EXERCISES FOR RANDOMIZATION AND CAUSATION (MATH-336)

Exercise Sheet 12

Exercise 1. In this exercise, you will study partial identification (bounds) of the average treatment effect. Suppose that Z, A, Y, U satisfy the single-world causal model corresponding to the graph below. Suppose that the measured variables $Z, A, Y \in \{0, 1\}$ are binary.

(a) Show that, without using Z, the average treatment effect of A and Y satisfies the following inequalities

$$-P(Y=0, A=1) - P(Y=1, A=0) \leq \mathbb{E}(Y^{a=1} - Y^{a=0}) \leq P(Y=1, A=1) + P(Y=0, A=0).$$

What is the difference between the upper and the lower bounds $(UB - LB)$?

(b) Suppose $A = 1$ if an individual elects to get the annual influenza vaccine and $A = 0$ otherwise. Let $Y^a = 1$ if an individual subsequently does develop flu-like symptoms when $A = a$, and $Y^a = 0$ otherwise. Suppose that the investigator is comfortable with assuming that each individual is more or as likely to develop flu-like symptoms if they are unvaccinated versus if they are vaccinated.[1]

(i) Formalize the investigator's assumption as a counterfactual inequality.

(ii) What is the upper bound on $\mathbb{E}(Y^{a=1} - Y^{a=0})$ under this assumption?

(iii) Can we derive a tighter lower bound without adding additional assumptions?

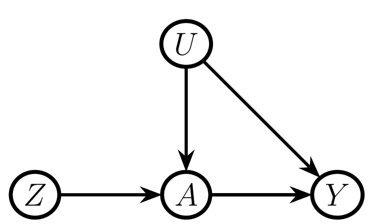
(c) Now you will show some famous bounds using the instrumental variable Z. Suppose that necessary consistency and positivity assumptions hold. Let $p(y, a \mid z)$ denote $P(Y = y, A = a \mid Z = z)$ and $p(y \mid z)$ denote $P(Y = y \mid Z = z)$. Show that

$$LB \leq \mathbb{E}(Y^{a=1} - Y^{a=0}) \leq UB,$$

where

$$\begin{aligned} LB = \max\{ & -p(0,1 \mid 0) - p(1,0 \mid 0), \\ & -p(0,1 \mid 1) - p(1,0 \mid 1), \\ & p(1 \mid 0) - p(1 \mid 1) - p(1,0 \mid 0) - p(0,1 \mid 1), \\ & p(1 \mid 1) - p(1 \mid 0) - p(1,0 \mid 1) - p(0,1 \mid 0)\}, \end{aligned}$$

[1] In this exercise we ignore interference, and suppose that individuals are iid and that positivity and consistency hold.



and

$$\begin{aligned}UB = \min\{&p(1,1 \mid 0) + p(0,0 \mid 0),\\ &p(1,1 \mid 1) + p(0,0 \mid 1),\\ &p(1 \mid 0) - p(1 \mid 1) + p(0,0 \mid 0) - p(1,1 \mid 1),\\ &p(1 \mid 1) - p(1 \mid 0) + p(0,0 \mid 1) + p(1,1 \mid 0)\},\end{aligned}$$

Conclude that

$$UB - LB \leq \min\{P(A=0 \mid Z=0) + P(A=1 \mid Z=1), P(A=0 \mid Z=1) + P(A=1 \mid Z=0)\} \leq 1.$$

and that $UB - LB = 1$ if and only if $A \perp\!\!\!\perp Z$.

Solution:

(a)

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

Now consider,

$$\begin{aligned}\mathbb{E}(Y^{a=1}) &= \mathbb{E}(Y^{a=1} \mid A=1)P(A=1) + \mathbb{E}(Y^{a=1} \mid A=0)P(A=0)\\ &= \mathbb{E}(Y \mid A=1)P(A=1) + \mathbb{E}(Y^{a=1} \mid A=0)P(A=0)\\ &= P(Y=1 \mid A=1)P(A=1) + P(Y^{a=1}=1 \mid A=0)P(A=0)\\ &= P(Y=1, A=1) + (1 - P(Y^{a=1}=0 \mid A=0))P(A=0)\\ &= P(Y=1, A=1) + P(A=0) - P(Y^{a=1}=0 \mid A=0)P(A=0)\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbb{E}(Y^{a=0}) &= P(Y=1 \mid A=0)P(A=0) + P(Y^{a=0}=1 \mid A=1)P(A=1)\\ &= P(A=0) - P(Y=0 \mid A=0)P(A=0) + P(Y^{a=0}=1 \mid A=1)P(A=1)\\ &= P(A=0) - P(Y=0, A=0) + P(Y^{a=0}=1 \mid A=1)P(A=1)\\ \therefore \mathbb{E}(Y^{a=1}) - \mathbb{E}(Y^{a=0}) &= P(Y=1, A=1) + P(A=0) - P(Y^{a=1}=0 \mid A=0))P(A=0)\\ &\quad - (P(A=0) - P(Y=0, A=0) + P(Y^{a=0}=1 \mid A=1)P(A=1))\\ &= P(Y=1, A=1) + P(Y=0, A=0) - P(Y^{a=1}=0 \mid A=0)P(A=0)\\ &\quad - P(Y^{a=0}=1 \mid A=1)P(A=1)\\ &\leq P(Y=1, A=1) + P(Y=0, A=0)\end{aligned}$$

We thus have the upper bound.

To get the lower bound

$$\begin{aligned}\mathbb{E}(Y^{a=1}) &= P(Y=1 \mid A=1)P(A=1) + P(Y^{a=1}=1 \mid A=0)P(A=0)\\ &= P(A=1) - P(Y=0 \mid A=1)P(A=1) + P(Y^{a=1}=1 \mid A=0)P(A=0)\\ \mathbb{E}(Y^{a=0}) &= P(Y=1 \mid A=0)P(A=0) + P(Y^{a=0} \mid A=1)P(A=1)\\ &= P(Y=1, A=0) + P(A=1) - P(Y^{a=0}=0 \mid A=1)P(A=1)\end{aligned}$$

$$\begin{aligned}\therefore \mathbb{E}(Y^{a=1}) - \mathbb{E}(Y^{a=0}) &= P(A=1) - P(Y=0 \mid A=1)P(A=1) + P(Y^{a=1}=1 \mid A=0)P(A=0) \\ &\quad - (P(Y=1, A=0) + P(A=1) - P(Y^{a=0}=0 \mid A=1)P(A=1)) \\ &= -P(Y=0, A=1) - P(Y=1, A=0) + P(Y^{a=1}=1 \mid A=0)P(A=0) \\ &\quad + P(Y^{a=0}=0 \mid A=1)P(A=1) \\ &\geq -P(Y=0, A=1) - P(Y=1, A=0).\end{aligned}$$

The difference between the bounds, $UB - LB = P(Y=1, A=1) + P(Y=0, A=0) + P(Y=1, A=0) + P(Y=0, A=1) = 1$

(b) (a) The investigator's assumption translates to $\mathbb{P}(Y^{a=1}=0) \geq \mathbb{P}(Y^{a=0}=0)$

(b) Using this assumption, the upper bound for $\mathbb{E}(Y^{a=1} - Y^{a=0})$ is 0.

(c) The naïve lower bound is $-1 \leq \mathbb{E}(Y^{a=1} - Y^{a=0})$, if everyone who takes the influenza vaccine has no flu-like symptoms, and everyone who does not take the vaccine develops flu-like symptoms. This bound cannot be made tighter.

(c) Since $Y^a \perp\!\!\!\perp Z$, we can do the same process as earlier for $\mathbb{E}(Y^{a=1} - Y^{a=0} \mid Z = z)$ instead of $\mathbb{E}(Y^{a=1} - Y^{a=0})$. We thus get the first two components of LB and UB. To get the other two components, consider the following:

$$\begin{aligned}\mathbb{E}(Y^{a=1}) &= \mathbb{E}(Y^{a=1} \mid Z) \\ &= P(Y^{a=1}=1 \mid Z) \\ &= P(Y^{a=1}=1, A=1 \mid Z) + P(Y^{a=1}=1, A=0 \mid Z) \\ &\geq P(Y^{a=1}=1, A=1 \mid Z) \\ &= P(Y=1, A=1 \mid Z)\end{aligned}$$

Also,

$$\begin{aligned}\mathbb{E}(Y^{a=1}) &= \mathbb{E}(Y^{a=1} \mid Z) \\ &= 1 - P(Y^{a=1}=0 \mid Z) \\ &= 1 - P(Y^{a=1}=0, A=1 \mid Z) - P(Y^{a=1}=0, A=0 \mid Z) \\ &\leq 1 - P(Y^{a=1}=0, A=1 \mid Z) \\ &= 1 - P(Y=0, A=1 \mid Z)\end{aligned}$$

So,

$$p(1,1 \mid z) \leq \mathbb{E}(Y^{a=1}) \leq 1 - p(0,1 \mid z)$$

Similarly,

$$\begin{aligned}p(1,0 \mid z') \leq \mathbb{E}(Y^{a=0}) \leq 1 - p(0,0 \mid z') \\ -(1 - p(0,0 \mid z')) \leq -\mathbb{E}(Y^{a=0}) \leq -p(1,0 \mid z')\end{aligned}$$

Take the cross terms, $z \neq z'$,

$$\begin{aligned}\mathbb{E}(Y^{a=1} - Y^{a=0}) &\geq p(1,1 \mid z) - (1 - p(0,0 \mid z')) \\ &= p(1,1 \mid z) + p(0,0 \mid z') - 1\end{aligned}$$

$$= p(1,1 \mid z) + p(1,0 \mid z) - p(1,0 \mid z) + p(0,0 \mid z') + p(0,1 \mid z') - p(0,1 \mid z') - 1$$
$$= p(1 \mid z) - p(1,0 \mid z) + p(0 \mid z') - p(0,1 \mid z') - 1$$
$$= p(1 \mid z) - (1 - p(0 \mid z')) - p(1,0 \mid z) - p(0,1 \mid z')$$
$$= p(1 \mid z) - p(1 \mid z')) - p(1,0 \mid z) - p(0,1 \mid z').$$

Exercise 2 (Efficiency of linear adjustment)**.** (Inspired by [1]) Consider 3 different linear models defined by population least squares,

$$\beta^* = \arg\min_{\beta} \mathbb{E}\big[(Y - \beta_1 - \beta_2 A)^2\big]$$
$$\beta' = \arg\min_{\beta} \mathbb{E}\big[(Y - \beta_1 - \beta_2 A - \beta_3^T L)^2\big] \text{ (ANCOVA model)}$$
$$\beta^\dagger = \arg\min_{\beta} \mathbb{E}\big[(Y - \beta_1 - \beta_2 A - \beta_3^T L - \beta_4^T AL)^2\big]$$

Suppose (L, A, Y) are i.i.d., $A \perp\!\!\!\perp L$, $\mathbb{E}(L) = 0$.

(a) Show that[2] $\beta_1^* = \beta_1' = \beta_1^\dagger$ and $\beta_2^* = \beta_2' = \beta_2^\dagger$.

(b) A classical result from M-estimation theory implies that

$$\sqrt{n}(\widehat{\beta}_1^m - \beta_1) \xrightarrow{d} N(0, V^m),$$

where $m \in \{*, ', \dagger\}$, $\pi = P(A = a \mid L)$, $V^m = \frac{E[(A-\pi)^2 \epsilon_m^2]}{\pi^2(1-\pi)^2}$ and $\epsilon_{i*}, \epsilon_{i'}, \epsilon_{i\dagger}$ are the error terms in the regression estimators, for example,

$$\epsilon_{i,\dagger} = Y_i - (\beta_1^\dagger + \beta_2^\dagger A_i + \beta_3^{\dagger T} L_i + \beta_4^{\dagger T} A_i L_i) .$$

Use this result to show that

$$V^\dagger \leq \min\{V', V^*\} .$$

In other words, asymptotically it is more efficient to use covariates L in the model indicated by $\dagger$.[3]

Solution:

(a) Consider first the largest model (specified by $\beta^\dagger$). By taking partial derivatives wrt. $\beta_1^\dagger$ and $\beta_2^\dagger$ we have

$$\mathbb{E}\big[Y - (\beta_1^\dagger + \beta_2^\dagger A + \beta_3^{\dagger T} L + \beta_4^{\dagger T} AL)\big] = 0.$$
$$\mathbb{E}\big[A(Y - (\beta_1^\dagger + \beta_2^\dagger A + \beta_3^{\dagger T} L + \beta_4^{\dagger T} AL))\big] = 0.$$

Next, using the fact that $A \perp\!\!\!\perp L$, $\mathbb{E}(L) = 0$ gives

$$\mathbb{E}\big[Y - \beta_1^\dagger - \beta_2^\dagger A\big] = 0,$$
$$\mathbb{E}\big[A(Y - \beta_1^\dagger - \beta_2^\dagger A)\big] = 0.$$

[2] We have not said anything about the linear model being correctly specified. We have not given an argument why $\mathbb{E}(L) = 0$. However, we could center L_i by using $L_i - \frac{1}{n}\sum_{i=1}^n L_i$, which will give the same point estimates of the β's but $\beta^\dagger$ has larger variance.

[3] Careful consideration is required to decide whether or not it is more efficient to use L in a finite sample.

We find exactly the same equations when we started by the models specified by β and β'. By WLLN we would expect, under regularity conditions, that the maximum likelihood estimator $\hat{\beta}^\dagger$ converges to $\beta^\dagger$.

(b) By taking partial derivatives $\frac{\partial}{\partial \beta_1}, \ldots, \frac{\partial}{\partial \beta_4}$ of $\mathbb{E}\big[(Y - \beta_1 - \beta_2 A - \beta_3^T L - \beta_4^T AL)^2\big]$, we find

(1) $$\mathbb{E}(\epsilon_{i\dagger}) = \mathbb{E}(A\epsilon_{i\dagger}) = \mathbb{E}(L\epsilon_{i\dagger}) = \mathbb{E}(AL\epsilon_{i\dagger}) = 0 \ .$$

Then, by the theorem on the equalities of the β's,

$$\epsilon_* = \epsilon_\dagger + \beta_3^{\dagger T} L + \beta_4^{\dagger T} AL$$
$$\epsilon_{\prime} = \epsilon_\dagger + (\beta_3^{\dagger T} - \beta_3^{\prime T})L + \beta_4^{\dagger T} AL$$

Eq. 1 implies that

$$\mathrm{Cov}(\epsilon_\dagger, \beta_3^{\dagger T} L + \beta_4^{\dagger T} AL) = 0$$
$$\mathrm{Cov}(\epsilon_\dagger, (\beta_3^{\dagger T} - \beta_3^{\prime T})L + \beta_4^{\dagger T} AL) = 0$$

Using the summation law of variances, for $m \in \{*, '\}$

$$\mathbb{E}(\epsilon_{i\dagger}^2) \leq \mathbb{E}(\epsilon_{im}^2),$$

which concludes the argument, because

$$V_m = \frac{E[(A-\pi)^2 \epsilon_m^2]}{\pi^2(1-\pi)^2} \stackrel{A \perp\!\!\!\perp \epsilon_m}{=} \frac{E[(A-\pi)^2]E[\epsilon_m^2]}{\pi^2(1-\pi)^2} \ .$$

The independence $A \perp\!\!\!\perp \epsilon_m$ follows from

$$\frac{\partial}{\partial \beta_1} E[(Y - \beta_1 - \beta_2 A] = 0,$$
$$\frac{\partial}{\partial \beta_1} E[(Y - \beta_1 - \beta_2 A - \beta_3^T L] = 0 \ .$$

References

[1] Qingyuan Zhao. Lecture Notes on Causal Inference. page 109.