Statistical Inference Project: Simulation Exercise Part1

Dipali Bagad

August 11, 2017

Part 1: Simulation Excercise Synopsis

In this exercise we 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. Set lambda = 0.2 for all of the simulations. We will investigate the distribution of averages of 40 exponentials. We need to do following points: 1. Show the sample mean and compare it to the theoretical mean of the distribution. 2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. 3. Show that the distribution is approximately normal.

Sample Mean and theoretical mean comparison

```
library(ggplot2)
# as per instructions, set lambda to 0.2 for all simulations
lambda <- 0.2
# we need to investigate distribution of averages of 40 exponentials so set
n=40
n <- 40
# we need to do thousand simulations
nosim <- 1000
# 1. Show the sample mean and compare it to the theoretical mean of the
distribution.
#  calculate sample mean for 1000 simulations and store it
set.seed(123)
SampleMean = NULL
for (i in 1 : 1000) {
    SampleMean = c(SampleMean, mean(rexp(n,lambda)))
    }
SampleFinalMean <- round(mean(SampleMean),3)
SampleFinalMean

## [1] 5.012

#calculate theoretical mean
TheoMean <- round(1/lambda,3)
TheoMean

## [1] 5
```

From sample mean 5.012 value and theoretical mean 5 we can say that both values are nearly same

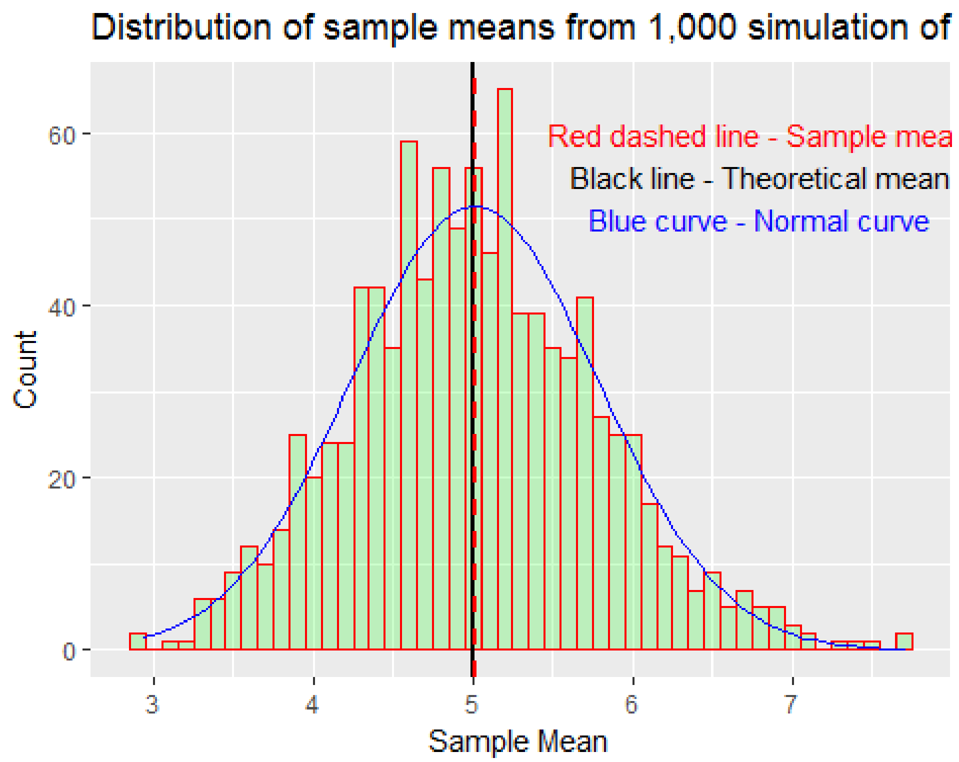
Graph to show Sample Mean and throretical mean comparison

Below graph shows the theoretical mean and sample mean values are nearly same

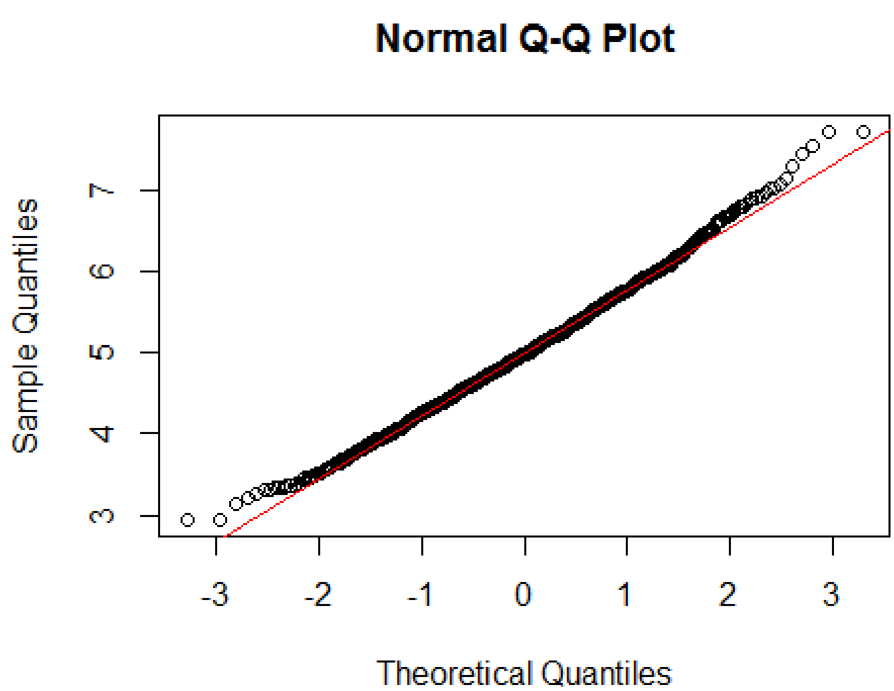
```
SampleMeanData <- as.data.frame(SampleMean)

# graph with normal, sample and theoretical mean with annotations

ggplot(data=SampleMeanData, aes(x=SampleMean)) +
geom_histogram(binwidth=.1, col="red", fill="green", alpha = .2) +
geom_vline(xintercept = 5,aes(fill="black"), size = 1) +
geom_vline(aes(xintercept= SampleFinalMean), color="red", linetype="dashed",
size=1) +
stat_function(
fun = function(x, mean, sd){
dnorm(x = x, mean = mean, sd = sd) * 1000 * .1
},
args = c(mean = mean(SampleMean), sd = sd(SampleMean)), color = "blue", size
= 0.5) +
guides(fill=FALSE) +
annotate(geom="text", x=6.8, y=60, color="red",label="Red dashed line -
Sample mean") +
annotate(geom="text", x=6.8, y=55, color="black",label="Black line -
Theoretical mean") +
annotate(geom="text", x=6.8, y=50, color="blue",label="Blue curve - Normal
curve") +
ggtitle("Distribution of sample means from 1,000 simulation of 40
exponentials vs theoretical mean") +
xlab("Sample Mean") + ylab("Count")
```




```
##Q-Q plot
qqnorm(SampleMean, main="Normal Q-Q Plot")
qqline(SampleMean, col="2")
```



Sample Variance versus Theoretical Variance

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

```
#calculate theoretical variance
TheoVar <- round((1/lambda)^2/n,3)
TheoVar

## [1] 0.625

#calculate sample variance
SampleVar <- round(var(SampleMean),3)
SampleVar

## [1] 0.6

#display the values of theoretical mean and sample mean with their respective
variance
output <- matrix(c(TheoMean, SampleFinalMean,
                TheoVar,SampleVar),
              ncol = 2, byrow=TRUE)
colnames(output) <- c("Theroetical","Sample")
rownames(output) <- c("Mean","Mariance")
output

##          Theroetical Sample
## Mean           5.000  5.012
## Mariance       0.625  0.600
```

from these results we can say that sample variance and theoretical variance of the distribution are very close.

Distribution

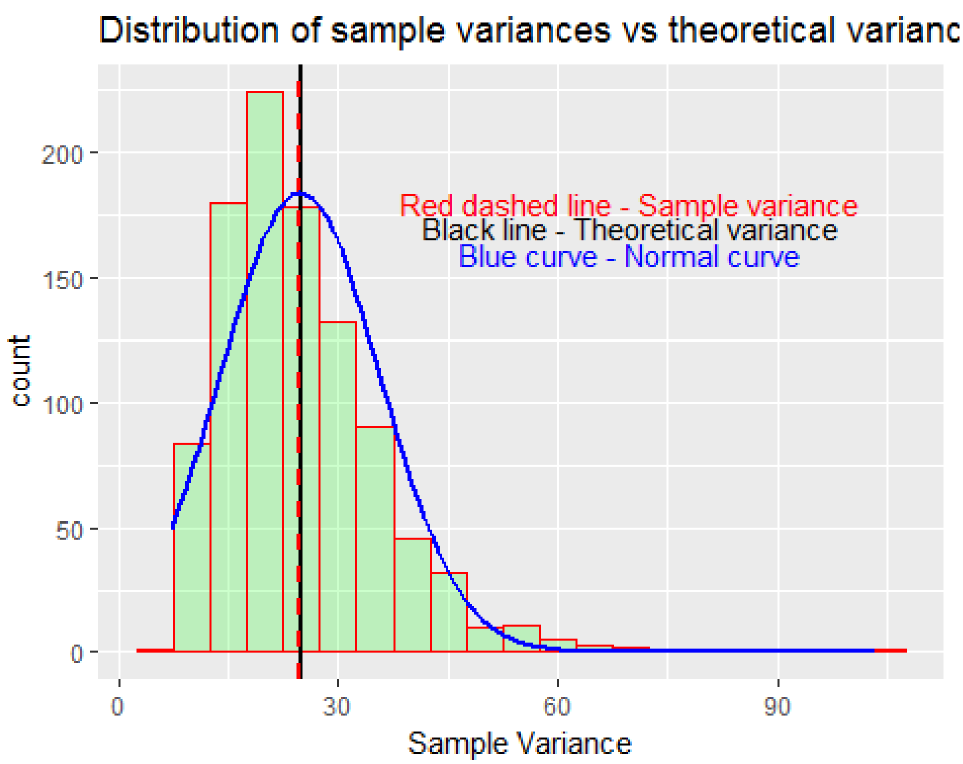
3. Show that the distribution is approximately normal.

The below plot shows the density curve is very similar as the normal distribution curve. This indicates the distribution is approximately normal.

```
 #calculate the variance
variance = NULL
for (i in 1 : 1000) variance = c(variance, var(rexp(40, 0.2)))

# graph with normal, sample and theoretical variance with annotations
ggplot(data=data.frame(variance), aes(x=variance)) +
geom_histogram(binwidth=5, col="red", fill="green", alpha = .2) +
geom_vline(xintercept = 25,aes(fill="black"), size = 1) +
geom_vline(aes(xintercept=mean(variance)),color="red", linetype="dashed",
size=1) +
stat_function(
fun = function(x, mean, sd){
dnorm(x = x, mean = mean, sd = sd) * 1000 * 5
```

```
},
args = c(mean = mean(variance), sd = sd(variance)),color = "blue", size = 1)
+
guides(fill=FALSE) +
annotate(geom="text", x=70, y=180, color="red",label="Red dashed line -
Sample variance") +
annotate(geom="text", x=70, y=170, color="black",label="Black line -
Theoretical variance") +
annotate(geom="text", x=70, y=160, color="blue",label="Blue curve - Normal
curve") +
ggtitle("Distribution of sample variances vs theoretical variance") +
xlab("Sample Variance")
```



Results

1. sample mean value and theoretical mean are nearly same
2. sample variance and theoretical variance of the distribution are very close
3. distribution is approximately normal.