Statistical inference course project | Part 1

Danna

May 28, 2017

## Overview

This project for the Statistical Inference class. In this, we will 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/lambda.

For the project, the following instructions were provided:

Set lambda = 0.2 for all of the simulations. Investigate the distribution of averages of 40 exponentials. Do a thousand simulations.

## Simulations

# set seed for reproducability  
set.seed(100)  
# set lambda to 0.2  
lambda <- 0.2  
# 40 samples  
n <- 40  
# 1000 simulations  
simulations <- 1000  
  
# simulate  
simulated\_exponentials <- replicate(simulations, rexp(n, lambda))  
  
# calculate mean of exponentials  
means\_exponentials <- apply(simulated\_exponentials, 2, mean)

## Question 1

distribution mean:

analytical\_mean <- mean(means\_exponentials)  
analytical\_mean

## [1] 4.999702

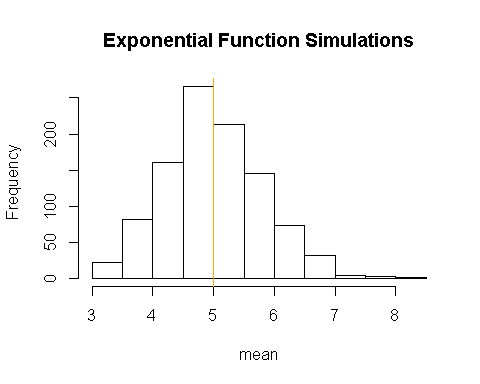
analytical mean:

theory\_mean <- 1/lambda  
theory\_mean

## [1] 5

visualization:

hist(means\_exponentials, xlab = "mean", main = "Exponential Function Simulations")  
abline(v = analytical\_mean, col = "red")  
abline(v = theory\_mean, col = "orange")



### Answer 1

The analytics mean is 4.999702 the theoretical mean 5. The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

## Question 2

Show how variable it is and compare it to the theoretical variance of the distribution

# standard deviation of distribution  
standard\_deviation\_dist <- sd(means\_exponentials)  
standard\_deviation\_dist

## [1] 0.8020251

# standard deviation from analytical expression  
standard\_deviation\_theory <- (1/lambda)/sqrt(n)  
standard\_deviation\_theory

## [1] 0.7905694

# variance of distribution  
variance\_dist <- standard\_deviation\_dist^2  
variance\_dist

## [1] 0.6432442

# variance from analytical expression  
variance\_theory <- ((1/lambda)\*(1/sqrt(n)))^2  
variance\_theory

## [1] 0.625

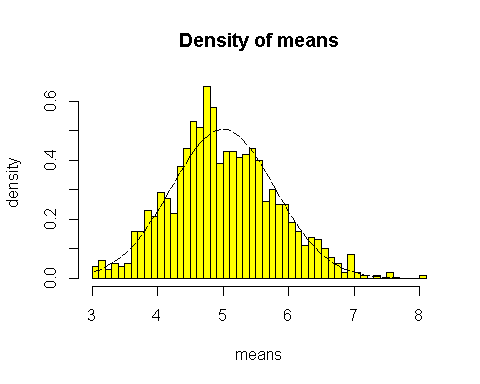
### Answer 2

Standard Deviation of the distribution is 0.8020251 with the theoretical SD calculated as 0.7905694. The Theoretical variance is calculated as (1λ∗1n√)2(1λ∗1n)2 = 0.625. The actual variance of the distribution is 0.6432442

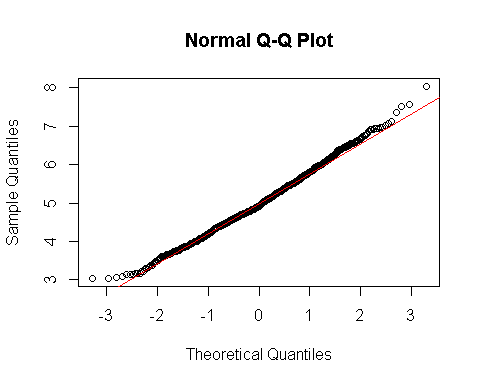
## Question 3

Show that the distribution is approximately normal.

xfit <- seq(min(means\_exponentials), max(means\_exponentials), length=100)  
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))  
hist(means\_exponentials,breaks=n,prob=T,col="yellow",xlab = "means",main="Density of means",ylab="density")  
lines(xfit, yfit, pch=22, col="black", lty=5)



# compare the distribution of averages of 40 exponentials to a normal distribution  
qqnorm(means\_exponentials)  
qqline(means\_exponentials, col = 2)



### Answer 3

Due to Due to the central limit theorem (CLT), the distribution of averages of 40 exponentials is very close to a normal distribution.