A simulation study

Investigate distribution of averages of 40 exponentials(0.2)

In line with the Statistical Inference Course Project, the distribution of the mean of exponential random variables was investigated. For this the exponential distribution was simulated in R with the command `rexp(n, lambda)` where `n` is the number of generated random values and `lambda` is the rate parameter (λ). The mean μ and standard deviation σ of $Exp(\lambda)$-random variables is $1/\lambda$.

```
set.seed(288)
lambda <- 0.2
nSimulates <- 1000
averages <- NULL
for(i in 1:nSimulates){
  rx <- rexp(40, rate=lambda)
  averages <- c(averages, mean(rx))
}
```

The 1000 generated averages are described by the next statistics which are also reflected in the figure 1.

```
summary(averages)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.87    4.48    4.95    5.00    5.51    8.19
```

In the figure 1 the distribution of averages of 40 exponential(0.2)s is displayed.

```
plot(density(averages), main="Density of simulated averages", col='grey')
hist(averages, 20, freq=FALSE, add=T, border='blue')
abline(v=mean(averages), col=2, lwd=2)
legend("topright", "mean", col=2, lwd=2)
```

1. Empirical distribution and the theoretical center of the distribution

The mean of the simulated averages of 40 exponentials with rate 0.2 is near to the theoretical center of the exponential distribution, i.e. 1/0.2=5.

```
mean(averages)
```

```
## [1] 5.001
```

From the law of large numbers we know that the empirical averages $\bar{X}_n$ converge in probability to the true mean $\mu = 5$. The law of large numbers states that if $X_1, \ldots X_n$ are iid from a population with mean μ and variance σ^2 then $\bar{X}_n$ converges in probability to μ. In this setting the conditions are fulfilled since we simulate numbers independently from the same distribution Exp(0.2).

We can construct a 95% confidence interval and see that the theoretical mean is included:

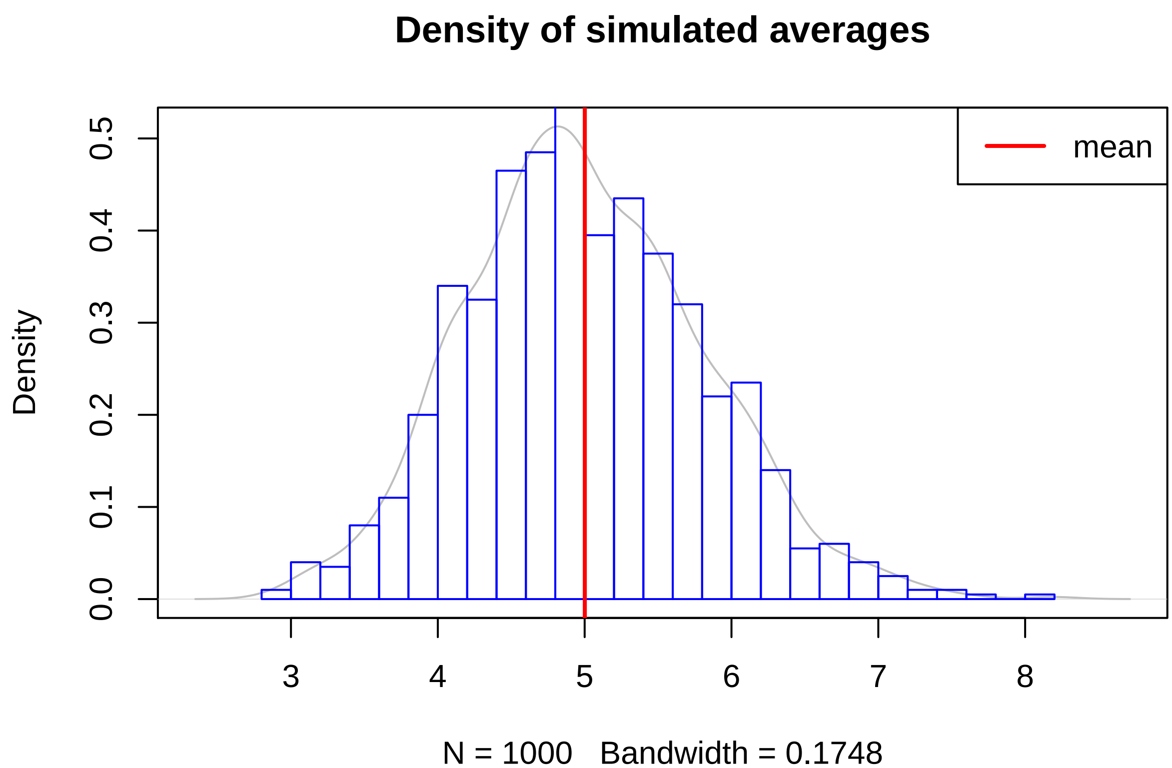

Figure 1: Distribution of 1000 simulated averages

```r
t.test(averages)$conf.int
```

```
## [1] 4.951 5.051
## attr(,"conf.level")
## [1] 0.95
```

2. Variability of estimation and the theoretical variance of the distribution

The theoretical distribution of averages should have a standard error of the form `standard deviation of exponential` devided by `square root of n`. In this case this should be:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{40}} = 0.7905.$$

In fact, the variability in the simulated data is very near to the theoretical value.

```r
sd(averages)
```

```
## [1] 0.8018
```

3. Check if the distribution is approximately normal.

From the first figure it can be seen that the distribution of the resulting averages looks relatively normal. Whether the distribution is approximately normal can be checked visually with `qqnorm` and with the Kolmogorov-Smirnov Test. This test has the null that the data comes from a normal distribution.

```r
qqnorm(averages)
qqline(averages, col=2)
```

```r
## Kolmogorov - Smirnov Test
ks.test(averages, "pnorm", 5, 5/sqrt(40))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  averages
## D = 0.0341, p-value = 0.1956
## alternative hypothesis: two-sided
```

In the KS-Test we could not reject the null. That does not mean that the simulated averages are truely normally distributed, but it is a good hint.

Using the **Central Limit Theorem** (CLT) we can assume that for large sample sizes n, $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ has a distribution like that of a standard normal.

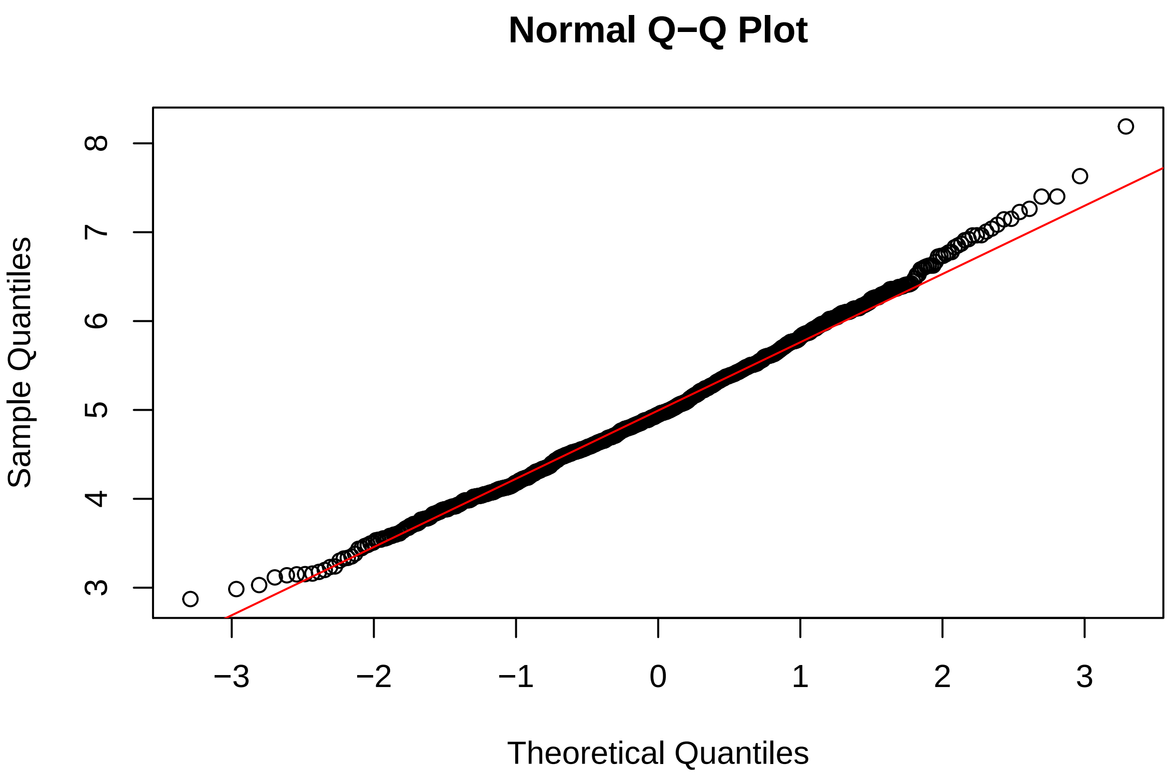

Figure 2: QQ-Plot