Comparison of Simulated means of an exponential distribution with the Central Limit Theorem

Jerry Harber

July 8, 2017

Overview

A collection of means from simulations of an exponential distribution is done. The distribution of these simulation means will be compared with a normal distibution. The purpose is to show that the Central Limit Theorem (i.e., CLT) does provide a normal distribution.

Simulations

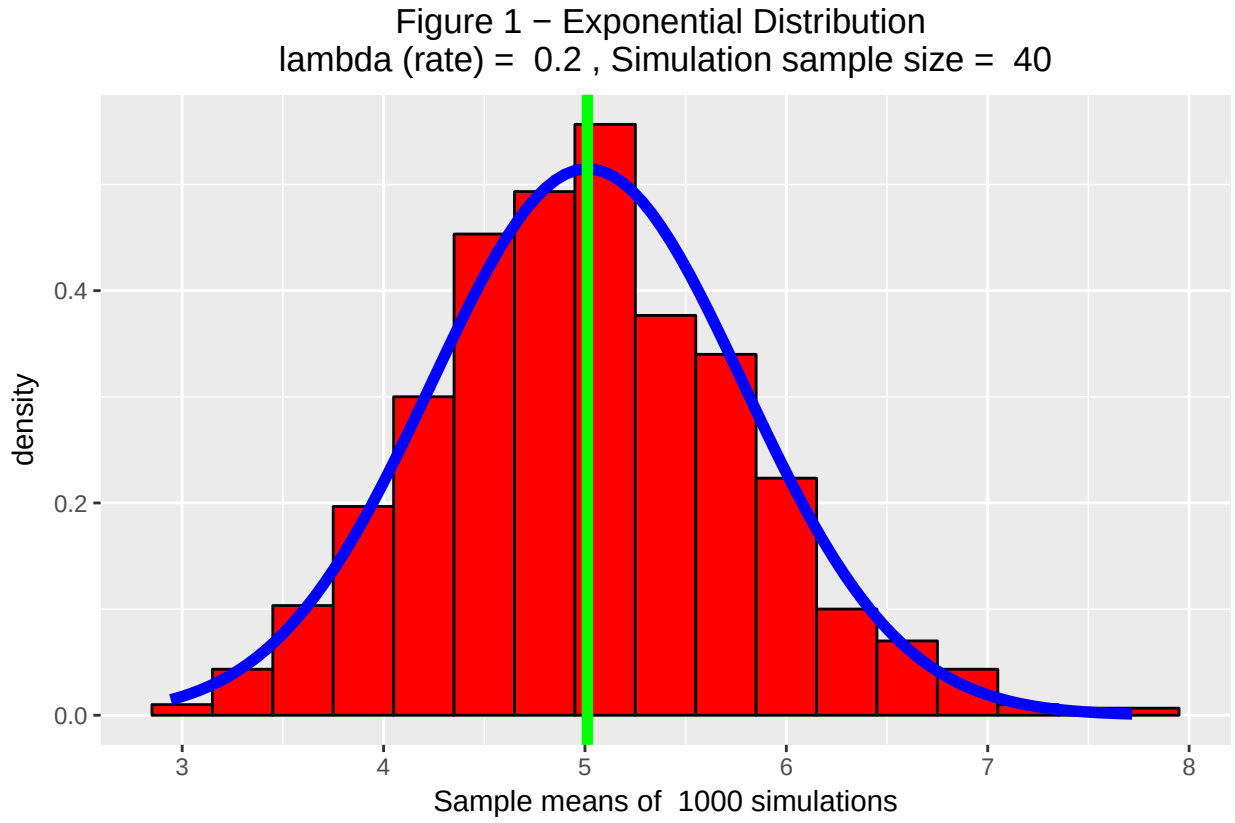
Simulations of the Exponential distribution were done using the rexp function in R. The rate parameter (i.e., lambda) is set to 0.2 and the number of simulations done is 1000. Each simulation sample size is 40. The means from each simulation sample will be plotted using a histogram. The histogram's results will be compared with a normal distribution which should show that the simulated sample means should follow the CLT.

Sample Mean versus Theoretical Mean.

R code and result graphics are as follows;

```
## Do a simulation of the expoential distribution
set.seed(123)
## Sample size
n = 40
## number of samples
nsim = 1000
lambda = 0.2
mean <- 1/lambda
sd <- 1/lambda
df <- data.frame()
for (i in 1 : nsim) {
    msample = rexp(n, lambda)
    df[i,"sample.size"] <- n
    df[i,"sample.mean"] <- mean(msample)
    df[i,"sample.variance"] <- var(msample)
}
## Determine the mean and sd of all the samples for later use in
## displaying a normal distribution for comparison to the distribution
## of the simulated sample means
smean <- mean(df[,"sample.mean"])
rsmean <- round(smean,3)
ssd <- sd(df[,"sample.mean"])
rssd <- round(ssd,3)
mytitle <- paste("Figure 1 - Exponential Distribution",
"\nlambda (rate) = ", lambda,", Simulation sample size = ",n)
myxlabels <- paste("Sample means of ", nsim, "simulations\n")
```

```
myylabels <- "density"
g <- ggplot(df,aes(x = df$sample.mean))
g <- g + geom_histogram(binwidth=.3, col = "black", fill = "red", aes(y = ..density..))
g <- g + stat_function(fun=dnorm,size=2,col="blue",args = list(mean=smean, sd=ssd) )
g <- g + labs(title=mytitle) + labs(x=myxlabels, y=myylabels)
g <- g + geom_vline(xintercept = smean, col="green", size = 2)
g <- g + theme(plot.title=element_text(hjust=0.5))
g
```



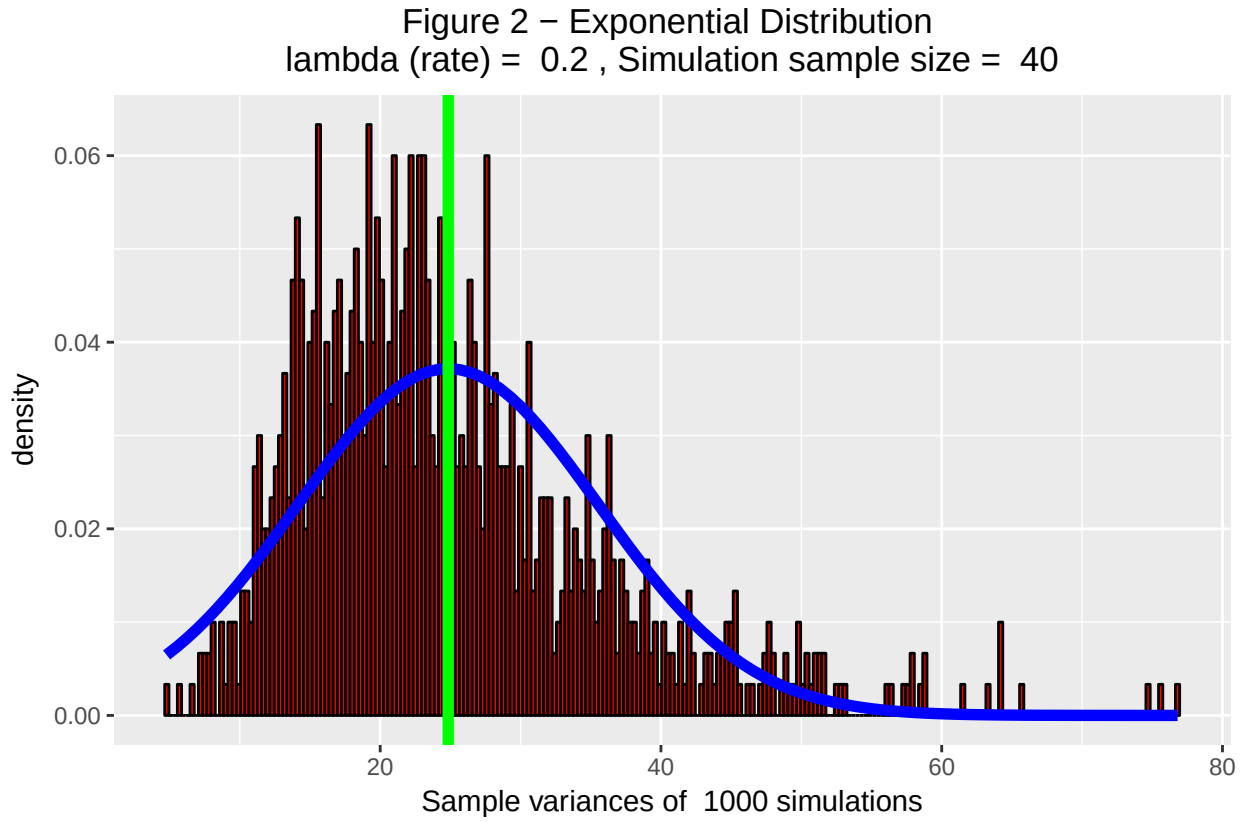
The "red" bars in the histogram show the distibution of the simulated sample means. The histogram of simulated means appears to follow a normal distribution. The "blue"" bell-shaped line is a normal distribution generated with parameters mean = 5.012 (i.e., mean of all the simulated sample means) and sd = 0.775 (i.e., the standard deviation of all the simulated sample means).The "green" vertical line shows the mean (i.e., 5.012) of all the simulated sample means. The mean of 5.012 is very close to the mean of 5.0 (i.e., 1/lambda), which was used to generate the samples.

Sample Variance versus Theoretical Variance

R code and result graphics are as follows;

```
svarmean <- mean(df[,"sample.variance"])
rsvarmean <- round(svarmean,3)
vssd <- sd(df[,"sample.variance"])
rvssd <- round(vssd,3)
```

```
mytitle <- paste("Figure 2 - Exponential Distribution",
"\nlambda (rate) = ", lambda,", Simulation sample size = ",n)
myxlabels <- paste("Sample variances of ", nsim, "simulations\n")
myylabels <- "density"
g <- ggplot(df,aes(x = df$sample.variance))
g <- g + geom_histogram(binwidth=.3, col = "black", fill = "red", aes(y = ..density..))
g <- g + stat_function(fun=dnorm,size=2,col="blue",args = list(mean=svarmean, sd=vssd) )
g <- g + labs(title=mytitle) + labs(x=myxlabels, y=myylabels)
g <- g + geom_vline(xintercept = svarmean, col="green", size = 2)
g <- g + theme(plot.title=element_text(hjust=0.5))
g
```



The "red" bars in the histogram show the distibution of the simulated sample variances. The histogram does not seem to follow a normal distribution. The "blue" bell-shaped line is a normal distribution generated with parameters mean = 24.843 (i.e., mean of all the simulated sample variances) and sd = 10.734 (i.e., the standard deviation of all the sample variances). The "green" vertical line shows the mean (i.e., 24.843) of all the simulated sample variances. The mean of 24.843 is very close to the variance of 25.0 (i.e., 1/lambda^2), which was used to generate the exponential samples.