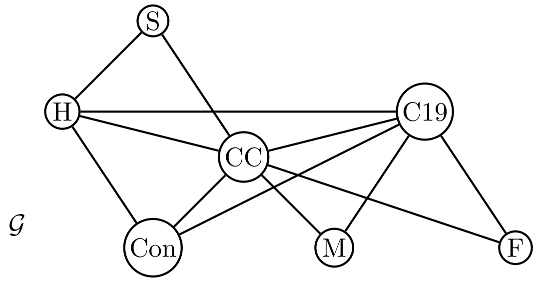
Probabilistic Graphical Models: Problem Set 1 (Solutions)

Svante Linusson, Liam Solus
KTH Royal Institute of Technology

27 September 2021

1. (Medical diagnosis revisited) Lets see how representing our distribution from our first example with a UG can help us reduce the size of our representation. Suppose we represent our model with the UG



(a) Suppose our distribution P over $(Con, M, F, H, CC, C19, S)$ satisfies the global Markov property with respect to $\mathcal{G}$. Do we need to make any assumptions about P in order to ensure that P factorizes according to $\mathcal{G}$?

(b) How many parameters are needed to represent our distribution P if we assume it factorizes according to $\mathcal{G}$?

(c) If we assume P satisfies the global Markov property with respect to $\mathcal{G}$, what CI relations $A \perp\!\!\!\perp B \mid C$ do we know hold in P where $A = \{S\}$ and $B = \{M\}$ are singletons?

(d) Does this model seem reasonable to you? For instance, do the CI relations you detected in the previous part make intuitive sense? What about the dependencies represented by the edges in $\mathcal{G}$?

Solution:

a) No further assumptions are needed. This is because $\mathcal{G}$ is chordal. So by Proposition 3 from Lecture 3 (1st lecture on Sept 27), since P satisfies the global Markov property with respect to $\mathcal{G}$ it also factors according to $\mathcal{G}$.

b) To determine the number of parameters needed to represent our distribution P, we need to determine the domain of the nonnegative functions $\psi_A(\mathbf{x}_A)$, where A is a maximal clique in $\mathcal{G}$. If $A = \{i_1, \dots, i_k\}$ is a maximal clique in $\mathcal{G}$ then we need a parameter that tells us the value of $\psi_A(x_{i_1}, \cdots, x_{i_k})$ for every realization $x_{i_1} \dots x_{i_k}$ of X_A. (Note that we do not get to subtract 1 from this number like we did in the DAG setting! This is because we do not know that the function $\psi_A(\mathbf{x}_A)$ is a distribution.) The table below specifies the maximal cliques in $\mathcal{G}$ and the number of parameters needed per clique.

Maximal cliques	Number of parameters needed
$\{S, H, CC\}$	$4 \cdot 2 \cdot 2 = 16$
$\{H, CC, C19, Con\}$	$2 \cdot 2 \cdot 2 \cdot 2 = 16$
$\{CC, C19, M\}$	$2 \cdot 2 \cdot 2 = 8$
$\{CC, C19, F\}$	$2 \cdot 2 \cdot 2 = 8$

Hence, the total number of parameters needed is $16+16+8+8=48$. Notice that 48 is still better than 63 (the number of parameters needed without any graphical representation) but not as good as 28, the number of parameters we needed for our DAG representation.

c) Any path from S to M contains either H or CC. If CC is not in the separating set there is a connecting path from S to H. Hence, any separating set must contain CC. If the separating A set also contains H, then adding in any other choice of nodes in $\{Con, C19, F\}$ yields a separating set. This gives $2^3 = 8$ CI relations that must hold for P to satisfy the global Markov property with respect to $\mathcal{G}$: $S \perp\!\!\!\perp H \mid A$, where $A = \{H, CC\} \cup B$ for $B \subset \{Con, C19, F\}$.

On the other hand, if H is not in the separating set, then $C19$ must be in the set in order for it to separate M and S. Adding in either of the nodes Con or F also yields a valid separating set. Hence, S is separated from M given any set $A = \{CC, C19\} \cup B$ where $B \subset \{Con, F\}$. This gives 2^2 additional CI relations that P must satisfy in order to satisfy the global Markov property with respect to $\mathcal{G}$: $S \perp\!\!\!\perp H \mid A$, where $A = \{CC, C19\} \cup B$ for $B \subset \{Con, F\}$.

d) Maybe the model should include an edge between S and $C19$, given that we have prior knowledge suggesting correlation between season and COVID-19. Perhaps the edge between H and $C19$ is also weird. This would suggest that S is independent of M given CC and $C19$, which seems reasonable. Other than that, the model seems to capture some believable relations amongst the variables in the system.

---

2. Suppose P is a distribution on $(X_1, \dots, X_m)$ satisfying the intersection axiom and suppose that we are given an undirected graph $\mathcal{G} = ([m], E)$. What is the quickest way to check that P satisfies the global Markov property with respect to $\mathcal{G}$? That is, what is the fewest number of CI relations that we need to check hold in P to verify that P is Markov to $\mathcal{G}$?

Solution:

Since P satisfies the intersection axiom then (G), (L) and (P) are all equivalent by Theorem 4 from lecture 3 (the 1st lecture on September 27). Hence, to check if P satisfies the global Markov property with respect to $\mathcal{G}$ we can check any one of these three properties. Checking (G) is at least as hard as checking (L) or (P), so checking either of the latter two is optimal. Checking (L) requires checking at most m relations: one per node (although we need not check relations for nodes whose set of nondescendants are precisely their parents, nor relations for nodes i where $j = \mathrm{nd}_{\mathcal{G}}(i) \setminus \mathrm{pa}_{\mathcal{G}}(i)$, $i = \mathrm{nd}_{\mathcal{G}}(j) \setminus \mathrm{pa}_{\mathcal{G}}(j)$ and $\mathrm{pa}_{\mathcal{G}}(i) = \mathrm{pa}_{\mathcal{G}}(j)$). Assuming we are not aware of such special structure of $\mathcal{G}$, we take m as the number of checks needed to verify (L). As for checking (P), we would need to verify one relation for every non-edge of (G). This yields $\binom{m}{2} - |E|$ checks. So if $\binom{m}{2} - |E| < m$, we should check (P), but if $\binom{m}{2} - |E| \geq m$ we should check (L). If one takes into consideration the above special cases, we could develop more complicated and specific formulas of the same nature.

---

3. Let $(X_1, \dots, X_m)$ have joint distribution P that is multivariate normally distributed as $\mathcal{N}(0, \Sigma)$. Suppose also that P is satisfies the global Markov property with respect to an undirected graph $\mathcal{G} = ([m], E)$. The **concentration matrix** of P is defined to be $K := \Sigma^{-1}$. For $i, j \in [m]$, show that the entry $K_{i,j}$ of K equals zero if and only if $X_i \perp\!\!\!\perp X_j \mid X_{[p]\setminus\{i,j\}}$. (Hint: How would Cramer take the inverse of a matrix?)

Solution:

By Cramer's Rule,

$$K = \frac{1}{\det(\Sigma)} \operatorname{Adj}(\Sigma),$$

where $\operatorname{Adj}(\Sigma)$ denotes the adjugate of Σ:

$$\operatorname{Adj}(\Sigma) = \left((-1)^{i+j} M_{j,i}\right), \qquad \text{where } M_{j,i} = \det(\Sigma_{[m]\setminus\{j\},[m]\setminus\{i\}}).$$

By Theorem 1.3 from the introductory notes, $X_i \perp\!\!\!\perp X_j \mid X_{[m]\setminus\{i,j\}}$ if and only if $\mathrm{Cov}[X_i, X_j | \mathbf{X}_{[m]\setminus\{i,j\}}] = \left(\Sigma_{\{i\}\cup\{j\}|[m]\setminus\{i,j\}}\right)_{i,j} = 0$. By the Guttman rank additivity formula for Schur Complements, we have that

$$\begin{aligned}
\operatorname{Rank}\left[\Sigma_{[m]\setminus\{j\},[m]\setminus\{i\}}\right] &= \operatorname{Rank}\left[\Sigma_{\{i\}\cup\{j\}|[m]\setminus\{i,j\}}\right], \\
&= \operatorname{Rank}\left[\Sigma_{[m]\setminus\{i,j\},[m]\setminus\{i,j\}}\right] + \operatorname{Rank}\left[\Sigma_{i,j} - \Sigma_{i,[m]\setminus\{i,j\}}\Sigma^{-1}_{[m]\setminus\{i,j\},[m]\setminus\{i,j\}}\Sigma_{[m]\setminus\{i,j\},j}\right], \\
&= \operatorname{Rank}\left[\Sigma_{[m]\setminus\{i,j\},[m]\setminus\{i,j\}}\right] + \operatorname{Rank}\left[\left(\Sigma_{\{i\}\cup\{j\}|[m]\setminus\{i,j\}}\right)_{i,j}\right], \\
&= \operatorname{Rank}\left[\Sigma_{[m]\setminus\{i,j\},[m]\setminus\{i,j\}}\right] + \operatorname{Rank}[0], \\
&= m-2.
\end{aligned}$$

Since $\Sigma_{[m]\setminus\{i,j\},[m]\setminus\{i,j\}}$ is an $(m-2)\times(m-2)$ positive definite matrix, we obtain $\operatorname{Rank}\left[\Sigma_{[m]\setminus\{j\},[m]\setminus\{i\}}\right] = m-2$ if and only if $\det(\Sigma_{[m]\setminus\{j\},[m]\setminus\{i\}}) = 0$, which happens if and only if $M_{j,i} = 0$. Hence, $X_i \perp\!\!\!\perp X_j \mid X_{[m]\setminus\{i,j\}}$ if and only if $K_{i,j} = 0$.

4. Suppose $P \sim (X_1, \dots, X_m)$ has probability density function $f(\mathbf{x})$, and let $A, B, C \subset [m]$ be disjoint subsets with $A, B \neq \emptyset$. Show that $X_A \perp\!\!\!\perp X_B \mid X_C$ if and only if there exist functions h, k such that

$$f(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C) = h(\mathbf{x}_A, \mathbf{x}_C)k(\mathbf{x}_B, \mathbf{x}_C).$$

Solution:

Suppose that $X_A \perp\!\!\!\perp X_B \mid X_C$. Then, by definition,

$$\begin{aligned}
f(x_A, x_B \mid x_C) &= f(x_A \mid x_C)f(x_B \mid x_C), \qquad \text{or equivalently,} \\
\frac{f(x_A, x_B, x_C)}{f(x_C)} &= \frac{f(x_A, x_C)}{f(x_C)}\frac{f(x_B, x_C)}{f(x_C)}.
\end{aligned}$$

So taking $h(x_A, x_C) = f(x_A, x_C)/\sqrt{f(x_C)}$ and $k(x_B, x_C) = f(x_B, x_C)/\sqrt{f(x_C)}$, proves the forward direction.

Suppose now that there exist functions h, k such that

$$f(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C) = h(\mathbf{x}_A, \mathbf{x}_C)k(\mathbf{x}_B, \mathbf{x}_C).$$

Then

$$\begin{aligned}
f(\mathbf{x}_A \mid \mathbf{x}_C) &= \frac{\int f(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C)d\,\mathbf{x}_B}{f(\mathbf{x}_C)}, \\
&= \frac{h(\mathbf{x}_A, \mathbf{x}_C)\int k(\mathbf{x}_A, \mathbf{x}_C)d\,\mathbf{x}_C}{f(\mathbf{x}_C)},
\end{aligned}$$

and similarly,

$$\begin{aligned}
f(\mathbf{x}_B \mid \mathbf{x}_C) &= \frac{\int f(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C)d\,\mathbf{x}_A}{f(\mathbf{x}_C)}, \\
&= \frac{k(\mathbf{x}_B, \mathbf{x}_C)\int h(\mathbf{x}_A, \mathbf{x}_C)d\,\mathbf{x}_C}{f(\mathbf{x}_C)}.
\end{aligned}$$

Hence,

$$\begin{aligned}
f(\mathbf{x}_A \mid \mathbf{x}_C)f(\mathbf{x}_B \mid \mathbf{x}_C) &= \frac{h(\mathbf{x}_A, \mathbf{x}_C)\int k(\mathbf{x}_A, \mathbf{x}_C)d\,\mathbf{x}_C}{f(\mathbf{x}_C)} \cdot \frac{k(\mathbf{x}_B, \mathbf{x}_C)\int h(\mathbf{x}_A, \mathbf{x}_C)d\,\mathbf{x}_C}{f(\mathbf{x}_C)}, \\
&= f(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C)\frac{\int\int h(\mathbf{x}_A, \mathbf{x}_C)k(\mathbf{x}_B, \mathbf{x}_C)d\,\mathbf{x}_A\,d\,\mathbf{x}_B}{f(\mathbf{x}_C)^2}, \\
&= f(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C)\frac{f(\mathbf{x}_C)}{f(\mathbf{x}_C)^2}, \\
&= f(\mathbf{x}_A, \mathbf{x}_B \mid \mathbf{x}_C),
\end{aligned}$$

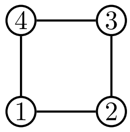
which completes the proof.

5. In this problem, our goal is to observe that if a distribution is nonpositive then it can obey the global Markov property for a nonchordal UG $\mathcal{G}$ and, at the same time, fail to factorize according to $\mathcal{G}$. In other words, we will see that the hypotheses of Theorem 5 or Proposition 3 from class are necessary.

   Consider the distribution P over (X_1, X_2, X_3, X_4), for X_i having outcomes $\{0, 1\}$ for all $i = 1, \ldots, 4$, that assigns probability $1/8$ to each of the following outcomes and probability zero to all other outcomes:

$$\begin{array}{cccc} (0,0,0,0) & (1,0,0,0) & (1,1,0,0) & (1,1,1,0) \\ (0,0,0,1) & (0,0,1,1) & (0,1,1,1) & (1,1,1,1) \end{array}$$

   Let $\mathcal{G}$ denote the UG



   (a) Show that P is satisfies the global Markov property with respect to $\mathcal{G}$.
   (b) Show that P does not factorize according to $\mathcal{G}$. (Hint: Use proof by contradiction)

Solution:

a) To show that P satisfies (G) with respect to $\mathcal{G}$ we need only check:

$$X_1 \perp\!\!\!\perp X_3 \mid X_{\{2,4\}} \qquad \text{and} \qquad X_2 \perp\!\!\!\perp X_4 \mid X_{\{1,3\}}.$$

We see that

$$\begin{aligned} P_{X_{\{2,4\}}}(0,0) &= 1/8 + 1/8 = 1/4, \\ P_{X_{\{2,4\}}}(0,1) &= 1/8 + 1/8 = 1/4, \\ P_{X_{\{2,4\}}}(1,0) &= 1/8 + 1/8 = 1/4, \\ P_{X_{\{2,4\}}}(1,1) &= 1/8 + 1/8 = 1/4, \end{aligned}$$

and so

$$\begin{aligned} P_{X_{\{1,3\}}|X_{\{2,4\}}}(0,0|0,0) &= 1/2, \\ P_{X_{\{1,3\}}|X_{\{2,4\}}}(0,1|0,0) &= 0, \\ P_{X_{\{1,3\}}|X_{\{2,4\}}}(1,0|0,0) &= 1/2, \\ P_{X_{\{1,3\}}|X_{\{2,4\}}}(1,1|0,0) &= 0, \end{aligned}$$

and we can check that $P_{X_1|X_{\{2,4\}}}(0|0,0) = 1/2$ and $P_{X_1|X_{\{2,4\}}}(0|0,0) = 1$. Hence,

$$P_{X_{\{1,3\}}|\{2,4\}}(0,0|0,0) = P_{X_1|X_{\{2,4\}}}(0|0,0)P_{X_3|X_{\{2,4\}}}(0|0,0).$$

Computing the remaining conditional probabilities we can see that $X_1 \perp\!\!\!\perp X_3 \mid X_{\{2,4\}}$ and $X_2 \perp\!\!\!\perp X_4 \mid X_{\{1,3\}}$. So P satisfies the global Markov property with respect to $\mathcal{G}$.

b) Suppose that $P(X_1, X_2, X_3, X_4)$ factors according to $\mathcal{G}$. Then

$$1/8 = P(0,0,0,0) = \psi_{\{1,2\}}(0,0)\psi_{\{2,3\}}(0,0)\psi_{\{3,4\}}(0,0)\psi_{\{1,4\}}(0,0),$$

and

$$0 = P(0,0,1,0) = \psi_{\{1,2\}}(0,0)\psi_{\{2,3\}}(0,1)\psi_{\{3,4\}}(1,0)\psi_{\{1,4\}}(0,0).$$

Hence, either $\psi_{\{2,3\}}(0,1) = 0$ or $\psi_{\{3,4\}}(1,0) = 0$ (or both). However,

$$1/8 = P(0,0,1,1) = \psi_{\{1,2\}}(0,0)\psi_{\{2,3\}}(0,1)\psi_{\{3,4\}}(1,1)\psi_{\{1,4\}}(1,0),$$

which implies $\psi_{\{2,3\}}(0,1) \neq 0$. However, we also have that

$$0 = P(1,1,1,0) = \psi_{\{1,2\}}(1,1)\psi_{\{2,3\}}(1,1)\psi_{\{3,4\}}(1,0)\psi_{\{1,4\}}(1,0),$$

which implies that $\psi_{\{3,4\}}(1,0) \neq 0$, which is a contradiction. Hence, P does not factor according to $\mathcal{G}$.