Advanced Statistical Methods: Assignment Solution

Problem: 1

In the first problem I need to find the effectiveness of new drug that is applied to the patients. Total patients is 38 and effectiveness of the drug benefit of the patients is 32. Plot the effectiveness of the benefited patients with the function of **dunif(),** where the benefited people is shown in orange in color and non-benefited people are in red color. This is the plot to understand the effectiveness of the drug.

Effectiveness of the new drug is computed over the new examination of Normal Distribution using the Prior experiment with the new drug of values. Prior experiment with the new drug of values of Distribution is the probability. It will give some insightful ideas and understanding about the data distribution and how it's work. Figure 2 will show the Normal Prior experiment with the new drug of values of Distribution with condition over 42 patients. To understand the effectiveness plotting a polygon where the orange section in figure 1 is the benefit people and red color at the right hand for the non-benefited people. Where the benefited person is 84.21 % and 15.78 % patients are not benefited.

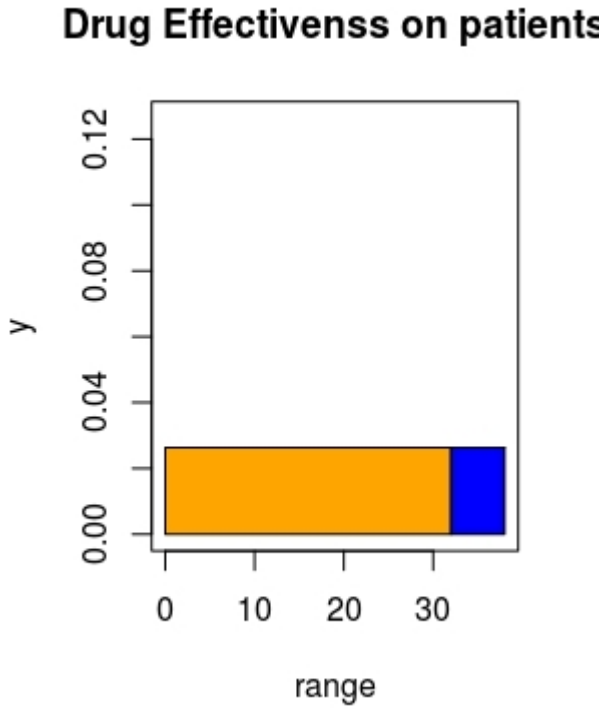


Figure Details 1: In the above figure it is showing the effectiveness of the benefited people out of 42 people. Where the green color is the benefited and red color is the not-benefited people.

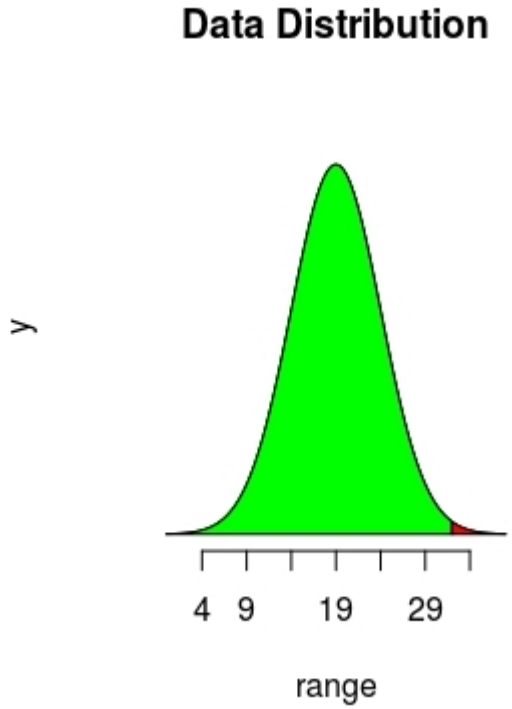


Figure Details 2: Normal Prior experiment with the new drug of values of Distribution of the given data as condition that 32 people will be get definit over 38 patients.

When a new data set is computed over the model of Prior experiment the data distribution is notice that 39 patients were benefited over 44 patients. So first to understand the data Distribution plot it in Figure 3. Benifited percentage of second drug over first one is: 4.425837 % it is computed form examine data. In the second drug the percentage of benefited patients are 88.63% and non-benefited patients are 11.63%.
The creditable interval difference for the Prior estimation of values of and posterior Distribution for 95% confident interval is 28.72427 % which is
calcutated from the normal distribution.

In the right hand side Figure 3: it's the Posterior Distribution over 39/44 patients
were benefited from the drug. It is the posterior Distribution of the given condition. X axis of the plot is the range of the data from 1 to 44 and after 39 in red in color which means that 5 people did not get benefited form the drug.

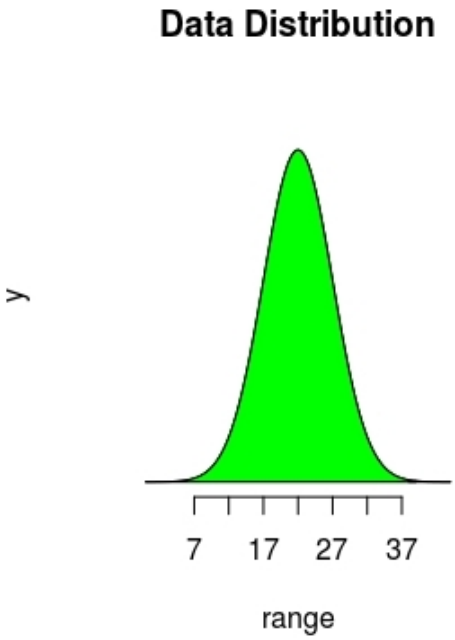


Figure Details 3: Posterior Distribution of the given condition where 39 patients get benefited from the applied new drug.

Problem 2:

In the second question data of 1000 people record are there who consume alcohol those record were collected as pd (Probability of taking alcohol) alcohol is consume over the weeks. The given data has only one data frame of name pd. Data frame values are in type of 'double'.

To plot and estimate Prior estimation of values of Distribution with Bayes Distribution using a MLE (Maximum Likelihood Estimation) function. To do this things first need to calculate the mean, standard deviation and variance of the distributed of the given data set and define a function to calculate the Beta Distribution **dbeta()** which is need to calculate for Prior estimation of values of Distribution. To find the value of alpha and beta from the Beta Distribution define function as name **Betaparam()** which will calculate the suitable values of alpha and beta. Alpha and beta value will be obtain from the function, where the function will accept two argument value of mean and variance of the data set. After computing this I got the alpha value is 0.066 and beta value is 0.13 got from the **Betaparam()** function.
Now plot the Prior estimation of values for distribution with the alpha and beta value using function dbeta() (beta Distribution function). Where the dbeta() function uses three values of distributed data , alpha and beta values.

Now plot a Prior estimation of values of Distribution of dbeta() function where x axis will represent the alpha and beta values and y axis is the density Distribution of the beta Distribution using MLE (Maximum likelihood estimation). Detail Figure Details is attached at bellow in figure 4.

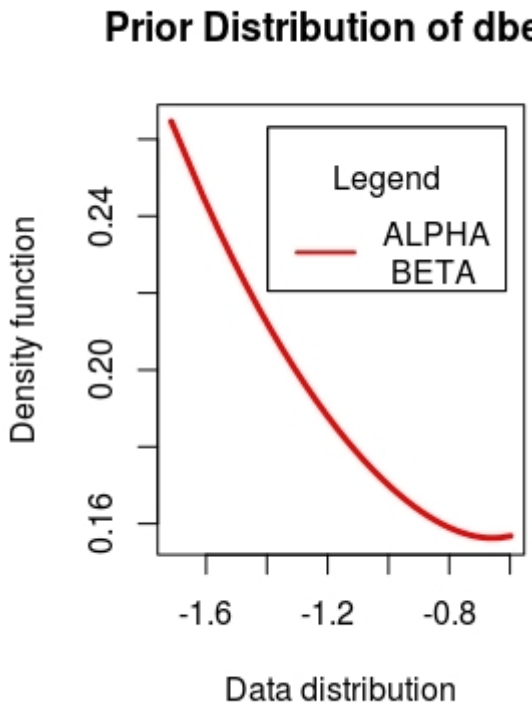


Figure Details 4: Prior estimation using Beta Distribution of MLE. Where x axis is the Distribution for alpha and beta values and y axis for Beta Distribution density.

Now need to make a Prior estimation of values of Distribution taking report from above that what will the probability for 13/20 successor record in the survey of pd data. To make Prior estimation of values of and make this report need to import library of **'klaR'** and **'LearnBayes'**. Make the Prior estimation of values of distribution using the values of alpha and beta which is calculated in the above of the 2.a problem and like the same use successor and total number of observation to make the posterior after that use function **triplot()** to plot Prior estimation of values of, likelihood and posterior distribute in the same plot. **Trplot()** function will take only two argument as a value of prior and posterior distribution values after that this function will automatically plot prior, maximum likelihood and posterior distribution. This function is very efficient that by given prior and posterior distributed value you can draw the maximum likelihood too over a single data distribution. To know more about **triplot()** function you can search for it in console as **?triplot()**. Before using of the function when you will define the value of the posterior distribution that time you will pass the successor and the total event that occurred. It will make sense to the distribution that what will be the posterior of the computed data.

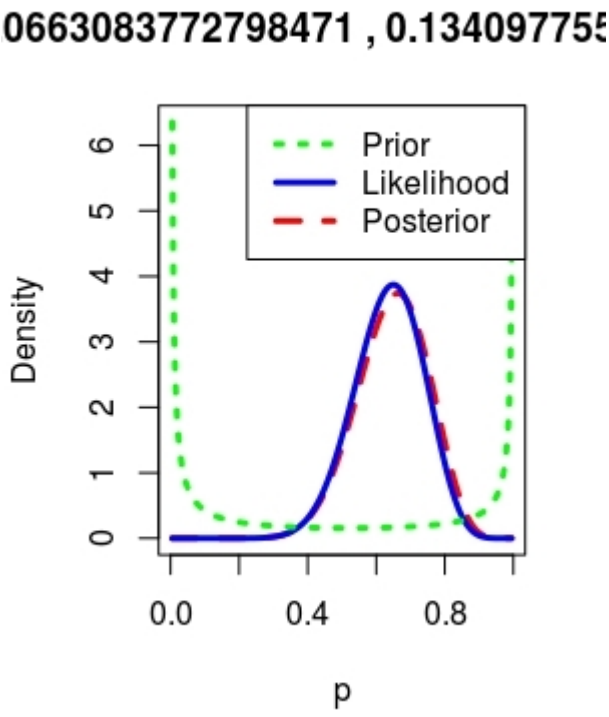


Figure Details 5: Prior estimation of values, posterior and maximum likelihood estimation Distribution for the 13 people who consume drug out of 20.

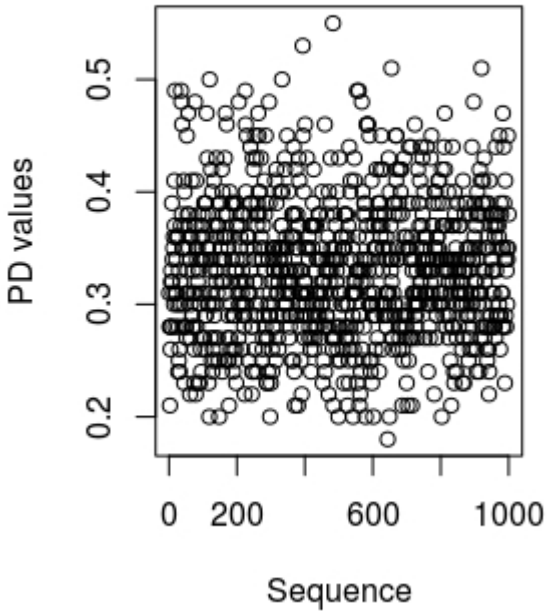


Figure Details 5.1: Visualization of the row data as the condition is given in the question.

When a person record is outdated before 20 of completion, on that record if 3 week he consume alcohol over 4 weeks. To find the MAP and Maximum Likelihood Estimation from part a. To calculate the difference and of alcohol consumption probability and the mean of the distributed data of pd of the given data set. Plot the data using Maximum Likelihood Estimation using the function of **likelihood()** and find the top most height of the data.

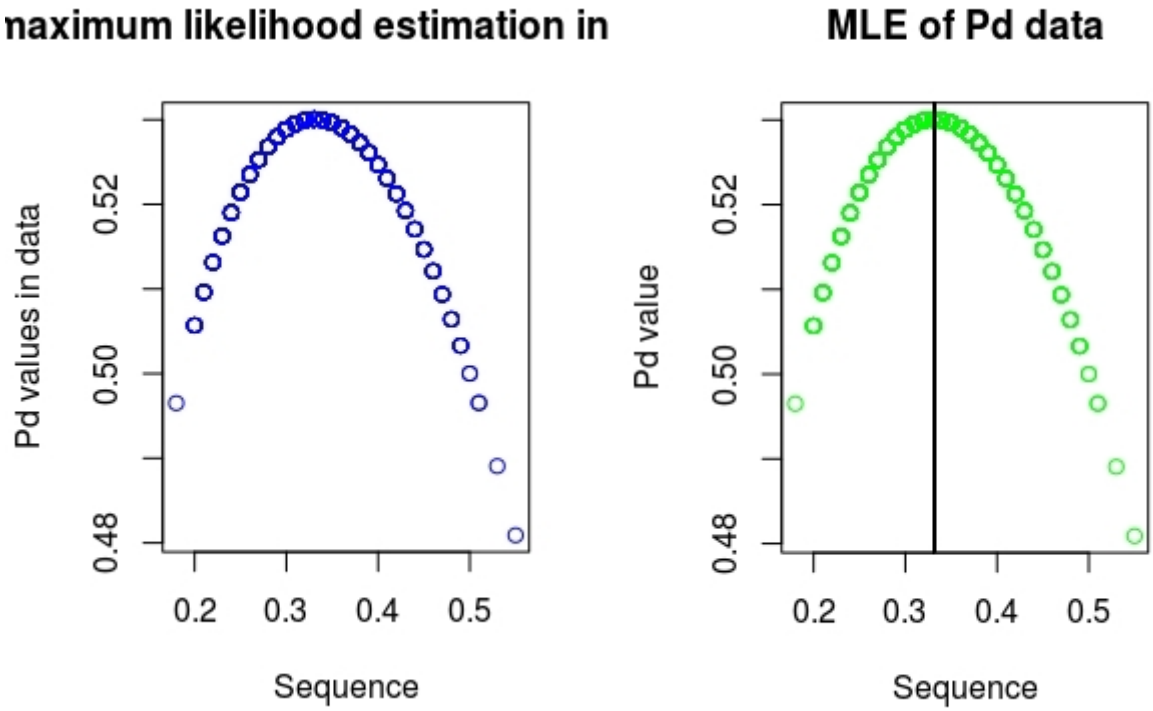


Figure Details: 6 This is the figure of MLE without using the Prior Estimation. Where the maximum Distribution at found at 0.33.

In the first case the outcome 3 out of 4 where the probability is 0.41 and in the Maximum Likelihood Estimation the obtain probability is 0.48.
To calculate the maximum likelihood estimation need to define three argument first those are length , mean and the data which is in distributed form. After that to find the peak distribution or maximum distribution of a particular point I am making a vertical plot in the maximum peak points or at maximum distributed location. To compute that find the division of mean of the data and length of the distributed data.

Problem: 3

Arrow throw data with condition that standard distribution for both of data will be 2 and I am taking those data are distributed as a mean of 0. Where the range of x and y axis value is given and radius is 5. Now with this data we will do computation Prior estimation.

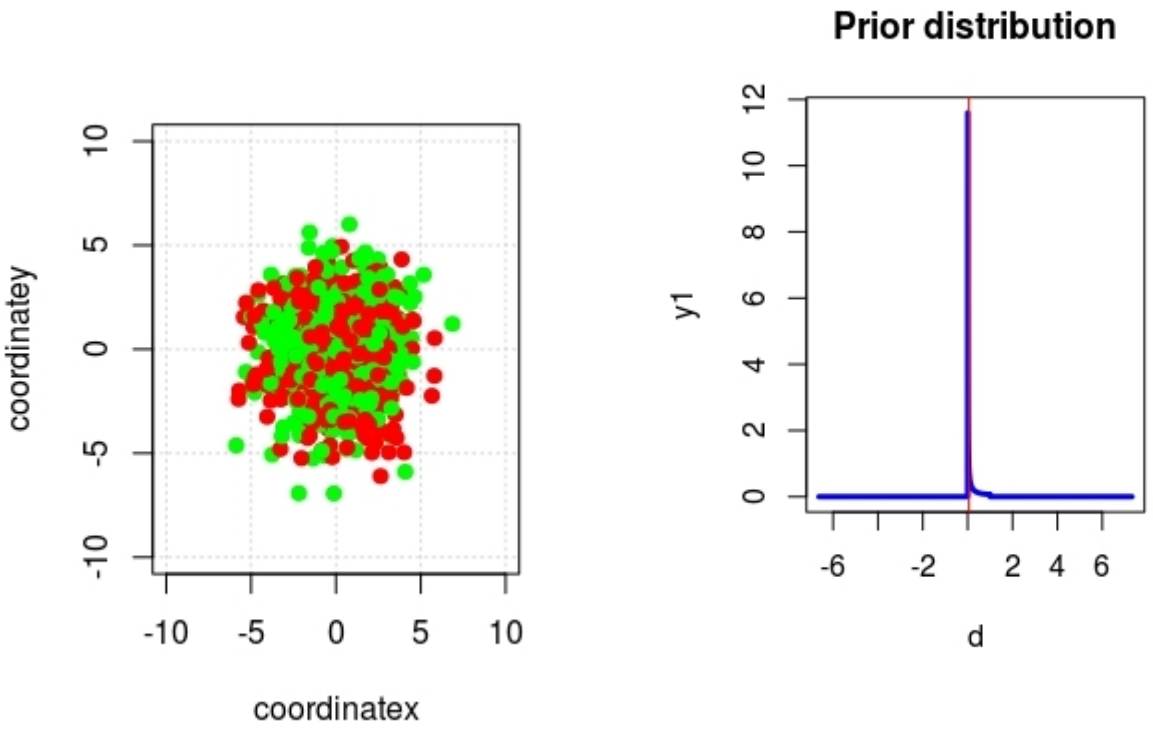


Figure Details 7: Left hand side plot is the visualization of the given conditional data. This plot is the Prior estimation of values of Distribution of the data. Where d is value of standard deviation 2 and y1 is the y coordinate value sd= 2 and for both case mean= 0.

Prior estimation of values for distribution of the data with the given condition after that we are merge both data data for x axis and y axis to make a single data set of data distribution. Where the maximum distribute is at 0.056 in the x axis over the y coordinate which is computed using the beta distribution form which is describe in above section.
Now, Maximum likelihood estimate for x=2.35 and y=0.95 as given in the question.

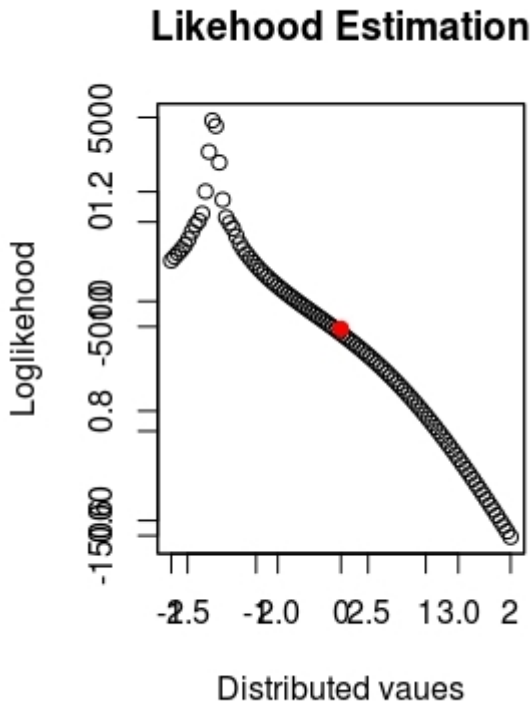
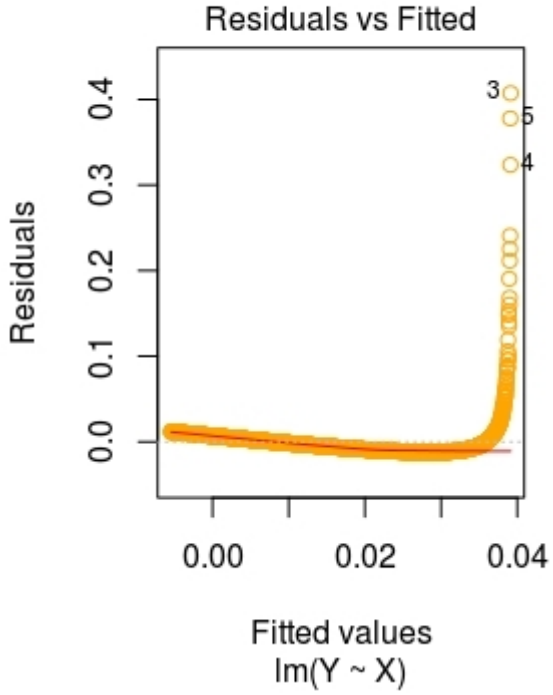


Figure Details 8: Plot of maximum likelihood estimation and the point over the plot in red in color to visualize it, whose coordinate values for x axis is 2.35 and y axis coordinate value is 0.95.

Now to compute the posterior Distribution values from the first section for this problem, need to compute first calculate using a beta Distribution which will useful to find the value of alpha and beta values. Now call the **triplot()** function to plot the posterior distribution and by simulation the posterior we found the mean of x axis as **mu(x)= posterior[1]** which is equal to 0.058 and mean of y axis as **mu(y)= posterior[2]** which is equal to 0.055

Now for the last case for 9 observation where x mean is 0.43 and y mean is 0.09 as given in the question.



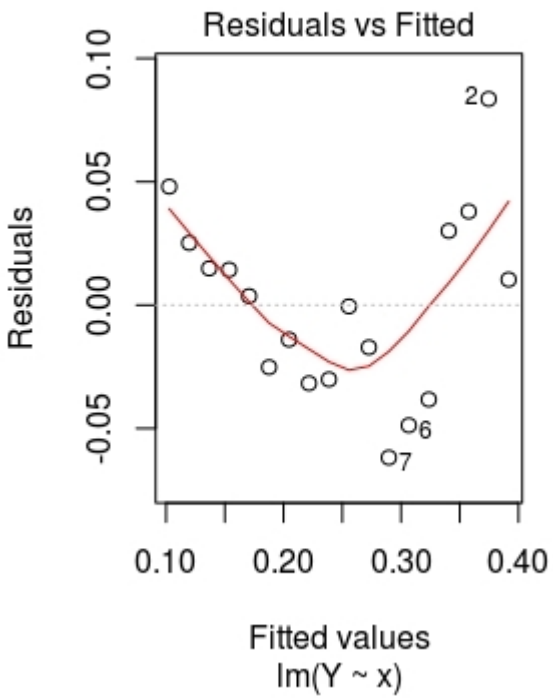


Figure Details 9: Visualization of the data for 9 observation. Where **mean(x)=0.43** and **mean(y)=0.09.**

Bellow Figure Details is the likelihood estimation and the data points visualization on part of the question(ii).

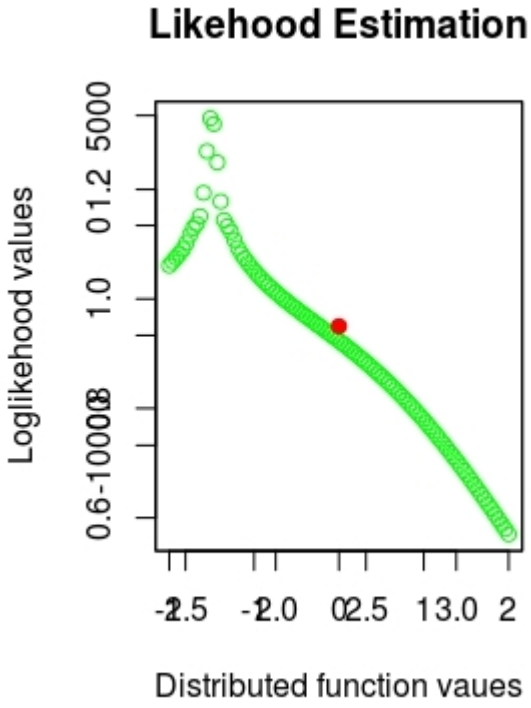


Figure Details: 10 Maximum Likelihood Estimation and the data points in red color which is mention in question section of ii. Where MEL given the value is -1.1515 and the maximum lambda value is 22 on likelihood of y axis.

Problem: 4

FTSE data set is given which have two data frame of **LogReturn** and **Day** of 2000 instances. Visualization of the data and bellow the two distribution of the data are Exponential Distribution and Normal Distribution.

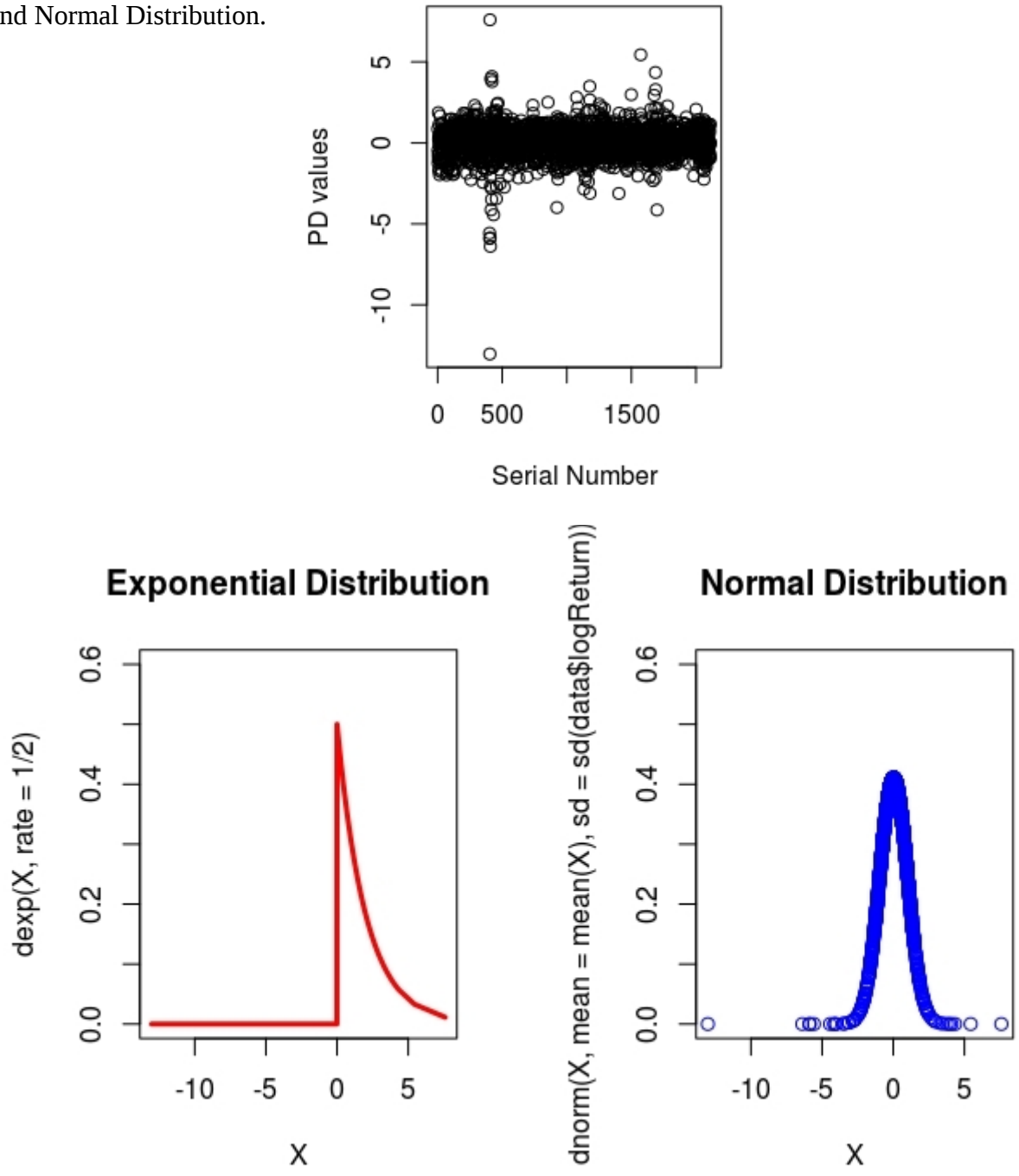
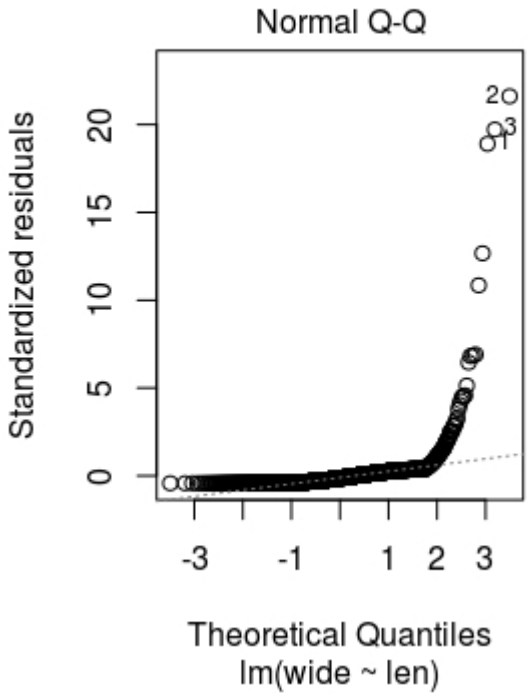


Figure Details: 11 Left hand figure is the exponential Distribution and right hand side figure is the Normal Distribution.

Plotting of two state Hidden Markov Model of Maximum Likelihood Estimation. When I used the the MLE I use the logarithm of the data to make it proper orientation and good visualization.



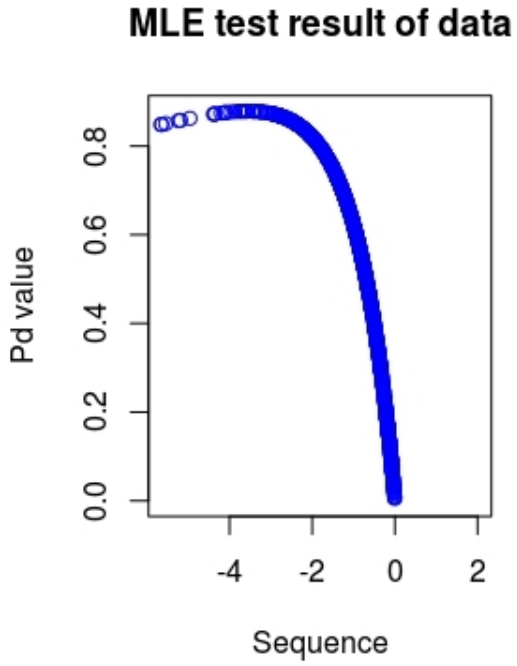


Figure Details: 12 In the plot y = probability density estimation value of the MLE and x axis is the coordinate value.

Conclusion: Both the case for Maximum Likelihood Estimation and AIC obtain value is -ve and their is significanc difference among their values. So the conclusion that large difference is there in-between AIC(-8431.79) and Maximum Likelihood Estimation (-1.51) which I examine in the given problem statement.

Posterior Distribute for the maximum mean value of problem section (b). To calculate, first need to obtain the values of Beta Distribution. Calculate max and mean from section (b). Finally Posterior distribute given the value of 0.0284 and 0.001056.
After calculating the Posterior Distribution with the big values of mean, it is notice that the AIC value in section b is 8 times bigger, where both the values is negative in nature.

Stationary Distribution (SD) is the probabilities Distribution that has no changes in the Markov Chain with respect to the time as increased. SD is represent in a row whose all the element are double in type and total summation will be 1 of the row of SD.