Causal Inference
Homework 2

- This assignment can be solved in groups of 1 up to 3 students. You must mention the name of all the participants. Note that all the students in a group will get the same grade.
- Deadline: Tuesday, April 12, 2022, 09:00 AM (No late submissions will be accepted.)
- Upload a single pdf file on Moodle.

Problem 1) Suppose $N_1, N_2 \sim Bern(0.5)$ and $N3 \sim U(\{0,1,2\})$, such that the three variables are jointly independent. That is, N_1 and N_2 have a Bernoulli distribution with parameter 0.5 and N_3 is uniformly distributed on $\{0,1,2\}$. We now define two different SEMs.

We construct the SEM $\mathcal{S}_1$ as:

$$\begin{aligned} X_1 &= N_1, \\ X_2 &= N_2, \\ X_3 &= (\mathbb{1}_{N_3>0}.X_1 + \mathbb{1}_{N_3=0}.X_2).\mathbb{1}_{X_1 \neq X_2} + N_3.\mathbb{1}_{X_1=X_2}, \end{aligned}$$

and the SEM $\mathcal{S}_2$ as:

$$\begin{aligned} X_1 &= N_1, \\ X_2 &= N_2, \\ X_3 &= (\mathbb{1}_{N_3>0}.X_1 + \mathbb{1}_{N_3=0}.X_2).\mathbb{1}_{X_1 \neq X_2} + (2 - N_3).\mathbb{1}_{X_1=X_2}, \end{aligned}$$

where $\mathbb{1}$ denotes the indicator function (e.g., $\mathbb{1}_{N_3>0}$ is 1 if $N_3 > 0$, and 0 otherwise.)

1. Using Markov factorization property, compute the observational distributions of both SEMs ($P_{\mathcal{S}_1}$ and $P_{\mathcal{S}_2}$) and show that $P_{\mathcal{S}_1} = P_{\mathcal{S}_2}$.
2. Suppose, we have seen a sample $(X_1, X_2, X_3) = (1, 0, 0)$ and we are interested in the counterfactual question "what X_3 would have been if X_1 had been 0?". Answer this counterfactual question for $\mathcal{S}_1$ and $\mathcal{S}_2$.

Problem 2) Consider the graph in Figure 1. Find all the valid parent and backdoor adjustment sets for the following pairs:

- (X_3, X_5) $(P(X_5|do(X_3)))$,
- (X_2, X_5) $(P(X_5|do(X_2)))$.

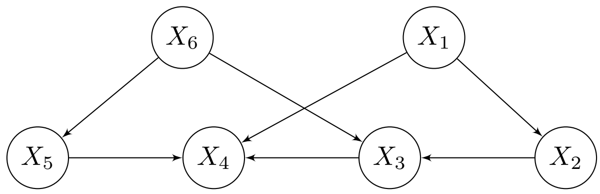

Figure 1: The graph of Problem 2.

Problem 3) The University of California, Berkeley was sued for gender discrimination over admission to graduate school in 1973. To back up this argument, the collected statistical data showed that of the 8,442 male applicants for the fall of 1973, 44 percent were admitted, whereas only 35 percent of the 4,351 female applicants were accepted. This difference is so large that makes it unlikely to be due to chance. The data of the 4 largest departments are listed in Table 1. You are hired to help the court to examine this claim through the lens of causal inference. To this end, denote by D, G and A, department, gender and admission respectively (D can take values in $\{1, 2, 3, 4\}$ for each department, G takes the values 0 for men and 1 for women, and A takes the values 0 for not admitted and 1 for admitted.) Considering the graph of Figure 2, calculate the values of $\mathbb{E}[A|do(G = 0)]$ and $\mathbb{E}[A|do(G = 1)]$. Did the graduate program make decisions in favor of men? How do you explain this phenomenon?

Department	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
1	933	64%	825	62%	108	82%
2	585	63%	560	63%	25	68%
3	918	35%	325	37%	593	34%
4	792	34%	417	33%	375	35%
Total	3,228	48%	2,127	53%	1,101	40%

Table 1: Statistics of the applicants of the 4 largest departments.

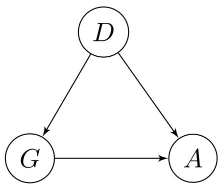

Figure 2: The graph of Problem 3.