Chapter 5

Priors

5.1 Dodgy coins

Suppose there are three coins in a bag. The first coin is biased towards heads, with a 75% probability of a heads occurring if the coin is flipped. The second is fair, so a 50% chance of heads occurring. The third coin is biased towards tails, and has a 25% probability of coming up heads. Assume that it is impossible to identify which coin is which from looking at or touching them.

Problem 5.1.1. Suppose we put our hand into the bag and pull out a coin. We then flip the coin and find it comes up heads. Let the random variable C=1,2,3 denote the identity of the coin, where the probability of heads is (0.75,0.50,0.25), respectively. Obtain the likelihood by using the equivalence relation (that a likelihood of a parameter value given data is equal to the probability of data given a parameter value), and show that the sum of the likelihood over all parameter values is 1.5.

Using the equivalence relation we obtain a likelihood of the form,

$$L(C|X=H) = \begin{cases} Pr(X=H|C=1) = 0.75, & \text{if } C=1 \\ Pr(X=H|C=2) = 0.50, & \text{if } C=2 \\ Pr(X=H|C=3) = 0.25, & \text{if } C=3 \end{cases}$$

So we obtain 1.50 by summing across all likelihoods.

Problem 5.1.2. What is the maximum likelihood estimate of the coin's identity?

From the likelihood we can see that $C=1$ maximises the likelihood.

Problem 5.1.3. Use Bayes' rule to prove that:

$$Pr(C=c|X=H) \propto Pr(X=H|C=c) \times Pr(C=c) \tag{5.1}$$

where $c=1,2,3$.

parameter C	likelihood $Pr(X = H \mid C = c)$	prior $Pr(C = c)$	likelihood $\times$ prior $Pr(X = H \mid C = c) \times Pr(C = c)$	posterior $Pr(C = c \mid X = H)$
1				
2				
3				
			$Pr(X = H) =$	

Table 5.1: A Bayes box for the coins example.

parameter C	likelihood $Pr(X = H \mid C = c)$	prior $Pr(C = c)$	likelihood $\times$ prior $Pr(X = H \mid C = c) \times Pr(C = c)$	posterior $Pr(C = c \mid X = H)$
1	3/4	1/3	3/12	1/2
2	2/4	1/3	2/12	1/3
3	1/4	1/3	1/12	1/6
			$Pr(X = H) = 6/12 = 1/2$	

Table 5.2: A Bayes box for the coins example.

Problem 5.1.4. Assume that since we cannot visually detect the coin's identity we use a uniform prior $Pr(C = c) = 1$ for $c = 1, 2, 3$. Use this to complete Table 5.1 (known as a Bayes' box) and determine the (marginal) probability of the data.

The completed Bayes box is shown in Table 5.2, where since the posterior probabilities are non-negative and their sum is 1, we can conclude that we have a valid probability distribution. The probability of the data is 0.5.

Problem 5.1.5. Confirm that the posterior is a valid probability distribution.

Since the individual probabilities are non-negative and their sum is 1, this is a valid probability distribution.

Problem 5.1.6. Now assume that we flip the same coin twice, and find that it lands heads up on both occasions. By using a Table similar in form to Table 5.1, or otherwise, determine the new posterior distribution.

Table 5.3 shows how we calculate the new posterior distribution. There is now an increased weighting towards $C = 1$, reflecting the increased chance of observing two heads given this state of the world.

Problem 5.1.7. Now assume that you believe that the tails-biased coin is much more likely to be drawn from the bag, and thus specify a prior: $Pr(C = 1) = 1/20$, $Pr(C = 2) = 5/20$ and $Pr(C = 3) = 14/20$. What is the posterior probability that $C = 1$ now?

The Bayes box for this example is shown in Table 5.4.

parameter C	likelihood $Pr(X_1 = H, X_2 = H \mid C = c)$	prior $Pr(C = c)$	likelihood $\times$ prior $Pr(X_1 = H, X_2 = H \mid C = c) \times Pr(C = c)$	posterior $Pr(C = c \mid X_1 = H, X_2 = H)$
1	9/16	1/3	9/48	9/14
2	4/16	1/3	4/48	4/14
3	1/16	1/3	1/48	1/14
			$Pr(X = H) = 14/48 = 7/24$	

Table 5.3: A Bayes box for the coins example where the coin is flipped twice, landing heads up both times.

parameter C	likelihood $Pr(X_1 = H, X_2 = H \mid C = c)$	prior $Pr(C = c)$	likelihood $\times$ prior $Pr(X_1 = H, X_2 = H \mid C = c) \times Pr(C = c)$	posterior $Pr(C = c \mid X_1 = H, X_2 = H)$
1	9/16	1/20	9/320	9/43
2	4/16	5/20	20/320	20/43
3	1/16	14/20	14/320	14/43
			$Pr(X = H) = 43/320$	

Table 5.4: A Bayes box for the coins example where the coin is flipped twice, landing heads up both times.

Problem 5.1.8. Continuing on from the previous example, calculate the posterior mean, maximum a posteriori (MAP) and maximum likelihood estimates. Does the posterior mean indicate much here?

Using the Bayes Box in Table 5.4, we calculate a posterior mean of,

$$\begin{aligned} \mathbb{E}(C \mid X_1 = H, X_2 = H) &= \sum_{C=1}^{3} 1 \times 9/43 + 2 \times 20/43 + 3 \times 14/43 && (5.2) \\ &= 91/43 \approx 2.12 && (5.3) \end{aligned}$$

The MAP estimate is $C = 2$ since this is the posterior mode. The maximum likelihood estimate is $C = 1$.

Problem 5.1.9. For the case when we flip the coin once and obtain $X = H$, using the uniform prior on C, determine the posterior predictive distribution for a new coin flip with result $\tilde{X}$, using the below expression,

$$Pr(\tilde{X} \mid X = H) = \sum_{C=1}^{3} Pr(\tilde{X} \mid C) \times Pr(C \mid X = H) \tag{5.4}$$

Using Table 5.2 we first determine the probability of a heads,

$$\begin{aligned} Pr(\tilde{X} = H \mid X = H) &= 3/4 \times 1/2 + 2/4 \times 1/3 + 1/4 \times 1/6 && (5.5) \\ &= 7/12 && (5.6) \end{aligned}$$

Then we calculate the probability of a tails,

$$\begin{aligned} Pr(\tilde{X} = H | X = H) &= 1/4 \times 1/2 + 2/4 \times 1/3 + 3/4 \times 1/6 && (5.7) \\ &= 5/12 && (5.8) \end{aligned}$$

These two results taken together form a valid probability distribution since the probabilities sum to 1.

Problem 5.1.10. (Optional) Justify the use of the expression in the previous question.

To do this we marginalise out C from the joint probability $Pr(\tilde{X}, C | X = H)$,

$$\begin{aligned} Pr(\tilde{X} | X = H) &= \sum_{C=1}^{3} Pr(\tilde{X}, C | X = H) && (5.9) \\ &= \sum_{C=1}^{3} Pr(\tilde{X} | C, X = H) \times Pr(C | X = H) && (5.10) \\ &= \sum_{C=1}^{3} \underbrace{Pr(\tilde{X} | C)}_{likelihood} \times \underbrace{Pr(C | X = H)}_{posterior} && (5.11) \end{aligned}$$

where we got from the first line to the second using the law of conditional probability, and from the second to the third by realising that once we know C the result $\tilde{X}$ is independent of $X = H$.

5.2 Left-handedness

Suppose that we are interested in the prevalence of left-handedness in a particular population.

Problem 5.2.1. We begin with a sample of one individual whose dexterity we record as $X = 1$ for left-handed, $X = 0$ otherwise. Explain why the following probability distribution makes sense here:

$$Pr(X | \theta) = \theta^X (1 - \theta)^{1-X}, \tag{5.12}$$

where θ is the probability that a randomly chosen individual is left-handed.

Under the two circumstances it yields the relevant probabilities, $Pr(X = 1 | \theta) = \theta$ and $Pr(X = 0 | \theta) = 1 - \theta$.

Problem 5.2.2. Suppose we hold θ constant. Demonstrate that under these circumstances the above distribution is a valid probability distribution. What sort of distribution is this?

Summing together the two possibilities here,

$$Pr(X = 1|\theta) + Pr(X = 0|\theta) = \theta + 1 - \theta = 1. \tag{5.13}$$

In this case we are dealing with a discrete distribution (a Bernoulli).

Problem 5.2.3. Now suppose we randomly sample a person who happens to be left-handed. Using the above function calculate the probability of this occurring.

$$Pr(X = 1|\theta) = \theta \tag{5.14}$$

Problem 5.2.4. Show that when we vary θ the above distribution does not behave as a valid probability distribution. Also, what sort of distribution is this?

The way to demonstrate this is by integrating the above over a range of θ,

$$\int_0^1 \theta \mathrm{d}\theta = \frac{1}{2} \tag{5.15}$$

So it is not a valid continuous probability distribution and hence we call it a 'likelihood'.

Problem 5.2.5. What is the maximum likelihood estimator for θ?

$\hat{\theta} = 1$ maximises the probability of obtaining one individual who is left-handed.

Bibliography