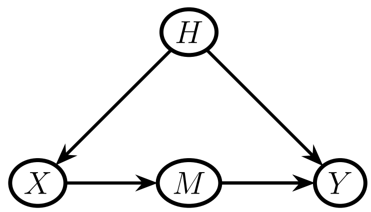
EXERCISES FOR RANDOMIZATION AND CAUSATION (MATH-336)

Exercise Sheet 10

Exercise 1 (Identification in another graph)**.** (From Technical Point 7.4 in [1]) Assume that variables X, M, Y satisfy the causal model $\mathcal{G}$ below, where we let H be an unmeasured variable. Furthermore, you can assume that all variables are discrete (and that Y is binary).



(a) Investigator 1 suggests the following identification formula (g-formula) for $E[Y^x]$:
$$E[Y^x] = E[Y \mid X = x] \ .$$
Show whether this identification formula holds or fails.

(b) Investigator 2 suggests another identification formula (not a g-formula) for a causal effect:
$$P(Y^x = 1) = \sum_m p(m \mid x) \sum_{x'} p(y \mid x', m) p(x') \ .$$
Show whether the identification formula holds or fails. You can assume that interventions on M are well-defined.

Hint: Draw several SWIGs corresponding on interventions on X, M, and both X and M. Next, remark that $Y^m = Y^x$ when $M^x = m$.

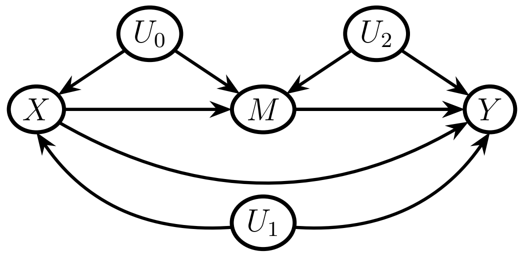
(c) State the positivity condition which is required for the identification formula in (b) to be well-defined.

(d) Prove that
$$E[Y^x] = \mathbb{E}\left[\frac{\pi(M \mid X = x)}{\pi(M \mid X)} Y\right] \ ,$$
where we defined π in the usual way as $\pi(\bullet \mid \circ) = P(M = \bullet \mid X = \circ)$. This is an IPW representation of the identification formula in part (b).

Exercise 2 (Mendelian randomization)**.** (Based on [2]) Consider a prospective Mendelian randomization study whose goal is to determine whether obesity is a cause of depression. Data are obtained on obesity ($M = 1$ indicates obese, $M = 0$ indicates non-obese), on incident depression ($Y = 1$ indicates depressed, $Y = 0$ otherwise), and on genetic variants in the FTO gene.[1]

For simplicity, we define $X = 1$ if both of a subject's genetic variants (more specifically, FTO alleles) are the minor variants (alleles); $X = 0$ otherwise (i.e. if the subject is heterozygous or homozygous for the major allele.) Consider the DAG $\mathcal{G}$:

[1] FTO is a gene which is associated with obesity.



$\mathcal{G}$

We assume that this is the causal DAG generating the data, except some of the arrows may not actually be present. Furthermore, we assume all counterfactuals are well-defined and the consistency assumption holds. Finally we assume we have a near infinite study population so sampling variability can be ignored.

(a) (i) What arrows would have to be missing in order to have

$$Y^x \perp\!\!\!\perp X \ ?$$

Justify your answer by creating an appropriate SWIG.

(ii) If these arrows are missing give the identifying formula for $E[Y^{x=1} - Y^{x=0}]$ in terms of the distribution of the observed data on (X, M, Y).

(b) (i) What arrows would have to be missing to have

$$Y^m \perp\!\!\!\perp M \mid X \ ,$$
$$Y^m \not\!\perp\!\!\!\perp M \ ?$$

Justify your answer using an appropriate SWIG.

(ii) If these arrows are missing give the identifying formula for $E[Y^{m=1} - Y^{m=0} \mid X = x]$ in terms of the distribution of the observed data on (X, M, Y). Also give the identifying formula for the unconditional effect $E[Y^{m=1} - Y^{m=0}]$ of M on Y.

(c) (i) What arrows would have to be missing in order for the following independence statements to hold:

$$Y^m \not\!\perp\!\!\!\perp M \mid X \ ,$$
$$Y^m \perp\!\!\!\perp M \ ?$$

Justify your answer using an appropriate SWIG.

(ii) If these arrows are missing, give the identification formula for the unconditional effect $E[Y^{m=1} - Y^{m=0}]$ of M on Y.

(d) (i) What arrows would have to be missing in order for the joint effect $E[Y^{x,m} - Y^{x=0,m=0}]$ to be unconfounded, i.e. for

$$Y^{x,m} \perp\!\!\!\perp M \mid X = x \ ,$$
$$Y^{x,m} \perp\!\!\!\perp X \ ?$$

Justify your answer using an appropriate SWIG.

(ii) If these arrows are missing, give the identification formula for $E[Y^{x,m}]$ in terms of the distribution of the observed data on (X, M, Y).

(e) What arrows would have to be missing for the exclusion restriction

$$Y^{x=1,m} = Y^{x=0,m} \quad \text{for} \quad m = 0, 1$$

to hold for all subjects?

(f) (i) What arrow would have to be missing to have

$$Y^{x,m} \perp\!\!\!\perp X \ ?$$

Justify your answer using an appropriate SWIG.

(ii) If these arrows are missing, is $E[Y^{x,m}]$ point identified and, if so, what is the identifying formula in terms of the distribution of the observed data on (X, M, Y)?

(g) (i) What arrows would have to be missing for both

$$Y^{x,m} \perp\!\!\!\perp X$$

and exclusion restriction

$$Y^{x=1,m} = Y^{x=0,m} \quad \text{for} \quad m = 0, 1$$

to hold?

(ii) If these arrows are missing, is $E[Y^{x,m}]$ point identified and what is the identifying formula in terms of the observed distribution on (X, M, Y)?

References

[1] Miguel Hernan and James M. Robins. *Causal Inference*. Chapman & Hall, 2018.

[2] J. M. Robins. EPI 207 (Harvard T.H. Chan School of Public Health).