ACTA ARITHMETICA
103.2 (2002)

On finite pseudorandom binary sequences VII: The measures of pseudorandomness

by

JULIEN CASSAIGNE (Marseille), CHRISTIAN MAUDUIT (Marseille)
and ANDRÁS SÁRKÖZY (Budapest)

1. Introduction. In this series we study finite pseudorandom binary sequences $E_N = \{e_1, \ldots, e_N\} \in \{-1, +1\}^N$. In particular, in Part I [MSá] we introduced the following measures of pseudorandomness: Write

$$U(E_N, t, a, b) = \sum_{j=0}^{t-1} e_{a+jb}$$

and, for $D = (d_1, \ldots, d_k)$ with non-negative integers $0 \le d_1 < \ldots < d_k$,

$$V(E_N, M, D) = \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \ldots e_{n+d_k}.$$

Then the *well-distribution measure* of E_N is defined as

$$W(E_N) = \max_{a,b,t} |U(E_N, t, a, b)| = \max_{a,b,t} \Big| \sum_{j=0}^{t-1} e_{a+jb} \Big| \tag{1.1}$$

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

$$C_k(E_N) = \max_{M,D} |V(E_N, M, D)| = \max_{M,D} \Big| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \ldots e_{n+d_k} \Big| \tag{1.2}$$

where the maximum is taken over all $D = (d_1, \ldots, d_k)$ and M such that $M + d_k \le N$.

2000 *Mathematics Subject Classification*: Primary 11K45.

Key words and phrases: pseudorandom, binary sequence, correlation.

Research of András Sárközy partially supported by the Hungarian National Foundation for Scientific Research, Grant No. T029759, MKM fund FKFP-0139/1997 and French–Hungarian APAPE–OMFB exchange program F-5/97. This paper was written while he was visiting the Institut de Mathématiques de Luminy.

In Part I first we discussed several elementary properties of these pseudorandom (briefly: PR) measures. In the second half of Part I and in Parts II–VI, we tested special sequences for pseudorandomness. This testing was based on the principle formulated in [CFMRS1] in the following way: "The sequence E_N" is considered as a "good" PR sequence if these measures $W(E_N)$ and $C_k(E_N)$ (at least for "small" k) are "small".

In this paper our goal is to return to the analysis of the general properties of these PR measures. First in Section 2 we will show that the testing principle quoted above is justified and, indeed, for a truly random $E_N \in \{-1,+1\}^N$ both PR measures $W(E_N)$ and $C_k(E_N)$ are "small". These results inspire a further question which, although less important from a practical point of view, seems to be of independent interest: for fixed N, k, what is the minimum of $W(E_N)$ and $C_k(E_N)$? This problem will be studied in Section 3. Finally, one might like to know whether it suffices to study correlation of order, say, 2, or correlations of higher order must be studied as well. This question can be answered by analyzing the connection between $C_k(E_N)$ and $C_l(E_N)$ for $k \neq l$; this analysis will be carried out in Section 4.

2. The PR measures for random binary sequences. In this section we will estimate $W(E_N)$ and $C_k(E_N)$ for "random" binary sequences $E_N \in \{-1,+1\}^N$, i.e., for choosing each $E_N \in \{-1,+1\}^N$ with probability $1/2^N$. We will show that for a random E_N both $W(E_N)$ and $C_k(E_N)$ are around $\sqrt{N}$:

THEOREM 1. *For all $\varepsilon > 0$ there are numbers $N_0 = N_0(\varepsilon)$ and $\delta = \delta(\varepsilon)$ such that for $N > N_0$ we have*

$$P(W(E_N) > \delta N^{1/2}) > 1 - \varepsilon, \tag{2.1}$$

$$P(W(E_N) > 6(N \log N)^{1/2}) < \varepsilon. \tag{2.2}$$

THEOREM 2. *For all $k \in \mathbb{N}$, $k \geq 2$ and $\varepsilon > 0$ there are numbers $N_0 = N_0(\varepsilon, k)$ and $\delta = \delta(\varepsilon, k)$ such that for $N > N_0$ we have*

$$P(C_k(E_N) > \delta N^{1/2}) > 1 - \varepsilon, \tag{2.3}$$

$$P(C_k(E_N) > 5(kN \log N)^{1/2}) < \varepsilon. \tag{2.4}$$

Thus with probability $> 1 - 2\varepsilon$ we have

$$\delta N^{1/2} < W(E_N) < 6(N \log N)^{1/2}, \tag{2.5}$$

$$\delta N^{1/2} < C_k(E_N) < 5(kN \log N)^{1/2}. \tag{2.6}$$

(2.5) could be improved with some effort but we did not force this since it is good enough for our purpose in this easier form. On the other hand, it seems to be more difficult to improve on (2.6).

Proof of Theorem 1. First we will prove (2.1). Since by (1.1),

$$W(E_N) \geq |U(E_N, N, 1, 1)| = \Big|\sum_{j=1}^{N} e_j\Big|,$$

we have

$$P(W(E_N) > \delta N^{1/2}) \geq P\Big(\Big|\sum_{j=1}^{N} e_j\Big| > \delta N^{1/2}\Big),$$

so that it suffices to prove

$$P\Big(\Big|\sum_{j=1}^{N} e_j\Big| > \delta N^{1/2}\Big) > 1 - \varepsilon. \tag{2.7}$$

If

$$|\{j : 1 \leq j \leq N,\ e_j = -1\}| = h, \tag{2.8}$$

then

$$\begin{aligned}\sum_{j=1}^{N} e_j &= |\{j : 1 \leq j \leq N,\ e_j = +1\}| - |\{j : 1 \leq j \leq N,\ e_j = -1\}| \\ &= (N - h) - h = N - 2h.\end{aligned}$$

(2.8) holds with probability $\frac{1}{2^N}\binom{N}{h}$, so that

$$\begin{aligned}P\Big(\Big|\sum_{j=1}^{N} e_j\Big| > \delta N^{1/2}\Big) &= \sum_{h:\, |N-2h|>\delta N^{1/2}} \frac{1}{2^N}\binom{N}{h} \\ &= \frac{1}{2^N} \sum_{h:\, |h-N/2|>(\delta/2)N^{1/2}} \binom{N}{h}.\end{aligned} \tag{2.9}$$

It is a well known property of the binomial distribution that for all $\varepsilon > 0$ there is an $\eta = \eta(\varepsilon) > 0$ such that

$$\sum_{h:\, |h-N/2|>\eta N^{1/2}} \binom{N}{h} > (1 - \varepsilon)2^N. \tag{2.10}$$

If we now choose $\delta = 2\eta(\varepsilon)$, then (2.7) follows from (2.9) and (2.10), and this completes the proof of (2.1).

Now we prove (2.2). Write $L = 6(N \log N)^{1/2}$. By (1.1) we have

$$\begin{aligned}P(W(E_N) > L) &= P(\max_{a,b,t} |U(E_N, t, a, b)| > L) \\ &\leq \sum_{a,b,t} P(|U(E_N, t, a, b)| > L)\end{aligned} \tag{2.11}$$

where both the maximum and the summation are taken over all $a, b, t \in \mathbb{N}$ such that

$$1 \le a \le a + (t-1)b \le N. \tag{2.12}$$

It follows that

$$a, b, t \in \{1, \ldots, N\}. \tag{2.13}$$

By (2.11) and (2.13) we have

$$P(W(E_N) > L) \le N^3 \max_{a,b,t} P(|U(E_N, t, a, b)| > L)$$

where again the maximum is taken over all a, b, t satisfying (2.12). Thus in order to prove (2.2), it suffices to show that for all a, b, t satisfying (2.12),

$$P(|U(E_N, t, a, b)| > L) = P\Big(\Big|\sum_{j=0}^{t-1} e_{a+jb}\Big| > L\Big) < \varepsilon/N^3. \tag{2.14}$$

If $t \le L$ then the probability in (2.14) is trivially 0 so that we may assume that

$$t > L = 6(N \log N)^{1/2}. \tag{2.15}$$

Write

$$M = 6(t \log t)^{1/2}, \tag{2.16}$$

$$|\{j : 0 \le j \le t-1,\ e_{a+jb} = -1\}| = h. \tag{2.17}$$

Then

$$\begin{aligned}\sum_{j=0}^{t-1} e_{a+jb} &= |\{j : 0 \le j \le t-1,\ e_{a+jb} = +1\}| \\ &\qquad - |\{j : 0 \le j \le t-1,\ e_j = -1\}| \\ &= (t-h) - h = t - 2h.\end{aligned}$$

(2.17) holds with probability $\frac{1}{2^t}\binom{t}{h}$ so that

$$P\Big(\Big|\sum_{j=0}^{t-1} e_{a+jb}\Big| > M\Big) = \sum_{h\colon |t-2h|>M} \frac{1}{2^t}\binom{t}{h} = \frac{1}{2^t} \sum_{h\colon |h-t/2|>M/2} \binom{t}{h}. \tag{2.18}$$

An easy computation shows that if $t \to \infty$ and $k \le t^{2/3}$, then

$$\binom{t}{[t/2]-k} = \binom{t}{[t/2]} \exp\Big(-\frac{2k^2}{t} + O\Big(\frac{k^3}{t^2}\Big)\Big).$$

If we also use the fact that $\binom{t}{i}$ is increasing in i for $0 \le i \le t/2$, it follows easily that for N large enough (so that t is also large by (2.15)),

$$\text{(2.19)}\qquad \sum_{h:\,|h-t/2|>M/2}\binom{t}{h}=\sum_{h:\,|h-t/2|>3(t\log t)^{1/2}}\binom{t}{h}$$
$$<\binom{t}{[t/2]}t\exp\Big(-2(3(t\log t)^{1/2})^2\frac{1}{t}+o(1)\Big)$$
$$=\binom{t}{[t/2]}t\exp(-18\log t+o(1))<\frac{2^t}{t^{16}}.$$

Since $M\le L$, it follows from (2.15), (2.18) and (2.19) that

$$P\Big(\Big|\sum_{j=0}^{t-1}e_{a+jb}\Big|>L\Big)\le P\Big(\Big|\sum_{j=0}^{t-1}e_{a+jb}\Big|>M\Big)$$
$$<\frac{1}{2^t}\cdot\frac{2^t}{t^{16}}=\frac{1}{t^{16}}<\frac{1}{L^{16}}=o\Big(\frac{1}{N^8}\Big)<\frac{\varepsilon}{N^3},$$

which proves (2.14) and this completes the proof of (2.2).

Proof of Theorem 2. First we prove (2.3). Since by (1.2), for $N>2k$,

$$C_k(E_N)\ge|V(E_N,[N/2]-k,(0,1,\dots,k-2,[N/2]))|$$
$$=\Big|\sum_{n=1}^{[N/2]-k}e_ne_{n+1}\dots e_{n+k-2}e_{n+[N/2]}\Big|,$$

we have

$$P(C_k(E_N)>\delta N^{1/2})\ge P\Big(\Big|\sum_{n=1}^{[N/2]-k}e_ne_{n+1}\dots e_{n+k-2}e_{n+[N/2]}\Big|>\delta N^{1/2}\Big)$$

so that it suffices to prove

$$\text{(2.20)}\qquad P\Big(\Big|\sum_{n=1}^{[N/2]-k}e_ne_{n+1}\dots e_{n+k-2}e_{n+[N/2]}\Big|>\delta N^{1/2}\Big)>1-\varepsilon.$$

For any fixed $(k-1)$-tuple $\mathbf{u}=(e_n,e_{n+1},\dots,e_{n+k-2})$, write $f_n=e_ne_{n+1}\dots$ $\dots e_{n+k-2}$, and define g_n by

$$e_{n+[N/2]}=f_ng_n$$

(so that $g_n\in\{-1,+1\}$). Then the sum in (2.20) can be rewritten as

$$\sum_{n=1}^{[N/2]-k}e_ne_{n+1}\dots e_{n+k-2}e_{n+[N/2]}=\sum_{n=1}^{[N/2]-k}g_n.$$

Since $e_{n+[N/2]}$ assumes the values -1 and $+1$, independently of $e_1,e_2,\dots$ $\dots,e_{[N/2]}$, with probability $1/2$, so clearly does g_n. Thus (2.20) can be written in the equivalent form

$$P\Big(\Big|\sum_{n=1}^{[N/2]-k} g_n\Big| > \delta N^{1/2}\Big) > 1-\varepsilon$$

where $g_1, \dots, g_{[N/2]-k}$ are independent and assume the values -1 and $+1$ with probability $1/2$. Writing again

$$|\{n : 1 \le n \le [N/2]-k,\ e_n = -1\}| = h,$$

we deduce in the same way as in the proof of (2.1) that

$$\sum_{n=1}^{[N/2]-k} g_n = [N/2]-k-2h$$

and

$$\begin{aligned}P\Big(\Big|\sum_{n=1}^{[N/2]-k} g_n\Big| > \delta N^{1/2}\Big) &= \sum_{h:\,|[N/2]-k-2h|>\delta N^{1/2}} \frac{1}{2^{[N/2]-k}}\binom{[N/2]-k}{h} \\ &= \sum_{h:\,|(1/2)([N/2]-k)-h|>(\delta/2)N^{1/2}} \frac{1}{2^{[N/2]-k}}\binom{[N/2]-k}{h}.\end{aligned}$$

For fixed k, small enough $\delta = \delta(\varepsilon)$ and $N > N_0(\varepsilon, k)$, this is, indeed, $> 1-\varepsilon$. Since this lower bound is uniform for any choice of $e_1, e_2, \dots, e_{[N/2]}$, (2.20) also holds and this completes the proof of (2.3).

Now we prove (2.4). This will be an easy consequence of an upper bound for the sum

$$S_{N,k}(l) = \sum_{E_N \in \{-1,+1\}^N} \sum_M \sum_D (V(E_N, M, D))^{2l}$$

where the inner sums are taken over all $M \in \mathbb{N}$, $D = (d_1, \dots, d_k)$ with $0 \le d_1 < \dots < d_k$, $M + d_k \le N$, and l will be fixed later in terms of k and N. The sum above can be rewritten as

$$S_{N,k}(l) = \sum_M \sum_D Z(M, D) \tag{2.21}$$

where

$$\begin{aligned}Z(M, D) &= \sum_{E_n \in \{-1,+1\}^N} (V(E_N, M, D))^{2l} \\ &= \sum_{E_n \in \{-1,+1\}^N} \Big(\sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \dots e_{n+d_k}\Big)^{2l}.\end{aligned}$$

If $M \le N^{1/4}$ then clearly

$$Z(M,D) = \sum_{E_n \in \{-1,+1\}^N} (V(E_N,M,D))^{2l} \le \sum_{E_n \in \{-1,+1\}^N} M^{2l} = 2^N M^{2l}. \tag{2.22}$$

Assume now that

$$N^{1/4} < M \le N. \tag{2.23}$$

Write $e_{n+d_1}e_{n+d_2}\dots e_{n+d_k} = x_n$. Then by the multinomial theorem we have

$$Z(M,D) = \sum_{E_N \in \{-1,+1\}^N} \sum_{t=1}^{2l} \sum_{1 \le i_1 < \dots < i_t \le M} \sum_{\substack{j_1+\dots+j_t=2l \\ 1 \le j_1,\dots,j_t}} \frac{(2l)!}{j_1!\dots j_t!} x_{i_1}^{j_1} \dots x_{i_t}^{j_t}.$$

Observe that each $x_i \in \{-1,+1\}$, and thus the value of x_i^j depends only on the parity of j: $x_i^j = 1$ if j is even and $x_i^j = x_i$ if j is odd. Let Z_1 denote the contribution of those terms for which at least one of $j_1,\dots,j_t$ is odd and let Z_2 denote the contribution of the terms such that each of $j_1,\dots,j_t$ is even, so that

$$Z(M,D) = Z_1 + Z_2. \tag{2.24}$$

All the terms in Z_1 can be replaced by a term of the form a constant times $x_{s_1}\dots x_{s_u}$ where $u \le 2l$, $1 \le s_1 < \dots < s_u \le M$. Thus Z_1 can be rewritten in the form

$$Z_1 = \sum_{u \le 2l} \sum_{1 \le s_1 < \dots < s_u \le M} a(s_1,\dots,s_u) \sum_{E_N \in \{-1,+1\}^N} x_{s_1}\dots x_{s_u} \tag{2.25}$$

(where the coefficients $a(s_1,\dots,s_u)$ are non-negative integers independent of E_N). Replace x_{s_i} again by $e_{s_i+d_1}e_{s_i+d_2}\dots e_{s_i+d_k}$ for each of $i=1,\dots,u$; then each term $x_{s_1}\dots x_{s_u}$ becomes of the form

$$x_{s_1}\dots x_{s_u} = e_{s_1+d_1} e_{v_2}^{q_2} e_{v_3}^{q_3} \dots e_{v_z}^{q_z}$$

where $s_1+d_1 < v_2 < \dots < v_z$ and $q_i \in \mathbb{N}$ for $i = 2,3,\dots,z$. Then the innermost sum in (2.25) is

$$\sum_{(e_1,\dots,e_{s_1+d_1-1},e_{s_1+d_1+1},\dots,e_N) \in \{-1,+1\}^{N-1}} e_{v_2}^{q_2}\dots e_{v_z}^{q_z} \sum_{e_{s_1+d_1} \in \{-1,+1\}} e_{s_1+d_1}.$$

Here the inner sum is 0 so that the innermost sum in (2.25) is always 0 and thus

$$Z_1 = 0. \tag{2.26}$$

In Z_2 we may replace each j_i by $2r_i$, and then we may use the fact that the inner sums are independent of E_N:

$$Z_2 = \sum_{E_N \in \{-1,+1\}^N} \sum_{t=1}^{2l} \sum_{1 \le i_1 < \ldots < i_t \le M} \sum_{\substack{r_1 + \ldots + r_t = l \\ 1 \le r_1, \ldots, r_t}} \frac{(2l)!}{(2r_1)! \ldots (2r_t)!}$$
$$= 2^N \sum_{t=1}^{2l} \sum_{1 \le i_1 < \ldots < i_t \le M} \sum_{\substack{r_1 + \ldots + r_t = l \\ 1 \le r_1, \ldots, r_t}} \frac{(2l)!}{(2r_1)! \ldots (2r_t)!}.$$

To compute this sum observe that, by a similar argument,

$$F(y_1, \ldots, y_M) := \sum_{\{f_1, \ldots, f_M\} \in \{-1,+1\}^M} (f_1 y_1 + \ldots + f_M y_M)^{2l}$$
$$= 2^M \sum_{t=1}^{2l} \sum_{1 \le i_1 < \ldots < i_t \le M} \sum_{\substack{r_1 + \ldots + r_t = l \\ 1 \le r_1, \ldots, r_t}} \frac{(2l)!}{(2r_1)! \ldots (2r_t)!} y_{i_1}^{2r_1} \ldots y_{i_t}^{2r_t}.$$

Substituting $y_1 = \ldots = y_M = 1$, we obtain $F(1, \ldots, 1) = 2^{M-N} Z_2$. On the other hand, $F(1, \ldots, 1)$ is easy to compute: if

$$|\{f_i : 1 \le i \le M,\ f_i = -1\}| = h, \tag{2.27}$$

then

$$f_1 + \ldots + f_M = M - 2h,$$

and there are $\binom{M}{h}$ M-tuples satisfying (2.27). Thus

$$2^{M-N} Z_2 = F(1, \ldots, 1) = \sum_{h=0}^{M} \binom{M}{h} (M - 2h)^{2l} = 2 \sum_{h=0}^{[M/2]} \binom{M}{h} (M - 2h)^{2l}.$$

Now we fix the value of l: let

$$l = [2k \log N]. \tag{2.28}$$

Write

$$A_h = \binom{M}{h} (M - 2h)^{2l} \quad \text{so that} \quad 2^{M-N} Z_2 = 2 \sum_{h=0}^{[M/2]} A_h.$$

A little computation shows that for $h < M/2$ we have

$$\frac{A_{h+1}}{A_h} = \frac{M-h}{h+1} \left(1 - \frac{2}{M-2h}\right)^{2l}$$

and clearly this is decreasing on the interval $0 < h \le M/2 - 1$. Thus writing $H = M/2 - \sqrt{lM}$, by (2.23) and (2.28), for $h \le H$ we have

$$\frac{A_{h+1}}{A_h} \ge \frac{M-H}{H+1} \left(1 - \frac{2}{M-2H}\right)^{2l}$$
$$= \frac{M/2 + \sqrt{lM}}{M/2 - \sqrt{lM} + 1} \left(1 - \frac{1}{\sqrt{lM}}\right)^{2l}$$

$$= (1 + (1 + o(1))4\sqrt{l/M})(1 - (1 + o(1))2\sqrt{l/M})$$
$$= (1 + (1 + o(1))2\sqrt{l/M}) > 1.$$

It follows that writing $H_0 = [M/2 - \sqrt{lM} + 1]$, we have $A_0 < A_1 < \ldots < A_{H_0}$, whence

$$2^{M-N} Z_2 = 2 \sum_{h=0}^{[M/2]} A_h = 2\Big(\sum_{h=0}^{H_0} A_h + \sum_{h=H_0+1}^{[M/2]} A_h\Big) \tag{2.29}$$
$$< 2\Big(\sum_{h=0}^{H_0} A_{H_0} + \sum_{h=H_0+1}^{[M/2]} \binom{M}{h}(M-2h)^{2l}\Big)$$
$$< 2\bigg(2H_0 A_{H_0} + (M - 2H_0)^{2l} \sum_{h=0}^{M} \binom{M}{h}\bigg)$$
$$< 2\bigg(M\binom{M}{H_0}(M - 2H_0)^{2l} + (M - 2H_0)^{2l} 2^M\bigg)$$
$$< 2^{M+1}(M+1)\bigg(M - 2\Big(\frac{M}{2} - \sqrt{lM}\Big)\bigg)^{2l}$$
$$< 2^{M+2} M (4lM)^l \quad \text{for } N^{1/4} < M \le N.$$

It follows from (2.21), (2.22), (2.24), (2.26) and (2.29) that

$$S_{N,k}(l) = \sum_D \Big(\sum_{M \le N^{1/4}} Z(M,D) + \sum_{N^{1/4} < M \le N} Z(M,D)\Big) \tag{2.30}$$
$$< \sum_D \Big(\sum_{M \le N^{1/4}} 2^N M^{2l} + \sum_{N^{1/4} < M \le N} 2^{N+2}(4l)^l N^{l+1}\Big)$$
$$< \sum_D \Big(\sum_{M \le N^{1/4}} 2^N N^{l/2} + N^{l+2} 2^{N+2}(4l)^l\Big)$$
$$< 2^N \sum_D (N^{l/2+1/4} + 4N^{l+2}(4l)^l)$$
$$< 5 \cdot 2^N N^{l+2}(4l)^l \sum_D 1.$$

Each d_i in $D = (d_1, \ldots, d_k)$ satisfies $d_i \in \{0, 1, \ldots, N-1\}$ thus it can be chosen in at most N ways so that

$$\sum_D 1 \le N^k. \tag{2.31}$$

It follows from (2.30) and (2.31) that

$$S_{N,k}(l) < 5 \cdot 2^N N^{k+l+2}(4l)^l. \tag{2.32}$$

On the other hand, writing $X = 5(kN\log N)^{1/2}$, we clearly have

$$S_{N,k}(l) = \sum_{E_N\in\{-1,+1\}^N}\sum_M\sum_D (V(E_N,M,D))^{2l} \ge \sum_{E_N\in\{-1,+1\}^N} (\max_{M,D}|V(E_N,M,D)|)^{2l} = \sum_{E_N\in\{-1,+1\}^N} (C_k(E_N))^{2l} \ge X^{2l}|\{E_N : E_N\in\{-1,+1\}^N,\ C_k(E_N) > X\}|. \tag{2.33}$$

It follows from (2.28), (2.32) and (2.33) that

$$\begin{aligned} P(C_k(E_N) > X) &= \frac{1}{2^N}|\{E_N : E_N\in\{-1,+1\}^N,\ C_k(N) > X\}| \\ &\le 5N^{k+l+2}(4l)^l X^{-2l} \\ &= 5N^{k+l+2}(4l)^l(25kN\log N)^{-l} < 5N^{k+2}3^{-l} \\ &= 15N^{k+2}3^{-l-1} < 15N^{2k}3^{-2k\log N} \\ &= 15N^{2k(1-\log 3)} < 15N^{1-\log 3} \end{aligned}$$

and this is $< \varepsilon$ if N is large enough in terms of ε (since $1-\log 3 < 0$), which completes the proof of (2.4).

3. The minimum of the PR measures. Write

$$m(N) = \min_{E_N\in\{-1,+1\}^N} W(E_N), \qquad M_k(N) = \min_{E_N\in\{-1,+1\}^N} C_k(E_N).$$

The estimate of $m(N)$ is a classical problem. In 1964 Roth [Ro] proved that $m(N) > c_1N^{1/4}$ for some positive absolute constant c_1. From the opposite side Erdős, Spencer, Sárközy and Beck estimated $m(N)$, and finally in 1996 Matoušek and Spencer [MSp] showed that $m(N) < c_2N^{1/4}$ so that now the order of magnitude of $m(N)$ is known.

On the other hand, as far as we know $M_k(N)$ has not been studied yet, not even $M_2(N)$ has been estimated. We will prove

THEOREM 3. (i) *For $k, N\in\mathbb{N}, 2\le k\le N$ we have*

$$M_k(N) < 27kN^{1/2}\log N. \tag{3.1}$$

(ii) *For $k\in\mathbb{N}$, $k\ge 2$ there is a number $N_0(k)$ such that if $N\in\mathbb{N}$, $N > N_0$, then also*

$$M_k(N) \le 5(kN\log N)^{1/2}. \tag{3.2}$$

On the other hand, we have only a very weak lower bound for $M_k(N)$ and only in the case when k is even:

Theorem 4. *If $k, N \in \mathbb{N}$, k is even, $2 \le k \le N$, then*

$$M_k(N) \ge \left[\frac{1}{\log 2}(\log N - \log k)\right]. \tag{3.3}$$

Note that if k is odd then there is no lower bound of type (3.3). More exactly we have $M_k(N) = 1$ for all $N \in \mathbb{N}$ and odd k with $1 < k \le N$. Indeed, $M_k(N) \ge 1$ is trivial. To see that also $M_k(N) \le 1$, consider the sequence $E_N = \{e_1, \dots, e_N\} \in \{-1, +1\}^N$ defined by $e_n = (-1)^n$ for $n = 1, \dots, N$. Then for all n and $D = (d_1, \dots, d_k)$ we have

$$\begin{aligned} e_{n+1+d_1} \dots e_{n+1+d_k} &= (-e_{n+d_1}) \dots (-e_{n+d_k}) \\ &= (-1)^k e_{n+d_1} \dots e_{n+d_k} = -e_{n+d_1} \dots e_{n+d_k} \end{aligned}$$

whence, for all M, D,

$$|V(E_N, M, D)| = \left| \sum_{n=1}^{M} e_{n+d_1} \dots e_{n+d_k} \right| = \begin{cases} 0 & \text{if } M \text{ is even,} \\ 1 & \text{if } M \text{ is odd,} \end{cases}$$

so that

$$C_k(E_N) = \max_{M,D} |V(E_N, M, D)| = 1 \quad \text{for } k \text{ odd,}$$

which proves $M_k(N) \le 1$.

We remark that it is easy to see that in this example if k is even, then $C_k(E_N)$ is large:

$$\begin{aligned} C_k(E_N) &= |V(E_N, N-k+1, (0, 1, \dots, k-1))| \\ &= \left| \sum_{n=1}^{N-k+1} (-1)^{n+(n+1)+\dots+(n-k+1)} \right| \\ &= |\pm (N-k+1)| = N-k+1 \quad \text{for } k \text{ even.} \end{aligned}$$

The contrast between the sizes of $C_2(E_N)$ and $C_3(E_N)$ in the example above inspires the following problem that we have not been able to settle:

Problem 1. For $N \to \infty$, are there sequences E_N such that $C_2(E_N) = O(\sqrt{N})$ and $C_3(E_N) = O(1)$ simultaneously?

We think that the upper bounds in Theorem 3 are much closer to the truth than the lower bound in Theorem 4 but, unfortunately, we have not been able to tighten the gap. In particular, we have not been able to settle the following problem:

Problem 2. Is it true that there is a $c > 0$ such that as $N \to \infty$,

$$M_2(N) \gg N^c? \tag{3.4}$$

We think that the answer is affirmative. We will return to this problem at the end of the proof of Theorem 4.

Proof of Theorem 3. (i) Let p denote the smallest prime with $p > N$ so that, by Chebyshev's theorem,

$$N < p \le 2N. \tag{3.5}$$

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

$$e_n = \left(\frac{n}{p}\right) \quad \text{for } n = 1, \dots, N,$$

where $\left(\frac{n}{p}\right)$ denotes the Legendre symbol. By Theorem 1, formula (3.1) in Part I [MSá] for this sequence E_N we have

$$C_k(E_N) \le 9kp^{1/2} \log p$$

whence, by (3.5),

$$C_k(E_N) \le 9k(2N)^{1/2} \log(2N) < 27kN^{1/2} \log N$$

which proves (3.1).

(ii) It follows from (2.4) in Theorem 2 (with, say, $\varepsilon = 1/2$) that for $N > N_0(k)$ there is at least one $E_N \in \{-1, +1\}^N$ with

$$C_k(E_N) \le 5(kN \log N)^{1/2},$$

which proves (3.2).

Proof of Theorem 4. First we remark that (3.3) is always true for $k \ge N/2$ so that we can now suppose

$$k \le N/2. \tag{3.6}$$

Write

$$Q = \left[\frac{1}{\log 2}(\log N - \log k)\right]$$

so that $Q \ge 1$ by (3.6). Let $E_N = \{e_1, \dots, e_N\} \in \{-1, +1\}^N$ and consider the $N - Q + 1$ Q-tuples

$$\mathbf{v}_1 = (e_1, \dots, e_Q),\ \mathbf{v}_2 = (e_2, \dots, e_{Q+1}),\ \dots,\ \mathbf{v}_{N-Q+1} = (e_{N-Q+1}, \dots, e_N).$$

We will show that there are subscripts

$$(1 \le)\ i_1 < \dots < i_k\ (\le N - Q + 1) \tag{3.7}$$

with

$$\mathbf{v}_{i_1} = \dots = \mathbf{v}_{i_k}. \tag{3.8}$$

The number of distinct Q-tuples in $\{-1, +1\}^Q$ is 2^Q. Thus if the number of the vectors $\mathbf{v}_i$ is greater than $(k-1)2^Q$, then by the pigeon-hole principle there is at least one Q-tuple occurring at least k times, so that it suffices to show that

$$N - Q + 1 > (k-1)2^Q. \tag{3.9}$$

By $Q \geq 1$ and the definition of Q we have

$$(k-1)2^Q + Q - 1 < (k-1)2^Q + 2^Q - 1 < k \cdot 2^Q$$
$$= k \exp(Q \log 2) \leq k \exp(\log N - \log k) = N,$$

whence (3.9) follows so that indeed there are $i_1, \dots, i_k$ satisfying (3.7) and (3.8).

Now write $d_j = i_j - 1$ for $j = 1, \dots, k$ and $D = (d_1, \dots, d_k)$, and consider the sum

$$V(E_N, Q, D) = \sum_{n=1}^{Q} e_{n+d_1} e_{n+d_2} \dots e_{n+d_k}.$$

By (3.8) and the definition of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_{N-Q+1}$, for $n = 1, \dots, Q$ we have

$$e_{n+d_1} e_{n+d_2} \dots e_{n+d_k} = e_{i_1+(n-1)} e_{i_2+(n-1)} \dots e_{i_k+(n-1)} = \left(e_{i_1+(n-1)}\right)^k.$$

But $e_{i_1+(n-1)} \in \{-1, +1\}$ and k is even and thus this equals 1. It follows that

$$V(E_N, Q, D) = \sum_{n=1}^{Q} 1 = Q$$

so that, by (1.2),

$$C_k(E_N) \geq |V(E_N, Q, D)| = Q$$

for all $E_N \in \{-1, +1\}^N$, which proves (3.3).

*

We have made a considerable effort to settle Problem 2 above. Even related computer calculations have been carried out. We computed $M_2(83) = 10$ (see also Table 1). However, to gather computer evidence of any value one has to consider much greater values of N, which again seems very difficult.

Of course, it is possible that a simple argument has been overlooked by us and by those who also approached the problem (we asked several people). However, there are certain signs which seem to indicate that, perhaps, the problem is really difficult. E.g., in the case $k = 2$ one might like to improve on the argument used in the proof of Theorem 4 in the following way: again consider the Q-tuples $\mathbf{v}_1, \dots, \mathbf{v}_{N-Q+1}$ defined there but now take Q much greater, say, $Q \sim N^{c'}$ for some $c' > 0$. A similar argument shows that it would suffice to find $i < j$ such that the scalar product of the vectors $\mathbf{v}_i, \mathbf{v}_j$ is "large". This is the same as finding two of the given vectors so that the angle between them is "significantly less" than $\pi/2$. However, this approach fails since "many" vectors can be given with the property that the angle between any two of them is "large"; see [EF] for further details.

Table 1

N	$M_2(N)$
$2 \le N \le 3$	1
$4 \le N \le 6$	2
$7 \le N \le 11$	3
$12 \le N \le 17$	4
$18 \le N \le 26$	5
$27 \le N \le 39$	6
$40 \le N \le 44$	7
$45 \le N \le 55$	8
$56 \le N \le 68$	9
$69 \le N \le 83$	10
$84 \le N \le 93$	10 or 11
$94 \le N \le 106$	between 10 and 12
$107 \le N \le 121$	between 10 and 13
$122 \le N \le 134$	between 10 and 14

Another warning sign is that Problem 1 is related to the classical and very difficult problem on the maximal absolute value of polynomials with -1 and $+1$ coefficients on the unit circle (see, e.g., [Kah, pp. 75–78]). Indeed, the study of the fourth mean of such a polynomial leads to sums of the type occurring in the definition of $C_2(E_N)$.

Our results above inspire a further problem:

PROBLEM 3. What is the connection between $M_2(N)$ and $M_4(N)$? Is it true that for $N > N_0$ we have $M_4(N) > M_2(N)$? Perhaps, even $M_4(N) - M_2(N) \to \infty$ as $N \to \infty$.

4. Comparison of correlations of different orders. First we will show that if $k \in \mathbb{N}$, $l \in \mathbb{N}$, $k \mid l$, $N \to \infty$ and $C_l(E_N)$ is "small", more exactly, $C_l(E_N) = o(N)$, then $C_k(E_N)$ is also small:

THEOREM 5. *For $k, l, N \in \mathbb{N}$, $k \mid l$, $E_N \in \{-1,+1\}^N$ we have*

$$C_k(E_N) \le N\biggl(\frac{(l!)^{k/l}}{k!}\biggl(\frac{C_l(E_N)}{N}\biggr)^{k/l} + \biggl(\frac{l^2}{N}\biggr)^{k/l}\biggr).$$

Next we will show that in the assertion of the first paragraph of this section the condition $k \mid l$ is necessary and, indeed, for any fixed k and for $N \to \infty$ there is an $E_N \in \{-1,+1\}^N$ such that $C_l(E_N)$ is small when $k \nmid l$, whereas $C_k(E_N)$ is large ($\gg N$):

THEOREM 6. *If $k, N \in \mathbb{N}$, and $k \le N$, then there is a sequence $E_N \in \{-1,+1\}^N$ such that if $l \in \mathbb{N}$, $l \le N/2$, then*

$$C_l(E_N) > (N-l)/k - 54k^2N^{1/2}\log N \quad \text{if } k \mid l, \tag{4.1}$$

$$C_l(E_N) < 27k^2 l N^{1/2}\log N \quad \text{if } k \nmid l. \tag{4.2}$$

Proof of Theorem 5. By (1.2), it suffices to prove that for all M and $D=(d_1,\dots,d_k)$ (with $0\le d_1<\dots<d_k,\ M+d_k\le N$) we have

$$|V(E_N,M,D)|\le\left(\frac{(l!)^{k/l}}{k!}\left(\frac{C_l(E_N)}{N}\right)^{k/l}+\left(\frac{l^2}{N}\right)^{k/l}\right). \tag{4.3}$$

Write $l/k=t$ so that $t\in\mathbb{N}$ because $k\,|\,l$. Then clearly,

$$\begin{aligned} V(E_N,M,D)^t &= \Big(\sum_{n_1=1}^{M} e_{n_1+d_1}\dots e_{n_1+d_k}\Big)\dots\Big(\sum_{n_t=1}^{M} e_{n_t+d_1}\dots e_{n_t+d_k}\Big)\\ &= \sum_{n_1=1}^{M}\dots\sum_{n_t=1}^{M} e_{n_1+d_1}\dots e_{n_1+d_k}\dots e_{n_t+d_1}\dots e_{n_t+d_k}\\ &= S_1+S_2, \end{aligned} \tag{4.4}$$

where S_1 denotes the contribution of those terms $e_{n_1+d_1}\dots e_{n_t+d_k}$ where there are two equal subscripts:

$$n_i+d_u=n_j+d_v, \tag{4.5}$$

while in S_2 all the subscripts are distinct.

First we estimate S_1. In (4.5), u and v can be chosen in at most k ways, i,j in t ways, n_j (for fixed j) in M ways, and u,v,n_j determine n_i uniquely. Each of the $t-2$ remaining n_h's can be chosen in at most M ways, so that S_1 has at most $k^2t^2M\cdot M^{t-2}=l^2M^{t-1}$ terms and thus

$$|S_1|\le l^2M^{t-1}. \tag{4.6}$$

Now we estimate S_2. Consider each of the terms $e_{n_1+d_1}\dots e_{n_t+d_k}$ in S_2, and rearrange the order of the factors $e_{n_i+d_u}$ so that the subscripts should be increasing:

$$e_{n_1+d_1}\dots e_{n_t+d_k}=e_{i_1}\dots e_{i_l},\qquad i_1<\dots<i_l.$$

Now we t-colour these factors $e_{i_1},\dots,e_{i_l}$: if the subscript of e_{i_u} is of the form $i_u=n_j+d_v$, then we colour e_{i_u} by the jth colour. Then to each term $e_{i_1}\dots e_{i_l}$ we may assign the sequence of the colours following each other in the order used to colour $e_{i_1},\dots,e_{i_l}$. In this way we get colour patterns of length l where each of the t colours occurs k times, so that the number of these colour patterns is $l!/(k!)^t$.

Now fix any of the colour patterns, and consider each of the terms $e_{i_1}\dots e_{i_l}$ with this fixed colour pattern. We define an equivalence relation among these terms: we say that

$$e_{i_1}\dots e_{i_l}\sim e_{j_1}\dots e_{j_l}\quad\text{if}\quad j_1-i_1=\dots=j_l-i_l.$$

Clearly, this is indeed an equivalence relation. Now fix a colour pattern and an equivalence class, and collect all the terms from this class. Let $e_{h_1}\dots e_{h_l}$

$(h_1 < \ldots < h_l)$ be the term for which the first subscript is minimal; it is easy to see that $h_1 = 1 + d_1$. Write

$$h_i - 1 = f_i \quad \text{for } i = 1, \ldots, l,$$

and let Q denote the number of terms in the given equivalence class. Then it is easy to see that the terms in this equivalence class are $e_{n+f_1} \ldots e_{n+f_l}$ with $n = 1, \ldots, Q$ so that, by (1.2), the absolute value of the sum of the terms in this class is

$$\Big|\sum_{n=1}^{Q} e_{n+f_1} \ldots e_{n+f_l}\Big| = |V(E_N, Q, (f_1, \ldots, f_l))| \le C_l(E_N).$$

It remains to estimate the number of equivalence classes. An equivalence class is uniquely determined by the colour pattern, which can be chosen in $l!/(k!)^t$ ways, and by the subscripts of the t e_{h_i}'s where these colours first appear. The first of these subscripts, $h_1 = 1 + d_1$, is fixed, while each of the other $t-1$ subscripts can be chosen in at most M ways. Thus the number of equivalence classes is $\le (l!/(k!)^t)M^{t-1}$, and thus the total sum is

$$|S_2| \le \frac{l!}{(k!)^t} M^{t-1} C_l(E_N). \tag{4.7}$$

It follows from (4.4), (4.6) and (4.7) that

$$\begin{aligned}
|V(E_N, M, D)| &\le |S_1 + S_2|^{1/t} \le (|S_1| + |S_2|)^{1/t} \\
&\le \Big(l^2 M^{t-1} + \frac{l!}{(k!)^t} M^{t-1} C_l(E_N)\Big)^{1/t} \le \Big(l^2 N^{t-1} + \frac{l!}{(k!)^t} N^{t-1} C_l(E_N)\Big)^{1/t} \\
&= N\Big(\frac{l^2}{N} + \frac{l!}{(k!)^t} \cdot \frac{C_l(E_N)}{N}\Big)^{1/t} \le N\bigg(\Big(\frac{l^2}{N}\Big)^{1/t} + \frac{(l!)^{1/t}}{k!}\Big(\frac{C_l(E_N)}{N}\Big)^{1/t}\bigg)
\end{aligned}$$

which proves (4.3) and this completes the proof of Theorem 5.

Proof of Theorem 6. We will construct a sequence $E_N \in \{-1, +1\}^N$ with the desired properties. The construction will be based on the following result which was the crucial tool also in [MSá]:

LEMMA 1. *Suppose p is a prime number, F_p denotes the field of residue classes modulo p, $\overline{F}_p$ denotes the algebraic closure of F_p, $f(x) \in F_p[x]$ is a polynomial of degree d which is not of the form $f(x) = b(g(x))^2$ with $b \in F_p$, $g(x) \in F_p[x]$ (in other words, if we factorize f in $\overline{F}_p$: $f(x) = b(x-x_1)^{d_1} \ldots (x-x_s)^{d_s}$, where $x_i \ne x_j$ for $i \ne j$, then there is at least one odd exponent d_i), X, Y are real numbers with $0 < Y \le p$, $\left(\frac{n}{p}\right)$ denotes the Legendre symbol for $p \nmid n$ and we write $\left(\frac{n}{p}\right) = 0$ for $p \mid n$. Then*

$$\Big|\sum_{X<n\le X+Y} \Big(\frac{f(n)}{p}\Big)\Big| < 9dp^{1/2}\log p.$$

Proof. This is Corollary 1 in [MSá] and, indeed, we derived it from A. Weil's theorem [We].

Now let p denote the smallest prime with $p > N$ so that, by Chebyshev's theorem,

$$N < p \leq 2N, \tag{4.8}$$

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

$$e_n = \begin{cases} \left(\dfrac{n}{p}\right) & \text{for } k \nmid n, \\ \left(\dfrac{(n-1)(n-2)\dots(n-k+1)}{p}\right) & \text{for } k \mid n. \end{cases} \tag{4.9}$$

First we prove (4.1). Assume that $k \mid l$. Then

$$\begin{aligned} C_l(E_N) &= \max_{M,D} |V(E_N, M, D)| \geq |V(E_N, N-l+1, (0, 1, \dots, l-1))| \\ &= \Big|\sum_{n=1}^{N-l+1} e_n e_{n+1} \dots e_{n+l-1}\Big| = \Big|\sum_{r=1}^{k} S(r)\Big| \end{aligned} \tag{4.10}$$

where $S(r)$ is defined by

$$S(r) = \sum_{\substack{1 \leq n \leq N-l+1 \\ n \equiv r \,(\mathrm{mod}\, k)}} e_n e_{n+1} \dots e_{n+l+1}.$$

Consider first the case $r = 1$:

$$\begin{aligned} S(1) &= \sum_{\substack{1 \leq n \leq N-l+1 \\ n \equiv 1 \,(\mathrm{mod}\, k)}} e_n e_{n+1} \dots e_{n+l+1} \\ &= \sum_{\substack{1 \leq n \leq N-l+1 \\ n \equiv 1 \,(\mathrm{mod}\, k)}} \prod_{\substack{n+k-1 \leq m \leq n+l-1 \\ k \mid m}} e_{m-k+1} e_{m-k+2} \dots e_m. \end{aligned}$$

For $k \mid m$, $1 \leq m \leq N < p$ we have

$$\begin{aligned} & e_{m-k+1} e_{m-k+2} \dots e_m \\ &= \left(\frac{m-k+1}{p}\right)\left(\frac{m-k+2}{p}\right) \dots \left(\frac{m-1}{p}\right) \\ &\quad \times \left(\frac{(m-1)(m-2)\dots(m-k+1)}{p}\right) \\ &= \left(\frac{(m-1)^2(m-2)^2\dots(m-k+1)^2}{p}\right) = +1 \quad \text{for } k \mid m \end{aligned} \tag{4.11}$$

so that

$$S(1)=\sum_{\substack{1\le n\le N-l+1\\ n\equiv 1\ (\mathrm{mod}\, k)}} 1=\left[\frac{N-l}{k}\right]+1>\frac{N-l}{k}. \tag{4.12}$$

Consider now the case $2 \le r \le k$. Write each n with $1 \le n \le N-l+1$, $n \equiv r \pmod k$ in the form $n = uk + r$ so that $S(r)$ can be rewritten as

$$S(r)=\sum_{0\le u\le (N-l+1-r)/k} e_{uk+r}e_{uk+r+1}\dots e_{uk+r+l-1}.$$

By (4.11), the product of the e_i's with $(u+1)k < i \le (u+l/k)k$ in the term $e_{uk+r}e_{uk+r+1}\dots e_{uk+r+l-1}$ is 1 so that these e_i's can be dropped. By using the definition of e_n we get

$$\begin{aligned}S(r)&=\sum_{0\le u\le (N-l+1-r)/k} e_{uk+r}\dots e_{(u+1)k}e_{uk+l+1}\dots e_{uk+r+l-1}\\ &=\sum_{0\le u\le (N-l+1-r)/k}\left(\frac{uk+r}{p}\right)\dots\left(\frac{((u+1)k-k+1)\dots((u+1)k-1)}{p}\right)\\ &\qquad\times\left(\frac{uk+l+1}{p}\right)\dots\left(\frac{uk+r+l-1}{p}\right).\end{aligned}$$

Using the multiplicativity of the Legendre symbol, and then dropping the square factors in the "numerator" of the Legendre symbol, we can rewrite the last sum as

$$S(r)=\sum_{0\le u\le (N-l+1-r)/k}\left(\frac{(uk+1)\dots(uk+r-1)(uk+l+1)\dots(uk+l+r-1)}{p}\right).$$

Since $k \le N < p$, there is an integer $\overline{k}$ with

$$k\overline{k}\equiv 1 \pmod p. \tag{4.13}$$

Then multiplying the sum above by $\left(\frac{\overline{k}^{\,2r-2}}{p}\right)=1$ we get

$$S(r)=\sum_{0\le u\le (N-l+1-r)/k}\left(\frac{f(u)}{p}\right)$$

where

$$f(u)=(u+\overline{k})(u+2\overline{k})\dots(u+(r-1)\overline{k})(u+(l+1)\overline{k})\dots(u+(l+r-1)\overline{k}).$$

As $l+r-1 < l+k \le 2l \le N < p$, here all the zeros $-\overline{k}, -2\overline{k}, \dots, -(l+r-1)\overline{k}$ are distinct modulo p, thus the polynomial f satisfies the conditions in Lemma 1 so that the lemma can be applied to estimate this sum $S(r)$.

The degree of $f(u)$ is $d = 2(r-1) < 2k$ so that by Lemma 1 and (4.8),

$$|S(r)| < 9 \cdot 2kp^{1/2} \log p \le 18k(2N)^{1/2} \log 2N < 54kN^{1/2} \log N \quad \text{for } 2 \le r \le k. \tag{4.14}$$

It follows from (4.10), (4.12) and (4.14) that

$$C_l(E_N) \ge S(1) - \sum_{r=2}^{k} |S(r)| > \frac{N-l}{k} - 54k^2 N^{1/2} \log N,$$

which completes the proof of (4.1).

(4.2) can be proved similarly. Assume that $k \nmid l$. By (1.2) it suffices to prove that for all $M \in \mathbb{N}$, $D = (d_1, \dots, d_l)$ with $0 \le d_1 < \dots < d_l$, $M + d_l \le N$ we have

$$|V(E_N, M, D)| = \Big| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \dots e_{n+d_l} \Big| < 27k^2 l N^{1/2} \log N. \tag{4.15}$$

As in the proof of (4.1), we write

$$V(E_N, M, D) = \sum_{r=1}^{k} S(r) \tag{4.16}$$

where

$$S(r) = \sum_{\substack{1 \le n \le M \\ n \equiv r \pmod{k}}} e_{n+d_1} e_{n+d_2} \dots e_{n+d_l}.$$

Again we substitute $n = uk + r$ so that

$$S(r) = \sum_{0 \le u \le (M-r)/k} e_{uk+r+d_1} e_{uk+r+d_2} \dots e_{uk+r+d_l} \tag{4.17}$$

where, by (4.9),

$$e_{uk+r+d_i} = \left(\frac{uk + r + d_i}{p} \right) \quad \text{for } k \nmid (r + d_i) \tag{4.18}$$

and

$$e_{uk+r+d_i} = \left(\frac{(uk + r + d_i - 1)(uk + r + d_i - 2) \dots (uk + r + d_i - k + 1)}{p} \right) \quad \text{for } k \mid (r + d_i). \tag{4.19}$$

Using again the multiplicativity of the Legendre symbol, we can write the general term in the sum (4.17) as a single Legendre symbol whose "numerator" is a polynomial in u which is the product of linear polynomials of the form $uk + a_z$ where, clearly, $1 \le a_z < p$ so that for distinct values of a_z these linear polynomials are also distinct modulo p. Clearly, each of these linear polynomials occurs at most twice, and $uk + a_z$ occurs twice if and only if it occurs in Legendre symbols of both forms (4.18) and (4.19), i.e., there are i, j, h with $1 \le i, j \le l$,

$$r + d_i = a_z, \quad k \nmid (r + d_i) \tag{4.20}$$

and

$$r + d_j - h = a_z, \quad k \mid (r + d_j), \quad 1 \le h \le k - 1. \tag{4.21}$$

Whenever this is the case, i.e., the factor $uk + a_z$ occurs twice, we drop the factor $(uk + a_z)^2$ in the "numerator" of the Legendre symbol in question. Doing this with all the factors occurring twice, we finally arrive at a representation of the form

$$e_{uk+r+d_1} e_{uk+r+d_2} \dots e_{uk+r+d_l} = \left(\frac{g(u)}{p} \right) \tag{4.22}$$

where either

$$g(u) = 1 \tag{4.23}$$

or $g(u)$ is a product of linear polynomials of the form $uk + b_i$ distinct modulo p:

$$g(u) = (uk + b_1)(uk + b_2) \dots (uk + b_y) \tag{4.24}$$

where

$$b_i \not\equiv b_j \pmod p \quad \text{for } 1 \le i < j \le y. \tag{4.25}$$

Note that the degree of $g(u)$ is at most the sum of the degrees of the polynomials in the "numerators" of the Legendre symbols corresponding to the numbers e_{uk+r+d_i} in the sense (4.18) and (4.19). Since the degree of each of these polynomials is at most k, and i may assume l distinct values, the degree of $g(u)$ is

$$\deg g(u) \le kl. \tag{4.26}$$

Now we show that it follows from the assumption $k \nmid l$ that case (4.23) cannot occur. We argue by contradiction: assume that (4.23) holds, i.e., each $uk + a_z$ mentioned above occurs twice so that each a_z can be represented in both forms (4.20) and (4.21). Fix a quadruple i, j, h and z satisfying (4.20) and (4.21). Then for all h' with $1 \le h' \le k - 1$, write $a_{z'} = r + d_j - h'$. For each of these numbers $a_{z'}$ the factor $uk + a_{z'}$ appears in the Legendre symbol corresponding to e_{uk+r+d_j} in sense (4.19). Since by our indirect assumption each $uk + a_{z'}$ occurs twice, each of the $a_{z'}$ must also have a representation in the form (4.20) so that each of the numbers $d_j - h'$ is a d_i, i.e., each of

$$e_{uk+r+d_j-1}, \quad e_{uk+r+d_j-2}, \quad \dots, \quad e_{uk+r+d_j-k+1} \tag{4.27}$$

occurs in the product in (4.22). Thus dropping all these factors $uk + a_z$, $uk + a_{z'}$ (the latter corresponding to the fixed a_z) means to eliminate the contribution ($= 1$) of the k e's in (4.27).

If not all the factors $uk + a_z$ have been dropped yet, then we may repeat this procedure again by dropping $k-1$ distinct factors $uk + a_z$ each occurring twice, and corresponding to the contribution of k further factors e_{uk+r+d_i} in (4.22). Repeating this procedure again and again, finally by our assumption (4.23) we drop all the factors $uk + a_z$. In each step we consider the contribution of k further factors e_{uk+r+d_i} in (4.22) so that the total number l of these factors must be an integer multiple of k. But this contradicts our assumption $k \nmid l$ and this contradiction proves that (4.23) cannot hold so that $g(u)$ must be of the form (4.24).

By (4.17) and (4.22), $S(r)$ can be rewritten as

$$S(r) = \sum_{0 \le u \le (M-r)/k} \left(\frac{g(u)}{p} \right)$$

where $g(u)$ is of the form (4.24). Defining $\overline{k}$ again by (4.13), we may write this sum as

$$S(r) = \left(\frac{k^y}{p} \right) \sum_{0 \le u \le (M-r)/k} \left(\frac{f(u)}{p} \right) \tag{4.28}$$

with

$$f(u) = (u + b_1\overline{k})(u + b_2\overline{k}) \ldots (u + b_y\overline{k}) \tag{4.29}$$

where

$$b_i\overline{k} \not\equiv b_j\overline{k} \pmod{p} \quad \text{for } 1 \le i < j \le y \tag{4.30}$$

by (4.25), and

$$\deg f(u) = y = \deg g(u) \le kl \tag{4.31}$$

by (4.26). By (4.29) and (4.30), we may apply Lemma 1 to estimate the sum in (4.28). By (4.8) and (4.31) we get

$$\begin{aligned} |S(r)| &= \left| \sum_{0 \le u \le (M-r)/k} \left(\frac{f(u)}{p} \right) \right| < 9klp^{1/2} \log p \\ &\le 9kl(2N)^{1/2} \log(2N) < 27klN^{1/2} \log N \quad \text{for } r = 1, \ldots, k. \end{aligned} \tag{4.32}$$

(4.15) follows from (4.16) and (4.32), and this completes the proof of Theorem 6.

5. Remarks to earlier papers of ours. Finally, we would like to make two remarks concerning our earlier papers [CFMRS1] and [CFMRS2]. In these two papers we studied the pseudorandom properties of the Liouville function $\lambda(n) = (-1)^{\Omega(n)}$ (where $\Omega(n)$ denotes the number of prime factors of n counted with multiplicity). Write $L_N = \{\lambda(1), \ldots, \lambda(N)\}$. In particular, in [CFMRS1] we showed that assuming the generalized Riemann hypothesis, we have

$$W(L_N) < N^{5/6+\varepsilon}. \tag{5.1}$$

It has been pointed out to us by Dr. Louis Goubin (Bull. PTS) that if one replaces the second half of our Lemma 1 there by a reference to a more recent result of Baker and Harman [BH], the exponent 5/6 in (5.1) can be improved to 3/4. We would like to thank Dr. Goubin for this comment.

Secondly, in [CFMRS2] we studied the behaviour of the Liouville function over polynomials and, in particular, over quadratic polynomials. We have learned recently that I. Kátai [Kát] had also studied the λ function over quadratic polynomials but his results are different from ours.

References

[BH] R. C. Baker and G. Harman, *Exponential sums formed with the Möbius function*, J. London Math. Soc. 43 (1991), 193–198.

[CFMRS1] J. Cassaigne, S. Ferenczi, C. Mauduit, J. Rivat and A. Sárközy, *On finite pseudorandom binary sequences III: The Liouville function, I*, Acta Arith. 87 (1999), 367–390.

[CFMRS2] —, —, —, —, —, *On finite pseudorandom binary sequences IV: The Liouville function, II*, ibid. 95 (2000), 343–359.

[EF] P. Erdős and Z. Füredi, *The greatest angle among n points in the d-dimensional Euclidean space*, in: Combinatorial Mathematics (Marseille-Luminy, 1981), North-Holland Math. Stud. 75, North-Holland, Amsterdam, 1983, 275–283.

[Kah] J.-P. Kahane, *Some Random Series of Functions*, 2nd ed., Cambridge Stud. Adv. Math. 5, Cambridge Univ. Press, Cambridge, 1985.

[Kát] I. Kátai, *Research problems in number theory II*, Ann. Univ. Sci. Budapest Sect. Comput. 16 (1996), 223–251.

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

[MSp] J. Matoušek and J. Spencer, *Discrepancy in arithmetic progressions*, J. Amer. Math. Soc. 9 (1996), 195–204.

[Ro] K. F. Roth, *Remark concerning integer sequences*, Acta Arith. 9 (1964), 257–260.

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

Institut de Mathématiques de Luminy
CNRS UPR 9016, FRUMAM
Université de la Méditerranée
163 av. de Luminy, Case 907
F-13288 Marseille Cedex 9, France
E-mail: cassaigne@iml.univ-mrs.fr
mauduit@iml.univ-mrs.fr

Department of Algebra and Number Theory
Eötvös Loránd University
Kecskeméti u. 10-12
H-1053 Budapest, Hungary
E-mail: sarkozy@cs.elte.hu

Received on 20.10.2000
and in revised form on 18.9.2001 (3912)