Central Limit Theorem vis Exponential Distribution

Nitin Aggarwal

September 26, 2015

Overview

In this analysis, central limit theorem is tested for exponential distribution using the simulated data. The averages of forty exponentials are investigated, and then the means and variances of the mean are compared with the theoretical values of the distribution. It is also showed that the distribution of the mean is approximately normal.

Simulation

We will use the function `rexp` the base stats package to produce the simulation for the exponential distribution. We want to create 1000 averages of 40 exponentials, which we do as follows:

First we initialize a vector of length `1000` to store the mean of the 40 simulated exponentials.

```
exp.means<-rep(0,1000)
```

The parameter `lambda` for the exponential distribution to be considered should be 0.2, and the mean is taken over 40 observations from exponentials, which we define as variables as follows for future use:

```
lambda = 0.2
n = 40
```

Then, we run a `for` loop to creat and store the mean of simulations into this variable.

```
set.seed(12121)
for(i in 1:1000){
  exp.means[i]<-mean(rexp(n=40, rate = lambda))
}
```

We used two functions `rexp` and `mean` above. The first function `rexp` takes two arguments: `n` the number of observations to be drawn from exponential distribution and `rate` which is just lambda for the exponential distribution. The second function `mean` from the base package is the generic function to produce the arithmetic mean.

Sample Mean versus Theoretical Mean

Theoretically, mean of this distribution should be same as the mean of the exponential distribution used which is 1/lambda. With `lambda = 0.2` the mean of the distribution would be:

```
distribution.mean <- 1 / lambda
theoretical.mean <- distribution.mean
theoretical.mean
```

```
## [1] 5
```

The sample mean is to be calculated from the sample as follows:

```r
sample.mean <- mean(exp.means)
sample.mean
```

```
## [1] 4.994445
```

The value of `theoretical.mean`, 5, is a close approximation to the value of `sample.mean`, 4.9944446. This validates the part of central limit theorem which asserts that the mean of samples from a distribution converges to the mean of the distribution itself.

Sample Variance versus Theoretical Variance

Theory suggests that the variance should be a the variance of the original distribution divided by the number of observation. So, the theoretical variance should be variance of the distribution, `1/lambda^2`, divided by the number of observation used to calculate the mean.

```r
distribution.varince <- 1 / (lambda ^ 2)
theoretical.variance <- distribution.varince / n
theoretical.variance
```

```
## [1] 0.625
```

The sample variance can be calculated by the function `var` from base `stats` package as follows:

```r
sample.variance <- var(exp.means)
sample.variance
```

```
## [1] 0.6034557
```

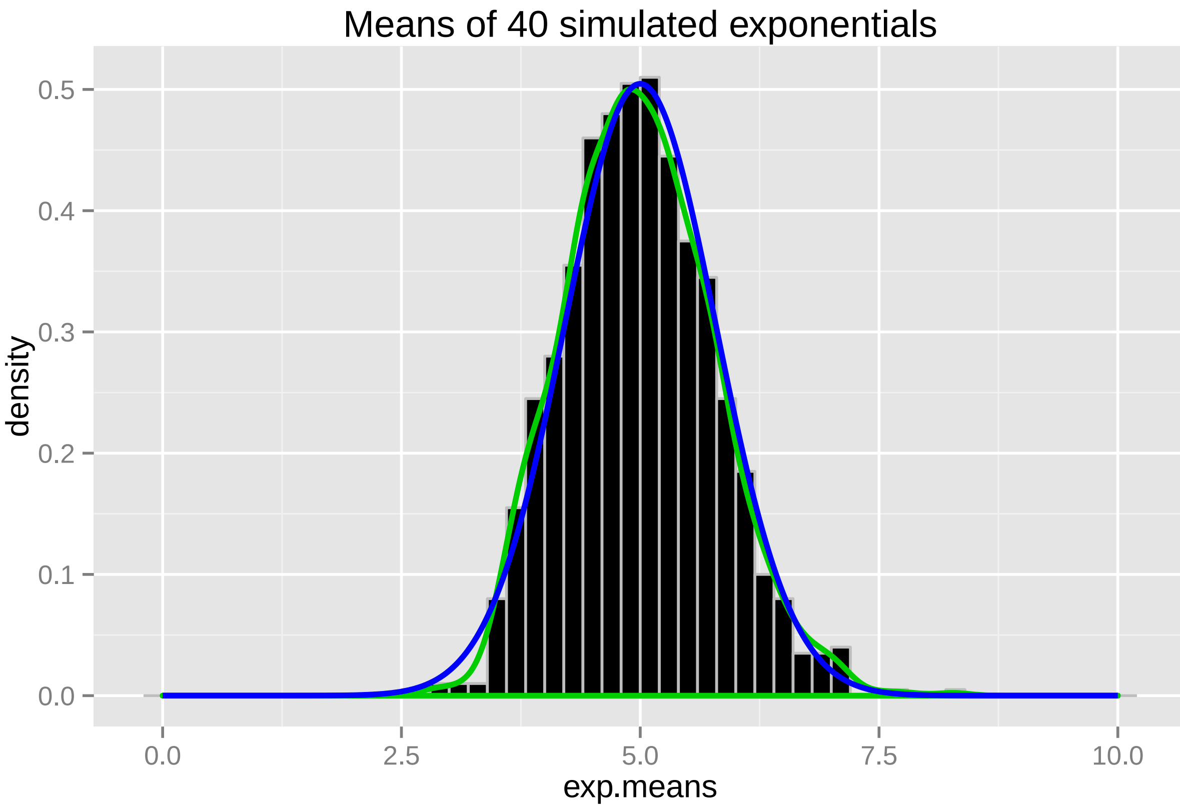
The closeness of theoretical variance, 0.625, to the sample variance, 0.6034557, suggests the argument of central limit theorem which suggests that the variance of mean of observations from a distribution converges to the variance of distribution divided by the number of observations in the mean.

Distribution

Let us plot the histogram of the `exp.means` and get the plot of density estimate of its distribution using `ggplot2` command called `geom_density`. Let us also plot simultaneously the density plot of the normal distribution with mean `theoretical.mean` and `theoretical.variance` found above.

```r
library(ggplot2)
g<-ggplot(data = data.frame(exp.means=exp.means), aes(x =exp.means))+
  geom_histogram(aes(y=..density..), binwidth = lambda, color = 8, fill = 1)+
  geom_density(color = 3, size =1) + ggtitle("Means of 40 simulated exponentials")

x<-seq(0,10, length = 10000)
y<-dnorm(x, mean = theoretical.mean, sd = sqrt(theoretical.variance))
g + geom_line(data = data.frame(x = x, y = y), aes(x = x, y = y), color = 4, size = 1)
```



Closeness of the density plot of normal distribution(in blue) with given theoretical parameters and the predicted density function for `exp.means` clearly suggests that the argument of central limit theorem which asserts that the distribution of mean of observations from a distribution with mean μ and variance σ^2 approaches in probability and almost surely to the normal distribution with mean μ and variance σ^2/n, with n the number of observations in the mean.