WINNING TICKETS FROM TWO COLLATZ GRAPHS' STRUCTURES

HUBERT SCHAETZEL

ABSTRACT. The Collatz conjecture states that, starting with any strictly positive integer, the (3x+1) algorithm leads systematically to the same cycle (1, 4, 2, 1, $\cdots$) after a finite number of steps. The only potential exceptions to this rule on the positive side of $\mathbb{Z}^*$ are the existence of either separate closed cycles or separate infinite divergent series. We will analyse the constraints and impediments on these types of objects using the underlying structures and laws linked to the Collatz algorithm.

CONTENTS

Date: October 17, 2023.
2020 *Mathematics Subject Classification.* 11Y55, 40A05.
Key words and phrases. Collatz conjecture, Syracuse conjecture, graph.

1. The algorithm

Let us have i an index and u_i an integer different from zero (positive or negative). The Collatz algorithm consist in dividing u_i by 2 if u_i is even and to multiply it by 3 and adding 1 otherwise in order to get u_{i+1} recursively.

$$u_{i+1} = \begin{cases} u_i/2 & \text{if } u_i \equiv 0 \pmod 2, \\ 3u_i + 1 & \text{if } u_i \equiv 1 \pmod 2. \end{cases}$$

According to the Collatz conjecture, starting with an integer in $\mathbb{N}^*$, this algorithm leads to the same cycle $(1,\, 4,\, 2,\, 1,\, \cdots)$.

2. Conventions and vocabulary

Vocabulary. *Part 1.*
We use standard graph vocabulary and a few additional conventions

- vertex *: any integer,*
- successor *: a successor vertex is obtain by applying the Collatz algorithm to an integer; an integer has one immediate successor; when speaking of successor in the singular we mean the immediate successor,*
- antecedent *: an antecedent vertex is obtain by applying the Collatz algorithm to an integer in the reverse way (upturn or upwards); an integer has one or two immediate antecedents according to its value; when speaking of antecedent in the singular we mean an immediate antecedent,*
- active vertex *: a green or blue integer (an integer equal to* 1 *or* 2 *modulo* 3*),*
- inactive vertex *: a yellow vertex (an integer equal to* 0 *modulo* 3*),*
- link *: an edge between two vertices,*
- active link *: a link giving a blue or green antecedent,*
- branch *: a set of vertices connected by some links,*
- inactive branch *: a branch with only yellow vertices,*
- graph *: an arbitrary initial choice of a vertex or cycle and then all the vertices and links formed by its successors and antecedents,*
- inactive graph *: a graph with ultimately only inactive ascendant branches; note : such graph has only a finite number of branches,*
- root *: the cycle or unique vertex (the latter being proven impossible) at the bottom of a graph,*
- rank *: the upturn step of antecedents from the root ; by extension, the term is also used for the number of upturn steps starting from some chosen integer,*
- graph crown *: the set of vertices and links of a graph except its root (or initial integer).*

Color code. *Let us consider the values of the integers modulo* 3*. We associate the green color to the* 1 (mod 3) *integers, the blue color to the* 2 (mod 3) *integers and the yellow color to the* 0 (mod 3) *integers.*

1 (mod 3)	2 (mod 3)	0 (mod 3)

Vocabulary. *Part 2.*

- integers *:* $\mathbb{N}^*$ *designates the natural numbers, while* $\mathbb{N}$ *includes* 0*. The same convention holds for* $\mathbb{Z}^*$ *and* $\mathbb{Z}$ *for integers.*
- stopping time *: executing the Collatz algorithm, the step when the absolute value of the resulting integer is equal or smaller than the chosen initial integer.*
- odd step *: a multiplicative operation* $(3x+1)$ *on* x*. The total number of odd steps is noted* v *at the stopping time.*
- even step *: a division operation* $(x/2)$ *on* x*. The total number of even steps is noted* w *at the stopping time.*

Part 1. The trees' structure

3. Graph crowns

Let us have the color code defined previously. There are 1 or 2 immediate antecedents for any integer and it is then straightforward to get the modulo values of these immediate antecedents and corresponding colors according to the six cases given in figure 1.

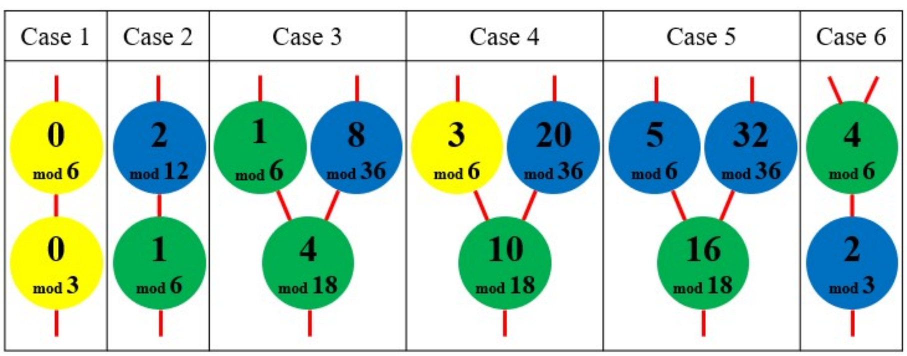

FIGURE 1. Collatz reverse algorithm : List of antecedents' modulo values.

The initial set of integers, of course, is chosen in order to cover a complete set of congruences:

Case 1	$0 \pmod 3$		
Case 2	$1 \pmod 3$	$1 \pmod 6$	
Case 3		$4 \pmod 6$	$4 \pmod{18}$
Case 4			$10 \pmod{18}$
Case 5			$16 \pmod{18}$
Case 6	$2 \pmod 3$		

It is likewise easy to verify that the antecedents' set is also a complete set of congruences:

Case 1	$0 \pmod 6$		
Case 3	$1 \pmod 6$		
Case 2	$2 \pmod 6$	$2 \pmod{12}$	
Case 3		$8 \pmod{12}$	$8 \pmod{36}$
Case 4			$20 \pmod{36}$
Case 5			$32 \pmod{36}$
Case 4	$3 \pmod 6$		
Case 6	$4 \pmod 6$		
Case 5	$5 \pmod 6$		

Lemma 1. *There is no blue antecedent to a blue vertex.*

Proof. Case 6 in figure 1 is the only alternative. It shows green vertices as antecedents. □

Lemma 2. *A green vertex may have a green antecedent vertex. But the later cannot have another green antecedent.*

Proof. Case 3 followed by case 2 in figure 1 is the only alternative and shows a blue vertex after two consecutive green vertices. □

Lemma 3. *Any branch is linked to an active branch.*

Proof. Recall, by our earlier definitions, a inactive branch contains only yellow vertices. The antecedent of a yellow vertex is unique (case 1 of figure 1) and is yellow, therefore the branch is inactive up to infinity once a yellow vertex appears. But the bottom yellow vertex of that branch has necessarily a green successor which is linked upwards to a blue vertex (case 4 of figure 1) and then again upwards to a green vertex (case 6 of figure 1). This last green vertex is equivalent to $4 \pmod 6$ and therefore has always 2 antecedents providing the début of an active branch as for there on, whatever follows, one encounters always at least a blue or a green vertex as antecedent, a yellow vertex never appearing alone (but with a blue vertex as shown in case 4). □

4. Roots

Figure 1 shows that any integer has two or three links. The 3 links' pattern is the one that allows roots to thrive with a crown graph. The objective of this section is to prove that this blossom will always occurs : Any root has a crown graph.

Lemma 4. *The two lemmas 1 and 2 apply also in a cycle.*

Proof. The Collatz algorithm is the same in any circumstances (in cycles as in linear branches). □

Lemma 5. *There is no root with one vertex. The only root with two vertices is $(-1,-2)$. The only root with three vertices is $(1,2,4)$. There is no root with four vertices. The only root with five vertices is $(-5,-14,-7,-20,-10)$. There is no root with six vertices.*

Proof. Let us verify all the 2^1 possibilities for one vertex $u_0 \to u_{-1} = u_0$ going from the initial integer upwards antecedents: we get $u_0 \to u_0 = \text{or}(2u_0,(u_0-1)/3)$. Hence $u_0 = \text{or}(0,-1/2)$ and therefore no solutions in $\mathbb{Z}^*$. The 2^2 possibilities for two vertices are $u_0 \to u_{-1} \to u_{-2} = u_0$ so that $u_0 \to \text{or}(2u_0,(u_0-1)/3) \to \text{or}(4u_0, 2(u_0-1)/3,\ (2u_0-1)/3,\ ((u_0-1)/3-1)/3)$. Hence $u_0 = \text{or}(0,-1,-2,-1/2)$. Only -1 and -2 are in our domain of definition $\mathbb{Z}^*$ and give effectively a cycle with two vertices. For three vertices, the 2^3 initial solutions are $(0,1,2,-4/7,4,-5/7,-8/7,-1/2)$ where only $(1,2,4)$ are in $\mathbb{Z}^*$ and is effectively a cycle with 3 vertices. For four vertices, the 2^4 initial solutions are $(0,1/5,2/5,\ -4/5,4/5,-1,\ -8/5,-13/25,8/5,-7/5,$ $-2,-14/25,-16/5,\ -17/25,-26/25,-1/2)$ where only $(-1,-2)$ are in $\mathbb{Z}^*$ and is effectively a cycle, but only with 2 vertices, a redundancy with the previous search. Similarly, one can resolve the five and six vertices' cases. □

Another cycle with 18 vertices is known in $\mathbb{Z}^-$ to this day. Of course, to solve it with the given previous method would be quite cumbersome and painful due to the $2^{18} = 262144$ equations to solve. The 4 known cycles are represented in figures 3 to 6.

Lemma 6. *There is no root containing a yellow vertex.*

Proof. The antecedent of an integer equal to $0 \pmod 3$ is unique and double its value (case 1 of figure 1). The next antecedent likewise and so up to infinity. Thus it cannot cycle back to its initial value. □

Lemma 7. *In a root, there cannot be 2 blue vertices next to each other. In a root, there can be possibly 2 green vertices next to each other, but not 3.*

Proof. This is an immediate result of lemma 4 and figure 1. □

Lemma 8. *In a root, a blue vertex has no link towards the outside of the root. In a root, an isolated green vertex has always a link towards the outside of the root. In a root, a pair of green vertex has one and only one of the vertex with a link towards the outside of the root.*

Proof. One gets again all the information from figure 1. For the blue vertex, which can only have two links (case 6), the two have to be inner links to get a cycle. For the isolated green vertex, the only contradictory case would be case 2, but then its value is equal to $1 \pmod 6$ and it is therefore linked to the green vertex on the top of case 3 which will have a blue link to the outside. □

Lemma 9. *Any root has active links towards the outside.*

Proof. For roots with less than six vertices, refer to figures 3 to 5 to confirm the claim. We know by figure 1 (cases 3, 4 and 5) that outside links from roots can only grow from green vertices. So then suppose that yellow vertices are growing out from two possible most narrow links. Figure 2 shows the two only possible cases of pieces of roots one can get on these premises. For each separate case in this figure, the vertices on the left side are those within the root itself and the vertices on the right side are the first items out of that root. White vertices may be green or blue, it doesn't matter. From figure 1, we know that the annotated green vertices must equal 10 (mod 18) as the only case with two antecedents of which one is yellow is case 4. We then consider the following two alternatives.

Case 1: Start from the first green vertex towards the top equal to $10+18k_1$. It is even thus the vertex underneath is equal to $5+9k_1$. The next vertex is then either equal to $(5+9k_1)/2$ or $16+27k_1$. This vertex in the same time must equal $10+18k_2$. Therefore either $3(k_1-4k_2)=5$ or $3(3k_1-2k_2)=-2$ which are both impossible with $k_1, k_2 \in \mathbb{Z}$.

Case 2: Start again from the first green vertex equal to $10+18k_1$. The vertex underneath is equal to $5+9k_1$. The next vertex is then either equal to $(5+9k_1)/2$ or $16+27k_1$. The next vertex is then either $(5+9k_1)/4$, $3(5+9k_1)/2+1$, $(16+27k_1)/2$ or $3(16+27k_1)+1$. This last vertex in the same time must equal $10+18k_2$. Therefore either $9(k_1-8k_2)=35$, $3(3k_1-4k_2)=1$, $9(3k_1-4k_2)=4$ or $3(9k_1-2k_2)=-13$ which are again all impossible. So we cannot have "adjacent" yellow vertices stemming from a root. We know also from lemma 8 that a supplementary intermediate blue-green-blue vertices' sequence without external link is impossible. Now considering a complete root, we know by lemma 7 that there are at least half of green vertices in any root. If these green vertices are systematically by pairs (which is certainly an absurd situation that we have not seek to object), there are anyway still at least 1/3 of the vertices in the root having links with a crown graph. Now from the above discussion, less then half of these links are yellow. Therefore, the number of active links is at least 1/6 of the root's cardinal. Thus with more than 6 vertices, a root has necessarily active links. □

For the four known cycles, as the reader can check directly, all the first links towards the outside of the root are active links. Shalom Eliahou [2] has proven that any unknown cycle in $\mathbb{N}^*$ would contain at least 17026679261 vertices (for the elements of the cycle only) and therefore would have, thanks to our own above study, more than 2837779877 active links.

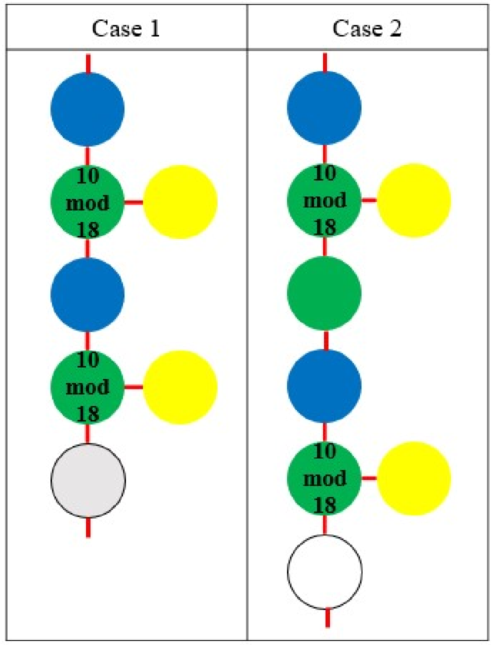

FIGURE 2. Pieces of roots.

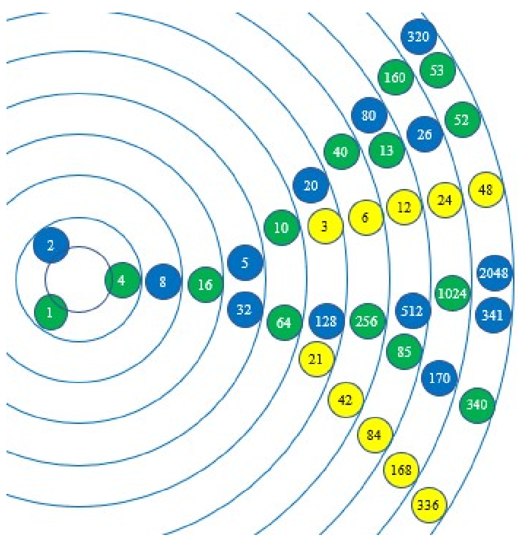

FIGURE 3. Cycle 1.

5. GRAPHS

Now that we know that any root has a crown graph, let us go further on this blossoming pattern.

Lemma 10. *A graph has one and only one root.*

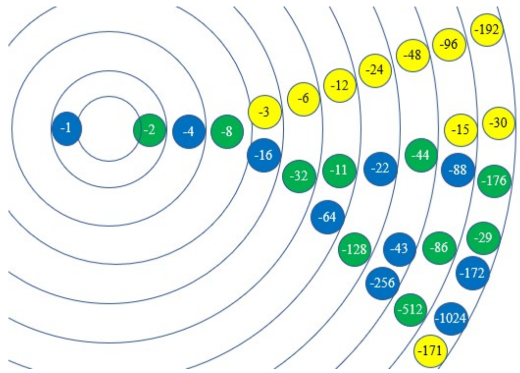

FIGURE 4. Cycle 2.

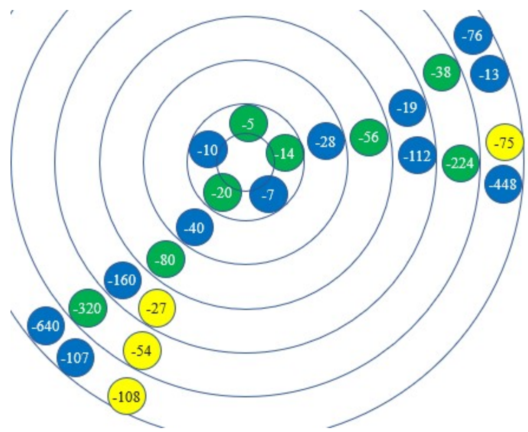

FIGURE 5. Cycle 3.

Proof. Let suppose the existence of a graph with two or more roots and let consider two of them. Applying the inverse algorithm, at some stage upwards among all the ramifications, there will be a common antecedent to two distinct vertices. As there is only one successor to a given integer, there is a contradiction to the way the Collatz algorithm works. Therefore only one unique root is the rule for any graph. □

Lemma 11. *Any integer belongs to a graph.*

Proof. This is lemma 3 □

Lemma 12. *There is no inactive graph.*

Proof. This is again lemma 3. According to figure 1, any active vertex has at least one active antecedent. Therefore:

Case 1: Starting from any active vertex, the cardinal of the successive antecedents is the same or increases, and this an infinite number of times. Hence, the graph is not inactive.

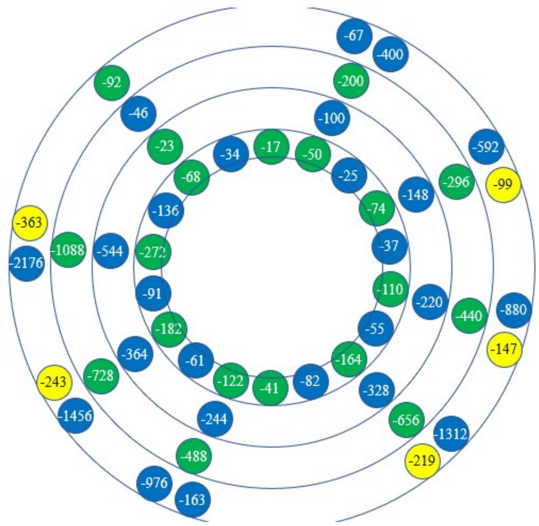

FIGURE 6. Cycle 4.

Case 2: Starting from any inactive vertex, its successors (in downwards Collatz algorithm) are divided by 2 as long as the result is even, otherwise the successor gets green and is hence an active vertex bringing us back to case 1. □

Theorem 1. *Any integer belongs to an active graph with a unique root.*

Proof. This is the aggregated result of lemmas 9, 10, 11 and 12. □

This proves our first point that there is no divergent series stemming from some given integer nor some series coming down from infinity and growing back up to infinity. There is always a root on the base of a Collatz tree structure and this root is unique.
Having established the first part of our proof, we have to concentrate also on the way the trees' structure grow.

Lemma 13. *The upwards asymptotic growth rate of the number of vertices between two consecutive ranks is equal to*

$$c = \frac{1 + \sqrt{1 + \frac{4k_2}{k_1 + k_2 + k_1.k_2}}}{2}$$

where k_1 is the ratio of the number of 0 mod 3 valued vertices to the number of 2 mod 3 vertices and k_2 is the ratio of the 0 mod 3 vertices to the number of 1 mod 3 vertices at a given stage of the upwards algorithm, for any Collatz graph in $\mathbb{Z}^*$, whatever the choice of the initial integer $\neq 0 \pmod 3$.

Proof. Any integer is linked upwards or downwards with an active branch according to theorem 1. But the choice of an integer $\neq 0 \pmod 3$ allows to apply exclusively the reverse Collatz algorithm to evaluate the growth rate, the said integer being right away part of an active branch that can bosom. This algorithm has only a finite number of alternatives (6 cases) according to figure 1. We insist here on the term "finite" which is crucial to conclude. Using the summary of the antecedents given in the said figure, let us suppose at rank r we have i_k proportion of vertices in class k. We deduce immediately the i_k' populations ratios for the antecedents underneath:

Cases	1	2	3	4	5	6
n	i_1	i_2	i_3	i_4	i_5	i_6
antecedent $2n$	i_1	i_2	i_3	i_4	i_5	i_6
antecedent $(n-1)/3$			i_3	i_4	i_5	

If asymptotically, the propositions between the modulo classes are stabilizing, it is necessarily according to the underneath relations, where c is the global multiplicative factor of the antecedents modulo classes populations using again figure 1:

$0 \pmod 3$	$c \cdot i_1 \simeq i_1 + i_4$
$1 \pmod 6$	$c \cdot i_2 \simeq i_3$
$4 \pmod 6$	$c \cdot (i_3 + i_4 + i_5) \simeq i_6$
$2 \pmod 3$	$c \cdot i_6 \simeq i_2 + i_3 + i_4 + 2i_5$

As the initial i_k are proportions, we write also $i_1+i_2+i_3+i_4+i_5+i_6 = 1$. In total, we have thus 5 equations for 7 unknowns (c and i_1 to i_6), leaving us with 2 degrees of freedom. These degrees of freedom are completed then with the two supplementary equations $i_1 \simeq k_1 * i_6 \simeq k_2 * (i_2 + i_3 + i_4 + i_5)$, where k_1 and k_2 are necessarily two positive factors. We will get the equation $c^2 - c - k_2/(k_1 + k_2 + k_1.k_2) = 0$. Hence the rate of growth $c = (1 + (1 + 4k_2/(k_1 + k_2 + k_1.k_2))^{1/2})/2$.

The dependency of c with the parameters k_1 and k_2 is given in figure 7. Except for $k_2 = 0$, which is asymptotically absurd, the ratio c is always strictly greater than 1 (with maximum value the golden ratio $\varphi = (1 + \sqrt{5})/2$). As the reverse algorithm is repeated infinitely, whatever the exact asymptotic value of c, or even if there is no asymptotic convergence to some constant value, so long as $c > 1$, which is the case, the Collatz graph will grow exponentially. □

Lemma 14. *The asymptotic proportion between the numbers of* 0, 1 *and* 2 *mod* 3 *vertices is identical in any graph.*

Proof. This means $k_1 = k_2 = 1$. It is obvious that the 0 and 1 (mod 2) populations cannot be equal as there is systematically an even antecedent to any integer by the reverse Collatz algorithm. On the opposite, if x is random (unbiased), the integers' transformations $x \longrightarrow 2x$ as well as $x \longrightarrow (x-1)/3$, the later when it has an integer solution, are unbiased modulo any odd number v. Insisting specifically on the modulo 3 case, let us consider $3k_1$,

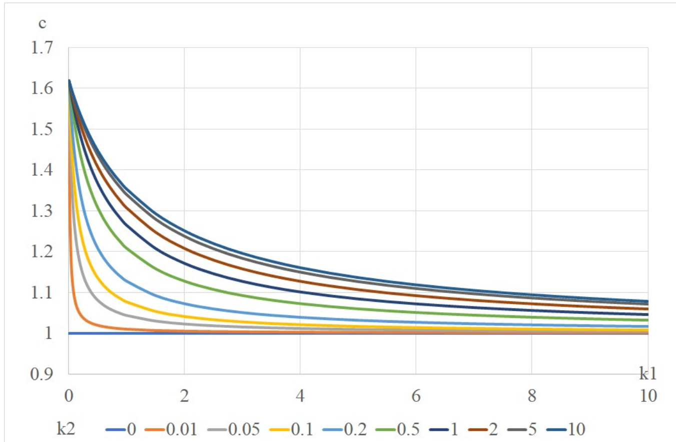

FIGURE 7. Evolution of c versus k_1 and k_2.

$3k_2 + 1$, $3k_3 + 2$ with equal populations, the k_i having random (mod 3) values. The antecedents are $6k_1$, $6k_2 + 2$, k_2, $6k_3 + 4$. Therefore as k_2 is initially unbiased (and $6k_1 \equiv 0 \bmod 3$, $6k_2 + 2 \equiv 2 \bmod 3$, $6k_3 + 4 \equiv 1 \bmod 3$), the new set of antecedents is unbiased modulo 3. If the initial sets k_i do not have the same populations (which is necessary the case starting from one integer or one root of integers), the k_2 values sampling, induced by the reverse algorithm, is collecting numbers from $\mathbb{Z}$ where some values have already being drawn out, therefore giving an advantage to the other modulo 3 samples, creating a progressive feedback compensation. The 0, 1 and 2 (mod 3) populations will therefore tend towards equal proportions by random draw in a growing and huge sample of $\mathbb{Z}$ and likewise for any odd integer v instead of 3 as further examined in the appendix A. It shows that the sampling in similar populations occurs rapidly (less than 50 rounds in the example). Subsequently, we get two supplementary equations closing down the said degrees of freedom: $i_1 \simeq i_6 \simeq i_2 + i_3 + i_4 + i_5$. $\square$

Theorem 2. *The asymptotic rate of growth of any Collatz graph is*

$$c = \frac{1 + \sqrt{7/3}}{2} \approx 1.26376262.$$

Proof. This is an immediate result of lemma 13 using $k_1 \to 1$ and $k_2 \to 1$ when the rank increases. $\square$

Note. Going back to figure 1, we are then able to make a summary of the modulo 6 's populations:

$0 \pmod 6$	$c-1$
$1 \pmod 6$	$\frac{4}{3}-c$
$2 \pmod 6$	$c-1$
$3 \pmod 6$	$\frac{4}{3}-c$
$4 \pmod 6$	$c-1$
$5 \pmod 6$	$\frac{4}{3}-c$

and therefore we get for the modulo 2 's populations:

$0 \pmod 2$	$3 \cdot (c-1)$
$1 \pmod 2$	$4 - 3 \cdot c$

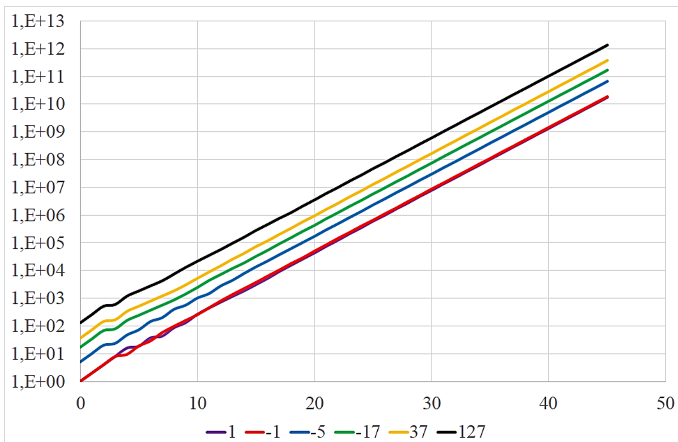

FIGURE 8. Approximate quantities of vertices versus rank r for several initial integers.

Note. The PARI/GP computer programming in appendix A provides the cardinals of the congruence classes of the integers at rank 48 of the reverse Collatz algorithm taking -67 as a sample for the initial integer. The data shows already less that 2 % deviation from the asymptotic equiprobable proportions at this early stage of calculation for odd v modulo operations, $v = 3$ to $v = 15$.

We provide also the hereby figure 8. The figure, thanks to the logarithmic ordinates, shows the same rate of growth for more initial samples.

With the said software, starting from some arbitrary initial integer $\neq 0 \bmod 3$, the reader can easily check the overall growth rate and the proportions in each congruence class using the resulting numbers of vertices.

The reader may also refer to [9] Sheet Terras Appendix 4 for the numerical values of growth rate and congruence class proportions. The data is given for the four known graphs in $\mathbb{Z}^*$ and the two additional partial graphs stemming from 37 and 127 show that the results are analogous for independent tree structures as for distinct parts of the same graph.

So far we checked the rate of growth of the Collatz graphs. But the equality of their asymptotic value is not sufficient to make sure that the density of one graph to another is not negligible allowing such graph to coexists on N*. We will have to show that the rates of growth evolve in a close enough way one to each other, that the domains in which the graphs develop are similar and that the repartition within these domains are also similar. The first objective will be achieved in lemma 16, the second in lemma 18 and the last one in lemma 17.

So let us start and consider two distinct graphs and the corresponding infinite products of positive real number series $\prod c1_i$ and $\prod c2_i$ where $c1_i \longrightarrow c$ and $c2_i \longrightarrow c$ when i tends towards infinity. The condition $\frac{c1_i}{c2_i} \longrightarrow 1$ is necessary for the infinite product $\prod \frac{c1_i}{c2_i}$ to converge (to a strictly positive value), but the reciprocal is false. Hence the need here for additional decisive arguments to meet our aim.

Lemma 15. *The absolute value of the ratio of the rates of growth $c1_i$ and $c2_i$, at rank i of the reverse Collatz algorithm, of two distinct Collatz graphs is asymptotically lesser than $1 + \frac{1}{i^n}$ where n is some constant > 1.*

Proof. The rate of growth c_i at rank i of a Collatz graph converges towards c in a random way. It takes asymptotically random inferior and superior values compared to c. The difference to 1 of the ratio $\frac{c1_i}{c2_i}$ cumulate the difference to c of $c1_i$ and $c2_i$. This induces a mere factor 2 in the uncertainty measurement. The population at rank i of a Collatz graph is similar to $\alpha \cdot c^i$ where α is some approximative initial value. But in random phenomena, average absolute differences compared to the populations fades away with the increase of populations. Exponential growth of populations means inverse exponential "growth" of the average absolute differences (see the binomial distribution's example in the note underneath) to the populations. The ratio of the latter, let us say $\lambda \cdot c^{-\epsilon \cdot i}$ with λ and ϵ strictly positive values, to the polynomial expression $\frac{1}{i^n}$, where n is constant, therefore tends towards 0 ($\lambda \cdot c^{-\epsilon \cdot i} \cdot i^n \longrightarrow 0$ when $i \longrightarrow +\infty$). □

Lemma 16. *The infinite product of the ratio of the rates of growth $c1_i$ and $c2_i$ of two distinct Collatz graphs is convergent (to a non-zero value) where i is the rank of the reverse Collatz algorithm.*

Proof. The asymptotic absolute ratio is lesser than $1 + \frac{1}{i}$, the multiplicative counterpart of the harmonic series. Let us say, it is smaller than $1 + \frac{1}{i^{1+\varepsilon}}$, $\varepsilon > 0$ after some rank k. Up to rank k the ratio gives some finite strictly positive value, and the remainder infinite product gives another finite strictly positive value which together give a finite strictly positive value. Hence the convergence. □

Note. Let us suppose the most likely standard binomial distribution of the vertices' values in each congruence class. For this kind of distribution, asymptotically, the mean absolute difference is $EM(|X - n|, p = 1/2) =$

$n\binom{2n}{n}/2^{2n} \simeq \sqrt{\frac{n}{\pi}}$ (see reference [6]) where n is the class' population. Here the class' population is $\alpha_j(c^i)$ where i is the rank upwards and α_j some constant value depending of the initial integer. The deviation at each rank from 1 is equal to $\frac{1}{n}\sqrt{\frac{n}{\pi}}$, hence the ratio is $1+2((\alpha_1/\alpha_2)\pi)^{-1/2}c^{-i/2}$ where the multiplicative factor 2 cumulates the uncertainties due to the numerator and the denominator of $|\frac{c1_i}{c2_i}|$ as previously mentioned.

For example, modulo 3, in appendix A, the average population in a congruence class is $n = 78932/3 \simeq 26311$, and implementing the comparison on the same graph (thus $\alpha_1/\alpha_2 = 1$), we get $2\frac{1}{n}\sqrt{\frac{n}{\pi}} \simeq 0.70\%$ which therefore corresponds to an order of magnitude of the mean absolute difference to population ratio of a binomial distribution. The other comparative values are given in the appendix A for the modulo v cases, $v = 5, 7, \ldots, 15$ with similar conclusion.

Similar conclusion would also be drawn with a non-binomial distribution as it would only scale the result by some finite factor for the ratio $\prod \frac{c1_i}{c2_i}$.

Now, let us turn to the distributions of values. One way to do it is to check that the average sum of the values of vertices at some common rank r are similar. Some constrain may however make it only "locally". A much decisive argument is that the result remains true for any exponential power affected to the vertices values.

Lemma 17. *The asymptotic medium growth of antecedents' values (absolute values of integers), to which a power n is applied, for the reverse Collatz algorithm, between two consecutive ranks, is equal to:*

$$t(n) = \frac{2^n + (c-1)\cdot(\frac{1}{3})^n}{c}$$

Proof. The obvious argument is that the same algorithm applies to the same types of entities (elements of $\mathbb{Z}^*$). Hence a unique result for given n. But more specifically, as usual, let us start with figure 1 and make the summary of the modulo distribution at stake (the k indices in i_k underneath are respectively 3, 4 and 5).

Congruencies	Initial proportions	Initial values	Final values		Dilution
0 (mod 3)	$\frac{1}{3}$	$(i_1)^n$	$(2i_1)^n$		$1/c$
1 (mod 6)	$\frac{4}{3} - c$	$(i_2)^n$	$(2i_2)^n$		$1/c$
4 (mod 6)	$c-1$	$(i_k)^n$	$(2i_k)^n$	$(\frac{1}{3}i_k)^n$	$1/c$
2 (mod 3)	$\frac{1}{3}$	$(i_6)^n$	$(2i_6)^n$		$1/c$

The dilution factor is the result of the factor c increase of populations between two successive ranks. Adding the terms, we get immediately the average value of the above proposition. □

In figure 9, the power value is equal to $n = 1$ and therefore $t(n)$ converges to the expected approximate value $t(1) = (2 + (c-1)/3)/c \approx 1,652$. It is indeed also the case for the $n = 1/4, n = 1/2$ and $n = 2$ powers that we studied without reproducing the data here (see our personal internet site for the data).

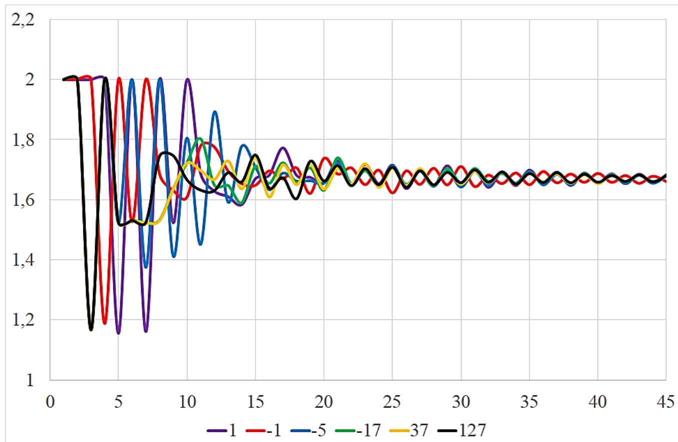

FIGURE 9. t(1) versus rank r for several initial integers.

Lemma 18. *Collatz graphs evolve in similar domains of definition.*

Proof. The further outer element (maximum absolute value vertex) at rank r is equal to $m.2^r$, m being the choice of the initial positive or negative integer ($\neq 0 \pmod 3$). Hence here the term similar domains of definition ($[0, m.2^r]$ if $m > 0$ and $[m.2^r, 0]$ if $m < 0$) is meant as the finite multiplicative ratio m of the domain for the graph stemming from m compared to the graph domain $[0, 2^r]$ involved in the standard graph stemming from 1. $\square$

Note. The global picture is that if one makes the assumption that some graph stemming from a given root would be able, instead of cycling at that root, to prolong further down to reach 1, then the data in figure 8 would fairly overlap with the data given for the standard graph from 1. Indeed, because the graph stemming from -1 has a final value which in absolute value is precisely equal to 1, the two data in the said figure do unsurprisingly asymptotically match. This being not the case in general, the curves (in y-logarithmic coordinates) are simply parallel (due to the m multiplicative factor). In these coordinates, one can see the very rapid diminution (around 20 ranks upwards is sufficient) of margin errors to straight lines. In other words any graph stemming at m ($\neq 0 \pmod 3$) would asymptotically overlap (in absolute values) with the standard graph stemming from root 1. For this, we would have just to adjust the initial positions of the two graphs, the former at rank r and the later at m such that $m \approx 2^r$.

Theorem 3. *The natural density of integers in the natural numbers set* $\mathbb{N}^*$ *with finite stopping time, that is using the notation of the introduction the integers* u_i *such that* $u_{i+r} < u_i$ *with finite* r, *is equal to* 1.

Proof. This is the theorem of Riho Terras [1]. □

Note. At this stage we have proven that there is no separate infinite divergent series (see page 9). Besides, we established that the Collatz trees' structures grow at a same rate which makes the existence of two such structures in $\mathbb{N}$ very unlikely as no trace of it exists. The Riho Terras' theorem establishes an important result but it may encompass in its realm the contribution of several trees, failing to be an ultimate argument. However the study of the numbers with finite stopping time is definitively useful as we will see.

Part 2. The parity vectors' structure

6. Cycles' classification

Lemma 19. *The composition of linear functions is a linear function.*

Proof. The lemma is obvious, but let us develop the precise result. Let us have $T_k(x) = a_k x + b_k$, $k = 1$ to i, a series of linear functions and let us consider $CT_i(x) = T_i \circ T_{i-1} \circ ... \circ T_1(x)$. Then $CT_1(x) = a_1 x + b_1$, $CT_2(x) = a_2 a_1 x + a_2 b_1 + b_2$, ... and $CT_i(x) = a_i a_{i-1} ... a_1 x + a_i a_{i-1} ... a_2 b_1 + a_i a_{i-1} ... a_3 b_2 + ... + a_i a_{i-1} ... a_{i-3} b_{i-4} + a_i a_{i-1} a_{i-2} b_{i-3} + a_i a_{i-1} b_{i-2} + a_i b_{i-1} + b_i$. □

Lemma 20. *The composition of j linear functions with $a_k = 1/2$ and $b_k = 0$ and $i - j$ linear functions with $a_k = 3/2$ and $b_k = 1/2$, in that specific order, is equivalent to the linear function :*

$$LT_i(x) = \frac{3^{i-j}}{2^i}(x + 1 - (\frac{2}{3})^{i-j}) \tag{1}$$

Proof. We have $T_k(x) = (3x + 1)/2$, $k = 1$ to $i - j$ and $T_k(x) = x/2$, $k = i - j + 1$ to i. Using the previous lemma, we get $LT_i(x) = (3^{i-j} x + 3^{i-j-1} 2^0 + 3^{i-j-2} 2^1 + 3^{i-j-3} 2^2 + ... + 3^3 2^{i-j-4} + 3^2 2^{i-j-3} + 3^1 2^{i-j-2} + 3^0 2^{i-j-1})/2^i = 3^{i-j} 2^{-i}(x + 3^{-1}(1 + (2/3)^1 + (2/3)^2 + ...(2/3)^{i-j-1}))$. Thus the former result. □

Lemma 21. *The function $HT_i(x) = LT_i(x)/x$ is an hyperbolic function, therefore strictly monotonous, defined everywhere except for $x = 0$. Its value is equal to 1 for the unique solution :*

$$x = -\frac{1 - (\frac{2}{3})^{i-j}}{1 - 2^j (\frac{2}{3})^{i-j}} \tag{2}$$

Posing $v = i - j$ the number of $3x + 1$ multiplications and $w = i$ the number of divisions by 2, we get also :

$$x = -\frac{1 - (\frac{2}{3})^v}{1 - 2^{w-v} (\frac{2}{3})^v} \tag{3}$$

Proof. We have $T_k(x) = 3^{i-j} 2^{-i}(1 + (1 - (2/3)^{i-j})(1/x))$ which is obviously a hyperbolic function. Its derivative is equal to $-3^{i-j} 2^{-i}(1 - (2/3)^{i-j})(1/x^2)$, therefore of the sign of the constant expression $-(1 - (2/3)^{i-j})$. Solving $T_k(x) = 1$ gives immediately the result x given in the lemma. □

Lemma 22. *Let us pose*

$$w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor + 1 - incr \tag{4}$$

using the floor function and incr being an integer. Then

$$\begin{aligned} &\text{if } incr > 1, \quad -2 < x < 0, \\ &\text{if } incr < 0, \quad 0 < x < 1. \end{aligned}$$

Proof. Let us go back to equation 3. We get immediately $x = -(1-(2/3)^v)/(1-2^w/3^v)$. For small values de v, we verify the proposition numerically and figure 10 illustrates the point. If $v >> 1$, as v diverges, the numerator $1-(2/3)^v$ will tend towards 1^-. Then $0 > x > -1/(1-1/2) = -2$ if $2^w/3^v < 1/2$. Solving $2^w/3^v < 1/2$, we get $w < (ln(3)/Ln(2))v - 1$. Then replacing w with the expression of the lemma, we get $\lfloor (ln(3)/Ln(2))v \rfloor + 1 - incr < (ln(3)/Ln(2))v - 1$, therefore $1 \leq \lfloor (ln(3)/Ln(2))v \rfloor - (ln(3)/Ln(2))v + 2 < incr$ which is the announced lower limit value of incr. Studying the second condition, we observe that $0 < x < -1/(1-2) = 1$ if $2^w/3^v > 2$. Solving $2^w/3^v > 2$, we get $w > (ln(3)/ln(2))v+1$. Then replacing w with the expression of the lemma, we get $\lfloor (ln(3)/Ln(2))v \rfloor + 1 - incr > (ln(3)/Ln(2))v + 1$, therefore $0 \geq \lfloor (ln(3)/Ln(2))v \rfloor - (ln(3)/Ln(2))v > incr$ which is this time the announced highest limit value of incr. $\square$

Note. The value of x tends towards -1 when *incr* increases asymptotically (incr $\to +\infty$). The value of x tends towards 0 when *incr* decreases asymptotically (incr $\to -\infty$).

Note. The figures 11 and 12 illustrate the two cases incr = 0 and incr = 1. The ordinates are in these two cases in logarithmic scales (ln(x) for incr = 0 and ln(-x) for incr = 1).

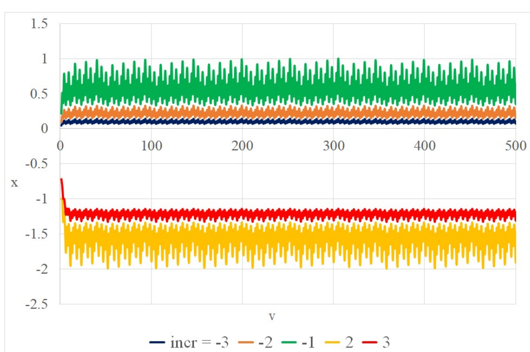

FIGURE 10.
Solutions to $HT_w(x) = 1$,
incr $= -3, -2, -1, 2$ and 3.

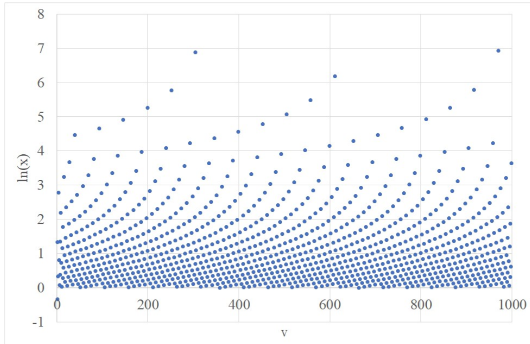

FIGURE 11. Solutions to $HT_w(x) = 1$, $incr = 0$.

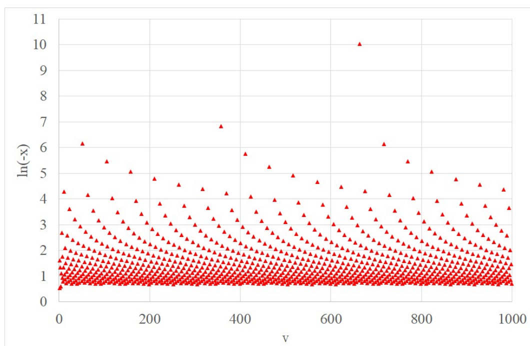

FIGURE 12. Solutions to $HT_w(x) = 1$, $incr = 1$.

Lemma 23. *The previous solution x is the largest in absolute value to the equation $PT_k(y)/y = 1$, where $PT_k(y)$ is any permutation of the composition $T_i \circ T_{i-1} \circ ... \circ T_1(x)$ keeping here the same number of $(3x+1)/2$ multiplications and $x/2$ divisions.*

Proof. Let us have $a = 3^{i-j-1}2^0 + 3^{i-j-2}2^1 + 3^{i-j-3}2^2 + ... + 3^3 2^{i-j-4} + 3^2 2^{i-j-3} + 3^1 2^{i-j-2} + 3^0 2^{i-j-1}$, thus $HT_i(x) = (3^{i-j}x + a)/2^i x = 1$ has solution $x = a/(2^i - 3^{i-j})$. Here the denominator has a fixed value and therefore the absolute value of x is maximal if the absolute value of a diminish when the permutation is applied (giving a smaller alternative value y). In order to get the final composition of the linear functions, we apply a finite number of elementary permutations such that each one switches two members $3^{n_1}2^{m_1}$ and $3^{n_2}2^{m_2}$ to $3^{n_1-1}2^{m_1+1}$ and $3^{n_2+1}2^{m_2-1}$, where $n_1 > n_2$ and $m_1 < m_2$, systematically reducing the value of the initial a (because $3 > 2$). Thus the result. $\square$

Theorem 4. *The Collatz algorithm may lead to a cycle in $\mathbb{N}^*$ if and only if the number of (3x+1) multiplications, noted v, to the number of (x/2)* *divisions, noted w, meets the condition*

$$Type1 : w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor + 1 \tag{5}$$

and may generate a cycle in $\mathbb{Z} - \mathbb{N}$ if and only if it meets the condition

$$Type0 : w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor. \tag{6}$$

Proof. According to the lemmas 22 and 23, which hold in $\Re$ and therefore also in $\mathbb{Z}$, the largest absolute value y to equation $PT_w(y) = 1$ is smaller then 2 (for any value of incr different from 0 or 1), a finite interval that one can easy totally explore numerically for solutions and check that there are none. One can then conclude on the value of w in $\mathbb{Z}^*$ using again lemma 22 which distinguish the two cases on *incr*. □

Note. Checking the known cycles (except 0 which meets the type 1), using $ln(3)/ln(2) \approx 1.58496$, we get the following numerical results
Type 1 : $y = 1$, $v = 1$, $w = 2$ and $2 = \lfloor (ln(3)/ln(2)).1 \rfloor + 1$,
Type 0 : $y = -1$, $v = 1$, $w = 1$ and $1 = \lfloor (ln(3)/ln(2)).1 \rfloor$,
Type 0 : $y = -5$, $v = 2$, $w = 3$ and $3 = \lfloor (ln(3)/ln(2)).2 \rfloor$,
Type 0 : $y = -17$, $v = 7$, $w = 11$ and $11 = \lfloor (ln(3)/ln(2)).7 \rfloor$.
The reader may refer to the figures 3 to 6 to check that the cycles meets the number of (3x+1) multiplications and (x/2) divisions.

At that stage, we know, v being the number of (3x+1) mutiplications and w the number of (x/2) divisions, that for cycles in $\mathbb{Z} - \mathbb{N}$ we will have systematically $w = \lfloor (ln(3)/ln(2)).v \rfloor$ and for cycles in $\mathbb{N}^*$ we will have $w = \lfloor (ln(3)/ln(2)).v \rfloor + 1$. We will study now the stopping time of integers over the whole domain $\mathbb{Z}$ before resuming arguments on the Collatz conjecture.

7. 2^w−periodicity in Z

Lemma 24. *Let us consider x_0 any positive integer. Applying v odd steps and w even steps in some order of the Collatz algorithm to x_0, the nearest result y_0 to the initial integer x_0 among all combination of the said odd and even steps, is equal to*

$$y_0 = \frac{3^v}{2^w} x_0 + \frac{1}{2^{w-v}} ((\frac{3}{2})^v - 1) \tag{7}$$

Proof. Applying first all the (3x+1)/2 multiplications, we get $y_0 = (3^v x_0 + 3^{v-1}2^0 + 3^{v-2}2^1 + 3^{v-3}2^2 + \ldots + 3^1 2^{v-2} + 3^0 2^{v-1})/2^v/2^{w-v} = (3^v/2^w)x_0 + (1/2^{w-v})((3/2)^v - 1)$. Applying first all the (x/2) divisions, we get $y_0 = (3^v x_0 + 3^{v-1}2^{w-v} + 3^{v-2}2^{w-v+1} + 3^{v-3}2^{w-v+2} + \ldots + 3^1 2^{w-2} + 3^0 2^{w-1})/2^{w-v}/2^v = (3^v/2^w)x_0 + ((3/2)^v - 1)$. The other combinations give intermediary values between these two results and the first expression is the nearest result y_0 to x_0 because of the additional ratio $(1/2^{w-v})$ smaller then 1 in front of $((3/2)^v - 1)$. □

Lemma 25. *The ratio x_w/x_0, where $x_0 > 0$ and x_w is resulting from a Collatz algorithm is systematically such that*

$$\frac{x_w}{x_0} > \frac{3^v}{2^w} \tag{8}$$

Proof. The nearest y_0 to x_0 means that $x_w/x_0 \geq y_0/x_0$. Thus using the previous lemma result

$$\frac{x_w}{x_0} \geq \frac{y_0}{x_0} = \frac{3^v}{2^w} + \frac{1}{x_0}\frac{1}{2^{w-v}}((\frac{3}{2})^v - 1) > \frac{3^v}{2^w}.$$

□

Lemma 26. *The ratio $ln(3)/ln(2)$ is irrational.*

Proof. Let us suppose $ln(2)/ln(3) = p/q$, where p and q are integers. Then $q.ln(2) = p.ln(3)$, so that $ln(2^q) = ln(3^p)$ and finally $2^q = 3^p$, which is obviously false. Thus $ln(3)/ln(2) \notin Q$. □

Lemma 27. *Let us have x_0, x_w, v and w some fixed strictly positive values meeting the condition of lemma 25, that is $3^v/2^w < x_w/x_0$. The function $f(k) = (x_w + k.3^v)/(x_0 + k.2^w)$ is continuous over positive or null k, decreasing monotonously from x_w/x_0 towards $3^v/2^w$, the later an asymptotic value.*

Proof. The function $f(k)$ is a hyperbolic function which is undefined at the unique strictly negative value $k = -x_0/2^w$, therefore is continuous on R^+. The derivative is $f'(k) = (x_0.3^v - x_w.2^w)/(x_0 + k.2^w)^2$, therefore of the sign of $x_0.3^v - x_w.2^w$ which is strictly negative by the chosen hypothesis. The function therefore evolves over R^+ monotonously from x_w/x_0 at $k = 0$ to the limit value $3^v/2^w$ when $k \to \infty$. □

Lemma 28. *Let us have some positive integer x_0 and x_w its result by the Collatz algorithm at its stopping time. If the stopping time is finite then the ratio x_w/x_0 is such that $1/2 < x_w/x_0 \leq 1$.*

Proof. By definition of the stopping time, we have $x_w/x_0 \leq 1$. The last step of the algorithm is necessarily an even step which either gives exactly the value of x_0 or a strictly greater value of its half. Using $(3x + 1)/2$ multiplications and $(x/2)$ divisions, each step of the process includes a division by 2, therefore w is the appropriate index to count them and v will be the number of multiplications. □

Lemma 29. *Let us consider the set $\{x_0+k.2^w, k \in Z\}$. Then, if x_w exist for a finite w, the elements of the set $\{x_w+k.3^v, k \in Z\}$ are the resulting values of the initial set at their respective stopping time and moreover $x_0 < 2^w$ (recall also that $0 < x_0$ by hypothesis) and $1/2 < (x_w + k.3^v)/(x_0 + k.2^w) \leq 1$.*

Proof. This is mostly a well-known result but is worth reviewing. By hypothesis, it is clear that the elements resulting from x_0 being divisible w times by 2 then $x_0 + k.2^w$ is also divisible in the same condition w times. Moreover at each step the distance 2^t between the intermediary results $x_i + k.2^t$ is constant and equal to $2^{w-s}3^m$ where s is the number of steps at that stage and m the number of $(x+1)/2$ multiplications, hence a distance $2^0 3^v$ at the stopping time. Now for $k = 0$, the number of steps v and w is necessarily such that $3^v/2^w \leq 1$ (in fact $3^v/2^w < 1$ by lemma 26) but in the closest way as a division by 2 is always the last step of the stopping time process and therefore $3^v/2^w > 1/2$. Now according to lemma 27, $x_w/x_0 > 3^v/2^w$ so that $x_0.3^v - x_0.x_w < x_w.2^w - x_0.x_w$ and so $x_0(3^v - x_w) < x_w(2^w - x_0)$ is equivalent to $(3^v - x_w)/(2^w - x_0) < x_w/x_0$ because $x_0 < 2^w$ providing the first sample of the $(x_w + k.3^v)/(x_0 + k.2^w)$ were k is negative. Here $k = -1 < -x_0/2^w$ which is the undefined abscissa of the hyperbolic function. Therefore, as we know that the function is strictly decreasing, the ratio is increasing from $(3^v - x_w)/(2^w - x_0)$ up asymptotically towards $3^v/2^w$ as $k \to -\infty$. □

Theorem 5. *At its stopping time, for a finite non-cyclic event, the number of (3x+1) multiplications, noted v, to the (x/2) divisions, noted w, is such that*

$$w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor + 1 \tag{9}$$

over the whole domain $\mathbb{Z}$.

Proof. According to lemma 29, $1/2 < (x_w + k.3^v)/(x_0 + k.2^w) \leq 1$ and the elements of set $x_0 + k.2^w, k \in \mathbb{Z}$ have all the same number of odd and even steps at the stopping time. We get also, for $k = 0$, $1/2 < 3^v/2^w \leq 1$ which is equivalent to $w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor + 1$. □

Lemma 30.

$$\frac{1}{2} < \frac{3^v}{2^w} < 1 \Leftrightarrow w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor + 1 \tag{10}$$

Proof. The only point to complete from the previous proof is that $\frac{2^w}{3^v} \neq 1$ which is obvious. □

Lemma 31. *Let us consider x_0 an integer such that $(3^v - 2^v)/(2^w - 3^v) < x_0 < 2^w$, where $w = \lfloor (ln(3)/ln(2)).v \rfloor + 1$, then applying the Collatz algorithm up to the stopping time w, the result being x_w, is such that $0 < x_0 - x_w < 2^w - 3^v$.*

Proof. The condition $w = \lfloor (ln(3)/ln(2)).v \rfloor + 1$ applies $0 < 2^w - 3^v$. Then x_w taking the place of y_0 in the lemma 24, we get $x_0 - x_w = (1 - (3^v/2^w))x_0 - (1/2^{w-v})((3/2)^v - 1)$. The proposition $0 < x_0 - x_w$ is equivalent to $0 < (1 - (3^v/2^w))x_0 - (1/2^{w-v})((3/2)^v - 1)$ that is $(1/2^{w-v})((3/2)^v - 1)/(1 - (3^v/2^w)) < x_0$ or $(3^v - 2^v)/(2^w - 3^v) < x_0$. The proposition $x_0 - x_w < 2^w - 3^v$ is equivalent to $(1 - (3^v/2^w))x_0 - (1/2^{w-v})((3/2)^v - 1) < 2^w - 3^v$ that is $((2^w - 3^v)/2^w)x_0 < (3^v/2^w - 2^v/2^w) + 2^w - 3^v$ or $x_0 < (3^v - 2^v)/(2^w - 3^v) + 2^w$ which is less constraint than the proposition of the lemma. □

Note. The limite $(3^v - 2^v)/(2^w - 3^v) < x_0$ can be replaced by $0 < x_0$ by direct numeric verification. The figure 13 shows that the ratio $r = \frac{3^v - 2^v}{2^w - 3^v} \frac{1}{x_0}$ tends towards 0 at an exponential pace.

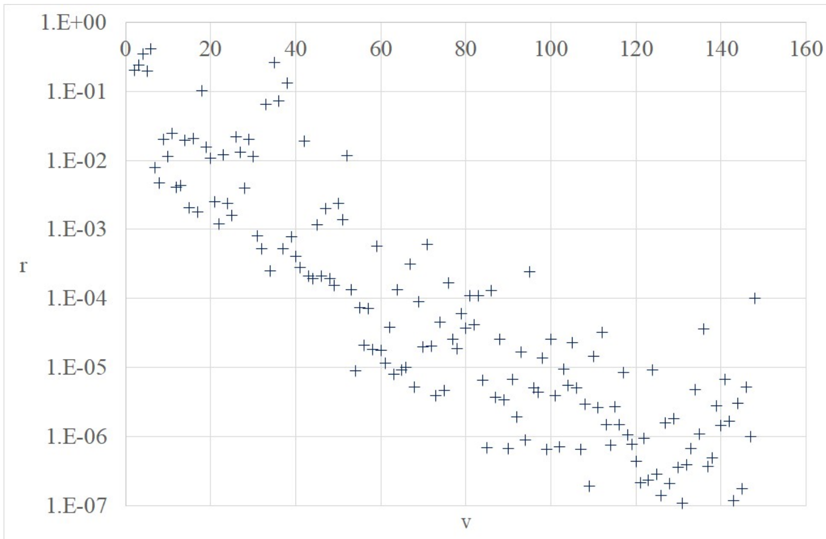

FIGURE 13. $r = \frac{3^v - 2^v}{2^w - 3^v} \frac{1}{x_0}$.

Example. $x = 7, v = 4, w = 7, 3^v = 81, 2^w = 128, w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor + 1$.

-249	-121	7	135	263
-373	-181	11	203	395
-559	-271	17	305	593
-838	-406	26	458	890
-419	-203	13	229	445
-628	-304	20	344	668
-314	-152	10	172	334
-157	-76	5	86	167

Counterexample. $x = -5, v = 2, w = 3, 3^v = 9, 2^w = 8, w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor$.

-21	-13	-5	3	11
-31	-19	-7	5	17
-46	-28	-10	8	26
-23	-14	-5	4	13

Counterexample. $x = -17, v = 7, w = 11, 3^v = 2187, 2^w = 2048, w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor$.

-4113	-2065	-17	2031	4079
-6169	-3097	-25	3047	6119
-9253	-4645	-37	4571	9179
-13879	-6967	-55	6857	13769
-20818	-10450	-82	10286	20654
-10409	-5225	-41	5143	10327
-15613	-7837	-61	7715	15491
-23419	-11755	-91	11573	23237
-35128	-17632	-136	17360	34856
-17564	-8816	-68	8680	17428
-8782	-4408	-34	4340	8714
-4391	-2204	-17	2170	4357

8. Numbering scheme

Having established the type of most of the integers, let us study the underlying structure derived from the stopping time w which depends on v by the relationship $w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor + 1$. Starting from the set $\mathbb{Z}$, let us remove all the elements such that $\{v = 0, w = 1\}$ that are in the interval $[0, 2^w - 1 = 1[$. Only 0 complies and the other elements satisfying $\{v = 0, w = 1\}$ are separated by a distance $2^w = 2$, therefore the even integers. Then we discard the elements such that $\{v = 1, w = 2\}$ in the interval $[0, 2^w - 1 = 3[$, where only 1 meets the requirement and those separated by a distance $2^w = 4$ from the formers, hence all $1 \mod 4$ integers. Going to step $\{v = 2, w = 4\}$, we consider the integers in interval $[0, 2^w - 1 = 15[$, where only 3 meets the requirement and the complement separated by a distance $2^w = 16$, hence all

3 mod 16 integers. At next step $\{v = 3, w = 5\}$, we consider the integers in interval $[0, 2^w - 1 = 31[$, where only 11 and 23 meets the requirement and all those separated by a distance $2^w = 32$, hence all 11 mod 32 and 23 mod 32 integers. This removal process is illustrate in the table underneath by lowering the initial integers to the corresponding $v-$indexed line. Of course the integers -1, -5 and -17 in red in the second line can never be affected as those comply with $w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor$ instead of $w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor + 1$.

Table 1

v	-19	-17	-15	-13	-11	-9	-7	-5	-3	-1	1	3	5	7	9	11
		-17						-5		-1						
1	-19		-15		-11		-7		-3		1		5		9	
2				-13								3				
3						-9										11

Definition. *We will call parity vector the binary representation* 1 *or* 0 *of the sequence of odd (as (3x+1)/2 multiplications) and even steps of the Collatz algorithm applied to some initial integer until the stopping time in the order of apparition. The parity vector is of size w.*

It includes v digits 1 and $w - v$ digits 0.

Definition. *We call a licit parity vector that one that doesn't break any rule of the Collatz algorithm during an altitude flight routine.*

Let us consider, for example, the parity vector 100. It is not licit because the number of even steps written here (that is 2) is greater than the correct value $w - v = \lfloor (ln(3)/ln(2).v) \rfloor + 1 - v$ here (that is $1+1-1 = 1$). The altitude flight time is exceeded in this writing. Similarly, writing 10111100 is not licit, even if we do have globally $w - v = \lfloor (ln(3)/ln(2).v) \rfloor + 1 - v = 7+1-5 = 3$, as the altitude flight time is met prematurely by writing 10 at the beginning of the sequence. For some given parity vector, in the same way, one has to check its validity at each new even intermediate step. For example, with parity vector 111111101000100, the three intermediate necessary checks for a premature non licit parity vector are the following ones between corresponding parentheses (((11111110)1000)100). We have $8-7 = 1 < \lfloor (ln(3)/ln(2).v1 \rfloor + 1 - v1 = 12+1-7 = 6$, $12-8 = 4 < \lfloor (ln(3)/ln(2).v2 \rfloor + 1 - v2 = 13+1-8 = 5$ and $15-9 = 6 = \lfloor (ln(3)/ln(2).v3 \rfloor + 1 - v3 = 14+1-9$ thus corresponding effectively to a licit parity vector.

This summarizes as follows.

Lemma 32. *The rule linking w to v being respected, there are two limit cases for the licit parity vectors. The first one is where the components* 1 *are all on the left and the components* 0 *follow.*

$$\underbrace{11...1}_{v\ times}\ \underbrace{00...0}_{w\text{-}v\ times}$$

The second is where the components 1 *are shifted on the right in such a way that at each step it stays a licit vector. For those, an easy algorithm is proposed in order to construct them. It consist to use the limit parity vector at step* $v-1$ *and replace the last* 0 *by* 1 *and complete to the right with the necessary number of* 0 *to get* $w-v$ *of them in total.*

10
1100
11010
1101100
11011010
1101101100
...

Proof. The proof is in the lemma's self-explanation. □

Lemma 33. *For the second limit case, the following applies. The successive increases of* w *is either* 1 *or* 2*. There can be one increase* $\Delta w = 1$ *but not two. There can be two successive increases* $\Delta w = 2$ *but not three.*

Proof. One increase from rank $v-1$ to v is $\Delta w = \lfloor (ln(3)/ln(2)).v \rfloor - \lfloor (ln(3)/ln(2)).(v-1) \rfloor = ln(3)/ln(2) - [((ln(3)/ln(2)).v - \lfloor (ln(3)/ln(2)).v \rfloor) - ((ln(3)/ln(2)).(v-1) - \lfloor (ln(3)/ln(2)).(v-1) \rfloor)]$. The term between the brackets [ ] is either equal to $-1 + ln(3)/ln(2)$ or $-2 + ln(3)/ln(2)$ and therefore Δw is either equal to 1 or 2. A jump from rank $v-2$ to v corresponds to $\Delta ws = \lfloor (ln(3)/ln(2)).v \rfloor - \lfloor (ln(3)/ln(2)).(v-2) \rfloor = 2.ln(3)/ln(2) - [((ln(3)/ln(2)).v - \lfloor (ln(3)/ln(2)).v \rfloor) - ((ln(3)/ln(2)).(v-2) - \lfloor (ln(3)/ln(2)).(v-2) \rfloor)]$. The term between the brackets [ ] is then either $-3+2.ln(3)/ln(2)$ or $-4+2.ln(3)/ln(2)$ and therefore Δws is either equal to 3 or 4. The jump from rank $v-3$ to v is equal to $\Delta wt = \lfloor (ln(3)/ln(2)).v \rfloor - \lfloor (ln(3)/ln(2)).(v-3) \rfloor = 3.ln(3)/ln(2) - [((ln(3)/ln(2)).v - \lfloor (ln(3)/ln(2)).v \rfloor) - ((ln(3)/ln(2)).(v-3) - \lfloor (ln(3)/ln(2)).(v-3) \rfloor)]$. The term between the brackets [ ] is then either $-4 + 3.ln(3)/ln(2)$ or $-5 + 3.ln(3)/ln(2)$ and therefore Δwt is either equal to 4 or 5. Combining the $\Delta w = 1$ or 2 and $\Delta ws = 1+2$ or $2+2$ constraints there can be only one increase of Δw of spacing 1 but it is possible to have two increases of spacing 2. Using the previous result, combining it with the $\Delta wt = 1+2+1$ or $2+2$ or $1+2+2$ or $2+1+2$ or $2+2+1$ constraints there can be only two successive increases of Δw of spacing 2. □

Definition. *We call a* 2^w-*seed an integer in the interval* $[0, 2^w - 1]$ *of stopping time* w*. The number of* 2^w-*seeds is noted* $\#s_w$*.*

A sample of the number of seeds is shown in the table underneath.

v	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
w	1	2	4	5	7	8	10	12	13	15	16	18	20	21	23
Δw		1	2	1	2	1	2	2	1	2	1	2	2	1	2
Δwt (to be added)				1 2 1	2 1 2	1 2 1	2 1 2	1 2 2	2 2 1	2 1 2	1 2 1	2 1 2	1 2 2	2 2 1	2 1 2
Δw_f			1	2	1	2	1	2	2	1	2	1	2	2	1
$i = 1$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2				1	2	3	4	5	6	7	8	9	10	11	12
3						3	7	12	18	25	33	42	52	63	75
4								12	30	55	88	130	182	245	320
5									30	85	173	303	485	730	1050
6											173	476	961	1691	2741
7													961	2652	5393
8														2652	8045
$\#s_w$	1	1	1	2	3	7	12	30	85	173	476	961	2652	8045	17637

Theorem 6. *The number of seeds $\#s_w$ at rank v is equal to the sum of terms $\#ps_w(v,i)$ where $\#ps_w(v,i) = \#ps_w(v,i-1) + \#ps_w(v-1,i)$ is applied so long that $\#ps_w(v-1,i) \neq 0$. Moreover the last term $\#ps_w(v,i)$ is then repeated when $\Delta w_f = 2$.*

Proof. It is easy to check the first steps when increasing incrementally the rank v. Let us then consider the rank $v-1$ and v for $v > 2$. Having collected the parity vectors up to that stage, the said vectors are licit parity vectors. The parity vectors are then constructed from the previous ones adding to them either 1 if $\Delta w = 1$ or 1 and 0 if $\Delta w = 2$. If those are added behind the vectors of rank $v-1$, one doesn't have to recheck the validity of the writing of the initial portion of the parity vectors but only what happens with the added part in regard of the said initial part. For easier understanding of the present proof, the reader will refer to the figures 14 and 15. The portions in red are added digits. The portions in blue are swapping of 10 in some initial portions to 01. The parity vectors containing red added portions are found in the same column as in the previous rank and their quantities are of course equal to those of the said previous rank. In the figures, we add them under to swapped parity vectors. The portions are included after the last digit 1 of the previous rank's items and are either 1 or 10 according to the value of Δw. The parity vectors with swapped part are deduced from the elements at their left. The swapping of the last 10 portion to 01 is continued to the right so long the vectors remains licit. Adding at some rank only 1 to the parity vectors follows in a "deficit" of 0 at the next rank. There won't be no possibility of an extra swapping creating a new column. The addition of 10 at the previous rank on the contrary allows it. A more precise way to check it is to exhaust all cases knowing by lemma 33 that there are only four possibilities to examine (f stands for former and f_f for the former of the former):

Δw_{f_f}	1	2	1	2
Δw_f	2	1	2	2
Δw	1	2	2	1

It is easy to check the compliance to the announced result by reviewing the tables up to rank $v = 8$.
To finish with,we have to check that all the licit parity vectors are reach in this way and that there are no redundancies. Therefore we collect the data from these rank v tables line by line in a peculiar way and put them in a vertical order. Figure 16 shows such an reordering. We start by the first line of some initial table corresponding to rank v. Then we pick the next line with one element less than the current line (that is the second line at the first iteration), then again the next line with one element less than the current one and so on. Once exhausted the downwards path, we restart the process on the first remaining line while changing column. This construction provides the parity vectors in an increasing order of its (binary) values, showing obvious exhaustiveness and non-redundancy. □

Note. The first construction is very practical for the enumeration of the parity vectors. The second construction has obvious advantage to check exhaustiveness. It has also specific properties that the interested reader may find in reference [9].

The density in $\mathbb{Z}$ of all number with a finite stopping time is equal to

$$ds_\infty = \sum_{v=0}^{\infty} \frac{\#s_w}{2^w}. \tag{11}$$

The figure 17 shows the evolution of the difference to 1 of the density of the said numbers up to the rank v :

$$dif_v = 1 - ds_v. \tag{12}$$

With logarithmic coordinates for the ordinates, the figure shows at an early stage a "linear" picture, a clear stand of its 0 asymptotic value. Of course the Riho Terras' theorem has established that limit in 1976 [9] as the natural density 1 result over $\mathbb{N}$ is equivalent to it over $\mathbb{Z}$.

Lemma 34. *Asymptotically, as v increases, the number of the 2^w−seeds in the interval $[0, 2^w[$ tends towards c^w where c is a constant and $c \approx 1.927$.*

Proof. The density of the seeds approaching 1 in $\mathbb{N}$ means in the same time that asymptotically $\#s_w \lesssim 2^w$ and that the trend must be of the same kind, that is exponential in w. A polynomial growth can in no way be strong enough because any such form w^c, c a constant, will trend towards 0 compared with 2^w whatever the huge size of c. Figure 18 shows how $\#s_w$ and 2^w do compare. Numerically, a fair approximation is $\#s_w \approx 1.927^w \simeq 2.828^v$ (with 1.92^w, as shown in the figure, the curves will ultimately cross). □

v	w	Δw	Δw_f	Col1	Col2	Col3	Col4
0	1	1		0			
1	2	1	1	10			
2	4	2	1	1100			
3	5	1	2	11100	11010		
4	7	2	1	1111000	1110100		
					1101100		
5	8	1	2	11111000	11110100	11110010	
					11101100	11101010	
					11011100	11011010	
6	10	2	1	1111110000	1111101000	1111100100	
					1111011000	1111010100	
					1110111000	1110110100	
					1101111000	1101110100	
						1111001100	
						1110101100	
						1101101100	
7	12	2	2	111111100000	111111010000	111111001000	111111000100
					111110110000	111110101000	111110100100
					111101110000	111101101000	111101100100
					111011110000	111011101000	111011100100
					110111110000	110111101000	110111100100
						111110011000	111110010100
						111101011000	111101010100
						111011011000	111011010100
						110111011000	110111010100
						111100111000	111100110100
						111010111000	111010110100
						110110111000	110110110100

FIGURE 14. Parity vectors of seeds.

Knowing the precise growth of the 2^w-seeds, which are the number of integers $\#s_w(v)$ with stopping time w within the interval $[0, 2^w[$, we are able immediately enumerate the cardinals $\#plf(v)$ and $\#plr(v)$ that are respectively the number of integers with stopping times strictly smaller than w and strictly greater than w. The computer program in appendix B provides the means to evaluate, at rank v, the 3 cardinals $\#s_w(v)$, $\#plf(v)$ and $\#plr(v)$. This allows us to get the very interesting following ratio.

Lemma 35. *Asymptotically, the average interval available for the seeds of stopping time equal or greater than w is approximatively equal to* $10^4(1.056921^v)$.

Proof. The number of seeds of stopping time strictly greater than $w(v-1)$, which is the stopping time at rank $v-1$, is $\#plr(v-1)$. This is in some way the number of empty places at rank $v-1$ available for the placement of

v	w	Δw	Δw_f	Col1	Col2	Col3	Col4	Col5
8	13	1	2	1111111100000	1111111010000	1111111001000	1111111000100	1111111000010
					1111110110000	1111110101000	1111110100100	1111110100010
					1111101110000	1111101101000	1111101100100	1111101100010
					1111011110000	1111011101000	1111011100100	1111011100010
					1110111110000	1110111101000	1110111100100	1110111100010
					1101111110000	1101111101000	1101111100100	1101111100010
						1111110011000	1111110010100	1111110010010
						1111101011000	1111101010100	1111101010010
						1111011011000	1111011010100	1111011010010
						1110111011000	1110111010100	1110111010010
						1101111011000	1101111010100	1101111010010
						1111100111000	1111100110100	1111100110010
						1111010111000	1111010110100	1111010110010
						1110110111000	1110110110100	1110110110010
						1101110111000	1101110110100	1101110110010
						1111001111000	1111001110100	1111001110010
						1110101111000	1110101110100	1110101110010
						1101101111000	1101101110100	1101101110010
							1111110001100	1111110001010
							1111101001100	1111101001010
							1111011001100	1111011001010
							1110111001100	1110111001010
							1101111001100	1101111001010
							1111100101100	1111100101010
							1111010101100	1111010101010
							1110110101100	1110110101010
							1101110101100	1101110101010
							1111001101100	1111001101010
							1110101101100	1110101101010
							1101101101100	1101101101010

FIGURE 15. Parity vectors of seeds.

the seeds at rank v taking however account of the following remark. At rank v, the previous interval $[0, 2^{w(v-1)}[$ in which we find the seeds is multiplied by 2 or 4 to the size $[0, 2^{w(v)}[$. We note this multiplicative factor $2^{\Delta w}$. We therefore have to evaluate the ratio $\frac{2^w}{2^{\Delta w}(\#plr(v-1)+1)}$ to get the mean distance available at rank v (note that $+1$, due to the difference between number of points and number of intervals, is negligible very soon as v increase). As the density of numbers of finite stopping time is 1 in $\mathbb{N}$, as previously for $\#s_w$, the growth is necessarily exponential in regard to v (or w). We are able then to compare this ratio with the approximation $10^4(1.056921^v)$, also equivalent to the approximation $10^4(1.035545^w)$. Figure 19 shows that comparison. $\square$

At this stage it is essential to our purpose to compare the position of the smallest seeds at rank k with the average distance obtained above. Let us note that it is difficult to get numerically a greater number of the smallest 2^w-seeds. Our sample of such elements is given in appendix C and figure 20 provides a close up of figure 19 in the zone of interest keeping the same colors for the corresponding curves. We see that most of the seeds are situated at higher ordinates than the one excepted by average distancing and that

v	w	Δw	Δw_f	C1	C2	C3	C4	C5	C6	C7
0	1	1		0						
1	2	1	1	10						
2	4	2	1	1100						
3	5	1	2	11100	11010					
4	7	2	1	1111000	1110100	1101100				
5	8	1	2	11111000	11110100	11101100	11011100			
					11110010	11101010	11011010			
6	10	2	1	1111110000	1111101000	1111011000	1110111000	1101111000		
					1111100100	1111010100	1110110100	1101110100		
						1111001100	1110101100	1101101100		
7	12	2	2	111111100000	111111010000	111110110000	111101110000	111011110000	110111110000	
					111111001000	111110101000	111101101000	111011101000	110111101000	
					111111000100	111110100100	111101100100	111011100100	110111100100	
						111110011000	111101011000	111011011000	110111011000	
						111110010100	111101010100	111011010100	110111010100	
							111100111000	111010111000	110110111000	
							111100110100	111010110100	110110110100	
8	13	1	2	1111111100000	1111111010000	1111110110000	1111101110000	1111011110000	1110111110000	1101111110000
					1111111001000	1111110101000	1111101101000	1111011101000	1110111101000	1101111101000
					1111111000100	1111110100100	1111101100100	1111011100100	1110111100100	1101111100100
					1111111000010	1111110100010	1111101100010	1111011100010	1110111100010	1101111100010
						1111110011000	1111101011000	1111011011000	1110111011000	1101111011000
						1111110010100	1111101010100	1111011010100	1110111010100	1101111010100
						1111110010010	1111101010010	1111011010010	1110111010010	1101111010010
						1111110001100	1111101001100	1111011001100	1110111001100	1101111001100
						1111110001010	1111101001010	1111011001010	1110111001010	1101111001010
							1111100111000	1111010111000	1110110111000	1101110111000
							1111100110100	1111010110100	1110110110100	1101110110100
							1111100110010	1111010110010	1110110110010	1101110110010
							1111100101100	1111010101100	1110110101100	1101110101100
							1111100101010	1111010101010	1110110101010	1101110101010
								1111001111000	1110101111000	1101101111000
								1111001110100	1110101110100	1101101110100
								1111001110010	1110101110010	1101101110010
								1111001101100	1110101101100	1101101101100
								1111001101010	1110101101010	1101101101010

FIGURE 16. Parity vectors of seeds.

the likely asymptotic positions is roughly equal to the ordinates $(1+\epsilon)^v$, $\epsilon \approx .056921$. These higher positions can be explained easily considering the ratio $\frac{2^w}{\#s_w}$ which is the ratio between the size of interval $[0, 2^w[$ by the number of seeds $\#s_w$ in the said interval at rank v. The close up figure 20 illustrates the framing by the two relevant ratios (in light blue and deep blue). Note that the upper framing is bumpy because of the two different cases $\Delta w = 1$ and $\Delta w = 2$.

Of course the effective values of the seeds are not limited to the said framing, which is only the typical framing if the 2^w−seeds would check at some exact equal spacing. There is no wonder that the effective positions are regularly off the scale, sometimes by more than a multiplicative factor of one decade (10^{-1} or 10^{+1}). It shows nevertheless a satisfactory match and a fair asymptotic

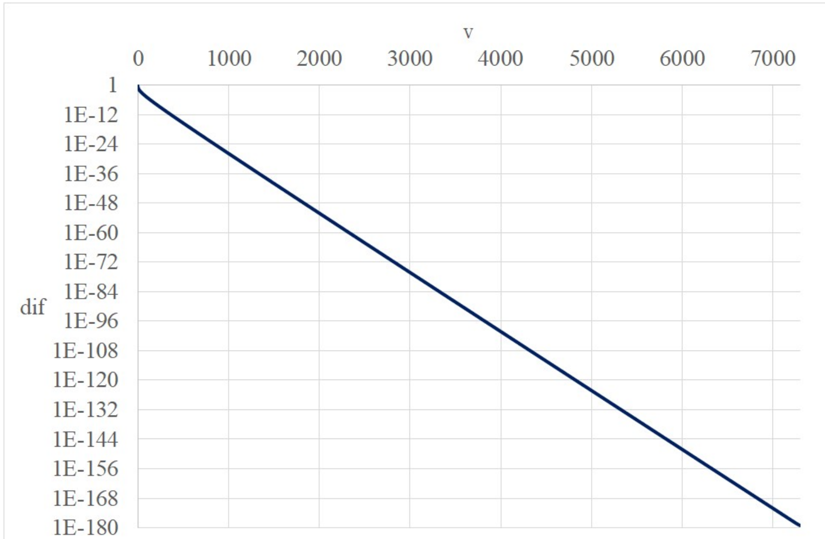

FIGURE 17. $dif_v = 1 - \sum_0^v \frac{\#s_w}{2^w}$.

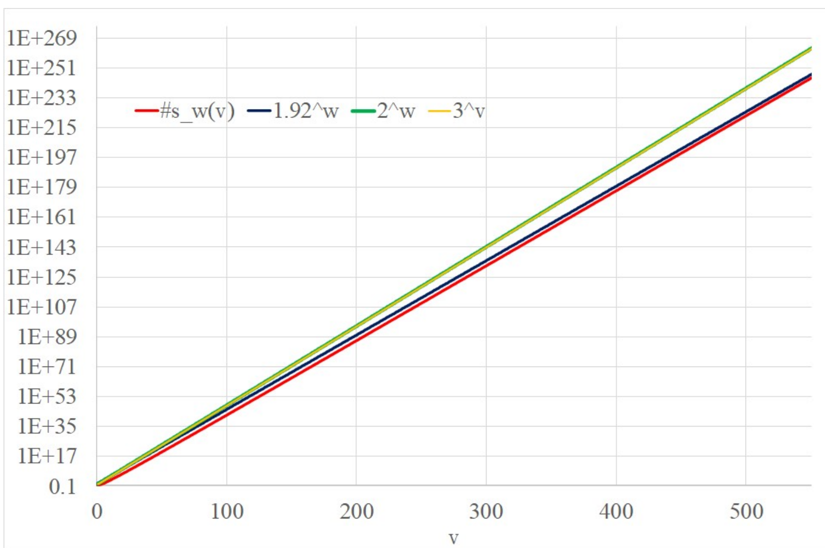

FIGURE 18. Comparison of $\#s_w$ and 2^w.

trend for our purpose, notably on the lower limit (no exception for $60 < v < 200$) which is the useful limit here.

Lemma 36. *For $v > 200$, the value of minimal seed, at rank v, is at most by a few decades near the approximation $10^4(1.035545^w)$*

Proof. For small v as shown in figure 20, the offset is hardly one decade. There are just a few exceptions around $v = 30$. The random nature of the repartition of the seeds increases with v necessarily. Moreover, the evolution

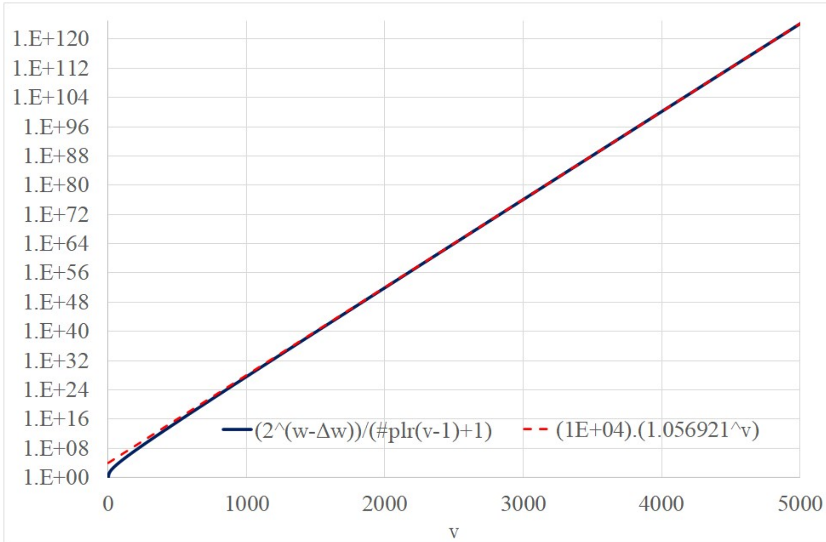

FIGURE 19. Comparison of the average distance of seeds of stopping time greater than w with $10^4(1.056921^v)$.

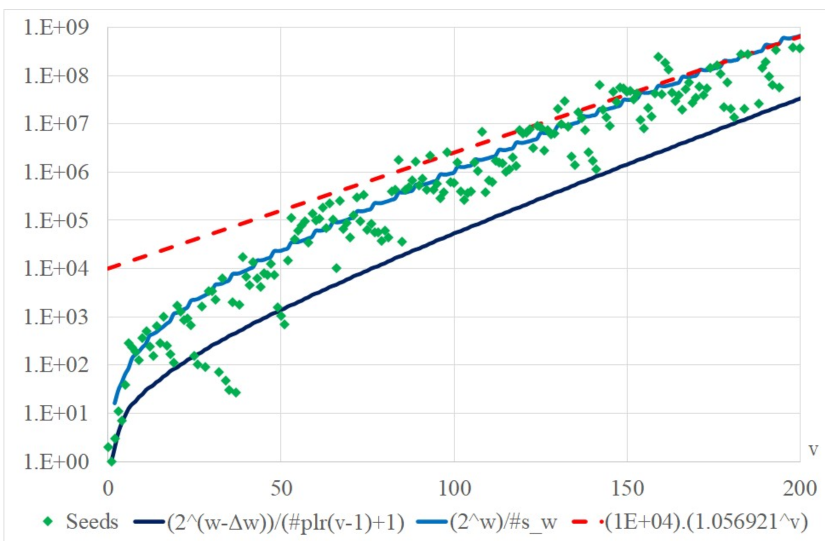

FIGURE 20. Comparison of the lower minimum and higher minimum average distance of seeds of stopping time greater than w with the exact abscissa of the known smallest seeds.

of the randomness is linked to an exponential growth of possible combinations of the seeds positions, therefore making the phenomena happening even much faster. $\square$

Lemma 37. *The evolution of the average distance* $\frac{2^{w-\Delta w}}{\#plr(v-1)+1}$ *available at rank* v *can be traced as* αv^{nc} *where* nc *is no more a constant but an increasing factor of* v. *With* $\alpha = \frac{1}{4ln(2)}$, *the exponent* nc *is greater than* 2 *as soon as rank* $v = 39$ *(and greater than* 3 *as soon as rank* $v = 146$*) and diverges as* $v \to +\infty$.

Proof. Figure 21 shows the numerical evolution of nc for v up to 5000. Besides any expression t^v, $t > 1$ diverges faster than any polynomial in v asymptotically. □

Note. The next section of this article will explain why we made the comparison precisely with $\frac{1}{4ln(2)}v^2$ (and $\frac{1}{4ln(2)}v^3$), the position of the smallest seeds being here a minima and the position of the smallest "generating" element of a cycle going to be there a maxima, at the same rank v, hence providing incompatible results.

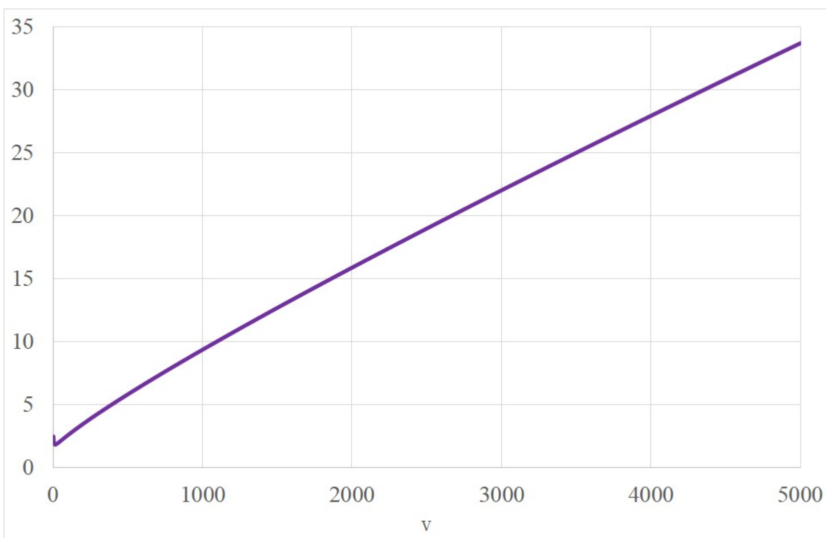

FIGURE 21. Value of nc such that v^{nc} approximates the exponential evolution of $\frac{2^{w-\Delta w}}{\#plr(v-1)+1}$.

9. THE COLLATZ CYCLES' SCARCITY

Lemma 38. *There are no known complete cycle in a unique column of the* 2^w-*period classification set except for seed* 1.

Proof. This is an immediate result of theorem 4. A complete cycle is included in a unique column of the 2^w-period classification table if and only if $w = \lfloor (ln(3)/ln(2))v \rfloor + 1$ which is false in the set $\mathbb{Z} - \mathbb{N}$. The summary for the known cycles is as follows :

v	1	0	0	0	1	0
x_0	1	4	2	-2	-7	-10
x_1	2	2	1	-1	-10	-5
x_2	1				-5	

v	6	5	1	0	3	2	1	0	0	0
x_0	-25	-37	-55	-82	-41	-61	-91	-136	-68	-34
x_1	-37	-55	-82	-41	-61	-91	-136	-68	-34	-17
x_2	-55	-82	-41		-91	-136	-68			
x_3	-82	-41			-136	-68				
x_4	-41	-61			-68	-34				
x_5	-61	-91			-34					
x_6	-91	-136								
x_7	-136	-68								
x_8	-68	-34								
x_9	-34									
x_{10}	-17									

No column of this 2^w- period classification set do contain the smallest value of the cycle as first generating element of the column except for 1 . □

Proposition. *The reason why the Collatz algorithm is true in $\mathbb{N}$ and the situation is different in $\mathbb{Z}-\mathbb{N}$.*

Argument. We proved in the first part of the article that any tree structure has a root. The root is a cycle, necessarily finite in size, and therefore has a smallest value item. This item is the smallest seed of the tree structure. In order to form a distinct tree, the "nature" of the smallest seed must be different from the other integers of the tree structure, that difference being the only "way out". This happens on the negative side of the integers, where -1, -5 and -17 have, unless almost all the other negative integers, the needed alternative property. The word "almost" is used in the previous phrase as one cannot be sure at that stage that there are no other cases in $\mathbb{Z}-\mathbb{N}$ than the 3 exceptions mentioned. That special property is $w = \lfloor (ln(3)/ln(2)).v \rfloor$ instead of $w = \lfloor (ln(3)/ln(2)).v \rfloor + 1$. On the positive side of the integers there is no possibility for a seed to escape the $w = \lfloor (ln(3)/ln(2)).v \rfloor + 1$ rule and therefore ultimately every integer will end its course at the smallest seed in $\mathbb{N}^*$ which is 1. The Collatz conjecture is true in $\mathbb{N}$ because a cycle is entirely part of the parity vectors' structure while a cycle is not totally in that structure in $\mathbb{Z}-\mathbb{N}$.

Knowing the rule $w = \lfloor (ln(3)/ln(2)).v \rfloor + 1$ for a solution to a Collatz cycle on the positive side of $\mathbb{Z}$, let us now consider x_0 such solution in $\Re^+$ (with v odd steps and w even steps).

Lemma 39. *The smallest rational value solution to the smallest element x_0 of a cycle on the $\mathbb{N}^*$ side of $\mathbb{Z}$ is equal to*

$$x_0 = \frac{3^v - 2^v}{2^w - 3^v}.$$

Proof. As solution to a hyperbolic equation, the cycle solution to some given order combination of odd and even steps is unique. In lemma 24, we got the following intermediary result $y_0 = (3^v/2^w)x_0 + 1/2^{w-v}((3/2)^v - 1)$ when applying first all the (3x+1)/2 multiplications while meeting the relationship $w = \lfloor (ln(3)/ln(2)).v \rfloor + 1$. The solution to $y_0 = x_0$ is therefore the proposed one because applying any other combinations of the even and odd steps will increase the solution x_0. □

Lemma 40. *Using a logarithmic scale on the ordinates, the expression $f(v) = \frac{3^v-2^v}{2^w-3^v} - 1$ is located around the horizontal axis 1, in some apparently symmetrical way, with diverging points (towards $+\infty$ or towards 0^+) depending largely on the rational approximations of $\frac{ln(2)}{ln(3)}$.*

Proof. The figure 22 visualize the lemma for the reader and provides an answer to the "symmetry" around ordinate 1−axis. The said expression diverges if and only if $2^w - 3^v \to 0$, which is equivalent to $2^w/3^v \to 1$. Replacing w by its value, we get

$$\begin{aligned} \frac{2^{\lfloor (ln(3)/ln(2))v \rfloor + 1}}{3^v} &= 2^{\lfloor (ln(3)/ln(2))v \rfloor + 1 - (ln(3)/ln(2))v} \ \frac{2^{(ln(3)/ln(2))v}}{3^v} \\ &= 2^{\lfloor (ln(3)/ln(2)).v \rfloor + 1 - (ln(3)/ln(2)).v} \\ &\to 1 \end{aligned}$$

and therefore $\lfloor \frac{ln(3)}{ln(2)} v \rfloor + 1 - \frac{ln(3)}{ln(2)} v \to 0$.
These kind of events occurs of course only if $\frac{ln(3)}{ln(2)} v$ approaches an integer value, so that $\frac{ln(3)}{ln(2)} v \to n$ equivalent to $\frac{ln(2)}{ln(3)} \to v/n$ for some $n \in \mathbb{N}$ (and $v \in \mathbb{N}$). Let us observe that $\frac{ln(2)}{ln(3)} \approx 0.63092975$ and that $\{53/84 \approx 0.63095238, 306/485 \approx 0.630927835, 665/1054 \approx 0.63092979\}$ and therefore the diverging locations' representation on the figure is only approximative with $v = 0$ mod 53 being one of the proposed locations (in green) and $v = 0$ mod 306 a stronger one (in red) and $v = 0$ mod 665 even more so (in yellow) as the fraction narrows the goal in a better way. □

Note. The previous lemma shows the importance to get the best rational approximations of the the real number $\frac{ln(2)}{ln(3)}$. For any continued fraction, the best rational approximations are also called the convergents of the continued fraction [7] and are represented by the Gaussian brackets [8].

Lemma 41. *The coefficients cf_i of the continued fraction of $\frac{ln(2)}{ln(3)}$ follow fairly a Gauss-Kuzmin discrete probability distribution. That is*

$$p(cf_i) \to -\log_2(1 - \frac{1}{(cf_i + 1)^2})$$

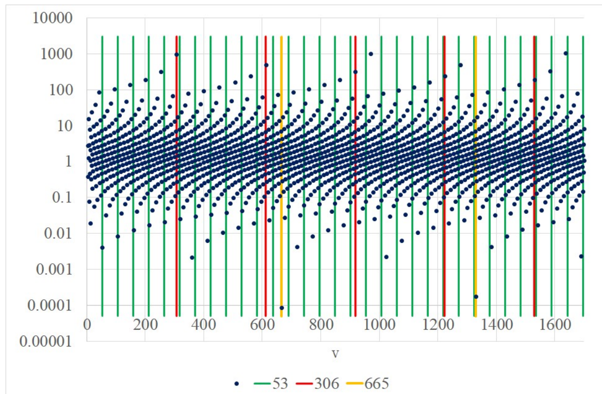

FIGURE 22. Data $\frac{3^v-2^v}{2^w-3^v}-1$.

Proof. The Gauss-Kuzmin discrete probability distribution $p(cf_i)$ arises as the limit probability distribution of the coefficients in the continued fraction expansion of a random variable uniformly distributed in $(0,1)$ [8]. A numerical verification, with a small size sample (2000 elements), shows that the coefficients of the continued fraction of $\frac{ln(2)}{ln(3)}$ follow fairly that distribution. ☐

Lemma 42. *The offset* Δr *of* $\frac{ln(2)}{ln(3)}$ *with its best rational approximations is approximatively*

$$\Delta r = (\frac{ln(2)}{ln(3)}\frac{1}{v})^2.$$

Proof. Here the continued fraction cf starts with the Gaussian bracket [0; 1, 1, 1, 2, 2, 3, 1, 5, 2, 23, 2, 2, 1, 1, 55, 1, 4, 3, 1, ...]. Table 2 gives the corresponding resulting extracted fractions. Using the theory of Diophantine approximations, the Dirichlet theorem states that there exists for any positive irrational number ir an infinity of couples $(p,q) \in N * N^*$ such that $\mid ir - \frac{p}{q} \mid < \frac{a}{q^2}$, a being some finite value. This theorem is of course optimum for the best rational approximations. Applying this to $ir = \frac{ln(2)}{ln(3)}$, we can thus find two infinite series of integers $\{v_i\}$, $\{wr_i\}$ and an infinite series of real numbers $\{a_i\}$ such that $\frac{ln(2)}{ln(3)} - \frac{v_i}{wr_i} = \frac{a_i}{wr_i^2}$. The index i is here a dummy index and the values of v_i are the one corresponding to v for which we get the said best approximations. Besides, the denominator wr_i is either equal to w or $w-1$ according to the cases where $\frac{ln(3)}{ln(2)}v_i$ tends to an integer from beneath or from above. An approximate value $|a_i| \approx 1$ is then obtained by numerical verification. Asymptotically $w_i = \lfloor \frac{ln(3)}{ln(2)}v_i \rfloor + 1 \approx \frac{ln(3)}{ln(2)}v_i$ and therefore $|\frac{ln(2)}{ln(3)} - \frac{v_i}{wr_i}| \approx (\frac{ln(2)}{ln(3)}\frac{1}{v_i})^2$. The figures 23, 24, 25 and 26 show the excellent match by making the choice $|a_i| \approx 1$ which is likely the asymptotic exact value. In the 3 last figures, we indicate the values of the coefficients of the continued fraction next to the representative points. We remark that the higher these coefficients the better the "fine tuning" at that step with the exact excepted value. $\square$

i	cf	fraction	v	w	approx	$x_0(v)$	$x_0(v, w^-)$
0	0	0/1	0	1	0	0	
1	1	1/1	1	2	1	1	-1
2	1	1/2	1	2	0.5	1	
3	1	2/3	2	4	0.66666667	0.71429	-5
4	2	5/8	5	8	0.62500000	16.23077	
5	2	12/19	12	20	0.63157895	1.01974	-73.72361
6	3	41/65	41	65	0.63076923	86.7389	
7	1	53/84	53	85	0.63095238	1.00419	-479.39702
8	5	306/485	306	485	0.63092784	977.7448	
9	2	665/1054	665	1055	0.63092979	1.00009	-22907.85023
10	23	15601/24727	15601	24727	0.63092975	54960.9	
11	2	31867/50508	31867	50509	0.63092975	1.00001	-137648.0025
12	2	79335/125743	79335	125743	0.63092975	272871.59	
13	1	111202/176251	111202	176252	0.63092975	1.00001	-277761.83
14	1	190537/301994	190537	301994	0.63092975	15502072.2	
15	55	10590737/16785921	10590737	16785922	0.63092975	1.00000	-19120269.3
16	1	10781274/17087915	10781274	17087915	0.63092975	81920324.8	
17	4	53715833/85137581	53715833	85137582	0.63092975	1.00000	-287969592.7
18	3	171928773/272500658	171928773	272500658	0.63092975	558903955.	
19	1	225644606/357638239	225644606	357638240	0.63092975	1.00000	$-594045517.$

TABLE 2

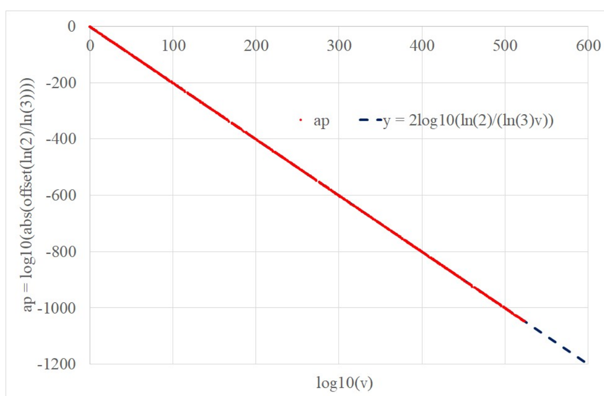

FIGURE 23. Comparison of $ap = log_{10}(abs(\frac{ln(2)}{ln(3)} - \frac{v}{wr}))$ with $2log_{10}(\frac{ln(2)}{ln(3)}\frac{1}{v})$.

Lemma 43. *The following strictly inequality is true for all $v > 2$*

$$\frac{1}{2} < \frac{2^{w-1} + 2^{v-1}}{3^v} < 1.$$

Proof. If $v = 0$, $\frac{2^{w-1}+2^{v-1}}{3^v} = \frac{3}{2}$. If $v = 1$, $\frac{2^{w-1}+2^{v-1}}{3^v} = 1$. If $v = 2$, $\frac{2^{w-1}+2^{v-1}}{3^v} = \frac{10}{9}$. These cases are excluded.
The equality $2^{w-1} + 2^{v-1} = 3^v$ is obviously false as soon as $v > 1$ for contradiction on parity between the two members of the equation, the first one

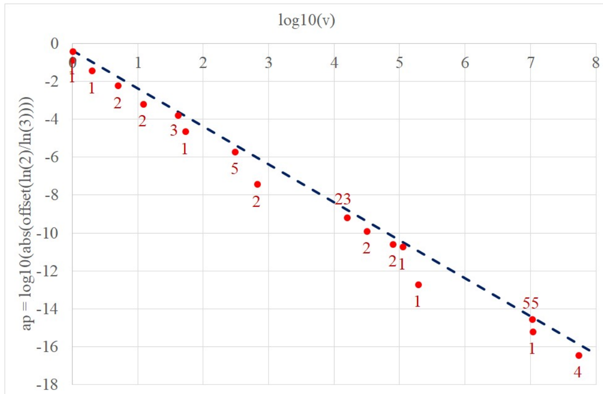

FIGURE 24. Detail relative to figure 23.

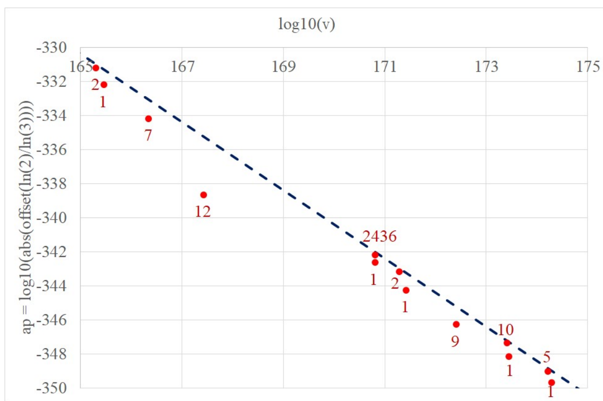

FIGURE 25. Detail relative to figure 23.

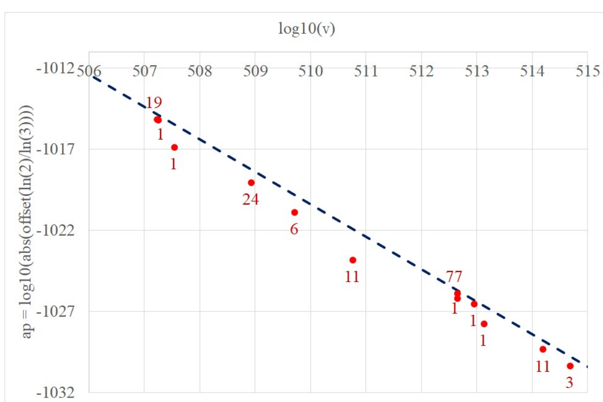

FIGURE 26. Detail relative to figure 23.

being even, the second odd. One can then check the inequality for a significant number of values of v. Using the same approach as in the proof of lemma 42, we can write rigorously $\frac{ln(3)}{ln(2)} - \frac{wr_i}{v_i} = \frac{c_i}{{v_i}^2}$, where $|c_i| \approx 1$ as soon as for example $v = 10$. That is equivalent to $\frac{ln(3)}{ln(2)}v_i - wr_i = \frac{c_i}{v_i}$ where wr_i is either equal to w or $w-1$ according to the cases where $\frac{ln(3)}{ln(2)}v_i$ tends to an integer from beneath or from above. Therefore, getting rid of indices, $w = \frac{ln(3)}{ln(2)}v + or(0,1) - \frac{c}{v}$, $|c| \approx 1$. As $w = \lfloor\frac{ln(3)}{ln(2)}v\rfloor + 1 = \frac{ln(3)}{ln(2)}v + \lfloor\frac{ln(3)}{ln(2)}v\rfloor - \frac{ln(3)}{ln(2)}v + 1$, we get $or(0,1) - \frac{c}{v} = \lfloor\frac{ln(3)}{ln(2)}v\rfloor - \frac{ln(3)}{ln(2)}v + 1$. Therefore if $\frac{ln(3)}{ln(2)}v$ tends to an integer from beneath, we get $\frac{c}{v} = \frac{ln(3)}{ln(2)}v - \lfloor\frac{ln(3)}{ln(2)}v\rfloor - 1$, $c \approx -1$ and if $\frac{ln(3)}{ln(2)}v$ tends to an integer from above, we get $\frac{c}{v} = \frac{ln(3)}{ln(2)}v - \lfloor\frac{ln(3)}{ln(2)}v\rfloor$, $c \approx 1$. This can be summarized, i being some positive integer that $\frac{ln(3)}{ln(2)}v$ is the nearest by, with

$$\begin{array}{ll} \frac{ln(3)}{ln(2)}v - \lfloor\frac{ln(3)}{ln(2)}v\rfloor \approx \frac{c}{v} & \text{if } \frac{ln(3)}{ln(2)}v \to i^+ \\ \frac{ln(3)}{ln(2)}v - \lfloor\frac{ln(3)}{ln(2)}v\rfloor \approx 1 - \frac{c}{v} & \text{if } \frac{ln(3)}{ln(2)}v \to i^- \end{array}$$

where $c \approx 1$.
Now $\frac{2^{w-1}}{3^v} = \frac{2^{\lfloor\frac{ln(3)}{ln(2)}v\rfloor}}{3^v} = \frac{2^{\lfloor\frac{ln(3)}{ln(2)}v\rfloor - \frac{ln(3)}{ln(2)}v}2^{\frac{ln(3)}{ln(2)}v}}{3^v} = 2^{\lfloor\frac{ln(3)}{ln(2)}v\rfloor - \frac{ln(3)}{ln(2)}v} = 2^{or(-\frac{c}{v}, -1+\frac{c}{v})}$.
Then $\frac{2^{w-1}+2^{v-1}}{3^v} = 2^{or(-\frac{c}{v}, -1+\frac{c}{v})} + \frac{2^{v-1}}{3^v}$, hence two cases.
If $\frac{ln(3)}{ln(2)}v \to i^+$ provides $2^{-\frac{c}{v}} + \frac{2^{v-1}}{3^v} = e^{-ln(2)\frac{c}{v}} + \frac{2^{v-1}}{3^v} \approx 1 - ln(2)\frac{c}{v} + \frac{1}{2}(\frac{2}{3})^v$. As v is exponentiated in the last term, this one will converges to 0 faster than the term preceding it. Therefore $\frac{2^{w-1}+2^{v-1}}{3^v} \to 1^-$.
If $\frac{ln(3)}{ln(2)}v \to i^-$ provides $2^{-1+\frac{c}{v}} + \frac{2^{v-1}}{3^v} = \frac{1}{2}e^{ln(2)\frac{c}{v}} + \frac{2^{v-1}}{3^v} \approx \frac{1}{2} + \frac{1}{2}ln(2)\frac{c}{v} + \frac{2^{v-1}}{3^v}$ $\to \frac{1}{2}^+$. These are of course limit cases but are the only ones we have to be concerned with as the intermediary cases' values will land between these two.

Let us be even more cautious and go back to the approximate value of c and have a look at the inconsistency of its value if the limit cases were to be met. If we reconsider the above first case, the limit situation would be to write the equality $2^{-\frac{c}{v}} + \frac{2^{v-1}}{3^v} = 1$. When v increases asymptotically, bringing the second term on the left side of the equation near 0, obviously to meet the equality would require to bring the first one up near 1 and therefore c nearer and nearer to 0. More precisely, using $ln(1+x) \to x$ for small x, will conduct to $c = -v\frac{ln(1-\frac{1}{2}(\frac{2}{3})^v)}{ln(2)} \to \frac{v}{2ln(2)}(\frac{2}{3})^v$ and therefore $log_{10}(c) \to log_{10}(\frac{v}{2ln(2)}) + v.log_{10}(\frac{2}{3}) \approx v.log_{10}(\frac{2}{3}) \approx -0.176v$. This is to be compared to $log_{10}(c) = log_{10}(1) = 0$ that we used for figure 23. The basis formula $|\frac{ln(2)}{ln(3)} - \frac{v}{wr}| \approx (\frac{ln(2)}{ln(3)}\frac{1}{v})^2$ for this figure is to be replaced by $|\frac{ln(2)}{ln(3)} - \frac{v}{wr}| \approx c.(\frac{ln(2)}{ln(3)}\frac{1}{v})^2$ which is equivalent to $log_{10}|\frac{ln(2)}{ln(3)} - \frac{v}{wr}| \approx log_{10}(c) + log_{10}((\frac{ln(2)}{ln(3)}\frac{1}{v})^2)$ $\approx -0.176v + log_{10}((\frac{ln(2)}{ln(3)}\frac{1}{v})^2)$. This gives an offset as represented in figure 27. The dark blue dashed line is now replaced by the clear blue dashed line. This is totally incompatible with the Gauss-Kuzmin discrete probability distribution's red (almost) line (remembering especially that the representation is in logarithmic scale). Let us not forget here, as mentioned previously [2], that any unknown cycle in $\mathbb{N}^*$ would contain at least 17026679261 vertices, that is $v \geq 10742638550$ or $log_{10}(v) \geq 10.03$. At this stage, to meet the limit case, $c \approx 1$ has to be replaced already by the way off value $c \approx 1.4\ 10^{-1891684748}$ to compensate (which is quite absurd).

We can resume the former argument for the second case. The result is then very close from the previous one, with only a change in sign of c and a slight change in absolute value which does not require further analysis. □

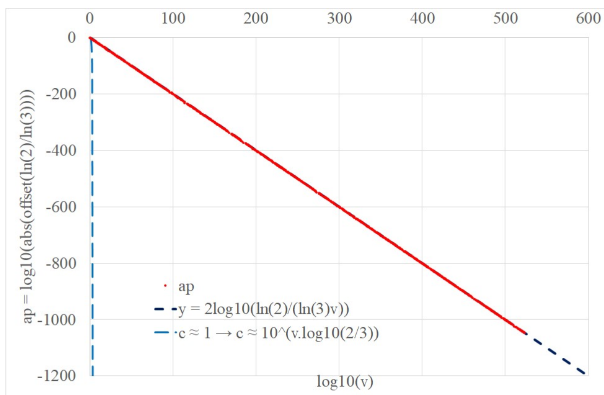

FIGURE 27. Comparison of $ap = log_{10}(abs(\frac{ln(2)}{ln(3)} - \frac{v}{wr}))$ with $-0.176v + 2log_{10}(\frac{ln(2)}{ln(3)}\frac{1}{v})$ (clear blue line).

Lemma 44. *Let us write num_i, den_i and cf_i the i^{th} numerator, denominator and continued fraction coefficient respectively in table 2 of the best approximations of $ln(2)/ln(3)$. Then*

$$num_i = cf_i \times num_{i-1} + num_{i-2}$$
$$den_i = cf_i \times den_{i-1} + den_{i-2}$$

Proof. This property is a general property of continued fractions (see reference [7] under title "Infinite continued fractions and convergents"). In Gaussian brackets language [8], it writes down precisely as $[cf_1, cf_2, \ldots, cf_i] = [cf_1, cf_2, \ldots, cf_{i-1}]cf_i + [cf_1, cf_2, \ldots, cf_{i-2}]$. □

The computer program in appendix E provides an algorithm to evaluate num_i and cf_i from the sole knowledge of num_{i-1} and num_{i-2}. The same holds for the denominators. It is based on an iterative test $cf = 1, cf = 2, \ldots$ on an equality $r_1 - r_2 = 0$ which fails when $cf = cf_i + 1$, therefore revealing the correct value $cf = cf_i$.

Lemma 45. *The absolute values of the extrema of the expression* $\frac{3^v-2^v}{2^w-3^v}-1$ *are linear with an approximate slope of* $\frac{1}{ln(2)}$ *according to the number of odd steps* v.

Proof. Asymptotically, the term -1 can of course be neglected. Again, using the same approach as in the proof of lemma 42, we can write rigorously $\frac{ln(3)}{ln(2)}-\frac{wr_i}{v_i}=\frac{c_i}{{v_i}^2}$, where $|c_i|\approx 1$. The asymptotic approximate value of $|c_i|$ again is guessed numerically and is quite suggestive again. Other way to obtain the result, but less rigorous, is to simply multiplying by $(ln(3)/ln(2))^2$ on each side of the previously obtained expression $\frac{ln(2)}{ln(3)}-\frac{v_i}{wr_i}=\frac{a_i}{{wr_i}^2}$, where $|a_i|\approx 1\approx|c_i|$. Indeed, asymptotically $(\frac{ln(3)}{ln(2)})^2\frac{v_i}{wr_i}=\frac{(ln(3)/ln(2))v_i}{(ln(2)/ln(3))wr_i}\approx\frac{wr_i}{v_i}$ and $(\frac{ln(3)}{ln(2)})^2\frac{a_i}{{wr_i}^2}=\frac{a_i}{((ln(2)/ln(3))wr_i)^2}\approx\frac{a_i}{{v_i}^2}\approx\frac{c_i}{{v_i}^2}$. Then $\frac{ln(3)}{ln(2)}v_i-\lfloor\frac{ln(3)}{ln(2)}v_i\rfloor-1=\frac{c_i}{v_i}+wr_i-\lfloor\frac{ln(3)}{ln(2)}v_i\rfloor-1\approx\frac{c_i}{v_i}+wr_i-w_i$.
The column $x_0(v)$ in table 2 is that of the extrema when there is an equality between denominator of the fraction and w, otherwise one picks the result in the column $x_0(v,w^-)$ with provides intermediate values but of negative sign (case denominator $=w-1$). For high values of v, and ignoring the index i, $x_0=\frac{3^v-2^v}{2^w-3^v}=\frac{1-(2/3)^v}{2^w/3^v-1}\approx\frac{1}{2^w3^{-v}-1}=\frac{1}{2^{w-\frac{ln(3)}{ln(2)}v}-1}=\frac{1}{2^{w-\lfloor\frac{ln(3)}{ln(2)}v\rfloor-1+(1+\lfloor\frac{ln(3)}{ln(2)}v\rfloor-\frac{ln(3)}{ln(2)}v)}-1}=\frac{1}{2^{-\frac{c}{v}+w-wr}-1}=\frac{1}{2^{w-wr}e^{-ln(2)\frac{c}{v}}-1}$. Then, again for large values of v, using $e^x\approx 1+x$ when $x\to 0$, we get $x_0\approx or(1,\frac{1}{-ln(2)\frac{c}{v}})$ $=or(1,-\frac{v}{ln(2)c})$, $|c|\approx 1$, giving the linear relationship for the values of v we are interested in. □

Figures 28 shows how the values of x_0 evolves at the best approximations of $\frac{ln(2)}{ln(3)}$ compared with $x_{0\_linear}=\frac{v}{ln(2)}$. A higher value of a coefficient of the continued fraction vector results in a higher value than $\frac{1}{ln(2)}v$ for the ordinates $x_0(v)$ or $x_0(v,w^-)$ respectively as illustrate in the figures 29, 30 and 31. Nevertheless more moderate values (within the 2000 first coefficients of the continued fraction, there are 41.65 % of 1, 16.15 % of 2, 9.80 % of 3, 6.0 % of 4, 4.05 % of 5 and only 22.35 % of higher value) coming next to such higher values will bring the resulting $x_0(v)$ or $x_0(v,w^-)$ values back to track rapidly.

Lemma 46. *The upper bound of absolute values of the extrema of the expression* $\frac{3^v-2^v}{2^w-3^v}-1$ *are roughly linear with an approximate slope of* $\frac{1}{ln(2)}$ *according to the number of odd steps* v

$$x_0\approx\frac{1}{ln(2)}v^{1+\chi},\chi\ll 1 \text{ as } v\to+\infty.$$

Proof. The multiplicative factor between the upper bound values line and the basis line, with the notation of the lemma, is v^χ. Taking account of the iteration i, let us write it ${v_i}^{\chi_i}$. Let us consider two consecutive iterations i

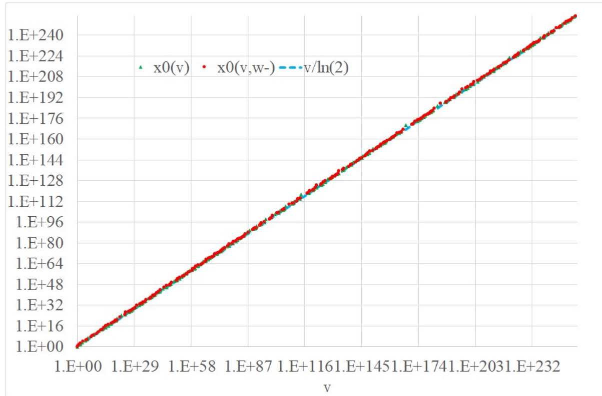

FIGURE 28. Evolution of $\frac{3^v-2^v}{2^w-3^v}$ for the best rational approximations of $\frac{ln(2)}{ln(3)}$ by its continued fraction.

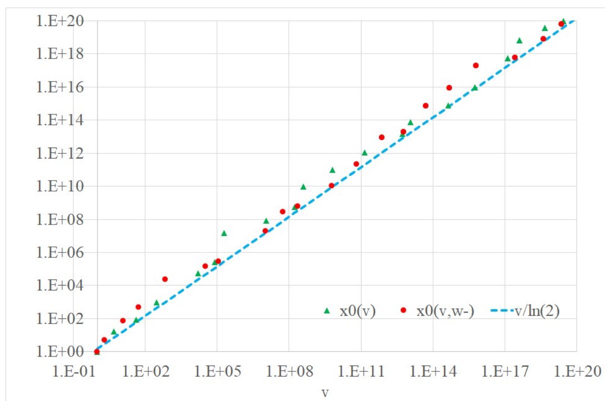

FIGURE 29. Evolution of $\frac{3^v-2^v}{2^w-3^v}$ for the best rational approximations of $\frac{ln(2)}{ln(3)}$. Detail around the origin.

and $i+1$ and the best approximations results x_{0_i} and $x_{0_{i+1}}$. The first one is at most equal to the second one, thus ${v_i}^{\chi_i} = x_{0_{i+1}}/x_{0_i}$. According to lemma 44, $v_i = cf_i \times v_{i-1} + v_{i-2}$ and therefore $x_{0_{i+1}} \approx cf_{i+1} \times x_{0_i}$ if $cf_{i+1} \gg 1$ and $v_i \gtrsim cf_i \times v_{i-1} > \prod_{k=1}^{i} cf_k$. Therefore $\chi_i \approx \frac{ln(cf_{i+1})}{ln(v_i)} < \frac{ln(cf_{i+1})}{ln(\prod_{k=1}^{i} cf_k)}$. The last expression can be greater than 1 only if $cf_{i+1} > \prod_{k=1}^{i} cf_k$ which is more and more unlikely as i increases. Indeed, according to lemma 41, the continued fraction's coefficients of $\frac{ln(2)}{ln(3)}$ follows fairly a Gauss-Kuzmin discrete probability distribution so that $p(cf_i) \to -\log_2(1 - \frac{1}{(cf_i+1)^2}) \approx \frac{1}{ln(2)} \frac{1}{cf_i^2}$, $i \to +\infty$. That last expression shows that the size of cf_i is not increasing with i (but only that higher values are possible with a larger sample). The squared inverse confirms the extremely small likelihood of the event $cf_{i+1} > \prod_{k=1}^{i} cf_k$. More precisely, an additional numerical check shows that the ratio $\frac{cf_i}{\prod_{k=1}^{i-1} cf_k}$

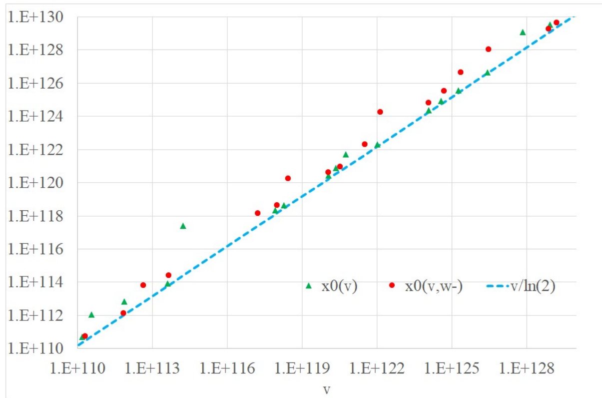

FIGURE 30. Evolution of $\frac{3^v-2^v}{2^w-3^v}$ for the best rational approximations of $\frac{ln(2)}{ln(3)}$. Detail around $i = 232$, continued fraction coefficient $= 964$.

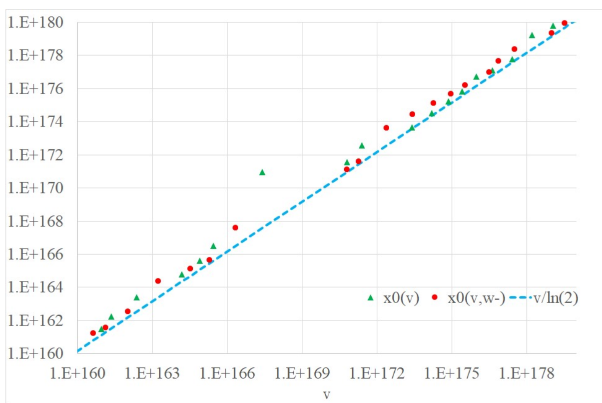

FIGURE 31. Evolution of $\frac{3^v-2^v}{2^w-3^v}$ for the best rational approximations of $\frac{ln(2)}{ln(3)}$. Detail around $i = 332$, continued fraction coefficient $= 2436$.

approximates $(e-1)e^{-i}$, a far-flung from 1 even with small values of the iteration index i (like $i > 15$). Figure 32 features the result χ of the equality ${v_i}^{1+\chi} = cf_i \times v_i$. If $cf_i = 1$, $\chi = 0$ and is not represented in that graphic in logarithmic scale. The representative points of some given constant value of the continued fraction coefficients are on a strictly decreasing curve, the lowest one here is the one corresponding to $cf_i = 2$, the second to $cf_i = 3$ and so on. This representatives curves of some fixed coefficient value cf are getting closer and closer as cf increases. $\square$

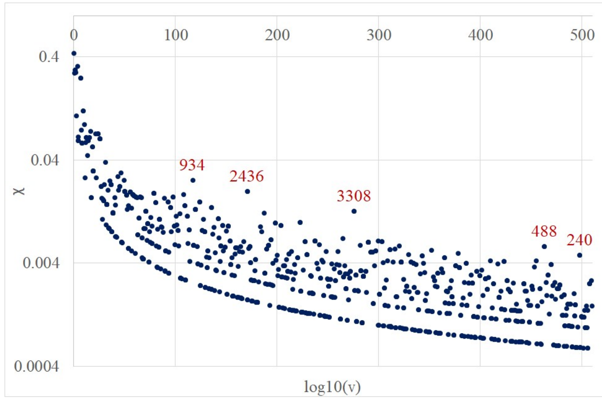

FIGURE 32. Evolution of χ. The numbers indicated are the highest continued fraction coefficients in the displayed range.

Note. An exceptional case $cf_{i+1} > \prod_{k=1}^{i} cf_k$ is not detrimental to our arguments unless $cf_{i+1} \ggg\ggg \prod_{k=1}^{i} cf_k$ so that $\chi \ggg 1$ as we will see later on.

Lemma 47. *The largest rational value, solution to the smallest element x'_0 of a cycle on the $\mathbb{N}^*$ side of $\mathbb{Z}$, is asymptotically approximately equal to*

$$x'_0 \approx \frac{v}{4}\,\frac{3^v - 2^v}{2^w - 3^v}.$$

Proof. Applying first all the $(x/2)$ divisions, referring again to lemma 24, we get $x_0 = (3^v/2^w)x_0 + ((3/2)^v - 1)$ so that $x_0 = 2^{w-v}\frac{3^v-2^v}{2^w-3^v}$. These cycles' solutions diverges rapidly and that general expression is therefore not particular useful to us. But there is no case were all the even steps are running first. The parity vectors needs to be licit which imposes left upwards limit positions of the 0 elements relatively to the elements 1 within the parity vectors as v increases. The first left elements of that resulting limit parity vector are 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, ... Going back to lemma 24, let us recall the equation to the solution x_0 corresponding to all the $(3x+1)/2$ multiplications applied first. We found $x_0(2^w - 3^v) = 3^{v-1}2^0 + 3^{v-2}2^1 + 3^{v-3}2^2 + \ldots + 3^1 2^{v-2} + 3^0 2^{v-1}$. Switching to the equation of the solution x'_0 corresponding to the above limit parity vector, we get $x'_0(2^w - 3^v) = 3^{v-1}2^0 + 3^{v-2}2^1 + 3^{v-3}2^3 + 3^{v-4}2^4 + 3^{v-5}2^6 + 3^{v-6}2^7 + \ldots + 3^0 2^{w-1-\mathrm{or}(1,2)}$, the term $or(1,2)$ depending on the limit parity vector at rank v ending with 1 or 2 zeroes. Each time the element 1 appears in the parity vector, the exponent of the factor 2 increases by 1 while it increases by $1+n$ if there are n elements 0 (thus $n+1 = 1+1 = 2$ for one digit 0). Meanwhile, the exponent of 3 decreases by 1 at each step. For easier understanding, a comparative example is given in the table underneath figuring the right part of the previous equations for $v = 7$. The representation is a little bit different as it consists in attributing a product value 0 for each occurrence of element 0 while offsetting the terms towards the bottom at the same time. The resulting ratio x_0'/x_0 is thus equal to

$$\frac{x_0'}{x_0} = \frac{3^{v-1}2^0 + 3^{v-2}2^1 + 3^{v-3}2^3 + 3^{v-4}2^4 + 3^{v-5}2^6 + ... + 3^0 2^{w-1-or(1,2)}}{3^{v-1}2^0 + 3^{v-2}2^1 + 3^{v-3}2^2 + ... + 3^1 2^{v-2} + 3^0 2^{v-1}}$$

The term $3^{v-1}2^0$ is the greatest term of the numerator on the right side of the equality and by the construction requirement imposed on the limit parity vector, the smallest term is superior to the half value, that is $3^{v-1}2^{-1}$. Figure 33 shows the values of the terms $3^{v-i}2^j$ within $3^{v-1}2^0 + 3^{v-2}2^1 + 3^{v-3}2^3 + 3^{v-4}2^4 + 3^{v-5}2^6 + 3^{v-6}2^7 + ... + 3^0 2^{w-1-or(1,2)}$ after being divided by $\frac{3}{4}3^{v-1}2^0$ for $v = 1$ up to 500. The fractions, according to the previous analysis, vary in the range $]\frac{2}{3},\ \frac{4}{3}]$. Note that this kind of figure is characteristic of those often found when studying the Collatz conjecture. The approximate average value shown in the figure is equal to $(4/3 + 2/3)/2 = 1$ with the hypothesis of a non biased distribution. Therefore, as there are v (non null) terms in the numerator (and denominator) of the expression, we get

$$\frac{x_0'}{x_0} \approx \frac{\frac{3}{4}3^{v-1}2^0 v}{2^{v-1}((\frac{3}{2})^{v-1} + (\frac{3}{2})^{v-2} + ... + +(\frac{3}{2})^0)} = \frac{\frac{1}{4}3^v v}{\frac{2^v}{2}(\frac{(\frac{3}{2})^v - 1}{\frac{3}{2}-1})} = \frac{1}{4}\frac{3^v v}{3^v - 2^v}$$

So that even for some quite small values of v :

$$\frac{x_0'}{x_0} \approx \frac{1}{4}\frac{v}{1-(\frac{2}{3})^v} \approx \frac{v}{4}$$

Hence the said lemma's result. □

In fact, looking carefully at figure 33 showing the repartition of the fractions within the range $]\frac{2}{3},\ \frac{4}{3}]$, the reader will see a little more concentrated population on the bottom part of the figure compared to the upper part. Therefore the average value of the ordinates should be smaller than 1 asymptotically. Figure 34 shows how the average evolves when v varies from 1 to 500. Note also that this offset from 1 is in no way critical to the main aim of this article.

In this case, the real solutions to the cycles are, according to numerical verification, compared to the case of the odd steps running first, in an approximate ratio $\approx 0.2405v$, as shown in figure 35. This asymptotic limit ratio is besides matched quite rapidly (as soon as $v = 10$) in a fairly good way.

Example. $v = 7, w = 12$,
$x'_0/x_0 = 3767/2059 \approx 1.83$ (to be compared to $7/4 = 1.75$).

w	p.v	3^{v-i}	2^j	product	p.v	3^{v-i}	2^j	product
1	1	729	1	729	1	729	1	729
2	1	243	2	486	1	243	2	486
3	1	81	4	324	0	0	0	0
4	1	27	8	216	1	81	8	648
5	1	9	16	144	1	27	16	432
6	1	3	32	96	0	0	0	0
7	1	1	64	64	1	9	64	576
8	0	0	0	0	1	3	128	384
9	0	0	0	0	0	0	0	0
10	0	0	0	0	1	1	512	512
11	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0
total				2059				3767

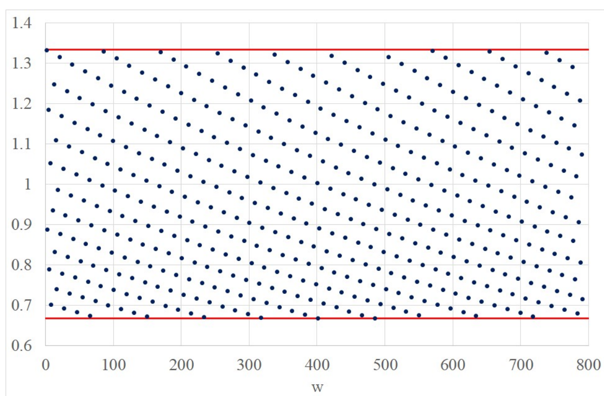

FIGURE 33. Ratios $\frac{3^{v-i}2^j}{\frac{3}{4}3^{v-1}2^0}$.

Lemma 48. *The extrema of the largest rational values, solutions to the smallest element x'_0 of a cycle in $\mathbb{N}^*$, is asymptotically approximately equal to*

$$x'_0 \approx \frac{1}{4ln(2)} v^{2+\chi}, \chi \ll 1.$$

Proof. This is an immediate result of lemmas 46 and 47. $\square$

Lemma 49. *The ratio of any solution to a cycle (in Re^+) to the smallest 2^w-seed is asymptotically tending towards 0 as the rank v tends towards infinity.*

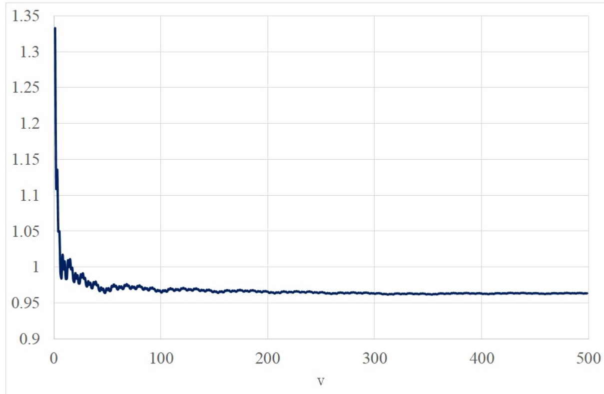

FIGURE 34. Average of ratios up to v (refer to figure 33).

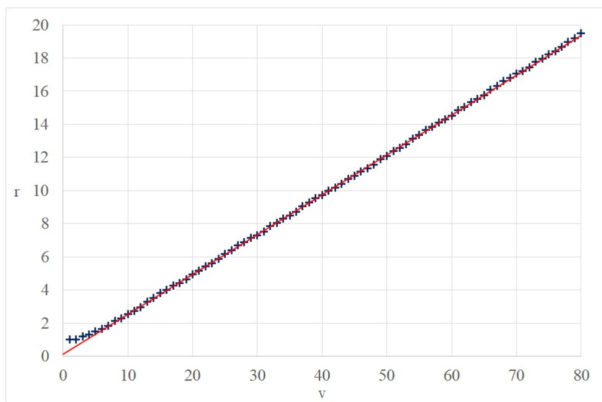

FIGURE 35. Ratio of a largest to the smallest cycles' solutions in Re^+: $r \approx 0.2405\ v + 0.1434$.

Proof. As mentioned earlier, a sample of data of the smallest 2^w-seeds is given in appendix C and figure 20. The ratio to the smallest cycle solution is given in figure 37. The ratio to the largest cycle solution is given in figure 38. This last limit case, which is the one to be considered, exhibits a "way afar distance" to the possibility of a solution as soon as $v \approx 100$ (by more than 3 to 4 decades ratio). This is an obvious consequence of the framing of the 2^w-seeds given in figure 20.
We know that asymptotically the "generating" element of a cycle (its smallest element) are growing polynomially in $v^{2+\chi}$, $\chi \ll 1$, while the smallest 2^w-seeds are expanding exponentially as $\approx 10^4(1.056921^v)$ according to the figure 20 and surpasses $\frac{1}{4ln(2)}v^{2+\chi}$ as soon as rank $v = 40$ even for $\chi = 1$. □

Theorem 7. *The lack of a non-trivial cycle in* $\mathbb{N}^*$ *is very unlikely.*

Proof. According to references [4] and [2], the length of a non-trivial cycle is at least $w = 186265759595$ which is equivalent to $v = 117520609800$. At that

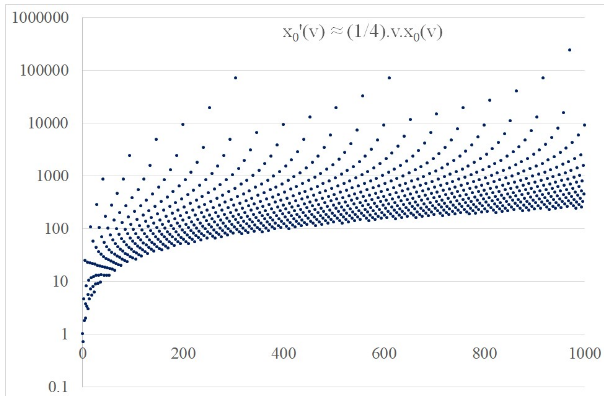

FIGURE 36. Largest solution rational values to the smallest cycles' solutions.

rank, the ratio of the smallest seed to the greatest cycle "generator", using the approximation $10^4(1.056921^v)$ for the first studying entity and $\frac{1}{4ln(2)}v^{2+\chi}$ for the second one, is already surpassing $10^{2.8\times10^9}$. Returning to lemma 36, there is no coherent imaginable common event between the two even if the said smallest and the said greatest might be at a maximum offset (which can hardly be more than a few decades) of their respective asymptotic expected values. In the absence of close up, referring to the figures 19, 28 and 35, the trends are almost straight lines (in linear or logarithmic scales) with totally different slopes and can't have any intersection when v starts to take "large" enough values. Of course, a literal proof of the former empirical assertion would be quite more elegant but is not necessary. A snail can't catch up with a cheetah especially if the double steroidic cheetah took the lead long ago. $\square$

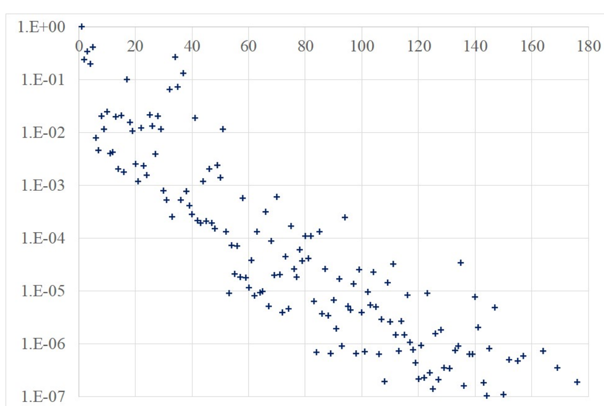

FIGURE 37. Ratio of a largest cycle solution in Re^+ to the smallest 2^w−seed (abscissa v).

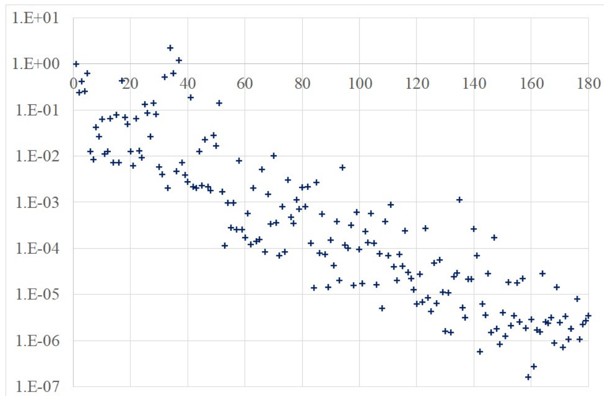

FIGURE 38. Ratio of a smallest cycle solution in Re^+ to the smallest 2^w−seed (abscissa v).

Note. Let us note forget that the seeds' size argument is only a necessarily condition. Even if the seeds would be small enough numbers (which they are not at all asymptotically), it doesn't make them likely candidates to deny the Collatz conjecture.

There is a somewhat simpler way, in which the 2^w−seeds are not involved, to get an equivalent result to the former one. Surprisingly, digging deeper, it enables us to get a rigorous proof of the Collatz conjecture. Let us see that underneath.

Lemma 50. *Let us consider the largest solution rational value of the smallest element of a cycle as provided by the lemmas 39 and 47. The ratios of the greatest common divisor between numerators and denominators to the value of the denominators are tending towards* 0 *as* v *tends towards* ∞.

Proof. Let recall that for both largest and smallest cycle solutions, the denominator is equal to $2^w - 3^v$. This term increase exponentially according to v. Taking the smallest cycle solution, the numerator is $3^v - 2^v$ is not rationally related to $2^w - 3^v$ as $ln(3)/ln(2)$ is in the expression of w. There is no reason to have peculiar common factors between numerator and denominator for some given v value. Therefore as v increases only small factors are expected to be common to both parts of the fraction. The *gcd* is expected to be small (and is often equal to 1) while the denominator is diverging. The reasoning is the same in the case of the largest cycle solution. □

Figure 39 and 40 show the evolution of the former ratios. Appendix D provides also the computer program to get additional data. The *gcd* being often 1 and sharing the same denominator, there is a great number of times where the results are the same for the ratios of the largest and the smallest cycle solutions. It would be likely so also for the intermediate cycle solutions (which all share the same denominator $2^w - 3^v$). The representation of the ordinate being in a logarithmic scale, the curves appear in a linear form. The evolution of the minima of both curves is given by the expression $\frac{1}{2^w-3^v}$. Now $w = \lfloor(ln(3)/ln(2)).v\rfloor +1 < (ln(3)/ln(2)).v + 1$ and therefore $2^w < 2^{(ln(3)/ln(2)).v+1} = 3^{v+1}$. So we get $\frac{1}{2^w-3^v} > \frac{1}{2}\frac{1}{3^v}$, the later expression being a minima of both curves. Of course, rather than the minima, we would like to get the maxima of the said evolution. The figures shows that it is located, using the logarithmic scaling, in a marginally near range of the minima and its trend is also asymptotically approaching the 0 limit.

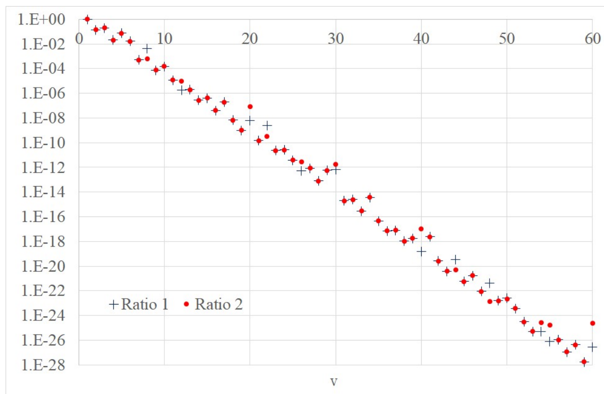

FIGURE 39. Ratio $\frac{gcd(num,den)}{den}$ of smallest (blue cross) and largest (red point) cycle rational solutions' numerators and denominators.

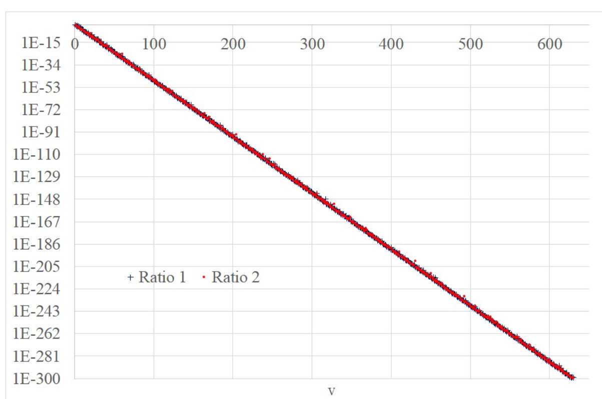

FIGURE 40. Ratio $\frac{gcd(num,den)}{den}$ of smallest and largest cycle rational solutions' numerators and denominators.

If we consider then that the said ratios are some kind of an indicator of the probability to have a natural number cycle solution (as the numerator has to be the exact product of the denominator by a natural number and therefore the ratio is at least equal to 1), we see that the said "probability indicator" is rapidly aiming towards 0. If we go back to the minimal known rank of a non-trivial cycle $v = 117520609800$, the approximative probability, using $\frac{1}{2}\frac{1}{3^v}$ as a basis, would be at that stage already less than $10^{-5.6\times 10^{10}}$.

Theorem 8. *There are no cycle of rank $v > 1$ in $\mathbb{N}^*$.*

Proof. Let us consider, for some given rank $v > 1$, the two limit cases x_0 and x_0' obtained within the domain of the licit parity vectors such as provided by lemma 32.
We get

$$x_0 = \frac{3^{v-1}2^0 + 3^{v-2}2^1 + 3^{v-3}2^2 + ... + 3^1 2^{v-2} + 3^0 2^{v-1}}{2^w - 3^v}$$

and

$$x_0' = \frac{3^{v-1}2^0 + 3^{v-2}2^1 + 3^{v-3}2^3 + 3^{v-4}2^4 + 3^{v-5}2^6 + 3^{v-6}2^7 + ... + 3^0 2^{w-1-\mathrm{or}(1,2)}}{2^w - 3^v},$$

the term $or(1,2)$ depending on the limit parity vector at rank v ending with 1 or 2 zeroes. We intend to study the greatest common divisor between numerator (num) and denominator (den) of these expressions. In order to make the understanding easier let us start with an example. Let us choose $v = 4$, $w = 7$ and one of the corresponding licit parity vector 1110100. Therefore we get $num = 3^3 2^0 + 2^1 3^2 + 2^2 3^1 + 2^4 3^0$ and $den = 2^w - 3^v$. Then

$$\begin{aligned}
gcd(den, num) &= gcd(2^7 - 3^4, 2^0 3^3 + 2^1 3^2 + 2^2 3^1 + 2^4 3^0) && (1)\\
&= gcd(2^7 - 3^4, 3(3^3 2^0 + 2^1 3^2 + 2^2 3^1 + 2^4 3^0))\\
&= gcd(2^7 - 3^4, 3^4 2^0 + 2^1 3^3 + 2^2 3^2 + 2^4 3^1)\\
&= gcd(2^7 - 3^4, 3^4 2^0 + 2^1 3^3 + 2^2 3^2 + 2^4 3^1 + 2^7 - 3^4)\\
&= gcd(2^7 - 3^4, 2^1 3^3 + 2^2 3^2 + 2^4 3^1 + 2^7 3^0)\\
&= gcd(2^7 - 3^4, 3^3 + 2^1 3^2 + 2^3 3^1 + 2^6 3^0) && (2)\\
&= gcd(2^7 - 3^4, 3^3 + 2^2 3^2 + 2^5 3^1 + 2^6 3^0) && (3)\\
&= gcd(2^7 - 3^4, 3^3 + 2^3 3^2 + 2^4 3^1 + 2^5 3^0) && (4)\\
&= gcd(2^7 - 3^4, 3^3 + 2^1 3^2 + 2^2 3^1 + 2^4 3^0) && (1)
\end{aligned}$$

We give the detail from step (1) to step (2), the other ones being entirely similar. In each main steps (1) to (4), the exponents of 3 are unchanged and decrease from $v - 1 = 3$ to 0. The only "challenge" is to handle the exponents of 2. One has to proceed as follows. Start with $(0, 1, 2, 4)$ which is the initial list of the exponents of 2 and add $w = 7$ at the end of the list $(0, 1, 2, 4; 7)$. Shift the list by one to the left and subtract the value of the first item to each number and add $w = 7$ at the end of the list again $(1-1, 2-1, 4-1, 7-1; 7) \equiv (0, 1, 3, 6; 7)$. Continue $(1-1, 3-1, 6-1, 7-1; 7) \equiv (0, 2, 5, 6; 7)$. Again $(2-2, 5-2, 6-2, 7-2; 7) \equiv (0, 3, 4, 5; 7)$ until going back to the initial expression $(3 - 3, 4 - 3, 5 - 3, 7 - 3; 7) \equiv (0, 1, 2, 4; 7)$. Here, during the process, the total number of subtractions is $v = 4$ and add to $w = 1 + 1 + 2 + 3 = 7$. Obviously it is a general pattern. Now, for any value v, if t_0 is some smallest elements of a cycle in $\mathbb{N}$, we will have $den = 2^w - 3^v$

and $num = 3^{v-1}2^0 + 3^{v-2}2^1 + ... + 3^0 2^{w-1-\mathrm{or}(1,2)} = 2^w - 3^v$. Let us have $(0, i_1, i_2, i_3, ..., i_{v-1}; w)$ the initial corresponding list of exponents. Then we inherit of a numerators' list containing v lines which each must correspond to exact multiples of $2^w - 3^v$

$$\begin{gathered}(0, i_1, i_2, i_3, ..., i_{v-2}, i_{v-1}; w)\\ (0, i_2 - i_1, i_3 - i_1, i_4 - i_1, ..., i_{v-1} - i_1, w - i_1; w)\\ (0, i_3 - i_2, i_4 - i_2, i_5 - i_2, ..., w - i_2, w + i_1 - i_2; w)\\ ...\\ (0, w - i_{v-1}, w + i_1 - i_{v-1}, w + i_2 - i_{v-1}, ..., w + i_{v-3} - i_{v-1}, w + i_{v-2} - i_{v-1}; w)\end{gathered}$$

before getting back to the initial

$$(0, i_1, i_2, i_3, ..., i_{v-2}, i_{v-1}; w).$$

The case that each of the corresponding *gcd* be equal to $2^w - 3^v$ bends largely to absurdity. But let us continue anyway.

We can step by step use these v expressions to reduce by (linear) subtractions and multiplications the powers of 3. That construction leads, in a systematic way, to expressions between commas containing each 4 members (if $v > 3$) as 4 terms disappear at each step being equal by pairs (and yes the 2 last ones $2^{w-i_4+i_1} - 2^{w-i_4+i_2}$ underneath emerge always at the end of the process whatever the value of v)

$$3 + 2^{i_{v-1}-1} - 2^{i_{v-2}-1} + 2^{w-i_4+i_1} - 2^{w-i_4+i_2}$$

and therefore to the following greatest common divisor to be studied

$$gcd(2^w - 3^v, 3 + 2^{i_{v-2}-1}(2^{i_{v-1}-i_{v-2}} - 1) - 2^{w-i_4+i_1}(2^{i_2-i_1} - 1)).$$

Although interesting by its simplicity, the second component of this *gcd* is not systematically smaller than $2^w - 3^v$, an event which would constituted a lucky first condition to prove our aim. Therefore, let us focus instead only on the two first items of the previous list:

$$\begin{gathered}(0, i_1, i_2, i_3, ..., i_{v-1}; w)\\ (0, i_2 - i_1, i_3 - i_1, i_4 - i_1, ..., w - i_1; w).\end{gathered}$$

As we mentioned, in order to get a Collatz' cycle, the first item corresponds to a value that equals exactly $2^w - 3^v$ while the second one must then correspond to an exact non-zero multiple of the same value. So let us have the two corresponding terms:

$$\begin{gathered}3^{v-1}2^0 + 3^{v-2}2^{i_1} + 3^{v-3}2^{i_2} + 3^{v-4}2^{i_3} + ... + 3^1 2^{i_{v-2}} + 3^0 2^{i_{v-1}}\\ 3^{v-1}2^0 + 3^{v-2}2^{i_2-i_1} + 3^{v-3}2^{i_3-i_1} + 3^{v-4}2^{i_4-i_1} + ...3^1 2^{i_{v-1}-i_1} + 3^0 2^{w-i_1}\end{gathered}$$

The ratio of the second term to the first one is

$$\begin{aligned} r &= \frac{3^{v-1}2^0 + 3^{v-2}2^{i_2-i_1} + 3^{v-3}2^{i_3-i_1} + 3^{v-4}2^{i_4-i_1} + ...3^1 2^{i_{v-1}-i_1} + 3^0 2^{w-i_1}}{3^{v-1}2^0 + 3^{v-2}2^{i_1} + 3^{v-3}2^{i_2} + 3^{v-4}2^{i_3} + ... + 3^1 2^{i_{v-2}} + 3^0 2^{i_{v-1}}}\\ &= \frac{1 + \frac{2^{i_2-i_1}}{3^1} + \frac{2^{i_3-i_1}}{3^2} + ... + \frac{2^{i_{v-1}-i_1}}{3^{v-2}} + \frac{2^{w-i_1}}{3^{v-1}}}{1 + \frac{2^{i_1}}{3^1} + \frac{2^{i_2}}{3^2} + \frac{2^{i_3}}{3^3} + ... + \frac{2^{i_{v-2}}}{3^{v-2}} + \frac{2^{i_{v-1}}}{3^{v-1}}}\end{aligned}$$

Now the exponents are issued from a parity vector and therefore we have systematically $i_1 = 1$. So that

$$\begin{array}{rcl} r & = & \frac{1+\frac{2^{i_2-1}}{3^1}+\frac{2^{i_3-1}}{3^2}+...+\frac{2^{i_{v-1}-1}}{3^{v-2}}+\frac{2^{w-1}}{3^{v-1}}}{1+\frac{2^{i_1}}{3^1}+\frac{2^{i_2}}{3^2}+\frac{2^{i_3}}{3^3}+...+\frac{2^{i_{v-2}}}{3^{v-2}}+\frac{2^{i_{v-1}}}{3^{v-1}}} \\ & = & \frac{1+\frac{2^{i_1-1}}{3^0}+\frac{2^{i_2-1}}{3^1}+\frac{2^{i_3-1}}{3^2}+...+\frac{2^{i_{v-1}-1}}{3^{v-2}}+(\frac{2^{w-1}}{3^{v-1}}-1)}{1+\frac{2^{i_1}}{3^1}+\frac{2^{i_2}}{3^2}+\frac{2^{i_3}}{3^3}+...+\frac{2^{i_{v-2}}}{3^{v-2}}+\frac{2^{i_{v-1}}}{3^{v-1}}}. \end{array}$$

Let us have

$$c_0 = \frac{2^{i_1}}{3^1} + \frac{2^{i_2}}{3^2} + \frac{2^{i_3}}{3^3} + ... + \frac{2^{i_{v-2}}}{3^{v-2}} + \frac{2^{i_{v-1}}}{3^{v-1}} \quad \text{and} \quad g = \frac{2^{w-1}}{3^{v-1}} - 1.$$

Then

$$r \quad = \quad \frac{1+3c_0/2+g}{1+c_0}.$$

Therefore the conditions underneath, if one of them is true, are equivalent

$$\begin{array}{rcl} 3/2 < r < 2 & \Leftrightarrow & \{g > 1/2 \text{ and } c_0 > 2(g-1)\} \\ & \Leftrightarrow & \{\frac{2^w}{3^v} > 1 \text{ and } c_0 > 3(\frac{2^w}{3^v} - \frac{4}{3})\}. \end{array}$$

Let us observe that starting with a lower and upper bound on r, we end with only a lower bound on c_0. This is a main simplifying event as we will see very soon. Meanwhile, let us examine the two limit cases.
For $i_1 = 1$, $i_2 = 2$, $i_3 = 3$, ..., $i_{v-1} = v-1$, we get

$$\begin{array}{rcl} c_0 & = & 1 + \frac{2^1}{3^1} + \frac{2^2}{3^2} + \frac{2^3}{3^3} + ... + \frac{2^{v-2}}{3^{v-2}} + \frac{2^{v-1}}{3^{v-1}} - 1 \\ & = & \frac{1-\frac{2^v}{3^v}}{1-\frac{2}{3}} - 1 \\ & = & 3(\frac{2}{3} - \frac{2^v}{3^v}) \end{array}$$

Now

$$\begin{array}{lcl} w = \lfloor \frac{ln(3)}{ln(2)} v \rfloor + 1 & \Rightarrow & \frac{2^w}{3^v} > 1 \\ \frac{2^{w-1}+2^{v-1}}{3^v} < 1 & \Rightarrow & 3(\frac{2}{3} - \frac{2^v}{3^v}) > 3(\frac{2^w}{3^v} - \frac{4}{3}) \end{array}$$

The first implication results from lemma 30. The second implication, an equivalence in fact, results from lemma 43. Therefore the term r cannot be an integer which proves the first limit case. Now for the second case (and all intermediary ones) the term

$$c_1 = \frac{2^{i_1}}{3^1} + \frac{2^{i_2}}{3^2} + \frac{2^{i_3}}{3^3} + ... + \frac{2^{i_{v-2}}}{3^{v-2}} + \frac{2^{i_{v-1}}}{3^{v-1}}$$

is necessary superior, as $i_k \geq k$ for all $k \in 1, 2, ..., v-1$, to the term

$$c_0 \quad = \quad \frac{2^1}{3^1} + \frac{2^2}{3^2} + \frac{2^3}{3^3} + ... + \frac{2^{v-2}}{3^{v-2}} + \frac{2^{v-1}}{3^{v-1}}$$

Hence $c_1 > c_0 > 3(\frac{2^w}{3^v} - \frac{4}{3})$ proving the second case and all intermediary cases. $\square$

Figure 41 shows a sample of the value of $r-1$. It shows the constraint $r \to \frac{3}{2}^+$ in the case of the second limit case which is a consequence of $\frac{1}{2} < \frac{2^{w-1}+2^{v-1}}{3^v}$ (see lemma 43). Note also the common value $r-1=\frac{6}{5}$ for $v=2$ in the first and second limit cases.

Appendix G provides a computer program to evaluate the ratio $r-1$ for some initial choice x_0.

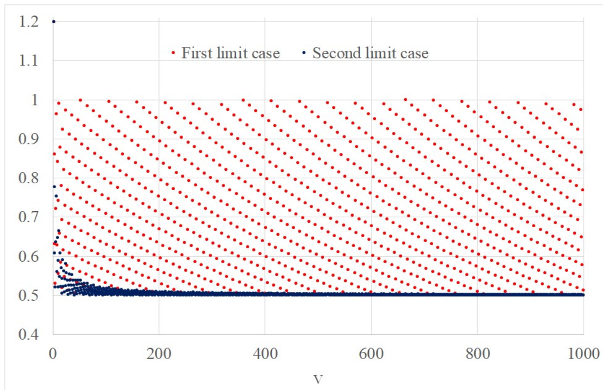

FIGURE 41. Ratio $r-1$.

Theorem 9. *The Collatz conjecture is true.*

Proof. This is an immediate consequence of the two theorems 1 and 8. □

10. THE ADMISSIBLE RATIONAL SOLUTIONS

We spend a long time looking on one kind of possible solutions for the Collatz cycles, the integers. Let us broaden here the perspective and have a glance on the "solutions" to the altitude flight time (thus including the possible cycles). Let us have some initial integer t_0. We are looking for the first event i such that $x_i = t_i/t_0 <= 1$, t_i being the iterated result of the Collatz algorithm, and $gcd(t_i, t_0) = 1$. Then if $t_0 > 1$ and $v > 0$ (at least one odd step is occurring in the process),

$$\frac{t_i}{t_0} = \lfloor \frac{t_i}{t_0} \rfloor + \frac{1}{2}(1 + \frac{k_1}{k_2}), k_1 = 1 \bmod 2, k_2 = 1 \bmod 2, 0 \le k_1 < k_2.$$

This, if proven so, excludes any integer solutions (but not only) and explains why there is no such solutions to the Collatz conjecture. Figure 42 shows graphically the corresponding result $\frac{t_i}{t_0} - \lfloor \frac{t_i}{t_0} \rfloor - \frac{1}{2} = \frac{k_1}{2k_2}$ for $t_0 = 3, 5, 7, ..., 2^{15} - 1$ classified by increasing values. Let us note that the plateau solutions are not composed of equal values but mostly of small distinct values.

These plateaus are explained easily as they originate mainly out of equal odd (and even) steps' solutions.

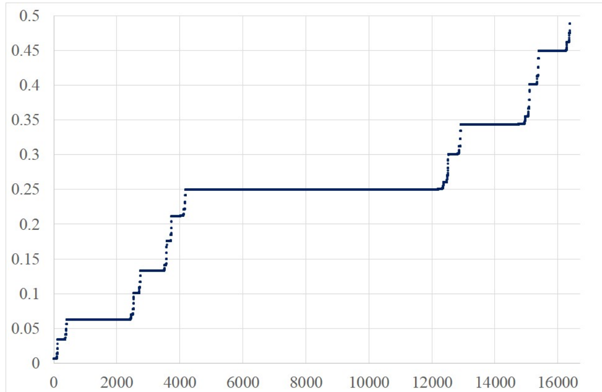

FIGURE 42. Ratio $\frac{k_1}{2k_2}$.

Lemma 51. *Let us consider some seed t_0 and its result t_i by the Collatz algorithm. Then, for increasing $k \in \mathbb{N}$, the ratio $\frac{t_i+k3^v}{t_0+k2^w}$, with the usual definition of (v,w) adopted in this article, is strictly decreasing from $\frac{t_i}{t_0}$ to $\frac{3^v}{2^w}$.*

Proof. By the property of the underlying hyperbolic function in k, the ratio $\frac{t_i+k3^v}{t_0+k2^w} = \frac{t_i/k+3^v}{t_0/k+2^w}$, as seen in previous lemmas, is strictly monotonous from $\frac{t_i}{t_0}$ to $\frac{3^v}{2^w}$. Now, in the Collatz algorithm, the multiplication is $3x+1$ instead of $3x$ while the division is exactly $x/2$. Therefore $\frac{t_i}{t_0} > \frac{3^v}{2^w}$ and therefore the ratio $\frac{t_i/k+3^v}{t_0/k+2^w}$ decreases systematically. As k increases, $\frac{t_0}{k}$ and $\frac{t_i}{k}$ tends towards 0, hence the rapid plateau's effect as $t_i < t_0 < 2^w$. $\square$

Lemma 52. *For fixed v, any ratio t_i/t_0 admits the strict lower bound $\frac{3^v}{2^w}$.*

Proof. This is another way to express lemma 51. $\square$

The bottom of the plateaus are given by the ratios $3/4 - 1/2 = 0.25$, $3^2/2^4 - 1/2 = 0.0625$, $3^3/2^5 - 1/2 = 0.34375$, $3^4/2^7 - 1/2 = 0.1328125$, $3^5/2^8 - 1/2 = 0.44921875$, $3^6/2^{10} - 1/2 \approx 0.21191406$, $3^7/2^{12} - 1/2 \approx 0.03393555$, $3^8/2^{13} - 1/2 \approx 0.30090332$, $3^9/2^{15} - 1/2 \approx 0.10067749$, $3^{10}/2^{16} - 1/2 = 0.401016235$ and so on. These plateaus are getting shorter as v increases. The best approximations of $ln(2)/ln(3)$ by $\frac{3^v}{2^{w-or(1,0)}}$ provide respectively, of course, the plateaus near 0 and those near 1/2.

v	w	$\frac{3^v}{2^w} - 1/2$		v	w	$1/2 - (\frac{3^v}{2^w} - 1/2)$
1	2	0.25		1	2	0.25
2	4	0.0625		5	8	0.05078125
12	20	$6.82163E-3$		41	65	$1.139745E-2$
53	85	$1.04516E-3$		306	485	$1.02172E-3$
665	1055	$2.18275E-5$		15601	24727	$1.81944E-5$
31867	50509	$3.63248E-6$		79335	125743	$3.66471E-6$
111202	176252	$1.80011E-6$		190537	301994	$6.45075E-8$
10590737	16785922	$2.61503E-8$		10781274	17087915	$1.22070E-8$
53715833	85137582	$1.73629E-9$		171928773	272500658	$1.78921E-9$
225644606	357638240	$8.41686E-10$				

TABLE 3

We conclude immediately, for $v > 0$,

$$\frac{t_i}{t_0} - \lfloor \frac{t_i}{t_0} \rfloor - \frac{1}{2} > 0.$$

The Collatz conjecture, for $t_0 > 0$, is nothing more than the same lemma on the upper bound.

$$\frac{t_i}{t_0} - \lfloor \frac{t_i}{t_0} \rfloor - \frac{1}{2} < \frac{1}{2}.$$

Appendix A. Trees' growth programming

Running the Collatz inverse algorithm without special care leads to duplicate graphs if the initial integer is chosen within a root. Let us start for example with integer 1 within the root $(1,2,4)$. We have, for the antecedents of x, one or two values respectively $(2x)$ or $(\frac{x-1}{3}, 2x)$, therefore here step by step upwards $(0,2)$, $(0,4)$, $(0,1,8)$, $(0,0,2,16)$, $(0,0,4,5,32)$, $(0,0,1,8,10,64)$, $(0,0,0,2,3,16,20,21,128)$, $(0,0,0,4,5,6,32,40,42,256)$, $(0,0,0,1,8,10,12,13,64,80,84,85,512)$, $\ldots$, compiling after only a few algorithmic steps already three graphs stemming from 1 at different stages of blossom (with an additional "0 graph" next step result of each integer 1 encounter) instead of $(\cancel{2})$, $(\cancel{4})$, (8), (16), $(5,32)$, $(10,64)$, $(3,20,21,128)$, $(6,40,42,256)$, $(12,13,80,84,85,512)$ starting beyond the root.

The following computer program give the cardinals at the successive upwards ranks of the total number of vertices and the number of vertices equal to i modulo v, $i = 0$ to $v-1$. It is executed without any particular constraint except it refuses even antecedent $\frac{x-1}{3}$ which is an imperative requirement. It therefore doesn't generate any antecedent equal to 0 but won't be able to avoid the unwanted additional integer 1 or any root member of some other cycle creating graphs' redundancies. Using it, the reader is advised to take the remark into account by preferably choosing an integer outside a cycle even if the asymptotic proportions of the cardinals will ultimately remain the same.

PARI/GP progamming code.

```
{pr = -67; \\initial integer choice
v = 3; \\modulo value choice
nm = vector(v);
k1 = k2 = vector(1000000); m2 = 1; k2[m2] = pr; for(x = 1, v, nm[x] = 0);
for(x = 1, v, if(k2[m2] % v == x-1, nm[x] = 1; break));
print(0" "m2" "nm);
for(x = 1, 48,\\last upwards rank choice
m1 = 0; for(y = 1, v, nm[y] = 0);
for(m = 1, m2, m1 = m1+1; k1[m1] = 2*k2[m];
for(y = 1, v, if(k1[m1] % v == y-1, nm[y] = nm[y]+1));
t = (k2[m]-1)/3; tr = truncate(t); r = t-tr;
if(r == 0, if(tr%2 == 1, m1 = m1+1; k1[m1] = t;
for(z = 1, v, if(k1[m1] % v == z-1, nm[z] = nm[z]+1)))));
print(x" "m1" "nm); m2 = m1; k2 = k1)}
```

Choosing for example as starting integer -67, the numerical data is, at upwards rank 48, a total of 78932 vertices with repartitions as follows:

v	3	5	7	9	11	13	15
0 mod v	26360	15788	11079	8744	7074	6099	5280
1 mod v	26257	15736	11373	8658	7210	6147	5164
2 mod v	26315	15786	11165	8738	7126	6100	5212
3 mod v		15773	11524	8744	7054	5962	5228
4 mod v		15849	10993	8772	7138	6067	5299
5 mod v			11349	8759	7122	6188	5282
6 mod v			11449	8872	7191	6140	5293
7 mod v				8827	7178	6127	5304
8 mod v				8818	7342	6096	5281
9 mod v					7242	6113	5289
10 mod v					7255	5922	5226
11 mod v						5902	5279
12 mod v						6069	5270
13 mod v							5264
14 mod v							5261

Expressing the offsets in percent, we get:

v	3	5	7	9	11	13	15
0 mod v	$0,19\%$	$0,01\%$	$-1,75\%$	$-0,30\%$	$-1,42\%$	$0,45\%$	$0,34\%$
1 mod v	$-0,20\%$	$-0,32\%$	$0,86\%$	$-1,28\%$	$0,48\%$	$1,24\%$	$-1,86\%$
2 mod v	$0,02\%$	$0,00\%$	$-0,98\%$	$-0,37\%$	$-0,69\%$	$0,47\%$	$-0,95\%$
3 mod v		$-0,08\%$	$2,20\%$	$-0,30\%$	$-1,70\%$	$-1,81\%$	$-0,65\%$
4 mod v		$0,40\%$	$-2,51\%$	$0,02\%$	$-0,52\%$	$-0,08\%$	$0,70\%$
5 mod v			$0,65\%$	$-0,13\%$	$-0,75\%$	$1,92\%$	$0,38\%$
6 mod v			$1,53\%$	$1,16\%$	$0,21\%$	$1,13\%$	$0,59\%$
7 mod v				$0,65\%$	$0,03\%$	$0,91\%$	$0,80\%$
8 mod v				$0,54\%$	$2,32\%$	$0,40\%$	$0,36\%$
9 mod v					$0,92\%$	$0,68\%$	$0,51\%$
10 mod v					$1,11\%$	$-2,47\%$	$-0,69\%$
11 mod v						$-2,79\%$	$0,32\%$
12 mod v						$-0,04\%$	$0,15\%$
13 mod v							$0,04\%$
14 mod v							$-0,02\%$
e.m.a.d/p*	$0,20\%$	$0,26\%$	$1,76\%$	$0,71\%$	$1,18\%$	$1,46\%$	$0,73\%$
$2\frac{1}{\sqrt{\pi}\sqrt{n}}$	$0,70\%$	$0,90\%$	$1,06\%$	$1,20\%$	$1,33\%$	$1,45\%$	$1,56\%$

* Effective mean absolute differences compared to populations' ratio

Appendix B. Equal stopping time seeds enumeration

The following program provides at rank v, within the interval $[0, 2^w[$, the cardinals $\#s_w(v)$, $\#plf(v)$ and $\#plr(v)$, that is respectively the number of integers with stopping times equal to w, smaller than w and greater than w. Note that sometimes the exponentiation sign ˆ won't copy successfully and has to be retyped manually (3 corrections in that case here).

PARI/GP progamming code.

```
{rank = 30; \\initial integer choice
dw = 1; da1 = da2 = vector(rank); da1[1] = 0; da1[2] = 1; da2[1] = 0;
print("v " "#sw(v) " "#plf(v) " "#plr(v) " "2 ^ w");
print(0" "1" "0" "1" "2); print(1" "1" "2" "1" "4); plf = 2+1;
for(v = 2, rank, pla =0; w = (log(3)/log(2)*v)\1+1;
wp = (log(3)/log(2)*(v-1))\1+1; wpp = (log(3)/log(2)*(v-2))\1+1;
ew = wp - (v-1); dw = w - wp; ddw = wp - wpp;
plf = (2 ^ dw)* plf;
for(k = 1, ew,
da2[k+1] = da2[k] + da1[k+1]; pla = da2[k+1]+pla);
if(ddw == 2, da2[ew+1] = da2[ew]; pla = da2[ew+2]+pla);
plf1 = plf+vecsum(da2); plr = 2 ^ w - plf1;
print(v" "pla" "plf" "plr" "pla+plf+plr); plf = plf1; da1 = da2)}
```

v	$\#s_w$	$\#plf(v)$	$\#plr(v)$	2^w
0	1	0	1	2
1	1	2	1	4
2	1	12	3	16
3	2	26	4	32
4	3	112	13	128
5	7	230	19	256
6	12	948	64	1024
7	30	3840	226	4096
8	85	7740	367	8192
9	173	31300	1295	32768
10	476	62946	2114	65536
11	961	253688	7495	262144
12	2652	1018596	27328	1048576
13	8045	2042496	46611	2097152
14	17637	8202164	168807	8388608
15	51033	16439602	286581	16777216
16	108950	65962540	1037374	67108864
17	312455	132142980	1762293	134217728
18	663535	529821740	6385637	536870912
19	1900470	2121941100	23642078	2147483648
20	5936673	4247683140	41347483	4294967296
21	13472296	17014479252	151917636	17179869184
22	39993895	34055903096	263841377	34359738368
23	87986917	136383587964	967378591	137438953472
24	257978502	545886299524	3611535862	549755813888
25	820236724	1092288556052	6402835000	1099511627776
26	1899474678	4372435171104	23711865322	4398046511104
27	5723030586	8748669291564	41700700058	8796093022208
28	12809477536	35017569288600	153993322696	35184372088832
29	38036848410	70060757532272	269949796982	70368744177664
30	84141805077	280395177522728	995657382851	281474976710656

Appendix C. Smallest stopping time seeds

The following data provides the smallest number of the 2^w−class of rank v (stopping time $w+v$). The empty spots corresponds to unknown values but greater than 400 000 000.

v	$sm(2^w)$	v	$sm(2^w)$	v	$sm(2^w)$	v	$sm(2^w)$	v	$sm(2^w)$	v	$sm(2^w)$
0	2	40	6887	80	60975	120	6255855	160	40814363	200	371871359
1	1	41	4591	81	45127	121	6492187	161	187375615	201	247914239
2	3	42	13439	82	393967	122	7849755	162	131801135	202	165276159
3	11	43	6383	83	423679	123	3137471	163	44186399	203	
4	7	44	4255	84	1759951	124	9294427	164	29457599	204	293824283
5	39	45	7963	85	35655	125	8484287	165	39276799	205	195882855
6	287	46	7527	86	434223	126	2788863	166	19638399	206	
7	231	47	12399	87	495687	127	7499935	167	53271551	207	348236187
8	191	48	7279	88	665215	128	6079559	168	71028735	208	
9	127	49	1583	89	1643759	129	6204543	169	27209575	209	
10	359	50	1055	90	528895	130	20808639	170	35514367	210	127456255
11	511	51	703	91	730559	131	9941863	171	60112511	211	245235559
12	239	52	15039	92	437247	132	29256191	172	40075007	212	
13	159	53	111259	93	2162111	133	8837211	173	53433343	213	217987163
14	639	54	41407	94	432923	134	2091647	174	143061311	214	290649551
15	283	55	62079	95	565247	135	1394431	175		215	193766367
16	991	56	77031	96	288615	136	17392879	176	162612223	216	145324775
17	251	57	94959	97	376831	137	13002751	177	107295983	217	96883183
18	167	58	34239	98	2548479	138	7460635	178	22649071	218	
19	111	59	138751	99	611455	139	2533535	179	71530655	219	
20	1695	60	99007	100	608111	140	1689023	180	20132507	220	
21	1307	61	106239	101	1585403	141	1126015	181	13421671	221	
22	871	62	187327	102	405407	142	64993051	182		222	326610023
23	927	63	69375	103	270271	143	19925503	183	279200511	223	
24	671	64	226767	104	362343	144	13774695	184	20638335	224	
25	155	65	104303	105	401151	145	9280639	185	272473947	225	
26	103	66	10087	106	1563647	146	46043247	186		226	
27	1639	67	256511	107	1042431	147	28290175	187		227	
28	91	68	67583	108	6721703	148	57330463	188	26716671	228	272896031
29	3431	69	90111	109	381727	149	54870655	189	144091295	229	181930687
30	3399	70	45055	110	667375	150	46355695	190	192121727	230	
31	2287	71	126575	111	626331	151	48773915	191	96060863	231	
32	71	72	299259	112	1691807	152	32515943	192	64040575	232	95592191
33	6395	73	96383	113	1564063	153	41946879	193	340208287	233	
34	47	74	336199	114	1541147	154	12132095	194	56924955	234	
35	31	75	64255	115	1027431	155	8088063	195		235	
36	2047	76	84383	116	1127871	156	21677295	196		236	
37	27	77	57115	117	1991615	157	14378779	197		237	63728127
38	1819	78	56255	118	1327743	158	41942559	198	383606875		
39	17691	79	37503	119	7303711	159	241682847	199			

Appendix D. Greatest common divisor linked to cycles

The following program provides the ratio of the smallest cycle solutions *gcd* to the denominator $2^w - 3^v$.
Note that sometimes the exponentiation sign ^ won't copy successfully and has to be retyped manually (5 corrections in that case).

PARI/GP progamming code.

```
{for(v = 1, 20, w = (log(3)/log(2)*v)\1+1;
n_x0 = 3 ^ v-2 ^ v;
d_x0 = 2 ^ w-3 ^ v;
gcdd = gcd(d_x0, n_x0);
x0 = n_x0/d_x0+0.0;
printf("%s %i %s %i %s %.3f %s %i %s %i %s %i",
"v "v" w "w" x0 "x0" denominator "d_x0" numerator "n_x0" gcd "gcdd);
print()) }
```

The following program provides the ratio of the largest cycle solutions *gcd* to the denominator $2^w - 3^v$.
Note that sometimes the exponentiation sign ^ won't copy successfully and has to be retyped manually (5 corrections in that case).

```
{w = 1; v = 0; d_x0 = 1; n_x0 = 2; tot = 2;
x0 = n_x0/d_x0+0.0; gcdd = 1;
printf("%s %i %s %i %s %.3f %s %i %s %i %s %i",
"v "v" w "w" x0 "x0" denominator "d_x0" numerator "n_x0" gcd "gcdd);
print();
for(w = 2, 20,
v = 1+((w-1)*log(2)/log(3))\1;
ww = (log(3)/log(2)*v)\1+1;
if(w == ww,
tot = tot+(3 ^ (-v))*(2 ^ w);
n_x0 = tot*(3 ^ (v))/2;
www = (log(3)/log(2)*(v+1))\1+1;
d_x0 = 2 ^ www-3 ^ (v+1);
gcdd = gcd(d_x0, n_x0);
x0 = n_x0/d_x0+0.0;
printf("%s %i %s %i %s %.3f %s %i %s %i %s %i",
"v "v" w "w" x0 "x0" denominator "d_x0" numerator "n_x0" gcd "gcdd);
print()))}
```

Appendix E. Continued fraction of Ln(3)/Ln(2)

The following program provides the continued fraction of $\frac{ln(2)}{ln(3)}$ and the corresponding successive fractions' approximations. For $\frac{ln(3)}{ln(2)}$, the continued fraction starts with [1; 1, 1, 2, ...] instead of [0; 1, 1, 1, 2, ...].

PARI/GP progamming code.

```
\p 5000
{nb = 20;
vct = vector(nb); x = log(2)/log(3);
for(i = 1, nb, vct[i] = x\1;
y = 1/(x-vct[i]); x = y);
print(vct);
for(i = 2, nb, t = i; x = vct[t];
for(j = 2, i, k = t-j+1; x = vct[k]+1/x);
print(numerator(x)" "denominator(x)))}
```

The following program provides an alternative way to get the continued fraction of $\frac{ln(2)}{ln(3)}$ and the corresponding successive fractions' approximations. It may be analogous to the Terence Jackson / Keith Matthews algorithm in the article "On Shanks' algorithm for computing the continued fraction of $\log_b$(a)" (see reference [3]).

PARI/GP progamming code.

```
{infty = 10000000000;
e1 = 1; e2 = 0;
v1 = 0; w1 = 1; v2 = 1; w2 = 1;
printtex(cf" "num" "den);
print(0" "v1" "w1);
print(1" "v2" "w2);
for(n = 1, 50, v3 = v1+v2;
for(cf = 1, infty, m = cf;
v3 = v3+v2;
w3 = (log(3)/log(2)*v3)\1+1-e2;
r1 = (w3-w1)/(v3-v1);
r2 = w2/v2;
if(r1 == r2, ,v3 = v3-v2; break));
w3 = (log(3)/log(2)*v3)\1+1-e2;
print(m" "v3" "w3);
e1 = 1-e1; e2 = 1-e2;
v1 = v2; v2 = v3;
w1 = (log(3)/log(2)*v1)\1+e1;
w2 = (log(3)/log(2)*v2)\1+e2)}
```

The continued fraction starts with the following coefficients.
[0; 1, 1, 1, 2, 2, 3, 1, 5, 2, 23, 2, 2, 1, 1, 55, 1, 4, 3, 1, 1, 15, 1, 9, 2, 5, 7, 1, 1, 4, 8, 1, 11, 1, 20, 2, 1, 10, 1, 4, 1, 1, 1, 1, 1, 37, 4, 55, 1, 1, 49, 1, 1, 1, 4, 1, 3, 2, 3, 3, 1, 5, 16, 2, 3, 1, 1, 1, 1, 1, 5, 2, 1, 2, 8, 7, 1, 1, 2, 1, 1, 3, 3, 1, 1, 1, 1, 5, 4, 2, 2, 2, 16, 8, 10, 1, 25, 2, 1, 1, 1, 2, 18, 10, 1, 1, 1, 1, 9, 1, 5, 6, 2, 1, 1, 12, 1, 1, 1, 6, 2, 12, 1, 1, 12, 1, 1, 2, 12, 1, 12, 3, 1, 5, 1, 14, 1, 1, 14, 2, 3, 1, 2, 2, 1, 4, 1, 4, 8, 1, 1, 1, 3, 5, 1, 1, 1, 1, 2, 1, 4, 3, 7, 5, 3, 1, 32, 1, 1, 1, 18, 1, 3, 2, 5, 2, 1, 3, 1, 8, 1, 1, 1, 2, 6, 6, 5, 33, 2, 2, 3, 1, 1, 1, 1, 29, 1, 3, 2, 1, 21, 1, 6, 52, 1, 8, 1, 4, 14, 9, 7, 1, 4, 18, 2, 2, 1, 1, 2, 100, 39, 1, 2, 1, 1, 19, 1, 5, 9, 1, 3, 964, 5, 1, 1, 1, 39, 1, 1, 1, 1, 5, 3, 1, 88, 1, 2, 1, 3, 1, 11, 1, 23, 11, 1, 1, 1, 2, 1, 1, 4, 3, 1, 5, 1, 4, 2, 1, 75, 1, 2, 1, 11, 17, 2, 5, 3, 1, 3, 34, 1, 10, 2, 4, 7, 1, 1, 23, 1, 6, 3, 1, 7, 1, 17, 2, 1, 24, 1, 1, 1, 10, 1, 4, 1, 1, 5, 3, 2, 1, 2, 1, 1, 3, 6, 8, 1, 8, 2, 1, 1, 4, 2, 7, 9, 2, 2, 2, 1, 7, 12, 2436, 1, 2, 1, 9, 10, 1, 5, 1, 3, 1, 2, 1, 2, 3, 1, 1, 3, 1, 4, 6, 1, 2, 1, 2, 2, 1, 2, 1, 1, 3, 46, 31, 196, 4, 1, 1, 3, 11, 1, 3, 14, 1, 1, 3, 2, 20, 1, 3, 6, 3, 85, 1, 7, 1, 9, 4, 5, 2, 1, 1, 78, 1, 4, 4, 2, 6, 6, 2, 4, 8, 4, 5, 1, 1, 11, 1, 2, 1, 5, 13, 2, 1, 3, 4, 2, 7, 5, 2, 2, 1, 2, 10, 1, 163, 1, 3, 1, 1, 1, 2, 1, 1, 2, 1, 6, 30, 1, 2, 2, 13, 1, 1, 2, 1, 2, 1, 1, 1, 3, 2, 5, 1, 5, 3, 1, 3, 1, 3, 2, 36, 1, 1, 1, 1, 9, 7, 1, 28, 2, 1, 1, 5, 1, 11, 10, 3, 1, 2, 1, 1, 2, 19, 2, 5, 5, 1, 4, 1, 1, 2, 1, 5, 3, 10, ...].

Besides, a small sample of the corresponding successive fractions' approximations is :
0/1, 1/1, 1/2, 2/3, 5/8, 12/19, 41/65, 53/84, 306/485, 665/1054, 15601/24727, 31867/50508, 79335/125743, 111202/176251, 190537/301994, 10590737/16785921, 10781274/17087915, 53715833/85137581, ...

Appendix F. Evaluation of the cycles' "generators"

The following program provides the evaluation of the ratio $x_0 = (3^v - 2^v)/(2^{wad} - 3^v)$ for very large value of v. It is designed to be able to get such high value for the best rational approximation of $\frac{ln(2)}{ln(3)}$. These value of v and wad are the numerator and denominator obtained in the previous appendix and we have to notice that $wad = or(w, w-1)$ which has to be correctly chosen underneath (or simply use the value on the denominator for w given by the previous computer program). Over a certain rank, the exponents get so high that straight calculations are out of reach and the exponents have to be addressed separately which is done here.

Note again that sometimes the exponentiation sign ^ won't copy successfully and has to be retyped manually (7 corrections in that case).

PARI/GP progamming code.

```
\p 1000
{v = 15601; zero_one = 0; \ \ initial choice
w = (log(3)/log(2)*v)\1+ 1 - zero_one;
if(v < 1000, z = (3 ^ v - 2 ^ v)/(2 ^ w - 3 ^ v)+0.0,
tot = 0; ajt = vev = vector(1);
nv = v;
if(nv/2 == nv\2, vev[1] = 0, vev[1] = 1); m = 1;
nv = (nv/2)\1;
for(i = 1, 1000,
if(nv/2 == nv\2, ajt[1] = 0, ajt[1] = 1);
vev = concat(ajt[1],vev); m = m+1; nv = (nv/2)\1;
if(nv < 2, ajt[1] = nv; vev = concat(ajt[1],vev); m = m+1; break));
for(i = 1, m, j = m-i; tot = tot+(2 ^ j)*vev[i]);
prodd = 1.5; prodp = vector(2); exposa = vector(2); ajtp = vector(1);
prodp[2] = 1.0; exposa[2] = 0; prodp[1] = 1.5; exposa[1] = 0; exposd = 0;
for(i = 2, m, prodd = prodd*prodd; exposd = exposd*2;
for(i = 1, 10, if(prodd > 10, prodd= prodd/10; exposd = exposd+1, break));
ajtp[1] = prodd; prodp = concat(ajtp[1],prodp);
ajtp[1] = exposd; exposa = concat(ajtp[1],exposa));
prodd = 1; exposd = 0;
for(i = 1, m, if(vev[i] == 1, prodd = prodd*prodp[i]; exposd = exposd+exposa[i]));
mem1 = prodd; mem2 = exposd;
tot = 0; ajt = vev = vector(1);
nv = w-v;
if(nv/2 == nv\2, vev[1] = 0, vev[1] = 1); m = 1;
nv = (nv/2)\1;
for(i = 1, 1000,
if(nv/2 == nv\2, ajt[1] = 0, ajt[1] = 1);
vev = concat(ajt[1],vev); m = m+1; nv = (nv/2)\1;
if(nv < 2, ajt[1] = nv; vev = concat(ajt[1],vev); m = m+1; break));
for(i = 1, m, j = m-i; tot = tot+(2 ^ j)*vev[i]);
prodd = 2.0; prodp = vector(2); exposa = vector(2); ajtp = vector(1);
prodp[2] = 1.0; exposa[2] = 0; prodp[1] = 2.0; exposa[1] = 0; exposd = 0;
for(i = 2, m, prodd = prodd*prodd; exposd = exposd*2;
for(i = 1, 10, if(prodd > 10, prodd= prodd/10; exposd = exposd+1, break));
ajtp[1] = prodd; prodp = concat(ajtp[1],prodp);
ajtp[1] = exposd; exposa = concat(ajtp[1],exposa));
prodd = 1; exposd = 0;
for(i = 1, m, if(vev[i] == 1, prodd = prodd*prodp[i]; exposd = exposd+exposa[i]));
prodd = prodd/mem1; exposd = exposd - mem2;
z = 1/(prodd*(10 ^ exposd) - 1));
print(z)}
```

Appendix G. Evaluation of the ratio r

The following program provides the evaluation of the ratio $r-1$ for some initial values x_0 equal to 3 mod 4. The ratio n°1 is always in the interval $]0,1[$ except here if $v=2$ for the said initial values.

Note again that sometimes the exponentiation sign ^ won't copy successfully and has to be retyped manually (4 corrections in that case).

PARI/GP progamming code.

```
{infty = 10000;
for(i = 1, 10, \\ make choice
x0 = 4*i+3;
parv = addp = shft = vector(1);
t = (3*x0+1)/2 ; parv[1] = 1; shft[1] = 0; v = 1;
for(k = 1, infty, if(t < x0, break,
if(t/2 == t\2, t = t/2;
addp[1] = 0, v = v+1; t = (3*t+1)/2; addp[1] = 1);
parv = concat(parv, addp[1])));
print("Initial value = "x0", v = "v);
print("Parity vector "parv);
w = (log(3)/log(2)*v)\1+1; addp[1] = 0; m = 1;
for(k = 1, v, n = m;
for(j = 1, w-n, if(parv[n+j] == 0, m = m+1, break)); m = m+1; addp[1] = m-1;
shft = concat(shft, addp[1])); shftp = shftn = shft; tot0 = 0.0;
for(k = 1, v, tot0 = tot0+(3 ^ (v-k))*(2 ^ shftn[k]));
for(j = 1, v-1,
for(k = 3, v+1, shftp[k-1] = shftn[k]-shftn[2]);
print("Exponents vector "shftp); tot1 = 0.0;
for(k = 1, v, tot1 = tot1+(3 ^ (v-k))*(2 ^ shftp[k]));
r = tot1/tot0; print("Ratio n°"j": "r-1); shftn = shftp))}
```

Literature and sources

[1] Riho Terras. A stopping time problem on the positive integers. Acta Arithmetica 30 (1976), 241–252
[2] Shalom Eliahou. 2011-2013. http://images.math.cnrs.fr/Le-probleme-3n-1-y-a-t-il-des-cycles-non-triviaux-III
[3] Terence Jackson, Keith Matthews. Journal of Integer Sequences, Vol. 5 (2002),Article 02.2.7. On Shanks' algorithm for computing the continued fraction of $\log_b$(a).
[4] Wikipedia. Collatz conjecture. Cycle length. https://en.wikipedia.org/wiki/Collatz_conjecture
[5] Wikipedia. Natural density. https://en.wikipedia.org/wiki/Natural_density
[6] Wikipedia. Ecart moyen. https://fr.wikipedia.org/wiki/Écart_moyen
[7] Wikipedia. Continued fraction. https://en.wikipedia.org/wiki/Continued_fraction
[8] Wikipedia. Gaussian brackets. https://en.wikipedia.org/wiki/Gaussian_brackets
[9] Hubert Schaetzel. https://hubertschaetzel.wixsite.com/website

INPG Grenoble
Email address: `hubert.schaetzel@wanadoo.fr`
URL: `https://hubertschaetzel.wixsite.com/website`