Muskaan Mahes

Dr. McGee

Nonparametric Statistics 4385

4 October 2024

Stimulation One

I. Introduction

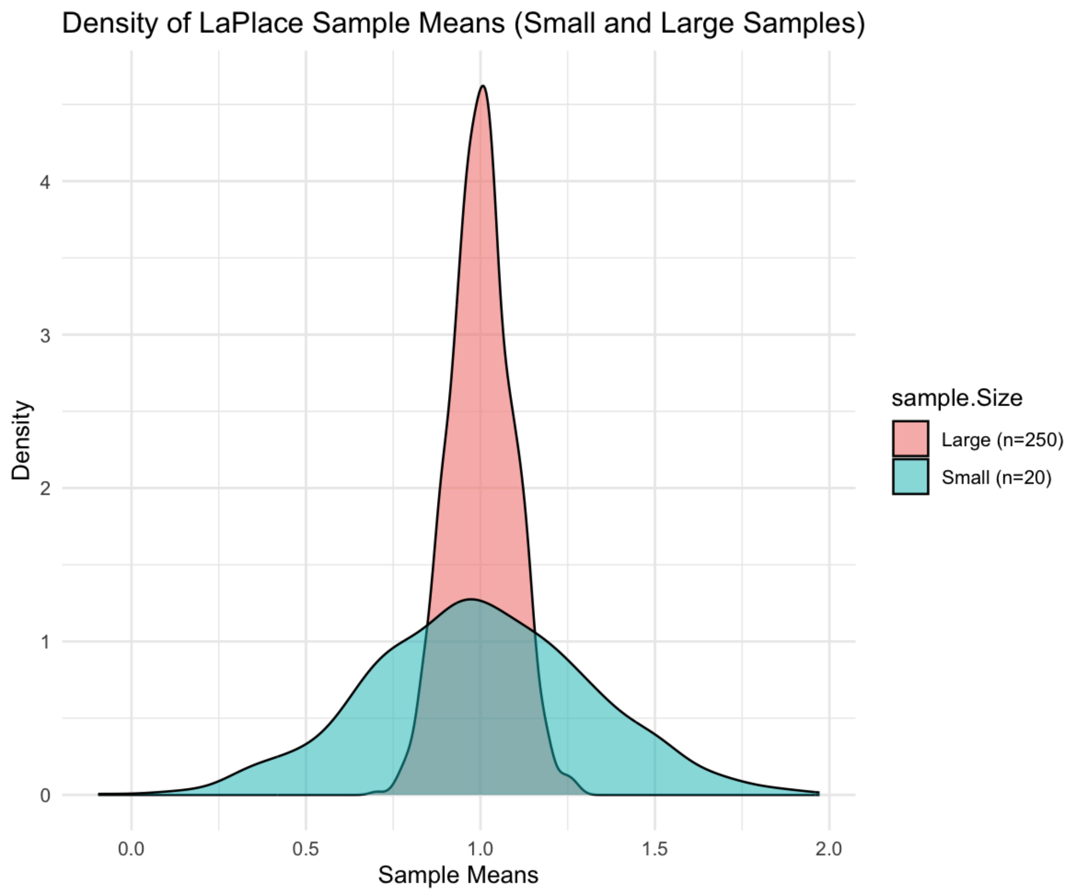
The central limit theorem is a widely used distribution in many areas of mathematics and statistics. The theorem is important because when we are working with a large data set, it allows us to assume that the sampling distribution of the mean will be normally distributed. Thus, we are allowed to make more statistical analysis and inference as we analyze the large data sets. The central limit theorem states that as the sampling distribution of the sample mean approaches the normal distribution the sample size gets smaller. Additionally, since the sample size is divided by the standard deviation, it will also be affected. If the sample size is small, the standard deviation will be larger; this occurs because the sample size is inversely related to the standard deviation. Additionally, the central limit theorem follows a few conditions which are the sample size has to be greater than or equal to 30, and the samples have to be independent. However, in this simulation, we will be testing two distributions that follow the central limit theorem; yet our sample sizes will vary as they will be less than 30 and greater than 200. The purpose of this stimulation is to analyze whether the central limit theorem will still work for distributions that are not normal, skewed, and have a different sample size that does not follow the conditions of the theorem.

II. Methods

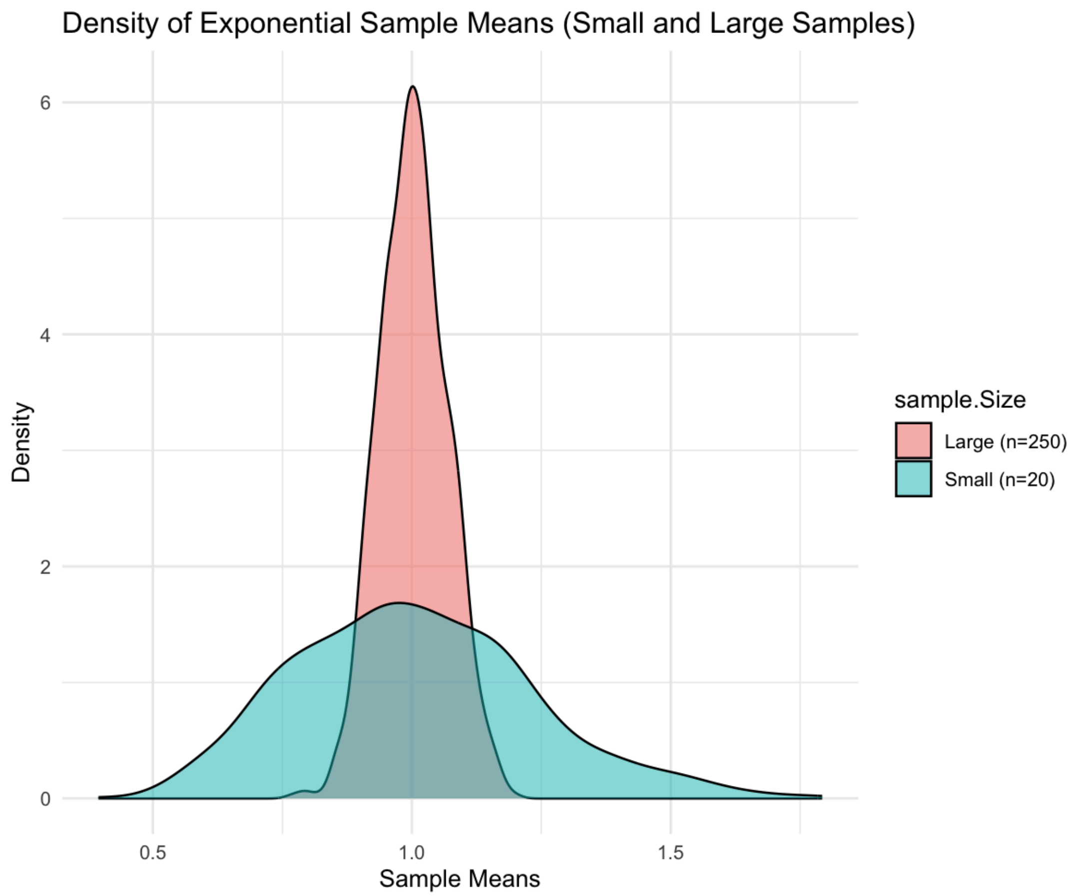
The first distribution I used was the Laplace distribution which is symmetric yet it is not a normal distribution because of its fatter and longer tails meaning that there is a high probability of producing extreme values far from the mean. Whereas, the normal distribution tails are thinner and have a smaller probability of producing extreme values far from the mean. To create this function in R I used “rlaplace” that uses a location parameter which is the mean, and a scale parameter which is the standard deviation. I placed this function in a “for loop” to stimulate 1000 sample means from 1000 samples. The loop also calculated the mean which is equal to 1 and the standard deviation equal to 1 from a sample size of 20. Then I used the mean and standard deviation function to get the values that I was observing. I repeated this procedure again but I changed the small sample size of 20 to a large sample size of 250. For the second distribution, I used the exponential distribution which is skewed because as the tail approaches the x-axis it diminishes as it reaches positive and negative infinity. In comparison, the tails of a normal distribution continue to extend infinitely. I used the “rexp” function which changed from 20 to 250 depending on the sample size, and a parameter rate of one. Then I used the mean and standard deviation function to get the corresponding values. Following the procedures for these two different distributions allowed me to see if the difference in the small and large sample sizes followed the central limit theorem.

III. Results

After running each distribution in R I noticed that for each sample size they were not too far off. For the Laplace distribution with sample size of 20, the sample mean was 0.9978674. Additionally, the stimulated standard deviation was 0.3164658, and to find the theoretical standard deviation I first found σ by using this equation $\sqrt{2}\sigma$ which gave me the value of 1.4142135. Then I used the standard error of the mean equation by dividing 1.4142135 by the $\sqrt{n}$, which would be in this case $\sqrt{20}$. Thus, doing so gave me the empirical standard deviation of 0.3162277 which is relatively close to to the stimulated standard deviation I found. For the Laplace distribution with a large sample size of 250, the stimulated sample mean was 1.0000706. To find the theoretical standard deviation I used the same procedure as above, and got the value of 0.0894427 which is close to the stimulated standard deviation I got of 0.08815852. For both of the sample means of each sample size they were very close to 1, but the larger sample size was ever closer. Therefore, from using the Laplace distribution it confirmed a multiude of things such as, if the sample size is small, the standard deviation will be larger. Additionally, the larger the sample size is the closer it is to the sample mean and population standard deviation you choose to observe. The density plot below, depicts this idea because the smaller sample size is more spread out and not as tall as the larger sample size. While the larger sample appears to look like the normal distribution as it is taler and not as widespread.



For the exponential distribution with a small sample size of 20, the sample mean was 1.003276. To find the theoretical standard deviation I first used $\sigma = 1/\lambda$ where $\lambda=1$ to find it equal to 1. Then I used the standard error of mean equation $\sigma / \sqrt{n}$ that translated to $1/\sqrt{20}$ which gave me a standard deviation of 0.2236067. This was relatively close to the stimulated standard deviation I found which was 0.2288488. For the exponential distribution with a large sample size of 250, the sample mean was 1.002426. I then used the same procedure as above to find the theoretical standard deviation to be 0.0632455, and this was also very close to the stimulated standard deviation I found which was 0.06514611. For each sample size the sample means was practically equal to my observed mean value of 1 that I choose.



This density plot shows again, that the smaller the sample size the more spread out it is, and the larger the sample size the taller it; which appears to be just like the normal distribution graph. The table below gives the calculation of each distribution for each sample size.

	Small Sample Size (n=20)	Large Sample Size (n=250)
LaPlace Theoretical Mean	1.000000	1.000000
LaPlace Stimulated Mean	0.9978674	1.000706
LaPlace Theoretical Standard Deviation	0.3162277	0.0894427
LaPlace Stimulated Standard Deviation	0.3164658	0.08815852
Exponential Theoretical Mean	1.000000	1.000000
Exponential Stimulated Mean	1.003276	1.002426
Exponential Theoretical Standard Deviation	0.2236067	0.0632455
Exponential Stimulated Standard Deviation	0.2288488	0.06514611

IV. Conclusion

After completing this stimulation we can confirm that the central limit theorem follows various distributions regardless of the sample size. As the sample size gets larger the distribution of the sample mean follows the normal distribution. Therefore, regardless the sample size, the central limit theorem allows us to analzye data sets and make it easier for statistical analysis and inference.