Final project - Part 1

Raphael Carvalho

In this paper we're going to 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. Set lambda = 0.2 for all of the simulations. A distribution of averages of 40 exponentials will be investigated.

Loading the required packages

```
library(data.table)
library(ggplot2)
```

Setting the variables of the problem:

```
  simulations <- 1000
  lambda <- 0.2
  quantile <- 1.96
  n <- 40
  set.seed(1000)
```

Matrix with 1000 simulations each with 40 samples drawn from the exponential distribution.

```
  data <- matrix(rexp(n * simulations, rate = lambda), simulations)
  #Mean of each row of the matrix
  row.mean <- rowMeans(data)
```

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

```
  sample.mean <- mean(row.mean)
  theoretical.mean <- 1 / lambda
  print(paste("Sample mean:", sample.mean))
  print(paste("Theoretical mean:", theoretical.mean))
```

```
## [1] "Sample mean: 4.98696338580556"
## [1] "Theoretical mean: 5"
```

Theoretical mean and sample mean are pretty close!

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

Comparing variances

```
sample.variance <- var(row.mean)
theoretical.variance  <- (1 / lambda)^2 / (n)
print(paste("Sample variance:", sample.variance))
print(paste("Theoretical variance:", theoretical.variance))
```

```
## [1] "Sample variance: 0.658355065045304"
## [1] "Theoretical variance: 0.625"
```

Theoretical variance and sample variance are not as close as when we compared the theoretical mean against the sample mean, but they're still pretty close!

Comparing standard deviations

```
sample.sd <- sd(row.mean)
theoretical.sd  <- 1/(lambda * sqrt(n))
print(paste("Sample standard deviation:", sample.sd))
print(paste("Theoretical standard deviation:", theoretical.sd))
```

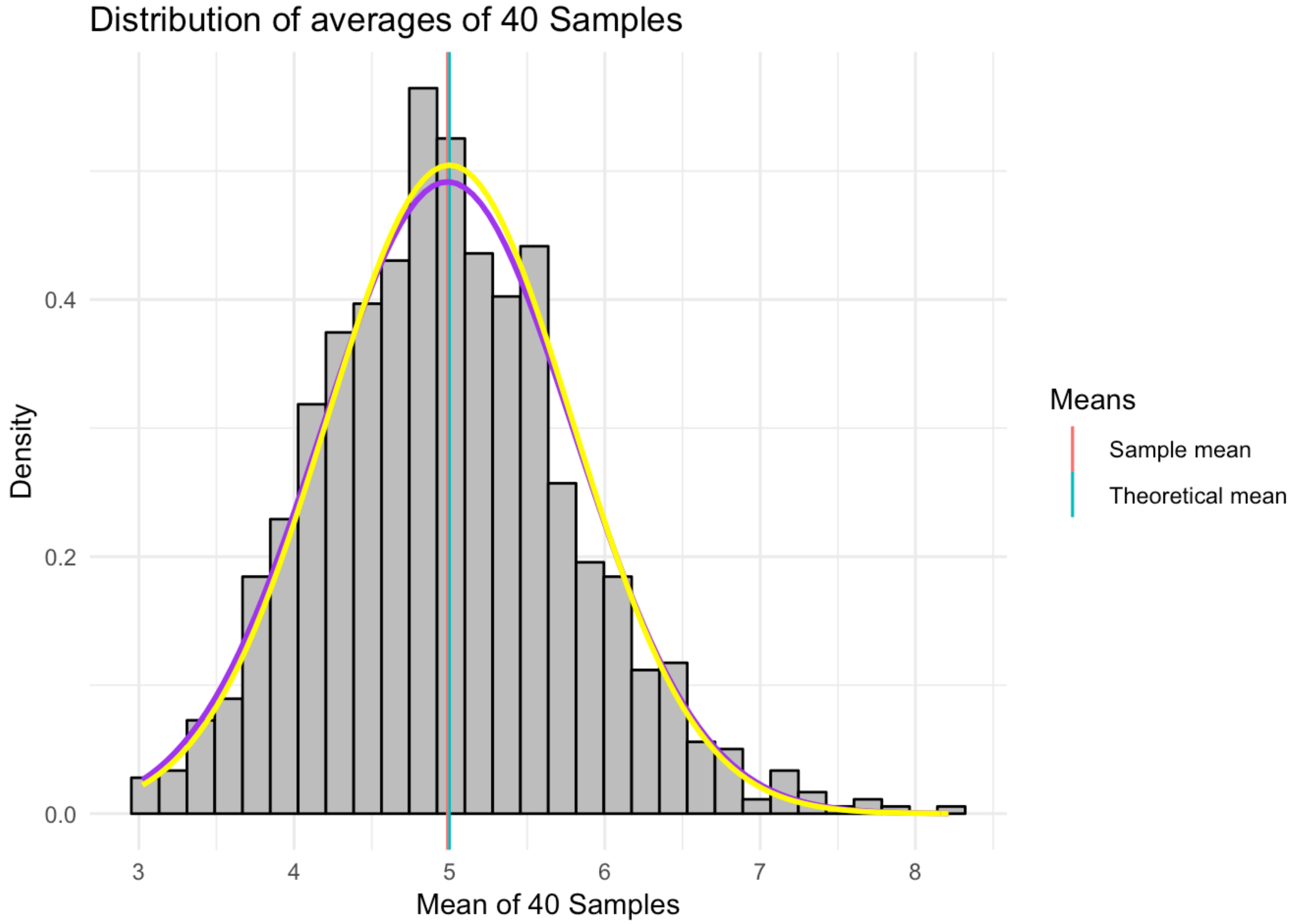
```
## [1] "Sample standard deviation: 0.811390821395771"
## [1] "Theoretical standard deviation: 0.790569415042095"
```

Again, the values are pretty close to each other.

3. Show that the distribuction is approximately normal.

```
  plotdata <- data.frame(row.mean)
  histogram <- ggplot(plotdata, aes(x = row.mean))
  histogram <- histogram + geom_histogram(aes(y=..density..), colour="black",
                          fill = "grey")
  histogram <- histogram + labs(title = "Distribution of averages of 40 Samples",
x = "Mean of 40 Samples", y = "Density", color = "Means")
  histogram <- histogram + geom_vline(aes(xintercept = sample.mean, colour = "Samp
le mean"))
  histogram <- histogram + geom_vline(aes(xintercept = theoretical.mean, colour =
"Theoretical mean"))
  histogram <- histogram + stat_function(fun = dnorm, args = list(mean = sample.me
an, sd = sample.sd), color = "purple", size = 1.0)
  histogram <- histogram + stat_function(fun = dnorm, args = list(mean = theoretic
al.mean, sd = theoretical.sd), colour = "yellow", size = 1.0)
  histogram <- histogram + theme_minimal()
  print(histogram)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



As you can see on the histogram above, de distribution follows a gaussian (normal) distribution. The density line in yellow shows the normal curve formed by the the theoretical mean and standard deviation and the density line in purple shows the normal curve formed by the the sample mean and standard deviation.

Confidence intervals

```
  sample.ci <- round (mean(row.mean) + c(-1,1)*1.96*sd(row.mean)/sqrt(n),3)
  theoretical.ci <- theoretical.mean + c(-1,1) * 1.96 * sqrt(theoretical.variance)
/sqrt(n)
  print(paste("Sample confidence interval:", sample.ci))
  print(paste("Theoretical confidence intercal:", theoretical.ci))
```

```
## [1] "Sample confidence interval: 4.736" "Sample confidence interval: 5.238"
## [1] "Theoretical confidence intercal: 4.755"
## [2] "Theoretical confidence intercal: 5.245"
```

The confidence levels also match closely, proving the distribution is approximately normal.