Coursera Statistical Inference Project

Jiajun Zhang

October 12, 2019

Part 1: Simulation Exercise

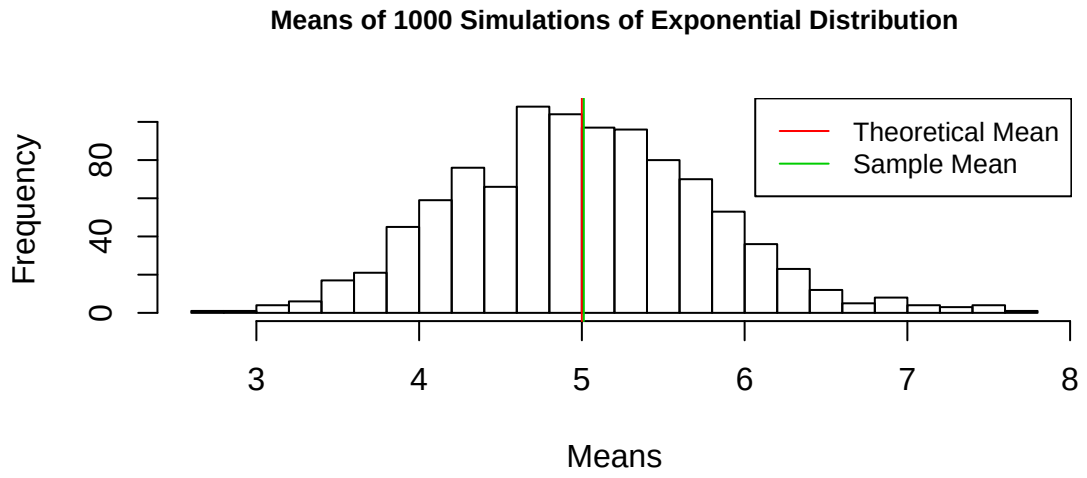
- The purpose of this exercise is to investigate the exponential distribution and make a simulated distribution and compare it to the theoretical one.

```
#Given Information
lambda <- 0.2; n <- 40
B <- 1000 #Thousand simulations
```

Sample Mean and Theoretical Mean of the Distribution

- Here, we will find out the averages of 40 exponentials and do a thousand simulations. Then calculate the average of 1000 means.

```
set.seed(123)
#Simulations
simu_data <- matrix(rexp(n*B, lambda), nrow=B, ncol=n)
#Calculate the mean of each simulation
simu_mean <- apply(simu_data, 1, mean) #1 indicates row calculation
hist(simu_mean, breaks=20, xlab="Means", cex.main=0.8,
     main="Means of 1000 Simulations of Exponential Distribution")
abline(v=1/lambda, col=2)
abline(v=mean(simu_mean), col=3)
legend("topright", c("Theoretical Mean", "Sample Mean"),
       col=c(2,3), lty=c(1,1), cex=0.85)
```




```
#Theoretical Mean
1/lambda
```

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

```r
#Simulation Mean
mean(simu_mean)
```

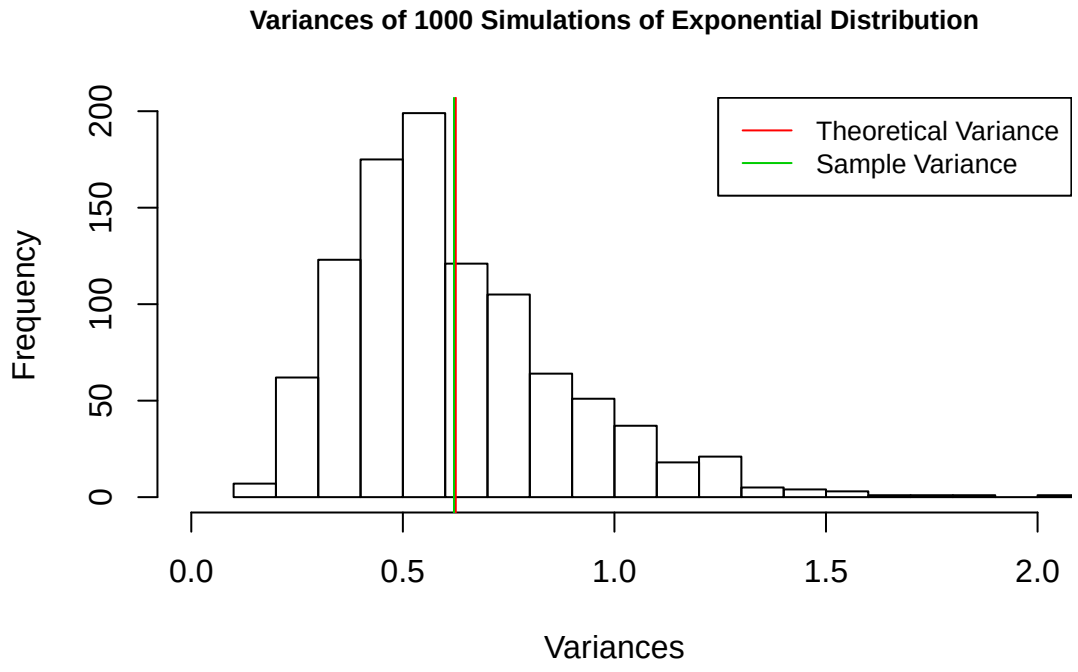
```
## [1] 5.011911
```

- We see that from the graph and calculation, the simulation mean and theoretical mean of the exponential distribution are really close to each other.

Sample Variance and Theoretical Variance of the Distribution

- Similar to the previous one, but this time we will compare their variances.

```r
# Calculate the variances of each simulation
simu_var <- apply(simu_data, 1, var)
simu_var <- simu_var/n
hist(simu_var, breaks=20, xlab="Variances", xlim=c(0, 2), cex.main=0.8,
     main="Variances of 1000 Simulations of Exponential Distribution")
abline(v=((1/lambda)^2)/n, col=2)
abline(v=mean(simu_var), col=3)
legend("topright", c("Theoretical Variance", "Sample Variance"),
       col=c(2,3), lty=c(1,1), cex=0.85)
```




```r
#Theoretical Variance
((1/lambda)^2)/n
```

```
## [1] 0.625
```

```r
#Simulation Variance
mean(simu_var)
```

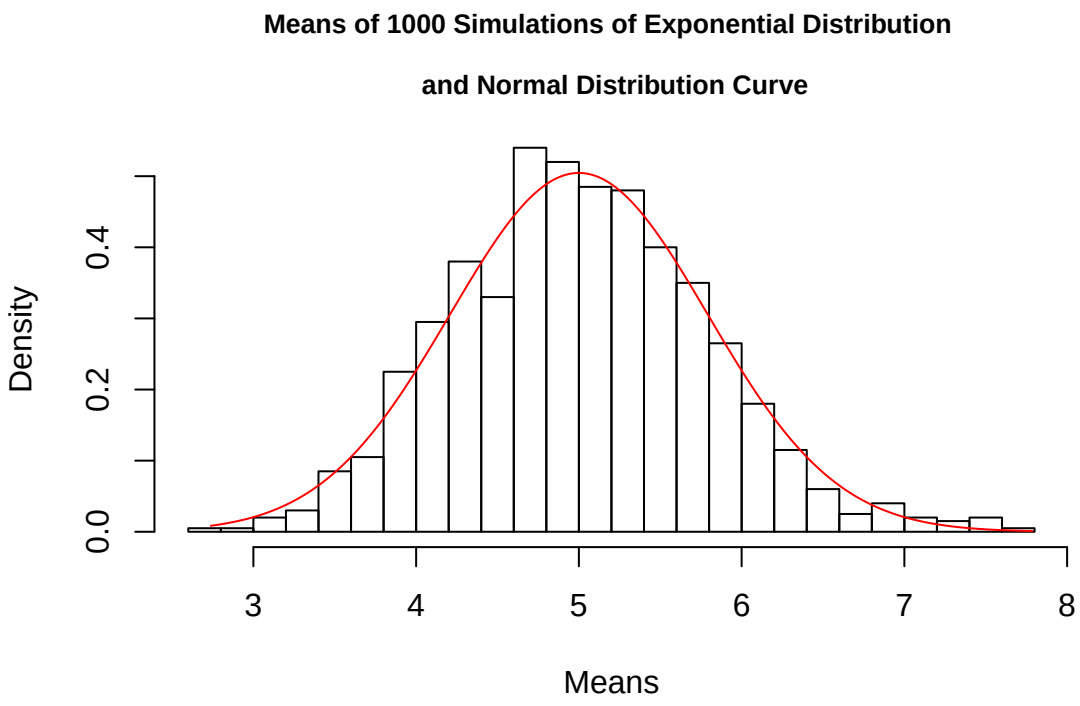
```
## [1] 0.6208656
```

- We see that from the graph and calculation, the simulation variance and theoretical variance of the exponential distribution are really close to each other.

Distribution Normality

- Now, we want to determine whether the simulation means follow a normal distribution.

```
#Plot a normal curve
x <- seq(min(simu_mean), max(simu_mean), length=200)
y <- dnorm(x, mean=1/lambda, sd=(1/lambda)/sqrt(n))
hist(simu_mean, breaks=20, prob=T, xlab="Means",
     main="Means of 1000 Simulations of Exponential Distribution
     \n and Normal Distribution Curve", cex.main=0.8)
lines(x, y, col=2)
```



- The histogram shows the distribution of the means of 1000 simulated results and we add a normal distribution curve on it. The graph tells us that the simulated distribution approximately follows a normal shape.