Statistical Inferences Course Project Part1

RAVI KUMAR YADAV

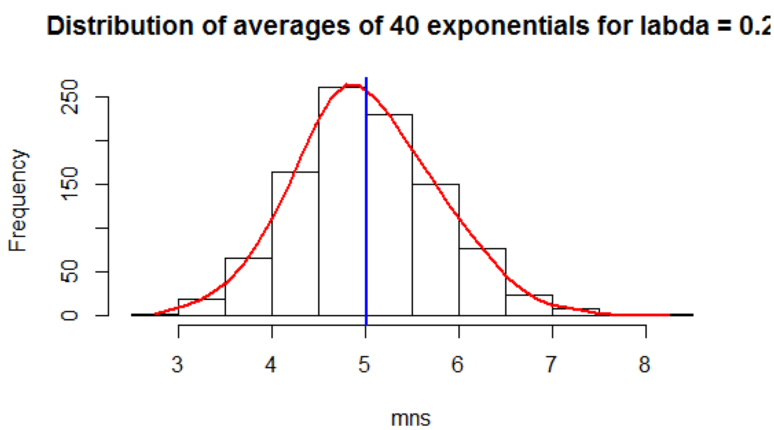
Overview:

Report includes the required theoretical information, simulation results, R code used for the simulation and the explanation for all the three parts of the project.

1. Sample Mean versus Theoretical Mean

According to central limit theorem (CLT) the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution.

In this part of the project thousand simulations are performed for distribution of the mean of 40 exponentials (lambda = 0.2 for all of the simulations). As per the theory mean of the distribution should be 1/lambda = 5. Distribution 1000 simulations for the means 40 exponentials is shown below in the plot:

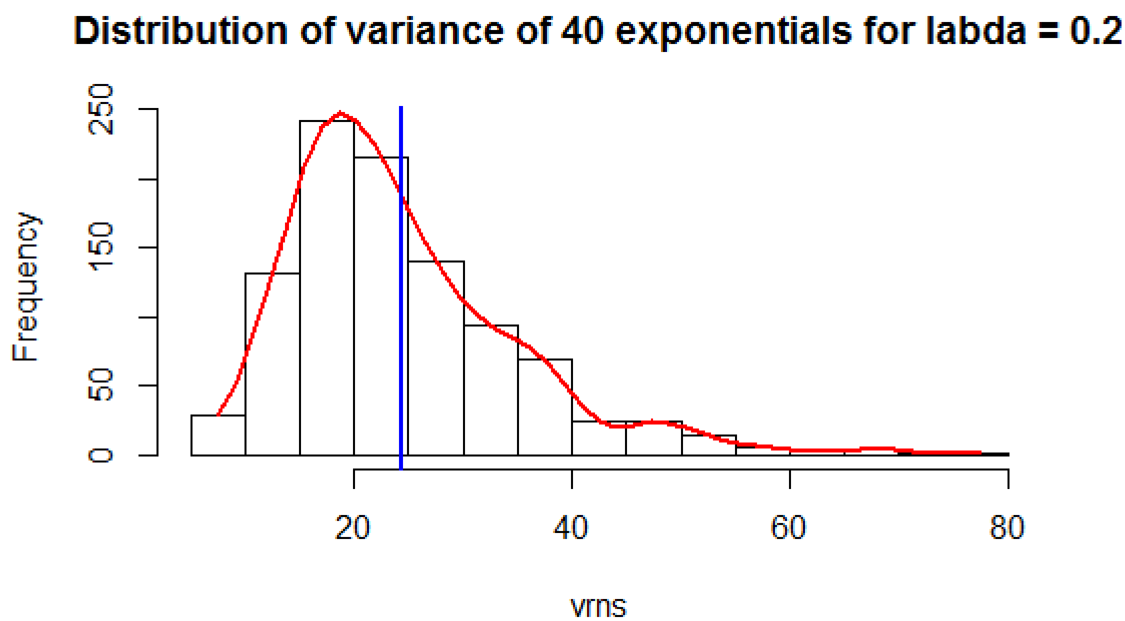



```
> mns <- NULL
> for (i in 1 : 1000) mns = c(mns, mean(rexp(n = 40, rate = 0.2)))
> samplehist <- hist(mns, main="Distribution of averages of 40 exponentials f
or labda = 0.2")
> samplespline <- spline(samplehist$count)
> lines(rescale(samplespline$x,range(samplehist$mids)),samplespline$y, col =
"red", lwd = 2)
> abline(v= mean(mns), col = "blue", lwd = 2)

> mean(mns)
[1] 5.014234
```

Mean for the simulation calculated to be 5.014234, which is fairly close to theoretical mean 5.0.

2. Sample Variance versus Theoretical Variance

Variance for 1000 simulations of 40 exponentials is plotted below:



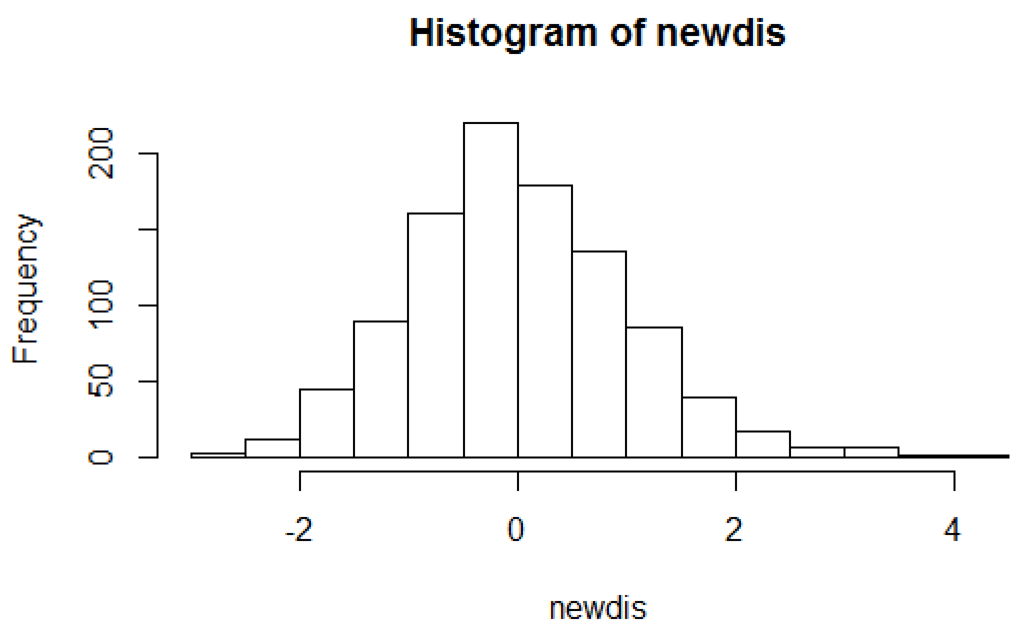

```
> vrns <- NULL
> for (i in 1 : 1000) vrns = c(vrns, var(rexp(n = 40, rate = 0.2)))
> samplehist <- hist(vrns, main="Distribution of variance of 40 exponentials 
for labda = 0.2")
> samplespline <- spline(samplehist$count)
> lines(rescale(samplespline$x,range(samplehist$mids)),samplespline$y, col = 
"red", lwd = 2)
> abline(v= mean(vrns), col = "blue", lwd = 2)

> mean(vrns)
[1] 24.40644
```

As per the theory variance of the distribution should be (1/lambd)^2 = 25 for this distribution. Mean variance obtained (24.40644) from the simulation is fairly close to the theoretical value. Match will improve with increase in number of simulations.

3. Distribution is approximately Normal

To prove that sample mean distribution is approximately normal one can change the distribution to the standard normal distribution and can plot the result and check the mean and the variance of the new distribution if the distribution mean is 0 and variance is 1, it can be said with confidence that distribution is normal.

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\text{Estimate} - \text{Mean of estimate}}{\text{Std. Err. of estimate}}$$




```
> mns <- NULL
> for (i in 1 : 1000) mns = c(mns, mean(rexp(n = 40, rate = 0.2)))
> newdis <- (mns-5)/(5/sqrt(40))
> hist(newdis)

> mean(newdis)
[1] 0.004074544
> var(newdis)
[1] 1.018116
```

We see that mean and variance obtained for the new distribution is close to 0 and 1 respectively. It proves that sample distribution was normal.