COMPLEMENTS OF FINITE SETS OF INTEGERS

D. J. Newman

Let A and B be sets of nonnegative integers, with $0 \in A$. The set B is called a *complement of* A if each nonnegative integer is expressible in the form $a + b$ ($a \in A$, $b \in B$). One of the basic problems in additive number theory is the determination, for a prescribed A, of a complement B that is in some sense minimal. Erdös [1] and Lorentz [2] have discussed some problems and concepts for the case where A is an infinite set; we shall deal with finite sets A.

For each set B, we denote by B(n) the number of elements of B that do not exceed n; the *upper* and *lower densities* of B are commonly defined as

$$\bar{d}(B) = \limsup_{n \to \infty} B(n)/n, \qquad \underline{d}(B) = \liminf_{n \to \infty} B(n)/n.$$

If the upper and lower densities coincide, we omit the bar and speak of the *density*.

The *codensity* of a set A is defined as the number

$$c(A) = \inf_{B} d(B),$$

where B ranges over all complements of A whose density exists.

An important result of Lorentz [2] is that $c(A) = 0$ if A is an infinite set. If A consists of k elements and B is complementary to A, then the union of the k translated sets $B + a_i$ ($a_i \in A$) contains all positive integers, and therefore $d(B) \geq 1/k$, that is, $c(A) \geq 1/k$. In this paper, we study the more difficult problem of finding the least upper bound c_k for the codensity c(A) as A ranges over all sets of k elements.

It is obvious that $c_1 = 1$ and $c_2 = 1/2$. The following two theorems constitute our main results.

THEOREM 1. $c_3 = 2/5$.

THEOREM 2. $c_k \sim \dfrac{\log k}{k}$.

Before proving these theorems, we make some simple but fundamental observations regarding codensity.

THEOREM 3. (i) $c(A) = \inf_B \underline{d}(B) = \inf_B \bar{d}(B)$, *where* B *ranges over all complements of* A.

(ii) $c(A) = \inf_S d(S)$, *where* S *ranges over all complements of* A *that are finite unions of residue classes.*

(iii) *There exists a complement* B_0 *of* A *such that* $d(B_0) = c(A)$.

Proof. Corresponding to any set B, we define the sets

$$B_n = B \cap [0, n] \oplus \{0, n, 2n, 3n, \cdots\} \qquad (n = 1, 2, \cdots).$$

Received October 26, 1965.

Clearly, if B is a complement of A, then each of the B_n is a complement of A. Also, $d(B_n)$ is $1/n$ times the number of elements in the set $B \cap [0, n]$, and we can make this as near $\underline{d}(B)$ as we like, by an appropriate choice of n. The first part of the assertion (i) is thus proved; (ii) and the second part of (i) follow because B_n is a finite union of residue classes.

For infinite sets A, (iii) was proved by Lorentz [2]. If A is finite, let a denote its greatest element. We suppose that $B^{(1)}, B^{(2)}, \cdots$ is a sequence of complements of A such that $d(B^{(k)}) \downarrow c(A)$, and for any increasing sequence $n_1, n_2, \cdots$ of integers we form the union B_0 of the sets

$$B^{(1)} \cap [0, n_1], \quad B^{(2)} \cap [n_1 - a, n_2], \quad B^{(3)} \cap [n_2 - a, n_3], \quad \cdots .$$

The set B_0 is a complement of A, and if $n_k \to \infty$ rapidly enough, then B_0 has density at most $d(B^{(k)})$, for each k. It follows that $d(B_0) = c(A)$, and Theorem 3 is proved.

Proof of Theorem 1. First we prove that $c_3 \geq 2/5$, by showing that $c(\{0, 1, 3\}) \geq 2/5$. Let S denote any finite union of residue classes such that the set

$$S \cup (S \oplus 1) \cup (S \oplus 3)$$

consists of all nonnegative integers. For convenience, we write

$$\alpha_{jk} = d[(S \oplus j) \cap (S \oplus k)]$$

(for example), and we note that

$$\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3, \quad \alpha_{02} = \alpha_{13}, \quad \alpha_{012} = \alpha_{123}.$$

We now apply the inclusion-exclusion principle (the logical rule that allows us to determine the number of elements in a *union* of sets from the number of elements in each of the possible *intersections*). Since the rule (together with our assumption on S) implies that

$$\alpha_0 + \alpha_1 + \alpha_3 - \alpha_{01} - \alpha_{13} - \alpha_{03} + \alpha_{013} = 1,$$

we deduce that

$$3\alpha_0 - \alpha_{01} - \alpha_{02} - \alpha_{03} + \alpha_{013} = 1. \tag{1}$$

Similarly, since $S \cup (S \oplus 1) \cup (S \oplus 2) \cup (S \oplus 3)$ is also the set of nonnegative integers, the rule implies that

$$4\alpha_0 - 3\alpha_{01} - 2\alpha_{02} - \alpha_{03} + 2\alpha_{012} + \alpha_{023} + \alpha_{013} - \alpha_{0123} = 1. \tag{2}$$

If we multiply both sides of (1) by 3 and subtract the corresponding sides of (2), we obtain the equation

$$5\alpha_0 - 2(\alpha_{03} - \alpha_{013}) - (\alpha_{023} - \alpha_{0123}) - \alpha_{02} - 2\alpha_{012} = 2.$$

Since the expressions in parentheses can not be negative, it follows that $5\alpha_0 \geq 2$, and our first assertion is proved.

It remains to show that $c_3 \leq 2/5$. Suppose that $0 < a < b$ and that S is a complement of the set $\{0, a, b\}$. If m is a positive integer and

$$T = \{ms + j \mid s \in S,\ j \in [0, m)\},$$

then T is a complement of the set $\{0, ma, mb\}$, and moreover $d(T) = d(S)$. Therefore it will be sufficient to prove the inequality for the sets $\{0, a, b\}$ with $(a, b) = 1$.

Corresponding to such a set $\{0, a, b\}$ we now consider the set S consisting of all the equivalence classes

$$3a, 6a, \cdots, 3ma \pmod{a+b}, \qquad \text{where } m = \left[\frac{a+b+2}{3}\right].$$

Inspection shows that the set $S \oplus \{0, a, b\}$ is the union of the residue classes $2a, 3a, \cdots, (3m+1)a \pmod{a+b}$. We note that $3m \geq a + b$, and that $a + b$ of the residue classes are distinct [otherwise, $a + b$ would divide some number ka $(0 < k < a + b)$]. It follows that S (augmented by a finite set, if necessary) is a complement of $\{0, a, b\}$. Also,

$$d(S) = \frac{m}{a+b} = \frac{\left[\frac{a+b+2}{3}\right]}{a+b}.$$

The last expression does not exceed 2/5, except if $a + b$ has one of the values 1, 2, 4, 7. The first two of these values are irrelevant. The other two arise only in the cases of the sets

$$\{0, 1, 3\}, \quad \{0, 1, 6\}, \quad \{0, 2, 5\}, \quad \{0, 3, 4\}.$$

These sets have the complements

$$\{0, 1\} \pmod 5, \quad \{0, 1, 4, 8\} \pmod{11}, \quad \{0, 1, 2\} \pmod 8, \quad \{0, 2\} \pmod 5,$$

respectively. Since none of these complements has density greater than 2/5, Theorem 1 is proved.

R. Graham has developed an algorithm for determining c(A) and has thereby obtained an independent proof of Theorem 1.

We now turn to the proof of Theorem 2.

An upper bound on c_k. Corresponding to any fixed set $A = \{a_1, a_2, \cdots, a_k\}$ $(0 = a_1 < a_2 < \cdots < a_k)$, we shall construct a complement whose density is at most $(1 + \log k)/k$. Let K and N be integers $(0 < K < N)$, and suppose that $N > a_k$, so that the a_i determine k distinct residue classes modulo N. For each number n, we denote by U_n the set of residue classes represented by

$$n - a_1, n - a_2, \cdots, n - a_k$$

(we use the symbol U to represent an unspecified class U_n). The symbol T will denote an unspecified set of K residue classes modulo N. Clearly, there are $\binom{N}{K}$ sets T, and for each n, exactly $\binom{N-k}{K}$ of these sets do not meet the set U_n. Since there are at most N different sets U_n, it follows that there are at most $N\binom{N-k}{K}$ disjoint pairs T, U. Consequently, at least one of the sets misses at

most $N\binom{N-k}{K}\Big/\binom{N}{K}$ of the sets U_n. Let S consist of such a set T, together with all residue classes n (mod N) for which $T \cap U_n = \emptyset$.

To see that the set S is a complement of A, let m be any integer. If $T \cap U_m = \emptyset$, we have the representation

$$m = 0 + m \qquad (0 \in A,\ m \in S).$$

If $T \cap U_m$ contains some element $m - a_i$, we have the representation

$$m = a_i + (m - a_i) \qquad (a_i \in A,\ m - a_i \in S).$$

To estimate the density of S, we observe that

$$d(S) \leq d(T) + N\binom{N-k}{K}\Big/\binom{N}{K}\cdot\frac{1}{N} = \frac{K}{N} + \binom{N-k}{K}\Big/\binom{N}{K} \leq \frac{K}{N} + \left(1 - \frac{k}{N}\right)^K \leq \frac{K}{N} + e^{-kK/N}.$$

By choosing K and N so that the ratio K/N is near $(\log k)/k$, we can bring the upper bound on d(S) arbitrarily near to $(1 + \log k)/k$. It follows that the codensity of A is at most $(1 + \log k)/k$.

Lower estimates on c_k. It remains to show that for each large integer k, some set A of k elements has codensity approximately $(\log k)/k$. Using probabilistic methods similar to those of Erdös, we shall show that "most" sets have the desired property.

LEMMA 1. *Let* K *be the union of* k *sets* $K_1, K_2, \cdots, K_k$, *each set* K_α *consisting of at most* j *elements, and each element lying in at most* i *of the sets. Let* K' *be formed by a random process in which each element of* K *has an independent probability* p *of being selected. Then the probability that* K' *contains at least one element of each of the sets* K_α *is at most* $[1 - (1 - p)^j]^{k/i}$.

Remark. If i = 1, the sets K are disjoint, and the upper bound in the lemma reduces to a familiar expression in probability theory. Similarly, if $i \mid k$, and if each of the k sets has exactly j elements and the k sets are comprised of k/i disjoint sets, each occurring i times, then our expression gives the exact value of the probability.

Proof. We use induction on k. Since the result is trivial for k = 1, we turn to the general case; that is, we suppose the lemma has been proved for sets of fewer than k elements. We select one of the k sets, say K_α, and we consider a success as coming about in the following way: first we pick an element in the selected set, and after that we pick an element in every remaining set *not containing* that element.

The first event occurs with a probability at most $1 - (1 - p)^j$. After the event (and conditional on it), there remain at least k - i sets from each of which we must pick an element. This latter event has probability at most

$$[1 - (1 - p)^j]^{(k-i)/j},$$

by the induction hypothesis, and therefore our upper bound for the probability of success is

$$[1 - (1 - p)^j][1 - (1 - p)^j]^{(k-i)/i} = [1 - (1 - p)^j]^{k/i}.$$

This completes the induction, and Lemma 1 is established.

Suppose now that p is a fixed number $(0 < p < 1)$, and let A be the result of choosing the number 0 and then choosing each of the integers in $\left[1, \left(\frac{1}{p} + 1\right)k\right]$ independently with probability p. We assert that, with high probability for large values of k,

(3) A has at least k elements,

(4) $$c(A) \geq (1 - 4p)\frac{\log k}{k}.$$

Because the existence of even one set A satisfying conditions (3) and (4) (with small p) is sufficient to complete the proof of our theorem, the proof of the assertion will achieve our purpose.

Since it is clear that (3) holds with high probability, we need only deal with (4).

LEMMA 2. *Suppose* $\varepsilon > 0$, $p < 1/2$, *and* $j \leq \left(\frac{1}{p} - 2\right)\log k$. *If* k *is large enough, then the probability that*

(5) $[1, k]$ *is covered by some set of* j *translates of* A

is less than ε.

Proof. Let $\{b_1, b_2, \cdots, b_j\}$ be any fixed set of integers. We first estimate the probability that the union of the translates of A by b_1, b_2, $\cdots$, and b_j covers $[1, k]$. For this to happen, A must contain at least one element from the set

$$\{1 - b_1, 1 - b_2, \cdots, 1 - b_j\},$$

one from the set $\{2 - b_1, \cdots, 2 - b_j\}$, $\cdots$, and one from the set $\{k - b_1, \cdots, k - b_j\}$. Now each of these k sets contains j elements, and no number m can lie in more than j of the sets (since this would give it two representations $m = r - b_i$ and $m = s - b_i$ with $r \neq s$). Lemma 1 is therefore applicable, and it tells us that $P \leq [1 - (1 - p)^j]^{k/j}$.

Clearly, we may assume that all the b_i lie in the half-open interval $\left(-\left(\frac{1}{p} + 1\right)k, k\right]$, so that the number of admissible j-tuples is bounded by $\left[\left(\frac{1}{p} + 2\right)k\right]^j$.

These two estimates give the upper bound

(6) $$Q = \left[\left(\frac{1}{p} + 2\right)k\right]^j [1 - (1 - p)^j]^{k/j}$$

for the probability of (5). To show that it is less than ε for sufficiently large values of k, we note that

$$\log Q = j \log\left[\left(\frac{1}{p} + 2\right)k\right] + \frac{k}{j}\log[1 - (1 - p)^j] \leq j \log\left[\left(\frac{1}{p} + 2\right)k\right] - \frac{k}{j}(1 - p)^j.$$

The last member is a decreasing function of j. Replacing j with $\left(\frac{1}{p} - 2\right)\log k$, we obtain the upper bound

$$\log Q \leq \left(\frac{1}{p} - 2\right) \log k \cdot \log\left[\left(\frac{1}{p} + 2\right) k\right] - \frac{k}{\left(\frac{1}{p} - 2\right) \log k} \cdot k^{\left(\frac{1}{p} - 2\right) \log (1-p)}$$

Expanding $1 + \left(\frac{1}{p} - 2\right) \log (1 - p)$ in powers of p, we see that $\log Q < -k^p$ when k is large, and we conclude that $Q \to 0$ as $k \to \infty$.

We now show that the failure of (5) implies (4). Indeed, suppose that A does not satisfy (5) and that B is a complement of A; we shall prove that each interval of length $\left(\frac{1}{p} + 2\right) k$ contains more than $\left(\frac{1}{p} - 2\right) \log k$ elements of B. This will then imply that

$$d(B) \geq \frac{\left(\frac{1}{p} - 2\right) \log k}{\left(\frac{1}{p} + 2\right) k} \geq (1 - 4p)(\log k)/k,$$

and the proof will be complete.

Consider the interval $I_n = \left(n - \left(\frac{1}{p} + 1\right) k, n + k\right]$. By the complement property, each integer in $[n + 1, n + k]$ is of the form $a + b$ ($a \in A$, $b \in B$), and the numbers b that arise in these representations automatically lie in I_n. Since the union of the translates of A by the numbers b in I_n covers $[n + 1, n + k]$, Lemma 2 implies that with probability greater than $1 - \varepsilon$, there are more than $\left(\frac{1}{p} - 2\right) \log k$ of these numbers. This completes the proof of Theorem 2.

Our theorems about sets of integers can easily be applied to other sets. Thus Theorem 1 implies the following proposition about point sets on the circle. *If* θ_1, θ_2, *and* θ_3 *are distinct* (modulo 2π), *then for each* $\varepsilon > 0$ *there exists a set* E *on the unit circle* C, *of measure less than* $\frac{4}{5}\pi + \varepsilon$, *such that*

$$e^{i\theta_1} E \cup e^{i\theta_2} E \cup e^{i\theta_3} E = C.$$

For some triples θ_1, θ_2, θ_3 *there does not exist such a set of measure* $\frac{4}{5}\pi$.

REFERENCES

1. P. Erdös, *Some results on additive number theory*, Proc. Amer. Math. Soc. 5 (1954), 847-853.

2. G. G. Lorentz, *On a problem of additive number theory*, Proc. Amer. Math. Soc. 5 (1954), 838-841.

Yeshiva University