Statistical Inference Course Project

Han Yunxi

16 January 2015

Investigation of Exponential Distribution in R vs Central Limit Theorem

This report investigates the distribution of averages of 40 numbers sampled from the exponential distribution with lambda = 0.2 .

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. We set lambda = 0.2 for all of the simulations.

Repeated over 1000 simulations, we examine the simulated results of the mean and variances versus the expected theoretical results.

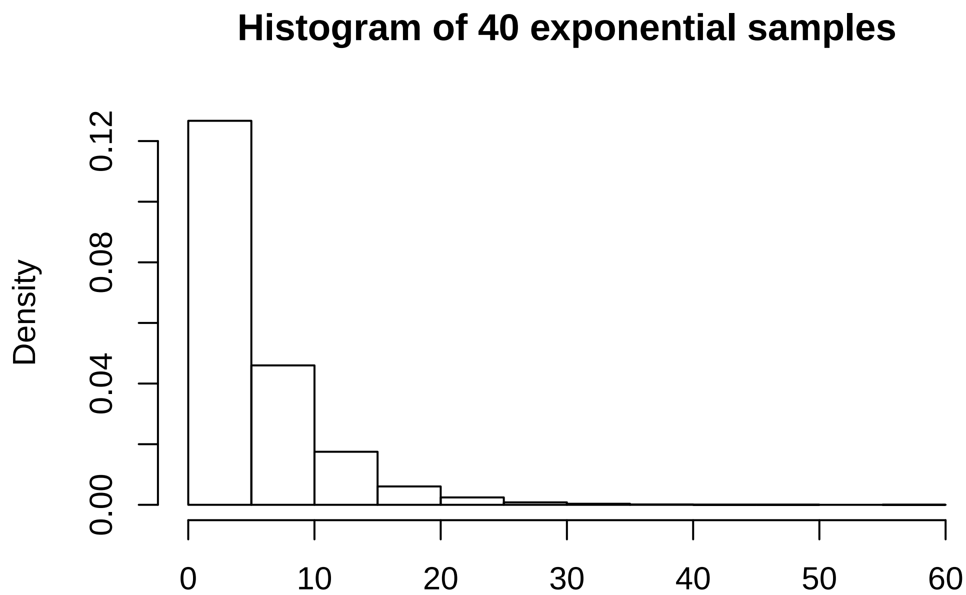
Assumptions are:

```
set.seed(1)

lambda <- 0.2       # Rate parameter
mean <- 1/ lambda
sd <- 1 / lambda    # Standard deviation

sample_size <- 40
```

Simulated Results

Let's first look at the shape of the exponential function

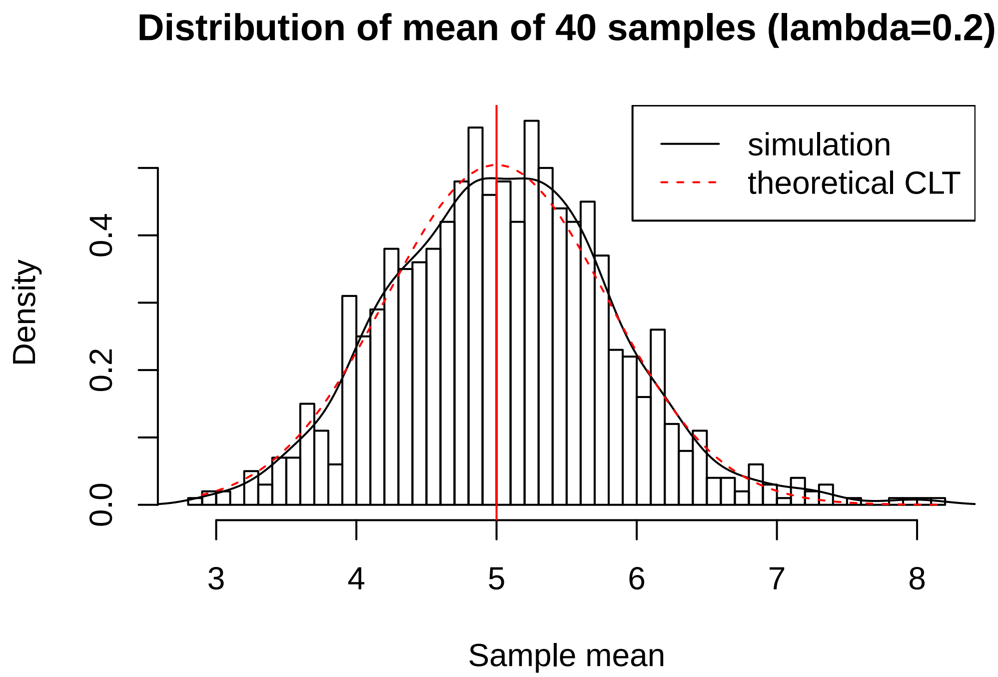
```
shape <- rexp(sample_size * 1000 , rate = lambda)
hist(shape, prob = TRUE, breaks = 20,
     main = "Histogram of 40 exponential samples" ,
     xlab ="")
```



Now, let's do 1000 simulations of 40 randomly selected exponential functions

```
# 1000 simulations of 40 random sample size
xbar <- rep(0,1000)
for (i in 1:1000) { xbar[i] <- mean (rexp(sample_size, rate = lambda))}
```

We will now model the sample population means ($\bar{x}$) distribution below. The distribution of the sample population means is then seen to be following a Normal distribution. The theoretical Normal curve and the theoretical mean, μ, is also overlaid. Diagram as follows:



Results

For the sample population, the mean and variance of the sample population ($\bar{X}$) is calculated to be $\mu = 5.05$ and $\sigma^2 = 0.63$ respectively.

For the theoretical CLT, the mean and variance of exponential distribution are given as 1/lambda. Hence the expected results are mean ,$\mu = 5$ and variance, $\sigma^2/n = 1/(\lambda^2 n) = 0.625$

Conclusion

Now, let compare the results of the simulation against the Central Limit Theorem.

1. CLT states that averages are approximately normal, with distributions centered at the population mean. From the histogram, we see that property to be **true**.
2. CLT states that the standard deviation is equal to the standard error of the mean. From the results, where sample variance is approximately population variance, that is also **true**.

However, as CLT gives no guarantee that sample size (n) is large enough, we run more simulations (1000) to ensure that the Law of Large Numbers requirement is met.

References

- http://msenux.redwoods.edu/math/R/CentralLimit2.php