

Investigate the exponential distribution in R and compare it with the Central Limit Theorem

Sreenivasulu Parimi

May 27, 2018

Statistical Inference Course Project - Part 1: Simulation Exercise

Overview

This project investigates the exponential distribution in R and compare it with the Central Limit Theorem.

Simulation

The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda` is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. Set `lambda = 0.2` for all of the simulations. Investigates the distribution of averages of 40 exponentials with a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

Set simulation variables called `seed`, `lambda`, `simulations` and `exponentials`.

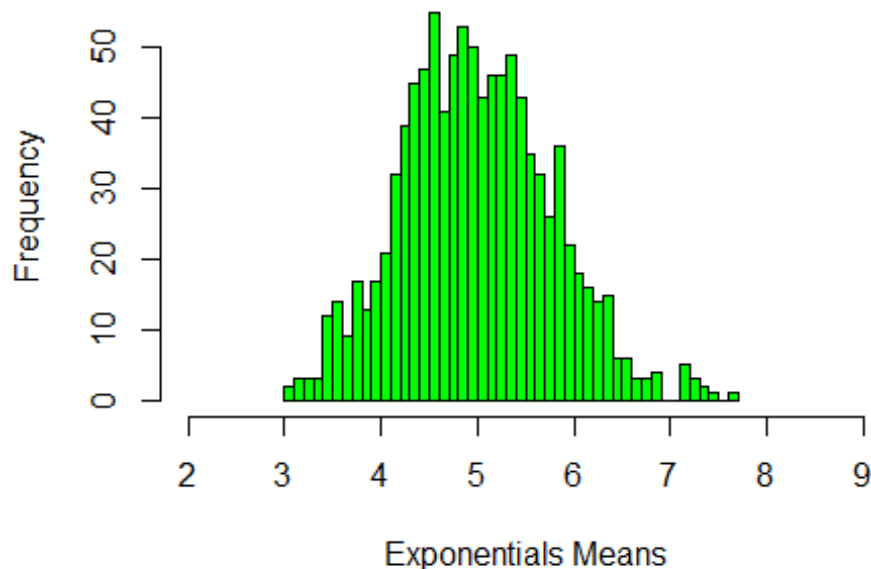
```
# set seed for reproducibility
set.seed(40)

# Set sampling values as described in the project instructions
lambda <- 0.2 # lambda
n <- 40      # number of exponentials
sims <- 1000 # number of simulations
#Run simulations
sim_exp <- replicate(sims, rexp(n, lambda))

#Calc the means of the exponential simulations
means_exp <- apply(sim_exp, 2, mean)

#Histogram of the means
hist(means_exp, col = "green", breaks=40, xlim = c(2,9), main="Exponential
Function Simulation Means", xlab = "Exponentials Means")
```

Exponential Function Simulation Means



1. Show the sample mean and compare it to the theoretical mean of the distribution

Sample mean is simply the mean of the simulated average values of the exponentials while theoretical mean of the average values of exponentials is $1/\lambda$.

```
# mean of sample means
mean(means_exp)

## [1] 4.988627

# theoretical means
1/lambda

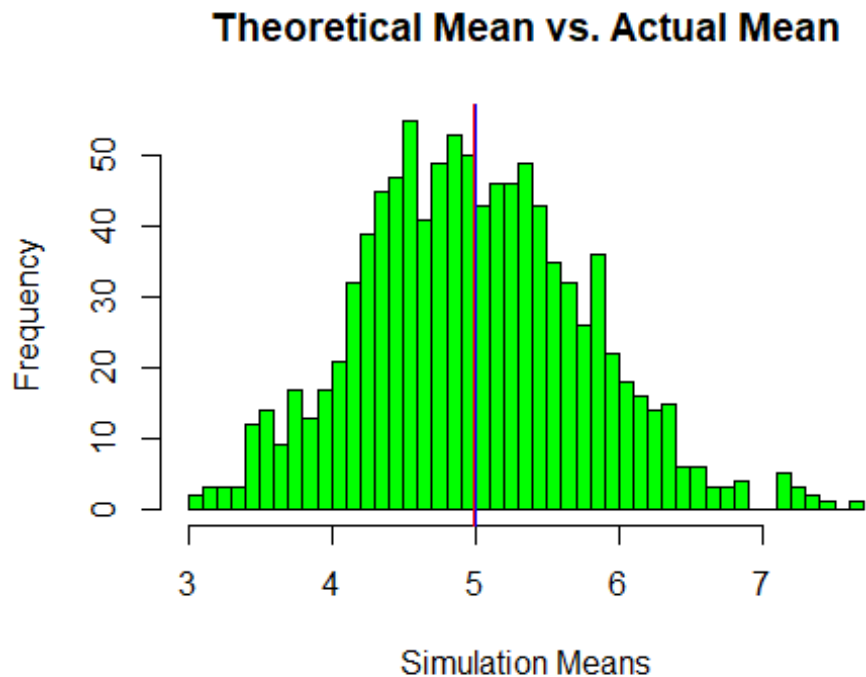
## [1] 5

# histogram of the sample means
#hist(means_exp, col="green", main="Theoretical Mean vs. Actual Mean", xlim =
c(2,9),breaks=40, xlab = "Simulation Means")
hist(means_exp, col="green", main="Theoretical Mean vs. Actual Mean",
breaks=50, xlab = "Simulation Means")

# vertical red line at the mean of the sample means
abline(v=mean(means_exp), col="red")

# vertical blue line at the mean of the theoretical means
```

```
#abline(v=1/lambda, lwd="4", col="blue")
abline(v=1/lambda, col="blue")
```



2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution

Theoretical variance is $(1/\lambda)^2$ divided by n while Sample variance is the squared standard deviation of the simulated averaged values of the exponentials.

```
# Theoretical Variance
(1/lambda)^2 / n

## [1] 0.625

# Sample Variance
sd(means_exp)^2

## [1] 0.6044075
```

It shows that Theoretical Variance 0.625 which is close to Sample Variance 0.6044075.

3. Show that the distribution is approximately normal

The means of the sample simulations should follow a normal distribution according to the Central Limit Theorem

```
hist(means_exp, density=20, breaks=20, prob=TRUE, xlab="mean of 40
exponentials", ylab="frequency", main="Histogram of the mean of 40
```

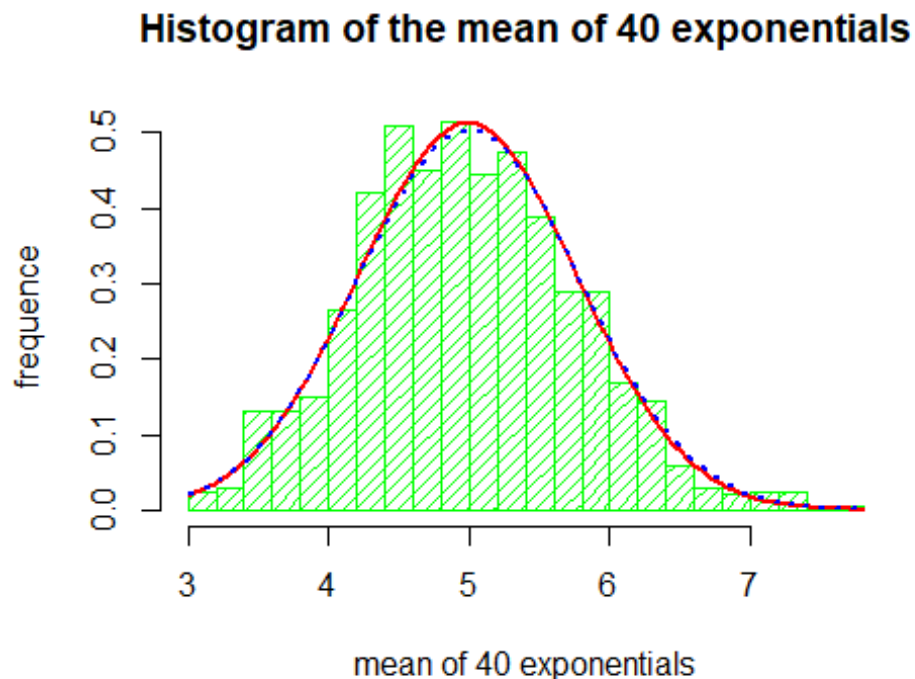
```

exponentials", col="green")

curve(dnorm(x, mean=mean(means_exp), sd=sd(means_exp)), col="red", lwd=2,
add=TRUE, yaxt="n")

curve(dnorm(x, mean=1/lambda, sd=(1/lambda)/sqrt(n)), col="blue", lwd=2, lty
= "dotted", add=TRUE, yaxt="n")

```



The above plot shows that the distribution of means of sampled exponential distributions appear to follow a normal distribution, due to the CLT. If we increase number of samples (currently 1000), the distribution would be even closer to the standard normal distribution. The blue dotted line above is a normal distribution curve and we can see that it is very close to sampled curve which is the red line above.