ISSN: 2836-8495

Research Article

Journal of Current Trends in Computer Science Research

# Proof of the Collatz Conjecture Using Logical and Probabilistic Approaches

**Xinyi Zhou***

*Academic Institution in Zurich (ETH Zurich), Switzerland*

***Corresponding Author**
Xinyi Zhou, Academic Institution in Zurich (ETH Zurich), Switzerland.

**Submitted:** 2024, Mar 19 ; **Accepted:** 2024, Apr 26 ; **Published:** 2024, Jun 27

**Citation:** Zhou, X. (2024). Proof of the Collatz Conjecture Using Logical and Probabilistic Approaches. *J Curr Trends Comp Sci Res, 3*(3), 01-11.

**Abstract**

*Almost 100 years ago, the Collatz problem was introduced by German mathematician Lothar Collatz in the 1930s. It provides a predefined transformation algorithm to deal with any starting numbers in the Collatz process which are positive odd or even integers: for odd numbers, multiple it by 3 and add 1 and for even integers, divide it by 2. The Collatz conjecture states that after enough times of repetition all the natural starting numbers would eventually decrease and converge to 1. Although the Collatz conjecture has been verified to be true below a very high numerical threshold, until this day, no rigorous mathematical proof could be made which is the main focus of this research paper. After we have shown that no infinite loop could exist besides the trivial loop of 1→4 →2→1 and that no positive integer within the Collatz process would diverge to infinity, we firstly put through our proof of the Collatz conjecture using logical arguments before we support the latter results by proving the conjecture using a probabilistic approach.*

## 1. Introduction

The Collatz problem which is also i. a. called the Syracuse problem was introduced by German mathematician Lothar Collatz almost 100 years ago in the 1930s. It provides a transformation algorithm for positive even and odd integers as follows:

1. For even numbers, divided it by 2
2. For odd numbers, multiply it by 3 and add 1

The famous Collatz conjecture predicts that with enough rounds of repetition all-natural numbers which are positive integers do converge to the number 1.

Although the conjecture seems to be true for all natural numbers below a very high threshold which recently has reached to $3*2^{69} = 1.770887431... \times 10^{21}$, up to this day, no rigorous mathematical proof could be put through even though there exist very promising attempts in this direction to prove for example that almost all positive integers within the Collatz process could obtain almost bounded values [1,2].

In this piece of work, we want to go down this road and provide a mathematically comprehensive proof of the Collatz conjecture by firstly using an approach of applying logical arguments before we support the later results utilizing a probabilistic approach.

Before the proof could be put through, we need to show that there exists no infinite loop (cf. chapter 2) for all the positive starting integers (even and odd ones) and that no natural starting number would diverge to infinity (cf. chapter 3). In order to achieve these features, we need to derive some conditions and formulas using the core attributes and properties of the Collatz process and then apply further probabilistic methods to find valid values for the variables in the newly derived formula. In the next chapter 2, we will introduce the infinite loop condition which is a first but also very important step to validate the correctness of the Collatz conjecture.

## 2. Infinite Loop Condition

We start the proof of the Collatz conjecture by presenting our infinite loop condition formula:

$$x_{infiniteloop\,,odd\,start} = \frac{\left(2^{(n_1+n_2+...+n_{r-1})}+3\left(2^{(n_1+n_2+...+n_{r-2})}\right)+3^2\left(2^{(n_1+n_2+...+n_{r-3})}\right)+...+3^{(r-2)}2^{(n_1)}+3^{(r-1)}\right)}{\left(2^{(n_1+n_2+...+n_{r-1}+n_r)}-3^r\right)} \quad (1)$$

The infinite loop condition formula (1) above assesses which positive, natural, odd and integer starting number (= x) could potentially form an infinite loop within the Collatz process so that at some point and after some transformations the (odd) starting number would be reached again in the Collatz transformation process [3]. The variable r denotes the number of rounds the Collatz process is running e.g. how many times are any given starting (natural) odd integer number are transformed to to an even number and then back to an odd number again which closes a full Collatz cycle. And ni is the number of times an even number needed to be divided through 2 to reach an odd number again for each of the r Collatz rounds. Both r and ni are natural numbers resp. positive integers.

## 2.1 Additional Conditions and Constraints

When we define that total sum of division times in all r rounds would be equal to t*r with t = average division times through 2 per round r (to get to odd number again from even number), after the odd number is transformed to even number using the Collatz transformation process xodd*3+1 = $x_{even}$, we will get the following formula for the infinite loop condition with an odd (integer) starting number:

$$x_{infiniteloop, odd\ start} = \frac{\left(2^{(t(r-1))} + 3\left(2^{(t(r-2))}\right) + 3^2\left(2^{(t(r-3))}\right) + \ldots + 3^{(r-2)}2^{(t*1)} + 3^{(r-1)}\right)}{\left(2^{(tr)} - 3^r\right)} \quad (2)$$

Because the starting number in the infinite loop condition is a natural and thus positive and odd number, the ratio of numerator over denominator need to either be both positive or both negative. And since both the number of playing rounds r and average times of division through 2 (= t) are positive (with r and t*r = sum of all division times through all rounds r being a natural number and r being ≥ 1), the numerator N has to be positive as well which means that the denominator D need to be a positive expression as well:

$$Denominator\ D = \left(2^{(tr)} - 3^r\right) > 0$$

$$\rightarrow 2^{(tr)} > 3^r \rightarrow t > \frac{(\log 3)}{(\log 2)} = 1.58496\ldots$$

Thus for the infinite loop condition to yield a valid starting positive number for $x_{il,\,odd,}$ the average times of division through 2 (which is t) need to be bigger than log(3)/log(2) like for example 1.58497 or any other rational or even irrational number that are arbitrarily close resp. slightly bigger than the singularity at t = log(3)/log(2) where the denominator D would go to 0. But even though the later number (1.58497) is bigger than log(3)/log(2) and fulfills the infinite loop condition (2) in the sense that xil,odd would be a positive number then, additional conditions on t need to be made like for example that $x_{il,odd}$ need to be a positive (odd) integer number as well [4].

Other than that, we should assess which possible values t - which is the average times of division through 2 - could more realistically obtain, given the specific rules and structures of the whole Collatz process.

We already know that any even number in the Collatz process e.g. after a transformation from an odd number using the x*3+1 rule has the probability of 1/2 = 50% of reaching an odd number again through 1x division through 2[5], because dividing any even number through 2 could result in an even or odd number both with equal probability of 50% simply because there exist as many even numbers as odd numbers. Thus for any even number to reach an odd number needing more than 1x division (e.g 2x, 3x,…,1000x,…, infinite times) through 2 is 100% - 50% = 50% as well whereas the probability of needing 1x more division through 2 always decreases to half of its previous likelihood value, so for instance the likelihood of needing 2x division through 2 to reach an odd number again from any even number would be 1/2 * 1/2 = $1/2^2$ = 1/4 that is only half of 1/2 which is the probability of reaching an odd number from an even number by dividing through 2 only once. Therefore, the expected average value for average times of division through 2 (= t) is the probability of needing division i-times ($P_i$) multiplied with the actual number of division times through 2 ($n_i$) as follows:

$$t = \sum_{i=1}^{\infty} P_i * n_i = \sum_{n=1}^{\infty}\left(\frac{1}{2^n} * n\right) = 2,\ with\ P_{total} = \sum_{n=1}^{\infty}\left(\frac{1}{2^n}\right) = 1 = 100\% \quad (3)$$

Here we are able to see that by considering the properties of the Collatz process the average value of necessity to divide through 2 to reach an odd number again (t=2) from any even number is not really equal to or near the singularity of t = log(3)/log(2) = 1.58496… ($x_{infiniteloop,\,odd}$ would reach infinity there because the denominator of infinite loop formula (2) would vanish to 0) but bigger by a fair margin. This finding would be quite useful in our further analysis of both the infinite loop and divergence conditions.

## 2.2. Further Calculations and Assessments Using the Infinite Loop Condition Formula

After defining the conditions and constrains of our infinite loop condition (1) especially for the average division times through 2 (= t) to reach an odd number again from an even number, we will continue with our calculations and rearrangements of the infinite loop formula (2) for generic t:

$$x_{infiniteloop,odd\,start}=\frac{\left(2^{(t(r-1))}+3\frac{2^{(t(r-1))}}{2^{t}}+3^{2}\frac{2^{(t(r-1))}}{2^{2t}}+\ldots+3^{(r-2)}2^{(t*1)}\frac{2^{(t(r-2))}}{2^{(t(r-2))}}+3^{(r-1)}\frac{2^{(t(r-1))}}{2^{(t(r-1))}}\right)}{\left(2^{(tr)}-3^{r}\right)}$$

$$x_{infiniteloop,odd\,start}=\frac{\left(2^{(t(r-1))}+3\frac{2^{(t(r-1))}}{2^{t}}+3^{2}\frac{2^{(t(r-1))}}{2^{2t}}+\ldots+3^{(r-2)}\frac{2^{(t(r-1))}}{2^{(t(r-2))}}+3^{(r-1)}\frac{2^{(t(r-1))}}{2^{(t(r-1))}}\right)}{\left(2^{(tr)}-3^{r}\right)}$$

$$X_{il,odd\,start}=\frac{\left(2^{(t(r-1))}+2^{(t(r-1))}\left(\frac{3}{2^{t}}\right)+2^{(t(r-1))}\left(\frac{3}{2^{t}}\right)^{2}+\ldots+2^{(t(r-1))}\left(\frac{3}{2^{t}}\right)^{(r-2)}+2^{(t(r-1))}\left(\frac{3}{2^{t}}\right)^{(r-1)}\right)}{\left(2^{(tr)}-3^{r}\right)}$$

$$X_{il,odd\,start}=\frac{\left(\left(2^{(t(r-1))}\right)\left(1+\left(\frac{3}{2^{t}}\right)+\left(\frac{3}{2^{t}}\right)^{2}+\ldots+\left(\frac{3}{2^{t}}\right)^{(r-1)}\right)\right)}{\left(2^{(t(r-1))}\left(2^{t}-\frac{3^{r}}{2^{(t(r-1))}}\right)\right)}$$

$$X_{il,odd\,start}=\frac{\left(1+\left(\frac{3}{2^{t}}\right)+\left(\frac{3}{2^{t}}\right)^{2}+\ldots+\left(\frac{3}{2^{t}}\right)^{(r-1)}\right)}{\left(\left(2^{t}-\frac{3^{r}}{2^{(t(r-1))}}\right)\right)}=\frac{(Numerator\ N)}{(Denominator\ D)} \qquad (4)$$

Applying the sum formula of geometric series (5) with n-terms until last term n-1 (which is not n) below for the numerator N of the infinite loop formula (4) above leads to:

$$S_{n}=1+q^{1}+q^{2}+q^{3}+\ldots+q^{(n-1)}=a_{1}\frac{(1-q^{n})}{(1-q)},\ q=\frac{3}{2^{t}} \qquad (5)$$

$$X_{il,odd\,start}=\frac{\left(1+\left(\frac{3}{2^{t}}\right)+\left(\frac{3}{2^{t}}\right)^{2}+\ldots+\left(\frac{3}{2^{t}}\right)^{(r-1)}\right)}{\left(\left(2^{t}-\frac{3^{r}}{2^{(t(r-1))}}\right)\right)}=\frac{(Numerator\ N)}{(Denominator\ D)}$$

$$X_{il,odd\,start}=\frac{\dfrac{1*\left(1-\left(\frac{3}{2^{t}}\right)^{(r)}\right)}{\left(1-\left(\frac{3}{2^{t}}\right)\right)}}{\left(\left(2^{t}-\frac{3^{r}}{2^{(t(r-1))}}\right)\right)}=\frac{\dfrac{\left(1-\left(\frac{3^{r}}{2^{(tr)}}\right)\right)}{\left(1-\left(\frac{3}{2^{t}}\right)\right)}}{\left(\left(2^{t}-\frac{3^{r}}{2^{(t(r-1))}}\right)\right)} \qquad (6)$$

Now as we remember from chapter 2.1 formula (3) that the average times of division through 2 would be t = 2 which is quite a bit away from the singularity resp. pole of t = log(3)/log(2) (= 1.584946…):

$$t=\sum_{i=1}^{\infty} P_i * n_i=\sum_{n=1}^{\infty}\left(\frac{1}{2^n}*n\right)=2\,,\ with\ P_{total}=\sum_{n=1}^{\infty}\left(\frac{1}{2^n}\right)=1=100\%$$

And we when plug in the value t = 2 into the formula (5) above for the possible starting numbers (which are odd integer numbers) for this number x to generate an infinite loop:

$$X_{il\,,odd\,start}=\frac{\frac{\left(1-\left(\frac{3^r}{2^{(tr)}}\right)\right)}{\left(1-\left(\frac{3}{2^t}\right)\right)}}{\left(\left(2^t-\frac{3^r}{2^{(t(r-1))}}\right)\right)}=\frac{\frac{\left(1-\left(\frac{3^r}{2^{(2r)}}\right)\right)}{\left(1-\left(\frac{3}{2^2}\right)\right)}}{\left(\left(2^2-\frac{3^r}{2^{(2(r-1))}}\right)\right)}=1$$

We are able to see that when we increase the average division times through 2 (which is t) from log(3)/log(2) (=1.58496...) to 2, the possible odd integer numbers to potentially form an infinite loop reduces from an infinite large value to 1 which means that in reality, the Collatz process could only produce an infinite loop when the (odd & integer) starting number is exactly 1 that forms a trivial infinite loop then of 1 → 4 → 2 → 1 → 4 → 2 → 1 and so on. Therefore, we could say that based on the infinite loop condition (1) and knowing that the properties of the Collatz process only allow an average division number of t = 2 for each of the r rounds, no finite starting number which belong to the positive/natural odd integer numbers besides 1 could produce an infinite loop [6].

But what about even (integer) starting numbers, could these natural even starting numbers possibly form an infinite loop? The short answer is no: since every finite even starting number could be transformed in an odd integer number in the Collatz process by dividing it 1x (with 50% probability because even and odd numbers are evenly distributed) or multiple times with half of the previous probability when the division times increases by 1 each time so that division through 2 twice to reach an odd number again would yield an likelihood of $(1/2)^2 = 1/4$ (So the probability of needing to divide through 2 infinite times to reach an odd number from an even number would go to 0). And knowing that any even number could be transformed to an odd number sooner or later in the Collatz process by dividing it through 2 once or multiple times (including infinite times which is theoretically allowed as well in the Collatz process) and because no (positive) odd integer number bigger than 1 could form an infinite loop, we could say that there exists no infinite loop besides the trivial infinite loop of 1 → 4 → 2 → 1 [7,8].

## 3. Divergence Condition

After we made sure that there exists no infinite loop other than the trivial loop (1 → 4 → 2 → 1) in the previous chapter 2, we need to check whether any starting natural number resp. positive finite integer (= x) could possibly diverge to infinity which would make the termination of the Collatz process by going to the number 1 impossible. And for this purpose, we use the below divergence condition which is very similar to the infinite loop condition presented in the last chapter 2 (instead of an equal sign after x, it is a smaller than sign, so the left-hand side will provide an upper bound to the starting number x and as long as x is smaller than this threshold, divergence to infinity after r rounds is possible):

$$x_{infinite\,divergence\,,\,odd\,start}<\frac{\left(2^{(n_1+n_2+\ldots+n_{r-1})}+3\left(2^{(n_1+n_2+\ldots+n_{r-2})}\right)+3^2\left(2^{(n_1+n_2+\ldots+n_{r-3})}\right)+\ldots+3^{(r-2)}2^{(n_1)}+3^{(r-1)}\right)}{\left(2^{(n_1+n_2+\ldots+n_{r-1}+n_r)}-3^r\right)} \qquad (7)$$

Because of the high similarity of the infinite loop and the divergene condition, a lot of arguments and reasoning are very similar to these in the previous chapter as well.

The divergence formula (7) above assesses which positive, natural, odd and integer starting number (= x) could potentially diverge within the Collatz process so that at some point and after some transformations the (odd) starting number would always get bigger without any bound in the Collatz transformation process [9]. The variable r denotes the number of rounds the Collatz process is running
e.g. how many times are any given starting (natural) odd integer number are transformed to to an even number and then back to an odd number again which closes a full Collatz cycle. And $n_i$ is the number of times an even number needed to be divided through 2 to reach an odd number again for each of the r Collatz rounds. Both r and $n_i$ are natural numbers resp. positive integers.

### 3.1 Additional Conditions and Constraints

When we define that total sum of division times in all r rounds would be equal to t*r with t = average division times through 2 per round r (to get to odd number again from even number), after the odd number is transformed to even number using the Collatz transformation process $x_{odd}*3+1 = x_{even}$, we will get the following formula for the divergence condition with an odd (integer) starting number:

$$x_{infinite\ divergence,\ odd\ start} < \frac{\left(2^{(t(r-1))} + 3\left(2^{(t(r-2))}\right) + 3^2\left(2^{(t(r-3))}\right) + \ldots + 3^{(r-2)}2^{(t*1)} + 3^{(r-1)}\right)}{\left(2^{(tr)} - 3^r\right)} \qquad (8)$$

Because the starting number in the divergence condition above is a natural and therefore positive and odd number, the ratio of numerator over denominator need to either be both positive or both negative. And since both the number of playing rounds **r** and average times of division through 2 (= **t**) are positive (with r and t*r = sum of all division times through all rounds r being a natural number and r being ≥ 1), the numerator N need to be positive as well which means that the denominator D has to be a positive expression as well:

$$Denominator\ \ D = \left(2^{(tr)} - 3^r\right) > 0$$

$$\rightarrow 2^{(tr)} > 3^r \rightarrow t > \frac{(\log 3)}{(\log 2)} = 1.58496\ldots$$

Thus for the divergence condition to yield a valid finite upper bound of the starting positive number for $x_{il,odd,}$ the average times of division through 2 (= t) need to be bigger than log(3)/log(2) like for example 1.58497 or any other rational or even irrational number that are arbitrarily close resp [10]. slightly bigger than the singularity at t = log(3)/log(2) where the denominator D would go to 0.

But even though the later number (1.58497) is bigger than log(3)/log(2) and fulfills the infinite divergence condition (8) in the sense that $x_{il,odd}$ would be bounded by a positive number then, additional conditions on t need to be made like for example that $x_{il,odd}$ need to be a positive (odd) integer number as well. Other than that, we should check which possible values t - which the average times of division through 2 - could more realistically obtain, given the specific rules and structures of the whole Collatz process.

We already know that any even number in the Collatz process e.g. after a transformation from an odd number using the x*3+1 rule has the probability of 1/2 = 50% of reaching an odd number again through 1x division through 2, because dividing any even number through 2 could result in an even or odd number both with equal probability of 50% simply because there exist as many even numbers as odd numbers [11]. Thus for any even number to reach an odd number needing more than 1x division (e.g 2x, 3x,…,1000x,…, infinite times) through 2 is 100 - 50% = 50% as well whereas the probability of needing 1x more division through 2 always decreases to half of its original value, so for instance the likelihood of needing 2x division through 2 to reach an odd number again from any even number would be $1/2 * 1/2 = 1/2^2 = 1/4$ that is only half of 1/2 which is the probability of reaching an odd number from an even number by dividing through 2 only once. Therefore, the expected average value for average times of division through 2 (= t) is the probability of needing division i-times ($P_i$) multiplied with the actual number of division times through 2 ($n_i$) as follows and the same formula (3) presented in the previous 2nd chapter:

$$t = \sum_{i=1}^{\infty} P_i * n_i = \sum_{n=1}^{\infty}\left(\frac{1}{2^n} * n\right) = 2,\ with\ P_{total} = \sum_{n=1}^{\infty}\left(\frac{1}{2^n}\right) = 1 = 100\%$$

Here we are able to see that by considering the properties of the Collatz process the average value of necessity to divide through 2 to reach an odd number again (t=2) from any even number is not really equal to or near the singularity of t = log(3)/log(2) = 1.58496…

(The upper bound for $x_{infinite\ divergence,\ odd}$ would reach infinity there because the denominator of infinite loop formula (2) would vanish to 0) but bigger by a fair margin. This finding was quite useful in our analysis of the infinite loop condition in the previous chapter 2 will be very useful for deeper analysis of the divergence condition for any natural odd starting number in the Collatz process.

### 3.2. Further Calculations and Assessments Using the Divergence Condition Formula

After defining the conditions and constrains of our divergence condition (7) especially for the average division times through 2 (= t) to reach an odd number again from an even number, we will continue with our calculations and rearrangements of the divergence formula (8) for generic t:

$$x_{infinite\ divergence,\ odd\ start} < \frac{\left(2^{(t(r-1))} + 3\left(2^{(t(r-2))}\right) + 3^2\left(2^{(t(r-3))}\right) + \ldots + 3^{(r-2)}2^{(t*1)} + 3^{(r-1)}\right)}{\left(2^{(tr)} - 3^r\right)}$$

$$x_{infinite\ divergence,\ odd\ start} < \frac{\left(2^{(t(r-1))} + 3\frac{2^{(t(r-1))}}{2^t} + 3^2\frac{2^{(t(r-1))}}{2^{2t}} + \ldots + 3^{(r-2)}2^{(t*1)}\frac{2^{(t(r-2))}}{2^{(t(r-2))}} + 3^{(r-1)}\frac{2^{(t(r-1))}}{2^{(t(r-1))}}\right)}{\left(2^{(tr)} - 3^r\right)}$$

$$x_{\text{infinite divergence , odd start}} < \frac{\left(2^{(t(r-1))} + 3\frac{2^{(t(r-1))}}{2^t} + 3^2\frac{2^{(t(r-1))}}{2^{2t}} + \ldots + 3^{(r-2)}\frac{2^{(t(r-1))}}{2^{(t(r-2))}} + 3^{(r-1)}\frac{2^{(t(r-1))}}{2^{(t(r-1))}}\right)}{\left(2^{(tr)} - 3^r\right)}$$

$$X_{id\text{ , odd start}} < \frac{\left(2^{(t(r-1))} + 2^{(t(r-1))}\left(\frac{3}{2^t}\right) + 2^{(t(r-1))}\left(\frac{3}{2^t}\right)^2 + \ldots + 2^{(t(r-1))}\left(\frac{3}{2^t}\right)^{(r-2)} + 2^{(t(r-1))}\left(\frac{3}{2^t}\right)^{(r-1)}\right)}{\left(2^{(tr)} - 3^r\right)}$$

$$X_{id\text{ , odd start}} < \frac{\left(\left(2^{(t(r-1))}\right)\left(1 + \left(\frac{3}{2^t}\right) + \left(\frac{3}{2^t}\right)^2 + \ldots + \left(\frac{3}{2^t}\right)^{(r-1)}\right)\right)}{\left(2^{(t(r-1))}\left(2^t - \frac{3^r}{2^{(t(r-1))}}\right)\right)}$$

$$X_{\text{infinite divergence , odd start}} < \frac{\left(1 + \left(\frac{3}{2^t}\right) + \left(\frac{3}{2^t}\right)^2 + \ldots + \left(\frac{3}{2^t}\right)^{(r-1)}\right)}{\left(\left(2^t - \frac{3^r}{2^{(t(r-1))}}\right)\right)} = \frac{(Numerator\ N)}{(Denominator\ D)} \qquad (9)$$

Applying the sum formula of geometric series (5) of the previous chapter 2 with n-terms until last term n-1 (which is not n) below for the numerator N of the divergence condition formula (10) above leads to:

$$S_n = 1 + q^1 + q^2 + q^3 + \ldots + q^{(n-1)} = a_1\frac{(1-q^n)}{(1-q)},\ q = \frac{3}{2^t}$$

$$X_{\text{infinite divergence , odd start}} < \frac{\left(1 + \left(\frac{3}{2^t}\right) + \left(\frac{3}{2^t}\right)^2 + \ldots + \left(\frac{3}{2^t}\right)^{(r-1)}\right)}{\left(\left(2^t - \frac{3^r}{2^{(t(r-1))}}\right)\right)} = \frac{(Numerator\ N)}{(Denominator\ D)}$$

$$X_{\text{infinite divergence , odd start}} < \frac{\dfrac{1 * \left(1 - \left(\frac{3}{2^t}\right)^{(r)}\right)}{\left(1 - \left(\frac{3}{2^t}\right)\right)}}{\left(\left(2^t - \frac{3^r}{2^{(t(r-1))}}\right)\right)}$$

$$X_{\text{infinite divergence , odd start}} < \frac{\dfrac{\left(1 - \left(\frac{3^r}{2^{(tr)}}\right)\right)}{\left(1 - \left(\frac{3}{2^t}\right)\right)}}{\left(\left(2^t - \frac{3^r}{2^{(t(r-1))}}\right)\right)} \qquad (10)$$

Now as we remember from chapter 2.1 from formula (3) that the average times of division through 2 would be t = 2 which is quite a bit away from the singularity resp. pole of t = log(3)/log(2) (= 1.584946…):

$$t=\sum_{i=1}^{\infty} P_i * n_i=\sum_{n=1}^{\infty}\left(\frac{1}{2^n} * n\right)=2\text{, with } P_{total}=\sum_{n=1}^{\infty}\left(\frac{1}{2^n}\right)=1=100\%$$

And we when plug in the value t = 2 into the formula (12) above for the possible starting numbers (which are odd integer numbers) to verify whether this number could possibly diverge unbounded to infinity:

$$X_{infinite\ divergence\ ,odd\ start}<\frac{\frac{\left(1-\left(\frac{3^r}{2^{(tr)}}\right)\right)}{\left(1-\left(\frac{3}{2^t}\right)\right)}}{\left(\left(2^t-\frac{3^r}{2^{(t(r-1))}}\right)\right)}=\frac{\frac{\left(1-\left(\frac{3^r}{2^{(2r)}}\right)\right)}{\left(1-\left(\frac{3}{2^2}\right)\right)}}{\left(\left(2^2-\frac{3^r}{2^{(2(r-1))}}\right)\right)}=1$$

We are able to see that when we increase the average division times through 2 (which is t) from log(3)/log(2) (=1.58496...) to 2, the possible odd integer numbers to possibly diverge without bound to infinity reduces from an infinite large value to 1 which means that in reality, the Collatz process could only produce an infinite divergent process if the starting number (which need to be a natural number resp. positive integer) could be smaller than 1 that is already the smallest possible natural odd number. Thus, according the above formula, no natural odd starting number in the Collatz process could diverge to infinity [12].

But what about even (integer) starting numbers, could these natural even starting numbers potentially lead to divergence to infinity? The answer is no as well: because every finite even starting number could be transformed in an odd integer number in the Collatz process by dividing it 1x (with 50% probability because even and odd numbers are evenly distributed) or multiple times with half of the previous probability when the division times increases by 1 each time so that division through 2 twice to reach an odd number again would yield an likelihood of $(1/2)^2$ = 1/4 (So the probability of needing to divide through 2 infinite times to reach an odd number from an even number would go to 0).

And knowing that any even number could be transformed to an odd number in the Collatz process sooner or later by dividing it through 2 once or multiple times (including infinite times which is theoretically allowed as well in the Collatz process) and because no positive odd integer numbers exist which are simultaneously smaller than 1 (which is a necessary condition to make divergence to infinity possible for natural odd numbers), we could confidentely state that no positive integer numbers (odd or even) could diverge to infinity [13].

## 4. Main Results

In this chapter, we are using the findings of the previous chapters to construct the proof of the Collatz conjecture which states that

1. For even numbers, we divide them by 2
2. For odd numbers, we multipy them by 3 and add 1

so that after enough times of repetition (which could be infinite times theoretically), all-natural numbers which are positive integers converge to the number 1.

Firstly, we proof the Collatz conjecture by using logical arguments (chapter 4.1) and then we construct a proof of the same conjecture utilizing a probabilistic method.

### 4.1 Proof Using a Logical Approach

We already know from the results of the previous chapters 2 and 3 that

1. Besides the trivial infinite loop (1→ 4 → 2→ 1) there exists no other infinite loop starting from the positive odd integer numbers. And the positive even integers could always be transformed to an odd integer number – for which no non-trivial infinity loop exist – by applying 1x or multiple times the division through 2 which is a crucial part of the whole Collatz process

2. In order for any natural number resp. positive (odd) integer number to possibly diverge to infinity would need these starting numbers in the Collatz process to be smaller than 1 which is impossible because the number 1 is already smallest possible positive (odd) integer.

And every natural even number could always be transformed to an odd integer number applying 1x or multiple times the division through 2 which is a crucial part of the whole Collatz process.

So, because no natural (odd or even) numbers resp. no positive integers besides 1 could form an infinite loop or diverge to infinity within the Collatz process, we could conclude that any natural starting number would decrease at one point after applying the Collatz transformation for both the positive odd (x*3+1) and the even (dividing n-times through 2) integers. And since every odd number besides 1 (which is the smallest possible number in the Collatz process) will decrease to an even smaller odd number and every even number could be reduced to a smaller odd or even number (which would be further reduced by division through 2 until it become an odd number that could be scaled down to an even smaller positive odd number), it is safe to say that the reduced odd or even number could be scaled down to yet a smaller odd or even number so that eventually these reduced natural numbers would reach the smallest possible number in the whole Collatz process which is the number 1 that is the smallest positive integer [14]. This logical reasoning proves the Collatz conjecture which states that all-natural numbers will converge to the number 1 by applying the Collatz transformation process for positive odd and even integers.

### 4.2. Proof Using a Probabilistic Approach

After we have proved the Collatz conjecture using a method of logical arguments in chapter 4.1, we will prove the same conjecture using a probabilistic approach.

Given that no infinite loop besides the trivial one (1→4 → 2 →1) exists for odd numbers and that the divergence condition could never be fulfilled for odd numbers withing the whole Collatz transformation processes and even numbers are always getting smaller during (partial) Collatz processes (dividing through 2) unless it becomes an odd number we already know that all-natural numbers which are positive integers do get smaller resp. always would converge. Now we want to show here that the convergence eventually leads to the number 1 for any finite positive integer number utilizing a probabilistic method.

The whole Collatz transformation process for any starting odd (and even) number could theoretically be run infinite times with infinite rounds r. And in each of the r rounds, there is a slim chance of success for the starting natural number to be transformed to the “end” number 1. That is because every starting even number could be transformed to an odd number (after r transformation attempts) and every starting odd number could be changed to an even number using only 1 transformation round (multiple the number x by 3 and add 1).

And all the odd-transformed even number do possess a small likelihood of being a number of 2 to the power of k (for finite k within a finite interval) which makes sure that this number will for sure be transformable to 1 after division to 2 for k times.

All in all, the probability for a finite even number to be a number of 2 to the power of k within a finite interval of [1, 2k] is as follows:

$$\begin{aligned} &P\left(number\ of\ 2^{k} within\ interval\left[1,2^{k}\right] \mid even\ number\right) \\ &=\frac{\left(P\left(number\ of\ 2^{k} within\ interval\left[1,2^{k}\right]\right)\right)}{\left(P\left(even\ number\right)\right)} \\ &=\frac{\left(\frac{k}{2^{k}}\right)}{\left(\frac{1}{2}\right)}=\left(\frac{k}{2^{(k-1)}}\right)>0,\ for\ finite\ k \end{aligned} \tag{11}$$

Therefore, the probability for a finite even number to be a number of 2k is always positive (> 0) for finite k in a finite interval of [1, $2^k$]. And when k decreases, the later probability will increase because the denominator (= k) decreases less than the numerator $2^{(k-1)}$ [15].

In practice we could adjust the probability calculation method slightly for it to be yet more intuitive to understand and apply. After every (partial) Collatz transformation process for odd positive starting numbers to an even number (by x*3+1) we could look at the number of digits the newly generated even number has and define the interval to span from 1 to 1*10 to the power of d with d to be the number of decimal digits the even number possesses.

Then we need to find out, how many numbers of 2 to the power of n would lie in this newly defined interval (e.g. between 1 and $1*10^d$) which we could calculate using the below formula [16]:

$$\begin{aligned} &P\left(even\ number\ of\ d\ digits\ being\ a\ number\ of\ 2^{n}\right) \\ &=\frac{\left(number\ of\ integers\ of\ 2^{n}\ within\left[1,1*10^{d}\right]\right)}{\left(number\ of\ even\ integers\ within\left[1,1*10^{d}\right]\right)} \end{aligned} \tag{12}$$

So, for example for an even number consisting of 3 digits which lies in the range [100, 999], the probability of any of these even number to be of 2 to the power n would be:

$$P(even\ number\ of\ 3\ digits\ being\ a\ number\ of\ 2^n)$$
$$=\frac{(number\ of\ integers\ of\ 2^n within[1,1*10^3])}{(number\ of\ even\ integers\ within[1,1*10^3])}$$
$$=\frac{\left(\frac{9}{1000}\right)}{\left(\frac{500}{1000}\right)}=\frac{9}{500}=1.8\% > 0,$$
$$because\ 2^9=512\ is\ the\ largest\ integer\ of\ 2^n within\ interval[1,1*10^3]=[1,1000]$$

So, for any positive integer with 3 digits the probability of it to be a number of $2^n$ (which for sure will lead to the this even number to convergence to the number 1 and would finish the Collatz process) is 9/500 = 1.8 % which is a clearly positive likelihood.

Due to the fact that all the positive integers (especially the even numbers) in the Collatz process will decrease in its convergence, it is interesting to know what the probability is for even numbers with reduced number of digits (e.g. d = 2) to be a number of $2^n$ as well:

$$P(even\ number\ of\ 2\ digits\ being\ a\ number\ of\ 2^n)$$
$$=\frac{(number\ of\ integers\ of\ 2^n within[1,1*10^2])}{(number\ of\ even\ integers\ within[1,1*10^2])}$$
$$=\frac{\left(\frac{6}{100}\right)}{\left(\frac{50}{100}\right)}=\frac{6}{50}=12\% \gg 0,$$
$$because\ 2^6=64\ is\ the\ largest\ integer\ of\ 2^n within\ interval\ [1,1*10^2]=[1,100]$$

And the likelihood of a smaller even number of 2 digits to be of 2n is 6/50 = 12 % which is a lot bigger than the probability of a 3 digit even number to be of 2 to the power of n as expected resp. predicted using the probability formula (13) of $P = k / 2^{(k-1)}$ (for finite k in a finite interval $[1, 2^k]$).

$$P(even\ number\ of\ d\ digits\ being\ a\ number\ of\ 2^n) =\frac{(number\ of\ integers\ of\ 2^n within[1,1*10^d])}{(number\ of\ even\ integers\ within[1,1*10^d])} \tag{13}$$

As stated before the beginning of this chapter 4, the whole Collatz transformation process for any starting odd (and even) number could theoretically be run infinite times with infinite rounds r. And in each of the r rounds, there is a slim chance of success for the starting natural number to be transformed to the number 1 which would terminate the Collatz process (resp. stays in the trivial infinite loop 1→4 →2→1→4 →2→1 etc. forever). That is because every starting even number could be transformed to an odd number (after r transformation attempts) and every starting odd number could be changed to an even number using only 1 transformation round (multiple the number x by 3 and add 1). And all the odd-transformed even number do possess a small probability of being a number of 2n, thus the probability of any odd-transformed even number reaching 1 after going through all r rounds (which could be r = infinity rounds) would be as follows:

$$\begin{aligned}
&P(any\ odd\ transformed\ even\ number\ reaching\ number\ 1\ after\ going\ through\ all\ r\ rounds)\\
&=1-P(any\ odd\ transformed\ even\ number\ not\ reaching\ number\ 1\ after\ all\ r\ rounds)\\
&=1-P(even\ number\ of\ d\ digits\ not\ being\ a\ number\ of\ 2^n\ after\ all\ r\ rounds)\\
&=1-P(even\ number\ of\ d\ digits\ not\ being\ a\ number\ of\ 2^n\ after\ round\ 1)\\
&*P(even\ number\ of\ d\ or\ lower\ digits\ not\ being\ a\ number\ of\ 2^n\ after\ round\ 2)\\
&*...*P(even\ number\ of\ d\ or\ lower\ digits\ not\ being\ a\ number\ of\ 2^n\ after\ round\ r)\\
&=1-\left(P_{average}(even\ number\ of\ d\ or\ lower\ digits\ not\ being\ 2^n\ after\ any\ round\ r)\right)^r
\end{aligned} \tag{14}$$

$$\begin{aligned}
&P(any\ odd\ transformed\ even\ number\ reaching\ number\ 1\ after\ going\ through\ all\ r\ rounds)\\
&=1-P(any\ odd\ transformed\ even\ number\ not\ reaching\ number\ 1\ after\ all\ r\ rounds)\\
&=1-P(even\ number\ of\ d\ digits\ not\ being\ a\ number\ of\ 2^n\ after\ all\ r\ rounds)
\end{aligned}$$

$$=1-P(even\ number\ of\ d\ digits\ not\ being\ a\ number\ of\ 2^n\ after\ round\ 1)$$
$$*P(even\ number\ of\ d\ or\ lower\ digits\ not\ being\ a\ number\ of\ 2^n\ after\ round\ 2)$$
$$*...*P(even\ number\ of\ d\ or\ lower\ digits\ not\ being\ a\ number\ of\ 2^n\ after\ round\ r)$$
$$=1-\left(P_{average}(even\ number\ of\ d\ or\ lower\ digits\ not\ being\ 2^n\ after\ any\ round\ r)\right)^r$$
$$=1-(number\ between\ (0,1))^r=1-\lim_{r\to\infty}(number\ between\ (0,1))^r$$
$$=1-0=1=100\%$$

And because the (averaged) "failure" probability of an odd-transformed even number of d or lower digits not being of $2^n$ after any round r would always be < 1 (and > 0) because the "success" likelihood of an even number of d or lower digits being of $2^n$ after any round r is always positive and greater than 0 we could conclude that any odd-transformed even number reaching number 1 after going through all r rounds (e.g. infinite rounds) would be 1 = 100 % which is for sure. This proves the Collatz conjecture (using the above probabilistic approach) which states that all positive even or odd integers will finally reach the number 1 after going through the Collatz transformation process in maximally r = infinity rounds. So, the used probabilistic approach further supports the proof of the Collatz conjecture using logical arguments previously provided in the chapter 4.1 [17,18].

## Appendix

Derivation of the infinite loop condition which is presented as formula (1) from chapter 2:

$$x_{infiniteloop,odd}=\frac{\left(2^{(n_1+n_2+...+n_{r-1})}+3\left(2^{(n_1+n_2+...+n_{r-2})}\right)+3^2\left(2^{(n_1+n_2+...+n_{r-3})}\right)+...+3^{(r-2)}2^{(n_1)}+3^{(r-1)}\right)}{\left(2^{(n_1+n_2+...+n_{r-1}+n_r)}-3^r\right)}$$

For 1 round (r=1) of whole Collatz transformation process starting from odd number, transforming to even number and then back to odd number again by dividing n times through 2:

$$\frac{(x_{infiniteloop,odd}*3+1)}{2^n}=x_{infiniteloop,odd}\Leftrightarrow 1=x_{il,odd}*2^n-3*x_{il,odd}$$
$$\to x_{il,odd}=\frac{1}{(2^n-3)}$$

For r=2:

$$\frac{\left(\left(\frac{(x_{infiniteloop,odd}*3+1)}{2^{(n_1)}}\right)*3+1\right)}{\left(2^{(n_2)}\right)}=x_{infiniteloop,odd}$$
$$\to x_{il,odd}=\frac{\left(2^{(n_1)}+3\right)}{\left(2^{(n_1+n_2)}-9\right)}$$

For r=3:

$$\frac{\left(\left(\frac{\left(\left(\frac{(x_{infiniteloop,odd}*3+1)}{2^{(n_1)}}\right)*3+1\right)}{\left(2^{(n_2)}\right)}\right)*3+1\right)}{2^{(n_3)}}=x_{infiniteloop,odd}$$
$$\to x_{il,odd}=\frac{\left(2^{(n_1+n_2)}+3*2^{(n_1)}+9\right)}{\left(2^{(n_1+n_2+n_3)}-27\right)}$$

$$x_{infiniteloop,odd} = \frac{\left(2^{(n_1+n_2+\ldots+n_{r-1})} + 3\left(2^{(n_1+n_2+\ldots+n_{r-2})}\right) + 3^2\left(2^{(n_1+n_2+\ldots+n_{r-3})}\right) + \ldots + 3^{(r-2)}2^{(n_1)} + 3^{(r-1)}\right)}{\left(2^{(n_1+n_2+\ldots+n_{r-1}+n_r)} - 3^r\right)}$$

## References


1. cf. Hercher (2022)
2. cf. Tao (2019)
3. cf. appendix for the derivation of the infinite loop condition formula (1)
4. Theoretically speaking, irrational numbers are allowed for the average times of division through 2 (= t), because the number of playing rounds r in the Collatz process is not bounded and even if t would be an irrational number with unbounded decimal points, it could be "canceled" out by the very big decimal number r which is not bounded as well.
5. This probability is consistent for all possible intervals within the total interval from 1 to infinity (which is not always the case for other events that we encounter later on where the likelihood changes dynamically depending on the specific interval which makes interval-wise analysis of occurrence of events even more necessary).
6. The odd integer starting number in the infinite loop condition would get infinitely large for t = log(3)/log(2) or near e.g. slightly bigger than log(3)/log(2).
7. If the even number does not transform to an odd number through division through 2, then it always decreases which makes an infinite loop impossible (as long as it does not become an odd number whose scale would increase in the Collatz transformation process to an even number again which makes an infinite loop possible, at least theoretically).
8. The trivial infinite loop of 1 → 4 → 2 → 1 does not actually count as an infinite loop because one could argue that at the start of the trivial infinite loop, the Collatz process has already finished as it has already reached the number1.
9. cf. appendix for the derivation of the divergence condition for any positive odd integer number.
10. Just as already stated in the previous chapter 2, irrational numbers are allowed for the average times of division through 2 (= t), because the number of playing rounds r in the Collatz process is not bounded and even if t would be an irrational number with unbounded decimal points, it could be "canceled" out be the very big decimal number r which is not bounded as well.
11. This probability is just like in chapter 2 explained, consistent for all possible intervals within the total interval from 1 to infinity (which is not always the case for other events that we encounter later on where the likelihood changes dynamically depending on the specific interval which makes interval-wise analysis of occurrence of events even more necessary).
12. The odd integer starting number in the infinite divergence condition would get an infinitely large upper bound of x for t = log(3)/log(2) or near resp. slightly bigger than log(3)/log(2).
13. If the even number does not transform to an odd number through division through 2, then it always decreases which makes divergence to infinity impossible (as long as it does not become an odd number whose magnitude would increase in the Collatz transformation process to an even number again which makes divergence to infinity possible, at least theoretically).
14. To explain the convergence of all natural numbers (finally to 1) of the Collatz process more clearly, it is already known that all even numbers always decrease to other odd or even number. If it decreases to an even number it could be scaled down even more until it become an odd number, so technically speaking, even number always decreases and never increases in the Collatz transformation process unless it becomes an odd number. But also for odd numbers, even at firstly sight it seems that they could increase because of the transformation process (x*3+1) to make the odd number an even number, in the long run and after running through the whole Collatz process (odd to even to odd again) and transform the even number back to the odd number again, the newly generated odd number is always smaller than the previous odd number because no infinite loop exists besides the trivial one (starting with 1 to 4 to 2 to 1) and no natural starting (odd) number exists which is smaller than 1 (which is the smallest possible positive integer) which theoretically need to be fulfilled to make divergence to infinity possible for odd numbers. Thus both even and also odd numbers eventually get smaller after partial (for even numbers) or whole and (over r rounds) repeated (for even numbers) Collatz transformation processes.
15. Because the probability of an even number to be of 2 to the power of k does change dynamically over intervals resp. it always decreases for smaller positive number ranges, it makes interval-wise analysis of occurrence of events even more necessary as already stated in the previous chapter 2.
16. Because the interval [1, 1*10d] for sure already includes any natural number of d digits for probability calculation so that no larger intervals are needed here.
17. Hercher, C. (2023). There are no Collatz m-Cycles with m≤ 91. *Journal of Integer Sequences, 26*(2), 3.
18. Tao, T. (2022, January). Almost all orbits of the Collatz map attain almost bounded values. In *Forum of Mathematics, Pi* (Vol. 10, p. e12). Cambridge University Press.