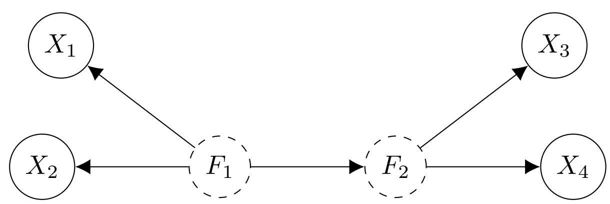
CAUSAL INFERENCE Part III / Michaelmas 2023
Example Sheet 2 (of 3)

In questions below, we either consider an acyclic directed mixed graph (ADMG) $\mathcal{G} = (\mathcal{V}, \mathcal{D}, \mathcal{B})$ with all bidirected self-loops (i.e. $(j, j) \in \mathcal{B}$ for all $j \in \mathcal{V}$) or an directed acyclic graph (DAG) $\mathcal{G} = (\mathcal{V}, \mathcal{D})$.

1. Recall that we say a walk from j to k is an *arc* if it has no collider, and we say it is a *trek* if it has no collider and exactly one bidirected edge.
   (a) How many bidirected edges can an arc have?
   (b) How many arrowheads can an arc have at the two endpoints? How many arrowheads can a trek have at the two endpoints?
   (c) Suppose $j, k \in \mathcal{V}$ are distinct and $j, k \notin \mathcal{U} \subset \mathcal{V}$. Show that there exists a trek from j to k via $\mathcal{U}$ (meaning all its non-endpoints are in $\mathcal{U}$) if and only if there exists an arc from j to k via $\mathcal{U}$ with two endpoint arrowheads, that is, $j \overset{t \text{ via } \mathcal{U}}{\leftrightsquigarrow} k$ if and only if $j \overset{\text{via } \mathcal{U}}{\leftrightsquigarrow} k$.

2. Consider the linear measurement model corresponding to graph below, where F_1 and F_2 are unobserved. Show that the model is generically identifiable up to some sign changes.



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?

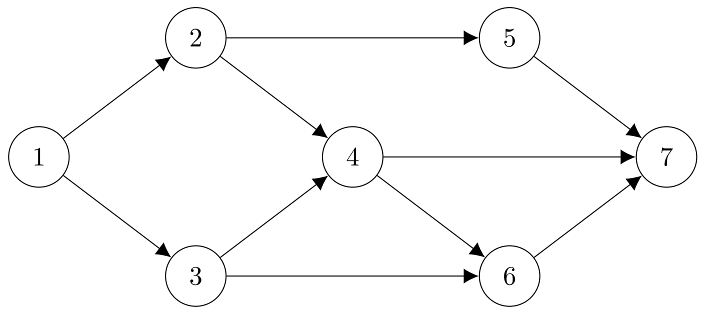
4. Prove that the following irrelevance notions satisfy the graphoid axioms as given in the lectures:
   (a) conditional independence between random variables (assuming that their joint density function is positive);
   (b) separation of vertex sets in undirected graphs;
   (c) d-separation of vertex sets in DAGs.

5. Suppose the distribution of $\boldsymbol{V}$ factorises according to the the DAG below. Use d-separation to check which of the following relations are true.
   (a) $V_2 \perp\!\!\!\perp V_6 \mid V_4$;
   (b) $V_2 \perp\!\!\!\perp V_6 \mid V_3$;

(c) $V_2 \perp\!\!\!\perp V_7 \mid \{V_4, V_5\}$;

(d) $V_5 \perp\!\!\!\perp V_6 \mid V_4$;

(e) $V_5 \perp\!\!\!\perp V_6 \mid \{V_3, V_4\}$;



6. Let j, k, l be three vertices in a DAG $\mathcal{G} = (\mathcal{V}, \mathcal{D})$. The IC/SGS algorithm is based on the following two observations about d-separation:

   (a) j and k are adjacent in $\mathcal{G}$ if and only if they cannot be d-separated by any subset of $\mathcal{V} \setminus \{j, k\}$.

   (b) Suppose j, l are adjacent, k, l are adjacent, and j, k are not adjacent in $\mathcal{G}$. Then they form an unshielded collider (i.e. $j \to l \leftarrow k$) if and only if j and k are not d-separated by any subset of $\mathcal{V} \setminus \{j, k\}$ containing l.

   Prove the first observation above.

7. Use the IC/SGS algorithm to derive the Markov equivalence class containing the DAG in Question 5. 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.

8. Let $\{i\}, J, K$ be disjoint subsets of vertices in a DAG. By using the recursive definition of potential outcomes, show that $\boldsymbol{V}_J(\boldsymbol{v}_K) = \boldsymbol{v}_J$ implies that $V_i(\boldsymbol{v}_J, \boldsymbol{v}_K) = V_i(\boldsymbol{v}_K)$ in the sense that the conditional probability of the second event given the first event is 1.

9. Consider an ADMG $\mathcal{G} = (\mathcal{V}, \mathcal{D}, \mathcal{B})$ and let $\mathcal{J}, \mathcal{K}, \mathcal{L}$ be disjoint subsets of $\mathcal{V}$ that satisfy $\mathcal{J} \rightsquigarrow \not\leftrightarrow \leftsquigarrow \mathcal{K} \mid \mathcal{L}$. Suppose $\mathcal{N} \subseteq \mathcal{V}$ is a subset of vertices with no outgoing edges. Show that

   (a) $\mathcal{J} \rightsquigarrow \not\leftrightarrow \leftsquigarrow \mathcal{K} \mid \mathcal{L} \setminus \mathcal{N}$.

   (b) $\mathcal{J} \rightsquigarrow \not\leftrightarrow \leftsquigarrow (\mathcal{L} \cap \mathcal{N} \cap \mathrm{ch}(K)) \mid \mathcal{L} \setminus \mathcal{N}$.

   (c) $\mathcal{J} \rightsquigarrow \not\leftrightarrow \leftsquigarrow \mathcal{K} \mid \mathcal{L} \cup (\mathrm{an}(\mathcal{J} \cup \mathcal{K} \cup \mathcal{L}) \setminus (\mathcal{J} \cup \mathcal{K} \cup \mathcal{L}))$.

10. Consider the causal model below representing a sequentially randomised experiment.

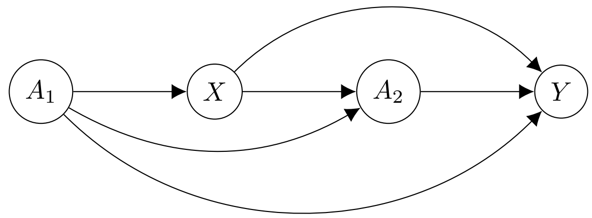
    (a) Use SWIGs to show that $Y(a_1, a_2) \perp\!\!\!\perp A_1$ and $Y(a_2) \perp\!\!\!\perp A_2 \mid A_1, X$.

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

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

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

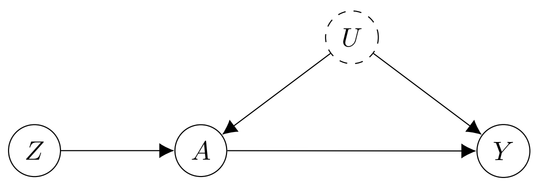
    (d) Does (1) still hold if there is an unmeasured common parent of X and Y?



11. 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: $\text{ATT} = \mathbb{E}[Y(1) - Y(0) \mid A = 1]$.

12. Let j and k be two vertices in an ADMG $\mathcal{G}$. Prove that there exists a m*-connected walk from j to k in $\mathcal{G}$ if and only if there exists a m-connected path from j to k in $\mathcal{G}$.

13. In this exercise, we 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

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

where

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

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

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