APMA 3100 Project 3

By Matthew Beck and Spencer Hernandez

Honor Pledge: On my honor as a student I have neither given nor received aid on this assignment

X: Matthew Beck & Spencer Hernandez

3. Model Analysis

1.

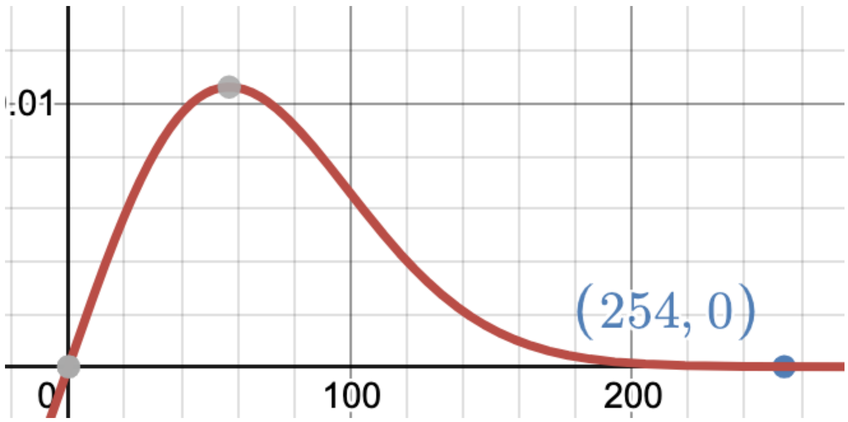

Figure 1: Graph of f_x

2.

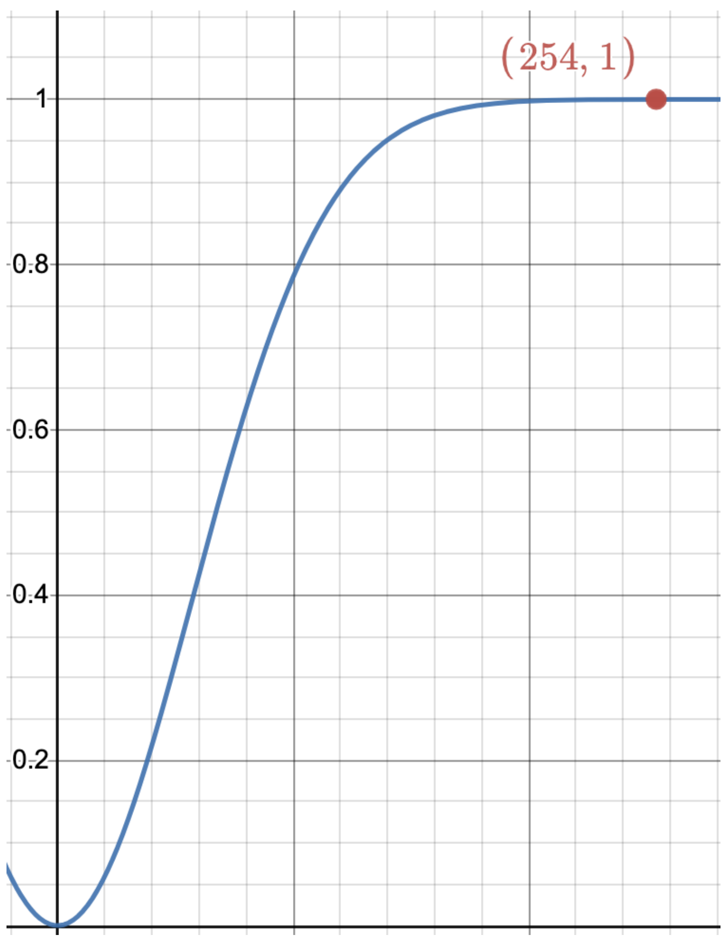

Figure 2: Graph of F_x

3.

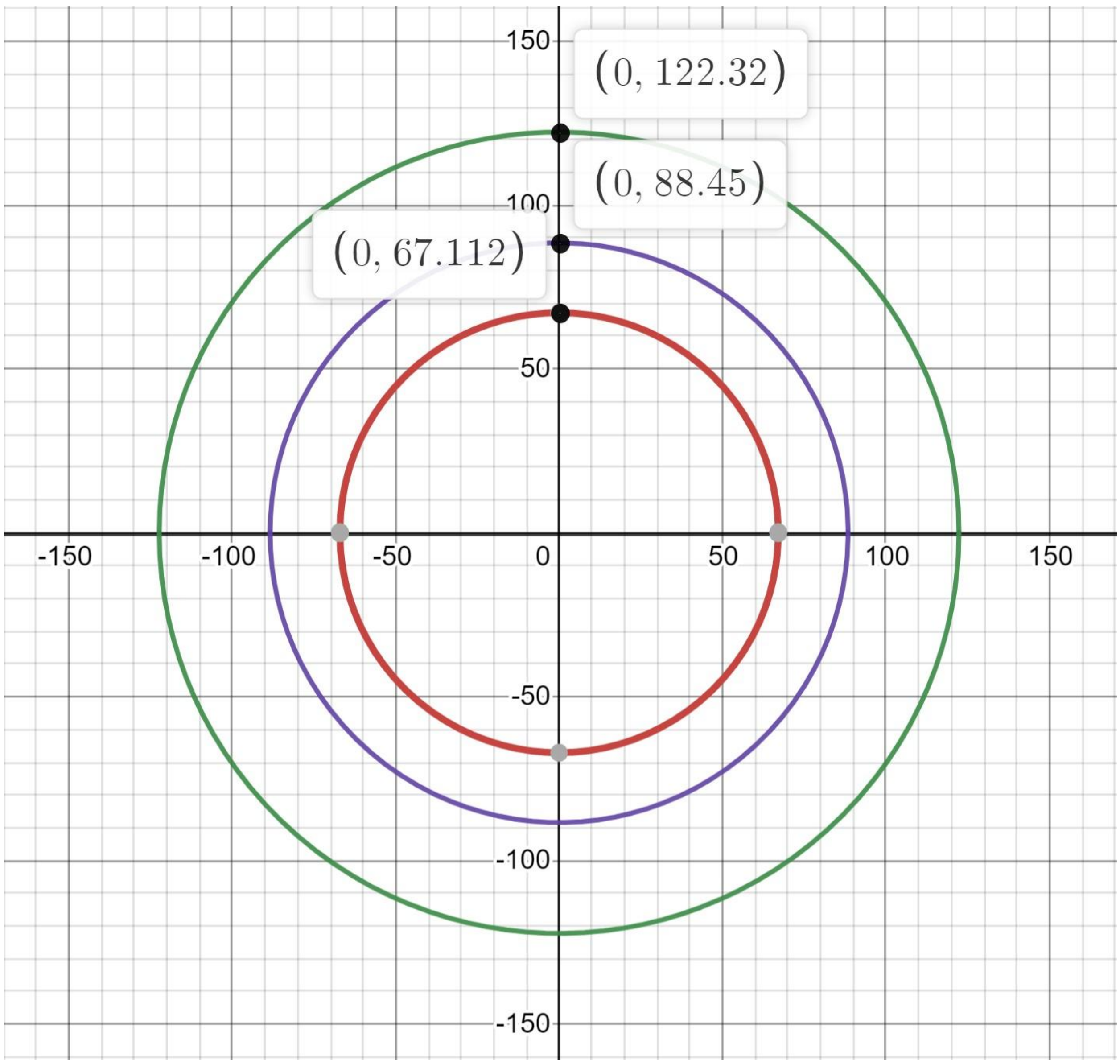

Figure 3: Three circles of radius x_p centered around point T

4. "While our drone delivery service manages to deliver newspapers with general accuracy, there are minor errors that occur due to speed, release angle, wind speed/direction, and so on. We can predict that the newspaper will drop less than or equal to some distance (radius, more specifically) as a result of this uncertainty. If we want to have more

confidence in where the newspaper will drop, our radius will have to be bigger because there is a tradeoff between how confident we are in the drop zone and the radius of the drop zone."

4. Law Of Large Numbers

1. The values for u_{51}, u_{52}, and u_{53} produced from the linear congruential random number generator are 0.1995, 0.2001, and 0.0469 respectively. More precise values can be seen in figure 4 which shows the output from the generator itself. We coded the random number generator and the monte carlo simulation in python. For the monte carlo simulation we had a function that took one parameter, n, which served as the n value in M_n. Each time the function is called it computed the expected value for M_n 110 ten times, returning the results in a list. Using this list we plotted the data in figure 5.

```
51
 LCG:  52297
 u:  0.19949722290039062
52
 LCG:  52444
 u:  0.2000579833984375
53
 LCG:  12299
 u:  0.04691696166992187 5

Process finished with exit code 0
```

Figure 4: u_{51}, u_{52}, and u_{53} values

4. The graph below is representative of the sample mean estimates, m_n, from the population mean M_n, with varying sample sizes, n, which are shown on the abscissa.

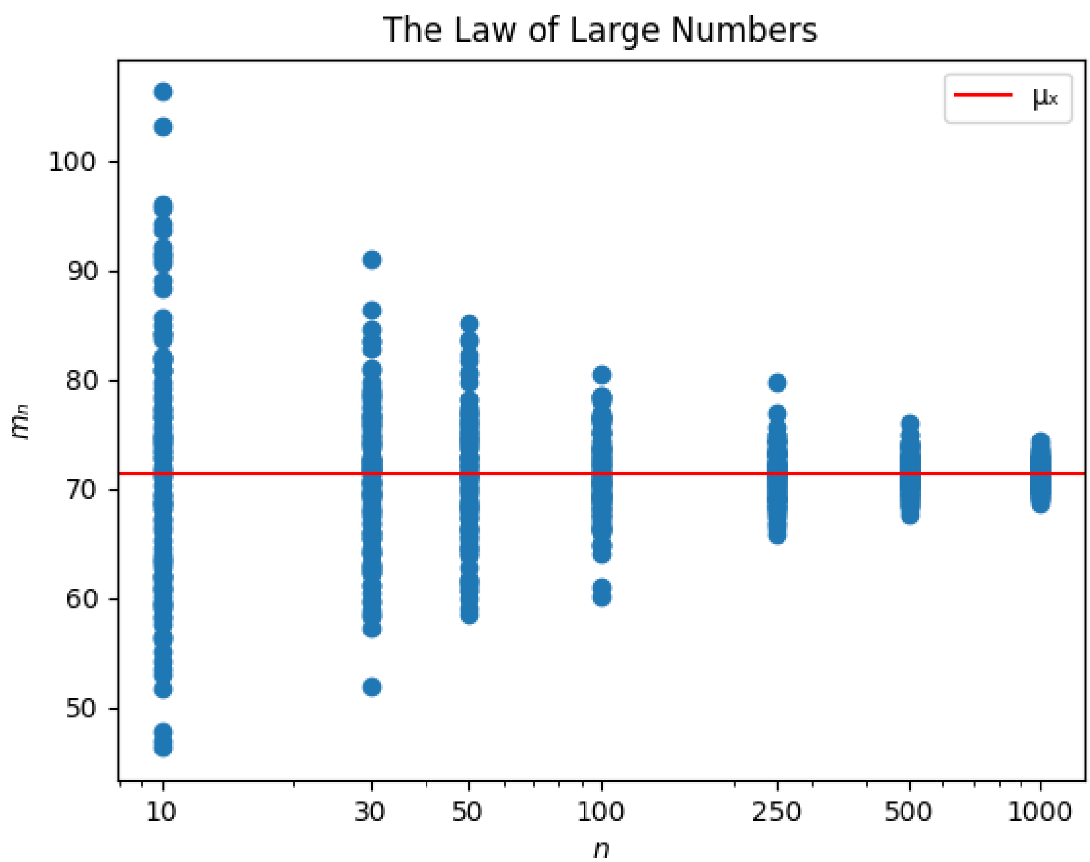

Figure 5: Scatter Plot of Simulation Results

5. From this graph, it is evident that the larger the sample size, the closer the sample mean becomes to the actual mean. However, we cannot be for certain that the sample mean is approaching the population mean, as it could be approaching the population mean with some small difference, c. Therefore, as our sample size grows to infinity, the probability that the difference between the sample mean and population mean becomes smaller than c approaches 1. This model represents the Weak Law of Large Numbers.

6. After running the monte-carlo simulation for various sample sizes, while it may be clear that the sample mean approximates the population mean, as n grows very large, the estimate tends to be accurate for small n as well. The sample mean for various n sizes is shown below in Figure 6.

```
n = 10: mₙ = 72.69101555201418
n = 100: mₙ = 70.73534839621395
n = 1000: mₙ = 70.70562910403378
n = 10000: mₙ = 70.7831599440629
```

Figure 6: Sample mean for various n

It is evident that the marginal gain of approximating the population mean becomes very small. Therefore, an n^* value in the range from 10 to 100 should be optimal so as to minimize the cost while still maintaining a close estimate to the population mean. To further analyze, we can graph a scatter plot of $P[|M_{n^*} - \mu_x| < c] \approx p$ (the probability that the difference of means is smaller than $c = 10$) vs. n (the sample size). The following plot is shown below in Figure 7.

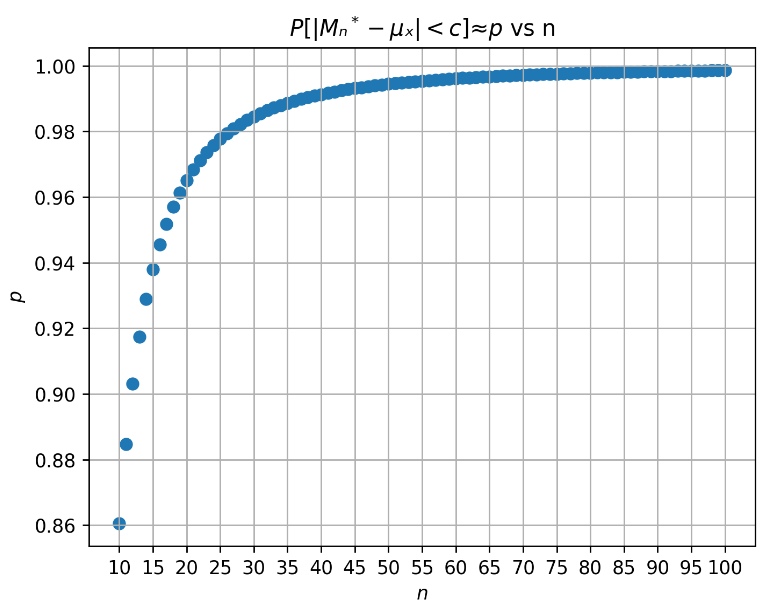

Figure 7: Probability Given By Chebyshev Inequality vs. Sample Size

It is evident that the probabilities converge to 1 as n grows large, however a large confidence interval can also be achieved with a far smaller n. Since minimizing cost is important, a value of $n^* = 25$ could be an appropriate sample size. While one could obviously choose a smaller n-value for a smaller cost, the probability drops off significantly faster since $n^* = 25$ is at the cusp of the curve, so there is less value added. A higher n^* could also be chosen, but the value added again is not worth it because a maximum probability of 0.02 can be gained compared to the probability at $n^* = 25$ (the actual value is calculated below and is approximately 0.9776). This plot was generated using the Chebyshev Inequality, as shown below. The equation can be written as $p \geq 1 - \frac{\sigma^2_{n^*}}{nc^2}$, or rather, $p \geq 1 - \frac{Var[X]}{n^2c^2}$. So by plugging in a sample size n, a minimum probability p can be realized, allowing the graph to be made.

7. Using the Chebyshev Inequality, we get the results shown below in Figure 8.

$P[|M_{n^*} - M_x| < c] \approx p$ $\quad n=50 \quad \sigma^2_{n^*} = \frac{Var[X]}{n} \quad c=10$

$1 - P[|M_{n^*} - M_x| < c] \leq \frac{\sigma_{n^*}^2}{nc^2}$ $\quad = \frac{1394.48}{25}$

$P[|M_{n^*} - M_x| < c] \geq 1 - \frac{\sigma_{n^*}^2}{nc^2}$ $\quad = 55.779$

$$p \geq 1 - \frac{\sigma^2_{n^*}}{nc^2}$$

$$p \geq 1 - \frac{55.779}{25 \cdot (10)^2}$$

$$p \geq 0.9776$$

Figure 8: Probability Given By Chebyshev Inequality

5. Central Limit Theorem

3.

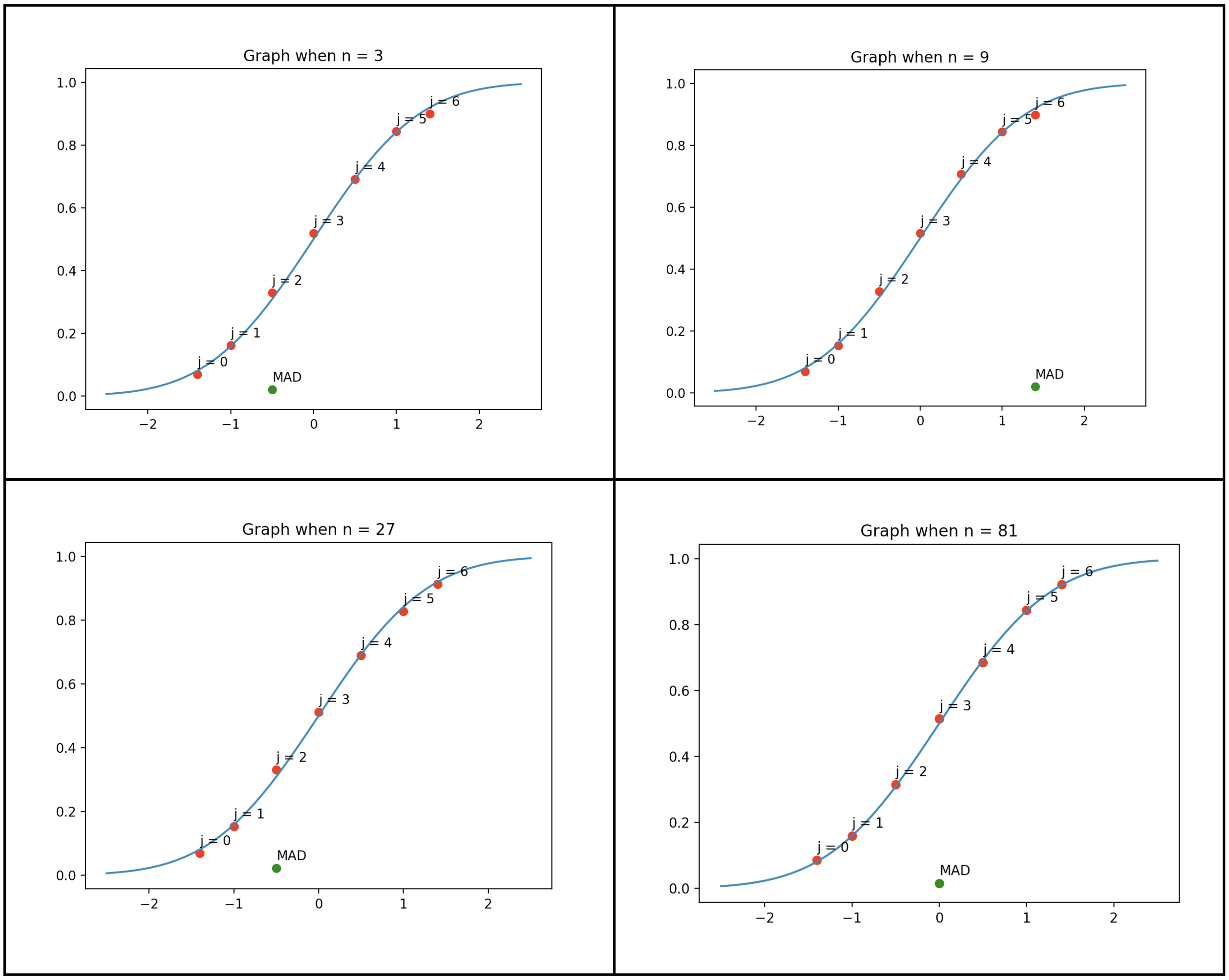

Figure 9: Standard Normal CDF vs. Empirical CDF with MAD

	n = 3	n = 9	n = 27	n = 81
$\widehat{\mu}_n$	71.6404	70.9968	71.3669	71.2738
$\widehat{\sigma}_n$	21.2227	12.5785	7.1443	4.1805
μ_x	71.4389	71.4389	71.4389	71.4389
$\sigma_x/\sqrt{n}$	21.5599	12.4476	7.1866	4.1492

Figure 10: Table of Estimates versus Population Values

	n = 3	n = 9	n = 27	n = 81
MAD_n	0.0206	0.0211	0.0224	0.0145
$\left\lvert \widehat{F_n}(z_1) - \phi(z_1) \right\rvert$	0.0117	0.0117	0.0117	0.0047
$\left\lvert \widehat{F_n}(z_2) - \phi(z_2) \right\rvert$	0.0032	0.0059	0.0059	0.005
$\left\lvert \widehat{F_n}(z_3) - \phi(z_3) \right\rvert$	0.0206	0.0187	0.0224	0.0060
$\left\lvert \widehat{F_n}(z_4) - \phi(z_4) \right\rvert$	0.0023	0.0164	0.109	0.0145
$\left\lvert \widehat{F_n}(z_5) - \phi(z_5) \right\rvert$	0.0006	0.0158	0.0024	0.0060
$\left\lvert \widehat{F_n}(z_6) - \phi(z_6) \right\rvert$	0.0023	0.0023	0.0141	0.0023
$\left\lvert \widehat{F_n}(z_7) - \phi(z_7) \right\rvert$	0.0192	0.0211	0.0065	0.0026

Figure 11: Table of MAD values and Absolute Differences

4. The empirical CDFs converge to the standard normal CDF as n gets larger as can be seen in figure 9. The realizations of the empirical CDF become closer and closer to the standard normal CDF as n grows larger, which is the definition of convergence. Overall, I would say this experiment does a really good job at demonstrating what the Central Limit Theorem is. It is already fairly intuitive to understand, but actually seeing graphical data converge to a standard normal Gaussian distribution is very useful for verification.

5. If we were to repeat this experiment, we believe that it would be beneficial to have different n values including larger values of n to further demonstrate that the empirical CDF converges to the standard normal CDF with the distance between the two decreasing for even larger values of n.