

Kamal Barghout

A probabilistic proof of the convergence of Collatz conjecture

Prince Mohammad Bin Fahd University

Alkhobar

Saudi Arabia

kbarghout@pmu.edu.sa

966552094285

Acknowledgments I am grateful for the help and encouragement I received from Prince Mohammad Bin Fahd University.

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 $x, 1, x, 1 \dots$, where x represents divisions by 2 more than once. The sequence presents a probability of 50:50 of division by 2 more than once as opposed to divisions by 2 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 and a ratio of 3:1 of high counts. Considering Collatz function producing random numbers and over sufficient iterations, this probability distribution produces numbers in descending order that lead to the convergence of the Collatz function to 1, assuming the only cycle of the function is 1-4-2-1.

Mathematics Subject Classification. 11K31, 11K36

Keywords. Collatz conjecture, probability distribution.

1. Introduction

Collatz conjecture concerns natural numbers treated as $(\text{mod } 2)$ of positive even integers. It is defined by the function,

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

It simply asks you to keep dividing any positive even 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. The conjecture has been widely studied [1-5]. It predicts that the recurring process will always form a sequence that descends 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 positive 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 probabilistically 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 it reaches a cycle, because statistically the sequence of all of the consecutive even integers of the elements of Collatz function over the natural numbers (validated by deducting 1 from every even positive integer and then checking divisibility by 3) has a recurrent pattern of $x, 1, x, 1 \dots$ of division by 2 more than once compared to division by 2 once of probability 50:50 and ratio of about 3:1, where x is division by 2 more than once.

Collatz 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 3:1. 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 positive even integers were sequenced by increase by 2, e.g., 2, 4, 6, 8, ..., division by 2 over the positive 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 hidden regularity produces a 50-50 probabilistic RF of division by 2 over the positive even integers and turns what seems a random distribution of division by 2 to a global process that makes the events of division by 2 recurrence over the whole positive even integers progress according to the sequence $x, 1, x, 1 \dots$, 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 sequence made by the function to a cycle. The proposed proof of the Collatz conjecture here is complete if its process only cycles about 1, 4, and 2, since the decrease of the sequence of the global Collatz process is assembled from a perfect correlated probabilistic events defined by the sequence $x, 1, x, 1 \dots$, 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 found in many references which states that the function averages division by 2 once 1/2 of the time and division by 2 twice 1/4 of the time and division by 2 three times 1/8 of the time ...etc., which produces a decrease of 3/4 of the preceding number each iteration on average. In this paper we claim that the function $3x + 1$ produces an increase of the start number of 50%, 1/2 of the time as opposed to a decrease of the start number of 62% the other 1/2 of the time, averaged over a sample of large number of Collatz even integers if we assume that the mixing properties of the function's even integers are truly picked at random in the process.

2. Division by 2 sequence of positive even integers

For comparison and to easy identify the RF sequence of division by 2 for Collatz function elements, we first generate the RF sequence of positive even integers.

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

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

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

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

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 , $s \in \mathbb{Z}^+$ 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., 4, 8, 16, 32. We then construct columns in ascending order by increase by 2 to produce all even positive integers with each column ending by an even number that is two less than the next integer on the collapsing symmetrical line. We observe that 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 positive integers to yield an odd number. It follows that consecutive Collatz function’s even elements (in bold) also follow the same pattern as those of the sequence of the table of $x, 1, x, 1 \dots$. We also constructed the table with the variable s spanning all positive 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.

#	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 positive even integers and their corresponding frequencies of division by 2. Any row has the same frequency.

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 in a deterministic process.

The symmetrical distribution of frequencies of divisions by 2 of even natural numbers as in the table exhibits a classical probability distribution 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), and the branches that are connected via the function with odd start numbers making new sub-branch on the Collatz tree (see Fig. 1) that if a branch is reached, the process will collapse to its start odd number i.e. on the trunk of the tree, 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.

3. Perfect symmetry of the table of positive even integers with pivotal frequencies

Looking for hidden symmetry in the background of even integers in terms of RFs is of prime importance to assign symmetry to the RFs of the $3x + 1$ function and determine their probability density. That is because Collatz function's elements occur sequentially every third element in the table of positive integers. Quick observation of the table of positive even integers reveals that each column is exactly symmetrical about a pivotal new frequency that is next number on the number line to the pivotal frequency in the preceding column and equals $(s - 1)$ of its own, i.e., the

column 2^5 (32) has double the frequencies of the preceding column of 2^4 about the pivotal RF of 4, with pivotal RF greater than the preceding one of 3 of the preceding column by 1 and equals $(S - 1)$, where s is 5. Those pivotal RFs increase to infinity by the increase of the numbers of the column about the symmetrical line of 2^s .

	F	Int.	F	Int.
2^s	4	16	5	32
1	1	18	1	34
2	2	20	2	36
3	1	22	1	38
4	3	24	3	40
5	1	26	1	42
6	2	28	2	44
7	1	30	1	46
8			4	48
9			1	50
10			2	52
11			1	54
12			3	56
13			1	58
14			2	60
15			1	62

Table 2 Representative of pivotal RF in the table of positive even integers.

4. Perfect symmetry of the table of $3x + 1$ elements

The perfect symmetrical RFs of the terms of the $3x + 1$ function allows us to determine their probability density distribution. Similar to the columns of the even integers in table 1, the columns of the terms of $3x + 1$ are symmetrical about a number at the center of their heights, i.e., the column of 2^{l+1} of RF = 6 has a pivotal F of 5 (see table 3). The symmetry allows taking the average of RFs of the terms of the function $3x + 1$ that divide by 2 more than once as a measure of the behavior of the function's complete process that leads to a descending trajectory. The ratio of 3:1 of division by 2 more than once of the RFs compared to division by 2 once was computed from a table with the first 1000 counts.

	F	$3x + 1$	F	$3x + 1$	F	$3x + 1$	F	$3x + 1$	F	$3x + 1$
2^{l+1}	4	16	6	64	8	256	10	1024	12	4096
1	1	22	1	70	1	262	1	1030	1	4102

2	2	28	2	76	2	268	2	1036	2	4108
3	1	34	1	82	1	274	1	1042	1	4114
4	3	40	3	88	3	280	3	1048	3	4120
5	1	46	1	94	1	286	1	1054	1	4126
6	2	52	2	100	2	292	2	1060	2	4132
7	1	58	1	106	1	298	1	1066	1	4138
8			4	112	4	304	4	1072	4	4144
9			1	118	1	310	1	1078	1	4150
10			2	124	2	316	2	1084	2	4156
11			1	130	1	322	1	1090	1	4162
12			3	136	3	328	3	1096	3	4168
13			1	142	1	334	1	1102	1	4174
14			2	148	2	340	2	1108	2	4180
15			1	154	1	346	1	1114	1	4186
16			5	160	5	352	5	1120	5	4192
17			1	166	1	358	1	1126	1	4198
18			2	172	2	364	2	1132	2	4204
19			1	178	1	370	1	1138	1	4210
20			3	184	3	376	3	1144	3	4216
21			1	190	1	382	1	1150	1	4222
22			2	196	2	388	2	1156	2	4228
23			1	202	1	394	1	1162	1	4234
24			4	208	4	400	4	1168	4	4240
25			1	214	1	406	1	1174	1	4246
26			2	220	2	412	2	1180	2	4252
27			1	226	1	418	1	1186	1	4258
28			3	232	3	424	3	1192	3	4264
29			1	238	1	430	1	1198	1	4270
30			2	244	2	436	2	1204	2	4276
31			1	250	1	442	1	1210	1	4282
32					6	448	6	1216	6	4288
33					1	454	1	1222	1	4294
34					2	460	2	1228	2	4300
35					1	466	1	1234	1	4306
36					3	472	3	1240	3	4312
37					1	478	1	1246	1	4318
38					2	484	2	1252	2	4324
39					1	490	1	1258	1	4330
40					4	496	4	1264	4	4336
41					1	502	1	1270	1	4342
42					2	508	2	1276	2	4348
43					1	514	1	1282	1	4354
44					3	520	3	1288	3	4360

45				1	526	1	1294	1	4366
46				2	532	2	1300	2	4372
47				1	538	1	1306	1	4378
48				5	544	5	1312	5	4384
49				1	550	1	1318	1	4390
50				2	556	2	1324	2	4396
51				1	562	1	1330	1	4402
52				3	568	3	1336	3	4408
53				1	574	1	1342	1	4414
54				2	580	2	1348	2	4420
55				1	586	1	1354	1	4426
56				4	592	4	1360	4	4432
57				1	598	1	1366	1	4438
58				2	604	2	1372	2	4444
59				1	610	1	1378	1	4450
60				3	616	3	1384	3	4456
61				1	622	1	1390	1	4462
62				2	628	2	1396	2	4468
63				1	634	1	1402	1	4474
64				7	640	7	1408	7	4480
65				1	646	1	1414	1	4486
66				2	652	2	1420	2	4492
67				1	658	1	1426	1	4498
68				3	664	3	1432	3	4504
69				1	670	1	1438	1	4510
70				2	676	2	1444	2	4516
71				1	682	1	1450	1	4522
72				4	688	4	1456	4	4528
73				1	694	1	1462	1	4534
74				2	700	2	1468	2	4540
75				1	706	1	1474	1	4546
76				3	712	3	1480	3	4552
77				1	718	1	1486	1	4558
78				2	724	2	1492	2	4564
79				1	730	1	1498	1	4570
80				5	736	5	1504	5	4576
81				1	742	1	1510	1	4582
82				2	748	2	1516	2	4588
83				1	754	1	1522	1	4594
84				3	760	3	1528	3	4600
85				1	766	1	1534	1	4606
86				2	772	2	1540	2	4612
87				1	778	1	1546	1	4618

88				4	784	4	1552	4	4624
89				1	790	1	1558	1	4630
90				2	796	2	1564	2	4636
91				1	802	1	1570	1	4642
92				3	808	3	1576	3	4648
93				1	814	1	1582	1	4654
94				2	820	2	1588	2	4660
95				1	826	1	1594	1	4666
96				6	832	6	1600	6	4672
97				1	838	1	1606	1	4678
98				2	844	2	1612	2	4684
99				1	850	1	1618	1	4690
100				3	856	3	1624	3	4696
101				1	862	1	1630	1	4702
102				2	868	2	1636	2	4708
103				1	874	1	1642	1	4714
104				4	880	4	1648	4	4720
105				1	886	1	1654	1	4726
106				2	892	2	1660	2	4732
107				1	898	1	1666	1	4738
108				3	904	3	1672	3	4744
109				1	910	1	1678	1	4750
110				2	916	2	1684	2	4756
111				1	922	1	1690	1	4762
112				5	928	5	1696	5	4768
113				1	934	1	1702	1	4774
114				2	940	2	1708	2	4780
115				1	946	1	1714	1	4786
116				3	952	3	1720	3	4792
117				1	958	1	1726	1	4798
118				2	964	2	1732	2	4804
119				1	970	1	1738	1	4810
120				4	976	4	1744	4	4816
121				1	982	1	1750	1	4822
122				2	988	2	1756	2	4828
123				1	994	1	1762	1	4834
124				3	1000	3	1768	3	4840
125				1	1006	1	1774	1	4846
126				2	1012	2	1780	2	4852
127				1	1018	1	1786	1	4858
128						8	1792	8	4864

Table 3 Representative of pivotal RF in the $3x + 1$ integers.

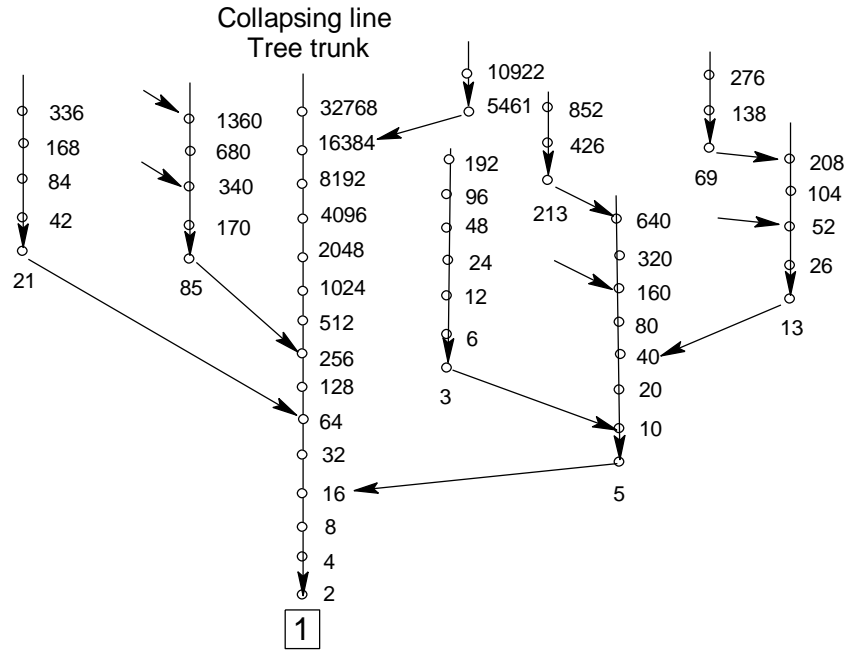


Fig. 1 Representative of Collatz tree

5. 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 positive 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. \square

6. 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 RFs of division by 2. 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 n be any Collatz even element that can be divided by 3 after deducting 1, then the next even integer $n + 2$ 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,

$$(n + k) - 1$$

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

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

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

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

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.

7. Probability density of RFs of Collatz even elements

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

Lemma 5 The sum of divisions by 2 more than once is on average 2.85 times the sum of divisions by 2 once over the Collatz even elements and the ratio stabilizes over large samples to about 3 at high counts.

Proof Quick inspection of table 1 verifies the prediction. The ratio is close to 3:1 and was approximated from table 1 above by taking a specimen of the first 1000 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 a perfect ordered distribution over the nonnegative even integers with ordered increase of pivotal RFs from one column to the next. A ratio of 3:1 was obtained with a sample of large counts since the RFs increase due to increasing pivotal RFs. \square

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

Proof Considering the apparent 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 distinctive operations of a start number of an odd Collatz element that ends up in an odd number. The first operation is to increment it by tripling it and adding 1 and then divide by 2 once, which increases the start number while the other operation is to increment the start number but divide by 2 three times, which decreases the start number by a larger magnitude than the increase, in line with Collatz conjecture. To see this, let n be the start-odd number. Then applying the function $3n + 1$ and dividing by 2 once gives an end number,

$$End = \frac{3n + 1}{2}$$

$$End = 1.5n + 0.5$$

For large enough numbers, this equation gives an increase of about 1/2 of the start number. If we divide the function by 2^3 instead, the process gives,

$$End = \frac{3n + 1}{2^3}$$

$$End = 0.375n + 0.125$$

This equation gives a decrease of about 3/5 of the start number, which is larger than the increase, and leads to successive average decrease of the start number. □

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 assuming the cycle 1-4-2-1 is the only cycle in the process.

Example 1 Start odd number is 3. A 50:50 probability of division by 2 more than once as opposed to division by 2 once with ratio 3:1 and a repeating pattern, the iteration of applying the function $3n + 1$ gives 10. Division by 2 once gives 5. If we divide by 2 three times the iteration gives 1.25. This is actually an increase of the start number. Fortunately, the process hits the symmetrical line and collapses to 1 upon increase of the start number only. If we evaluate the function with a start number 5, the process also hits the symmetrical line upon increase of the start number only. Starting with the number 7 and up on the number line, in general, the collective decrease of the start number is larger than the collective increase, except for those start numbers that hit the symmetrical line before the decrease occurs, e.g., start numbers that make the main branches on the Collatz tree (see Fig. 1) such as 5, 21, 85, 5461..etc, in line with theory 1.

Example 2 Start odd number is 9999. Incrementing it by Collatz function $3n + 1$ and dividing by 2 once yields 14999. This is an increase of the number of 50%. Repeating the process but divide by 2 three times you get 3749.75. That is a decrease of the start number of about 62%. Obviously, the percentage decrease is larger than the percentage increase of the start number in line with Collatz conjecture.

8. 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, i.e., 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 for other generalized functions. i.e., besides the fact that 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 $3n + 1$ leading to the divergence of the function.

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

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

9. 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 n be any positive odd integer. Then the only trivial cycle of the Collatz function is 1-4-2-1.

Proof The equation,

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

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

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

For positive integer solution, l must be 2, x must be 1 and $3n + 1$ 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. \square

10. 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 trajectory of Collatz conjecture is represented by the sequence that defines the trajectory of a start odd number n ,

$$n, T(n), T^{(2)}(n), T^{(3)}(n), \dots$$

The sequence $T^{(k)}(n)$ of the function must become periodic with an end number equals the start number n for any $k \geq 1$ and $n \neq 1$.

Collatz conjecture suggests the nonexistence of a nested cycle with the start number equals the end number. 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 with high probability by the same reasoning of stochastic distribution of large number of elements in a sample, to hit the same number twice with a very low probability.

In comparison with the function $5n + 1$ that has no elements on the trunk of the Collatz tree and therefore any of its cycles must end up with an odd number other than 1 (see Fig. 1), the function $3n + 1$ has elements on the trunk and it collapses to 1 if it happens that the function's trajectory reaches the trunk to cycle around the trivial cycle 1-4-2-1. Two known cycles of $5n + 1$ are 17-27-43-17 and 13-83-33-13. The higher degree of zigzagging up and down of the function (compared with the function $3n + 1$) starts with the column with a start odd number 13 of Collatz tree of the specified function, 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 as well. Obviously, the three columns involved must contain elements of the function that are spaced wide enough for the trajectory to return to the same column (tree branch) it starts from. The generalized function $5n + 1$ statistically has a larger chance to hit the start branch after its launch than $3n + 1$ because the degree of zigzagging about its launch branch is higher because it has a wider spacing (multiplying by 5) and also that division by 2 once as well as twice contributing to the increase of the function as opposed to only division by 2 once that contribute to the increase of the start number for the function $3n + 1$. Overall, probabilistically, it doesn't seem there exists what prohibits a non-trivial cycle for the function $3n + 1$.

11. Conclusion

The convergence process of the $3x + 1$ function was proven by identifying a sequence of the function's positive even integers that produces a probability of 50:50 of the division of the integers by 2 more than once as opposed to their division by 2 once with a ratio of 3:1. For any positive odd integer, the collective divisions by 2 more than once that produced a total decrease of the start number in the function's trajectory 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, presuming the function yields no other cycle.

References

- [1] Lagarias, J. *The $3x + 1$ Problem and its Generalization*, Amer. Math. Monthly, 92(1) 3-23 (1985)
- [2] Manfred, T., *The Collatz Problem in the Light of an Infinite Free Semigroup*, Chinese Journal of Mathematics Research , Article ID 756917, Volume 2014 (2014)
- [3] Lagarias, J. *The $3x + 1$ Problem: An Annotated Bibliography (1963–1999)*. Department of Mathematics, University of Michigan Ann Arbor, 48109–1109 (2011)
- [4] Lagarias, J. *The $3x + 1$ Problem: An Annotated Bibliography, II (2000–2009)*. Department of Mathematics, University of Michigan Ann Arbor, 48109–1109 (2011)
- [5] Lagarias, J. C. and Weiss A., *The $3x+1$ problem: two stochastic models*, The Annals of Applied Probability, vol. 2, no. 1, pp. 229-261 (1992)