Journal of Integer Sequences, Vol. 15 (2012),
Article 12.5.2

The $3n+1$ Problem: A Probabilistic Approach

Darrell Cox
dlcox@graysoncable.com

Abstract

The Poisson probability distribution is used to model the number of cycles in the generalized Collatz problem. First, *interrelated cycles* are defined and used as a criterion in counting the cycles for a given q value. Initially, archived data in the mathematical literature (giving the known $3n+q$ cycles) is analyzed. For large samples, the Poisson probability distribution gives a poor fit of the data (there are too many cycles for the large x values). *Associated cycles* are defined and used as an additional criterion in counting cycles; this improves the data fit substantially. Some theory and empirical results are given in an attempt to explain the origin of this distribution of cycle counts. Degrees of freedom in probability distributions involving the difference between the number of odd and even elements in a cycle are shown to be a partial explanation for the distribution of cycle counts. *(L, K) trees* (generalized associated associated cycles) are defined and used to account for the smallest difference between the number of odd and even elements in the cycles for a given q value. The article consists entirely of analysis of empirical results; no proofs are given.

1 Introduction

Let n be a natural number. If n is odd, let the next number in a sequence be $3n+1$, otherwise let the next number be $n/2$. The Collatz conjecture states that starting with any initial n value, the sequence always ends in the cycle (4, 2, 1). A standard technique in mathematics is to study generalizations of apparently intractable problems. Let d be an odd natural number not divisible by 3. In the following, the $3n+d$ sequence is considered. As is standard in the mathematical literature, the element after an odd element n is defined to be $(3n+d)/2$. Let K be the number of odd elements in a $3n+d$ cycle and L the number of even elements. For a given d value, cycles having the same length $(K+L)$ and the same number of odd elements will be said to be interrelated. When counting cycles for a given d value, only one of the interrelated cycles will be included. In this case, a Poisson probability

distribution ($P(x) = (\lambda^x/x!)e^{-\lambda}$, $x = 0, 1, 2, \ldots$) where $\lambda = 1$ can be used to model the number of $3n + d$ cycles for a given d value. (For this parameter value, the $x = 1$ count should equal the $x = 0$ count, the $x = 2$ count should equal half the $x = 1$ count, the $x = 3$ count should equal a third of the $x = 2$ count, etc.) Belaga and Mignotte [1] have tabulated the known primitive $3n + d$ cycles for d less than 20000 (a cycle is said to be primitive if its elements are not a common multiple of the elements of another cycle). They state that they believe these are the only such cycles that exist. For d less than 100, the Poisson probability distribution models the number of cycles quite well; there are 14 d values where there is only 1 cycle, there are 12 d values where there are exactly 2 cycles, there are 6 d values where there are exactly 3 cycles, and there is 1 d value where there are exactly 4 cycles (again, taking interrelated cycles into account). (They also include $3n-1$ cycles in their table. Here, these cycles are counted with those for $d = 1$.) Note that a cycle count of 1 corresponds to $x = 0$, a cycle count of 2 corresponds to $x = 1$, etc.

In the remainder of this article, larger samples are considered and n is allowed to be an integer (positive or negative). (Consequently, the cycles up to 10000 have been recomputed by the author [2, "test23h" in the software section]. There are new cycles that do not appear in Belaga and Mignotte's table.) There does not appear to be any mention of Poisson probability distributions in Lagarias' [3, 4] annotated bibliographies of the $3x + 1$ problem. (Lagarias [5] also gives the history of the $3x+1$ problem and surveys the literature concerning the problem. By way of a collection of papers, Lagarias [6] reports on what is currently known on the $3x + 1$ problem.) Davison [7] presented a probabilistic argument concerning the number of circuits required to iterate from a number n to 1; this appears to be the only probabilistic argument concerning potential cycles (or cycles) in the literature.

2 Interrelated Cycles

A parity vector gives the order of even and odd elements in a $3n + d$ sequence, a "1" for an odd element and a "0" for an even element. Let $k_0, k_1, k_2, \ldots, k_{K-1}$ denote the positions of the 1's in a parity vector containing K 1's and L 0's. Let

$$Z := 3^{K-1}2^{k_0} + 3^{K-2}2^{k_1} + 3^{K-3}2^{k_2} + \cdots + 3^0 2^{k_{K-1}}.$$

A cycle exists if and only if $dZ/(2^{K+L} - 3^K)$ is an integer (see Bohm and Sontacchi [8]). Note that every parity vector containing at least one 1 corresponds to a cycle for some d value (duplicated parity sub-vectors correspond to duplicated sub- cycles). For example, for the $d = 17$ cycle of (2, 1, 10, 5, 16, 8, 4), $k_0 = 1$ and $k_1 = 3$, $Z = 3^1 2^1 + 3^0 2^3 = 14$, $119 = 2^7 - 3^2$, and $(17 \cdot 14)/119$ is an integer. There do not appear to be any factors of d left over after the division by $2^{K+L} - 3^K$. This has been confirmed experimentally (for the 7429 distinct (d, L, K) values of the $3n + d$ cycles for $d < 10000$):

Conjecture 1. A $3n + d$ cycle exists only if d divides $2^{K+L} - 3^K$.

If this conjecture holds, Z need not be considered when searching for cycles; only the factors of $2^{K+L} - 3^K$ need be considered. When $d = |2^{K+L} - 3^K|$, an arbitrarily constructed parity vector with a length of $K + L$ and containing K 1's corresponds to a cycle (or possibly

duplicated sub-cycles if L and K are not relatively prime), but the cycle is not guaranteed to be primitive. When reduced, such a cycle corresponds to a cycle for a d value that is a proper divisor of $2^{K+L} - 3^K$. In this sense, there is no problem of finding cycles; they're all well-defined and determined by the parity vectors. Even the number of interrelated cycles is determined by the combinatorics of generating parity vectors that are distinct under rotation. For example, for L and K greater than or equal to 1 and less than or equal to 10, the numbers of distinct parity vectors are as follows:

	$K=1$	2	3	4	5	6	7	8	9	10
$L=1$	1	1	1	1	1	1	1	1	1	1
2	1	2	2	3	3	4	4	5	5	6
3	1	2	4	5	7	10	12	15	19	22
4	1	3	5	10	14	22	30	43	55	73
5	1	3	7	14	26	42	66	99	143	201
6	1	4	10	22	42	80	132	217	335	504
7	1	4	12	30	66	132	246	429	715	1144
8	1	5	15	43	99	217	429	810	1430	2438
9	1	5	19	55	143	335	715	1430	2704	4862
10	1	6	22	73	201	504	1144	2438	4862	9252

There is, however, the problem of determining which d values the primitive cycles map to. For example, for $K + L = 1$ and $K = 1$, the parity vector is (1) (corresponding to the cycle (-1) for $d = 1$), for $K + L = 2$ and $K = 1$, the parity vector is (0, 1) (corresponding to the cycle (2, 1) for $d = 1$), and for $K + L = 3$ and $K = 2$, the parity vector is (1, 1, 0) (corresponding to the cycle $(-5, -7, -10)$ for d=1). For $K + L = 11$ and $K = 7$, there are 30 distinct parity vectors, the first few of which are (0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0), (1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0), (1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0), ... (corresponding to cycles for $d = 139$). The cycle corresponding to the first parity vector is not primitive (when $d = 139$) and corresponds to the remaining known $d = 1$ cycle of $(-34, -17, -25, -37, -55, -82, -41, -61, -91, -136, -68)$. Another example factorization will be given. For ($K + L$, K)=(8, 4), there are 10 distinct parity vectors (corresponding to cycles for $d = 175$). The parity vector (1, 1, 0, 0, 1, 1, 0, 0) corresponds to two duplicated primitive cycles for $d = 7$, the parity vector (0, 1, 0, 1, 0, 1, 0, 1) corresponds to four duplicated primitive cycles for $d = 1$, the parity vector (1, 0, 1, 1, 0, 1, 0, 0) corresponds to a primitive cycle for $d = 25$, the parity vectors (0, 0, 1, 1, 1, 1, 0, 0) and (0, 1, 1, 0, 1, 1, 0, 0) correspond to primitive cycles for $d = 35$, and the parity vectors (1, 0, 0, 1, 1, 1, 0, 0), (0, 1, 0, 1, 1, 1, 0, 0), (1, 0, 1, 0, 1, 1, 0, 0), (1, 1, 0, 1, 0, 1, 0, 0), and (0, 1, 1, 1, 0, 1, 0, 0) correspond to primitive cycles for $d = 175$. Although 5 also divides 175, there are no cycles for this d value when ($K + L$, K)=(8, 4) (the parity vectors have been used up by the cycles for larger d values and $d = 1$).

Apparently, every d value is covered in this mapping. A more appropriate question to ask than why cycles do not exist for a given d value is why they do exist and what the expected number of cycles is. For the 1667 d values less than or equal to 4999, the number of d values having 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11 cycles (counting only one of interrelated cycles)

are 535, 607, 310, 115, 57, 22, 9, 5, 4, 2, and 1 respectively. This distribution has a mean of 1.235. For the 3333 d values less than or equal to 9997, the number of d values having 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15 cycles (counting only one of interrelated cycles) are 1117, 1193, 591, 228, 104, 48, 25, 11, 5, 5, 2, 2, 0, 1, and 1 respectively. This distribution has a mean of 1.229. This resembles a Poisson probability distribution where $\lambda = 1$, but the counts for the larger x values are too big.

3 Associated Cycles

This raises the question of whether there are other ways for cycles to be interrelated so that the data would more closely fit a Poisson probability distribution. There are frequently (L, K) values of the cycles for a given d value that are multiples of other (L, K) values. For example, for $d = 269$, the (L, K) values of the cycles are (4, 5), (8, 10), (12, 15), and (16, 20). Also, there are frequently (L, K) values of the cycles for a given d value that are in arithmetic progression (and are not multiples of other (L, K) values). For example, for $d = 1843$, the sorted (L, K) values (by increasing L values) of the cycles are (6, 11), (12, 22), (18, 9), (24, 20), (30, 31), (36, 42), (42, 53), (48, 64), (54, 51), and (90, 117) ((18, 9), (24, 20), (30, 31), (36, 42), (42, 53), and (48, 64) are in arithmetic progression with an increment of (6, 11)). Empirical evidence supports the following

Conjecture 2. For a given d value, there are cycles with (L, K) values that are in arithmetic progression (excluding arithmetic progressions consisting of multiples of an (L, K) value) only if one of the $|L - K|$ values is relatively small.

The term "relatively small" will be defined in more detail in the next section. The (L, K) values of just a pair of cycles are considered to be in arithmetic progression if the difference between them is the (L, K) value of another cycle. Let (L_1, K_1), (L_2, K_2), (L_3, K_3), ..., denote the sorted (L, K) values. Cycles for a given d value with (L, K) values that are in arithmetic progression with an increment of (L_1, K_1) (and aren't multiples of (L_1, K_1)) will be said to be associated with each other. (Increments of (L_2, K_2), (L_3, K_3), etc. will be considered later.) In the case of $d = 1843$, the cycles with (L, K) values of (24, 20), (30, 31), (36, 42), (42, 53), and (48, 64) will not be counted when computing the frequency distribution of cycle counts. For $d = 4879$, the sorted (L, K) values of the cycles are (11, 14), (22, 28), (25, 10), (33, 42), (36, 24), (44, 56), (58, 52), (66, 84), (69, 66), (77, 98), and (80, 80). Note that the (L, K) values that are in arithmetic progression need not be contiguous; (25, 10) and (36, 24) are in arithmetic progression and (58, 52), (69, 66), and (80, 80) are in arithmetic progression (with an increment of (11, 14)). In this case, the cycles with (L, K) values of (36, 24), (69, 66), and (80, 80) will not be counted. ($2^{K_1+L_1} - 3^{K_1}$ is frequently negative, in which case the corresponding cycles do not appear in Belaga and Mignotte's table.)

Let (L_b, K_b) denote the first (L, K) value in the sorted (L, K) values for which the arithmetic progression(s) begins. Of the 3333 d values less than or equal to 9997, associated cycles occur for 328 d values and $L_1 > K_1$ and $L_b < K_b$ or $L_1 < K_1$ and $L_b > K_b$ in all but 20 instances. For 17 of these d values, there are only two (L, K) values that are in arithmetic

progression (with an increment of (L_1, K_1)). For example, for $d = 1679$, the sorted (L, K) values are (20, 28), (37, 32), (97, 116), and (117, 144) ((97, 116) and (117, 144) are in arithmetic progression with an increment of (20, 28)). For the remaining 3 d values, there are two pairs of (L, K) values each that are in arithmetic progression. When $d = 2305$, the sorted (L, K) values are (30, 17), (40, 38), (50, 59), (70, 55), and (80, 76). The pairs of (L, K) values that are in arithmetic progression are ((40, 38), (70,55)) and ((50, 59), (80, 76)) (note that $2 \cdot (40, 38) = (80, 76)$). In this case, $30 > 17$ ((30, 17) equals (L_1, K_1)) and $50 < 59$ ((50, 59) is the (L, K) value starting the second arithmetic progression). When $d = 5113$, the sorted (L, K) values are (39, 27), (42, 40), (45, 53), (48, 66), (81, 67), (84, 80), and (180, 212). The pairs of (L, K) values that are in arithmetic progression are ((42, 40), (81, 67)) and ((45, 53), (84, 80)) (note that $2 \cdot (42, 40) = (84, 80)$). In this case, $39 > 27$ ((39, 27) equals (L_1, K_1)) and $45 < 53$ ((45, 53) is the (L, K) value starting the second arithmetic progression). When $d = 6989$, the sorted (L, K) values are (32, 50), (52, 55), (72, 60), (84, 105), (104, 110), (124, 115), and (188, 215). The pairs of (L, K) values that are in arithmetic progression are ((52, 55), (84, 105)) and ((72, 60), (104, 110)) (note that $2 \cdot (52, 55) = (104, 110)$). In this case, $32 < 50$ ((32, 50) equals (L_1, K_1)) and $72 > 60$ ((72, 60) is the (L, K) value starting the second arithmetic progression).

Even including the 20 d values where $L_1 > K_1$ and $L_b > K_b$ or $L_1 < K_1$ and $L_b < K_b$, $|L - K|$ is relatively small for some (L, K) value in the arithmetic progression(s) or $|L_1 - K_1|$ is relatively small. For example, for $d = 601$, the sorted (L, K) values of the cycles are (1, 6), (13, 3), (14, 9), (15, 15), (17, 27), (29, 24), (32, 42), and (33, 48) ((15, 15) is in the arithmetic progression). For $d = 2231$, the sorted (L, K) values of the cycles are (30, 31), (48, 32), (54, 91), (60, 62) and (84, 122) and in this case the $|L - K|$ values in the arithmetic progression(s) ($|54 - 91|$ and $|84 - 122|$) do not appear to be relatively small, but $|30 - 31|$ ($|L_1 - K_1|$) is relatively small.

These properties of associated cycles are the rationale for not considering the cycles with (L, K) values that are in arithmetic progression to be random. For the 1667 d values less than or equal to 4999, the numbers of d values having 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 cycles (counting only one of interrelated or associated cycles) are 535, 664, 324, 89, 39, 13, 1, 1, 1, and 0 respectively. This distribution has a mean of 1.092. For the 3333 d values less than or equal to 9997, the numbers of d values having 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12 cycles (counting only one of interrelated or associated cycles) are 1117, 1319, 594, 185, 76, 31, 6, 2, 1, 0, 1, and 1 respectively. This distribution has a mean of 1.080. The associated cycles then account for most of the deviation of the distribution of cycle counts away from a Poisson probability distribution where $\lambda = 1$. A few of the counts for the larger x values are still too big, but this is due to cycles where there are many (L, K) values that are multiples of other (L, K) values. For example, for $d = 7153$, the (L, K) values of the cycles are (7, 12), (14, 24), (21, 36), (28, 48), (35, 60), (42, 72), (56, 96), (63, 108), (271, 221), (292, 257), (341, 341), and (362, 377).

4 The Minimum $|L - K|$ Value of the $3n + d$ Cycles for a Given d Value

For the d values less than or equal to 9997, there is only 1 cycle (counting only one of interrelated cycles) for 1117 d values and the mean and standard deviation of the $L - K$ values are 1.271 and 17.729 respectively. For the d values less than or equal to 9997, there are exactly 2 cycles (counting only one of interrelated cycles) for 1193 d values and the mean and standard deviation of the $L-K$ values of the smaller $|L-K|$ values are 0.082 and 9.116 respectively. For the d values less than or equal to 9997, there are exactly 3 cycles (counting only one of interrelated cycles) for 591 d values and the mean and standard deviation of the $L - K$ values of the smallest $|L - K|$ values are 0.509 and 6.421 respectively. For the d values less than or equal to 9997, there are exactly 4 cycles (counting only one of interrelated cycles) for 228 d values and the mean and standard deviation of the $L - K$ values of the smallest $|L - K|$ values are -0.355 and 4.576 respectively. In this case, the numbers of d values for which $L - K$ equals -11, -10, -9, ..., 13 are 1, 4, 4, 6, 4, 12, 8, 25, 10, 7, 12, 54, 13, 11, 13, 12, 8, 14, 1, 2, 1, 0, 1, 4, and 1 respectively. This appears to be a normal probability distribution; the 95% confidence interval of the mean is (-0.952, 0.242) and the 95% confidence interval of the standard deviation is (4.191, 5.039). The corresponding standard deviations when there are 5, 6, 7, and 8 cycles (counting only one of interrelated cycles) are 3.474, 2.944, 2.934, and 2.401 respectively (when there are j cycles, the standard deviation is about $1/j$ times the standard deviation when there is 1 cycle).

The decrease in standard deviation of these distributions as the number of cycles increases appears to account for the existence of a cycle for every d value (the probability that there are cycles where $|L - K|$ is small is still relatively high). Apparently, if there is not exactly one cycle (counting only one of interrelated cycles) for a given d value, then there are j cycles where $|L - K|$ for one of the cycles is relatively small (the distribution of $L - K$ values of cycles for d values where there is only one cycle determines what "relatively small" is). For example, for the d values less than or equal to 97, there is only one cycle for $d = 7$, 35, 37, 41, 43, 53, 65, 67, and 89 and the respective (L, K) values are (2, 2), (4, 4), (3, 3), (12, 8), (8, 3), (12, 17), (12, 12), (14, 16), and (9, 8). The mean and standard deviation of the distribution of $L - K$ values are 0.333 and 2.958 respectively. For the d values less than or equal to 97, there are exactly two cycles for $d = 19$, 31, 49, 59, 61, 71, 77, 79, 83, 85, and 91 and the respective (L, K) values of the pairs of cycles are ((6, 5), (6, 11)), ((3, 6), (11, 12)), ((1, 4), (16, 22)), ((4, 6), (17, 11)), ((5, 1), (25, 41)), ((5, 5), (10, 17)), ((8, 14), (22, 16)), ((9, 14), (24, 20)), ((5, 7), (10, 14)), ((4, 8), (44, 56)), and ((6, 6), (24, 24)). The mean and standard deviation of the $L - K$ values of the smaller $|L - K|$ values are -0.818 and 3.093 respectively. The standard deviation is larger than expected, but if the cycles for $d = 77$ (where $|L-K| = 6$) are considered to be outliers, the standard deviation is 2.710 (still larger than expected, but at least smaller than the standard deviation of the distribution of $L-K$ values when there is only one cycle). For the d values less than or equal to 97, there are exactly three cycles for $d = 5$, 11, 13, 17, 23, 25, 47, 55, 73, 95, and 97 and the respective (L, K) values of the triplets of cycles are ((2, 1), (2, 3), (10, 17)), ((1, 3), (4, 2), (6, 8)), ((3, 1), (3, 5), (9, 15)), ((2, 4), (5, 2), (13, 18)), ((3, 2), (7, 12), (17, 26)), ((4, 4), (8, 8), (12, 22)), ((3, 4), (11, 7), (12, 16)), ((4, 2), (8, 4), (12, 16)), ((7, 8), (8, 4), (28, 32)), ((6,

5), (6, 11), (36, 36)), and ((6, 3), (6, 11), (12, 6)). The mean and standard deviation of the $L - K$ values of the smallest $|L - K|$ values are 0.273 and 1.679 respectively. The standard deviation is somewhat larger than expected but is smaller than the standard deviation of the distribution of $L - K$ values of the smaller $|L - K|$ values when there are two cycles. For the d values less than or equal to 97, there are exactly four cycles for $d = 1$ and 29 and the respective (L, K) values of the quadruplets of cycles are ((0, 1), (1, 1), (1, 2), (4, 7)) and ((4, 1), (4, 8), (8, 9), (24, 41)). The mean and standard deviation of the $L - K$ values of the smallest $|L - K|$ values are -0.500 and 0.707 respectively. This is approximately equal to the expected standard deviation. (Of the d values less than or equal to 97, these are the only d values for which associated cycles occur.) For the d values less than or equal to 9997, the standard deviations of the $L - K$ values of the smallest $|L - K|$ values when (L_1, K_1), (L_2, K_2), (L_3, K_3), ..., (L_i, K_i), $i = 1$, 2, 3, and 4, are not included and there are 2, 3, 4, 5, 6, 7, and 8 cycles are as follows:

# cycles=	2	3	4	5	6	7	8
$i = 1$	19.450	10.005	6.957	4.624	3.845	3.265	2.618
2	n/a	18.561	10.298	6.569	5.240	3.736	2.945
3	n/a	n/a	16.630	8.623	6.520	5.585	3.165
4	n/a	n/a	n/a	16.618	9.038	6.055	6.472

Not including the sorted (L, K) values does not change the ratios of the standard deviations much.

For the 3333 d values less than or equal to 9997, there are 472 d values where $L = K$ for at least one cycle. For the d values less than or equal to 9997, there are 1090 d values where $|L - K| \leq 3$ for at least one cycle. This is additional evidence that for a given d value, a cycle always exists for some (L, K) value where $|L - K|$ is relatively small.

In general, the $L - K$ values appear to be normally distributed (although there is a tendency for $L - K$ to be a multiple of 6). For example, the $L - K$ values of the 380 cycles (counting only one of interrelated cycles) for the 167 d values less than or equal to 499 have a normal probability distribution with a mean of -2.087 (the 95% confidence interval is $(-2.893, -1.280)$) and a standard deviation of 7.996 (the 95% confidence interval is (7.465, 8.609)).

5 Associated Cycles and the Minimum $|L - K|$ Value of the Cycles for a Given d Value

The standard deviation of the $L - K$ values of the smallest $|L - K|$ values of the cycles for d values where there aren't associated cycles is about the same as that for d values where there are associated cycles (for d values having at least three cycles). For the d values less than or equal to 9997 where there are no associated cycles, the means and standard deviations of the $L - K$ values of the smallest $|L - K|$ values when there are 3, 4, 5, 6, 7, and 8 cycles are (0.559, 6.422), $(-0.830, 4.501)$, $(-1.157, 3.367)$, $(-0.833, 2.230)$, $(-3.600, 3.782)$, and $(-2.000$, n/a) respectively (the numbers of cycles are 483, 135, 51, 18, 5, and 1 respectively). For the d values less than or equal to 9997 where there are associated cycles, the means and

standard deviations of the $L-K$ values of the smallest $L-K$ values when there are 3, 4, 5, 6, 7, and 8 cycles are (0.287, 6.442), (0.333, 4.619), (0.566, 3.394), (−0.100, 3.305), (−0.500, 2.417), and (0.000, 2.450) respectively (the numbers of cycles are 108, 93, 53, 30, 20, and 10 respectively). In this respect, there is nothing exceptional about associated cycles. In all but 21 of the 328 d values where d is less than or equal to 9997 and there are associated cycles, $|L_1-K_1|$ or the smallest $|L-K|$ value in the arithmetic progression(s) is the smallest $|L-K|$ value of all the cycles for the d value.

When $d = 643$, the sorted (L, K) values of the cycles are (14, 5), (16, 21), (30, 26), (46, 47) and (64, 84). In this case, (16, 21) ((L_2, K_2)) is the increment between (30, 26) and (46, 47) and $|46-47|$ is smaller than $|30-26|$ (the smallest $|L-K|$ value in the arithmetic progression(s) with an increment of (L_1, K_1)). Similar exceptions occur for d values of 3881, 4297, 4393, 5233, 5257, 5605, 6499, and 8189. When $d = 2045$, the sorted (L, K) values of the cycles are (10, 1), (16, 22), (26, 23), (42, 45), and (68, 68). In this case, (26, 23) ((L_3, K_3)) is the increment between (42, 45) and (68, 68) and $|68-68|$ is smaller than $|26-23|$ (the smallest $|L-K|$ value in the arithmetic progression(s) with an increment of (L_1, K_1)). Similar exceptions occur for $d = 3005$ and 3493. When $d = 6563$, the sorted (L, K) values of the cycles are (11, 15), (22, 30), (33, 45), (47, 29), (55, 75), (58, 44), and (102, 104). In this case (47, 29) ((L_4, K_4)) is the increment between (55, 75) and (102, 104) and $|102-104|$ is smaller than $|11-15|$ ($|L_1-K_1|$). When $d = 3985$, the sorted (L, K) values of the cycles are (50, 27), (54, 61), (104, 88), (106, 105), and (108, 122). The (L, K) values of (104, 88), (106, 105), and (108, 122) are in arithmetic progression with an increment of (2, 17) (half of (54, 61)−(50, 27)) and $|106-105|$ is smaller than $|54-61|$ (the smallest $|L-K|$ value in the arithmetic progression(s) with an increment of (L_1, K_1)). When $d = 5555$, the sorted (L, K) values of the cycles are (40, 30), (80, 60), (80, 110), (100, 100), (120, 140), and (160, 170). In this case, (40, 30) ((L_1, K_1)) is the increment between (80, 110) and (120, 140), (80, 60) ((L_2, K_2)) is the increment between (80, 110) and (160, 170), and (100, 100) does not occur in these arithmetic progressions.

When $d = 6497$, the sorted (L, K) values of the cycles are (77, 88), (79, 80), (81, 72), and (158, 160). In this case, $|79-80|$ ($|L_2-K_2|$) is smaller than $|158-160|$ (the smallest $|L-K|$ value in the arithmetic progression(s) with an increment of (L_1, K_1)). When $d = 9673$, the sorted (L, K) values of the cycles are (72, 64), (79, 86), (86, 108), (93, 130), and (158, 172). In this case, the smallest $|L-K|$ values occur for (L_1, K_1) and (L_2, K_2) and $|L_2-K_2|$ is smaller than $|L_1-K_1|$. In both of these cases, there is no arithmetic progression of (L, K) values with an increment of (L_2, K_2).

When $d = 503$, the sorted (L, K) values of the cycles are (7, 2), (25, 43), (32, 45), (39, 47), and (53, 51). The smallest $|L-K|$ value occurs for (53, 51) where the gap between (39, 47) and (53, 51) is $2 \cdot (L_1, K_1)$. Similar exceptions occur for $d = 1487$, 1679, and 9997.

6 The Minimum $|L-K|$ Value of the Cycles for a Given d Value and (L, K) Trees

For $d = 121$, the sorted (L, K) values of the cycles are (8, 14), (16, 28), (26, 18), and (42, 46). In this case, (26, 18) and (42, 46) are in arithmetic progression with an increment of

(16, 28) ((L_2, K_2)) and $|42-46|$ is the smallest $|L-K|$ value. Note that no associated cycles occur for this d value. In general, (L_2, K_2) can be used as the beginning of a "branch" in a tree of (L, K) values, the branch consisting of arithmetic progressions of (L, K) values with an increment of (L_2, K_2) (excluding arithmetic progressions consisting of multiples of (L_2, K_2)). Cycles to be included when computing the frequency distribution of cycle counts are determined just as they are for associated cycles. When comparing $|L - K|$ values, the (L, K) values in the arithmetic progression(s) and (L_2, K_2) are considered (just as they are for associated cycles). Similarly, (L_3, K_3) can be used as the beginning of a branch, etc., so that the totality of these arithmetic progressions (including associated cycles) comprise the (L, K) tree.

When it exists, the (L, K) tree for a given d value almost always accounts for the minimum $|L - K|$ value of the cycles if in the case where there are no associated cycles $|L_1 - K_1|$ is also considered (when there are no associated cycles, the first few sorted (L, K) values are usually multiples of (L_1, K_1)). (L, K) trees do not account for the minimum $|L - K|$ values of the cycles for d values of 503, 1487, 1679, 3985, 5555, 6497, 9673, and 9997 (discussed in the previous section) and the minimum $|L - K|$ value of the cycles for $d = 6095$ (where the sorted (L, K) values of the cycles are (4, 10), (8, 20), (192, 194), (204, 224), and (212, 244)). For the d values less than or equal to 9997, there are (L, K) trees for 396 d values. For the 1667 d values less than or equal to 4999, the numbers of d values having 1, 2, 3, 4, 5, 6, and 7 cycles (taking interrelated cycles and the cycles in (L, K) trees into account) are 535, 670, 353, 85, 20, 4, and 0 respectively. This distribution has a mean of 1.0384. For the 3333 d values less than or equal to 9997, the numbers of d values having 1, 2, 3, 4, 5, 6, 7, and 8 cycles (taking interrelated cycles and the cycles in (L, K) trees into account) are 1117, 1338, 635, 177, 58, 6, 2, and 0 respectively. This distribution has a mean of 1.0240. For a Poisson probability distribution with this parameter, the expected numbers of d values having 1, 2, 3, 4, 5, 6, 7, and 8 cycles would be 1197, 1226, 628, 214, 55, 11, 2, and 0 respectively.

7 The Number of Prime Factors of $2^{K+L} - 3^K$

In the following, duplicated prime factors of $2^{K+L} - 3^K$ are also counted. For example, $2^{10} - 3^8 = -7^2 \cdot 113$ is considered to have 3 prime factors. The number of prime factors of a natural number n where the prime factors are not necessarily distinct is denoted by $\Omega(n)$ (see Hardy and Wright [9, p. 354]). For $1 \le L, K \le 10$, the numbers of values of $2^{K+L} - 3^K$ where $\Omega(|2^{K+L} - 3^K|)$ equals 1, 2, 3, 4, and 5 are 37, 36, 22, 3, and 0 respectively (there are no prime factors of $2^{K+L} - 3^K$ in two instances, so the number of samples is $98 = 10^2 - 2$). A Poisson probability distribution where $\lambda = 0.91$ can be used to model this data. Corresponding data for larger L, K upper bounds is

# prime factors=	1	2	3	4	5	6	7
L, K upper bound=12	48	54	28	8	3	1	0
14	56	75	41	17	4	1	0
16	61	100	54	28	7	4	0

Poisson probability distributions where $\lambda = 1.06$, 1.18, and 1.34 can be used to model the data for respective L, K upper bounds of 12, 14, and 16. These Poisson probability distributions model the number of prime factors of $2^{K+L} - 3^K$ quite well. For example, for a Poisson probability distribution where $\lambda = 1.34$, the relative frequencies for $x = 0$, 1, 2, 3, 4, 5, and 6 are 0.262, 0.351, 0.235, 0.105, 0.035, 0.009, and 0.002 respectively (for $1 \leq L, K \leq 16$, the corresponding relative frequencies of the $\Omega(|2^{K+L} - 3^K|)$ values are 0.240, 0.393, 0.213, 0.110, 0.028, 0.016, and 0.000 respectively). The number of prime factors of $2^{K+L} - 3^K$ affects the number of d values covered by the parity vectors corresponding to $(K + L, K)$.

References

[1] Edward G. Belaga and Maurice Mignotte, The Collatz problem and its generalizations: Experimental data. Table 1. Primitive cycles of $3n + d$- uappings, available at http://hal.archives-ouvertes.fr/hal-00129727.

[2] Darrell Cox, The $3n + 1$ problem, http://home.graysoncable.com/dkcox.

[3] Jeffrey C. Lagarias, The $3x + 1$ problem: an annotated bibliography (1963-1999), http://arxiv.org/abs/math.NT/0309224.

[4] Jeffrey C. Lagarias, The $3x + 1$ problem: an annotated bibliography. II (2000–2009), http://arxiv.org/abs/math.NT/0608208.

[5] J. C. Lagarias, The $3x + 1$ problem and its generalizations, *Amer. Math. Monthly* **92** (1985), 3–23.

[6] J. C. Lagarias, ed., *The Ultimate Challenge: The* $3x+1$ *Problem*, American Mathematical Society, 2010.

[7] Leslie J. Davison, Some comments on an iteration problem, *Proc. 6-th Manitoba Conf. On Numerical Mathematics and Computing (Univ. of Manitoba-Winnipeg 1976)*, Congressus Numerantium **18**, Utilitas Math., 1977, pp. 55–59.

[8] C. Böhm and G. Sontacchi, On the existence of cycles of given length in integer sequences like $x_{n+1} = x_n/2$ if x_n even, and $x_{n+1} = 3x_n + 1$ otherwise, *Atti Accad. Naz. Lincei*, **8** *Ser., Rend., Cl. Sci. Fis. Mat. Nat.* **64** (1978), 260–264.

[9] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford, 1960.

2010 *Mathematics Subject Classification*: Primary 11B99.
Keywords: Collatz conjecture, $3x + 1$ problem.

(Concerned with sequence A008876.)

Received February 20 2012; revised version received May 22 2012. Published in *Journal of Integer Sequences*, May 28 2012.

Return to Journal of Integer Sequences home page.