ACTA ARITHMETICA
XCV.4 (2000)

On finite pseudorandom binary sequences IV: The Liouville function, II

by

Julien Cassaigne (Marseille), Sébastien Ferenczi (Marseille),
Christian Mauduit (Marseille), Joël Rivat (Nancy)
and András Sárközy (Budapest)

1. Introduction. Throughout this paper, we shall use the following notations: p_i for the ith prime number ($p_1 = 2$, $p_2 = 3$, $p_3 = 5, \ldots$), $\pi(x)$ for the number of primes $\le x$, $\omega(n)$ for the number of distinct prime factors of n, $\Omega(n)$ for the number of prime factors of n counted with multiplicity. We 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

$$L_N = \{\lambda(1), \ldots, \lambda(N)\} \quad \text{and} \quad G_N = \{\gamma(1), \ldots, \gamma(N)\}.$$

For $y \ge 1$ 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

$$L_N(y) = \{\lambda_y(1), \ldots, \lambda_y(N)\} \quad \text{and} \quad G_N(y) = \{\gamma_y(1), \ldots, \gamma_y(N)\}.$$

In this series we study pseudorandom properties of binary sequences. As measures of pseudorandomness of the binary sequence

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

2000 *Mathematics Subject Classification*: Primary 11N64.

Research partially supported by Hungarian National Foundation for Scientific Research, Grant No. T017433 MKM fund FKFP-0139//1997 and by French-Hungarian APAPE-OMFB exchange program F-5//97.

the *well-distribution measure*:

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

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

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

(where the maximum is taken over all $M \in \mathbb{N}$ and non-negative integers $d_1 < \ldots < d_k$ such that $M + d_k \le N$) are used.

We also need the notion of *complexity*. Consider a finite set $\mathcal{S}$ of symbols, also called *letters*, and form a, finite or infinite, sequence $w = s_1 s_2 \ldots$ of these letters; such a sequence w is also called a *word*. The *concatenation* of words is defined in the following way: if $w = w_1 \ldots w_r$, $w' = w'_1 \ldots w'_s$ are two words then we set $ww' = w_1 \ldots w_r w'_1 \ldots w'_s$. If $v = t_1 t_2 \ldots t_k$ is a finite word and there is an $n \in \mathbb{N}$ such that $s_n = t_1$, $s_{n+1} = t_2, \ldots, s_{n+k-1} = t_k$, 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" pseudorandom 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 Part I [CFMRS] of this paper, we first studied the well-distribution measure and correlation of the sequences L_N, $L_N(y)$. Next we analyzed the connection between correlation and complexity. Finally, we proved a conditional result on the complexity of the Liouville function: we showed that assuming Schinzel's "Hypothesis H" [Sc], [ScSi], for $k \in \mathbb{N}$, $N > N_0(k)$ we have

$$f(k, L_N) = 2^k. \tag{1.1}$$

We remark that in all the problems studied in Part I, there is no significant difference between the behaviour of the functions λ and γ, and the behaviour of their truncated versions is also similar.

Since (1.1) is a conditional result, one might like to prove unconditional results on the complexity of the functions studied by us as well. Since this seems to be hopeless in the case of the functions λ and γ, instead we will study their truncated versions. This will be done in Section 2 and it will turn out that, unlike the cases studied so far, there is a quite striking contrast between the behaviour of the functions λ_y and γ_y. In Section 3 we will return to the analysis of the structure of the sequence $\{\lambda(1), \lambda(2), \ldots\}$.

First we will formulate a conjecture on the behaviour of the λ function over polynomials $f(n) \in \mathbb{Z}[n]$. Next we will prove this conjecture in the special case when $f(n)$ is the product of certain linear polynomials. In Section 4 we will prove the same conjecture for certain quadratic polynomials $f(n)$. In Section 5 we will pose several related unsolved problems and conjectures. Finally, in Section 6 we will present numerical data obtained by computer.

2. The complexity of the truncated functions. Let $2 \le y \le N$, and write $P_y = \prod_{p \le y} p$.

Consider first the sequence $G_N(y)$. Clearly, the value of $\gamma_y(n)$ depends only on the number of primes $p \le y$ with $p \mid n$, and the number of these primes is a periodic function of n with period P_y:

$$\gamma_y(n + P_y) = \gamma_y(n) \quad \text{for } n = 1, 2, \ldots$$

It follows trivially that for all $k \in \mathbb{N}$, the complexity $f(k, G_N(y))$ is at most the period length:

$$f(k, G_N(y)) \le P_y = \prod_{p \le y} p$$

so that it is bounded as $N \to \infty$ for fixed k, and then we let $k \to \infty$. (Indeed it can be shown with a little work that for fixed y, $k > k_0(y)$ and $N \to \infty$ there is equality here.)

On the other hand, we will show that the complexity $f(k, L_N(y))$ grows as fast as a constant times $k^{\pi(y)}$ (for every fixed k and $N \to \infty$):

Theorem 1. *For $y \ge 2$, $r = \pi(y)$, there are positive numbers c_r, c'_r (depending only on r) such that if $k \in \mathbb{N}$ and N is large enough in terms of r and k then*

$$c_r k^r < f(k, L_N(y)) < c'_r k^r. \tag{2.1}$$

We do not know whether the quotient $f(k, L_N(y))//k^r$ has a limit or not. We remark that when $y = 2$ then $L_N(y)$ is an automatic sequence and explicit formulas for $f(k, L_N(y))$ can be found with standard techniques. In particular

$$\liminf_{k\to\infty} \lim_{N\to\infty} f(k, L_N(y)) = 3//2, \qquad \limsup_{k\to\infty} \lim_{N\to\infty} f(k, L_N(y)) = 5//3,$$

so that $f(k, L_N(y))//k^r$ has no limit in this case. For $y = 3$ computations (see Table 3) seem to indicate that it has no limit either.

In order to prove Theorem 1 we will first prove

Lemma 1. *For $n, i \in \mathbb{N}$, we define $\alpha_i(n) = \max\{\alpha \ge 0,\ p_i^\alpha \mid n\}$, and write*

$$s_i(n) = (-1)^{\alpha_i(n)}, \qquad S_N(i) = \{s_i(1), \ldots, s_i(N)\}.$$

(i) *For all* $i \in \mathbb{N}$ *there is an* (*explicitly computable*) *constant* b_i *such that if* $k \in \mathbb{N}$ *and* N *is large enough in terms of* i *and* k, *then*

$$k \le f(k, S_N(i)) \le b_i k.$$

(ii) *If* $i, k \in \mathbb{N}$ *and* w *is a factor of length* k *of the sequence*

$$s_i(1), s_i(2), \ldots, \tag{2.2}$$

then there are $j, m \in \mathbb{N}$ *such that for each* $q = 0, 1, 2, \ldots,$ *the word occurring at place* $j + p_i^m q$ *is* w.

(iii) *If* $j, m, k \in \mathbb{N}$ *and* N *is large enough in terms of* m *and* k, *then the number of different factors of length* k *of* $S_N(i)$ *occurring at places* $\equiv j \pmod{p_i^m}$ *is at most the number of different factors of length* $[k//p_i^m] + 2$ *of* $S_N(i)$.

Proof. (i) Define the operation σ_i on the set of the words on the letters -1, $+1$ by

$$\sigma_i(1) = \underbrace{1 \ldots 1}_{p_i - 1}(-1), \qquad \sigma_i(-1) = \underbrace{1 \ldots 1}_{p_i}$$

and

$$\sigma_i(ww') = \sigma_i(w)\sigma_i(w').$$

Then the word $S_{p_i^m}(i)$ is the image of the word 1 by σ_i^m. σ_i is called a *primitive substitution*, and the upper bound for $f(k, S_N(i))$ is standard (see [Que], proof of Proposition V.19), while the lower bound follows from the fact that the infinite sequence (2.2) is not ultimately periodic (see [HM]).

(ii) This follows from the fact that the sequence (2.2) is a concatenation of the words $\sigma_i^m(1)$ and $\sigma_i^m(-1)$, which both have length p_i^m. Every factor of the sequence (2.2) must occur at place j in $\sigma_i^m(1)$ for some j and m; it will then occur at place j in both $\sigma_i^{m+1}(1)$ and $\sigma_i^{m+1}(-1)$, and hence at all places $j + qp_i^{m+1}$ in (2.2).

(iii) This follows from the relation $S_{Np_i^m}(i) = \sigma_i^m S_N(i)$. Assume $0 \le j < p_i^m$. Let $q = [(k + j - 1)//p_i^m]$. A word w of length k occurring at a place congruent to $j \pmod{p_i^m}$ can be decomposed as $w = f\sigma_i^m(e_1) \ldots \sigma_i^m(e_{q-1})d$, f being a suffix of $\sigma_i^m(e_0)$, d a prefix of $\sigma_i^m(e_q)$, and $e_0 \ldots e_q$ a factor of length $q + 1$ of $S_N(i)$. As w is uniquely determined by j and $e_0 \ldots e_q$ and $q + 1 \le [k//p_i^m] + 2$, the assertion follows.

Proof of Theorem 1. We have $\lambda_y(n) = s_1(n) \ldots s_r(n)$ if $r = \pi(y)$. Hence

$$f(k, L_N(y)) \le \prod_{i=1}^{r} f(k, S_N(i)),$$

which gives the upper bound in Theorem 1 if we use the one in Lemma 1.

We will prove the lower bound by induction. For $y \le p_1$ it holds by Lemma 1. Assume now that it holds for $y < p_r$, and take a y such that

$p_r \le y < p_{r+1}$. Then $\lambda_y(n) = \lambda_{y'}(n)s_r(n)$ for $p_{r-1} \le y' < p_r$. For N large enough, the sequence $(L_N(y'), S_N(r))=\{(\lambda_{y'}(1), s_r(1)),\ldots,(\lambda_{y'}(N), s_r(N))\}$ on 4 letters has at least $c_{r-1}k^r$ factors, as factors of $L_N(y')$ occur at places $j+p_1^{m_1}\ldots p_{r-1}^{m_{r-1}}a$, $a=0,1,2,\ldots$ (by Lemma 1 and $\lambda_{y'}(n)=s_1(n)\ldots s_{r-1}(n)$), while factors of $S_N(r)$ occur at places $j' + p_r^m a'$, $a' = 0,1,2,\ldots$ Hence all pairs (w', w''), where w' is a factor of $L_N(y')$ and w'' is a factor of $S_N(r)$, occur at factors of $(L_N(y'), S_N(r))$.

Now, with a factor w of $L_N(y)$, we associate all the factors (w', w'') of $(L_N(y'), S_N(r))$ such that

$$w = \lambda_y(m)\ldots\lambda_y(m'), \quad w' = \lambda_{y'}(m)\ldots\lambda_{y'}(m'), \quad w'' = s_r(m)\ldots s_r(m')$$

with $m < m'$. We will show that to w there correspond at most K_r different factors (w', w'') for a fixed constant K_r; clearly, this will complete the proof that the lower bound also holds for $p_r \le y < p_{r+1}$.

To simplify the notation we put $q = p_1\ldots p_{r-1}$, $p = p_r$, $s(n) = s_r(n)$. Let w be a factor of length k of $L_N(y)$ where N is large enough. We shall control the places where w can occur. Suppose

$$w = \lambda_y(m_1)\ldots\lambda_y(m_1') = \lambda_y(m_2)\ldots\lambda_y(m_2')$$

and

$$m_2 \equiv m_1 \pmod{pq},$$

i.e., $m_2 = m_1 + apq$ with some $a \in \mathbb{N}$. Then $\lambda_y(m + apq) = \lambda_y(m)$ for $m \in \mathcal{A}$ where $\mathcal{A}$ is an interval of length k. But $\lambda_{y'}(m + apq) = \lambda_{y'}(m) = 1$ whenever $m \equiv \pm 1 \pmod q$, hence $s(m+apq) = s(m)$ when $m \equiv \pm 1 \pmod q$. Thus we choose m such that $m \equiv \pm 1 \pmod q$, $m \equiv 0 \pmod p$. Then $s(m + apq) = s(m)$ whence $s(m//p + aq) = s(m//p)$.

We choose $m \in \mathcal{A}$ such that $m \equiv 1 \pmod r$, $m \equiv 0 \pmod{p^2}$, $m \not\equiv 0 \pmod{p^3}$. This is possible if $k \ge 2p^2r$, since in an interval of length p^2r, there is m such that $m \equiv 1 \pmod r$ and $m \equiv 0 \pmod{p^2}$, and if it happens that $m \equiv 0 \pmod{p^3}$, then $m + p^2r$ satisfies the condition.

Then $s(m//p + aq) = s(m//p) = -1$ so that $m//p + aq \equiv 0 \pmod p$ whence $a \equiv 0 \pmod p$. Thus we have shown that if $k \ge 2p^2q$, then $m_2 - m_1$ must be a multiple of p^2.

We can iterate the process: writing $a = pa'$, we have

$$s\left(\frac{m}{p} + pa'q\right) = s\left(\frac{m}{p}\right)$$

so that for those m which are $\equiv 0 \pmod{p^2}$, we have

$$s\left(\frac{m}{p^2} + a'q\right) = s\left(\frac{m}{p^2}\right),$$

and the same reasoning shows that if $k \geq 2p^3q$ then $m_2 - m_1$ is a multiple of p^3.

Similarly, $m_2 - m_1$ is a multiple of p^r whenever $k \geq 2p^r q$. Then the last assertion of Lemma 1 shows that there are at most $\overline{K}_r = f(2p^2q+2, S_N(r))$ possible factors w'' of $S_N(r)$ such that (w', w'') correspond to w. But when w and w'' are known, so is w'. We have to multiply $\overline{K}_r$ by pq to get K_r, to take into account the possible congruences $\pmod{pq}$ of the occurrences of w. Hence the result, with explicitly computable values of c_r and c'_r.

3. The λ function over a product of linear polynomials. If we try to prove something unconditional on the structure of the sequence $\{\lambda(1), \lambda(2), \ldots\}$, the first question to decide is whether the sequence is ultimately periodic. It follows from a result of Sárközy [Sá] that the answer to this question is negative. Namely, an ultimately periodic arithmetic function $g(n)$ satisfies a linear recursion. By a special case of the main theorem in [Sá], a completely multiplicative function $g(n)$ with $g(n) \not\equiv 0$, $g(n) = o(n)$ satisfies a linear recursion if and only if $g(n) = \chi(n)$ is a (multiplicative) character modulo m for some $m \in \mathbb{N}$ so that either $g(n) = 0$ infinitely often (for $m > 1$) or $g(n) = 1$ for all n. Since $\lambda(n)$ is never 0 and it is -1 infinitely often, it follows that $\lambda(n)$ cannot be ultimately periodic.

Next one might like to know whether this statement can be sharpened in the following way: the function $\lambda(n)$ cannot be constant over an arithmetic progression, i.e., there are no $a \in \mathbb{N}$, $b \in \mathbb{Z}$ such that $\lambda(an + b)$ is constant for $n > n_0$. The affirmative answer follows easily from the following

LEMMA 2. *If $a \in \mathbb{N}$, $b \in \mathbb{Z}$, and $g(n)$ is a complex-valued multiplicative arithmetic function such that $g(an + b)$ is a non-zero constant for $n > n_0$, then there is a Dirichlet character $\chi(n)$ modulo a so that $g(n) = \chi(n)$ for every $n \in \mathbb{N}$ with $(a, n) = 1$.*

This lemma can be derived easily from Sárközy's result [Sá], and it is stated as Lemma (19.3) in Elliott's book [Ell] where a simple direct proof is given.

COROLLARY 1. *There are no $a \in \mathbb{N}$, $b \in \mathbb{Z}$ such that $\lambda(an+b)$ is constant for $n > n_0$.*

Proof. Assume that contrary to the assertion, there are $a \in \mathbb{N}$, $b \in \mathbb{Z}$ such that $\lambda(an + b)$ is constant for $n > n_0$. Then by Lemma 2 there is a (multiplicative) character $\chi(n)$ modulo a so that $\lambda(n) = \chi(n)$ for $(a, n) = 1$. It follows that

$$\lambda(ak + 1) = \chi(1) = 1 \quad \text{for all } k \in \mathbb{N}. \tag{3.1}$$

However, by Dirichlet's theorem there are infinitely many primes p of the form $p = ak + 1$. By the definition of the λ function for these primes p we have $\lambda(p) = -1$, which contradicts $p = ak + 1$ and (3.1), and this completes the proof of Corollary 1.

One might like to extend the problem by studying the λ function over polynomials. In this direction we conjecture:

CONJECTURE 1. *If $f(n) = a_0 n^k + \ldots + a_k \in \mathbb{Z}[n]$, $a_0 > 0$ then $\lambda(f(n))$ is constant for $n > n_0$ if and only if $f(n)$ is of the form $f(n) = b(g(n))^2$ where $b \in \mathbb{N}$, $g(n) \in \mathbb{Z}[n]$.*

This is a weaker form of a conjecture of Chowla [Ch]. He writes:

CONJECTURE 2. *Let $f(x)$ be an arbitrary polynomial with integer coefficients, which is not, however, of the form $cg^2(x)$ where c is an integer and $g(x)$ is a polynomial with integer coefficients. Then*

$$\sum_{n=1}^{x} \lambda(f(n)) = o(x).$$

If $f(x) = x$ this is equivalent to the Prime Number Theorem. If the degree of $f(x)$ is at least 2, this seems an extremely hard conjecture.

Clearly, Conjecture 1 would follow from Conjecture 2. While indeed Conjecture 2 seems hopelessly difficult, we have been able to settle certain special cases of our easier Conjecture 1. First in this section we will study the case when $f(n)$ is the product of certain linear polynomials.

THEOREM 2. *If $a, k \in \mathbb{N}$, $b_1, \ldots, b_k$ are distinct integers with*

$$b_1 \equiv \ldots \equiv b_k \pmod a, \tag{3.2}$$

$g(n)$ is a completely multiplicative arithmetic function such that $g(n) \in \{-1, +1\}$ for all $n \in \mathbb{N}$ and, writing $f(n) = (an + b_1)\ldots(an + b_k)$, $g(f(n))$ is constant for $n \geq n_0$, then

(i) *for any b with $b \equiv b_1 \equiv \ldots \equiv b_k \pmod a$, $g(an + b)$ is ultimately periodic;*

(ii) *there is an $a' \in \mathbb{N}$ with $a \mid a'$ and a real character $\chi(n)$ modulo a' so that*

$$g(n) = \chi(n) \tag{3.3}$$

for every $n \in \mathbb{N}$ with $(a', n) = 1$.

COROLLARY 2. *There are no $a, k \in \mathbb{N}$ and distinct integers $b_1, \ldots, b_k$ with*

$$b_1 \equiv \ldots \equiv b_k \pmod a$$

such that $\lambda((an + b_1)\ldots(an + b_k))$ is constant for $n > n_0$.

Note that Corollary 1 is a special case of Corollary 2.

Proof of Theorem 2. We may assume that $b_1 < \ldots < b_k$. Write $l = (b_k - b_1)//a$ (so that l is an integer by (3.2)). Consider the l-tuple $(g(an+b_1), g(an+b_1+a), \ldots, g(an+b_1+(l-1)a))$ (whose last element is $g(an+b_k-a)$) for all $n \in \mathbb{N}$ with $n \geq n_0$. This l-tuple may assume only finitely many (2^l) distinct values, thus there are $n_1, n_2 \in \mathbb{N}$ with

$$n_0 \leq n_1 < n_2 \tag{3.4}$$

and

$$g(an_1 + b_1 + ja) = g(an_2 + b_1 + ja) \quad \text{for } j = 0, 1, \ldots, l-1. \tag{3.5}$$

Now we show by straight induction that

$$g(m) = g(m + a(n_2 - n_1)) \quad \text{for } m \geq an_1 + b_1,\ m \equiv b \ (\operatorname{mod} a). \tag{3.6}$$

If $an_1 + b_1 \leq m < an_1 + b_k$, $m \equiv b \ (\operatorname{mod} a)$, then (3.6) holds by (3.5). Assume now that

$$m' \geq an_1 + b_k, \tag{3.7}$$

$$m' \equiv b \ (\operatorname{mod} a) \tag{3.8}$$

and (3.6) holds for all m with

$$an_1 + b_1 \leq m < m', \quad m \equiv b \ (\operatorname{mod} a).$$

We have to show that this assumption implies that (3.6) also holds with m' in place of m.

If m is one of the numbers

$$m = m' - b_k + b_j \quad \text{with } 1 \leq j \leq k-1,$$

then by (3.2), (3.7), (3.8) and the definition of b we have

$$m \geq (an_1 + b_k) - b_k + b_1 = an_1 + b_1,$$
$$m = m' - (b_k - b_j) < m'$$

and

$$m = m' - b_k + b_j \equiv b - b + b \equiv b \ (\operatorname{mod} a),$$

so that by the induction hypothesis, (3.6) holds for each of these numbers:

$$g(m' - b_k + b_j) = g(m' - b_k + b_j + a(n_2 - n_1)) \quad \text{for } 1 \leq j \leq k-1. \tag{3.9}$$

Writing $n = (m' - b_k)//a$ (which is an integer by (3.8) and the definition of b) by (3.4) and (3.7) we have

$$n \geq \frac{(an_1 + b_k) - b_k}{a} = n_1 \geq n_0.$$

Thus by the assumption of the theorem we have

$$g(f(n)) = g(f(n + n_2 - n_1)).$$

By the definition of $f(n)$, and since $g(n)$ is completely multiplicative, this can be rewritten as

$$\prod_{j=1}^{k} g(an + b_j) = \prod_{j=1}^{k} g(an + a(n_2 - n_1) + b_j)$$

or, by the definition of n,

$$\prod_{j=1}^{k} g(m' - b_k + b_j) = \prod_{j=1}^{n} g(m' - b_k + b_j + a(n_2 - n_1)). \tag{3.10}$$

Since $g(n) \neq 0$ for $n \in \mathbb{N}$, it follows from (3.9) and (3.10) that (3.6) also holds with m' in place of m, which completes the proof of (3.6).

By (3.6), $g(an + b)$ is ultimately periodic with period $n_2 - n_1$, which proves (i).

Since $g(an + b)$ is ultimately periodic with period $n_2 - n_1$, it follows that $g(a(n_2 - n_1)m + b)$ is constant in m for m large enough, and since $g(n) \in \{-1, +1\}$ for all n, this constant is non-zero. Thus by Lemma 2 there is a Dirichlet character $\chi(n)$ modulo $a' = a(n_2 - n_1)$ so that (3.3) holds for every $n \in \mathbb{N}$ with $(a', n) = 1$. By $g(n) \in \{-1, +1\}$ this is a real character, and this completes the proof of (ii).

4. The λ function over quadratic polynomials. In this section we will settle Conjecture 1 for certain quadratic polynomials:

THEOREM 3. *Let $a \in \mathbb{N}$, $b, c \in \mathbb{Z}$, and write $f(n) = an^2 + bn + c$, $D = b^2 - 4ac$. Assume that a, b and c satisfy the following conditions*:

(i) $2a \mid b$,

(ii) $D < 0$,

(iii) *there is a positive integer k with*

$$\lambda\left(-\frac{D}{4}k^2 + 1\right) = -1. \tag{4.1}$$

(*Note that $-D//4 \in \mathbb{N}$ by* (i) *and* (ii).) *Then $\lambda(f(n))$ assumes both values $+1$ and -1 for infinitely many $n \in \mathbb{N}$.*

Proof. Assume that, contrary to assertion, (i)–(iii) hold, but

$$\lambda(f(n)) \text{ is constant for } n \geq n_0. \tag{4.2}$$

Writing $m = n + b//(2a)$ (note that $b//(2a) \in \mathbb{Z}$ by (i)) we clearly have

$$f(n) = an^2 + bn + c = a\left(n + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a} \tag{4.3}$$
$$= am^2 - \frac{D}{4a}.$$

By (4.2) and (4.3),

$$\lambda\left(am^2 - \frac{D}{4a}\right) \text{ is constant for } m \geq m_0. \tag{4.4}$$

It follows that

$$\lambda\left(a\left(-\frac{D}{4a}t\right)^2 - \frac{D}{4a}\right) = \lambda\left(-\frac{D}{4a}\right)\lambda\left(-\frac{D}{4}t^2 + 1\right) \text{ is constant for } t \geq t_0,$$

whence

$$\lambda\left(-\frac{D}{4}t^2 + 1\right) \text{ is constant for } t \geq t_0. \tag{4.5}$$

By (i) and (ii), $-\frac{D}{4}k^2 + 1$ is a positive integer, and by (iii), it is not a square; thus the Pell equation

$$x^2 - \left(-\frac{D}{4}k^2 + 1\right)y^2 = 1 \tag{4.6}$$

has infinitely many solutions in positive integers x, y. Consider solutions x, y with

$$x \geq t_0, \qquad y \geq t_0. \tag{4.7}$$

Multiplying (4.6) by $-\frac{D}{4}k^2$ we get

$$-\frac{D}{4}(kx)^2 + \frac{D}{4}\left(-\frac{D}{4}k^2 + 1\right)(ky)^2 = -\frac{D}{4}k^2,$$

whence

$$-\frac{D}{4}(kx)^2 + 1 = \left(-\frac{D}{4}k^2 + 1\right)\left(-\frac{D}{4}(ky)^2 + 1\right).$$

Since the function $\lambda(n)$ is completely multiplicative, by (4.1) it follows that

$$\lambda\left(-\frac{D}{4}(kx)^2 + 1\right) = \lambda\left(-\frac{D}{4}k^2 + 1\right)\lambda\left(-\frac{D}{4}(ky)^2 + 1\right)$$
$$= -\lambda\left(-\frac{D}{4}(ky)^2 + 1\right).$$

By (4.7) this contradicts (4.5) so that, indeed, the indirect assumption (4.2) leads to a contradiction which completes the proof of Theorem 3.

THEOREM 4. *Let* $a \in \mathbb{Z}$, $b \in \mathbb{N}$, $c \in \mathbb{Z}$ *and*

$$ab \neq c. \tag{4.8}$$

Write

$$f(n) = (n+a)(bn+c).$$

Then $\lambda(f(n))$ *assumes both values* $+1$ *and* -1 *for infinitely many* $n \in \mathbb{N}$.

Note that it follows from (4.8) that the discriminant of the polynomial $f(n)$ is $D = (ab-c)^2 > 0$.

Proof. We will prove the assertion of the theorem in several steps: first we will prove it in a special case, then we will extend it further and further, obtaining finally the result stated.

STEP 1. *Let* $A \in \mathbb{N}$ *and write*

$$g(n) = n(An+1).$$

Then $\lambda(g(n))$ *assumes both values* $+1$ *and* -1 *for infinitely many* $n \in \mathbb{N}$.

Assume that contrary to assertion, $\lambda(g(n)) = \lambda(n(An+1))$ is constant for $n \geq n_0$, with some $n_0 \in \mathbb{N}$. Since the λ function is completely multiplicative, it follows that $\lambda(An(An+1))$ is also constant for $n \geq n_0$, i.e.

$$\lambda(An(An+1)) = \varepsilon \quad \text{for } n \geq n_0 \tag{4.9}$$

(where $\varepsilon \in \{-1,+1\}$).

Now we prove by induction on i that, for all $i \in \mathbb{N}$,

$$\lambda(An+i) = \varepsilon\,\lambda(An) \quad \text{for } n \geq n_0. \tag{4.10}$$

By the multiplicativity of λ, (4.9) can be rewritten as

$$\lambda(An)\lambda(An+1) = \varepsilon.$$

Since $\lambda(An) \in \{-1,+1\}$, (4.10) follows with 1 in place of i.

Assume now that (4.10) holds with j in place of i for all $j \leq i$:

$$\lambda(An+j) = \varepsilon\,\lambda(An) \quad \text{for } j = 1,\dots,i \text{ and } n \geq n_0. \tag{4.11}$$

We have to show that it also holds with $i+1$ in place of i:

$$\lambda(An+i+1) = \varepsilon\,\lambda(An) \quad \text{for } n \geq n_0. \tag{4.12}$$

By (4.11) we have

$$\lambda((An+1)(An+i)) = \lambda(An+1)\lambda(An+i) = (\varepsilon\,\lambda(An))^2 = +1$$

or, in equivalent form,

$$\lambda(A^2n^2 + A(i+1)n + i) = +1,$$
$$\lambda(A(An^2 + (i+1)n) + i) = +1.$$

From (4.11) with $An^2+(i+1)n$ and i in place of n, resp. j, it follows that

$$\varepsilon\,\lambda(A(An^2+(i+1)n)) = \lambda(A(An^2+(i+1)n)+i) = +1,$$

whence

$$\lambda(A(An^2+(i+1)n)) = \varepsilon,$$
$$\lambda(An)\lambda(An+i+1) = \varepsilon$$

for $n \ge n_0$, which, by $\lambda(An) \in \{-1,+1\}$, proves (4.12).

By (4.10), $\lambda(m)$ is constant for $m > An_0$, which contradicts to the fact that $\lambda(n)$ assumes both -1 and $+1$ for infinitely many $n \in \mathbb{N}$, and this completes the proof of the assertion of Step 1.

Step 2. *Let $A \in \mathbb{N}$ and write*

$$h(n) = n(An-1).$$

Then $\lambda(h(n))$ assumes both values -1 and $+1$ for infinitely many $n \in \mathbb{N}$.

Again we argue by contradiction: assume that $\lambda(h(n))$ is constant for $n \ge n_1$. It follows that $\lambda(A)\lambda(h(n)) = \lambda(An(An-1))$ is also constant for $n \ge n_1$, i.e.,

$$\lambda(An(An-1)) = \varepsilon \quad \text{for } n \ge n_1 \tag{4.13}$$

(where $\varepsilon \in \{-1,+1\}$). Replace n by An^2:

$$\lambda(A^2n^2(A^2n^2-1)) = \varepsilon.$$

Then

$$\lambda(A^2n^2)\lambda(An-1)\lambda(An+1) = \varepsilon,$$
$$\lambda(An+1) = \varepsilon\,\lambda(An-1) \quad \text{for } n \ge n_1. \tag{4.14}$$

It follows from (4.13) and (4.14) that

$$\begin{aligned}\lambda(n(An+1)) &= \lambda(n)\lambda(An+1) = \varepsilon\,\lambda(n)\lambda(An-1)\\ &= \lambda(An(An-1))\lambda(n)\lambda(An-1)\\ &= \lambda(A) \quad \text{for } n \ge n_1,\end{aligned}$$

which contradicts the assertion of Step 1.

Step 3. *Let $B \in \mathbb{N}$, $C \in \mathbb{Z}$, $C \ne 0$, and write*

$$k(n) = n(Bn+C).$$

Then $\lambda(k(n))$ assumes both values -1 and $+1$ for infinitely many $n \in \mathbb{N}$.

Assume that contrary to assertion, $\lambda(k(n))$ is constant for $n \ge n_2$. It follows that

$$\begin{aligned}\lambda(k(|C|m)) &= \lambda(|C|m(B|C|m+C))\\ &= \lambda(|C|^2)\lambda\left(m\left(Bm+\frac{C}{|C|}\right)\right)\\ &= \lambda\left(m\left(Bm+\frac{C}{|C|}\right)\right)\end{aligned}$$

is also constant for $m \geq n_2$, which is impossible by Steps 1 and 2.

We are now ready to prove Theorem 4. Assume that contrary to assertion, a, b, c and $f(n)$ are defined as in the theorem, but $\lambda(f(n))$ is constant for $n \geq n_3$. Then, writing $l(m) = f(m-a)$ we have $l(m) = m(bm+(c-ab))$, and $\lambda(l(m))$ is constant for $m \geq n_3 + a$, which is impossible by Step 3, and this completes the proof of Theorem 4.

5. Further problems. The problems and results above could be extended in various directions. In particular, one might like to study general multiplicative functions $g(n)$ with $g(n) \in \{-1, +1\}$.

Conjecture 3. *If $g(n)$ is a multiplicative function with $g(n) \in \{-1, +1\}$ for all $n \in \mathbb{N}$ and such that*

$$\sum_{\substack{g(p)=-1\\ p\equiv h \pmod m}} \frac{1}{p}$$

is divergent for all $h \in \mathbb{Z}$, $m \in \mathbb{N}$, $(h, m) = 1$, then, writing

$$E_N = \{g(1), \dots, g(N)\},$$

we have

$$W(E_N) = o(N) \tag{5.1}$$

and

$$C_2(E_N) = o(N). \tag{5.2}$$

While (5.1) seems to be difficult but not hopeless, (5.2) is beyond reach at present.

Moreover, we conjecture, and Tables 1 and 2 seem to indicate, that the contrast between the complexities of the sequences $L_N(y)$ and $G_N(y)$ disappears as y grows (for fixed N), and L_N and G_N are of equally high complexity:

Conjecture 4. *If $k, N \in \mathbb{N}$, $N \to \infty$ and $k = o(\log N)$ then*

$$f(k, L_N) = 2^k \tag{5.3}$$

and

$$f(k, G_N) = 2^k. \tag{5.4}$$

As we mentioned in Section 1, we proved that assuming Schinzel's "Hypothesis H", (5.3) holds for fixed k and $N \to \infty$. However, in this general form and without unproved hypotheses, (5.3) and (5.4) seem to be beyond our reach, thus we will present certain related numerical data in the next section.

6. Numerical data. We computed $W(L_N(y))$, $C_2(L_N(y))$, $W(G_N(y))$ and $C_2(G_N(y))$ for several values of N and y. In particular, if $y > N$ then $L_N(y) = L_N$ and $G_N(y) = G_N$. Thus those lines in the tables below where $y = \infty$ appears in the second column correspond to the sequences L_N and G_N. We also studied the complexities of the sequences $L_N(y)$ and

Table 1. Correlation and complexity of $L_N(y)$

N	y	$W(L_N(y))$	$C_2(L_N(y))$	$k(L_N(y))$	$f(k, L_N(y))$
100	2	50	78	2	3
100	3	26	44	4	15
100	5	17	23	5	30
100	∞	11	19	5	31
1000	2	500	928	2	3
1000	3	251	700	4	15
1000	5	167	428	6	63
1000	7	132	221	8	253
1000	∞	46	150	8	254
10000	2	5000	9770	2	3
10000	3	2502	8439	4	15
10000	5	1667	6557	6	63
10000	7	1256	4450	9	511
10000	11	1046	2923	10	1021
10000	13	915	2015	11	2020
10000	∞	155	446	11	2032
100000	2	50000	99228	2	3
100000	3	25000	92666	4	15
100000	5	16665	76954	6	63
100000	7	12492	62248	10	1023
100000	11	10426	41762	13	8183
100000	13	8938	29760	13	8190
100000	∞	453	1380	14	16352
1000000	2	500000	997676	2	3
1000000	3	249999	967224	4	15
1000000	5	166667	878822	6	63
1000000	7	124994	737476	10	1023
1000000	11	104124	600614	13	8191
1000000	13	89287	429055	14	16383
1000000	17	79440	334077	16	65529
1000000	19	71579	268037	16	65534
1000000	∞	1423	4635	17	131011

$G_N(y)$. For a sequence $E_N \in \{-1,+1\}^N$, $k(E_N)$ denotes the smallest k value such that $f(k, E_N) < 2^k$. The values of $k(L_N(y))$ and $k(G_N(y))$ are presented in the respective tables, and in the next column the value of $f(k(L_N(y)), L_N(y))$, resp. $f(k(G_N(y)), G_N(y))$ is given.

In Table 3 we compare $f(k, L_N(y))$ and $k^{\pi(y)}$ for $y = 3$ and N large ($N = 2000000$) to illustrate Theorem 1. The ratio $f(k, L_N(3))//k^2$ does not seem to have a limit. We have also included values of $S_N(1)$ and $S_N(2)$ (defined in Lemma 1) and observe that $f(k, L_N(3))$ is almost equal to $f(k, S_N(1))f(k, S_N(2))$. We selected values of k that correspond to local extrema of $f(k, S_N(1))//k$, $f(k, S_N(2))//k$ or $f(k, L_N(3))//k^2$.

Table 2. Correlation and complexity of $G_N(y)$

N	y	$W(G_N(y))$	$C_2(G_N(y))$	$k(G_N(y))$	$f(k, G_N(y))$
100	2	50	99	2	2
100	3	18	97	3	6
100	5	11	85	5	22
100	∞	24	29	5	31
1000	2	500	999	2	2
1000	3	168	997	3	6
1000	5	101	985	5	22
1000	7	78	895	7	104
1000	∞	81	312	8	246
10000	2	5000	9999	2	2
10000	3	1668	9997	3	6
10000	5	1001	9985	5	22
10000	7	719	9895	7	104
10000	11	593	8845	9	510
10000	13	511	6123	10	1023
10000	∞	395	2054	11	2027
100000	2	50000	99999	2	2
100000	3	16668	99997	3	6
100000	5	10001	99985	5	22
100000	7	7148	99895	7	104
100000	11	5856	98845	9	510
100000	13	4966	84985	12	4062
100000	∞	1181	10445	14	16345
1000000	2	500000	999999	2	2
1000000	3	166668	999997	3	6
1000000	5	100001	999985	5	22
1000000	7	71435	999895	7	104
1000000	11	58448	998845	9	510
1000000	13	49470	984985	12	4062
1000000	17	43646	753225	15	32716
1000000	19	39068	594645	15	32767
1000000	∞	4113	38526	17	131014

For larger values of y, much larger values of N and k would be needed to observe oscillations of the ratio $f(k, L_N(y))//k^{\pi(y)}$.

Table 3. Complexity for $y = 3$

k	$f(k,S_N(1))$	$f(k,S_N(2))$	$f(k,L_N(3))$	$\frac{f(k,L_N(3))}{k^2}$	$\frac{f(k,L_N(3))}{f(k,S_N(1))f(k,S_N(2))}$
1	2	2	2	2.000	0.500
2	3	3	4	1.000	0.444
3	5	4	8	0.889	0.400
4	6	6	15	0.938	0.416
5	8	8	28	1.120	0.437
6	10	9	47	1.306	0.522
7	11	10	71	1.449	0.645
8	12	11	103	1.609	0.780
9	14	12	142	1.753	0.845
10	16	14	188	1.880	0.839
11	18	16	238	1.967	0.826
12	20	18	296	2.056	0.822
13	21	20	352	2.083	0.838
14	22	22	416	2.122	0.859
15	23	24	484	2.151	0.876
16	24	25	544	2.125	0.906
17	26	26	624	2.159	0.923
18	28	27	708	2.185	0.936
19	30	28	788	2.183	0.938
24	40	33	1240	2.153	0.939
27	43	36	1474	2.022	0.952
32	48	46	2124	2.074	0.961
45	74	72	5062	2.500	0.950
48	80	75	5752	2.497	0.958
64	96	91	8608	2.102	0.985
81	130	108	13810	2.105	0.983
96	160	138	21640	2.348	0.980
103	167	152	24930	2.350	0.982
128	192	202	38340	2.340	0.988
135	206	216	43954	2.412	0.987
192	320	273	86720	2.352	0.992
243	371	324	119666	2.027	0.995
256	384	350	133836	2.042	0.995
300	472	438	205556	2.284	0.994

References

[CFMRS] 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.

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

[Ell] P. D. T. A. Elliott, *Arithmetic Functions and Integer Products*, Springer, New York, 1985.

[HM] G. A. Hedlund and M. Morse, *Symbolic dynamics*, Amer. J. Math. 60 (1938), 815–866.

[Que] M. Queffélec, *Substitution Dynamical Systems—Spectral Analysis*, Lecture Notes in Math. 1294, Springer, New York, 1987.

[Sá] A. Sárközy, *On multiplicative arithmetic functions satisfying a linear recursion*, Studia Sci. Math. Hungar. 13 (1978), 79–104.

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

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

Institut de Mathématiques de Luminy
CNRS-UPR 9016
163 avenue de Luminy, Case 907
13288 Marseille Cedex 9, France
E-mail: cassaigne@iml.univ-mrs.fr
ferenczi@iml.univ-mrs.fr
mauduit@iml.univ-mrs.fr

Institut Élie Cartan
Université Henri Poincaré, B.P. 239
54506 Vandœuvre-lès-Nancy Cedex, France
E-mail: rivat@iecn.u-nancy.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 9.3.1999
and in revised form 8.3.2000 (3569)