# Statistical Inference Course Project - Part 1

*Artem Braun*

**Statistical Inference Course Project - Part 1**

**1. Overview**

**In this project we will investigate the exponential distribution in R and compare 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.**

**The following constants were provided in the task: lambda = 0.2, number of simulations = 1000, investigate the distribution of averages of 40 exponentials.**

**2. Simulations**

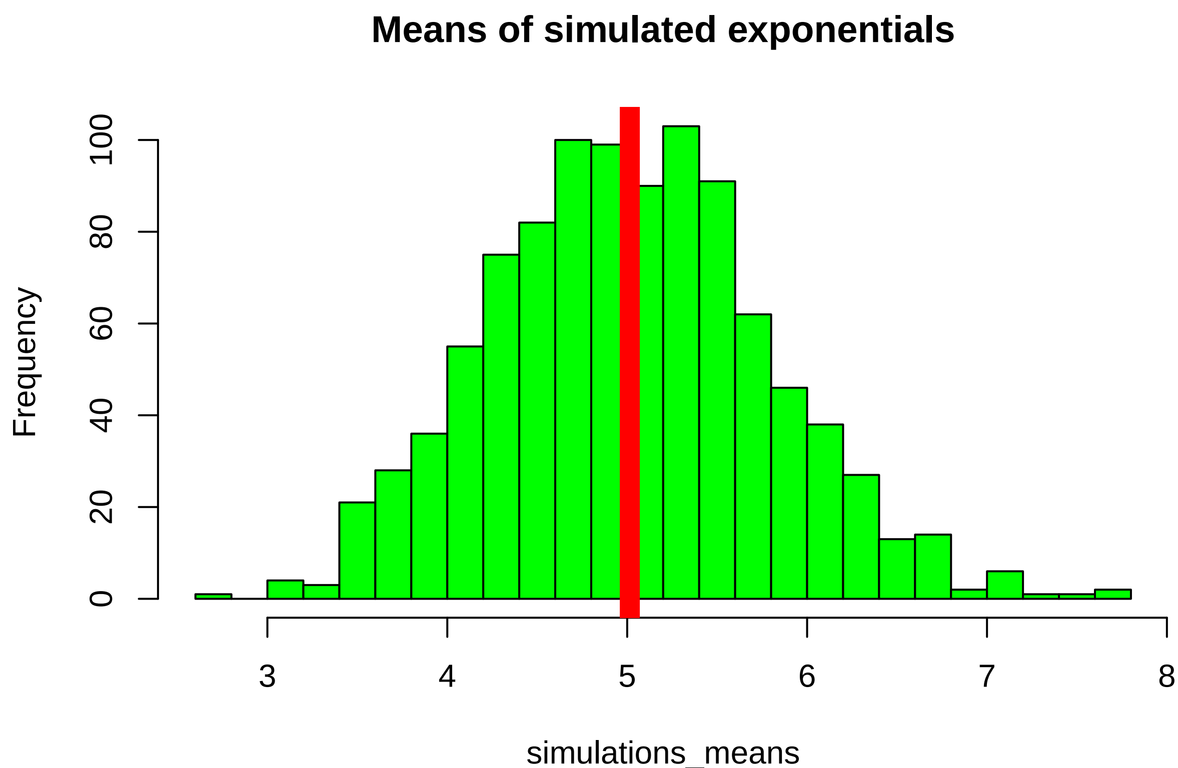
**In the following code we define constants provided in the task. Then we simulate exponential distributions 1000 times with 'replicate' function and create a data frame with 1000 columns and 40 rows (number of exponentials in each simulation). Then we make a vector with 1000 means of exponentials.**

```
set.seed(3000)

num_exp <- 40
lambda <- 0.2
num_sim <- 1000

exp_simulations <- replicate(num_sim, rexp(num_exp, lambda))
simulations_means <- apply(exp_simulations, 2, mean)
```

**3. Sample Mean versus Theoretical Mean**

**According to the task, the mean of the exponential distribution is 1/lambda. Since lambda = 0.2., theoretical mean should equal 5 (= 1 / 0.2). The following code plots a histogram of the means of simulated exponentials and defines a mean line. Calculated mean is appeared to be 5.01, which is very close to theoretical mean 5.**

```
hist(simulations_means, breaks=30, main="Means of simulated exponentials", col = "green")
abline(v = mean(simulations_means), lwd="10", col="red")
```




```
mean(simulations_means)
```

```
## [1] 5.014507
```

**4. Sample Variance versus Theoretical Variance**

**Theoretical standard deviation for given constants is as follows:**

```
print((1/lambda)/sqrt(num_exp))
```

```
## [1] 0.7905694
```

**Simulated standard deviation is appeared to be**

```r
print(sd(simulations_means))
```

```
## [1] 0.7756039
```

**which is quite close to the theoretical one!**

**Theoretical variance for given constants is as follows:**

```r
print(((1/lambda)/sqrt(num_exp))^2)
```

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

**Simulated variance is appeared to be**

```r
print(sd(simulations_means)^2)
```

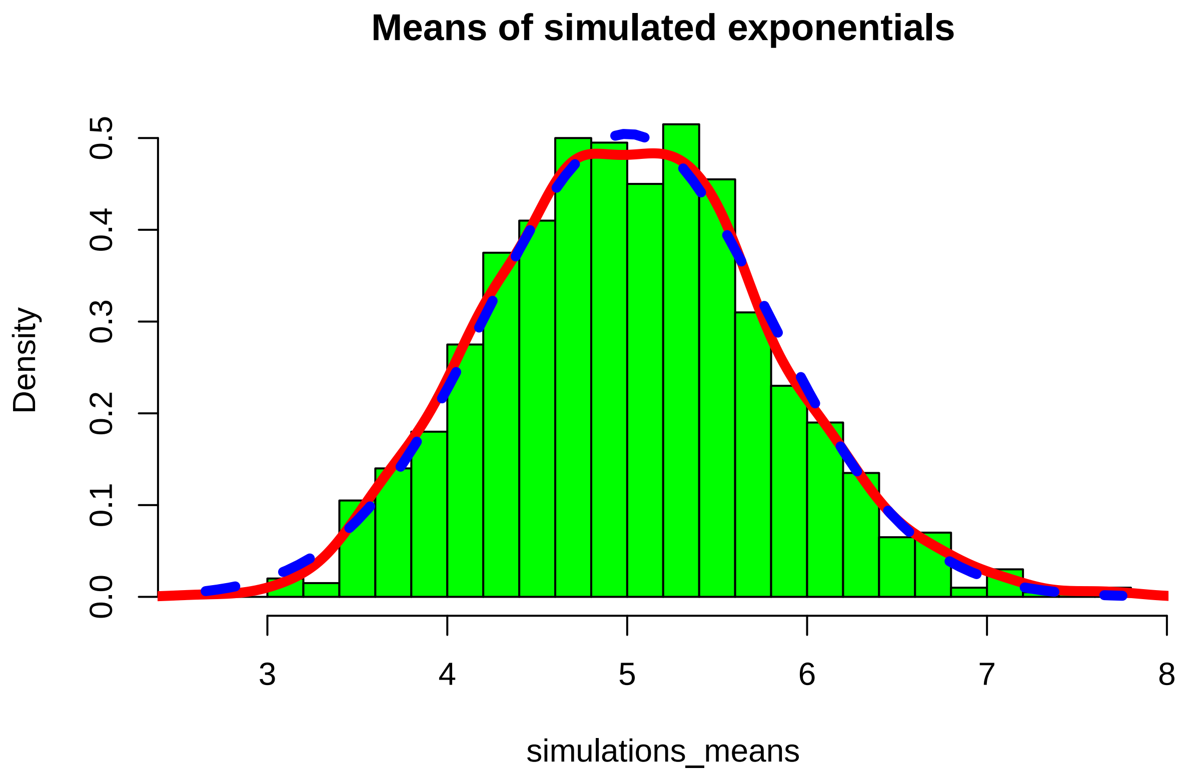
```
## [1] 0.6015614
```

**which is quite close to the theoretical one as well!**

**5. Distribution**

**In this section we should investigate the exponential distribution in regards to its "normality". According to Central Limit Theorem, the distribution of averages of iid variables becomes that of a standard normal as the sample size increases even if the original variables themselves are not normally distributed. Since we have significant number of exponentials, we can expect that their means follow normal distribution. Let us test it graphically:**

```r
hist(simulations_means, prob=TRUE, col="green", main="Means of simulated exponentials", breaks=30)
lines(density(simulations_means), lwd=5, col="red")
x <- seq(min(simulations_means), max(simulations_means), length=2*num_exp)
y <- dnorm(x, mean = 1/lambda, sd = sqrt(((1/lambda)/sqrt(num_exp))^2))
lines(x, y, pch=22, col="blue", lwd=5, lty = 2)
```




**We added a smoothing red line. As we can see, this is almost perfect bell-shaped curve. Additionally, to reaffirm our inference, we added real normal distribution line, which is dashed and blue. Obviously, these lines are close enough to infere that our simulated distrubution is almost normal.**