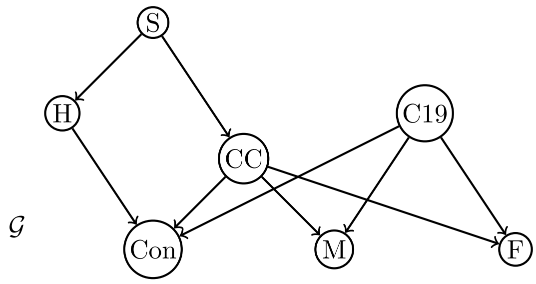
Probabilistic Graphical Models: Problem Set 2 (Solutions)

Svante Linusson, Liam Solus
KTH Royal Institute of Technology

8 oktober 2021

1. (Medical diagnosis revisited again) Suppose that the random variables in our medical diagnosis model from the first lesson $(Con, M, F, H, CC, C19, S)$ have joint distribution $\mathbb{P} \in \mathcal{M}_F(\mathcal{G})$ for the following DAG $\mathcal{G}$:



   (a) What is the relationship between this graph $\mathcal{G}$ and the undirected graph from problem 1 on the first practice problem set?

   (b) Do you think this model is better or worse than our undirected graphical model for the same system of variables on practice problem set 1? Why or why not?

   (c) Suppose we were unsure if $\mathbb{P}$ was in $\mathcal{M}_F(\mathcal{G})$. What is the fewest number of CI relations we would need to check to convince ourselves that it is?

Solution:

a) The undirected graph from problem 1 is the moral graph of this DAG.
b) In general, undirected graphs and DAGs are good in different ways. In this case we can e.g. deduce the undirected graph form the DAG, but not vice versa.
c) By the theorem in Lecture 2 we can check either (DL) or (DG). The first has fewer CI relations to check, one for each node i: $X_i \perp\!\!\!\perp X_{\mathrm{nd}(i)\setminus \mathrm{pa}(i)} \mid X_{\mathrm{pa}(i)}$. That is, 7 CI relations in this case.

2. There are four DAGs on three nodes whose underlying undirected graph (i.e. **skeleton**) is a path. What are the d-separation statements for each of these four DAGs? What do you notice? What are the implications of what you notice for distributions Markov to anyone of these DAGs?
Solution:
The four possible DAGs are $X \leftarrow Y \leftarrow Z, X \leftarrow Y \rightarrow Z, X \rightarrow Y \rightarrow Z$ and $X \rightarrow Y \leftarrow Z$. The first three have X, Z d-separated given $\{Y\}$, whereas the last has X, Z d-separated given $\emptyset$. The conclusion we wanted you to draw was that many DAGs can have the same d-separation statements.

3. A DAG $\mathcal{G} = ([m], E)$ with linear extension $\pi = 12\cdots p$ is called **perfect** if for all $i \in [m]$, the set of nodes $\mathrm{pa}_{\mathcal{G}}(i)$ form a clique in $\mathcal{G}$. Let $\overline{\mathcal{G}}$ denote the skeleton of $\mathcal{G}$. Show that $\mathcal{M}_{\mathrm{Glo}}(\mathcal{G}) = \mathcal{M}_{\mathrm{Glo}}(\overline{\mathcal{G}})$ whenever $\mathcal{G}$ is perfect.
Solution:
That $\mathrm{pa}_{\mathcal{G}}(i)$ form a clique in $\mathcal{G}$ implies that $G^m = \overline{G}$ and $\overline{G}$ is chordal. If G satisfies (DG) then by Proposition 1 in Lecture 2 $\overline{G}$ obeys the condition (UG).

For the other direction one may use induction over m. For $m = 1$ it is obvious.
Induction step: Assume true for all graphs on $m-1$ vertices. Given DAG G such that $\overline{G}$ satisfies (UG) we let j be a sink in G, so $\mathrm{pa}_G(j) = N_{\overline{G}}(j)$ and form a complete set. Clearly, $\overline{G}\backslash\{j\} = \overline{G \setminus \{j\}}$ also satisfies (UG) and since $G \setminus \{j\}$ is also a perfect graph the induction hypothesis gives that it satisfies (DG). We must now show that adding back j (DG) is still satisfied. For the (UG) to imply the (DG) we need to prove that A, B d-separated given C in G implies that C separates A and B in $\overline{G}$. If $j \notin A \cup B \cup C$ we are done by induction.

There are two cases left to consider. If $j \in A$ we will prove the contrapositive statement. Assume C does not separate A and B and not that if C does not separate $A\backslash j$ and B then we are done by induction. So we have a path starting at j ending in $b \in B$. Let i be the neighbor of j on this path. If $i = b$ then clearly A and B are not d-separated given C either. Otherwise we deduce that C does not separate i and B and by induction i and B are not d-separated given C in G. But $i \in \mathrm{pa}_G(j)$ and thus we can use the same path to prove that j (and thus A) is not d-separated from B given C, since i can never be a collider node.

Second, if $j \in C$ and C does not separate A and B in $\overline{G}$ then $C\backslash j$ does not separate A and B either. By induction, A and B are not d-separated given $C\backslash j$. Since j is a sink it must be a collider node on any path and can thus not interrupt a d-connecting path by being added to the conditioning set C. Hence, A and B are not d-separated given C and we are done.

---

4. Consider three discrete random variables X_1, X_2, X_3.
   (a) Suppose that X_1 and X_2 are both binary with state space $\{0, 1\}$. Show that if $f_{X_1,X_2}(0, 0) = f_{X_1}(0)f_{X_2}(0)$ then $X_1 \perp\!\!\!\perp X_2$.
   (b) Is it the case that for binary X_1, X_2, and X_3 with state space $\{0, 1\}$ that $X_1 \perp\!\!\!\perp X_2 \mid X_3 = 0$ implies $X_1 \perp\!\!\!\perp X_2 \mid X_3$? Provide either a proof or a counterexample. In either case, determine the DAG with the fewest edges with respect to which the distribution(s) in your result satisfy the global Markov property.

Solution:
a) Use e.g. that

$$\begin{aligned}
f_{X_1,X_2}(1,0) &= f_{X_2|X_1}(1 \mid 0)f_{X_2}(0),\\
&= (1 - f_{X_2|X_1}(0 \mid 0))f_{X_2}(0),\\
&= f_{X_2}(0) - f_{X_1,X_2}(0,0),\\
&= f_{X_2}(0) - f_{X_1}(0)f_{X_2}(0), \qquad \text{(by assumption in problem)}\\
&= f_{X_2}(0)(1 - f_{X_1}(0)),\\
&= f_{X_2}(0)f_{X_1}(1).
\end{aligned}$$

The rest follows similarly.

b) no, give counterexample

---

5. Let $\mathcal{G} = ([m], E)$ a DAG with linear extension $\pi = 12\cdots m$ and consider the **linear Gaussian DAG model**

$$\mathbf{X} = A\,\mathbf{X} + E$$

for the lower triangular $m \times m$ matrix $A = [a_{i,j}]_{i,j=0}^m$ in which $a_{i,j} \neq 0$ if and only if $i \in \mathrm{pa}_{\mathcal{G}}(j)$, and for the vector $E = [\varepsilon_1, \ldots, \varepsilon_m]^T$ of independent standard normal random variables $\varepsilon_i \sim \mathcal{N}(\mathbf{0}, 1)$.

   (a) Show that $\mathbf{X}$ has a multivariate normal distribution $N(0, \Sigma')$, where $\Sigma' = (1-A)^{-1}(1-A)^{-T}$ and 1 denotes the $m \times m$ identity matrix.
   (b) Consider a linear Gaussian DAG model $\mathbf{X}$ for the DAG $\mathcal{G} = ([4], E)$ where

$$E = \{1 \to 2, 1 \to 3, 2 \to 4, 3 \to 4\}.$$

Does $\mathbf{X}$ factorize according to $\mathcal{G}$?

Solution:

(a) We apply Theorem 1.2 from the introductory notes, noting that $E = (\mathbf{1} - A)^{-1}\mathbf{X}$.

(b) Yes, applying the result from part (a) and the formula for checking conditional independence in Theorem 1.3 of the introductory notes.