Binary search trees based on Weyl and Lehmer sequences

Luc Devroye
School of Computer Science
McGill University
Montreal, Canada H3A 2K6
luc@cs.mcgill.ca

1 Introduction

This paper is based upon the presentation at the meeting in Salzburg. As a courtesy to those who attended the meeting, I will try to faithfully reproduce—with minor omissions and additions—what I said at that meeting. There are two basic background references for mathematical details, Devroye (1987) and Devroye and Goudjil (1996).

The purpose of the study is to amass even more evidence than already exists regarding the inherent dangers of using certain random number generators in simulations. For example, linear congruential generators (based upon the recurrences $x_n = (ax_n + b) \bmod m$ for a, b, m, x_n all integer) derive their popularity from easily understood properties of Weyl and Lehmer sequences. As such generators produce integer sequences that are necessarily periodic, they can strictly speaking not be compared with aperiodic sequences such as $\{n\theta\}$ for θ irrational. However, they are fundamental akin to Weyl sequences, which may be obtained by the recurrence

$$x_{n+1} = \{x_n + \theta\} \ , \ x_0 = 0 \ ,$$

and Lehmer sequences, which may be obtained by the recurrence

$$x_{n+1} = \{ax_n\} \ , x_0 \in (0,1) \ ,$$

We study binary search trees formed by consecutive (standard) insertions of numbers $\{\theta\}, \{2\theta\}, \{3\theta\}, \ldots$, where $\theta \in (0,1)$ is an irrational number, and $\{.\}$ denotes "mod 1". The sequence in question is called the <u>Weyl sequence</u> for θ, after Weyl (but see also Bohl, 1909), who showed that for all irrational θ the sequence is equi-distributed. We recall that a

H. Niederreiter et al. (eds.), *Monte Carlo and Quasi-Monte Carlo Methods 1996*

sequence x_n, $n \geq 1$, is equi-distributed if for all $0 \leq a \leq b \leq 1$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_{x_i \in [a,b]} = b - a$$

(see Freiberger and Grenander (1971), Hlawka (1984) or Kuipers and Niederreiter (1974)).

The equi-distribution property makes Weyl sequences, or suitable generalizations of them, prime candidates for pseudo-random number generation. Of course, various regularities in the sequence make them rather unsuitable for most purposes (see, e.g., Knuth, 1981). Let $\mathcal{T}_n(\theta)$ be the binary search tree based upon the first n numbers in the Weyl sequence for θ. This tree, called the Weyl tree, captures a lot of refined information regarding the structure of the Weyl sequence, and is a fundamental tool for the analysis of algorithms involving Weyl sequences in the input stream. Computer scientists are mostly concerned with the following structural qualities: the average depth of a node (the depth is the path distance from a node to the root), the height (the maximal depth), and the number of leaves (the number of nodes with no children). The discussion in this paper focuses on these quantities. We relate various properties of the structure of these trees to the continued fraction expansion of θ. If H_n is the height of the tree with n nodes when θ is chosen at random and uniformly on $[0, 1]$, then we show that in probability, $H_n \sim (12/\pi^2) \log n \log\log n$. We conclude with some properties of binary search trees for Lehmer sequences $\{x_0 a^n\}$ where x_0 is a real number and $a > 1$ is an integer.

2 Random binary search trees

Binary search trees are binary trees in which each datum is associated with a node, and each node has the search tree property, that is, all nodes in its left subtree have smaller values, and all nodes in its right subtree have larger values. Nodes have two possible children. There are actual children (which are nodes) and potential children (places for future placement of nodes). Potential children are called external nodes. A binary tree with n nodes has $n + 1$ external nodes. Nodes without children are called leaves. Standard insertion of datum x proceeds by finding the unique external node that could accept x, given the binary search tree property, and placing x there. If a tree is constructed in this manner from an i.i.d. sequence $X_1, \ldots, X_n$ (drawn from a uniform distribution on $[0, 1]$), or from a random permutation of $\{1, \ldots, n\}$, it is called a random binary search tree, and will be denoted by $\mathcal{R}_n$. These trees may easily be drawn by placing X_i at $(X_i, -i)$, as is done below. In this manner, one easily sees the binary search tree property in action.

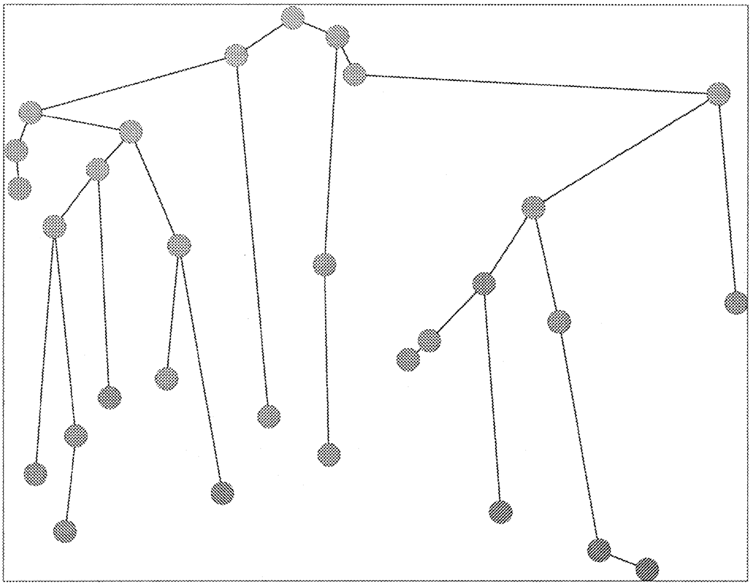

FIGURE 1. A random binary search tree in which X_i is placed at $(X_i, -i)$ in the plane. As the projections of the points show the data in order, the binary search tree is a tool for sorting and for unraveling permutations.

Mostly everything is known about the behavior of $\mathcal{R}_n$ (see Mahmoud, 1992, for a survey). For example, the depth D_n of X_n (that is, the path distance to the root) satisfies

$$\frac{D_n}{2 \log n} \to 1 \text{ in probability}$$

(Lynch, 1965, and Knuth, 1973). In fact, $(D_n - 2\log n)/\sqrt{2 \log n} \xrightarrow{\mathcal{L}}$ normal $(0, 1)$, where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution (Devroye, 1988). Therefore, most nodes may be found at distance $2 \log n \pm O(\sqrt{\log n})$ from the root. The height H_n' of $\mathcal{R}_n$ satisfies

$$\frac{H_n'}{\log n} \to 4.31107\ldots \text{ almost surely}$$

(Robson, 1979, 1982; Devroye, 1986, 1987). All levels of the tree are full up to about level $0.3711\ldots \log n$ (Devroye, 1986). The constants given above are the two solutions of the equation $c \log(2e/c) = 1$. These properties may be used for the purpose of comparison, as they exemplify ideal behavior. Departures of randomness will very likely result in trees with radically different behavior. All of the above properties are summarized in the figure below.

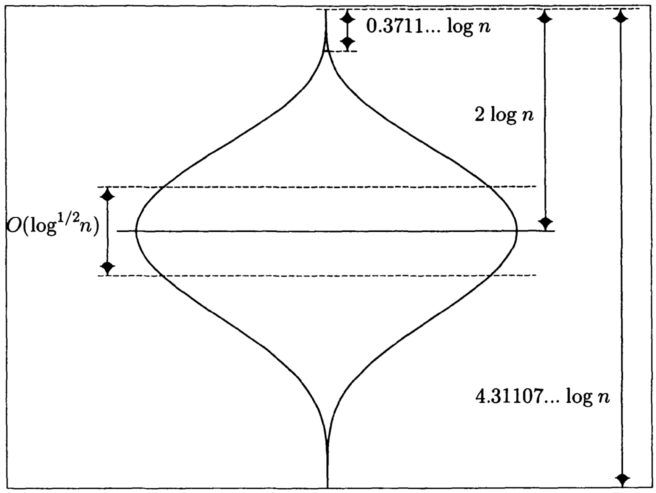

FIGURE 2. The properties of a random binary search tree are summarized. The width of the blob is proportional to the number of nodes at a given level.

3 The structure of Weyl trees

The following notation will be used. The height of $\mathcal{T}_n(\theta)$ is $H_n(\theta)$. The set of leaves of $\mathcal{T}_n(\theta)$ is $\mathcal{L}_n(\theta)$. The collection of $n+1$ possible positions for a new node to be added to $\mathcal{T}_n(\theta)$ is called the set of external nodes, and is denoted by $\mathcal{E}_n(\theta)$. When θ is understood, the suffix (θ) will be dropped from the notation. The collection $\mathcal{E}_n$ may be split into $\mathcal{E}_n^R$ and $\mathcal{E}_n^L$, where $\mathcal{E}_n^R$ has those nodes that are right children, and $\mathcal{E}_n^L$ collects all left children in $\mathcal{E}_n$.

In this section, we look at the structure of a Weyl tree for fixed irrational θ. Crucial connections are made with the continued fraction for θ. Well-known properties of continued fractions then permit us to deduce results about H_n and $|\mathcal{L}_n|$ without too much trouble.

Let $1 = T_1 < T_2 < \cdots$ be the record times, i.e., the times at which $x_n = \{n\theta\}$ is minimum or maximum among $x_1, \ldots, x_n$. The times of occurrence of a minimum or maximum are denoted by L_n and R_n, and the indices of these sequences are synchronized with the T_i's as follows:

$$(L_n, R_n) = \begin{cases} (L_{n-1}, T_n) & \text{if at } T_n \text{ there is a maximum;} \\ (T_n, R_{n-1}) & \text{if at } T_n \text{ there is a minimum.} \end{cases}$$

As it turns out, there is a lot of structure in these sequences. The fundamental property in this respect is the following.

Lemma 1 (Ellis and Steele, 1981). *We have*

$$(L_n, R_n) = \begin{cases} (L_{n-1}, L_{n-1} + R_{n-1}) & \text{if at } T_n \text{ there is a maximum;} \\ (L_{n-1} + R_{n-1}, R_{n-1}) & \text{if at } T_n \text{ there is a minimum.} \end{cases}$$

Let k be the smallest integer such that $n < L_k + R_k$. Then, if $x_{(1)} < \dots < x_{(n)}$ denotes the ordered sequence for $x_1, \dots, x_n$, then the indices $(1), \dots, (n)$ coincide with

$$\{L_k, 2L_k, 3L_k, \dots,\} \pmod{(L_k + R_k)} \cap \{1, \dots, n\} .$$

Also, (L_n, R_n) are relatively prime for all n.

A quick verification: if $n = L_k + R_k - 1$, then the index of the maximum is $(L_k + R_k - 1)L_k \pmod{(L + k + R_k)} = -L_k \pmod{(L_k + R_k)} = R_k$, as was expected. This Lemma says that at $n = L_k + R_k - 1$, the shape of the binary search tree for $x_1, \dots, x_n$ is entirely determined by the two numbers L_k and R_k. In fact, then, there are only $O(n^2)$ possible Weyl search trees with n elements, even though there are $\frac{1}{n+1}\binom{2n}{n} = \Theta(4^n/n^{3/2})$ possible binary search trees on n nodes. As the simplest example, consider the five binary search trees on 3 nodes. Two of these are impossible to obtain as Weyl trees (the ones in which the root has one child and the child has one child but of different polarity). This fact was used by Ellis and Steele to derive a method that would sort any Weyl sequence using comparisons only (thus, without being capable of numerically inspecting entries) in $O(\log n)$ comparisons.

There is a natural way of looking at the growth of the Weyl search tree in layers. The $(i + 1) - st$ layer consists of all x_j with $T_i \le j \le T_{i+1} - 1$. A special role is played also by the ancestor tree $\mathcal{T}_{T_i - 1}$. A layer can be considered as a new coat of leaves painted on the ancestor tree. Each layer adds one and just one coat, as the next Lemma explains.

Lemma 2 (Devroye and Goudjil, 1996). *All nodes in the $(i + 1)$-st layer are leaves, and all leaves of $\mathcal{T}_{T_{i+1}-1}$ are in the $(i + 1)$-st layer. All nodes in the $(i + 1) - st$ layer are either right children or left children, but not both. In fact,*

$$|\mathcal{E}^L_{T_{i+1}-1}| = R_i \ , |\mathcal{E}^R_{T_{i+1}-1}| = L_i \ ,$$

and

$$|\mathcal{T}_{T_{i+1}-1}| = T_{i+1} - 1 = L_i + R_i - 1 \ .$$

Lemma 3. *We have $|\mathcal{L}_{T_{i+1}-1}| = \min(L_i, R_i)$, and $H_{T_{i+1}-1} = i$. Put differently, $k - 1 \le H_n \le k$ if k is the unique integer with $T_k \le n < T_{k+1}$.*

PROOF. The first statement is an immediate corollary of Lemma 2. Also, as each layer destroys all the leaves of the ancestor tree, it is clear by induction

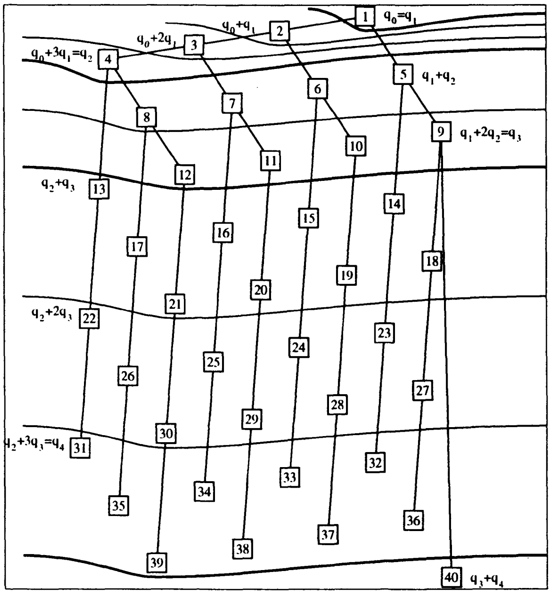

FIGURE 3. This figure shows the Weyl tree for $\theta = \sqrt{77} = [8; 1, 3, 2, 3, \ldots]$. The nodes are at positions $(x_i, -i)$ and i is shown in the boxes that represent the nodes. Layers are separated by wiggly lines. Thicker lines separate layers of different polarity. Note that just before an extremum, all leaves may be found in the last layer. Using Lemma 1, can the user guess who the parent will be of x_{41}? And how many nodes are there in the layer started with x_{40}? The heads of the layers have indices that satisfy a certain recurrence; the integers in this recurrence are $q_0 = 1, q_1 = 1, q_2 = 4, q_3 = 9, q_4 = 31$.

that the height of the tree is exactly equal to the number of layers minus one. $\square$

The study of the height and of the number of leaves reduces to the study of the sequence (L_i, R_i). For the height, the growth of T_k as a function of k is important. This is closely related to the continued fraction expansion

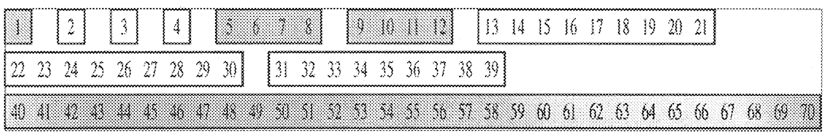

FIGURE 4. The figure shows the break-up of the Weyl sequence for $\theta = \sqrt{77}$ into layers. The shaded layers are all leaves in the tree of right polarity. Each layer covers the existing tree with a new layer of leaves.

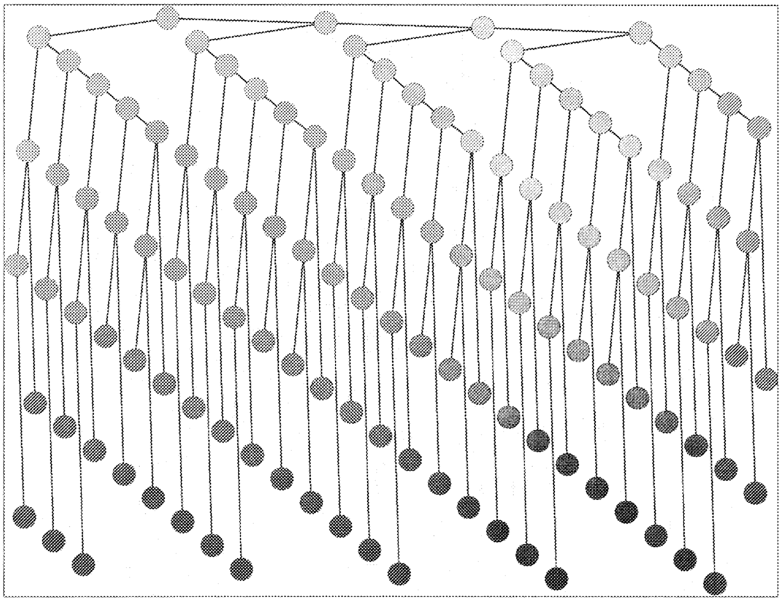

FIGURE 5. The Weyl tree $\mathcal{T}_n(\sqrt{1311})$ in which each X_i is placed at $(X_i, -i)$ exhibits an obvious orchard pattern. This phenomenon is observed for all θ. Note that $\sqrt{1311} = [36; (4, 1, 4, 2, 1, 2, 4, 1, 4, 72)^*]$.

of θ. To understand the rest of the paper, we recall a few basic facts from the theory of continued fractions.

4 Continued fractions

Let θ be irrational, and define the Weyl sequence with n-th term $x_n = \{n\theta\}$, $n \geq 1$, where $\{.\}$ denotes the "modulo 1" operator: $\{u\} = u - \lfloor u \rfloor$. Denote the continued fraction expansion of θ by

$$\theta = [a_0; a_1, a_2, \ldots] \ ,$$

where the a_i's are the partial quotients, $a_i \geq 1$ for $i \geq 1$ (see Lang (1966) or LeVeque (1977)). Thus, we have

$$\theta = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \frac{1}{a_3 + \cdots}}} \ ,$$

with $a_0 = \lfloor \theta \rfloor$. The i-th convergent of θ is

$$r_i = [a_0; a_1, \ldots, a_i] \ .$$

It can be computed recursively as $r_i = \frac{p_i}{q_i}$, where $\gcd(p_i, q_i) = 1$, and

$$p_{-2} = 0, p_{-1} = 1, p_i = a_i p_{i-1} + p_{i-2} \ , \ i \geq 0 \ ,$$

and

$$q_{-2} = 1, q_{-1} = 0, q_i = a_i q_{i-1} + q_{i-2} \ , \ i \geq 0 \ .$$

Note that $r_0 = a_0$ and $r_1 = a_0 + 1/a_1$. The r_i's alternately underestimate and overestimate θ. The denominators q_i of the convergents play a special role as

$$1 = q_0 \leq q_1 \leq q_2 \leq \cdots$$

and

$$\left| \theta - \frac{p_i}{q_i} \right| \leq \frac{1}{q_i q_{i+1}} \ , \ i > 0 \ .$$

To study the number of records and the the evolution of the layers, the following result is essential. It extends a theorem of Lang (1966).

Lemma 4 (Boyd and Steele, 1978). *In a Weyl sequence for an irrational θ with partial quotients a_n, and convergents p_n/q_n, there are a_1 right extrema, followed by a_2 left extrema, a_3 right extrema and so forth. These extrema occur when n is in the following list:*

a_1 right extremes $q_{-1} + q_0$, $q_{-1} + 2q_0$, $\ldots$, $q_{-1} + a_1 q_0 = q_1$

a_2 left extremes $q_0 + q_1$, $q_0 + 2q_1$, $\ldots$, $q_0 + a_2 q_1 = q_2$

a_3 right extremes $q_1 + q_2$, $q_1 + 2q_2$, $\ldots$, $q_1 + a_3 q_2 = q_3$

a_4 left extremes $q_2 + q_3$, $q_2 + 2q_3$, $\ldots$, $q_2 + a_4 q_3 = q_4$

$\ldots$

The list above is the list of indices of the first elements in each layer.

Lemma 4 shows that we start with a_1 right extremes, followed by a_2 left extremes, then a_3 right extremes, and so forth. This description, together with Lemma 1 and Lemma 2 should suffice to completely describe the tree. In fact, it suffices to show the right and left extrema in groups of sizes a_1, a_2, and so forth, and let the user paint in the layers of leaves one by one. In obvious notation, we may represent θ as $R^{a_1} L^{a_2} R^{a_3} L^{a_4} \ldots$, where L and R stand for left extremes and right extremes respectively. We note that this is precisely a representation of an infinite path in the Stern-Brocot tree (L means that you should take a left turn and R a right turn), which was found in the nineteenth century and is described in detail in Graham, Knuth and Patashnik (1989). Basically, the Stern-Brocot tree is an infinite binary tree which holds all possible convergents so that irrational numbers correspond to infinite paths.

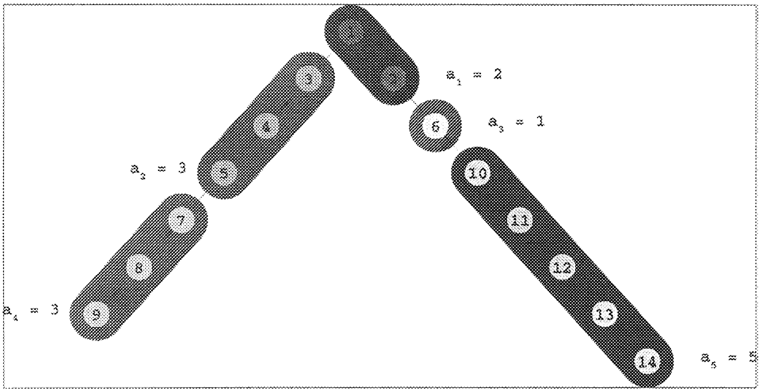

FIGURE 6. The groups of right and left extrema are shown. To construct the Weyl tree, first paint in a_1 layers of right leaves, then a_2 layers of left leaves, the a_3 layers of right leaves, and so forth. The "guts" of the tree are not shown in the figure.

5 Height of random Weyl trees

In this section, we look at random Weyl trees, that is, binary search trees for $\theta = U$, where U is a uniform $[0,1]$ random variable. This study allows us to make statements that are true for almost all θ. The probabilistic setting comes in handy for the purpose of analysis. The main result of the next three sections, for example, shows that

$$\frac{H_n}{\log n \ \log\log n} \to \frac{12}{\pi^2} \text{ in probability,}$$

where $c > 0$ is a universal constant. This shows that $\mathcal{T}_n$ differs greatly from $\mathcal{R}_n$.

From Lemma 3 and Lemma 4, we easily determine the relationship between height and partial quotients.

Proposition 1. *Let θ be irrational. Let $k \geq 2$. If $n = q_k - 1$, then there are exactly $\sum_{i=1}^{k} a_i - 1$ full layers, and the Weyl tree $\mathcal{T}_n$ has height $H_n = \sum_{i=1}^{k} a_i - 2$. In general, if $q_k \leq n < q_{k+1}$, then*

$$\sum_{i=1}^{k} a_i \leq H_n + 2 \leq \sum_{i=1}^{k+1} a_i \ .$$

Remark (Discrepancy). There is another field in which the behavior of the partial sums $S_n = \sum_{i=1}^{n} a_i$ matters. In quasi-random number generation, the notion of discrepancy is important. In general, the discrepancy for a sequence $x_1, \ldots, x_n$ is

$$D_n = \sup_{A \in \mathcal{A}} \left| \frac{\sum_{i=1}^n I_{x_i \in A}}{n} - \lambda(A) \right| ,$$

where $\lambda(.)$ denotes Lebesgue measure, and $\mathcal{A}$ is a suitable subclass of the Borel sets. For example, if we take the intervals, then (Schmidt, 1972; Béjian, 1982)

$$D_n \geq \frac{0.12 \log n}{n}$$

infinitely often. From Niederreiter (1992, p. 24), we note that for a Weyl sequence for irrational θ,

$$D_n \leq \frac{1}{n} \sum_{i=1}^{l(n)+1} a_i = \frac{S_{l(n)+1}}{n} ,$$

where $l(n)$ is the unique integer with the property that

$$q_{l(n)} \leq n < q_{l(n)+1} .$$

For example, Niederreiter's bound implies that if θ is such that $\sum_{i=1}^m a_i = O(m)$ (as when all a_i's are bounded), then

$$D_n = O\left(\frac{\log n}{n}\right) .$$

Thus, Weyl sequences with small partial quotients behave well in this sense. We will see that the same is true for random search trees based on Weyl sequences.

6 Partial quotients of random irrationals

Replace θ by a uniform $[0, 1]$ random variable. Its continued fraction has been studied in depth by Khintchine (1963), Philipp (1970) and the references found there.

Lemma 5 (Borel-Bernstein theorem; see Khintchine, 1924). *For almost all θ, $a_n \geq \varphi(n)$ infinitely often if and only if $\sum_n 1/\varphi(n) = \infty$. (Thus, if θ is uniform $[0, 1]$, then with probability one, $a_n \geq n \log n \log \log n$ infinitely often, for example.)*

This shows that the a_n's necessarily have large oscillations. The result can also be used to show that certain subclasses of θ's have zero measure. Examples include:

a. The θ's with bounded partial quotients. The extreme example here is $\theta = \frac{1+\sqrt{5}}{2}$, which has $a_0 = a_1 = a_2 = \cdots = 1$.

b. The θ's that are quadratic irrationals (solutions of quadratic equations). It is known that the a_i's are eventually periodic and thus bounded (in fact, the periodicity characterizes the quadratic irrationals, see Khintchine, 1963, p. 56).

Lemma 6 (Kusmin, 1928; Lévy, 1929). *Let z_n denote the value of the continued fraction*

$$[0; a_{n+1}, a_{n+2}, \ldots] .$$

(That is, $z_n = r_n - a_n = \{r_n\}$, where

$$r_n = [a_n; a_{n+1}, a_{n+2}, \ldots] .)$$

Then, if θ is uniform $[0,1]$, then z_n tends in distribution to the so-called Gauss-Kusmin distribution with distribution function

$$F(x) = \log_2(1+x) \ , \ \ 0 \le x \le 1 \ .$$

This limit theorem is easy to interpret if we consider convergents. Indeed, $r_0 = \theta$, and in general, $a_{n+1} = \lfloor 1/z_n \rfloor$. Thus, Lemma 6 also gives an accurate description of the limit law for a_n. In fact, as a corollary, one obtains another result of Lévy (1929), which states that the proportion of a_i's taking value k tends for almost all θ to a finite constant only depending upon k. If θ is uniform $[0,1]$, then $a_1 = \lfloor 1/\theta \rfloor$ is a discretized version of a uniform $[0,1]$ random variable. As n grows, the distribution gradually shifts to a discretized version of one over a Gauss-Kusmin random variable. As the latter law has a density $f(x) = 1/((1+x)\log 2)$ on $[0,1]$ which varies monotonically from $1/\log 2$ to $1/\log 4$, for practical purposes, it is convenient to think of the a_n's as having a law close to that of $1/U$. For example, the Borel-Bernstein law holds also for the sequence $1/U_n$ where $U_1, U_2, \ldots$ are i.i.d. uniform $[0,1]$.

There is stability if we start the process with θ having the Gauss-Kusmin law, just if we were firing up a Markov chain by starting with the stationary distribution: if θ has the Gauss-Kusmin law, then all z_n's have the Gauss-Kusmin law, and all the a_n's have the same distribution (however, they are not independent.)

We must master the dependence between the a_n's. This was done by Philipp in 1970:

Lemma 7 (Philipp, 1970). *Let θ have the Gauss-Kusmin distribution. Let $\mathcal{M}_{u,v}$ be the smallest σ-algebra with respect to which the coefficients $a_u, \ldots, a_v$ are measurable. Then for any sets $A \in \mathcal{M}_{1,t}$ and $B \in \mathcal{M}_{t+n,\infty}$,*

$$|\mathbf{P}\{AB\} - \mathbf{P}\{A\}\mathbf{P}\{B\}| < c\rho^n \mathbf{P}\{A\}\mathbf{P}\{B\} \ ,$$

where $\rho \in (0,1)$ and c is a constant.

This result states that in effect the a_n's are almost independent, with the dependence decreasing in an exponential fashion. Next we consider the behavior of partial sums of the partial quotients of a random Weyl sequence, and obtain a limit law. More precisely, we study the behavior of

$$S_n = \sum_{i=1}^{n} a_i$$

when θ is replaced by U, a uniform $[0,1]$ random variable. Proposition 2 below was obtained without the help of modern probability theoretical tools such as Lemma 7 by Khintchine (1935). A short proof of it and of Theorem 1 below is given in Devroye and Goudjil (1996).

Proposition 2. *If θ is Gauss-Kusmin distributed, then*

$$\frac{\sum_{i=1}^{n} a_i}{n \log_2 n} \to 1$$

in probability.

Proposition 2 may be rephrased as follows, if A_n denotes the collection of all θ's on $[0,1]$ with $|\sum_{i=1}^{n} a_i/(n\log_2 n) - 1| > \epsilon$:

$$\lim_{n\to\infty} \mathbf{P}\{\theta \in A_n\} = 0 .$$

Theorem 1. *If θ has a distribution with a density on $[0,1]$, then*

$$\frac{\sum_{i=1}^{n} a_i}{n \log_2 n} \to 1$$

in probability.

7 The behavior of the denominator of the convergents

Lemma 8 (Khintchine, 1935, and Lévy, 1937; see Khintchine, 1963, p. 75). *There exists a universal constant $\gamma = \pi^2/12\ln 2 \approx 1.186569111$ such that for almost all θ,*

$$q_n = e^{(\gamma + o(1))n} .$$

Lemma 8 is related to the property (Khintchine, 1963, p. 101) that

$$\left(\prod_{i=1}^{n} a_i\right)^{1/n} \to c \stackrel{\text{def}}{=} \prod_{j=1}^{\infty}\left(1 + \frac{1}{j(j+2)}\right)^{\frac{\ln j}{\ln 2}}$$

for almost all θ. Indeed, to get this intuition, recall from the recurrences for the q_n's that

$$q_{n+1} = a_{n+1}q_n + q_{n-1} \leq (a_{n+1}+1)q_n \ ,$$

so that

$$q_n \leq \prod_{j=1}^{n}(1+a_j) \leq 2^n \prod_{j=1}^{n} a_j \ . \square$$

We also note that q_n must grow faster than a Fibonacci sequence, as $q_{n+1} \geq q_n + q_{n-1}$. This implies that $q_n \geq \rho^{n-1}$ for all n, where $\rho = (1+\sqrt{5})/2$ is the golden ratio. Another simple lower bound is $q_n \geq 2^{(n-1)/2}$ (Khintchine, 1963, p. 18).

Theorem 2 (Devroye and Goudjil, 1996). *If θ has any density on* $[0,1]$, *then*

$$\frac{H_n}{\frac{1}{\gamma}\log n \log_2 \log n} = \frac{H_n}{\frac{12}{\pi^2}\log n \log\log n} \to 1$$

in probability. Note that $12/\pi^2 \approx 1.215854203$.

PROOF. By Theorem 1, as $k \to \infty$,

$$\sum_{i=1}^{k} a_i \sim k \log_2 k$$

in probability. Next, $\log q_k \sim \gamma k$ in probability. The latter fact implies that in probability, $k \sim (1/\gamma)\log n$ if k is the unique integer such that $q_k \leq n < q_{k+1}$. But Theorem 1 and Proposition 1 then imply that

$$\frac{H_n}{k \log_2 k} \sim \frac{H_n}{(1/\gamma)\log n \log_2 \log n} \to 1$$

in probability. $\square$

This theorem does not describe the behavior as $n \to \infty$ for a single θ (the "strong" behavior). Rather, it refers to a metric property and takes for each n a cross-section of θ's that give a height in the desired range, and confirms that the measure (probability) of these θ's tends to one. For oscillations and strong behavior, a bit more is required. By the Borel-Bernstein theorem, with probability one,

$$a_n \geq n \log n \log\log n$$

infinitely often (the Borel-Bernstein theorem yields a better lower bound, but the one given here suffices to make our point). Since with probability one, $q_k^{1/k} \to e^\gamma$ as $k \to \infty$, we see from Proposition 1 that with probability one,

$$H_n \geq (1/\gamma)\log n \log\log n \log\log\log n$$

infinitely often. Thus, Theorem 2 cannot be strengthened to almost sure convergence, as the oscillations are too wide.

It is of interest to bound the oscillations in the strong behavior as well. Also, again by the Borel-Bernstein theorem, with probability one, for all but finitely many n, $a_n \leq n \log n \log^{1+\epsilon} \log n$ for $\epsilon > 0$. This implies that with probability one, for all but finitely many n, $\sum_{j=1}^{n} a_j \leq n^2 \log n \log^{1+\epsilon} \log n$. But then, by Proposition 1 and Lemma 8, with probability one, for all but finitely many n,

$$H_n \leq \frac{2}{\gamma^2} (\log n)^2 (\log \log n)(\log \log n \log n)^{1+\epsilon} .$$

7.1 Very good trees

From the inequality of Proposition 1, we recall that $H_n = O(\log n)$ if $\sum_{i=1}^{n} a_i = O(n)$. Such irrationals have zero probability. As the most prominent member with the smallest partial sums of partial quotients, we have the golden ratio ($a_n \equiv 1$ for $n \geq 0$). Indeed, as for these sequences, $q_n \leq \prod_{i=1}^{n}(1 + a_i) \leq \exp(\sum_{i=1}^{n} a_i) = \exp(O(n))$, we have the claimed result on H_n without further ado. In fact, for the golden ratio, we have $q_n \sim c\rho^n$, where $\rho = (1 + \sqrt{5})/2$ and $c > 0$ is a constant. As $\sum_{i=1}^{n} a_i \equiv n$, we see that $H_n \sim \frac{\log n}{\log \rho}$. The Weyl tree is simply not high enough compared to typical random Weyl trees or truly random binary search trees.

If $a_n \equiv a$ for all n, then $q_n = aq_{n-1} + q_{n-2}$ for all n. This simple linear recurrence has a solution $q_n \sim c\left(\frac{a+\sqrt{a^2+4}}{2}\right)^n$ for some constant c. As $\sum_{i=1}^{n} a_i = an$, we see that

$$H_n \sim \frac{a}{\log\left(\frac{a+\sqrt{a^2+4}}{2}\right)} \log n .$$

Note that the coefficient can be made as large as desired by picking a large enough.

Quadratic algebraic numbers θ are solutions of quadratic equations with integer-valued coefficients. Such numbers have periodic partial quotients, and thus satisfy the boundedness property discussed here. If the period length is p, and the sum of the partial quotients over a period is s, then $\sum_{i=1}^{k} a_i \sim sk/p$ as $k \to \infty$. Furthermore, $q_k = (\rho + o(1))^k$ where ρ depends upon the periodic elements only. This implies that

$$H_n \sim \frac{s \log n}{p \log \rho} .$$

7.2 Very bad trees

We first show that Weyl trees can be almost of arbitrary height.

Theorem 3. *Let h_n be a monotone sequence of numbers decreasing from 1 to 0 at any slow rate. Then there exists an irrational θ such that for the Weyl tree, $H_n \geq nh_n$ infinitely often.*

PROOF. We exhibit a monotonically increasing sequence a_n of partial quotients to describe θ. The inequality will be satisfied at instants when the tree size $n = q_k$ for some k. Thus, we will have for all k large enough,

$$H_{q_k} \geq q_k h_{q_k} .$$

Now, $H_{q_k} \geq \sum_{i=1}^{k} a_i \geq a_k$, and

$$a_k \leq q_k \leq 2^k \prod_{i=1}^{k} a_i \leq 2^k a_k (a_{k-1})^{k-1} .$$

Thus,

$$\frac{H_{q_k}}{q_k} \geq \frac{1}{2^k (a_{k-1})^{k-1}} \geq h_{a_k} \geq h_{q_k}$$

by choosing a_k large enough (note that k and a_{k-1} are fixed). □

EXAMPLE 1. Take $a_k = 2^k$. Then

$$2^{k(k+1)/2} \leq q_k \leq 2^{k+k(k+1)/2} ,$$

so that $k = \sqrt{2 \log_2 n} - K - o(1)$, where $K \in [1/2, 3/2]$ and possibly depends upon n. As $\sum_{i=1}^{k} a_i = 2^{k+1} - 1$, we have at those times when $n = q_k$ for some k,

$$H_n = 2^{k+1} - 1 = \Theta\left(2^{\sqrt{2 \log_2 n}}\right) .$$

This grows much faster than any power of the logarithm.

EXAMPLE 2. If we set $a_k = 2^{2^k}$, then $q_k \leq 2^k \prod_{i=1}^{k} a_i \leq 2^{k+2^{k+1}-1} \leq \log_2(a_k) a_k^2 / 2$. Combine this with $H_{q_k} \geq a_k$, and note that when $n = q_k$ for some k,

$$H_n \geq \sqrt{\frac{2n}{\log_2 H_n}} ,$$

and therefore,

$$H_n \geq (1 + o(1)) \sqrt{\frac{4n}{\log_2 n}} .$$

EXAMPLE 3. By considering $a_k = b^{b^k}$ for integer b, the height increases at least as $(n / \log_2 n)^{1-1/b}$.

7.3 Trees for a few selected transcendental numbers

The partial quotients are known for just a few transcendental numbers. For example

$$\tan(1/2) = [0; 1, 1, 4, 1, 8, 1, 12, 1, 16, \ldots] .$$

Thus, $a_{2k} = 1$, $a_{2k+1} = 4k$, $k \geq 1$. From $q_{2k+1} = 4kq_{2k} + q_{2k-1}$ and $q_{2k} = q_{2k-1} + q_{2k-2}$, one can show (see Boyd and Steele, 1978, p. 57) that

$$4^k k! < q_{2k+1} < 8^k (k+1)!$$

and

$$q_{2k+1} \approx q_{2k+2} \approx (ck)^k$$

for some constant c. In fact, then, we see that the k for Theorem 2 satisfies

$$k \sim \frac{\log n}{\log \log n} .$$

But then

$$H_n \sim \sum_{j=1}^{k/2} (4j) \sim \frac{k^2}{2} \sim \frac{\log^2 n}{2 \log^2 \log n} .$$

The Weyl tree is much higher than that of a typical random Weyl tree.

In a second example, consider

$$e = [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \ldots]$$

so that $a_0 = 2$, $a_{3m} = a_{3m-2} = 1$ and $a_{3m-1} = 2m$ for $m \geq 1$. Then (Lang, 1966, p. 74) there exist constants C_1 and C_2 such that

$$C_1 4^n \Gamma(n + 3/2) \leq q_{3n+1} \leq C_2 4^n \Gamma(n + 3/2) .$$

This shows that $k \sim \log n / \log \log n$. Thus,

$$H_n \sim \frac{\log^2 n}{9 \log^2 \log n} .$$

Again, the Weyl tree has an excessive height.

8 Sorting Weyl sequences

Ellis and Steele (1981) have shown that the first n elements of any Weyl sequence can be sorted with the aid of $O(\log(n))$ comparisons only, even though these sequences too are equi-distributed for any irrational θ. This shows that such sequences possess a lot of structure. Of course, the fact that discrete random Weyl sequences and random Lehmer sequences are imperfect is because they can be "described" very simply by a small number of bits. The randomness of a sequence has been related by several authors to the length of the descriptors (see e.g. Martin-Löf (1966), Knuth (1973), Bennett (1979)). For surveys and discussions on the topic of uniform random variate generation, one could consult Niederreiter (1991) or L'Ecuyer (1989).

It is well-known that the number of comparisons needed in quicksort is equal to the sum of the depths of all the nodes in the binary search tree constructed from the data by ordinary insertion. As this sum is bounded from below by $H_n(H_n + 1)/2$ (just by summing over the path leading to the furthest node), we see that the number of comparisons in quicksort is infinitely often at least equal to $nh_n(nh_n + 1)/2$ for any sequence h_n decreasing to zero, and some irrational θ. Yet, for i.i.d. data drawn from the same nonatomic distribution, the expected number of comparisons is asymptotic to $2n \log n$ (Sedgewick, 1977). Therefore, Weyl sequences are not appropriate for generating test data for sorting algorithms. With a uniform $[0,1]$ θ, the expected number of comparisons grows as $n \log n \ \log\log n$. In fact, we have the following.

Proposition 3 (Devroye and Goudjil, 1996). *Let θ be uniform $[0,1]$. For any constant C, with probability one, the number of comparisons for quicksort-ing the first n numbers of a random Weyl sequence is infinitely often larger than*

$$Cn \log n \log\log n \log\log\log n \ .$$

9 Lehmer sequences

When a is integer, the sequence

$$x_n = \{a^n x_0\}$$

will be called a Lehmer sequence. Lehmer sequences are equi-distributed for all integer $a > 1$ and almost all $x_0 \in (0,1)$ (see Freiberger and Grenander (1971), Hlawka (1984) or Kuipers and Niederreiter (1974)). The behavior of these sequences models the behavior of multiplicative sequences on infinite-wordsize machines. In particular, if Lehmer sequences don't have a certain desirable property, then it is just about impossible that multiplicative sequences possess that property. In addition, the multiplier a in the Lehmer sequence is inherited from the multiplicative generator, because it determines the same geometric increase in the sequence of values (before truncation due to the $\{.\}$ operation). In fact, the "complexity" of the generator can be measured by a. This continuous model is at the same time the genesis and an abstraction of the discrete finite wordsize sequences. It was employed by many to explain certain properties, gain insight, and advance the understanding of certain generators.

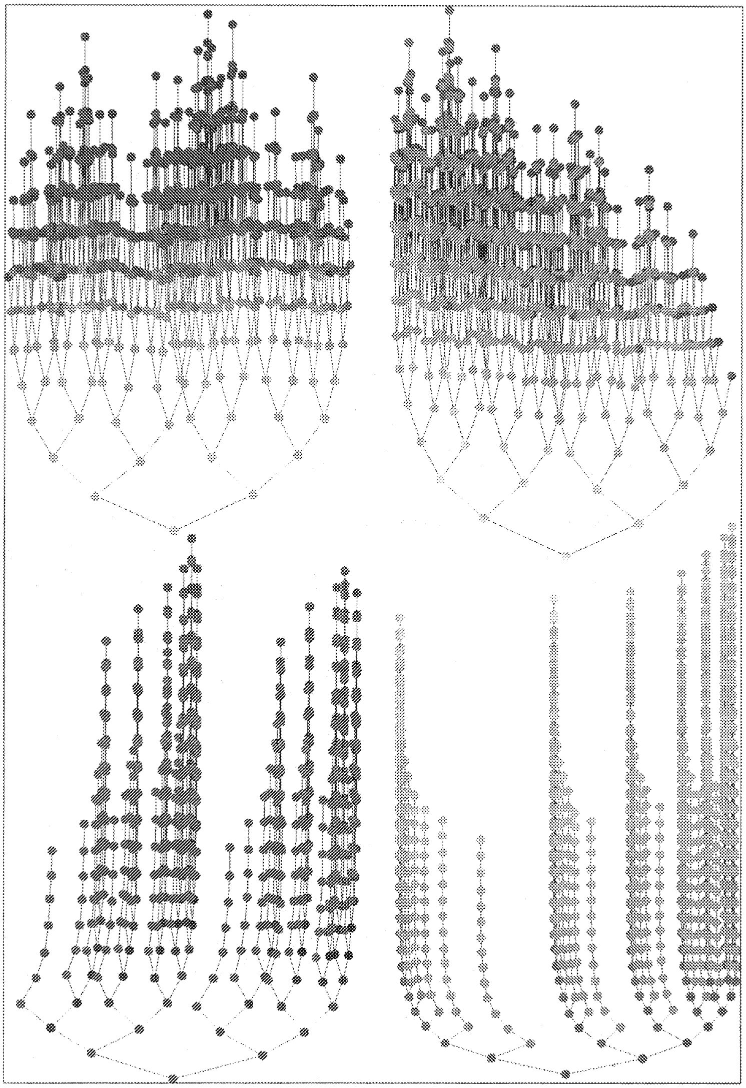

FIGURE 7. Four Weyl trees with θ, from top left to bottom right ranging from $\sqrt{2}$, $\sqrt{3}$, $\tan(1/2)$ and π. Colors (gray-levels) represent the various layers of paint. The points are slightly perturbed so that one gets a better impression of the number of nodes at each level.

For a not integer, the sequence obtained by $x_{n+1} = \{ax_n\}$ does not coincide with $x_n = \{x_0 a^n\}$. In fact, while the latter sequence is equi-distributed for almost all x_0, it is easy to show that the former sequence is not equi-distributed for any value x_0.

EXAMPLE 7 (x_0 IS RATIONAL). In computers, we can only store rational numbers. If we begin with a rational $x_0 = k/m$ for integers k and m that are relatively prime, then x_0 has a periodic decomposition base a:

$$\{x_0\} = \sum_{i=1}^{\infty} \frac{x_{0i}}{a^i} \ ,$$

where $x_{0i} \in \{0, 1, \ldots, a-1\}$, and the sequence $\{x_{0i}\}$ is periodic for all $i \geq N$ and some N. Therefore, the sequence $x_N, x_{N+1}, \ldots$ is periodic, so that $\{x_n, n \geq 1\}$ is not equi-distributed.

EXAMPLE 8 (POOR CHOICES FOR x_0). Decompose x_0 base a as in the previous example. Let $N_{0,n}, \ldots, N_{a-1,n}$ denote the number of values in $\{x_{01}, \ldots, x_{0n}\}$ equal to $0, 1, \ldots, a-1$ respectively. Equi-distribution implies

$$\lim_{n \to \infty} \frac{N_{i,n}}{n} = \frac{1}{a} \ , \ 0 \leq i < a \ .$$

All integers in the decomposition must occur equally frequently. In fact, equi-distribution occurs if and only if for all integer $k \geq 1$, and all $b = (b_1, \ldots, b_k) \subseteq \{0, \ldots, a-1\}^k$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I_{(x_{0,i+1}, \ldots, x_{0,i+k}) = b} = \frac{1}{a^k} \ .$$

Thus, all pairs, triples, and so forth must occur equally often. Thus, x_0 itself is very restricted. If the bits in a binary expansion of x_0 are independent Bernoulli (p), then the only acceptable p is $1/2$. It is also the only value for which x_0, the random variable with this bit expansion, has a density (the uniform density). For all other values, x_0 has a singular distribution. Considered in this manner, we see that the vast majority of x_0's will in fact be poor. It is easy to construct an irrational x_0 for which there is no equi-distribution.

A random Lehmer sequence has an integer multiplier a and a uniform $[0, 1]$ random seed X_0. How such a seed is obtained is of course problematic. In fact, having X_0 is like having an infinite i.i.d. sequence of uniform $[0, 1]$ random variables $X_1, X_2, \ldots$. To see this, consider the bit representation of X_0:

$$X_0 = 0.X_{01} X_{02} X_{03} \cdots ,$$

and distribute the bits in a triangular fashion among the X_i's: X_{01} goes to X_1, X_{02} and X_{03} go to X_1 and X_2 respectively, X_{04} through X_{06} go to X_1 through X_3 and so forth. This establishes a one-to-one mapping between the bits of X_0 and all the bits of all X_i's. But since an i.i.d. sequence of bits (with 1's and 0's occurring with equal probability), when considered as a fraction of a real number, defines a uniform $[0, 1]$ random variable, we see that $X_1, X_2, \ldots$ are indeed i.i.d. uniform $[0, 1]$ as claimed. This perfect generator is, however, not of the form $X_{n+1} = f(X_n)$, so that we need not celebrate prematurely.

A random Lehmer sequence has obvious imperfections. For example, the pairs (X_n, X_{n+1}) fall on one of many parallel lines in the plane because $X_{n+1} - aX_n$ is integer-valued. What we would like to discuss here is how the sequence behaves globally. Knuth (1973, p. 88) spends some time on the present model. He shows for example that while $\mathbf{P}(X_{n+1} < X_n) = 1/2$,

$$\mathbf{P}(X_{n+2} < X_{n+1} < X_n) = \frac{1}{6} + \frac{1}{6a} + O(a^{-2}).$$

This value should be $1/6$ and is thus wrong. Starting with X_0, a monotone run of length k occurs if $X_0 \geq X_1 \geq \cdots \geq X_{k-1} < X_k$. The expected length of such a run is shown by Knuth to be

$$\left(\frac{a}{a-1}\right)^a - \frac{a}{a-1} = e - 1 + \frac{e/2 - 1}{a} + O(a^{-2}),$$

which is too big—for an i.i.d. sequence the answer is $e - 1$.

10 Binary search trees for Lehmer sequences

A Lehmer tree is a binary search tree for a Lehmer sequence, and thus has two parameters, an integer multiplier a and a real-valued seed x_0. A random Lehmer tree is a Lehmer tree in which x_0 is uniformly chosen on $[0, 1]$ (and thus has one parameter, a). We will point out a few flaws of particular Lehmer trees. For example, in view of the fact that the height of $\mathcal{R}_n$ has expected value $\Theta(\log n)$, it is puzzling that even with good multipliers, binary search trees based upon Lehmer sequences have heights that can take any value $o(n)$ for all n large enough.

Theorem 4. *Let $a > 1$ be an arbitrary integer. Let g_n be a monotonically increasing sequence of positive integers with $g_n = o(n)$. Then there exist infinitely many seeds x_0 for which the Lehmer sequence $\{x_0 a^n\}$ is equidistributed, yet $H_n \geq g_n$ for all n large enough, where H_n is the height of the binary search tree with n nodes obtained by inserting the first n numbers of the Lehmer sequence.*

PROOF. We proceed by showing that for a specially constructed random seed x_0, the Theorem is true with probability one, and thus, that there must

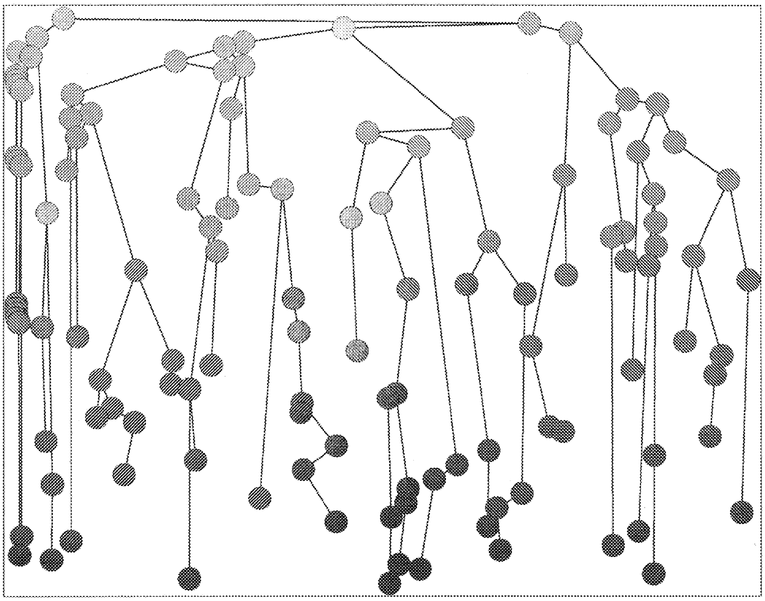

FIGURE 8. A random Lehmer tree with multiplier a, in which each X_i is placed at $(X_i, -i)$. The orchard pattern from Weyl trees has disappeared.

exist infinitely many deterministic seeds x_0 for which the Theorem is true. Let x_0 be a uniformly distributed random variable on $[0, 1]$. It is well-known that its fraction has independent digits that are uniformly distributed on $\{0, \ldots, a-1\}$. With probability one, all k-tuples occur equally often in the limit (see above), and the Lehmer sequence is thus equi-distributed. Define $n_i = 2^i$. Define an integer sequence h_{n_i} recursively as follows: $h_{n_0} = 0$, and for $i \geq 0$,

$$h_{n_{i+1}} = h_{n_i} + 2i + g_{n_{i+2}} \ .$$

Observe that h_{n_i} is increasing, and that $h_{n_i} = o(n_i)$. That follows from telescoping the recurrence relation and noting that

$$h_{n_i} = g_{n_{i+1}} + g_{n_i} + \cdots + g_{n_2} + i(i-1) \ .$$

Replace $h_{n_{i+1}}$ digits starting at digit $n_i + 1$ by 0, for $i = 1, 2, \ldots$. Among the first n_i digits, at most $h_{n_2} + \cdots + h_{n_i}$ digits have been thus modified. Clearly, this is $o(n_i)$ as $i \to \infty$. Therefore, all k-tuples still occur equally often in the limit, and the Lehmer sequence for the new x_0 is equi-distributed.

In a sequence of i.i.d. random digits $X_1, \ldots, X_n$, with $\mathbf{P}\{X_i = i\} = 1/a$, $0 \leq i < a$, let L_n be the length of the longest run of zeros. Clearly,

$$\mathbf{P}\{L_n \geq k\} \leq \frac{n}{a^k} \ .$$

Let L_n now denote the same thing for the first n digits of x_0. We have added a few sequences of consecutive zeros, but obvious calculations show that

$$\mathbf{P}\{L_{n_i} \geq k + h_{n_i}\} \leq \frac{n_i}{a^k} .$$

The infinite string starting at position $n_i + 1$ starts with $h_{n_{i+1}}$ zeros. It and the next $h_{n_{i+1}} - h_{n_i} - k$ infinite strings form a father/right child chain in the binary search tree if $L_{n_i} < k + h_{n_i}$. Therefore,

$$H_{n_i+h_{n_{i+1}}} \geq h_{n_{i+1}} - h_{n_i} - k$$

with probability at least $1 - n_i/a^k$. We assume without loss of generality that $n_i + h_{n_{i+1}} < n_{i+1}$. Thus,

$$H_{n_{i+1}} \geq h_{n_{i+1}} - h_{n_i} - 2i$$

with probability at least $1 - n_i/a^{2i}$. Let A be the event that all these inequalities hold except finitely often. Then, by the Borel-Cantelli lemma, $\mathbf{P}\{A\} = 1$ if

$$\sum_{i=1}^{\infty} \frac{n_i}{a^{2i}} < \infty .$$

As $n_i = 2^i$, this is indeed the case. Therefore, with probability one, for all i large enough,

$$H_{n_{i+1}} \geq h_{n_{i+1}} - h_{n_i} - 2i .$$

Assume i is such a large integer. Take $n \in [n_{i+1}, n_{i+2})$. Then

$$\begin{aligned} H_n &\geq H_{n_{i+1}} \\ &\geq h_{n_{i+1}} - h_{n_i} - 2i \\ &= g_{n_{i+2}} \quad \text{(by definition of } h\text{)} \\ &\geq g_n \quad \text{(by monotonicity of } g\text{)} \end{aligned}$$

so that $H_n \geq g_n$ for all n large enough. This concludes the proof. □

11 Virtual trees

The random trees introduced in this note may also be used as crude compact representations of real-life trees for possible use in graphics. One could take each tree and associate a branch with each node, and add nodes at the end of the leaf branches. If the branches are randomly rotated with respect to their parent branches, and if the branches have a direction that is somehow forced to grow towards the sun (north, up), and if we attach

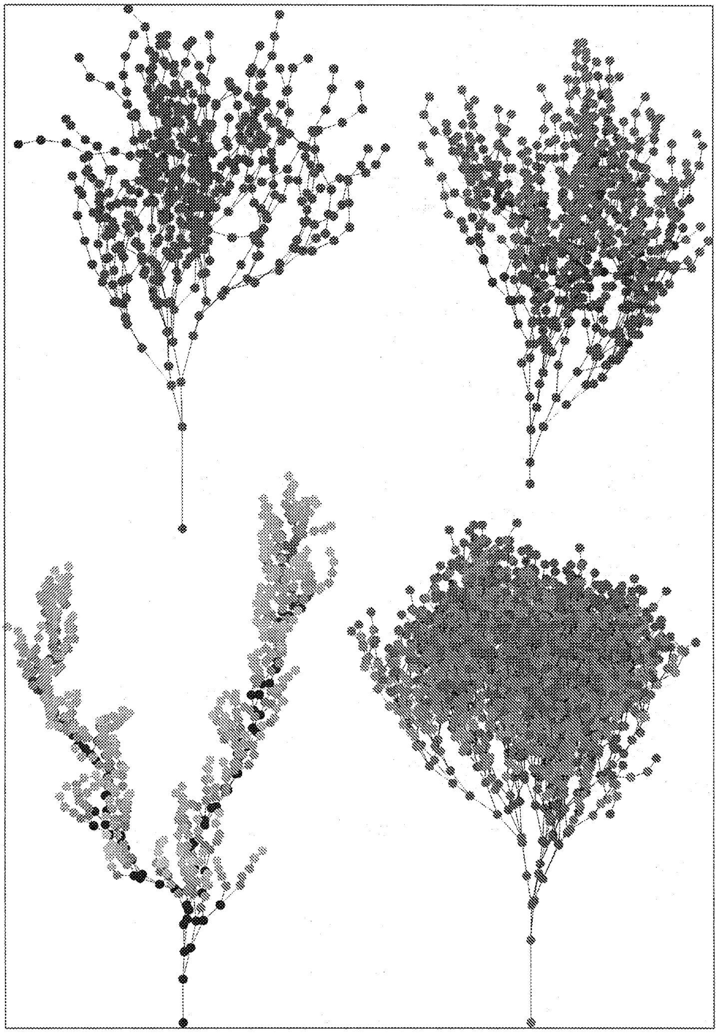

FIGURE 9. Four Weyl trees with θ, from top left to bottom right ranging from e, $\tan(1/2)$, $e^{2\pi}$ and $\sqrt{7}$. Colors (gray-levels) represent the various layers of paint.

random branch lengths, very simple but sometimes interesting tree shapes occur, as is apparent from Figure 9.

References

Aho, A.V., Hopcroft, J.E., and Ullman, J.D., 1983. *Data Structures and Algorithms*. Reading, Mass.: Addison-Wesley.

Beck, J., 1991. Randomness of $n\sqrt{2}$ (mod 1) and a Ramsey property of the hyperbola. *Colloquia Mathematica Societatis János Bolyai* 60. Budapest.

Béjian, R., 1982. Minoration de la discrépance d'une suite quelconque sur T. *Acta Arithmetica* 41. 185–202.

Bennett, C.H., 1979. *On random and hard-to-describe numbers*. IBM Watson Research Center Report RC 7483 (no 32272). Yorktown Heights, N.Y.

Bohl, P., 1909. Über ein in der Theorie der säkularen Störungen vorkommendes Problem. *Jl. Reine und Angewandte Mathematik* 135. 189–283.

Boyd, D.W. and Steele, J.M., 1978. Monotone subsequences in the sequence of fractional parts of multiples of an irrational. *Jl. Reine und Angewandte Mathematik* 204. 49–59.

Chatterji, S.D., 1966. Masse, die von regelmässigen Kettenbrüchen induziert sind. *Mathematische Annalen* 164. 113–117.

Chow, Y.S., and Teicher, H., 1978. *Probability Theory*. New York, N.Y.: Springer–Verlag.

del Junco, A. and Steele, J.M., 1979. Hammersley's law for the van der Corput sequence: an instance of probability theory for pseudorandom numbers. *Annals of Probability* 7. 267–275.

Devroye, L., 1986. A note on the height of binary search trees. *Journal of the ACM* 33. 489–498.

Devroye, L., 1987. Branching processes in the analysis of the heights of trees *Acta Informatica* 24. 277–298.

Devroye, L., 1988. Applications of the theory of records in the study of random trees. *Acta Informatica* 26. 123–130.

Devroye, L., and Goudjil, A., 1996. *A study of random Weyl trees*. Technical Report, School of Computer Science, McGill University, Montreal.

Diaconis, P., 1977. The distribution of leading digits and uniform distribution mod 1. *Annals of Probability* 5. 72–81.

Ellis, M.H., and Steele, J.M., 1981. Fast sorting of Weyl sequences using comparisons. *SIAM Journal on Computing* 10. 88–95.

Franklin, J.N., 1963. Deterministic simulation of random sequences. *Mathematics of Computation* 17. 28–59.

Freiberger, W., and Grenander, U., 1971. *A Short Course in Computational Probability and Statistics*. New York: Springer-Verlag.

Fushimi, M., and Tezuka, S., 1983. The k-distribution of generalized feedback shift register pseudorandom numbers. *Communications of the ACM* 26. 516–523.

Fushimi, M., 1988. Designing a uniform random number generator whose subsequences are k-distributed. *SIAM Journal on Computing* 17. 89–99.

Galambos, J., 1972. The distribution of the largest coefficient in continued fraction expansions. *Quarterly Journal of Mathematics Oxford Series* 23. 147–151.

Graham, R.L., Knuth, D.E., and Patashnik, O., 1989. *Concrete Mathematics—A Foundation for Computer Science*. Reading, MA: Addison-Wesley.

Hlawka, E., 1984. *The Theory of Uniform Distribution*. Berkhamsted, U.K.: A B Academic.

Kesten, H., 1960. Uniform distribution mod 1. *Annals of Mathematics* 71. 445–471.

Khintchine, A., 1924. Einige Sätze über Kettenbrüche, mit Anwendungen auf die Theorie der Diophantischen Approximationen, *Mathematische Annalen* 92. 115–125.

Khintchine, A., 1935. Metrische Kettenbruchprobleme. *Compositio Mathematica* 1. 361–382.

Khintchine, A., 1936. Metrische Kettenbruchprobleme. *Compositio Mathematica* 2. 276–285.

Khintchine, A., 1963. *Continued Fractions*. Groningen: Noordhoff.

Knuth, D.E., 1973. *The Art of Computer Programming, Vol. 3 : Sorting and Searching*. Reading, MA: Addison-Wesley.

Knuth, D.E., 1981. *The Art of Computer Programming, Vol. 2, 2nd Ed.*. Reading, Mass.: Addison-Wesley.

Kuipers, L., and Niederreiter, H., 1974. *Uniform Distribution of Sequences*. New York: John Wiley.

Kusmin, R.O., 1928. On a problem of Gauss. *Reports of the Academy of Sciences* A. 375–380.

Lang, S., 1966. *Introduction to Diophantine Approximations*. Reading, MA: Addison-Wesley.

L'Ecuyer, P. and Blouin, F., 1988. Linear congruential generators of order $k > 1$. *Proceedings of the 1988 Winter Simulation Conference*, ed. M.Abrams, P.Haigh and J.Comfort. 432–439. ACM.

L'Ecuyer, P., 1989. A tutorial on uniform variate generation. *Proceedings of the 1989 Winter Simulation Conference*, ed. E.A. MacNair, K.J. Musselman and P. Heidelberger. 40–49. ACM.

L'Ecuyer, P., 1990. Random numbers for simulation. *Communications of the ACM* 33. 85–97.

LeVeque, W.J., 1977. *Fundamentals of Number Theory*. Reading, MA: Addison-Wesley.

Lévy, P., 1929. Sur les lois de probabilité dont dépendent les quotients complets et incomplets d'une fraction continue. *Bulletin de la Société Mathématique de France* 57. 178–193.

Lévy, P., 1937. *Théorie de l'addition des variables aléatoires*. Paris.

Lynch, W.C., 1965. More combinatorial problems on certain trees. *Computer Journal* 7. 299–302.

Mahmoud, H.M., 1992. *Evolution of Random Search Trees*. New York: John Wiley.

Martin-Löf, P., 1966. The definition of random sequences. *Information and Control* 9. 602–619.

Niederreiter, H., 1977. Pseudo-random numbers and optimal coefficients. *Advances in Mathematics* 26. 99–181.

Niederreiter, H., 1978. Quasi-Monte Carlo methods and pseudo-random numbers. *Bulletin of the American Mathematical Society* 84. 957–1042.

Niederreiter, H., 1991. Recent trends in random number and random vector generation. *Annals of Operations Research* 31. 323–346.

Niederreiter, H., 1992. *Random Number Generation and Quasi-Monte Carlo Methods* 63. SIAM CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia: SIAM.

Peskun, P.H., 1980. Theoretical tests for choosing the parameters of the general mixed linear congruential pseudorandom number generator. *Journal of Statistical Computation and Simulation* 11. 281–305.

Philipp, W., 1969. The central limit problem for mixing sequences of random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 12. 155–171.

Philipp, W., 1970. Some metrical theorems in number theory II. *Duke Mathematical Journal* 38. 477–485.

Robson, J.M., 1979. The height of binary search trees. *The Australian Computer Journal* 11. 151–153.

Robson, J.M., 1982. The asymptotic behaviour of the height of binary search trees. *Australian Computer Science Communications* 88–88.

Schmidt, W.M., 1972. Irregularities of distribution. *Acta Arithmetica* 21. 45–50.

Sedgewick, R., 1977. The analysis of quicksort programs. *Acta Informatica* 4. 327–355.