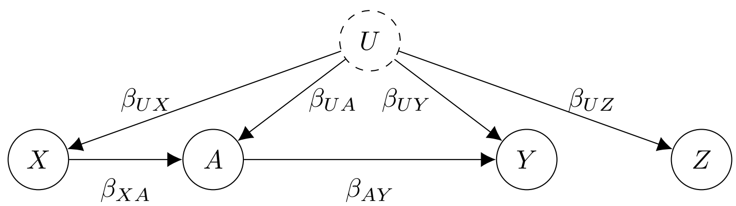
1. (Exercises 3.27, 3.28) In the linear SEM corresponding to the graph below (all variables but U are observed), suppose $\beta_{UX}\beta_{UA}\beta_{UY} \neq 0$. Show that

   (a) β_{AY} is non-identifiable if $\beta_{XA} = 0$ and $\beta_{UZ} = 0$.

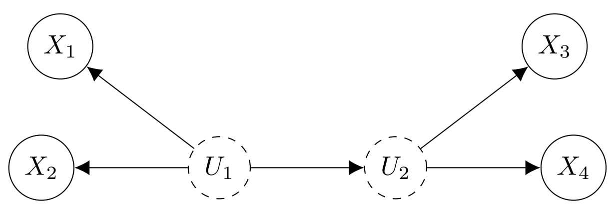
   (b) β_{AY} is identifiable if $\beta_{XA} \neq 0$ and $\beta_{UZ} \neq 0$;

   Hint: Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$ be three random variables/vectors. The partial covariance *between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ given $\boldsymbol{X}_3$ is defined as* $\mathrm{PCov}(\boldsymbol{X}_1, \boldsymbol{X}_2 \mid \boldsymbol{X}_3) = \mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) - \mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_3)\,\mathrm{Var}(\boldsymbol{X}_3)^{-1}\,\mathrm{Cov}(\boldsymbol{X}_3, \boldsymbol{X}_2)$. *Show that*

$$\beta_{AY} = \mathrm{Cov}(A, Y) - \frac{\mathrm{PCov}(X, Y \mid A)}{\mathrm{PCov}(X, Z \mid A)}\,\mathrm{Cov}(A, Z).$$



2. (Exercise 3.31) Consider the linear measurement model corresponding to graph below, where U_1 and U_2 are unobserved. Show that the model is identifiable (up to sign changes) if all the path coefficients are nonzero.



3. Academic self-concept is the perception that a student has about his/her own academic abilities and is believed to have significant influence on learning and cognitive functioning. Read the abstract and the section "Tests of Initial a Priori Model" (page 649–651) in the paper titled "Causal Ordering of Academic Self-Concept and Academic Achievement: A Multiwave, Longitudinal Panel Analysis" (click here for the paper). Then answer the following questions:

   (a) What is the research question in this study and what is the author's main conclusion?

   (b) In Figure 1, which variables are observed and which are not? What is the meaning of the double-headed arrows?

   (c) How are the three models mentioned by the author different from each other?

   (d) The three latent grades (GRADES-T1, GRADES-T2, GRADES-T3) only have one measurement. Is this measurement model identifiable without additional assumptions? Locate where the author discusses this issue.

   (e) Suppose the latent grades are accurately observed. Is the author's Model 3 actually identifiable?

4. (Exercises 4.14 and 4.22) Suppose the distribution $\mathbb{P}$ of $\boldsymbol{X}$ factorises according to the the DAG below. Use separation in the moralised (ancestral) graphs to check which of the following relations are true. Then use d-separation to check them again.

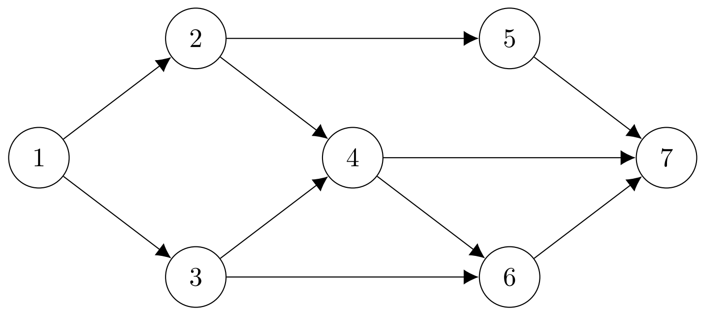
(a) $X_2 \perp\!\!\!\perp X_6 \mid X_4$;

(b) $X_2 \perp\!\!\!\perp X_6 \mid X_3$;

(c) $X_2 \perp\!\!\!\perp X_7 \mid \{X_4, X_5\}$;

(d) $X_5 \perp\!\!\!\perp X_6 \mid X_4$;

(e) $X_5 \perp\!\!\!\perp X_6 \mid \{X_3, X_4\}$;



5. (Exercise 4.30) Use the IC/SGS algorithm to derive the Markov equivalence class containing the DAG in Question 4. Give the conditional independences and dependences you used in the IC algorithm to discover the skeleton and immoralities, assuming the distribution of $\boldsymbol{X}$ is faithful to the DAG.

6. (Exercises 5.16, 5.23) Consider the causal model below representing a sequentially randomised experiment.

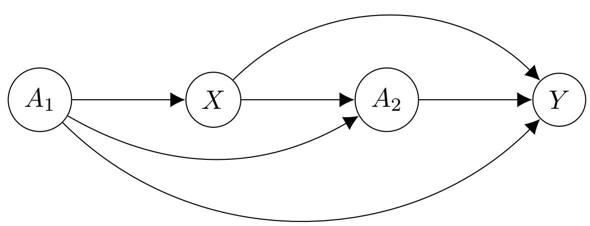
   (a) Use SWIGs to show that $Y(a_1, a_2) \perp\!\!\!\perp A_1$ and $Y(a_2) \perp\!\!\!\perp A_2 \mid A_1, X$.

   (b) By applying the g-computation formula, show that

$$\mathbb{E}[Y(a_1, a_2)] = \sum_x \mathbb{P}(X = x \mid A_1 = a_1) \cdot \mathbb{E}[Y \mid A_1 = a_1, A_2 = a_2, X = x]. \tag{1}$$

   (c) Prove (1) again by using the condition independences in part (a) and the consistency of counterfactuals.

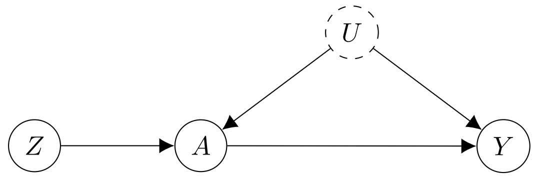
   (d) Does (1) still hold if there is an unmeasured common parent of X and Y?



7. (Exercise 5.32) Prove the counterfactual calculus (Proposition 5.31 in the notes).

8. Suppose the treatment A is binary and the no unmeasured confounders assumption $A \perp\!\!\!\perp Y(a) \mid X$, $a = 0, 1$ is satisfied. Derive an identification formula for the average treatment effect on the treated: $\text{ATT} = \mathbb{E}[Y(1) - Y(0) \mid A = 1]$.

9. In this exercise, you will explore (partial) identification of the average treatment effect using instrumental variables. Suppose Z, A, Y, U satisfy the single-world causal model corresponding to the graph below.



(a) Suppose the variables satisfy a linear structural equation model according to this graph. Show that the causal effect of A on Y is identifiable if the effect of Z on A is non-zero.

(b) For the rest of this question, suppose $Z, A, Y \in \{0, 1\}$ are binary. Show that, without using Z, the average treatment effect of A on Y satisfies the following inequalities:

$$-\mathbb{P}(Y=0, A=1) - \mathbb{P}(Y=1, A=0) \le \mathbb{E}[Y(1) - Y(0)] \le \mathbb{P}(Y=1, A=1) + \mathbb{P}(Y=0, A=0).$$

Conclude that the gap between the lower and upper bounds is 1.

(c) Let $p(y, a \mid z)$ denote $\mathbb{P}(Y = y, A = a \mid Z = z)$ and $p(y \mid z)$ denote $\mathbb{P}(Y = y \mid Z = z)$. Show that

$$\mathrm{LB} \le \mathbb{E}[Y(1) - Y(0)] \le \mathrm{UB},$$

where

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

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

Conclude that $\mathrm{UB} - \mathrm{LB} \le \min\{\mathbb{P}(A = 0 \mid Z = 0) + \mathbb{P}(A = 1 \mid Z = 1), \mathbb{P}(A = 0 \mid Z = 1) + \mathbb{P}(A = 1 \mid Z = 0)\} \le 1$ and $\mathrm{UB} - \mathrm{LB} = 1$ if and only if $A \perp\!\!\!\perp Z$.

10. In this exercise, you will apply the methods covered in the lectures to a real randomised trial to evaluate three strategies of improving the academic performance of college freshmen.

(a) Download the dataset at `http://www.statslab.cam.ac.uk/~qz280/teaching/causal-2019/examples/Angrist2009.zip`.

(b) Unzip the file, you will find a pdf report of the trial, a dta file of the original data, a csv file converted from the dta file, and a txt file describing variables recorded in the study (columns in the dta/csv file).

(c) An arm of a randomised trial is a group of units receiving a specific treatment (or no treatment). Besides a control arm, the trial had three treatment arms: SSP, SFP, and SFSP. Find out what they are and the relationship between the three treatments.

(d) Read the dataset into `R` or using your preferred programming language. How many students were assigned to each of the four arms?

(e) We will preprocess the data in the same way as the report. First, remove student who enrolled less than 2 courses in Fall 2005. You will find that there are 1579 students left in the dataset.

(f) Some students did not have fall grades because they took no one-semester courses. What is the percentage of students without fall grade in each treatment arm? Remove students without fall grades, you will end up with 1404 observations.

(g) Identify columns in the data corresponding to the "basic controls" in footnote 11, page 143 of the report (the jargon "control" refers to the covariates used in the regression adjustment).

(h) Use the "basic controls" and the three regression-based estimators in Example sheet 1, Question 5 to estimate the average treatment effect of SFSP (compared to the control). Obtain heteroskedasticity-robust standard errors by replacing the asymptotic variance formula in Example sheet 1, Question 5 by sample estimates. In `R`, you can also obtain the standard errors using the `vcovHC` function in the `sandwich` package. Is the standard error you computed the same as the output of `vcovHC`?

(i) Compute the Mann-Whitney U (Wilcoxon's rank sum) statistic that compares the SFSP treatment with the control. Compute a two-sided p-value using Monte-Carlo simulations. If you are using `R`, compare your result with the output of `wilcox.test`.