Stat Inference project part 1

Asfa Lohani

Overview

This project aims to investigate the exponential distribution 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. For this project lambda = 0.2 for all of the simulations. We will investigate the distribution of averages of 40 exponentials. Note that we will be performing a thousand simulations.

Simulation

Let's start off by simulating 40 random exponentials and taking their mean 1000 times. We set our variables accordingly as seen below and set lambda to .2 as we mentioned before. We then create a list called mns which lists the means of the 40 random exponentials 1000 times. We set the seed to ensure reproducibility.

```
set.seed(11)
mns = NULL
nosim <- 1000
samplesize <- 40
lambda <- 0.2
for (i in 1 : nosim) mns = c(mns, mean(rexp(samplesize, lambda)))
```

Sample Mean Vs. Theoretical Mean

Now that we have simulated the data, let's take a look at it. First we will look at the means. The theoretical mean of an exponential distribution is simply 1/lambda which is 5:

```
tmean<-1/lambda
tmean
```

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

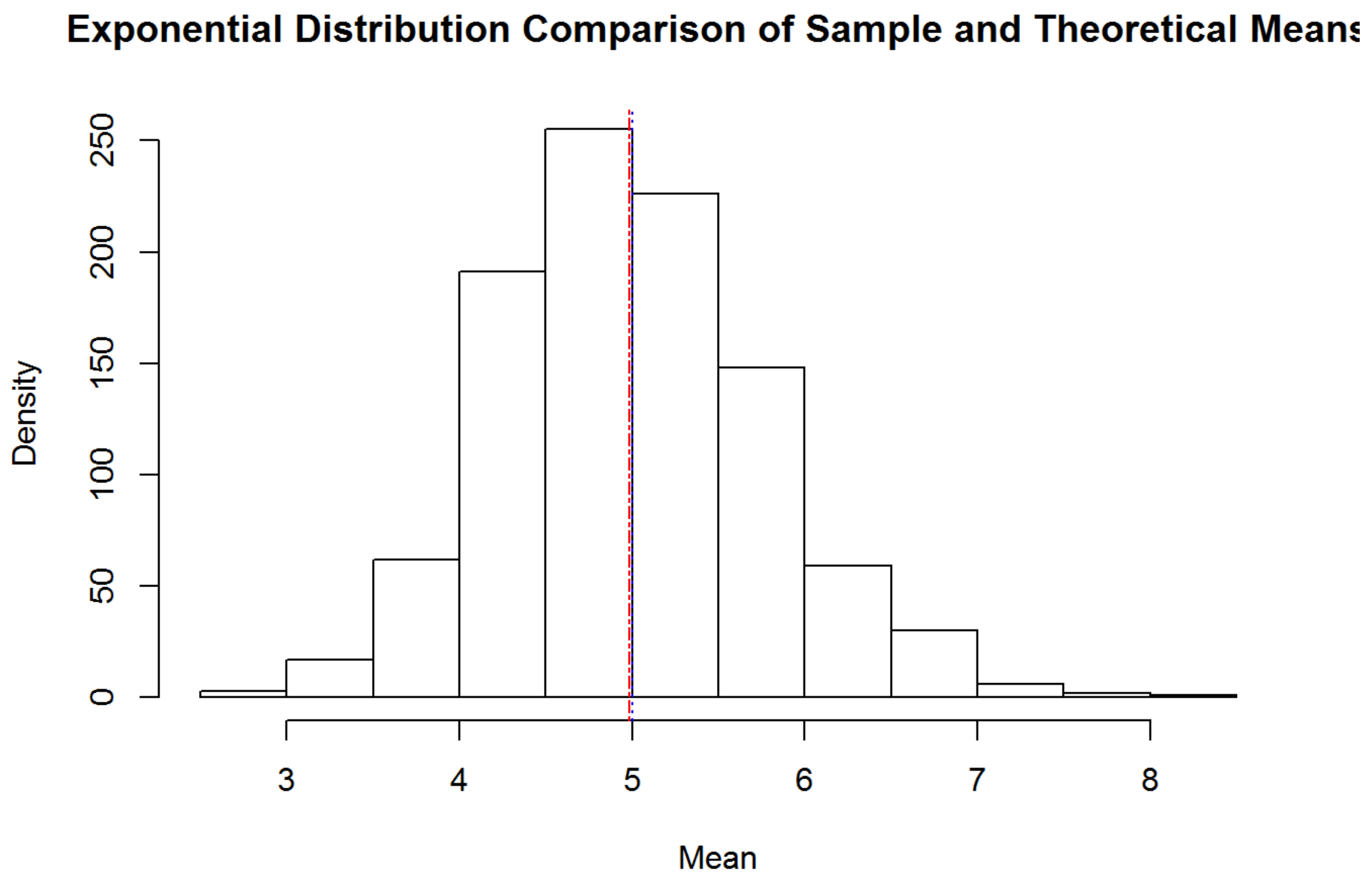
The sample mean is 4.987

```
smean<-mean(mns)
smean
```

```
## [1] 4.987157
```

Both means are very close and according to the CLT (Central Limit Theorem) the average of sample means converges to theoretical mean. Therefore we can conclude the exponential distribution follows the CLT. Lets look at the means graphically to get a better understanding.

```
hist(mns, main = "Exponential Distribution Comparison of Sample and Theoretical Means", xlab = "Mean", ylab = "Density" )
abline(v=tmean, col = "BLUE", lty = 3)
abline(v=smean, col = "RED", lty = 6)
```



The graphic above helps us visually understand that the sample mean and theoretical mean are very close as they are color coded in Blue and Red. They are so close that the lines are almost the same. So, to restate what we have already done previously - the avergae of sample means of an exponential distribution converges to the theoretical mean and therefore follows the CLT.

Sample Variance vs. Theoretical Variance

Now lets look at the variance of the sample and the theoretical distribution. The standard deviation of an exponential distribution is 1/lambda, therefore the variance is (1/lambda)^2/n. N here is equal to sample size. Let's calculate this:

```
tvar<- (1/lambda)^2/samplesize
tvar
```

```
## [1] 0.625
```

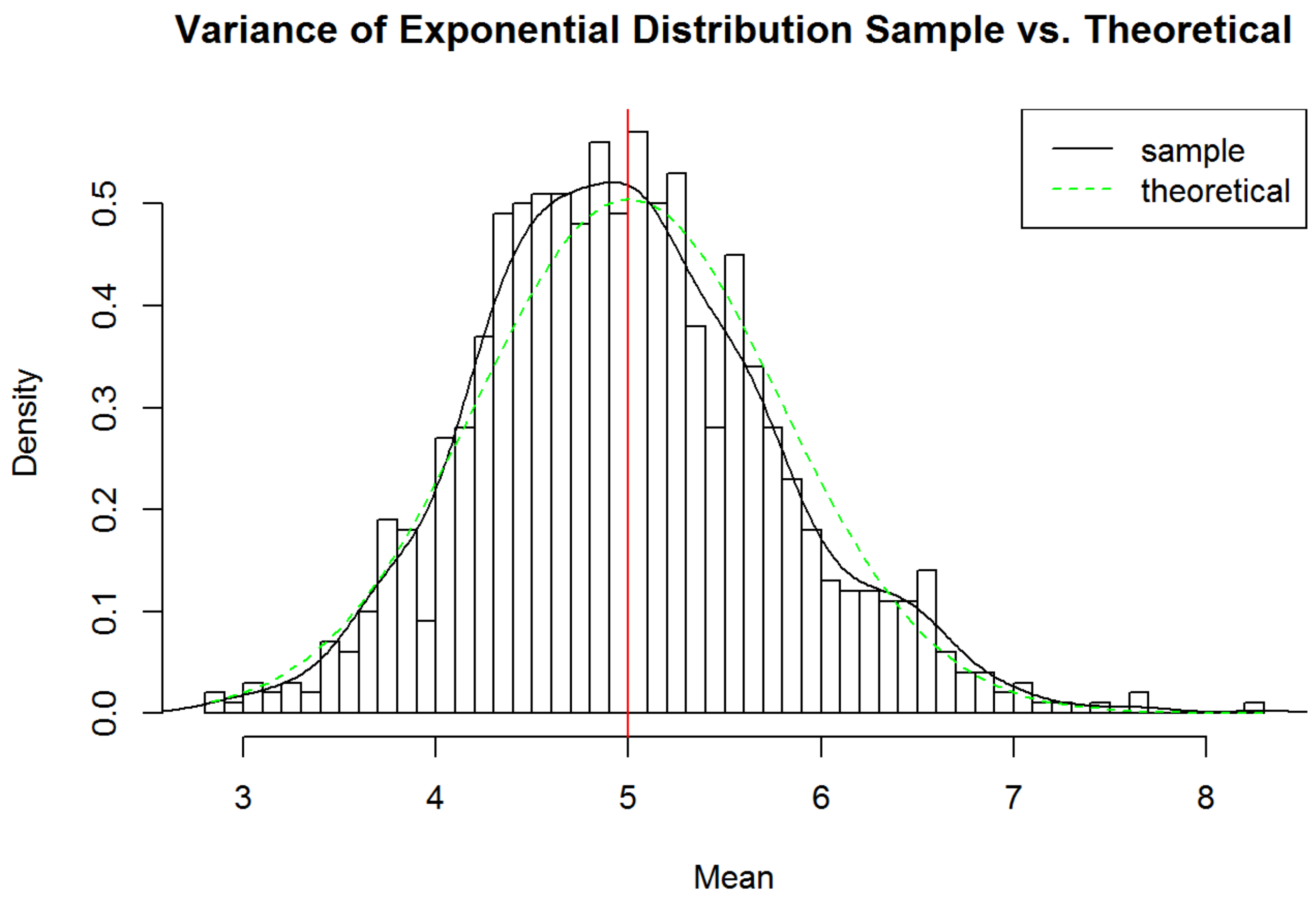
Variance of the sample is calculated as follows:

```
svar<- var(mns)
svar
```

```
## [1] 0.6009383
```

Just like the means, the variance of sample and the theoretical distributions are also quite close and therefore very similar. We can visually see the variance of the sample and compare to the theoretical in the graphic below.

```
hist(mns, breaks=50, prob=TRUE,
     main="Variance of Exponential Distribution Sample vs. Theoretical",
     xlab="Mean")
lines(density(mns))
abline(v=1/lambda, col="red")
x <- seq(min(mns), max(mns), length=100)
y <- dnorm(x, mean=1/lambda, sd=(1/lambda/sqrt(samplesize)))
lines(x, y, pch=20, col="green", lty=2)
legend('topright', c("sample", "theoretical"), lty=c(1,2), col=c("black", "green"))
```



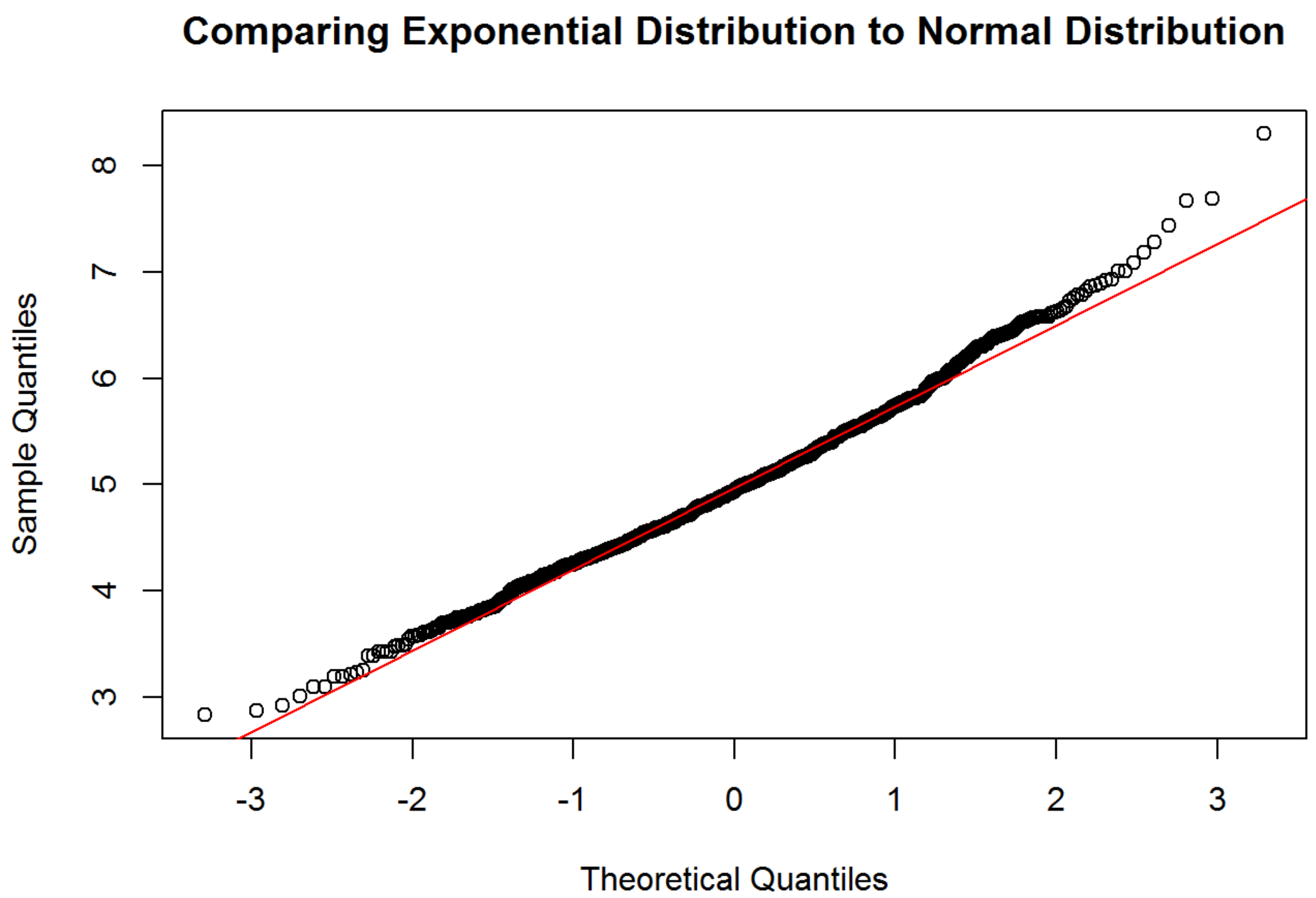
The graphic above shows how close the variance of the sample and theoretical distributions are with the black and green lines. The sample distribution has a slight left shift in comparison with the theoretical. The theoretical distribution also looks quite Gaussian (normal) which we will discuss in the next section.

Distribution

The histogram gives us a pretty good idea that the exponential distribution is very close to that of a normal distribution. Here we will further prove that the distribution is approximately normal.

We will use the normal probability plot with the qqnorm function and qqline to compare the exponential distribution to a normal distribution.

```
qqnorm(mns, main = "Comparing Exponential Distribution to Normal Distribution")
qqline(mns, col = 2)
```



From the above graphic we can conclude that the sample exponential distribution is a very close approximate to the normal distribution.