~~Question~~:

To build: (1) Bayesian Networks – Strictly Statistical
(2) Causality – an Intervention on top of a Bayesian network

We'll think of BNs causally (though you don't have to)

- We want to model a joint distribution.
- We want to assume some independences to simplify it

To model $P(X_1, \dots, X_N)$ on N variables

To assume rules of the form : $\{ P(X_1 | X_2, X_3) = P(X_1 | X_2), \dots \}$

Solution! Chain rule for conditional independence

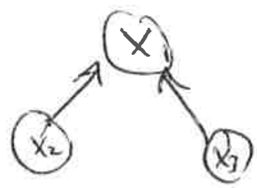
$$P(X_1, \dots, X_n) = P(X_1 | X_2, \dots, X_n) P(X_2 | X_3, \dots, X_n) \dots P(X_n)$$
$$= \prod_{i=1}^{n} P(X_i | X_{i+1}, \dots, X_n)$$

example $P(X_1, X_2, X_3)$ where $P(X_1 | X_2, X_3) = P(X_1 | X_2)$

$$P(X_1, X_2, X_3) = P(X_1 | X_2, X_3) P(X_2 | X_3) P(X_3)$$
$$= P(X_1 | X_2) P(X_2 | X_3) P(X_3)$$

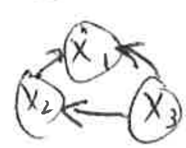
consider a visual for this.
(1) Each term is a node
(2) Each variable conditioned on is a parent.

$P(X_1 | X_2, X_3) \Rightarrow$



* No Causal Interpretation Yet *

$P(X_1 | X_2, X_3) P(X_2 | X_3) P(X_3) \Rightarrow$



use the assumption now...

$P(X_1 | X_2) P(X_2 | X_3) \; P(X_3) \Rightarrow$ $\quad X_1 \leftarrow X_2 \leftarrow X_3$

* Our independence assumption implies we cut an edge! *

A couple Notes! (1) Complete graph $\Rightarrow$ no assumptions.

(2) Variable ordering is arbitrary! So any complete graph can represent any joint distribution.

(Statistical representation, not a causal representation. Causality comes from extra structure: interventions)

(3) $P(X_1, X_2, X_3)$ where $P(X_1 | X_2, X_3) = P(X_1 | X_2)$ can be represented by

$X_1 \rightarrow X_2 \rightarrow X_3$ $\qquad$ $1 \leftarrow 2 \leftarrow 3$

$X_1 \leftarrow X_2 \rightarrow X_3$

but not

$X_1 \rightarrow X_2 \leftarrow X_3$

"V-structure" or "collider".

Cutting the $1 \leftrightarrow 3$ edge above implies $P(X_1 | X_2, X_3) = P(X_1 | X_2)$ but this graph doesn't imply that. Instead:

$P(X_1, X_3) = P(X_1) P(X_3)$

$\Rightarrow$ Graphs with the same skeleton can represent the same joint distributions only if they have the same V-structures

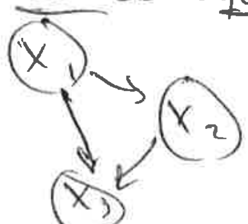
- What we're talking about up to this point is models for joint distributions. These summarize the statistical properties of a system.
- If we only care about statistics, we can use any graph that is "observationally equivalent", i.e. it has the same colliders and skeleton, that can express the dependences in the joint distribution.

— END Bayesian Networks, BEGIN Causality —

Key assumption: One of these graphs is the right one to model the set of possible interventions.

Intervention: In real life, disrupt the system to fix the value of a node, independently from what the value would have been if you hadn't disrupted it.

model for observed system

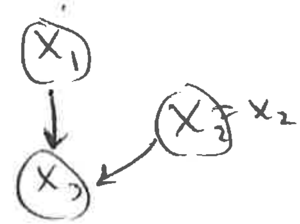


G

now we intervene on X_2

$\Longrightarrow$

model for intervention



Steps: (1) Cut all edges in G going into the node you're intervening on, producing a new graph $G_{\overline{X_2}}$

notation: $\left(\begin{array}{l} \overline{X_2} \Rightarrow \text{causes of } X_2 \text{ deleted} \\ \underline{X_2} \Rightarrow \text{effects of } X_2 \text{ deleted} \end{array} \right)$

(2) Set $X_2 = x_2$, the value you choose for your intervention

(3) The rest happens as normal.

observational system	interventional system
$P(X_3, X_2, X_1) = P(X_3 \mid X_2, X_1) P(X_2 \mid X_1) P(X_1)$	$P(X_3, X_1 \mid do(X_2 = x_2)) =$
$\Rightarrow P(X_3, X_1 \mid X_2) = \dfrac{P(X_3 \mid X_2, X_1) P(X_2 \mid X_1) P(X_1)}{P(X_2)}$	$P(X_3 \mid X_1, X_2 = x_2) P(X_1)$

$\Rightarrow$ different from conditioning.

* but still estimable from the old joint + marginals!!! *

Notice

$$P(X_3, X_1 \mid do(X_2 = x_2)) = \frac{P(X_3, X_2, X_1)}{P(X_2 \mid X_1)} = \cancel{P(X_1, \dots, X_n)}$$

In general,

$$P(X_1, \dots, X_n \mid do(X_i = x_i)) = \frac{P(X_1, \dots, X_n)}{P(X_i \mid Pa(X_i))}$$

Bayesian Network together with this notion of intervention are called "Causal Bayesian Networks".

p. 110-111 SS. example of bad collider practice.

Does X cause Y? What does cause mean?

(Randomized control!)

~~$P(Y|do(X=x'))$~~

$P(Y|do(X=hi)) \neq P(Y|do(X=lo))$

assume stability

$X \to Z \to Y$

intercepted!

yes. (indirectly, through mechanism Z, X, Y correlated)

$X \leftarrow Z \to Y$

see (18)

no ("confounding" lemonade example correlated X & Y)

$X \to Z \leftarrow Y$

no (not even correlated)

~~Regression Example for~~ which

Direct vs. Indirect Effects, mediation

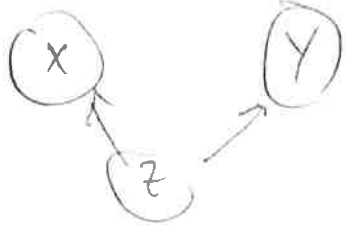
Direct effect: X a parent of Y

Indirect effect: X upstream from Y. (← there is a directed path from X to Y, $X \notin Pa(Y)$)

- Are there real "direct" effects? mediators, suppressing mediators.
- We often (legitimately) ignore mechanisms.
- Okay as long as you don't suppress confounders. Then you have to introduce arcs, e.g.

$W \to Z,\ Z \to X,\ Z \to Y \quad \longrightarrow \quad W \to X,\ W \to Y \quad \longrightarrow \quad X \leftarrow\cdots\to Y$

example: what does confounding look like? (p. 5)



$$Y = \beta_{yz} Z + \varepsilon_Y$$

$$X = \beta_{xz} Z + \varepsilon_X$$

$$Z = \varepsilon_Z$$

$$\cancel{Y = \varepsilon_Z \beta_{YZ} + \varepsilon_Y =}$$

$$\beta_{yx} = \frac{\operatorname{cov}(x, y)}{\sigma_x^2} = \frac{\operatorname{cov}(\beta_{yz} Z, \beta_{xz} Z)}{\sigma_x^2}$$

$$= \beta_{yz} \beta_{xz} \frac{\sigma_z^2}{\sigma_x^2}$$

nonzero (unconditional) regression coefficient, with no directed causal path from X to Y! (or vice versa)

- Fix Z, then $\sigma_z^2 = 0$, and $\beta_{yx} \to 0$
  → controlling ~~breaks~~ breaks confounding
- set either β_{yz} or β_{xz} (or both) to 0, and you break confounding
  ⇒ due to common causes.

How do paths ~~leadto~~ lead to dependence?

$$X \to Y \to Z$$

$$P(X,Z) = \sum_{Y} P(Z|Y) \underbrace{P(Y|X)}_{\text{coupled}} P(X)$$

$$Y \to Z \leftarrow Y$$

marginalized.

$$P(X,Y) = \sum_{Z} P(Z|X,Y) \cancel{P(Z|X)} P(X)P(Y)$$

$$= P(X,Y)$$

$$X \to A \to Z \leftarrow B \leftarrow Y$$

$$P(X,Y) = \sum_{A,B,Z} \cancel{P(A|X)} P(X)P(A|X)P(Z|A,B)P(B|Y)P(Y)$$

$$= \sum_{A,B} P(X) P(A|X) P(B|Y) P(Y)$$

$$= P(X)P(Y)$$

$$X \leftarrow Z \to Y$$

$$P(X,Y) = \sum_{Z} P(X|Z) \underbrace{P(Z) P(Y|Z)}_{\text{coupled}}$$

(7)

Regressions just give conditional expectations (when properly specified, and using MSE loss), so if

$Y \perp X \mid Z$ (Z d-separates X and Y), then

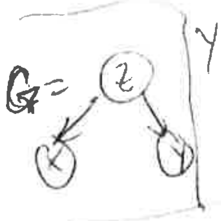
$$Y(x,z) = \sum_y y\, P(Y \mid X, Z) = \sum_y y\, \frac{P(Y \mid Z)\, P(X \mid Z)\, P(Z)}{P(X,Z)}$$

$$= \sum_y y\, P(Y \mid Z)\, \frac{P(X,Z)}{P(Z)}\, \frac{P(Z)}{P(X,Z)} = \sum_y P(Y \mid Z)\, P(Y)$$

the regression estimate is independent of X when you include Z! (but not in general, since $P(Y \mid X, Z) = P(Y \mid Z) \not\Rightarrow P(Y \mid X) = P(Y)$ i.e. conditional independence does not imply independence)

We can look at this in the context of confounding, and see

$$P(X,Y,Z) = P(Y \mid Z)\, P(X \mid Z)\, P(Z)$$ (from factorization implied by G)



$$Y(x,z) = \sum_y y\, \frac{P(X,Y,Z)}{P(X,Z)} = \sum_y y\, \frac{P(Y \mid Z)\, P(X \mid Z)\, P(Z)}{P(X,Z)}$$

$$= \sum_y y\, P(Y \mid Z)\, \frac{\cancel{P(X,Z)}}{\cancel{P(Z)}}\, \frac{\cancel{P(Z)}}{\cancel{P(X,Z)}} = \sum_y y\, P(Y \mid Z)$$

independent of X!, but without Z,

$$Y(x) = \sum_y y\, P(Y \mid X) = \sum_{z,y} y\, P(Y, Z \mid X) = \sum_{z,y} y\, \frac{P(X,Y,Z)}{P(X)}$$

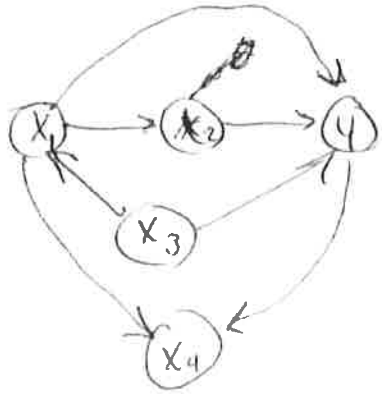
$$= \sum_{z,y} y\, \frac{P(Y \mid Z)\, P(X \mid Z)\, P(Z)}{P(X)}$$

So X and Y are coupled by the sum over Z! (won't factor in general)

Wanted: Total effect

How can we block all the bad paths?

which paths should we block? which are good?



× bad $X_1 \leftarrow X_3 \rightarrow Y$ Spurious not downstream!

✓ good $X_1 \rightarrow Y$ direct effect

✓ good $X_1 \rightarrow X_2 \rightarrow Y$ mediated effect.

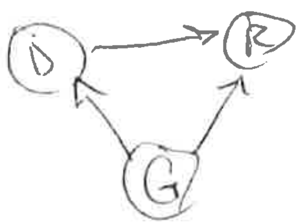
× bad $X_1 \rightarrow X_4 \leftarrow Y$ collider bias (selection effect ... admissions example)

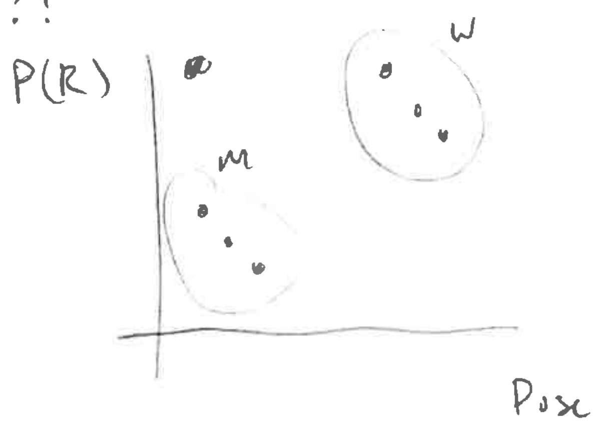
Knowing when to condition resolves Paradoxes!

A drug ~~trial~~ study (Simpson's Paradox)

Examples ~~Simpson's Paradox~~

- Positive effect overall
- Negative effect on ~~ppl~~ men / women
- Condition or not ?!

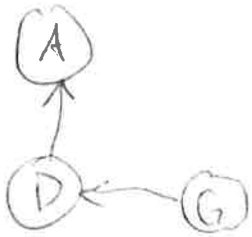




~~Goal~~

- Women comply more, have naturally higher recovery rates
- Men comply less, have naturally lower rates.

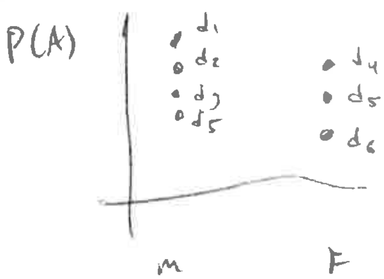
Berkely Admission Paradox



$P(A \mid G = F) < P(A \mid G = M)$

is there gender ~~bias~~ discrimination?

- Because of mediation, there is no discrimination.



You really need no spurious paths to establish causation.

→ Guaranteed in randomized trial

Ignorability → Guaranteed when theyre "blocked"

$(Y^0, Y^1) \perp D$ $= \begin{pmatrix} \text{assignment} \\ \text{ignorable} \end{pmatrix} =$ an RCT

$(Y^0, Y^1) \perp D \mid S$ = conditionally ignorable = as good as RCT.

$S \supset Z$, improve precision of effect measurement.

(regression example in jupyter) → controlling breaks dependence b/t D and Y^0, Y^1 gives unbiased estimator for δ

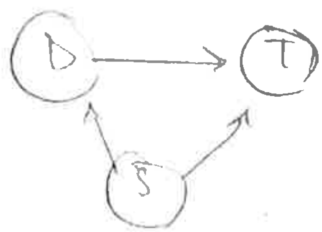
weaker version:

$Y^0 \perp D$, but not $(Y^1, Y^0) \perp D$; still identify ATT

vice versa for ATC

Matching + Exposure modeling : Next time!

$P(D \mid S) = ?$



Same $P(D \mid S) \Rightarrow$ effectively fixed S!