Central Limit Theorem Simulations on the Exponential Distribution

Statistical Inference – Class Project – Part 1
Author: Josh Jensen

Overview:
This is a practical demonstration of the central limit theorem as it applies to the exponential distribution. We will conduct 1000 simulations of 40 exponentially distributed random values (with λ=0.2). We will compare the sample means and sample variances to the asymptotic theoretical values.

Simulations:
Below is the code for the simulation ran in R which:

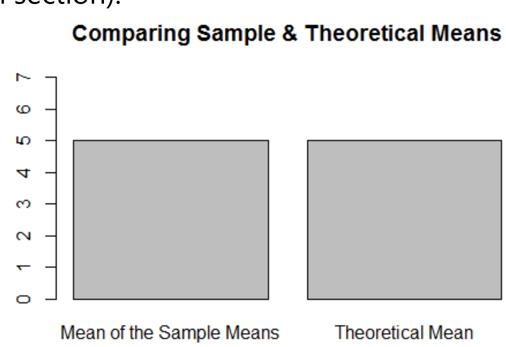
1. Takes the initial parameters of the simulation.
2. Stores the theoretical values for later use.
3. Simulates `n` exponentially distributed random variables and calculates the mean. It repeats this simulation `nosim` times. The code outputs vector `sample_means` containing the sample mean for each simulation.

```
# Set parameters
n <- 40
lambda <- 0.2
nosim <- 1000

# Store Theoretical Values
theoretical_mean <- 1/lambda
theoretical_sd <- 1/lambda


## Run simulation on sample means
sample_means <- NULL
for(i in 1:nosim){
  sample_means <- c(sample_means, mean(rexp(n,lambda)))
}
```

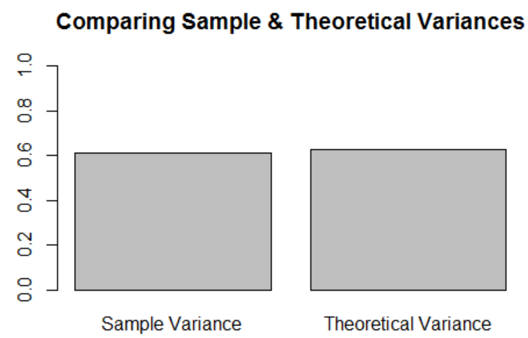
Sample Mean versus Theoretical Mean:
The below code and figure compare the mean of the sample means to the theoretical mean (1/λ). This demonstrates that while individual sample means vary they are distributed around the population mean (also see the histogram in the distribution section).



```
> # Mean of sample means
> mean(sample_means)
[1] 5.012506
> # Theoretical mean
> theoretical_mean
[1] 5
```

Sample Variance versus Theoretical Variance:

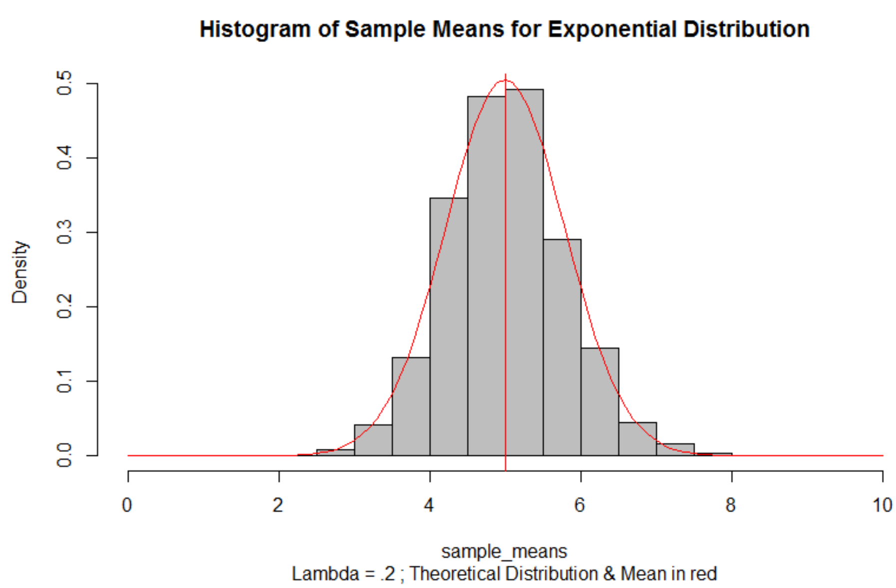
The below code and figure compare the variance of the sample means to the theoretical variance ($\frac{\sigma^2}{n} = \frac{1}{n\lambda^2}$). This demonstrates that the sample means approach being distributed in accordance to the population variance (also see the histogram in the distribution section).




```
> # Variance of Sample Means
> var(sample_means)
[1] 0.6112414
> # Theoretical Variance
> theoretical_sd^2/n
[1] 0.625
```

Distribution:

Below is the histogram of the sample means for the 1000 simulations. The theoretically corresponding normal distribution is overlaid in red, with mean 1/λ and variance $\frac{\sigma^2}{n} = \frac{1}{n\lambda^2}$. As we can see below, the distribution of the sample means roughly conforms to the to normal distribution. If n approached ∞, according to the central limit theorem we could expect the distribution of the sample means to be perfectly normally distributed.



Appendix

Full R Script:

```
#
### 1. Sample Mean versus Theoretical Mean
#

# Set parameters
n <- 40
lambda <- 0.2
nosim <- 1000

# Store Theoretical Values
theoretical_mean <- 1/lambda
theoretical_sd <- 1/lambda


## Run simulation on sample means
sample_means <- NULL
for(i in 1:nosim){
  sample_means <- c(sample_means, mean(rexp(n,lambda)))
}


## Compare distribution of sample means to theoretical means

# Mean of sample means
mean(sample_means)
# Theoretical mean
theoretical_mean

# Plot
barplot(c(mean(sample_means),theoretical_mean),
        names.arg = c("Sample Mean","Theoretical Mean"),
        ylim = c(0,7),
        main = "Comparing Sample & Theoretical Means")



#
### 2. Sample Variance versus Theoretical Variance
#

# Variance of Sample Means
var(sample_means)
# Theoretical Variance
theoretical_sd^2/n

# Plot
barplot(c(var(sample_means),theoretical_sd^2/n),
        names.arg = c("Sample Variance","Theoretical Variance"),
        ylim = 0:1,
        main = "Comparing Sample & Theoretical Variances")
```

```
#
### 3. Show that the distribution is approximately normal
#

# Plot histogram and normal distribution
plot.new()
hist(sample_means,
     breaks = (0:20)/2,
     freq = FALSE,
     col = "grey",
     main = "Histogram of Sample Means for Exponential Distribution",
     sub = "Lambda = .2 ; Theoretical Distribution & Mean in red")
abline(v = theoretical_mean, col = "red")
curve(dnorm(x, mean = theoretical_mean, sd = theoretical_sd/sqrt(n))
      ,0,10, add = TRUE, col = "red")
```