A simulation exercise

Vitalii Diakonov

2024-03-21

Overview:

This report delves into the properties of the exponential distribution in R and its comparison with the Central Limit Theorem (CLT). The investigation focuses on the distribution of averages of 40 exponentials across a thousand simulations.

In summary, this report will:

- Show the sample mean and compare it to the theoretical mean of the distribution
- Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution
- Show that the distribution is approximately normal

Simulation:

The simulation process involves utilizing the **`rexp(n, lambda)`** function in R, where lambda represents the rate parameter. Notably, the mean and standard deviation of the exponential distribution are both equivalent to 1/lambda. Throughout this report, lambda remains constant at 0.2.

Generating Sample Exponential Distribution:

```
set.seed(123) # Set seed for reproducibility
num_simulations <- 1000  # Number of simulations
num_exponentials <- 40   # Number of exponentials
rate_param <- 0.2        # Rate parameter

simulated_data <- matrix(rexp(num_simulations * num_exponentials, rate = rate_param), num_
```

```r
simulated_means <- rowMeans(simulated_data)

# Calculate sample mean, standard deviation, and variance
sample_mean <- mean(simulated_means)
sample_std_dev <- sd(simulated_means)
sample_variance <- var(simulated_means)
```

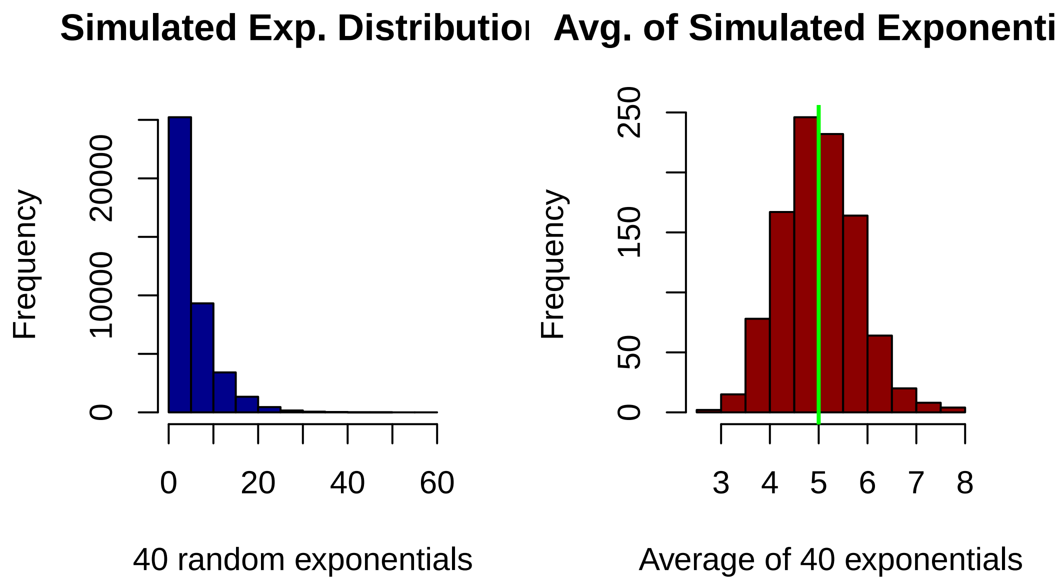
Theoretical Exponential Distribution:

```r
# Calculate theoretical mean, standard deviation, and variance
theoretical_mean <- 1/rate_param
theoretical_std_dev <- (1/rate_param) * (1/sqrt(num_exponentials))
theoretical_variance <- theoretical_std_dev^2
```

Analysis

Histogram Comparison:

```r
# Set up multi-panel plotting
par(mfrow = c(1, 2))

# Plot the simulated exponential distribution
hist(simulated_data, col = "darkblue", main = "Simulated Exp. Distribution", xlab = "40 ra

# Plot the averages of simulated exponentials
hist(simulated_means, col = "darkred", main = "Avg. of Simulated Exponentials", xlab = "Av

# Add vertical line for theoretical mean
abline(v = theoretical_mean, col = "green", lwd = 2)
```



The left histogram displays the simulated exponential distribution, while the right histogram represents the means of simulated exponentials. The green line indicates the theoretical mean, which closely aligns with the simulated distribution.

Comparison of Sample and Theoretical Statistics:

```
Sample_statistics <- c(sample_mean, sample_std_dev, sample_variance)
Theoretical_statistics <- c(theoretical_mean, theoretical_std_dev, theoretical_variance)
differences <- c(abs(theoretical_mean - sample_mean), abs(theoretical_std_dev - sample_std
names <- c("Mean", "Standard Deviation", "Variance")
data.frame(Sample_statistics, Theoretical_statistics, differences, row.names = c("Mean", "
```

```
                   Sample_statistics Theoretical_statistics differences
Mean                       5.0119113              5.0000000  0.01191128
Standard Deviation         0.7802751              0.7905694  0.01029432
Variance                   0.6088292              0.6250000  0.01617077
```

- The sample mean is slightly higher than the theoretical mean by approximately 0.012 units.
- The sample standard deviation is slightly lower than the theoretical standard deviation by approximately 0.010 units.

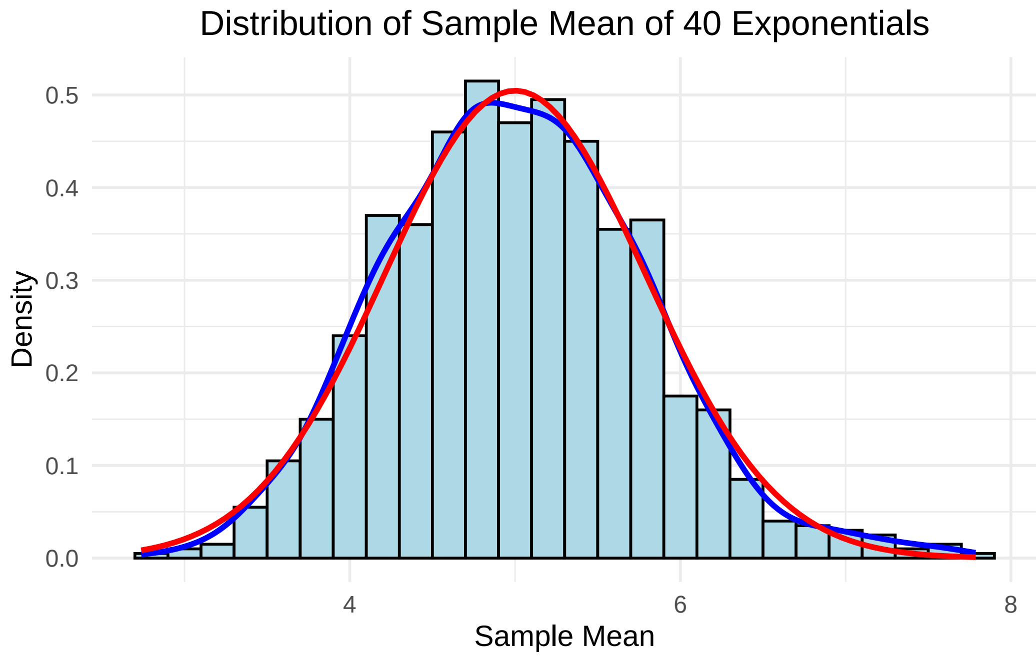
- The sample variance is slightly lower than the theoretical variance by approximately 0.016 units.

Overall, the sample statistics are very close to their theoretical counterparts, indicating that the simulated data aligns well with the expected values from the theoretical distribution.

Distribution Assessment

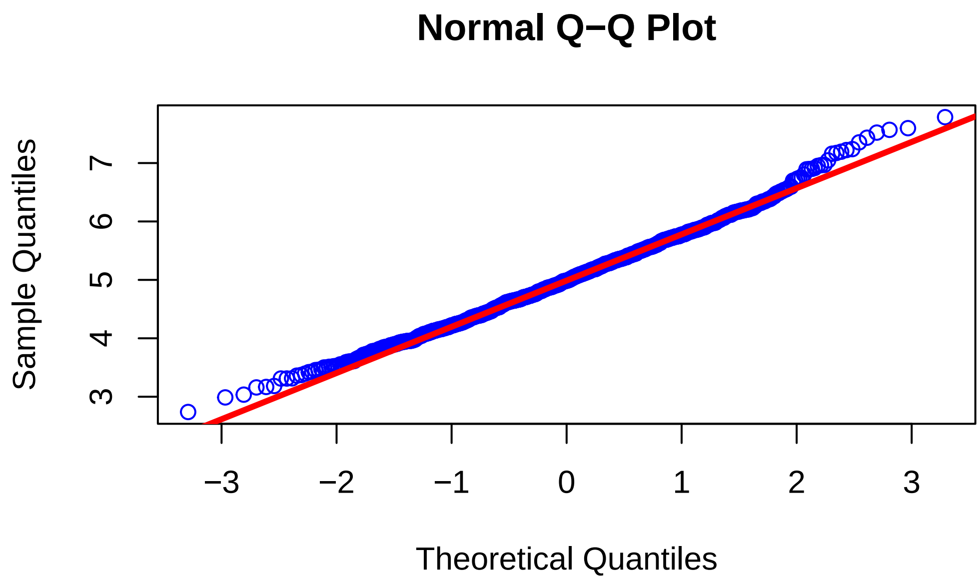
Histogram and Density Plot:

```
# Load the ggplot2 library
library(ggplot2)

# Create a data frame for the simulated means
simulated_means_df <- data.frame(simulated_means)

# Create a histogram and density plot
ggplot(simulated_means_df, aes(x = simulated_means)) +
    geom_histogram(binwidth = .2, fill = "lightblue", color = "black", aes(y = after_stat(
    geom_density(color = "blue", lwd = 1) +  # Density plot
    labs(title = "Distribution of Sample Mean of 40 Exponentials", x = "Sample Mean", y =
    stat_function(fun = dnorm, args = list(mean = theoretical_mean, sd = theoretical_std_d
    theme_minimal() +
    theme(plot.title = element_text(hjust = 0.5))
```



Q-Q Plot:

```
qqnorm(simulated_means, col = "blue") # Sample distribution
qqline(simulated_means, col = "red", lwd = 3) # Theoretical
```



The Q-Q plot indicates a close approximation between the sample distribution and the theoretical normal distribution, particularly in the central region.

Conclusion

Based on the analysis and comparison of the simulated sample distribution with the theoretical distribution, it is evident that the Central Limit Theorem holds true. Despite slight discrepancies in mean and variance, the overall distributions exhibit close alignment, thereby validating the applicability of the CLT.