Statistical Inference - Course Project Part 1

Eduardo Alves

August 09, 2019

Overview

This project investigate the exponential distribution in R and compare it with the Central Limit Theorem.

Simulations

The below code setups the paramters as outline in the course project instructions. This includes the rate (lambda), number of exponentials and the number of simulations we wish to run.

In this part we will investigate the distribution of averages of 40 exponentials over a thousand simulations.

Finally, we plot a histogram of the simulated mean values (Means).

```
# loading necessary R packages
if(!require(ggplot2)){install.packages('ggplot2')}; library(ggplot2)
```

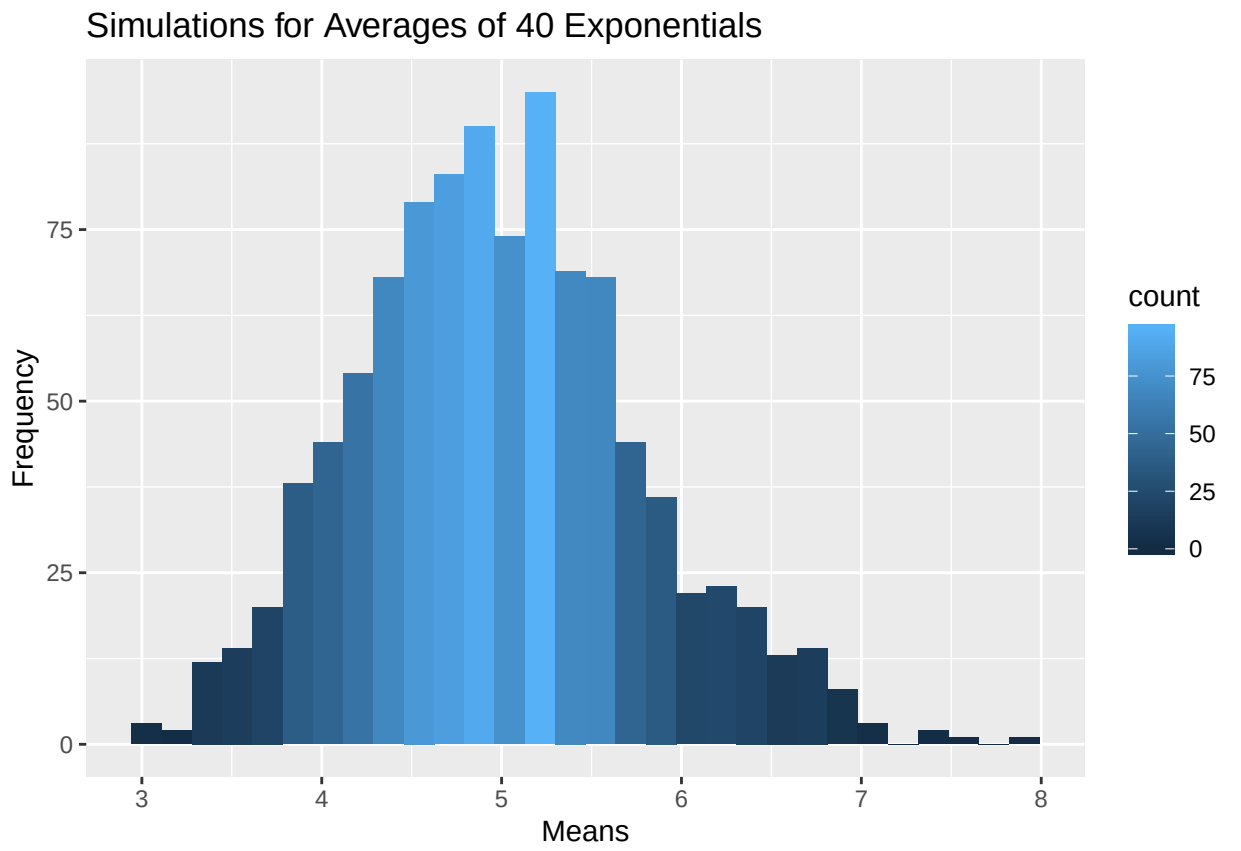
```
## Loading required package: ggplot2
```

```
# parameters
set.seed(135)
lambda <- 0.2
n <- 40
sims <- 1:1000

# Simulate exponential distribution
main.means <- sapply(sims,
         function(x) {
           mean( rexp(n, lambda) )
         })

population <- data.frame(x=main.means)

# Plot the histogram
ggplot(population, aes(x=x)) +
  geom_histogram(aes(y=..count.., fill=..count..), bins = 30) +
  labs(
    title="Simulations for Averages of 40 Exponentials",
    y="Frequency",
    x="Means")
```

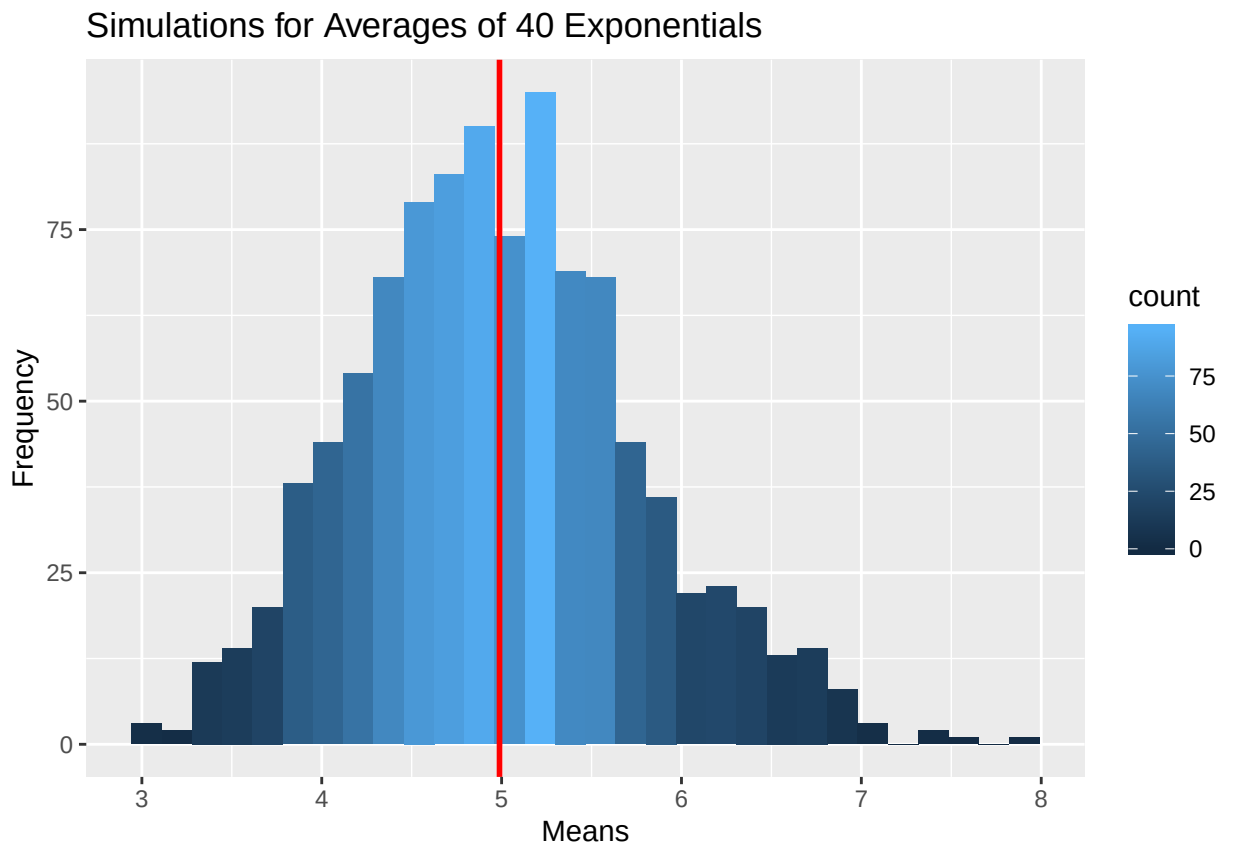


Sample Mean versus Theoretical Mean

The mean of the exponential distribution is 1/lambda. In this case, lambda is 0.2. Therefore, the theoretical mean should result as 5.

```
# calculate sample.means
sample.means <- mean(main.means)

# plot a vertical line at the mean of the sample means
ggplot(population, aes(x=x)) +
  geom_histogram(aes(y=..count.., fill=..count..), bins = 30) +
  geom_vline(xintercept=sample.means, colour="red", linetype="solid", size=1) +
  labs(
    title="Simulations for Averages of 40 Exponentials",
    y="Frequency",
    x="Means")
```




```
# calculate the mean of main.means
mean(main.means)
```

```
## [1] 4.988491
```

The function above shows us that our sample mean is 4.988491 which is pretty close to our theoretical mean of 5.

Sample Variance versus Theoretical Variance

```
# calculate variance
sample.variance <- var(population$x)
theoretical.variance <- ((1/lambda)^2)/n
cbind(sample.variance, theoretical.variance)
```

```
##      sample.variance theoretical.variance
## [1,]       0.6123671                0.625
```

The variance of sample means is 0.6123671 while the theoretical variance is 0.625. They are close to each other.

Distribution

As shown in the graph below, the calculated distribution of means of random sampled exponantial distributions overlaps with the normal distribution, due to the Central Limit Theorem. The more samples we would get, the closer will the density distribution be to the normal distribution bell curve.

```
# plot a density line at the sample means
ggplot(population, aes(x=x)) +
  geom_histogram(aes(y=..density.., fill=..density..), bins = 30) +
  geom_density(colour="red") +
  labs(
    title="Distribution of the Average Exponentials",
    y="Density",
    x="Means")
```