Exercises 1

Probabilistic modelling

Exercise 1.1 (adapted from [1])
Consider two boxes with white and black balls. Box 1 contains three black and five white balls and box 2 contains two black and five white balls. First a box is chosen at random with a prior probability $p(\texttt{box} = 1) = p(\texttt{box} = 2) = 0.5$, secondly a ball picked at random from that box. This ball turns out to be black. What is the posterior probability that this black ball came from box 1?

Exercise 1.2 (adapted from [1])
Two fair coins are tossed. If two head occurs two white balls are placed in a box, if one head and one tail occurs one white ball and one black ball are placed in a box, and if two tails occurs two black balls are placed in the box. Balls are then drawn from the box three times in succession, always with replacing the drawn ball back in the box. It is found that on all three occasions a black ball is drawn. What is the probability that both balls in the box are black?

Exercise 1.3 (adapted from [1])
The weather in Uppsala (based on data from 2018)[1] can be summarised as: if it rains or snows one day there is a 60% chance it will also rain or snow the following day; if it does not rain or snow one day there is an 80% chance it will not rain or snow the following day either.

a) Assuming that the prior probability it rained or snowed yesterday is 50%, what is the probability that it was raining or snowing yesterday given that it does not rain or snow today?

b) If the weather follows the same pattern as above, day after day, what is the probability that it will rain or snow on any day (based on an effectively infinite number of days of observing the weather)?

c) Use the result from part 2 above as a new prior probability of rain/snow yesterday and recompute the probability that it was raining/snowing yesterday given that it's does not rain or snow today.

Exercise 1.4 Bernoulli random variables
Consider a Bernoulli random variable with

$$X \sim \text{Bern}(x; \mu), \qquad \text{where} \qquad x \in \{0,1\}, \quad \text{Bern}(x;\mu) = \mu^x (1-\mu)^{1-x}.$$

(a) Show that the mean of X is μ.

(b) Show that the variance of X is $\mu(1-\mu)$.

Exercise 1.5 (adapted from [1])
For a variable $x \in \{0,1\}$, and $p(x = 1) = \mu$, show that in N independent draws $x_1, \ldots, x_N$ from this distribution, the number of 1 among these draws, denoted $m = \sum_{n=1}^{N} x_n$, is Binomial distributed (1)

$$\text{Bin}(m; N, \mu) = \binom{N}{m} \mu^m (1-\mu)^{N-m}, \qquad \text{where} \qquad m \in \{0, 1, \ldots, N\}, \quad \binom{N}{m} = \frac{N!}{(N-m)!m!}. \tag{1}$$

[1] `http://celsius.met.uu.se/rapport/`

Exercise 1.6 (adapted from [2])
Prove that the binomial distribution (1) is correctly normalized, i.e. that

$$\sum_{m=0}^{N} \mathrm{Bin}\,(m; N,\, \mu) = 1$$

See ex 2.3 in [2] for hints.

Exercise 1.7 (adapted from [2])
Prove that the Beta distribution

$$\mathrm{Beta}\,(\mu; a,\, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}, \qquad \text{where} \quad \mu \in [0,1], \quad \Gamma(a) = \int_0^{\infty} e^{-x}x^{a-1}dx \tag{2}$$

is correctly normalized, i.e. that

$$\int_0^1 \mathrm{Beta}\,(\mu; a,\, b)\, d\mu = 1 \quad \Leftrightarrow \quad \int_0^1 \mu^{a-1}(1-\mu)^{b-1}d\mu = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{3}$$

See ex 2.5 in [2] for hints.

Exercise 1.8 (adapted from [2])
Consider a Beta distributed random variable $\mu \in [0, 1]$ with

$$\mu \sim \mathrm{Beta}\,(\mu; a,\, b)\,.$$

Make use of (3) to show that

(a) ... the mean of μ is $\frac{a}{a+b}$. *Hint:* $\Gamma(a+1) = a\Gamma(a)$.

(b) ... the variance of μ is $\frac{ab}{(a+b)^2(a+b+1)}$.

Exercise 1.9 Bayesian vs maximum likelihood inference for a coin
Consider a variable $x \in \{0, 1\}$, and $p(x = 1) = \mu$ and N independent draws $x_1, \ldots, x_N$ from this distribution; that is, N flips of a (possibly unfair) coin. According to Exercise 1.5, the number of 1 among these draws, denoted $m = \sum_{n=1}^{N} x_n$, is Binomial distributed (1)

a) Show that the maximum likelihood solution of μ is the mean of $x_1, \ldots, x_N$, i.e. that

$$\hat{\mu} = \arg\max_{\mu} p(m|\mu) = \frac{1}{N}\sum_{n=1}^{N} x_n,$$

where $p(m|\mu) = \mathrm{Bin}\,(m; N,\, \mu)$.

b) Consider the prior $p(\mu) = \mathrm{Beta}\,(\mu; a,\, b)$ where the Beta distribution is defined in (2). Show that the posterior is $p(\mu|m) = \mathrm{Beta}\,(\mu; a^*,\, b^*)$ where

$$a^* = a + m,$$
$$b^* = b + N - m.$$

c) Consider $N = 3$ observations where all of them are 1, i.e. $x_1 = 1$, $x_2 = 1$, $x_3 = 1$. What is the probability of the next coin flip being heads up $x_4 = 1$ according to the solution in **(a)** and **(b)**, respectively. For the Bayesian solution, use $a = 1$ and $b = 1$ in the prior. Which of the two solutions provides the most realistic model? Motivate!

Exercise 1.10 Fair or unfair coin
Consider a variable $x \in \{0, 1\}$ and $p(x = 1) = \mu$ representing flipping of a coin. With 50% probability we think the coin is fair, i.e. that $\mu = 0.5$, and with 50% probability we think that is unfair, i.e. that $\mu \neq 0.5$. We encode this prior belief with the following prior

$$p(\mu) = \frac{1}{2}\mathrm{Beta}\,(\mu; 1,\, 1) + \frac{1}{2}\delta(\mu - 0.5).$$

a) Assume that we get one observation $x_1 = 1$. What is the posterior $p(\mu|x_1)$? In particular, how does the belief of the fairness of the coin change under this observation? *Hint:* $\Gamma(a) = (a-1)!$ *for positive integers.*

b) Assume that we get one additional observation $x_2 = 1$. What is the posterior $p(\mu|x_1, x_2)$? In particular, how does the belief of the fairness of the coin change under this observation?

c) Compute the probability of the coin being fair by instead defining an event `fair` where with the prior probability $p(\texttt{fair}) = 0.5$. Compute $p(\texttt{fair}|x_1, x_2)$ using Bayes' theorem, based on the observations $x_1 = 1$, $x_2 = 1$.

Solutions 1

Probabilistic modelling

Solution to Exercise 1.1
We know that

$$\begin{aligned}p(\texttt{box}=1)=p(\texttt{box}=2)&=1/2,\\ p(\texttt{black ball}|\texttt{box}=1)&=3/8,\\ p(\texttt{black ball}|\texttt{box}=2)&=2/7.\end{aligned}$$

and we want to know $p(\texttt{box}=1|\texttt{black ball})$. With Bayes' theorem we get

$$\begin{aligned}p(\texttt{box}=1|\texttt{black ball})&=\frac{p(\texttt{black ball}|\texttt{box}=1)p(\texttt{box}=1)}{p(\texttt{black ball})}\\ &=\frac{p(\texttt{black ball}|\texttt{box}=1)p(\texttt{box}=1)}{p(\texttt{black ball}|\texttt{box}=1)p(\texttt{box}=1)+p(\texttt{black ball}|\texttt{box}=2)p(\texttt{box}=2)}\\ &=\frac{(3/8)\cdot(1/2)}{(3/8)\cdot(1/2)+(2/7)\cdot(1/2)}=\frac{3/8}{3/8+2/7}=\frac{3\cdot 7}{3\cdot 7+2\cdot 8}=\frac{21}{37}.\end{aligned}$$

Solution to Exercise 1.2
Let

$$\begin{aligned}&b \quad \text{number of black balls in the box,}\\ &d \quad \text{number of black balls drawn from the box.}\end{aligned}$$

We want to find $p(b=2|d=3)$. Using Bayes' rule we get

$$p(b=2|d=3)=\frac{p(d=3|b=2)p(b=2)}{p(d=3)}. \tag{24}$$

There are in total three possibilities of what balls can be in the box.

$$\begin{aligned}p(b=0)&=\frac{1}{4},\\ p(b=1)&=\frac{1}{2},\\ p(b=2)&=\frac{1}{4}.\end{aligned}$$

We also know that

$$\begin{aligned} p(d=3|b=0) &= 0, \\ p(d=3|b=1) &= \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}, \\ p(d=3|b=2) &= 1. \end{aligned}$$

Hence, we can compute

$$\begin{aligned} p(d=3) &= p(d=3|b=0)p(b=0) + p(d=3|b=1)p(b=1) + p(d=3|b=2)p(b=2) \\ &= 0\cdot\frac{1}{4} + \frac{1}{8}\cdot\frac{1}{2} + 1\cdot\frac{1}{4} = \frac{5}{16} \end{aligned}$$

That together with (24) we get

$$p(b=2|d=3) = \frac{1\cdot\frac{1}{4}}{\frac{5}{16}} = \frac{4}{5}$$

Solution to Exercise 1.3
Let

$$\begin{aligned} w_t &\in \{1,0\} && \text{the weather today,} \\ w_{t-1} &\in \{1,0\} && \text{the weather yesterday.} \end{aligned}$$

where 1 represent rain/snow and 0 represent no rain/snow.
We know that

$$\begin{aligned} p(w_t=1|w_{t-1}=1) &= 0.6, \\ p(w_t=0|w_{t-1}=0) &= 0.8, \end{aligned}$$

and consequently

$$\begin{aligned} p(w_t=0|w_{t-1}=1) &= 0.4, \\ p(w_t=1|w_{t-1}=0) &= 0.2. \end{aligned}$$

a) We assume that $p(w_{t-1}=1) = p(w_{t-1}=0) = 0.5$. We then get

$$p(w_{t-1}=1|w_t=0) = \frac{p(w_t=0|w_{t-1}=1)p(w_{t-1}=1)}{p(w_t=0|w_{t-1}=1)p(w_{t-1}=1) + p(w_t=0|w_{t-1}=0)p(w_{t-1}=0)} = 1/3 \quad (25)$$

b) We have

$$p(w_t=1) = p(w_t=1|w_{t-1}=1)p(w_{t-1}=1) + p(w_t=1|w_{t-1}=0)p(w_{t-1}=0)$$

Define $r = p(w_t=1) = p(w_{t-1}=1)$ and consequently $1-r = p(w_t=0) = p(w_{t-1}=0)$ Then we get the equation

$$r = 0.6r + 0.2(1-r),$$

which has the solution $r = 1/3$.

c) Using $p(w_{t-1}=1) = 1/3$ and $p(w_{t-1}=0) = 2/3$ in (25) we get $p(w_{t-1}=1|w_t=0) = 1/5$.

Solution to Exercise 1.4 **a)** The mean is

$$\mathbb{E}[X] = \sum_{x=0}^{1} x\mathrm{Bern}\,(x;\mu) = 0\cdot(1-\mu) + 1\cdot\mu = \mu.$$

b) The variance is

$$\mathrm{Var}[X] = \sum_{x=0}^{1}(x - \mathbb{E}[X])^2 \text{Bern}\,(x; \mu) = (0-\mu)^2(1-\mu) + (1-\mu)^2\mu = (1-\mu)(\mu^2 + (1-\mu)\mu) = (1-\mu)\mu.$$

Solution to Exercise 1.6
See solution to exercise 2.3 in [2] which is available as a PDF from the book web site.

Solution to Exercise 1.7
See solution to exercise 2.5 in [2] which is available as a PDF from the book web site.

Solution to Exercise 1.8 a) The mean is

$$\begin{aligned}
\mathbb{E}[(\mu - \mathbb{E}[\mu])^2] &= \int_{x=0}^{1} \mu \text{Beta}\,(\mu; a,\, b)\, d\mu \\
&= \int_{x=0}^{1} \mu \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1} d\mu \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_{x=0}^{1} \mu^{a+1-1}(1-\mu)^{b-1} d\mu \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)} \\
&= \frac{\Gamma(a+b)}{\Gamma(a)} \frac{a\Gamma(a)}{(a+b)\Gamma(a+b)} \\
&= \frac{a}{a+b}.
\end{aligned}$$

b) The variance is

$$\begin{aligned}
\mathbb{E}[\mu] &= \int_{x=0}^{1} (\mu - \mathbb{E}[\mu])^2 \text{Beta}\,(\mu; a,\, b)\, d\mu \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_{x=0}^{1} \left(\mu^2 - 2\frac{a}{a+b}\mu + \frac{a^2}{(a+b)^2}\right) \mu^{a-1}(1-\mu)^{b-1} d\mu \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\int_{x=0}^{1} \mu^{a+2-1}(1-\mu)^{b-1} d\mu - 2\frac{a}{a+b}\int_{x=0}^{1} \mu^{a+1-1}(1-\mu)^{b-1} d\mu + \frac{a^2}{(a+b)^2}\int_{x=0}^{1} \mu^{a-1}(1-\mu)^{b-1} d\mu\right) \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+2+b)} - 2\frac{a}{a+b}\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)} + \frac{a^2}{(a+b)^2}\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\right) \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{a(a+1)\Gamma(a)\Gamma(b)}{(a+b)(a+b+1)\Gamma(a+b)} - 2\frac{a}{a+b}\frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)} + \frac{a^2}{(a+b)^2}\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\right) \\
&= \frac{a(a+1)}{(a+b)(a+b+1)} - 2\frac{a}{a+b}\frac{a}{a+b} + \frac{a^2}{(a+b)^2} \\
&= \frac{a(a+1)}{(a+b)(a+b+1)} - \frac{a^2}{(a+b)^2} \\
&= \frac{a(a+1)(a+b)}{(a+b)^2(a+b+1)} - \frac{a^2(a+b+1)}{(a+b)^2(a+b+1)} \\
&= \frac{a(a^2+ab+a+b-a^2-ab-a)}{(a+b)^2(a+b+1)} = \frac{ab}{(a+b)^2(a+b+1)}.
\end{aligned}$$

Solution to Exercise 1.9 a) Maximizing the likelihood gives

$$\begin{aligned}
\hat{\mu} &= \arg\max_{\mu} \binom{N}{m} \mu^m (1-\mu)^{N-m} \\
&= \arg\max_{\mu} \mu^m (1-\mu)^{N-m} \\
&= \arg\max_{\mu} \log(\mu^m (1-\mu)^{N-m}) \\
&= \arg\max_{\mu} \underbrace{m\log(\mu) + (N-m)\log(1-\mu)}_{=L(\mu)}.
\end{aligned}$$

If we set the derivative of this expression equal to zero we get

$$\begin{aligned}
0 &= \frac{dL(\mu)}{d\mu} \\
0 &= \frac{m}{\mu} - \frac{N-m}{1-\mu} \\
0 &= m(1-\mu) - (N-m)\mu \\
\mu &= \frac{m}{N} = \frac{1}{N}\sum_{n=1}^{N} x_n.
\end{aligned}$$

This is a maximum since

$$\frac{d^2L(\mu)}{d\mu^2} = -\frac{m}{\mu^2} - \frac{N-m}{(1-\mu)^2} < 0.$$

b) The posterior can be computed as

$$\begin{aligned}
p(\mu|m) &\propto p(m|\mu)p(\mu) \\
&\propto \text{Bin}\,(m; N,\ \mu)\,\text{Beta}\,(\mu; a,\ b) \\
&= \mu^m(1-\mu)^{N-m}\mu^{a-1}(1-\mu)^{b-1} \\
&= \mu^{m+a-1}(1-\mu)^{N-m+b-1}.
\end{aligned}$$

Hence, the posterior is also a Beta distribution

$$p(\mu|m) = \text{Beta}\,(\mu; a^*,\ b^*)\,,$$

where

$$\begin{aligned}
a^* &= m + a, \\
b^* &= N - m + b.
\end{aligned}$$

c) With the maximum likelihood solution in **(a)** we get the model

$$\hat{\mu} = \frac{1+1+1}{3} = 1.$$

Hence , under this model we would be certain that the next coin flip is $x_4 = 1$ since $\hat{\mu} = 1$.
With the Bayesian solution in **(b)** we get the posterior

$$p(\mu|x_1 = 1, x_2 = 1, x_3 = 1) = \text{Beta}\,(\mu; 4,\ 1)\,,$$

which does not have all of its mass at $\mu = 1$. More specifically,

$$\begin{aligned}
p(x_4 = 1|x_1 = 1, x_2 = 1, x_3 = 1) &= \int_0^1 p(x_4 = 1|\mu)p(\mu|x_1 = 1, x_2 = 1, x_3 = 1)d\mu \\
&= \int_0^1 \mu \cdot \text{Beta}\,(\mu; 4,\ 1)\,d\mu \\
&= \mathbb{E}[\mu] = \frac{4}{4+1} = \frac{4}{5}.
\end{aligned}$$

where we have used the expression of the mean for the Beta distribution.

Based on the spare data of only three observation, the Bayesian solution seems more realistic than the maximum likelihood solution. Another way to view it is that the maximum likelihood solution has overfitted the data in comparison to the Bayesian solution. This is a general drawback of maximum likelihood solutions in comparison to the Bayesian solution when N is small.

Solution to Exercise 1.10 a) Consider the Bernoulli likelihood Bern $(x;\mu) = \mu^x(1-\mu)^{1-x}$

$$\begin{aligned}
p(\mu|x_1=1) &\propto p(x_1=1|\mu)p(\mu) \\
&\propto \text{Bern}\,(1;\mu)\left(\frac{1}{2}\text{Beta}\,(\mu;1,\,1)+\frac{1}{2}\delta(\mu-0.5)\right) \\
&= \mu(1-\mu)^{1-1}\left(\frac{1}{2}\frac{\Gamma(2)}{\Gamma(1)\Gamma(1)}\mu^{1-1}(1-\mu)^{1-1}+\frac{1}{2}\delta(\mu-0.5)\right) \\
&= \frac{1}{2}\frac{\Gamma(2)}{\Gamma(1)\Gamma(1)}\frac{\Gamma(2)\Gamma(1)}{\Gamma(3)}\frac{\Gamma(3)}{\Gamma(2)\Gamma(1)}\mu^{2-1}(1-\mu)^{1-1}+\frac{1}{2}\mu\delta(\mu-0.5) \\
&= \frac{1}{2}\cdot\frac{1}{2}\frac{\Gamma(2+1)}{\Gamma(2)\Gamma(1)}\mu^{2-1}(1-\mu)^{1-1}+\frac{1}{2}\cdot\frac{1}{2}\delta(\mu-0.5) \\
&= \frac{1}{4}\text{Beta}\,(\mu;2,\,1)+\frac{1}{4}\delta(\mu-0.5).
\end{aligned} \tag{26}$$

Since $\int_0^1 \text{Beta}\,(\mu;2,\,1)\,d\mu = 1$ and $\int_0^1 \delta(\mu-0.5)d\mu = 1$, we get a correctly normalized posterior if we multiply (26) by 2

$$p(\mu|x_1=1) = \frac{1}{2}\text{Beta}\,(\mu;2,\,1)+\frac{1}{2}\delta(\mu-0.5).$$

In particular, we still have 50% belief that the coin is fair, which is natural since the fact that one of the sides of the coin has turned up does not provide any information about the fairness. However, if the coin would be unfair, our belief is that μ is closer to 1 than to 0.

Alternative solution

You can also consider the full Bayesian theorem

$$p(\mu|x_1=1) = \frac{p(x_1=1|\mu)p(\mu)}{p(x_1=1)} = \frac{p(x_1=1|\mu)p(\mu)}{\int_0^1 p(x_1=1|\mu)p(\mu)d\mu}$$

where

$$p(x_1=1|\mu)p(\mu) = \cdots = \frac{1}{2}\mu+\frac{1}{2}\mu\delta(\mu-0.5)$$

by following the first lines of the previous solution and then get the normalization constant by computing the integral

$$\int_0^1 p(x_1=1|\mu)p(\mu)d\mu = \int_0^1 \frac{1}{2}\mu+\frac{1}{2}\mu\delta(\mu-0.5)d\mu = \frac{1}{4}+\frac{1}{4}=\frac{1}{2}$$

This gives

$$p(\mu|x_1=1) = \frac{p(x_1=1|\mu)p(\mu)}{\int_0^1 p(x_1=1|\mu)p(\mu)d\mu} = \frac{1}{2}2\mu+\frac{1}{2}\delta(\mu-0.5) = \frac{1}{2}\text{Beta}\,(\mu;2,\,1)+\frac{1}{2}\delta(\mu-0.5).$$

b) Again consider the Bernoulli likelihood Bern $(x; \mu) = \mu^x(1-\mu)^{1-x}$ and now use the previous posterior as our prior

$$\begin{aligned}
p(\mu|x_1 = 1, x_1 = 2) &\propto p(x_2 = 1|\mu)p(\mu|x_1 = 1)\\
&\propto \text{Bern}\,(1;\mu)\left(\frac{1}{2}\text{Beta}\,(\mu; 2,\ 1) + \frac{1}{2}\delta(\mu - 0.5)\right)\\
&= \mu(1-\mu)^{1-1}\left(\frac{1}{2}\frac{\Gamma(3)}{\Gamma(2)\Gamma(1)}\mu^{2-1}(1-\mu)^{1-1} + \frac{1}{2}\delta(\mu - 0.5)\right)\\
&= \frac{1}{2}\frac{\Gamma(3)}{\Gamma(2)\Gamma(1)}\frac{\Gamma(3)\Gamma(1)}{\Gamma(4)}\frac{\Gamma(4)}{\Gamma(3)\Gamma(1)}\mu^{3-1}(1-\mu)^{1-1} + \frac{1}{2}\mu\delta(\mu - 0.5)\\
&= \frac{1}{2}\cdot\frac{2}{3}\frac{\Gamma(3+1)}{\Gamma(3)\Gamma(1)}\mu^{3-1}(1-\mu)^{1-1} + \frac{1}{2}\cdot\frac{1}{2}\delta(\mu - 0.5)\\
&= \frac{1}{3}\text{Beta}\,(\mu; 3,\ 1) + \frac{1}{4}\delta(\mu - 0.5). \qquad (27)
\end{aligned}$$

The integral of (27) from 0 to 1 is equal to $\frac{1}{3} + \frac{1}{4} = \frac{7}{12}$. Hence, we need to multiply that expression with $\frac{12}{7}$ for the posterior to be correctly normalized.

$$p(\mu|x_1 = 1, x_1 = 2) = \frac{4}{7}\text{Beta}\,(\mu; 3,\ 1) + \frac{3}{7}\delta(\mu - 0.5)$$

Consequently, now belief is slightly less likely than 50% that the coin is fair. This is natural since both observations had the same outcome.

Alternative solution Similar to the alternative solution in **(a)**, computing the normalization constant by integrating.

c) We want to compute

$$p(\texttt{fair}|x_1 = 1, x_2 = 1) = \frac{p(x_1 = 1, x_2 = 1|\texttt{fair})p(\texttt{fair})}{p(x_1 = 1, x_2 = 1|\texttt{fair})p(\texttt{fair}) + p(x_1 = 1, x_2 = 1|\texttt{unfair})p(\texttt{unfair})} \qquad (28)$$

where

$$\begin{aligned}
p(\texttt{fair}) &= \frac{1}{2},\\
p(\texttt{unfair}) &= \frac{1}{2},\\
p(x_1 = 1, x_2 = 1|\texttt{fair}) &= \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}.
\end{aligned}$$

For $p(x_1 = 1, x_2 = 1|\texttt{unfair})$ we have to do some more work

$$\begin{aligned}
p(x_1 = 1, x_2 = 1|\texttt{unfair}) &= \int_0^1 p(x_1 = 1, x_2 = 1|\mu, \texttt{unfair})p(\mu|\texttt{unfair})d\mu\\
&= \int_0^1 \text{Bin}\,(2; 1,\ \mu)\,\text{Beta}\,(\mu; 1,\ 1)\,d\mu\\
&= \int_0^1 \mu^2(1-\mu)^{2-2}\mu^{1-1}(1-\mu)^{1-1}d\mu\\
&= \int_0^1 \mu^{3-1}(1-\mu)^{1-1}d\mu\\
&= \int_0^1 \frac{1}{3}\frac{3!}{2!0!}\mu^{3-1}(1-\mu)^{1-1}d\mu\\
&= \int_0^1 \frac{1}{3}\text{Beta}\,(\mu; 3,\ 1)\,d\mu = \frac{1}{3}.
\end{aligned}$$

This inserted in (28) gives

$$p(\texttt{fair}|x_1 = 1, x_2 = 1) = \frac{\frac{1}{4}\frac{1}{2}}{\frac{1}{4}\frac{1}{2} + \frac{1}{3}\frac{1}{2}} = \frac{3}{7}.$$

Bibliography

[1] David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

[2] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[3] Kevin B Korb and Ann E Nicholson. *Bayesian artificial intelligence*. CRC press, 2010.

[4] Steffen L Lauritzen and David J Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):157–194, 1988.

[5] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.