The Exponential Distribution and the Central Limit Theorem.

Carlos Martinez Reyes

4/10/2020

Part 1: Simulation Exercise

Overview.

In this project we use R to investigate the behavior of the distribution of the sample mean of 40 iid Exponential distributions with parameter lambda = 0.2. And based on a thousand simulations of this sample, we empirically verify the Central Limit Theorem.

Simulations.

The following code is used to generate a random sample of 40 exponential distributions with the same parameter of 0.20, subsequently we obtain a thousand simulations of each of these distributions to find the distribution of the sample mean.

```
set.seed(13000)
# Setting a rate lambda parameter, the sample n and the 1000 simulations.
lambda <- 0.2
n <- 40
sims <- 1000
simexp <- replicate(sims, rexp(n, lambda))
```

simexp is a 40 x 1000 matrix where each column contains a random sample of the 40 exponential distributions. Now we calculate the mean of each column to obtain a thousand sample means.

```
# calculate means of simulations
expmeans <- apply(simexp, 2, mean)
```

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

Theoretical mean: the mean of the exponential distribution is 1/lambda, if lambda = 0.2 it follows that the theoretical mean turns out to be 1 / 0.2 = 5. Sample mean:

```
# theoretical mean vs simulated mean
sample_mean <- mean(expmeans)
sample_mean
```

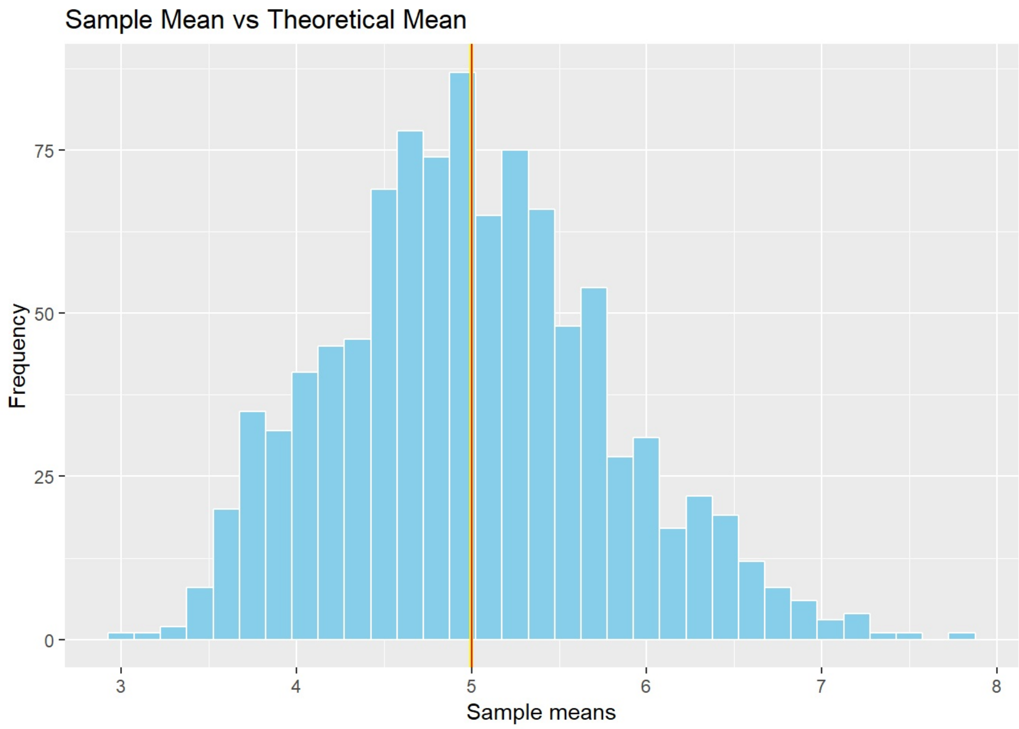
```
## [1] 4.998317
```

```
theory_mean <- 1/lambda
theory_mean
```

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

So the sample mean is = 4.998317 is a good approximation to the theoretical mean is = 5. Let’s see this graphically from the frequency distribution of the sample mean.

```
library(ggplot2)
ggplot(as.data.frame(expmeans), aes(x= expmeans)) +
geom_histogram(binwidth =.15, fill = "skyblue" ,col = "white") +
geom_vline(xintercept = sample_mean, color = "yellow", size = 1) +
geom_vline(xintercept = theory_mean, color = "red", size = 0.5) +
labs(title = "Sample Mean vs Theoretical Mean", x = "Sample means", y = "Frequency")
```



The yellow vertical line corresponds to the sample mean, while the red vertical line is the theoretical mean.

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

```
# theoretical variance vs simulated variance
print(paste ("Theoretical variance is: ", tvar<-round( (1/lambda)^2/n, 3)))
```

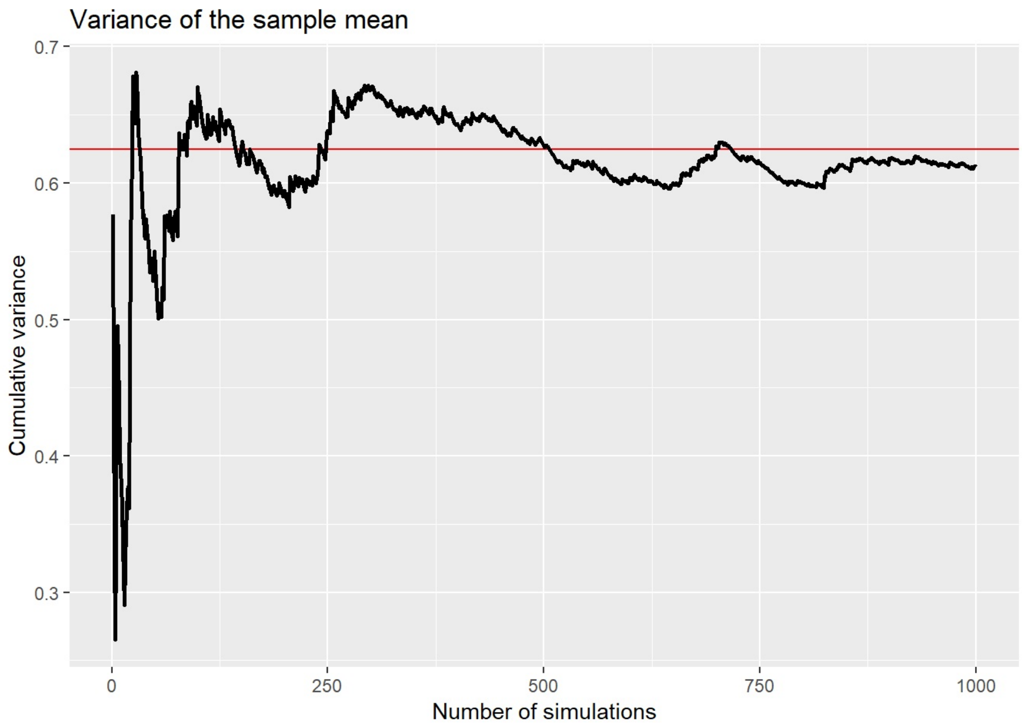
```
## [1] "Theoretical variance is:  0.625"
```

```
print(paste("Sample mean variance is: ", svar<-round(var(expmeans),3)))
```

```
## [1] "Sample mean variance is:  0.612"
```

Both values of the theoretical and sample variance assume values close to each other. The variance of the sample mean and the theoretical variance are similar. Let's see graphically the asymptotic behavior of the variance of the sample mean.

```
varac <- cumsum((expmeans - sample_mean)^2)/(seq_along(expmeans - 1))
ggplot(data.frame(x = 1:sims, y = varac), aes(x = x, y = y))+
 geom_hline(yintercept = tvar, colour = 'red') + geom_line(size = 1)+
labs(x = "Number of simulations", y = "Cumulative variance")+
ggtitle('Variance of the sample mean')
```

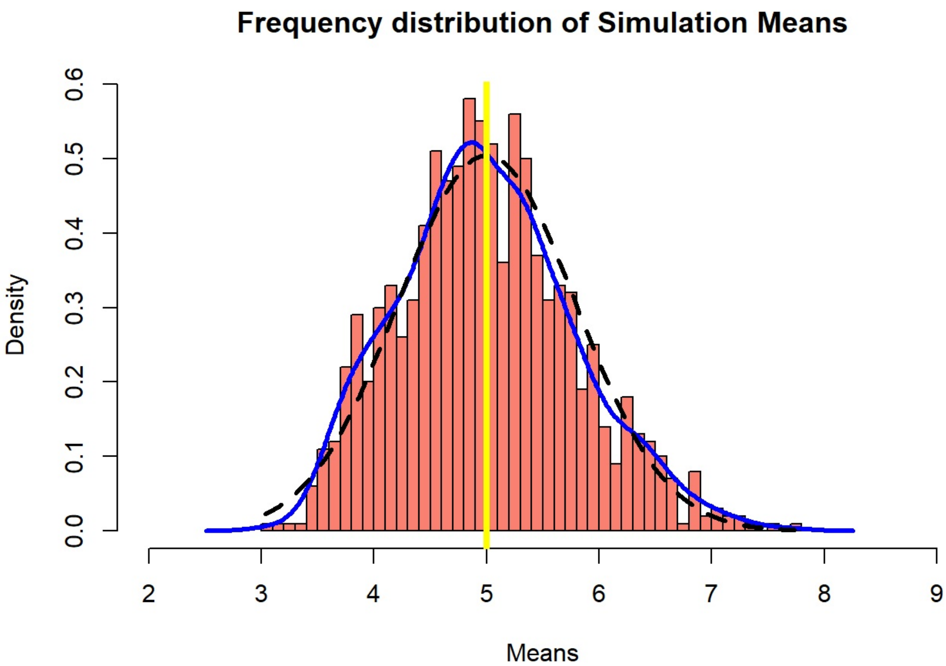


Question 3: Show that the distribution is approximately normal.

Due to the Central Limit Theorem, the means of the sample simulations should follow a normal distribution.

```
# Frequency distribution of the sample mean
hist(expmeans, prob=TRUE, col="salmon", main="Frequency distribution of Simulation Means", breaks=40, xlim=c(2,9),
xlab = "Means")
lines(density(expmeans), lwd=3, col="blue")

# Normal distribution
x <- seq(min(expmeans), max(expmeans), length=2*n)
y <- dnorm(x, mean=1/lambda, sd=sqrt(((1/lambda)/sqrt(n))^2))
lines(x, y, pch=22, col="black", lwd=3, lty = 2)
abline(v=mean(expmeans), lwd="4", col="yellow ")
```



From the previous graph we can conclude that the distribution of the means of our exponentials seems to follow a normal distribution with mean equal to 5, thus fulfilling the central limit theorem. By increasing the sample size, the distribution would be even closer to the normal distribution. The dotted black line on the graph is a normal distribution curve and we can see that it is very close to our sampled curve, which is the solid blue line.