Statistical Inference Assignment - Part 1

Tomás A. Maccor

23/3/2020

Synopsis

In this assigment we are going to verify if the postulates of the Central Limit Theorem (CLT) stand for an exponential distribution density. That is, as we obtain a higher **N** of random samples from an an Exponention distribution, the distribution of the mean of each of these N random sampless will approximate a Normal distribution (regardless of the fact that the samples are obtained from an exponential distribution density).

We will also try to prove another of the CLT postulates: that the average of the sample means will be the population mean. In other words, if we add up the means from all of the samples & calculate the average —-> that average will be the actual population mean. The same can be applied for finding the standard deviation of your population (distribution, in this case)

Basic Facts

EXPONENTIAL distribution parameters

Mean = $1/\lambda$
λ (lambda) = rate parameter for the distribution
St. dev. = $1/\lambda$

We will work with an exponential distribution that has $\lambda = 0.2$
Therefore, $\mu = 0.5$ (1/0.2)

Simulation

We will simulate obtaining 1000 samples from an exponential distribution with λ=0.2, each of size 40 (n=40). So here we go:

```r
averages <- NULL

for (i in 1:1000)
{
  averages <- c(averages, mean (rexp (40, 0.2) ) )
}
```

```r
str(averages)
```

```
##  num [1:1000] 4.35 4.66 5.51 6.75 5.55 ...
```

"averages" is a vector of length = 1000 —> each element of vector the contains the MEAN of a sample of size n = 40.

Mean & SD of SAMPLE approximate Mean & Sd of POPULATION

Mean

The average of the 1000 sample means is

```r
print(mean(averages))
```

```
## [1] 5.031213
```

This is almost identical to the theoretical mean for an exponential distribution with $\lambda = 0.2$ (which is 0.5, as we saw earlier).

Standard deviation

The standard deviation is the same as the mean in this case.
So we just have to exponentiate it, to get the variance for the exponential distribution with $\lambda = 0.2$:

```r
theoreticalVariance <- (1/0.2)^2 / 40
print(theoreticalVariance)
```

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

If we then calculate the variance of our own sampling data, we get an extremely close approximation:
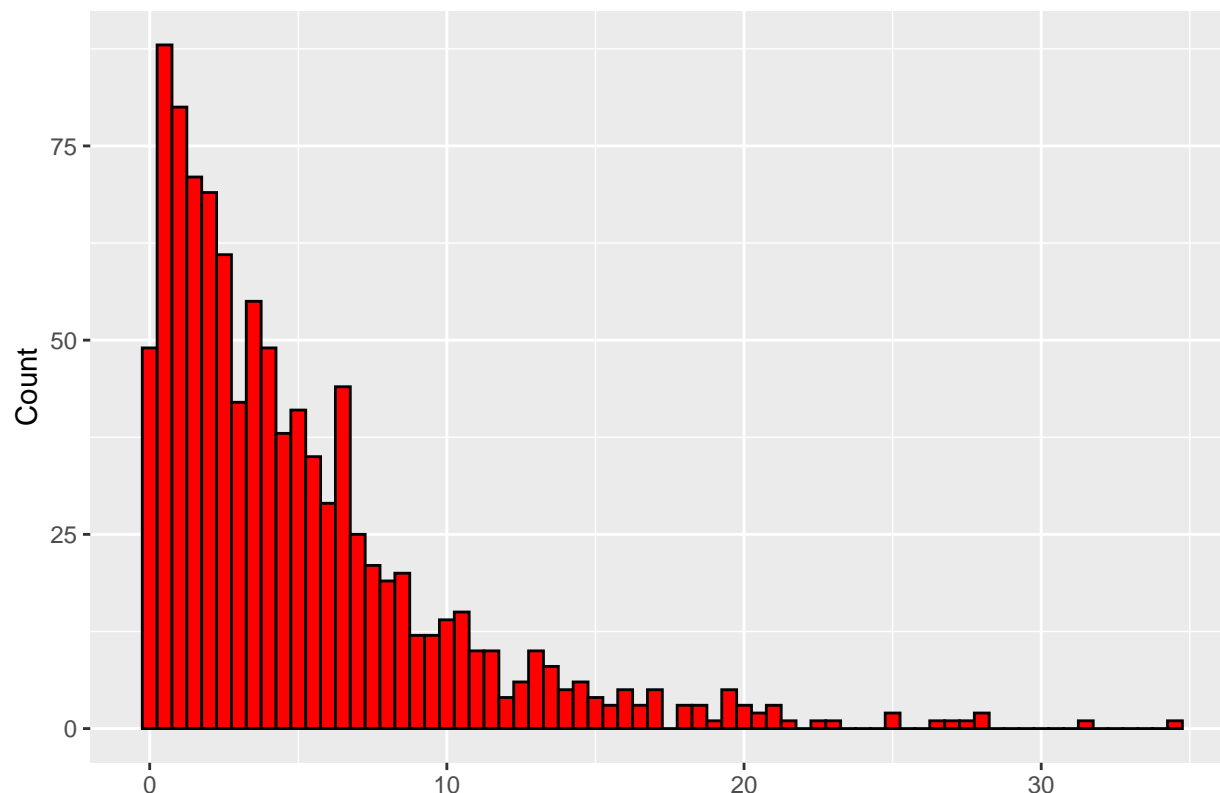
```r
print(var(averages))
```

```
## [1] 0.6119691
```

Now we will show that the distribution of the means of 1000 samples (size n=40) taken from an exponential distribution, actually follow a Normal distribution!

We will first plot how does an exponential distribution of size=1000 looks like:

Exponential Distribution – 1000 random exponentials

And now we plot the distribution of the means of our 1000 samples, and compare it to the curve of a Normal distribution with u and sd equal to the theoretical exponential distribution u and sd:

```
mean_averages <- mean(averages)

sample_plot <- ggplot(averages_df, aes(x_average)) +
               geom_histogram(aes(y = ..density..), binwidth = 0.15,
               color = "black", fill = "lightblue") +
               ggtitle("Distribution of the Mean of 40 random exponentials - 1000 samples") +
               xlab("Sample Means") +
               stat_function(fun = dnorm, args = list(mean = theoreticalMean,
               sd = sqrt(theoreticalVariance)), size = 1.2) +
               geom_vline(xintercept = mean_averages, size = 1.5, color = "red") +
               geom_text(x = 7, y = 0.45, label = "- AVERAGE of the MEANS",
                        color="red") +
               geom_text(x = 7, y = 0.50, label = "- Normal Distribution density",
                        color="black")


sample_plot
```

Distribution of the Mean of 40 random exponentials – 1000 samples

– Normal Distribution density

– AVERAGE of the MEANS