Central Limit Theorem Revisited: Non-Uniform Distributions

B. C. Chap *

Winter Quarter 2021

Course: Computational Methods of Mathematical Physics
Instructor: Professor D. Ferenc
University of California, Davis
One Shields Ave., Davis, CA 95616, USA

Abstract

In the previous project we explored the results of the Central Limit Theorem and the way various data are distributed depending on the distribution of the population sample. In this project, we do the same for a non-uniform linear and Mickey Mouse distribution.

1 Central Limit Theorem

The Central Limit Theorem (CLT) states a sufficiently large number of observed events yields a normal distribution [1]. Probability distributions describe how likely different possible outcomes of an event are and are characterized by a few parameters: the mean μ, the standard deviation σ, and the variance σ^2 as previously discussed. The mean and standard deviation of the probability distribution for a variable x of sample size N can then be calculated as:

$$\mu_{\bar{x}} = \mu \tag{1}$$

$$\sigma_{\bar{x}} = \sigma/\sqrt{N} \tag{2}$$

The CLT holds true when either of two conditions are met: the sample population is normal or the sample size is sufficiently large [1], as seen in the previous project solving Problem 6.3 in Wong [2]. The latter is the condition of interest. To further prove the CLT, we will examine it in two additional scenarios, non-uniform distributions for a straight sloped line and a Mickey Mouse distribution.

*bcchap@ucdavis.edu

1.1 Problem 6.3 Revisted

Problem 6.3 is as follows: **Use a random number generator with an even distribution in the range [—1, +1] to produce n = 6 values and store the sum as x. Collect 1000 such sums and plot their distribution. Compare the results with a normal distribution of the same mean and variance as the x collected. Calculate the χ^2 value. Repeat the calculations with n = 50. Compare the two chi^2 obtained.**

1.2 General Procedure

To solve this problem, we must to create uniformly distributed data for different n values and compare them to Gaussian distributions, and calculate their χ^2 values. The procedure to accomplish this is as follows:

Argument List:

- n: number of values to be created
- nsum: the number of sums of n values, normalized by the mean
- lower_bound, upper_bound: interval for the random number generator
- bins: number of data groupings
- mu: mean value for the normal distribution
- sigma: standard deviation of the normal distribution

1. Uniform Probability Distribution: Linear Interval = [-1,1], Mickey Mouse Interval = [-2,2]

    2. Create an array of n number of values with uniformly random generated numbers, input them into a confined random distribution

        3. Linear: $y = m*x+b$ where m and b are randomly generated within $\pm .5$ of a specified slope and y-intercept

        4. Mickey Mouse: Three circles of the form $coefficient * ((x - shift_{left/right})^2 + (y - shift_{up/down})^2)$

    5. Create an array where each value sums the n number of values previously generated $nsum$ times

        6. Linear: Sum all values created

        7. Mickey Mouse: Sum all values within the Mickey Mouse outline

    8. Put the values of $nsum$ into a histogram with $bins$ bins

    9. Plot the histogram

10. Normal Probability Distribution: $\mu = 0$, $\sigma = \sqrt{\sigma_i^2}$, interval = [-1,1]

    11. Plot a normal distribution centered at μ with standard deviation σ on the same plot as the uniform histogram

1.3 Discussion

The CLT is once against demonstrated. Regarldess of the type of distribution for a sample, if there is a sufficient sample size, the distribution of measurements will approach normal. This is depicted for both the linear and Mickey Mouse distributions for n = 1, n = 50, and n = 500.

1.3.1 Linear Data

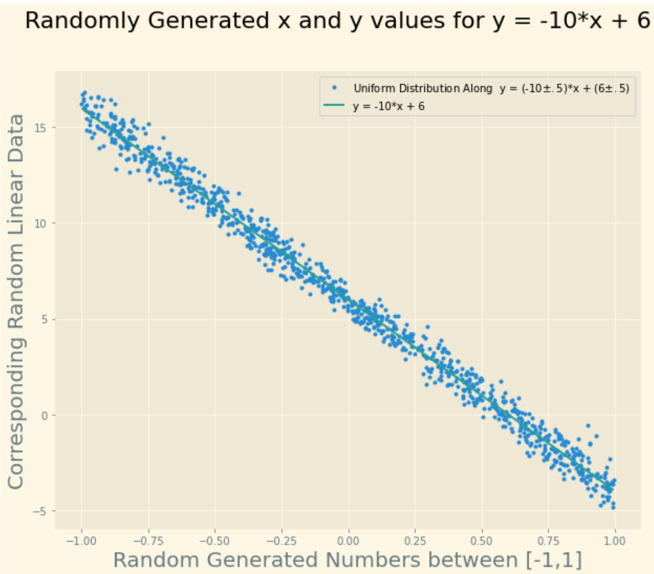

Figure 1: Randomly Generated Numbers Along a Linear Line for $n = 1$ and $nsum = 1000$

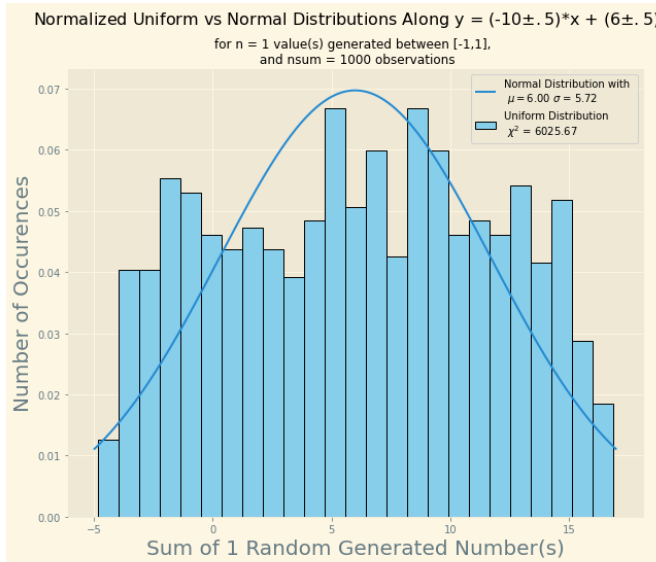

Figure 2: Normalized Distribution of $nsum = 1000$ sums of $n = 1$ confined to a Linear Distribution

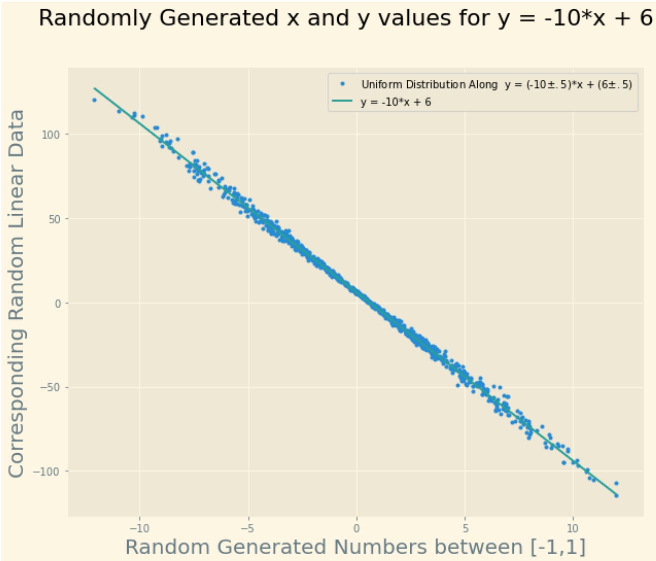

Figure 3: Randomly Generated Numbers Along a Linear Line for $n = 50$ and $nsum = 1000$

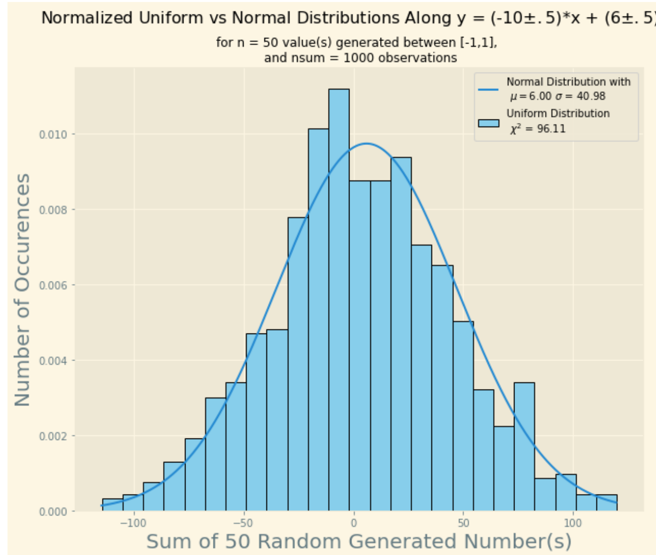

Figure 4: Normalized Distribution of $nsum = 1000$ sums of $n = 50$ confined to a Linear Distribution

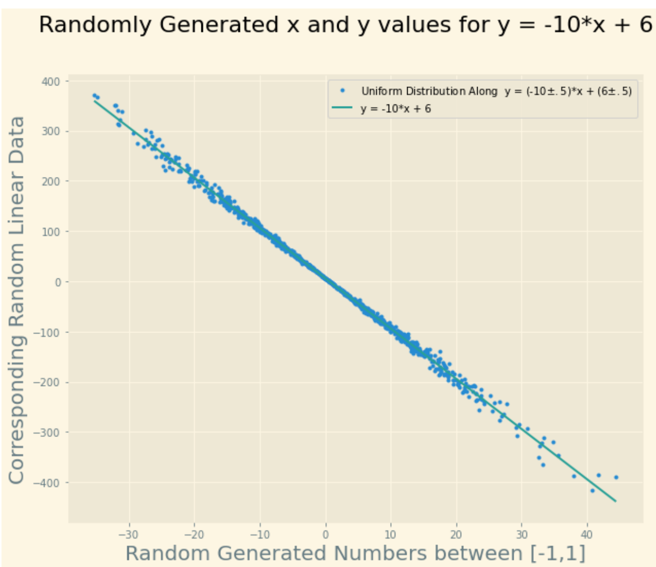

Figure 5: Randomly Generated Numbers Along a Linear Line for $n = 500$ and $nsum = 1000$

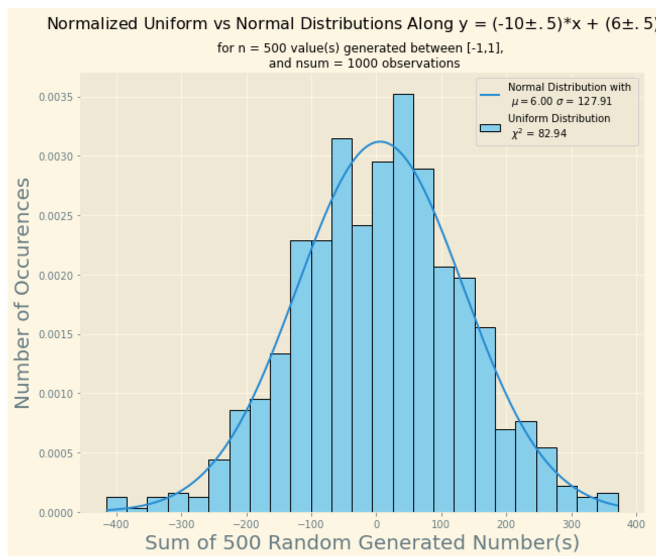

Figure 6: Normalized Distribution of $nsum = 1000$ sums of $n = 500$ confined to a Linear Distribution

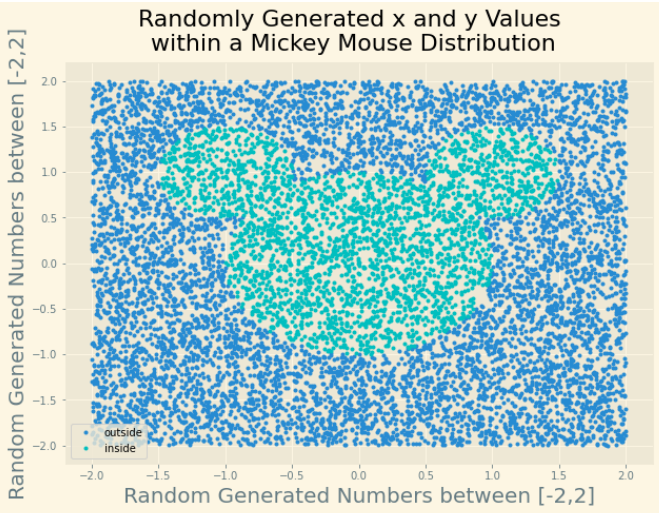

Figure 7: Randomly Generated Numbers Within a Mickey Mouse Distribution $n = 1$ and $nsum = 1000$

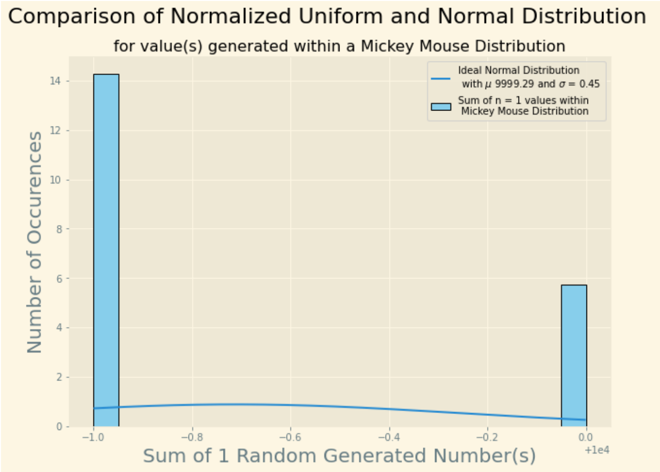

Figure 8: Normalized Distribution of $nsum = 1000$ sums of $n = 1$ confined to a Mickey Mouse Distribution

1.3.2 Mickey Mouse Data

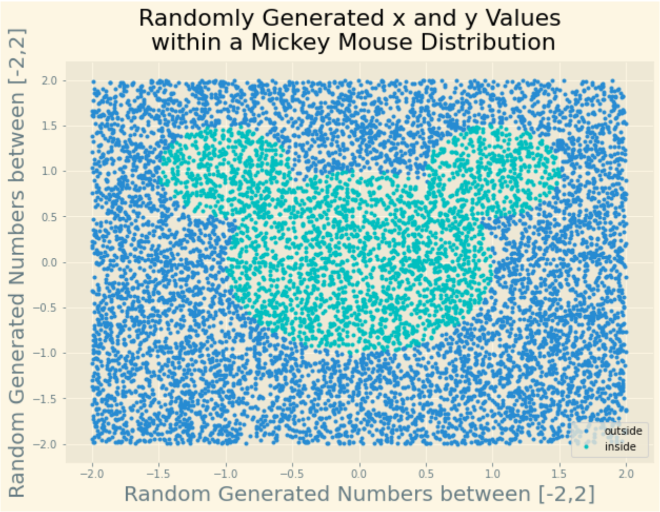

Figure 9: Randomly Generated Numbers Within a Mickey Mouse Distribution $n = 50$ and $nsum = 1000$

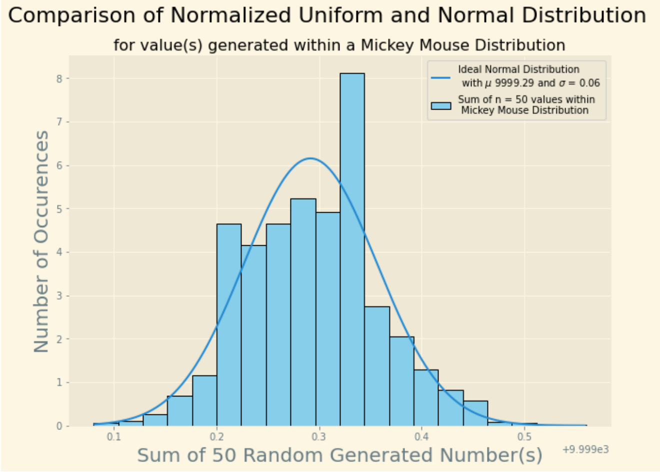

Figure 10: Normalized Distribution of $nsum = 1000$ sums of $n = 50$ confined to a Mickey Mouse Distribution

Figure 11: Randomly Generated Numbers Within a Mickey Mouse Distribution $n = 500$ and $nsum = 1000$

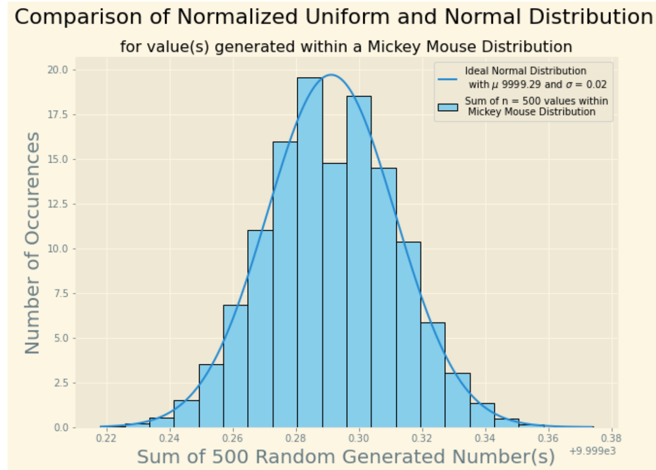

Figure 12: Normalized Distribution of $nsum = 1000$ sums of $n = 500$ confined to a Mickey Mouse Distribution

It was previously stated that the distribution of values will approach a normal distribution if one of two conditions are met: either the population data is normal or there is a sufficiently large population. Using a uniform random number generator and confining it to a function, we are able to examine the latter. For n = 1, we see that the distribution is not Gaussian at all, it is uniform as expected. However, as n increases, we see that the distribution becomes more Gaussian, once again confirming the CLT. This is the case for both the Linear and Mickey Mouse distributions. Regardless of the distribution of the population, for a sufficiently large population, the distribution approaches a Gaussian distribution.

2 Conclusion

In this project, we examined the CLT again, and as expected, the CLT holds true for large values of n and regardless of how the initial population is distributed. Specifically, the distribution of data approaches a Gaussian distribution regardless of the sample population distribution. Overall, the CLT allows us to determine whether a certain set of measurements is valid in relation to a Gaussian distribution.

References

[1] LaMorte Wayne W. *Central Limit Theorem*. URL: https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_probability/BS704_Probability12.html.

[2] Wong S.S.M. *Computational Methods in Physics and Engineering*.