Causal Directed Acylic Graphs

introduction

Wouter van Amsterdam

2024-08-06

Day 2 intro: Causal Directed Acyclic Graphs and Structural Causal Models

Today's lectures

- introduce 1.5 new framework based on
  - causal Directed Acyclic Graphs (DAGs)
  - Structral Causal Models (SCMs)
- counterfactuals and Pearl's Causal Hierarchy of questions
- lectures will follow Pearl's book Causality Pearl (2009), specifically chapters 3 (DAGs) and 7 (SCMs)

Causal inference frameworks

What are they for?

Mathematical language to

- define *causal* quantities
- express *assumptions*
- derive how to *estimate* causal quantities

Causal inference frameworks

Why learn more than one?

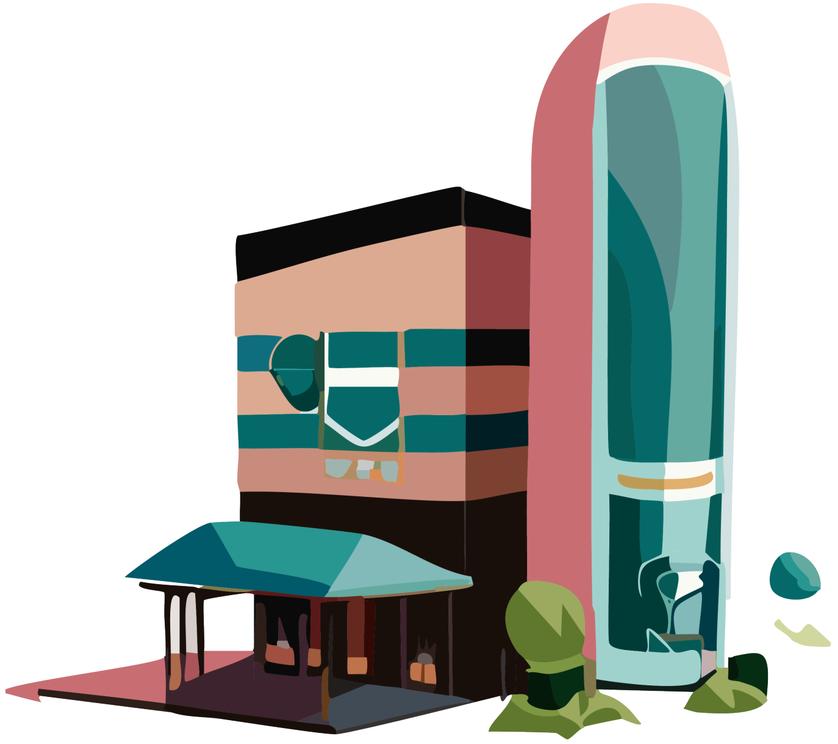
- On day 1 we learned about the Potential Outcomes framework
  - Defines causal effects in terms of (averages of) *individual potential outcomes*
  - Estimation requires assumptions of (conditional) exchangeability and positivity / overlap and consistency
- There isn't only 1 way to think about causality, find one that '*clicks*'
- Now we will learn another framework: *Structural Causal Models* and *causal graphs*
  - causal relations and manipulations of *variables*
  - Developed by different people initially - Judea Pearl, Peter Spirtes, Clark Glymour
  - SCM approach is broader in that it can define more different types of causal questions
- Equivalence: given the same data and assumptions, get the same estimates

Lecture 1 & 2 topics

- motivating examples for DAGs
- what are DAGs
- causal inference with DAGs
  - what is an intervention
  - DAG-structures: confounding, mediation, colliders
  - d-separation
  - back-door criterion

Motivating examples

Example task: are hospital deliveries good for babies?



Example task: are hospital deliveries good for babies?

- You’re a data scientist in a children’s hospital
- Have data on
  - delivery location (home or hospital)
  - neonatal outcomes (good or bad)
  - pregnancy risk (high or low)
- Question: do hospital deliveries result in better outcomes for babies?

Observed data

percentage of good neonatal outcomes

		location	
		home	hospital
risk	low	648 / 720 = 90%	19 / 20 = 95%
	high	40 / 80 = 50%	144 / 180 = 80%

- better outcomes for babies delivered in the hospital for *both risk groups*

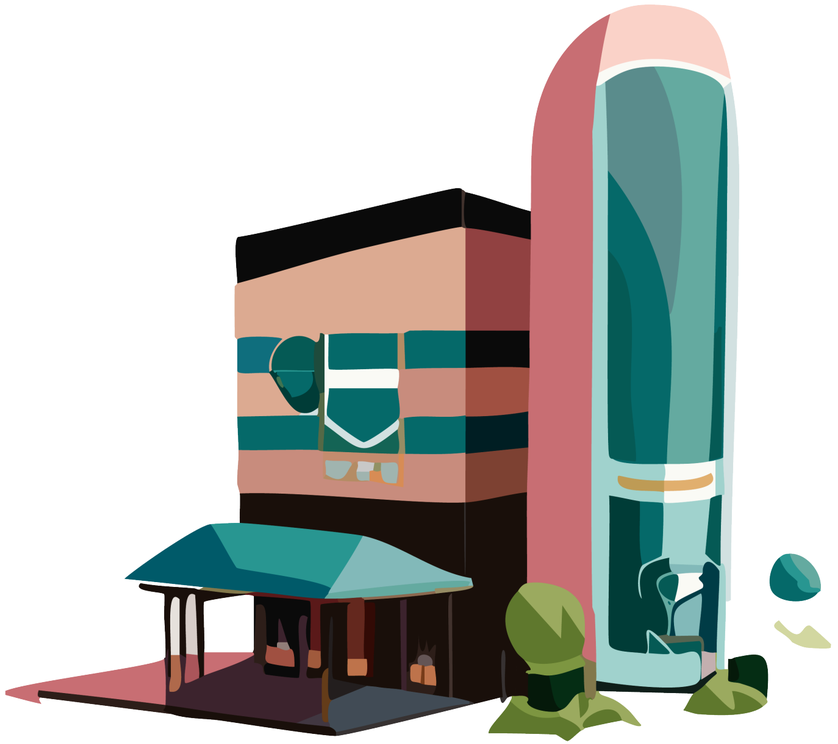
Observed data

		location	
		home	hospital
risk	low	648 / 720 = 90%	19 / 20 = 95%
	high	40 / 80 = 50%	144 / 180 = 80%
	marginal	688 / 800 = 86%	163 / 200 = 81.5%

- better outcomes for babies delivered in the hospital for *both risk groups*
- but not better *marginal* (‘overall’)
- how is this possible? (a.k.a. *simpsons paradox*)
- what is the correct way to estimate the effect of delivery location?

New question: hernia

- for a patient with a hernia, will they be able to walk sooner when recovering at home or when recovering in a hospital?
- observed data: location, recovery, bed-rest



Observed data 2

		location	
		home	hospital
bedrest	no	648 / 720 = 90%	19 / 20 = 95%
	yes	40 / 80 = 50%	144 / 180 = 80%
	marginal	688 / 800 = 86%	163 / 200 = 81.5%

- more bed rest in hospital
- what is the correct way to estimate the effect of location?

How to unravel this?

- we got two questions with exactly the same data
- in one example, ‘stratified analysis’ seemed best
- in the other example, ‘marginal analysis’ seemed best
- with *Directed Acyclic Graphs* we can make our decision

Causal Directed Acyclic Graphs

diagram that represents our assumptions on causal relations

1. nodes are variables
2. arrows (directed edges) point from cause to effect

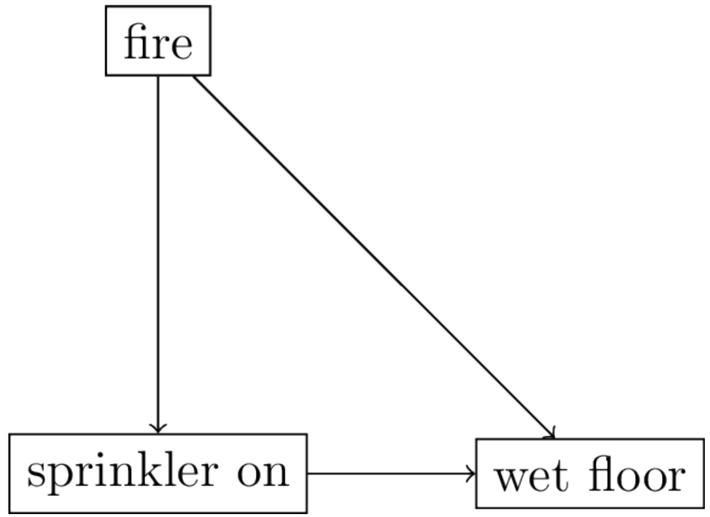
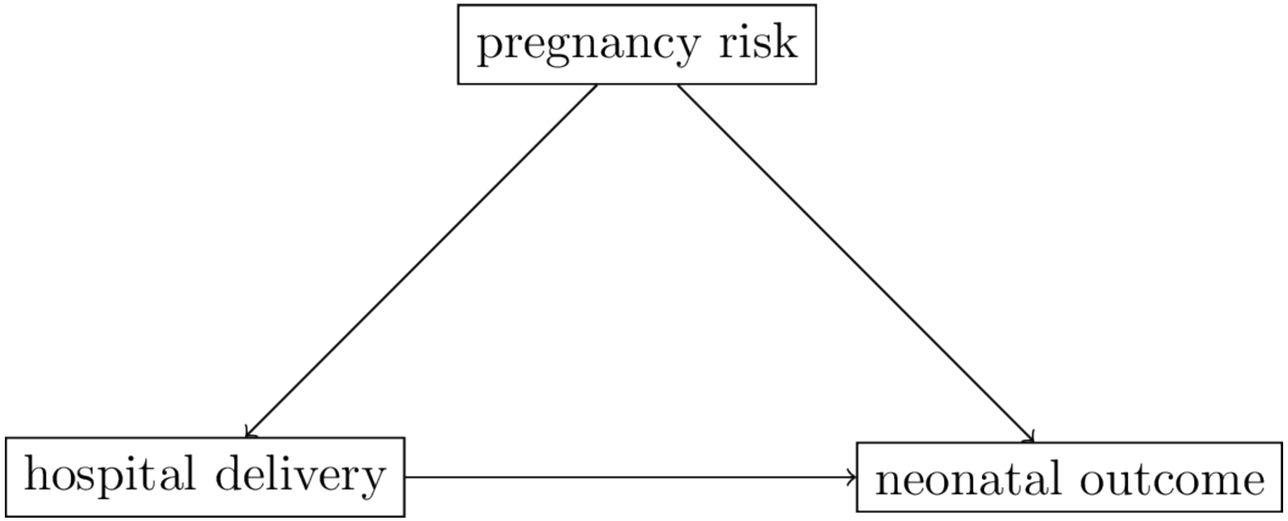


Figure 1: Directed Acyclic Graph

- when used to convey causal assumptions, DAGs are ‘causal’ DAGs
- this is not the only use of DAGs (see day 4)

Making DAGs for our examples:

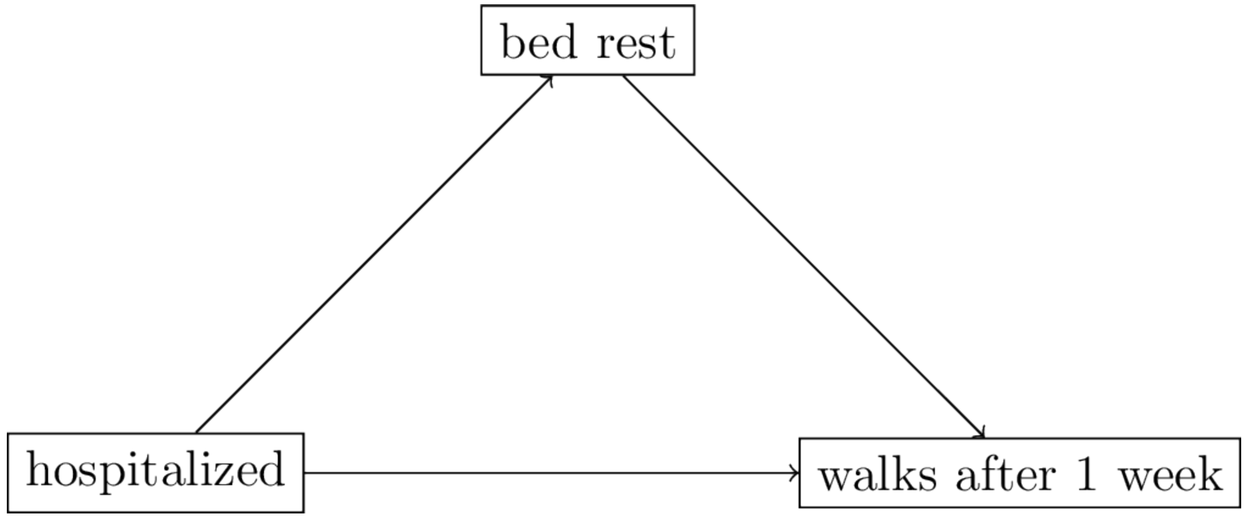
The pregnancy DAG



- assumptions:
  - women with high risk of bad neonatal outcomes (`pregnancy risk`) are referred to the hospital for delivery
  - hospital deliveries lead to better outcomes for babies as more emergency treatments possible
  - both `pregnancy risk` and `hospital delivery` cause `neonatal outcome`
- the *other variable* `pregnancy risk` is a common cause of the treatment (`hospital delivery`) and the outcome (this is what's called a confounder)

Making DAGs for our examples:

The hernia DAG



- assumptions:
  - patients admitted to the hospital keep more `bed rest` than those who remain at home
  - `bed rest` leads to lower recovery times thus less walking patients after 1 week
- the *other variable* `bed rest` is a *mediator* between the treatment (`hospitalized`) and the outcome

Causal DAGs to the rescue

- the *other variable* was:
  - a **common cause** of the treatment and outcome in the pregnancy example
  - a **mediator** between the treatment and the outcome in the hernia example
- using our background knowledge we could see *something* is different about these examples
- next: ground this in causal theory and see implications for analysis

Some math background: probabilites and assumptions

Why math???

- need probability for estimation
- need conditional independence for causal inference
- need to understand ‘strength’ of assumptions

oh no math

Marginal, Joint and Conditional probabilites

Probability statements about *random events* A and B

- A: patient diest ($\mathrm{A} = 1$)
- B: patient has cancer ($\mathrm{B} = 1$)

statement	**interpretation**
$\mathrm{P}(\mathrm{A})$	*marginal* probability that event A occurs
$\mathrm{P}(\mathrm{B})$	*marginal* probability that event B occurs

Marginal, Joint and Conditional probabilites

Probability statements about *random events* A and B:

statement	interpretation
$\mathrm{P}(\mathrm{A})$	*marginal* probability that event A occurs
$\mathrm{P}(\mathrm{A}, \mathrm{B})$	*joint* probability of A and B

Marginal, Joint and Conditional probabilites

Probability statements about *random events* A and B:

statement	interpretation
$\mathrm{P}(\mathrm{A})$	*marginal* probability that event A occurs
$\mathrm{P}(\mathrm{A}, \mathrm{B})$	*joint* probability of A and B
$\mathrm{P}(\mathrm{A} \mid \mathrm{B})$	*conditional* probability of A given B

		A		
		dies	lives	
B	has cancer	5	5	10
	has no cancer			

- *marginal* $\mathrm{P}(\mathrm{A} = 1) = 15/100$
- *conditional* $\mathrm{P}(\mathrm{A} = 1 \mid \mathrm{B} = 1) = 5/10$

conditional probabilities require dividing by the denominator of the conditioning set

This is why we need *positivity* (as dividing by 0 is not defined)

Probability rules and identities

statement	interpretation
$P(A) = \sum_b P(A, B = b)$	marginal is sum over joint

		A	
		dies	lives
B	has cancer	5	
	has no cancer	10	
		15	100

$$
\begin{aligned}
P(A = 1) &= P(A = 1, B = 0) + P(A = 1, B = 1) \\
&= 5/100 + 10/100 \\
&= 15/100
\end{aligned}
$$

Probability rules and identities

statement	interpretation
$P(A) = \sum_b P(A, B = b)$	marginal is sum over joint
$P(A, B) = P(A\|B)P(B)$	product rule

		A	
		dies	lives
B	has cancer	5	10
	has no cancer		
			100

$$
\begin{aligned}
P(A = 1, B = 1) &= P(A = 1|B = 1)P(B = 1) \\
&= 5/10 * 10/100 \\
&= 5/100
\end{aligned}
$$

Probability rules and identities

statement	interpretation
$\mathrm{P}(\mathrm{A}) = \sum_{\mathrm{b}} \mathrm{P}(\mathrm{A}, \mathrm{B} = \mathrm{b})$	marginal is sum over joint
$\mathrm{P}(\mathrm{A}, \mathrm{B}) = \mathrm{P}(\mathrm{A}\|\mathrm{B})\mathrm{P}(\mathrm{B})$	product rule
$\mathrm{P}(\mathrm{A}\|\mathrm{B}) = \frac{\mathrm{P}(\mathrm{A},\mathrm{B})}{\mathrm{P}(\mathrm{B})}$	conditional is joint over marginal (follows from product rule)

Probability rules and identities

statement	interpretation
$\mathrm{P}(\mathrm{A}) = \sum_{\mathrm{b}} \mathrm{P}(\mathrm{A}, \mathrm{B} = \mathrm{b})$	marginal is sum over joint
$\mathrm{P}(\mathrm{A}, \mathrm{B}) = \mathrm{P}(\mathrm{A}\|\mathrm{B})\mathrm{P}(\mathrm{B})$	product rule
$\mathrm{P}(\mathrm{A}\|\mathrm{B}) = \frac{\mathrm{P}(\mathrm{A},\mathrm{B})}{\mathrm{P}(\mathrm{B})}$	conditional is joint over marginal (follows from product rule)
$\mathrm{P}(\mathrm{A}\|\mathrm{C}) = \sum_{\mathrm{b}} \mathrm{P}(\mathrm{A}\|\mathrm{B} = \mathrm{b}, \mathrm{C})\mathrm{P}(\mathrm{B} = \mathrm{b}\|\mathrm{C})$	total expectation (consequence of marginal vs joint and product rule)

Marginal and conditional independence:

statement	interpretation
$\mathrm{P}(\mathrm{A}, \mathrm{B}) = \mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B})$	(marginal) independence of A and B

- knowing A has no information on what to expect of B
- If I roll a die, the result of that die (A) has no information on the weather in the Netherlands (B)

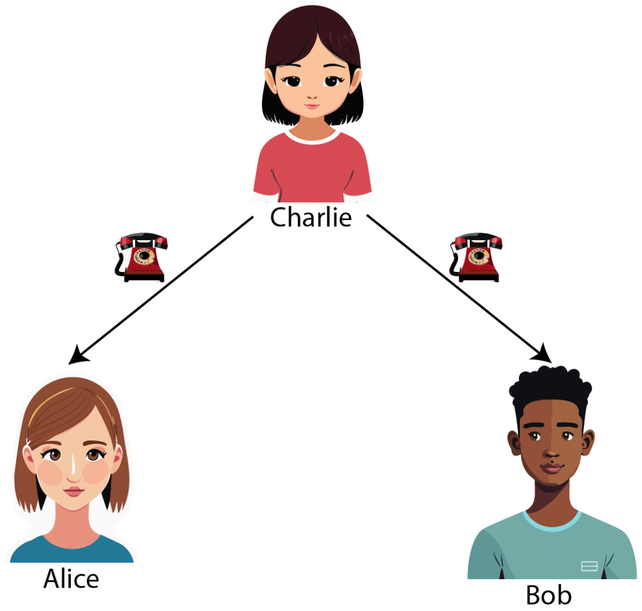
Marginal and conditional independence:

statement	interpretation
$\mathrm{P}(\mathrm{A}, \mathrm{B}) = \mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B})$	(marginal) independence of A and B
$\mathrm{P}(\mathrm{A}, \mathrm{B} \mid \mathrm{C}) = \mathrm{P}(\mathrm{A} \mid \mathrm{C})\mathrm{P}(\mathrm{B} \mid \mathrm{C})$	conditional independence of A and B given C
$\mathrm{P}(\mathrm{A} \mid \mathrm{B}, \mathrm{C}) = \mathrm{P}(\mathrm{A} \mid \mathrm{C})$	conditional independence of A and B given C

- C has all the information that is shared between A and B

Conditional Independence in an example

- Charlie calls Alice and reads her script C, then she calls Bob and reads him the same
- A week later we ask Alice to repeat the story Charlie told her, she remembered A, a noisy version of C
- We ask Bob the same, he recounts B, a different noisy version of C
- Are A and B independent? No! $\mathrm{P}(\mathrm{A}, \mathrm{B}) \neq \mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B})$
  - If we learn A from Alice, we can get a good guess about B from Bob
- If we knew C, would hearing A give use more information about B?
  - No, because all the shared information between A and B is explained by C, so:
  - $\mathrm{P}(\mathrm{A}, \mathrm{B}) \neq \mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B})$
  - $\mathrm{P}(\mathrm{A}, \mathrm{B}|\mathrm{C}) = \mathrm{P}(\mathrm{A}|\mathrm{C})\mathrm{P}(\mathrm{B}|\mathrm{C})$
- Variables can be marginally dependent but conditionally independent (and vice-versa)



ABC

Assumption parlance

- necessary assumption:
  - A **must** hold for B to be true
- sufficient assumption:
  - B is always true when A holds
- strong assumption:
  - requires *strong* evidence, we'd rather not make these
- weak assumption:
  - requires *weak* evidence
- strong vs weak assumption are judged on relative terms
  - if assumption A is sufficient for B, B cannot be a stronger assumption that A

DAG definitions

DAGs convey two types of assumptions:

causal direction and conditional independence

1. causal direction: what causes what?

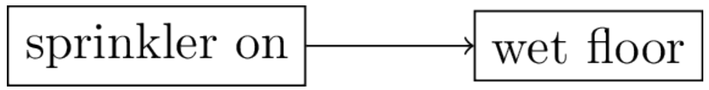
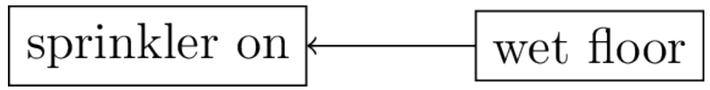


Figure 4: DAG 1



DAG 2

- read Figure 4 as
  - `sprinkler on` **may** (or may not) cause `wet floor`
  - `wet floor` **cannot** cause `sprinkler on`

DAGs convey two types of assumptions:

causal direction and conditional independence

1. conditional indepence (e.g. exclusion of influence / information)

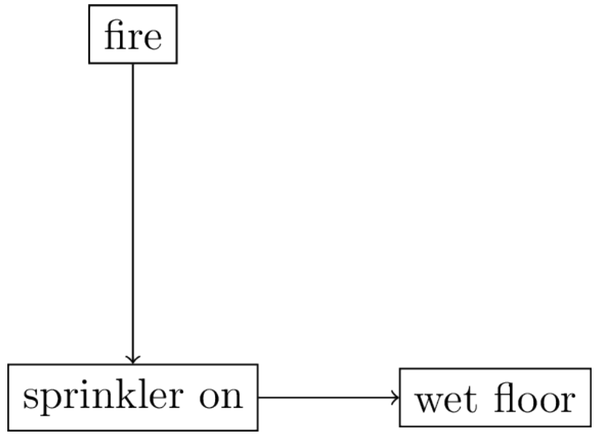


Figure 5: DAG 1

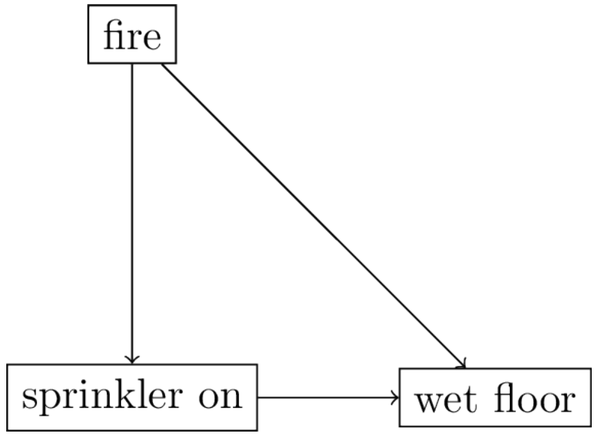


Figure 6: DAG 2

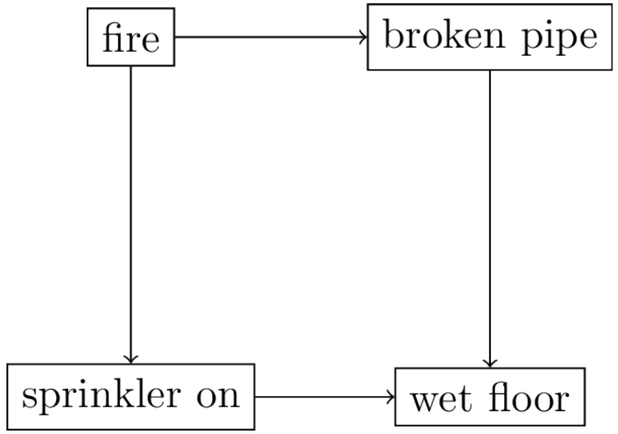
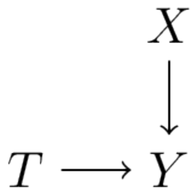


Figure 7: DAG 3

- Figure 5 says `fire` can **only** cause `wet floor` through `sprinkler on`
  - this implies `fire` is independent of `wet floor` given `sprinkler on` and can be tested!
- Figure 6 says *there may be other ways through which `fire` causes `wet floor`*
  - Figure 6 is thus a *weaker* assumption than Figure 5
- Figure 7 is also compatible with Figure 6

DAGs are ‘non-parametric’

They relay what variable ‘listens’ to what, but not in what way



DAG

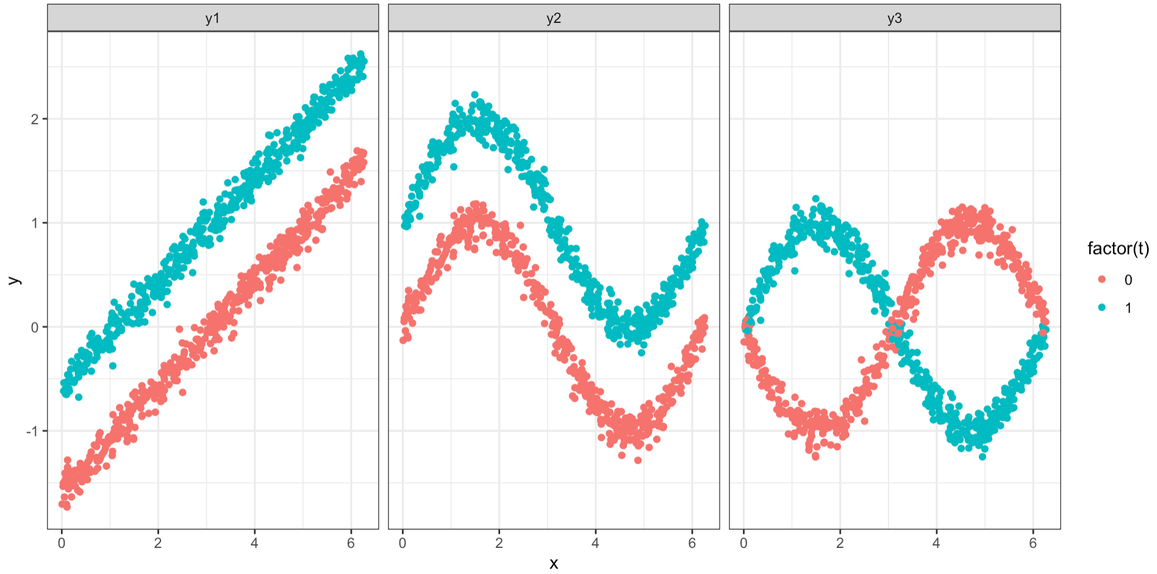
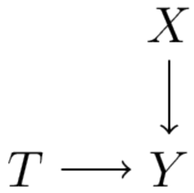


Figure 8: Three datasets with the same DAG

1. $\mathrm{Y} = \mathrm{T} + 0.5(\mathrm{X} - \pi) + \epsilon$ (linear)
2. $\mathrm{Y} = \mathrm{T} + \sin(\mathrm{X}) + \epsilon$ (non-linear additive)
3. $\mathrm{Y} = \mathrm{T} * \sin(\mathrm{X}) - (1 - \mathrm{T})\sin(\mathrm{x}) + \epsilon$ (non-linear + interaction)

DAGs are ‘non-parametric’

They relay what variable ‘listens’ to what, but not in what way



DAG

- this DAG says Y is a function of X, T and external noise $\mathrm{U_Y}$, or:
- $\mathrm{Y} = \mathrm{f_Y}(\mathrm{X}, \mathrm{T}, \mathrm{U_Y})$
- in the next lecture we’ll talk more about these ‘structural equations’

DAGs imply a causal factorization of the joint distribution

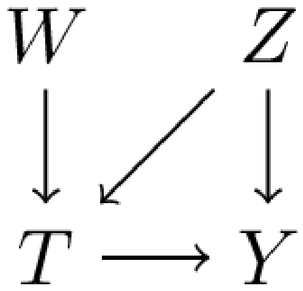


Figure 9: observational data

$$
\begin{aligned}
\mathrm{P}(\mathrm{Y},\mathrm{T},\mathrm{Z},\mathrm{W}) &=^{1} \mathrm{P}(\mathrm{Y}\,|\,\mathrm{T},\mathrm{Z},\mathrm{W})\mathrm{P}(\mathrm{T},\mathrm{Z},\mathrm{W}) \\
&=^{2} \mathrm{P}(\mathrm{Y}\,|\,\mathrm{T},\mathrm{Z})\mathrm{P}(\mathrm{T},\mathrm{Z},\mathrm{W}) \\
&=^{3} \mathrm{P}(\mathrm{Y}\,|\,\mathrm{T},\mathrm{Z})\mathrm{P}(\mathrm{T}\,|\,\mathrm{Z},\mathrm{W})\mathrm{P}(\mathrm{Z},\mathrm{W}) \\
&=^{4} \mathrm{P}(\mathrm{Y}\,|\,\mathrm{T},\mathrm{Z})\mathrm{P}(\mathrm{T}\,|\,\mathrm{Z},\mathrm{W})\mathrm{P}(\mathrm{Z})\mathrm{P}(\mathrm{W})
\end{aligned}
$$

1. product-rule
2. Y independent of W given T,Z per DAG
3. product-rule
4. Z,W marginally independent per DAG

- If this looks complicated: just follow the arrows

The DAG definition of an intervention

assume this is our DAG for a situation and we want to learn the effect T has on Y

- this is denoted $\mathrm{P}(\mathrm{Y}\,|\mathrm{do}(\mathrm{T}))$: a hypothetical intervention in the system
- in the graph, intervening on variable T means removing all incoming arrows
- this assumes such a *modular* intervention is possible: i.e. leave everything else unaltered

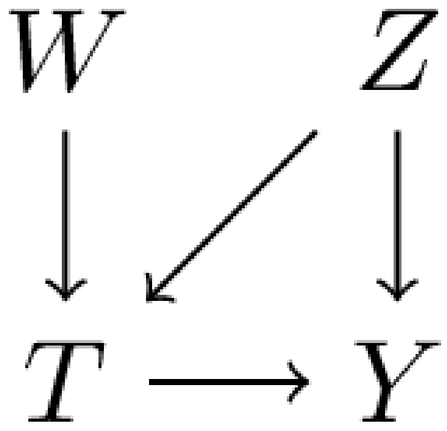


Figure 10: observational data

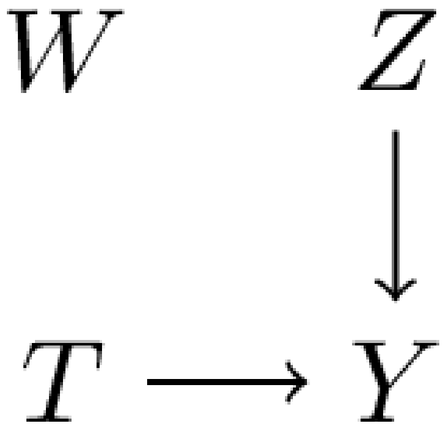


Figure 11: intervened DAG

- which means T does not *listen* to other variables anymore, but is set at a particular value, like in an experiment
- imagining this scenario requires a well-defined treatment variable (akin to consistency)

Intervention as graph surgery - changed distribution

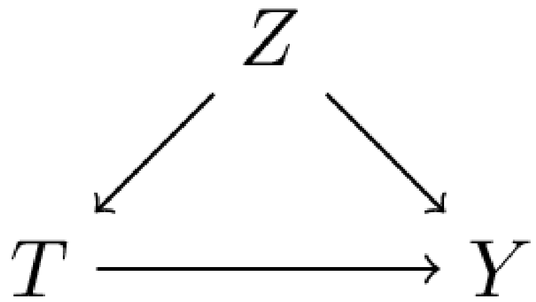


Figure 12: observational data

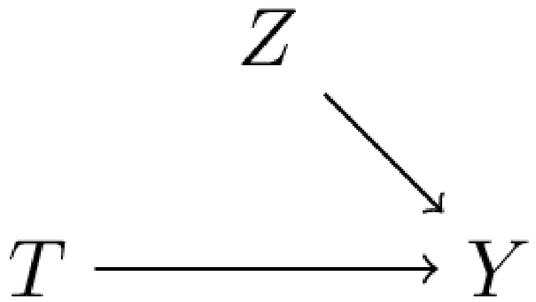


Figure 13: intervened DAG

$$P_{obs}(Y,T,Z) = P(Y|T,Z)P(T|Z)P(Z)$$

$$P_{obs}(Y|T) = \sum_{z} P(Y|T,Z=z)P(Z=z|T)$$

$$P_{int}(Y,T,Z) = P(Y|T,Z)P(T)P(Z)$$

$$P_{int}(Y|T) = \sum_{z} P(Y|T,Z=z)P(Z=z|T)$$

Intervention as graph surgery - changed distribution

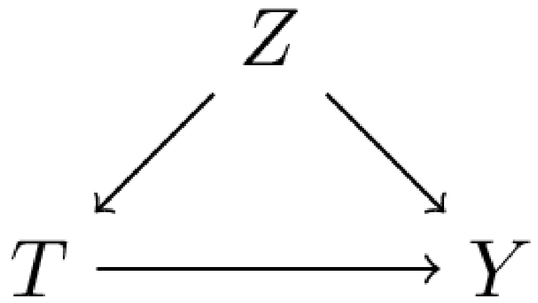


Figure 14: observational data

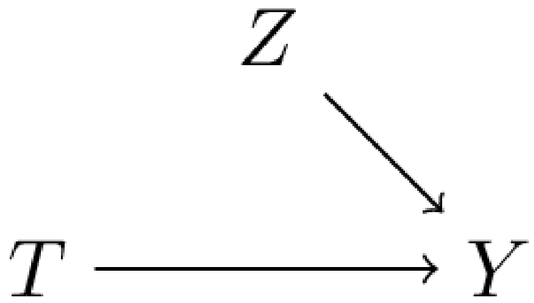
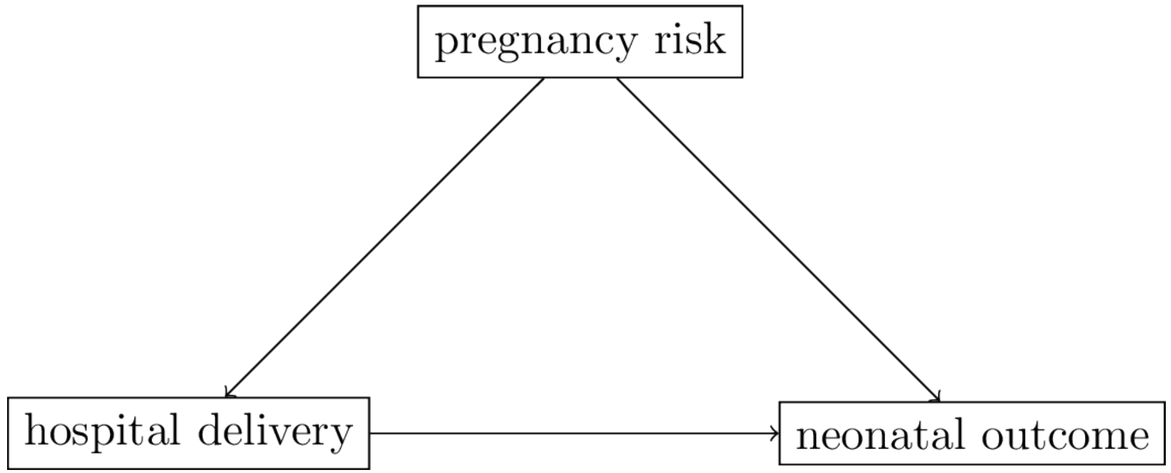


Figure 15: intervened DAG

$$\mathrm{P}_{\mathrm{obs}}(\mathrm{Y}\,|\mathrm{T}) = \sum_{\mathrm{z}} \mathrm{P}(\mathrm{Y}\,|\mathrm{T},\mathrm{Z}=\mathrm{z})\mathrm{P}(\mathrm{Z}=\mathrm{z}|\mathrm{T}) \qquad \mathrm{P}_{\mathrm{int}}(\mathrm{Y}\,|\mathrm{T}) = \sum_{\mathrm{z}} \mathrm{P}(\mathrm{Y}\,|\mathrm{T},\mathrm{Z}=\mathrm{z})\mathrm{P}(\mathrm{Z}=\mathrm{z}) \tag{1}$$

- in $\mathrm{P}_{\mathrm{obs}}$, $\mathrm{P}(\mathrm{Z}|\mathrm{T}) \neq \mathrm{P}(\mathrm{Z})$
- in $\mathrm{P}_{\mathrm{int}}$, $\mathrm{P}(\mathrm{Z}|\mathrm{T}) = \mathrm{P}(\mathrm{Z})$
- thereby $\mathrm{P}_{\mathrm{obs}}(\mathrm{Y}\,|\mathrm{T}) \neq \mathrm{P}_{\mathrm{int}}(\mathrm{P}(\mathrm{Y}\,|\mathrm{T})) = \mathrm{P}(\mathrm{Y}\,|\mathrm{do}(\mathrm{T}))$
- **seeing is not doing**
- **looking at Equation 1, we can compute these from $\mathrm{P}_{\mathrm{obs}}$!** (this is what is called an *estimand*)

Back to example 1



DAG

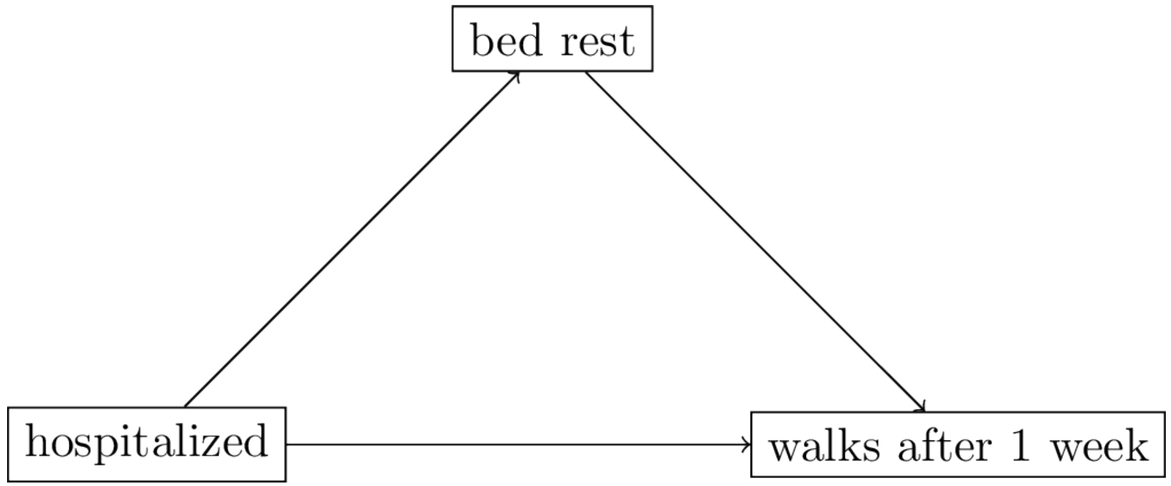
		location	
		home	hospital
risk	low	648 / 720 = 90%	19 / 20 = 95%
	high	40 / 80 = 50%	144 / 180 = 80%
	marginal	688 / 800 = 86%	163 / 200 = 81.5%

- estimand: $\mathrm{P}(\mathrm{outcome}|\mathrm{do}(\mathrm{location})) = \sum_{\mathrm{risk}} \mathrm{P}(\mathrm{outcome}|\mathrm{location}, \mathrm{risk})\mathrm{P}(\mathrm{risk})$
- $\mathrm{P}(\mathrm{risk} = \mathrm{low}) = 74\%$

$$\mathrm{P}(\mathrm{outcome}|\mathrm{do}(\mathrm{hospital})) = 95 * 0.74 + 80 * 0.26 = 91.1\%$$
$$\mathrm{P}(\mathrm{outcome}|\mathrm{do}(\mathrm{home})) = 90 * 0.74 + 50 * 0.26 = 79.6\%$$

- **conclusion**: sending all deliveries to the hospital leads to better neonatal outcomes

Back to example 2



DAG

- removing all arrows going in to T results in the same DAG
- so $\mathrm{P(Y\,|T)} = \mathrm{P(Y\,|do(T))}$
- i.e. use the marginals

The gist of observational causal inference

is to take data we have to make inferences about data from a different distribution (i.e. the intervened-on distribution)

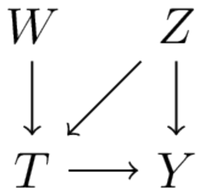


Figure 16: observational data: data we have

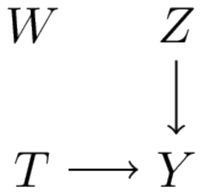
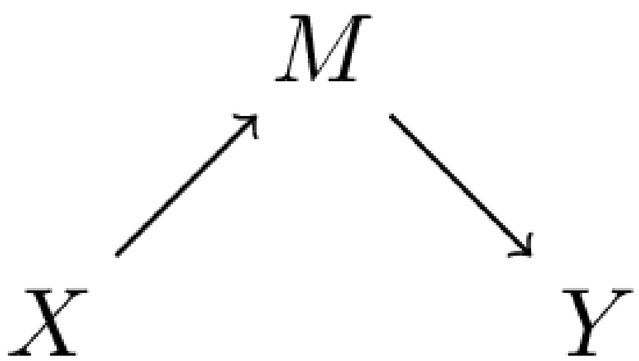


Figure 17: intervened DAG: what we want to know

- causal inference frameworks provide a language to express assumptions
- based on these assumptions, the framework tells us whether such an inference is possible
  - this is often referred to as *is the effect identified*
- and provide formula(s) for how to do so based on the observed data distribution (*estimand(s)*)
- (one could say this is essentially assumption-based extrapolation, some researchers think this entire enterprise is anti-scientific)
- not yet said: *how* to do statistical inference to estimate the estimand (much can still go wrong here)
  - can also be part of identification, see the following lecture on SCMs

DAG rules

Basic DAG patterns: chain



- M *mediates* effect of X on Y
- $\mathrm{X} \perp \mathrm{Y} | \mathrm{M}$
- do not want to adjust for M when estimating total effect of X on Y

Figure 18: chain / mediation

Basic DAG patterns: fork

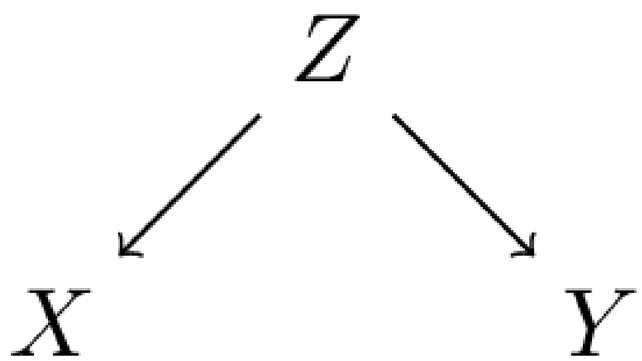


Figure 19: fork / confounder

- Z *causes* both X and Y (common cause / confounder)
- $\mathrm{X} \perp \mathrm{Y} | \mathrm{Z}$
- $\mathrm{Z} \rightarrow \mathrm{X}$ is a *back-door*: a path between X and Y that starts with an arrow into X
- typically want to adjust for Z (see later 5.9)

Basic DAG patterns: collider

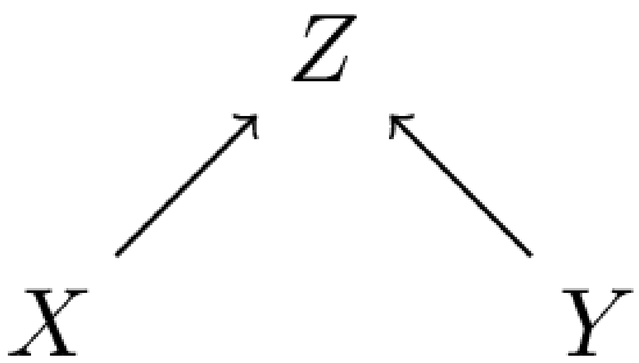
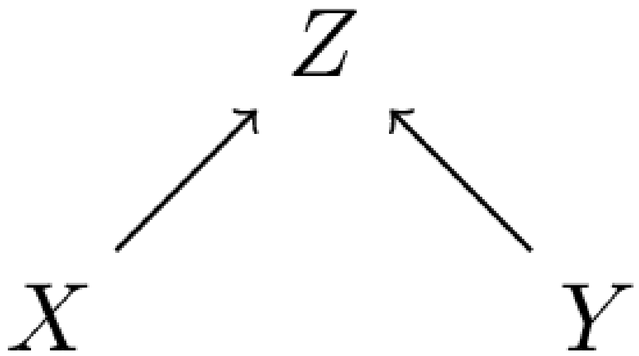


Figure 20: collider

- X and Y *both cause* Z
- $\mathrm{X} \perp \mathrm{Y}$ (but *NOT* when conditioning on Z)
- often do not want to condition on Z as this induces a correlation between X and Y

Collider bias - Tinder



(a) collider

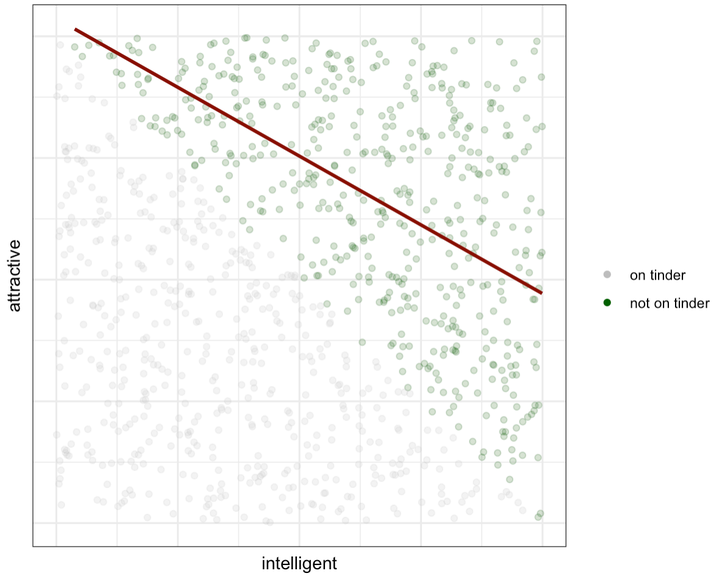


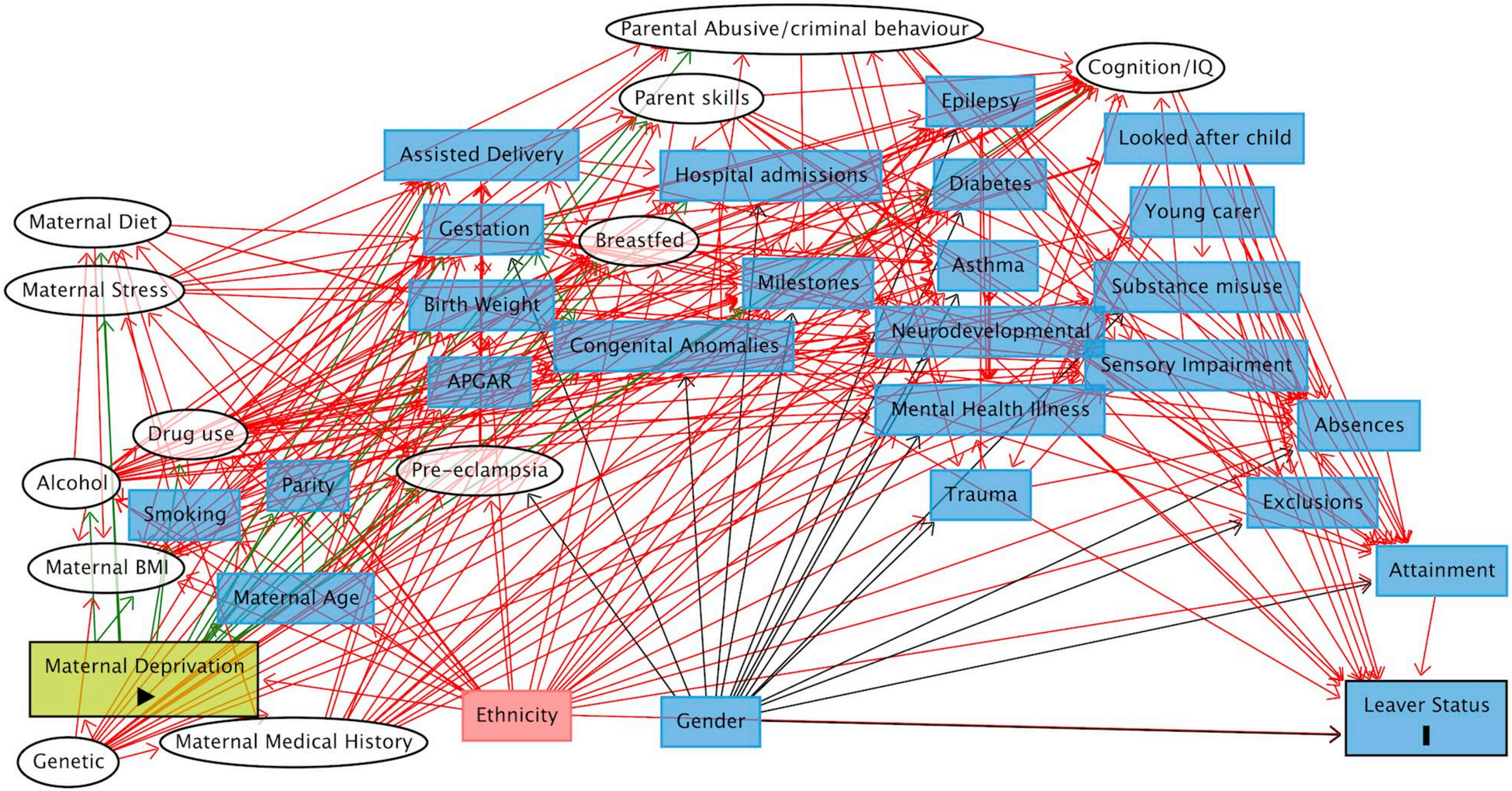
Figure 21:

$$\begin{aligned} \text{intelligent} &\sim \text{U}[0,1] \\ \text{attractive} &\sim \text{U}[0,1] \\ \text{on tinder} &= \text{I}_{\text{intelligent+attractive<1}} \end{aligned}$$

Conditioning on a collider creates dependence of its parents

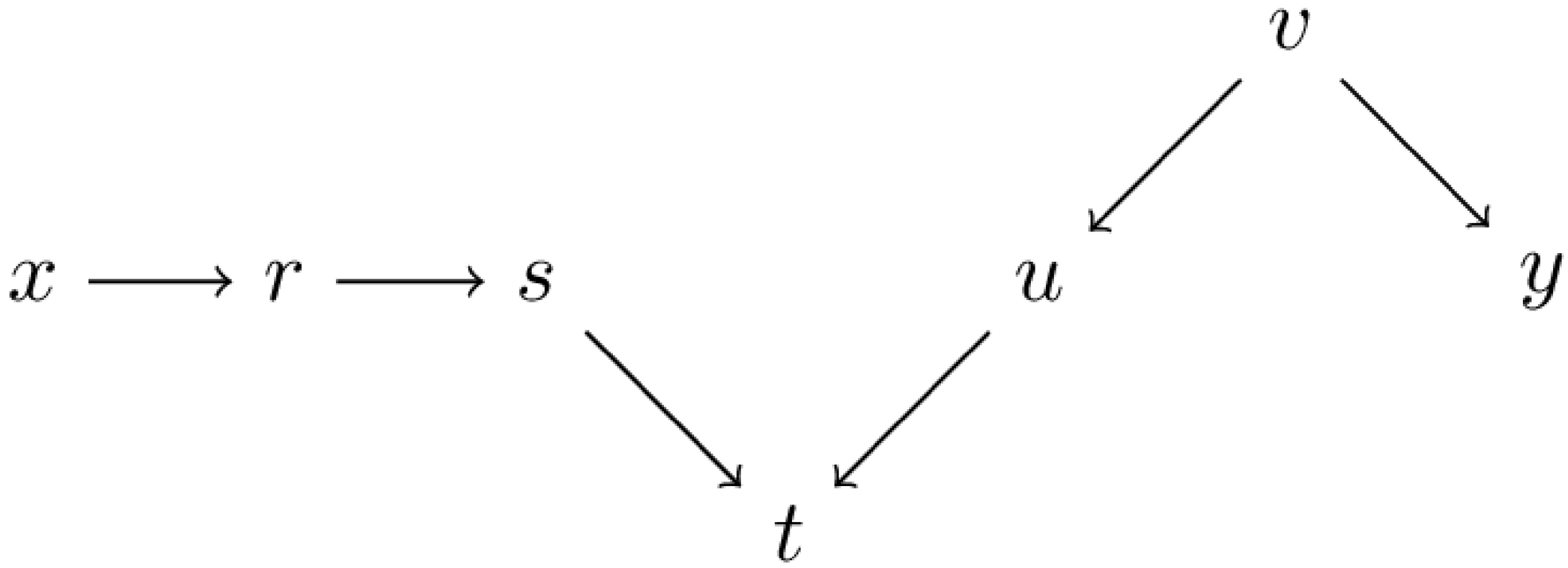
- may not be too visible: doing an analysis in a selected subgroup is a form of (‘invisible’) conditioning)
- e.g. when selecting only patients in the hospital
  - being admitted to the hospital is a collider (has many different causes, e.g. traffic accident or fever)
  - usually only one of these is the reason for hospital admission
  - the causes for hospital admission now seem anti-correlated
- collider conditioning *might* be an explanation for the *obsesity paradox* (i.e. obesity is correlated with better outcomes in diverse medical settings) (e.g. Banack and Stokes 2017)

When life gets complicated / real



Bogie, James; Fleming, Michael; Cullen, Breda; Mackay, Daniel; Pell, Jill P. (2021). Full directed acyclic graph.. PLOS ONE. Figure. https://doi.org/10.1371/journal.pone.0249258.s003

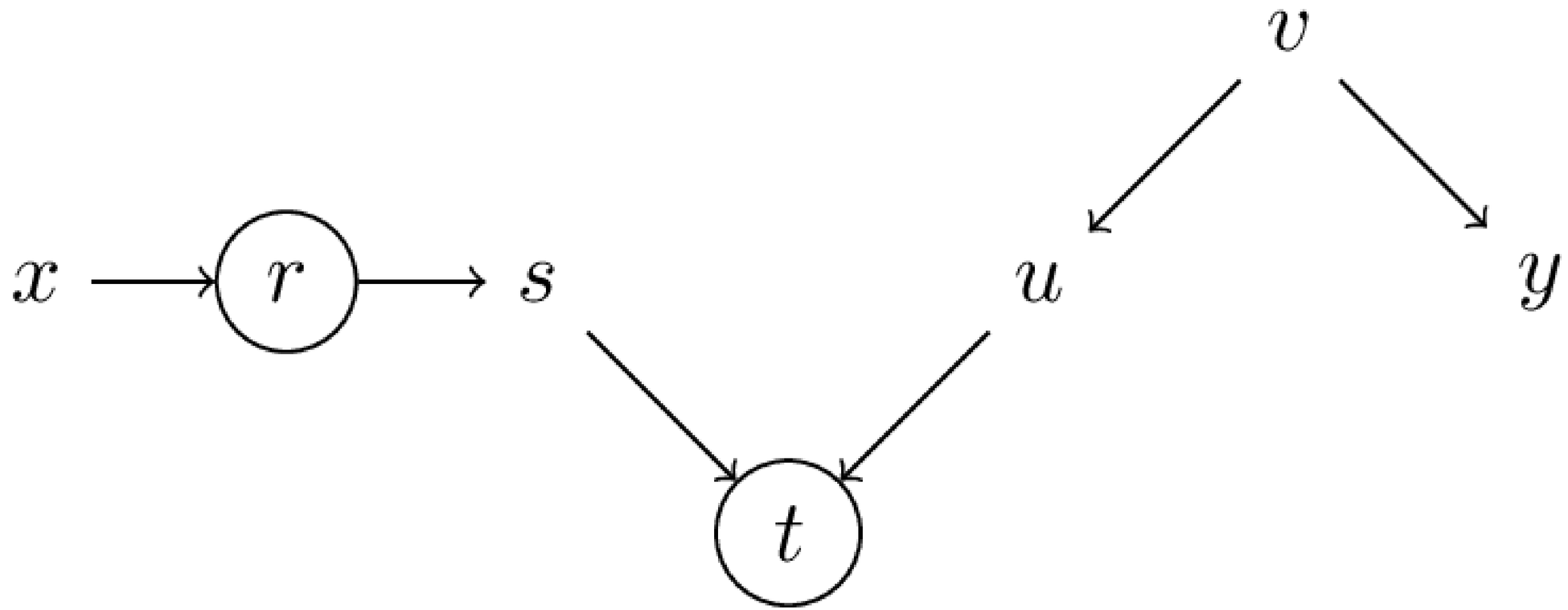
d-separation (directional-separation)



paths

- a *path* is a set of nodes connected by edges $(\mathrm{x} \ldots \mathrm{y})$
- a *directed-path* is a path with a constant direction $(\mathrm{x} \ldots \mathrm{t})$
- an *unblocked-path* is a path without a collider $(\mathrm{t} \ldots \mathrm{y})$
- a *blocked-path* is a path with a collider $(\mathrm{s}, \mathrm{t}, \mathrm{u})$
- *d(irectional)-separation* of x, y means there is no unblocked path between them

d-separation when conditioning



paths with conditioning variables r, t

- conditioning on variable:
  - when variable is a collider: *opens a path* (t opens $\mathrm{s}, \mathrm{t}, \mathrm{u}$ etc.)
  - otherwise: *blocks a path* (e.g. r blocks $\mathrm{x}, \mathrm{r}, \mathrm{s}$)
- conditioning *set* $\mathrm{Z} = \{\mathrm{r}, \mathrm{t}\}$: set of conditioning variables

The back-door criterion and adjustment

Definition 3.3.1 (Back-Door) (for pairs of variables)

A set of variables Z satisfies the *back-door* criterion relative to an ordered pair of variables (X, Y) in a DAG if:

1. no node in Z is a descendant of X *(e.g. mediators)*
2. Z blocks every path between X and Y that contains an arrow into X

Theorem 3.2.2 (Back-Door Adjustment)

If a set of variables Z satisfies the back-door criterion relative to (X, Y), then the causal effect of X on Y is identifiable and is given by the formula

$$\mathrm{P(y|do(x))} = \sum_{\mathrm{z}} \mathrm{P(y|x,z)P(z)} \qquad (2)$$

Did we see this equation before?

- Yes! When computing the effect of hospital deliveries on neonatal outcomes Equation 1
- DAGs tell us what to adjust for
- automatic algorithms tell use whether an estimand exists and what it is
- several point-and-click websites for making DAGs that implement these algorithms:
  - dagitty.net
  - causalfusion.net

How about positivity

- backdoor adjustment with z requires computing $\mathrm{P}(\mathrm{y}|\mathrm{x},\mathrm{z})$
- by the product rule:

$$\mathrm{P}(\mathrm{y}|\mathrm{x},\mathrm{z}) = \frac{\mathrm{P}(\mathrm{y},\mathrm{x},\mathrm{z})}{\mathrm{P}(\mathrm{x},\mathrm{z})}$$

- this division is only defined when $\mathrm{P}(\mathrm{x},\mathrm{z}) > 0$
- which is the same as the positivity assumption from Day 1 in Potential Outcomes

References

Banack, H. R., and A. Stokes. 2017. “The ‘Obesity Paradox’ May Not Be a Paradox at All.” *International Journal of Obesity* 41 (8): 1162–63. https://doi.org/10.1038/ijo.2017.99.

Pearl, Judea. 2009. *Causality*. Cambridge University Press.