Part 1 - Simulation Exercise

Praffull Agrawal

July 7, 2020

Overview

In this project, we undertake a simulation exercise to investigate the **Exponential** distribution and compare it with the Central Limit Theorem. We will invesigate the distribution of the **averages** of *40* exponentials. And, we will run *1000* simulations.

Simulations

Initially, set the **seed** for reproducibility.

```
set.seed(1)
```

Define the various parameters.

```
lambda <- 0.2    # Rate parameter
n <- 40          # Number of samples
B <- 1000        # Number of simulations
```

Sample the exponentials *1000* times (*40* samples each) and store in `mat`.

```
mat <- matrix(rexp(n*B, lambda), B, n)
```

Calculate the **average** for each of the *1000* simulations (*40* samples each) and store in `avg`.

```
avg <- apply(mat, 1, mean)
```

Sample Mean versus Theoretical Mean

Sample mean.

```
sam_mean <- mean(avg)
print(round(sam_mean, 2))
```

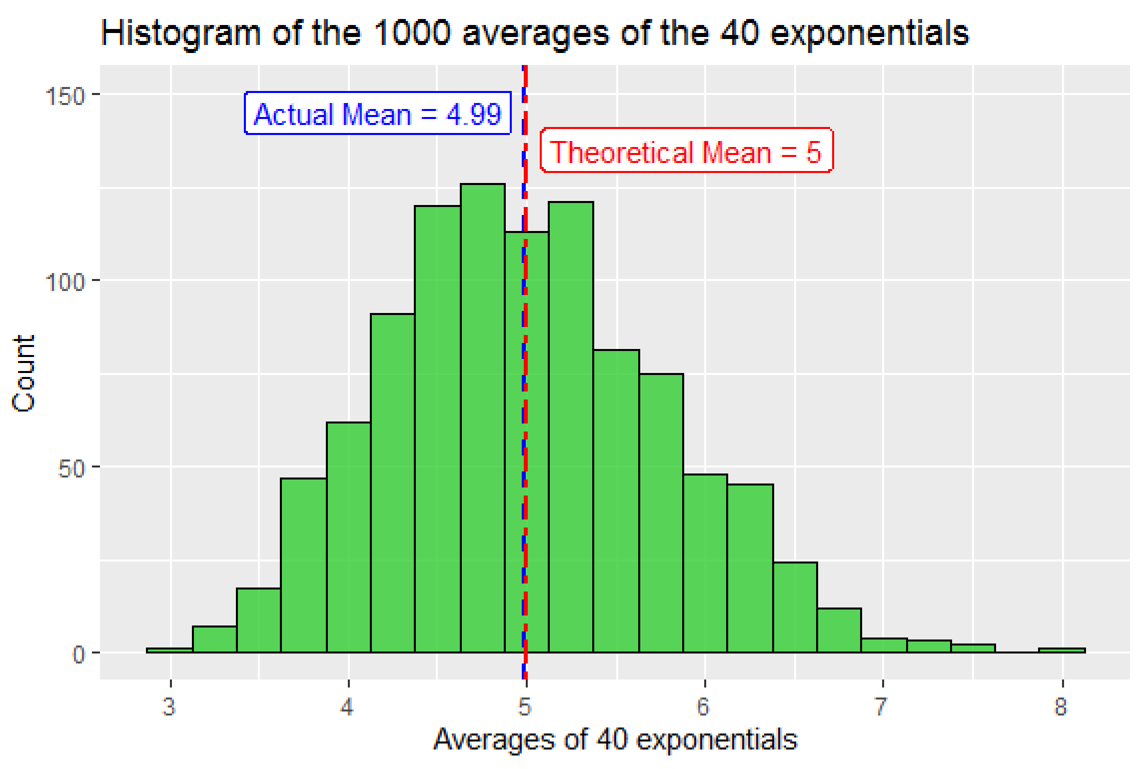
```
## [1] 4.99
```

Theoretical mean.

```
theo_mean <- 1/lambda
print(theo_mean)
```

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

Look at the *histogram* of the *1000* **averages** of the *40* exponentials.



Hence, the *sample mean* and the *theoretical mean* are approximately equal.

Sample Variance versus Theoretical Variance

Sample variance.

```
sam_var <- var(avg)
print(round(sam_var, 2))
```

```
## [1] 0.62
```

Theoretical variance.

```
theo_var <- (1/lambda)^2
print(theo_var)
```

```
## [1] 25
```

The relation between the *sample variance* and *theoretical variance* is given as:

$$S^2 = \frac{\sigma^2}{n}$$

The calculated sample variance:

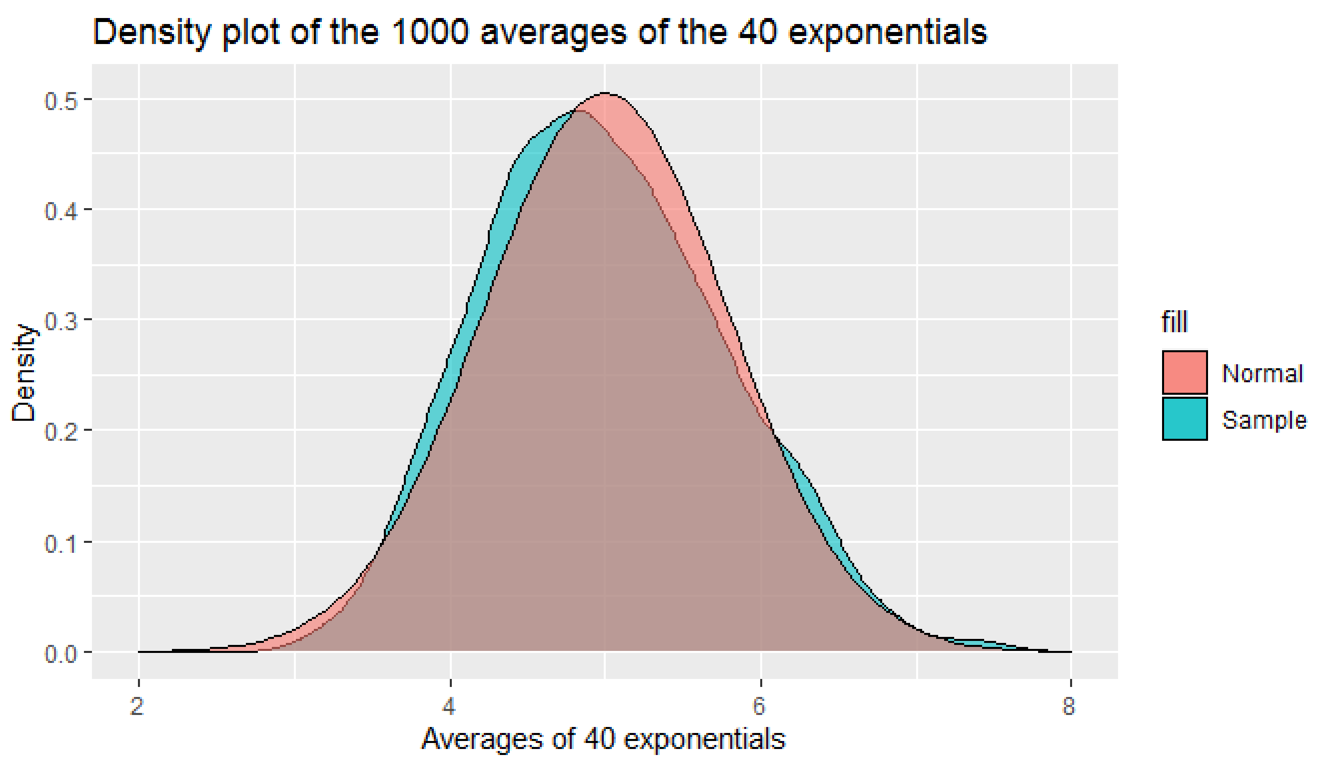
```
cal_sam_var <- ((1/lambda)^2)/n
print(round(cal_sam_var, 2))
```

```
## [1] 0.62
```

Hence, the *sample variance* and the *theoretical variance* follow the above relation approximately.

Distribution

Look at the *density plot* of the *1000* **averages** of the *40* exponentials.



There is a large overlap between the *Sample density curve* and *Normal density curve* and therefore, the distribution is approximately normal.

Appendix

Load the packages

```
library(ggplot2)
library(dplyr)
```

Code for `Plot-01.png`.

```
g <- as.data.frame(avg) %>%
      ggplot(aes(avg)) +
        geom_histogram(binwidth = 0.25, colour = "black", alpha = 0.8, fill="limegreen") +
        geom_vline(xintercept = mean(avg), lwd = 1, lty = 2, colour = "blue") +
        geom_label(aes(x = mean(avg), y = 145, label = paste0("Actual Mean = ", round(mean(avg), 2))),
                  colour="blue", hjust = 1.05, text=element_text(size=10)) +
        geom_vline(xintercept = 1/lambda, lwd = 1, lty = 6, colour = "red") +
        geom_label(aes(x = 1/lambda, y = 135, label = paste0("Theoretical Mean = ", round(1/lambda, 2))),
                  colour="red", hjust = -0.05, text=element_text(size=10)) +
        ylim(c(0, 150)) +
        labs(title = "Histogram of the 1000 averages of the 40 exponentials",
             x = "Averages of 40 exponentials",
             y = "Count")

print(g)
```

Code for `Plot-02.png`.

```
g <- as.data.frame(avg) %>%
      ggplot(aes(avg)) +
        geom_density(alpha = 0.6, aes(fill = "Sample")) +
        geom_area(stat = "function", fun = dnorm, args = list(mean = 1/lambda, sd = (1/lambda)/sqrt(n)),
                  colour = "black", alpha = 0.6, aes(fill = "Normal")) +
        xlim(c(2, 8)) +
        scale_colour_manual(name="fill", values=c(Sample="lightblue", Normal="salmon")) +
        labs(title = "Density plot of the 1000 averages of the 40 exponentials",
             x = "Averages of 40 exponentials",
             y = "Density")

print(g)
```