CAUSALITY : PART II

Recap :

Causal models

"Jupyter notebooks for generating data"

Example :

$$X := N$$
$$Z := 2X + N'$$
$$Y := (X + Z)^2$$

N, N' indep. noise

Formally :

$$X_1 := f_1(V_1, N_1)$$
$$X_2 := f_2(V_2, N_2)$$
$$\vdots$$
$$X_d := f_d(V_d, N_d)$$

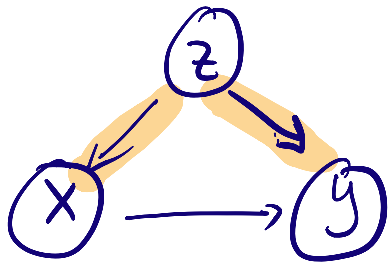
List of assignments to generate distribution $(X_1, X_2 \dots X_d)$ from independent noise variables $(N_1 \dots N_d)$

Here, $V_i \subseteq \{X_1, \dots, X_d\}$ called the parents of X_i.

$do(X_i := x) \iff$ Replace i-th assignment with $X_i := x$

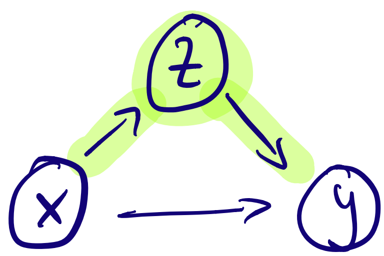
Causal graphs:

Dependence structure of causal model



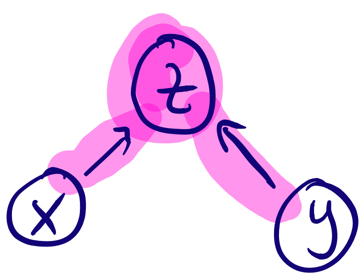
Z is called a confounder

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



Z is a mediator

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



Z is a collider

Conditioning on a collider can create (anti-)correlation between X and Y

Berkson's Law

Example:

Z hospital admission

X broken leg

Y pneumonia

Causal effect

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

Treatment effect when $X \in \{0, 1\}$

$$\mathbb{E}[Y \mid do(X := 1)] - \mathbb{E}[Y \mid do(X := 0)]$$

Fundamental question:

$\boxed{\text{When/How can we estimate causal effects from data?}}$

Equivalent: When/how can we express do-intervention with a formula that involves only conditional probabilities

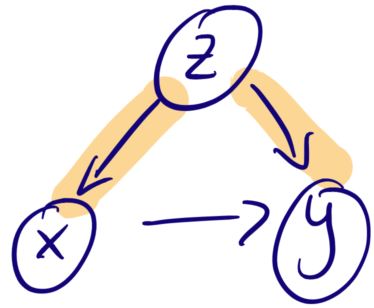
Recall: In general

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

due to confounding.

Simplest case:

Discrete rv Z



Adjustment formula

"One seperate analysis for each z"

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

averaging

$$\sum_z \Pr\{Y=y \mid X=x, Z=z\} \cdot \Pr\{Z=z\}$$

(Not to be confused with Law of Total Probability)

Proof: Note

$$\Pr\{Y=y \mid do(X:=x), Z=z\}$$
$$\overset{(*)}{=} \Pr\{Y=y \mid X=x, Z=z\}$$

Why?

Hence,

(1) law of total prob applied to model where $do(X:=x)$

$$\Pr\{Y=y \mid do(X:=x)\}$$
$$\overset{(1)}{=} \sum_z \Pr\{Y=y \mid do(X:=x), Z:=z\} \Pr\{Z=z\}$$
$$\overset{(*)}{=} \sum_z \Pr\{Y=y \mid X=x, Z=z\} \Pr\{Z=z\}$$

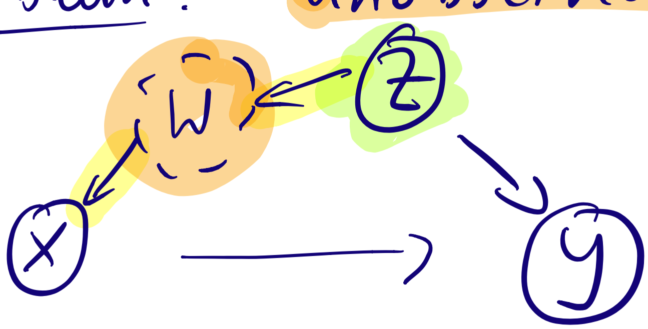
It turns out this directly generalizes to arbitrary graphs

Idea: Replace Z with parents of X

Use adjustment formula with $Z = (P_1, \dots, P_k)$.

This in principle solves all problems where all nodes are observed.

Problem: Unobserved confounding



In this case, it's enough to adjust for Z alone.

Intuitively, this **blocks** the "backdoor path" $X \leftarrow W \leftarrow Z \rightarrow Y$.

$X \leftarrow W \rightarrow Z \leftarrow Y$

On the homework we'll see the backdoor criterion which gives a general method to figure out what variables we can adjust for in the adjustment formula.

Another problem:

What do we do when the adjustment variable Z is continuous or its support is too large to have data in each slice?

Idea: Propensity scores

$$e(z) = \mathbb{E}[X \mid Z = z]$$

binary treatment $X \in \{0, 1\}$

Claim: $\mathbb{E}[Y \mid do(X:=1)] = \mathbb{E}\left[\frac{Y \cdot X}{e(Z)}\right]$ (*)

assuming the adjustment formula holds for Z, and $e(z) \neq 0$ for all z.

This is good because we can first learn a model

$$\hat{e}(z) \approx e(z)$$

from data, e.g. usin logistic regression.

We can then estimate (*) from samples (x_i, y_i, z_i) as:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{x_i y_i}{\hat{e}(z_i)} \approx \mathbb{E}\left[\frac{YX}{e(Z)}\right]$$

Inverse propensity score weighting

Proof of Claim:

Using the adjustment formula:

$$\mathbb{E}[Y \mid do(X:=1)] = \sum_y y \cdot \Pr\{Y=y \mid do(X:=1)\}$$

$$= \sum_y y \cdot \sum_z \Pr\{Y=y \mid X=1, Z=z\} \Pr\{Z=z\}$$

Multiply numerator and denominator by

$$e(z) = \Pr\{X=1 \mid Z=z\} \neq 0 :$$

$$\sum_y y \cdot \sum_z \frac{\Pr\{Y=y \mid X=1, Z=z\} \Pr\{Z=z\} \Pr\{X=1 \mid Z=z\}}{\Pr\{X=1 \mid Z=z\}}$$

$$= \sum_y \sum_z \frac{\Pr\{Y=y, X=1, Z=z\}}{\Pr\{X=1 \mid Z=z\}}$$

$$= \sum_{y,z,x\in\{0,1\}} y \cdot \mathbb{1}\{x=1\} \frac{\Pr\{Y=y, X=x, Z=z\}}{\Pr\{X=1 \mid Z=z\}}$$

$$= \mathbb{E}\left[\frac{YX}{e(Z)}\right] \qquad \blacksquare$$

Note: Same proof shows $\mathbb{E}[Y \mid do(X:=0)] = \mathbb{E}\left[\frac{Y(1-X)}{e(Z)}\right]$