Exponential Distribution and Central Limit Theorem

CJ

Thursday, January 08, 2015

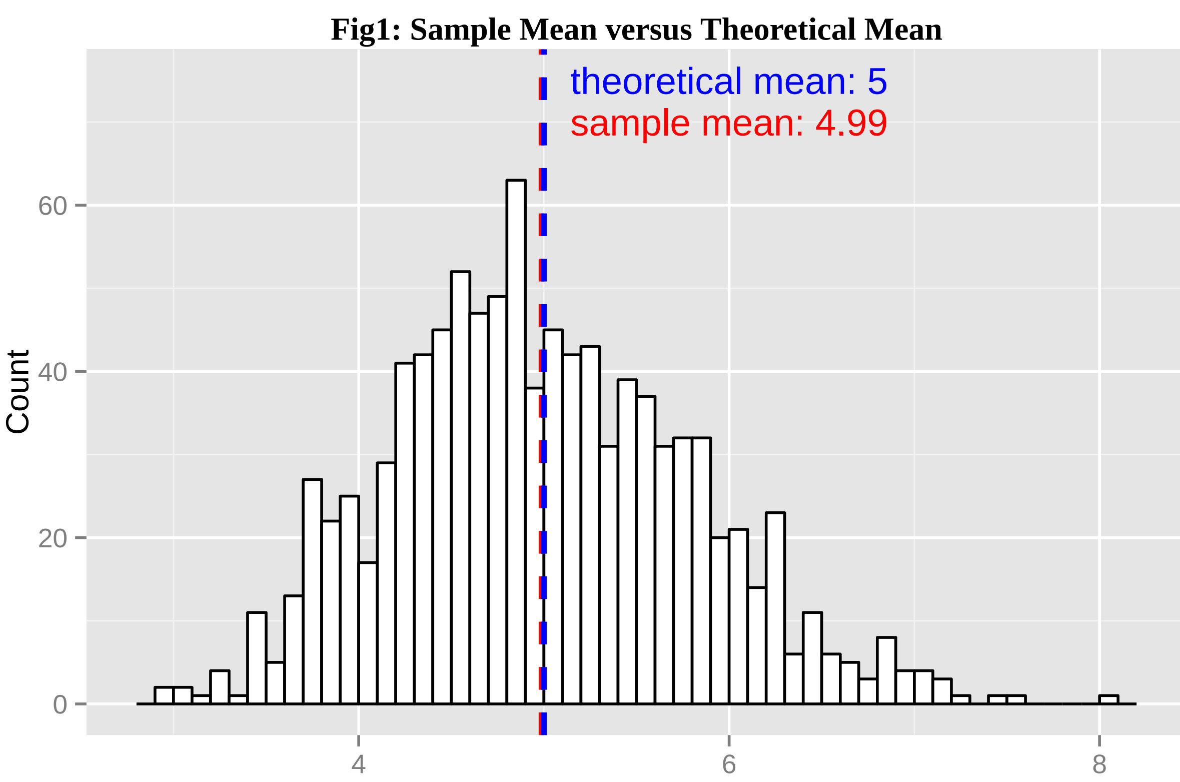
Overview

In this project we investigate the exponential distribution and compare it with the Central Limit Theorem. We sample 40 exponentials and calculate their mean and standard deviation. After a thousand of simulations, we compare the average of such mean and standard deviation with its theoretical values, whose value should be both 1/lambda for the mean and the standard deviation. Finally, the distributions of the sample mean and standard deviation are normalized and compared with the standard normal.

Simulation

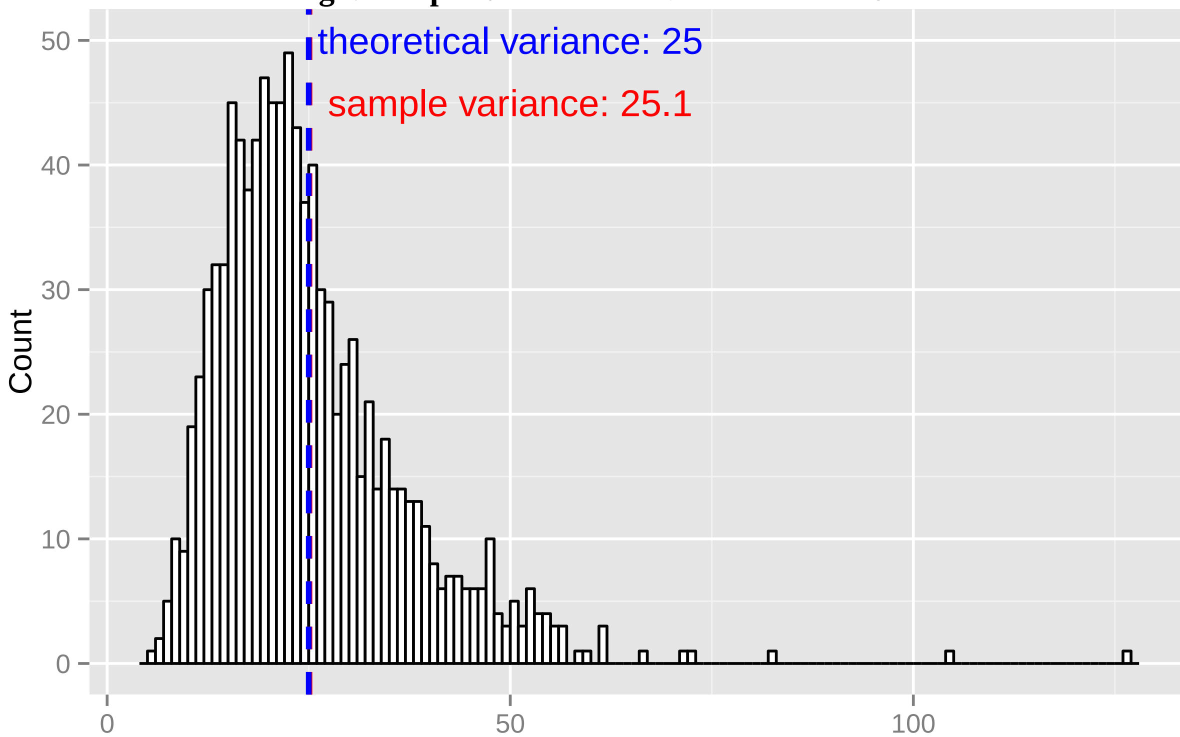
Firstly, calculate the theoretical values of exponential mean and standard deviation. When `lambda`= 0.2, both mean and standard deviation equal to `5`. We generate 1000 averages and variances of 40 random exponentials using fucntion `exp` and calculate their means. Both values are close to their theoretical values as shown in `Fig1` and `Fig2`. In `Fig1`, the center of sample means is 4.99 with a standard error 0.02558013. In `Fig2`, the center of sample variance is 25.1, which is close to its theoretical values,25.

Plot the histogram of a thousand of simulated means and compare the average of these means with its theoretical value.



Plot the histogram of a thousand of simulated variances and compare the average of these variances with its theoretical value.

Fig2: Sample Variance versus Theoretical Variance

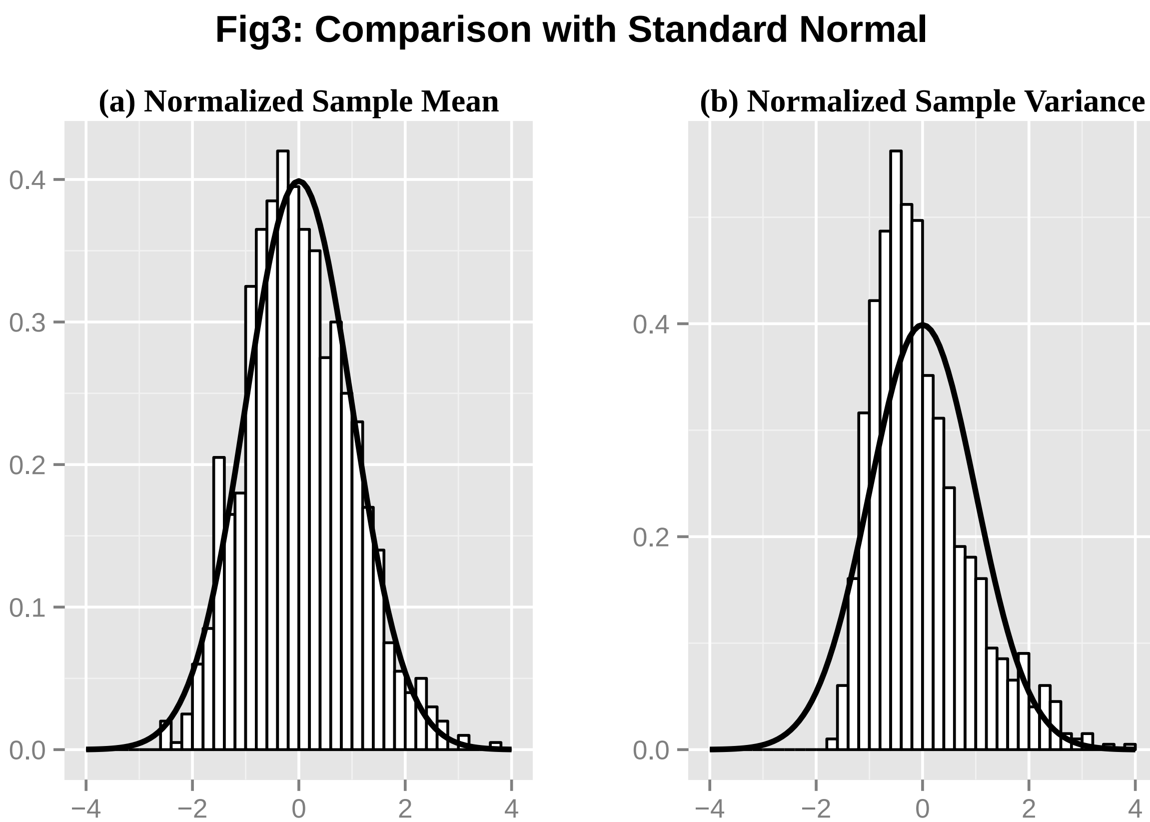


Comparison with the Central Limit Theorem

The **law of large numbers (LLN)** states that the average of the results obtained from a large number of trials should be close to the expected value. As shown in both `Fig1` and `Fig2`, mean and variance are close to their theoretical values. Futher more, the `Central Limit Theorem` (CLT) states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases.

Nomarlization Here we normalize both sample mean and variance and compare with the standard normal distribution.

Now plot both distributions of sample mean and variance and overlay with the standard normal distribution.



As we can see in `Fig3`, both `(a)` and `(b)` are close to standard normal distribution, which center at 0 and most of the data are within 3 times σ. But variance distribution is more skewing to the right. This might due to the limitation of the CLT that doesn't give the exact sample size.