Exponential distribution and the Central Limit Theorem

Hugo Janssen

15 sep, 2015

Overview

This project investigates the exponential distribution in R and compares it with the Central Limit Theorem. To this aim, the **mean** and **variance** of a simulated exponential distribution are compared to the theoretical mean and variance of the normal distribution.

Simulations

The sample distribution is obtained by running 1000 simulations of averages of 40 exponentials. The value of the rate parameter lambda for the distribution is set to 0.2 for all simulations.

```
nosim <- 1000
n <- 40
lambda <- 0.2
sim <- apply(matrix(rexp(n * nosim, rate = lambda), nosim), 1, mean)
```

Sample Mean versus Theoretical Mean

The theoretical mean of the exponential distribution with given rate parameter lambda is:

$$E[X] = \mu = \frac{1}{\lambda} = \frac{1}{0.2} = 5.0$$

The simulated sample mean is: **5.04**

Observation: the sample distribution seems to be centered around the theoretical mean.

Sample Variance versus Theoretical Variance

For exponential distributions, the standard deviation σ is equal to the mean μ:

$$\sigma = \mu = 5.0$$

We also know that the variance of the sample mean is $\frac{\sigma^2}{n}$ and that its logical estimate is $\frac{s^2}{n}$. Therefore, the theoretical variance of the exponential distribution with the given rate parameter lambda is:

$$Var[X] = \frac{\sigma^2}{n} = \frac{\frac{1}{\lambda^2}}{n} = \frac{25}{40} = 0.625$$

The simulated variance is: **0.66**

Observation: the simulated variance is **slightly higher** than the theoretical variance. The intuition behind this observation is that the more simulations are performed, the more data is available and the more the sample variance will converge to the theoretical variance. So, running more simulations would drive the variance even closer to the theoretical value.

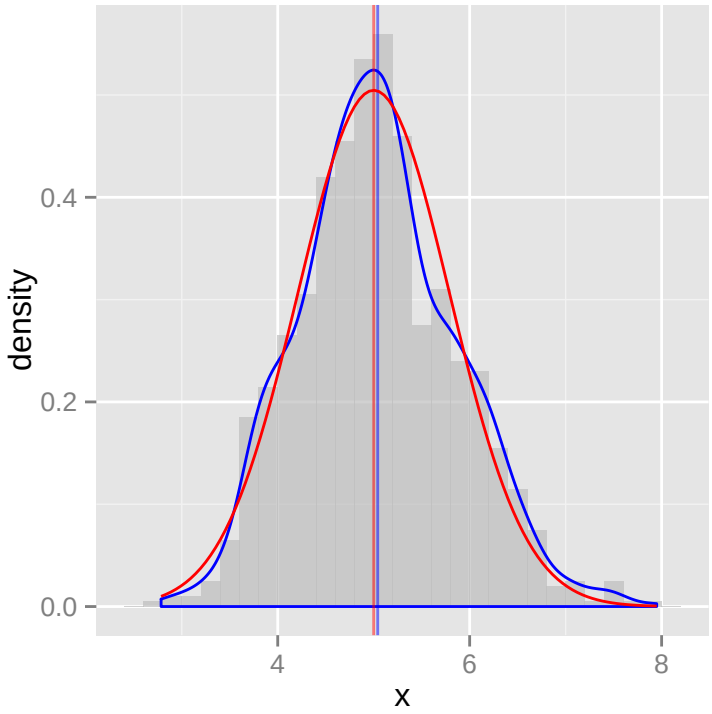
Distribution

The figure below shows an overlay of a normal distribution (red) on top of the histogram and density plot of the sample distribution (blue). The sample mean and theoretical mean are shown as vertical lines.

Observations:

- The shapes of the red and blue bell curve look very similar.
- The means of both distributions are centered around $x = 5$ and are very close together.
- The height of the simulated distribution is slightly higher, due to its higher variance.

From these observations, one can conclude that the distribution of the simultated sample is approximately normal.



Conclusion

The sample mean is centered around the theoretical mean 5. The sample variance is slightly higher than the theoretical variance of 0.62. This is explained due to the fact that the more data is gathered, the closer the variance will approximate the the theoretical variance. From the analyses in this report one can conclude that the simulated sample distribution is approximately normally distributed.