Exercises MLSS 2019: *Causality*

Joris Mooij

August 27, 2019

1 In-class exercises

1.1 Simpson's Paradox

You are investigating the effectiveness of a drug against a deadly disease. You are given access to data collected by health insurance companies about their customers. You divide the diseased customers into two groups: those that took the drug, and those that didn't take the drug. Some of the customers recovered, others unfortunately didn't recover. The reasons why some patients were treated and others were not, are unknown to you. You find the following numbers:

	Recovery	No recovery	Total	Recovery rate
Drug	20	20	40	...%
No drug	16	24	40	...%
Total	36	44	80	

a) *Calculate the recovery rates (in %) for both groups ("drug" and "no drug").* **Answer:** Drug: 50%; No drug: 40%.

b) *If you were diseased, would you take the drug, or not?* **Answer:** The data provided does not contain enough information to decide rationally (at first sight, one might think that taking the drug is advantageous).

Upon closer inspection of the data, you notice something peculiar when you group patients according to gender:

Males	Recovery	No recovery	Total	Recovery rate
Drug	18	12	30	...%
No drug	7	3	10	...%
Total	25	15	40	

Females	Recovery	No recovery	Total	Recovery rate
Drug	2	8	10	...%
No drug	9	21	30	...%
Total	11	29	40	

c) *Calculate the recovery rates (in %) for both groups ("drug" and "no drug"), for each subpopulation (males and females) separately.* **Answer:** Males: Drug 60%; No drug 70%.
Females: Drug 20%; No drug 30%.

d) *In light of these numbers, would you take the drug if you were diseased, or not?* **Answer:** The data provided does not contain enough information to decide rationally (at first sight, one might think that taking the drug is disadvantageous for both males and females).

e) *What would be your advice to a diseased patient with unknown gender?* **Answer:** The data provided does not contain enough information to decide rationally.

This phenomenon is known as "Simpson's paradox". A lot has been written about this paradox, but it dissolves once you recognize that you should not make the mistake of interpreting correlations as causations, as we'll see later today.

1.2 Paths, colliders, d-blocked paths and d-separation

Definition 1 (Paths, Ancestors) *Let $\mathcal{G}$ be a directed mixed graph.*

- *A **path** q in $\mathcal{G}$ is a sequence of adjacent edges in $\mathcal{G}$ in which no node occurs more than once.*
- *A path consisting of directed edges $X_{i_1} \to X_{i_2} \to X_{i_3} \to \cdots \to X_{i_k}$ that all point in the same direction is called a **directed** path.*
- *If there is a directed path from X to Y (or if $X = Y$), X is called a **ancestor** of Y.*
- *The ancestors of Y are denoted $\mathrm{an}_{\mathcal{G}}(Y)$, and include Y.*

Definition 2 (Colliders, Blocked Paths, d-separation) *Let $\mathcal{G}$ be a directed mixed graph, and q a path on $\mathcal{G}$.*

- *A **collider** on q is a (non-endpoint) node X on q with precisely two arrowheads pointing towards X on the adjacent edges:*

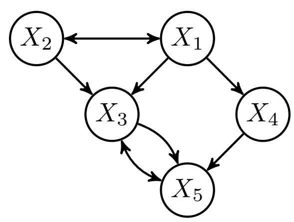
$$\to X \leftarrow, \quad \to X \leftrightarrow, \quad \leftrightarrow X \leftarrow, \quad \leftrightarrow X \leftrightarrow$$

- *A **non-collider** on q is any node on the path which is not a collider.*

*A set of nodes $\boldsymbol{S}$ in $\mathcal{G}$ is said to **d-block** q if q contains a non-collider which is in $\boldsymbol{S}$, or a collider which is* not *an ancestor of $\boldsymbol{S}$.*

*For three sets $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ of nodes in $\mathcal{G}$, we say that **$\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-separated by $\boldsymbol{Z}$** iff all paths between a node in $\boldsymbol{X}$ and a node in $\boldsymbol{Y}$ are d-blocked by $\boldsymbol{Z}$, and write $\boldsymbol{X} \perp_{\mathcal{G}} \boldsymbol{Y} \mid \boldsymbol{Z}$.*

Consider the following directed mixed graph $\mathcal{G}$:



a) *Is $X_3 \to X_5 \leftrightarrow X_3$ a path? Is it a directed path?* **Answer:** It is neither, as the node X_3 appears more than once.

b) *Is $X_3 \leftrightarrow X_5$ a path? Is it a directed path?* **Answer:** It is a path (a single edge counts as a sequence of edges) but it is not directed, as it contains a bidirected edge.

c) *Is $X_5 \leftarrow X_3 \leftarrow X_1$ a path? Is it a directed path?* **Answer:** Yes, it is a path, and it is a directed path (from X_1 to X_5).

d) *What are the ancestors of X_4?* **Answer:** $\mathrm{an}_{\mathcal{G}}(X_4) = \{X_1, X_4\}$.

Consider the path $X_2 \leftrightarrow X_1 \to X_3 \leftrightarrow X_5 \leftarrow X_4$ on $\mathcal{G}$.

e) *Which nodes on the path are colliders?* **Answer:** X_3 and X_5 are colliders on this path.

f) *Which nodes on the path are non-colliders?* **Answer:** X_2, X_1, X_4 are non-colliders on this path.

g) *Does $\{X_3\}$ d-block this path? Does $\{X_5\}$ d-block this path? Does $\{X_3, X_5\}$ d-block this path?* **Answer:** Yes, Yes, No.

h) *Does X_1 d-separate X_2 from X_4?* **Answer:** Yes.

i) *Is $X_1 \perp X_5 \mid \{X_3, X_4\}$?* **Answer:** No: the path $X_1 \to X_3 \leftrightarrow X_5$ is not d-blocked by $\{X_3, X_4\}$.

2 Tutorial exercises

2.1 Seeing vs. Doing in SCMs

Consider a simple causal model of a car. The endogenous variables are (all binary):

X: the battery is charged
Y: the start engine is operational
S: the car starts

The exogenous variables (latent, independent, binary) have Bernoulli distributions:

$$E_X \sim \mathrm{Ber}(0.95)$$
$$E_Y \sim \mathrm{Ber}(0.99)$$
$$E_Z \sim \mathrm{Ber}(0.999)$$

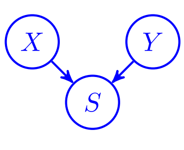
The structural equations are specified by:

$$X = E_X$$
$$Y = E_Y$$
$$S = X \wedge Y \wedge E_S$$

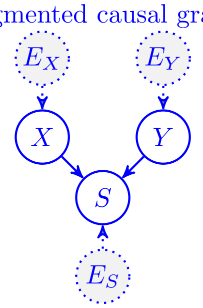
where "$\wedge$" is the logical AND.

a) *Draw the corresponding augmented causal graph and the causal graph.* **Answer:**

Causal graph:



Augmented causal graph:



b) *Write pseudocode to draw a sample (x, y, s) from this model.* **Answer:** Draw $x \sim \mathrm{Ber}(0.95)$. Draw $y \sim \mathrm{Ber}(0.99)$. Draw $e_S \sim \mathrm{Ber}(0.999)$. Calculate $s := x \wedge y \wedge e_S$.

c) *How does the model change under a perfect intervention* $\mathrm{do}(S = 0)$*? Write down the intervened model. How does the pseudocode to sample from the model change?* **Answer:** The exogenous distribution remains the same. The intervened structural equations are:

$$X = E_X$$
$$Y = E_Y$$
$$S = 0.$$

The change to the sampler is the last step, which is replaced by $s := 0$.

d) *Is $p(X = x|S = s) = p(X = x|\,\mathrm{do}(S = s))$? Motivate your answer.* **Answer:** No: observing S provides information about the value of X (e.g., observing that the car doesn't start makes it more likely that the battery is empty). Since S is not a cause of X, intervening on S does not lead to a change of the value of X.

e) *Is $p(S = s|X = x) = p(S = s|\,\mathrm{do}(X = x))$? Motivate your answer.* **Answer:** Yes: S is a direct effect of X, and the two do not have an observed or unobserved confounder, hence seeing is the same as doing in this case.

f) *Calculate:*

i) $p(S=1)$ *(the probability that the car starts)*

ii) $p(S=1|X=1)$ *(the probability that the car starts, given that the battery is charged)*

iii) $p(S=1|\operatorname{do}(X=1))$ *(the probability that the car starts when we charge the battery)*

Answer:

i) $p(S=1) = \sum_{x=0}^{1}\sum_{y=0}^{1} p(S=1, X=x, Y=y) = 0.95 \times 0.99 \times 0.999 \approx 0.934$

ii) $p(S=1|X=1) = \sum_{y=0}^{1} p(S=1|X=1, Y=y)p(Y=y) = 0.999 \times 0.99 \approx 0.989$

iii) $p(S=1|\operatorname{do}(X=1)) = \sum_{y=0}^{1} p(S=1|X=1, Y=y)p(Y=y) = 0.999 \times 0.99 \approx 0.989$

g) *Calculate:*

i) $p(X=0)$ *(the probability that the battery is empty)*

ii) $p(X=0|S=0)$ *(the probability that the battery is empty given that the car fails to start)*

iii) $p(X=0|\operatorname{do}(S=0))$ *(the probability that the battery is empty if we lost the key)*

Answer:

i) $p(X=0) = 0.05$

ii)

$$\begin{aligned}
p(X=0|S=0) &= \frac{p(S=0, X=0)}{p(S=0)} \\
&= \frac{p(S=0|X=0)p(X=0)}{\sum_{x=0}^{1}\sum_{y=0}^{1} p(X=x, Y=y, S=0)} \\
&= \frac{1 \times 0.05}{0.05 \times 0.01 + 0.95 \times 0.01 + 0.05 \times 0.99 + 0.95 \times 0.99 \times 0.001} \\
&\approx 0.827
\end{aligned}$$

iii) $p(X=0|\operatorname{do}(S=0)) = 0.05$

2.2 The Back-Door Criterion

Theorem 1 (Back-Door Criterion (Pearl, 2000)) *For an acyclic SCM $\mathcal{M}$, variables X, Y and set of variables $\boldsymbol{H}$: Let $\hat{\mathcal{G}}$ be $\mathcal{G}(\mathcal{M})$ extended with an intervention node $I_X \to X$. If*

1. $X, Y \notin \boldsymbol{H}$;

2. $H \perp_{\hat{\mathcal{G}}} I_X$;

3. $Y \perp_{\hat{\mathcal{G}}} I_X \,|\, X, H$,

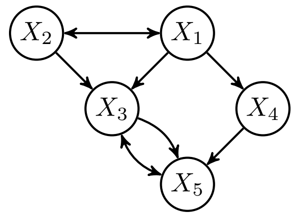
then $\boldsymbol{H}$ is called admissible for adjustment *to find the causal effect of X on Y, and this causal effect is given by:*

$$p_{\mathcal{M}}\big(y \,|\, \operatorname{do}(X=x)\big) = \int p_{\mathcal{M}}(y \,|\, x, \boldsymbol{h}) p_{\mathcal{M}}(\boldsymbol{h})\, d\boldsymbol{h} \left(= \sum_{\boldsymbol{h}} p_{\mathcal{M}}(y \,|\, x, \boldsymbol{h}) p_{\mathcal{M}}(\boldsymbol{h}) \right).$$

where the last equation only applies if $\boldsymbol{H}$ is discrete-valued. For the special case $\boldsymbol{H} = \emptyset$, this should be read as:

$$p_{\mathcal{M}}\big(y \,|\, \operatorname{do}(X=x)\big) = p_{\mathcal{M}}(y \,|\, x).$$

Consider an SCM $\mathcal{M}$ with the following causal graph $\mathcal{G}(\mathcal{M})$:



a) *Give a set that is admissible for adjustment to find the causal effect of X_4 on X_5.* **Answer:** It must contain X_1 (so $\{X_1\}$, $\{X_1, X_2\}$, $\{X_1, X_3\}$, $\{X_1, X_2, X_3\}$).

b) *Provide an expression for this causal effect in terms of the observational distribution.* **Answer:**

$$p_{\mathcal{M}}\big(x_5 \mid \mathrm{do}(X_4 = x_4)\big) = \int p_{\mathcal{M}}(x_5 \mid x_4, x_1) p_{\mathcal{M}}(x_1)\, dx_1.$$

c) *Give a set that is admissible for adjustment to find the causal effect of X_1 on X_5.* **Answer:** The only set that satisfies the sufficient condition of the back-door criterion is $\{X_2\}$.

d) *Provide an expression for this causal effect in terms of the observational distribution.* **Answer:**

$$p_{\mathcal{M}}\big(x_5 \mid \mathrm{do}(X_1 = x_1)\big) = \int p_{\mathcal{M}}(x_5 \mid x_1, x_2) p_{\mathcal{M}}(x_2)\, dx_2.$$

e) *Is $\emptyset$ admissible for adjustment to find the causal effect of X_1 on X_4? If so, provide an expression for this causal effect in terms of the observational distribution.* **Answer:** Yes. $p_{\mathcal{M}}\big(x_4 \mid \mathrm{do}(X_1 = x_1)\big) = p_{\mathcal{M}}(x_4 \mid x_1)$.

f) *Which sets are admissible for adjustment to find the causal effect of X_3 on X_5?* **Answer:** None: the path $X_3 \leftrightarrow X_5$ cannot be blocked.

g) *Which sets are admissible for adjustment to find the causal effect of X_5 on X_4?* **Answer:** None: the path $X_4 \to X_5$ cannot be blocked.

2.3 Simpson's Paradox: resolution

This exercise continues where exercise 1 ended. We will make use of causal reasoning with SCMs to resolve Simpson's paradox.

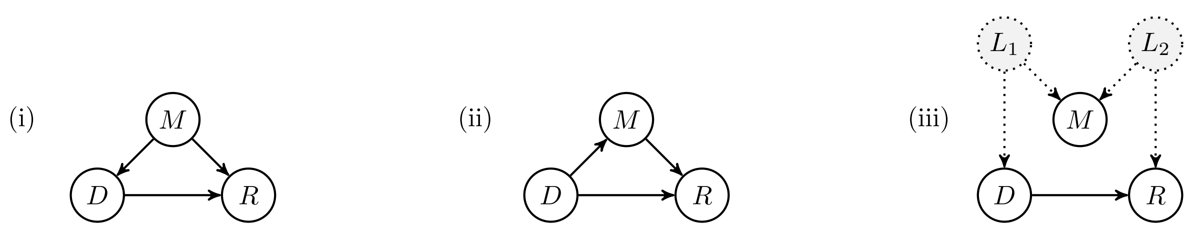


Figure 1: Different hypothetical causal graphs, where R stands for *Recovery*, D for taking the *Drug*, and M has different interpretations in cases (i), (ii) and (iii).

Suppose you believe that an SCM with causal graph as in Figure 1(i) applies, where M denotes gender of the patient (male/female).

a) *Apply the back-door criterion to obtain a formula that expresses $p(r \mid \mathrm{do}(D = d))$ in terms of observable quantities (i.e., in terms of marginal or conditional distributions where the* do*-operator does not appear).* **Answer:** We are looking for a subset of covariates that satisfies the assumptions of the back-door criterion. There are only two possible subsets of covariates in this simple case: $\emptyset$ and $\{M\}$. According to the back-door criterion, we can adjust for $\{M\}$:

$$p(R \mid \mathrm{do}(D)) = \sum_M p(R \mid D, M) p(M).$$

b) *Is $p(r \mid \mathrm{do}(D = d)) = p(r \mid d)$ in this case?* **Answer:** No, for this causal model $p(R \mid \mathrm{do}(D)) \neq p(R \mid D)$ (except possibly for specifically tuned parameters of the model). Indeed, note that $p(R \mid D) = \sum_M p(R \mid D, M) p(M \mid D)$, and for the model in Figure 1(i), in general $p(M) \neq p(M \mid D)$ (because of Bayes' theorem).

c) *What would be your advice for a patient with unknown gender?* **Answer:** First of all, one should realize that treating the patient is an *intervention*: we are actively setting the random variable D to some value. Therefore, we are interested in the quantity $p(R \mid \mathrm{do}(D))$, and not in $p(R \mid D)$ (which lead to the paradox). We can calculate $p(R \mid \mathrm{do}(D))$ using the adjustment formula in a). As $p(R|D = 1, M) < p(R|D = 0, M)$ for both genders, the same will hold for a mixture: $p(R \mid \mathrm{do}(D = 1)) < p(R \mid \mathrm{do}(D = 0))$. So assuming the causal structure in Figure 1(i), the advice would be not to take the drug.

Now suppose that instead, you believe that an SCM with causal graph as in Figure 1(ii) applies. Intuitively, this would be quite unlikely, as we know that most drugs don't change gender, but we could have used a slightly different story where the variable M has a different interpretation (for example, "blood pressure"), and then this causal structure would also be a plausible one.

d) *Again, use the back-door criterion to express $p(r \mid \mathrm{do}(D = d))$ in terms of observable quantities.* **Answer:** Now M is a descendant of D, so we cannot use $\{M\}$ for adjustment. There are no back-door paths $D \leftarrow \dots\dots R$, so $\emptyset$ is sufficient for adjustment:

$$p(R \mid \mathrm{do}(D)) = p(R \mid D)$$

e) *Is $p(r \mid \mathrm{do}(D = d)) = p(r \mid d)$ in this case?* **Answer:** Yes.

f) *What would be your advice for a patient with unknown M (say, blood pressure) in this case?* **Answer:** As $p(R \mid D = 1) > p(R \mid D = 0)$, this means that now $p(R \mid \mathrm{do}(D = 1)) > p(R \mid \mathrm{do}(D = 0))$. So assuming the causal structure is as in Figure 1(ii), the advice for (all) patients is to take the drug.

Finally, suppose that you believe that the SCM has the causal graph of Figure 1(iii).

g) *Invent an interpretation of M and the two latent variables L_1, L_2 yourself that could match the causal model depicted in Figure 1(iii).* **Answer:** Left to the imagination of the reader.

h) *Express $p(r \mid \mathrm{do}(D = d))$ in terms of* observable *quantities.* **Answer:** $\emptyset$ is sufficient for adjustment according to the back-door criterion, so

$$p(R \mid \mathrm{do}(D)) = p(R \mid D).$$

i) *Is $p(r \mid \mathrm{do}(D = d)) = p(r \mid d)$ in this case?* **Answer:** Yes.

j) *Again, what would be your advice for a patient with unknown M in this case?* **Answer:** Same as in f): all patients should take the drug.

Conclusion: whether or not you should prescribe the drug depends on which causal model you believe to apply to this situation. The fact that different causal models will lead to different conclusions should not be paradoxical, it is another illustration that "correlation does not imply causation".

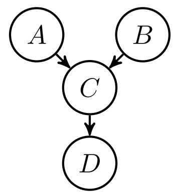
2.4 Y-structure

Theorem 2 (Global Markov Property) *For an acyclic SCM, the following* Global Markov Property *holds:*

$$\boldsymbol{X}, \boldsymbol{Y} \underset{\mathcal{G}(\mathcal{M})}{\perp} \boldsymbol{Z} \qquad \Longrightarrow \qquad \boldsymbol{X} \underset{p_{\mathcal{M}}}{\perp\!\!\!\perp} \boldsymbol{Y} \mid \boldsymbol{Z}$$

for all subsets $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ of nodes.

Given an SCM $\mathcal{M}$ with the following causal graph:



Which (conditional) independences in $p_{\mathcal{M}}$ are implied by the Global Markov Property?

2.5 LCD

Theorem 3 (Cooper, 1997) *Given an acyclic SCM $\mathcal{M}$ with three endogenous variables C, X, Y. If:*

1. *$X \to C \notin \mathcal{G}(\mathcal{M})$,*
2. *$Y \to C \notin \mathcal{G}(\mathcal{M})$,*
3. *$C \not\perp\!\!\!\perp_{p_{\mathcal{M}}} X$,*
4. *$X \not\perp\!\!\!\perp_{p_{\mathcal{M}}} Y$,*
5. *$C \perp\!\!\!\perp_{p_{\mathcal{M}}} Y \mid X$,*
6. *Faithfulness holds, i.e., the Global Markov Property gives* all *(conditional) independences in $p_{\mathcal{M}}$.*

Then $X \to Y \in \mathcal{G}(\mathcal{M})$, $C \to Y \notin \mathcal{G}(\mathcal{M})$, $X \leftrightarrow Y \notin \mathcal{G}(\mathcal{M})$, $C \leftrightarrow Y \notin \mathcal{G}(\mathcal{M})$.

Prove this theorem by considering for each possible ADMG $\mathcal{G}(\mathcal{M})$ whether it satisfies the assumptions. (Hint: could there be an edge between C and Y?) **Answer:**