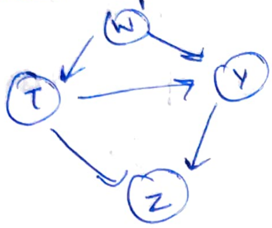
Puneet-Mangla (CS17BTECH11029)

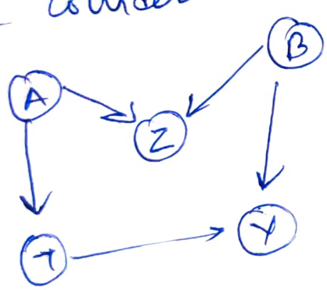
Q.1 Collider-bias: when the treatment, T and the outcome, Y variable cause a third variable, that variable is called collider. Conditioning on a collider opens the indirect path from T to Y



Here Z is a collider. to estimate P(Y| do(T)), we condition or adjust only using {W}.

$$P(Y|do(t)) = \sum_{w} P(Y \mid T=t, W=w)$$

M-bias: It is introduced by conditioning on a pre-treatment covariate due to particular M-structure between two latent factors, an observed treatment (T) and outcome (Y) and a collider (Z)

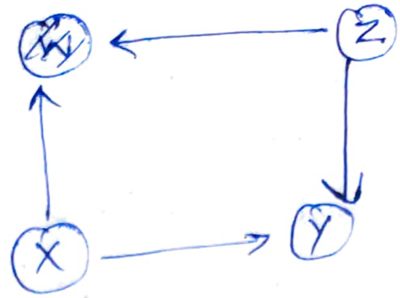


If we condition on Z then association path: T→A→Z→B→Y opens.

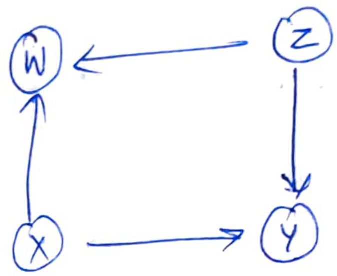
So we have to adjust over {A, B}

Selection-bias: Selection bias is introduced by the selection of individuals, groups or data for analysis in such a way that proper randomization is not acheied, thereby ensuring that sample obtained is not representative of population intended to be analyzed.

Here, we have to adjust over {W} which introduces this selection bias.



~~It~~ W is a collider here. If we condition on W, the path $X \to W \to Z \to Y$ opens, thus providing an association effect from X to Y. It will introduce a collider-bias.



$$P(Y=y \mid do(X=x), W=w) = \sum_{z} P(Y=y \mid X=x, W=w, Z=z)$$

Q.2 a) Actual ATE = $ATE = E[Y(1) - Y(0)] = 2$

Assuming uncounfounddness is present

Estimated ATE = $\hat{ATE} = E_X[Y_{obs} \mid X = x, W = 1]$
$- E_X[Y_{obs} \mid X = x, W = 0]$

$$\cancel{= (0.5 \cdot 6 + 0.5 \cdot 6 + 0.5 \times 12 + 0.5 \times 12)}$$
$$\cancel{- (0.5 \cdot 4 + 0.5 \cdot 4 + 0.5 \times 10 + 0.5 \times 10}$$

$$= E_X[E[Y \mid X = x, W = 1]] - E_X[E[Y \mid X = x, W = 0]]$$

$$= \left[\frac{1}{2} \cdot 6 + \frac{1}{2} \cdot 12\right] - \left[\frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 10\right] = 2$$

Since $ATE = \hat{ATE}$, unconfoundedness is present.

b) $e(X=0) = 0.5$
$e(X=1) = 0.75$

c) $$E\left[\frac{1(W=1) \cdot Y}{e(X)}\right] - E\left[\frac{1(W=0) \cdot Y}{e(X)}\right]$$

$$\frac{1}{200}\left[\frac{6 \times 30}{0.5} + \frac{6 \times 30}{0.75} + \frac{12 \times 20}{0.5} + \frac{45 \times 12}{0.75}\right]$$

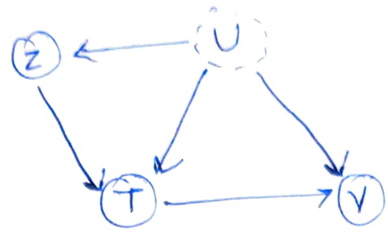
$$-\frac{1}{200}\left[\frac{30 \times 4}{1-0.5} + \frac{10 \times 4}{1-0.75} + \frac{20 \times 10}{1-0.5} + \frac{15 \times 10}{1-0.75}\right]$$

$$= \frac{1}{200} \times 1800 - \frac{1}{200} 1400$$

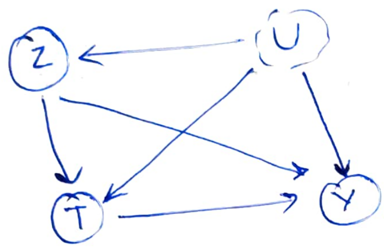
$$= 2$$

Q.3 a) Exclusion Restriction: The causal effect of Z on Y is not direct. It affects outcome variable, Y only through a mediator X.

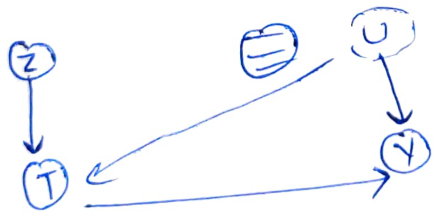
For eg:



b) Relevance: Z has a causal effect on Y



c) Exogeneity: Z is unconfounded i.e there are no unblockable backdoor paths to Y



Q.4 $e(Z=1) = P(X=1 \mid Z=1) = \dfrac{P(X=1, Z=1)}{P(Z=1)}$

$$= \frac{\sum_{Y} P(X=1, Y, Z=1)}{\sum_{X Y} P(X, Y, Z=1)}$$

$$= \frac{0.116 + 0.009}{0.51}$$

$$= 0.245$$

$$e(Z=0) = P(X=1 \mid Z=0) = \frac{\sum_{Y} P(X=1, Y, Z=0)}{\sum_{X,Y} P(X, Y, Z=0)}$$

$$= \frac{0.274 + 0.101}{0.49}$$

$$= 0.765$$

$$ATE = E\left[\frac{Y \cdot 1(X=1)}{e(Z)}\right] - E\left[\frac{Y \cdot 1(X=0)}{1-e(Z)}\right]$$

$$= \left[\frac{0.116 \times 1}{0.245} + \frac{0.274 \times 1}{0.765} + \frac{0.009 \times 0}{0.245} + \frac{0.101 \times 0}{0.765}\right]$$

$$- \left[\frac{0.334 \times 1}{1-0.245} + \frac{0.079 \times 1}{1-0.765} + \frac{0.051 \times 0}{1-0.245} + \frac{0.036 \times 0}{1-0.765}\right]$$

$$= [0.473 + 0.358] - [0.442 + 0.336]$$

$$= 0.053$$

Q.5 $\psi = E_X\left[E[Y|X=x, W=1] - E[Y|X=x, W=0]\right]$

Take first term i.e $E_X[E[Y|X=x, W=1]]$

$$E_X[E[Y|X=x, W=1]] = \sum_x \left(\sum_y y \cdot p(y|X=x, W=1)\right) p(x)$$

$$= \sum_x \sum_y y \cdot p(y|X=x, W=1)\, p(x)$$

now $p(W=1|x) = \dfrac{p(W=1, X=x)}{p(x)}$

$$\Rightarrow p(x) = \frac{p(W=1, X=x)}{p(W=1|x)} = \frac{p(W=1, X=x)}{e(x)}$$

$$E_X[E[Y|X=x, W=1]] = \sum_x \sum_y y \cdot \frac{p(y|X=x, W=1)\, p(W=1, X=x)}{e(x)}$$

$$= \sum_x \sum_y y \cdot \frac{p(X=x, y, W=1)}{e(x)}$$

$$= \sum_W \sum_x \sum_y y \cdot 1(W=1) \frac{p(X=x, y, W)}{e(x)}$$

$$= E\left[\frac{y \cdot 1(W=1)}{e(x)}\right]$$

Take Second term i.e $E_X[E[Y|X=x, W=0]]$

$$E_X[E[Y|X=x, W=0]] = \sum_x \sum_y y \cdot p(y|X=x, W=0)\, p(x)$$

$$p(W=0|x) = 1 - p(W=1|x) = \frac{p(W=0, X=x)}{p(x)}$$

$$p(x) = \frac{p(W=0, X=x)}{1 - e(x)}$$

$$\underset{X}{E}\left[E[Y \mid X=x, W=0]\right] = \frac{\sum_x \sum_y y \cdot p(y \mid X=x, W=0) \cdot p(W=0, X=x)}{1-e(x)}$$

$$= \sum_x \sum_y y \cdot \frac{p(y, X=x, W=0)}{1-e(x)}$$

$$= \sum_{x,y,w} y \cdot \frac{p(y,x,w) \cdot \mathbb{1}(w=0)}{1-e(x)}$$

$$= E\left[\frac{y \cdot \mathbb{1}(w=0)}{1-e(x)}\right]$$

Hence,

$$\phi = E\left[\frac{Y \cdot 1(W=1)}{e(x)}\right] - E\left[\frac{Y \cdot \mathbb{1}(W=0)}{1-e(x)}\right]$$

Q8 a) $E[Y_{X=x} \mid Z=1]$

b)

u_1	u_2	$X_{U_1=u_1, U_2=u_2}$	$Z_{U_1=u_1, U_2=u_2}$	Y_{u_1, u_2}	$Z_{X=0, U_1=u_1, U_2=u_2}$	$Z_{X=1, U_1=u_1, U_2=u_2}$	$Y_{X=0, U_1=u_1, U_2=u_2}$	$Y_{X=1, U_1=u_1, U_2=u_2}$
0	0	0	0	0	0	a	0	ab
0	1	0	1	b	1	a+1	b	ab+b
1	0	1	a	ab	0	a	0	ab
1	1	1	a+1	ab+b.	1	a+1	b	ab+b

Q.7 Let m be the slope of total effect of X on Y.

$$m = E[Y \mid do(x+1)] - E[Y \mid do(x)]$$

Now for any evidence e.

$$E[Y_{X=x} \mid e] = E[Y \mid e] + m(x - E[X \mid e])$$

This means $E[Y_{X=x} \mid e]$ can be computed by first calculating the best estimate of Y conditioned on the evidence e, $E[Y \mid e]$ and then adding to it whatever change is expected in Y when X is shifted from its current best estimate $E[X \mid e]$ to its hypothetical value, x.