Annales Univ. Sci. Budapest., Sect. Comp. 14 (1994) 13-22

DISTRIBUTION OF q-ARY DIGITS IN SOME SEQUENCES OF INTEGERS

N. L. Bassily (Cairo, Egypt)

Dedicated to Professor Karl-Heinz Indlekofer on the occasion of his fiftieth birthday

1. Introduction

Let $\mathbb{N}$, $\mathbb{R}$ denote the set of natural, real numbers, resp. Let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For a real number x let $\{x\}$ be the fractional part of x. We shall write $e(x)$ instead of $e^{2\pi ix}$. $\sum_p$ denotes summation over the prime numbers, $\pi(x)$ is the number of primes up to x.

Let us fix an integer $q \geq 2$ and let $\mathbb{A} = \{0, 1, \ldots, q-1\}$. Then each $n \in \mathbb{N}_0$ can be uniquely written as

$$n = \sum_{r=0}^{\infty} a_r(n) q^r, \quad a_r(n) \in \mathbb{A} \quad (r = 0, 1, \ldots). \tag{1.1}$$

In fact, $a_r(n) = 0$ if $r > \dfrac{\log\ n}{\log\ q}$.

Let $k \in \mathbb{N}$, $F_k : A^k \to \mathbb{R}$ be a function satisfying the condition $F_k(0, \ldots, 0) = 0$, and let

$$\alpha(n) := \sum_{j=0}^{\infty} F_k(a_j(n), a_{j+1}(n), \ldots, a_{j+k-1}(n)). \tag{1.2}$$

$\alpha(n)$ is a generalization of the sum of digit function.

Let

$$M := \frac{1}{q^k} \sum_{\substack{b_\nu \in \mathbb{A} \\ \nu = 0, \ldots, k-1}} F_k(b_0, \ldots, b_{k-1}).$$

It is well-known that

$$\alpha(n) = (1 + o(1))M\frac{\log\ n}{\log\ q} \tag{1.3}$$

holds for $n \to \infty$, neglecting a set of integers having zero density. (1.3) can be obtained by the following purely probabilistic argument. Let $\Omega = \Omega_N = \{n \mid 0 \le n < q^N\}$, $\mathcal{A}_N$ the set of all subsets of Ω_N, and $P = P_N$ be the measure defined by $P_N(A) = \frac{1}{q^N}\sharp(A)$, $\forall A \in \mathcal{A}$. Then $a_j(n)$ $(j = 0, \ldots, N-1)$ are random variables with the distribution $P_N(a_j(n) = b) = \frac{1}{q}$, furthermore they are independent. Let

$$\alpha_1(n) := \sum_{j=0}^{N-k} F_k(a_j(n),\ a_{j+1}(n), \ldots,\ a_{j+k-1}(n)).$$

The mean value of $\Theta_j := E_k(a_j(n), \ldots, a_{j+k-1}(n)) - M$ is 0. Thus the variance of $\alpha_1(n) - M(N-k+1)$ can be written as

$$E((\alpha_1(n) - M(N-k+1))^2) = \sum_{j_1, j_2} E(\Theta_{j_1}\Theta_{j_2}).$$

Since Θ_{j_1} and Θ_{j_2} are independent if $\mid j_1 - j_2 \mid > k$, and $E(\Theta_j) = 0$, we obtain that the right hand side is less than $\ll N$. Since $\alpha(n) - \alpha_1(n)$ is bounded,

$$\frac{1}{q^N} \sum_{n<q^N} (\alpha(n) - M \cdot N)^2 \ll N$$

follows.

(1.3) is an immediate consequence of this inequality. Our purpose in this paper is to prove that (1.3) remains valid on subsequences of integers like polynomial values, or polynomial values at prime places.

Theorem. *Let $P(x) = c_r x^r + \ldots + c_0$ be a polynomial with integer coefficients taking on positive values for $x > 0$, such that $(c_r, q) = 1$. Then*

$$\sum_{n \le x} \left(\alpha(P(n)) - Mr\frac{\log\ x}{log\ q}\right)^2 \ll x \log\ x, \tag{1.4}$$

$$\sum_{p \le x} \left(\alpha(P(p)) - Mr\frac{\log\ x}{\log\ q}\right)^2 \ll \pi(x) \log\ x, \tag{1.5}$$

where the constant, implied by $\ll$ may depend on F_k, and p runs over the set of primes.

The proof is based upon the theorems of I.M. Vinogradov and L.K. Hua for trigonometric sums and on the theorem of Erdős-Turán for the discrepancy of sequences mod 1.

Acknowledgement. The author is indebted to I. Kátai for suggesting the problem.

2. Auxiliary results

2.1. The discrepancy D_N of the real numbers $x_1, \dots, x_N$ is defined by

$$\sup \left| \frac{1}{N} \sum_{\substack{n=1 \\ \{x_n\} \in [\alpha, \beta)}}^{N} 1 - (\beta - \alpha) \right| ,$$

where the supremum is taken for all interval $[\alpha, \beta) \subseteq [0,1]$. Let $s_N(h) :=$ $:= \frac{1}{N} \sum_{l=1}^{N} e(hx_l)$.

Lemma 1. (Erdős-Turán [1]) *We have*

$$D_N \leq c \left(\sum_{0<h\leq M} \frac{1}{h} \mid s_N(h) \mid \right) ,$$

$$D_N \leq c \left(\frac{1}{M} + \sum_{h=1}^{M} \frac{\mid s_N(h) \mid}{h} \right)$$

for any positive integer M. c is an absolute constant.

2.2. Let $f(x) = \frac{U}{V} x^r + \alpha_1 x^{r-1} + \ldots + \alpha_r$, $\alpha_j \in \mathbb{R}$ $(j = 1, \dots, r)$, U, V be coprime integers,

$$T := \sum_{p \leq X} e(f(p)), \ Y = \sum_{n \leq X} e(f(n)).$$

Lemma 2. (Hua [2], Theorem 10 and Lemma 6.2). *Assume that* $(\log X)^\sigma < V < X^r (\log X)^{-\sigma}$ *holds with some* $\sigma > 2^{6r}$. *Let* $\sigma_0 := \sigma \cdot 2^{-6r} - 1$. *Then*

$$| T | \le c(r) X (\log X)^{-\sigma_0}, \quad | Y | \le c(r) X (\log X)^{-\sigma_0},$$

where $c(r)$ *is a constant that depends only on* r.

2.3. Let $B = b_0 + b_1 q + \ldots + b_{k-1} q^{k-1}$, $b_j \in \mathbb{A}$, and $\varphi_B(x)$ be a periodic function mod 1, defined in [0,1) by

$$\varphi_B(x) = \begin{cases} 1 & \text{if } x \in \left(\frac{B}{q^k}, \frac{B+1}{q^k} \right), \\ \frac{1}{2} & \text{if } x = \frac{B}{q^k} \text{ or } x = \frac{B+1}{q^k}, \\ 0 & \text{otherwise.} \end{cases}$$

The Fourier-expansion of $\varphi_B(x)$ has the following form:

$$\varphi_B(x) = \sum_{m=-\infty}^{\infty} c_m(B) e(mx),$$

$$c_0(B) = \frac{1}{q^k}, \quad c_m(B) = -\frac{e(-\frac{mB}{q^k})}{2\pi i m} \left[e\left(-\frac{m}{q^k} \right) - 1 \right].$$

Let $0 < \Delta < \frac{1}{2q^k}$,

$$f_B(x) := \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} \varphi_B(x+z) dz = \sum_{m=-\infty}^{\infty} d_m(B) e(mx).$$

By an easy computation we have

(2.1) $$d_0(B) = \frac{1}{q^k}, \quad d_m(B) = c_m(B) \frac{e(\frac{m\Delta}{2}) - e(-\frac{m\Delta}{2})}{2\pi i m \Delta},$$

(2.2) $$d_m(B) = 0 \text{ if } m \equiv 0 \ (mod\ q^k) \text{ and } m \neq 0,$$

furthermore that

(2.3) $$| d_m(B) | \le \min \left(\frac{1}{\pi |m|}, \frac{1/\Delta}{\pi m^2} \right).$$

It is clear that $0 \le f_B(x) \le 1$ for every x and that

$$f_B(x) = \begin{cases} 1 & \text{if } x \in \left[\frac{B}{q^k} + \Delta,\ \frac{B+1}{q^k} - \Delta\right], \\ 0 & \text{if } x \in [0,1] \setminus \left[\frac{B}{q^k} - \Delta,\ \frac{B}{q^k} + \Delta\right]. \end{cases} \tag{2.4}$$

3. Proof of the theorem

3.1. If (1.4), (1.5) are valid for $F_k^{(1)}$, $F_k^{(2)}$, then they are also true for $F_k = d_1 F_k^{(1)} + d_2 F_2^{(k)}$, where d_1, d_2 are arbitrary constants. Consequently, it is enough to prove them for the special functions

$$F_k(e_0, \dots, e_{k-1}) = \begin{cases} 1 & \text{if } (e_0, \dots, e_{k-1}) = (b_0, \dots, b_{k-1}), \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

We assume that F_k is of form (3.1). Let $B = b_0 + b_1 q + \ldots + b_{k-1} q^{k-1}$.

3.2. Let $X \ge q$, $N = \left[\frac{\log X}{\log q}\right]$, d be such an integer for which $\max_{n<X} P(n) < q^{N+d}$. Let $t_1 = [\sqrt{N}]$, $t_2 = N - [\sqrt{N}]$. Let us write $P(n)$ in the form

$$P(n) = l_n^{(0)} + l_n^{(1)} q^{t_1} + l_n^{(2)} q^{t_2} \tag{3.3}$$

where $0 \le l_n < q^{t_1}$, $0 \le l_n^{(1)} < q^{t_2 - t_1}$, $0 \le l_n^{(2)} < q^{N+d-t_2}$. Since $\alpha(P(n)) =$ $= \alpha(l_n^{(0)}) + \alpha(l_n^{(1)}) + \alpha(l_n^{(2)}) + O(1)$, and $\alpha(l_n^{(0)})$, $\alpha(l_n^{(2)}) \ll \sqrt{N}$, therefore it is enough to prove that

$$E := \sum_{n \le X} \left(\alpha(l_n^{(1)}) - \frac{t_2 - t_1}{q^k} \right)^2 \ll X \log X, \tag{3.1}$$

$$F := \sum_{p \le X} \left(\alpha(l_p^{(1)}) - \frac{t_2 - t_1}{q^k} \right)^2 \ll \pi(X) \log X. \tag{3.2}$$

We shall prove only (3.2), the proof of (3.1) is almost the same.

3.3. Let $R_m(x)$ be the number of primes $p \le X$ satisfying $\Big(a_m(P(p)), \ldots, a_{m+k-1}(P(p))\Big) = (b_0, b_1, \ldots, b_{k-1})$, and $R_{m_1,m_2}(x)$ be the number of those primes $p \le X$ for which $\Big(a_{m_i}(P(p)), \ldots, a_{m_i+k-1}(P(p))\Big) = (b_0, \ldots, b_{k-1})$ ($i = 1, 2$) simultaneously hold.

It is clear that

$$A := \sum_{p \le X} \alpha(l_p^{(1)}) = \sum_{m=t_1}^{t_2-1} R_m(x), \tag{3.3}$$

$$C := \sum_{p \le X} \alpha^2(l_p^{(2)}) = \sum_{m=t_1}^{t_2-1} R_{m,m}(x) + 2 \cdot \sum_{t_1 \le m_1 < m_2 \le t_2 - t_1} R_{m_1,m_2}(x) \tag{3.4}$$

and that

$$F = C - 2A\frac{(t_2 - t_1)}{q^k} + \frac{(t_2 - t_1)^2}{q^{2k}}\pi(X). \tag{3.5}$$

3.4. A prime p is counted in $R_m(x)$ if and only if the q-ary expansion of $P(p)$ is of the form

$$P(p) = u + q^m B + v q^{m+k},$$

where $0 \le u < q^m$, $v \ge 0$, i.e. if

$$x_p := \left\{ \frac{P(p)}{q^{m+k}} \right\} \in \left[\frac{B}{q^k}, \frac{B+1}{q^k} \right).$$

Applying Lemma 1 to the sequence x_p ($p \le X$), we obtain

$$R_m(X) = \frac{\pi(X)}{q^k} + \Delta_m(X), \tag{3.6}$$

where

$$\Delta_m(X) \ll \frac{\pi(X)}{M} + \pi(X) \sum_{1 \le h \le M} \left| \sum_{p \le X} e\left(\frac{hP(p))}{q^{m+k}} \right) \right|.$$

Let $M = (\log\ X)^{10}$, and observe that the leading coefficient of the polynomial $\frac{hP(n)}{q^{m+k}}$ is a rational number $\frac{h\ c_r}{q^{m+k}}$ the reduced form $\frac{U}{V}$ of which satisfies

$$\exp\left(\frac{1}{2}\sqrt{N}\right) \le V \le X^r \exp\left(-\frac{1}{2}\sqrt{N}\right)$$

for every large X. Consequently the condition of Lemma 2 is satisfied with an arbitrary σ, and so with $\sigma = 12 \cdot 2^{6r}$, say. Hence $\Delta_m(X) \ll \pi(X) \cdot (\log\ X)^{-10}$, and

$$A = \frac{\pi(X)}{q^k}(t_2 - t_1) + O\left(\frac{\pi(X)(t_2 - t_1)}{(\log\ X)^{10}}\right). \tag{3.8}$$

3.5. Let us evaluate now $R_{m_1,m_2}(X)$ under the condition $t_1 \le m_1 < < m_1 + k \le m_2 \le t_2 - t_1$.

We start from the formula

$$R_{m_1,m_2}(X) = \sum_{p \le X} \varphi_B\left(\frac{P(p)}{q^{m_1+k}}\right) \varphi_B\left(\frac{P(p)}{q^{m_2+k}}\right),$$

where φ_B is defined in section (2.3). Then

$$R_{m_1,m_2}(X) = \sum_{p \le X} f_B\left(\frac{P(p)}{q^{m_1+k}}\right) f_B\left(\frac{P(p)}{q^{m_2+k}}\right) + O(\Sigma), \tag{3.9}$$

where Σ is the number of primes p for which at least one of $\left\{\frac{P(p)}{q^{m_j+k}}\right\}$ $(j = = 1, 2)$ belongs to the set

$$\left[\frac{B}{q^k} - \Delta,\ \frac{B}{q^k} + \Delta\right] \cup \left[\frac{B+1}{q^k} - \Delta,\ \frac{B+1}{q^k} + \Delta\right].$$

By using Lemmas 1,2 arguing as in 3.4 for the choice $\Delta = (\log\ X)^{-10}$ we obtain that

$$\sum = O\left(\pi(X)(\log\ X)^{-10}\right).$$

By using the Fourier-expansion of f_B, the sum on the right hand side of (3.9) can be written as

$$\sum_{h_1=-\infty}^{\infty} \sum_{h_2=-\infty}^{\infty} d_{h_1}(B) d_{h_2}(B) T_{h_1,h_2}, \tag{3.10}$$

where

$$T_{h_1,h_2} = \sum_{p\le X} e\left(\left(\frac{h_1}{q^{m_1+k}} + \frac{h_2}{q^{m_2+k}}\right) P(p)\right). \tag{3.11}$$

The leading coefficient of the polynomial in the exponent in (3.11) is

$$D_{h_1,h_2} = \left(\frac{h_1}{q^{m_1+k}} + \frac{h_2}{q^{m_2+k}}\right) c_r.$$

Let $D_{h_1,h_2} = \frac{c_r U}{V}$, $(c_r U, V) = 1$. If $q^k | h_j$ and $h_j \neq 0$, then $d_{h_j}(B) = 0$. Hence (3.10) equals to

$$d_0^2(B)T_{0,0} + \sum_{h_1\not\equiv 0 \pmod{q^k}} d_{h_1}(B)d_0(B)T_{h_1,0} + \sum_{h_2\not\equiv 0 \pmod{q^k}} d_0(B)d_{h_2}(B)T_{0,h_2} +$$

$$+ \sum_{h_1\not\equiv 0 \pmod{q^k}} \sum_{h_2\not\equiv 0 \pmod{q^k}} d_{h_1}(B)d_{h_2}(B)T_{h_1,h_2}. \tag{3.12}$$

Let the prime decomposition of $q = \pi_1^{e_1}\dots\pi_t^{e_t}$. If $h_1 q^{m_2-m_1} + h_2 \equiv$ $\equiv 0 \pmod{\pi_j^{e_j k}}$ holds for $j = 1,\dots,t$, then $q^k | h_2$. Thus for $h_2 \not\equiv 0 \pmod{q^k}$ we have $V \ge \min\left(\pi_1^{e_1(m_2-k)},\dots,\pi_t^{e_t(m_2-k)}\right)$, consequently $V \gg q^{\delta\sqrt{N}}$ with a suitable constant δ which depends only on q. Applying Lemma 2, hence we obtain that the last sum in (3.12) is less than

$$\ll \frac{\pi(X)}{(\log X)^{10}}\left(\sum_{h\not\equiv 0 \pmod{q^k}} | d_h(B) |\right)^2.$$

For $h_1 = 0$, $h_2 \not\equiv 0 \pmod{q^k}$ and for $h_2 = 0$, $h_1 \not\equiv 0 \pmod{q^k}$, similarly we have $T_{h_1,0}$, $T_{0,h_1} \ll \frac{\pi(X)}{(\log X)^{10}}$, consequently (3.12) equals to

$$\frac{\pi(X)}{q^{2k}} + O\left(\frac{\pi(X)}{(\log X)^{10}}\left(1 + 2\sum_{h\ge 1} | d_h(B) |\right)^2\right).$$

From (2.2) we have

$$\left(1+2\sum_{h\geq 1}\mid d_h(B)\mid\right)^2 \ll (\log\log X)^2.$$

Collecting our inequalities we obtain

$$R_{m_1,m_2}(X)=\frac{\pi(X)}{q^{2k}}+O\left(\frac{\pi(X)}{(\log X)^9}\right). \tag{3.13}$$

Since $R_{m_1,m_2}(X)\leq \pi(X)$, the contribution of the choices $m_1\leq m_2\leq$ $\leq m_1+k-1$ to C is $O((t_2-t_1)\pi(X))$, thus

$$C=\frac{(t_2-t_1)^2}{q^{2k}}\pi(X)+O(\pi(X)(t_2-t_1)).$$

Hence, from (3.5) and (3.8) we get immediately that

$$F\ll \pi(X)\log X.$$

This proves (3.2). By this the proof of (1.5) is completed.

4. Corollary

Let P be a polynomial satisfying the condition stated in Theorem. Then, for every fixed $\varepsilon>0$,

$$\frac{1}{X}\sharp\left\{n\leq X\left|(\alpha(P(n))-Mr\frac{\log X}{\log q}\right|>(\log X)^{1/2+\varepsilon}\right\}\to 0,$$

$$\frac{1}{\pi(X)}\sharp\left\{p\leq X\left|(\alpha(P(p))-Mr\frac{\log X}{\log q}\right|>(\log X)^{1/2+\varepsilon}\right\}\to 0.$$

References

[1] **Erdős P. and Turán P.**, On a problem in the theory of uniform distribution I, II., *Indigationes math.*, **10** (1948), 370-378, 406-413.

[2] **Hua L. K.**, *Additive theory of prime numbers,* Translations of Mathematical Momographs **13**, AMS Providence, Rhode Island, 1965.

N. L. Bassily
Department of Mathematics
Faculty of Science, Ain Shams University
Abbasia, Cairo, Egypt