Case study on Exponential Distribution

Aman Bhagat

21/04/2020

Overview

In this report we aim to investigate the exponential distribution in R and compare it with the Central Limit Theorem. We have generated a thousand simulation with the mean of exponential distribution which is 1/lambda and standard deviation 1/lambda. For all the simulation we have set the lambda = 0.2. Basically we will be investigating the distribution of averages of 40 exponentials.

Simulations

We are trying to generate a thousand exponentials where we kept lambda as 0.2. In R it can be achieved by using `rexp()`.After that we converted that thousand as a matrix so that each row consists of 40 random exponentials and there are 1000 rows in total which gives a total of 40000 exponentials to investigate.

```
set.seed(111)
lambda <- 0.2
n <- 40
sample.size <- 1000
sample.exponentials <- matrix(rexp(40000,0.2),sample.size,n)
sample.exponentials[1:5,1:5]
```

```
##           [,1]      [,2]       [,3]      [,4]     [,5]
## [1,] 0.9298128  1.364895 15.6583781 1.8972460 1.837269
## [2,] 2.2648112  1.458486  7.2567785 0.3423393 3.075893
## [3,] 5.8741760  2.060109  0.5767754 5.6395793 1.760570
## [4,] 0.1492383  6.476079  0.5697798 1.9120896 2.322804
## [5,] 6.0190002 11.305384  0.7265282 1.3835789 7.021825
```

Here we can see few simulated exponentials which follows exponential distribution. The total number of rows is 1000 and the total number of columns in a row is 40

Sample Mean versus Theoritical Variance

Now we will find the average of 40 exponentials that is we will take average of each row and then investigate further.

```
sample.mean <- apply(sample.exponentials, 1, mean)
```

Now the sample mean from the average of each 40 averaged exponential is

```
averaged.mean <- mean(sample.mean)
averaged.mean
```

```
## [1] 5.02562
```

While the mean from the original simulated can be calculated from the lambda which was 0.2, which gives us:

```
theo.mean <- 1/lambda
theo.mean
```

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

If we compare the simulated mean (i.e. 5) with the averaged mean(i.e. 5.03) which is very close to each other.

Sample Variance versus Theoretical Variance

Now the varaince and standard distribution from the averaged 40 exponential is

```
sample.variance <- var(sample.mean)
sample.sd <- sd(sample.mean)
sample.variance
```

```
## [1] 0.6312296
```

```
sample.sd
```

```
## [1] 0.7944996
```

While the variance and standard distribution from the simulated can be calculated from the lambda which was 0.2, which gives us:

```
theo.variance <- (1/lambda^2)/n
theo.variance
```

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

```
theo.sd <- (1/lambda)/sqrt(n)
theo.sd
```

```
## [1] 0.7905694
```

Now if we compare the sample variance(i.e.0.6312296) with the calculated variance from the simulation(i.e. 0.625), we observe that it is very close and we get the same result if we compare the standard deviation from the sample standard deviation(i.e. 0.7944996) with the calculated standard deviation(0.7905694).

Distribution Analysyis

Now we can further investigate the data to find out the relation between distribution of the averaged mean of the exponentials to the central limit theorem by plotting a histogram of the averaged mean.

```
mean.dataframe <- data.frame(sample.mean)
forcolour <- c("")
g <- ggplot(mean.dataframe,aes(x = sample.mean))
g <- g + geom_histogram(aes(y = ..density..),fill = "grey66", color = "grey")
g <- g + labs(title = "Distribution of means of 40 exponentials", x = "Means of 40 samples" , y = "Dens
g <- g + geom_vline(aes(xintercept =averaged.mean, colour = "sample") )
g <- g + geom_vline(aes(xintercept = theo.mean, colour = "theoretical"))
g <- g + stat_function(fun = dnorm, args = list(mean = averaged.mean, sd = sample.sd),aes(color = "sampl
```

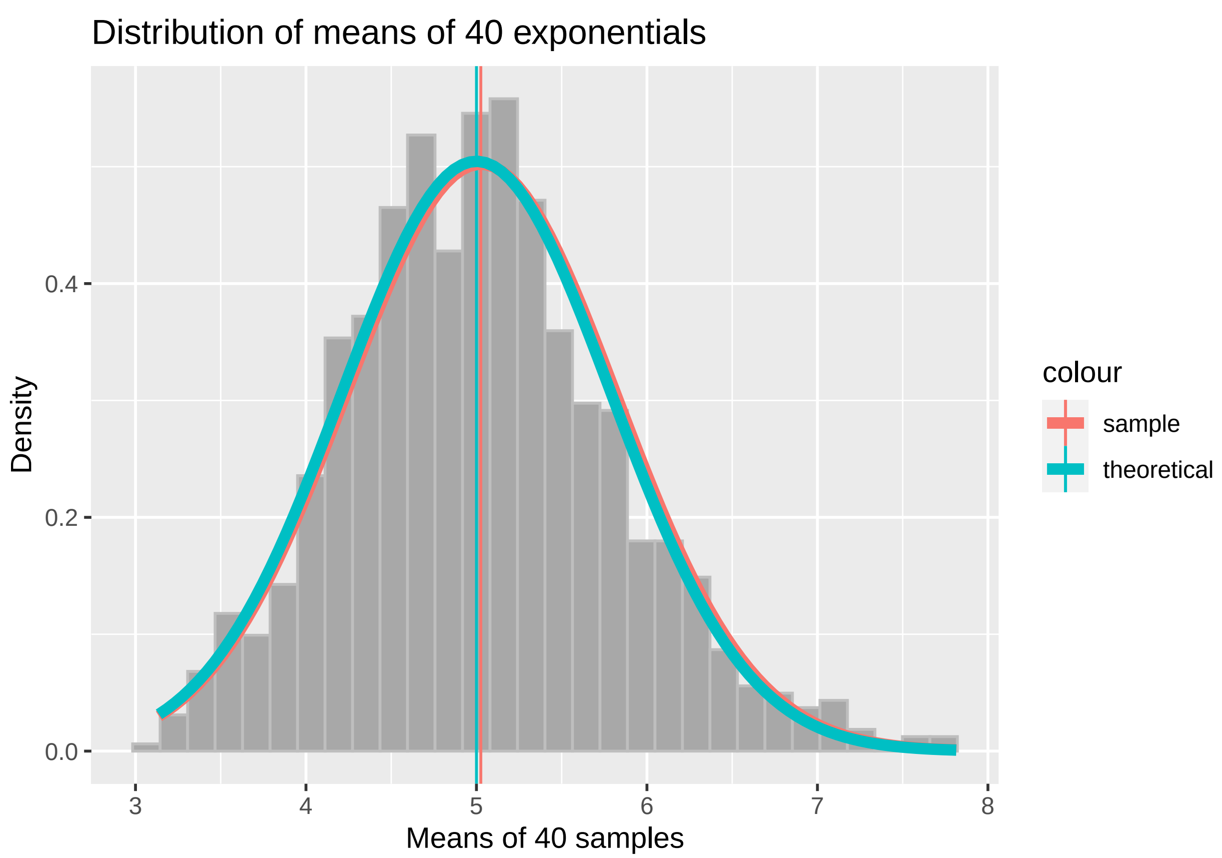
```
## Warning: `mapping` is not used by stat_function()
```

```
g <- g + stat_function(fun = dnorm, args = list(mean = theo.mean, sd = theo.sd),aes(colour = "theoretica
```

```
## Warning: `mapping` is not used by stat_function()
```

```
print(g)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



From the histogram we can observe that the bars is concentrated towards the mean of the histogram which is exactly the central theorem states that the distribution of the averages of i.i.d variables becomes that of a standard normal. Therefore it shows that the sample averaged value is a random distribution which is concentrated towards the mean. And we can also see the observed mean and the theoretical mean is very close to each other.

Calculating Confidence Interval

Now we will calculate the 95% confidence interval and we will check whether the sample mean lies in between the interval or we get any error.

```
conf <- theo.mean + c(-1,1) * qnorm(0.975) * theo.sd/sqrt(n)
conf
```

```
## [1] 4.755005 5.244995
```

From the above calculation we can see that the 95% confidence interval of the simulated exponential is (4.76, 5.24).Now we can see our averaged mean i.e. 5.03 lies in between the confidence interval; Hence we can conclude that the average mean of 40 samples gives an estimate of the simulated mean correctly as the central theorem states, which again proves the distribution is approximately normal.