A Project Report

On

Generating The Posterior Distribution Using Likelihood estimator and Prior Distribution

BY

Kshitij Verma (2017B1A71145H)

Rahul Prakash (2018AAPS0893H)

Siddharth Raj (2018AAPS0398H)

Under the supervision of

Dr. NL Bhanu Murthy

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

HYDERABAD CAMPUS

(OCTOBER , 2020)

ASSIGNMENT 1: FINDING THE POSTERIOR DISTRIBUTIONS

Introduction:

Dataset And Parameters: In this assignment , the prior distribution follows a beta distribution and has parameter values a=6 and b=9. Using these as the parameter values, we get a mean of 0.4 as specified in the question. The dataset used has 160 entries consisting of 112 instances of zeroes(0) corresponding to tails and 48 instances of ones(1) corresponding to heads. Upon creation of the array, it's shuffled using Numpy's shuffling method. The maximum likelihood estimator of mean is therefore equal to 0.3.

For the prior beta distribution, we have used a=6 and b=9 so that the prior mean is 0.4

Prior distribution mean = a / (a + b) = 0.4

Plots For Bob and Lisa:

Bob: As specified in the problem statement, Bob used a sequential learning method for finding out the maximum likelihood estimator. He analyzed the data generated by the 160 flips considering one flip at a time. The total number of iterations performed by him are therefore 160. For each 1/0 he observed in the dataset, he updated the values of a, b which were initially taken as 6, 9 respectively.
As with any sequential learning problem, the posterior distribution after viewing the first example becomes the prior for finding the next posterior distribution . Following this approach he was able to get to the final plot of the maximum likelihood estimator μ_{ML}

The Final plot of the posterior distribution attained by Bob is as follows:

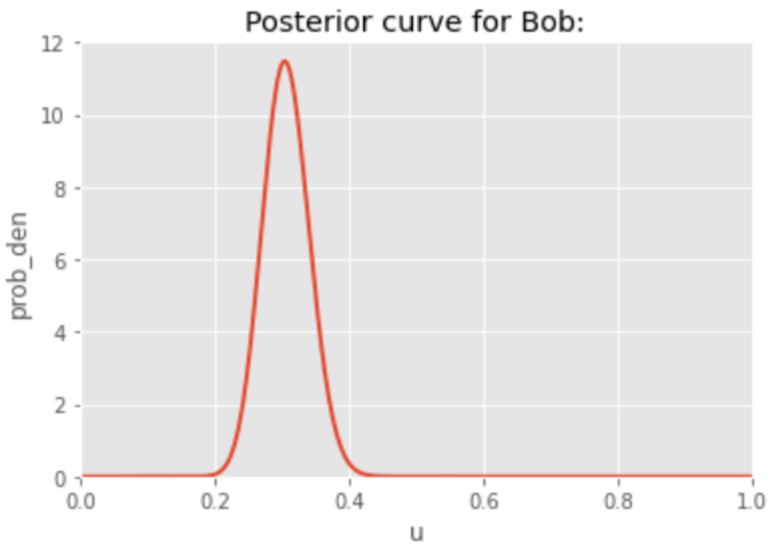

Fig. Posterior distribution for Bob

Lisa: Lisa directly takes into account the complete dataset at once by calculating the total number of heads and tails in all 160 flips. Likelihood in this case is taken directly over the whole dataset. The values of a and b change only once and directly get updated to a= (a + total heads in dataset), b=(b +total number of tails). We then use these updated values to calculate the posterior probabilities.

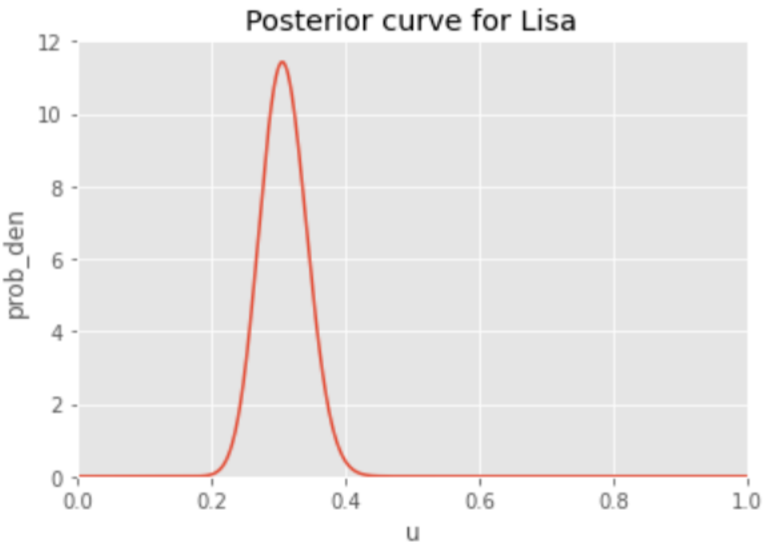

Fig. Posterior distribution for Lisa

- The sequential approach incorporates a Bayesian viewpoint. It is independent of the choice of prior distribution and of the likelihood function. The observations are recorded one at a time and deletes them before the next observation is recorded. In a sense it can be used for extrapolation of data.
- Therefore this method becomes useful when we are handling large datasets, the reason being that it does not need the whole dataset to work with. On observing the plot, the beta distribution curve becomes increasingly sharply peaked with a corresponding increase in the number of observations. This is in direct agreement that with an increase in known data uncertainty represented by the posterior steadily decreases.
- When the whole dataset is available at once, the number of ones and zeros in the dataset respectively updates the parameter values simultaneously. The final posterior beta distribution has its parameters a = 6+48 = 54 and b = 112+9= 121 which gives a mean value equal to 54/175 = 0.308 .
- Both the approaches give the same posterior value(0.308) which lies between our likelihood mean(0.3) and the prior mean(0.4).
- Both the results verify that the final posterior distribution for both the approaches is the same, i.e., with parameters a=54 and b=121.

QUESTIONS TO PONDER:

1. **The size of the dataset has been restricted to 160 data points. What happens if more points are added (say ~10^5)? What would the posterior distribution look like if μ_{ML} = 0.5? Which model, Bob's or Lisa's, would be more helpful and easier while working with large real time data and why?**

- As we increase the number of observations , the graphs' uncertainity decreases and the mean becomes sharper and more sharply peaked and the probability increases which converged to 0.308 in our case.
- By the addition of number of more data points(more tosses), the variance decreases and becomes sharper. The mean probability also increases. So, if there are ~10^5 points, the mean converges towards the likelihood with variance tending to zero.
- If the MLE = 0.5, the posterior mean converges towards 0.5 as the dataset increases.
- Therefore, sequential way(Bob's method) can be used for larger datasets as the posterior distribution can be calculated in batches and then can be combined later which takes less computational cost.

2. **What if another distribution like Gamma, Guassian or Pareto were to be chosen as the prior? Would the posterior computation be easier or difficult and why?**

- If another distribution like Gamma, Gaussian or Pareto were used as the prior distribution, then the posterior function would be different and may not be proportional to that of the present functional form and hence the computation becomes complex.