The density of the set of sums

by

IMRE Z. RÜZSA* (Budapest)

Let \( 1 \leq a_1 \leq a_2 \leq \ldots \) be a sequence of integers, and let \( S \) be the set of all sums of the form \( \sum e_i a_i \), where \( e_i = 0 \) or \( 1 \).

**Theorem.** If

\[
a_{n+1} \leq 2a_n
\]

for all but at most finitely many values of \( n \), then \( S \) has an asymptotic density.

This problem was proposed by U. Zannier at the 1989 September number theory conference in Amalfi. In [1], he proves the same conclusion under the stronger assumption that \( a_{n+1} \sim a_n \). I heard it from P. Erdős at the DIMACS conference in October 1989. He also asked how (1) can be weakened, in particular, whether

\[
a_n \leq a_1 + a_2 + \ldots + a_{n-1} + c
\]

is sufficient. My proof makes a heavy use of (1). It is easy to see that if we do not impose any restriction on the sequence \( (a_i) \), then \( S \) need not have a density.

Takisaki [2] took long intervals and large gaps in \( (a_i) \) one can easily achieve \( d(S) = 1 \) and \( d(S) = c \) for an arbitrary prescribed number \( 0 \leq c \leq 1 \), and I believe even \( d(S) = c, d(S) = C \) is possible with an arbitrary pair of numbers \( 0 \leq c \leq C \leq 1 \).

(1) or (2) implies that \( d(S) > 0 \), even that \( S \) has bounded gaps.

\( S(x) \) will denote the number of integers \( s \in S, 1 \leq s \leq x \).

**Lemma 1.** If \( x = a_{i_1} + a_{i_2} + \ldots + a_{i_k} \), where \( i_1 > i_2 > \ldots > i_k \) and \( y < a_{i_k} \), then

\[
S(x) + S(y) \geq S(x+y).
\]

Indeed, all the numbers \( x+s \), where \( s \in S, 1 \leq s \leq y \), are elements of \( S \) between \( x \) and \( x+y \).

Write

\[
u = d(S), \quad v = \overline{d(S)}.
\]
Lemma 2.

\[ \lim S(a_n)/a_n = u. \]

Proof. Write

\[ w = \limsup S(a_n)/a_n. \]

We are going to deduce a contradiction from the assumption \( w > u. \) If \( S(a_n) \) is large, we shall show that every \( S(x) \) is large. We shall use the definition of \( w \) in the following form: for every \( \varepsilon > 0 \) there is a \( K = K(\varepsilon) \) such that

\[ S(x) > (u - \varepsilon)x - K \quad \text{for all } x. \]

If \( S(a_n) \) is large, first we find an \( m \) such that \( S(t) \) is large for all \( a_m/2 < t < a_m. \) Let \( m \) be the smallest subscript for which \( a_m > 2a_n. \) By definition, we have \( a_{m-1} < 2a_m, \) consequently \( a_m < 4a_n. \)

By Lemma 1 and (4) we have

\[ S(a_{m-1}) > S(a_m) + (u - \varepsilon)(a_{m-1} - a_m) - K. \]

In particular, we find

\[ S(a_{m-1}) > S(a_m) + (u - \varepsilon)(a_{m-1} - a_m) - K. \]

For \( y < a_{m-1} \) the same argument and an application of (6) yields

\[ S(a_{m-1} + y) > S(a_{m-1}) + S(y) > S(a_{m}) + (u - \varepsilon)(a_{m-1} - a_m) + K. \]

For every \( a_m/2 < t < 2a_m, \) which includes our desired interval \([a_m/2, a_m],\) we can apply either (5) or (7), and we have anyway

\[ S(t) = S(a_m) + (u - \varepsilon)(t - a_m) - 2K. \]

If we take an \( a_n \) such that \( S(a_n) > (w - \varepsilon)a_n \) while also \( a_n > 2K/\varepsilon, \) then this implies

\[ S(t) > (u - \varepsilon)t - (w - \varepsilon)a_n, \]

and taking into account that \( t < 4a_n, \) we conclude

\[ S(t)/t > u - \varepsilon + (w - \varepsilon)/4 = z > u \]

if \( \varepsilon < (w - \varepsilon)/5. \) For this \( z \) we have infinitely many \( m \) such that (8) holds for \( a_m/2 < t < a_m. \)

Now take such an \( m \) and let \( 0 < d_1 < d_1 < \ldots \) be the sequence of integers representable in the form

\[ \sum_{i=1}^{m} e_i d_i, \quad e_i = 0 \text{ or } 1. \]

By (1) we have \( d_{i+1} - d_i \leq a_m. \) If there are two consecutive differences whose sum is at most \( a_m, \) then we can omit a \( d_i \) while keeping the gaps below \( a_m \); by repeating this process we get a subsequence \((D_i)\) which satisfies

\[ D_{i+1} - D_i \leq a_m, \quad D_{i+2} - D_i > a_m \]

for all \( i. \)

A repeated application of (3) yields

\[ S(D_k) > \sum_{i=0}^{k-1} S(D_{i+1} - D_{i}) - 1 \]

(by \( -1 \) is necessary to make it work also in the extremal case \( D_{i+1} - D_i = a_m \)). To the summands we apply the inequality

\[ S(D_{i+1} - D_i - 1) > z(D_{i+1} - D_i - 1) \]

if \( D_{i+1} - D_i > a_m/2, \) while for \( D_{i+1} - D_i \leq a_m/2 \) we cannot say anything better than

\[ S(D_{i+1} - D_i - 1) > (u - \varepsilon)(D_{i+1} - D_i - 1) - K \]

by (4). Since at least every second gap is large, the total length of the large gaps is at least half the total length of all gaps, minus possibly the last one. This means that from (10), (11) and (12) we can conclude

\[ S(D_k) > \frac{z + u - \varepsilon}{2} - D_k - a_m. \]

To estimate \( k, \) we apply the second inequality of (9) to find

\[ k \leq 2D_k/a_m + 1. \]

For a general number \( y, D_k < y < D_{k+1}, \) we get

\[ S(y) > S(D_k) > \frac{z + u - \varepsilon}{2} - \frac{2y}{a_m} + 1 \]

(13) \( S(y) > \frac{z + u - \varepsilon}{2} - \frac{2y}{a_m} + 1. \)

If we take for instance \( \varepsilon = (z - u)/6, \) which determines a \( K, \) then we choose an \( a_n \) such that \( (K + 1)/a_m < \varepsilon; \) then (13) implies that

\[ S(y) > (u + \varepsilon)y - \text{ constant}, \]

a contradiction to the definition of \( u. \)

Proof of the Theorem. We estimate a general \( S(y), a_n < y < a_{n+1}, \) from above.

\[ S(y) - S(y - a_n) \]

is the number of integers \( m \in S, m \leq y, \) for which \( m - a_n \notin S, \) thus it is a monotonic function of \( y. \) Comparing its values at \( y \) and \( a_{n+1}, \) we find

\[ S(y) - S(y - a_n) \leq S(a_{n+1}) - S(a_n - a_n), \]

that is,

\[ S(y) \leq S(y - a_n) + S(a_{n+1}) - S(a_{n+1} - a_n). \]
Now take an arbitrary positive $\varepsilon$. By definition of $u$, $v$ and Lemma 2 we can find a $K$ such that we have

$$S(x) \leq (u+\varepsilon)x + K,$$

$$S(a_n) \leq (u+\varepsilon)a_n + K,$$

$$S(x) \geq (u-\varepsilon)x - K$$

for all $x$ and $k$. Applying these inequalities in this order to the terms of (14) we obtain

$$S(y) \leq vy - (v-u)a_n + \varepsilon(y + 2a_{n+1} - 2a_n) + 3K.$$ 

Taking into account that $a_n \geq y/2$ and $a_{n+1} \leq 2y$, this gives

$$S(y) \leq \frac{u+v}{2}y + 5\varepsilon y + 3K.$$

Dividing by $y$ and taking the lim sup we get

$$v \leq (u+v)/2 + 5\varepsilon.$$

Since this holds for every $\varepsilon > 0$, we have $v \leq u$. ■

Remark (1) was used several times in the course of the proof. The crucial one seems to be that in the proof of Lemma 2 to infer $d_{n+1} - d_n \leq a_n$: in the rest weaker assumptions would also work.

Acknowledgement. I have profited much from discussing the problem with Professor P. Erdős.

Reference