Introduction to causal discovery

Anne Helby Petersen

Program

- Welcome + introduction to DAGs
- Human generated DAG
- Causal discovery: PC algorithm and CPDAGs
- Work with PC in R
- TPC and TPDAGs
- Work with TPC in R
- LUNCH (11-12)

Program

- Welcome + introduction to DAGs
- Human generated DAG
- Causal discovery: PC algorithm and CPDAGs
- Work with PC in R
- TPC and TPDAGs
- Work with TPC in R
- LUNCH (11-12)
- More on TPC
- Human generated DAG with latents
- Introduction to PAGs, FCI algorithm and TFCI algorithm
- Work with (T)FCI in R
- Final remarks + further perspectives

Program

- Welcome + introduction to DAGs
- Human generated DAG
- Causal discovery: PC algorithm and CPDAGs
- Work with PC in R
- TPC and TPDAGs
- Work with TPC in R
- LUNCH (11-12)
- More on TPC
- Human generated DAG with latents
- Introduction to PAGs, FCI algorithm and TFCI algorithm
- Work with (T)FCI in R
- Final remarks + further perspectives

Manage your own breaks during exercise bits.

Directed acyclic graphs (DAGs)

$X \rightarrow Y$ means that X is a **direct cause** of Y relative to the other variables in the DAG:

Cause: Manipulating X affects Y, but not vice versa

Direct cause: There does not exist any mediator variable M such that if M is further manipulated, Y is no longer affected by changes in X.

Directed acyclic graphs (DAGs)

$X \rightarrow Y$ means that X is a **direct cause** of Y relative to the other variables in the DAG:

Cause: Manipulating X affects Y, but not vice versa

Direct cause: There does not exist any mediator variable M such that if M is further manipulated, Y is no longer affected by changes in X.

We use **"family language"** to discuss structures in DAGs: parent, child, ancestor, descendant, ...

Directed acyclic graphs (DAGs)

$X \rightarrow Y$ means that X is a **direct cause** of Y relative to the other variables in the DAG:

Cause: Manipulating X affects Y, but not vice versa

Direct cause: There does not exist any mediator variable M such that if M is further manipulated, Y is no longer affected by changes in X.

We use **"family language"** to discuss structures in DAGs: parent, child, ancestor, descendant, ...

Two specific structures have **additional names**:

Confounder: Z is a confounder of the X-Y relationship if it is an ancestor of both X and Y.

Collider: W is a collider of the X-Y relationship if it is a descendant of both X and Y.

d-separation

We use **d-separation** to understand causal flow in a DAG: Two variables X and Y are d-separated by at set of variables $\mathbf{Z} = \{Z_1, ..., Z_k\}$ if the following two conditions hold:

1. All causal paths or confounder paths between X and Y include a variable Z_i from $\mathbf{Z}$:

$$X \to ... \to Z_i \to ... \to Y$$

$$X \leftarrow ... \leftarrow Z_i \leftarrow ... \leftarrow Y$$

$$X \leftarrow ... \leftarrow Z_i \to ... \to Y$$

... or combinations of the above.

2. No collider paths between X and Y include a variable Z_i from $\mathbf{Z}$, nor a descendant of any variable in $\mathbf{Z}$:

$X \to ... \to Z_i \leftarrow ... \leftarrow Y$ not allowed!

Why DAGs?

- Tool used for visualizing and summarizing causal assumptions and their implications
- Mathematically neat: A lot of mathematical theory exists, e.g., d-separation implies conditional independence (Markov property): If X and Y are d-separated by Z, it implies that X and Y are conditionally independent given Z.
- Useful for guiding causal inference (effect identifiability, adjustment sets)
- But note: (Almost) always a crude simplification of the real world!

Exercise 1: Human generated DAG

Go to the course website and find Exercise: Human Generated DAG.

Follow the instructions to create a DAG individually – we will follow up with a plenary discussion.