4.

Intro

In this question, we have to infer a posterior for the parameter θ based on limited information. In the usual situation, we would have the results of the $n = 5$ tosses, which correspond to one element of the sample space Ω. If this were the case, the result of the inference would be a posterior with a Beta distribution. The corresponding parameters would be a combination based on the parameters of the likelihood and the prior, as showed in Chapter 3. However, we are not working with this simple situation. The only information that we have about the result is that the number of head is less than three ($X < 3$). So we have an event which have more than one possible outcome. Fortunately, it is simple to decompose this particular event in individual results of the sample space. Being more specific, $p(X < 3|\theta) = p(X = 0|\theta)+p(X = 1|\theta)+p(X = 2|\theta)$. Therefore, our problem is nothing more than a combination of the simpler case mentioned above. As such, we should expect a similar combination to appear in the result.

Solution

Random variables

$X - number of heads$

The posterior is given by Bayes' rule:

$$p(\theta|X < 3) = \frac{p(X < 3|\theta)p(\theta)}{p(X < 3)} \tag{1}$$

Since we want derive an expression proportional to the posterior, we can ignore $p(X < 3)$ from the calculation. So, what is left is to calculate the likelihood, calculate the prior and combine both of them. We shall start with the likelihood:

$$\begin{aligned}&p(X < 3|\theta) = p(X = 0|\theta) + p(X = 1|\theta) + p(X = 2|\theta) = \\ &\binom{5}{0}\theta^0(1-\theta)^5 + \binom{5}{1}\theta^1(1-\theta)^4 + \binom{5}{2}\theta^2(1-\theta)^3 = \\ &Bin(0|\theta, 5) + Bin(1|\theta, 5) + Bin(2|\theta, 5)\end{aligned} \tag{2}$$

As (2) shows, the likelihood is a mixture distribution composed by 3 Binomial distributions. Now, let's focus our attention in the prior $p(\theta)$. As the question inform us, the prior is given by:

$$p(\theta) = Beta(\theta|1, 1) = \frac{1}{B(1, 1)}\theta^0(1-\theta)^0 = 1 = U(0, 1) \tag{3}$$

Thus, the prior is a uniform distribution over the interval $[0, 1]$. Therefore, it became easy to calculate the posterior (up to a normalization constant):

$$\begin{aligned}&p(\theta|X < 3) \propto p(X < 3|\theta)p(\theta) = \\ &(Bin(0|\theta, 5) + Bin(1|\theta, 5) + Bin(2|\theta, 5))\end{aligned} \tag{4}$$

Therefore, $p(\theta|X < 3) \propto (Bin(0|\theta, 5) + Bin(1|\theta, 5) + Bin(2|\theta, 5))$.

Conclusion

In this question, we showed how to infer the posterior of the parameter θ from a censored likelihood. The final result was a mixture model formed by three binomial distributions with the same weight. A mixture model was expected, since our censored likelihood could be decomposed in a sum of non censored likelihoods.

The only question left open is the following: the posterior of the non censored case gives us a posterior with a Beta distribution. So why don't we arrive at a mixture model involving Beta distributions on this exercise? To answer this, one should note that the binomial and the beta distribution have the same general form $\theta^{\gamma_1}(1-\theta)^{\gamma_2}$. To be more precise, we have the following relationship between the models: $Bin(k|\theta, n) \propto Beta(\theta|k+1, n-k+1)$. Unfortunately, the equality doesn't hold because the Beta function is not equal to the binomial coefficient. I will left to the reader to see that the proportion coefficient between our three binomial models and their corresponding Beta functions is the same. Therefore, our result could also be expressed as $p(\theta|X < 3) \propto (Beta(\theta|1, 6) + Beta(\theta|2, 5) + Beta(\theta|3, 4))$.