SimulationExercise_Part1

Nazmi Anik

Thursday, October 22, 2015

Overview

In this project the exponential distribution will be investigated in R and a comparison with the Central Limit Theorem will be provided. Lambda parameter will be set at 0.2 for all of the simulations. The distribution of averages of 40 exponentials will be investigated. A thousand simulations will be carried out.

Simulations

40 random exponentials with lambda = 0.2 are created using the function *rexp()*. This constitues as one experiment. This experiment is repeated 1000 times. All the random numbers are stored in a matrix called *sim*. Then the averages of each experiment is taken using *apply* function with *mean* and rows of the matrix as the input.

```
#number of simulations
nosim <- 1000
#lambda value given
lambda <- 0.2
#number of samples
n <- 40
#set seed for reproducibility
set.seed(42)
#run exponential simulations
sim <- matrix(data = rexp(nosim * n, rate = lambda), nrow = nosim)
#apply the mean to get the mean of each row
sim_means <- apply(sim, 1, mean)
```

Plot the histogram to visualize.

```
#plot the histogram
hist(sim_means, main="Histogram of Sample Means", xlab="Sample Means", ylab="Frequency",
col = "steelblue")
```

Histogram of Sample Means
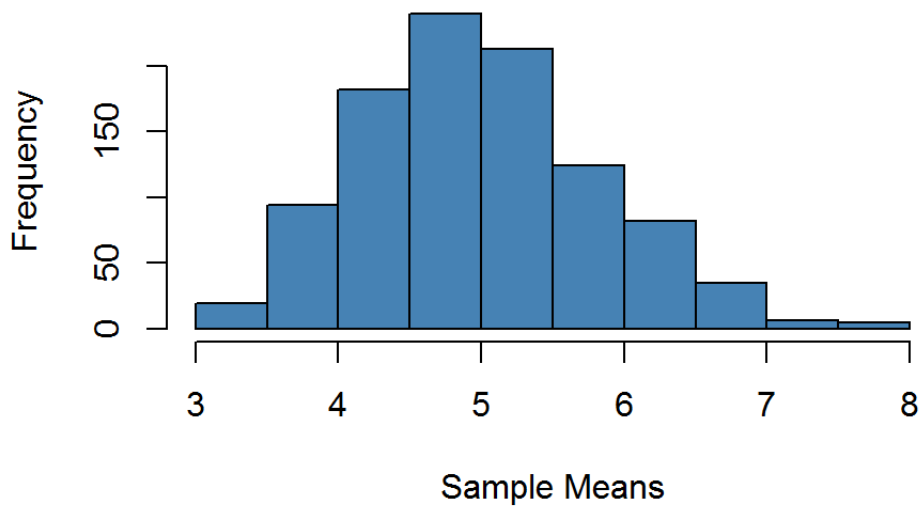
When we generate the histogram, we can see that the distribution of means of 40 exponentials resemble the normal distibution.

Now we will make comparisons to see how similar the distribution is with normal distribution.

Sample Mean versus Theoretical Mean

The theoretical mean is calculated by 1/lambda. Here's the comparison of sample mean and theoretical mean:

```
sample_mean <- mean(sim_means)
theoretical_mean <- 1/lambda
sample_mean
```

```
## [1] 4.986508
```

```
theoretical_mean
```

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

The sample mean is approximately equal to the theoretical mean.

Sample Variance versus Theoretical Variance

The theoretical standard deviation is also calculated by 1/lambda. Since we have 40 samples in each experiment, theoretical standard deviation is (1/lambda)/sqrt(n). Here's the comparison of sample variance and theoretical variance:

```
sample_variance <- (sd(sim_means))^2
theoretical_variance <- ((1/lambda)/sqrt(n))^2
sample_variance
```

```
## [1] 0.6793521
```

```
theoretical_variance
```

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

The sample mean is approximately equal to the theoretical mean.

Distribution

To check if the distribution is similar to a normal distribution, let's provide the theoretical mean and variance into the *dnorm* function.

```
#generate quantiles
x <- seq(min(sim_means), max(sim_means), length = 100)
#plot the histogram
hist(sim_means, prob = TRUE, breaks = 20, main="Density of Sample Means", xlab="Sample Me
ans", ylab="Density", col = "steelblue")
#plot the normal distribution curve
curve(dnorm(x, mean = 1/lambda, sd = (1/lambda)/sqrt(n)),col="orange",lwd=2,add=TRUE)
```


Density of Sample Means
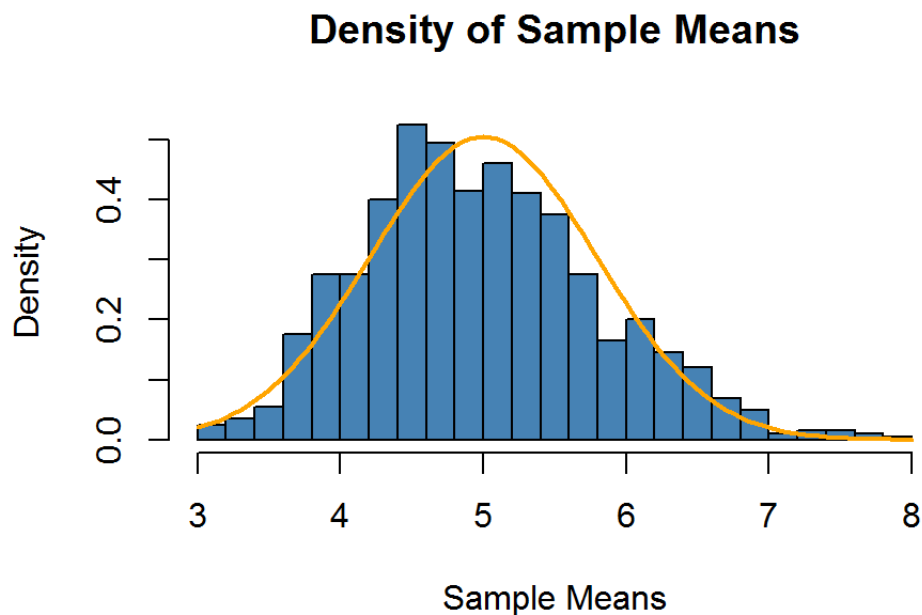
As it is visible from the above figure, the distribution of the exponential approximates to a normal distribution.

Supporting Images

We can view that the sample mean and theoretical mean are very close in the following histogram:

```
#plot the histogram
hist(sim_means, main="Histogram of Sample Means", xlab="Sample Means", ylab="Frequency",
col = "steelblue")
abline(v = sample_mean, col = "orange")
abline(v = theoretical_mean, col = "green")
legend(x = "topright", c("Distribution", "Sample Mean", "Theoretical Mean"), col = c("ste
elblue", "orange", "green"), lwd = c(2, 2, 2))
```
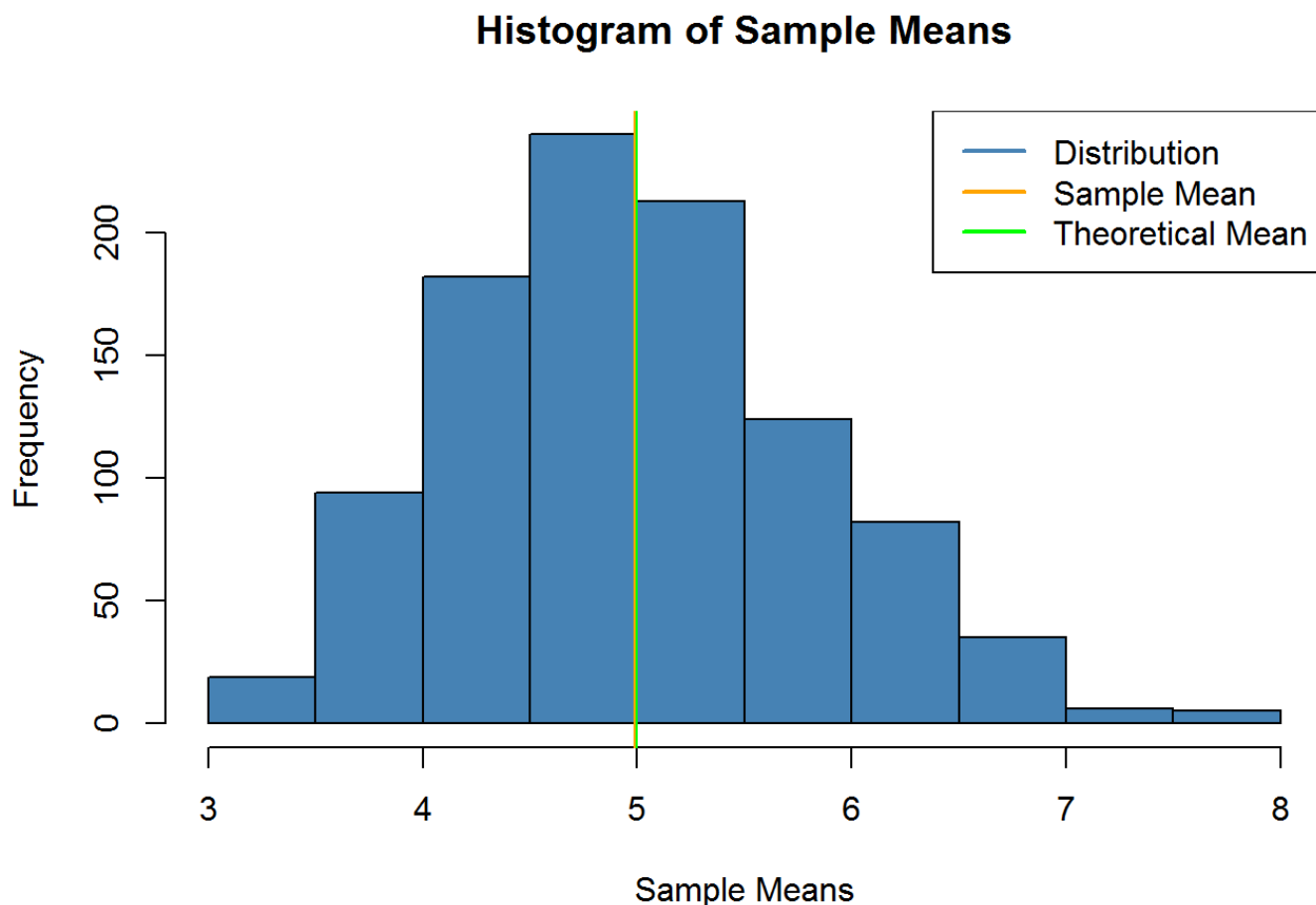

We can also view that the sample mean and theoretical standard deviations (and therefore variances) are very close in the following histogram:

```
#calculate standard deviations
sample_sd <- sd(sim_means)
theoretical_sd <- (1/lambda)/sqrt(n)
#plot the histogram
hist(sim_means, main="Histogram of Sample Means and Standard Deviations", xlab="Sample Me
ans", ylab="Frequency", col = "steelblue")
abline(v = sample_mean + sample_sd, col = "orange")
abline(v = sample_mean - sample_sd, col = "orange")
abline(v = theoretical_mean + theoretical_sd, col = "green")
abline(v = theoretical_mean - theoretical_sd, col = "green")
legend(x = "topright", c("Distribution", "Sample SD", "Theoretical SD"), col = c("steelbl
ue", "orange", "green"), lwd = c(2, 2, 2))
```


Histogram of Sample Means and Standard Deviations