Causal Inference for Health Data (Winter 2026)
STATS C160/C260
Homework 3 Solutions

Exercise 1. Counterfactual Quantities

Consider an SCM $\mathcal{M}$ below.

$$\mathcal{M}^* = \begin{cases} \mathbf{U} & = \{U_1, U_2, U_3\}, \text{ all binary} \\ \mathbf{V} & = \{X, W, Y\} \\ \mathcal{F} & = \begin{cases} f_X(u_1) & = u_1 \\ f_W(x, u_1, u_2) & = (u_1 \wedge u_2) \oplus (\neg u_2 \wedge \neg x) \\ f_Y(x, w, u_3) & = (x \wedge w) \oplus u_3 \end{cases} \\ P(\mathbf{U}) & \text{defined such that } P(U_1 = 0) = 0.4, P(U_2 = 0) = 0.2, P(U_3 = 0) = 0.6, \end{cases}$$

We let X be the treatment (whose effect we seek to assess), Y be the outcome variable and W for all the observed intermediate variables between X and Y (called mediators);

(a) Calculate the total variation $P(Y = 1 \mid X = 1)$ and the total effect $P(Y = 1 \mid do(X = 1))$.

We can derive a probability table:

$P(\mathbf{U})$	U_1	U_2	U_3	X	W	Y	$W_{X=0}$	$Y_{X=0}$	$W_{X=1}$	$Y_{X=1}$	$Y_{X=1,W_{X=0}}$	$Y_{X=0,W_{X=1}}$
$p_0 = 0.048$	0	0	0	0	1	0	1	0	0	0	1	0
$p_1 = 0.032$	0	0	1	0	1	1	1	1	0	1	0	1
$p_2 = 0.192$	0	1	0	0	0	0	0	0	0	0	0	0
$p_3 = 0.128$	0	1	1	0	0	1	0	1	0	1	1	1
$p_4 = 0.072$	1	0	0	1	0	0	1	0	0	0	1	0
$p_5 = 0.048$	1	0	1	1	0	1	1	1	0	1	0	1
$p_6 = 0.288$	1	1	0	1	1	1	1	0	1	1	1	0
$p_7 = 0.192$	1	1	1	1	1	0	1	1	1	0	0	1

Thus, $P(Y = 1 \mid X = 1) = (p_5 + p_6)/(p_4 + p_5 + p_6 + p_7) = 0.56$. And $P(Y = 1 \mid do(X = 1)) = p_1 + p_3 + p_5 + p_6 = 0.496$.

(b) Calculate $ETT_{x,x'}(y)$ where $y = 1$, $x = 1$ and $x' = 0$.

From the definition of ETT, we have

$$ETT_{x,x'}(y) = P(Y_{X=1} = 1 \mid X = 0) \tag{1}$$

The evidence $X = 0$ implies $U_1 = 0$ from the probability table in (a). Then, $ETT_{x,x'}(y) = (p_1 + p_3)/(p_0 + p_1 + p_2 + p_3) = 0.4$.

(c) Calculate $PN/PS_{(x,y)(x',y')}(X; Y)$ where $x = 1, y = 0$ and $x' = 0, y' = 1$.

From the definition of PN/PS, we have

$$PN/PS_{(x,y)(x',y')}(X;Y) = P(Y_{X=1} = 0 \mid X = 0, Y = 1) \tag{2}$$

The evidence implies $U_1 = 0, U_3 = 1$. Then, $PN/PS_{(x,y)(x',y')}(X;Y) = 0$.

(d) Calculate the $NDE_{x_0,x_1}(y)$ and $NIE_{x_0,x_1}(y)$ where $x_0 = 0$, $x_1 = 1$, and $y = 1$.

From the definition of NDE and NIE, we have

$$\begin{aligned} NDE_{x_0,x_1}(y) &= P(Y_{X=1,W_{X=0}} = 1) - P(Y_{X=0} = 1) \\ NIE_{x_0,x_1}(y) &= P(Y_{X=0,W_{X=1}} = 1) - P(Y_{X=0} = 1) \end{aligned} \tag{3}$$

Then, $NDE_{x_0,x_1}(y) = (p_0 + p_3 + p_4 + p_6) - (p_1 + p_3 + p_5 + p_7) = 0.136$.
$NIE_{x_0,x_1}(y) = (p_1 + p_3 + p_5 + p_7) - (p_1 + p_3 + p_5 + p_7) = 0$.

(e) Calculate the counterfactual quantity $DE_{x_0,x_1}(y \mid x) := P(y_{x_1,W_{x_0}} \mid x) - P(y_{x_0}|x)$ where $x_0 = 0, x_1 = 1, x = 1, y = 1$.

The evidence $x = 1$, implies that $U_1 = 1$. Then, $DE_{x_0,x_1}(y \mid x) = (p_4 + p_6)/(p_4 + p_5 + p_6 + p_7) - (p_5 + p_7)/(p_4 + p_5 + p_6 + p_7) = 0.2$

Exercise 2. Counterfactual Evaluation

The team of data scientists is tasked with evaluating the psychological effects of a new health program. Consider the variable X, which represents whether the individual signed up for the gym membership in month 1 ($X = 1$ represents sign-up), Z is the individual's body mass index (BMI) after month 6 ($Z = 1$ represents a healthy BMI), and Y whether the individual ranks above 5 on a mood-scale after month 9 ($Y = 1$ represents a positive mood).

The true, underlying SCM is given as follows: Consider an SCM $\mathcal{M}$ below.

$$\mathcal{M}^* = \begin{cases} \mathbf{U} & = \{U_1, U_2, U_3, U_4, U_5\}, \text{ all binary} \\ \mathbf{V} & = \{X, Z, Y\} \\ \mathcal{F} & = \begin{cases} X & \leftarrow u_1 \wedge u_2 \\ Z & \leftarrow (u_1 \oplus u_3) \wedge ((x \vee u_2) \oplus u_4) \\ Y & \leftarrow (z \vee (\neg u_2) \vee (\neg x)) \oplus u_5 \end{cases} \\ P(\mathbf{U}) & \text{ defined such that } P(U_1 = 1) = 0.5, \\ & P(U_2 = 1) = 0.5, P(U_3 = 1) = 0.7, P(U_4 = 1) = 0.4, P(U_5 = 1) = 0.2 \end{cases}$$

(a) First, write the counterfactual quantity that given an individual who signed up for gym membership, this same individual would not have reported a positive mood had they not reported a healthy BMI. Second, evaluate this quantity directly from $\mathcal{M}$.

The corresponding quantity is $P(Y_{Z=0} = 0 \mid X = 1) = 0.8$.

(b) The team is asked about the necessity and sufficiency of X regarding Z.

(i) Write and evaluate the quantity that describes how much the absence of X is necessary to make $Z = 0$ given an individual who signed up for gym membership and reported a healthy BMI.

The corresponding quantity is $P(Z_{X=0} = 0 \mid X = 1, Z = 1) = 0$.

(ii) Write and evaluate the quantity that describes how much the presence of X is sufficient to make $Z = 1$ given an individual who did not sign up for gym membership and reported an unhealthy BMI.

The corresponding quantity is $P(Z_{X=1} = 1 \mid X = 0, Z = 0) = 0.275$.

(c) The team is further asked how sign-up for the gym, directly and indirectly, affects the mood of the population (Y).

(i) Write in counterfactual notation the full expression for the natural direct effect, written $\text{NDE}_{X=0,X=1}(Y = 1)$, and evaluate it directly from $\mathcal{M}$.

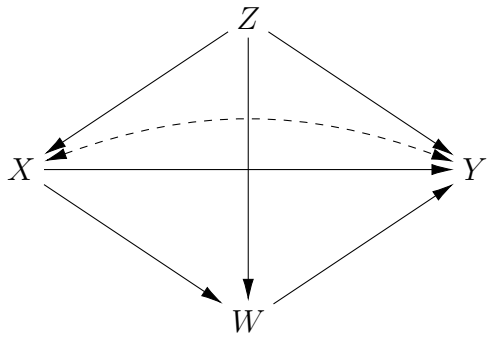
The corresponding quantity is $P(Y_{X=1,Z_{X=0}} = 1) - P(Y_{X=0} = 1) = -0.21$.

(ii) Write in counterfactual notation the full expression for the natural indirect effect, written $\text{NIE}_{X=0,X=1}(Y = 1)$, and evaluate it directly from $\mathcal{M}$.

The corresponding quantity is $P(Y_{X=0,Z_{X=1}} = 1) - P(Y_{X=0,Z_{X=0}} = 1)$, and the same is equal to zero since both factors simplify to $P(Y_{X=0})$.

Exercise 3. Network Construction

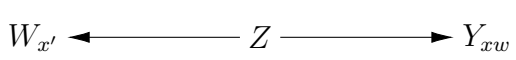
Consider the following graph.



(a) What is $An(Y_{xw}, W_{x'}, Z)$?

$An(Y_{xw} = \{Y_{xw}, Z\}$, $An(W_{x'}) = \{W_{x'}, Z\}$, $An(Z) = \{Z\}$.
Thus, $An(Y_{xw}, W_{x'}, Z) = \{Y_{xw}, W_{x'}, Z\}$.

(b) Draw the ancestral multi-world network to check $Y_{xw} \perp W_{x'} \mid Z$.



Thus, $Y_{xw} \perp W_{x'} \mid Z$.

(c) Let's consider deriving the target query $P(y_{x,W_{x'}})$ given $P(\mathbf{V})$, $P(\mathbf{V} \mid do(x))$ and $P(\mathbf{V} \mid do(x'))$ and the causal diagram.

(i) Unnest the target query $P(y_{x,W_{x'}})$.

Apply the counterfactual unnesting theorem, we have

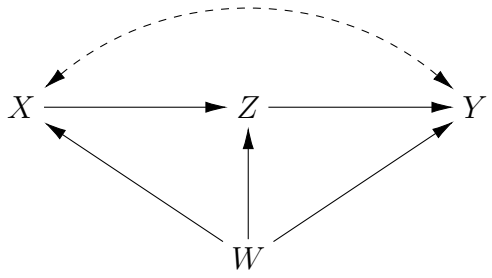
$$P(y_{x,W_{x'}}) = \sum_{w} P(y_{xw}, w_{x'}) \tag{4}$$

(ii) [Optional] Derive an expression for the target query $P(y_{x,W_{x'}})$. Suppose you have acces to both observational data $P(\mathbf{V})$, and also interventional data $P(\mathbf{V} \mid do(x))$, $P(\mathbf{V} \mid do(x'))$. Can the causal query be computed? *Hint: Use the results from (a)-(b).*

$$\begin{aligned}
P(y_{x,W_{x'}}) &= \sum_{w} P(y_{xw}, w_{x'}) && \text{Unnesting} && (5)\\
&= \sum_{w,z} P(y_{xw}, w_{x'} \mid z)P(z) && \text{Conditioing on } Z && (6)\\
&= \sum_{w,z} P(y_{xw} \mid z)P(w_{x'} \mid z)P(z) && Y_{xw} \perp W_{x'} \mid Z && (7)\\
&= \sum_{w,z} P(y_{xw} \mid z)P(w_{x'} \mid z, x')P(z) && W_{x'} \perp X \mid Z && (8)\\
&= \sum_{w,z} P(y_{xw} \mid z)P(w \mid z, x')P(z) && \text{Consistency} && (9)\\
&= \sum_{w,z} P(y_{xw} \mid z_x)P(w \mid z, x')P(z) && \text{Rule 3: } \{X\} \cap An(Z) = \emptyset && (10)\\
&= \sum_{w,z} P(y_{xw} \mid z_x, w_x)P(w \mid z, x')P(z) && Y_{xw} \perp W_x \mid Z && (11)\\
&= \sum_{w,z} P(y \mid z, w, do(x))P(w \mid z, x')P(z) && \text{Consistency} && (12)
\end{aligned}$$

Exercise 4. Counterfactual Identification

Consider the following graph.



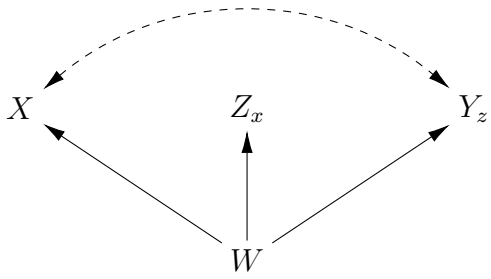
(a) Compute the following counterfactual ancestors:

(a) $An(Y_z)$
(b) $An(Z_x)$
(c) $An(W)$
(d) $An(X)$

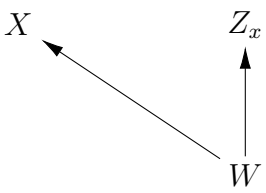
$An(Y_z) = \{Y_z, W\}$, $An(Z_x) = \{W, Z_x\}$, $An(W) = \{W\}, An(X) = \{W, X\}$.

(b) Compute the ancestral multi-world network to evaluate the following counterfactual independencies:

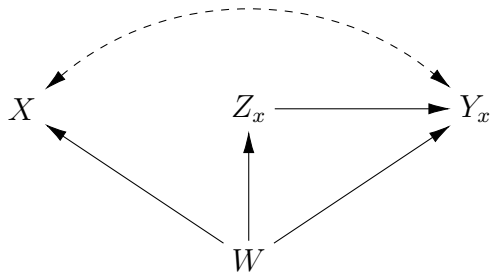
(a) $Y_z \perp Z_x \mid W, X$;
(b) $Z_x \perp X \mid W$;
(c) $Y_x \perp X \mid W$.



Thus, $Y_z \perp Z_x \mid W, X$.



Thus, $Z_x \perp X \mid W$.



Thus, $Y_x \not\perp\!\!\!\perp X \mid W$.

(c) Using consistency, exclusion, and independence rules, decide whether the query $P(y_x \mid x')$ is identifiable from $P(\mathbf{V})$ and the causal diagram above. If identifiable, provide the derivation step by step.

$$P(y_x \mid x') \tag{13}$$

$$= \sum_{w,z} P(y_x \mid z_x, w, x') P(z_x \mid w, x') P(w \mid x') \qquad \text{Conditioning on } Z_x \text{ and } W \tag{14}$$

$$= \sum_{w,z} P(y_{xz} \mid z_x, w, x') P(z_x \mid w, x') P(w \mid x') \quad \text{ctf-calculus rule 1: } Z_x = z \Rightarrow Y_x = Y_{xz} \tag{15}$$

$$= \sum_{w,z} P(y_z \mid z_x, w, x') P(z_x \mid w, x') P(w \mid x') \qquad \text{ctf-calculus rule 3: } X \cap An(Y) \text{ in } \mathcal{G}_{\overline{Z}} \tag{16}$$

$$= \sum_{w,z} P(y_z \mid w, x') P(z_x \mid w, x') P(w \mid x') \qquad \text{ctf-calculus rule 2: } Y_z \perp Z_x \mid W, X \tag{17}$$

$$= \sum_{w,z} P(y_z \mid z_{x'}, w, x') P(z_x \mid w, x') P(w \mid x') \qquad \text{ctf-calculus rule 2: } Y_z \perp Z_{x'} \mid W, X \tag{18}$$

$$= \sum_{w,z} P(y_z \mid z, w, x') P(z_x \mid w, x') P(w \mid x') \qquad \text{ctf-calculus rule 1: } X = x' \Rightarrow Z_{x'} = Z \tag{19}$$

$$= \sum_{w,z} P(y \mid z, w, x') P(z_x \mid w, x') P(w \mid x') \qquad \text{ctf-calculus rule 1: } Z = z \Rightarrow Y_z = Y \tag{20}$$

$$= \sum_{w,z} P(y \mid z, w, x') P(z_x \mid w, x) P(w \mid x') \qquad \text{ctf-calculus rule 2: } Z_x \perp X \mid W \tag{21}$$

$$= \sum_{w,z} P(y \mid z, w, x') P(z \mid w, x) P(w \mid x') \qquad \text{ctf-calculus rule 1: } X = x \Rightarrow Z_x = Z \tag{22}$$

Exercise 5. Counterfactual

(a) Consider the following SCM $M = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$:

$$\mathbf{V} = \{X, Y, Z, W\}$$
$$\mathbf{U} = \{U_1, U_2, U_3\}$$
$$\mathcal{F} = \begin{cases} f_Z = U_1 \oplus U_2, \\ f_W = Z \oplus U_3, \\ f_X = U_1 \oplus W, \\ f_Y = X \oplus U_2, \end{cases}$$
$$P(\mathbf{u}) = \left\{ P(U_1 = 1) = \frac{1}{2}, P(U_2 = 1) = \frac{1}{10}, P(U_3 = 1) = \frac{1}{4} \right\}$$

Compute following counterfactual queries using the three-step algorithm consisting of abduction, action and prediction.

(i) $P(Y_{x=0} = 1)$.

$$P(Y_{x=0} = 1) = P(0 \oplus U_2 = 1) = P(U_2 = 1) = \frac{1}{10}.$$

(ii) $P(Y_{x=0} = 1 | X = 1)$.

Given $X = 1$, we have

$$P(U_2 = 1 \mid X = 1) = \frac{P(U_2 = 1, X = 1)}{P(X = 1)} = \frac{P(U_2 = 1, U_2 \oplus U_3 = 1)}{P(U_2 \oplus U_3 = 1)}$$

The above equation gives

$$P(U_2 = 1 \mid X = 1) = \frac{P(U_2 = 1, U_3 = 0)}{P(U_2 = 1, U_3 = 0) + P(U_2 = 0, U_3 = 1)}$$
$$\frac{P(U_2 = 1)P(U_3 = 0)}{P(U_2 = 1)P(U_3 = 0) + P(U_2 = 0)P(U_3 = 1)}$$
$$= \frac{1}{4}$$

We thus have

$$P(Y_{x=0} = 1 \mid X = 1) = P(0 \oplus U_2 = 1 \mid X = 1) = P(U_2 = 1 \mid X = 1) = \frac{1}{4}.$$

(iii) $P(Y_{x=0} = 1 | X = 1, W = 1, Z = 1)$.

$X = 1, W = 1, Z = 1$ implies

$$U_1 \oplus U_2 = 1, \;\; U_1 \oplus U_2 \oplus U_3 = 1, \;\; U_2 \oplus U_3 = 1.$$

The above equation gives

$$U_1 = 0, \;\; U_2 = 1, \;\; U_3 = 0.$$

That means that $P(U_2 = 1|X = 1, W = 1, Z = 1) = 1$. We thus have

$$\begin{aligned} P(Y_{x=0} = 1 \mid X = 1, W = 1, Z = 1) &= P(0 \oplus U_2 = 1 \mid X = 1, W = 1, Z = 1) \\ &= P(U_2 = 1 \mid X = 1, W = 1, Z = 1) \\ &= 1. \end{aligned}$$

(b) Consider the following causal diagram

$$X \longrightarrow Y$$

Is the distribution $P(y_x, y_{x'})$ for $x \neq x'$ identified from $P(x, y)$ and $P(y|do(x))$? If yes, provide a derivation. Otherwise, provide a counterexample.

Consider following SCMs M_1, M_2.

$$M_1 = \begin{cases} \mathbf{U} = \{U_1, U_2\} \\ X = U_1, \\ Y = U_2, \\ P(U_1 = 1) = P(U_2 = 1) = 0.5 \end{cases} \qquad M_2 = \begin{cases} \mathbf{U} = \{U_1, U_2\} \\ X = U_1, \\ Y = X \cdot U_2 + (1 - X)(1 - U_2), \\ P(U_1 = 1) = P(U_2 = 1) = 0.5 \end{cases}$$

It is verifiable in both M_1, M_2, for $i, j = 0, 1$

$$P_{M_i}(X = i, Y = j) = 0.25, \qquad P_{M_i}(Y = j|do(X = i)) = 0.5.$$

Since X and Y have no functional relationship in M_1, we have $Y_{x=0}(\mathbf{u}) = Y_{x=1}(\mathbf{u})$ for any $\mathbf{u}$. We thus have

$$P_{M_1}(Y_{x=0} = 1, Y_{x=1} = 1) = P(U_2 = 1) = 0.5.$$

On the other hand, in M_2, given any $U_2 = u_2$,

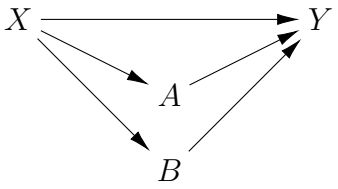
$$Y_{x=0}(\mathbf{u}) = 1 - u_2, \qquad Y_{x=1}(\mathbf{u}) = u_2.$$

That is, potential outcomes $Y_{x=0}(\mathbf{u}) \neq Y_{x=1}(\mathbf{u})$ for any $\mathbf{u}$. This implies

$$P_{M_2}(Y_{x=0} = 1, Y_{x=1} = 1) = 0$$

Exercise 6. Direct and Indirect Effects

In this problem, we will investigate how counterfactual queries can allow more fine-tuned explanations of cause and effect. Consider the following graph $\mathcal{G}$ and observational data $P(\mathbf{V})$.



(a) Graph $\mathcal{G}$

X	A	B	Y	P
0	0	0	0	$p_0 = 0.324$
0	0	0	1	$p_1 = 0.036$
0	0	1	0	$p_2 = 0.02$
0	0	1	1	$p_3 = 0.02$
0	1	0	0	$p_4 = 0.045$
0	1	0	1	$p_5 = 0.045$
0	1	1	0	$p_6 = 0.005$
0	1	1	1	$p_7 = 0.005$
1	0	0	0	$p_8 = 0.005$
1	0	0	1	$p_9 = 0.005$
1	0	1	0	$p_{10} = 0.009$
1	0	1	1	$p_{11} = 0.081$
1	1	0	0	$p_{12} = 0.004$
1	1	0	1	$p_{13} = 0.036$
1	1	1	0	$p_{14} = 0.036$
1	1	1	1	$p_{15} = 0.324$

(b) Observational Data $P(\mathbf{V})$. Probability of ith row is labeled p_i.

(a) Identify the following expressions from observational data $P(\mathbf{V})$ and the graph $\mathcal{G}$.

(i) $P(Y_{X=x} = y)$

$P(Y_{X=x} = y) = P(Y = y \mid X = x)$ by Rule 2 of do-calculus.

(ii) $P(Y_{X=x,A_{X=x'},B_{X=x'}} = y)$

$$P(Y_{X=x,A_{X=x'},B_{X=x'}} = y) \tag{23}$$
$$= \sum_{a,b} P(Y_{X=x,A=a,B=b} = y, A_{X=x'} = a, B_{X=x'} = b) \tag{24}$$

by counterfactual unnesting

$$= \sum_{a,b} P(Y_{X=x,A=a,B=b} = y \mid A_{X=x'} = a, B_{X=x'} = b) P(A_{X=x'} = a, B_{X=x'} = b) \tag{25}$$

chain rule

$$= \sum_{a,b} P(Y_{X=x,A=a,B=b} = y) P(A_{X=x'} = a, B_{X=x'} = b) \tag{26}$$

ctf-calculus rule 2: $Y_{x,a,b} \perp \{A_{x'}, B_{x'}\}$ in $\mathcal{G}_A(Y_{x,a,b}, A_{x'}, B_{x'})$

$$= \sum_{a,b} P(Y = y \mid X = x, A = a, B = b) P(A = a, B = b \mid X = x') \tag{27}$$

do-calculus rule 2

$$= \sum_{a,b} P(Y = y \mid X = x, A = a, B = b) P(A = a \mid X = x') P(B = b \mid X = x') \tag{28}$$

$A \perp B \mid X$

(b) Using the results from part (a), compute the following queries using the table in Figure (b).

(i) Average Treatment Effect (ATE): $P(Y_{X=1} = 1) - P(Y_{X=0} = 1)$

$$P(Y_{X=1} = 1) - P(Y_{X=0} = 1) \tag{29}$$
$$= P(Y = 1 \mid X = 1) - P(Y = 1 \mid X = 0) \tag{30}$$
$$= \frac{p_9 + p_{11} + p_{13} + p_{15}}{p_8 + p_9 + p_{10} + p_{11} + p_{12} + p_{13} + p_{14} + p_{15}} - \frac{p_1 + p_3 + p_5 + p_7}{p_0 + p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7} \tag{31}$$
$$= \frac{0.446}{0.5} - \frac{0.106}{0.5} = 0.68 \tag{32}$$

(ii) Natural Direct Effect (NDE): $P(Y_{X=1,A_{X=0},B_{X=0}} = 1) - P(Y_{X=0} = 1)$

To keep the solution concise, note that from the table, the following can be calculated

$$P(A = a \mid X = x) = \begin{cases} 0.8 & a = x \\ 0.2 & a \neq x \end{cases} \tag{33}$$

$$P(A = b \mid X = x) = \begin{cases} 0.9 & b = x \\ 0.1 & b \neq x \end{cases} \tag{34}$$

$$P(Y = 1 \mid A = a, B = b, X = x) = \begin{cases} 0.1 & a = b = x = 0 \\ 0.9 & x = 1 \wedge (a = 1 \vee b = 1) \\ 0.5 & \text{otherwise.} \end{cases} \tag{35}$$

Now we can compute

$$
\begin{aligned}
&P(Y_{X=1,A_{X=0},B_{X=0}} = 1) \\
&= \sum_{a,b} P(Y = 1 \mid X = 1, A = a, B = b) P(A = a \mid X = 0) P(B = b \mid X = 0) \\
&= P(Y = 1 \mid X = 1, A = 0, B = 0) P(A = 0 \mid X = 0) P(B = 0 \mid X = 0) \\
&+ P(Y = 1 \mid X = 1, A = 0, B = 1) P(A = 0 \mid X = 0) P(B = 1 \mid X = 0) \\
&+ P(Y = 1 \mid X = 1, A = 1, B = 0) P(A = 1 \mid X = 0) P(B = 0 \mid X = 0) \\
&+ P(Y = 1 \mid X = 1, A = 1, B = 1) P(A = 1 \mid X = 0) P(B = 1 \mid X = 0) \\
&= (0.5)(0.8)(0.9) + (0.9)(0.8)(0.1) + (0.9)(0.2)(0.9) + (0.9)(0.2)(0.1) \\
&= 0.612.
\end{aligned}
$$

Recall that $P(Y_{X=0} = 1) = 0.212$ from the previous answer, so $P(Y_{X=1,A_{X=0},B_{X=0}} = 1) - P(Y_{X=0} = 1) = 0.612 - 0.212 = 0.4$.

(iii) Natural Indirect Effect (NIE): $P(Y_{X=0,A_{X=1},B_{X=1}} = 1) - P(Y_{X=0} = 1)$

$$
\begin{aligned}
&P(Y_{X=0,A_{X=1},B_{X=1}} = 1) \\
&= \sum_{a,b} P(Y = 1 \mid X = 0, A = a, B = b) P(A = a \mid X = 1) P(B = b \mid X = 1) \\
&= P(Y = 1 \mid X = 0, A = 0, B = 0) P(A = 0 \mid X = 1) P(B = 0 \mid X = 1) \\
&+ P(Y = 1 \mid X = 0, A = 0, B = 1) P(A = 0 \mid X = 1) P(B = 1 \mid X = 1) \\
&+ P(Y = 1 \mid X = 0, A = 1, B = 0) P(A = 1 \mid X = 1) P(B = 0 \mid X = 1) \\
&+ P(Y = 1 \mid X = 0, A = 1, B = 1) P(A = 1 \mid X = 1) P(B = 1 \mid X = 1) \\
&= (0.1)(0.2)(0.1) + (0.5)(0.2)(0.9) + (0.5)(0.8)(0.1) + (0.5)(0.8)(0.9) \\
&= 0.492.
\end{aligned}
$$

Once again, recall that $P(Y_{X=0} = 1) = 0.212$, so $P(Y_{X=0,A_{X=1},B_{X=1}} = 1) - P(Y_{X=0} = 1) = 0.492 - 0.212 = 0.28$.

(c) Query (b)(ii) can be interpreted as specifically the effect of intervening $X = 1$ on Y through the direct path $X \to Y$, ignoring the indirect paths through A and B. On the other hand, query (b)(iii) is the opposite: the effect of intervening $X = 1$ on Y through the indirect paths through A and B and ignoring the direct path. Query (b)(i) aggregates the causal effect of $X = 1$ on Y through all paths [1]. What query should be calculated if the goal is to get the causal effect of $X = 1$ on Y specifically through the path $X \to A \to Y$, ignoring both the direct path $X \to Y$ and the indirect path $X \to B \to Y$?

Answer: $P(Y_{X=0,A_{X=1}} = 1) - P(Y_{X=0} = 1)$

$P(Y_{X=0,A_{X=1},B_{X=0}} = 1) - P(Y_{X=0} = 1)$ is acceptable, but the $B_{X=0}$ term is redundant.

(d) Identify the terms of the query in part (c) and calculate its value using $P(\mathbf{V})$ from Table (b).

[1] Despite this observation, it will typically not be the case that ATE = NDE + NIE (i.e., the sum of effects on all paths) because the causes are not necessarily disjoint.

We can identify $P(Y_{X=0,A_{X=1}} = 1)$ as follows:

$$
\begin{aligned}
&P(Y_{X=0,A_{X=1}} = 1) \\
&= \sum_a P(Y_{X=0,A=a} = 1, A_{X=1} = a) \\
&\text{by counterfactual unnesting} \\
&= \sum_a P(Y_{X=0,A=a} = 1 \mid A_{X=1} = a) P(A_{X=1} = a) \\
&\text{chain rule} \\
&= \sum_a P(Y_{X=0,A=a} = 1) P(A_{X=1} = a) \\
&\text{Ctf-calculus Rule 2: } Y_{x,a} \perp A_{x'} \text{ in } \mathcal{G}_A(Y_{x,a}, A_{x'}) \\
&= \sum_a P(Y = 1 \mid X = 0, A = a) P(A = a \mid X = 1) \\
&\text{do-calculus Rule 2}
\end{aligned}
$$

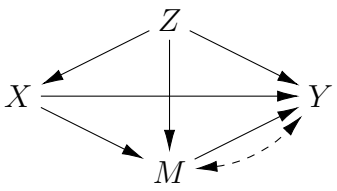
We can first calculate from the table that $P(Y = 1 \mid X = 0, A = 0) = 0.14$ and $P(Y = 1 \mid X = 0, A = 1) = 0.5$. Then we can compute

$$
\begin{aligned}
P(Y_{X=0,A_{X=1}} = 1) & \\
&= \sum_a P(Y = 1 \mid X = 0, A = a) P(A = a \mid X = 1) \\
&= P(Y = 1 \mid X = 0, A = 0) P(A = 0 \mid X = 1) \\
&+ P(Y = 1 \mid X = 0, A = 1) P(A = 1 \mid X = 1) \\
&= (0.14)(0.2) + (0.5)(0.8) = 0.428.
\end{aligned}
$$

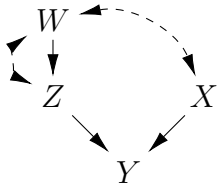
Combine this with the value of $P(Y_{X=0} = 1) = 0.212$ and we get $P(Y_{X=0,A_{X=1}} = 1) - P(Y_{X=0} = 1) = 0.428 - 0.212 = 0.216$

Exercise 7. [Optional] Counterfactual Identification

Consider the following graphs. Assume all variables have finite discrete domains (but not necessarily binary).



(a)



(b)

(a) Is $P(Y_{X=x_1} = y_1 \mid X = x_0)$ identifiable from $P(\mathbf{V})$ and the graph $\mathcal{G}$ in Figure (a)? If yes, then provide the identification expression (and show your work). If no, then provide a counterexample.

$$P(Y_{X=x_1} = y_1 \mid X = x_0)$$
$$= \sum_z P(Y_{X=x_1} = y_1 \mid Z = z, X = x_0) P(Z = z \mid X = x_0)$$
Conditioning on Z
$$= \sum_z P(Y_{X=x_1} = y_1 \mid Z_{X=x_1} = z, X = x_0) P(Z = z \mid X = x_0)$$
Rule 3: $\{X\} \cap An(Z) = \emptyset$
$$= \sum_z P(Y_{X=x_1, Z=z} = y_1 \mid Z_{X=x_1} = z, X = x_0) P(Z = z \mid X = x_0)$$
Rule 1: $Z_{x_1} = z \Rightarrow Y_{x_1} = Y_{x_1,z}$
$$= \sum_z P(Y_{X=x_1, Z=z} = y_1 \mid Z = z, X = x_0) P(Z = z \mid X = x_0)$$
Rule 3: $\{X\} \cap An(Z) = \emptyset$
$$= \sum_z P(Y_{X=x_1, Z=z} = y_1 \mid Z = z, X = x_1) P(Z = z \mid X = x_0)$$
Rule 2: $X \perp Y_{x_1,z} \mid Z$ in $\mathcal{G}_A$
$$= \sum_z P(Y \mid Z = z, X = x_1) P(Z = z \mid X = x_0)$$
Rule 1: $Z = z, X = x \Rightarrow Y_{x_1,z} = Y$

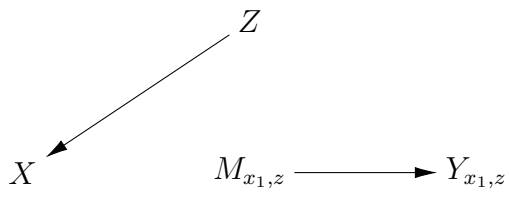


Figure 3: $\mathcal{G}_A$

(b) Are the following queries identifiable from $P(\mathbf{V})$ and $P(X, Z, Y \mid do(W))$ and the graph $\mathcal{G}$ in Figure (b)? If yes, then provide the identification expression (and show your work). If no, then provide a counterexample.

(i) $P(Y_{X=x_0, W=w_0} = y_0, X_{W=w_1} = x_1)$

It is identifiable.

$$P(Y_{X=x_0,W=w_0} = y_0, X_{W=w_1} = x_1) \tag{36}$$
$$= \sum_z P(Y_{X=x_0,W=w_0} = y_0, Z_{X=x_0,W=w_0} = z, X_{W=w_1} = x_1) \tag{37}$$
$$= \sum_z P(Y_{X=x_0,Z=z,W=w_0} = y_0, Z_{X=x_0,W=w_0} = z, X_{W=w_1} = x_1) \tag{38}$$

by counterfactual consistency

$$= \sum_z P(Y_{X=x_0,Z=z} = y_0, Z_{X=x_0,W=w_0} = z, X = x_1) \tag{39}$$

ctf-calculus rule 3: $W \cap An(Y) = \emptyset$ in $\mathcal{G}_{\overline{XZ}}$

$$= \sum_z P(Y_{X=x_0,Z=z} = y_0, Z_{W=w_0} = z, X = x_1) \tag{40}$$

ctf-calculus rule 3: $X \cap An(Z) = \emptyset$ in $\mathcal{G}_{\overline{W}}$

$$= \sum_z P(Y_{X=x_0,Z=z} = y_0 \mid Z_{W=w_0} = z, X = x_1) P(Z_{W=w_0} = z \mid X = z_1) P(X = x_1) \tag{41}$$

chain rule

$$= \sum_z P(Y_{X=x_0,Z=z} = y_0) P(Z_{W=w_0} = z) P(X = x_1) \tag{42}$$

ctf-calculus rule 2: $Y_{x,z} \perp \{Z_w, X\}, Z_w \perp X$ in $\mathcal{G}_{Y_{x,z},Z_w,X}$

$$= \sum_z P(Y = y_0 \mid X = x_0, Z = z) P(Z_{W=w_0} = z) P(X = x_1) \tag{43}$$

by do-calculus Rule 2.

Note that $P(Z_{W=w_0} = z)$ can be obtained from $P(X, Z, Y \mid do(W))$, so the result is an identification expression.

(ii) $P(Y_{X=x_0} = y_0, Z_{W=w_0} = z_0)$

The query is non-ID. Consider the following pair of models.

$$\mathcal{M}_1 = \begin{cases} \mathbf{U} & = \{U_W, U_Z, U_X, U_Y\} \\ \mathbf{V} & = \{W, Z, X, Y\} \\ \mathcal{F} & = \begin{cases} f_W(u_W) & = u_W \\ f_Z^1(w, u_Z) & = w \oplus u_Z \\ f_X(u_X) & = u_X \\ f_Y(z, x, u_Y) & = z \oplus u_Y \end{cases} \\ P(\mathbf{U}) & = \begin{cases} P(U_W = 1) = P(U_X = 1) = P(U_Z = 1) = 0.5 \\ P(U_Y = 1) = 0.1 \end{cases} \end{cases} \tag{44}$$

$$\mathcal{M}_2 = \begin{cases} \mathbf{U} & = \{U_W, U_{Z0}, U_{Z1}, U_X, U_Y\} \\ \mathbf{V} & = \{W, Z, X, Y\} \\ \mathcal{F} & = \begin{cases} f_W(u_W) & = u_W \\ f_Z^1(w, u_{Z0}, u_{Z1}) & = \begin{cases} u_{Z0} & w = 0 \\ u_{Z1} & w = 1 \end{cases} \\ f_X(u_X) & = u_X \\ f_Y(z, x, u_Y) & = z \oplus u_Y \end{cases} \\ P(\mathbf{U}) & = \begin{cases} P(U_W = 1) = P(U_X = 1) = P(U_{Z0} = 1) = P(U_{Z1} = 1) = 0.5 \\ P(U_Y = 1) = 0.1 \end{cases} \end{cases} \tag{45}$$

Note that both $\mathcal{M}_1$ and $\mathcal{M}_2$ match in $\mathcal{G}$. They also match in $P(\mathbf{V})$, as they are identical except for f_Z, which outputs a value of 0 or 1 with probability 0.5 regardless of the value of w. However, calculating the query (with $w_i = i$, $z_i = i$, $x_i = i$, $y_i = i$), we see that

$$P^{\mathcal{M}_1}(Y_{X=0} = 0, Z_{W=0} = 0) \tag{46}$$
$$= P(U_W = 0, U_Z = 0, U_Y = 0) + P(U_W = 1, U_Z = 0, U_Y = 1) \tag{47}$$
$$= 0.225 + 0.025 = 0.25 \tag{48}$$

while

$$P^{\mathcal{M}_2}(Y_{X=0} = 0, Z_{W=0} = 0) \tag{49}$$
$$= P(U_W = 0, U_{Z0} = 0, U_Y = 0) + P(U_W = 1, U_{Z0} = 0, U_{Z1} = 0, U_Y = 0) \tag{50}$$
$$+ P(U_W = 1, U_{Z0} = 0, U_{Z1} = 1, U_Y = 1) \tag{51}$$
$$= 0.225 + 0.1125 + 0.0125 = 0.4 \tag{52}$$