Report - Simulation of Exponential Distribution

Hauer-Glocke

12 Juli 2017

Overview of report

I conducted the following steps, which I will elaborate in the following:

1. Create the dataset and the standard setting for the subsequent analysis
2. Discuss the sample mean
3. Discuss the sample variance
4. Discuss the sample distributions

The analysis further focuses on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

Step 1 - Standard setting and creating of data set

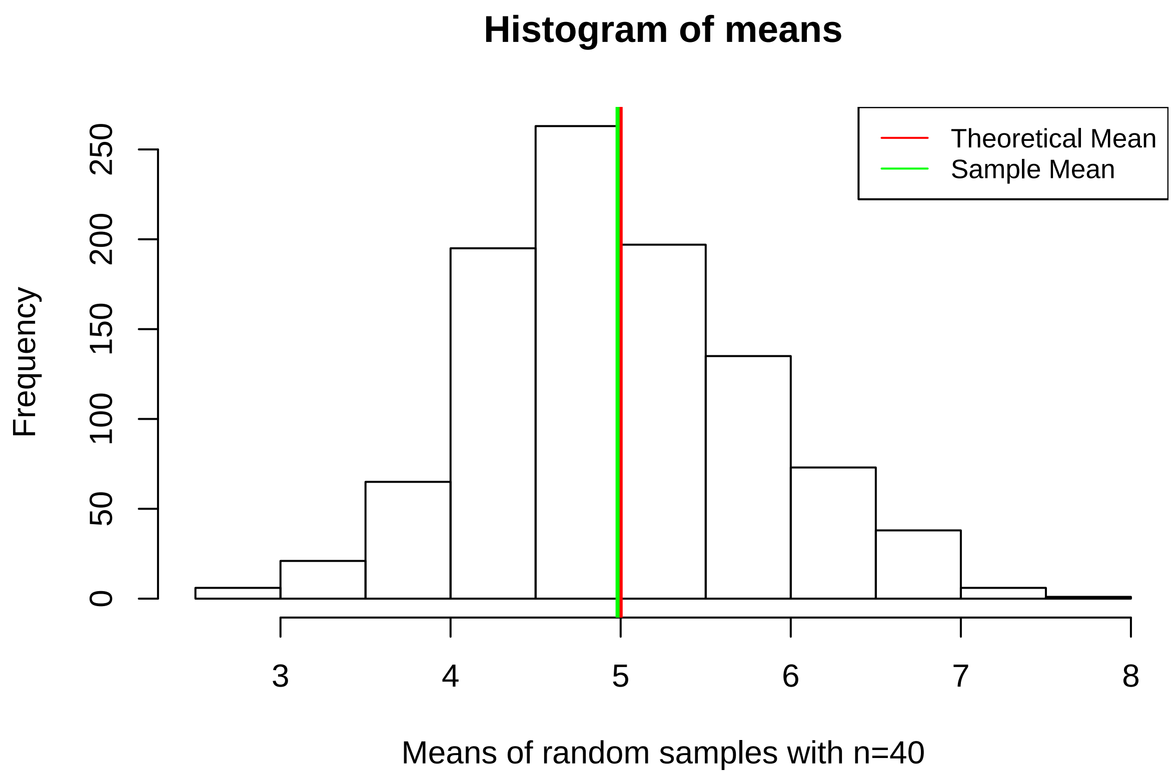
```
#Standard Setting
lambda <- 0.2
n <- 40
t <- 1000

#Means of 1,000 repititions of a random samples of n=40 exponentially distributed
means=NULL
for (i in 1:t){
        means[i] <- mean(rexp(n, lambda))
}

#Standard Deviations of 1,000 repititions of a random samples of n=40 exponentially
#distributed
sds=NULL
for (i in 1:t){
        sds[i] <- sd(rexp(n, lambda))
}

#General Values of the Means-dataset
theoreticalMean <- (1/lambda)
theoreticalSD <-(1/lambda)/ sqrt((40-1))

means_avrg <- mean(means)
means_sd <- sd(means)
```

Step 2 - Discussion of the sample mean

The histogramm shows the distribution of means over repeatedly generated samples with n=40 for 1,000 times. The expected mean (red line) and the mean of the sample shows that there are only marginal differences.

The shown difference is by chance and would move further towards the theoretical mean, when we add more observations (random samples with n=40).

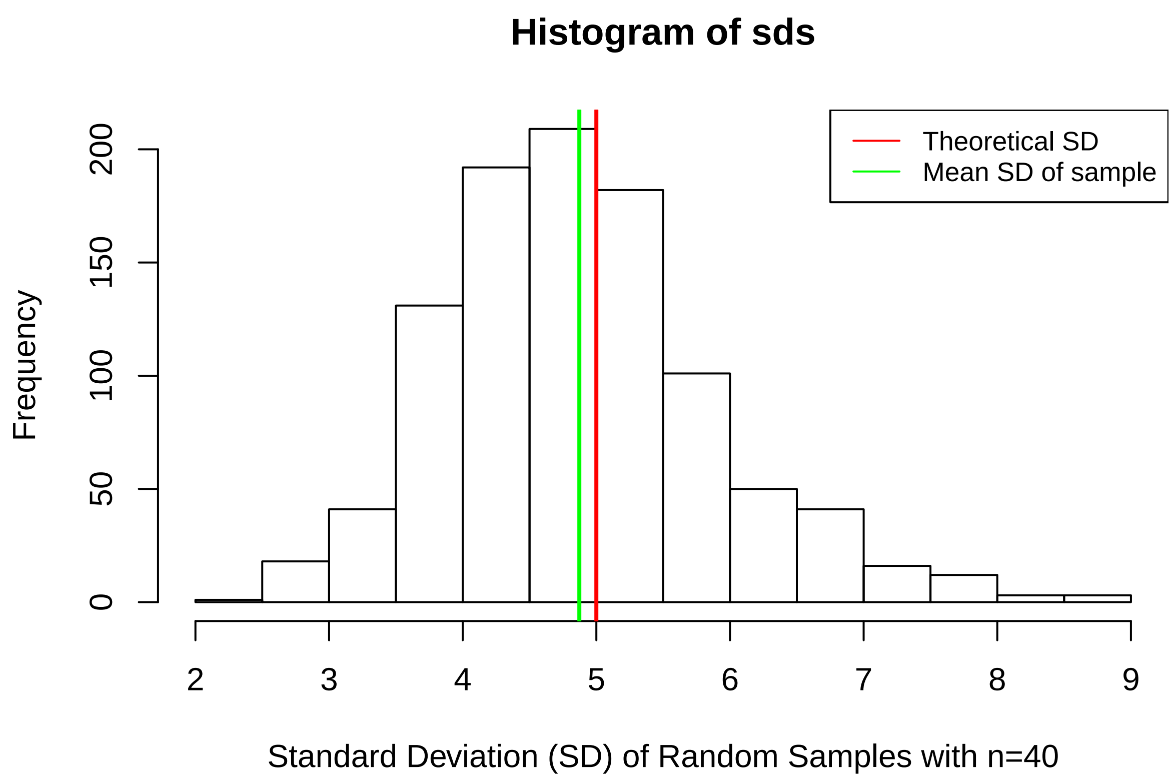
```
myhist <- hist(means,xlab="Means of random samples with n=40",15)
abline(v = theoreticalMean, col = "red", lwd = 2)
abline(v = means_avrg, col = "green", lwd = 2)
legend(x="topright", legend=c("Theoretical Mean", "Sample Mean"), lty=1, lwd=1,
       col = c("red", "green"), cex=0.8)
```



Step 3 - Discussion of the sample variance

The histogramm shows the distribution of standard distributions over repeatedly generating samples with n=40 for 1,000 times. The expected mean (red line) and the mean of the sample shows that there are only marginal differences. The shown difference is by chance and would move further towards the theoretical mean, when we add more observations (random samples with n=40).

```
hist(sds,xlab="Standard Deviation (SD) of Random Samples with n=40",20)
abline(v = 1/lambda, col = "red", lwd = 2)
abline(v = mean(sds), col = "green", lwd = 2)
legend(x="topright", legend=c("Theoretical SD", "Mean SD of sample"), lty=1, lwd=1,
       col = c("red", "green"), cex=0.8)
```

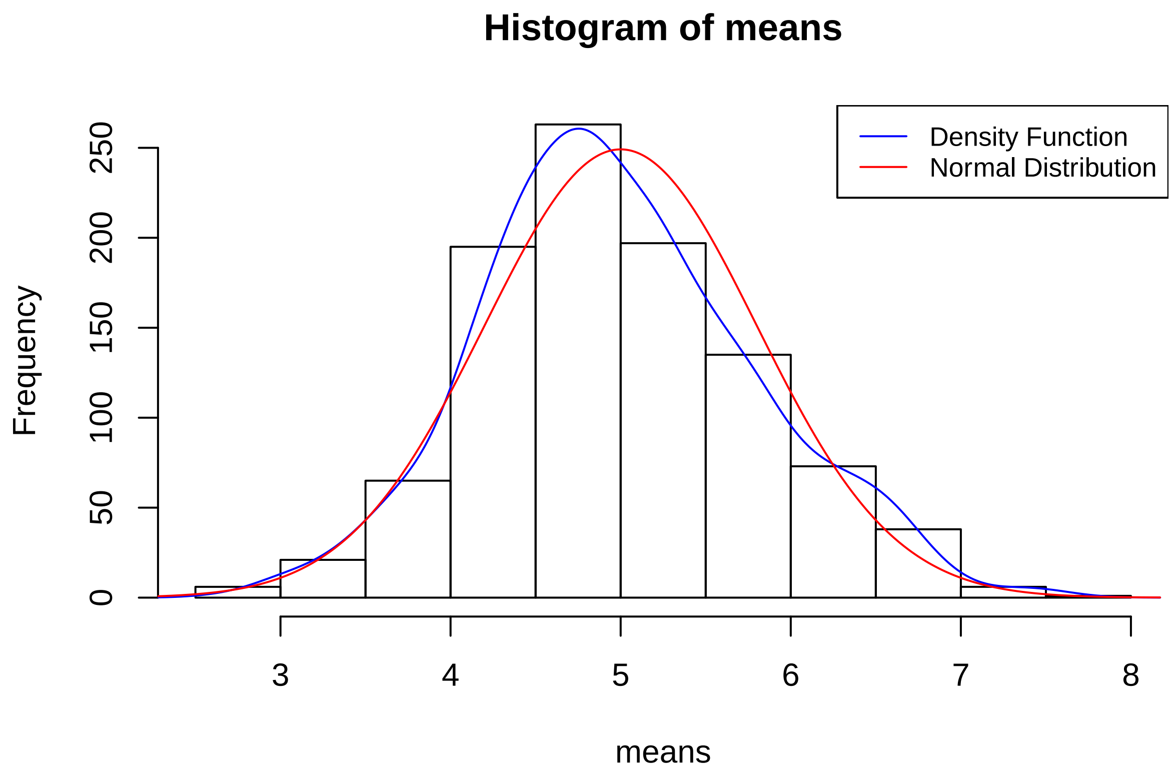


Step4 - Discuss the distribution

The graph below compares the distribution of means, its density function (blue) and the normal distribution (red). The similarity of the two functions suggests that the underlying distribution of means is indeed normally distributed. This can be furter tested by a two-sided t-test between the density function and the normal distribution (xhx.

```
#Creation of supporting variables for Historgram, Density and Bell Curve
multiplier <- myhist$counts / myhist$density #Multiplyer to transfer into frequency table.
means_dens <- density(means) #Calculated the underlying density of means
means_dens$y <- means_dens$y * multiplier[1] #Mutate the density function in refers to
        #the underlying frequency table

x <- means_dens$x
hx <- dnorm(x ,mean = theoreticalMean, sd = theoreticalSD)*multiplier[1]
xhx <- as.data.frame(cbind(x,hx)) #This is the normal distribution
```

```
#Plotting of relevant curves

plot(myhist)
lines(means_dens, col = "blue")
lines(xhx,col = "red")
legend("topright", legend=c("Density Function", "Normal Distribution"), col =
               c("blue", "red"), lty=1, lwd = 1, cex=0.8)
```



As suggested I want to execute the two-sided t-test and prove that the density function is normally distributed on a 95%-confidence level.

```
t.test(means_dens$y, xhx$hx, alternative = "two.sided",
       paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  means_dens$y and xhx$hx
## t = 0.017503, df = 1021.8, p-value = 0.986
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.62409  10.81532
## sample estimates:
## mean of x mean of y
##  82.41262  82.31700
```

The two-sided t-test suggusts a confidence interval of -4.286 to 4.383 in difference between the two means, before the null hypothesis (Normally distributed) can be rejected. The P-Value of 0.9825 suggusts that in 98.25% of all cases the mean of the means-sample is normally distributed.