Simulation - Course Project - Part I

D. Nienhold

14 Feb 2015

Requirements

Library ggplot2 is required.

```
library(ggplot2)
```

Overview

In this report samples from a given exponential distribution with mean 1/lambda and standard deviation 1/lambda are taken in order to show the behaviour of the sample mean, the sample variance and their respective distribution. Lamda is 0.2 for the given exponential distribution. The sample mean and variance are also compared to the theoretical mean and theoretical variance.

Exponential Distribution

The given exponential distribution has the following properties:

$$Mean \ = 1/\lambda$$

$$Standard \ \ Deviation \ = 1/\lambda$$

$$with \ \ \lambda \ = 0.2$$

Simulation

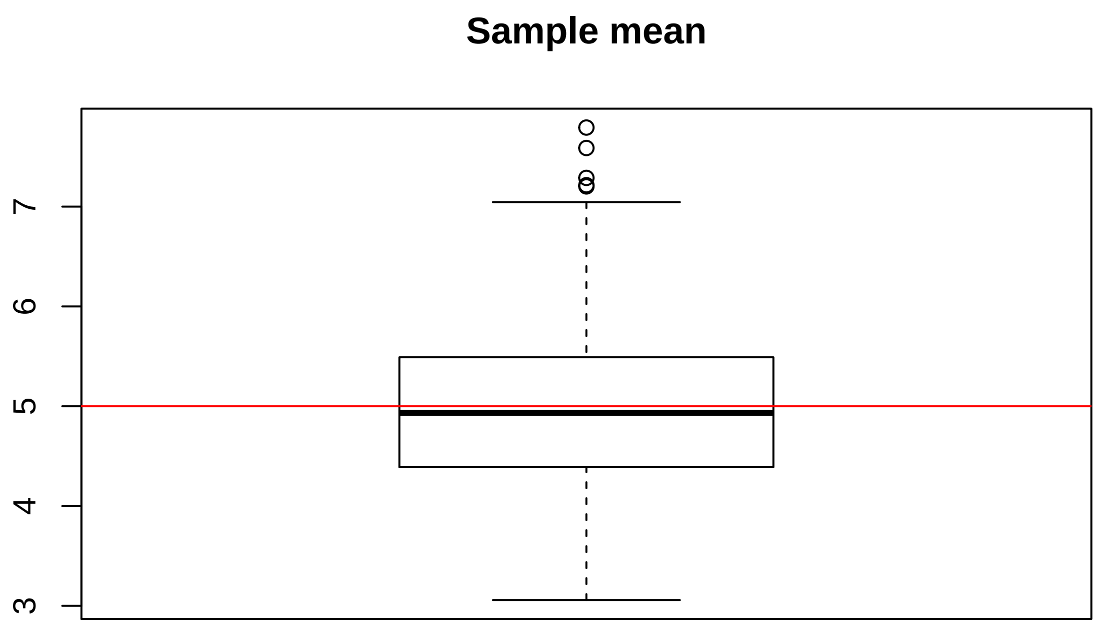
1000 simulations are run with 40 samples each. The function rexp is used to create the simulation of an exponential distribution using a lambda value of 0.2. Each simulation is stored in a column of the matrix data.

```
numberSimulation <- 1000
numberSample <- 40
lambda <- 0.2
data <- matrix(rexp(numberSample * numberSimulation, lambda), numberSample)
```

Sample Mean versus Theoretical Mean

In this section the sample mean of the simulation is shown and compared to the theoretical mean. The theoretical mean is equal to 1/lambda = 5. The barplot shows that the sample mean is very near to this value (marked with red line) which is confirmed by the summary calculation for the simulations

```r
df <- data.frame(x=apply(data,2,mean))
boxplot(df$x,main="Sample mean")
abline(h = 1/lambda, col = "red")
```



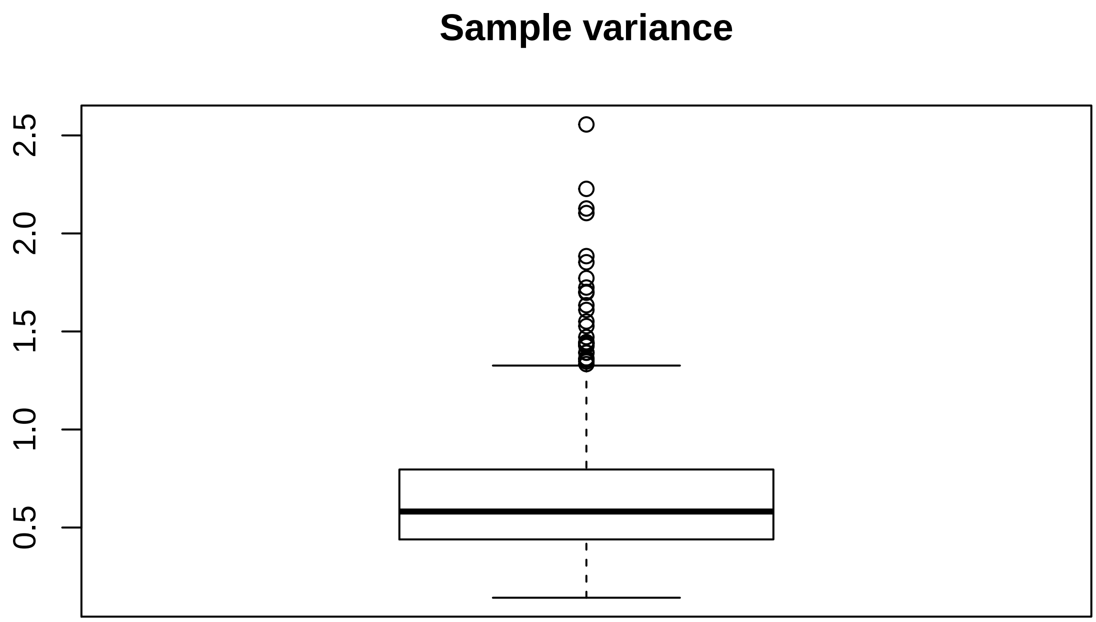

```r
summary(df)
```

```
##        x
##  Min.   :3.058
##  1st Qu.:4.390
##  Median :4.932
##  Mean   :4.976
##  3rd Qu.:5.491
##  Max.   :7.792
```

Sample Variance versus Theoretical Variance

The theoretical variance is given with 1/lambda^2. The sample variance is given by sigma^2/(n-1). As stated in the central limit theorem the variance should converge to the sample mean which is shown in the boxplot and summary statistics

```r
df <- data.frame(x=apply(data,2,function(x) sd(x)^2/(numberSample-1)))
boxplot(df$x,main="Sample variance")
```




```
summary(df)
```

```
##        x         
##  Min.   :0.1420  
##  1st Qu.:0.4394  
##  Median :0.5817  
##  Mean   :0.6446  
##  3rd Qu.:0.7955  
##  Max.   :2.5559
```

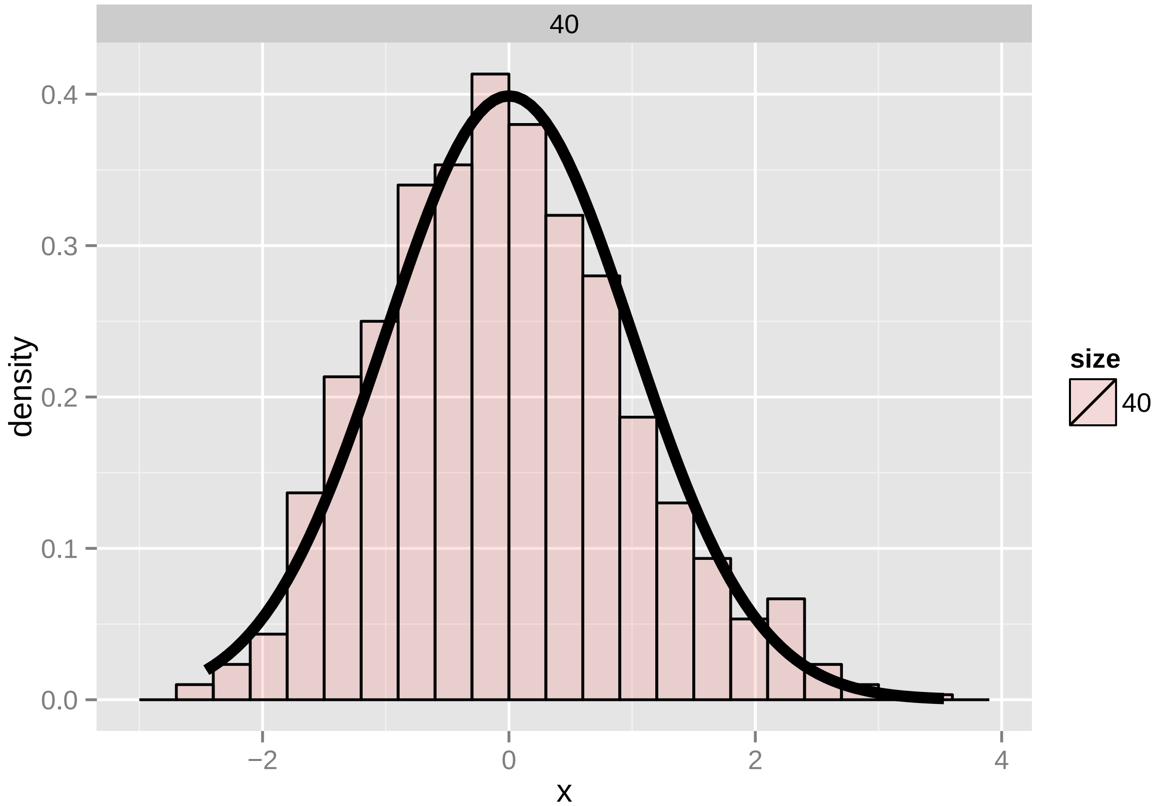
Distribution

Based on the central limit theorem we would expect that the sample means are distributed normally. The result is that

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\text{Estimate} - \text{Mean of estimate}}{\text{Std. Err. of estimate}}$$

has a distribution like that of a standard normal for large n. As shown in the diagram.

```
cfunc <- function(x, n) (sqrt(n) * (mean(x) - (1/lambda))) / (1/lambda)
dat <- data.frame(x = apply(data, 2, cfunc, 40),size = factor(rep(c(40), rep(numberSimulation, 1))))
g <- ggplot(dat, aes(x = x, fill = size)) + geom_histogram(alpha = .20, binwidth=.3, colour = "black", a
g <- g + stat_function(fun = dnorm, size = 2)
g + facet_grid(. ~ size)
```



Sources

Caffo, B. 2014. https://github.com/bcaffo/courses/blob/master/06_StatisticalInference/07_Asymptopia/index.Rmd