Statistical Inference Course Project - Simulation Exercise

Vijay Vaidya

Wednesday, July 22, 2015

Overview

Investigate exponential distribution in R and compare with Central Limit Theorem. Illustrate via simulation that mean and variance approach the theoratical values for large number of simulations. We will take 1000 simulations.

```
rm(list = ls()) # clear the environment
lambda <- 0.2   # set the rate parameter
n <- 40         # exponentials' count
simCount <- 1000 # simulations' count
```

Compare Sample mean with theoratical mean.

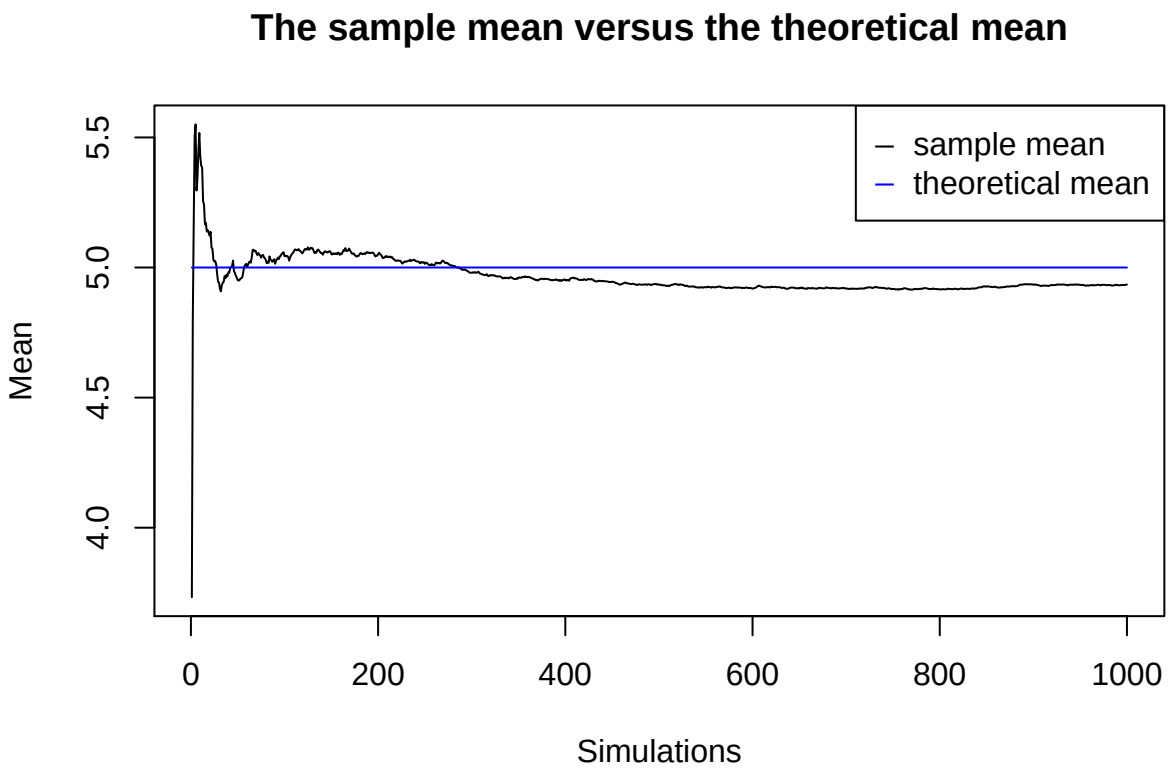
We are given theoratical mean as 5 (1/0.2). We will create a vector "means" of 1000 elements. Each element is the mean of 40 exponentials. We take the average of "means" and then plot the chart.

```
theoreticalMean <- 1 / lambda
means <- c()
for(i in 1:simCount)
{
    means <- c(means, mean(rexp(n, lambda)))
}
m <- cumsum(means) / (1:simCount)

plot(x = 1:simCount,
     y = m,
     type = "l",
     main = "The sample mean versus the theoretical mean",
     xlab = "Simulations",
     ylab = "Mean",
     col = "black")

lines(x = c(1, simCount),
      y = c(theoreticalMean, theoreticalMean),
      type = "l",
      col = "blue")

legend("topright", pch = "_",
       col = c("black", "blue"),
       legend = c("sample mean", "theoretical mean"))
```



Compare Sample variance with theoratical variance.

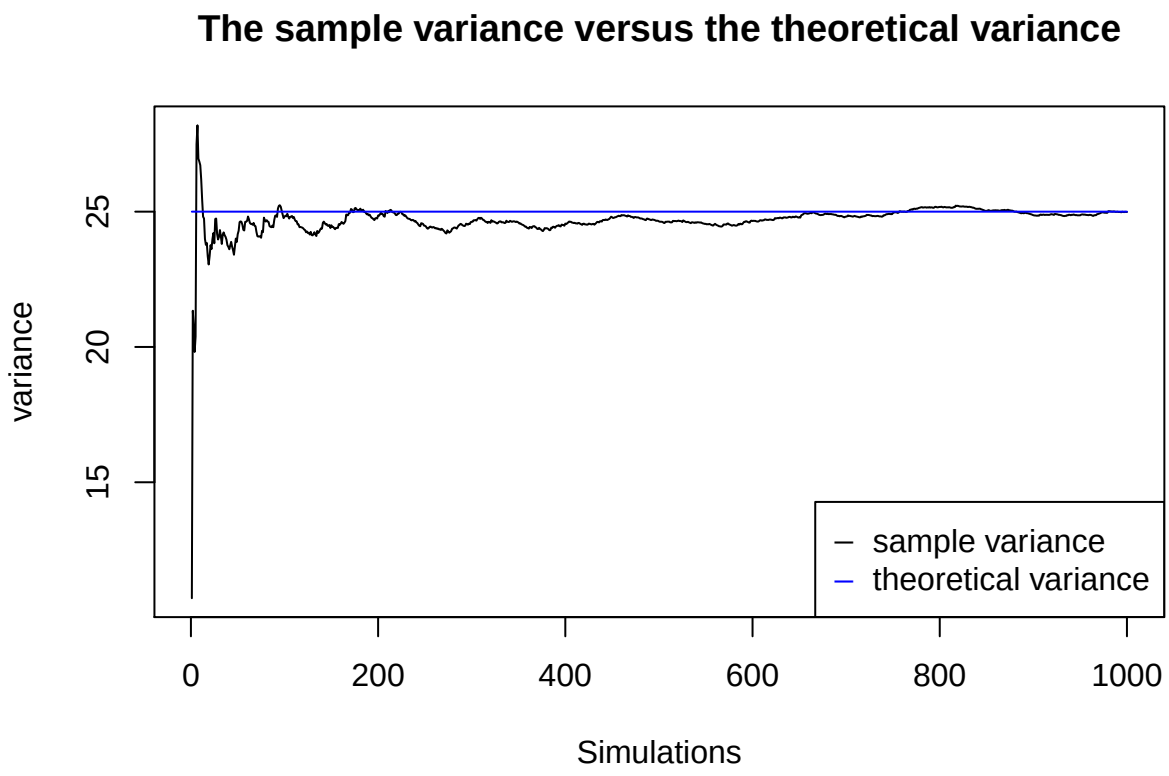
We have theoratical variance as 25 (1/0.04) will create a vector "var" of 1000 elements. Each element is the variance of 40 exponentials. We take the average of "vars" and then plot the chart

```
theoreticalVar <- 1 / lambda^2
vars <- c()
for(i in 1:simCount)
{
    vars <- c(vars, var(rexp(n, lambda)))
}
v <- cumsum(vars) / (1:simCount)

plot(x = 1:simCount,
     y = v,
     type = "l",
     main = "The sample variance versus the theoretical variance",
     xlab = "Simulations",
     ylab = "variance",
     col = "black")

lines(x = c(1, simCount),
      y = c(theoreticalVar, theoreticalVar),
      type = "l",
      col = "blue")
```

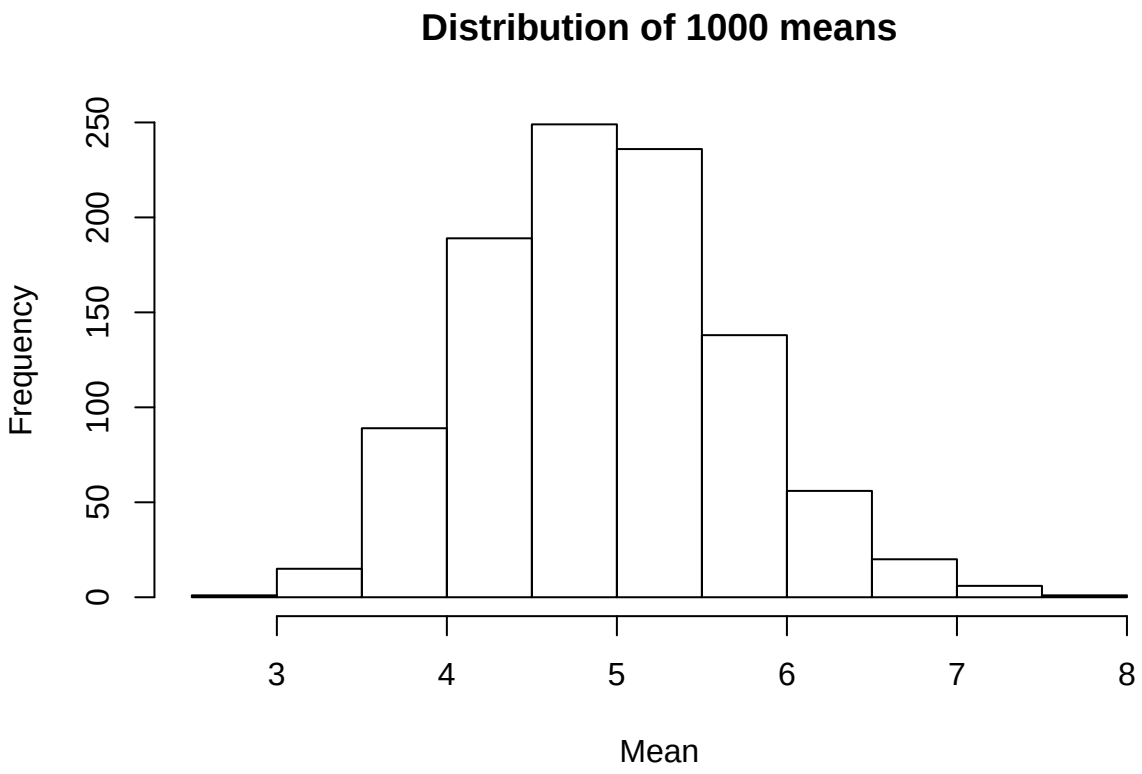
```
legend("bottomright", pch = "_",
       col = c("black", "blue"),
       legend = c("sample variance", "theoretical variance"))
```



Distribution.

Now we prove that as the sample size increases, distribution becomes standard normal.

```
hist(means,
     main = paste("Distribution of", simCount, "means"),
     xlab = "Mean")
```


Distribution of 1000 means
Frequency
0
50
100
150
200
250
3
4
5
6
7
8
Mean