Exercises for Causal Thinking (Math-352)

September 30, 2024

Exercise Sheet 3

Exercise 1 (Causal effects)**.** Based on the definition of a causal effect in the lecture slides, argue whether the following statements about a covariate $L \in \mathbb{R}$, a treatment $A \in \{0,1\}$ and an outcome $Y \in \{0,1\}$ are right or wrong (there is no guarantee that A is randomly assigned).

(i) $\mathbb{E}(Y^{a=1} \mid L = l) - \mathbb{E}(Y^{a=0} \mid L = l)$ is a causal effect.

(ii) $\mathbb{E}(Y \mid A = 1, L = l) - \mathbb{E}(Y \mid A = a, L = l)$ is a causal effect.

(iii) $\mathbb{E}(Y^{a=1} \mid A = 1, L = l) - \mathbb{E}(Y^{a=0} \mid A = 1, L = l)$ is a causal effect.

(iv) $\frac{\mathbb{E}(Y^{a=1})}{\mathbb{E}(Y^{a=0})}$ can be expressed as an expectation of individual level (additive) causal effects.

Solution:

(i) Yes, this is the causal effect in the group with $L = l$

(ii) No, this is not a causal effect. Causal effects should be differences between different counterfactual distributions, where both distributions have the same conditioning set.

(iii) Yes, this is the causal effect in the group with $A = 1, L = l$, i.e. the effect of treatment in the treated, conditional on $L = l$.

(iv) This is not an average of individual level causal effects. There is really no reason it should be such an average. One argument that could be used to prove the existence of a counterexample, is that $\mathbb{E}(Y^{a=1})$ may exceed $\mathbb{E}(Y^{a=0})$ and $\frac{\mathbb{E}(Y^{a=1})}{\mathbb{E}(Y^{a=0})}$ may therefore take a value greater than 1, whereas individual level causal effects are in the range $[-1, 1]$.

Exercise 2 (Translating into maths)**.** Consider an infinite superpopulation of interest. Translate these English sentences into mathematical (counterfactual) statements.

(i) The average causal effect of receiving statins ($A = 1$) vs. placebo ($A = 0$) on mortality after one year ($Y = 1$ is death, $Y = 0$ is alive) in the entire population.

(ii) The average causal effect of receiving statins ($A = 1$) vs. placebo ($A = 0$) on mortality after one year ($Y = 1$ is death, $Y = 0$ is alive) among those who received placebo in the observed data.

(iii) The average causal effect of receiving statins ($A = 1$) vs. placebo ($A = 0$) on mortality after one year ($Y = 1$ is death, $Y = 0$ is alive) among those who received treatment in the observed data.

(iv) The average causal effect of receiving statins ($A = 1$) vs. placebo ($A = 0$) on mortality after one year ($Y = 1$ is death, $Y = 0$ is alive) in men ($L = 1$).

(v) Are your answers in (i)-(v) estimands, estimators or estimates?

Solution:

(i) $\mathbb{E}(Y^{a=1}) - \mathbb{E}(Y^{a=0})$

(ii) $\mathbb{E}(Y^{a=1} \mid A = 0) - \mathbb{E}(Y^{a=0} \mid A = 0)$

(iii) $\mathbb{E}(Y^{a=1} \mid A = 1) - \mathbb{E}(Y^{a=0} \mid A = 1)$

(iv) $\mathbb{E}(Y^{a=1} \mid L = 1) - \mathbb{E}(Y^{a=0} \mid L = 1)$

(v) These quantities are estimands.

Exercise 3 (Conditionally randomized experiment)**.** Suppose investigators had access to data from a study where they observed a binary outcome Y, a binary treatment A and a 4-level baseline covariate L. Certain parameters of the joint distribution of (L, A, Y) were computed from the observed data, which aresummarized in Table 1. We suppose that the sample size was so large, that sampling variability is not a concern.

(i) From the parameters in Table 1, compute $\mathbb{E}[Y]$.

(ii) Suppose now that the data were collected from a special type of randomized trial. Upon recruitment into this trial, each patient's covariate L was measured and then the patients were sorted into groups based on that covariate's value. In each group, the investigators conducted a separate experiment, where a special coin was used to randomly assign patients to either treatment ($A = 1$) or control ($A = 0$), with "heads" corresponding to treatment and "tails" corresponding to control. The probabilities for heads for each of these sub-trials is given by the column labeled $P(A = 1 \mid L = l)$. Assume consistency holds ($A = a \implies Y^a = Y$), and that patients perfectly complied with their assignments. With the information in the table, compute the effect of treatment $\mathbb{E}[Y^{a=1} - Y^{a=0} \mid L = l]$ for each subgroup $L = l$ that was targeted in each of the sub-trials. What additional assumptions did you use along the way, that was justified given the source of the data?

(iii) From the quantities computed in part (i), use laws of probability to compute the average treatment effect in the whole population, $\mathbb{E}[Y^{a=1} - Y^{a=0}]$.

(iv) Suppose that the data analyst for the study approached you and said they made a terrible mistake: when preparing the column $P(A = 1 \mid L = l)$ in Table 1, they reverse coded the treatment variable, so in fact the true values of the treatment propensities are 1 minus those listed in the table. What will be the values of the previously computed parameters, and explain in words why these changes did (or did not occur).

	$P(Y = 1 \mid A = a, L = l)$		$P(A = 1 \mid L = l)$	$P(L = l)$
	$a = 1$	$a = 0$		
$l = 1$	0.1	0.8	0.2	0.2
$l = 2$	0.2	0.7	0.4	0.4
$l = 3$	0.3	0.6	0.6	0.1
$l = 4$	0.4	0.5	0.8	0.3

Table 1: Parameters of $P_{L,A,Y}$ observed in the conditionally randomized trial.

Solution:

(i) We compute $\mathbb{E}[Y]$ as follows:

$$\mathbb{E}[Y] = P(Y = 1) \qquad \text{as } Y \text{ is binary}$$

$$= \sum_{l \in \{1,2,3,4\}} P(Y = 1 \mid L = l) P(L = l) \qquad \text{by LOTP}$$

$$= \sum_{l \in \{1,2,3,4\}} P(L = l) \left[\sum_{a \in \{0,1\}} P(Y = 1 \mid A = a, L = l) P(A = a \mid L = l) \right] \qquad \text{by LOTP}$$

Entering the values from table 1 into this expression yields:

$\mathbb{E}[Y] = 0.2 \times (0.1 \times 0.2 + 0.8 \times 0.8)$
$+ 0.4 \times (0.2 \times 0.4 + 0.7 \times 0.6)$
$+ 0.1 \times (0.3 \times 0.6 + 0.6 \times 0.4)$
$+ 0.3 \times (0.4 \times 0.8 + 0.5 \times 0.2)$
$= 0.5$

(ii)
- For $l = 1$: 0.1-0.8 = -0.7
- For $l = 2$: 0.2-0.7 = -0.5
- For $l = 3$: 0.3-0.6 = -0.3
- For $l = 4$: 0.4-0.5 = -0.1
- We relied on an assumption of conditional exchangeability, $Y^a \perp\!\!\!\perp A|L$ for all a. When exchangeability holds, $\mathbb{E}[Y^a \mid L = l] = \mathbb{E}[Y^a \mid L = l, A = a]$ (which is equal to $\mathbb{E}[Y \mid L = l, A = a]$ assuming consistency). Exchangeability is justified because treatment assignment was randomly assigned conditional on l.

(iii) $-0.7 \times 0.2 - 0.5 \times 0.4 - 0.3 \times 0.1 - 0.1 \times 0.3 = -0.4$

(iv) For part (i), $0.2 \times (0.1 \times 0.8 + 0.8 \times 0.2)$
$+ 0.4 \times (0.2 \times 0.6 + 0.7 \times 0.4)$
$+ 0.1 \times (0.3 \times 0.4 + 0.6 \times 0.6)$
$+ 0.3 \times (0.4 \times 0.2 + 0.5 \times 0.8)$
$= 0.4$. The conditional effects of treatment from from part (ii) are unaffected. The marginal effect of treatment from part (iii) is also unaffected. These quantities are not functions of the propensities.

Exercise 4 (Marginal vs. conditional independence)**.** Consider a covariate $L \in \mathbb{R}$, a treatment $A = 0, 1$ and an outcome $Y \in \{0, 1\}$.

(i) Investigator 1 claims that $A \perp\!\!\!\perp Y \implies A \perp\!\!\!\perp Y \mid L$. Argue (or show) that the statement is wrong.

(ii) Investigator 2 claims that $A \perp\!\!\!\perp Y \mid L \implies A \perp\!\!\!\perp Y$. Argue (or show) that the statement is wrong.

Solution:

(i) Any numerical counterexample (i.e. where $A \perp\!\!\!\perp Y$ but $A \perp\!\!\!\perp Y|L$ does not hold) is sufficient to answer this question. You can also provide an intuitive answer, for example that in a dataset that contains data about whether it rains, whether the sprinkler is on, and whether the grass is wet (in a setting where the sprinkler has been set to turn itself on or off automatically while disregarding the weather), the state of the sprinkler will be independent of whether it rains. However, when conditioning on whether the grass is wet, these variables will be highly dependent (because it the grass is wet when it does not rain, the sprinkler is very likely to be on). For example, consider the joint distribution with binary L, A, Y in Table 2.

(ii) For part (ii) any numerical counterexample is again sufficient, For example, consider the joint distribution with binary L, A, Y in Table 3.

Table 2: Counterexample to $A \perp\!\!\!\perp Y \implies A \perp\!\!\!\perp Y \mid L$

	$P(Y=1 \mid A=a)$		$P(Y=1 \mid A=a, L=l)$		$P(L=l \mid A=a)$	
	$a=1$	$a=0$	$a=1$	$a=0$	$a=1$	$a=0$
$l=1$	0.5	0.5	0.25	0.75	0.5	0.5
$l=2$			0.75	0.25	0.5	0.5

Table 3: Counterexample to $A \perp\!\!\!\perp Y \mid L \implies A \perp\!\!\!\perp Y$

	$P(Y=1 \mid A=a)$		$P(Y=1 \mid A=a, L=l)$		$P(L=l \mid A=a)$	
	$a=1$	$a=0$	$a=1$	$a=0$	$a=1$	$a=0$
$l=1$	0.65	0.5	0.2	0.2	0.25	0.5
$l=2$			0.8	0.8	0.75	0.5

Exercise 5 (Effect modification)**.** Let $Y \in \{0,1\}$ be an indicator of having a heart attack within 5 years, and suppose that V and L are two different binary random variables denoting sex and underlying cardiovascular disease respectively. We denote treatment with $A \in \{0,1\}$ ($A=1$ indicates statins and A=0 indicates no treatment). Suppose that

1. for every $a \in \{0,1\}$, $Y^a \perp\!\!\!\perp A \mid L, V$ (Exchangeability).

2. for every $a \in \{0,1\}, l \in \mathcal{L}, v \in \mathcal{V}$

$$P(L=l, V=v) > 0 \implies P(A=a \mid L=l, V=v) > 0 \text{ (Positivity)}.$$

3. $A=a \implies Y^a = Y$ (Consistency).

(a) Show or argue that the average causal effect of $A=a$ in the stratum defined by $V=v$ is

$$\mathbb{E}(Y^a \mid V=v) = \sum_l \mathbb{E}(Y \mid L=l, V=v, A=a) P(L=l \mid V=v).$$

(b) Next, suppose L and V are binary random vectors and that $V \subset L$. Suppose conditions 2.3 are still valid with L and V as random vectors. State a sufficient exchangeability condition to identify $\mathbb{E}(Y^a \mid V = v)$, and write out the identification formula for $\mathbb{E}(Y^a \mid V = v)$.

Solution:

(a)

$$
\begin{aligned}
&\mathbb{E}(Y^a \mid V = v) \\
&\overset{\text{LOTP}}{=} \sum_l \mathbb{E}(Y^a \mid L = l, V = v) P(L = l \mid V = v) \\
&\overset{Y^a \perp\!\!\!\perp A \mid L, V}{=} \sum_l \mathbb{E}(Y^a \mid L = l, V = v, A = a) P(L = l \mid V = v) \\
&\overset{Consistency}{=} \sum_l \mathbb{E}(Y \mid L = l, V = v, A = a) P(L = l \mid V = v)
\end{aligned}
$$

(b) $Y^a \perp\!\!\!\perp A \mid V$ is a sufficient condition:

$$
\begin{aligned}
&\mathbb{E}(Y^a \mid V = v) \\
&\overset{Y^a \perp\!\!\!\perp A \mid V}{=} \mathbb{E}(Y^a \mid V = v, A = a) \\
&\overset{Consistency}{=} \mathbb{E}(Y \mid V = v, A = a)
\end{aligned}
$$

Exercise 6 (Average effect in the treated)**.** Consider a scenario where adults with appendicitis are either treated with appendectomy (surgical removal of the appendix) or antibiotics. Let $A \in \{0, 1\}$ be an indicator for the actual treatment received ($A = 1$ indicates surgery), $Y \in \{0, 1\}$ an indicator for the presence of an allergic reaction ($Y = 1$ indicates an allergic reaction), and $L \in \mathcal{L}$ a discrete random vector representing measured baseline covariates. Assume that allergic reactions can only arise from anesthetics *used in surgery* (these anesthetics are not given to those who receive antibiotic treatment). Finally, let $U \in \{0, 1\}$ be an unmeasured indicator for the presence of a gene which gives increased risk of allergic reactions against anesthetics (presence of the gene is indicated by $U = 1$). Some individuals with $U = 1$ choose to receive antibiotic therapy instead of surgery, because they have family members who had allergic reactions during surgery. Therefore, assume that U has a direct effect on the treatment received.

Suppose we observe A, L, Y: that is, allergic reactions Y under treatment A with baseline covariates L. The average effect in the treated is defined as the causal contrast

$$\mathbb{E}[Y^{a=0} \mid A = 1] \text{ vs. } \mathbb{E}[Y^{a=1} \mid A = 1] \ .$$

Throughout the question, you may assume that consistency and positivity hold:

1. For every $a \in \{0, 1\}, l \in \mathcal{L}$, $P(L = l) > 0 \Longrightarrow P(A = a \mid L = l) > 0$ (Positivity).
2. $A = a \Longrightarrow Y^a = Y$ (Consistency).

(a) Show that the estimand $E[Y^a \mid A = 1]$ is identified under the weaker exchangeability condition $Y^{a=0} \perp\!\!\!\perp A \mid L$ (without assuming that $Y^{a=1} \perp\!\!\!\perp A \mid L$), and show that the corresponding identification formula is

$$\mathbb{E}[Y^a \mid A = 1] = \sum_l \mathbb{E}[Y \mid A = a, L = l]P(L = l \mid A = 1) \ . \tag{1}$$

(b) *Challenge:* For $a = 0$, show that the previous equation can be written on inverse probability weighted form as

$$\mathbb{E}[Y^a \mid A = 1] = \frac{\mathbb{E}\left[\frac{I(A=a)Y}{\pi(A|L)} P(A = 1 \mid L)\right]}{\mathbb{E}\left[\frac{I(A=a)}{\pi(A|L)} P(A = 1 \mid L)\right]} \ . \tag{2}$$

(c) What is the average treatment effect in the treated in a study where A is randomized?

(d) *Challenge:* Let A, L, Y be binary random variablers denoting treatment, a baseline covariate, and outcome, respectively. An investigator conducts a thought experiment where she matches every individual with $A = 1, L = 0$ to a randomly chosen individual with $A = 0, L = 0$ (and likewise matches every individual with $A = 1, L = 1$ to an individual with $A = 0, L = 1$) in an infinitely large sample of the population. Finally, the investigator computes the expectation of Y in the matched population. Which estimand is this equivalent to?

Solution:

(a) We begin by using the law of total probability:

$$P(Y^0 = 1 \mid A = 1) = \sum_l P(Y^0 = 1, L = l \mid A = 1)$$

$$= \sum_l P(Y^0 = 1 \mid L = l, A = 1) P(L = l \mid A = 1)$$

$$\overset{Y^{a=0} \perp\!\!\!\perp A \mid L}{=} \sum_l P(Y^0 = 1 \mid L = l) P(L = l \mid A = 1)$$

$$\overset{Y^{a=0} \perp\!\!\!\perp A \mid L}{=} \sum_l P(Y^0 = 1 \mid L = l, A = 0) P(L = l \mid A = 1)$$

$$\overset{\text{consistency, positivity}}{=} \sum_l P(Y = 1 \mid L = l, A = 0) P(L = l \mid A = 1) \ .$$

The estimand $P(Y^1 = 1 \mid A = 1)$ is identified straightforwardly from consistency and the law of total probability:

$$P(Y^1 = 1 \mid A = 1) \overset{\text{consistency}}{=} P(Y = 1 \mid A = 1)$$

$$\overset{\text{LOTP}}{=} \sum_l P(Y = 1 \mid L = l, A = 1) P(L = l \mid A = 1) \ .$$

We have thus derived Eq. 1.

(b) The denominator of Eq. 2 is equal to

$$\mathbb{E}\left[\frac{I(A=0)}{\pi(A \mid L)} P(A = 1 \mid L)\right] = \sum_{\tilde{a},l} P(A = \tilde{a}, L = l) \frac{I(\tilde{a} = 0)}{P(A = \tilde{a} \mid L = l)} P(A = 1 \mid L = l)$$

$$= \sum_{\tilde{a},l} P(L = l) I(\tilde{a} = 0) P(A = 1 \mid L = l)$$

$$= \left(\sum_{\tilde{a}} I(\tilde{a} = 0)\right) \left(\sum_l P(A = 1 \mid L = l) P(L = l)\right)$$

$$= P(A = 1)$$

and the numerator is $\mathbb{E}\left[\frac{I(A=0)Y^0}{\pi(A|L)} P(A = 1 \mid L)\right]$ by consistency. Furthermore, let us denote $P(L = l \mid A = 1)$ with $p(l \mid 1)$ and $P(L = l)$ with $p(l)$:

$$\frac{\mathbb{E}\left[\frac{I(A=0)Y^0}{\pi(A|L)} P(A = 1 \mid L)\right]}{P(A = 1)} = \mathbb{E}\left\{\mathbb{E}\left[\frac{I(A = 0)Y^0}{\pi(A \mid L)} \frac{P(A = 1 \mid L)}{P(A = 1)} \;\middle|\; L\right]\right\}$$

$$\overset{\text{Bayes}}{=} \mathbb{E}\left\{\mathbb{E}\left[\frac{I(A = 0)Y^0}{\pi(A \mid L)} \frac{p(L \mid 1)}{p(L)} \;\middle|\; L\right]\right\}$$

$$= \mathbb{E}\left\{\mathbb{E}\left[\frac{I(A = 0)Y^0}{\pi(A \mid L)} \;\middle|\; L\right] \frac{p(L \mid 1)}{p(L)}\right\}$$

$$\overset{Y^{a=0} \perp\!\!\!\perp A \mid L}{=} \mathbb{E}\left\{\underbrace{\mathbb{E}\left[\frac{I(A = 0)}{\pi(A \mid L)} \;\middle|\; L\right]}_{=1} E[Y^0 \mid L] \frac{p(L \mid 1)}{p(L)}\right\}$$

$$
\begin{aligned}
&= \mathbb{E}\left\{\mathbb{E}\left[Y^0 \frac{p(L \mid 1)}{p(L)} \;\middle|\; L\right]\right\} \\
&= \sum_l P(L=l) \sum_y P(Y^0 = y \mid L = l) \cdot y \cdot \frac{P(L = l \mid A = 1)}{P(L = l)} \\
&= \sum_l \sum_y y \cdot P(Y^0 = y \mid L = l) P(L = l \mid A = 1) \\
&\overset{\text{LOTP}}{=} \mathbb{E}[Y^0 \mid A = 1] \ .
\end{aligned}
$$

(c) In a randomized trial, the average treatment effect of the treated is equal to the total effect since individuals with $A = 0$ and $A = 1$ are exchangeable, i.e. $Y^a \perp\!\!\!\perp A$ for $a \in \{0, 1\}$. We show this below:

$$
\begin{aligned}
\mathbb{E}[Y^{a=1} - Y^{a=0} \mid A = 1] &\overset{Y^a \perp\!\!\!\perp A}{=} E[Y^{a=1} - Y^{a=0}] \\
&\overset{Y^a \perp\!\!\!\perp A}{=} \mathbb{E}[Y^{a=1} \mid A = 1] - \mathbb{E}[Y^{a=0} \mid A = 0] \\
&\overset{\text{consistency}}{=} \mathbb{E}[Y \mid A = 1] - \mathbb{E}[Y \mid A = 0] \ .
\end{aligned}
$$

(d) $\mathbb{E}[Y^{a=0} \mid A = 1]$.
You can see Section 4.5 of Hernan and Robins [2020] for more details about matching, which is another form of adjustment. The goal is to construct a subset of the population in which the variables L have the same distribution in both the treated ($A = 1$) and the untreated ($A = 0$). For each individual i with $A_i = 1$ and $L = \ell$, sample individuals j with $A_j = 0$ among those who have $L = \ell$, and look at the outcome Y_j. Repeat this n times and average on Y_j. Let n goes to infinity, and use consistency and exchangeability.

References

Miguel Hernan and James M. Robins. *Causal Inference: What If*. Chapman & Hall, 2020.