Statistical Inference - Simulation project

H.Harvey

25 December 2015

## Introduction

The purpose of this project is 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/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials.

## Setting the variables

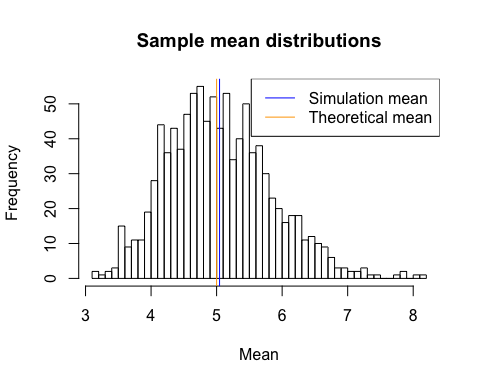
## R Preparation for reproducible results  
set.seed(10)  
# set lambda to 0.2  
lambda <- 0.2  
# 40 samples  
Num\_Sample <- 40  
# 1000 simulations  
Num\_Sim <- 1000

## Calculations of the sample mean

# Replicate 1000 instances of the random 40 exponentials  
Sim\_Exp <- replicate(Num\_Sim, rexp(Num\_Sample, lambda))  
# Calculate the mean of the sample of exponentials  
mean\_Sim\_Exp <- apply(Sim\_Exp, 2, mean)  
# Calculate the mean of the sample mean  
mean\_Sample <- mean(mean\_Sim\_Exp)  
# Calculate the theoritical mean  
mean\_Theory <- 1/lambda

## Graphical representation of the central/theoritical mean

# Represent the sample mean distribution  
hist(mean\_Sim\_Exp, breaks=50, xlab = "Mean", main = "Sample mean distributions")  
# Represent the sample mean of the random simulation  
abline(v = mean\_Sample , col = "blue")  
# Represent the theoritical mean  
abline(v = mean\_Theory, col = "orange")  
legend('topright', c("Simulation mean", "Theoretical mean"),  
 lty=c(1,1), col=c("blue", "orange"))



round(mean\_Sample, digits = 3)

## [1] 5.045

mean\_Theory

## [1] 5

The distribution of sample means is centered at 5.045 and the theoretical center of the distribution is calculated at 5. The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

## Calculation of the variance and standard variation

# Calculate the standard deviation of the sample distribution  
stddev\_Sample <- sd(mean\_Sim\_Exp)  
# Calculate the theoritical standard deviation  
stddev\_Theory <- (1/lambda)/sqrt(Num\_Sample)  
# Calcultate the variance of the sample distribution  
var\_Sample <- stddev\_Sample^2  
# Calculate the theoritical variation  
var\_Theory <- ((1/lambda)\*(1/sqrt(Num\_Sample)))^2

round(stddev\_Sample, digits = 3)

## [1] 0.798

round(stddev\_Theory, digits = 3)

## [1] 0.791

round(var\_Sample, digits = 3)

## [1] 0.637

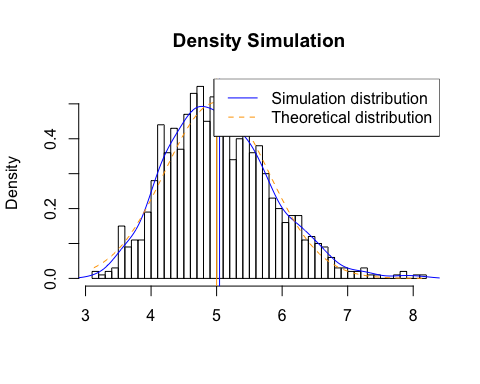
var\_Theory

## [1] 0.625

The standard variation is 0.798 where the theoeritical standard deviation is calculated at 0.791. The variance of sample means is 0.637 where the theoretical variance of the distribution is calculated at 0.625.

## Graphical representation of the distribution

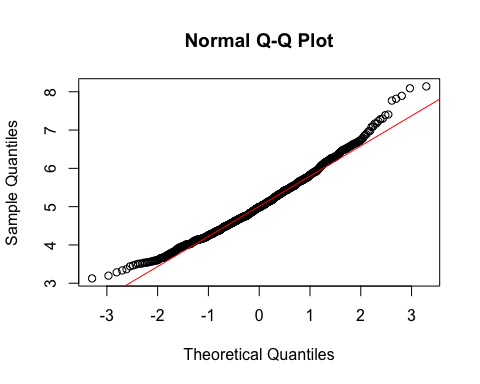
hist(mean\_Sim\_Exp, breaks=50, prob=TRUE,  
 main="Density Simulation",  
 xlab="")  
# Calculate and identify the density of the sample mean  
lines(density(mean\_Sim\_Exp), col="blue")  
# Identify the mean of the sample mean  
abline(v=mean(mean\_Sim\_Exp), col="blue")  
# Identify the theoretical mean  
abline(v=1/lambda, col="orange")  
  
# Calculate the theoretical density of the averages of samples  
xfit <- seq(min(mean\_Sim\_Exp), max(mean\_Sim\_Exp), length=100)  
yfit <- dnorm(xfit, mean=mean\_Theory, sd=stddev\_Theory)  
lines(xfit, yfit, pch=22, col="orange", lty=2)  
  
# Legend  
legend('topright', c("Simulation distribution", "Theoretical distribution"),  
 lty=c(1,2), col=c("blue", "orange"))



As represented in the graph, the mean of the random sampled exponantials distributions overlaps with the theoritical mean of the distribution, due to the Central Limit Theorem.

## Validation

qqnorm(mean\_Sim\_Exp)  
qqline(mean\_Sim\_Exp, col = 2)



## Conclusion

Following the central limit theorem (CLT), we can claim that the distribution of averages of 40 exponentials is very close to a normal distribution.