An investigation in the Hailstone function

Jack van der Elsen, vde@oce.nl
Rob van den Tillaart, rvdt@oce.nl

Introduction

One of the most intriguing functions in algorithms-alley is the Hailstone function. This Hailstone function is known on other names like "3X+1" or "the Collatz problem" [3], [4], [5] as well. For me the first encounter with this function was in the book "Programming Pearls" [1]. The function looks very simple and for those unfamiliar we give the definition:

```
int Hailstone(int n)
{
  if ((n % 2) == 0)// n = even
  {
    return n/2;
  }
  else           // n = odd
  {
    return n * 3 + 1;
  }
}
```

There exists a hypothesis that for every integer value of n > 1 this function will converge to 1 when called in a loop.

```
iterations = 0;
while (n > 1)
{
 iterations ++;
  n = Hailstone(n);
  Console.Write(iterations);
  Console.WriteLine(n);
}
```

This loop will count the number of iterations needed to reach 1 and prints all intermediate results. A list of these results is called a Hailstone sequence.

On the Internet there are many sites describing aspects of this function but no one has proven the conjecture above. This essay doesn't describe this proof either, however it will give an explanation for one of the attributes of the Hailstone function.

Problem description

In February 2005 Jack and I discussed the Hailstone function and he showed me the web site [2] of Greg Keogh. On this web site Greg presents a Mathematica program with which he has tested if very large numbers also converge to 1. And as far as the numbers tested they do. However no proof of convergence only another indication that all numbers do.

Most interesting on the web is page was the experiment to calculate the length of Hailstone sequences for very large numbers. With the help of Mathematica Greg found a logarithmic relationship between the number z and the length of Hailstone sequence of z. The formula found was:

```
Length = 10.3104 log[z]
```

The question [the author] asked was if anyone could explain the magic number 10.3104. The next paragraph of this essay will discuss the Hailstone function and the length of Hailstone sequences. The result of this discussion is an explanation for this number.

Approximation calculation of Hailstone run lengths

The problem of calculating an approximation of the length of Hailstone sequences is split in two. First we derive a formula for the odd numbers, and derive from this the formula for even numbers.

If one looks at the Hailstone function one sees 2 calculations, a multiply with 3 + 1 (call it O for Odd) , and a division by 2 (call it E for even). A typical Hailstone sequence for an odd number can symbolically be written as follows:

$$\underline{O \swarrow E^{\wedge}n_1} \swarrow \underline{O \swarrow E^{\wedge}n_2} \swarrow \ . \ . \ . \ \swarrow \underline{O \swarrow E^{\wedge}n_{i-1}} \swarrow \underline{O \swarrow E^{\wedge}n_i} \swarrow 1$$

We start with the O-rule, do n_1 times the E-rule, do one O rule, n_2 times the E rule etc. The difficulty is that we do not know how often the E rule is to be applied. A thought experiment will help. Let us define the execution of one O rule followed by n times the E-rule a block. In the sequence every block is underlined.

After applying the O rule to an odd number, the number becomes even and therefor divisible by 2. The result of applying the E-rule has a 50% chance it is even or odd. Formulated in other words there is a 50 % chance that the E rule is executed twice. After applying the E-rule again there is again a 50% chance for even and odd. So there is a 25% chance the E rule is executed three times, 12.5% that it is executed 4 times, 6.25% chance that it is executed 5 times etc. The expected number of times the E-rule is = 2 [see appendix].

This means that the average block has one O-rule followed by the E-rule twice and therefor a length of three. This means that on average any odd number is multiplied by 3/4 in three iterations of the Hailstone function. As this factor is smaller than one we may conclude on statistical grounds that the number will converge to one. *Please note that the constant +1 is ignored. For large numbers this constant is not significant. For the algorithm of course it is!*

We define the sequence length as N and the starting number X. With the knowledge above we can state that after 'executing' N/3 blocks of the Hailstone function, the sequence is converged to one. Written down as a formula:

    (3/4) ^(N/3) * X = 1

Solving this formula for N gives us the following steps.

    (4/3) ^(N/3) = X   // note the exchange of 4 and 3

    N/3 = ln(X) / ln(4/3)   // Logarithmus Naturalis

    N = 3 / (ln(4) – ln(3)) * ln(X)

    N = 10.428178490 * ln(X)

So from our though experiment we derived an approximation formula that calculates the length of a Hailstone sequence for an odd number. The good news is that the magic number in this formula is very close to the number 10.3104 mentioned on the web site of Greg. The difference is less than 1.15%. The bad news is that we still have to do the even numbers so lets start the second thought experiment.

For an even number the Hailstone sequence can be written as follows

$$E\verb|^|n_0 \; \diagup \; \underline{O \; \diagup \; E\verb|^|n_1} \; \diagup \; \underline{O \; \diagup \; E\verb|^|n_2} \; \diagup \; . \; . \; . \; \diagup \; \underline{O \; \diagup \; E\verb|^|n_i} \; \diagup \; 1$$

This sequence looks quite similar to the sequence for odd numbers but there are a number of calls to the E-rule first. It is easy to convert this sequence for even numbers into a sequence for an odd number. Following the same reasoning as above about the times the E-rule is executed we can state that on average n_0 will be 2. This brings us to the following calculation for the even numbers:

```
(3/4) ^((N-2)/3) * X/4 = 1
```

Note that X has become X/4 in the two E-rule steps.

Solving this formula for N gives us the following steps.

```
(4/3) ^((N-2)/3) = X/4

(N-2)/3 = ln(X/4) / ln(4/3)

N-2 = 3 / (ln(4) - ln(3)) * ln(X/4)

N-2 = 10.428178490 * (ln(X) + ln(1/4))

N-2 = 10.428178490 * ln(X) - 14.45652504

N = 10.428178490 * ln(X) - 12.45652504
```

To get the average formula we add the formulas for the odd and even numbers and divide the result by two. This gives us in the following final formula:

N = 10.428178490 * ln(X) - 6.228262519

Discussion points

From the thought experiment we derived a formula that gives an approximation of the expected sequence length for a given number X. The formula is certainly not exact. In the thought experiment the factor +1 in the O-rule (3X+1) is ignored. The effect of adding this constant would give a far more complex thought experiment and therefor a far more complex formula. Incorporating the constant +1 would make the resulting factor 10.4281… slightly larger and it would become a complex function of the number N. It is expected that only for relative small numbers (< 100) the difference will be significant (>1%).

Now we compare the derived formula with the one found by Greg Keogh:

```
Length = 10.3104 log[z]
```

When writing the functions as `N = C`$_1$` * ln(X) + C`$_2$` ` we see that the constant C_1 is identical within 1.15%. As it is an approximation function these numbers may be considered the same for large values of X. The constant C_2 is responsible for a better fit for small values of X as C_1 and C_2 are about the same size.

Finally the two functions are (approx.) equal for `X = 1.107276E+23.`

Conclusions

The conclusion is that the formula derived gives an approximation of the length of the Hailstone sequence for any given number. What is important is that this theoretically derived number does match the experimental number within 1.15%.

We do not know if this formula helps proving the longstanding conjecture of the Hailstone function. As it has a statistical basis it is definitely not a proof. The model of blocks can be used to get some grip on the conjecture. One could calculate the needed ratio between the E and O-rule for a number that would not converge to 1 anymore (~ ln(3)/ln(2) : 1 ≅ 1.58496 : 1) and compare this to how odd and even numbers are distributed.

References

[1] John Bentley, Programming Pearls, Addison Wesley 2000, ISBN 0-201-65788-0.
[2] Greg Keogh, http://www.orthogonal.com.au/hobby/mm/hail/; Greg Keogh
[3] Eric W. Weisstein. "Collatz Problem." From *MathWorld*--A Wolfram Web Resource.
        http://mathworld.wolfram.com/CollatzProblem.html
[4] http://sprott.physics.wisc.edu/pickover/Hailstone.html
[5] http://perso.wanadoo.fr/patrick.davalan/Liens/liens_syracuse.html

Appendix

The average number of times the E-rule is executed is equal to the average number of factor 2's when a even number is written as a product of its prime factors.

Of the even numbers
        50% has one factor 2
        25% has two times the factor 2
        12.5% has three times the factor 2
        etc.

So the average number of factors 2 in even numbers is:
Lim(n-> infinity) Sum (k=1..n, k*(0.5)^k) / Sum(k=1..n, (0.5)^k) = 2 / 1 = 2