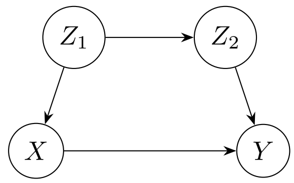
Causal and Probabilistic Graphical Models
ICS 276 (Winter 2024)
Homework 3

Due: Monday, February 26, 2024

Problem 1. Query Estimation [10 points]

Consider the following graphical model:



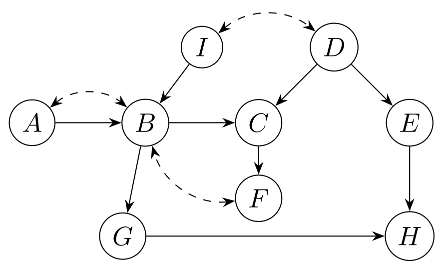
The target query is $Q = \sum_{Z_1} P(y \mid x, z_1)P(z_1)$.

(a) [5 points] Is $Q = P(y \mid x)$? Justify your answer.

(b) [5 points] Suppose that only $P(X, Y, Z_2)$ is given as input. Is Q estimable? If so, show how to do it. Otherwise, explain why is that the case.

Problem 2. d-Separation [16 points]

(a) Consider the following causal diagram:



For each case find a set that, when conditioned, on d-separates the given pair of (sets) variables

(i) [4 points] A and F

(ii) [4 points] A and C

(iii) [4 points] D and $\{F, H\}$

(iv) [4 points] I and H

Notice that Problems 4 has a (*****). This means that you are encouraged to do this question, but I will not grade this closely.

Problem 3. Modeling [5 points]

Consider the recent study of the connection between sleep quality and dementia presented here and discussed in class. https://www.nytimes.com/2021/04/20/health/sleep-dementia-risk.html?referringSource=articleShare.

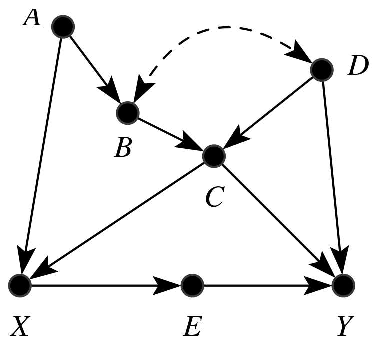
(a) [2 points] Provide a structural causal diagram based on your understanding of the study and assumptions made.

(b) [1 point] Discuss the suitability of the different conclusions proposed by the study. You can focus on one or two statements.

(c) [2 points] The article talks about associations rather than causation. In your opinion, can they claim causation and under what assumptions.

Problem 4. Understanding the Model's Granularity [10 points]

Consider the causal diagram G below.



(a) [1 point] Determine whether the causal effect $P(y \mid do(x))$ is identifiable from G and $P(\mathbf{V})$, where $\mathbf{V}$ is the set of endogenous variables. If so, show how; otherwise, provide a counter-example.

(b) [2 points] Write an SCM that induces G and a probability distribution $P(\mathbf{V})$, with $P(\mathbf{v}) > 0$ for every $\mathbf{v}$. You don't need to show $P(\mathbf{V})$ in your answer.

Suppose that the same system (represented by the SCM) is investigated in another study. However, in this case, only the variables $\mathbf{V}' = \{X, Y, B, C\}$ are measured.

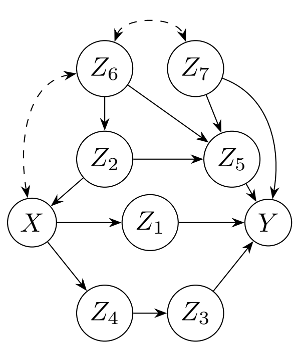
(c) [3 points] Write a new SCM $M' = \langle \mathbf{V}', \mathbf{U}', \mathcal{F}', P(\mathbf{u}') \rangle$ corresponding to this different cut of reality, consistent with your answer to the previous question (i.e., departing from SCM written in (b)).

(d) [1 point] Draw the causal diagram G' induced by the SCM M'.

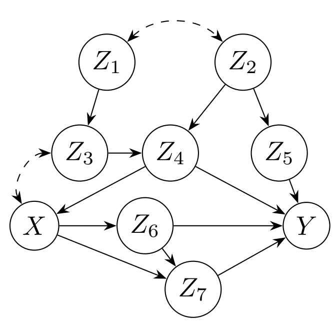
(e) [3 points] Is the effect $P(y \mid do(x))$ identifiable from $P(\mathbf{V}')$ and G'? Is there a back-door or front-door adjustment? Can it be solved with do-calculus?

Problem 5. Optimal Experiment Design [10 points]

An advertisement company is trying to identify the effect of a new campaign X on the click through rate Y. They have two hypotheses about how the strategy relates to a possibly measured set of covariates $\mathbf{Z}$. The hypotheses are represented in the causal diagrams (a) and (b) shown below:



(a)



(b)

Variable	**Cost**
X	2
Y	1
Z_1	4
Z_2	2
Z_3	4
Z_4	5
Z_5	5
Z_6	2
Z_7	1

(c)

(a) [4 points] If it exists, find a minimal admissible backdoor set for adjustment in each of the graphs.

(b) [6 points] The company wants to minimize the measurement cost for identifying $P(y \mid do(x))$. Find the minimum cost ID expression based on the table (c) and justify your answer.

Problem 6. Back-door Adjustment as a Substitute for the Direct Parents [1 point]

The causal effect of the intervention $do(X = x)$ on a variable Y can be identified if all parents of X are observed and is given by

$$P(y \mid do(x)) = \sum_{pa_X} P(y \mid x, pa_X)P(pa_X). \tag{1}$$

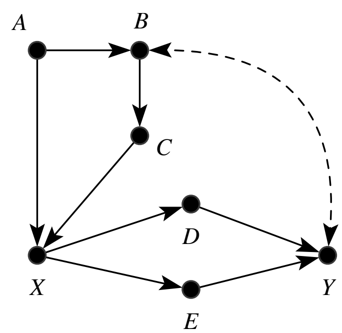
Based on this result, prove that if a set $\mathbf{Z}$ satisfies the back-door criterion relative to X and Y in the graph , it follows that

$$P(y \mid do(x)) = \sum_{\mathbf{z}} P(y \mid x, \mathbf{z})P(\mathbf{z}). \tag{2}$$

This question is asking you to leverage eq3.1 to prove the backdoor identification formula in eq3.2.

Problem 7. Many Paths Lead to ID [10 points]

Consider the following causal diagram.



Give **three** different functions of the observational distribution $P(\mathbf{V})$ that are equal to the effect $P(y \mid do(x))$. At least one answer should correspond to a front-door case and one to a back-door case. Justify each one of the expression showing its do-calculus derivation.