On the distribution of values of additive number-theoretical functions

By A. RÉNYI (Budapest)

Introduction

In our paper with P. TURÁN [1] we have given a new proof for the theorem of P. ERDŐS and M. KAC [2] according to which the distribution of the number of prime factors of the natural number n ($1 \leqq n \leqq N$) tends for $N \to +\infty$ to the normal distribution. Our method enabled us also to prove the conjecture of J. W. V. LE VEQUE concerning the exact order of magnitude of the remainder term in the theorem of ERDŐS—KAC. Our method consisted in using the usual methods of analytical number theory, i. e. expressing the sum of coefficients of a Dirichlet series by a contour integral, and in using certain function-theoretical properties of the ζ-function.

It is natural to guess that this powerful method yields also other known results on the distribution of values of additive number-theoretical functions. This is in fact true: however, for this purpose the technique has to be developed somewhat further. In the present paper we shall give a new proof running on the same lines as our proof of the Erdős—Kac theorem in [1], for the following theorem of P. ERDŐS [3]: *Let $f(n)$ be a real valued additive number-theoretical function, and put*

$$f^*(n) = \begin{cases} f(n) & \text{for} \quad |f(n)| \leqq 1 \\ 0 & \text{for} \quad |f(n)| > 1. \end{cases}$$

Put

(1) $$F_N(x) = \frac{1}{N} \sum_{\substack{f(n)<x \\ n \leqq N}} 1.$$

Then the distribution-functions $F_N(x)$ tend for $N \to +\infty$ to a limiting distribution function $F(x)$ at all points of continuity of the latter[1]*), if the following three conditions are satisfied:*

1. $$\sum_p \frac{f^*(p)}{p} \quad \textit{is convergent,}$$

2. $$\sum_p \frac{(f^*(p))^2}{p} < +\infty,$$ 3. $$\sum_{|f(p)|>1} \frac{1}{p} < +\infty.$$

---

[1]) It has been shown also by ERDŐS that if the series $\sum_{f(p) \neq 0} \frac{1}{p}$ is divergent then $F(x)$ is continuous, and thus $F_N(x)$ tends for $N \to +\infty$ to $F(x)$ for every value of x.

(Here and in what follows p runs over the sequence of primes). In this case the distribution function $F(x)$ has the characteristic function

$$(2)\qquad g(u)=\int_{-\infty}^{+\infty} e^{iux}\,dF(x)=\prod_p\left(1+\sum_{k=1}^{\infty}\frac{e^{iuf(p^k)}-e^{iuf(p^{k-1})}}{p^k}\right).$$

It has been proved later by P. ERDŐS and A. WINTNER [4] that the conditions 1), 2) and 3) are also necessary for the validity of the assertion of the theorem of ERDŐS.

Before giving in § 1 a new proof for this theorem let us mention that the theorem of ERDŐS and WINTNER shows a clear-cut resemblance to the three-series theorem of KOLMOGOROFF in probability theory. However, a proof of the Erdős—Wintner theorem, reducing it to the three-series theorem, has not been found up to now. An attempt in this direction has been made recently by E. M. PAUL [5] but he succeeded only in proving by purely probabilistic considerations that under the above conditions 1)—3) the *logarithmic* density of the numbers n, for which $f(n)<x$, exists and is equal to $F(x)$. This is, however, essentially less than the assertion that the ordinary density of these numbers equals $F(x)$. A new proof of the theorem of ERDŐS and WINTNER has recently been given by H. DELANGE [10].

In § 2. we make some remarks on the Erdős—Wintner theorem. We intend to return to other applications of the method in question (e. g. to prove a theorem of I. P. KUBILIUS [6]) in an other paper.

§ 1. Analytic proof of the theorem of Erdős

Before going into details, we give a sketch of the proof. Let us put[2])

$$(1.1)\qquad f_N(n)=\sum_{\substack{p^k\| n\\ p^k\leqq N}}{}' f(p^k).$$

Then clearly

$$(1.2)\qquad f_N(n)=f(n)\quad\text{for}\quad n\leqq N$$

In order to prove the theorem in question, it would be sufficient, — according to the well known continuity theorem concerning characteristic functions — to show that for all real u one has

$$(1.3)\qquad \lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}e^{iuf(n)}=g(u),$$

where $g(u)$ is defined by (2). In view of (1.2) we can write (1.3) in the form

$$(1.4)\qquad \lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}e^{iuf_N(n)}=g(u).$$

---

[2]) $p^k\|n$ means that n is divisible by p^k but not by p^{k+1}.

Now clearly

$$\lambda_N(s,u)=\sum_{n=1}^{\infty}\frac{e^{iuf_N(n)}}{n^s}=\prod_p\left(1+\sum_{\substack{p^k\leqq N\\ k\geqq 1}}\frac{e^{iuf(p^k)}}{p^{ks}}+\sum_{p^k>N}\frac{1}{p^{ks}}\right),\tag{1. 5}$$

and thus

$$\lambda_N(s,u)=\zeta(s)\mu_N(s,u)\tag{1. 6}$$

where, putting $k(N,p)=\left[\frac{\log N}{\log p}\right]$, we have

$$\mu_N(s,u)=\prod_p\left(1+\sum_{\substack{p^k\leqq N\\ k\geqq 1}}\frac{e^{iuf(p^k)}-e^{iuf(p^{k-1})}}{p^{ks}}+\frac{1-e^{iuf(p^{k(N,p)})}}{p^{(k(N,p)+1)s}}\right).\tag{1. 7}$$

As $\mu_N(s,u)$ is a Dirichlet-polynomial, it is an entire function, and its values can be easily estimated. Thus the behaviour of $\lambda_N(s,u)$ is sufficiently known in order to evaluate the sum of its first N coefficients. However carrying out this program in a straightforward manner leads to a proof of (1. 4) only in the special case if instead of 1) the much more restrictive condition

$$\sum_p\frac{|f^*(p)|}{p}<+\infty\tag{1'.}$$

is satisfied, i. e. we get in this way not the full theorem of ERDŐS, only the special case which was previously obtained by I. J. SCHÖNBERG [7].

To get the full theorem we have to introduce the following modification of the method: we put

$$N^*=N^{\varepsilon_N}\tag{1. 8}$$

where ε_N is a sequence of positive numbers, tending to 0 for $N\to+\infty$, such that

$$\varepsilon_N\sim\frac{\alpha}{e^{(\log\log N)^\beta}}\qquad\text{with}\quad\alpha>0,\ \beta>\frac{1}{2}\tag{1. 9}$$

and prove first

$$\lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}e^{iuf_{N^*}(n)}=g(u).\tag{1. 10}$$

As a matter of fact, we prove more, namely we obtain also an estimate (see (1. 17)) of the remainder term in (1. 10), and from this relation we can already deduce (1. 4). The situation is still somewhat complicated, because for technical reasons we consider instead of the sum $\frac{1}{N}\sum_{n=1}^{N}e^{iuf_{N^*}(n)}$ the related sum

$$g_N(u,N^*)=\frac{1}{N}\sum_{n=1}^{N}e^{iuf_{N^*}(n)}\cdot\log\frac{N}{n}.\tag{1. 11}$$

Let us pass now to the details. We have (see [8])

$$g_N(u,N^*)=\frac{1}{2\pi i}\int_{(c)}\frac{\zeta(s)\mu_{N^*}(s,u)N^{s-1}\,ds}{s^2}\tag{1. 12a}$$

and thus

$$g_N(u, N^*) = \frac{\mu_{N^*}(1, u)}{2\pi i} \int_{(c)} \frac{N^{s-1}\, ds}{s^2(s-1)} + \tag{1.12b}$$

$$+\frac{1}{2\pi i} \int_{(c)} \left(\zeta(s) - \frac{1}{s-1}\right) \frac{\mu_{N^*}(s, u) N^{s-1}\, ds}{s^2} + \frac{1}{2\pi i} \int_{(c)} \frac{(\mu_{N^*}(s, u) - \mu_{N^*}(1, u) N^{s-1}\, ds}{(s-1)s^2}$$

where $\int_{(c)}$ denotes that the integration has to be carried out on the line $s = c+it$ from $t = -\infty$ to $t = +\infty$ and c is a real number, $c>1$. Now according to the theorem of residues

$$\frac{1}{2\pi i} \int_{(c)} \frac{N^{s-1}\, ds}{(s-1)s^2} = 1 - \frac{1+\log N}{N}. \tag{1.13}$$

On the other hand evidently under conditions 1)—3)

$$\lim_{N \to +\infty} \mu_{N^*}(1, u) = g(u) \tag{1.14}$$

for all real u. Thus the first term on the right hand side of (1. 12) tends to $g(u)$; it remains to prove that the other two terms tend sufficiently rapidly to 0. As the integrand in both terms is regular for $s = \sigma + it$, $\sigma > 0$, the path of integration can in both integrals be shifted to the line $s = 1+it$ i. e. we may take $c=1$.

Now we have evidently for $s = 1+it$

$$|\mu_{N^*}(s, u)| = O\left(e^{|u| \cdot \sum_{p \leqq N^*} \frac{|f^*(p)|}{p}}\right)$$

and by the Cauchy inequality and condition 2) we obtain

$$\sum_{p \leqq N^*} \frac{|f^*(p)|}{p} = O(\sqrt{\log\log N}).$$

Thus

$$|\mu_{N^*}(s, u)| \leqq e^{|u| A \sqrt{\log\log N}} \tag{1.15}$$

where A is a positive constant. It is easy to see, that (1. 15) holds not only for $s = 1+it$ but also for $s = \sigma + it$ where $1 - \frac{1}{\log N^*} \leqq \sigma \leqq 1$.

Similarly we obtain (denoting by $\mu'_{N^*}(s, u)$ the derivative of $\mu_{N^*}(s, u)$ with respect to s) for $s = 1+it$

$$|\mu'_{N^*}(s, u)| = O\left(\varepsilon_N \cdot \log N \cdot e^{A|u| \sqrt{\log\log N}}\right) \tag{1.16}$$

where A is a positive constant. Thus we obtain by partial integration, taking into account the well-known estimates (see [8])

$$|\zeta(1+it)| = O(\log |t|)$$

$$\left|\frac{\zeta'(1+it)}{\zeta(1+it)}\right| = O(\log |t|)$$

that the second term on the right of (1. 12) is of order $O\left(\varepsilon_N \cdot e^{A|u|\sqrt{\log\log N}}\right)$ and thus tends to 0 for $N \to +\infty$. As regards the third term of the right of (1. 12) the path of integration can be transformed into the broken line $\mathfrak{L}$ consisting of the 5 segments

$$\mathfrak{L}_1: \quad s = 1+it \quad -\infty < t \leqq -1,$$

$$\mathfrak{L}_2: \quad s = \sigma - i \quad 1 \geqq \sigma \geqq 1 - \frac{1}{\log N^*},$$

$$\mathfrak{L}_3: \quad s = 1 - \frac{1}{\log N^*} + it \quad -1 \leqq t \leqq +1,$$

$$\mathfrak{L}_4: \quad s = \sigma + i \quad 1 - \frac{1}{\log N^*} \leqq \sigma \leqq 1,$$

$$\mathfrak{L}_5: \quad s = 1+it \quad 1 \leqq t \leqq +\infty.$$

Now as (1. 15) holds for $s = 1 - \dfrac{1}{\log N^*} + it$ the contribution of the line $\mathfrak{L}_3$ to the third term on the right of (1. 12) is of order

$$O\left(\log N \cdot e^{-\frac{1}{\varepsilon_N}}\right) = O\left(\varepsilon_N e^{A|u|\sqrt{\log\log N}}\right).$$

On $\mathfrak{L}_1$ and $\mathfrak{L}_5$ we can again apply partial integration, and use the estimates (1. 15) resp. (1. 16) and thus obtain that the contributions of the segments $\mathfrak{L}_1$ and $\mathfrak{L}_5$ are also of order $O\left(\varepsilon_N e^{A|u|\sqrt{\log\log N}}\right)$.

Finally, in view of

$$\int\limits_{1-\frac{1}{\log N^*}}^{1} N^{\sigma-1}\, d\sigma = O\left(\frac{1}{\log N}\right)$$

the contribution of the segments $\mathfrak{L}_3$ and $\mathfrak{L}_5$ is of order $O\left(\dfrac{e^{A|u|\sqrt{\log\log N}}}{\varepsilon_N \log N}\right) =$ $= O\left(\varepsilon_N e^{A|u|\sqrt{\log\log N}}\right)$. Thus collecting all estimates, we obtain

$$g_N(u, N^*) = \mu_{*N}(1, u) + O\left(\varepsilon_N \cdot e^{A|u|\sqrt{\log\log N}}\right). \tag{1. 16}$$

Now let us apply (1. 16) for N and $N' = \left[N\left(1+\dfrac{1}{\log N}\right)\right]$ but with the same N^* and consider the difference of the two expressions. We obtain

$$\frac{1}{N}\sum_{n=1}^{N} e^{iuf_{N^*}(n)} = \mu_{N^*}(1, u) + O\left(\varepsilon_N e^{A|u|\sqrt{\log\log N}}\right). \tag{1. 17}$$

Now let us consider the difference

$$d_N = \frac{1}{N}\sum_{n=1}^{N} e^{iuf_{N^*}(n)} - \frac{1}{N}\sum_{n=1}^{N} e^{iuf(n)}. \tag{1. 18}$$

We have evidently

$$d_N = \frac{1}{N}\sum_{n=1}^{N} e^{iuf_{N^*}(n)}\cdot\left(e^{iu\sum\limits_{\substack{p^k\|n\\ N^*<p^k\le N}} f(p^k)}-1\right).$$

Now let us collect the terms in which there occurs in the exponent a prime power p^k with $k\geqq 2$; the sum of these terms does not exceed

$$\frac{1}{N}\sum_{N^*\leqq p^k,\, k\geqq 2}\left[\frac{N}{p^k}\right]=O\left(\sum_{p\geqq\sqrt{N^*}}\frac{1}{p^2}\right).$$

Thus the sum of these terms tends to 0. It follows that

$$d_N = \frac{1}{N}\sum_{n=1}^{N} e^{iuf_{N^*}(n)}\left(e^{iu\sum\limits_{\substack{N^*<p\le N\\ p\mid n}} f(p)}-1\right)+o(1). \tag{1.19}$$

We can further disregard the terms in which there occurs in the exponent a p such that $|f(p)|>1$, because the contribution of these terms is of order $O\left(\sum\limits_{\substack{|f(p)|>1\\ p>N^*}}\frac{1}{p}\right)$.

Thus we have

$$d_N = \frac{1}{N}\sum_{n=1}^{N} e^{iuf_{N^*}(n)}\left(e^{iu\sum\limits_{\substack{p\mid n\\ N^*<p\le N}} f^*(p)}-1\right)+o(1). \tag{1.20}$$

As

$$e^{i\alpha}=1+i\alpha+\frac{\vartheta\alpha^2}{2}\quad\text{where}\quad|\vartheta|\leqq 1 \tag{1.21}$$

we have

$$d_N = D_N+R_N+o(1), \tag{1.22}$$

where

$$D_N=\frac{iu}{N}\sum_{n=1}^{N} e^{iuf_{N^*}(n)}\Big(\sum_{\substack{p\mid n\\ N^*<p\le N}} f^*(p)\Big)$$

and

$$|R_N|\leqq\frac{u^2}{2N}\sum_{n=1}^{N}\Big(\sum_{\substack{p\mid n\\ N^*<p\le N}} f^*(p)\Big)^2.$$

Clearly

$$|R_N|=O\left(\sum_{N^*<p}\frac{f^{*2}(p)}{p}+\left(\sum_{N^*<p<N}\frac{f^*(p)}{p}\right)^2\right)$$

Thus in view of conditions 1), 2) and 3) it follows that

$$\lim_{N\to+\infty} R_N=0, \tag{1.23}$$

and thus

$$d_N = D_N+o(1). \tag{1.24}$$

Let us consider now the term D_N. We have evidently

$$(1.25)\qquad D_N = \frac{iu}{N} \sum_{N^* < p \leqq N} f^*(p) \cdot \left[\sum_{m=1}^{\left[\frac{N}{p}\right]} e^{iuf_{N^*}(m)} \right].$$

Now let us consider first the contribution of those terms to D_N in which $p > \sqrt{N}$. The contribution of these terms clearly does not exceed $|u| \cdot \sum_{\sqrt{N} < p \leqq N} \frac{|f^*(p)|}{p}$. As we have

$$\sum_{\sqrt{N} < p \leqq N} \frac{|f^*(p)|}{p} \leqq \sqrt{\left(\sum_{\sqrt{N} < p} \frac{f^{*2}(p)}{p} \right) \cdot \left(\sum_{\sqrt{N} < p \leqq N} \frac{1}{p} \right)} = o(1),$$

we obtain

$$(1.26)\qquad D_N = \frac{iu}{N} \sum_{N^* < p \leqq \sqrt{N}} f^*(p) \cdot \left(\sum_{m=1}^{\left[\frac{N}{p}\right]} e^{iuf_{N^*}(m)} \right) + o(1).$$

Now we apply (1. 17) to each sum $\sum_{m=1}^{\left[\frac{N}{p}\right]} e^{iuf_{N^*}(m)}$ in (1.26). Taking into account that for $N^* < p \leqq \sqrt{N}$

$$N^* = N^{\varepsilon_N} = \left(\frac{N}{p} \right)^{\varepsilon(N,P)}$$

where

$$(1.27)\qquad \varepsilon(N,P) = \frac{\varepsilon_N \log N}{\log N - \log p} = \vartheta \varepsilon_N \quad \text{with} \quad 1 \leqq \vartheta \leqq 2$$

it follows that

$$(1.28)\qquad D_N = \frac{iu}{N} \sum_{N^* < p \leqq \sqrt{N}} f^*(p) \left[\frac{N}{p} \right] \mu_{N^*}(1, u) + \varrho_N,$$

where

$$|\varrho_N| \leqq \left(\sum_{N^* < p \leqq \sqrt{N}} \frac{|f^*(p)|}{p} \right) \cdot O\left(\varepsilon_N e^{A|u|\sqrt{\log\log N}}\right).$$

Now, applying the estimate

$$\sum_{N^* < p \leqq \sqrt{N}} \frac{|f^*(p)|}{p} = O(\sqrt{\log\log N}),$$

we obtain

$$(1.29)\qquad \lim_{N \to +\infty} \varrho_N = 0.$$

On the other hand we have

$$\left(\sum_{N^* < p \leqq \sqrt{N}} f^*(p) \left[\frac{N}{p} \right] \frac{1}{N} \right) = \sum_{N^* < p \leqq \sqrt{N}} \frac{f^*(p)}{p} + O\left(\frac{1}{\sqrt{N}} \right).$$

As by condition 1) $\sum_{N^*<p\leq\sqrt{N}} \frac{f^*(p)}{p}$ tends to 0 and by (1. 14) $\mu_{N^*}(1, u)$ tends to the limit $g(u)$ and thus is bounded, it follows that

(1. 30)
$$\lim_{N\to+\infty} D_N = 0.$$

Thus we have proved that

(1. 31)
$$\lim_{N\to+\infty} d_N = 0$$

and therefore (1. 17) implies the validity of (1. 3), which was to be proved. Thus the proof of the theorem of ERDŐS is completed.

§ 2. Some remarks on the Erdős—Wintner theorem

In the course of the proof of the theorem of ERDŐS in § 1 we have made use of the evident fact, that in case the conditions 1)—3) are satisfied the infinite product

(2. 1)
$$g(u) = \prod_p \left(1 + \sum_{k=1}^{\infty} \frac{e^{iuf(p^k)} - e^{iuf(p^{k-1})}}{p^k}\right)$$

is convergent for every real u. Conversely it is easy to see that the convergence of this product for all real u implies the fulfilment of conditions 1)—3).

This can be shown as follows. If the infinite product (2. 1) is convergent, then clearly the series

(2. 2)
$$\sum_p \frac{e^{iuf(p)} - 1}{p}$$

is convergent too. Thus the real part of the series, i. e. the series with non-negative terms

(2. 3)
$$\sum_p \frac{1 - \cos uf(p)}{p}$$

is convergent for every real u. Thus both of the series

(2. 4)
$$\sum_{|f(p)|>1} \frac{1 - \cos uf(p)}{p}$$

and

(2. 5)
$$\sum_p \frac{1 - \cos uf^*(p)}{p}$$

are convergent too.

Now the convergence of (2. 5) clearly implies the convergence of the series $\sum_p \frac{f^{*2}(p)}{p}$. On the other hand, it can be shown that the convergence of the series (2. 4) implies the convergence of the series $\sum_{|f(p)|>1} \frac{1}{p}$. To show this we need the following

Lemma. *If the series* $\sum_{n=1}^{\infty} b_n(1-\cos a_n x)$ *where* $b_n \geqq 0$ *and* $a_n \geqq 1$ $(n=1, 2, \ldots)$ *is convergent for every real* x, *then the series* $\sum_{n=1}^{\infty} b_n$ *is convergent.*

PROOF OF LEMMA [3]). Let us put

$$S(x) = \sum_{n=1}^{\infty} b_n(1-\cos a_n x). \tag{2.6}$$

As $S(x)$ is finite for every real x, and is a measurable function of x, there exists a set E of positive measure lying in a finite interval I on which $S(x)$ is bounded, $S(x) \leqq K$. Denoting by $e(x)$ the indicator of the set E (i. e. $e(x)=1$ for $x \in E$ and $e(x)=0$ otherwise), and integrating the series (2. 6) on the set E (the integration can be carried out term by term owing to BEPPO LEVI's theorem) we obtain

$$\sum_{n=1}^{\infty} b_n \int_I (1-\cos a_n x) e(x)\, dx \leqq K|E|$$

where $|E|$ denotes the measure of the set E. Now evidently $G(a) = \int_I (1-\cos ax) e(x) dx$ as a function of a is bounded from below by a positive constant for $1 \leqq a < +\infty$. (As a matter of fact, $G(a)$ is everywhere positive, continuous and by Riemann 's lemma $\lim_{a \to +\infty} G(a) = |E| > 0$.) It follows, putting $\min_{1 \leqq a < +\infty} G(a) = g$, that

$$g \sum_{n=1}^{\infty} b_n \leqq K \cdot |E|.$$

This proves the Lemma.

Now the convergence of the series $\sum_p \frac{f^{*2}(p)}{p}$ and $\sum_{|f(p)|>1} \frac{1}{p}$, together with that of the series (2. 2) imply by (1. 21) the convergence of the series $\sum_p \frac{f^*(p)}{p}$. Thus conditions 1)—3) are satisfied.

Therefore the theorem of ERDŐS and WINTNER can also be formulated as follows: *for the existence of the limiting distribution of the values of an additive arithmetic function $f(n)$ the convergence of the infinite product* (2. 1) *is necessary and sufficient.*

This reformulation of their theorem has been already given by ERDŐS and WINTNER themselves in [4]; however, they deduced this form of the theorem from the original one in another, indirect way, by using the theory of infinite convolutions.

---

[3]) This Lemma can be proved in exactly the same way as the theorem of DENJOY—LUSIN (see [9], p. 232); nevertheless, for the sake of the convenience of the reader, I give here the proof.

Bibliography

[1] A. RÉNYI—P. TURÁN, On a theorem of Erdős and Kac, *Acta Arithmetica* **1** (1957), 71—84.
[2] P. ERDŐS—M. KAC, The Gaussian law of errors in the theory of additive number-theoretic functions, *Amer. J. Math.* **62** (1940), 738—742.
[3] P. ERDŐS, On the density of some sequences of numbers, III. *J. London Math. Soc.* **13** (1938), 119—127.
[4] P. ERDŐS—A. WINTNER, Additive arithmetic functions and statistical independence, *Amer. J. Math.* **61** (1939), 713—721.
[5] I. P. KUBILIUS, Probabilistic methods in number theory, 2nd edition (in Russian), *Vilnius*, 1962.
[6] E. M. PAUL, Density in the light of probability theory, *Sankhyā* **24** (1962), 103—114.
[7] I. J. SCHÖNBERG, On the asymptotic distribution of arithmetical functions, *Trans. Amer. Math. Soc.* **39** (1936), 315—330.
[8] E. C. TITCHMARSH, The theory of the Riemann Zeta-function, *Oxford*, 1951.
[9] A. ZYGMUND, Trigonometric series, I., *Cambridge*, 1959.
[10] H. DELANGE, Un théorème sur les fonctions arithmétiques multiplicatives et ses applications, *Ann. Sci. Ecole Norm. Sup.* **78** (1961), 1—29.

(Received April 2, 1963.)