SMOOTHNESS OF THE DISTRIBUTIONS OF ARITHMETIC FUNCTIONS

G. J. BABU*
Department of Statistics, 219 Pond Laboratory, Pennsylvania State University, University Park, PA 16802, U.S.A.

ABSTRACT

It is well known that the distribution of a real-valued additive arithmetic function $f$, if it exists, is either purely discrete, purely continuous singular, or purely absolutely continuous. The distribution of $f$ is shown to be absolutely continuous if $f$ is not "too small" on primes. This result is generalized to additive functions on the set of pairs of positive integers. Conditions for existence of singular distributions for additive functions on the set of pairs of positive integers are also given. These are nontrivial extensions, as the convolution of two singular distributions need not be singular. A result on the distributions of multiplicative functions on the set of pairs of positive integers is also given.

1. INTRODUCTION

A real-valued arithmetic function $f$ is called additive if

$$f(mn) = f(m) + f(n),$$

whenever $(m,n) = 1$. A real-valued additive arithmetic function is said to have distribution $F$, if the density of $(m: f(m) \leq x)$ exists and equals to $F(x)$, whenever $x$ is a continuity point of $F$. Conditions, under which the distribution of a real-valued additive arithmetic function is absolutely continuous, are given in the next section. This result complements the work of (Babu, 1973), where among other results, it is established that every bounded additive arithmetic function has a singular distribution.

Similar results for the distribution of additive functions on the set of pairs of positive integers are given in Section 3. Though such distributions are related to the distributions of the associated "marginal" additive arithmetic functions, the results of Section 3 are nontrivial extensions. This is because the convolution of two singular distributions need not be singular. A result on multiplicative functions on the set of pairs of positive integers is also stated.

Throughout this paper let $p$ and $q$ denote prime numbers, $m$ and $n$ denote natural numbers, $\sum_p$ denote the sum over prime numbers and let $\prod_p$ denote the product over prime numbers.

© G. J. Babu, 1991

*Research supported in part by NSA Grant MDA 904-90-M-1001.
2. ABSOLUTE CONTINUITY

To obtain sufficient conditions, under which a real-valued additive arithmetic function has an absolutely continuous distribution, we introduce the following:

**Definition.** A positive real valued function $h$ on the interval $[2, \infty)$ is said to satisfy Condition K, if for some $x_0 \geq 3$,

i) $h$ is decreasing in $[x_0, \infty)$, \( \lim_{x \to \infty} h(x) = 0 \),

and

ii) for some $0 < \alpha < 2$, \( h(x)(\log x)^\alpha \) is nondecreasing on $[x_0, \infty)$.

**Theorem 1.** Let $Q$ be a set of primes satisfying

\[
\sum_{p \in Q} \frac{1}{p} < \infty.
\]

Let $f$ be a real-valued additive arithmetic function such that $|f(p)| = h(p)$ for all $p \in Q$, where $h$ satisfies Condition K. Then the distribution of $f$, if it exists, is absolutely continuous.

**Sketch of the Proof.** It is well known that $f$ has a distribution $F$, if and only if the product

\[
L(t) = \prod_p \left( 1 - \frac{1}{p} \right) \left( 1 + \sum_{k=1}^{\infty} \frac{1}{p^k} e^{itf(p^k)} \right)
\]

converges. If it converges, then $L$ denotes the characteristic function of $F$. See Theorem 4.5 in (Kubilius, 1964) or the proof of Theorem 1 in (Babu, 1974). By the Plancherel theorem (see, for instance, (Rudin, 1973, Thm. 7.9)), it is enough to show that $L$ is square integrable. This is established using an estimate of

\[
|L(2t)|^2 = \exp \left( -4 \sum_p \frac{1}{p} \sin^2 (\theta h(p)) + O(1) \right),
\]

and the well-known estimate that for some $\gamma > 0$ and $d > 0$,

\[
\sum_{p \leq x} \frac{1}{p} = \log \log x + \gamma + O \left( \exp \left( -d\sqrt{\log x} \right) \right).
\]

Towards this end, an appropriate lower bound for the sum in the exponent of (1) is obtained by expressing it as several sums over different intervals of primes, and estimating each sum as in (Babu, 1975a). See (Babu, 1991) for details.
Examples
The conditions of Theorem 1 hold if $f$ is a real-valued additive arithmetic function given by one of the following definitions:

a) $|f(p)| = (\log p)^{-\alpha}$, $0 < \alpha < 2$,

b) $|f(p)| = (\log \log p)^{-\alpha}$, $\alpha > 1/2$,

c) $|f(p)| = (\log \log p)^{-1/2}(\log \log \log p)^{-\alpha}$, $\alpha > 1/2$,

d) $|f(p)| = e^{-d(\log \log p)^{\alpha}}$, $0 < \alpha < 1$ and $d > 0$.

The arithmetic function $f$ in the examples above has absolutely continuous distribution, whenever it exists. The special case a) was considered in (Babu, 1973a). Theorem 1 complements the results of (Babu, 1973), where conditions for an additive arithmetic function to have a singular distribution are given.

3. FUNCTIONS ON THE SET OF PAIRS OF POSITIVE INTEGERS

A real-valued arithmetic function $g$ on the set of pairs of positive integers is said to be additive if $g(1, 1) = 0$ and

$$g(mi, nj) = g(m, n) + g(i, j),$$

whenever $(mn, ij) = 1$. It is called multiplicative if $g(1, 1) = 1$ and

$$g(mi, nj) = g(m, n)g(i, j),$$

whenever $(mn, ij) = 1$. A real-valued arithmetic function $g$ on the set of pairs of positive integers is said to have a distribution if for some probability distribution function $F$,

$$\frac{1}{MN} \# \{1 \leq m \leq M, 1 \leq n \leq N: g(m, n) \leq x\} \longrightarrow F(x),$$

as $M$ and $N$ tend to infinity independently, for all continuity points $x$ of $F$. See (Delange, 1969).

It is well known that the distribution of an additive function $g$ on the set of pairs of positive integers exists if and only if both the marginals $g_1(m) = g(m, 1)$ and $g_2(n) = g(1, n)$ have distributions. Further the distribution of $g$ is closely related to the convolution of the distributions of $g_1$ and $g_2$. Since convolution of two singular distributions is not necessarily singular the following result is of interest.

**Theorem 2.** Let positive numbers $\theta$ and $\beta$ satisfy the inequality

$$\theta^{-1} + \beta^{-1} > \frac{1}{2}.$$

Let $h_1$ and $h_2$ be two functions satisfying Condition K with $\alpha = \theta$ and $\alpha = \beta$, respectively. Suppose $g$ is a real-valued additive function on the set of pairs of
positive integers such that \(|g(p, 1)| = h_1(p)\) and \(|g(1, p)| = h_2(p)\) for all \(p \in Q\), where

\[
\sum_{p \not\in Q} \frac{1}{p} < \infty.
\]

Then the distribution of \(g\), if it exists, is absolutely continuous.

Proof. Using the estimates in the proof of Theorem 1, the characteristic function of the distribution of \(g\) can be shown to be square integrable under the given conditions on \(\theta\) and \(\beta\). Now the result follows from the Plancherel theorem. If \(\theta = 3\) and \(\beta = 4\) in Theorem 2, then we cannot conclude absolute continuity of the distributions of \(g_1\) and \(g_2\) from Theorem 1. But on the other hand \(g\) in Theorem 2 has an absolutely continuous distribution, as \(3^{-1} + 4^{-1} > 2^{-1}\).

Now we consider conditions for singularity of the distributions of additive functions on the set of pairs of positive integers. We require a result from probability theory, which is stated as Lemma 1 below. A proof of the lemma is given for easy reference.

**Lemma 1.** Let \(\{X_n\}\) and \(\{Y_n\}\) be two sequences of independent discrete random variables such that \(X = \sum X_n\) converges almost surely. If

\[
\sum_{n} \sum_{x} |P(X_n = x) - P(Y_n = x)| < \infty, \tag{3}
\]

then \(Y = \sum Y_n\) converges almost surely. Further, \(Y\) has absolutely continuous distribution if and only if \(X\) has absolutely continuous distribution. The distribution of \(Y\) is discrete if and only if the distribution of \(X\) is discrete.

Proof. In view of Theorem E of (Babu, 1978), it is enough to prove that \(\{X_n\}\) and \(\{Y_n\}\) can be defined on a single probability space without changing their distributions and such that

\[
\sum_{n} P(X_n \neq Y_n) < \infty.
\]

To show this, we fix \(n\). Suppose \(X_n\) and \(Y_n\) take values only in \(\{x_1, x_2, \ldots\}\). Define

\[
u_i = P(X_n = x_i) \quad \text{and} \quad v_i = P(Y_n = x_i).
\]

Let \(\lambda\) denote the Lebesgue measure on the unit interval.

\[
C = \{i: u_i \geq v_i\} = \{i_1 < i_2 < \ldots\},
\]

\[
B = \{i: u_i < v_i\} = \{j_1 < j_2 < \ldots\},
\]

\[
a_0 = 0, \quad a_j - a_{j-1} = u_j, \quad c_0 = \sum_j u_j, \quad c_i - c_{i-1} = u_j,
\]

and \(b_j = a_{j-1} + v_j\). Note that \(a_i\) and \(c_j\) make partitions of the unit interval.
If 
\[ A = \bigcup_j (b_j, a_j], \]
then 
\[ \lambda(A) = \sum_j (u_i - v_i), \]
and the set \( A \) can be partitioned into union of intervals \( I_i, i \in B \), such that 
\[ A = \bigcup_{i \in B} I_i \quad \text{and} \quad \lambda(I_i) = v_i - u_i. \]

Now for \( \omega \in (0, 1] \) we define 
\[ X'_n(\omega) = \begin{cases} x_i & \text{if } \omega \in (a_{j-1}, a_j], \\ x_{j+1} & \text{if } \omega \in (c_{j-1}, c_j]. \end{cases} \]
\[ Y'_n(\omega) = \begin{cases} x_i & \text{if } \omega \in (a_{j-1}, b_j], \\ x_{j+1} & \text{if } \omega \in (c_{j-1}, c_j \cup I_i]. \end{cases} \]

Since 
\[ \sum_{i \in C} (u_i - u_i) = \sum_{i \in B} (v_i - u_i), \]
\( X'_n \) and \( Y'_n \) have the same distributions as those of \( X_n \) and \( Y_n \), respectively. Further, 
\[ \lambda(X'_n \neq Y'_n) = \sum_{i \in C} (u_i - v_i) = \frac{1}{2} \sum_i |u_i - v_i| \]
\[ = \frac{1}{2} \sum_x |\mathbb{P}(X_n = x) - \mathbb{P}(Y_n = x)|. \] (4)

Let 
\[ \Omega = (0, 1] \times (0, 1] \times \ldots, \]
and let \( \mathbb{P}^0 \) denote the product measure on \( \Omega \) given by the Lebesgue measure on \( (0, 1] \). For 
\[ \omega = (\omega_1, \omega_2, \ldots) \in \Omega, \quad \omega_n \in (0, 1], \]
define 
\[ X^0_n(\omega) = X'_n(\omega_n) \quad \text{and} \quad Y^0_n(\omega) = Y'_n(\omega_n). \]
Clearly \( \{X^0_n\} \) and \( \{Y^0_n\} \) are two sequences of independent random variables such that for each \( n \), \( X^0_n \) and \( Y^0_n \) have the same distributions as those of \( X_n \) and \( Y_n \), respectively. From (3) and (4), we have 
\[ \sum_n \mathbb{P}^0(X^0_n \neq Y^0_n) < \infty. \]

This completes the proof.
The probabilistic part of the proof of the next theorem is extracted as the following
Lemma 2. Let \( r \geq 2 \) be an integer. Let \( \{a_{sp}: p \text{ is prime, } 1 \leq s \leq r\} \) be a sequence of real numbers satisfying

\[
\sum_{s=1}^{r} \sum_{p \leq N} \frac{1}{p} a_{sp} = O(N^{-d})
\]

for some \( d > 0 \). Let \( \{X_{sp}: p \text{ is prime, } 1 \leq s \leq r\} \) be independent random variables satisfying

\[
P(X_{sp} = 1) = 1 - P(X_{sp} = 0) = p^{-1}.
\]

Then

\[
\sum_{s=1}^{r} \sum_{p} a_{sp}(X_{sp} - p^{-1})
\]

converges almost surely to a random variable \( X \) and the distribution of \( X \) is singular with respect to the Lebesgue measure on the line.

Proof. The proof is adopted from (Erdős, 1939). Without loss of generality, we assume that \( 0 < d < 1 \). Let \( a = d/(4r) \) and \( b = d/3 \). Define

\[
A_N = \sum_{s=1}^{r} \sum_{p \leq N} \frac{1}{p} a_{sp},
\]

\[
B_N = \{m \leq N^a: m \text{ is square-free}\},
\]

\[
C_N = \left\{ \sum_{s=1}^{r} \sum_{p|m_s} a_{sp} - A_N: m_s \in B_N \text{ for } s = 1, \ldots, r \right\}.
\]

Let \( \bigcup' \) and \( \sum' \) denote the union and sum over all

\[
\{(m_1, \ldots, m_r): m_s \in B_N, s = 1, \ldots, r\}.
\]

Note that

\[
P\left( \left( \sum_{s=1}^{r} \sum_{p \leq N} a_{sp}X_{sp} - A_N \right) \in C_N \right)
\]

\[
\geq P\left( \bigcup' \left( \left( \sum_{s=1}^{r} \sum_{p \leq N} a_{sp}X_{sp} - A_N \right) \in C_N, \right.ight.
\]

\[
X_{sp} = 1 \text{ for } p \mid m_s \text{ and } X_{sp} = 0 \text{ for } (p, m_s) = 1,
\]

\[
\text{for all } s = 1, \ldots, r \text{ and } p \leq N\right)\)
\]

\[
= P\left( \bigcup' (X_{sp} = 1 \text{ for } p \mid m_s \text{ and } X_{sp} = 0 \text{ for } (p, m_s) = 1; \right.ight.
\]

\[
\left. \sum' \left( \sum_{p \leq N} a_{sp}X_{sp} - A_N \right) \in C_N \right)\)
\]

\[
= P\left( \bigcup' (X_{sp} = 1 \text{ for } p \mid m_s \text{ and } X_{sp} = 0 \text{ for } (p, m_s) = 1; \right.ight.
\]

\[
\left. \sum' \left( \sum_{p \leq N} a_{sp}X_{sp} - A_N \right) \in C_N \right).
\]
for all $s = 1, \ldots, r$ and $p \leq N$)

\[
\sum_p P(X_{sp} = 1 \mid m_s \text{ and } X_{sp} = 0 \text{ for } (p, m_s) = 1; \\
\text{for all } s = 1, \ldots, r \text{ and } p \leq N)
\]

\[
= \prod_{p \leq N} \left[ \left(1 - \frac{1}{p}\right) \left(\sum_{m \in B_N} \frac{1}{m} \prod_{p \mid m} \left(1 - \frac{1}{p}\right)^{-1}\right) \right]^r
\]

\[
\geq \prod_{p \leq N} \left[ \left(1 - \frac{1}{p}\right) \sum_{m \in B_N} \frac{1}{m} \right]^r \geq \delta
\]

for some $\delta > 0$ and for all large $N$. The second equality above follows from the fact that the events considered are disjoint. The last inequality follows from

\[
(a \log N)^{-1} \sum_{m \in B_N} \frac{1}{m} \to \prod_p \left(1 - \frac{1}{p}\right) > 0.
\]

Further, observe that by Chebyshev's inequality,

\[
P\left(\left| \sum_{s=1}^{r} \sum_{p > N} a_{sp} \left(X_{sp} - \frac{1}{p}\right) \right| > N^{-b}\right)
\]

\[
\leq N^{2b} \sum_{s=1}^{r} \sum_{p > N} \frac{1}{p} a_{sp}^2 = O(N^{2b-d}) = O(N^{-b}).
\]

This leads to

\[
P(X \in G_N) \geq \delta - O(N^{-b})
\]

for all large $N$, where

\[
G_N = \bigcup_{x \in C_N} \{x - N^{-b}, x + N^{-b}\}.
\]

On the other hand the Lebesgue measure of $G_N$ is bounded by

\[
2N^{-b} N^{ar} \leq 2N^{-d/12} \to 0,
\]

as $N \to \infty$. It follows that the distribution of $X$ is singular. This completes the proof.

In addition to the conditions of Lemma 2, if

\[
\sum_p \frac{1}{p} a_{sp}
\]

converges for $1 \leq s \leq r$, then

\[
\sum_{s=1}^{r} \sum_p a_{sp} X_p
\]
converges almost surely and it has a singular distribution.

**Theorem 3.** Let \( g \) be an additive arithmetic function on the set of pairs of positive integers having a distribution \( F \). If

\[
\sum_{p > N} \frac{1}{p} \left[ (g(1,p))^2 + (g(p,1))^2 \right] = O(N^{-d})
\]

for some \( d > 0 \) then \( g \) has a singular distribution.

**Proof.** Let \( X_{1p} \) and \( X_{2p} \) be as in Lemma 2. From Theorem 2 of (Babu, 1976), \( g \) has a distribution if and only if

\[
\sum_p (g(p,1)X_{1p} + g(1,p)X_{2p})
\]

converges almost surely. Now the result follows from Lemmas 1 and 2. Theorem 3 extends the results of (Babu, 1973) to arithmetic functions on the set of pairs of positive integers.

**Corollary.** If an additive function on the set of pairs of positive integers is bounded, then it has a singular distribution.

The derivation of the Corollary from Theorem 3 is similar to the proof of Corollary 2 of (Babu, 1973).

We now state a result on the distributions of multiplicative functions on the set of pairs of positive integers. This can be proved following the lines of proof of Theorems 1 and 3 of (Babu, 1980).

**Theorem 4.** Let \( g \) be a real-valued multiplicative function on the set of pairs of positive integers. Let \( g_1 \) and \( g_2 \) be the multiplicative functions given by \( g_1(m) = g(m,1) \) and \( g_2(n) = g(1,n) \). Then \( g \) has a distribution if and only if both \( g_1 \) and \( g_2 \) have distributions. Further, the distribution of \( g \) is concentrated at zero if and only if either the distribution of \( g_1 \) or \( g_2 \) is concentrated at zero. This occurs if and only if

\[
\sum_{p \in Q} \frac{1}{p} = \infty,
\]

where \( Q = \{ p : g(p,1)g(1,p) = 0 \} \).

The conditional distribution of \( g \) given \( g \neq 0 \) is of pure type. It is absolutely continuous if and only if the conditional distribution of \( g_1(m)g_2(n) \) given \( g_1(m)g_2(n) \neq 0 \) is absolutely continuous.

In addition, if

\[
\sum_{g(p,1) < 0} \frac{1}{p} + \sum_{g(1,p) < 0} \frac{1}{p} = \infty
\]

then \( g \) has a symmetric distribution.

Proof of a special case of this result, using characteristic transforms, was given in (Babu, 1975b). We only mention that, in proving that \( g \) has a symmetric
distribution under (6), we use the following lemma, instead of Lemma 4 of (Babu, 1980).

**Lemma 3.** For $k > 1$, let $B_k$ consist of all pairs $(i, j)$ of positive integers such that $(ij, p) = 1$ for all $p > k$, and $A_k$ consist of all pairs $(m, n)$ of positive integers such that $(mn, p) = 1$ for all $p \leq k$. Let

$$L_k \subseteq A_k, \quad M_k \subseteq B_k,$$

$$R_k = \{(im, jn): (i, j) \in B_k, (m, n) \in L_k\}$$

and

$$S_k = \{(im, jn): (i, j) \in M_k, (m, n) \in A_k\}.$$ 

Then $S_k$ has a density. If $R_k$ has density, then the density of $R_k \cap S_k$ exists and is given by the product of the densities of $R_k$ and $S_k$. That is $R_k$ and $S_k$ are independent."

Proof is similar to that of Lemma 4 of (Babu, 1980).

**REFERENCES**


