A probabilistic proof of Collatz conjecture

Abstract: A probabilistic proof of the Collatz conjecture is described via identifying a sequential permutation of even natural numbers by divisions by 2 that follows a recurrent pattern of the form $x, 1, x, 1$ ..., where x represents divisions by 2 more than once. The sequence presents a probability of 50:50 of divisions by 2 once as opposed to divisions by 2 more than once over the even natural numbers. The sequence also gives the same 50:50 probability of consecutive Collatz even elements when counted for division by 2 more than once as opposed to division by 2 once ratio of 2.8. Considering Collatz function producing random numbers, over sufficient iterations this probability distribution produces a descending order of its elements that leads to convergence of the Collatz function to the cycle 1-2-4-1.

Introduction

Colatz conjecture concerns natural numbers treated as mod 2 of non-negative even integers. It is defined by the function,

$$f(x) = \begin{cases} \frac{x}{2} & if \quad x \equiv 0 \pmod 2 \\ 3x + 1 & if \quad x \equiv 1 \pmod 2 \end{cases}$$

It simply asks you to keep dividing any even non-negative integer repeatedly by 2 until it becomes an odd integer, then convert it to even integer by tripling it and adding 1 to it then repeat the process. Many studies of the conjecture have been discussed and summarized. [1][2][3][4][5] The conjecture predicts that the recurring process will always descend on the natural numbers to cycle around the trivial cycle 1-4-2-1. The conjecture involves the natural even numbers and it simply asks, under any complete process of the conjecture, why is it always the case that over statistically sufficient number of iterations, the decrease made by the divisions by 2 exceeds the increase made by the conversions from oddness to evenness? It has been noticed here that, from a start odd integer, one iteration either increases the number when the result-even number is divided by 2 ***only once*** to obtain an odd number or decreases the start number when the result-even number is divided by 2 ***more than once***. Therefore, we seek here to quantify the decrease and the increase probabilities of the start number after every iteration and generalize that over a sufficient number of iterations to check convergence of the function. It is claimed here that the function decreases the start number until reaches a cycle, because statistically the sequence of the consecutive even integers of the elements of Collatz function (validated by deducting 1 from the even integer and dividing by 3 to produce odd positive integer) has a recurrent pattern of $x, 1, x, 1$ ... of division by 2 more than once compared to division by 2 once of probability 50:50 and ratio of about 2.8:1, where x is division by 2 more than once.

Modular arithmetic is mainly concerned with comparing mathematical processes involving division by an integer (mod) and its remainder, largely disregarding the number of iterations involved. While congruence properties help much in identifying a point on a graph that complies with a modular function, it largely disregards the length of the path to that point which in some cases is a leading characteristic to identifying an output of the function. To illustrate the concept, if a function requires to compare how much juice is in two boxes of oranges with comparable number of oranges governed by the modularity of the function that identifies how oranges are "sequenced" in each box, it is the number of iterations in the mathematical process is what is at stake here that quantifies the weight of the function rather than the mod or the remainder involved, i.e., two boxes containing 40 and 20 oranges with characteristic $\bmod (2) \equiv 0$ function won't help in quantifying the juice content of the two boxes as the functions of both of the boxes evaluate to $0 \mod 2$. But a function that defines the number of iterations of division by 2 for each box can compare the amount of juice in each box and defines it as 2:1 for the example above. Another example is two competitors on a race track with marks every 2 meters and length 200 m. A modular function of their performance with mod 2 will conclude that both will eventually finish the track but unless the function is a power function that describes the number of iterations (velocity, or how many of the two meters every second they can cover), the function cannot tell who wins the race.

Colatz conjecture seems to produce random numbers and generate a random walk process locally but globally converges to 1. Therefore, to prove the conjecture probabilistically we can show that globally the recurrence of divisions of Collatz even elements by 2 more than once to reach an odd number has the same probability as that of their recurrent divisions by 2 once, denoted here as recurrent frequency (RF), and averages by the ratio of about 2.8:1. Summing over the respective divisions will always lead by a margin that offsets the increase of the recurrent sum made by the recursive conversion process of the odd Collatz number to even number by tripling it and adding 1 to it. This is easily noticeable if we recognize that if the non-negative even integers were sequenced by increase by 2, e.g., $2, 4, 6, 8, \ldots$., division by 2 over the non-negative even integers follows a sequential order that is described as follows: if any of the sequence's even elements produces an odd number when divided by 2 once, the following element in the sequence must produce an odd number by division by 2 more than once. This produces a 50-50 probabilistic RF of division by 2 over the non-negative even integers and turns what seems a random distribution of division by 2 to a global process that makes the event of divisions by 2 recurrence over the whole non-negative even integers progress according to the sequence $x, 1, x, 1$ ..., where x is the number of divisions of the even number by 2 more than once to produce an odd number. Here we prove that Collatz-even numbers also follow the same 50:50 probability distribution that leads to descent convergence of the function to a cycle. The proposed proof of the Colatz conjecture here is complete if its process only cycles about 1, 4, and 2, since the decrease of the sequence of the global Colatz process is assembled from a correlated probabilistic events defined by the sequence $x, 1, x, 1$ ...., over the function's even elements. This probabilistic correlation is not heuristically derived as opposed to the well-known heuristic argument of the function by Lagarias and found in many references and states that the function produces division by 2 once $1/2$ of the time and division by 2 twice $1/4$ of the time and by three times $1/8$ …etc.

Division by 2 sequence of non-negative even integers

For comparison and to easy identify the RF sequence for Collatz function elements, we first generate the RF of positive even integers.

Lemma 1 Let x be any non-negative even integer that can be divided by 2 only once to yield an odd positive integer, then the next even integer $x+2$ must be divided by 2 by more than once to yield an odd positive integer.

Proof If $\frac{x}{2} \rightarrow odd$ by initial definition, then $\frac{x+2}{2} \rightarrow Parity$. Adding the LHS expressions yields $x+1$, an odd number. This necessitates that $\frac{x+2}{2} \rightarrow even$ and the term $x+2$ divisible by 2 more than once.

Lemma 2 Let x be any non-negative even integer that can be divided by 2 only once to yield an odd positive integer, then the second to next even integer $x+4$ must be divided by 2 only once to yield an odd positive integer.

Proof If $\frac{x}{2} \rightarrow odd$ by initial definition, then $\frac{x+4}{2} \rightarrow Parity$. Adding the LHS expressions yields $x+2$, an even number. This necessitates $\frac{x+4}{2} \rightarrow odd$ and the term $x+4$ divisible by 2 only once to obtain an odd number.

From Lemma 1 and 2, we generate a table of positive even integers and their corresponding frequencies of division by 2 until reaching an odd parity. Starting with the first row as the even integers made by the term 2^s, $s \in z^+$ with elements as the frequencies of division by 2 and spans the natural numbers we can construct a "RF table" over all positive integers that identifies Collatz elements with the back-bone as the line of integers that collapse to 1 by repeated divisions by 2 made by the even numbers 2^s, as Collatz function requires. This row makes a symmetrical line that contains all even numbers made by collatz function that collapse to the trivial cycle 1-4-2-1, e.g., 2,4,16.. We then construct columns in ascending order by increase by 2 to produce all even non-negative integers with each column ending by an even number that is two less than the next integer on the collapsing symmetrical line. The symmetrical line in the table has symmetrical sequential frequencies for all of the columns to infinity along the rows and makes rows with equal frequencies because of the ordered repeated frequencies for each column, which allows us to estimate relative RFs, a key probability distribution that allows us to conclude that Collatz conjecture converges probabilistically to a cycle. The table is constructed in this order mainly to be able to count frequencies of divisions by 2 and approximate the relative RFs of even non-negative integers to yield an odd number. It follows that consecutive Collatz function even elements (in bold) also follows the same pattern as those of the sequence of the table of $x, 1, x, 1$ .... We also constructed the table with the variable s spanning all non-negative integers of 2^s on the symmetrical line, not just even Collatz function elements that contribute to the collapse process of Collatz function, to produce a line of all powers of twos.

Starting with any natural number, Collatz function produces numbers in seemingly random way locally but globally the numbers decrease and the process proceeds toward the collapsing symmetrical line and to the left on the table and it eventually hits a number on the symmetrical line and then collapses to 1 and cycles around 1-4-2-1.

#	F	Int.	F	Int.	F	Int.	F	Int.	F	Int.	F	Int.	F	Int.	F	Int.	F	Int.
2^s	3	8	4	**16**	5	32	6	**64**	7	128	8	**256**	9	512	10	**1024**	30	**1073741824**
2	1	**10**	1	18	1	**34**	1	66	1	**130**	1	258	1	514	1	1026	1	1073741826
3	2	12	2	20	2	36	2	68	2	132	2	260	2	516	2	1028	2	1073741828
4	1	14	1	**22**	1	38	1	**70**	1	134	1	**262**	1	518	1	**1030**	1	**1073741830**
5			3	24	3	**40**	3	72	3	**136**	3	264	3	**520**	3	1032	3	1073741832
6			1	26	1	42	1	74	1	138	1	266	1	522	1	1034	1	1073741834
7			2	**28**	2	44	2	**76**	2	140	2	**268**	2	524	2	**1036**	2	**1073741836**
8			1	30	1	**46**	1	78	1	**142**	1	270	1	**526**	1	1038	1	1073741838
9					4	48	4	80	4	144	4	272	4	528	4	1040	4	1073741840
10					1	50	1	**82**	1	146	1	**274**	1	530	1	**1042**	1	**1073741842**
11					2	**52**	2	84	2	**148**	2	276	2	**532**	2	1044	2	1073741844
12					1	54	1	86	1	150	1	278	1	534	1	1046	1	1073741846
13					3	56	3	**88**	3	152	3	**280**	3	536	3	**1048**	3	**1073741848**
14					1	**58**	1	90	1	**154**	1	282	1	**538**	1	1050	1	1073741850
15					2	60	2	92	2	156	2	284	2	540	2	1052	2	1073741852
16					1	62	1	**94**	1	158	1	**286**	1	542	1	**1054**	1	**1073741854**
17							5	96	5	**160**	5	288	5	**544**	5	1056	5	1073741856
18							1	98	1	162	1	290	1	546	1	1058	1	1073741858
19							2	**100**	2	164	2	**292**	2	548	2	**1060**	2	**1073741860**
20							1	102	1	**166**	1	294	1	**550**	1	1062	1	1073741862
21							3	104	3	168	3	296	3	552	3	1064	3	1073741864
22							1	**106**	1	170	1	**298**	1	554	1	**1066**	1	**1073741866**
23							2	108	2	**172**	2	300	2	**556**	2	1068	2	1073741868
24							1	110	1	174	1	302	1	558	1	1070	1	1073741870
25							4	**112**	4	176	4	**304**	4	560	4	**1072**	4	**1073741872**
26							1	114	1	**178**	1	306	1	**562**	1	1074	1	1073741874
27							2	116	2	180	2	308	2	564	2	1076	2	1073741876
28							1	**118**	1	182	1	**310**	1	566	1	**1078**	1	**1073741878**
29							3	120	3	**184**	3	312	3	**568**	3	1080	3	1073741880
30							1	122	1	186	1	314	1	570	1	1082	1	1073741882
31							2	**124**	2	188	2	**316**	2	572	2	**1084**	2	**1073741884**
32							1	126	1	**190**	1	318	1	**574**	1	1086	1	1073741886
33									6	192	6	320	6	576	6	1088	6	1073741888
34									1	194	1	**322**	1	578	1	**1090**	1	**1073741890**
35									2	**196**	2	324	2	**580**	2	1092	2	1073741892
36									1	198	1	326	1	582	1	1094	1	1073741894
37									3	200	3	**328**	3	584	3	**1096**	3	**1073741896**
38									1	**202**	1	330	1	**586**	1	1098	1	1073741898
39									2	204	2	332	2	588	2	1100	2	1073741900
40									1	206	1	**334**	1	590	1	**1102**	1	**1073741902**
41									4	**208**	4	336	4	**592**	4	1104	4	1073741904
42									1	210	1	338	1	594	1	1106	1	1073741906

43									2	212	2	**340**	2	596	2	**1108**	2	**1073741908**
44									1	**214**	1	342	1	**598**	1	1110	1	1073741910
45									3	216	3	344	3	600	3	1112	3	1073741912
46									1	218	1	**346**	1	602	1	**1114**	1	**1073741914**
47									2	**220**	2	348	2	**604**	2	1116	2	1073741916
48									1	222	1	350	1	606	1	1118	1	1073741918
49									5	224	5	**352**	5	608	5	**1120**	5	**1073741920**
50									1	**226**	1	354	1	**610**	1	1122	1	1073741922
51									2	228	2	356	2	612	2	1124	2	1073741924
52									1	230	1	**358**	1	614	1	**1126**	1	**1073741926**
53									3	**232**	3	360	3	**616**	3	1128	3	1073741928
54									1	234	1	362	1	618	1	1130	1	1073741930
55									2	236	2	**364**	2	620	2	**1132**	2	**1073741932**
56									1	**238**	1	366	1	**622**	1	1134	1	1073741934
57									4	240	4	368	4	624	4	1136	4	1073741936
58									1	242	1	**370**	1	626	1	**1138**	1	**1073741938**
59									2	**244**	2	372	2	**628**	2	1140	2	1073741940
60									1	246	1	374	1	630	1	1142	1	1073741942
61									3	248	3	**376**	3	632	3	**1144**	3	**1073741944**
62									1	**250**	1	378	1	**634**	1	1146	1	1073741946
63									2	252	2	380	2	636	2	1148	2	1073741948
64									1	254	1	**382**	1	638	1	**1150**	1	**1073741950**
65											7	384	7	**640**	7	1152	7	1073741952
											..	..	..	..	..	..	..	..

Table 1 Table of non-negative even integers and their corresponding frequencies of division by 2. Any row has the same frequency.

The symmetrical distribution of frequencies of divisions by 2 of even natural numbers as in the table exhibits a classical probability distribution function about the collapsing symmetrical line over the natural numbers. Only those numbers on the symmetrical line that satisfy Collatz function can branch out and contribute to the collapse process to 1 (those numbers with s an even integer), i.e. the number 2^8 (256) contributes to the collapse process because you can deduct 1 from it and divide by 3 to get a whole number, but the number 2^9 (512) does not, and the number 341 leads to 2^{10} (1024) on the symmetrical line that collapses to 1 while the odd number 357913941 ends with 2^{30} (1073741824) on the symmetrical line as well. Those numbers on the symmetrical line that can be traced backward by the function $3x + 1$ act as points for branching out to trace the Collatz tree where the symmetrical line is the tree's trunk.

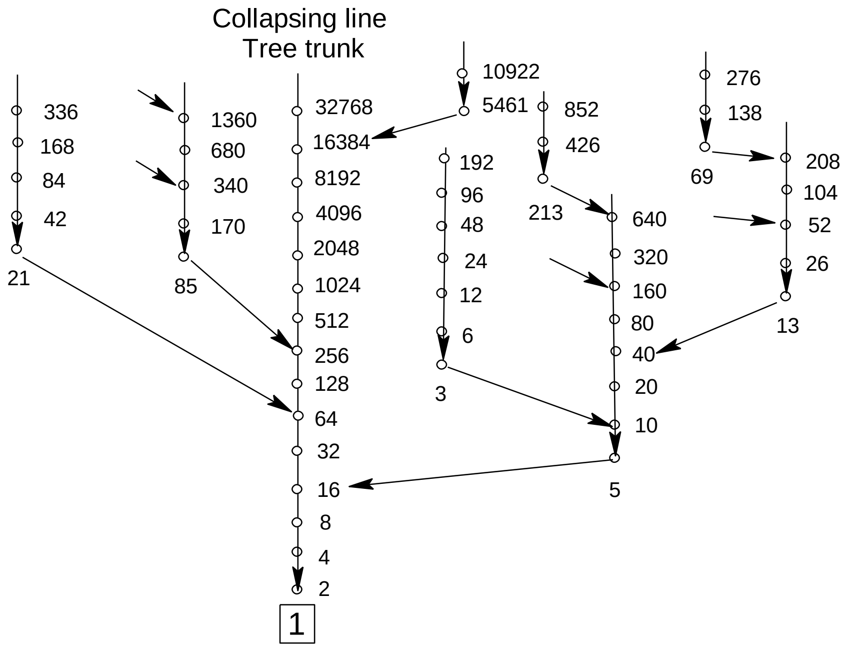

Fig. 1 Representative of Collatz tree

Probability distribution of even natural numbers in terms of division by 2

The ordered distribution of even natural numbers in terms of division by 2 about the symmetrical line represents a classical probability distribution.

Lemma 3 The probability of division by 2 more than once and division by 2 once for a randomly chosen non-negative even number is $P(\text{more than once}) = P(\text{ once}) = 1/2$.

Proof It follows from the ordered distribution of division by 2 once followed by division by 2 more than once by Lemma 1 and Lemma 2. The probability is easy to check in table 1.

Probability distribution of Collatz function's even elements

Table 1 represents a 50:50 probability of the Collatz function's even elements over the even natural numbers in terms of their RF. The equal probability distributions of both validates the adoption of table 1 as representative to the distribution of even elements of Collatz function.

Lemma 4 Collatz even elements are sequenced every three consecutive numbers on the sequence of the even nonnegative integers with probability of division by 2 more than once as opposed to division by 2 once, to obtain an odd number for a randomly chosen Collatz element, is $P(\text{more than once}) = P(\text{ once}) = 1/2$.

Proof Let x be any Collatz even element that can be divided by 3 after deducting 1, then the next even integer $x + 2$ is not a Collatz even element since it doesn't follow that restriction, neither is the next one, but the one follows is, since a Collatz element is restricted by,

$$(x + k) - 1$$

Which is divisible by 3 if the variable k is multiples of 6 only and leads to the fourth consecutive integer after the variable x on the table of nonnegative even integers.

Further, let x be any Collatz even element that is only divisible by 2 once, the following set of logical equations describes RF of all of Collatz even elements by obtaining their parity,

First, if $\frac{x}{2} \rightarrow odd$ by initial definition, then $\frac{x+6}{2} \rightarrow Parity$. Adding the LHS expressions yields $x + 3$, an odd number. This necessitates that $\frac{x+6}{2} \rightarrow even$ and the term $x + 6$ is even number and therefore it is divisible by 2 more than once to obtain and odd number.

Second, if $\frac{x}{2} \rightarrow odd$ by initial definition, then $\frac{x+12}{2} \rightarrow Parity$. Adding the LHS expressions yields $x + 6$, an even number. This necessitates that $\frac{x+12}{2} \rightarrow odd$ and that the term $x + 12$ is divisible by 2 only once to obtain an odd number.

This is shown by quick inspection of the sequence of the even natural numbers by deducting 1 followed by division by 3 (bold face).

Note: Lemma 4 can be generalized to any generalized Collatz function in the form of $kx + 1$, where k is an odd number and their corresponding sequence on the even nonnegative integers sequence can be derived accordingly.

The ratio of frequencies of divisions by 2 of Collatz even elements

Since RF is ordered perfectly among all non-negative even integers as well as among Collatz even elements as table 1 indicates, any large sample is a true representation to compute RF of Collatz function's elements.

Lemma 5 The sum of divisions by 2 more than once is on average 2.8 times the sum of divisions by 2 once over the Collatz even elements and the ratio increases over yet larger samples.

Proof Quick inspection of table 1 verifies the prediction. The ratio is close to 2.8:1 and was approximated from table 1 above by taking a specimen of the first 2000 counts excluding Collatz elements present in the symmetrical line since the function collapses there. The specimen is large enough to describe the distribution of Collatz elements since it exhibits an ordered distribution over the nonnegative even integers. With larger sample it was found that the ratio increases slightly to 3. END.

Theory 1 Collatz function process must produce a descending order of numbers over adequate number of iterations.

Proof Considering a random distribution of integers produced by Collatz function processes, the probability of division by 2 more than once to division by 2 once of Collatz elements is then $1/2$. Therefore we can average Collatz processes as two consecutive steps from a start number of an odd Collatz element that ends up in an odd number. The first step is to increment it by tripling it and adding 1 and then divide by 2 once, which increase the start number while the second step is to increment the result again but divide by two, 2.8 times on average, which decreases the start number by a larger magnitude than the increase. This will always result in an end number less than the start number confirming the prediction of Collatz conjecture. To see this, let n be the start-odd number. Then, the two processes yield an end number according to the nested step,

$$End = \frac{3\left(\frac{3n+1}{2}\right)+1}{2^{l}}$$

Where l the RF of 2.8. Simplifying yields,

$$End = 0.646n + 0.359\ldots\ldots\ldots\ldots\ldots(1)$$

Equation (1) always results in end number less than the start number, $End < n$ for $n > 1$, because the coefficient of n is a fraction. Also, for large numbers, RF grows slowly as can be verified in the table. This produces even smaller coefficient of n and brings the process to higher descending rate. END

Collatz function then produces steps that end with even numbers that descends tracing down a line of multiples of an odd number by the process of division by 2 until it eventually reaches an odd number whose ascending step is on the symmetrical line and collapses to the ultimate odd number of 1.

Example 1 A representative example of the average iterations is the odd number 3 as the start number. With 50:50 probability of division by 2 more than once as opposed to division by 2 once with ratio 2.8:1 and a repeating pattern, the first iteration of the function $3x + 1$ gives 10. Division by 2 once give 5. Since the first iteration gives division by 2 once, the next one must divide by 2 more than once. The second iteration gives 16 and division by two 2.8 times (division by $2^{2.8}$) gives 2.297. Since the number of iterations involved is two and represents a 50:50 probability, the process is a true representation and leads to end number less than the start number in line with theory 1.

Example 2 Start odd number is 9999. Incrementing it by Collatz function $3x + 1$ and divide by 2 once yields 14999. Repeating the process with 14999 as the start number but divide by $2^{2.8}$ yields and end number of 6461. The end number is less than the start number.

Comparison with other generalized Collatz functions

Other generalized Collatz functions such as $5x + 1$ and $7x + 1$ do have probability distributions of their even elements in terms of division by 2, e.g., Inspection of table 1 reveals that the function $5x + 1$ has even integers every four consecutive integers over the even integers sequence with a 50:50 ratio of division by 2 more than once as opposed to once. The same goes with the function $7x + 1$ with spacing of seven consecutive integers. Therefore, to check for the divergence of those functions, we must compute the relative frequencies of divisions by 2 that contribute to the rise of the start number as opposed to those that contribute to its descend. It is noticed here that, unlike the function $3x + 1$, division by 2 contribute differently to the increase or decrease of the start number. i.e., Besides the fact the you multiply the start number by 5 instead of 3, division by 2 once as well as twice with the function $5x + 1$ contributes to the rise of the start number and its process then produces an equation with coefficient of n much large compared to the function $3x + 1$ leading to the divergence of the function.

Question: why the function as $5x + 1$ does not eventually reach the symmetric line on its ascending path and collapse to 1?

Answer: because none of the even numbers on the symmetrical line belongs to the function's elements since deducting by 1 from all its even integers don't produce odd integers that evaluate to 0 mod 5.

The only trivial cycle of the Collatz function is 1-4-2-1

It is easy to prove that the cycle 1-4-2-1 is the only trivial cycle for Collatz function.

Lemma 6 Let x be any positive odd integer. Then the only trivial cycle of the Collatz function is 1-4-2-1.

Proof The equation,

$$x = \frac{3x + 1}{2^l}$$

describes a trivial cycle where l is an integer equals the number of divisions by 2. Solving for x yields,

$$x = \frac{1}{2^l - 3}$$

For non-negative integer solution, l must be 2, x must be 1 and $3x + 1$ must be 4. That is because if $l > 2$, the expression yields x as fraction and if $l < 2$, the expression yields negative value. This leaves $l = 2$ the only solution to the equation with a start number 1.

Non-trivial nested cycle

Collatz conjecture forbids looping anywhere on the Collatz tree except at the bottom of the trunk as clear in Fig 1. Starting at any point on the tree, Collatz function allows the process only to head in one direction from one point to another on a sub-branch to another sub-branch leading to a main branch and then to the trunk and finally collapsing to the loop 1-4-2-1. A global nested path of Collatz conjecture is represented by equation 1. The only closed (nested) cycle with the start number equals the end number as equation 1 requires is for $n = 1$. This may not be generalized locally for any relatively small degree of nesting of the function which, according to the conjecture, prohibits the return to the same start number different than 1. Since we assume that the function heads to stochastic behavior very fast, we may assume with small degree of certainty that the function does not trace back to the same start number and hit a cycle somewhere on the sequence of integers. In comparison with the function $5x + 1$ that has no elements on the trunk of the tree and therefore any of its cycles must end up with an odd number other than 1 (see figure 1), the function $3x + 1$ has elements on the trunk and it collapses to 1 if it happens that the function's path reaches the trunk to cycle around the trivial cycle 1-4-2-1. Two known cycles of $5x + 1$ is 17-27-43-17 and 13-83-33-13. The zigzaging up and down of the function starts with the column with start number 13 of Collatz tree, then alternates between other columns, here 81 and 33 for the second cycle, and returns to the column of start number 13. Any number of the four numbers can be the starting and ending number. Obviously, the three columns involved must contain elements of the function.

Conclusion

The convergence of Collatz conjecture function was proven by identifying a sequence of the non-negative even integers that produces a probability of $1/2$ of the division of the integers by 2 more than once and their division by 2 once with a ratio of 2.8:1. For any positive odd integer, the collective divisions by 2 more than once that produced a total decrease of the start number in the Collatz process was found to exceed the total increase of the start number produced by division by 2 once. The process indicates a systematic global decrease until one event matches an even number on the symmetrical line and collapses to 1 and loops the cycle 1-4-2-1.

References

[1] J. Lagarias, The $3x + 1$ Problem and its Generalization, Amer. Math. Monthly, 92(1)

(1985) 3-23

[2] The Collatz Problem in the Light of an Infinite Free Semigroup, Manfred Trümper

Chinese Journal of Mathematics Research Article (21 pages), Article ID 756917, Volume 2014 (2014)

[3] J. Lagarias. The 3x + 1 Problem: An Annotated Bibliography (1963–1999). Department of Mathematics, University of Michigan Ann Arbor, (2011), 48109–1109, lagarias@umich.edu.

[4] J. Lagarias. The 3x + 1 Problem: An Annotated Bibliography, II (2000–2009). Department of Mathematics, University of Michigan Ann Arbor, (2011), 48109–1109, lagarias@umich.edu.

[5] J. C. Lagarias and A. Weiss, *The 3x+1 problem: two stochastic models*, The Annals of Applied Probability, vol. 2, no. 1, pp. 229-261, 1992.