Statistical Inference Course Project part 1

Anton Titov

5-April-2016

Syllabus

This report will cover the analysis of exponential distribution. We will compare parameters of sample distribution and theoretical one. ## Loading and preprocessing the data We are loading `data.table` for fast data handling and `ggplot2` . Also we change system time display to English by `Sys.setlocale` .

```
Sys.setlocale("LC_TIME", "English")
library(data.table)
library(ggplot2)
set.seed(1234521)
```

Playing with exponential distribution

We set parameters for exponential distributions. Then we generate vector of means for 50 samples. After that we calculate sample mean and sample standart deviation.

```
smean <- sexpo <- NULL
lambda <- 0.2
nexp <- 50 #number of experiments
n <- 1000 #number of simulations
for (i in 1:n)
  {
  expounit <- rexp(nexp, lambda)
  sexpo <- c(sexpo, expounit)
  smean <- c(smean, mean(expounit))

}
simumean <- mean(smean)
simusd <- sd(smean)
```

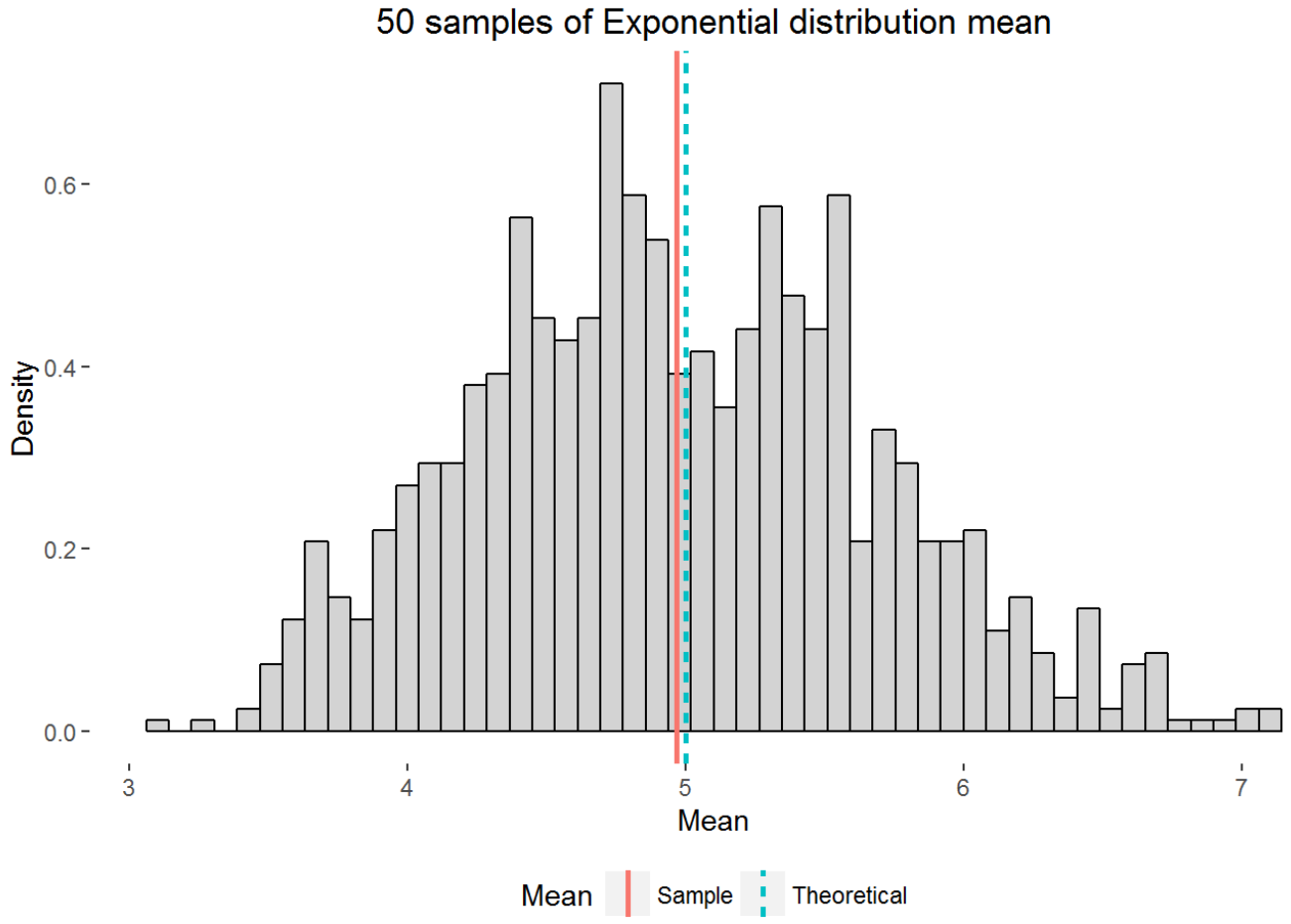
We create data.frame of sample and theoretical means for easy plotting

```
bothmeans <- data.frame(Mean=c("Sample","Theoretical"), vals = c(simumean,1/lambda))
```

Question 1.

Show the sample mean and compare it to the theoretical mean of the distribution

```
g = ggplot(data.frame(smean = smean), aes(smean))
g = g + geom_histogram(aes(y = ..density..), fill = "lightgrey", colour = "black", bins
=50)
g = g +  geom_vline(data = bothmeans, aes(xintercept = vals, color = Mean, linetype = M
ean), size = 1)
g = g +  theme(panel.background = element_rect(fill = 'white'), legend.position = "bott
om")
g = g + labs(x="Mean", y="Density", title="50 samples of Exponential distribution mea
n")
print(g)
```



As we can see sample mean is pretty close to theoretical.
Sample mean: `simumean` = 4.9659
Theoretical mean: `1/lambda` 5
Sample SD of the mean: `simusd` 0.7197

Question 2.

Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

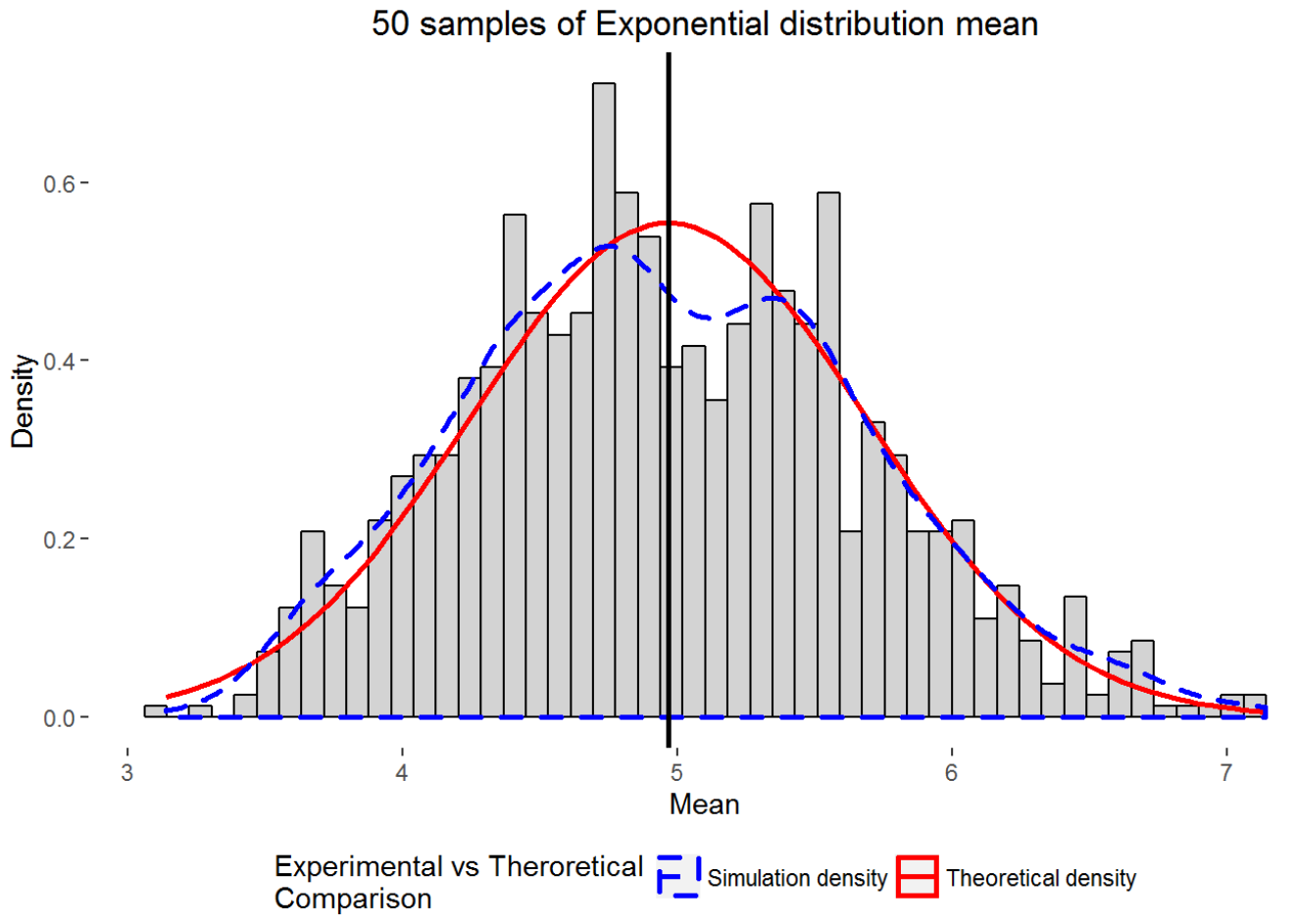
As we can see sample variance is pretty close to theoretical.
Sample varinace: 0.518
Theoretical variance: 0.5

Sample SD of the mean: 0.7197
Theoretical SD of the mean 0.7071

Question 3.

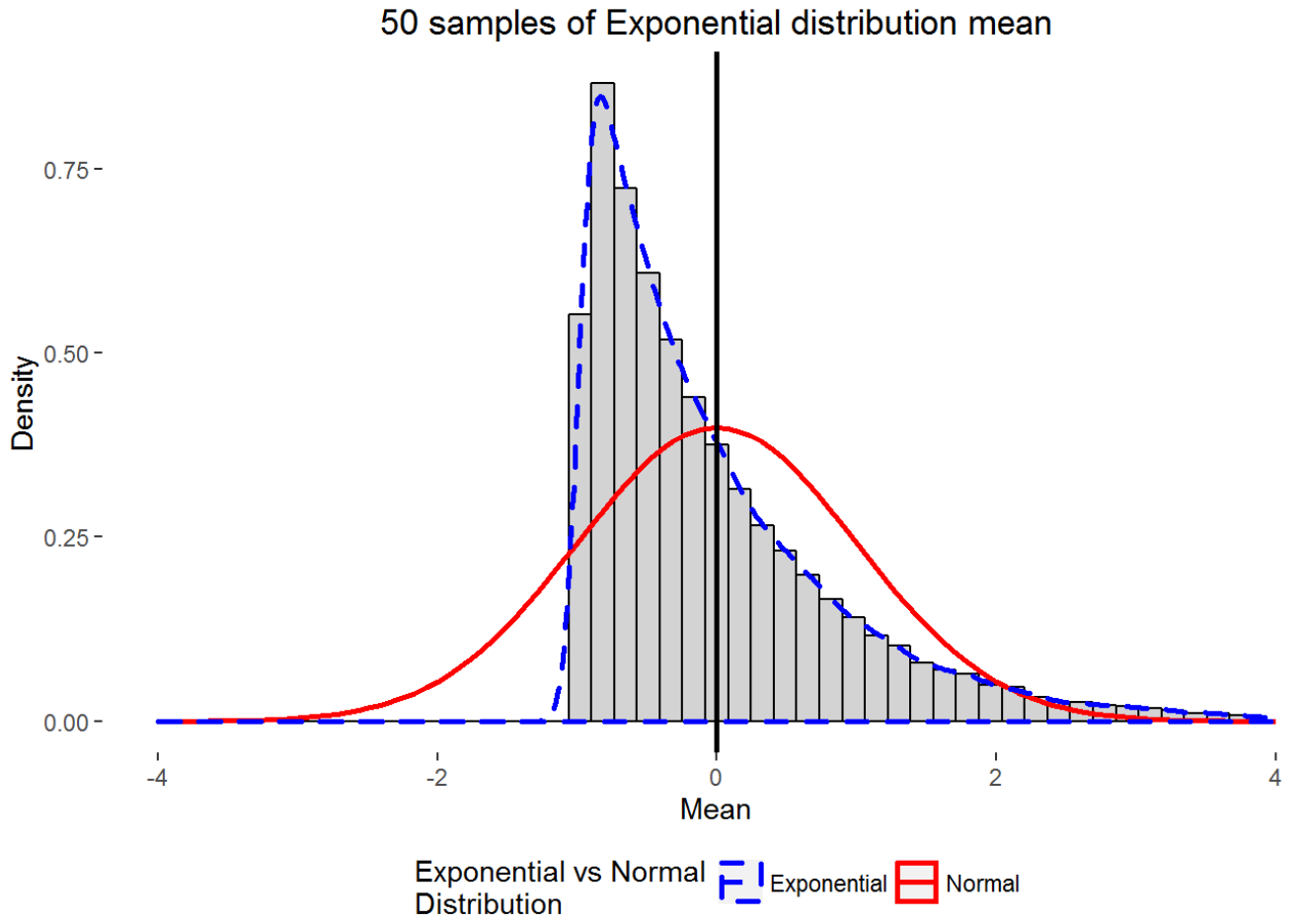
Show that the distribution is approximately normal.

```
g = ggplot(data.frame(smean = smean), aes(smean))
g = g + geom_histogram(aes(y = ..density..), fill = "lightgrey", colour = "black", bins
=50) +
      stat_function(fun = dnorm, size = 1, args = list(mean = simumean, sd = simusd), a
es(color="Theoretical density") ) +
      geom_density( size=1, aes(color="Simulation density"), linetype = 2) +
      scale_color_manual(values=c("Simulation density" = "blue", "Theoretical density"
= "red"),
                         name="Experimental vs Theroretical\nComparison") +
      guides(colour = guide_legend(override.aes = list(linetype=c(2,1))),
             fill = guide_legend(override.aes = list(colour = NULL)))

g = g + geom_vline(xintercept = simumean, size = 1)
g = g +  theme(panel.background = element_rect(fill = 'white'), legend.position = "bott
om")
g = g + labs(x="Mean", y="Density", title="50 samples of Exponential distribution mea
n")
print(g)
```



Distribution of the sample mean looks like normal. If we increase nubmer of experiments then it will look more normal.

```
g = ggplot(data.frame(sexpo = scale(sexpo)), aes(sexpo))
g = g + geom_histogram(aes(y = ..density..), fill = "lightgrey", colour = "black", bins
=50) +
      stat_function(fun = dnorm, size = 1, args = list(mean = 0, sd = 1), aes(color="No
rmal") ) +
      geom_density( size=1, aes(color="Exponential"), linetype = 2) +
      scale_color_manual(values=c("Exponential" = "blue", "Normal" = "red"),
                         name="Exponential vs Normal\nDistribution") +
      guides(colour = guide_legend(override.aes = list(linetype=c(2,1))),
             fill = guide_legend(override.aes = list(colour = NULL)))

g = g + geom_vline(xintercept = 0, size = 1)
g = g +  theme(panel.background = element_rect(fill = 'white'), legend.position = "bott
om")
g = g + labs(x="Mean", y="Density", title="50 samples of Exponential distribution mea
n") +xlim(-4,4)
print(g)
```



Sample distribution does not look like normal at all. That means we succesfully simulated exponential distribuion. Increasing number of experiments we will see it becoming less normal.