

Causal Inference: Tracing History and 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



1 Association

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

2 Intervention

- Causal graphs
- Causal reasoning
- Causal discovery

3 Counterfactuals

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

4 Conclusion

1

Association

Bayesian networks

d-separation

Prediction

Do we need more than associations?

1

Association

Bayesian networks

d-separation

Prediction

Do we need more than associations?

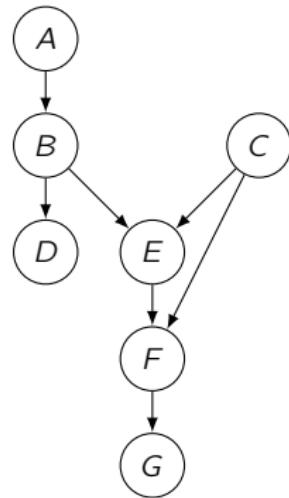
- Sets: $\mathbb{A} = \{X, Y, Z\}$
- Statistical independence: $\perp\!\!\!\perp_P$
- Statistical dependence: $\not\perp\!\!\!\perp_P$
- $P(Y = y | X = x) \equiv P(y | x)$

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

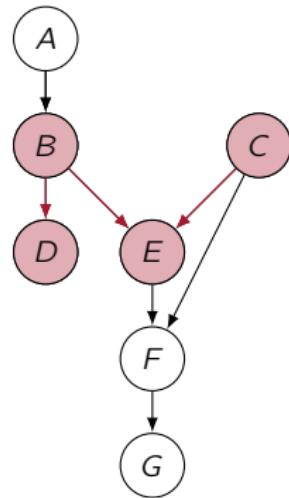
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



Directed graphs: basic concepts

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

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

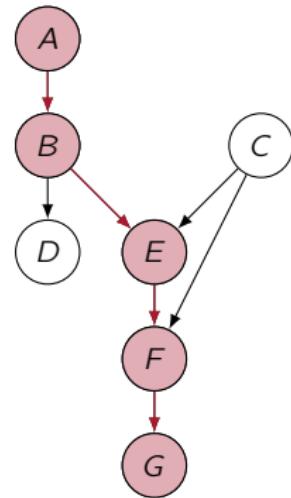


Directed graphs: basic concepts

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

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

Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$



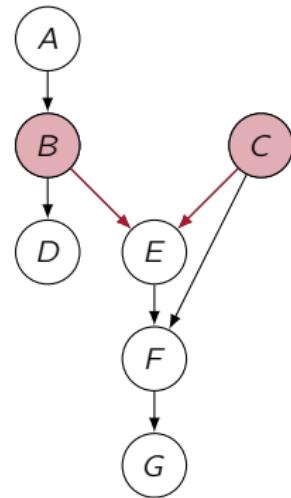
Directed graphs: basic concepts

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

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

Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$

Parents: $Pa(E) = \{B, C\}$



Directed graphs: basic concepts

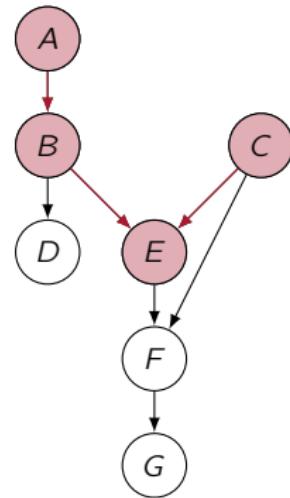
Consider the following directed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:

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

Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$

Parents: $Pa(E) = \{B, C\}$

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



Directed graphs: basic concepts

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

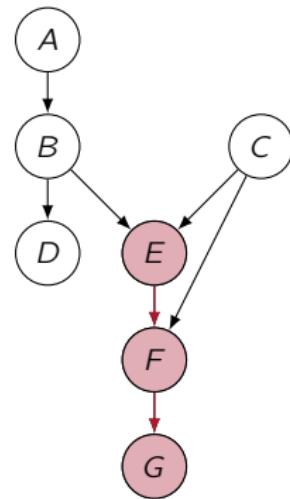
Path: $D \leftarrow B \rightarrow E \leftarrow C$

Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$

Parents: $Pa(E) = \{B, C\}$

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

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



Directed graphs: basic concepts

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

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

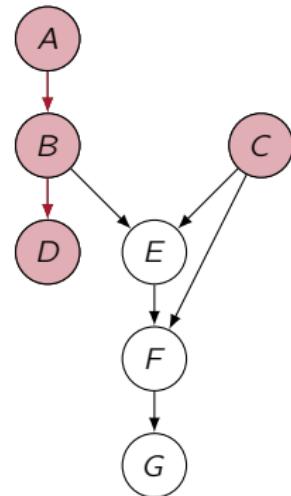
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$

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: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$

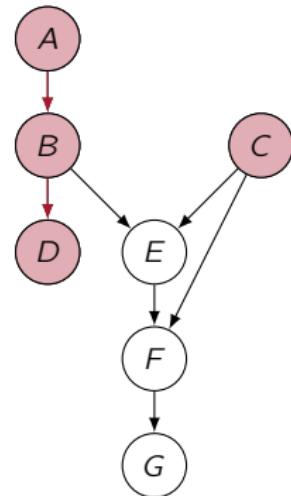
Parents: $Pa(E) = \{B, C\}$

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 $\forall X \in \mathbb{S}, An(X) \subseteq \mathbb{S}$



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

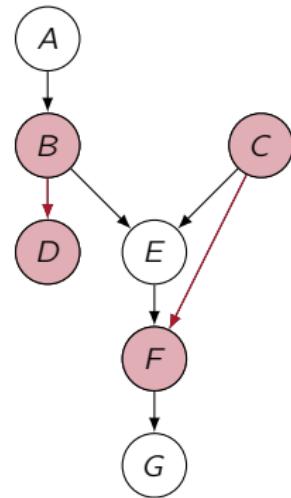
Ancestors: $An(E) = \{A, B, C, E\}$

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 $\mathcal{G}[\mathbb{S}]$: $\mathcal{G}[\{B, C, D, F\}]$



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

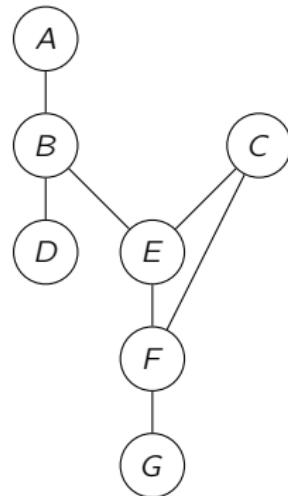
Descendants: $Desc(E) = \{E, F, G\}$

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}



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} .

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

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

$\Leftrightarrow \forall X \in \mathbb{V}, X \perp\!\!\!\perp_{\mathcal{G}} Nd(X) \mid Pa(X)$ (**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

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

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

generalization of the common cause principle

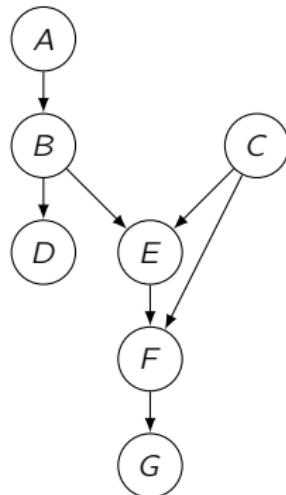
A DAG $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ is a **Bayesian network** iff there exists a joint distribution $P(\mathbb{V})$ that is compatible with \mathcal{G} .

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 $\mathcal{G}[\mathbb{S}]$.

Theorem

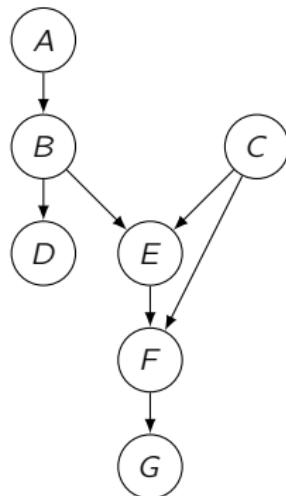
If P is compatible with \mathcal{G} and $\mathbb{S} \subseteq \mathbb{V}$ is an ancestral set, then $P(\mathbb{S})$ is compatible with $\mathcal{G}[\mathbb{S}]$.



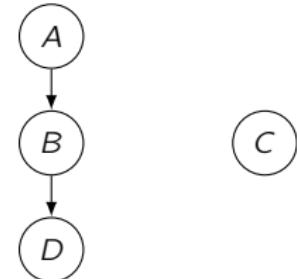
Bayesian network
compatible with P

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

1

Association

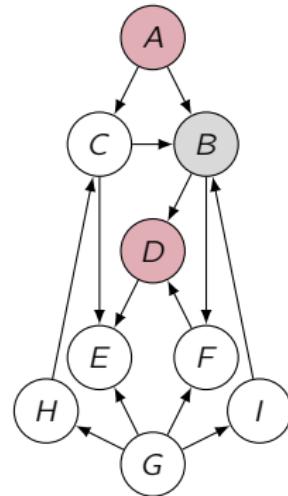
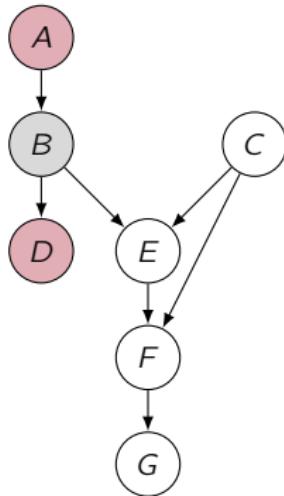
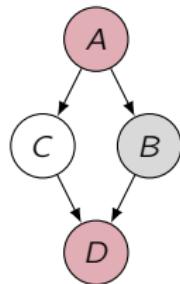
Bayesian networks

d-separation

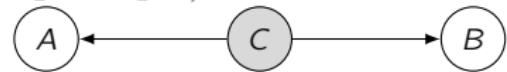
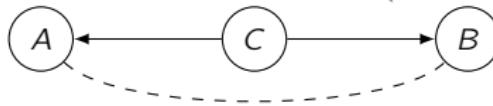
Prediction

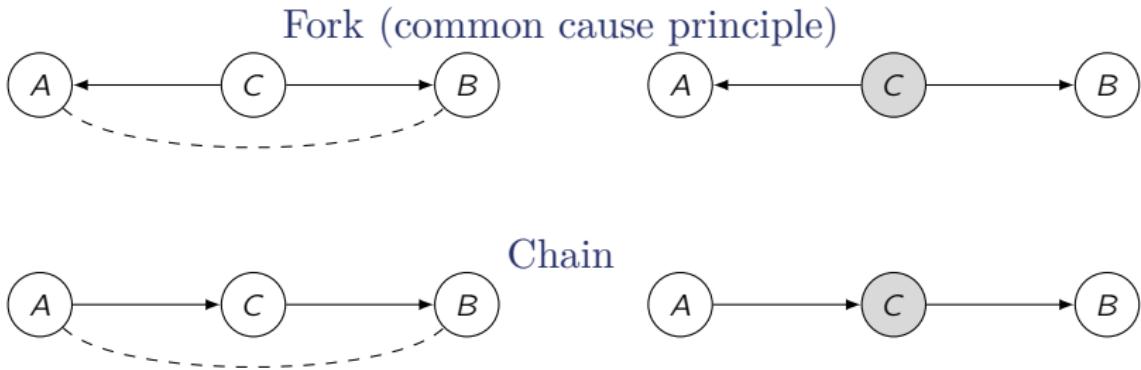
Do we need more than associations?

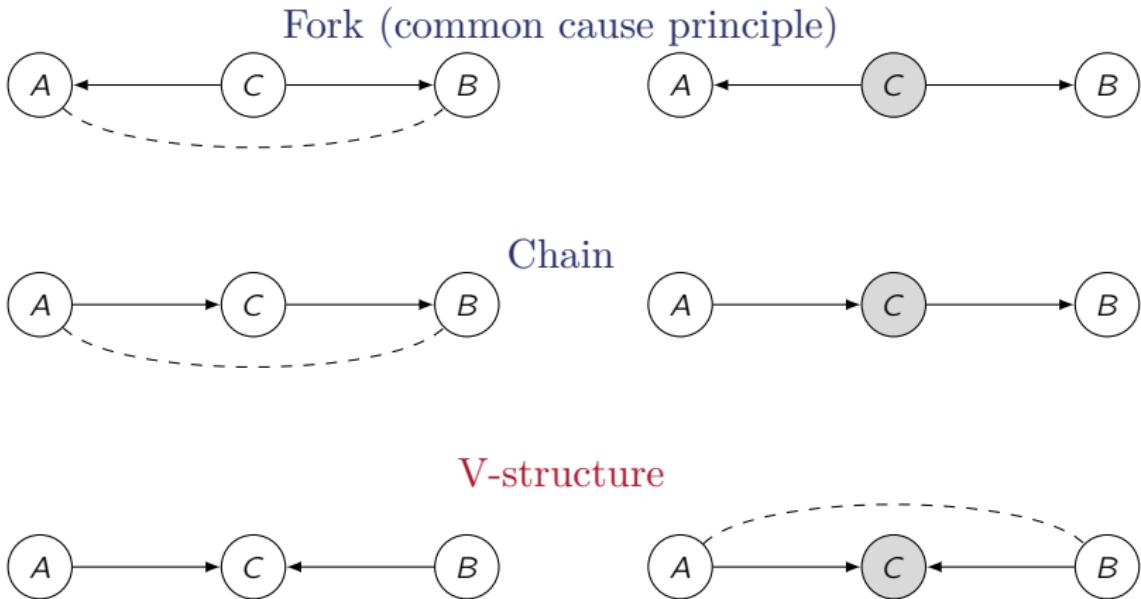
$$A \perp\!\!\!\perp_P D | B \quad ?$$



Fork (common cause principle)

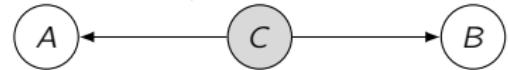
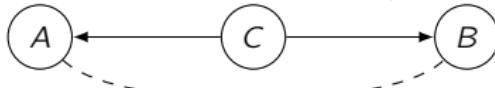




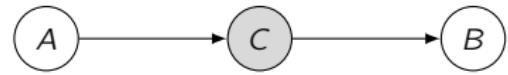
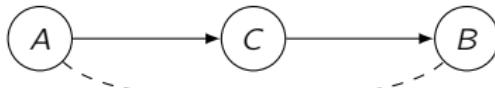


Basic structures

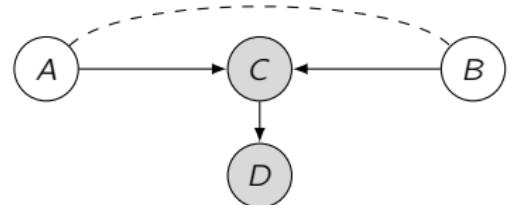
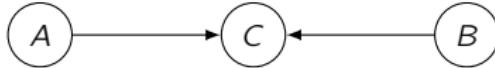
Fork (common cause principle)



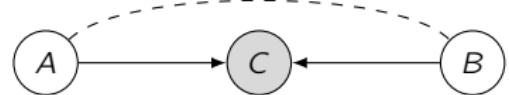
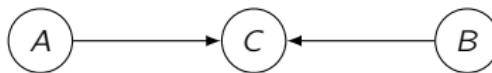
Chain



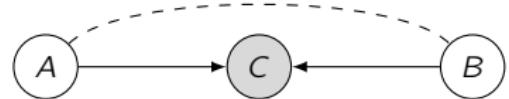
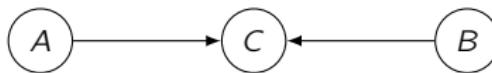
V-structure



Artificial correlation in V-structures



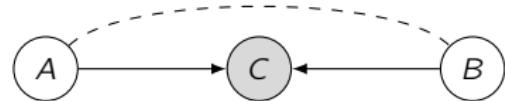
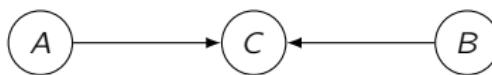
Artificial correlation in V-structures



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

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

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



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

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

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

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

- { If $A = \text{Mother not carrier}$ then $B = \text{Father carrier}$
- { If $B = \text{Father not carrier}$ then $A = \text{Mother carrier}$

Artificial correlation in V-structures



$$A = \begin{cases} \text{Asthmatic} \\ \text{Not asthmatic} \end{cases}$$

$$B = \begin{cases} \text{Sleep apnea} \\ \text{No sleep apnea} \end{cases}$$

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



$$A = \begin{cases} \text{Asthmatic} \\ \text{Not asthmatic} \end{cases}$$

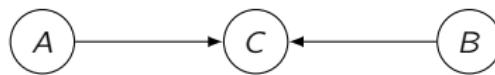
$$B = \begin{cases} \text{Sleep apnea} \\ \text{No sleep apnea} \end{cases}$$

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

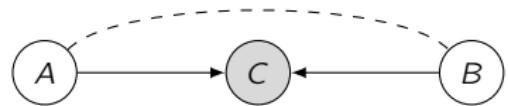
If $C = \text{Visit pulmonologist} \implies$

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

Artificial correlation in V-structures



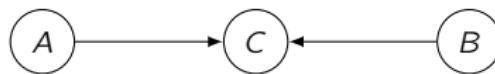
$$A, B \sim U(-1, 1)$$



$$\xi_c \sim N(0, \frac{1}{2})$$

$$C = 2AB + \xi_c$$

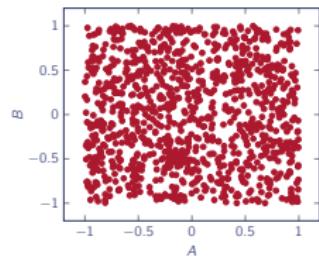
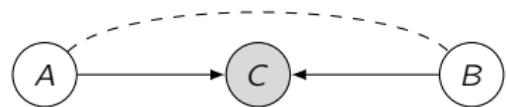
Artificial correlation in V-structures



$$A, B \sim U(-1, 1)$$

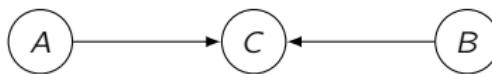
$$\xi_c \sim N(0, \frac{1}{2})$$

$$C = 2AB + \xi_c$$



$$\text{Corr}(A; B) = 0.002$$

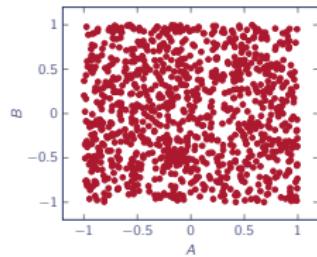
Artificial correlation in V-structures



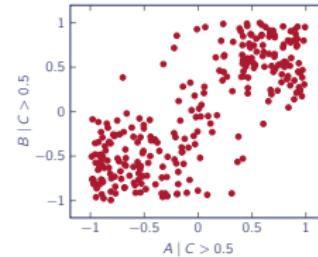
$$A, B \sim U(-1, 1)$$

$$\xi_c \sim N(0, \frac{1}{2})$$

$$C = 2AB + \xi_c$$



$$\text{Corr}(A; B) = 0.002$$



$$\text{Corr}(A; B | C > 0.5) = 0.8$$

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.

¹We also refer to Z as the collider

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} .

¹We also refer to Z as the collider

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**.

¹We also refer to Z as the collider

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

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

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

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

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

$$\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} | \mathbb{Z} \Rightarrow \mathbb{X} \perp\!\!\!\perp_P \mathbb{Y} | \mathbb{Z}$$

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

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

$$\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} | \mathbb{Z} \Rightarrow \mathbb{X} \perp\!\!\!\perp_P \mathbb{Y} | \mathbb{Z}$$

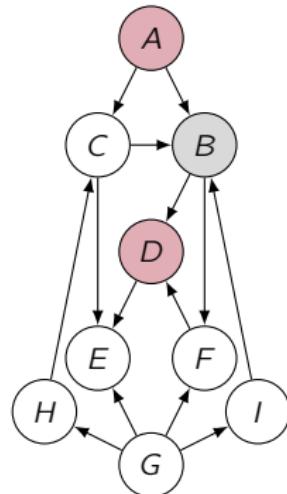
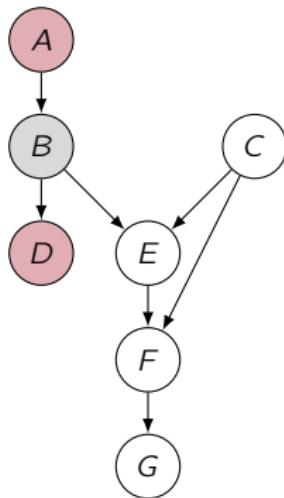
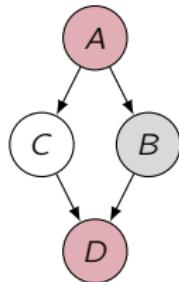
but

$$\mathbb{X} \perp\!\!\!\perp_P \mathbb{Y} | \mathbb{Z} \not\Rightarrow \mathbb{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} | \mathbb{Z}$$

$$\mathbb{X} \not\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} | \mathbb{Z} \not\Rightarrow \mathbb{X} \not\perp\!\!\!\perp_P \mathbb{Y} | \mathbb{Z}$$

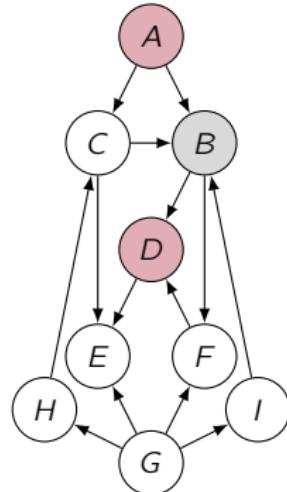
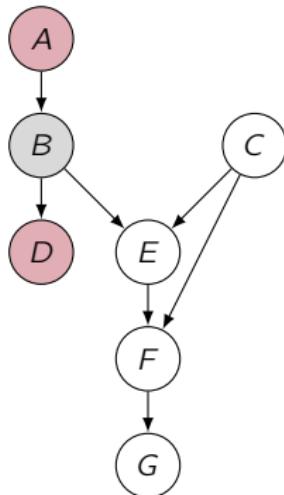
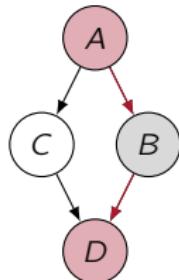
$$\mathbb{X} \not\perp\!\!\!\perp_P \mathbb{Y} | \mathbb{Z} \not\Rightarrow \mathbb{X} \not\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} | \mathbb{Z}$$

$$A \stackrel{?}{\perp\!\!\!\perp}_P D | B$$



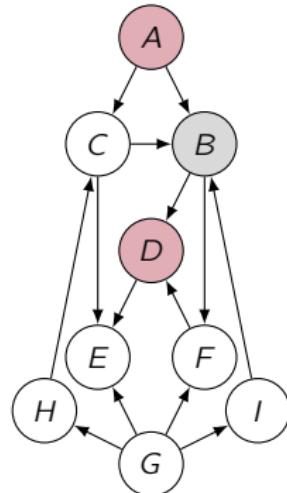
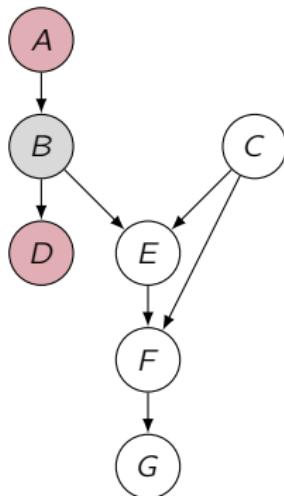
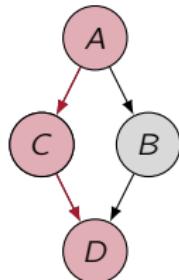
Reading conditional independencies using d-separation

$$A \perp\!\!\!\perp_P D | B$$

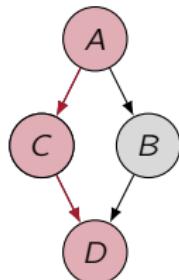


Reading conditional independencies using d-separation

$$A \stackrel{?}{\perp\!\!\!\perp}_P D | B$$

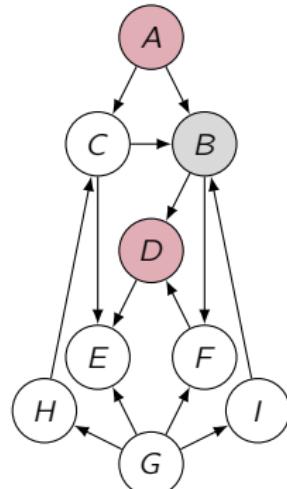
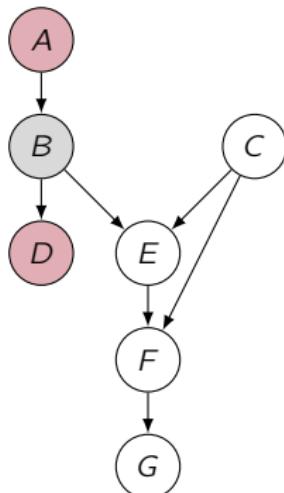


$$A \perp\!\!\!\perp_P D | B \quad ?$$

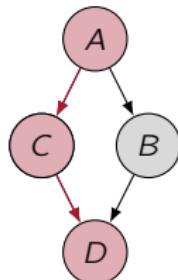


$\langle A, B, D \rangle$ is not
blocked

?
 $\implies A \perp\!\!\!\perp_P D | B$

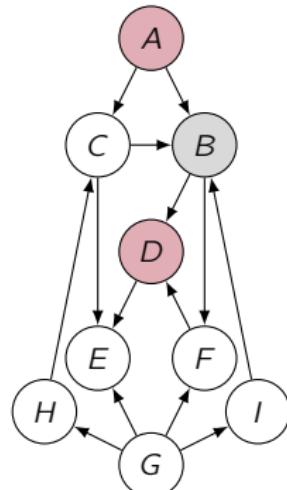
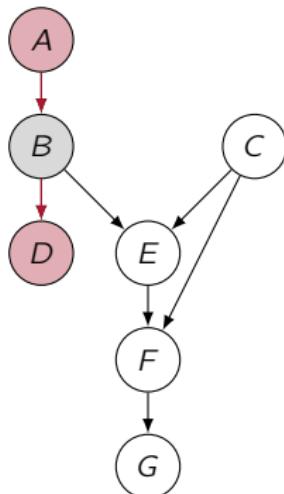


$$A \perp\!\!\!\perp_P D | B \quad ?$$

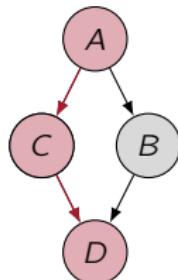


$< A, B, D >$ is not
blocked

?
 $\implies A \perp\!\!\!\perp_P D | B$

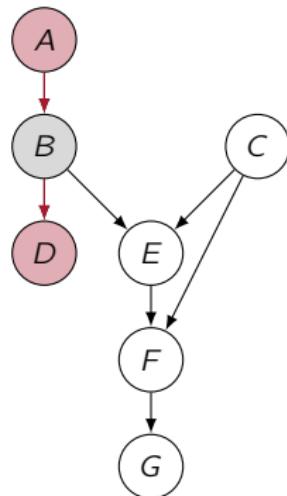


$$A \perp\!\!\!\perp_P D | B \quad ?$$

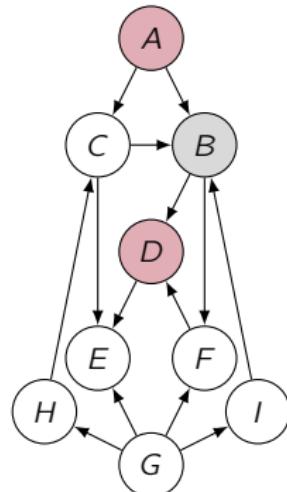


$\langle A, B, D \rangle$ is not
blocked
?

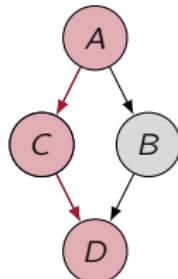
$$\implies A \perp\!\!\!\perp_P D | B$$



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

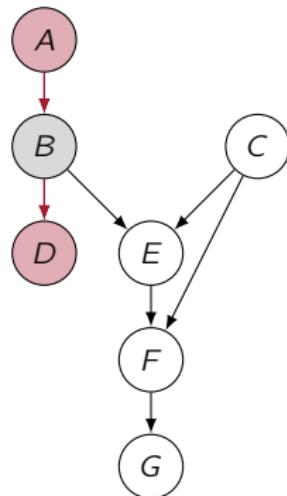


$$A \perp\!\!\!\perp_P D | B \quad ?$$

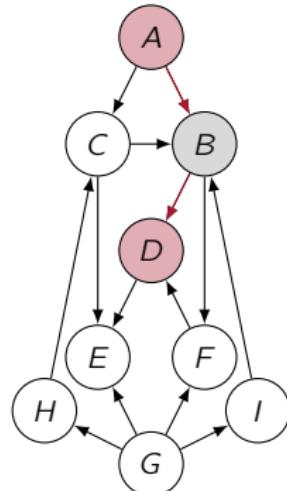


$\langle A, B, D \rangle$ is not
blocked
?

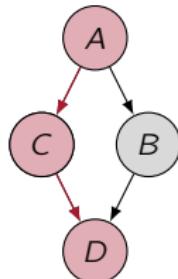
$$\implies A \perp\!\!\!\perp_P D | B$$



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

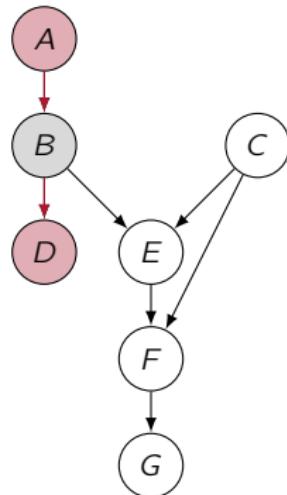


$$A \perp\!\!\!\perp_P D | B \quad ?$$

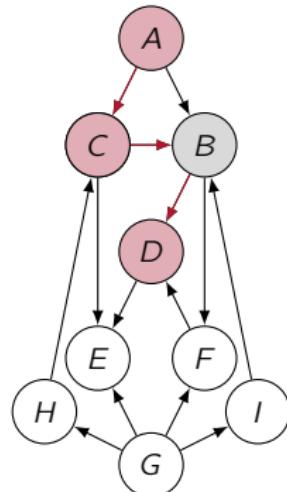


$\langle A, B, D \rangle$ is not
blocked
?

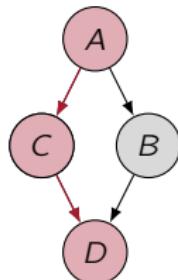
$$\implies A \perp\!\!\!\perp_P D | B$$



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

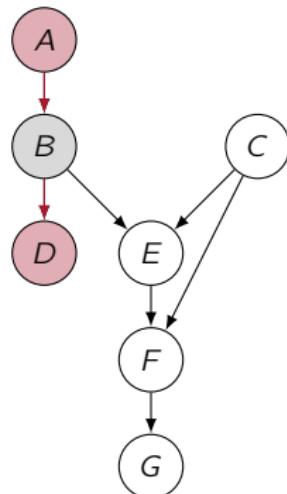


$$A \perp\!\!\!\perp_P D | B \quad ?$$

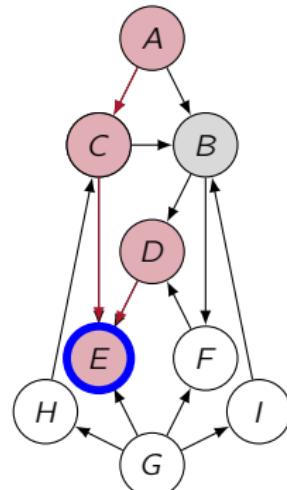


$\langle A, B, D \rangle$ is not
blocked
?

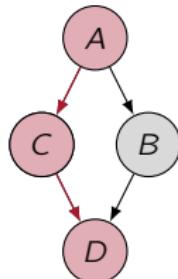
$$\implies A \perp\!\!\!\perp_P D | B$$



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

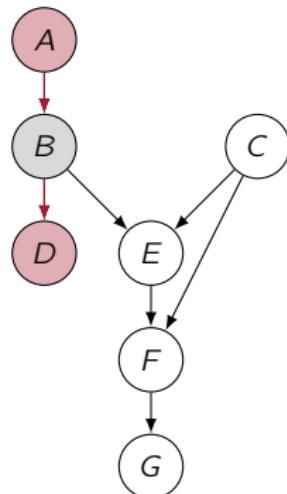


$$A \perp\!\!\!\perp_P D | B \quad ?$$

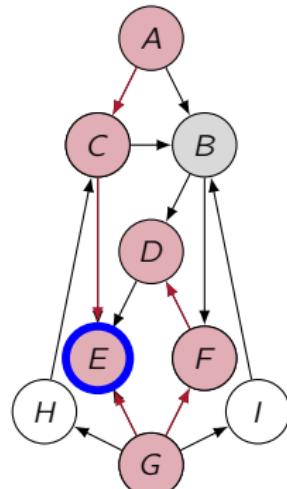


$\langle A, B, D \rangle$ is not
blocked
?

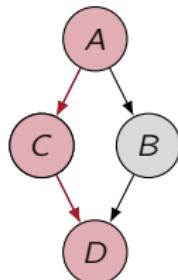
$$\implies A \perp\!\!\!\perp_P D | B$$



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

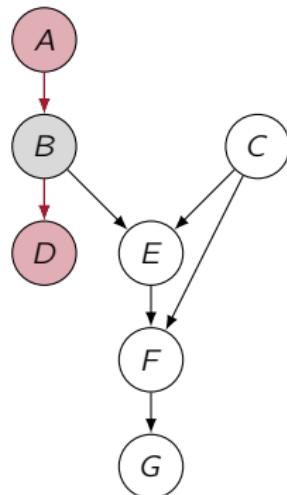


$$A \perp\!\!\!\perp_P D | B \quad ?$$

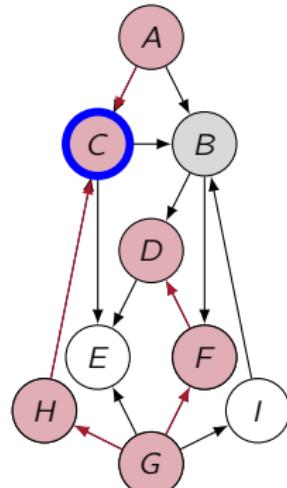


$\langle A, B, D \rangle$ is not
blocked
?

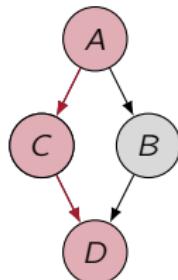
$$\implies A \perp\!\!\!\perp_P D | B$$



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

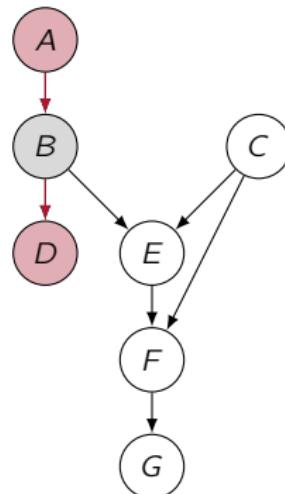


$$A \perp\!\!\!\perp_P D | B \quad ?$$

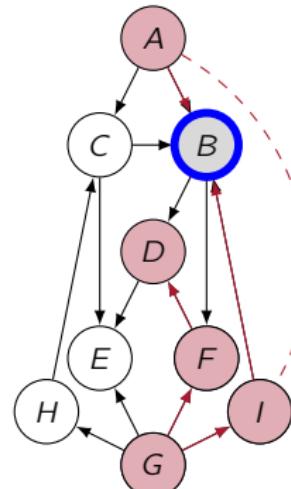


$\langle A, B, D \rangle$ is not
blocked
?

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

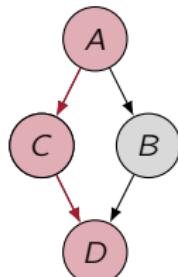


All paths are
blocked
 $\implies A \perp\!\!\!\perp_P D | B$



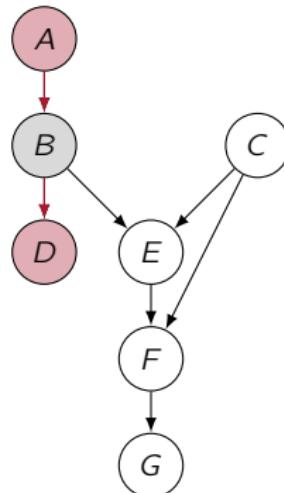
d-separation

$$A \perp\!\!\!\perp_P D | B \quad ?$$



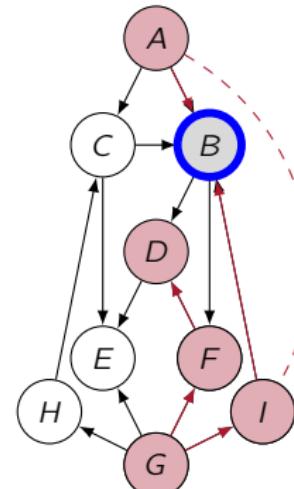
$\langle A, B, D \rangle$ is not blocked

$$\Rightarrow A \perp\!\!\!\perp_P D | B \quad ?$$



All paths are blocked

$$\Rightarrow A \perp\!\!\!\perp_P D | B$$



$\langle A, I, G, F, D \rangle$ is not blocked

$$\Rightarrow A \perp\!\!\!\perp_P D | B \quad ?$$

1

Association

Bayesian networks

d-separation

Prediction

Do we need more than associations?

Bayesian networks can be used to detect relevant variables that contains all the useful information 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)

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

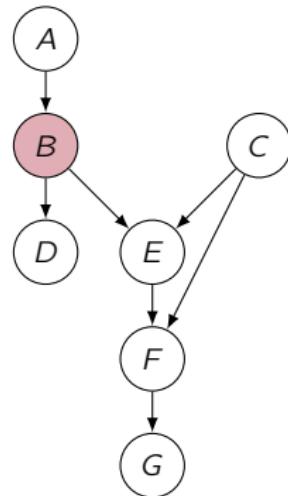
Relevant variables for predicting Y :

- 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 predicting Y :

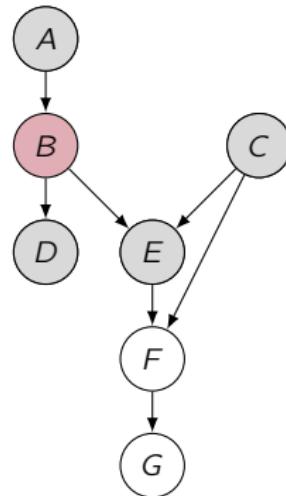
- Parents of Y ;
- Children of Y ;
- Other parents of the children of Y .



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 predicting Y :

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



1

Association

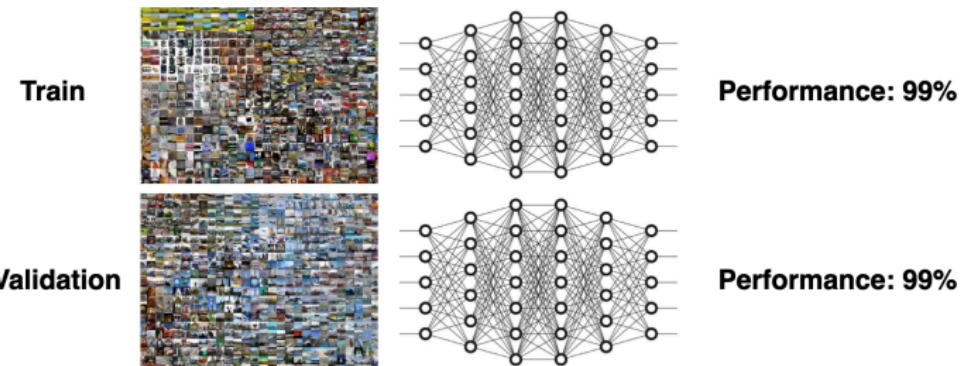
Bayesian networks

d-separation

Prediction

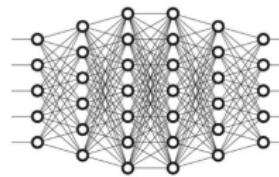
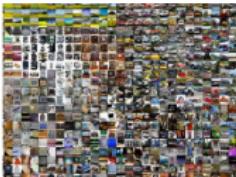
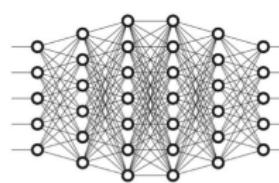
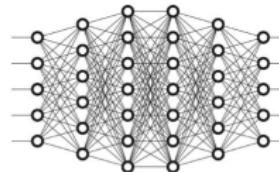
Do we need more than associations?

Distribution shifts



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

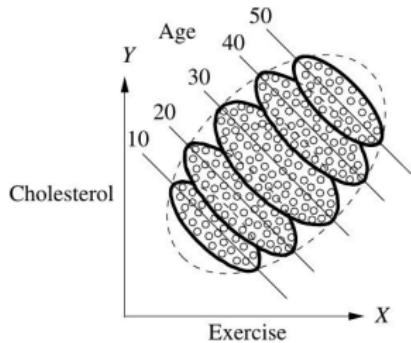
Distribution shifts

Train**Performance: 99%****Validation****Performance: 99%****New image****It's a fish :)**

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

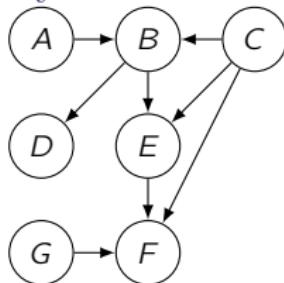
Simpson paradox [13]

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

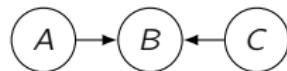


Exercise 1

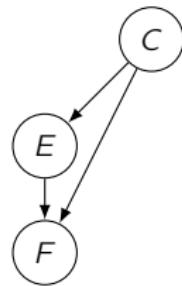
Suppose this DAG is a bayesian network compatible with P :



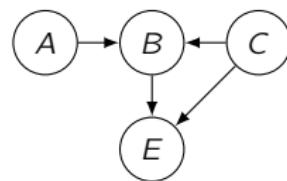
Do the following subgraphs represent Bayesian networks that are compatible with their respective distributions?



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

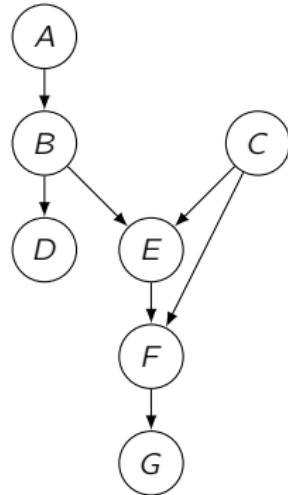


$$P(C, E, F)$$



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

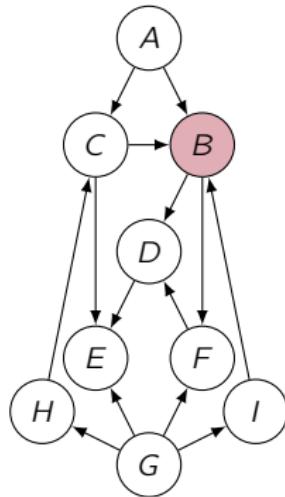
Exercise 2



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

Exercise 3

What are the relevant variables for predicting B ?



2

Intervention

Causal graphs

Causal reasoning

Causal discovery

2

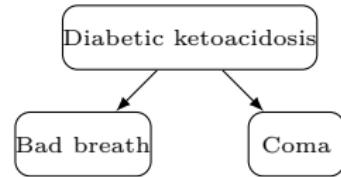
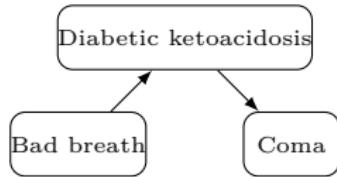
Intervention

Causal graphs

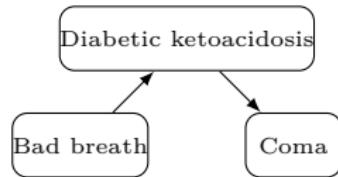
Causal reasoning

Causal discovery

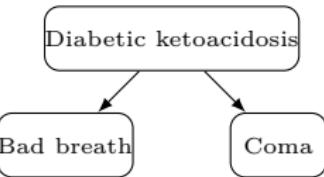
Bayesian networks vs causal graph



Bayesian networks vs causal graph

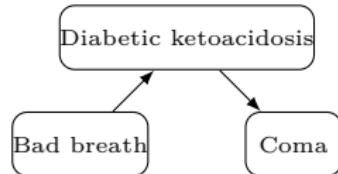


Bad breath $\perp\!\!\!\perp_P$ Coma |
Diabetic ketoacidosis
Bayesian network



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

Bayesian networks vs causal graph

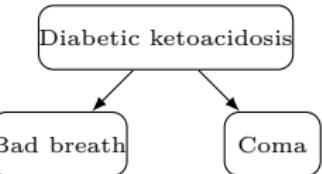


Bad breath $\perp\!\!\!\perp_P$ Coma |

Diabetic ketoacidosis

Bayesian network

Not a causal graph



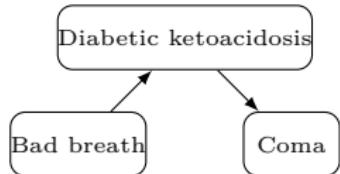
Bad breath $\perp\!\!\!\perp_P$ Coma |

Diabetic ketoacidosis

Bayesian network

Causal graph

Bayesian networks vs causal graph

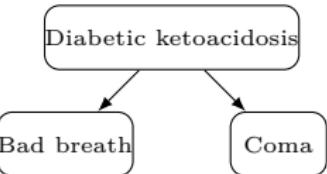


Bad breath $\perp\!\!\!\perp_P$ Coma |

Diabetic ketoacidosis

Bayesian network

Not a causal graph



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

Diabetic ketoacidosis

Bayesian network

Causal graph

Oracle for conditional
independence

Oracle for intervention

Conditioning vs Intervening (1/2)

Population

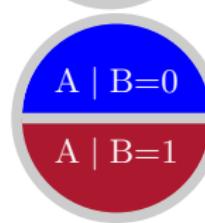


Conditioning vs Intervening (1/2)

Population



Sub-populations

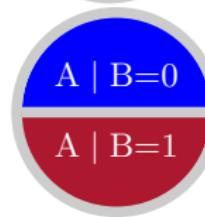


Conditioning vs Intervening (1/2)

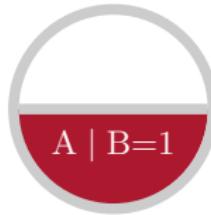
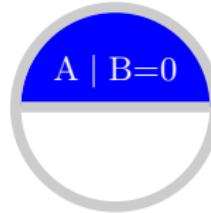
Population



Sub-populations



Conditioning

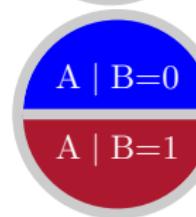


Conditioning vs Intervening (1/2)

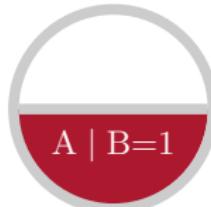
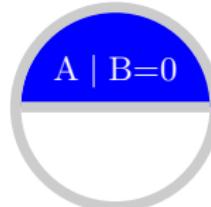
Population



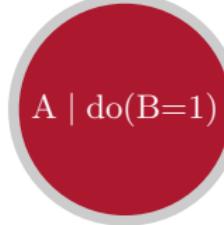
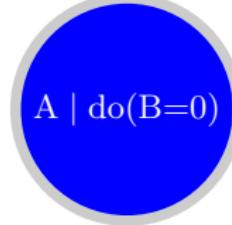
Sub-populations



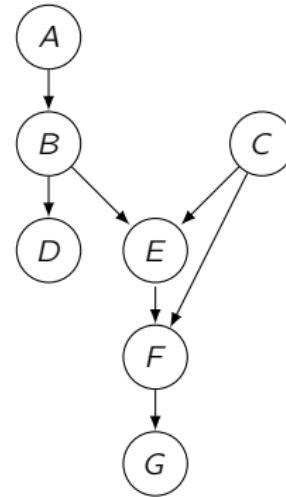
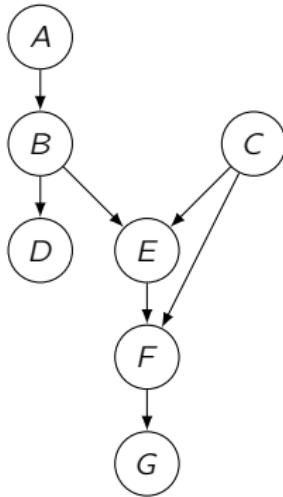
Conditioning



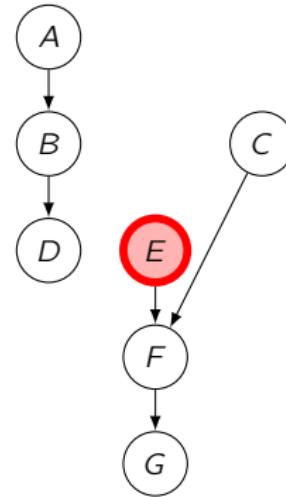
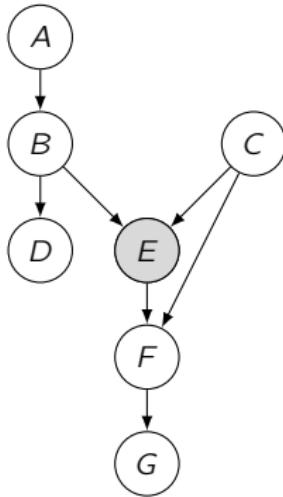
Intervening



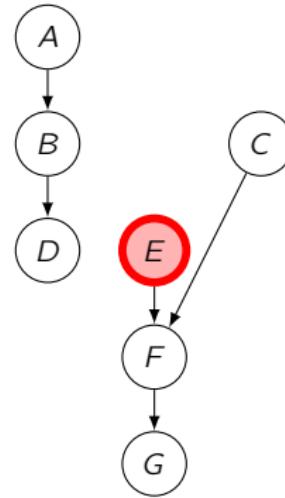
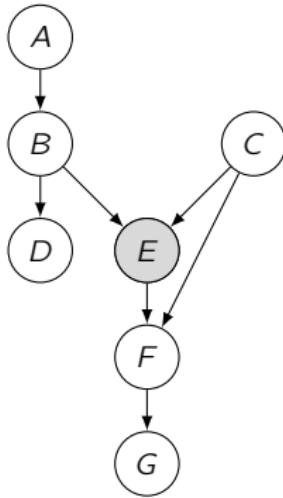
Conditioning vs Intervening (2/2)



Conditioning vs Intervening (2/2)



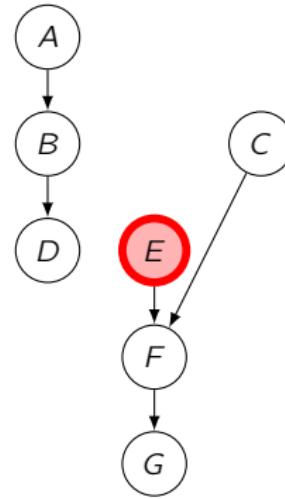
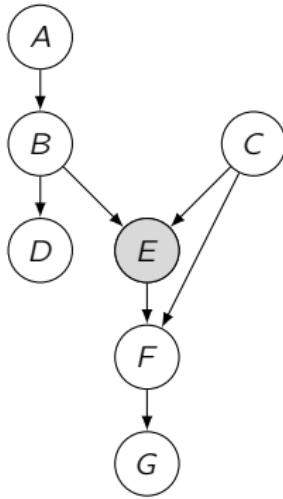
Conditioning vs Intervening (2/2)



Note that there are two types of interventions:

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

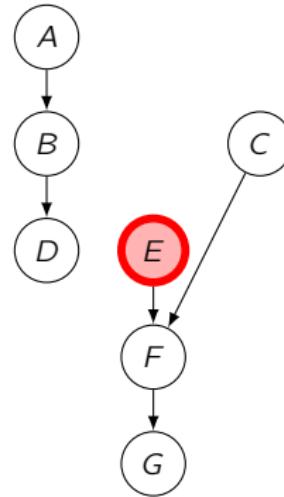
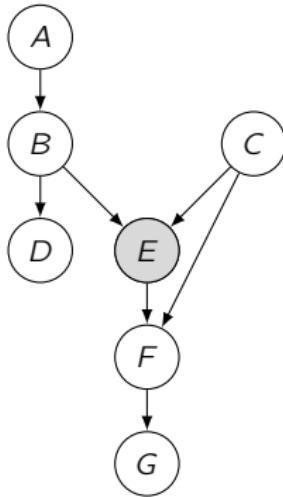
Conditioning vs Intervening (2/2)



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)

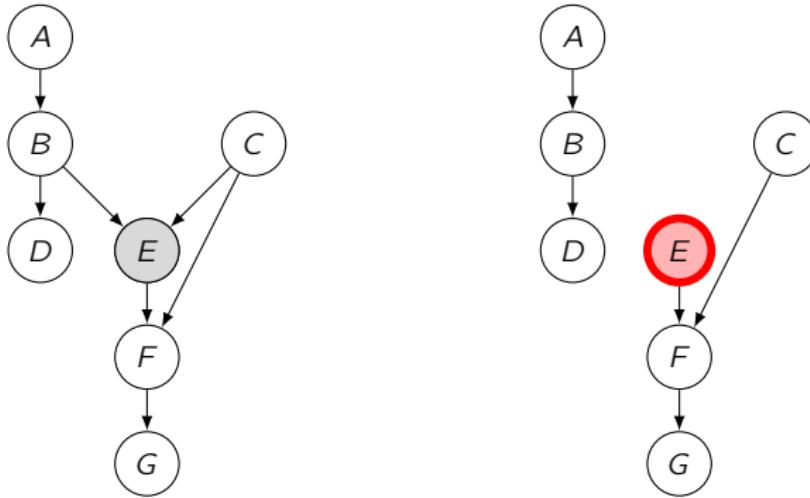


Note that there are two types of interventions:

- 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 [11]

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

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

Truncated factorization [11]

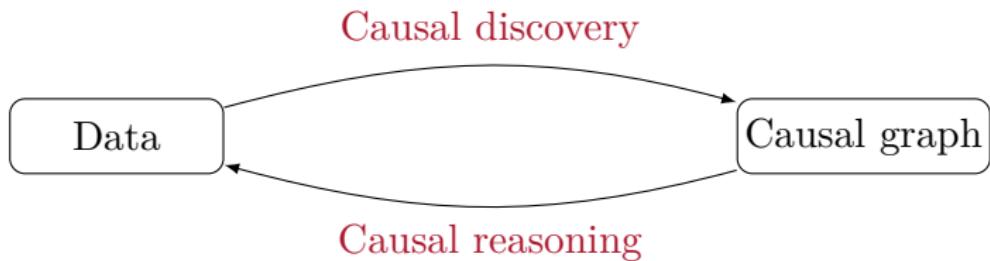
If we intervene on a subset $\mathbb{S} \subset \mathbb{V}$, then

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

Also known as G-computation formula [5] or Manipulation theorem [19].

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

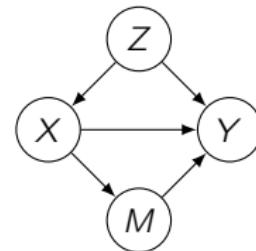
Causal graphs

Causal reasoning

Causal discovery

Different types of causal effects

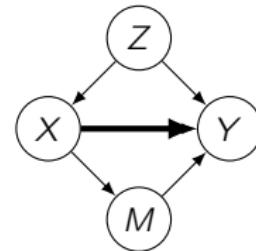
There exists three main types of causal effects:



Different types of causal effects

There exists three main types of causal effects:

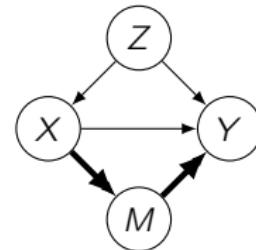
- Direct effect



Different types of causal effects

There exists three main types of causal effects:

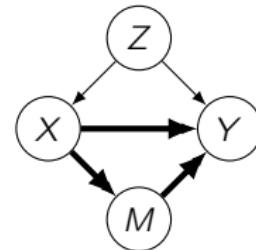
- Direct effect
- Indirect effect



Different types of causal effects

There exists three main types of causal effects:

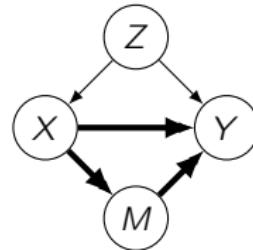
- Direct effect
- Indirect effect
- Total effect



Different types of causal effects

There exists three main types of causal effects:

- Direct effect
- Indirect effect
- Total effect



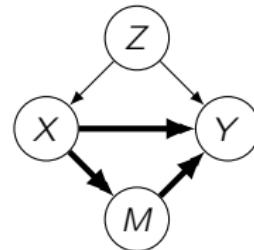
Total effect of X on $Y = P(y | do(x)) - P(y | do(x'))$

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



Total effect of X 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.

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 $P(y | do(x))$ is uniquely computable from $P(\mathbb{V})$.

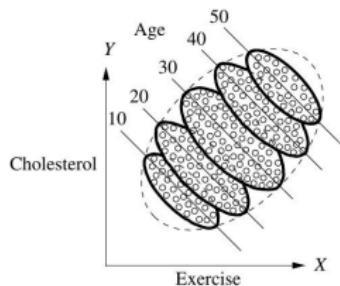
Identifiability

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 $P(y | do(x))$ is uniquely computable from $P(\mathbb{V})$.

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

Simpson paradox 1 and a simple solution

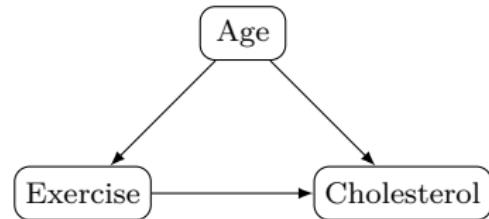
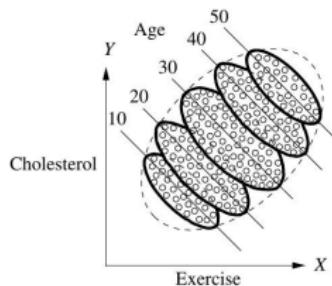
In a study, we measure weekly exercise and cholesterol levels for various age groups.



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

Simpson paradox 1 and a simple solution

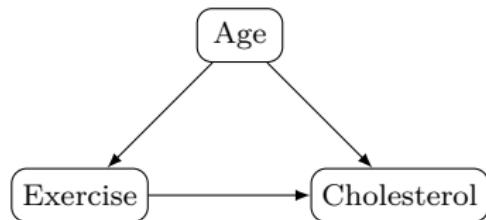
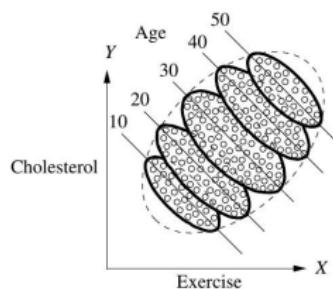
In a study, we measure weekly exercise and cholesterol levels for various age groups.



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.



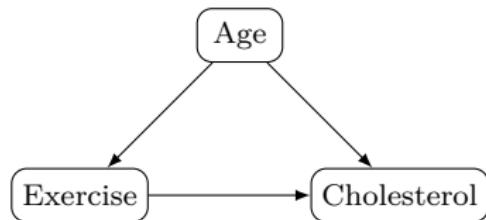
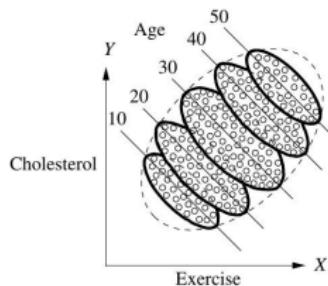
What is the effect of exercise on cholesterol $P(c \mid \text{do}(e))$?

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

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

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 $P(c | \text{do}(e))$?

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

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

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

Theorem

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

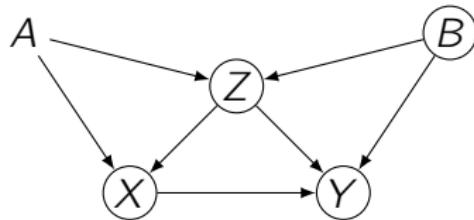
$$P(y \mid \text{do}(x)) = \sum_{z \in \text{Parents}(x)} P(y \mid x, z) \Pr(z)$$

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?



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

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

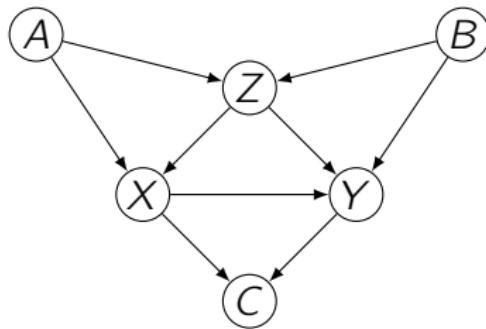
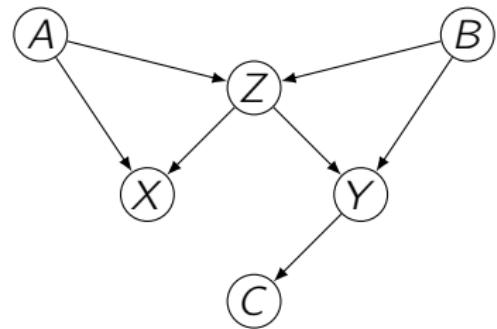
Back-door criterion: using d-separation

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 $P(y \mid \text{do}(x))$. A set of variables Z satisfies the **back-door criterion** iff:

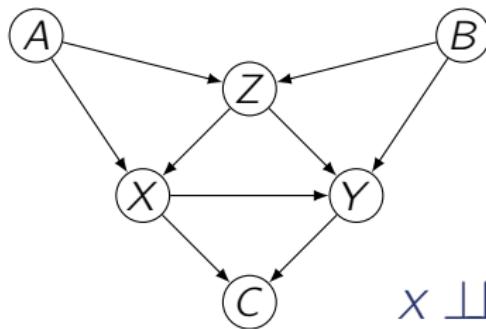
- $X \perp\!\!\!\perp_{\mathcal{G}_{\underline{X}}} Y \mid Z$

Back-door criterion: example

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

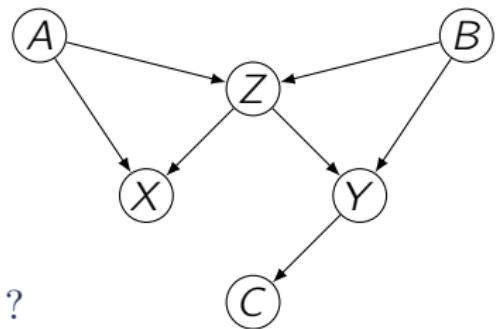
Back-door criterion: example

Causal graph \mathcal{G}



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

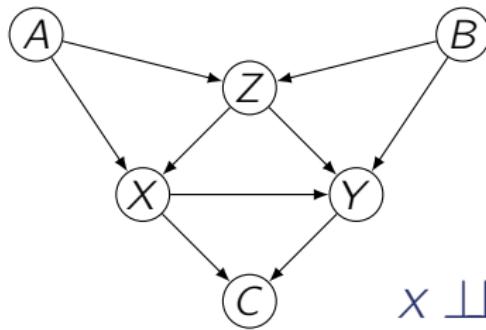
Mutilated graph \mathcal{G}_X



$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z ?$

Back-door criterion: example

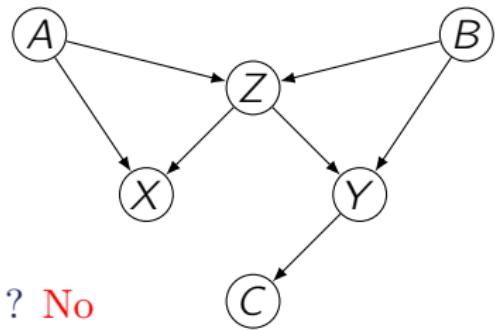
Causal graph \mathcal{G}



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

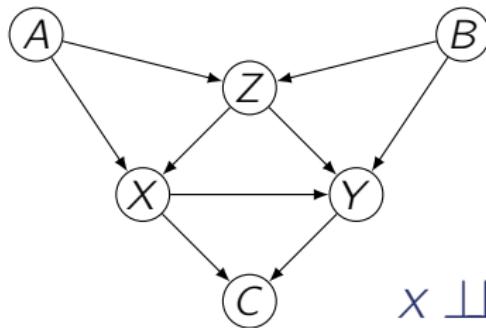
$$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z ? \text{No}$$

Mutilated graph \mathcal{G}_X



Back-door criterion: example

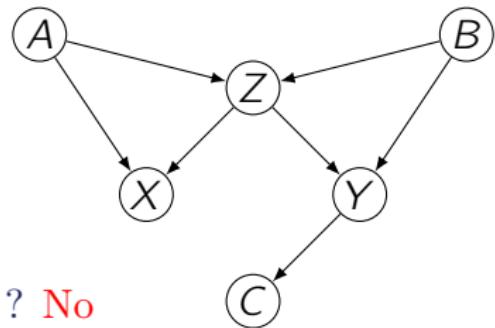
Causal graph \mathcal{G}



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

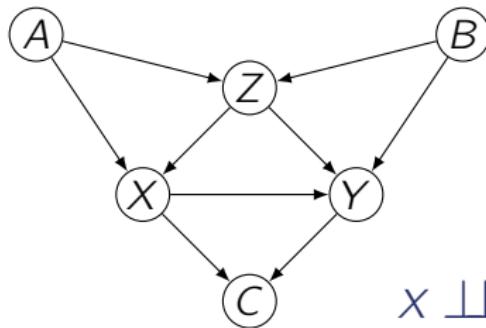
$$\begin{aligned} & x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z ? \text{No} \\ & x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A ? \end{aligned}$$

Mutilated graph \mathcal{G}_X



Back-door criterion: example

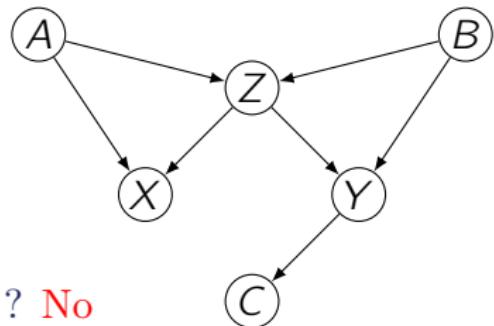
Causal graph \mathcal{G}



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

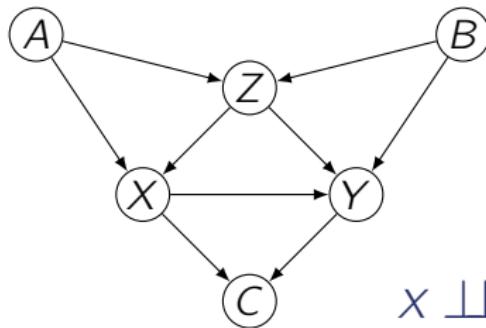
$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A$? No

Mutilated graph \mathcal{G}_X



Back-door criterion: example

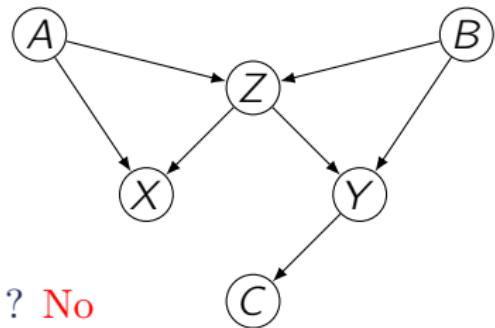
Causal graph \mathcal{G}



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

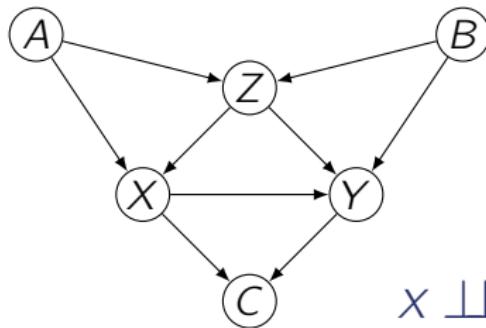
$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid B$?

Mutilated graph \mathcal{G}_X



Back-door criterion: example

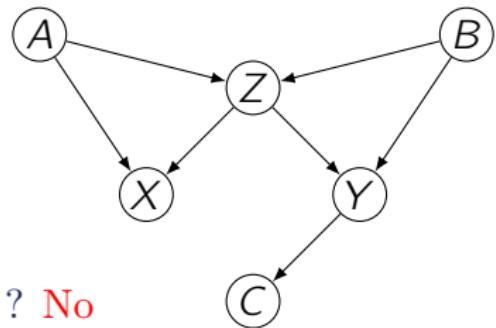
Causal graph \mathcal{G}



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

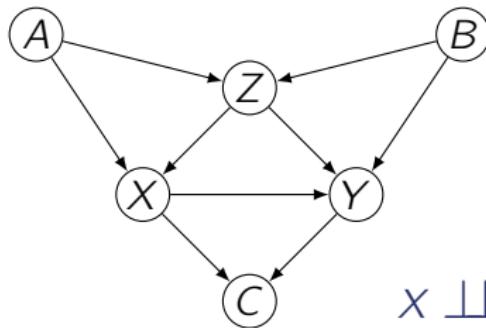
$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid B$? No

Mutilated graph \mathcal{G}_X



Back-door criterion: example

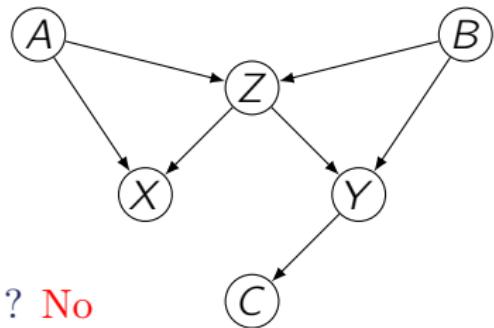
Causal graph \mathcal{G}



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

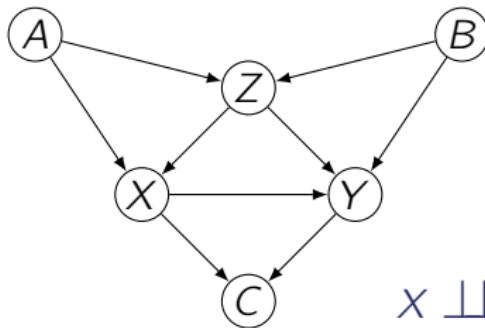
$$\begin{aligned}
 & x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z ? \text{ No} \\
 & x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A ? \text{ No} \\
 & x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid B ? \text{ No} \\
 & x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid C ?
 \end{aligned}$$

Mutilated graph \mathcal{G}_X



Back-door criterion: example

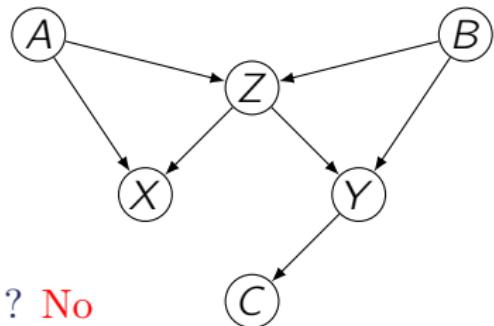
Causal graph \mathcal{G}



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

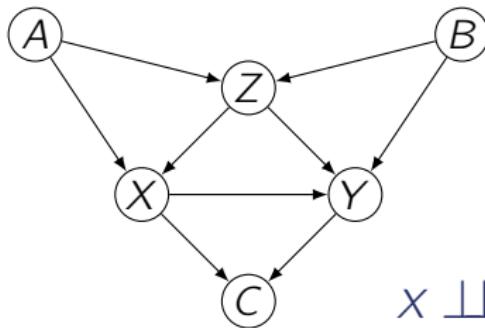
$$\begin{aligned}
 & X \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z ? \text{No} \\
 & X \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A ? \text{No} \\
 & X \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid B ? \text{No} \\
 & X \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid C ? \text{No}
 \end{aligned}$$

Mutilated graph \mathcal{G}_X



Back-door criterion: example

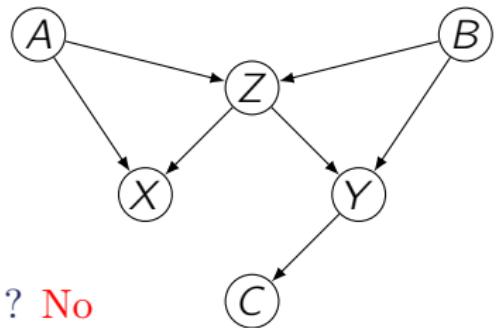
Causal graph \mathcal{G}



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

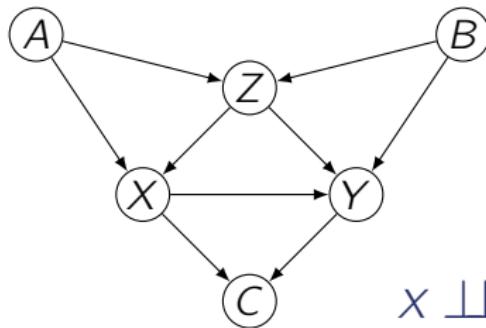
$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid B$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid C$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A, B$?

Mutilated graph \mathcal{G}_X



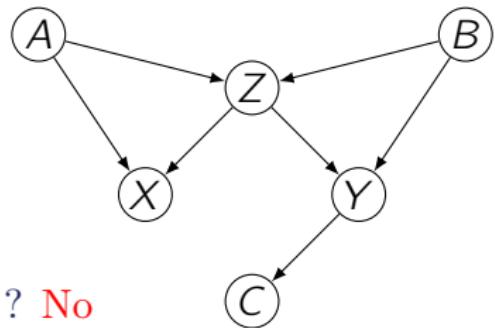
Back-door criterion: example

Causal graph \mathcal{G}



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

Mutilated graph \mathcal{G}_X



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

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

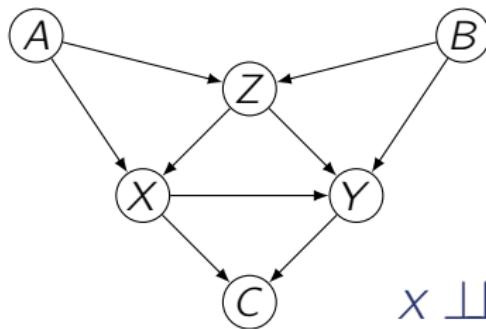
$X \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid B ?$ No

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

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

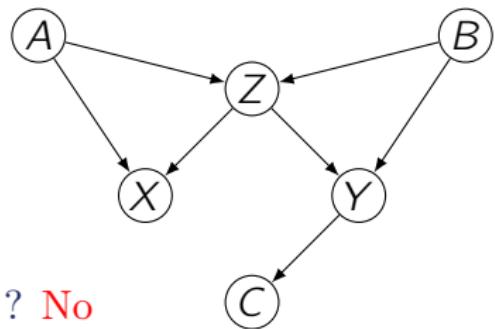
Back-door criterion: example

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

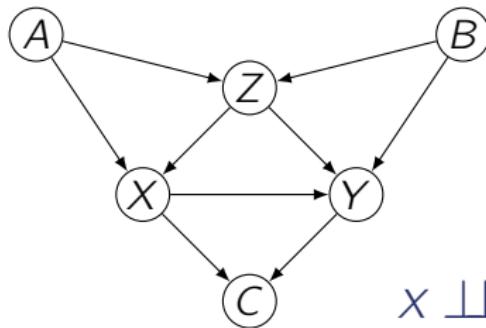
$X \perp\!\!\!\perp_{\mathcal{G}_\underline{x}} Y \mid C ?$ No

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

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

Back-door criterion: example

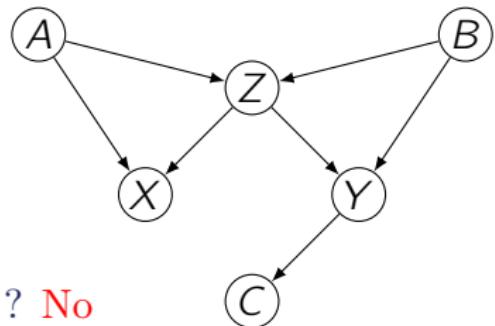
Causal graph \mathcal{G}



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

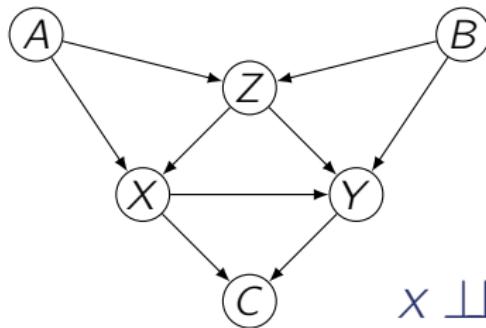
$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid B$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid C$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A, B$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z, A$? Yes

Mutilated graph \mathcal{G}_X



Back-door criterion: example

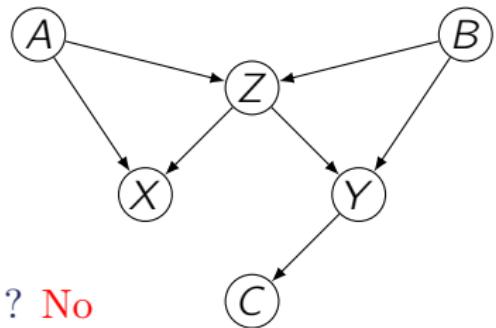
Causal graph \mathcal{G}



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

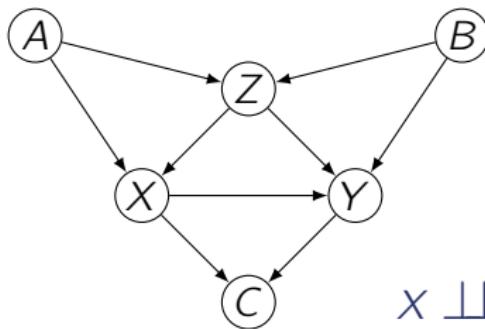
$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid B$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid C$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A, B$? No
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z, A$? Yes
 $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z, B$?

Mutilated graph \mathcal{G}_X



Back-door criterion: example

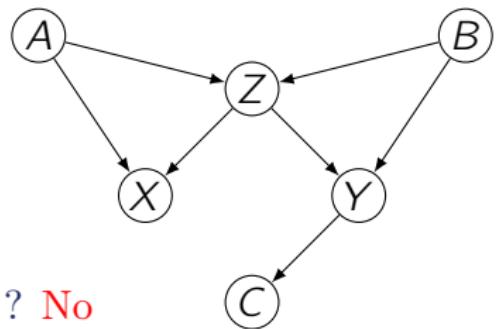
Causal graph \mathcal{G}



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

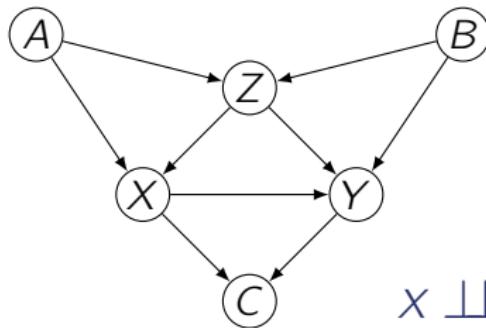
- $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z ?$ No
- $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A ?$ No
- $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid B ?$ No
- $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid C ?$ No
- $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A, B ?$ No
- $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z, A ?$ Yes
- $x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z, B ?$ Yes

Mutilated graph \mathcal{G}_X



Back-door criterion: example

Causal graph \mathcal{G}



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

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z ?$ No

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A ?$ No

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid B ?$ No

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid C ?$ No

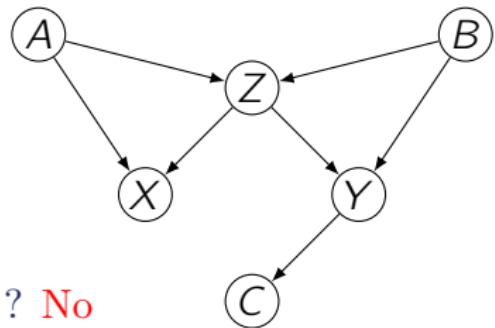
$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A, B ?$ No

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z, A ?$ Yes

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z, B ?$ Yes

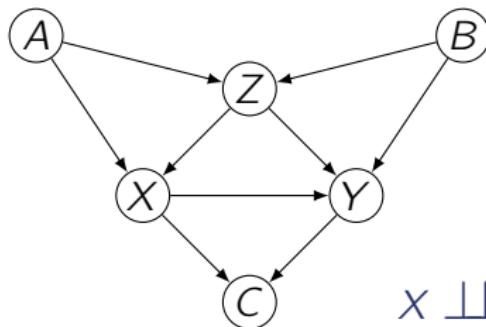
$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z, A, B ?$

Mutilated graph \mathcal{G}_X



Back-door criterion: example

Causal graph \mathcal{G}



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

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z ?$ No

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A ?$ No

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid B ?$ No

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid C ?$ No

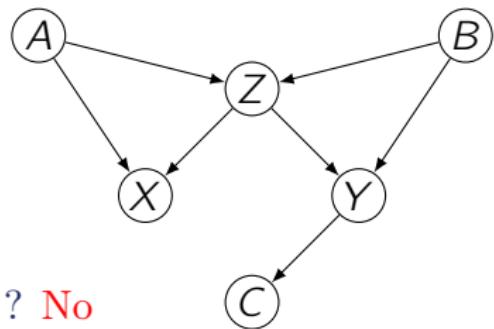
$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid A, B ?$ No

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z, A ?$ Yes

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z, B ?$ Yes

$x \perp\!\!\!\perp_{\mathcal{G}_X} Y \mid Z, A, B ?$ Yes

Mutilated graph \mathcal{G}_X



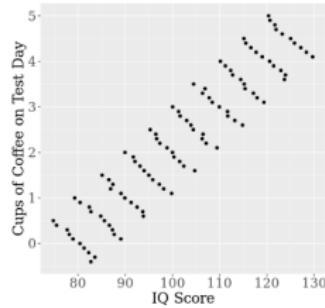
Theorem

If 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

$$P(y \mid \textcolor{red}{do}(x)) = \sum_z P(y \mid x, z)P(z).$$

Simpson paradox 2 and the back-door in action

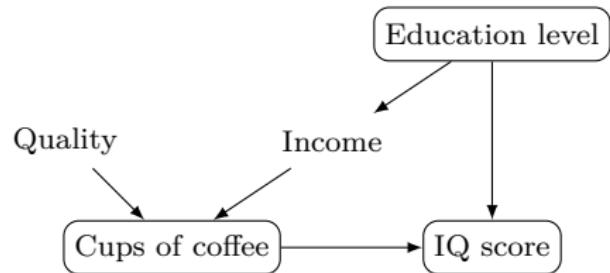
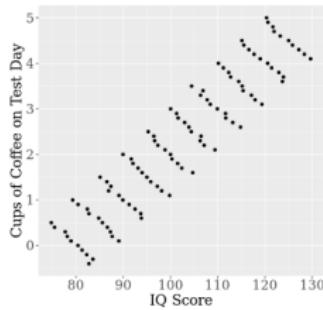
In a study, we measure the number of coffee intake and IQ score for a sample of a population with various education level.



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

Simpson paradox 2 and the back-door in action

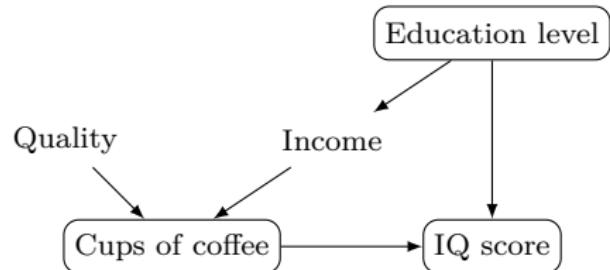
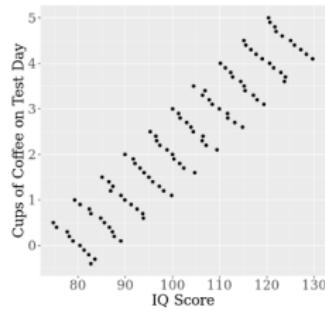
In a study, we measure the number of coffee intake and IQ score for a sample of a population with various education level.



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

Simpson paradox 2 and the back-door in action

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
 $P(i | \text{do}(c))$?

$$P(i | \text{do}(c)) = \sum_e P(i | c, e)P(e)$$

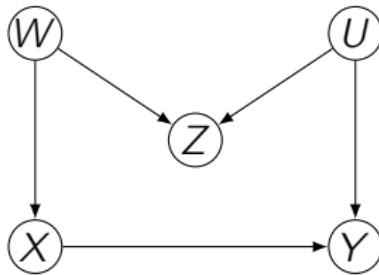
Let \mathbb{Z} be the set of observed variables in a problem that are not affected by X . \mathbb{Z} satisfies the associational criterion if each member Z of \mathbb{Z} satisfied at least one of the following conditions:

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

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 .

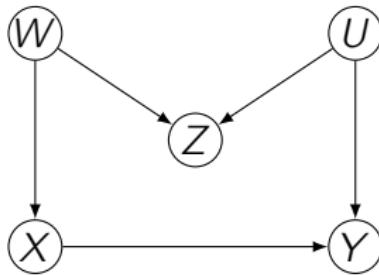
Counter-example



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

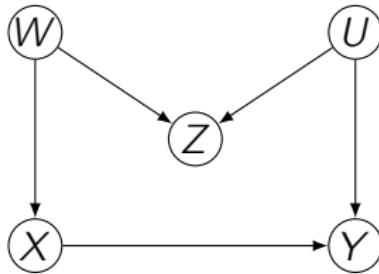


Z is associated with X and Z is associated with Y , conditional on X

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

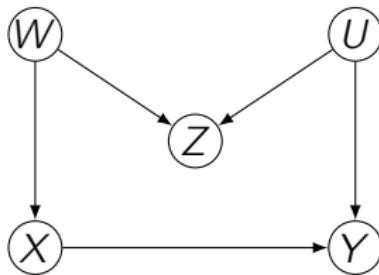


Z is associated with X and Z is associated with Y , conditional on X
 $\implies Z$ should be adjusted on

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



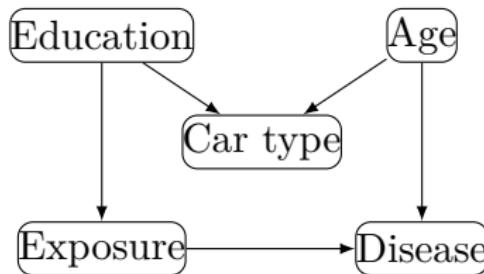
Z is associated with X and Z is associated with Y , conditional on X

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

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
~~⇒ 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 divided into two sets \mathbb{Z}_1 and \mathbb{Z}_2 that satisfy the following conditions:

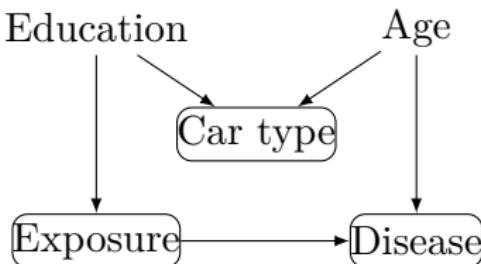
- \mathbb{Z}_1 is not associated with X ; and
- \mathbb{Z}_2 is not associated with Y , conditional on X and \mathbb{Z}_1 .

Modified associational criterion for finding non-confounding

Let Z be the set of variables in a problem that are not affected by X . Z satisfies the associational criterion if it can be divided into two sets Z_1 and Z_2 that satisfy the following conditions:

- Z_1 is not associated with X ; and
- Z_2 is not associated with Y , conditional on X and Z_1 .

Counter-example

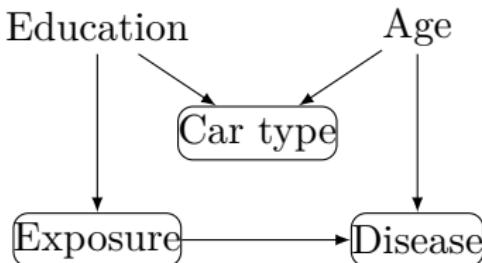


Modified associational criterion for finding non-confounding

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 divided 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



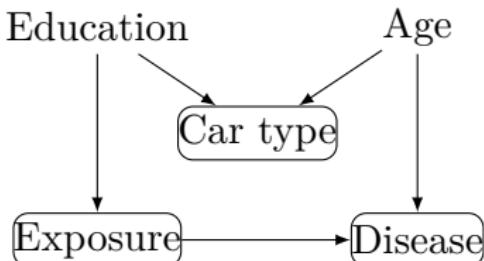
Car type is associated with Exposure

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 divided 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



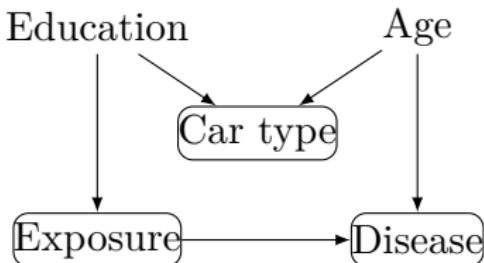
Car type is associated with Exposure
 \implies Car type should be adjusted on

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 divided 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



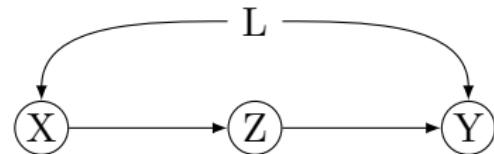
Car type is associated with Exposure

~~Car type should be adjusted on~~

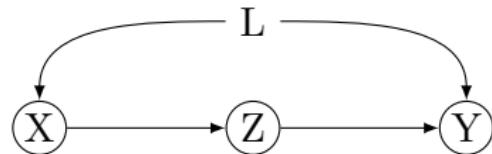
- The back-door criterion is sound but not complete:

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

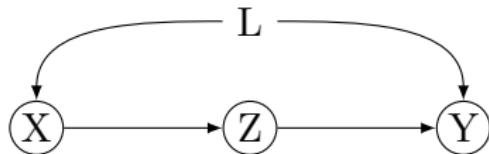
Going beyond the back-door



Going beyond the back-door

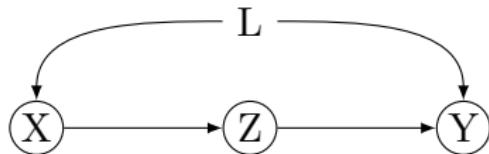


- $P(z \mid \textcolor{red}{do}(x)) = P(z \mid x)$ (No back-door path)



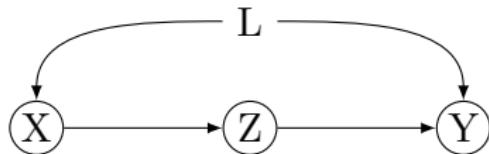
- $P(z | \text{do}(x)) = P(z | x)$ (No back-door path)
- $P(y | \text{do}(z)) = \sum_x P(y | z, x)P(x)$ (x blocks the back-door)

Going beyond the back-door



- $P(z | \text{do}(x)) = P(z | x)$ (No back-door path)
- $P(y | \text{do}(z)) = \sum_x P(y | z, x)P(x)$ (x blocks the back-door)

$$P(y | \text{do}(x)) = \sum_z P(y | \text{do}(z))P(z | \text{do}(x))$$



- $P(z | \text{do}(x)) = P(z | x)$ (No back-door path)
- $P(y | \text{do}(z)) = \sum_x P(y | z, x)P(x)$ (x blocks the back-door)

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

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

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

Theorem ([10])

If 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

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

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

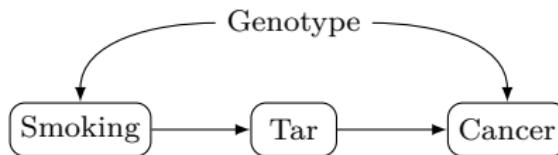
Simpson paradox 3 and the front-door in action

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 $P(c | do(s))$?



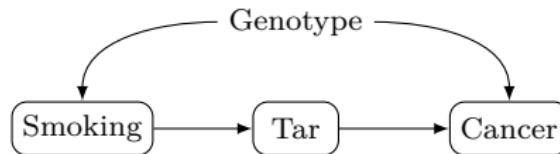
Simpson paradox 3 and the front-door in action

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 $P(c | do(s))$?



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

Incompleteness of the front-door criterion

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

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

- The front-door criterion is sound but not complete:
 - ▶ If there exists a set that satisfy the front-door criterion for $P(y | \text{do}(x))$, then $P(y | \text{do}(x))$ is identifiable;
 - ▶ If there exists a no set that satisfy the front-door criterion for $P(y | \text{do}(x))$, then $P(y | \text{do}(x))$ 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 [10]

The do calculus consists of three rules :

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

The do calculus consists of three rules :

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

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

The do calculus consists of three rules :

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

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

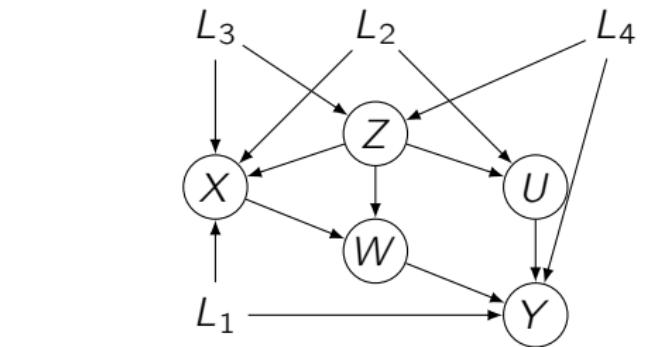
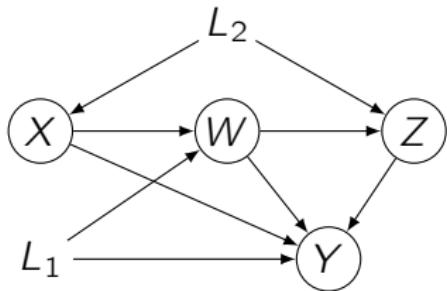
The do calculus is sound and complete!

The 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!

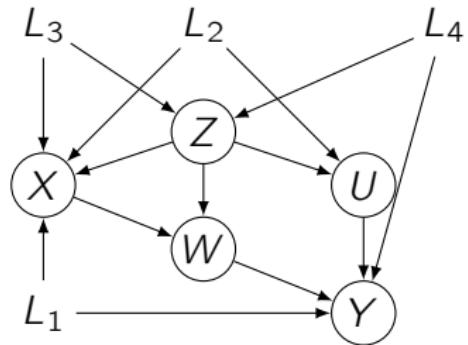
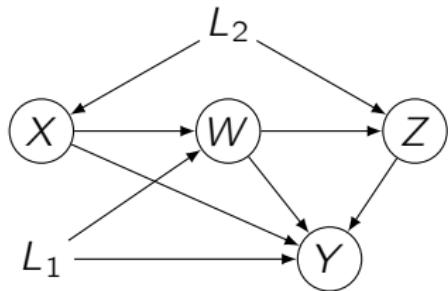


The do calculus consists of three rules :

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

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

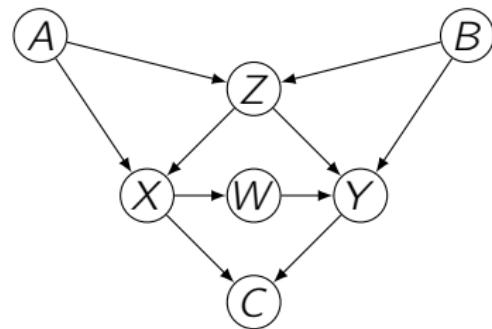
The do calculus is sound and complete!



The ID algorithm [18] uses the do-calculus to automatically uncover the do-free expression.

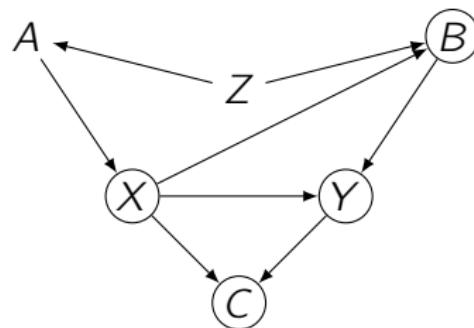
Exercise 4

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 $P(y | \text{do}(x))$;

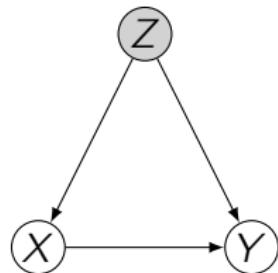


Exercise 6

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

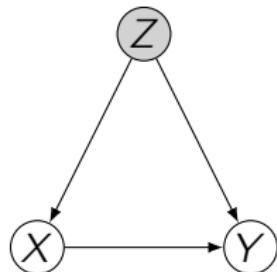
Exercise 6

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



Exercise 6

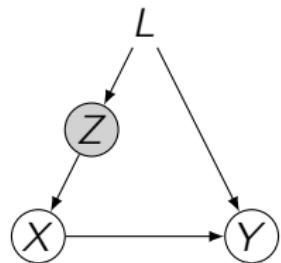
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(x))$?



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

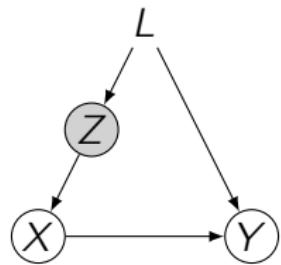
Exercise 6

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



Exercise 6

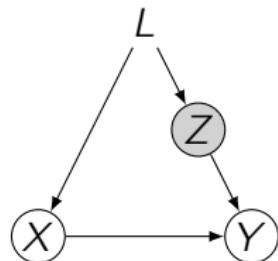
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y | do(x))$?



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

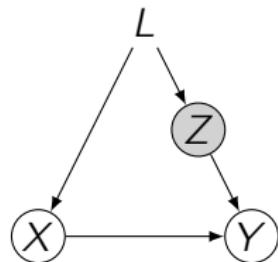
Exercise 6

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



Exercise 6

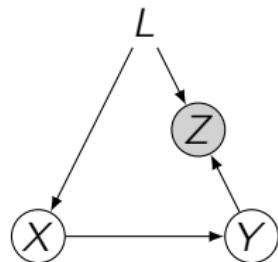
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y | do(x))$?



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

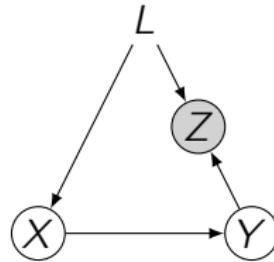
Exercise 6

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



Exercise 6

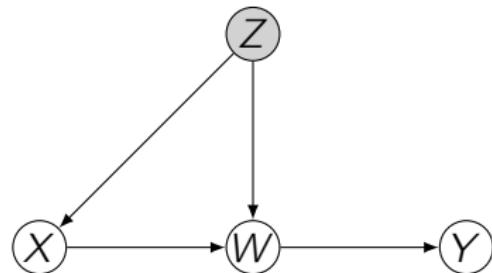
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(x))$?



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

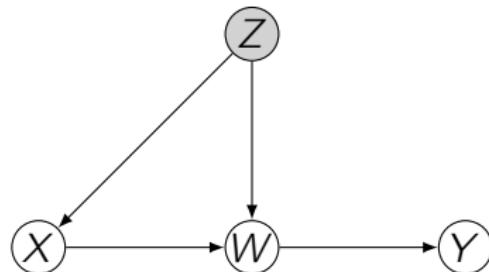
Exercise 6

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



Exercise 6

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



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

Exercise 6

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



Exercise 6

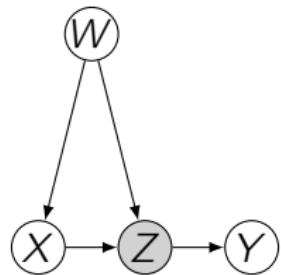
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y | \text{do}(x))$?



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

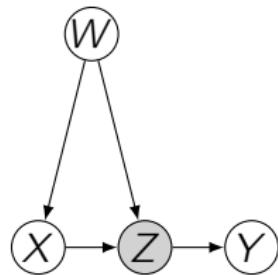
Exercise 6

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



Exercise 6

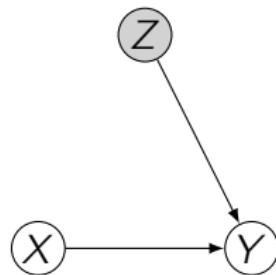
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y | do(x))$?



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

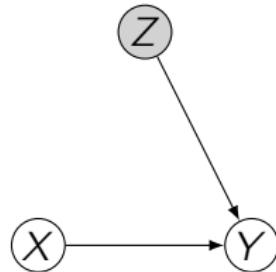
Exercise 6

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



Exercise 6

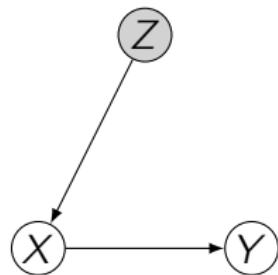
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y | do(x))$?



- Z does not open any back-door paths from X to Y
 $\implies \{Z\}$ is a neutral adjustment set;
- Adjusting on Z can reduce the variation of Y , and helps improve the precision of the estimate in finite samples.

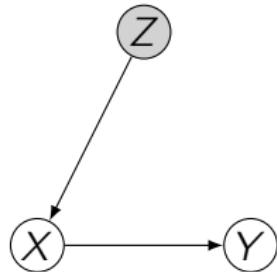
Exercise 6

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



Exercise 6

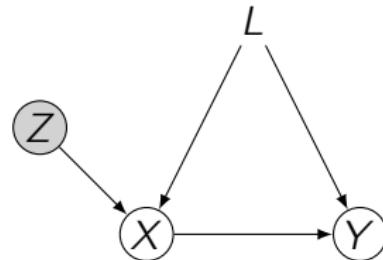
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y | do(x))$?



- Z does not open any back-door paths from X to Y
 $\implies \{Z\}$ is a neutral adjustment set;
- Adjusting on Z can reduce the variation of X and so may hurt the precision of the estimate in finite samples.

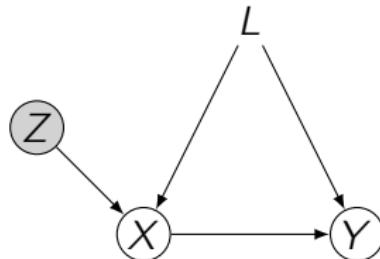
Exercise 6

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



Exercise 6

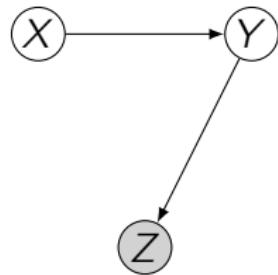
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y | do(x))$?



- 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.

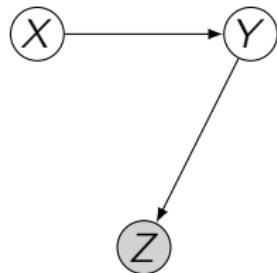
Exercise 6

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



Exercise 6

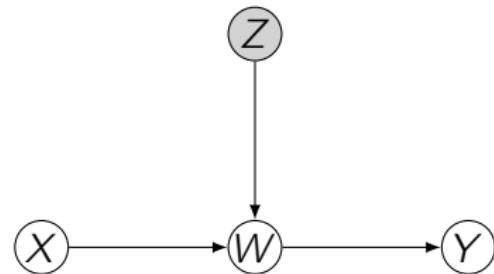
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y \mid do(x))$?



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

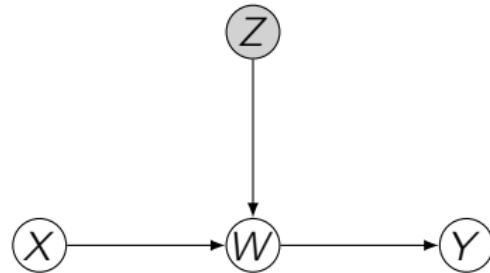
Exercise 6

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



Exercise 6

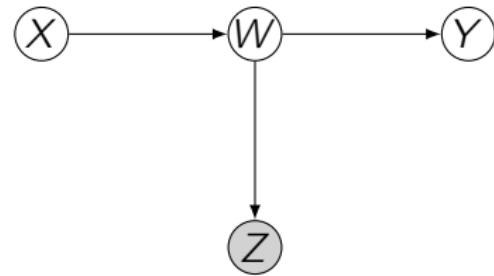
Is $\{Z\}$ a good, bad or neutral adjustment set for $P(y | do(x))$?



- Z does not open any back-door paths from X to Y
 $\implies \{Z\}$ is a neutral adjustment set;
- Adjusting on Z can reduce the variation of W , and helps improve the precision of the estimate in finite samples.

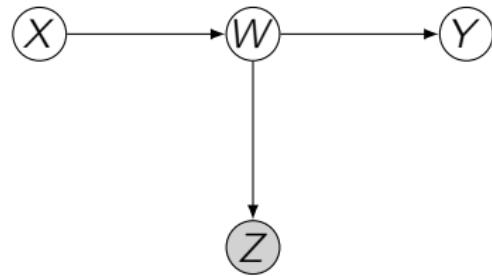
Exercise 6

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



Exercise 6

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



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

2

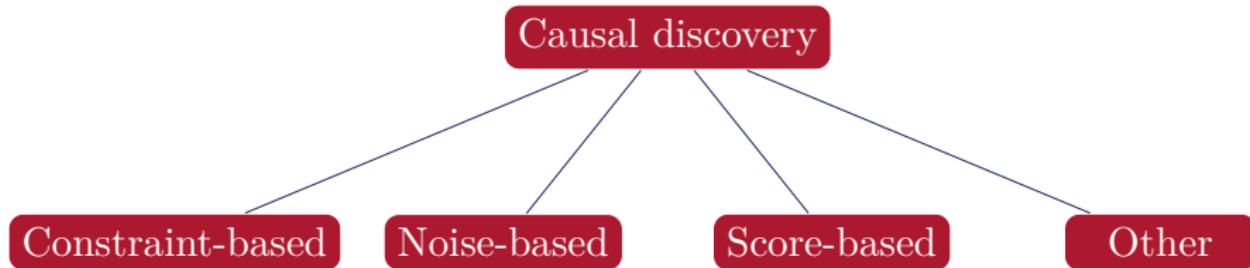
Intervention

Causal graphs

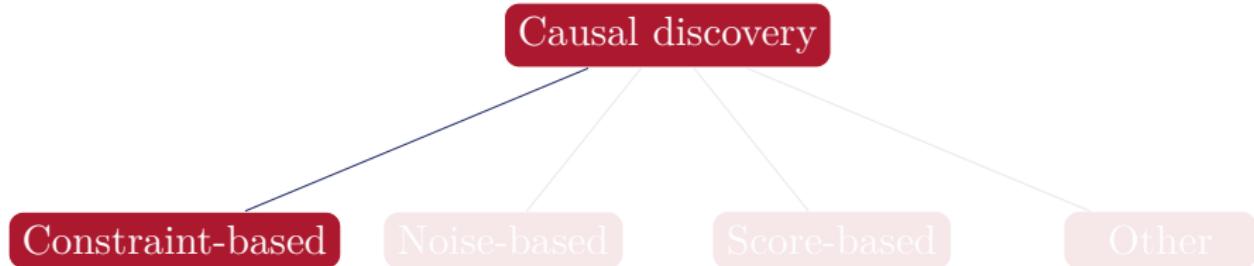
Causal reasoning

Causal discovery

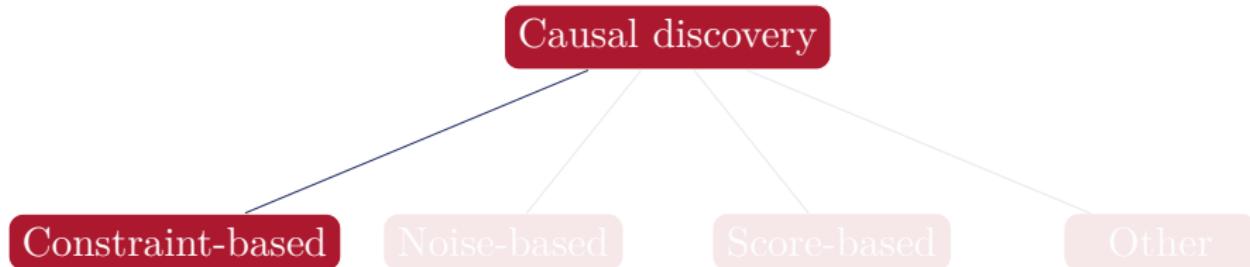
Families of causal discovery methods



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?

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!**

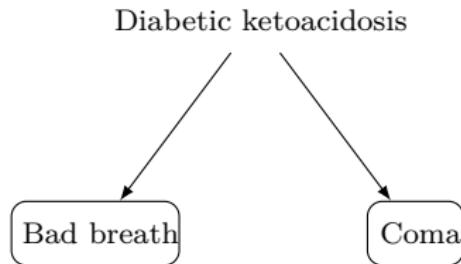
We cannot even construct the skeleton of the graph because

- $\not\perp\!\!\!\perp_P \not\Rightarrow \not\perp\!\!\!\perp_G$
- $\perp\!\!\!\perp_P \not\Rightarrow \perp\!\!\!\perp_G$

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

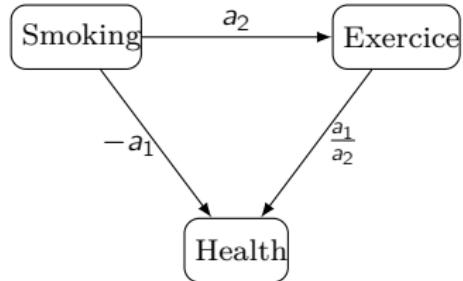
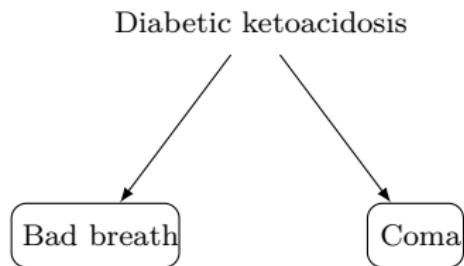
- $\not\perp\!\!\!\perp_P \not\Rightarrow \not\perp\!\!\!\perp_G$
- $\perp\!\!\!\perp_P \not\Rightarrow \perp\!\!\!\perp_G$



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

- $\not\perp\!\!\!\perp_P \not\Rightarrow \not\perp\!\!\!\perp_G$
- $\perp\!\!\!\perp_P \not\Rightarrow \perp\!\!\!\perp_G$



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} .

Causal sufficiency: No hidden confounding.

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

$$\implies A \perp\!\!\!\perp_{\mathcal{G}} B | S \iff A \perp\!\!\!\perp_P B | S$$

Causal sufficiency: No hidden confounding.

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

$$\implies A \perp\!\!\!\perp_{\mathcal{G}} B | S \iff A \perp\!\!\!\perp_P B | S$$

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

Causal sufficiency: No hidden confounding.

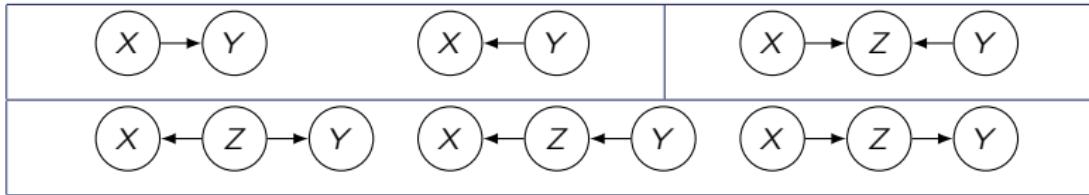
Faithfulness: All conditional independence relations true in P are entailed by the Markov condition applied to \mathcal{G} .

$$\implies A \perp\!\!\!\perp_{\mathcal{G}} B | S \iff A \perp\!\!\!\perp_P B | S$$

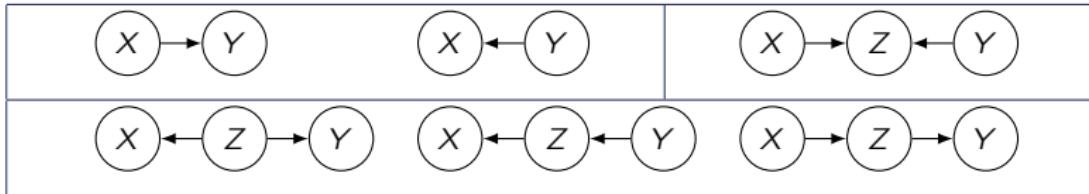
Given observational data, is it possible to infer a causal graph using conditional independencies under the assumptions of faithfulness and causal sufficiency? In general no!

Markov equivalence class [23]

Equivalence in terms of conditional independence



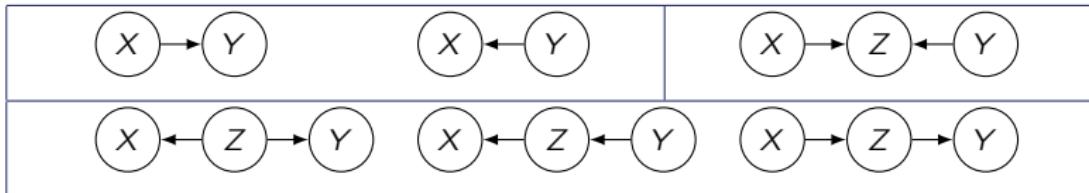
Equivalence in terms of conditional independence



Theorem

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

Equivalence in terms of conditional independence

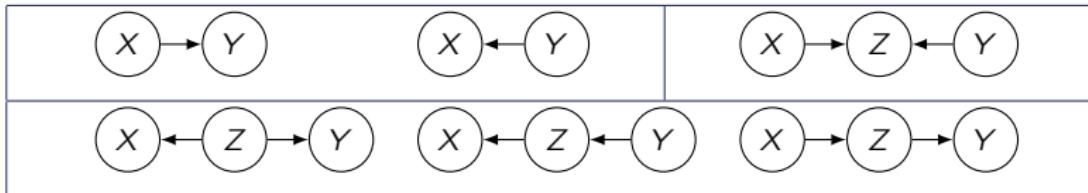


Theorem

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

Equivalence in terms of conditional independence



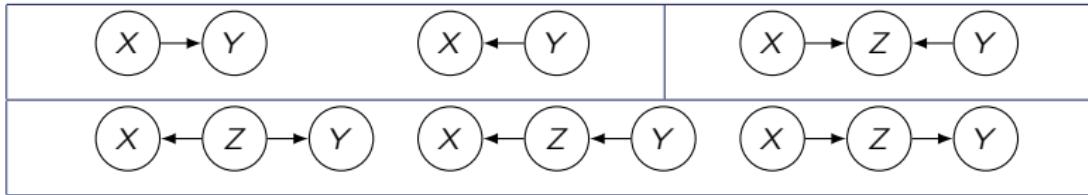
Theorem

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?

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 $\forall \mathbb{S} \subseteq \mathbb{V} \setminus \{X, Y\}$,
 $X \not\perp\!\!\!\perp_P Y \mid \mathbb{S}$;

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 $\forall \mathbb{S} \subseteq \mathbb{V} \setminus \{X, Y\}$,
 $X \not\perp\!\!\!\perp_P Y \mid \mathbb{S}$;
- 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 $\forall \mathbb{S} \subseteq \mathbb{V} \setminus \{X, Y\}$ such that $Z \in \mathbb{S}$, $X \not\perp\!\!\!\perp_P Y \mid \mathbb{S}$.

Theorem

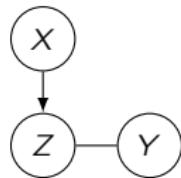
If $P(\mathbb{V})$ is faithful to some causal graph \mathcal{G} with vertex \mathbb{V} then:

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

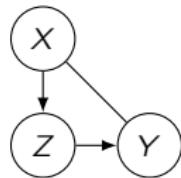
- Point 1 can be used to discover the skeleton of \mathcal{G} from $P(\mathcal{V})$;
- Given the skeleton of \mathcal{G} , point 2 can be used to find all v-structures.

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

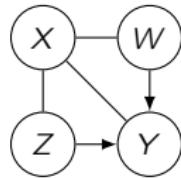
PC-Rule 1:



PC-Rule 2:



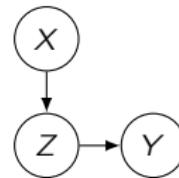
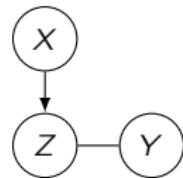
PC-Rule 3:



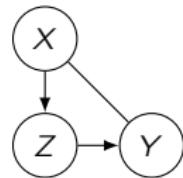
Orientation rules

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

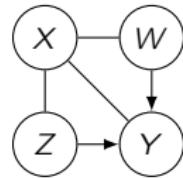
PC-Rule 1:



PC-Rule 2:



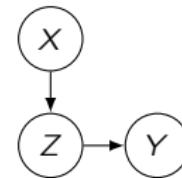
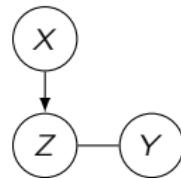
PC-Rule 3:



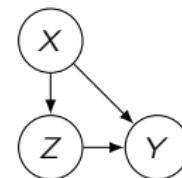
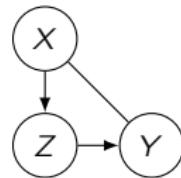
Orientation rules

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

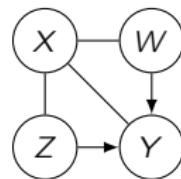
PC-Rule 1:



PC-Rule 2:



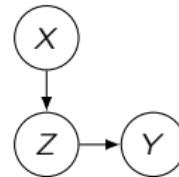
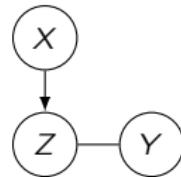
PC-Rule 3:



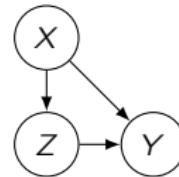
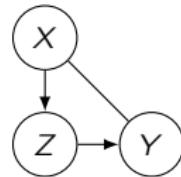
Orientation rules

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

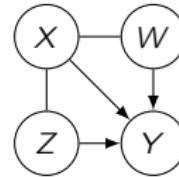
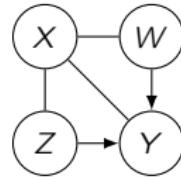
PC-Rule 1:



PC-Rule 2:



PC-Rule 3:



- Step 1: skeleton construction:

- ▶ Construct a complete non oriented graph
- ▶ Prune unnecessary edges (**optimal**) from the skeleton using
 $\perp\!\!\!\perp_P$

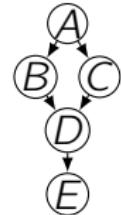
- Step 1: skeleton construction:
 - ▶ Construct a complete non oriented graph
 - ▶ Prune unnecessary edges (**optimal**) from the skeleton using $\perp\!\!\!\perp_P$
- Step 2: orientation
 - ▶ PC-Rule 0: find V-structures
 - ▶ PC-Rule 1, 2, 3: deduce other orientations by contradiction to V-structures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

- 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 ([19])

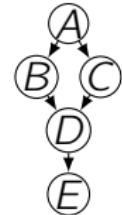
Assume the distribution P is compatible and faithful to some causal graph \mathcal{G} and assume that we are given perfect conditional independence information about all pairs of variables. The PC algorithm returns the CPDAG of \mathcal{G} .

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

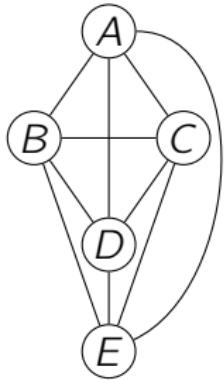


PC in action

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

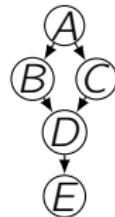


Skeleton construction:

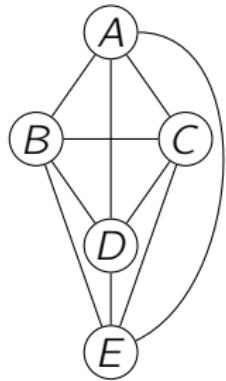


$$\text{card} = 0$$

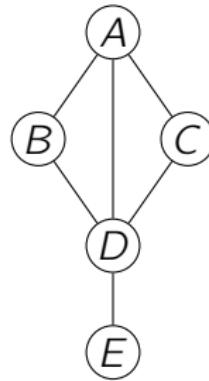
- Suppose the causal graph on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: Markov condition, causal sufficiency, faithfulness



Skeleton construction:

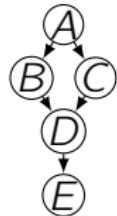


card = 0

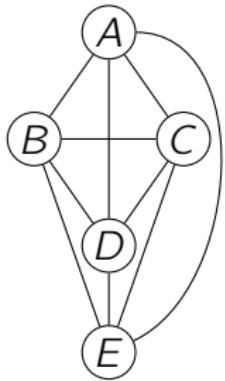


card = 1

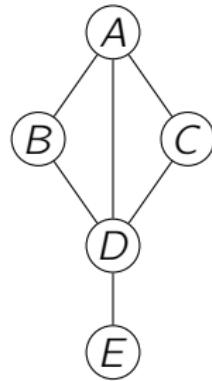
- Suppose the causal graph on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: Markov condition, causal sufficiency, faithfulness



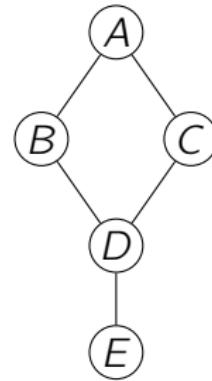
Skeleton construction:



card = 0

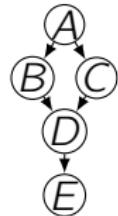


card = 1

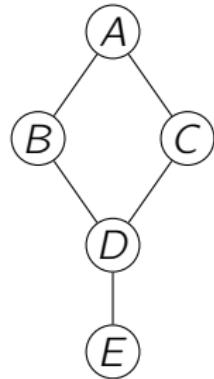


card = 2

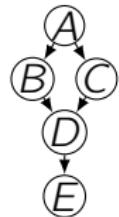
- Suppose the causal graph on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: Markov condition, causal sufficiency, faithfulness



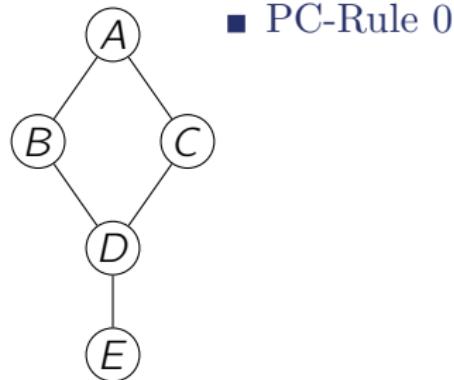
Orientation:



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



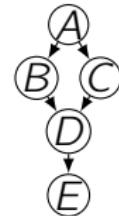
Orientation:



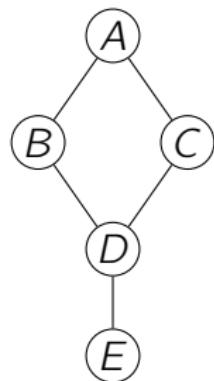
- PC-Rule 0

PC in action

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



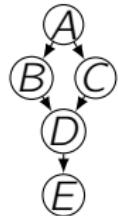
Orientation:



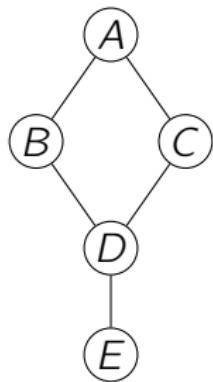
- PC-Rule 0

$$\blacktriangleright \begin{array}{c} B \perp\!\!\!\perp_P C \mid A \\ \implies B \rightarrow D \leftarrow C \end{array}$$

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

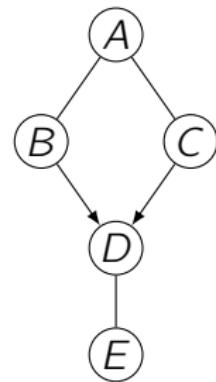


Orientation:

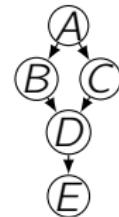


- PC-Rule 0

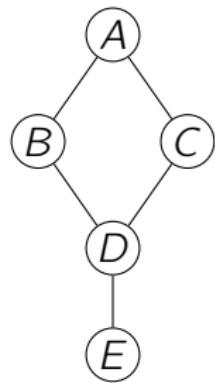
$$\blacktriangleright \begin{array}{c} B \perp\!\!\!\perp_P C \mid A \\ \implies B \rightarrow D \leftarrow C \end{array}$$



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

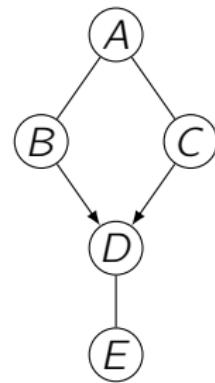


Orientation:

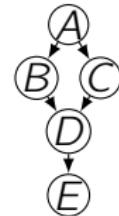


■ PC-Rule 0
 $\blacktriangleright \frac{B \perp\!\!\!\perp_P C | A}{C} \implies B \rightarrow D \leftarrow C$

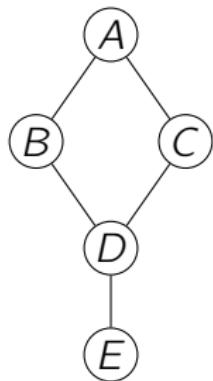
■ PC-Rule 1



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

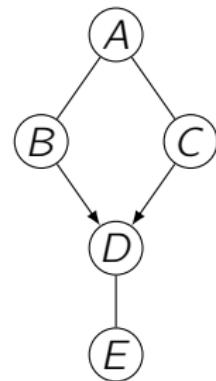


Orientation:

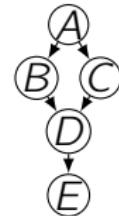


■ PC-Rule 0
 $\blacktriangleright \begin{array}{c} B \perp\!\!\!\perp_P C \mid A \\ \implies B \rightarrow D \leftarrow C \end{array}$

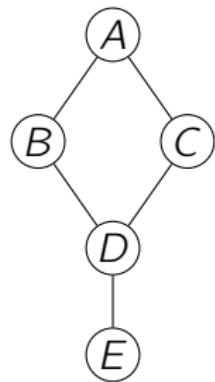
■ PC-Rule 1
 $\blacktriangleright \begin{array}{c} B \rightarrow D \ \& \ D \rightarrow E \\ \implies D \rightarrow E \end{array}$



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

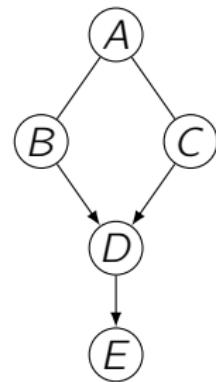


Orientation:



■ PC-Rule 0
 $\blacktriangleright \frac{B \perp\!\!\!\perp_P C \mid A}{B \rightarrow D \leftarrow C}$

■ PC-Rule 1
 $\blacktriangleright \frac{B \rightarrow D \text{ & } D \rightarrow E}{D \rightarrow E}$

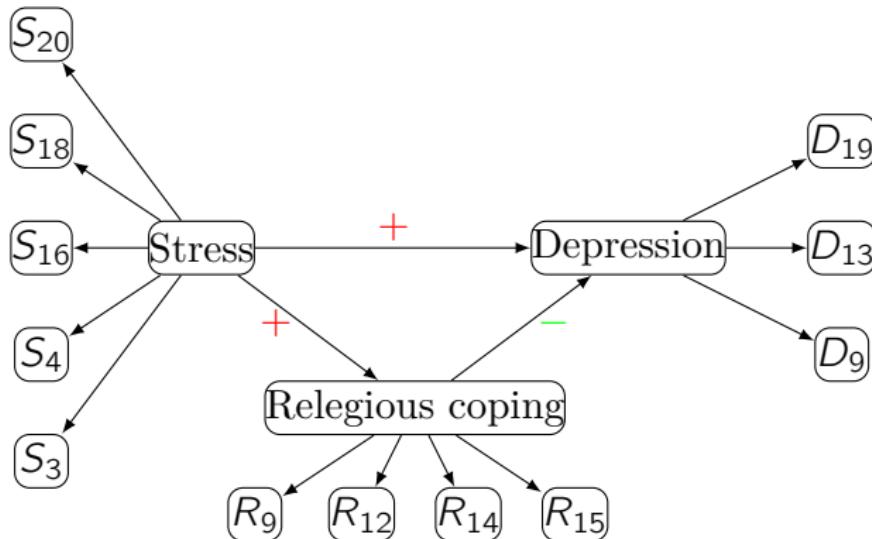


The PC algorithm can effectively incorporate background knowledge in the form of:

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

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

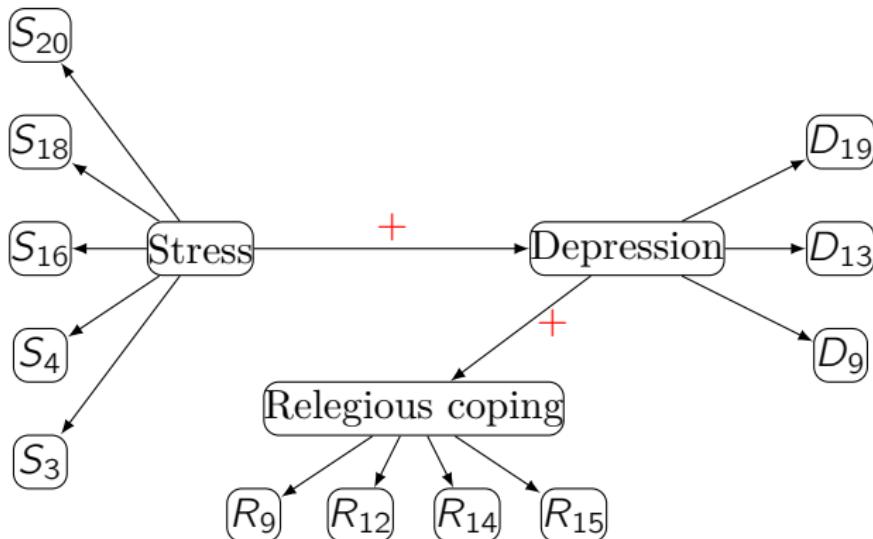
Specified graph



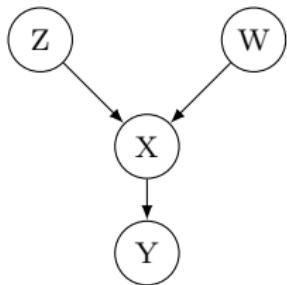
Real application of PC [16]

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

Inferred graph (assuming stress is temporally prior)



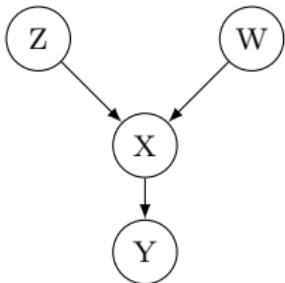
Beyond causal sufficiency (1/2)



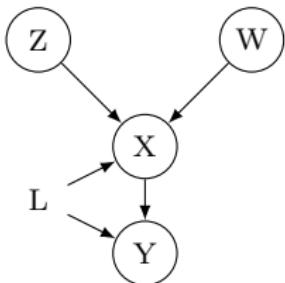
Y-structure

$$\begin{aligned}
 & z \perp\!\!\!\perp_P w \\
 & z \not\perp\!\!\!\perp_P w \mid x \\
 & y \not\perp\!\!\!\perp_P z \\
 & y \perp\!\!\!\perp_P z \mid x \\
 & y \not\perp\!\!\!\perp_P w \\
 & y \perp\!\!\!\perp_P w \mid x
 \end{aligned}$$

Beyond causal sufficiency (1/2)



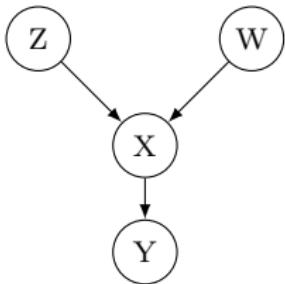
Y-structure



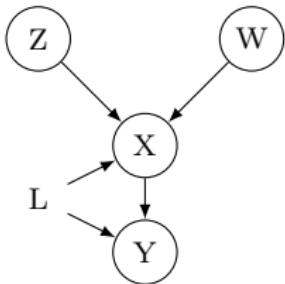
$$\begin{aligned}
 & z \perp\!\!\!\perp_P w \\
 & z \not\perp\!\!\!\perp_P w \mid x \\
 & y \perp\!\!\!\perp_P z \\
 & y \perp\!\!\!\perp_P z \mid x \\
 & y \not\perp\!\!\!\perp_P w \\
 & y \perp\!\!\!\perp_P w \mid x
 \end{aligned}$$

$$\begin{aligned}
 & z \perp\!\!\!\perp_P w \\
 & z \not\perp\!\!\!\perp_P w \mid x \\
 & y \perp\!\!\!\perp_P z \\
 & y \not\perp\!\!\!\perp_P z \mid x \\
 & y \perp\!\!\!\perp_P w \\
 & y \not\perp\!\!\!\perp_P w \mid x
 \end{aligned}$$

Beyond causal sufficiency (1/2)



Y-structure

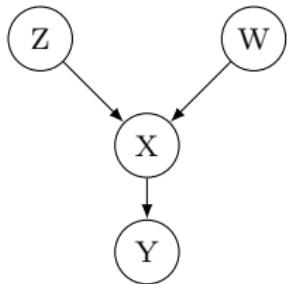


Intervention

$$\begin{aligned}
 & z \perp\!\!\!\perp_P w \\
 & z \not\perp\!\!\!\perp_P w \mid x \\
 & y \perp\!\!\!\perp_P z \\
 & y \perp\!\!\!\perp_P z \mid x \\
 & y \not\perp\!\!\!\perp_P w \\
 & y \perp\!\!\!\perp_P w \mid x
 \end{aligned}$$

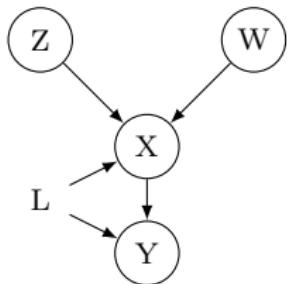
$$\begin{aligned}
 & z \perp\!\!\!\perp_P w \\
 & z \not\perp\!\!\!\perp_P w \mid x \\
 & y \perp\!\!\!\perp_P z \\
 & \cancel{y \perp\!\!\!\perp_P z \mid x} \\
 & \cancel{y \perp\!\!\!\perp_P w} \\
 & \cancel{y \perp\!\!\!\perp_P w \mid x}
 \end{aligned}$$

Beyond causal sufficiency (1/2)



Y-structure

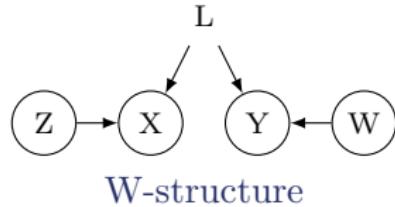
$$\begin{aligned}
 & z \perp\!\!\!\perp_P w \\
 & z \not\perp\!\!\!\perp_P w | x \\
 & y \perp\!\!\!\perp_P z \\
 & y \perp\!\!\!\perp_P z | x \\
 & y \not\perp\!\!\!\perp_P w \\
 & y \perp\!\!\!\perp_P w | x
 \end{aligned}$$



$$\begin{aligned}
 & z \perp\!\!\!\perp_P w \\
 & z \not\perp\!\!\!\perp_P w | x \\
 & y \perp\!\!\!\perp_P z \\
 & \cancel{y \perp\!\!\!\perp_P z | x} \\
 & \cancel{y \perp\!\!\!\perp_P w} \\
 & \cancel{y \perp\!\!\!\perp_P w | x}
 \end{aligned}$$

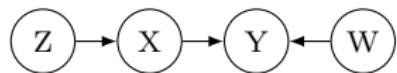
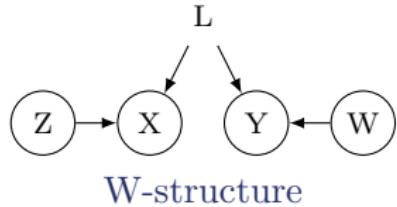
Pattern of independence can rule out hidden confounding.

Beyond causal sufficiency (2/2)



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

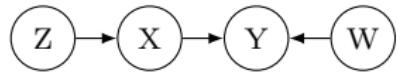
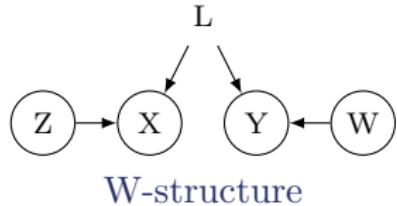
Beyond causal sufficiency (2/2)



$$\begin{array}{l} z \not\perp\!\!\!\perp_P x \\ x \not\perp\!\!\!\perp_P y \\ y \not\perp\!\!\!\perp_P w \\ z \perp\!\!\!\perp_P w \\ z \perp\!\!\!\perp_P y \\ x \perp\!\!\!\perp_P w \\ z \not\perp\!\!\!\perp_P y \mid x \\ x \not\perp\!\!\!\perp_P w \mid y \end{array}$$

$$\begin{array}{l} z \not\perp\!\!\!\perp_P x \\ x \not\perp\!\!\!\perp_P y \\ y \not\perp\!\!\!\perp_P w \\ z \perp\!\!\!\perp_P w \\ z \not\perp\!\!\!\perp_P y \\ x \perp\!\!\!\perp_P w \\ z \perp\!\!\!\perp_P y \mid x \\ x \not\perp\!\!\!\perp_P w \mid y \end{array}$$

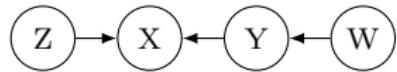
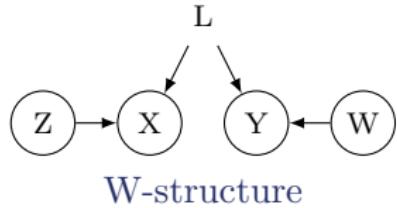
Beyond causal sufficiency (2/2)



$$\begin{array}{l} z \not\perp\!\!\!\perp_P x \\ x \not\perp\!\!\!\perp_P y \\ y \not\perp\!\!\!\perp_P w \\ z \perp\!\!\!\perp_P w \\ z \perp\!\!\!\perp_P y \\ x \perp\!\!\!\perp_P w \\ z \not\perp\!\!\!\perp_P y | x \\ x \not\perp\!\!\!\perp_P w | y \end{array}$$

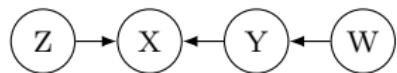
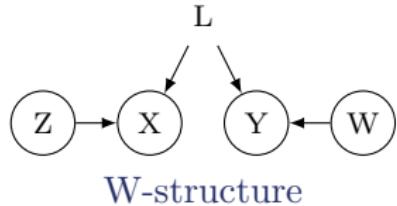
$$\begin{array}{l} z \not\perp\!\!\!\perp_P x \\ x \not\perp\!\!\!\perp_P y \\ y \not\perp\!\!\!\perp_P w \\ z \perp\!\!\!\perp_P w \\ z \not\perp\!\!\!\perp_P y \\ x \perp\!\!\!\perp_P w \\ z \perp\!\!\!\perp_P y | x \\ x \not\perp\!\!\!\perp_P w | y \end{array}$$

Beyond causal sufficiency (2/2)



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

Beyond causal sufficiency (2/2)



$Z \not\perp\!\!\!\perp_P X$
 $X \not\perp\!\!\!\perp_P Y$
 $Y \not\perp\!\!\!\perp_P W$
 $Z \perp\!\!\!\perp_P W$
 $Z \perp\!\!\!\perp_P Y$
 $\textcolor{red}{X \perp\!\!\!\perp_P W}$
 $Z \not\perp\!\!\!\perp_P Y | X$
 $\textcolor{red}{X \not\perp\!\!\!\perp_P W | Y}$

$Z \not\perp\!\!\!\perp_P X$
 $X \not\perp\!\!\!\perp_P Y$
 $Y \not\perp\!\!\!\perp_P W$
 $Z \perp\!\!\!\perp_P W$
 $Z \perp\!\!\!\perp_P Y$
 $\textcolor{red}{X \not\perp\!\!\!\perp_P W}$
 $Z \not\perp\!\!\!\perp_P Y | X$
 $\textcolor{red}{X \not\perp\!\!\!\perp_P W | Y}$

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

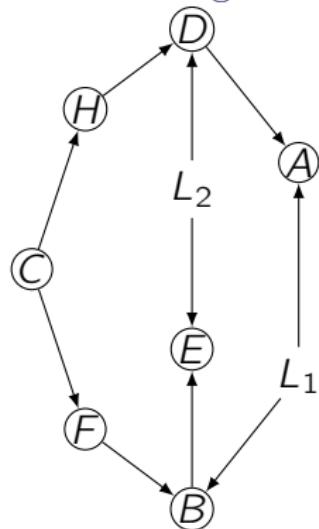
The FCI algorithm extends the PC algorithm to accommodate hidden confounders:

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

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:



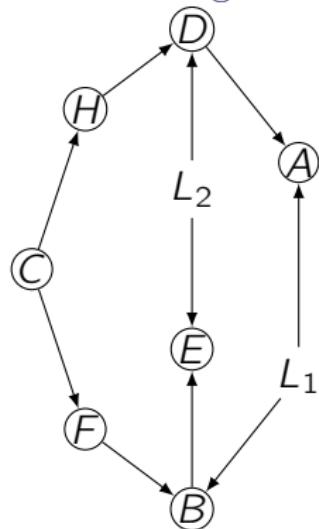
True graph

A glimpse of the FCI algorithm

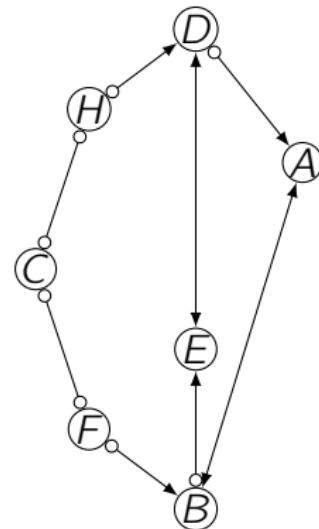
The FCI algorithm extends the PC algorithm to accommodate hidden confounders:

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

The FCI algorithm in action:



True graph



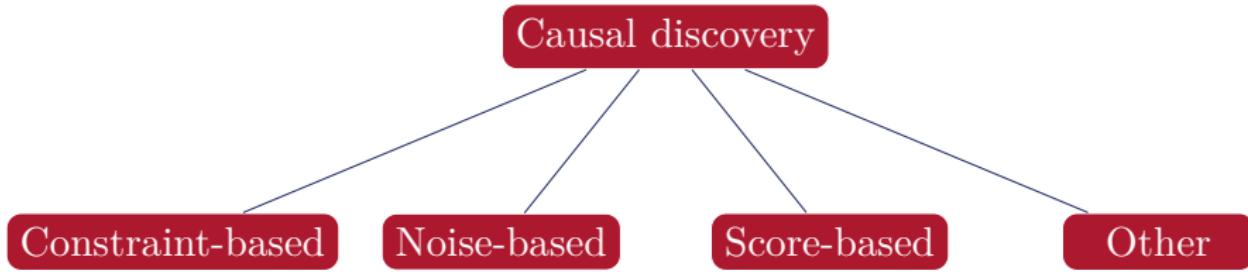
Inferred graph

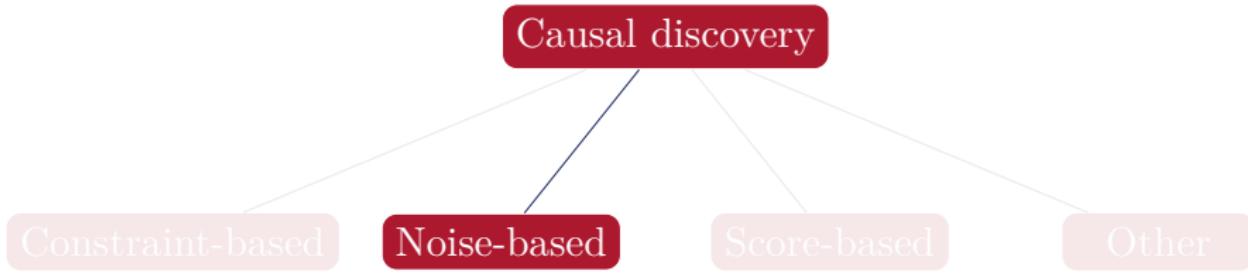
Pros:

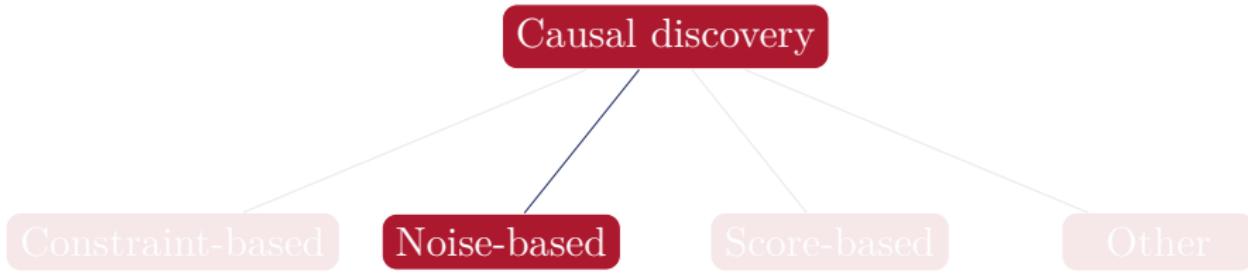
- Non-parametric (in practice, it depends on the selected independence test)
- Intuitive

Cons:

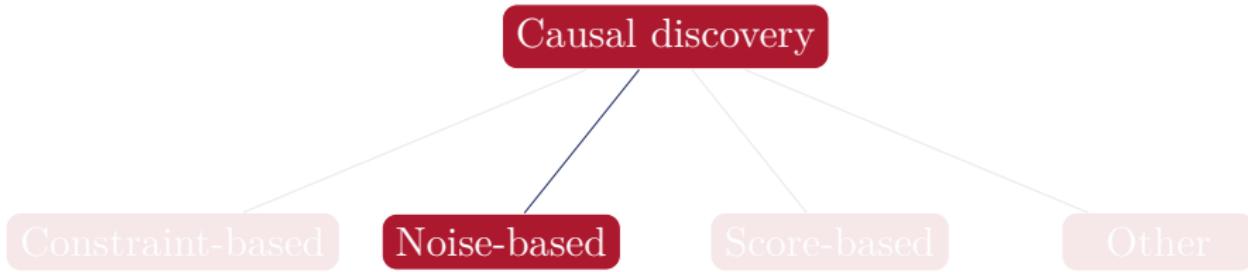
- Recover a partially oriented graph
- The faithfulness assumption is not always accepted







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



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

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?

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

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

Without further assumption we cannot know.

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 Without further assumption we cannot know.

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

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

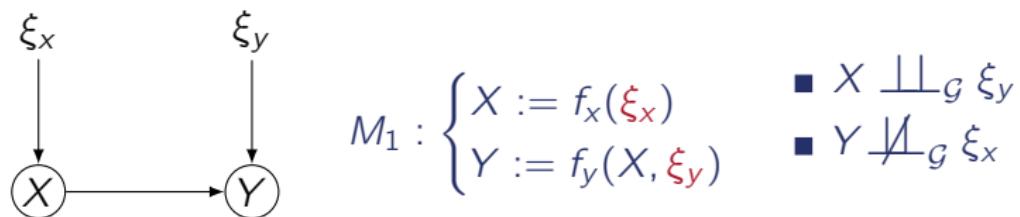
Without further assumption we cannot know.

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

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

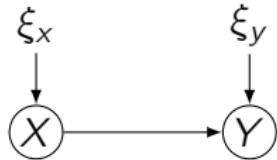
X	Y	$\xi_y = Y - 2X$	$\hat{\xi}_x = X - Y/2$
1	2	$0 \in \{-1, 0, 1\}$	$0 \in \{-1, 0, 1\}$
3	6	$0 \in \{-1, 0, 1\}$	$0 \in \{-1, 0, 1\}$
4	9	$1 \in \{-1, 0, 1\}$	$-0.5 \notin \{-1, 0, 1\}$

The intuition behind the noise (2/2)



The experiment that changed everything

Suppose



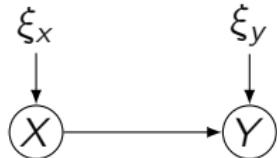
$$X \sim N(0, 1)$$

$$\xi_y \sim N(0, 1)$$

$$Y := 2X + \xi_y$$

The experiment that changed everything

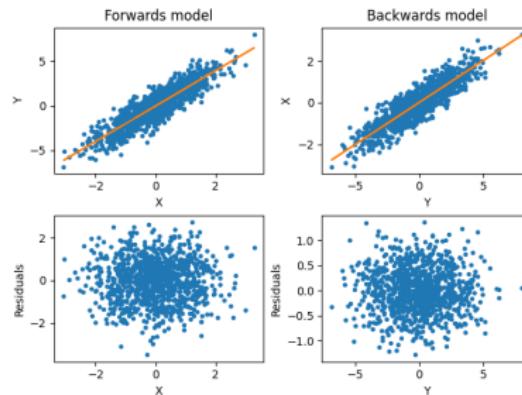
Suppose



$$X \sim N(0, 1)$$

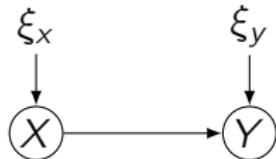
$$\xi_y \sim N(0, 1)$$

$$Y := 2X + \xi_y$$



The experiment that changed everything

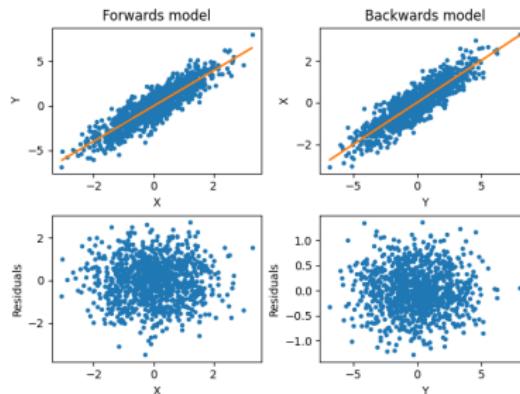
Suppose



$$X \sim N(0, 1)$$

$$\xi_y \sim N(0, 1)$$

$$Y := 2X + \xi_y$$



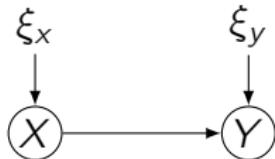
$$X \sim U(0, 1)$$

$$\xi_y \sim U(0, 1)$$

$$Y := 2X + \xi_y$$

The experiment that changed everything

Suppose



$$X \sim N(0, 1)$$

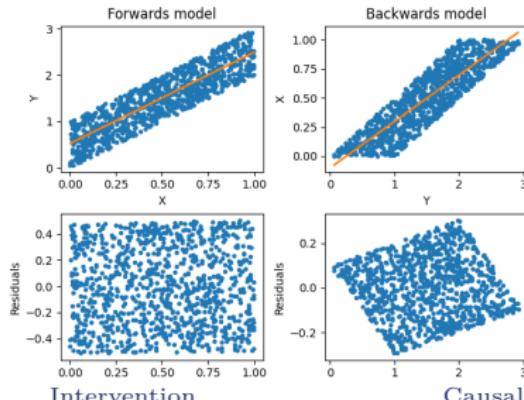
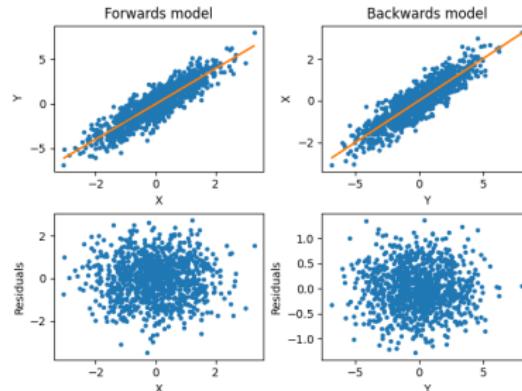
$$\xi_y \sim N(0, 1)$$

$$Y := 2X + \xi_y$$

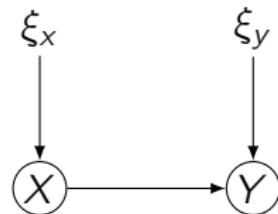
$$X \sim U(0, 1)$$

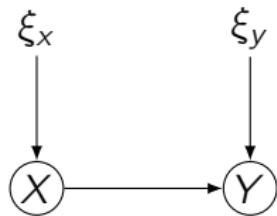
$$\xi_y \sim U(0, 1)$$

$$Y := 2X + \xi_y$$



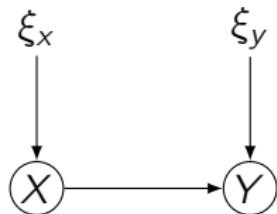
The linear case (1/2)





True model:

$$M_1 : \begin{cases} X := \xi_x \\ Y := aX + \xi_y \end{cases} \quad \blacksquare \quad X \perp\!\!\!\perp_{\mathcal{G}} \xi_y$$

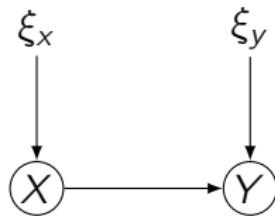


True model:

$$M_1 : \begin{cases} X := \xi_x \\ Y := aX + \xi_y \end{cases} \quad \blacksquare \quad X \perp\!\!\!\perp_{\mathcal{G}} \xi_y$$

Backwards model:

$$M_2 : \begin{cases} Y := \hat{\xi}_y \\ X := bY + \hat{\xi}_x \end{cases} \quad \blacksquare \quad Y \perp\!\!\!\perp_{\mathcal{G}} \hat{\xi}_x ?$$



True model:

$$M_1 : \begin{cases} X := \xi_x \\ Y := aX + \xi_y \end{cases} \quad \blacksquare \quad X \perp\!\!\!\perp_{\mathcal{G}} \xi_y$$

Backwards model:

$$M_2 : \begin{cases} Y := \hat{\xi}_y \\ X := bY + \hat{\xi}_x \end{cases} \quad \blacksquare \quad Y \perp\!\!\!\perp_{\mathcal{G}} \hat{\xi}_x?$$

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

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

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

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

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

Theorem (Darmois-Skitovich)

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

$$l_1 = a_1 X_1 + \dots + a_n X_n$$
$$l_2 = b_1 X_1 + \dots + b_n X_n$$

are independent, then each X_i is normally distributed.

Theorem

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

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

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

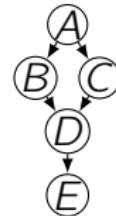
$$X := bY + \hat{\xi}_x, \quad Y \perp\!\!\!\perp_P \hat{\xi}_x,$$

iff ξ_y and X are Gaussian.

Similar result for the multivariate case

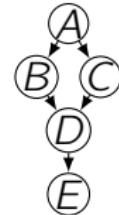
A glimpse of LiNGAM in action

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

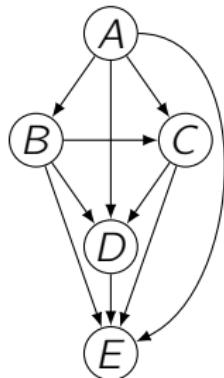


A glimpse of LiNGAM in action

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

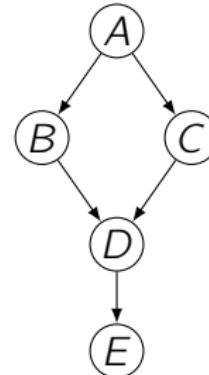


Causal order:



Causal Inference

Pruning:



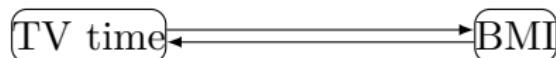
Intervention

Causal discovery

Real application of LiNGAM [4]

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

Specified graphs



Real application of LiNGAM [4]

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

Inferred graphs



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

Pros:

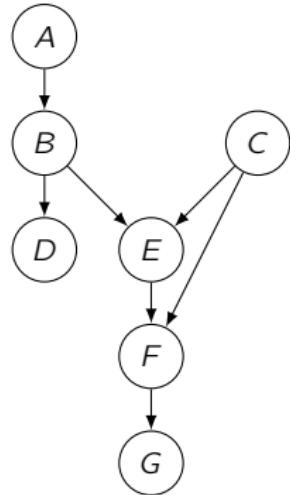
- Recover the complete causal graph

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

Structural causal models

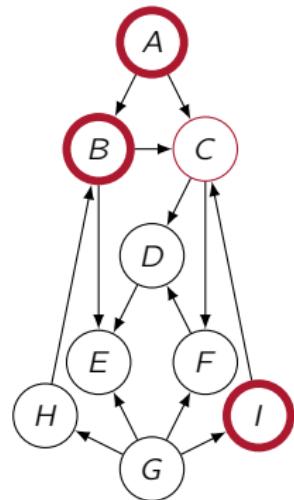
Counterfactual reasoning

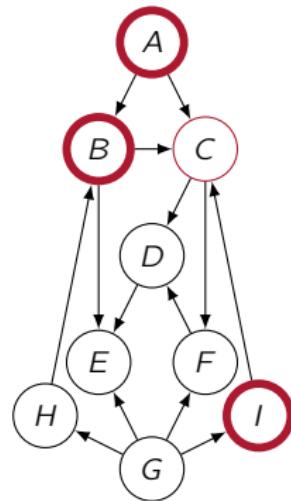
Mediation analysis

Difference with the potential outcome framework

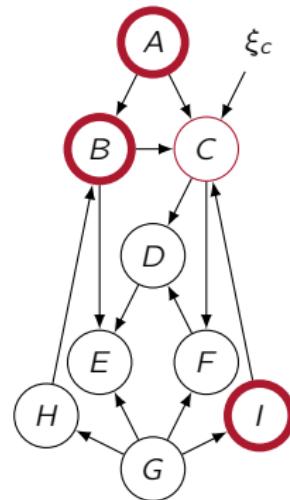
https:

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

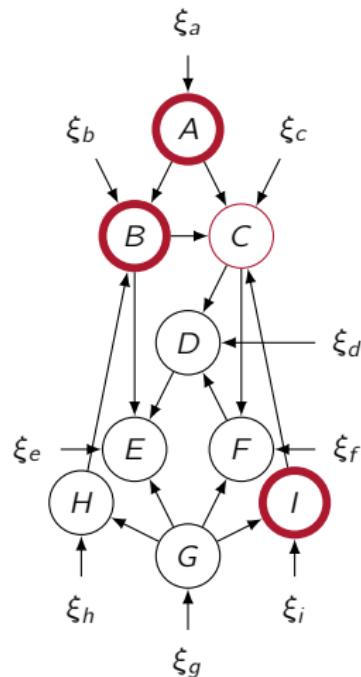




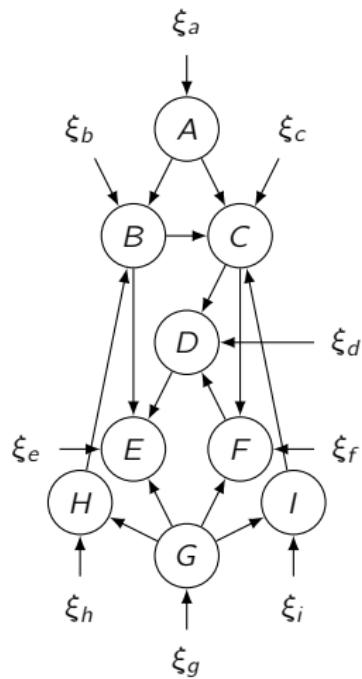
$$C = f(A, B, I)$$



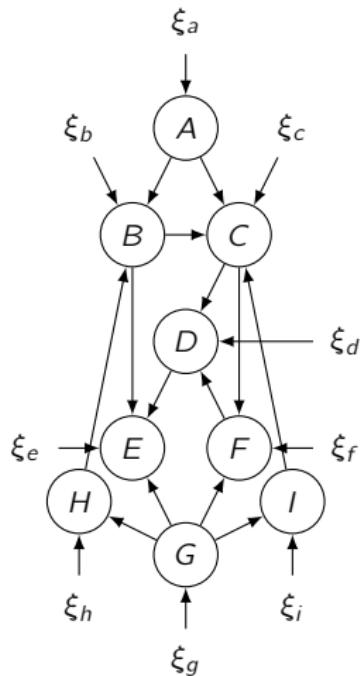
$$C := f_c(A, B, I, \xi_c)$$



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

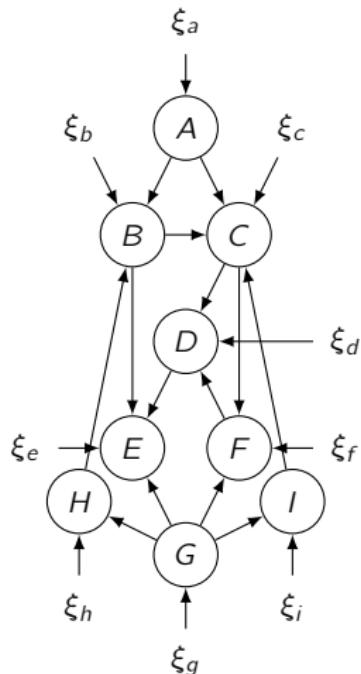


Structural causal model



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

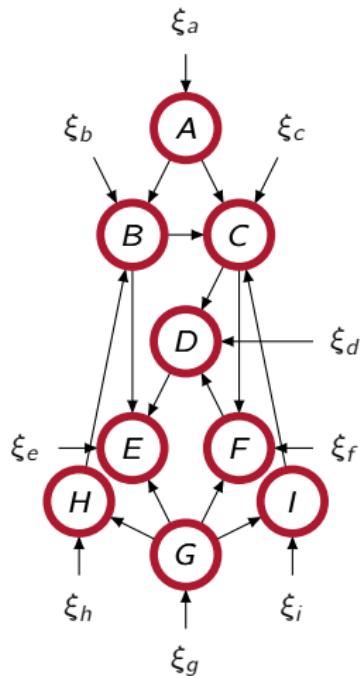
Structural causal model



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

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

Structural causal model

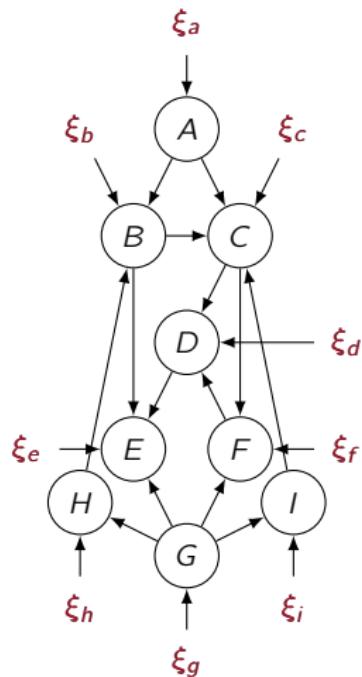


$$M : \left\{ \begin{array}{l} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{array} \right.$$

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

- Endogenous variables

Structural causal model

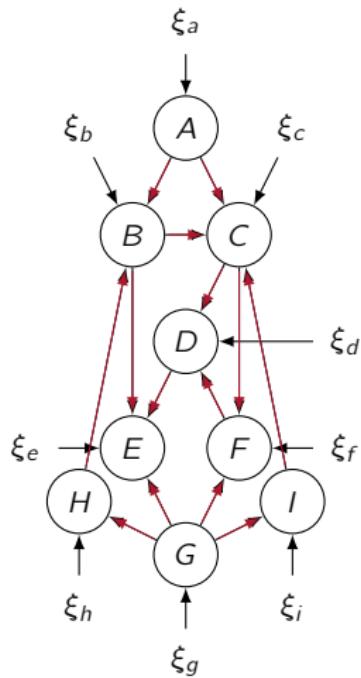


$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

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

- Endogenous variables
- Exogenous variables

Structural causal model



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

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

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

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

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

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

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

$$M_c : \left\{ \begin{array}{l} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := c \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{array} \right.$$

Interventional SCMs naturally helps us in computing counterfactual queries:

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

3

Counterfactuals

Structural causal models

Counterfactual reasoning

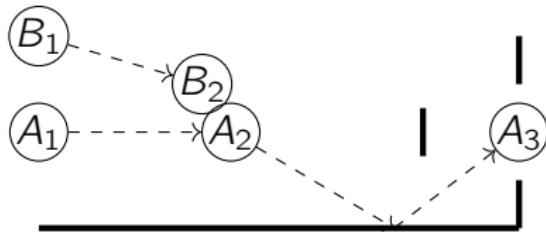
Mediation analysis

Difference with the potential outcome framework

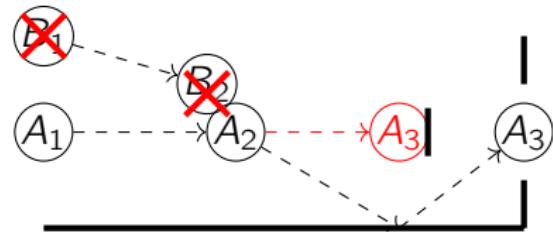
1. Abduction: use the observations to determine the value of ξ_c
2. Action: modify the model M by removing the structural equations for the variable C and replacing it with the appropriate functions $C = c$, to obtain the interventional SCM M_c
3. Prediction: use M_c and the value of ξ_c to compute the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

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

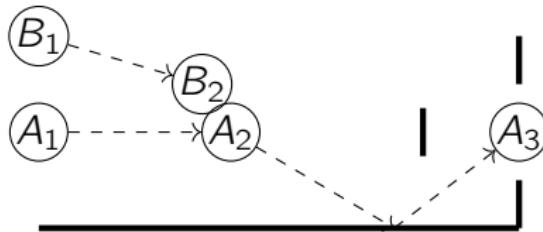
Factual world



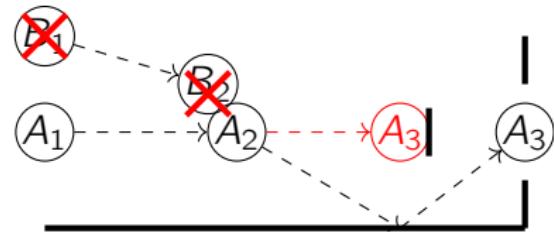
Conterfactual world



Factual world



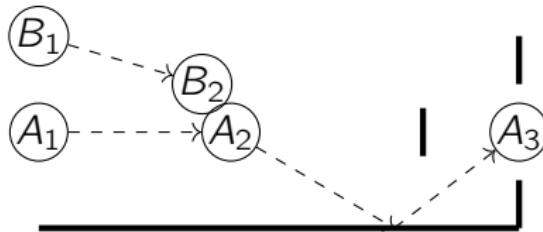
Conterfactual world



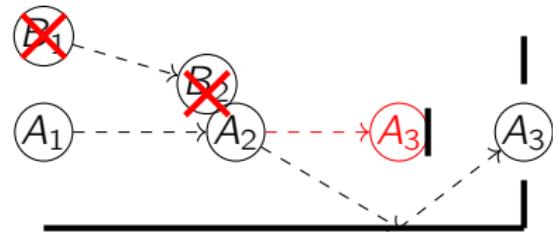
Toy example (without noise on C , W , and G):

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

Factual world



Conterfactual world

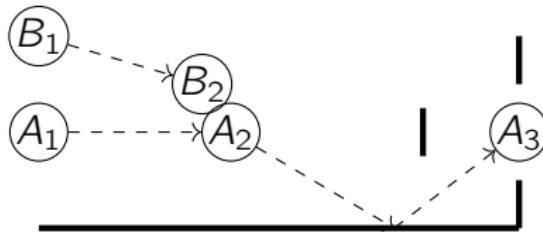


Toy example (without noise on C , W , and G):

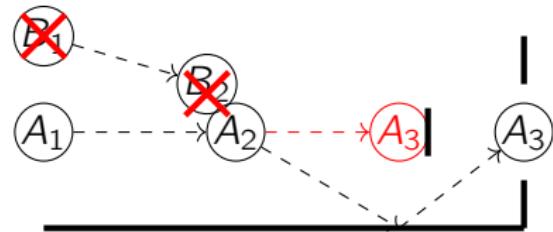
$$M : \begin{cases} H & (B \text{ hits } A) := \xi_h \\ C & (A \text{ changes trajectory}) := H \\ W & (A \text{ hits the wall}) := 1 - C \\ G & (A \text{ reaches the goal}) := 1 - W \end{cases}$$

$$\quad \quad \quad \begin{cases} H := 1 \\ C := 1 \\ W := 0 \\ G := 1 \end{cases}$$

Factual world



Conterfactual world

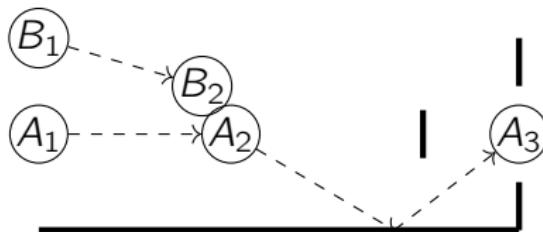


Toy example (without noise on C , W , and G):

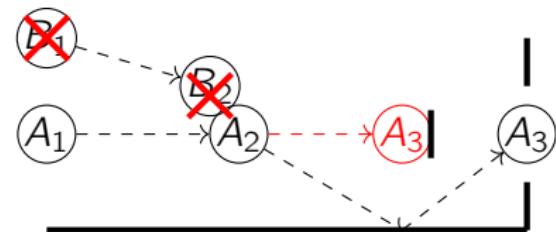
$$M : \begin{cases} H & (B \text{ hits } A) := \xi_h \\ C & (A \text{ changes trajectory}) := H \\ W & (A \text{ hits the wall}) := 1 - C \\ G & (A \text{ reaches the goal}) := 1 - W \end{cases}$$

$$M_{H=0} : \begin{cases} H := 0 \\ C := 1 \\ W := 0 \\ G := 1 \end{cases}$$

Factual world



Conterfactual world



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

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

$$\quad \quad \quad \begin{cases} H := 1 \\ C := 1 \\ W := 0 \\ G := 1 \end{cases}$$

$$M_{H=0} : \begin{cases} H := 0 \\ C := 0 \\ W := 1 \\ G := 0 \end{cases}$$

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

3

Counterfactuals

Structural causal models

Counterfactual reasoning

Mediation analysis

Difference with the potential outcome framework

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 omitted factors and α_{xy} which are known as a structural coefficient represents the strength of the causal relation.

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 omitted 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 .

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 omitted factors and α_{xy} which are known as a structural coefficient represents the strength of the causal relation.

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

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

Consider a causal 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 Z satisfies the **single-door criterion** iff:

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

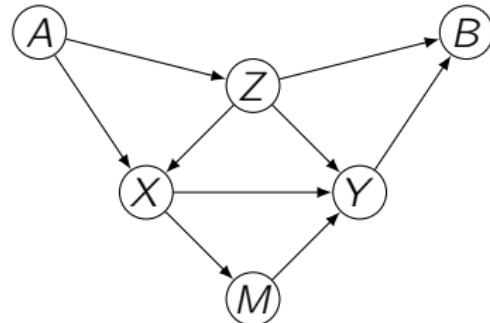
Theorem

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

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

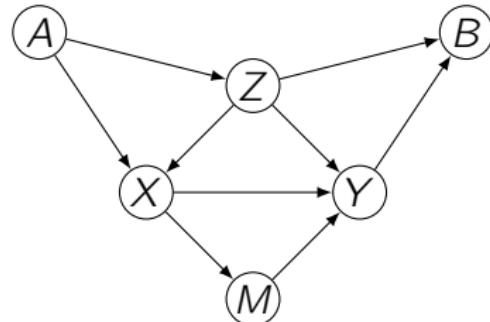
where $r_{XY.Z}$ represents the residual correlation between Y and X after \mathbb{Z} is "partialled out".

Single-door in action



Does the following set satisfy the single-door criterion?

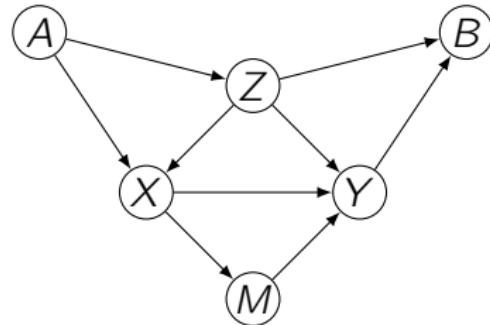
Single-door in action



Does the following set satisfy the single-door criterion?

$$\{Z\} ?$$

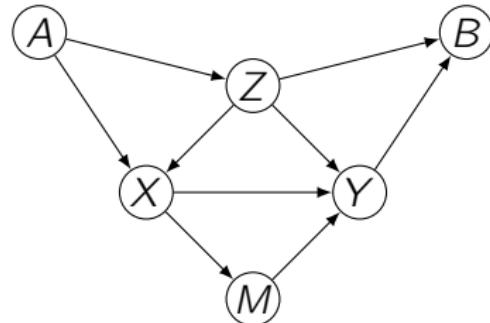
Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No

Single-door in action

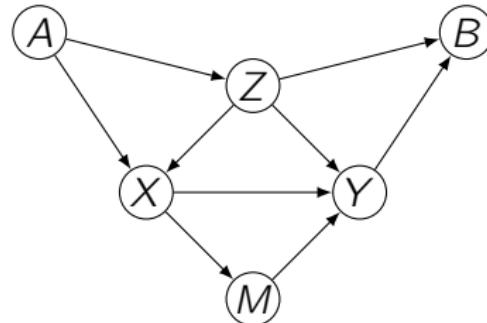


Does the following set satisfy the single-door criterion?

$\{Z\}$? No

$\{A\}$?

Single-door in action

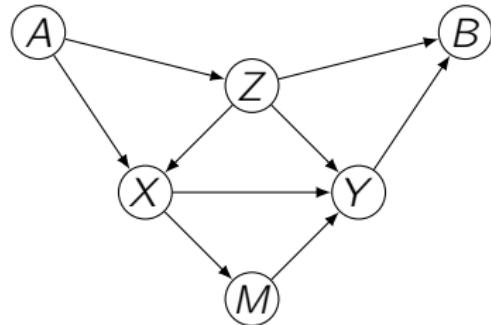


Does the following set satisfy the single-door criterion?

$\{Z\}$? No

$\{A\}$? No

Single-door in action



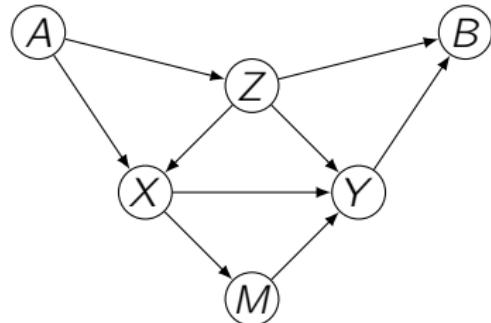
Does the following set satisfy the single-door criterion?

$\{Z\}$? No

$\{A\}$? No

$\{B\}$?

Single-door in action



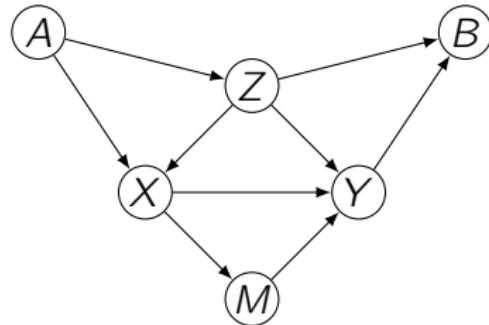
Does the following set satisfy the single-door criterion?

$\{Z\}$? No

$\{A\}$? No

$\{B\}$? No

Single-door in action



Does the following set satisfy the single-door criterion?

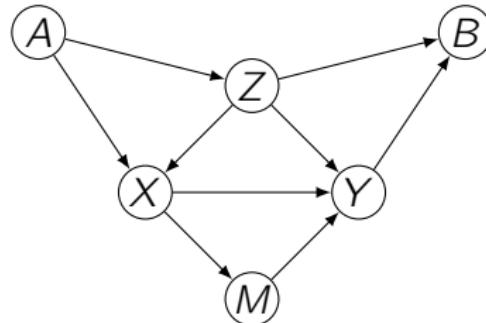
$\{Z\}$? No

$\{A\}$? No

$\{B\}$? No

$\{M\}$?

Single-door in action



Does the following set satisfy the single-door criterion?

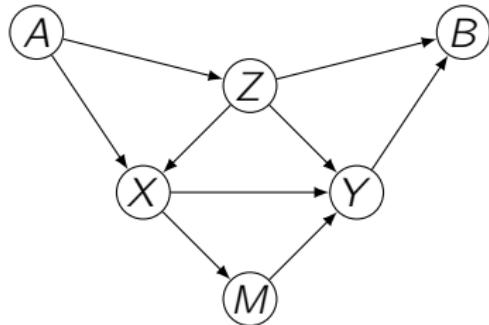
$\{Z\}$? No

$\{A\}$? No

$\{B\}$? No

$\{M\}$? No

Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No

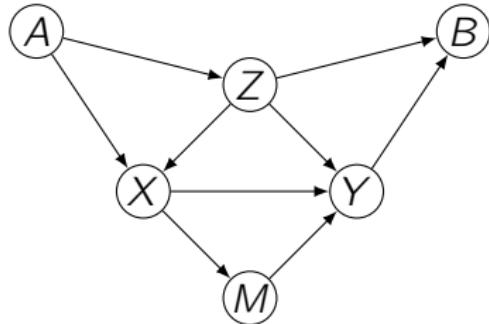
$\{A\}$? No

$\{B\}$? No

$\{M\}$? No

$\{A, B\}$?

Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No

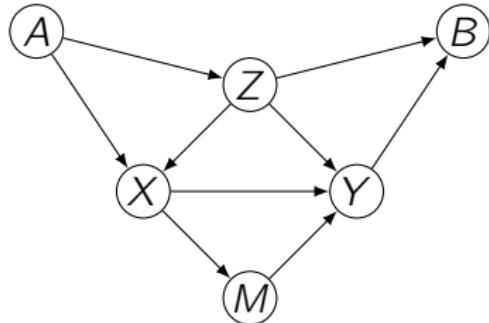
$\{A\}$? No

$\{B\}$? No

$\{M\}$? No

$\{A, B\}$? No

Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No

$\{A\}$? No

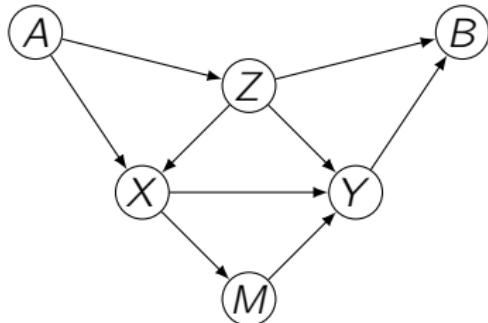
$\{B\}$? No

$\{M\}$? No

$\{A, B\}$? No

$\{Z, M\}$?

Single-door in action



Does the following set satisfy the single-door criterion?

$\{Z\}$? No

$\{A\}$? No

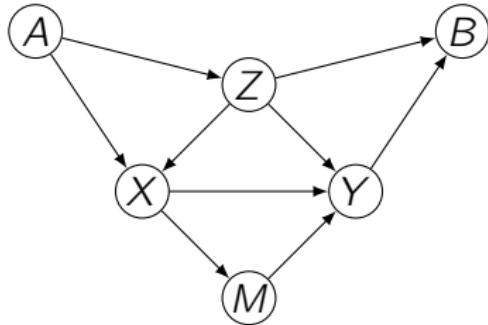
$\{B\}$? No

$\{M\}$? 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

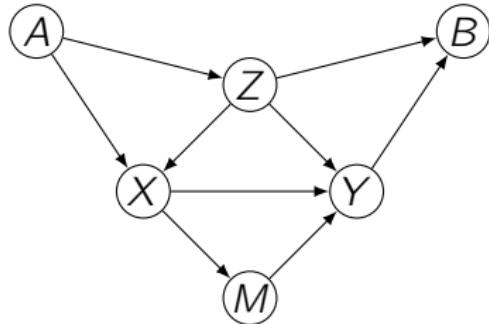
$\{M\}$? No

$\{A, B\}$? 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

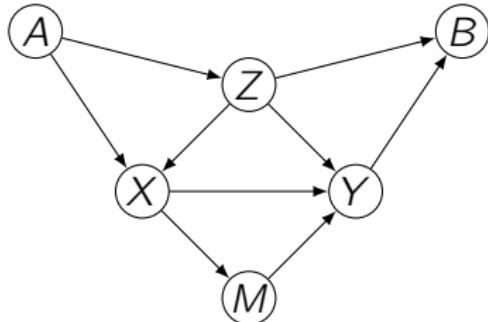
$\{M\}$? No

$\{A, B\}$? 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

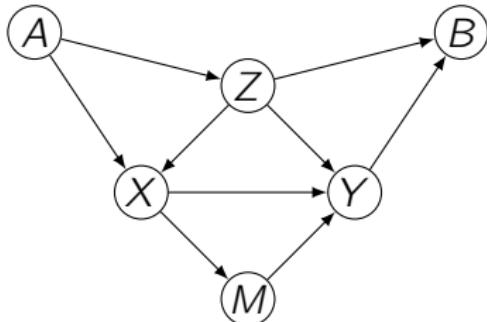
$\{A, B\}$? No

$\{Z, M\}$? 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

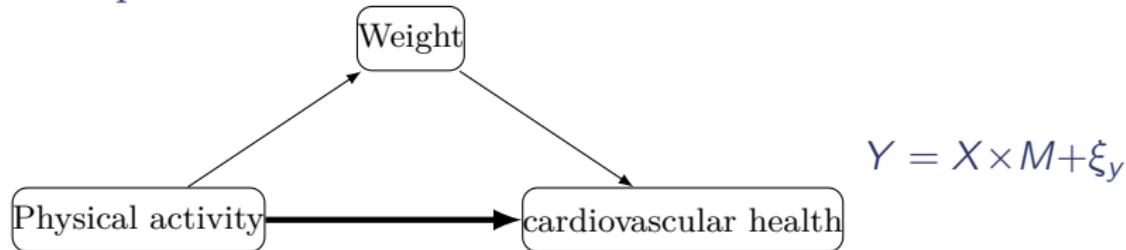
Indirect effect = total effect – direct effect

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

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

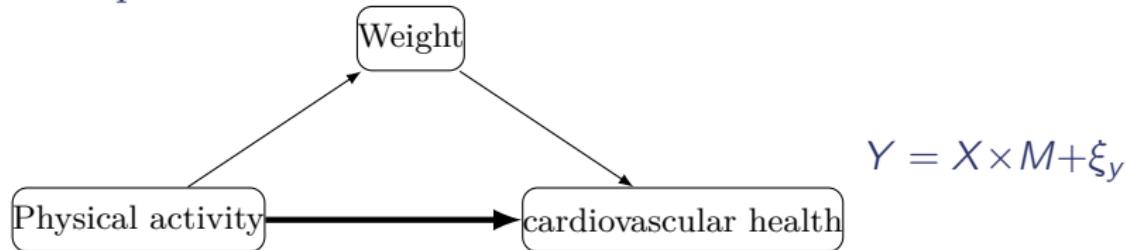
It is possible to use the single-door criterion in a non-parametric model? **No!**

Example



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

Example



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):

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

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

Natural direct effect (NDE)

Natural direct effect (NDE):

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

where $pa_{Y \setminus X}^{x'}$ 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):

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

where $pa_{Y \setminus X}^{x'}$ 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 counterfactual expression

3

Counterfactuals

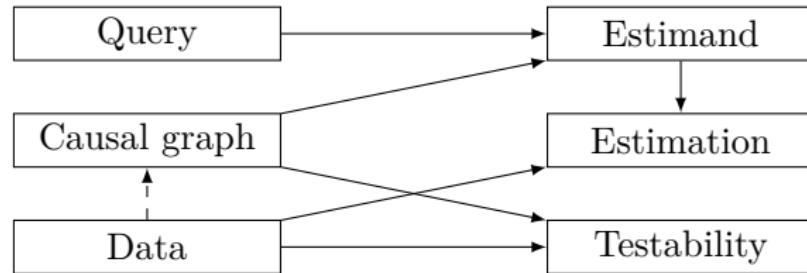
Structural causal models

Counterfactual reasoning

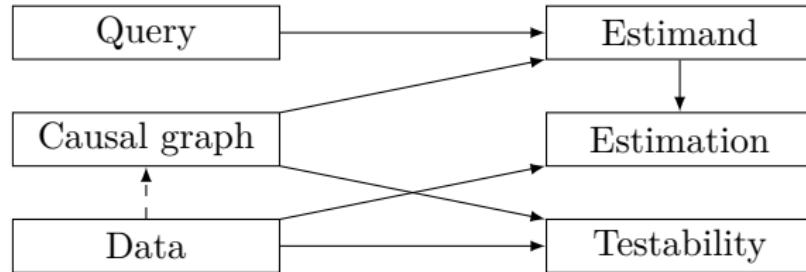
Mediation analysis

Difference with the potential outcome framework

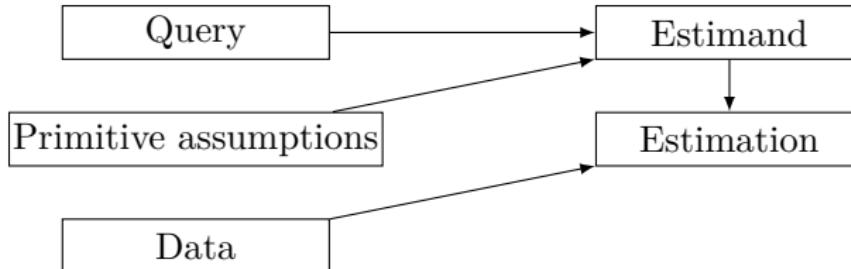
Structural causal models



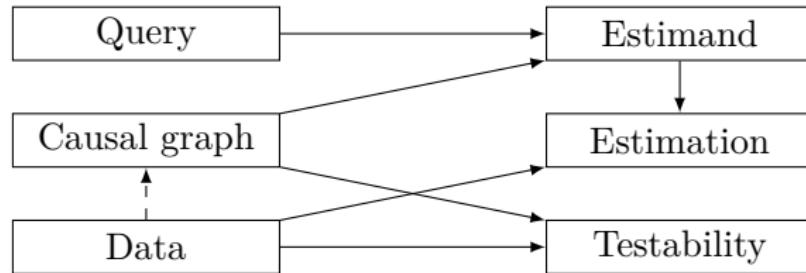
Structural causal models



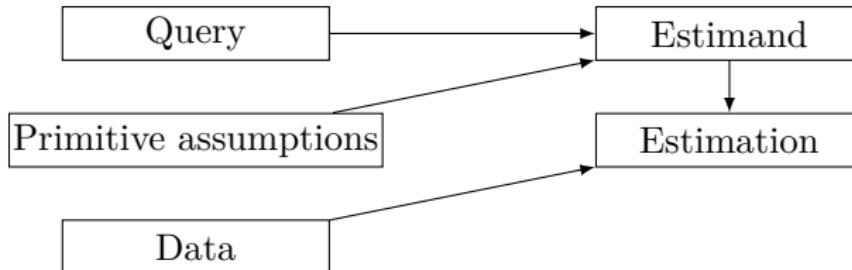
Potential outcome



Structural causal models



Potential outcome



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

4

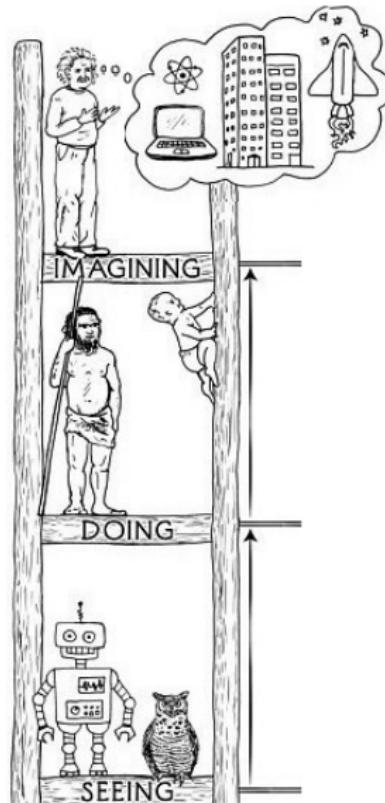
Conclusion

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

6

References

References I

- [1] Aristotle and C. D. C. Reeve. Physics. Hackett Publishing Company, 2018.
- [2] Francis Bacon, Robert L. Ellis, and James Spedding. Novum Organum. Muses' library. Translation of: Novum Organum. 1620. George Routledge and Sons, 1905. isbn: 9780598711663.
- [3] Elias Bareinboim et al. “On Pearl’s Hierarchy and the Foundations of Causal Inference (1st edition)”. In: Probabilistic and Causal Inference: the Works of Judea Pearl. Ed. by Hector Geffner, Rita Dechter, and Joseph Halpern. ACM Books, 2022, pp. 507–556.

References II

- [4] 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.
- [5] M.A. Hernan and J.M. Robins. Causal Inference: What If. Chapman & Hall/CRC Monographs on Statistics & Applied Probab. CRC Press, 2023. isbn: 9781420076165.
- [6] David Hume. A Treatise of Human Nature. Oxford University Press, 1738.
- [7] David Hume. An Enquiry Concerning Human Understanding. Oxford University Press, 1748.
- [8] Guido W. Imbens and Donald B. Rubin.
Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press, 2015.

References III

- [9] David Lewis. “Causation”. In: Journal of Philosophy 70.17 (1973), pp. 556–567.
- [10] 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.
- [11] Judea Pearl. Causality: Models, Reasoning, and Inference. Cambridge University Press, 2000.
- [12] Judea Pearl. “Direct and Indirect Effects”. In: Probabilistic and Causal Inference (2001).
- [13] Judea Pearl and Dana Mackenzie. The Book of Why: The New Science of Cause and Effect. Basic Books, 2018.

References IV

- [14] F. Redi, M. Bigelow, and R.P. Bigelow.
Experiments on the Generation of Insects. Translation of:
Esperienze intorno alla generazione degl'insetti. 1668. Open court publishing Company, 1909.
- [15] Hans Reichenbach. The Direction of Time. Dover Publications, 1956.
- [16] 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.

References V

- [17] 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.
- [18] 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>.
- [19] Peter Spirtes, Clark Glymour, and Richard Scheines. Causation, Prediction, and Search. 2nd. MIT press, 2000.
- [20] 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.

References VI

- [21] Jerzy Splawa-Neyman, D. M. Dabrowska, and T. P. Speed. “On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.”. In: Statistical Science 5.4 (1990), pp. 465–472. issn: 08834237.
- [22] Patrick Suppes. A Probabilistic Theory of Causality. Amsterdam: North-Holland Pub. Co., 1970.
- [23] Thomas Verma and Judea Pearl. “Equivalence and synthesis of causal models”. In: UAI '90: Proceedings of the Sixth Annual Conference on Uncertainty Ed. by Piero P. Bonissone et al. Elsevier, 1990, pp. 255–270.
- [24] Sewall Wright. “Correlation and causation”. In: Journal of agricultural research 20.7 (1921), pp. 557–585.

References VII

- [25] Sewall Wright. “The Relative Importance of Heredity and Environment in Determining the Piebald Pattern of Guinea-Pigs”. In: Proceedings of the National Academy of Sciences of the United States 6.6 (1920), pp. 320–332. issn: 00278424. (Visited on 11/02/2022).

MERCI POUR VOTRE ATTENTION

