Exponential Distribution and CLT

Yadder Aceituno

January 23, 2018

Overview

This project will show a comparison between the Exponential Distribution and CLT (Central Limit Theorem). I will calculate and compare the mean, the variance, and I will show how the distribution looks approximately normal.

Have in mind that CLT states that the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, if the sample size is large enough.

Simulations

First, I need to create 1000 simulations of the mean from 40 exponential values using lambda = 0.2. I will use the rexp function to get those values. Over that function I will use the replicate function to get the 1000 simulations.

```r
# Seed for reproducibility
SEED <- 0123;
set.seed(SEED);

# Lambda value
lambda <- 0.2;
# Number of values
n <- 40;
# Number of simulations
sim <- 1000;

# Creating the simulated values
simulated_values <- replicate(sim, mean(rexp(n, lambda)));

# Looking the firsts values in the vector
head(simulated_values);
```

```
## [1] 4.811212 5.360077 4.592871 4.900051 5.516619 5.612835
```

Sample Mean vs Theorical Mean

We know that we can calculate the theoric mean with this equation: $E(x) = 1/\lambda = \beta$

```r
# Calculating theoric mean.
theoric_mean <- 1 / lambda;

# Looking the value
theoric_mean;
```

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

Now, I will calculate the mean from my 1000 simulations.

```r
# Calculating simulations' mean
simulation_mean <- mean(simulated_values);

# Looking the value
simulation_mean;
```

```
## [1] 5.011911
```

We can check that both are approximately similar, the theoric mean is equal 5, and the simulations' mean is equal 5.01.

Sample Variance vs Theorical Variance

We know that we can calculate the theoric standard deviation with this equation: $\sigma_x = \sigma/\sqrt{n}$ We also know that $\sigma = \beta$ for exponential distribution. So, we can calculate standard deviation with $\sigma_x = (1/\lambda)/\sqrt{n}$

```r
# Calculating theoric standard deviation.
theoric_standard_deviation <- (1 / lambda) / sqrt(n);

# Looking the value
theoric_standard_deviation;
```

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

Now, I will calculate the standard deviation from my 1000 simulations.

```r
# Calculating simulations' standard deviation
simulation_standard_deviation <- sd(simulated_values);

# Looking the value
simulation_standard_deviation;
```

```
## [1] 0.7749147
```

With those values, we can calculate both variances.

```r
# Calculating  theoric variance
theoric_variance <- theoric_standard_deviation ^ 2;
theoric_variance;
```

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

```r
# Calculating  simulations' variance
simulation_variance <- simulation_standard_deviation ^ 2;
simulation_variance;
```

```
## [1] 0.6004928
```

Notice that both variables (standard deviation and variance) are approximately equal. With those results we can draw the distribution and compare it with a normal distribution.

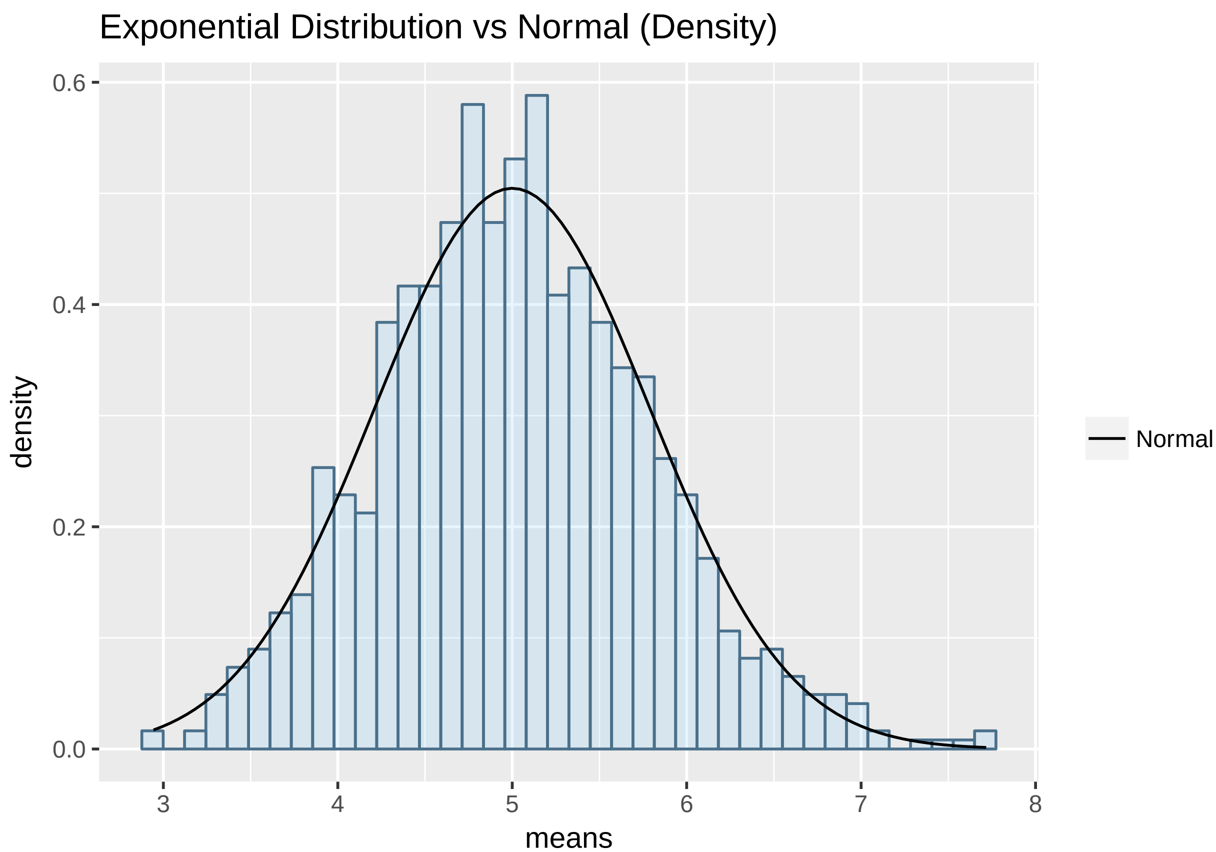
Distribution

Lets compare the distribution of the simulated means and the normal distribution.

```r
# Creating the exponential density histogram.
plot <- ggplot(data = data.frame(means = simulated_values), aes(x = means));
plot <- plot +
```

```
  geom_histogram(aes(y = ..density..),
                 color = "skyblue4",
                 fill = "skyblue1",
                 bins = n,
                 alpha = 0.2) +
  labs(title="Exponential Distribution vs Normal (Density)");

# Creating the normal line density distribution.
plot <- plot +
  stat_function(aes(colour = "Normal"),
                fun = dnorm,
                args = list(mean = theoric_mean, sd = theoric_standard_deviation)) +
  scale_colour_manual("",values = c("black"));
# Plotting
plot;
```



Conclusion

Using random exponential sample of values and a long enough size of data, the sample gets more like a normal distribution. This confirm the central limit theorem.