Statistical Inference course project - Part 1

Vignesh C Iyer

8/8/2020

Overview

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. 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.

Given instructions

- Set lambda = 0.2 for all of the simulations($lambda = 0.2$)
- Investigate the distribution of averages of 40 exponentials($exponentials = 40$)
- Do a thousand simulations($simulations = 1000$)

Simulating the mean

- First we set the seed for reproducibility using the set.seed() function
- We then initialize the variables lambda, exponentials and the simulations

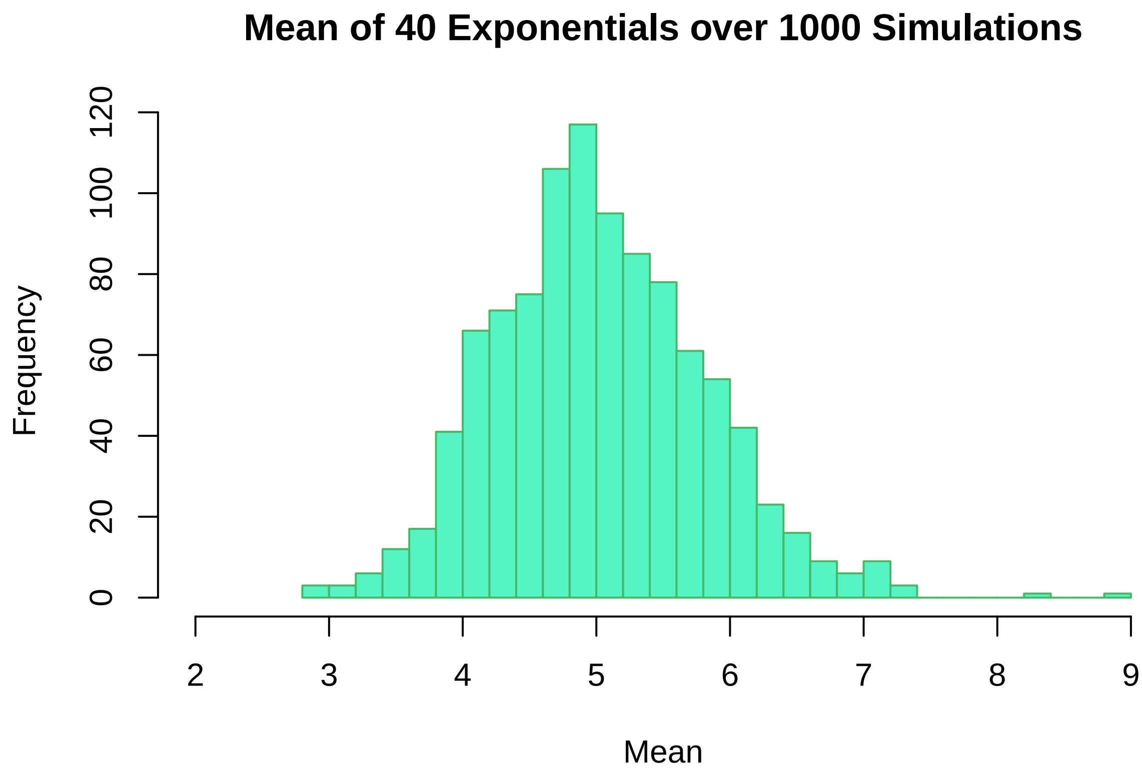
```r
# Setting the seed for reproducibility
set.seed(2000)

# As per given details initializing the variables
lambda <- 0.2
exponentials <- 40
simulations <- 1000

# Calculating the mean over 1000 simulations
meanexponents <- apply(replicate(simulations, rexp(exponentials, lambda)), 2, mean)

#Plotting the Mean of 40 Exponentials

hist(meanexponents, breaks = 40, xlim = c(2,9),
     main = "Mean of 40 Exponentials over 1000 Simulations",
     xlab = "Mean", ylab = "Frequency", col = "#56F4C4",
     border = "#45BA64")
```



We first compare the sample mean generated out of the 1000 simulations with the theoretical mean.

```
#Theoretical mean
theo_mean <- 1/lambda
theo_mean
```

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

```
#Sample mean of simulations
sample_mean <- mean(meanexponents)
sample_mean
```

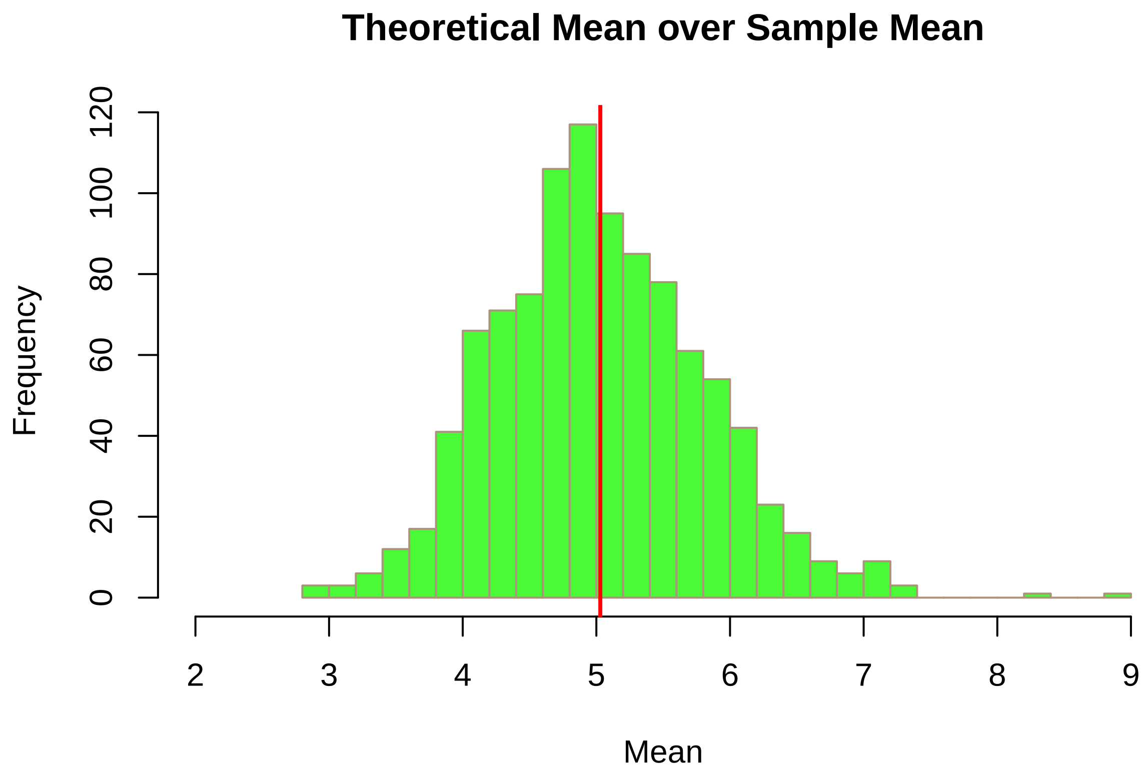
```
## [1] 5.02941
```

From the above observations we find the Sample mean and the theoretical mean are nearly same, now we will go ahead and plot the theoretical mean over the sample mean

```
# Plotting theoretical mean over sample mean

hist(meanexponents, breaks = 40, xlim = c(2,9),
     main = "Theoretical Mean over Sample Mean",
     xlab = "Mean", ylab = "Frequency", col = "#49FA34",
     border = "#AC947A")

abline(v= mean(sample_mean), lwd = "2", col = "#FF0000")
```



Now we compare the Theoretical Variance over the Sample Variance

```
# Sample Variance is the square of the Standard deviation
sample_variance <- (sd(meanexponents))^2
sample_variance
```

```
## [1] 0.6168082
```

```
# Theoretical Variance is Theoretical mean squared divided by the exponentials
theo_variance <- (1/lambda)^2/exponentials
theo_variance
```

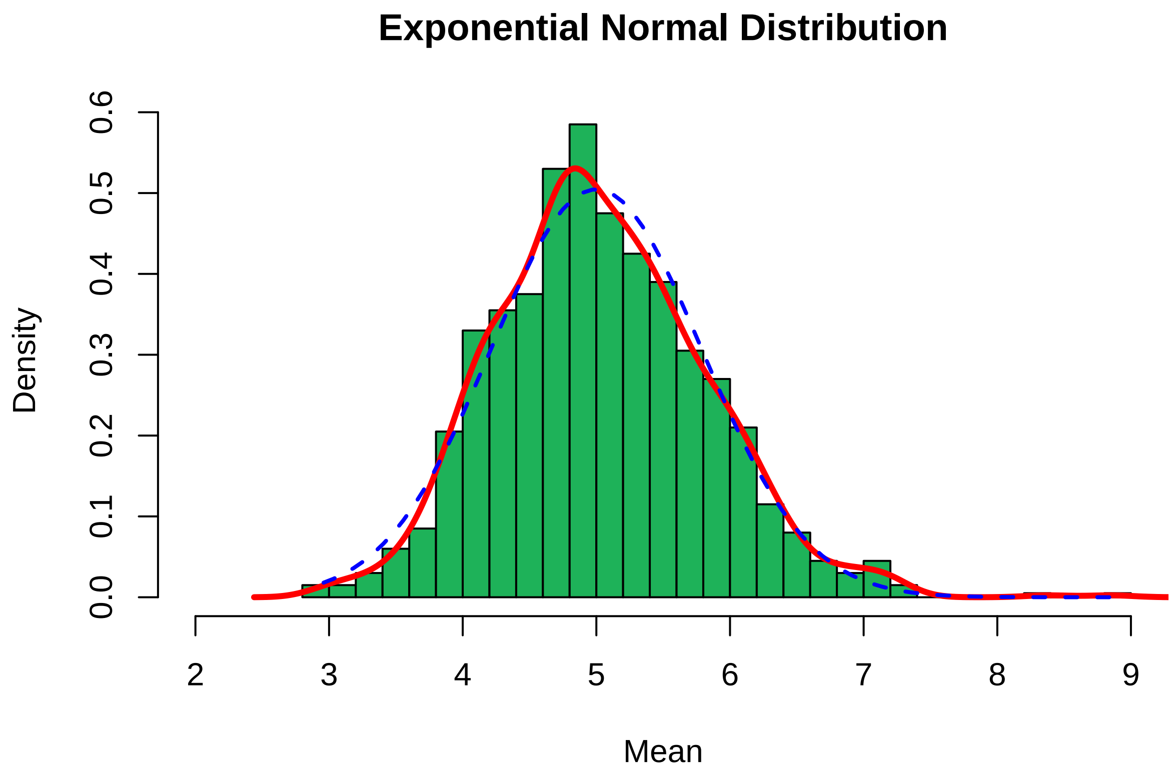
```
## [1] 0.625
```

Plotting the Distributions

While plotting the distribution we will plot the histogram and then extrapolate the normal distribution.

```
#General Plot with distribution curve drawn
hist(meanexponents, prob=TRUE, col="#1EB259",
     main="Exponential Normal Distribution", breaks=40, xlim=c(2,9),
     xlab = "Mean")
lines(density(meanexponents), lwd=3, col="red")

# Normal distribution line creation
x <- seq(min(meanexponents), max(meanexponents), length=2*exponentials)
y <- dnorm(x, mean=1/lambda, sd=sqrt(((1/lambda)/sqrt(exponentials))^2))
lines(x, y, pch=22, col="blue", lwd=2, lty = 2)
```



From the above calculation and graph we see that the distribution of means of our sample exponential distributions appear to follow a normal distribution, due to the Central Limit Theorem. Increasing the sample size would result in the distribution being even closer to the standard normal distribution.The dotted line above is a normal distribution curve and we can see that it is very close to our sampled curve, which is the red line above.