CS 477/677 Causal Inference: Homework 4
Graphical Causal Models

Ha Bui
hbui13@jhu.edu

1 Analytical (34 Points)

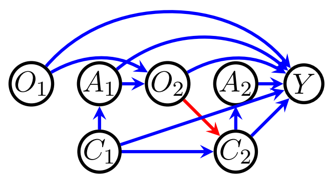


Figure 1: A causal model corresponding to a variation of the simulation study in *Statistical Methods for Dynamic Treatment Regimes*. The red edge is absent in the original simulation study and is added to create *time-dependent confounding* by C_2 and O_2.

1 (4 Points) Assume the red arrow was absent in Fig. 1. If this is true, can we use standard covariate adjustment to estimate $p(Y(a_1, a_2))$? That is, does there exist a set $\mathbf{C} \subseteq \{O_1, C_1, O_2, C_2\}$ such that

$$p(Y(a_1, a_2)) = \sum_{\mathbf{C}} p(Y \mid a_1, a_2, \mathbf{C}) p(\mathbf{C}).$$

Explain.

Solution: Yes. Because:

If we remove the red arrow, then O_2, C_2, C_1 is not a triplet collider. Therefore, C_1, C_2 blocks all backdoor paths (non-causal association) from A_1, A_2 to Y and C_1, C_2 doesn't contain any descendants of A_1, A_2, i.e., if we do SWIGs, we have:

$$A_1, A_2 \perp\!\!\!\perp Y(a_1, a_2) \mid C_1, C_2 \tag{1}$$

Consider $\mathbf{C} = \{C_1, C_2\} \subseteq \{O_1, C_1, O_2, C_2\}$, then:

$$\begin{aligned}
p(Y(a_1, a_2)) &= \sum_{c_1, c_2} p(Y(a_1, a_2), c_1, c_2) \\
&= \sum_{c_1, c_2} p(Y(a_1, a_2) \mid c_1, c_2) p(c_1, c_2 | do(a_1), do(a_2)) \\
&= \sum_{c_1, c_2} p(Y(a_1, a_2) \mid A_1, A_2, c_1, c_2) p(c_1, c_2) \text{ (since 1)} \\
&= \sum_{c_1, c_2} p(Y \mid a_1, a_2, c_1, c_2) p(c_1, c_2) \text{ (consistency).}
\end{aligned}$$

2 ID algorithm (15 points) In the following problem you do not have to submit drawings of intermediate graphs used by the **ID** algorithm. However, please show the derivations (you do not have to re-express the final answer explicitly in terms of the original observed data distribution, and can instead report the answer as a kernel – as long as it is clear what the mapping from the kernel to the observed data distribution is).

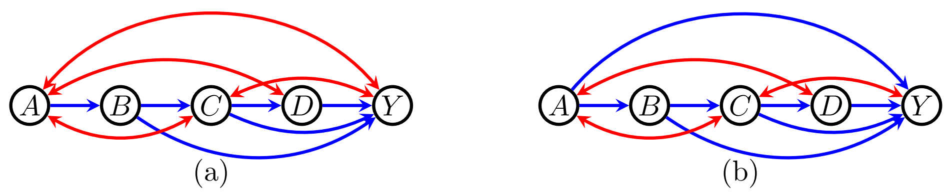


Figure 2: Latent projections representing hidden variable causal DAG models where we wish to identify $p(Y(b,d))$ as a function of the observed data distribution $p(A,B,C,D,Y)$.

- Use the one line **ID** algorithm to try to identify $p(Y(b,d))$ from the observed data distribution $p(A,B,C,D,Y)$in Fig. 2 (a).

  Solution:

  Construct SWIG G(b,d) for fixing b and d, then $\vec{Y}^* = \{Y, C\}$.
  Construct $G_{\vec{Y}*}$, we have one district $\{Y, C\}$.
  Because we can fix $\phi_{A,B,D}(G)$, then:

$$
\begin{aligned}
p(Y(b,d)) &= \sum_{\vec{Y}*\setminus\vec{Y}} \prod_{\vec{D}\in D(G_{\vec{Y}*})} \phi_{\vec{V}\setminus\vec{D}}(p(\vec{V};G) \\
&= \sum_C \phi_{\{A,D\}}\left(\frac{p(A,B,C,D,Y)}{p(B|A)};\phi_B(G)\right) \\
&= \sum_C \phi_{\{A,D\}}\left(q(Y,A,C,D|B);\phi_B(G)\right) \\
&= \sum_C \phi_{\{D\}}\left(\frac{q(Y,A,C,D|B)}{q(A|B,C,D,Y)};\phi_{\{A,B\}}(G)\right) \\
&= \sum_C \phi_{\{D\}}\left(\frac{p(Y,C,D|A,B)p(A)}{\frac{p(Y,C,D|A,B)p(A)}{\sum_A p(Y,C,D|A,B)p(A)}};\phi_{\{A,B\}}(G)\right) \\
&= \sum_C \phi_{\{D\}}\left(\sum_A p(Y,C,D|A,B)p(A);\phi_{\{A,B\}}(G)\right) \\
&= \sum_C \phi_{\{D\}}\left(q(Y,C,D|A,B);\phi_{\{A,B\}}(G)\right) \\
&= \sum_C \frac{q(Y,C,D|A,B)}{q(D|A,B,C)} = \frac{q(Y,D|A,B)}{q(D|A,B,C)} = \frac{q(Y,D|A,B)q(C|A,B)}{q(C,D|A,B)} \\
&= \frac{p(Y|A,B,D)\sum_D p(C|A,B,D)p(D)}{p(C|A,B,D)}.
\end{aligned}
$$

- Use the one line **ID** algorithm to try to identify $p(Y(b,d))$ from the observed data distribution $p(A,B,C,D,Y)$in Fig. 2 (b).

  Solution:

  Construct SWIG G(b,d) for fixing b and d, then $\vec{Y}^* = \{Y, C, A\}$.
  Construct $G_{\vec{Y}*}$, we have two district $\{A, C\}$, $\{C, Y\}$.
  Consider:

$$\sum_{\vec{Y}*\setminus\vec{Y}} \prod_{\vec{D}\in D(G_{\vec{Y}*})} \phi_{\vec{V}\setminus\vec{D}}(p(\vec{V};G)$$
$$= \sum_{A,C} \phi_{\{B,D,Y\}}\left(p(A,B,C,D,Y);G\right) \phi_{\{A,B,D\}}\left(p(A,B,C,D,Y);G\right).$$

  Consider $\phi_{\{B,D,Y\}}(G)$, we can fix Y, then B, then D.
  However, consider $\phi_{\{A,B,D\}}(G)$, we can fix B, but then: can not fix A in $\phi_{\{A,D\}}$ because there is a blue path from A to Y and red path between A, C, and Y. Similarly, we can not fix D in $\phi_{\{A,D\}}$ because there is a blue path from D to Y and red path between D, A, C, and Y.
  Therefore, $p(Y(b,d))$ is not identified.

3 Front-Door and Mediation (5 points) Prove that $p(Y(A, M(a)))$ in Fig. 3 (a) and $p(Y(a))$ in Fig. 3 (b) are identified by the same functional. Note that $p(Y(A, M(a)))$ is not the same distribution as $p(Y(a', M(a)))$. In the latter distribution we evaluate A as a' (for the purposes of Y) and as a (for the purposes of M). In the former distribution, we let A assume its natural value for the purposes of Y, and set A to a for the purposes of M.

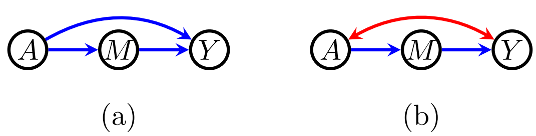


Figure 3: (a) The standard mediation graph. (b) A latent projection corresponding to the front-door criterion.

Solution:
Construct SWIG from Fig. 3 (a), we have:

$$\begin{aligned}
p(Y(A,M(a))) &= \sum_m p(Y(A,M(a))|M(a)=m)p(M(a)=m) \\
&= \sum_m p(Y(A,M(a))|M(a)=m)p(M(a)=m|A) \text{ (since } M(a) \perp\!\!\!\perp A \text{ in SWIG)} \\
&= \sum_m p(Y(A,M(a))|M(a)=m)p(m|a) \text{ (consistency: } A=a \Rightarrow M(a)=m) \\
&= \sum_m p(Y(A,m)|m)p(m|a) \text{ (consistency: } M(a)=m \Rightarrow Y(A,M(a))=Y(A,m)) \\
&= \sum_m \left(\sum_{a'} p(Y|m,a')p(a')\right) p(M=m|a) \text{ (consistency: } Y(A)=Y).
\end{aligned} \tag{2}$$

Construct SWIG from Fig. 3 (b), we have:

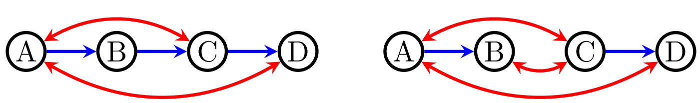
$$\begin{aligned}
p(Y(a)) &= \sum_m p(Y(a)|M(a)=m)p(M(a)=m) \\
&= \sum_m p(Y(a)|M(a)=m)p(m|a) \text{ (since } M(a) \perp\!\!\!\perp A \text{ in SWIG)} \\
&= \sum_m p(Y(a,m)|M(a)=m)p(m|a) \text{ (consistency: } M(a)=m \Rightarrow Y(a,m)=Y(a)) \\
&= \sum_m p(Y(m)|M(a)=m)p(m|a) \text{ (consistency: } Y(m,a)=Y(m)) \\
&= \sum_m p(Y(m))p(m|a) \text{ (since } Y(m) \perp\!\!\!\perp M(a) \text{ in SWIG)} \\
&= \sum_m \left(\sum_{a'} p(Y|m,a')p(a')\right) p(M=m|a) \text{ (since } Y(m) \perp\!\!\!\perp M|A \text{ in SWIG).}
\end{aligned} \tag{3}$$

From 2 and 3, we obtain: $p(Y(A, M(a)))$ in Fig. 3 (a) is the same with $p(Y(A, M(a)))$ in Fig. 3 (b).

4 Verma constraints (10 Points)

- Consider the latent projection in Fig. 4 (a). Does a missing edge between B and D in this graph correspond to a constraint in the observed data distribution $p(A, B, C, D)$?
- Repeat the problem for Fig. 4 (b).

Hint: think about a set of intervention operations and conditioning operations such that in the resulting SWIG, the d-separation corresponding to the missing edge holds. Try to use **ID** to see if the corresponding (possibly conditional) interventional distribution is identified from $p(A, B, C, D)$ *without assuming the missing edge between* B *and* D *is absent.*



(a) (b)

Figure 4: Latent projection ADMGs. with non-adjacent vertices B and D.

Solution:

Assume there is an edge between B and D in Fig. 4 (a) and consider the latent projection, we have:

$$\begin{aligned}
p(A, B, C, D) &= q(A, C|B)q(A, D|B, C) \\
&= p(C|A, B)p(A)p(D|A, B, C)p(A).
\end{aligned}$$

If missing the edge between B and D, we can not fix C, if we fix B, then $D \not\perp\!\!\!\perp B$ in $q(A, D|B, C)$ because there is still a path $B \to C \to D$. Therefore, the missing edge between B and D does not correspond to a Verma constraint in Fig. 4 (a).

Assume there is an edge between B and D in Fig. 4 (b) and consider the latent projection, we have:

$$\begin{aligned} p(A, B, C, D) &= q(A, C)q(B, C|A)q(A, D|B, C) \\ &= p(A, C|\cdot)p(B|A)p(C)p(D|B, C)p(A). \end{aligned}$$

If missing the edge between B and D, we can not fix C or A, if we fix B, then $D \perp\!\!\!\perp B$ in $q(A, D|B, C)$ because there is no path between B and D. Therefore, $q(D|B, C)$ is not a function of B because $q(D|B, C) = q(D|C)$, so the missing edge between B and D corresponds to a Verma constraint in Fig. 4 (b).