3.17

Intro

In this question we will calculate the marginal likelihood for the Beta Binomail model under uniform prior. But hold on. Didn't we derived a general expression for the this marginal likelihood in exercise 3.2? Yes, we did... Unfortunately, if we just use the expression derived in 3.2 in our current exercise, we get the wrong answer! This might induce the reader to think that one of the two questions is wrong in its formulation. However, that is not the case at all. Let's jump into the resolution and see what is going on.

Solution

First let's write the full posterior expression for the Beta-binomial model:

$$\begin{aligned} p(\theta|D) &= \frac{p(D|\theta)p(\theta)}{p(D)} \\ Beta(\theta|N_1+1, N-N_1+1) &= \frac{p(D|\theta)Beta(\theta|1,1)}{p(D)} \end{aligned} \tag{1}$$

Equation 1 is just the expression for the Beta-binomial posterior seen in this chapter. Note that the pseudo counts are: $a = 1$, $b = 1$ and that $N_0 = N - N_1$. For the next step we should be very carefull. We are working with the **sufficient statistics** N and N_1, therefore the likelihood is given by:

$$p(D|\theta) = \binom{N}{N_1}\theta^{N_1}(1-\theta)^{N-N_1} = Bin(N_1|N,\theta) \tag{2}$$

Substituting (2) in (1) and putting $p(D) = p(N_1|N)$ in evidence, we get:

$$p(N_1|N) = \frac{Bin(N_1|N,\theta)Beta(\theta|1,1)}{Beta(\theta|N_1+1, N-N_1+1)} \tag{3}$$

Now, remember that $p(N_1|N)$ is a **marginalization**. Thus, it does not depends on θ. Thus, we can work only with coefficients of each term.

$$\begin{aligned} p(N_1|N) &= \frac{\binom{N}{N_1}B(N_1+1, N-N_1+1)}{B(1,1)} = \\ \frac{N!}{(N-N_1)!N_1!} & \frac{\Gamma(N_1+1)\Gamma(N-N_1+1)}{\Gamma(N+2)} = \frac{1}{N+1} \end{aligned} \tag{4}$$

Conclusion

Ok, we arrived at the desired result. But **why** it is different from the result deduced in 3.2? In order to fully understand it, we should quote one phrase from question 2.1: "Probabilities are sensitive to the form of the question that was used to generate the answer". So what is the difference from the questions made here and in 3.2? In 3.2 we are working with the raw dataset, i.e. the exact sequence of results of the coin tosses. Here, on the other hand, we are working with the sufficient statistics N and N_1. Despite being equivalent to infer the posterior, these two datasets have different likelihood and marginal likelihoods.

Regarding the result, we have a uniform distribution $p(N_1|N) = \frac{1}{N+1}$ saying that each of the outcomes is equally probable. This makes a lot of sense, because our prior is uniform $\theta \sim Beta(1,1)$, so there is no combinations of heads that is more favorable than the rest.