Statistical Inference Course Project

m_denboeft

15-3-2018

PART 1 For this project I investigated the exponential distribution in R and compared 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. Lambda = 0.2 was set for all of the simulations. I investigated the distribution of averages of 40 exponentials. I was needed to do a thousand simulations. Assignment: Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

Load plotting library

```
library(ggplot2)
```

Set variables

```
n <- 40
lambda <- 0.2
nrsim <- 1000
quantileCI <- 1.96
set.seed(234)
```

Create a matrix of 1000 rows with colums corresponding to 40 times random simulations

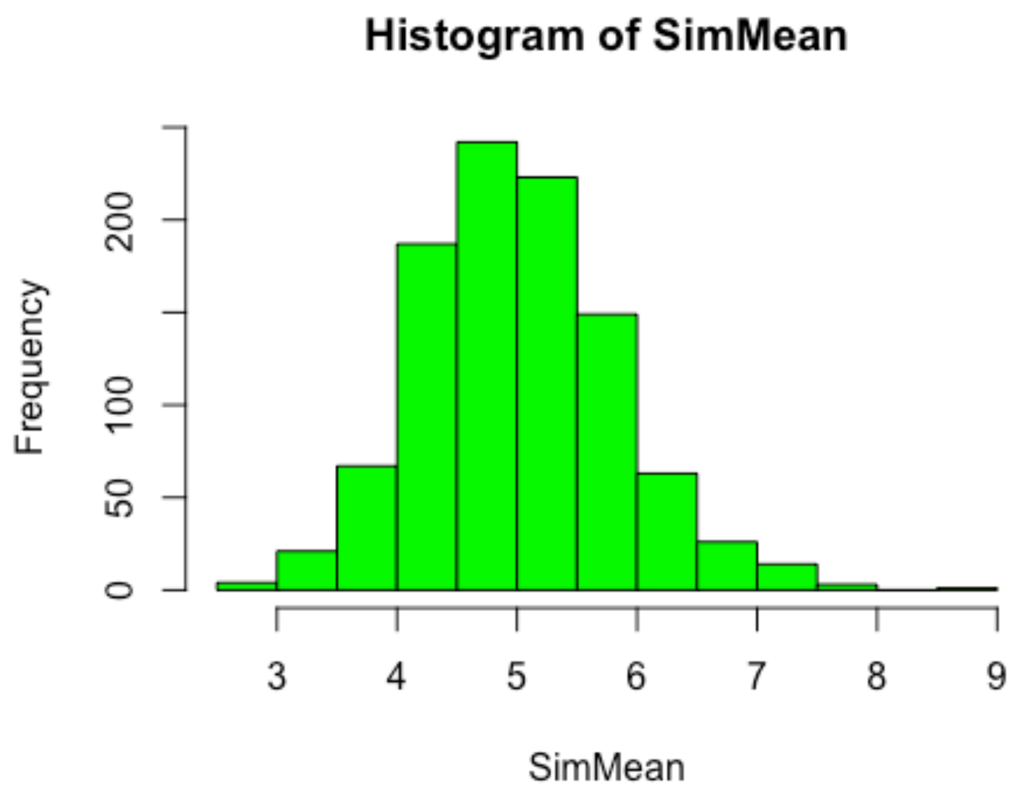
```
SimData <- matrix(rexp(n * nrsim, rate = lambda), nrsim)
```

Calculate the averages across the 40 values for each of the 1000 simulations

```
SimMean <- rowMeans(SimData)
```

Create a histogram to explore the plot

```
hist(SimMean, col="green")
```



SHOW THE SAMPLE MEAN AND COMPARE IT TO THE THEORETICAL MEAN

Calculate the sample data mean

```
SampleMean <- mean(SimMean)
SampleMean

## [1] 5.001573
```

Calculate the theoretical mean

```
TheoreticalMean <- 1 / lambda
TheoreticalMean

## [1] 5
```

==> So the sample mean is 5.001 and the Theoretical Mean is 5 so almost the same

SHOW THE SAMPLE VARIANCE AND COMPARE IT TO THE EXPECTED VARIANCE

Calculate the variance

```
SampleVariance <- var(SimMean)
SampleVariance

## [1] 0.6631504
```

Calculate the theoretical (expected) variance

```
TheoreticalVar  <- (1 / lambda)^2 / (n)
TheoreticalVar

## [1] 0.625
```

==> So the sample variance is 0.663 and the expected variance is 0.625 so also almost the same

SHOW THE SAMPLE MEAN SD AND COMPARE IT TO THE THEORETICAL MEAN SD

Calculate the sample mean standard deviation

```
SampleSD <- sd(SimMean)
SampleSD

## [1] 0.8143405
```

Calculate the theoretical standard deviation

```
TheoreticalSD <- 1/(lambda * sqrt(n))
TheoreticalSD

## [1] 0.7905694
```

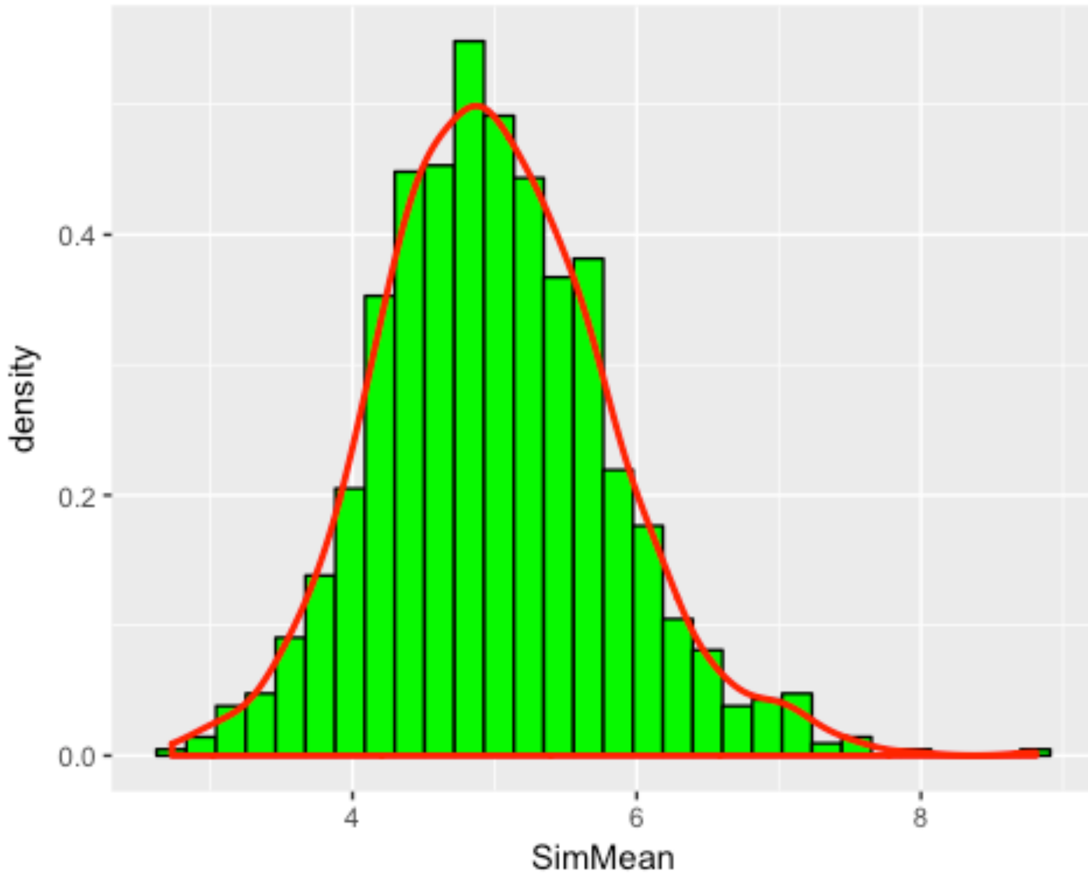
==> So the sample SD is 0.814 and the theoretical SD is 0.791 so also almost the same

SHOW THAT THE DISTRIBUTION IS APPROXIMATELY NORMAL BY

1. By plotting the data
2. By calculating and matching the confidence intervals
3. By calculating and matching the theoretical quantiles in a plot

4. Plotting the data

```
plotdata <- data.frame(SimMean)
m <- ggplot(plotdata, aes(x = SimMean))
m <- m + geom_histogram(aes(y=..density..), colour="black",
                        fill = "green")
m + geom_density(colour="red", size=1);

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



==> The plot shows an approximate normal distribution

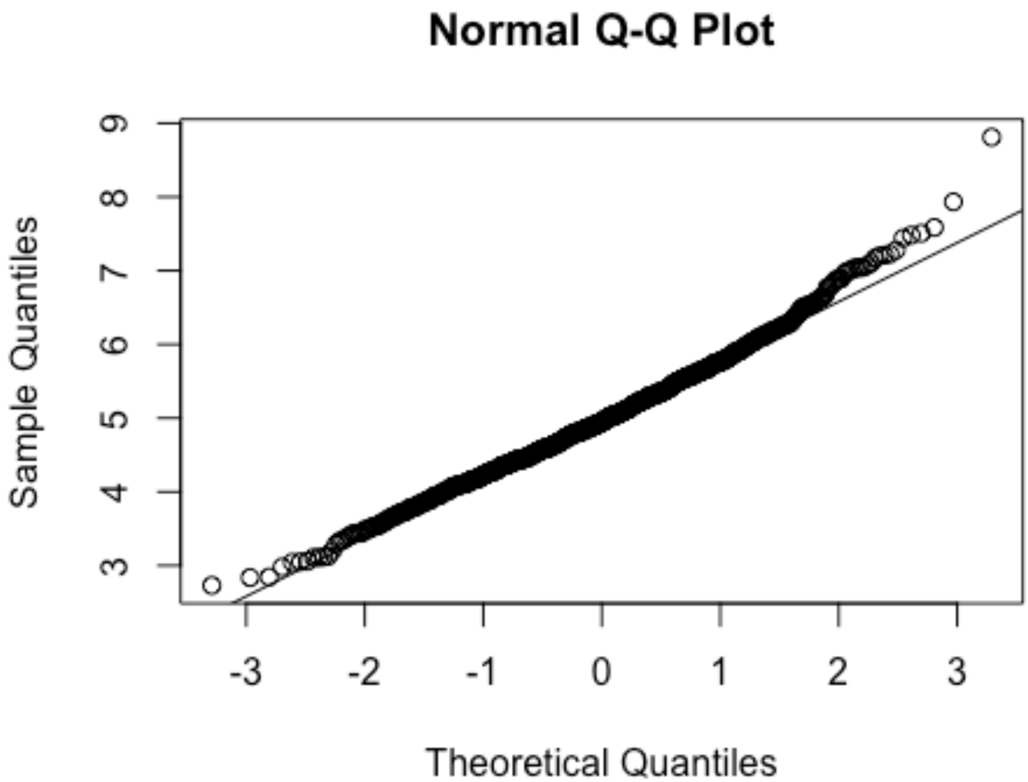
2. Match the confidence intervals

```
SampleCI <- round (mean(SimMean) + c(-1,1)*1.96*sd(SimMean)/sqrt(n),3)
TheoreticalCI <- TheoreticalMean + c(-1,1)*1.96*sqrt(TheoreticalVar)/sqrt(n);

SampleCI

## [1] 4.749 5.254

TheoreticalCI

## [1] 4.755 5.245
```

==> The sample CIs are 4.75 - 5.25 and the theoretical CIs are 4.75 - 5.25 so even, matching an normaldistribution

3. Matching the theoretical quantiles

```
qqnorm(SimMean); qqline(SimMean)
```



==> The sample quantiles match with the theoretical quantiles, so the distribution is approximately normal