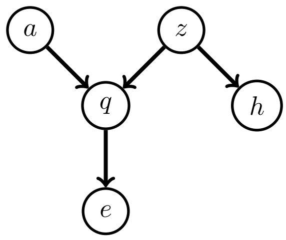
Exercise 1. ***Directed graph concepts***

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

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

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

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

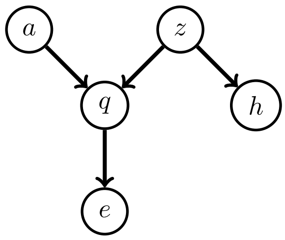
- *(a,z,h,q,e)*
- *(a,z,e,h,q)*
- *(z,a,q,h,e)*
- *(z,q,e,a,h)*

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

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

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

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

The ordered Markov property reads $x_i \perp\!\!\!\perp \text{pre}_i \setminus \text{pa}_i \mid \text{pa}_i$ where the x_i refer to the ordered variables, e.g. $x_1 = z, x_2 = h, x_3 = a, etc.$
With

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

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

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

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

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

Verify them by d-separation

Solution. For $a \perp\!\!\!\perp \{z, h\}$: 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.

For $q \perp\!\!\!\perp h \mid \{a, z\}$: 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. 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 first compute the joint distribution of (a, z, q, h) (use integrals in case of continuous random variables)

$$p(a, z, q, h) = \sum_e p(a)p(z)p(q|a, z)p(h|z)p(e|q) \tag{S.1}$$

$$= p(a)p(z)p(q|a, z)p(h|z) \sum_e p(e|q) \tag{S.2}$$

$$= p(a)p(z)p(q|a, z)p(h|z) \tag{S.3}$$

We further have

$$p(a, z) = \sum_{q,h} p(a)p(z)p(q|a, z)p(h|z) \tag{S.4}$$

$$= p(a)p(z) \sum_q p(q|a, z) \sum_h p(h|z) \tag{S.5}$$

$$= p(a)p(z) \tag{S.6}$$

so that

$$p(q, h|a, z) = \frac{p(a, z, q, h)}{p(a, z)} \tag{S.7}$$

$$= p(q|a, z)p(h|z) \tag{S.8}$$

Furthermore, $p(q|a, z)$ and $p(h|z)$ are the marginals of $p(q, h|a, z)$, i.e.

$$p(q|a, z) = \sum_h p(q, h|a, z) \tag{S.9}$$

$$p(h|z) = \sum_q p(q, h|a, z) \tag{S.10}$$

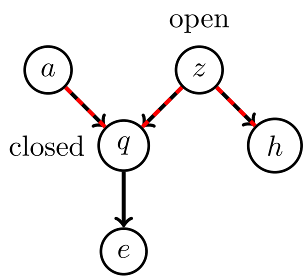
so that $p(q, h|a, z) = p(q|a, z)p(h|a, z)$, which shows that $q \perp\!\!\!\perp h|a, z$. We see that using the graph to determine the independency is easier than manipulating the pmf/pdf.

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

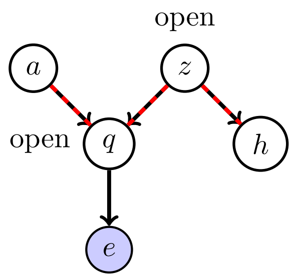
Solution. The independencies that follow from the ordered or local Markov property require conditioning on parent sets. However, e is not a parent of any node so that the above independence assertion cannot be checked via the ordered or local Markov property.

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

Solution. 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 $a \not\perp\!\!\!\perp h \mid e$ for some distributions that factorise over the graph.

(g) *Determine the Markov blanket of z.*

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

(h) *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

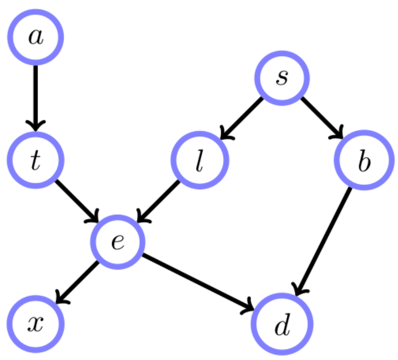
$$p(a, z, q, h, e) = p(a)p(z)p(q|a, z)p(h|z)p(e|q)$$

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.



$x =$ 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).

(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.

3. $a \perp\!\!\!\perp s \mid l$

   Solution.
   - There are two trails from a to s: (a, t, e, l, s) and (a, t, e, d, b, s)
   - The trail (a, t, e, l, s) features a collider node e that blocks the trail (the trail is also blocked by l).
   - The trail (a, t, e, d, b, s) is blocked by the collider node d.
   - All trails are blocked so that the independence relation holds.

4. $a \perp\!\!\!\perp s \mid l, d$

Solution.

- There are two trails from a to s: (a, t, e, l, s) and (a, t, e, d, b, s)
- The trail (a, t, e, l, s) features a collider node e that is opened by the conditioning variable d but the l node is closed by the conditioning variable l: the trail is blocked
- The trail (a, t, e, d, b, s) features a collider node d that is opened by conditioning on d. On this trail, e is not in a head-head (collider) configuration) so that all nodes are open and the trail active.
- Hence, the independence relation does generally not hold.

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

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

(c) *Let g be a (deterministic) function of x and t. Is the expected value $\mathrm{E}[g(x, t) \mid l, b]$ equal to $\mathrm{E}[g(x, t) \mid l]$?*

Solution. The question boils down to checking whether $x, t \perp\!\!\!\perp b \mid l$. For the independence relation to hold, all trails from both x and t to b need to be blocked by l.

- For x, we have the trails (x, e, l, s, b) and (x, e, d, b)
- Trail (x, e, l, s, b) is blocked by l
- Trail (x, e, d, b) is blocked by the collider configuration of node d.
- For t, we have the trails (t, e, l, s, b) and (t, e, d, b)
- Trail (t, e, l, s, b) is blocked by l.
- Trail (t, e, d, b) is blocked by the collider configuration of node d.

As all trails are blocked we have $x, t \perp\!\!\!\perp b \mid l$ and $\mathrm{E}[g(x, t) \mid l, b] = \mathrm{E}[g(x, t) \mid l]$.

Exercise 4. *Independencies*

This exercise is on further properties and characterisations of statistical independence.

(a) *Without using d-separation, show that $x \perp\!\!\!\perp \{y, w\} \mid z$ implies that $x \perp\!\!\!\perp y \mid z$ and $x \perp\!\!\!\perp w \mid z$.*

Solution. We consider the joint distribution $p(x, y, w|z)$. By assumption

$$p(x, y, w|z) = p(x|z)p(y, w|z) \tag{S.11}$$

We have to show that $x \perp\!\!\!\perp y|z$ and $x \perp\!\!\!\perp w|z$. For simplicity, we assume that the variables are discrete valued. If not, replace the sum below with an integral.

To show that $x \perp\!\!\!\perp y|z$, we marginalise $p(x, y, w|z)$ over w to obtain

$$\begin{aligned} p(x, y|z) &= \sum_w p(x, y, w|z) && \text{(S.12)} \\ &= \sum_w p(x|z)p(y, w|z) && \text{(S.13)} \\ &= p(x|z) \sum_w p(y, w|z) && \text{(S.14)} \end{aligned}$$

Since $\sum_w p(y, w|z)$ is the marginal $p(y|z)$, we have

$$p(x, y|z) = p(x|z)p(y|z), \tag{S.15}$$

which means that $x \perp\!\!\!\perp y|z$.

To show that $x \perp\!\!\!\perp w|z$, we similarly marginalise $p(x, y, w|z)$ over y to obtain $p(x, w|z) = p(x|z)p(w|z)$, which means that $x \perp\!\!\!\perp w|z$.

(b) *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).

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

$$p(x, y|z)p(z) = a(x, z)b(y, z). \tag{S.16}$$
$$\tag{S.17}$$

Summing over y gives

$$\sum_y p(x, y|z)p(z) = p(z) \sum_y p(x, y|z) \tag{S.18}$$
$$= p(z)p(x|z) \tag{S.19}$$

Moreover

$$\sum_y p(x, y|z)p(z) = \sum_y a(x, z)b(y, z) \tag{S.20}$$
$$= a(x, z) \sum_y b(y, z) \tag{S.21}$$

so that

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

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

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

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

$$\sum_x p(x, y|z)p(z) = p(z) \sum_x p(x, y|z). \tag{S.24}$$
$$= p(y|z)p(z) \tag{S.25}$$

We also have

$$\sum_x p(x, y|z)p(z) = \sum_x a(x, z)b(y, z) \tag{S.26}$$
$$= b(y, z) \sum_x a(x, z) \tag{S.27}$$
$$\overset{(S.23)}{=} b(y, z) \frac{p(z)}{\sum_y b(y, z)} \tag{S.28}$$

so that

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

We thus have

$$p(x, y, z) = a(x, z)b(y, z) \tag{S.30}$$
$$\overset{(S.22)}{=} \frac{p(z)p(x|z)}{\sum_y b(y, z)} b(y, z) \tag{S.31}$$
$$= p(x|z)p(z) \frac{b(y, z)}{\sum_y b(y, z)} \tag{S.32}$$
$$\overset{(S.29)}{=} p(x|z)p(y|z)p(z) \tag{S.33}$$

which is Equation (1).

(c) *For the directed graphical model below, show that $x \perp\!\!\!\perp y$ and generally $x \not\perp\!\!\!\perp y \mid w$ without using d-separation.*

$$\begin{array}{ccccc} x & \rightarrow & z & \leftarrow & y \\ & & \downarrow & & \\ & & w & & \end{array}$$

The exercise shows that not only conditioning on a collider but also on a descendent of it activates the trail between x and y.

Solution. The graphical model corresponds to the factorisation

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

For the marginal $p(x, y)$ we have to sum (integrate) over all (z, w)

$$p(x, y) = \sum_{z,w} p(x, y, z, w) \tag{S.34}$$

$$= \sum_{z,w} p(x)p(y)p(z|x, y)p(w|z) \tag{S.35}$$

$$= p(x)p(y) \sum_{z,w} p(z|x, y)p(w|z) \tag{S.36}$$

$$= p(x)p(y) \underbrace{\sum_{z} p(z|x, y)}_{1} \underbrace{\sum_{w} p(w|z)}_{1} \tag{S.37}$$

$$= p(x)p(y) \tag{S.38}$$

Since $p(x, y) = p(x)p(y)$ we have $x \perp\!\!\!\perp y$.

For $x \not\perp\!\!\!\perp y|w$, compute $p(x, y, w)$ and use the result from Question (b), namely $x \perp\!\!\!\perp y|w \Leftrightarrow p(x, y, w) = a(x, w)b(y, w)$.

$$p(x, y, w) = \sum_{z} p(x, y, z, w) \tag{S.39}$$

$$= \sum_{z} p(x)p(y)p(z|x, y)p(w|z) \tag{S.40}$$

$$= p(x) \underbrace{p(y) \sum_{z} p(z|x, y)p(w|z)}_{k(x,y,w)} \tag{S.41}$$

Since $p(x, y, w)$ cannot be factorised as $a(x, w)b(y, w)$, the relation $x \perp\!\!\!\perp y|w$ cannot generally hold.