Exponential distribution Simulation and Analysis

Ion Suruceanu

September 24th, 2015

Overview

The main aim of this analysis project is to demonstrate Central Limit Theorem by simulation for Exponential distribution. Exponential distribution is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate.

The interesting fact is that mean and standard deviation is equal to $1/\lambda$.

Simulation

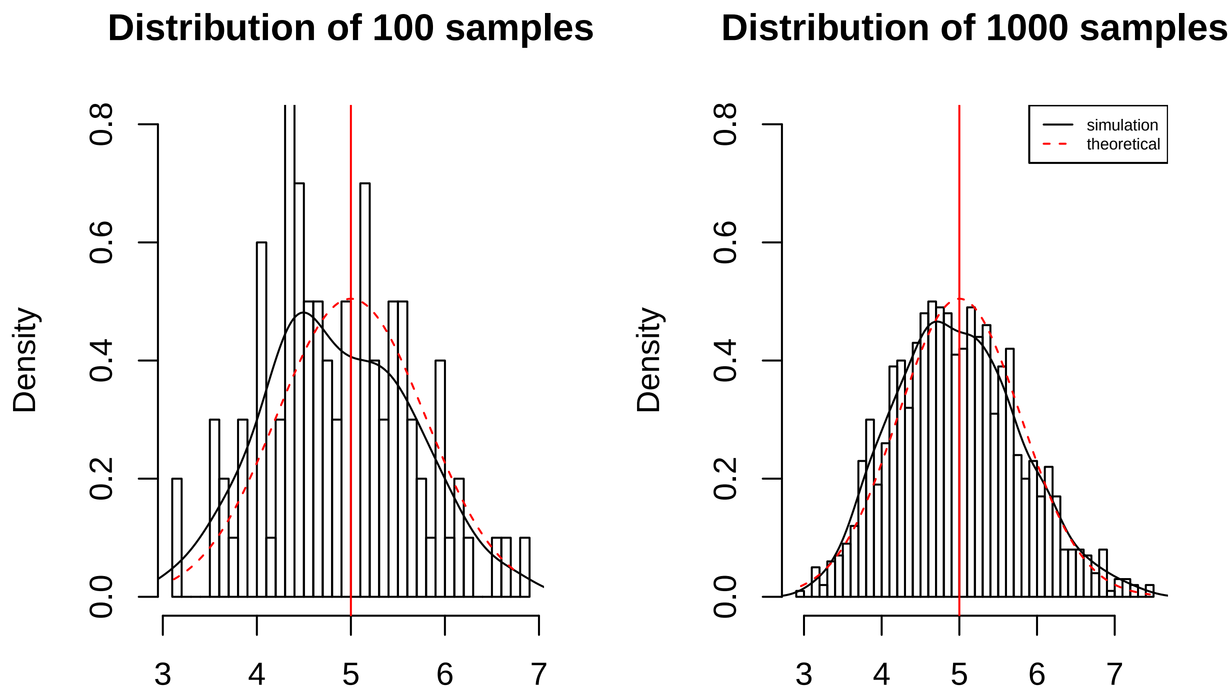
For a better understanding of Central Limit Theorem (CLT) we are going to run two simulations with a default $batchSize = 40$. First with 100 random uniforms and second with 1000 random uniforms and will compare the results. The default value of λ will set to 0.2 *(See figure bellow)*

```
set.seed(17)
lambda <- 0.2
sim100 <- simulation(100)
sim1000 <- simulation(1000)

par(mfrow = c(1,2))

plotHistogram(sim100)
plotHistogram(sim1000)

#only with recordGraphics was able to draw a nice legend
    recordGraphics(
 legend('topright', c("simulation", "theoretical")
           , lty=c(1,2)
           , col=c("black", "red")
           , cex=0.5)
    ,list(), getNamespace("graphics"))
```



For more details on R functions *simulation* and *plotHistogram* please consult the Appendex

Analysis

The Central Limit Theorem (CLT) states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution.

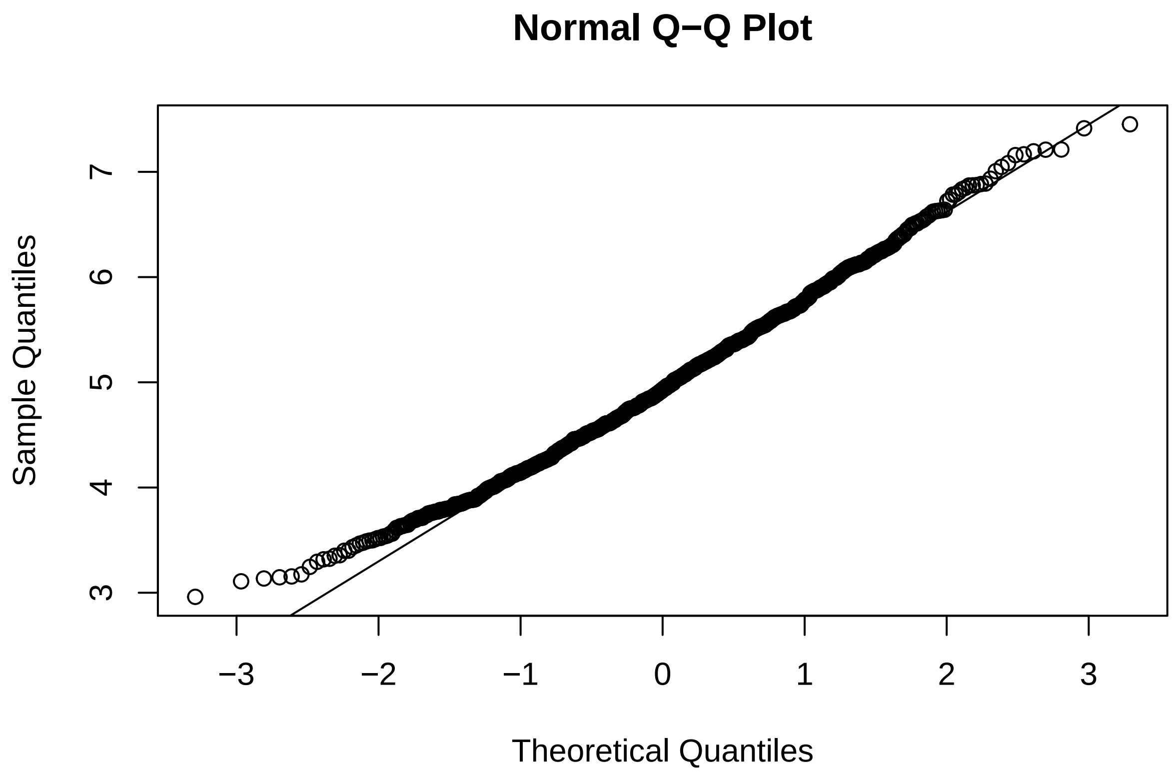
From figure above we can see that for 100 randoms the distribution has a significant deviation from normal one. As we increase the number of simulations to 1000 it approaches to normal distribution.

The distribution of 1000 sample means is 4.9723784 and the theoretical value of the distribution is $\lambda^{-1} = 5$.

The variance of sample means is 0.6465382 where the theoretical value is $\sigma^2/n = 1/(\lambda^2 n) = 1/(0.04 \times 40) = 0.625$.

The figure bellow also indicates to normality

```r
qqnorm(sim1000); qqline(sim1000)
```



See source files and documentation on githib:

Appendix

R functions for simulation and graphical representation.

The R functions bellow will allow us to accomplish this.

```
### Function to run exponential distribution simulation
### Arguments
###   simulationCount - number of simulations to run
###   batchSize - number of random uniforms
###   rate - rate parameter or lambda
simulation <- function(simulationCount, batchSize = 40, rate = 0.2){
    sim <- matrix(
        rexp(simulationCount * batchSize, rate = rate)
        , simulationCount
        , batchSize)
    simMeans <- rowMeans(sim)
    simMeans
}
```

And

```r
### Function to plot exponential distribution histogram
### Arguments
###  meanX - a vector of simulated averages
###  lambda - rate parameter
###  batchSize - number of random uniforms
plotHistogram <- function(meanX, lambda = 0.2, batchSize = 40){

    len = length(meanX)
    hist(meanX, breaks = 50, prob = T, xlab=""
         , ylim = c(0, 0.8)
         , main = paste("Distribution of", len, "samples"))
    lines(density(meanX))
    #theoretical
    abline(v = 1/lambda, col="red")
    x <- seq(min(meanX), max(max(meanX)), length = len)
    y <- dnorm(x, mean = 1/lambda, sd = (1/lambda/sqrt(batchSize)))
    lines(x, y, col="red", lty=2)

}
```