8.

Intro

The next couple of questions will follow the same pattern of 3.6 - 3.7. In this question we will perform MLE for the uniform distribution. In the next question, we will see how to perform a full Bayesian analysis on the same distribution.

Solution

a)

Let's calculate the maximum likelihood estimator of the uniform distribution on the dataset $D = \{x_1, x_2, ..., x_n\}$

$$\begin{aligned} p(D|a) = \prod_{i=1}^{N} p(x_i|a) = \prod_{i=1}^{N} \frac{1}{2a} I(x_i \in [-a, a]) = \\ \frac{1}{(2a)^N} \prod_{i=1}^{N} I(x_i \in [-a, a]) \end{aligned} \tag{1}$$

From (1), we see that the likelihood is a combination of two terms. The first term $\frac{1}{(2a)^N}$ is a monotonic decreasing function in a. Thus, as we decrease a we get bigger likelihoods. The second term $\prod_{i=1}^{N} I(x_i \in [-a, a])$ is equal to 0 if any of the points of the dataset are outside the interval $[-a, a]$, and equal to 1 otherwise. So, in order to have the maximum likelihood, we have to take the minimum value $\hat{a}$ such that $D \subset [-\hat{a}, \hat{a}]$. Let's define $D_{abs} = |x_1|, |x_2|, ..., |x_n|$. Thus, the maximum likelihood estimator is given by:

$$\hat{a} = sup(D_{abs}) = max(|x_i|) \tag{2}$$

b)

Given a new datapoint x_{n+1}, the model would assign it the following probability:

$$\text{p}(\text{x}_{n+1}) = \frac{1}{2\hat{a}} I(x_{n+1} \in [-\hat{a}, \hat{a}]) = \begin{cases} \frac{1}{2\hat{a}} & \text{if } \hat{a} \geq x_{n+1} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

c) The problem exposed in b is the zero count problem, where we have an unrealiable MLE, given the nature of our dataset. The MLE is unrealiable because it gives 0 probability to every datapoint bigger than $max(|x_i| \in D_{abs})$. This becomes even more unrealiable when we have small datasets. A better approach would be using Bayesian analysis.

Conclusion

In this question we saw how to derive the MLE for the uniform distribution and its limitations. The MLE is given by $max(|x_i| \in D_{abs})$, which makes sense. After all, as we already know, the MLE is all about maximizing the chance of our dataset to occur.

Let's make a deeper analysis of our result. Since all datapoints were sampled from the uniform distribution that we are trying to infer, it is reasonable to assume that $a \geq max(|x_i|)$. In another words, if x_i was sampled from a uniform distribution, then $p(x_i) \neq 0$ and $x_i \in [-a, a]$. Note that this result has nothing to do with MLE yet.

Now, if we want to maximize the chance of our data to occur, we already saw in item a, that we need to minimize the parameter a, given the restriction above. The problem with this assumption, as we saw in itens b and c, is that it cannot explain the apperance of any new sample bigger then $max(|x_i|)$, which consist of the zero count problem. The Bayesian analysis of the uniform distribution in the next question will show us how to overcome this limitation. The main idea behind the Bayesian analysis is to infer a distribution for the parameter a that gives zero probability to values $a < max(|x_i| \in D_{abs})$ and gives decreasing probabilities for values $a \geq max(|x_i| \in D_{abs})$. As you might suspect, the perfect distribution to model this is the Pareto distribution.