Week 1: Causal Inference and Directed Acyclic Graphs

Causal Inference & Structural Equation Modeling

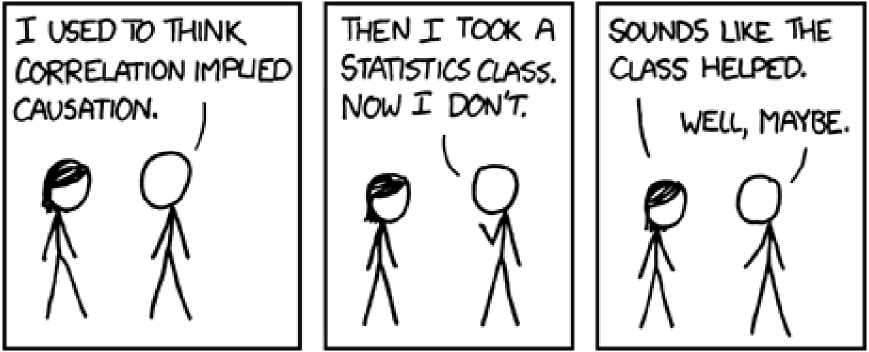
Noémi K. Schuurman
based on slides by Oisín Ryan

February 2022

Overview

- **Causal inference - intro**
- Causal Graphs, DAGs and SCMs
- Statistical dependencies implied by DAG structures
- Causal Discovery

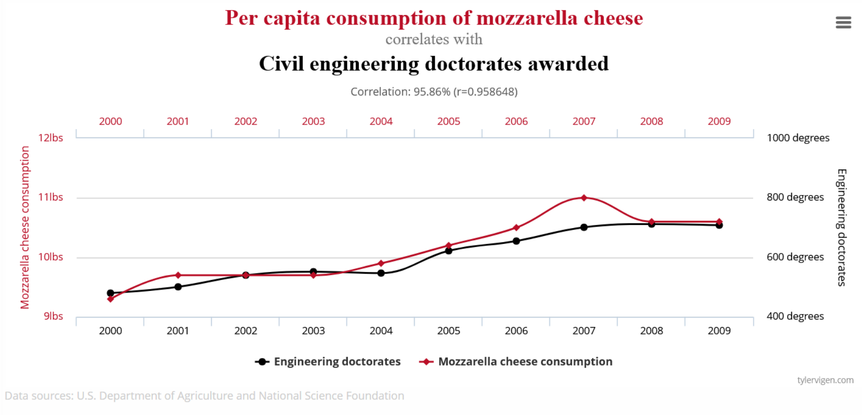
Correlation =/= Causation



https://imgs.xkcd.com/comics/correlation.png

Spurious Assocations

A **spurious association** is a non-causal association.



Check out: http://tylervigen.com/spurious-correlations

Correlation =/= Causation


bewonen. Het is dan ook steviger en duurzamer gebouwd, beter geïsoleerd, functioneler en heeft de uitstraling van een woonhuis – maar dan mini. Hoewel veel tiny houses op wielen staan, zijn het volgens de wet geen woonwagens. In een woonwagen mag je ook niet permanent wonen, terwijl dat bij een tiny house juist het doel is.

9 **Klein wonen (en ontspullen) maakt gelukkiger.** Wie klein gaat wonen, denkt ook vanzelf na over de vraag wat je nou echt nodig hebt en wat belangrijk voor je is. En dat brengt je dichter bij wie je bent en hoe je wilt leven. Je hoeft er niet meteen kleiner voor te gaan wonen, maar onderzoekers hebben wel een link ontdekt tussen mensen die veel spullen in huis hebben en het stresshormoon cortisol. Hoe meer spullen, hoe meer stress – zo bleek. En hoe paradoxaal het ook klinkt: minder woonoppervlakte levert volgens tiny-house-bewoners juist méér ruimte op. Namelijk: ruimte in je hoofd. Zonder de ballast van een groot huis, een hypotheek, al die bezittingen en verplichtingen, houd je meer tijd

Taken from magazine Flow "Het grote boek van minder"

Correlation =/= Causation

Living small scale (and getting rid of your "stuff") makes you happier. Who starts living on a small scale, automatically thinks about what they really need and what is important to them. And that brings you closer to who you are, and how you want to live. You don't have to immediately move to a smaller house, but researchers did find a link between people who have a lot of stuff in their house and the stress-hormone cortisol. The more stuff, the more stress - they found. And although it may sound paradoxical: according to tiny-house-residents less living space actually result in móre space. That is: space in your head. Without the burden of a large house, a morgage, all those possesions and obligations, you have more time and energy left to do what makes you happy. Reading, gardening, volunteer work, walking,

Translated from magazine Flow "Het grote boek van minder"

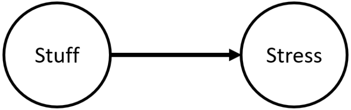
Causal interpretation by Flow:

Correlation =/= Causation

Living small scale (and getting rid of your "stuff") makes you happier. Who starts living on a small scale, automatically thinks about what they really need and what is important to them. And that brings you closer to who you are, and how you want to live. You don't have to immediately move to a smaller house, but researchers did find a link between people who have a lot of stuff in their house and the stress-hormone cortisol. The more stuff, the more stress - they found. And although it may sound paradoxical: according to tiny-house-residents less living space actually resulst in móre space. That is: space in your head. Without the burden of a large house, a morgage, all those possesions and obligations, you have more time and energy left to do what makes you happy. Reading, gardening, volunteer work, walking,

Translated from magazine Flow "Het grote boek van minder"

Causal interpretation by Flow:

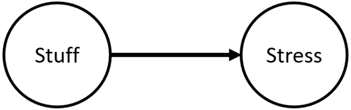


Correlation =/= Causation

Living small scale (and getting rid of your "stuff") makes you happier. Who starts living on a small scale, automatically thinks about what they really need and what is important to them. And that brings you closer to who you are, and how you want to live. You don't have to immediately move to a smaller house, but researchers did find a link between people who have a lot of stuff in their house and the stress-hormone cortisol. The more stuff, the more stress - they found. And although it may sound paradoxical: according to tiny-house-residents less living space actually result in móre space. That is: space in your head. Without the burden of a large house, a morgage, all those possesions and obligations, you have more time and energy left to do what makes you happy. Reading, gardening, volunteer work, walking,

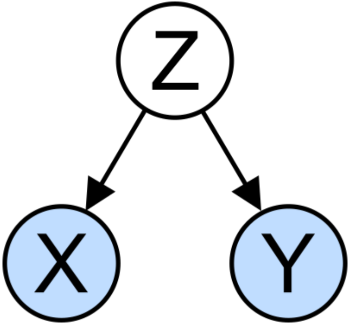
Translated from magazine Flow "Het grote boek van minder"

Causal interpretation by Flow:



Alternative Explanations?

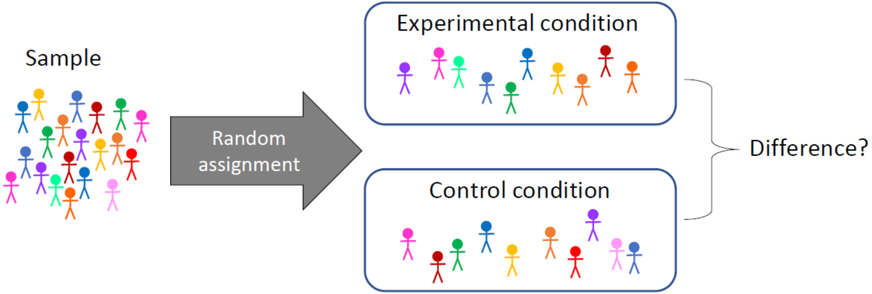
Confounding



Confounder: A variable (Z) that influences both the independent variable (X) and dependent variable (Y), causing a spurious association between them.

Solution a.: Experiments/ "Randomized Controlled Trials"

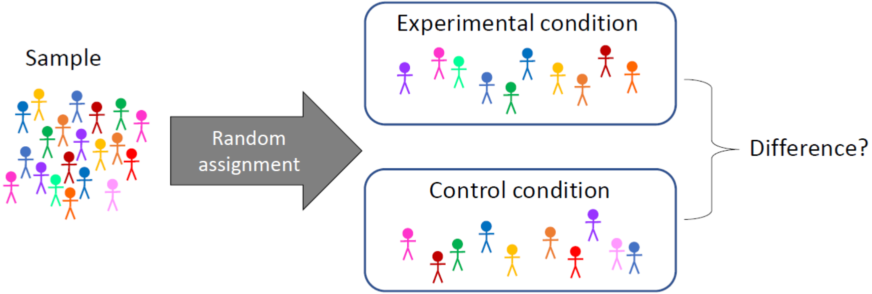
Random assignment ensures that the independent variable is not affected by confounders.



Hence: Difference between groups = effect of the treatment.

Solution a.: Experiments/ "Randomized Controlled Trials"

Random assignment ensures that the independent variable is not affected by confounders.



Hence: Difference between groups = effect of the treatment.

Problems:

- Can go wrong: Drop-out, switching groups, contamination, etc.
- Often infeasible.

Solution b.: Avoid doing causality at all costs.

"I just care about description".

"I just care about prediction".

Solution b.: Avoid doing causality at all costs.

"I just care about description".

"I just care about prediction".

Solution c.: Secretly do causality, hope people won't notice.

Avoid explicitly talking about causality in your research.

- Hernán, M. A. (2018). The C-word: scientific euphemisms do not improve causal inference from observational data. *American journal of public health.*
- Hernán, M. (2018). The C-word: the more we discuss it, the less dirty it sounds.
- Haber, N., Smith, E. R., Moscoe, ... & CLAIMS research team. (2018). Causal language and strength of inference in academic and media articles shared in social media (CLAIMS): A systematic review. *PloS one.*
- Hamaker, E. L., Mulder, J. D., & van IJzendoorn, M. H. (2020). Description, prediction and causation: Methodological challenges of studying child and adolescent development. *Developmental cognitive neuroscience.*

the underlying dynamics between X and Y
"X would engage increased Y"
"to what extent do genes drive an association between X and Y"
"can associations between X and Y be attributed to variations in"
"X would modulate Y"
"is X able to protect against Y"
"X is affected by Y"
"X may form a target of intervention"
"X can be primed, depending on Y"
"are differences in X accounted for by Y"
"does X elicit Y"
"the spillover from X to Y"
"(...) processes (...)"
"the impact of X on Y"
"the amount of X explained by Y"
"the interplay between X and Y"

Solution d.: Do causal inference explicitly, in a principled way.

The Only Thing That Can Stop Bad Causal Inference Is Good Causal Inference

AUTHORS
Julia M. Rohrer, Stefan Schmukle, Richard McElreath

Explicitly:

- Be open about your causal interests, research questions.
- Specify your causal questions as clearly as possible: e.g. what kind of treatment/intervention are you interested in exactly.

Solution d.: Do causal inference explicitly, in a principled way.

The Only Thing That Can Stop Bad Causal Inference Is Good Causal Inference

AUTHORS
Julia M. Rohrer, Stefan Schmukle, Richard McElreath

Explicitly:

- Be open about your causal interests, research questions.
- Specify your causal questions as clearly as possible: e.g. what kind of treatment/intervention are you interested in exactly.

Principled way:

- Be clear about what assumptions you are making for your causal inference.
- Do you inference in a systematic way, for example using a causal inference framework.
- Make use of triangulation: Do different studies that require different assumptions, to fill the gaps.

Causal Inference Frameworks & Techniques

- Structural Causal Models & Causal Graphs (Pearl) - lecture 1
- Potential Outcomes Framework (Rubin) - lecture 2

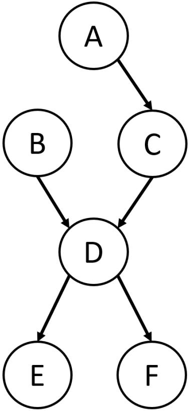
From wiki "Structural Causal Models - History":
"Sociologists originally called causal models **structural equation modeling**, but once it became a rote method, it lost its utility, leading some practitioners to reject any relationship to causality. Economists adopted the algebraic part of path analysis, calling it simultaneous equation modeling. However, economists still avoided attributing causal meaning to their equations (Pearl, Book of Why)."

Overview

- Causal inference - intro
- Causal Graphs - DAGS
- Statistical dependencies implied by DAG structures
- Causal Discovery

Causal Graphs

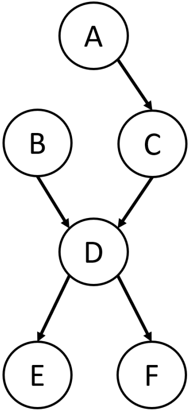
Example:



- The circles, or 'nodes' represent variables.
- Arrows going directly from variable X to Y, represent that X directly causes Y.
- No arrow, means no causal effect.

Causal Graphs

Example:

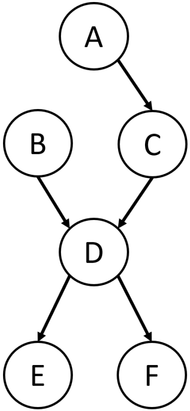


- The circles, or 'nodes' represent variables.
- Arrows going directly from variable X to Y, represent that X directly causes Y.
- No arrow, means no causal effect.
- 'Parents' directly cause 'Children'.
- 'Ancestors' directly or indirectly cause 'descendants'.

Take a causal effect of X on Y as: 'if we intervene and change the value of X, then as a result of this the value of Y will change'.

Directed Acyclical Graphs

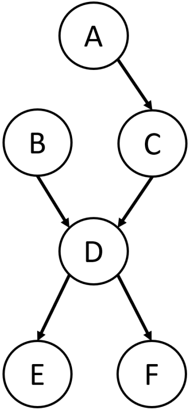
Example:



- Not necessarily causal graphs - but we will use them as causal graphs.

Directed Acyclical Graphs

Example:

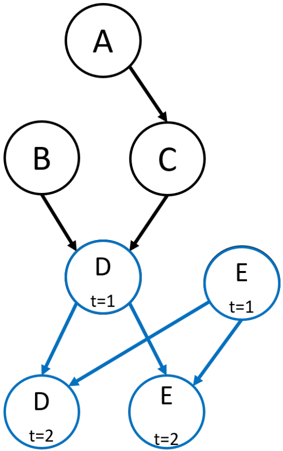


The example in this figure is a DAG.

- Not necessarily causal graphs - but we will use them as causal graphs.
- Directed - Only directed edges/arrows allowed!
- Acyclic - no bidirectional effects or 'loops'

Directed Acyclical Graphs

Example:



- Not necessarily causal graphs - but we will use them as causal graphs.
- Directed - Only directed edges/arrows allowed!
- Acyclic - no bidirectional effects or 'loops'

In this figure you see the loophole to have 'loops' and 'bidirectional' effects in a DAG: Include time-specific variables!

Tying Causal Directed Acyclical Graphs to Statistical Relationships

From a causal DAG, we can read of causal relationships, but also implied statistical relationships among the variables.
To do this, we use the following condition:

Global Markov Condition:

Every variable (node) is conditionally independent of its non-descendants, given its parents:
X $\perp\!\!\!\perp$non-descendants$(X)|$parents(X).

Tying Causal Directed Acyclical Graphs to Statistical Relationships

From a causal DAG, we can read of causal relationships, but also implied statistical relationships among the variables.
To do this, we use the following condition:

Global Markov Condition:

Every variable (node) is conditionally independent of its non-descendants, given its parents:
X $\perp\!\!\!\perp$non-descendants$(X)|$parents(X).

This means, to read of statistical dependencies from the graph, we can use markov factorization:

Markov Factorization:

The joint density of the variables $P(X_1, \ldots X_n)$ is given by $\prod_{i=1}^{n} P(X_i \mid Parents(X_i))$

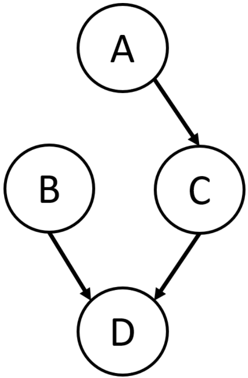
Note: This is based on the chain rule for random variables! Check out the wiki for "Chain rule probability"

Markov Factorization Example

To read of statistical dependencies from the graph, we can use markov factorization to make things simpler:

Example:

Lets get the joint distribution of A, B, C, and D.



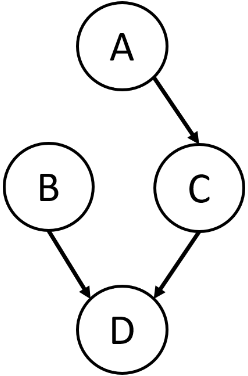
Markov Factorization Example

To read of statistical dependencies from the graph, we can use markov factorization to make things simpler:

Example:

Lets get the joint distribution of A, B, C, and D.

- Chain rule: $P(A, B, C, D) = P(D|A, B, C)P(C|A, B)P(B|A)P(A)$
- Markov Factorization: $P(A, B, C, D) = P(D|B, C)P(C|A)P(B)P(A)$



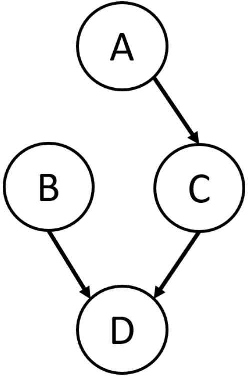
Markov Factorization Example

To read of statistical dependencies from the graph, we can use markov factorization to make things simpler:

Example:

Lets get the joint distribution of A, B, C, and D.

- Chain rule: $P(A, B, C, D) = P(D|A, B, C)P(C|A, B)P(B|A)P(A)$
- Markov Factorization: $P(A, B, C, D) = P(D|B, C)P(C|A)P(B)P(A)$



Markov Factorization based on the DAG makes things simpler.

Structural Causal Model

Based on a DAG, we know exactly what conditional probability densities we need, to directly calculate the joint probability of two or more variables in the DAG.[1]

Of course, to actually calculate this, we also need to specify the functional forms of those densities in some way. We do this with a Structural Causal Model.

[1] This is also super handy in the context of Bayesian statistics...to figure what conditional distributions we need to obtain a joint posterior. In Bayesian stats context, DAGs are called 'Bayesian Networks'.

Structural Causal Models

A Structural Causal Model:
a set of equations describing causal relations between variables, and include noise terms ϵ that are independent of other noise terms and variables.

Notes:

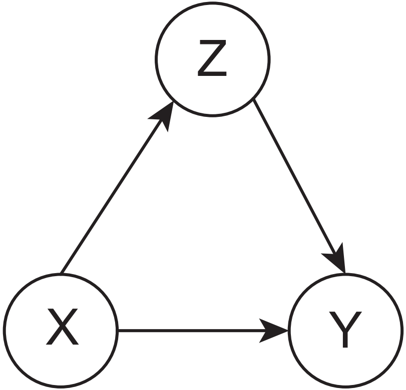
- We consider variables that are stochastic/probabalistic.
- X causing Y means something like if I intervene and change the value of X, this changes the probabilities of the outcomes of Y.
- The noise terms are often not explicitly drawn in causal graphs (but in SEM path models they often are!). This is possible because they are (assumed) independent of all other variables.
- Specifying precise functional forms of the relationships, usually means introducing more assumptions (e.g., linearity).

Example: SCM with normal noise terms and linear conditional relationships

$$
\begin{aligned}
X &:= \epsilon_X \\
Z &:= 2X + \epsilon_Z \\
Y &:= 1X + 2Z + \epsilon_Y
\end{aligned}
$$

where

- $\epsilon_X, \epsilon_Z, \epsilon_Y$ are independently and identically distributed $\sim \mathcal{N}(0,1)$



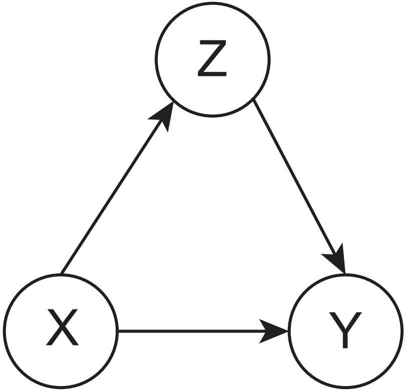
Example: SCM with normal noise terms and linear conditional relationships

$$X := \epsilon_X$$
$$Z := 2X + \epsilon_Z$$
$$Y := 1X + 2Z + \epsilon_Y$$

where $\epsilon_X, \epsilon_Z, \epsilon_Y$ are independently and identically distributed $\sim \mathcal{N}(0, 1)$

Conditional distributions:

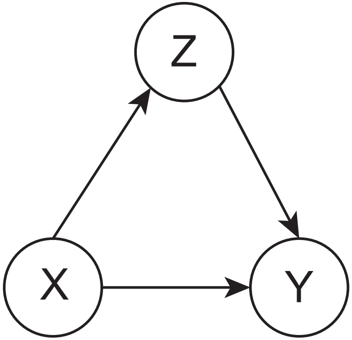
- $X \sim \mathcal{N}(0, 1)$
- $Z|X \sim \mathcal{N}(2X, 1)$
- $Y|Z, X \sim \mathcal{N}(1X + 2Z, 1)$



Example: SCM with normal noise terms and linear conditional relationships

Conditional distributions:

- $X \sim \mathcal{N}(0, 1)$
- $Z|X \sim \mathcal{N}(2X, 1)$
- $Y|Z, X \sim \mathcal{N}(1X + 2Z, 1)$



Markov factorization:
$P(Y, X, Z) = P(Y|Z, X)P(Z|X)P(X)$
This results in the following multivariate normal:

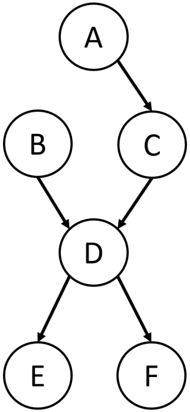
$$\begin{pmatrix} X \\ Z \\ Y \end{pmatrix} = \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 5 \\ 2 & 5 & 12 \\ 5 & 12 & 30 \end{pmatrix}\right]$$

Overview

Now, we are going to use DAGs to more intuitively read of statistical dependencies that are implied by causal structures.

- Causal inference - intro
- Causal Graphs, DAGs and SCMs
- Statistical dependencies implied by DAG structures
- Causal Discovery

Paths between variables in Graphs



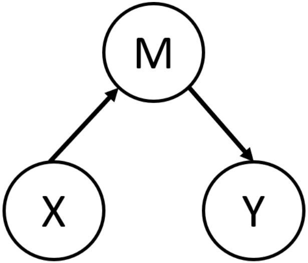
Arrows (any direction) connecting two variables?
$\rightarrow$ path between those variables.

"Open paths": Imply statistical dependency between the variables.
"Blocked paths": Imply statistical independency between the variables.
No paths: Implies statistical independency between the variables.

Three Types of Paths in DAGs: 1) Chains

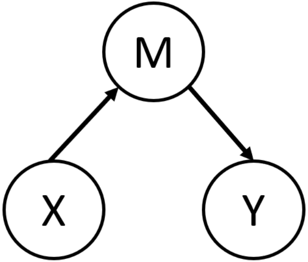
The Chain, aka Mediation.

M is the "mediator".



Three Types of Paths in DAGs: 1) Chains

The Chain, aka Mediation.
M is the "mediator".

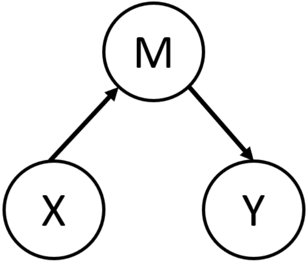


M transmits a causal association of X on Y.

- This results in a marginal association between X and Y: $X \not\perp\!\!\!\perp Y$
- Hence, a chain is an open path (between X and Y).

Three Types of Paths in DAGs: 1) Chains

The Chain, aka Mediation.
M is the "mediator".



M transmits a causal association of X on Y.

- This results in a marginal association between X and Y: $X \not\perp\!\!\!\perp Y$
- Hence, a chain is an open path (between X and Y).

Controlling for M blocks transmission of the causal effect:

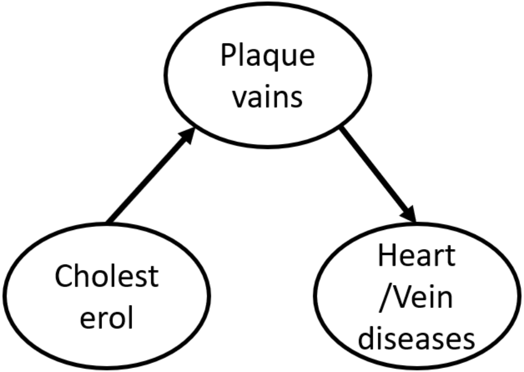
- By conditioning on M, X and Y become *in*dependent: $X \perp\!\!\!\perp Y \mid M$
- We can block the open path between X and Y by conditioning on M.

3 Type of Paths in DAGs: 1) Chains - Example

- Cholesterol $\not\perp\!\!\!\perp$ Heart-vein Diseases
- Cholesterol and Heart-vein Diseases are marginally dependent

- Cholesterol$\perp\!\!\!\perp$Heart-vein Diseases | Plaque
- Cholesterol and Heart-vein Diseases are *in*dependent conditional on the amount of plaque in veins

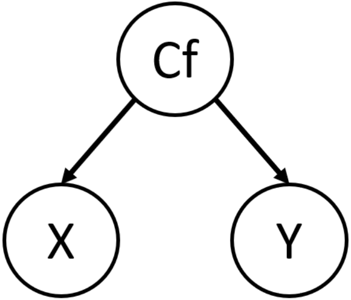
The Chain, aka Mediation.
Plaque is the "mediator".



Three Types of Paths in DAGs: 2) Forks

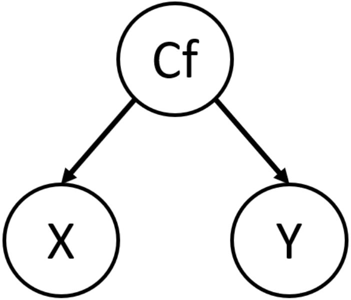
The Fork, aka Confounding.

Cf is the "confounder".



Three Types of Paths in DAGs: 2) Forks

The Fork, aka Confounding.
Cf is the "confounder".

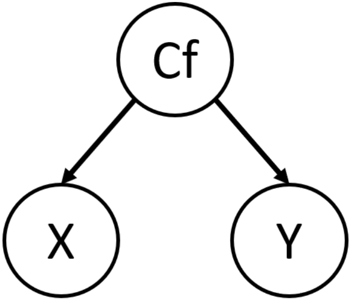


Cf transmits NON-causal/spurious association between X and Y:

- This results in a marginal association between X and Y: $X \not\perp\!\!\!\perp Y$
- Hence, a fork is an open path (between X and Y).

Three Types of Paths in DAGs: 2) Forks

The Fork, aka Confounding.
Cf is the "confounder".



Cf transmits NON-causal/spurious association between X and Y:

- This results in a marginal association between X and Y: $X \not\perp\!\!\!\perp Y$
- Hence, a fork is an open path (between X and Y).

Controlling for Cf blocks transmission of the spurious effect:

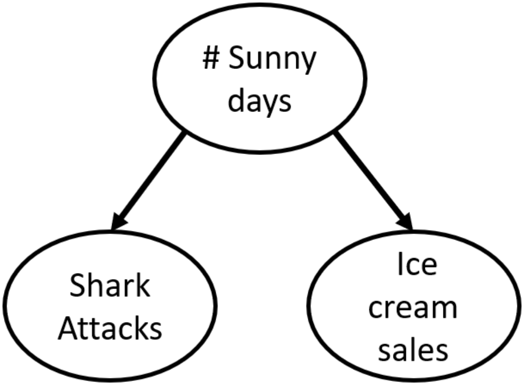
- By conditioning on Cf, X and Y become *in*dependent: $X \perp\!\!\!\perp Y \mid Cf$
- We can block the open path between X and Y by conditioning on Cf.

3 Type of Paths in DAGs: 2) Forks - Example

The Fork, aka Confounding. Number of Sunny Days is the "confounder".

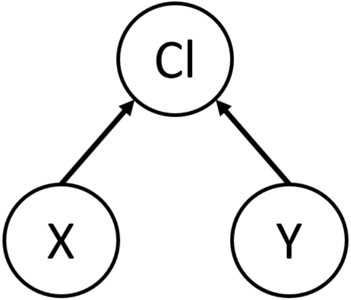
- Shark Attacks $\not\perp\!\!\!\perp$ Ice Cream Sales
- Shark Attacks and Ice Cream Sales are marginally dependent

- Shark Attacks $\perp\!\!\!\perp$ Ice Cream Sales | Sunny Days
- The amount of shark attacks and ice cream sales are *in*dependent conditional on the number of sunny days



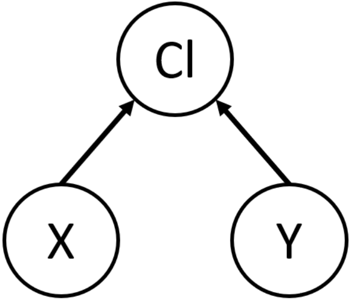
The Inverted Fork, aka Collider structure.

CI is the "collider".



Three Types of Paths in DAGs: 3) Inverted Forks

The Inverted Fork, aka Collider structure.
CI is the "collider".



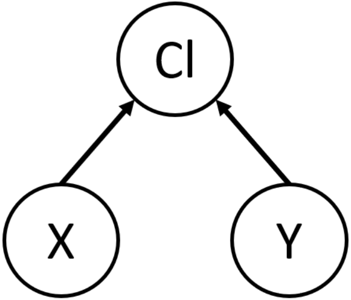
CI does NOT transmit association between X and Y:

- This results in marginal *in*dependence between X and Y: $X \perp\!\!\!\perp Y$
- The collider blocks the path between X and Y.

Three Types of Paths in DAGs: 3) Inverted Forks

The Inverted Fork, aka Collider structure.

Cl is the "collider".



Cl does NOT transmit association between X and Y:

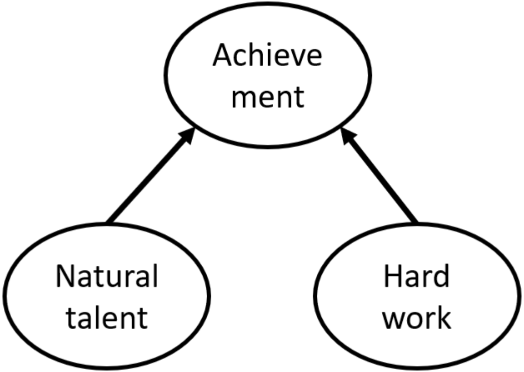
- This results in marginal *in*dependence between X and Y: $X \perp\!\!\!\perp Y$
- The collider blocks the path between X and Y.

Controlling for Cl transmiths a spurious association between X and Y:

- By conditioning on Cl, X and Y become dependent: $X \not\perp\!\!\!\perp Y \mid Cl$
- Conditioning on the collider opens a path between X and Y.

Three Types of Paths in DAGs: 3) Inverted Forks

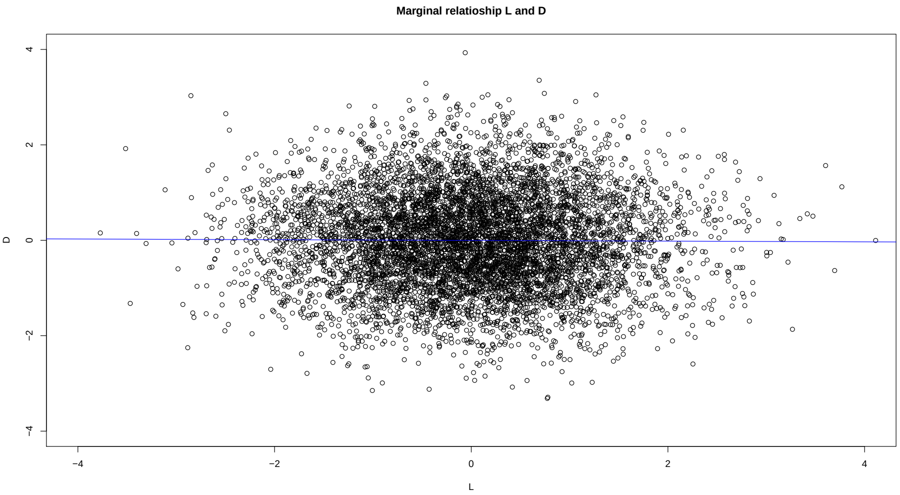
The Inverted Fork, aka Collider structure.
where Achievement is the "collider".



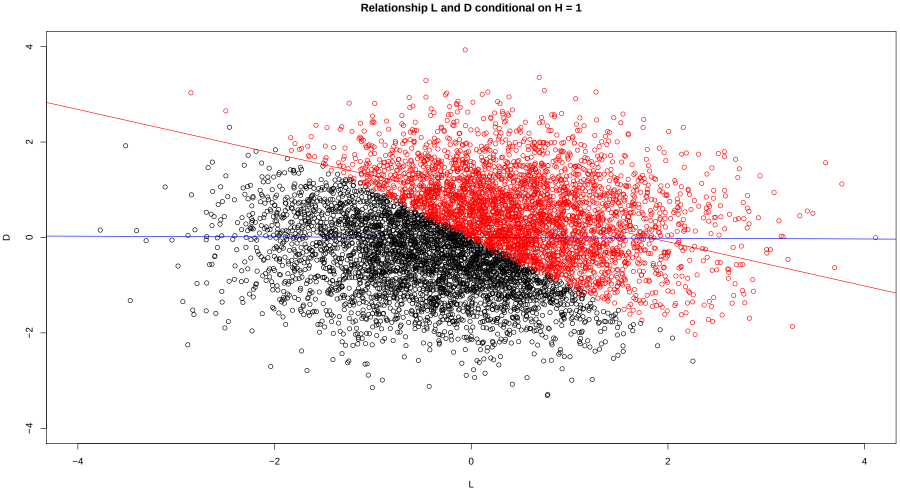
- Natural Talent $\perp\!\!\!\perp$ Working Hard
- Natural Talent and Working Hard are marginally *in*dependent

- Natural Talent $\not\!\perp\!\!\!\perp$ Working Hard | Achievement
- Natural Talent and Working Hard are dependent conditional on the level of Achievement.

Collider Bias: Example



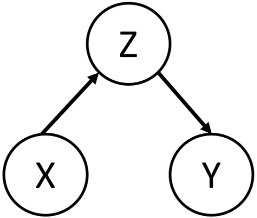
Collider Bias: Example



In red: only the high achievers.

Three types of paths

Chain

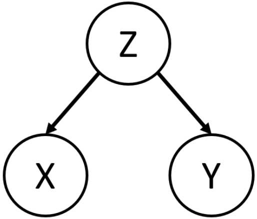


$X \not\perp\!\!\!\perp Y$

$X \perp\!\!\!\perp Y \mid Z$

X: cholesterol
Z: plaque
Y: heart-vein disease

Fork

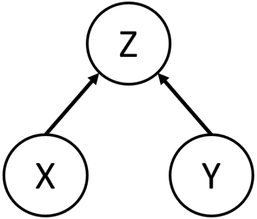


$X \not\perp\!\!\!\perp Y$

$X \perp\!\!\!\perp Y \mid Z$

X: shark attacks
Z: sunny days
Y: ice cream sales

Collider



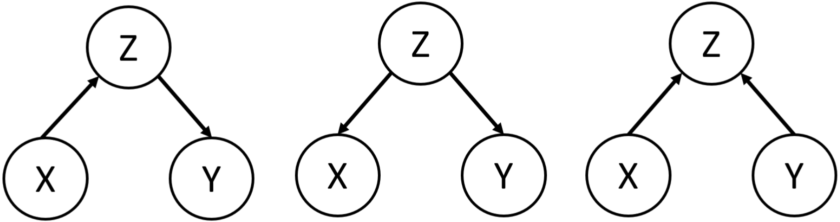
$X \perp\!\!\!\perp Y$

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

X: natural talent
Z: achievement
Y: working hard

When do causal (in)dependency and statistical (in)dependency align?

Three types of paths



Chain	Fork	Collider
$X \not\perp\!\!\!\perp Y$ ♥	$X \not\perp\!\!\!\perp Y$ ⚡	$X \perp\!\!\!\perp Y$ ♥
$X \perp\!\!\!\perp Y \mid Z$ ⚡	$X \perp\!\!\!\perp Y \mid Z$ ♥	$X \not\perp\!\!\!\perp Y \mid Z$ ⚡
X: cholesterol Z: plaque Y: heart-vein disease	X: shark attacks Z: sunny days Y: ice cream sales	X: natural talent Z: achievement Y: working hard

When do causal (in)dependency and statistical (in)dependency align?

D-Separation Rules

Conditional (in)dependence, also for larger graphs, can be read off using *d-seperation rules*

Open path between variables: Variables are ‘d-connected’.
No or Blocked path between variables: Variables are ‘d-separated’.

D-Separation Rules

Conditional (in)dependence, also for larger graphs, can be read off using *d-seperation rules*

Open path between variables: Variables are 'd-connected'.
No or Blocked path between variables: Variables are 'd-separated'.

Marginal (in)dependencies

- Chains and Forks are *open paths* $\rightarrow$ marginal dependence $X \not\perp\!\!\!\perp Y$
- Collider structures are *blocked paths*. $\rightarrow$ marginal independence $X \perp\!\!\!\perp Y$

D-Separation Rules

Conditional (in)dependence, also for larger graphs, can be read off using *d-seperation rules*

Open path between variables: Variables are ‘d-connected’.
No or Blocked path between variables: Variables are ‘d-separated’.

Marginal (in)dependencies

- Chains and Forks are *open paths* $\rightarrow$ marginal dependence $X \not\perp\!\!\!\perp Y$
- Collider structures are *blocked paths*. $\rightarrow$ marginal independence $X \perp\!\!\!\perp Y$

Conditional (in)dependencies

- Conditioning on Mediators and Confounders *or their descendants block a path* $\rightarrow$ conditional *in*dependence $X \perp\!\!\!\perp Y|Z$
- Conditioning on Colliders *or their descendants* open a path. $\rightarrow$ conditional dependence $X \not\perp\!\!\!\perp Y|Z$

d-seperation: Exercise

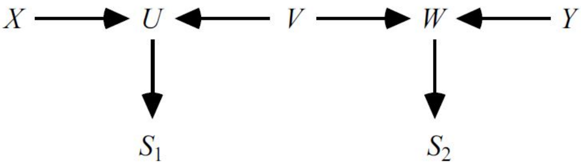


Figure 2.10

- Are X and Y marginally dependent?
- Given which sets of variables are X and Y conditionally dependent?

d-seperation Exercise: Are X and Y marginally dependent?

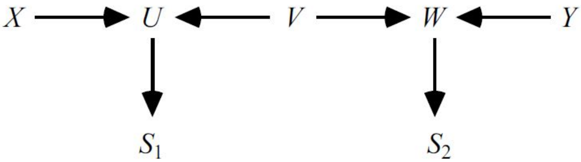


Figure 2.10

When not conditioning, both U and W block the path between X and Y. There are no open paths.
X and Y are d-seperated "given the empty set" (given no variables)

This implies:
$X \perp\!\!\!\perp Y$
X and Y are *marginally* independent.

d-seperation exercise: Given which sets of variables are X and Y conditionally dependent?

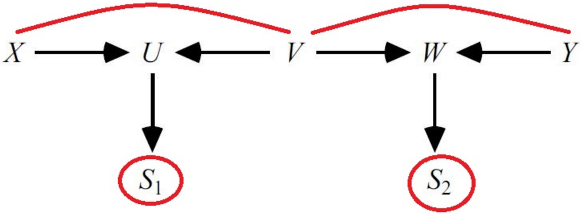


Figure 2.10

Conditioning on S_1 (see figure) or U creates an open path between X and V.
Conditioning on S_2 (see figure) or W creates an open path between V and Y.
So: X and Y are d-connected given a member of the set $\{U, S_1\}$ AND a member of the set $\{W, S_2\}$

For example: $X \not\perp\!\!\!\perp Y \mid \{S_1, S_2\}$; X and Y are dependent conditional on S_1 and S_2

d-seperation example: How can we close the path again?

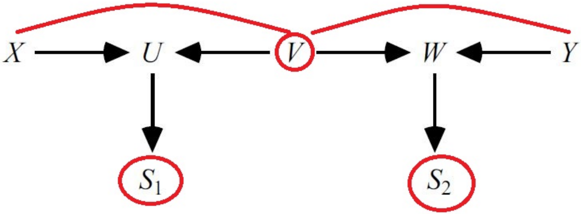


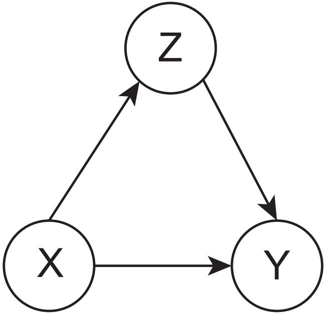
Figure 2.10

Conditioning on V in addition to S_1 or U and S_2 or W closes the path we opened by conditioning on colliders again.
X and Y are d-separated given a member of the set $\{U, S_1\}$ AND a member of the set $\{W, S_2\}$ AND V.

This for example implies: $X \perp\!\!\!\perp Y \mid \{S_1, S_2, V\}$
X and Y are independent conditional on S_1, S_2 and V

Using d-seperation to guide our causal analyses for observational data

I'm interested in the total causal effect of X on Y - both direct and indirect effects.
Interventions are not an option.
What variables should and shouldn't I control for to obtain the causal effect?



Using d-seperation to guide our causal analyses for observational data

I'm interested in the causal effect of X on Y. What variables should and shouldn't I control for?

- Block all *backdoor paths*: Paths from $X \to Y$ *that contain an arrow into* X (...$\to X$)
- Don't open up any new spurious paths - don't condition on colliders *or on their descendants*
- Leave all the directed paths you care about in-tact

Using d-seperation to guide our causal analyses for observational data

I'm interested in the causal effect of X on Y. What variables should and shouldn't I control for?

- Block all *backdoor paths*: Paths from $X \rightarrow Y$ *that contain an arrow into* X (...$\rightarrow X$)
- Don't open up any new spurious paths - don't condition on colliders *or on their descendants*
- Leave all the directed paths you care about in-tact

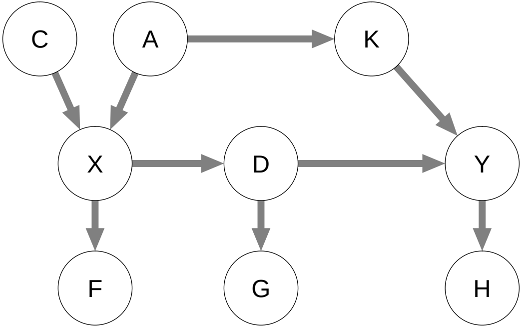
Backdoor criterion

If conditioning on a set of variables $\boldsymbol{Z}$ meets these goals, we say $\boldsymbol{Z}$ fulfills the 'backdoor criterion'. Adjusting for $\boldsymbol{Z}$ yields the causal effect of X on Y.

Valid Adjustment Set

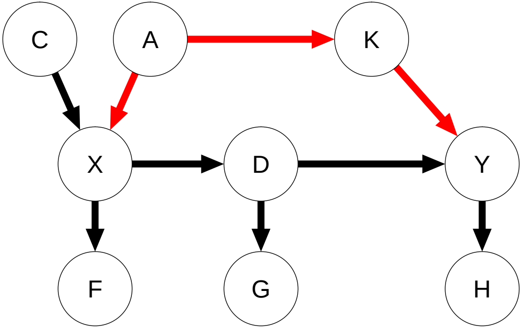
A set of variables $\boldsymbol{Z}$ that by conditioning on them allows us to correctly estimate the effect of X on Y , we call the 'Valid Adjustment Set' for the effect of X on Y.

Using d-seperation to guide our causal analyses for observational data



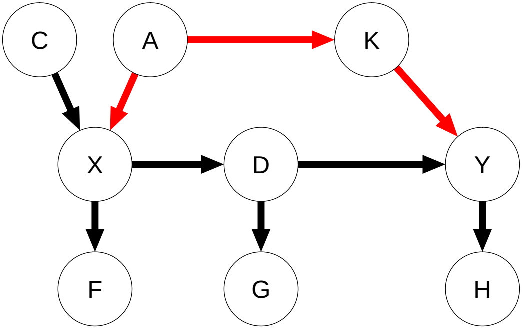
We want to estimate the causal effect of X on Y.

Using d-seperation to guide our causal analyses for observational data



$X \leftarrow A \rightarrow K \rightarrow Y$ is a "backdoor path" from X to Y

DAGs tell us how to estimate causal effects from observational data



Valid Adjustment Sets: $\{A\}$, $\{K\}$, $\{A, K\}$, $\{F, C, K\}$

Overview

But what if we do not know the DAG?!

- Causal inference - intro
- Causal Graphs, DAGs and SCMs
- Statistical dependencies implied by DAG structures
- **Causal Discovery**

Causal Discovery or Causal Learning

Can we infer the causal structure from (observational) data?

Short answer:

- In general, **no**
- There is usually more than one SCM and DAG that can generate the same dataset.

Long answer:

- Yes, or at least, we can learn something about the causal structure.
- But only if we are willing to make certain assumptions about the causal system.
- Using triangulation - using different methods with different (causal) assumptions, we may learn even more.

Causal Discovery using DAGS and d-separation rules

Basic Idea:

1. Find all marginal and conditional independence relations present in the data
2. Draw the DAG in which all (and only) those independencies hold up based on the d-seperation rules.

Causal Discovery using DAGS and d-separation rules

Basic Idea:

1. Find all marginal and conditional independence relations present in the data
2. Draw the DAG in which all (and only) those independencies hold up based on the d-seperation rules.

Typical Assumptions:

- The causal system of interest can be captured in a DAG.
- No *unobserved* common causes. (**Sufficiency**)
- No conditioning on *unobserved* colliders (no selection bias).
- Faithfulness (tbd later)
- various statistical assumptions for evaluating statistical dependencies

Example 1: CI-based discovery

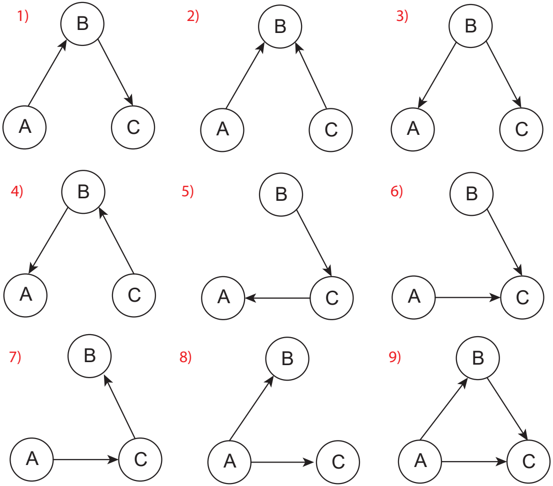
We have three variables in our dataset, A, B and C.

We know of the following (in)dependencies:

$$A \perp\!\!\!\perp C$$
$$A \not\perp\!\!\!\perp C \mid B$$

All other combinations of variables are dependent (e.g. $A \not\perp\!\!\!\perp B$ and $B \not\perp\!\!\!\perp C \mid A$)

What is the true (data-generating) DAG?



Example 1: CI-based discovery

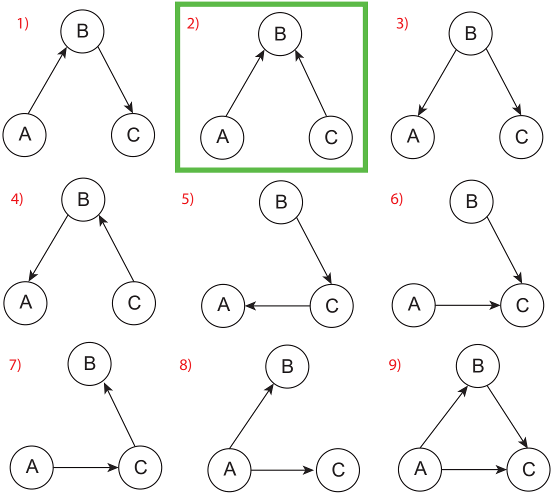
We have three variables in our dataset, A, B and C.

We know of the following (in)dependencies:

$$A \perp\!\!\!\perp C$$
$$A \not\!\perp\!\!\!\perp C \mid B$$

All other combinations of variables are dependent (e.g. $A \not\!\perp\!\!\!\perp B$ and $B \not\!\perp\!\!\!\perp C \mid A$)

What is the true (data-generating) DAG?



Causal Discovery with Conditional Independencies and DAGs

Draw undirected edges between variables you are sure should be there.

Principle 1

Two variables A and B are directly connected in the DAG (either A$\rightarrow$B OR B$\rightarrow$A) if, and only if, they are dependent conditional on every possible subset of the other variables

This includes marginal relationships, that is, when you condition on no other variables.

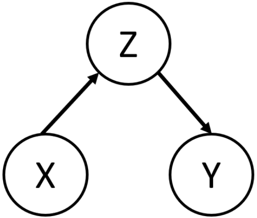
Get all (in)dependencies and only directly link those variables that are always dependent: the result is the *skeleton* of the DAG.

Causal Discovery with Conditional Indepencies and DAGs

Infer the direction of as many of the undirected edges as possible.

Reminder Intermezzo: Types of Paths

Chain

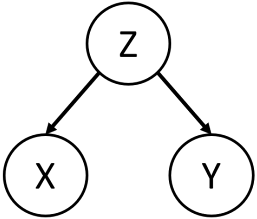


$X \not\!\perp\!\!\!\perp Y$

$X \perp\!\!\!\perp Y \mid Z$

X: cholesterol
Z: plaque
Y: heart-vein disease

Fork

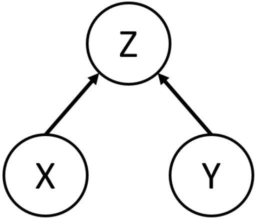


$X \not\!\perp\!\!\!\perp Y$

$X \perp\!\!\!\perp Y \mid Z$

X: shark attacks
Z: sunny days
Y: ice cream sales

Collider



$X \perp\!\!\!\perp Y$

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

X: natural talent
Z: achievement
Y: working hard

Which of these causal structures can be statistically distinguished from each other?

Types of Paths

Which of these causal structures can and cannot be statistically distinguished from each other?

Which have different conditional (in)dependence relations?

- Chains and forks are statistically equivalent
- This applies to chains in either direction: $X \rightarrow Z \rightarrow Y$ is equivalent to $X \leftarrow Z \leftarrow Y$.
- Colliders are distinct from chains and forks.

Causal Discovery with Conditional Indepencies and DAGs

Infer the direction of as many of the undirected edges as possible.

Principle 2

If our skeleton contains a triplet X - Z - Y, where X and Y are marginally independent, we can orientate the arrows as $X \rightarrow Z \leftarrow Y$ if and only if X and Y are dependent conditional on every set of variables containing Z.

This works for 'immoral' colliders, where parents have a common child but are 'unmarried' - don't directly cause each other.

So, find variables you are sure have to be colliders, and provide the relevant directed arrows.

Example 2: CI-based discovery

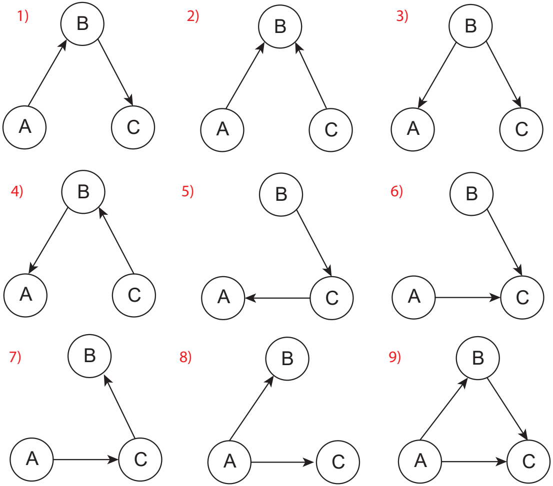
We have three variables in our dataset, A, B and C.

We know of the following (in)dependencies:

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

All other combinations of variables are dependent

What is the data-generating DAG?



Example 2: CI-based discovery

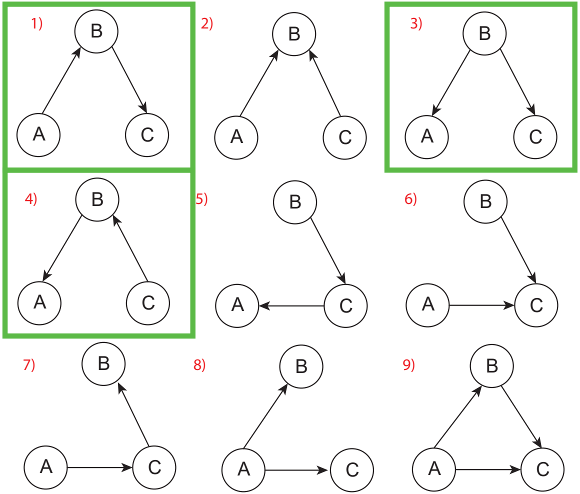
We have three variables in our dataset, A, B and C.

We know of the following (in)dependencies:

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

All other combinations of variables are dependent

What is the data-generating DAG?



Markov Equivalence

Typically CI-based methods find more than one DAGs that match the observational data. The set of possible DAGs is called the *Markov Equivalence set*

Markov Equivalence:

Two DAGs are *Markov Equivalent* if they satisfy the same d-separation statements, that is, the same set of (conditional) (in)dependence relations.

Markov Equivalence

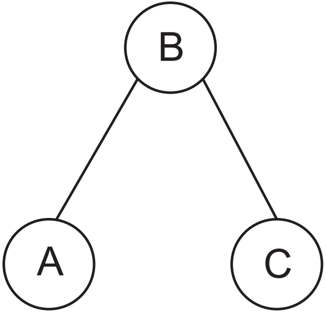
Typically CI-based methods find more than one DAGs that match the observational data. The set of possible DAGs is called the *Markov Equivalence set*

Markov Equivalence:

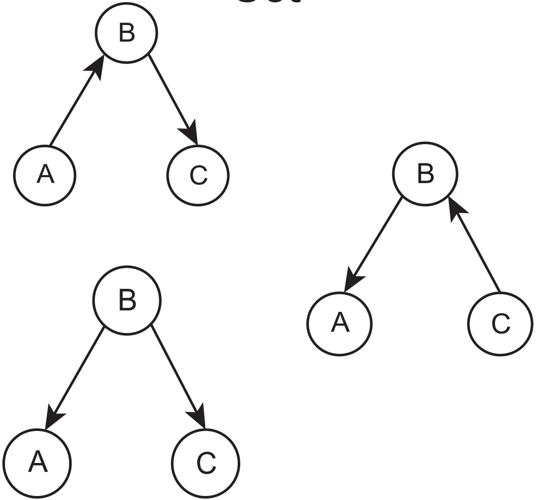
Two DAGs are *Markov Equivalent* if they satisfy the same d-separation statements, that is, the same set of (conditional) (in)dependence relations.

The Markov-Equivalance set is sometimes simply represented by a ‘Complete Partially-Oriented Directed Acyclic Graph’ **(CPDAG)**

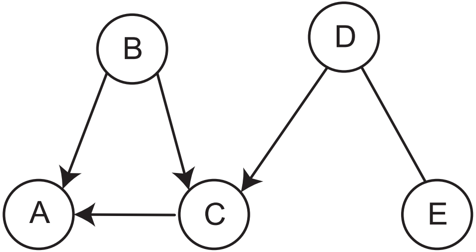
CPDAG



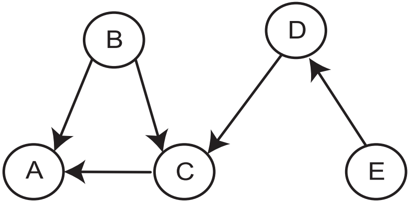
Marov-Equivalence Set

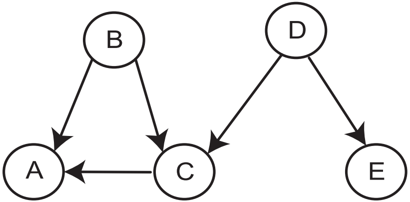


CPDAG



Marov-Equivalence Set





In Practice

PC algorithm; FCI algorithm (Spirtes et al. 2000)

- Do a quicker search without having to test all (in)dependencies
- Various extensions exist that deal with violations of *sufficiency*.
- Still rely on faithfulness.

Important Assumptions for CI-based discovery

For (sets of) variables X, Y, and Z:

Global Markov Condition:

If X and Y are d-separated by Z then $X \perp\!\!\!\perp Y \mid Z$

"The d-separation rules are appropriate" (which is the case if the causal structure is a DAG).

Important Assumptions for CI-based discovery

For (sets of) variables X, Y, and Z:

Global Markov Condition:

If X and Y are d-separated by Z then $X \perp\!\!\!\perp Y \mid Z$

"The d-separation rules are appropriate" (which is the case if the causal structure is a DAG).

Faithfulness:

If $X \perp\!\!\!\perp Y \mid Z$ then X and Y are d-separated by Z

"The other way around works as well" - not always true even if the structure is a DAG.

Important Assumptions for CI-based discovery

For (sets of) variables X, Y, and Z:

Global Markov Condition:

If X and Y are d-separated by Z then $X \perp\!\!\!\perp Y \mid Z$

"The d-separation rules are appropriate" (which is the case if the causal structure is a DAG).

Faithfulness:

If $X \perp\!\!\!\perp Y \mid Z$ then X and Y are d-separated by Z

"The other way around works as well" - not always true even if the structure is a DAG.

Also: Various statistical assumptions, e.g., those we need to estimate and decide whether two variables can be considered (in)dependent.

Violations of Faithfulness

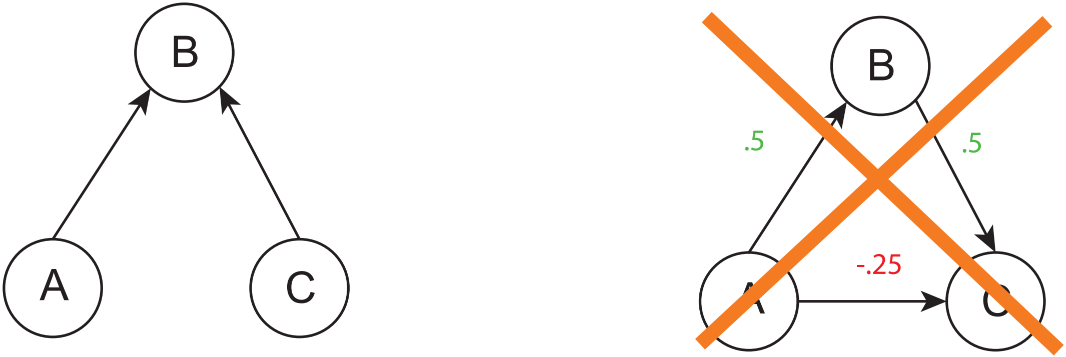
$$
\begin{aligned}
A &:= \epsilon_A \\
B &:= .5A + \epsilon_B \\
C &:= -.25A + .5B + \epsilon_C
\end{aligned}
$$

where

- $\epsilon_A, \epsilon_B, \epsilon_C$ are iid, $\sim \mathcal{N}(0,1)$

$$
\begin{pmatrix} A \\ B \\ C \end{pmatrix} = \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & .5 & 0 \\ .5 & 1.25 & .5 \\ 0 & .5 & 1.25 \end{pmatrix}\right]
$$

Violations of Faithfulness



$$A \perp\!\!\!\perp C$$
$$A \not\!\perp\!\!\!\perp C \mid B$$

Assume Faithfulness

Assumptions for CI-based discovery

Global Markov Condition:

X and Y are d-seperated by $Z \implies \text{X} \perp\!\!\!\perp Y \mid Z$

Faithfulness:

$X \perp\!\!\!\perp Y \mid Z \implies X$ and Y are d-seperated by Z

Essentially: Paths never "perfectly cancel out"
Statistical (conditional) Independence $\implies$ causal independence (d-seperation)

In Practice

PC algorithm; FCI algorithm (Spirtes et al. 2000)

- Do a quicker search without having to test all (in)dependencies
- Various extensions exist that deal with violations of *sufficiency*.
- Still rely on faithfulness.

Disadvantages:

- Population (in)dependencies are estimated: uses sample data + statistical tests
  - All of statistics is relevant here, e.g., sample size considerations
  - In a given sample : Type I and II errors
  - Faithfulness does NOT mean there are no false negatives!

In Practice

PC algorithm; FCI algorithm (Spirtes et al. 2000)

- Do a quicker search without having to test all (in)dependencies
- Various extensions exist that deal with violations of *sufficiency*.
- Still rely on faithfulness.

Disadvantages:

- Population (in)dependencies are estimated: uses sample data + statistical tests
  - All of statistics is relevant here, e.g., sample size considerations
  - In a given sample : Type I and II errors
  - Faithfulness does NOT mean there are no false negatives!
- CI testing easy if linear + Normal (partial correlation / regression) or discrete
  - Can be difficult in other cases (Shah & Peters, 2020)
  - Non-parametric methods - difficult and requires large sample size

In Conclusion...

DAG/SCM causal modeling allows us *in theory* to make causal statements from observational data

- Rests on our beliefs/assumptions the causal structure (e.g., the DAG being correct).
- And as soon as we estimate things (including DAGs), many more (statistical) assumptions pop up.
- Often we may not be able to evaluate these assumptions.
- BUT... by being explicit about our assumptions, we can openly debate and learn about them, and about our causal inferences of interest.

In Conclusion...

DAG/SCM causal modeling allows us *in theory* to make causal statements from observational data

- Rests on our beliefs/assumptions the causal structure (e.g., the DAG being correct).
- And as soon as we estimate things (including DAGs), many more (statistical) assumptions pop up.
- Often we may not be able to evaluate these assumptions.
- BUT... by being explicit about our assumptions, we can openly debate and learn about them, and about our causal inferences of interest.

- Next Week: Practicing all of the above in an R lab.
- Next Week: More info on the assignment(s).
- Next Lecture: Rubin's causal inference framework - very explicit about assumptions needed for estimating causal effects - and based on ideas related to missing data.