7.

Intro

In this exercise, we will perform Bayesian analysis of the Poisson distribution. This can be thought as an extension of the last exercise, where we calculate the MLE for the same distribution. Let's see if the results of this exercise have any relationship with the MLE result from 3.6.

Solution

a)

Let's derive the posterior $p(\lambda|D)$. We shall start using Bayes' rule:

$$p(\lambda|D) \propto p(D|\lambda)p(\lambda) \tag{1}$$

We know from the question statement that $p(\lambda) = Ga(\lambda|a,b) \propto \lambda^{a-1}e^{-\lambda b}$. Also, we know how to calculate the likelihood:

$$p(D|\lambda) = \prod_1^N e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = \frac{e^{-N\lambda}\lambda^{\sum_1^N x_i}}{\prod_1^N x_i!} \tag{2}$$

Notice that $\prod_1^N x_i!$ is a number and not a function of λ. Therefore, we can exclude it from the calculation.

$$\begin{aligned} &p(D|\lambda)p(\lambda) \propto e^{-N\lambda}\lambda^{\sum_1^N x_i}\lambda^{a-1}e^{-\lambda b} = e^{-(N+b)\lambda}\lambda^{\sum_1^N x_i + (a-1)} \propto \\ &Ga(\lambda|\sum_1^N x_i + a, N + b) \end{aligned} \tag{3}$$

As the author hinted, the posterior is also a Gamma function.

b)

Let's discover the tendency of the posterior when the hyperparameters a and b go to zero. First, we need to calculate the mean of the posterior.

$$E[\lambda|D] = \frac{\sum_1^N x_i + a}{N + b} \tag{4}$$

The mean is well behaved function of a and b, so it will be easy to take the limit:

$$\lim_{a\to 0, b\to 0} E[\lambda|D] = \lim_{a\to 0, b\to 0} \frac{\sum_1^N x_i + a}{N + b} = \frac{\sum_1^N x_i}{N} = \lambda_{MLE} \tag{5}$$

As we can see, (5) tell us that in the limit, the mean of the posterior becomes the maximum likelihood estimator.

Conclusion

In this exercise, we performed Bayesian analysis of the Poisson distribution. We discovered two things. First, assuming a Gamma prior, the posterior also takes the form of a Gamma distribution. Second, in the extreme case where the parameters go to zero, the expected value goes to the MLE.

The posterior result of item a was no surprise, because the author hinted the answer. Nevertheless, it is important to take a look on how the parameters are updated. The shape (first parameter) is a function of the **values** of the dataset. The rate (second parameter), on the other hand, is a function only of the **size** of the dataset. So, we can think on the hyperparameter b as a pseudo-count. Furthermore, we can think on a as the sum of the "content" of the pseudo count.

The extreme case analysis of item b showed something very interesting. The result says that when we make a and b approach zero, the MLE can substitute the expected value. If you think about it, this make a lot of sense. In the mean expression for the posterior Gamma distribution, we can see that the hyper-parameters behave like pseudo counts and actual values. If we erase these hyper-parameters, then all that is left is the sample average, which we know is equal the the MLE.