# Statistical Inference Project

*Andrej Schoeke*

*2015-Oct- 21*

## Statistical Inference Project - Part 1

The purpose of the project is to apply practically the knowledge in this course means and variances both in populations and sample distributions.

First, we define the basic parameters of the distributuin in question. To keep reproducible results, we set the seed of the random generator.

```
set.seed(599)

lambda <- 0.2
sample_size <- 40
num_sim <- 1000
```

Next, we run the simulation and save the result in a matrix:

```
sims <- matrix(nrow = 1000, ncol = 40, byrow = T)
for (i in 1:num_sim){
    sims[i,] <- rexp(sample_size, lambda)
}
sim_means <- rowMeans(sims)
```

The theoretical distribution has an estimated mean of $\frac{1}{\lambda} = \frac{1}{0.2} = 5$.

```
mean_of_distribution = 5
mean_of_sim_means <- mean(sim_means)

abs(mean_of_sim_means - mean_of_distribution)
```

```
## [1] 0.02073787
```

The theoretical distribution has an estimated variance of $\sigma = \frac{\frac{1}{\lambda}}{\sqrt{n}} = \frac{\frac{1}{0.2}}{\sqrt{40}} = 0.7905694$

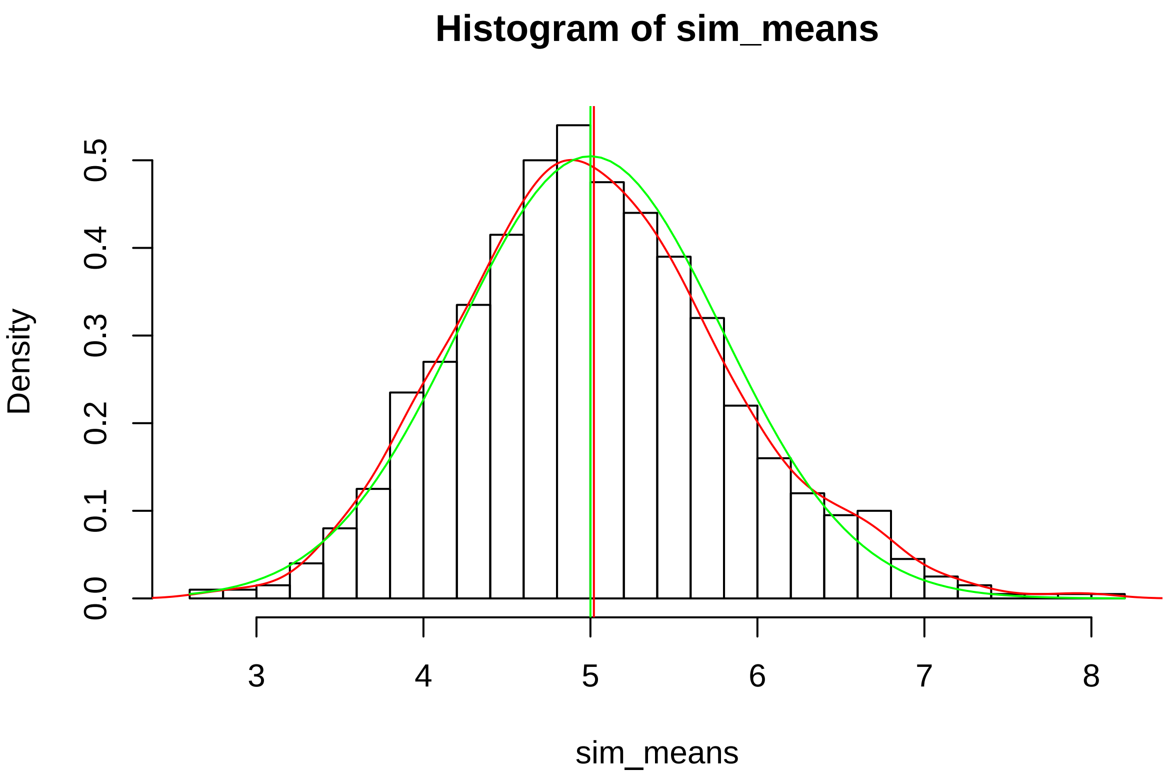
```
sd_of_distribution = 0.7905694
sd_of_sim_means <- sd(sim_means)

abs(sd_of_distribution - sd_of_sim_means)
```

```
## [1] 0.027416
```

Besides calculating the differences, let's have a look at the theoretical and simulated distribution by graphing it. We will have a look if the simulated sample distribution conforms to the nomal distribution.

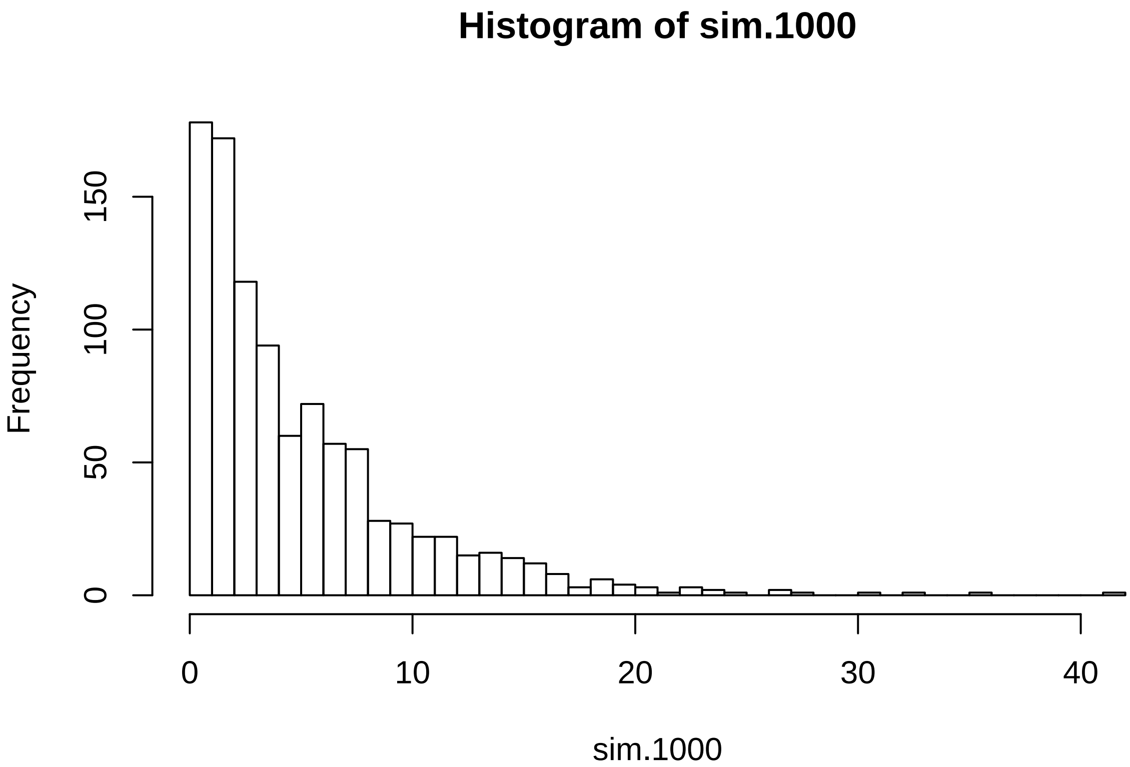
```r
hist(sim_means, breaks = 25, prob = T)
lines(density(sim_means), col = "red")
abline(v = mean_of_sim_means, col = "red")
abline(v = 1/lambda, col = "green")
curve(dnorm(x, mean = 5, sd = 0.7905694), add = T, col = "green")
```



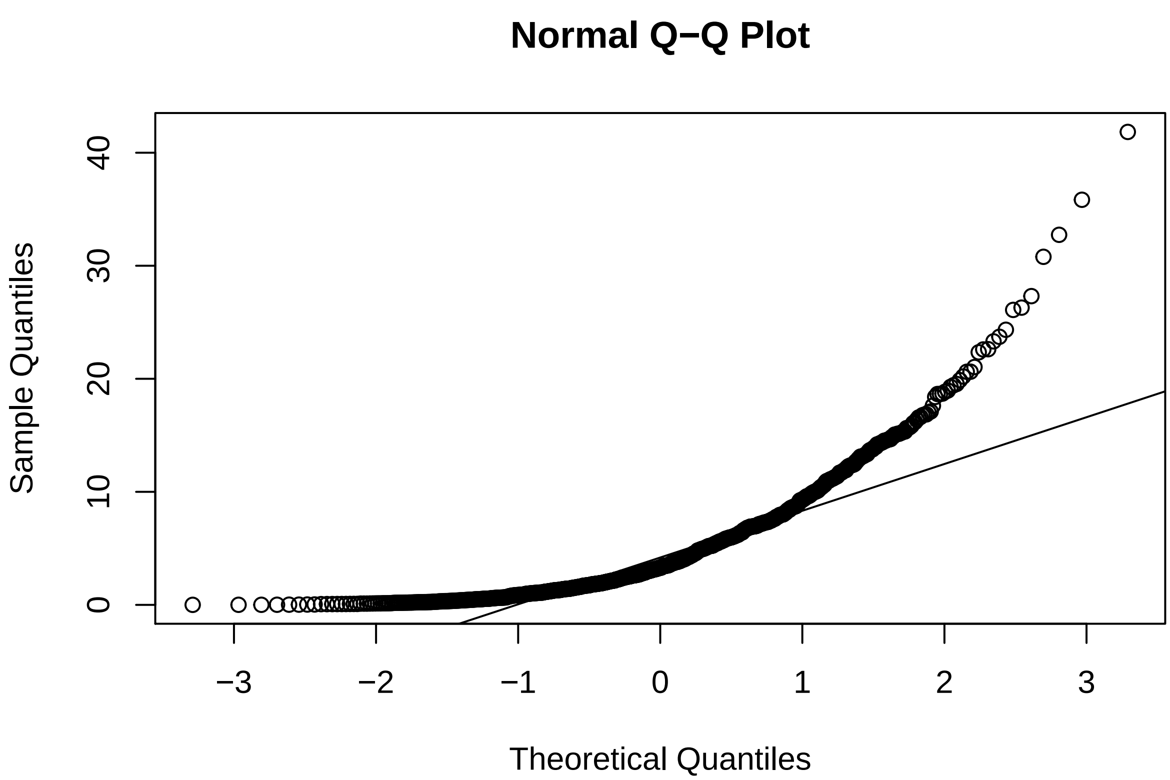

We can see that the simulated averages conform quite well to the theoretical normal distribution of the mean and variance.

If we compare it to 1000 simulated draws from the same distribution, we find a very different picture (or rather plot):

```r
sim.1000 <- rexp(1000, lambda)
hist(sim.1000, breaks = 50)
```




```r
qqnorm(sim.1000)
qqline(sim.1000)
```




Clearly, the histogram as well as the quantile plot indicated that this sample is not normal distributed. This is not surprising, as the Central Limit Theory states that the distribution of the sum or average of a large number of independent, identically distributed variables will be approximately normal. Not, that the distributions themselves become normal.