Yash Sharma
CSE13S, Winter 2021

Assignment 1 - WRITEUP.pdf

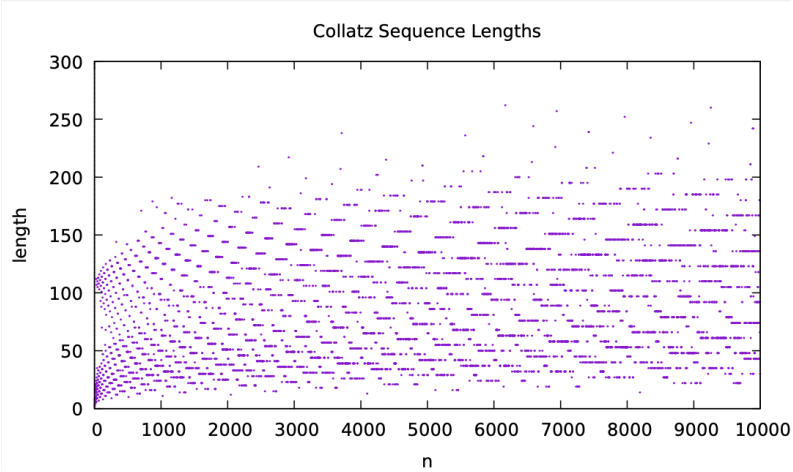
Introduction:

For this writeup, we'll be focusing on how the collatz sequence is utilized by the GnuPlot bash script to print & display the length and max values of a collatz sequence. In addition, the collatz sequence lengths generate a histogram as well using the GnuPlot bash script.

The Collatz Sequence creates a sequence of integers with some positive integers. These integers are then iterated through a for loop based on the previous term of the sequence. This collatz sequence data also helps us produce interesting plots using GnuPlot, which is all automated through the means of a bash script.

Collatz Sequence Lengths:

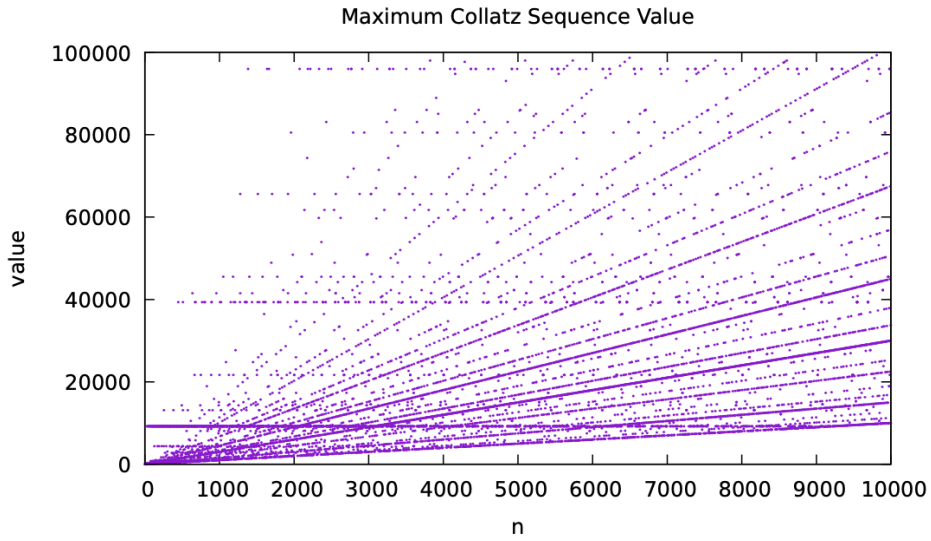
The Correlation between n and length:



This was the data produced by the correlation between n and length. To produce this graph, I wrote a for loop that iterated through any integer between 2 and 10000. This for loop proceeds to the next integer after it prints a new line for each byte and then counts for each file. At the end, all this data is sent to a separate length.dat file which then GNUPlot uses to create a graph. GNU Plot takes in the xlabel, ylabel and plots the graph according to the output file (length.pdf).

Maximum Collatz Sequence Value:

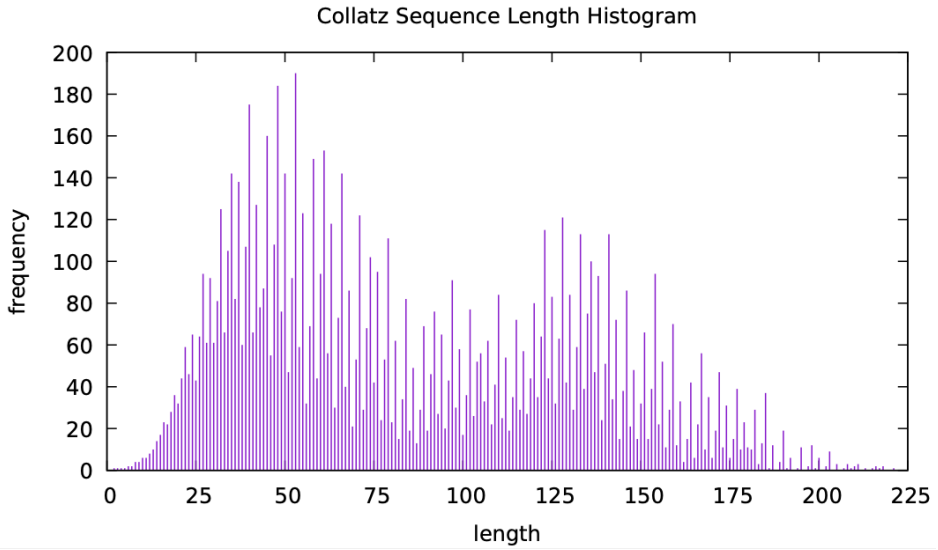
The Correlation between n and value:



This was the data produced by the correlation between n and value. To produce this graph, I once again utilized a for loop that iterated through any integer between 2 and 10000. Alongside this, the loop sent all its values to a maxvalues.dat file which sorted the integers in numeric order with the 1st integer at the tail end of the data being at the top. After this, the data is then sent through GNUPlot which plots this data based on the xlabel, ylabel, and a yrange [0:10000] in the maxvalues.dat file.

Collatz Sequence Length Histogram:

The Correlation between length and frequency:



This was the data produced by the correlation between length and frequency. To produce this graph, I also utilized a for loop that iterated through any integer between 2 and 10000. Along with this, the loop sends all values into histogram.dat after sorting the collatz sequence in numeric order and counting the number of unique occurrences. This data is then sent to hisort.dat (a sorted file created for GNUPlot to pull data and graph). Once the data is in GNUPlot, I added a feature in GNUPlot called “xtics” in which I could set the x-axis tics to be within a range of 25. Finally, GNUPlot charted the graph using impulses on a 2:1 basis as the frequency and length were correlated in the graph.

Conclusion:

From the data shown here, we can make 3 major conclusions:

1. Increasing the collatz sequence lengths stabilizes the value of n throughout the sequence.
2. Increasing the size of n continues to increase the maximum collatz sequence values on the y-axis.
3. Generating a histogram of the Collatz Sequence Lengths shows that there was more frequency exposure within the 25-75 length range than any other length.