Statistical inference
Simulating exponential distribution in R and analyzing statistics

Aigars Kvists
MSc.Financial Engineering
davidkvists@yahoo.com
Aliso Viejo, CA 92656

January 11, 2015

1 Abstract

In this project we simulate exponential distribution in R and examine its mean and variance using properties of Central Limit Theorem. The underlying idea is that given sufficiently large sample from population, the mean and variance of population can be approximated performing only simulations on sample set. Also when number of simulation is going to infinity, the distributions of sample mean and variance will follow normal (Gaussian distribution) regardless of the underlying distribution.

2 Mathematical interpretation of exponential distribution

The exponential distribution has probability density function (pdf) as

$$f(x) = \lambda exp^{(-\lambda x)}, \quad x \geq 0$$

The mean of exponential distribution is: $\mathbb{E}[X] = 1/\lambda$
The variance of exponential distribution is: $Var[X] = 1/\lambda^{-2}$

3 Simulation of exponential distribution in R

For the scope of this project, we simulate $n = 40000$ variables with $\lambda = 0.20$ using R function $rexp(n, 0.2)$
The histogram of simulation is given in Figure 1.

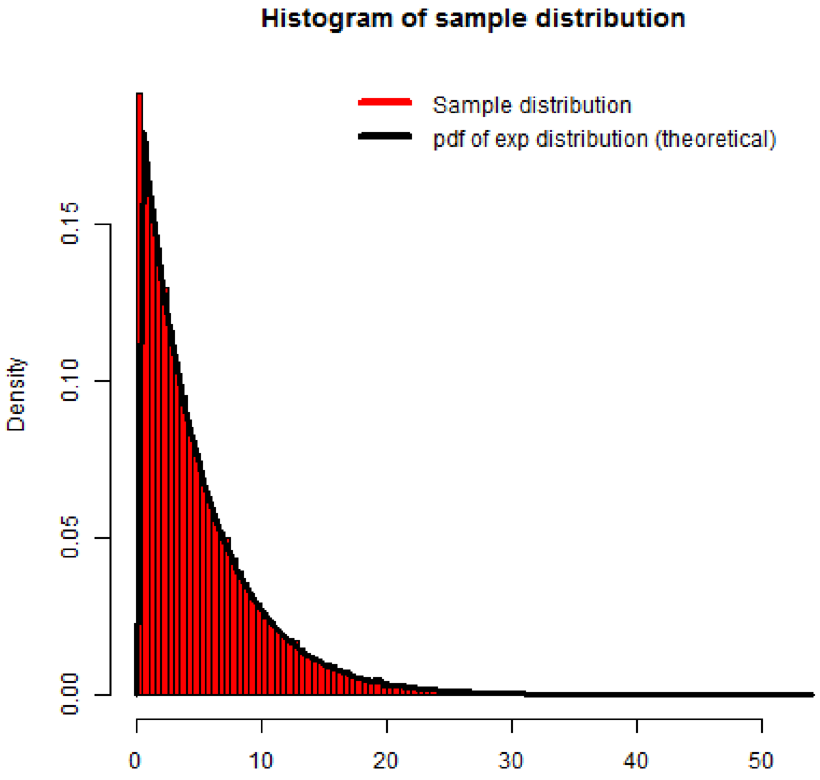

Figure 1.

4 Calculation of population mean and variance using simulated sample

The theoretical mean for exp distribution $\lambda = 0.2$ is $\mu = 1/\lambda^{-1} = 5$
The theoretical variance of exp distribution $\lambda = 0.2$ is $var = 1/\lambda^{-2} = 25$

First, the simulated data is stored in $1000x40$ matrix replicating 1000 simulation of 40 random variables exponentially distributed. For sample mean calculation, the mean of each row is calculated, then the expected value of mean for all 1000 simulations is calculated giving the answer of population mean. For sample variance calculation, the variance of each row is calculated, then the expected value of variance for all 1000 simulations is calculated giving the answer of population variance. The output from R console that shows how sample mean and sample variance are distributed around the mean. As seen the values are pretty close to theoretical values: $\mu = 5$ and $var = 25$.

```
> summary(sample_mean)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.695   4.460   4.917   4.999   5.486   7.872

> summary(sample_var)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  6.331  16.870  22.480  24.540  29.350 113.700

>
```

Standard deviation of the average of random sample is called Standard Error (SE). It shows how averages of random sample n deviates from population mean. Theoretical SE is computed as $SE = \sigma/sqrt(n)$ The R console output below shows theoretical SE and SE from the simulation.

```
> SE_theoretical
[1] 0.7905694

> SE_sample
[1] 0.7845405
```

5 Histograms of sample distributions

The one important aspect that follows from CLT is that sample means and variance are also random variables and when number of observation tends to be large follows the normal (Gaussian) distribution. Figure 2 and Figure 3 shows this important property.

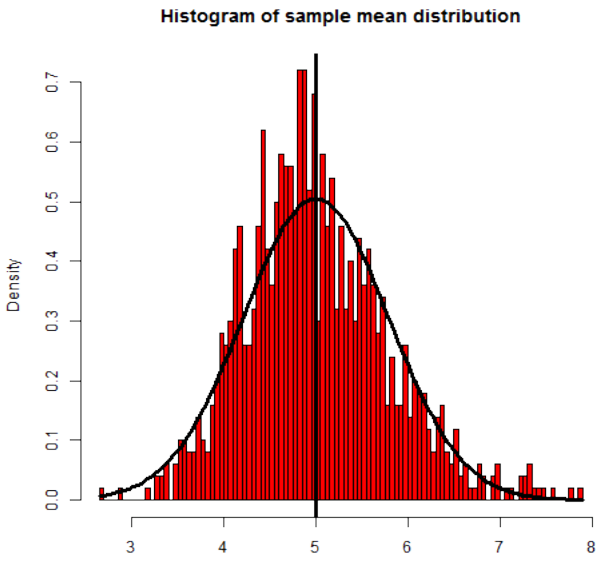

Figure 2. Distribution of sample mean

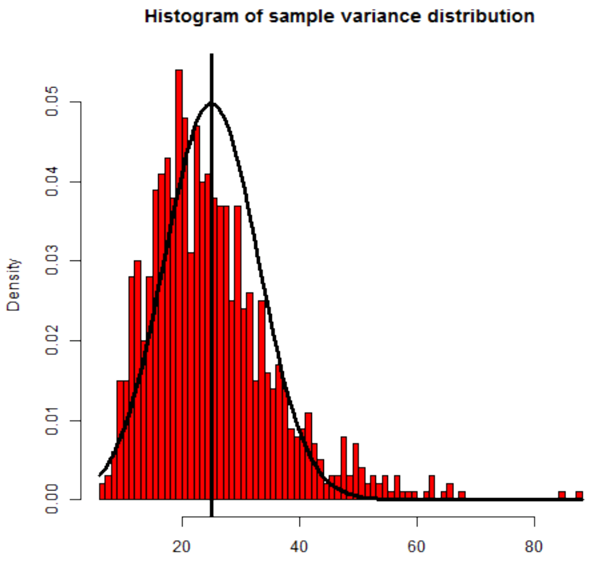

Figure 3. Distribution of sample variance

As seen from the above graphs distributions of sample mean and variance are indeed close to normal.

6 Appendix

R code

```
 ## Statistical inference course
#Course project
#Investigate exponantial distribution and compare it to
    Central Limit Theorem

##Setting proper working directory
setwd(file.path("C:","Users","David","Documents","
    Financial Engineering","Coursera","Statistical
    Inference"))

# simulate exponantial distribution and compare it to
    theoretical density function

hist(rexp(40000,0.20),100,main="Histogram of sample
    distribution", col="red",freq=FALSE,xlab="")
curve(dexp(x,0.20), add=TRUE,lwd=3)
legend("topright",c("Sample distribution","pdf of exp
    distribution (theoretical)"),
         col=c("red","black"),lty=c(1,1),lwd=c(4,4),bty="n
            ")
dev.copy(png,file="st_inf_plot1.png")
dev.off()


#Theoretical mean of exp distribution lambda^(-1) = 5
#Theoretical variance of exp distribution is lambda^(-2)
    =25

#calculating mean of the sample distribution
sample_mean<-(apply(matrix(rexp(40000,0.2),1000),1,mean))
SE_sample<-sd(apply(matrix(rexp(40000,0.2),1000),1,mean))
     #measures sample mean deviation


#calculating variance of sample mean
sample_var<-(apply(matrix(rexp(40000,0.20),1000),1,var))
#SE of variance distribution in sample
SE_var<-sd(apply(matrix(rexp(40000,0.20),1000),1,var))

#calculating population mean
dist_mean<-mean(sample_mean)
```

```
#calculating population variance
dist_var<-mean(sample_var)


#calculating theoretical SE for sample mean
SE_theoretical<-5/sqrt(40)
summary(sample_mean)
summary(sample_var)
SE_theoretical
SE_sample


#plotting histogram of distribution  of sample mean
hist(sample_mean,100,col="red",xlab="",main="Histogram of
     sample mean distribution",freq=FALSE)
curve(dnorm(x,5,SE_theoretical), add=TRUE,lwd=3)
abline(v=5,col="black",lwd=3)
dev.copy(png,file="st_inf_plot2.png")
dev.off()


#plotting histogram of distribution  of sample variance
hist(sample_var,100,col="red",xlab="",main="Histogram of
    sample variance distribution",freq=FALSE)
curve(dnorm(x,25,8), add=TRUE,lwd=3)

abline(v=25,col="black",lwd=3)

dev.copy(png,file="st_inf_plot3.png")
dev.off()
```