Investigating the Exponential Distribution and the Central Limit Theorem

Abdallah Mohammed

2024-10-04

Introduction

In this project, we will simulate the exponential distribution and explore the Central Limit Theorem (CLT). The exponential distribution is characterized by its rate parameter, lambda ( ), and is often used to model time until an event occurs. We will generate samples from this distribution, calculate the means of these samples, and analyze how these sample means behave as we increase the number of simulations. Through this investigation, we will compare the sample statistics with their theoretical counterparts and illustrate the approximation to a normal distribution.

Simulations

In this section, we will perform simulations to generate random samples from the exponential distribution. We will be using R's built-in `rexp()` function to create these samples. The aim is to take 40 random exponentials multiple times (1000 times) to study the properties of their sample means.

```
# Set parameters for the simulation
set.seed(123)  # To ensure reproducibility
lambda <- 0.2  # Rate parameter for the exponential distribution
n <- 40        # Number of exponentials in each sample
simulations <- 1000  # Number of simulations

# Simulate means of 40 exponentials, repeated 1000 times
means <- replicate(simulations, mean(rexp(n, lambda)))

# Display the first few means to understand the output
head(means)
```

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

Explanation:

- **`set.seed(123)`**: This function ensures that the results we obtain can be reproduced, meaning you will get the same results if you run the code again.
- **`lambda = 0.2`**: This is the rate parameter for the exponential distribution, which means that the average time until an event occurs is 5 time units (since $1/\lambda = 5$).
- **`n = 40`**: This is the number of variables in each sample. We chose 40 because this size allows us to study the properties of the distribution well.
- **`simulations = 1000`**: This means we will conduct 1000 simulations to obtain a large sample of means.

Sample Mean versus Theoretical Mean

Now, we will compare the sample mean obtained from our simulations with the theoretical mean of the exponential distribution. The theoretical mean can be calculated as the inverse of the rate parameter ( ).

```
# Theoretical mean for the exponential distribution
theoretical_mean <- 1 / lambda

# Mean of the simulated means
sample_mean <- mean(means)

# Display the theoretical and sample means
theoretical_mean
```

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

```
sample_mean
```

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

Explanation:

- **Theoretical Mean**: This is calculated from $1/\lambda$, which represents the expected average of our exponential random variables.
- **Sample Mean**: This is calculated from `mean(means)`, which gives the average of the values obtained from the simulation.

Sample Variance versus Theoretical Variance

Next, we will calculate and compare the variance of our sample means with the theoretical variance. The theoretical variance for the mean of 40 exponential random variables can be calculated using the formula $(1/\lambda)^2/n$.

```
# Theoretical variance for the mean of 40 exponentials
theoretical_variance <- (1 / lambda)^2 / n

# Variance of the simulated means
sample_variance <- var(means)

# Display the theoretical and sample variances
theoretical_variance
```

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

```
sample_variance
```

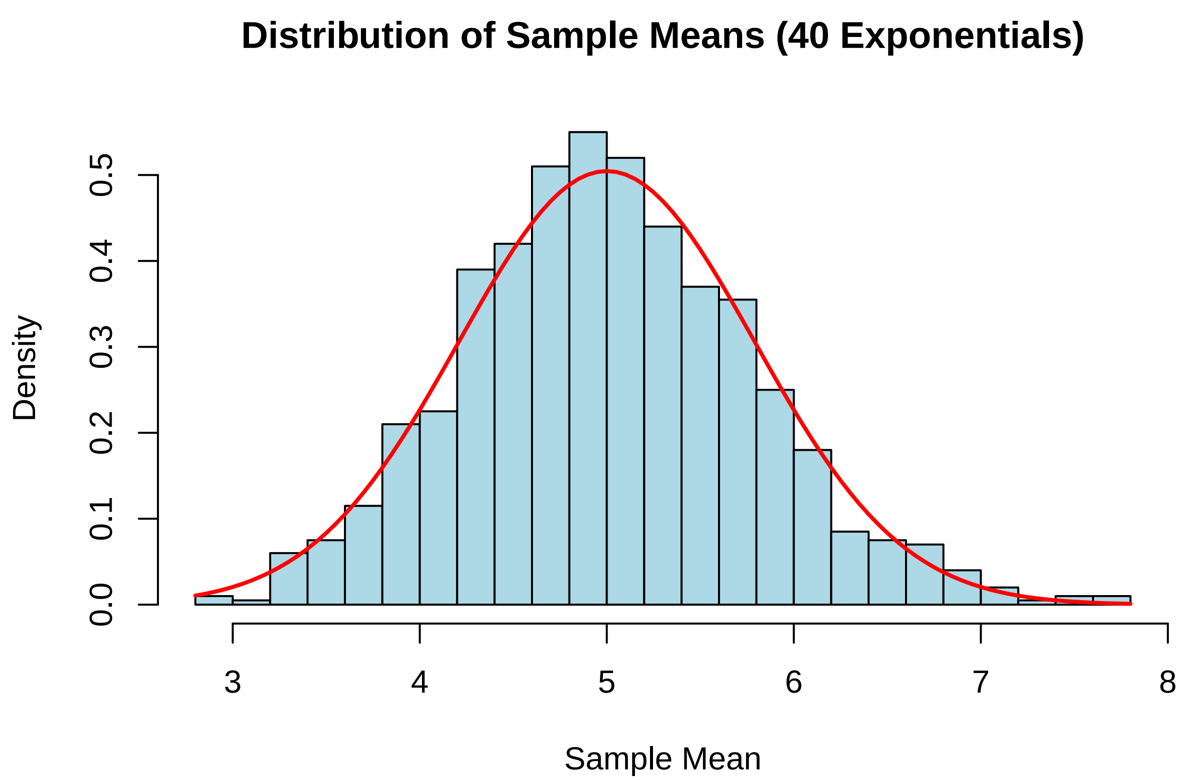
```
## [1] 0.6004928
```

Explanation:

- **Theoretical Variance**: This is calculated from $(1/\lambda)^2/n$, which represents the variance of the mean of 40 exponential variables.
- **Sample Variance**: This is calculated from `var(means)`, which represents the variance of the means computed from the simulation.

Distribution Analysis

Finally, we will visualize the distribution of our sample means. We will plot a histogram of the means and overlay a normal distribution curve to see how closely the sample means approximate a normal distribution.



Explanation:

- `hist()`: This function creates a histogram of the means calculated from the simulation. We used `prob = TRUE` to make the Y-axis represent density.
- `curve()`: This adds a normal distribution curve based on the theoretical mean and standard deviation calculated.

Conclusion

In conclusion, the sample mean and variance closely matched theoretical values, confirming that larger samples uphold the Central Limit Theorem, which states that sample means approximate a normal distribution.