ON AN EXTREMAL PROBLEM CONCERNING PRIMITIVE SEQUENCES

P. ERDŐS, A. SÁRKÖZI *and* E. SZEMERÉDI

A sequence $a_1 < \ldots$ of integers is called primitive if no a divides any other. ($a_1 < \ldots$ will always denote a primitive sequence.) It is easy to see that if $a_1 < \ldots < a_k \leqslant n$ then $\max k = [(n+1)/2]$. The following question seems to be very much more difficult. Put

$$f(n) = \max\left(\Sigma \frac{1}{a_i}\right),$$

where the maximum is taken over all primitive sequences all of whose terms are not exceeding n. Determine, or obtain an asymptotic formula for $f(n)$. The explicit determination of $f(n)$ is probably hopeless but we will obtain an asymptotic formula for $f(n)$. In fact we will prove the following:

THEOREM.
$$f(n) = \left(1 + o(1)\right) \frac{\log n}{(2\pi \log\log n)^{\frac{1}{2}}}. \tag{1}$$

Behrend [**2**] proved that ($c_1, \ldots$ will denote positive absolute constants)

$$f(n) < c_1 \frac{\log n}{(\log\log n)^{\frac{1}{2}}}$$

and Pillai showed that

$$f(n) > c_2 \frac{\log n}{(\log\log n)^{\frac{1}{2}}}.$$

P. Erdős [**3**] stated without giving a detailed proof that (1) holds. He proves in [**3**] that

$$f(n) \geqslant \left(1 + o(1)\right) \frac{\log n}{(2\pi \log\log n)^{\frac{1}{2}}} \tag{2}$$

but the proof of the upper bound is only indicated. I. Anderson [**1**] showed that the proof suggested in [**3**] only gives

$$f(n) \leqslant \left(1 + o(1)\right) \frac{\log n}{(\pi \log\log n)^{\frac{1}{2}}}.$$

In the present paper we will prove (1), but our proof will be completely different than envisaged in [**3**]. In view of (2) it will suffice to prove that

$$f(n) \leqslant \left(1 + o(1)\right) \frac{\log n}{(2\pi \log\log n)^{\frac{1}{2}}} \tag{3}$$

Received 7 February, 1966.

[JOURNAL LONDON MATH. SOC., 42 (1967), 484–488]

and in the rest of our paper we will mainly be concerned with the proof of (3).

Denote by $\alpha(m)$ the number of prime factors of m multiple factors counted multiply. $v(m)$ denotes the number of distinct prime factors of m. Put

$$\sum_r^{(n)} = \sum \frac{1}{t}, \quad t \leqslant n, \quad \alpha(t) = r.$$

Denote $[\log\log n] = x$. In [**3**] it is proved that

$$\sum_x^{(n)} = \left(1+o(1)\right)\frac{\log n}{(2\pi x)^{\frac{1}{2}}}. \tag{4}$$

Thus (3) and hence our theorem will be proved if we show that

$$f(n) \leqslant \left(1+o(1)\right)\sum_x^{(n)}, \tag{5}$$

Instead of (5) we could prove

$$f(n) < \left(1+\frac{c_3}{x}\right)\sum_x^{(n)}. \tag{6}$$

We do not discuss the proof of (6) since perhaps very much more is true. Possibly

$$f(n) - \max_r \sum_r^{(n)}$$

is much smaller (we can show that it is not bounded). The value of r for which $\sum_r^{(n)}$ assumes its maximum is estimated very accurately in [**3**].

Now we prove (5). We need the following:

Lemma.
$$\Sigma_1 \frac{1}{t} = o\left(\frac{\log n}{x^{\frac{1}{2}}}\right)$$

where in Σ_1 $1 \leqslant t \leqslant n$ *and* $\alpha(t) - v(t) > 100 \log x$.

Let t be an integer for which $\alpha(t)-v(t) > 100\log x$. Then t is clearly divisible by a square m for which $\alpha(m)-v(m) > 25 \log x$. Hence we obtain by a simple argument (p runs through the primes)

$$\Sigma_1 \frac{1}{t} < \left(\sum_p \frac{1}{p^2}\right)^{10\log x} \sum_{l\leqslant n} \frac{1}{t} < \left(\frac{3}{4}\right)^{10\log x} 2\log n = o\left(\frac{\log n}{x^{\frac{1}{2}}}\right),$$

which proves the Lemma.

Let $a_1 < \ldots < a_k \leqslant n$ be a primitive sequence for which

$$\alpha(a_i) - v(a_i) < 100 \log x.$$

Now we show

$$\sum_{i=1}^{k} \frac{1}{a_i} \leqslant \left(1+o(1)\right)\sum_x^{(n)}. \tag{7}$$

(7), and our Lemma implies (5). Thus to prove our theorem it will suffice to prove (7).

Denote by $a_j^{(r)}$ the set of those a's which have r prime factors $\left(i.e.\ \alpha(a_j^{(r)})=r\right)$. Write

$$\sum_{i=1}^{k}\frac{1}{a_i}=\sum_{r>x}\sum_j\frac{1}{a_j^{(r)}}+\sum_j\frac{1}{a_j^{(x)}}+\sum_{r<x}\sum_j\frac{1}{a_j^{(r)}}=\Sigma_1+\Sigma_2+\Sigma_3. \tag{8}$$

Some of the sums on the right-hand side of (8) may be empty, an empty sum is 0.

Consider first the $a_j^{(r)}$ with $r>x$. Replace each such $a_j^{(r)}$ by all its divisors having exactly x prime factors. Thus we obtain the sequence $b_1<\ldots$. In other words the b's are those integers with $\alpha(b_i)=x$ which are divisors of some $a_j^{(r)}$ with $r>x$. Similarly $d_1<\ldots$ are those integers not exceeding n with $\alpha(d_i)=x$ which are multiples of some $a_j^{(r)}$ with $r<x$. Since $a_1<\ldots<a_k\leqslant n$ is a primitive sequence the three sequences $b_1<\ldots;\ a_1^{(x)}<\ldots;\ d_1<\ldots$ are disjoint, hence

$$\sum_i\frac{1}{b_i}+\sum_j\frac{1}{a_j^{(x)}}+\sum_i\frac{1}{d_i}\leqslant\sum_x^{(n)}. \tag{9}$$

In view of (8) and (9), (7) [and hence (5) and (1)] will follow if we show

$$\Sigma_1\leqslant\left(1+o(1)\right)\sum_i\frac{1}{b_i} \tag{10}$$

and

$$\Sigma_3\leqslant\left(1+o(1)\right)\sum_i\frac{1}{d_i}+o\left(\frac{\log n}{x^{\frac{1}{2}}}\right). \tag{11}$$

Thus to prove our Theorem we have to show (10) and (11). First we prove (10) the proof of (11) will be similar but slightly more involved. Put

$$\max_i\alpha(a_i)=r_1,\quad \min_i\alpha(a_i)=r_2.$$

We can assume that $r_1>x$, for it not then $\Sigma_1=0$ and (10) is trivial. We will transform the set of a's satisfying $\alpha(a_j)>x$ into the b's by an induction process. The first step is to consider all the integers $u_i^{(r_1-1)}$ [$u^{(k)}$ denotes an integer with $\alpha(u^{(k)})=k$] which divides some $a^{(r_1)}$. These integers clearly all differ from the $a_j^{(r_1-1)}$ (since the a's are primitive). Now consider all the $u_i^{(r_1-2)}$ which divide either one of the $u_i^{(r_1-1)}$ or one of the $a_j^{(r_1-1)}$. These $u_i^{(r_1-2)}$ all differ from the $a_j^{(r_1-2)}$. If we apply this process r_1-x times we clearly obtain the b's (in other words the b's are the $u_i^{(r-x)}$'s. We have for every l

$$\sum_i\frac{1}{u_i^{(r_1-l)}}\sum\frac{1}{p}\geqslant(r_1-l+1-100\log x)\left(\sum_i\frac{1}{u_i^{(r_1-l+1)}}+\sum_j\frac{1}{a_j^{(r_1-l+1)}}\right). \tag{12}$$

The proof of (12) follows easily from the definition of the $u^{(r)}$'s. The integers $u_i^{(r_1-l)}$ are defined as the set of all divisors, having r_1-l prime factors, of the integers $u_i^{(r_1-l+1)}$ and $a_j^{(r_1-l+1)}$. Hence if we multiply each integer $u_i^{(r_1-l)}$ by all the primes $p \leqslant n$ we obtain each integer $m = u_i^{(r_1-l+1)}$ or $m = a_j^{(r_1-l+1)}$ at least $v(m)$ times and by Lemma 1

$$v(m) \geqslant r_1 - l + 1 - 100 \log x.$$

This completes the proof of (12).

From (12) and the theorem of Mertens,

$$\sum_{p \leqslant n} \frac{1}{p} < x + c_4,$$

we obtain

$$\sum_i \frac{1}{u_i^{(r_1-l)}} \geqslant \frac{r_1 - l + 1 - 100 \log x}{x + c_4} \left(\sum_i \frac{1}{u_i^{(r_1-l+1)}} + \sum_j \frac{1}{a_j^{(r_1-l+1)}} \right). \tag{13}$$

Clearly

$$\frac{r_1 - l + 1 - 100 \log x}{x + c_4} > 1 \quad \text{if } r_1 - l > x + 200 \log x \tag{14}$$

and for every $r_1 - l \geqslant x$ $(x \geqslant x_0)$

$$\frac{r_1 - l + 1 - 100 \log x}{x + c_4} > 1 - \frac{200 \log x}{x}. \tag{15}$$

From (13), (14) and (15) we obtain by a simple induction argument with respect to l

$$\sum_i \frac{1}{u_i^{(x)}} = \sum_i \frac{1}{b_i} > \left(1 - \frac{200 \log x}{x}\right)^{200 \log x} \sum_{r > x} \sum_j \frac{1}{a_j^{(r)}} = \bigl(1 + o(1)\bigr) \Sigma_1 \tag{16}$$

hence (10) is proved.

We now prove (11). We can assume that $r_2 < x$. As in the proof of (10) we start with the integers $a_i^{(r_2)}$. Denote by $u_i^{(r_2+1)}$ the set of all (distinct) integers of the form $p a_i^{(r_2)}$, $p < n^{1/x^2}$. The $u_i^{(r_2+1)}$ and $a_j^{(r_2+1)}$ are distinct as in the proof of (10). By $u_i^{(r_2+2)}$ we denote the numbers of the form $p u_i^{(r_2+1)}$ and $p a_j^{(r_2+1)}$, $p < n^{1/x^2}$, etc. We repeat this process $x - r_2$ times. The u's are all less than $n . n^{(1/x^2)(x-r_2)} \leqslant n^{1+1/x}$.

The numbers $u_i^{(x)}$ consist of some (perhaps all) the d's and also some (or all) the integers in the interval $(n, n^{1+1/x})$ having x prime factors. We have

$$\left(\sum_j \frac{1}{a_i^{(r_2+l)}} + \sum_i \frac{1}{u_i^{(r_2+l)}} \right) \sum_{p < m^{1/x^2}} \frac{1}{p} \leqslant (r_2 + l + 1) \sum_i \frac{1}{u_i^{(r_2+l+1)}}, \tag{17}$$

(17) is evident since each integer having $r_2 + l + 1$ prime factors has at most $r_2 + l + 1$ divisors having $r_2 + l$ prime factors.

By the theorem of Mertens we obtain from (17)

$$\sum_i \frac{1}{u_i^{(r_2+l+1)}} \geqslant \frac{x - 3 \log x}{r_2 + l + 1} \left(\sum \frac{1}{u_i^{(r_2+l)}} + \sum \frac{1}{a_i^{(r_2+l)}} \right). \tag{18}$$

If $r_2+l+1 \leqslant x-3\log x$ then

$$\frac{x-3\log x}{r_2+l+1} \geqslant 1$$

and since $r_2+l+1 \leqslant x$ we always have

$$\frac{x-3\log x}{r_2+l+1} \geqslant \frac{x-3\log x}{x} = 1-\frac{3\log x}{x}.$$

Thus as in the proof of (10) we have by induction with respect to l

$$\sum \frac{1}{u_i^{(x)}} \geqslant \left(1-\frac{3\log x}{x}\right)^{3\log x} \Sigma_3 = \big((1+o(1)\big)\Sigma_3. \tag{19}$$

On the other hand we have

$$\sum_i \frac{1}{u_i^{(x)}} \leqslant \sum \frac{1}{d_i} + \sum_{l=n}^{n^{1+1/x}} \frac{1}{t} = \sum_i \frac{1}{d_i} + O\left(\frac{\log n}{x}\right). \tag{20}$$

(11) immediately follows from (19) and (20) and hence (3) and (1) are proved.

References

1. I. Anderson, " On primitive sequences ", *Journal London Math. Soc.*, 42 (1967) 137–148.
2. F. Behrend, " On sequences of numbers not divisible one by another ", *Journal London Math. Soc.*, 10 (1935), 42–44.
3. P. Erdős, " On the integers having exactly k prime factors ", *Annals of Math.*, 49 (1948), 53–66.

Mathematical Institute,
Hungarian Academy of Science,

and

University of Budapest.

Printed by C. F. Hodgson & Son, Ltd., Pakenham Street, London, W.C.1