Statististical Inference Course Project - Part 1

Jose Arturo Mora Soto

May 10, 2016

Overview

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Simulation

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.

```
# Simulation settings
lambda = 0.2
num_samples = 40
num_simulations = 1000

# Simulate exponentials
simulated_exp <- replicate(num_simulations, rexp(num_samples, lambda))
# Calculate the mean of exponentials
mean_exp <- apply(simulated_exp, 2, mean)
```

Sample Mean Vs. Theoretical Mean

In this part we will corroborate if the mean of the simulations (sample mean) is similar to the theoretical mean of the distribution.

```
sample_mean <- mean(mean_exp)
print(paste("Sample Mean: ", sample_mean))
```

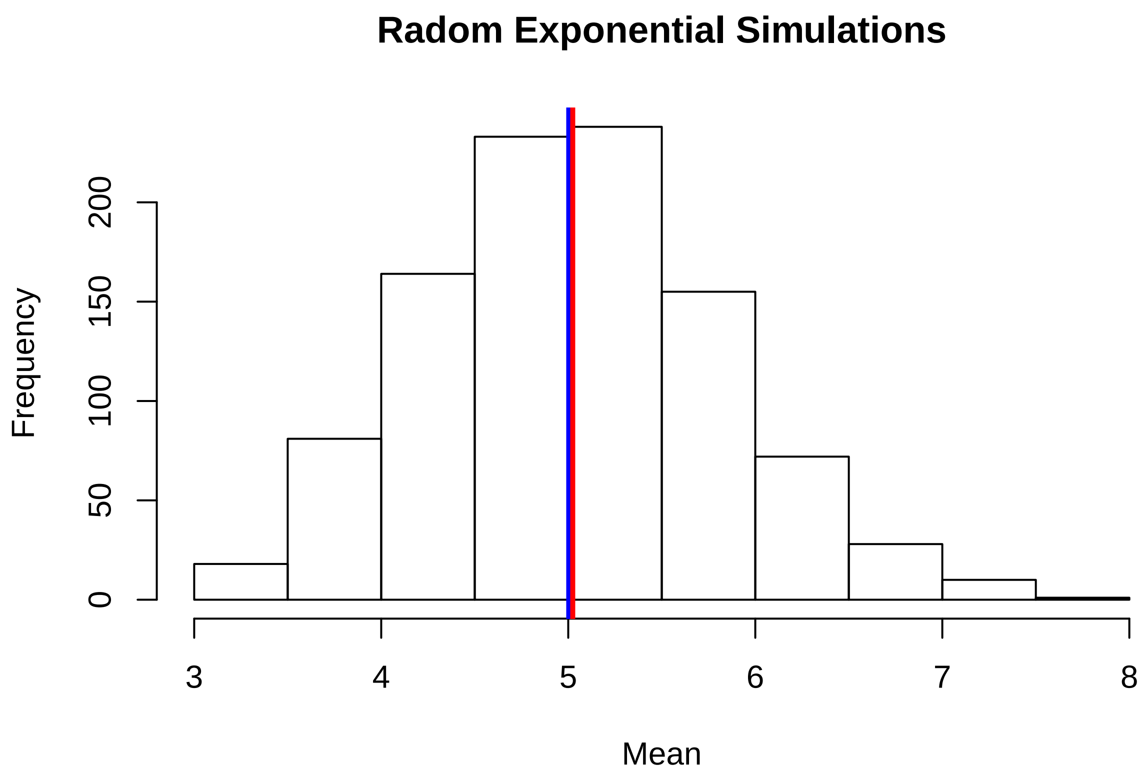
```
## [1] "Sample Mean:  5.02135550982861"
```

```
theoretical_mean <- 1/lambda
print(paste("Theoretical Mean: ", theoretical_mean))
```

```
## [1] "Theoretical Mean:  5"
```

As it can be seen the Central Limit Theorem (CLT) is corroborated since the sample and theoretical means are very similar, now we will plot this results along with an histogram of the sample distribution. The **red line** corresponds to the sample mean, whereas the **blue line** corresponds to the theoretical mean.

```r
# Means visualization
hist(mean_exp, xlab = "Mean", main = "Radom Exponential Simulations")
abline(v = sample_mean, col = "red", lwd = 3)
abline(v = theoretical_mean, col = "blue", lwd = 2)
```



Sample Variance Vs. Theoretical Variance

Now we will compare the sample and theoretical variances to show how variable they are.

```r
sample_variance <- sd(mean_exp)^2
print(paste("Sample Variance: ", sample_variance))
```

```
## [1] "Sample Variance:  0.621703648940988"
```

```r
theoretical_variance <- ((1/lambda) * (1/sqrt(num_samples)))^2
print(paste("Theoretical Variance: ", theoretical_variance))
```

```
## [1] "Theoretical Variance:  0.625"
```

As it can be seen the variance are very similar, this confirms that the sample and theoretical variance are very similar in large sample sizes, so, the mean and variance of the population can be infered by the sample data.

Aproximation to normal distribution

Finally we will show how for large sample sizes, according to the CLT the distributions tend to be normal.

```
x <- seq(min(mean_exp), max(mean_exp), length = 100)
y <- dnorm(x, mean = 1/lambda, sd = (1/lambda) * (1/sqrt(num_samples)))

hist(mean_exp, breaks = num_samples, prob = T, col = "blue", xlab = "Means", ylab = "Density", main = "
lines(x, y, col = "red", lty = 5, lwd = 3)
```