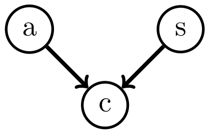
Exercise 1. *Cancer-asbestos-smoking example: MLE*

Consider the model specified by the DAG



The distribution of a and s are Bernoulli distributions with parameter (success probability) θ_a and θ_s, respectively, i.e.

$$p(a; \theta_a) = \theta_a^a (1 - \theta_a)^{1-a} \qquad p(s; \theta_s) = \theta_s^s (1 - \theta_s)^{1-s}, \tag{1}$$

and the distribution of c given the parents is parameterised as specified in the following table

$p(c = 1 \mid a, s; \theta_c^1, \ldots, \theta_c^4)$	a	s
θ_c^1	0	0
θ_c^2	1	0
θ_c^3	0	1
θ_c^4	1	1

The free parameters of the model are $(\theta_a, \theta_s, \theta_c^1, \ldots, \theta_c^4)$.

Assume we observe the following iid data (each row is a data point).

a	s	c
0	1	1
0	0	0
1	0	1
0	0	0
0	1	0

(a) Determine the maximum-likelihood estimates of θ_a and θ_s.

(b) Determine the maximum-likelihood estimates of $\theta_c^1, \ldots, \theta_c^4$.

Exercise 2. *Cancer-asbestos-smoking example: Bayesian inference*

We here perform Bayesian inference for the model from Question 1.

We assume that the prior over the parameters of the model, $(\theta_a, \theta_s, \theta_c^1, \ldots, \theta_c^4)$, factorises and is given by Beta distributions with hyperparameters $\alpha_0 = 1$ and $\beta_0 = 1$ (same for all parameters). The posterior then factorises too, with each parameter i following a Beta distribution with hyperparameters equal to

$$\alpha_{i,n}^k = \alpha_{i,0}^k + n_{x_i=1}^k, \qquad \beta_{i,n}^k = \beta_{i,0}^k + n_{x_i=0}^k. \tag{2}$$

Here x_i is the random variable associated with the parameter, e.g. a for θ_a or c for θ_c^k, and k enumerates the possible configurations of its parents. $n^k_{x_i=1}$ denotes the number of times variable x_i equals 1 when its parents are in configuration k, and $n^k_{x_i=0}$ is defined analogously.

(a) Determine the posteriors for θ_a and θ_s.

(b) Determine the posteriors for θ_c^k, $k = 1, \dots, 4$.

(c) Determine the posterior predictive probabilities $p(a = 1|\mathcal{D}), p(s = 1|\mathcal{D})$, and $p(c = 1|\text{pa}, \mathcal{D})$ for all possible parent configurations.

Exercise 3. ***Independent component analysis***

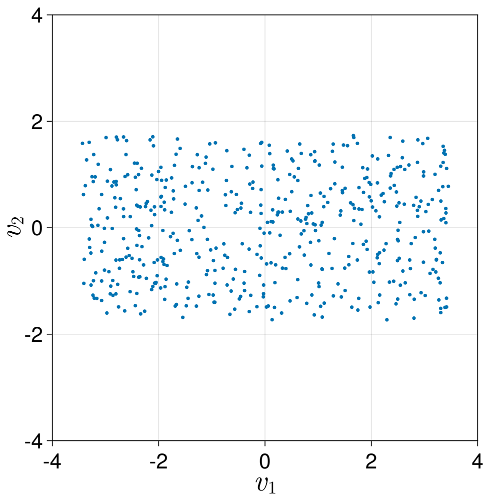
The two scatter plots below show two-dimensional data $\mathbf{v} = (v_1, v_2)^\top$ that were generated by sampling from the noise-free square ICA model

$$\mathbf{v} = \mathbf{A}\mathbf{h}, \tag{3}$$

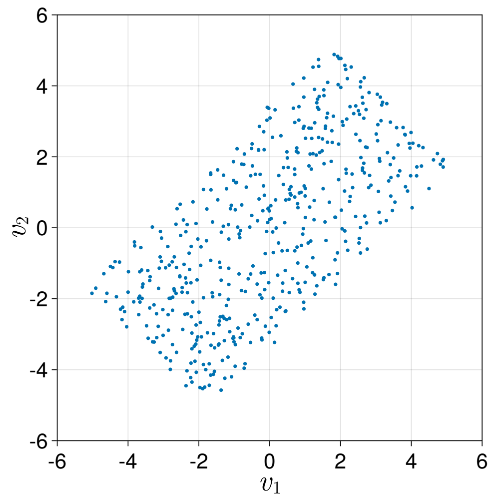
where $\mathbf{A}$ is a 2×2 matrix and $\mathbf{h} = (h_1, h_2)^\top$ contains the independent sources. Both h_1 and h_2 were sampled from a uniform distribution of mean zero and variance one.

For each scatter plot, select among the following 4 mixing matrices the one that has most likely generated the data. Justify your answer.

$$\mathbf{A}_1 = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix} \quad \mathbf{A}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \quad \mathbf{A}_3 = \begin{pmatrix} 2 & 2 \\ 1 & 2 \end{pmatrix} \quad \mathbf{A}_4 = \begin{pmatrix} -1 & 2 \\ 1 & 2 \end{pmatrix} \tag{4}$$



(a)



(b)