SOME PROBLEMS AND RESULTS ON ADDITIVE AND MULTIPLICATIVE NUMBER THEORY

P. Erdős

To my old friend Emil Groswald, in friendship and admiration.

In this note I discuss some problems of a somewhat unconventional nature which recently occupied me and my collaborators. I will deal with divisors and prime factors of integers, some additive problems of a combinatorial nature and on differences of consecutive primes, squarefree numbers and more general sequences defined by divisibility properties.

1. Let $1 = d_1 < d_2 < \dots < d_{\tau(n)} = n$ be the sequence of consecutive divisors of n. Put

$$h_\alpha(n) = \sum_{i=1}^{\tau(n)-1} \left(\frac{d_{i+1}}{d_i} - 1\right)^\alpha. \tag{1.1}$$

Is it true that for every $\alpha > 1$ there is a constant C_α and infinitely many integers n for which $h_\alpha(n) < C_\alpha$? This question occurred to me a few weeks ago but I was unable to make any progress. In fact I could not prove the existence of C_α for any α. $n!$ or the least common multiple of the integers not exceeding n seem to be good candidates for integers with (1.1) bounded above.

I came to (1.1) by considering the sum $\sum_{i=1}^{\tau(n)-1} d_{i+1}/d_i$.

It is easy to see that

$$\sum_{i=1}^{\tau(n)-1} d_{i+1}/d_i > \tau(n) + \log n$$

and I asked myself the question whether it is true that

$$\liminf_{n \to \infty} \left(\sum d_{i+1}/d_i - \tau(n) - \log n\right) < \infty. \tag{1.2}$$

(1.2) would follow if (1.1) is bounded for an infinite set of n.

Srinivasan calls a number n practical if every $m \le n$ is the sum of distinct divisors of n. It is well known and easy to see that the density of practical numbers is 0. Let $S(n)$ be the smallest integer so that every $1 \le m \le n$ is the sum of $S(n)$ or fewer distinct divisors of n ($S(n) = 0$ if n is not practical). In connection with problems on representation of the form

$$\frac{a}{b} = \frac{1}{x_1} + \ldots + \frac{1}{x_k}, \quad 1 \le a < b, \ k \text{ minimal},$$

I needed integers n for which $S(n)$ is small. I easily observed

$$S(n!) < n \text{ or } S(m) < \frac{\log m}{\log \log m}$$

for infinitely many m. I conjectured that for infinitely many n

$$(1.3) \qquad S(n) < (\log \log n)^C$$

but I could make no progress with (1.3), which is unsolved for more than 30 years. I offer 250 dollars for a proof or disproof of (1.3). In itself (1.3) is perhaps somewhat artificial and isolated but a proof or disproof of (1.3) might throw some light on more important problems.

I just notice that the investigation of $\max_{n \le x} S(n)$ might lead to nontrivial questions. At first I thought that $S(n) < c\log n$ holds for all n but this is easily seen to be false. Let m_k be the product of the first k primes and let q_k be the greatest prime less than $\sigma(m_k)$. It is easy to see that $n_k = q_k m_k$ is practical but $q_k - 1$ needs for its representation $n_k^{c/\log \log n_k}$ divisors of n_k. Perhaps one could try to obtain an asymptotic formula for $\sum_{n=1}^{x} S(n)$.

My most interesting unsolved problem on divisors states that almost all integers have two consecutive divisors

$$(1.4) \qquad d_{i+1} < 2d_i$$

or in a sharper form: For almost all n, (and every $\varepsilon > 0$)

$$(1.5) \qquad \min_i d_{i+1}/d_i < 1 + c^{-\log \log n(\log 3 - 1 - \varepsilon)}$$

R.R. Hall and I proved that the exponent in (1.5) if true is best possible.

Denote by $\tau^+(n)$ the number of integers k for which n has a divisor in $(2^k,2^{k+1})$. I conjectured that for almost all n $\tau^+(n)/\tau(n) \to 0$, which of course would have implied (1.4). Tenenbaum and I recently disproved this, and we also proved a recent conjecture of Montgomery which stated that if $\tau^{(d)}(n)$ denotes the number of indices i for which $d_i | d_{i+1}$ then $\tau^{(d)}(n)/\tau(n) > \varepsilon$ holds for a sequence of positive density. Very likely $\tau^{(d)}(n)/\tau(n)$ has a distribution function, but this question we have not yet settled.

Denote by $\tau_r(n)$ the number of indices i for which $(d_i,d_{i+1}) = 1$. R.R. Hall and I studied $\tau_r(n)$ and we obtained various asymptotic inequalities for it, but we are very far from settling all the interesting questions which can be posed here. One of our questions stated: Let n be squarefree and $v(n) = k$ ($v(n)$ denotes the number of distinct prime factors of n). How large is $\max_{v(n)=k} \tau_r(n)$? Simonovits and I proved

$$(1.6) \qquad (2^{1/2}+o(1))^k < \max_{v(n)=k} \tau_r(n) < (2-c)^k.$$

We proved (1.6) by the following lemma: Let $0 < x_1 < \ldots < x_k$, assume that the 2^k sums $\sum_{i=1}^{k} \varepsilon_i x_i$, $\varepsilon_i = 0$ or 1, are all distinct and order the sums $\sum_{i=1}^{k} \varepsilon_i x_i$ by size. Denote by $g(k)$ the maximum number of consecutive sums $\sum_{i=1}^{k} \varepsilon_i x_i$, $\sum_{i=1}^{k} \varepsilon_i' x_i$, $\varepsilon_i \varepsilon_i' = 0$, for every $1 \le i \le k$. Clearly $g(k) = \max_{v(n)=k} \tau_r(n)$. Simonovits and I proved that $g(k)$ satisfies (1.6).Perhaps $g(k)$ can be determined explicitly.

Let $p_1^{(n)} < \ldots < p_{v(n)}^{(n)}$ be the sequence of consecutive prime factors of n. Our knowledge of the properties of the prime factors of almost all integers is much more satisfactory than our knowledge of the divisors of n. Here I state only one result which can easily be obtained by the methods of probabilistic number theory: Put

$$\varepsilon_r = \frac{\log \log p_r^{(n)} - r}{r^{1/2}}.$$

The sequence

$$\frac{1}{\log \log \log n} \sum_{\varepsilon_r > c} \frac{1}{r} \; ; \; r = 1,2,\dots,v(n),$$

has Gaussian distribution; $\varepsilon_r > 0$ does not have a distribution function. Also roughly speaking for almost all n

$$r^{1/2}\varepsilon_r = \log \log p_r^{(n)} - r$$

is dense in $(-C\, r^{1/2}, C\, r^{1/2})$. Here is a more exact special case. An old theorem of mine states that the r-th prime factor of n is for almost all n between exp exp$(r(1-\varepsilon))$ and exp exp$(r(1+\varepsilon))$. How close can in fact $p_r^{(n)}$ come to exp exp r for almost all n? It is easy to see that for almost all n the number of solutions of

$$|\log \log p_r^{(n)} - r| < \frac{1}{f(r) r^{1/2}}$$

tends to infinity if and only if $\sum_r \frac{1}{r f(r)} = \infty$. The proof is an easy consequence of sieve methods and elementary independence arguments.

It seems impossible to obtain similarly sharp estimates for the divisors of n; in fact such results are almost certainly not true mainly due to the lack of independence.

REFERENCES

1. P. Erdös and R.R. Hall, The propinquity of divisors, Bull. London Math. Soc. 11 (1979), 304-307.
2. P. Erdös and R.R. Hall, On some unconventional problems on the divisors of integers, J. Austral. Math. Soc., (Series A) 25 (1978), 479-485. See also P. Erdös, Some unconventional problems in number theory, Math. Magazine, 52 (1979), 67-70.

2. Let $1 \le a_1 < \dots < a_k \le n$. Assume that the sums $a_i + a_j$ are all distinct. Denote $g(n) = \max k$. Turán and I conjectured

(2.1) $$g(n) = n^{1/2} + O(1),$$

but we are very far from being able to prove (2.1). The sharpest result known about g(n) states:

(2.2) $$n^{1/2}-n^{\frac{1}{2}-c} < g(n) < n^{1/2}+n^{1/4}+1.$$

Our original proof of the upper bound gives without much difficulty the following slightly sharper theorem: Let $1 \leq a_1 < \ldots < a_k \leq n$, $k = [(1+c)n^{1/2}]$. Then the number of distinct differences of the form $a_i - a_j$, $a_i > a_j$ is less than $(1-\varepsilon_c)\binom{k}{2}$. I do not know the best possible value of ε_c and probably the determination of the best possible value of ε_c will not be easy. This problem is perhaps of some interest but I have not investigated it carefully. A problem of Graham and Sloane in graph theory led me to conjecture that if $k > (1+c)n^{1/2}$, then the number of distinct sums a_i+a_j is also less than $(1-\varepsilon_c')\binom{k}{2}$. Unfortunately I noticed a few days ago that my conjecture is completely wrongheaded. To see this we define $k = [(1+o(1))\frac{2}{3^{1/2}}n^{1/2}]$ a's not exceeding n so that if $a_i+a_j = a_r+a_s$ then $a_i+a_j = n$. Let $1 \leq a_1 < \ldots < a_\ell \leq \frac{n}{3}$ be a maximal sequence for which all the sums a_i+a_j, $1 \leq i \leq j \leq \ell$ are distinct. By our result with Turán we have $\ell = [(1+o(1))\ (\frac{n}{3})^{1/2}]$. Now put $a_{\ell+i} = n-a_{\ell-i+1}$. Our sequence has $(1+o(1))2(\frac{n}{3})^{1/2}$ terms and it is easy to see that all the sums a_i+a_j are distinct unless $a_i+a_j = n$.

The problem now remains: What is the largest value of c for which there is a sequence $1 \leq a_1 < \ldots < a_k \leq n$, $k = (1+o(1))cn^{1/2}$, so that the number of distinct sums a_i+a_j is $(1+o(1))\binom{k}{2}$? Trivially $c \leq 2$ and it is not hard to show that $c < 2$. Perhaps $c < 2^{1/2}$ but at the moment I do not see how to show this. For the problem of Graham and Sloane it was more natural to assume that the number of distinct sums mod n should be $(1+o(1))\binom{k}{2}$. Here of course trivially $k < (2n)^{1/2}$ and probably $k < (1-c)(2n)^{1/2}$ but I have not yet been able to settle this problem.

REFERENCES

1. P. Erdös and P. Turán, On a problem of Sidon in additive number theory and on some related problems, J. London Math. Soc. 16 (1941), 212-216. For the sharpest result, see B. Lindström, An inequality for B_2-sequences, Journal

Comb. Theory, 6 (1969), 211-212. See also H. Halberstam and K.F. Roth, Sequences, Vol. I, Oxford Univ. Press. Oxford 1966.

3. Let $1 = q_1 < q_2 < \dots$ be the sequence of squarefree numbers. Many mathematicians investigated them from various points of view. Denote by $Q(x)$ the number of squarefree numbers not exceeding x. It is easy to see that $Q(x) = \frac{6}{\pi^2}x + O(x^{1/2})$; the prime number theorem gives $Q(x) - \frac{6}{\pi^2}x = o(x^{1/2})$. It is known that the error term cannot be $o(x^{1/4})$ and it was known for a long time that the Riemann hypothesis implies that $Q(x) - \frac{6}{\pi^2}x = o(x^{2/5})$ and this has been recently improved to $o(x^{1/3})$. We will not deal with these problems here.

The difference $q_{k+1}-q_k$ has been investigated a great deal. No doubt $q_{k+1}-q_k = o(q_k^{\varepsilon})$ for every $\varepsilon > 0$ if $k > k_0(\varepsilon)$, but we are very far from being able to prove this. The sharpest results are due to Richert, Rankin and Schmidt. They proved it for ε a little less than 2/9. I proved that for every $\alpha \leq 2$

$$\sum_{q_k < x} (q_{k+1}-q_k)^{\alpha} = c_{\alpha}x + o(x). \tag{3.1}$$

Hooley proved that (3.1) holds for every $\alpha \leq 3$. There is no doubt that (3.1) holds for every $\alpha > 0$ but this seems hopeless at present. Put

$$f(x,c) = \sum_{q_k < x} \exp(c(q_{k+1}-q_k)).$$

I expect that

$$f(x,c)/x \to \infty \tag{3.2}$$

for every $c > 0$ but cannot prove it for any c.

The reason for the difficulty of proving (3.2) is that I cannot give a uniform estimation for the density α_t of the indices k for which $q_{k+1}-q_k > t$. It is not difficult to show that $\alpha_t^{1/t} \to 0$, i.e. α_t tends to 0 faster than exponentially, but I have no uniform estimation for α_t and as far as I know there is no such estimation available in the literature; i.e. I have no good estimation for the number of indices k for which $q_{k+1}-q_k > t_n$, $q_k < n$, when t_n tends to infinity together with n.

I observed nearly 30 years ago that for infinitely many k,

$$q_{k+1}-q_k > (1+o(1))\frac{\pi^2}{12}\frac{\log k}{\log\log k}. \tag{3.3}$$

(3.3) follows easily from the Chinese remainder theorem, the prime number theorem and the sieve of Eratosthenes. I never was able to improve (3.3) and cannot exclude the unlikely possibility that (3.3) is best possible. More generally let $u_1 < u_2 < \ldots$ be a sequence of integers satisfying

$$(u_i.u_j) = 1, \quad \sum_i \frac{1}{u_i} < \infty, \tag{3.4}$$

and denote by $a_1 < a_2 < \ldots$ the set of integers not divisible by any of the u's. Put

$$u_1.u_2\ldots u_{t_x} \le x < u_1\ldots u_{t_x}u_{t_x+1}.$$

Analogously to (3.3) we obtain

$$\max_{a_k < x}(a_{k+1}-a_k) > (1+o(1))t_x\pi(1-\frac{1}{u_i})^{-1}. \tag{3.3'}$$

We will show that there are sequences satisfying (3.4) for which (3.3') is best possible. I stated this result in a previous paper. In fact we shall prove it in the following slightly stronger form: There is an infinite sequence of primes $p_1 < p_2 < \ldots$, $\sum_i \frac{1}{p_i} < \infty$, so that for all $k > k_0(\varepsilon)$

$$a_{k+1}-a_k < (1+\varepsilon)t_x\prod_i(1-\frac{1}{p_i})^{-1}. \tag{3.5}$$

The proof of (3.5) is indeed easy. Let $p_1 < p_2 < \ldots$ be an infinite sequence of primes which tend to infinity sufficiently fast. Let

$$p_k < x < x+L < p_{k+1},\ L = (1+\varepsilon)t_x\prod_i(1-\frac{1}{p_i})^{-1}. \tag{3.6}$$

To prove (3.5) we only have to show that there is at least one integer T,

$x < T < x+L$, which is not a multiple of any of the primes $p_1, \ldots, p_k$. If the p's increase sufficiently fast then $t_x = k-1$ or $t_x = k$. Let r be large but small compared to k. Then the number of integers in $(x, x+L)$ which are not multiples of any of the p's is by the sieve of Eratosthenes at least

$$(3.7) \qquad L \prod_{i=1}^{r} (1-\frac{1}{p_i}) - 2^r - k \sum_{i > r} \frac{1}{p_i} - k > 0,$$

by (3.6) and $t_x \geq k-1$, which completes the proof of (3.5).

The real problem here is: Is there a sequence $u_1 < u_2 < \ldots$, $(u_i, u_j) = 1$, $\sum_i \frac{1}{u_i} < \infty$, which satisfies (3.5) and the u_i do not tend to infinity very fast, say $u_i < i^C$ for some absolute constant C? I do not expect that such a sequence exists. I am fairly sure that there is a sequence satisfying (3.5) for which $u_i^{1/i} \to 1$.

I proved that every irreducible cubic polynomial represents infinitely many squarefree integers and Hooley that the set of integers n for which the cubic polynomial f(n) is squarefree has positive density. It seems hopeless at present to extend this result to quartic polynomials,; in fact there is no quartic polynomial about which we can prove that it represents infinitely many squarefree integers and of course it seems hopeless to prove that $2^n \pm 1$, $2^{2^n} \pm 1$ or $n! \pm 1$ represents infinitely many squarefree numbers. The sharpest results on the representation of power free numbers are due to Nair and to Huxley and Nair.

The analogue of the prime k tuple conjecture is true and was certainly known to L. Mirsky for a long time. It states: let $a_1, \ldots, a_k$ be a set of integers which does not contain a complete set of residues mod p^2 for every p;

then the density of integers n,for which the integers $n+a_i$, $i = 1,\ldots,k$, are all squarefree, is positive. There seems to be no possibility of extending this result for infinite sequences $A = \{a_1 < a_2 < \ldots\}$, where we assume that A does not contain a complete set of residues mod p^2. A is said to have property P if for every integer n, $n+a_i$ is squarefree for only a finite number of indices i. It is easy to see that there are sequences having property P. The simple proof is left to the reader. Probably a sequence having property P must increase fairly fast, but I have no results in this direction.

A is said to have property $\bar{P}$ (respectively $\bar{P}_\infty$) if there are infinitely many n for which $n+a_i$ is squarefree for all (respectively for all but finitely many) $a_i \in A$. A sequence having property $\bar{P}$ or $\bar{P}_\infty$ must no doubt also increase fast.

A is said to have property Q if for infinitely many n, $n+a_i$ is squarefree for all $a_i < n$. It is easy to see that if A increases sufficiently fast then it has property Q and in fact there is an n, $a_k < n < n_{7k+1}$ for which $n+a_i, i = 1,\ldots,k$, is always squarefree. I have no precise information about the rate of increase a sequence having property Q must have.

It would of course be interesting to investigate which special sequences (e.g. $2^n \pm 1$, $n! \pm 1$ etc.) have properties P, $\bar{P}$, $\bar{P}_\infty$ or Q, but as far as I know nothing is known here. These problems can of course be stated for other sequences than p^2, but we formulate only one such question: Is there an infinite sequence $a_1 < a_2 < \ldots$ so that there are infinitely many n for which for all $a_k < n$, $\{n+a_k\}$ always is a prime?

The prime k-tuple conjecture implies the existence of such a sequence. It would be of interest to obtain some estimates about the rate of growth of such a sequence.

It is easy to see that there is an infinite sequence A for which a_i+a_j, $1 \le i \le j$, is always squarefree. In fact one can find such a sequence which grows exponentially. Must such a sequence really increase so fast? I do not expect that there is such a sequence of polynomial growth.

Is there a sequence of integers $1 \le a_1 < a_2 < \ldots$ so that for every i,

$a_i \equiv t \pmod{p^2}$ implies $1 \le t < p^2/2$? If such a sequence exists then clearly a_i+a_j is always squarefree, but I am doubtful if such a sequence exists. I formulated this problem while writing these lines and must ask the indulgence of the reader if it turns out to be trivial.

Let $A(X)$ be the largest integer for which there is a sequence $1 \le a_1 < \dots < a_k \le X$, $k = A(X)$, which does not form a complete set of residues mod p^2 (for every p). Trivially $A(X) = (1+o(1))\,\frac{6}{\pi^2} X$ and Ruzsa pointed it out to me that for infinitely many X $A(X) > O(X)$. Probably this holds for all large X. It would be of some interest to estimate $A(X)$ as accurately as possible. This problem is of course of interest for other sequences than p^2. The sensational results of Hensley and Richards for the sequence of all primes are well known.

One final problem of this type: Let $(u_i,u_j) = 1$ and $a_1 < a_2 < \dots$ an infinite sequence with the property R: for every u_i and $a_k > u_i$ there is an $a_j < u_i$ for which $a_k \equiv a_j \pmod{u_i}$. The set of all integers clearly always has property R, and if the u's are the set of all primes then no other set has property R. It is easy to see that if the u's are sufficiently thin then there are nontrivial sequences with property R. I am not sure if property R leads to interesting and fruitful questions.

Let $u_1 < u_2 < \dots$ be a sequence of integers. I conjectured long ago that if $u_n/n \to \infty$ then $\sum_n u_n/2^n$ is irrational. Recently I proved this if we assume the slightly stronger hypothesis $u_{n+1}-u_n \to \infty$. I know of no example of a sequence $u_1 < \dots$ for which $\limsup (u_{n+1}-u_n) = \infty$, and $\sum_n u_n/2^{u_n}$ is rational. I am sure that such sequences exist and perhaps I overlook an obvious idea. To my surprise and disappointment I could not prove that $\sum_n q_n/2^{q_n}$ is irrational where $q_1 < \dots$ is the sequence of all squarefree numbers. In fact if $q_{i_1} < q_{i_2} < \dots$ is any subsequence of the squarefree numbers then surely $\sum_n q_{i_m}/2^{q_{i_m}}$ is always irrational. Here again I perhaps overlook a trivial point.

In trying unsuccessfully to prove these conjectures I found a result which perhaps is of some interest:

THEOREM. *Let $c > 0$ be a sufficiently small absolute constant. Then for every $x > x_0(c)$ there are integers $y_1 < y_2 < y_3 < y_4 < x$ satisfying*

$$y_2-y_1 = y_4-y_3 = t > c(\log x)^2 \tag{3.8}$$

for which the squarefree numbers in (y_1,y_2) and (y_3,y_4) are congruent by translation by $y_3-y_1 = y_4-y_2$.

Denote by t_x the longest such interval. Unfortunately I have no good upper bound for t_x; surely $t_x = o(x^{\varepsilon})$ and perhaps $t_x < (\log x)^c$. I. Ruzsa pointed out that it is unlikely that one can get a good result without some really new idea since we cannot exclude the existence of large gaps between the y's.

The proof of our Theorem will not be difficult. Denote by $f(n,t)$ the number of integers $m, n < m < n+t$, for which there is a $p > \frac{1}{100}\log x$, satisfying $m \equiv 0 \pmod{p^2}$. Clearly by the prime number theorem and (3.8)

$$\sum_{n=1}^{x} f(n,t) < tx \sum_{p > \frac{\log x}{100}} \frac{1}{p^2} < \frac{200\, t\, x}{\log x \log\log x} < \frac{200\, cx \log x}{\log\log x}. \tag{3.9}$$

Thus from (3.9) there are clearly at least $\frac{x}{2}$ values of n for which

$$f(n,t) < \frac{400\, c \log x}{\log\log x} = L. \tag{3.10}$$

Henceforth we will only consider these (at least) $\frac{x}{2}$ values of n which satisfy (3.10). We now give an upper bound for the number of patterns the integers $m \equiv 0 \pmod{q^2}$ can form in $(n,n+t)$ (q runs through all primes).

The number of these patterns is clearly less than

$$\binom{t}{L} \prod_{p \le \frac{\log x}{100}} p^2 < \left(\frac{t}{L}\right)^L c^L x^{1/10} = o(x^{1/2}) \tag{3.11}$$

for sufficiently small c. To prove (3.11) observe that the factor $\binom{t}{L}$ comes from the primes $p > \frac{\log x}{100}$ and the factor $\prod_{p \le \frac{\log x}{100}} p^2$ from the primes $\le \frac{\log x}{100}$.

Thus by (3.10) there are two intervals of length t, which by (3.11) can be assumed to be disjoint, in which the squarefree numbers are congruent.

It would be easy to get an explicit bound for c, but this is hardly worth the trouble since at the moment there is no reason to assume that the true order of magnitude of t_x is $(\log x)^2$.

REFERENCES

1. N. Nair, Power free values of polynomials I and II, Mathematika 23, (1976), 159-183 and Proc. London Math. Soc. (3) 38 (1979), 353-368.
2. H.M. Huxley and N. Nair, Power free values of polynomials III, Proc. London Math. Soc. (3) 41,(1980), 66-82. These papers have extensive references to the earlier literature.
3. C. Hooley, On the intervals between consecutive terms of sequences, Proc. Symp. Pure Math., vol. 24, Analytic number theory, Amer. Math. Soc., 1973, 129-140. This paper contains extensive references.
4. P. Erdös, On the difference of consecutive terms of sequences defined by divisibility properties, Acta Arithmetica 12,(1966-67), 175-182; E. Szemerédi, same title II, ibid 23 (1973), 359-361.