McGill University

School of Computer Science

DATA STRUCTURES

BINARY SEARCH TREES

A STUDY OF RANDOM WEYL TREES

BY

AMAR GOUDJIL

A Thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

0-612-50778-5

October 25, 1998

Abstract

DATA STRUCTURES

BINARY SEARCH TREES

A STUDY OF RANDOM WEYL TREES

BY

AMAR GOUDJIL

This thesis covers the study of a particular class of binary search trees, the Weyl trees formed by consecutive insertion of numbers $\{\theta\}$, $\{2\theta\}$, $\{3\theta\}$, $\ldots$, where θ is an irrational number from $(0,1)$, and $\{x\}$ denotes the fractional part of x. Various properties of the structure of these trees are explored and a relationship with the continued fraction expansion of θ is shown. Among these properties, an approximation of the height H_n of a Weyl tree with n nodes is given when θ is chosen at random and uniformly on $(0, 1)$. The main result of this work is that in probability, $H_n \sim (12/\pi^2) \log n \log \log n$.

STRUCTURES DE DONNEES

ARBRES DE RECHERCHE BINAIRE

UNE ETUDE DES ARBRES ALEATOIRES DE WEYL

PAR

AMAR GOUDJIL

Résumé

Cette thèse est une contribution à l'important travail de recherche sur les structures de données du Prof. Luc Devroye. Elle couvre une classe particulière d'arbres de recherche binaire : Les arbres de Weyl construits à partir d'insertions consécutives des éléments de la suite $\{\theta\}, \{2\theta\}, \{3\theta\}, \ldots$, où θ est un nombre irrationnel de l'intervalle $[0, 1]$, et où $\{x\}$ désigne la partie fractionnaire de x. Différentes propriétés de la structure de ces arbres sont explorées et une relation avec l'expansion en fractions continues de θ est exhibée. Parmi ces propriétés, une approximation de la hauteur H_n de l'arbre de Weyl à n noeuds est donnée lorsque θ est choisi de facon aléatoire selon la loi uniforme de $[0, 1]$. Le résultat principal de ce travail est qu'en probabilité, $H_n \sim (12/\pi^2) \log n \log \log n$.

Acknowledgements

My gratitude goes first to my advisor, Prof. LUC DEVROYE, who suggested the topic of my thesis, provided me with helpful suggestions and guided me throughout the research work. I would like also to thank him also for his financial support that helped me so much as a newcomer with my family to Canada.

I express my gratitude to Prof. HYDER, one of the founders of McGill School of Computer Science, for his contribution and his financial support during more than two years.

I am grateful to Prof. RATZER and Prof. MAQUELIN for their technical help.

Finally, I would like to thank all the administrative staff and the technical team of the school of computer science for their invaluable help.

Contents

1 Introduction.

This thesis is based essentially on a work published[1] with my advisor, Prof. Luc Devroye. It is a study of binary search trees formed by consecutive insertions of numbers $x_1 = \{\theta\}, x_2 = \{2\theta\}, x_3 = \{3\theta\}, \ldots$, where $\theta \in (0, 1)$ is an irrational number, and $\{x\}$ denotes "mod 1" which is the fractional part of the number x. The sequence in question is called the Weyl sequence for θ, after Weyl, who showed that for all irrational θ the sequence is equidistributed e.g.: for all $0 \le a \le b \le 1$,

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

which means that the average number of x_i that fall into $[a, b]$ is equal to the measure of the interval when n goes to infinity. (see [12], [14] or [18]).

The equidistribution 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 ([17a] and Sós ([31]) have interesting accounts of this. Let $T_n(\theta)$ 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 quantities:

[1] Random Structures and Algorithms, Vol 12, Issue 3 1998–John Wiley & Sons, Inc.

- The average depth of a node (the depth is the path distance from a node to the root).
- The height (the maximal depth).
- The number of leaves (the number of nodes with no children).

In this thesis we will focus on these quantities. The following notation borrowed from Prof. Luc Devroye's course notes [8b] 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 thesis we consider two cases of Weyl trees:

- Weyl trees for fixed θ.
- Weyl trees when θ is a random variable.

In the first case important connections with the continued fractions are settled and then some algebraic theory of numbers properties are used to deduce easily results about H_n and $|\mathcal{L}_n|$. In the latter case we are in presence of random Weyl trees. In fact we consider that $\theta = U$, 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, in this thesis, 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 $X_1, \ldots, X_n$. The height H'_n of $\mathcal{R}_n$ satisfies

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

(Robson [28, 28a], Devroye [8, 8a], Mahmoud [24]).

2 Random binary search trees

Let us consider an iid sequence of random variables $X_1, \ldots, X_n$ defined on a probability space $(\Omega, \mathcal{A}, P) \mapsto R$, and let's suppose that there are no ties which means that

$$\mathbf{P}\{\omega \in \Omega \textit{ such that } X_i(\omega) = X_j(\omega) \textit{ for some } i \neq j\} = 0$$

It is sufficient to have a density for the law of X_i to avoid ties since that kind of measures doesn't charge points. Now, briefly we explain how the tree is built. For any $\omega \in \Omega$ let's put $X_1(\omega)$ on the root of the tree. For $i > 1$ all $X_i(\omega)$ that are lower(resp. greater) than $X_1(\omega)$ are nodes of the left subtree(resp. right subtree). With the index representing the time of insertion we repeat this operation recursively for each subtree until all elements of the sequence have been processed. 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 x proceeds by finding the unique external node that could accept x, given the binary search tree property, and placing x there. A tree 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 $\{1, \ldots, n\}$ is called a random binary search tree, and will be denoted by $\mathcal{R}_n$. Mostly everything is known about the behaviour of $\mathcal{R}_n$ (see [24]). 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, $\frac{(D_n - 2\log n)}{\sqrt{2\log n}} \xrightarrow{\mathcal{L}}$ normal$(0,1)$, where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution(Devroye 1988).

Before starting study of Weyl trees, we need to define what is a record.

Given and iid sequence $X_1, \ldots, X_n$ we say that X_i, for $i \geq 1$ is an up-record (resp. a down-record) if $X_i = \max\{X_1, \ldots, X_i\}$ (resp. $X_i = \min\{X_1, \ldots, X_i\}$) and that X_i is a record if it is either an up-record or a down-record. If N is the number of up-records, we know from [8b] that $E[N] = H_n$ where H_n stands for the n-th harmonic e.g. $\sum_{i=1}^{n} \frac{1}{i}$ and that

$$\log(n+1) < E[N] < 1 + \log(n).$$

Example

i	X_i	Up Record	Down Record
1	8.5	Yes	Yes
2	13.2	Yes	—
3	5.1	—	Yes
4	9.0	—	—
5	11.1	—	—
6	10.9	—	—
7	17.2	Yes	—
8	6.5	—	—
9	5.0	—	Yes
10	18.3	Yes	—

At time $i = 1$, X_1 is both an up and down record. At time $i = 2$, $X_2 = 13.2 > X_1 = 8.5$. So, we have an up-record at that time.

The records for this very small sample happen at times 1,2,3,7,9,10. The binary search tree for this sample is obtained by first putting $X_1 = 8.5$ at the root. The second value $X_2 = 13.2$ is greater than X_1 so it is a right children. The third value $X_3 = 5.1$ is lower than X_1 so it is a left children. The fourth value $X_4 = 9.0$ is greater than X_1, so it is a node of the right subtree of the tree rooted a X_1. It is also inserted immediately after $X_3 = 13.2$ which is greater so X_4 is a left children of X_3. The process is continued until all the nodes have been introduced. The result is shown by fig. 1.

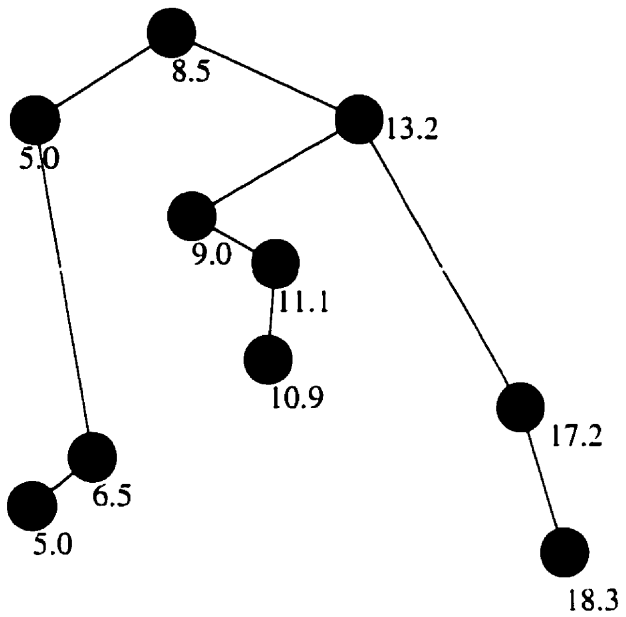

figure 1: Tree obtained by consecutive insertion of data from the sample above.

We also need some inequalities to understand the technical part where we will use Devroye's demonstration to show that $\frac{\sum_{i=1}^{n} a_i}{n \log_2 n} \to 1$ in probability when θ is Gauss-Kusmin distributed, a result due originally to Khintchine.

- Bonferroni's inequalities: Let $\mathbf{A} = \{A_i,\ 1 \le i \le n\}$ be a set of events in some probability space $(\Omega, \mathcal{A}, P)$, and define S_k as

$$\begin{aligned} S_0 &= \mathbf{P}(\Omega) \\ S_k &= \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} \mathbf{P}[A_{i_1} A_{i_2} \ldots A_{i_k}] \end{aligned}$$

then the sum $\sum_{k=1}^{n}(-1)^{k-1}S_k$ satisfies the alternating inequalities

$$(-1)^k \quad [\mathbf{P}(\cup_{i=1}^{n} A_i) + \sum_{j=1}^{k}(-1)^j S_j] \geq 0.\ 1 \leq k \leq n$$

$$(-1)^{k+1} \quad [\mathbf{P}(\cap_{i=1}^{n} \overline{A_i}) + \sum_{j=0}^{k}(-1)^{j+1} S_j] \geq 0.\ 1 \leq k \leq n.$$

- Chebychev's inequality:
  $\mathbf{P}[|X| \geq t] \leq t^{-2}\mathbf{E}\{X^2\}$.

3 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:

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

As it turns out, there is a lot of structure in these sequences. For instance the next record in a Weyl sequence is the element with indice equal to the sum of previous up and down records indices. More precisely we have the following fundamental property:

Lemma 3.1 (Ellis and Steele, 1981 [10]) . *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

$$\{i * L_k \pmod{L_k + R_k} \text{ for } i \geq 1\} \cap \{1, \ldots, n\} .$$

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). Indeed, let's suppose that one of these trees is a Weyl tree. We have $L_2 + R_2 = 3$ which implies that time T_3 is time for an extrema but for these two trees the nodes are ordered either as $X_1 < X_3 < X_2$ or $X_2 < X_3 < X_1$ which means that X_3 is not an extrema.

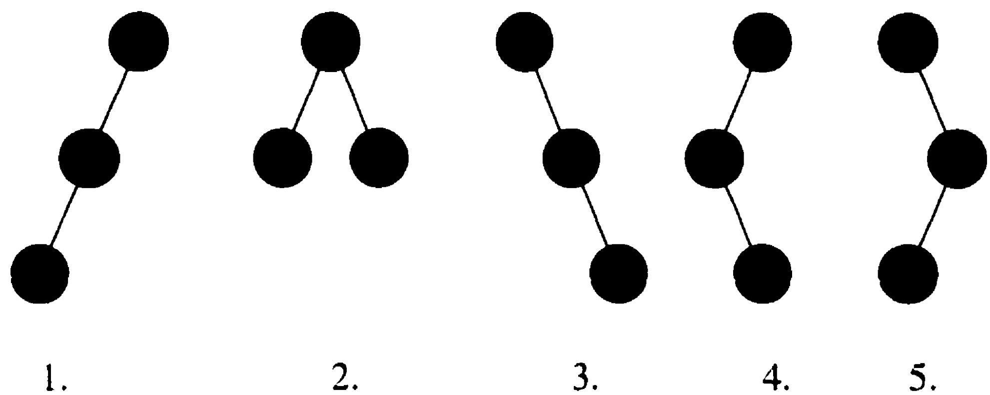

figure 2: Trees 4 and 5 are impossible to obtain as Weyl trees.

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 new result explains.

Lemma 3.2 *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 3.1 on page 14, 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.

Example: Let $\theta = \sqrt{2} \approx 1.41421\ldots$. T_i for $i \geq 1$ are times of records and L_i, for $i \geq 1$ (resp. R_i, for $i \geq 1$) are times of minimas (resp. maximas).

T_i	n	$n\theta$	Extremas
T_1	1	0.414214	$L_1 = R_1 = 1$
T_2	2	0.828427	$R_2 = L_1 + R_1 = 2,\ L_2 = L_1 = 1$
T_3	3	0.242641	$L_3 = R_2 + L_2 = 3,\ R_3 = R_2 = 2$
	4	0.656854	
T_4	5	0.071068	$L_4 = R_3 + L_3 = 5,\ R_4 = R_3 = 2$
	6	0.485281	
T_5	7	0.899495	$R_5 = R_4 + L_4 = 7,\ L_5 = L_4 = 5$
	8	0.313708	
	9	0.727922	
	10	0.142136	
	11	0.556349	
T_6	12	0.970563	$R_6 = R_5 + L_5 = 12,\ L_6 = L_5 = 5$
	13	0.384776	
	14	0.798989	
	15	0.213203	
	16	0.627417	
T_7	17	0.041631	$L_7 = R_6 + L_6 = 17,\ R_7 = R_6 = 12$
	18	0.455844	

The next extremum is the element with index $L_7 + R_7 = 17 + 12 = 29$. Furthermore, sorting the sequence $x_1, \ldots, x_n$ needs only the index of the minimum which is 5 for $T_6 \leq n < T_7$. The indices of the ordered sequence are computed according to lemma 3.1:

$5*1 = 5$	(mod 17) = 5	$5*9 = 45$	(mod 17)	= 11
$5*2 = 10$	(mod 17) = 10	$5*10 = 50$	(mod 17)	= 16
$5*3 = 15$	(mod 17) = 15	$5*11 = 55$	(mod 17)	= 4
$5*4 = 20$	(mod 17) = 3	$5*12 = 60$	(mod 17)	= 9
$5*5 = 25$	(mod 17) = 8	$5*13 = 65$	(mod 17)	= 14
$5*6 = 30$	(mod 17) = 13	$5*14 = 70$	(mod 17)	= 2
$5*7 = 35$	(mod 17) = 1	$5*15 = 75$	(mod 17)	= 7
$5*8 = 40$	(mod 17) = 6	$5*16 = 80$	(mod 17)	= 12

The ordered sequence is $x_5 < x_{10} < x_{15} < \ldots < x_2 < x_7 < x_{12}$. Finally if $T_i = T_6$ and $j \in \{T_6, T_6+1, \ldots, T_7-1\}$ we can compute indices of immediate predecessor and successor of x_j, for instance, if $j = 14$ the index of the immediate predecessor is $j - L_6 = 14 - 5 = 9$ and the index of the immediate successor is $j + L_6 - (R_6 + L_6) = 14 + 5 - 17 = 2$ and we can effectively check these results from the list above. The tree obtained is shown by fig.3 below:

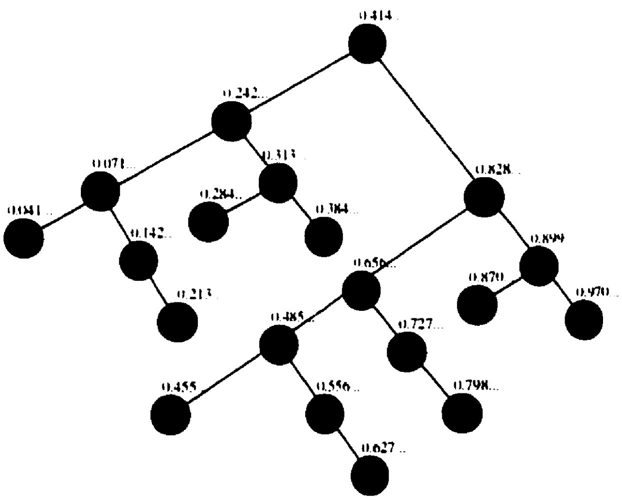

figure 3: Weyl tree from a seed $\theta = \sqrt{2}$.

Lemma 3.3 *We have*

$$|\mathcal{L}_{T_{i+1}-1}| = \min(L_i,\ R_i)\ .$$

and

$$H_{T_{i+1}-1} = i\ .$$

Put differently,

$$k - 1 \leq H_n \leq k$$

if k is the unique integer with $T_k \leq n < T_{k+1}$.

Proof.
The first statement is an immediate corollary of the Lemma 3.2. 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 thesis, we recall a few basic facts from the theory of continued fractions.

4 Continued fractions.

To define a continued fraction we consider the irrationnal number $\pi = 3.14159\ldots$ The first step is to write $\pi = 3 + 0.14159\ldots$. Next we consider the fractional part $0.14159\ldots$ and rewrite it as $1/x$ for some irrational value x:

$\pi = 3 + \frac{1}{7.0625...}$. The denominator may be rewritten as $7 + 0.0625\ldots$ and we can repeat this process n times for any $n \geq 1$. The result is:

$$\pi = 3 + \frac{1}{7.0625\ldots} = 3 + \frac{1}{7 + \frac{1}{15.996...}} = 3 + \frac{1}{7 + \frac{1}{15 + \frac{1}{1.0034...}}}$$

So, the representation of π as a continued fraction will be noted $[3;7,15,1,292,...]$.

Let, now 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 <u>continued fraction</u> expansion of θ by

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

where the a_i's are the <u>partial quotients</u>, $a_i \geq 1$ for $i \geq 1$(see Lang [21] or LeVeque [22]). Thus, we have

$$\theta = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \cdots}} \ .$$

with

$$a_0 = \lfloor \theta \rfloor .$$

The <u>i-th convergent</u> 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 evolution of the layers, the following result is essential. It extends a theorem of Lang [21].

Theorem 4.1 (Boyd and Steele [5].) *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*

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

$q_1 + q_2,\ q_1 + 2q_2,\ \ldots,\ q_1 + a_3 q_2 = q_3;$

$q_3 + q_4,\ q_3 + 2q_4,\ \ldots,\ q_3 + a_5 q_4 = q_5;$

...

and the (left extrema) occurs when n is in the list $q_0 + q_1,\ q_0 + 2q_1,\ \ldots,$ $q_0 + a_2 q_1 = q_2;$

$q_2 + q_3,\ q_2 + 2q_3,\ \ldots,\ q_2 + a_4 q_3 = q_4;$

$q_4 + q_5,\ q_4 + 2q_5,\ \ldots,\ q_4 + a_6 q_5 = q_6;$

... .

Lemma 4.1 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 3.1 and Lemma 3.2 should suffice to completely reconstruct the shape of the tree (see figure 4).

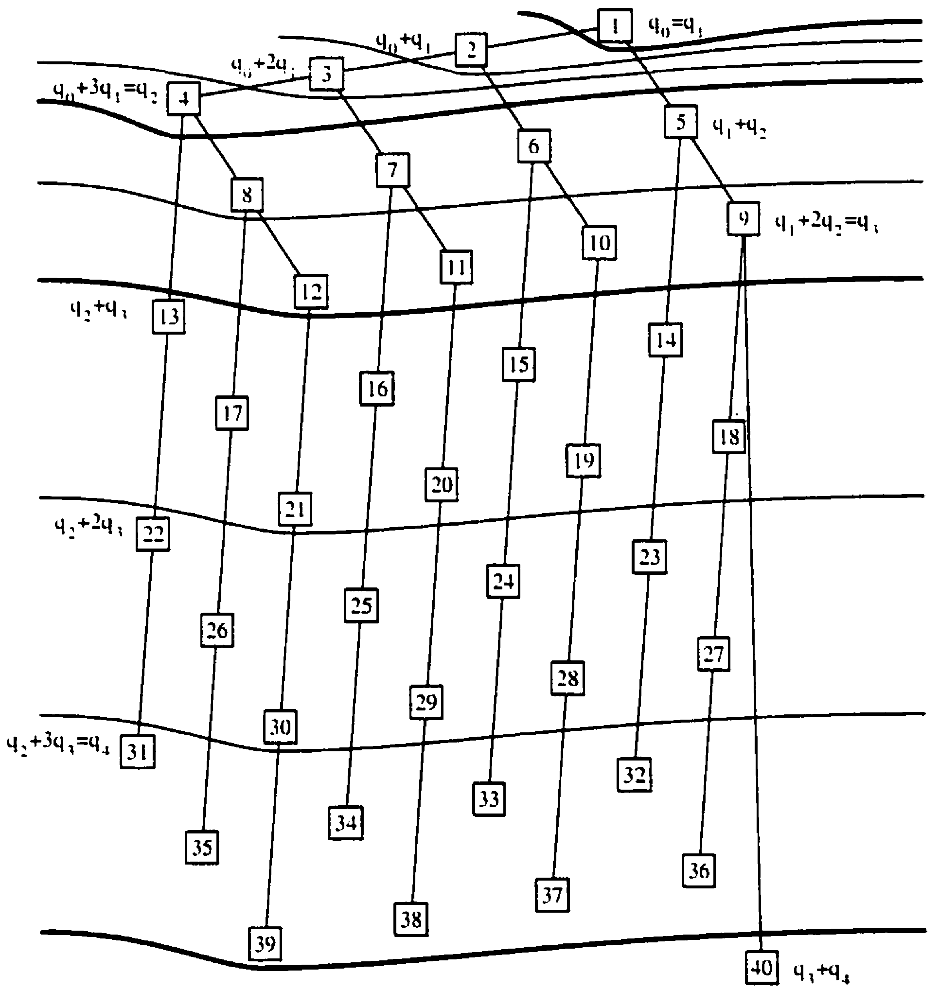

figure 4: This figure shows the Weyl tree for $\theta = \sqrt{77} = [8; 1, 3, 2, 3, \ldots]$.

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 3.1, can the reader guess who the parent is of point 41? The last node is a maxima, hence $R = 40$. The next layer is a set of left nodes, to find the index i of the parent of node $j = 41$ we use relation $j - L + k(R + L)$ with $L = 31$, $R = 40$, $k = 0$:

$$i = 41 - 31 + 0 \times (31 + 40) = 10.$$

5 Height of random Weyl trees.

From the lemma 3.3 and lemma 4.1, we easily determine the relationship between height and partial quotients.

Proposition 5.1 *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

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

6 Discrepancy.

There is another field in which the behaviour 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 [29]; Béjian [6])

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

infinitely often. From Niederreiter ([26c], 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.

7 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 [16b], Philipp [27a], or the references found there.

Theorem 7.1 (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 [16b]).

Lemma 7.1 (Kusmin [19]; Lévy [23]) . *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 $[0,1]$, 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 7.1 also gives an accurate description of the limit law for a_n. In fact, as a corollary, one obtains another result of Lévy ([23]), 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 ([7]) showed that any law with independent a_n's corresponds to a random θ with a singular distribution.)

Lemma 7.2 (Galambos [13].) *Let θ have the Gauss-Kusmin law. Then*

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

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 7.2 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 7.3 (Philipp [27a].) *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 7.4 *If P and Q are two probability measures and $\alpha > 0$ is a number such that for all rectangular Borel sets (products of intervals), $P \geq \alpha Q$, then $P \geq \alpha Q$ 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.

8 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 8.1 *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 $\epsilon > 0$,*

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

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 7.3, 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 7.3,

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

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

$$\begin{aligned}
\mathbf{P}\left\{\left|\frac{\sum_{i=1}^{n} X_i}{n\log_2 n}-1\right|>2\epsilon\right\} &= \mathbf{P}\left\{\left|\frac{\sum_{j=1}^{N}\sum_{i=0}^{n/N-1} X_{Ni+j}}{n\log_2 n}-1\right|>2\epsilon\right\}\\
&\le \sum_{j=1}^{N}\mathbf{P}\left\{\left|\frac{\sum_{i=0}^{n/N-1} X_{Ni+j}}{(n/N)\log_2 n}-1\right|>2\epsilon\right\}\\
&= N\mathbf{P}\left\{\left|\frac{\sum_{i=1}^{n/N} X_{Ni}}{(n/N)\log_2 n}-1\right|>2\epsilon\right\}\\
&\le N\mathbf{P}\left\{\left|\frac{\sum_{i=1}^{n/N} X_{Ni}}{(n/N)\log_2 n}-\frac{\sum_{i=1}^{n/N} X_{Ni}}{(n/N)\log_2(n/N)}\right|>\epsilon\right\}\\
&\quad +N\mathbf{P}\left\{\left|\frac{\sum_{i=1}^{n/N} X_{Ni}}{(n/N)\log_2(n/N)}-1\right|>\epsilon\right\}\\
&\le N\mathbf{P}\left\{\left|\frac{\sum_{i=1}^{n/N} X_{Ni}}{(n/N)\log_2(n/N)}\right|>\frac{\epsilon\log n}{\log N}\right\}\\
&\quad +N\left(1+c\rho^N\right)^{n/N}\varphi\left(\frac{n}{N}\right)\\
&\le 2N\left(1+c\rho^N\right)^{n/N}\varphi\left(\frac{n}{N}\right)\\
&\quad \text{(as soon as } \epsilon\log n/\log N > 1+\epsilon\text{)}\ .
\end{aligned}$$

Now, assume that N is chosen such that

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

Then

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

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.

8.1 Generalization.

Note also that in Lemma 8.1, 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

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

and apply Lemma 8.1 to the $g(X_i)$'s.

Proposition 8.1 *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,

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

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

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

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

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

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

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

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:

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

Thus, in Lemma 8.1, 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 8.1,

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

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

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

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

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

By Lemmas 7.3 and 7.4, we have

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

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

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

Proposition 8.1 was proved by analytical methods by Khintchine [16]. The Proof given here provides explicit estimates of rates of convergence as well. Proposition 8.1 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 8.1 *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.

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

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

Thus, Proposition 8.1 remains true for the uniform distribution and for any distribution with a density on $[0,1]$.

9 The behavior of the denominator of the convergents.

Lemma 9.1 (Khintchine [16] and Lévy [23a]; see Khintchine [16b], 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 9.1 is related to the property (Khintchine [16b], p. 101) that

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

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

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

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 [16b], p. 18). Finally, we note that a normal limit law for $(\log q_k - \gamma k)/\sqrt{k}$ was obtained by Philipp [27].

Theorem 9.1 *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 8.1, 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 $q_k \le n < q_{k+1}$. But Theorem 8.1 and Proposition 5.1 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.

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,

$$a_n \ge n \log n \log\log n$$

infinitely often. Since with probability one, $q_k^{1/k} \to e^{\gamma}$ as $k \to \infty$, we see from Lemma 5.1 that with probability one,

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

infinitely often. Thus, Theorem 9.1 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,

$$a_n \le n \log n \log^{1+\epsilon} \log n$$

for $\epsilon > 0$. This implies that with probability one, for all but finitely many n,

$$\sum_{j=1}^{n} a_j \le n^2 \log n \log^{1+\epsilon} \log n \ .$$

But then, by Theorem 5.1 and lemma 9.1, with probability one, for all but finitely many n,

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

9.1 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 $n \ge 0$). Indeed, as for these sequences, $q_n \le \prod_{i=1}^{n}(1 + a_i) \le \exp(\sum_{i=1}^{n} a_i) = \exp(O(n))$, 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, 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.

9.2 Very bad trees.

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

Theorem 9.2 *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 $(n/\log_2 n)^{1-1/b}$.

9.3 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 cam show (see Boyd and Steele [5], 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.

10 Sorting Weyl sequences

Ellis and Steele [10] 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 equidistributed 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 [25], Knuth [17], Bennett [3]). For surveys and discussions on the topic of uniform random variate generation, one could consult Niederreiter [26, 26a, 26b] or L'Ecuyer [20, 20a]).

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 [30]). 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 $n \log n \quad \log \log n$. In fact, we have the following.

Proposition 10.1 *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 9.1, $k \sim (1/\gamma) \log n$ almost surely. Combining all this gives the result.

11 The number of leaves.

For a random binary search tree, the expected number of leaves is asymptotic to $n/3$ (see Mahmoud [24]). However, for Weyl trees, the behavior of the number of leaves is much more erratic. We refer to Lemma 3.3. 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.

12 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 [8]). 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.

13 Examples

13.1 Example 1

By lemma 9.1, 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 $\epsilon > 0$,

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

13.2 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.

13.3 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 $F_n \le b_n$ for all n large enough.

14 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 9.1, 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 [8a]).

15 Conclusion

Working on this thesis led me to investigate data structures, algebraic theory of numbers, discrepancy and technical computations involving bounding of expressions in probability theory.

In addition to the new results, particular attention is given to link to continued fractions of irrational numbers, which is not obvious at first sight. This has suggested to go further and design Weyl trees starting now from given values for partial quotients. For instance, we can fix for each step the number of layers of right and left polarity. Also, we realized that there are many structures in the continued fractions expansion of numbers and that the field is wide open to research.

The investigations in this thesis are important and worthwhile as the new results are based on ideas of Ellis and Steele, Boyd and Steele, Levy, and Galambos. These results were not easy to find and to understand.

Finally, this thesis was of prime interest for me since it strengthened my theoretical background on data structures. In addition, all the material comes in handy for further research.

References

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

[2] J. Beck. Randomness of $n\sqrt{2}$ (mod 1) and a Ramsey property of the hyperbola. Vol. 60.Budapest, 1991.

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

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

[5] 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.

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

[7] S. D. Chatterji. Masse, die von regelmässigen Kettenbrüchen induziert sind. Mathematische. Annalen.vol. 164, pp. 113–117, 1966.

[8] L. Devroye.A note on the height of binary search trees. Journal of the ACM, vol. 33, pp. 489–498, 1986.

[8a] L. Devroye,Branching processes in the analysis of the heights of trees, Acta Informatica, vol. 24, pp. 277–298, 1987.

[8b] L. Devroye, Course notes on Probabilistic Analysis of Algorithms, McGill University 1997.

[9] P. Diaconis, The distribution of leading digits and uniform distribution mod 1, Annals of Probability, vol. 5, pp. 72–81, 1977.

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

[11] J. N. Franklin, Deterministic simulation of random sequences, Mathematics of Computation, vol. 17, pp. 28–59, 1963.

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

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

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

[15] H. Kesten, Uniform distribution mod 1, Annals of Mathematics, vol. 71, pp. 445–471, 1960.

[16] A. Khintchine, Metrische Kettenbuchprobleme, Compositio Mathematica, vol. 1, pp. 361–382, 1935.

[16a] A. Khintchine, Metrische Kettenbuchprobleme, Compositio Mathematica, vol. 2, pp. 276–285, 1936.

[16b] A. Khintchine, Continued Fractions, P. Noordhoff,Groningen, The Netherlands, 1963.

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

[17a] The Art of Computer Programming, Vol. 2, 2nd Ed., Addison-Wesley,Reading, Mass., 1981.

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

[19] R. O. Kusmin, On a problem of Gauss, Reports of the Academy of Sciences, vol. A, pp. 375–380, 1928.

[20] 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.

[20a] Random numbers for simulation, Communications of the ACM, vol. 33, pp. 85—97, 1990.

[21] S. Lang, Introduction to Diophantine Approximations, Addison-Wesley, Reading, MA, 1966.

[22] W. J. LeVeque, Fundamentals of Number Theory, Addison-Wesley, Reading, MA, 1977.

[23] 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.

[23a] P. Lévy, Théorie de l'addition des variables aléatoires, Paris, 1937.

[24] H. M. Mahmoud, Evolution of Random Search Trees, John Wiley, New York, 1992.

[25] P. Martin-Löf, The definition of random sequences, Information and Control, vol. 9, pp. 602-619, 1966.

[26] H. Niederreiter, Pseudo-random numbers and optimal coefficients, Advances in Mathematics, vol. 26, pp. 99-181, 1977.

[26a] H. Niederreiter, Quasi-Monte Carlo methods and pseudo-random numbers, Bulletin of the American Mathematical Society, vol. 84, pp. 957-1042, 1978.

[26b] H. Niederreiter, Recent trends in random number and random vector generation, Annals of Operations Research, vol. 31, pp. 323-346, 1991.

[26c] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, vol. 63, SIAM CBMS-NFS Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1992.

[27] W. Philipp, The central limit problem for mixing sequences of random variables, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, vol. 12, pp. 155-171, 1969.

[27a] W. Philipp, Some metrical theorems in number theory II, Duke Mathematical Journal, vol. 38, pp. 477–458, 1970.

[28] J. M. Robson, The height of binary search trees, The Australian Computer Journal, vol. 11, pp. 151–153, 1979.

[28a] J. M. Robson, The asymptotic behaviour of the height of binary search trees, Australian Computer Science Communications, p. 88, 1982.

[29] W. M. Schmidt, Irregularities of distribution, Acta Arithmetica, vol. 21, pp. 45–50, 1972.

[30] R. Sedgewick, The analysis of quicksort programs, Acta Informatica, vol. 4, pp. 327–355, 1977.

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