Proc. of the 4th Pannonian Symp.
on Math. Stat., Bad Tatzmannsdorf, Austria 1983
W. Grossmann, G. Pflug, I. Vincze, W. Wertz, eds

SOME APPLICATIONS OF PROBABILITY METHODS TO NUMBER THEORY

P. ERDŐS

Budapest

I wrote many papers on these subjects. In this note I will mainly state some unsolved problems which have perhaps been somewhat neglected and which perhaps cannot be solved completely by probability methods, but where slightly weaker results can be obtained by these methods. I will mainly discuss my own problems and those of my coworkers not because I consider these more important but because I hope I know more about them than the reader. Also since my memory is still good I will occasionally add some historical remarks.

First a few words about the literature, this list of references will be very incomplete.

P.D.T.A. Elliott, Probabilistic number theory, Springer Verlag, Grundlehren der Math. Wissenschaften, Vol. 239 and 240 (1980). This comprehensive handbook deals with additive and multiplicative arithmetical functions, it contains also many references to the earlier literature and many unsolved problems and interesting historical remarks.

H. Halberstam and K. F. Roth, Sequences, Springer Verlag 1982. The third chapter of this excellent book contains applications of probability methods to additive number theory.

The first survey paper on probability methods in number theory is M. Kac, Probability methods in some problems of analysis and number theory, Bull. Amer. Math. Soc. 55 (1949), 641-665. I feel that this interesting paper deserves careful study even now.

I will give references only if they are not contained in these books.

Before I start my subject I just remark that probabilistic ideas are often useful in making plausible conjectures which cannot be attacked by our methods which are at our disposal at present. The best known such conjecture is due to Cramer: Let $p_1<p_2<\dots$ be the sequence of consecutive primes. Then

$$(1) \qquad \overline{\lim} \frac{p_{n+1}-p_n}{(\log n)^2} = 1.$$

The Riemann hypothesis would only imply $p_{n+1}-p_n<p_n^{\frac{1}{2}+\varepsilon}$ and the currently known best inequality from below due to Rankin states that there is a $c>0$ so that for infinitely many n

$$(2) \qquad p_{n+1}-p_n > \frac{c \log n \ \log\log n \ \log\log\log\log n}{(\log\log\log n)^2}$$

Rankin obtained (2) 45 years ago, no progress has been made since then except that Schönhage and Rankin himself improved the value of c. This prompted me to offer a reward of 10000 dollars for a proof that (2) holds for every c and infinitely many n. I am so sure that (2) holds that for a disproof I offer 25000 dollars. The only reason that I do not

offer 10^6 dollars that is the very unlikely event that I am wrong and (2) does not hold for every c I could not pay my debt.

Let me state two less well known conjectures of myself: Denote by $P(m)$ the greatest prime factor of m. Is it true that for every $n>n_0(\varepsilon)$ $P(n(n+1)) > (\log n)^{2-\varepsilon}$, but for infinitely many n $P(n(n+1)) < (\log n)^{2+\varepsilon}$. Also is it true that every $n > n_0(\varepsilon)$ can be written in the form $n=a+b$ $P(a\cdot b)< < (\log n)^{2+\varepsilon}$? It is easy to see that the result fails if we replace $(\log n)^{2+\varepsilon}$ by $(\log n)^{2-\varepsilon}$. Very much weaker positive results have been proved by analytic methods by Balog and Sárközy. Balog and Sárközy will publish several papers on this and related subjects.

A. Balog and A. Sárközy, On sums of integers having small prime factors, I-II, Studia Sci. Math. Hung., and On sums of sequences of integers, I-III, Acta Arithmetica and Acta Math. Acad. Sci. Hung., to appear.

H. Cramer, On the order of magnitude of the difference between consecutive prime numbers, Acta Arith. 2 (1936), 23-46.

R. A. Rankin, The difference between consecutive prime numbers, J. London Math. Soc. 13 (1938), 242-247.

First a few words about additive and multiplicative number theoretic functions, this chapter will be very short in view of the book of Elliott. I proved many years ago that the density of integers n for which $\varphi(n) > \varphi(n+1)$ is $\frac{1}{2}$ and the same result holds for $\sigma(n)$ and for $d(n)$ the number of divisors of n. This later result was a conjecture of Chowla. I could never prove that the density of integers with

$P(n) > P(n+1)$ is $\frac{1}{2}$ and this conjecture is probably unattackable by methods at our disposal. Very much weaker results have been proved by Pomerance and myself.

On the other hand easy independence arguments will give that the density of integers for which $\varphi(n) > \varphi(n+1)$ and $d(n) > d(n+1)$ is $\frac{1}{4}$ since the two inequalities are asymptotically independent. Put $f(n) = \sum_{p|n} \frac{1}{\log\log p}$. With very little more trouble I can prove that the three inequalities $d(n) > d(n+1)$, $f(n) > f(n+1)$ and $\varphi(n) > \varphi(n+1)$ are asymptotically independent. On the other hand $\varphi(n) > \varphi(n+1)$ and $\sigma(n) < \sigma(n+1)$ are strongly correlated. The density of integers which satisfy both inequalities is strictly between $\frac{1}{4}$ and $\frac{1}{2}$. Finally the density of integers with $d(n) > d(n+1)$, $\omega(n) > \omega(n+1)$ is $\frac{1}{2}$ ($\omega(n)$ denotes the number of distinct prime factors of n). I am not sure if these results are in the literuture but anyone familiar with the methods of probabilistic number theory can easily supply the proofs. I just want to state one of my old problems which does not seem quite hopeless but which so far resisted all attacks. Let $f(n)$ be an additive function and assume that for a pair a, b of real numbers the density of integers n for which $a<f(n)<b$ exists and is positive. Does it then follow that $f(n)$ has a limiting distribution? It is not difficult to prove that our condition implies that the two series

$$\sum_{|f(p)|>1} \frac{1}{p}, \qquad \sum_{|f(p)|<1} \frac{f(p)^2}{p}$$

both converge, but I could not prove that

$$\sum_{|f(p)|<1} \frac{f(p)}{p}$$

also converges. If this would also be proved then by the theorem of Wintner and my conjecture would be settled. Elliott gave a purely probabilistic formulation of my conjecture (see Vol. 2, p. 331 of Elliott's book).

An old conjecture of mine on multiplicative functions stated

Let $f(n)$ be a multiplicative function which only takes the values ± 1. I conjectured that $f(n)$ has a mean value, i.e. that

$$\lim \frac{1}{x} \sum_{n=1}^{x} f(n) \tag{3}$$

exists, and is 0 if and only if $\sum_{f(p)=-1} \frac{1}{p} = \infty$. (3) was first proved by Wirsing and later in a more general form by Halász. Tchudakoff and I conjectured (independently) that for such a multiplicative function

$$\overline{\lim} \left| \sum_{n=1}^{x} f(n) \right| = \infty. \tag{4}$$

In fact I have a more general conjecture: Let $f(n) = \pm 1$ be any number theoretic function (not necessarily multiplicative) then to every c there is a d and m so that

$$\left| \sum_{k=1}^{m} f(kd) \right| > c. \tag{5}$$

(5) was one of my first conjectures and is now more than 50 years old. I offer 500 dollars for a proof or disproof of (5).

Choose $f(p) = +1$ or $f(p) = -1$ with probability $\frac{1}{2}$ and for $n = \prod p_i^{\alpha_i}$ put $f(n) = \prod_{p|n} f(p)^{\alpha_i}$. Wintner asked what can be said about $\sum_{n=1}^{x} f(n)$ for almost all choices of $f(p) = \pm 1$. He proved that for almost all functions $f(n)$ we have

$$\sum_{n=1}^{x} f(n) < x^{\frac{1}{2}+\varepsilon}. \tag{6}$$

I improved (6) and showed that for almost all functions $f(n)$

$$c_1\left(\frac{x}{\log x}\right)^{1/2} < \sum_{n=1}^{x} f(n) < x^{1/2}(\log x)^{c_2}$$

and conjectured that for almost all functions

$$\overline{\lim_{x}} \frac{1}{x^{1/2}} \sum_{n=1}^{x} f(n) = \infty \tag{7}$$

but

$$\lim \frac{\sum_{n=1}^{x} f(n)}{x^{1/2}(\log x)^{\varepsilon}} = 0 \tag{8}$$

I could not even guess the analog of the law of the iterated logarithm. Halász proved (8), but only proved a slightly weaker result than (7). As far as I know nobody has a plausible guess for the true order of magnitude of $\left|\sum_{n=1}^{x} f(n)\right|$.

Many further interesting questions could be asked e.g.: What can be said about the number of 0-s of the partial sums $\sum_{n<x} f(n)$ for the random multiplicative function? By analogy with the Rademacher functions one would expect that the number of zeros is between $x^{\frac{1}{2}-\varepsilon}$ and $x^{\frac{1}{2}+\varepsilon}$,

of course more precise results would be very desirable.

G. Halász, On random multiplication functions, Publ. Math. D'Orsay 1983, Journées Arith. Coll. H. Delange 79-96.

P. Erdös and Carl Pomerance, On the largest prime factors of n and n+1, Aequationes Math. 17 (1978), 311-321.

Now I discuss some problems and results on additive number theory. These problems were first stated by Sidon more than 50 years ago, he was led to these problems by his study of lacunary trigonometric series. Let $a_1 < a_2 < \ldots$ be an infinite sequence of integers, denote by $f(n)$ the number of solutions of $n=a_i+a_j$. Sidon asked me in 1932 when we first met whether there is a sequence $A = \{a_1 < a_2 < \ldots\}$ for which $f(n) > 0$ for all n but for which $f(n)/n^{\varepsilon} \to 0$ for every $\varepsilon>0$ i.e. A is a basis of order 2 but $f(n)$ is small. I at first thought that the problem will not be hard and that it will be easy to construct such an A. I never succeeded in constructing such an A but about 20 years later I proved by probabilistic methods that there is a sequence A for which

$$(9) \qquad c_1 \log n < f(n) < c_2 \log n$$

holds for every n. An outstanding problem here is whether there is a sequence A for which

(10) $f(n) = (1 + o(1))\log n$

I expect that such a sequence A does not exist. As a first step one should prove that for every sequence A

(11) $\overline{\lim}\ |f(n) - \log n| = \infty$

(11) will perhaps not be hard to prove. I offer 500 dollars for a proof or disproof of (10). If there is no A satisfying (10) then one could ask: Put

$$\overline{\lim}\ f(n)/\log n = C_1,\ \underline{\lim}\ f(n)/\log n = C_2$$

Is there an $\varepsilon>0$ so that for every A, $C_1/C_2 > 1+\varepsilon$?

This question just occured to me while I write these lines and I hope it will not turn out to be trivial. Let $g(n)$ be a monotonic function which tends to infinity arbitrarily slowly. It is easy to prove by the probability method that there is a sequence A for which

$$f(n)/g(n)\ \log n \to 1.$$

An old conjecture of Turán and myself states that if

(12) $f(n) > 0$ for all n then $\overline{\lim}\ f(n) = \infty$

I offer 500 dollars for a proof or disproof of (12).

Perhaps $f(n) > 0$ for all $n > n_0$ already implies that there is an absolute constant c so that $f(n) > c \log n$. If this conjecture is true one can again ask: Is it true that in fact $c > c_0$?

Another possible strengthening of my conjecture with Turán would be: Assure $a_k < c k^2$. Is it then true that $\overline{\lim} f(n) = \infty$ and perhaps even $f(n) > c \log n$ for infinitely many n?

Sidon calls an infinite sequence A a B_2 sequence if the integers $a_i + a_j$ are all distinct. When we first met Sidon asked for a B_2 sequence for which a_n increases as slowly as possible. The greedy algorithm easily gives that there is a B_2 sequence for which $a_k < c k^3$ and we both conjectured that in fact there is a B_2 sequence for which $a_k < k^{2+\varepsilon}$. Rényi and I proved by probabilistic methods that there is a sequence A with $a_k < k^{2+\varepsilon}$ and $f(n) < C_\varepsilon$ for all n. For nearly 50 years we could not prove that there is a B_2 sequence for which $a_k = o(k^3)$. A few years ago Ajtai, Komlós and Szemerédi proved by an ingenious combination of combinatorial and probabilistic methods that there is a B_2 sequence for which

$$a_k < c k^3 / \log k. \tag{13}$$

(13) at the moment seems to be the natural boundary of their method.

Let me state an interesting problem in connection with the greedy algorithm. We construct a B_2 sequence as follows:

Assume $a_1<a_2<\ldots<a_{k-1}$ has already been constructed. Then a_k is the smallest integer for which $\{a_1,\ldots,a_k\}$ is a B_2 sequence. It is easy to see that $a_k<c\,k^3$ and Chowla and Mian carried out extensive calculations on the basis of which they suggest that $a_k \sim k^{2+c}$. It is not even known if

(14) $a_k/k^2 \to \infty$ and $a_k/k^3 \to 0$.

I am not an expert on algorithms but find (14) a fascinating conjecture and offer 250 dollars for a proof or disproof.

I proved that if A is a B_2 sequence then

(15) $$\overline{\lim}\ \frac{a_k}{k^2 \log k} > c > 0$$

for a certain $c>0$. Is (15) best possible? I have no information and offer 500 dollars for a proof or disproof. It is best possible in the following much weaker sense. I proved (15) by showing that if $a_k < c_1 k^2 \log k$ for all $k > k_o$, c_1 sufficiently small, then the number of solutions $g(x)$ of $0 < a_j - a_i < x$ satisfies $g(x) > x$. Thus our sequence cannot be a B_2 sequence. It is not hard to show that (15) is best possible in this case i.e. if $a_k = C\,k^2 \log k$, C sufficiently large, then $g(x)<x$. Krickeberg and I proved that there is a B_2 sequence for which

(16) $$\underline{\lim}\ a_k/k^2 < C$$

for some C. The best result in (16) is due to Krickeberg:

$C > \frac{1}{\sqrt{2}}$. The "Truth" is probably C=1. Perhaps if the $\underline{\lim}$ in '16) is finite then the $\overline{\lim}$ in (15) is infinite. As far as I know this question has not yet been investigated.

Straus and I conjectured that if A is any infinite sequence then there always is a sequence B of density 0 so that every integer is of the form A+B (i.e. of the form a_i+b_j). In fact Lorentz proved the following stronger result: Denote $A(x) = \sum_{a_i<x} 1$. Then there always is a sequence B for which A+B is I e.i. the sequence of all integers and

$$B(x) < c \sum_{n=1}^{x} \frac{\log A(n)}{A(n)} \tag{17}$$

The proof of (17) is nonconstructive and Lorentz remarks that it often is close to being best possible. Lorentz in particular deduces from (17) that if A is the sequence of primes then $B(x) < c(\log x)^3$. I proved by probabilistic methods that in fact $B(x) < c(\log x)^2$ is also possible. I think it would be very interesting to decide if for every additive complement of the primes we in fact have

$$\overline{\lim} \frac{B(x)}{\log x} = \infty \tag{18}$$

and if there is an additive complement for which

$$\lim B(x)/(\log x)^2 = 0 \tag{19}$$

I could not even prove that there is no additive complement of the primes with

(20) $\lim B(x)/\log x = 1$

I am certain that no such sequence exists.

I offer 100 dollars for a proof that there is no additive complement satisfying (20) and 1000 dollars for a proof that such a sequence exists. Probability methods probably will not help and at present no methods seem to be available.

Lorentz observed that (17) implies that if A has positive density then there is an additive complement B for which $B(x) \leq c(\log x)^2$. I showed that this is best possible in general. Let Q be the sequence of squarefree numbers. It is easy to see from the chinese remainder theorem that if F is an additive complement of Q then for a certain $c>0$

$$(21) \qquad B(x) > \frac{c \log x}{\log\log x}.$$

It is not difficult to prove by probabilistic methods that there is an additive complement B for which $B(x) < c \log x$. It would be of some interest to determine the exact order of $B(x)$ in this case.

It is easy to show by the probability method that almost all sequences A have an additive complement B for which $B(x) = (1+o(1))\frac{\log x}{\log 2}$. The measure in the space of sequences is the Lebesgue measure i.e. we make correspond to the sequence A the real number $\sum_{i=1}^{\infty} \frac{1}{2^{a_i}}$

Just a few words about problems and results of Nathanson and myself. We proved by probabilistic methods that there is a sequence A for which

$$c_1 \log n < f(n) < c_2 \log n$$

and if $A(n)$ denotes the set of integers $a+b=n$ then $|A(n) \cap A(m)| < 4$ for every n and m, we could not decide if 4 can be replaced by 3 and perhaps even by 2. More generally the probabilistic method gives that there is a basis A of order k for which $A(x) < cn^{1/k}(\log n)^{1/k}$ and for which if $A(n)$ denotes the set of a's which occur in the set of solutions

$$n = \sum_{i=1}^{s} \varepsilon_i a_i, \quad \sum_{i=1}^{s} \varepsilon_i < k, \quad 0 < \varepsilon_i < k, \quad \varepsilon_i \text{ integers}$$

then

$$|M_k(n) \cap M_k(m)| < C_k$$

The exact value of C_k is unknown, as just stated $2<C_2<4$.

We further asked: Is there a basis of order 2 of density 0 for which the equation $a_i+a_j = a_r$ has only a finite number of solutions? The odd numbers show that the condition "density 0" cannot entirely be dropped. More generally we asked: Is there for every k a basis of order k for which all the sums $\sum_{i=1}^{r} \varepsilon_i a_i$, $\varepsilon_i=0$ or 1, are all distinct (except

for a finite number of cases) as long as $r \leq k-1$?

Finally I want to mention some early work of Atkin. Littlewood posed the following problem: Does there exist a sequence of integers $a_1<a_2<\ldots$ for which $|a_k-k^2| < (\log k)^c$ and for which the sequence a_i+a_j has positive density? Atkin proved sometime before 1950 that the answer is positive and that in fact we can choose $c=1$. Atkin uses semi-probabilistic methods. He told me of his results in 1949 during my visit to England. I unfortunately forgot completely about this and made therefore no reference ot Atkin's unpublished results in my papers on the probability method around 1954-55. Later I remembered my conversations with Atkin and we referred to his work in my papers with Rényi. It is of course impossible to tell if my conversations with Atkin had influenced my later papers. I believe I can prove that the best possible value of c in Littlewood's problem is log 2 but I never worked out all details in full.

A.O.L. Atkin, On pseudo-squares. Proc. London Math. Soc., Third series 14 (1965), 22-27.

M. Ajtai, J. Komlós and E. Szemeredi, On dense infinite Sidon sequences, European J. Comb. 2 (1981), 1-11.

See also

M. Ajtai, P. Erdös, T. Komlós and E. Szemerédi, On Turán's theorem for sparse graphs, Combinatorica 1 (1981), 313-317.

J. Komlós, J. Pintz and E. Szemerédi, A lower bound for Heilbronn's problem, J. London Math. Soc. 25 (1982), 13-24.

To end this paper I discuss some problems which I have somewhat neglected. First of all an old problem of mine stated Divide the integers $1 \le k \le 2n$ into two disjoint subsets $a_1 < a_2 < \ldots < a_n$; $b_1 < b_2 < \ldots < b_n$. Put

$$M_k = \sum_{a_i - b_j = k} 1, \quad M = M(n) = \min \max_k M_k$$

where the minimum is to be taken over all possible divisions of $1 \le k \le 2n$ into two sets of size n A and B . I first thought that $M = \frac{n}{2}$ but showed by the probability method that $M < \frac{4}{9} n$. Independently and about simultaneously Selfridge, Motzkin and Ralston showed by aid of an early electronic computer SWAC that for $n=15$ $M=6$ and they observed that this implies that for infinitely many n $M \le 0.4\, n$. I further showed $M \ge 0.25\, n$ and Scherk improved this to $M > (1 - \frac{1}{\sqrt{2}})n$. Finally L. Moser showed that

$$M > (4-15^{1/2})^{1/2} (n-1) > 0.3970\ (n-1).$$

It would perhaps be worthwhile to get the best possible value of M or at least to determine the smallest c for which $M \ge cn + o(n)$.

L. Moser: On the overlap problem of Erdös, Acta Arith. 5 (1959), 117-119.

An old result of Tchebisheff states that the probability that n and m are relatively prime is $\frac{6}{\pi^2}$. One can expect that this will remain true in general for a large class of

number theoretic functions $g(n)$. E. g. R. R. Hall proved this if $g(n) = \omega(n)$, the number of distinct prime factors of n. Probably it will be true for every $[n\alpha]$, $0<\alpha<\infty$, α not an integer and this will probably not be very difficult to prove (if it is not already in the literature). On the other hand it is probably hopeless to prove that for all $\alpha>1$, α not an integer, the density of integers n for which $(n,[\alpha^n])=1$ is $\frac{6}{\pi^2}$. I do not see how to show this for almost all α, but it is doubtful if probability methods will be of any help here.

P. Erdős and G. G. Lorentz, On the probability that n and $g(n)$ are relatively prime, Acta Arith. 5 (1959), 35-44.

R. R. Hall, On the probability that n and $f(n)$ are relatively prime, Acta Arith. 17 (1970), 169-183.

Rényi and I proved the following theorem. Let G be an additively written Abelian group of order n . Let us choose k elements of G at random, consider all the 2^k sums

$$\sum_{i=1}^{k} \varepsilon_i a_i, \quad \varepsilon_i = 0 \text{ or } 1. \tag{22}$$

Rényi and I proved that if

$$k > \frac{\log n + \log\log n + \omega_n}{\log 2} \tag{23}$$

where $\omega_n \to \infty$ arbitrarily slowly then almost surely all elements of G can be represented in the form (22). We further showed that if

$$k > \frac{(2 + o(1))\log n}{\log 2} \tag{24}$$

then almost surely all elements of G have

$$(1 + o(1))\frac{2^k}{n} \tag{25}$$

representations in the form (22). K. Bognár improved (24) and later R. R. Hall and I proved that in (24) $2+o(1)$ can be replaced by $1+o(1)$.

It is curious that these results hold for every Abelian group of order n, in other words the results are independent of the structure of G. Almost certainly (23) is best possible if we insist that the result should hold for all Abelian groups of order n independent of the structure, but as far as I know this question has not yet been completely cleared up. Also the best possible error term in the result of R. R. Hall and myself is very far from being known.

These results have a surprising number theoretic application. Denote by $f(x,k,\ell)$ the number of integers $n<x$ which have a divisor

$$d \equiv \ell \pmod{k}, \quad (k,\ell) = 1,$$

and $F(x;k)$ denotes the number of integers $n<x$ which have a divisor in every residue class $\ell \pmod{k}$ $(\ell,k) = 1$. Clearly $F(x;k) \leq f(x;k,\ell)$. I proved that for $k<2^{(1-\varepsilon)\,\mathrm{loglog}\,x}$

(26) $$F(x;k) = x + o(x)$$

uniformly in k.

The following problem is perhaps of some interest: Let $k(n) = k$ be the smallest integer for which there is an $\ell, (\ell,k) = 1$ for which n has no divisor $d \equiv \ell \pmod{k}$. (26) implies that the normal order of $k(n)$ is $(\log n)^{\log 2}$, but one could try to obtain sharper results and one could try to estimate $\max_{n<x} k(n)$. I hope to return to these and related questions in the future if there is a future for me.

P. ERDÖS, On the distribution of divisors of integers in the residue classes mod d.

P. ERDŐS and A. RÉNYI, Probalilistic methods in group theory, J. Analyse Math. 14 (1965), 127-138.

K. BOGNÁR, On a problem of statistical group theory, Studia Sci. Math. Hungar. 5 (1970), 29-36.

P. ERDŐS and R. R. HALL, Probabilistic methods in group theory, Houston J. of Math. 2 (1976), 173-180,

and Some results in probabilistic group theory, Comment. Math. Helv. 53 (1978), 448-457. This paper contains many further references.

Pál Erdős
Mathematical Institute of the
Hungarian Academy of Science
Reáltanoda utca 13-15.
H-1364 BUDAPEST