Some additive and multiplicative problems
in number theory

by

S. L. G. Choi (Vancouver, B. C.), P. Erdős and E. Szemerédi (Budapest)

Introduction. In this paper we consider various additive and multiplicative problems concerning sets of integers. The major aim of our investigation is in exhibiting the relationship between the number of elements in a given set of positive integers not exceeding \( n \) and the number of integers that can always be chosen (with or without the restriction that these should lie in the given set) so that their sums (or products), taken two at a time, should all lie in the given set. We shall only once consider the analogous question relating to sums formed with a variable number of summands.

Notation. The letters \( c_1, c_2, \ldots \) denote positive absolute constants, unless otherwise indicated. A sum or product in this paper will mean, unless otherwise indicated, one formed with distinct integers. A sequence will always mean a strictly increasing sequence of positive integers.

1. Let \( A \) denote a set of \( n + t \) integers not exceeding \( 2n \). It is clear that if \( t = 1 \) then in general one cannot choose three integers from \( 1, 2, \ldots, 2n \) whose sums, taken two at a time, all appear in \( A \); for instance we may let \( A \) consist of \( 2 \) and all the odd integers not exceeding \( 2n \). It turns out, however (as Theorems 1-4 below show), that corresponding to \( t = 2, c_1, c_2 \log n, c_3 n^{1/2} \) respectively, we can always choose three, four, five, or six integers respectively so that in each case all sums, taken two at a time, will appear in the given sequence \( A \); further, these results are essentially best possible. Theorems 5 and 6 below give us some idea of the rate of growth of the smallest integer \( t_k \) \((k \geq 3)\) such that for any sequence of \( n + t_k \) integers not exceeding \( 2n \), we can always choose \( k \) integers all whose sums, taken two at a time, appear in the sequence.

**Theorem 1.** Suppose \( n \geq 4 \) and let \( A \) denote a sequence of \( n + 2 \) positive integers not exceeding \( 2n \). Then there are integers \( b_1, b_2, b_3 \) such that \( b_i + b_j \) \((1 \leq i < j \leq 3)\) are all in \( A \).
Proof. We assume the theorem false and proceed to deduce a contradiction. Accordingly, suppose there exists \( n \geq 4 \) and a sequence \( A \) of \( n+2 \) integers not exceeding \( 2n \) such that one can never choose \( b_1, b_2, b_3 \) with \( b_i + b_j \) \((1 \leq i < j \leq 3)\) in \( A \).

Let \( 2m + 1 \) be the smallest odd integer \( \geq 3 \) in \( A \). Then \( 3 \leq 2m + 1 < 2n \).

Since the sum of \( m + 1 \) and \( m \) is \( 2m + 1 \), for each integer \( j = m + 2, \ldots, 2n - (m + 1) \), at most one of the sums \( m + j, m + 1 + j \) belongs to \( A \). In other words, no two consecutive integers from \( 2m + 2, \ldots, 2n \) can belong to \( A \). In view of the choice of \( 2m + 1 \), there are at most \( m + 1 \) integers (i.e. consisting of 1 and the even integers) from 1, 2, \ldots, 2m belong to \( A \), which implies that at least \( n - m \) of the integers \( 2m + 2, \ldots, 2n \) belong to \( A \). As we have already shown that no two consecutive integers from \( 2m + 2, \ldots, 2n \) can belong to \( A \), the last sentence implies that there are precisely \( n - m \) integers from \( 2m + 2, \ldots, 2n \) belonging to \( A \) and that these are simply the even integers from \( 2m + 2, \ldots, 2n \); further all the even integers from 1, 2, \ldots, 2n also belong to \( A \). Consequently all the even integers from 1, 2, \ldots, 2n belong to \( A \) and these include the numbers 4, 6, 8 since \( n \geq 4 \). But then the number \( b_1 = 1, b_2 = 3, b_3 = 5 \) have all sums \( b_i + b_j \) \((1 \leq i < j \leq 3)\) belonging to \( A \). This gives the desired contradiction.

We remark that \( n \geq 4 \) in the above theorem is best possible since we cannot choose \( b_1, b_2, b_3 \) all whose sums \( b_i + b_j \) \((1 \leq i < j \leq 3)\) appear in 1, 2, 3, 4, 6.

The proofs of Theorems 2–4 below depend on the following lemma (cf. [3], Lemma p(\( \delta, b \))) and its corollary.

**Lemma A.** Suppose \( B \) denotes a sequence of positive integers not exceeding \( 2n \)
\[ y_1 < \ldots < y_t, \]
then, provided \( t \geq 2^k n^{1-2^{-k}} \), there exist positive integer \( x_0 \) and distinct positive integers \( x_1, \ldots, x_k \) such that \( B \) contains the subset:
\[ \{x_0\} + \{0, x_1\} + \ldots + \{0, x_k\}. \]

**Proof.** The proof is by induction on \( k \). Clearly the theorem is true for \( k = 1 \) or 2. Let now \( k \geq 2 \) and assume theorem holds for \( k \). We proceed to prove that the theorem holds also for \( k + 1 \). Accordingly let \( B \) denote a sequence (1) of integers not exceeding \( 2n \), where \( t \geq 2^{k+1} n^{1-2^{-k(k+1)}} \). Since there are \( \frac{1}{2}(t-1)t \) differences \( y_i - y_j \) \((1 \leq j < i \leq t)\) there exists some integer \( m \) such that there are \( t_1 > \{(t(t-1))/8n\} > t^2/(16n) \) distinct pairs \( y_i^* < y_i^{**} \) \((i = 1, \ldots, t_1)\) such that
\[ y_i^{**} - y_i^* = m \quad (i = 1, 2, \ldots, t_1). \]

It is clear that \( y_i^* \) \((i = 1, \ldots, t_1)\) are distinct and
\[ t_1 \geq 2^{2k+2} n^{2-2^{-k}} (16n)^{-1} \geq 2^k n^{1-2^{-k}}. \]
But then by the induction hypothesis there exists a subset of form

\[\{x_0\} + \ldots + \{0, x_k\}\]

in the set \(\{y_i^n; i = 1, \ldots, t_l\}\). Since for each \(y_i^n, y_i^n + m = y_i^n\) is also in \(B\) we conclude, by taking \(x_{k+1} = m\), that the set

\[\{x_0\} + \{0, x_1\} + \ldots + \{0, x_{k+1}\}\]

is also a subset of \(B\).

As a consequence of Lemma A, we prove the following

**Corollary.** Suppose \(n > n_o(k)\) and let \(A\) denote a sequence of \(t\) even integers not exceeding \(2n\), where \(t > 2^k n^{1 - 2^{-k}}\). Then there exist integers \(b_0, \ldots, b_k\) such that all \(b_i + b_j\) \((0 < i < j < k)\) appear in \(A\).

**Proof.** By the lemma, \(A\) possesses a subset of type

\[\{x_0\} + \{0, x_1\} + \ldots + \{0, x_k\}\]

We take \(b_0 = \frac{1}{2} x_0, b_1 = \frac{1}{2} x_0 + x_1, \ldots, b_k = \frac{1}{2} x_0 + x_k\). Since \(x_0\) is an even integer, \(b_0, b_1, \ldots, b_k\) are integers whose sums \(b_i + b_j\) \((0 < i < j < k)\) are all in \(A\).

**Theorem 2.** There exists a positive integer \(c_1\) such that if \(n > n_0(c_1)\) and \(A\) denotes a sequence of \(n + c_1\) positive integers not exceeding \(2n\), then there are \(b_1, b_2, b_3, b_4\) so that all sums \(b_i + b_j\) \((1 < i < j < 4)\) are in \(A\).

**Proof.** Let \(c_1\) be a sufficiently large integer. Let \(t\) denote the number of even integers in \(A\). Then

\[c_1 \leq t \leq 10^{-2} n,\]

the latter inequality holding in view of the corollary to Lemma A, if \(n\) is chosen large enough. By the same corollary, provided \(c_1\) is chosen large enough, we may assert that there exists an even integer \(2m\) in \([20t, 2n - 20t]\). Let \(b_1\) and \(b_2\) be even integers defined by

\[b_1 + b_2 = 2m,\]

\[b_2 - b_1 = \begin{cases} 2 & \text{if } m \text{ is odd,} \\ 4 & \text{if } m \text{ is even.}\end{cases}\]

If \(a\) is any integer in \([m - 10t, m + 10t]\) then certainly

\[0 < a + b_1 \leq 2n,\]

\[0 < a + b_2 \leq 2n.\]

Now there are \(5t\) pairs of odd integers \(x, y\) in \([m - 10t, m + 10t]\) such that \(x + y = 2m\). For each \(i = 1, 2\), there are at most \(t\) odd integers \(a\) in \([m - 10t, m + 10t]\) such that \(b_i + a\) is not an integer in \(A\). Thus there exist at least \(3t\) pairs of odd integers \(x, y\) in \([m - 10t, m + 10t]\) with \(x + y = 2m\) and such that \(b_1 + x, b_2 + x, b_1 + y, b_2 + y\) are all in \(A\). Let \(b_3, b_4\) be one such pair.
Then \( b_1, b_2, b_3, b_4 \) are integers such that all the sums \( b_i + b_j \) \((1 \leq i < j \leq 4)\) are in \( A \).

**Theorem 3.** There exists an absolute constant \( c_2 > 0 \) such that if \( n \geq n_0(c_2) \) and \( A \) is a sequence of \( n + m \) positive integers not exceeding \( 2n \), where \( m \geq c_2 \log n \), then there are integers \( b_1, b_2, b_3, b_4, b_5 \) such that \( b_i + b_j \) \((1 \leq i < j \leq 5)\) are all in \( A \). Further, the result no longer holds if \( c_2 \) is replaced by \( c'_2 \), where \( c'_2 \) is sufficiently small.

**Proof.** Let \( t \) denote the number of even integers in \( A \). Then, provided \( n \) is sufficiently large, we may assume

\[
c_2 \log n \leq t \leq 10^{-2} n,
\]

the right-hand side inequality holding since otherwise an application of the corollary to Lemma A (with \( k = 4 \)) gives the theorem. In view of the corollary again, provided \( c_2 > 0 \) is sufficiently large, and \( n \geq n_0(c_2) \), there are at least \( 2 \log_2 c \log n \) even integers from our sequence \( A \) falling into the interval \([40t, 2n - 40t]\). Therefore there exists a subinterval \([n_t, 2n]\) containing three even integers \( a_1, a_2 < a_3 < a_4 \) from \( A \). Let the integers \( b_1, b_2, b_3 \) be determined by

\[
\begin{align*}
b_1 + b_2 &= a_1^*, \\
b_1 + b_3 &= a_2^*, \\
b_2 + b_3 &= a_3^*.
\end{align*}
\]

We thus obtain

\[
\begin{align*}
b_1 &= \frac{1}{2}(a_1^* + a_2^* - a_3^*), \\
b_2 &= \frac{1}{2}(a_1^* - a_2^* + a_3^*), \\
b_3 &= \frac{1}{2}(a_3^* - a_1^* + a_2^*).
\end{align*}
\]

It is clear that \( b_1, b_2, b_3 \) are either all odd or all even. Suppose they are all odd (the case when \( b_1, b_2, b_3 \) are all even can be treated similarly). There are \( 10t \) pairs of even integers \( b_4 < b_5 \) in \([a_i^*/2 - 20t, a_i^*/2 + 20t]\) such that \( b_4 + b_5 = a_i^* \). We note that for any \( a \) in \([a_i^*/2 - 20t, a_i^*/2 + 20t]\), \( a + b \leq 2n \) \((i = 1, 2, 3)\). We choose a pair \( b_4 < b_5 \) such that \( b_4 + b_1, b_4 + b_2, \\
b_4 + b_3, b_5 + b_1, b_5 + b_2, b_5 + b_3, \) are all in \( A \). This is possible since for each \( i = 1, 2, 3 \) there are at most \( t \) even integers \( a \) in \([a_i^*/2 - 20t, a_i^*/2 + 20t]\) such that \( a + b \) is not in \( A \). This proves the main part of the theorem.

Finally, if \( A \) consists of all the odd integers and the integers \( 2, 2^2, 2^3, \ldots, \) in \([1, 2n]\) then one cannot choose \( b_1, \ldots, b_5 \) such that \( b_i + b_j \) \((1 \leq i < j \leq 5)\) are all in \( A \). This completes the proof of Theorem 3.

**Theorem 4.** There exists \( c_3 > 0 \) such that if \( n \geq n_0(c_3) \), and \( A \) is a sequence of \( n + m \) positive integers not exceeding \( 2n \), where \( m \geq c_3 n^{1/2} \), then one can find six integers \( b_1, \ldots, b_6 \) whose sums \( b_i + b_j \) \((1 \leq i < j \leq 6)\) are
all in \( A \). Further, the results becomes false if \( c_3 \) is replaced by a sufficiently small constant \( c'_3 \).

**Proof.** Let \( t \) denote the number of even integers in \( A \). Then we can assume, in view of the corollary to Lemma A, that

\[
c_3 n^{1/2} \leq t \leq 10^{-2} n,
\]

and that there are at least \( \frac{1}{2} t \geq 12n^{1/2} \) even integers of \( A \) falling into the interval \([40t, 2n - 40t]\). Thus, if \( c_3 \) is sufficiently large, there exists a subinterval \([n_1, 2n_1]\) containing at least \( 3n_1^{1/2} \) even integers of \( A \). Since the sum of any two integers in \([n_1, 2n_1]\) lies between \( 2n_1 \) and \( 4n_1 \), there exist even integers \( z_1, z_2, z_3, z_4, z_5, z_6 \) of \( A \) such that

\[
z_1 + z_2 = z_3 + z_4 = z_5 + z_6
\]

and

\[
n_1 \leq z_6 < z_3 < z_1 < z_2 < z_4 < z_6 \leq 2n_1.
\]

We determine integers \( b_1, b_2, b_3, b_4 \) such that

\[
b_1 + b_2 = z_1,
b_3 + b_4 = z_2,
b_1 + b_3 = z_3,
b_1 + b_4 = z_5
\]

and thus also

\[
b_2 + b_4 = z_4,
b_2 + b_3 = z_6.
\]

It is clear that \( b_1 < b_4 < b_3 < b_2 \) and that they are all odd or all even. Solving for \( b_1, b_2, b_3, b_4 \) gives

\[
b_1 = \frac{1}{2}(z_3 - z_2),
b_2 = \frac{1}{2}(2z_1 - z_3 + z_2 - z_6),
b_3 = \frac{1}{2}(z_2 + z_3 - z_5),
b_4 = \frac{1}{2}(z_3 - z_2 + z_6).
\]

Since clearly \( b_1 > 0 \) we have \( b_2 > 0, b_3 > 0, b_4 > 0 \) as well.

If \( b_1, \ldots, b_4 \) are all odd (even) then we determine even (odd) integers \( b_5, b_6 \) in \([\frac{1}{2} z_3 - 20t, \frac{1}{2} z_3 + 20t]\) such that

\[
b_5 + b_6 = z_5
\]

and such that \( b_i + b_6 \) \((i = 1, 2, 3, 4)\) and \( b_i + b_3 \) \((i = 1, 2, 3, 4)\) are all in \( A \). This is possible since for each \( i = 1, 2, 3, 4 \) there exist at most \( t \) even
(odd) integers \( a \) in \([\frac{1}{2}z_5 - 20t, \frac{1}{2}z_5 + 20t]\) such that \( b_i + a \) does not belong to \( A \); but there are initially 10\( t \) possible choices for \( b_5, b_6 \) such that \( b_5 + b_6 = z_5 \).

To prove the last part of the theorem we let \( A \) consist of all the odd integers \(< 2nt \) and \( c_3 n^{1/2} \) even integers \( = 2(4) \) so that the sums taken two at a time of these even integers are distinct. Suppose in fact there exist \( b_1, \ldots, b_6 \) such that \( b_i + b_j \ (1 \leq i < j \leq 6) \) are all in \( A \). We shall deduce a contradiction. Clearly at most two of the integers \( b_i \) can be even for otherwise we have a sum \( = 0 \ (4) \). Thus there are four odd \( b_i \), say \( b_1, b_2, b_3, b_4 \). The sums \( b_1 + b_2, b_3 + b_4, b_1 + b_3, b_2 + b_4 \) are in \( A \). But then

\[
(b_1 + b_2) + (b_3 + b_4) = (b_1 + b_3) + (b_2 + b_4),
\]

violating our choice of the even numbers in \( A \).

We summarize the results contained in Theorems 1–4 as follows. We first recall the definition of \( t_k \) in the opening paragraph of this section. For large \( n \), Theorems 1–4 reveal that the order of magnitude of \( t_k \) (\( k = 3, 4, 5, 6 \)) is known. More precisely

\[
t_3 = 2, \quad 2 < t_4 \leq c_1, \quad c_2 \log n \leq t_5 \leq c_2 \log n, \quad c_3 n^{1/2} \leq t_6 \leq c_3 n^{1/2},
\]

where \( c_1, c_2, c_3, c_4, c_5, c_6 \) are positive absolute constants. It might be of interest to determine these constants precisely. For \( k \geq 7 \), the order of magnitude of \( t_k \) is not known, but Theorems 5 and 6 below give some indication of the possible rate of growth of \( t_k \). We mention that a slightly more precise form of Theorem 5 below is possible; but as there is no indication that Theorem 5 is anywhere near the best possible we shall not aim at precision here.

**Theorem 5.** Let \( k \) be a positive integer and \( n \geq n_0(k) \), and suppose \( A \) is a sequence of \( n + t \) positive integers not exceeding \( 2n \), where \( t \geq 2^k n^{1-2^{-k}} \). Then there exist integers \( b_0, \ldots, b_n \) all whose sums \( b_i + b_j \ (0 \leq i < j \leq k) \) are in \( A \).

**Proof.** Since there are at least \( 2^k n^{1-2^{-k}} \) even integers in \( A \), the theorem follows from the corollary of Lemma A.

**Corollary.** If \( A \) is a sequence of \( n + t \) positive integers not exceeding \( 2n \), where \( t \geq 5n \), and \( n \geq n_0(\delta) \), then we can find integers \( b_1, \ldots, b_k \) where \( k \leq \log\log n \), with the implied constant depending on \( \delta \), such that all sums \( b_i + b_j \ (1 \leq i < j \leq k) \) are in \( A \).

**Proof.** By Theorem 5 we can always choose \( b_1, \ldots, b_k \) if

\[
n^{-2^{-k}} 2^k \leq \delta
\]

which is valid if \( k \leq \log\log n \).
Before stating our next theorem, we prove a result concerning the frequency of occurrence of sequences with few distinct sums (taken two at a time).

**Lemma B.** Suppose \( a_1 \) is given. Then there exist \( k_1 = k_1(a_1) \) and \( a_2 \) depending only on \( a_1 \), such that, if \( k \geq k_1 \) and \( n \geq n_1(k, a_1) \), the number of choices of sequences \( A \)

\[
a_1 < \ldots < a_k \leq n
\]
each with \( \leq a_1 k \) distinct sums (taken two at a time), does not exceed \( n^{a_2} \).

We deduce the lemma from the following theorem of Freiman (see [2], p. 134) reworded to suit our present purposes.

**Theorem A.** Suppose the sequence \( A \)

\[
a_1 < \ldots < a_k
\]
is such that there are at most \( c k \) distinct sums \( a_i + a_j \ (1 \leq i < j \leq k) \), then there exist \( k^* \), \( c^* \) depending only on \( c \), and an integer \( m \leq c - 1 \), such that, if \( k \geq k^* \), there are arithmetic progressions \( B_0, B_1, \ldots, B_m \) each of length at most \( c^* k \) such that \( A \) is contained in the set \( S_m \), where the sets \( S_i \ (i = 0, 1, \ldots, m) \) are defined inductively by

\[
\begin{align*}
S_0 &= B_0, \\
S_i &= \bigcup_{b \in B_i} (S_{i-1} + b), \ i \geq 1.
\end{align*}
\]

**Proof of Lemma B.** We apply Theorem A with \( c = a_1 \). Then we have \( k_1 = k^* \) such that if \( k \geq k_1 \), the sequence \( A \) is contained in \( S_m \), with \( S_m \) defined by (2). The number of choices for \( B_i \ (i = 0, \ldots, m) \) is at most \( n^2 \). Thus the total number of choices for \( S_m \) is \( \leq n^{2 a_2} \). Now the number of choices of \( A \) corresponding to each choice of \( S_m \) is \( \leq (c^* k)^{2 a_2} \). Therefore, the total number of choices of \( A \) is

\[
\leq n^{2 a_2} (c^* k)^{2 a_2} \leq n^{a_2},
\]

where \( a_2 \) depends only on \( a_1 \) if we choose \( n \geq n_1(k, a_2) \).

**Theorem 6.** Suppose \( 0 < \varepsilon < 1 \) is given. Then there exist \( k_0(\varepsilon) \) and \( n_0(k_0) \) such that if \( n \geq n_0 \), there exists a sequence \( A \) of \( n+t \) positive integers consisting of all the odd integers \( \leq 2n \) and \( t \) positive even integers \( \leq 2n \), where \( t = [n^{1-\varepsilon}] \), such that there are at most \( k_0(\varepsilon) - 1 \) integers

\[
b_1, \ldots, b_{k_0(\varepsilon) - 1}
\]
all whose sums \( b_i + b_j \ (1 \leq i < j \leq k_0(\varepsilon) - 1) \) are in \( A \).

**Proof.** Let \( a = [2/\varepsilon] + 1 \). We apply Lemma B with \( a_1 = a^2 \). Let \( k_0 = 2 a_2 a k_1 \), where \( k_1 \) and \( a_2 \) are the numbers in Lemma B corresponding to \( a_1 = a^2 \). Finally let \( n_0 = n_0(k_0, a_1) \) be the choice of \( n_1 \) in Lemma B.
corresponding to \( k = k_0 \). We shall establish the theorem with these choices of \( k_0 \) and \( n_0 \). Accordingly let \( n \geq n_0 \) and we proceed to establish the existence of a sequence \( A \) with the desired property.

We determine first the number of choices of sequences \( B \)

\[
b_1 < \ldots < b_{k_0} \leq n
\]

so that the number of distinct even sums \( b_i + b_j \) is \( \leq ak_0 \). We let \( B^*, B^{**} \) denote the subsequences of \( B \) consisting of respectively the odd and even integers of \( B \). Further we denote by \( T(B^*), T(B^{**}) \) the number of distinct sums (taken two at a time) formed from the integers of \( B^* \) and \( B^{**} \) respectively. We have

\[
T(B^*) < ak_0, \\
T(B^{**}) < ak_0.
\]

We consider two cases according as both \( B^*, B^{**} \) have each \( \geq a^{-1}k_0 > k_1 \) integers or otherwise. Take the first case and let \( M_1 \) denote the number of choices of \( B \) in this case. Then

\[
T(B^*) < a^2|B^*| = a_1|B^*|
\]

and similarly

\[
T(B^{**}) < a_1|B^{**}|.
\]

Since \( |B^*| > k_1 \) and \( |B^{**}| > k_1 \) we may apply Lemma B to \( B^* \) and \( B^{**} \) to conclude that

\[
M_1 \leq n^{2a_2}.
\]

We next consider the second case. Let \( M_2 \) denote the number of choices of \( B \) in this case. One of the sets \( B^*, B^{**} \) has \( \leq a^{-1}k_0 \) integers and thus the number of choices for this set is \( \leq n^{k_0/a} \). The number of choices for the other set is \( \leq n^{a_2} \), by an application of Lemma B. Thus

\[
M_2 \leq 2n^{k_0a^{-1}+a_2}.
\]

Thus the number of choices of \( B \) each with \( \leq ak_0 \) distinct even sums is

\[
M_1 + M_2 \leq n^{2a_2} + 2n^{k_0a^{-1}+a_2} \leq n^{3a_2 - 1}k_0,
\]

since \( k_0 = 2a_2ak_1 \).

Each such sequence \( B \) determines at least \( k_0 - 3 \) even sums, so corresponding to a given \( B \), there exist \( \leq \binom{n-k_0-3}{t-k_0-3} \) choices of \( A \) containing these sums. Let \( N_1 \) denote the number of choices of \( A \) corresponding to these \( B \). Then

\[
N_1 \leq \binom{n-k_0+3}{t-k_0+3} n^{3a_2 - 1}k_0.
\]
We now consider sequences $B$ having each at least $a k_0$ distinct even sums. Each such sequence determines at least $a k_0$ distinct even sums and thus the number of choices of $A$ containing these even sums is at most $\binom{n}{t-[a k_0]}$. As there are $\leq \binom{n}{k_0}$ choices for such $B$, the number $N_2$ of choices of $A$ corresponding to all such $B$ satisfies

\[ N_2 \leq \binom{n-[a k_0]}{t-[a k_0]} \binom{n}{k_0}. \]

Since the total number of possible choices of $A$ is $\binom{n}{t}$ we have our theorem if we can prove

\[ \binom{n}{t} > N_1 + N_2. \]

We shall establish this by showing that

\[ N_1 < \frac{1}{2} \binom{n}{t}, \quad N_2 < \frac{1}{2} \binom{n}{t}. \]

We have

\[ \binom{n}{t} \binom{n-[a k_0]+3}{t-[a k_0]+3} \geq n^{\epsilon k_0 + O(1)} \geq 2n^{\frac{3}{2}a^{-1}k_0} \]

on recalling $a = \lceil 2/e \rceil + 1$ and $t = \lceil n^1/e \rceil$. The above inequality implies $N_1 < \frac{1}{2} \binom{n}{t}$ in view of (3).

Next

\[ \binom{n}{t} \binom{n-[a k_0]}{t-[a k_0]} \geq n^{\epsilon \epsilon k_0 + O(1)} \geq 2 \binom{n}{k_0} \]

on using $a = \lceil 2/e \rceil + 1$ and $t = \lceil n^{1/e} \rceil$. We have $N_2 < \frac{1}{2} \binom{n}{t}$ in view of (4). This completes the proof of Theorem 6.

2. In this section we consider the question of estimating the number of integers that can be chosen from a given sequence so that all sums, taken two at a time, should appear in the sequence. We shall prove three theorems (Theorem 7, 8, and 9) of which the last depends on the following theorem which has just been established by Szemerédi.

**Theorem B.** For any given integer $k \geq 2$ let $r_k(n)$ denote the largest number of integers that can be chosen from 1, 2, ..., $n$ with no $k$ terms in arithmetic progression. Then $n^{-1}r_k(n) \to 0$ as $n \to \infty$.

We further remark that Theorem 7 would also follow from Szemerédi's result though we give a proof which uses only a theorem of Varnavides.
THEOREM 7. For any given \( \epsilon > 0 \) and any integer \( k > 1 \), there exists \( n_0(\epsilon, k) \) so that if \( n \geq n_0 \) and \( A \) is a sequence of \( t \) integers not exceeding \( n \), where \( t \geq (\frac{2}{3} + \epsilon)n \), then we can find \( k \) integers

\[
a_1, a_2, \ldots, a_k
\]

in \( A \) whose sums \( a_i + a_j \) (\( 1 \leq i < j \leq k \)) are all in \( A \).

Proof. Since \( t \geq (\frac{2}{3} + \epsilon)n \) there exist \( s \) integers, where \( s \geq \epsilon_1 n \), in the sequence \( A \), say \( a_1, a_2, \ldots, a_s \) such that \( 2a_1, 2a_2, \ldots, 2a_s \) are also in \( A \). By a theorem of Varnavides (see [4]) there are \( e_2 n^2 \) triples \( a_1, a_2, a_3 \) which form an arithmetic progression. Thus there is an integer, say \( a_{i_1} \), for which there are \( \geq \epsilon_2 n \) integers \( a_{i_j}'s \) so that

\[
\frac{1}{2}(a_{i_1} + a_{i_j}) = a_{i_j};
\]

but then \( a_{i_1} + a_{i_j} = 2a_{i_j} \) is also in \( A \). Now repeat the same argument with these \( \epsilon_2 n \) \( a_{i_j}'s \), and so on. In this way one can find integers \( a_{i_1}, a_{i_2}, \ldots, a_{i_k} \) in \( A \) such that \( a_{i_u} + a_{i_v} \) (\( 1 \leq u < v < k \)) are all in \( A \).

The following theorem is a refinement of Theorem 7.

THEOREM 8. Suppose \( k \) is given. Then there exists \( \epsilon_k > 0 \) such that if \( n \geq n_0(\epsilon_k, k) \) and \( A \) is a sequence of \( t \) integers not exceeding \( n \), where \( t \geq (\frac{2}{3} - \epsilon_k)n \), then one can find \( k \) integers in \( A \)

\[
a_1, a_2, \ldots, a_k
\]

whose sums \( a_i + a_j \) (\( 1 \leq i < j \leq k \)) are all in \( A \).

Proof. Let \( \epsilon_k > 0 \) be a sufficiently small number. In view of Theorem 7, we may assume there are at most \( \epsilon_\kappa n \) integers \( a \) in \( A \) such that \( 2a \) is also in \( A \). Thus there exists a subset \( B \) of \( A \) with at least \((\frac{2}{3} - 2\epsilon_k)n \) integers and with the property that whenever \( a \) belongs to \( B \) then \( 2a \) does not belong to \( B \). This property is crucial in our proof and we refer to it as property \( P \).

For \( j = 1, \ldots, k \), let

\[
I_j = (n2^{-j}, n2^{-j+2}], \quad I_j^* = (n2^{-j-1}, n2^{-j}],
\]

\[
B_j = B \cap I_j, \quad B_j^* = B \cap I_j^*.
\]

As property \( P \) implies that

\[
|B_j| + |B_j^*| \leq 2^{-j}n \quad (j = 1, 2, \ldots, k)
\]

and as

\[
|B| > (\frac{2}{3} - 2\epsilon_k)n,
\]

we conclude that

\[
|B_j| + |B_j^*| \geq (2^{-j} - 2\epsilon_k)n.
\]
Further, by repeated application of property P and using (5), we may assert that for each $j = k, k-1, \ldots, 1$, and $i = 0, 1, \ldots, k-j$, $B_j$ contains all but at most $2(i+1)\varepsilon_k n$ integers of type $4^j x$ (where $x$ is odd) in $I_j$. By now choosing $\varepsilon_k$ small enough, we can find an integer $b_1$ of type $x_1$ in $B_k$, an integer $b_2$ of type $4x_2$ in $B_{k-1}$, \ldots, and an integer $b_k$ of type $4^{k-1}x_k$ in $B_1$, where $x_1, \ldots, x_k$ are all odd, such that $b_i + b_j$ $(1 \leq i < j \leq k)$ are all in $B$ and thus in $A$. This completes the proof.

**Theorem 9.** For any integer $r \geq 2$, and any integer $k$, there exist $\delta_r > 0$ and $n_0(\delta_r, k)$ such that if $n \geq n_0(\delta_r, k)$ and $A$ is a sequence of $t$ positive integers not exceeding $n$, where $t \geq (1 - \delta_r)n$, then there exists a subsequence

$$a_1 < \ldots < a_k$$

such that all sums of the form

$$\sum_{j=1}^{k} \varepsilon_j a_j \quad (\varepsilon_j = 0, 1; 1 \leq \sum_{j=1}^{k} \varepsilon_j \leq r)$$

are in $A$.

**Proof.** We choose $\delta_r = 1/(2r^2)$ and suppose $n \geq n_0(\delta_r, k)$. Then there exist $s \geq n/(2r^2)$ and a subsequence of $A$

$$a_1 < \ldots < a_s$$

in $[(r-1)r^{-2}n, r^{-3}n]$ such that $2a_j, 3a_j, \ldots, ra_j$ ($j = 1, 2, \ldots, s$) are all in $A$. By Theorem B, we can find an arithmetic progression

$$a, a+b, \ldots, a+r!(k-1)^2 b$$

within $a_1, \ldots, a_s$. Now we take

$$b_1 = a, \quad b_2 = a+r!b, \quad \ldots, \quad b_k = a+r!(k-1)b.$$ 

Clearly $\sum_{i=1}^{k} \varepsilon_i b_i$, subject to $1 \leq \sum_{i=1}^{k} \varepsilon_i \leq r$, are all in $A$. This completes the proof of the theorem.

3. In this section we consider some aspects of the multiplicative analogue of the additive problems in § 1 and 2. Theorems 10–12 below represent the type of results that can be established by probabilistic arguments.

**Theorem 10.** Suppose $c_4$ is any positive integer and $n \geq n_0(c_4)$. Then there exists a sequence $A$ of $k$ positive integers not exceeding $n$, where $k \geq n(1 - e^{-c_4 \log n/\log \log n})$ such that for any $s$, where $s$ is an integer or the reciprocal of one, there exist at most $t \leq e^{c_4 \log n/\log \log n}$ integers

$$b_1 < \ldots < b_t,$$

where $c_3$ depends only on $c_4$, such that all products $s^{-1}b_i b_j$ $(1 \leq i < j \leq t)$ are in $A$. 

---

Additive and multiplicative problems in number theory
Proof. Let \( t = [e^{c_s \log n} \log \log n] + 1 \), where \( c_s \) is a sufficiently large constant depending on \( c_s \). Let \( k = n - [nm^{-1}] \), where \( m = e^{c_s \log n} \log \log n \). Suppose \( B \) is a sequence of \( t \) integers

\[ b_1 < \ldots < b_t. \]

We first estimate the number of sequences \( A \)

\[ a_1 < \ldots < a_k \leq n \]

which contain all products \( b_i b_j s^{-1} (1 \leq i < j \leq t) \) for a given \( s \), where \( s \) is an integer or the reciprocal of one.

Since \( d(l) < 2^{(1+\varepsilon) \log \log l} \) for \( l \geq l_1(\varepsilon) \), where \( d(l) \) denotes the divisor function, the number of distinct products \( b_i b_j \) determined by (6) is

\[ \geq 2^{-1} t(t-1) 2^{-(1+\varepsilon) \log n} \log \log n + O(1) \geq t^{3/2}, \]

if \( c_s \) is chosen large enough. Thus, if \( A \) contains all \( s^{-1} b_i b_j \) for a fixed \( s \), at least \( h = [t^{3/2}] \) of its integers are fixed by \( B \) and thus the number of choices of \( A \) is at most

\[ \binom{n-h}{k-h}. \]

Hence, on allowing \( s \) to vary, the number of possible choices of \( A \) corresponding to a given \( B \) is at most

\[ n^2 \binom{n-h}{k-h}. \]

The number of choices of \( B \) is \( \binom{n}{t} \). Since the number of choices of \( A \) (without restriction) is \( \binom{n}{k} \), the theorem would follow if we can prove

\[ (n) \geq (n) n^2 (k-h). \]

We have

\[ \binom{n}{k} / \binom{n-h}{k-h} = \frac{n \ldots (n-h+1)}{k \ldots (k-h+1)}. \]

For each \( i = 0, \ldots, h-1, \)

\[ (n-i)(k-i)^{-1} \geq (n-h)(n-2nm^{-1})^{-1} \geq 1 + m^{-1}. \]

Therefore,

\[ \frac{n}{k-h} \geq (1 + m^{-1})^h \geq e^{h/(2m)} \geq e^{t^{1/3}} \]

since \( h = [t^{3/2}] \) and \( m = e^{c_s \log n} \log \log n \leq t^{1/2} c_s. \) But \( \binom{n}{t} n^2 \leq t^{1/2} n^t \leq e^{t^{1/3}}. \)

Thus we have (7) as required.
The following lemma, whose proof is somewhat involved (see [4]) enables us to strengthen Theorem 10.

**Lemma C.** Suppose $k$ is any positive integer, and $t \geq (\log_2 n)^k$, where $\log_2 n$ denotes the logarithmic function to the base 2. Then for any sequence $a_1 < \ldots < a_t < n$ of $t$ positive integers, there are at least $c k^2 t$ distinct products $a_i a_j$ ($1 \leq i < j \leq t$) where $c$ is a positive absolute constant.

Using the above lemma we obtain the following

**Theorem 11.** Suppose $0 < a < 1$, and $n > n_0(a)$. Then there exists a sequence of $k$ positive integers not exceeding $n$, where $k \geq an$, such that for any $s$, where $s$ is an integer or the reciprocal of one, there exist at most

$$t = \left[ e^{c_6 (\log n)^k} \right] = \left[ (\log n)^{c_6 (\log n)^k} \right]$$

integers $b_1 < \ldots < b_t$, where $c_6$ depends only on $a$, such that all products $s^{-1} b_i b_j$ ($1 \leq i < j \leq t$) are in $A$.

**Proof.** Arguing as in proof of Theorem 10 and using Lemma C instead of $d(l) < 2^{(1 + \varepsilon) \log l / \log \log l}$, we need only prove that

$$\left( \frac{n}{[an]} \right) / \left( \frac{n - [c_7 (\log n) t]}{[an] - [c_7 (\log n) t]} \right) \geq n^2 \binom{n}{t}.$$

We note that the left hand side is $\geq e^{(\log n)^k \log \log n}$ which is greater than $n^2 \binom{n}{t}$, if $c_6$ and hence also $c_8$ is large enough in terms of $a$.

It seems quite plausible that the following conjecture is true:

Suppose $a_1 < \ldots < a_t < n$, $t > (\log_2 n)^k$. Then there are $(1 + \varepsilon)^k t$ distinct products $a_i a_j$ ($1 \leq i < j \leq t$), where $c$ is some positive absolute constant.

The above conjecture, if true, would imply the following

**Theorem 12.** Suppose $0 < a < 1$ and $n > n_0(a)$. Then there exists a sequence $\mathcal{A}$ of $k$ positive integers not exceeding $n$, where $k \geq an$, such that for any $s$, where $s$ is an integer or the reciprocal of one, there exist at most

$$t = \left[ e^{c_9 (\log \log n)^2} \right]$$

integers $b_1 < \ldots < b_t$, where $c_9$ depends only on $a$, such that $b_i b_j s^{-1}$ ($1 \leq i < j \leq t$) are all in $\mathcal{A}$.

The proof, which we omit, is an adaptation of the probabilistic argument used in the proof of Theorem 10. The theorems in this paragraph can undoubtedly be sharpened considerably. We hope to return to these questions at another occasion.