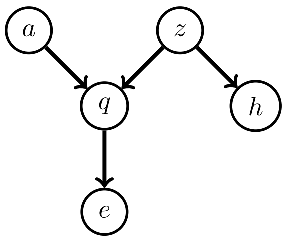
Exercises for the tutorials: 1, 2(a-b), 3.

The other exercises are for self-study and exam preparation. All material is examinable unless otherwise mentioned.

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?*

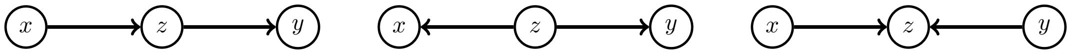
- *(a,z,h,q,e)*
- *(a,z,e,h,q)*
- *(z,a,q,h,e)*
- *(z,q,e,a,h)*

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. *Canonical connections*

We here derive the independencies that hold in the three canonical connections that exist in DAGs, shown in Figure 1.



(a) Serial connection (b) Diverging connection (c) Converging connection

Figure 1: The three canonical connections in DAGs.

(a) *For the serial connection, use the ordered Markov property to show that $x \perp\!\!\!\perp y \mid z$.*

Solution. The only topological ordering is x, z, y. The predecessors of y are $\text{pre}_y = \{x, z\}$ and its parents $\text{pa}_y = \{z\}$. The ordered Markov property

$$y \perp\!\!\!\perp (\text{pre}_y \setminus \text{pa}_y) \mid \text{pa}_y \tag{S.1}$$

thus becomes $y \perp\!\!\!\perp (\{x, z\} \setminus z) \mid z$. Hence we have

$$y \perp\!\!\!\perp x \mid z, \tag{S.2}$$

which is the same as $x \perp\!\!\!\perp y \mid z$ since the indepedency relationship is symmetric.

This means that if the state or value of z is known (i.e. if the random variable z is "instantiated"), evidence about x will not change our belief about y, and vice versa. We say that the z node is "closed" and that the trail between x and y is "blocked" by the instantiated z. In other words, knowing the value of z blocks the flow of evidence *between* x and y.

(b) *For the serial connection, show that the marginal $p(x, y)$ does generally not factorise into $p(x)p(y)$, i.e. that $x \perp\!\!\!\perp y$ does not hold.*

Solution. There are several ways to show the result. One is to present an example where the independency does not hold. Consider for instance the following model

$$x \sim \mathcal{N}(x; 0, 1) \tag{S.3}$$

$$z = x + n_z \tag{S.4}$$

$$y = z + n_y \tag{S.5}$$

where $n_z \sim \mathcal{N}(n_z; 0, 1)$ and $n_y \sim \mathcal{N}(n_y; 0, 1)$, both being statistically independent from x. Here $\mathcal{N}(\cdot; 0, 1)$ denotes the Gaussian pdf with mean 0 and variance 1, and $x \sim \mathcal{N}(x; 0, 1)$ means that we sample x from the distribution $\mathcal{N}(x; 0, 1)$. Hence $p(z|x) = \mathcal{N}(z; x, 1)$, $p(y|z) = \mathcal{N}(y; z, 1)$ and $p(x, y, z) = p(x)p(z|x)p(y|z) = \mathcal{N}(x; 0, 1)\mathcal{N}(z; x, 1)\mathcal{N}(y; z, 1)$.

Whilst we could manipulate the pdfs to show the result, it's here easier to work with the generative model in Equations (S.3) to (S.5). Eliminating z from the equations, by plugging the definition of z into (S.5) we have

$$y = x + n_z + n_y, \tag{S.6}$$

which describes the marginal distribution of (x, y). We see that $\mathbb{E}[xy]$ is

$$\mathbb{E}[xy] = \mathbb{E}[x^2 + xn_z + xn_y] \tag{S.7}$$
$$= \mathbb{E}[x^2] + \mathbb{E}[x]\mathbb{E}[n_z] + \mathbb{E}[x]\mathbb{E}[n_y] \tag{S.8}$$
$$= 1 + 0 + 0 \tag{S.9}$$

where we have use the linearity of expectation, that x is independent from n_z and n_y, and that x has zero mean. If x and y were independent (or only uncorrelated), we had $\mathbb{E}[xy] = \mathbb{E}[x]\mathbb{E}[y] = 0$. However, since $\mathbb{E}[xy] \neq \mathbb{E}[x]\mathbb{E}[y]$, x and y are not independent.

In plain English, this means that if the state of z is unknown, then evidence or information about x will influence our belief about y, and the other way around. Evidence can flow through z between x and y. We say that the z node is "open" and the trail between x and y is "active".

(c) *For the diverging connection, use the ordered Markov property to show that $x \perp\!\!\!\perp y \mid z$.*

Solution. A topological ordering is z, x, y. The predecessors of y are $\mathrm{pre}_y = \{x, z\}$ and its parents $\mathrm{pa}_y = \{z\}$. The ordered Markov property

$$y \perp\!\!\!\perp (\mathrm{pre}_y \setminus \mathrm{pa}_y) \mid \mathrm{pa}_y \tag{S.10}$$

thus becomes again

$$y \perp\!\!\!\perp x \mid z, \tag{S.11}$$

which is, since the independence relationship is symmetric, the same as $x \perp\!\!\!\perp y \mid z$.

As in the serial connection, if the state or value z is known, evidence about x will not change our belief about y, and vice versa. Knowing z closes the z node, which blocks the trail between x and y.

(d) *For the diverging connection, show that the marginal $p(x, y)$ does generaly not factorise into $p(x)p(y)$, i.e. that $x \perp\!\!\!\perp y$ does not hold.*

Solution. As for the serial connection, it suffices to give an example where $x \perp\!\!\!\perp y$ does not hold. We consider the following generative model

$$z \sim \mathcal{N}(z; 0, 1) \tag{S.12}$$
$$x = z + n_x \tag{S.13}$$
$$y = z + n_y \tag{S.14}$$

where $n_x \sim \mathcal{N}(n_x; 0, 1)$ and $n_y \sim \mathcal{N}(n_y; 0, 1)$, and they are independent of each other and the other variables. We have $\mathbb{E}[x] = \mathbb{E}[z + n_x] = \mathbb{E}[z] + \mathbb{E}[n_x] = 0$. On the other hand

$$\mathbb{E}[xy] = \mathbb{E}[(z + n_x)(z + n_y)] \tag{S.15}$$
$$= \mathbb{E}[z^2 + z(n_x + n_y) + n_x n_y] \tag{S.16}$$
$$= \mathbb{E}[z^2] + \mathbb{E}[z(n_x + n_y)] + \mathbb{E}[n_x n_y] \tag{S.17}$$
$$= 1 + 0 + 0 \tag{S.18}$$

Hence, $\mathbb{E}[xy] \neq \mathbb{E}[x]\mathbb{E}[y]$ and we do not have that $x \perp\!\!\!\perp y$ holds.

In a diverging connection, as in the serial connection, if the state of z is unknown, then evidence or information about x will influence our belief about y, and the other way around. Evidence can flow through z between x and y. We say that the z node is open and the trail between x and y is active.

(e) *For the converging connection, show that $x \perp\!\!\!\perp y$.*

Solution. We can here again use the ordered Markov property with the ordering y, x, z. Since $\text{pre}_x = \{y\}$ and $\text{pa}_x = \varnothing$, we have

$$x \perp\!\!\!\perp (\text{pre}_x \setminus \text{pa}_x) \mid \text{pa}_x = x \perp\!\!\!\perp y. \tag{S.19}$$

Alternatively, we can use the basic definition of directed graphical models, i.e.

$$p(x, y, z) = k(x)k(y)k(z \mid x, y) \tag{S.20}$$

together with the result that the kernels (factors) are valid (conditional) pdfs/pmfs and equal to the conditionals/marginals with respect to the joint distribution $p(x, y, z)$, i.e.

$$k(x) = p(x) \tag{S.21}$$
$$k(y) = p(y) \tag{S.22}$$
$$k(z|x, y) = p(z|x, y) \quad \text{(not needed in the proof below)} \tag{S.23}$$

Integrating out z gives

$$\begin{aligned} p(x, y) &= \int p(x, y, z)\mathrm{d}z && \text{(S.24)} \\ &= \int k(x)k(y)k(z \mid x, y)\mathrm{d}z && \text{(S.25)} \\ &= k(x)k(y) \underbrace{\int k(z \mid x, y)\mathrm{d}z}_{1} && \text{(S.26)} \\ &= p(x)p(y) && \text{(S.27)} \end{aligned}$$

Hence $p(x, y)$ factorises into its marginals, which means that $x \perp\!\!\!\perp y$.

Hence, when we do not have evidence about z, evidence about x will not change our belief about y, and vice versa. For the converging connection, if no evidence about z is available, the z node is closed, which blocks the trail between x and y.

(f) For the converging connection, show that $x \perp\!\!\!\perp y \mid z$ does generally not hold.

Solution. We give a simple example where $x \perp\!\!\!\perp y \mid z$ does not hold.

Consider three independent Bernoulli random variables x, y, u with success probability 0.5 each. That is, $p(x = 1) = p(x = 0) = 0.5$, and analogous for y and u. Assume further that we deterministically generate z from (x, y, u) as

$$z = xy + u. \tag{S.28}$$

Note that z can take on the values 0, 1, or 2. The following table summarises the possible outcomes and probabilities:

x	y	u	$p(x, y, u)$	z
0	0	0	1/8	0
0	1	0	1/8	0
1	0	0	1/8	0
1	1	0	1/8	1
0	0	1	1/8	1
0	1	1	1/8	1
1	0	1	1/8	1
1	1	1	1/8	2

While z is deterministic given (x, y, u), it is a random variable if we only know (x, y). In other words, we consider the transformation of random variables (x, y, u) to (x, y, z). The table above allows us to read out the probabily mass function of (x, y, z).

x	y	z	$p(x, y, z)$
0	0	0	1/8
0	1	0	1/8
1	0	0	1/8
1	1	0	0
0	0	1	1/8
0	1	1	1/8
1	0	1	1/8
1	1	1	1/8
0	0	2	0
0	1	2	0
1	0	2	0
1	1	2	1/8

If the independency $x \perp\!\!\!\perp y \mid z$ would hold, we would have that $p(x, y|z) = p(x|z)p(y|z)$ for all values of x, y, z (where $p(z) > 0$). Let us consider the case where $z = 0$. We have $p(z = 0) = 3/8$ and the conditional pmf $p(x, y|z = 0) = p(x, y, z = 0)/p(z = 0)$ is given in the following table:

x	y	$p(x, y\|z = 0)$
0	0	1/3
0	1	1/3
1	0	1/3
1	1	0

Using the sum rule, we have $p(x = 1|z = 0) = p(x = 1, y = 0|z = 0) + p(x = 1, y = 1|z = 0) = 1/3$ and similarly $p(y = 1|z = 0) = 1/3$. Hence

$$p(x = 1|z = 0)p(y = 1|z = 0) = \frac{1}{9} \tag{S.29}$$

but $p(x = 1, y = 1|z = 0) = 0$. Since we have found one configuration of the random variables where $p(x, y|z) \neq p(x|z)p(y|z)$, the independency cannot hold.

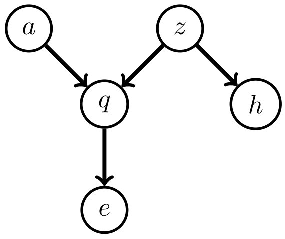
The intuition here is that knowing $z = 0$ imposes constraints on what configurations of x, y are possible—$x = y = 1$ is not possible any more—which means that x and y are not independent given z.

More generally, for converging connections, if evidence or information about z is available, evidence about x will influence the belief about y, and vice versa. We say that information about z opens the z-node, and evidence can flow between x and y.

Note: information about z means that z *or one of its descendents* is observed, see exercise 9.

Exercise 3. *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.

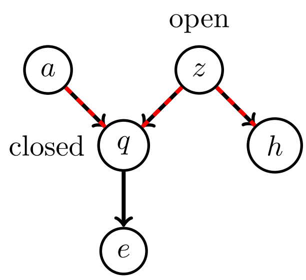
$$a \perp\!\!\!\perp \{z, h\} \qquad z \perp\!\!\!\perp a \qquad h \perp\!\!\!\perp \{a, q, e\} \mid z \qquad q \perp\!\!\!\perp h \mid \{a, z\} \qquad e \perp\!\!\!\perp \{a, z, h\} \mid q$$

(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.*

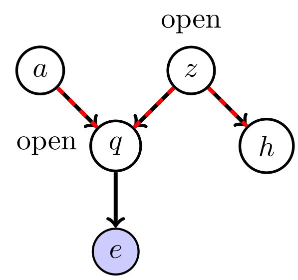
Solution. The only trail from q to h goes through z which is in a tail-tail configuration. Since z is part of the conditioning set, the trail is blocked and the result follows.

(d) *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.

(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

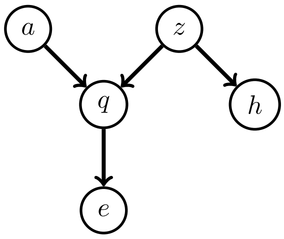
$$p(a, z, q, h, e) = p(a)p(z)p(q|a, z)p(h|z)p(e|q)$$

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 4. ***More on ordered and local Markov properties, d-separation***

We continue with the investigation of the graph below



(a) *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.

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

Solution. 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) *Determine the Markov blanket of z.*

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

(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,

$$x_1 \perp\!\!\!\perp x_2 \mid x_3 \quad \Longleftrightarrow \quad p(x_1, x_2 \mid x_3) = \left(\sum_{x_2} p(x_1, x_2 \mid x_3)\right)\left(\sum_{x_1} p(x_1, x_2 \mid x_3)\right) \tag{S.30}$$

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

$$p(q, h|a, z) = \left(\sum_{h} p(q, h|a, z)\right)\left(\sum_{q} p(q, h|a, z)\right) \tag{S.31}$$

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

$$p(a, z, q, h) = \sum_e p(a, z, q, h, e) \tag{S.32}$$
$$= \sum_e p(a)p(z)p(q|a, z)p(h|z)p(e|q) \tag{S.33}$$
$$= p(a)p(z)p(q|a, z)p(h|z) \underbrace{\sum_e p(e|q)}_{1} \tag{S.34}$$
$$= p(a)p(z)p(q|a, z)p(h|z), \tag{S.35}$$

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

$$p(a, z) = \sum_{q,h} p(a, z, q, h) \tag{S.36}$$
$$= \sum_{q,h} p(a)p(z)p(q|a, z)p(h|z) \tag{S.37}$$
$$= p(a)p(z) \sum_q p(q|a, z) \sum_h p(h|z) \tag{S.38}$$
$$= p(a)p(z) \tag{S.39}$$

so that

$$p(q, h|a, z) = \frac{p(a, z, q, h)}{p(a, z)} \tag{S.40}$$
$$= \frac{p(a)p(z)p(q|a, z)p(h|z)}{p(a)p(z)} \tag{S.41}$$
$$= p(q|a, z)p(h|z). \tag{S.42}$$

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

$$p(q|a, z) = \sum_h p(q, h|a, z) \tag{S.43}$$
$$p(h|z) = \sum_q p(q, h|a, z). \tag{S.44}$$

This means that

$$p(q, h|a, z) = \left(\sum_h p(q, h|a, z)\right)\left(\sum_q p(q, h|a, z)\right), \tag{S.45}$$

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.

Exercise 5. ***Chest clinic (based on Barber's exercise 3.3)***

The directed graphical model in Figure 2 is about the diagnosis of lung disease (t=tuberculosis or l=lung cancer). In this model, a visit to some place "a" is thought to increase the probability of tuberculosis.

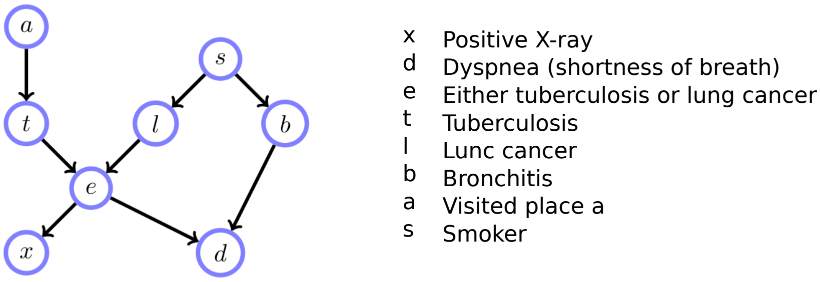


Figure 2: Graphical model for Exercise 5 (Barber Figure 3.15).

(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.

(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 6. *More on the chest clinic (based on Barber's exercise 3.3)*

Consider the directed graphical model in Figure 2.

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

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

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.

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

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

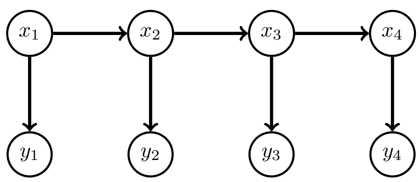
Solution. The question boils down to checking whether $x, t \perp\!\!\!\perp b \mid l$. For the independence relation to hold, all trails from both x and t to b need to be blocked by l.

- 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 $\mathbb{E}[g(x, t) \mid l, b] = \mathbb{E}[g(x, t) \mid l]$.

Exercise 7. *Hidden Markov models*

This exercise is about directed graphical models that are specified by the following DAG:



These models are called "hidden" Markov models because we typically assume to only observe the y_i and not the x_i that follow a Markov model.

(a) Show that all probabilistic models specified by the DAG factorise as

$$p(x_1, y_1, x_2, y_2, \ldots, x_4, y_4) = p(x_1)p(y_1|x_1)p(x_2|x_1)p(y_2|x_2)p(x_3|x_2)p(y_3|x_3)p(x_4|x_3)p(y_4|x_4)$$

Solution. From the definition of directed graphical models it follows that

$$p(x_1, y_1, x_2, y_2, \dots, x_4, y_4) = \prod_{i=1}^{4} p(x_i|\text{pa}(x_i)) \prod_{i=1}^{4} p(y_i|\text{pa}(y_i)).$$

The result is then obtained by noting that the parent of y_i is given by x_i for all i, and that the parent of x_i is x_{i-1} for $i = 2, 3, 4$ and that x_1 does not have a parent ($\text{pa}(x_1) = \varnothing$).

(b) *Derive the independencies implied by the ordered Markov property with the topological ordering* $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$

Solution.

$$y_i \perp\!\!\!\perp x_1, y_1, \dots, x_{i-1}, y_{i-1} \mid x_i \qquad x_i \perp\!\!\!\perp x_1, y_1, \dots, x_{i-2}, y_{i-2}, y_{i-1} \mid x_{i-1}$$

(c) *Derive the independencies implied by the ordered Markov property with the topological ordering* $(x_1, x_2, \dots, x_4, y_1, \dots, y_4)$.

Solution. For the x_i, we use that for $i \geq 2$: $\text{pre}(x_i) = \{x_1, \dots, x_{i-1}\}$ and $\text{pa}(x_i) = x_{i-1}$. For the y_i, we use that $\text{pre}(y_1) = \{x_1, \dots, x_4\}$, that $\text{pre}(y_i) = \{x_1, \dots, x_4, y_1, \dots, y_{i-1}\}$ for $i > 1$, and that $\text{pa}(y_i) = x_i$. The ordered Markov property then gives:

$$\begin{aligned} x_3 &\perp\!\!\!\perp x_1 \mid x_2 & x_4 &\perp\!\!\!\perp \{x_1, x_2\} \mid x_3 \\ y_1 &\perp\!\!\!\perp \{x_2, x_3, x_4\} \mid x_1 & y_2 &\perp\!\!\!\perp \{x_1, x_3, x_4, y_1\} \mid x_2 \\ y_3 &\perp\!\!\!\perp \{x_1, x_2, x_4, y_1, y_2\} \mid x_3 & y_4 &\perp\!\!\!\perp \{x_1, x_2, x_3, y_1, y_2, y_3\} \mid x_4 \end{aligned}$$

(d) *Does* $y_4 \perp\!\!\!\perp y_1 \mid y_3$ *hold?*

Solution. The trail $y_1 - x_1 - x_2 - x_3 - x_4 - y_4$ is active: none of the nodes is in a collider configuration, so that their default state is open and conditioning on y_3 does not block any of the nodes on the trail.

While $x_1 - x_2 - x_3 - x_4$ forms a Markov chain, where e.g. $x_4 \perp\!\!\!\perp x_1 \mid x_3$ holds, this not so for the distribution of the y's.

Exercise 8. *Alternative characterisation of 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).

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

$$p(x,y|z)p(z) = a(x,z)b(y,z). \tag{S.46}$$
$$\tag{S.47}$$

Summing over y gives

$$\sum_y p(x,y|z)p(z) = p(z)\sum_y p(x,y|z) \tag{S.48}$$
$$= p(z)p(x|z) \tag{S.49}$$

Moreover

$$\sum_y p(x,y|z)p(z) = \sum_y a(x,z)b(y,z) \tag{S.50}$$
$$= a(x,z)\sum_y b(y,z) \tag{S.51}$$

so that

$$a(x,z) = \frac{p(z)p(x|z)}{\sum_y b(y,z)} \tag{S.52}$$

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

$$\sum_x a(x,z) = \frac{p(z)}{\sum_y b(y,z)}. \tag{S.53}$$

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

$$\sum_x p(x,y|z)p(z) = p(z)\sum_x p(x,y|z). \tag{S.54}$$
$$= p(y|z)p(z) \tag{S.55}$$

We also have

$$\sum_x p(x,y|z)p(z) = \sum_x a(x,z)b(y,z) \tag{S.56}$$
$$= b(y,z)\sum_x a(x,z) \tag{S.57}$$
$$\overset{(S.53)}{=} b(y,z)\frac{p(z)}{\sum_y b(y,z)} \tag{S.58}$$

so that

$$p(y|z)p(z) = p(z)\frac{b(y,z)}{\sum_y b(y,z)} \tag{S.59}$$

We thus have

$$\begin{aligned}
p(x,y,z) &= a(x,z)b(y,z) && \text{(S.60)}\\
&\overset{\text{(S.52)}}{=} \frac{p(z)p(x|z)}{\sum_y b(y,z)} b(y,z) && \text{(S.61)}\\
&= p(x|z)p(z)\frac{b(y,z)}{\sum_y b(y,z)} && \text{(S.62)}\\
&\overset{\text{(S.59)}}{=} p(x|z)p(y|z)p(z) && \text{(S.63)}
\end{aligned}$$

which is Equation (1).

Exercise 9. *More on 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 $x \perp\!\!\!\perp w \mid z$.*
Hint: use the definition of statistical independence in terms of the factorisation of pmfs/pdfs.

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

$$p(x,y,w|z) = p(x|z)p(y,w|z) \tag{S.64}$$

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

$$\begin{aligned}
p(x,y|z) &= \sum_w p(x,y,w|z) && \text{(S.65)}\\
&= \sum_w p(x|z)p(y,w|z) && \text{(S.66)}\\
&= p(x|z)\sum_w p(y,w|z) && \text{(S.67)}
\end{aligned}$$

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

$$p(x,y|z) = p(x|z)p(y|z), \tag{S.68}$$

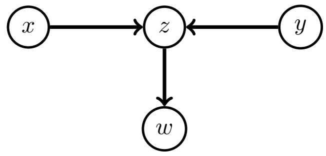
which means that $x \perp\!\!\!\perp y|z$.

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) *For the directed graphical model below, show that the following two statements hold without using d-separation:*

$$x \perp\!\!\!\perp y \quad \textit{and} \tag{3}$$

$$x \not\!\perp\!\!\!\perp y \mid w \tag{4}$$



The exercise shows that not only conditioning on a collider node but also on one of its descendents activates the trail between x and y. You can use the result that $x \perp\!\!\!\perp y|w \Leftrightarrow p(x,y,w) = a(x,w)b(y,w)$ for some non-negative functions $a(x,w)$ and $b(y,w)$.

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)

$$\begin{aligned}
p(x,y) &= \sum_{z,w} p(x,y,z,w) && \text{(S.69)}\\
&= \sum_{z,w} p(x)p(y)p(z|x,y)p(w|z) && \text{(S.70)}\\
&= p(x)p(y)\sum_{z,w} p(z|x,y)p(w|z) && \text{(S.71)}\\
&= p(x)p(y)\underbrace{\sum_{z} p(z|x,y)}_{1}\underbrace{\sum_{w} p(w|z)}_{1} && \text{(S.72)}\\
&= p(x)p(y) && \text{(S.73)}
\end{aligned}$$

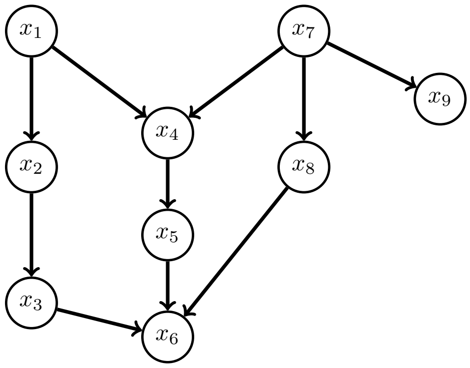
Since $p(x,y) = p(x)p(y)$ we have $x \perp\!\!\!\perp y$.
For $x \not\perp\!\!\!\perp y|w$, compute $p(x,y,w)$ and use the result $x \perp\!\!\!\perp y|w \Leftrightarrow p(x,y,w) = a(x,w)b(y,w)$.

$$\begin{aligned}
p(x,y,w) &= \sum_{z} p(x,y,z,w) && \text{(S.74)}\\
&= \sum_{z} p(x)p(y)p(z|x,y)p(w|z) && \text{(S.75)}\\
&= p(x)\underbrace{p(y)\sum_{z} p(z|x,y)p(w|z)}_{k(x,y,w)} && \text{(S.76)}
\end{aligned}$$

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.

Exercise 10. *Independencies in directed graphical models*

Consider the following directed acyclic graph.



For each of the statements below, determine whether it holds for all probabilistic models that factorise over the graph. Provide a justification for your answer.

(a) $p(x_7|x_2) = p(x_7)$

Solution. Yes, it holds. x_2 is a non-descendant of x_7, $\mathrm{pa}(x_7) = \varnothing$, and hence, by the local Markov property, $x_7 \perp\!\!\!\perp x_2$, so that $p(x_7|x_2) = p(x_7)$.

(b) $x_1 \not\perp\!\!\!\perp x_3$

Solution. No, does not hold. x_1 and x_3 are d-connected, which only implies independence for *some* and not all distributions that factorise over the graph. The graph generally only allows us to read out independencies and not dependencies.

(c) $p(x_1, x_2, x_4) \propto \phi_1(x_1, x_2)\phi_2(x_1, x_4)$ *for some non-negative functions* ϕ_1 *and* ϕ_2.

Solution. Yes, it holds. The statement is equivalent to $x_2 \perp\!\!\!\perp x_4 \mid x_1$. There are three trails from x_2 to x_4, which are all blocked:

1. $x_2 - x_1 - x_4$: this trail is blocked because x_1 is in a tail-tail connection and it is observed, which closes the node.
2. $x_2 - x_3 - x_6 - x_5 - x_4$: this trail is blocked because x_3, x_6, x_5 is in a collider configuration, and x_6 is not observed (and it does not have any descendants).
3. $x_2 - x_3 - x_6 - x_8 - x_7 - x_4$: this trail is blocked because x_3, x_6, x_8 is in a collider configuration, and x_6 is not observed (and it does not have any descendants).

Hence, by the global Markov property (d-separation), the independency holds.

(d) $x_2 \perp\!\!\!\perp x_9 \mid \{x_6, x_8\}$

Solution. No, does not hold. Conditioning on x_6 opens the collider node x_4 on the trail $x_2 - x_1 - x_4 - x_7 - x_9$, so that the trail is active.

(e) $x_8 \perp\!\!\!\perp \{x_2, x_9\} \mid \{x_3, x_5, x_6, x_7\}$

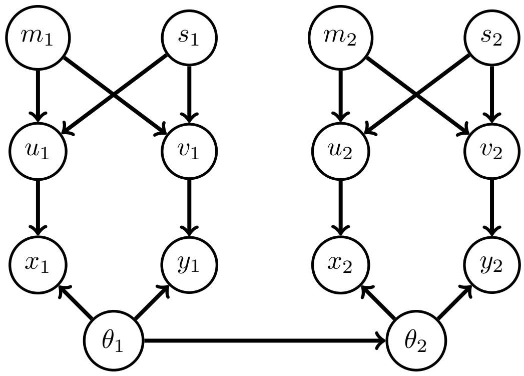
Solution. Yes, it holds. $\{x_3, x_5, x_6, x_7\}$ is the Markov blanket of x_8, so that x_8 is independent of remaining nodes given the Markov blanket.

(f) $\mathbb{E}[x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_8 \mid x_7] = 0$ *if* $\mathbb{E}[x_8 \mid x_7] = 0$

Solution. Yes, it holds. $\{x_2, x_3, x_4, x_5\}$ are non-descendants of x_8, and x_7 is the parent of x_8, so that $x_8 \perp\!\!\!\perp \{x_2, x_3, x_4, x_5\} \mid x_7$. This means that $\mathbb{E}[x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_8 \mid x_7] = \mathbb{E}[x_2 \cdot x_3 \cdot x_4 \cdot x_5 \mid x_7]\mathbb{E}[x_8 \mid x_7] = 0$.

Exercise 11. ***Independencies in directed graphical models***

Consider the following directed acyclic graph:



For each of the statements below, determine whether it holds for all probabilistic models that factorise over the graph. Provide a justification for your answer.

(a) $x_1 \perp\!\!\!\perp x_2$

Solution. Does not hold. The trail $x_1 - \theta_1 - \theta_2 - x_2$ is active (unblocked) because none of the nodes is in a collider configuration or in the conditioning set.

(b) $p(x_1, y_1, \theta_1, u_1) \propto \phi_A(x_1, \theta_1, u_1)\phi_B(y_1, \theta_1, u_1)$ *for some non-negative functions* ϕ_A *and* ϕ_B

Solution. Holds. The statement is equivalent to $x_1 \perp\!\!\!\perp y_1 \mid \{\theta_1, u_1\}$. The conditioning set $\{\theta_1, u_1\}$ blocks all trails from x_1 to y_1 because they are both only in serial configurations in all trails from x_1 to y_1, hence the independency holds by the global Markov property. Alternative justification: the conditioning set is the Markov blanket of x_1, and x_1 and y_1 are not neighbours which implies the independency.

(c) $v_2 \perp\!\!\!\perp \{u_1, v_1, u_2, x_2\} \mid \{m_2, s_2, y_2, \theta_2\}$

Solution. Holds. The conditioning set is the Markov blanket of v_2 (the set of parents, children, and co-parents): the set of parents is $\text{pa}(v_2) = \{m_2, s_2\}$, y_2 is the only child of v_2, and θ_2 is the only other parent of y_2. And v_2 is independent of all other variables given its Markov blanket.

(d) $\mathbb{E}[m_2 \mid m_1] = \mathbb{E}[m_2]$

Solution. Holds. There are four trails from m_1 to m_2, namely via x_1, via y_1, via x_2, via y_2. In all trails the four variables are in a collider configuration, so that each of the trails is blocked. By the global Markov property (d-separation), this means that $m_1 \perp\!\!\!\perp m_2$ which implies that $\mathbb{E}[m_2 \mid m_1] = \mathbb{E}[m_2]$.

Alternative justification 1: m_2 is a non-descendent of m_1 and $\text{pa}(m_2) = \varnothing$. By the directed local Markov property, a variable is independent from its non-descendents given the parents, hence $m_2 \perp\!\!\!\perp m_1$.

Alternative justification 2: We can choose a topological ordering where m_1 and m_2 are the first two variables. Moreover, their parent sets are both empty. By the directed ordered Markov, we thus have $m_1 \perp\!\!\!\perp m_2$.