Exponential Simulation Exercises for Sample Mean and Variance

Sebastian Gomez Gomez

Report - Part I: Simulation Exercises

Overview

This report includes simulation exercises for an exponential distribution, and its sample mean and variance. This includes generating random data, with a specific seed for reproducibility purposes, and their statistical comparison based on the Law of Large Numbers and the Central Limit Theorem.

Simulations

As per the instructions provided, all exponential simulations will have the following parameters:

```
set.seed(100); lambda <- 0.2; n <- 1000; navg <- 40
```

NOTE: I have taken the liberty of adding a specific seed for reproducibility purposes.

Below you will find the code to simulate n exponential observations, which would help establish a baseline of what an empiric exponential distribution should look like:

```
exp_obs <- rexp(n, rate = lambda)
```

Also, the following code represents a simulation of n sample means and sample variances obtained from (n x navg) exponential observations:

```
exp_sim <- matrix(rexp(n*navg, rate = lambda), nrow = n, ncol = navg)
exp_avg <- exp_var <- NULL
for (i in 1:n) {
        exp_avg <- c(exp_avg, mean(exp_sim[i, ])); exp_var <- c(exp_var, var(exp_sim[i, ]))
}
```

Sample Mean vs. Theoretical Mean

As part of the supporting material, you will find the code and plotting of the “Baseline” simulated observations as well as the sample mean frequency distributions (histogram). They include also a line for their respective mean (in red), as well as the theoretical population mean (1/lambda - in green) values.

NOTE: It is important to mention that because the sample mean mean and theoretical mean are so close to each other, the only visible line is the theoretical one (in blue).

Now, here are the actual sample mean and theoretical mean values:

```
c(Sample_Mean = mean(exp_avg), Theoretical_Mean = (1/lambda))
```

```
##      Sample_Mean Theoretical_Mean
##         4.997191         5.000000
```

Sample Variance vs. Theoretical Variance

As part of the supporting material, you will find the code and plotting of the sample variance frequency distribution (Histogram), as well as the sample variance mean (in red) and theoretical exponential distribution population variance (1/lambda^2 - in green) values.

Now, here are the actual values for sample variances mean and theoretical population variance:

```
c(Sample_Variance_Mean = mean(exp_var), Theoretical_Variance = (1/lambda)^2)
```

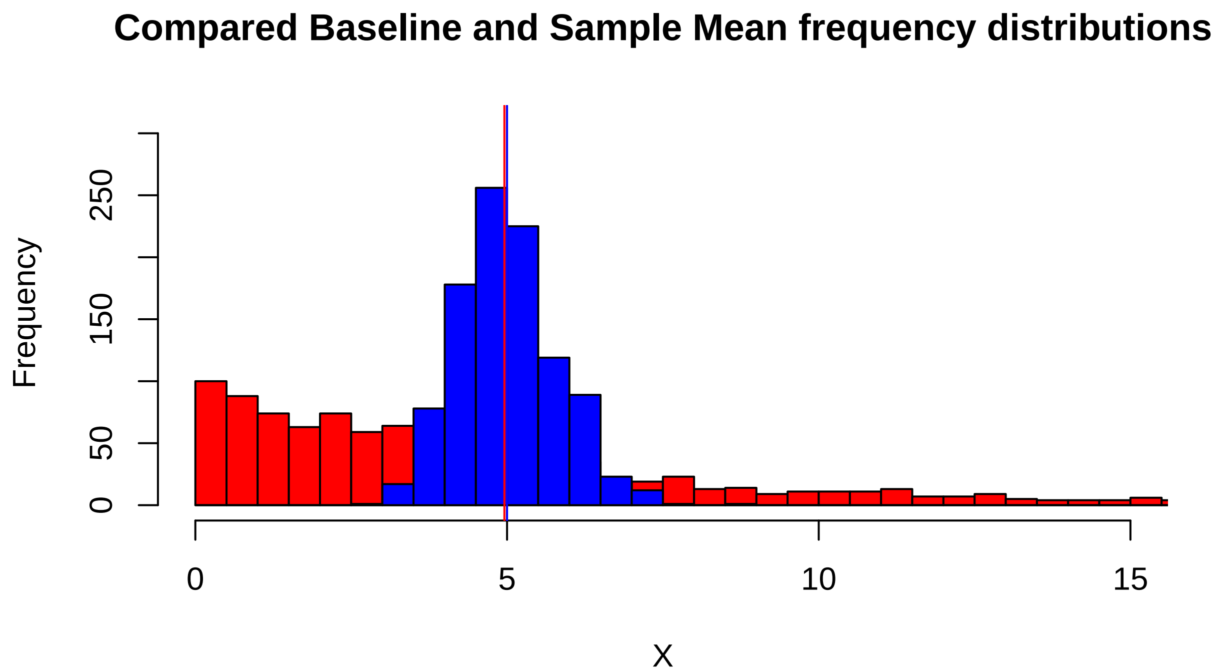
```
## Sample_Variance_Mean Theoretical_Variance
##             25.35603             25.00000
```

Distribution

Below is the code and figure of the actual overlapping plot of the "Baseline" and Sample mean frequency distributions, including vertical lines for sample mean mean, and median values, and of the theoretical (1/lambda) exponential distribution mean.

It is important to highlight the fact that the "Baseline" frequency is shaped approximately as a theoretical exponential distribution whereas the sample mean frequency histogram reflects a shape similar to a normal distribution.

```
bas <- hist(exp_obs, breaks = seq(0, max(exp_obs)+1, 0.5), plot = FALSE)
sam_mean <- hist(exp_avg, seq(0, max(exp_obs)+1, 0.5), plot = FALSE)
plot(bas, col = 'red', ylim = c(0, 310), xlim = c(0, 15),
     main = 'Compared Baseline and Sample Mean frequency distributions', xlab = 'X')
plot(sam_mean, col = 'blue', add = TRUE)
abline(v = c(mean(exp_avg), median(exp_avg), 1/lambda),
     col = c('red', 'red', 'blue'), lwd=c(1, 1, 1))
```



Now, here are the actual values for the "baseline", sample mean and theoretical population means:

```
c(Baseline_Mean = mean(exp_obs), Baseline_Median = median(exp_obs),
        Sample_Mean_Mean = mean(exp_avg), Sample_Mean_Median = median(exp_avg),
        Theoretical_Mean = 1/lambda)
```

```
##      Baseline_Mean    Baseline_Median   Sample_Mean_Mean Sample_Mean_Median
##           4.857406           3.367034           4.997191           4.956790
##   Theoretical_Mean
##           5.000000
```

Conclusions

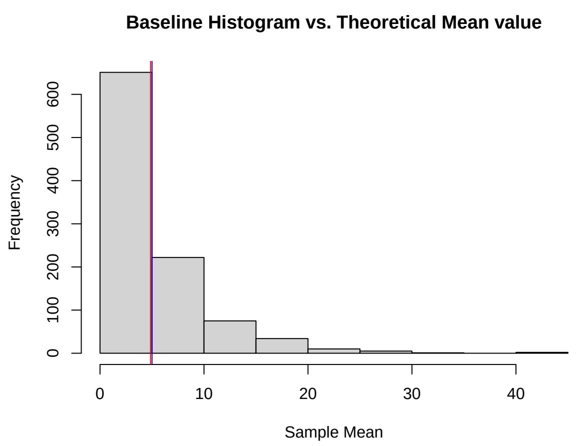
As you can see, following the idea of the Law of Large Numbers, both the "baseline" and sample means are consistent (approximately equal) to that of the population theoretical mean.

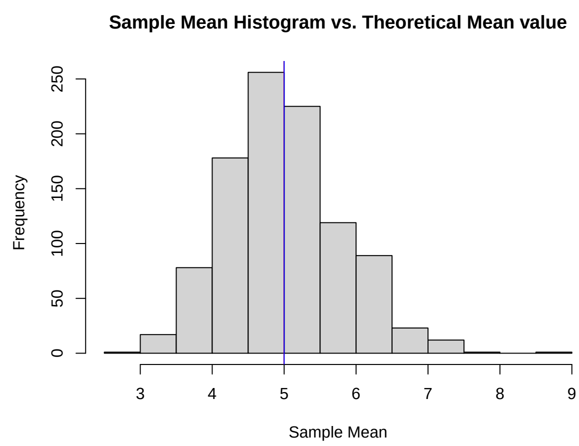
Similarly, as per the Central Limit Theorem, the distribution of sample means for this nvg sample size appears to be approximately normal. This is demonstrated also by the fact that the sample mean median is almost the same as its average, as well as the theoretical population average, which reflects the symmetry of the normal distribution (in comparison to the "Baseline" exponential values median, which is significantly lower than its mean).

Supporting Appendix Material

Sample Mean vs. Theoretical Mean

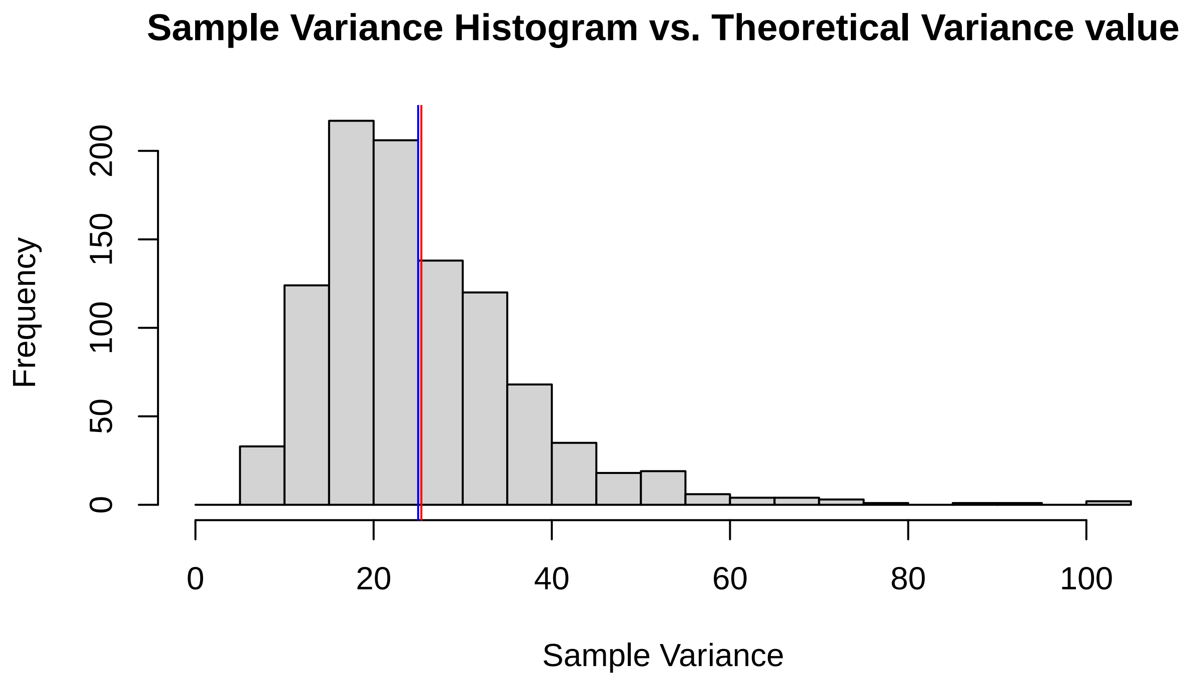
```
print(hist(exp_obs, main = 'Baseline Histogram vs. Theoretical Mean value',
           xlab = 'Sample Mean'));
abline(v = c(mean(exp_obs), 1/lambda), col = c('red', 'blue'), lwd=c(1, 1))
print(hist(exp_avg, main = 'Sample Mean Histogram vs. Theoretical Mean value',
           xlab = 'Sample Mean'))
abline(v = c(mean(exp_avg), 1/lambda), col = c('red', 'blue'), lwd=c(1, 1))
```





Sample Variance vs. Theoretical Variance

```
print(hist(exp_var, breaks = seq(0, 105, 5),
           main = 'Sample Variance Histogram vs. Theoretical Variance value',
           xlab = 'Sample Variance'))
abline(v = c(mean(exp_var), (1/lambda)^2), col = c('red', 'blue'), lwd=c(1, 1))
```



Sample Variance Histogram vs. Theoretical Variance value
Frequency
0
50
100
150
200
0
20
40
60
80
100
Sample Variance