Comparison of the exponential distribution and the central limit theorem

Andra T.

31st of January 2016

Summary

In this project we aim to 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/lamda**. We will set **lambda = 0.2** for all of the simulations. We wil investigate the distribution of averages of 40 exponentials, with a 1000 simulations

Central limit theorem Given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well defined expected value and a well defined variance, will be approximately normally distributed. Thus, if the distribution of independent observations is not strongly skewed, the sampling distribution is well approximated by a normal model : $\bar{x}_n \sim N(\mu, \frac{\sigma}{\sqrt{n}})$.

Data

```
set.seed(1000)

# set lambda to 0.2
lambda <- 0.2

# number of exponentials (40)
n <- 40

# Number of simulations (1000)
simulations <- 1000

# simulate
sim.exp<- replicate(simulations, rexp(n, lambda))

# calculate mean of exponentials
means.exp <- apply(sim.exp, 2, mean)
```

Comparing the sample mean with the theoretical mean of the distribution

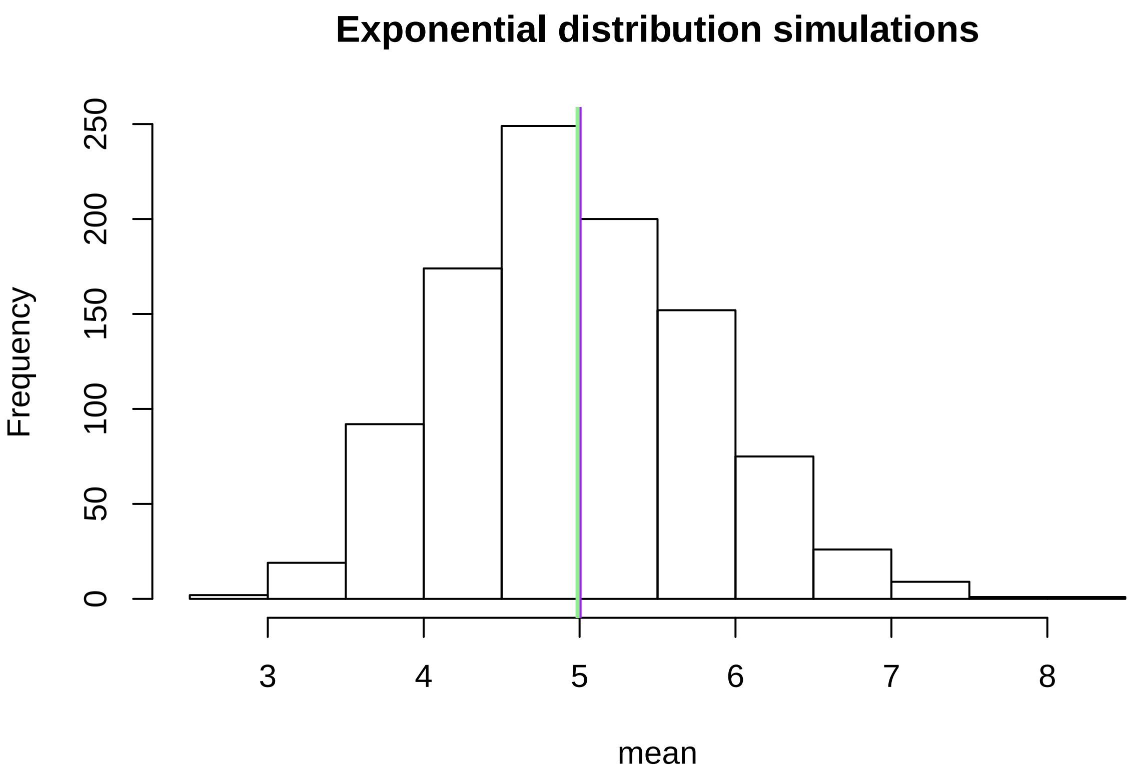
```
sample.mean<-mean(means.exp)
sample.mean
```

```
## [1] 4.986963
```

```
theoretical.mean<-1/lambda
theoretical.mean
```

```
## [1] 5
```

```
hist(means.exp,xlab="mean", main="Exponential distribution simulations")
abline(v=theoretical.mean, col='purple', lwd=2)
abline(v=sample.mean, col='lightgreen',lwd=2)
```



Conclusion : in this case we can see that the theoretical mean (4.97) is very close to the mean of the theoretical distribution (5) .

Comparing the sample variance with the theoretical variance of the distribution

From CLT : the variance of the sample of the 1000 means (sample.mean) is : $Var = 1/0.2^2 * 1/40 =>$ $Var = 0.625$

```
#standard deviation of the means distribution
sd.exp.dist<-sd(means.exp)
sd.exp.dist
```

```
## [1] 0.8089147
```

```
#variance of the means distribution
var.exp.dist<-var(means.exp)
var.exp.dist
```

```
## [1] 0.654343
```

```r
#theoretical standard deviation
theoretical.sd<-(1/lambda)/sqrt(n)
theoretical.sd
```

```
## [1] 0.7905694
```

```r
#theoretical variance
theoretical.var<-((1/lambda)*(1/sqrt(n)))^2
theoretical.var
```

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

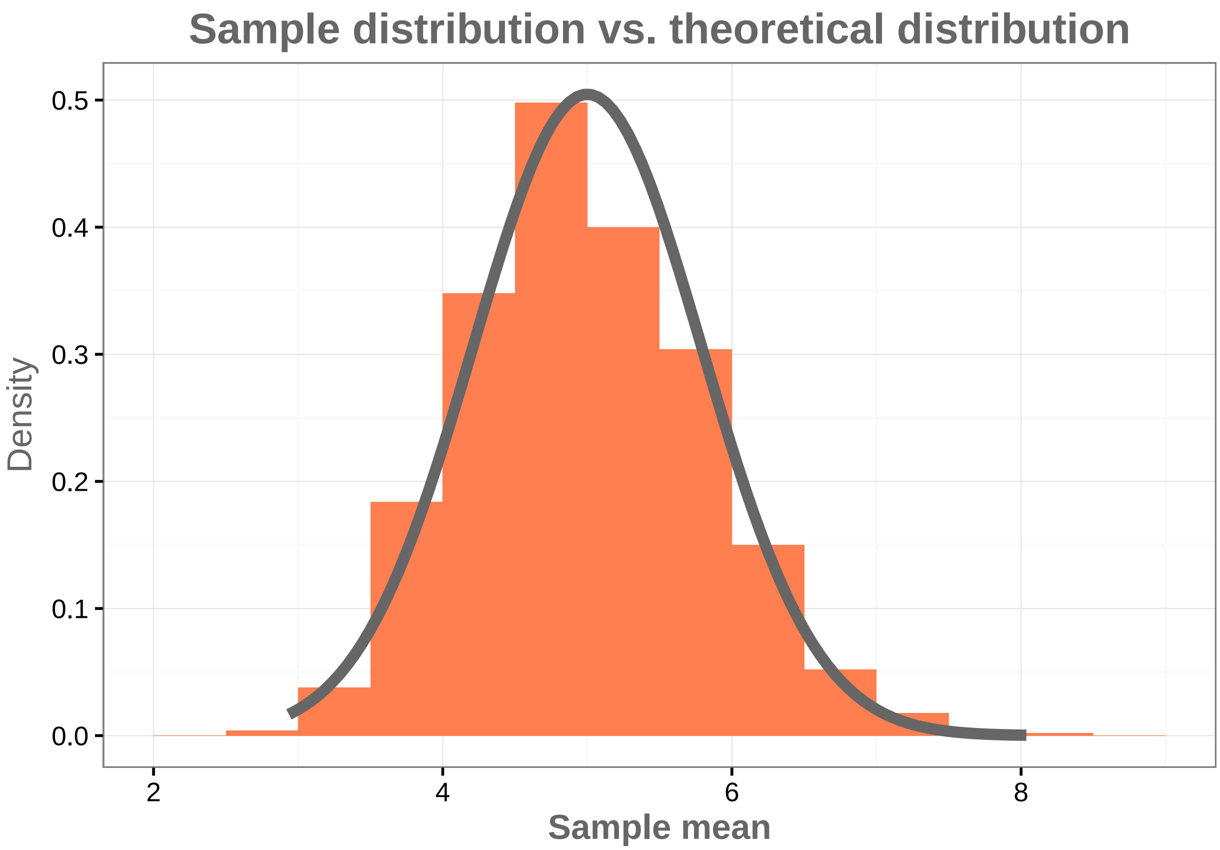
Conclusion : The standard deviation of the distribution is approximately 0.80, with a theoretical standard deviation of 0.79 On the other hand the variance of the distribution is approximately 0.65 , whereas the theoretical variance is 0.625.

Showing that the data is normally distributed

In order to show that the data is normally distributed, we can plot our samples and fit a density function that follows the normal distribution . In this case the mean of the normal distribution is 1/lambda with a sandard deviation of $\frac{1/lambda}{sqrt(n)}$

```r
library(ggplot2)
samples<-as.data.frame(means.exp)

ggplot(samples,aes(x=means.exp)) +
geom_histogram(binwidth=0.5,fill='coral',aes(y=..density..)) + theme_bw()+
stat_function(fun = dnorm, colour = "#666666", size=2, arg = list(mean = 5, sd=sqrt(0.625)))+
labs(x="Sample mean", y=expression("Density")) +
ggtitle("Sample distribution vs. theoretical distribution")+
theme(plot.title = element_text(color="#666666", face="bold", size=16, hjust=0.5)) +
theme(axis.title = element_text( color="#666666", face="bold", size=13))
```



Conclusion The sampling distribution of the mean of the exponential distribution (40 observations, $\lambda = 0.2$) is approximately normaly distributed with the mean $\frac{1}{0.2}$ and a sandard deviation of $\frac{1/0.2}{sqrt(40)}$