CAUSAL INFERENCE - Example Sheet 2 Solutions

J. Hera Shi   Part III Michaelmas 2023

Q 1 Recall the definition of m-connectedness:

$$W(V \leftrightsquigarrow V) = W(V \leftsquigarrow V) + W(V \rightsquigarrow V) + W(V \leftrightsquigarrow V) \tag{1}$$

And trek rule:

$$\begin{aligned} W(V \overset{t}{\leftrightsquigarrow} V) =& W(V \leftrightarrow V) + W(V \leftsquigarrow V \leftrightarrow V) \\ &+ W(V \leftrightarrow V \rightsquigarrow V) + W(V \leftsquigarrow V \leftrightarrow V \rightsquigarrow V) \end{aligned} \tag{2}$$

(a) An arc can have zero or one bidirected edge. Two or more will create colliders.

(b) An arc can have one or two arrowheads at the two endpoints. A trek has two arrowheads at the two endpoints.

(c) "$\Rightarrow$" a trek is an arc, thus the sufficiency is trivial.
"$\Leftarrow$" If an arc π is not already a trek, then we can construct the following:

$$\pi' = j \leftsquigarrow \mathcal{U} \overset{t}{\leftrightarrow} \mathcal{U} \rightsquigarrow k, \tag{3}$$

and π' is a trek (we can do so because the arc we start with has two endpoint arrowheads, which avoids creating extra colliders.

Q 2 We use the naming convention of Figure 1 and apply the result of Lemma 3.24 w.l.o.g. assuming that X_1, X_2, X_3, X_4 are standardised. By the path analysis formula, we get the covariance matrix

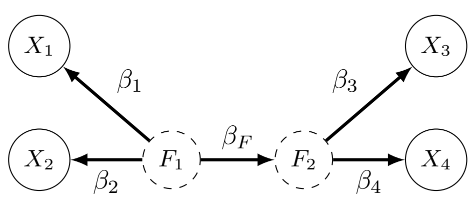


Figure 1: Graphical model of Question 2.

$$\mathrm{cov}\left((X_1, X_2, X_3, X_4)^T\right) = \begin{pmatrix} 1 & & & \\ \beta_1\beta_2 & 1 & & \\ \beta_1\beta_F\beta_3 & \beta_2\beta_F\beta_3 & 1 & \\ \beta_1\beta_F\beta_4 & \beta_2\beta_F\beta_4 & \beta_3\beta_4 & 1 \end{pmatrix}.$$

Using these relationships we can identify β_F and β_1 up to the sign as follows

$$\beta_F^2 = \frac{\mathrm{cov}(X_1, X_3)\,\mathrm{cov}(X_2, X_4)}{\mathrm{cov}(X_1, X_2)\,\mathrm{cov}(X_3, X_4)},$$
$$\beta_1^2 = \frac{\mathrm{cov}(X_1, X_2)\,\mathrm{cov}(X_1, X_3)}{\mathrm{cov}(X_2, X_3)}.$$

The path coefficients β_2, β_3 and β_4 can be identified similarly.

Q 3 (a) Research question: Does academic self-concept have an impact on academic performance or the other way around? Hence, the authors are interested in causal ordering.
Main conclusion: Based on this study, the academic self-concept influences grades, but grades don't influence the academic self-concept.

(b) The variables in ovals are latent, that is not observed; the variables represented by boxes are observed. Double-headed arrows represent the correlation between two variables.

(c)
  * Model 1: no correlation between academic self-concepts
  * Model 2: correlation between academic self-concepts
  * Model 3: correlation between academic self-concepts and between two variables of academic ability

(d) According to Proposition 3.29 the path coefficients in this case aren't identifiable because each factor has to have at least three measurements. The author agrees and sets this coefficient to 0.9. This issue is discussed in the last paragraph of the section "Tests of initial a priori models" and in section "Sensitivity analysis".

(e) If the latent grades are observed, then model 3 is identifiable. (Two measurements on academic self-concept T4 balance with four in ability T1.)

Q 4 (a) First, we prove that conditional independence fulfills the graphoid axioms. Let X, Y, Z, and W be random variables; we denote (conditional) densities by f.

  * Symmetry: Assume $Y \perp\!\!\!\perp X|Z$. Then,

$$f(x,y|z) = f(x|z) \cdot f(y|z) = f(y|z) \cdot f(x|z) = f(y,x|z).$$

    Hence, $X \perp\!\!\!\perp Y|Z$.

  * Decomposition: Assume $X \perp\!\!\!\perp (Y,W)|Z$. Then,

$$f(x,y|z) = \int f(x,y,w|z)\,dw = \int f(x|z)f(y,w|z)\,dw = f(x|z) \cdot f(y|z).$$

    Hence, $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp W|Z$ holds by the same argument.

  * Weak Union: Assume $X \perp\!\!\!\perp (Y,W)|Z$. By the decomposition property, we have $f(x|z) = f(x|z,w)$. Hence,

$$f(x,y|w,z) = \frac{f(x,y,w|z)}{f(w|z)} = \frac{f(x|z)f(y,w|z)}{f(w|z)} = f(x|z,w)f(y|w,z).$$

    Therefore, $X \perp\!\!\!\perp Y|Z,W$.

  * Contraction: Assume $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp W|Z,Y$. Then,

$$f(x,y,w|z) = f(x,w|z,y)f(y|z) = f(x|z,y)f(w|z,y)f(y|z) = f(x|z)f(y,w|z),$$

    which yields $X \perp\!\!\!\perp (Y,W)|Z$.

  * Intersection: Assume $X \perp\!\!\!\perp Y|Z,W$ and $X \perp\!\!\!\perp W|Z,Y$. This implies,

$$f(x,y,w|z) = f(x|z,w)f(y,w|z), \qquad f(x,y,w|z) = f(x|z,y)f(y,w|z), \tag{4}$$

    and, thus, $f(x|z,y) = f(x|z,w)$ for all x, y and w. Therefore, we get

$$f(x|z) = \int f(x,w|z)\,dw = \int f(x|z,w)f(w|z)\,dw = \int f(x|z,y)f(w|z)\,dw = f(x|z,y).$$

    Plugging this relationship into (4) proves $X \perp\!\!\!\perp (Y,W)|Z$.

(b) Second, we prove that the separation of vertex sets fulfills the graphoid axioms. Let I, J, K and L be disjoint sets of vertices in the undirected graph $\mathcal{G}$.

  * Symmetry: Assume $I \perp\!\!\!\perp J|K\,[\mathcal{G}]$. Since every path from a node in I to J is also a path from a node in J to I, it holds $J \perp\!\!\!\perp I|K\,[\mathcal{G}]$.
  * Decomposition: Assume $I \perp\!\!\!\perp (J,L)|K\,[\mathcal{G}]$. The set of paths between I and (J,L) is a superset of the set of paths between I and J and I and L. Hence, $I \perp\!\!\!\perp J|K\,[\mathcal{G}]$ and $I \perp\!\!\!\perp L|K\,[\mathcal{G}]$.

* Weak Union: Assume $I \perp\!\!\!\perp (J, L)|K\,[\mathcal{G}]$. All paths between I and J that don't contain a node in L are blocked by K, and also by (K, L), due to the decomposition property. All paths between I and J that contain at least one node in L are blocked by L and, therefore, also by (K, L). Hence, we have $I \perp\!\!\!\perp J|K, L\,[\mathcal{G}]$.
* Contraction: Assume $I \perp\!\!\!\perp J|K\,[\mathcal{G}]$ and $I \perp\!\!\!\perp L|K, J\,[\mathcal{G}]$. All paths between I and (J, L) that don't contain at least one vertex in J are blocked by K; otherwise, we could have an open path between that vertex and I that contradicts $I \perp\!\!\!\perp J|K\,[\mathcal{G}]$. All paths between I and (J, L) that don't contain a vertex in J are a subset of paths between I and L and, therefore, blocked by (K, J). Since these paths don't contain nodes in J, they are also blocked by K. In summary, we have $I \perp\!\!\!\perp (J, L)|K\,[\mathcal{G}]$.
* Intersection: Assume $I \perp\!\!\!\perp J|K, L\,[\mathcal{G}]$ and $I \perp\!\!\!\perp L|K, J\,[\mathcal{G}]$. Any path p starting at a vertex in (J, L) and ending at a vertex in I has a "subpath" p_s that starts at the last vertex of p that is contained in (J, L). Clearly, if p_s is blocked by K, so is p. Since p_s contains only one vertex in (J, L), it is either an element of J or L. In the former case, the subpath is blocked by (K, L) and, because it doesn't contain a vertex in L, also blocked by K. In the latter case, it is blocked by (K, J) and, by the same argument, also blocked by K. Hence, p is blocked by K and we get $I \perp\!\!\!\perp (J, L)|K\,[\mathcal{G}]$.

(c) Lastly, we prove that the d-separation of vertex sets in DAGs satisfies the graphoid axioms:

* **Symmetry**: if $J \rightsquigarrow \not\star \leftsquigarrow K|L$, then the reverse should hold by definition $K \rightsquigarrow \not\star \leftsquigarrow J|L$.
* **Weak Union**: if $J \rightsquigarrow \not\star \leftsquigarrow K \cup M|L$, then $J \rightsquigarrow \not\star \leftsquigarrow K|L \cup M$:
  [Proof by contradiction] Suppose $\pi \in W(J \rightsquigarrow \star \leftsquigarrow K|L \cup M)$.
  If π has no collider in M, then $J \rightsquigarrow \star \leftsquigarrow K|L$ should hold, which contradicts $J \rightsquigarrow \not\star \leftsquigarrow K|L$.
  If π has a collider in M, let π' be the subpath of π from J to the first vertex on π in M. Then $\pi' \in W(J \rightsquigarrow \star \leftsquigarrow M|L)$, which contradicts $J \rightsquigarrow \not\star \leftsquigarrow M|L$.
  In conclusion, this type of walk π doesn't exist, and $J \rightsquigarrow \not\star \leftsquigarrow K|L \cup M$.
* **Contraction**: Assume $J \rightsquigarrow \not\star \leftsquigarrow K|M$ and $J \rightsquigarrow \not\star \leftsquigarrow L|M \cup K$, then $J \rightsquigarrow \not\star \leftsquigarrow L \cup K|M$:
  [Proof by contradiction] By assumption, $J \rightsquigarrow \not\star \leftsquigarrow K|M$, now suppose $\pi \in W(J \rightsquigarrow \star \leftsquigarrow L|M)$.
  By this setup, we claim that π doesn't have any vertex in K. Otherwise, considering the subpath construction method mentioned in the previous proof, we can construct a $\pi' \in W(J \rightsquigarrow \star \leftsquigarrow K|M)$.
  Based on this claim, the m-connectedness would still hold if we condition on an irrelevant vertex set: $\pi \in W(J \rightsquigarrow \star \leftsquigarrow L|M \cup K)$, which contradicts the assumption.
  In conclusion, this type of walk π doesn't exist, and $J \rightsquigarrow \not\star \leftsquigarrow L \cup K|M$.
* **Intersection**: Assume $J \rightsquigarrow \not\star \leftsquigarrow K|L \cup M$ and $J \rightsquigarrow \not\star \leftsquigarrow L|M \cup K$, then $J \rightsquigarrow \not\star \leftsquigarrow L \cup K|M$:
  [Proof by contradiction] Suppose $\pi \in W(J \rightsquigarrow \star \leftsquigarrow L \cup K|M)$.
  Let π' be the subpath of π from J to the first vertex on $L \cup K$. Without loss of generality, suppose the endpoint of π' is in L. Then $\pi \in W(J \rightsquigarrow \star \leftsquigarrow L|M \cup K)$, which contradicts the assumption.
  In conclusion, this type of walk π doesn't exist, and $J \rightsquigarrow \not\star \leftsquigarrow L \cup K|M$.

Q 5 We only present the detailed solution for (a) and give the conclusions for (b) - (e).
First, we use moralisation to investigate $X_2 \perp\!\!\!\perp X_6|X_4$. The ancestors of X_2, X_4 and X_6 are X_4, X_1 and X_3 and the moralized ancestral graph is depicted in Figure 2. We see that the path $2-3-6$ is unblocked. Turning to the d-separation perspective, we notice that the path $2 \leftarrow 1 \rightarrow 3 \rightarrow 6$ isn't blocked by 4 and thus X_2 and X_6 aren't d-separated by X_4. Both criteria agree that $X_2 \perp\!\!\!\perp X_6|X_4$ doesn't hold.
Likewise for (b), (c) and (d) we can't conclude that the respective conditional independence statements hold. However, the relation $X_5 \perp\!\!\!\perp X_6|\{X_3, X_4\}$ in (e) is true.

Q 6 For the IC/SGS algorithm, we start with a fully connected undirected graph and remove the edge between j and k if they are d-separated. Now we prove the first observation:

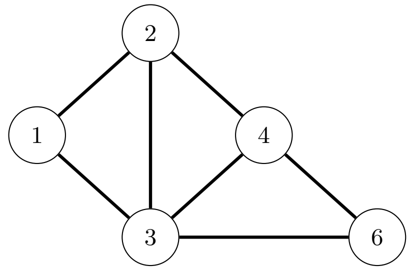


Figure 2: Moralised graph of Question 4 (a).

"⇒" If j and k are adjacent, there exists an edge between them, i.e., j and k are directly connected, without loss of generality, we assume $j \to k$, which is a d-connecting path between j and k that can't be blocked by any subset of $\mathcal{V}\backslash\{j,k\}$. Thus j and k cannot be d-separated by any subset of $\mathcal{V}\backslash\{j,k\}$.

"⇐" We can write $\mathcal{V}$ in a topological order, and WLOG, we assume $j \notin \mathrm{des}(k)$. For $j \rightsquigarrow \star \leftrightsquigarrow k$, if j and k are not adjacent, we claim that we can find a subset of $\mathcal{V}\backslash\{j,k\}$ to block j to k, namely the parents of k:

$$j \rightsquigarrow \not\star \leftrightsquigarrow k|\ \mathrm{pa}(k)$$

Consider the case when there is no collider between j and k, then we could write the path between j and k as $j \rightsquigarrow \star \leftrightsquigarrow k$, and condition on the parents of k can effectively block the pathway:

$$j \rightsquigarrow \not\star \leftrightsquigarrow k|\ \mathrm{pa}(k)$$

which contradicts the assumption that no subset of $\mathcal{V}\backslash\{j,k\}$ can block j to k.

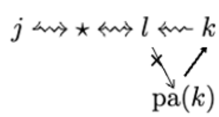
When $j \rightsquigarrow \star \leftsquigarrow k$, there is (at least) one collider l on the pathway. In this case, $\{l\} \in \mathcal{V}\backslash\{j,k\}$ naturally blocks j to k.

$$j \rightsquigarrow \star \leftrightsquigarrow l \leftsquigarrow k,$$

which is equivalent to:

$$j \rightsquigarrow \not\star \leftsquigarrow k$$

Besides, we want to prove that in this case, d-separation still holds after we condition on the parents of k, which is equivalently saying that l is not an ancestor of k: Therefore, we can always condition



on $\mathrm{pa}(k)$ to block the walk from j to k, which contradicts that j and k cannot be d-separated by any subset of $\mathcal{V}\backslash\{j,k\}$. Thus, j and k are adjacent.

Q 7 **(0)** We start with a fully connected, undirected graph.

(1) We remove all edges between i and j if $X_i \perp\!\!\!\perp X_j | X_K$ for some $K \subseteq V \setminus \{i,j\}$. For instance, we remove the edges $1-4$, $1-5$, $1-6$ and $1-7$ because of $X_1 \perp\!\!\!\perp \{X_4, X_5, X_6, X_7\} | \{X_2, X_3\}$. Proceeding with this method, we arrive at the skeleton depicted in Figure 3.

(2) We orient all paths $i-k-j$ such that i and j aren't adjacent as $i \to k \leftarrow j$ if $X_i \not\perp\!\!\!\perp X_j | X_K$ for all $K \subseteq V \setminus \{i,j\}$ containing k. We find three such cases, namely $2 \to 4 \leftarrow 3$, $4 \to 7 \leftarrow 5$ and $5 \to 7 \leftarrow 6$, and orient the graph accordingly, see Figure 4.

(3) Last, we orient edges to avoid cycles or the introduction of new immoralities. We find that $4-6$ must be oriented $4 \to 6$ to avoid the immorality $2 \to 4 \leftarrow 6$. Subsequently, we can also orient $3-6$ as $3 \to 6$ to avoid a cycle $3 \to 4 \to 6 \to 3$. Thus, get the graph depicted in Figure 5.

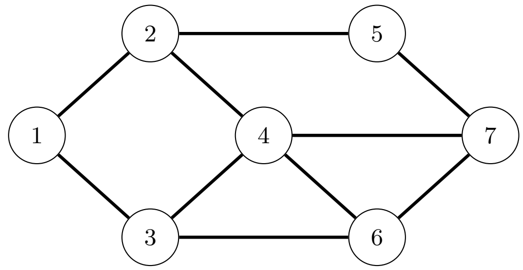


Figure 3: Skeleton (Step 1).

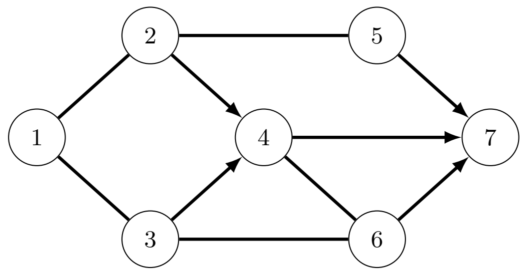


Figure 4: Orientied immoralities (Step 2).

Q 8 Let $(t_1, \dots, t_p)$ be a topological ordering of the DAG and define $(\tilde{t}_1, \dots, \tilde{t}_{\tilde{p}}) = (t_i)_{i \in V \setminus (J \cup K)}$, where $\tilde{p} = \#(V \setminus (J \cup K))$.
Base case For $\tilde{t}_1$, we have $pa(\tilde{t}_1) \setminus (J \cup K) = \emptyset$. Hence, by definition we get

$$X_{\tilde{t}_1}(\mathbf{x}_J, \mathbf{x}_K) = X_{\tilde{t}_1}(\mathbf{x}_{pa(\tilde{t}_1)\cap J}, \mathbf{x}_{pa(\tilde{t}_1)\cap K}) = X_{\tilde{t}_1}(\mathbf{X}_{pa(\tilde{t}_1)\cap J}(\mathbf{x}_K), \mathbf{x}_{pa(\tilde{t}_1)\cap K}) = X_{\tilde{t}_1}(\mathbf{x}_K)$$

because of $\mathbf{X}_J(\mathbf{x}_K) = \mathbf{x}_J$.
Induction assumption For all $l \in [\tilde{p}], l \le L < \tilde{p}$, it holds $X_{\tilde{t}_l}(\mathbf{x}_J, \mathbf{x}_K) = X_{\tilde{t}_l}(\mathbf{x}_K)$.
Induction step Since we have $pa(\tilde{t}_{L+1}) \setminus (J \cup K) \subseteq \{\tilde{t}_s\}_{1 \le s \le L}$, we can use the induction assumption and the definition of counterfactuals to complete the proof as follows

$$\begin{aligned} X_{\tilde{t}_{L+1}}(\mathbf{x}_J, \mathbf{x}_K) &= X_{\tilde{t}_{L+1}}\big(\mathbf{x}_{pa(\tilde{t}_{L+1})\cap J}, \mathbf{x}_{pa(\tilde{t}_{L+1})\cap K}, \mathbf{X}_{pa(\tilde{t}_{L+1})\setminus(J\cup K)}(\mathbf{x}_J, \mathbf{x}_K)\big) \\ &= X_{\tilde{t}_{L+1}}\big(\mathbf{X}_{pa(\tilde{t}_{L+1})\cap J}(\mathbf{x}_K), \mathbf{x}_{pa(\tilde{t}_{L+1})\cap K}, \mathbf{X}_{pa(\tilde{t}_{L+1})\setminus(J\cup K)}(\mathbf{x}_K)\big) = X_{\tilde{t}_{L+1}}(\mathbf{x}_K). \end{aligned}$$

Q 9 (a) Given the m-separation between sets J and K:

$$J \rightsquigarrow\!\!\not\;\leftsquigarrow K|L$$

If $L \cap N = \emptyset$, then the conclusion trivially holds. Now we assume $L \cap N \neq \emptyset$. Here we first prove a claim that the walk between J and K doesn't contain any vertices in $N \cap L$.
Since N has no outgoing edges, all the vertices in N can only be colliders. By the m-separation between sets J and K given L, we can make the following observations: (1) if there is no collider

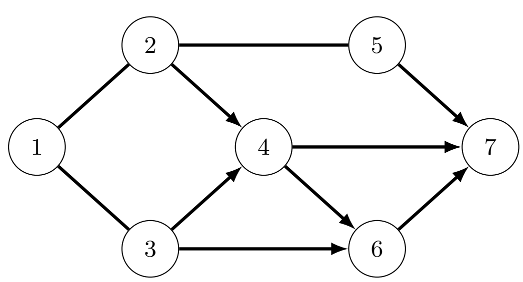


Figure 5: Markov equivalence class (Step 3).

on the walk from J to K, then the walk between J and K doesn't contain any vertices in N, thus no vertices in $N \cap L$; (2) if there is (at least) one collider on the walk from J to K, then it can't be in L, thus no vertices in $N \cap L$.

To summarize, the walk between J and K doesn't contain any vertices in $N \cap L$. Therefore, by removing $N \cap L$ from L in the condition set, we won't change the m-separation between sets J and K, i.e.:

$$J \rightsquigarrow \not\star \leftsquigarrow K | L \backslash N$$

(b) Suppose there exists a walk $\pi \in W(J \rightsquigarrow \star \leftsquigarrow m | L \backslash N)$, where $m \in L \cap N \cap \mathrm{ch}(K)$ and the non-endpoints in π have no overlap with $L \cap N \cap \mathrm{ch}(K)$ (intuitively, the shortest walk from J to any element in $L \cap N \cap \mathrm{ch}(K)$).

In this case, because $m \in N \cap \mathrm{ch}(K)$, we can construct another walk by adding K:

$$\pi' = J \rightsquigarrow \star \leftrightsquigarrow m \leftarrow K$$

In addition, because $m \in L$, then π' is m-connected given L, which contradicts the original assumption. Therefore, $J \rightsquigarrow \not\star \leftsquigarrow L \cap N \cap \mathrm{ch}(K) | L \backslash N$

(c) See the written proof.

Q 10 (a) Performing interventions on A_1 and A_2, we get the SWIG is depicted in Figure 6. We can read $Y(a_1, a_2) \perp\!\!\!\perp A_1$ directly off the graph as there is no path between A_1 and $Y(a_1, a_2)$. If

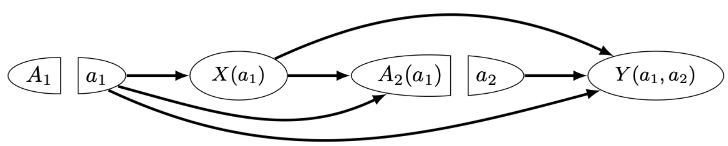


Figure 6: SWIG for interventions on A_1 and A_2.

we only intervene on A_2, we obtain the SWIG in Figure 7. We see that A_1 and X block all

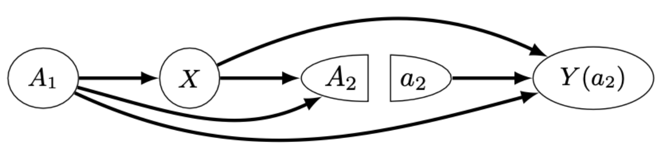


Figure 7: SWIG for intervention on A_2.

paths between A_2 and $Y(a_2)$ which implies that A_2 and $Y(a_2)$ are d-separated and $A_2 \perp\!\!\!\perp Y(a_2) | A_1, X$.

(b) Using Corollary 5.21 with $\mathbf{X}_I = Y$, $\mathbf{X}_J = (A_1, A_2)$ and $\mathbf{X}_K = X$, we get

$$\mathbb{P}(Y(a_1, a_2) = y) = \sum_x \mathbb{P}(Y = y | A_1 = a_1, A_2 = a_2, X = x)\, \mathbb{P}(X = x | A_1 = a_1)$$

and thus

$$\begin{aligned}\mathbb{E}[Y(a_1, a_2)] &= \sum_y y \sum_x \mathbb{P}(Y = y | A_1 = a_1, A_2 = a_2, X = x)\, \mathbb{P}(X = x | A_1 = a_1) \\ &= \sum_x \mathbb{P}(X = x | A_1 = a_1)\, \mathbb{E}[Y | A_1 = a_1, A_2 = a_2, X = x].\end{aligned}$$

(c)

$$\begin{aligned}
\mathbb{E}[Y(a_1,a_2)] &\overset{\text{cond. ind.}}{=} \mathbb{E}[Y(a_1,a_2)|A_1=a_1] \\
&\overset{\text{consist.}}{=} \mathbb{E}[Y(a_2)|A_1=a_1] \\
&\overset{\text{tower rule}}{=} \sum_x \mathbb{P}(X=x|A_1=a_1)\,\mathbb{E}[Y(a_2)|A_1=a_1,X=x] \\
&\overset{\text{cond. ind.}}{=} \sum_x \mathbb{P}(X=x|A_1=a_1)\,\mathbb{E}[Y(a_2)|A_1=a_1,X=x,A_2=a_2] \\
&\overset{\text{consist.}}{=} \sum_x \mathbb{P}(X=x|A_1=a_1)\,\mathbb{E}[Y|A_1=a_1,X=x,A_2=a_2]
\end{aligned}$$

(d) If there is an unmeasured parent of X and Y the (conditional) independences $A_1 \perp\!\!\!\perp Y(a_1,a_2)$ and $A_2 \perp\!\!\!\perp Y(a_2)|A_1,X$ still hold. Therefore, the derivation in (c) is still valid.

Q 11 To identify ATT, we apply usual tools such as consistency and the tower law as well as the relationship $A \perp\!\!\!\perp Y(0)\,|\,X$.

$$\begin{aligned}
\text{ATT} = \mathbb{E}[Y(1)-Y(0)|A=1] &= \mathbb{E}[Y|A=1] - \sum_x \mathbb{E}[Y(0)|A=1,X=x]\mathbb{P}(X=x|A=1) \\
&= \mathbb{E}[Y|A=1] - \sum_x \mathbb{E}[Y(0)|A=0,X=x]\mathbb{P}(X=x|A=1) \\
&= \mathbb{E}[Y|A=1] - \sum_x \mathbb{E}[Y|A=0,X=x]\mathbb{P}(X=x|A=1).
\end{aligned}$$

Q 12 The question is equivalent to:

$$W(j \rightsquigarrow \star \leftsquigarrow k) \neq \emptyset \Leftrightarrow P(j \rightsquigarrow \star \leftsquigarrow k) \neq \emptyset \tag{5}$$

"$\Leftarrow$" A path is a walk, so this direction of the proof is trivial.

"$\Rightarrow$" suppose $\pi \in W(j \rightsquigarrow \star \leftsquigarrow k)$, and let r be the repeated vertex closest to the end. Note here if r is an empty set, then "$\Rightarrow$" holds. Now we restrict $r \neq \emptyset$, and rewrite π as:

$$\pi = j \rightsquigarrow r \rightsquigarrow \star \leftsquigarrow r \leftsquigarrow k$$

where we can treat $r \rightsquigarrow \star \leftsquigarrow r$ as a backtrack walk, and replace it by simply r. Thus, we can construct another walk π' such that:

$$\pi' = j \rightsquigarrow r \leftsquigarrow k$$

Notice that π' has no repeated vertexes, thus we obtain a **path** from j to k, i.e. $\pi' \in P(j \rightsquigarrow \star \leftsquigarrow k)$.

Q 13 (a) W.l.o.g. we can assume that Z, A and Y are standardised. Hence, the covariance matrix is given by

$$\text{cov}\left((Z,A,Y)^T\right) = \begin{pmatrix} 1 & & \\ \beta_{ZA} & 1 & \\ \beta_{ZA}\beta_{AY} & \beta_{AY}+\beta_{UA}\beta_{UY} & 1 \end{pmatrix}.$$

We can identify β_{AY} using $\beta_{AY} = \frac{\beta_{ZA}\beta_{AY}}{\beta_{ZA}} = \frac{\text{cov}(Z,Y)}{\text{cov}(Z,A)}$.

(b) First, we decompose the average treatment effect, use consistency and the fact that $\mathbb{E}[Y|A=a] = \mathbb{P}(Y=1|A=a)$ for $a \in \{0,1\}$:

$$\begin{aligned}
\text{ATE} &= \mathbb{E}[Y(1)-Y(0)] \\
&= \mathbb{E}[Y(1)|A=1]\mathbb{P}(A=1) + \mathbb{E}[Y(1)|A=0]\mathbb{P}(A=0) - \mathbb{E}[Y(0)|A=1]\mathbb{P}(A=1) - \mathbb{E}[Y(0)|A=0]\mathbb{P}(A=0) \\
&= \mathbb{E}[Y|A=1]\mathbb{P}(A=1) + \mathbb{E}[Y(1)|A=0]\mathbb{P}(A=0) - \mathbb{E}[Y(0)|A=1]\mathbb{P}(A=1) - \mathbb{E}[Y|A=0]\mathbb{P}(A=0) \\
&= \mathbb{P}(Y=1,A=1) + \mathbb{E}[Y(1)|A=0]\mathbb{P}(A=0) - \mathbb{E}[Y(0)|A=1]\mathbb{P}(A=1) - \mathbb{P}(Y=1,A=0)
\end{aligned}$$

To get lower and upper bounds, we use $0 \leq \mathbb{E}[Y(1)|A=0] \leq 1$ and $0 \leq \mathbb{E}[Y(0)|A=1] \leq 1$:

$$\text{ATE} \geq \mathbb{P}(Y=1,A=1) - \mathbb{P}(A=1) - \mathbb{P}(Y=1,A=0) = -\mathbb{P}(Y=0,A=1) - \mathbb{P}(Y=1,A=0),$$
$$\text{ATE} \leq \mathbb{P}(Y=1,A=1) + \mathbb{P}(A=0) - \mathbb{P}(Y=1,A=0) = \mathbb{P}(Y=1,A=1) + \mathbb{P}(Y=0,A=0).$$

From the first estimation we see that the width between the upper and lower bound is $\mathbb{P}(A=0) + \mathbb{P}(A=1) = 1$.

(c) By examining the SWIG resulting from an intervention on A, we see that $Z \perp\!\!\!\perp Y(a)$ for $a \in \{0,1\}$. Applying the first rule of counterfactual calculus and consistency, we get

$$\mathbb{P}(Y(1)=1) = \mathbb{P}(Y(1)=1|Z=z) \geq \mathbb{P}(Y(1)=1, A=1|Z=z) = p(1,1|z)$$

for $z \in \{0,1\}$. Likewise, we can get a bound in the opposite direction

$$\mathbb{P}(Y(1)=1) = 1 - \mathbb{P}(Y=0,A=1|Z=z) - \mathbb{P}(Y(1)=0,A=0|Z=z) \leq 1 - p(0,1|z).$$

Using the same reasoning to $\mathbb{P}(Y(0)=1)$, we obtain

$$p(1,0|z) \leq \mathbb{P}(Y(0)=1) \leq 1 - p(0,0|z)$$

and thus

$$p(1,1|z) \leq \mathbb{E}[Y(1)] \leq 1 - p(0,1|z),$$
$$p(1,0|z) \leq \mathbb{E}[Y(0)] \leq 1 - p(0,0|z).$$

Consequently, we can bound the average treatment effect by

$$\max_{z,z' \in \{0,1\}} p(1,1|z) + p(0,0|z') - 1 \leq \mathbb{E}[Y(1) - Y(0)] \leq \min_{z,z' \in \{0,1\}} 1 - p(0,1|z) - p(1,0|z').$$

Spelling out the expressions for the lower and upper bound for all four combinations of z and z', we arrive at the bounds stated in the exercise. From the estimations above, we see

$$\begin{aligned}
\text{UB} - \text{LB} &= \min_{z_1,z_2,z_3,z_4 \in \{0,1\}} 2 - \big[p(1,1|z_1) + p(0,0|z_2) + p(0,1|z_3) + p(1,0|z_4)\big] \\
&\leq \min_{z \in \{0,1\}} 2 - \big[p(1,1|z) + p(0,0|\,1-z) + p(0,1|z) + p(1,0|\,1-z)\big] \\
&= \min_{z \in \{0,1\}} 2 - \big[\mathbb{P}(A=1|Z=z) + \mathbb{P}(A=0|Z=1-z)\big] \\
&= \min_{z \in \{0,1\}} \mathbb{P}(A=0|Z=z) + \mathbb{P}(A=1|Z=1-z) \\
&= \min\Big\{2 - \big(\mathbb{P}(A=0|Z=0) + \mathbb{P}(A=1|Z=1)\big), \mathbb{P}(A=0|Z=0) + \mathbb{P}(A=1|Z=1)\Big\} \\
&\leq 1.
\end{aligned}$$

It follows that $\text{UB} - \text{LB} = 1$ implies $\mathbb{P}(A=0|Z=0) + \mathbb{P}(A=1|Z=1) = 1$. Thus, we have $\mathbb{P}(A=0|Z=0) = \mathbb{P}(A=0|Z=1)$ which leads to $A \perp\!\!\!\perp Z$. The opposite direction follows from the first of the upper system of equations. If $A \perp\!\!\!\perp Z$, we obtain $\text{UB} - \text{LB} = 2 - 1 = 1$.

d-separation: How to determine which variables are independent in a Bayes net

(This handout is available at http://web.mit.edu/jmn/www/6.034/d-separation.pdf)

The Bayes net assumption says:

"Each variable is conditionally independent of its non-descendants, given its parents."

It's certainly possible to reason about independence using this statement, but we can use **d-separation** as a more formal procedure for determining independence. We start with an independence question in one of these forms:

- "Are X and Y conditionally independent, given {givens}?"
- "Are X and Y marginally independent?"

For instance, if we're asked to figure out "Is P(A|BDF) = P(A|DF)?", we can convert it into an independence question like this: "Are A and B independent, given D and F?"

Then we follow this procedure:

1. Draw the ancestral graph.
Construct the "ancestral graph" of all variables mentioned in the probability expression. This is a reduced version of the original net, consisting only of the variables mentioned and all of their ancestors (parents, parents' parents, etc.)

2. "Moralize" the ancestral graph by "marrying" the parents.
For each pair of variables with a common child, draw an undirected edge (line) between them. (If a variable has more than two parents, draw lines between every pair of parents.)

3. "Disorient" the graph by replacing the directed edges (arrows) with undirected edges (lines).

4. Delete the givens and their edges.
If the independence question had any given variables, erase those variables from the graph and erase all of their connections, too. Note that "given variables" as used here refers to the question "Are A and B conditionally independent, given D and F?", not the equation "P(A|BDF) =? P(A|DF)", and thus does not include B.

5. Read the answer off the graph.

- If the variables are **disconnected** in this graph, they are guaranteed to be independent.
- If the variables are **connected** in this graph, they are not guaranteed to be independent.* Note that "are connected" means "have a path between them," so if we have a path X-Y-Z, X and Z are considered to be connected, even if there's no edge between them.
- If one or both of the variables are missing (because they were givens, and were therefore deleted), they are independent.

* We can say "the variables are dependent, as far as the Bayes net is concerned" or "the Bayes net does not require the variables to be independent," but we cannot guarantee dependency using d-separation alone, because the variables can still be numerically independent (e.g. if P(A|B) and P(A) happen to be equal for all values of A and B).

Practicing with the d-separation algorithm will eventually let you determine independence relations more intuitively. For example, you can tell at a glance that two variables with no common ancestors are marginally independent, but that they become dependent when given their common child node.

Here are some examples of questions we can answer about the Bayes net below, using d-separation:

1. Are A and B conditionally independent, given D and F?

(Same as "P(A|BDF) =? P(A|DF)" or "P(B|ADF) =? P(B|DF)")

2. Are A and B marginally independent? (Same as "P(A|B) =? P(A)" or "P(B|A) =? P(B)")

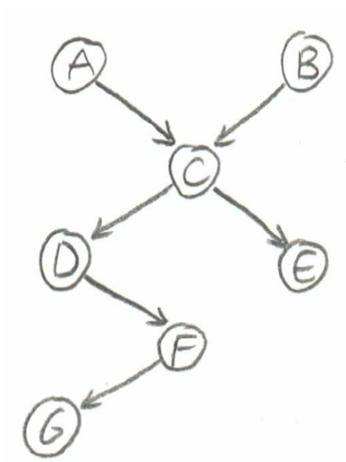
3. Are A and B conditionally independent, given C?

4. Are D and E conditionally independent, given C?

5. Are D and E marginally independent?

6. Are D and E conditionally independent, given A and B?

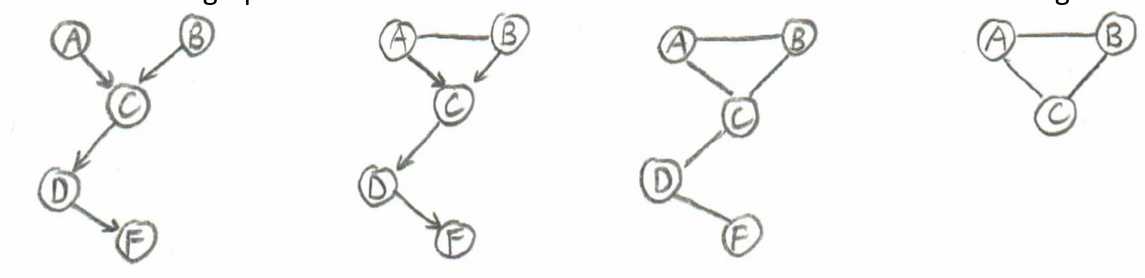
7. P(D|BCE) =? P(D|C)



Solutions are on the following pages.

1. Are A and B conditionally independent, given D and F?
(Same as "P(A|BDF) =? P(A|DF)" or "P(B|ADF) =? P(B|DF)")

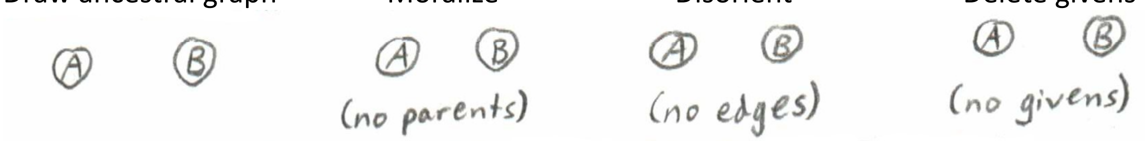
Draw ancestral graph | Moralize | Disorient | Delete givens



Answer: No, A and B are connected, so they are not required to be conditionally independent given D and F.

2. Are A and B marginally independent? (Same as "P(A|B) =? P(A)" or "P(B|A) =? P(B)")

Draw ancestral graph | Moralize | Disorient | Delete givens



Answer: Yes, A and B are not connected, so they are marginally independent.

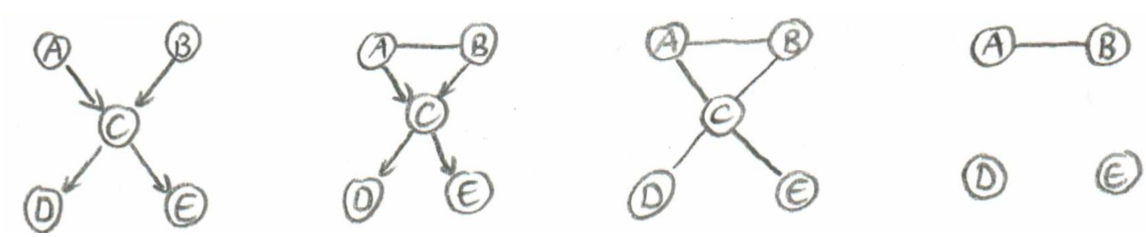
3. Are A and B conditionally independent, given C?

Draw ancestral graph | Moralize | Disorient | Delete givens

Answer: No, A and B are connected, so they are not required to be conditionally independent given C.

4. Are D and E conditionally independent, given C?

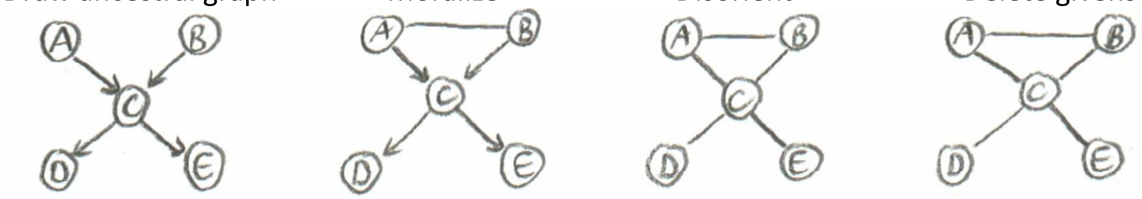
Draw ancestral graph | Moralize | Disorient | Delete givens



Answer: Yes, D and E are not connected, so they are conditionally independent given C.

5. Are D and E marginally independent?

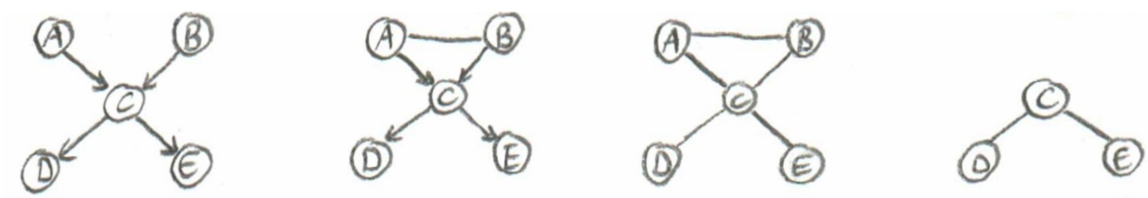
Draw ancestral graph | Moralize | Disorient | Delete givens



Answer: No, D and E are connected (via a path through C), so they are not required to be marginally independent.

6. Are D and E conditionally independent, given A and B?

Draw ancestral graph | Moralize | Disorient | Delete givens



Answer: No, D and E are connected (via a path through C), so they are not required to be conditionally independent given A and B.

7. P(D|CEG) =? P(D|C)

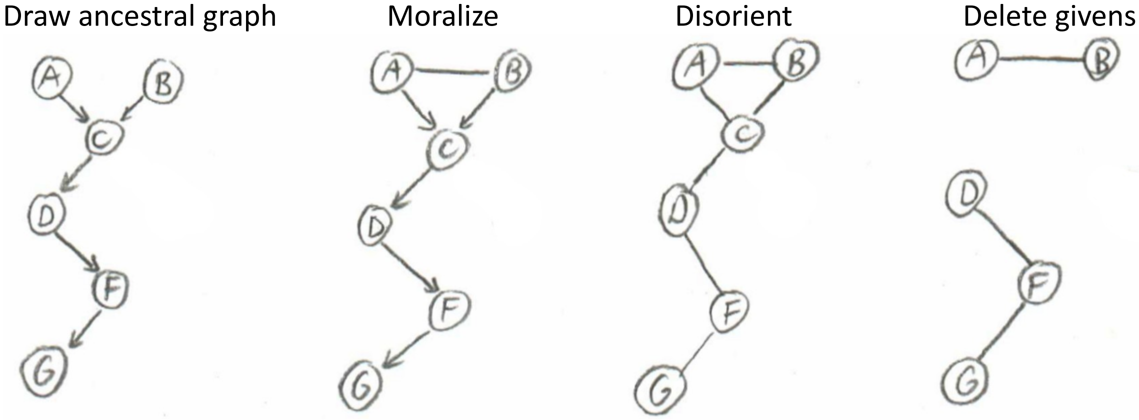
Rewrite as independence questions “Are X and Y conditionally independent, given {givens}?”:

- Are D and E conditionally independent, given C? AND
- Are D and G conditionally independent, given C?

(a) Are D and E conditionally independent, given C? Yes; see example 4.

(b) Are D and G conditionally independent, given C? No, because they are connected (via F):

Draw ancestral graph | Moralize | Disorient | Delete givens



Overall answer: No. D and E are conditionally independent given C, but D and G are not required to be. Therefore we cannot assume that P(D|CEG) = P(D|C).

Thm $J \rightsquigarrow \not\ast \leftsquigarrow K \mid L \iff J \rightsquigarrow \not\ast \leftsquigarrow K \mid M$

where $M = L \cup \mathrm{an}(J \cup K \cup L) \setminus (J \cup K)$.

Pf Suppose this is not true and let π be the shortest walk in $W(J \rightsquigarrow \ast \leftsquigarrow K \mid M)$.

So $\text{non-endpoint}(\pi) \cap (J \cup K) = \emptyset$.

Noee: $M \cup (J \cup K) = \overline{\mathrm{an}}(J \cup K \cup L)$.

① π has no collider

Then π is like $J \rightsquigarrow K$, $J \leftsquigarrow K$, $J \leftrightsquigarrow K$.

π cannot contain a non-end point, otherwise it must contain an edge like $J \leftarrow \mathrm{pa}(J)$ or $\mathrm{pa}(K) \rightarrow K$ and π cannot be m-connected given M.

So we are down to $J \rightarrow K$, $J \leftarrow K$, $J \leftrightarrow K$.

But none of this is possible because $J \rightsquigarrow \not\ast \leftsquigarrow K \mid L$.

② π has ≥ 1 colliders.

Then π looks like $J \rightsquigarrow M \leftsquigarrow K$,

$J \rightsquigarrow M \leftrightsquigarrow M \leftrightsquigarrow K$,

...

A similer argument shows that each arc in π must be a single edge, because $\text{non-endpoint}(\pi) \subseteq \overline{\mathrm{an}}(J \cup K \cup M)$

$= \overline{\mathrm{an}}(J \cup K \cup L)$.

This shows π is the shortest in $W(J \hookrightarrow \ast \hookleftarrow K \mid M)$.

Consider any collider m on π. Define the following map from π to π':

1. If $m \in L$, $\pi' = \pi$.

2. If $m \in an(L) \setminus L$, let $m \rightsquigarrow l$ be the shortest

$$\pi' = J \hookrightarrow * \leftrightarrow m \rightsquigarrow l \leftsquigarrow m \leftrightarrow * \hookleftarrow K.$$

3. If $m \in an(J) \setminus J \setminus \widetilde{an}(L)$,

$$\pi' = J \underbrace{\leftsquigarrow}_{\text{no } L \text{ because } m \notin \widetilde{an}(L)} m \leftrightarrow * \hookleftarrow K.$$

4. If $m \in an(K) \setminus K \setminus \widetilde{an}(L)$

$$\pi' = J \hookrightarrow * \leftrightarrow m \rightsquigarrow K.$$

Repeat this for every collider on π, we obtain a walk in $W(J \rightsquigarrow * \leftsquigarrow K \mid L)$, which contradicts the assumption.