Exploring the Collatz Conjecture

Introduction

The Collatz conjecture is a simple yet unproven mathematical conjecture about a function that operates on integers. Specifically, it concerns the following rules:

1) If a number is even, divide it by 2
2) If a number is odd, multiply it by 3 and add 1

The conjecture states that if you apply these two rules repeatedly, any starting integer will eventually reach 1. Despite its simple description, the Collatz conjecture remains an open problem in mathematics.

Prior Work

In 1937, Lothar Collatz first proposed this conjecture. Since then, mathematicians have verified its truth for vast ranges of starting integers (Tao, 2019). Using a computer, Krasikov and Lagarias (2003) even proved a lower bound that at least x0.84 integers below x eventually reach 1. However, a general proof still eludes mathematicians.

Methods

To explore patterns in the conjecture, I wrote code examining the "stopping times" required for different starting integers. Specifically, I measured how many iterations of the Collatz function are needed before various integers reach 1. My code utilized three sampling techniques:

1) Log-spaced sampling across intervals
2) Evenly-spaced sampling with binning
3) Random sampling of very large integers (up to 10^250)

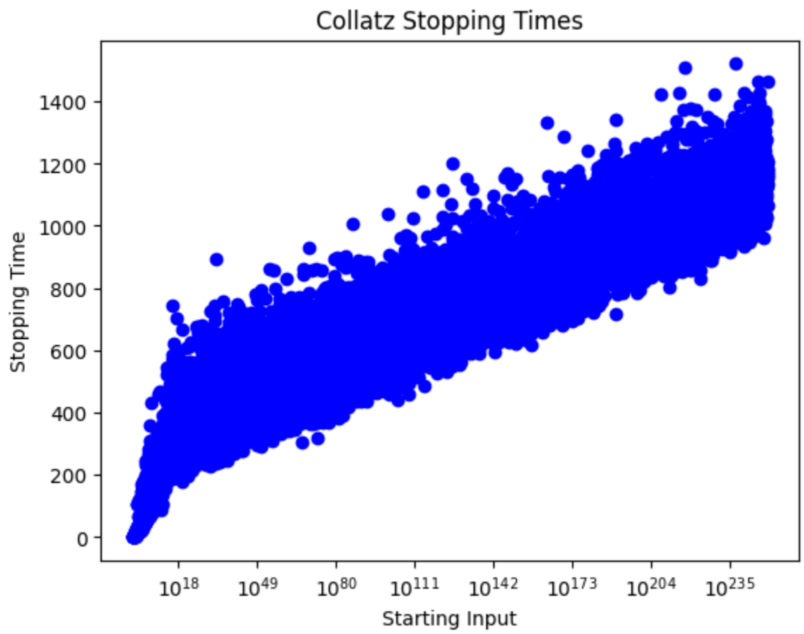
Results

My explorations revealed intriguing structure in Collatz stopping times. First, starting integers seem to follow one of two distributions of stopping times. For one distribution, I identified an approximation formula for the stopping times. However, there is also an odd staggering effect between *close* (say 1000 and 1001) inputs and their stopping times, where too close numbers may have largely different stopping times (say 1500 and 7500).

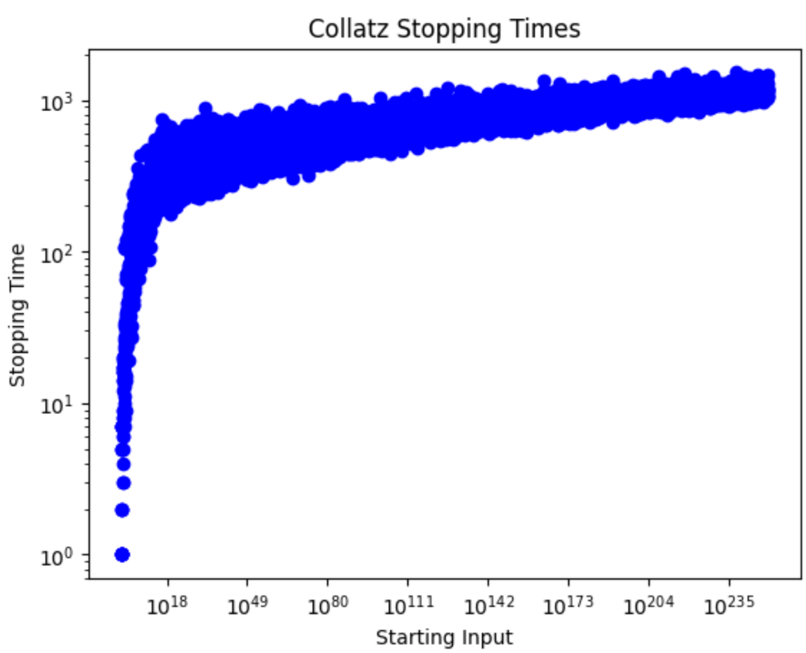
Note:

I've attached some figures from my exploration below.

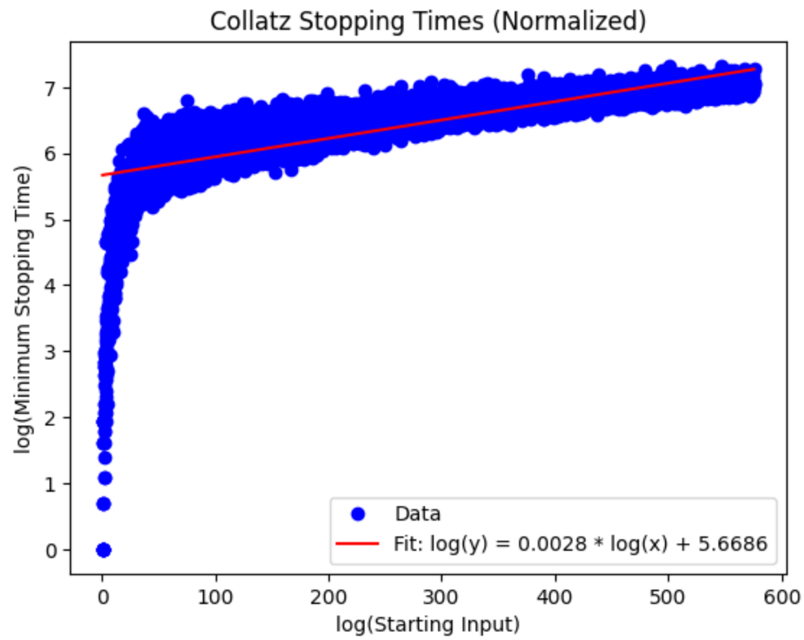
Figures



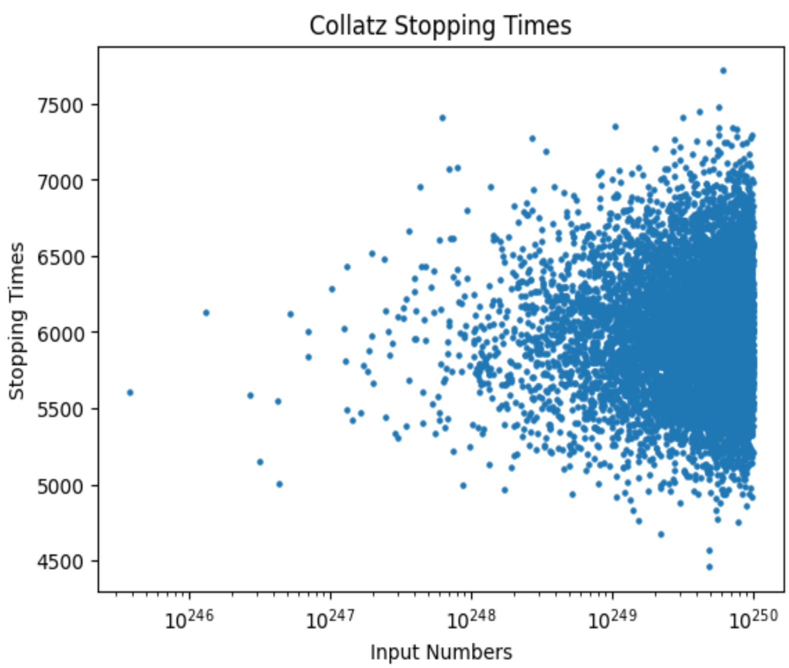
Description: Log-spaced sampling for starting inputs and stopping times
Using log - scale y(axis)



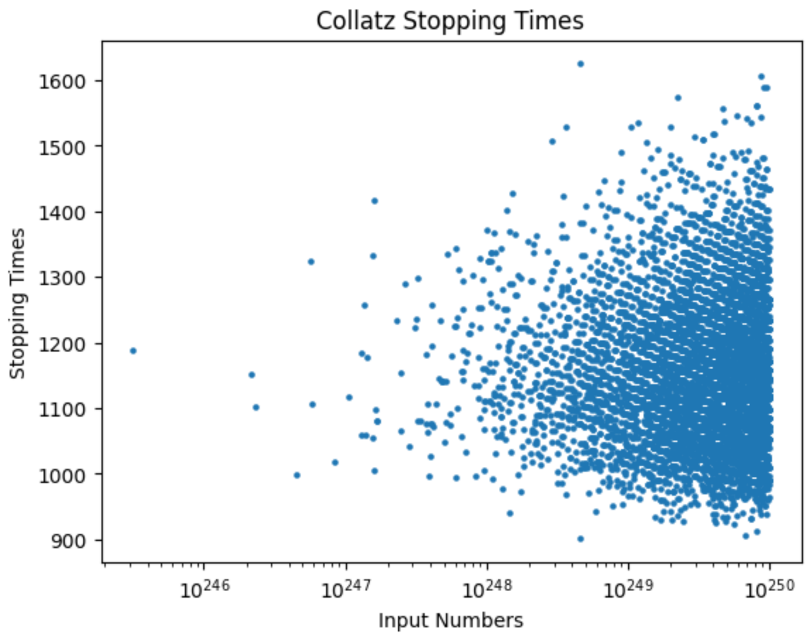
Description: Log-scale x and y-axis



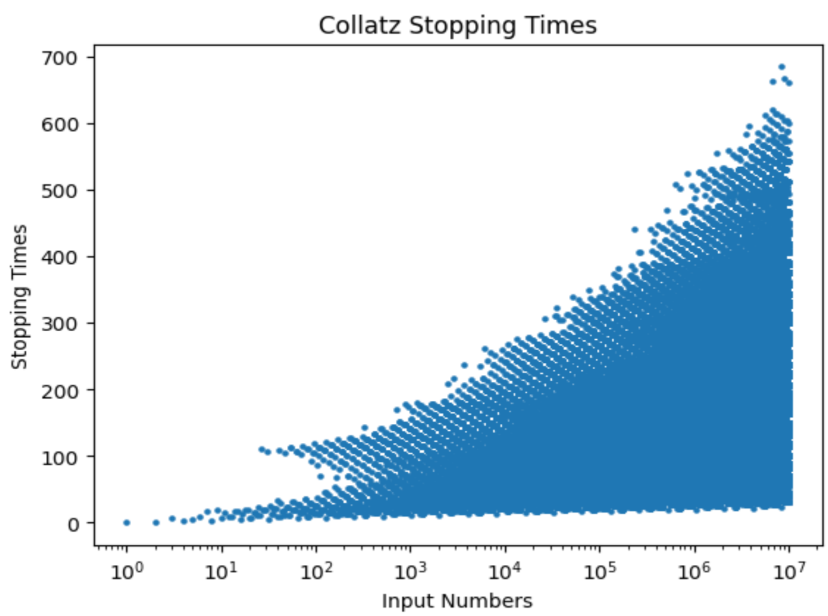
Description: Fit a trend line to the data



Description: Used random sampling technique



Description: Used random sampling with evenly spaced binning technique



Description: Found stopping times for all input numbers