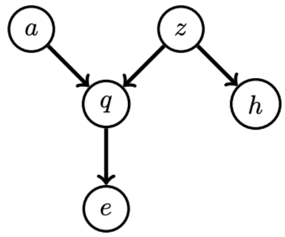
COMP64101 - Lecture 3 - Exercise sheet with answers

Exercise 1 [Directed graph concepts] For this exercise, consider the directed graph shown below.



(a) List all trails in the graph (of maximal length).

(b) List all directed paths in the graph (of maximal length).

(c) What are the descendants of z?

(d) What are the non-descendants of q?

(e) Which of the following orderings are topological with respect to the graph?

- (a, z, h, q, e)
- (a, z, e, h, q)
- (z, a, q, h, e)
- (z, q, e, a, h)

Answers:

(a) We have the maximal-length trails (a, q, e), (a, q, z, h), (h, z, q, e); and the corresponding ones with swapped start and end nodes.

(b) We have the maximal-length paths (a, q, e), (z, q, e), and (z, h).

(c) $\text{desc}(z) = \{q, e, h\}$.

(d) $\text{nondesc}(q) = \{a, z, h, e\} \setminus \{e\} = \{a, z, h\}$.

(e)
- (a, z, h, q, e): yes.
- (a, z, e, h, q): no (q is a parent of e and thus has to come before e in the ordering).
- (z, a, q, h, e): yes.
- (z, q, e, a, h): no (a is a parent of q and thus has to come before q in the ordering).

Exercise 2 [Ordered and local Markov properties, d-separation] For this exercise, we continue considering the directed graph shown in Exercise 1.

(a) The ordering (z, h, a, q, e) is topological to the graph. What are the independencies that follow from the ordered Markov property?

(b) What are the independencies that follow from the local Markov property?

(c) The independency relations obtained via the ordered and local Markov property include $q \perp\!\!\!\perp h \mid \{a, z\}$. Verify the independency using d-separation.

(d) Use d-separation to check whether $a \perp\!\!\!\perp h \mid e$ holds.

(e) Assume all variables in the graph are binary. How many numbers do you need to specify, or learn from data, in order to fully specify the probability distribution?

Answers:

(a) The predecessor sets are:

$$\mathrm{pre}(z) = \varnothing, \quad \mathrm{pre}(h) = \{z\}, \quad \mathrm{pre}(a) = \{z, h\}, \quad \mathrm{pre}(q) = \{z, h, a\}, \quad \mathrm{pre}(e) = \{z, h, a, q\}.$$

The parent sets are independent from the topological ordering chosen. In the lecture, we have seen that they are:

$$\mathrm{pa}(z) = \varnothing, \quad \mathrm{pa}(h) = \{z\}, \quad \mathrm{pa}(a) = \varnothing, \quad \mathrm{pa}(q) = \{a, z\}, \quad \mathrm{pa}(e) = \{q\}.$$

The ordered Markov property reads $x_i \perp\!\!\!\perp \mathrm{pre}(i) \setminus \mathrm{pa}(i) \mid \mathrm{pa}(i)$ where the x_i refer to the ordered variables, e.g. $x_1 = z$, $x_2 = h$, $x_3 = a$, etc. We have:

$$\mathrm{pre}(h) \setminus \mathrm{pa}(h) = \varnothing \quad \mathrm{pre}(a) \setminus \mathrm{pa}(a) = \{z, h\} \quad \mathrm{pre}(q) \setminus \mathrm{pa}(q) = \{h\} \quad \mathrm{pre}(e) \setminus \mathrm{pa}(e) = \{z, h, a\}.$$

We thus obtain:

$$h \perp\!\!\!\perp \varnothing \mid z \qquad a \perp\!\!\!\perp \{z, h\} \qquad q \perp\!\!\!\perp h \mid \{a, z\} \qquad e \perp\!\!\!\perp \{z, h, a\} \mid q$$

The relation $h \perp\!\!\!\perp \varnothing \mid z$ should be understood as "there is no variable from which h is independent given z" and should thus be dropped from the list. Generally, having a variable later in the topological ordering allows one to possibly obtain a stronger independence relation because the set $\mathrm{pre} \setminus \mathrm{pa}$ can only increase when the predecessor set pre becomes larger.

(b) The sets of non-descendants are:

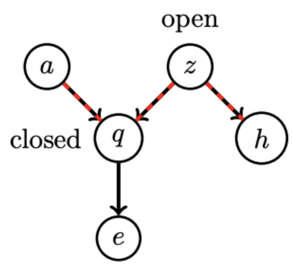
$$\begin{aligned}
\mathrm{nondesc}(a) &= \{z, h\} \\
\mathrm{nondesc}(z) &= \{a\} \\
\mathrm{nondesc}(h) &= \{a, z, q, e\} \\
\mathrm{nondesc}(q) &= \{a, z, h\} \\
\mathrm{nondesc}(e) &= \{a, q, z, h\}.
\end{aligned}$$

With the parent sets as before, the independencies that follow from the local Markov property are $x_i \perp\!\!\!\perp \mathrm{nondesc}(x_i) \setminus \mathrm{pa}(i) \mid \mathrm{pa}(i)$, i.e.

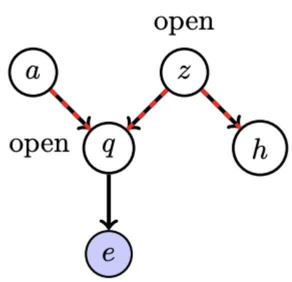
$$a \perp\!\!\!\perp \{z, h\} \qquad z \perp\!\!\!\perp a \qquad h \perp\!\!\!\perp \{a, q, e\} \mid z \qquad q \perp\!\!\!\perp h \mid \{a, z\} \qquad e \perp\!\!\!\perp \{a, z, h\} \mid q.$$

(c) The only trail from q to h goes through z which is in a tail-tail configuration. Since z is part of the conditioning set, the trail is blocked and the result follows.

(d) The trail from a to h is shown below in red together with the default states of the nodes along the trail.



Conditioning on e opens the q node since q in a collider configuration on the path.



The trail from a to h is thus active, which means that the relationship does not hold because $a \not\perp h \mid e$ for some distributions that factorise over the graph.

(e) The graph defines a set of probability mass functions (pmf) that factorise as

$$p(a, z, q, h, e) = p(a)p(z)p(q|a, z)p(h|z)p(e|q).$$

To specify a member of the set, we need to specify the (conditional) pmfs on the right-hand side. The (conditional) pmfs can be seen as tables, and the number of elements that we need to specified in the tables are:

- 1 for $p(a)$
- 1 for $p(z)$
- 4 for $p(q|a, z)$
- 2 for $p(h|z)$
- 2 for $p(e|q)$

In total, there are 10 numbers to specify. This is in contrast to $2^5 - 1 = 31$ for a distribution without independencies. Note that the number of parameters to specify could be further reduced by making parametric assumptions.

Exercise 3 **[More on ordered and local Markov properties, d-separation]** For this exercise, we continue considering the directed graph shown in Exercise 1.

(a) Why can the ordered or local Markov property not be used to check whether $a \perp\!\!\!\perp h \mid e$ may hold?

(b) The independency relations obtained via the ordered and local Markov property include $a \perp\!\!\!\perp \{z, h\}$. Verify the independency using d-separation.

(c) Determine the Markov blanket of z.

(d) Verify that $q \perp\!\!\!\perp h \mid \{a, z\}$ holds by manipulating the probability distribution induced by the graph.

Answers:

(a) The independencies that follow from the ordered or local Markov property require conditioning on parent sets. However, e is not a parent of any node so that the above independence assertion cannot be checked via the ordered or local Markov property.

(b) All paths from a to z or h pass through the node q that forms a head-head connection along that trail. Since neither q nor its descendant e is part of the conditioning set, the trail is blocked and the independence relation follows.

(c) The Markov blanket is given by the parents, children, and co-parents. Hence: $\mathrm{MB}(z) = \{a, q, h\}$.

(d) A basic definition of conditional independence $x_1 \perp\!\!\!\perp x_2 \mid x_3$ is that the (conditional) joint $p(x_1, x_2|x_3)$ equals the product of the (conditional) marginals $p(x_1|x_3)$ and $p(x_2|x_3)$. For discrete random variables:

$$x_1 \perp\!\!\!\perp x_2 \mid x_3 \quad \Longleftrightarrow \quad p(x_1, x_2|x_3) = \left(\sum_{x_2} p(x_1, x_2|x_3)\right)\left(\sum_{x_1} p(x_1, x_2|x_3)\right)$$

(Use integrals in case of continuous random variables.) We thus answer the question by showing that

$$p(q, h|a, z) = \left(\sum_{h} p(q, h|a, z)\right)\left(\sum_{q} p(q, h|a, z)\right). \qquad (\star)$$

First, note that the graph defines a set of probability density or mass functions that factorise as

$$p(a, z, q, h, e) = p(a)p(z)p(q|a, z)p(h|z)p(e|q)$$

We then use the sum-rule to compute the joint distribution of (a, z, q, h), i.e. the distribution of all the variables that occur in $p(q, h|a, z)$, as shown next:

$$p(a, z, q, h) = \sum_e p(a, z, q, h, e) = \sum_e p(a)p(z)p(q|a, z)p(h|z)p(e|q) = p(a)p(z)p(q|a, z)p(h|z)\sum_e p(e|q)$$

and since $\sum_e p(e|q) = 1$ we get $p(a, z, q, h) = p(a)p(z)p(q|a, z)p(h|z)$. We further have

$$p(a, z) = \sum_{q,h} p(a, z, q, h) = \sum_{q,h} p(a)p(z)p(q|a, z)p(h|z) = p(a)p(z)\sum_q p(q|a, z)\sum_h p(h|z) = p(a)p(z)$$

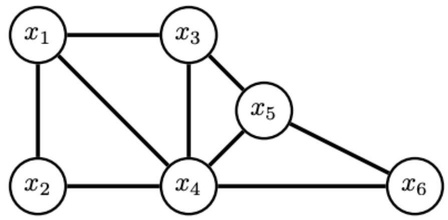
and hence, using Bayes' rule,

$$p(q, h|a, z) = \frac{p(a, z, q, h)}{p(a, z)} = \frac{p(a)p(z)p(q|a, z)p(h|z)}{p(a)p(z)} = p(q|a, z)p(h|z).$$

Now, we see that $p(q|a, z)$ and $p(h|z)$ are the marginals of $p(q, h|a, z)$, i.e. $p(q|a, z) = \sum_h p(q, h|a, z)$ and $p(h|z) = \sum_q p(q, h|a, z)$. Thus, equation $(\star)$ holds, proving that $q \perp\!\!\!\perp h \mid \{a, z\}$ holds.

Notice that using the graph to determine the independence property is much easier than manipulating the pmf/pdf. This is even more so for larger numbers of variables.

Exercise 4 [Factorisation and independencies for undirected graphical models] Consider the undirected graphical model defined by the following graph.



(a) What is the form of Gibbs distributions induced by the graph?

(b) Let p be a pdf that factorises according to the graph. Does $p(x_3|x_2, x_4) = p(x_3|x_4)$ hold?

(c) Explain why $x_2 \perp\!\!\!\perp x_5 \mid \{x_1, x_3, x_4, x_6\}$ holds for all distributions that factorise over the graph.

(d) Assume you would like to approximate $\mathbb{E}(x_1x_2x_5|x_3, x_4)$ with a sample average. Do you need to have joint observations for all five variables $x_1, \ldots, x_5$? Explain.

Answers:

(a) The graph in the figure has four maximal cliques:

$$(x_1, x_2, x_4), \quad (x_1, x_3, x_4), \quad (x_3, x_4, x_5), \quad (x_4, x_5, x_6).$$

The Gibbs distributions are thus of the form

$$p(x_1, \ldots, x_6) \propto \psi_{c_1}(x_1, x_2, x_4)\psi_{c_2}(x_1, x_3, x_4)\psi_{c_3}(x_3, x_4, x_5)\psi_{c_4}(x_4, x_5, x_6)$$

where ψ_c are nonnegative functions, one for each maximal clique c.

(b) $p(x_3|x_2, x_4) = p(x_3|x_4)$ means that $x_3 \perp\!\!\!\perp x_2 \mid x_4$. We can use the graph to check whether this generally holds for pdfs that factorise according to the graph. There are multiple trails from x_3 to x_2, including the trail (x_3, x_1, x_2), which is not blocked by x_4. From the graph, we thus cannot conclude that $x_3 \perp\!\!\!\perp x_2 \mid x_4$, and $p(x_3|x_2, x_4) = p(x_3|x_4)$ will generally not hold. (The relation may hold for some carefully defined factors ψ_c, but not in general.)

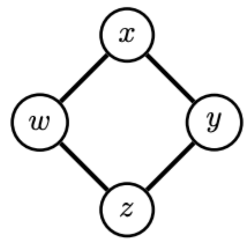
(c) Distributions that factorise over the graph satisfy the pairwise Markov property. Since x_2 and x_5 are not neighbours, and x_1, x_3, x_4, x_6 are the remaining nodes in the graph, the independence relation follows from the pairwise Markov property.

(d) In the graph, all trails from $\{x_1, x_2\}$ to x_5 are blocked by $\{x_3, x_4\}$, so that $x_1, x_2 \perp\!\!\!\perp x_5 \mid x_3, x_4$. We thus have

$$\mathbb{E}(x_1x_2x_5|x_3, x_4) = \mathbb{E}(x_1x_2|x_3, x_4)\mathbb{E}(x_5|x_3, x_4).$$

Hence, we only need joint observations of (x_1, x_2, x_3, x_4) and (x_3, x_4, x_5) for the said approximation. Notice, in particular, that variables x_1, x_2 and x_5 do not need to be jointly measured.

Exercise 5 [Factorisation and independencies for undirected graphical models] Consider the undirected graphical model defined by the following graph.



(a) How do the pdfs/pmfs corresponding to this undirected graphical model factorise?

(b) List all independencies that hold for this undirected graphical model.

Answers:

(a) The maximal cliques are (x, w), (w, z), (z, y) and (x, y). The undirected graphical model thus consists of pdfs/pmfs that factorise as follows

$$p(x, w, z, y) \propto \psi_{c_1}(x, w)\psi_{c_2}(w, z)\psi_{c_3}(z, y)\psi_{c_4}(x, y)$$

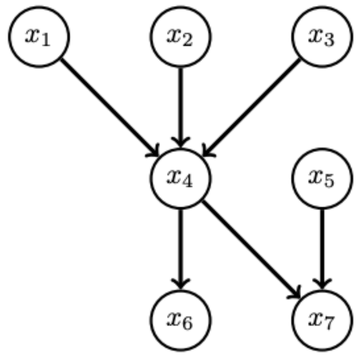
where ψ_c are nonnegative functions, one for each maximal clique c.

(b) We can generate the independencies by conditioning on progressively larger sets. Since there is a trail between any two nodes, there are no unconditional independencies. If we condition on a single variable, there is still a trail that connects the remaining ones. Let us thus consider the case where we condition on two nodes. By graph separation, we have

$$w \perp\!\!\!\perp y \mid x, z \qquad x \perp\!\!\!\perp z \mid w, y$$

These are all the independencies that hold for the model, since conditioning on three nodes does not lead to any independencies in a model with four variables.

Exercise 6 **[Moralisation]** For the DAG below, call it G, find the minimal undirected I-map for $I(G)$.



Answer: To derive an undirected minimal I-map from a directed one, we have to construct the moralised graph where the "unmarried" parents are connected by a covering edge. This is because each conditional $p(x_i|\text{pa}(i))$ corresponds to a factor $\phi_i(x_i, \text{pa}(i))$ and we need to connect all variables that are arguments of the same factor with edges.

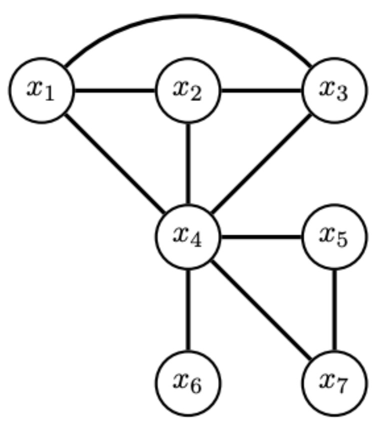
Statistically, the reason for marrying the parents is as follows: An independence property

$$x \perp\!\!\!\perp y \mid \{\text{child, other nodes}\}$$

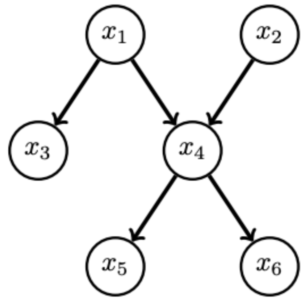
does not hold in the directed graph in case of collider connections but would hold in the undirected graph if we didn't marry the parents. Hence links between the parents must be added.

It is important to add edges between all parents of a node. Here, $p(x_4|x_1, x_2, x_3)$ corresponds to a factor $\phi(x_4, x_1, x_2, x_3)$ so that all four variables need to be connected. Just adding edges $x_1 - x_2$ and $x_2 - x_3$ would not be enough.

The moral graph, which is the requested minimal undirected I-map, is shown below.



Exercise 7 [Conversion to factor graphs for polytrees] Consider the directed graph shown below. (This kind of graph is called a polytree: there are no loops but a node may have more than one parent.)

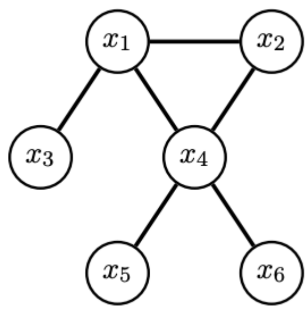


(a) Draw the moralised graph for this directed graph.

(b) Draw an undirected factor graph for this directed graph.

Answers:

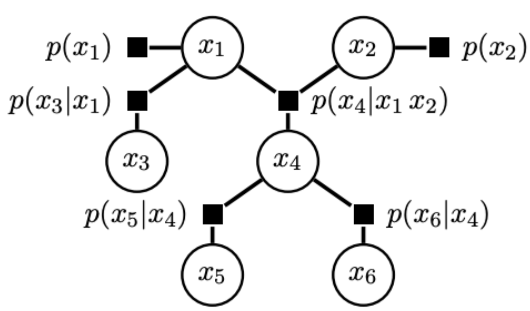
(a) The moral graph is obtained by connecting the parents of the collider node x_4. See the graph in the figure below.



(b) For the factor graph, we note that the directed graph defines the following class of probabilistic models

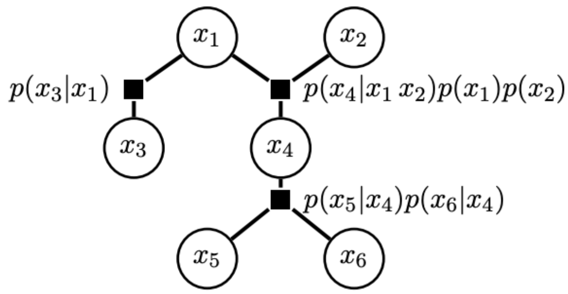
$$p(x_1, \ldots, x_6) = p(x_1)p(x_2)p(x_3|x_1)p(x_4|x_1, x_2)p(x_5|x_4)p(x_6|x_4)$$

This gives the factor graph in the figure below.

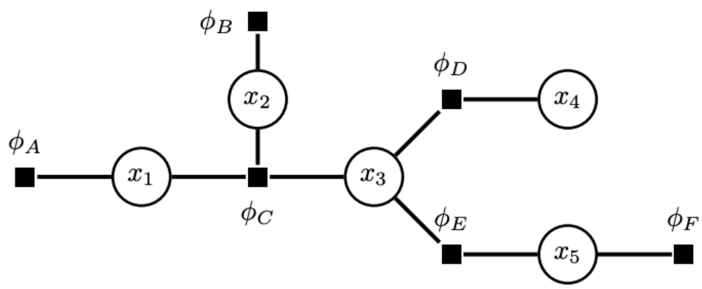


Comments:

- The moral graph contains a loop while the factor graph does not. The factor graph is still a polytree. This can be exploited for inference.
- One may choose to group some factors together in order to obtain a factor graph with a particular structure (see factor graph below)



Exercise 8 [Message passing] For this exercise, consider the following factor tree:



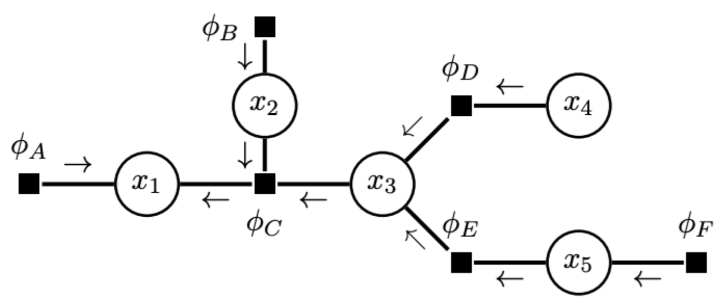
(a) Mark the graph with arrows indicating all messages that need to be computed for the computation of $p(x_1)$ via sum-product message passing.

(b) Mark the graph with arrows indicating all messages that need to be computed for the computation of $\arg\max_{x_1,\ldots,x_5} p(x_1, \ldots, x_5)$ via max-sum message passing with x_1 as root.

(c) Mark the graph with arrows indicating all messages that need to be computed for the computation of $\arg\max_{x_1,\ldots,x_5} p(x_1, \ldots, x_5)$ via max-sum message passing with x_3 as root.

Answers:

(a) The graph with marked arrows is:



(b) Same as the graph marked with arrows wshown in part (a).

(c) Similar to the graph with arrows in part (a), but with two arrows reverted: $x_1 \to \phi_C$ and $\phi_C \to x_3$.