Project-part1

Willem Hogewoning

7-2-2021

Part 1: Simulation Exercise Instructions

In this project you 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. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Stimulation

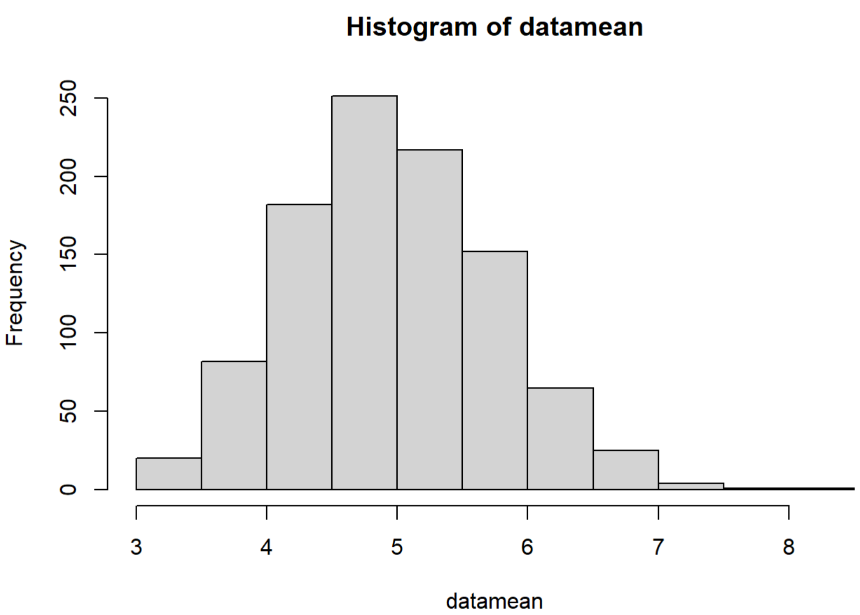
Bot means for the sample and the one according the Central Limit Theorem (CLT) are calculated below.

```
#parameters
lambda <- 0.2
mean <- 1/lambda
sd <- 1/lambda
n <- 40
simulations <- 1000
set.seed(12345)

#stimulation
datamatrix <- matrix(rexp(simulations * n, rate=lambda), simulations, n)
datamean <- rowMeans(datamatrix)
```

Show the plot of the stimulation

```
hist(datamean)
```



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

```
stimulationmean <- mean(datamean)
clt_mean <- 1/lambda


stimulationvariance <- var(datamean)
clt_varieance <- (1/lambda)^2/n
```

The mean

The actual mean versus the CLT mean:

```
#Mean
stimulationmean
```

```
## [1] 4.971972
```

```
clt_mean
```

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

The actual mean is 4.97 versus the CLT mean of 5.00; quite close to each other

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

The variance

The actual mean versus the CLT mean:

```
#Variance
stimulationvariance
```

```
## [1] 0.6157926
```

```
clt_varieance
```

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

The actual varience is 0.62 verus the CLT variance of 0.63; quite close each other

3. Show that the distribution is approximately normal.

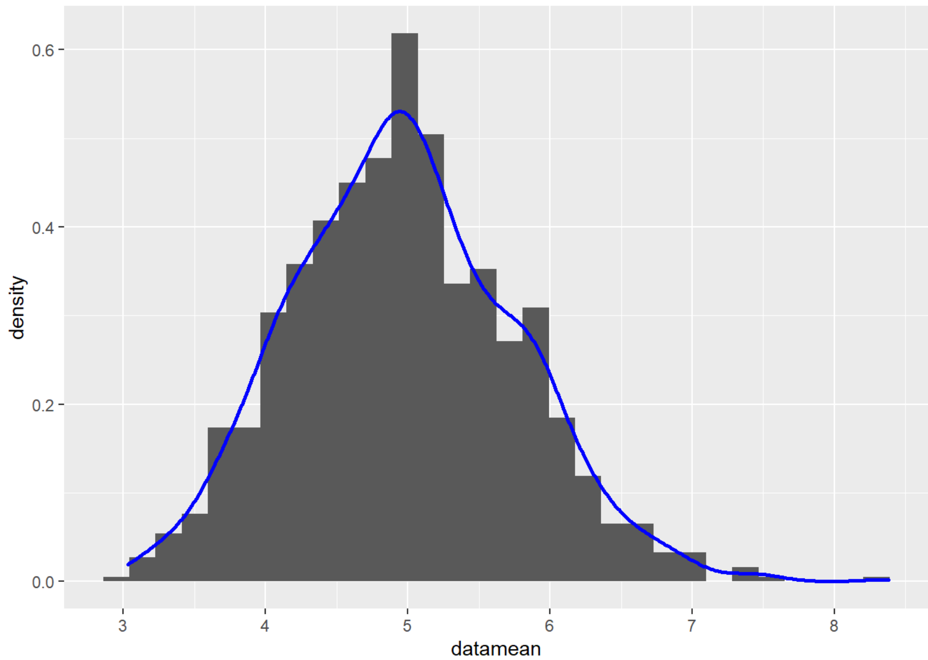
First we show approximately by normal by looking at the shape:

```
library(ggplot2)

library(ggplot2)

plotdata <- data.frame(datamean);

plot <- ggplot(plotdata, aes(x =datamean))
plot <- plot + geom_histogram(aes(y=..density..),bins=30)
plot + geom_density(colour="blue", size=1)
```



as one can see the shape looks like a normal distribution

Second we look at the confidence intervals

```
stimulation_confidence_interval  <- round (mean(datamean) + c(-1,1)*1.96*sd(datamean)/sqrt(n),3)
clt_confidence_interval <- clt_mean + c(-1,1)*1.96*sqrt(clt_varieance)/sqrt(n)

stimulation_confidence_interval
```

```
## [1] 4.729 5.215
```

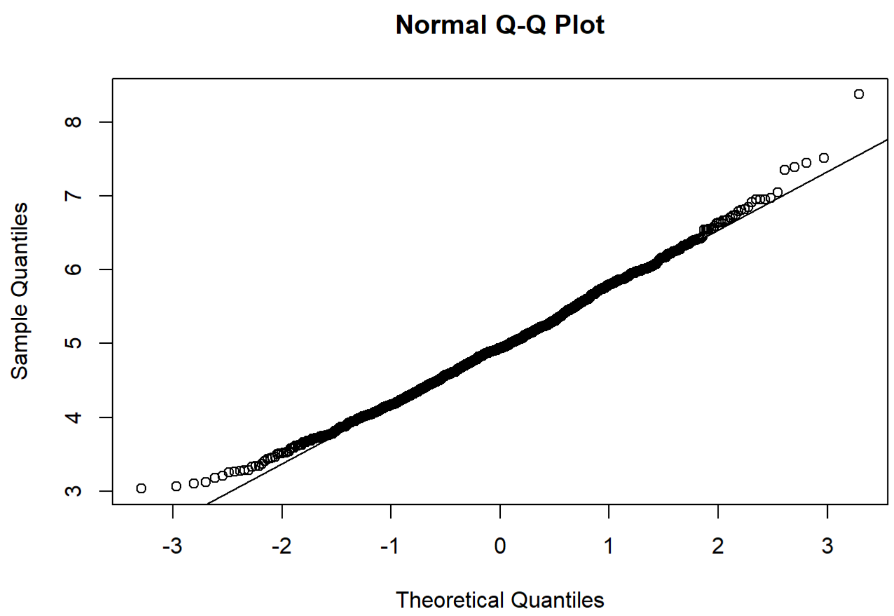
```
clt_confidence_interval
```

```
## [1] 4.755 5.245
```

The actual 95% confidence interval is [4.729 5.215] the CTL 95% confidence interval is [4.749 5.254] both quite close to each other

Third we look at the Q-Q plot

```
qqnorm(datamean); qqline(datamean)
```



Also here we can see it lies quite inline.