Homework 4 — Data Science II
Causal Inference

Instructions

Please prepare a writeup addressing the following questions and submit on Brightspace. Please show all your work.

Problem 1. *Counterfactuals*. Recall the example population dataset from class:

X	Y	C_0	C_1
0	0	0	0*
0	0	0	0*
0	0	0	0*
0	0	0	0*
1	1	1*	1
1	1	1*	1
1	1	1*	1
1	1	1*	1

where asterisks denote unobserved values. In class, we showed for this example that the *average causal effect* $\theta = E[C_1] - E[C_0] = 0$ and that the *association* $\alpha = E[Y \mid X = 1] - E[Y \mid X = 0] = 1$, i.e., $\theta \neq \alpha$.

Create an example like this one in which $\alpha > 0$ and $\theta < 0$. (Include the computation of α and θ for your example.) What is the "intuition" of your example?

Problem 2. Suppose the variables X, Y and Z have the following joint distribution:

	Z = 0		Z = 1	
	Y = 0	Y = 1	Y = 0	Y = 1
X = 0	0.405	0.045	0.125	0.125
X = 1	0.045	0.005	0.125	0.125

(1) Find the conditional distribution of X and Y given $Z = 0$ and the conditional distribution of X and Y given $Z = 1$.

(2) Show that $X \perp\!\!\!\perp Y \mid Z$.

(3) Find the marginal distribution of X and Y.

(4) Show that X and Y are not marginally independent.

Problem 3. Consider the following DAG:

$$X \longrightarrow Y \longleftarrow Z$$

Prove that $X \perp\!\!\!\perp Z$ and that X and Z are dependent given Y. Use these results to interpret the meaning of this DAG. Do you know the *name* of this DAG?

Problem 4. Let $V = (X, Y, Z)$ be distributed as follows:

$$\begin{aligned} X &\sim \text{Bernoulli}\left(\frac{1}{2}\right) \\ Y \mid X = x &\sim \text{Bernoulli}\left(\frac{e^{4x-2}}{1+e^{4x-2}}\right) \\ Z \mid X = x, Y = y &\sim \text{Bernoulli}\left(\frac{e^{2(x+y)-2}}{1+e^{2(x+y)-2}}\right) \end{aligned}$$

(1) Make a diagram showing the DAG corresponding to this model. Please do not report a hand-drawn diagram.

(2) Derive a mathematical expression for $\Pr(Z = z \mid Y = y)$. What is $\Pr(Z = 1 \mid Y = 1)$?

(3) Write a program to simulate this model. Conduct simulations to compute $\Pr(Z = 1 \mid Y = 1)$ empirically. Plot this probability as a function of the simulation size N and show that it converges to the theoretical value you derived in (2).

(4) *Interventions.* Derive a mathematical expression for $\Pr(Z = 1 \mid Y := y)$. What is $\Pr(Z = 1 \mid Y := 1)$?

(5) Modify your program to simulate the intervention "fix $Y = 1$". Use simulations to compute $\Pr(Z = 1 \mid Y := 1)$. Plot this probability as a function of simulation size N and show that it converges to the theoretical value you derived in (4).

Your plots should be included in your writeup as figures (with captions). You do not need to submit your simulation code.