

# MA4740 - Introduction to Bayesian Statistics

1<sup>st</sup> Krishna Teja Chilukuri  
Mathematics and Computing  
Indian Institute of Technology, Hyderabad  
Hyderabad, India  
ma20btech11005@iith.ac.in

2<sup>nd</sup> Anita Dash  
Mathematics and Computing  
Indian Institute of Technology, Hyderabad  
Hyderabad, India  
ma20btech11001@iith.ac.in

3<sup>rd</sup> KN Vardhan  
Mathematics and Computing  
Indian Institute of Technology, Hyderabad  
Hyderabad, India  
ma20btech11006@iith.ac.in

**Abstract**—This report is based on a group project part of the MA4740 - Introduction to Bayesian Statistics course material. The primary goal of this project is to acquire a real-world data set (not synthetic) and execute Beta Binomial Bayesian Analysis and other approaches presented in class.

## I. INTRODUCTION

The project includes collecting a real-life data set and applying the statistical methods while also performing a beta-binomial bayesian analysis. The real-life data set that we have taken is based on the stocks of MSFT (Microsoft Corp) from the past 2010 to 2022, for prior.

The real-life data set, MSFT stocks have been taken from the Python3 *yfinance* package. The data includes opening price, closing price, maximum price, minimum price and date. Amongst which, the attributes that are of our interest are closing price and date.

## II. PRIOR DATA

As stated in the introduction, our prior data is on the stocks of MSFT from the year 2000 to 2022. The closing price of stocks' of MSFT on each and every day from all the years is taken into consideration. Based on this prior data, we try to guess if the stock of MSFT increases or decreases on a certain day after the year 2022. This data is taken as the training set.

We take the observation sample as either 0 or 1. 0 if the stock has decreased compared to the previous day and 1 if the stock has increased compared to the previous day. We choose a random variable  $X$ , where  $X$  is the proportion of days the stock price increased in  $n$  days, and  $P(X = x)$  representing the probability of the proportion,  $x$ .

E.g. Suppose in the past  $n$  days, the MSFT stock has increased  $k$  times, then

$$x = \frac{k}{n}$$

The plot between these proportions with its frequency gives us a near normal distribution. This is our prior data and it's probability mass function. The plot has been depicted below.

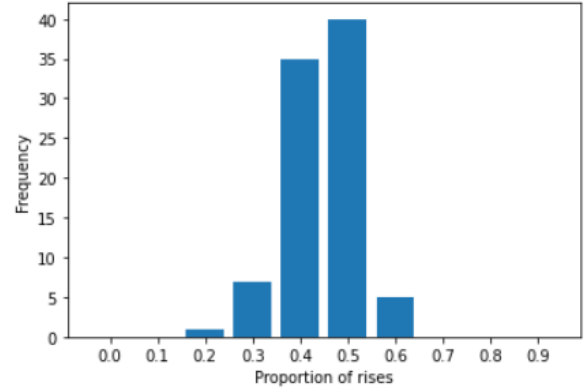


Fig. 1. Prior Data

## III. APPLYING THE METHODS

From the prior data collected, we apply MOM and MLE to fit the normal distribution of the above data. We have

$$E[X] = \mu = \frac{1}{n} \sum x_i$$
$$E[X^2] = \mu^2 + \sigma^2 = \frac{1}{n} \sum x_i^2$$

where,

- $x_i$  are i.i.d realized values of  $X$ .
- $N(\mu, \sigma)$  is the normal distribution that fits the sample data.

Since our distribution is a normal distribution, the MOM and MLE yield the same result. Below is the plot depicted for MOM and MLE.

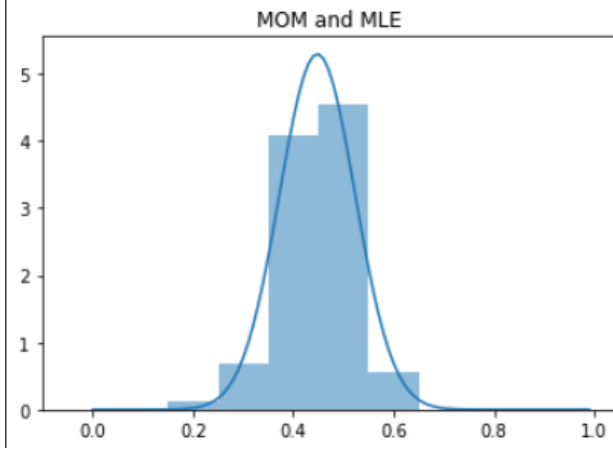


Fig. 2. MOM and MLE

The calculations yield us the results

$$\begin{aligned} \text{First moment} &= \text{Mean } (\mu) = 0.4647727, \\ \text{Second moment} &= 0.2235227, \\ \text{Variance} &= 0.0075090, \\ \text{Standard Deviation} &= 0.0866547, \end{aligned}$$

#### IV. BETA-BAYESIAN

##### A. Beta

We then apply the beta-binomial Bayesian analysis to this data. We try to fit the prior data distribution to a beta distribution. In a beta distribution, we get

$$X \sim (\alpha, \beta)$$

where,

$$\begin{aligned} \text{mean} &= \frac{\alpha}{\alpha + \beta} \\ \text{var} &= \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \end{aligned}$$

By calculating  $\alpha, \beta$  using above MOM estimators we get,

$$\begin{aligned} \alpha &= 19.05468607825279 \\ \beta &= 23.50400363967223 \end{aligned}$$

$$\text{Prior Mean} = 0.4477272727272727$$

Given below is the plot of pdf of beta distribution of the prior data.

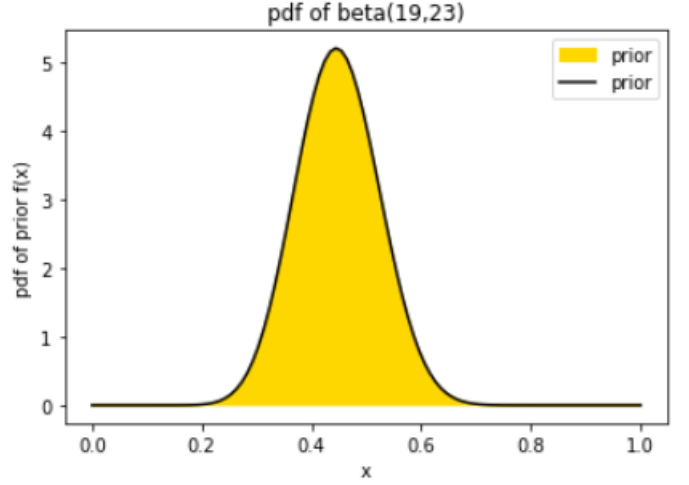


Fig. 3. PDF of Beta(19,23)

##### B. Likelihood

Now we require a likelihood function to perform a beta-binomial analysis on the generated beta distribution. We choose  $L | \pi \sim \text{Bin}(n, \pi)$  where,

- $n$ : Number of days we are checking if the stock price increased or decreased
- $\pi$ : Probability of stock getting increased

After performing the calculations, we get the value of probability of stock increasing as  $p = 0.4661354$ . And the plot of the likelihood function and the prior distributions are as follows

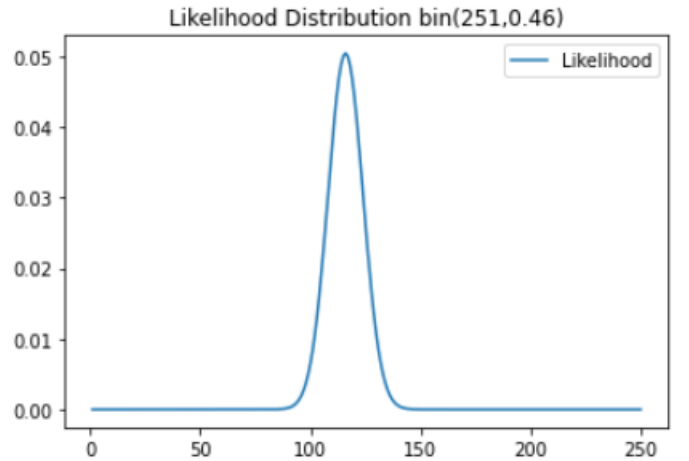


Fig. 4. Likelihood Distribution

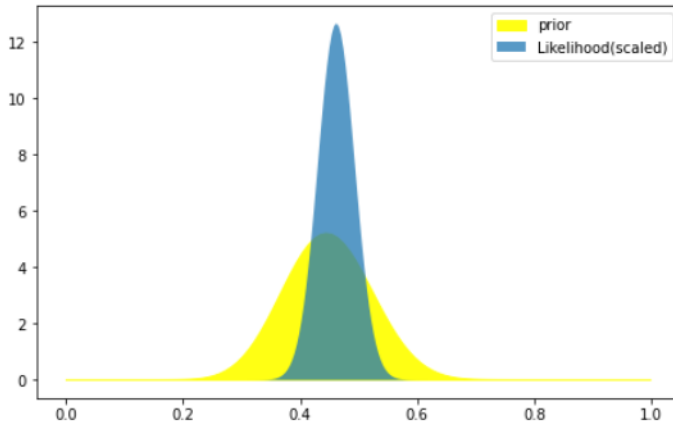


Fig. 5. Combination of Prior and Likelihood(scaled)

### C. Posterior

Finally, we find the posterior distribution from the prior data and the likelihood function. In a posterior distribution, we have

$$\pi | L = y \sim \text{beta}(\alpha', \beta')$$

where,

$$\begin{aligned}\alpha' &= \alpha + y, \\ \beta' &= \beta + n - y\end{aligned}$$

$y$  is the realized value of number of times stocks increased in  $n$  days

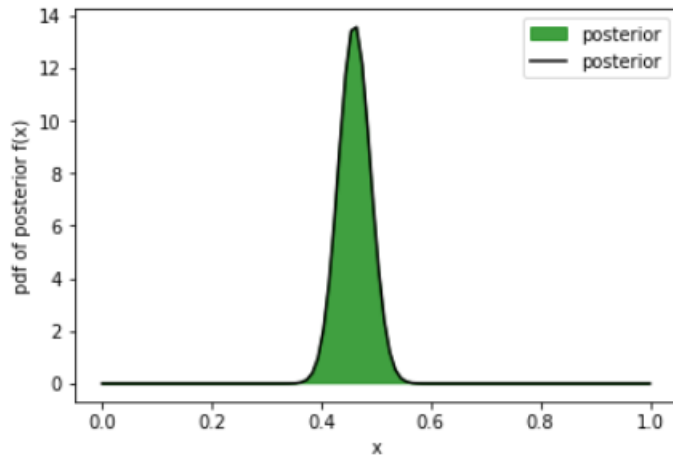
Thus, after performing the necessary calculations, we get

$$\alpha' = 135.0546860782528$$

$$\beta' = 158.50400363967225$$

$$\text{Posterior mean} = 0.46006025646191656$$

And the plot is as shown below



### D. Conclusion

From the below combined plot we can see that the mean of our prior and posterior differ by 0.02 which is not that high. But the variance of prior and posterior differ by a lot.

$$\text{Prior std} = 0.07534356570114019$$

$$\text{Post std} = 0.029039831153004316$$

$$\text{Prior estimation of proportion} \approx (29.704\%, 59.841\%)$$

$$\text{Posterior estimation of proportion} \approx (40.198\%, 51.813\%)$$

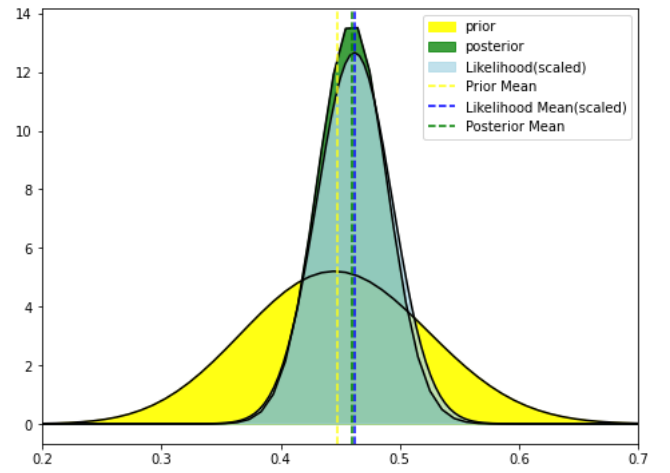


Fig. 6. Combination of Prior, Data-Likelihood and Posterior Distribution