Simulation of Central Limit Theorem

Dorota

21 June 2015

## Overview

By performing simulations in R we will show that average of n exponentialy distributed variables is approximately normal with mean 1/lambda and the standard deviation 1/(lambda\*sqrt(n)). It will visually present the central limit theorem, that states that the distribution of average of a large number of independent, identically distributed variables will be approximately normal, regardless of the underlying distribution.

## Simulations:

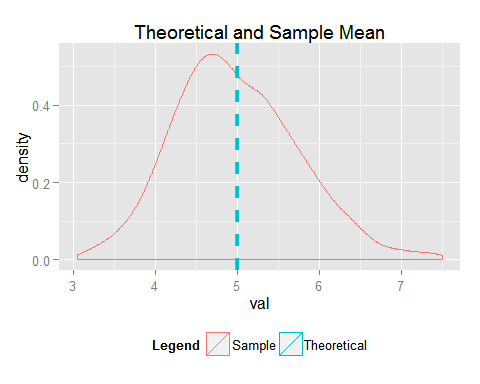
We wil perform 1000 simulation. The variable we simulate is average of 40 variables, exponentialy distributed with lambda 0.2.

set.seed(7)  
mns = NULL  
lambda=0.2  
n=40  
for (i in 1 : 1000) mns = c(mns, mean(rexp(n,lambda)))  
mns<-as.data.frame(merge('Sample',mns))  
colnames(mns)<-c('Legend','val')

## Sample Mean versus Theoretical Mean

Below we can compare sample mean (4.98) vs theoretical mean of exponentialy distributed variables (r round(as.numeric( 1/lambda),2)

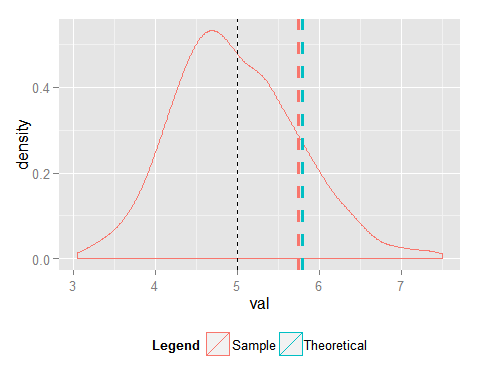
# calculate mean  
mean=as.data.frame(merge('Sample', as.numeric(mean(mns$val,na.rm=T))))  
mean=rbind(mean,merge('Theoretical',as.numeric( 1/lambda)))  
colnames(mean)<-c('Legend','mean\_val')  
  
#plot  
library(ggplot2)  
ggplot(data=mns,aes(x=val,colour=Legend)) +  
geom\_density() +  
geom\_vline(data=mean, aes(xintercept=mean\_val, colour=Legend),linetype="dashed", size=1.1) +  
theme(legend.position="bottom") +   
labs(title="Theoretical and Sample Mean")



## Sample Variance versus Theoretical Variance

Below we can compare sample standard deviation (0.76) vs theoretical standard deviation of exponentialy distributed variables (r round(as.numeric( 1/(n\*lambda)),2), Chart below shows mean + one standrd deviation.

# calculate mean  
sdev=as.data.frame(merge('Sample', as.numeric(sd(mns$val,na.rm=T))))  
sdev=rbind(sdev,merge('Theoretical',as.numeric( 1/(lambda\*sqrt(n))) ))  
sdev<-cbind(sdev,as.data.frame(sdev[,2]+mean[,2]))  
colnames(sdev)<-c('Legend','sdev\_val','sm\_val')  
  
  
#plot  
ggplot(data=mns,aes(x=val,colour=Legend)) +  
geom\_density() +  
geom\_vline(data=mean[2, ], aes(xintercept=mean\_val), colour='black',linetype="dashed", size=0.25) +  
geom\_vline(data=sdev, aes(xintercept=sm\_val, colour=Legend),linetype="dashed", size=1.1) +  
theme(legend.position="bottom")

 ## Distribution

Comparing the shapes we can see that the distribution is approximately normal with mean 1/lambda and the standard deviation 1/(lambda\*sqrt(n).

#normal distribution  
set.seed(7)  
rnormal = rnorm(1000,1/lambda,1/(sqrt(n)\*lambda))  
rnormal<-as.data.frame(merge('Normal',rnormal ))  
colnames(rnormal)<-c('Legend','val')  
dist<-rbind(mns,rnormal)  
  
#plot  
ggplot(data=dist,aes(x=val,colour=Legend)) +  
geom\_density() +  
theme(legend.position="bottom") +   
labs(title="Sample and Normal Distribution")

