Pseudorandom binary functions on almost uniform trees

Katalin Gyarmati, Pascal Hubert, András Sárközy

Abstract

First the notion of r-almost s-uniform tree is introduced which includes both the case of finitely generated free groups and uniform binary trees as special cases. The goal of the paper is to study pseudorandomness of binary functions defined on r-almost s-uniform trees. The measures of pseudorandomness of binary functions are introduced; the connection between these measures is analyzed; the size of these measures for truly random binary functions is studied; binary functions with strong pseudorandom properties are constructed.

1 Introduction

Recently in a series of papers a new constructive approach has been developed to study pseudorandomness of binary sequences

$$E_N = \{e_1, \dots, e_N\} \in \{-1, +1\}^N.$$

2000 Mathematics Subject Classification: Primary: 05C05, Secondary: 11K45.

Key words and phrases: pseudorandom, binary function, uniform tree, correlation, Legendre symbol.

Research partially supported by Hungarian National Foundation for Scientific Research, Grants No. K67676, K72731 and PD72264, French-Hungarian exchange program F-06/48, and the János Bolyai Research Fellowship.

In particular, in [25] first the following measures of pseudorandomness are introduced: the *well-distribution measure* of E_N is defined by

$$W(E_N) = \max_{a,b,t} \left| \sum_{j=0}^{t-1} e_{a+jb} \right|$$

where the maximum is taken over all $a, b, t \in \mathbb{N}$ with $1 \le a \le a+(t-1)b \le N$, the *correlation measure of order* k of E_N is defined as

$$C_k(E_N) = \max_{M,\mathbf{D}} \left| \sum_{n=1}^{M} e_{n+d_1} \dots e_{n+d_k} \right|$$

where the maximum is taken over all $\mathbf{D} = (d_1, \dots, d_k)$ and M such that $0 \le d_1 < \dots < d_k \le N - M$, and the *normality measure of order* k of E_N is defined as

$$N_k(E_N) = \max_{\mathbf{X} \in \{-1,+1\}^k} \max_{0<M\le N+1-k} \Big| |\{n : \ 0 \le n < M, \ (e_{n+1}, e_{n+2}, \dots, e_{n+k}) = \mathbf{X}\}| - \frac{M}{2^k} \Big|.$$

Then the sequence E_N is considered to be a "good" pseudorandom sequence if both $W(E_N)$ and $C_k(E_N)$ (at least for "small" k) are "small" in terms of N; in particular, both are $o(N)$ as $N \longrightarrow \infty$ (they showed that the normality measures can be estimated in terms of the correlation measures). Indeed, later Cassaigne, Mauduit and Sárközy [5] proved that this terminology is justified since for almost all $E_N \in \{-1, +1\}^N$, both $W(E_N)$ and $C_k(E_N)$ are less than $N^{1/2}(\log N)^c$; see also [3]. It was also shown in [25] that the Legendre symbol forms a "good" pseudorandom sequence. In some other paper further sequences were tested for pseudorandomness, further constructions were given for sequences with good pseudorandom properties, and the measures of pseudorandomness were studied (in the next sections we will present some further details of these results).

Later this theory of pseudorandomness of binary sequences has been extended in various directions: pseudorandomness of binary vectors [31], binary lattices [17], [18], [19], [20], [21], [22], [24], [27], [29], [30], subsets of

$\{1, 2, \ldots, n\}$ [6], [7], [8], subsets of $\mathbb{Z}_n$ [9], [10], sequences of k symbols [1], [4], [13], [26], [23], etc. have been studied. In this paper our goal is to continue this work by studying pseudorandomness of *binary functions on finite almost uniform trees*. More precisely, we will study finite r-almost s-uniform trees with $r \geq 2$, $s \geq 2$:

Definition 1 *If $r, s \in \mathbb{N}$ and $r \geq 2,\ s \geq 2$, then a tree is called an r-almost, s-uniform tree if the root has r children and, except for the vertices in the last row, all the other vertices have s children.*

If $r = s$ then the tree is called s-uniform tree, and in the $r = s = 2$ special case the tree is called uniform binary tree.

First we show a uniform binary tree:

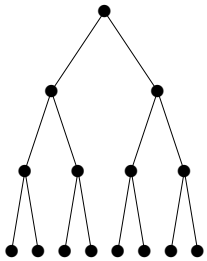

Figure 1.

Next we present a 4-almost 3-uniform tree:

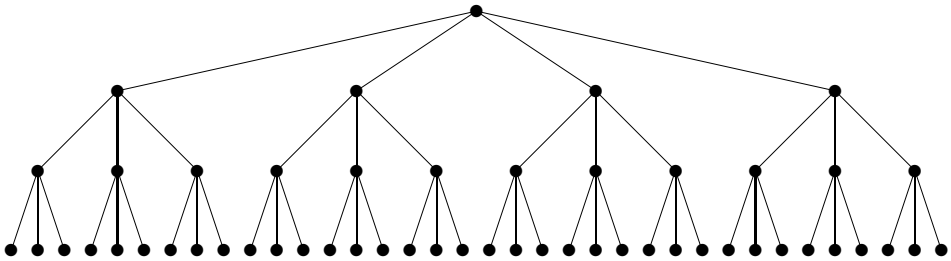

Figure 2.

These are the most important special cases: the uniform binary trees and the 4-almost 3-uniform trees. The importance of the uniform binary trees is clear, while the significance of the 4-almost 3-uniform trees is that there is a

bijection between the k-letter reduced words of a free group on two generators and the paths (of length k) connecting the root of a 4-almost 3-uniform tree of "height" k (of $k+1$ rows) with one of the vertices in the last row. (See [34].) Indeed, this connection with the free groups is the reason of that we are also considering trees where the number of the children of the root is not necessarily the same as the common number of children of all other vertices.

Denote the set of the vertices of a tree T by $\mathcal{P}(T)$.

Definition 2 *A binary function on the tree T is a function f of the type $f:\ \mathcal{P}(T) \to \{-1,+1\}$.*

In this paper *our goal is to study pseudorandomness of binary functions defined on r-almost s-uniform trees.* We will introduce measures of pseudorandomness of binary functions of this type; we will analyze the connection between these measures; we will study the size of these measures for truly random binary functions; and we will construct binary functions with strong pseudorandom properties.

2 Notations, terminology, connection with binary sequences, the measures of pseudorandomness

Throughout this paper we will use the following notations:

Tree will always mean an r-almost s-uniform rooted tree for some r, s. We will use the words path, distance, height, subtree in the usual sense.

Definition 3 *If T is an r-almost s-uniform tree, then a rooted subtree T' of T is called a* proper subtree *of T if either its root is the root of T and it is an r'-almost s-uniform tree for some $r' \le r$, or its root is different from the root of T and it is an s-uniform tree.*

We will use the word *row* in the following sense: the root forms the first row, the children of the root form the second row, and in general, the i-th row consists of the children of the vertices in the $i-1$-st row. Thus an r-almost s-uniform tree of height k has $k+1$ rows.

Consider an r-almost s-uniform tree, and draw it as in Figure 1 and Figure 2: the root is on the top, and the rows are drawn in horizontal lines. The j-th vertex from the left in the i-th row will be denoted by $P(i,j)$ (so that, starting from the left, the vertices in the i-th row are $P(i,1)$, $P(i,2)$, $P(i,3),\dots$).

From now on, we will formulate some basic facts in forms of *propositions*. It is an easy exercise to prove these facts (see [32], [35] for an approach of this type), thus we will not present the proofs.

Proposition 1 *The number of vertices in the i-th row of an r-almost s-uniform tree is 1 if $i=1$ (then this single vertex is the root $P(1,1)$), and rs^{i-2} if $i\geq 2$ (then these vertices are denoted by $P(i,1)$, $P(i,2),\dots,P(i,rs^{i-2})$).*

Proposition 2 *The total number $N=N(T)$ of vertices of an r-almost s-uniform tree T of height $k(\geq 1)$ is*

$$N=N(T)=1+r+rs+rs^2+\cdots+rs^{k-1}=1+r\frac{s^k-1}{s-1}.$$

(The number of vertices of a tree T will be always denoted by $N=N(T)$.)

We will also use the following alternative notation for the vertices: The root is denoted by Q_1 : $Q_1=P(1,1)$, the vertices in the second row by $Q_2,Q_3,\dots,Q_{r+1}$: $Q_2=P(2,1)$, $Q_3=P(2,2),\dots,Q_{r+1}=P(2,r)$; the vertices in the third row by $Q_{r+2},Q_{r+3},\dots,Q_{1+r(s+1)}$: $Q_{r+2}=P(3,1)$, $Q_{r+3}=P(3,2),\dots,Q_{1+r(s+1)}=P(3,rs)$, and so on; finally Q_N denotes the last vertex in the last row: $Q_N=P(k+1,rs^{k-1})$.

Proposition 3 *Define $y(i)$, $y'(i)$ so that $Q_{y(i)}=P(i,1)$ is the first vertex in*

the i-th row and $Q_{y'}(i) = P(i, rs^{i-2})$ is the last one. Then we have

$$\begin{aligned} y'(1) &= y(1) = 1 \\ y'(i) &= y(i) + rs^{i-2} - 1 \text{ for } i > 1, \\ y'(i) &= y(i+1) - 1 \text{ for } i \leq k, \\ y'(k+1) &= N, \\ y(i) &= 2 + r\frac{s^{i-2}-1}{s-1} \text{ for } i > 1, \\ y'(i) &= 1 + r\frac{s^{i-1}-1}{s-1} \text{ for } i > 1 \end{aligned}$$

and the j-th vertex in the i-th row is

$$P(i,j) = Q_{y(i)+j-1} = Q_{r\frac{s^{i-2}-1}{s-1}} \text{ for } i > 1.$$

To any binary function $f : \mathcal{P}(T) \to \{-1, +1\}$ defined on an r-almost s-uniform tree T of height k one may assign the unique binary sequence

$$E_N = E_N(f,T) = (e_1, e_2, \ldots, e_N) \in \{-1, +1\}^N \tag{2.1}$$

defined by

$$e_n = f(Q_n) \text{ for } n = 1, 2, \ldots, N. \tag{2.2}$$

Now we are ready to introduce the measures of pseudorandomness of binary functions defined on r-almost s-uniform trees T.

The most natural way of measuring the uniformity of the distribution of the binary function f over T relative to arithmetic progressions is to study the analogous property of $E_N(f,T)$.

Definition 4 *The* well-distribution measure *of the binary function f over T is defined by*

$$\overline{W}(f,T) = W(E_N(f,T)).$$

Note that k-ary functions on uniform trees (uniform trees coloured by k colours instead of using 2 colours as we do) and arithmetic progressions

also appear in a paper of Furstenberg and Weiss [11]. However, there are two basic differences between their problem and the problem studied by us here. First, they consider arithmetic progressions along paths only, while we consider arithmetic progressions lying in (one or more) rows. Secondly, they are looking for the extreme case of monochromatic arithmetic progressions, while we are interested whether the distribution in arithmetic progression is "typical", "random type" or it is not.

The definition of correlation is not so simple. We propose the following definition:

Definition 5 *For $k \geq 2$ and $\ell \geq 2$ the* correlation measure *$C_{k,\ell}(f,T)$ of height k and order ℓ of f over T is defined in the following way: consider ℓ different isomorphic proper subtrees $T_1, T_2, \ldots, T_\ell$ of height k of T, denote the set of their vertices by $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_\ell$, and for $t = 1, 2, \ldots, \ell$ let $\mathcal{P}_t = \{P_t(i,j) : i = 1, 2, \ldots, k+1,\ j = 1, 2, \ldots, q(i)\} = \{Q_{t,n} :\ n = 1, 2, \ldots, N(T_t)\}$ (note that both the number of vertices in the i-th row and $N(T_t)$ are independent of t by the isomorphism), and write*

$$U(T_1, T_2, \ldots, T_\ell) = \sum_{i=1}^{k+1} \sum_{j=1}^{q(i)} f(P_1(i,j)) f(P_2(i,j)) \cdots f(P_\ell(i,j))$$
$$= \sum_{n=1}^{N(T_t)} f(Q_{1,n}) f(Q_{2,n}) \ldots f(Q_{\ell,n}).$$

Then

$$\overline{C}_{k,\ell}(f,T) = \max_{T_1, T_2, \ldots, T_\ell} |U(T_1, T_2, \ldots, T_\ell)|$$

where the maximum is taken over all ℓ-tuples $T_1, T_2, \ldots, T_\ell$ of proper subtrees of the type described above.

Definition 6 *The* universal correlation measure of order *ℓ of f over T is defined by*

$$\tilde{C}_\ell(f,T) = \max_k \overline{C}_{k,\ell}(f,T).$$

(The normality measure will be introduced and studied later.)

3 The measures of f and the associated sequence E_N. The measures of f in the truly random case.

We will prove the following theorem:

Theorem 1 *If T is an r-almost s-uniform tree of height K, f is a binary function on T, and $k, \ell \in \mathbb{N}$ $1 \leq k \leq K$ and $\ell \geq 2$, then*

$$\overline{C}_{k,\ell}(f,T) \leq (k+1)C_\ell(E_N(f,T)). \tag{3.1}$$

Corollary 1 *For $\ell \geq 2$ we have*

$$\tilde{C}_\ell(f,T) < \left(\frac{\log N/r}{\log s} + 2\right) C_\ell(E_N(f,T)). \tag{3.2}$$

Proof of Theorem 1. Defining the isomorphic proper subtrees $T_1, T_2, \ldots, T_\ell$ as in Definition 5, we have

$$U(T_1, T_2, \ldots, T_\ell) = \sum_{i=1}^{k+1} \sum_{j=1}^{q(i)} f(P_1(i,j))f(P_2(i,j)) \ldots f(P_\ell(i,j)). \tag{3.3}$$

For $1 \leq t \leq \ell$, let e_{m_t} denote the element of E_N assigned to $P_t(i,1)$. Then $m_1, m_2, \ldots, m_\ell$ are distinct, since $T_1, T_2, \ldots, T_\ell$ are distinct; we may assume that $m_1 < m_2 < \cdots < m_\ell$. Define d_t by $d_t = m_t - 1$ for $t = 1, 2, \ldots, \ell$. Then by Proposition 3 the absolute value of the inner sum in (3.3) is

$$\begin{aligned}\left|\sum_{j=1}^{q(i)} f(P_1(i,j))f(P_2(i,j)) \ldots f(P_\ell(i,j))\right| &= \left|\sum_{j=1}^{q(i)} e_{m_1+j-1}e_{m_2+j-1} \ldots e_{m_\ell+j-1}\right| \\ &= \left|\sum_{j=1}^{q(i)} e_{j+d_1}e_{j+d_2} \ldots e_{j+d_\ell}\right| \\ &\leq C_\ell(E_N(f,t)). \end{aligned} \tag{3.4}$$

It follows from (3.3) and (3.4) that

$$U(T_1, T_2, \ldots, T_\ell) \le (k+1) C_\ell(E_N(f,T))$$

which proves (3.1).

Proof of Corollary 1. By Proposition 2 we have

$$N > r s^{K-1}$$

whence

$$K < \frac{\log n/r}{\log s} + 1$$

and using this, (3.2) follows from (3.1).

It was shown in [5] (by using Weil's theorem [36], that for almost all $E_N \in \{-1,+1\}^N$, both $W(E_N)$ and $C_\ell(E_N)$ are around $N^{1/2}$ (in a slightly sharper form); these results were sharpened in [3] where it was proved that for every $\ell \in \mathbb{N},\ \varepsilon > 0$ there is a $\delta > 0$ such that for large N we have

$$P\left(\delta N^{1/2} < W(E_N) < \frac{1}{\delta} N^{1/2}\right) > 1 - \varepsilon \tag{3.5}$$

and

$$P\left(\delta \left(\ell N \log N\right)^{1/2} < C_\ell(E_N) < \frac{1}{\delta}(\ell N \log N)^{1/2}\right) > 1 - \varepsilon. \tag{3.6}$$

It follows from these results, Definition 4 and Corollary 1 that for every $\ell \in \mathbb{N},\ \varepsilon > 0$ there is a $\delta > 0$ such for a large N and more than $(1-\varepsilon)2^N$ binary functions $f:\ \mathcal{P}(T) \to \{-1,+1\}$ (where again $N = N(T)$) we have

$$\delta N^{1/2} < W(f,T) < \frac{1}{\delta} N^{1/2}$$

and

$$\tilde{C}_\ell(f,T) < \frac{1}{\delta}(\ell N)^{1/2} (\log N)^{3/2}. \tag{3.7}$$

On the other hand, it does not follow from the earlier results that for more than $(1-\varepsilon)2^N$ binary functions $f:\ \mathcal{P}(T) \to \{-1,+1\}$ we have

$$\tilde{C}_\ell(f,T) > \delta N^{1/2}.$$

Thus we will prove

Theorem 2 *For every $\varepsilon > 0,\ \ell \in \mathbb{N},\ \ell \geq 2,\ r, s \in \mathbb{N},\ r, s \geq 2$ there are numbers $\delta = \delta(\varepsilon, \ell, r, s)$ and $N_0 = N_0(\varepsilon, \ell, r, s)$ such that if T is an r-almost s-uniform tree with $N = N(T) > N_0$, then for more than $(1-\varepsilon)2^N$ binary functions $f:\ \mathcal{P}(T) \to \{-1, +1\}$ we have*

$$\delta N^{1/2} < \tilde{C}_\ell(f, T). \tag{3.8}$$

(Note that there is a logarithm power gap between the lower bound (3.8) and upper bound (3.7). One might like to close this gap; it is not clear which one of the two bounds is closer to the truth.)

Proof of Theorem 2. Denote the height of T by k, and let $f(Q_1) = f(P(1,1)) = e_1$, $f(Q_2) = f(P(2,1)) = e_2, \ldots, f(Q_N) = f(P(k+1, rs^{k-1})) = e_N$ be independent random variables of distribution

$$P(e_i = +1) = P(e_i = -1) = \frac{1}{2}. \tag{3.9}$$

We have to show that

$$P\left(\delta N^{1/2} < \tilde{C}_\ell(f, T)\right) > 1 - \varepsilon \tag{3.10}$$

(for δ small enough and N large enough in terms of ε, ℓ, r and s).

Now define the integer i so that the i-th row is the first row of T which contains at least ℓ vertices. In other words, we have

$$i = 2, \quad \text{if } \ell \leq r \tag{3.11}$$

and i is defined by

$$rs^{i-3} < \ell \leq rs^{i-2} \quad \text{if } \ell > r. \tag{3.12}$$

Let $T_1, T_2, \ldots, T_\ell$ denote the s-uniform subtrees of T whose roots are $P(i,1)$, $P(i,2), \ldots, P(i,\ell)$ (the first ℓ vertices of the i-th row of T) and whose height is $k - i + 1$ (so that the vertices in their last rows belong to the last row of T). Moreover, for $1 \leq t \leq \ell$ let T_t' denote the s-uniform subtree of T that we

get from T_t by dropping its last row (so that the root of T'_t is $P(i,t)$ and its height is $k-i$). Finally, define m by

$$Q_{m+1} = P(k+1,1)$$

(in other words, the last vertex in the k-th row of T is Q_m).

First we select the values of $e_1, e_2, \ldots, e_m$ and $e_{m+s^{k-i+1}+1}, e_{m+s^{k-i+1}+2}, \ldots, e_N$ (each of them are selected according to the law (3.9)). We will show that independently of the choice of these e_j's, the remaining e_j's, i.e.,

$$e_{m+1}, e_{m+2}, \ldots, e_{m+s^{k-i+1}} \tag{3.13}$$

can be selected with probability $> 1-\varepsilon$ so that the event

$$\delta N^{1/2} < \tilde{C}_\ell(f,T) \tag{3.14}$$

holds.

For $j = 1, 2, \ldots, s^{k-i+1}$, write

$$g_j = \prod_{t=1}^{\ell-1} f\left(P\left(k+1, j+ts^{k-i+1}\right)\right) \left(= \prod_{t=1}^{\ell} e_{m+j+ts^{k-i+1}}\right),$$

and define h_j by

$$h_j = e_{m+j} g_j \quad \left(\text{for } j = 1, 2, \ldots, s^{k-i+1}\right). \tag{3.15}$$

For every fixed choice of the e_j's not in (3.13) i.e., for any $g_1, g_2, \ldots, g_{s^{k-i+1}} \in \{-1, +1\}$, the h_j's in (3.15), together with the e_{m+j}'s are distributed according to the law in (3.9) so that

$$P(h_j = -1) = P(h_j = +1) = \frac{1}{2} \quad \left(\text{for } j = 1, 2, \ldots, s^{k-i+1}\right). \tag{3.16}$$

Write $S = h_1 + h_2 + \cdots + h_{s^{k-i+1}}$. By (3.16), the expectation of S is 0, its standard deviation is $\frac{1}{2}\left(s^{k-i+1}\right)^{1/2}$, thus by the central limit theorem the limit distribution of $S/\frac{1}{2}\left(s^{k-i+1}\right)^{1/2}$ for $s^{k-i+1} \to \infty$ (which follows from $N \to \infty$)

is Gaussian distribution. It follows that if N is large enough and $\delta(>0)$ is small enough in terms of ε, ℓ, r and s, then the probability of the event

$$\frac{|S|}{\frac{1}{2}\left(s^{k-i+1}\right)^{1/2}} > 8\delta(\ell s)^{1/2} \tag{3.17}$$

is greater, than $1-\varepsilon$ uniformly for any fixed choice of the e_j's not in (3.13), so that for a random choice of the binary function $f:\ \mathcal{P}(T) \to \{-1,+1\}$ we have

$$P\left(\frac{|S|}{\frac{1}{2}\left(s^{k-i+1}\right)^{1/2}} > 8\delta(\ell s)^{1/2}\right) > 1-\varepsilon.$$

Thus in order to prove (3.10), it suffices to show that the event (3.14) follows from the event (3.17).

Assume that (3.17) holds. For $t = 1, 2, \ldots, \ell$, denote the vertices of the subtrees T_t and T'_t by $P_t(r,j)$ (=the j-th vertex in the r's row) and $P'_t(r,j)$, respectively (so that $P_t(r,j) = P'_t(r,j)$ for $r = 1, 2, \ldots, k-i+1$, $j = 1, 2, \ldots, s^{r-1}$). Then by (3.12),(3.17) and Proposition 2, we have

$$|U(T_1, T_2, \ldots, T_\ell) - U(T'_1, T'_2, \ldots, T'_\ell)| = \left| \sum_{r=1}^{k-i+2} \sum_{j=1}^{s^{r-1}} f(P_1(r,j)) f(P_2(r,j)) \ldots \right.$$

$$\left. f(P_\ell(r,j)) - \sum_{r=1}^{k-i+1} \sum_{j=1}^{s^{r-1}} f(P'_1(r,j)) f(P'_2(r,j)) \ldots f(P'_\ell(r,j)) \right|$$

$$= \left| \sum_{j=1}^{s^{k-i+1}} f(P_1(k-i+2,j)) f(P_2(k-i+2,j)) \ldots f(P_\ell(k-i+2,j)) \right|$$

$$= \left| \sum_{j=1}^{s^{k-i+1}} f(P(k+1,j)) f(P(k+1, j+s^{k-i+1})) \ldots \right.$$

$$\left. f(P_\ell(k+1, j+(\ell-1)s^{k-i+1})) \right| = \left| \sum_{j=1}^{s^{k-i+1}} e_{m+j} g_j \right| = |S|$$

$$> 4\delta\left(\ell s^{k-i+2}\right)^{1/2} > 4\delta\left(r s^{i-3} s^{k-i+2}\right)^{1/2} = 4\delta(r s^{k-1})^{1/2} = 2\delta\left(2r\frac{s^k}{s/2}\right)^{1/2}$$

$$> 2\delta\left(2r\frac{s^k-1}{s-1}\right)^{1/2} \geq 2\delta N^{1/2}.$$

It follows from this inequality that

$$\max\{U(T_1, T_2, \ldots, T_\ell), U(T'_1, T'_2, \ldots, T'_\ell)\} \geq \delta N^{1/2}$$

whence by Definitions 5 and 6,

$$\begin{aligned}\tilde{C}_\ell(f,T) = \max_j \overline{C}_{j,\ell}(f,T) &\geq \max\{\overline{C}_{k-i+1,\ell}(f,T), \overline{C}_{k-i,\ell}(f,T)\} \\ &\geq \max\{U(T_1, T_2, \ldots, T_\ell), U(T'_1, T'_2, \ldots, T'_\ell)\} \geq \delta N^{1/2}\end{aligned}$$

which proves (3.14) (assuming (3.17)) and thus also (3.10).

Based on these facts, we may say that a binary function $f : \mathcal{P}(T) \to \{-1, +1\}$ is a "good pseudorandom binary function", "it possesses strong pseudorandom properties" if $\overline{W}(f,T)$ and $\tilde{C}_\ell(f,T)$ (at least for small ℓ) are small in terms of $N = N(T)$; they must be $o(N)$, and ideally they are $O(N^{1/2+\varepsilon})$.

Note that if T is an r-almost s-uniform tree with $N = N(T)$, then any "good" pseudorandom binary sequence $E_N = (e_1, e_2, \ldots, e_N) \in \{-1, +1\}^N$ induces a "good" pseudorandom binary function $f : \mathcal{P}(T) \to \{-1, +1\}$. Namely, if we define f by $f(Q_n) = e_n$ (for $n = 1, 2, \ldots, N$) and E_N possesses good pseudorandom properties, then by Definition 4 and Corollary 1 the binary function f also possesses good pseudorandom properties.

Referring to the statement of the last paragraph, the referee of this paper asked the following question: "What can be said about the converse?" The answer to this question is that the converse of this statement is not true: it may occur that the binary function $f : \mathcal{P}(T) \to \{-1, +1\}$ possesses good pseudorandom properties but the associated binary sequence $E_N(f,T)$ does not; more precisely, it may occur that both $\overline{W}(f,T)(= W(E_N(f,T)))$ and $\tilde{C}_\ell(f,T)$ (for almost all "small" ℓ) are "small"$(= o\left(N^{1/2+\varepsilon}\right))$, however $C_2(E_N(f,T))$ is "large" $(\gg N)$. This can be shown by the following example:

Example 1 Consider the uniform binary tree T of height k, so that it has $N = N(T) = 2^{k+1} - 1$ vertices, and its vertices in the last row are $Q_{2^k} =$

$P(k+1,1),\ Q_{2^k+1} = P(k+1,2), \ldots, Q_N = Q_{2^{k+1}-1} = P(k+1,2^k)$. Define the binary function $f : \mathcal{P}(T) \to \{-1,+1\}$ so that $f(Q_1) = f(P(1,1)) = e_1\ (\in \{-1,+1\})$, $f(Q_2) = f(P(2,1)) = e_2\ (\in \{-1,+1\}), \ldots, f(Q_{2^k+2^{k-1}}) = f(P(k+1,2^{k-1}+1)) = e_{2^k+2^{k-1}}\ (\in \{-1,+1\})$ in a truly random way, and let

$$\begin{aligned} f(Q_{2^k+2^{k-1}+i}) &= f(P(k+1,2^{k-1}+1+i)) = e_{2^k+2^{k-1}+i} = e_{2^k-1+i} \\ &= f(P(k+1,i)) = f(Q_{2^k+i-1}). \end{aligned}$$

Then we have

$$\begin{aligned} C_2(E_N(f,T)) &\geq \left| \sum_{i=1}^{2^{k-1}-1} e_{2^k-1+i} e_{2^k+2^{k-1}+i} \right| = \left| \sum_{i=1}^{2^{k-1}-1} 1 \right| \\ &= 2^{k-1} - 1 \gg 2^{k+1} - 1 = N. \end{aligned}$$

On the other hand, it is easy to see that with probability $> 1-\varepsilon$ we have $\overline{W}(f,T) \ll N^{1/2}(\log N)^c$ for some $c > 0$, and it could be shown with a little work that for every fixed ℓ with probability $> 1-\varepsilon$ we also have $\overline{C}_\ell(f,T) \ll N^{1/2}(\log N)^c$ for some $c = c(\ell) > 0$. (Observe that for this f and e.g., $\ell = 2$ the estimate of the sum in Definition 5 can be reduced to $O(\log N)$ sums of form

$$\sum_{n=m}^{M} e_n e_{n+d}$$

where $1 \leq m < m+M \leq 2^k + 2^{k-1}$, and by the structure of the construction we have $d \neq 0$.)

Thus, e.g., we may construct a "good" pseudorandom binary function f on a given almost uniform tree T by using the Legendre symbol. Let $N = N(T)$, and let p denote the smallest odd prime with $N < p$ so that by Tchebycheff's theorem we have $p \leq 2N$. Then consider the Legendre symbol sequence $E_{p-1} = (e_1, e_2, \ldots, e_{p-1})$ defined by

$$e_n = \left(\frac{n}{p}\right) \quad \text{for } n = 1,2,\ldots,p-1. \tag{3.18}$$

It was shown in [25] that for this binary sequence E_{p-1} we have

$$W(E_{p-1}) \ll p^{1/2} \log p \tag{3.19}$$

and

$$C_\ell(E_{p-1}) \ll \ell p^{1/2} \log p \tag{3.20}$$

(in [25] slightly different notation was used). Then consider the binary function

$$\lambda : \ \mathcal{P}(T) \to \{-1, +1\} \tag{3.21}$$

associated with the truncated binary sequence $E_N = (e_1, e_2, \ldots, e_N)$. It follows from definition 4, Corollary 1, (3.19) and (3.20) that for this binary function we have

$$\overline{W}(\lambda, T) = W(E_N(\lambda, T)) \ll p^{1/2} \log p \ll N^{1/2} \log N$$

(recall that $p < N < 2p$) and

$$\tilde{C}_\ell(\lambda, T) \ll (\log N) C_\ell(E_N(\lambda, T)) \ll (\log N) \ell p^{1/2} \log p \ll \ell N^{1/2} (\log N)^2. \tag{3.22}$$

By Corollary 1, the correlation of order ℓ of a binary function f can be estimated from above in terms of the correlation of order ℓ of the associated sequence E_N, i.e., if $C_\ell(E_N(f, T))$ is small, then $\tilde{C}_\ell(f, T)$ also must be small. One might like to know whether the converse of this is also true, i.e., if $\tilde{C}_\ell(f, T)$ is small, then $C_\ell(E_N(f, T))$ also must be small? We will show that the answer to this question is negative. To simplify the discussion we will restrict ourselves to uniform binary trees and to the special case $\ell = 2$ but the proof could be extended to the general case.

We will start out from a slight modification of the Legendre symbol construction above. Consider a uniform binary tree T, write $N = N(T)$, and let p denote the smallest prime p with $4N < p$ (so that now p is slightly greater than above). Then we have

$$4N < p < 8N. \tag{3.23}$$

Define the Legendre symbol sequence (3.18) and the binary function λ in (3.21) as above. Note that (3.22) also holds for this modified construction (only the implicit constant changes).

Let T' denote the uniform binary tree obtained from T by adding one more row: Denote the height of T' by H so that $N = N(T) = 2^H - 1$ and, writing $N' = N(T')$, we have $N' = 2^{H+1} - 1$. Let $\lambda' : \mathcal{P}(T') \to \{-1, +1\}$ denote the extension of λ from T to T' defined by

$$\lambda'(P(i,j)) = \lambda(P(i,j)) \text{ for } 1 \le i \le H,\ 1 \le j \le 2^{i-1}$$

and

$$\lambda'(P(H+1,2j-1)) = \lambda'(P(H+1,2j)) = \lambda(P(H,j)) \text{ for } 1 \le j \le 2^{H-1}$$

(so that λ' assumes the same value on the vertices of the last row of T as on their children). Then the binary sequences $E_N = E_N(\lambda, T)$ and $E_{N'} = E_{N'}(\lambda', T')$ are

$$E_N = (e_1, e_2, \dots, e_N) = \left(\left(\frac{1}{p}\right)\left(\frac{2}{p}\right), \dots, \left(\frac{N}{p}\right)\right) \tag{3.24}$$

and

$$E_{N'} = (e_1, e_2, \dots, e_{N'}) = \left(\left(\frac{1}{p}\right)\left(\frac{2}{p}\right), \dots, \left(\frac{N}{p}\right), \left(\frac{2^{H-1}}{p}\right), \left(\frac{2^{H-1}}{p}\right), \left(\frac{2^{H-1}+1}{p}\right), \left(\frac{2^{H-1}+1}{p}\right), \dots, \left(\frac{2^H-1}{p}\right), \left(\frac{2^H-1}{p}\right)\right). \tag{3.25}$$

Theorem 3 *Defining T', f' and N' in this way we have*

$$\tilde{C}_2(\lambda', T') \ll N^{1/2}(\log N)^2 \tag{3.26}$$

and

$$C_2(E_{N'}(\lambda', T')) \gg N'. \tag{3.27}$$

Proof of Theorem 3. In order to prove (3.26) we have to show that defining T_1, T_2 and $U(T_1, T_2)$ as in Definition 5 with T', λ' and 2 in place of T, f and ℓ, the inequality

$$\left| \sum_{i=1}^{k+1} \sum_{j=1}^{q(i)} \lambda'(P_1(i,j))\lambda'(P_2(i,j)) \right| \ll N^{1/2}(\log N)^2 \tag{3.28}$$

holds. To prove this we have to distinguish three cases.

CASE 1. T_1, T_2 are subtrees of T, i.e., no vertex of the last row of T' is among the vertices of T_1 and T_2. Then by the construction we have

$$\left| \sum_{i=1}^{k+1} \sum_{j=1}^{q(i)} \lambda'(P_1(i,j))\lambda'(P_2(i,j)) \right| = \left| \sum_{i=1}^{k+1} \sum_{j=1}^{q(i)} \lambda(P_1(i,j))\lambda(P_2(i,j)) \right| \le \tilde{C}_2(\lambda, T)$$

whence (3.28) follows by (3.22).

CASE 2. Assume that one of T_1 and T_2, say T_1 is a subtree of T, but the last row of the other one, T_2 consists of consecutive elements of the last row of T'. Then we have

$$\begin{aligned} \left| \sum_{i=1}^{k+1} \sum_{j=1}^{q(i)} \lambda'(P_1(i,j))\lambda'(P_2(i,j)) \right| \le & \left| \sum_{i=1}^{k} \sum_{j=1}^{q(i)} \lambda'(P_1(i,j))\lambda'(P_2(i,j)) \right| \\ & + \left| \sum_{j=1}^{q(k+1)} \lambda'(P_1(k+1,j))\lambda'(P_2(k+1,j)) \right|. \end{aligned} \tag{3.29}$$

Here the first double sum can be estimated as the one in Case 1:

$$\left| \sum_{i=1}^{k+1} \sum_{j=1}^{q(i)} \lambda'(P_1(i,j))\lambda'(P_2(i,j)) \right| \le \tilde{C}_2(\lambda, T) \ll N^{1/2}(\log N)^2. \tag{3.30}$$

By (3.24), (3.25) and our assumption on T_1 and T_2, the last sum is of the form

$$\left|\sum_{j=1}^{q(k+1)}\left(\frac{a+j}{p}\right)\lambda'(P'(H+1,2b+j))\right| = \left|\sum_{u=1}^{\frac{q(k+1)}{2}}\left(\frac{a+2u}{p}\right)\left(\frac{d+u}{p}\right)\right.$$
$$\left.+\sum_{u=1}^{\frac{q(k+1)}{2}}\left(\frac{a+2u-1}{p}\right)\left(\frac{d+u}{p}\right)\right| \le \left|\sum_{u=1}^{\frac{q(k+1)}{2}}\left(\frac{(a2^{-1}+u)\,(d+u)}{p}\right)\right|$$
$$+\left|\sum_{u=1}^{\frac{q(k+1)}{2}}\left(\frac{((a-1)2^{-1}+u)\,(d+u)}{p}\right)\right| \tag{3.31}$$

where a, d are integers with

$$0 \le a \le N - q(k+1) = 2^H - 1 - q(k+1) \tag{3.32}$$
$$2^{H-1} - 1 \le d \le N' - 1 = 2^{H+1} - 1 \tag{3.33}$$

(and 2^{-1} denotes the multiplicative inverse of 2 modulo p).

Now we need the following well-known consequence of Weil's theorem [36]:

Lemma 1 *If p is a prime, χ is a non-principal character modulo p of order D, $g(x) \in \mathbb{F}_p[x]$ is not of the form $a(h(x))^d$ with $a \in \mathbb{F}_\imath$, $h(x) \in \mathbb{F}_p[x]$, $y \in \mathbb{Z}$, $z \in \mathbb{N}$ and $z \le p$, then we have*

$$\left|\sum_{n=y}^{y+z}\chi(g(n))\right| \le 10tp^{1/2}\log p$$

where t denotes the number of distinct zeros of $g(x)$.

(See e.g. [12], [25], [33] for different versions of this lemma.)

To use this lemma (with $\chi(n) = \left(\frac{n}{p}\right)$) for estimating the sums in (3.31), we have to show that the polynomials of u there are not perfect squares, i.e.,

$$a2^{-1} \not\equiv d \pmod p \text{ and } (a-1)2^{-1} \not\equiv d \pmod p$$

or, in equivalent form,

$$a \not\equiv 2d \pmod{p} \text{ and } a-1 \not\equiv 2d \pmod{p}. \tag{3.34}$$

By (3.23), (3.32) and (3.33) we have

$$-1 \le a-1 < a \le 2^H - 1 - q(k+1) \tag{3.35}$$

and

$$2^H - 2 = 2\left(2^{H-1}-1\right) \le 2d \le N' - 1 = 2N+1 \le 4N-1 < p-1. \tag{3.36}$$

The case $q(k+1) = 1$ is trivial (then the upper bound in (3.31) is $O(1)$). Thus we may assume that $q(k+1) > 1$. Then by (3.35) and (3.36) we have

$$-1 \le a-1 < a < 2d < p-1$$

and (3.34) follows from this. Thus the lemma can be applied, and then we obtain from (3.31) that

$$\left|\sum_{j=1}^{q(k+1)} \lambda'(P_1(k+1,j))\lambda'(P_2(k+1,j))\right| \ll p^{1/2}\log p \ll N^{1/2}\log N \tag{3.37}$$

(which also holds for $q(k+1) = 1$).

(3.28) follows from (3.30) and (3.37).

CASE 3. Assume that the last row of both T_1 and T_2 consists of consecutive elements of the last row of T'. Then as in (3.29) and (3.30) in Case 2, we obtain

$$\begin{aligned}\left|\sum_{i=1}^{k+1}\sum_{j=1}^{q(i)} \lambda'(P_1(i,j))\lambda'(P_2(i,j))\right| &\ll N^{1/2}(\log N)^2 \\ + \left|\sum_{j=1}^{q(k+1)} \lambda_1'(P_1(k+1,j))\lambda_2'(P_2(k+1,j))\right|&. \end{aligned}\tag{3.38}$$

The last term can be rewritten as

$$\left|\sum_{u=1}^{\frac{q(k+1)}{2}}\left(\frac{a+u}{p}\right)\left(\frac{b+u}{p}\right)\right| = \left|\sum_{u=1}^{\frac{q(k+1)}{2}}\left(\frac{(a+u)(b+u)}{p}\right)\right|$$

with some integers a, b such that

$$0 \le a < b < p.$$

Then again we may apply Lemma 1 to estimate this sum, and we obtain

$$\left|\sum_{j=1}^{q(k+1)} \lambda_1'(P_1(k+1,j))\lambda_2'(P_2(k+1,j))\right| \ll p^{1/2}\log p \ll N^{1/2}\log N. \quad (3.39)$$

(3.28) follows from (3.38) and (3.39), and this completes the proof of (3.26).

Now we will prove (3.27). $e_1, e_2, \ldots, e_{N'}$ are defined as in (3.25), then

$$\begin{aligned} C_2(E_{N'}(\lambda', T')) &\ge \left|\sum_{n=N+1}^{N'-1} e_n e_{n+1}\right| \\ &= \left|\sum_{\substack{N+1\le n\le N'-1\\ 2|n}} e_n e_{n+1} + \sum_{\substack{N+1\le n\le N'-1\\ 2\nmid n}} e_n e_{n+1}\right| \\ &= \left|\sum_{\substack{N+1\le n\le N'-1\\ 2|n}} 1 + \sum_{m=2^{H-1}}^{2^H-2}\left(\frac{m}{p}\right)\left(\frac{m+1}{p}\right)\right| \\ &= \left|\frac{N'-N}{2} + \sum_{m=2^{H-1}}^{2^H-2}\left(\frac{m(m+1)}{p}\right)\right|. \end{aligned} \quad (3.40)$$

Since $N = \frac{1}{2}N' + O(1)$ and, by (3.23) and Lemma 1,

$$\left|\sum_{m=2^{H-1}}^{2^H-2}\left(\frac{m(m+1)}{p}\right)\right| \ll p^{1/2}\log p \ll N^{1/2}\log N,$$

it follows from (3.40) that

$$C_2(E_{N'}(\lambda', T')) \gg N$$

which completes the proof of Theorem 3.

Numerous other properties of the measures of pseudorandomness of binary sequences have been studied [3], [2], [5], [14], [15], [16], [28]. In particular, $\min_{E_N \in \{-1,+1\}^N} W(E_N)$ and $\min_{E_N \in \{-1,+1\}^N} C_k(E_N)$ have been estimated; an inequality between $W(E_N)$ and $C_2(E_N)$ (and later $C_k(E_N)$) has been proved; for fixed k and ℓ, the connection between $C_k(E_N)$ and $C_\ell(E_N)$ has been analyzed; an inequality between $C_2(E_N)$ and $C_3(E_N)$ has been proved. One might like to look for analogues of these results for binary functions. However, these problems seem to be much more difficult for binary functions than for binary sequences, and we have not been able to settle them. The constructions, resp. the crucial ideas of the proofs which were used in the case of binary sequences fail in the case of binary functions completely. Thus we just do not know what to expect in the latter case. Before formulating any conjectures in this directions, first one has to answer the following basic questions (that we have not been able to settle either): Is it true, that if r, s are fixed integers, $2k+1$ is a fixed odd integer, and $N = N(T) \longrightarrow \infty$, then $\min \tilde{C}_{2k+1}(f, T) = O(1)$ where the minimum is taken over all binary functions defined on r-almost s-uniform trees of N vertices? Is it true, that it follows from $\tilde{C}_2(f, T) = o(N)$ that $\overline{W}(f, T) = o(N)$?

4 The normality measure

So far we have introduced and studied the well-distribution and correlation measures of binary functions defined on r-almost s-uniform trees. In this section we will introduce and study the *normality measure of order k* of binary functions of this type. A simple way of defining the normality measure of order k of the binary function $f : \mathcal{P}(T) \to \{-1, +1\}$ would be to define it in the same manner as the well-distribution measure (Definition 4),

i.e., by

$$\overline{N}_k(f,T) = N_k(E_N(f,T)).$$

However, this seems to be a rather artificial definition of not much use. Instead, we propose to use the following more natural definition:

Definition 7 *The* normality measure $\overline{N}_k(f,T)$ of order k *($k \in \mathbb{N},\ k \geq 2$) of the binary function f over the r-almost s-uniform tree T is defined in the following way: Let τ_k denote the set of uniform binary subtrees of height k of T. If $G_{2^{k+1}-1} = (g_1, g_2, \ldots, g_{2^{k+1}-1}) \in \{-1,+1\}^{2^{k+1}-1}$, then let $\phi(f,T,G_{2^{k+1}-1})$ denote the number of the subtrees $T' \in \tau_k$ such that the binary sequence $E_{2^{k+1}-1} = E_{2^{k+1}-1}(f,T')$ assigned to the binary function $f:\ \mathcal{P}(T') \to \{-1,+1\}$ (restricted to T' and defined after Proposition 2) is the given $2^{k+1}-1$ tuple $G_{2^{k+1}-1}$:*

$$\phi(f,T,G_{2^{k+1}-1}) = \left|\{T':\ T' \in \tau_k,\ E_{2^{k+1}-1}(f,T') = G_{2^{k+1}-1}\}\right|.$$

Then define $\overline{N}_k(f,T)$ by

$$\overline{N}_k(f,T) = \max_{G_{2^{k+1}-1} \in \{-1,+1\}^{2^{k+1}-1}} \left|\phi\left(f,T,G_{2^{k+1}-1}\right) - \frac{|\tau_k|}{2^{2^{k+1}-1}}\right|.$$

(So that $\overline{N}_k(f,T)$ is defined as the maximal deviation between $\phi(f,T,G_{2^{k+1}-1})$ and its expected value for all the possible choices of $G_{2^{k+1}-1}$.)

As (3.6) shows, for fixed $k, N \to \infty$ and a truly random binary sequence $E_N \in \{-1,+1\}^N$ one has

$$C_\ell(E_N) = O(N^{1/2+\varepsilon}). \tag{4.1}$$

It was proved in [25] (see Proposition 1) that for all N, E_N and $k < N$ we have

$$N_k(E_N) \leq \max_{1 \leq t \leq k} |C_t(E_N)|. \tag{4.2}$$

Combining (4.1) and (4.2) we get that for fixed k, $N \to \infty$ and a truly random $E_N \in \{-1,+1\}^N$ we have

$$N_k(E_N) = O(N^{1/2+\varepsilon}).$$

One may expect that the measure $\overline{N}_k(f,T)$ possesses an analogous property: if k, r, s are fixed and we consider r-almost s-uniform trees T with $N(T) \to \infty$, then for a truly random binary function f over T one has

$$\overline{N}_k(f,T) = O\left(N(T)^{1/2+\varepsilon}\right). \tag{4.3}$$

This is probably true but, unfortunately, we have not been able to prove it. The difficulty is that the above argument cannot be adopted to binary functions over trees: namely, there is no inequality of type (4.2) in other words, it may occur that $\tilde{C}_t(f,T)$ is small for all $t \le k$, however, $\overline{N}_k(f,T)$ is large. Indeed, we will show that this is the case for the especially important Legendre symbol construction studied in (3.18) and (3.21) in Section 3. As (3.22) shows, for this construction $\tilde{C}_t(\lambda,T)$ is small for every fixed t, thus it remains to show that $\overline{N}_k(\lambda,T)$ is large (for fixed k). To simplify the discussion we will restrict ourselves to uniform binary trees.

Theorem 4 *Let $k, K \in \mathbb{N}$, $k \le K$, let T denote uniform binary tree of height K, and define the binary function $\lambda : \mathcal{P}(T) \to \{-1,+1\}$ by* (3.18) *and* (3.21)*. Then we have*

$$\overline{N}_k(\lambda,T) > \frac{2^{2^{k+1}-1-k}-1}{2^{2^{k+1}-1}}\left(2^{K-k+1}-1\right). \tag{4.4}$$

Note that for fixed k and $K \to \infty$ this implies that

$$\overline{N}_k(\lambda,T) \gg N(T)$$

(where the implicit constant factor depends on k).

Proof of Theorem 4. First we will compute $|\tau_k|$. Denote the height of T by K so that by Proposition 2 we have

$$N = N(T) = 2^{K+1} - 1. \tag{4.5}$$

A subtree $T' \in \tau_k$ is uniquely determined by its root, and the root must belong to one of the first $K-k+1$ rows of T. Again by Proposition 2, the total number of vertices of these rows is 2^{K-k+1} so that

$$|\tau_k| = 2^{K-k+1} - 1. \tag{4.6}$$

Let $\mathcal{G}$ denote the set of the $2^{k+1}-1$ tuples $G_{2^{k+1}-1}$ such that $\phi(\lambda, T, G_{2^{k+1}-1}) > 0$, i.e., there is at least one $T' \in \tau_k$ with

$$E_{2^{k+1}-1}(\lambda, T') = (e'_1, e'_2, \ldots, e'_{2^{k+1}-1}) = G_{2^{k+1}-1}. \tag{4.7}$$

Now we will give an upper bound for $|\mathcal{G}|$. Assume that $G_{2^{k+1}-1} = (g_1, g_2, g_3, \ldots, g_{2^{k+1}-1}) \in \mathcal{G}$, and there is a $T' \in \tau_k$ satisfying (4.7). We will show that

$$g_1, g_3, g_5, \ldots, g_{2^{k+1}-1} \tag{4.8}$$

uniquely determine

$$g_2, g_4, g_6, \ldots, g_{2^{k+1}-2}.$$

Indeed, assume that we are looking for $g_{2\ell}$ with $\ell \in \{1, 2, \ldots, 2^k - 1\}$. It follows from (4.2) that

$$e'_m = g_m \quad \text{for } m = 1, 2, \ldots, 2^{k+1} - 1 \tag{4.9}$$

and, in particular,

$$e'_{2\ell} = g_{2\ell}. \tag{4.10}$$

Now consider a uniform binary tree $\tilde{T}$. Denote its j-th vertex in the i-th row by $\tilde{P}(i,j)$. We will say that $\tilde{P}(i,j)$ is an even vertex if j is even, and it is said to be an odd vertex if j is odd. Note that if $\tilde{P}(i,j)$ is an odd vertex of $\tilde{T}$, and it is also a vertex of a binary uniform subtree $\tilde{T}'$ of $\tilde{T}$, then $\tilde{P}(i,j)$ is also an odd vertex in $\tilde{T}'$. It follows from Proposition 3 that if a binary function $f : \mathcal{P}(\tilde{T}) \to \{-1, +1\}$ is given, and $\tilde{E_N} = \tilde{E_N}(f, \tilde{T}) = (\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_N)$ (with $N = N(\tilde{T})$) is the binary sequence associated with it (in the sense described after Proposition 2), then we have

$$f\left(\tilde{P}(i,j)\right) = \tilde{e}_{2^{i-1}+j-1}. \tag{4.11}$$

It follows from this that the vertex $\tilde{P}(i,j)$ is odd if and only if the subscript of the associated $\tilde{e}_n$ is even:

$$f\left(\tilde{P}(i,j)\right) = \tilde{e}_{2q} \quad (\text{iff } \tilde{P}(i,j) \text{ is odd}). \tag{4.12}$$

Let us consider an odd vertex $P(i,2u-1)$ of T which is not the root; this is the "left" child of the vertex $P(i-1,u)$. Then by (3.18) and (4.11) we have

$$\begin{aligned}\lambda(P(i,2u-1)) = e_{2^{i-1}+2u-2} &= \left(\frac{2^{i-1}+2u-2}{p}\right) = \left(\frac{2}{p}\right)\left(\frac{2^{i-2}+u-1}{p}\right) \\ &= \left(\frac{2}{p}\right) e_{2^{i-2}+u-1} = \left(\frac{2}{p}\right)\lambda(P(i-1,u)).\end{aligned} \tag{4.13}$$

Now we return to (4.10); then it follows from the discussion above that $e'_{2\ell}$ is associated with an odd vertex $P'(i,j)$ of T'. Let us consider the directed path starting from $P'(i,j)$ and ending with the root $P'(1,1)$ of T'. The root $P'(1,1)$ is an even vertex; let $P'(i,x,v)$ be the first even vertex along this path. Denote the subscript of the element of $E_{2^{k+1}-1}(\lambda,T')$ associated with it by $2t-1$ (it must be odd by (4.12)) so that, by (4.9),

$$\lambda(P'(i-x,v)) = e'_{2t-1} = g_{2t-1}. \tag{4.14}$$

Then using (4.13) repeatedly, we obtain from (4.14) that

$$g_{2\ell} = e'_{2\ell} = \lambda\left(P'(i,j)\right) = \left(\frac{2}{p}\right)^x \lambda\left(P'(i-x,v)\right) = \left(\frac{2}{p}\right)^x g_{2t-1}$$

so that, indeed, $g_{2\ell}$ is uniquely determined by g_{2t-1} (note that p is fixed, and x is determined by ℓ).

The sequence (4.8) can be selected in at most 2^k ways (since we have $g_i \in \{-1,+1\}$ for all i) so that

$$|\mathcal{G}| \le 2^k. \tag{4.15}$$

It follows from (4.15) by the pigeon hole principle that there is at least one $G_{2^{k+1}-1}$ such that

$$\phi(\lambda,T,G_{2^{k+1}-1}) \ge \frac{|\tau_k|}{|G|} \ge \frac{|\tau_k|}{2^k}$$

whence, by (4.6),

$$\phi(\lambda,T,G_{2^{k+1}-1})-\frac{|\tau_k|}{2^{2^{k+1}-1}}\ge\frac{|\tau_k|}{2^k}-\frac{|\tau_k|}{2^{2^{k+1}-1}}=\frac{2^{2^{k+1}-1-k}-1}{2^{2^{k+1}-1}}\,|\tau_k|$$
$$=\frac{2^{2^{k+1}-1-k}-1}{2^{2^{k+1}-1}}\left(2^{K-k+1}-1\right)$$

which proves (4.4).

As we mentioned earlier, we conjecture that for truly random binary function f defined on a fixed r-almost s-uniform tree the normality measure of it is small (more precisely, (4.3) holds) but we have not been able to prove this. Thus one might like to present at least one example for a binary function with small normality measure. In case of binary sequences the Legendre symbol construction (3.18) is known to possess the best pseudorandom properties (and in some other situations also the Legendre symbol provides the best examples). However, as Theorem 4 shows here, somewhat unexpectedly, the most natural Legendre symbol construction defined by (3.18) and (3.21) fails, its normality measure is large. On the other hand, we will present another, slightly more complicated Legendre symbol construction where the normality measure (and the other pseudorandom measures as well) are small:

Theorem 5 *Let p be a prime such that $p>3$ and 2 is primitive root mod p. Define the positive integer K by*

$$2^{K+1}\le p<2^{K+2},\tag{4.16}$$

and let T denote the uniform binary tree of height K. Let c be a quadratic non-residue modulo p, and define the binary function $\rho:\ \mathcal{P}(T)\to\{-1,+1\}$ (where $\mathcal{P}(T)=\{Q_1,Q_2,\dots,Q_N\}$) by

$$\rho\left(Q_n\right)=\left(\frac{n^2-c}{p}\right)\quad\text{for } n=1,2,\dots,N=N(T)=2^{K+1}-1(\le p-1)\tag{4.17}$$

(note that $p\nmid n^2-c$ since c is quadratic non-residue modulo p, thus $\left(\frac{n^2-c}{p}\right)$ is defined). Then we have

$$\overline{N}_k(\rho,T)\le 20\left(2^{k-1}-1\right)p^{1/2}\log p(\ll 2^kN^{1/2}\log N)\tag{4.18}$$

for every $k \in \mathbb{N}$ with $k \leq K$.

Note that it also follows from the results in [12], Definition 4 and Corollary 1 that $\overline{W}(\rho, T)$ and $\tilde{C}(\rho, T)$ are also small. Moreover, we remark that it could be shown that if (4.17) is replaced by

$$\rho\left(Q_n\right) = \left(\frac{n-c}{p}\right)$$

where $|c| = O(1)$ and $\ell \to +\infty$ slowly, then the normality measure of order ℓ is large.

Proof of Theorem 5. We will use the notations of Section 1 and Definition 7. For

$$G_{2^{k+1}-1} = (g_1, g_2, \ldots, g_{2^{k+1}-1}) \in \{-1, +1\}^{2^{k+1}-1},$$

we have to estimate $\phi(\rho, T, G_{2^{k+1}-1})$, i.e., the number of the uniform binary subtrees $T' \in \tau_k$ such that

$$\begin{aligned} E_{2^{k+1}-1}(\rho, T') &= \left(e'_1, e'_2, \ldots, e'_{2^{k+1}-1}\right) = \left(\rho(Q'_1), \rho(Q'_2), \ldots, \rho(Q'_{2^{k+1}-1})\right) \\ &= G_{2^{k+1}-1} = (g_1, g_2, \ldots, g_{2^{k+1}-1}) \end{aligned} \tag{4.19}$$

where $\mathcal{P}(T') = \{Q'_1, Q'_2, \ldots, Q'_{2^{k+1}-1}\}$ is the vertex set of T'. To simplify the notation, write $2^{k+1} - 1 = H$. Then (4.19) holds if and only if

$$e'_t = \rho(Q'_t) = g_t \quad \text{for } 1 \leq t \leq H. \tag{4.20}$$

Now consider a subtree

$$T' \in \tau_k$$

whose root is the vertex Q_n of T so that

$$Q'_1 = Q_n, \tag{4.21}$$

and, by (4.6) here the possible values of n are

$$n \in \{1, 2, \ldots, |\tau_k|\} = \{1, 2, \ldots, 2^{K-k+1} - 1\}. \tag{4.22}$$

Since T and T' are uniform binary trees, it follows from (4.21) by Proposition 3 that the vertices of T' are

$$\begin{aligned}(Q'_1, Q'_2, Q'_3, Q'_4, \dots, Q'_H) &= (Q_n, Q_{2n}, Q_{2n+1}, Q_{4n}, \dots, Q_{2^k n+2^{k-1}}) \\ &= (Q_{p_1(n)}, Q_{p_2(n)}, Q_{p_3(n)}, Q_{p_4(n)}, \dots, Q_{p_H(n)},)\end{aligned}$$

where $p_i(x)$ denotes the i-th polynomial in the sequence $x, 2x, 2x+1, 4x, \dots, 2^k x+2^{k-1}$. Then by (4.17), (4.20) can be rewritten as

$$e_{t'} = \rho(Q'_t) = \rho(Q_{p_t(n)}) = \left(\frac{p_t(n)^2 - c}{p}\right) = g_t \ \text{ for } 1 \le t \le H. \tag{4.23}$$

Clearly for fixed T', i.e., for a fixed n satisfying (4.22) we have

$$\frac{1}{2^H}\prod_{t=1}^{H}(e_{t'}g_t + 1) = \frac{1}{2^H}\prod_{t=1}^{H}\left(\left(\frac{p_t(n)^2 - c}{p}\right)g_t + 1\right) = \begin{cases} 1 & \text{if (4.23) holds,} \\ 0 & \text{otherwise.}\end{cases}$$

Since n may run over the integers in (4.22), thus writing $M = |\tau_k| = 2^{K-k+1} - 1$ we have

$$\begin{aligned}&\left|\phi(\rho, T, G_{2^{k+1}-1}) - \frac{|\tau_k|}{2^{2^{k+1}-1}}\right| = \left|\frac{1}{2^H}\sum_{n=1}^{M}\prod_{t=1}^{H}\left(\left(\frac{p_t(n)^2 - c}{p}\right)g_t + 1\right) - \frac{M}{2^H}\right| \\ &= \left|\frac{1}{2^H}\sum_{\ell=1}^{H}\sum_{1\le t_1<\dots<t_\ell\le H}\sum_{n=1}^{M-1}\left(\frac{p_{t_1}(n)^2 - c}{p}\right)g_{t_1}\dots\left(\frac{p_{t_\ell}(n)^2 - c}{p}\right)g_{t_\ell}\right| \\ &= \left|\frac{1}{2^H}\sum_{\ell=1}^{H}\sum_{1\le t_1<\dots<t_\ell\le H} g_{t_1}\dots g_{t_\ell}\sum_{n=1}^{M-1}\left(\frac{(p_{t_1}(n)^2 - c_1)\dots(p_{t_\ell}(n)^2 - c)}{p}\right)\right| \\ &\le \frac{1}{2^H}\sum_{\ell=1}^{H}\sum_{1\le t_1<\dots<t_\ell\le H}|g_{t_1}\dots g_{t_\ell}|\left|\sum_{n=1}^{M-1}\left(\frac{(p_{t_1}(n)^2 - c_1)\dots(p_{t_\ell}(n)^2 - c)}{p}\right)\right| \\ &= \frac{1}{2^H}\sum_{\ell=1}^{H}\sum_{1\le t_1<\dots<t_\ell\le H}\left|\sum_{n=1}^{M-1}\left(\frac{(p_{t_1}(n)^2 - c_1)\dots(p_{t_\ell}(n)^2 - c)}{p}\right)\right|. \end{aligned}\tag{4.24}$$

Now we will apply Lemma 1 for estimating the innermost sum. To be able to use this lemma with $\chi(n) = \left(\frac{n}{p}\right)$, we have to show that here a typical polynomial

$$f(n) = (p_{t_1}(n)^2 - c_1)\dots(p_{t_\ell}(n)^2 - c) \ \text{ with } 1 \le t_1 < \dots < t_\ell \le H (< p) \tag{4.25}$$

is not of the form $a(g(n))^2$ with $a \in \mathbb{F}_p$, $g(x) \in \mathbb{F}_p[x]$.

The polynomial $n^2 - c$ is irreducible over $\mathbb{F}_p$ since c is a quadratic non-residue modulo p. Thus for $A \neq 0$ a polynomial of form $(An+B)^2 - c$ is also irreducible. It follows that the polynomials $p_i(n)^2 - c$ are irreducible. Then writing $p_i(n)^2 - c$ as constant times a monic polynomial:

$$p_i(n)^2 - c = C_i q_i(n)$$

(where the coefficient of the highest degree term of $q_i(n)$ is 1), the polynomials $q_i(n)$ are irreducible. Now we will show:

Lemma 2 *For $1 \leq i < j \leq H$ the polynomials $q_i(n), q_j(n)$ are different.*

Proof of Lemma 2 Let

$$p_i(n) = 2^u n + v \quad (\text{with } 1 \leq u \leq k < p,\ 0 \leq v < 2^u),$$
$$p_j(n) = 2^y n + z \quad (\text{with } 1 \leq y \leq k < p,\ 0 \leq z < 2^y),$$

so that

$$p_i(n)^2 - c = 2^{2u}\left(n^2 + 2^{1-u}vn + 2^{-2u}(v^2 - c)\right) = 2^{2u} q_i(n)$$

and

$$p_j(n)^2 - c = 2^{2y}\left(n^2 + 2^{1-y}zn + 2^{-2y}(z^2 - c)\right) = 2^{2y} q_j(n)$$

If $u = y$ then, by $i \neq j$, we must have $v \neq z$, so that clearly the coefficients of n are different in $q_i(n)$ and $q_j(n)$, thus $q_i(n)$ and $q_j(n)$ are different.

If $u \neq y$, then we may assume that $u < y$. Write $y - u = s$ (> 0) so that

$$2^s < 2^y \leq 2^k \leq 2^K \leq \frac{1}{2}p. \tag{4.26}$$

Then q_j can be rewritten as

$$q_j(n) = n^2 + 2^{1-u-s}zn + 2^{-2u-2s}(z^2 - c).$$

Assume that contrary to the conclusion of the lemma we have $q_i(n) = q_j(n)$. Then comparing the coefficients of n in $q_i(n)$ and $q_j(n)$ we get

$$2^{1-u}v \equiv 2^{1-u-s}z \pmod{p}, \tag{4.27}$$

$$2^{-2u}(v^2 - c) \equiv 2^{-2u-2s}(z^2 - c) \pmod{p}. \tag{4.28}$$

It follows from (4.27) that

$$z \equiv 2^s v \pmod{p}. \tag{4.29}$$

By (4.28) and (4.29) we have

$$2^{2s}(v^2 - c) \equiv z^2 - c \equiv 2^{2s}v^2 - c \pmod{p}$$

whence

$$2^{2s}c \equiv c \pmod{p},$$

$$2^{2s} \equiv 1 \pmod{p}.$$

Since 2 is a primitive root modulo p, it follows that $p - 1 \mid 2s$ so that by $s > 0$, we have

$$p - 1 \leq 2s$$

$$\frac{p-1}{2} \leq s$$

which contradicts (4.26), and this completes the proof of the lemma.

The polynomial $f(n)$ in (4.25) can be written as

$$f(n) = C_1 C_2 \dots C_\ell q_1(n) q_2(n) \dots q_\ell(n)$$

where $C_1 C_2 \dots C_\ell$ is constant and, by Lemma 2, $q_1(n), q_2(n), \dots, q_\ell(n)$ are different irreducible polynomials. Then clearly, $f(n)$ cannot be of the form $a(g(n))^2$ so that, indeed, we may apply Lemma 1 to estimate the innermost sum in (4.24). Then we get from (4.24) that

$$\left|\phi(\rho, T, G_{2^{k+1}-1}) - \frac{|\tau_k|}{2^{2^{k+1}-1}}\right| \leq \frac{1}{2^H} \sum_{\ell=1}^{H} \sum_{1 \leq t_1 < \dots < t_\ell \leq H} 10(2\ell)p^{1/2} \log p$$

$$< 20Hp^{1/2} \log p = 20(2^{k-1} - 1)p^{1/2} \log p$$

which proves (4.18).

Acknowledgement. We would like to thank the (anonymous) referee for a couple of good questions and for an important reference.

References

[1] R. Ahlswede, C. Mauduit and A. Sárközy, *Large families of pseudorandom sequences of k symbols and their complexity, I, II,* General Theory of Information Transfer and Combinatorics, eds. R. Ahlswede et al., LNCS 4123, Springer, Berlin, 2006, pp. 293-307 and 308-325.

[2] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira and V. Rödl, *Measures of pseudorandomness for finite sequences: minimal values*, Combin. Probab. Comput. 15 (2006), 1-29.

[3] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira and V. Rödl, *Measures of pseudorandomness for finite sequences: typical values*, Proc. London Math. Soc. 95 (2007), 778-812.

[4] G. Bérczi, *On finite pseudorandom binary sequences of k symbols*, Periodica Math. Hungar. 47 (2003), 29-44.

[5] J. Cassaigne, C. Mauduit and A. Sárközy, *On finite pseudorandom binary sequences VII: The measures of pseudorandomness*, Acta Arith. 103 (2002), 97-118.

[6] C. Dartyge, E. Mosaki and A. Sárközy, *On large families of subsets of the set of the integers not exceeding N*, Ramanujan J. 18 (2009), 209-229.

[7] C. Dartyge and A. Sárközy, *Large families of pseudorandom subsets formed by power residues*, Unif. Distrib. Theory 2 (2007), 73-88.

[8] C. Dartyge and A. Sárközy, *On pseudorandom subsets of the set of integers not exceeding N*, Periodica Math. Hungar. 54 (2007), 183-200.

[9] C. Dartyge and A. Sárközy, *On pseudo-random subsets of* $\mathbb{Z}_n$, Monatshefte Math., to appear.

[10] C. Dartyge, A. Sárközy and M. Szalay, *On the pseudo-randomness of subsets related to primitive roots*, Combinatorica, submitted.

[11] H. Furstenberg and B. Weiss, *Markov processes and Ramsey theory for trees*, Special issue on Ramsey theory, Combin. Probab. Comput. 12 (2003), 547-563.

[12] L. Goubin, C. Mauduit and A. Sárközy, *Construction of large families of pseudorandom binary sequences*, J. Number Theory 106 (2004), 56-69.

[13] E. Grant, J. Shalit and T. Stoll, *Bounds for the discrete correlation of infinite sequences on k symbols and generalized Rudin-Shapiro sequences*, Acta Arith., to appear.

[14] K. Gyarmati, *An inequality between the measures of pseudorandomness*, Annales Univ. Sci. Budapest. Eötvös 46 (2003), 157-166.

[15] K. Gyarmati, *On a fast version of a pseudorandom generator*, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, Vol. 4123, Springer, Berlin/Heidelberg, 2006, 326-342.

[16] K. Gyarmati, *On the correlation of binary sequences*, Studia Sci. Math. Hungar. 42 (2005), 59-75.

[17] K. Gyarmati, C. Mauduit and A. Sárközy, *Constructions of pseudorandom binary lattices*, Uniform Distribution Theory, submitted.

[18] K. Gyarmati, A. Sárközy and C. L. Stewart, *On Legendre symbol lattices*, Uniform Distribution Theory, to appear.

[19] P. Hubert, C. Mauduit and A. Sárközy, *On pseudorandom binary lattices*, Acta Arith. 125 (2006), 51-62.

[20] H. Liu, *A large family of pseudorandom binary lattices*, Proc. Amer. Math. Soc. 137 (2009), 793-803.

[21] H. Liu, T. Zhan and X. Wang, *Large families of elliptic curve pseudorandom binary sequences*, Acta Arith., to appear.

[22] H. Liu, T. Zhan and X. Wang, *On the correlation of pseudorandom binary sequences with composite moduli*, Publ. Math. Debrecen 74 (2009), 195-214.

[23] R. Marzouk and A. Winterhof, *On the pseudorandomness of binary and quaternary sequences linked by the Gray mapping*, Periodica Math. Hungar., to appear.

[24] C. Mauduit and A. Sárközy, *Construction of pseudorandom binary lattices by using the multiplicative inverse*, Monatshefte Math. 153 (2008), 217-231.

[25] C. Mauduit and A. Sárközy, *On finite pseudorandom binary sequences I: Measure of pseudorandomness, the Legendre symbol*, Acta Arith. 82 (1997), 365-377.

[26] C. Mauduit and A. Sárközy, *On finite pseudorandom binary sequences of k symbols*, Indag. Mathem. 13 (2002), 89-101.

[27] C. Mauduit and A. Sárközy, *On large families of pseudorandom binary lattices*, Uniform Distribution Theory 2 (2007), 23-37.

[28] C. Mauduit and A. Sárközy, *On the measures of pseudorandomness of binary sequences*, Discrete Math. 271 (2003), 195-207.

[29] L. Mérai, *Construction of pseudorandom binary lattices based on multiplicative characters*, Periodica Math. Hungar., to appear.

[30] L. Mérai, *On finite pseudorandom lattices of k symbols*, Monatshefte Math., submitted.

[31] H. Niederreiter, J. Rivat and A. Sárközy, *Pseudorandom sequences of binary vectors*, Acta Arith. 133 (2008), 109-125.

[32] B. R. Preiss, *Data Structures and Algorithms with Object Oriented Design Patterns in Java*, Wiley, 1999.

[33] W. M. Schmidt, *Equations over Finite Fields:* An Elementary Approach, Lecture Notes in Math. 536, Springer, Berlin, 1976.

[34] J.-P. Serre, *Trees*, Springer Monographs in Math., 2nd ed., Springer, Berlin, 2003.

[35] J. A. Storer, *An Introduction to Data Structures and Algorithms*, Springer, 2002.

[36] A. Weil, *Sur les courbes algébriques et les variétés qui s'en déduisent*, Act. Sci. Ind. 1041, Hermann, Paris, 1948.

Eötvös Loránd University
Department of Algebra and Number Theory
H-1117 Budapest, Pázmány Péter sétány 1/C, Hungary
email: gykati@cs.elte.hu

Laboratorie d'Analyse, Topologie et Probabilités
Faculté des Sciences de Saint Jérôme
Avenue Escadrille Normandie-Niemen
F-13397 Marseille Cedex 20, France
email: hubert@cmi.univ-mrs.fr

Eötvös Loránd University
Department of Algebra and Number Theory
H-1117 Budapest, Pázmány Péter sétány 1/C, Hungary
email: sarkozy@cs.elte.hu