ACTA ARITHMETICA
XI (1966)

On the divisibility properties of sequences of integers (I)

by

P. Erdös, A. Sárközy and E. Szemerédi (Budapest)

Let $a_1 < a_2 < \ldots$ be a sequence A of integers. Put $A(x) = \sum_{a_i \leqslant x} 1$. The sequence is said to *have positive lower density* if

$$\varliminf_{x=\infty} \left(A(x)/x\right) > 0,$$

it is said to *have positive upper logarithmic density* if

$$\varlimsup_{x=\infty} \frac{1}{\log x} \sum_{a_i \leqslant x} \frac{1}{a_i} > 0.$$

The definition of upper density and lower logarithmic density is selfexplanatory.

Besicovitch ([2]) was the first to construct a sequence of positive upper density no term of which divides any other. Behrend ([1]) and Erdös ([4]) on the other hand proved that in a sequence of positive lower density there are infinitely many couples satisfying $a_i | a_j$, Behrend in fact proved this if we only assume that the upper logarithmic density is positive.

Davenport and Erdös ([3]) proved that if A has positive upper logarithmic density there is an infinite subsequence a_{i_j}, $1 \leqslant j < \infty$ satisfying $a_{i_j} | a_{i_{j+1}}$.

Put

$$f(x) = \sum_{\substack{a_i | a_j \\ a_j < x}} 1.$$

It is reasonable to conjecture that if A has positive density then

$$\lim \frac{f(x)}{x} = \infty. \tag{1}$$

We have proved (1) and in fact obtained a fairly accurate determination of the speed with which $f(x)/x$ has to tend to infinity, this strongly depends on the numerical value of the density of A. We will prove (1) in a subsequent paper.

Throughout this paper $c_1, c_2, \ldots$ will denote positive absolute constants, not necessarily the same at each occurence, $\log_k x$ denotes the k-fold iterated logarithm. In the present paper we shall prove the following

THEOREM 1. *Assume that the sequence A has positive upper logarithmic density and put*

$$\overline{\lim}\, \frac{1}{\log x} \sum_{a_i < x} \frac{1}{a_i} = c_1. \tag{2}$$

Then there is a c_2 depending only on c_1 so that for infinitely many x

$$f(x) > x e^{c_2 (\log_2 x)^{1/2} \log_3 x}. \tag{3}$$

On the other hand there is a sequence A satisfying (2) *so that for all x*

$$f(x) < x e^{c_3 (\log_2 x)^{1/2} \log_3 x}. \tag{4}$$

First we prove (3). Our principal tool will be the following purely combinatorial

THEOREM 2. *Let $\mathscr{S}$ be a set of n elements and let $B_1, \ldots, B_z$, $z > c_4 2^n$ ($c_4 < 1$) be subsets of $\mathscr{S}$. Then if $n > n_0(c_4)$ one of the B's contains at least $e^{c_5 n^{1/2} \log n}$ of the B's, where c_5 depends only on c_4.*

Before we prove Theorem 2 we show that apart from the value of c_5 it is best possible. To see this let the B's be all subsets of $\mathscr{S}$ having t elements where $\frac{1}{2}n + c_6 n^{1/2} > t > \frac{1}{2}n - c_6 n^{1/2}$. A simple computation shows that for suitable c_6, $z > c_4 2^n$ and every B contains fewer than $e^{c_7 n^{1/2} \log n}$ other B's.

To prove Theorem 2 we first note the well known fact that for suitable c_8

$$\sum\nolimits_1 \binom{n}{j} + \sum\nolimits_2 \binom{n}{j} < \frac{c_4}{2} 2^u, \tag{5}$$

where in $\sum_1$, $j < \frac{1}{2}n - c_8 n^{1/2}$ and in $\sum_2$, $j > \frac{1}{2}n + c_8 n^{1/2}$. Because of (5) we can assume without loss of generality (replacing c_4 by $\frac{1}{2}c_4$) that $|B|$ denotes the number of elements of B

$$\tfrac{1}{2}n - c_8 n^{1/2} < |B_i| < \tfrac{1}{2}n + c_8 n^{1/2}. \tag{6}$$

Denote by $\mathscr{S}^{(j)}$ the family of these B's which have precisely j elements (j satisfies (6)) and denote by $B_1^{(j)}, \ldots, B_{g(j)}^{(j)}$ the sets of $\mathscr{S}^{(j)}$. Clearly

$$\sum\nolimits' g^{(j)} g(j) \leqslant \frac{c_4}{2} 2^n \leqslant \frac{z}{2}, \tag{7}$$

where in $\sum'$ the summation is extended over those j's for which $g(j) \leqslant \frac{c_4}{2}\binom{n}{j}$. By (7) and $\binom{n}{j} < \frac{c2^n}{\sqrt{n}}$ we can assume without loss of generality that either $g(j) = 0$ or $g(j) > \frac{1}{2}c_4$ and that

$$\sum g(j) > c_9\sqrt{n}. \tag{8}$$

We obtain this by considering only the B's which have j elements where $g(j) > \frac{1}{2}c_4$.

Put

$$s = \left[\frac{2}{c_4}\right] + 2.$$

From (8) we obtain by a simple argument that for a suitable c_{10} there is a sequence $j_1 < j_2 < \ldots < j_s$ satisfying

$$g(j_r) > \tfrac{1}{2}c_4, \quad r = 1, \ldots, s \tag{9}$$

and

$$j_{r+1} - j_r > c_{10}n^{1/2}, \quad r = 1, \ldots, s-1. \tag{10}$$

From (10) we obtain by a simple computation that

$$\binom{j_r}{j_{r-1}} > e^{c_{11}n^{1/2}\log n}, \quad r = 1, \ldots, s. \tag{11}$$

We are going to show that c_5 can be chosen as $\frac{1}{2}c_{11}$. In fact we shall show that if we consider only the set of $\mathscr{S}^{(j_r)}, r = 1, \ldots, s$ and denote these sets by $B_1', \ldots, B_{r_1}'$ then there is a B' which contains at least

$$e^{c_5 n^{1/2}\log n}, \quad c_5 = \tfrac{1}{2}c_{11} \tag{12}$$

B's. Assume that (12) is false for sufficiently large n, we will arrive at a contradiction. Denote by $I^{(j_r)}$ the subsets of $\mathscr{S}$ having j_r elements which contain at least $e^{c_5 n^{1/2}\log n}$ of the sets B. By our assumption the families $I^{(j_r)}$ and $\mathscr{S}^{(j_r)}$ are disjoint. Denote $I^{(j_r)} \cup \mathscr{S}^{(j_r)} = V^{(j_r)}$. Put

$$|I^{(j_r)}| = h(j_r), \quad |V^{(j_r)}| = \varphi(j_r).$$

By our assumption we have

$$\varphi(j_r) = h(j_r) + |\mathscr{S}^{(j_r)}| \geqslant h(j_r) + \tfrac{1}{2}c_4\binom{n}{j_r}. \tag{13}$$

We will obtain our contradiction by showing that for a suitable r

$$\varphi(j_r) > \binom{n}{j_r}. \tag{14}$$

Now we estimate $\varphi(j_r)$ from below. First of all we evidently have

$$\varphi(j_1) = |\mathscr{S}^{(j_1)}| > \tfrac{1}{2}c_4\binom{n}{j_1}. \tag{15}$$

Now we show that for every $r \leqslant s$ $\left(s = \left[\frac{2}{c_4}\right]+2\right)$

$$\varphi(j_r) > \big(r+o(1)\big)\tfrac{1}{2}c_4\binom{n}{j_r}. \tag{16}$$

To prove (16) we use induction with respect to r. By (15), (16) holds for $r = 1$. Assume that it holds for $r-1$, we will deduce it for r. To show this we will prove that if (16) holds for $r-1$ then

$$h(j_r) > \big(r-1+o(1)\big)\binom{n}{j_r}. \tag{17}$$

By (13), (17) implies (16) for r and thus we only have to prove (17). Consider now all the subsets of $\mathscr{S}$ having j_r elements which contain one of the sets of $V^{(j_{r-1})}$. We will estimate $h(j_r)$ from below by counting in two ways the number of times a subset of $\mathscr{S}$ having j_r elements can contain a set of $V^{(j_{r-1})}$. First of all there are clearly $\varphi(j_{r-1})\binom{n-j_{r-1}}{j_r-j_{r-1}}$ such relations, since to each of the $\varphi(j_{r-1})$ sets of $V^{(j_{r-1})}$ there are clearly $\binom{n-j_{r-1}}{j_r-j_{r-1}}$ subsets of $\mathscr{S}$ having j_r elements which contain it. On the other hand the $h(j_r)$ sets of $I^{(j_r)}$ each contain at most $\binom{j_r}{j_{r-1}}$ sets of $V^{(j_{r-1})}$ (since they contain at most $\binom{j_r}{j_{r-1}}$ subsets having j_{r-1} elements). The other $\binom{n}{j_r}-h(j_r)$ subsets of $\mathscr{S}$ having j_r elements contain fewer than $e^{c_5 n^{1/2}\log n}$ sets of $V^{(j_{r-1})}$. To see this observe that such a set can not contain a set of $I^{(j_{r-1})}$ since otherwise it would belong to $I^{(j_r)}$ and since it does not belong to $I^{(j_r)}$ it contains fewer than $e^{c_5 n^{1/2}\log n}$ sets of $\mathscr{S}^{(j_r)}$. Thus we evidently have

$$\varphi(j_{r-1})\binom{n-j_{r-1}}{j_r-j_{r-1}} < h(j_r)\binom{j_r}{j_{r-1}} + \binom{n}{j_r}e^{c_5 n^{1/2}\log n}. \tag{18}$$

From (18) we obtain by a simple computation using (11) and $c_5 = \frac{1}{2}c_{11}$

$$\begin{aligned} h(j_r) &> \varphi(j_{r-1})\binom{n-j_{r-1}}{j_r-j_{r-1}}\binom{j_r}{j_{r-1}}^{-1} - \binom{n}{j_r}e^{c_5 n^{1/2}\log n}\binom{j_r}{j_{r-1}}^{-1} \\ &\geqslant \varphi(j_{r-1})\binom{n}{j_{r-1}}^{-1}\binom{n}{j_r} - \binom{n}{j_r}e^{-c_5 n^{1/2}\log n}. \end{aligned} \tag{19}$$

In (19) we use

$$\binom{n-j_{r-1}}{j_r-j_{r-1}}\binom{j_r}{j_{r-1}}^{-1}=\binom{n}{j_{r-1}}^{-1}\binom{n}{j_r}.$$

From (19) and the fact that (16) holds for $r-1$ we have

$$h(j_r)>\big(r-1+o(1)\big)\binom{n}{j_r},$$

which proves (17), and hence (16) holds for all $r\leqslant s$.

But (16) implies that (14) holds for $r=s$. This contradiction proves Theorem 2.

By the same method we would prove the following

THEOREM 3. *Let $\mathcal{S}$ be a set of n elements and let $B_1,\ldots,B_z$, $z>c\dfrac{2^n}{\sqrt{n}}x$, where $x>1$, $z\leqslant 2^n$ and c is a sufficiently large constant. Then if $n>n_0$ one of the B' contains at least $e^{c'x\log n}$ of the B's.*

Theorem 3 clearly contains Theorem 2. The proof of Theorem 3 is similar but somewhat more complicated then that of Theorem 2. We supress the proof of Theorem 3.

The proof of (3) is now a simple task. In fact we shall prove the following slightly stronger

THEOREM 1′. *Let $a_1<\ldots<a_l\leqslant N$ be a sequence of integers satisfying*

$$\sum_{i=1}^{l}\frac{1}{a_i}>c_{12}\log N. \tag{20}$$

Then there is a constant c_{13} depending only on c_{12} so that if $N>N_0(c_{11},c_{12})$ then

$$\sum{}^{+}\frac{1}{a_i}>\frac{1}{2}c_{12}\log N \tag{21}$$

where in (21) the summation is extended over the a's, which have at least $\exp\big(c_{13}(\log_2 N)^{1/2}\log_3 N\big)$ divisors among the a's.

It is easy to see that Theorem 1′ implies Theorem 1. To see this observe that if (2) holds then (20) holds for infinitely many N. But if (21) holds a simple computation shows that to each N which satisfies (21) there is an $M=M(N)<N$ which tends to infinity with N and for which the number of $a_l<M$ which have at least $\exp\big(c_{13}(\log_2 N)^{1/2}\log_3 N\big)$ divisors among the a's is greater than $\frac{1}{4}c_{12}M$. Thus M satisfies (3) and hence Theorem 1′ implies (3).

Thus we only have to prove Theorem 1′. Assume that Theorem 1′ is false. Then for arbitrarily large values of n there exists a sequence

$a_1 < \ldots < a_l \leqslant N$ satisfying (20) which does not satisfy (21). Then there clearly exists a subsequence of the sequence $a_1 < \ldots$, say $b_1 < \ldots < b_r \leqslant N$ satisfying

$$\sum_{i=1}^{r} \frac{1}{b_i} > \tfrac{1}{2} c_{12} \log N \tag{22}$$

so that each b has fewer than $\exp\left(c_{13}(\log_2 N)^{1/2} \log_3 N\right)$ divisors among the b's. We now show that this conclusion leads to a contradiction.

First we observe that by using

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6} < 2$$

we obtain that there is a t so that there is a subsequence $b_{i_1} < \ldots < b_{i_s}$ of the b's each of which can be written in the form

$$b_{i_r} = t^2 q_r, \quad 1 \leqslant r \leqslant s$$

where the q_r are squarefree integers and where

$$\sum_{r=1}^{s} \frac{1}{q_r} > \frac{1}{4} c_{12} \log N. \tag{23}$$

(23) immediately follows from the fact that every integer can be written (uniquely) as the product of a square and a squarefree number.

$d(n)$ (as usual) will denote the number of divisors of n. $d^+(n)$ denotes the number of q's which divide n. By our assumption we have for all r $(r = 1, \ldots, s)$

$$d^+(q_r) < \exp\left(c_{13}(\log_2 N)^{1/2} \log_3 N\right). \tag{24}$$

From (23) we have for $N > N_0$

$$\sum_{m=1}^{N} d^+(m) = \sum_{r=1}^{s} \left[\frac{N}{q_r}\right] \geqslant N \sum_{r=1}^{s} \frac{1}{q_r} - N > \frac{1}{5} c_{12} N \log N. \tag{25}$$

Denote by $\nu(m)$ the number of distinct prime factors of m. Since the q's are squarefree we have $d^+(n) \leqslant 2^{\nu(n)}$.

Thus from (25) we obtain (the dash indicates that the summation is extended over the $n \leqslant N$ for which $\nu(n) > \log_2 N$)

$$\sum_{m=1}^{N}{}' d^+(m) > \frac{1}{5} c_{12} N \log N - N 2^{\log_2 N} > \frac{1}{10} c_{12} N \log N. \tag{26}$$

On the other hand we evidently have

$$\sum_{m=1}^{N} d(m) = \sum_{m=1}^{N} \left[\frac{N}{m}\right] < 2N\log N.$$

Thus by (26) there is an m satisfying $\nu(m) > \log_2 N$ for which

$$d^+(m) > \frac{c_{12}}{20} d(m) \geqslant \frac{1}{20} c_{12} 2^{\nu(m)}. \tag{27}$$

The last equality of (27) follows from the fact that since the q's are squarefree we can assume that m is squarefree.

Now we can apply Theorem 2. The set $\mathscr{S}$ is the set of prime divisors of m, $\nu(m) = n$. The B's are the q's which divide m, $c_{12}/20 = c_4$. We thus obtain by Theorem 2 that there is a q/m for which

$$d^+(q) > \exp\left(c_5(\log_2 N)^{1/2}\log_3 N\right)$$

which contradicts (24) if c_{13} is sufficiently small.

This completes the proof of Theorem 1′ and hence (3) is proved. It is clear from the above proof that (21) would remain true with $1-\varepsilon$ instead of $\frac{1}{2}$.

To complete the proof of Theorem 1 we now have to show (4). (We do not give the proof in full detail.) In fact we shall prove the following stronger

THEOREM 4. *There is an infinite sequence A of positive density for which for all x*

$$f(x) < x\exp\left(c_{14}(\log_2 x)^{1/2}\log_3 x\right). \tag{28}$$

Our principal tool for the proof of Theorem 4 will be the following result from probabilistic number theory:

THEOREM 5. *Let n be squarefree. Let $n = \prod_k p_k^{(n)}$, $p_1^{(n)} < \ldots < p_{\nu(n)}^{(n)}$, be the decomposition of n into primes. Then for every $c_{15} > 0$ there is a $k_0 = k_0(c_{14})$ so that the density of integers n which satisfy for all $k_0 < k \leqslant \nu(n)$*

$$e^{e^{k-c_{15}(\log_2 n)^{1/2}}} < p_k < e^{e^{k+c_{15}(\log_2 n)^{1/2}}} \tag{29}$$

is positive.

Theorem 5 can be proved by the methods of probabilistic number theory ([5], [6]). We do not give here the proof of Theorem 5.

Now we show that the sequence of integers which satisfy (29) for all $k > k_0(c_{15})$ also satisfy (28) and if this is accomplished Theorem 4 and therefore (4) is proved. Thus the proof of Theorem 1 will be complete.

Let $a_1 < \ldots < a_l \leqslant x$ be the sequence of integers satisfying (29). From (29) we obtain by a simple computation that for every r, $1 \leqslant r \leqslant l$

(30) $$\log_2 a_r - 2c_{14}(\log_2 a_r)^{1/2} < \nu(a_r) < \log_2 a_r + 2c_{14}(\log_2 a_r)^{1/2}.$$

Denote as before by $d^+(a_r)$ the number of a's dividing a_r. To prove (28) it will suffice to show that for every r

(31) $$d^+(a_r) < \exp\left(c_{14}(\log_2 x)^{1/2}\log_3 x\right).$$

Denote by $p_1 < \ldots < p_{\nu(a_r)}$ the prime factors of a_r. Assume $a_t | a_r$. If $\nu(a_t) \leqslant k_0$ then by (30) there are clearly fewer than $\nu(a_r)^{k_0+1} \leqslant (\log_2 x)^{k_0+2}$ choices for a_t, thus these can be ignored. If $\nu(a_t) > k_0$, let p_s be the greatest prime factor of a_t. Since a_t and a_r both satisfy (29) and (30) a simple computation shows that

(32) $$s - 3c_{14}(\log_2 a_r)^{1/2} \leqslant \nu(a_t) \leqslant s.$$

Thus by an easy argument and simple computation

$$\begin{aligned} d^+(a_r) &\leqslant (\log_2 x)^{k_0+2} + \sum_{s=k_0+1}^{\nu(a_r)} \sum_{s-3c_{15}(\log_2 a_r)^{1/2}}^{s} \binom{s}{u} \\ &< (\log_2 x)^{k_0+2} + \nu(a_r)\left(\nu(a_r)\right)^{4c_{15}(\log_2 a_r)^{1/2}} \\ &< \nu(a_r)^{5c_{15}(\log_2 a_r)1/2} < \exp\left(c_{16}(\log_2 x)^{1/2}\log_3 x\right). \end{aligned}$$

Thus (31) is proved (with $c_{16} = c_{14}$).

References

[1] F. Behrend, *On sequences of numbers not divisible one by another*, J. London Math. Soc. 10 (1935), pp. 42-44.

[2] A. S. Besicovitch, *On the density of certain sequences*, Math. Ann. 110 (1934), pp. 336-341.

[3] H. Davenport and P. Erdös, *On sequences of positive integers*, Acta Arith. 2 (1936), pp. 147-151.

[4] P. Erdös, *Note on sequences of integers no one of which is divisible by any other*, J. London Math. Soc. 10 (1935), pp. 126-128.

[5] — *On the distribution function of additive functions*, Ann. Math. 47 (1946), pp. 1-20.

[6] J. Kubilius, *Probabilistic methods in the theory of numbers*, Translation of Math. Monographs, Amer. Math. Soc. 1964, vol. 11.

Reçu par la Rédaction le 2. 7. 1965