Coursera Statistical Inference Project Part 1

Nicholas Pairolero

Overview

This project analyzes the distribution of the sample mean through simulation. 40 Exponentials are simulated 1000 times and means/variances are compared to theoretical values. Finally, it is shown that even though the underlying distribution is exponential, the distribution of the sample mean is approximately normal.

Simulations

First specify the rate parameter

```r
lambda <- .2
```

Next, run 1000 simulations of 40 simulations and record the mean of each. The first line of code initializes the variable mean to store the means of each simulation. The for loop simulates 40 exponential random variables 1000 times and records the means. The function rexp within the for loop generates 40 exponential draws at rate lambda.

```r
mean <- numeric(40)
for (i in 1:1000){
  mean[i]<- mean(rexp(40, rate = lambda))
}
```

Sample Mean versus Theoretical Mean

The **theoretical mean of 40 exponentials is 1/lambda=5**, which is the same as the exponential distribution mean. The mean of the simulation means is

```r
mean(mean)
```

```
## [1] 4.956784
```

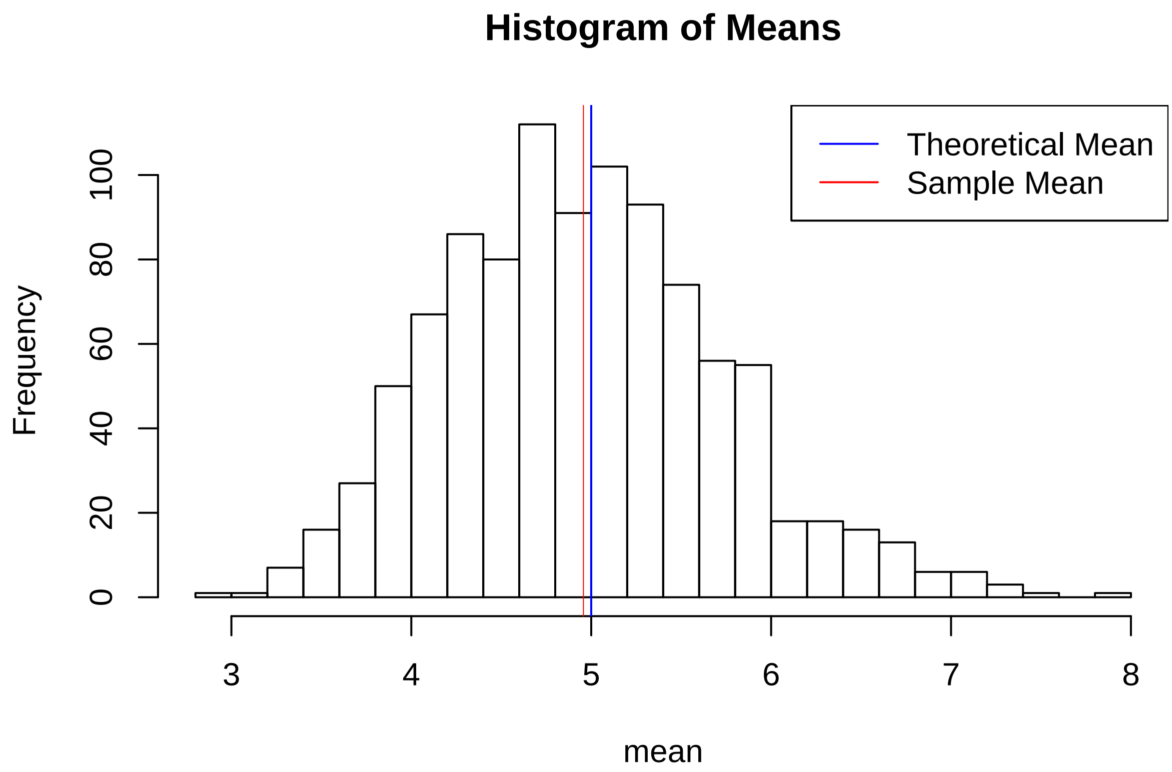
Therefore, the **absolute value in the difference in the sample mean and theoretical mean** is

```r
abs(mean(mean)-5)
```

```
## [1] 0.04321579
```

In words, a sample size of 1000 is enough to get very close to the theoretical mean. The following shows the distrbution of sample means, and notes the theoretical mean and mean of sample.

```r
hist(mean, breaks=20, main="Histogram of Means")
abline(v = 5, col = "blue", lwd = 1)
abline(v = mean(mean), col = "red", lwd = .5)
legend("topright", col = c("blue","red") , lty=c(1,1),  legend = c("Theoretical Mean", "Sample Mean"))
```



Sample variance versus Theoretical Variance

The **theoretical variance of the mean of 40 exponentials** is

```
(1/(lambda)^2)/40
```

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

The variance of the simulation means is

```
var(mean)
```

```
## [1] 0.5971921
```

The **absolute value in the difference between theoretical variance and sample variance** is

```
abs((1/(lambda)^2)/40-var(mean))
```

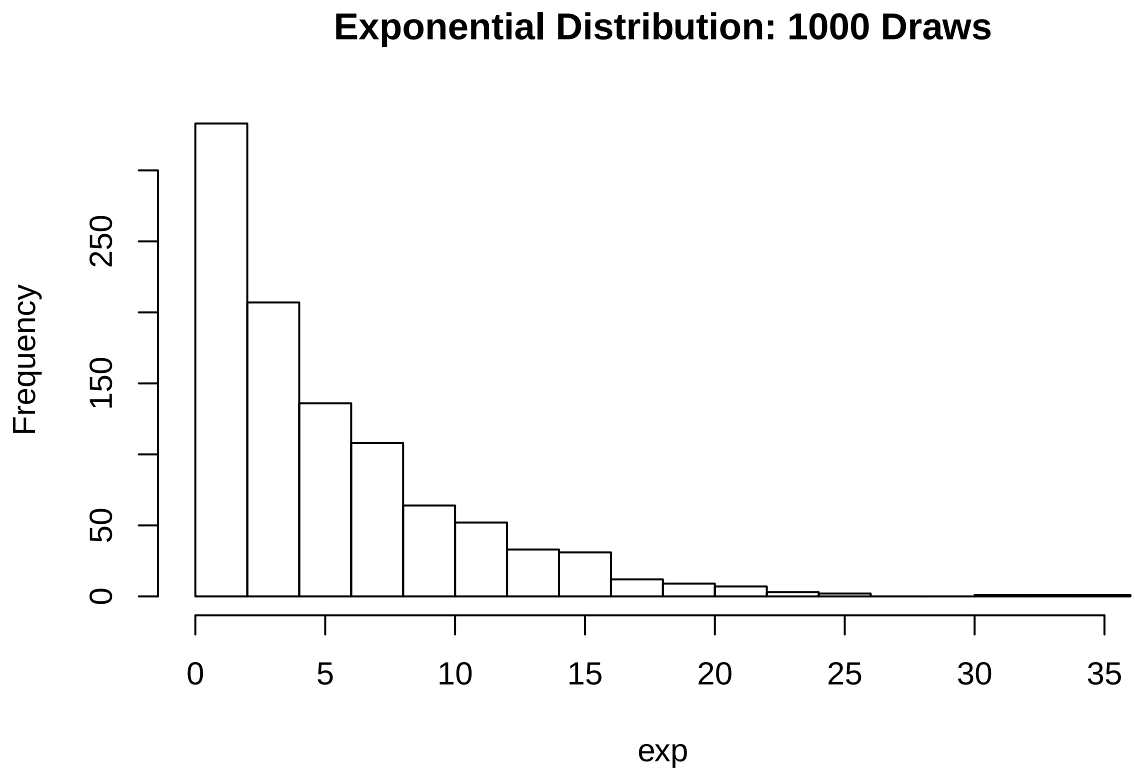
```
## [1] 0.02780793
```

As with the sample and theoretical mean, a sample size of 1000 is enough to get very close to the theoretical variance.

Distribution

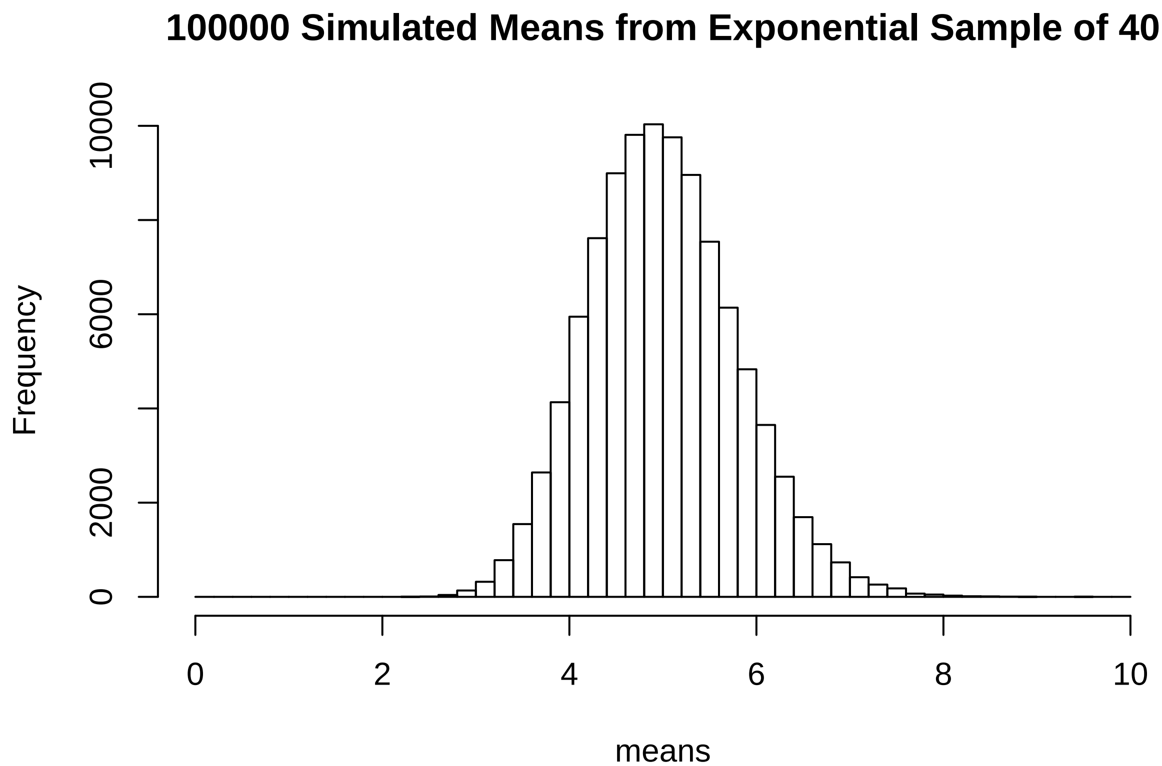
First, simulate 1000 exponentials and plot the histogram

```
exp <- numeric(1000)
exp <- rexp(1000, lambda)
hist(exp, main = "Exponential Distribution: 1000 Draws", breaks = 20)
```



To show that the distribution of the sample mean looks normal, simulate 100000 times and construct a histogram of means. 100000 is used to give a more accurate picture of the sample mean distribution.

```
means <- numeric(100000)
for (i in 1:100000){
  means[i] <- mean(rexp(40, lambda))
}
hist(means, main="100000 Simulated Means from Exponential Sample of 40", breaks=seq(from=0, to=10, by=.
```



As can be seen, the base distribution (exponential) is not normal, but the distribution of the sample mean looks much more normal. By the central limit theorem, this should become more pronounced as the sample size (here n=40) increases.