ACTA ARITHMETICA
LXXXVII.4 (1999)

On finite pseudorandom binary sequences III: The Liouville function, I

by

Julien Cassaigne (Marseille), Sébastien Ferenczi (Marseille),
Christian Mauduit (Marseille),
Jöel Rivat (Lyon) and András Sárközy (Budapest)

1. Introduction. In this series, we are constructing and testing finite pseudorandom (briefly, PR) sequences. In [MS1] we proposed the use of the following measures of pseudorandomness:

For a binary sequence

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

write

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

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

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

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

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

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

1991 *Mathematics Subject Classification*: Primary 11K45.

Research of A. Sárközy partially supported by Hungarian National Foundation for Scientific Research, Grant No. T017433 and MKM fund FKFP-0139/1997. This paper was written while he was visiting the Institut de Mathématiques de Luminy.

defined as

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

where the maximum is taken over all $D = (d_1, \dots, d_k)$ and M such that $M + d_k \le N$. 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". Motivation and background of these definitions was given in [MS1] and [MS2].

In Part I [MS1] of this series we showed that if p is a prime number, $N = p - 1$, and the sequence $E_N = \{e_1, \dots, e_N\}$ is defined 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), then E_N is a "good" PR sequence and, indeed,

$$W(E_N) \ll N^{1/2} \log N \quad \text{and} \quad C_k(E_N) \ll k N^{1/2} \log N.$$

One may guess that, perhaps, this result can be extended and generalized in the following way: if $f(n)$ is a completely multiplicative function such that $f(n) = -1$ or $+1$ and $f(p) = -1$ often enough in every residue class $r \pmod q$, $(r, q) = 1$, then $\{f(1), \dots, f(n)\}$ is a "good" PR sequence. The most important special case is when $f(n)$ is the Liouville function. Let $\omega(n)$ denote the number of distinct prime factors of n, and let $\Omega(n)$ denote the number of prime factors of n counted with multiplicity. Write $\lambda(n) = (-1)^{\Omega(n)}$ (this is the Liouville function) and $\gamma(n) = (-1)^{\omega(n)}$ so that $\lambda(n)$ is completely multiplicative and $\gamma(n)$ is multiplicative, and let

$$L_N = \{l_1, \dots, l_N\} = \{\lambda(1), \dots, \lambda(N)\}$$

and

$$G_N = \{g_1, \dots, g_N\} = \{\gamma(1), \dots, \gamma(N)\}.$$

Hildebrand [Hi1] writes: "It is natural to expect that the sequence $\gamma(n)$ $(n \ge 1)$ behaves like a random sequence of $\pm$ signs." Indeed, first in Sections 2 and 3 we will study the PR properties of the sequence L_N. We will show that the well-distribution measure of the sequence is small (depending on the Riemann hypothesis). On the other hand, only very weak estimates can be given for the correlation of the sequence; in Section 3 we improve slightly on the earlier results of this type. Since the estimate of the correlation is so difficult, we provide partial results in three directions: first in Sections 4 and 5 we study the "truncated" Liouville function. Secondly, we study a PR property which is weaker than the small correlation but it points to the same direction: namely, we study the complexity of the given

sequences. More exactly, in Section 6 we study the connection between correlation and complexity while in Section 7 we estimate the complexity of the sequence L_N under a certain hypothesis. In Part II we will compare the complexities of the "truncated" λ and γ functions (unconditionally); we will formulate a conjecture on the structure of the sequence $\{\lambda(1), \lambda(2), \ldots\}$ and we will prove special cases of it; we will pose several unsolved problems and conjectures; finally, we will present numerical data obtained by computers.

2. The well-distribution measure for the Liouville function. In this section we prove the following theorem:

THEOREM 1. (i) *For any real number $A > 0$ and for $N > N_0(A)$ we have*

$$W(L_N) < N(\log N)^{-A}.$$

(ii) *Under the generalized Riemann hypothesis, for $\varepsilon > 0$ and $N > N_1(\varepsilon)$ we have*

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

P r o o f. Write

$$F_x(\alpha) = \sum_{n \le x} \lambda(n) e(n\alpha).$$

The proof will be based on the following lemma:

LEMMA 1. (i) *For any real number $H > 0$, for $x > x_0(H)$ we have*

$$|F_x(\alpha)| < x(\log x)^{-H} \quad \text{for all } 0 \le \alpha \le 1.$$

(ii) *Under the generalized Riemann hypothesis, for $\varepsilon > 0$ and $x > x_1(\varepsilon)$ we have*

$$|F_x(\alpha)| < x^{5/6+\varepsilon} \quad \text{for all } 0 \le \alpha \le 1.$$

Indeed, this is Lemma 2 of [Sa].

By Lemma 1, for large enough x we have

$$|F_M(\alpha)| < x(\log x)^{-H} \tag{2.1}$$

unconditionally and, under the generalized Riemann hypothesis,

$$|F_M(\alpha)| < x^{5/6+\varepsilon} \tag{2.2}$$

uniformly for $M \le x$ and $0 \le \alpha \le 1$.

Clearly we have

$$\begin{aligned}|U(L_N,t,a,b)| &= \Big|\sum_{j=1}^{t}\lambda(a+jb)\Big| = \Big|\sum_{\substack{a<n\le a+tb\\ n\equiv a \pmod b}}\lambda(n)\Big| \\ &= \Big|\sum_{a<n\le a+tb}\lambda(n)\cdot\frac{1}{b}\sum_{h=1}^{b}e((n-a)h/b)\Big| \\ &= \frac{1}{b}\Big|\sum_{h=1}^{b}e(-ah/b)\sum_{a<n\le a+tb}\lambda(n)e(nh/b)\Big| \\ &= \frac{1}{b}\Big|\sum_{h=1}^{b}e(-ah/b)(F_{a+tb}(h/b)-F_a(h/b))\Big| \\ &\le \frac{1}{b}\sum_{h=1}^{b}(|F_{a+tb}(h/b)|+|F_a(h/b)|),\end{aligned}$$

whence, by using (2.1) and (2.2) with N, $2A$ and $\varepsilon/2$ in place of x, H and ε, respectively,

$$|U(L_N,t,a,b)| < 2N(\log N)^{-2A}$$

unconditionally and, under GRH,

$$|U(L_N,t,a,b)| < 2N^{5/6+\varepsilon/2}.$$

It follows that, for N large enough,

$$W(L_N) = \max_{a,b,t}|U(L_N,t,a,b)| < N(\log N)^{-A}$$

unconditionally and, under GRH,

$$W(L_N) = \max_{a,b,t}|U(L_N,t,a,b)| < N^{5/6+\varepsilon},$$

which completes the proof of Theorem 1.

3. A further remark and the correlation. In Section 2 we showed that if the generalized Riemann hypothesis is true, then the well-distribution measure $W(L_N)$ of the Liouville function is small. The GRH and $W(L_N)$ are so closely connected that if the GRH fails then this fact implies that $W(L_N)$ is "large" for infinitely many values of N. Chowla [Ch, p. 95] writes: "The RH for the ordinary ζ-function is equivalent to

$$L(x) = \sum_{n=1}^{x}\lambda(n) = O(x^{1/2+\varepsilon})$$

where ε is an arbitrary positive number." Littlewood [Li] showed that if the supremum of the real parts of the zeros of the zeta function in the critical

strip is denoted by θ, then for all $\varepsilon > 0$ there are infinitely many $N \in \mathbb{N}$ with

$$\Big|\sum_{n\le N} \mu(n)\Big| > N^{\theta-\varepsilon}.$$

One expects that the same holds with λ in place of μ. This would imply, e.g., that if the RH fails so badly that $\theta = 1$, then for all $\varepsilon > 0$ we have

$$W(L_N) > N^{1-\varepsilon}$$

infinitely often.

While we have a limited control over the well-distribution measure of the Liouville function, the estimate of the correlation measure of it is a hopelessly difficult problem. Numerous papers have been written on the estimate of sums of the form

$$\sum_{n\le x} g_1(n)g_2(n+1)$$

where g_1 and g_2 are multiplicative functions; see [Ell3] and [St1] for references. However, as Hildebrand writes in his review [Hi3] written on Elliott's paper [Ell3]: "For example, in the case when the functions $g_i(n)$ are both equal to the Möbius function $\mu(n)$ or the Liouville function $\lambda(n)$, one would naturally expect that the above sum is of order $o(x)$ when $x \to \infty$, but even the much weaker relation

$$\liminf_{x\to\infty} \frac{1}{x} \sum_{n\le x} \lambda(n)\lambda(n+1) < 1$$

is not known and seems to be beyond reach of the present methods." Indeed, the best known estimates given for the sum above by Graham and Hensley [GH], resp. Harman, Pintz and Wolke [HPW] are

$$-(1+o(1))\frac{1}{3} < \frac{1}{x}\sum_{n\le x} \lambda(n)\lambda(n+1) < 1 - \frac{1}{(\log x)^{7+\varepsilon}} \tag{3.1}$$

for $x > x_0(\varepsilon)$ (the lower bound is a trivial consequence of Theorem 2 in [HPW]). In the case of correlation of order 3 the situation is slightly better: Elliott [Ell2] proved that

$$\limsup_{x\to\infty} \frac{1}{x}\Big|\sum_{n\le x} \lambda(n)\lambda(n+1)\lambda(n+2)\Big| \le \frac{20}{21}. \tag{3.2}$$

The value of the constant on the right hand side of (3.2) has been improved slightly by Ruzsa (unpublished yet). In this section we generalize and slightly improve the lower bound in (3.1) and inequality (3.2):

THEOREM 2. *Let $g(n)$ be a completely multiplicative arithmetic function such that $g(n) \in \{-1, +1\}$ for all $n \in \mathbb{N}$ and $g(2) = -1$. Let $k, d \in \mathbb{N}$. Then*

for all $x \geq 2$ we have

$$\text{(3.3)}\qquad \Big|\sum_{n\leq x} g(n)g(n+d)\dots g(n+2kd)\Big| \leq \begin{cases} \left(1-\dfrac{2}{3(2k+1)}\right)x+O(\log x) & \text{if } d \text{ is even},\\ \left(1-\dfrac{2}{3(k+1)}\right)x+O(\log x) & \text{if } d \text{ is odd}, \end{cases}$$

where the constant factor implied by the $O(\dots)$ notation depends on k and d only (but not on g and x).

In the $g=\lambda$ special case we get

Corollary 1. *For $k,d\in\mathbb{N}$ we have*

$$\Big|\sum_{n\leq x} \lambda(n)\lambda(n+d)\dots\lambda(n+2kd)\Big| \leq \begin{cases} \left(1-\dfrac{2}{3(2k+1)}\right)x+O(\log x) & \text{if } d \text{ is even},\\ \left(1-\dfrac{2}{3(k+1)}\right)x+O(\log x) & \text{if } d \text{ is odd}, \end{cases}$$

and, in particular,

$$\Big|\sum_{n\leq x} \lambda(n)\lambda(n+d)\lambda(n+2d)\Big| \leq \begin{cases} \frac{7}{9}x+O(\log x) & \text{if } d \text{ is even,}\\ \frac{2}{3}x+O(\log x) & \text{if } d \text{ is odd.} \end{cases}$$

Proof (of Theorem 2). Write

$$t(n)=g(n)g(n+d)\dots g(n+2kd),$$

and, for $\varepsilon\in\{-1,+1\}$,

$$T(x,\varepsilon)=|\{n : n\leq x,\ t(n)=\varepsilon\}|$$

so that

$$\sum_{n\leq x} g(n)g(n+d)\dots g(n+2kd)=\sum_{n\leq x} t(n)=T(x,+1)-T(x,-1) = \begin{cases} [x]-2T(x,-1),\\ -[x]+2T(x,+1), \end{cases}$$

whence

$$\text{(3.4)}\qquad \Big|\sum_{n\leq x} g(n)g(n+d)\dots g(n+2kd)\Big| \leq [x]-2\min\{T(x,-1),T(x,+1)\}.$$

Thus it remains to give a lower bound for $T(x,\varepsilon)$ for both $\varepsilon=-1$ and $+1$.

Clearly, for all $n \in \mathbb{N}$ we have

$$t(2n)t(2n+d)\dots t(2n+2kd) = \prod_{j=0}^{2k} t(2n+jd) = \prod_{j=0}^{2k}\prod_{l=0}^{2k} g((2n+jd)+ld) = \prod_{u=0}^{4k} (g(2n+ud))^{\varphi(u)} \tag{3.5}$$

where $\varphi(u)$ denotes the number of pairs (j,l) with

$$j+l=u, \quad j,l \in \{0,1,\dots,2k\}.$$

Since this is

$$\varphi(u) = \varphi(4k-u) = u+1 \quad \text{for } u \in \{0,1,\dots,2k\},$$

we have

$$\varphi(u) \equiv \begin{cases} 0 \pmod 2 & \text{for } u = 1,3,\dots,4k-1, \\ 1 \pmod 2 & \text{for } u = 0,2,\dots,4k. \end{cases}$$

Thus, if we use also $g(2) = -1$ and the complete multiplicativity of $g(n)$, it follows from (3.5) that

$$\begin{aligned} t(2n)t(2n+d)\dots t(2n+2kd) &= g(2n)g(2n+2d)\dots g(2n+4kd) \\ &= (g(2))^{2k+1} g(n)g(n+d)\dots g(n+2kd) \\ &= -t(n). \end{aligned}$$

This clearly implies that for both $\varepsilon = -1$ and $+1$,

(3.6) at least one of $t(n), t(2n), t(2n+d), \dots, t(2n+2kd)$ is ε.

Consider now a number $y \geq 1$, let m denote the greatest positive integer such that

$$2m + 2kd \leq 4y,$$

let $C = C(k,d)$ be a large but fixed number, and write

$$h = \begin{cases} \left[\dfrac{m}{(2k+1)d} - C\right] & \text{if } d \text{ is even}, \\ \left[\dfrac{m}{2(k+1)d} - C\right] & \text{if } d \text{ is odd}. \end{cases}$$

Let us write

$$\mathcal{S}(n) = \{n, 2n, 2n+d, \dots, 2n+2kd\}$$

(for all $n \in \mathbb{N}$) and

$$\mathcal{T}(y) = \begin{cases} \displaystyle\bigcup_{j=1}^{h}\bigcup_{l=1}^{d/2} \mathcal{S}\left(m-(j-1)(2k+1)\frac{d}{2}-(l-1)\right) & \text{if } d \text{ is even}, \\ \displaystyle\bigcup_{j=1}^{h}\bigcup_{l=1}^{d} \mathcal{S}(m-(j-1)(k+1)d-(l-1)) & \text{if } d \text{ is odd}. \end{cases} \tag{3.7}$$

A simple computation shows that if C is large enough in terms of k and d, then for $(j,l) \neq (j',l')$ the sets $\mathcal{S}$ on the right hand side of (3.7) are not overlapping, and all these sets are covered by the interval $(y, 4y]$. Since by (3.6) each of these sets $\mathcal{S}$ contains an integer r with $t(r) = \varepsilon$, it follows that

$$\begin{aligned} T(4y,\varepsilon) - T(y,\varepsilon) &= |\{r : y < r \le 4y,\ t(r) = \varepsilon\}| \\ &\ge \begin{cases} |\{(j,l) : 1 \le j \le h,\ 1 \le l \le d/2\}| \\ \quad = \dfrac{hd}{2} = \dfrac{m}{2(2k+1)} + O(1) = \dfrac{y}{2k+1} + O(1) & \text{if } d \text{ is even,} \\ |\{(j,l) : 1 \le j \le h,\ 1 \le l \le d\}| \\ \quad = hd = \dfrac{m}{2(k+1)} + O(1) = \dfrac{y}{k+1} + O(1) & \text{if } d \text{ is odd.} \end{cases} \end{aligned}$$

Thus we have

$$\begin{aligned} (3.8)\quad T(x,\varepsilon) &= \sum_{4^j \le x} \left(T\left(\frac{x}{4^{j-1}}, \varepsilon\right) - T\left(\frac{x}{4^j}, \varepsilon\right)\right) + O(1) \\ &\ge \begin{cases} \displaystyle\sum_{4^j \le x} \left(\frac{x}{(2k+1)4^j} + O(1)\right) + O(1) & \text{if } d \text{ is even,} \\ \displaystyle\sum_{4^j \le x} \left(\frac{x}{(k+1)4^j} + O(1)\right) + O(1) & \text{if } d \text{ is odd,} \end{cases} \\ &\ge \begin{cases} \dfrac{x}{4(2k+1)} \displaystyle\sum_{l=0}^{\infty} \frac{1}{4^l} + O(\log x) = \frac{x}{3(2k+1)} + O(\log x) \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } d \text{ is even,} \\ \dfrac{x}{4(k+1)} \displaystyle\sum_{l=0}^{\infty} \frac{1}{4^l} + O(\log x) = \frac{x}{3(k+1)} + O(\log x) \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } d \text{ is odd.} \end{cases} \end{aligned}$$

(3.3) follows from (3.4) and (3.8) and this completes the proof of Theorem 2.

THEOREM 3. *Let $g(n)$ be a completely multiplicative arithmetic function such that $g(n) \in \{-1,+1\}$ for all $n \in \mathbb{N}$. Let $k, d \in \mathbb{N}$. Then for all $x \ge 2$ we have*

$$\begin{aligned} &\sum_{n \le x} g(n)g(n+d)\dots g(n+(2k-1)d) \\ &\qquad \ge \begin{cases} -\left(1 - \dfrac{2}{3k}\right)x + O(\log x) & \textit{if } d \textit{ is odd,} \\ -\left(1 - \dfrac{1}{3k}\right)x + O(\log x) & \textit{if } d \textit{ is even,} \end{cases} \end{aligned}$$

where the constant factor implied by the $O(\dots)$ notation depends on k and d only.

Note that the lower bound $-\frac{1}{3}x + O(\log x)$ for $k = 1$, d odd is best possible as the completely multiplicative function $f(n)$ defined by $f(2) = -1$, $f(p) = +1$ for $p > 2$ shows.

In the $g = \lambda$ special case we get

COROLLARY 2. *For $k, d \in \mathbb{N}$ we have*

$$\sum_{n\le x} \lambda(n)\lambda(n+d)\dots\lambda(n+(2k-1)d) \ge \begin{cases} -\left(1-\dfrac{2}{3k}\right)x + O(\log x) & \text{if } d \text{ is odd,} \\ -\left(1-\dfrac{1}{3k}\right)x + O(\log x) & \text{if } d \text{ is even,} \end{cases}$$

and, in particular,

$$\sum_{n\le x} \lambda(n)\lambda(n+d) \ge \begin{cases} -\frac{1}{3}x + O(\log x) & \text{if } d \text{ is odd,} \\ -\frac{2}{3}x + O(\log x) & \text{if } d \text{ is even.} \end{cases}$$

P r o o f (of Theorem 3). Since the proof is similar to that of Theorem 2, we leave some details to the reader.

Again we write

$$t(n) = g(n)g(n+d)\dots g(n+(2k-1)d)$$

and

$$T(x,+1) = |\{n : n \le x,\ t(n) = +1\}|$$

so that

$$\sum_{n\le x} g(n)g(n+d)\dots g(n+(2k-1)d) = \sum_{n\le x} t(n) = -[x] + 2T(x,+1).$$

To give a lower bound for $T(x,+1)$, we use

$$\begin{aligned} t(2n)t(2n+d)\dots t(2n+(2k-1)d) &= g(2n)g(2n+2d)\dots g(2n+4k-2) \\ &= (g(2))^{2k} g(n)g(n+d)\dots g(n+(2k-1)d) = t(n), \end{aligned}$$

so that

(3.9) at least one of $t(n), t(2n), t(2n+d), \dots, t(2n+(2k-1)d)$ is equal to $+1$.

Now for some $y \ge 1$, let m denote the greatest positive integer such that

$$2m + (2k-1)d \le 4y,$$

let C be large enough in terms of k and d, and write

$$h = \left[\frac{m}{2kd} - C\right].$$

Let

$$\mathcal{S}(n) = \{n, 2n, 2n+d, \dots, 2n+(2k-1)d\}$$

(for all $n \in \mathbb{N}$) and

$$\mathcal{T}(y) = \begin{cases} \displaystyle\bigcup_{j=1}^{h}\bigcup_{l=1}^{d} \mathcal{S}(m-(j-1)kd-(l-1)) & \text{if } d \text{ is odd,} \\ \displaystyle\bigcup_{j=1}^{h}\bigcup_{l=1}^{d/2} \mathcal{S}(m-(j-1)kd-(l-1)) & \text{if } d \text{ is even.} \end{cases} \tag{3.10}$$

Again the sets $\mathcal{S}$ in (3.10) are not overlapping, and by (3.9), each of them contains an $r \in \mathbb{N}$ with $t(r) = +1$. Thus

$$\begin{aligned} T(4y,+1) - T(y,+1) &= |\{r : y < r \le 4y,\ t(r) = +1\}| = |\mathcal{T}(y)| \\ &= \begin{cases} hd = \dfrac{m}{2k} + O(1) = \dfrac{y}{k} + O(1) & \text{if } d \text{ is odd,} \\ h\dfrac{d}{2} = \dfrac{m}{4k} + O(1) = \dfrac{y}{2k} + O(1) & \text{if } d \text{ is even.} \end{cases} \end{aligned}$$

The proof can be completed in the same way as the proof of Theorem 2.

4. The well-distribution measure of the truncated Liouville function. Since one cannot control the PR properties of the Liouville function satisfactorily, one might like to look for partial results in other directions; the remaining part of this paper is devoted to results of this type. First we study functions "close" the Liouville function but easier to handle. For $y \le 1$ let $\lambda_y(n)$ and $\gamma_y(n)$ denote the multiplicative functions defined by

$$\lambda_y(p^\alpha) = \begin{cases} (-1)^\alpha\ (= \lambda(p^\alpha)) & \text{for } p \le y, \\ +1 & \text{for } p > y, \end{cases}$$

and

$$\gamma_y(p^\alpha) = \begin{cases} -1\ (= \gamma(p^\alpha)) & \text{for } p \le y, \\ +1 & \text{for } p > y, \end{cases}$$

respectively, and write

$$L_N(y) = \{\lambda_y(1), \dots, \lambda_y(N)\}$$

and

$$G_N(y) = \{\gamma_y(1), \dots, \gamma_y(N)\}.$$

In this paper we restrict ourselves to the sequence $L_N(y)$ since $G_N(y)$ could be handled similarly, and its properties studied here are also similar (on the other hand, in Part II we will also study the sequence $G_N(y)$ since the comparison of a certain other property of the two sequences will show an interesting contrast).

First we prove

THEOREM 4. *There is a positive absolute constant c_1 such that for $3 < y \le N$ we have*

$$W(L_N(y)) < c_1 \frac{N}{(\log\log y)^{1/4}}. \tag{4.1}$$

We remark that the point of this result is the uniformity in y. On the other hand, the upper bound in (4.1) is weak and certainly far from the truth; this is the price paid for the uniformity.

For small values of y (for $y \ll \log N$), this upper bound could easily be improved considerably. This could be done by reducing the problem to the estimate of the sum

$$\sum_{n\le x} \lambda_y(n)\chi(n)$$

(as in the proof of the theorem below), then writing $\lambda_y(n)$ as

$$\lambda_y(n) = \sum_{d|n} h_y(d)$$

where h_y is the Möbius inverse of λ_y and, finally, changing the order of summation over n and d. We leave the details of this to the reader; here we restrict ourselves to the deeper uniform version presented above.

P r o o f (of Theorem 4). If y_0 is large but fixed then (4.1) holds trivially for $3 < y < y_0$ if c_1 is large enough; thus we may assume that y is large.

If $a \in \mathbb{Z}$, $b \in \mathbb{N}$ and we write $d = (a,b)$, $a = da'$, $b = db'$ then

$$\Big|\sum_{j=1}^{t} \lambda_y(a+jb)\Big| = \Big|\lambda_y(d)\sum_{j=1}^{t} \lambda_y(a'+jb')\Big| = \Big|\sum_{j=1}^{t} \lambda_y(a'+jb')\Big|$$

since $\lambda_y(n)$ is completely multiplicative. Here in the last sum we have $(a',b') = 1$ and thus

$$W(L_N(y)) = \max_{a,b,t} \Big|\sum_{j=1}^{t} \lambda_y(a+jb)\Big| = \max_{\substack{a,b,t\\(a,b)=1}} \Big|\sum_{j=1}^{t} \lambda_y(a+jb)\Big|$$

so that we may restrict ourselves to a, b with $(a, b) = 1$. Moreover, clearly we have

$$\Big|\sum_{j=1}^{t} \lambda_y(a+jb)\Big| = \Big| \sum_{\substack{n\le a+tb\\ n\equiv a \,(\mathrm{mod}\, b)}} \lambda_y(n) - \sum_{\substack{n\le a\\ n\equiv a \,(\mathrm{mod}\, b)}} \lambda_y(n)\Big|$$
$$\le \Big| \sum_{\substack{n\le a+tb\\ n\equiv a \,(\mathrm{mod}\, b)}} \lambda_y(n)\Big| + \Big| \sum_{\substack{n\le a\\ n\equiv a \,(\mathrm{mod}\, b)}} \lambda_y(n)\Big|.$$

Thus in order to prove (4.1), it suffices to show that

$$\Big| \sum_{\substack{n\le x\\ n\equiv a \,(\mathrm{mod}\, b)}} \lambda_y(n)\Big| < c_2 \frac{N}{(\log\log y)^{1/4}} \tag{4.2}$$
$$\text{for } y_0 < y \le N,\ x \le N,\ (a,b) = 1.$$

If $x \le \sqrt{N}$ then this is trivial (since the left hand side is $\le x$), while for $\sqrt{N} < x \le N$, $x \le y$ we have $\lambda_y(n) = \lambda(n)$ for all $n \le x$ and thus (4.2) holds by Theorem 1(i). Thus we may assume that

$$y_0 < y \le x \le N.$$

Assume first that

$$b \ge c_3(\log\log y)^{1/4}$$

where c_3 is a positive absolute constant which will be fixed later. Then clearly

$$\Big| \sum_{\substack{n\le x\\ n\equiv a \,(\mathrm{mod}\, b)}} \lambda_y(n)\Big| \le \sum_{\substack{n\le x\\ n\equiv a \,(\mathrm{mod}\, b)}} |\lambda_y(n)| = \sum_{\substack{n\le x\\ n\equiv a \,(\mathrm{mod}\, b)}} 1$$
$$\le \frac{x}{b} + 1 < \frac{2x}{b} \le \frac{2N}{c_3(\log\log y)^{1/4}}$$

so that (4.2) holds trivially in this case.

Assume now that

$$b < c_3(\log\log y)^{1/4}. \tag{4.3}$$

By $(a, b) = 1$ we have

$$\sum_{\substack{n\le x\\ n\equiv a \,(\mathrm{mod}\, b)}} \lambda_y(n) = \frac{1}{\varphi(b)} \sum_{\chi \,(\mathrm{mod}\, b)} \overline{\chi}(a) \sum_{n\le x} \lambda_y(n)\chi(n).$$

Thus writing

$$G_y(x,\chi) = \sum_{n\le x} \lambda_y(n)\chi(n), \tag{4.4}$$

we have

$$(4.5)\quad \Big|\sum_{\substack{n\le x\\ n\equiv a \pmod b}} \lambda_y(n)\Big| \le \frac{1}{\varphi(b)} \sum_{\chi \pmod b} |\overline{\chi(a)} G_y(x,\chi)|$$
$$= \frac{1}{\varphi(b)} \sum_{\chi \pmod b} |G_y(x,\chi)| \le \max_{\chi \pmod b} |G_y(x,\chi)|.$$

It remains to estimate $|G_y(x,\chi)|$ for a character χ mod b. To do this, we will use Halász' [Ha] mean value theorem in the slightly more general form proved by Elliott [Ell1, p. 211]:

LEMMA 2. *Let $g(n)$ be a completely multiplicative function which for some $\eta > 0$ satisfies $g(p) = 0$ or $\eta \le |g(p)| \le 2 - \eta$ for each prime p. When $g(p)$ is non-zero let θ_p denote a value of its argument. Assume that there are numbers θ_0 and $\delta > 0$ so that*

$$(4.6)\qquad |e^{i\theta_p} - e^{i\theta_0}| \ge \delta$$

is always satisfied. Then there are positive numbers c_4 and ν so that the inequality

$$(4.7)\quad \Big|\sum_{n\le x} g(n)\Big| \le c_4 x \exp\Big\{\sum_{p\le x} \frac{|g(p)|-1}{p} - \nu \sum_{p\le x} \frac{|g(p)| - \operatorname{Re} g(p)}{p} + 2\eta \sum_{p\le x,\, g(p)=0} \frac{1}{p}\Big\}$$

holds uniformly for $x \ge 3$. Here c_4^{δ} is bounded in terms of η alone, and $\nu = c_5\delta^3\eta$ for a certain positive absolute constant c_5.

We will use this theorem with

$$(4.8)\qquad g(n) = \lambda_y(n)\chi(n).$$

Then clearly either $\chi(p) = 0$ so that $g(p) = 0$, or $|\chi(p)| = 1$ so that $|g(p)| = 1$. Thus $\eta = 1$ can be chosen in Lemma 2. Moreover, if $g(p) \ne 0$ then

$$(g(p))^{2b} = ((\lambda_y(p))^2)^b((\chi(p))^b)^2 = 1 \cdot 1 = 1$$

so that $g(p)$ is a $2b$th root of unity. Thus choosing $\theta_0 = 2\pi/(4b)$, by the inequality

$$|1 - e^{2\pi i\alpha}| \ge 4\|\alpha\|$$

(where $\|\alpha\|$ denotes the distance from α to the nearest integer: $\|\alpha\| = \min(\{\alpha\}, 1 - \{\alpha\})$) we have

$$|e^{i\theta_p} - e^{i\theta_0}| \ge |1 - e^{2\pi i/(4b)}| \ge 4\Big\|\frac{1}{4b}\Big\| = \frac{1}{b}$$

so that $\delta = 1/b$ can be chosen in Lemma 2. Then there is a positive absolute constant $c_6 > 1$ such that in (4.7) we have $c_4^\delta = c_4^{1/b} < c_6$ whence

$$c_4 < c_6^b. \tag{4.9}$$

Moreover, we have

$$\nu = c_5\delta^3\eta = c_5/b^3. \tag{4.10}$$

Let $\ell_k(x)$ denote the k-fold logarithm of x so that $\ell_k(x) = \log \ell_{k-1}(x)$ for $k = 2, 3, \dots$ Then by (4.3) and (4.10), the exponent on the right hand side of (4.7) can be estimated in the following way:

$$\begin{aligned}
\sum_{p\le x}\frac{|g(p)|-1}{p} - \nu\sum_{p\le x}\frac{|g(p)|-\operatorname{Re} g(p)}{p} + 2\eta\sum_{p\le x,\, g(p)=0}\frac{1}{p}
&\le 0 - \frac{c_5}{b^3}\sum_{p\le x,\, g(p)=-1}\frac{2}{p} + 2\sum_{p\le x,\, g(p)=0}\frac{1}{p}\\
&\le -\frac{c_5}{b^3}\biggl(\sum_{p\le y}\frac{1}{p} - \sum_{p\mid b}\frac{1}{p}\biggr) + 2\sum_{p\mid b}\frac{1}{p}\\
&= -\frac{c_5}{b^3}\sum_{p\le y}\frac{1}{p} + \biggl(\frac{c_5}{b^3}+2\biggr)\sum_{p\mid b}\frac{1}{p}\\
&\le -\frac{c_5}{b^3}(\ell_2(y) - c_7) + c_8\ell_3(2b) < -\frac{c_5\ell_2(y)}{2b^3} + c_9\ell_5(y).
\end{aligned}$$

Thus it follows from (4.3)–(4.5) and (4.7)–(4.9) that

$$\begin{aligned}
&\Bigl|\sum_{\substack{n\le x\\ n\equiv a \,(\mathrm{mod}\, b)}}\lambda_y(n)\Bigr| \\
&\quad\le \max_{\chi \,(\mathrm{mod}\, b)}\Bigl|\sum_{n\le x}\lambda_y(n)\chi(n)\Bigr| \le c_4 x\exp\biggl(-\frac{c_5\ell_2(y)}{2b^3} + c_9\ell_5(y)\biggr)\\
&\quad< x\exp\biggl(b\log c_6 - \frac{c_5\ell_2(y)}{2b^3} + c_9\ell_5(y)\biggr)\\
&\quad< x\exp\biggl(c_3\log c_6(\ell_2(y))^{1/4} - \frac{c_5\ell_2(y)}{2c_3^3(\ell_2(y))^{3/4}} + c_9\ell_5(y)\biggr)\\
&\quad= x\exp\biggl(\biggl(c_3\log c_6 - \frac{c_5}{2c_3^3}\biggr)(\ell_2(y))^{1/4} + c_9\ell_5(y)\biggr).
\end{aligned} \tag{4.11}$$

Now we fix the value of c_3: we define c_3 by

$$c_3\log c_6 = c_5/(4c_3^3)$$

(note that $c_6 > 1$). Then for large y it follows from (4.11) that

$$\Big|\sum_{\substack{n\le x\\ n\equiv a \pmod b}} \lambda_y(n)\Big| < \exp\Big(-\frac{c_5}{4c_3^3}(\ell_2(y))^{1/4} + \frac{c_5}{8c_3^3}(\ell_2(y))^{1/4}\Big)$$

$$= x\exp(-c_{10}(\ell_2(y))^{1/4}) < \frac{N}{(\log\log y)^{1/4}}$$

and this completes the proof of (4.2) and thus also of Theorem 4.

5. The correlation measure of the truncated Liouville function. We restrict ourselves to the study of correlation of order 2; higher order correlations could be studied similarly. We prove

THEOREM 5. *There is a positive absolute constant c_{11} such that if $x \ge 2$,*

$$2 \le y \le (\log x)^2/(\log\log x)^2 \tag{5.1}$$

and b is a positive integer with

$$b \le y, \tag{5.2}$$

then

$$\Big|\sum_{n\le x} \lambda_y(n)\lambda_y(n+b)\Big| < c_{11}\frac{(\log\log(b+1))^4}{(\log y)^4}x. \tag{5.3}$$

P r o o f. We derive the theorem from a result of Stepanauskas [St2] (see also [Ell3] and [St1]). This result is too complicated and technical to present it here in its most general form. Therefore we restrict ourselves to formulating the special case that we need here:

LEMMA 3. *Assume that $g : \mathbb{N} \to \mathbb{C}$ is a multiplicative function, $b \in \mathbb{N}$,*

$$x \ge r \ge 2, \tag{5.4}$$

$$1/2 < \alpha < 1, \tag{5.5}$$

$$b \le r, \tag{5.6}$$

$$\sum_{r<p\le x} (\operatorname{Re} g(p) - 1)/p \le C, \tag{5.7}$$

$$|g(n)| = O(1) \tag{5.8}$$

and

$$S(r,x) := \sum_{r<p\le x+b} |g(p)-1|^2/p \le 1/8. \tag{5.9}$$

Let $h(n)$ denote the Möbius inverse of $g(n)$:

$$h(n) = \sum_{d|n} g(d)\mu(n/d),$$

and write

$$w_p = \sum_{\substack{k,l\ge 0\\ (p^k,p^l)\mid b}} \frac{h(p^k)h(p^l)}{[p^k,p^l]}, \qquad P(x) = \prod_{p\le x} W_p.$$

Then

$$\left|\frac{1}{x}\sum_{n\le x} g(n)g(n+b) - P(x)\right| \ll x^{1-2\alpha}\exp\left(c\frac{r^\alpha}{\log r}\right) + (S(r,x))^{1/2} + \frac{1}{r\log r} + \frac{1}{x^{1/2}\log x} \tag{5.10}$$

where the constant c and the constant implied by the $\ll$ symbol may depend on the constants in (5.7) *and* (5.8) *only.*

Indeed, this is the $s = 2$, $g_1 = g_2 = g$, $a_1 = a_2 = 1$, $b_1 = 0$, $b_2 = b$, $A_1 = A_2 = 0$ special case of the theorem in [St2].

Note that Stepanauskas remarks in [St2] that (5.6) (together with other conditions) could be relaxed considerably. This would lead to a much weaker condition than (5.2) so that we could study long range correlation as well.

To derive Theorem 5 from Lemma 3, we use the lemma with $g(n) = \lambda_y(n)$, $r = y$, and

$$\alpha = \frac{1}{2} + \frac{\log\log\log x}{2\log\log x}. \tag{5.11}$$

Then (5.4)–(5.6) and (5.8) hold trivially, and

$$\sum_{r<p\le x} (\operatorname{Re} g(p) - 1)/p = \sum_{y<p\le x} (\operatorname{Re}\lambda_y(p) - 1)/p = 0$$

and

$$S(r,x) = \sum_{y<p\le x+b} |\lambda_y(p) - 1|^2/p = 0 \tag{5.12}$$

so that (5.7) and (5.9) also hold and thus the lemma can be applied.

Moreover, $h(1) = 1$ and

$$h(p^\alpha) = \lambda_y(p^\alpha) - \lambda_y(p^{\alpha-1}) = \begin{cases} (-1)^\alpha - (-1)^{\alpha-1} = 2(-1)^\alpha & \text{for } p\le y,\\ 1-1=0 & \text{for } p>y,\end{cases}$$

so that

$$w_p = h(1) = 1 \quad \text{for } p > y,$$

and, defining the non-negative integer β_p by $p^{\beta_p} \| b$, we get

$$\begin{aligned} w_p &= \sum_{k=0}^{\beta_p} \frac{(h(p^k))^2}{p_k} + 2\sum_{k=0}^{\beta_p} h(p^k) \sum_{l=k+1}^{+\infty} \frac{h(p^l)}{p^l} \\ &= \left(1 + 4\sum_{k=1}^{\beta_p} \frac{1}{p^k}\right) + 4\sum_{l=1}^{+\infty} \frac{(-1)^l}{p^l} + 8\sum_{k=1}^{\beta_p} \frac{(-1)^{2k+1}}{p^{k+1}} \sum_{j=0}^{+\infty} \frac{(-1)^j}{p^j} \\ &= 1 + \frac{4(p^{\beta_p}-1)}{p^{\beta_p}(p-1)} - \frac{4}{p+1} - \frac{8(p^{\beta_p}-1)}{p^{\beta}(p+1)(p-1)} \\ &= 1 - 4\frac{1}{p^{\beta}(p+1)} \quad \text{for } p \le y. \end{aligned}$$

It follows that

$$\begin{aligned} P(x) &= \prod_{p \le x} w_p \\ &= \prod_{p \le y} \left(1 - \frac{4}{p+1}\right) \prod_{\substack{p \le y \\ p \mid b}} \left(1 - \frac{4}{p^{\beta_p}(p+1)}\right)\left(1 - \frac{4}{p+1}\right)^{-1} \\ &\ll \exp\left(-4\sum_{p \le y} \frac{1}{p} + 4\sum_{p \le y} \frac{1}{p}\right) \ll \frac{(\log\log(b+1))^4}{(\log y)^4}. \end{aligned} \tag{5.13}$$

By (5.1), (5.11) and (5.12), the upper bound in (5.10) can be estimated in the following way (writing again $\ell_k(x)$ for the k-fold logarithm):

$$\begin{aligned} & x^{1-2\alpha} \exp\left(c\frac{r^\alpha}{\log r}\right) + (S(r,x))^{1/2} + \frac{1}{r\log r} + \frac{1}{x^{1/2}\log x} \\ &\ll \exp\left(-\frac{\log x \ell_3(x)}{\ell_2(x)} + c\frac{y^\alpha}{\log y}\right) + 0 + \frac{1}{y\log y} + \frac{1}{x^{1/2}\log x} \\ &\ll \exp\left(-\frac{\log x \ell_3(x)}{\ell_2(x)} + c\left(\frac{\log x}{\ell_2(x)}\right)^{2\alpha}\left(\log\frac{(\log x)^2}{(\ell_2(x))^2}\right)^{-1}\right) + \frac{1}{y\log y} \\ &= \exp\left(-\frac{\log x \ell_3(x)}{\ell_2(x)} + O\left(\frac{\log x}{\ell_2(x)}\right)\right) + \frac{1}{y\log y} \\ &\ll \exp\left(-\frac{\log x \ell_3(x)}{2\ell_2(x)}\right) + \frac{1}{y\log y} \\ &= \exp\left(-\left(\frac{1}{2} + o(1)\right) y^{1/2}\ell_2(y)\right) + \frac{1}{y\log y} \ll \frac{1}{y\log y}. \end{aligned} \tag{5.14}$$

(5.3) follows from (5.2), (5.13) and (5.14), and this completes the proof of Theorem 5.

6. Complexity and correlation. Another often used measure of pseudorandomness of binary sequences is *complexity*. Consider a finite set $\mathcal{S}$ of finitely many symbols, also called *letters*, and form a, finite or infinite, sequence $w = s_1 s_2 \dots$ of these letters; such a sequence w is also called a *word*. If $v = t_1 \dots t_k$ is a finite word and there is an $n \in \mathbb{N}$ such that $s_n = t_1$, $s_{n+1} = t_2, \dots, s_{n+k-1} = t_k$, i.e., the word v *occurs in* w *at place* n, then v is said to be a *factor* (of length k) of w. The *complexity* of the word w is characterized by the function $f(k, w)$ defined in the following way: for $k \in \mathbb{N}$, let $f(k, w)$ denote the number of different factors of length k occurring in w. In particular, for a "good" PR sequence $E_N \in \{-1, +1\}^N$ one expects high complexity, more exactly, one expects that $f(k, E_N) = 2^k$ for "small" k, and $f(k, E_N)$ is "large" for k growing not faster than $\log N$.

In the previous parts of this series we did not study the complexity of the given sequences. The reason is that, as Theorem 6 will show, small correlation implies high complexity (but, clearly, it is not so the other way round); thus if we are able to control the correlation then estimating it, we obtain information superior to the one obtained by studying complexity. As pointed out in Section 3, in the case of Liouville's function it is hopeless to give a good estimate for the correlation; on the other hand, we shall be able to estimate the complexity at least hypothetically. Moreover, the comparison of the complexities of the "truncated" λ and γ functions (to be carried out in Part II) will reflect an interesting contrast in their structures.

First we prove

THEOREM 6. *If $k, N \in \mathbb{N}$, and the sequence $E_N \in \{-1, +1\}^N$ satisfies*

$$C_l(E_N) \le \frac{N}{2^{2k+1}} \quad \text{for } l = 1, \dots, k, \tag{6.1}$$

then

$$f(k, E_N) = 2^k \tag{6.2}$$

(i.e., E_N contains every word of length k).

Proof. The proof will be based on the following lemma:

LEMMA 4. *If $k, N \in \mathbb{N}$, $k \le N$ and $E_N \in \{-1, +1\}^N$, then for all $(\varepsilon_1, \dots, \varepsilon_k) \in \{-1, +1\}^k$ we have*

$$\begin{aligned} &\big|\big|\{n : 1 \le n \le N - k + 1, \\ &\qquad (e_n, e_{n+1}, \dots, e_{n+k-1}) = (\varepsilon_1, \dots, \varepsilon_k)\}| - (N - k + 1)/2^k\big| \\ &\qquad\qquad \le \sum_{l=1}^{k} \binom{k}{l} C_l(E_N). \end{aligned} \tag{6.3}$$

P r o o f. Clearly we have

$$\begin{aligned}
&|\{n : 1 \le n \le N-k+1,\ (e_n, e_{n+1}, \dots, e_{n+k-1}) = (\varepsilon_1, \dots, \varepsilon_k)\}| \\
&\quad = \sum_{n=1}^{N-k+1} \frac{\varepsilon_1 \dots \varepsilon_k}{2^k}(e_n + \varepsilon_1)(e_{n+1} + \varepsilon_2) \dots (e_{n+k-1} + \varepsilon_k) \\
&\quad = \frac{1}{2^k} \sum_{n=1}^{N-k+1} (\varepsilon_1 e_n + 1)(\varepsilon_2 e_{n+1} + 1) \dots (\varepsilon_k e_{n+k-1} + 1) \\
&\quad = \frac{N-k+1}{2^k} + \sum_{l=1}^{k} \sum_{0 \le d_1 < \dots < d_l \le k-1} \varepsilon_{d_1+1} \dots \varepsilon_{d_l+1} \sum_{n=1}^{N-k+1} e_{n+d_1} \dots e_{n+d_l},
\end{aligned}$$

whence

$$\begin{aligned}
&||\{n : 1 \le n \le N-k+1,\ (e_n, e_{n+1}, \dots, e_{n+k-1}) = (\varepsilon_1, \dots, \varepsilon_k)\}| \\
&\qquad\qquad -(N-k+1)/2^k| \\
&\quad = \Big|\sum_{l=1}^{k} \sum_{0 \le d_1 < \dots < d_l \le k-1} \varepsilon_{d_1+1} \dots \varepsilon_{d_l+1} V(E_N, N-k+1, (d_1, \dots, d_l))\Big| \\
&\quad \le \sum_{l=1}^{k} \sum_{0 \le d_1 < \dots < d_l \le k-1} |V(E_N, N-k+1, (d_1, \dots, d_l))| \\
&\quad \le \sum_{l=1}^{k} \sum_{0 \le d_1 < \dots < d_l \le k-1} C_l(E_N) = \sum_{l=1}^{k} \binom{k}{l} C_l(E_N),
\end{aligned}$$

which completes the proof of the lemma.

To derive the theorem from the lemma, first observe that by (6.1) we have

$$k \le \tfrac{1}{2} \cdot 2^{2k+1} \le \tfrac{1}{2} \cdot 2^{2k+1} C_l(E_N) \le N/2.$$

By (6.1) and (6.3), it follows from Lemma 4 that for all $(\varepsilon_1, \dots, \varepsilon_k) \in \{-1, +1\}^k$ we have

$$\begin{aligned}
&|\{n : 1 \le n \le N-k+1, (e_n, e_{n+1}, \dots, e_{n+k-1}) = (\varepsilon_1, \dots, \varepsilon_k)\}| \\
&\qquad \ge \frac{N-k+1}{2^k} - \sum_{l=1}^{k} \binom{k}{l} C_l(E_N) \\
&\qquad \ge \frac{N - N/2 + 1}{2^k} - \sum_{l=1}^{k} \binom{k}{l} \cdot \frac{N}{2^{2k+1}} > \frac{N}{2^{k+1}} - \frac{N}{2^{2k+1}} \sum_{l=1}^{k} \binom{k}{l} = 0
\end{aligned}$$

so that, indeed, E_N contains every word $(\varepsilon_1, \dots, \varepsilon_k) \in \{-1, +1\}^k$, which proves (6.2).

7. Complexity of the Liouville function. To estimate the complexity of the sequence L_N (Liouville function) seems to be as hopeless as the estimate of the correlation of it. Chowla [Ch, p. 95] formulates the following related conjecture: "Let $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_g$ be arbitrary numbers each equal to $+1$ or -1, where g is a fixed (but arbitrary) number. Then the equations (in n)

$$\lambda(n+m) = \varepsilon_m \quad (1 \le m \le g)$$

have infinitely many solutions. For $g \ge 3$ this seems an extremely hard conjecture." The $g = 3$ special case of this conjecture has been proved by Hildebrand [Hi2].

Unlike in the case of correlation, here we shall be able to establish at least a hypothetical result, more exactly, we derive Chowla's conjecture from a very well-known and widely used hypothesis. This hypothesis is Schinzel's "Hypothesis H" [Sc], [ScSi] (see also [HR, p. 2]) which generalizes the twin prime conjecture:

Hypothesis H. *If $k \in \mathbb{N}$, $F_1, \dots, F_k$ are distinct irreducible polynomials in $\mathbb{Z}[x]$ (with positive leading coefficients) and the product polynomial $F = F_1 \dots F_k$ has no fixed prime divisor, then there exist infinitely many integers n such that each $F_i(n)$ $(i = 1, \dots, k)$ is a prime.*

We prove

Theorem 7. *Assuming Hypothesis* H, *for all $k \in \mathbb{N}$ and $\{\varepsilon_1, \dots, \varepsilon_k\} \in \{-1, +1\}^k$ there are infinitely many positive integers m such that*

$$\lambda(m+1) = \varepsilon_1, \quad \lambda(m+2) = \varepsilon_2, \ \dots, \ \lambda(m+k) = \varepsilon_k. \tag{7.1}$$

(We remark that the analogous result with γ in place of λ could be proved similarly.)

It follows trivially from this theorem that

Corollary 3. *Assuming Hypothesis* H, *for all $k \in \mathbb{N}$ there is a number $N_0 = N_0(k)$ such that for $N > N_0$ we have*

$$f(k, L_N) = 2^k.$$

Proof (of Theorem 7). If p is a prime and $i \in \mathbb{N}$, then define $\alpha_p(i)$ by $p^{\alpha_p(i)} \| i$ so that

$$i = \prod_{p \mid i} p^{\alpha_p(i)} \tag{7.2}$$

(and $\alpha_p(i) = 0$ for $p \nmid i$). For $i = 1, \dots, k$, define the number δ_i by $\delta_i \in \{0, 1\}$ and

$$\delta_i \equiv \begin{cases} \sum_{p \le k} \alpha_p(i) - 1 \pmod 2 & \text{for } \varepsilon_i = +1, \\ \sum_{p \le k} \alpha_p(i) \pmod 2 & \text{for } \varepsilon_i = -1. \end{cases} \tag{7.3}$$

Let $q_1, \dots, q_k$ be k distinct primes greater than k (e.g., we may choose the first k primes greater than k), and write

$$h_i = q_i^{\delta_i} = \begin{cases} q_i & \text{if } \delta_i = 1, \\ 1 & \text{if } \delta_i = 0. \end{cases}$$

Write $M = [1, \dots, k]$ (= the least common multiple of $1, \dots, k$). Consider the linear congruence system

$$\frac{M}{i}x + 1 \equiv 0 \pmod{h_i}, \qquad i = 1, \dots, k. \tag{7.4}$$

Here $(M/i, h_i) = 1$ for $i = 1, \dots, k$, and thus each of these congruences can be solved. Moreover, the moduli $h_1, \dots, h_k$ are pairwise coprime and thus by the Chinese remainder theorem, the system (7.4) can be solved, and the solutions form a residue class modulo

$$H := \prod_{i=1}^{k} h_i,$$

i.e., there is an $x_0 \in \mathbb{Z}$ such that all the solutions are

$$x \equiv x_0 \pmod{H}.$$

In other words, x is of the form

$$x = Hn + x_0 \quad \text{with } n \in \mathbb{Z}.$$

For $i = 1, \dots, k$, write

$$a_i = \frac{M}{i} \cdot \frac{H}{h_i}, \qquad b_i = \frac{1}{h_i}\left(\frac{M}{i}x_0 + 1\right) \tag{7.5}$$

and

$$F_i(n) = a_i n + b_i. \tag{7.6}$$

Now we show that the polynomials (7.6) satisfy the assumptions in Hypothesis H. $a_i \in \mathbb{Z}$ holds trivially, and since x_0 is a solution of (7.4), $b_i \in \mathbb{Z}$ also holds. The polynomials (7.6) are clearly distinct since $a_i \neq a_j$ for $i \neq j$, and they are irreducible since they are linear. It remains to show that the product polynomial

$$F(n) = F_1(n) \dots F_k(n)$$

has no fixed prime divisor. We prove this by contradiction: assume that there is a prime p such that

$$F(n) = F_1(n) \dots F_k(n) \equiv 0 \pmod{p} \quad \text{for all } n \in \mathbb{Z}. \tag{7.7}$$

We have to distinguish three cases.

Case 1. Assume first that $p > k$ and $p \notin \{q_1, \dots, q_k\}$. It follows that $p \nmid H$ and thus, since the prime factors of M do not exceed k, we have $p \nmid a_i$ for $i = 1, \dots, k$. Then $F(n) \in \mathbb{Z}[n]$ is a polynomial of degree k which is less

than p, and its leading coefficient is $\not\equiv 0 \pmod p$; these facts imply that (7.7) cannot hold.

Case 2. Assume now that $p = q_i$ for some $1 \le i \le k$. If $p \nmid (a_j, b_j)$ for $j = 1, \dots, k$, then again $f(n)$ is not the zero polynomial modulo p and its degree is $\le k < q_i = p$ so that (7.7) cannot hold. Thus there is a j with $p \,|\, (a_j, b_j)$. Since clearly $p = q_i \nmid \frac{M}{i} \cdot \frac{H}{h_i} = a_i$, we have $j \ne i$. Then $p = q_i \,|\, a_j = (M/j) \prod_{l \ne j} h_l$ implies that $h_i = q_i$. Since $h_i = q_i$ and x_0 satisfies (7.4), we have

$$Mx_0 + i = i\left(\frac{M}{i}x_0 + 1\right) \equiv 0 \pmod{q_i}, \tag{7.8}$$

and as $p = q_i \,|\, b_j$,

$$Mx_0 + j = jh_jb_j \equiv 0 \pmod{q_i}. \tag{7.9}$$

It follows from (7.8) and (7.9) that

$$i - j = (Mx_0 + i) - (Mx_0 + j) \equiv 0 \pmod{q_i}$$

but this is impossible since $i \ne j$, $1 \le i, j \le k$ and $q_i > k$.

Case 3. Assume finally that $p \le k$. Then it follows from (7.7) that p is also a fixed prime divisor of the polynomial

$$Q(n) := \Big(\prod_{i=1}^{k} h_i\Big) F(n) = \prod_{i=1}^{k} h_i F_i(n) = \prod_{i=1}^{k} Q_i(n)$$

where

$$Q_i(n) = h_i F_i(n) = \frac{M}{i} Hn + \left(\frac{M}{i}x_0 + 1\right) \quad \text{for } i = 1, \dots, k.$$

Clearly,

$$p \,\Big|\, \frac{M}{i}$$

implies that

$$Q_i(n) \equiv 1 \pmod p \quad \text{for all } n \in \mathbb{N}.$$

Thus p is also a fixed prime divisor of the polynomial

$$Q^*(n) := \prod_{\substack{1 \le i \le k \\ (p, M/i) = 1}} Q_i(n). \tag{7.10}$$

It follows from $p \le k$, $1 \le i \le k$ and $(p, M/i) = (p, [1, \dots, k]/i) = 1$ that, defining β_p by

$$p^{\beta_p} \le k < p^{\beta_p + 1} \tag{7.11}$$

(so that $\beta_p = \alpha_p(M)$), we have $p^{\beta_p} \,|\, i$ so that i is of the form

$$i = jp^p \quad \text{with } j \in \mathbb{N},\ j \le k/p^{\beta_p}.$$

Thus i in (7.10) may assume at most k/p^{β_p} values, so that the degree of the polynomial $Q^*(n)$ is, by (7.11), at most $k/p^{\beta_p} < p$. Moreover, it follows from $(p, M/i) = 1$ that the leading coefficient of $Q_i(n)$ is also coprime to p:

$$(h_i a_i, p) = \left(\frac{M}{i} H, p\right) = 1.$$

Thus the leading coefficient of $Q^*(n)$ is also coprime to p. Then the polynomial $Q^*(n)$ is not identically zero modulo p and its degree is less than p, which contradicts the fact that p is a fixed prime divisor of it.

This completes the proof that there is no prime p satisfying (7.7) so that, indeed, the polynomials (7.6) satisfy the assumptions in Hypothesis H. Since now this hypothesis is assumed, there are infinitely many integers n such that each $F_i(n)$ $(i = 1, \dots, k)$ is a prime. For such an integer n define $m = m(n)$ by

$$m = MHn + Mx_0.$$

It remains to show that m satisfies (7.1).

For $i = 1, \dots, k$ we have

$$m + i = MHn + (Mx_0 + i) = ih_i\left(\frac{M}{i} \cdot \frac{H}{h_i} n + \frac{1}{h_i}\left(\frac{M}{i} x_0 + 1\right)\right) = ih_i F_i(n).$$

By (7.2) and (7.3), it follows that

$$\Omega(m+i) = \Omega(i) + \Omega(h_i) + \Omega(F_i(n)) = \sum_{p \mid i} \alpha_p(i) + \delta_i + 1$$

$$= \sum_{p \le k} \alpha_p(i) + \delta_i + 1 \equiv \begin{cases} 0 \pmod 2 & \text{for } \varepsilon_i = +1, \\ 1 \pmod 2 & \text{for } \varepsilon_i = -1, \end{cases}$$

whence $\lambda(m+i) = (-1)^{\Omega(m+i)} = \varepsilon_i$ for $i = 1, \dots, k$, which completes the proof of the theorem.

Acknowledgements. We would like to thank Dr. Louis Goubin (Bull. PTS) for his valuable comments.

References

[Ch] S. Chowla, *The Riemann Hypothesis and Hilbert's Tenth Problem*, Gordon and Breach, New York, 1965.

[Ell1] P. D. T. A. Elliott, *Probabilistic Number Theory*, Vol. II, Springer, New York, 1980.

[Ell2] —, *On the correlation of multiplicative functions*, Notas Soc. Mat. Chile 11 (1992), 1–11.

[Ell3] —, *On the correlation of multiplicative and the sum of additive arithmetic functions*, Mem. Amer. Math. Soc. 538 (1994).

[GH] S. W. Graham and D. Hensley, *Problem E3025*, Amer. Math. Monthly 90 (1983), 707.

[Ha] G. Halász, *Über die Mittelwerte multiplikativer zahlentheoretischer Funktionen*, Acta Math. Acad. Sci. Hungar. 19 (1968), 365–403.

[HR] H. Halberstam and H.-E. Richert, *Sieve Methods*, Academic Press, London, 1974.

[HPW] G. Harman, J. Pintz and D. Wolke, *A note on the Möbius and the Liouville function*, Studia Sci. Math. Hungar. 20 (1985), 295–299.

[Hi1] A. Hildebrand, *Multiplicative functions at consecutive integers*, Math. Proc. Cambridge Philos. Soc. 100 (1986), 229–236.

[Hi2] —, *On consecutive values of the Liouville function*, Enseign. Math. 32 (1986), 219–226.

[Hi3] —, Math. Reviews, review no. 95d:11099.

[Li] J.-E. Littlewood, *Quelques conséquences de l'hypothèse que la fonction $\zeta(s)$ de Riemann n'a pas de zéros dans le demi-plan $R(s) > \frac{1}{2}$*, C. R. Acad. Sci. Paris 154 (1912), 263–266.

[MS1] C. Mauduit and A. Sárközy, *On finite pseudorandom binary sequences I: Measure of pseudorandomness, the Legendre symbol*, Acta Arith. 82 (1997), 365–377.

[MS2] —, —, *On finite pseudorandom binary sequences II: The Champernowne, Rudin–Shapiro and Thue–Morse sequences. A further construction*, J. Number Theory, to appear.

[Sa] A. Sárközy, *On the number of prime factors of integers of the form $a_i + b_j$*, Studia Sci. Math. Hungar. 23 (1988), 161–168.

[Sc] A. Schinzel, *Remarks on the paper "Sur certaines hypothèses concernant les nombres premiers"*, Acta Arith. 7 (1961/1962), 1–8.

[ScSi] A. Schinzel et W. Sierpiński, *Sur certaines hypothèses concernant les nombres premiers*, ibid. 4 (1958), 185–208; Corrigendum: ibid. 5 (1959), 259.

[St1] G. Stepanauskas, *The mean values of multiplicative functions, II*, Liet. Mat. Rink. 37 (1997), 212–223 (in Russian).

[St2] —, *The mean values of multiplicative functions, III*, in: Analytic and Probabilistic Methods in Number Theory, New Trends in Probability and Statistics, Vol. 4, A. Laurinčikas *et al.* (eds.), VSP BV, Utrecht, 1997, 371–387.

Institut de Mathématiques de Luminy
CNRS-UPR 9016
163 av. de Luminy, Case 930
13288 Marseille Cedex 9, France
E-mail: cassaigne@iml.univ-mrs.fr
ferenczi@iml.univ-mrs.fr
mauduit@iml.univ-mrs.fr

Département de Mathématiques
Institut Girard Desargues, UPRES-A 5028
Université Claude Bernard, Lyon 1
43, Bd du 11 Novembre 1918, Bât. 101
69622 Villeurbanne Cedex, France
E-mail: Joel.Rivat@desargues.univ-lyon1.fr

Department of Algebra and Number Theory
Eötvös Loránd University
Múzeum krt. 6-8
1088 Budapest, Hungary
E-mail: sarkozy@cs.elte.hu

Received on 31.7.1998 (3429)