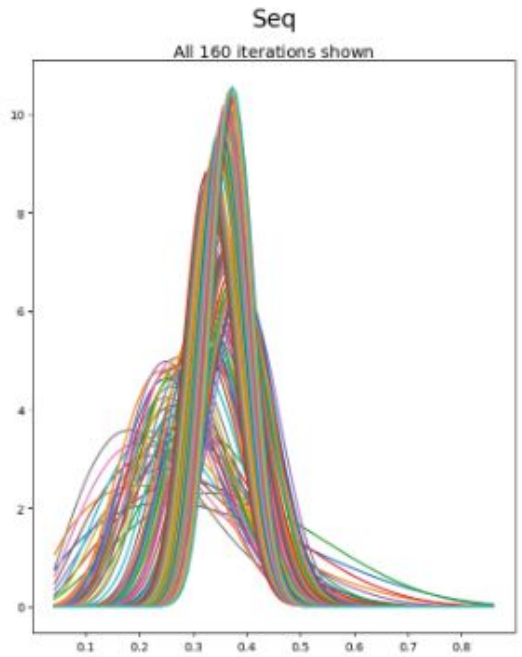
Assignment 3

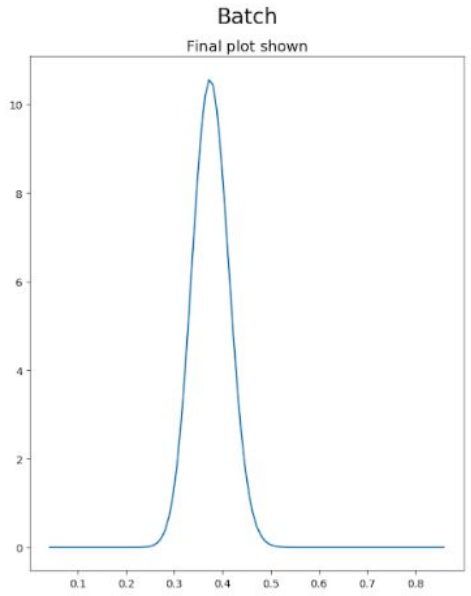
INTRODUCTION

DATASET and PARAMETERS

- In this assignment our prior probability follows a beta distribution and has parameter values a=2 and b=3. Using these parameters we get a mean of 0.4. The dataset consisting of 160 entries has 60 instances of zeroes corresponding to tails and 100 instances of ones corresponding to heads. These entries are shuffled using numpy's shuffling method. The maximum likelihood estimator of mean is therefore equal to 0.375.
- For the prior beta distribution, we have used a=2 and b=3 so that the prior mean=a/(a+b)= 0.4. For generation of the dataset in order to ensure the maximum likelihood estimator of mean is not in the range (0.4,0.6), we added 53 ones(representing heads) and 107 zeros(representing tails) which makes the estimator of mean = 53/160=0.33125 and then shuffled the data so that there's a random distribution of ones and zeros.

Plots for Part A and B)





Part C)

- The sequential approach incorporates a Bayesian viewpoint. It is independent of the choice of prior and of the likelihood function. It records observations 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 handy for 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 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= 55 and b=110 which gives a mean value equal to 0.33333.

- It can be observed that the posterior mean lies between the likelihood mean(0.375) and prior mean(0.4) which is expected for a finite dataset

.

- Both the approaches is that the final posterior distribution for both the approaches is same, i.e., with parameters a=55 and b=110.

- With the addition of data points , parameter values of a and b will be increased in relation to the number of ones and zeros respectively. The curve will become more sharply peaked, due to the expression for variance of the beta distribution.

- In other words we can say that the variance decreases with an increase in the parameter values. Having a very large dataset implies that variance almost tends to zero. Therefore the hurdle we face here is as we get a larger dataset, we are more certain of our prediction and hence the variance decreases.
- When the mean estimated using the maximum likelihood estimator is kept constant at 0.5. The mean of the posterior function will shift rightwards and an approximate estimation of its value will be in the range of 0.4 to 0.5, as it was previously in the range from 0.375 and 0.4
- The prior mean is 0.4 while the likelihood mean is changed from 0.375 to 0.5. For our dataset consisting of 160 entries the likelihood of 0.5 will imply equal number of ones and zeros and the posterior mean will attain a value equal to 0.497.