Homework 1

Barbara Aleksandrov
337844252

Matan Solomon
209339894

April 30, 2025

1 Question 1

1.1 Example

In this part, we give an DGP example in the table given below (as requested with binary covariate X and treatment T).

Index	**X**	**T**	$\mathbf{Y_0}$	$\mathbf{Y_1}$	**Y**
1	0	0	2	3	2
2	0	0	1	2	1
3	0	1	2	3	3
4	1	1	4	5	5
5	1	1	3	4	4
6	1	0	3	4	3

1.2 $E[Y_t]$ and ATE calculation

$$E[Y_0] = \frac{1}{6}(2+1+2+4+3+3) = 2.5$$
$$E[Y_1] = \frac{1}{6}(3+2+3+5+4+4) = 3.5$$
$$ATE = E[Y_1 - Y_0] = 1$$

1.3 DGP exmaple with different ATE based on ignorability

Notation: $Y = T \cdot Y_1 - (1-T) \cdot Y_0$

Recap from Tutorial 2: Conditional ignorability means that the covariate X affects both the treatment assignment and the outcome - $Y(0), Y(1) \perp T\,|X$, while the full ignorability lets us assume that treatment is unaffected by X (meaning it is random): $Y(0), Y(1) \perp T$

In our example in Section 1.1, X affects both the treatment assignment and the outcome. For X = 0, we have that only $\frac{1}{3}$ of the items received treatment and overall Y values are lower in both outcomes. And for X = 1, we have that $\frac{2}{3}$ of items recived the treatment, with higher outcome values Y.

1.4 ATE estiamtion

First let's assume that full ignorability assumption holds for this example, then we have:

$$\hat{ATE} = \hat{E}[Y|T=1] - \hat{E}[Y|T=0] = \frac{1}{3}(3+5+4-2-1-3) = 2$$

As we can see it overestimates an ATE by 1 unit.

Now let's assume that only conditional ignorability holds, meaning:

$$\hat{ATE} = E[E[Y|T=1,X] - E[Y|T=0,X]]$$

For that we need to calculate:

$$\hat{E}[Y|T=1,X=1] = \frac{1}{2}(5+4) = 4.5$$

$$\hat{E}[Y|T=1,X=0] = 3$$

$$\hat{E}[Y|T=0,X=1] = 3$$

$$\hat{E}[Y|T=0,X=0] = \frac{1}{2}(2+1) = 1.5$$

In the end we have: In the end we have:

$$\hat{ATE} = \frac{1}{2}(\hat{E}[Y|T=1,X=1] + \hat{E}[Y|T=1,X=0]) - \frac{1}{2}(\hat{E}[Y|T=0,X=1] + \hat{E}[Y|T=0,X=0])$$
$$= \tfrac{1}{2}(4.5+3-3-1.5) = 1.5$$

Clearly that the ATE that was adjusted to affect of X on both Y and T was closer to real ATE value. If we had a bigger data size (with similiar correlation between the variables) the result would have been even more accurate.

Question 2:

1. Let

$$e_t(x) = E[\mathbb{I}\{T = t\}|X = x] = P(T = t|X = x)$$
$$q_t(x) = \frac{\mathbb{I}(T = t)}{e_t(x)} = \frac{\mathbb{I}(T = t)}{E[\mathbb{I}\{T = t\}|X = x]}$$

We assume Consistency: $Y = TY_1 + (1 - T)Y_0$ so for T=t we get that $\mathbb{I}\{T = t\} \cdot Y = \mathbb{I}\{T = t\} \cdot Y_t$ (*)

Ignorability: $(Y_0, Y_1) \perp T|X$ (**)

$$E[q_t(x)Y] = E\left[\frac{\mathbb{I}\{T = t\}}{e_t(x)} Y\right] =^{(*)} E\left[\frac{\mathbb{I}\{T = t\}}{e_t(x)} Y_t\right] =^{(Law\ of\ total\ expectation)}$$
$$= E_{x \sim p(x)}\left[E\left[\frac{\mathbb{I}\{T = t\}}{e_t(x)} Y_t|X\right]\right].$$

The inner expression:

$$E\left[\frac{\mathbb{I}\{T = t\}}{e_t(x)} Y_t|X\right] =^{e_t(x)\ is\ const} \frac{1}{e_t(x)} E[\mathbb{I}\{T = t\} \cdot Y_t|X] =^{(**)} \frac{1}{e_t(x)} \cdot E[\mathbb{I}\{T = t\}|X]$$
$$\cdot E[Y_t|X] =^{(e_t(x)\ defenition)}= \frac{1}{e_t(x)} \cdot e_t(x) \cdot E[Y_t|X] = E[Y_t|X].$$

We plug back into the original expression and we get:

$$E[q_t(x)Y] = E_{x \sim p(x)}\left[E\left[\frac{\mathbb{I}\{T = t\}}{e_t(x)} Y_t|X\right]\right] = E_{x \sim p(x)}\big[E[Y_t|X]\big] = E[Y_t].$$

2. We will denote the covariate vector for a pregnant woman $X = x_{preg}$ and we know that in the trial there were no pregnant women, so: $Pr\left(T = 1\middle|X = x_{preg}\right) = 0 \Rightarrow e_1\left(x_{preg}\right) = 0.$ This means $q_1\left(x_{preg}\right)$ is undefined (division by zero), and the same for $q_0\left(x_{preg}\right)$.

   That information violates the positivity of the trial, and we cannot say anything about the causal effect that the drug has on pregnant women.

Question 3:

1) The SUTVA assumption does **NOT** hold.
   For example, in treatments Y00000010 and Y00000011, the only change is that 'Steph Curry' got a basketball, but it affects 'Seth Curry', who got 0 in the first treatment and 1 with the second one.

2) Because each family in our data has 2 kids, and we assume the SUTVA holds between the different families, we can define the treatment as the number of basketballs each family got.
   For a family with kids i,j we can write the new treatment as $T_i + T_j \in \{0,1,2\}$.
   And the potential outcomes, that is, the number of children who got a full scholarship in each family - $Y_f = Y_i + Y_j \in \{0,1,2\}$.

3) First, we will create an updated table:

	$Y_f(0)$	$Y_f(1)$	$Y_f(2)$
Ball	0	1	0
Plumlee	0	2	0
Lopez	0	1	0
Curry	0	0	2

$$ATE(1-0) = \frac{1}{4}\sum_f [Y_f(1) - Y_f(0)] = \frac{1+2+1+0}{4} = 1$$

$$ATE(2-0) = \frac{1}{4}\sum_f [Y_f(2) - Y_f(0)] = \frac{0+0+0+2}{4} = \frac{1}{2}$$

$$ATE(2-1) = \frac{1}{4}\sum_f [Y_f(2) - Y_f(1)] = \frac{-1-2-1+2}{4} = -\frac{1}{2}$$

Question 4

Example

We were interested in finding causal relation between listening to classical music and cognitive performance, since as a student we are highly motivated to perform our best.

Causal question:
Does listening to classical music while studying improve cognitive performance on a test?

- Treatment (T): Listening to classical music while studying (binary)
- Outcome (Y): Score on a some cognitive test taken after the study session with or without listening to classical music
- Covariates (X): age, study time, IQ, musical preference

Is there confounding between Y and T

If we perform an observational study instead of RCT there could be people that will chose classical music. The choice will be dependent on their background and personal preferences which will affect both the outcome and treatment. A lot of times people from more prosperous households are exposed to classical music from childhood in addition their education upbringing is usually much better. In this case it's better to include such confounder such as socio-economic status.

In addition Students who choose to listen to classical music may already be more focused, disciplined, or have different study habits.

Prediction problem

Prediction of average grade in math course of a student based on his musical preferences.

In this case we aren't interested in causal relation but in predicting the outcome, there could be some additional confounders or no relation at all which could impact the model's performance.