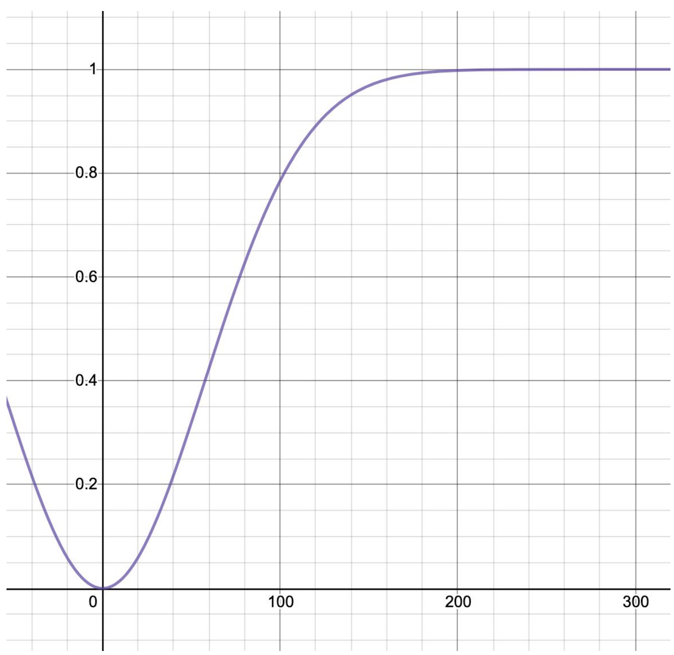
Saiteja Bevara (sb2xf) and Rithik Yelisetty (ry9bf)

APMA 3100: Project 3

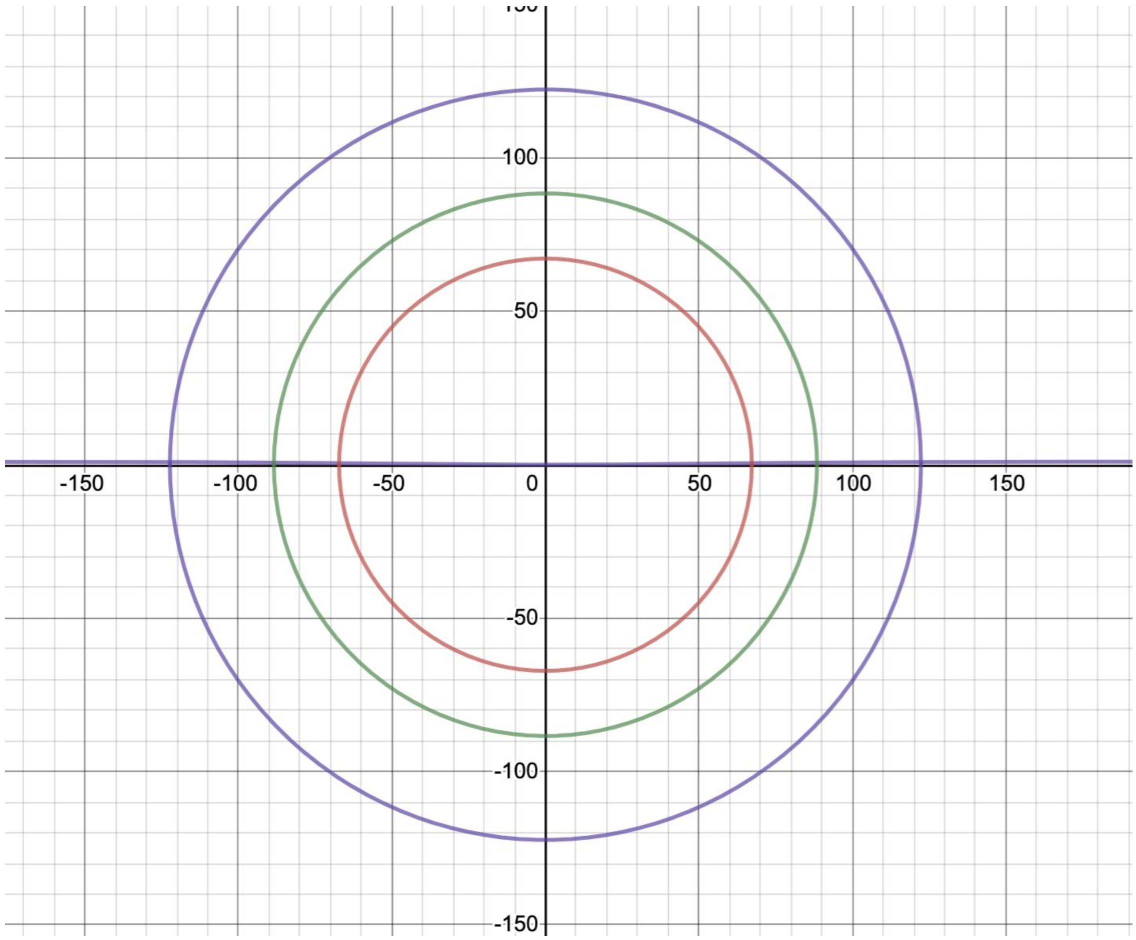
Section 3

1. Below is the graph of function f_x, where $f_x = a^2xe^{-1/2 * a^2x^2}$. This represents the PDF of the Rayleigh Distribution, where $a = 1/\tau$, where τ is equal to 57 inches.

2. Below is the graph of function F_x, where $F_x = 1 - e^{-1/2 * a^2x^2}$. This represents the CDF of the Rayleigh Distribution, where $a = 1/\tau$, where τ is equal to 57 inches.



3. Next, we found values of x_p where the following statement was true: $P[x \leq x_p] = p$, where p is equal to 0.5, 0.7, 0.9. We then used the x_p values as the radius of circles that are based at the origin. The equations of the circles are:



$X_p = x^2 + y^2 = 67.22^2$

$X_p = x^2 + y^2 = 88.45^2$

$X_p = x^2 + y^2 = 122.32^2$

4. The circles indicate the probability the newspaper will drop within a certain distance from the target location. There is a 50% chance the newspaper will fall within the red circle from the target, or 67.22 inches from the target. There is a 70% chance it falls within 88.45 inches, and a 90% chance it falls within 122.32 inches. This means that the homeowner should be aware that the newspaper will most likely (~90%) fall within 10 feet and 2.32 inches of the target location.

Section 4.2

In order to complete this part of the project, the team used the Python programming language to simulate the results of the experiment. The team started out by creating the random number generator with the values supplied in the directions. The team found the following random values: U_{51} = 0.398994445800781, U_{52} = 0.4001159667968 and U_{53} = 0.093833923339843. We then calculated 110 estimates of m_n in order to determine M_n. The data belows shows the relationship between n (on the left) and M_n (on the right):

10: 59.91095043760961

30: 64.7255501241526

50: 62.49111986027385

100: 63.22820475124484

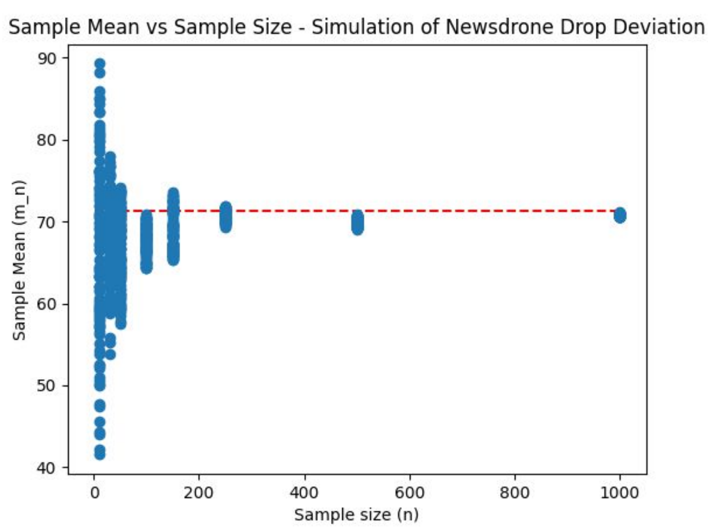
150: 66.84336567974064

250: 69.44822598591331

500: 69.09465584075485

1000: 70.5980696423357

The team then plotted every single value of m_n on a graph with the x-axis as n to understand the distribution of the data. The graph is shown below:



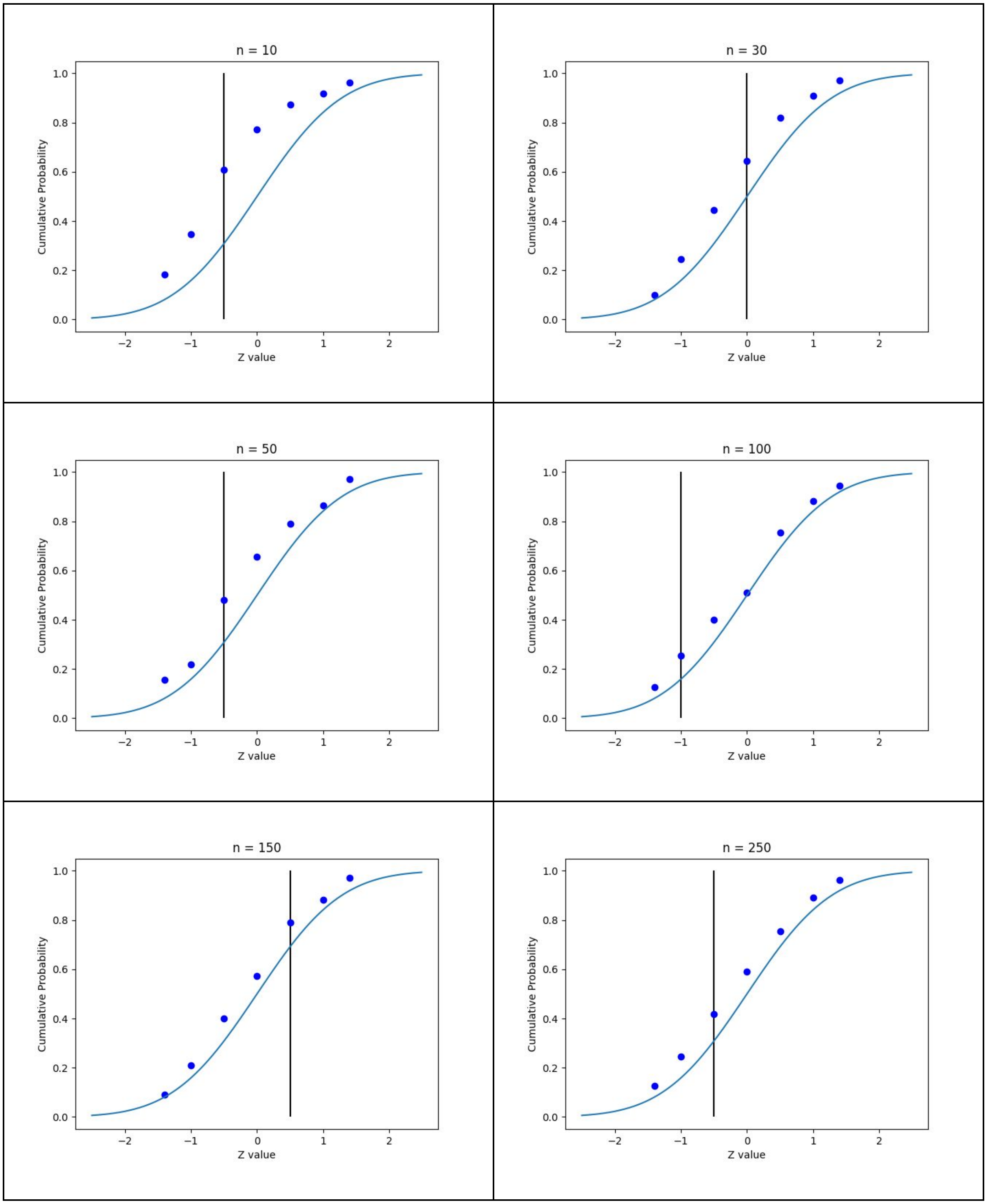
As shown in the graph, for small values of n, there is large variance in the values of m_n. However, as the values grow larger and larger, that is values such as n = 500 and n= 1000, the values for m_n begin to converge closer to the expected mean μ_x. Therefore, this shows the behavior for the strong law of large numbers, as M_n converges to μ_x as the value for n increases without bound. However, even before we reach values closer to infinity, we see on our graph that the values are converging for n = 500, 1000, etc., which means we can pick a sample size prior to infinity, which is not practical, that would converge the mean to the expected mean.

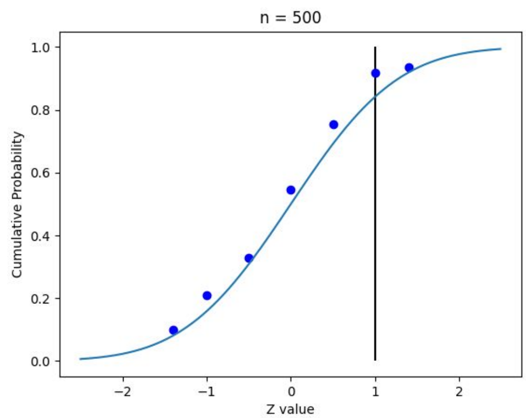
For the real life experiment, we recommend a sample size of n* = 100 to calculate the sample mean of the distance X. Based on our observed data, we find that amongst the given sample sizes, 100 is the first sample size where all of the data points are confined within the bounds of 61 to 81 inches, determined as 10 inches from the expected value of 71 inches. The sample size measured before 100, n = 50, had data points below 61, and thus was not fit as the optimal sample size for this investigation. Samples sizes larger would suffice, but 100 was the lowest for which the statement held true.

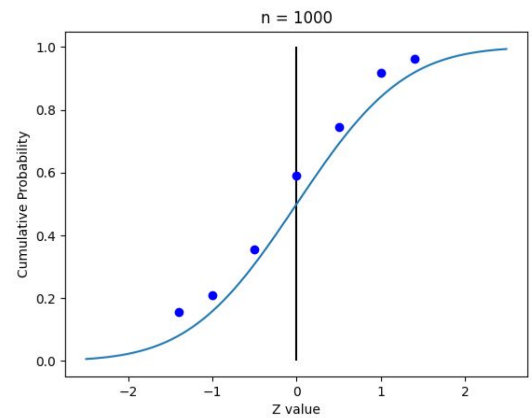
$$P[\,|M_{100} - \mu_x| < 10\,] \approx 0.936$$

The probability p as estimated above is derived from the probability that values of estimates fall out of the range of 61 to 81 inches. For sample sizes smaller than 100, there is a possibility from the observed data for 7/110 estimates to fall out of this range. As n increases, less estimates will fall out of this range, so the probability would increase, which the weak law of large numbers explains.

Section 5.2


n = 10
n = 30
n = 50
n = 100
n = 150
n = 250
Cumulative Probability
Z value





		J							
		1	2	3	4	5	6	7	MAD_N
N	10	0.10102	0.18675	0.30059	0.27273	0.18123	0.07688	0.04244	0.30059
	30	0.01920	0.08675	0.13695	0.14545	0.12668	0.06779	0.05153	0.14545
	50	0.07375	0.05948	0.17332	0.15455	0.09941	0.02234	0.05153	0.17332
	100	0.04647	0.09585	0.09150	0.00909	0.06305	0.04052	0.02425	0.09585
	150	0.01011	0.05039	0.09150	0.07273	0.09941	0.04052	0.05153	0.09941
	250	0.04647	0.08675	0.10968	0.09091	0.06305	0.04961	0.04244	0.10968
	500	0.01920	0.05039	0.01877	0.04545	0.06305	0.07688	0.01516	0.07688
	1000	0.07375	0.05039	0.04605	0.09091	0.05395	0.07688	0.04244	0.09091

Conclusion

The standard normal distribution CDF and the observed CDF are very similar for all values of n, with slight variations displayed by the values for MAD_n. However, for all values of n, the CDFs and MAD_n values were not converging to any particular point, as the variation between $\widehat{F_n}$ and $\varphi(n)$ are relatively stagnant across all values of n and MAD_n was not converging to 0 as n approaches larger numbers although it does overall show a decreasing trend. The CLT is used to approximate the standard normal distribution, which our data shows as the observed CDF was similar to the standard normal CDF.

Reflection

The experiment evidently was accurate at displaying the strong law of large numbers as the sample mean converged to the expected mean for larger values of sample size, n. However, there was no evidence for convergence of the CDF or MAD_n. This could be a result of the iteration number of 110 used for calculating estimates for the sample mean. An improvement would be to choose more varying values of K to better achieve the standard normal distribution.

Code

Our code can be found at https://github.com/rithik/Newsdrone/.

On my honor as a student, I have neither given nor received aid on this assignment.

Saiteja Bevara

Rithik Yelisetty