Causal Inference: a Tutorial

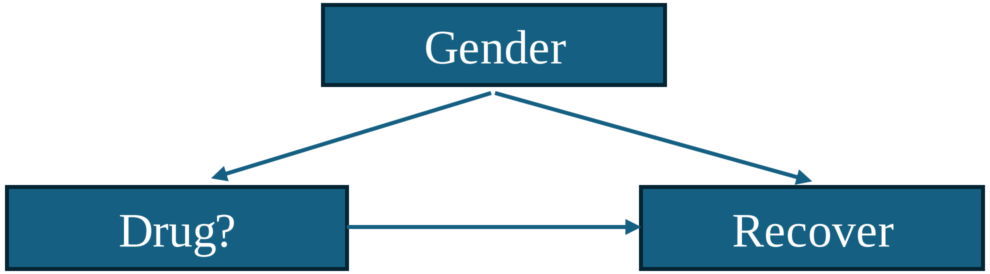
Haozhe Tian
Imperial College London
haozhe.tian21@imperial.ac.uk
https://haozhetian.netlify.app/

Table of Content

Benefit of knowing a covariate

Table 1.1 Results of a study into a new drug, with gender being taken into account

		Drug	No drug
Drug helpful	Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
	Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Drug harmful	Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)



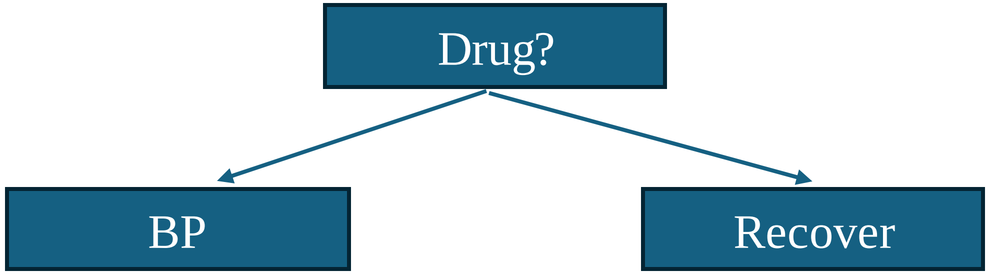
Women are more likely to take the drug, but have lower recovery rate

Should draw the conclusion from data segregated by gender.

The toxic effect of knowing a covariate

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No drug	Drug
Low BP	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High BP	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)



Drug reduces the number of High BP people

??Should draw the conclusion from data segregated by BP?? **NO!**

Structural Causal Models

SCM (Basketball Performance Based on Height and Sex)

$$V = \{\text{Height},\ \text{Sex},\ \text{Performance}\}, \quad U = \{U_1, U_2, U_3\}, \quad F = \{f1, f2\}$$

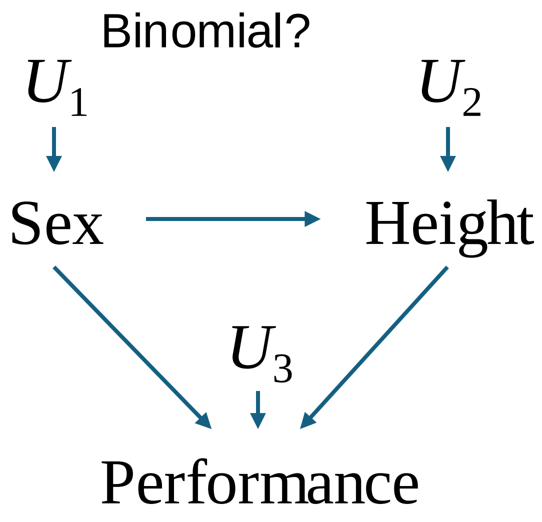
$$\text{Sex} = U_1$$

$$\text{Height} = f_1(\text{Sex},\ U_2)$$

$$\text{Performance} = f_2(\text{Height},\ \text{Sex},\ U_3)$$

V: endogenous variable

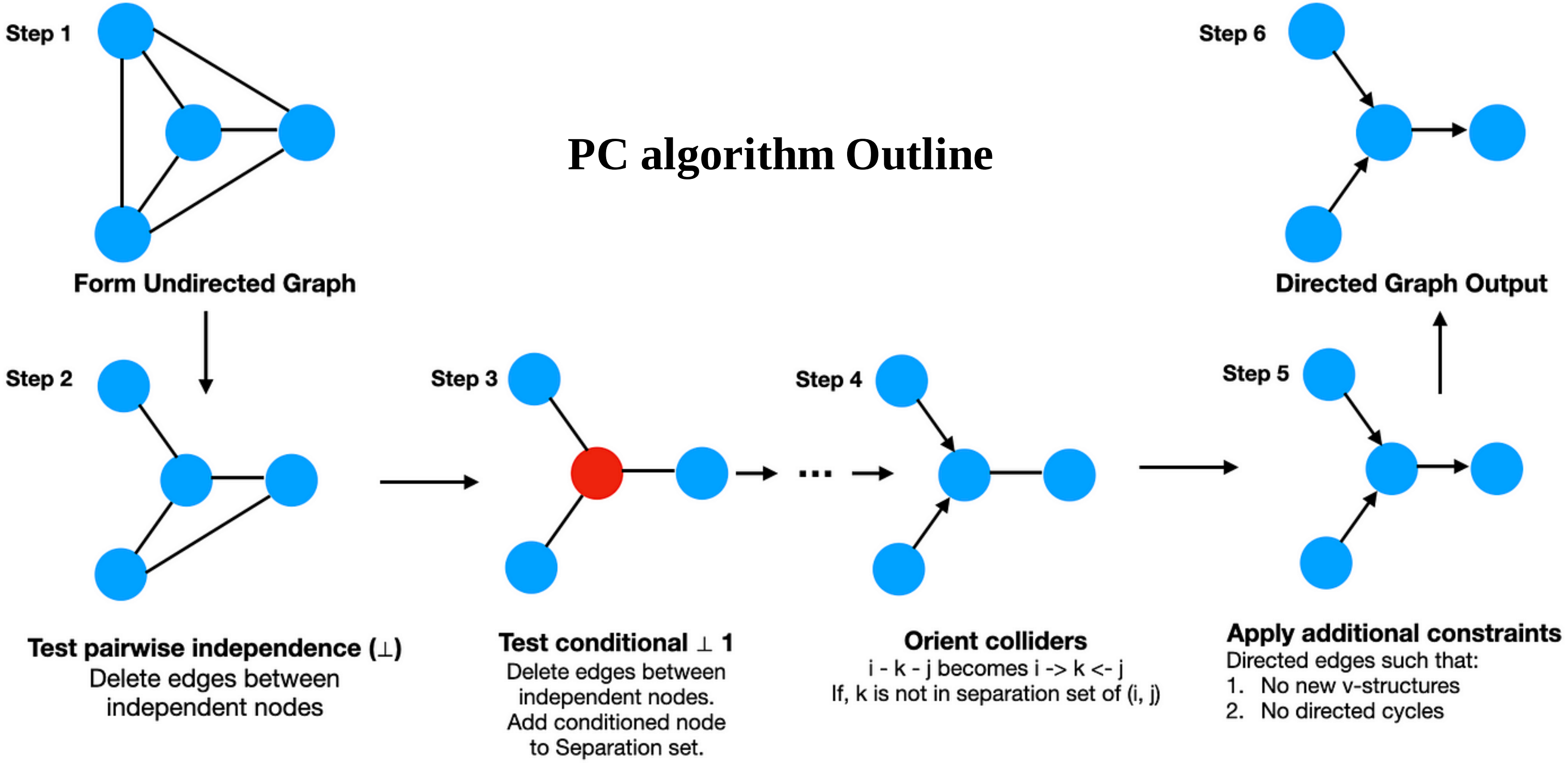
U: exogenous variable



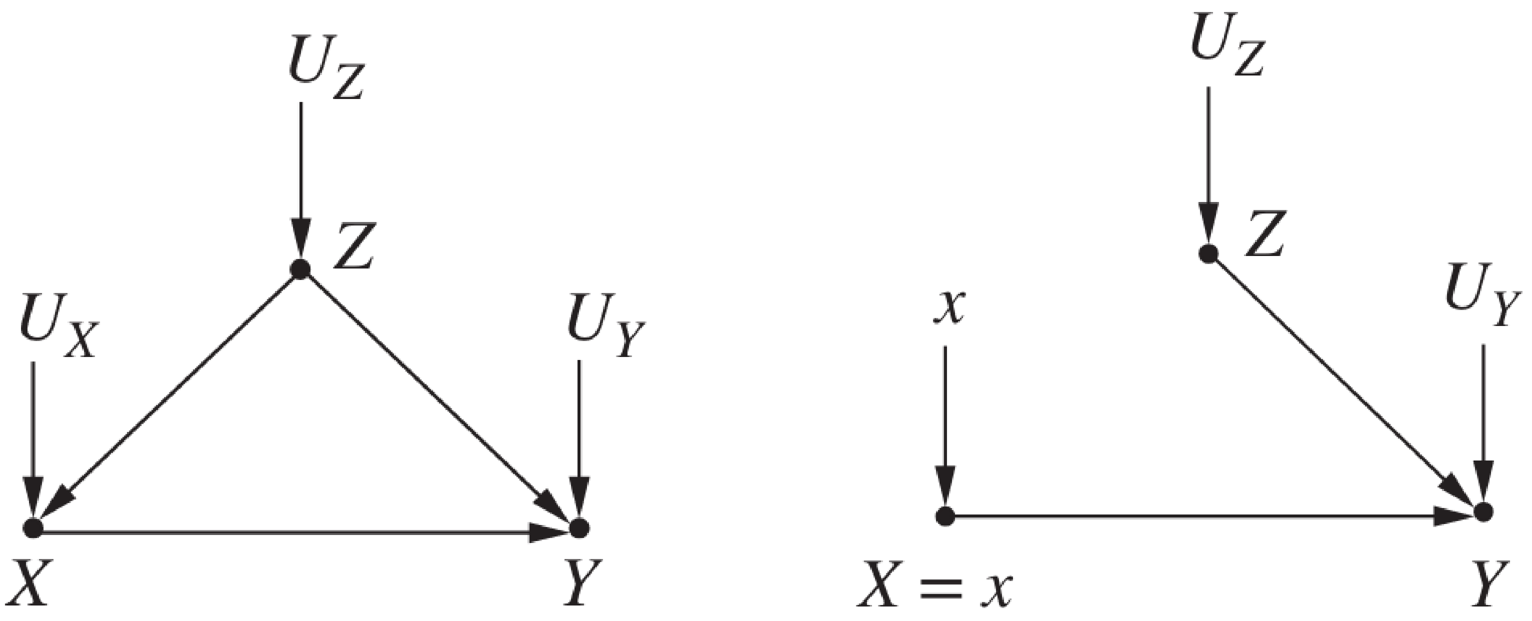
Causal Discovery

- Constraint-based Methods
  - PC algorithm
- Score-based Methods
  - GES (Greedy Equivalence Search)
- Deep Learning
  - CausalGAN
- etc.

PC algorithm Outline



Interventions



Graph before and after intervention. X is the application of drug, Y is whether the person is cured, and Z is the gender.

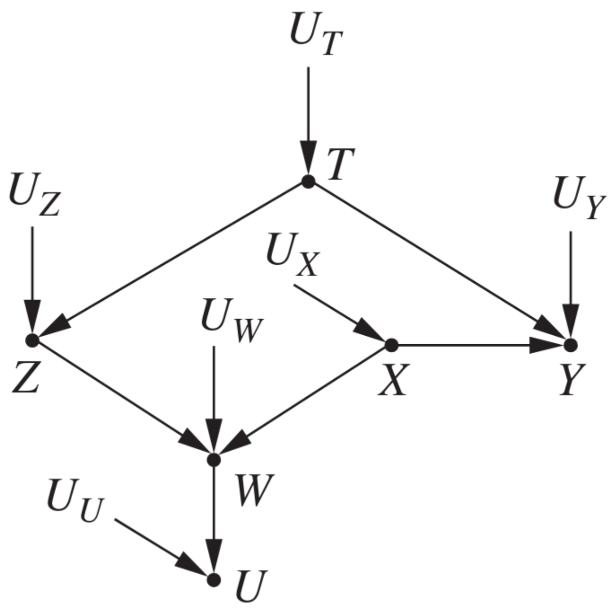
- Intervention: $P(Y=y \mid X=1)$ vs. $P(Y=y \mid do(X=1))$
- Want to know: $P(Y = 1|do(X = 1)) - P(Y = 1|do(X = 0))$
- Adjustment Formula: $P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, \boxed{Z = z)P(Z = z)}$

Backdoor Criterion

- Block all spurious paths
- Leave direct paths
- Create no new spurious paths

Interventions: Backdoor Criterion

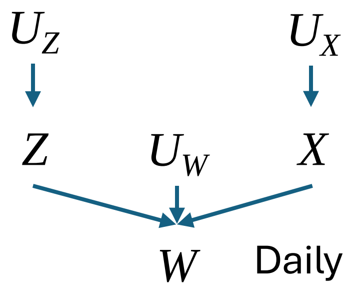
Backdoor Criterion
- Block all spurious paths between X and Y
- Leave direct paths from X to Y
- Create no new spurious paths



$P(Y{=}y \mid do(X{=}1)) = P(y \mid x)$

$P(Y{=}y \mid do(X{=}1), W{=}w)$ **?**

creates spurious path X→W←Z↔T →Y



$P(Z) = P(Z \mid X)$ **<u>but</u>** $P(Z \mid W) \neq P(Z \mid X, W)$

Interventions: Backdoor Criterion

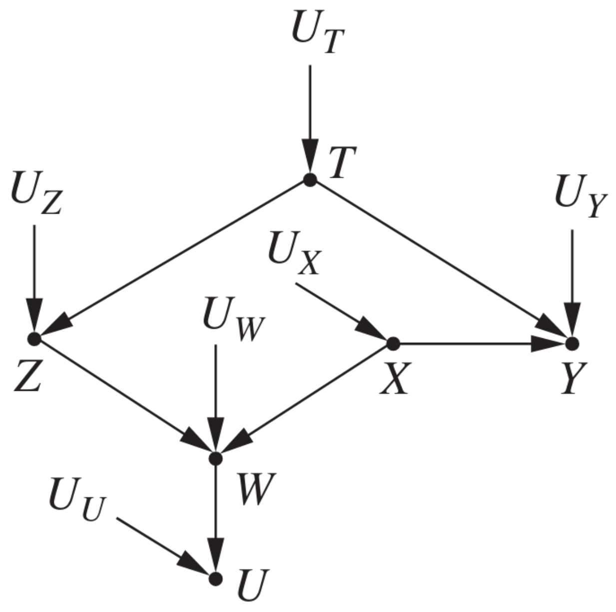
Backdoor Criterion

- Block all spurious paths between X and Y
- Leave direct paths from X to Y
- Create no new spurious paths

$P(Y{=}y \mid do(X{=}1), W{=}w)$ **?**

creates spurious path X→W←Z↔T →Y

Block spurious path by **conditioning on T**



$$P(Y=y|do(X=x), W=w) = \sum_t P(Y=y|X=x, W=w, T=t)P(T=t|X=x, W=w)$$

Interventions: Practical estimation

- Adjustment Formula: $P(Y=y|do(X=x)) = \sum_z \boxed{P(Y=y|X=x,Z=z)} P(Z=z)$

Too few data per strata when there are too many covariates

- Inverse Probability Weighing:

  * Propensity Score: $P(X{=}x \mid Z{=}z) \approx g(x, z)$

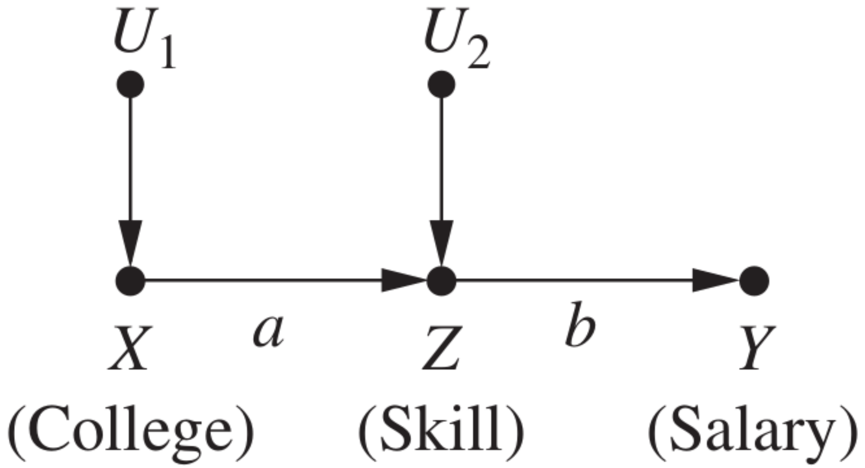
$$
\begin{aligned}
P(y|do(x)) &= \sum_z P(Y=y|X=x,Z=z)P(Z=z) \\
&= \sum_z \frac{P(Y=y|X=x,Z=z)P(X=x|Z=z)P(Z=z)}{P(X=x|Z=z)} \\
&= \sum_z \frac{P(Y=y,X=x,Z=z)}{P(X=x|Z=z)}
\end{aligned}
$$

Counterfactuals

- Intervention: $E(Y \mid do(X{=}x))$
- Counterfactual: $E(\boxed{Y_{X=1}} \mid X{=}0, \boxed{Y_{X=0}{=}y})$ **What if ?**

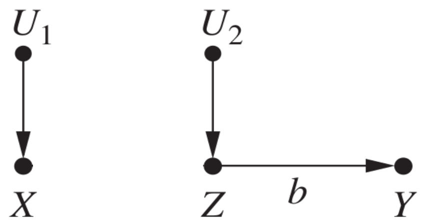
New world Original World

$X = U_1 \quad Z = aX + U_2, Y = bZ$



- $E(Y \mid do(X{=}1), Z{=}1) = E(Y \mid do(X{=}0), Z{=}1)$

For those who have skill level 1, no matter obtained in college or not

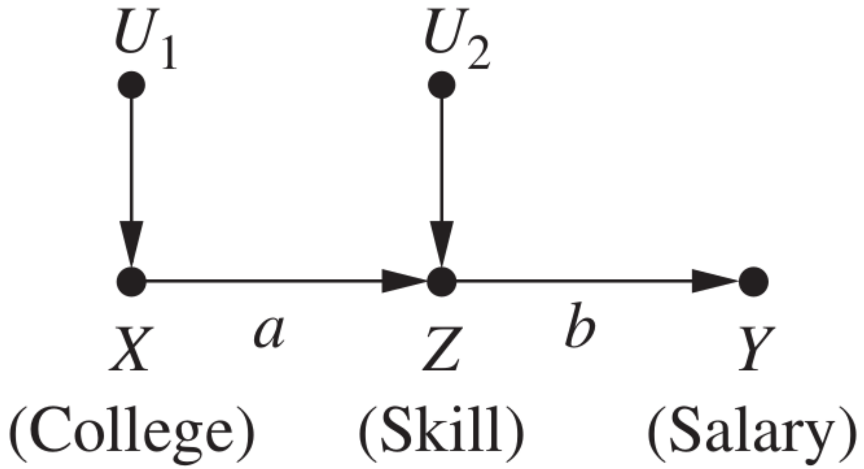


- $E(Y_{X=1} \mid Z{=}1) \neq E(Y_{X=0} \mid Z{=}1)$

For those who have skill level 1, what if they went to college? (Z will be bigger)

Counterfactuals

$X = U_1 \quad Z = aX + U_2, Y = bZ$



Three steps in computing counterfactuals:

1. **Abduction:** solve for U_i
2. **Action:** modify X
3. **Prediction:** predict Y_X

1. Abduction: Z=1, thus U_1=X=0, U_2=1.
2. Action: X=1
3. Prediction: Y=(a+1)b

$$E[Y_1 | Z = 1] = (a+1)b$$

$$E[Y_0 | Z = 1] = b$$

$$E[Y | do(X = 1), Z = 1] = b$$

$$E[Y | do(X = 0), Z = 1] = b$$

Counterfactuals: Practical Estimation

We have a powerlifting program. Let $X=1$ represent taking the program and Y represent strength after two weeks. We find:

$$E[Y|X=1] > E[Y|X=0]$$

$X=1$ people might be more interested in training to begin with. What we are interested in is actually:

$$E[Y_1|X=1] - E[Y_0|X=1]$$

*** Model parameters unknown**
*** Time travel not allowed**

$E[Y_0|X=1]$ can be estimated if we know some covariates.

$$P(Y_x = y|X = x') = \sum_z P(Y = y|X = x, Z = z)P(Z = z|X = x')$$

Thus $E[Y_0|X=1]$

$$= \sum_z E[Y|X = 0, Z = z]P(Z = z|X = 1)$$

References

- Pearl J, Glymour M, Jewell NP. Causal inference in statistics: A primer. John Wiley & Sons; 2016 Jan 25.
- Glymour C. Causal discovery. Towards Data Science [Internet]. 2020 Jun 12 [cited 2024 May 31]. Available from: https://towardsdatascience.com/causal-discovery-6858f9af6dcb.