Statistical Inference Project - Part 1

Andrew Mendonca

July 28, 2017

Overview

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. You will also investigate the distribution of averages of 40 exponentials.

Simulations

Load the following library for creating graphs.

```
library(ggplot2)
```

Set the variables for controlling the simulation.

```
set.seed(234)
num_simulation <- 1000
lambda <- 0.2
n <- 40
```

Create a matrix of 1000 simulations each with 40 samples being drawn from the exponential distribution and then calculate the averages.

```
matrix_simulation <- matrix(rexp(num_simulation * n, rate = lambda), num_simulation, n)
mean_simulation <- rowMeans(matrix_simulation)
```

Sample Mean versus Theoretical Mean

Calculate both the actual mean for the sample data and the theoretical mean and compare both of them.

```
sample_mean <- mean(mean_simulation)
sample_mean
```

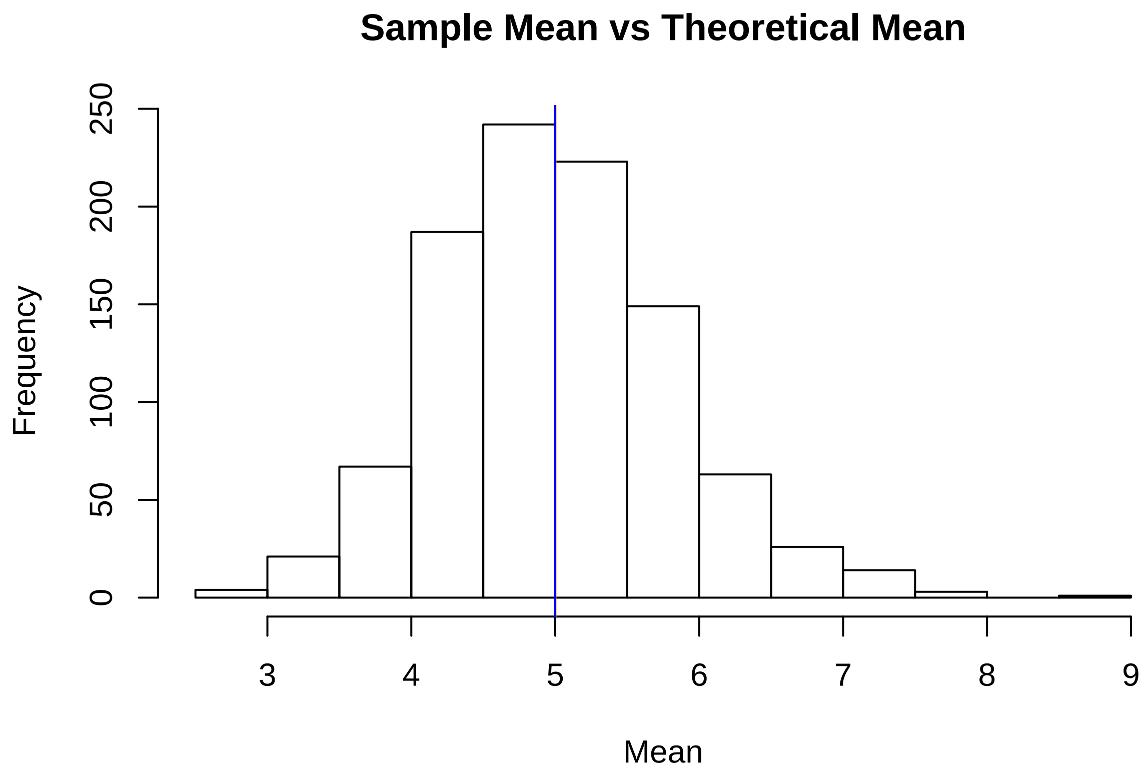
```
## [1] 5.001573
```

```
theoretical_mean <- 1 / lambda
theoretical_mean
```

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

Create a graph to show the comparison between the sample mean and the theoretical mean.

```
hist(mean_simulation, main = "Sample Mean vs Theoretical Mean", xlab = "Mean", ylab = "Frequency")
abline(v = sample_mean, col = "orange")
abline(v = theoretical_mean, col = "blue")
```



The actual mean is 5.001573 and the theoretical mean is 5. This shows that the actual mean from the sample data and the theoretical mean of the normal data are very close to each other.

Sample Variance versus Theoretical Variance

Calculate both the actual variance for the sample data and the theoretical variance and compare both of them.

```
sample_variance <- var(mean_simulation)
sample_variance
```

```
## [1] 0.6631504
```

```
theoretical_variance <- (1 / lambda)^2 / n
theoretical_variance
```

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

The actual variance is 0.6631504 and the theoretical variance is 0.625. This shows that both variances are very close to each other.

Calculate the sample standard of deviation and theoretical standard of deviation.

```
sample_stdDev <- sd(mean_simulation)
sample_stdDev
```

```
## [1] 0.8143405
```

```
theoretical_stdDev <- 1 / (lambda * sqrt(n))
theoretical_stdDev
```

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

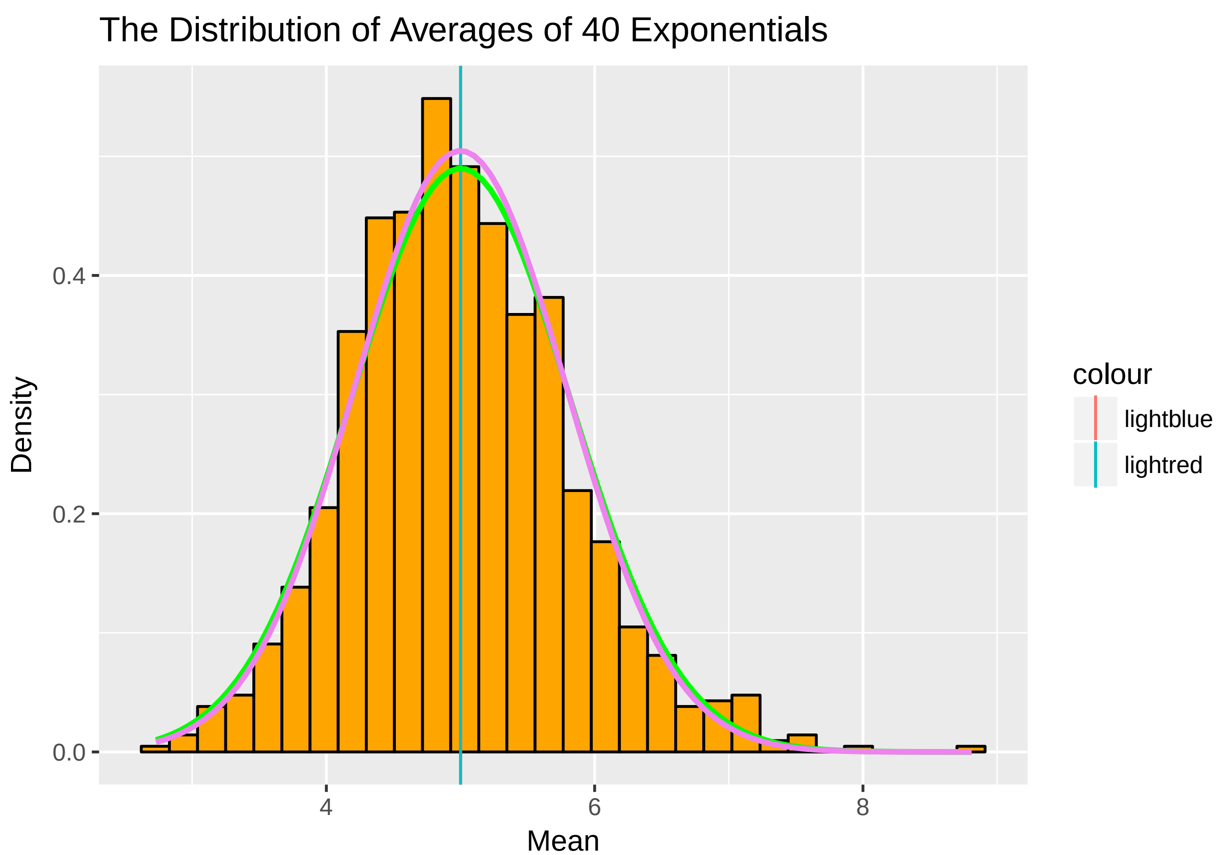
The sample standard of deviation is 0.8143405 and the theoretical deviation is 0.7905694.

Distribution

Create a graph to show that the distribution is approximately normal.

```
data_display <- data.frame(mean_simulation)
ggplot(data_display, aes(x = mean_simulation)) + geom_histogram(aes(y = ..density..),
  colour = "black", fill = "orange") + ggtitle("The Distribution of Averages of 40 Exponentials") +
  xlab("Mean") + ylab("Density") + geom_vline(aes(xintercept = sample_mean, colour = "lightblue")) +
  geom_vline(aes(xintercept = theoretical_mean, colour = "lightred")) + stat_function(fun = dnorm,
  args = list(mean = sample_mean, sd = sample_stdDev), color = "green", size = 1.0) +
  stat_function(fun = dnorm, args = list(mean = theoretical_mean, sd = theoretical_stdDev),
  color = "violet", size = 1.0)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The density of the actual data is shown by orange bars. The sample mean and theoretical mean nearly overlap each other. The violet line shows the normal curve of the theoretical mean and standard deviation. The violet line shows the curve of the sample mean and standard deviation. So, the distribution of averages of 40 exponential distributions is close to the normal distribution.