JOURNAL OF NUMBER THEORY 10, 451–474 (1978)

On a Class of Relatively Prime Sequences

PAUL ERDÖS

The Hungarian Academy of Sciences, Budapest, Hungary

AND

D. E. PENNEY AND CARL POMERANCE

Department of Mathematics, The University of Georgia, Athens, Georgia 30602

Communicated by H. Zassenhaus

Received October 15, 1977

For each natural number n, let $a_0(n) = n$, and if $a_0(n),\ldots, a_i(n)$ have already been defined, let $a_{i+1}(n) > a_i(n)$ be minimal with $(a_{i+1}(n), a_0(n) \cdots a_i(n)) = 1$. Let $g(n)$ be the largest $a_i(n)$ not a prime or the square of a prime. We show that $g(n) \sim n$ and that $g(n) > n + cn^{1/2}\log(n)$ for some $c > 0$. The true order of magnitude of $g(n) - n$ seems to be connected with the fine distribution of prime numbers. We also show that "most" $a_i(n)$ that are not primes or squares of primes are products of two distinct primes. A result of independent interest comes of one of our proofs: For every sufficiently large n there is a prime $p < n^{1/2}$ with $[n/p]$ composite.

1. INTRODUCTION

In a recent paper [3], one of us (P.E.) considered the following family of sequences. For each natural number n, let $a_0(n) = n$, and if $a_0(n),\ldots, a_i(n)$ have already been defined, let $a_{i+1}(n) > a_i(n)$ be minimal with

$$(a_{i+1}(n), a_0(n) \cdots a_i(n)) = 1.$$

So, for example, if $n = 31$, then the sequence is

$$31, 32, 33, 35,\ldots,$$

where the succeeding terms, other than 13^2, 17^2, 19^2, 23^2, and 29^2, are just the primes. The following facts were established in [3]:

1. Every prime $p \geqslant n$ appears in n's sequence and every $a_i(n) > n^2$ is prime. For every prime p, there is a unique member of n's sequence divisible by p. Denote this number by $a^{(p)}(n)$.

0022-314X/78/0104-0451$02.00/0

2. Let $f_0(n)$ denote the number of $a_i(n)$ which are squares of primes and let $f_1(n)$ denote the number of remaining composite $a_i(n)$. Then

$$0 \leqslant \pi(n) - \pi(n^{1/2} - 1) - f_0(n) \leqslant f_1(n), \tag{1.1}$$

$$0 \leqslant f_1(n) \leqslant \pi(n^{1/2}). \tag{1.2}$$

Hence, $f_0(n) = \pi(n) + O(\pi(n^{1/2}))$.

3. The largest n for which every $a_i(n)$, $i > 0$, is a prime power is $n = 70$.

Also stated without proof in [3]:

4. For all sufficiently large n, some $a_i(n)$ with $i > 0$ is the product of two distinct primes.

In addition, the following two problems were raised in [3]:

5. Can one do better than (1.2) in estimating $f_1(n)$?

6. For $n > 70$, let $g_1(n)$ denote the largest $a_i(n)$ which is not a prime power. Is $g_1(n) \sim n$?

In this paper we deal with these and related questions. In particular, relevant to (4), we show that

$$f_2(n) = \pi(n^{1/2}) + O(\pi(n^{1/3})), \tag{1.3}$$

where $f_2(n)$ denotes the number of $a_i(n)$ which are products of two distinct primes (Section 2). We also show that $n = 272$ is the largest n for which no $a_i(n)$, $i > 0$, is the product of two distinct primes (Section 5).

Since $f_2(n) \leqslant f_1(n)$, (1.2) and (1.3) show that

$$f_1(n) = \pi(n^{1/2}) + O(\pi(n^{1/3})), \tag{1.4}$$

which deals with (5). Note that (1.1), (1.4), and the fact that $f_0(n) \leqslant \pi(n) - \pi(n^{1/2} - 1) - f_2(n)$ give

$$f_0(n) = \pi(n) - 2\pi(n^{1/2}) + O(\pi(n^{1/3})).$$

For each $n > 4$, let

$$g(n) \stackrel{\text{def}}{=} \max\{a^{(p)}(n)\colon \text{primes } p < n^{1/2}\}.$$

Then $g(n)$ is the largest $a_i(n)$ that is neither a prime nor the square of a prime. In our opening example we have $g(31) = 35$. Note that $g_1(n) \leqslant g(n)$ for all $n > 70$. We show that $g(n) \sim n$, which answers (6) affirmatively (Section 2). We conjecture that $g_1(n) = g(n)$ for all sufficiently large n. In fact we conjecture $g(n)$ is the product of two distinct primes but for finitely many n

(Section 6). Our numerical work suggests that 118 is the largest value of n for which $g_1(n) < g(n)$ and that 1478 is the largest value of n for which $g(n)$ is not in the form pq. We prove that $g(n) < 2n$ for all but 10 exceptional values of n, the largest being $n = 371$ (Section 5).

Many of these problems seem to be intimately connected with some deep questions in the distribution of primes. For example, we show a relationship between $g(n) - n = o(n)$ and the order of magnitude of the error term in the prime number theorem (Section 2). In addition, the above-mentioned question on whether every sufficiently large $g(n)$ is in the form pq is related to the order of magnitude of the difference between consecutive primes.

We show how a certain result of Selberg [13], which says that the distribution of primes in very small intervals is "usually well-behaved," shows that the set of n for which some $a_i(n)$ is a prime power with exponent at least 3 has density 0 (Section 5). We use a new result of Warlimont [15] that is similar to Selberg's theorem to show that the set of values of $g(n)$ has density 0 (Section 6).

We show that the asymptotic density $d(t)$ of the set of n for which $n + t = a_i(n)$ for some i exists and that $d(t) \sim e^{-\gamma}/\log t$, where γ is Euler's constant (Section 4). The proof uses a result of Hooley [7] on the mean square of the differences of the members in a reduced residue system modulo an integer.

We prove that $(g(n) - n)/n^{1/2} \to \infty$ (Section 3). Our proof uses the upper bound obtained from Brun's method for the number of representations of a number as a sum of two primes.

Many of the theorems, arguments, and conjectures of this paper carry over almost intact to the family of sequences $\{b_i(n)\}$, where $b_0(n) = n > b_1(n) > \cdots > b_i(n)$ and $b_{i+1}(n) < b_i(n)$ is maximal with $(b_{i+1}(n), b_0(n) \cdots b_i(n)) = 1$. This family of sequences is studied in a forthcoming paper of Eggleton, Erdös, and Selfridge. Some other somewhat related papers are those by Erdös and Selfridge [4, 5] and Eggleton *et al.* [2].

2. Upper Bounds for $g(n)$

THEOREM 2.1. $g(n) \sim n$.

Proof. Let $\epsilon > 0$ be arbitrary and let $p < n^{1/2}$ be a prime. We now show that if

$$\pi((1 + \epsilon)\, n/p) - \pi(n/p) \geqslant \pi(p) - \pi(p/(1 + \epsilon)) + \pi(p^{1/2}) \tag{2.1}$$

holds, then $a^{(p)}(n) \leqslant (1 + \epsilon)n$. Let $q_1, \ldots, q_s$ be the primes in $(n/p, (1 + \epsilon)\, n/p]$, $p_1, \ldots, p_t$ the primes in $(p/(1 + \epsilon), p)$, and $r_1, \ldots, r_u$ the primes below $p^{1/2}$. Then (2.1) implies $s > t + u$. For $1 \leqslant i \leqslant s$, consider pq_i. If $pq_i = a^{(p)}(n)$, then $a^{(p)}(n) \leqslant (1+\epsilon)\, n$. So say no $pq_i = a^{(p)}(n)$. Then for each i, there is a number $j(i)$ with $a_{j(i)}(n) < pq_i$ and $(a_{j(i)}(n), pq_i) > 1$. If $p \mid a_{j(i)}(n)$, then

$$a^{(p)}(n) = a_{j(i)}(n) < pq_i \leqslant (1+\epsilon)\, n.$$

So say no $a_{j(i)}(n)$ is a multiple of p. Then $q_i \mid a_{j(i)}(n)$ for each i. Now $a_{j(i)}(n)/q_i \in (p/(1+\epsilon), p)$, so if $a_{j(i)}(n)/q_i$ is a prime, it is one of $p_1, \ldots, p_t$. If it is composite, it is divisible by one of $r_1, \ldots, r_u$. Hence there can be at most $t+u$ choices for $a_{j(i)}(n)$. But $i \to j(i)$ is one-to-one, since if $q_i q_{i'} \mid a_{j(i)}(n)$, then $a_{j(i)}(n) > pq_i$, a contradiction. Thus there are at most $t+u$ choices for i, contradicting $s > t+u$. Hence if (2.1) holds, $a^{(p)}(n) \leqslant (1+\epsilon)n$.

Thus $g(n) \leqslant (1+\epsilon)n$ will follow if we can prove (2.1) holds for every prime $p < n^{1/2}$. Now by the prime number theorem we have

$$\begin{aligned} \pi((1+\epsilon)\, n/p) - \pi(n/p) &> (\epsilon - \epsilon^2/4)\, n/(p \log(n/p)) \\ &> (\epsilon - \epsilon^2/4)\, n^{1/2}/\log n^{1/2} \end{aligned}$$

for all sufficiently large n. We also have

$$\begin{aligned} \pi(p) - \pi(p/(1+\epsilon)) &< (\epsilon - \epsilon^2/2)\, p/\log p \\ &< (\epsilon - \epsilon^2/2)\, n^{1/2}/\log n^{1/2}, \\ \pi(p^{1/2}) < p^{1/2} < n^{1/4} &< \epsilon^2 n^{1/2}/(4 \log n^{1/2}) \end{aligned}$$

for all sufficiently large p and n. Hence there is a p_0 so that (2.1) holds for all sufficiently large n and all p with $p_0 < p < n^{1/2}$. But for $p \leqslant p_0$, the right side of (2.1) is bounded, so (2.1) holds for all sufficiently large n and all $p < n^{1/2}$. As we have seen this implies that $a^{(p)}(n) \leqslant (1+\epsilon)n$, and so $g(n) \leqslant (1+\epsilon)n$.

We recall now that $f_2(n)$ is the number of $a_i(n)$ that are products of two distinct primes. We have the following.

COROLLARY. *For each $\epsilon > 0$, there is an $n_0(\epsilon)$ so that for all $n > n_0(\epsilon)$,*

$$0 \leqslant \pi(n^{1/2}) - f_2(n) < (3+\epsilon)\, \pi(n^{1/3}). \tag{2.2}$$

Proof. Let $n_1(\epsilon)$ be such that for all $n > n_1(\epsilon)$, $g(n) < (1+\epsilon)n$. Let $n > n_1(\epsilon)$. For each prime $r < n^{1/3}$, we have $a^{(r)}(n)$ divisible by no more than two primes in $I = ((1+\epsilon)\, n^{1/3}, n^{1/2}]$, since $r((1+\epsilon)\, n^{1/3})^3 > rn > g(n) \geqslant a^{(r)}(n)$. Hence with at most $2\pi(n^{1/3})$ exceptions, for every prime $p \in I$, $a^{(p)}(n)$ is not divisible by any prime $r < n^{1/3}$. Then for these p, $a^{(p)}(n)$ is the product of two distinct primes. Indeed,

$$a^{(p)}(n)/p \leqslant g(n)/p < (1+\epsilon)\, n/p < n^{2/3},$$

so that $a^{(p)}(n)/p$ is prime. Thus (2.2) now follows from the fact that $\pi((1 + \epsilon) n^{1/3}) < (1 + \epsilon) \pi(n^{1/3})$ for all sufficiently large n.

In Section 5 we show that the largest n for which no $a_i(n)$, $i > 0$, is the product of two distinct primes is $n = 272$. Now, however, we return to the topic of this section. The following theorem shows we can get good upper bounds for $a^{(p)}(n)$ provided p "keeps its distance" from $n^{1/2}$.

THEOREM 2.2. *For every $\delta > 0$, $\epsilon > 0$, there is an $n_0(\delta, \epsilon)$ such that for all $n > n_0(\delta, \epsilon)$ and primes $p < (1 - \delta) n^{1/2}$, we have*

$$a^{(p)}(n) < n + p(n/p)^{7/12+\epsilon} < n + n^{19/24+\epsilon}. \tag{2.3}$$

Proof. Let $\tau = (n/p)^{-5/12+\epsilon}$. It follows from the proof of Theorem 2.1 that (2.3) will hold if we have

$$\pi((1 + \tau) n/p) - \pi(n/p) \geqslant \pi(p) - \pi(p/(1 + \tau)) + \pi(p^{1/2}). \tag{2.4}$$

We now use a recent result of Huxley [8] that, when combined with results of Hoheisel and Tchudakoff as reported by Ingham [9], yields

$$\pi(x + x^{\theta}) - \pi(x) \sim x^{\theta}/\log x \qquad \text{as} \quad x \to \infty \tag{2.5}$$

if $\theta > 7/12$. Now $(1 + \tau) n/p = n/p + (n/p)^{7/12+\epsilon}$, so that (2.5) implies

$$\pi((1 + \tau) n/p) - \pi(n/p) \sim \tau n/(p \log(n/p)) \qquad \text{as} \quad n/p \to \infty.$$

Then using $p < (1 - \delta) n^{1/2}$, we have

$$\pi((1 + \tau) n/p) - \pi(n/p) > (1 - \delta/2) \tau n/(p \log(n/p)) \tag{2.6}$$

for all $n > n_1(\delta, \epsilon)$. Now using $(1 + \tau)^{-1} > 1 - \tau$ and (2.5), we have

$$\pi(p) - \pi(p/(1 + \tau)) \leqslant \pi(p) - \pi((1 - \tau)p) \sim \tau p/\log p \qquad \text{as} \quad p \to \infty.$$

Thus since $p < (1 - \delta) n^{1/2}$, we have

$$\pi(p) - \pi(p/(1 + \tau)) < (1 + \delta) \tau p/\log p < (1 - \delta) \tau n/(p \log(n/p)) \tag{2.7}$$

for all $p > p_1(\delta, \epsilon)$. Now using the trivial estimate $\pi(p^{1/2}) < p^{1/2} < n^{1/4}$ and the fact that $\tau n/p > n^{7/24}$, we have from (2.6), (2.7),

$$\pi((1 + \tau) n/p) - \pi(n/p) - \pi(p) + \pi(p/(1 + \tau)) - \pi(p^{1/2}) > \tau \, \delta n/(4p \log(n/p))$$

for all $n > n_2(\delta, \epsilon)$, $p > p_1(\delta, \epsilon)$; (2.4) follows for these p, n. Now for $p \leqslant p_1(\delta, \epsilon)$, the right side of (2.4) is bounded. Hence it follows from (2.6) that there is an $n_0(\delta, \epsilon)$ such that for all $n > n_0(\delta, \epsilon)$ and all $p < (1 - \delta) n^{1/2}$, we have (2.4).

Remark 2.1. It is known that if the Riemann hypothesis holds, then (2.5) is true for all $\theta > \frac{1}{2}$. Hence on the Riemann hypothesis, we have: For each $\delta > 0$, $\epsilon > 0$, there is an $n_0(\delta, \epsilon)$ such that for all $n > n_0(\delta, \epsilon)$ and all primes $p < (1 - \delta) n^{1/2}$,

$$a^{(p)}(n) < n + p(n/p)^{1/2+\epsilon} < n + n^{3/4+\epsilon}. \tag{2.8}$$

We remark that even if (2.5) is true for some $\theta < \frac{1}{2}$, we cannot by our method improve (2.8) very much. This is due to the term $\pi(p^{1/2})$ in (2.4) which would no longer be negligible.

THEOREM 2.3. *For each $\epsilon > 0$ there is an $n_0(\epsilon)$ so that for all $n > n_0(\epsilon)$ and primes $p < n^{5/17-\epsilon}$, we have $a^{(p)}(n) \leqslant pq$, where q is the first prime above n/p.*

Proof. By (2.5) we have $q < n/p + (n/p)^{7/12+\epsilon}$, all $n > n_0(\epsilon)$. Then by a simple calculation we have

$$pq - n < p(n/p)^{7/12+\epsilon} < n/p < q.$$

Hence $(p - q)\, q < n$, so that q divides no $a_i(n) < pq$. Thus $a^{(p)}(n) \leqslant pq$.

Remark 2.2. If the Riemann hypothesis is valid, the conclusion of Theorem 2.3 is true for all primes $p < n^{1/3-\epsilon}$. Moreover, from the conjecture of Cramér [1] (in slightly weaker form),

$$\limsup_{n\to\infty} (p_{n+1} - p_n)/\log^2 n < \infty. \tag{2.9}$$

where p_n denotes the nth prime, we have the conclusion of Theorem 2.3 true for all $p < cn^{1/2}/\log n$ and all $n > 1$, where $c > 0$ is an absolute constant. Thus for these p we would have

$$a^{(p)}(n) - n \ll p \log^2 n \ll n^{1/2} \log n. \tag{2.10}$$

We now turn to an improvement of Theorem 2.1. Let $E(x)$ be a concave function for all $x > x_0$ such that

$$|\pi(x) - li(x)| \leqslant E(x) \qquad \text{for all} \quad x > x_0 .$$

We omit the details, but following the proof of Theorem 2.1 for the case $p > n^{1/2}/2$ and the proof of Theorem 2.2 for the case $p \leqslant n^{1/2}/2$, we have the following.

THEOREM 2.4. *Given E as above, there is a constant c such that for all $n > 4$ we have*

$$g(n) < n + cn^{3/4}(\log n)^{1/2} (E(n^{1/2}))^{1/2}.$$

Remark 2.3. Since it is known [14, Chap. V] that we can take

$$E(x) = c_1 x \cdot \exp(-c_2(\log x)^{3/5} (\operatorname{loglog} x)^{-1/5}),$$

where c_1, $c_2 > 0$ are constants, we have for all $n > 4$

$$g(n) < n + c_3 n \cdot \exp(-c_4(\log n)^{3/5} (\operatorname{loglog} n)^{-1/5}), \tag{2.11}$$

where c_3, $c_4 > 0$ are constants. If the Riemann hypothesis is true, it would follow that we can take $E(x) = cx^{1/2} \log x$, so that in this case we would have

$$g(n) - n \ll n^{7/8} \log n. \tag{2.12}$$

It is known [10] that

$$E(x) \neq o(x^{1/2} \operatorname{logloglog} x/\log x),$$

so no improvement in the error term in the prime number theorem could establish by our methods that $g(n) - n \ll n^{7/8}$.

Remark 2.4. Although we cannot do better than (2.11) for all n, one might try to do better for infinitely many n. In particular, is

$$\liminf_{n\to\infty} \log(g(n) - n)/\log n < 1?$$

At present we cannot answer this question (see Remark 3.2 and Fig. 1).

Remark 2.5. Let $\mathcal{Cl}$ be the set of all subsets A of the natural numbers such that the terms in A are pairwise relatively prime and such that each prime divides some member of A. For each $n > 4$ let

$$g(A, n) \overset{\text{def}}{=} \max\{|a - n| : a \in A \text{ and } \exists \text{ prime } p < n^{1/2} \ni p \mid a\},$$

$$G(n) \overset{\text{def}}{=} \min\{g(A, n) : A \in \mathcal{Cl}\}.$$

For each n, let $A(n) = \{a_0(n), a_1(n),...\}$. Then $A(n) \in \mathcal{Cl}$ and $g(A(n), n) = g(n) - n$ if $n > 4$. Hence $g(n) - n \geq G(n)$. Thus from (2.11) we have

$$G(n) \ll n \cdot \exp(-c_4(\log n)^{3/5} (\operatorname{loglog} n)^{-1/5}). \tag{2.13}$$

We cannot do better than (2.13), not even for infinitely many n. From (2.12) we would have $G(n) \ll n^{7/8} \log n$ if the Riemann hypothesis holds. But we conjecture that $G(n) \ll n^{1/2+\epsilon}$ for every $\epsilon > 0$ (compare with Remark

3.2). It will follow from the proof of Theorem 3.1 that there is a constant $c > 0$ such that

$$G(n) > cn^{1/2} \log n$$

for all sufficiently large n.

3. Lower Bounds for $g(n)$

Because of the many constants in this section, we have numbered them $c_1, c_2, \ldots$. From the corollary to Theorem 2.1 we easily obtain

$$g(n) > n + c_1 n^{1/2}/\log n \tag{3.1}$$

for all large n where $c_1 > 0$ is a constant. The following short argument removes the "$\log n$": Let $\epsilon > 0$ be small and suppose that $g(n) < n + \epsilon n^{1/2}$. The set $\{a^{(p)}(n): p \leqslant n^{1/2}\}$ lies in $[n, n + \epsilon n^{1/2})$ and has cardinality asymptotic to $\pi(n^{1/2})$ (again using the corollary to Theorem 2.1). Delete from this set those $a^{(p)}(n)$ with $p \leqslant n^{1/4}$. The cardinality of the resulting set is still asymptotic to $\pi(n^{1/2})$. Also, this set still lies in $[n, n + \epsilon n^{1/2})$ and its members are not divisible by any prime up to $n^{1/4}$. By Brun's method, an upper bound for its cardinality is $c_2 \epsilon n^{1/2}/\log n^{1/2}$. Hence we cannot choose $\epsilon < 1/c_2$. This proves that

$$g(n) > n + c_3 n^{1/2}$$

for all large n, where $c_3 > 0$ is a constant.

We now show that $(g(n) - n)/n^{1/2}$ tends to infinity.

Theorem 3.1. *There is a constant $c_4 > 0$ such that for all large n*

$$g(n) > n + c_4 n^{1/2} \log n.$$

Proof. Let n be large. Consider the function

$$F(x) = x + [n/x] - [2n^{1/2}]$$

defined for integers $x \in [1, n^{1/2})$. Then $F(x)$ is integer valued and decreasing (but not strictly). Say $j = F(x_0) > F(x_0 + 1)$. Then define $b_j = x_0 + 1/2$. We have

$$x + [n/x] = [2n^{1/2}] + j \qquad \text{for integers} \quad x \in (b_{j+1}, b_j). \tag{3.2}$$

Let m be maximal, so that

$$n^{1/2} > b_1 > b_2 > \cdots > b_m > n^{1/2} - n^{1/3}.$$

We note that

$$m \sim n^{1/6}, \qquad b_j - b_{j+1} \sim \tfrac{1}{2}n^{1/4}j^{-1/2} \qquad \text{for} \quad 1 \leqslant j \leqslant m-1, \\ n^{1/2} - b_1 \ll n^{1/4}, \qquad b_m - n^{1/2} + n^{1/3} \ll n^{1/6}. \tag{3.3}$$

Let $\epsilon > 0$ be small and let $t = [\epsilon \log n]$. Assume that $g(n) < n + (t-1)\, n^{1/2}$. We shall show that this assumption leads to a contradiction. For every prime $p \in (n^{1/2} - n^{1/3}, n^{1/2})$ with $p \nmid n$, we have

$$a^{(p)}(n) = p([n/p] + i) \qquad \text{for some } i, \quad 1 \leqslant i \leqslant t. \tag{3.4}$$

We now consider a subset S of these primes: S is the set of primes p for which $p \nmid n$, $a^{(p)}(n)/p$ is prime, and $b_m < p < b_1$. If $a^{(p)}(n)/p$ is composite, it is divisible by a prime $q < 2n^{1/4}$, so the number of such p is less than $n^{1/4}$. Then from (2.5) and (3.3) we have

$$|S| = (2 + o(1))\, n^{1/3}/\log n, \tag{3.5}$$

where $|S|$ denotes the cardinality of S.

Let $1 \leqslant j \leqslant m-1$ and suppose $p \in (b_{j+1}, b_j) \cap S$ is prime. Then by (3.4) we have $q = [n/p] + i$ prime for some i, $1 \leqslant i \leqslant t$, and so by (3.2),

$$p + q = [2n^{1/2}] + i + j. \tag{3.6}$$

By Brun's method we have for fixed j, i that the number of primes $p \in (b_{j+1}, b_j)$ for which there is a prime q satisfying (3.6) is at most

$$\frac{c_5(b_j - b_{j+1})}{\log^2(b_j - b_{j+1})} \cdot \frac{[2n^{1/2}] + i + j}{\varphi([2n^{1/2}] + i + j)},$$

where c_5 is an absolute constant. Hence using (3.3) we have

$$|S| < \frac{(18 + o(1))\, c_5 n^{1/4}}{\log^2 n} \sum_{i=1}^{t} \sum_{j=1}^{m-1} \frac{[2n^{1/2}] + i + j}{j^{1/2}\varphi([2n^{1/2}] + i + j)}. \tag{3.7}$$

Similar to the old result of Landau, $\sum_{s \leqslant x} 1/\varphi(s) \sim c_6 \log x$, we can prove

$$\sum_{s \leqslant x} s/\varphi(s) = c_6 x + O(\log x).$$

Hence (using $m \sim n^{1/6}$)

$$\sum_{j=1}^{m-1} \frac{[2n^{1/2}] + i + j}{j^{1/2}\varphi([2n^{1/2}] + i + j)} = (c_6 m + O(\log n))/(m-1)^{1/2} \\ + \frac{1}{2}\int_1^{m-1} y^{-3/2} \sum_{j=1}^{y} \frac{[2n^{1/2}] + i + j}{\varphi([2n^{1/2}] + i + j)}\, dy$$

$$= c_6 m^{1/2} + O\left(\int_1^{m^{1/2}} y^{-1/2} \operatorname{loglog} n \, dy\right)$$

$$+ \frac{c_6}{2} \int_{m^{1/2}}^{m-1} y^{-1/2}\, dy + O\left(\int_{m^{1/2}}^{m} y^{-3/2} \log n \, dy\right)$$

$$\sim 2c_6 m^{1/2}.$$

Thus from (3.7) we have

$$|S| < \frac{(36 + o(1))\, c_5 c_6 n^{1/4}}{\log^2 n} \sum_{i=1}^{t} m^{1/2},$$

so that there is a constant c_7 with

$$|S| < c_7 t n^{1/3}/\log^2 n \leqslant c_7 \epsilon n^{1/3}/\log n.$$

This upper bound contradicts (3.5) if ϵ is sufficiently small.

Hence our assumption that $g(n) < n + (t - 1)\, n^{1/2}$ is false, and so

$$g(n) \geqslant n + (t - 1)\, n^{1/2} > n + \tfrac{1}{2}\epsilon n^{1/2} \log n.$$

This proves our theorem.

Remark 3.1. Using a method similar to the proof of Theorem 3.1 we can show that if $\epsilon > 0$ is sufficiently small then a positive proportion of the primes in $(n^{1/2} - n^{1/3}, n^{1/2})$ do not have a multiple in $[n, n + \epsilon n^{1/2})$. No doubt this is true for a positive proportion of all the primes up to $n^{1/2}$. That is, we conjecture that for a positive proportion $p(\epsilon)$ of $p < n^{1/2}$ we have $n/p - [n/p] \geqslant \epsilon$, where $\epsilon > 0$ is fixed, but small. In fact we conjecture that $p(\epsilon)$ is continuous, monotonic, and that $p(0+) = 1$, $p(1-) = 0$. The same should be true if we replace "$n^{1/2}$" in the definition of $p(\epsilon)$ with "n^c" for any c with $0 < c < \frac{1}{2}$. The method of proof of Theorem 3.1 also demonstrates that for almost all primes $p \in (n^{1/2} - n^{1/3}, n^{1/2})$, we have $[n/p]$ composite. We conjecture that, except for $o(\pi(n^{1/2}))$ primes $p < n^{1/2}$, $[n/p]$ is composite.

Theorem 3.1 is not best possible for all n. Indeed we have

THEOREM 3.2. *There is a constant $c_8 > 0$ such that, for infinitely many n, we have*

$$g(n) > n + c_8 n^{1/2} (\log n)(\operatorname{loglog} n)(\operatorname{logloglogloglog} n)(\operatorname{logloglog} n)^{-2}.$$

Proof. From Rankin [11], we know that for each $r > e^{e^e}$, there is a sequence of at least

$$\alpha(r) \overset{\text{def}}{=} c_9 r (\log^2 r)(\operatorname{loglogloglog} r)(\operatorname{loglog} r)^{-2}$$

consecutive integers, each divisible by one of the first r primes. Let $m_r \in [0, P_r]$ be 1 less than the first member of Rankin's string, where P_r denotes the product of the first r primes. Then

$$(m_r, P_r) = 1, \qquad (m_r + i, P_r) > 1 \qquad \text{for} \quad 1 \leqslant i \leqslant \alpha(r). \tag{3.8}$$

Let p denote the first prime in the arithmetic progression $m_r \pmod{P_r}$ for which $p > P_r$. Then p is either the first or second prime in the progression. By a theorem of Fogels [6] generalizing Linnik's well-known work, there is an absolute constant c_{10} so that $p < P_r^{c_{10}}$. Let x be such that if

$$n = p^2 + x$$

then $P_r \mid n$ and $0 < x < P_r < p$. Finally, let i be such that $a^{(p)}(n) = p(p + i)$. Since $p \equiv m_r \pmod{P_r}$ and $P_r \mid n$, we see by (3.8) that $i > \alpha(r)$. Hence,

$$g(n) \geqslant a^{(p)}(n) > p^2 + p\alpha(r) > n + \tfrac{1}{2}n^{1/2}\alpha(r). \tag{3.9}$$

Now note that $\log P_r \sim r \log r$, so that $n < 2p^2 < 2P_r^{2c_{10}}$ implies

$$r > c_{11} \log n/\operatorname{loglog} n, \tag{3.10}$$

where $c_{11} > 0$ is a constant. Our theorem follows from (3.9) and (3.10).

Remark 3.2. Let

$$h(n) \stackrel{\text{def}}{=} \log(g(n) - n)/\log n.$$

From Theorem 3.1 we have $h(n) > \frac{1}{2}$ for all but finitely many n. We can show that the problem of finding the largest n for which $h(n) \leqslant \frac{1}{2}$ is effectively computable. We have not rigorously determined this value of n, but our numerical work suggests that it is 1331. We have computed $h(n)$ for every $n \leqslant 27{,}500$ and for many other larger values of n. These data suggest that $n = 4366$ is the largest n for which $h(n) < 0.6$. The largest n we found with $h(n) > 0.95$ was $n = 12{,}834$. We cannot prove $h(n)$ has any limit points exceeding $\frac{1}{2}$ (compare with Remark 2.4). All we know for sure is that all the limit points of $h(n)$ lie in $[\frac{1}{2}, 1]$. From (2.12) we would have all limit points in $[\frac{1}{2}, \frac{7}{8}]$ if the Riemann hypothesis is true. Let

$$\alpha = \liminf_{n \to \infty} h(n), \qquad \beta = \limsup_{n \to \infty} h(n).$$

Our numerical work suggests that $\alpha > \frac{1}{2}$; perhaps α is as large as $\frac{3}{4}$. We also believe that β is near $\frac{7}{8}$. For $\alpha \leqslant x \leqslant \beta$, let $\delta(x)$ denote the asymptotic density of the set of n for which $h(n) \leqslant x$. We conjecture that $\delta(x)$ exists for each x, $\delta(x)$ is monotonic and continuous, and that $\delta(\alpha) = 0$, $\delta(\beta) = 1$. In Fig. 1, we have three numerical experiments recorded that may approximate the graph of $\delta(x)$.

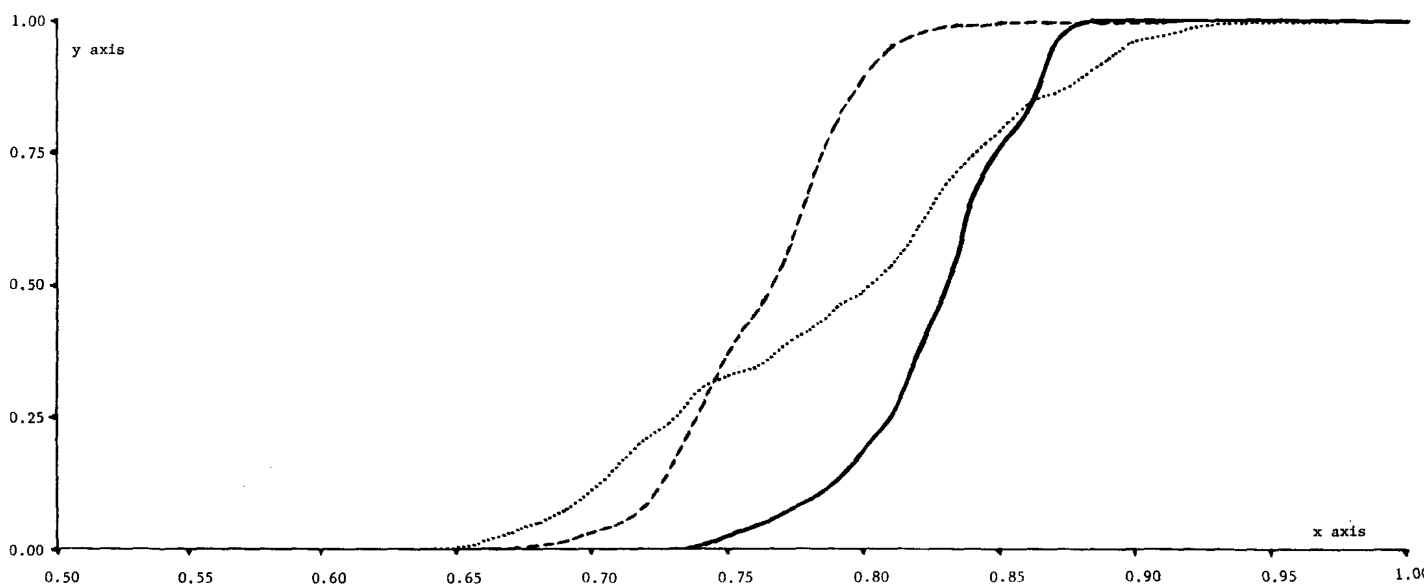

FIG. 1. Cumulative distributions of $h(n)$ in three experiments.

EXPERIMENT 1. We computed $h(n)$ for every $n \in (10^4, 2 \times 10^4] = I_1$. A point (x, y) on the dotted curve in Fig. 1 means that $y/10^4$ of the $n \in I_1$ have $h(n) \leqslant x$.

EXPERIMENT 2. We computed $h(n)$ for one random n in each consecutive subinterval of length 10 in $(10^5, 2 \times 10^5] = I_2$. The dashed curve represents the approximate distribution of $h(n)$ for $n \in I_2$.

EXPERIMENT 3. We computed $h(n)$ for one random n in each consecutive subinterval of length 1000 in $(10^6, 2 \times 10^6] = I_3$. The solid curve in Fig. 1 represents the approximate distribution of $h(n)$ for $n \in I_3$.

4. THE SIZE OF $a^{(p)}(n)$ FOR FIXED p

We now say a word about fixed p: If p is a fixed prime, what can be said about $a^{(p)}(n)$ as $n \to \infty$? If $p = 2$, we meet with immediate success, for $a^{(2)}(n) = n$ or $n + 1$. But already for $p = 3$ we have a difficult problem. It is clear from Theorem 2.2 that $a^{(3)}(n) < n + n^{7/12+\epsilon}$ for every $\epsilon > 0$ and every $n > n_0(\epsilon)$. Moreover if Cramér's conjecture is true, Remark 2.2 gives us $a^{(3)}(n) - n \ll \log^2 n$. On the other hand, if n is the product of the primes $p \leqslant x$ with $p \equiv 2 \pmod 3$ and if $n \equiv 1 \pmod 3$ then $a^{(3)}(n) > n + x$. This proves there are infinitely many n for which

$$a^{(3)}(n) > n + c \log n,$$

where $c > 0$ is a constant. No doubt this can be improved slightly using a Rankin-type result, as in Theorem 3.2. These comments for the case $p = 3$ can be generalized easily for any odd prime.

For each integer $t \geqslant 0$, let $M_t = P_{\pi(t)}$ denote the product of the primes up to t. Let p be an odd prime. It is possible to determine whether $a^{(p)}(n) > n + t$ solely by considering to what class n belongs modulo pM_t. Moreover, there is at least one of these classes for which $a^{(p)}(n) > n + t$, namely, if $p \mid n - 1$ and n is divisible by every other prime up to t. Hence $D(p, t)$, the asymptotic density of the set of n for which $a^{(p)}(n) > n + t$, exists and is positive. In the next theorem we shall in addition insist that n is in a fixed residue class modulo p.

THEOREM 4.1. *For each two integers $t \geqslant 2$, a, and each odd prime p, let $D(p, t, a)$ denote the asymptotic density of the set of $n \equiv a \pmod p$, with $a^{(p)}(n) > n + t$. Then there is an absolute constant c such that*

$$D(p, t, a) < c \log^2 t/t.$$

We shall use the following:

LEMMA. *There is an absolute constant c' such that for any $y \geqslant 1$, $t \geqslant 2$, we have*

$$f(y, t) < c'(\log^2 t)/y.$$

Here $f(y, t)$ denotes the asymptotic density of the set of n with each $n + i$, $0 \leqslant i \leqslant y$, divisible by a prime not exceeding t.

Proof. Let $N = M_t$ and let $1 = b_1 < b_2 < \cdots < b_{\varphi(N)} = N - 1$ be the integers in $[1, N]$ relatively prime to N. Say $(n + i, N) > 1$ for each i, $0 \leqslant i \leqslant y$. Then every $n' \equiv n \pmod N$ has the same property, so we shall assume that $0 < n < N$. It is clear that $1 < n < N - 1$, so that there is some j with $b_j < n < b_{j+1}$. Hence, $b_{j+1} - b_j > b_{j+1} - n > y$. Thus the number of such $n \leqslant N$ is less than

$$\sum{}' b_{j+1} - b_j < \frac{1}{y} \sum{}' (b_{j+1} - b_j)^2 \leqslant \frac{1}{y} \sum_{j=1}^{\varphi(N)-1} (b_{j+1} - b_j)^2 \leqslant \frac{c''}{y} N(\operatorname{loglog} N)^2,$$

where $\sum'$ denotes the sum over all $b_{j+1} - b_j > y$ and where for the last inequality, we use a theorem of Hooley [7] (c'' is an absolute constant). The lemma now follows, since $\operatorname{loglog} N \sim \log t$.

We note that Hooley's is not the best result known on the mean square gaps in a reduced residue system. Certain improvements have been obtained independently by Hausman and Shapiro [16] and Norton [18]. These improvements, however, do not appear to be of help in a possible strengthening of the lemma.

Proof of Theorem 4.1. Since we trivially have $D(p, t, a) \leqslant 1/p$, we may assume that $t \geqslant 3p$. Say $a^{(p)}(n) > n + t$. Let $m = [n/p] + 1$ and let $m + y = [(n + t)/p]$. Then $y \geqslant 1$. Say that for some i, $1 \leqslant i \leqslant y$, $m + i$ is divisible by no prime up to t. Then $a^{(p)}(n) \leqslant (m + i)p \leqslant n + t$, contradicting our assumption. Thus by the lemma,

$$D(p, t, a) < \frac{1}{p} \cdot \frac{c' \log^2 t}{y} < \frac{1}{p} \cdot \frac{c' \log^2 t}{t/p - 2} \leqslant \frac{3c' \log^2 t}{t},$$

which proves our theorem.

Remark 4.1. Theorem 4.1 implies that for each prime p, $D(p, t) < cp(\log^2 t)/t$, so that $\lim_{t\to\infty} D(p, t) = 0$. Thus for each $\epsilon > 0$, there is a $t = t(\epsilon, p)$ such that $a^{(p)}(n) \leqslant n + t$ but for a set of n of density at most ϵ.

THEOREM 4.2. *Let $d(t)$ denote the asymptotic density of the set of n for which $a_i(n) = n + t$ for some i. Then (γ is Euler's constant)*

$$d(t) \sim e^{-\gamma}/\log t \quad \text{as} \quad t \to \infty.$$

Proof. We first observe that whether $a_i(n) = n + t$ for some i is determined solely by what class n is in modulo M_t, so that $d(t)$ exists. Moreover, if $(n + t, M_t) = 1$, then $n + t = a_i(n)$ for some i. Hence

$$d(t) \geqslant \prod_{p \leqslant t} (1 - 1/p) \sim e^{-\gamma}/\log t$$

by Mertens' theorem. Now assume that $n + t = a_i(n)$ for some i. Either $n + t$ is divisible by a prime $p < t/\log^3 t$ or not. For p in the former case we have, by definition, $a^{(p)}(n) = n + t$. Hence by Theorem 4.1 the asymptotic density of such n is at most

$$D(p, t - 1, -t) < c \log^2(t - 1)/(t - 1).$$

Then summing over $p < t/\log^3 t$ and considering n in the latter case,

$$d(t) < \frac{c \log^2(t-1)}{t-1} \cdot \pi\left(\frac{t}{\log^3 t}\right) + \prod_{p \leqslant t/\log^3 t} (1 - 1/p) \sim \frac{e^{-\gamma}}{\log t},$$

since the first term is $O(1/\log^2 t)$. This completes the proof of our theorem.

Remark 4.2. Even though Theorem 4.2 shows that $d(t) \to 0$ as $t \to \infty$, the local behavior of $d(t)$ is probably irregular. The values of $d(t)$ for $0 \leqslant t \leqslant 11$ are 1, 1, 1/2, 1/3, 1/3, 2/5, 4/15, 2/7, 2/7, 1/3, 4/15, 2/11.

5. Multiplicative Properties of the $a_i(n)$

Our first goal is to show that $n = 272$ is the largest n such that $a_i(n)$ is never the product of two distinct primes for all $i > 0$. To show this, the following result of independent interest will be useful.

Theorem 5.1. *The set of n for which* $g(n) > 2n$ *is* {10, 27, 51, 52, 151, 152, 170, 367, 368, 371}.

Proof. From the proof of Theorem 2.1 we see that $g(n) \leqslant 2n$ will follow if for every prime $p < n^{1/2}$ we have

$$\pi(n/p) + \pi(p) + \pi(p^{1/2}) \leqslant \pi(2n/p) + \pi(p/2). \tag{5.1}$$

We consider separately the following cases:

(i) $9n^{1/2}/10 \leqslant p < n^{1/2}$,

(ii) $3n^{1/2}/4 \leqslant p < 9n^{1/2}/10$,

(iii) $3 < p < 3n^{1/2}/4$,

(iv) $p \leqslant 3$.

We suppress the details, but using the estimates

$$\frac{x}{\log x - \frac{1}{2}} < \pi(x) \qquad \text{for all} \quad x \geqslant 67, \tag{5.2}$$

$$\pi(x) < \frac{x}{\log x}\left(1 + \frac{3}{2 \log x}\right) \qquad \text{for all} \quad x > 1, \tag{5.3}$$

due to Rosser and Schoenfeld [12], we are able to show that (5.1) holds for all p in case (i) if $n \geqslant 13{,}111$, all p in case (ii) if $n \geqslant 3476$, all p in case (iii) if $n \geqslant 311$, and all p in case (iv) if $n \geqslant 17$. Hence, we conclude $g(n) \leqslant 2n$ for all $n \geqslant 13{,}111$. A computer check up to this point reveals the 10 values of n stated in the theorem.

THEOREM 5.2. *The set of n for which no $a_i(n)$, $i > 0$, is the product of two distinct primes is* {1, 2, 3, 4, 6, 7, 8, 11, 12, 15, 17, 18, 22, 23, 24, 29, 30, 35, 39, 43, 44, 69, 70, 103, 104, 119, 268, 271, 272}.

Proof. Let $S(n)$ denote the set of primes $p \in ((2n)^{1/3}, n^{1/2}]$ such that $a^{(p)}(n)/p$ is composite. By Theorem 4.1, for each $n > 371$ and each $p \in S(n)$, we have $a^{(p)}(n)$ divisible by a prime $r \leqslant (2n)^{1/3}$. Moreover, $a^{(p)}(n)$ is divisible by at most one other prime in $S(n)$. Hence

$$|S(n)| \leqslant 2\pi((2n)^{1/3}). \tag{5.4}$$

Now if p_1, p_2 are two primes in $((2n)^{1/3}, n^{1/2}]$ and not in $S(n)$, then both $a^{(p_1)}(n)$, $a^{(p_2)}(n)$ are the product of two distinct primes and are unequal. Hence, one is an $a_i(n)$ for $i > 0$. Thus we would like to show two such primes exist; that is, that

$$\pi(n^{1/2}) - 3\pi((2n)^{1/3}) \geqslant 2, \tag{5.5}$$

using (5.4). From (5.2) and (5.3) we have (5.5) for all $n \geqslant 108{,}037$. Using a table of primes, we have (5.5) for all $n \geqslant 26{,}569$. Hence for these n, there is some $a_i(n)$, $i > 0$, the product of two distinct primes. A computer check for $n < 26{,}569$ reveals the 29 cases reported in the theorem.

Remark 5.1. From the corollary to Theorem 2.1, the number of $a_i(n)$ in the form pq is asymptotic to $\pi(n^{1/2})$, and hence tends to infinity. Thus for each k, the set of n for which fewer than k of the $a_i(n)$ are in the form pq is finite. In fact, the above proof shows that for any such $n > 371$, $\pi(n^{1/2}) - 3\pi((2n)^{1/3}) < k$. The set of such n can then be computed using (5.2) and (5.3).

It has been conjectured by Erdös [3] (see also [4]) that for every k and all sufficiently large n, there is a square-free integer m with exactly k prime factors such that

$$n < m < n + p(m), \tag{5.6}$$

where $p(m)$ denotes the least prime factor of m. Since every m satisfying (5.6) must be an $a_i(n)$, this conjecture would imply that for every k and all sufficiently large n there is some $a_i(n)$, $i > 0$, that is the product of k distinct primes. Another conjecture is that for all sufficiently large n there is some $a_i(n)$, $i > 0$, composed entirely of primes below $n^{1/2}$. Much weaker than these conjectures is this: For all sufficiently large n there is some $a_i(n)$, $i > 0$, not in the form p, p^2, or pq. We tested this last conjecture numerically and found there are fairly large choices for n, where every $a_i(n)$, $i > 0$, is in the form p, p^2, or pq (e.g., $n = 362{,}610$). We also found that $n = 1{,}021{,}482$ has no $a_i(n)$, $i > 0$, divisible by three distinct primes.

In a somewhat different direction, we conjecture that the set of n for which every $a_i(n)$, $i > 0$, is square-free or the square of a prime is infinite. In fact we think that this set has positive asymptotic density. Our numerical work suggests this density may be larger than 1/10. If true, this conjecture would imply that the set of n for which every m satisfying (5.6) is square-free has positive lower density. We can, however, give a direct proof of this last statement and in fact show the density exists, but we do not present the details here.

Another conjecture supported by our calculations is that there are infinitely many n for which no $a_i(n) < g(n)$ is the square of a prime.

For each n, let $M(n)$ denote the set of m satisfying (5.6). As a corollary to Theorem 3.1 we have that for all large n there is a prime $p < n^{1/2}$ with $p \nmid n$ and p dividing no member of $M(n)$. The analogous statement is also true if we replace (5.6) with the inequality $n - p(m) < m < n$.

We now prove the following.

THEOREM 5.3. *The number of $n \leqslant x$ for which some $a_i(n) = p^3$ for some prime p is $O(x/\log x)$.*

Proof. We shall use the following result of Selberg [13]: If $\Phi(x)$ is a positive increasing function with

$$\liminf_{x\to\infty} \log \Phi(x)/\log x > 19/77,$$

then $\pi(y + \Phi(y)) - \pi(y) \sim \Phi(y)/\log y$ for all values of $y \leqslant x$ but for an exceptional set of measure $O(x/\log x)$. That is, for each $\epsilon > 0$, the set of $y \leqslant x$ for which

$$| \pi(y + \Phi(y)) - \pi(y) - \Phi(y)/\log y | \geqslant \epsilon\Phi(y)/\log y$$

is $O_\epsilon(x/\log x)$.

Let S be the set of $n \leqslant x$ for which

$$\pi(n^{2/3} + 6n^{1/3}) - \pi(n^{2/3}) < 6n^{1/3}/\log n.$$

Then for each $n \in S$, we have

$$[(n-1)^{2/3}, n^{2/3}] \subset \{y: 0 \leqslant y \leqslant x^{2/3}, \pi(y + 6y^{1/2}) - \pi(y) < 4y^{1/2}/\log y + 1\},$$

so that by Selberg's theorem, we have

$$\sum_{n \in S} (n^{2/3} - (n-1)^{2/3}) = O(x^{2/3}/\log x).$$

A simple calculation then shows the number of $n \in S$ is $O(x/\log x)$. Hence except for at most $O(x/\log x)$ choices of $n \leqslant x$ we have

$$\pi(n^{2/3} + 6n^{1/3}) - \pi(n^{2/3}) \geqslant 6n^{1/3}/\log n.$$

Another calculation shows that except for $O(x/\log x)$ choices of $n \leqslant x$, there are no primes in the interval $[n^{1/3}, n^{1/3} + 3]$.

Let $\epsilon > 0$ be small and let $p_1 < \cdots < p_s$ denote the primes in $[n^{1/3}, (1+\epsilon)\, n^{1/3})$. By Theorem 2.1, if p is another prime, then $a^{(p)}(n) \neq p^3$. Let $q_1 < \cdots < q_t$ be the primes in $(n^{2/3}, n^{2/3} + 6n^{1/3})$. By the above considerations, except for $O(x/\log x)$ choices of $n \leqslant x$, we have

$$p_1 > n^{1/3} + 3 \qquad \text{and} \qquad t \geqslant 6n^{1/3}/\log n. \tag{5.7}$$

Now

$$q_t < n^{2/3} + 6n^{1/3} < (n^{1/3} + 3)^2 < p_1^2,$$

so that each $p_i q_t < p_t p_1^2 \leqslant p_1^3$. Hence to show no $a^{(p)}(n) = p^3$, it will suffice to show each $a^{(p_i)}(n) \leqslant p_i q_t$. But if $a^{(p_i)}(n) > p_i q_t$, then each $a^{(q_j)}(n) < p_i q_j$. Thus each $a^{(q_j)}(n)$ is divisible by a prime below p_i. No $a^{(q_j)}(n)$ is divisible by a $q_{j'}$ with $j' \neq j$. Hence

$$\pi(n^{1/3}) + i \geqslant t.$$

But $\pi(n^{1/3}) \sim 3n^{1/3}/\log n$ and $i \leqslant s \sim 3\epsilon n^{1/3}/\log n$. Thus, we have contradicted (5.7) if n is sufficiently large.

Remark 5.2. If Cramér's conjecture (2.9) is true, then Theorem 5.3 can be considerably strengthened. Indeed if some $a_i(n) = p^3$, then Theorem 2.1 and (2.10) imply $n^{1/3} \leqslant p \leqslant n^{1/3} + O(n^{-1/3} \log^2 n)$. Hence for all sufficiently large n there would be at most one $a_i(n) = p^3$. Moreover the number of $n \leqslant x$ with such an $a_i(n)$ would be $O(x^{2/3} \log x)$.

THEOREM 5.4. *There is a constant $c > 0$ such that the number of $n \leqslant x$*

for which some $a_i(n) = p^k$ *for some prime* p *and some integer* $k \geqslant 4$ *is at most* x^{1-c} *for all sufficiently large* x.

Proof. If $a^{(p)}(n) = p^k$ where $k \geqslant 4$, then by Theorem 2.1, we have $p < 2n^{1/4}$ for all large n. Then by Theorem 2.2, we have

$$n \leqslant p^k = a^{(p)}(n) < n + n^{11/16+\epsilon}$$

for all $n \geqslant n_0(\epsilon)$. Then since $k \geqslant 4$, we have for $n \geqslant n_0(\epsilon)$,

$$n^{1/k} \leqslant p < n^{1/k} + n^{-1/16+\epsilon}. \tag{5.8}$$

Since $k \ll \log n$, a simple computation shows that the number of $n \leqslant x$ for which some prime p satisfies (5.8) for some integer k is at most $x^{15/16+o(1)}$.

6. The Numbers $g(n)$

In this section we shall look at the distribution of the numbers $g(n)$ as well as their multiplicative properties.

Theorem 6.1. *The number of values of* $g(n) \leqslant x$ *is* $O(x/\log x)$.

Proof. We shall use the following recent result of Warlimont [15]: Let p_m denote the mth prime and let $d_m = p_{m+1} - p_m$. Then there is an absolute constant $K > 0$ such that for all $\epsilon > 0$ we have

$$\sum_{\substack{m \leqslant x \\ d_m \geqslant p_m^{1/6+\epsilon}}} d_m \ll x^{1-K\epsilon}.$$

The actual value "$\frac{1}{6}$" does not appear in Warlimont's paper, but we obtain this number by using Huxley [8]. We shall apply this result with $\epsilon = 1/12$, so let $K/12 = c$. Then

$$\sum_{\substack{m \leqslant x \\ d_m \geqslant p_m^{1/4}}} d_m \ll x^{1-c}. \tag{6.1}$$

Let p be an arbitrary prime. From (6.1) we immediately have that the number $s_p(x)$ of $n \leqslant x$ for which there are no primes in the interval $[n/p, n/p + (n/p)^{1/4}]$ satisfies $s_p(x) \ll p(x/p)^{1-c}$ uniformly. Let $c' > 0$ satisfy $c' < c/(1 + c)$ and $c' < 5/17$. Then

$$\sum_{p < x^{c'}} s_p(x) \ll x^{1-c+c'(1+c)} \ll x/\log x.$$

So we may assume that if $n \leqslant x$, $p < x^{c'}$, there is a prime $q \in [n/p, n/p +$

$(n/p)^{1/4}]$. Theorem 2.3 then implies for each such n and each prime $p < n^{c'}$,

$$a^{(p)}(n) \leqslant n + p(n/p)^{1/4} < n + n^{8/17} < g(n),$$

the last inequality coming from (3.1). Hence except for $O(x/\log x)$ choices of $n \leqslant x$ we have $g(n)$ not divisible by any primes $p < n^{c'}$. Since $n < g(n) < 2n$ for large n, our theorem now follows.

If $g(n)$ is not in the form pq, then it is divisible by a prime $r < g(n)^{1/3} \leqslant (2n)^{1/3}$ for $n > 371$ (using Theorem 5.1). Suppose that Cramér's conjecture (2.9) holds. Then by (2.10) we have $a^{(r)}(n) - n \ll n^{1/3} \log^2 n$. Then (3.1) contradicts $a^{(r)}(n) = g(n)$ if n is large. Hence Cramér's conjecture implies that all but finitely many values of $g(n)$ are products of two distinct primes. Moreover, if the prime factorization of $g(n)$ is pq where $p < q$, then (2.10) implies that if $p < n^{1/2}/\log^4 n$, then $g(n) = a^{(p)}(n) \leqslant n + cn^{1/2}/\log^2 n$ again contradicting (3.1). Hence $p \geqslant n^{1/2}/\log^4 n$. Then by Theorem 2.1 and a simple computation we find that Cramér's conjecture implies the number of $g(n) \leqslant x$ is $O(x \log\log x/\log^2 x)$.

For the numbers below 27,500, the largest value of $g(n)$ not in the form pq is $1519 = 7^2 \cdot 31$. We conjecture that $1519 = g(1478)$ is the largest such value of $g(n)$. We cannot, however, even prove that $g(n)$ is infinitely often in the form pq. We also conjecture that the ratio of the two conjectured primes in $g(n)$ approaches 1 (or, at least, is bounded).

From the proof of Theorem 6.1 we have the following: There are positive constants ϵ and N such that for all large x there are fewer than $x^{1-\epsilon}$ choices of $n \leqslant x$ for which $g(n)$ has more than N prime factors. Indeed, we just choose $N = 1 + [1/c']$. We are not sure what the exact value of K is in Warlimont's theorem. If this value of K were large enough we could prove the above statement for $N = 2$ thus obtaining infinitely many $g(n)$ in the form pq. Also if K were large enough we could improve the estimate in Theorem 5.3 to $x^{1-\epsilon}$.

For each n, let $\gamma(n)$ be the number of integers m with $g(m) = n$. Then from Theorem 6.1, we have $\gamma(n) = 0$ on a set of density 1. Are there infinitely many n for which $\gamma(n) = 1$? Our numerical data suggest the answer is yes, but that these n have relative density 0 among all n for which $\gamma(n) > 0$. In fact, our data suggest that 1 is the second most popular nonzero value for γ, the most popular being 2.

Theorems 2.1 and 6.1 imply that $\gamma(n)$ is unbounded. Our numerical work has uncovered some values of n for which $\gamma(n)$ is very large. For example, $\gamma(2623) = 190$ and $\gamma(23,381) = 514$. We conjecture that

$$c = \limsup_{n\to\infty} \log \gamma(n)/\log n > 0;$$

perhaps $c > \frac{3}{5}$. We note that (2.12) implies that on the Riemann hypothesis, $c \leqslant \frac{7}{8}$.

7. Additional Comments

Let $P(n)$ denote the largest prime factor of an integer $n > 1$. For each n, let $C(n)$ denote the set of $P(a_i(n))$ for all $i > 0$ such that $a_i(n)$ is not a prime nor a square of a prime. Then for $n > 30$, $C(n)$ is not empty. There is a positive constant c and an n_0 such that

$$\max C(n) > cn^{2/3} \qquad \text{for all} \quad n > n_0, \tag{7.1}$$

where, of course, $\max C(n)$ denotes the largest member of $C(n)$. Indeed, if $\epsilon > 0$, it follows from the corollary to Theorem 2.1 that for all large n, there are primes $p < (3 + \epsilon)\, n^{1/3}$ with $a^{(p)}(n)/p$ prime. Hence, c can be taken as any number less than $\frac{1}{3}$. We can prove that c can be taken a little larger, but we cannot show $\max C(n)/n^{2/3} \to \infty$. It is an easy consequence of Theorem 2.1 that for all sufficiently large n

$$\max C(n) \leqslant (n + 1)/2. \tag{7.2}$$

Equality holds for all sufficiently large n of the form $2p - 1$, where p is prime. However, if n is the product of the first k primes, then as $k \to \infty$,

$$\max C(n) \leqslant (1 + o(1))\, n/\log n. \tag{7.3}$$

We do not know how to narrow the gap between (7.1) and (7.3).

In Section 5 we conjectured that for all sufficiently large n, there are some $a_i(n)$, $i > 0$, composed entirely of primes below $n^{1/2}$; that is, $\min C(n) < n^{1/2}$. We now conjecture that for every $\epsilon > 0$, there is an $n_0(\epsilon)$ such that

$$\min C(n) < n^{\epsilon} \qquad \text{for all} \quad n > n_0(\epsilon). \tag{7.4}$$

Perhaps it is possible to prove (7.4) for almost all n, but we cannot quite show this. It is easy to see that for each prime p, there are infinitely many n for which $\min C(n) = p$. However, if $p_2(n)$ denotes the second smallest member of $C(n)$, then

$$p_2(n) \to \infty \qquad \text{as} \quad n \to \infty. \tag{7.5}$$

Indeed, let K be large. If $p_2(n) \leqslant K$, then there are $0 < i < j$ with $a_j(n) \leqslant g(n)$ and $P(a_i(n)) \leqslant K$, $P(a_j(n)) \leqslant K$. From a result of Mahler, for each $\epsilon > 0$, there is an $n_0(K, \epsilon)$ such that for all $n > n_0(K, \epsilon)$, $a_j(n) - a_i(n) > n^{1-\epsilon}$. But from Theorem 2.2, $a_j(n) - a_i(n) < n^{19/24+\epsilon}$. Then for small ϵ we have a contradiction. Thus for $n > n_0(K, \epsilon)$, $p_2(n) > K$.

Now let $s(n)$ be the largest $a_i(n)$, $i > 0$, not the square of a prime but not square free. In Section 5 we conjectured that $s(n)$ does not exist for a positive density of n. Now we conjecture that the upper asymptotic density of the set of n for which $s(n) \geqslant n + t$ tends to 0 as $t \to \infty$.

In another direction, we conjecture that all sufficiently large integers m are of the form $g(n) - n$. In fact this may be true for every nonnegative integer m. We have verified that every m in the interval [0, 1000] is so representable. We cannot even show that such m contain a positive density of integers.

We now say a word about $a_i(n)$ for fixed i. Clearly $a_i(n) = n + i$ for $i = 0, 1$ for all n. It is not difficult to show $a_2(n) = n + p$ where p is the least prime which does not divide n. Moreover, if n is odd, then $a_3(n) = n + 1 + q$, where q is the least prime that does not divide $n + 1$. If n is even, we do not have a simple formula for $a_3(n)$, but we do note that $a_3(n) \leqslant n + p^2$ and that equality can hold for every p and infinitely many n. We conjecture that for every fixed i,

$$a_i(n) \leqslant n + (1 + o(1)) \log n \qquad \text{as} \quad n \to \infty. \tag{7.6}$$

By the above comments we have (7.6) for $i = 0, 1, 2$ and for $i = 3$ in the case n is odd. We note that from recent work of Iwaniec [17] we have

$$a_i(n) \leqslant n + ci^2 \log^2 n,$$

where c is an absolute constant. Finally we conjecture that if $f_k(n)$ is the least integer larger than $n + k$, with $(\prod_{k=0}^{i} (n + i), f_k(n)) = 1$, then $f_k(n) \leqslant n + (1 + o(1)) \log n$ as $n \to \infty$. We note that if we ignore a sequence of n of density ϵ_k (where $\epsilon_k \to 0$) we have $f_k(n) = F_k(n)$, where $F_k(n)$ is the least number larger than $n + k$ and relatively prime to $k!$. Furthermore, except for density ϵ_k choices of n, $F_k(n) < n + k + c \log k$.

It is clear that $a_i(n) - n \sim i \log i$ for fixed n as $i \to \infty$. It might be interesting to determine for which range of i, n this result becomes true.

8. Programming Note

This is an abstract of the computer program used for the majority of the results in the paper "On a Class of Relatively Prime Sequences," by Erdös, Penney, and Pomerance. The heart of the program is the construction of the integer $a_{i+1}(n)$ mentioned in the Abstract and on the first page of the paper itself. We remind the reader that for a given natural number n, we put $a_0(n) = n$, and if $a_0(n),..., a_i(n)$ have been defined, then $a_{i+1}(n)$ is the least integer exceeding $a_i(n)$ that is relatively prime to all the previously constructed terms $a_0(n),..., a_i(n)$.

The results of the paper require the computation of no term of the sequence beyond $g(n)$, the largest $a_i(n)$ that is neither prime nor the square of a prime. In addition, no prime members of the sequence need actually be computed.

It is easy to see that in the finite sequence actually computed, each prime $p \leqslant n^{1/2}$ will appear as a factor of at most one term of the sequence, and that terms of the sequence not divisible by such primes will be squares of primes $q > n^{1/2}$.

The program used takes advantage of these observations as follows. Suppose that n is given. Note that we always have $a_1 = n + 1$. Initially store the primes $p \leqslant n^{1/2}$ in array A. Move to array B those primes in A that are divisors of n or of $n + 1$. The other prime factors of n and $n + 1$ are stored in a third array, say, C.

The inductive step proceeds as follows. Suppose that $a_0, a_1, \ldots, a_i$ have been chosen. Form a fourth array D by this method: For each prime p remaining in A, let mp be the least multiple of p exceeding a_i. The array D consists of all such multiples, together with q^2 where q is the least prime exceeding $n^{1/2}$ that is not already stored in C.

The array D is then examined; its least element extracted. If this is q^2, we have found the value of a_{i+1}. Otherwise, this element must be tested; if it is divisible by any prime in each B or C, it must be replaced by the next larger multiple of its "corresponding" prime, and D reexamined. If not, we have found a_{i+1}.

The process terminates when A is exhausted. Running records are kept of various information needed for the paper, including statistics on $g(n)$ and $h(n)$, existence of terms in each sequence of forms other than pq and p^2, and several other related records. Certain accelerating options—such as use of only *odd* multiples of the primes p in A in construction of D—were used, since this program does not run rapidly with values of $n > 10^7$. We omit data on such accelerators from this abstract, since they are many in number and about as trivial as the example cited. All programs were run in FORTRAN on the University of Georgia's CDC Cyber 70 Model 74, under Batch mode in the interactive system NOS 1.2, release 446. The program consists of a driver/prime generator and seven subroutines, amounting to about 350 lines of FORTRAN, and a photocopy may be obtained from the author of this note (D.E.P.).

References

1. H. Cramér, On the order of magnitude of the difference between consecutive prime numbers, *Acta Arith.* **2** (1936), 396–403.
2. R. B. Eggleton, P. Erdös, and J. Selfridge, Computation of sequences maximizing least common multiples, *in* "Proceedings, Fifth Manitoba Conference on Numerical Mathematics, University of Manitoba, 1975," pp. 293–303.
3. P. Erdös, On a property of 70, *Math. Mag.* **51** (1978), 238–240.
4. P. Erdös and J. Selfridge, Some problems on the prime factors of consecutive integers, *Illinois J. Math.* **11** (1967), 428–430.

5. P. Erdös and J. Selfridge, Complete prime subsets of consecutive integers, *in* "Proceedings, Manitoba Conference on Numerical Mathematics, University of Manitoba, 1971," pp. 1–14.
6. E. Fogels, On the existence of primes in short arithmetical progressions, *Acta Arith.* **6** (1960/61), 295–311.
7. C. Hooley, On the difference of consecutive numbers prime to *n*, *Acta Arith.* **8** (1963), 343–347.
8. M. N. Huxley, On the difference between consecutive primes, *Invent. Math.* **15** (1972), 164–170.
9. A. E. Ingham, On the difference between consecutive primes, *Quart. J. Math. Oxford Ser.* **8** (1937), 255–266.
10. J. E. Littlewood, Sur la distribution des nombres premiers, *C. R. Acad. Sci. Paris, Sér. A* **158** (1914), 1869–1872. Also see G. H. Hardy and J. E. Littlewood, Contributions to the theory of the Riemann zetafunction and the theory of the distribution of primes, *Acta Math.* **41** (1918), 119–196.
11. R. A. Rankin, The difference between consecutive prime numbers, V, *Proc. Edinburgh Math. Soc.* (2) **13** (1962/63), 331–332.
12. J. B. Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers, *Illinois J. Math.* **6** (1962), 64–89.
13. A. Selberg, On the normal density of primes in small intervals, and the difference between consecutive primes, *Arch. Math. Naturvid.* **47**, No. 6 (1943), 87–105.
14. A. Walfisz, "Weylsche Exponentialsummen in der neueren Zahlentheorie," Mathematische Forschungsberichte, Vol. XV, VEB Deutscher Verlag der Wissenschaften, Berlin, 1963.
15. R. Warlimont, Über die Häufigkeit grosser Differenzen konsekutiver Primzahlen, *Monatsh. Math.* **83** (1977), 59–63.
16. M. Hausman and H. N. Shapiro, On the mean square distribution of primitive roots of unity, *Comm. Pure Appl. Math.* **26** (1973), 539–547.
17. H. Iwaniec, On the error term in the linear sieve, *Acta Arith.* **19** (1971), 1–30.
18. K. K. Norton, On character sums and power residues, *Trans. Amer. Math. Soc.* **167** (1972), 203–226; Erratum, **174** (1972), 507.