Journal of Integer Sequences, Vol. 16 (2013),
Article 13.2.12

Construction of Normal Numbers via Generalized Prime Power Sequences

M. G. Madritsch and R. F. Tichy
Department for Analysis and Computational Number Theory
Graz University of Technology
A-8010 Graz
Austria
madritsch@math.tugraz.at
tichy@tugraz.at

Dedicated to Jean-Paul Allouche on the occasion of his 60th birthday

Abstract

We construct normal numbers in base q by concatenating q-adic expansions of prime powers $\lfloor \alpha p^{\theta} \rfloor$ with $\alpha > 0$ and $\theta > 1$.

1 Introduction

Let $q \geq 2$ be a fixed integer and $\sigma = 0.a_1a_2\ldots$ be the q-ary expansion of a real number σ with $0 < \sigma < 1$. We write $d_1 \cdots d_\ell \in \{0, 1, \ldots, q-1\}^\ell$ for a block of ℓ digits in the q-ary expansion. By $\mathcal{N}(\sigma; d_1 \cdots d_\ell; N)$ we denote the number of occurrences of the block $d_1 \cdots d_\ell$ in the first N digits of the q-ary expansion of σ. We call σ normal to the base q if for every fixed $\ell \geq 1$

$$\mathcal{R}_N(\sigma) = \mathcal{R}_{N,\ell}(\sigma) = \sup_{d_1 \cdots d_\ell} \left| \frac{1}{N} \mathcal{N}(\sigma; d_1 \cdots d_\ell; N) - \frac{1}{q^\ell} \right| = o(1)$$

as $N \to \infty$, where the supremum is taken over all blocks $d_1 \cdots d_\ell \in \{0, 1, \ldots, q-1\}^\ell$.

A slightly different, however equivalent definition of normal numbers is due to Borel [6] who also showed that almost all numbers are normal (with respect to the Lebesgue measure) to any base. However, despite their omnipresence among the reals, all numbers currently known to be normal are established by ad hoc constructions. In particular, we do not know whether given numbers, such as π, e, $\log 2$ and $\sqrt{2}$, are normal.

In this paper we consider the construction of normal numbers in base q as concatenation of q-ary integer parts of certain functions. A first result was achieved by Champernowne [8], who showed that

$$0.1\,2\,3\,4\,5\,6\,7\,8\,9\,10\,11\,12\,13\,14\,15\,16\,17\,18\,19\,20\dots$$

is normal in base 10. This construction can be easily generalized to any integer base q. Copeland and Erdös [9] proved that

$$0.2\,3\,5\,7\,11\,13\,17\,19\,23\,29\,31\,37\,41\,43\,47\,53\,59\,61\,67\dots$$

is normal in base 10.

This construction principle has been generalized in several directions. In particular, Dumont and Thomas [12] used transducers in order to rewrite the blocks of the expansion of a given normal number to produce another one. Such constructions using automata yield to q-automatic numbers, i.e., real numbers whose q-adic representation is a q-automatic sequence (cf. Allouche and Shallit [1]). By these means one can show that for instance the number

$$\sum_{n\geq 0} 3^{-2^n} 2^{-3^{2^n}}$$

is normal in base 2.

In the present paper we want to use another approach to generalize Champernowne's construction of normal numbers. In particular, let f be any function and let $[f(n)]_q$ denote the base q expansion of the integer part of $f(n)$. Then define

$$\sigma_q = \sigma_q(f) = 0.\left\lfloor f(1)\right\rfloor_q \left\lfloor f(2)\right\rfloor_q \left\lfloor f(3)\right\rfloor_q \left\lfloor f(4)\right\rfloor_q \left\lfloor f(5)\right\rfloor_q \left\lfloor f(6)\right\rfloor_q \dots, \tag{1}$$

where the arguments run through all positive integers. Champernowne's example corresponds to the choice $f(x) = x$ in (1). Davenport and Erdös [10] considered the case where $f(x)$ is an integer valued polynomial and showed that in this case the number $\sigma_q(f)$ is normal. This construction was subsequently extended to polynomials over the rationals and over the reals by Schiffer [23] and Nakai and Shiokawa [21], who were both able to show that $\mathcal{R}_N(\sigma_q(f)) = \mathcal{O}(1/\log N)$. This estimate is best possible, as proved by Schiffer [23]. Furthermore, Madritsch et al. [19] gave a construction for f being an entire function of bounded logarithmic order.

Nakai and Shiokawa [20] constructed a normal number by concatenating the integer part of a pseudo-polynomial sequence, i.e., a sequence $(\lfloor p(n)\rfloor)_{n\geq 1}$ where

$$p(x) = \alpha_0 x^{\theta_0} + \alpha_1 x^{\theta_1} + \dots + \alpha_d x^{\theta_d} \tag{2}$$

with $\alpha_0, \theta_0, \dots, \alpha_d, \theta_d \in \mathbb{R}$, $\alpha_0 > 0$, $\theta_0 > \theta_1 > \dots > \theta_d > 0$ and at least one $\theta_i \notin \mathbb{Z}$.

This method of construction by concatenating function values is in strong connection with properties of q-additive functions. We call a function f strictly q-additive, if $f(0) = 0$ and the function operates only on the digits of the q-adic representation, i.e.,

$$f(n) = \sum_{h=0}^{\ell} f(d_h) \quad \text{for} \quad n = \sum_{h=0}^{\ell} d_h q^h.$$

A very simple example of a strictly q-additive function is the sum of digits function s_q, defined by

$$s_q(n) = \sum_{h=0}^{\ell} d_h \quad \text{for} \quad n = \sum_{h=0}^{\ell} d_h q^h.$$

Refining the methods of Nakai and Shiokawa the first author obtained the following result.

Theorem 1 ([18, Theorem 1.1])**.** *Let $q \geq 2$ be an integer and f be a strictly q-additive function. If p is a pseudo-polynomial as defined in (2), then there exists $\varepsilon > 0$ such that*

$$\sum_{n \leq N} f\left(\lfloor p(n) \rfloor\right) = \mu_f N \log_q(p(N)) + NF\left(\log_q(p(N))\right) + \mathcal{O}\left(N^{1-\varepsilon}\right), \tag{3}$$

where

$$\mu_f = \frac{1}{q} \sum_{d=0}^{q-1} f(d)$$

and F is a 1-periodic function depending only on f and p.

The aim of the present paper is to extend the above results to prime power sequences. Let f be a function and set

$$\tau_q = \tau_q(f) = 0.\left\lfloor f(2) \right\rfloor_q \left\lfloor f(3) \right\rfloor_q \left\lfloor f(5) \right\rfloor_q \left\lfloor f(7) \right\rfloor_q \left\lfloor f(11) \right\rfloor_q \left\lfloor f(13) \right\rfloor_q \ldots, \tag{4}$$

where the arguments of f run through the sequence of primes.

Letting f be a polynomial with rational coefficients, Nakai and Shiokawa [22] could show that $\tau_q(f)$ is normal. Moreover, letting f be an entire function of bounded logarithmic order, Madritsch et al. [19] showed that $\mathcal{R}_N(\tau_q(f)) = \mathcal{O}(1/\log N)$.

At this point we want to mention the connection of normal numbers with uniform distribution. In particular, a number $x \in [0,1]$ is normal to base q if and only if the sequence $\{q^n x\}_{n \geq 0}$ is uniformly distributed modulo 1 (cf. Drmota and Tichy [11]). Here $\{y\}$ stands for the fractional part of y. Let us mention Kaufman [17] and Balog [4, 3], who investigated the distribution of the fractional part of $\sqrt{p}$ and p^θ respectively. Harman [14] gave estimates for the discrepancy of the sequence $\sqrt{p}$. In his papers Schoissengeier [24, 25] connected the estimation of the discrepancy of αp^θ with zero free regions of the Riemann zeta function. This allowed Tolev [29] to consider the multidimensional variant of this problem as well as to provide an explicit estimate for the discrepancy. This result was improved for different special cases by Zhai [31]. Since the results above deal with the case of $\theta < 1$ Baker and Kolesnik [2] extended these considerations to $\theta > 1$ and provided an explicit upper bound for the discrepancy in this case. This result was improved by Cao and Zhai [7] for $\frac{5}{3} < \theta < 3$. A multidimensional extension is due to Srinivasan and Tichy [27].

Combining the methods for proving uniform distribution mentioned above with a recent paper by Bergelson et al. [5] we want to extend the construction of Nakai and Shiokawa [20] to prime numbers. Our first main result is the following theorem.

Theorem 2. *Let $\theta > 1$ and $\alpha > 0$. Then*

$$\mathcal{R}_N(\tau_q(\alpha x^\theta)) = \mathcal{O}(1/\log N).$$

Remark 3. This estimate is best possible, as Schiffer [23] showed.

In our second main result we use the connection of this construction of normal numbers with the arithmetic mean of q-additive functions as described above. Known results in this area are due to Shiokawa [26], who was able to show the following theorem.

Theorem 4 ([26, Theorem])**.** *We have*

$$\sum_{p\le x} s_q(p) = \frac{q-1}{2}\frac{x}{\log q} + \mathcal{O}\left(x\left(\frac{\log\log x}{\log x}\right)^{\frac{1}{2}}\right),$$

where the sum runs over the primes and the implicit $\mathcal{O}$-constant may depend on q.

Similar results concerning the moments of the sum of digits function over primes have been established by Kátai [16]. An extension to Beurling primes is due to Heppner [15].

Let $\pi(x)$ stand for the number of primes less than or equal to x. Adapting these ideas to our method we obtain the following theorem.

Theorem 5. *Let $\theta > 1$ and $\alpha > 0$. Then*

$$\sum_{p\le N} s_q(\lfloor \alpha p^\theta \rfloor) = \frac{q-1}{2}\pi(N)\log_q N^\theta + \mathcal{O}(\pi(N)),$$

where the sum runs over the primes and the implicit $\mathcal{O}$-constant may depend on q and θ.

Remark 6. With simple modifications Theorem 5 can be extended to completely q-additive functions replacing s_q.

The proof of the two theorems is divided in three parts. In the following section we rewrite both statements and state the central theorem, which combines them and which we prove in the rest of the paper. In Section 3 we present all the tools we need in the proof of the central theorem. Finally, in Section 4 we proof the theorem.

2 Preliminaries

Throughout the paper, an interval denotes a set

$$I = (\alpha, \beta] = \{x : \alpha < x \le \beta\} \quad \text{with} \quad \beta > \alpha \ge \frac{1}{2}.$$

We will often subdivide a interval into smaller ones. In particular we use the observation that if $\log(\beta/\alpha) \ll \log N$, then $(\alpha, \beta]$ is the union of, say, s intervals of the type $(\gamma, \gamma_1]$ with $s \ll \log N$ and $\gamma_1 \le 2\gamma$. Given any complex function F on I, we have

$$\left|\sum_{x\in I} F(x)\right| \ll (\log N)\left|\sum_{\gamma<x\le\gamma_1} F(x)\right|, \tag{5}$$

for some such $(\gamma, \gamma_1]$.

In the proof p will always denote a prime. We fix the block $d_1 \cdots d_\ell$ and write $\mathcal{N}(f(p))$ for the number of occurrences of this block in the q-ary expansion of $\lfloor f(p) \rfloor$. By $\ell(m)$ we denote the length of the q-ary expansion of an integer m.

In the first step we want to get rid of the blocks that may occur between two expansions. To this end we define an integer N by

$$\sum_{p \leq N-1} \ell\left(\lfloor p^\theta \rfloor\right) < L \leq \sum_{p \leq N} \ell\left(\lfloor p^\theta \rfloor\right), \tag{6}$$

where $\sum$ indicates that the sum runs over all primes. Thus we get that

$$\begin{aligned} L &= \sum_{p \leq N} \ell(\lfloor p^\theta \rfloor) + \mathcal{O}(\pi(N)) + \mathcal{O}(\theta \log_q(N)) \\ &= \frac{\theta}{\log q} N + \mathcal{O}\left(\frac{N}{\log N}\right). \end{aligned} \tag{7}$$

Here we have used the prime number theorem in the form

$$\pi(x) = \mathrm{Li}\, x + \mathcal{O}\left(\frac{x}{(\log x)^G}\right),$$

where G is an arbitrary positive constant and

$$\mathrm{Li}\, x = \int_2^x \frac{\mathrm{d}t}{\log t}.$$

Let $\mathcal{N}(n; d_1 \cdots d_\ell)$ be the number of occurrences of the block $d_1 \cdots d_\ell$ in the expansion of n. Since we have fixed the block $d_1 \cdots d_\ell$ we will write $\mathcal{N}(n) = \mathcal{N}(n; d_1 \cdots d_\ell)$ for short. Then (7) implies that

$$\left| \mathcal{N}(\tau_q(x^\theta); d_1 \cdots d_\ell; L) - \sum_{p \leq N} \mathcal{N}(p^\theta) \right| \ll \frac{L}{\log L}. \tag{8}$$

For the next step we collect all the values that have a certain length of expansion. Let j_0 be a sufficiently large integer. Then for each integer $j \geq j_0$ we get that there exists an N_j such that

$$q^{j-2} \leq f(N_j) < q^{j-1} \leq f(N_j + 1) < q^j.$$

We note that this is possible since f asymptotically grows as its leading coefficient. This implies that

$$N_j \asymp q^{\frac{j}{\beta}}.$$

Furthermore for $N \geq q^{j_0}$ we set J to be the greatest length of the q-ary expansions of $f(p)$ over the primes $p \leq N$, i.e.,

$$J := \max_{p \leq N} \ell(\lfloor f(p) \rfloor) = \log_q(f(N)) + \mathcal{O}(1) \asymp \log N. \tag{9}$$

In the next step we want to perform the counting by adding the leading zeroes to the expansion of $f(p)$. For $N_{j-1} < p \le N_j$ we may write $f(p)$ in q-ary expansion, i.e.,

$$f(p) = b_{j-1}q^{j-1} + b_{j-2}q^{j-2} + \cdots + b_1 q + b_0 + b_{-1}q^{-1} + \ldots .$$

Then we denote by $\mathcal{N}^*(f(p))$ the number of occurrences of the block $d_1, \ldots, d_\ell$ in the string $0 \cdots 0 b_{j-1} b_{j-2} \cdots b_1 b_0$, where we filled up the expansion with zeroes such that it has length J. The error of doing so can be estimated by

$$\begin{aligned} 0 &\le \sum_{p \le N} \mathcal{N}^*(f(p)) - \sum_{p \le N} \mathcal{N}(f(p)) \\ &\le \sum_{j=j_0+1}^{J-1} (J-j)\left(\pi(N_{j+1}) - \pi(N_j)\right) + \mathcal{O}(1) \\ &\le \sum_{j=j_0+2}^{J} \pi(N_j) + \mathcal{O}(1) \ll \sum_{j=j_0+2}^{J} \frac{q^{j/\beta}}{j} \ll \frac{N}{\log N} \ll \frac{L}{\log L}. \end{aligned} \tag{10}$$

In the following two sections we will estimate this sum of indicator functions in order to prove the following proposition.

Proposition 7. *Let $\theta > 1$ and $\alpha > 0$. Then*

$$\sum_{p \le N} \mathcal{N}^*\left(\left\lfloor \alpha p^\theta \right\rfloor\right) = q^{-k}\pi(N) \log_q N^\theta + \mathcal{O}\left(\frac{N}{\log N}\right) \tag{11}$$

Proof of Theorem 2. We insert (11) into (8) and get the desired result. □

Proof of Theorem 5. For this proof we have to rewrite the statement. In particular, we use that the sum of digits function counts the number of 1s, 2s, etc. and assigns weights to them, i.e.,

$$s_q(n) = \sum_{d=0}^{q-1} d \cdot \mathcal{N}(n; d).$$

Thus

$$\begin{aligned} \sum_{p \le N} s_q(\lfloor p^\theta \rfloor) &= \sum_{p \le N} \sum_{d=0}^{q-1} d \cdot \mathcal{N}(p^\theta) = \sum_{p \le N} \sum_{d=0}^{q-1} d \cdot \mathcal{N}^*(p^\theta) + \mathcal{O}\left(\frac{N}{\log N}\right) \\ &= \frac{q-1}{2}\pi(N) \log_q(N^\theta) + \mathcal{O}\left(\frac{N}{\log N}\right) \end{aligned}$$

and the theorem follows. □

3 Tools

In this section we want to present all the tools we need on the way of proof of Proposition 7. We start with an estimation which essentially goes back to Vinogradov. This will provide us with Fourier expansions for the indicator functions used in the proof. As usual, for a real number y, the expression $e(y)$ denotes $\exp\{2\pi i y\}$.

Lemma 8 ([30, Lemma 12])**.** *Let α, β, Δ be real numbers satisfying*

$$0 < \Delta < \frac{1}{2}, \quad \Delta \leq \beta - \alpha \leq 1 - \Delta.$$

Then there exists a periodic function $\psi(x)$ with period 1, satisfying

1. *$\psi(x) = 1$ in the interval $\alpha + \frac{1}{2}\Delta \leq x \leq \beta - \frac{1}{2}\Delta$,*
2. *$\psi(x) = 0$ in the interval $\beta + \frac{1}{2}\Delta \leq x \leq 1 + \alpha - \frac{1}{2}\Delta$,*
3. *$0 \leq \psi(x) \leq 1$ in the remainder of the interval $\alpha - \frac{1}{2}\Delta \leq x \leq 1 + \alpha - \frac{1}{2}\Delta$,*
4. *$\psi(x)$ has a Fourier series expansion of the form*

$$\psi(x) = \beta - \alpha + \sum_{\substack{\nu=-\infty \\ \nu \neq 0}}^{\infty} A(\nu) e(\nu x),$$

where

$$|A(\nu)| \ll \min\left(\frac{1}{\nu}, \beta - \alpha, \frac{1}{\nu^2 \Delta}\right). \tag{12}$$

After we have transformed the sums under consideration into exponential sums we want to split the interval by the following lemma.

Lemma 9. *Let $I = (a, b]$ be an interval and F be a complex function defined on I. If $\log(b/a) \ll L$, then I is the union of ℓ intervals of the type $(c, d]$ with $\ell \ll L$ and $d \leq 2c$. Furthermore we have*

$$\left| \sum_{n \in I} F(n) \right| \ll L \left| \sum_{n \in (c,d]} F(n) \right|,$$

for some such $(c, d]$.

Proof. For $i = 1, \ldots, \ell$ let I_i be the ℓ splitting intervals. Then

$$\left| \sum_{n \in I} F(n) \right| = \left| \sum_{i=1}^{\ell} \sum_{n \in I_i} F(n) \right| \leq \ell \max_{1 \leq i \leq \ell} \left| \sum_{n \in I_i} F(n) \right| \ll L \left| \sum_{n \in I_i} F(n) \right|$$

$\square$

We will apply the following lemma in order to estimate the occurring exponential sums provided that the coefficients are very small. This corresponds to the case of the most significant digits in the expansion.

Lemma 10 ([28, Lemma 4.19])**.** *Let $F(x)$ be a real function, k times differentiable, and satisfying $\left|F^{(k)}(x)\right| \geq \lambda > 0$ throughout the interval $[a,b]$. Then*

$$\left|\int_a^b e(F(x))\mathrm{d}x\right| \leq c(k)\lambda^{-1/k}.$$

A standard tool for estimating exponential sums over the primes is Vaughan's identity. In order to apply this identity we have to rewrite the exponential sum into a normal one having von Mangoldt's function as weights. Therefore let Λ denote von Mangoldt's function, i.e.,

$$\Lambda(n) = \begin{cases} \log p, & \text{if } n = p^k \text{ for some prime } p \text{ and an integer } k \geq 1; \\ 0, & \text{otherwise.} \end{cases}$$

In the next step we may subdivide this weighted exponential sum into several sums of Type I and II. In particular, let $P \geq 2$ and $P_1 \leq 2P$, then we define Type I and Type II sums by the expressions

$$\sum_{X<x\leq X_1} a_x \sum_{\substack{Y<y\leq Y_1\\ P<xy\leq P_1}} f(xy) \quad \text{(Type I)} \tag{13}$$

$$\sum_{X<x\leq X_1} a_x \sum_{\substack{Y<y\leq Y_1\\ P<xy\leq P_1}} (\log y) f(xy)$$

$$\sum_{X<x\leq X_1} a_x \sum_{\substack{Y<y\leq Y_1\\ P<xy\leq P_1}} b_y f(xy) \quad \text{(Type II)} \tag{14}$$

with $X_1 \leq 2X$, $Y_1 \leq 2Y$, $|a_x| \ll P^\varepsilon$, $|b_y| \ll P^\varepsilon$ for every $\varepsilon > 0$ and

$$P \ll XY \ll P,$$

respectively. The following lemma provides the central tool for the subdivision of the weighted exponential sum.

Lemma 11 ([2, Lemma 1])**.** *Let $f(n)$ be a complex valued function and $P \geq 2$, $P_1 \leq 2P$. Furthermore let U, V, and Z be positive numbers satisfying*

$$2 \leq U < V \leq Z \leq P, \tag{15}$$

$$U^2 \leq Z, \quad 128UZ^2 \leq P_1, \quad 2^{18}P_1 \leq V^3. \tag{16}$$

Then the sum

$$\sum_{P\leq n\leq P_1} \Lambda(n) f(n)$$

may be decomposed into $\ll (\log P)^6$ sums, each of which is either a Type I sum with $Y \geq Z$ or a Type II sum with $U \leq Y \leq V$.

The next tool is an estimation for the exponential sum. After subdividing the weighted exponential sum we use Vinogradov's method in order to estimate the occurring unweighted exponential sums.

Lemma 12 ([20, Lemma 6])**.** *Let k, P and N be integers such that $k \geq 2$, $2 \leq N \leq P$. Let $g(x)$ be real and have continuous derivatives up to the $(k+1)$th order in $[P+1, P+N]$; let $0 < \lambda < 1/(2c_0(k+1))$ and*

$$\lambda \leq \frac{g^{(k+1)}(x)}{(k+1)!} \leq c_0\lambda \quad (P+1 \leq x \leq P+N),$$

or the same for $-g^{(k+1)}(x)$, and let

$$N^{-k-1+\rho} \leq \lambda \leq N^{-1}$$

with $0 < \rho \leq k$. Then

$$\sum_{n=P+1}^{P+N} e(g(n)) \ll N^{1-\eta},$$

where

$$\eta = \frac{\rho}{16(k+1)L}, \quad L = 1 + \left\lfloor \frac{1}{4}k(k+1) + kR \right\rfloor, \quad R = 1 + \left\lfloor \frac{\log\left(\frac{1}{\rho}k(k+1)^2\right)}{-\log\left(1-\frac{1}{k}\right)} \right\rfloor. \tag{17}$$

4 Proof of Proposition 7

We will apply the estimates of the preceding sections in order to estimate the exponential sums occurring in the proof. We will proceed in four steps.

1. In the first step we use a method of Vinogradov [30] in order to rewrite the counting function into the estimation of exponential sums. Then we will distinguish two cases in the following two steps.

2. First we assume that we are interested in a block which occurs among the most significant digits. This corresponds to a very small coefficient in the exponential sum and we may use the method of van der Corput (cf. [13]).

3. For the blocks occurring among the least significant digits we apply Vaughan's identity together with ideas from a recent paper by Bergelson et al. [5].

4. Finally we combine the estimates of the last two steps in order to end the proof.

In this proof, the letter p will always denote a prime and we set $f(x) := \alpha x^\theta$ for short. Furthermore we set

$$\delta := \min\left(\frac{1}{4}, 1-\theta\right). \tag{18}$$

4.1 Rewriting the sum

Throughout the rest of the paper we fix a block $d_1 \cdots d_\ell$. In order to count the occurrences of this block in the q-ary expansion of $\lfloor f(p) \rfloor$ $(2 \le p \le P)$ we define the indicator function

$$\mathcal{I}(t) = \begin{cases} 1, & \text{if } \sum_{i=1}^{\ell} d_i q^{-i} \le t - \lfloor t \rfloor < \sum_{i=1}^{\ell} d_i q^{-i} + q^{-\ell}; \\ 0, & \text{otherwise}; \end{cases} \tag{19}$$

which is a 1-periodic function. Indeed, we have

$$\mathcal{I}(q^{-j} f(n)) = 1 \iff d_1 \cdots d_\ell = b_{j-1} \cdots b_{j-\ell}.$$

Thus we can write our block counting function as follows

$$\mathcal{N}^*(f(p)) = \sum_{j=l}^{J} \mathcal{I}\left(q^{-j} f(p)\right). \tag{20}$$

Following Nakai and Shiokawa [20] we want to approximate $\mathcal{I}$ from above and from below by two 1-periodic functions having small Fourier coefficients. In particular, we set $H = N^{\delta/3}$ and

$$\begin{aligned} \alpha_- &= \sum_{\lambda=1}^{\ell} d_\lambda q^{-\lambda} + (2H)^{-1}, \quad \beta_- = \sum_{\lambda=1}^{\ell} d_\lambda q^{-\lambda} + q^{-\ell} - (2H)^{-1}, \quad \Delta_- = H^{-1}, \\ \alpha_+ &= \sum_{\lambda=1}^{\ell} d_\lambda q^{-\lambda} - (2H)^{-1}, \quad \beta_+ = \sum_{\lambda=1}^{\ell} d_\lambda q^{-\lambda} + q^{-\ell} + (2H)^{-1}, \quad \Delta_+ = H^{-1}. \end{aligned} \tag{21}$$

We apply Lemma 8 with $(\alpha, \beta, \delta) = (\alpha_-, \beta_-, \delta_-)$ and $(\alpha, \beta, \delta) = (\alpha_+, \beta_+, \delta_+)$, respectively, in order to get two functions $\mathcal{I}_-$ and $\mathcal{I}_+$. By the choices of $(\alpha_\pm, \beta_\pm, \delta_\pm)$ it is immediate that

$$\mathcal{I}_-(t) \le \mathcal{I}(t) \le \mathcal{I}_+(t) \qquad (t \in \mathbb{R}). \tag{22}$$

Lemma 8 also implies that these two functions have Fourier expansions

$$\mathcal{I}_\pm(t) = q^{-\ell} \pm H^{-1} + \sum_{\substack{\nu=-\infty \\ \nu \neq 0}}^{\infty} A_\pm(\nu) e(\nu t) \tag{23}$$

satisfying

$$|A_\pm(\nu)| \ll \min(|\nu|^{-1}, H |\nu|^{-2}).$$

In a next step we want to replace $\mathcal{I}$ by $\mathcal{I}_+$ in (20). For this purpose we observe, using (22), that

$$|\mathcal{I}(t) - \mathcal{I}_+(t)| \le |\mathcal{I}_+(t) - \mathcal{I}_-(t)| \ll H^{-1} + \sum_{\substack{\nu=-\infty \\ \nu \neq 0}}^{\infty} A_\pm(\nu) e(\nu t).$$

Thus subtracting yields the main part, and summing over $p \le N$ gives

$$\left| \sum_{p \le N} \mathcal{I}(q^{-j} f(p)) - \frac{\pi(N)}{q^\ell} \right| \ll \pi(N) H^{-1} + \sum_{\substack{\nu=-\infty \\ \nu \ne 0}}^{\infty} A_\pm(\nu) \sum_{p \le N} e\left(\frac{\nu}{q^j} f(p) \right). \tag{24}$$

Now we consider the coefficients $A_\pm(\nu)$. Noting (12) one observes that

$$A_\pm(\nu) \ll \begin{cases} \nu^{-1}, & \text{for } |\nu| \le H; \\ H\nu^{-2}, & \text{for } |\nu| > H. \end{cases}$$

Estimating all summands with $|\nu| > H$ trivially we get

$$\sum_{\substack{\nu=-\infty \\ \nu \ne 0}}^{\infty} A_\pm(\nu) e\left(\frac{\nu}{q^j} f(p) \right) \ll \sum_{\nu=1}^{H} \nu^{-1} e\left(\frac{\nu}{q^j} f(p) \right) + H^{-1}.$$

Using this in (24) yields

$$\left| \sum_{p \le N} \mathcal{I}(q^{-j} f(p)) - \frac{\pi(N)}{q^\ell} \right| \ll \pi(N) H^{-1} + \sum_{\nu=1}^{H} \nu^{-1} \sum_{p \le N} e\left(\frac{\nu}{q^j} f(p) \right). \tag{25}$$

Finally we sum over all js and get

$$\left| \sum_{p \le N} \mathcal{N}^*(f(p)) - \frac{\pi(N)}{q^\ell} J \right| \ll \pi(N) H^{-1} J + \sum_{j=\ell}^{J} \sum_{\nu=1}^{H} \nu^{-1} S(N, j, \nu), \tag{26}$$

where we have set

$$S(N, j, \nu) := \sum_{p \le N} e\left(\frac{\nu}{q^j} f(p) \right).$$

The crucial part is the estimation of the exponential sums over the primes. In the following we will distinguish two cases according to the size of j. This corresponds to the position in the expansion of $f(p)$. In particular, let $\rho > 0$ be arbitrarily small then we want to distinguish between the most significant digits and the least significant digits, i.e., between the ranges

$$1 \le q^j \le N^{\theta-1+\rho} \quad \text{and} \quad N^{\theta-1+\rho} < q^j \le N^\theta.$$

4.2 Most significant digits

In this subsection we assume that

$$N^{\theta-1+\rho} < q^j \le N^\theta,$$

which means that we deal with the most significant digits in the expansion. We start by rewriting the sum into an integral.

$$S(N, j, \nu) = \sum_{p \le N} e\left(\frac{\nu}{q^j} f(p) \right) = \int_2^N e\left(\frac{\nu}{q^j} f(t) \right) \mathrm{d}\pi(t) + \mathcal{O}(1).$$

In the second step we then apply the prime number theorem. Thus

$$S(N,j,\nu) = \int_{N(\log N)^{-G}}^{N} e\left(\frac{\nu}{q^j} f(t)\right) \frac{\mathrm{d}t}{\log t} + \mathcal{O}\left(\frac{N}{(\log N)^G}\right).$$

Now we use the second mean-value theorem together with Lemma 10 and $k = \lfloor \theta \rfloor$ to get

$$\begin{aligned} S(N,j,\nu) &\ll \frac{1}{\log N} \sup_{\xi} \left| \int_{N(\log N)^{-G}}^{\xi} e\left(\frac{\nu}{q^j} f(t)\right) \mathrm{d}t \right| + \mathcal{O}\left(\frac{N}{(\log N)^G}\right) \\ &\ll \frac{1}{\log N} \left(\frac{|\nu|}{q^j}\right)^{-\frac{1}{k}} + \mathcal{O}\left(\frac{N}{(\log N)^G}\right). \end{aligned} \tag{27}$$

4.3 Least significant digits

For the digits in this range we want to apply Vaughan's identity in order to transfer the sum over the primes into two special types of sums involving products of integers. Before we may apply Vaughan's identity we have to weight the exponential sum under consideration by the von Mangoldt function. By an application of Lemma 9, it suffices to consider an interval of the form $(P, 2P]$. Thus

$$|S(N,j,\nu)| \ll (\log N) \left| \sum_{P<p\leq 2P} e\left(f(p)\right) \right|.$$

Using partial summation we get

$$|S(N,j,\nu)| \ll (\log N) \left| \sum_{P<p\leq 2P} e\left(f(p)\right) \right| \ll (\log N) P^{\frac{1}{2}} + (\log N) \left| \sum_{P<n\leq P_1} \Lambda(n) e\left(f(n)\right) \right|$$

for some P_1 with $P < P_1 \leq 2P$. From now on we may assume that $P > N^{1-\eta}$.

Then an application of Lemma 11 with $U = P^{\frac{\delta}{3}}$, $V = P^{\frac{1}{3}}$, $Z = P^{\frac{1}{2}-\frac{\delta}{3}}$ yields

$$S(N,j,\nu) \ll P^{\frac{1}{2}} + (\log P)^7 |S_1|, \tag{28}$$

where S_1 is either a Type I sum as in (13) with $Y \geq P^{\frac{1}{2}-\frac{\delta}{3}}$ or a Type II sum as in (14) with

$$P^{\frac{\delta}{3}} \leq Y \leq P^{\frac{1}{3}}.$$

Suppose first that S_1 is a Type II sum, i.e.,

$$S_1 = \sum_{X<x\leq X_1} a_x \sum_{\substack{Y<y\leq Y_1 \\ P<xy\leq P_1}} b_y e\left(f(xy)\right).$$

Then an application of the Cauchy-Schwarz inequality yields

$$
\begin{aligned}
|S_1|^2 &\le \sum_{X<x\le X_1} |a_x|^2 \sum_{X<x\le X_1} \left| \sum_{\substack{Y<y\le Y_1 \\ P<xy\le P_1}} b_y e\left(\frac{\nu}{q^j} f(xy)\right) \right|^2 \\
&\ll XP^{2\varepsilon} \sum_{Y<y\le Y_1} \sum_{Y<z\le Y_1} b_y \overline{b_z} \sum_{\substack{X<x\le X_1 \\ P<xy,xz\le P_1}} e\left(\frac{\nu}{q^j}\left(f(xy)-f(xz)\right)\right),
\end{aligned}
$$

where we have used that $|a_x| \ll P^\varepsilon$. Collecting all the terms where $y=z$ and using $|b_y| \ll P^\varepsilon$ yields

$$
|S_1|^2 \ll XP^{4\varepsilon} \left(XY + \sum_{Y<y<z\le Y_1} \left| \sum_{\substack{X<x\le X_1 \\ P<xy,xz\le P_1}} e\left(\frac{\nu}{q^j}\left(f(xy)-f(xz)\right)\right) \right| \right). \tag{29}
$$

There must be a pair (y,z) with $Y<y<z<Y_1$ such that

$$
|S_1|^2 \ll P^{2+4\varepsilon} Y^{-1} + P^{4\varepsilon} XY^2 \left| \sum_{X_2<x\le X_3} e(g(x)) \right|, \tag{30}
$$

where $X_2 = \max(X, Py^{-1})$, $X_3 = \min(X_1, P_1 z^{-1})$ and

$$
g(x) = \frac{\nu}{q^j}\left(f(xy)-f(xz)\right) = \frac{\nu}{q^j}\alpha(y^\theta - z^\theta)x^\theta.
$$

We will apply Lemma 12 to estimate the exponential sum. Setting

$$
k := \lceil 2\theta \rceil + 1
$$

we get that $g^{(k+1)}(x) \sim \nu q^{-j}\alpha\theta(\theta-1)\cdots(\theta-k)x^{\theta-(k+1)}$. Thus

$$
\lambda \le \frac{g^{(k+1)}(x)}{(k+1)!} \le c_0\lambda \quad (X_2 < x \le X_3)
$$

or similarly for $-g^{(k+1)}(x)$, where

$$
\lambda = c\nu q^{-j}\alpha(y^\theta - z^\theta)X^{\theta-(k+1)}
$$

and c depends only on θ and α.

Since $\theta > 1$ we get

$$
\lambda \ge P^{\delta-\theta} Y^{\theta-1} X^{\theta-(k+1)} \ge X^{-k-\frac{1}{2}}.
$$

Similarly we obtain

$$
\lambda \le P^{2\delta} Y^\theta X^{\theta-(k+1)} \ll P^{\theta+2\delta} X^{-(k+1)} \le X^{-1}.
$$

Thus we get that $X^{-k-\frac{1}{2}} \leq \lambda \leq X^{-1}$. Therefore an application of Lemma 12 yields

$$\sum_{X_2<x\leq X_3} e(g(x)) \ll X^{1-\eta},$$

where η depends only on k an therefore on θ. Inserting this in (30) we get

$$|S_1|^2 \ll P^{2+4\varepsilon}Y^{-1} + P^{4\varepsilon}XY^2X^{1-\eta} \ll P^{2+4\varepsilon}\left(P^{-\delta/3}+P^{-2\eta/3}\right). \tag{31}$$

The case of S_1 being a type I sum is similar but simpler. We have

$$|S| \leq \sum_{X<x\leq X_1} |a_x| \left|\sum_{\substack{Y<y\leq Y_1\\ P<xy\leq P_1}} (\log y) e\left(f(xy)\right)\right| \ll XP^{\varepsilon}\left|\sum_{\substack{Y<y\leq Y_1\\ P<xy\leq P_1}} (\log y) e\left(f(xy)\right)\right|$$

for some x with $X < x \leq X_1$. By a partial summation we get

$$|S| \ll XP^{\varepsilon}\log P\left|\sum_{\substack{Y_2<y\leq Y_3\\ P<xy\leq P_1}} e\left(f(xy)\right)\right| \tag{32}$$

for some $Y \leq Y_2 < Y_3 \leq Y_1$. Now we set

$$g(y) = f(xy) = \frac{\nu}{q^j}\alpha x^{\theta}y^{\theta}.$$

Again the idea is to apply Lemma 12 for the estimation of the exponential sum. We set

$$k := \lceil 3\theta \rceil + 2$$

and get for the $k+1$-st derivative

$$\lambda \leq \frac{g^{(k+1)}(x)}{(k+1)!} \leq c_0\lambda \quad (X_2 < x \leq X_3)$$

or similarly for $-g^{(k+1)}(x)$, where

$$\lambda = c\frac{\nu}{q^j}\alpha x^{\theta}Y^{\theta-(k+1)}$$

and c again depends only on α and θ. We may assume that N and hence P is sufficiently large, then we get that

$$Y^{-(k+1)} \ll P^{-\theta}X^{\theta}Y^{\theta-(k+1)} \leq \lambda \leq P^{2\delta}X^{\theta}Y^{\theta-(k+1)} \ll P^{\theta+2\delta}Y^{-(k+1)} \leq Y^{-1}.$$

Now an application of Lemma 2.5 yields

$$\sum_{Y_2<y\leq Y_3} e(g(y)) \ll Y^{1-\eta},$$

where η depends only on k and thus on θ. Inserting this in (32) we get

$$|S_1| \ll (\log P)XP^{\varepsilon}Y^{1-\eta} \ll (\log P)P^{1+\varepsilon-\eta(1/2-\delta/3)}. \tag{33}$$

Combining (33) and (31) in (28) yields

$$\begin{aligned} |S(N,j,\nu)| &\ll P^{\frac{1}{2}} + (\log P)^7\left(P^{1+2\varepsilon}\left(P^{-\delta/6}+P^{-\eta/3}\right) + (\log P)P^{1+\varepsilon-\eta(1/2-\delta/3)}\right) \\ &\ll P^{\frac{1}{2}} + (\log P)^8 P^{1-\sigma}. \end{aligned} \tag{34}$$

4.4 Conclusion

On the one hand summing (27) over j and ν yields

$$\begin{aligned}
&\sum_{1\leq|\nu|\leq H}|\nu|^{-1}\sum_{N^{\theta-\delta}<q^j\leq N^\theta}S(N,j,\nu)\\
&\quad\ll\sum_{1\leq|\nu|\leq H}|\nu|^{-1}\sum_{N^{\theta-\delta}<q^j\leq N^\theta}\left(\frac{1}{\log N}\left(\frac{|\nu|}{q^j}\right)^{-\frac{1}{k}}+\mathcal{O}\left(\frac{N}{(\log N)^G}\right)\right)\\
&\quad\ll\frac{1}{\log N}\sum_{1\leq|\nu|\leq H}|\nu|^{-1-\frac{1}{k}}\sum_{N^{\theta-\delta}<q^j\leq N^\theta}q^{-\frac{j}{k}}+\mathcal{O}\left(\frac{N}{(\log N)^{G-2}}\right)\\
&\quad\ll\frac{N}{\log N}.
\end{aligned}$$

On the other hand in (34) we sum over j and ν and get

$$\sum_{1\leq|\nu|\leq H}|\nu|^{-1}\sum_{q^\ell\leq q^j\leq N^{\theta-\delta}}S(N,j,\nu)\ll(\log N)^2N^{\frac{1}{2}}+(\log N)^{10}N^{1-\sigma'}.$$

Combining these estimates in (26) finally yields

$$\left|\sum_{p\leq N}\mathcal{N}^*(f(p))-\frac{\pi(N)}{q^\ell}J\right|\ll\frac{N}{\log N}$$

and the proposition is proved.

5 Acknowledgment

The authors thank the anonymous referee, who read the manuscript very carefully and his/her suggestions considerably improved the presentation of the results.

References

[1] J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, 2003.

[2] R. C. Baker and G. Kolesnik. On the distribution of p^α modulo one. *J. Reine Angew. Math.* **356** (1985), 174–193.

[3] A. Balog. On the fractional part of p^θ. *Arch. Math. (Basel)* **40** (1983), 434–440.

[4] A. Balog. On the distribution of $p^\theta \bmod 1$. *Acta Math. Hungar.* **45** (1985), 179–199.

[5] V. Bergelson, G. Kolesnik, M. Madritsch, Y. Son, and R. Tichy. Uniform distribution of prime powers and applications to van der Corput sets. Submitted, 2012.

[6] E. Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Palermo Rend.* **27** (1909), 247–271.

[7] X. Cao and W. Zhai. On the distribution of p^α modulo one. *J. Théor. Nombres Bordeaux* **11** (1999), 407–423.

[8] D. G. Champernowne. The construction of decimals normal in the scale of ten. *J. Lond. Math. Soc.* **8** (1933), 254–260.

[9] A. H. Copeland and P. Erdös. Note on normal numbers. *Bull. Amer. Math. Soc.* **52** (1946), 857–860.

[10] H. Davenport and P. Erdös. Note on normal decimals. *Canadian J. Math.* **4** (1952), 58–63.

[11] M. Drmota and R. F. Tichy. *Sequences, Discrepancies and Applications*, Vol. 1651 of *Lecture Notes in Mathematics*. Springer-Verlag, 1997.

[12] J.-M. Dumont and A. Thomas. Modifications de nombres normaux par des transducteurs. *Acta Arith.* **68** (1994), 153–170.

[13] S. W. Graham and G. Kolesnik. *Van der Corput's Method of Exponential Sums*, Vol. 126 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 1991.

[14] G. Harman. On the distribution of $\sqrt{p}$ modulo one. *Mathematika* **30** (1983), 104–116.

[15] E. Heppner. Über die Summe der Ziffern natürlicher Zahlen. *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **19** (1976), 41–43.

[16] I. Kátai. On the sum of digits of primes. *Acta Math. Acad. Sci. Hungar.* **30** (1977), 169–173.

[17] R. M. Kaufman. The distribution of $\{\sqrt{p}\}$. *Mat. Zametki* **26** (1979), 497–504, 653.

[18] Manfred G. Madritsch. The summatory function of q-additive functions on pseudo-polynomial sequences. *J. Théor. Nombres Bordeaux* **24** (2012), 153–171.

[19] M. G. Madritsch, J. M. Thuswaldner, and R. F. Tichy. Normality of numbers generated by the values of entire functions. *J. Number Theory* **128** (2008), 1127–1145.

[20] Y. Nakai and I. Shiokawa. A class of normal numbers. *Japan. J. Math.* **16** (1990), 17–29.

[21] Y. Nakai and I. Shiokawa. Discrepancy estimates for a class of normal numbers. *Acta Arith.* **62** (1992), 271–284.

[22] Y. Nakai and I. Shiokawa. Normality of numbers generated by the values of polynomials at primes. *Acta Arith.* **81** (1997), 345–356.

[23] J. Schiffer. Discrepancy of normal numbers. *Acta Arith.* **47** (1986), 175–186.

[24] J. Schoissengeier. Eine neue Diskrepanz für gewisse Primzahlfolgen. *Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II* **187** (4-7) (1978), 219–224.

[25] J. Schoissengeier. The connection between the zeros of the ζ-function and sequences $(g(p))$, p prime, $\bmod 1$. *Monatsh. Math.* **87** (1979), 21–52.

[26] I. Shiokawa. On the sum of digits of prime numbers. *Proc. Japan Acad.* **50** (1974), 551–554.

[27] S. Srinivasan and R. F. Tichy. Uniform distribution of prime power sequences. *Anz. Österreich. Akad. Wiss. Math.-Natur. Kl.* **130** (1993), 33–36.

[28] E. C. Titchmarsh. *The Theory of The Riemann Zeta-Function*, second edition. Oxford University Press, 1986.

[29] D. I. Tolev. On the simultaneous distribution of the fractional parts of different powers of prime numbers. *J. Number Theory* **37** (1991), 298–306.

[30] I. M. Vinogradov. *The Method of Trigonometrical Sums in The Theory of Numbers*. Translated from the Russian, revised and annotated by K. F. Roth and Anne Davenport, reprint of the 1954 translation. Dover Publications Inc., 2004.

[31] W. Zhai. On the simultaneous distribution of the fractional parts of different powers of prime numbers. *J. Number Theory* **86** (2001), 133–155.

2010 *Mathematics Subject Classification*: Primary 11N37; Secondary 11A63.
Keywords: normal number, pseudo-polynomial.

Received July 23 2012; revised version received October 31 2012. Published in *Journal of Integer Sequences*, March 2 2013.

Return to Journal of Integer Sequences home page.