# ON ADDITIVE ARITHMETICAL FUNCTIONS AND APPLICATIONS OF PROBABILITY TO NUMBER THEORY

P. Erdös

The main topic of this lecture will be the study of the distribution function of additive arithmetical functions. Before proceeding with the subject I want to call attention to the excellent review of Kac [13] on this subject, were a very clear and detailed discussion of the interplay of probability and number theory is given. In the present review I try to emphasize those results which were not treated in detail by Kac.

A real valued function $f(n)$, $n = 1, 2, \ldots$ is called additive if

$$f(m.n) = f(m) + f(n) \text{ whenever } (m,n) = 1.$$

The function $g(n)$ $n = 1, 2, \ldots$ is called multiplicative if

$$g(m.n) = g(m).g(n) \text{ whenever } (m,n) = 1.$$

Additive and multiplicative functions are thus completely determined if one knows their values for all powers of primes $p^{\alpha}$.

Examples for additive functions are $\log n$, $\nu(n)$ (the number of distinct prime factors of $n$); examples for multiplicative functions are $n$, Euler's function $\varphi(n)$, and $\delta(n)$ (the sum of divisors of $n$).

Let $h(n)$ be any real valued function defined for $n = 1, 2, \ldots$. We say that $h(n)$ has a distribution function $\psi(c)$ if for every $c$ the density of integers satisfying $h(n) < c$ exists. Denoting this density by $\psi(c)$ we further must have $\psi(-\infty) = 1$, $\psi(+\infty) = +1$.

The question whether an arithmetic function has a distribution function received considerable attention. As far as I know the first result in this direction is due to Schoenberg [15] who proved that $\varphi(n)/n$ has a distribution function. A little later almost simultaneously Davenport[2], Behrend and Chowla proved that $\sigma(n)/n$ also has a distribution function. This result was especially interesting since for $c = 2$ we abtain the so called deficient numbers. (If $\sigma(n) < 2n$, $n$ is called deficient, with $\sigma(n) \geqq 2n$, $n$ is called abundant and with $\sigma(n) = 2n$, $n$ is called perfect).

All these papers used the theory of Fourier transforms. I attacked the problem in a somewhat different way. An integer u is called primitive abundant if it is abundant and all its divisors are deficient. Clearly we obtain all abundant numbers by considering the set of all multiples of all the primitive abundant numbers. Thus if we can show that the sum of the reciprocals of the primitive abundant numbers converges, it will follow by a simple argument that the density of the abundant numbers exists. Denote by $N(n)$ the number of primitive abundant numbers not exceeding $n$, then I proved that [4, 5]

$$(1) \qquad \frac{n}{e^{25\sqrt{\log n \operatorname{loglog} n}}} < N(n) < \frac{n}{e^{\frac{1}{8}\sqrt{\log n \operatorname{loglog} n}}}.$$

(1) easily implies that the sum of the reciprocals of the primitive abundant numbers converges and thus the density of the abundant numbers exists.

Unfortunately this method can be applied only in very special cases. Define $n$ to be primitive $\alpha$-abundant if $\sigma(n) \geqq \alpha n$ but for all divisors $d$ of $n$ $\sigma(d) < \alpha d$. It is not hard to prove that for almost all $\alpha$ the sum of the reciprocals of the primitive $\alpha$-abundant numbers converges, but it is also easy to show that every interval contains $c\ \alpha - s$ for which the sum of the reciprocals of the primitive $\alpha$-abundant numbers diverges (these results are unpublished). Further Besicovitch constructed a sequence of integers $a_1 < a_2 < \ldots$ so that the set of their multiples does not have a density. Thus it is clear that the above method does not lead to general results.

In a subsequent paper Schoenberg [14] deduced the existence of the distribution function of additive functions under very general conditions. Since the logarithm of a multiplicative function is additive it usually will suffice to consider additive functions.

Generalising several previous results [15, 6] I proved that the convergence of the two series

$$(2) \qquad \sum_p \frac{f^+(p)}{p}, \quad \sum_p \frac{(f^+(p))^2}{p}$$

where

$$(3) \qquad f^+(p) = \begin{cases} f(p) \text{ if } |f(p)| \leqq 1 \\ 1 \quad \text{if } |f(p)| > 1 \end{cases}$$

is a sufficient condition for the existence of the distribution function. In a subsequent paper Wintner and I [7]) proved that the convergence of (2) is the necessary and sufficient condition for the existence of the distribution function of $f(n)$.

Our principal tool was the following theorem of Kac [8]) and myself: Let $f(n)$ be additive. Put

$$(4) \qquad A_n = \sum_{p<n} \frac{f(p)}{p}, \quad B_n = \sum_{p<n} \frac{f^2(p)}{p}$$

Assume that

(5) $$B_n \to \infty,\ f(p) = 0(1).$$

Denote by $K_n(\omega)$ the number of integers $m, 1 \leq m \leq n$ for which

(5) $$f(m) < A_m + \omega B_m^{\frac{1}{2}}.$$

Then

(6) $$\lim_{n \to \infty} \frac{k_n(\omega)}{n} = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\omega} e^{-u^2/2} du.$$

Our proof of this result uses a combination of probabilistic and number theoretic methods, we use the central limit theorem and Brun's method. As stated in the introduction Kac gives a more detailed discussion of the connections between probability and number theory. As far as I know Kac and Wintner were the first who called attention to the probabi listic nature of these theorems. I also omit here the connection of these results with those of Hardy-Ramanujan and Turan since Kac discusses them in detail. Recently new proofs were given to our theorem with Kac by Delange [1]) and Halberstam which avoid both the central limit theorem and Brun's method but uses instead the method of moments.

A simple argument shows that (6) indeed implies that $B_n \to \infty$ is incompatible with the existence of the distribution function. The fact that the convergence of $\sum_p \frac{f(p)}{p}$ is also necessary for the existence of the distribution function requires some new and complicated arguments.

Using Brun's method and the estimates of Cramer and Berry on the error term in the central limit theorem I [9]) proved the following result:
Let $f(n)$ be additive.

$$B_n \to \infty,\ f(p) \to 0.$$

Then the density of integers for which

$$0 < f(m) < c (\text{mod } 1)$$

equals $c$, and is thus independent of the function $f(n)$.

Let us now assume that the series in (2) converges. It follows from a result of Paul Levy that the necessary and sufficient condition for the continuity of the distribution function is that

(7) $$\sum_{f(p) \neq 0} \frac{1}{p} = \infty$$

I obtained for this result a direct number-theoretical proof. In a subsequent paper I extended this result considerably, in fact I proved the following theorem: [9])

Let $f(n)$ be any additive function which satisfies (7). Then to every $\varepsilon$ there exists a $\delta$ so that if $x$ is sufficiently large and if $a_1 < a_2 < \ldots < a_t \leqq x$ is a sequence of integers satisfying

$$| f(a_i) - f(a_j) | < \delta$$

then $t < \varepsilon x$. A slichtly inaccurate but suggestive formulation of this result is that the distribution function tries to be continuous even if it does not exist.

The above result is a consequence of the following one:

An additive function is said to be finitely distributed if there exist two constants $c_1$ and $c_2$ and a sequence $n_1 < n_2 < \ldots$ tending to infinity so that for every $i$ there exists a sequence

$$a_1^{(i)} < a_2^{(i)} < \ldots < a_{t_i}^{(i)} < n_i$$

satisfying

$$|f(a_r^i) - f(a_s^i)| < c_1,\ t_i > c_2 n_i\ (1 \leqq r, s \leqq t_i).$$

If $f(n)$ has a distribution function it is clearly finitely distributed, also $\log n$ is finitely distributed.

The necessary and sufficient condition that $f(n)$ should be finitely distributed is that there should exist a constant $c$ and an additive function $g(n)$ so that

$$f(n) = c \log n + g(n), \sum_p \frac{(g^+(p))^2}{p} < \infty. \tag{8}$$

The proof that (8) is sufficient is not difficult, but the proof of the necessity is rather involved.

(7) gives the necessary and sufficient condition for the continuity of the distribution function, it would be of interest to obtain necessary and sufficient conditions for the absolute sontinuity of the distribution function. Here I [10]) proved using results of Jessen and Wintner that if $| f(p) | < \frac{1}{p^\alpha}$ for some $\alpha > 0$ and (7) holds, then the distribution function is continuous but purely singular. On the other hand I gave examples of absolutely continuous distribution functions, in fact if

$$f(p^\alpha) = (-1)^{(p-1)/2} \ (\text{loglog}\ p)^{-3/4}$$

the distribution function is an entire function. It does not seem easy to give a necessary and sufficient condition for the absolute continuity of the distribution function. Another difficult problem seems to be to decide which distribution functions can be the distribution functions of additive functions. It is not hard to prove that the distribution function $\psi(x) = 0$ for $-\infty < x < 0$, $\psi(x) = x$ for $0 \leqq x \leqq 1$, $\psi(x) = 1$ for $1 < x < \infty$, can not be the distribution function of an additive function.

Some time ago Chowla conjectured that the density of the integers for which $d(n+1) > d(n)$ is $\frac{1}{2}$. It is not difficult to deduce from the results of Kac and myself[1]) that if $|f(p)| < c$ (7) is satisfied, the density of integers satisfying $f(n+1) > f(n)$ is $\frac{1}{2}$. It would be of interest to find out to what extent the condition $|f(p)| < c$ can be weakened. The additive function $f(n) = \log n$ shows that it can not entirely be omitted. If $f(p) = (\log p)^\alpha$, $\alpha \neq 1$, $\alpha > 0$ or $f(p) = p$ it presumably is true that the density of integers for which $f(n+1) > f(n)$ is $\frac{1}{2}$, but I am unable to prove these conjectures. The conjecture for the case $f(p) = p$ is easily seen to be equivalent with the following conjecture: The density of integers $n$ for which the greatest prime factor of $n+1$ is greater than the greatest prime factor of $n$ is $\frac{1}{2}$. Leveque[14]) proved that if $|f(p)| < c$ and $\sum_p \frac{f^+(p)^2}{p}$ diverges $f(n)$ and $f(n+1)$ are statistically independent. It would be interesting to know to what extent the condition $|f(p)| < c$ can be omitted here, again $f(n) = \log n$ shows that it is not entirely superfluous (the condition $\sum_p \frac{(f^+(p))^2}{p} = \infty$ can not be omitted).

The proof of all these conjectures seems to present great difficulties. The methods used here by Kac, Wintner and myself and some other investigators which combine the central limit theorem with Brun's method works very well if the effect of the very large primes (the primes greater than $n^\varepsilon$) on the value of $f(n)$ is negligible. But if this condition is not satisfied (like e.g. for $f(p) = p$) our method fails completely.

*Some further problems.* I[9]) proved that if $f(n+1) - f(n) \to 0$, then $f(n) = c \log n$. Let us now only assume that $|f(n+1) - f(n)| < c$. I conjecture then that $f(n) = c_1 \log n + g(n)$, $|(gn)| < c_2$.

Is it true that almost all integers $n$ have two divisors $d_1$ and $d_2$ satisfying

(9) $\quad d_1 < d_2 < 2d_1$ (or more generally $d_1 < d_2 < (1+\varepsilon)d_1$)?

I[11]) proved that the density of integers satisfying (9) exists, but I can not prove that it is 1.

Let $g(n)$ be a multiplicative function for which

$$\sum_p \frac{g(p) - 1}{p}$$

and

$$(11) \qquad \sum_p \frac{(g(p)-1)^2}{p}$$

converges. Then

$$(12) \qquad \lim_{n=\infty} \frac{1}{n} \sum_{k=1}^{n} g(k)$$

exists and is different from 0 and $\infty$ [12]). Does the limit (12) still exist if we only assume the convergence of (10)?

It might be remarked that the method used to prove the existence of the distribution function of additive functions often can be used to prove the existence of the distribution function of non additive functions e.g. Shapiro and I proved that (will be published in the Canadian Journal of Math.)

$$\sum_{k=1}^{n} \frac{\varphi(k)}{k} - \frac{6}{\pi^2} n$$

has a continuous and steadily increasing distribution function. Also

$$(14) \qquad f(a) = \sum \frac{1}{k l_a(k)}$$

where $l_a(k)$ is the exponent of $a (\mathrm{mod}\ k)$, has a distribution function.

Many aspects of this subject were not considered in this short note e.g. the interesting results of Van Kampen, Wintner, Kac and de Bruijn on connections between the theory of almost periodic functions and additive and multiplicative functions, perhaps I gave an undue amount of place to my own papers, but I may be excused since not unnaturally I am best acquainted with them. The references to the litterature are not complete and again I refer to the review article by Kac quoted above.

In closing I would like to call attention to a few other applications of probability to number theory. First of all Cramer used probability theory to make plausible conjectures about primes. Further Rényi by using probabilistic methods considerably strengthened the large sieve of Linnik. He himself derived from these results that every integer is the sum of a prime and an integer having a bounded number of prime factors. His results have several other applications. Finally I would like to call attention to the fact that one often can prove the existence of sequences with certain properties by probability methods. In fact one then proves that almost all sequences have the required property, perhaps without being able to construct a single sequence having the property.

FOOTNOTES.

[1]) A. S. BESICOVITCH, On the density of certain sequences of integers, Math. Annalen **110** (1934) 336—341.

[2]) H. DAVENPORT, Über die Numeri Abundantes, Preuss. Akad. Wiss. Sitzungsber. (1933) 830—837.

[3]) H. DELANGE, C. R. Acad. Sci. Paris **237**, (1953) 542—544. Halberstam's paper was presented to the international congress in Amsterdam and has not yet been published in detail. P. Erdos,

[4]) On the density of the abundant numbers, Journal London Math. Soc. **9** (1934) 278—282.

[5]) On primitive abundant numbers, ibid **10** (1935) 49—58.

[6]) On the density of some sequences of numbers, ibid. **10** (1935) 120—125, **12** (1937) 7—11, **13** (1938) 119—127.

[7]) (with A. WINTNER) Additive arithmetical functions and statistical independence, Amer. Journal of Math. **61** (1939) 713—722.

[8]) (with M. KAC) The Gaussian law of errors in the theory of additive arithmetical functions, ibid. **62** (1940) 738—742.

[9]) On the distribution function of additive functions, Annals of Math. **47** (1946) 1—20

[10]) On the smoothness of the asymptotic distribution function of additive arithmetical functions, Amer. Journal of Math. **61** (1939) 722—725.

[11]) On the density of some sequences of integers, Bull. Amer. Math. Soc. **54** (1948) 685—692.

[12]) Some asymptotic formulas for multiplicative functions, ibid. **53** (1947) 536—544.

[13]) M. KAC, Probability methods in some problems of analysis and number theory, Bull. Amer. Math. Soc. **55** (1949) 641—665.

[14]) W. LEVEQUE, Trans. Amer. Math. Soc. **66** (1949) 440—463. I. J. Schoenberg,

[15]) Über asymptotische Verteilung reeller Zahlen mod 1, Math. Zeitschrif **28** (1928) 171—200.

[16]) On the asymptotic distribution of arithmetical functions, Trans. Amer. Math. Soc. **39** (1936) 315—330.