Contents

Introduction……………………………………………….3

Creation of dataset……………………………………...3

Bob’s sequential approach…………………………………4

Lisa’s one shot approach……………………………………5

Comparison between Bob’s and Lisa’s Approach……………..7

Points to ponder on…………………………………………..7

**INTRODUCTION**

In this assignment, we followed the principles of Bayes’ theorem to find a posterior probability distribution for a biased coin toss experiment. Bayes’ theorem states that probability distributions follow

**𝑃𝑜𝑠𝑡𝑒𝑟𝑖𝑜𝑟 ∝ 𝑙𝑖𝑘𝑒𝑙𝑖ℎ𝑜𝑜𝑑 × 𝑝𝑟𝑖𝑜𝑟**

**𝑃(𝜇|𝐷, 𝑎, 𝑏) ∝ 𝑃(𝐷|𝜇) × 𝑃(𝜇|𝑎, 𝑏)**

Here µ is the probability of getting heads from the dataset, D represents the data(the coin toss experiments) and a and b are parameters of the Beta Distribution as in our case the prior and posterior probability distributions follow beta distributions.

𝐵𝑒𝑡𝑎(𝜇| 𝑎, 𝑏) = (𝛤(𝑎 + 𝑏)/( 𝛤(𝑎)\*𝛤(𝑏))) 𝜇(a-1) (1 − 𝜇) (b-1).

Where a and b are the parameters of the beta distribution.

𝛤’ represents the gamma function.

**CREATION OF THE DATASET:**

We randomly generated a dataset of 160 data points with the outcome of each experiment being either a head or a tail where a 0 represents a tail and a 1 represents a head. We created the dataset such that roughly 69 percent of the outcomes were heads. So mean of the maximum likelihood estimator was = 0.69.

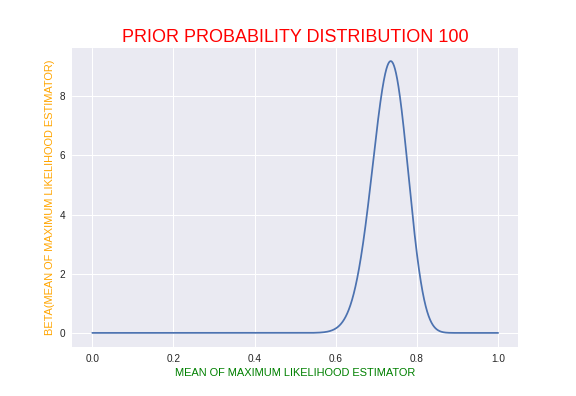
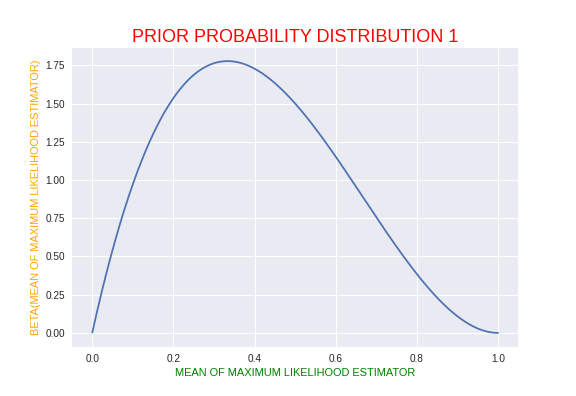
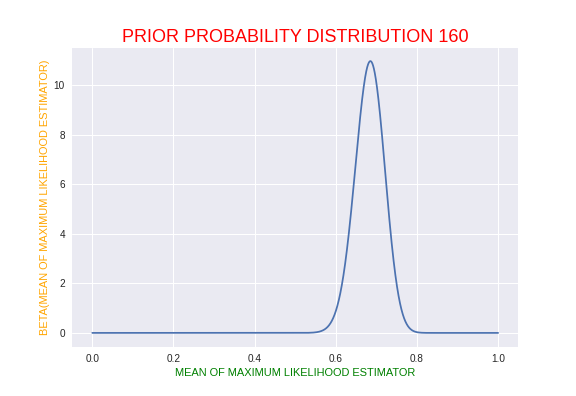
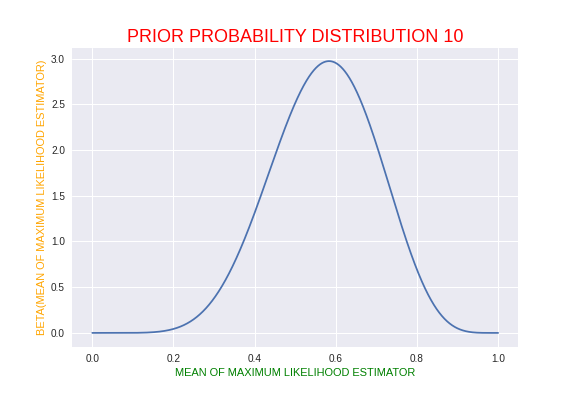
To find the posterior probability distribution, we used two approaches -

* Bob’s sequential approach
* Lisa’s one shot approach.

**BOB’S SEQUENTIAL APPROACH:**

* We took the parameters a and b of the prior distribution(which is a beta distribution) as 2 and 3 respectively.
* We then sequentially update the posterior distribution looking at one data point at a time.
* This posterior distribution becomes the prior distribution for updating the posterior distribution the next time and this process is repeated iteratively 160 times.
* OBSERVATIONS FROM BOBS APPROACH:

1. The mean of the beta distribution shifted from 0.4 to 0.6829268292682927.
2. A few plots of the posterior distribution are given below.



1. A GIF of all the 160 posterior distributions can be found in this assignment file.

**LISA’S ONE SHOT APPROACH:**

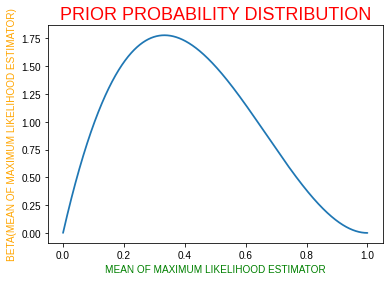
* We took the parameters a and b of the prior distribution(which is a beta distribution) as 2 and 3 respectively.
* The posterior distribution is found out by considering all the 160 data points at one time.
* The updated posterior distribution is given by Beta(m+a,l+b)

where a and b are the parameters of the prior distribution, m is the number of heads obtained from 160 coin tosses and l is the number of tails obtained from the 160 coin tosses

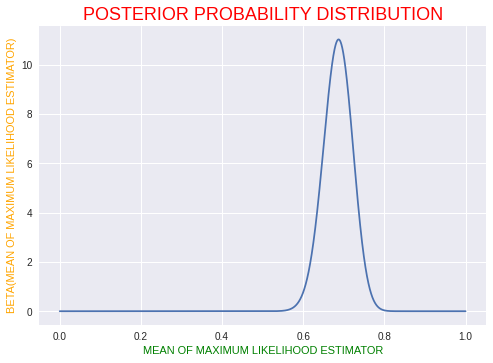
* OBSERVATIONS FROM LISA’S APPROACH:

1 .The mean of the beta distribution shifted from **0.4** to **0.6848484848484848**.

2.The plot of the prior probability distribution is shown below -



3 The plot of the posterior probability distribution is shown below -



**COMPARISON BETWEEN BETWEEN BOB’S APPROACH AND LISA’S APPROACH -**

* Both Bob and Lisa’s approach give us roughly the same probability of getting a head from the posterior distribution which is around 0.68.
* Lisa’s approach was not very time consuming so it might be useful for small datasets for faster computation but for real time applications it may be a better idea to go with Bob’s approach especially when considering very large datasets.

**POINTS TO PONDER ON-**

* As dataset size reaches infinity, we will notice the posterior curve as a vertical line passing through the mean. This means zero variance. If the mean obtained from the maximum likelihood estimator is 0.5, then the posterior graph will look like a graph whose peaks will be based on the size of our dataset. This means that the coin here is biased. The graph will resemble a normal distribution with mean 0.5.
* Since bob’s model follows sequential learning, it makes use of observations one at a time or in small batches and discards them before the next observations are used. They can be used in real time learning scenarios where a steady stream of data is arriving, and precautions must be made before all the data is seen. As they do not require the whole dataset to be stored or loaded into memory, sequential methods are also useful for large datasets.
* While choosing the prior distribution we have to choose a distribution which has certain analytical properties. For this, we take the likelihood function of the form of the product of factors of the form µ^x∗ (1 − µ)^1-x . If we choose a prior to be proportional to powers of µ and (1 − µ), then the posterior distribution, which is proportional to the product of the prior and the likelihood function, will have the same functional form as the prior distribution. This is known as **Conjugacy.** Now, since Beta distribution is also proportional to the above likelihood function, we take this distribution whereas Gamma, Gaussian or Pareto all these distributions do not follow the above-mentioned property. Hence using these distributions as the prior distribution will not give a convenient and easy posterior distribution computation compared to that of the beta distribution. Hence Conjugacy is followed when we take a beta distribution as the prior distribution but not in the case of a Gamma, Gaussian or Pareto distribution.