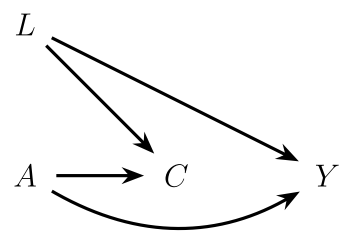
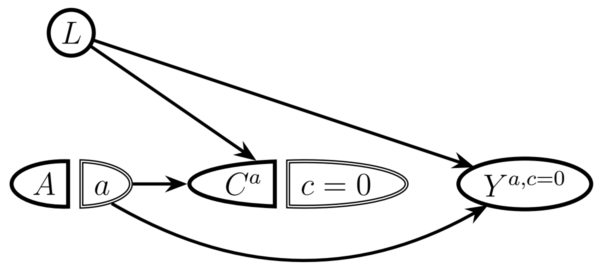
EXERCISES FOR RANDOMIZATION AND CAUSATION (MATH-336)

Exercise Sheet 6

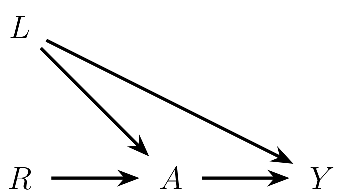
Exercise 1 (Censoring)**.** Consider the DAG and SWIG below, reproduced from the lectures. Let $A, C, Y \in \{0, 1\}$ be indicators of treatment, loss to follow-up and outcome respectively.

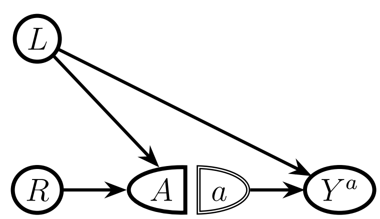




(a) Give one English sentence that explains the interpretation of $E[Y^{a,c=0}]$. Can we identify $E[Y^{a,c=1}]$ from the observed data distributions?
(b) Write down the positivity and conditional exchangeability assumption required to identify $E[Y^{a,c=0}]$.
(c) Find an identification formula for $E[Y^{a,c=0}]$.

Exercise 2 (Imperfect adherence)**.** Consider a randomized trial where patients are assigned to one of two treatments $R \in \{0, 1\}$ by randomization (flipping an unbiased coin) but do not necessarily adhere to their assigned treatment, such that their observed treatment level $A \in \{0, 1\}$ may differ from R. Let L be a baseline covariate and let Y be the outcome. Suppose that all variables are binary, and assume that the causal model in the DAG and corresponding SWIG below are valid.



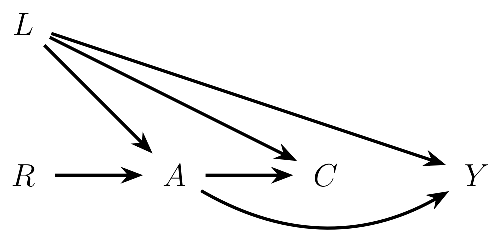


(a) (i) Write down an estimand for the per protocol effect (causal effect of A on Y). Write down the exchangeability conditions which allow us to identify the per protocol effect in a study with imperfect adherence.
(ii) Find an identification formula for this causal effect.

(b) (i) Write down an estimand for the intention-to-treat effect (the causal effect of R on Y). Write down the positivity and conditional exchangeability conditions which allow us to identify the intention-to-treat effect in a study with imperfect adherence (here, we assume no censoring)? Compare this to your answer in part (a)-(i)

(ii) Find an identification formula for this causal effect.

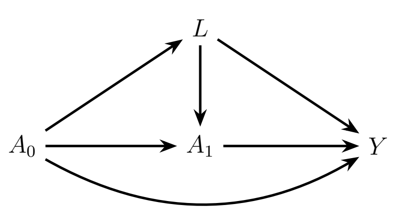
Next, we will consider a setting with imperfect adherence and losses to follow-up, depicted in the DAG below:



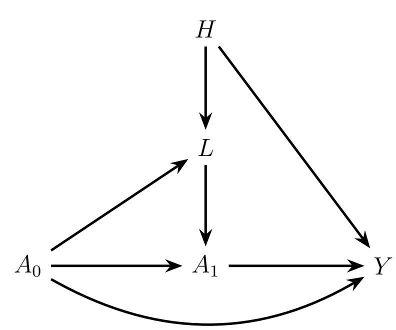
$\mathcal{G}$

(c) Write down the estimand for the causal effect of A on Y if we were to intervene to eliminate loss to follow-up, and draw the SWIG corresponding to this estimand. Write down the positivity and conditional exchangeability conditions which allow us to identify this estimand, and find an identification formula for this estimand.

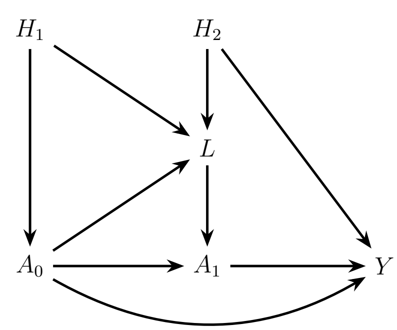
Exercise 3 (Identification with hidden variables)**.** Consider another example of a sequentially randomized experiments, where the following measured variables are temporally (and topologically) ordered from left to right $\langle A_0, L, A_1, Y\rangle$, and any variable can depend on any other variable measured in its past.[1]



$\mathcal{G}_c$



$\mathcal{G}_H$



$\mathcal{G}_{H_1,H_2}$

(a) Draw the SWIG $\mathcal{G}_c(a_0, a_1)$. By assessing the conditional exchangeability assumptions for every path between treatments $A_0, A_1^{a_0}$ and outcome Y^{a_0,a_1}, convince yourself that $\mathcal{G}_c(a_0, a_1)$ satisfies the conditional exchangeability conditions

$$Y^{a_0,a_1} \perp\!\!\!\perp A_0 \ ,$$
$$Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid A_0, L^{a_0} \ .$$

[1]The motivation in this question is to show that we can identify causal effects in the presence of unmeasured variables. This follows straightforwardly from the identification theorem, which says that causal estimands are equal to the g-formula under certain conditions (positivity, conditional exchangeability and consistency). The idea is to check this manually in two special cases without using the identification theorem. As you can see, even these simple cases require some uses of algebra and independencies.

Use these conditions to show that

$$P(Y^{a_0,a_1} = y) = \sum_l P(y \mid a_0, a_1, l)p(l \mid a_0) \ . \tag{1}$$

(b) Suppose next that there is a common cause H of L and Y. Draw the SWIG $\mathcal{G}_H(a_0, a_1)$. Convince yourself that it satisfies the conditional exchangeability conditions

$$Y^{a_0,a_1} \perp\!\!\!\perp A_0 \ ,$$
$$Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid A_0, L^{a_0}, H \ .$$

Using these conditions, show that

$$P(Y^{a_0,a_1} = y) = \sum_l \sum_h P(y \mid a_1, a_0, l, h)p(l \mid a_0, h)p(h) \ .$$

(c) Draw the SWIG $\mathcal{G}_{H_1,H_2}(a_0, a_1)$ and convince yourself that it satisfies the conditional exchangeability conditions

$$Y^{a_0,a_1} \perp\!\!\!\perp A_0 \mid H_1 \ ,$$
$$Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid A_0, L^{a_0}, H_1, H_2 \ .$$

Using these conditions, show that

$$P(Y^{a_0,a_1} = y) = \sum_l \sum_{h_1} \sum_{h_2} p(y \mid a_1, a_0, l, h_1, h_2)p(l \mid h_1, h_2, a_0)p(h_1)p(h_2) \ .$$

(d) By manipulating the conditional probabilities on the RHS of parts (b) and (c), show that both the right hand sides are equal to Eq. 1. Deduce that it is not necessary to measure H (or conversely H_1 and H_2) in order to identify $E[Y^{a_0,a_1}]$.[2]

Hint: Use the laws of probability and independencies in the graphs $\mathcal{G}_H$ and $\mathcal{G}_{H_1,H_2}$ in order to express RHS on the form

$$\text{RHS}_{(\text{b})} = \sum_l \sum_h P(y, h \mid a_0, a_1, l)p(l \mid a_0)$$

and

$$\text{RHS}_{(\text{c})} = \sum_l \sum_{h_1} \sum_{h_2} P(y, h_1, h_2 \mid a_0, a_1, l)p(l \mid a_0)$$

in order to marginalize out the hidden variables by summing over them.

(e) Do the graphs in (b) and (c) satisfy the exchangeability conditions in part (a) (reproduced below)?

$$Y^{a_0,a_1} \perp\!\!\!\perp A_0 \ ,$$
$$Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid A_0, L^{a_0} \ .$$

[2] More broadly, it is not necessary to measure all causes of all variables in a causal model in order to identify causal effect, if positivity, conditional exchangeability and consistency hold. This is an important result, which tells us that we can study isolated parts of complex systems without knowing the full causal structure. In fact, can we ever be certain that we have captured all causes in our causal model?