EXPONENTIAL DISTRIBUTION

Oleg Tsarev

Overview

In this report we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate. It is a particular case of gamma distribution (from Wikipedia: https://en.wikipedia.org/?title=Exponential_distribution).

We aim to illustrate via simulation and associated explanatory text the properties of the distribution. Our report based on the following subtasks:
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

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. In this project we set lambda equals 0.2 for all of the simulations and we will investigate the distribution of averages of 40 exponentials with a thousand simulations.
So, let's investigate the exponential distribution in R.

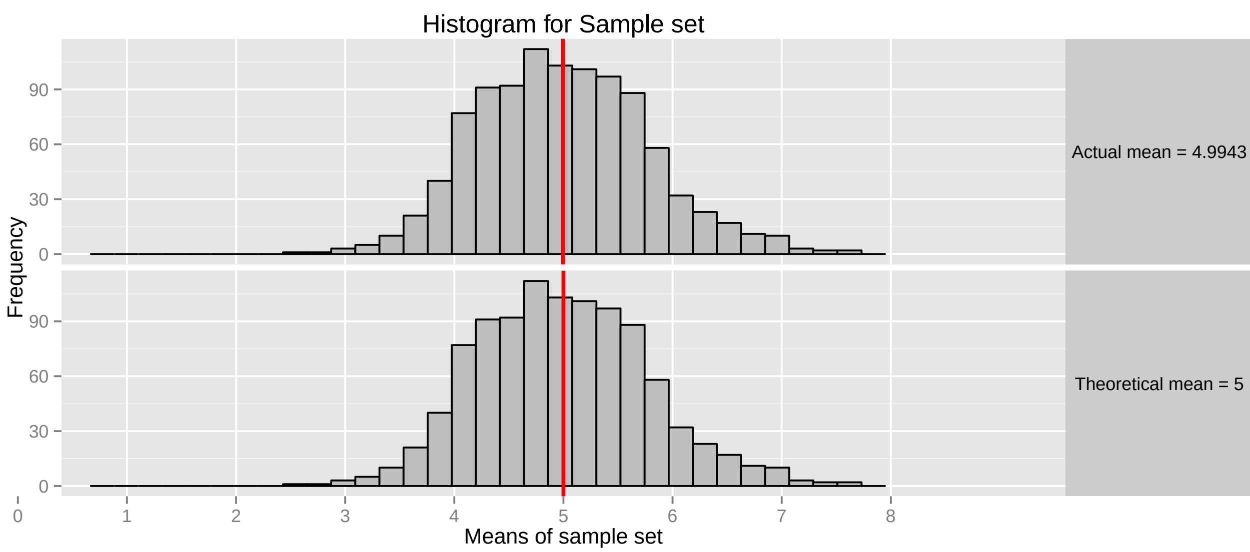
```
set.seed(5375)       # set state for random number generation for reproducibility
lambda <- 0.2        # set lambda
n <- 40              # set number of variables
nosim <- 1000        # set number of simulations
# data frame with means of 1000 simulations of exponential distributions
ed_df <- data.frame(x = replicate(nosim, mean(rexp(n = n, rate = lambda))))
ed_df$m <- round(mean(ed_df$x),4)    # calculate sample mean
ed_mean <- round(mean(ed_df$x),4)    # save sample mean in variable
ed_var <- round(var(ed_df$x),4)      # save sample variance in variable
```

Sample Mean versus Theoretical Mean

The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Let's plot mean of our sample and compare it with the theoretical mean that equals 1/lambda = 5.

```
library(ggplot2)                    # load graph package
tmp <- ed_df                        # make copy in order to use facets mode
tmp$m <- 5                          # initialize second half of data frame by theoretical mean
ed_df_mean <- rbind(ed_df,tmp)      # merging original and temp data frame
# create function for control of facets labels
f_labeller <- function(var, value){
    value <- as.character(value)
    if (var == "m") {
        value[value == "5"] <- "Theoretical mean = 5"
        value[value == ed_mean]   <- paste("Actual mean =",ed_mean)
```

```r
    }
    return(value)
}
# create plot with name and labels
q <- qplot(ed_df_mean$x,data = ed_df_mean,geom = "histogram",main = "Histogram for Sample set",xlab = "M
      ,col = I("black"),fill = I("grey"),ylab = "Frequency")
# add facet mode and change default labels
q <- q + facet_grid(m ~ .,labeller = f_labeller) + theme(strip.text.y = element_text(size = 10, angle = 
# change labels of x axis
q <- q + scale_x_discrete(labels = 0:8)
# add lines for actual and theoretical means
q + geom_vline(aes(xintercept = ed_df_mean$m),lwd = 1,col = "red",show_guide = T)
```



As shown above actual mean 4.9943 (upper plot) is very close to the theoretical mean 5 (lower plot).

Sample Variance versus Theoretical Variance

The standard deviation of exponential distribution is also 1/lambda.
So, theoretical variance equals

$$\frac{(1/\lambda)^2}{n}$$

```r
th_var <- (1/lambda)^2/n     # calculate theoretical variance
th_var                       # print theoretical variance
```

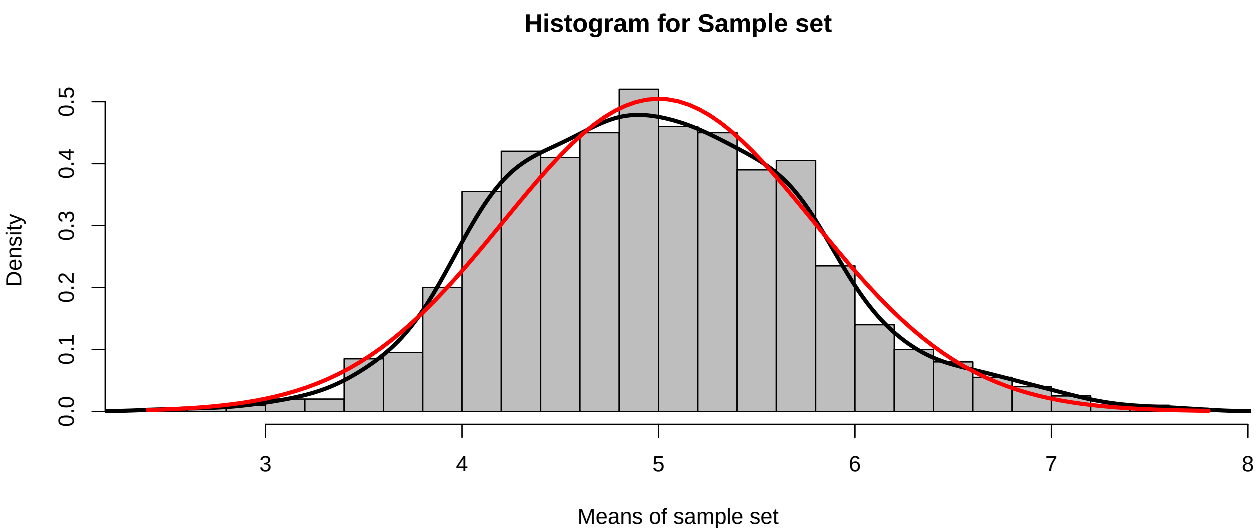
```
## [1] 0.625
```

So, actual variance equals 0.6087 which is close to the theoretical mean 0.625.

Distribution

Now let's have a look again on our histogram.
Let's add on the plot actual sample distribution curve and normal distribution curve with known mean and standard deviation.

```
# create histogram with name and labels
hist(ed_df$x,probability = T,main = "Histogram for Sample set",xlab = "Means of sample set",col = "grey
# add distribution curve of actual sample
lines(density(ed_df$x),col="black",lwd=3)
# add distribution curve of normal distribution with known mean and standard deviation
curve(dnorm(x,mean = 1/lambda,sd = (1/lambda)/sqrt(n)),add = T,col = "red",lwd = 3)
```



As you can see, plot above shows that distribution of our sample set (black curve) is symmetric around the mean with a bell shape and can be adeqautely approximated with the normal distribution (red curve). Therefore we can say that **distribution of our sample set can be defined as a normal distribution**.