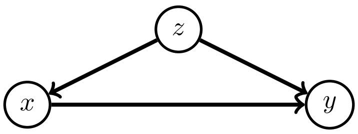
THE UNIVERSITY of EDINBURGH informatics

Probabilistic Modelling and Reasoning
Tutorial Solutions 5

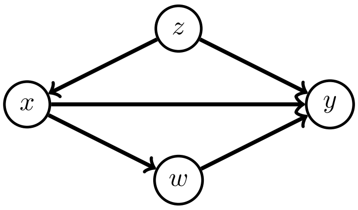
Autumn 2025
Michael Gutmann
Chris Williams

Exercise 1. *Adjustment for direct causes*

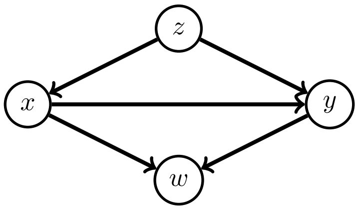
We would like to compute $p(y; do(x) = a)$ for the models represented by the following three causal DAGs. Assume that all variables are discrete.



(a)



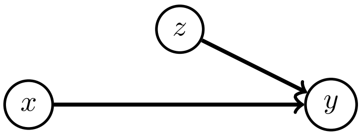
(b)



(c)

(a) *For the causal DAG in (a), specify the factorization of $p(x, y, z; do(x) = a)$ and derive $p(y; do(x) = a)$ in terms of the conditional probability distributions $p(x_i|\mathrm{pa}_i)$ of the graphical model defined the DAG.*

Solution. Intervening on x deletes all incoming edges into node x, so in this case, this is the edge $z \to x$, which gives the graph



Thus we have

$$p(x, y, z; do(x) = a) = p(z)p(y|x, z)\delta(x - a) \tag{S.1}$$

with

$$\delta(x - a) = \begin{cases} 1 & \text{if } x = a \\ 0 & \text{otherwise} \end{cases} \tag{S.2}$$

To obtain $p(y; do(x) = a)$ we marginalise out x and z, which gives

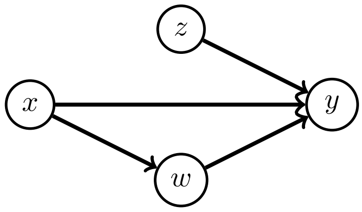
$$p(y; do(x) = a) = \sum_{x,z} p(z)p(y|x, z)\delta(x - a) \tag{S.3}$$

$$= \sum_{z} p(z) \sum_{x} p(y|x, z)\delta(x - a) \tag{S.4}$$

$$= \sum_{z} p(z)p(y|x = a, z) \tag{S.5}$$

(b) *For the causal DAG in (b), specify the factorization of $p(x, y, z, w; do(x) = a)$ and derive $p(y; do(x) = a)$ in terms of the conditional probability distributions $p(x_i|\mathrm{pa}_i)$ of the graphical model defined by the DAG.*

Solution. As in (a) we delete all incoming edges into x to model the intervention. This gives



We read out the factorisation, using that the interventional distribution is $p'(x) = \delta(x-a)$,

$$p(x, y, z, w; do(x) = a) = p(z)\delta(x - a)p(w|x)p(y|x, w, z) \tag{S.6}$$

We next sum out x, z, w to obtain the desired result:

$$\begin{aligned} p(y; do(x) = a) &= \sum_{x,z,w} p(z)\delta(x - a)p(w|x)p(y|x, w, z) &\text{(S.7)}\\ &= \sum_{z,w} p(z)p(w|x = a)p(y|x = a, w, z) &\text{(S.8)}\\ &= \sum_{z} p(z) \sum_{w} p(w|x = a)p(y|x = a, w, z) &\text{(S.9)} \end{aligned}$$

The same result could have been obtained from the general formula

$$p(y; do(x) = a) = \mathbb{E}_{p(\mathrm{pa}_x)} [p(y|x = a, \mathrm{pa}_x)] \tag{S.10}$$

where pa_x denotes the parents of x, which is here z. The distribution $p(y|x = a, z)$ can be obtained as

$$\begin{aligned} p(y|x = a, z) &= \sum_{w} p(y, w|x = a, z) &\text{(S.11)}\\ &= \sum_{w} p(y|x = a, w, z)p(w|x = a, z) &\text{(S.12)} \end{aligned}$$

From the original graph (before intervening on x), we see that $w \perp\!\!\!\perp z|x$, hence $p(w|x = a, z) = p(w|x = a)$ so that

$$p(y|x = a, z) = \sum_{w} p(y|x = a, w, z)p(w|x = a) \tag{S.13}$$

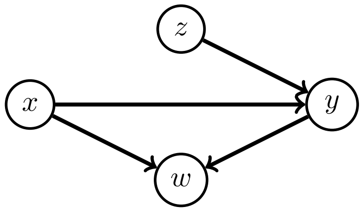
and

$$\begin{aligned} p(y; do(x) = a) &= \mathbb{E}_{p(\mathrm{pa}_x)} [p(y|x = a, \mathrm{pa}_x)] &\text{(S.14)}\\ &= \sum_{z} p(z) \sum_{w} p(y|x = a, w, z)p(w|x = a) &\text{(S.15)} \end{aligned}$$

as in (S.9)

(c) *Answer the same questions for the causal DAG in (c).*

Solution. The graph modelling the intervention on x is given by



(c)

We read out the factorisation, using that the interventional distribution is $p'(x) = \delta(x-a)$,

$$p(x, y, z, w; do(x) = a) = p(z)\delta(x - a)p(w|x, y)p(y|x, z) \tag{S.16}$$

We next sum out x, z, w to obtain the desired result:

$$\begin{aligned} p(y; do(x) = a) &= \sum_{x,z,w} p(z)\delta(x - a)p(w|x, y)p(y|x, z) && \text{(S.17)} \\ &= \sum_{z,w} p(z)p(w|x = a, y)p(y|x = a, z) && \text{(S.18)} \\ &= \sum_{z} p(z)p(y|x = a, z) \sum_{w} p(w|x = a, y) && \text{(S.19)} \\ &= \sum_{z} p(z)p(y|x = a, z) && \text{(S.20)} \end{aligned}$$

For the interventional distribution, we can also start from

$$p(y; do(x) = a) = \mathbb{E}_{p(\mathrm{pa}_x)}\left[p(y|x = a, \mathrm{pa}_x)\right] \tag{S.21}$$

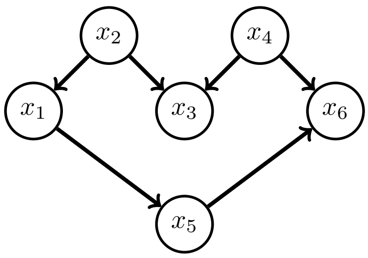
with $\mathrm{pa}_x = z$. Since the distribution $p(y|x = a, z)$ equals $p(y|\mathrm{pa}_y)$ (with x set to a) we thus have

$$p(y; do(x) = a) = \sum_{z} p(z)p(y|x = a, x), \tag{S.22}$$

as before.

Exercise 2. *Intervening and conditioning*

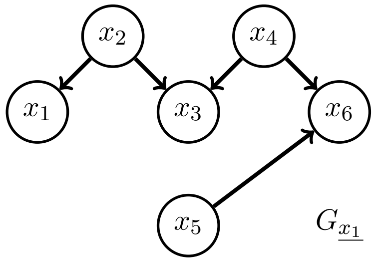
Consider the following graph G:



For each question below, provide an answer and a brief justification.

(a) *Do we have $p(x_6; do(x_1)) = p(x_6|x_1)$?*

Solution. We need to check whether there are open backdoor paths from x_1 to x_6. For that we first form the graph $G_{\underline{x_1}}$ obtained by removing all outgoing edges from x_1:



We next check whether $x_6 \perp\!\!\!\perp x_1$ holds for $G_{\underline{x_1}}$. If so, all backdoor path are closed in G. Given that x_3 is in a collider configuration and we do not condition on it, or any of its descendants, the independency holds. We can thus exchange observation and action, and $p(x_6; do(x_1)) = p(x_6|x_1)$ holds.

(b) *Do we have $p(x_6|x_3; do(x_1)) = p(x_6|x_1, x_3)$?*

Solution. Here, we need to check whether $x_6 \perp\!\!\!\perp x_1|x_3$ holds for $G_{\underline{x_1}}$. Since we condition on the collider node x_3, the independency does not hold. Hence, there is an open backdoor path and $p(x_6|x_3; do(x_1)) \neq p(x_6|x_1, x_3)$.

(c) *Do we have $p(x_6|x_5, x_2; do(x_1)) = p(x_6|x_5)$?*

Solution. We first determine whether $p(x_6|x_5, x_2; do(x_1)) = p(x_6|x_5, x_2, x_1)$ holds. For that we check whether $x_6 \perp\!\!\!\perp x_1|x_5, x_2$ for $G_{\underline{x_1}}$. As in (a), the collider node x_3 blocks the path from x_1 to x_6 and conditioning on x_5 or x_2 does not change that.

Next, we check whether $x_6 \perp\!\!\!\perp \{x_2, x_1\}|x_5$ holds in G. We have two paths from x_6 to $\{x_2, x_1\}$. The first one goes via x_3 which is blocked (collider node and we do not condition on it or its descendants). The second goes via x_5, which is blocked because x_5 is in a head-tail configuration and we condition on it. Hence, we have that $p(x_6|x_5, x_2, x_1) = p(x_6|x_5)$ and $p(x_6|x_5, x_2; do(x_1)) = p(x_6|x_5)$ holds.

Exercise 3. *Reject option*

[Murphy PML1 (2022) Ex 5.1] Consider a K-class discrete variable Y with labels $\mathcal{Y} = \{1, \ldots, C\}$. The actions are $\mathcal{A} = \mathcal{Y} \cup \{0\}$, where $a = 0$ denotes the reject option, and choosing action $a = i$ for $i \in \mathcal{Y}$ denotes selecting label i. Define the loss function as follows:

$$\ell(y = j, a = i) = \begin{cases} 0 & \text{if } i = j \text{ and } a \in \{1, \ldots, C\} \\ \lambda_r & \text{if } a = 0 \\ \lambda_c & \text{otherwise,} \end{cases} \tag{1}$$

where λ_r is the cost of a reject, λ_c the cost of an error.

Given information $\mathbf{x}$ we obtain the posterior $p(Y|\mathbf{x})$. Show that the minimum posterior risk is obtained for the following decision rule: we decide $Y = j$ if $p(Y = j|\mathbf{x}) \geq p(Y = k|\mathbf{x})$ for all k (i.e. j is the most probable label) and *if $p(Y = j|\mathbf{x}) \geq 1 - \lambda_r/\lambda_c$, otherwise we decide to reject.*

Solution. The posterior risk for the action $a = i$ for $i = 1, \ldots, C$ is given by

$$\begin{aligned} R(a = i|\mathbf{x}) &= \sum_k p(Y = k|\mathbf{x})\ell(Y = k, a = i) \\ &= p(Y = i|\mathbf{x}) \cdot 0 + \lambda_c \sum_{k \neq i} p(Y = k|\mathbf{x}) \\ &= \lambda_c(1 - p(Y = i|\mathbf{x})). \end{aligned}$$

This risk is minimized by choosing the most probable label j. On the other hand, the risk for the rejection option ($a = 0$) is $R(a = 0|\mathbf{x}) = \lambda_r$. We choose not to reject if

$$\begin{aligned} R(a = 0|\mathbf{x}) &\geq R(a = j|\mathbf{x}), \text{i.e.} \\ \lambda_r &\geq \lambda_c(1 - p(Y = j|\mathbf{x})), \end{aligned}$$

which can be rearranged to give $p(Y = j|\mathbf{x}) \geq 1 - \lambda_r/\lambda_c$. Hence, the minimal risk is obtained for the following decision rule: decide $Y = j$ if $p(Y = j|\mathbf{x}) \geq p(Y = k|\mathbf{x})$ for all k (i.e. j is the most probable label) *and* if $p(Y = j|\mathbf{x}) \geq 1 - \lambda_r/\lambda_c$, otherwise decide to reject.

