Publ. Math. Debrecen
40 / 3–4 (1992), 211–217

Additive functions vanishing outside a given set of primes. Part I: Limiting distribution on short intervals

By JÁNOS GALAMBOS* (Philadelphia, PA) and
KARL-HEINZ INDLEKOFER (Paderborn)

The main results

Let $f(n)$ be a real valued strongly additive arithmetical function; that is, $f(nm) = f(n) + f(m)$ for coprime n and m, and $f(p^k) = f(p)$ for all primes p and integers $k \geq 1$. Throughout the paper p, as a variable, goes through the prime numbers. For $x \geq 2$, let E_x be a given set of primes satisfying

$$\text{for all} \quad p \in E_x, \quad p \leq x^{g(x)} \tag{1}$$

where

$$0 < g(x) \to 0 \quad \text{and} \quad x^{g(x)} \to +\infty \quad \text{as} \quad x \to +\infty. \tag{2}$$

Define the strongly additive arithmetical function $f_x(n)$ by $f_x(p) = f(p)$ if $p \in E_x$ and $f_x(p) = 0$ otherwise. Our aim is to investigate the distribution of the values of $f_x(n)$ as n goes through the integers between $x - y$ and x, where $y = y(x) \to +\infty$ with x but with some limitations as to how small it may be. We say that $f_x(n)$ has a limiting distribution $F(z)$ on the interval $x - y < n \leq x$, $0 < y \leq x$, if $F(z)$ is a proper distribution function and, for all continuity points z of $F(z)$,

$$\lim \nu_{x,y}\{f_x(n) \leq z\} = F(z) \quad \text{as} \quad x \to +\infty, \tag{3}$$

where $\nu_{x,y}\{\dots\} = N_{x,y}\{\dots\}/y$ and $N_{x,y}\{\dots\}$ signifies the number of integers $x - y < n \leq x$ for which the property stated in the dotted space

*Work was partially done while visiting the University of Paderborn as a Fellow of the Humboldt Foundation.

holds. Notice that $\lim \nu_{x,y}\{A\} = d(A)$ becomes the natural density of A if $y = x$.

One of our results is a necessary and sufficient condition for (3) under the sole assumption $y \geq x^{h(x)}$, where

$$h(x) = r(x) \max\left[g(x), \frac{2e(\log\log x)(\log\log\log x)}{\log x}\right] \tag{4}$$

with an arbitrary $r(x) \nearrow +\infty$ with x such that $0 < h(x) \leq 1$. Hence, in view of (2) and (4), we can choose $y(x)$ over a wide range, including $y(x) = x$, thus extending and unifying a variety of results previously known for the special case of natural density only. Note also that, if all prime divisors of n are outside the set E_x, $f_x(n) = 0$, and thus the essential contribution to the set of values of $f_x(n)$ comes from a subset of the consecutive integers rather than from the set of all consecutive integers. In this sense, the work is also related to, although not overlapping with, the papers of KÁTAI [7] and [8] dealing with the distribution of the values of additive functions on the particular subset "primes plus one" of the integers.

Before the formulation of the main results, let us explain the assumption at (1) and (2). It is known (DICKMAN [2]) that the set of integers n whose largest prime divisor $P(n)$ satisfies

$$\lim[\log P(n)]/\log n \leq a \quad \text{as} \quad n \to +\infty$$

has positive density, where $0 < a < 1$ is arbitrary. Therefore, if the restriction at (1) and (2) on E_x were dropped, the single term $f(P(n))$ would have an influence on the limiting distribution of $f_x(n)$ when E_x is a "rare set" (such as in the case of Corollary 2 below), while $f(P(n))$ has no influence on Corollary 1 below. Consequently, a unified approach were not possible to these cases.

The main results of the present paper are as follows.

Theorem 1. *Let E_x satisfy (1) and (2). Let $r(x)$ be an increasing function tending to infinity with x and such that $h(x)$ of (4) satisfies $0 < h(x) \leq 1$. Then, for any $y = y(x)$ satisfying $x^{h(x)} \leq y \leq x$, (3) holds if, and only if, as $x \to +\infty$,*

$$\varphi_x(t) = \prod_{p \in E_x} \left(1 - \frac{1}{p}\right)\left(1 + \frac{e^{it\, f(p)}}{p-1}\right) \longrightarrow \varphi(t), \tag{5}$$

where $\varphi(t)$ is a function continuous at $t = 0$.

Notice that $y(x)$ does not appear in (5). Hence, $f_x(n)$ has the same asymptotic distributional properties for all permissible y of the theorem. The following corollaries extend the results of ERDÖS and WINTNER [4] and of DE KONINCK and GALAMBOS [1] to short intervals; consequently, these quoted results are unified by our approach.

Corollary 1. *Let E_x be the set of all primes satisfying (1) and (2). Then, for all y of Theorem 1, (3) holds if, and only if, each of the series*

$$\sum_{|f(p)|\le 1} \frac{f(p)}{p}, \quad \sum_{|f(p)|\le 1} \frac{f^2(p)}{p}, \quad \sum_{|f(p)|>1} \frac{1}{p} \tag{6}$$

converges.

Corollary 2. *Let E_x be such that, in addition to (1) and (2), if $p \in E_x$ then $p \to +\infty$ with x. Let $f(p) = 1$ for all p and assume that, as $x \to +\infty$,*

$$\sum_{p\in E_x} \frac{1}{p} \to \lambda, \quad \text{where} \quad 0 < \lambda < +\infty. \tag{7}$$

Then, for all y of Theorem 1, as $x \to +\infty$,

$$\nu_{x,y}\{f_x(n) = k\} \longrightarrow \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \ge 0. \tag{8}$$

It is well known in probability theory, for which in this number theoretic context the most convenient reference is ELLIOTT [3], pp. 27–28, that (3) is equivalent to the limit relation

$$\lim_{x=+\infty} \frac{1}{y} \sum_{x-y<n\le x} \exp[it\, f_x(n)] = \varphi(t), \ t \text{ real}, \tag{9}$$

where $\varphi(t)$ is a continuous function at $t = 0$. Hence, the following result immediately implies Theorem 1.

Theorem 2. *With the notation and assumption of Theorem 1, as $x \to +\infty$,*

$$\sum_{x-y<n\le x} \exp[it\, f_x(n)] = y \prod_{p\in E_x} \left(1 - \frac{1}{p}\right)\left(1 + \frac{e^{it\, f(p)}}{p-1}\right) + o(y). \tag{10}$$

Tools from the literature

In this section we collect a number of results from the literature which we need in the proof of Theorem 2.

Lemma 1. *For all real $x \ge 2$,*

$$\sum_{p\le x} \frac{1}{p} \exp\left(-\frac{\log x}{\log p}\right) \le c,$$

where $c > 0$ is a universal constant.

PROOF. Since, for $p \leq x$, $(\log p)^{-1} \exp\left[-(\log x)/\log p\right]$ is strictly increasing in p, the asymptotic formula (see Chapter 22 in HARDY and WRIGHT [5])

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1) \tag{11}$$

entails the inequality of the lemma.

For the next statements we introduce further notation. Let $r^*(x)$ be a positive and increasing function which tends to infinity with x. We set

$$g^*(x) = \max\left(g(x), \frac{2e(\log\log x)(\log\log\log x)}{\log x}\right) \tag{12}$$

and

$$G(x) = x^{g(x)}, \quad T = T(x) = x^{r^*(x)g^*(x)}. \tag{13}$$

Lemma 2. *There exists a constant c_1 such that, for $\log x \leq w \leq G(x)$,*

$$\sum_{\substack{T<m \\ P(m)\leq w}} \frac{1}{m} \ll \exp\left(\sum_{p \leq w} \frac{1}{p} - c_1\ u \log u\right),$$

where $u = (\log T)/\log w$ and $P(m)$ is the largest prime divisor of m.

PROOF. By an idea of RANKIN [9] one obtains (see INDLEKOFER [6], p. 268 for further detail)

$$\sum_{\substack{T<m \\ P(m)\leq w}} \frac{1}{m} < \sum_{\substack{T<m \\ P(m)\leq w}} \left(\frac{m}{T}\right)^{\varepsilon} \frac{1}{m} \ll \exp\left(\sum_{p \leq w} \frac{1}{p} + c_2 w - \varepsilon \log T\right),$$

where $0 < \varepsilon < 1/3$. By an appropriate choice of ε, the lemma now follows.

Lemma 3. *For a given $h(x)$ of (4), let $r^*(x)$ satisfy $r^*(x) < r(x)/2$. Let $y = x^{h(x)}$. Let a_n be the n-th integer exceeding $x - y$ (i.e., $a_n = x - y + n$), and let b_n be the product of all prime divisors of a_n which do not exceed $G(x)$. Denote by $K(y; \mu, s)$ the number of integers $n \leq y$ for which μ divides b_n and each prime divisor of b_n/μ is at least s. Then, for $\mu < T^{3/2}$ and $s \leq G(x)$,*

$$K(y; \mu, s) = \frac{y}{\mu} \prod_{p \leq s} \left(1 - \frac{1}{p}\right)(1 + R_1 + R_2),$$

where
$R_1 = O\Big\{ \exp\big[-u(\log u - \log\log 3u - 2)\big]\Big\}$, $R_2 = O\Big\{ \exp\big[-\log^{\frac{1}{2}}(y/\mu)\big]\Big\}$,
and $u = \big[\log(y/\mu)\big]/\log s$.

PROOF. Such a conclusion is a typical consequence of Selberg's sieve. The details of proof are identical to the one given by ELLIOTT [3], p. 79 and pp. 84–86, which we do not repeat here. Note that

$$s \le G(x) \le T^{1/2} \le T^2/\mu < y/\mu.$$

The proof of the main results

Put

$$F_{x,y}(z) = \nu_{x,y}\{f_x(n) \le z\}.$$

Then

$$\int_{-\infty}^{+\infty} e^{itz} dF_{x,y}(z) = \frac{1}{y} \sum_{x-y<n\le x} \exp\big[it\, f_x(n)\big],$$

and thus, as mentioned at (9), Theorem 1 is an immediate consequence of Theorem 2.

PROOF of Theorem 2. Let a_n be the n-th integer exceeding $x - y$. Write $a_n = b_n B_n$, where each prime factor of B_n exceeds $G(x)$, while for $p \mid b_n$ we have $p \le G(x)$. Hence, $f_x(a_n) = f_x(b_n)$. Consequently, we have to establish an asymptotic formula for

$$\varphi_{x,y}(t) = \frac{1}{y} \sum_{n\le y} \exp\big[it\, f_x(b_n)\big]. \tag{14}$$

With the notations at (12) and (13), we split up the above sum according as $b_n \le T$ or $b_n > T$, and denote the respective sums by $\sum_1$ and $\sum_2$. We first estimate $\sum_2$, and show that its asymptotic value is $o(y)$. Clearly, in view of the assumptions on T and the prime divisors of b_n, every $b_n > T$ can be written as $b_n = \mu m$, $T \le \mu < T^{3/2}$, and $p \mid m$ entails $p \ge P(\mu) = s$, say. Hence, upon denoting by $Q(d)$ the smallest prime divisor of d, we get

$$\sum\nolimits_2 \le \sum_{T\le\mu<T^{3/2}} \sum_{\substack{n\le y \\ b_n=\mu m,\ Q(m)\ge s}} 1 \le \sum_{s\le G(x)} \sum_{\substack{T/s<\mu_1<T^{3/2}/s \\ P(\mu_1)\le s}} \sum_{\substack{n\le y \\ \mu_1 s\mid b_n \\ Q(b_n/\mu_1 s)\ge s}} 1,$$

where the last estimate is obtained by first writing $\mu = s\mu_1$ and then letting s run through all primes upto $G(x)$. We now apply Lemma 3 to

the innermost sum in the last expression. Hence, by changing s to our customary notation p for primes, and dropping the subscript, we have

$$\sum_2 \ll y \sum_{p \le G(x)} \frac{1}{p} \prod_{p^* \le p} \left(1 - \frac{1}{p^*}\right) \sum_{\substack{T/p < \mu < T^{3/2}/p \\ P(\mu) \le p}} \frac{1}{\mu}. \tag{15}$$

We further split the summation over p: in $\sum_{21}$ we sum for $p \le \log x$ and in $\sum_{22}$ for $\log x < p \le G(x)$. Now, when estimating $\sum_{21}$, we once more utilize Rankin's method and estimate the inner sum by

$$\left(\frac{T}{p}\right)^{-\varepsilon} \prod_{p^* \le p} \left[1 + \sum_{k \ge 1} (p^*)^{-k(1-\varepsilon)}\right],$$

where we choose $\varepsilon = 1/\log\log x$. By applying the elementary inequalities $1 - a < e^{-a}$ and $1 + a < e^a$, $a > 0$, and appealing to

$$\sum_{p \le z} \frac{1}{p} \sim \log\log\ z,$$

we get

$$\sum_{21} \ll \sum_{p \le \log x} \frac{1}{p\ \log\ p} \exp\left[-\frac{\log\ T - \log\ p}{\log\log\ x} + e\ \log\log\log\ x\right],$$

which is $o(1)$ in view of (12) and (13). The asymptotic formula $\sum_{22} = o(1)$ is immediate from Lemmas 2 and 1.

Finally, we turn to $\sum_1$. Note that the same value $b = b_n$ comes up several times. In fact, the exact form of $\sum_1$ is as follows

$$\sum_1 = \sum_{\substack{b \le T \\ P(b) \le G(x)}} \exp[it\ f_x(b)] \sum_{\substack{n \le y \\ b \mid a_n,\ Q(a_n/b) > G(x)}} 1\ .$$

Now, the inner sum, by the sieve formula of Lemma 3 (by appropriately changing its notation), equals

$$\frac{y}{b} \left[\prod_{p \le G(x)} \left(1 - \frac{1}{p}\right)\right] [1 + o(1)]\ .$$

Therefore, upon noting that, by Lemma 2, the sum of

$$b^{-1} |\exp[it\ f_x(b)]| = 1/b$$

over $b > T$ and $P(b) \le G(x)$ is small, we can sum for all b with $P(b) \le$

$\leq G(x)$, which, by the strongly additive property of $f_x(b)$, leads to the product

$$\prod_{p \leq G(x)} \left(1 + \frac{e^{it\, f_x(p)}}{p-1}\right).$$

This completes the proof, because $f_x(p) = f(p)$ for $p \leq G(x)$.

The corollaries are immediate from Theorem 1. In the case of Corollary 1, the limit in (5) becomes an infinite product whose convergence is equivalent to the convergence of the three series of (6) (see ELLIOTT [3], Chapter 5). In Corollary 2, $f(p) = 1$, the smallest $p \to +\infty$ with x, and (7) applies. By taking logarithm and Taylor's expansion in (5), we get that the limit in (5) does exist and equals $\varphi(t) = \exp[\lambda(e^{it} - 1)]$. This specific form of the limit makes (5) and (8) equivalent (see ELLIOTT [3], pp. 28 and 52).

References

[1] J. M. DE KONINCK and J. GALAMBOS, The intermediate prime divisors of integers, *Proc. Amer. Math. Soc.* **101** (1987), 213–216.

[2] K. DICKMAN, On the frequency of numbers containing prime factors of a certain relative magnitude, *Ark. Mat. Astr. Fys.* **22** No. A10 (1930).

[3] P.D.T.A. ELLIOTT, Probabilistic number theory I, *Springer, Berlin*, 1979.

[4] P. ERDÖS and A. WINTNER, Additive arithmetical functions and statistical independence, *Amer. J, Math.* **61** (1939), 713–721.

[5] G.H. HARDY and E.M. WRIGHT, An introduction to the theory of numbers, 4th ed., *Oxford*, 1960.

[6] K.-H. INDLEKOFER, A mean-value theorem for multiplicative functions, *Math. Z.* **172** (1980), 255–271.

[7] I. KÁTAI, On distribution of arithmetical functions on the set of prime plus one, *Compositio Math.* **19** (1968), 278–289.

[8] I, KÁTAI, Some remarks on additive arithmetical functions, *Liet. mat. rinkinys* **9** (1969), 515–518.

[9] R.A. RANKIN, The difference between consecutive prime numbers, *J. London Math. Soc.* **13** (1938), 242–247.

JANOS GALAMBOS
DEPT. OF MATHEMATICS
TU 038-16
TEMPLE UNIVERSITY
PHILADELPHIA, PA 19122
USA

KARL-HEINZ INDLEKOFER
FACHBEREICH MATHEMATIK
DER UNIVERSITÄT
WARBURGER STRASSE 100
D-4790 PADERBORN
GERMANY

(Received February 15, 1990)