Statistical Inference Course project Part 1.

Aleksandr Beloushkin (aka gudun-ku)

October 22, 2015.

The gist

Our goal is to Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. To do this we need to:

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. (that difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials is almost irrelevant)

Necessary R packages (ggplot2)

```
library(ggplot2)
```

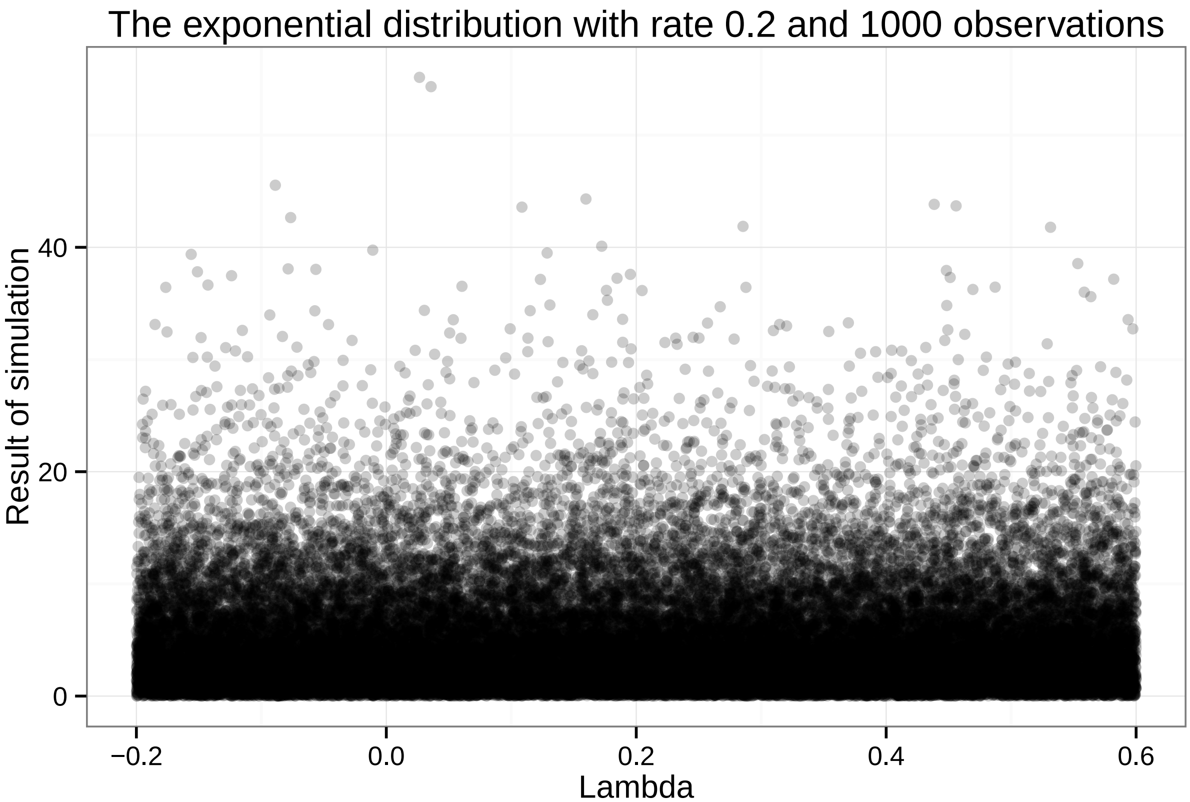
Simulation stage

Let's simulate a distribution of 1000 averages of 40 random exponentials and make dataset for further

```
set.seed(137)
lambda = 0.2
n_draws = 40  # no. of draws of random exponentials in each simulation
n_sim = 1000  # no. of simulations
# Function to generate 40 exponentials during each iterations.
cfunc <- function(sim_no) {
    sim_result <- rexp(n_draws, lambda)
    return(data.frame(simno = sim_no, simResult = sim_result, simMean =
                        mean(sim_result)))
}
df = do.call(rbind,lapply(1:1000, cfunc))
```

The overall picture of the our exponential distribution example:

```
qplot(lambda,simResult,data=df, geom="jitter", alpha=I(.2),
main="The exponential distribution with rate 0.2 and 1000 observations",
xlab="Lambda",
ylab="Result of simulation")+
theme_bw()
```



Sample mean and theoretical mean

Calculating and comparing sample Mean against theoretical mean

```
theoretical_mean = 1/lambda
sample_mean = mean(df$simMean)
```

- **Theoretical mean: 5**
- **Sample mean: 4.9776028**

Sample variance and theoretical variance

Calculating and comparing sample Variance against theoretical Variance

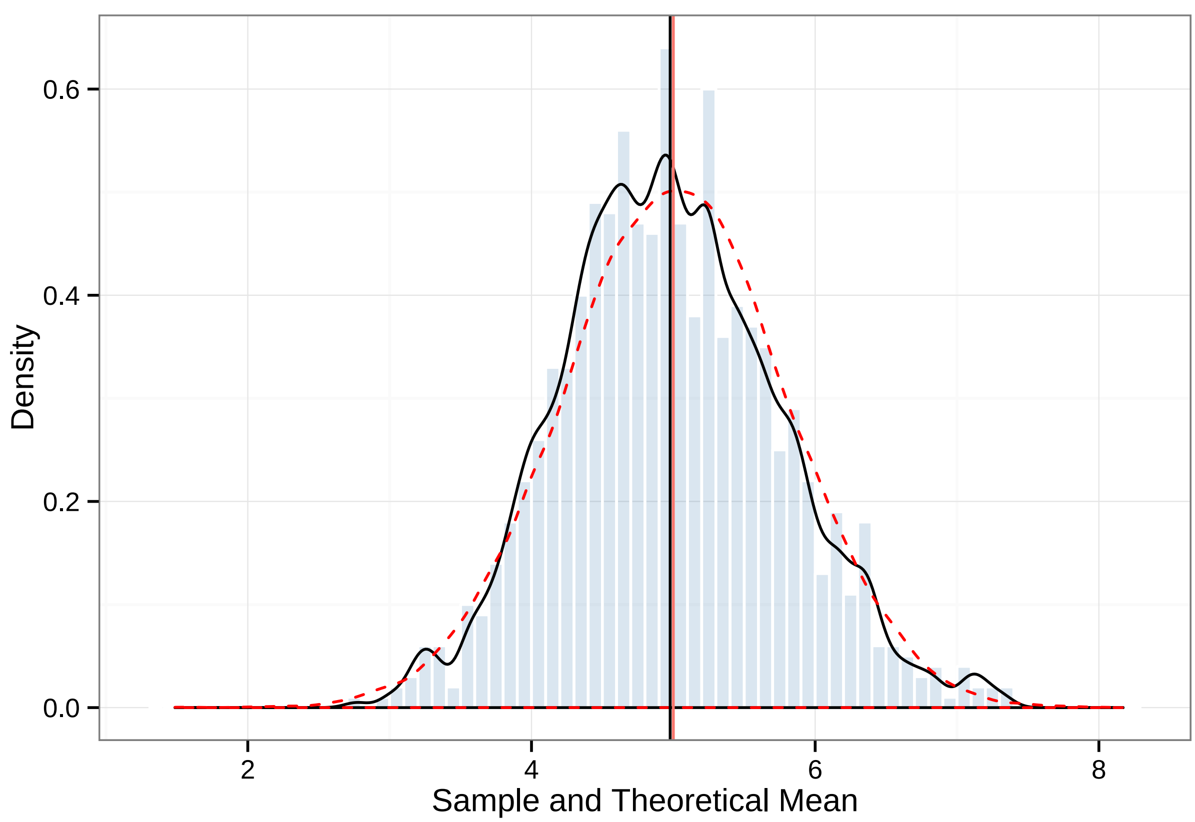
```
theoretical_variance = ((1/lambda)/sqrt(n_draws))^2
sample_variance = var(df$simMean)
```

- **Theoretical variance: 0.625**
- **Sample variance: 0.5913138**

Distribution properties

Given lambda be set equal 0.2, we can notice that the distribution of sample means of 40 exponentials, 4.9871567, is very similar to but not exactly the same as the theoretical mean which equals to 5. Now our task is to show that the distribution is approximately normal

```
g <- ggplot(df, aes(simMean))
g <- g +  geom_histogram(alpha = .20, binwidth=.1, aes(y=..density..),
                         color='white',fill="steelblue")
g <- g + geom_density()
g <- g + geom_density(inherit.aes=F,
                        aes(rnorm(n=1000 * n_draws,
                                   mean=theoretical_mean,
                                   sd = sqrt(theoretical_variance))),
                        linetype='dashed',col = "red")
g <- g + geom_vline(aes(xintercept=sample_mean))
g <- g + geom_vline(aes(xintercept=theoretical_mean,col = "red"))
g <- g + theme_bw()
g <- g + xlab("Sample and Theoretical Mean")
g <- g + ylab("Density")
g
```



Densities of the two distributions different slightly (normal distribution represented by dashed red line). This showing us that the distribution of means of an exponential distribution follows that of a normal distribution and tends to be closer to it with growing number of simulations. This behavior visualises proof of the Central Limit Theorem.