Causal discovery

Daria Bystrova daria.bystrova@inserm.fr

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

June 2025

Santé Sorbonne Université

Sommaire

SANTÉ SORBONNE UNIVERSITÉ

1

Preliminaries

Reminder: Causal DAGs

A causal DAG $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ is a directed acyclic graph where

- $\mathbb{V}$ is the set of vertices representing random variables.
- $\mathbb{E}$ is the set of directed edges representing causal relations between these variables.

$\mathcal{G}$ should be compatible with the probability distribution of $\mathbb{V}$, meaning:

Not all DAGs are causal!

Basic structures:

Fork

$A \longleftarrow C \longrightarrow B$

Chain

$A \longrightarrow C \longrightarrow B$

V-structure

$A \longrightarrow C \longleftarrow B$

Reminder: Blocked path and d-separation

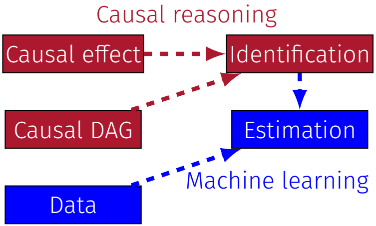
A path is said to be blocked by a set of vertices $\mathbb{Z} \in \mathbb{V}$ if:

- 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 $\mathbb{Z}$.

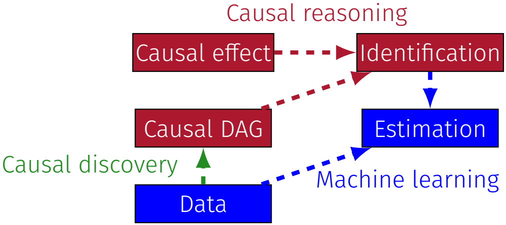
Given disjoint sets $\mathbb{X}, \mathbb{Y}, \mathbb{Z} \subseteq \mathbb{V}$, we say that $\mathbb{X}$ and $\mathbb{Y}$ are d-separated by $\mathbb{Z}$ if every path between a vertex in $\mathbb{X}$ and a vertex in $\mathbb{Y}$ is blocked by $\mathbb{Z}$ and we write $\mathbb{X} \perp\!\!\!\perp_{\mathcal{G}}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

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

What if the causal DAG is unknown?



What if the causal DAG is unknown?

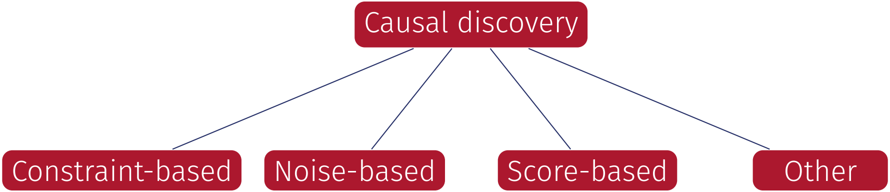


SANTÉ SORBONNE UNIVERSITÉ

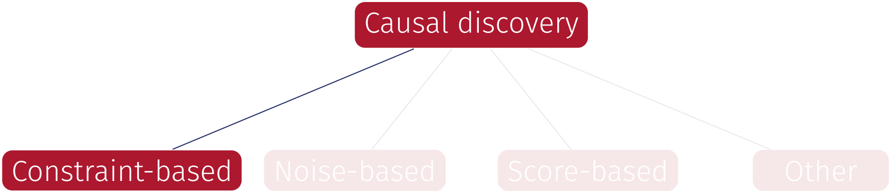
2

Constraint-based causal discovery

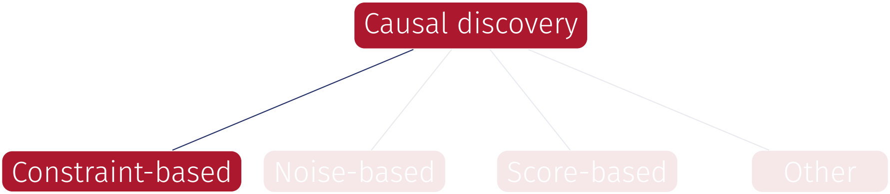
Families of causal discovery methods



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.

Causal discovery using conditional independencies

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

Causal discovery using conditional independencies

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

Causal discovery using conditional independencies

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

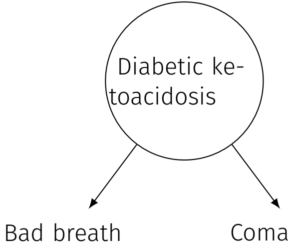
We cannot even construct the skeleton of the graph because

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

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

We cannot even construct the skeleton of the graph because

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

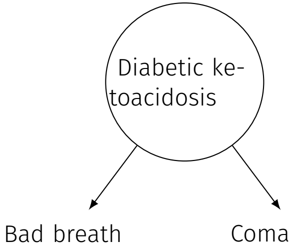


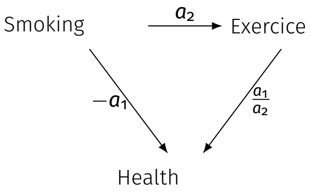
Causal discovery using conditional independencies

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

We cannot even construct the skeleton of the graph because

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





Additional assumptions

Causal sufficiency: No hidden confounding.

SANTÉ SORBONNE UNIVERSITÉ

Additional assumptions

Causal sufficiency: No hidden confounding.

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

Additional assumptions

Causal sufficiency: No hidden confounding.

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

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

Additional assumptions

Causal sufficiency: No hidden confounding.

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

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

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

Additional assumptions

Causal sufficiency: No hidden confounding.

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

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

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

Markov equivalence class [7]

Equivalence in terms of conditional independence

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

Markov equivalence class [7]

Equivalence in terms of conditional independence

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

Theorem

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

Markov equivalence class [7]

Equivalence in terms of conditional independence

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

Theorem

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

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

Markov equivalence class [7]

Equivalence in terms of conditional independence

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

Theorem

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

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

Finding skeleton and v-structures

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

Finding skeleton and v-structures

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 DAG $\mathcal{G}$ with vertex $\mathbb{V}$ then:

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

Finding skeleton and v-structures

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

Theorem

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

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

Finding skeleton and v-structures

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

Theorem

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

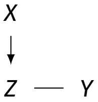
- *For $X, Y \in \mathbb{V}$, X and Y are adjacent iff $\forall \mathbb{S} \subseteq \mathbb{V} \setminus \{X, Y\}$, $X \not\perp\!\!\!\perp_P Y \mid \mathbb{S}$;*
- *For $X, Y, Z \in \mathbb{V}$ such that X is adjacent to Z and Z is adjacent to Y and X and Y are not adjacent, $X \rightarrow Z \leftarrow Y$ 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 $P(\mathbb{V})$;
- Given the skeleton of $\mathcal{G}$, point 2 can be used to find all v-structures.

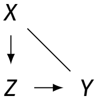
Meek orientation rules [3]

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

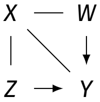
Meek-Rule 1:



Meek-Rule 2:



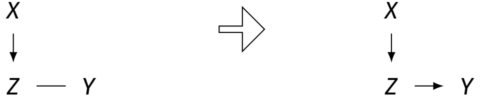
Meek-Rule 3:



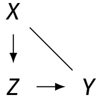
Meek orientation rules [3]

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

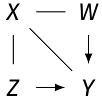
Meek-Rule 1:



Meek-Rule 2:



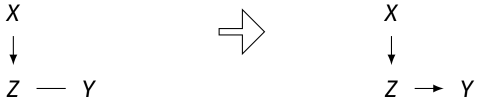
Meek-Rule 3:



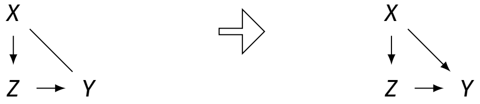
Meek orientation rules [3]

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

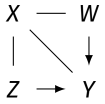
Meek-Rule 1:



Meek-Rule 2:



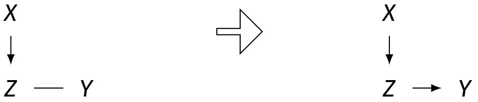
Meek-Rule 3:



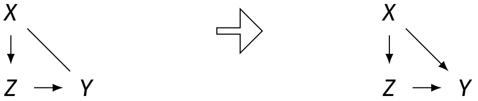
Meek orientation rules [3]

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

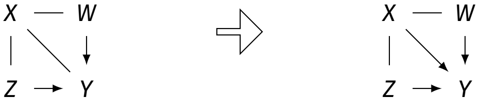
Meek-Rule 1:



Meek-Rule 2:



Meek-Rule 3:



The PC algorithm [6]

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

The PC algorithm [6]

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

The PC algorithm [6]

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

Theorem ([6])

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

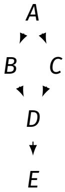
The optimal strategy of the PC algorithm [6]

Pruning unnecessary edges in optimal way:

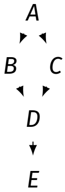
- use a sequential conditioning strategy, where the size of the conditioning set increases progressively from 1 to p-2
- the conditioned set is the subset of the set of variables adjacent to tested variables
- storing the separation sets, which can later be used for orienting v-structures and other triplets

PC in action

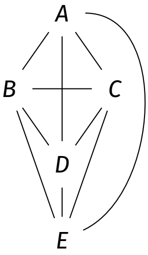
- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness



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

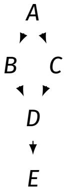


Skeleton construction:

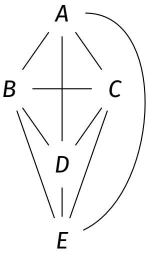


$card = 0$

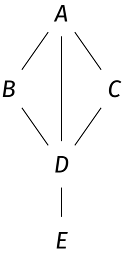
- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness



Skeleton construction:



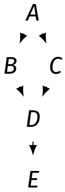
$card = 0$



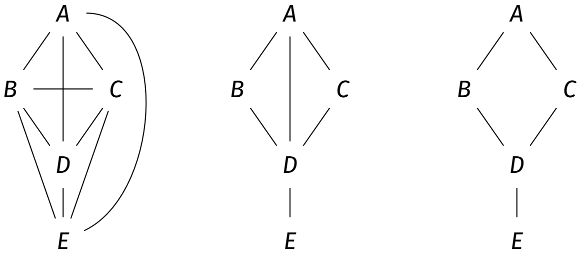
$card = 1$

PC in action

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

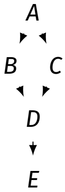


Skeleton construction:

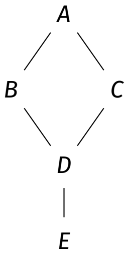


$card = 0$ $card = 1$ $card = 2$

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

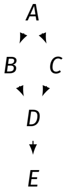


Orientation:

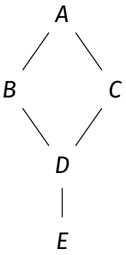


PC in action

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

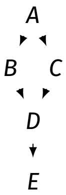


Orientation:

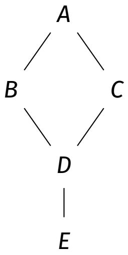


- Finding V-structures

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



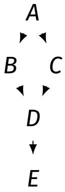
Orientation:



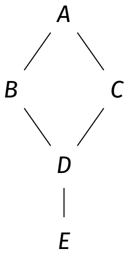
- Finding V-structures
  - $B \perp\!\!\!\perp_P C \mid A \implies B \rightarrow D \leftarrow C$

PC in action

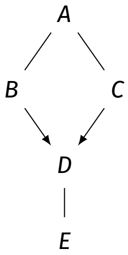
- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness



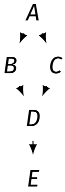
Orientation:



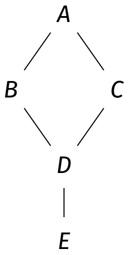
- Finding V-structures
  - $B \perp\!\!\!\perp_P C \mid A \implies B \rightarrow D \leftarrow C$



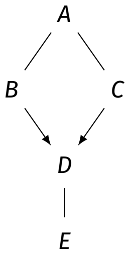
- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness



Orientation:

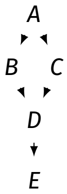


- Finding V-structures
  - $B \perp\!\!\!\perp_P C \mid A \implies B \to D \leftarrow C$
- Meek-Rule 1

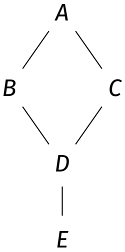


PC in action

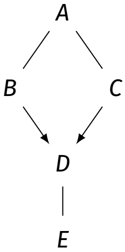
- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness



Orientation:

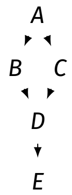


- Finding V-structures
  - $B \perp\!\!\!\perp_P C \mid A \implies B \rightarrow D \leftarrow C$
- Meek-Rule 1
  - $B \rightarrow D \;\&\; D - E \implies D \rightarrow E$

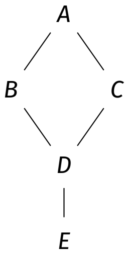


PC IN ACTION

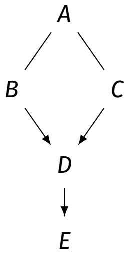
- Suppose the causal DAG on the right
- Input: Observational data
- Output: CPDAG
- Assumptions: causal sufficiency, faithfulness



Orientation:



- Finding V-structures
  - $B \perp\!\!\!\perp_P C \mid A \implies B \to D \leftarrow C$
- Meek-Rule 1
  - $B \to D \;\&\; D - E \implies D \to E$



Incorporating background knowledge

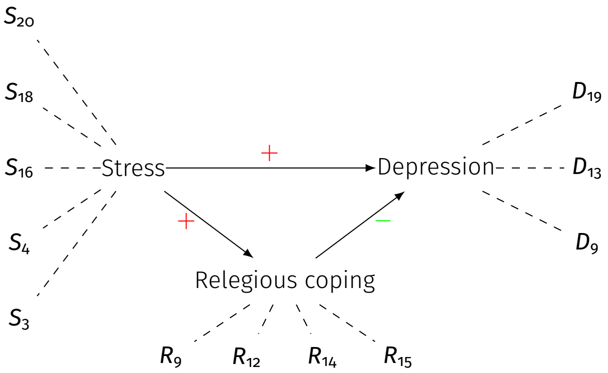
The PC algorithm can effectively incorporate background knowledge in the form of:

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

Real application of PC [4]

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

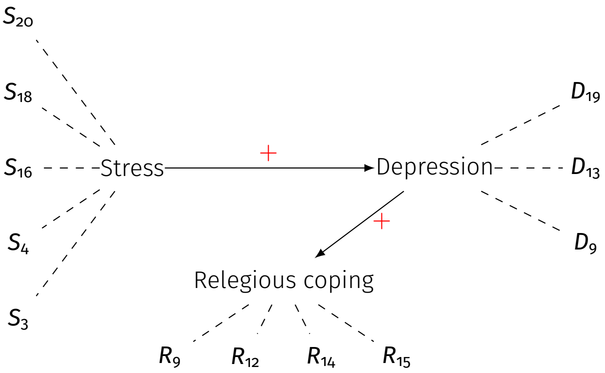
Specified graph



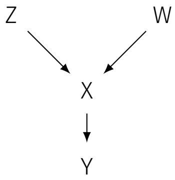
Real application of PC [4]

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

Inferred graph (assuming stress is temporally prior)



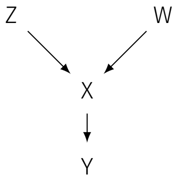
Beyond causal sufficiency (1/2)



Y-structure

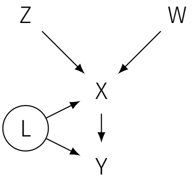
$$Z \perp\!\!\!\perp_P W$$
$$Z \not\perp\!\!\!\perp_P W \mid X$$
$$Y \not\perp\!\!\!\perp_P Z$$
$$Y \perp\!\!\!\perp_P Z \mid X$$
$$Y \not\perp\!\!\!\perp_P W$$
$$Y \perp\!\!\!\perp_P W \mid X$$

Beyond causal sufficiency (1/2)



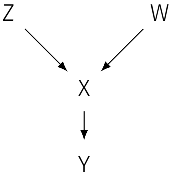
Y-structure

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



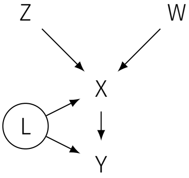
$$Z \perp\!\!\!\perp_P W$$
$$Z \not\!\perp\!\!\!\perp_P W \mid X$$
$$Y \not\!\perp\!\!\!\perp_P Z$$
$$Y \not\!\perp\!\!\!\perp_P Z \mid X$$
$$Y \not\!\perp\!\!\!\perp_P W$$
$$Y \not\!\perp\!\!\!\perp_P W \mid X$$

Beyond causal sufficiency (1/2)



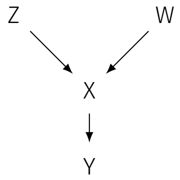
Y-structure

$$Z \perp\!\!\!\perp_P W$$
$$Z \not\!\perp\!\!\!\perp_P W \mid X$$
$$Y \not\!\perp\!\!\!\perp_P Z$$
$$\color{red}{Y \perp\!\!\!\perp_P Z \mid X}$$
$$Y \not\!\perp\!\!\!\perp_P W$$
$$\color{red}{Y \perp\!\!\!\perp_P W \mid X}$$



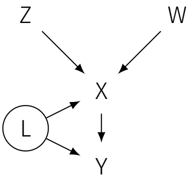
$$Z \perp\!\!\!\perp_P W$$
$$Z \not\!\perp\!\!\!\perp_P W \mid X$$
$$Y \not\!\perp\!\!\!\perp_P Z$$
$$\color{red}{Y \not\!\perp\!\!\!\perp_P Z \mid X}$$
$$Y \not\!\perp\!\!\!\perp_P W$$
$$\color{red}{Y \not\!\perp\!\!\!\perp_P W \mid X}$$

Beyond causal sufficiency (1/2)



Y-structure

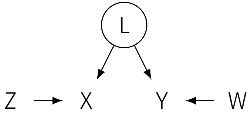
$$Z \perp\!\!\!\perp_P W$$
$$Z \not\!\perp\!\!\!\perp_P W \mid X$$
$$Y \not\!\perp\!\!\!\perp_P Z$$
$$Y \perp\!\!\!\perp_P Z \mid X$$
$$Y \not\!\perp\!\!\!\perp_P W$$
$$Y \perp\!\!\!\perp_P W \mid X$$



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

Pattern of independence can rule out hidden confounding.

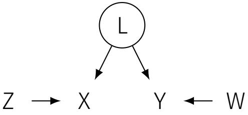
Beyond causal sufficiency (2/2)



W-structure

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

Beyond causal sufficiency (2/2)

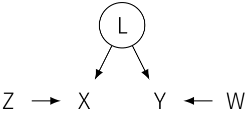


W-structure

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

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

Beyond causal sufficiency (2/2)

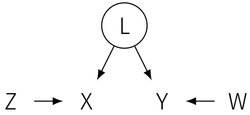


W-structure

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

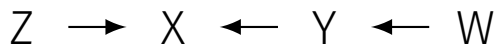
$Z \not\perp\!\!\!\perp_P X$
$X \not\perp\!\!\!\perp_P Y$
$Y \not\perp\!\!\!\perp_P W$
$Z \perp\!\!\!\perp_P W$
$Z \not\perp\!\!\!\perp_P Y$
$X \perp\!\!\!\perp_P W$
$Z \perp\!\!\!\perp_P Y \mid X$
$X \not\perp\!\!\!\perp_P W \mid Y$

Beyond causal sufficiency (2/2)



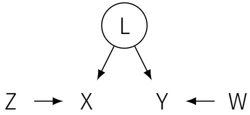
W-structure

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



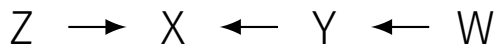
$Z \not\perp\!\!\!\perp_P X$
$X \not\perp\!\!\!\perp_P Y$
$Y \not\perp\!\!\!\perp_P W$
$Z \perp\!\!\!\perp_P W$
$Z \perp\!\!\!\perp_P Y$
$X \not\perp\!\!\!\perp_P W$
$Z \not\perp\!\!\!\perp_P Y \mid X$
$X \perp\!\!\!\perp_P W \mid Y$

Beyond causal sufficiency (2/2)



W-structure

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



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

Pattern of independence can suggest hidden confounding.

A glimpse of the FCI algorithm

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

A glimpse of the FCI algorithm

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

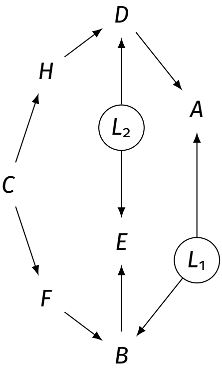
- Skeleton construction is much more complicated;
- Orientation is done using 10 different rules.

A glimpse of the FCI algorithm

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

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

The FCI algorithm in action:



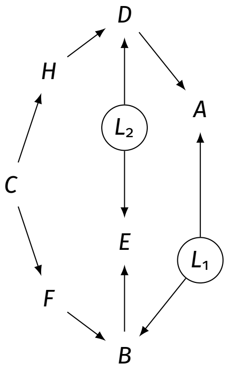
True graph

A glimpse of the FCI algorithm

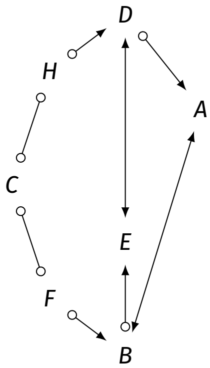
The FCI algorithm extends the PC algorithm to accommodate hidden confounders:

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

The FCI algorithm in action:



True graph



Inferred graph

Pros and cons of constraint-based methods

Pros:

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

Cons:

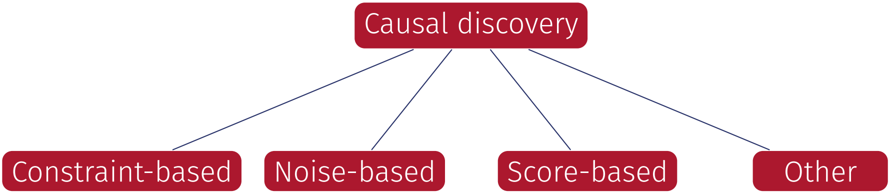
- Recover a partially oriented graph
- The faithfulness assumption is not always accepted

SANTÉ SORBONNE UNIVERSITÉ

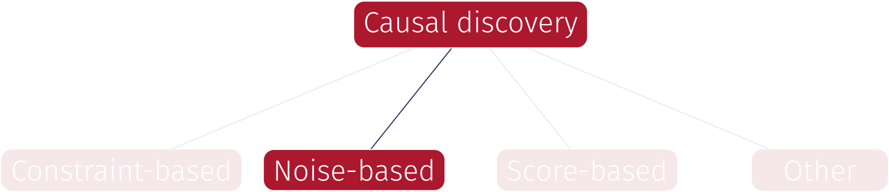
3

Noise-based causal discovery

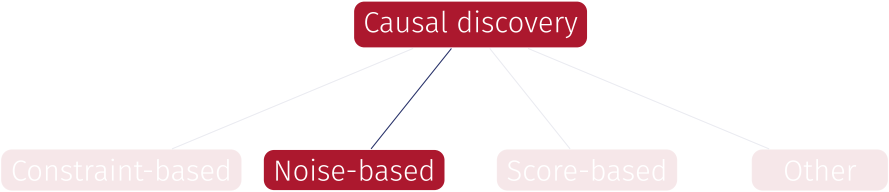
Noise based


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

Noise based

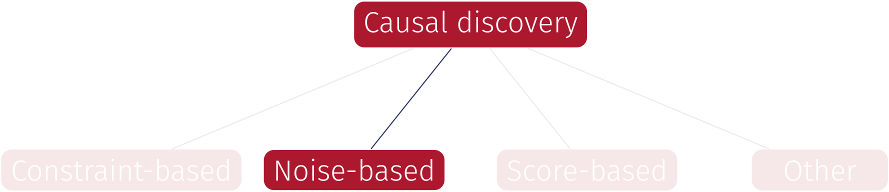

Causal discovery
Constraint-based
Noise-based
Score-based
Other

Noise based



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

Noise based



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

The intuition behind the noise (1/2)

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

The intuition behind the noise (1/2)

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

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.

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.

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

The intuition behind the noise (1/2)

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

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

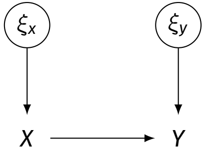
$Y := 2X + \xi_y$?
or
$X := \frac{Y}{2} + \hat{\xi}_x$?

Without further assumption we cannot know.

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

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

The intuition behind the noise (2/2)

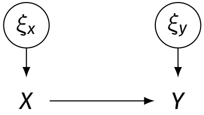


$$M_1 : \begin{cases} X := f_x(\xi_x) \\ Y := f_y(X, \xi_y) \end{cases}$$

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

The experiment that changed everything

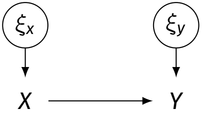
Suppose



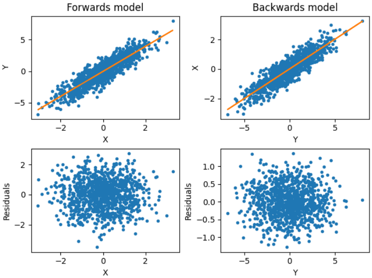
$$X \sim N(0, 1)$$
$$\xi_y \sim N(0, 1)$$
$$Y := 2X + \xi_y$$

The experiment that changed everything

Suppose

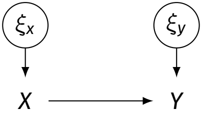


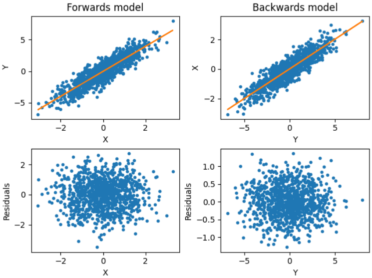
$$X \sim N(0, 1)$$
$$\xi_y \sim N(0, 1)$$
$$Y := 2X + \xi_y$$



The experiment that changed everything

Suppose



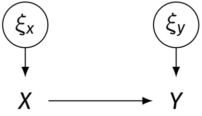


$$X \sim N(0, 1)$$
$$\xi_y \sim N(0, 1)$$
$$Y := 2X + \xi_y$$

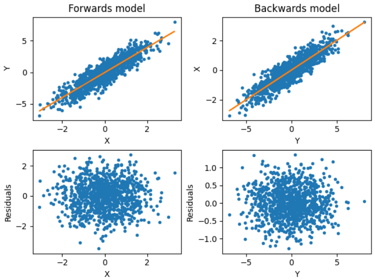
$$X \sim U(0, 1)$$
$$\xi_y \sim U(0, 1)$$
$$Y := 2X + \xi_y$$

The experiment that changed everything

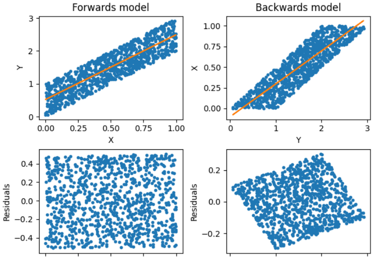
Suppose



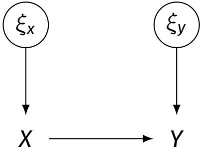
$$X \sim N(0, 1)$$
$$\xi_y \sim N(0, 1)$$
$$Y := 2X + \xi_y$$



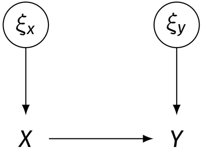
$$X \sim U(0, 1)$$
$$\xi_y \sim U(0, 1)$$
$$Y := 2X + \xi_y$$



The linear case (1/2)



The linear case (1/2)

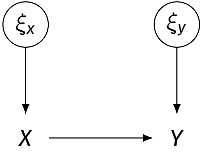


True model:

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

- $X \perp\!\!\!\perp_P \xi_y$

The linear case (1/2)



True model:

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

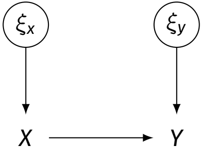
- $X \perp\!\!\!\perp_P \xi_y$

Backwards model:

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

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

The linear case (1/2)



True model:

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

- $X \perp\!\!\!\perp_P \xi_y$

Backwards model:

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

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

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

The linear case (2/2)

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

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

The linear case (2/2)

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

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

Theorem (Darmois-Skitovich)

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

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

are independent, then each X_i is normally distributed.

Linear non-Gaussian models (LiNGAM) [5]

Theorem

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

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

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

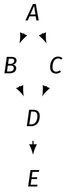
$$X := bY + \hat{\xi}_x, \qquad Y \perp\!\!\!\perp_P \hat{\xi}_x,$$

iff ξ_y *and* X *are Gaussian.*

Similar result for the multivariate case

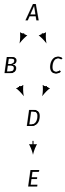
A GLIMPSE OF LiNGAM IN ACTION

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

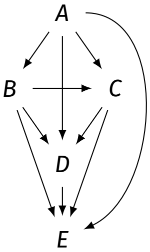


A glimpse of LiNGAM in action

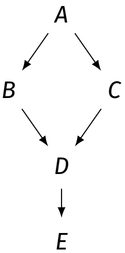
- Suppose the causal DAG on the right
- Input: Observational data
- Output: Causal DAG
- Assumptions: causal sufficiency, minimality, linearity, non-gaussianity



Causal order:



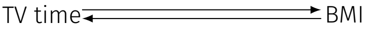
Pruning:

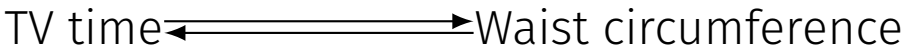


Real application of LiNGAM [2]

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

Specified graphs

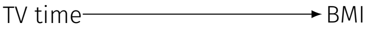


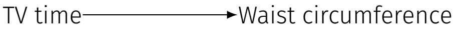


Real application of LiNGAM [2]

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

Inferred graphs





Extensions

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

Pros and cons of noise-based methods

Pros:

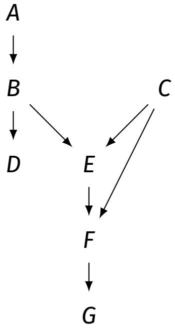
- Recover the complete causal DAG

Cons:

- Semi-parametric

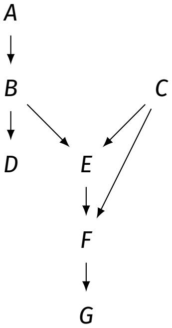
Exercise

What is the CPDAG of the following causal DAG?



Exercise

What will be the output of PC if $A = B$?



4

Additional notes

Practical considerations

To be specified:

- The introduced implementation of PC is order dependent, however there exists order independent implementation introduced by [1]
- A conditional independence measure and test (constaint-based)
- An independece measure and tests (noise-based)
- A significance level in the statistical tests (constaint-based and noise-based)
- Optional: Existing expert knowledge (temporal ordering, forbiden orientations,...)

Usefulness of causal discovery when assumptions are violated

If assumptions are violated, then most causal discovery methods would give a Bayesian network
$\implies$ this Bayesian network could be used for the selection of the variables for the prediction task.

5

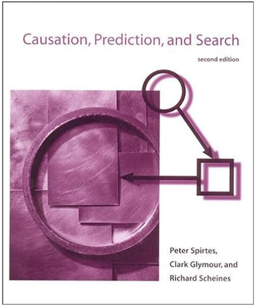
Conclusion

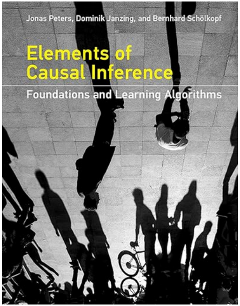
Conclusion

- In general, a causal DAG cannot be discovered from observational data alone.
- However, it can be discovered under certain untestable assumptions.
- If you can construct the causal DAG manually based on domain knowledge, then avoid using causal discovery methods (except for prediciton).
- Causal discovery is useful only when the causal DAG is unknown and needs to be inferred from data.

CAUSAL DISCOVERY REFERENCES

- "Review of causal discovery methods based on graphical models" by Clark Glymour, Kun Zhang, and Peter Spirtes,2019
- "Causal discovery algorithms: A practical guide" by Daniel Malinsky and David Danks, 2018





6

References

References I

[1] Diego Colombo, Marloes H Maathuis, et al. "Order-independent constraint-based causal structure learning.". In: J. Mach. Learn. Res. 15.1 (2014), pp. 3741–3782.

[2] 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.

[3] Christopher Meek. "Causal Inference and Causal Explanation with Background Knowledge". In: Proceedings of the Eleventh Conference on Uncertainty in Artificial Inte UAI'95. Montréal, Qué, Canada: Morgan Kaufmann Publishers Inc., 1995, 403–410. ISBN: 1558603859.

References II

[4] 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.

[5] 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.

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

References III

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

MERCI POUR VOTRE ATTENTION

SANTÉ
SORBONNE
UNIVERSITÉ