Simulation Exercise of the Exponential Distribution

by Diana Hanania

Overview

This report is submitted as an assignment for the "Statistical Inference" course by "John Hopkins University"on Coursera.

The exponential distribution (lambda=0.2) will be investigated in R and compared with the Central Limit Theorem. The investigation involves the distribution of averages of 1000 simulated samples of 40 random exponentials each. The distribution of the averages appears Gaussian even though the distribution of the sample is not Gaussian but exponential and his is due to the Central Limit Theorem.

Exponential Distribution in R

Quoted from the course:

"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."

Setting Glowbal Options

```r
options(warn = -1)
knitr::opts_chunk$set(comment = NA)
knitr::opts_chunk$set(message = FALSE)
knitr::opts_chunk$set(opts.label="kill_prefix")
knitr::opts_template$set("kill_prefix"=list(comment=NA, null_prefix=TRUE))
knitr::opts_chunk$set(echo=TRUE)
```

Simulations

The simulation consists of sampling 1000 different random samples of size 40. For each of these samples, the mean and variance will be calculated. Finally, a distribution of each of the sample means and sample variances will be plotted.

Setting the needed parameters

As mentioned in the overview, the rate of the exponential distribution from which random samples are taken is lambda=0.2. The simulation size is 1000 and the sample size is 40.

```r
# sample size
n<- 40
# Simulation size
B<- 1000
# rate parameter for exponential distribution
lambda <- 0.2
# theoretical values,mean and standard deviation, of exponential distribution with l
# ambda=0.2
mu <- 1/lambda
sd <- 1/lambda
var<- sd^2
```

Sample Mean versus Theoretical Mean

The theoretical mean, red dotted line, is plotted so that a comparison can be made with the sample averages. This plot shows the distribution of calculated means of 1000 samples each of size 40. Each sample has random values from the exponential distribution. Figure 1 shows:

- The average sample means, green line, which is the average of all 1000 sample means, is approximately equal to the theoretical mean, 1/lambda.
- The distribution of the sample means are centered around and almost symmetrically distributed around the average sample means.

```r
# Monte Carlo simulation (B=1000) of the average of 40 values picked randomly from
# exponential distribution with lambda=0.2
expmu<- sapply(1:B, function(n=40, lambda=0.2){
  y<- rexp(n, lambda)
  muy<- mean(y)  #round(mean(y),2)
})

{hist(expmu, col="grey",main="Histogram of Samples' Means", breaks=70,
      xlab="Samples' Means")
  abline(v= mean(expmu), col="#336600",lwd="2")
  abline(v= mu, col="#990000",lwd="2",lty=2)
  legend("topright", c("Average sample means","Theoretical mean"),bty = "n", lwd = 2,
        col = c("#336600","#990000"),lty=c(1,2))}
```

```
The average of the samples' means is 4.994587544172. And the absolute difference between
the  average of the samples' means and the theoretical mean is 0.00541245582799554 which
is very small.
```

Sample Variance versus Theoretical Variance

Include figures (output from R) with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

Sample variance vs theoretical variance The theoretical variance, red dotted line, is plotted so that a comparison can be made with the sample variances. This plot shows the distribution of calculated variances of 1000 samples each of size 40. Each sample has random values from the exponential distribution. Figure 2 shows:

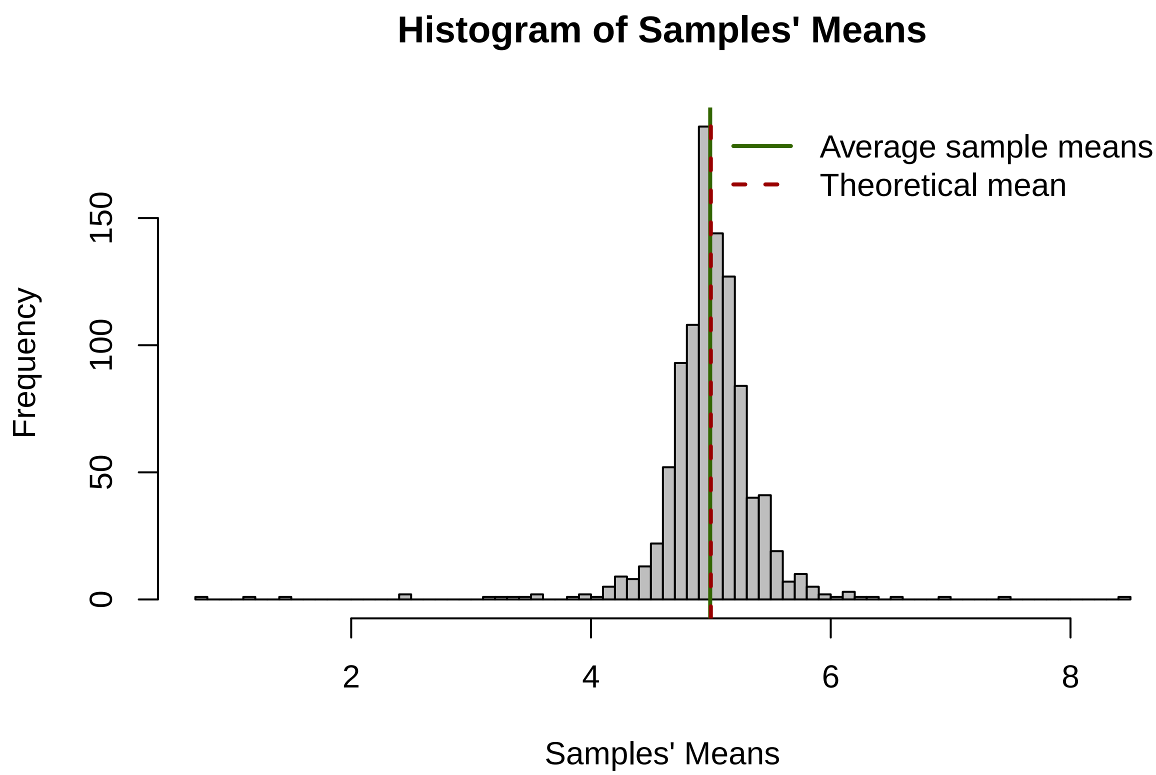

Figure 1: Figure 1: Histogram plot of the averages of 1000 samples of size 40

- The average variance, green line, which is the average of all 1000 variances, is approximately equal to the theoretical variance, (1/lambda)^2.
- The distribution of the sample variances are centered around and almost symmetrically distributed around the average sample variances.

```
# Monte Carlo simulation (B=1000) of the average of 40 values picked randomly from
# exponential distribution with lambda=0.2
expvar<- sapply(1:B, function(n=40, lambda=0.2){
  y<- rexp(n, lambda)
  vary<- var(y)
})

{hist(expvar, col="grey",main="Histogram of Samples' Variance", breaks=70,
      xlab="Samples' Standard Deviations")
  abline(v= mean(expvar,na.rm=T), col="#336600",lwd="2")
  abline(v= var, col="#990000",lwd="2",lty=2)
  legend("topright", c("Average sample Variances","Theoretical Variance"),
         bty = "n", lwd = 2, col = c("#336600","#990000"),lty=c(1,2))}
```

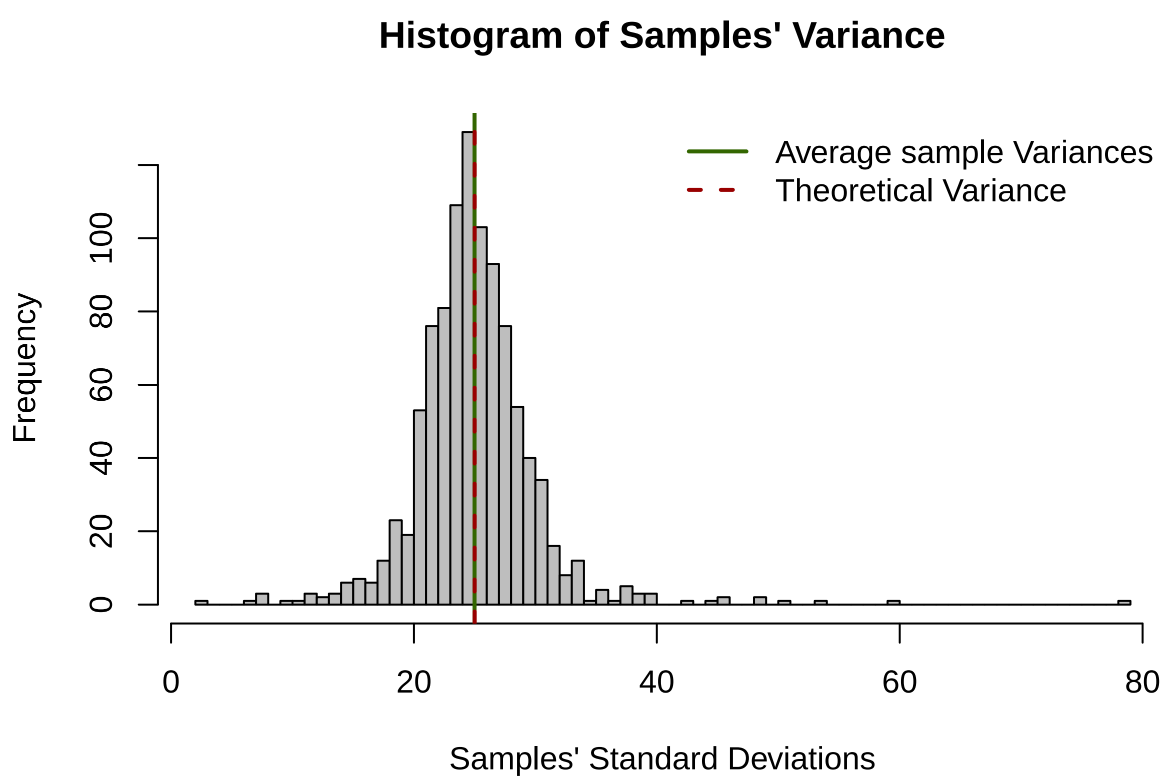

Figure 2: Figure 2: Histogram plot of the varianve of 1000 samples of size 40

```
The average of the samples' variances is 24.9913355296575. And the absolute difference between
the  average of the samples' variances and the theoretical variance is 0.00866447034254847 which
is very small.
```

Distribution

This plot shows approximately the shape of an exponential distribution by generating the plot from 1000 random samples.

From Figure 3, the distribution looks exponential and does not look normal at all. Its values are not centered around the mean, neither are they symmetrical but more skewed to the left. The mean of this sample is approximately equal to the theoretical mean. This is expected because of the large sample size, 1000.

While this sample size showed an exponential distribution, yet Figures 2 and 3 show more of a normal/ Gaussian distribution for both the means and variances generated from the multiple samples of the exponential distribution.

```
# plotting a sample of 1000 values of an exponential distribution
sample<- rexp(1000, lambda)
mu_s <- mean(sample)
sd_s <- sd(sample)

{hist(sample, col="grey",main="Histogram of the Sample", breaks=70, xlab="Sample Values")
 abline(v= mu_s, col="#336600",lwd="2")
 abline(v= mu, col="#990000",lwd="2",lty=2)
 legend("topright", c("Sample mean","Theoretical mean"),bty = "n", lwd = 2,
        col = c("#336600","#990000"),lty=c(1,2))
 }
```

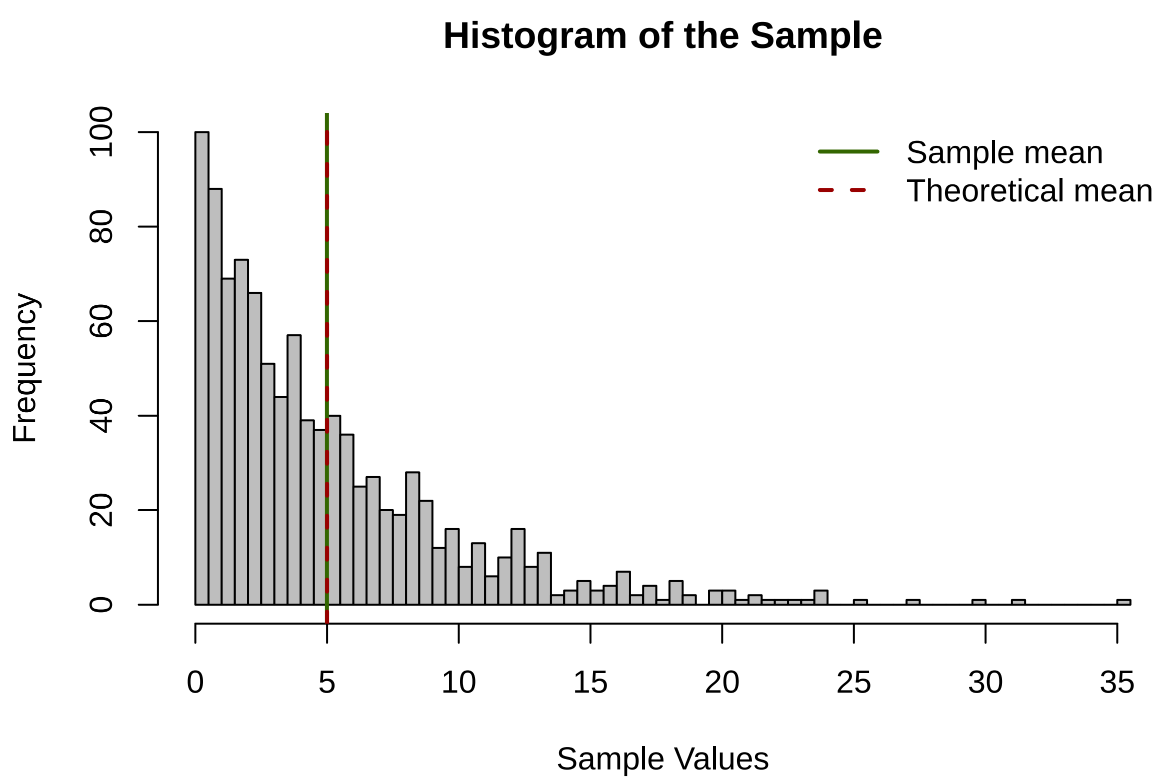

Figure 3: Figure 3: Histogram plot of 1000 random samples of an exponential distribution

Conclusion

So, distribution of the means can be Gaussian centered around the theoretical mean even if the original distribution of the data is not Gaussian/normal as seen in this exponential case. This example is justified by and is also a validation of the CLT(central limit theorem).