

CS 476/676 (Spring 2021): Homework 4

Due: Mar 17, 2021 at 11:59pm EST

Name: Josh Popp

Instructions: This homework requires answering some open-ended questions, proofs, and analysis. This is an individual assignment, not group work. Though you may discuss the problems with your classmates, you must solve the problems and write the solutions independently. As stated in the syllabus, copying code from a classmate or the internet (even with minor changes) constitutes plagiarism. You are required to submit your answers in pdf form (use \LaTeX) in a file called `<your-JHED>-hw4.pdf` to Gradescope under “HW4”. Late submissions will be penalized, except in extenuating circumstances such as medical or family emergency. Submissions submitted 0-24 hours late will be penalized 10%, 24-48 hours late by 20%, 48-72 hours late by 30%, and later than 72 hours by 100%. Late days may be used (if available) to avoid these penalties.

The total assignment is worth 90 points.

Important note: For portions of the assignment that require drawing graphs, we encourage you to use the \LaTeX template provided to draw graphs using the `tikz` package. However, hand drawing graphs/using other software and then including the images into the pdf using `includegraphics` is also acceptable as long as it is neat.

Problem 1 (10 points)

Recall our motivation for using an ADMG $\mathcal{G}(V)$ to summarize the joint distribution on observed variables V in a latent variable DAG $\mathcal{G}(V \cup U)$ was to preserve all **causal** and **statistical** information on the observed variables V without explicitly thinking about the latent variables U . Refer to the lecture slides for precise definitions of what is meant by preserving all causal and statistical information.

Imagine the second rule of the latent projection operator was modified to read as follows:

$V_i \leftrightarrow V_j$ exists in $\mathcal{G}(V)$ if there exists a path $V_i \leftarrow \cdots \rightarrow V_j$ in $\mathcal{G}(V \cup U)$ such that all intermediate vertices are unmeasured variables in U .

That is, we do not have the additional check that there are no colliders along such paths. Answer the following questions.

1) Prove, via a suitable counterexample of a latent variable DAG and the ADMG obtained using the modified latent projection operator, that the absence of this additional check results in ADMGs

that do not correctly preserve statistical information among the observed variables. (5 points)

Latent variable DAG

$$V_1 \leftarrow U_1 \rightarrow U_2 \leftarrow U_3 \rightarrow V_2$$

Proposed ADMG

$$V_1 \leftrightarrow V_2$$

Under the latent variable DAG, $V_1 \perp\!\!\!\perp V_2$, but under the proposed ADMG, this conditional independence does not hold because we failed to account for the collider along the path from V_1 to V_2 when determining whether to place a bidirected edge between the two nodes.

2) Does the absence of the additional check also affect the preservation of causal information in the ADMG $\mathcal{G}(V)$? Why or why not? (5 points)

Yes, this does also affect the preservation of causal information. The latent variable DAG suggests distinct unobserved causes for V_1 and V_2 (and has no open path from V_1 to V_2 since we don't observe the collider U_2), while the proposed ADMG suggests a common unobserved cause for the two variables (and has an open bidirected path $V_1 \leftrightarrow V_2$).

Remember to rely on the precise formal definitions as stated in the lecture slides of what we mean by preserving causal and statistical information when answering 1) and 2).

Problem 2 (10 points)

Please ask any questions regarding Problem 2 (so only for this problem) via **private** Piazza posts, as it is easy to accidentally give away the solution to this problem.

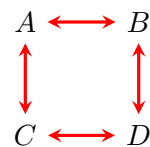
Draw all ADMGs that are Markov equivalent (that is, imply the exact same m-separation relations) to,

1) the ADMG shown in Figure 1a (5 points)

2) the ADMG shown in Figure 1b (5 points)

$$A \leftrightarrow B \leftrightarrow C$$

(a)



(b)

Figure 1

For the first one, constraining to bow-free ADMGs, the following are all Markov equivalent

$$A \longrightarrow B \longleftarrow C$$

(a1)

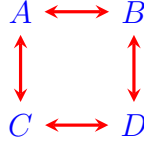
$$A \longrightarrow B \longleftrightarrow C$$

(a2)

$$A \longleftrightarrow B \longleftarrow C$$

(a3)

If we're expanding beyond bow-free ADMGs, any bidirected edge (in any of these four graphs) can be accompanied by a directed edge pointing towards B . For the second one, there are no bow-free Markov equivalent ADMGs.



(b)

Directed edges can be added in any way that avoids a directed path of length greater than 1 (ie, $\circ \rightarrow \circ \rightarrow \circ$ is not allowed)

Problem 3 (10 points)

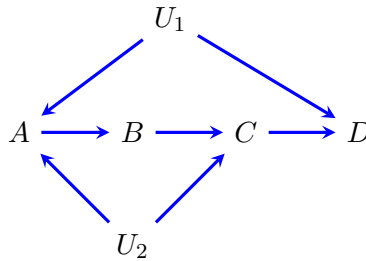


Figure 2

Consider the latent variable DAG shown in Figure 2 where U_1 and U_2 are unmeasured/latent. Prove, using the proof technique shown in Lecture 9 for deriving generalized equality constraints, that the kernel defined below is not a function of B .

$$q_D(D \mid A, B, C) \equiv \frac{\sum_A p(A) \times p(C, D \mid A, B)}{\sum_A p(A) \times p(C \mid A, B)}$$

$$\begin{aligned}
\sum_A p(A) \times p(C, D \mid A, B) &= \sum_A p(A) \sum_{U_1} p(C, D, U_1 \mid A, B) \\
&= \sum_A p(A) \sum_{U_1} p(D, U_1 \mid A, B, C) p(C \mid A, B) \\
&= \sum_A p(A) \sum_{U_1} p(D \mid A, B, C, U_1) p(U_1 \mid A, B, C) p(C \mid A, B) \\
&= \sum_A p(A) p(C \mid A, B) \sum_{U_1} p(D \mid C, U_1) p(U_1 \mid A, B, C) \\
&= \sum_A p(A) p(C \mid A, B) \sum_{U_1} p(D \mid C, U_1) \sum_{U_2} p(U_1, U_2 \mid A, B, C) \\
&= \sum_A p(A) p(C \mid A, B) \sum_{U_1} p(D \mid C, U_1) \sum_{U_2} p(U_1 \mid A, B, C, U_2) \\
&\quad \times p(U_2 \mid A, B, C) \\
&= \sum_A p(A) p(C \mid A, B) \sum_{U_1} p(D \mid C, U_1) \sum_{U_2} p(U_1 \mid A, U_2) \\
&\quad \times p(U_2 \mid A, B, C) \\
&= \sum_{U_1} p(D \mid C, U_1) \\
&\quad \times \sum_A p(A) p(C \mid A, B) \sum_{U_2} p(U_1 \mid A, U_2) p(U_2 \mid A, B, C)
\end{aligned}$$

Problem 4 (20 points)

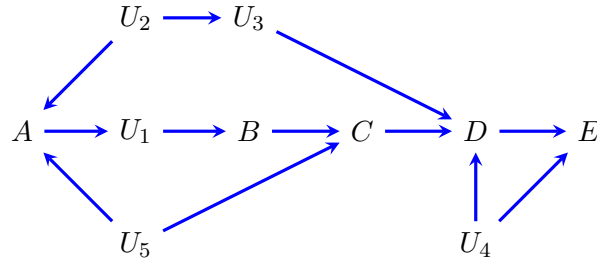


Figure 3

Consider the latent variable DAG in Figure 3 where U_1, U_2, U_3, U_4, U_5 are latent.

1) Draw the latent variable ADMG $\mathcal{G}(V)$ on the observed variables $V = \{A, B, C, D, E\}$. (5 points)



2) Apply m-separation to the latent variable ADMG $\mathcal{G}(V)$ from part 1) to answer the following questions. Justify each of your answers, i.e., list at least one open path if there is one. Is, (3 points each)

- a) $C \perp\!\!\!\perp E \mid D$?
- b) $B \perp\!\!\!\perp D \mid A, C$?
- b) $B \perp\!\!\!\perp E \mid A, C, D$?

a) No, D is a collider, opening the path

$$C \rightarrow D \leftrightarrow E$$

b) No, since A and C are both colliders we open up a path from B to D by conditioning on them

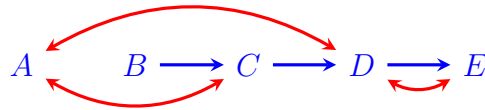
$$B \rightarrow C \leftrightarrow A \leftrightarrow D$$

c) No, once again, C , A , and D are colliders so we open up

$$B \rightarrow C \leftrightarrow A \leftrightarrow D \leftrightarrow E$$

3) Find and describe, using the algorithm provided in Lecture 11, a Verma constraint implied by the ADMG $\mathcal{G}(V)$ from part 1). If you are unable to find a constraint, this might be an indication that you did not derive the ADMG $\mathcal{G}(V)$ in part 1) correctly, or you may not be applying the algorithm correctly. That is, there exists at least one Verma constraint in the correct latent projection ADMG. (6 points)

1. There is an inducing path from A to E , $A \leftrightarrow D \rightarrow E$
2. B is fixable, since it has no bidirected edges. Fixing it breaks the inducing path by removing the direct edge from A to B , such that E is no longer a descendant of A .
3. (In the slides, this is an extra step for when our fixable vertex is one of the endpoints of the inducing path, which it's not here)
4. Since $\text{omb}(B) = \{A\}$, we take $q = p(V)/p(B \mid A)$ and \mathcal{G}^q is the following graph



So we have the Verma constraint $A \perp\!\!\!\perp_q E$

Problem 5 (15 points)

A mixed graph \mathcal{G} on d variables V_1, \dots, V_d consisting of directed (\rightarrow) and bidirected (\leftrightarrow) edges can be summarized via two binary adjacency matrices D and B as follows.

$D_{ij} = 1$ if $V_i \rightarrow V_j$ exists in \mathcal{G} and $D_{ij} = 0$ otherwise.

$B_{ij} = B_{ji} = 1$ if $V_i \leftrightarrow V_j$ exists in \mathcal{G} and $B_{ij} = B_{ji} = 0$ otherwise.

Recall in HW3 you proved that \mathcal{G} is acyclic (contains no directed cycles) if and only if $\text{trace}(e^D) - d = 0$. That is, \mathcal{G} is an acyclic directed mixed graph (ADMG) if and only if $\text{trace}(e^D) - d = 0$. Often, we also want to impose certain requirements on the way bidirected edges appear in \mathcal{G} . This is because when there are no restrictions imposed, \mathcal{G} is not a simple graph (could have more than one edge between each pair of vertices), which complicates statistical inference in the underlying models.¹

Important note: For simplicity, in the following proofs you may assume that the diagonal elements of D and B are always 0, i.e., “self edges” of the form $V_i \rightarrow V_i$ and $V_i \leftrightarrow V_i$ are ruled out apriori.

a) One popular subclass of ADMGs, is the class of bow-free ADMGs. Besides acyclicity, bow-free ADMGs have the additional restriction that for any pair of vertices V_i, V_j , both $V_i \rightarrow V_j$ and $V_i \leftrightarrow V_j$ do not appear in \mathcal{G} .² That is, bow-free ADMGs form the class of all simple ADMGs. Prove that an ADMG \mathcal{G} is bow-free if and only if $\text{sum}(D \circ B) = 0$, where $X \circ Y$ denotes the elementwise matrix product a.k.a Hadamard product of X and Y , and $\text{sum}(X)$ represents taking the sum of all elements of the matrix X . (4 points)

Since $D_{ij} \in \{0, 1\}$ and $B_{ij} \in \{0, 1\} \forall \{i, j\}$, $D_{ij} * B_{ij} \in \{0, 1\} \forall \{i, j\}$. Then $\text{sum}(D \circ B) = \sum_{i,j} D_{ij} * B_{ij} = 0$ IFF $\nexists \{k, l\}$ s.t. $D_{kl} = B_{kl} = 1$. $D_{kl} = B_{kl} = 1$ specifically corresponds to the condition of having a bow between V_k and V_l in \mathcal{G} (both $V_k \rightarrow V_l$ and $V_k \leftrightarrow V_l$), so \mathcal{G} is bow-free IFF $\text{sum}(D \circ B) = 0$

b) Another popular subclass of ADMGs, is the class of ancestral ADMGs. Besides acyclicity, ancestral ADMGs have the additional restriction that for any pair of vertices V_i, V_j , both a directed path $V_i \rightarrow \dots \rightarrow V_j$ and $V_i \leftrightarrow V_j$ do not appear in \mathcal{G} . In other words, for every V_i it is the case that $\text{an}_{\mathcal{G}}(V_i) \cap \text{sib}_{\mathcal{G}}(V_i) = \emptyset$. Given the elementary operations we have used so far in proving various constraints on graphical structures – the matrix exponential e^D and its relation to directed paths in \mathcal{G} (see HW3, Problem 4.1), the elementwise matrix product $X \circ Y$ between two matrices X and Y , and the operation $\text{sum}(X)$ where we sum up all elements of a matrix X – design a function $f(D, B)$ based on these operations that is zero if and only if \mathcal{G} is ancestral. That is, design and prove your function $f(D, B) = 0$ if and only if \mathcal{G} is ancestral. (11 points)

¹In the sense that the corresponding model may not be identified from observed data. That is, its parameters cannot be computed in an unbiased fashion. We will formalize what this means in class, but it is tangential to solving Problem 5.

²The term bow-free comes from the fact that when both $V_i \rightarrow V_j$ and $V_i \leftrightarrow V_j$ are present, the edges often look like a bow ready to shoot (in the archery sense) – you can refer to the lecture slides to see what I mean.

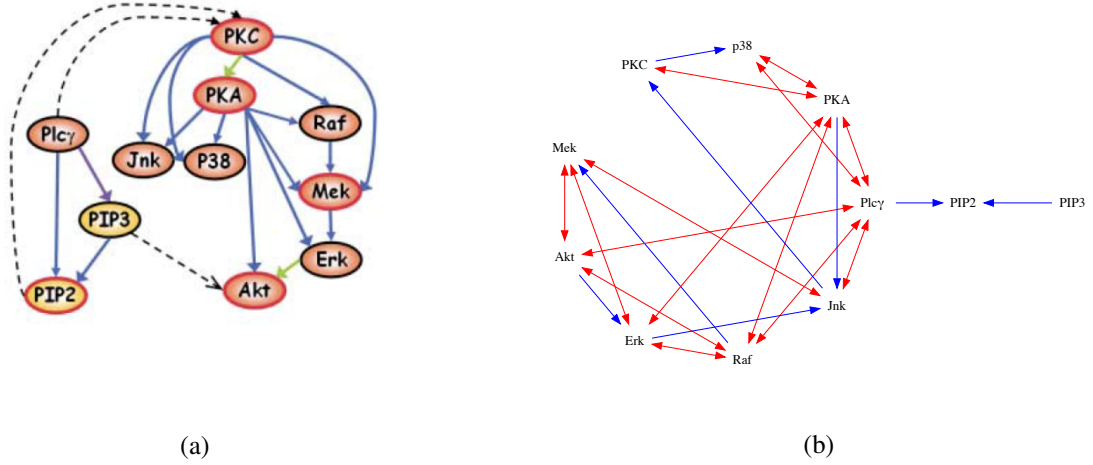


Figure 4

$$f(D, B) = \text{sum}(e^D \circ B) = 0 \text{ iff } \mathcal{G} \text{ ancestral}$$

Proof: $\text{sum}(e^D \circ B) = 0$ iff $\nexists \{k, l\}$ s.t. $(e_{kl}^D \neq 0) \cap (B_{kl} \neq 0)$. So we want to show that $\nexists \{k, l\}$ s.t. $(e_{kl}^D \neq 0) \cap (B_{kl} \neq 0)$ iff \mathcal{G} ancestral.

1. \mathcal{G} ancestral $\implies \nexists \{k, l\}$ s.t. $(e_{kl}^D \neq 0) \cap (B_{kl} \neq 0)$

e_{kl}^D is a normalized count of the total number of paths $V_k \rightarrow \dots \rightarrow V_l$, of any length, in \mathcal{G} . Since \mathcal{G} is ancestral, for any vertices between which a path $V_k \rightarrow \dots \rightarrow V_l$ (of length greater than or equal to 1) exists ($\implies e_{kl}^D \neq 0$), we know there's no bidirected edge $V_k \leftrightarrow V_l$ allowed ($\implies B_{kl} = 0$), and vice versa ($B_{kl} \neq 0 \implies e_{kl}^D = 0$). A path of length zero can only exist from V_k to V_k (and does exist for all k), but we imposed an a priori constraint that $B_{kk} = 0$ for all k .

2. $\neg \mathcal{G}$ ancestral $\implies \exists \{k, l\}$ s.t. $(e_{kl}^D \neq 0) \cap (B_{kl} \neq 0)$

If \mathcal{G} is not ancestral, there exists some pair of nodes V_k and V_l , $k \neq l$ such that $V_k \rightarrow \dots \rightarrow V_l$ and $V_k \leftrightarrow V_l$ both exist in \mathcal{G} . Then $e_{kl}^D > 0$ since this is a normalized count of such paths in D , and $B_{kl} = 1$ by the definition of the matrix B .

Together, these prove that \mathcal{G} ancestral $\iff \nexists \{k, l\}$ s.t. $(e_{kl}^D \neq 0) \cap (B_{kl} \neq 0) \iff \text{sum}(e^D \circ B) = 0$

Problem 6 (25 points)

In the last homework you learned how you might design a greedy causal discovery procedure to learn a DAG from data. If you suspect that there are unobserved variables in the problem, you may design a similar procedure to learn an ADMG from data using the constraints you have derived in Problem 5. The output of such a procedure when applied to the same data from HW4, i.e., from the

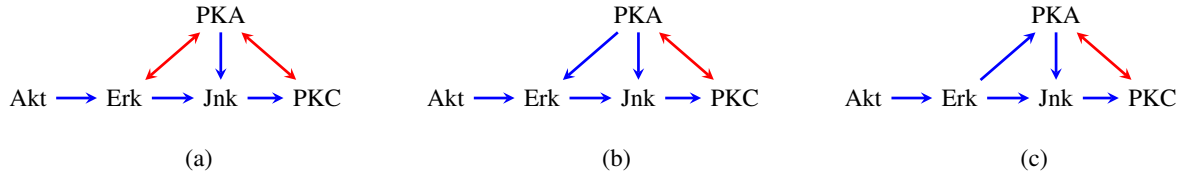


Figure 5: (a) A subgraph of the protein network in Figure 4b

paper by Sachs et al. (2005), is displayed in Figure 4b.³ The DAG learned by the original authors is displayed in Figure 4a.

As you can see, the ADMG in Figure 4b posits that there may exist many relations that arise due to unmeasured confounding rather than causation – this possibility was also acknowledged by the authors of Sachs et al. (2005). Some of these bidirected edges miss the mark, e.g., Figure 4b has $\text{Raf} \rightarrow \text{Mek} \leftrightarrow \text{Erk}$ rather than $\text{Raf} \rightarrow \text{Mek} \rightarrow \text{Erk}$, but others present interesting alternatives to the hypotheses suggested by Sachs et al. (2005). We will study one of these in this problem.

In Figure 4a, there is an edge $\text{PKA} \rightarrow \text{Erk}$. The authors of Sachs et al. (2005) performed an experiment where they manipulated expression of Erk to confirm that this has no downstream effect on PKA, and took this as evidence for the edge $\text{PKA} \rightarrow \text{Erk}$. However, in the absence of an experiment to show that manipulation of PKA does indeed cause downstream change in Erk, we cannot be sure if the association between them is due to causation ($\text{PKA} \rightarrow \text{Erk}$) as suggested by Figure 4a or unmeasured confounding ($\text{PKA} \leftrightarrow \text{Erk}$) as suggested by Figure 4b. Figure 5a zooms in on a subgraph of the ADMG in Figure 4b that is relevant to our attempt to distinguishing these mechanisms from each other. Figure 5b corresponds to an ADMG that would support the Sachs et al. (2005) hypothesis of $\text{PKA} \rightarrow \text{Erk}$ and Figure 5c corresponds to an ADMG that depicts the causal hypothesis $\text{Erk} \rightarrow \text{PKA}$ that was rejected via real biological experiments performed by Sachs et al. (2005). Based on these figures, answer the following questions.

- 1) Consider the pair of variables Akt and PKC. Describe a graphical structure that allows us to recognize the absence of conditional independence relations between Akt and PKC in Figure 5a. (5 points)

The inducing path $\text{Akt} \rightarrow \text{Erk} \leftrightarrow \text{PKA} \leftrightarrow \text{PKC}$ is sufficient to rule out any CI relations between Akt and PKC

- 2) What happens to the same graphical structure from part 1) when the edge $\text{Erk} \leftrightarrow \text{PKA}$ is oriented as either $\text{PKA} \rightarrow \text{Erk}$ or $\text{Erk} \rightarrow \text{PKA}$ as in Figures 5b, c respectively. What are the implications with respect to conditional independences between Akt and PKC? (5 points)

³We will learn the formal details of such procedures in the coming weeks, but they are not important for this problem.

Both of these break the inducing path by making it so one of the nodes on the path is no longer a collider (whichever becomes a parent, specifically). As a result, both of these changes will introduce a CI relation between Akt and PKC (ie $\text{Akt} \perp\!\!\!\perp \text{PKC} \mid Z$ for some subset $Z \in \{\text{Erk}, \text{Jnk}, \text{PKA}\}$). This is because there exists such a subset if and only if there's no inducing path between Akt and PKC. For the case shown in 5b, $Z = \{\text{PKA}, \text{Jnk}\}$. For 5c, $Z = \{\text{Erk}\}$.

3) Assume faithfulness holds. Derive equality constraints between Akt and PKC (which could be a Verma constraint if no ordinary conditional independence is possible) for the following scenarios (3 points each)

a) Find an equality constraint in the ADMG shown in Figure 5b that can be used to distinguish it from the ones shown in Figures 5a and c.

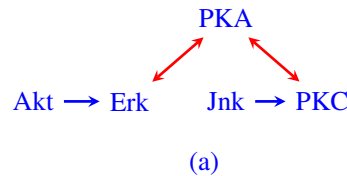
When $\text{Erk} \leftrightarrow \text{PKA}$ is oriented as $\text{PKA} \rightarrow \text{Erk}$, we break the inducing path, since PKA is no longer a collider. This introduces the CI relationship, $\text{Akt} \perp\!\!\!\perp \text{PKC} \mid \text{PKA}, \text{Jnk}$. This can be rephrased in the equality constraint $p(\text{Akt} \mid \text{PKA}, \text{Jnk}, \text{PKC}) = p(\text{Akt} \mid \text{PKA}, \text{Jnk})$. We can test this equality constraint using the odds ratio.

b) Say testing the above constraint leads us to reject the ADMG model in Figure 5b. Find an equality constraint in the ADMG shown in Figure 5c that can be used to distinguish it from the one in Figure 5a.

When $\text{Erk} \leftrightarrow \text{PKA}$ is oriented as $\text{Erk} \rightarrow \text{PKA}$, we break the inducing path, since Erk is no longer a collider. This introduces the CI relationship, $\text{Akt} \perp\!\!\!\perp \text{PKC} \mid \text{Erk}$. This can be rephrased in the equality constraint $p(\text{Akt} \mid \text{Erk}, \text{PKC}) = p(\text{Akt} \mid \text{Erk})$. We can test this equality constraint using the odds ratio.

c) Say testing the above constraint also leads us to reject the ADMG model in Figure 5c. Find an equality constraint in the ADMG shown in Figure 5a that would allow us to reject or fail to reject this model.

As previously mentioned, the ADMG in 5a doesn't have any conditional independences, so this time we do need to use a Verma constraint. Jnk is fixable, and fixing it does break the inducing path $\text{Akt} \rightarrow \text{Erk} \leftrightarrow \text{PKA} \leftrightarrow \text{PKC}$ since Erk and PKA are no longer ancestors of PKC in \mathcal{G}^q shown below



In the reweighted $q = \frac{p(\text{Akt}, \text{Erk}, \text{PKA}, \text{Jnk}, \text{PKC})}{p(\text{Jnk} \mid \text{Erk}, \text{PKA})}$, we have $\text{PKC} \perp\!\!\!\perp_q \text{Akt}$, corresponding to a generalized equality constraint $q(\text{PKC} \mid \text{Akt}) = q(\text{PKC})$

Remember all constraints for parts a), b), and c) must be between Akt and PKC.

4) Given the above exercise in part 3), argue that the three causal hypotheses of $\text{Erk} \leftrightarrow \text{PKA}$, $\text{PKA} \rightarrow \text{Erk}$, and $\text{Erk} \rightarrow \text{PKA}$ can be distinguished using observed data without the need for a physical lab experiment. In doing so, briefly describe how to conduct a statistical test based on a Verma constraint (you may use the odds ratio as a measure of association) and how this allows us to avoid performing new (potentially expensive or difficult) experiments in the lab. (6 points)

Testing the equality constraints from (a) and (b) is pretty straightforward, we can apply the methods we've used previously in the class. Compute the odds ratio $\text{OR}(\text{Akt}, \text{Erk} \mid Z)$ (filling in the appropriate Z as described above), apply bootstrapping to find a 95% confidence interval, and reject the model if 1 falls outside of the interval. We can also, however, test the Verma constraint with only the data that was already collected using the reweighted distribution $q = \frac{p(\text{Akt}, \text{Erk}, \text{PKA}, \text{Jnk}, \text{PKC})}{p(\text{Jnk} \mid \text{Erk}, \text{PKA})}$ (note the denominator that we need for reweighting relies only on data that we've measured) and then applying the same odds ratio-based hypothesis testing scheme to test $\text{PKC} \perp\!\!\!\perp_q \text{Akt}$ with this reweighted distribution.

References

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A., and Nolan, G. P. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529.