d-separation

The three rules from last lecture (adding 'simple'; relabelling):

Rule 1: conditional independence in chains

If the only simple path between two variables X and Y is directed, and $\mathbf{Z}$ is a set of variables at least one of which is on that path, then X and Y are conditionally independent given $\mathbf{Z}$

Rule 2: conditional independence in forks

If the only simple path between two variables X and Y passes through a common cause Z, then X and Y are conditionally independent given Z

Rule 3: conditional independence in colliders

If the only simple path between two variables X and Y passes through a collider Z, then X and Y are marginally (i.e. unconditionally) independent, but likely dependent given Z, or given any descendant of Z

d-separation: simple paths

General case of a simple path between X and Y: it is open (i.e. gives a likely dependence between X and Y) unless at some node on the path between X and Y it is blocked

A non-collider ($\leftarrow\rightarrow$, $\leftarrow\leftarrow$, or $\rightarrow\rightarrow$) blocks the path iff:

- it is conditioned on

A collider ($\rightarrow\leftarrow$) blocks the path iff:

- it is *not* conditioned on, and none of its descendants are conditioned on

d-separation: multiple paths

Now for the general case:

Intuition: if *any* path between X and Y is open (= not blocked), X and Y are likely dependent

Definition: d-separation

X and Y are d-separated given (set of) variable(s) Z if conditioning on ZS blocks *every* path between X and Y

d-separation: multiple paths

Now for the general case:

Intuition: if *any* path between X and Y is open ($=$ not blocked), X and Y are likely dependent

Definition: d-separation

X and Y are d-separated given (set of) variable(s) Z if conditioning on ZS blocks *every* path between X and Y

Trick: instead of remembering the exception about descendants of colliders, we can look at all paths (not just simple paths). Follow the path to a collider, down to a descendant that's conditioned on, and back up to the collider: this path is open!

d-separation and conditional independence

If X and Y are d-separated given Z in a structural causal model, then in the distribution of data coming from that model, X and Y are conditionally independent given Z

Otherwise, they are likely independent

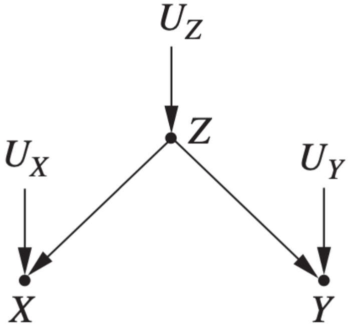
Formalizing interventions

Last week, we defined the terms ‘cause’ and ‘direct cause’ for SCMs. Today we will justify those terms.

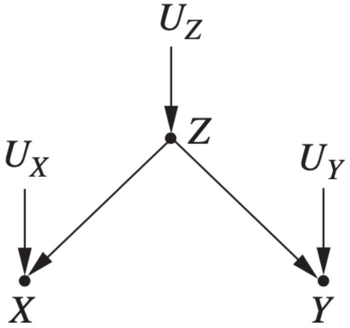
We distinguish:

- $P(Y \mid X = x)$
  - the *conditional distribution* of Y given $X = x$
  - obtained by ‘filtering’ the population, but the system remains the same
- $P(Y \mid \mathrm{do}(X = x))$
  - the distribution of Y after an intervention that sets X to x
  - obtained by changing the system

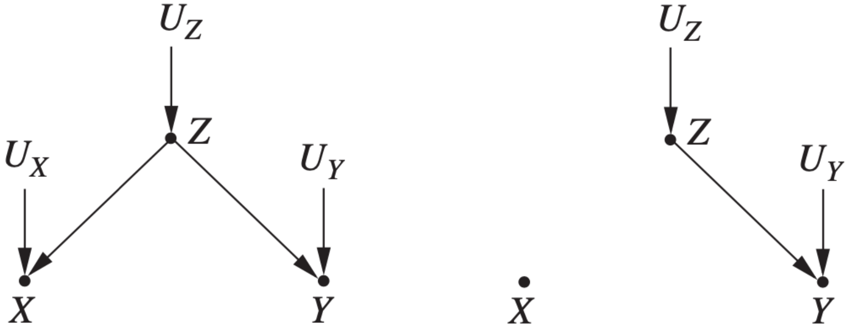
Interventions graphically



Interventions graphically



Interventions graphically



General rule: an intervention on X removes all arrows pointing at X

- So unlike conditioning, what happens with an intervention depends on the graph!

Perfect interventions

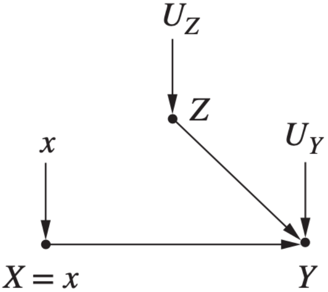
What we look at here is called a perfect intervention:

- It affects only the intended variable
- It sets that variable's value exactly to the value we want

This is not always realistic, and many extensions have also been studied. But all have in common that the result depends on the causal graph

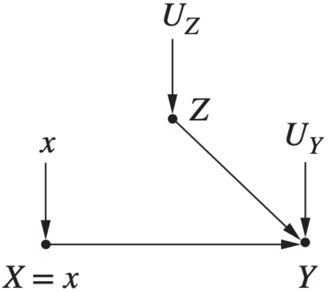
Randomized controlled trials

In a randomized controlled trial, patients are assigned to a treatment group X (e.g. treatment vs. placebo) at random, independently of anything that might affect the outcome Y (e.g. gender Z). So X might be set by a coin flip.



Randomized controlled trials

In a randomized controlled trial, patients are assigned to a treatment group X (e.g. treatment vs. placebo) at random, independently of anything that might affect the outcome Y (e.g. gender Z). So X might be set by a coin flip.



- Then intervening on X doesn't remove any edges, so that $P(Y \mid X = x) = P(Y \mid \text{do}(X = x))$

Randomized controlled trials

RCTs are the golden standard for measuring the causal effect $P(Y \mid \mathrm{do}(X = x))$: can just measure it directly from the data

- Same idea by another name: *A/B testing*

Randomized controlled trials

RCTs are the golden standard for measuring the causal effect $P(Y \mid \mathrm{do}(X = x))$: can just measure it directly from the data

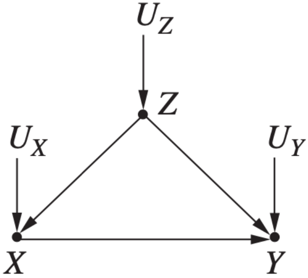
- Same idea by another name: *A/B testing*

In case of binary X, the quantity we want to know may not be the causal effect $P(Y \mid \mathrm{do}(X = x))$, but the causal effect difference (more commonly known as the average causal effect:

$$P(Y \mid \mathrm{do}(X = 1)) - P(Y \mid \mathrm{do}(X = 0))$$

Revisiting Simpson's paradox

In the Simpson's paradox case from the first lecture, X was not chosen independently at random:



Problem: we can't go back and change how the data was gathered, but have to work with the data that is there. Can we still find $P(Y \mid \mathrm{do}(X = x))$?

Relations between P and P_m

Write P_m for the distribution of the distribution after $\mathrm{do}(X = x)$

Under a perfect intervention,

- the structural equations for Y and Z don't change, and
- the distributions of U_Y and U_Z don't change

So,

$$\begin{aligned} P_m(Z = z) &= P(Z = z) \\ P_m(Y = y \mid Z = z, X = x) &= P(Y = y \mid Z = z, X = x) \end{aligned}$$

An expression for the causal effect

$$
\begin{aligned}
&P(Y = y \mid \mathrm{do}(X = x)) \\
&= P_m(Y = y \mid X = x) && \text{by definition} \\
&= \sum_z P_m(Y = y \mid X = x, Z = z) P_m(Z = z \mid X = x) \\
&= \sum_z P_m(Y = y \mid X = x, Z = z) P_m(Z = z) && \text{by } * \\
&= \sum_z P(Y = y \mid X = x, Z = z) P(Z = z)
\end{aligned}
$$

($*$: by independence of X and Z in P_m, because they are d-separated (given the empty set) in the intervened graph)

An expression for the causal effect

$$
\begin{aligned}
& P(Y = y \mid \mathrm{do}(X = x)) \\
& = P_m(Y = y \mid X = x) && \text{by definition} \\
& = \sum_z P_m(Y = y \mid X = x, Z = z) P_m(Z = z \mid X = x) \\
& = \sum_z P_m(Y = y \mid X = x, Z = z) P_m(Z = z) && \text{by } * \\
& = \sum_z P(Y = y \mid X = x, Z = z) P(Z = z)
\end{aligned}
$$

($*$: by independence of X and Z in P_m, because they are d-separated (given the empty set) in the intervened graph)

We have expressed the causal effect in terms of the original distribution P: we can compute it without actually performing the intervention!

Adjustment formula

Conclusion: In the graph from Simpson's paradox,

$$P(Y = y \mid \mathrm{do}(X = x)) = \sum_z P(Y = y \mid X = x, Z = z) P(Z = z)$$

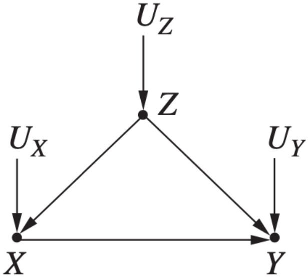
This is called the adjustment formula: it 'adjusts' for Z

So: to determine the causal effect of X on Y, we need to look (in the binary case) at $P(Y = y \mid X = x, Z = 0)$ and $P(Y = y \mid X = x, Z = 1)$, and take an average weighted by $P(Z)$.

This confirms what we saw informally in the first lecture: we need to look at the *segragated* data

Simpson's paradox: blood pressure case

Now let's look at the version of the paradox where gender was replaced by (post-treatment) blood pressure:



Causal direction between X and Z reversed! Different conclusion:

$$P(Y = y \mid \mathrm{do}(X = x)) = P(Y = y \mid X = x)$$

So here we should look at the aggregated data

Causal effect rule

We can compute $P(Y = y \mid \mathrm{do}(X = x))$ by adjusting for the parents of X

Then the derivation above still works (let's check!)

Truncated product rule

Recall the 'product decomposition rule' (for paired exogenous and endogenous variables; let $\mathrm{pa}(x_i)$ be just the endogenous parents now):

$$P(x_1, \ldots, x_n) = \prod_i P(x_i \mid \mathrm{pa}(x_i))$$

Truncated product rule

Recall the 'product decomposition rule' (for paired exogenous and endogenous variables; let $\mathrm{pa}(x_i)$ be just the endogenous parents now):

$$P(x_1, \ldots, x_n) = \prod_i P(x_i \mid \mathrm{pa}(x_i))$$

An intervention $\mathrm{do}(X_j = x_j)$ removes just the factor for x_j:

$$P(x_1, \ldots, x_n) = \prod_{i \neq j} P(x_i \mid \mathrm{pa}(x_i) \setminus u_i)$$

(or rather: the factor is replaced by the constant 1)

Backdoor paths

What if we can't adjust for all parents of X (e.g. because some weren't measured)?

Backdoor path

A simple path from X to Y that starts with $X \leftarrow$ is called a backdoor path

Backdoor criterion

For a graph with $X \rightarrow Y$, a set Z satisfies the backdoor criterion if

- no node in Z is a descendant of X, and
- Z blocks all backdoor paths from X to Y

If we have such a set Z of measured variables, we can use this Z in the adjustment formulae

Do-calculus

Many sets that are valid adjustment sets don't satisfy the backdoor criterion (and the same is true for the more complicated frontdoor criterion that's also in the book)
The do-calculus is a set of three rules for working with formulas involving do($\cdot$)'s

- Shown to be complete: if you can't find compute it with those three rules, you can't compute it! (unless you make additional assumptions)
- Can be done automatically by the ID algorithm: it either returns an expression for the causal effect, or reports that it can't be done