Causal Inference

Sofia Olhede

EPFL

December 7, 2020

1. Causal Inference

2. Directed graphs

Exam

- Have just requested from SAC to have a 4 h exam so there is time online to upload.
- Today there is a talk from Bin Yu Verdical data science for the practice of responsible data analysis and decision making at 6 pm
  https://www.epfl.ch/research/domains/cis/prof-bin-yu/ .

Causal Inference V

- Last lecture we started to study causal effects.
- We determined the association and the causal effect were different.
- Random assignments of treatments can help us do so.
- Theorem: suppose that the treatments are assigned randomly to the experimental units. Suppose that $\Pr\{X=0\}>0$ and $\Pr\{X=1\}>0$. In this instance $\alpha=\theta$. Thus any consistent estimator of α is also a consistent estimator of θ. We can therefore take as estimator

$$\widehat{\theta}=\widehat{\mathbb{E}}\{Y|X=1\}-\widehat{\mathbb{E}}\{Y|X=0\}=\overline{Y}_1-\overline{Y}_0.$$

- This is a consistent estimator of θ where

$$\overline{Y}_1=\frac{1}{n_1}\sum_{i=1}^{n}Y_iX_i,$$

$$\overline{Y}_0=\frac{1}{n_0}\sum_{i=1}^{n}Y_i\{1-X_i\}.$$

Causal Inference VI

- With (assuming neither n_0 nor n_1 being zero)

$$n_1 = \sum_{i=1}^{n} X_i, \quad n_0 = \sum_{i=1}^{n} \{1 - X_i\}.$$

- Proof: Since X is being randomly assigned it follows that X is independent of (C_0, C_1).
- Thus we can write

$$\begin{aligned}
\theta &= \mathbb{E}\{C_1\} - \mathbb{E}\{C_0\} \\
&= \mathbb{E}\{C_1 \,|\, X = 1\} - \mathbb{E}\{C_0 \,|\, X = 0\} \\
&= \mathbb{E}\{Y \,|\, X = 1\} - \mathbb{E}\{Y \,|\, X = 0\} \\
&= \alpha.
\end{aligned} \tag{1}$$

Consistency then follows directly from the LLN. □

Causal Inference VII

- The causal effect is defined as:

$$\theta = \mathbb{E}\{C_1\} - \mathbb{E}\{C_0\}.$$

We can define the conditional causal effect as

$$\theta_z = \mathbb{E}\{C_1 \mid Z = z\} - \mathbb{E}\{C_0 \mid Z = z\}.$$

- For example Z might denote the age of the participants in a study. Age is not a value that can be made to take a certain number, but it would be important to capture differences in causal effect given age.

Causal Inference VIII

- Now suppose $X \notin \{0, 1\}$ but instead $X \in \mathcal{X}$.
- We then have a counterfactual function $C(z)$ rather than the counterfactual vector (C_0, C_1).
- The observed response now gives us a consistency relation:

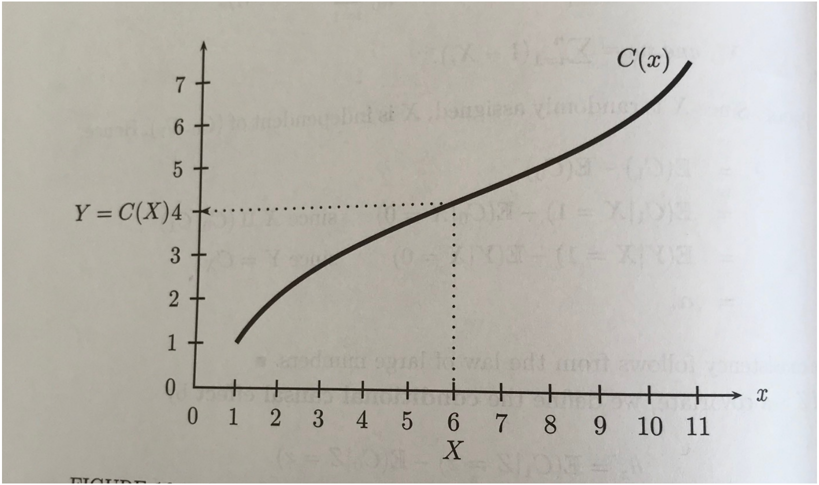
$$Y \equiv C(X).$$

- The causal regression function is

$$\theta(x) = \mathbb{E}\{C(x)\}.$$

  In contrast the regression function that measures association is $r(x) = \mathbb{E}\{Y \mid X = x\}$.
- Theorem: In general $\theta(x) \neq r(x)$. However when X is randomly assigned then $\theta(x) = r(x)$.

Casual regression function



example from 'All of Statistics' by Wasserman

Causal Inference IX

- A study in which the treatment (or exposure) is not randomly assigned, is called an observational study. In such cases, subjects may select or set their own value of the exposure X.
- Many of the health studies we read about in newspapers are of this form.
- Association and causation are in general quite different. This discrepancy is clear in non-randomized studies because the potential outcome C is not independent of the treatment X.
- Suppose that we could find grouping of subjects so that within a group X and $\{C(x) : x \in \mathcal{X}\}$ are independent?
- This would happen if the subjects are similar within a group, e.g. if we find people that have similar ages, gender, educational and ethnic backgrounds.
- For such people we might find it reasonable to assume that the choice of X is essentially random.

Causal Inference X

- Other variables are then called confounding variables. If we write these other variables as Z we can express this idea by saying

ie the observables have to be uncorrelated of X given those other variables.

$$\{C(x) : x \in \mathcal{X}\} \perp\!\!\!\perp X | Z.$$

- This means that within groups of Z the choice of X does not depend on type, written as $\{C(x) : x \in \mathcal{X}\}$.
- If this holds then we say that there is no unmeasured confounding.

Causal Inference XI

- Theorem: suppose that $\{C(x) : \; x \in \mathcal{X}\} \overset{\perp}{\cancel{\amalg}} X|Z$. In this case it follows that

$$\theta(x) = \int \mathbb{E}\{Y|X = x, \, Z = z\}\underline{dF_Z(z)}.$$

Furthermore if $\widehat{r}(x, z)$ is a consistent estimate of the regression function $\mathbb{E}\{Y|X = x, \, Z = z\}$ then a consistent estimator of θ is

$$\widehat{\theta}(x) = \frac{1}{n}\sum_{i=1}^{n} \widehat{r}(x, Z_i).$$

In particular if $r(x, z) = \beta_0 + \beta_1 x + \beta_2 z$ is linear then a consistent estimator of $\theta(x)$ is

$$\widehat{\theta}(x) = \widehat{\beta}_0 + \widehat{\beta}_1 x + \widehat{\beta}_2 \overline{Z}_n,$$

where $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)$ is estimated using least squares.

Causal Inference XII

- $\theta(x)$ is often called the adjusted treatment effect. The process of computing adjusted treatment effects is called adjusting for confounding.
- The selection of what confounders Z to measure and control for requires scientific insight.
- Even after adjusting for confounders we cannot be sure that there are not other confounders that we have missed.
- This is why observational studies must be treated with a lot of scepticism.
- Results from observational studies start to become believable when 1) we can replicate the same result in many studies, 2) each of the studies controls for possible confounders, 3) there is a plausible scientific explanation for the causal relationship.

Causal Inference XIII

- We shall now discuss another common phenomenon in binary data, namely that of Simpson's paradox.
- Let X be a binary treatment variable, Y a binary outcome and Z a third binary variable.

Casual regression function

	$Y=1$	$Y=0$	$Y=1$	$Y=0$
$X=1$	.1500	.2250	.1000	.0250
$X=0$	.0375	.0875	.2625	.1125
	$Z=1$ (men)		$Z=0$ (women)	

The marginal distribution for (X,Y) is

	$Y=1$	$Y=0$	
$X=1$	.25	.25	.50
$X=0$	.30	.20	.50
	.55	.45	1

From these tables we find that,

$$
\begin{aligned}
\mathbb{P}(Y=1|X=1)-\mathbb{P}(Y=1|X=0) &= -0.1\\
\mathbb{P}(Y=1|X=1,Z=1)-\mathbb{P}(Y=1|X=0,Z=1) &= 0.1\\
\mathbb{P}(Y=1|X=1,Z=0)-\mathbb{P}(Y=1|X=0,Z=0) &= 0.1.
\end{aligned}
$$

example from ‘All of Statistics’ by Wasserman

Causal Inference XIV

- There appears to be the case that overall the treatment is negative, but irrespective of your Z it is beneficial. What?
- The issue here is how we translate mathematics into English. $\Pr\{Y = 1 \mid X = 1\} < \Pr\{Y = 1 \mid X = 0\}$ does not mean the treatment is harmful!!!
- The phrase of "the treatment is harmful" should be written mathematically as $\Pr\{C_1 = 1\} < \Pr\{C_0 = 1\}$. The treatment is harmful to $Z = 1$ subjects should be written as$\Pr\{C_1 = 1 \mid Z = 1\} < \Pr\{C_0 = 1 \mid Z = 1\}$. Thus the maths is not contradictory but our interpretation into English is!

Conditional independence and graphs

- Graphs are useful for representing independence relations between variables.
- They can also be used to represent causal relationships.
- We shall use directed acyclic graphs to represent dependence between variables.
- Defn: Let X, Y and Z be random variables, X and Y are conditionally independent given Z (recall lecture 2) written $X \perp\!\!\!\perp Y \mid Z$ if

$$f_{X,Y|Z}(x,y|z) = f_{X|Z}(x|z) f_{Y|Z}(y|z),$$

  for all x, y and z.
- Intuitively once you know Z then Y provides no extra information about X.

Conditional independence and graphs

- Theorem: if all events have positive probability then

$$X \perp\!\!\!\perp Y \mid Z \Rightarrow Y \perp\!\!\!\perp X \mid Z \tag{2}$$

$$X \perp\!\!\!\perp Y \mid Z \,\text{and}\, U = h(X) \Rightarrow U \perp\!\!\!\perp Y \mid Z \tag{3}$$

$$X \perp\!\!\!\perp Y \mid Z \,\text{and}\, U = h(X) \Rightarrow X \perp\!\!\!\perp Y \mid (Z, U) \tag{4}$$

$$X \perp\!\!\!\perp Y \mid Z \,\text{and}\, X \perp\!\!\!\perp W \mid (Y, Z) \Rightarrow X \perp\!\!\!\perp (W, Y) \mid Z \tag{5}$$

$$X \perp\!\!\!\perp Y \mid Z \,\text{and}\, X \perp\!\!\!\perp Z \mid Y \perp\!\!\!\perp (Y, Z). \tag{6}$$