Statistical Inference Project 1:Part 1

Author: Aditya Gudal

THE PROBLEM STATEMENT AND INSTRUCTIONS

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

1.Show the sample mean and compare it to the theoretical mean of the distribution. 2.Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. 3.Show that the distribution is approximately normal. In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

THE OVERVIEW:

I have Taken 100000 Samples From the set.We need 40 Exponentials and need to take the Average of these values,lambda is set to 0.2 as given.I have sampled the Data Points obtained from *REXP* Function in R and Kept REPLACEMENT =TRUE. I have fed it to a matrix with Columns equal to the NUMBER OF EXPONENTIALS and Rowsequal to the NUMBER OF SAMPLES. I have then taken The MEAN OF Each Row of this dataset.

```
samples<-100000
n<-40
lambda<-0.2
resamples<-matrix(sample(rexp(n, lambda),n*samples,replace=TRUE),samples,n)
sample_means<-apply(resamples,1,mean)
```

OBSERVATIONS

THIS BELOW IS THE OBSERVED MEAN OF THE DATASET AFTER RESAMPLING OR BOOTSTRAPPING.

```
mean(sample_means)

## [1] 4.210102
```

THIS IS THE OBSERVED STANDARD DEVIATION OF THE DATASET AFTER RESAMPLING.

```r
sd(sample_means)

## [1] 0.5954997
```

THIS BELOW IS THE theoretical MEAN OF THE DATASET.

```r
t_mean=1/lambda
t_mean

## [1] 5
```

THIS IS THE theoretical STANDARD DEVIATION OF THE DATASET.

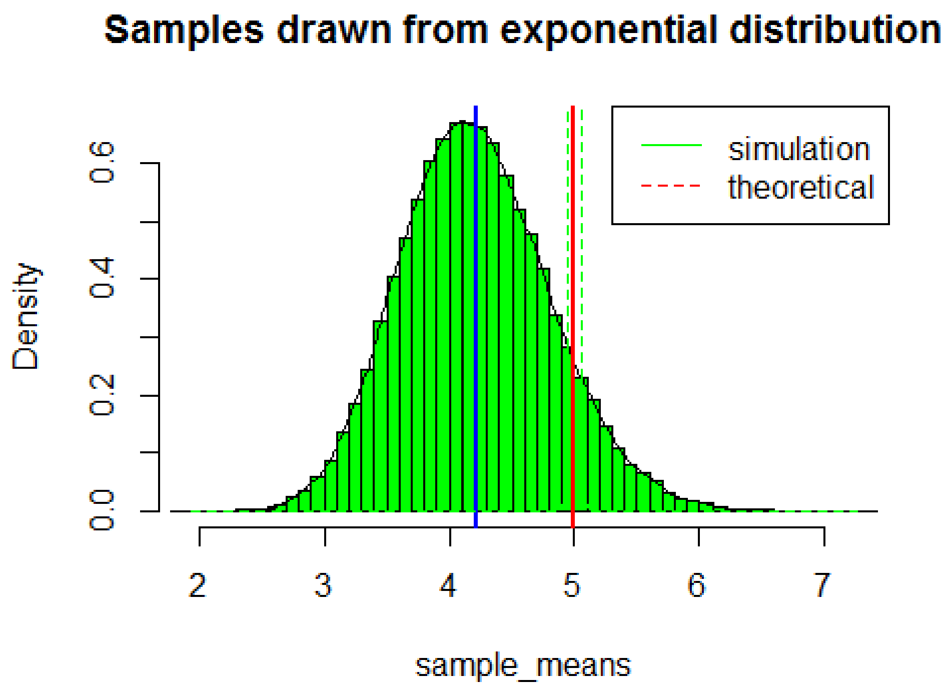
```r
t_std=1/lambda
t_std

## [1] 5
```

PLOTTING AND FIGURES

PLOTTED A HISTOGRAM OUT OF THE RESAMPLED DATA FILLED WITH COLOR GREEN,THE SIMULATED DATA IS SHOWN IN GREEN AND THE THEORETICAL IS SHOWN THROUGH DOTTED LINES.THEORETICAL MEAN=1/LAMBDA , THEORETICAL STD =1/LAMBDA.I HAVE SHOWN THE THEORETICAL MEAN AND OBSERVED MEAN WITH HORIZONTAL LINES.

```r
hist(sample_means, breaks=50, prob=TRUE
     ,main="Samples drawn from exponential
distribution",col='green',border='black')
# density of the averages of samples
lines(density(sample_means))
# theoretical center of distribution
abline(v=1/lambda, col="red",lwd=2)
#observed center of ditribution
abline(v=mean(sample_means),col='blue',lwd=2)
# theoretical density of the averages of samples
xfit <- seq(min(sample_means), max(sample_means), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(samples)))
lines(xfit, yfit, pch=22, col="green", lty=2)
# add legend
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("green",
"red"))
```



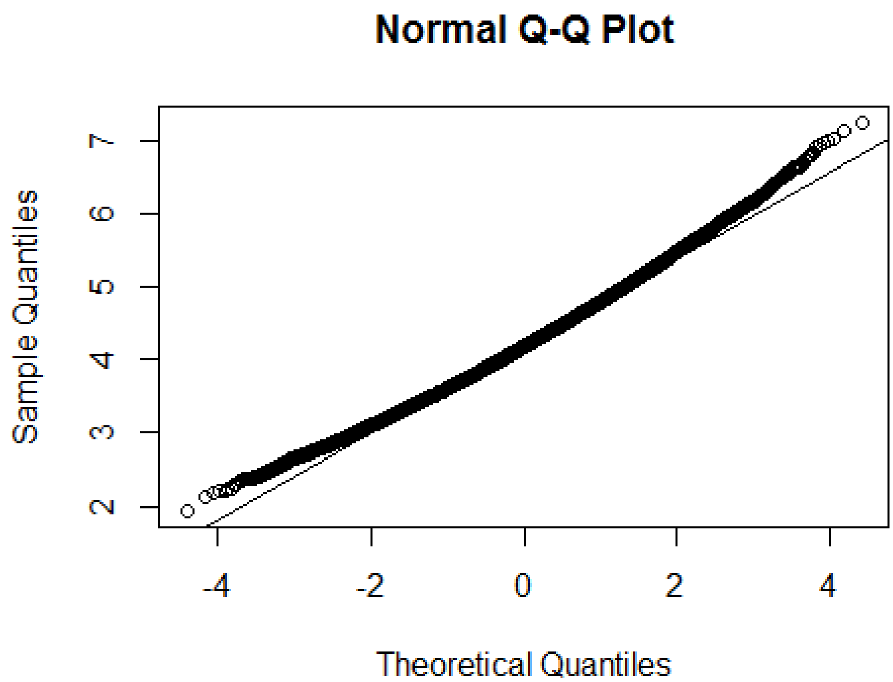
##Check if its a normal distribution. Confidence interval test of atleast 95%

```
quantile(sample_means,c(0.025,0.975))

##     2.5%    97.5%
## 3.126037 5.457272
```

QQPLOT SHOWING HOW CLOSE THE SAMPLED DATA DISTRIBUTION IS TO NORMAL DISTRIBUTION.

```
qqnorm(sample_means);qqline(sample_means)
```



##CONCLUSION

WE SEE THAT THE DISTRIBUTION IS APPROXIMATELY EQUIVALENT TO A NORMAL DISTRIBUTION THROUGH THE HISTOGRAM PLOTS AND QQNORM PLOTS.THE OBSERVED MEAN LIES WITHIN THE CONFIDENCE INTERVAL.