PENDING TITLE - Statistical Inference - Course Project

Sebastian Gomez Gomez

Overview

Part I: Simulation Exercises

Simulations

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

```
## Setting seed and other simulation 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:

```
## Simulating n exponential observations with rate = lambda
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:

```
## Simulating n averages and variances of navg exponential observations with rate = lambda
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

Below is the code and plotting of the sample mean frequency distribution (Histogram) as well as the sample mean (in blue) and theoretical exponential distribution population mean (1/lambda - in green) values.

```
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('blue', 'green'), lwd=c(1, 3))
```

Sample Mean Histogram vs. Theoretical Mean value
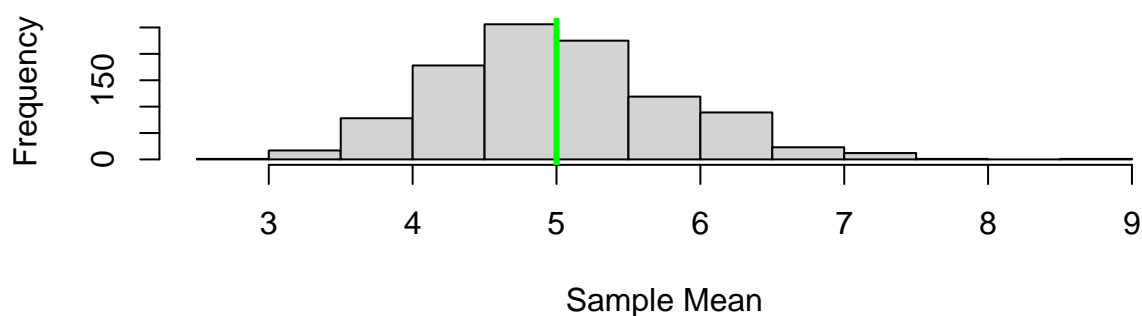
Here are the actual sample mean and theoretical mean values:

```r
c(Sample_Variance_Mean = mean(exp_var), Theoretical_Variance = (1/lambda)^2)
```
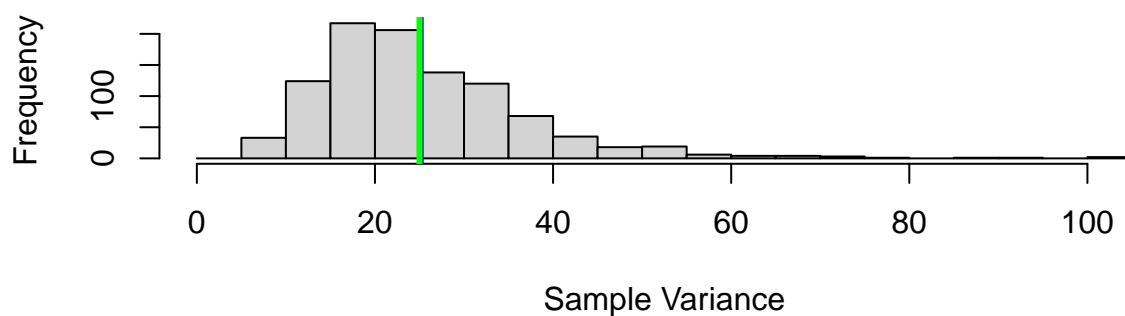
```
## Sample_Variance_Mean Theoretical_Variance
##             25.35603             25.00000
```

Sample Variance vs. Theoretical Variance

Below is the code and plotting of the sample variance frequency distribution (Histogram) as well as the sample variance mean (in blue) and theoretical exponential distribution population variance (1/lambda^2 - in green) values.

```r
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('blue', 'green'), lwd=c(1, 3))
```

Sample Variance Histogram vs. Theoretical Variance value

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 their mean values, and of the theoretical (1/lambda) exponential distribution mean.

```
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_obs), median(exp_obs), mean(exp_avg), median(exp_avg), 1/lambda),
       col = c('red', 'purple', 'blue', 'purple', 'green'), lwd=c(1, 1, 1, 1, 3))
```

Compared Baseline and Sample Mean frequency distributions
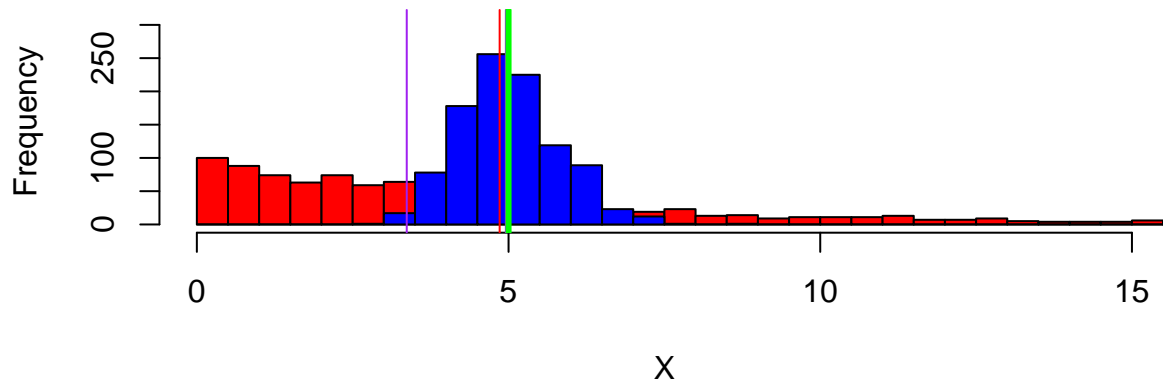
Now, here are the actual values for the "baseline", sample mean and theoretical population means:

```
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).

Part II: Basic Inferential Data Analysis

Loading data

Basic data summary

Loading data

Comparing tooth growth by supp and dose

Conclusions

Part III: Supporting Appendix Material

xxxxx

```
print(hist(exp_obs));abline(v = c(mean(exp_obs), 1/lambda), col = c('red', 'green'), lwd=c(1, 3))
print(hist(exp_avg))
abline(v = c(mean(exp_avg), 1/lambda), col = c('blue', 'green'), lwd=c(1, 3))
```


Histogram of exp_obs

Histogram of exp_avg