Foundations of Data Science
(CS F320)
Assignment 3

Team Members :

1. **Shreeya Bharat Neelekar - 2017A7PS0093H**
2. **Shriya Choudhary - 2017AAPS0409H**
3. **Prateek Agarwal - 2017A7PS0075H**

Introduction:

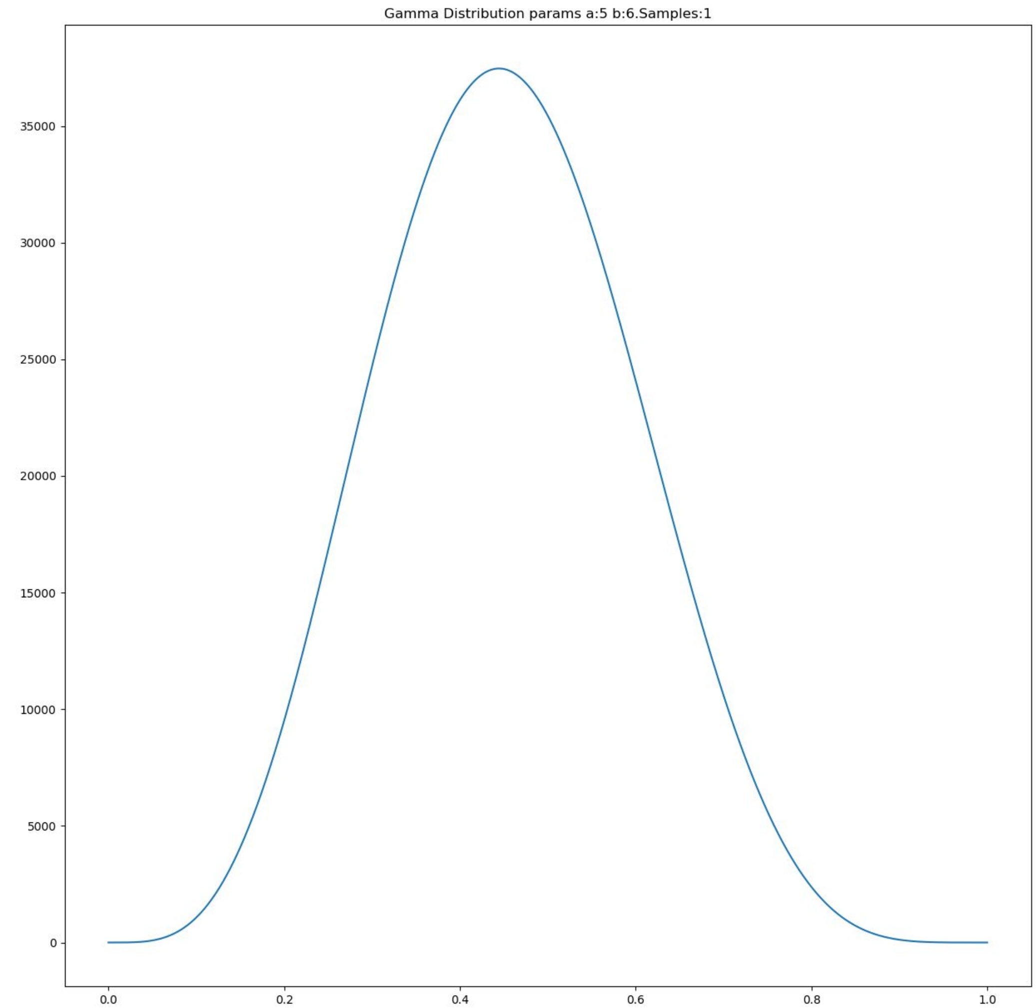
For this assignment an experiment based on tossing coins and predicting the distribution of the probability of getting heads is conducted.The dataset has 160 tosses with 70% of the tosses being heads and rest being tails, i.e. μ = 0.7.. The data assumes Bernoulli's distribution and we have taken our prior distribution to be that of the gamma function. Predicting is then done in two different methods i.e.:

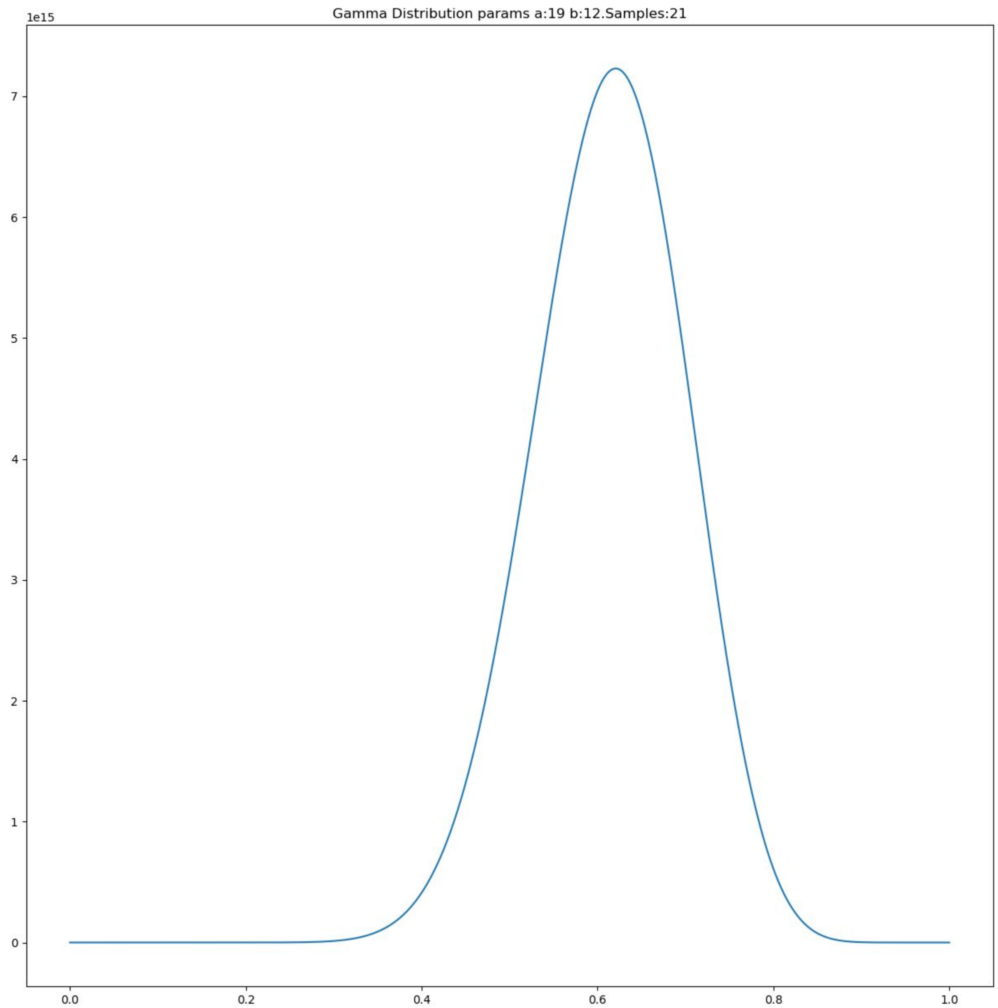
1. Sequential learning
2. Complete data learning

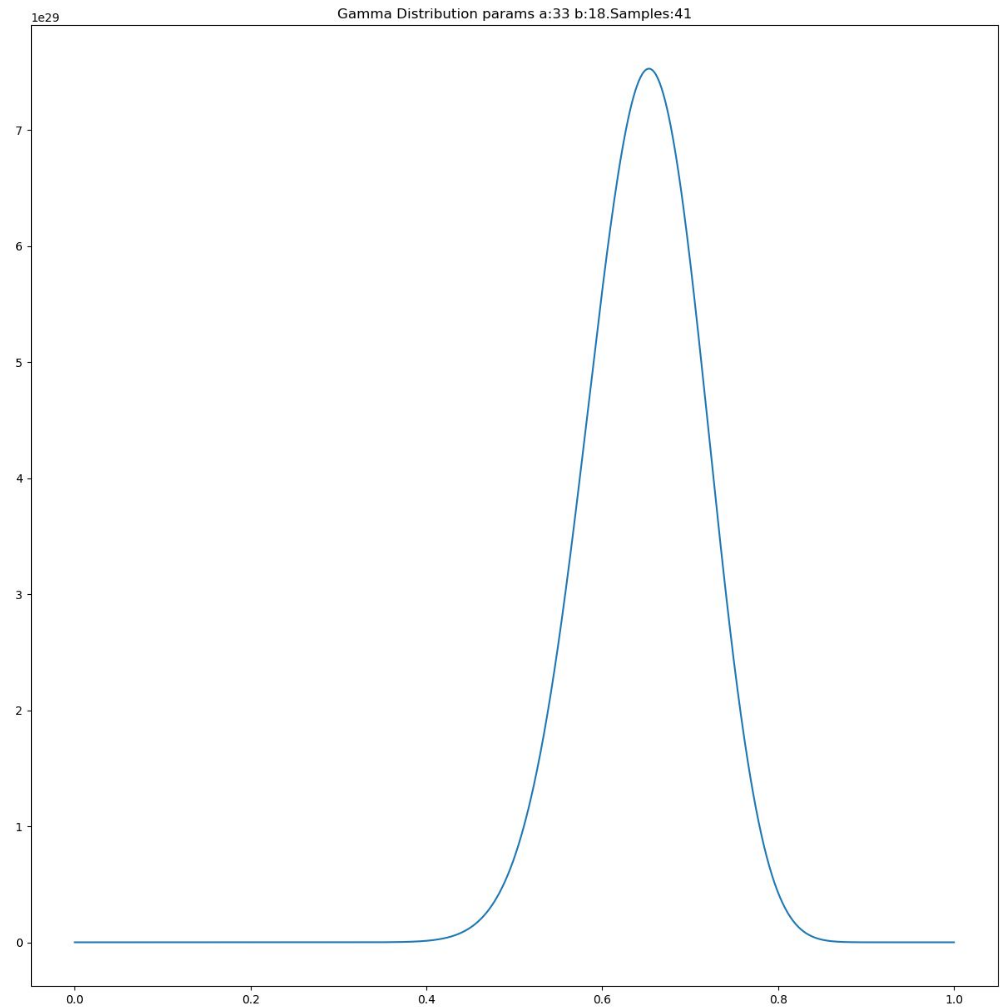
The results are shown as follows.

Part A (Sequential Learning)

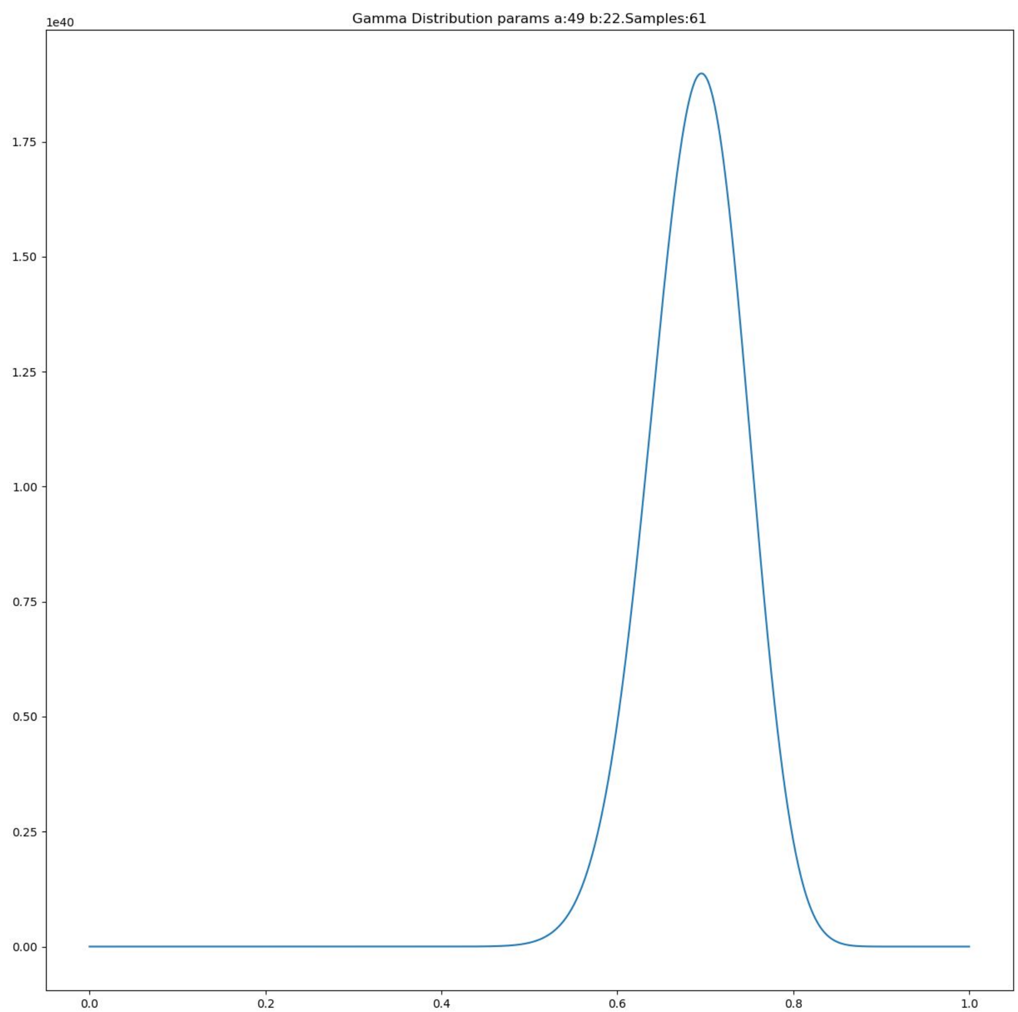
The μML chosen is 0.7. The distribution is seen to be converging to having a mean at 0.7 from its prior mean which starts at 0.4.

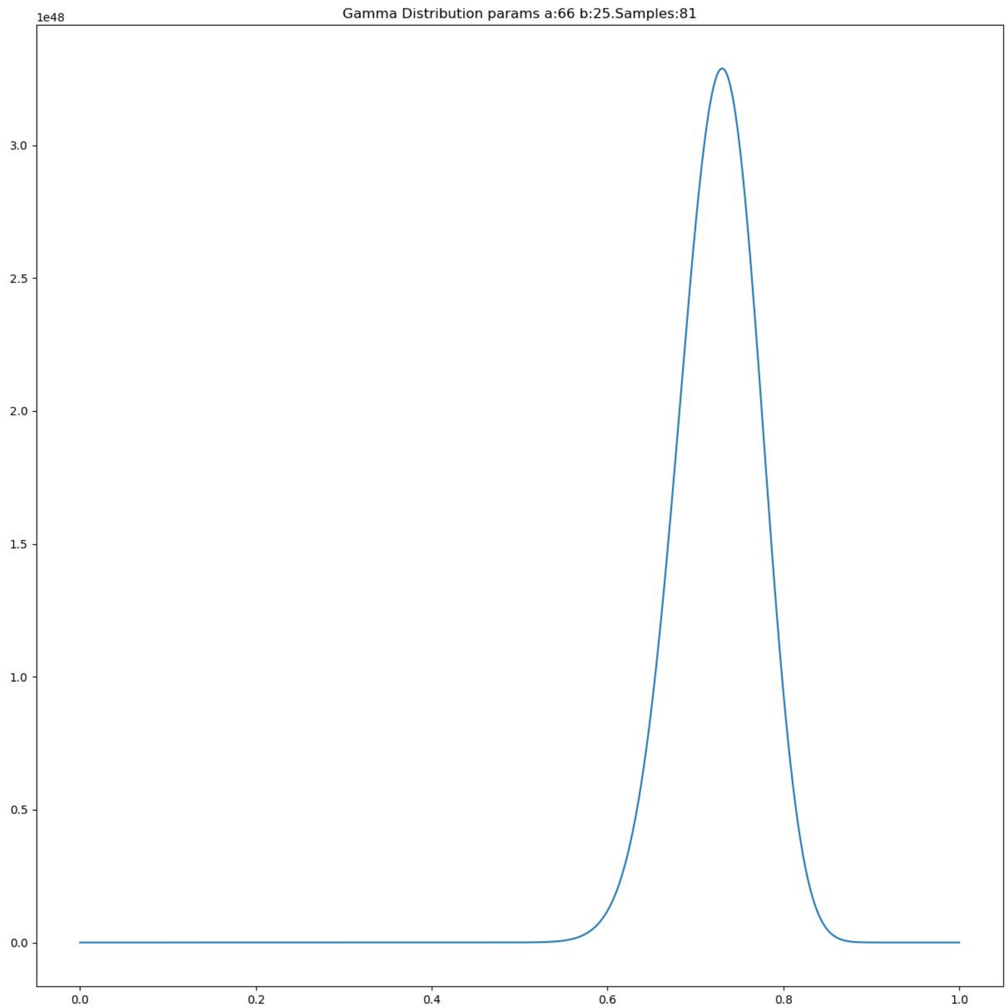

Gamma Distribution params a:5 b:6.Samples:1
35000
30000
25000
20000
15000
10000
5000
0
0.0
0.2
0.4
0.6
0.8
1.0

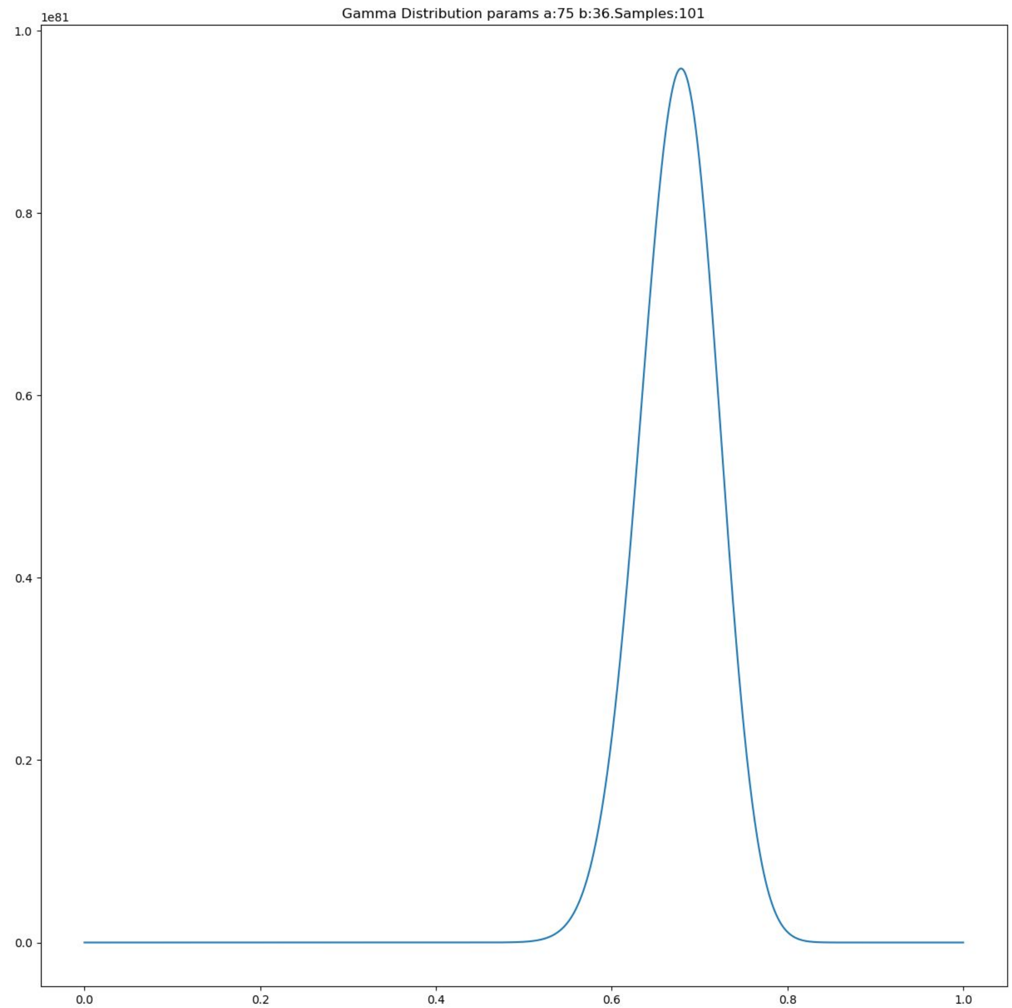

Gamma Distribution params a:19 b:12.Samples:21
1e15
0
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
0.8
1.0



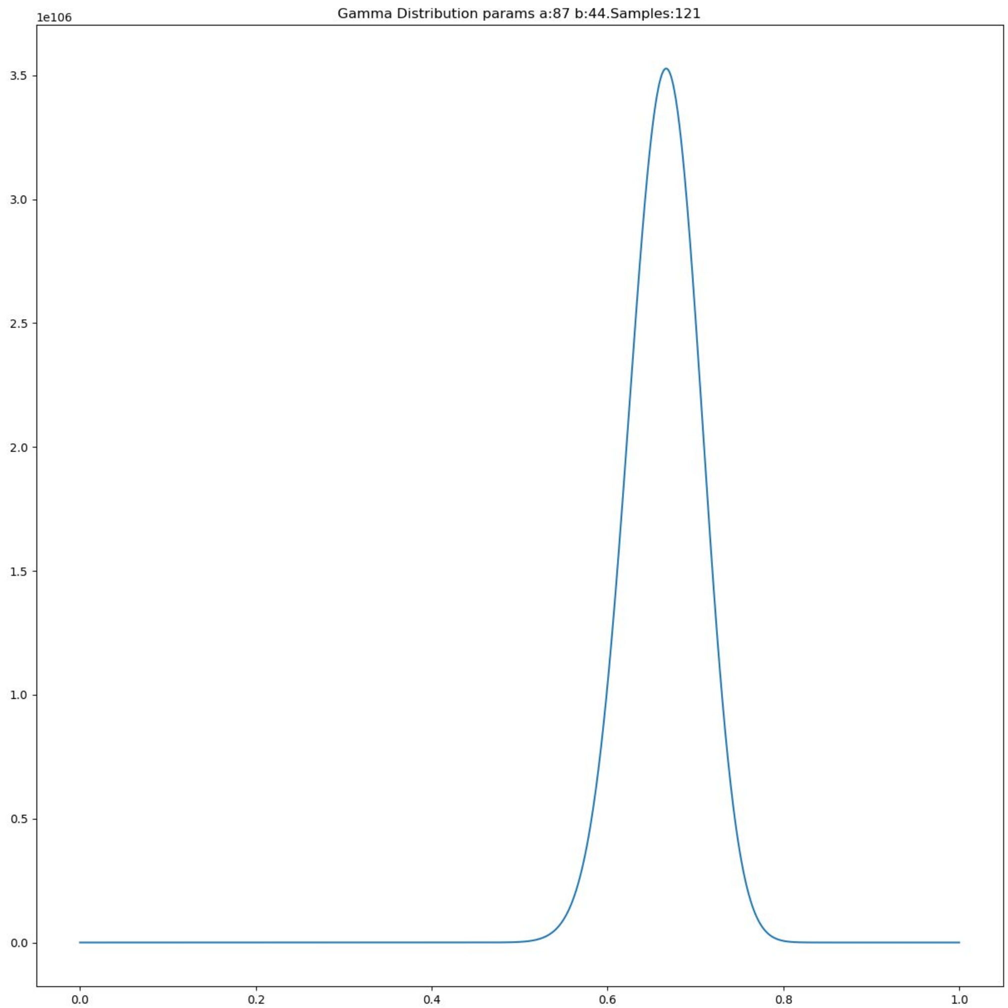
Gamma Distribution params a:33 b:18.Samples:41
1e29
0
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
0.8
1.0

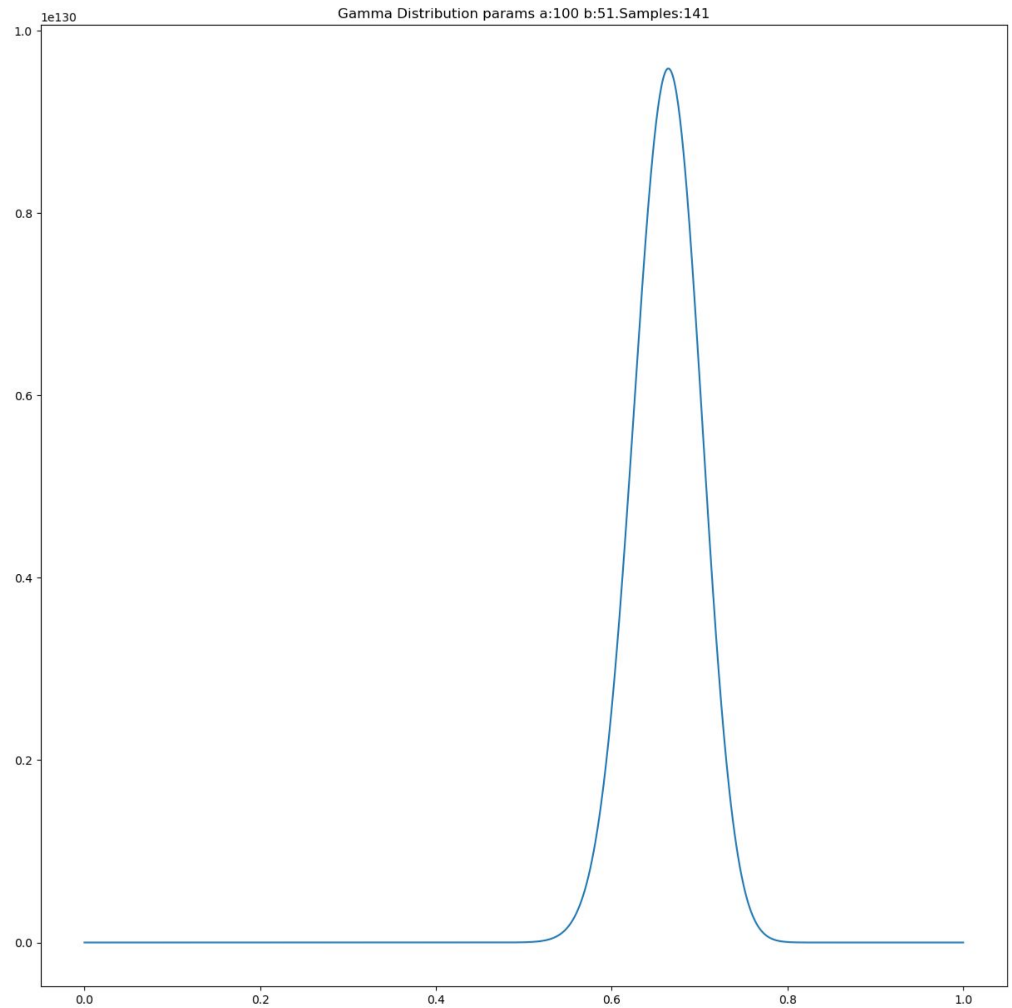

Gamma Distribution params a:49 b:22.Samples:61
1e40
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

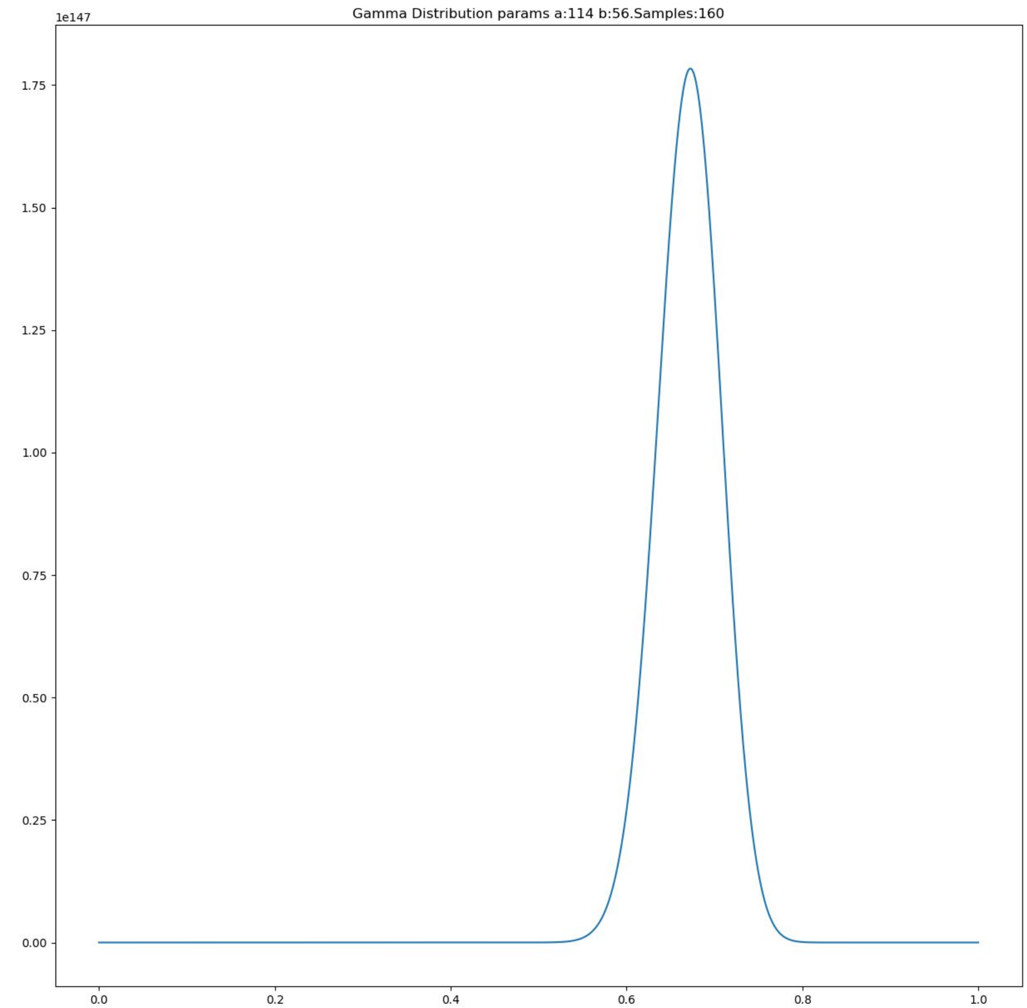

Gamma Distribution params a:66 b:25.Samples:81
1e48
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0



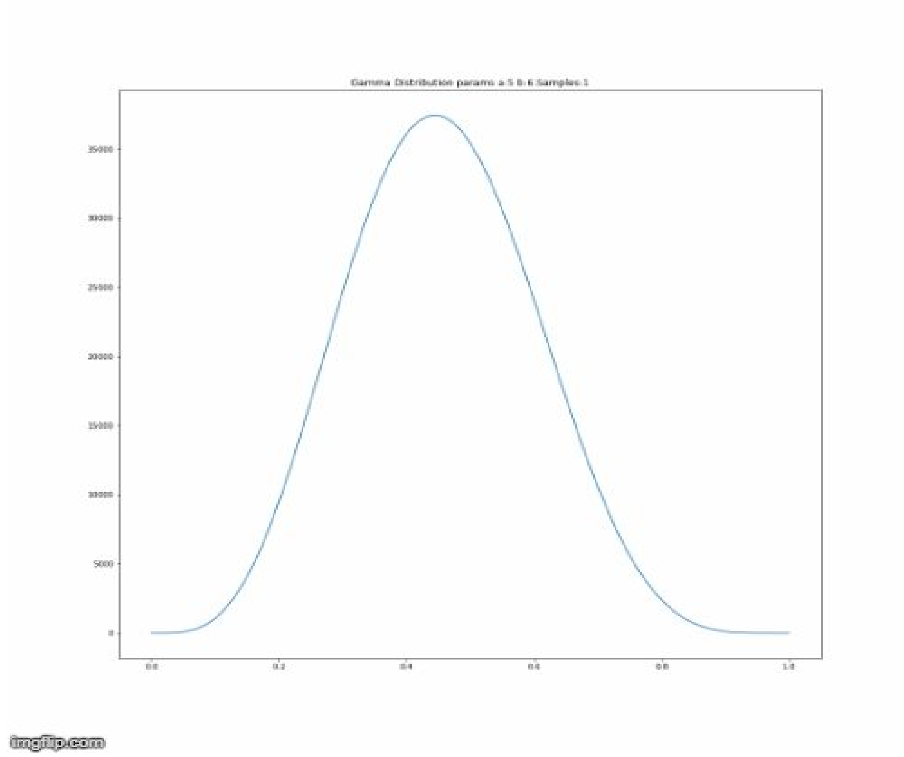
Gamma Distribution params a:75 b:36.Samples:101
1e81
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


Gamma Distribution params a:87 b:44.Samples:121
1e106
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0


Gamma Distribution params a:100 b:51.Samples:141
1e130
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

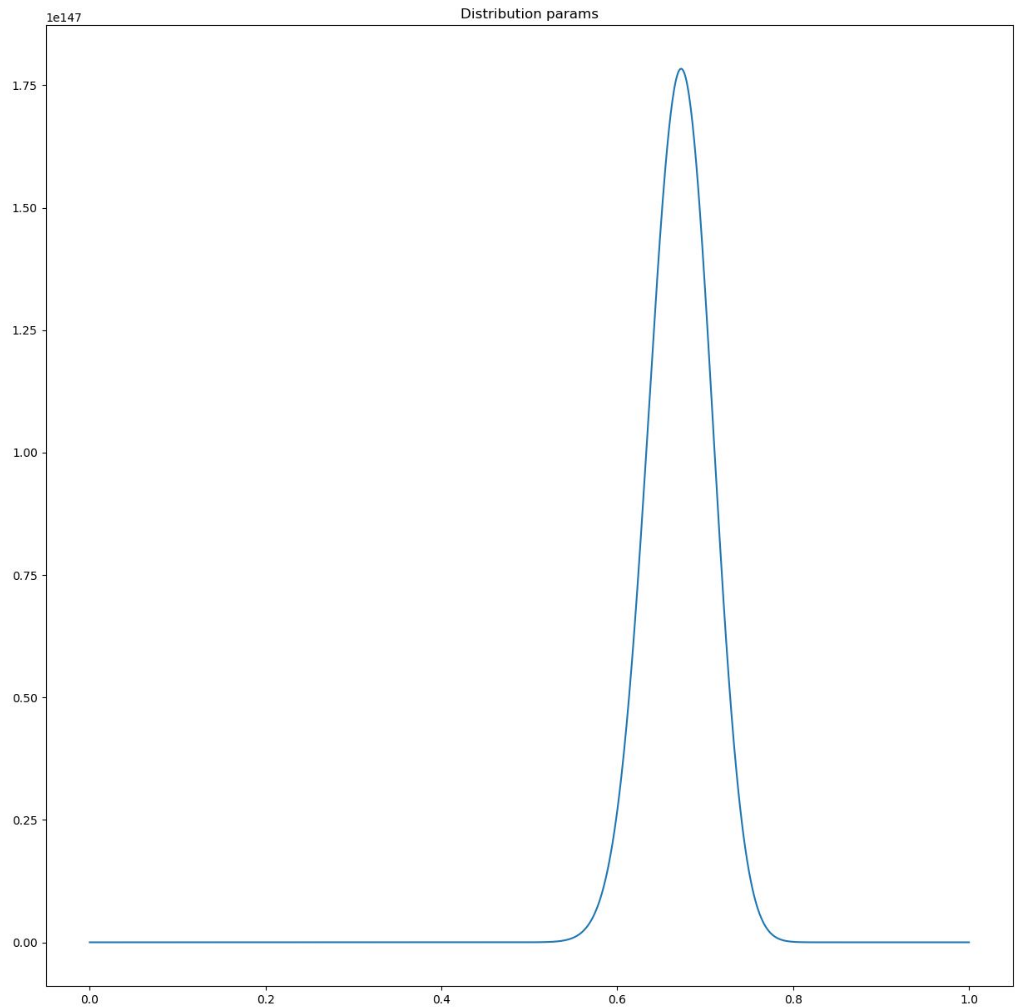


All the plots combined together :



Part B

The posterior distribution after viewing the entire dataset at once :



Part C

Both the above two models give the same final output as at the end, both models consider all 160 data points.

If the no. of points is increased, the posterior distribution graph becomes more sharply peaked. It is a general property that he Bayesian and maximum likelihood results will agree in the limit of an infinitely large dataset. So, as the dataset becomes infinitely large, the result reduces to the maximum likelihood result. For a finite data set, the posterior mean for μ always lies between the prior mean and the maximum likelihood estimate for μ.

The posterior distribution if $\mu ML = 0.5$ would be :