

Assignment 1

ALECK ZHAO

March 8, 2019

Problem 1

Consider the DAG in Figure 1.

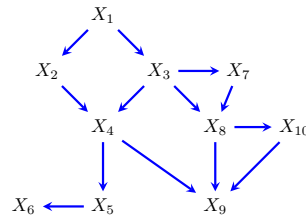


Figure 1

- a) List all the independencies that comprise the local Markov property for this DAG.

Answer. We have the independencies

$$\begin{aligned}
 X_2 &\perp\!\!\!\perp \{X_3, X_7, X_8, X_{10}\} \mid X_1 \\
 X_3 &\perp\!\!\!\perp X_2 \mid X_1 \\
 X_4 &\perp\!\!\!\perp \{X_1, X_7, X_8, X_{10}\} \mid \{X_2, X_3\} \\
 X_5 &\perp\!\!\!\perp \{X_1, X_2, X_3, X_7, X_8, X_9, X_{10}\} \mid X_4 \\
 X_6 &\perp\!\!\!\perp \{X_1, X_2, X_3, X_4, X_7, X_8, X_9, X_{10}\} \mid X_5 \\
 X_7 &\perp\!\!\!\perp \{X_1, X_2, X_4, X_5, X_6\} \mid X_3 \\
 X_8 &\perp\!\!\!\perp \{X_1, X_2, X_4, X_5, X_6\} \mid \{X_3, X_7\} \\
 X_9 &\perp\!\!\!\perp \{X_1, X_2, X_3, X_5, X_6, X_7\} \mid \{X_4, X_8, X_{10}\} \\
 X_{10} &\perp\!\!\!\perp \{X_1, X_2, X_3, X_4, X_5, X_6, X_7\} \mid X_8
 \end{aligned}$$

- b) Is $X_2 \perp_d X_9 \mid X_4$? Is $X_7 \perp_d X_5 \mid \{X_3, X_8\}$? Is $\{X_2, X_4\} \perp_d X_7 \mid \{X_6, X_9, X_{10}\}$?

Answer. $X_2 \perp_d X_9 \mid X_4$: This is not true, there exists a d -connecting path $X_2 \rightarrow X_1 \rightarrow X_3 \rightarrow X_8 \rightarrow X_9$ where none of the intermediate vertices are colliders, yet none are X_4 .

$X_7 \perp_d X_5 \mid \{X_3, X_8\}$: This is true. Every path from X_7 to X_5 must pass through either X_3 or X_8 , and neither are colliders, so there are no d -connecting paths.

$\{X_2, X_4\} \perp_d X_7 \mid \{X_6, X_9, X_{10}\}$: This is not true, there exists a d -connecting path $X_4 \rightarrow X_3 \rightarrow X_7$ and X_3 is not a collider but is not in $\{X_6, X_9, X_{10}\}$.

Problem 2

Given a DAG \mathcal{G} define: $MB(X_i, \mathcal{G}) \equiv pa(X_i) \cup ch(X_i) \cup pa(ch(X_i))$. Prove, using d-separation, that this set satisfies $X_i \perp\!\!\!\perp \mathbf{X} \setminus \{MB(X_i, \mathcal{G}), X_i\} | MB(X_i, \mathcal{G})$.

Proof. Consider $X \in \mathbf{X} \setminus \{pa(X_i) \cup ch(X_i) \cup pa(ch(X_i)) \cup X_i\}$, and consider any path π from X to X_i .

If π passes through $X_p \in pa(X_i)$ right before X_i , then it cannot be d -connecting because X_p is a non-collider yet is in $MB(X_i)$.

If π passes through $X_c \in ch(X_i)$ right before X_i , then it either went through $X_{pc} \in pa(ch(X_i))$ right before, or it did not. If it did, then X_{pc} is a non-collider yet is in $MB(X_i)$. If it did not, then X_c is a non-collider yet is in $MB(X_i)$.

There are no other possible paths, so since all possible π are d -separated, it follows that $X_i \perp_d \mathbf{X} \setminus \{MB(X_i) \cup X_i\} | MB(X_i)$. In DAGs, d -separation implies conditional independence, so we have the result as desired. \square

Problem 3

Consider an undirected graph \mathcal{G} which satisfies the local Markov property for UGs. Prove that it also satisfies the pairwise Markov property using the graphoid axioms. That is, assume that $X_i \perp\!\!\!\perp \mathbf{X} \setminus cl(X_i, \mathcal{G}) | ne(X_i, \mathcal{G})$ and that $X_i \not\sim X_j$ in \mathcal{G} , and derive that $X_i \perp\!\!\!\perp X_j | \mathbf{X} \setminus \{X_i, X_j\}$. (Hint: this is quite easy if you pick the right graphoid axiom!)

Proof. Since $X_i \not\sim X_j$, we know that $X_j \notin cl(X_i) \implies X_j \in \mathbf{X} \setminus cl(X_i)$. We write $cl(X_i) = X_i \cup ne(X_i)$ and

$$\mathbf{X} \setminus cl(X_i) = X_j \cup [\mathbf{X} \setminus (X_j \cup X_i \cup ne(X_i))]$$

so by the graphoid axiom of weak union, we have

$$\begin{aligned} X_i \perp\!\!\!\perp \mathbf{X} \setminus cl(X_i) | ne(X_i) &\implies X_i \perp\!\!\!\perp X_j \cup [\mathbf{X} \setminus (X_j \cup X_i \cup ne(X_i))] | ne(X_i) \\ &\implies X_i \perp\!\!\!\perp X_j | \mathbf{X} \setminus (X_j \cup X_i \cup ne(X_i)) \cup ne(X_i) \\ &\implies X_i \perp\!\!\!\perp X_j | \mathbf{X} \setminus \{X_i, X_j\} \end{aligned}$$

as desired. □

Problem 4

a) Consider the DAG \mathcal{G} in Figure 2. Which of the DAGs in Figure 3 are Markov equivalent to \mathcal{G} ?

Answer. (a) is not equivalent because $A \rightarrow B \leftarrow C$ was unshielded in the original but is not a v -structure here.

(b) is not equivalent because $A \rightarrow B \leftarrow C$ was unshielded in the original but is not here.

(c) is not equivalent because $A \rightarrow E \leftarrow C$ was unshielded in the original but is not here.

(d) is not equivalent because $C \rightarrow E \leftarrow A$ was unshielded in the original but is not here.

(e) is equivalent. The original unshielded colliders $A \rightarrow E \leftarrow C, A \rightarrow E \leftarrow D, B \rightarrow E \leftarrow D, A \rightarrow B \leftarrow C$ are all preserved here

b) How many DAGs are Markov equivalent to the chain $X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n$ (excluding itself)?

Solution. Any Markov equivalent DAG cannot have any colliders $X_i \rightarrow X_{i+1} \leftarrow X_{i+2}$. We can construct equivalent graphs as such: beginning from (X_1, X_2) , we can choose the orientation of the edge, and then continue to the right. However, as soon as an edge is assigned \rightarrow , all subsequent edges must be assigned \rightarrow , otherwise we will have a collider. Thus, the first k edges are \leftarrow , the last $n - k - 1$ edges must be \rightarrow , where k ranges from 1 (since we are excluding the original DAG) to $n - 1$, so there are $n - 1$ equivalent DAGs \square

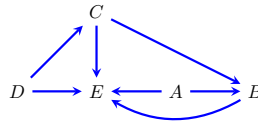


Figure 2

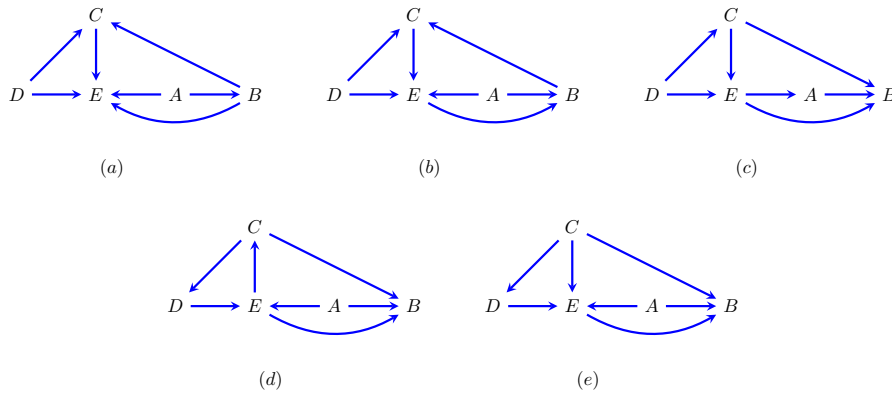
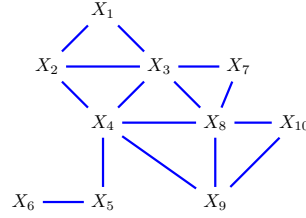


Figure 3

Problem 5

Consider the DAG \mathcal{G} in Figure 1 and its moralized graph \mathcal{G}^m .



- a) How many factor potentials appear in the factorization property for \mathcal{G}^m , when the factorization is by *maximal* cliques?

Answer. The moralized graph is shown above. Here, the maximal cliques are the 3-cliques and the pairs (X_4, X_5) and (X_5, X_6) . Thus, there are 8 factor potentials.

- b) If each factor is parameterized by one parameter for each variable which appears in its scope, plus one parameter for each pair in the scope, and one for each three-way interaction (each triple in the scope), and so on... how many parameters are there in total in this representation of the joint distribution? (You can assume that the single-variable parameters appear only once for each variable, so you're not "double counting" when the same variable appears in multiple cliques.)

Answer. There are 10 parameters for the individual variables, 15 parameters for the pairwise interactions, and 6 parameters for the three-way interactions, for a total of 31 parameters.

- c) How many parameters would be required to specify the joint without a graphical model, assuming binary variables?

Answer. There are 10 variables, and thus $2^{10} - 1 = 1023$ parameters required.

- d) How many parameters would be required to specify the joint using the DAG factorization, assuming binary variables?

Answer. We have the factorization

$$p(\mathbf{x}) = p(x_1)p(x_2 | x_1)p(x_3 | x_1)p(x_4 | x_2, x_3)p(x_5 | x_4)p(x_6 | x_5)p(x_7 | x_3) \\ \times p(x_8 | x_3, x_7)p(x_9 | x_4, x_8, x_{10})p(x_{10} | x_8)$$

In general, for a conditional probability of the form $p(x_i | x_{j_1}, x_{j_2} \dots x_{j_n})$, we require $2(2^n - 1)$ parameters because there are $2^n - 1$ parameters for the conditioned variables (since the last one is determined) for each possible value of x_i . Thus, the total number of parameters needed is

$$1 + 2(2 - 1) + 2(2 - 1) + 2(2 - 1) + 2(2 - 1) + 2(2 - 1) + 2(2 - 1) + 2(2^2 - 1) + 2(2^3 - 1) + 2(2 - 1) = 35$$

- e) How many factor potentials appear when the MRF factorization is by pairwise cliques?

Answer. There are 15 pairwise cliques.

- f) How many parameters are there total in this latter representation of the joint distribution, with same scheme as above?

Answer. There are 10 parameters for the individual variables and 15 parameters for the pairwise interactions, for a total of 25 parameters.

Problem 6

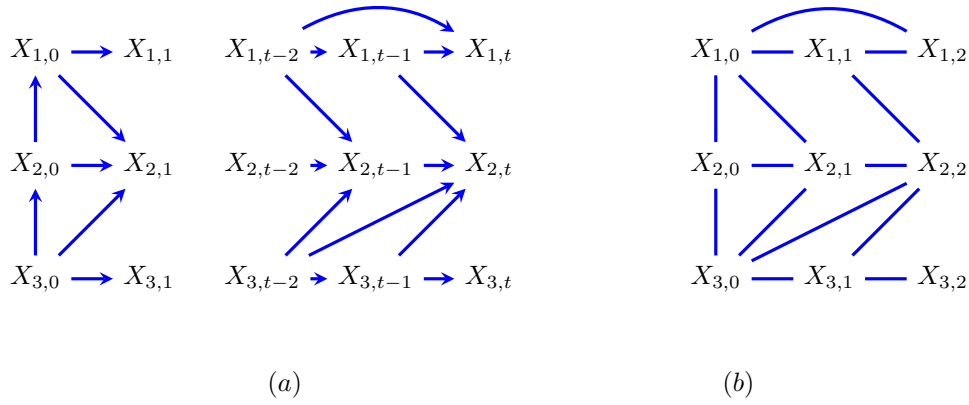


Figure 4

Consider the dynamic Bayesian network (DBN) in Figure 4(a). As we discussed, DBNs are typically specified in two parts: the “prior network” \mathcal{B}_0 (left) and the “transition network” \mathcal{B}_\rightarrow (right). Here the prior network spans 2 time-steps because the DBN has a Markov order of 2 (edges from $t - 2$ to t). In Figure 4(b) we have a seemingly similar undirected graph, which we can imagine being extended into the future, following the same structure over time. (Note: this is not the moralized version of the DBN in Figure 4(a).) Both of these models can be used to represent a set of variables evolving over time, with $t \in \{0, 1, \dots, T\}$.

- a) Consider a parameterization of the DBN where all variables are binary, and there is some repeating structure: $p(x_{i,t} | pa(X_{i,t}, \mathcal{B}_\rightarrow))$ is the same for all $t > 1$. How many parameters does one need to specify this DBN model? If someone gave you data generated from this model, and asked: “are $X_{1,50}$ and $X_{3,50}$ conditionally independent given $X_{2,50}$?” what would you answer? What if one *also* conditioned on the entire history (all variables prior to $t = 50$)?

Solution. To describe the entire DBN, we need to describe the starting states, and then the probabilities for each transition are the same. The starting states are

$$p(x_{1,0} | x_{2,0})p(x_{2,0} | x_{3,0})p(x_{3,0})p(x_{1,1} | x_{0,0})p(x_{2,1} | x_{1,0}, x_{2,0}, x_{3,0})p(x_{3,1} | x_{3,0})$$

which requires $2 + 2 + 1 + 2 + 2(2^3 - 1) + 2 = 23$ parameters. Then the probabilities for the transition at time t for $t \geq 2$ are

$$p(x_{1,t} | x_{1,t-2}, x_{1,t-1})p(x_{2,t} | x_{1,t-1}, x_{2,t-1}, x_{3,t-1}, x_{3,t-2})p(x_{3,t} | x_{3,t-1})$$

note that $p(x_{3,t} | x_{3,t-1})$ is already accounted for by $p(x_{3,1} | x_{3,0})$ from the starting state, this transition requires $2(2^2 - 1) + 2(2^4 - 1) = 36$ parameters, for a total of $23 + 36 = 59$ parameters to describe the entire DBN.

We have the active trail $X_{1,50} \leftarrow X_{1,49} \leftarrow \dots \leftarrow X_{1,0} \leftarrow X_{2,0} \leftarrow X_{3,0} \rightarrow X_{3,1} \rightarrow \dots \rightarrow X_{3,50}$ which does not have any colliders and does not pass through $X_{2,50}$, so there is no conditional independence here.

If we conditioned on the entire history, then any trail going backwards in time is blocked since any node adjacent to $X_{1,50}$ going backwards is a non-collider and thus a blocker. Then any path going forward in time must collide at $X_{2,n}$ for some n , which is a collider with no observed descendants, and is thus blocked. Thus the conditional independence holds. \square

- b) Consider an MRF parameterization of Figure 4(b) in terms of maximal cliques. Using the same parameterization scheme as in question 5 (i.e., for cliques of size k , consider one parameter for the k -way interaction, one for

each k -1-way interactions, and so on... again without double-counting), how many parameters does one need to specify this MRF model? If someone gave you data generated from this model, and asked: “are $X_{1,50}$ and $X_{3,50}$ conditionally independent given $X_{2,50}$?” what would you answer? What if one *also* conditioned on the entire history (all variables prior to $t = 50$)?

Solution. We have the repeating 3-cliques $(X_{1,0}, X_{1,1}, X_{1,2})$, $(X_{2,1}, X_{2,2}, X_{3,0})$, and $(X_{2,2}, X_{3,0}, X_{3,1})$. Then we have the repeating pairwise interactions $X_{1,0} - X_{1,1}$, $X_{2,0} - X_{2,1}$, $X_{3,0} - X_{3,1}$, $X_{1,0} - X_{2,1}$, $X_{2,1} - X_{3,0}$, $X_{1,0} - X_{1,2}$ and $X_{2,2} - X_{3,0}$. We have the initial 3-cliques $(X_{1,0}, X_{2,0}, X_{2,1})$ and $(X_{2,0}, X_{2,1}, X_{3,0})$, and the pairwise interactions $X_{1,0} - X_{2,0}$ and $X_{2,0} - X_{3,0}$. And finally we can describe the individual nodes with a single parameter. Thus, the total number of parameters needed is 15.

As in part (a), we can easily find a path from $X_{1,50}$ to $X_{3,50}$ that doesn't pass through $X_{2,50}$ so there is no conditional independence here. Even if we conditioned on the entire history, we can take the path $X_{1,50} - X_{2,51} - X_{3,50}$ that does not pass through any history and thus there is no conditional independence here either. \square

Problem 7

Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ be three disjoint sets of nodes in a DAG \mathcal{G} . Prove that \mathbf{C} d-separates \mathbf{A} from \mathbf{B} in \mathcal{G} if and only if \mathbf{C} separates \mathbf{A} from \mathbf{B} in $(\mathcal{G}_{an(\mathbf{A}, \mathbf{B}, \mathbf{C})})^m$. (Recall that we use the convention that $X_i \in an(X_i, \mathbf{G})$.)

Proof. (\implies) : Suppose there exists a path π in $(\mathcal{G}_{an(\mathbf{A}, \mathbf{B}, \mathbf{C})})^m$ from \mathbf{A} to \mathbf{B} that does not pass through \mathbf{C} . Consider this same directed path in \mathcal{G} .

If the directed analogue has a v -structure $X \rightarrow Z \leftarrow Y$, then if $Z \in an(\mathbf{C})$, this is active. Otherwise, if $Z \notin an(\mathbf{C})$, we must have $Z \in an(\mathbf{A}, \mathbf{B})$, and WLOG $Z \in an(\mathbf{B})$. Then we can find a directed path from Z to \mathbf{B} that does not pass through \mathbf{C} , since otherwise Z would be in $an(\mathbf{C})$. Furthermore, all nodes along this path are in $an(\mathbf{B})$, so it would also be a valid path in the moralized graph.

If π contains an edge that isn't in the directed analogue, this must be a moralized edge between $X - Y$ corresponding to the v -structure $X \rightarrow Z \leftarrow Y$ in the original graph. If $Z \notin an(\mathbf{A}, \mathbf{B}, \mathbf{C})$ for all v -structures containing X and Y , then we would not have had the moralized edge between $X - Y$ in the first place. If $Z \in an(\mathbf{C})$ then this is active and we are done. Otherwise, if $Z \in an(\mathbf{A}, \mathbf{B})$, we proceed by the previous argument.

Using the above, we can always construct an active trail from \mathbf{A} to \mathbf{B} if we have a path in the moralized graph from \mathbf{A} to \mathbf{B} not passing through \mathbf{C} , a contradiction.

(\impliedby) : Suppose there exists an active trail π from \mathbf{A} to \mathbf{B} in \mathcal{G} . Then on this trail, if there exists a v -structure $X \rightarrow Z \leftarrow Y$, it must be that $Z \in an(\mathbf{C})$, and therefore $X, Y \in an(\mathbf{C})$ as well. However, we must have $X, Y \notin \mathbf{C}$ since otherwise we would have a cycle since there exists a directed path from Z to either X or Y .

Otherwise, on any other part of the trail $X \rightarrow Z \rightarrow Y$ or $X \leftarrow Z \rightarrow Y$ or $X \leftarrow Z \leftarrow Y$ we must have $Z \notin \mathbf{C}$ in all cases in order for the trail to be active. Consider the case $X \rightarrow Z \rightarrow Y$. WLOG the rest of the path is from Y to \mathbf{B} , but suppose $Y \notin an(\mathbf{B})$. Then, somewhere along this path, we must have a collider, but as we established earlier, such a collider must be in $an(\mathbf{C})$, so $Y \in an(\mathbf{C})$, and thus $X, Z \in an(\mathbf{C})$ as well. Otherwise, if $Y \in an(\mathbf{B})$, it follows that $X, Z \in an(\mathbf{B})$. A similar argument can be made for the other structures, so it follows that $X, Y, Z \in an(\mathbf{A}, \mathbf{B}, \mathbf{C})$.

Combining these two facts, we can construct a path in the moralized graph from \mathbf{A} to \mathbf{B} using this active trail. If we have the v -structure from earlier $X \rightarrow Z \leftarrow Y$, we can use the moralized edge $X - Y$ since $X, Y \in an(\mathbf{C})$. If we are not in a v -structure, all the nodes must be in $an(\mathbf{A}, \mathbf{B}, \mathbf{C})$ anyway, and thus all nodes along this constructed path are in $an(\mathbf{A}, \mathbf{B}, \mathbf{C})$ and none of them are in \mathbf{C} , a contradiction. \square

Problem 8

Implement a method/function which takes a matrix representation of a DAG as input, along with a d-separation query, and outputs a value TRUE or FALSE. Specifically, there is an attached dag.txt file on Piazza that accompanies this assignment, which is a 100x100 matrix of zeroes and ones. This matrix represents a DAG in the following way: if $X_i \rightarrow X_j$ then the ij entry is 1, otherwise it is 0. Your function should come with an executable file which takes the matrix as input, along with a query of the form: `number number list`, and returns TRUE iff the variables corresponding to those numbers are d-separated given the numbers in `list`. You may implement d-separation anyway you like, but *do not use implementations of d-separation that are already out there in available software*. The point of this exercise is to get practice implementing this concept yourself, not to copy some code off the internet or out of the source of some available software package. Furthermore, your implementation should work for matrices of any (reasonable) size, not just 100x100.

Use your function to determine if:

- a) $61 \perp_d 68 | 4, 19, 90$?
- b) $55 \perp_d 27 | 4, 8, 9, 12, 29, 32, 40, 44, 45, 48, 50, 52$?

Along with your answers to the above queries, write 2-5 sentences explaining your implementation, and why you chose to go about it that way.

Answer. I chose the implementation based on the pseudocode presented in the book. It seemed like an efficient method since it was linear time in the size of the graph, and is also correct because it came from the book.

What to hand in: The code for the programming assignment will be handed in separately on github. Please create a **private** repository called “d-separation”, and add the TA Noam Finkelstein (username: nsfinkelstein) as a collaborator to the repository on github. The repository should include three files:

1. The dag.txt data file, available on Piazza.
2. Your code, which reads in the data file and prints your answers to problem 8 parts a) and b). Exactly two lines should be printed, and the text of each line should be either ‘TRUE’ or ‘FALSE’.
3. A bash script called ‘run.sh’ which compiles (if necessary) and runs your code.

Please make sure your variables and functions are aptly named, and your code is clear and readable, with comments where appropriate. We may try to execute your code on a new graph