Statistical Inference Course Project

Part 1: Simulation Excercise

It is good practice to set a defined seed on an experiment to improve reproducibility.

```
set.seed(240598)
```

We then assign the data given to us for this experiment

```
lambda <- 0.2
n <- 40
nsim <- 1000
```

Then the experiment is made through the replicate function and the mean of the exponentials is calculated

```
sim <- replicate(nsim, rexp(n, lambda))
eMean <- apply(sim,2,mean)
```

Comparison of the Means

The first question of the first part is to compare the sample mean and the theoretical mean. We start by calculating both:

```
sMean <- mean(eMean)
sMean
```

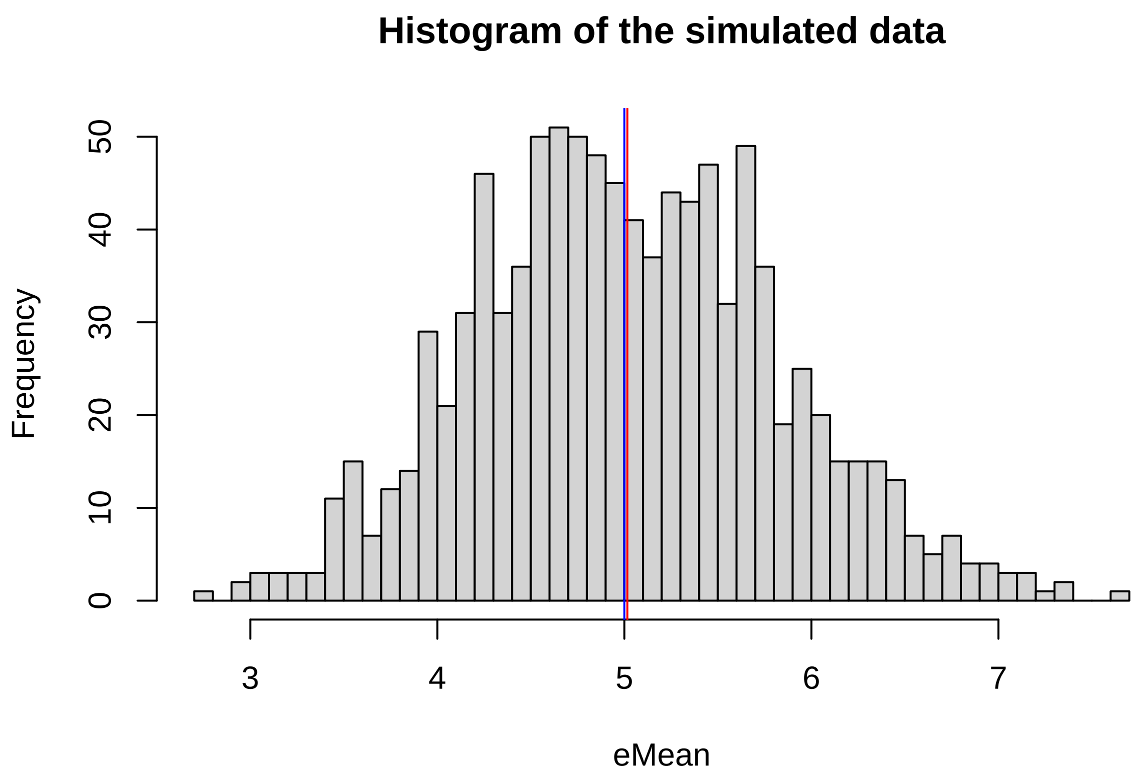
```
## [1] 5.016071
```

```
tMean <- 1/lambda
tMean
```

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

The histogram is then plotted. Both means are also plotted as lines to do the comparison

```
hist(eMean, breaks = 50, main = "Histogram of the simulated data")
abline(v = sMean, col = "red")
abline(v = tMean, col = "blue")
```



We can clearly see that both means almost overlap eachother, since both are very close to 5.

Comparison of the Variances

As we did before, we calculate both variances:

```
sSd <- sd(eMean)
sVar <- sSd^2
sVar
```

```
## [1] 0.6599058
```

```
tSd <- (1/lambda)/sqrt(n)
tVar <- tSd^2
tVar
```

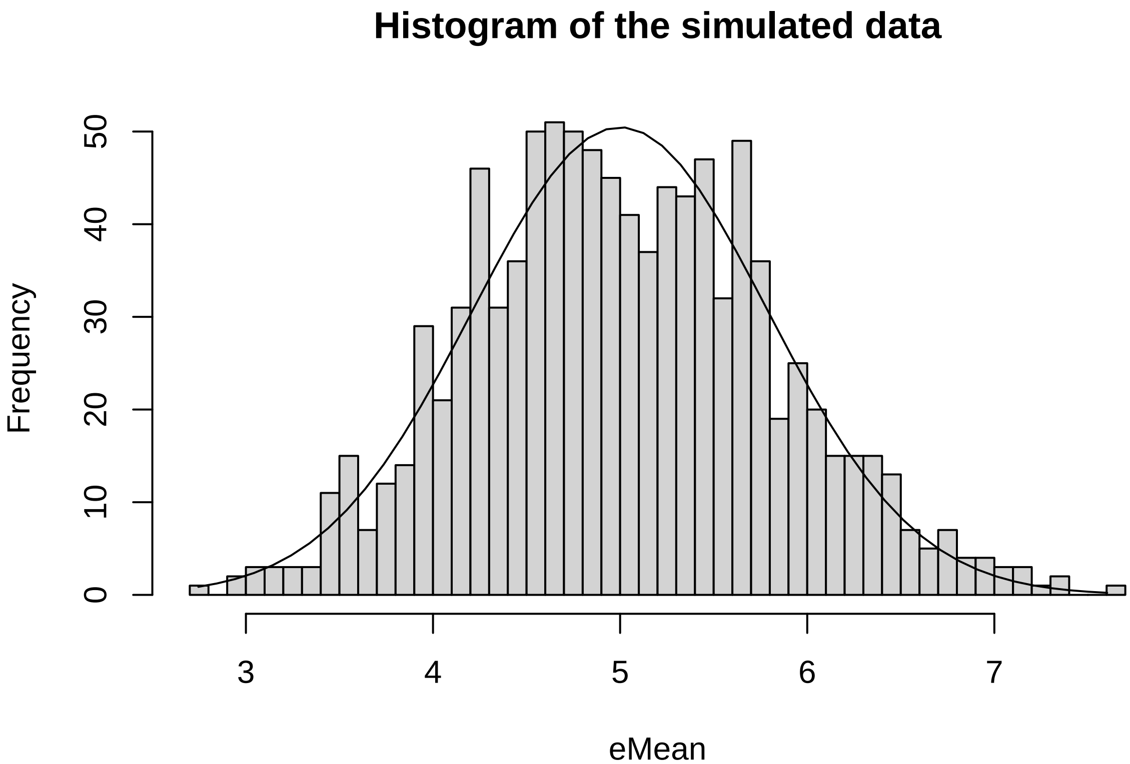
```
## [1] 0.625
```

We therefore proved that both variances are very close to each other

Distribution

The Central Limit Theorem states that the simulation should follow a normal distribution

```
hist(eMean, breaks = 50, main = "Histogram of the simulated data")
xfit <- seq(min(eMean),max(eMean), length = 50)
yfit <- dnorm(xfit, mean = tMean, sd = tSd)
lines(xfit, yfit * 100, lty = 1)
```



This figure proves that the distribution of means of the sampled exponential distributions follow a normal distribution.