9.

Intro

In this question, we will expand on the suggestion we made on question 8.c. Being more specific, we will make a Bayesian analysis of the Uniform distribution. Remeber our discussion in the Conclusion of the last exercise? So, it seems we are indeed using the Pareto distribution to make Bayesian analysis. I hope the posterior result will help to further explain the conclusions that we made in the last exercise.

Solution

From Bayes' rule:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} = \frac{p(D,\theta)}{p(D)} = \frac{Kb^K}{\theta^{N+K+1}}\mathcal{I}(\theta \geq max(D,b))\frac{1}{p(D)} \tag{1}$$

Now, we can dismember (1) into two cases. The first case correspond to $m \leq b$:

$$p(\theta|D) = \frac{Kb^K}{\theta^{N+K+1}}\mathcal{I}(\theta \geq max(D,b))\frac{(N+K)b^N}{K} = \frac{(N+K)b^{N+K}}{\theta^{N+K+1}}I(\theta \geq b) = Pareto(\theta|N{+}K, b) \tag{2}$$

The second case correspond to $m > b$:

$$p(\theta|D) = \frac{Kb^K}{\theta^{N+K+1}}\mathcal{I}(\theta \geq max(D,b))\frac{(N+K)m^{N+K}}{Kb^K} = \frac{(N+K)m^{N+K}}{\theta^{N+K+1}}I(\theta \geq m) = Pareto(\theta|N{+}K, m) \tag{3}$$

Therefore, the posterior is given by the following Pareto distribution:

$$p(\theta|D) = Pareto(\theta|N+K, max(m,b)) \tag{4}$$

Conclusion

Well, well, well, it seems the posterior gives us more information that we initially supected in 3.8. First, let's take a look on the prior $Pareto(\theta|K,b)$. This prior is encoding the following belief: the parameter θ **must** be bigger than b. Moreover, the value of K is encoding our belief that the true result is near b. For instance, if K goes to infinity, we are basically saying that $\theta = b$. On the other hand, if K goes to 0 then there is a fair chance that the parameter might be a value distant from b.

With the fresh analysis on a Pareto distribution, let's take a new look at the posterior $p(\theta|D) = Pareto(\theta|N+K, max(m,b))$. The distribution tell us two new things. First, the MLE (m) and our prior hyper-parameter b will dispute to see who will be the most probable value for θ. The winner will be the bigger value. To see why this is true think this way. The prior states "I believe that the parameter is equal or bigger than b". At the same time, our dataset states: "The parameter must be equal or bigger than $max(|x_i|, x_i \in D)$". So it is only natural that the parameter will be equal or bigger than the maximum between them.

The second discovery is about the size of the dataset. The first parameter of the Pareto distribution goes from K to $N+K$. This is essentially saying the following: "The more data we have, the more certainty we know that the parameter is close to $max(m,b)$".