Test CLT with Exponentials

JeremiShane

11/24/2017

Summary

In testing the Central Limit Theorem (CLT) with means from Exponential distributions, the evidence shows the CLT to hold true.

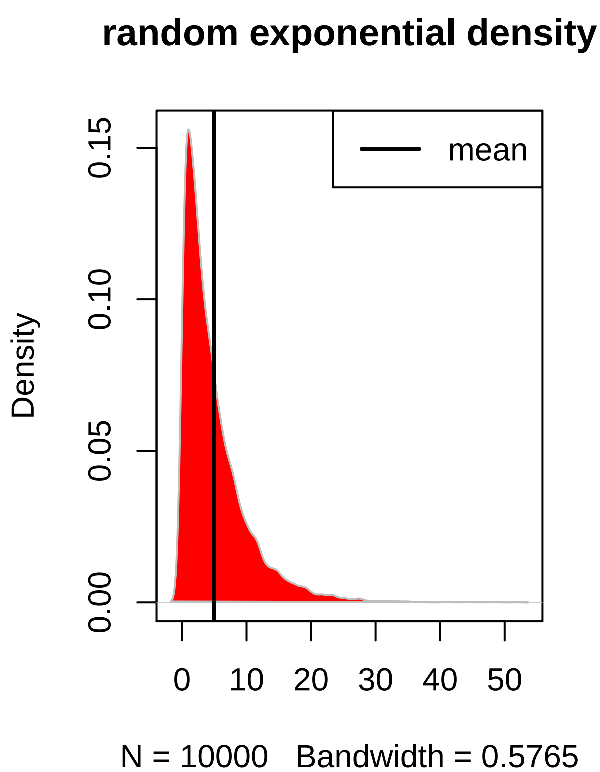
Overview

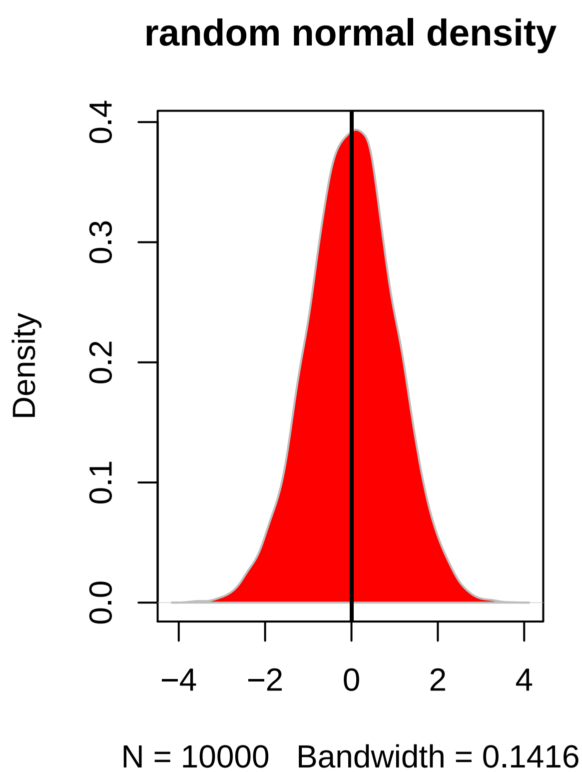
The central limit theorem (CLT) establishes that, in most situations, when independent random variables are added, their properly normalized sum tends toward a normal distribution (informally a "bell curve") even if the original variables themselves are not normally distributed. We will evaluate this theory with the averages/means of simulated exponentials.

From Wikipedia, the exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a homogeneous Poisson process. The exponential distribution describes the time for a continuous process to change state

This analysis is organized as follows:
1. Simulations
2. Sample Mean versus Theoretical Mean
3. Sample Variance versus Theoretical Variance
4. Distribution

First, let's have a quick look at an exponential distribution and a normal distribition.





We can see from these two simple histograms that an exponential distribution does not look normal. To carry out our experiment, we will look at the means of multiple exponential distributions. The Central limit theorem tells us that even though the exponential distribution is not normal, the means from large enough number of exponential distributions should approach normal.

Simulation of Exponentials

Below is the R code for simulating the exponential distribution with size = 40 and 100. In each case we perform 1000 simulations and take the mean of each simulation.

```
set.seed(23)  ## set seed for reproducability
## run 1000 simulations
nsims <- 1000
n <- 40 ## our default number of exponentials
forty_mean = sapply(1:nsims, function(x) {mean(rexp(n, lambda))})
hundred_mean = sapply(1:nsims, function(x) {mean(rexp(100, lambda))})
```

Simulated Mean vs. Theoretical Mean

With rate lambda = .2, an exponential distribution should have mean = 1/lambda = 5, and standard deviation = 1/lambda as well. We would expect the means from 1000 of these exponential distributions to be a normal distribution around 5, from the Central Limit Theorem.

```
##   theoreticalMean fortyMean
## 1               5   5.01425
```

These calculations show that the theoretical mean and the mean from each of our simulations are very close.

Simulated Variance vs. Theoretical Variance

Variance…
- of an exponential distribution is 1/lambda^2
- for a normal distribution is sd^2

```
##   theoreticalVariance simulatedVariance
## 1               0.625         0.6862199
```

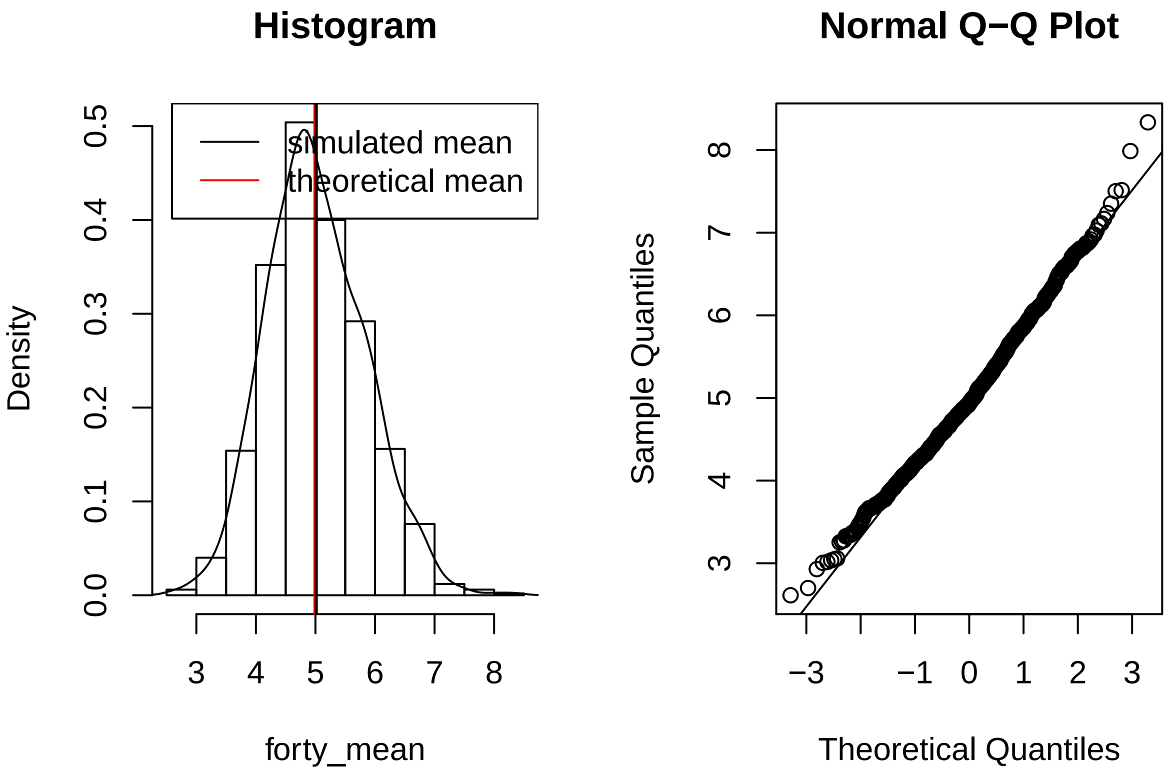
We see here that the theoretical variance and calculated variance from our simulated distribution is very close.

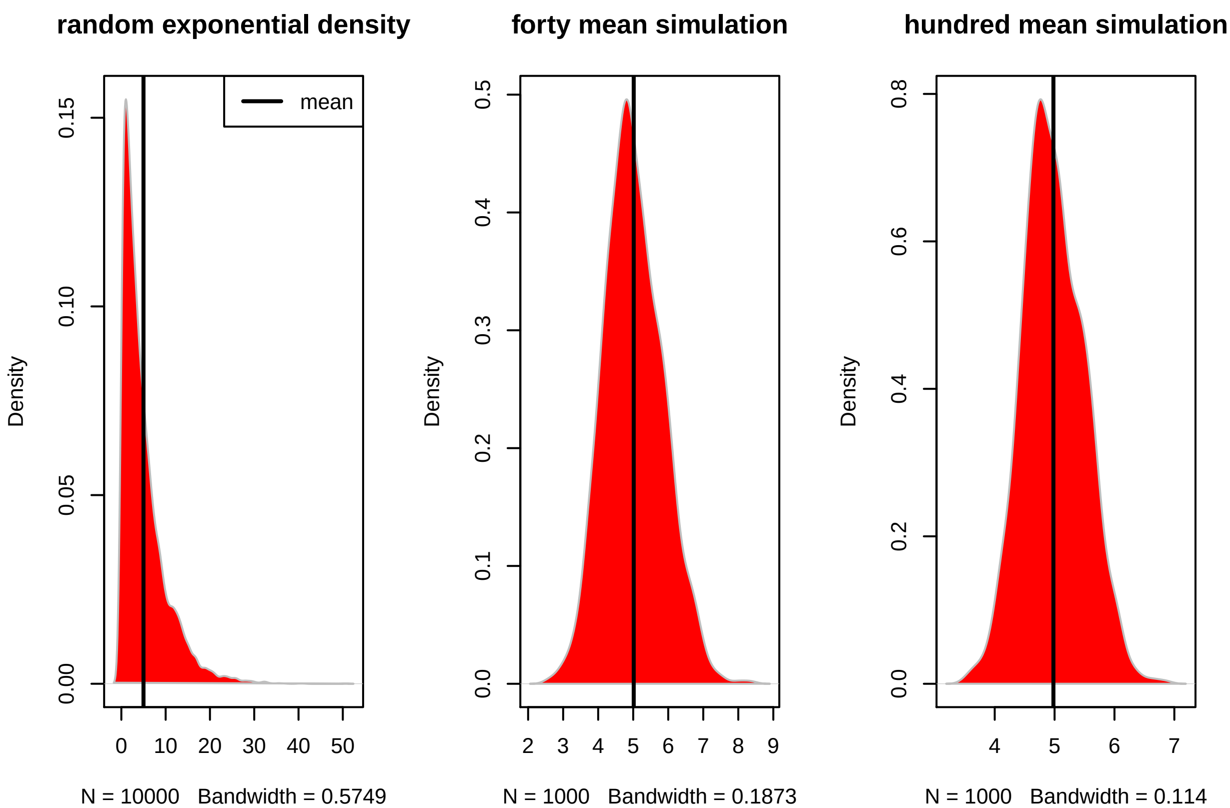
Distribution

Let's look at a histogram distribution and a qq plot of our forty_mean simulation. Both of these indicate the Central Limit Theorem is holding true. The distribution is forming a normal bell curve around 5.

Then, let's refer back to our exponential distribution of size = 1000, and look at it alongside our distribution of the means of our simulations as n gets larger.

The larger n gets, the closer to normal it gets. It is safe to say that our experiment shows clearly the Central Limit Theorem holding true.


Histogram
Density
forty_mean
simulated mean
theoretical mean
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles


random exponential density
forty mean simulation
hundred mean simulation
Density
mean
N = 10000   Bandwidth = 0.5749
N = 1000   Bandwidth = 0.1873
N = 1000   Bandwidth = 0.114