Introduction to Casual Inference - 097400

Homework 1 Winter 2020/2021

Asaf Gendler 301727715

5 בנובמבר 2020

Question 1:

1.

let's describe a data generating process:

assume we have a binary covariate x and a treatment T, and there is a linear relation between the possible outcomes from the treatment and x, hence:

$$Y_0 = 2x$$

$$Y_1 = 2x + 3$$

The true treatment effect is:

$$\mathbb{E}\left[Y_1 - Y_0\right] = \mathbb{E}\left[2x + 3 - 2x\right] = \mathbb{E}\left[3\right] = 3$$

let's assume that there is the same probability for x to receive any of it's values:

$$P\left(x = 0\right) = P\left(x = 1\right) = \frac{1}{2}$$

and also that there is the same probability to receive or not receive the treatment, no matter what is the value of x:

$$P\left(T = 1|x = 0\right) = P\left(T = 1|x = 1\right) = \frac{1}{2}$$

let's calculate:

$$\begin{aligned} P\left(T = 1\right) = P\left(T = 1|x = 0\right)P\left(x = 0\right) + P\left(T = 1|x = 1\right)P\left(x = 1\right) = \\ = \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4} + \frac{1}{4} = \frac{2}{4} = \frac{1}{2} \end{aligned}$$

$$P\left(T = 0\right) = 1 - P\left(T = 1\right) = 1 - \frac{1}{2} = \frac{1}{2}$$

$$\begin{aligned} \mathbb{E}\left[Y|T = 1\right] &= \mathbb{E}\left[Y|T = 1, x = 0\right]P\left(x = 0|T = 1\right) + \mathbb{E}\left[Y|T = 1, x = 1\right]P\left(x = 1|T = 1\right) = \\ &= \mathbb{E}\left[Y|T = 1, x = 0\right]\frac{P\left(T = 1|x = 0\right)P\left(x = 0\right)}{P\left(T = 1\right)} + \mathbb{E}\left[Y|T = 1, x = 1\right]\frac{P\left(T = 1|x = 1\right)P\left(x = 1\right)}{P\left(T = 1\right)} = \\ &= 3\cdot\frac{\frac{1}{2}\cdot\frac{1}{2}}{\frac{1}{2}} + 5\frac{\frac{1}{2}\cdot\frac{1}{2}}{\frac{1}{2}} = 3\cdot\frac{1}{2} + 5\cdot\frac{1}{2} = \frac{8}{2} = 4 \end{aligned}$$

$$\begin{aligned} \mathbb{E}\left[Y|T = 0\right] &= \mathbb{E}\left[Y|T = 0, x = 0\right]P\left(x = 0|T = 0\right) + \mathbb{E}\left[Y|T = 0, x = 1\right]P\left(x = 1|T = 0\right) = \\ &= \mathbb{E}\left[Y|T = 0, x = 0\right]\frac{P\left(T = 0|x = 0\right)P\left(x = 0\right)}{P\left(T = 0\right)} + \mathbb{E}\left[Y|T = 0, x = 1\right]\frac{P\left(T = 0|x = 1\right)P\left(x = 1\right)}{P\left(T = 0\right)} = \\ &= 0\cdot\frac{\frac{1}{2}\cdot\frac{1}{2}}{\frac{1}{2}} + 2\frac{\frac{1}{2}\cdot\frac{1}{2}}{\frac{1}{2}} = 2\cdot\frac{1}{2} = 1 \end{aligned}$$

$$\mathbb{E}\left[Y|T=1\right]-\mathbb{E}\left[Y|T=0\right]=4-1=3$$

hence the theorem holds:

$$\boxed{\mathbb{E}\left[Y_1-Y_0\right]=\mathbb{E}\left[Y|T=1\right]-\mathbb{E}\left[Y|T=0\right]=3}$$

2.

now let's change only one thing:

$$P\left(x=0\right)=\frac{1}{3},P\left(x=1\right)=\frac{2}{3}$$

and calculate again:

$$\begin{aligned}P\left(T=1\right)=P\left(T=1|x=0\right)P\left(x=0\right)+P\left(T=1|x=1\right)P\left(x=1\right)=\\=\frac{1}{2}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{2}{3}=\frac{1}{6}+\frac{1}{3}=\frac{1}{2}\end{aligned}$$

$$P\left(T=0\right)=1-P\left(T=1\right)=1-\frac{1}{2}=\frac{1}{2}$$

$$\begin{aligned}\mathbb{E}\left[Y|T=1\right]&=\mathbb{E}\left[Y|T=1,x=0\right]P\left(x=0|T=1\right)+\mathbb{E}\left[Y|T=1,x=1\right]P\left(x=1|T=1\right)=\\&=\mathbb{E}\left[Y|T=1,x=0\right]\frac{P\left(T=1|x=0\right)P\left(x=0\right)}{P\left(T=1\right)}+\mathbb{E}\left[Y|T=1,x=1\right]\frac{P\left(T=1|x=1\right)P\left(x=1\right)}{P\left(T=1\right)}=\\&=3\cdot\frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}}+5\frac{\frac{1}{2}\cdot\frac{2}{3}}{\frac{1}{2}}=3\cdot\frac{1}{3}+5\cdot\frac{2}{3}=1+\frac{5}{3}=\frac{8}{3}\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[Y|T=0\right]&=\mathbb{E}\left[Y|T=0,x=0\right]P\left(x=0|T=0\right)+\mathbb{E}\left[Y|T=0,x=1\right]P\left(x=1|T=0\right)=\\&=\mathbb{E}\left[Y|T=0,x=0\right]\frac{P\left(T=0|x=0\right)P\left(x=0\right)}{P\left(T=0\right)}+\mathbb{E}\left[Y|T=0,x=1\right]\frac{P\left(T=0|x=1\right)P\left(x=1\right)}{P\left(T=0\right)}=\\&=0\cdot\frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}}+2\frac{\frac{1}{2}\cdot\frac{2}{3}}{\frac{1}{2}}=2\cdot\frac{2}{3}=\frac{4}{3}\end{aligned}$$

$$\mathbb{E}\left[Y|T=1\right]-\mathbb{E}\left[Y|T=0\right]=\frac{8}{3}-\frac{4}{3}=\frac{4}{3}$$

hence the theorem doesn't hold:

$$\boxed{\mathbb{E}\left[Y_1-Y_0\right]\neq\mathbb{E}\left[Y|T=1\right]-\mathbb{E}\left[Y|T=0\right]}$$

the reason for this is that our assumption that $(Y_0,Y_1)\perp\!\!\!\perp T$ breaks in this scenario (we can prove that).

Question 2:

from the assumption that the treatment is completely random and independent in the possible outcomes, $(Y_0, Y_1) \perp\!\!\!\perp T$ we can write:

$$\mathbb{E}\left[Y_1 - Y_0 | T = 1\right] = \mathbb{E}\left[Y_1 - Y_0\right]$$

from the linearity of the expected value:

$$\mathbb{E}\left[Y_1 - Y_0\right] = \mathbb{E}\left[Y_1\right] - \mathbb{E}\left[Y_0\right]$$

from the proof in the lecture using the consistency assumption $(Y_{obs} = TY_1 + (1 - T)\,Y_0)$ and the independence again $((Y_0, Y_1) \perp\!\!\!\perp T)$ we get:

$$\mathbb{E}\left[Y_1\right] - \mathbb{E}\left[Y_0\right] = \mathbb{E}\left[Y_{obs} | T = 1\right] - \mathbb{E}\left[Y_{obs} | T = 0\right]$$

denote:

$$A_0 = \{i : t_i = 0\}$$

$$|A_0| = n_0$$

$$A_1 = \{i : t_i = 1\}$$

$$|A_1| = n_1$$

$$n_0 + n_1 = n$$

using the law of large numbers:

$$\lim_{n_1 \to \infty} \frac{1}{|A_1|} \sum_{i \in A_1} y_i = \mathbb{E}\left[Y_{obs} | T = 1\right]$$

$$\lim_{n_0 \to \infty} \frac{1}{|A_0|} \sum_{i : \in A_0} y_i = \mathbb{E}\left[Y_{obs} | T = 0\right]$$

hence if the sample size is large enough (in both groups) we get:

$$\mathbb{E}\left[Y_{obs} | T = 1\right] - \mathbb{E}\left[Y_{obs} | T = 0\right] = \frac{1}{|A_1|} \sum_{i \in A_1} y_i - \frac{1}{|A_0|} \sum_{i : \in A_0} y_i$$

thus:

$$\mathbb{E}\left[Y_1 - Y_0 | T = 1\right] = \frac{1}{|A_1|}\sum_{i\in A_1} y_i - \frac{1}{|A_0|}\sum_{i:\in A_0} y_i$$

meaning we can estimate the ATT from the samples, by using three assumptions:

- Consistency: $Y_{obs} = TY_1 + (1-T)\,Y_0$
- Random treatment: $(Y_0, Y_1) \perp\!\!\!\perp T$
- The sample is sufficiently large (in both groups)

Question 3:

let's imagine a data-set of football strikers.height weight

for each striker in the data-set there exists a feature vector which consists of:

$$\phi\left(x_i\right)=\left(\begin{array}{c}\text{age}\\ \text{height}\\ \text{weight}\\ \text{speed}\\ \text{muscle mass}\\ \text{average goals per season through the career}\\ \text{left footer/right footer}\\ \text{main focus of training last year(heading/kicking)}\end{array}\right)$$

and two observed outcome variables will be:

$$y_i=\left(\begin{array}{c}\text{nuber of goals score by foot last season}\\ \text{nuber of goals score by head last season}\end{array}\right)$$

1.

a)

one interesting casual question can be, which type of training is better for the striker, hence will cause him to score more goals next season.

the treatment will be:

$$T=1 : \text{the striker recieves heading training}$$

$$T=0 : \text{the striker recieves kicking training}$$

the potential outcomes will be:

$$Y_0 : \text{number of goals scored by the striker had he recieved kicking training}$$

$$Y_1 : \text{number of goals scored by the striker had he recieved heading training}$$

b)

another interesting question can be whether there is a connection between the striker kicking foot and the number of goals he will score by foot, hence does right/left footers score more goals than the other.

the treatment will be:

$$T = 1 : \text{the striker is a right footer}$$

$$T = 0 : \text{the striker is a left footer}$$

the potential outcomes will be:

$$Y_0 : \text{number of goals scored by foot had the striker been left footer}$$

$$Y_1 : \text{number of goals scored by foot had the striker been right footer}$$

2.

a)

one interesting prediction question can be how many goals, a given striker will score in the next season based on all of the features.

b)

another interesting question can to predict the number of goals the striker will score by head next season, based on his height and speed alone.