Statistical Inference - Course Project 1

Author: Danijel Bara

```
## Run time: 2015-08-22 20:13:04
## R version: R version 3.2.0 (2015-04-16)
```

Overview

This is the project for the statistical inference class. In it, we will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

1. A simulation exercise.
2. Basic inferential data analysis. Each pdf report should be no more than 3 pages with 3 pages of supporting appendix material if needed.

Instructions

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. We will illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. Tasks:
1. Show where the distribution is centered at and compare it to the theoretical center of the distribution. 2. Show how variable it is and compare it to the theoretical variance of the distribution. 3. Show that the distribution is approximately normal. Note that for point 3, we will focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

Load libraries

```r
library(knitr)
library(ggplot2)
```

Setting global variables

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. We will perform thousand simulations.

```r
set.seed(1234)
lambda <- 0.2
sample <- 40
nsimul <- 1000
```

Results

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

```
# Create a matrix of 1000 rows with the columns corresponding to random simulation 40 times
matrix.simulation <- matrix(rexp(nsimul * sample, rate=lambda), nsimul, sample)
mean.simulation <- rowMeans(matrix.simulation)
# Sample mean
sample.mean <-mean(mean.simulation)
sample.mean
```

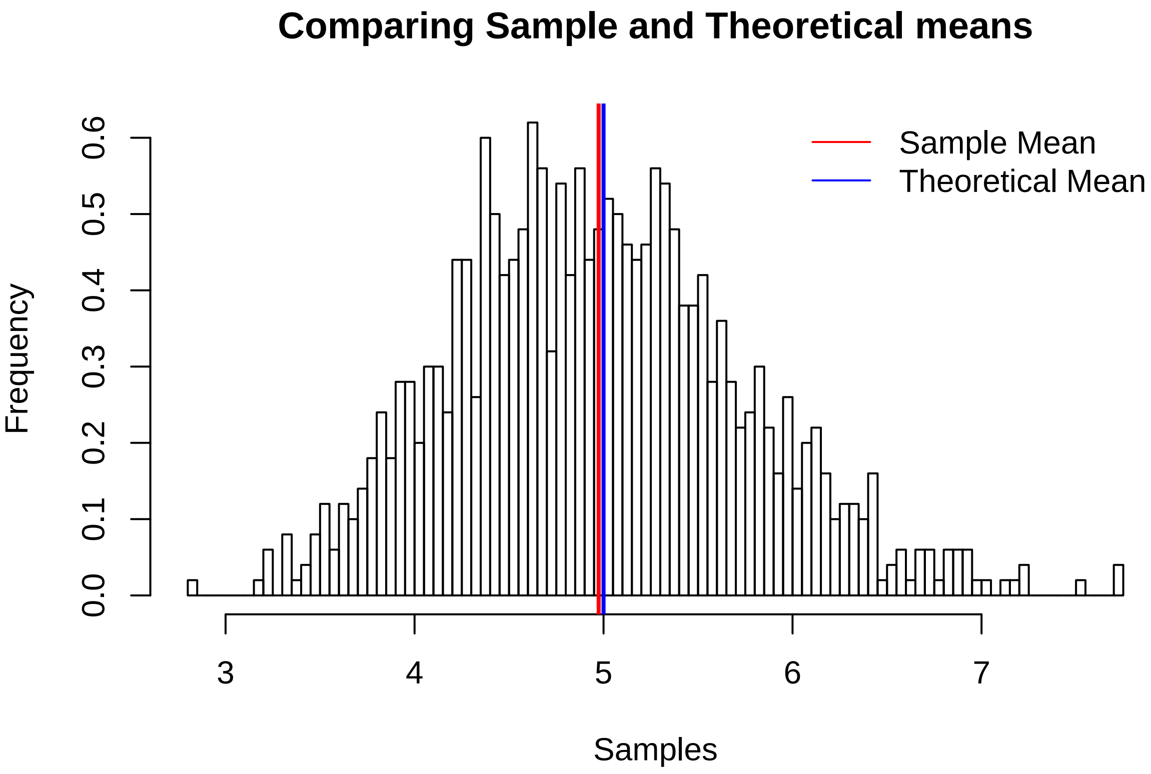
```
## [1] 4.974239
```

```
# Theoretical mean
theoretical.mean <- 1/lambda
theoretical.mean
```

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

If we compare sample mean, which is 4.974239 and theooretical mean, which is 5, we can conlude that the sample mean of the distribution is close to the center of the distribution. It also can be seen on the plot where red line presented sample mean and blue line presented theoretical mean.

```
hist(mean.simulation, breaks = 100, prob = TRUE,
     main="Comparing Sample and Theoretical means", xlab="Samples", ylab="Frequency")
     abline(v = theoretical.mean, col= "blue", lwd = 2)
     abline(v = sample.mean, col = "red", lwd = 2)
     legend('topright', c("Sample Mean", "Theoretical Mean"), bty = "n",
     lty = c(1,1), col = c(col = "red", col = "blue"))
```



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

We can calulate variance of the sample

```
variance.simulation <- var(mean.simulation)
variance.simulation
```

```
## [1] 0.5949702
```

The theoretical variance of the distribution is:

```
theoretical.variance <- ((1/lambda)/sqrt(40))^2
theoretical.variance
```

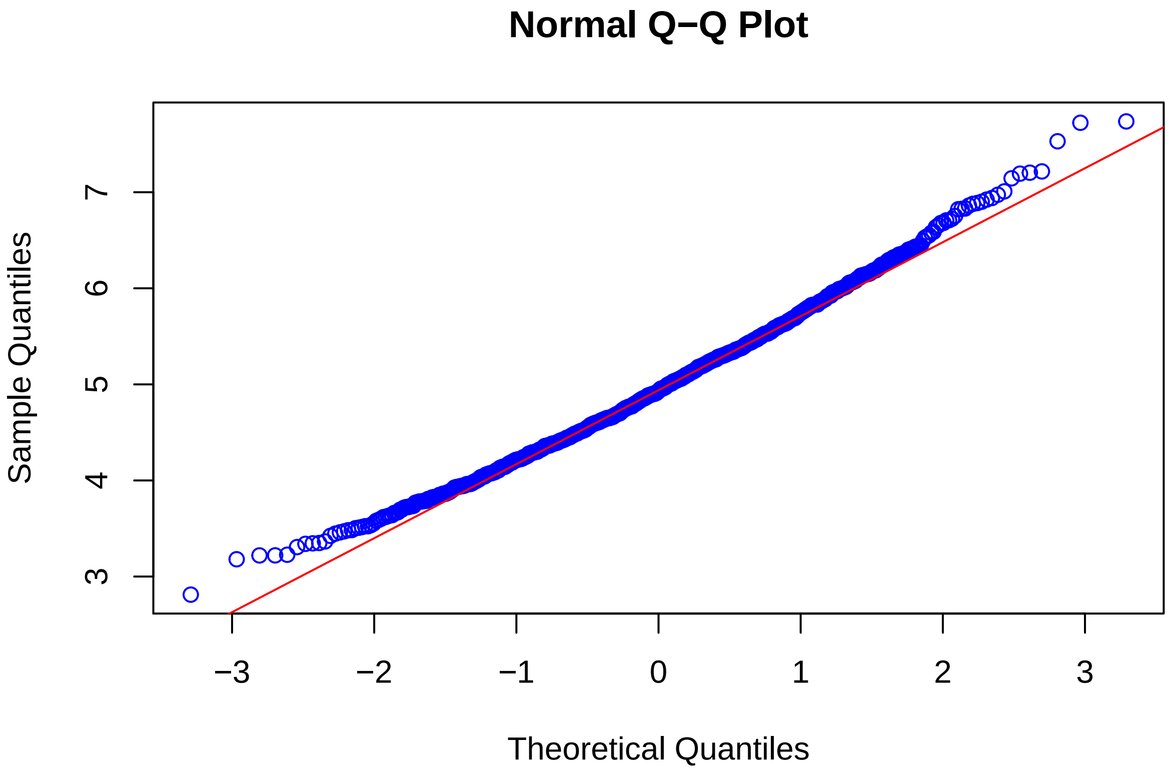
```
## [1] 0.625
```

The results of the theoretical variance of 1/lambdaˆ2, which is 0.625, as compared to actual variance, which is 0.5949072, are very close.

Question 3. Show that the distribution is approximately normal.

The theoretical quantiles also match closely with the actual quantiles. Thus we can conlcuded that the distribution is approximately normal.

```
qqnorm(mean.simulation, main ="Normal Q-Q Plot", col="blue")
qqline(mean.simulation, col = "red")
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles