Causal Inference - HW3

Afek Adler

December 23, 2019

* Along the excercise I will note G as the group that blocks paths (Z in the lectures)

Q1 Consider the DAG in Figure 1. For each of the following pairs of nodes, list: (1) all possible paths between them (2) what set of nodes is required (in a d-separation sense) to block all of their paths:

1. $W \to S$
2. $X \to T$
3. $Y \to H$

Q1 solution

For $W \to S$, the following paths exist:

1. $W \to Y \leftarrow X \to Z \to S$
2. $W \to Y \leftarrow X \to Z \to R \to S$
3. $W \to Y \to R \leftarrow Z \to S$
4. $W \to Y \to R \to S$

And in order to block each path (in d- separation), we need to fulfill:

1. $Y, R, R \notin G \vee X \in G \vee Z \in G$
2. $Y, R, R \notin G \vee X \in G \vee Z \in G \vee R \in G$
3. $T, R \notin G \vee Y \in G \vee Z \in G$
4. $Y \in G \vee R \in G$

So for example $G := \{Z, R\}$ or $\{Y, X, R\}$

For $X \to T$, the following paths exist:

1. $X \to Y \to R \to T$
2. $X \to Z \to R \to T$
3. $X \to Z \to S \leftarrow R \to T$

And in order to block each path (in d- separation), we need to fulfill:

1. $Y \in G \vee R \in G$
2. $Z \in G \vee R \in G$
3. $S \notin G \vee Z \in G \vee R \in G$

So for example $G := \{T, Z\}$

For $Y \to H$, the following paths exist:

1. $Y \leftarrow X \to Z \leftarrow H$
2. $Y \to R \to S \leftarrow Z \leftarrow H$
3. $Y \to R \leftarrow Z \leftarrow H$

And in order to block each path (in d- separation), we need to fulfill:

1. $X \in G \vee Z, R, S \notin G$
2. $Z \in G \vee R \in G \vee S \notin G$
3. $Z \in G \vee R, S \notin G$

So for example G := {X,Z}

Q2 *Consider the casual graph in Figure 2.*

1. *List all of the sets of variables that satisfy the backdoor criterion to determine the causal effect of T on Y .*
2. *List all of the minimal sets of variables that satisfy the backdoor criterion to determine the casual effect of T on Y (i.e., any set of variables such that, if you removed any one of the variables from the set, it would no longer meet the criterion).*
3. *Give a minimal set of variables that need to be measured in order to identify the effect of D on Y.*

Q2 solution

a. W is a descendent of T so it can't be in G.

1. $T \leftarrow A \leftarrow B \rightarrow Z \rightarrow Y$
2. $T \leftarrow A \leftarrow B \rightarrow Z \leftarrow C \rightarrow D \rightarrow Y$
3. $T \leftarrow Z \rightarrow Y$
4. $T \leftarrow Z \leftarrow C \rightarrow D \leftarrow Y$

And in order to block each path (in d- separation), we need to fulfill:

1. $A \in G \vee B \in G \vee Z \in G$
2. $A \in A \vee B \in G \vee C \in G \vee D \in G \vee Z \notin G$
3. $Z \in G$
4. $Z \in G \vee C \in G \vee D \in G$

b. from 3 we need $Z \in G$, from 2 we need to add one of $\{A, B, C, D\}$ to be in G. When $Z \in G$ 1 and 4 are True those this is the final solution.
c. We need to block all paths between Y and D such that there is an arrow to D. There are two of those:

1. $Y \leftarrow W \leftarrow T \leftarrow Z \leftarrow C \rightarrow D$
2. $Y \leftarrow W \leftarrow T \leftarrow A \leftarrow B \rightarrow Z \leftarrow C \rightarrow D$
3. $Y \leftarrow Z \leftarrow C \rightarrow D \rightarrow D$

And in order to block each path (in d- separation), we need to fulfill:

1. $W \in G \vee T \in G \vee Z \in G \vee C \in G$
2. $W \in G \vee T \notin G \vee A \in G \vee B \in G \vee Z \notin G \vee C \in G$
3. $Z \in G \vee C \in G \vee D \in G$

So C is enough to block those paths.

Q3 *the data (See Table 1), you have binary indicators for: prior education (Z), whether the annual real income in 2024 is higher than 50K\$ (X), had job training (T), whether the annual real income in 2026 is higher than 100K\$ (Y), and whether the citizen bought a house in 2026 (W). We know the following:*

1. *The income in 2024 depends solely on the prior education.*
2. *A person is selected to the job training program based on her prior education, and income in 2024.*

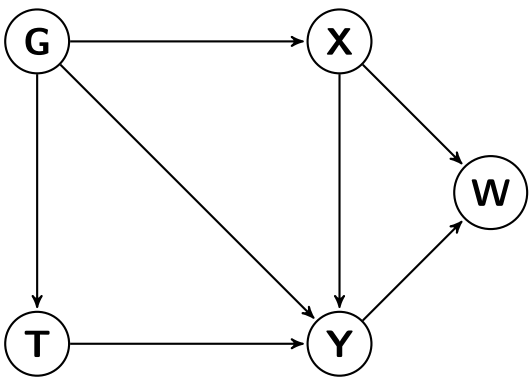
3. *The income in 2026 depends on the job training, prior education and income in 2024.*
4. *Whether a citizen purchased a house is directly based on her income in 2024 and 2026.*

Your task is as follows:

1. *Draw the causal graph that describe the above experiment.*
2. *Calculate the ATE of the experiment (derive the necessary probabilities from Table 1)*

Q3 solution

a. *The causal graph -*



b. *We can see that we need to measure Z and X in order to block all paths from T and Y .*
Now:
$ATE = E(Y|do(T=1)) - E(Y|do(T=0)) = E_{X,Y}(Y|x,y,T=1) - E_{X,Y}(Y|x,y,T=0)$

X	Z	p(X,Z)	N	T=1	T=0
0	0	0.3	6	4	2
0	1	0.35	7	4	3
1	0	0.1	2	1	1
1	1	0.25	5	2	3

The Column N stands for the number of observations in the bin. This number is divided to either being T = 0 or T = 1 (for each row (T = 0) + (T = 1) = N).
So:
$E_{X,Y}(Y|x,y,T=1) = 0.3 * \frac{4}{6} + 0.35 * \frac{4}{7} + 0.1 * \frac{1}{2} + 0.25 * \frac{2}{5} = 0.55$
$E_{X,Y}(Y|x,y,T=0) = 0.3 * \frac{2}{6} + 0.35 * \frac{3}{7} + 0.1 * \frac{1}{2} + 0.25 * \frac{3}{5} = 0.45$
and $E_{X,Y}(Y|x,y,T=1) - E_{X,Y}(Y|x,y,T=0) = 0.55 - 0.45 = 0.1$