Validating Collatz Conjecture through Binary Representation and Probabilistic Path Analysis

Budee U Zaman

February 2024

Abstract

The Collatz conjecture, a longstanding mathematical puzzle, posits that, regardless of the starting integer, iteratively applying a specific formula will eventually lead to the value 1. This paper introduces a novel approach to validate the Collatz conjecture by leveraging the binary representation of generated numbers. Each transition in the sequence is predetermined using the Collatz conjecture formula, yet the path of transitions is revealed to be intricate, involving alternating increases and decreases for each initial value.

The study delves into the global flow of the sequence, investigating the behavior of the generated numbers as they progress toward the termination value of 1. The analysis utilizes the concept of probability to shed light on the complex dynamics of the Collatz conjecture. By incorporating probabilistic methods, this research aims to unravel the underlying patterns and tendencies that govern the convergence of the sequence.

The findings contribute to a deeper understanding of the Collatz conjecture, offering insights into the inherent complexities of its trajectories. This work not only validates the conjecture through binary representation but also provides a probabilistic framework to elucidate the global flow of the sequence, enriching our comprehension of this enduring mathematical mystery.

1 Introduction

The Collatz conjecture, a perplexing enigma in the realm of mathematics, postulates that, regardless of the initial integer chosen, the iterative application of a specific formula will inevitably lead to the value 1. While this longstanding puzzle has captivated the mathematical community for decades, this paper introduces a groundbreaking approach to validate the Collatz conjecture by delving into the binary representation of the generated numbers.

In this study, we explore the intricacies of the Collatz conjecture's behavior by scrutinizing the binary patterns inherent in the sequence. The transitions in the sequence are predetermined by the Collatz conjecture formula, yet the path-way of these transitions unveils a fascinating interplay of alternating increases and decreases for each initial value.

Our research goes beyond the traditional examination of the Collatz problem, offering a unique perspective that leverages binary representation to elucidate the meaning of operations within the sequence. The investigation then extends to scrutinize whether the observed behavior aligns with the expectations set forth by the Collatz conjecture.

Furthermore, this paper employs the concept of probability to illuminate the complex dynamics governing the convergence of the sequence towards the termination value of 1. By integrating probabilistic methods, our research aims to unravel the underlying patterns and tendencies that shape the global flow of the Collatz sequence.

The insights derived from this study contribute to a deeper understanding of the Collatz conjecture, offering a dual perspective that not only validates the conjecture through binary representation but also provides a novel probabilistic framework. This framework, in turn, sheds light on the inherent complexities of the trajectories within the Collatz sequence, enriching our comprehension of this enduring mathematical mystery. [1] [2] [4] [5] [3]

2 Unveiling the Power of Positive Integers in Digital Form

The digital form of positive integers is commonly represented in binary, where each digit is called a bit. A bit can take on one of two possible values: 0 or 1. The rightmost digit holds the least significant value, with each subsequent digit representing higher powers of 2 as you move towards the left.

In the example "bin11001," the rightmost digit is 1, followed by 0, 0, 1, and the leftmost digit is 1. To convert this binary representation to decimal, you sum the contributions of each digit multiplied by the corresponding power of 2. In this case.

$$1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 16 + 8 + 0 + 0 + 1 = 25$$

Therefore, "bin11001" in binary is equivalent to the positive integer 25 in decimal notation. This binary-to-decimal conversion relies on the positional value of each bit, with higher bits contributing to larger powers of 2 in the final decimal result.

2.1 achievement

Collatz Conjecture Operations Breakdown:

Starting Point

Begin with a positive odd integer, denoted as n.

(i) Doubling Operation: $n_1 = 2n..(Doubling)$

(ii) Increment Operation: $n_2 = n_1 + n......(Increment)$

(iii) Add One Operation: $n_3 = n_2 + 1$..(*Add One*)

(iv) Even Division Operation: If n_3 is even, perform m times division until n_4 becomes odd: $n_4 = \frac{n_3}{2^m}$ (Even Division) Iteration Process: Repeat the operations (ii) and (iii) until n_4 becomes 1.

Summary of Steps

1. $n_1 = 2n$..(*Doubling*)

2. $n_2 = n_1 + n$......(*Increment*)

3. $n_3 = n_2 + 1$..(*Add One*)

4. If n_3 is even, perform times division until n_4 becomes odd: $n_4 = \frac{n_3}{2^m}$ (Even Division)

5. Repeat steps 2-4 until $n_4$becomes 1.

This sequence of operations continues until the number reaches 1, and the Collatz Conjecture posits that, regardless of the starting odd integer, this iterative process will always eventually lead to the value 1

2.2 Behavior and meanings of each calculation

Problem Statement

Investigate the behavior related to the calculation of (i), (ii), (iii), and (iv) using binary representation. The investigation is focused on the relationship between n_1, n_2, n_3, and n_4, where $n_1 = 2n$.

Definitions

• n_1: The number obtained by shifting n too left by 1 digit in binary characterization.

• n_2: Obtained by adding 1 to an odd number, making n_3.

• n_3: Always an even number, with a carry-over to digit 2.

• n_4: Obtained by shifting n_3 one digit to the right.

Operations

(i) Multiple Make Odd Number (MMON):

• $\frac{1}{2^m}$ where m is an integer.

Define: MMON = R_1 (ii) Sample Calculation: Given: $n = \text{bin}1011$ Calculate: $n_1 = \text{bin}101110$

Other way of writing

• $n_3 = \text{bin}10110$

• $n_4 = \text{bin}1011$

• $R_1 = \frac{1}{2}$

Additional Information

• Adding 1 to an odd number n_2 always results in an even number n_3.

• $n_4 = \frac{n_3}{2}$ involves shifting n_3 one digit to the right.

		1	1	1	
	1	1	1	0	add
1	0	1	0	1	

3 Highlight and Recognition of the Operations

Let's break down the information into (i)(ii)(iii) and express it in a more structured way:

(a) Identified Function

The function can be represented as follows: y=3x+1=2x+x+1 (v) Where:

- x is a positive odd integer.
- y is a positive even integer.

(b) Linear Congruential Generator (LCG)

The LCG can be expressed as follows: $R_2 = 3$, $R_1 = \frac{1}{2^3}$, Here, R_2 Number of Most Correct Continuous Value 0 Digits (NMRZD)

denoted as NMRZD= R_2

(iii) **Random Number Generator**

The next random number generator utilizes a Linear Congruential Generator (LCG). The value of the next input x is causally related to bit two , to i-1 of the current input the x here.

(c) Sample

Sample: 101011000

R_2=3, , $R_1 = \frac{1}{2^3}$, This implies that the random number generator is utilizing the sample with for this R_2

4 Quality of y Generated

Beyond the numbers , A comprehensive analysis of function y results.
Let +ve O(odd) integer x has i , digits so we calculate as following.

digit no.	$i+2$	$i+1$	i	$i-1$				1
$x+$			1	.	.	.	.	1
$2x+1$		1	.	.	.	.	1	1
$=y$	.	.	.	.	.	.	.	0

TABLE ...1

we examine the quality of the output (y) generated by a function applied to a positive odd integer x with i digits. The calculation involves specific rules, and the resulting y exhibits distinct characteristics.

Processing Rules

1.**Digit Equality**

The first digit of x is equal to the second digit of 2x, both being 1.

2. **Initial Digits**

The first digit of 2x is the same as the first digit of y, and it is set to 0.

3. **Pattern in Digits**

For digits i and (i+1) of x, and for all digits j where $2 \leq j \leq i-1$, certain **Conditions apply**

- Digit i of x equals , digit (i+1) of 2x and is set to 1.
- Digit j of x equals , digit (j+1) of 2x, with values of 0 or 1.

Characteristics of y

The output y is a result of the specified processing rules. It is constructed based on the digit relationships outlined above, providing a unique pattern to the output. The adherence to these rules is essential for the correct functioning of the given algorithm.

The outlined rules govern the transformation of a positive odd integer x into the output y through a defined process. Evaluating the quality of y involves ensuring that these rules are meticulously followed, leading to a reliable and consistent outcome. This assessment lays the foundation for understanding the performance and reliability of the function in transforming input data.

Analysis for y Structure when Carry Over or x's Digit i-1 is 1

The structure of y follows a specific pattern under certain conditions. If there's a carry-over from the previous digit (i-1) or if the digit (i-1) of x is 1, y takes on a distinct form from digit (i+2) to 1, represented as y = bin 1...0.

This pattern becomes more noticeable when the number of digits (i) in x is even, and the rightmost i-1 digits of y are all set to 0. The reason behind this is to ensure that the digit structure of x aligns with table 2, leading to those specific zeros in y.

Additionally, when the digit (i-1) in y is equal to 0, the value of digit i for y is consistently 1. This is because the structure of x needs to match with table 2, requiring the rightmost i-1 digits of y to be zeros. Consequently, digit i of y is always guaranteed to be 1, never taking the value of 0. This condition ensures that the maximum possible value for R_2 in this context is i-1.

Therefore
maximum R_2 is i-1.

digit no.	$i+2$	$i+1$	i	$i-1$	8	7	6	5	4	3	2	1
$x+$			1	1	0	1	0	1	0	1	0	1
$2x+1$		1	1	0	1	0	1	0	1	0	1	1
$=y$	1	0	1	0	0	0	0	0	0	0	0	0

TABLE2

Explanation of the Conditional Relationship between the Digits of x and y for Maximum R_2

In the given scenario, the relationship between the number of digits in x and specific digits in y leads to a distinctive pattern, ultimately influencing the maximum value of R_2.

Conditions

(a)The number of digits (i) in x is odd.
(b)The rightmost i-1 digits of y are all 0.
(c)The digit i and i+1 for y are always 0

Logical Structure

The structure of the digits in x is carefully designed, ensuring that the rightmost i-1 digits of y are exclusively 0. Consequently, this condition sets the stage for specific characteristics in the digit values of y.

Inference

The critical observation is that when the above conditions are met, digit i and i+1 of y are guaranteed to be 0. The significance of this lies in the prevention of these digits from being 1, thereby establishing a pattern with implications for the maximum value of R_2.

The intricate relationship between the odd number of digits in x and the specific arrangement of zeros in the rightmost i-1 digits of y leads to a predictable structure. This structured pattern ensures that digit i and i+1 of y are always 0, strategically influencing the outcome of R_2 in a manner conducive to the specified conditions.

digit no.	$i+2$	$i+1$	i	$i-1$	7	6	5	4	3	2	1
$x+$			1	0	1	0	0	1	0	1	1
$2x+1$		1	0	1	0	1	0	1	0	1	1
$=y$	1	0	0	0	0	0	0	0	0	0	0

TABLE 3

Analyzing digits, A structured approach to calculating R_2 for each variable y.

5 Consider the four Hypothesis

Hypothesis 1

Let's clarify the information:

Given the structure of the digits in a number y:

Digit 2 of y is 1. Digit 1 of y is 0.

You have defined two conditions:

R_2 (which seems to represent the result when shifting the number to the right) is 1. R_1 is $\frac{1}{2}$. Now, let's discuss the meaning and the number of hypotheses for x:

Meaning of Digits Structure

.The structure suggests that the number y can be divided by 2, as indicated by the condition $R_1 = \frac{1}{2}$.

.The fact that Digit 2 of y is 1 suggests that the number can be shifted one digit to the right, as indicated by the condition $R_2 = 1$.

Number of Hypotheses for x

Hypotheses satisfying the conditions

The number of hypotheses that meet the criteria, where R_2 equals 1, is given by the expression 2^{i-3}. In this context, the leftmost digit is 1, the rightmost digit is also 1, and the digit at position 2 is 1. The remaining i-3 digits can take on values of either 0 or 1.

Total Possible Hypotheses

Total number of possible hypotheses:2^{i-3}

(Most , L(light) digit i = one, most, R(right) digit one = one, and so Other i-2 digits can have values zero or one.) The NO(number) of hypotheses satisfying the given conditions is 2^{i-3} , and the total possible hypotheses are 2^{i-2}.

Please note that the variable 'i' is mentioned in your explanation, but its specific value or context is not defined. If 'i' represents the total number of digits in the number y, then the explanations provided would be appropriate. If you have a specific value for 'i', you can substitute it into the expressions accordingly.

Therefore

Prospect rate for this ;

$$\frac{number\ of\ hypothesis\ which\ is\ satisfied\ this\ hypothesis}{number\ of\ all\ possible\ hypothesis} = \frac{2^{i-3}}{2^{i-2}} = \frac{1}{2}$$

Hypothesis 2

Let given

- Digit 3 of y is one
- Digit 2 of y is zero
- Digit 1 of y is zero

Binary representation of y: bin1...100bin1...100 Now, let's calculate $R_2$for this hypothesis: R_2 =2
Similarly, following the same approach as Hypothesis 1: $R_1 = \frac{1}{2^2}$
The possibility rate of Hypothesis 2 is given by: Possibility rate= $\frac{1}{2^2}$
This hypothesis suggests that the binary representation of y is bin1...100bin1...100, and the likelihood of this hypothesis is$\frac{1}{4}$. Further analysis and comparisons can be made based on these values to determine the most probable hypothesis.

Hypothesis 3

In general (including Hypothesis 1, Hypothesis 2, so that $1 \leq j \leq i-2$ digit(j+1) of y is equal to 1, digit 1 to j of y is zero,

R_2=j $\quad R_1 = \frac{1}{2^j}$

Possibility rate of this hypothesis =$\frac{1}{2^j}$

Hypothesis 4

Describing a mathematical scenario involving two sequences, x and y, where the digits of y are determined based on the parity (even or odd) of the corresponding digits in x. You've introduced two cases:

1. If digit number i of x is even, digit i of y is 1.

- R_2 is i-1.
- $R_1 = \frac{1}{2^{i-1}}$

2. If digit number i of x is odd, digit i and digit i+1 of y are 0.
• Actual R_2 is i+1.
• $R_1 = \frac{1}{2^{i+1}}$

It appears that you are considering the probability of these cases, denoted as "Possibility rate," and you have mentioned Table 2 and Table 3 in relation to this.

6 Analysis of the Evidence

Evidence of the analysis for digits shape of y Total possibility of rate S is this
Sum=$\frac{1}{2^1} + \frac{1}{2^2} + ... + \frac{1}{2^{i-2}} + \frac{1}{2^{i-2}}$
Sum=1 So that analysis lists up all hypothesis regarding to R_2

7 Expected R_2

So that Hypothesis 1,2,3,4 and the fact that number of digits for x can be even or odd equally, expected R_2 is calculated on following formula.

Expected

$$R_2 = \frac{1}{2}[expected\ R_2\ for\ even\ i + expected\ R_2\ for\ odd\imath]$$

Possibility rate for = Q (denoted this Q)

$$R_2 = \frac{1}{2}\sum_{j=1}^{i-)}(R_{2_j} \times wR_{2_j}) + (R_{2_{(i-1)}} \times wR_{2_{(i-1)}}) + \frac{1}{2}\sum_{j=1}^{(i-2)}(R_{2_j} \times wR_{(2_j)}) + (R_{2_{(i+1)}} \times wR_{2_{(i+1)}})$$

$$= \sum_{j=1}^{i-2}(R_{2_j} \times wR_{2_j}) + \frac{1}{2}(R_{2_{(i-1)}} \times wR_{2_{(i-1)}}) + \frac{1}{2} + (R_{2_{(i+1)}} \times wR_{2_{(i+1)}})$$

So that

$$R_2 = \frac{1}{2^1} + \frac{1}{2^2} + ... + \frac{i-2}{2^{i-2}} + \frac{1}{2}(\frac{i-1}{2^{i-2}} + \frac{i+1}{2^{i-2}})$$

$$R_2 = (2 + \frac{i}{2^{i-2}}) + \frac{i}{2^{i-2}} = 2$$

This calculation result does not depend on i, it means expected R_2 of output from result A is B for all positive odd integers. Therefore after result A is executed, y can be divided by $2^2 (R_2 = \frac{1}{2^2})$ as an average during result B.Then next n becomes

$$\frac{3n+1}{2^2}........(C)$$

Generally for integer $n > 1$, following relation can be satisfied

$$\frac{3n+1}{2^2} < n.........(D)$$

8 Proving Collatz Conjecture by Analyzing Operations A and B

The Collatz Conjecture, a long-standing mathematical puzzle, posits that iterating two operations, A and B, on any positive integer will eventually lead to the number 1. In this discussion, we delve into the dynamics of these operations and demonstrate how their combination statistically reduces the average value of the given integer, reinforcing the conjecture.

Operations A and B

Operation A is defined as 3n + 1, which tends to increase the value of n or number of digit's. On the other hand, Operation B is the division of n by 2, effectively decreasing the number of digits when $R_2 > 0$.

Analyzing the Impact

Repetition of A followed by B consistently decreases the value of the n on average. The iterative nature of these operations implies that the process continues until n reaches the minimum positive integer, 1. This is achieved by dividing by 2^m, ultimately resulting in the minimum number of digits, 1.

Expectation and Reality

While the expected value is statistically one, it is crucial to note that it becomes a certainty only after a sufficient number of iterations. Operation A, with its 3n + 1 increment, contributes to both an up in the value of n and an increase in the number of digits. work B, the division by 2, decreases the number of digits, particularly when $R_2 > 0$.

Exceptions and Looping

Despite the compelling logic, exceptions can occur due to the possibility of looping. If a loop occurs, it prevents the completion of enough iterations of A and B, obstructing the realization of the conjectured target. Therefore, while the presented argument provides strong evidence for the validity of the Collatz Conjecture, the potential for looping introduces an element of caution.

The combination of operations A and B, when iterated sufficiently, supports the Collatz Conjecture by demonstrating a consistent decrease in the average value of n. The exception, in the form of potential looping, acknowledges the need for further exploration and verification. Nonetheless, the overall analysis strengthens the case for the Collatz Conjecture's correctness.

Investigation of Looping Conditions

The given loop is represented as follows:

$$n_1, n_2, n_3, \ldots, n_{m-1}, n_m \quad n_m = n_1 \quad (E)$$

There are two required conditions for looping:

On (C), the relation is $n_1 > n_m$ on average, but $n_1 \cong n_m$ is also required for $n_1 = n_m$ to hold.

Regarding (1), on (C), m iterations make n_m in an average as follows from n_1:

$$n_m \cong \left(\frac{3n_1+1}{2^2}\right)^m \cong \left(\frac{3}{2^2}\right)^m \cdot n_1 = \left(\frac{3}{4}\right)^m \cdot n_1 \quad (n \gg 1)$$

This decrease in n requires an increasing factor to accomplish (E). The increase factor is only for $R_2 = \frac{1}{2}$, based on hypothesis 3. In such a hypothesis, the next n is given by $\frac{3n+1}{2}$.

When the existence rate of this factor is about 0.4 of the total iteration m, $n_1 = n_m$ is realized with the following calculation:

$$n_m \cong \left(\frac{3n_1+1}{2^2}\right)^l \left(\frac{3n_1+1}{2^2}\right)^{m-l}$$

$$n_m \cong \left(\frac{3}{2^2}\right)^l \cdot \left(\frac{3}{2}\right)^{m-l} \cdot n_1$$

$$n_m \cong \left(\frac{3}{2^2}\right)^{0.6m} \cdot \left(\frac{3}{2}\right)^{0.4m} \cdot n_1 \quad (F)$$

Because on (1), the occurrence possibility of $\frac{3n+1}{2^2}$ is $\frac{1}{2}$, the possibility for $n_1 = n_m$ is $\left(\frac{1}{2}\right)^{0.4m}$. This depends on the total iterations.

For illustration

$$(\frac{1}{2})^{0.4m} = (\frac{1}{2})^{10} = 0.06 \qquad when\ total\ iteration = 10$$

$$(\frac{1}{2})^{0.4m} = (\frac{1}{2})^{100} = \frac{9}{10^3} \qquad when\ total\ iteration = 100$$

Probability Analysis of Digit Patterns in Binary Sequences

In this analysis, we explore the probability of specific digit patterns occurring in binary sequences, particularly focusing on sequences where the number of digits is denoted by 'i.' The hypothesis suggests that if the i-th digit of a sequence is 1, and all other digits can take values of 0 or 1, the probability of obtaining a sequence close to 2^{i-1} diminishes as i increases.

Probability Calculation

For a given i, the probability of having the same digit sructure ($n_1 = n_m$)is determined by the likelihood of the remaining i-2 digits (excluding the i-th digit and the first digit) taking values of 0 or 1. This probability is expressed as $(\frac{1}{2})^{i-2}$.

Impact of i on Probability

As i increases, the probability of obtaining the same digit structure decreases. For instance, if i is large, such as i = 35 (representing sequences around a billion), the probability becomes significantly small, approximately $\frac{1}{10^{10}}$. This observation implies that longer spans in looping sequences have a lower likelihood of occurring.

Relation to Number Size

The analysis suggests that larger numbers with more digits have a smaller probability of exhibiting the same digit structure. This conclusion aligns with the intuition that as the magnitude of the number increases, the chances of specific digit patterns repeating decrease exponentially.

Formation of Loops

The analysis also hints at the entry of the value 1 into looping sequences. These loops manifest as span 1, where i = 1, creating a unique case in the overall probability analysis.

The probability of obtaining specific digit patterns in binary sequences is inversely proportional to the number of digits in the sequence. As the digit span increases, the likelihood of identical digit structures decreases, emphasizing the diminishing probability of longer span looping occurrences. The relationship between the size of the number and the probability of specific digit patterns provides insights into the distribution and behavior of binary sequences.

9 Conclusion

In conclusion, the Collatz conjecture, while theoretically correct, introduces a fascinating element of uncertainty due to the possibility of looping. Although no instances of looping have been discovered despite extensive trials with large initial numbers, the systematic correctness of the conjecture remains a point of consideration. The conjecture can be deemed correct as long as no looping cases are found, and its statistical correctness is supported by the lack of evidence to the contrary. The degree of correctness could be enhanced by further exploration and analysis to more accurately determine the non-existence possibility of looping. Nevertheless, the Collatz conjecture stands as an intriguing mathematical puzzle, highlighting the delicate balance between its theoretical correctness and the infinitesimally small chance of encountering a looping scenario.

References

[1] Jose Capco. Odd collatz sequence and binary representations. 2019.

[2] Olinto de Oliveira Santos. Proving the collatz conjecture with binaries numbers.

[3] Patrick Chisan Hew. Working in binary protects the repetends of 1/3h: Comment on colussi's 'the convergence classes of collatz function'. *Theoretical Computer Science*, 618:135–141, 2016.

[4] Venkatesulu Mandadi and Devi Paramwswari. Verification of collatz conjecture: An algorithmic approach based on binary representation of integers. *arXiv preprint arXiv:1912.05942*, 2019.

[5] Tristan Stérin. Binary expression of ancestors in the collatz graph. In *International Conference on Reachability Problems*, pages 115–130. Springer, 2020.