Programming task, COMP533 CA3 component

Challenge Description: Exploring the Collatz Conjecture and its Variations

In this challenge, you will investigate the behavior of sequences generated by three variants of the famous *Collatz conjecture*. In each variant you are asked to compute the lengths of a *sequences* (*paths*) generated for all integers n in the range 1 to 1000000 (one million).

You will also need to compute the averages and standard deviations defined below.

The **average value** (or mean) of a set of numbers is the sum of the numbers divided by the total count of numbers. The formula for the average value of a set of numbers $x_1, \ldots, x_n$ is:

$$\mu = \frac{1}{n}\sum_{i=0}^{n} x_i$$

Where:

- $x_1, \ldots, x_n$ are the individual data points, and
- n is the total number of data points.

The **standard deviation** measures the spread of a set of values. It indicates how much the values deviate from the mean. The formula for the **standard deviation** σ is:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(x_i - \mu)^2}$$

In all challenges try to replicate the plots provided. Your codes should work for any $n \leq 1M$.

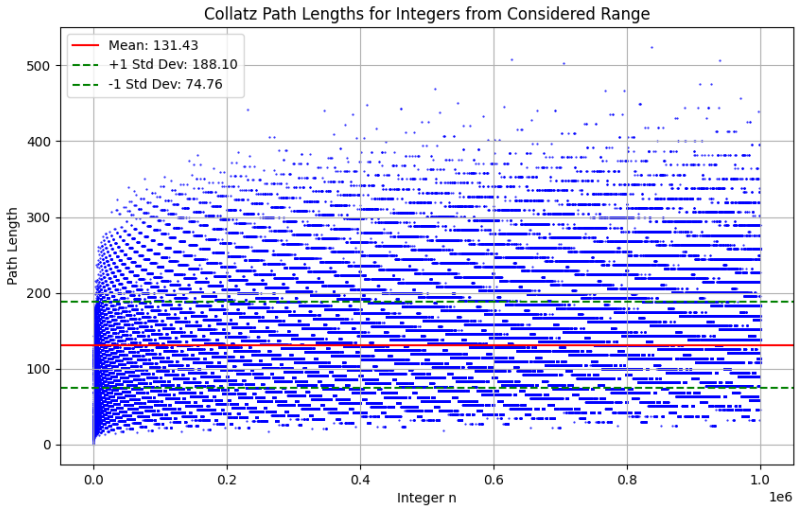
Challenge 1: In the first variant we will be working with the original Collatz conjecture. Namely, given any integer n we generate a sequence (*path*), by applying three rules:

1) if n is even, divide it by 2,
2) if n is odd, multiply by 3 and add 1,
3) if $n = 1$, STOP.

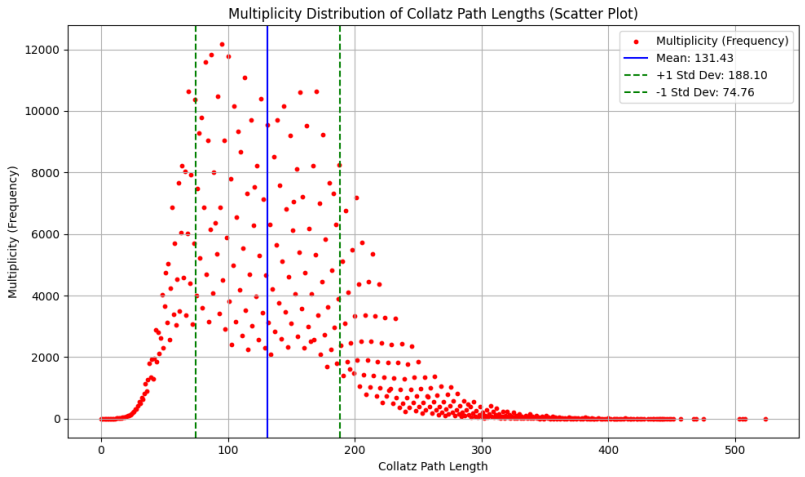
For example, for $n = 1$, the sequence of transitions (applying rules 1 and 2) is of length 0. For $n = 2$, the sequence $2 \to 1$ is of length 1. But for $n = 5$, the sequence $5 \to 16 \to 8 \to 4 \to 2 \to 1$ is of length 5.

Task 1: You are asked to write two python codes which compute the lengths of the sequences associated with all integers n in the range 1 to 1000000 (one million), and

a) plot these lengths against integers by a python code named *L.py*



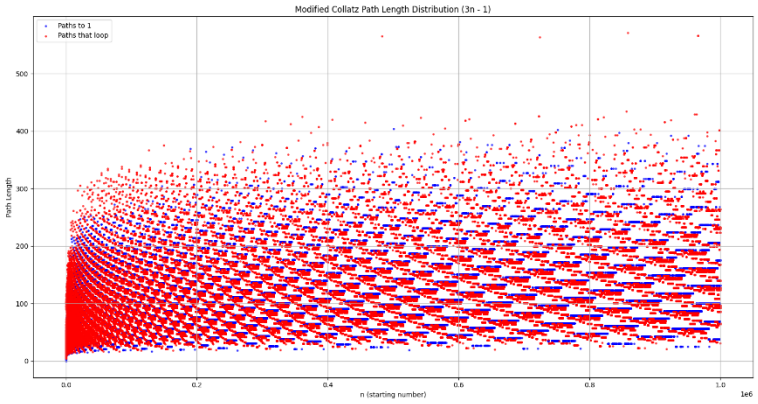
b) plot multiplicities of lengths against available lengths by a python code named *M.py*



Your plots must include the average value based on all lengths (with repetitions) and the relevant standard deviation (same in both cases). **(2 + 2 marks)**

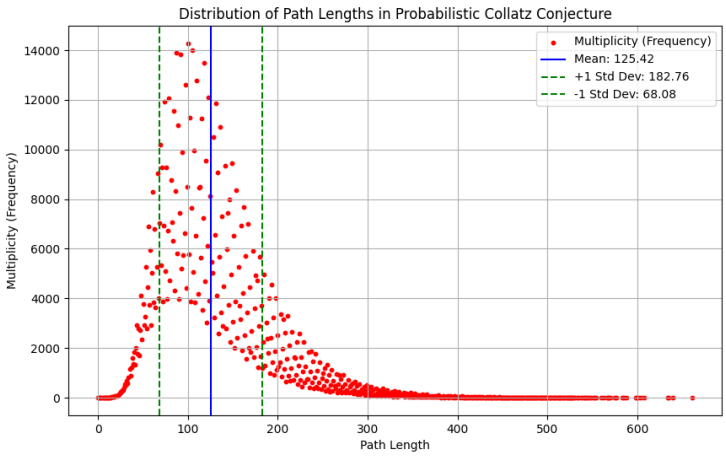
Challenge 2: In the second variant we assume that rule 2) is modified to $n = 3n - 1$. In this case the conjecture about path leading to 1 does not apply, i.e., there are some integer n for which their path loop eventually. We define two groups of integers: **blue group** with paths reaching 1, and **red group** with paths that loop. In the red group the length of any path is the number of transitions (application of rule 1 or 2) needed to loop the path.

Task 2: You are asked to write a python code *N.py* which computes the lengths of the sequences associated with all integers n in the range 1 to 1000000 (one million) and visualizes in the same plot the lengths of paths in the blue group as blue points and in the red group in as red points. **(2 marks)**



Challenge 3: In the third probabilistic variant we assume that rule 2) is now probabilistic. Namely, for odd n we use formula $n = 3n + 1$ with probability p and formula $n = 3n - 1$ with the remaining probability $1 - p$.

Task 3: You are asked to write a python code *P.py* which computes the lengths of the sequences associated with all integers n in the range 1 to 1000000 (one million), and plot multiplicities of lengths against available lengths. In your experiment assume that the probability p=0.5. **(2 marks)**



Task 4: Comment on how and why this distribution changes with different choices of the probability p. **(2 marks)**

Submit zipped codes and the answer sheet with your plots and comments via Canvas by

Monday December 2nd 23:59 hours, 2024.