13.

Intro

In this question, we will calculate the posterior predictive for a batch of data, using the Dirichlet-multinomial model. This is a straightfoward math question, that will follow the same pattern as the previous posterior predictives we saw before. Nevertheless, there is easier and hard ways to arrive at the result. We will see two easy ways to arrive at the final expression.

Solution

Conventions

$N = N^{old} + N^{new}$

$N_j = N_j^{old} + N_j^{new}$

First solution

Let's start by writing down the posterior predictive for one single trial:

$$p(X = j|D, \alpha) = \frac{\alpha_j + N_j}{\alpha + N} \tag{1}$$

Now, we need to express the batch of data as a series of single trials:

$$p(\widetilde{D}|D, \alpha) = p(\widetilde{x_1}|D)p(\widetilde{x_2}|\{D, \widetilde{x_1}\})p(\widetilde{x_3}|\{D, \widetilde{x_1}, \widetilde{x_2}\})... \tag{2}$$

Now we just have to substitute (1) in (2), updating the number of empirical counts of the total amount and of the spefic trial for each instance:

$$\begin{gathered} p(\widetilde{D}|D, \alpha) = \frac{1}{\prod_{i=0}^{N-1}(\alpha + N^{old} + i)} \prod_{j=1}^{K} \prod_{i=0}^{N_j^{new}-1} (\alpha_j + N_j^{old} + i) = \\ \frac{\Gamma(\alpha + N^{old})}{\Gamma(\alpha + N)} \prod_{j=1}^{K} \frac{\Gamma(\alpha_j + N_j)}{\Gamma(\alpha_j + N_j^{old})} \end{gathered} \tag{3}$$

Second solution

The posterior distribution, given the original dataset D, is given by:

$$p(\theta|D, \alpha) = Dir(\theta|\alpha + N^{old}) \tag{4}$$

which have the same form as the prior. So, we may think of the posterior as a prior with updated hyper parameters ($\alpha_j^{new} = N_j^{old} + \alpha_j$).

Now, we can use the marginal likelihood expression given by the author, using the new dataset $\widetilde{D}$ as input and the new hyper parameters as the parameter:

$$\begin{gathered} p(\widetilde{D}|\alpha^{new}) = \frac{\Gamma(\alpha^{new})}{\Gamma(N^{new} + \alpha^{new})} \prod_j \frac{\Gamma(N_j^{new} + \alpha_j^{new})}{\Gamma(\alpha_j^{new})} = \\ \frac{\Gamma(N^{old} + \alpha)}{\Gamma(N^{new} + N^{old} + \alpha)} \prod_j \frac{\Gamma(N_j^{new} + N_j^{old} + \alpha_j)}{\Gamma(N_j^{old} + \alpha_j)} = \\ \frac{\Gamma(\alpha + N^{old})}{\Gamma(\alpha + N)} \prod_{j=1}^{K} \frac{\Gamma(\alpha_j + N_j)}{\Gamma(\alpha_j + N_j^{old})} \end{gathered} \tag{5}$$

Conclusion

In this question, we derived the expression for the posterior predictive of a batch of data using the Dirichlet-multinomial model. We saw two ways of deriving the result. The first one use the chain product rule of probability and the second one used updated hyper parameters together with the expression for the marginal likelihood.