Statistical Inference Course Project - The Exponential Distribution

Maciej Nowak

Monday, July 20, 2015

Overview

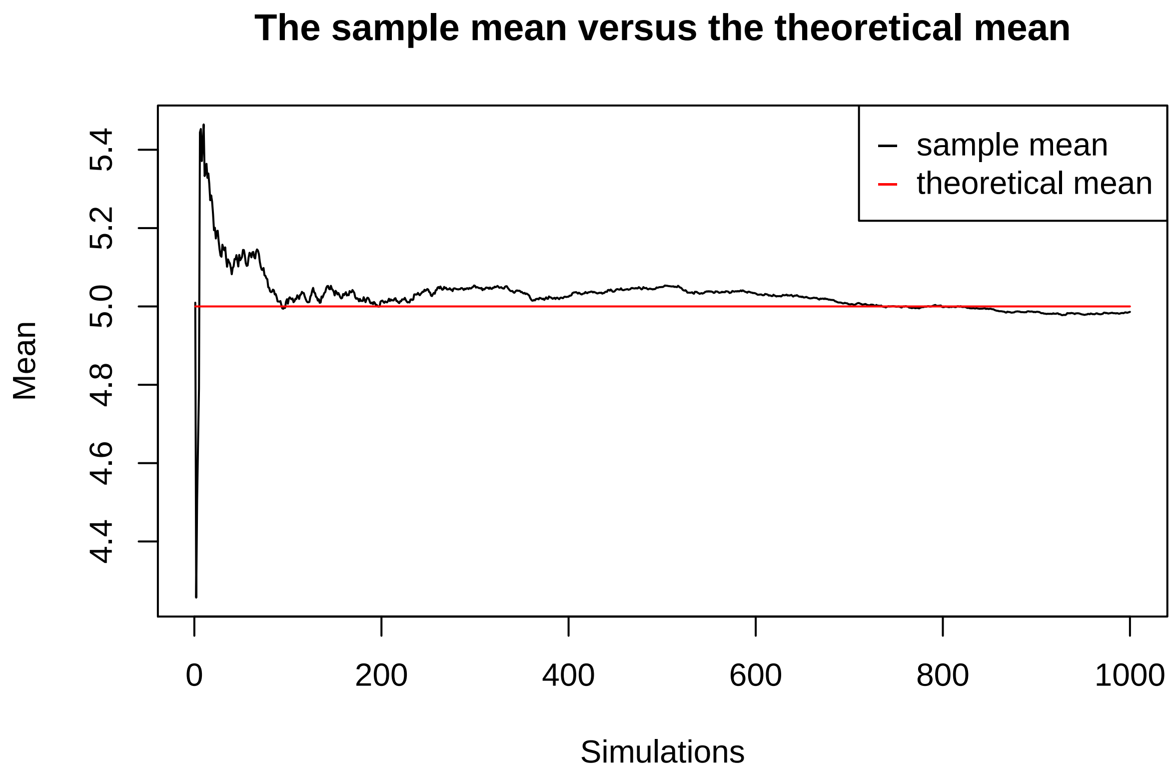
This project concentrates on the law of large numbers and the central limit theorem both illustrated by the exponential distribution with rate parameter set here to 0.2. We will run 1000 simulations and show that the mean and variance approach their theoretical values for adequately large number of simulations. The first step is to set up the driving variables.

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

Sample Mean versus Theoretical Mean

Here the theoretical mean is 5 (1/0.2), we will generate a vector (*means*) of 1000 means of 40 exponentials then cumulate them and calculate the cumulative averages. Then we plot the sample means and the theoretical mean. As we can see both lines converge as the sample count grows.

```
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 = "red")

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

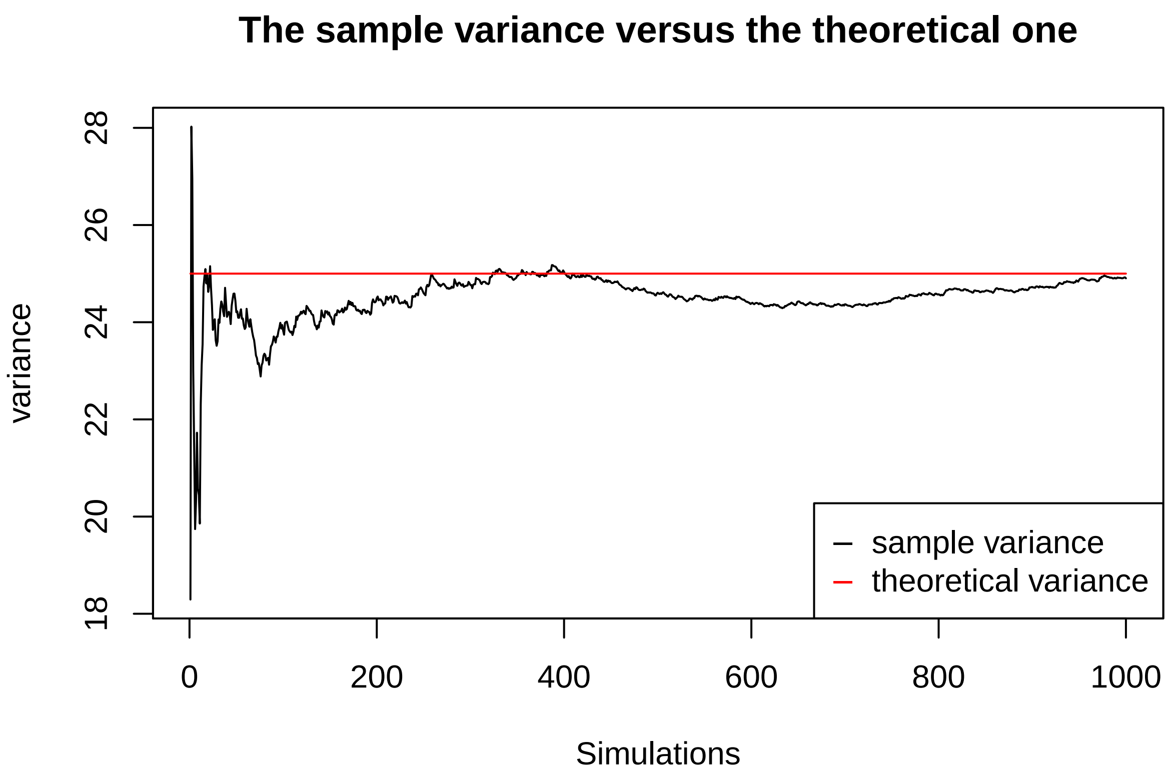


Sample Variance versus Theoretical Variance

Here the theoretical variance is 25 (1/0.04), we will generate a vector (*vars*) of 1000 variances of 40 exponentials then cumulate them and calculate the cumulative averages. Then we plot the sample variance and the theoretical one. As we can see both lines converge as the sample count grows.

```
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 one",
     xlab = "Simulations",
     ylab = "variance",
     col = "black")

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

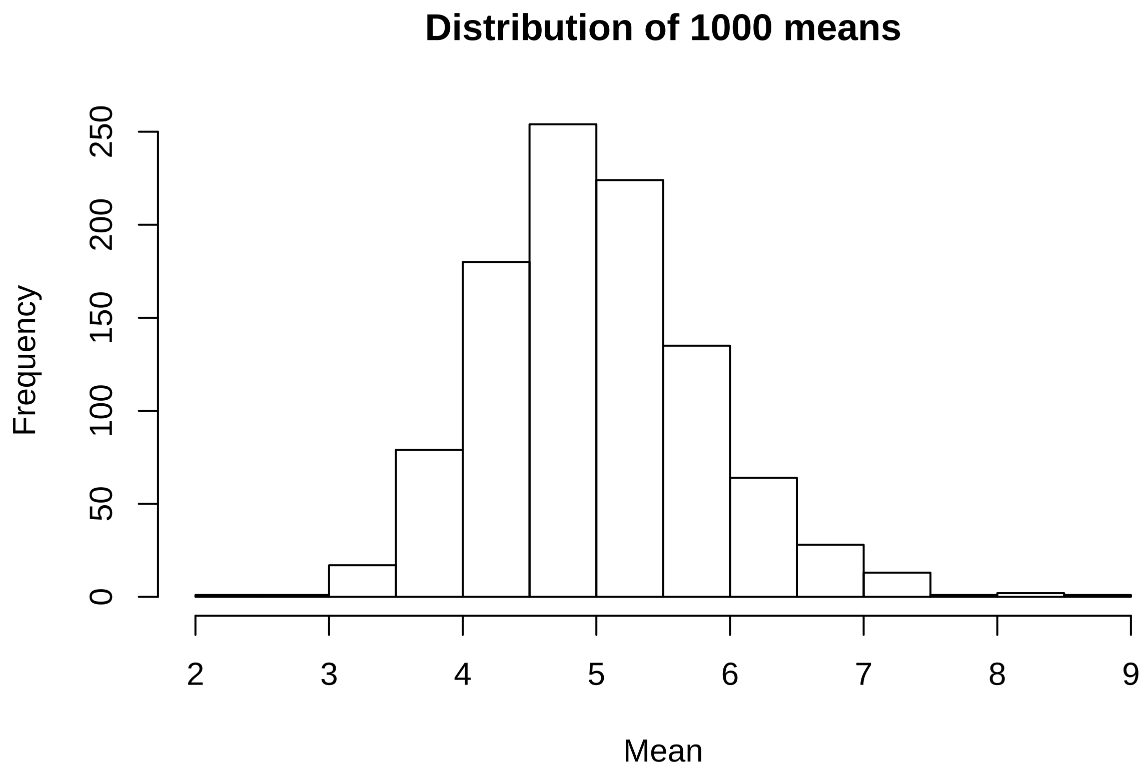
```
       col = "red")

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



Distribution

Finally we want to show that with accordance to the central limit theorem as the sample size increases the distribution becomes that of a standard normal. We use the previously generated mean vector (*means*) to draw a histogram that resembles the standard normal one, as we can see the distribution "concentrates" around the theoretical mean of 5 (1/0.2).

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


Distribution of 1000 means
Frequency
0
50
100
150
200
250
2
3
4
5
6
7
8
9
Mean