Density Bounds for the 3x+1 Problem
I. Tree-Search Method

David Applegate
Jeffrey C. Lagarias

AT&T Bell Laboratories
Murray Hill, NJ 07974

(December 15, 1993)

(Dedicated to the memory of D. H. Lehmer)

1. Introduction

The $3x+1$ problem concerns the iteration of the function $T : \mathbb{Z} \to \mathbb{Z}$ defined by

$$T(x) = \begin{cases} \dfrac{3x+1}{2} & \text{if } x \equiv 1 \pmod 2 , \\ \dfrac{x}{2} & \text{if } x \equiv 0 \pmod 2 . \end{cases} \tag{1.1}$$

The $3x+1$ Conjecture asserts that, for all $n \geq 1$, some iterate $T^{(k)}(n) = 1$. More generally, it is conjectured that T has finitely many cycles under iteration, and that every $n \in \mathbb{Z}$ eventually enters a cycle, cf. Lagarias [6]. The $3x+1$ Conjecture has been verified for all $n < 5.6 \times 10^{13}$ by Leavens and Vermeulen [8].

One approach to these questions is to study how many integers n below a given bound x have some $T^{(k)}(n) = 1$. More generally, for any $a \in \mathbb{Z}$, set

$$\pi_a(x) = \#\{n : |n| \leq x \text{ and some } T^{(k)}(n) = a , \ \ k \geq 0\} . \tag{1.2}$$

It is well-known that the growth of $\pi_a(x)$ depends on the residue class $a \pmod 3$. If $a \equiv 0 \pmod 3$, then the preimages of a under iterates of T are exactly $\{2^k a : k \geq 1\}$; hence $\pi_a(x)$ grows logarithmically with x. The other cases are covered by:

Conjecture A. *For each $a \not\equiv 0 \pmod 3$, there is a positive constant c_a such that*

$$\pi_a(x) \ge c_a x \quad \textit{for all } x \ge |a| \ .$$

In any case one has, for $a \not\equiv 0 \pmod 3$,

$$\pi_a(x) \ge x^{\gamma} \quad \text{for } x \ge x_0(a) \ , \tag{1.3}$$

for some constant $\gamma > 0$, as was first shown by Crandall [3], with $\gamma = .05$. Crandall's approach directly studies the tree of preimages of a under T. Sander [9] strengthened Crandall's approach to obtain $\gamma = .30$. Krasikov [5] introduced a different method which derives a system of difference inequalities with variables associated to congruence classes (mod 3^k). Using these inequalities for $k = 2$, he obtained $\gamma = .43$. Wirsching [10] used Krasikov's inequalities with $k = 3$ to obtain $\gamma = .48$.

In studying $\pi_a(x)$, a related problem concerns the size of the tree of preimages of a under T. Let

$$n_k(a) := \#\{n : T^{(k)}(n) = a\} \ . \tag{1.4}$$

Lagarias and Weiss (1992) prove a result implying that, for $a \not\equiv 0 \pmod 3$, the average size of $n_k(a)$ as a varies is $\frac{3}{2}\left[\frac{4}{3}\right]^k$. They conjectured:

Conjecture B. *For each $a \not\equiv 0$ (mod 3),*

$$n_k(a) = \left[\frac{4}{3}\right]^{k(1+o(1))} \quad \textit{as } k \to \infty \ . \tag{1.5}$$

For a not in a cycle, they showed that

$$\frac{1}{2}(\sqrt[4]{2})^k \le n_k(a) \le 2 \cdot (\sqrt{3})^k \ , \tag{1.6}$$

by studying all possible trees of backward iterates of depth 4.

The object of this paper and its sequel is to obtain improved bounds for $\pi_a(x)$ and $n_k(a)$, using computer-assisted proofs. This paper obtains bounds based on the tree-search approach started by Crandall, while the sequel obtains bounds for $\pi_\alpha(x)$ derived from Krasikov's difference inequalities.

In §2 we study the trees $\mathcal{T}^*_k(a)$ containing all $n \not\equiv 0 \pmod 3$ with $T^{(j)}(n) = a$ for some $j \le k$. The structure of this tree depends only on $a \pmod{3^{k+1}}$. Each leaf n of the tree is assigned a *weight* which counts the number of iterates $T^{(i)}(n) \equiv 1 \pmod 2$, for $0 \le i \le k-1$. By computer search we find, for all $k \le 30$, upper and lower bound statistics concerning the number of leaves of such trees having a fixed weight. An immediate consequence is:

Theorem 1.1. *For any $a \not\equiv 0$ (mod 3), and for all sufficiently large k,*

$$(1.302053)^k \le n_k(a) \le (1.358386)^k \ . \tag{1.7}$$

The proof of Theorem 1.1 is unavoidably computer-intensive; in effect it searches all trees of depth 30.

The upper bound and lower bound statistics for number of leaves lie within a small constant factor of $(\frac{4}{3})^k$. They appear to have a much narrower distribution than that predicted by branching process models for $3x+1$ trees studied in [7], as we show in detail elsewhere [2].

In §3 we use Chernoff bounds to obtain lower bounds for the number of leaves in such trees having a large weight and use this to get lower bounds for the exponent γ in (1.3). Using trees of depth k we obtain a bound γ_k^* by optimizing a ''large deviations'' bound for the number of heavily weighted leaves in a ''worst-case'' tree of depth k. In this fashion using $k = 30$ we obtain:

Theorem 1.2. *For each $a \not\equiv 0$ (mod 3), there is a positive constant c_a such that*

$$\pi_a(x) \ge c_a \, x^{.65} \ \text{for } all \ x \ge |a| \ . \tag{1.8}$$

This exponent improves on previous bounds; however in part II we will show that Krasikov's inequalities give still better exponents.

In §3 we also obtain upper bounds for the number of leaves in any tree $7_k^*(a)$ that have a large weight. Korec [4] showed that the set $\{n : \text{some } |T^{(k)}(n)| < |n|^{\beta}\}$ has density one for all $\beta > \beta_c := \frac{\log 3}{\log 4}$. We describe an approach to lower the bound β_c using such upper bound estimates. This approach becomes effective, however, only if a certain threshold is exceeded, and it is not reached by tree depth $k = 30$.

We are indebted to T. H. Foregger and an anonymous referee for a critical reading and helpful comments.

2. 3x+1 Trees

In this section we always suppose that $a \not\equiv 0 \pmod 3$. The preimages under T^{-1} of any integer form an infinite labelled tree $7(a)$, whose root node is labelled a and whose nodes at the k-th level are labelled $\{n : T^{(k)}(n) = a\}$. Note that if a is not in a cycle, then no two nodes of $7(a)$ have the same label, while if a is in a cycle then labels will be repeated. The tree $7(a)$ is constructed recursively using the multivalued operator

$$T^{-1}(n) = \begin{cases} \{2n\} & \text{if } n \equiv 0,1 \pmod 3 \\ \{2n\,,\ \dfrac{2n-1}{3}\} & \text{if } n \equiv 2 \pmod 3\ . \end{cases}$$

Each node n at level k of the tree is connected to one or two nodes, labelled with the labels in $T^{-1}(n)$, at depth $k+1$ of the tree.

In studying asymptotic properties of $n_k(a)$, it proves convenient to throw out all preimages $n \equiv 0 \pmod 3$, and to estimate instead the quantity

$$n_k^*(a) := \#\{ n : T^{(k)}(n) = a \ \text{ and } \ n \not\equiv 0 \pmod 3 \} \ . \tag{2.1}$$

It is easy to show that

$$n_k^*(a) \le n_k(a) \le k\, n_k^*(a) \ ,$$

see Lemma 3.1 of [7], hence $n_k^*(a)$ and $n_k(a)$ have similar exponential growth in k as $k \to \infty$.

Thus, following [7], we study the smaller tree $\mathcal{T}^*(a)$ resulting by deleting all nodes $n \equiv 0 \pmod 3$ from $\mathcal{T}(a)$. The inverse operator $(T^*)^{-1}$ to T on the restricted domain $\{ n : n \not\equiv 0 \pmod 3 \}$ is:

$$(T^*)^{-1}(n) = \begin{cases} \{2n\} & \text{if } \ n \equiv 1,4,5 \text{ or } 7 \pmod 9 \ , \\ \{2n, \dfrac{2n-1}{3}\} & \text{if } \ n \equiv 2 \text{ or } 8 \pmod 9 \ . \end{cases} \tag{2.2}$$

Now let $\mathcal{T}_k^*(a)$ denote the depth k subtree of $\mathcal{T}(a)$, see Figure 2.1 for $\mathcal{T}_5(4)$ and $\mathcal{T}_5^*(4)$.

Insert Figure 2.1 about here

We next assign *weights* to each edge of the tree which keep track of $3x+1$ iterates (mod 2): An edge connecting $2n$ and n is assigned weight 0, while one connecting $\dfrac{2n-1}{3}$ and n is assigned weight 1. Each node of a tree (except the root) is then assigned weight equal to the sum of the weights of the edges connecting it to the root node. Thus a leaf l of $\mathcal{T}_k^*(a)$ has

$$\text{weight}(l) := \#\{ i : T^{(i)}(l) \equiv 1 \pmod 2 \ , \ 0 \le i \le k-1 \} \ . \tag{2.3}$$

The weight approximately measures the size of the node label, namely

$$l \le 3^{-\text{weight}(l)}\, 2^k a \ . \tag{2.4}$$

In addition it can be shown that

$$l \;=\; (1 + o(1))3^{-\text{weight}(l)}2^k a \tag{2.5}$$

as $k \to \infty$, for all those l having weight$(l) \le \frac{6}{10}k$.

The branching structure of the tree $7_k^*(a)$, together with all the weights of all its nodes and edges, is completely determined by the congruence class a (mod 3^{k+1}); thus the number of distinct tree structures $7_k^*(a)$ is at most $2 \cdot 3^k$.

We will study various statistics concerning the leaves of the trees $7_k^*(a)$. Let $w_j^k(a)$ count the number of leaves of $7_k^*(a)$ having weight j, yielding the vector of weights

$$\mathbf{w}_k^*(a) := (w_0^k(a),\, w_1^k(a), \ldots, w_k^k(a)) \ . \tag{2.6}$$

Now let $N_k^*(a)$ count the number of leaves of $7_k^*(a)$, whence

$$N_k^*(a) \;=\; w_0^k(a) \;+\; w_1^k(a) \;+\ldots+\; w_k^k(a) \ . \tag{2.7}$$

It is obvious that

$$n_k^*(a) \;\le\; N_k^*(a) \ ,$$

and equality holds whenever a is not in a cycle of T. Theorem 3.1 of [7] showed that the expected size $E[N_k^*(a)]$ averaged over residue classes a(mod 3^{k+1}) with $a \not\equiv 0$ (mod 3) is

$$E[N_k^*(a)] \;=\; \left[\frac{4}{3}\right]^k \ . \tag{2.8}$$

The quantities we study are

$$N^+(k) := \max\{N_k^*(a) : a \ (\text{mod } 3^{k+1}) \ \text{with } a \not\equiv 0 \ (\text{mod } 3)\}$$
$$N^-(k) := \min\{N_k^*(a) : a \ (\text{mod } 3^{k+1}) \ \text{with } a \not\equiv 0 \ (\text{mod } 3)\}$$

and the majorizing and minorizing vectors:

$$\mathbf{w}^{+}(k) := \text{majorize}\{\mathbf{w}_k^*(a) : a (\text{mod } 3^{k+1}) \text{ with } a \not\equiv 0 \pmod 3\}$$
$$\mathbf{w}^{-}(k) := \text{minorize}\{\mathbf{w}_k^*(a) : a \pmod{3^{k+1}} \text{with } a \not\equiv 0 \pmod 3\} .$$

Here we say that a vector $\mathbf{w} = (w_0, \ldots, w_k)$ *majorizes* a vector $\mathbf{w}' = (w_0', \ldots, w_k')$ if

$$\sum_{j=0}^{i} w_{k-j} \geq \sum_{j=0}^{i} w_{k-j}' , \quad 0 \leq i \leq k ,$$

while $\mathbf{w}$ *minorizes* $\mathbf{w}'$ if

$$\sum_{j=0}^{i} w_{k-j} \leq \sum_{j=0}^{i} w_{k-j}' , \quad 0 \leq i \leq k .$$

Now

$$\mathbf{w}^{+}(k) := (w_0^{+}(k) , w_1^{+}(k), \ldots, w_k^{+}(k))$$

is the smallest vector majorizing all the $\mathbf{w}_k^*(a)$, and is determined by the conditions

$$\sum_{j=0}^{i} w_{k-j}^{+}(k) = \max\left\{\sum_{j=0}^{i} w_{k-j}^{k}(a) : a \pmod{3^{k+1}} \text{ with } a \not\equiv 0 \pmod 3\right\} , \quad 0 \leq i \leq k . \tag{2.9}$$

Similarly

$$\mathbf{w}^{-}(k) := (w_0^{-}(k) , w_1^{-}(k), \ldots, w_k^{-}(k))$$

is determined by the conditions

$$\sum_{j=0}^{i} w_{k-j}^{-}(a) = \min\left\{\sum_{j=0}^{i} w_{k-j}^{k}(a) : a (\text{mod } 3^{k+1}) \text{ with } a \not\equiv 0 \pmod 3\right\} , \quad 0 \leq i \leq k \tag{2.10}$$

It is easy to see that these definitions imply that

$$N^{+}(k) = \sum_{j=0}^{k} w_j^{+}(k) , \tag{2.11a}$$
$$N^{-}(k) = \sum_{j=0}^{k} w_j^{-}(k) . \tag{2.11b}$$

In view of (2.8), we have

$$N^{-}(k) \le \left[\frac{4}{3}\right]^{k} \le N^{+}(k) \;, \qquad k \ge 1 \;. \tag{2.12}$$

We computed the vectors $\mathbf{w}^{+}(k)$ and $\mathbf{w}^{-}(k)$ for $1 \le k \le 30$; the data for $\mathbf{w}^{-}(k)$ and $N^{-}(k)$ appear in Table 2.1, and that for $\mathbf{w}^{+}(k)$ and $N^{+}(k)$ in Table 2.2. Details on the computational method are given at the end of the section.

Insert Tables 2.1 and 2.2 about here

The associated growth rates are

$$g^{-}(k) = N^{-}(k)^{1/k} \;; \;\; g^{+}(k) = N^{+}(k)^{1/k} \;. \tag{2.13}$$

They are tabulated for $1 \le k \le 30$ in Tables 2.1 and 2.2.

Theorem 2.1. *For any* $k \ge 1$, *and any* $a \not\equiv 0$ *(mod 3),*

$$g^{-}(k) \le \liminf_{j \to \infty} N_j^{*}(a)^{1/j} \le \limsup_{j \to \infty} N_j^{*}(a)^{1/j} \le g^{+}(k) \;. \tag{2.14}$$

In addition,

$$g^{-}(k) \le \liminf_{j \to \infty} n_j^{*}(a)^{1/j} \le \limsup_{j \to \infty} n_j(a)^{1/j} \le g^{+}(k) \;. \tag{2.15}$$

Proof. Since each tree of depth jk splits into trees of depth k attached to each leaf of the tree of depth $j(k-1)$, we get by an easy induction

$$N^{-}(k)^{j} \le N_{jk}^{*}(a) \le N^{+}(k)^{j} \;.$$

For $0 \le l \le k$, we obviously have

$$N^{-}(k)^{j} \le N_{jk+l}^{*}(a) \le N^{+}(k)^{j+1} \;.$$

Taking jk-th roots and letting $j \to \infty$ yields (2.14).

To prove the upper bound in (2.15), use

$$n_j(a) \le j n_j^*(a) \le j N_j^*(a) ,$$

and (2.14). The lower bound in (2.15) is also immediate if a is not in a cycle of T, since $n_j^*(a) = N_j^*(a)$ in this case. If a is in a cycle, then the tree $7^*(a)$ contains some a' not in a cycle, say at level l. Then

$$n_j^*(a) \ge n_{j-l}^*(a') = N_{j-l}^*(a') ,$$

and the lower bound follows from the lower bound (2.14) for $N_{j-l}^*(a')$. ■

Theorem 1.1 follows immediately from this result, using the $k = 30$ entries of Tables 1.1 and 1.2.

How fast do $N^+(k)$ and $N^-(k)$ grow? In order for Conjecture B to be derivable from Theorem 2.1, it is necessary that

$$\lim_{k \to \infty} g^+(k) = \lim_{k \to \infty} g^-(k) = \frac{4}{3} .$$

We restate this as the following conjecture.

Conjecture C. *Both $N^+(k)$ and $N^-(k)$ are $\left[\frac{4}{3}\right]^{k(1+o(1))}$ as $k \to \infty$.*

This conjecture is stronger than Conjecture B, because it concerns extreme values over all trees of depth k, while Conjecture B applies to the quantities $n_k(a)$, which as $k \to \infty$ should behave like ‘‘random’’ trees. To further compare the data with this conjecture, we give in Table 2.3 the quantities $(\frac{4}{3})^k$ and the ratios $(\frac{4}{3})^k (N^-(k))^{-1}$ and $(\frac{4}{3})^{-k} N^+(k)$. Formula (2.8) implies that for all $k \ge 1$ both these ratios must be at least 1.

Insert Table 2.3 about here.

The data support Conjecture C, and even suggest the following stronger conjecture.

Conjecture C′. *There are positive constants C^+ and C^- such that*

$$C^- \left[\frac{4}{3} \right]^k \le N^-(k) < N^+(k) \le C^+ \left[\frac{4}{3} \right]^k$$

for all sufficiently large k.

Lagarias and Weiss [7] developed branching process models intended to mimic the behavior of $3x+1$ trees. It can be proved for the branching process models @$[3^j]$ for $j \ge 2$ discussed in [7] that the analogue of Conjecture C is true, but that the analogue of the stronger Conjecture C′ is false, see [2]. That is, $3x+1$ trees empirically have a narrower variation of leaf counts than that predicted by such stochastic models. This is the first significant deviation found for the $3x+1$ iteration from being as random as possible consistent with obvious constraints. It merits on explanation, and we raise this as an open question.

The computation of Tables 2.1 and 2.2 was based on a simple observation: For a given $a \pmod{3^{k+1}}$ with $a \not\equiv 0 \pmod{3}$, let $mw_k(a)$ denote the maximum weight of a leaf of the tree $7_k^*(a)$. Then all trees $7_k^*(a')$ with $a' \equiv a \pmod{3^{mw_k(a)+1}}$ have identical branching structure and node weights. Thus in doing the computation we may group all these trees together, specifying them by a single congruence class $a \pmod{3^{l+1}}$ where $l = mw_k(a)$, which we call a *clone*. Let R_l^k count the number of distinct clones of depth k having maximum weight leaf l. The values of R_l^k up to $k = 23$ are given in Table 2.4.

Insert Table 2.4 about here.

The R_l^k satisfy the identity

$$\sum_{l=0}^{k} R_l^k 3^{k-l} = 2 \cdot 3^k . \qquad (2.16)$$

The total number of clones of depth k,

$$R(k) \ := \ \sum_{l=0}^{k} R_l^k \ , \tag{2.17}$$

counts all possible tree structures of depth k that occur using the $3x+1$ function. Data on $R(k)$ and on $R(k)^{1/k}$ also appear in Table 2.4. Using

$$w_i^k(a) \ = \ \begin{cases} w_i^{k-1}(2a) & \text{if } a \equiv 1,4,5, \text{ or } 7 \pmod 9 \ , \\ w_i^{k-1}(2a) \ + \ w_{i-1}^{k-1}(\frac{2a-1}{3}) & \text{if } a \equiv 2,\ 8 \pmod 9 \ , \end{cases}$$

for $0 \le i \le k$, and

$$mw_k(a) \ = \ \begin{cases} mw_{k-1}(2a) & \text{if } a \equiv 1,4,5, \text{ or } 7 \pmod 9 \ , \\ \max\{ mw_{k-1}(2a) \ , \ mw_{k-1}(\frac{2a-1}{3}) \ + \ 1 \} & \text{if } a \equiv 2 \text{ or } 8 \pmod 9 \ , \end{cases}$$

all clones of depth k can be identified and $\mathbf{w}_k^*(\cdot)$ and $mw_k(\cdot)$ computed for them in $O(kR(k))$ operations from a hashtable containing $\mathbf{w}_{k-1}^*(\cdot)$ and $mw_k(\cdot)$ for all clones of depth $k - 1$. In the actual computation, memory was exhausted by the hashtable at $k = 21$, so $\mathbf{w}_l^*(\cdot)$ and $mw_l(\cdot)$ for clones of depth $l \ge 21$ were recomputed as needed.

The quantity $R(k)$ grows at a somewhat slower exponential growth rate than $2 \cdot 3^k$, which makes the computation feasible up to $k = 30$. By analogy with a branching process model in Lagarias and Weiss [7] one expects that there is a constant θ such that $R(k) = \theta^{k(1+o(1))}$ as $k \to \infty$, and empirically we estimate $1.87 < \theta < 1.92$. Here the lower bound 1.87 comes from $R(k)^{1/(k+1)}$, which is monotonically increasing for $8 \le k \le 28$. Observe also that $R_l^k = 0$ for small l, which occurs because branching of the tree is unavoidable. By analogy with branching process models, one expects that there exists a positive constant ϕ such that $R_l^k = 0$ for $l < (\phi + o(1))k$ and $R_l^k > 0$ for $(\phi + o(1))k \le l \le k$, as $k \to \infty$.

3. Large Deviation Estimates: Lower Bounds and Upper Bounds

We can use minorizing vectors $\mathbf{w}^-(k)$ to get lower bounds for γ in (1.3), as follows. For any constant $\alpha \in (0,1]$, set

$$N_j^*(a;\,\alpha) := \#\{l : l \text{ is a leaf in } 7_j^*(a) \text{ and weight}(l) \geq \alpha j\}\ .$$

By (2.4) all such leaves satisfy the bound

$$l \leq \exp(j(\log 2 - \alpha \log 3))\, a\ . \tag{3.1}$$

Consequently, if we set $x = \exp(j(\log 2 - \alpha \log 3))\, a$, and let $j \to \infty$, then we obtain

$$\pi_a(x) \geq x^{\gamma - \varepsilon}\ ,$$

where

$$\gamma = \frac{1}{\log 2 - \alpha \log 3} \liminf_{j \to \infty} \frac{1}{j} (\log N_j^*(a;\alpha))\ . \tag{3.2}$$

Next we use the minorizing vector $\mathbf{w}^-(k)$ to obtain an asymptotic lower bound for $N_j^*(a;\alpha)$. Form a *minorizing tree* 7_k^- consisting of $N^-(k)$ leaves of depth one, with exactly $w_i^-(k)$ of these leaves having edges assigned the weight i, for $0 \leq i \leq k$. Now, for all $j \geq 1$, recursively construct the *concatenated minorizing tree** $7_k^-(j)$ by setting $7_k^-(1) = 7_k^-$ with root node labelled 1, and then forming $7_k^-(j)$ from $7_k^-(j-1)$ by attaching copies of the tree 7_k^- to each leaf of $7_k^-(j-1)$. Each leaf of $7_k^-(j)$ is assigned a weight consisting of the sum of edge weights from it to the root node. Let

$$\mathbf{w}^-(k)^{(*j)} := (x_0^k(j), \ldots, x_{jk}^k(j)) \tag{3.3}$$

be a vector counting the number of leaves of $7_k^-(j)$ of weight i, for $0 \leq i \leq jk$. (The notation is

---

* The tree $7_k^-(j)$ has depth j, but its leaf counts will minorize those of a $3x+1$ tree of depth jk.

intended to indicate repeated convolution of $\mathbf{w}^-(k)$, as explained below.) Note also that the number of leaves of $7^-_k(j)$ is $N^-(k)^j$. We claim that:

$$\mathbf{w}^-(k)^{(*j)} \quad \textit{minorizes} \quad \mathbf{w}^-(jk) \ . \tag{3.4}$$

To prove the claim, it suffices to show that $\mathbf{w}^-(k)^{(*j)}$ minorizes each $\mathbf{w}^*_{jk}(a)$. We proceed by induction on j, it being obviously true for $j = 1$. Take any tree $7_{jk}(a)$ and view it as a tree $7_{(j-1)k}(a)$ with various trees $7_k(b)$ attached to its leaves. By the induction hypothesis (3.4), the tree $7^-_k(j-1)$ can have its leaves paired with those of $7_{(j-1)k}(a)$ in such a way that each leaf of $7^-_k(j-1)$ has a weight smaller than the corresponding leaf of $7_{(j-1)k}(a)$, and $7_{(j-1)k}(a)$ has some unpaired leaves left over. Then replace $7_{(j-1)k}(a)$ with $7^-_k(j-1)$ and throw away all trees $7_k(b)$ attached to the unpaired nodes, and the weight vector of the resulting new tree minorizes that of the old tree $7_{jk}(a)$. Next, in the resulting tree, replace each tree $7_k(b)$ with the tree 7^-_k, and the weight vector of the resulting tree minorizes the one before. This final tree is $7^-_k(j)$, hence we have shown that $\mathbf{w}^-(k)^{(*j)}$ minorizes $\mathbf{w}^*_{jk}(a)$, and the induction step follows.

Now (3.4) yields the lower bound

$$N^*_{jk}(a;\alpha) \ \geq \ P^-_{j,k}(\alpha) \ := \sum_{i > jk\alpha} x^k_i(j) \ . \tag{3.5}$$

The right side of (3.5) depends only on $\mathbf{w}^-(k)$, and can be estimated in a standard fashion, see Lemma 3.1 below. We can then interpolate estimates for $N^*_{jk+l}(a;\alpha)$ using

$$N^*_{jk+l}(a;\alpha) \ \geq \ N^*_{(j+1)k}(a;\alpha + \frac{1}{jk}) \,, \quad 0 \leq l \leq k \ .$$

It is convenient to interpret this estimation as a ‘‘large deviations’’ bound in probability theory. To do this, we assign node labels to the tree 7^-_k, by giving each leaf of weight i the label

$$l \ = \ 2^k 3^{-i} \ .$$

(This label actually represents the ratio of a leaf label to the root label.) We can use this scheme to recursively label all the nodes of the trees $7_k^{-}(j)$, starting by assigning the root node the label 1. Next, let Z_k^{-} be a random variable which draws a leaf l of $7_k^{-}(1)$ uniformly and then takes the value

$$\begin{aligned} Z_k^{-} &:= \log l \ . \\ &= k \log 2 - i \log 3 \end{aligned} \tag{3.6}$$

The convolved random variable $(Z_k^{-})^{(*j)}$ then describes the value $\log l$ of a leaf of $7_k^{-}(j)$ drawn uniformly. Now, the right side of (3.5) counts exactly those leaves of $7_k^{-}(j)$ with $l = 2^{jk} 3^{-i} \le 2^{jk} 3^{-jk\alpha}$, hence

$$P_{j,k}^{-}(\alpha) = (N^{-}(k))^{j} \ \mathrm{Prob}[(Z_k^{-})^{(*j)} < jk(\log 2 - \alpha \log 3)] \ , \tag{3.7}$$

The estimation of (3.7) is a standard ‘‘large deviations’’ result.

Lemma 3.1. *The random variable* $Z = Z_k^{-}$ *has moment generating function*

$$M_k^{-}(\theta) = E[e^{\theta Z}] = \sum_{i=0}^{k} \frac{w_i^{-}(k)}{N^{-}(k)} 2^{k\theta} 3^{-i\theta} \ ,$$

whose Legendre transform is

$$g_k^{-}(\beta) := \sup_{\theta \in \mathrm{R}} [\beta\theta - \log M_k^{-}(\theta)] \ .$$

If $0 < \log 2 - \alpha \log 3 < \dfrac{1}{k} E[Z_k^{-}]$ *, then*

$$\lim_{j \to \infty} \frac{1}{jk} (\log P_{j,k}^{-}(\alpha)) = \frac{1}{k} (\log N^{-}(k) - g_k^{-}(k(\log 2 - \alpha \log 3))) \ . \tag{3.8}$$

Proof. This is just an application of Chernoff’s theorem, see [7], Lemma 2.1. ■

Combining (3.2), (3.5), (3.7) and (3.8) yields the bound

$$\gamma \geq \frac{\frac{1}{k}(\log N^-(k) - g_k^-(k(\log 2 - \alpha \log 3)))}{\log 2 - \alpha \log 3} , \tag{3.9}$$

provided

$$0 < \log 2 - \alpha \log 3 < \frac{1}{k} E[Z_k^-] = \frac{1}{k} \sum_{i=0}^{k} \frac{i\, w_i^-(k)}{N^-(k)} .$$

For each value of k it remains to optimize the bound (3.9) by choosing the optimal $\alpha = \alpha_k^*$.

Data on the expected value $\frac{1}{k} E[Z_k^-]$, the optimal cutoff value α_k^*, and the resulting lower bound γ_k^*, are given in Table 3.1 below. The quantity $\frac{1}{k} E[Z_k^-]$ is always greater than the expected growth rate of labels on a random branch of a ‘‘random’’ tree $7_k(a)$, which is $\log 2 - \frac{1}{4} \log 3 \doteq .418494$, cf. [7], Theorem 3.3. Note that $\frac{1}{k} E[Z_k^-]$ is not a monotonically decreasing function of k, though it tends to decrease as k increases. Consequently the estimates γ_k^* are also not monotonically increasing, but tend to increase. The largest value we found was $\gamma_{30}^* = .654717$; this proves Theorem 1.2. It is natural to conjecture that $\frac{1}{k} E[Z_k^-] \to \log 2 - \frac{1}{4} \log 3$ and that $\gamma_k^* \to 1$ as $k \to \infty$.

Insert Table 3.1 about here

We can similarly use majorizing vectors $\mathbf{w}^+(k)$ to get upper bounds on $N_j^*(a;\alpha)$. We construct trees 7_k^+ and $7_k^+(j)$ analogously to the lower bound case, using $\mathbf{w}^+(k)$ instead of $\mathbf{w}^-(k)$. Let

$$\mathbf{w}^+(k)^{(*j)} := (y_0^k(j), \ldots, y_{jk}^k(j))$$

enumerate the number of leaves in the tree $7_k^+(j)$ of different weights. We then show, analogously to the lower bound case, that

$$\mathbf{w}^+(k)^{(*j)} \quad \textit{majorizes} \quad \mathbf{w}^+(jk)\ , \tag{3.10}$$

from which we conclude

$$N^*_{jk}(a;\alpha) \le P^+_{j,k}(\alpha) := \sum_{i > jk\alpha} y_i^k(j)\ . \tag{3.11}$$

The right side of (3.11) is estimated by a Chernoff inequality argument. Let Z_k^+ be a random variable which draws a leaf l from $\mathcal{T}_k^+(1)$ uniformly and assigns it the value $\log(l)$, similarly to (3.6). The convolution $(Z_k^+)^{(*j)}$ then describes the value $\log(l)$ for a random leaf of $\mathcal{T}_k^+(j)$ and we have

$$P^+_{j,k}(\alpha) = (N^+(k))^j \ \mathrm{Prob}[(Z_k^+)^{(*j)} < jk(\log 2 - \alpha \log 3)]\ .$$

The Chernoff bound formula is analogous to Lemma 3.1.

Lemma 3.2. *The random variable* $Z = Z_k^+$ *has moment generating function*

$$M_k^+(\theta) = \sum_{i=0}^{k} \frac{w_i^+(k)}{N^+(k)}\ 2^{k\theta} 3^{-i\theta}\ ,$$

whose Legendre transform is

$$g_k^+(\beta) := \sup_{\theta \in \mathbb{R}} [\beta\theta - \log M_k^+(\theta)].$$

If $\log 2 - \alpha \log 3 > \frac{1}{k} E[Z_k^+]$, *then*

$$\lim_{j \to \infty} \frac{1}{jk}(\log P^+_{j,k}(\alpha)) = \frac{1}{k}(\log N^+(k) - g_k^+(k(\log 2 - \alpha \log 3)))\ . \tag{3.12}$$

Table 3.1 presents data on $\frac{1}{k} E[Z_k^+]$. It is always less than the expected growth rate $\log 2 - \frac{1}{4}\log 3 \doteq .418494$ of labels on a random branch of a ‘‘random’’ tree $\mathcal{T}_k(a)$. Empirically, it appears to be a monotone function of k, unlike the lower bound case. It is natural to conjecture that $\frac{1}{k} E[Z_k^+] \to \log 2 - \frac{1}{4} \log 3$ as $k \to \infty$.

Upper bound estimates for $N_j^*(a;\alpha)$ are also relevant to proving results saying that ‘‘almost all’’ integers decrease under iteration by T. Currently the best quantitative result of this kind is that of Korec [4].

Theorem 3.1. (Korec) *For any* $\beta > \beta_c := \frac{\log 3}{\log 4} \doteq .7925$ *the set*

$$S(\beta) := \{ n : \textit{some } |T^{(k)}(n)| < |n|^{\beta} \}$$

has density one.

Korec’s method actually shows that almost all $\{ n : |n| \le x \}$ satisfy

$$|T^{(k)}(n)| \le x^{\beta} \ , \quad \text{for} \quad k = \left\lceil \frac{\log x}{\log 2} \right\rceil , \tag{3.13}$$

as $x \to \infty$, for any fixed $\beta > \beta_c$.

We show below that one can get improved bounds for β_c in Theorem 3.1 provided that the quantity

$$\chi_k^* := \frac{1}{k} (\log N^+(k) - g_k^+(k(\log 2 - 1/2(\log 3)))) \tag{3.14}$$

is sufficiently small. This quantity is the upper bound (3.12) with $\alpha = 1/2$, and its values are given in Table 3.1.

Consider the set of ‘‘bad elements’’

$$R_{\delta}(x) : = \{ n : |n| < x \ \text{ and no } \ T^{(j)}(x) < x^{1-\delta} \ \text{ for } \ 1 \le j < \left\lceil \frac{\log x}{\log 2} \right\rceil \} \ .$$

The cardinality of $R_{\delta}(x)$ decreases as $\delta \to 0$ and

$$\lim_{\delta \to 0} \frac{\log \#(R_{\delta}(x))}{\log x} = H \left[\frac{\log 2}{\log 3} \right] \doteq .94995 \ , \tag{3.15}$$

where $H(t) = -t \log_2 t - (1-t) \log_2 (1-t)$ is the binary entropy function, cf. [6], Theorem D.

Almost all $\{n : |n| \le x\}$ satisfy (3.13), and we can get an improvement if furthermore almost all such $T^{(k)}(n)$ with $k = \left\lceil \frac{\log x}{\log 2} \right\rceil$ not lie in a ''bad element'' set $R_\delta(x^\beta)$, for some fixed $\delta > 0$. How many such n can hit a particular ''bad'' element y? They must lie in the tree of preimages of y, at height $j = \frac{\log x}{\log 2}$, so we need an upper bound for the number of leaves l in such a tree, at this height, having $y \approx x^\beta$ and $l \le x$. Such leaves correspond to paths having $\alpha \ge \frac{1}{2}$ (as explained in [7], §2), hence we can apply* the upper bounds (3.11)–(3.13) to bound the number of such leaves by $\exp\left[\chi_k \frac{\log x}{\log 2}\right]$. Now the number of such ''bad elements'' as $\beta \to \beta_c$ and $\delta \to 0$ satisfies

$$\log \#(R_\delta(x^\beta)) = \left[.94995 \, \frac{\log 3}{\log 4} + o(1) \right] \log x ,$$

hence the number of preimages $n \le x$ which these generate is at most

$$\exp\left[\left[.94995 \, \frac{\log 3}{\log 4} + \frac{\chi_k}{\log 2} \right] \log x \right] .$$

This bound will be $O(x^{1-\varepsilon'})$ for some $\varepsilon' > 0$, if and only if

$$\chi_k < \log 2 - \frac{1}{2} H\left[\frac{\log 2}{\log 3} \right] \log 3 \doteq .171331 . \tag{3.16}$$

As the data of Table 3.1 show, however, for $k \le 30$ we never attain the bound (3.16).

---

* To get a rigorous bound, one must also count a few extra leaves having $\alpha < \frac{1}{2}$, which creep in because T^{-1} has $\frac{2x-1}{3}$ instead of $\frac{2x}{3}$. However a rigorous variant of (2.5) can be used to show that these leaves make an asymptotically negligible contribution.

The assumption that $3x+1$ trees behave like the branching process models of [7] leads to the heuristic prediction that $\chi_k \to 0$ as $k \to \infty$. If so, this approach to lowering β_c should eventually work for large enough k. The data of Table 3.1 strongly indicate that the smallest k for which (3.16) holds will however be so large that it will be impossible to compute by an exhaustive tree search.

References

[1] D. Applegate and J. C. Lagarias, Density Bounds for the $3x+1$ Problem II. Krasikov Inequalities, Math. Comp., submitted.

[2] D. Applegate and J. C. Lagarias, On Branching Process Models for the $3x+1$ Problem, preprint.

[3] R. E. Crandall, On the ‘‘3x+1’’ Problem, Math. Comp. **32** (1978), 1281-1292.

[4] I. Korec, An Estimation in the 3x+1 Problem, Math. Slovaca, to appear. (preprint 1992).

[5] I. Krasikov, How many numbers satisfy the 3x+1 conjecture?, Inter. J. Math. & Math. Sci. **12** (1989), 791-796.

[6] J. C. Lagarias, The 3x+1 Problem and Its Generalizations, Amer. Math. Monthly. **92** (1985), 3-21.

[7] J. C. Lagarias and A. Weiss, The 3x+1 Problem: Two Stochastic Models, Ann. Applied Prob. **2** (1992), 229-261.

[8] G. T. Leavens and M. Vermeulen, $3x+1$ Search Programs, Computers & Mathematics, with Applications **24,** No. 11, (1992), 79-99.

[9] J. W. Sander, On the $(3N+1)$-conjecture, Acta Arith. **55** (1990), 241-248.

[10] G. Wirsching, An Improved Estimate Concerning 3N+1 Predecessor Sets, Acta Arith., **63** (1993), 205-210.

$7_5(4)$ $7^*_5(4)$

Figure 2.1. $3x + 1$ *tree* $7_5(4)$ *and pruned tree* $7^*_5(4)$.

Density Bounds for the 3x+1 Problem
I. Tree-Search Method

David Applegate
Jeffrey C. Lagarias

AT&T Bell Laboratories
Murray Hill, NJ 07974

(December 15, 1993)

(Dedicated to the memory of D. H. Lehmer)

ABSTRACT

The $3x+1$ function $T(x)$ takes the values $(3x+1)/2$ if x is odd and $x/2$ if x is even. Let $n_k(a)$ count the number of n with $T^{(k)}(n) = a$. Then for any $a \not\equiv 0 \pmod 3$ and sufficiently large k, $(1.302)^k \le n_k(a) \le (1.359)^k$. Let $\pi_a(x)$ count the number of n with $|n| \le x$ which eventually reach a under iteration by T. Then for any $a \not\equiv 0 \pmod 3$ and sufficiently large x, $\pi_a(x) \ge x^{.65}$. The proofs are computer-intensive.

4. Branching Process Model for 3x+1 Trees

Lagarias and Weiss (1992) developed branching process models intended to mimic the behavior of $3x+1$ trees. Detailed rigorous results can be obtained for such models, in contrast to the $3x+1$ problem itself. We ask: How do the data in Table 2.1 compare with predictions for such a model?

We consider the multi-type Galton-Watson branching process @[9] described in [8], §3, Table 2. It has individuals of six types, labelled with congruence classes 1,2,4,5,7 and 8 (mod 9), and these evolve as pictured in Table 3.1. Individuals labelled 1,4,5 and 7 evolve deterministically, having one child of specified type, while individuals of type 2 or 8 always have two children, one of specified type, while the other's type is specified with probability 1/3 each.

Insert Table 3.1 about here.

Let X_k denote the distribution of the number of leaves at depth k of a sample tree drawn from this branching process, starting from a single individual of type drawn uniformly from $\{1,2,4,5,7,8\}$. The data in Table 2.1 is analogous to extreme value statistics for of the quantity $\left[\frac{4}{3}\right]^{-k} X_k$ for repeated independent draws of such trees at depth k.

How many independent draws should we allow in such a branching process model? The naive model is to take $2 \cdot 3^k$ draws, corresponding to all $a \pmod{3^{k+1}}$. An alternative is to take $R(k)$ draws, where $R(k)$ is number of different possible $3x$ tree structure $7_{\,a}$ of depth k possible. The quantities $R(k)$ grow exponentially in k, and based on the data for $k \le 30$ in Applegate and Lagarias (1993), we conjecture that

$$1.87 < \lim_{k \to \infty} inf R(k)^{1/k} < 1.92 \ .$$

We therefore consider for any fixed $\theta > 1$ the model quantities

$$\tilde{N}_\theta^-(k) \ := \ E[\min\{X_k : [\theta^k] \text{ i.i.d. draws}\}] \tag{3.1a}$$

$$\tilde{N}_\theta^-(k) \ := \ E[\max\{X_k : [\theta^k] \text{ i.i.d. draws}\}] \ . \tag{3.1b}$$

Then the quantities $\left[\frac{4}{3}\right]^k (\tilde{N}_\theta^-(k))^{-1}$ and $\left[\frac{4}{3}\right]^{-k} \tilde{N}_\theta^+(k)$ are analogous to the quantities given in Table 2.1.

For this branching process models the analogue of Conjecture C′ is false.

Theorem 3.1. *For any fixed* $\theta > 1$,

$$\lim_{k \to \infty} \left[\frac{4}{3}\right]^k (\tilde{N}_\theta^-(k))^{-1} \ = \ +\infty \ ,$$

$$\lim_{k \to \infty} \left[\frac{4}{3}\right]^{-k} (\tilde{N}_\theta^+(k))^{-1} \ = \ +\infty \ .$$

Proof. Let W_k^m for $m \pmod 9$ enumerate the number of leaves of type m of a random tree of depth k drawn from @$[9]$, with root node drawn uniformly from $\{1,2,4,5,7,8\}$. Set

$$\mathbf{W}_k := (W_k^1, W_k^2, W_k^4, W_k^5, W_k^7, W_k^8) \ , \tag{3.2}$$

so that $X_k = W_k^1 + W_k^2 + W_k^4 + W_k^5 + W_k^7 + W_k^8$. Now let $\mathbf{w}_k$ denote the probability distribution of the random vector $\left[\frac{4}{3}\right]^{-k} \mathbf{W}_k$. One has $E[X_1 \log X_1] < \infty$, hence a well-known result for a multitype Galton-Watson process (Theorem 1 of Sect. V.6 of Athreiya and Ney (1972)) implies that the distributions $\mathbf{w}_k$ converge weakly to a limiting distribution $\mathbf{w}_\infty$, where

$$\mathbf{w}_\infty = w \cdot \mathbf{v} \tag{3.3}$$

and $\mathbf{v} = (1,1,1,1,1,1)$ is a left-eigenvector of the mean value matrix $\mathbf{M}$ in Table 4.2, and w is a one-dimensional positive random variable which is absolutely continuous, except for a possible jump at the origin. The distribution w depends on the starting individual's type and

$$E[w : \text{initial type } i] = u_i\ , \tag{3.4}$$

where $\mathbf{u}$ is a right eigenvector of $\mathbf{M}$, and the jump q_i at the origin depends on i. For this special case there are no jumps (all $q_i = 0$), and *each distribution* $w_i = \{w |$ *initial type* $i\}$ *is strictly positive on* R^+, by Theorem 2(iv) of Chapter V.6 of Athreiya and Ney (1972).* Now the random variables $\left[\frac{4}{3}\right]^{-k} \tilde{N}_\theta^-(k)$ and $\left[\frac{4}{3}\right]^{k} \tilde{N}_\theta^+(k)$ essentially sample values in the tails of the distributions $\mathbf{w}_k$, i.e. values that lie outside any fixed region $(\varepsilon, 1-\varepsilon)$ in the cumulative distribution for large enough k. Since $\mathbf{w}_k$ converge weakly to $\mathbf{w}_\infty$ it follows from the strict positivity of w on R^+ that

$$\left[\frac{4}{3}\right]^{-k} \tilde{N}_\theta^-(k) \to 0 \qquad \text{as } k \to \infty\ ,$$

$$\left[\frac{4}{3}\right]^{-k} \tilde{N}_\theta^+(k) \to \infty \qquad \text{as } k \to \infty\ ,$$

so Theorem 3.1 follows. ■

To what extent does the asymptotic behavior given by Theorem 3.1 show up for $k \leq 30$? To

---

* A detailed proof of the positivity of w for the single-type Galton-Watson process appears as Theorem 2 of Sect. II.5 of Athreiya and Ney (1972).

obtain as exact a numerical comparison with Table 2.1 as possible, we computed the quantities

$$\tilde{N}^{-}(k) := E[\min\{X_k : R(k) \text{ i.i.d. draws}\}]$$
$$\tilde{N}^{+}(k) := E[\max\{X_k : R(k) \text{ i.i.d. draws}\}]$$

using the exact values of $R(k)$ computed in Applegate and Lagarias (1993). Table 3.2 gives the results.

Insert Table 3.2 about here.

In this table the qualitative increase of these quantities with k is evident. If we used the model which takes $2 \cdot 3^k$ draws instead of $R(k)$ draws, the disagreement with Table 2.1 would be even greater. Note the non-monotonicity in k of $\left[\frac{4}{3}\right]^k (\tilde{N}^{-}(k))^{-1}$ and $\left[\frac{4}{3}\right]^{-k} \tilde{N}^{+}(k)$ for small values of k; this is apparently due to initial irregularities in the distribution $\mathbf{w}_k$ for small k.

Although there are a double-exponential number of different trees possible at depth k of such a branching process, the data $\tilde{N}^{-}(k)$ and $\tilde{N}^{+}(k)$ in Table 3.2 were computed in single-exponential time as follows: Let X_k^i for $i \pmod 9$ be a random variable counting the number of leaves at depth k of a sample tree drawn from the branching process @[9], starting from a single individual of type i. Then, the distributions of X_k^i and X_k were computed from

$$P[X_0^i = 1] = 1 ,$$
$$P[X_k^i = x] = P[X_{k-1}^{2i} = x] \quad \text{if} \quad i = 1,4,5, \quad \text{or} \quad 8 ,$$
$$P[X_k^2 = x] = \sum_{y=0}^{\infty} P[X_{k-1}^{2i} = x - y] + \frac{P[X_{k-1}^1 = y] + P[X_{k-1}^4 = y] + P[X_{k-1}^7 = y]}{3} ,$$
$$P[X_k^8 = x] = \sum_{y=0}^{\infty} P[X_{k-1}^{2i} = x - y] + \frac{P[X_{k-1}^2 = y] + P[X_{k-1}^5 = y] + P[X_{k-1}^8 = y]}{3} ,$$

and

$$P[X_k = x] = \frac{1}{6} \sum_{i \pmod 9} P[X_k^i = x] .$$

The cumulative distribution function $f_k(t)$ of the number of leaves was then computed. Finally the cumulative distributions of the minimum and maximum of $R(k)$ draws were computed using $(1-(1-f_k(t)))^{R(k)}$ and $f_k(t)^{R(k)}$, respectively.

The analogue of Conjecture C is certainly true for this model. We shall not rigorously prove it here, but present a heuristic argument for its truth. The important feature of the branching process @$[9]$ is that any branch must split after at most 4 steps, hence *all* subtrees of a tree grown by this process must grow exponentially. (This property fails for the simpler models @$[1]$ and @$[3]$ considered in [8], which is one reason we used @$[9]$ here. In addition [8], p. 259 gives another reason to use @$[9]$.) As a consequence the tails of the total leaf distribution drop off double-exponentially away from their mean. Thus minimizing over an exponential number of i.i.d. draws of random trees is insufficient to change the growth exponent $\frac{4}{3}$.

To summarize: empirically the $3x+1$ trees show a more compact distribution of total leaf counts than that predicted by this branching process model.

5. Branching Process Model for 3x+1 Trees

In this section we examine the growth of the quantities $N^+(k)$ and $N^-(k)$ in more detail and compare them with predictions made using a branching process model introduced in [7].

Recall that $N^+(k)$ and $N^-(k)$ are extreme values of the quantities $N_k^*(a)$, which have expected value $\left[\frac{4}{3}\right]^k$ by (2.8). It is natural to compare these quantities multiplicatively with $\left[\frac{4}{3}\right]^k$, and Table 4.1 below gives data for the quantities $\left[\frac{4}{3}\right]^k (N^-(k))^{-1}$ and $\left[\frac{4}{3}\right]^{-k} N^+(k)$, which measure multiplicative deviations from $\left[\frac{4}{3}\right]^k$.

Insert Table 4.1 about here.

This data supports Conjecture C, and even seem to support the following stronger conjecture.

Conjecture C′. *There are positive constants C^+ and C^- such that*

$$C^- \left[\frac{4}{3}\right]^k \le N^-(k) < N^+(k) \le C^+ \left[\frac{4}{3}\right]^k$$

for all sufficiently large k.

This conjecture asserts that the multiplicative deviations are incredibly small, as explained below. It is even conceivable that $N^-(k)\left[\frac{4}{3}\right]^{-k}$ and $N^+(k)\left[\frac{4}{3}\right]^{-k}$ have limiting values as $k \to \infty$.

Lagarias and Weiss (1992) developed branching process models intended to mimic the behavior of $3x+1$ trees. Detailed rigorous results can be obtained for such models, in contrast to the $3x+1$ problem itself. We ask: How do the data in Table 4.1 compare with predictions for such a model?

We consider the multi-type Galton-Watson branching process $@[9]$ described in [7], §3, Table 2. It has individuals of six types, labelled with congruence classes 1,2,4,5,7 and 8 (mod 9), and these evolve as pictured in Table 4.2. Individuals labelled 1,4,5 and 7 evolve deterministically, having one child of specified type, while individuals of type 2 or 8 always have two children, one of specified type, while the other's type is specified with probability 1/3 each.

Insert Table 4.2 about here.

Let X_k denote the distribution of the number of leaves at depth k of a sample tree drawn from this branching process, starting from a single individual of type drawn uniformly from $\{1,2,4,5,7,8\}$. The data in Table 4.1 is analogous to estimating extreme values of the quantity

$(\frac{4}{3})^{-k} X_k$ for repeated independent draws of such trees at depth k. How many independent draws should we allow in the branching process model? The naive model is to take $2 \cdot 3^k$ draws, corresponding to all $a \pmod{3^{k+1}}$. An alternative is to take $R(k)$ draws, corresponding to the number of different tree structures of depth k allowed for the $3x+1$ function. We therefore take as the model quantities:

$$\tilde{N}^{-}(k) := E[\min \left\{ X_k : R(k) \text{ i.i.d. draws} \right\}] \tag{4.1a}$$

$$\tilde{N}^{+}(k) := E[\max \left\{ X_k : R(k) \text{ i.i.d. draws} \right\}] \ . \tag{4.1b}$$

Here $(\frac{4}{3})^{k} (\tilde{N}^{-}(k))^{-1}$ and $(\frac{4}{3})^{-k} \tilde{N}^{+}(k)$ are analogous to the quantities in Table 4.1.

For this branching process model the analogue of Conjecture C is almost certainly true, while the analogue of Conjecture C′ is false.

We first discuss the analogue of Conjecture C′. Let W_k^m for $m \pmod 9$ enumerate the number of leaves of type m of a random tree of depth k drawn from $@[9]$, with root node drawn uniformly from $\{1,2,4,5,7,8\}$. Set

$$\mathbf{W}_k := (W_k^1, W_k^2, W_k^4, W_k^5, W_k^7, W_k^8) \ , \tag{4.2}$$

so that $X_k = W_k^1 + W_k^2 + W_k^4 + W_k^5 + W_k^7 + W_k^8$. Now let $\mathbf{w}_k$ denote the probability distribution of the random vector $\left[\frac{4}{3} \right]^{-k} \mathbf{W}_k$. One has $E[X_1 \log X_1] < \infty$, hence a well-known result for a multitype Galton-Watson process (Theorem 1 of Sect. V.6 of Athreiya and Ney (1972)) implies that the distributions $\mathbf{w}_k$ converge weakly to a limiting distribution $\mathbf{w}_\infty$, where

$$\mathbf{w}_\infty = w \cdot \mathbf{v} \tag{4.3}$$

and $\mathbf{v} = (1,1,1,1,1,1)$ is a left-eigenvector of the mean value matrix $\mathbf{M}$ in Table 4.2, and w is a one-dimensional positive random variable which is absolutely continuous, except for a possible jump at the origin. The distribution w depends on the starting individual's type and

$$E[w : \text{initial type } i] = u_i \,, \tag{4.4}$$

where $\mathbf{u}$ is a right eigenvector of $\mathbf{M}$, and the jump q_i at the origin depends on i. For this special case there are no jumps (all $q_i = 0$), and *each distribution* $w_i = \{w \mid \text{initial type } i\}$ *is strictly positive on* R^+, by Theorem 2(iv) of Chapter V.6 of Athreiya and Ney (1972).* Now the random variables $\tilde{N}^-(k)$ and $\tilde{N}^+(k)$ essentially sample values in the tails of the distributions w_k, i.e. values that lie outside any fixed region $(\varepsilon, 1-\varepsilon)$ in the cumulative distribution for large enough k. Since $\mathbf{w}_k$ converge weakly to $\mathbf{w}_\infty$ it follows from the strict positivity of w on R^+ that

$$\begin{aligned} \tilde{N}^-(k) &\to 0 \qquad \text{as } k \to \infty \,, \\ \tilde{N}^+(k) &\to \infty \qquad \text{as } k \to \infty \,, \end{aligned}$$

so the analogue of Conjecture C′ is false.

We obtain a numerical comparison with Table 4.1 by computing $(\frac{4}{3})^k (\tilde{N}^-(k))^{-1}$ and $(\frac{4}{3})^{-k} \tilde{N}^+(k)$ for small k, which is given in Table 4.3.

Insert Table 4.3 about here.

In this table the qualitative increase of these quantities with k is evident. If we used the model which takes $2 \cdot 3^k$ draws instead of $R(k)$ draws, the disagreement with Table 4.1 would be even greater. Also note the non-monotonicity in k of $(\frac{4}{3})^k (\tilde{N}^-(k))^{-1}$ and $(\frac{4}{3})^{-k} \tilde{N}^+(k)$ for small values of k; this is apparently due to initial irregularities in the distribution $\mathbf{w}_k$ for small k. Even

* A detailed proof of the positivity of w for the single-type Galton-Watson process appears as Theorem 2 of Sect. II.5 of Athreiya and Ney (1972).

though there are a double-exponential number of different trees possible at depth k, the values of $\tilde{N}^-(k)$ and $\tilde{N}^+(k)$ for Table 4.3 were computed in single exponential time as follows: Let X_k^i for i (mod 9) be a random variable counting the number of leaves at depth k of a sample tree drawn from the branching process @[9], starting from a single individual of type i. Then, the distributions of X_k^i and X_k were computed from

$$P[X_0^i = 1] = 1 \ ,$$

$$P[X_k^i = x] = P[X_{k-1}^{2i} = x] \quad \text{if} \quad i = 1, 4, 5, \quad \text{or} \quad 8 \ ,$$

$$P[X_k^2 = x] = \sum_{y=0}^{\infty} P[X_{k-1}^{2i} = x - y] + \frac{P[X_{k-1}^1 = y] + P[X_{k-1}^4 = y] + P[X_{k-1}^7 = y]}{3} \ ,$$

$$P[X_k^8 = x] = \sum_{y=0}^{\infty} P[X_{k-1}^{2i} = x - y] + \frac{P[X_{k-1}^2 = y] + P[X_{k-1}^5 = y] + P[X_{k-1}^8 = y]}{3} \ ,$$

and

$$P[X_k = x] = \frac{1}{6} \sum_{i (\mathrm{mod}\ 9)} P[X_k^i = x] \ .$$

The cumulative distribution function $f_k(t)$ of the number of leaves was then computed. Finally the cumulative distributions of the minimum and maximum of $R(k)$ draws were computed using $(1 - (1 - f_k(t)))^{R(k)}$ and $f_k(t)^{R(k)}$, respectively.

The analogue of Conjecture C appears to be true for this model. We shall not rigorously prove it here, but present a heuristic argument for its truth. The important feature of the branching process @[9] is that any branch must split after at most 4 steps, hence *all* subtrees of a tree grown by this process must grow exponentially. (This property fails for the simpler models @[1] and @[3] in [6], which is why we use @[9] here.) As a consequence the tails of the total leaf distribution drop off double-exponentially away from their mean. Thus minimizing over an exponential number of i.i.d. draws of random trees is insufficient to change the growth exponent $\frac{4}{3}$.

In conclusion: empirically the $3x+1$ trees show an unusually sharp distribution of total leaf counts compared to that predicted by this branching process model.