SOME RESULTS AND PROBLEMS IN PROBABILISTIC NUMBER THEORY

Imre Kátaí and Bui Minh Phong (Budapest, Hungary)

Le Manh Thanh (Hue, Vietnam)

Communicated by László Germán

(Received August 5, 2013; accepted February 1, 2014)

Abstract. We formulate some results, open problems and conjectures in probabilistic number theory.

1. Introduction

Notation. Let, as usual, \( \mathcal{P}, \mathbb{N}, \mathbb{N}_0, \mathbb{Z}, \mathbb{R}, \mathbb{C} \) be the set of primes, positive integers, non-negative integers, integers, real and complex numbers, respectively. We say that \( f : \mathbb{N} \rightarrow \mathbb{R} \) is an additive function, if \( f(1) = 0 \) and \( f(mn) = f(m) + f(n) \) for all \( (m,n) = 1 \). Let \( \mathcal{A} \) denote set of all additive functions. A function \( g : \mathbb{N} \rightarrow \mathbb{C} \) is multiplicative, if \( g(1) = 1 \) and \( g(mn) = g(m) \cdot g(n) \) for all \( (m,n) = 1 \). We denote by \( \mathcal{M} \) the set of all multiplicative functions.

Key words and phrases: multiplicative functions, additive functions, continuous homomorphism.

2010 Mathematics Subject Classification: 11A07, 11A25, 11N25, 11N64.

This work was completed with the support of the Hungarian and Vietnamese TET (grant agreement no. TET 10-1-2011-0645).
Let \( q \geq 2 \), integer, \( A_q = \{0, \ldots, q - 1\} \). Every \( n \in \mathbb{N}_0 \) can be written as
\[
n = \sum_{j=0}^{k} \varepsilon_j(n)q^j, \quad \varepsilon_k(n) \neq 0, \quad \varepsilon_j(n) \in A_q,
\]
and this expansion is unique.

In the following let \( A_q \) denote the set of \( q \)-additive functions, i.e. a function \( f : \mathbb{N}_0 \to \mathbb{R} \) belongs to \( A_q \), if \( f(0) = 0 \), and \( f(n) = \sum_{j=0}^{k} f(\varepsilon_j(n)q^j) \quad (\forall n \in \mathbb{N}_0) \).

Similarly, we say that \( g : \mathbb{N}_0 \to \mathbb{C} \) is \( q \)-multiplicative, if \( g(0) = 1 \), and \( g(n) = \prod_{j=0}^{k} g(\varepsilon_j(n)q^j) \quad (\forall n \in \mathbb{N}) \). Let \( M_q \) be the set of \( q \)-multiplicative functions.

Let \( \pi(x) \) be the number of the primes up to \( x \), and \( \pi(x, k, l) \) be the number of the primes \( p \leq x \) satisfying \( p \equiv l \pmod{k} \).

**Examples.**
- \( \omega(n) = \) number of prime factors of \( n \), \( \omega(n) \in A \),
- \( \tau(n) = \) number of divisors of \( n \), \( \tau(n) \in M \),
- \( \text{lcm } f(n) = \log n \in A \),
- \( \text{lcm } \varphi(n) = \) Euler's totient function, \( \varphi(n) \in M \),
- \( \text{lcm } \sigma(n) = \) sum of divisors of \( n \), \( \sigma(n) \in M \).
- If \( f(n) \in A \), \( g(n) = z^{f(n)}, \quad z \in \mathbb{C} \), then \( g \in M \).
- If \( \alpha(n) = \sum_{j=0}^{k} \varepsilon_j(n) = \) sum of digits functions, then \( \alpha(n) \in A_q \).
- \( f(n) = n \in A_q \),
- \( g(n) = z^n \in M_q \).

The following wellknown results can be found in [8].

**Definition 1.** We say that \( f \in A \) has a limit distribution (on the set \( \mathbb{N} \)) if
\[
F_N(y) := \frac{1}{N} \#\{n \leq N \mid f(n) < y\}
\]
has a limit \( \lim_{N \to \infty} F_N(y) = F(y) \) for almost all \( y \) (or in all continuity points of \( F \)).
Theorem 1 (Erdős-Wintner). \( f \in \mathcal{A} \) has a limit distribution if and only if the next three series converge

\[(1.1) \quad \sum_{|f(p)| > 1} \frac{1}{p}, \]
\[(1.2) \quad \sum_{|f(p)| \leq 1} \frac{f(p)}{p}, \]
\[(1.3) \quad \sum_{|f(p)| \leq 1} \frac{f^2(p)}{p}. \]

Theorem 2. \( f \in \mathcal{A} \) has a limit distribution with a suitable centralization \( c(N) \), i.e.

\[(1.4) \quad \lim_{N \to \infty} \frac{1}{N} \# \{n \leq N \mid f(n) - c(N) < y\} = F(y) \]

exists for almost all \( y \), if and only if (1.1), (1.3) converge.

Let \( a(N) := \sum_{p \leq N} \frac{f(p)}{p} \). If (1.4) holds, then \( \lim(a(N) - c(N)) = \alpha \) exists, and \( \alpha \) is finite.

Theorem 3 (Erdős-Kac). Let \( f \in \mathcal{A} \),

\[(1.5) \quad A_N := \sum_{p \leq N} \frac{f(p)}{p}, \quad B_N^2 = \sum_{p \leq N} \frac{f^2(p)}{p^\alpha}. \]

Assume that \( f(p) = O(1) \quad (p \in \mathcal{P}) \). Then

\[ \lim_{N \to \infty} \frac{1}{N} \# \left\{ n \leq N \mid \frac{f(n) - A(N)}{B_N} < y \right\} = \Phi(y), \]
\[ \Phi = \text{Gaussian law.} \]

Turán-Kubilius inequality. Let \( f \in \mathcal{A} \). Then

\[ \sum_{n \leq N} (f(n) - A_N)^2 \leq cN B_N^2, \]

\( A_N, B_N \) are defined in (1.5).
2. Distribution of additive functions on some subsets of integers

Let $B \subseteq \mathbb{N}$, $B(x) = \# \{ n \leq x, n \in B \}$.

We say that $f \in A$ is distributed in limit with centralization $a_N$ and normalization $b_N$ on the set $B$, if

$$
\lim_{N \to \infty} \frac{1}{B(N)} \# \left\{ n \leq N, n \in B \left| \frac{f(n) - a_N}{b_N} < y \right. \right\} = F(y)
$$

exists a.a.

Results:

1. $B = \mathcal{P}_{+1} = \{ p+1 \mid p \in \mathcal{P} \}$ = set of shifted primes, $a_N = 0$, $b_N = 0$. $f \in A$ has a limit distribution on $\mathcal{P}_{+1}$ if the 3 series (in Erdős-Wintner theorem) are convergent ([6]).

The proof is based on the method of characteristic functions: $g(n) = e^{i \pi f(n)}$.

(2.1) \[
\lim_{x \to \infty} \frac{1}{\pi(x)} \sum_{p < x} g(p+1) = M_r(x) = \prod_p \left( 1 - \frac{1}{p-1} + \frac{g(p)}{p} + \frac{g(p^2)}{p^2} + \cdots \right).
\]

To prove (2.1), the Siegel-Walfisz theorem was used, namely that

$$
\pi(x, k, l) = \sum_{\substack{p \leq x \atop p \equiv l \ (\mod k)}} 1 = li_x \frac{x}{\phi(k)} \left( 1 + O \left( e^{-c \sqrt{\log x}} \right) \right)
$$

uniformly as $(l, k) = 1$, $k < (\log x)^{c_1}$, where

$$
li_x = \int_2^x \frac{du}{\log u},
$$

and the A. I. Vinogradov - E. Bombieri inequality, according to it

$$
\sum_{\substack{k \leq \log x \atop (l,k)=1}} \max_{y \leq x} \left| \pi(y, k, l) - \frac{ly}{\phi(k)} \right| \leq C \frac{x}{(\log x)^{B_r}},
$$
$B = 2A + 5, \ C$ is a suitable (non-effective) constant.

J. Kubilius and P. Erdős asked on the necessity of the convergence of the 3 series (1.1), (1.2), (1.3). Partial results were obtained by P.D.T.A. Elliott, I. Kátai, and N. Timofeev. Finally more than 20 years later it was proved by A. Hidebrand that they are necessary ([4]).

2. L. Germán proved the analog of Erdős-Wintner theorem on the set

$$B = \{ n + 1 \mid \omega(n) \leq \varepsilon(n) \sqrt{\log \log n} \},$$

where $\varepsilon(n) \to 0$. (See [2].)

3. **On $q$-additive functions**

Let $f \in A_q$, $\xi_j$ be independent random variables,

$$P(\xi_j = f(aq^j)) = \frac{1}{q} \quad (a \in A_q),$$

$$\eta_N = \xi_0 + \xi_1 + \ldots + \xi_{N-1},$$

$$m_j = E(\xi_j) = \frac{1}{q} \sum_{a \in A_q} f(aq^j), \quad \sigma_j^2 = \frac{1}{q} \sum_{a \in A_q} f^2(aq^j) - m_j^2.$$  

By using the standard method of probability theory one can prove

**Theorem 4.** Let $f \in A_q$. Then

$$\lim_{x \to \infty} \frac{1}{x} \# \{ n < x \mid f(n) < y \} = F(y)$$

for almost all $y$ exists, if and only if (3.1), (3.2) are convergent:

(3.1) \[ \sum_{j=0}^{\infty} \sum_{a \in A_q} f(aq^j), \]

(3.2) \[ \sum_{j=0}^{\infty} \sum_{a \in A_q} f^2(aq^j). \]
What is happening if we consider the distribution of $f$ over a subset $B$ in $N_0$?

Let $1 \leq j_1 < \ldots < j_h < N$, $b_1, \ldots, b_h \in A_q$,

$$B_N \left( \frac{j_1, \ldots, j_h}{b_1, \ldots, b_h} \right) = \{m < q^N \mid m \in B, \ e_{j_l}(m) = b_l, l = 1, \ldots, h\},$$

$$B_N \left( \frac{j_1, \ldots, j_h}{b_1, \ldots, b_h} \right) = \# \{m < q^N \mid m \in B, \ e_{j_l}(m) = b_l, l = 1, \ldots, h\},$$

$$B(x) = \# \{m < x \mid m \in B\}.$$

Let $P(u) \in \mathbb{Z}[u]$ be a polynomial $r = \deg P$, $P(u) = a_ru^r + \ldots + a_1u + a_0$, $a_r > 0$.

Let $\Sigma_1 := A \left( x \left| \frac{l_1, \ldots, l_h}{b_1, \ldots, b_h} \right\} = \#\{n \leq x \mid e_{l_j}(P(n)) = b_j, j = 1, \ldots, h\},$

$$\Sigma_2 := \prod \left( x \left| \frac{l_1, \ldots, l_h}{b_1, \ldots, b_h} \right\} = \#\{p \leq x \mid e_{l_j}(P(p)) = b_j, j = 1, \ldots, h\}.$$

**Lemma 1.** Let $q^N \leq x < q^{N+1}$. Let $h$ be fixed, $\lambda$ be an arbitrary constant, $N^\frac{1}{4} \leq l_1 < \ldots < l_h \leq rN - N^\frac{1}{4}$.

Then

$$\Sigma_1 = \frac{x}{q^h} + O \left( \frac{x}{(\log x)^\lambda} \right),$$

$$\Sigma_2 = \frac{\pi(x)}{q^h} + O \left( \frac{x}{(\log x)^\lambda} \right)$$

uniformly as $l_1, \ldots, l_h$ in (3.3), $b_1, \ldots, b_h \in A_q$.

The proof depends on the next Lemmas 2 and 3. Let $e(u) := e^{2\pi i u}$.

**Lemma 2** (Hua Loo Keng). Let $0 < Q \leq c_1(k)(\log x)^{r_1}$ and

$$S = \sum_{p \leq x \atop p \equiv a \pmod{Q}} e(f(p))$$

in which

$$f(y) = \frac{h}{q}y^k + \alpha_1y^{k-1} + \cdots + \alpha_k, \ (h, q) = 1.$$
Suppose that \((\log x)^{\tau} \leq q \leq x^{k}(\log x)^{-\tau}\). For arbitrary \(\tau_0 > 0\), when \(\tau > 2^{6k}(\tau_0 + \tau_1 + 1)\), we always have
\[
|S| \leq c_2(k)x \cdot (\log x)^{-\tau_0} \cdot Q^{-1}.
\]
c_2(k) depends only on \(k\).

**Lemma 3** (I. M. Vinogradov). Let \(f\) be as in Lemma 2, \(S_1 = \sum_{n \leq x} e(f(n))\).

Let \(\tau_0, \tau_3, \tau_4\) be arbitrary positive numbers,
\[
\tau \geq 2^k(\tau_0 + \tau_3) + 2k\tau_4 + 2^{3(k-2)}.
\]
Suppose
\[
(\log x)^{\tau} < q \leq x^{k}(\log x)^{-\tau}.
\]
Then
\[
S_1 \ll x(\log x)^{-\tau}.
\]
The constant standing implicitly in \(\ll\) may depend on \(\tau_3, \tau\).

**Theorem 5.** ([1]) Let \(f \in A_q\), \(f(bq^j) = O(1)\) as \(j \to \infty\), \(b \in E\). Furthermore let \(\frac{D(x)}{(\log x)^{\frac{r}{2}}} \to \infty\). Here
\[
D^2(x) = \sum_{j=0}^{N} \sigma_j^2 \quad (q^N \leq x < q^{N+1}).
\]
Let \(P(u) \in \mathbb{Z}[u]\), \(\deg P = r\), \(P(u) \to \infty\) \((u \to \infty)\). Then
\[
\lim_{x \to \infty} \frac{1}{x} \# \left\{ n < x \ \mid \frac{f(P(n)) - M(x^r)}{D(x^r)} < y \right\} = \Phi(y),
\]
\[
\lim_{x \to \infty} \frac{1}{\pi(x)} \# \left\{ n < x \ \mid \frac{f(P(p)) - M(x^r)}{D(x^r)} < y \right\} = \Phi(y),
\]
\[
c_k(x) = \frac{1}{x^r} \sum_{n \leq x^r} \left( \frac{f_1(n) - M_1(x^r)}{D_1(x^r)} \right)^k.
\]
From Lemma 1 one can deduce that
\[
\lim a_k(x) = \lim b_k(x) = \lim c_k(x).
\]
It is clear that (3.4) is true if $P(u) = u$, furthermore

$$\lim_{x \to \infty} e_k(x) = \int u^k d\phi(u) = \mu_k.$$ 

Consequently $\lim a_k(x) = \mu_k$, $\lim b_k(x) = \mu_k$, and the Frechet-Shohat theorem implies the fulfilment of (3.4) and (3.5).

How to prove Lemma 1?

Let

$$\varphi_{b_k}(x) = \begin{cases} 1 & \text{if } x \in \left[ \frac{b}{q}, \frac{b+1}{q} \right), \\ 0 & \text{otherwise in } [0,1). \end{cases}$$

Let $\varphi_b(x + n) = \varphi_b(x)$ ($n \in \mathbb{Z}$).

Let $b_1, \ldots, b_h \in A_q$, $(1 \leq l_1 < \ldots < l_h)$ be arbitrary integers. Then

$$F(x_1, \ldots, x_h) = \varphi_{b_1}(x_1) \ldots \varphi_{b_h}(x_h),$$

$$\varepsilon_j(n) = b \iff \left( \frac{n}{q^{j+1}} \right) \in \left[ b - \frac{b+1}{q}, \frac{b}{q} \right).$$

Let $t(y) = F \left( \frac{y}{q^{l_1+1}}, \ldots, \frac{y}{q^{l_h+1}} \right)$.

Then

$$t(m) = \begin{cases} 1 & \text{if } \varepsilon_j(m) = b_j, j = 1, \ldots, h, \\ 0 & \text{otherwise.} \end{cases}$$

Let $0 < \Delta < \frac{1}{2q}$, $f_b(x) = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} \varphi_b(x + u) du,$

$$\tilde{t}(y) = f_{b_1} \left( \frac{y}{q^{l_1+1}} \right) \ldots f_{b_h} \left( \frac{y}{q^{l_h+1}} \right).$$

We can prove that

$$\sum_{n \leq x} |t(P(n)) - \tilde{t}(P(n))|$$

is small, furthermore

$$\tilde{t}(y) = \sum_M T_M e(MVy),$$

$$M = [m_1, \ldots, m_h], \quad V = \left[ \frac{1}{q^{l_1+1}}, \ldots, \frac{1}{q^{l_h+1}} \right],$$

$$\sum |T_M| < \infty.$$
Thus
\[ \sum \tilde{t}(P(n)) = \sum_{n \leq x} T_M \sum_{u \leq x} e \left( \frac{A_M}{H_M} P(n) \right), \]
\[ \sum \tilde{t}(P(p)) = \sum_{p \leq x} T_M \sum_{u \leq x} e \left( \frac{A_M}{H_M} P(p) \right), \]

\[ VM = \frac{A_M}{H_M}, \quad (H_M, A_M) = 1. \]

We can apply Lemma 1 and 2. This completes the proof.

**Theorem 6.** Let \( f \in A_q \). Assume that \( \lim M(x) \) exists and is finite, \( \lim D^2(x) < \infty \). Then there exist suitable distribution functions \( F_1(y), F_2(y) \) such that

\[ \lim \frac{1}{x} \# \left\{ n < x \left| f(P(n)) < y \right. \right\} = F_1(y) \quad a.a. \ y, \quad (3.6) \]

\[ \lim \frac{1}{\pi(x)} \# \left\{ p < x \left| f(P(p)) < y \right. \right\} = F_2(y) \quad a.a. \ y. \quad (3.7) \]

**Theorem 7.** If \( P(n) = n \), and (3.7) holds true, then \( \lim M(x) \) exists and is finite, furthermore \( \lim_{x \to \infty} D(x) \) is finite.

4. Linear combinations of \( q \)-additive functions

Here we mention some theorems without proof.

In this section \( \lambda \) is the Lebesgue measure defined in \( \mathbb{R} \).

Let
\[ f_1, \ldots, f_k \in A_q; \quad (1 \leq a_1 < \ldots < a_k < q), \quad (a_i, q) = 1, \quad (a_i, a_j) = 1 \quad \text{if} \quad i \neq j, \]
\[ l(n) = f_1(a_1 n) + \ldots + f_k(a_k n). \]

**Definition 2.** We say that \( l(n) \) is a tight sequence if \( \exists A_N \) such that
\[ C(K) := \limsup \frac{1}{q^N} \# \{ n < q^N \left| |l(n) - A_N| > K \right. \} \to 0 \]
as \( K \to \infty \).

**Theorem 8.** (K.-H. Indlekofer and I. Kátaï, [5])
1. \( l(n) \) is tight if and only if \( \exists \, \gamma_1, \ldots, \gamma_k \) for which \( a_1\gamma_1 + \ldots + a_k\gamma_k = 0 \) and for \( \psi_l(n) := f_l(n) - \gamma_l(n) \) we have

\[
\sum_{j=0}^{\infty} \sum_{a=0}^{q-1} \psi_l(aq^j)^2 < \infty \quad (l = 1, \ldots, k).
\]

\((4.1)_l\)

2. If the conditions of 1) hold, then

\[
\lim_{N \to \infty} \frac{1}{q^N} \#\{n < q^N \mid l(n) - E_N < y\} = F(y) \quad \text{a.a.} \, y,
\]

where \( F \) is a suitable distribution function,

\[
A_N^{(l)} := \frac{1}{q} \sum_{j=0}^{N} \sum_{a\in A_q} \psi_l(aq^j),
\]

\[
E_N = \sum_{l=1}^{k} A_N^{(l)}.
\]

3. \( l(n) \) has a limit distribution, if and only if the conditions of 1) are satisfied, and if \( E_N \) has a finite limit.

**Theorem 9.** 1. The sequence \( l(p) \) \( (p \in \mathcal{P}) \) is tight if it is tight on the whole set \( \mathbb{N} \).

2. If \((4.1)_l\) hold, then

\[
\lim_{N \to \infty} \frac{1}{q^N} \pi(q^N) \#\{p < q^N \mid l(p) - E_N < y\} = F^*(y) \quad \text{a.a.} \, y.
\]

3. \( l(p) \) has a limit distribution if and only if \( l(p) \) is tight and if \( E_N \) has a finite limit.

Let \( \mu_l(u) = \frac{1}{q} \sum_{c\in A_q} f_l(cq^u) \), \( p(u) = \sum_{l=1}^{k} \mu_l(u) \),

\[
F(N) := \sum_{u=0}^{N-1} p(u),
\]

\[
\pi_u(c_1, \ldots, c_k) := \frac{1}{q^{u+1}} \#\{n < q^{u+1} \mid \varepsilon_u(a_jn) = c_j\},
\]

\[
\mu_l(u) = \frac{1}{q} \sum_{c\in A_q} f_l(cq^u).
\]
\[ \pi(c_1, \ldots, c_k) = \lambda \left( \left\{ x \in [0,1] \mid \{a_jx\} \in \left[\frac{c_j}{q}, \frac{c_j + 1}{q}\right], j = 1, \ldots, k \right\} \right). \]

We have \( \pi_u(c_1, \ldots, c_k) = \pi(c_1, \ldots, c_k) + \mathcal{O}\left(\frac{1}{q}\right) \) \( (u \to \infty) \).

Let
\[
\tau_u := \sum_{c_1, \ldots, c_k \in A_q} (f_1(c_1q^u) + \ldots + f_k(c_kq^u) - p(u))^2 \pi_u(c_1, \ldots, c_k),
\]
and
\[ \varrho^2_N = \sum_{u=0}^{N-1} \tau_u. \]

**Theorem 10.** Assume that
\[ \frac{|f_l(cq^M)|}{\varrho^2_M} \to 0 \quad \text{as} \quad M \to \infty \]
for \( l = 1, \ldots, k \). Then
\[ \lim_{N \to \infty} \frac{1}{q^N} \# \left\{ n < q^N \left| \frac{l(n) - F_N}{\varrho_N} < y \right. \right\} = \phi(y). \]

**Theorem 11.** Assume that
\[ \sup_M \max_{c \in A_q} \max_l |f_l(cq^M)| < \infty, \]
and that
\[ \varrho^2_{K_N}/\varrho^2_N \to 0, \quad (\varrho^2_N - \varrho^2_{N-K_N})/\varrho^2_N \to 0 \]
as \( K_N = \lceil \log N \rceil, \quad N \to \infty \). Then
\[ \lim_{N \to \infty} \frac{1}{\pi(q^N)} \# \left\{ p < q^N \left| \frac{l(p) - F_N}{\varrho_N} < y \right. \right\} = \phi(y). \]

5. **On a theorem of G. Harman and I. Kátai ([3])**

Let \( 0 \leq j_1 < j_2 < \ldots < j_r \leq N - 1, \quad x > \exp(q^2), \quad q^{N-1} \leq x \leq q^N - 1 \). Let \( b = (b_1, \ldots, b_r), \quad j = (j_1, \ldots, j_r) \),
\[
\prod \left( x \mid \frac{j}{b} \right) := \#\{ p \leq x \mid \epsilon_{l_j}(p) = b_j, j = 1, \ldots, r \},
\]
and
\[
A \left( x \mid \frac{j}{b} \right) := \#\{ n \leq x \mid \epsilon_{l_j}(n) = b_j, j = 1, \ldots, r \}.\]
Theorem 12. Suppose that $1 \leq r < C\sqrt{N} / \log N$. Then

$$\prod \left( x \left| \frac{j}{b} \right. \right) = q^r f(b,j) \log x A \left( x \left| \frac{j}{b} \right. \right) + O \left( \frac{x \log \log x}{\phi(q) q^{r-1} (\log x)^2} \right),$$

where

$$f(b,j) = \begin{cases} 
q^{-r} & \text{if } j_1 > 0, \\
0 & \text{if } j_1 = 0, (b_1, q) > 1, \\
q^{1-r} \phi(q)^{-1} & \text{if } j_1 = 0, (b_1, q) = 1.
\end{cases}$$

A similar theorem is claimed in the paper ([7]) of I. Kátai, but there were some mistakes in the proof.

Our guess: Theorem 12 remains valid up to $r < \frac{1}{3} N$.

References


Some results and problems in probabilistic number theory


---

Imre Kátaí and Bui Minh Phong
Eötvös Loránd University
H-1117 Budapest
Pázmány Péter Sétány I/C,
Hungary
katai@compalg.inf.elte.hu
bui@compalg.inf.elte.hu

Le Manh Thanh
Hue University
Hue City
3 Le Loi
Vietnam
lmanh@hueuni.edu.vn