Statistical Inference, Assigment 1

Part 1: A Simullation Exercise

In this project we investigate the exponential distribution in R and compare it with the Central Limit Theorem. For the simulations, we set lambda = 0.2. We investigate the distribution of averages of 40 exponentials. We do 1000 simulations.

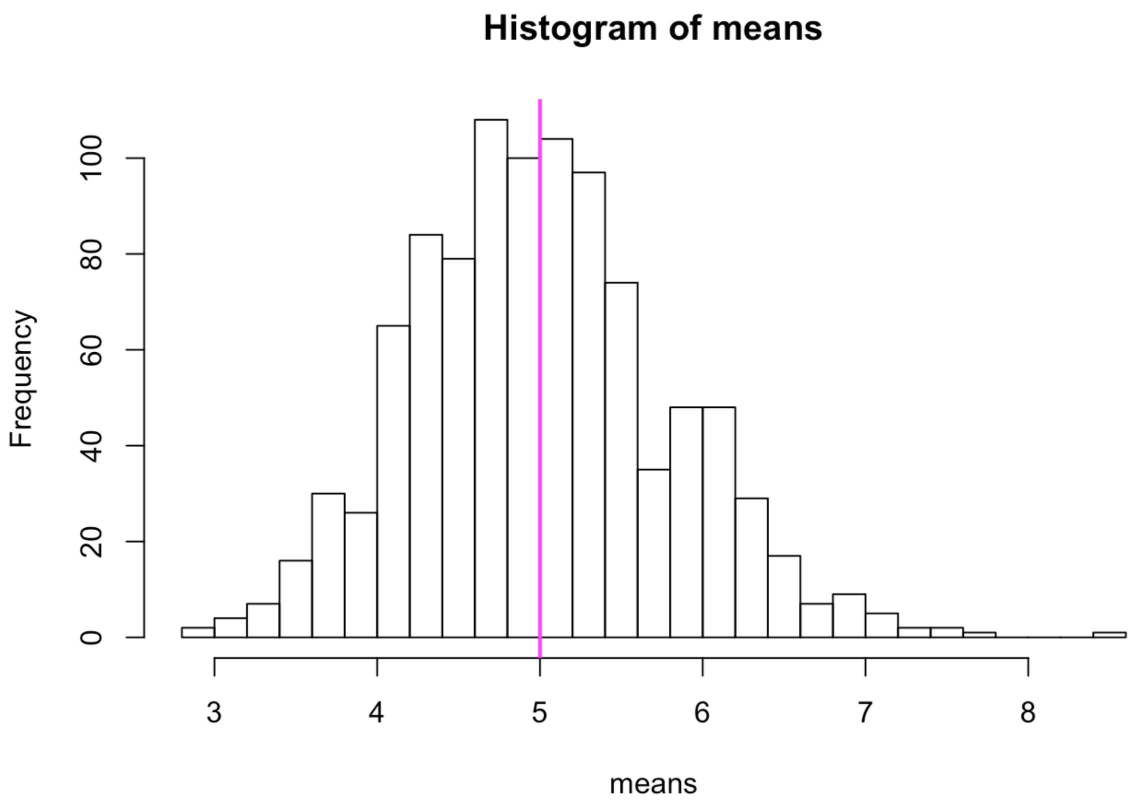
First we set the seed for replicability, then initialize the variables and we do the simulations using a code from the slides.

```
set.seed(100)
n <- 1000
num <- 40
lambda <- .2
simulations <- matrix(rexp(n*num, lambda), n, num)
```

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

We calculate the means of the 1000 simulations and then we plot a histogram

```
# get the means of the simulations
means <- apply(simulations,1, mean)
# plot the results
hist(means, breaks = 30)
abline(v=1/lambda, col= "magenta", lwd=2)
```



Then we calculate the mean of this new distribution, which is the Sample Mean and compare it to the theoretical mean, which is 1/lambda:

```
# simulation mean
sample_mean <- mean(means)
print(sample_mean)
```

```
## [1] 4.999702
```

```
# theoretical mean
population_mean <- 1/lambda
print(population_mean)
```

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

We can see that the means are very close.

Compare the sample variance to the theoretical variance of the distribution.

Now we calculate the sample variance using sd function from R and then we calculate the theoretical variance.

```
    # simulation variance
    sample_sd <- sd(means)
    sample_var <- sample_sd^2
    print(sample_var)
```

```
## [1] 0.6335302
```

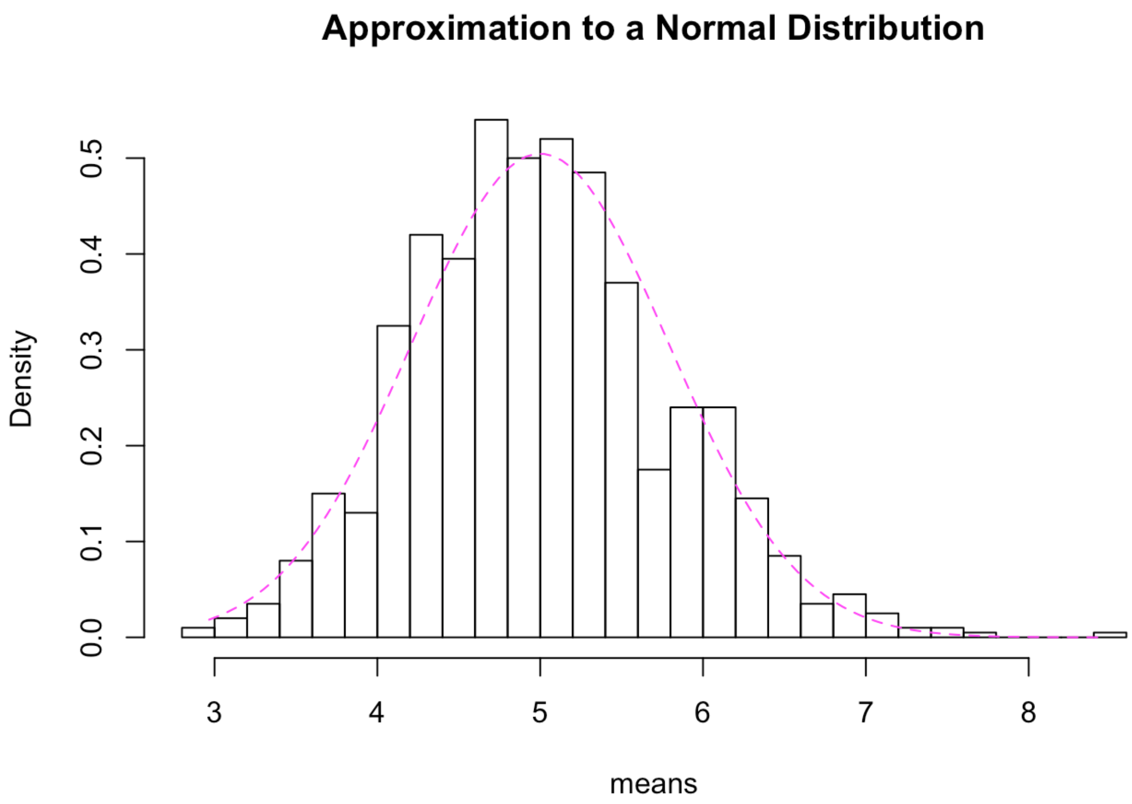
```
    # theoretical variance
    population_sd <- (1/lambda)/sqrt(num)
    population_var <- population_sd^2
    print(population_var)
```

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

We can see again that the variances are very similar.

Show that the distribution is approximately normal

```
    # plot the histogram of averages
hist(means, breaks = 30, prob=T, main = "Approximation to a Normal Distribut
ion")
# theoretical density of the averages of samples
xfit <- seq(min(means), max(means), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(num)))
lines(xfit, yfit, pch=22, col="magenta", lty=2)
```



We can see in the plot that the shape of the distribution is very similar to a Normal Distribution.