On the arithmetical density of the sum of two sequences one of which forms a basis for the integers.

By

P. Erdös (Manchester).

Let $a_1, a_2, \ldots$ be any given sequence of positive steadily increasing integers and suppose there are $x = f(n)$ of them not exceeding a number n, so that

$$a_x \leqq n < a_{x+1}.$$

The density δ of the sequence is defined by Schnirelmann as the lower bound of the numbers $f(n)/n$, $n = 1, 2, \ldots$. Thus if $a_1 \neq 1$, $\delta = 0$.

Clearly $f(n) \geqq \delta n$.

Suppose also that the steadily increasing set

$$A_0 = 0, A_1, A_2, \ldots$$

forms a basis of order l of the positive integers. This means that every positive integer can be expressed as the sum of at most l of the A's. I prove the following

Theorem: *If δ' is the density of the sequence $a + A$, i. e. of the integers which can be expressed as the sum of an a and an A, then*

$$\delta' \geqq \delta + \frac{\delta(1-\delta)}{2l}.$$

Particular cases of this theorem have been proved by Khintchine and Buchstab in an entirely different and more complicated way.

I prove my theorem as a particular case of a more general one. Let the positive integers $\leqq n$ not included among the a's be denoted by $b_1, b_2, \ldots$, and let

$$b_y \leqq n < b_{y+1}.$$

Put

$$E = b_1 + b_2 + \ldots + b_y - \frac{1}{2} y(y+1).$$

Clearly $E \geqq 0$, since $b_1 \geqq 1$, $b_2 \geqq 2$ etc. Then there exist at least $x + \frac{E}{ln}$ integers $\leqq n$ of the form $a + A$, where in fact we need only use $A = 0$ and a single other A. This theorem is deduced from the one that there are at least $\frac{E}{ln}$ of the b's $\leqq n$ which can be represented in the form $a + A$, and in fact only a single A is used.

We require the

Lemma: *An integer $J > 0$ exists such that there are at least $\frac{E}{n}$ of the b's among the integers $\leqq n$ in the set $a_1 + J$, $a_2 + J, \ldots$.*

For the number of solutions of the equation

$$a + v = b$$

in positive integers v and a's, b's $\leqq n$ is E. Thus for given $b = b_r$, there are $b_r - r$ solutions since the number of a's $< b_r$ is clearly $b_r - r$ and every such a gives a solution v. Hence summing for $r = 1, 2, \ldots, y$, the total number of solutions is

$$E = \sum_{r=1}^{y} (b_r - r).$$

But there are at most n possible values of v, namely $1, 2, \ldots, n$ and so for at least one value of v, say J, there are not less than $\frac{E}{n}$ solutions of $a + J = b$ in a's and b's not greater than n.

Now express J as a sum of exactly l A's, say

$$J = A_1 + A_2 + \ldots + A_l,$$

by including a sufficient number of A_0's among the A's if need be and where A_1 need not denote the first, A_2 the second etc.

Denote by μ_s fhe number of b's in the set $a + A_s$, $s = 1, 2, \dots, l$. I prove now that

$$\mu_1 + \mu_2 + \dots + \mu_l \geqq \frac{E}{n}.$$

For in the set of integers given by

$$a + A_1 + A_2,$$

there are at most $\mu_1 + \mu_2$ of the b's. Thus the set $a + A_1$ contains μ_1 of the b's together with some a's. When we add A_2 to the numbers of the set $a + A_1$, the μ_1 b's give at most μ_1 b's, while the a's give at most μ_2 b's. Now take the set $a + A_1 + A_2 + A_3$. This contains at most $\mu_1 + \mu_2 + \mu_3$ of the b's by precisely the same argument applied to the sum of $a + A_1 + A_2$ and A_3. Similarly the set $a + A_1 + A_2 + \dots + A_l$, *i. e.* $a + J$ will contain at most $\mu_1 + \mu_2 + \dots + \mu_l$ of the b's. But since the set $a + J$ contained at least $\frac{E}{n}$ of the b's, clearly one of the μ's say $\mu_k \geqq \frac{E}{ln}$, and so the set $a + A_k$ contains at least $\frac{E}{ln}$ of the b's $\leqq n$. Now the set $a + A_0$, since $A_0 = 0$, consists of exactly the x a's. Hence the set $a + A$ including $A_k = 0$ contains at least $x + \frac{E}{ln}$ different integers $\leqq n$.

Suppose now the a's have a density δ with $\delta < 1$ which is no loss of generality. We have $f(b_r) \geqq \delta b_r$ hence $b_r - r = f(b_r) \geqq \delta b_r$, $b_r \geqq \frac{r}{1 - \delta}$. and therefore

$$E \geqq \frac{1 + 2 + \dots + y}{1 - \delta} - \frac{y(y+1)}{2} \geqq \frac{\delta}{2(1 - \delta)} y(y+1).$$

Hence for the number N of integers $\leqq n$ in the set $a + A$

$$N \geqq x + \frac{\delta}{2(1 - \delta)} \frac{y^2}{ln} = \varphi(x) \qquad (y = n - x),$$

say. For $x \geqq \delta n$

$$\varphi'(x) = 1 - \frac{\delta}{2(1 - \delta)} \frac{2(n - x)}{ln} \geqq 1 - \frac{\delta}{l} > 0,$$

i. e.

$$N \geqq \varphi(x) \geqq \varphi(\delta n) = \delta n + \frac{\delta}{2(1 - \delta)} \frac{(1 - \delta)^2 n^2}{ln}.$$

hence
$$N \geqq n\left(\delta + \frac{\delta(1-\delta)}{2l}\right),$$
and this is the theorem.

I can prove in the same way that if a sequence $a_1, a_2, \ldots$ is given and there are $f(n)$ of the a's not exceeding n, then in the set $|a \pm A|$, there are at least
$$f(n) + \frac{f(n)(n - f(n))}{2l}$$
numbers not exceeding n.

Before closing my paper I would express my sincere gratitude to Prof. L. J. Mordell for having so kindly helped me with my ms.

(Received 11 March, 1935.)