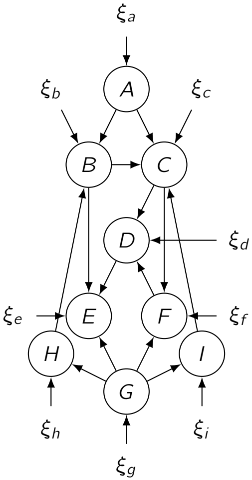
$$C := f_c(A, B, I, \xi_c)$$

Causal mechanism

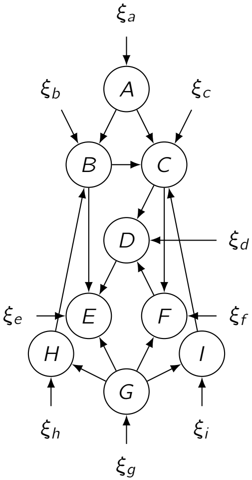


$$C := f_c(A, B, I, \xi_c)$$

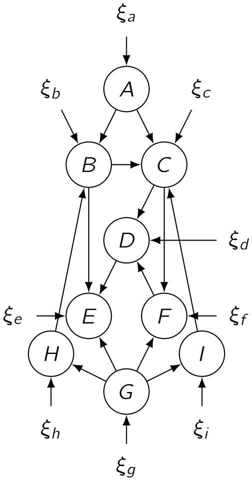
Structural causal model



Structural causal model



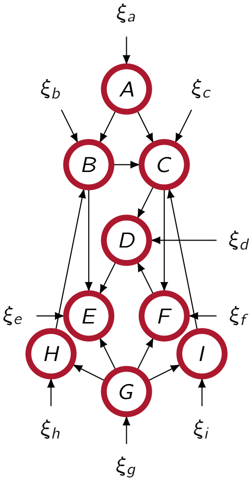
$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$



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

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

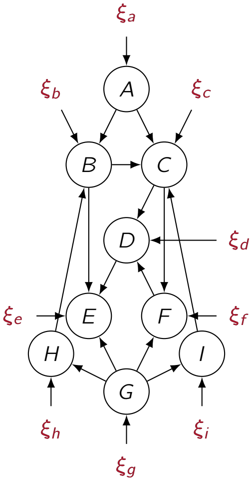
Structural causal model



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

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

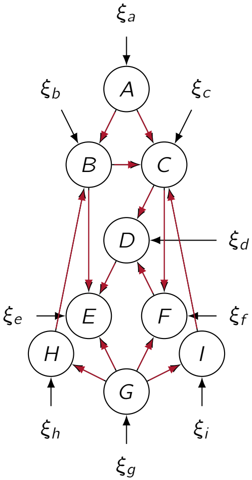
- Endogenous variables



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

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

- Endogenous variables
- Exogenous variables



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

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

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

Interventional SCMs naturally helps us in computing counterfactual queries:

$$P(D_{C=c} \mid C = c_0, D = d_0)$$

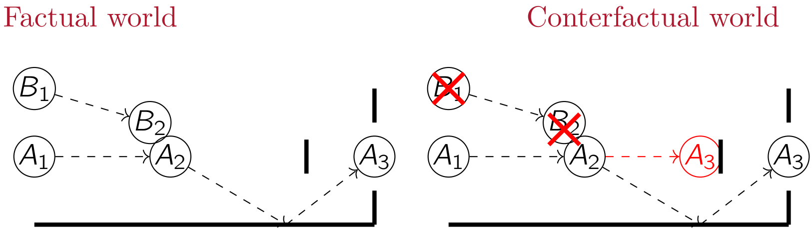
3 Counterfactuals

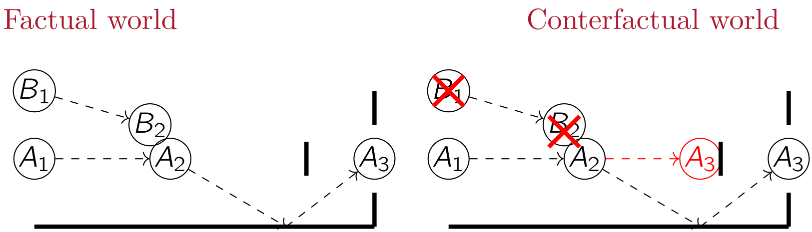
Computing counterfactuals using deterministic SCMs

1. Abduction: use the observations to determine the value of ξ_c
2. Action: modify the model M by removing the structural equations for the variable C and replacing it with the appropriate functions $C = c$, to obtain the interventional SCM M_c
3. Prediction: use M_c and the value of ξ_c to compute the value of $D_{C=c}$

If $D_{C=c} \neq d_0$ then c_0 is a cause of d_0

Computing counterfactuals using deterministic SCMs



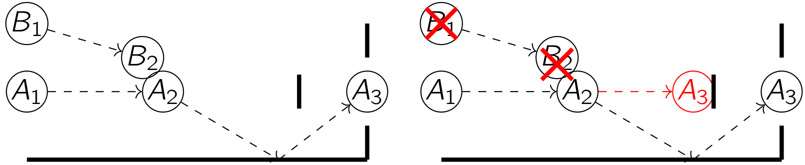


Toy example (without noise on C, W, and G):

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

Factual world

Conterfactual world



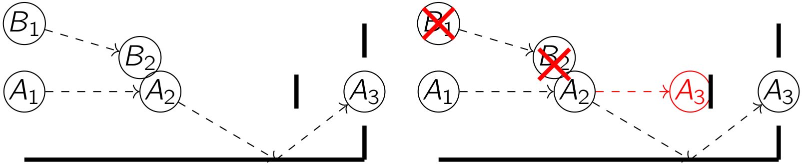
Toy example (without noise on C, W, and G):

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

Computing counterfactuals using deterministic SCMs

Factual world

Conterfactual world

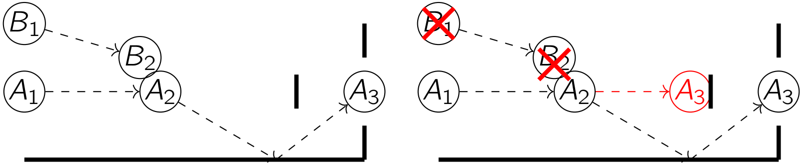


Toy example (without noise on C, W, and G):

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

Computing counterfactuals using deterministic SCMs

Factual world | Conterfactual world



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

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

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

3

Counterfactuals

Structural causal models

Counterfactual reasoning

Mediation analysis

Difference with the potential outcome framework

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 <u>structural coefficient</u> 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.

Single-door criterion

Consider a causal graph $\mathcal{G}$ and a direct effect α_{xy} associated with link $X \rightarrow Y$. Let $\mathcal{G}_{\alpha_{xy}}$ denote the causal graph that results when $X \rightarrow Y$ 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]

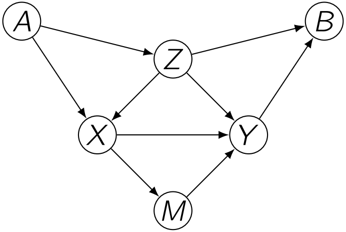
Theorem

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

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

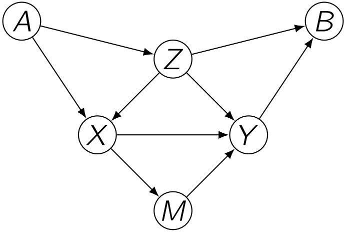
where $r_{XY.\mathbb{Z}}$ represents the residual correlation between Y and X after $\mathbb{Z}$ is "partialled out".

Single-door in action



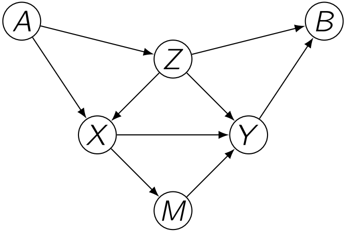
Does the following set satisfy the single-door criterion?

Single-door in action



Does the following set satisfy the single-door criterion?

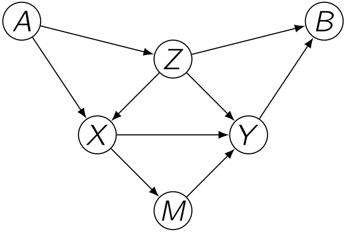
$$\{Z\} \text{ ?}$$



Does the following set satisfy the single-door criterion?

$\{Z\}$? No

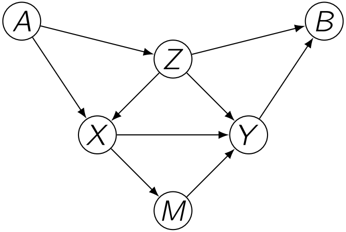
Single-door in action



Does the following set satisfy the single-door criterion?

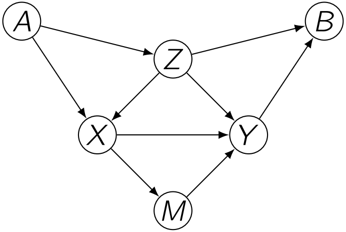
$\{Z\}$? No
$\{A\}$?

Single-door in action



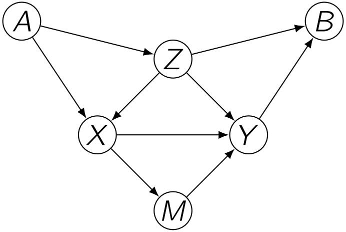
Does the following set satisfy the single-door criterion?

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



Does the following set satisfy the single-door criterion?

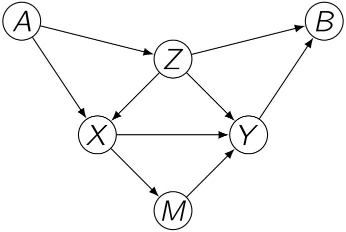
$\{Z\}$? No
$\{A\}$? No
$\{B\}$?



Does the following set satisfy the single-door criterion?

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

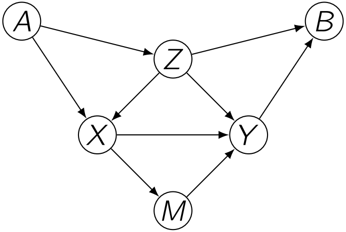
Single-door in action



Does the following set satisfy the single-door criterion?

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

Single-door in action



Does the following set satisfy the single-door criterion?

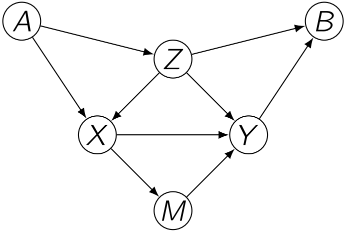
$\{Z\}$? No

$\{A\}$? No

$\{B\}$? No

$\{M\}$? No

Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No

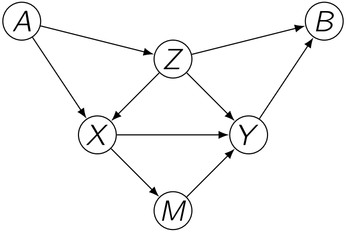
$\{A\}$? No

$\{B\}$? No

$\{M\}$? No

$\{A, B\}$?

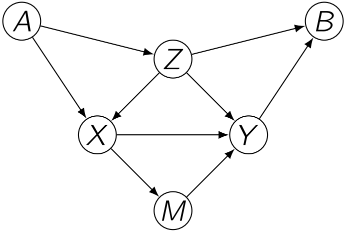
Single-door in action



Does the following set satisfy the single-door criterion?

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

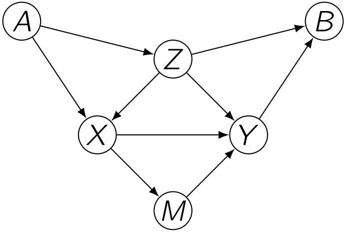
Single-door in action



Does the following set satisfy the single-door criterion?

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

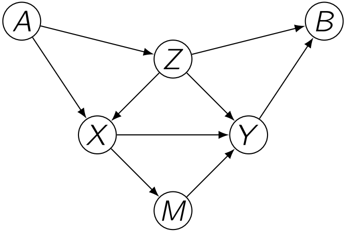
Single-door in action



Does the following set satisfy the single-door criterion?

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

Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No

$\{A\}$? No

$\{B\}$? No

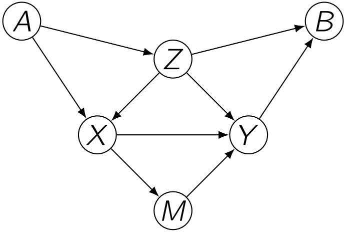
$\{M\}$? No

$\{A, B\}$? No

$\{Z, M\}$? Yes

$\{Z, M, B\}$?

Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No

$\{A\}$? No

$\{B\}$? No

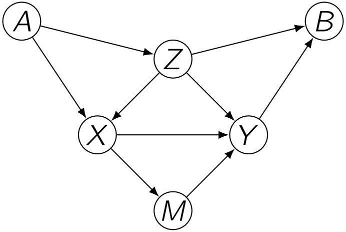
$\{M\}$? No

$\{A, B\}$? No

$\{Z, M\}$? Yes

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

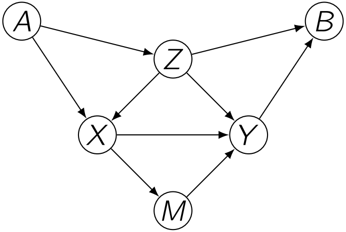
Single-door in action



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

Indirect effect

Indirect effect = total effect − direct effect

It is possible to use the single-door criterion in a non-parametric model?

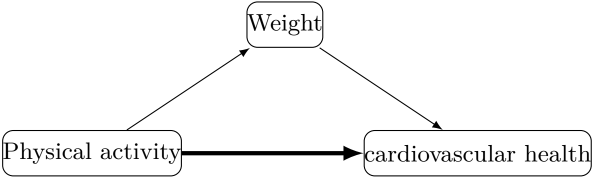
The single-door criterion in the non-parametric case

It is possible to use the single-door criterion in a non-parametric model? No!

The single-door criterion in the non-parametric case

It is possible to use the single-door criterion in a non-parametric model? No!

Example

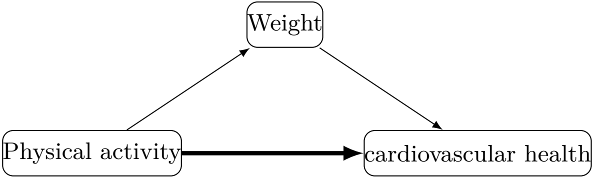


$$Y = X \times M + \xi_y$$

The single-door criterion in the non-parametric case

It is possible to use the single-door criterion in a non-parametric model? No!

Example



$$Y = X \times M + \xi_y$$

The question "What is the direct effect of X on Y?" in a non-parametric setting is not simple.

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)

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

Identifiability of NDE from experiments [5]

Theorem

Let $\mathbb{M}$ stand for the set of all intermediate variables between X and Y. If there exists a set $\mathbb{Z}$ such that

- No member of $\mathbb{Z}$ is a descendant of X or $\mathbb{M}$
- $\mathbb{Z}$ blocks all back-door paths from $\mathbb{M}$ to Y not traversing X

then the natural average direct effect of X on Y is experimentally identifiable, and it is given by

$$NDE(x, x'; Y) = \sum_m \sum_z (\mathbb{E}(y \mid do(x, m), z) - \mathbb{E}(y \mid do(x', m), z)) \times \Pr(m, do(x'), z) \Pr(z)$$

Identifiability of NDE from observations [5]

Theorem

Let $\mathbb{M}$ stand for the set of all intermediate variables between X and Y. If there exists a set $\mathbb{Z}$ such that

- No member of $\mathbb{Z}$ is a descendant of of X or $\mathbb{M}$
- $\mathbb{Z}$ blocks all back-door paths from $\mathbb{M}$ to Y not traversing X
- $\Pr(m \mid do(x), z)$ is identifiable
- $\Pr(y \mid do(x, m), z)$ is identifiable

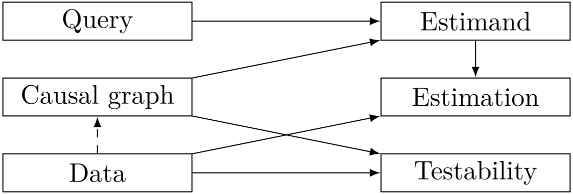
then the natural average direct effect of X on Y is identifiable, and it is given by

$$NDE(x, x'; Y) = \sum_m \sum_z (\mathbb{E}(y \mid do(x, m), z) - \mathbb{E}(y \mid do(x', m), z)) \\ \times \Pr(m, do(x'), z) \Pr(z)$$

3

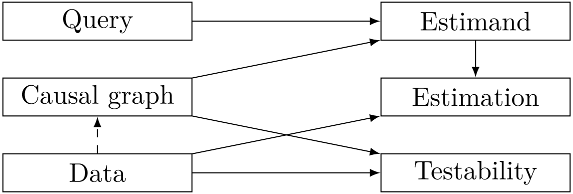
Counterfactuals

Structural causal models

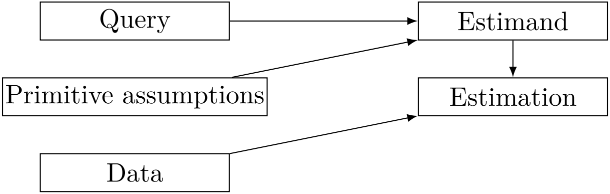


Structural causal models vs potential outcome

Structural causal models

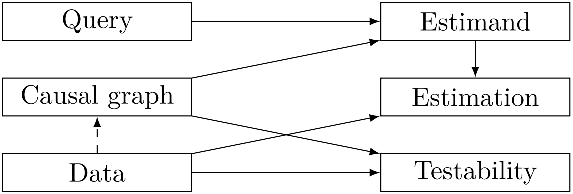


Potential outcome

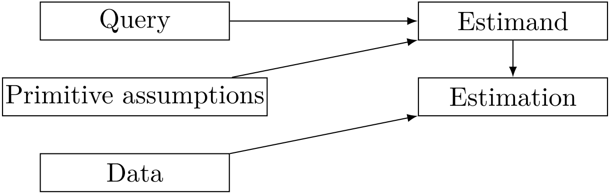


Structural causal models vs potential outcome

Structural causal models



Potential outcome



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

4

Conclusion

Counterfactual data

Experimental data

Observational data

SCMs
counterfactual reasoning
mediation analysis

Causal graphs
causal reasoning
causal discovery

Bayesian networks
prediction

5

Acknowledgements

Acknowledgements

- David Hajage for the template;
- Benjamin Glemain, Jeanne Belly-Le Guilloux, Samuel Nilusmas, Achille Raviez, and Harold Tankpinou for discussions about examples related to health;
- Simon Ferreira for proofreading

SANTÉ SORBONNE UNIVERSITÉ

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).

References II

[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

SANTÉ SORBONNE UNIVERSITÉ