CSE 107
Lab Assignment 6

The goal of this project is to demonstrate the content and generality of the Central Limit Theorem, which is found in section 5.4 of the text Introduction to Probability (2nd edition) by Bertsekas and Tsitsiklis. This theorem states that if $X_1, X_2, \ldots$ is a sequence of independent and identically distributed (i.i.d.) random variables with common mean μ and variance σ^2, and if we define

$$Z_n = \frac{(X_1 + X_2 + \cdots + X_n) - n\mu}{\sigma\sqrt{n}}$$

then the CDF $F_{Z_n}(z)$ of Z_n converges to the standard normal CDF $\Phi(z)$, i.e.

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z)$$

for each $z \in \mathbb{R}$.

To understand what is going on, observe that the sum $S_n = X_1 + X_2 + \cdots + X_n$ has mean $E[S_n] = n\mu$, and variance $\mathrm{Var}(S_n) = n\sigma^2$ (using independence of the X_i). Thus, the random variable

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

is a shifted and scaled version of the sum, having mean 0 and variance 1. The content of the theorem is that the distribution of Z_n approaches the standard normal distribution as $n \to \infty$. The most amazing feature of this theorem is its generality. For reference, let X be a random variable having the common distribution of all the X_i. We can think of the X_i then as just independent repetitions of the experiment defining X. Then the convergence to normal of Z_n does not depend in any way on the distribution of X. In fact, X can be any random variable of any kind: discrete, continuous, or even mixed. This explains why the normal distribution is found everywhere in nature. It appears to be a law of nature because it is in fact a law of arithmetic.

In this project, X will be a discrete random variable with $\mathrm{range}(X) = \{1, 2, 3, 4, 5, 6\}$. We can think of X as being the result of rolling a single loaded die. The loading can be chosen arbitrarily, since your results will not depend on it. Just choose any numbers $p_1, p_2, p_3, p_4, p_5, p_6 \in (0, 1)$ satisfying $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$ as your probabilities for each outcome. The example

https://people.ucsc.edu/~ptantalo/cse107/Winter26/Examples/LoadedDice.py

illustrates one way to simulate loaded dice in Python. Once you have your weights $p_1, p_2, p_3, p_4, p_5, p_6$, the mean μ and variance σ^2 of X are easily computed. We should emphasize that the distribution of X is fixed, so you will choose the weights only once in this project.

To sample a value from Z_n, roll your weighted die n times, compute the sum $S_n = X_1 + X_2 + \cdots + X_n$, then compute Z_n using the above formula. Your simulation will use the value $n = 500$. This procedure constitutes a single trial that will be repeated 20,000 times. Using these trials, compute the relative frequencies of the events $\{Z_n \le z\}$ for all $z \in \{0.00, 0.01, 0.02, \ldots, 3.47, 3.48, 3.49\}$, which are 350 equally spaced points in the range 0.00 to 3.49 spaced 0.01 apart. These are exactly the set of points used to present the standard normal CDF $\Phi(z)$ as a table here

https://people.ucsc.edu/~ptantalo/cse107/Winter26/Handouts/StandardNormalTable.pdf

and in the text. Your relative frequencies are then your estimates of the CDF $F_{Z_n}(z)$ of Z_n. Present your estimates in a table with the same layout as the standard normal table, i.e.

Experimental CDF

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
0.1	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
0.2	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
0.3	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
0.4	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
0.5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
0.6	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
0.7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
0.8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
0.9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1.0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1.1	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1.2	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1.3	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1.4	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1.5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1.6	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1.7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1.8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1.9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2.0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2.1	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2.2	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2.3	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2.4	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2.5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2.6	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2.7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2.8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2.9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
3.0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
3.1	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
3.2	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
3.3	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
3.4	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

As usual, replace $.0000$ everywhere with your estimate of $F_{Z_n}(z) = P(Z_n \leq z)$ for $z = r + c$, where r is a row label and c a column label. Once you have your table of relative frequencies, observe how close it is to the standard normal table, as asserted by the Central Limit Theorem. With the parameters specified here ($n = 500$ and 20,000 trials), most of your table entries should match the standard normal table to 2 or 3 decimal places. All entries should match to at least one decimal place. Turn in a pdf file called lab6.pdf containing your table, and any other information you think is pertinent.

Note the similarity between the output table in this project and that from lab5. You should therefore be able to re-use much the code you wrote for that project. This project is about the same level of difficulty as lab5, so get started as soon as possible, and get help in discussion sections and office hours.