A Simple Probabilistic Heuristic Supporting the Collatz Conjecture

Shreyansh Jaiswal

7 August 2024

Abstract

The Collatz Conjecture, also known as the 3x+1 problem, posits that for any positive integer n, the sequence defined by the Collatz function will eventually reach the number 1. This conjecture has been extensively tested for a vast range of values, consistently supporting its validity. In this paper, we explore a probabilistic perspective to provide additional support for the conjecture. We focus on the probability that the Collatz sequence $T(n)$, for any starting value n, reaches a power of 2—an essential step in the sequence's progression toward 1. Our approach suggests that as n tends to infinity, the likelihood of the Collatz conjecture being satisfied becomes very high. This probabilistic argument aligns with the extensive empirical evidence supporting the conjecture and offers a novel perspective on its validity. While not a formal proof, our findings contribute to the broader understanding of the Collatz Conjecture and reinforce the conjecture's plausibility through probabilistic reasoning.

1 Introduction to the Collatz Conjecture

The $3n + 1$ problem concerns the following innocent seeming arithmetic procedure applied to integers: If an integer n is odd then 'multiply by three and add one", while if it is even then divide by two". This operation is described by the *Collatz function*

$$C(n) = \begin{cases} 3n + 1 & \text{if } n \equiv 1 \pmod 2, \\ \\ \frac{n}{2} & \text{if } n \equiv 0 \pmod 2. \end{cases}$$

The $3n + 1$ problem, which is often called the *Collatz problem*, concerns the behavior of this function under iteration, starting with a given positive integer n.

$3n + 1$ **Conjecture.** *Starting from any positive integer n, iterations of the function $C(n)$ will eventually reach the number 1. Thereafter iterations will cycle, taking successive values $1, 4, 2, 1, \ldots$.*

A commonly used reformulation of the $3n+1$ problem iterates a different function, the $3n+1$ *function*, given by

$$T(n) = \begin{cases} \frac{3n+1}{2} & \text{if } n \equiv 1 \pmod 2, \\ \frac{n}{2} & \text{if } n \equiv 0 \pmod 2. \end{cases}$$

From the viewpoint of iteration the two functions are simply related; iteration of $T(n)$ simply omits some steps in the iteration of the Collatz function $C(n)$ $(3n+1)$ for odd n always results in even outputs. The relation of the $3n+1$ function $T(n)$ to the Collatz function $C(n)$ is that:

$$T(n) = \begin{cases} C(C(n)) & \text{if } n \equiv 1 \pmod 2 \text{ ,} \\ C(n) & \text{if } n \equiv 0 \pmod 2 \text{ .} \end{cases}$$

As it turns out, the function $T(n)$ proves more convenient for analysis of the problem in a number of significant ways.

1.1 The hailstone numbers

By the trajectory of n under a function T, we mean the forward orbit of n, that is, the sequence of its forward iterates $(n, T(n), T^2(n), T^3(n), \ldots)$. Figure 1 displays the $3n+1$-function iterates of $n = 27$ plotted on a standard scale. We see an irregular series of increases and decreases,thus leading to these numbers getting the name of "Hailstone numbers" as hailstones form by repeated upward and downward movements in a thunderhead. Our heuristic argument will be regarding these hailstone numbers. (Function of $T^{\alpha}(n)$)

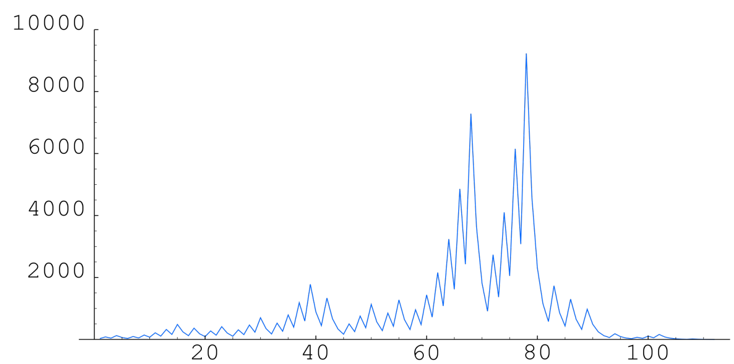

Figure 1: Trajectory of $n = 27$ under the $3n+1$ function plotted on a standard scale. Observe the variablility in the numbers.

A connection to probability theory and stochastic processes arises when one attempts to model the behavior of the $3x+1$ iteration on

large sets of integers. This leads to heuristic probabilistic models for the iteration, which allow predictions of its behavior. Some authors have argued that the iteration can be viewed as a kind of pseudo-random number generator, viewing the input as being given by a probability distribution, and then asking how this probability distribution evolves under iteration.One can model this by a stochastic model corresponding to random tree growth, e.g. a branching random walk.

2 Empirical Analysis, Heuristic Arguments, and Proofs Supporting the Collatz Conjecture

The Collatz Conjecture, also known as the $3n + 1$ problem, has been a subject of extensive empirical investigation and heuristic exploration. This section elaborates on the empirical evidence, heuristic arguments, and partial proofs that collectively support the conjecture.

2.1 Empirical Analysis

Empirical analysis plays a crucial role in understanding the validity of the Collatz Conjecture. To date, the conjecture has been verified for an extensive range of integers. Specifically, computational checks have been performed up to 2^{68} (approximately 2.95×10^{20}), demonstrating that every tested integer eventually reaches the number 1. These verifications involve iterating the Collatz function for an enormous number of starting values and observing the behavior of the resulting sequences.

The results of these empirical tests provide strong evidence in favor of the conjecture.

2.2 Heuristic Arguments

Heuristic arguments offer insights into why the Collatz Conjecture might be true, even though they do not constitute formal proofs. These arguments often involve probabilistic models and statistical observations to provide a plausible rationale for the conjecture's validity.

When considering only the odd numbers in the sequence generated by the Collatz process, it is observed that each odd number is, on average, $\frac{3}{4}$ of the previous one. More precisely, the geometric mean of the ratios of outcomes is $\frac{3}{4}$. This observation yields a heuristic argument suggesting that every Hailstone sequence should decrease in the long run. However, this argument does not address the possibility of other cycles within the sequence but specifically targets the issue of divergence.

It is important to note that this argument does not constitute a formal proof. The heuristic is based on the assumption that Hailstone sequences are constructed from uncorrelated probabilistic events. Thus, while this heuristic provides insight into the behavior of the sequences, it is not a definitive proof of the Collatz Conjecture.

3 Our Basic argument

(The reader must note that, $\sigma_\infty(n)$ here denotes the total number of terms required to reach $T(n) = 1$, or the total stopping time for n, while counting n as the 0-th iterate.)

Our argument states that, for any $\sigma_\infty(n)$ to be:

$$\sigma_\infty(n) < \infty$$

The condition:

$$T^\alpha(n) = 2^m \quad \alpha \in \mathbb{N},\ m \in \mathbb{Z}^+$$

Must be satisfied. This can be understood by the fact that if $T^\alpha(n) = 2^m$, then after applying the rules of the Collatz conjecture, $T(n)$ surely goes down to the 1 - 2 - 4 loop. Also 1,2 and 4 themselves are of the form 2^m. If we rewrite this condition, we get:

$$\log_2 T^\alpha(n) \in \mathbb{Z}^+$$

And, we can also write this condition as:

$$\log_{10} T^\alpha(n) = a(\log_{10} 2), \quad a \in \mathbb{N}$$

The first condition will be the main condition for probabilistic analysis. While the reason the condition $\log_{10} T^\alpha(n) = a(\log_{10} 2)$ is mentioned, will be explained in the further sections.

4 Probabilistic Heuristic analysis to get a final result

The Collatz Conjecture can essentially be restated in the terms of the first condition,

$$\log_2 T^\alpha(n) \in \mathbb{Z}^+$$

If and only if this condition is satisfied for any αth iteration for any seed number n, then only the Collatz conjecture would hold for that n.

The main concept which will be analyzed in this approach is the probability of any term (Hailstone number) satisfying $\log_2 T^\alpha(n) \in \mathbb{Z}^+$ out of all the $T^\alpha(n)$. $(1 \leq \alpha \leq \sigma_\infty(n))$. Even one term satisfying the condition would make the Collatz conjecture true for that n. So, using this information, we can define ourselves an event:

$$A_{i=\alpha} = \{T^\alpha(n) \mid T^\alpha(n) = 2^m \text{ for some } m \in \mathbb{Z}^+\}$$

If we do an simple analysis, we can understand the fact that, for any i which satisfies the condition $T^{\alpha}(n) = 2^m$, the next i also satisfies the condition, the next also, and so on till the last event/number. $(i = \sigma_\infty(n))$. This implies that the events A_i are not mutually exclusive nor independent. Hence, directly adding probabilities of each $T^{\alpha}(n)$ satisfying the condition would definitely overestimate our results. Hence, we apply the Inclusion Exclusion principle regarding probability. However, the reader must note that, these arguments (or more specifically, any generalised arguments regarding this approach) are not a proof because it assumes that Hailstone sequences are assembled from uncorrelated probabilistic events. These arguments are wise heuristics.

4.0.1 Applying Inclusion Exclusion principle regarding probability

The Inclusion-Exclusion Principle is a fundamental concept in probability theory and combinatorics used to find the probability of at least one event occurring among a given set of events. For n events $A_1, A_2, \ldots, A_n$, the Inclusion-Exclusion Principle generalizes to:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right)$$

where:

- The first sum $\sum_{k=1}^{n}$ iterates over the number of events included in the intersection.
- The second sum $\sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n}$ runs over all possible intersections of k events.
- $P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right)$ is the probability of the intersection of k events.

4.0.2 Our specific application for the concept

If we recall section 3, then we can say the following points:

- If A_{i_1} is true (Probability is 1), then $A_{i_2}, \ldots, A_{i_k = \sigma_\infty(n)}$ are also true. - Thus, the intersection of events where $i_1 \leq i_2 \leq \cdots \leq i_k$ simplifies to:

$$P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) = P\left(A_{i_1}\right)$$

This is because if A_{i_1} is true, all subsequent events in the intersection are also true. Applying the Formula

The Inclusion-Exclusion Principle simplifies to:

$$P\left(\bigcup_{i=1}^{\sigma_\infty(n)} A_i\right) = \sum_{k=1}^{\sigma_\infty(n)} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq s} P\left(A_{i_1}\right)$$

This is effectively summing the probabilities of the earliest events in each combination of intersections, adjusting for our specific case. Thus,we can effectively say that for any n satisfying the Collatz's Conjecture, the following condition is true:

$$P\left(\bigcup_{i=\alpha=1}^{\sigma_\infty(n)} A_i\right) = 1$$

4.0.3 Developing a more generalised result/interpretation

Before starting this subsection, the reader must note the use of another wise heuristic in this section, which is that $P(A_i) > 0$ for all i. This heuristic is obviously supported by the large number of n which satisfy the Collatz's conjecture. (Each n which satisfies the Collatz's conjecture, must have at least one A_i event being satisfied, thus making $P(A_i) > 0$ for that n.) This is also the reason why the author expressed the original conditions as $\log_{10} T^\alpha(n) = a(\log_{10} 2), \quad a \in \mathbb{N}$, because they believe that this observation could help to convert this heuristic into a formal statement.

Behavior as $\sigma_\infty(n) \to \infty$

If $P(A_i) > 0$ for all i, then:

$$\lim_{\sigma_\infty(n) \to \infty} P\left(\bigcup_{i=1}^{\sigma_\infty(n)} A_i\right) = 1$$

This is because the probability of missing out on all events tends to zero as $\sigma_\infty(n)$ increases. In other words, the probability of at least one of the events occurring becomes almost certain as the number of events goes to infinity.
The reader must note the fact that $\sigma_\infty(n) \to \infty$ also is an equivalent statement to $n \to \infty$, as both $\sigma_\infty(n)$ and $n \to \infty$

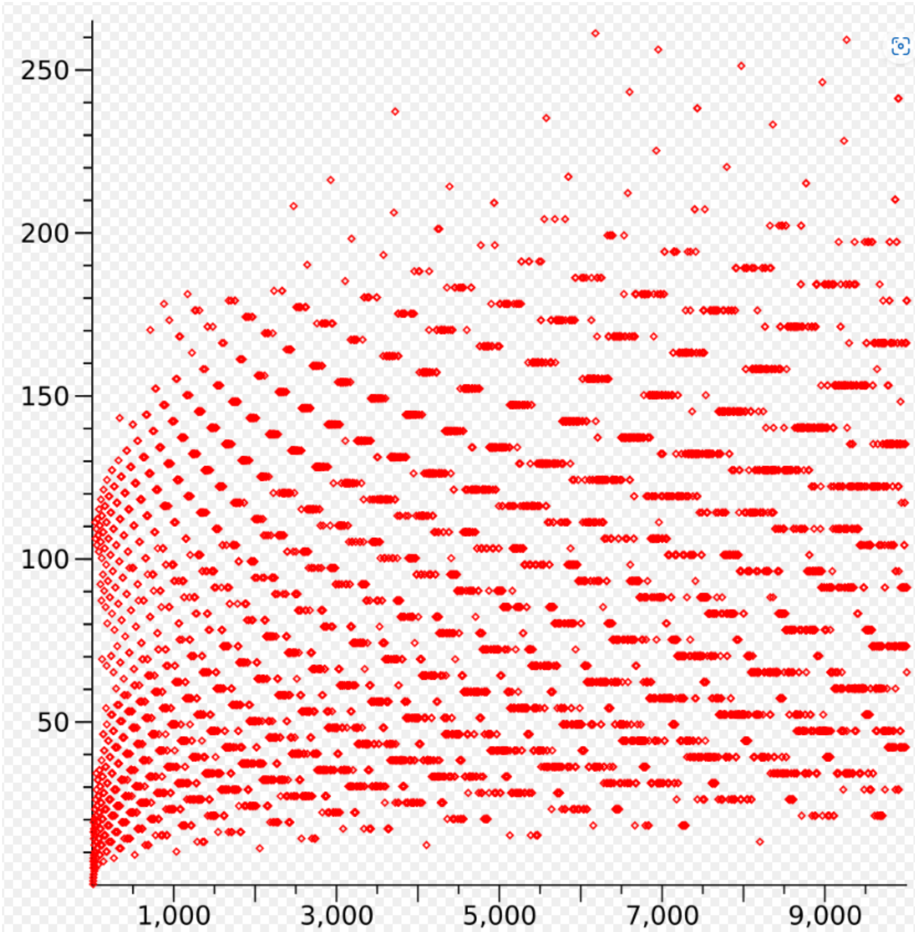

Figure 2: Iteration time versus n for numbers from 1 to 9999. The graph shows the total stopping time for each number in this range. Observe how $\sigma_\infty(n)$ increases with n

This also means:

$$\lim_{\sigma_\infty(n)\to\infty} P\left(\bigcap_{i=1}^{\sigma_\infty(n)} A_i^c\right) = 0$$

This is because the probability of not having any of these events occur becomes very small as the number of events grows infinitely.
These results/heuristics suggest a fact that the Collatz conjecture holds true even as $n \to \infty$.

5 An argument in favour of the conjecture based on a probabilistic contradiction

This section works upon building a framework for a potential "proof" of the conjecture by contradiction. First consider that:

If $P(A_i) > 0$ for all i, then:

$$\lim_{\sigma_\infty(n)\to\infty} P\left(\bigcup_{i=1}^{\sigma_\infty(n)} A_i\right) = 1$$

So, for starting the proof, let's consider a hypothetical seed, n_d, for which:

$$T^\alpha(n_d) \neq 1.$$

This means that, for this n_d seed, the function never converges to one. Thus,

$$\sigma_\infty(n_d) = \infty$$

If the existence for such n_d is true, then, this asserts that:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = 0$$

as 0 as the probability asserts the impossibility of any $T^\alpha(n)$ being equal to $2^m, m \in \mathbb{N}$. However, if you look at the first condition/limit of this section, we have:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = 1 \neq 0$$

The contradiction in values of the probability has arisen due to the fact that the assumption of any seed $n_d, n_d \in \mathbb{N}$ is done. Thus, our analysis suggest that, at least by probability,that the collatz conjecture holds true for all n where $1 < n < \infty$ and $n \in \mathbb{N}$

6 Towards a more justified and rigorous argument

This section will just discuss the main arguments from the previous section, but in more detail, depth and justifications.

If we remember the concepts/discussions of Section 5, we know that for any n_d, $n_d \in \mathbb{N}$ which makes $T^\alpha(n_d)$ diverge to ∞, must have,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = 0$$

strictly as, the probability being 0 asserts the impossibility of any $T^{\alpha}(n_d)$ being equal to $2^m, m \in \mathbb{N}$. So, in order to move forward, we need to show that:

$$P(A_i) = 0$$

to prove the existence of any n_d. Vice-versa, we will have to prove that:

$$P(A_i) > 0$$

to disprove the existence of any n_d. An important thing to note is that, there isn't any lower bounds for $P(A_i)$, $(P(A_i)) > \gamma > 0)$, as if $P(A_i) > 0$, then these non zero values would definitely add up in $P\left(\bigcup_{i=1}^{\infty} A_i\right)$, as $\sigma_{\infty}(n)$ for n_d is ∞, thus making the probability converge to 1 as $\sigma_{\infty}(n) \to \infty$ if $P(A_i) > 0$.

6.0.1 Working on proving definite values for $P(A_i)$

Look at $T^{\alpha}(n)$,

$$T(n) = \begin{cases} \frac{3n+1}{2} & \text{if } n \equiv 1 \pmod 2, \\ \frac{n}{2} & \text{if } n \equiv 0 \pmod 2. \end{cases}$$

If we look at any iteration of $T^{\alpha}(n)$ where $\alpha > 1$

$$T^{\alpha}(n) = \begin{cases} \frac{3T^{\alpha-1}(n)+1}{2} & \text{if } T^{\alpha-1}(n) \equiv 1 \pmod 2, \\ \frac{T^{\alpha-1}(n)}{2} & \text{if } T^{\alpha-1}(n) \equiv 0 \pmod 2. \end{cases}$$

if we look at A_i, we are essentially looking for the probability that

$$T^{\alpha}(n) = 2^m (m \in \mathbb{N})$$

is satisfied. So, let's look again at $T^{\alpha}(n)$ related with A_i. Before that, let's establish some notions that we must be aware of.

$$T^{\alpha}(n) \in \mathbb{N}$$

and

$$2^m \in \mathbb{N}$$

If $P(A_i)$ could be 0, then:

$$\frac{3(T^{\alpha-1}(n)) + 1}{2} \neq 2^m$$

and

$$\frac{T^{\alpha-1}(n)}{2} \neq 2^m$$

as,$T^{\alpha}(n) \in \mathbb{N}$ and $m \in \mathbb{N}$, we can use $\mathbb{N}$ as a 'variable' to represent the natural numbers. Therefore:

$$\frac{3(\mathbb{N}) + 1}{2} \neq 2^{\mathbb{N}}$$

and:

$$\frac{\mathbb{N}}{2} \neq 2^{\mathbb{N}}$$

So, for validating the 'claim' that $P(A_i) = 0$, we need to show that the conditions (at least one of them) is true. If we look at:

$$\frac{\mathbb{N}}{2} \neq 2^{\mathbb{N}}$$

$$\mathbb{N} \neq 2^{\mathbb{N}+1}$$

As this condition is blatantly wrong, we can look at the other condition,

$$\frac{3(\mathbb{N})+1}{2} \neq 2^{\mathbb{N}}$$

$$3(\mathbb{N})+1 \neq 2^{\mathbb{N}+1}$$

$$(\mathbb{N}) \neq \frac{2^{\mathbb{N}}-1}{3}$$

In order to disprove this false statement, we need to show that, there exist $\mathbb{N}$, which satisfy:

$$2^{\mathbb{N}} - 1 \equiv 0 \pmod 3$$

The simplest way to do this is by presenting a counterexample. Look at $\mathbb{N} = 2$, which gives 4 - 1 = 3, satisfying the condition. This asserts that, the last remaining condition for asserting $P(A_i) = 0$ is wrong. However, some may argue that, for some $\mathbb{N}$, the condition $2^{\mathbb{N}} - 1 \not\equiv 0 \pmod 3$ is satisfied. However, we must remember that, in the case of potential existence of n_d, $\sigma_\infty(n) = \infty$, thus the number of $T^\alpha(n) = \infty$, thus asserting that there exist $\mathbb{N}$, which obviously are not the whole of the majority of the natural numbers, satisfying the condition is not enough. The whole of the natural numbers must satisfy the condition in order to confidently assert the notion that $P(A_i) = 0$.

Hence, as the minimum conditions for confidently asserting that $P(A_i) = 0$ are wrong, we can now confidently assert that $P(A_i) > 0$. Applying this in:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)$$

we get the limit as:

$$\lim_{\sigma_\infty(n)\to\infty} P\left(\bigcup_{i=1}^{\infty} A_i\right) = 1 \neq 0$$

These results align with the general notion within the math community that the Collatz conjecture is "satisfied" for majority of the numbers. However, before concluding this section, we would like to provide a brief summary for the heuristics used in this text.

7 Heuristics

1. **Heuristic 1** The first heuristic used in the text is that, $P(A_i) > 0$ is satisfied for **all** $\mathbb{N}$. A more analytic approach can help us to modifiy this statement to make it more correct for different n. However, as we are majorly dealing with $n \to \infty$, we can say that this heuristic is acceptable.

2. **Heuristic 2** The second heuristic is that we assume a entirely random system of the hailstone numbers.Again, this can also be improved through more detailed analysis. However, as mentioned before, as we are dealing primarily with $n \to \infty$, we presume that this heuristic is also acceptable.

Hence, through **probabilistic analysis and carefully considered heuristics**, we arrive at the conclusion that **the likelihood of the Collatz conjecture remaining satisfied as $n \to \infty$ is extremely high.**

8 Conclusion

This paper presents a probabilistic heuristic to support the Collatz Conjecture by analyzing the likelihood of the sequence $T(n)$ reaching a power of 2. We have demonstrated that:

$$\lim_{\sigma_\infty(n) \to \infty} P\left(\bigcup_{i=1}^{\sigma_\infty(n)} A_i \right) = 1$$

where A_i denotes the event that $T^\alpha(n)$ equals a power of 2. This result implies that, as n and $\sigma_\infty(n)$ increase indefinitely, the probability of the Collatz Conjecture being satisfied converges to 1.

Our findings reinforce the conjecture's plausibility, aligning with extensive empirical evidence and providing a probabilistic framework for understanding its validity. While this heuristic approach does not offer a formal proof, it strengthens the case for the Collatz Conjecture and highlights the need for further research. Future work could aim to refine these probabilistic arguments or seek more rigorous proofs to advance our understanding of the Collatz Conjecture.

9 References

1. **Jeffrey C. Lagarias**, *The 3x+1 Problem: An Overview*, arXiv:2111.02635v1, 2021. Available at: `https://arxiv.org/abs/2111.02635v1`

2. **Wikipedia Contributors**, *Collatz Conjecture*, Wikipedia, 2024. Available at: `https://en.wikipedia.org/wiki/Collatz_conjecture#Supporting_arguments`