Simulation Exercise

Emilie Smith

November 14, 2016

### Overview

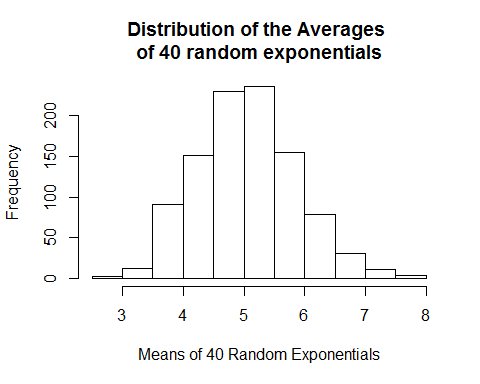
##### In this project I investigated the distribution of averages of 40 exponentials with a large number of simulations to illustrate the Central Limit Theorem. I demonstrated that as the sample size increases the distribution of averages of independent and identically distributed (IID) variables becomes that of a standard normal.

### Simulation

##### In R, the exponential distribution can be simulated with rexp(n, lambda) where lambda is the rate parameter. We were asked to investigate the distribution of averages of 40 exponentials, with lambda set to 0.2, and to perform 1000 simulations.

##### First I will take a look at the distribution of 1000 averages of 40 random exponentials.

set.seed(10) # set the seed for reproducibility  
lambda <- 0.2 # set lambda  
n <- 40 # set number of exponentials  
ns <- 1000 #set number of simulations  
distribution <- matrix(data=rexp(n \* ns, lambda), nrow = ns)  
averages <- data.frame(means=apply(distribution, 1, mean))  
  
hist(averages$means, xlab = "Means of 40 Random Exponentials",main = "Distribution of the Averages \nof 40 random exponentials")



#### 1)Show the sample mean and compare it to the theoretical mean of the distribution.

theor\_mean <- 1/lambda  
theor\_mean

## [1] 5

sample\_mean <- mean(averages$means)  
sample\_mean

## [1] 5.04506

##### The sample mean 5.0450596 is pretty close to the theoretical mean 5.

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

##### The theoretical standard deviation for an exponential distribution is 1/lambda/sqrt(n), and the theoretical variance is equal to the square of the theoretical standard deviation.

theor\_sd <- 1/lambda/sqrt(n)  
theor\_var <- theor\_sd^2  
theor\_var

## [1] 0.625

sample\_var <- var(averages$means)  
sample\_var

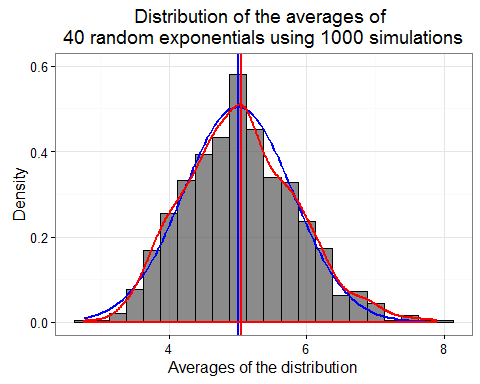
## [1] 0.6541736

##### The theoretical variance 0.625 is close to the sample variance 0.6541736.

#### Show that the distribution is approximately normal.

##### I will plot the distribution of of the averages of 40 random exponentials and overlay it with the normal distribution.

library(ggplot2)  
ggplot(data=averages,aes(x=means))+  
 theme\_bw()+  
 geom\_histogram(aes(y=..density..),alpha=0.7,binwidth=.25,col="black") +   
 ylim(c(0,0.6))+  
 stat\_function(fun=dnorm,args=list(mean=theor\_mean,sd=theor\_sd), color="blue", size=1) +  
 geom\_vline(xintercept=theor\_mean,color="blue", size=1) +   
 geom\_density(colour="red", size=1) +  
 geom\_vline(xintercept=sample\_mean,color="red", size=1) +  
 xlab("Averages of the distribution") + ylab("Density")+  
 ggtitle("Distribution of the averages of \n40 random exponentials using 1000 simulations")



### Conclusion

##### As illustrated in the graph above, the distribution of the averages of 40 random exponentials using 1000 simulations (in red) looks very similar to the Bell curve of the theoretical normal distribution (in blue). The sample mean (red vertical line) and the theoretical mean (blue vertical line) are very close. This illustrates the Central Limit Theorem that states that that the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, if the sample size is large enough.