2.

Intro

In this question, we will derive an alternative expression for the marginal likelihood of the Beta-Bernoulli model $p(D)$. In another words, we will calculate the probability of a given dataset to occur. It is important to note that the original expression is only derived in chapter 5, so there is not much to comment on it right now. Fortunately, the author gives all that is needed to solve the exercise in the question's statement. Therefore, the only thing left to do is to show that the marginal likelihood can be experssed as the product of several gamma functions.

Solution

It is showed in the question statement that the marginal likelihood can expressed as:

$$p(D) = \frac{[(\alpha_1)...(\alpha_1 + N_1 - 1)][(\alpha_0)...(\alpha_0 + N_0 - 1)]}{(\alpha)...(\alpha + N - 1)} \tag{1}$$

Now, we only need to realize that each of the three terms of (1) (the two products over squared brackets and the denominator) can be expressed as gamma functions. Being more specific, we have:

$$(\alpha_1)...(\alpha_1 + N_1 - 1) = \frac{\Gamma(\alpha_1 + N_1)}{\Gamma(\alpha_1)} \tag{2}$$

$$(\alpha_0)...(\alpha_0 + N_0 - 1) = \frac{\Gamma(\alpha_0 + N_0)}{\Gamma(\alpha_0)} \tag{3}$$

$$(\alpha)...(\alpha + N - 1) = \frac{\Gamma(\alpha + N)}{\Gamma(\alpha)} \tag{4}$$

Where $\alpha = \alpha_0 + \alpha_1$. Substituting (2), (3) and (4) in (1) we arrive at the desired result:

$$p(D) = \frac{\Gamma(\alpha_1 + N_1)}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_0 + N_0)}{\Gamma(\alpha_0)} \frac{\Gamma(\alpha)}{\Gamma(\alpha + N)} = \frac{\Gamma(\alpha_1 + N_1)\Gamma(\alpha_0 + N_0)}{\Gamma(\alpha_1 + \alpha_0 + N)} \frac{\Gamma(\alpha_1 + \alpha_0)}{\Gamma(\alpha_1)\Gamma(\alpha_0)} \tag{5}$$

Conclusion

In this question, we derived an alternative expression for the marginal likelihood of the Beta-Bernoulli model. The original expression was not showed in Chapter 3 and we will only see it when we arrive at Chapter 5 (Bayesian statistics). Two things should be noticed in this result. First of all, the marginal distribution $p(D)$ is the normalization constant that we usually ignore while doing inference (we usually work only with the prior and the likelihood when calculating the posterior). We can ignore $p(D)$ because it is a **number** and not a function of θ, like the likelihood and the prior. Second, the marginal distribution is a function both of the dataset results(N_1 and N_0) and the hyperparameters (α_0 and α_1). This reinforces the bayseian creed: the probability of an event is a combination of our prior beliefs and the actual results.