A Statistical approach to Collatz Conjecture

The Collatz conjecture is a conjecture in mathematics that concerns a sequence defined as follows: start with any positive integer n. Then each term is obtained from the previous term as follows: if the previous term is even, the next term is one half of the previous term. If the previous term is odd, the next term is 3 times the previous term plus 1. The conjecture is that no matter what value of n, the sequence will always reach 1.

Some papers associate Collatz Conjecture to Prime Numbers properties. Others consider it a Number Theory problem. Since my mathematics don't go that far, I started to see Collatz Conjecture as a battle between even and odd numbers. That battle didn't make much sense to me, until I noticed the "hidden rules" of the sequence:

The "classic" definition

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

can be expanded to

$$f(n)=\begin{cases} \frac{n}{2^x} & if\, x=\infty \text{ to } 1, n \equiv 0 (mod\, 2^x) \\ 3n+1 & if\, n \equiv 1 (mod\, 2) \end{cases}$$

This way, the conjecture turns to be a battle between odd and "power of 2" numbers.

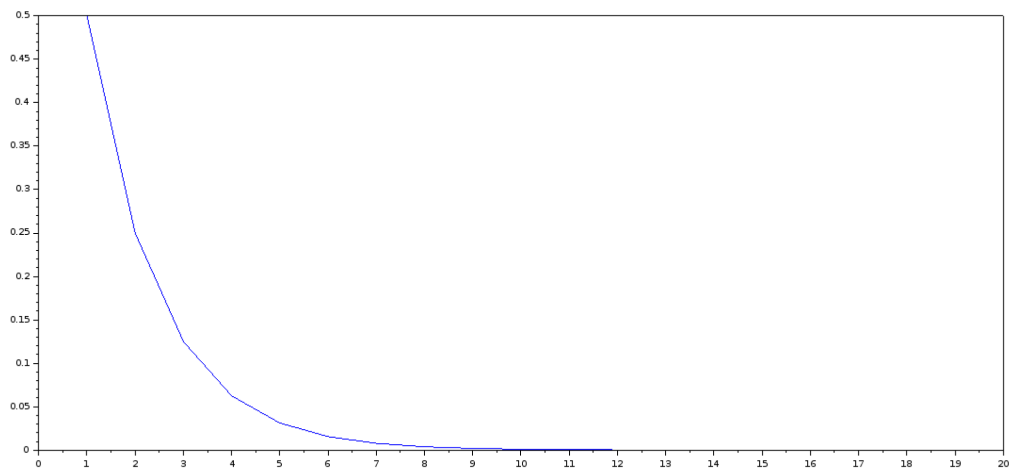
One advantage of this approach is that this 2 groups are evenly distributed on the new Collatz sequence: after a odd number there is a "power of 2" number, and so on. Whether the sequence tends to infinity or 1 depends of the randomness distribution of "power of 2" group at the sequence and its effective "weight" – by what factor is n divided.

We can be sure that the odd factor is 3 – each odd number is multiplied by 3.

About "power of 2" numbers, we know their distribution. Each "power of 2" 2^x is 2^{-x} of all even numbers.

For x=1, $2^x = 2$, $2^{-x} = 1/2$

For x=2, $2^x = 4$, $2^{-x} = 1/4$



The sum converges to

$$\sum_{x=1}^{\infty} 2^{-x} = 1$$

The division factor for each 2^x are given by the rules : it's 2^x.

For x=1, $2^x = 2$
For x=2, $2^x = 4$

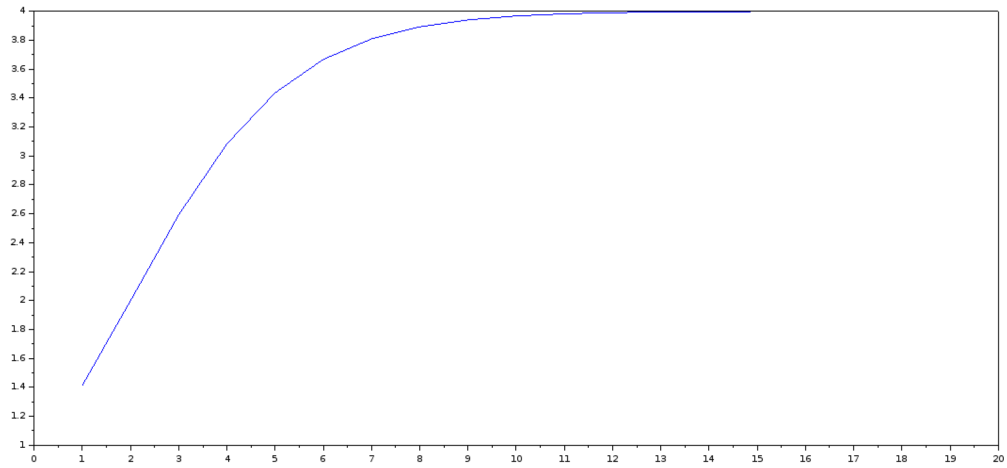
The "weight" of each "power of 2" is given by its $\text{factor}^{\text{distribution}}$

For x=1, $2^x = 2$ => weight = $2^{0.5} \approx 1{,}4142135623731$
For x=2, $2^x = 4$ => weight = $4^{0.25} \approx 1{,}4142135623731$
For x=3, $2^x = 8$ => weight = $8^{0.125} \approx 1{,}29683955465101$

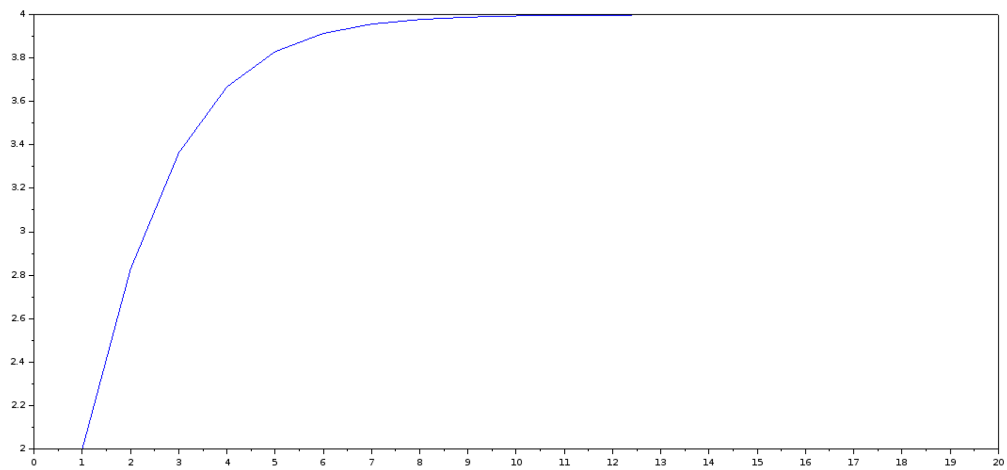
The total weight for all "power of 2" numbers (which is the factor we are looking for) is given by the product of the series

$$\prod_{x=1}^{\infty} (2^x)^{2^{-x}} = 4$$

and the corresponding graphic is



but if the range is (for example) {1,…, 64}, there are no powers of 2 bigger then 64 and the 64 distribution is equal to the previous – 32 – distribution. The corrected graphic is



For all "power of 2", the maximum factor is 4, so we calculate the ratio odd/"power of 2" as 3/4.

Conclusion

Considering that all odds are multiplied by 3 and all “power of 2” are (statistically) divided by 4, Collatz function can’t diverge to infinity because of the Law of large numbers. Does it always reach number 1? Maybe there is a repeating cycle that excludes 1, different from $4 \rightarrow 2 \rightarrow 1$ (which in the expanded Collatz function is reduced to the cycle $4 \rightarrow 1$), but bear in mind that at range {1..4} the “power of 2” factor is ≈ 2.83, so the ratio odd/”power of 2” is close to 1. That never happens at other ranges.

It can be assumed that at Collatz function variations like 5n+1 most of the numbers will diverge to infinity, because of the 5/4 ratio. Why not all? If a sequence hits a “power of 2” like $1 \rightarrow 6 \rightarrow 3 \rightarrow 16$ then you’re stuck in a cycle. How there is a cycle with such a divergent ratio? You can find similar cycles for 7n+1 and bigger ratios. I again call for the Law of large numbers to affirm that this happens solely with small samples, where the “power of 2” distribution is not respected by the sequence. The bigger the sequence, the most improbable to find such cycle. The bigger the numbers, the more sparse are the infamous “power of 2”.

Since you can find some 5n+1 sequences that reach 1, I guess the next step is to find out if 5n+1 sequences always hit a power of 2. If true, that function will also not never diverge to infinity. That’s a tough one to prove. Infinity is really big!

Caldas Lopes

February 15, 2021