A Demonstration of the Central Limit Theorem Using the Exponential Distribution

Russ Boucher

March 7, 2015

Overview

The purpose of this project is to investigate the exponential distribution in **R** and compare it with the Central Limit Theorem. The Central Limit Theorem states that the means of random samples drawn from any distribution with mean m and variance s^2 will have an approximately normal distribution with a mean equal to m and a variance equal to s^2/n. This project involves running a simulation to generate a distribution of exponentials, comparing the sample mean and variance to the theoretical mean and variance, and showing that the distribution is approximately normal.

Simulations

The exponential distribution can be simulated in **R** using the `rexp(n, lambda)` function, where n is the number of observations and *lambda* is the rate parameter. First, set the random number seed so the results are reproducible. Next, generate a distribution of 1000 means of $n = 40$ exponentials with a rate parameter of $lambda = 0.2$.

```
set.seed(42)
mns = NULL
n <- 40
lambda <- 0.2
for (i in 1:1000){
   mns  = c(mns, mean(rexp(n, lambda)))
}
```

Sample Mean versus Theoretical Mean

How does the sample mean compare to the theoretical mean?

```
mean(mns) # sample mean
```

```
## [1] 4.986508
```

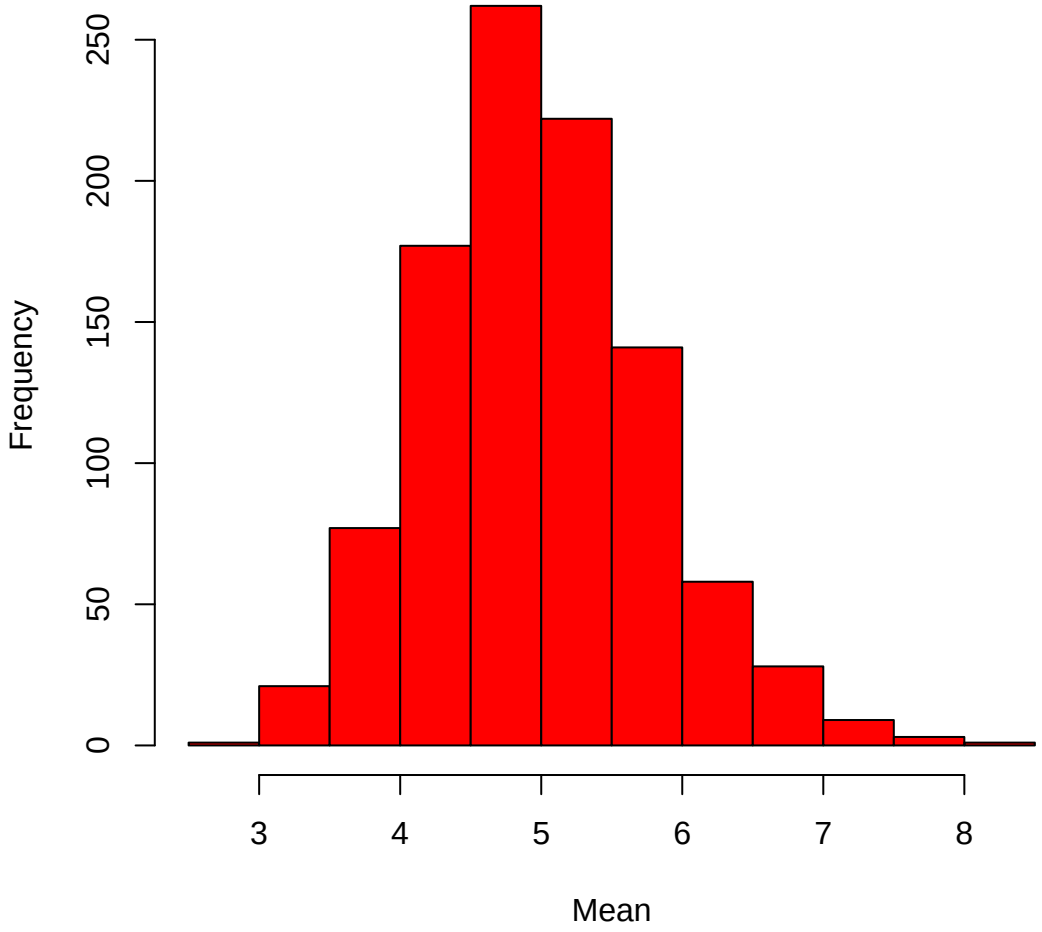
```
print(1 / lambda) # theoretical mean
```

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

The sample mean is very close to the theoretical mean, as would be expected under the Central Limit Theorem.

```r
# plot a histogram of the 40 means
hist(mns, main = "Figure 1. Histogram of Means of 40 Exponentials",
     xlab = "Mean", col = "red")
```

Figure 1. Histogram of Means of 40 Exponentials



As shown in Figure 1, the means appears to be normally distributed, with the distribution centered around the theoretical mean of 5.

Sample Variance versus Theoretical Variance

How does the sample variance compare to the theoretical variance?

```r
var(mns) # sample variance
```

```
## [1] 0.6344405
```

```r
print(1 / (lambda^2 * n)) # theoretical variance
```

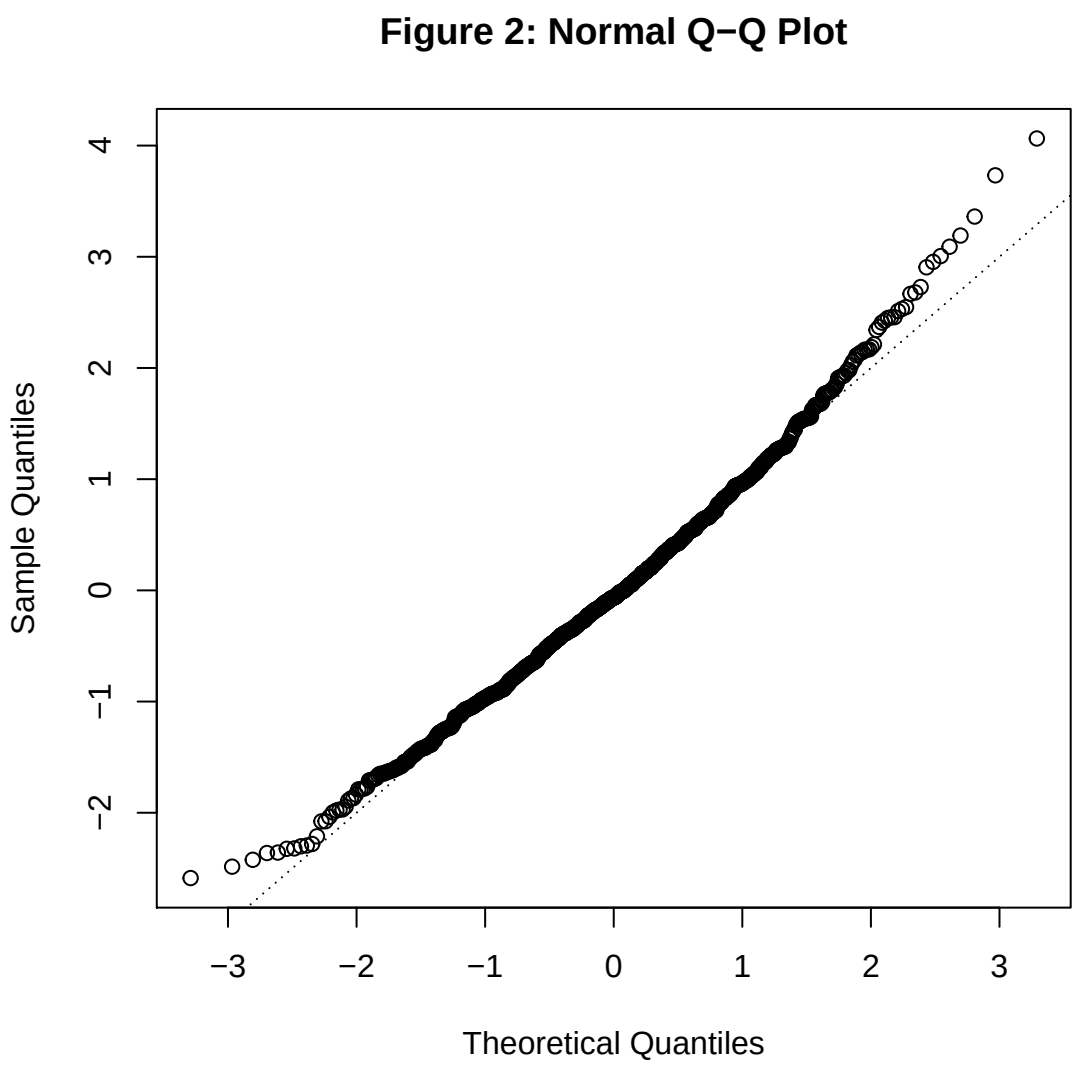
```
## [1] 0.625
```

The sample variance is very close to the theoretical variance, again as would be expected under the Central Limit Theorem.

Distribution

Finally, is the distribution of means approximately normal? Figure 2 shows a Q-Q plot of sample quantiles from the distribution of means against theoretical quantiles from a normal distribution. Q-Q plots are used to see how well a theoretical distribution models empirical data. The linear trend indicates that the distribution of means is approximately normal, as would be expected under the Central Limit Theorem.

```
mns_scale <- scale(mns) # convert the means to standard Z scores
qqnorm(mns_scale, main = "Figure 2: Normal Q-Q Plot")
abline(0, 1, lty = 3)
```



The results of this project demonstrate that means of random samples drawn from the exponential distribution with mean m and variance s^2 have an approximately normal distribution with a mean equal to m and a variance equal to s^2/n.