Exponential Distribution Simulation

Andres Camilo Zuñiga Gonzalez

23/5/2020

As a first step, it is necessary to load the necessary packages and set the working directory.

```
setwd('./Statistical_Inference')
```

```
library(ggplot2) #Plotting system
library(cowplot) #Panel for ggplot2
```

Overview

In the following exercise, I am going to show the properties of the exponential distribuiton and how to simulate this distribution and if the mean of exponential distribution simulation follows a normal distribution.

Simulations

First, we define λ, the number of observations for each simulation and the number of simulations for the `rexp()` function.

```
lambda <- .2
n <- 40
sim <- 1000
```

Then we simulate the data and build a matrix where each row is a simulation and the columns are the observations (sim x n). The `replicate()` function performs `rexp()` the number of times defind in `sim`, while `matrix()` builds the matrix.

```
set.seed(941104)
simulation <- matrix(replicate(sim, rexp(n, lambda)), sim, n)
```

Sample Mean versus Theoretical Mean

First, we must define the theoretical mean of the exponential distribution $(1/\lambda)$ and find the mean for each simulation.

```
teo_mean <- 1 / lambda
sim_mean <- rowMeans(simulation)
```

Then, we build a histogram of the mean for each simulation and draw a vertical line showcasing the theoretical mean, and another line showing the simulated mean of the means.

```
mean_hist <- ggplot() +
  #Histogram
  geom_histogram(aes(x = sim_mean), color = 'gray', fill = 'lightgrey') +
```
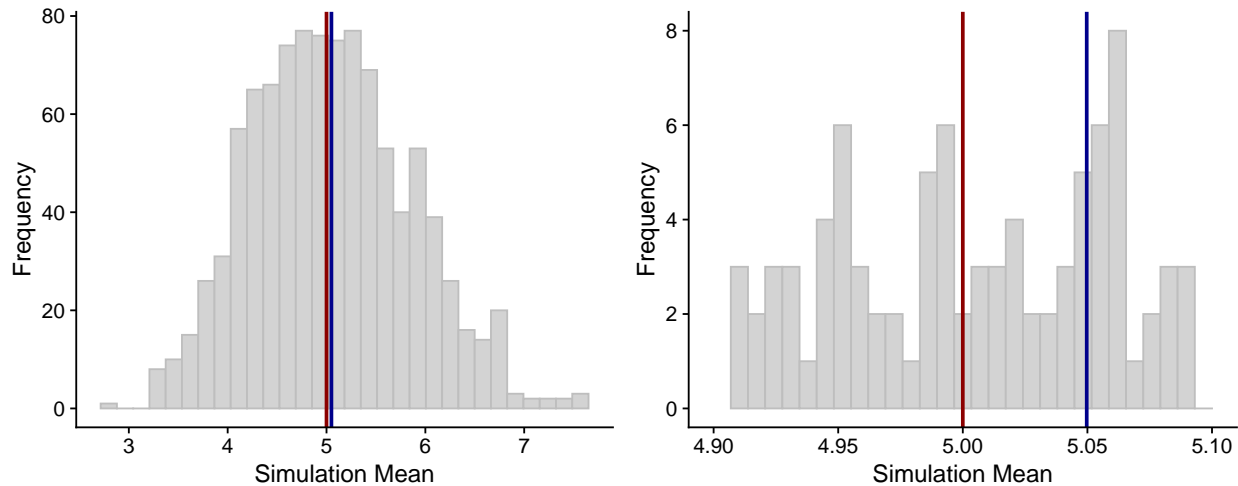
```
  #Simulated and theoretical means
  geom_vline(aes(xintercept = teo_mean), color = 'red4', size = 1) +
  geom_vline(aes(xintercept = mean(sim_mean)), color = 'blue4', size = 1) +

  labs(x = 'Simulation Mean', y = 'Frequency') +
  theme_cowplot()

mean_hist_close <- mean_hist + scale_x_continuous(limits = c(4.9, 5.1))

plot_grid(mean_hist, mean_hist_close)
```


As shown in the histograms above, where the blue line represents the simulated mean 5.0497235 and the red shows the theoretical mean 5, they are very close to each other. In fact, they differ by just -0.0497235. The histogram on the left shows the distribution of the mean for each simulation, while the histogram on the right is a zoomed in view of the former showing the difference between the expected mean and the simulated mean.

Sample Variance versus Theoretical Variance

For this part, first we must define the theoretical and simulated variances and standard deviations

```
sim_var <- var(sim_mean)
sim_sd <- sd(sim_mean)

teo_var <- (1 / lambda) ** 2 / n
teo_sd <- (1 / lambda) / sqrt(n)
```

According to these calculations, as shown in the table below, it is possible to see that the simulated values are very similar to the theoretical ones of the distribution.

	Simulated	Theoretical
Variance	0.649	0.625
Standard Deviation	0.806	0.791

Distribution

Finally, we find the 95% confidence intervals for both the simulated and theoretical values of the mean using the following formula:

$$C.I_{95\%} = \mu \pm 1.96 * \frac{SD}{\sqrt{n}},$$

where 1.96 is the 0.975 quantile of a normal distribution, SD is the standard deviation and n is the number of observations.

```
sim_ci95 <- round(mean(sim_mean) + c(-1, 1) * 1.96 *(sim_sd) / sqrt(n) , 3)
teo_ci95 <- round(teo_mean + c(-1, 1) * 1.96 *(teo_sd) / sqrt(n) , 3)
```

The 95% confidence intervals are:
* Simulation: [**4.8, 5.299**]
* Theoretical: [**4.755, 5.245**]

For the final part, we take the distribution shown above, but convert the frequency to density. In addition, with the `dnorm()` we create density values of a normal distribution that spans the same range as the simulated. Besides, this the density plot is drawn using the simulated data. Just like the histogram above, the red lines show the mean and density of a normal distribution, and blue lines show the same parameters for the simulated means.

```
norm_x <- seq(min(sim_mean), max(sim_mean), 0.01)
sim_normal <- dnorm(norm_x, mean = teo_mean, sd = teo_sd)

ggplot() +
  #Density histogram plot
  geom_histogram(aes(x = sim_mean, y = ..density..), color = 'gray', fill = 'lightgrey') +

  #Simulated and theoretical means
  geom_vline(aes(xintercept = teo_mean), color = 'red4', size = 1) +
  geom_vline(aes(xintercept = mean(sim_mean)), color = 'blue4', size = 1) +

  #Simulated and normal distribution density plots
  geom_density(aes(x = sim_mean), color = "blue3", size = 1) +
  geom_line(aes(x = norm_x, y = sim_normal), color = "red3", size = 1) +

  #95% Confidence intervals
  geom_vline(aes(xintercept = sim_ci95), color = 'red4', size = 1, linetype = 2) +
  geom_vline(aes(xintercept = teo_ci95), color = 'blue4', size = 1, linetype = 2) +

  labs(x = 'Mean', y = 'Density') +
  theme_cowplot()
```
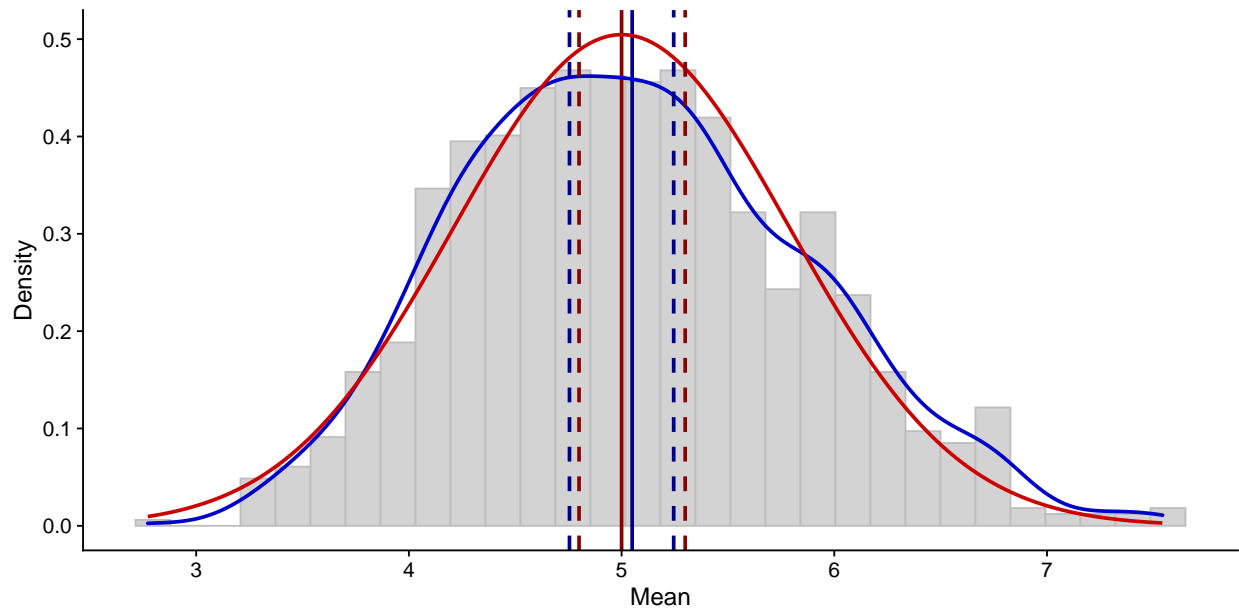
As shown in this plot, it is possible to see that the distribution of the means of the simulated exponential distributions is close to a normal distribution. Nonetheless, more simulations or more values per simulation are necessary to make it closer to the normal distribution. Notice the 95% confidence intervals in dashed lines.