Central Limit Theorem Exponent Distribution

Richie Liew

Nov 15, 2016

Part 1: Simulation Exercise Instructionsless

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.

Data Preparation

The Central Limit Theorem (CLT) states that the sample mean of identically distributed independent random variables is approximately normally distributed if the sample size is large. This is true for virtually any distribution! The following we will sample and simulate to get a CLT resuls.

```r
set.seed(33) # Set Seed
lambda <- 0.2 # We set to perform 1000 simulations with 40 samples
sam_size <- 40
sim_num <- 1000
sim_exp <- replicate(sim_num, mean(rexp(sam_size,lambda))) # This will do a 1000 simulations
sim_mean <- mean(sim_exp)
sim_median <- median(sim_exp)
sim_sd <- sd(sim_exp)
```

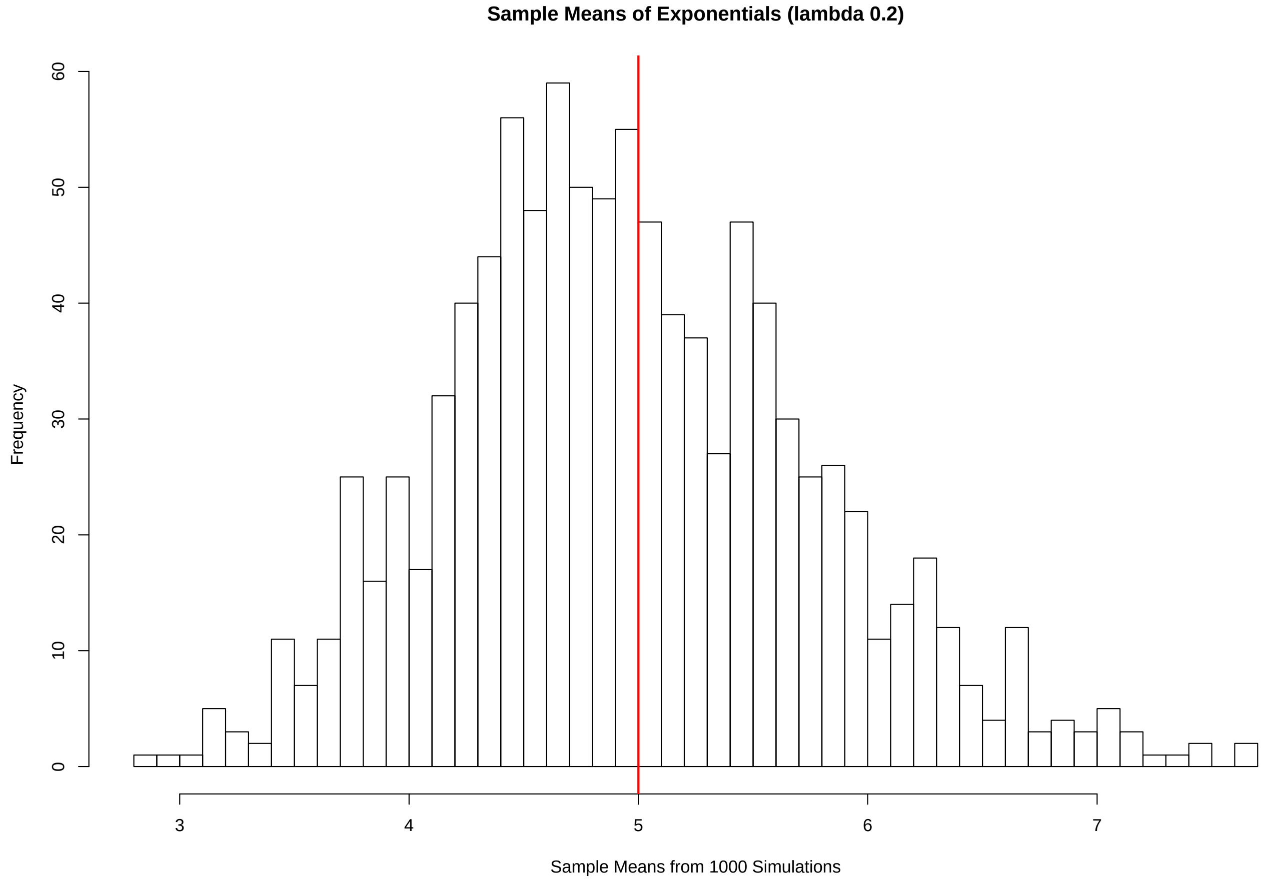
1. Show the sample mean and compare it to the theoretical mean of the distribution.

The theoretical mean and sample mean for the exponential distribution.

```r
tm <- 1/lambda  # calculate theoretical mean
sm <- mean(sim_exp) # calculate avg sample mean
```

Theoretical mean is 5 and the Sampling mean is 4.9644306, this shows that the sample mean is very close to the theoretical mean at 5.

```r
hist(sim_exp,  freq=TRUE, breaks=50, main="Sample Means of Exponentials (lambda 0.2)",
     xlab="Sample Means from 1000 Simulations")
abline(v=5, col="red", lwd=2)
```



2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
theor_sd <- (1/lambda)/sqrt(40) # Calculation of the theoretical sd
theor_var <- theor_sd^2 # Calculation of the theoretical variance for sampling data
sim_var    <- var(sim_exp)
```

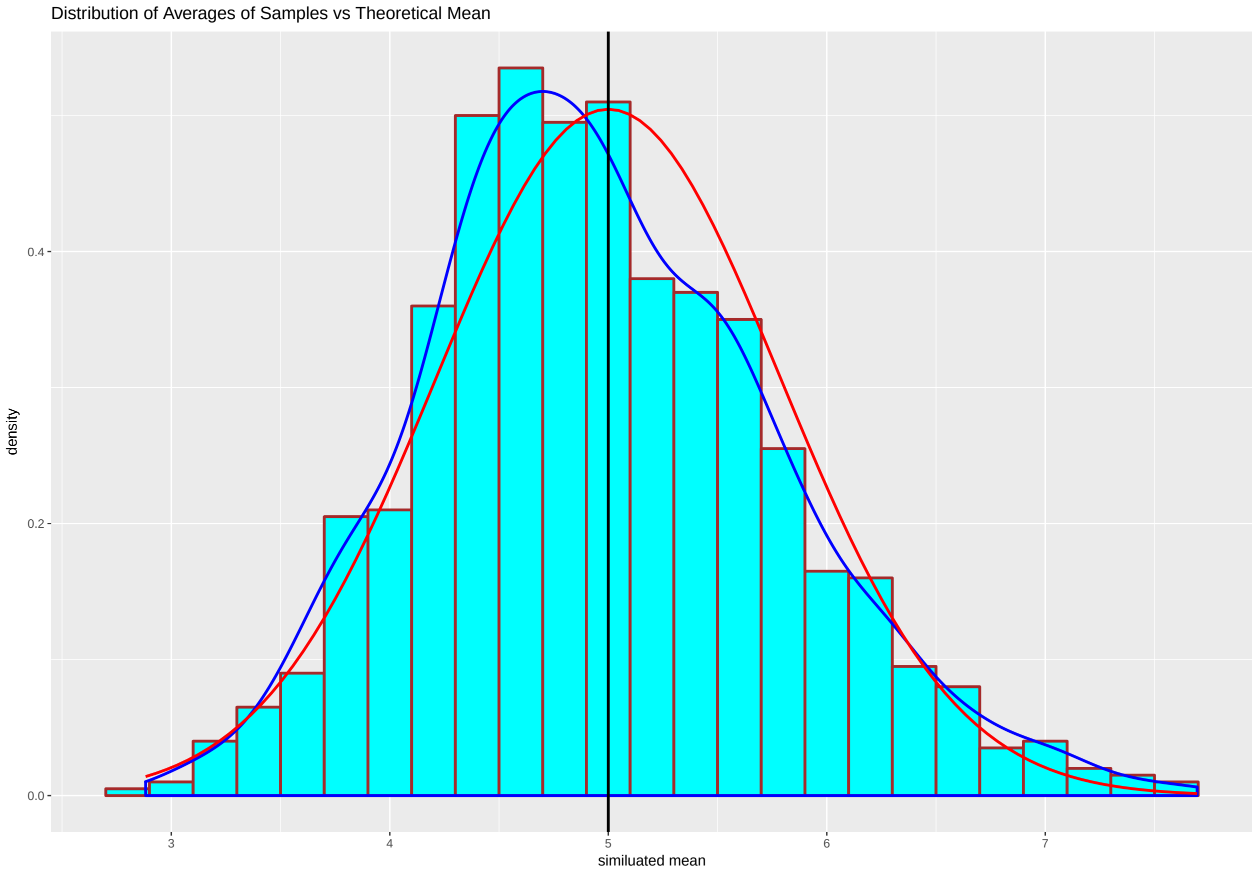
The variance for the sample data is 0.6456619 and the theoretical variance is 0.625, both these values are close to each other.

3. Show that the distribution is approximately normal.

To show the distribution is normal, we plot the distribution of simulated sample data and overlay the normal distribution (a bell curve) with lambda=0.2. The green line is the distribution of averages of the simulated samples. The red line is the normal distribution with lambda = 0.2. The figure shows that the 2 distribution lines, blue and red, are in the similar bell shape and thus the distribution of simulated data is approximately normal.

```
sim_mean_df <- data.frame(sim_exp);
names(sim_mean_df) <- c("sim_exp")
ggplot(sim_mean_df, aes(x=sim_exp)) +
       labs(x="similuated mean", title="Distribution of Averages of Samples vs Theoretical Mean") +
       geom_histogram(aes(y=..density..), color="brown", fill="cyan", size=1, binwidth=0.2) +
       geom_density(color="blue", size=1) +
```

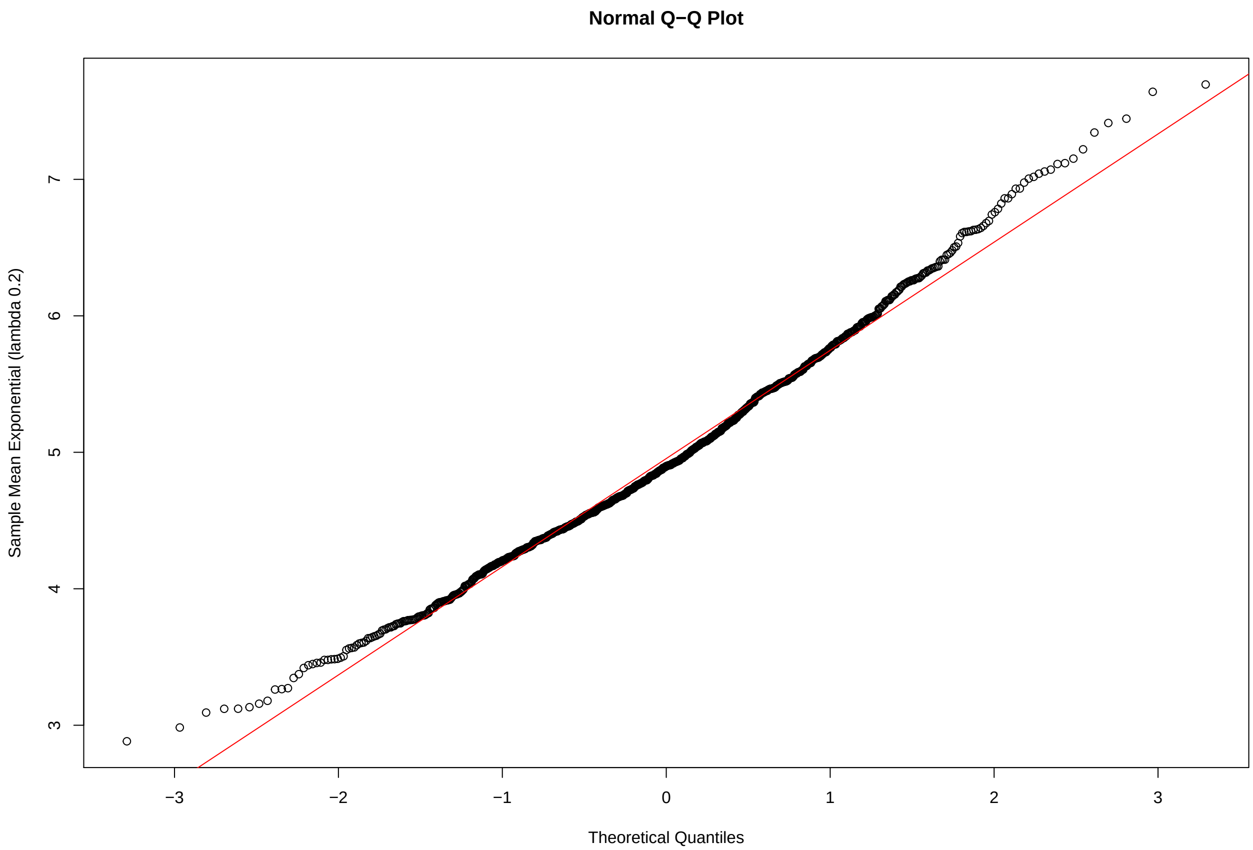
```r
    stat_function(fun=dnorm, args=list(mean=tm, sd=theor_sd), color = "red", size=1) +
    geom_vline(xintercept=tm, color="black", size=1);
```



Conclusion

We see that the distribution is approximately normal as the straight line is closer to the points using a QQ plot to illustrate.

```r
qqnorm(sim_exp, ylab = "Sample Mean Exponential (lambda 0.2)")
qqline(sim_exp, col = 2)
```



Normal Q−Q Plot
Sample Mean Exponential (lambda 0.2)
Theoretical Quantiles
3
4
5
6
7
−3
−2
−1
0
1
2
3