Comparison of the exponential distribution and the central limit theorem

Gbenro Selere

August 22, 2015

Overview

In this project we will be investigating the exponential distribution in R and comparing it with the Central Limit Theorem. The exponential distribution can be simulated in R with `rexp(n, lambda)` where λ is the rate parameter. The mean(μ) of exponential distribution is $\frac{1}{/}\lambda$ and the standard deviation(σ) is also $\frac{1}{/}\lambda$.

We would also be Illustrating via simulation and associated explanatory text the properties of the distribution of the mean(μ) of 40 exponentials and λ would be set to 0.2 for all the simulations. The outcomes from this report are to: 1. Show the sample mean and compare it to the theoretical mean of the distribution. 2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. 3. Show that the distribution is approximately normal.

Simulations

Comparison of the sample mean with the theoretical mean of the distribution

In the following we will draw 1000 samples of size 40 from an $Exp(\frac{1}{0.2}, \frac{1}{0.2})$ distribution. For each of the 1000 samples we will calculate the mean. Theoretically, this the same as drawing a single sample of size 1000 from the corresponding sampling distribution with $N(\frac{1}{0.2}, \frac{\frac{1}{0.2}}{\sqrt{40}})$.

According to the CLT we would expect that each single mean of those 1000 means is already approximately $\frac{1}{\lambda} = \frac{1}{0.2} = 5$. Since we now caluculate the mean of 1000 sampled means we expect the output to be very close to 5.

```
set.seed(133233)
lambda <- 0.2
nobserv <- 1000
sample_size <- 40

exp_sample_means <- c()

for(i in 1:nobserv) {
  exp_sample_means <- c(exp_sample_means, mean(rexp(sample_size, lambda)))
}
mean(exp_sample_means)
```

```
## [1] 4.999618
```

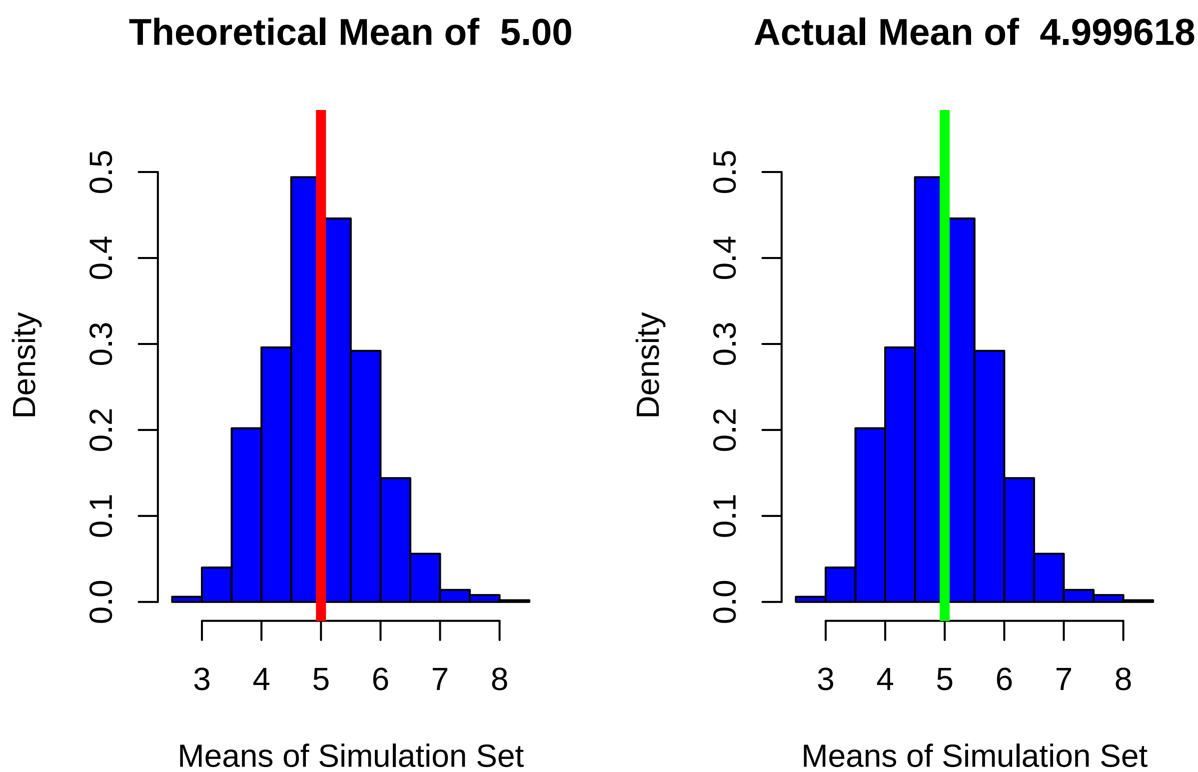
By rounding off the number we have a value of 5 which is very close to the mean

```r
round(mean(exp_sample_means), 2)
```

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

which is very close to the mean of the theoretical distribution namely $\mu = \frac{1}{0.2} = 5$.

Plotting the comparison



From the plot above, it can be seen that the histograms look almost identical

Comparison of the sample variance with the theoretical variance of the distribution

According to the CLT we would expect that the variance of the sample of the 1000 means is approximately $\frac{\frac{1}{0.2^2}}{40} = 0.625$.

From calculating the actuall variance it can be seen that the theoretical and actual variance from the distribution are close.

```r
actual_var <- var(exp_sample_means)

actual_var
```

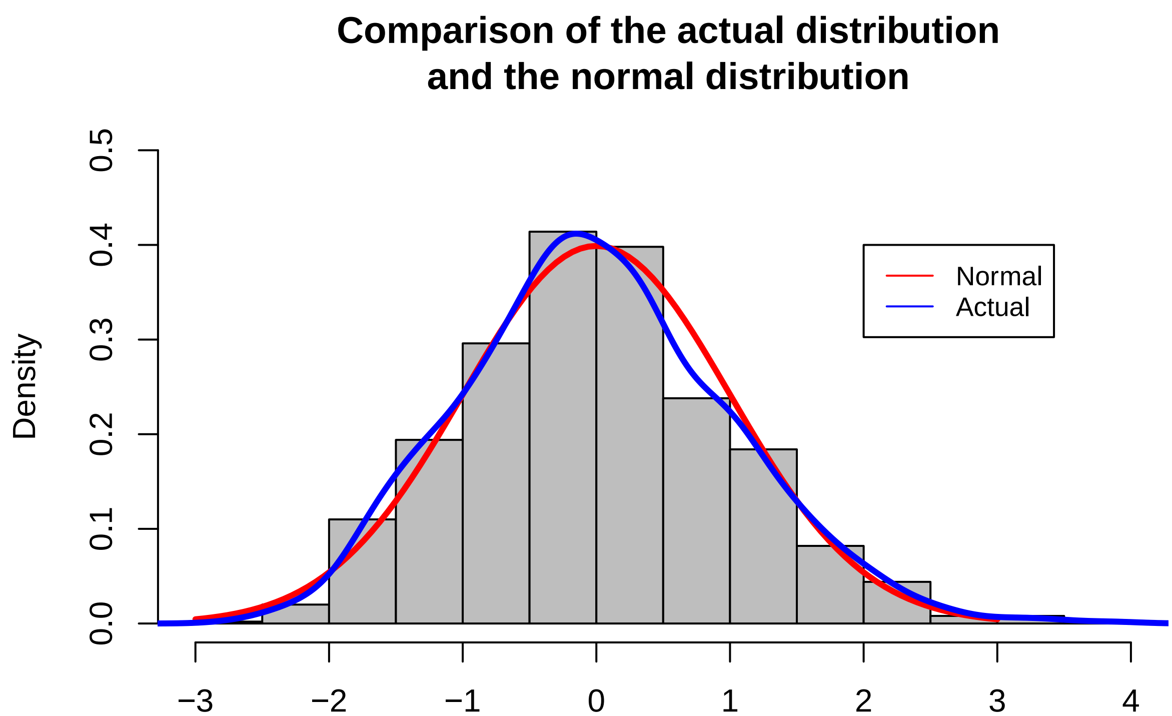
```
## [1] 0.6833211
```

This means that the sample data is spread-out to proportion as to what is expected.

Showing that the sample distribution is approximately normal

Finally, if we plotted the a normal curve and overlayed to the sample set along with the actual distribution curve, we can see very clearly that they are very similar; in turn, the sample set can be defined as normal, in terms of distribution.

```
par(mfrow=c(1,1))
hist(scale(exp_sample_means),probability=T,main=" Comparison of the actual distribution\n and the normal
curve(dnorm(x,0,1),-3,3, col='red',add=TRUE, lwd = 3) # normal distribution
lines(density(scale(exp_sample_means)),col='blue', lwd = 3) # actual distribution
legend(2,0.4,c('Normal','Actual'),cex=0.8,col=c('red','blue'),lty=1)
```



Appendix

The full source code to this report can be found here