Statistical Inference Project

A Project By Plato Karageorgis

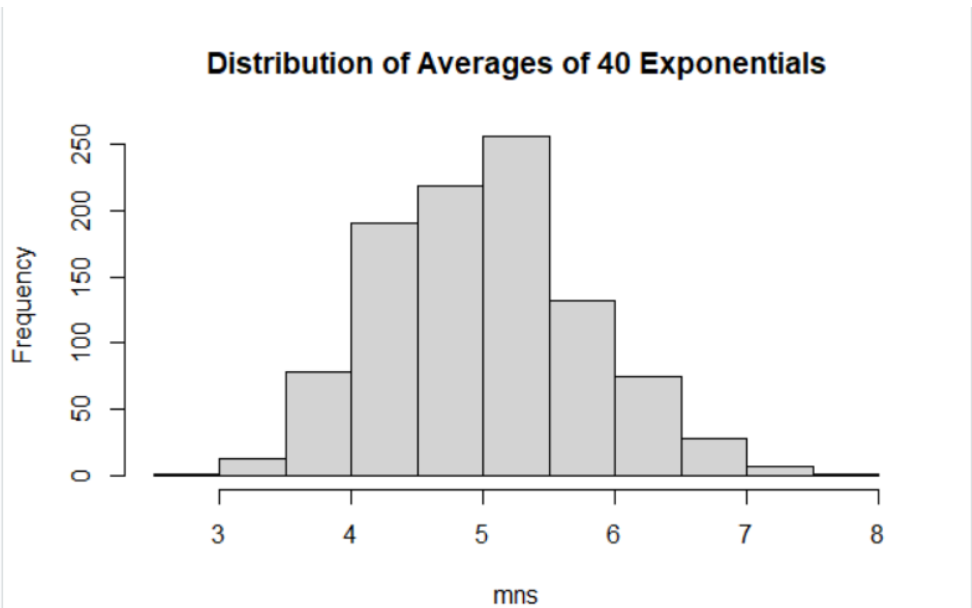
Part 1

The first part of this project revolves around the exponential distribution. As the exercise instructs, the lambda value will be 0.2 throughout the calculations.

To begin with, I made some calculations in R with the following code:

- mns = NULL
- for (i in 1 : 1000) mns = c(mns, mean(rexp(40, 0.2)))
- hist(mns)

Before I explain what I am trying to do here, I will first demonstrate you the histogram that these lines of code generated.

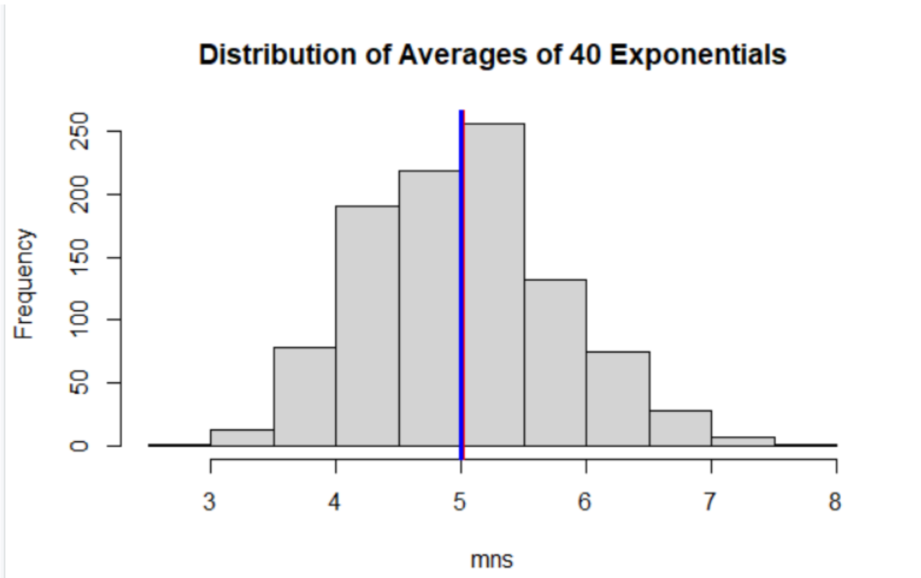


Now, what I am actually doing is this. Instead of using 1 exponential distribution, I use 40 and do the simulation 1000 times. To clarify, I get the average value of 40 exponential distributions, 1000 times. The reason I am doing this, is because the result is satisfying since the distribution is very close to the normal distribution but I will get back to that further below. For now, let's focus on the mean of the distribution.

Sample Mean VS Theoretical Mean

We expect that the Sample Mean will be around 5 but let's keep that aside for a second and calculate the Theoretical Mean. We know from theory that the mean of the exponential distribution is 1/lambda and we stated at the beginning of this assignment that the lambda value is 0.2, so the Theoretical mean is 5. To calculate the sample mean exactly I run the command mean(mns) and the result is 5.002573. I do not know about you but I find the approximation spectacular! In order to see the result more clearly here is the histogram again, the blue line represents the theoretical mean and the red one the sample mean. I also provide you below the extra lines of code that I needed to add these lines.

- abline(v = mean(mns), col = "red", lwd= 3)
- abline(v = 1/0.2, col = "blue", lwd= 3)



Sample Variance VS Theoretical Variance

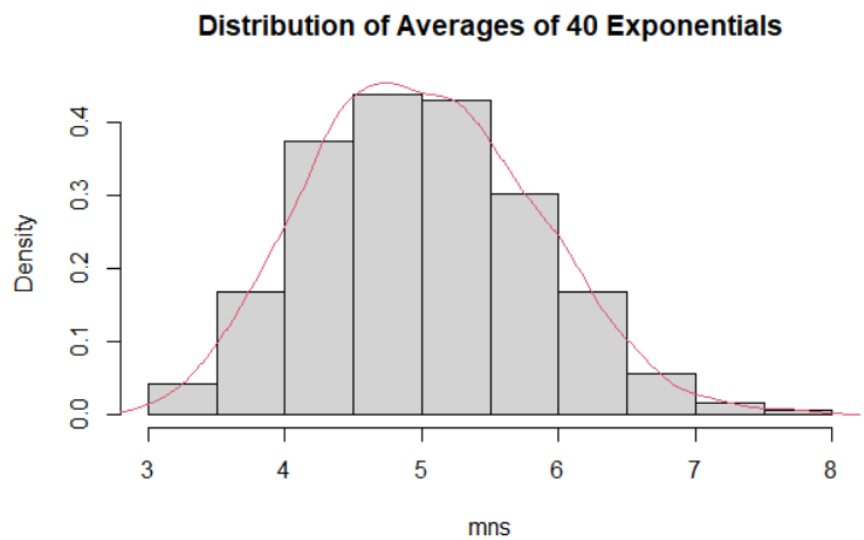
The value of the variance helps us determine the variability of the data and it is crucial for experiments such as this one. Let's do the same thing as before, compare the 2 variances and then we'll discuss the result. The Sample Variance is calculated through the execution of the command "var(mns)" and the result is 0.6195. On the other hand, the Theoretical Variance has the formula " (1/lambda)^2 ". The result is 25. So, it is obvious that the sample variance and the theoretical variance are very different. The sample variance is by far better because it has much less variability and the data are more concentrated.

Normal Distribution

Finally we want to decide if the above distribution fits the normal distribution. We'll draw the density curve on the histogram in order to compare it with the Gaussian bell curve.

- hist(mns,probability=T,main="Distribution of Averages of 40 Exponentials")

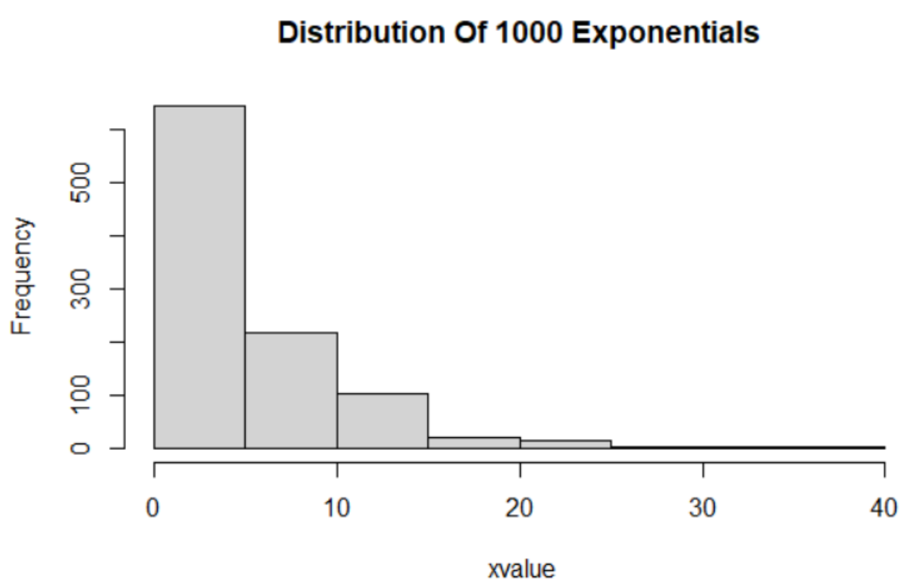
➢ lines(density(mns),col=2)



As we can see from the screenshot above, the density curve of the distribution is very similar to the bell curve of the normal distribution so we can say confidently that the distribution fits the normal one.

Moreover, we'll do the same to a distribution that uses 1000 random exponential calculations (instead of 40) but this time without using the average calculations. Then, we'll compare the result with the previous distribution.

➢ hist(rexp(1000, 0.2),xlab = " xvalue",main = "Distribution Of 1000 Exponentials")



The distribution reminds us of f(x) = e^x and it should because it’s the exponential function. On the other hand, the result obviously does not fit the normal distribution and we can see that the previous distribution was much better.