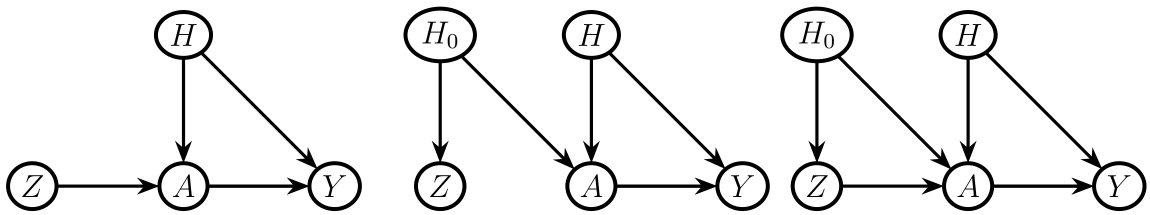
EXERCISES FOR RANDOMIZATION AND CAUSATION (MATH-336)

Exercise Sheet 11

Exercise 1 (Instrumental variables)**.** (From [1]) Consider an instrumental variable setting which is described by one of the following three DAGs.



(a) Can we use the main IV assumptions (1)-(3) to infer any (conditional) independencies between the observed variables A, Z, Y, that is, any factorization of the law $p(y, a, z)$ that describes the observed data? We reproduce the main IV assumptions below for convenience:
(1) $\text{cor}(Z, A) \neq 0$ (instrument strength)
(2) $Y^{z,a} = Y^a$ for all a, z (exclusion restriction)
(3) $Z \perp\!\!\!\perp Y^a$ for all a (unconfoundedness of Z).

(b) Consider the following structural equation model for Y:

(1) $$Y = f_Y(A, H, \epsilon_Y) = h(\epsilon_Y)A + g(H, \epsilon_Y) \ .$$

The model does allow certain effect heterogeneity, because the individual level causal effect

$$Y^a - Y^{a'} = h(\epsilon_Y)(a - a')$$

is a random variable. The average causal effect is defined as

$$E[Y^a] - E[Y^{a'}] = E[h(\epsilon_Y)](a - a') \ .$$

Assume that the linear structural equation model Eq. 1 holds, that $Y^{a=0} \perp\!\!\!\perp Z$ and that $E[h(\epsilon_Y) \mid Z, A] = E[h(\epsilon_Y)]$. Show that the additive average causal effect is then given by

$$E[h(\epsilon_Y)] = \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, A)} \ .$$

(c) Assume that the model in Eq. 1 holds, and that $E[h(\epsilon_Y) \mid Z, A] = E[h(\epsilon_Y)]$. Show that then, there exists a constant β such that

$$E[Y \mid Z, A] - E[Y^0 \mid Z, A] = \beta A \ .$$

Solution:

(a) We need to verify the following six (conditional) independencies:

(a) $Y \perp\!\!\!\perp A$ (fails because of the path $A \to Y$)
(b) $Y \perp\!\!\!\perp A \mid Z$ (fails because of the path $A \to Y$)
(c) $Y \perp\!\!\!\perp Z$ (fails because of the path $Z \to A \to Y$ in the left DAG)
(d) $Y \perp\!\!\!\perp Z \mid A$ (fails because of the path $Z \to A \leftarrow H \to Y$)
(e) $A \perp\!\!\!\perp Z$ (fails because of the path $Z \to A$ in the left DAG)
(f) $A \perp\!\!\!\perp Z \mid Y$ (fails because of the path $Z \to A$ in the left DAG)

As there exist counterexamples which violate every one of the independencies above, while satisfying the IV assumptions (1)-(3), we can conclude that no such independencies are implied by the main IV assumptions and thus the law $p(a, z, y)$ factorizes as for a complete DAG:

$$p(a, z, y) = p(y \mid a, z)p(a \mid z)p(z) \ .$$

(b) Under Eq. 1,

$$Y - Y^{a=0} = h(\epsilon_Y)A$$

so

$$Y - h(\epsilon_Y)A = Y^{a=0}$$

and consequently

$$E[Y - h(\epsilon_Y)A \mid Z] = E[Y^{a=0} \mid Z] \overset{Y^{a=0} \perp\!\!\!\perp Z}{=} E[Y^{a=0}] \ .$$

Using the law of total expectation,

$$\begin{aligned} E[(Z - E[Z])(Y - h(\epsilon_Y)A)] &= E[E[(Z - E[Z])(Y - h(\epsilon_Y)A) \mid Z]] \\ &= E[(Z - E[Z])E[\underbrace{(Y - h(\epsilon_Y)A)}_{=Y^{a=0}} \mid Z]] \\ &\overset{Y^{a=0} \perp\!\!\!\perp Z}{=} E[(Z - E[Z])]E[(Y - h(\epsilon_Y)A)] \\ &= 0 \ . \end{aligned}$$

Likewise,

$$\begin{aligned} E[(Z - E[Z])(Y - h(\epsilon_Y)A)] &= \underbrace{E[(Z - E[Z])Y]}_{=\mathrm{Cov}(Z,Y)} - E[(Z - E(Z))h(\epsilon_Y)A] \\ &= \mathrm{Cov}(Z, Y) - E[E[(Z - E(Z))h(\epsilon_Y)A \mid Z, A]] \\ &= \mathrm{Cov}(Z, Y) - E[A(Z - E(Z)) \underbrace{E[h(\epsilon_Y) \mid Z, A]}_{\overset{\text{assumption}}{=} E[h(\epsilon_Y)]}] \\ &= \mathrm{Cov}(Z, Y) - E[h(\epsilon_Y)] \cdot E[A(Z - E(Z))] \\ &= \mathrm{Cov}(Z, Y) - E[h(\epsilon_Y)]\mathrm{Cov}(Z, A) \ . \end{aligned}$$

Thus,

$$E[h(\epsilon_Y)] = \frac{\mathrm{Cov}(Z, Y)}{\mathrm{Cov}(Z, A)} \ .$$

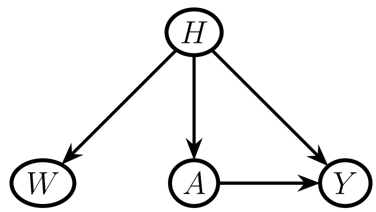
(c)

$$E[Y - Y_0 \mid A, Z] = AE[h(\epsilon_Y) \mid A, Z] - E[g(H, \epsilon_Y) \mid A, Z] + E[g(H, \epsilon_Y) \mid A, Z]$$

$$= AE[h(\epsilon_Y) \mid A, Z]$$
$$= AE[h(\epsilon_Y)] \ .$$

The result follows because $AE[h(\epsilon_Y)]$ does not depend on Z.

Exercise 2 (A sensitivity analysis)**.** Consider the treatment A, outcome Y, unmeasured variable H and measured pre-treatment variable W satisfying the graph below.



As we can see from the graph, both W and A are confounded for Y by H. Suppose that

$$E[Y^{a=0} \mid A = 1] - E[Y^{a=0} \mid A = 0] = E[W \mid A = 1] - E[W \mid A = 0] \ .$$

(a) Use this assumption to find an identification formula for $E[Y^{a=1} - Y^{a=0} \mid A = 1]$ in terms of the observed data A, W, Y.
(b) Can we interpret this as an average total effect in the entire population?

Solution:

In this problem, we derive the difference in differences estimand for negative outcome control.

(a) From the assumption,

$$E[Y^{a=0} \mid A = 1] = E[Y^{a=0} \mid A = 0] + E[W \mid A = 1] - E[W \mid A = 0]$$
$$\overset{\text{consistency}}{=} E[Y \mid A = 0] + E[W \mid A = 1] - E[W \mid A = 0]$$

We also have that

$$E[Y^{a=1} \mid A = 1] \overset{\text{consistency}}{=} E[Y \mid A = 1] \ .$$

Combining these two results gives

$$E[Y^{a=1} - Y^{a=0} \mid A = 1] = E[Y - W \mid A = 1] - E[Y - W \mid A = 0] \ .$$

(b) This is the total effect in the treated subset of the population. This may be different from the effect of treatment in the untreated, $E[Y \mid A = 0]$, and thus cannot be interpreted as an average effect in the population.

Exercise 3 (Sensitivity analysis with IVs)**.** Consider a binary instrument Z, a binary treatment A and a binary outcome Y satisfying:
(1) Exclusion restriction: $Y^{z,a} = Y^a$
(2) IV exchangeability: $Y^a \perp\!\!\!\perp Z$

Show that under assumptions (1)-(2),

$$P(Y = 0, A = 1 \mid Z = 0) + P(Y = 1, A = 1 \mid Z = 1) \leq 1 \ .$$

Hint: Use the fact that $p(x_1, x_2 \mid x_3) \leq p(x_1 \mid x_3)$. Likewise, it can also be shown that

- $P(Y = 0, A = 1 \mid Z = 0) + P(Y = 1, A = 1 \mid Z = 1) \leq 1$

- $P(Y=0, A=1 \mid Z=1) + P(Y=1, A=1 \mid Z=0) \leq 1$
- $P(Y=0, A=0 \mid Z=1) + P(Y=1, A=0 \mid Z=0) \leq 1$

These inequalities can be used to falsify IV exchangeability assumption. With some more arguments, it is also possible to use the IV inequalities to obtain bounds on causal effects.

Solution:

In this problem, we derive instrumental variable bounds.

$$\begin{aligned} P(Y^{a=1}=1) &\overset{Y^a \perp\!\!\!\perp Z}{=} P(Y^{a=1}=1 \mid Z=1) \\ &\geq P(Y^{a=1}=1, A=1 \mid Z=1) \\ &\overset{\text{consistency}}{=} P(Y=1, A=1 \mid Z=1) \ . \end{aligned}$$

Likewise,

$$\begin{aligned} P(Y^{a=1}=0) &\overset{Y^a \perp\!\!\!\perp Z}{=} P(Y^{a=1}=0 \mid Z=0) \\ &\geq P(Y^{a=1}=0, A=1 \mid Z=0) \\ &\overset{\text{consistency}}{=} P(Y=0, A=1 \mid Z=0) \end{aligned}$$

which is equivalent to

$$1 - P(Y^{a=1}=1) \geq P(Y=0, A=1 \mid Z=0) \ .$$

Using the previously derived inequality for the probability on LHS gives

$$1 - P(Y=1, A=1 \mid Z=1) \geq P(Y=0, A=1 \mid Z=0) \ .$$

References

[1] Andrea Rotnitzky. BST 257 (Harvard T.H. Chan School of Public Health).