Statistical Inference Course Project #1

Nikolaos Perdikis

18 7 2019

Simulation Exercise

Exercise scope, rephrased from requirements: This project 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. Parameter lambda will be set to 0.2 for all of the simulations. The distribution of averages of 40 exponentials is investigated and for this, a thousand simulations will be needed.

First, create the distribution:

```
# Known parameters
lambda <- 0.2
sims <- 1000
#create uniqueness of our random generator
set.seed(555);
# Generate 40 random exponentials
n <- seq(0,40,length.out = 1000)
# Define the Exponential Distro
my_dexp <- dexp(n,rate=lambda)
#...and put it in a dataframe (required for ggplot)
df_dexp <- data.frame(x=n,my_dexp)
```

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

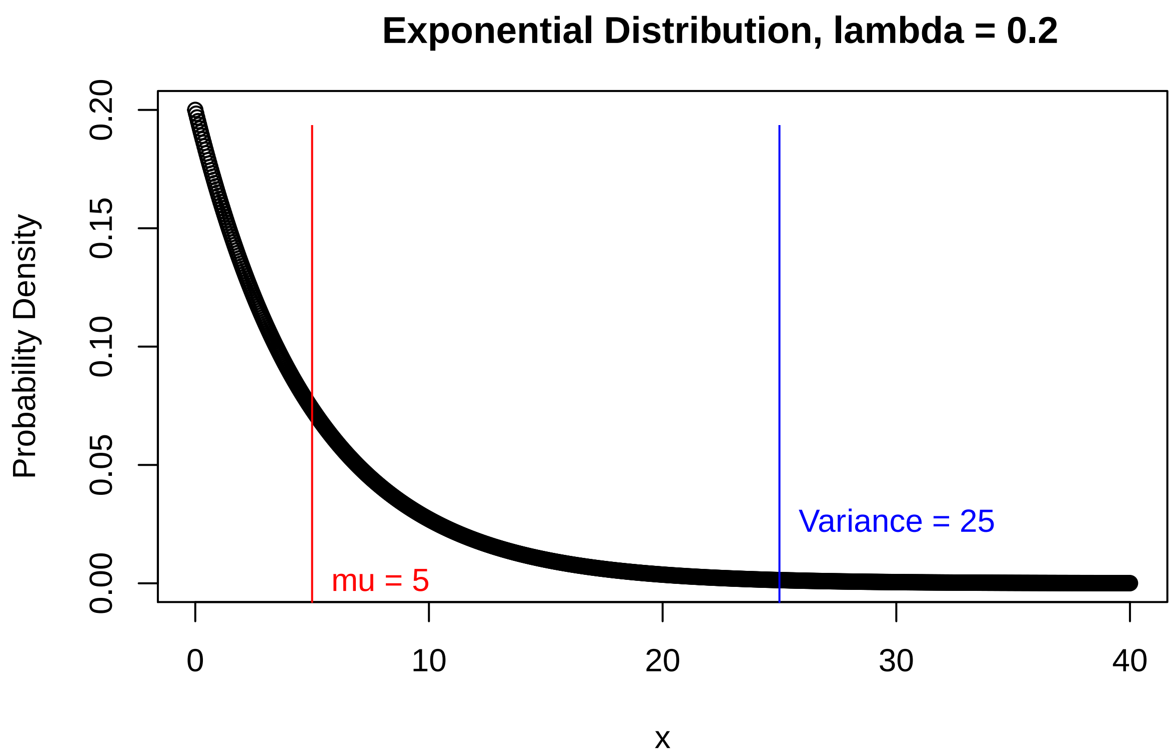
The formulae for the computations below are available in Wikipedia https://en.wikipedia.org/wiki/Exponential_distribution

```
# Based on link
mu <- sigma <- 1 / lambda
#Variance is the square of stdev
varExp <- sigma^2
varExp
```

```
## [1] 25
```

Let's plot the distribution:

```
with(df_dexp, plot(n, my_dexp,xlab="x",ylab="Probability Density"))
par(mar = c(3, 5, 5, 0))
title(main = "Exponential Distribution, lambda = 0.2")
text(0, lambda, labels = "lambda", pos = 4)
abline(v = mu, col = "red")
muLabel <- paste("mu =", mu)
text(mu, 0, labels = muLabel, pos = 4, col = "red")
abline(v = varExp, col = "blue")
varExpLabel <- paste("Variance =", varExp)
text(varExp, 0.025, labels = varExpLabel, pos = 4,
     col = "blue")
```



Result: So far we have calculated the mean mu of the distribution to be 5 and the variance to be 25

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

On to create the 1000 simulated trials of the means of 40 exponentials:

```
set.seed(575);
sm1Kn40 <- replicate(1000,mean(rexp(40,lambda)))
summary(sm1Kn40)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.124   4.505   5.019   5.039   5.538   7.884
```

…and get its measurements, so the mean, the variance using R function var:

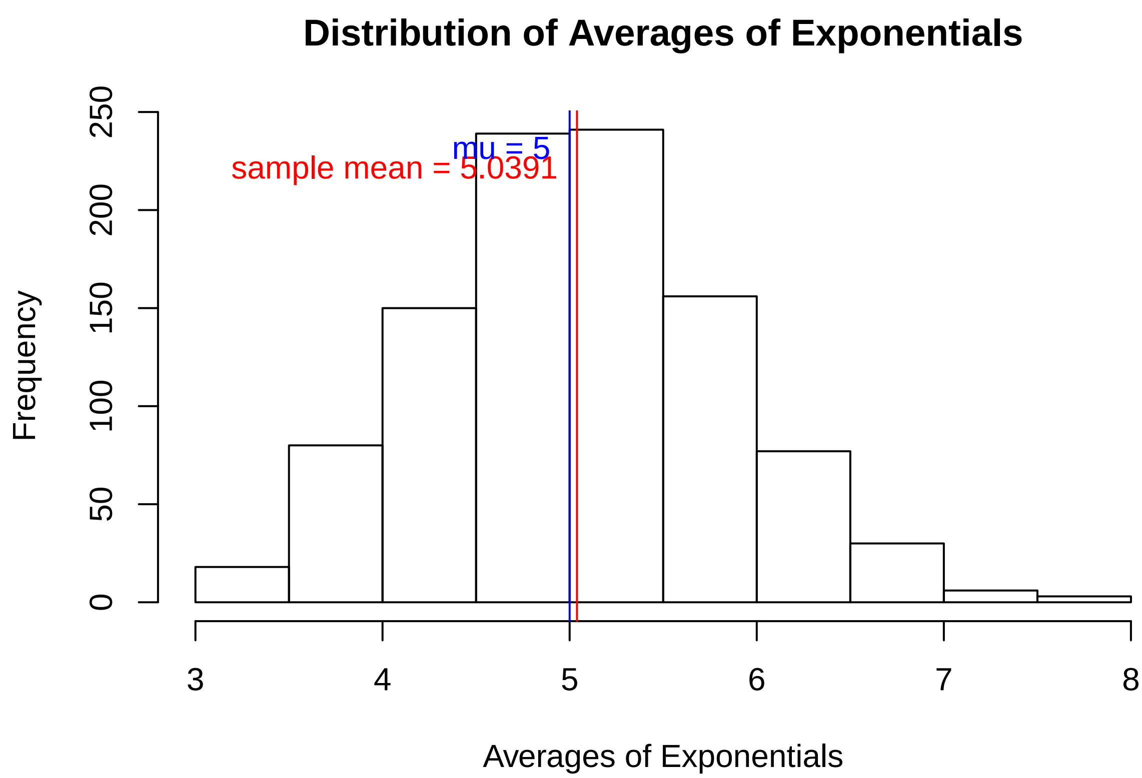
```
mu_sim <- round(mean(sm1Kn40),4)
VarSim <- var(sm1Kn40)
```

Show that the distribution is approximately normal:

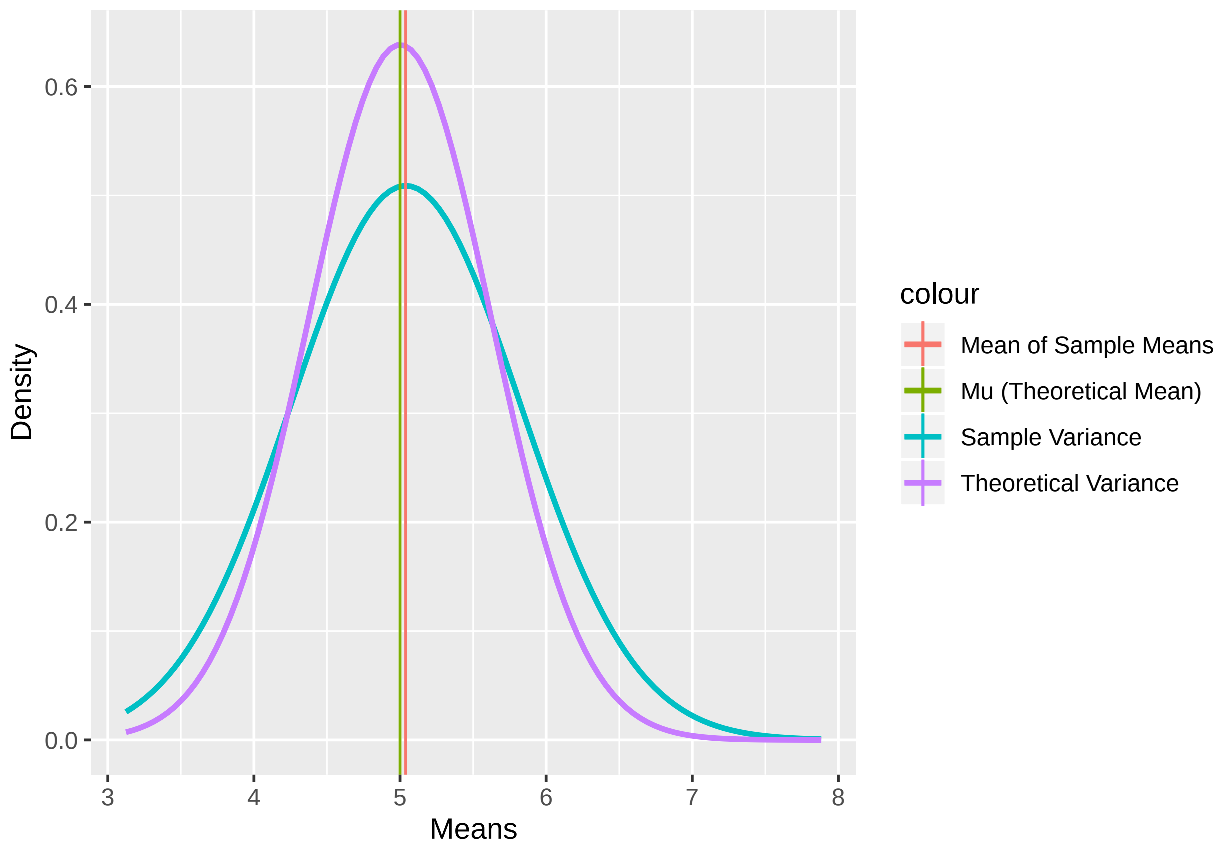
Let's plot the histogram of the averages of the 40 exponentials, simulated 1000 times:

```
hist(sm1Kn40,xlab="Averages of Exponentials",main="Distribution of Averages of Exponentials")
abline(v = round(mu_sim,4), col = "red")
muSimLabel <- paste("sample mean =", mu_sim)
text(mu_sim, 220, labels = muSimLabel, pos =2,
     col = "red")
```

```r
abline(v = mu, col = "blue")
text(mu, 230, labels = muLabel, pos = 2, col = "blue")
```




```r
g <- ggplot(as.data.frame(sm1Kn40), aes(sm1Kn40)) +
        xlab("Means") +
        ylab("Density")
g <- g + stat_function(fun = dnorm,
                       args = list(mean = mu_sim,
                                   sd = sqrt(VarSim)),
                       aes(color = "Sample Variance"),
                       size = 1) +
        stat_function(fun = dnorm,
                      args = list(mean = mu,
                                  sd = sigma^2/40),
                      aes(color = "Theoretical Variance"),
                       size = 1)
g <- g + geom_vline(aes(xintercept = mu_sim, color =
                                "Mean of Sample Means")) +
        geom_vline(aes(xintercept = mu, color =
                               "Mu (Theoretical Mean)"))
g
```



Conclusions

It was expected, that the mean of sample means would approach the theoretical mean mu. This is due to the Central Limit Theorem. We also see the sample variance having similar pattern as the theoretical variance divided by the number of exponentials. Based on the histogram of the distribution of the means, we observe that their distribution looks distinctively Gaussian