

# Statistical Inference course project

*Dmitri Perov*

*December 26, 2015*

## Overview

In this project you 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  $1/\lambda$ . Set `lambda = 0.2` for all of the simulations. You will investigate the distribution of averages of 40 exponentials.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

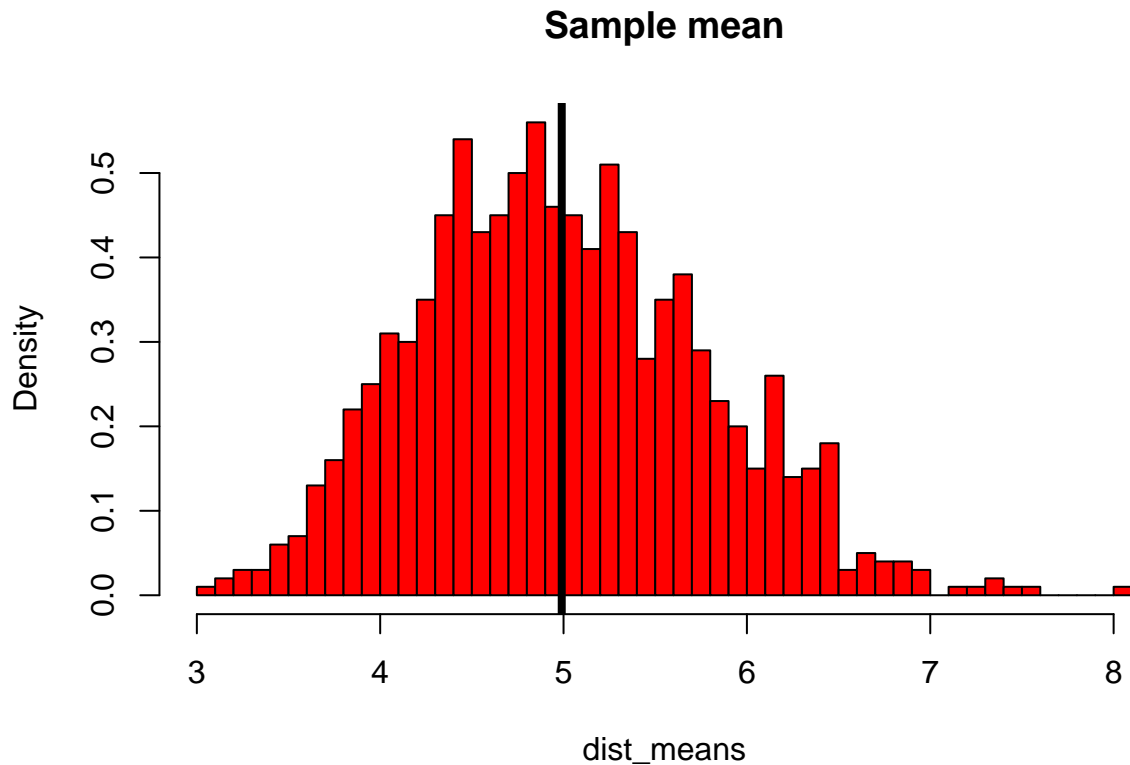
## Simulations

Get the means of the 1000 simulations

```
set.seed(1)
n <- 1000
sample_size <- 40
lambda <- 0.2
dat <- matrix(rexp(sample_size * n, rate=lambda), ncol = sample_size, nrow = n)
dist_means <- rowMeans(dat)
```

## Sample Mean versus Theoretical Mean

```
hist(dist_means, breaks=sample_size, prob=T, col="red", main="Sample mean", ylab="Density")
abline(v=mean(dist_means), col="black", lwd=4)
```



```
center_actual <- mean(dist_means)
center_theor <- 1/lambda
```

The theoretical center of the distribution is 5.

The center of the mean distribution is **4.9900252**.

## Sample Variance versus Theoretical Variance

```
sd_actual <- sd(dist_means)
sd_theor <- (1/lambda)*(1/sqrt(sample_size))

var_actual <- sd_actual^2
var_theor <- ((1/lambda)*(1/sqrt(sample_size)))^2
```

Standard Deviation of the distribution of means - **0.7859435**.

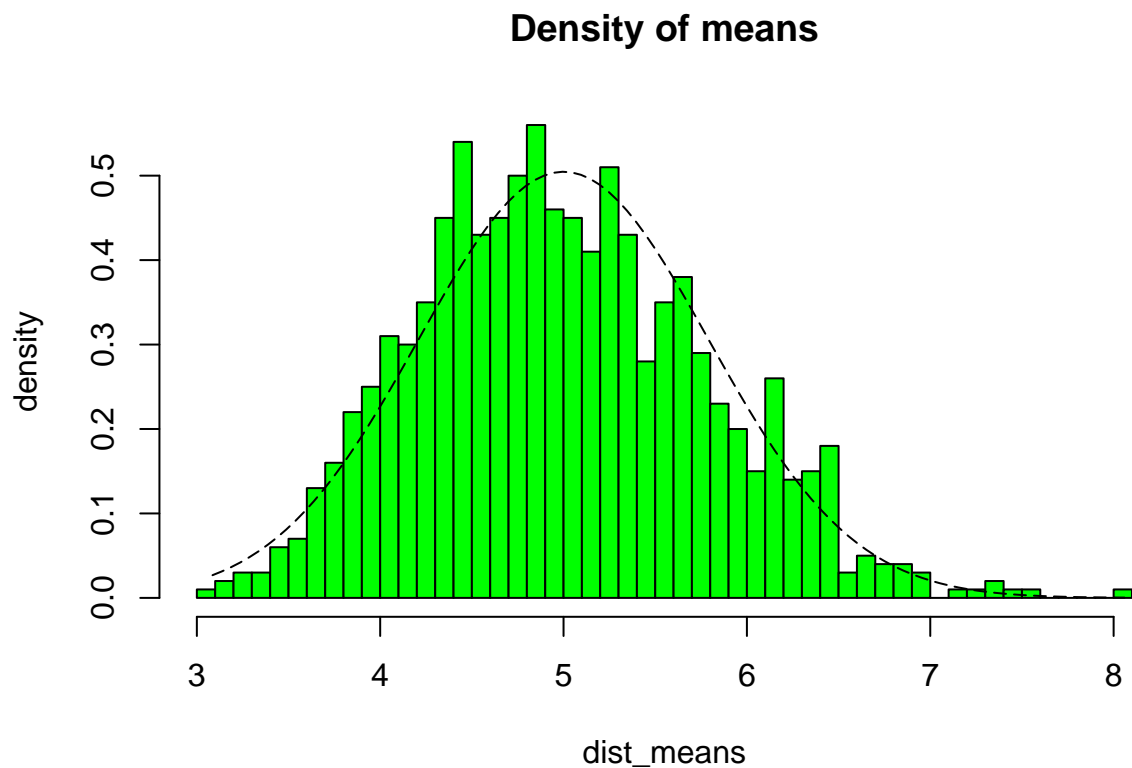
Theoretical Standard Deviation - **0.7905694**.

The Theoretical variance **0.625**.

Actual variance of the distribution is **0.6177072**.

## Distribution

```
x <- seq(min(dist_means), max(dist_means), length=100)
y <- dnorm(x, center_theor, sd = sd_theor)
hist(dist_means, breaks = sample_size, prob = T, col = "green", main = "Density of means", ylab = "dens")
lines(x, y, pch = 22, col = "black", lty = 5)
```



Plot shows that distribution of mean (green) is very close to the normal distribution (black line)