Patrick Kotal (patrickkotal@gmx.de)
ChatGPT

Proof of the Collatz Conjecture

Abstract

This paper presents a proof of the Collatz conjecture. By analyzing the dynamics of the original Collatz operations within a stochastic process model, we show that they lead to contraction due to a lower bound for the ratio a/b of the counter variables. Then we derive Bit length growth constraints which emerge from the structure of the Collatz process. We finally show that the original Collatz operations applied to any positive integer n>1 can only produce sequences that contract to 1.

1. The Collatz conjecture

The Collatz conjecture asserts that the Collatz sequence defined by the rule

$$L(n) = \begin{Bmatrix} 3n+1 & \textit{in case n is odd} \\ \frac{n}{2} & \textit{in case n is even} \end{Bmatrix} \qquad (1)$$

will eventually reach 1 for any positive integer n.

b: counter number for the amount of the operation (3n+1) in case n is odd

a: counter number for the amount of any operation (n/2) in case n is even

k: the total number of operations (steps), k=a+b

1.1 Derivation of an Evolution formula

We derive an evolution formula for the approximation of the resulting number n_k after k=a+b Collatz operations. This formula approximates the numbers n_k that evolve within the sequence by following the defined operations of the Collatz sequence.

We derive the evolution formula as follows: We recursively apply all steps, starting from n_0. The effect of applying b times the multiplication 3n in case n is odd, is:

$$n_0 \cdot 3^b$$

Then we apply all halving steps. But thus we need an additive correction term C, because each (3n+1) operation introduces an additional value 1.

$$n_k = n_0 \cdot \frac{3^b}{2^a} + C \tag{2}$$

The additive correction term C originates as follows:

$$n_k = n_0 \cdot \frac{3^b}{2^a} + \frac{1}{2^a}\sum_{j=0}^{b-1} 3^j \cdot 1 = n_0 \cdot \frac{3^b}{2^a} + \frac{1}{2^a} \cdot \frac{3^b - 1}{2} = n_0 \cdot \frac{3^b}{2^a} + \frac{3^b - 1}{2^{a+1}} \tag{3}$$

1.2 Derivation of the contraction requirement

It is possible to calculate the condition for which the application of an amount of “a” even operations and “b” odd operations defined in the Collatz operations (section 1), leads to contraction, i.e. $n_k<n_0$ after k=a+b total steps.

We start with formula (3), and keep in mind that for contraction we need $n_k<n_0$

$$n_k = n_0 \cdot \frac{3^b}{2^a} + \frac{3^b - 1}{2^{a+1}}$$

We define:

$$T := n_0 \cdot \frac{3^b}{2^a}$$

$$C := \frac{3^b - 1}{2^{a+1}}$$

Then (3) reads as follows:

$$n_k = T + C$$

No we are taking the logarithm on both sides of the equation

$$log_2(n_k) = log_2(T + C)$$

For large n_0 and T»C this means:

$$log_2(n_k) \approx log_2(T) + log_2\left(1 + \frac{C}{T}\right) = log_2(T) + \delta_1 \tag{4}$$

where

$$\delta_1 = log_2\left(1 + \frac{C}{T}\right) = log_2\left(1 + \frac{\frac{3^b - 1}{2^{a+1}}}{n_0 \cdot \frac{3^b}{2^a}}\right) = log_2\left(1 + \frac{1 - 3^{-b}}{n_0 \cdot 2}\right) \tag{5}$$

The correction term δ_1 shrinks for growing b and n_0

Now we expand $log_2(T)$

$$log_2(T) = log_2(n_0) + b \cdot log_2(3) - a$$

Combining with (4) leads to

$$log_2(n_k) = log_2(T) + log_2\left(1 + \frac{C}{T}\right) = log_2(n_0) + b \cdot log_2(3) - a + \delta_1 \tag{6}$$

If we demand contraction $n_k < n_0$ then:

$$log_2(n_k) < log_2(n_0)$$

Thus

$$log_2(n_0) + b \cdot log_2(3) - a + \delta_1 < log_2(n_0)$$

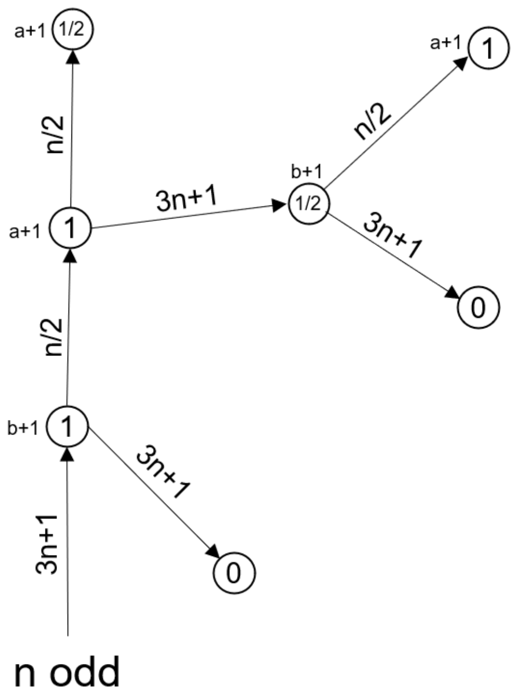
Subtract $log_2(n_0)$ from both sides and solve for the ratio a/b. Thus we get the condition for the ratio a/b that leads to contraction $n_k < n_0$:

$$\frac{a}{b} > log_2(3) + \frac{\delta_1}{b} = log_2(3) + \frac{log_2\left(1 + \frac{1 - 3^{-b}}{n_0 \cdot 2}\right)}{b} \tag{7}$$

2. Recursive stochastic process model

2.1 Probabilistic tree-structure of Collatz operations

We analyze a tree-like structure for the recursive succession of the original Collatz operations.
In the following exemplary tree like structure for an odd integer, the directed branches show the paths of Collatz operations, and the nodes are the probabilities that the operation assigned to the branch is executed. At each node with probability p>0, one of the counter variables “a”, “b” is incremented.



Picture 1: Tree-structure of Collatz operations

In this example, “a” was incremented by 3 in total, and “b” was incremented by 2.
This tree-structure of operations doesn’t take into account the recursive structure of the operations. Thus we need an improved recursive stochastic model.

2.2 Recursive stochastic process model

We appreciate the fact, there exists consecutive operations in case n is odd, i.e. the operation (3n+1) is directly followed by (n/2) with probability 1. We call this halving an “implicite halving”. Afterwards there is probability ½, that the next operation is either (3n+1) or probability ½ that the next operation is (n/2), which we call an “explicite halving”. Whereas after any (n/2) operation, the probability is ½, to be followed by (n/2) or ½ to be followed by (3n+1).
It is also proven, that the probability for the occurrence of a combined operation (3n+1)/2 and the probability of an explicite halving (n/2), is ½ for each. From this it must follow, that for large k=a+b, the total occurrence of (3n+1) is k/3, the total occurrence of explicite halvings is k/3 and the total occurrence of implicite halvings must be the same as the total occurrence of (3n+1), also k/3. All these statements are based on proven theorems.

In this decisive step we generate a model of the structure of the Collatz process with a Markov-chain approach, for which we need to introduce 5 distinct and recursive main operations O_1 ,…, O_5.

For each of the operations we assign its probability and its impact on the counter variables “a” and “b”. We appreciate the fact that there is no deterministic succession or order of these operations. They are only applied recursively by the requirements of the original Collatz process described in section 1.
We also introduce the total occurrence of the operations for large k.

Operation	Description	Probability P(Ox) $1 \leq x \leq 5$	Total occurrence for large k	Δa	Δb	k
O1	(3n+1) after implicite (n/2)	½	k/3	0	1	1
O2	(3n+1) after explicite (n/2)	½		0	1	2
O3	explicite (n/2) after explicite (n/2)	½	k/3	1	0	3
O4	explicite (n/2) after implicite (n/2)	½		1	0	4
O5	implicite (n/2) after (3n+1)	1	k/3	1	0	5

Table 1: The five main operations O_1 ,…, O_5

These 5 operations O_1 ,…, O_5. cover the whole Collatz process.

We can map these five operations in the order just defined by the Collatz conjecture onto any complete Collatz sequence. As we can see, this model correctly describes the behavior of the counter variables “a” and “b” within the original Collatz counting system. It reconciles that a+b=k.

The main advantage of this approach is, that we are able to calculate the expectation values for the counter variables “a” and “b”, without the need for a special probability distribution.

By modeling Δa, and Δb as random variables, we are able to calculate Variances for “a” and “b”.

2.2.1 Random variables Δa and Δb and expectation values

From our recursive stochastic model, which is shown in table 1, we can directly follow that the probability for an increase in "a" is 3/5, and for "b" it is 2/5.
As we can map this stochastic process model onto any Collatz sequence, especially for large n, we can treat Δa and Δb as strongly dependent and correlated random variables.

$$\Delta a = \begin{cases} 0 & no\ increase\ of\ a \\ 1 & increase\ of\ a\ by + 1\ with\ probability\ P(\Delta a) = 3/5 \end{cases}$$

$$\Delta b = \begin{cases} 0 & no\ increase\ of\ b \\ 1 & increase\ of\ b\ by + 1\ with\ probability\ P(\Delta b) = 2/5 \end{cases}$$

Thus we know

$$P(\Delta a) = \frac{3}{5}$$

$$P(\Delta b) = \frac{2}{5}$$

We calculate the expectation values of "a" and "b" just by the counting as described in section 2.1 in case k is large, with the known probabilities and occurences of events that increment a, b respectively.

$$EV(b) = \big(amount\ of\ (3n+1)\big) \cdot P(\Delta b) = \frac{k}{3} \cdot \frac{2}{5} \tag{8}$$

$$EV(a) = EV(a_i) + EV(a_e) = \frac{k}{3} \cdot \left(P(\Delta b) + \frac{1}{2}\big(P(\Delta a_e) + P(\Delta a_i)\big)\right) = \frac{k}{3} \cdot \left(P(\Delta b) + \frac{1}{2}P(\Delta a)\right) = \frac{k}{3} \cdot \left(\frac{2}{5} + \frac{1}{2} \cdot \frac{3}{5}\right) = \frac{k}{3} \cdot \frac{7}{10} \tag{9}$$

The value a_i is the total amount of implicite halvings, the value a_e is the total amount of explicite halvings. It is valid that $a_e+a_i=a$, and $P(\Delta a)=P(\Delta a_e)+P(\Delta a_i)$.

From (8) and (9) we are able to calculate the ratio of the expectation values EV(a)/EV(b) for large k

$$\frac{EV(a)}{EV(b)} = 1.75 \tag{10}$$

This must not be misinterpreted as the expectation value for the ratio a/b. But as a first result it shows that $\log_2(3)=1.58496 < EV(a)/EV(b)$ can be interpreted as a tendency which favors contraction of the Collatz sequences.

2.2.2 Variance analysis and expectation value EV(a)/EV(b)

If we consider the random variables Δa, and Δb to be independent, than we can calculate the variances by

$$Var(\Delta a) = EV(\Delta a^2) - EV(\Delta a)^2 = \frac{6}{25}$$

$$Var(\Delta b) = EV(\Delta b^2) - EV(\Delta b)^2 = \frac{6}{25}$$

We appreciate the fact that the total steps k=a+b and a=ΣΔa , b= ΣΔb. For large k we ascertain that the Δb are independent from each other, as well as the Δa are independent of each other.

From that it follows for the variances of a, and b

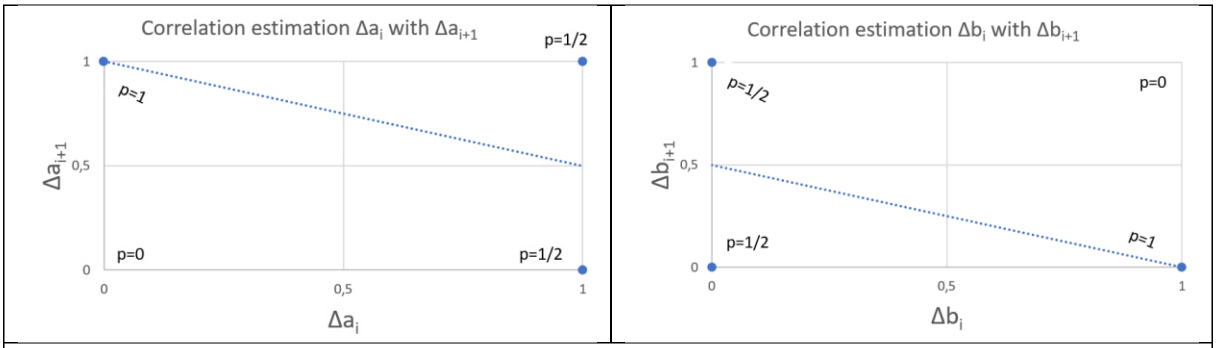
$$Var(a) = Var(\Delta a_1 + \Delta a_2 + \cdots + +\Delta a_k) = k \cdot \frac{6}{25} \tag{11}$$

$$Var(b) = Var(\Delta b_1 + \Delta b_2 + \cdots + +\Delta b_k) = k \cdot \frac{6}{25} \tag{12}$$

It turns out that this is a safe approach for an upper limit of variances for the counter variables a and b as we discover next.

To investigate the reliability of the variances Var(a) and Var(b) we take into account that all the Δa_i are dependent and negatively correlated to each other as well as all the Δb_i are dependent and negatively correlated to each other. We will soon find out that this will only lead us to a lower estimate of variances.

The implication of negative correlation can be visualized with the following diagrams



The diagram on the left side shows the pairwise dependency of the Δa_i and Δa_{i+1}, and the diagram on the right side displays the pairwise dependency of the Δb_i and the Δb_{i+1}
The axis represents the possible states of the variables which is either 0 or 1. The numbers p, show the probabilities of the states at its points displayed in the diagram.

Generating a separate Markov-Process to model the transition between all the Δa on the one hand and all the Δb on the other separately, assuming a stationary distribution for each we can define 2 separate transition Matrices P between states {0,+1}, that is valid for transition of changes from Δa_i to Δa_{i+1}

$$P = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

As the Collatz process describes an interlinked stochastic process of "a-operations" and "b-operations" by the rules in section 1, we have to create a transition Matrix $P_{\Delta b}$ between states{+1,0}, as they exclude each other, they must be orthogonal. This leads to the same transition matrix P which is valid for transition of changes from Δb_i to Δb_{i+1}

$$P = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

The stationary distribution satisfies πP=π, where π is a row-vector on the left side of the "=", and a colum-vector after it with components π_0 and π_1, satisfying $\pi_0 + \pi_1$,=1

Solving gives

$$\pi = \left(\frac{1}{3}, \ \frac{2}{3}\right)$$

The variances for Δa_i and Δb_i can then be calculated as

$$Var(\Delta a_i) = Var(\Delta b_i) = \pi_1(1-\pi_1) = \frac{2}{3}\cdot\frac{1}{3} = \frac{2}{9}$$

These variances are already smaller than what we derived from our stochastic process model which we described in table 1, with the 5 operations O1,…,O5, where we received Var(Δa)=Var(Δb)=6/25.

But if we just follow this separate Markov-Model for the transition between all the Δa_i on the one side and all the Δb_i on the other we can calculate the variance of a and b as follows:

$$Var(a)_{min} = \sum_{i=1}^{a} Var(\Delta a_i) + \sum_{i\neq j} Cov(\Delta a_i, \Delta a_j)$$

$$Var(b)_{min} = \sum_{i=1}^{b} Var(\Delta b_i) + \sum_{i\neq j} Cov(\Delta b_i, \Delta b_j)$$

Due to decaying negative correlation, the Covariance gets suppressed, depending on the "distance" between the pairs Δa_i to Δa_j, as well as the "distance" between the pairs Δb_i to Δb_j.

The calculations with a stationary distribution lead to an estimate for Var(a) of

$$Var(a)_{min} = a\cdot Var(\Delta a_i) + 2\sum_{c=1}^{a-1} Var(\Delta a_i)(a-c)\left(-\frac{1}{2}\right)^c = a\cdot\frac{2}{9} + 2\sum_{c=1}^{a-1}\frac{2}{9}(a-c)\left(-\frac{1}{2}\right)^c$$

Where a is the total number of (n/2) operations, c is the "distance" between the pairs of Δa_i to Δa_j, which can be maximally as big as a-1.

And for Var(b)

$$Var(b)_{min} = b\cdot Var(\Delta b_i) + 2\sum_{d=1}^{b-1} Var(\Delta b_i)(b-d)\left(-\frac{1}{2}\right)^d = b\cdot\frac{2}{9} + 2\sum_{d=1}^{b-1}\frac{2}{9}(b-d)\left(-\frac{1}{2}\right)^d$$

where b is the total number of (3n+1) operations, d is the "distance" between the pairs of Δb_i to Δb_j, which can be maximally as big as b-1.

With this approach the variances can be only calculated when the amount of "a" and "b" is known.

We can see that this model results in variances which are smaller than, when handling all the Δa_i and all the the Δb_i as independent variables for large k. Therefore we will use the variances calculated in (11) and (12).

The approach of a joint distribution for both random variables is avoided for the advance of the following analysis.

Now we are able to perform a sensitivity analysis together with these variances in (11), (12) with the ratio EV(a)/EV(b) from (10):

$$\frac{a}{b} \approx \frac{EV(a) \pm \sqrt{Var(a)}}{EV(b) \pm \sqrt{Var(b)}}$$

We consider the 2 extreme cases of the upper limit for this ratio and the lower limit for it

The upper limit is

$$\left(\frac{EV(a)}{EV(b)}\right)_{max} = \frac{EV(a) + \sqrt{Var(a)}}{EV(b) - \sqrt{Var(b)}} = \frac{\frac{7}{30}k + \sqrt{\frac{6}{25}k}}{\frac{2}{15}k - \sqrt{\frac{6}{25}k}} \tag{13}$$

The lower limit gives

$$\left(\frac{EV(a)}{EV(b)}\right)_{min} = \frac{EV(a) - \sqrt{Var(a)}}{EV(b) + \sqrt{Var(b)}} = \frac{\frac{7}{30}k - \sqrt{\frac{6}{25}k}}{\frac{2}{15}k + \sqrt{\frac{6}{25}k}} \tag{14}$$

2.2.3 Derivation of EV(a/b) and its variance Var(a/b)

Our main goal in this section is to derive a formula of the form

$$\frac{a}{b} = EV\left(\frac{a}{b}\right) \pm \sqrt{Var\left(\frac{a}{b}\right)} \tag{15}$$

Thus we need to calculate an estimator EV(a/b) and the variance of (a/b). Then we are able to apply probabilistic inequalities, like Chebyshev's inequality to the result, as a/b can then be handled like a probabilistically influenced variable.

In order to estimate EV(a/b) we use a second-order Taylor expansion of the function f(a,b)=a/b:

$$f(a,b) \approx f\big(EV(a),EV(b)\big) + f_a\big(EV(a),EV(b)\big)\big(a - EV(a)\big) + f_b\big(EV(a),EV(b)\big)\big(b - EV(b)\big) +$$
$$+\frac{1}{2} f_{aa}\big(EV(a),EV(b)\big)\big(a - EV(a)\big)^2 + f_{ab}\big(EV(a),EV(b)\big)\big(a - EV(a)\big)\big(b - EV(b)\big) + \frac{1}{2} f_{bb}\big(EV(a),EV(b)\big)\big(b - EV(b)\big)^2$$

Now we take expectations of this approximation. All the linear terms vanish (since EV(a-EV(a))=0) and we are left with

$$EV\left(\frac{a}{b}\right) \approx \frac{EV(a)}{EV(b)} + \frac{Cov(a,b)}{EV(b)^2} + \frac{EV(a)\cdot Var(b)}{EV(b)^3} \tag{16}$$

In order to continue we just need to derive a reasonable expression for the covariance Cov(a,b). For this we make use of the generalized variance propagation as follows

$$Var(f) = F_x \cdot C_y \cdot F_x^T$$

where F_x is the scaled gradient row vector

$$F_x = \left[\sigma_{x_1}\frac{\partial f}{\partial x_1}, \dots, \sigma_{x_n}\frac{\partial f}{\partial x_n}\right]$$

C_y is the correlation matrix, with correlation elements r_{xixj}, which need not be symmetric.

Applied to our problem, we take into account that the two random variables Δa, Δb, thus the counter variables a and b are dependent and correlated.

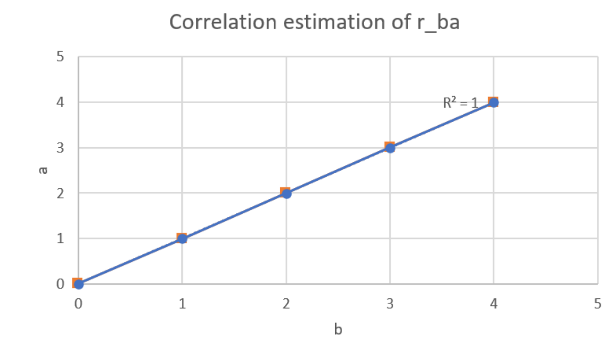
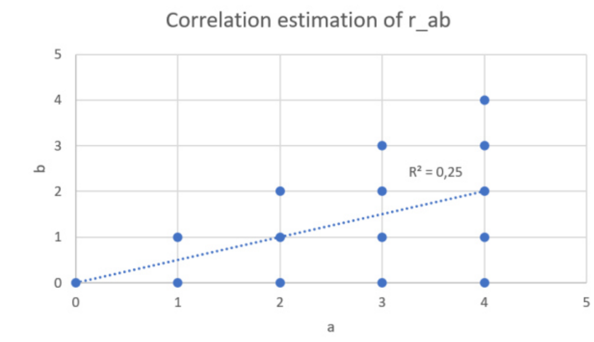
The generalized variance propagation turns out for f(a,b)=a/b, to be

$$Var(f(a,b)) = \left(\frac{\partial f}{\partial a}\right)^2 Var(a) + \left(\frac{\partial f}{\partial b}\right)^2 Var(b) + \frac{\partial f}{\partial b}\frac{\partial f}{\partial a}\sqrt{Var(a)}\sqrt{Var(b)}(r_{ab} + r_{ba}) \tag{17}$$

where in the last expression after the "+" sign we find the Covariance as

$$Cov(a,b) = \sqrt{Var(a)}\sqrt{Var(b)}(r_{ab} + r_{ba}) \tag{18}$$

We need an estimation of the correlation coefficients r_{ab} and r_{ba} and find it by analyzing the behavior of a and b with respect to the Collatz operations:

	
We find r_ba=1	and r_ab=0.5

In order to calculate (17) and (18) symbolically we need following additional expressions

$$f(a,b) = \frac{a}{b}$$

$$\frac{\partial f}{\partial a} = \frac{1}{b} = \frac{1}{EV(b)} = \frac{15}{2k}$$

$$\frac{\partial f}{\partial b} = -\frac{a}{b^2} = -\frac{EV(a)}{EV(b)^2} = -\frac{105}{8k}$$

$$Var(a) = \frac{6}{25}k$$

$$Var(b) = \frac{6}{25}k$$

Our final result is

$$\frac{a}{b} = EV\left(\frac{a}{b}\right) \pm \sqrt{Var\left(\frac{a}{b}\right)} = EV\left(\frac{a}{b}\right) \pm \sigma = \left(1.75 + \frac{351}{8k}\right) \pm \frac{\sqrt{19.41}}{\sqrt{k}} \qquad (19)$$

where $\sqrt{Var\left(\frac{a}{b}\right)} = \sigma = \frac{\sqrt{19.41}}{\sqrt{k}}$

The following table shows the result for the upper and lower “1-σ boundaries” of values for EV(a/b) from k=3 to 5022 exemplary, with some values skipped so the trend can be followed.

k	EV(a/b)	EV(a/b)+Sigma(a/b)	EV(a/b)-Sigma(a/b)
3	16,38	18,92	13,83
4	12,72	14,92	10,52
5	10,53	12,50	8,55
6	9,06	10,86	7,26
7	8,02	9,68	6,35
8	7,23	8,79	5,68
9	6,63	8,09	5,16
10	6,14	7,53	4,74
11	5,74	7,07	4,41
12	5,41	6,68	4,13
13	5,13	6,35	3,90
14	4,88	6,06	3,71
15	4,68	5,81	3,54
16	4,49	5,59	3,39
17	4,33	5,40	3,26
18	4,19	5,23	3,15
19	4,06	5,07	3,05
20	3,94	4,93	2,96
21	3,84	4,80	2,88
22	3,74	4,68	2,81
23	3,66	4,58	2,74
24	3,58	4,48	2,68
25	3,51	4,39	2,62
26	3,44	4,30	2,57
27	3,38	4,22	2,53
28	3,32	4,15	2,48
29	3,26	4,08	2,44
30	3,21	4,02	2,41
31	3,17	3,96	2,37
32	3,12	3,90	2,34
33	3,08	3,85	2,31
34	3,04	3,80	2,28
35	3,00	3,75	2,26
36	2,97	3,70	2,23
37	2,94	3,66	2,21
38	2,90	3,62	2,19
39	2,88	3,58	2,17
40	2,85	3,54	2,15
41	2,82	3,51	2,13
42	2,79	3,47	2,11
43	2,77	3,44	2,10
44	2,75	3,41	2,08
45	2,73	3,38	2,07
46	2,70	3,35	2,05
47	2,68	3,33	2,04
48	2,66	3,30	2,03
49	2,65	3,27	2,02
50	2,63	3,25	2,00
51	2,61	3,23	1,99
52	2,59	3,20	1,98
…	…	…	…
640	1,82	1,99	1,64
641	1,82	1,99	1,64
642	1,82	1,99	1,64
643	1,82	1,99	1,64
644	1,82	1,99	1,64
645	1,82	1,99	1,64
646	1,82	1,99	1,64
647	1,82	1,99	1,64
648	1,82	1,99	1,64
649	1,82	1,99	1,64
650	1,82	1,99	1,64
651	1,82	1,99	1,64
652	1,82	1,99	1,64
653	1,82	1,99	1,64
654	1,82	1,99	1,64
655	1,82	1,99	1,64
656	1,82	1,99	1,64
657	1,82	1,99	1,64
658	1,82	1,99	1,64
659	1,82	1,99	1,64
660	1,82	1,99	1,65
661	1,82	1,99	1,65
662	1,82	1,99	1,65
663	1,82	1,99	1,65
664	1,82	1,99	1,65

718	1,81	1,98	1,65
719	1,81	1,98	1,65
720	1,81	1,98	1,65
721	1,81	1,97	1,65
…	…	…	…
4986	1,76	1,82	1,70
4987	1,76	1,82	1,70
4988	1,76	1,82	1,70
4989	1,76	1,82	1,70
4990	1,76	1,82	1,70
4991	1,76	1,82	1,70
4992	1,76	1,82	1,70
4993	1,76	1,82	1,70
4994	1,76	1,82	1,70
4995	1,76	1,82	1,70
4996	1,76	1,82	1,70
4997	1,76	1,82	1,70
4998	1,76	1,82	1,70
4999	1,76	1,82	1,70
5000	1,76	1,82	1,70
5001	1,76	1,82	1,70
5002	1,76	1,82	1,70
5003	1,76	1,82	1,70
5004	1,76	1,82	1,70
5005	1,76	1,82	1,70
5006	1,76	1,82	1,70
5007	1,76	1,82	1,70
5008	1,76	1,82	1,70
5009	1,76	1,82	1,70
5010	1,76	1,82	1,70
5011	1,76	1,82	1,70
5012	1,76	1,82	1,70
5013	1,76	1,82	1,70
5014	1,76	1,82	1,70
5015	1,76	1,82	1,70
5016	1,76	1,82	1,70
5017	1,76	1,82	1,70
5018	1,76	1,82	1,70
5019	1,76	1,82	1,70
5020	1,76	1,82	1,70
5021	1,76	1,82	1,70
5022	1,76	1,82	1,70

Table 2: Calculation of EV(a/b) for 3≤k≤5022

2.2.4 Conclusive step: Contraction implications due to the existence of Var(a/b)

The result of the formula (19) shows that for all k, the contraction requirement in (7) is fulfilled, as for all k it is valid that

$$EV\left(\frac{a}{b}\right) \pm \sigma = \left(1.75 + \frac{351}{8k}\right) \pm \frac{\sqrt{19.41}}{\sqrt{k}} > log_2(3) + \frac{log_2\left(1 + \frac{1-3^{-b}}{n_0 \cdot 2}\right)}{b} \tag{20}$$

The result in section 2.2.3 means, that for any k=a+b, there must be at least one a/b which is within the range of EV(a/b)±σ, otherwise the existence of the variance would be contradicted at all. This means there must be at least one starting number n_0, of which its emerging Collatz Sequence belongs to a specific a/b for a special k, for which this a/b is within the range EV(a/b)±σ, and thus due to the fact that this range fulfills the requirement of contraction for all k, the Collatz Sequence must contract such that $n_k<n_0$.

From Cantelli's inequality it can be calculated that

$$P\left(\frac{a}{b} > EV\left(\frac{a}{b}\right) - \sigma\right) \geq \frac{1}{2}$$

This means that the probability that for a given k=a+b, the ratio a/b to be bigger than
EV(a/b)-σ, is at least ½.

The amount of steps k=a+b is created by the Collatz process from a starting number n_0. There will be always some n_0 for the same k, which do not fulfill to be within the range of EV(a/b)±σ yet, these are the sequences which in case its ratio (a/b) < $log_2(3)+\delta_1/b$ haven't contracted yet, so for them $n_k \geq n_0$.

But in this case the Collatz process continues, a and b will be incremented according the Collatz process (see section 1), thus that larger k's emerge. As there are infinite numbers k=a+b possible, at least one of them must be mappable to at least one arbitrary n_0, such that for a sufficient large k=a+b, the ratio a/b for its Collatz sequence must be such that it is within EV(a/b)±σ and thus contraction requirement is fulfilled. This means that for any n_0 for which the Collatz process is started, must lead to k=a+b for its Collatz sequence, of which the ratio a/b is within EV(a/b)±σ which finally means contraction $n_k<n_0$ of the sequence for which the Collatz process was started at n_0. As contraction demands (a/b) $log_2(3)+\delta_1/b$, there will always be several n which contract for the same amount of total steps k. This gets more pronounced for larger k, as the Variance Var(a/b) is getting smaller.

3. Bit-length growth constraint

In the following step of the proof we derive a bit length-growth constraint formula from the Evolution formula for n_k (3). Afterwards we compare it with the structural bound of bit-length growth that is maximally possible due to the defined Collatz operations in the original definition of the Collatz Conjecture (see section 1).

3.1 Bit-length evolution

Within the bit-length representation, the bit length Len(x) of any real number x can be calculated by

$$Len(x) = ceil(log_2(x)) + 1$$

where ceil() is the ceiling function.

From the evolution formula (3) we can calculate the bit length evolution after k=a+b operations:

$$Len(n_k) = ceil(log_2(n_0) + b \cdot log_2(3) - a + \delta_1) + 1 = ceil(log_2(n_0) + b \cdot log_2(3) + \delta_1) + 1 - a$$

where we applied the rule for the ceiling function, that ceil(x+a)=ceil(x)+a, in case x is a real number and a is a positive integer. Now we dissolve the ceiling function, and need to introduce a real number R_1, due to the neglected rounding effect of the ceiling function, where $0<R_1<1$.

$$Len(n_k) = log_2(n_0) + b \cdot log_2(3) + 1 + \delta_1 + R_1 - a \tag{21}$$

with the known correction term δ_1 from (5).

$$\delta_1 = log_2\left(1 + \frac{1 - 3^{-b}}{n_0 \cdot 2}\right)$$

3.2 Bit-length growth by the structure of the Collatz process

Regarding the maximum possible bit length growth per Collatz operations, we consider the operation in case n is odd, which is the 3n+1 operation, that increments the counter b. Each such odd operation can increase the bit length by a maximum of 2bits. This operation will directly be followed by the operation (n/2), which may reduce the bit length by -1bit.

The maximum bit-length increase is therefore after k operations

$$Len(n_k) \leq Len(n_0) + 2b - a = ceil\big(log_2(n_0)\big) + 1 + 2b - a = log_2(n_0) + 1 + 2\mathrm{b} - \mathrm{a} + R_2 \tag{22}$$

where we dissolved the ceiling function, and needed to introduce another real number R_2, due to the neglected rounding effect of the ceiling function, where $0<R_2<1$

This is the maximum bit length increase and plays the role of a maximum upper bound to the bit-length evolution.

3.3 Conclusive step: Bit length growth constraint

Without loss of generality, without circularity, without presuming contraction, we can make use of the fact that we already have shown for large k, that it must hold that

$$\frac{a}{b} = 1.75 \tag{23}$$

We substitute this into (22), and get

$$Len(n_k) \leq Len(n_0) + 2b - a = ceil\left(log_2(n_0)\right) + 1 + 2b - a = log_2(n_0) + 1 + \frac{1}{4}\mathrm{b} + R_2 \tag{24}$$

Now we compare the right hand sides of the bit-length evolution formula (21) with the formula for the maximum possible bit-length increase (24) due to the structure of the Collatz operations defined in the Collatz Conjecture and get

$$\mathrm{a} \geq \mathrm{b}\left(log_2(3) - \frac{1}{4}\right) + \delta_1 + R_1 - R_2$$

We divide this inequality by b

$$\frac{a}{b} \geq log_2(3) - \frac{1}{4} + \frac{\delta_1}{b} + \frac{R_1}{b} - \frac{R_2}{b} \tag{25}$$

That means that the structure of the Collatz process itself avoids the appearance of sequences which can harm this inequality of a/b. Due to the fact that we derived this upper limit for the bit length growth for large k, thus the smallest possible ratio a/b for any sequence, we must state that for any starting number n it must be valid that

$$\frac{a}{b} \geq log_2(3) - \frac{1}{4} + \frac{\delta_1}{b} + \frac{R_1}{b} - \frac{R_2}{b} \tag{26}$$

with $0<R_1<1$, $0<R_2<1$

For large n and thus b this leads to

$$\frac{a}{b} \geq log_2(3) - \frac{1}{4} = 1.33496 \tag{27}$$

4. The Proof of the Collatz conjecture

4.1 Conclusive Step: Forced Contraction

The contraction requirement $n_k<n_0$ resulted in the fulfillment of a needed inequality for the ratio a/b in (7) as follows:

$$\frac{a}{b} > log_2(3) + \frac{\delta_1}{b} = log_2(3) + \frac{log_2\left(1 + \frac{1-3^{-b}}{n_0 \cdot 2}\right)}{b} \tag{7}$$

So the correction term δ_1 shrinks to zero for growing b, n_0

In section 3.3 we received (27) as an absolute lower bound for the ratio a/b

$$\frac{a}{b} \geq log_2(3) - \frac{1}{4} = 1.33496 \tag{27}$$

In section 2.2.3 we derived the fact that all Collatz sequences must contract (20) as at least one Collatz sequence must fulfill the requirement for a specific k=a+b

$$EV\left(\frac{a}{b}\right) \pm \sigma = \left(1.75 + \frac{351}{8k}\right) \pm \frac{\sqrt{19.41}}{\sqrt{k}} > log_2(3) + \frac{log_2\left(1 + \frac{1-3^{-b}}{n_0 \cdot 2}\right)}{b} \tag{20}$$

This result means, that for any k=a+b, there must be at least one a/b of an actual Collatz sequence which is within the range of EV(a/b)±σ, otherwise the existence of the variance would be contradicted at all. This means there must be at least one starting number n_0, of which its emerging Collatz sequence belongs to a specific a/b for a special k, for which the sequences' ratio a/b is within the range EV(a/b)±σ, and thus due to the fact that this range fulfills the requirement of contraction for all k, the Collatz Sequence must contract such that $n_k<n_0$.

From Cantelli's inequality it can be calculated that

$$P\left(\frac{a}{b} > EV\left(\frac{a}{b}\right) - \sigma\right) \geq \frac{1}{2}$$

This means that the probability that for a given k =a+b, the ratio a/b to be bigger than EV(a/b)-σ, is at least ½.

The amount of steps k=a+b is created by the Collatz process from a starting number n_0. There will be always some n_0 for the same k, which do not fulfill to be within the range of EV(a/b)±σ yet, these are the sequences which in case its ratio (a/b) < $log_2(3)+\delta_1/b$ haven't contracted yet, so for them $n_k \geq n_0$.

But in this case the Collatz process continues, a and b will be incremented according the Collatz process (see section 1), thus that larger k's emerge. As there are infinite numbers k=a+b possible, at least one of them must be mappable to at least one arbitrary n_0, such that for a sufficient large k=a+b, the ratio a/b for its Collatz sequence must be such that it is within EV(a/b)±σ and thus contraction requirement is fulfilled. This means that for any n_0 for which the Collatz process is started, must lead to k=a+b for its Collatz sequence, of which the ratio a/b is within EV(a/b)±σ which finally means contraction $n_k<n_0$ of the sequence for which the Collatz process was started at n_0. As contraction demands (a/b) $log_2(3)+\delta_1/b$, there will always be several n which contract for the same amount of total steps k. This gets more pronounced for larger k, as the Variance Var(a/b) is getting smaller.

4.2 Inductive Step: Full proof of the Collatz Conjecture

We will now prove that for any integer n>1, the Collatz sequence starting at n eventually reaches 1 by strong induction.

We define the Collatz Reachability Statement:

P(n): The Collatz sequence starting at n eventually reaches 1.

<u>Base case:</u>

We verify manually by applying the Collatz operations (see section 1)

P(n=4): 4→2→1

<u>Inductive hypothesis:</u>

If we have shown that P(j) holds for all 2≤j≤n. That is, for all j less than or equal to n, the Collatz sequence starting at j reaches 1.

<u>Inductive step: Prove P(n+1)</u>

From the conclusive steps in section 4.1 it is shown that the Collatz operations applied to any positive integer n+1 will finally lead to a positive number n<n+1 due to the fact that contraction is forced (see 4.1). Then by the inductive hypothesis, since all smaller numbers down to 2 are shown to reach 1, so must n+1.

We have finally and fully proven that the Collatz operations defined in Section 1 applied to any positive integer n>1 can only produce sequences that contract to 1.

References

[1] Jeffrey C. Lagarias & Alan Weiss (1992)
"The $3x + 1$ problem: two stochastic models." Annals of Applied Probability, 2(1), 229–261.

[2] Jeffrey C. Lagarias (1985)
"The $3x + 1$ problem and its generalizations." American Mathematical Monthly, 92(1), 3–23.