INTRODUCTION:

Assignment 1 project deals with the concept of discrete and continuous probability. Discrete probability distribution as the name suggests it describes the probability of occurrence of each value for distinct random variables whereas Continuous probability distribution is a distribution that describes the probability of occurrence of set of possible values which are infinite.

We are going to perform:

- Exponential Distribution: The exponential distribution is one of the widely used continuous distributions. It is often used to model the time elapsed between events
- Gamma Distribution: The Gamma distribution is widely used in engineering, science, and business, to model continuous variables that are always positive and have skewed distributions
- Normal Distribution: Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

ANALYSIS:

Problem 1: In this particular problem, random 1000 variables have to be generated and then following operations are performed over it using R

1. Relative frequency Histogram
2. Selection of Probability distribution
3. Plotting of Probability Plot
4. Chi squared test

Answer 1:

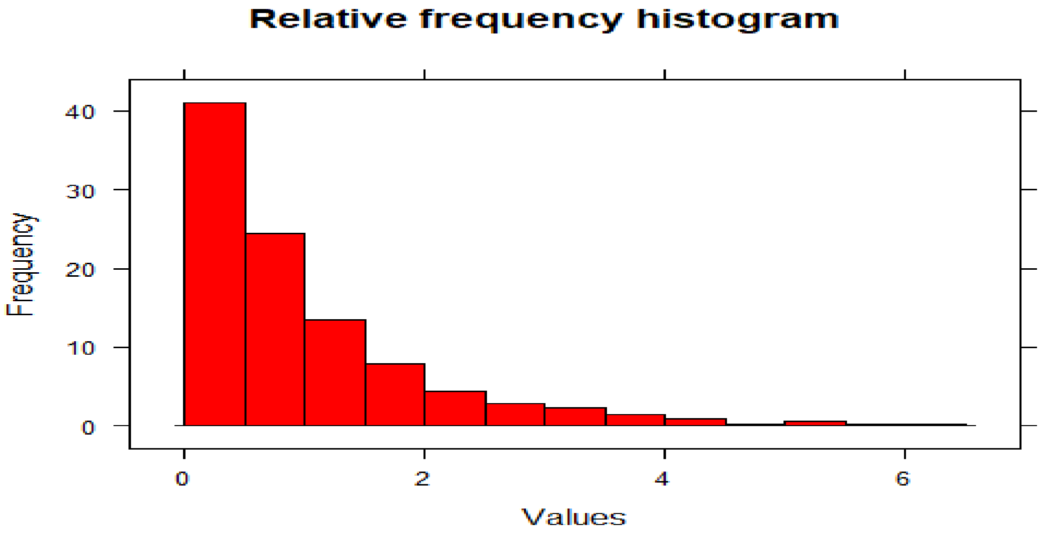
1000 Random numbers are generated between 0 and 1 using **RUNIF()** function.

The value of **X** is calculated using logarithmic function **X<- -log(r).**

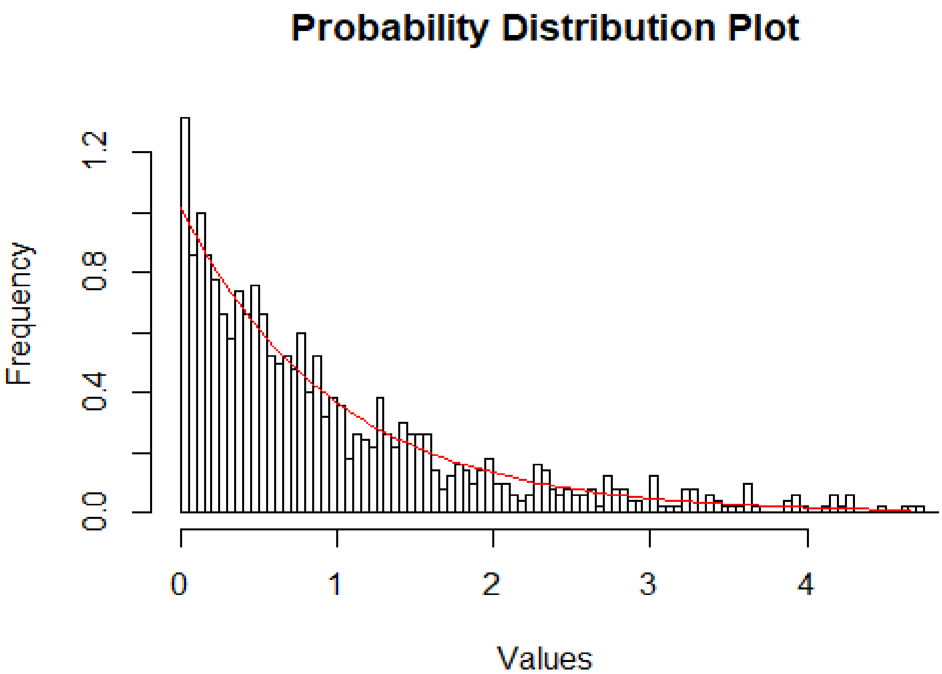
Then retrieved the descriptive statistics using **describe(X).** It describes all the statistics such as mean, median , standard deviation , minimum , maximum, range , skewness , kurtosis etc.

```
> describe(X)
   vars    n mean   sd median trimmed mad min  max range skew kurtosis   se
X1    1 1000 0.98 1.02   0.66     0.8 0.7   0 6.42  6.42  1.9     4.35 0.03
```

The **Histogram()** method from the library(lattice) is used to plot the values of X in the histogram.



Used method **fit<-fitdistr(X,"exponential")** to check whether the observed data fits the exponential distribution. Then plotted the highest and lowest limit in the distribution plot.



Chi-Square test is calculated using chisq.test(X,p=prob.exp,rescale.p = T).

```
> chisq.test(X,prob.exp,rescale.p = T)

        Pearson's Chi-squared test

data:  X and prob.exp
X-squared = 999000, df = 998001, p-value = 0.2397
```

Where, X is a numeric vector

P is the vector of probability of same length of X.

Chi square Goodness of Fit Test	
X-squared	999000
Level of Significance	0.05
Df	998001
p-value	0.2397
Null Hypothesis	Data is Exponentially Distributed
Alternate Hypothesis	Data is not Exponentially Distributed
p-value>level of significance	Data is Exponentially Distributed

<u>Problem 2:</u> In this, random 10000 values have to generated for 3 different variables, r1,r2,r3 and then the operations performed over the Problem 1 are performed.

<u>Answer 2:</u>

10000 Random numbers are generated between 0 and 1 for variables r1,r2,r3 using **RUNIF()** function.

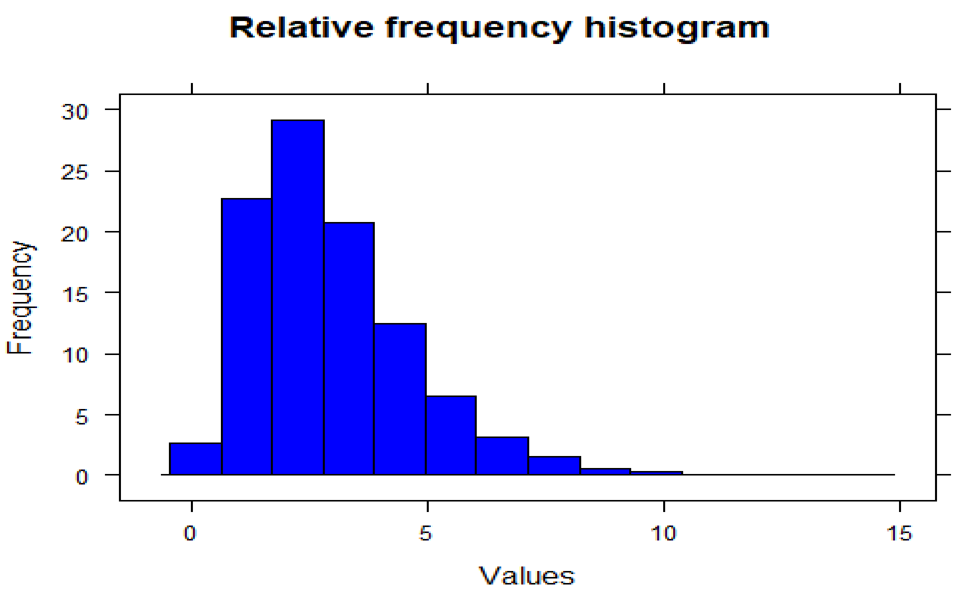
The value of **X** is calculated using logarithmic function:

X<- -log(r1), X<- -log(r2) and X<- -log(r3).

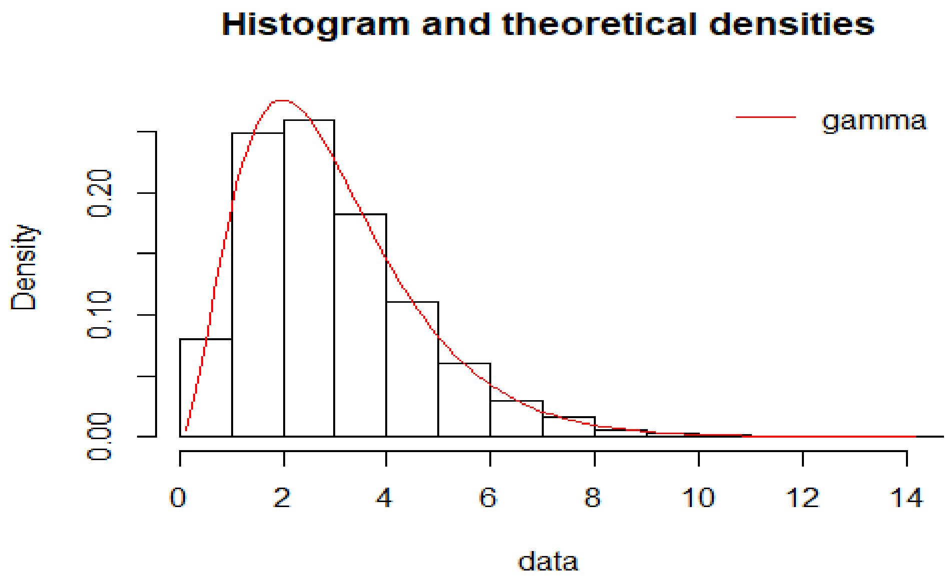
Then retrieved the descriptive statistics using **describe(X).** It describes all the statistics such as mean, median , standard deviation , minimum , maximum, range , skewness , kurtosis etc.

```
> r1<-runif(10000,min=0,max=1)
> r2<-runif(10000,min=0,max=1)
> r3<-runif(10000,min=0,max=1)
> x_p2<- -log(r1*r2*r3)
> describe(x_p2)
   vars     n mean  sd median trimmed  mad  min   max range skew kurtosis   se
X1    1 10000 2.96 1.7   2.62    2.78 1.54 0.12 14.14 14.02 1.16     1.98 0.02
```

The **Histogram()** method from the library(lattice) is used to plot the values of X in the histogram.



Used method **fit<-fitdistr(X," gamma")** from library(fitdistrplus) to check whether the observed data fits the exponential distribution. Then plotted the highest and lowest limit in the distribution plot.



Chi-Square test is calculated using chisq.test(x).

```
        Chi-squared test for given probabilities

data:  x_p2
X-squared = 9773.6, df = 9999, p-value = 0.9454
```

Chi square Goodness of Fit Test	
X-squared	9773.6
Level of Significance	0.05
Df	9999
p-value	0.9454
Null Hypothesis	Data Shows Gamma Distributed
Alternate Hypothesis	Data does not Shows Gamma Distributed
p-value>level of significance	Data Shows Gamma Distributed

Problem 3 : In the third problem, random values are generated, and certain algorithms are applied over the data.

Answer 3:

1000 Random numbers are generated between 0 and 1 for variables r1,r2 using **RUNIF()** function.

The value of **X** is calculated using logarithmic function **X<- -log(r1) and X<- -log(r2).**

Then retrieved the descriptive statistics using **describe(y).** It describes all the statistics such as mean, median , standard deviation , minimum , maximum, range , skewness , kurtosis etc.

```
> describe(y)
   vars   n  mean   sd median trimmed  mad   min  max range  skew kurtosis   se
X1    1 758 -0.03 1.02  -0.03   -0.03 1.09 -2.72 3.03  5.75 -0.01    -0.41 0.04
```

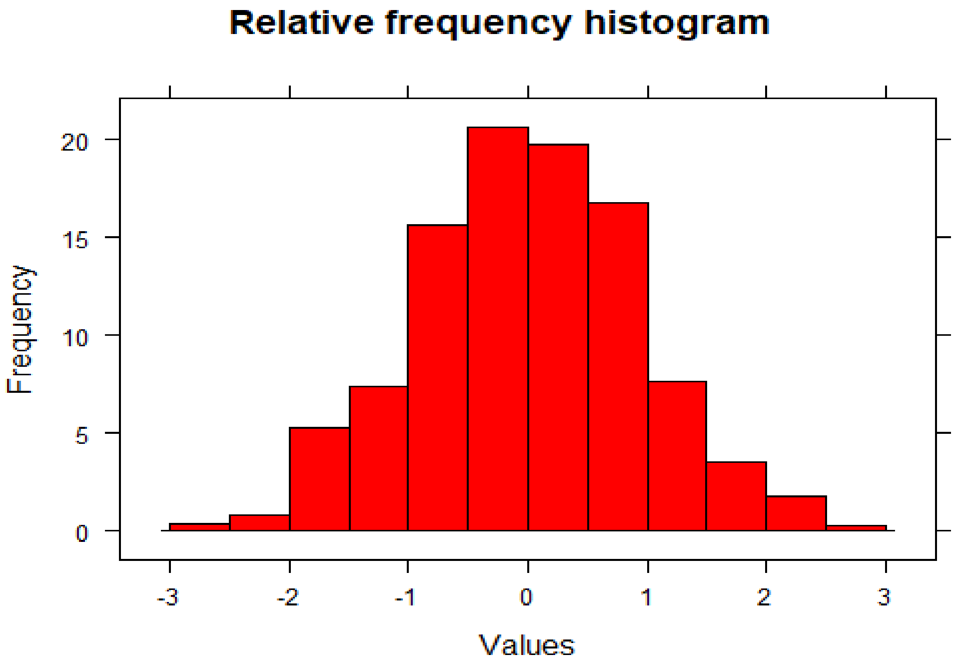
Conditions are applied over the values:

k=(x1-1)^2/2
If x2>=k
Then generate random number r3
Y=X1 Or -X1

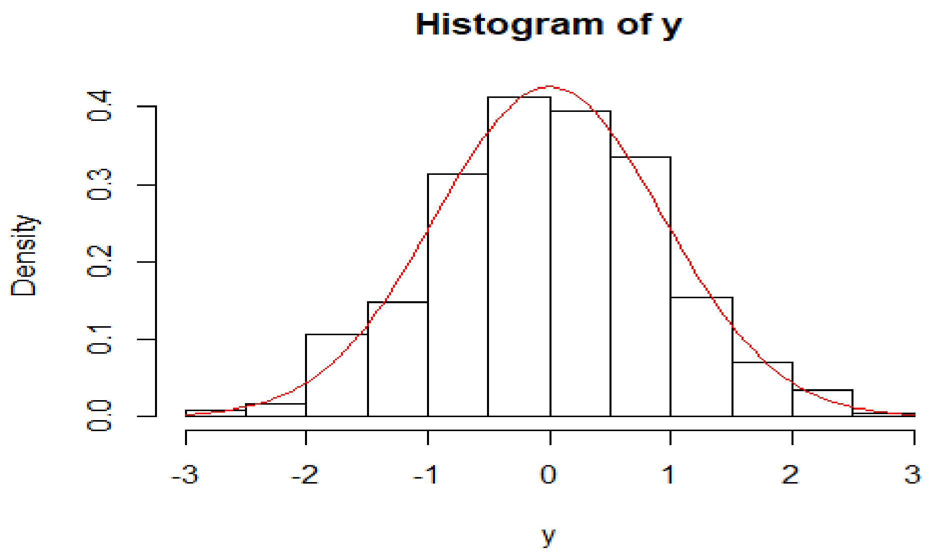
The Y value is generated by,

```
r3_p3<-runif(cnt)
if (r3_p3[length(r3_p3)]>0.5){
    y=c(y,x1_p3[i])
} else if (r3_p3[length(r3_p3)]<=0.5){
    y=c(y,-x1_p3[i])
}
```

The **Histogram()** method from the library(lattice) is used to plot the values of y in the histogram.



Used method **fit<-fitdistr(X," normal")** from library(fitdistrplus) to check whether the observed data fits the exponential distribution. Then plotted the highest and lowest limit in the distribution plot.



Chi-Square test is calculated using **pearson.test(y).**

```
> pearson.test(y)

        Pearson chi-square normality test

data:  y
P = 25.678, p-value = 0.4809
```

Pearson Chi-square Normality test	
Level of Significance	0.05
p-value	0.4809
Null Hypothesis	Data is Normally Distributed
Alternate Hypothesis	Data is not Normally Distributed
p-value>level of significance	Data is Normally Distributed

Problem 4 : In the fourth problem, iteration concept is introduced for the values that does not satisfy the condition in the Problem 3.

Answer 4:

Iteration is performed in R. Output after 1000 iterations came out as around 700.

W is calculated:

$$W=M/N$$

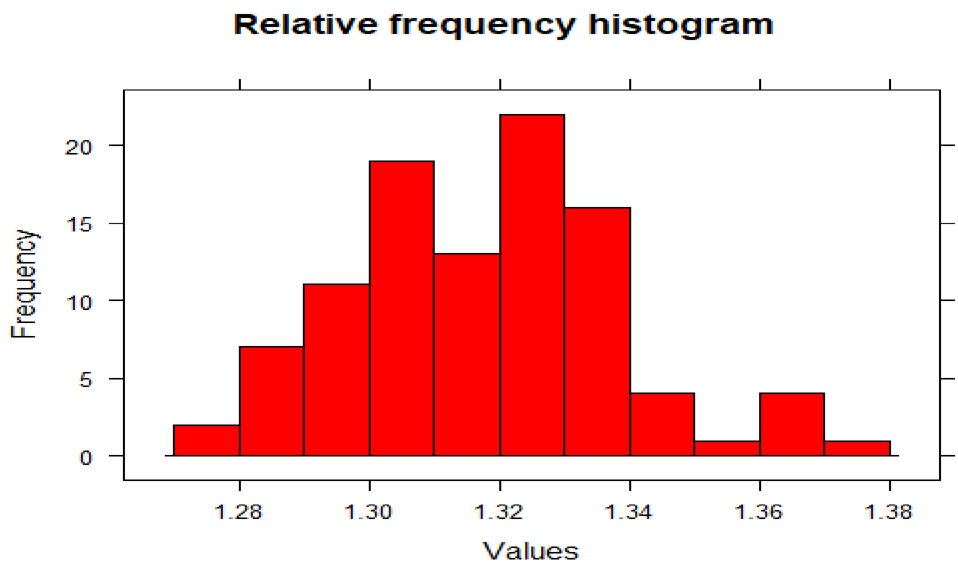
```
> m_p4<-c(10,20,30,40,50,60,70,80,90,100,200,300,400,500,600,700,800,900,1000)
> n_p4<-c()
> for (i in m_p4) {
+     n_p4<-c(n_p4,i/(i-sum(selct.y[1:i],na.rm = T)))
+ }
> n_p4
 [1] 1.1203071 0.9373801 0.9070356 0.8881703 0.8954861 0.8655452 0.8494076 0.8825998 0.9030171
[10] 0.9699521 1.0187937 1.0344172 0.9700213 0.9794831 0.9957999 1.0113037 1.0289240 1.0261232
[19] 1.0220153
```

Rejection algorithm is used for calculating counts using iteration function and count function.

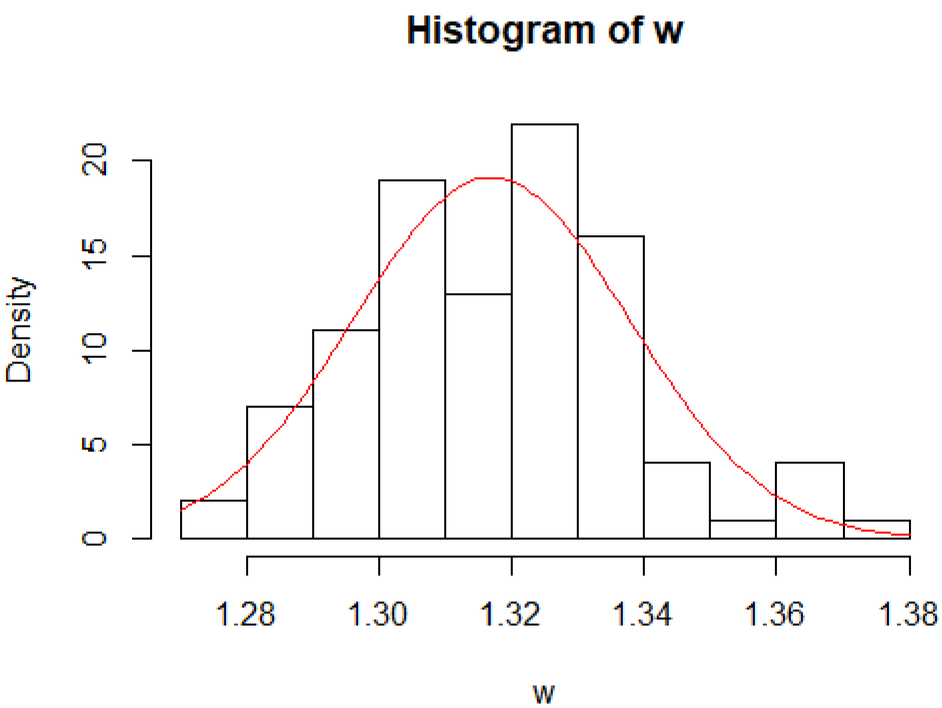
Then retrieved the descriptive statistics using **describe(w).** It describes all the statistics such as mean, median , standard deviation , minimum , maximum, range , skewness , kurtosis etc.

```
> describe(w)
   vars   n mean   sd median trimmed  mad  min  max range skew kurtosis se
X1    1 100 1.32 0.02   1.32    1.32 0.02 1.27 1.38   0.1  0.4     0.11  0
```

The **Histogram()** method from the library(lattice) is used to plot the values of w in the histogram.



Used method **fit<-fitdistr(w," normal")** from library(fitdistrplus) to check whether the observed data fits the exponential distribution. Then plotted the highest and lowest limit in the distribution plot.



Chi-Square test is calculated using pearson.test(w)

```
        Pearson chi-square normality test

data:  w
P = 8.94, p-value = 0.5378
```

Chi-square Goodness of Fit test	
X-Squared	0.17
Level of Significance	0.05
Df	247
p-value	1
Null Hypothesis	Data is Normally Distributed
Alternate Hypothesis	Data is not Normally Distributed
p-value>level of significance	Data is Normally Distributed

CONCLUSION:

1. If r is a standard uniform random variable, then −ln(r) has the ______**Exponential**______ probability distribution.
2. The sum of three independent and identically distributed ______**Standard Uniform**______ random variables have the ________**Gamma**________ probability distribution.
3. The output of the algorithm of problem 3 has a ________**Normal**________ probability distribution.
4. In step 2 of the algorithm of problem 3, random variables X1 and X2 , each of whose probability distribution is ______________**Exponential**______________ are used to generate a random value Y that has the __________**Normal**______ probability distribution.
5. The random value W that was discussed in problem 4, has the _______**Normal**____________ probability distribution. The expected value of W is: _____**1.3**_______.