A STUDY OF RANDOM WEYL TREES

Luc Devroye and Amar Goudjil
School of Computer Science
McGill University

ABSTRACT. We study binary search trees constructed from Weyl sequences $\{n\theta\}, n \geq 1$, where θ is an irrational and $\{.\}$ denotes "mod 1". We explore various properties of the structure of these trees, and relate them 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, $H_n \sim (12/\pi^2) \log n \log\log n$.

KEYWORDS AND PHRASES. Weyl sequence, random number generation, binary search trees, diophantine approximation, random continued fractions, probabilistic analysis, tree height.

Authors' address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. The authors' research was sponsored by NSERC Grant A3456 and FCAR Grant 90-ER-0291.

Introduction.

In this note, we study binary search trees formed by consecutive 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 Weyl sequence for θ, after Weyl, who showed that for all irrational θ the sequence is equi-distributed. We recall that a 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)). Mike Steele has pointed out that the credit should perhaps go to Bohl (1909), but we will nevertheless retain the commonly accepted name of Weyl sequence.

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. Knuth (1981) and Sós (1983) have interesting accounts of this. 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 permutation 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. 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 the first half of the paper, we look at the structure of a Weyl tree for fixed 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.

In the second half, 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 section, for example, shows that

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

This shows that the random Weyl tree differs greatly from the standard random binary search tree, $\mathcal{R}_n$, obtained by insertion of an i.i.d. uniform $[0,1]$ sequence $X_1, \ldots, X_n$. We recall that the height H_n' of $\mathcal{R}_n$ satisfies

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

(Robson, 1977, 1982; Devroye, 1985, 1986; Mahmoud, 1992).

Structure of Weyl trees.

In this section, an irrational θ is fixed. 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 $x_{(1)} < \ldots < x_{(n)}$ denotes the ordered sequence for $x_1, \ldots, x_n$, then the indices $(1), \ldots, (n)$ coincide with

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

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 $x_1, \ldots, x_n$ 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, of the five binary search trees on 3 nodes, two 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. We refer to the subsection on sorting later on in the paper.

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 $T_i \leq j \leq T_{i+1} - 1$. 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. *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

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

PROOF. Recall that $\mathcal{L}_{T_i-1}$ is the collection of leaves of the ancestor tree, and that the left and right external nodes of the ancestor tree are collected in sets $\mathcal{E}^L_{T_i-1}$ and $\mathcal{E}^R_{T_i-1}$ respectively. Fix $j \in \{T_i, T_i + 1, \ldots, T_{i+1} - 1\}$, so that j is an index of a point in the current $(i+1)$-st layer. Without loss of generality, assume $T_i = R_i$ (the last record was a maximum). Thus,

$$R_i \leq j < L_i + R_i \ .$$

To determine the place x_j occupies in the search tree, it is important to find out which points are the immediate predecessors and successors of x_j.

Consider first the immediate predecessor of x_j in $\{x_1, \ldots, x_{j-1}\}$. By Lemma 1, the index of this node is

$$j - L_i + k(R_i + L_i)$$

for some integer $k \geq 0$. But

$$j + R_i > j + L_i \geq R_i + L_i \ ,$$

so k must be 0, and thus, the index of the immediate predecessor is $j - L_i$, which is in the ancestor tree, as $j > L_i$ and $j - L_i < R_i$.

Similarly, the immediate successor of j has index

$$j + L_i - k(R_i + L_i)$$

for some $k \geq 0$. It cannot have index $j + L_i$ as

$$j + L_i \geq R_i + L_i \ .$$

Thus, it must have index $j + L_i - (R_i + L_i)$ or smaller, i.e., $j - R_i$ or smaller. But

$$j - R_i < L_i + R_i - R_i = L_i < R_i \ ,$$

so that $j - R_i$ belongs to the ancestor tree (if $j - R_i > 0$) or is nonexistent (if $j = R_i$).

Thus, the immediate neighbors in the ordered sequence have indices that put them in the ancestor tree (the right neighbor may not exist if $j = R_i$). As $L_i < R_i$, it is clear then that j is a right child of its left neighbor. Note also that at the end of the construction of the $(i+1)$-st layer, all nodes in it are leaves, and are right children of nodes in the ancestor tree. Thus, the $(i+1)$-st layer paints a collection of leaves on the ancestor tree. In fact, it destroys all existing leaves of the ancestor tree, as we will now prove.

We prove by induction the following:

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

As

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

we verify that indeed, at all times, the number of external nodes is equal to the tree size plus one. The statement is quickly verified for $i = 1$ as $L_1 = R_1 = 1$, $T_2 = 2$, and T_1 has one left and one right external node. Assuming the hypothesis to be satisfied for $j < i$, we look at the $(i+1)$-st layer. All nodes in this layer are leaves of $\mathcal{T}_{T_{i+1}-1}$, and if $T_i = R_i$ (without loss of generality; a symmetric statement for $T_i = L_i$ is easily obtained as well), then all these leaves fill right-external nodes of the ancestor tree $\mathcal{T}_{T_i-1}$. But by the induction hypothesis,

$$|\mathcal{E}^R_{T_i-1}| = L_{i-1} = L_i \ .$$

Also, the $(i+1)$-st layer has size

$$T_{i+1} - T_i = R_i + L_i - R_i = L_i \ ,$$

so that we can conclude that all right-external nodes of the ancestor tree are filled in. But then,

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

which was to be shown. Because all left externals survive from the ancestor tree,

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

and the proof is complete. □

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 that the height of the tree is exactly equal to the number of layers minus one. □

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

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

$$\theta = a_0 + 1/(a_1 + 1/(a_2 + \cdots)) ,$$

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, the right extrema occur when n is in the following list*

$$q_{-1} + q_0, q_{-1} + 2q_0, \ldots, q_{-1} + a_1 q_0 = q_1;$$
$$q_1 + q_2, q_1 + 2q_2, \ldots, q_1 + a_3 q_2 = q_3;$$
$$q_3 + q_4, q_3 + 2q_4, \ldots, q_3 + a_5 q_4 = q_5;$$
$$\ldots$$

and the left extrema occur when n is in the list

$$q_0 + q_1, q_0 + 2q_1, \ldots, q_0 + a_2 q_1 = q_2;$$
$$q_2 + q_3, q_2 + 2q_3, \ldots, q_2 + a_4 q_3 = q_4;$$
$$q_4 + q_5, q_4 + 2q_5, \ldots, q_4 + a_6 q_5 = q_6;$$
$$\ldots .$$

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 reconstruct the shape of the tree (see figure 1).

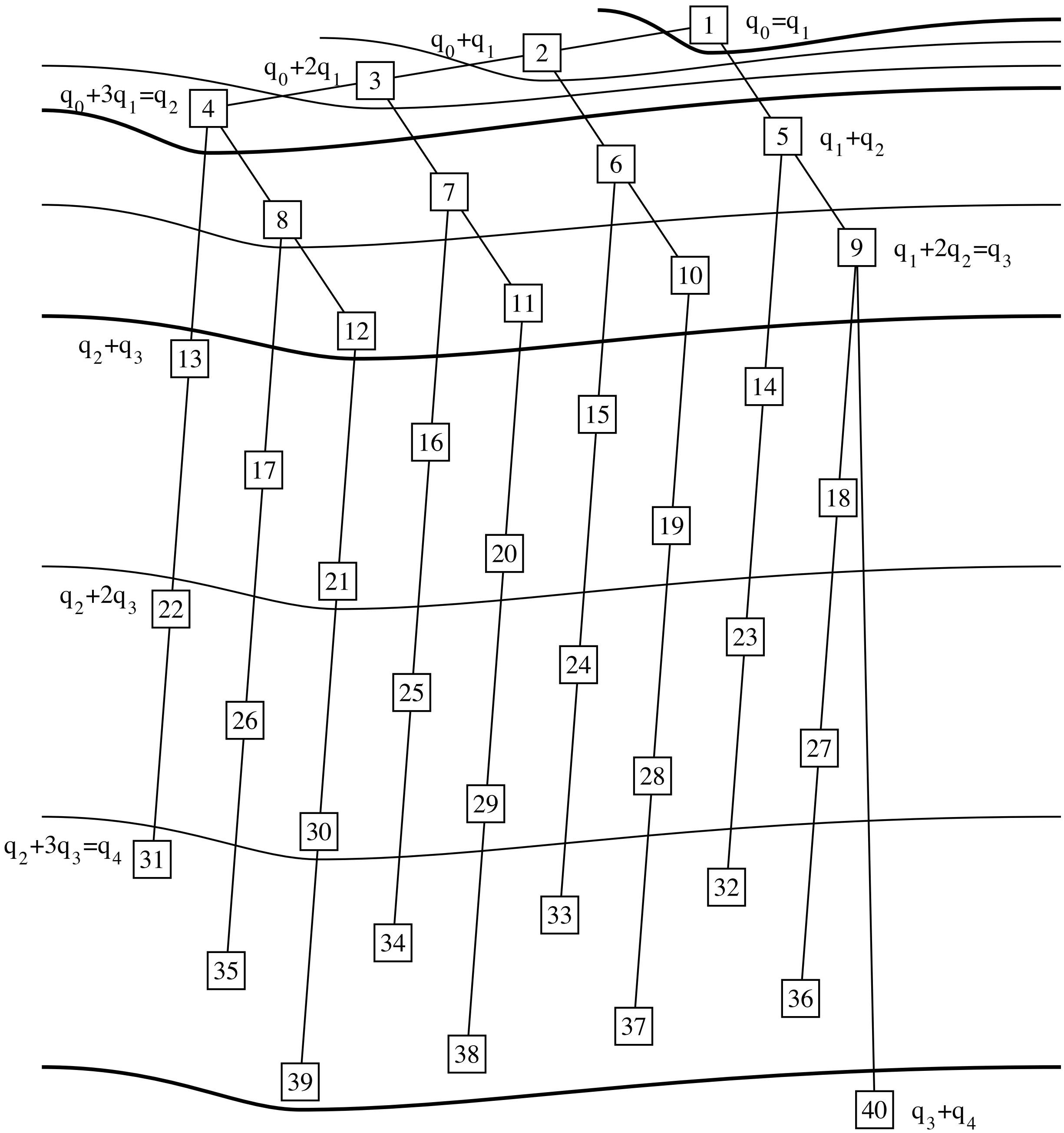

Figure 1. This figure shows the Weyl tree for $\theta = \sqrt{77} = [8; 1, 3, 2, 3, \ldots]$. Note that $q_0 = 1, q_1 = 1, q_2 = 4, q_3 = 9, q_4 = 31$. Layers are separated by wiggly lines. Thicker lines separate layers of different polarity. Note that there are first a_1 layers of right polarity, followed by a_2 layers of left polarity, and so forth. Also note that just before an extremum, all leaves may be found in the last layer. The x-coordinates of the points are geometrically exact, to facilitate interpretation. Using Lemma 1, can the reader guess who the parent is of point 41?

Height of random Weyl trees.

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

PROPOSITION 5. *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

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

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,

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

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

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

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. □

Partial quotients of random irrationals.

Now, replace θ by a uniform $[0,1]$ random variable, and consider its continued fraction expansion. Several results are known about this, and most may be found in Khintchine (1963), Philipp (1970) or the references found there.

LEMMA 6 (THE BOREL-BERNSTEIN THEOREM). *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 coefficients. The extreme example here is $\theta = (1 + \sqrt{5})/2$, which has $a_0 = a_1 = a_2 = \cdots = 1$.

B. The θ's that are quadratic irrationals (non-rational 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 7 (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 7 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 Gaus-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; in fact, Chatterji (1966) showed that any law with independent a_n's corresponds to a random θ with a singular distribution.)

LEMMA 8 (GALAMBOS, 1972). *Let θ have the Gauss-Kusmin law. Then*

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{\max_{1\le i\le n} a_i}{n} < \frac{y}{\log 2}\right\} = e^{-1/y},\ y > 0\ .$$

Galambos's result says that the excursions predicted by Borel-Bernstein are rather rare, as the maximal a_i up to time n typically has magnitude $\Theta(n)$. Of course, the difference is easily explained by the different natures of strong and weak convergence. Note that Lemma 8 remains valid if θ has the uniform distribution on $[0,1]$. The important technical contribution of Galambos is that he has mastered the dependence between the a_n's. We are faced with the same problem, and cite the fundamental result needed to make things click.

LEMMA 9 (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 $A \in M_{1,t}$ and $B \in M_{t+n,\infty}$,*

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

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. One last Lemma concludes the technical introduction.

LEMMA 10. *If P and Q are two probability measures and $\alpha > 0$ is a number such that for all rectangular Borel sets (products of intervals), $P \ge \alpha Q$, then $P \ge \alpha Q$ for all Borel sets.*

PROOF. This result should be standard. Let A be a Borel set. For $\epsilon > 0$, we find N and rectangles A_i and B_j, $1 \leq i, j \leq N$, such that

$$\left| Q(A) - \sum_{i=1}^{N} Q(A_i) \right| < \epsilon \ , \left| P(A) - \sum_{j=1}^{N} P(B_i) \right| < \epsilon \ .$$

Clearly, then,

$$\left| Q(A) - \sum_{i,j=1}^{N} Q(A_i \cap B_j) \right| < \epsilon \ ,$$

and similarly for P. Therefore,

$$\begin{aligned} P\{A\} &\geq \sum_{i,j} P\{A_i \cap B_j\} - \epsilon \\ &\geq \alpha \sum_{i,j} Q\{A_i \cap B_j\} - \epsilon \\ &\geq \alpha(Q(A) - \epsilon) - \epsilon \\ &= \alpha Q(A) - \epsilon(\alpha + 1) \\ &\geq \alpha Q(A) - 2\epsilon \ . \end{aligned}$$

Let $\epsilon \to 0$, and the inequality follows. □

Partial sums of partial quotients.

Here 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. The following Lemma relates bounds for sums of (dependent) partial quotients to bounds for sums of independent partial quotients.

Lemma 11. *Let $X_1, \ldots, X_n$ be the first n partial quotients when θ is Gauss-Kusmin distributed, and let $Y_1, \ldots, Y_n$ be i.i.d. with common distribution that of X_1. Define, for $\epsilon > 0$,*

$$\varphi(m) \stackrel{\text{def}}{=} \sup_{n \geq k \geq m} \mathbf{P}\left\{\left|\frac{\sum_{j=1}^k Y_j}{k \log_2 k} - 1\right| > \epsilon\right\} .$$

Then there exists n_0 depending upon ϵ only such that for $n \geq n_0$,

$$\mathbf{P}\left\{\left|\frac{\sum_{i=1}^n X_i}{n \log_2 n} - 1\right| > 2\epsilon\right\} \leq \frac{4e \log(cn)}{\log(1/\rho)} \varphi\left(\frac{n \log(1/\rho)}{\log(cn)}\right) .$$

Proof. Let N be a positive integer and let $\epsilon > 0$ be arbitrary. Let A_j denote a generic Borel set. Then, if a_j is replaced by X_j to denote the fact that it is a random variable, and if θ has the Gauss-Kusmin law, then by repeated application of Lemma 9, for $k \geq 1$,

$$\mathbf{P}\{\cap_{j=1}^k [X_{Nj} \in A_{Nj}]\} \leq (1 + c\rho^N)^{k-1} \prod_{j=1}^k \mathbf{P}\{X_{Nj} \in A_{Nj}\}$$

Let $Y_1, Y_2, \ldots$ be an i.i.d. sequence with the same distribution as X_1. In particular, then, by Lemma 9,

$$\begin{aligned}
\mathbf{P}\left\{\left|\frac{\sum_{j=1}^k X_{Nj}}{k \log_2 k} - 1\right| > \epsilon\right\} &\leq (1 + c\rho^N)^{k-1} \mathbf{P}\left\{\left|\frac{\sum_{j=1}^k Y_{Nj}}{k \log_2 k} - 1\right| > \epsilon\right\} \\
&= (1 + c\rho^N)^{k-1} \mathbf{P}\left\{\left|\frac{\sum_{j=1}^k Y_j}{k \log_2 k} - 1\right| > \epsilon\right\} \\
&\leq (1 + c\rho^N)^{k-1} \varphi(k) .
\end{aligned}$$

Note that φ is a nonincreasing function. Clearly, assuming that n is a multiple of N to avoid messy expressions,

$$\begin{aligned}
&\mathbf{P}\left\{\left|\frac{\sum_{i=1}^n X_i}{n \log_2 n} - 1\right| > 2\epsilon\right\} \\
&= \mathbf{P}\left\{\left|\frac{\sum_{j=1}^N \sum_{i=0}^{n/N-1} X_{Ni+j}}{n \log_2 n} - 1\right| > 2\epsilon\right\} \\
&\leq \sum_{j=1}^N \mathbf{P}\left\{\left|\frac{\sum_{i=0}^{n/N-1} X_{Ni+j}}{(n/N) \log_2 n} - 1\right| > 2\epsilon\right\}
\end{aligned}$$

$$
\begin{aligned}
&= N\mathbf{P}\left\{\left|\frac{\sum_{i=1}^{n/N} X_{Ni}}{(n/N)\log_2 n} - 1\right| > 2\epsilon\right\} \\
&\le N\mathbf{P}\left\{\left|\frac{\sum_{i=1}^{n/N} X_{Ni}}{(n/N)\log_2 n} - \frac{\sum_{i=1}^{n/N} X_{Ni}}{(n/N)\log_2(n/N)}\right| > \epsilon\right\} \\
&\qquad + N\mathbf{P}\left\{\left|\frac{\sum_{i=1}^{n/N} X_{Ni}}{(n/N)\log_2(n/N)} - 1\right| > \epsilon\right\} \\
&\le N\mathbf{P}\left\{\left|\frac{\sum_{i=1}^{n/N} X_{Ni}}{(n/N)\log_2(n/N)}\right| > \frac{\epsilon \log n}{\log N}\right\} \\
&\qquad + N\left(1 + c\rho^N\right)^{n/N} \varphi\left(\frac{n}{N}\right) \\
&\le 2N\left(1 + c\rho^N\right)^{n/N} \varphi\left(\frac{n}{N}\right) \\
&\quad \text{(as soon as } \epsilon \log n / \log N > 1 + \epsilon) \ .
\end{aligned}
$$

Now, assume that N is chosen such that

$$
\frac{\log(cn)}{\log(1/\rho)} \le N < \frac{2\log(cn)}{\log(1/\rho)} \ .
$$

Then

$$
\begin{aligned}
\mathbf{P}\left\{\left|\frac{\sum_{i=1}^{n} X_i}{n \log_2 n} - 1\right| > 2\epsilon\right\} &\le 2\,\frac{2\log(cn)}{\log(1/\rho)} \times \left(1 + \frac{1}{n}\right)^{n/N} \varphi\left(\frac{n\log(1/\rho)}{\log(cn)}\right) \\
&\le \frac{4e\log(cn)}{\log(1/\rho)} \varphi\left(\frac{n\log(1/\rho)}{\log(cn)}\right) \ .
\end{aligned}
$$

Recall that this bound is valid under the condition $\epsilon \log n / \log N > 1 + \epsilon$. This in turn is valid for all n large enough by our choice of N. □

GENERALIZATION. Note also that in Lemma 11, the X_i's and Y_i's may be replaced by $g(X_i)$ and $g(Y_i)$ for any mapping g. In what follows below, we fix n, and define

$$
g(u) = \begin{cases} 0 & \text{if } u \ge n/\log\log n \\ u & \text{otherwise,} \end{cases}
$$

and apply Lemma 11 to the $g(X_i)$'s. □

PROPOSITION 12. *If θ is Gauss-Kusmin distributed, then*

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

in probability.

PROOF. By Bonferroni's inequality, if g is as in the remark above,

$$\begin{aligned}
&\mathbf{P}\left\{\left|\frac{\sum_{i=1}^n X_i}{n \log_2 n} - 1\right| > 3\epsilon\right\} \\
&\le \mathbf{P}\left\{\left|\frac{\sum_{i=1}^n g(X_i)}{n \log_2 n} - 1\right| > 2\epsilon\right\} \\
&\quad + n\mathbf{P}\{X_1 \ge n \log\log n\} \\
&\quad + \mathbf{P}\left\{\sum_{i=1}^n X_i I_{[n/\log\log n, n \log\log n]}(X_i) > \epsilon n \log_2 n\right\} \\
&= I + II + III \ .
\end{aligned}$$

<u>TERM II.</u> If Z is a Gauss-Kusmin random variable,

$$II \le n\mathbf{P}\left\{1/Z \ge n\sqrt{\log n}\right\} = n \log_2\left(1 + \frac{1}{n \log\log n}\right) \le \frac{1}{\log_2 \log n} \to 0 \ .$$

<u>TERM I.</u> I is bounded as above with a slight change in the definition of φ:

$$\varphi(m) \stackrel{\text{def}}{=} \sup_{n \ge k \ge m} \mathbf{P}\left\{\left|\frac{\sum_{j=1}^k g(Y_j)}{k \log_2 k} - 1\right| > \epsilon\right\} \ .$$

Let us compute the mean μ and variance σ^2 of $g(Y_1)$.

$$\begin{aligned}
\mathbf{E}\left\{\lfloor (1/Z) I_{1/Z < n/\log\log n} \rfloor\right\} &\le \mathbf{E}\left\{(1/Z) I_{1/Z < n/\log\log n}\right\} \\
&= \int_{\log\log n/n}^1 (1/z) dF(z) \\
&\quad \text{(where } F(z) = \log_2(1+z)\text{)} \\
&= \frac{1}{\log 2} \int_{\log\log n/n}^1 \frac{1}{z(1+z)}\, dz \\
&\le \log_2(n/\log\log n) \ .
\end{aligned}$$

Similarly,

$$\begin{aligned}
1 + \mathbf{E}\left\{\lfloor (1/Z) I_{1/Z < n/\log\log n} \rfloor\right\} &\geq \mathbf{E}\left\{(1/Z) I_{1/Z < n/\log\log n}\right\} \\
&\geq \int_{\log\log n/n}^{1} (1/z) dF(z) \\
&= \frac{1}{\log 2} \int_{\log\log n/n}^{1} \frac{1}{z(1+z)}\, dz \\
&= \log_2(n/\log\log n) - \log_2\left(\frac{2}{1+\log\log n/n}\right) \\
&\geq \log_2(n/\log\log n) - 1 \ .
\end{aligned}$$

Therefore,

$$|\mu - \log_2(n/\log\log n)| \leq 2 \ ,$$

and thus, $|\mu - \log_2 n| \leq 2 + \log_2 \log\log n$. Next, to compute an upper bound for the variance, we argue simply as follows:

$$\begin{aligned}
\sigma^2 &\leq \mathbf{E} g^2(Y_1) \\
&\leq \mathbf{E}\left\{(1/Z)^2 I_{1/Z < n/\log\log n}\right\} \\
&= \int_{\log\log n/n}^{1} (1/z^2) dF(z) \\
&= \frac{1}{\log 2} \int_{\log\log n/n}^{1} \frac{1}{z^2(1+z)}\, dz \\
&\leq \frac{n}{\log 2 \log\log n} \ .
\end{aligned}$$

We are finally ready to apply Chebyshev's inequality:

$$\begin{aligned}
&\mathbf{P}\left\{\left|\frac{\sum_{j=1}^k g(Y_j)}{k\log_2 k} - 1\right| > \epsilon\right\} \\
&\quad \leq \mathbf{P}\left\{\left|\frac{\sum_{j=1}^k (g(Y_j) - \mu)}{k\log_2 k}\right| > \frac{\epsilon}{2}\right\} + \mathbf{P}\left\{\left|\frac{\sum_{j=1}^k (\mu - \log_2 k)}{k\log_2 k}\right| > \frac{\epsilon}{2}\right\} \\
&\quad \leq \frac{4\sigma^2}{k\log_2^2 k\, \epsilon^2} + I_{(2+\log_2\log\log n + \log_2(n/k))/\log_2 k > \epsilon/2} \\
&\quad \leq \frac{4n}{k\log_2^2 k\ \log 2\, \epsilon^2\ \log\log n} + I_{(2+\log_2\log\log n + \log_2(n/k))/\log_2 k > \epsilon/2} \ .
\end{aligned}$$

Thus, in Lemma 11, applied to $g(X_i)$'s, we may take

$$\varphi(m) = \begin{cases} 1 & \text{if } 2(2+\log_2\log\log n + \log_2(n/m)) > \epsilon \log_2 m \\ \frac{4n}{m\ \log_2^2 m\ \log 2\, \epsilon^2\ \log\log n} & \text{otherwise.} \end{cases}$$

Therefore, by Lemma 11,

$$
\begin{aligned}
I &\le \frac{4e\log(cn)}{\log(1/\rho)}\varphi\left(\frac{n\log(1/\rho)}{\log(cn)}\right) \\
&= O(\log n) \times \frac{O(n)}{(n/\log n)\log^2(n/\log n)\log\log n} \\
&= O(1/\log\log n)\ ,
\end{aligned}
$$

which tends to zero. Thus, $I \to 0$ as well.

<u>TERM III.</u> Define $B = [n/\log\log n, n\log\log n]$. We bound $\mathbf{P}\{A\}$, where

$$
A \stackrel{\text{def}}{=} \left[\sum_{i=1}^{k} X_i I_{X_i \in B} > \epsilon n \log n\right]\ .
$$

Let N be the number of X_i's in B. Clearly, $A \subseteq [N > \epsilon\log n/\log\log n]$. Note that

$$
\begin{aligned}
p &\stackrel{\text{def}}{=} \mathbf{P}\{X_1 \ge n/\log\log n\} \\
&\le \mathbf{P}\{1/Z \ge n/\log\log n\} \\
&= \log_2(1 + \log\log n/n) \\
&\le \frac{\log_2\log n}{n}\ .
\end{aligned}
$$

By Lemmas 9 and 10, we have

$$
\begin{aligned}
\mathbf{P}\{N \ge u\} &\le \mathbf{P}\{\exists(i_1,\ldots,i_u) \subseteq \{1,\ldots,n\} : X_{i_1} \in B,\ldots,X_{i_u} \in B\} \\
&\le (1+c\rho)^u \mathbf{P}\{\exists(i_1,\ldots,i_u) \subseteq \{1,\ldots,n\} : Y_{i_1} \in B,\ldots,Y_{i_u} \in B\} \\
&\quad \text{(where the } Y_i\text{'s are i.i.d. and distributed as the } X_i\text{'s)} \\
&\le (1+c\rho)^u \binom{n}{u} \mathbf{P}^u\{Y_{i_1} \in B\} \\
&\le \left(\frac{(1+c\rho)enp}{u}\right)^u \\
&\le \left(\frac{(1+c\rho)e\log_2\log n}{u}\right)^u \\
&\to 0
\end{aligned}
$$

if we set $u = \lceil 2(1+c\rho)e\log_2\log n\rceil$. As $u = o(\epsilon\log n/\log\log n)$, we have shown (with room to spare) that

$$
III = \mathbf{P}\{A\} \to 0\ .\ \square
$$

Proposition 12 was proved by analytical methods by Khintchine (1935). The proof given here provides explicit estimates of rates of convergence as well. Proposition 12 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 13. *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.

PROOF. If the Gauss-Kusmin θ is replaced by a uniform $[0,1]$ random variable U, then, as the density f of θ decreases monotonically from $1/\log 2$ to $1/\log 4$ on $[0,1]$, we have

$$\begin{aligned}\mathbf{P}\{U \in A_n\} &= \int_{A_n} du \\ &\le \int_{A_n} \frac{2}{(1+u)} du \\ &= \log 4 \int_{A_n} \frac{1}{(1+u)\log 2} du \\ &= \log 4\, \mathbf{P}\{\theta \in A_n\} \\ &\to 0 \ .\end{aligned}$$

Thus, Proposition 12 remains true for the uniform distribution and for any distribution with a density on $[0,1]$. □

The behavior of the denominator of the convergents.

LEMMA 14 (KHINTCHINE, 1935 AND LÉVY, 1936; 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 14 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

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

so that

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

We also note that q_n must grow faster than a Fibonacci sequence, as $q_{n+1} \ge q_n + q_{n-1}$. This implies that $q_n \ge \rho^{n-1}$ for all n, where $\rho = (1+\sqrt{5})/2$ is the golden ratio. Another simple lower bound is $q_n \ge 2^{(n-1)/2}$ (Khintchine, 1963, p. 18). Finally, we note that a normal limit law for $(\log q_k - \gamma k)/\sqrt{k}$ was obtained by Philipp (1969).

THEOREM 15. *If θ has any density on $[0,1]$, then*

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

in probability. Note that $12/\pi^2 \approx 1.215854203$ and $12\ln 2/\pi^2 \approx 0.8427659130$.

Proof. By Theorem 13, 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 13 and Proposition 5 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. □

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. Since with probability one, $q_k^{1/k} \to e^{\gamma}$ as $k \to \infty$, we see from Lemma 5 that with probability one,

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

infinitely often. Thus, Theorem 15 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 Theorem 5 and Lemma 14, with probability one, for all but finitely many n,

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

Very good trees. From the inequality of Theorem 2, 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 hight enough compared to typical random Weyl trees, and also with respect to true random binary search trees.

If $a_n \equiv a$ for all n, then $q_n = aq_{n-1} + q_{n-2}$ for all n. From this, $q_n \sim c \left(\frac{a+\sqrt{a^2+4}}{2}\right)^n$ 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.

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

THEOREM 16. *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, for $k \geq 2$, $H_{q_k} \geq \sum_{i=1}^k a_i - 1 \geq a_k$, 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). □

A few examples suffice to drive our point home. 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 $K \in [1/2, 3/2]$. 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.

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,

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

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

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 $q_{2k+1} = 4k q_{2k} + q_{2k-1}$ and $q_{2k} = q_{2k-1} + q_{2k-2}$, one cam 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.

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 b. 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 (1977, 1978, 1991) or L'Ecuyer (1989, 1990).

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

$$\frac{nh_n(nh_n + 1)}{2}$$

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 17. *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 exceeds*

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

infinitely often.

PROOF. Consider only $n = q_k$ for some k. Note that the sum of the depths of the nodes in the Weyl tree is at least q_{k-1} (the number of leaves) times $(a_k + 1)a_k/2$ (as each leaf is the end of a path of a_k all-left or all-right edges and these paths are thus disjoint). But $a_k q_{k-1} = q_k - q_{k-2} \geq q_k/2 = n/2$, because $2q_{k-2} \leq q_{k-2} + q_{k-1} \leq q_k$. Therefore, the number of comparisons in quicksort is at least

$$\frac{n(a_k + 1)}{4} \; .$$

But by the Borel-Bernstein Theorem,

$$a_k \geq 4C\gamma k \log k \log\log k$$

infinitely often almost surely, while by Lemma 14, $k \sim (1/\gamma) \log n$ almost surely. Combining all this gives the result. □

The number of leaves.

For a random binary search tree, the expected number of leaves is asymptotic to $n/3$ (see Mahmoud, 1992). However, for Weyl trees, the behavior of the number of leaves is much more erratic. We refer to Lemma 3, and note that at time $q_k - 1$, the number of leaves is exactly q_{k-1}:

$$|\mathcal{L}_{q_k - 1}| = q_{k-1} \ .$$

Thus, at that instant in the tree construction (the last node to complete a layer), the proportion of leaves is

$$\frac{q_{k-1}}{q_k - 1} \sim \frac{q_{k-1}}{q_k} \ .$$

Just to show how this interesting relationship explains the erratic behavior of typical Weyl trees, consider the recurrence $q_k = a_k q_{k-1} + q_{k-2}$, and observe that

$$\frac{q_{k-1}}{q_k} \leq \frac{1}{a_k} \ .$$

The behavior of a_k was discussed in an earlier section. It suffices to note that $a_k > k \log k$ infinitely often with probability one, so that, with probability one, the proportion of leaves is infinitely often less than $1/\log n$, for example.

The fill-up level.

The fill-up level F_n of a search tree is the maximal number of full levels. For a random binary search tree, this is known to be asymptotic to $0.3711\ldots \log n$ in probability (Devroye, 1986). Again, random Weyl trees deviate from this substantially. While we will not study F_n in detail, we would like to note one inequality:

$$\prod_{i=1}^{F_n} a_i \leq q_{F_n} \leq n \ .$$

Indeed, to get a path in the tree of polarity $+-+-+-\ldots$ of length k, by the way layers are painted on, we must have $n \geq q_k$. But $q_k \geq \prod_{i=1}^{k} a_i$, which proves the inequality.

Example 1. By Lemma 14, we have without further work

$$F_n \leq (1/\gamma + o(1)) \log n$$

in probability when θ is uniform $[0,1]$. In fact, then, we have for all $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbf{P} \left\{ \frac{H_n}{F_n} \geq (1-\epsilon) \log_2 \log n \right\} = 0 \ .$$

EXAMPLE 2. If $a_k \equiv k$, then $F_n! \leq n$, so that

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

This result applies also when $\theta = \tan(1/2)$, and $\theta = e$, two examples cited earlier.

EXAMPLE 3. When $a_k = 2^k$, simple calculations show that

$$F_n = O(\sqrt{\log n}) .$$

In fact, for any slowly increasing sequence b_n, it is possible to find a θ such that $F_n \leq b_n$ for all n large enough.

Other characteristics.

Let the left height of a tree be the maximal number of left edges seen on any path from a node to the root. Let the right height be defined similarly. Clearly, the left height is one less than the number of layers of left polarity, and this grows as $\sum_{i=1}^{k/2} a_{2i}$ where k is the solution of $n = q_k$. Using arguments as in Theorem 15, it is easy to prove that if H_n^L and H_n^R are the left and right heights of $\mathcal{T}_n$, then

$$\frac{H_n^L}{\log n \log\log n} \to \frac{6}{\pi^2}$$

and

$$\frac{H_n^R}{\log n \log\log n} \to \frac{6}{\pi^2}$$

in probability.

The distance from the root to the minimum is equal to H_n^L, and is thus also covered by the result above. In random binary search trees, these quantities are $\Theta(\log n)$ in probability: the left height grows as $e \log n$, while the distance from the minimum to the root grows as $\log n$ in probability (Devroye, 1987).

References

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

J. Beck, *Randomness of* $n\sqrt{2} \pmod 1$ *and a Ramsey property of the hyperbola*, vol. 60, Budapest, 1991.

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

P. Bohl, “Über ein in der Theorie der säkularen Störunger vorkommendes Problem,” *Jl. Reine und Angewandte Mathematik*, vol. 135, pp. 189–283, 1909.

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

R. Béjian, “Minération de la discrépance d'une suite quelconque sur T,” *Acta Arithmetica*, vol. 41, pp. 185–202, 1982.

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

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

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

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

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

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

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

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

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

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

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

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

A. Khintchine, “Metrische Kettenbuchprobleme,” *Compositio Mathematica*, vol. 1, pp. 361–382, 1935.

A. Khintchine, “Metrische Kettenbuchprobleme,” *Compositio Mathematica*, vol. 2, pp. 276–285, 1936.

A. Ya. Khintchine, *Continued Fractions*, P. Noordhoff, Groningen, The Netherlands, 1963.

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

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

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

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

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

P. L’Ecuyer, “Random numbers for simulation,” *Communications of the ACM* , vol. 33, pp. 85—97, 1990.

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

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

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

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

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

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

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

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

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

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

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

W. Philipp, “Some metrical theorems in number theory II,” *Duke Mathematical Journal*, vol. 38, pp. 477–458, 1970.

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

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

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

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

V. T. Sós, *Studies in Pure Mathematics*, pp. 685–700, Birkhauser, Basel, 1983.