Statistical Inference Course Project - Simulation

Naman Gandhi

February 4, 2017

Overview

In this project we'll investigate the distribution of averages of 40 exponentials in R and compare it with the Central Limit Theorem by doing a thousand simulations.

We'll break it into 5 broad sections -
1) **Simulations**
2) **Sample Mean versus Theoretical Mean**
3) **Sample Variance versus Theoretical Variance**
4) **Understanding the Distribution**
5) **Conclusion**

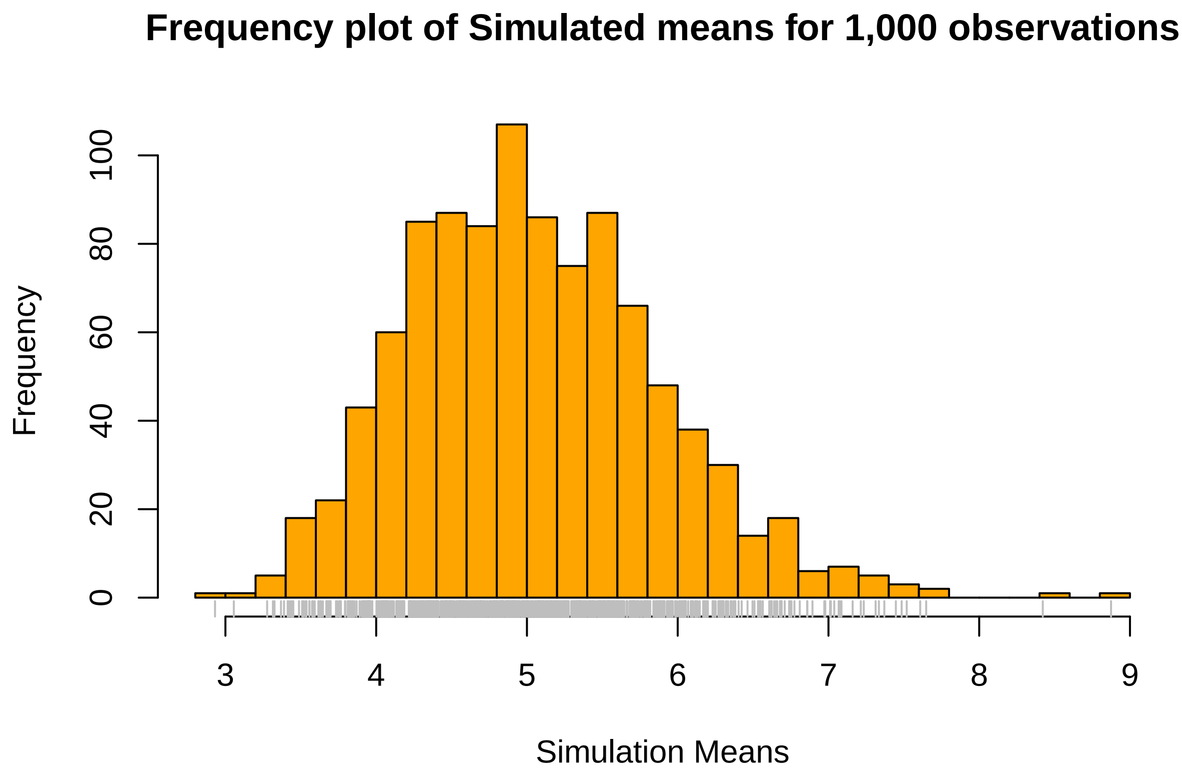
Let's dive in!!

Section 1: Simulations

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.

```r
# Setting seed for reproducibility
set.seed(7815)
# Setting simulation parameters
n = 40
lambda = 0.2
simulations = 1000
# Generating 1,000 simulations
simulated_exp <- replicate(simulations, rexp(n, lambda))
# calculate columnar mean of exponentials
means_exp <- apply(simulated_exp, 2, mean)

# Visualizing the means using histogram and adding rugs to see finer density scatter
hist(means_exp, col = "orange", xlab = "Simulation Means",
     main = "Frequency plot of Simulated means for 1,000 observations", breaks = 40)
rug(means_exp, col = "grey",lwd = 1)
```



Section 2: Sample Mean versus Theoretical Mean

We know that the theoretical mean of an exponential distribution is 1/lambda, which implies for our distribution its 5

```
(sample_mean <- mean(means_exp))          # Sample Mean
```

```
## [1] 5.046637
```

```
(theoretical_mean <- 1/lambda)            # Theoretical Mean
```

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

```
(delta_mean <- abs(sample_mean - theoretical_mean))      # Delta
```

```
## [1] 0.0466371
```

With the delta of 0.046, we can say that The sample mean is very close to the theoretical mean

Section 3: Sample Variance versus Theoretical Variance

```
(sample_var <- var(means_exp))            # Sample variance
```

```
## [1] 0.6888332
```

```
(theoretical_var <- (1/lambda)^2/n)       # Theoretical variance
```

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

```r
(delta_var <- abs(sample_var - theoretical_var))     # Delta
```

```
## [1] 0.06383324
```

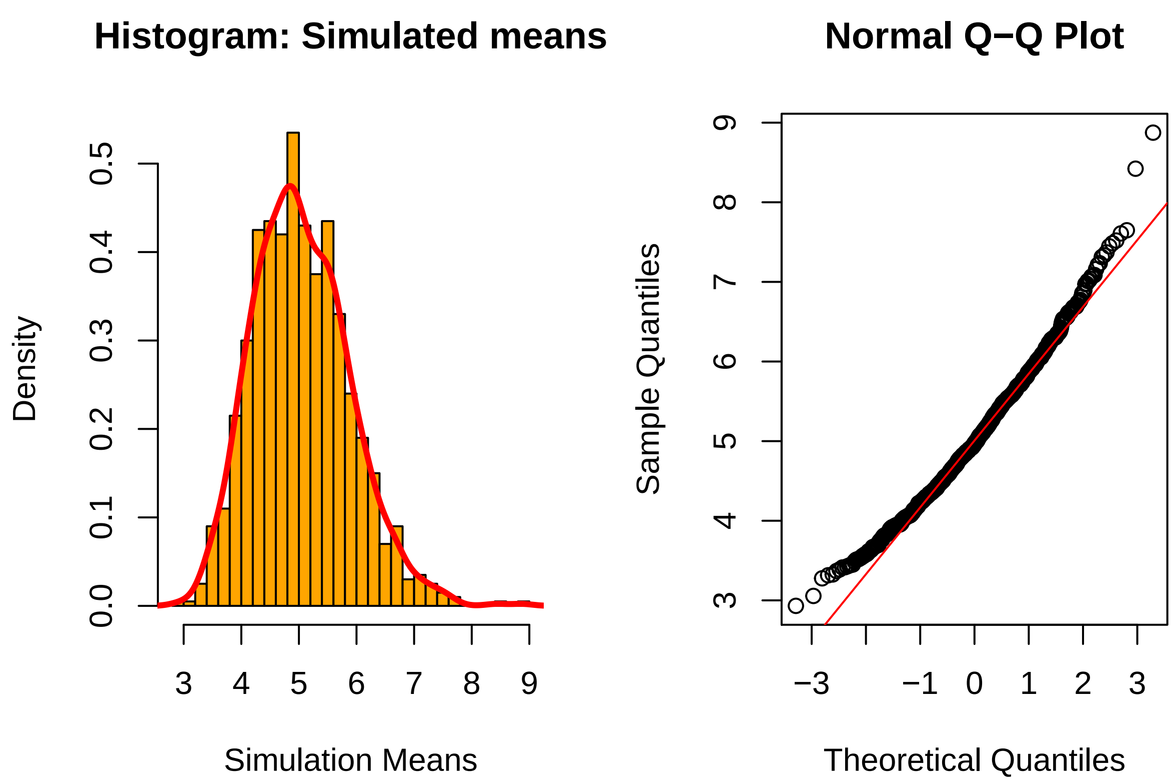
With the delta of 0.064, we can say that the sample variance is very close to the theoretical variance

Section 4: Understanding the Distribution

The exponential distribution is approximately normal. Due to the Central Limit Theorem, the means of samples should also follow a normal distribution. Let's see if this holds true.

```r
par(mfrow = c(1,2))
hist(means_exp, probability = TRUE, col = "orange", xlab = "Simulation Means",
     main = "Histogram: Simulated means", breaks = 40)
lines(density(means_exp), lwd=3, col = "red")

qqnorm(means_exp, main="Normal Q-Q Plot", xlab="Theoretical Quantiles", ylab="Sample Quantiles")
qqline(means_exp, col="red")
```



Section 5: Conclusion

Yes, indeed the graphs show the distribution as a close approximation of Gaussian and as we simulate for higher n sizes, the graph will tend to become more closer to the Gaussian Normal!