Causal Inference for Health Data (Winter 2026)
STATS C160/C260
Homework 2

Exercise 1. SCMs and the Truncated Product

Consider the structural causal model below[1], where $a, b \in (0, 1)$ are parameters.

$$\mathbf{V} = \{Z, X, Y\}$$
$$\mathbf{U} = \{U_x, U_y, U_z, U_{zy}\}$$
$$\mathcal{F} = \begin{cases} f_Z = U_z \oplus U_{zy} \\ f_X = U_x \oplus Z \\ f_Y = (U_y \wedge (X \oplus U_{zy})) \vee (\neg U_y \wedge (X \oplus \neg U_{zy})) \end{cases}$$
$$P(\mathbf{u}) = \{P(U_z = 1) = a, P(U_x = 1) = b, P(U_y = 1) = 1/5, P(U_{zy} = 1) = 3/4\}$$

(a) Draw the corresponding causal diagram.

(b) Write the following queries as functions of a and b:

   (a) $P(Y = 1)$
   (b) $P(Y = 1 \mid X = 1)$
   (c) $P(Y = 1 \mid X = 1, Z = 1)$
   (d) $P(Y = 1 \mid do(X = 1)) - P(Y = 1 \mid do(X = 0))$

(c) If any of the probability distributions in the previous question are independent of a or b, explain if that independence can be inferred from the causal diagram alone (not knowing the actual function, just their observable and unobservable arguments). Your answer should address each query.

Exercise 2. Applying the Backdoor Criterion

Consider the following definition:
Definition 1 (Minimal adjustment set)**.** A minimal adjustment set $\mathbf{Z}$ relative to a pair of variables X and Y is a set of variables that satisfies the back-door criterion to find the causal effect of X on Y, such that no proper subset of $\mathbf{Z}$ satisfies the criterion.

Consider SCMs compatible with the graphs in Fig. 1. Find a *minimal* adjustment set to compute the causal effect of X on Y and write the corresponding expression for $P(y \mid do(x))$.

[1] $\neg, \wedge, \vee, \oplus$ represents the logical operators negation, and, or, xor respectively.

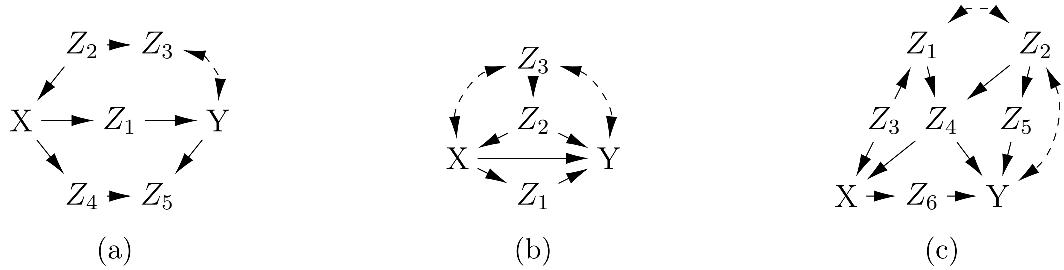


Figure 1: Graphs for the SCM used in this problem.

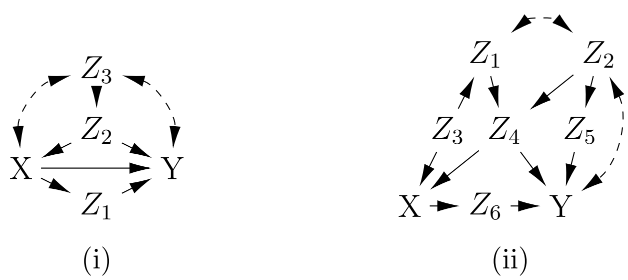
Exercise 3. Back-door Adjustment – I

(a) Construct a causal diagram $\mathcal{G}$ such that a set $\mathbf{Z}$ does not satisfy the Back-door criterion relative to (X, Y), but the adjustment formula with the same $\mathbf{Z}$ still holds

$$P(y \mid do(x)) = \sum_{z} P(y \mid x, z)P(z)$$

for every SCM $\mathcal{M}$ with causal diagram $\mathcal{G}$). For the constructed example, prove that that the adjustment formula holds using do-calculus.

(b) Consider the following causal diagrams. For each case find a *minimal* back-door admissible (if one exists) set to compute the causal effect of X on Y, and write the corresponding expression for $P(y \mid do(x))$.



(i) $\mathbf{Z} = \{ \quad \}$

$P(y \mid do(x)) =$

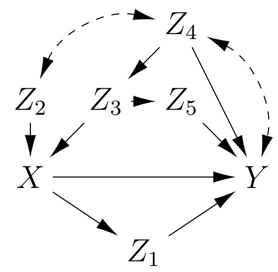
(ii) $\mathbf{Z} = \{ \quad \}$

$P(y \mid do(x)) =$

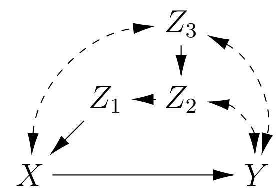
(c) Prove the Back-door criterion using do-calculus.

Exercise 4. Back-door Adjustment – II

(a) Consider the following causal diagrams. For each case find a *minimal* back-door admissible (if one exists) set to compute the causal effect of X on Y, and write the corresponding expression for $P(y \mid do(x))$.



(i)



(ii)

(i) $\mathbf{Z} = \{ \quad \}$

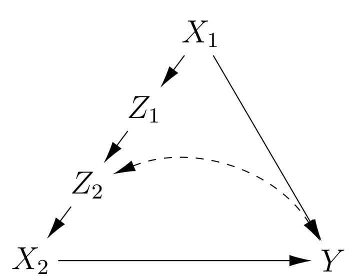
$P(y \mid do(x)) =$

(ii) $\mathbf{Z} = \{ \quad \}$

$P(y \mid do(x)) =$

(b) Consider the causal diagram $\mathcal{G}$ below. The goal is to identify the effect of $\mathbf{X} = \{X_1, X_2\}$ on Y. First, determine whether there exists a set $\mathbf{Z}$ that satisfies the back-door criterion. Second, using the rules of do-calculus, show that

$$P(y \mid do(x_1, x_2)) = \sum_{z_1, z_2} P(y \mid x_1, x_2, z_1, z_2) P(z_1, z_2).$$



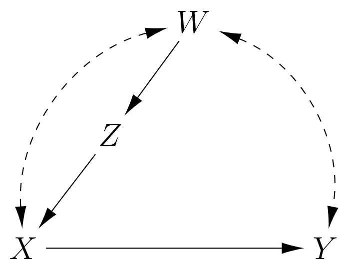
Exercise 5. Do-Calculus Rules

(a) For each of the rules of Do-Calculus, provide an example of a (invalid) use of the rule where the corresponding, licensing separation in the graph does not hold.

Specifically, provide a fully specified SCM and compute the left and right hand side of the equation in the rule in each case, showing that they are different for some particular value of the variables.

You are allowed to construct three different SCMs or re-use them throughout the rules.

(b) Consider the following causal diagram:

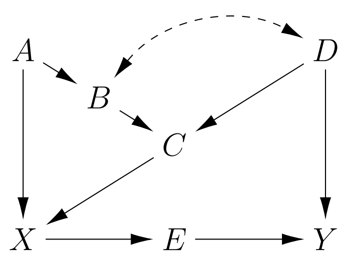


(i) Explain why neither Z, W, nor $\{Z, W\}$ are back-door admissible sets for the causal effect of X on Y.

(ii) Derive the identification expression for $P(y \mid do(x))$ using the rules of do-calculus. *Hint: in the first step, justify adding $do(w), do(z)$ to $P(y \mid do(x))$. Then, note that $do(x)$ in the $do(w), do(z)$ is the same as $see(x)$. Finally, use the law of conditional probability to compute the numerator and the denominator.*

Exercise 6. Changing the Granularity of the Model

Consider the causal diagram $\mathcal{G}$ below.



(a) Determine whether the causal effect $P(y \mid do(x))$ is identifiable; if so, show how.

(b) Write a SCM $\mathcal{M}$ that induces this causal diagram and a probability distribution $P(\mathbf{V})$ such that for every $\mathbf{v}, P(\mathbf{v}) > 0$. You don't need to show $P(\mathbf{v})$ in your answer.

Suppose that in a different study the same system (represented by the SCM $\mathcal{M}$) is observed, but only the variables $\mathbf{V}' = \{X, Y, B, C\}$ are measured.

(c) Write a new SCM $\mathcal{M}' = \langle \mathbf{V}', \mathbf{U}', \mathcal{F}', P(\mathbf{u}') \rangle$ corresponding to the model for this study, based on your answer to the previous question.

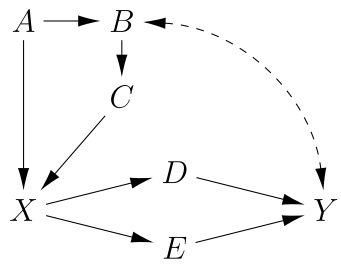
(d) Draw the causal diagram $\mathcal{G}'$ corresponding to $\mathcal{M}'$.

Naturally, since we are talking about the same underlying reality (the SCM you proposed), $\mathcal{M}'$ and $\mathbf{V}'$ have to preserve all the (probabilistic) independencies among the variables in $\mathbf{V}'$ that were implied in $\mathcal{M}$ and $\mathbf{V}$.

(e) Is the effect $P(y \mid do(x))$ identifiable from $P(\mathbf{v}')$ and $\mathcal{G}'$? Specifically, is there a back-door or front-door adjustment set?

Exercise 7. Many Paths Lead to ID

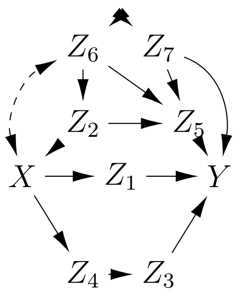
The target causal quantity $P(y \mid do(x))$ may not be identifiable from $P(\mathbf{V})$ depending on the causal diagram. On the other hand, for identifiable effects, there could be more than one expression that is equal to the effect of interest. Consider the following causal diagram:



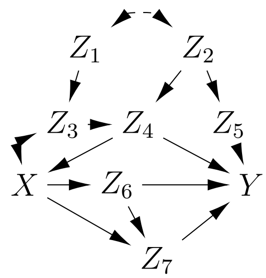
Give 5 different functions of the observational distribution $P(\mathbf{v})$ that are equal to the effect Q. That is, find 5 different identfication expressions. Justify the validity of each expression using the Back-door Criterion, the Front-door criterion, or showing a derivation in do-calculus.

Exercise 8. [Optional] Optimal Experimental Design

An advertisement company aims to predict 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 below (a,b):



(a)



(b)

Variable	**Cost**
X	2
Y	1
Z_1	3
Z_2	1
Z_3	3
Z_4	1
Z_5	6
Z_6	2
Z_7	2

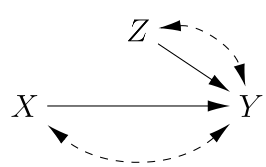
(c)

(a) If it exists, find a minimal admissible set for adjustment in each of the graphs.

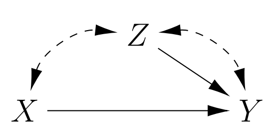
(b) The company wants to minimize the measurement cost for identifying $P(y \mid do(x))$. Find the minimum cost ID expression based on the table (c) and justify your selection.

Exercise 9. [Optional] Non-Identifiability

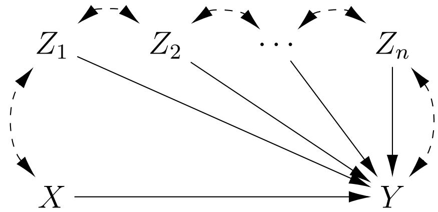
(a) Prove that the effect $P(y \mid do(x))$ is not identifiable from the following causal diagram and $P(X, Y, Z)$, where the domain of X, Y, Z is $\{0, 1\}$ and $P(x, y, z) > 0, \forall x, y, z \in \{0, 1\}$.



(b) Prove that the effect $P(y \mid do(x))$ is not identifiable from the following causal diagram and $P(X, Y, Z)$, where the domain of X, Y, Z is $\{0, 1\}$ and $P(x, y, z) > 0, \forall x, y, z \in \{0, 1\}$.



(c) Prove that the effect $P(y \mid do(x))$ is not identifiable from the following causal diagram and $P(X, Y, Z_1, \ldots, Z_n)$, $1 \leq n < \infty$, where the domain of $X, Y, Z_1, \ldots, Z_n$ is $\{0, 1\}$ and $P(x, y, z_1, \ldots, z_n) > 0, \forall x, y, z_1, \ldots, z_n \in \{0, 1\}$. *Note: We discussed in class different strategies for constructing such counter-examples. You are allowed to leverage this understanding but need to construct your own pair of models.*


Z_1
Z_2
$\cdots$
Z_n
X
Y