Exponential Distribution: Simulated and Theoretical Properties

Cha Li

06/20/2015

Overview

In this report, we'll compare simulated means and variances, created through simulation, with their theoretical values. Numerous simulations will be executed with each instance performing a random draw and calculating the sample mean and variance. The objective is to show that the mean and variance of the simulated values follow the Law of Large Numbers (LLN) and the Cental Limit Theorem (CLT).

Simulations

I'll be illustrating that the mean of simulated means and the mean of simulated variances converge to the theoretical values of the mean and variance of the exponential distribution. The simulation consists of 1000 instances where 40 random values are drawn from the target distribution and then their mean and variance calculated. Below is the code to accomplish this task.

```
lambda = 0.2
trials = 1000
n = 40

# the theorerical values of the mean and variance of exp. dist with lambda=.2
true_mean = 1 / lambda
true_variance = (1 / lambda)^2

# set a constant seed so rerunning this code produces the same simulations
set.seed(12345)

# run 1000 simulations and for each simulation calculate the mean and variance of 40 exponential
# random variables.
sim_means = replicate(trials, mean(rexp(n, lambda)))
sim_variances = replicate(trials, var(rexp(n, lambda)))

# put everything into a dataframe for easy plotting
simulation_data = data.frame(means=sim_means, variances=sim_variances)
```
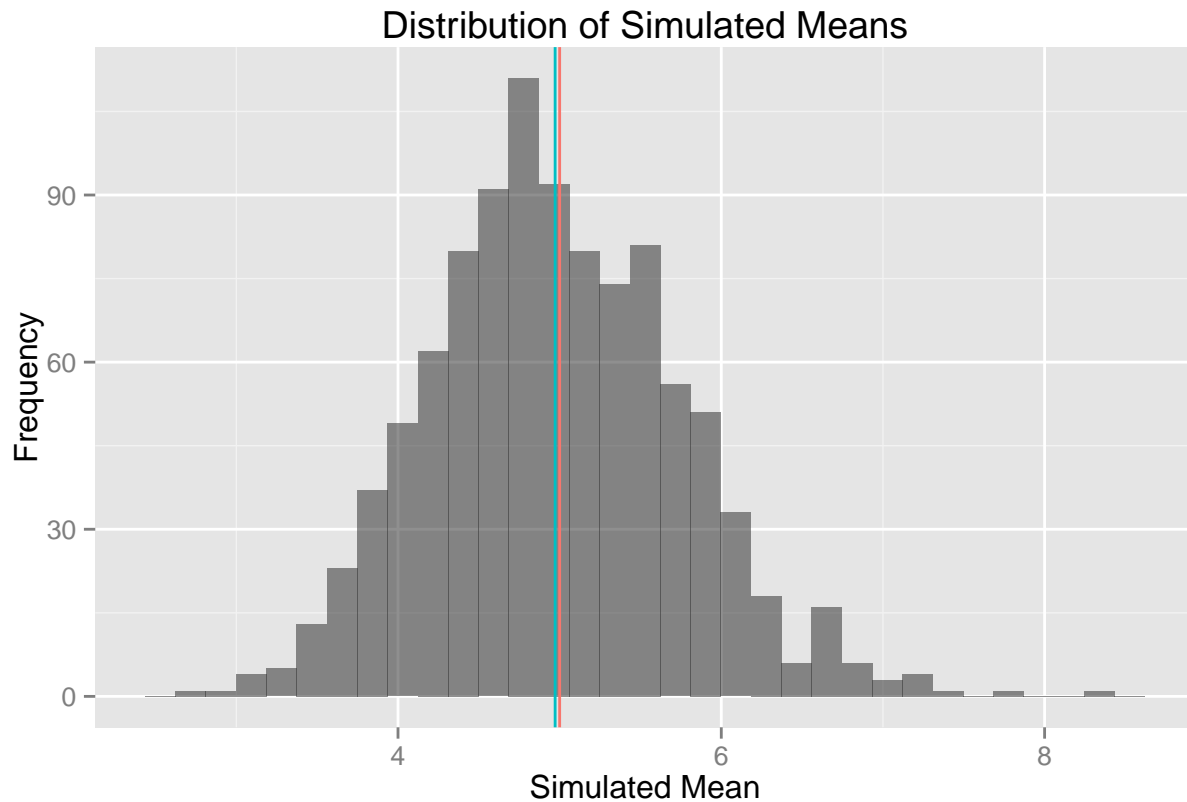
Simulated Means versus Theoretical Mean

The following plot shows the distribution of means generated by each simulation of the exponential distribution. The mean of the simulations is represented by the red line (4.971972) and the theoretical mean is represented by the blue line (5).

```
ggplot(simulation_data, aes(x=means, alpha=.5)) +
    geom_histogram() +
    geom_vline(aes(xintercept=mean(means), colour="red")) +
    geom_vline(aes(xintercept=true_mean, colour="blue", linetype="dashed")) +
    labs(x="Simulated Mean", y="Frequency") +
    ggtitle("Distribution of Simulated Means") +
    guides(alpha=F, fill=F)
```

Distribution of Simulated Means

Simulated Variance versus Theoretical Variance

Similar to before, the following graph shows the distribution of simulated variances, their mean 25.4214007 (red line), and the theoretical variance, 25 (blue line). One thing to notice is that the distribution is not quite normal and seems to be skewed slightly. This is a result of the simulations being too small as increasing n to 100 creates a more normal-like distribution.

```
ggplot(simulation_data, aes(x=variances, alpha=.5)) +
    geom_histogram() +
    geom_vline(aes(xintercept=true_variance, colour="red")) +
    geom_vline(aes(xintercept=mean(variances), colour="blue", linetype="dashed")) +
    labs(x="Simulated Variance", y="Frequency") +
    ggtitle("Distribution of Simulated Variances") +
    guides(alpha=F, fill=F)
```

Distribution of Simulated Variances
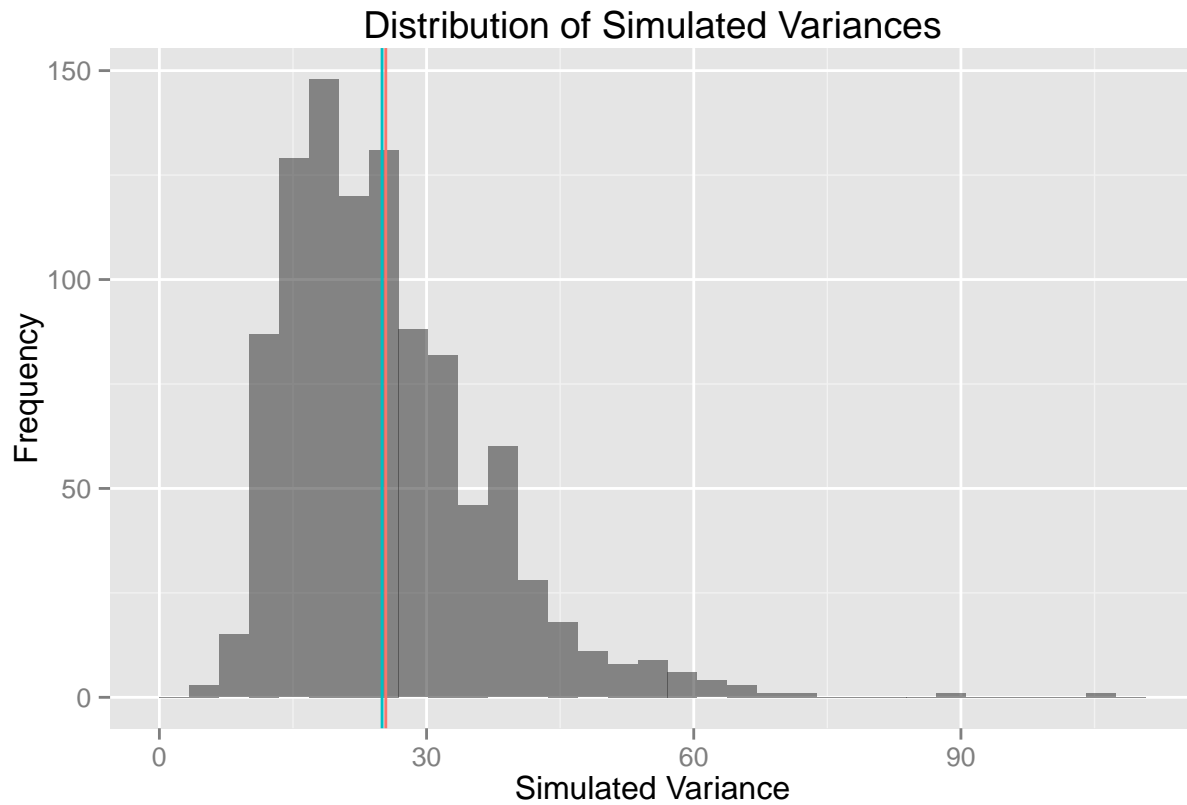
Conclusion

This exercise illustrated the LLN and the CLT. The former states that sample means, of iid random variables, converge towards their population mean while the latter states that the distribution of sample means converges towards a normal distribution. Our random variables were independent and identically distributed so the expected distribution of their means is normal, which the above plots show.