Some Extremal Problems in Combinatorial Number Theory

Paul Erdős

Nemetvolgyi UT 72C Budapest XII Hungary and Department of Mathematics University of Colorado Boulder, Colorado 80302 U.S.A.

In this note I discuss problems in number theory most of which have a combinatorial flair. Section 2 is a joint work with A. Sárkőzi and E. Szemerédi.

First we introduce some notations which will be used frequently in this paper. The sequence a_1, a_2, \ldots , will be denoted by $A, A(x) = \sum_{a_i \leq x} 1$. The limit, $\lim_{x \to \infty} A(x)/x$, if it exists, is called the density of A (the upper density is the lim sup of the same expression). The term V(n) denotes the number of prime factors of n, and V(n, l), the number of prime factors of n not exceeding l (in both cases multiple factors are counted multiply). The symbols c, c_1, \ldots , will denote positive absolute constants not necessarily the same at each occurrence; ε , δ , η denote positive numbers which can be chosen arbitrarily small. The letters a, b, t, l, \ldots denote integers; p is a prime; P(t) is the greatest and p(t) the least prime factor of t.

1.

Denote by f(k, x) the maximum number of integers $a_1 < \cdots < a_r \le x$ so that no k of them have pairwise the same common divisor. I have proved [6] that for every k if $x > x_0(k)$

Some Extremal Problems in Combinatorial Number Theory

(1)
$$\exp\left(c_k \frac{\log x}{\log \log x}\right) < f(k, x) < x^{3/4+\epsilon}$$
, where $\exp z = e^z$.

It was conjectured in [6] that the lower bound seems to give the right order of magnitude for f(k, x).

Denote by F(k, x) the maximum number of integers $a_1 < \cdots < a_s \leq x$ so that no k of them have pairwise the same least common multiple. I conjectured that F(k, x) = o(x) for every $k \geq 3$. Recently, I proved that for $k \geq 4$ this conjecture is certainly false. At present I cannot disprove this conjecture for k = 3.

The falseness of the conjecture will easily follow from the following result which is of independent interest:

Theorem 1. The density of integers having three relatively prime divisors satisfying $b_1 < b_2 < b_3 < 2b_1$ exists and is less than 1.

I have proved [7] that the density of integers having two relatively prime divisors $b_1 < b_2 < 2b_1$ is 1. The proof has not been published and is quite complicated, but we will not need this result here.

Let us assume that Theorem 1 is already proved. Then consider the integers $x/2 < a_1 < \cdots < a_s < x$ no one of which has three pairwise relatively prime divisors $b_1 < b_2 < b_3 < 2b_1$. By our theorem s > cx. Now we show that there are no four a's, say $a_1 < a_2 < a_3 < a_4$, satisfying

(2)
$$[a_i, a_j] = T, \quad 1 \le i < j \le 4.$$

To see this assume that (2) holds. Put $T/a_i = b_i$, $1 \le i \le 4$. Clearly $b_j | a_i$ for $j \ne i$ and $(b_i, b_j) = 1$, $1 \le i < j \le 4$. Finally from $x/2 < a_1 < a_2 < a_3$ $< a_4 < x$ we obtain $b_2 < b_3 < b_4 < 2b_2$. Thus a_1 would have three divisors $b_2 < b_3 < b_4 < 2b_2$, $(b_i, b_j) = 1$, $2 \le i < j \le 4$, which contradicts our assumptions. Hence F(4, x) > cx as stated.

Thus we only have to prove Theorem 1. First we show the following:

Lemma 1. Let $1 < u_1 < \cdots$ be any sequence of integers. Denote by d the density, and by $\overline{d}(u_1, \ldots)$ the upper density of the integers having at least one divisor amongst the u's. Assume that for every $\varepsilon > 0$ there is a k satisfying

$$(3) \qquad \qquad \vec{a}(u_{k+1},\ldots) < \varepsilon.$$

Then $d(u_1, \ldots)$ exists and is less than 1.

A theorem of Behrend [2] states that if $a_1 < \cdots < a_k$ and $b_1 < \cdots < b_l$ are any two sequences of integers then

(4) $1 - d(a_1, \ldots, a_k, b_1, \ldots, b_l) \ge (1 - d(a_1, \ldots, a_k))(1 - d(b_1, \ldots, b_l)).$

Paul Erdös

From (3) and (4) we obtain by a simple limiting process that

(5)
$$1 - d(u_1, \ldots) > (1 - \eta)(1 - d(u_1, \ldots, u_k)).$$

Inequality (5) easily implies Lemma 1. (The term $d(u_1, \ldots, u_k)$ clearly exists for every finite set u_1, \ldots, u_k .)

Now let $u_1 < \cdots$ be the sequence of integers which can be written in the form

(6)
$$b_1b_2b_3, b_1 < b_2 < b_3 < 2b_1, (b_i, b_j) = 1, 1 \le i < j \le 3.$$

To prove Theorem 1 it suffices to show that the u's satisfy (3). Denote by $m_1 < \cdots$ the integers which are divisible by at least one u_i , i > k; we have to show that for $k > k_0(\varepsilon)$ the upper density of the m's is less than ε . A theorem of mine states [8] that for every ε and δ there is an *l* such that the density of integers *n* which for some t > l do not satisfy

$$(1 - \delta)\log \log t < V(n, t) < (1 + \delta)\log \log t$$

is less than $\varepsilon/2$. Thus to prove that the *u*'s satisfy (3) it will suffice to show that for sufficiently small δ and $k > k_0(\varepsilon, \delta, l)$ the upper density of the *m*'s satisfying

(7)
$$V(m_i, t) < (1+\delta)\log\log t$$

for every t > l is less than $\varepsilon/2$.

Showing that this statement is true will be the main difficulty of our proof. First of all observe that for $k > k_0(l)$ (6) implies that every m_j has a divisor of the form

(8) $b_1b_2b_3, l < 2^s < b_1 < b_2 < b_3 < 2^{s+2}, (b_i, b_j) = 1, 1 \le i < j \le 3.$

We now prove

Lemma 2. The upper density of the integers m_j satisfying (7) and having a divisor of the form (8) is $O(1/s^{1+c})$.

Since $\sum_{s=1}^{\infty} 1/s^{1+c}$ converges, it immediately follows from Lemma 2 that the upper density of the *m*'s satisfying (7) is less than $\varepsilon/2$. Thus to complete the proof of Theorem 1 we only have to prove Lemma 2. Some of the elementary computations needed in this proof, we will not carry out in full detail.

Clearly every m_i satisfying (7) can be written in the form

(9)
$$b_1b_2b_3t_1t_2, P(t_1) < 2^{s+2}, p(t_2) > 2^{s+2}$$

where the b's satisfy (8) and

(10)
$$V(b_1b_2b_3t_1) < (1+2\delta)\log s,$$

Some Extremal Problems in Combinatorial Number Theory

It easily follows from the sieve of Eratosthenes and the well-known theorems of Mertens that the upper density of the integers of the form (9) and (10) is at most

(11)
$$\sum' \frac{1}{b_1 b_2 b_3 t_1} \prod_{p < 2^{\frac{1}{p}+2}} \left(1 - \frac{1}{p}\right) < \frac{c}{s} \sum' \frac{1}{b_1 b_2 b_3 t_1}$$

where the prime indicates that $b_1b_2b_3$ satisfies (8) and b_1, b_2, b_3t_1 satisfies (10).

Thus to complete the proof of Lemma 2 we only have to prove that for a sufficiently small c

(12)
$$\sum' \frac{1}{b_1 b_2 b_3 t_1} = O\left(\frac{1}{s^c}\right).$$

Now clearly (in $\sum_{r'} V(t) = r$, $t < 2^{s+2}$)

(13)
$$\sum_{r} \frac{1}{t} < \left(\sum_{p^{\alpha} < 2^{\sigma+2}} \frac{1}{p^{\alpha}}\right) / r! < (\log s + c_1)^r / r!.$$

A well-known theorem of Hardy and Ramanujan states [13] that

(14)
$$\Pi_r(x) < c_2 x \frac{(\log \log x + c_3)^{r-1}}{(r-1)! \log x},$$

where $\Pi_r(x)$ denotes the number of integers t < x satisfying V(t) = r.

From (14) we obtain (in $\sum_{r} V(b) = r$, $2^{s} < b < 2^{s+2}$)

(15)
$$\sum_{r}^{\prime} \frac{1}{b} < \prod_{r} (2^{s+2})/2^{s} < c_{4} \frac{(\log s + c_{3})^{r-1}}{(r-1)!s}.$$

From (13) and (15) we obtain

(16)
$$\sum_{i}^{\prime} \frac{1}{b_{1}b_{2}b_{3}t^{\prime}} \leq \sum_{1} \left(\sum_{r_{1}}^{\prime} \frac{1}{b_{1}} \sum_{r_{2}}^{\prime} \frac{1}{b_{2}} \sum_{r_{3}}^{\prime} \frac{1}{b_{3}} \sum_{r_{4}} \frac{1}{t_{1}} \right)$$

where in $\sum_{1} r_1 + r_2 + r_3 + r_4 < (1 + 2\delta)\log s$. Using (13) and (15) we obtain by a simple calculation that the terms of the inner sum on the right side of (16) are maximal if

(17)
$$r_i = (1 + o(1))(\frac{1}{4} + \delta/2)\log s, \quad i = 1, 2, 3, 4.$$

From (13), (14), (16), and (17) we easily obtain by a simple computation $(r_i = (1 + o(1))(\frac{1}{4} + \delta/2)\log s, \eta = \eta(\delta)$ tends to 0 as $\delta \to 0$) (18)

$$\sum' \frac{1}{b_1 b_2 b_3 t_1} < c_5 \frac{(\log s)^{c_6}}{s^3} \prod_{i=1}^4 \frac{(\log s)^{r_i}}{r_i!} < c_5 \frac{(\log s)^{c_6}}{s^3} \prod_{i=1}^4 \frac{(\log s)^{r_i} e^{r_i}}{r_i^{r_i}} < \frac{s^{\eta}}{s^2} 4^{(1+O(1))(1+2\delta)\log s} < \frac{1}{s^c}$$

Paul Erdős

for sufficiently large s if δ and $\eta = \eta(\delta)$ are sufficiently small. Relation (18) proves (12) and thus the proof of Theorem 1 is complete.

It is easy to see that by our method we can construct a sequence A of positive upper density so that there are no four integers $a_i \in A$ which have pairwise the same least common multiple. On the other hand, it is easy to see that if $x > x_0(c, k)$ then

(19)
$$\sum_{a_i < x} \frac{1}{a_i} > c \log x$$

implies that there are k a's which have pairwise the same least common multiple. In fact (19) implies that for $x > x_0(c, k)$ there is a t such that

$$(20) t = a_i p$$

has at least k solutions. To see this observe that if (20) had fewer than k solutions we would have

$$k\sum_{i\sum_{a_ic\log x\log\log x,$$

an evident contradiction.

I do not know how much (19) can be weakened so that there should always be k a's every two of which have the same least common multiple. This question seems connected with the following combinatorial problem: Let \mathscr{S} be a set of *n* elements, $A_i \subset \mathscr{S}$, $1 \leq i \leq m(n, k)$. What is the smallest value of m(n, k) for which we can be sure that there are k A's which have pairwise the same union? An asymptotic formula for m(n, k) would also be of some interest.

Before concluding this section I would like to say a few words about equation (20). Assume first that our sequence is such that (19) has only one solution for every t, in other words the integers a_i/p_j , $p_j|a_i$ are distinct for all i and j. It is not difficult to prove that in this case

(21)
$$\max A(x) = \frac{x}{\exp((c + o(1))(\log x \log \log x)^{1/2})}.$$

The proof of (21) uses methods similar to those in [9] and will not be discussed here.

By the methods used in proving Theorem 1, it is not difficult to prove that there is a sequence A of positive upper density such that (20) has for every t at most two solutions.

It would be of some interest to obtain best possible (or at least good) inequalities on $\sum_{a_i < x} 1/a_i$ which ensure that (20) has at least k solutions for some t.

2.

Let $a_1 < \cdots$ be a sequence of integers no one of which divides any other. I proved [10] that there exists an absolute constant c such that

(22)
$$\sum_{i} \frac{1}{a_i \log a_i} < c$$

and Behrend [3] proved that

(23)
$$\sum_{a_i < x} \frac{1}{a_i} < \frac{c \log x}{(\log \log x)^{1/2}}.$$

Alexander [1] and later Sárkőzi, Szemerédi, and I strengthened [22] in the following sense: There is an absolute constant c_1 such that if $a_1 < \cdots$ is any sequence such that

is unsolvable, then

(25)
$$\sum_{i} \frac{1}{a_i \log a_i} < c_i.$$

Inequality (25) easily implies that if a sequence of integers satisfies (24) then it also satisfies

(26)
$$\sum_{a_i < x} \frac{1}{a_i} = o(\log x).$$

Now we show that (26) is best possible. In other words if $f(x) \to \infty$ as slowly as we wish there always exists an infinite sequence satisfying (24) such that for infinitely many x

(27)
$$\sum_{a_i < x} \frac{1}{a_i} > \frac{\log x}{f(x)}.$$

Equation (27) is indeed very easy to see. Let $x_1 < x_2 < \cdots$ tend to infinity sufficiently fast. Let our sequence A consist of the integers in $(x_i^{1/2}, x_i)$ which have no prime factor less than x_{i-1} but have a prime factor greater than $x_i^{1/2}$. A simple argument shows that our sequence satisfies (24), and if $x_i \to \infty$ sufficiently fast then it also satisfies (27) for $x = x_i$.

We can now ask, if $a_1 < \cdots < a_k < x$ satisfy (24) what is the value of

$$\max_{A} \frac{1}{\log x} \sum_{a_i < x} \frac{1}{a_i},$$

where the maximum is taken over all such sequences? The maximum is clearly less than 1.

Paul Erdös

It is well known that the upper density of any sequence of integers no one of which divides any other is less than $\frac{1}{2}$ and any number $\alpha < \frac{1}{2}$ can be the upper density of such a sequence [4]. Similarly one can show that the upper density of any sequence satisfying (24) has upper density less than 1 and any $\beta < 1$ can be the upper density of such a sequence.

It is well known and is easy to see [10] that if $a_1 < \cdots < a_k \leq x$ is such that no *a* divides any other then

$$\max A(x) = \left[\frac{x+1}{2}\right].$$

Now let $a_1 < \cdots < a_l \le x$ be a sequence which satisfies (24). We outline the proof that

(28)
$$\max A(x) = x - \frac{x}{\exp((\log x)^{1/2 + o(1)})}.$$

We can in fact easily write down the sequence $A = \{a_1 < \cdots < a_i \le x\}$ which maximizes $l, a_i \in A$ if and only if $a_i = p_1, p_2, \ldots, p_j, p_1 \le \cdots \le p_j$ and $p_1 \cdots p_j \le x < p_1 \cdots p_j p_{j+1}$ where p_{j+1} is the least prime greater than p_j . Our sequence clearly satisfies (24). To show that it maximizes l, let $A' = \{a_1' < \cdots < a_{l'} \le x\}$ be a sequence of integers satisfying (24). It suffices to show that if A' contains r integers not contained in A then A contains at least r integers not in A'. To see this let $u_j < \cdots < u_s$ be the integers not in A. There clearly is a $p^{(i)} > p(u_i)$ so that $u_i p^{(i)} \in A$. Now these integers must be all distinct. To see this observe that

(29)
$$u_i p_1 \neq u_j p_2$$
 where $p_1 > p(u_i), p_2 > p(u_j).$

To prove (29) observe that we can assume $p_1 \neq p_2$. Thus without loss of generality we have $p_2 > p_1$. But then if (29) did not hold we would have $p_2 | u_i$ which contradicts $p_2 > p_1 > p(u_i)$.

Now it is easy to prove (28). On the one hand consider all the integers n satisfying

(30)
$$n < \frac{x}{\exp(2(\log x)^{1/2})}, \quad p(n) < \exp((\log x)^{1/2}).$$

It is easy to see that none of the integers (30) belong to A and a simple computation gives that their number is greater than $x/\exp((\log x)^{\frac{1}{2}+\epsilon})$ for every $\epsilon > 0$ if $x > x_0(\epsilon)$.

To prove the opposite inequality split the integers not in A into two classes. In the first class are the integers n with $P(n) < \exp((\log x)^{1/2})$. By the results of de Bruijn [5] and others the number of these integers not exceeding x is less than $x/\exp((\log x)^{1/2+o(1)})$. If n is in the second class we have $P(n) \ge \exp((\log x)^{1/2})$. But then since n is not in A we must have nP(n) < 2x, or $n < 2x/\exp((\log x)^{1/2})$, which completes the proof of (28).

Some Extremal Problems in Combinatorial Number Theory

3.

In this section we investigate some properties of the divisors of n. Let $1 = u_1 < \cdots < u_{d(n)} = n$ be the net of all divisors of n. Denote by A_t the set of those n for which t can be represented as the distinct sum of divisors of n. Clearly if n is in A_t then any multiple of n is also in A_t , and it is easy to see that every integer in A_t is a multiple of an integer in A_t not exceeding t!. Thus it easily follows that A_t has a density d_t . It is a little less easy to see that $d_t \to 0$ as $t \to \infty$. To see this we split the integers of A_t into two classes. In the first class are the integers which have a divisor in $(t/(\log t)^2, t)$. I proved [11] that the density of these integers tends to 0 as $t \to \infty$ (in fact the density is $O(1/(\log t)^{c_1})$. The integers of the second class have no divisor in

 $(t/(\log t)^2, t),$

Thus if t is the sum of divisors of n we must have $(d_t(n)$ denotes the number of divisors of n not exceeding t)

$$(31) d_t(n) > (\log t)^2$$

But clearly

(32)
$$\sum_{n=1}^{x} d_{t}(n) \leq \sum_{u=1}^{t} \frac{x}{u} < 2x \log t.$$

From (32) we obtain that the number of integers $n \le x$ satisfying (31) is less than $2x/\log t$, or the density of integers of the second class is not greater than $2/\log t$. Hence $d_t \to 0$ (and in fact $d_t < 1/(\log t)^{c_1}$ for $t > t_0$. We can prove that for $t > t_0$, $d_t > 1/(\log t)^{c_2}$. Perhaps

(33)
$$d_t = (1 + o(1))c_3/(\log t)^{c_4},$$

but (33) if true may not be quite easy to prove.

An integer n is said to have property P if all the $2^{d(n)}$ distinct sums formed from its d(n) divisors are distinct. One's first guess might be that the integers having property P have density 0. But we prove

Theorem 2. The density of integers having property P exists and is positive.

The proof will be similar to [12]. Clearly if m does not have property P then all the multiples of m also do not have property P. Let $m_1 < m_2 < \cdots$ be the sequence of integers which do not have property P but every divisor of them has property P. $(m_1 = 6.) n$ has property P if and only if it is not divisible by any of the m's. Thus to prove Theorem 2 we have to show that the density of the integers not divisible by any of the m's exists and is less than 1.

If we could prove that

$$(34) \qquad \qquad \sum_{i=1}^{\infty} \frac{1}{m_i} < \infty$$

then as in [12] it would follow that the density of integers having property P exists and is greater than 0. Inequality (34) is quite possibly true but I cannot prove it. Thus we have to argue in a more roundabout way. We split the *m*'s into two classes. In the first class are the $m_i^{(1)}$'s satisfying

(35)
$$V(m_i^{(1)}) > (1 + \varepsilon) \log \log m_i^{(1)}$$
.

The $m_i^{(2)}$'s of the second class do not satisfy (35).

Now we prove (see [3])

(36)
$$d(m_1^{(1)}, m_2^{(1)}, \ldots) = \alpha < 1$$

and

(37)
$$d(m_1^{(2)}, m_2^{(2)}, \ldots) = \beta < 1.$$

Using (4) (as in Section 1) we obtain from (36) and (37) that $d(m_1, m_2, ...)$ exists and satisfies

(38)
$$1 - d(m_1, m_2, \ldots) \ge (1 - \alpha)(1 - \beta) > 0.$$

In other words the density of integers having property P exists and is positive.

Thus to prove Theorem 2 we only have to prove (36) and (37). Expression (36) indeed follows from my result in [8] as in Section 1. Expression (37) will follow as in [12] from

(39)
$$\sum_{i} \frac{1}{m_i''} < \infty.$$

To prove (39) it will suffice to show that

(40)
$$\sum_{m_i^{\prime\prime} < x} 1 = O\left(\frac{x}{(\log x)^2}\right).$$

To prove (40) we split the $m_i'' < x$ again into two classes. In the first class are the m_i'' satisfying

(41)
$$P(m_i'') < \exp(\log x/(\log \log x)^2).$$

It is well known [12] that the number of integers $m_i'' \leq x$ satisfying (41) is $O(x/(\log x)^2)$.

Thus henceforth it suffices to consider the integers of the second class (not satisfying (41)). Consider the integers $m_i''/P(m_i'')$. They are all less than $x(\exp(\log x/(\log \log x)))^{-1} = x/L$.

Now we prove that for every t < x/L the number of solutions of

(42)
$$m_i''/P(m_i'') = t$$

is less than $\exp(\log x/2(\log \log x)^2) = L_1$.

Suppose we already proved that (42) has fewer than L_1 solutions; then we evidently have (in \sum' , m_i'' belong to the second class, i.e., they do not satisfy (41))

(43)
$$\sum_{m_1 = x}^{\infty} 1 < \frac{xL_1}{L} = O(x/(\log x)^2).$$

Expression (43) completes the proof of (40) and hence of Theorem 2.

Let

(44)
$$m_{i_r}"/P(m_{i_r}") = t, \quad r = 1, ..., s$$

be the set of all solutions of (42). Put $P(m_{i_r}") = p_r$, r = 1, ..., s. These s primes are clearly all distinct. By our assumptions t has property P but the integers

$$m_{ir}'' = tp_r, \qquad r = 1, \ldots, s$$

do not have property P. Hence for every r there are divisors $d_j^{(r)}$ of tp_r satisfying

(45)
$$\sum_{j} E_{j} d_{j}^{(r)} = 0, E_{j} = \pm 1,$$

and in the sum (45) at least one $d_j^{(r)}$ must be a multiple of p_r (for otherwise all the $d_j^{(r)}$ would be divisors of t and t would not have property P). Thus for every p_r there is a sum (different from 0) satisfying

(46)
$$\sum_{u} E_{u} d_{u}^{(r)} \equiv 0 \pmod{p_{r}}, \quad d_{u}^{(r)} \mid t, \quad E_{u} = \pm 1.$$

Now since m_i does not satisfy (35) we have $V(t) < (1 + \varepsilon)\log \log x$. Hence the number of sums (46) is less than

(47)
$$3^{d(r)} < 3^{2^{(1+\epsilon)\log\log x}} < \exp((\log x)^{1-c}).$$

Each of the sums (46) has fewer than $\log x$ prime divisors, thus from (46) and (44) we have

$$s < \log x \exp((\log x)^{1-c}) < L_1$$

which completes the proof of Theorem 2.

By the same method we can prove that the density of integers n for which n is the sum of distinct proper divisors of n exists and is between 0 and 1. Several other related results can be proved by this method.

REFERENCES

- 1. R. Alexander, Density and multiplicative structure of sets of integers, Acta Arith. 12 (1967), 321-332.
- F. Behrend, Generalization of an inequality of Heilbronn and Rohrbach, Bull. Amer. Math. Soc. 54 (1948), 681-684.
- 3. ——, On sequences of numbers not divisible one by another, J. London Math. Soc. 10 (1935), 42-44.
- 4. A. S. Besicovitch, On the density of certain sequences, Math. Ann. 110 (1934), 336-341.
- 5. N. G. de Bruijn, On the number of positive integers $\leq x$ and free of prime factors $\geq y$, Nederl. Akad. Wetensch. Ind. Math. 13 (1951), 50-60.
- P. Erdős, On a problem in elementary number theory and a combinatorial problem, Math. Comp. 18 (1964), 644-646.
- , On some applications of probability to analysis and number theory, J. London Math. Soc. 39 (1964), 692-696.
- 8. ____, On the distribution function of additive functions, Ann. of Math. 47 (1946), 1-20.
- 9. ____, On primitive abundant numbers, J. London Math. Soc. 10 (1935), 49-58.
- Mote on sequences of integers no one of which is divisible by any other, J. London Math. Soc. 10 (1935), 126-128.
- 11. _____, A generalization of a theorem of Besicovitch, J. London Math. Soc. 11 (1936), 92-98.
- 12. -----, On the density of abundant numbers, J. London Math. Soc. 9 (1934), 278-282.
- G. H. Hardy, P. V. Seshu Aiyar, and B. M. Wilson, Eds., Collected Papers of Srinivasa Ramanujan, Cambridge University Press, Cambridge, England, 1927, pp. 262–275.

(Received March 20, 1968)