Statistical Inference Project - Part I

Overview:

In this report, we will generate a dataset of 1000 simulations performed using random exponential distribution (rexp R function) with *n= 40* observations for each simulations and *lambda or rate = 0.2*. We take mean & variance of each simulation and plot the distribution function and observe the type of distibution visually. Also, compare the theoritical mean & standard deviation/Variance *1/lambda, 1/lambda, 1/lambda ^2* respectively against the sample distribution mean, sd & variance.

Simulations:

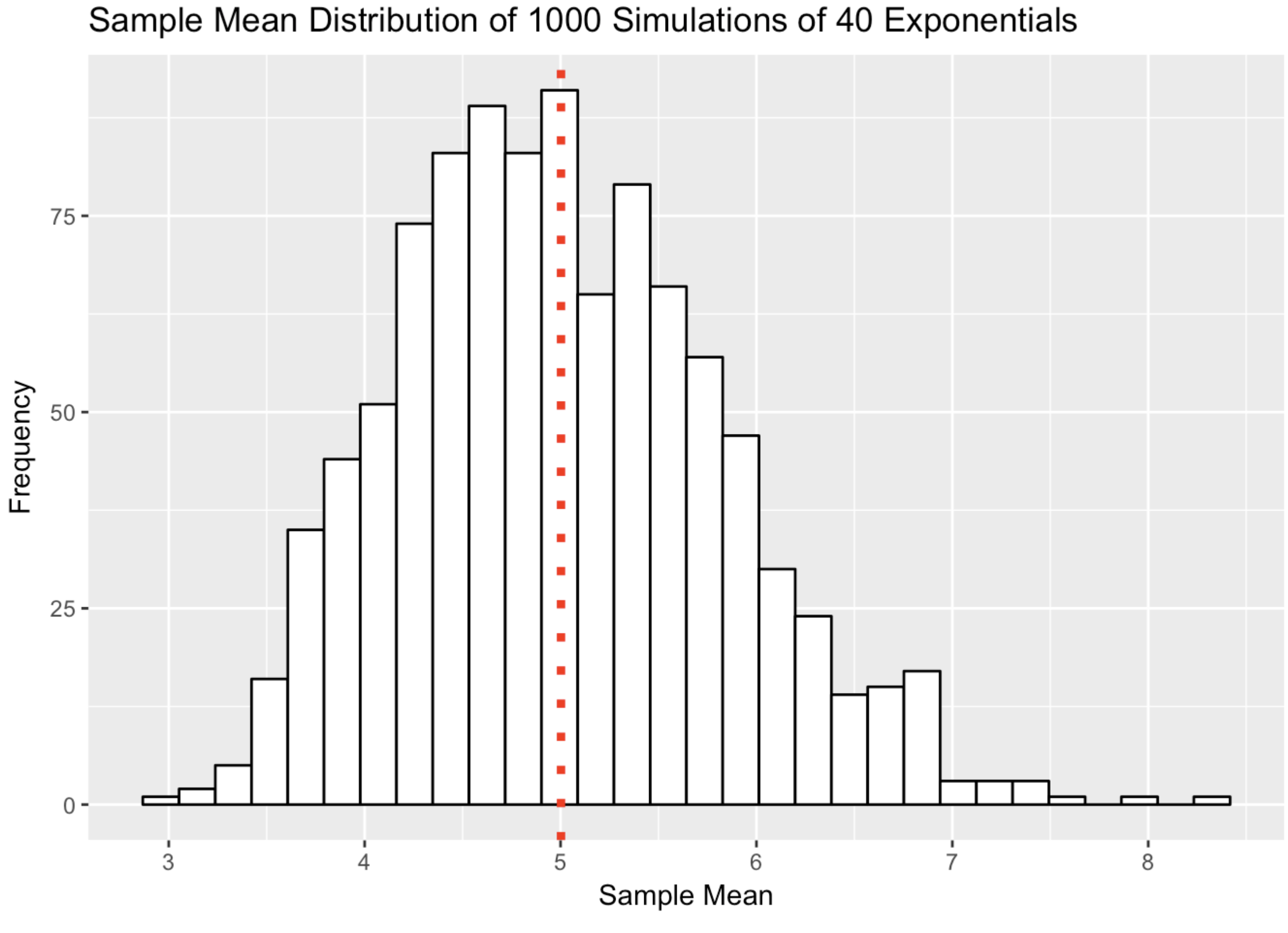
With *lambda / rate set as 0.2* and *n = 40 observations*, rexp function is executed 1000 times to generate 40 observations in each iteration. Mean and variance of 40 observations, for each iteration is appended in a separate vector. We also find average mean and variance of the distribution or 1000 simulations.

```
lambda <- 0.2
n<- 40 # Number of exponentials to be generated per simulation

mn <- NULL
variance <- NULL

for(i in 1:1000){ # 1000 simulations with each simulation having 40 exponentials
  expon <- rexp(n,lambda)
  mn <- c(mn,mean(expon)) # mean of each simulation appended to a common vector/list
  variance <- c(variance, var(expon)) # find variance for each simulation
}

# Find average of means of all simulations
pMean <- mean(mn)
# Find variance of means of all simulations
pVariance <- mean(variance) # average variance or point where variances are centered toward
```

Sample Mean versus Theoretical Mean:

We know that theoritical mean for an exponential distribution is *1/lambda*, in this case, *5*. We compare this value with average mean of all simulations which is *5.0023103*. It is clear that both are **very close**. We also plot the qplot to visually see how and where it is centered. The dotted line shows the population mean of the sample means distribution.

```
# Plot the sample means distribution
gg <- qplot(mn,ylab="Frequency", geom = "histogram", xlab="Sample Mean", main="Sample Mean Distribution of 1000 S
imulations of 40 Exponentials")
gg <- gg + geom_histogram(fill="white", color="black")
gg <- gg + geom_vline(lwd=1.5, xintercept = pMean, colour = "red", lty=3)
gg
```




```
print(1/lambda) # Theoritical Mean (& Standard deviation) of the sample mean-exponential distribution
```

```
## [1] 5
```

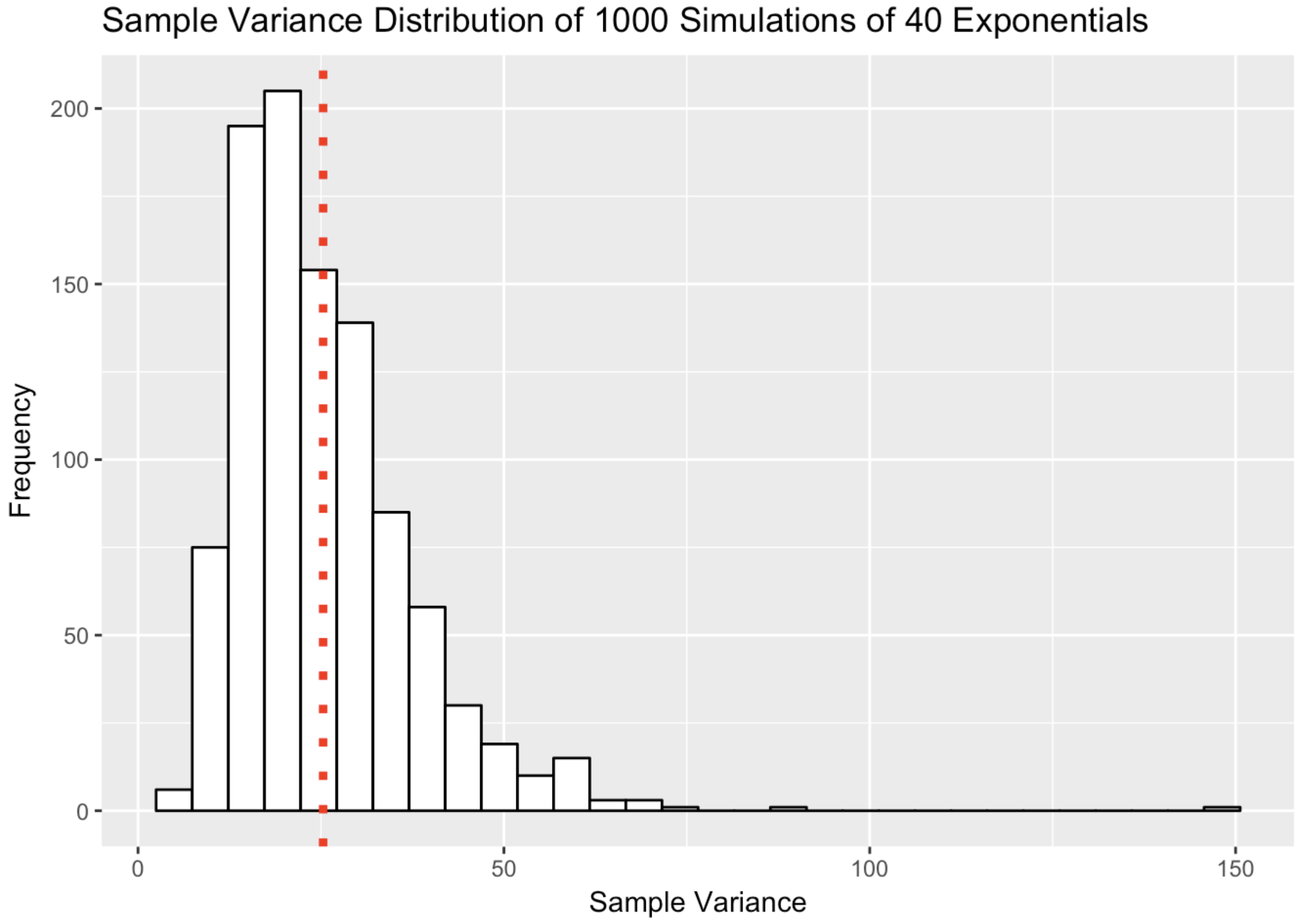
```
print(pMean) #sample Mean
```

```
## [1] 5.00231
```

Sample Variance versus Theoretical Variance:

We know that theoritical variance for an exponential distribution is *1/lambda ^2*, in this case, *25*. We compare this value with average variance of all simulations which is *25.2940238*. It is clear that both are **very close**. We also plot the qplot to visually see how and where it is centered. The dotted line shows the population average of the sample variances of the distribution.

```
# Plot the sample variances distribution
gg2 <- qplot(variance,ylab="Frequency", geom = "histogram", xlab="Sample Variance", main="Sample Variance Distrib
ution of 1000 Simulations of 40 Exponentials")
gg2 <- gg2 + geom_histogram(fill="white", color="black") + geom_vline(lwd=1.5, xintercept = pVariance, colour =
"red", lty=3)
gg2
```




```
# Standard deviation of the distribution

print(1/(lambda^2)) # Theoritical Standard deviation of the sample mean-exponential distribution
```

```
## [1] 25
```

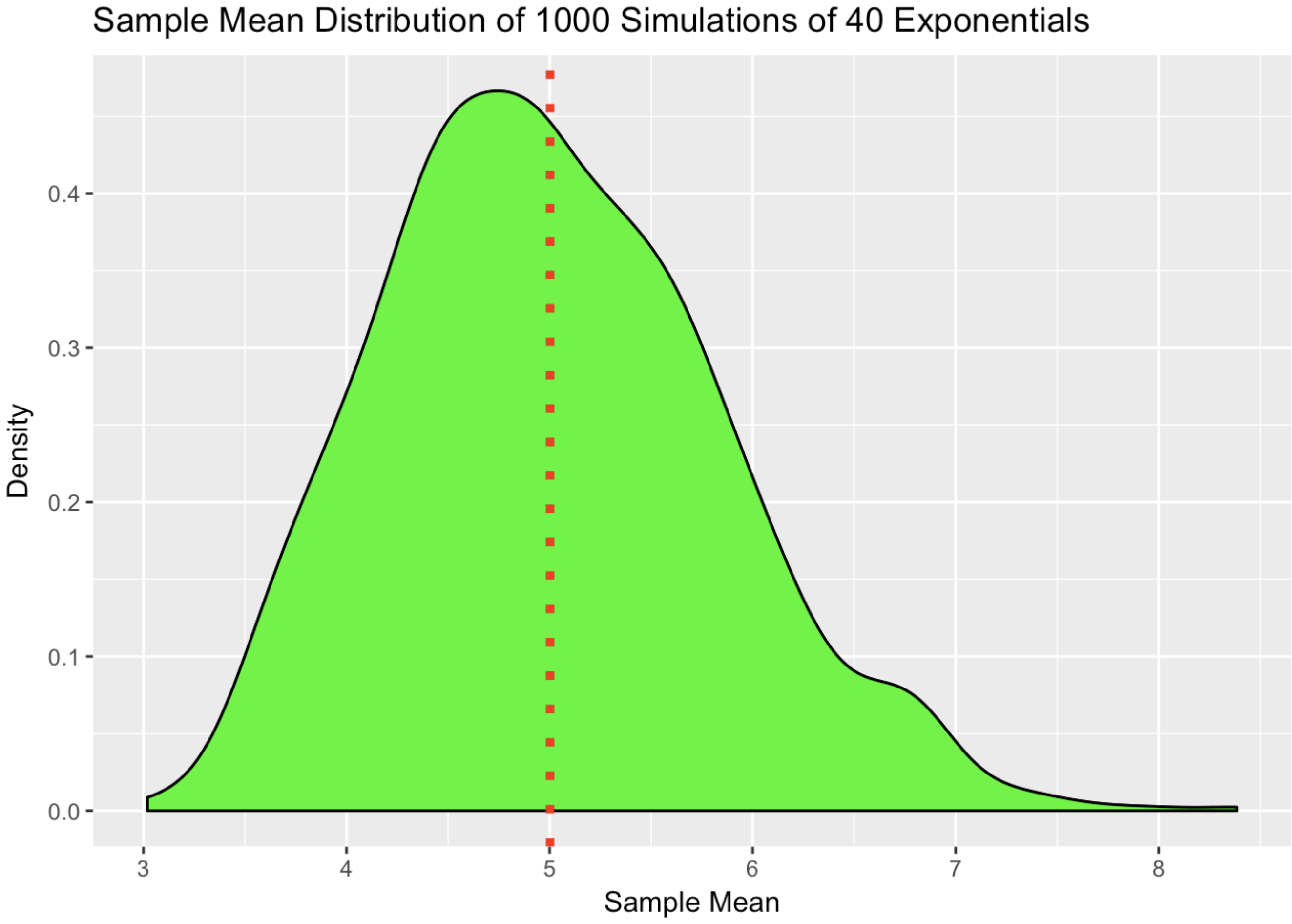
```
print(pVariance) # sample SD
```

```
## [1] 25.29402
```

Distribution:

From the histograms it is evident that the sample mean distribution of 40 exponentials generated over 1000 simulations follows a normal distribution, as per the central limit theorem. We can further prove the same by drawing a density function of sample means, which also follows approximately a normal distribution. The dotted line shows the population mean of the sample means distribution.

```
# Plot the density graph for the sample means of the  distribution
gg3 <- qplot(mn,ylab="Density", geom = "density", xlab="Sample Mean", main="Sample Mean Distribution of 1000 Simu
lations of 40 Exponentials") + geom_density(fill="green", color = "black")
gg3 <- gg3 + geom_vline(lwd=1.5, xintercept = pMean, colour = "red", lty=3)
gg3
```



Conclusion:

This concludes the report with statistical analysis of the random sample created using random exponential functions.