CS F320 – FODS

Assignment 1 Report

In this assignment, we find the probability distribution of μ(the probability of getting head while tossing a biased coin) using Bayes' Formula. According to the Bayes' formula:

Posterior distribution ∝ Likelihood distribution x Prior distribution

i.e., **P (μ | D, a, b) ∝ P (D |μ) x P (μ | a, b)**

1. **Dataset**

   As we know, coin tossing follows a Bernoulli distribution. It's probability density function is given by:

   $$\boldsymbol{Bern(x|\mu) = \mu^{x}(1-\mu)^{1-x}}$$

   where μ is the mean of the Bernoulli distribution.

   Thus, for a dataset D of N points, we get the likelihood function as:

   $$\boldsymbol{P(D|\mu) = \prod_{n=1}^{N} P(x_n\,|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{(1-x_n)}}$$

   Upon differentiating the logarithm(P(D |μ)) we get $\boldsymbol{\mu_{ML} = \sum_{n=1}^{N} \frac{x_n}{N}}$

   We have generated 160 data points whose μ_{ML} = 0.7203 using the numpy.random library.

2. **Prior Distribution**

   We will take the prior to be a beta distribution. The PDF for a beta distribution is given by:

   $$\boldsymbol{Beta(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}}$$

   We have taken a=4 and b=6 such that the mean of the Beta distribution is 0.4.

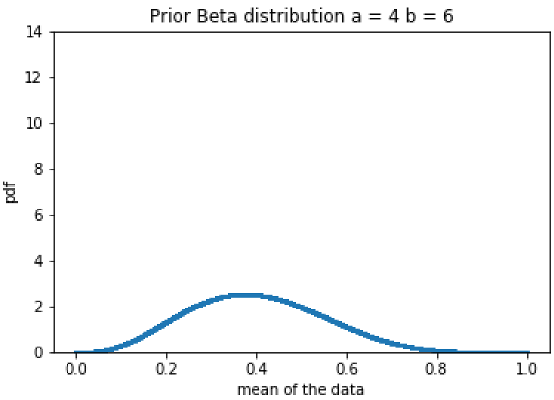
3. **Results –**

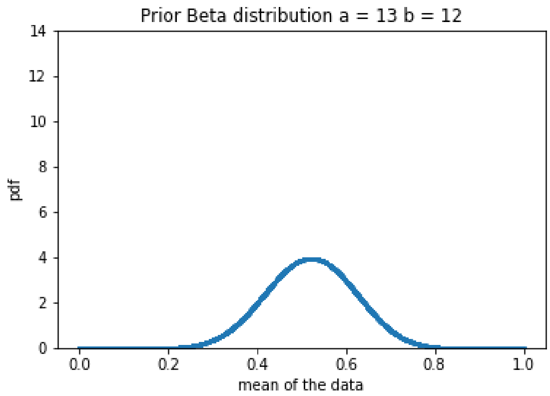
Actual mean of data = **0.7203244934421581**

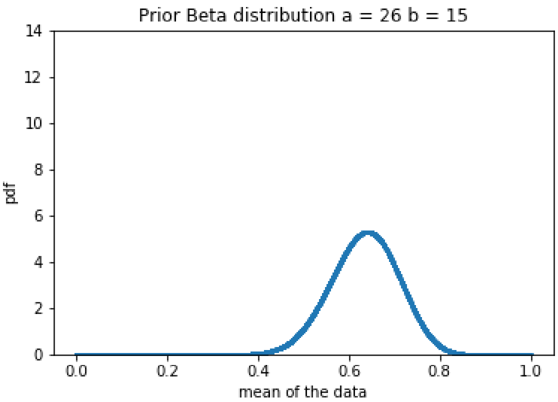
Using sequential approach we get the final mean as **0.7058823529411765.**

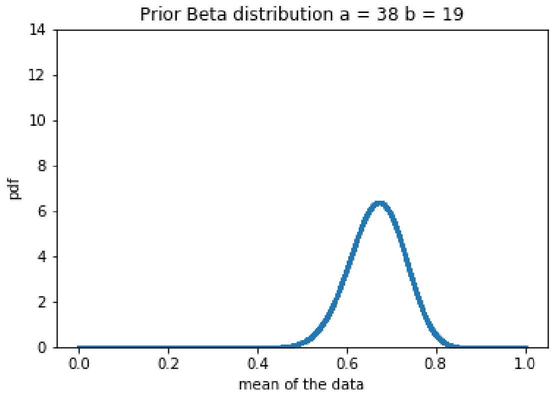
Using concurrent learning approach we get the final mean as **0.7058823529411765.**

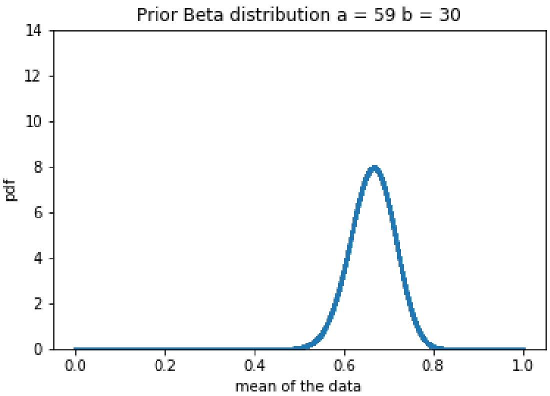
Part A – Sequential Learning

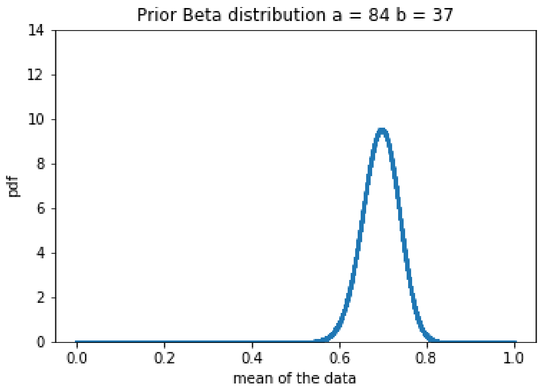


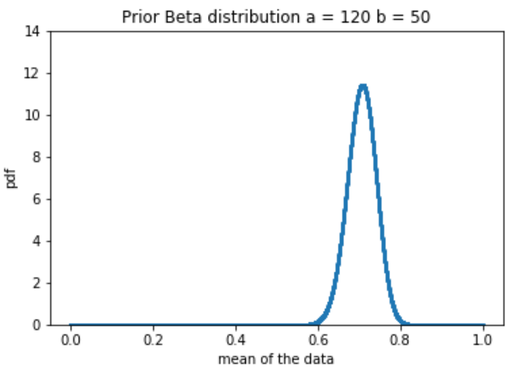




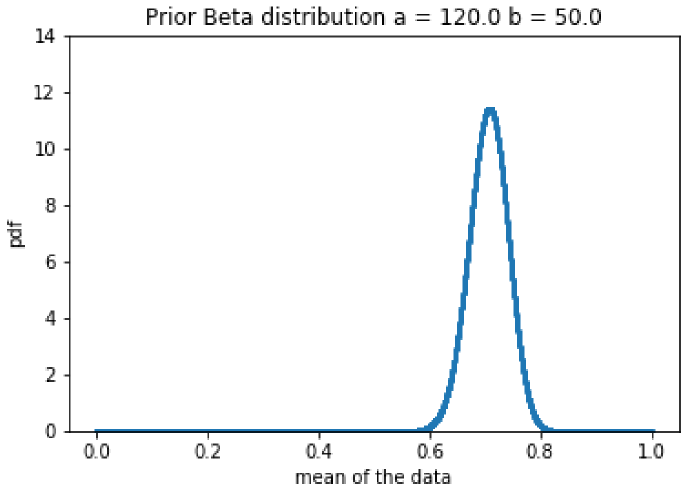






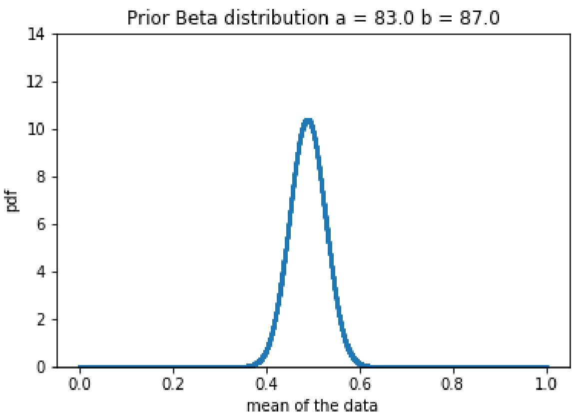


Part B – Concurrent Learning



Conclusions –

- Both sequential and batch/concurrent learning gave same predictions for the number of heads. But sequential learning takes more time to calculate likelihood of each point individually, whereas in batch/concurrent learning likelihood is calculated once and the final posterior distribution is obtained. Sequential approach is used when a lot of data has to be processed.
- As the number of data points is increased the error in predicting the number of heads decreases in both cases. By adding more data points the mean of the posterior distribution shifts towards μ_{ML} of the data set and the variance decreases gradually.
- If $\mu_{ML} = 0.5$ the final distribution approximately becomes a Gaussian distribution with mean = 0.5 and the Central Limit Theorem will hold. This also means that coin will behave like an ideal coin.



- The main reason behind choosing beta function as our priori is that it is proportional to powers of μ and $(1 - \mu)$ just like our Likelihood function. Hence we obtain conjugacy so the posterior distribution will have the same functional form as of priori i.e. Beta distribution. This can't be achieved with other distributions such as Gamma, Gaussian or Pareto distribution.

Team Details –

- **Parth Krishna Sharma – 2017B3A70907H**
- **Siddhi Burse – 2017B3A70972H**
- **Maurya Grover – 2017B3A71433H**