Lecture 13: Intro to Causality

Suppose we look at hospital data about kidney stone treatments:

	Treatment A	Treatment B
Small stones	93% success rate (81/87)	87% (234/270)
Large stones	73% (192/263)	69% (55/80)
Combined	78% (273/350)	83% (289/350)

What's going on?

Treatment A seems to be more effective for small and large stones, but not in the two groups combined.

This is an example of <u>Simpson's paradox</u>.

Formally,

$$\Pr\{y \mid A\} < \Pr\{y \mid B\}$$

$$\Pr\{y \mid A, X\} > \Pr\{y \mid B, X\}$$

$$\Pr\{y \mid A, \neg X\} > \Pr\{y \mid B, \neg X\}$$

Mathematically, there is no contradiction. Yet, Simpson's paradox causes discomfort.

Why?

We tend to read conditional events as <u>actions</u>, but they are not.

Conditional events are <u>observations</u>.

We observe doctors in a hospital.

We see who gets treatment A (or B) according to the doctor's natural inclination.

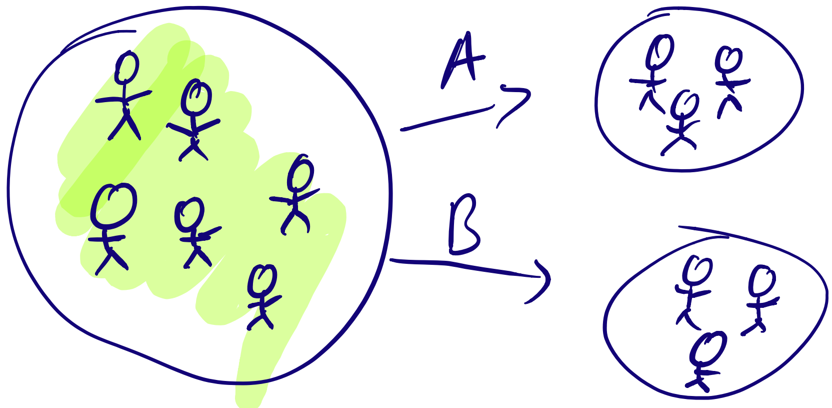
There is no intervention, no action. just passive observation.

In our example, one possible story is that doctors assign treatment B more often to mild cases (small stones) who generally have a higher success rate.

Thus the size (X) influences the choice of treatment.

Contrast this with a randomized controlled trial (RCT)

We randomly assign treatment to patients regardless of size, therefore breaking the natural practise of the doctor



This active assignment is an action.

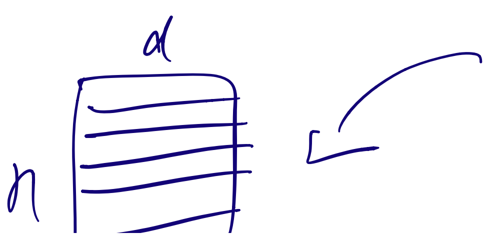
Its effects are in general not given by conditional probability (observation).

Question:

How can we formalize actions?

Once we formalized actions, we can talk about causation.

E.g. "does A cause Y?"



datagen.ipynb

1. ___
2. ___
3.

Intuition from programming

Suppose you have a program to generate a distribution step by step.

1. Sample Bernoulli random vars
   $U_1 \sim B(1/2)$, $U_2 \sim B(1/3)$, $U_3 \sim B(1/3)$
2. $X := U_1$ (exercise)
3. $W :=$ if $X=1$ then 0 else U_2 (overweight)
4. $H :=$ if $X=1$ then 0 else U_3 (heart disease)

This defines a joint distribution over X, W, H. We can compute probabilities in this joint distribution.

Let's compute a few:

$\Pr\{H\} = 1/2 \cdot 1/3 = 1/6$

$\Pr\{H \mid W\} = 1/3$

$\Pr\{H \mid \neg W\} =$

Consider substituting:

2. $X := U_1$

3. $W := 1$

4. $H :=$ if $X = 1$ then 0 else U_3

In this new program, the probability of H is still $1/6$.

We write this as

$$\Pr\{H \mid do(W=1)\} = \frac{1}{6}$$

do-substitution

called do-intervention.

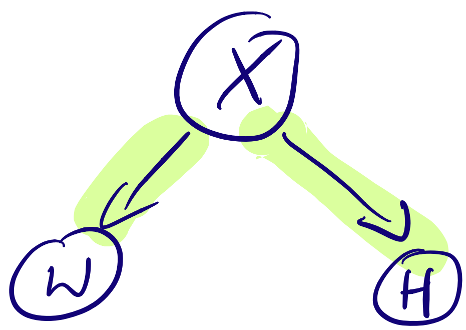
Note: In general,

$$\Pr\{H \mid \underbrace{W=1}_{\text{"observing } W=1\text{"}}\} \neq \Pr\{H \mid \underbrace{do(W=1)}_{\text{"doing } W=1\text{"}}\}$$

The "programs" we saw are called <u>structural causal models</u>.

They come with an acyclic assignment graph, called <u>causal graph</u>.

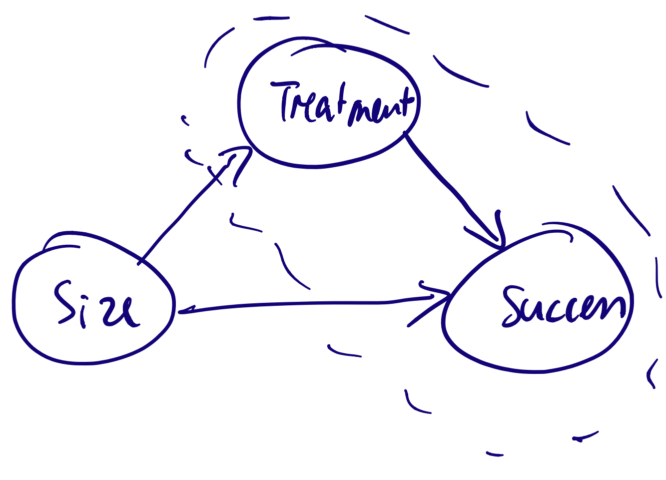
Here,



X has a <u>causal effect</u> on W and H.

But W does not have a causal effect on H. (there is no path $W \rightsquigarrow H$)

Kidney example



Confounding

We say two variables X and Y are confounded if:

$$\Pr\{Y=y \mid X=x\} \neq \Pr\{Y=y \mid do(X=x)\}$$

This corresponds to the graph structure

$$\begin{tikzcd}
 & Z \arrow[dl] \arrow[dr] & \\
X \arrow[rr] & & Y
\end{tikzcd}$$

We call Z the confounding variable.

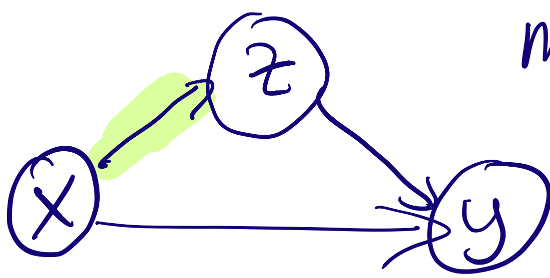
In our kidney example, size was a confounder between treatment and outcome.

In our weight example, exercise was a confounder.

To eliminate confounding, we need to hold the confounding variable constant in our analysis.

In a study this means we need to control for all possible confounders between treatment and outcome.

But we need to be careful not to control for mediators:



Mediator

- Don't control for mediators
- Do control for confounders

Causal graphs/models represent <u>assumptions</u> you make.