Coursera Statistical Inference Project - Part 1
Exponential Distribution Simulation

Filomena Ciccarelli

29 March 2016

Overiew

The purpose of this paper is to simulate an exponential distribution in R and compare it with the Central Limit Theorem (CLT). The exponential distribution can be simulated in R with the *rexp(n,lambda)* where *lambda* λ is the rate parameter. For the exponential distribution mean=standard deviation=1/lambda. For this simulation, we investigate the statistical properties of the distribution of the mean of 40 exponentials with $\lambda = 0.2$ (sample mean) and assess them against the theoretical mean of the distribution. The analysis will show that the sample distribution of the exponential distribution with n=40 observations and $\lambda = 0.2$ is distributed approximately as a $N(\frac{1}{0.2}, \frac{\frac{1}{0.2}}{\sqrt{40}})$.

Simulation of forty exponentials

The simulation will consist of 1000 samples of 40 Exponential distribution with $\lambda = 0.2$.

```
set.seed (1938)
n<-1000
lambda<-0.2
nval<-40
exp_mean<-0
for(i in 1:n) {
        exp_mean<-c(exp_mean, mean(rexp(nval,lambda)))
}
```

Comparison - Sample mean and theoretical distribution mean

The sample mean is:

```
round(mean(exp_mean),2)
```

```
## [1] 5.02
```

The sample mean is very close to the theoretical mean of the distribution $\mu = \frac{1}{0.2} = 5$ in accordance with the CLT.

Comparison - Sample variance and theoretical distribution variance

The sample variance is:

```
round(var(exp_mean),3)
```

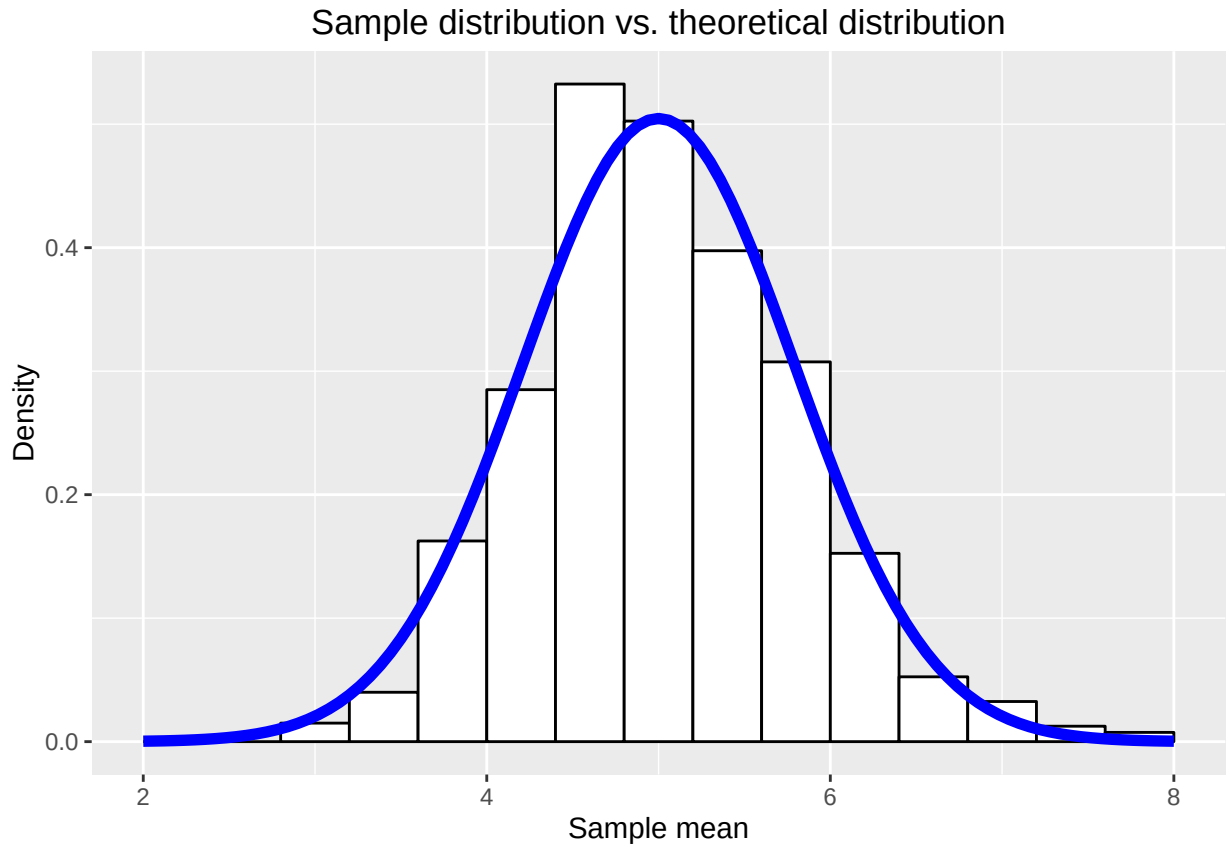
```
## [1] 0.631
```

The sample variance of the exponetial distribution is close to the variance of the theoretical distribution $\frac{\frac{1}{0.2^2}}{40} = 0.625$.

The sample distribution is approximately normal

In order to demonstrate that the sample distribution is approximately normal, in this section we plot the histogram of the sampled means and we overlay the density function of the theoretical sampling normal distribution $N(\frac{1}{0.2}, \frac{\frac{1}{0.2}}{\sqrt{40}})$

```
library(ggplot2)
data<-as.data.frame(exp_mean)
ggplot(data,aes(x=exp_mean))+
        geom_histogram(aes(y=..density..),binwidth = 0.4,color="black",fill="white")+
        stat_function(fun=dnorm,color="blue",
                      arg=list(mean=1/lambda,sd=(lambda*sqrt(nval))^-1),size=2)+
        xlim(2,8)+
        xlab("Sample mean")+
        ylab("Density")+
        ggtitle("Sample distribution vs. theoretical distribution")
```



Conclusion

The analysis confirmed that the sample distribution of the exponential distribution with n=40 observations and $\lambda = 0.2$ is distributd approximately as a $N(\frac{1}{0.2}, \frac{\frac{1}{0.2}}{\sqrt{40}})$.

The source code and the associated files of this report can be found on GitHub.