Statistical Inference Course Project - Part 1

Mike McFarren

November 26, 2016

Synopsis

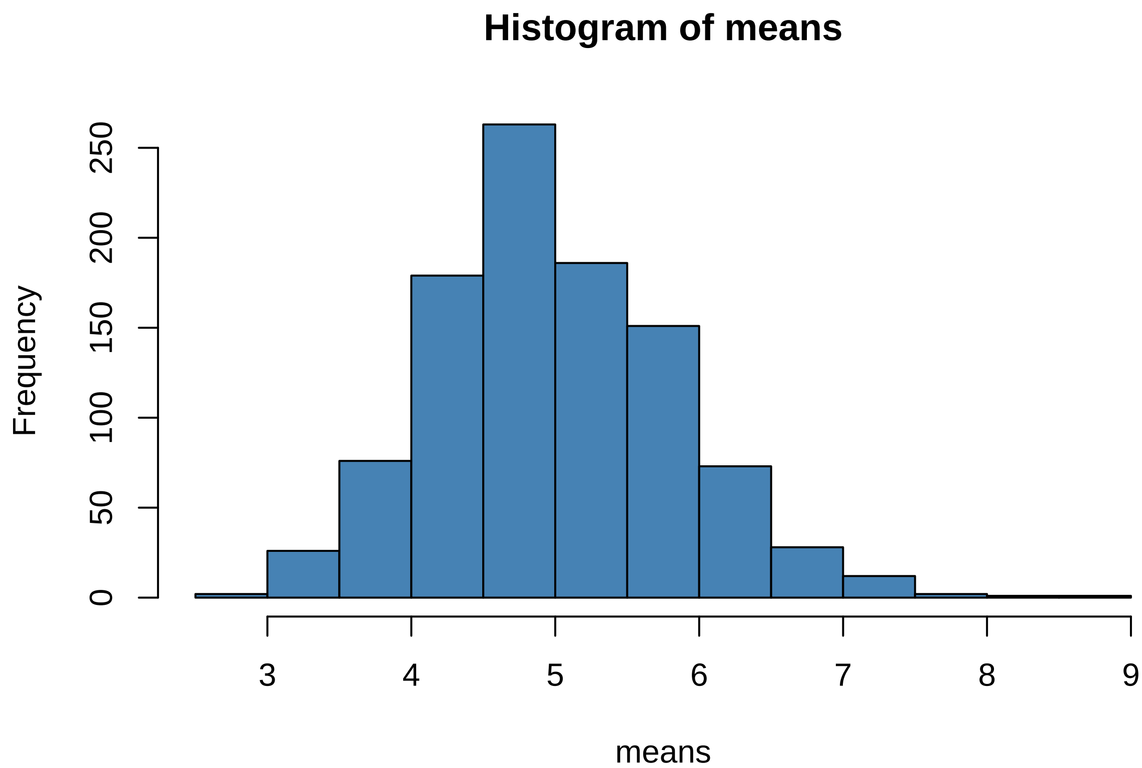
In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. I will also investigate the distribution of averages of 40 exponentials. I will do these calculations over a thousand simulations.

Setup

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```r
simCount <- 1000 # # of simulations
n <- 40 # # of exponentials
lambda <- 0.2 # rate parameter

set.seed(37027)

simulateExponentials <- matrix(rexp(simCount * n, lambda), simCount);
means <- apply(simulateExponentials, 1, mean) # also referred to as X bar

hist(means, col="steel blue", main="Histogram of means", xlab = "means")
```



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

```
mu <- 1/lambda # theoretical mean of the distribition  (aka - expected mean)
mu
```

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

```
mu.x <- mean(means) # sample mean
mu.x
```

```
## [1] 4.999394
```

We can see that the expected mean (5) and the sample mean (4.9993942) are extremely close to each other

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

```
sigma <- (1/lambda) / sqrt(n)  # theoretical stddev
sigma
```

```
## [1] 0.7905694
```

```r
sigma.x <- sd(means) # actual (sample) stddev
sigma.x
```

```
## [1] 0.8423335
```

The theoretical standard deviation = 0.7905694 and the actual standard deviation for our simulations = 0.8423335

```r
var <- sigma^2   # theoretical variance
var
```

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

```r
var.x <- var(means) # actual (sample) variance
var.x
```

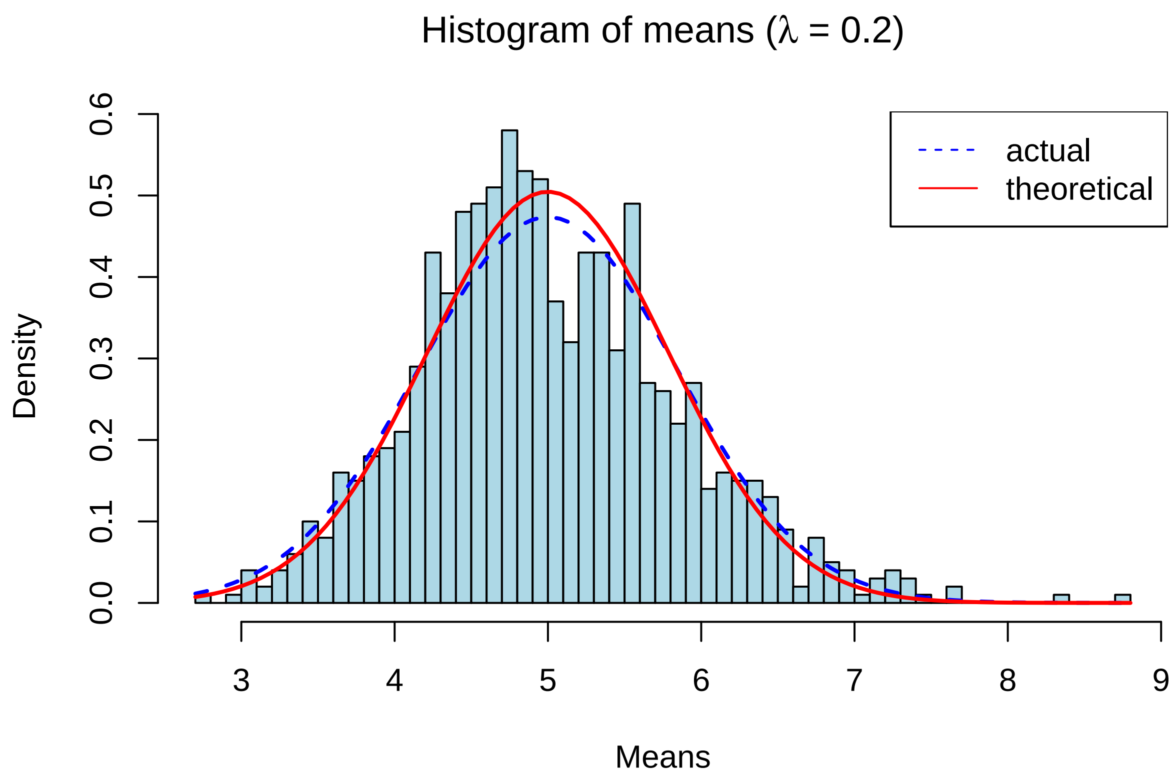
```
## [1] 0.7095258
```

The theoretical variance = 0.625 and the actual variance for our simulations = 0.7095258

3.Show that the distribution is approximately normal.

```r
# draw the histogram
hist(means,
     breaks=50,
     prob=TRUE,
     xlab="Means",
     ylab="Density",
     main=expression(paste("Histogram of means (",lambda, " = 0.2)")),
     col="light blue")

# abline(v = mu.x, col = "blue", lwd = 2, lty = 2, add = TRUE, yaxt = "n")

# add  normal distribution for actual values
curve(dnorm(x,
      mean = mu.x,
      sd = sqrt(var.x)),
      col = "blue",
      lwd = 2,
      lty = 2, add = TRUE,
      yaxt = "n")

# add  normal distribution for theoretical values
curve(dnorm(x,
            mean = mu,
            sd = sqrt(var)),
      col = "red",
      lwd = 2,
```

```r
      add = TRUE,
      yaxt = "n")

# add legend
legend("topright",
       c("actual", "theoretical"),
       lty = c(2,1),
       col = c("blue", "red"))
```



Confidence Intervals

```r
# using 97.5% CI

confidence <- mu + c(-1, 1) * qnorm(.975) * sigma / n

cat("Theoretical Confidence Interval: ", confidence)
```

```
## Theoretical Confidence Interval:  4.961263 5.038737
```

```r
confidence.x <- mu.x + c(-1, 1) * qnorm(.975) * sigma.x / sqrt(n)

cat("Actual Confidence Interval: ", confidence.x)
```

```
## Actual Confidence Interval:  4.738357 5.260431
```

Conclusion

Based on the 1000 simulations performed and the illustrations shown above, we can conclude that the calculated distribution of means of 40 random sampled exponential distributions approximates the normal distribution.