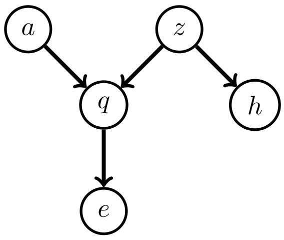
Exercise 1. ***Directed graph concepts***

We here consider the directed graph below that was partly discussed in the lecture.



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

Solution. We have

$$(a, q, e) \qquad (a, q, z, h) \qquad (h, z, q, e)$$

and the corresponding ones with swapped start and end nodes.

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

Solution. $(a, q, e) \qquad (z, q, e) \qquad (z, h)$

(c) *What are the descendants of z?*

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

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

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

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

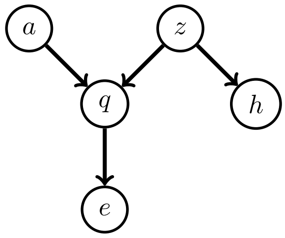
- *(a,z,h,q,e)*
- *(a,z,e,h,q)*
- *(z,a,q,h,e)*
- *(z,q,e,a,h)*

Solution.

- (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***

We continue with the investigation of the graph from Exercise 1 shown below for reference.



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

Solution. We proceed as in the lecture slides: The predecessor sets are

$$\mathrm{pre}_z = \varnothing, \mathrm{pre}_h = \{z\}, \mathrm{pre}_a = \{z, h\}, \mathrm{pre}_q = \{z, h, a\}, \mathrm{pre}_e = \{z, h, a, q\}$$

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

$$\mathrm{pa}_z = \varnothing, \mathrm{pa}_h = \{z\}, \mathrm{pa}_a = \varnothing, \mathrm{pa}_q = \{a, z\}, \mathrm{pa}_e = \{q\},$$

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. $x_1 = z, x_2 = h, x_3 = a, etc.$
With

$$\mathrm{pre}_h \setminus \mathrm{pa}_h = \varnothing \quad \mathrm{pre}_a \setminus \mathrm{pa}_a = \{z, h\} \quad \mathrm{pre}_q \setminus \mathrm{pa}_q = \{h\} \quad \mathrm{pre}_e \setminus \mathrm{pa}_e = \{z, h, a\}$$

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. Compared to the relations obtained for the orderings in the lecture, the new one here is $a \perp\!\!\!\perp \{z, h\}$. Generally, having a variable later in the topological ordering allows one to possibly obtain a stronger independence relation because the set pre $\setminus$ pa can only increase when the predecessor set pre becomes larger.

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

Solution. The non-descendants are

$$\mathrm{nondesc}(a) = \{z, h\} \quad \mathrm{nondesc}(z) = \{a\} \quad \mathrm{nondesc}(h) = \{a, z, q, e\}$$

$$\mathrm{nondesc}(q) = \{a, z, h\} \quad \mathrm{nondesc}(e) = \{a, q, z, h\}$$

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

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

Solution. A basic definition of conditional statistical independence $x_1 \perp\!\!\!\perp x_2 \mid x_3$ is that the (conditional) joint $p(x_1, x_2 \mid x_3)$ equals the product of the (conditional) marginals $p(x_1 \mid x_3)$ and $p(x_2 \mid x_3)$. In other words, for discrete random variables,

$$x_1 \perp\!\!\!\perp x_2 \mid x_3 \quad \Longleftrightarrow \quad p(x_1, x_2 \mid x_3) = \left(\sum_{x_2} p(x_1, x_2 \mid x_3)\right)\left(\sum_{x_1} p(x_1, x_2 \mid x_3)\right) \tag{S.1}$$

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

$$p(q, h|a, z) = \left(\sum_{h} p(q, h|a, z)\right)\left(\sum_{q} p(q, h|a, z)\right) \tag{S.2}$$

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

$$\begin{aligned}
p(a, z, q, h) &= \sum_e p(a, z, q, h, e) &\text{(S.3)}\\
&= \sum_e p(a)p(z)p(q|a, z)p(h|z)p(e|q) &\text{(S.4)}\\
&= p(a)p(z)p(q|a, z)p(h|z)\underbrace{\sum_e p(e|q)}_{1} &\text{(S.5)}\\
&= p(a)p(z)p(q|a, z)p(h|z), &\text{(S.6)}
\end{aligned}$$

where $\sum_e p(e|q) = 1$ because (conditional) pdfs/pmfs are normalised so that the integrate/sum to one. We further have

$$\begin{aligned}
p(a, z) &= \sum_{q,h} p(a, z, q, h) &\text{(S.7)}\\
&= \sum_{q,h} p(a)p(z)p(q|a, z)p(h|z) &\text{(S.8)}\\
&= p(a)p(z)\sum_q p(q|a, z)\sum_h p(h|z) &\text{(S.9)}\\
&= p(a)p(z) &\text{(S.10)}
\end{aligned}$$

so that

$$\begin{aligned}
p(q, h|a, z) &= \frac{p(a, z, q, h)}{p(a, z)} &\text{(S.11)}\\
&= \frac{p(a)p(z)p(q|a, z)p(h|z)}{p(a)p(z)} &\text{(S.12)}\\
&= p(q|a, z)p(h|z). &\text{(S.13)}
\end{aligned}$$

We further see that $p(q|a,z)$ and $p(h|z)$ are the marginals of $p(q,h|a,z)$, i.e.

$$p(q|a,z) = \sum_h p(q,h|a,z) \tag{S.14}$$

$$p(h|z) = \sum_q p(q,h|a,z). \tag{S.15}$$

This means that

$$p(q,h|a,z) = \left(\sum_h p(q,h|a,z)\right)\left(\sum_q p(q,h|a,z)\right), \tag{S.16}$$

which shows that $q \perp\!\!\!\perp h|a,z$.

We see that using the graph to determine the independency is easier than manipulating the pmf/pdf.

(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?*

Solution. 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. ***Chest clinic (based on Barber's exercise 3.3)***

The directed graphical model in Figure 1 is the "Asia" example of Lauritzen and Spiegelhalter (1988). It concerns the diagnosis of lung disease (T=tuberculosis or L=lung cancer). In this model, a visit to some place in A=Asia is thought to increase the probability of tuberculosis.

(a) *Explain which of the following independence relationships hold for all distributions that factorise over the graph.*

1. $t \perp\!\!\!\perp s \mid d$

Solution.

- There are two trails from t to s: (t,e,l,s) and (t,e,d,b,s).
- The trail (t,e,l,s) features a collider node e that is opened by the conditioning variable d. The trail is thus active and we do not need to check the second trail because for independence all trails needed to be blocked.

- The independence relationship does thus generally not hold.

2. $l \perp\!\!\!\perp b \mid s$

Solution.

- There are two trails from l to b: (l, s, b) and (l, e, d, b)
- The trail (l, s, b) is blocked by s (s is in a tail-tail configuration and part of the conditioning set)
- The trail (l, e, d, b) is blocked by the collider configuration for node d.
- All trails are blocked so that the independence relation holds.

(b) Can we simplify $p(l|b, s)$ to $p(l|s)$?

Solution. Since $l \perp\!\!\!\perp b \mid s$, we have $p(l|b, s) = p(l|s)$.

Exercise 4. ***Independencies***

We have seen that $x \perp\!\!\!\perp y|z$ is characterised by $p(x, y|z) = p(x|z)p(y|z)$ or, equivalently, by $p(x|y, z) = p(x|z)$. Show that further equivalent characterisations are

$$p(x, y, z) = p(x|z)p(y|z)p(z) \quad \textit{and} \tag{1}$$

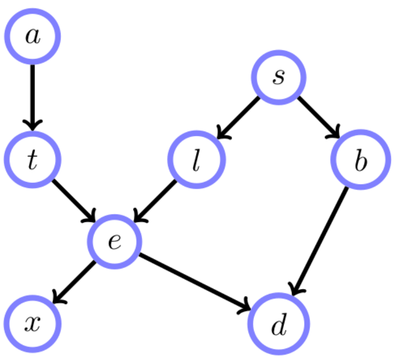
$$p(x, y, z) = a(x, z)b(y, z) \quad \textit{for some non-neg. functions } a(x, z) \textit{ and } b(x, z). \tag{2}$$

The characterisation in Equation (2) will be important for undirected graphical models.

Solution. We first show the equivalence of $p(x, y|z) = p(x|z)p(y|z)$ and $p(x, y, z) = p(x|z)p(y|z)p(z)$: By the product rule, we have

$$p(x, y, z) = p(x, y|z)p(z).$$

If $p(x, y|z) = p(x|z)p(y|z)$, it follows that $p(x, y, z) = p(x|z)p(y|z)p(z)$. To show the opposite direction assume that $p(x, y, z) = p(x|z)p(y|z)p(z)$ holds. By comparison with the decomposition in the product rule, it follows that we must have $p(x, y|z) = p(x|z)p(y|z)$ whenever $p(z) > 0$ (it suffices to consider this case because for z where $p(z) = 0$, $p(x, y|z)$ may not be uniquely defined in the first place).



x = Positive X-ray
d = Dyspnea (Shortness of breath)
e = Either Tuberculosis or Lung Cancer
t = Tuberculosis
l = Lung Cancer
b = Bronchitis
a = Visited Asia
s = Smoker

Figure 1: Graphical model for Exercise 3 (Barber Figure 3.15).

Equation (1) implies (2) with $a(x, z) = p(x|z)$ and $b(y, z) = p(y|z)p(z)$. We now show the inverse. Let us assume that $p(x, y, z) = a(x, z)b(y, z)$. By the product rule, we have

$$p(x, y|z)p(z) = a(x, z)b(y, z). \tag{S.17}$$

$$\tag{S.18}$$

Summing over y gives

$$\sum_y p(x, y|z)p(z) = p(z) \sum_y p(x, y|z) \tag{S.19}$$

$$= p(z)p(x|z) \tag{S.20}$$

Moreover

$$\sum_y p(x, y|z)p(z) = \sum_y a(x, z)b(y, z) \tag{S.21}$$

$$= a(x, z) \sum_y b(y, z) \tag{S.22}$$

so that

$$a(x, z) = \frac{p(z)p(x|z)}{\sum_y b(y, z)} \tag{S.23}$$

Since the sum of $p(x|z)$ over x equals one we have

$$\sum_x a(x, z) = \frac{p(z)}{\sum_y b(y, z)}. \tag{S.24}$$

Now, summing $p(x, y|z)p(z)$ over x yields

$$\sum_x p(x, y|z)p(z) = p(z) \sum_x p(x, y|z). \tag{S.25}$$

$$= p(y|z)p(z) \tag{S.26}$$

We also have

$$\sum_x p(x, y|z)p(z) = \sum_x a(x, z)b(y, z) \tag{S.27}$$

$$= b(y, z) \sum_x a(x, z) \tag{S.28}$$

$$\overset{\text{(S.24)}}{=} b(y, z) \frac{p(z)}{\sum_y b(y, z)} \tag{S.29}$$

so that

$$p(y|z)p(z) = p(z) \frac{b(y, z)}{\sum_y b(y, z)} \tag{S.30}$$

We thus have

$$p(x, y, z) = a(x, z)b(y, z) \tag{S.31}$$

$$\overset{\text{(S.23)}}{=} \frac{p(z)p(x|z)}{\sum_y b(y, z)} b(y, z) \tag{S.32}$$

$$= p(x|z)p(z) \frac{b(y, z)}{\sum_y b(y, z)} \tag{S.33}$$

$$\overset{\text{(S.30)}}{=} p(x|z)p(y|z)p(z) \tag{S.34}$$

which is Equation (1).