COMPOSITIO MATHEMATICA

Primes with an average sum of digits

Michael Drmota, Christian Mauduit and Joël Rivat

Compositio Math. **145** (2009), 271–292.

doi:10.1112/S0010437X08003898

FOUNDATION
COMPOSITIO
MATHEMATICA

The London
Mathematical
Society

Compositio Math. 145 (2009) 271–292
doi:10.1112/S0010437X08003898

Primes with an average sum of digits

Michael Drmota, Christian Mauduit and Joël Rivat

Abstract

The main goal of this paper is to provide asymptotic expansions for the numbers $\#\{p \le x : p \text{ prime}, s_q(p) = k\}$ for k close to $((q-1)/2)\log_q x$, where $s_q(n)$ denotes the q-ary sum-of-digits function. The proof is based on a thorough analysis of exponential sums of the form $\sum_{p\le x} e(\alpha s_q(p))$ (where the sum is restricted to p prime), for which we have to extend a recent result by the second two authors.

1. Introduction

In this paper the letter p will denote a prime number and $e(x)$ the exponential function $e^{2\pi i x}$.

For an integer $q \ge 2$, let $s_q(n)$ denote the q-ary sum-of-digits function of a non-negative integer n; that is, if n is given by its q-ary digital expansion $n = \sum_{j\ge 0} \varepsilon_j(n) q^j$ with digits $\varepsilon_j(n) \in \{0, 1, \dots, q-1\}$, then

$$s_q(n) = \sum_{j\ge 0} \varepsilon_j(n).$$

The statistical behaviour of the sum-of-digits function and, more generally, of q-additive functions has been intensively studied by several authors. It is, for example, well-known (see, for instance, the paper of Delange [Del75]) that the average sum-of-digits function is given by

$$\frac{1}{x}\sum_{n<x} s_q(n) = \frac{q-1}{2}\log_q x + \gamma(\log_q x),$$

where γ is a continuous, nowhere-differentiable and periodic function with period 1. Similar relations are known for higher moments (see [GKPT], as well as [Sto77] and [Coq86], for the case $q = 2$). Furthermore, the distribution of the sum-of-digits function can be approximated by a normal distribution

$$\frac{1}{x}\#\left\{n < x : s_q(n) \le \mu_q \log_q x + y\sqrt{\sigma_q^2 \log_q x}\right\} = \Phi(y) + o(1), \tag{1}$$

where

$$\mu_q := \frac{q-1}{2}, \quad \sigma_q^2 := \frac{q^2-1}{12}$$

and $\Phi(y)$ denotes the normal distribution function (see [KM68]).

A local version of these results can be found in [MS97], where a uniform estimate of $\#\{n < q^\nu : s_q(n) = k\}$ is provided for any $k \le \mu_q \nu$; also, in [FM05] it is proved that for any

Received 17 December 2007, accepted in final form 31 August 2008, published online 19 February 2009.
2000 Mathematics Subject Classification 11A63, 11L03, 11N05 (primary); 11N60, 11L20, 60F05 (secondary).
Keywords: sum-of-digits function, primes, exponential sums, central limit theorem.
The first author was supported by the Austrian Science Foundation (FWF), grant S9604, which is part of the National Research Network 'Analytic Combinatorics and Probabilistic Number Theory'.

fixed $k \geq 1$, we have

$$\#\{n < x : s_q(n) = \mu_q \lfloor \log_q n \rfloor + b(\lfloor \log_q n \rfloor)\} = \sqrt{\frac{6}{\pi(q^2-1)}} \frac{x}{\sqrt{\log x}} + O_K\left(\frac{x}{\log_q x}\right)$$

uniformly for any $x \geq 2$ and $b : \mathbb{N} \to \mathbb{R}$ such that $|b(n)| \leq K n^{1/4}$ and $\mu_q n + b(n) \in \mathbb{N}$ for any $n \geq 1$.

The first result on the asymptotic behaviour of the sum-of-digits function restricted to prime numbers is a consequence of the famous theorem of Copeland and Erdős in [CE46], which concerns the normality of the real number whose q-adic representation is 0 followed by the concatenation of the increasing sequence of prime numbers written in base q. Indeed, it follows from Copeland and Erdős's theorem that

$$\frac{1}{\pi(x)} \sum_{p<x} s_q(p) = \frac{q-1}{2} \log_q x + o(\log_q x), \tag{2}$$

and it has been shown by Shiokawa in [Shi74] that

$$\frac{1}{\pi(x)} \sum_{p<x} s_q(p) = \frac{q-1}{2} \log_q x + O(\sqrt{\log x \log\log x})$$

(see also [Kat67] for a related result).

Interestingly, these results suggest that the overall behaviour of the sum-of-digits function is, in principle, the same as when the average is taken over primes $p \leq x$. For example, Katai showed in [Kat77] that

$$\sum_{p \leq x} |s_q(p) - \mu_q \log_q x|^k \ll x (\log x)^{k/2-1} \quad \text{for } k = 1, 2, \ldots$$

and in [Kat86] that there is a central limit theorem similar to the statement above, namely,

$$\frac{1}{\pi(x)} \#\left\{p < x : s_q(p) \leq \mu_q \log_q x + y\sqrt{\sigma_q^2 \log_q x}\right\} = \Phi(y) + o(1) \tag{3}$$

(see also [KM68] for a related result).

The first aim of this paper is to prove Theorem 1.1, which is a local version of these results.

Theorem 1.1. *We have, uniformly for all integers $k \geq 0$ with $(k, q-1) = 1$,*

$$\#\{p \leq x : s_q(p) = k\} = \frac{q-1}{\varphi(q-1)} \frac{\pi(x)}{\sqrt{2\pi\sigma_q^2 \log_q x}} \left(\exp\left(-\frac{(k - \mu_q \log_q x)^2}{2\sigma_q^2 \log_q x}\right) + O((\log x)^{-1/2+\varepsilon})\right), \tag{4}$$

where $\varepsilon > 0$ is arbitrary but fixed.

Remark 1. The condition $(k, q-1) = 1$ is necessary: since $s_q(p) \equiv p \bmod q-1$, it follows that

$$\{p \leq x : s_q(p) = k\} \subset \{p \leq x : p \equiv k \bmod (q-1)\},$$

which is finite in the case where $(k, q-1) > 1$.

Such a local version of (2) or (3) was considered by Erdős to be 'hopelessly difficult', and the first breakthrough in this direction was made by Mauduit and Rivat, who proved in [MR05] the Gelfond conjecture concerning the sum of digits of prime numbers: for $(m, q-1) = 1$, there exists $\sigma_{q,m} > 0$ such that for every $a \in \mathbb{Z}$ we have

$$\#\{p \leq x,\ s_q(p) \equiv a \bmod m\} = \frac{1}{m}\pi(x) + O_{q,m}(x^{1-\sigma_{q,m}}).$$

However, the method involved in proving this theorem is not enough to provide a proof of Theorem 1.1.

If we consider primes p for which the sum-of-digits function $s_q(p)$ equals precisely the 'expected value' $\lfloor \mu_q \log_q p \rfloor$, then we get the following result that can be deduced from Theorem 1.1.

THEOREM 1.2. *We have, as $x \to \infty$,*

$$\#\{p \le x : s_q(p) = \lfloor \mu_q \log_q p \rfloor\} = Q\left(\frac{\mu_q}{q-1} \log_q x\right) \frac{x}{(\log_q x)^{3/2}} (1 + O_\varepsilon((\log x)^{-1/2+\varepsilon})), \tag{5}$$

where $Q(t)$ denotes a positive periodic function with period 1 and $\varepsilon > 0$ is arbitrary but fixed.

The proof of Theorem 1.1 relies on a precise analysis of the generating function

$$T(z) = \sum_{p \le x} z^{s_q(p)}$$

for complex numbers z of modulus $|z| = 1$ (Propositions 2.1 and 2.2). It is, however, an interesting and probably very difficult problem to obtain, in addition, some asymptotic information on $T(z)$ for z with $|z| \neq 1$. For example, we are not able to provide any non-trivial bounds for the sum

$$T(2) = \sum_{p \le x} 2^{s_q(p)}.$$

Such bounds could be used to obtain estimates for *tail distributions*, i.e. bounds on the numbers

$$\#\{p \le x : s_q(p) \le c_1 \log_q(x)\} \quad \text{and} \quad \#\{p \le x : s_q(p) \ge c_2 \log_q(x)\}$$

for $0 < c_1 < \mu_q$ and $\mu_q < c_2 < 2\mu_q$, respectively. As a matter of curiosity, we mention that Fermat primes and Mersenne primes correspond to the extremal cases in base $q = 2$ defined, respectively, by $s_2(p) = 2$ and $s_2(p) = \lfloor \log_2 p \rfloor$.

2. Plan for the proof of the main theorems

The proof of Theorem 1.1 uses two main ingredients, Propositions 2.1 and 2.2, which we prove in §§ 3 and 4.

The aim of Proposition 2.1, whose proof is based on a method from [MR05], is to provide a bound for $\sum_{p \le x} e(\alpha s_q(p))$ which is uniform in terms of α and x. This will enable us to apply a saddle-point-type method in § 5.1 to obtain asymptotics for the numbers $\#\{p \le x : s_q(p) = k\}$.

PROPOSITION 2.1. *For every fixed integer $q \ge 2$, there exists a constant $c_1 > 0$ such that*

$$\sum_{p \le x} e(\alpha s_q(p)) \ll (\log x)^3 x^{1 - c_1 \|(q-1)\alpha\|^2} \tag{6}$$

uniformly for real α.

The main idea of Proposition 2.2 is to approximate the sum-of-digits function by a sum of independent random variables. In fact, we shall adapt the moment method due to Bassily and Kátai [BK95] (see also [KM68] and [Kat77]). The difference from [BK95] is that we provide bounds for the dth moments (of a certain random variable) that are uniform for all $d \ge 1$. Note that the generalization of [BK95] provided in [BK96] is not sufficient for our purposes here;

therefore we need to adapt all of the main steps. As usual, $\pi(x; k, q-1)$ denotes the number of primes $p \le x$ with $p \equiv k \bmod q-1$.

Proposition 2.2. *Suppose that $0 < \nu < 1/2$ and $0 < \eta < \nu/2$. Then, for every k with $(k, q-1) = 1$, we have*

$$\sum_{p\le x,\ p\equiv k \bmod q-1} e(\alpha s_q(p)) = \pi(x; k, q-1)\, e(\alpha \mu_q \log_q x) \\ \times (e^{-2\pi^2\alpha^2\sigma_q^2 \log_q x}(1 + O(\alpha^4 \log x)) + O(|\alpha|\,(\log x)^{\nu})) \qquad (7)$$

uniformly for all real α with $|\alpha| \le (\log x)^{\eta-1/2}$.

Finally, the proof of Theorem 1.1 is obtained in § 5 by evaluating asymptotically the integral

$$\#\{p \le x : s_q(p) = k\} = \int_{-1/2}^{1/2} \biggl(\sum_{p\le x} e(\alpha s_q(p))\biggr) e(-\alpha k)\, d\alpha, \qquad (8)$$

using both the analytic estimates coming from Proposition 2.1 and the probabilistic ideas contained in Proposition 2.2.

Theorem 1.2 is then a corollary of Theorem 1.1.

3. Proof of Proposition 2.1

We denote by $\Lambda(n)$ the von Mangoldt function defined by $\Lambda(n) = \log p$ if $n = p^k$ with p prime and k a positive integer, and $\Lambda(n) = 0$ otherwise.

The proof of Proposition 2.1 is based on methods from [MR05]. More precisely, we need to obtain a bound for $\sum_{p\le x} e(\alpha s_q(p))$ that is uniform in terms of α and x.

First, note that by partial summation (see, for example, [MR05, Lemma 11]), it suffices to prove that for every fixed integer $q \ge 2$ there exists a constant $c_1 > 0$ such that

$$\biggl|\sum_{n\le x} \Lambda(n) e(\alpha s_q(n))\biggr| \ll (\log x)^4 x^{1-c_1\|(q-1)\alpha\|^2} \qquad (9)$$

uniformly for real α.

Actually, we will prove (9) only for α with $\|(q-1)\alpha\| \ge c_2(\log x)^{-1/2}$, where $c_2 > 0$ is a suitably chosen constant. If $\|(q-1)\alpha\| < c_2(\log x)^{-1/2}$, then (9) is trivially satisfied.

3.1 A combinatorial identity

A classical method [Hoh30, Vin54] for dealing with sums of the form $\sum_n \Lambda(n) g(n)$ is to transform them into sums like

$$\sum_{n_1,\dots,n_k} a_1(n_1)\cdots a_k(n_k) g(n_1\cdots n_k),$$

where $n_1, \dots, n_k$ satisfy multiplicative conditions. Vaughan gave an elegant formulation of this method [Vau80], which was later generalized by Heath-Brown [Hea82].

A drawback of these methods in their original setting is the emergence of several arithmetic functions involving divisors, which cannot be individually majorized by a logarithmic factor. We will use a slight variant of Vaughan's method [IK04] which allows us to circumvent this difficulty.

LEMMA 3.1. *Let $q \geq 2$, $x \geq q^2$, $0 < \beta_1 < 1/3$ and $1/2 < \beta_2 < 1$. Let g be an arithmetic function. Suppose that, uniformly for all complex numbers a_m, b_n with $|a_m| \leq 1$ and $|b_n| \leq 1$, we have*

$$\sum_{M/q<m\leq M} \max_{x/(qm)\leq t\leq x/m} \Big| \sum_{t<n\leq x/m} g(mn) \Big| \leq U \quad \text{for } M \leq x^{\beta_1} \quad \text{(type I)}, \tag{10}$$

$$\Big| \sum_{M/q<m\leq M} \sum_{x/(qm)<n\leq x/m} a_m b_n g(mn) \Big| \leq U \quad \text{for } x^{\beta_1} \leq M \leq x^{\beta_2} \quad \text{(type II)}. \tag{11}$$

Then

$$\Big| \sum_{x/q<n\leq x} \Lambda(n) g(n) \Big| \ll U\, (\log x)^2.$$

Proof. This is [MR05, Lemma 1]. □

Thus, in order to obtain upper bounds for (9), it is sufficient to get bounds for sums of types I and II, i.e. (10) and (11), for $g(n) = e(\alpha s_q(n))$. The next lemma reduces the problem of bounding type-II sums to a slightly simpler problem.

LEMMA 3.2. *Let g be an arithmetic function, and take $q \geq 2$, $0 < \delta < \beta_1 < 1/3$ and $1/2 < \beta_2 < 1$. Suppose that, uniformly for all complex numbers b_n with $|b_n| \leq 1$, we have*

$$\sum_{q^{\mu-1}<m\leq q^{\mu}} \Big| \sum_{q^{\nu-1}<n\leq q^{\nu}} b_n\, g(mn) \Big| \leq V \tag{12}$$

whenever

$$\beta_1 - \delta \leq \frac{\mu}{\mu+\nu} \leq \beta_2 + \delta. \tag{13}$$

Then, for $x > x_0 := \max(q^{1/(1-\beta_2)}, q^{3/\delta})$ we have, uniformly for all M such that

$$x^{\beta_1} \leq M \leq x^{\beta_2}, \tag{14}$$

the estimate (11) with $U = (12/\pi)(1 + \log 2x)\, V$.

Proof. This is [MR05, Lemma 3]. □

3.2 Type-I sums

Fortunately, type-I sums are easy to deal with because the corresponding upper bounds obtained in [MR05] are already uniform in α and x.

PROPOSITION 3.1. *For $q \geq 2$, $x \geq 2$ and every α such that $(q-1)\alpha \in \mathbb{R} \setminus \mathbb{Z}$, we have*

$$\sum_{M/q<m\leq M} \max_{x/(qm)\leq t\leq x/m} \Big| \sum_{t<n\leq x/m} e(\alpha\, s_q(mn)) \Big| \ll_q x^{1-\kappa_q(\alpha)} \log x \tag{15}$$

for $1 \leq M \leq x^{1/3}$ and

$$0 < \kappa_q(\alpha) := \min(\tfrac{1}{6}, \tfrac{1}{3}(1 - \gamma_q(\alpha))), \tag{16}$$

where $1/2 \leq \gamma_q(\alpha) < 1$ is defined by

$$q^{\gamma_q(\alpha)} = \max_{t\in\mathbb{R}} \sqrt{\varphi_q(\alpha+t)\, \varphi_q(\alpha+qt)}$$

with

$$\varphi_q(t) = \begin{cases} |\sin \pi q t|/|\sin \pi t| & \text{if } t \in \mathbb{R} \setminus \mathbb{Z}, \\ q & \text{if } t \in \mathbb{Z}. \end{cases}$$

Proof. This is [MR05, Proposition 2]. □

3.3 Type-II sums

To verify (11) we use Lemma 3.2, that is, we will prove the following proposition (which is a variant of [MR05, Proposition 1]).

Proposition 3.2. *For $q \geq 2$ and any α with $(q-1)\alpha \in \mathbb{R} \setminus \mathbb{Z}$, there exist β_1, β_2 and δ satisfying $0 < \delta < \beta_1 < 1/3$ and $1/2 < \beta_2 < 1$ together with $\xi_q(\alpha) > 0$ such that, uniformly for all complex numbers b_n with $|b_n| \leq 1$, we have*

$$\sum_{q^{\mu-1} < m \leq q^\mu} \left| \sum_{q^{\nu-1} < n \leq q^\nu} b_n\, e(\alpha s_q(mn)) \right| \ll_q (\mu+\nu) q^{(1-\xi_q(\alpha)/2)(\mu+\nu)} \tag{17}$$

whenever

$$\beta_1 - \delta \leq \frac{\mu}{\mu+\nu} \leq \beta_2 + \delta.$$

We note that the constants β_1, β_2, δ and $\xi_q(\alpha)$ can be stated explicitly in terms of α, as shown in (24)–(28), so that (17) is actually an explicit estimate that is uniform in α.

The proof of Proposition 3.2 is divided into several steps. We first apply the Cauchy–Schwarz inequality and a Van der Corput-type inequality in order to *smooth the sums*.

For $q \geq 2$ and $\alpha \in \mathbb{R}$, let

$$f(n) = \alpha s_q(n).$$

Further, let μ, ν and ρ be integers such that $\mu \geq 1$, $\nu \geq 1$ and $0 \leq \rho \leq \nu/2$, and let b_n be complex numbers with $|b_n| \leq 1$. We consider the sum

$$S = \sum_{q^{\mu-1} < m \leq q^\mu} \left| \sum_{q^{\nu-1} < n \leq q^\nu} b_n\, e(f(mn)) \right|.$$

By the Cauchy–Schwarz inequality,

$$|S|^2 \leq q^\mu \sum_{q^{\mu-1} < m \leq q^\mu} \left| \sum_{q^{\nu-1} < n \leq q^\nu} b_n\, e(f(mn)) \right|^2. \tag{18}$$

This sum will be further estimated by applying the following version of Van der Corput's inequality.

Lemma 3.3. *Let $z_1, \ldots, z_N$ be complex numbers. For any integer $R \geq 1$, we have*

$$\left| \sum_{1 \leq n \leq N} z_n \right|^2 \leq \frac{N+R-1}{R} \sum_{|r|<R} \left(1 - \frac{|r|}{R}\right) \sum_{\substack{1 \leq n \leq N \\ 1 \leq n+r \leq N}} z_{n+r} \overline{z_n}.$$

Proof. See, for example, [MR05, Lemme 4]. □

Taking $R = q^\rho$, $N = q^\nu - q^{\nu-1}$ and $z_n = b_{q^{\nu-1}+n} e(f(m(q^{\nu-1}+n)))$ in Lemma 3.3 and observing that $\rho \le \lfloor \nu/2 \rfloor \le \nu - 1$, we obtain

$$\begin{aligned}\Big| \sum_{q^{\nu-1}<n\le q^\nu} & b_n\, e(f(mn)) \Big|^2 \\ &\le q^{\nu-\rho} \sum_{|r|<q^\rho} \Big(1 - \frac{|r|}{q^\rho}\Big) \Big(\sum_{q^{\nu-1}<n\le q^\nu} b_{n+r}\, \overline{b_n}\, e(f(m(n+r)) - f(mn)) + O(q^\rho) \Big),\end{aligned}$$

where the term $O(q^\rho)$ comes from the removal of the condition of summation $q^{\nu-1} < n + r \le q^\nu$ introduced by Lemma 3.3. Indeed, this removal potentially gives $O(q^\rho)$ values of n, and each term in the sum is of modulus less than or equal to 1, leading to an error of at most $O(q^\rho)$. We separate the cases $r = 0$ and $r \ne 0$, obtaining

$$|S|^2 \ll q^{2(\mu+\nu)-\rho} + q^{\mu+\nu} \max_{1\le|r|<q^\rho} \sum_{q^{\nu-1}<n\le q^\nu} \Big| \sum_{q^{\mu-1}<m\le q^\mu} e(f(m(n+r)) - f(mn)) \Big|,$$

where we have taken into account the fact that the contribution of $O(q^\rho)$ is $O(q^{2\mu+\nu+\rho})$, which is negligible in comparison with $O(q^{2(\mu+\nu)-\rho})$ since $\rho \le \nu/2$.

In order to continue the proof, we will show that only the digits of low weight in the difference $f(m(n + r)) - f(mn)$ make a significant contribution. We therefore introduce the notion of *truncated sum of digits* and show that, in sums of type II, we can replace the function f by this truncated function.

For any integer $\lambda \ge 0$, we define f_λ by the formula

$$f_\lambda(n) = \sum_{k<\lambda} f(\varepsilon_k(n)\, q^k) = \alpha \sum_{k<\lambda} \varepsilon_k(n), \tag{19}$$

where the $\varepsilon_k(n)$ are integers representing the digits of n in base q. The function f_λ is clearly periodic with period q^λ. This truncated function appears in a different context in [DR05], where Drmota and Rivat study some properties of $f_\lambda(n^2)$ with λ being of order $\log n$. The following lemma is a variant of [MR05, Lemme 5].

Lemma 3.4. *For all integers μ, ν, ρ with $\mu > 0$, $\nu > 0$, $0 \le \rho \le \nu/2$ and all $r \in \mathbb{Z}$ with $|r| < q^\rho$, we denote by $E(r, \mu, \nu, \rho)$ the number of pairs $(m, n) \in \mathbb{Z}^2$ such that $q^{\mu-1} < m \le q^\mu$, $q^{\nu-1} < n \le q^\nu$ and*

$$f(m(n + r)) - f(mn) \ne f_{\mu+2\rho}(m(n + r)) - f_{\mu+2\rho}(mn).$$

Then, if μ and ν satisfy the condition

$$\frac{27}{82} < \frac{\mu}{\mu+\nu}, \tag{20}$$

we have

$$E(r, \mu, \nu, \rho) \ll (\mu + \nu)(\log q)\; q^{\mu+\nu-\rho}. \tag{21}$$

Proof. Suppose $0 \le r < q^\rho$. In this case, $0 \le mr < q^{\mu+\rho}$. When we compute the sum $mn + mr$, the digits of the product mn with index greater than or equal to $\mu + \rho$ cannot be modified unless there is a carry propagation. Hence we must count the number of pairs (m, n) such that the digits a_j in basis q of the product $a = mn$ satisfy $a_j = q - 1$ for $\mu + \rho \le j < \mu + 2\rho$. Therefore, grouping the products mn according to their value a, we obtain

$$E(r,\mu,\nu,\rho) \le \sum_{q^{\mu+\nu-2}<a\le q^{\mu+\nu}} \tau(a)\,\chi(a);$$

here $\tau(a)$ denotes the number of divisors of a, and χ is defined by $\chi(a)=1$ if the digits a_j in base q of a satisfy $a_j=q-1$ for $\mu+\rho\le j<\mu+2\rho$, and $\chi(a)=0$ in the opposite case, i.e. if there exists an index j with $\mu+\rho\le j<\mu+2\rho$ for which $a_j\ne q-1$. We deduce that

$$E(r,\mu,\nu,\rho) \le \sum_{b<q^{\mu+\rho}} \sum_{c<q^{\nu-2\rho}} \tau(b+(q-1)q^{\mu+\rho}+\cdots+(q-1)q^{\mu+2\rho-1}+q^{\mu+2\rho}c).$$

For each fixed c, we apply Lemma 3.5 below with

$$x = q^{\mu+\rho}-1+(q-1)q^{\mu+\rho}+\cdots+(q-1)q^{\mu+2\rho-1}+q^{\mu+2\rho}c \le q^{\mu+\nu},$$
$$y = q^{\mu+\rho}$$

(by (20) we have $x^{27/82}\le q^{(27/82)(\mu+\nu)}\le y\le x$), to obtain

$$E(r,\mu,\nu,\rho) \ll q^{\nu-2\rho}q^{\mu+\rho}\log q^{\mu+\nu} = (\mu+\nu)(\log q)q^{\mu+\nu-\rho}.$$

The same argument can be applied whenever $-q^\rho<r<0$, counting the pairs (m,n) such that the digits a_j of the product $a=mn$ satisfy $a_j=0$ for $\mu+\rho\le j<\mu+2\rho$, and we obtain the same upper bound (21). □

Lemma 3.5. *For $x^{27/82}\le y\le x$, we have*

$$\sum_{x-y<n\le x} \tau(n) = O(y\log x).$$

Proof. It follows from Van der Corput's method of exponential sums (see, for example, [GK91, Theorem 4.6]) that

$$\sum_{n\le x}\tau(n) = x\log x+(2\gamma-1)x+O(x^{27/82}) = \int_0^x \log t\,dt + 2\gamma\,x + O(x^{27/82}),$$

where γ is Euler's constant. As a consequence, we have

$$\sum_{x-y<n\le x}\tau(n) = \int_{x-y}^x \log t\,dt + 2\gamma\,y + O(x^{27/82}) + O((x-y)^{27/82}) = O(y\log x). \qquad \square$$

Using Lemma 3.4, we may now replace f in the upper bound (18) by the truncated function $f_{\mu+2\rho}$ defined in (19), at the price of a total error $O((\mu+\nu)(\log q)\,q^{2(\mu+\nu)-\rho})$. Thus, if (20) holds, then

$$|S|^2 \ll (\mu+\nu)(\log q)\,q^{2(\mu+\nu)-\rho} + q^{\mu+\nu}\max_{1\le|r|<q^\rho} S_2(r,\mu,\nu,\rho), \tag{22}$$

where

$$S_2(r,\mu,\nu,\rho) := \sum_{q^{\nu-1}<n\le q^\nu}\left|\sum_{q^{\mu-1}<m\le q^\mu} e(f_{\mu+2\rho}(m(n+r))-f_{\mu+2\rho}(mn))\right|. \tag{23}$$

The sum $S_2(r, \mu, \nu, \rho)$ has been studied in [MR05]. For $q \geq 2$ and $(q-1)\alpha \in \mathbb{R} \setminus \mathbb{Z}$, let us introduce some notation from [MR05]. We write

$$\omega_2 = 1 - \frac{\log(2+\sqrt{2})}{2\log 2},$$
$$\omega_q = \left(\frac{3}{2} - \frac{\log 5}{\log 3}\right)\frac{\log 2}{\log q} \quad \text{for } q \geq 3,$$
$$\tau_q(\alpha) = \min\left(\omega_q, -\frac{2\log(\varphi_q(\alpha)/q)}{\log q}\right) \quad \text{for } q \geq 2,$$

where $\varphi_q(t)$ is defined as in Proposition 3.1; also, let

$$\epsilon_q(\alpha) := \min(\tau_q(\alpha), 1 - \gamma_q(\alpha)) \quad \text{for } q \geq 2,$$

where $\gamma_q(t)$ is defined in Proposition 3.1. In addition, define

$$\xi_q(\alpha) := \frac{\epsilon_q(\alpha)}{14}, \quad \delta := \frac{\epsilon_q(\alpha)}{28}, \tag{24}$$
$$\beta_1 := \frac{(3 - 2\epsilon_q(\alpha))\xi_q(\alpha)}{\epsilon_q(\alpha)} + \delta \quad \text{for } q = 2, \tag{25}$$
$$\beta_1 := \frac{(4 - 2\epsilon_q(\alpha))\xi_q(\alpha)}{\epsilon_q(\alpha)} + \delta \quad \text{for } q \geq 3, \tag{26}$$
$$\beta_2 := \frac{1 - (5 - 2\epsilon_q(\alpha))\xi_q(\alpha)}{2 - \epsilon_q(\alpha)} - \delta \quad \text{for } q = 2, \tag{27}$$
$$\beta_2 := \frac{1 - (6 - 2\epsilon_q(\alpha))\xi_q(\alpha)}{2 - \epsilon_q(\alpha)} - \delta \quad \text{for } q \geq 3. \tag{28}$$

It was shown in [MR05, Paragraph 7.3] that $0 < \delta < \beta_1 < 1/3$, $1/2 < \beta_2 < 1$ and that for any integers $\mu > 0$ and $\nu > 0$ satisfying

$$\beta_1 - \delta < \frac{\mu}{\mu + \nu} \leq \beta_2 + \delta$$

we have, for every $\rho \leq \xi_q(\alpha)(\mu + \nu)$,

$$S_2(r, \mu, \nu, \rho) \ll_q (\mu + \nu)^2 q^{\mu+\nu-\rho}. \tag{29}$$

Let us remark that for any $\alpha \in \mathbb{R}$, we have $\varphi_q(\alpha) \leq q^{\gamma_q(\alpha)}$ so that

$$\begin{aligned}\tau_q(\alpha) &= \min\left(\omega_q, -\frac{2\log(\varphi_q(\alpha)/q)}{\log q}\right) \\ &\geq \min\left(\omega_q, -\frac{2\log(q^{\gamma_q(\alpha)-1})}{\log q}\right) = \min(\omega_q, 2(1 - \gamma_q(\alpha)))\end{aligned}$$

and

$$\xi_q(\alpha) = \frac{1}{14}\min(\omega_q, 1 - \gamma_q(\alpha)). \tag{30}$$

Furthermore, by [MR07, Lemma 7],

$$\gamma_q(\alpha) \leq 1 - \frac{\pi^2}{12}\frac{q-1}{(q+1)\log q}\|(q-1)\alpha\|^2,$$

so that

$$\xi_q(\alpha) \geq \frac{1}{14}\min\left(\omega_q, \frac{\pi^2}{12}\frac{q-1}{(q+1)\log q}\|(q-1)\alpha\|^2\right) \geq 2c_1\|(q-1)\alpha\|^2 \tag{31}$$

for

$$c_1 := \frac{1}{28} \min\left(4\omega_q, \frac{\pi^2}{12} \frac{q-1}{(q+1)\log q}\right).$$

It follows from (22) that

$$|S|^2 \ll_q (\mu+\nu)^2 q^{2\mu+2\nu-\rho}$$

for $\rho \le 2c_1 \|(q-1)\alpha\|^2(\mu+\nu)$; so

$$|S| \ll_q (\mu+\nu)\ q^{(1-c_1\|(q-1)\alpha\|^2)(\mu+\nu)},$$

which ends the proof of Proposition 3.2.

We are now able to complete the estimate for type-II sums. It follows from Proposition 3.2 that we can apply Lemma 3.2 with $g(n) = e(\alpha s_q(n))$ and some V such that

$$V \ll_q (\mu+\nu)\ q^{(1-c_1\|(q-1)\alpha\|^2)(\mu+\nu)} \ll_q (\log x)\ x^{1-c_1\|(q-1)\alpha\|^2}.$$

This shows that for $x > x_0 = \max(q^{1/(1-\beta_2)}, q^{3/\delta})$ we have, uniformly for M such that

$$x^{\beta_1} \le M \le x^{\beta_2},$$

the estimate

$$\left|\sum_{M/q<m\le M}\ \sum_{x/(qm)<n\le x/m} a_m b_n g(mn)\right| \le \frac{12}{\pi}(1+\log 2x)\ V \ll_q (\log x)^2\ x^{1-c_1\|(q-1)\alpha\|^2}. \tag{32}$$

It now follows from [MR05, Paragraph 7.3] that the values of β_1, β_2 and δ in Proposition 3.2 lead to taking $x_0 \ge q^{6/\xi_q(\alpha)}$. By (31), we have $6/\xi_q(\alpha) \le 3/(c_1\|(q-1)\alpha\|^2)$; thus we can take

$$x_0 := q^{3/(c_1\|(q-1)\alpha\|^2)}. \tag{33}$$

3.4 Proof of Proposition 2.1

In order to prove Proposition 2.1, we apply Lemma 3.1. Indeed, Proposition 3.1 shows that (10) holds for any $x \ge 2$ with some U such that

$$U \ll_q x^{1-\kappa_q(\alpha)} \log x \ll_q x^{1-c_1\|(q-1)\alpha\|^2} \log x$$

(the second upper bound follows from (31), (30) and (16)), while (32) shows that (11) holds for any $x > x_0$ with some U such that

$$U \ll_q x^{1-c_1\|(q-1)\alpha\|^2} (\log x)^2.$$

From Lemma 3.1 it follows that for $x > x_0$,

$$\left|\sum_{x/q<n\le x} \Lambda(n) g(n)\right| \ll_q x^{1-c_1\|(q-1)\alpha\|^2} (\log x)^4.$$

By (33), the condition $x > x_0$ is equivalent to $\|(q-1)\alpha\| \ge c_2(\log x)^{-1/2}$ with $c_2 = \sqrt{3\log q/c_1}$; so we have established (9), which completes the proof of Proposition 2.1.

4. Proof of Proposition 2.2

To prove Proposition 2.2, we will approximate the sum-of-digits function by a sum of independent random variables.

4.1 Approximation of $s_q(p)$ by sums of independent random variables

We fix some residue class $k \bmod q-1$ with $(k, q-1) = 1$, and for (sufficiently large) $x \geq 2$ we consider the set of primes

$$\{p \in \mathbb{P} : p \leq x,\ p \equiv k \bmod q-1\}.$$

The cardinality of this set is denoted by $\pi(x; k, q-1)$, and it is well-known that asymptotically,

$$\pi(x; k, q-1) = \frac{\pi(x)}{\varphi(q-1)}(1 + O((\log x)^{-1})) = \frac{1}{\varphi(q-1)}\frac{x}{\log x}\,(1 + O((\log x)^{-1})).$$

If we assume that every prime in this set is equally likely, then the sum-of-digits function $s_q(p)$ can be interpreted as a random variable

$$S_x = S_x(p) = s_q(p) = \sum_{j \leq \log_q x} \varepsilon_j(p).$$

Of course, $D_j = D_{j,x} = \varepsilon_j$, the jth digit, is also a random variable.

We can now reformulate Proposition 2.2. Set $L = \log_q x$. Then the asymptotic formula (7) is equivalent to the relation

$$\varphi_1(t) := \mathbb{E}\, e^{it(S_x - L\mu_q)/(L\sigma_q^2)^{1/2}} = e^{-t^2/2}\left(1 + O\left(\frac{t^4}{\log x}\right)\right) + O\left(\frac{|t|}{(\log x)^{\frac{1}{2}-\nu}}\right), \tag{34}$$

which holds uniformly for $|t| \leq (\log x)^{\eta}$. We just have to set $\alpha = t/(2\pi\sigma_q(\log_q x)^{1/2})$.

For technical reasons, we need to truncate this sum-of-digits expression appropriately. Set $L' = \#\{j \in \mathbb{Z} : L^{\nu} \leq j \leq L - L^{\nu}\} = L - 2L^{\nu} + O(1)$, where $0 < \nu < 1/2$ is fixed, and let

$$T_x = T_x(p) = \sum_{L^{\nu} \leq j \leq L - L^{\nu}} \varepsilon_j(p) = \sum_{L^{\nu} \leq j \leq L - L^{\nu}} D_j.$$

First, we observe that $\varphi_1(t)$ and

$$\varphi_2(t) := \mathbb{E}\, e^{it(T_x - L'\mu_q)/(L'\sigma_q^2)^{1/2}}$$

do not differ essentially.

Lemma 4.1. *We have, uniformly for all real t,*

$$|\varphi_1(t) - \varphi_2(t)| = O\left(\frac{|t|}{(\log x)^{1/2-\nu}}\right).$$

Proof. We only have to observe that $|L - L'| \ll L^{\nu}$, $\|S_x - T_x\|_\infty \ll L^{\nu}$, $\|S_x\|_\infty \ll L$ and $|e^{it} - e^{is}| \leq |t - s|$. Consequently,

$$\begin{aligned}
|\varphi_1(t) - \varphi_2(t)| &\leq |t|\,\mathbb{E}\left|\frac{S_x - L\mu_q}{(L\sigma_q^2)^{1/2}} - \frac{T_x - L'\mu_q}{(L'\sigma_q^2)^{1/2}}\right| \\
&\ll |t|\left(\frac{\|S_x - T_x\|_\infty}{L^{1/2}} + \frac{|L - L'|}{L^{1/2}} + \|S_x\|_\infty\left(\frac{1}{L'^{1/2}} - \frac{1}{L^{1/2}}\right)\right) \\
&\ll \frac{|t|}{(\log x)^{1/2-\nu}}.
\end{aligned}$$

This proves the lemma. □

We shall now approximate T_x by a sum $\overline{T}_x$ of independent random variables. Let Z_j ($j \ge 0$) be a sequence of independent random variables with range $\{0, 1, \ldots, q-1\}$ and uniform probability distribution

$$\mathbb{P}\{Z_j = \ell\} = \frac{1}{q}.$$

We then set

$$\overline{T}_x := \sum_{L^\nu \le j \le L - L^\nu} Z_j.$$

Note that the expected value and the variance of $\overline{T}_x$ are given exactly by

$$\mathbb{E}\, \overline{T}_x = L'\mu_q \quad \text{and} \quad \mathbb{V}\, \overline{T}_x = L'\sigma_q^2.$$

Since $\overline{T}_x$ is the sum of independent identically distributed random variables, it is clear that $\overline{T}_x$ satisfies a central limit theorem. For the reader's convenience, we state the following well-known property.

Lemma 4.2. *The characteristic function of the normalized random variable $\overline{T}_x$ is given by*

$$\varphi_3(t) := \mathbb{E}\, e^{it(\overline{T}_x - L'\mu_q)/(L'\sigma_q^2)^{1/2}} = e^{-t^2/2}\left(1 + O\left(\frac{t^4}{\log x}\right)\right), \tag{35}$$

which also holds uniformly for $|t| \le (\log x)^{1/4}$.

Proof. First, note that

$$\begin{aligned}\mathbb{E}\, v^{\overline{T}_x} &= \prod_{L^\nu \le j \le L - L^\nu} \mathbb{E} v^{Z_j} \\ &= q^{-L'}(1 + v + v^2 + \cdots + v^{q-1})^{L'}.\end{aligned}$$

Now (35) follows upon setting

$$v = e^{it/(L'\sigma_q^2)^{1/2}}$$

and using the Taylor expansion

$$\log\left(\frac{1 + e^{is} + \cdots + e^{is(q-1)}}{q}\right) = i\mu_q s - \frac{1}{2}\sigma_q^2 s^2 + O(s^4).$$

Note that there are no odd powers of s (besides the linear one), since the random variables Z_j are symmetric with respect to their mean. □

Thus, it remains to compare $\varphi_2(t)$ and $\varphi_3(t)$. To do this, we first prove the following bound.

Proposition 4.1. *Suppose that η and κ satisfy $0 < 2\eta < \kappa < \nu$. Then we have, uniformly for all real t with $|t| \le L^\eta$,*

$$|\varphi_2(t) - \varphi_3(t)| = O(|t| e^{-c_1 L^\kappa}),$$

where c_1 is a certain positive constant that depends on η and κ.

Note that $e^{-c_1 L^\kappa} \ll L^{-1}$. Therefore, Proposition 4.1 (together with Lemmas 4.1 and 4.2) immediately implies (34) and hence Proposition 2.2.

4.2 Comparision of moments

In what follows, we will use the well-known bound on exponential sums over primes given in the next lemma.

Lemma 4.3. *For $x > 0$, $0 \le K \le \frac{2}{5}\log_q x$, Q an integer with $q^K \le Q \le x\, q^{-K}$ and A an integer that is coprime with Q, we have*

$$\sum_{p \le x} e\left(\frac{A}{Q}\, p\right) \ll (\log x)^2\, x\, q^{-K/2},$$

where the implied constant is absolute.

Proof. We just need to apply a partial summation and the estimate in [IK04, Theorem 13.6]. □

Lemma 4.4. *Let $0 < \Delta < 1$ and*

$$U_\Delta := [0, \Delta] \cup \bigcup_{\ell=1}^{q-1} \left[\frac{\ell}{q} - \Delta, \frac{\ell}{q} + \Delta\right] \cup [1-\Delta, 1].$$

Then, for $L^\nu \le j \le L - L^\nu$ and $0 < \Delta < 1/(2q)$ we have, uniformly, that

$$\frac{1}{\pi(x; k, q-1)} \#\left\{p < x : p \equiv k \bmod q-1,\ \left\{\frac{p}{q^{j+1}}\right\} \in U_\Delta\right\} \ll \Delta + e^{-c_3 L^\nu} \tag{36}$$

as $x \to \infty$, where c_3 is a certain positive constant.

Proof. It suffices to show that the discrepancy D between the sequence (pq^{-j-1}), where p ranges over all primes $p \le x$, and $p \equiv k \bmod q-1$ is bounded above, with $D \ll e^{-c_3 L^\nu}$. The bound (36) then follows immediately.

We use the Erdős–Turán inequality which says that

$$D \ll \frac{1}{H} + \sum_{h=1}^{H} \frac{1}{h} \left| \frac{1}{\pi(x; k, q-1)} \sum_{p \le x, p \equiv k \bmod q-1} e\left(\frac{h}{q^{j+1}} p\right)\right|,$$

where $H > 0$ can be arbitrarily chosen. For our purpose here, we will use $H = \lfloor e^{cL^\nu} \rfloor$ (for a suitable constant $c > 0$).

First of all, recall that

$$\sum_{p \le x, p \equiv k \bmod q-1} e(\alpha p) = \frac{1}{q-1} \sum_{\ell=0}^{q-2} e\left(-\frac{k\ell}{q-1}\right) \sum_{p \le x} e\left(\left(\alpha + \frac{\ell}{q-1}\right) p\right).$$

Thus, we actually need to estimate exponential sums of the particular form

$$\sum_{p \le x} e\left(\left(\frac{h}{q^{j+1}} + \frac{\ell}{q-1}\right) p\right).$$

Let us write the rational number in the exponent as

$$\frac{h}{q^{j+1}} + \frac{\ell}{q-1} = \frac{A}{Q},$$

where $(A,Q)=1$. Then $Q \ge q^{j+1}/H$. Hence we can apply Lemma 4.3 with $K = 2L^\nu/3$ and finally obtain, with $H = \lfloor q^{L^\nu/3} \rfloor$, that

$$\begin{aligned} D &\ll \frac{1}{H} + \frac{L}{x} \sum_{h=1}^{H} \frac{1}{h} L^2 \, x \, q^{-L^\nu/3} \\ &\ll \frac{1}{H} + L^4 q^{-L^\nu/3} \\ &\ll e^{-c_3 L^\nu}, \end{aligned}$$

where $c_3 < (\log q)/3$. This completes the proof of the lemma. □

The key property to be used for comparing moments of T_x and $\overline{T}_x$ is given in the following lemma. Note that the essential difference from [BK95] is that the estimate in Lemma 4.5 is uniform for all $1 \le d \le L'$.

Lemma 4.5. *Let $1 \le d \le L'$, and let $j_1, j_2, \ldots, j_d$ and $\ell_1, \ell_2, \ldots, \ell_d$ be integers satisfying*

$$L^\nu \le j_1 < j_2 < \cdots < j_d \le L - L^\nu$$

and

$$\ell_1, \ell_2, \ldots, \ell_d \in \{0, 1, \ldots, q-1\}.$$

Then, uniformly, we have

$$\begin{aligned} &\frac{1}{\pi(x; k, q-1)} \#\{p \le x : p \equiv k \bmod q-1, \epsilon_{j_1}(p) = \ell_1, \ldots, \epsilon_{j_d}(p) = \ell_d\} \\ &\quad = q^{-d} + O((4L^\nu)^d e^{-c_4 L^\nu}), \end{aligned}$$

where c_4 is a certain positive constant.

Remark 2. Note that Lemma 4.5 can also be interpreted as

$$\begin{aligned} &\mathbb{P}\{D_{j_1,x} = \ell_1, \ldots, D_{j_d,x} = \ell_d\} \\ &\quad = \mathbb{P}\{Z_{j_1} = \ell_1, \ldots, Z_{j_d} = \ell_d\} + O((4L^\nu)^d e^{-c_4 L^\nu}). \end{aligned} \tag{37}$$

This means that the joint probability distribution of the summands of T_x and that of the summands of $\overline{T}_x$ are very close. Note further that (37) remains valid when $j_1, j_2, \ldots, j_d$ are not ordered and even when they are not distinct.

Proof. Let $f_{\ell,\Delta}(x)$ be defined by

$$f_{\ell,\Delta}(x) := \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} \mathbf{1}_{[\ell/q,\,(\ell+1)/q]}(\{x+z\}) \, dz,$$

where $\mathbf{1}_A$ denotes the characteristic function of the set A. The Fourier coefficients of the Fourier series $f_{\ell,\Delta}(x) = \sum_{m\in\mathbb{Z}} d_{m,\ell,\Delta} e(mx)$ are given by

$$d_{0,\ell,\Delta} = \frac{1}{q}$$

and, for $m \ne 0$,

$$d_{m,\ell,\Delta} = \frac{e(-m\ell/q) - e(-m(\ell+1)/q)}{2\pi i m} \cdot \frac{e(m\Delta/2) - e(-m\Delta/2)}{2\pi i m \Delta}.$$

Note that $d_{m,\ell,\Delta} = 0$ if $m \neq 0$ and $m \equiv 0 \bmod q$; also note that

$$|d_{m,\ell,\Delta}| \leq \min\left(\frac{1}{\pi|m|}, \frac{1}{\Delta\pi m^2}\right).$$

By definition, we have $0 \leq f_{\ell,\Delta}(x) \leq 1$ and

$$f_{\ell,\Delta}(x) = \begin{cases} 1 & \text{if } x \in \left[\frac{\ell}{q} + \Delta, \frac{\ell+1}{q} - \Delta\right], \\ 0 & \text{if } x \in [0,1] \setminus \left[\frac{\ell}{q} - \Delta, \frac{\ell+1}{q} + \Delta\right]. \end{cases}$$

So if we set

$$t_{\mathbf{l},\mathbf{j}}(y_1, \ldots, y_d) := \prod_{i=1}^{d} f_{\ell_i,\Delta}\left(\frac{y_i}{q^{j_i+1}}\right)$$

where $\mathbf{l} = (\ell_1, \ldots, \ell_d)$ and $\mathbf{j} = (j_1, \ldots, j_d)$, then we get, for $\Delta < 1/(2q)$, that

$$\left|\#\{p \leq x : p \equiv k \bmod q-1, \epsilon_{j_1}(p) = \ell_1, \ldots, \epsilon_{j_d}(p) = \ell_d\} - \sum_{p<x,\, p\equiv k \bmod q-1} t_{\mathbf{l},\mathbf{j}}(p, \ldots, p)\right|$$
$$\leq d \cdot \max_{L^\nu \leq j \leq L - L^\nu} \#\left\{p \leq x : p \equiv k \bmod q-1, \left\{\frac{p}{q^{j+1}}\right\} \in U_\Delta\right\}$$
$$\ll d\, \pi(x)(\Delta + e^{-c_3 L^\nu}).$$

The third line above follows from Lemma 4.4.

For convenience, let $\mathbf{m} = (m_1, \ldots, m_d)$,

$$\mathbf{v_j} = (q^{-j_1-1}, \ldots, q^{-j_d-1})$$

and

$$d_{\mathbf{m},\mathbf{l},\Delta} := \prod_{i=1}^{d} d_{m_i,\ell_i,\Delta}.$$

Then $t_{\mathbf{l},\mathbf{j}}(y_1, \ldots, y_d)$ has Fourier series expansion

$$t_{\mathbf{l},\mathbf{j}}(y_1, \ldots, y_d) = \sum_{\mathbf{m}} d_{\mathbf{m},\mathbf{l},\Delta}\, e(m_1 q^{-j_1-1} y_1 + \cdots + m_d q^{-j_d-1} y_d).$$

Thus, we are led to consider the exponential sum

$$\begin{aligned} S &= \sum_{p<x,\, p\equiv k \bmod q-1} t_{\mathbf{l},\mathbf{j}}(p, \ldots, p) \\ &= \sum_{\mathbf{m}} d_{\mathbf{m},\mathbf{l},\Delta} \sum_{p<x,\ p\equiv k \bmod q-1} e((m_1 q^{-j_1-1} + \cdots + m_d q^{-j_d-1})p) \\ &= \frac{1}{q-1} \sum_{r=0}^{q-2} e\left(-\frac{rk}{q-1}\right) \sum_{\mathbf{m}} d_{\mathbf{m},\mathbf{l},\Delta} \sum_{p \leq x} e\left(\left(\mathbf{m} \cdot \mathbf{v_j} + \frac{r}{q-1}\right)p\right). \end{aligned}$$

If $\mathbf{m} = (0, \ldots, 0)$, then

$$d_{\mathbf{0},\mathbf{l},\Delta} \sum_{p<x,\ p\equiv k \bmod q-1} e(0) = \frac{\pi(x; k, q-1)}{q^d},$$

which provides the leading term. Furthermore, if there exists i with $m_i \neq 0$ and $m_i \equiv 0 \bmod q$, then $d_{\mathbf{m},\mathbf{l}} = 0$. So it remains to consider the case where $\mathbf{m} \neq \mathbf{0}$ and either $m_i = 0$ or $m_i \not\equiv 0 \bmod q$ for all i. We write the exponent in the form

$$\mathbf{m} \cdot \mathbf{v_j} + \frac{r}{q-1} = \frac{A}{Q}$$

with $(A, Q) = 1$. In order to apply Lemma 4.3, we need a proper lower bound for Q. Note first that $\mathbf{m} \cdot \mathbf{v_j}$ can be written as mq^{-j-1}, where $j \geq j_1$ and $m \not\equiv 0 \bmod q$. Suppose that the prime decompositions of q and m are given by

$$q = p_1^{e_1} \cdots p_k^{e_k} \quad \text{and} \quad m = p_1^{f_1} \cdots p_k^{f_k} m',$$

where $p_1, \ldots, p_k$ are primes with $p_1 < p_2 < \cdots < p_k$, m' has no prime factors $p_1, \ldots, p_k$, and we have $e_i > 0$ and $f_i \geq 0$ for $i = 1, \ldots, k$. Since $m \not\equiv 0 \bmod q$, there is some i with $f_i < e_i$. Thus, if we write

$$\mathbf{m} \cdot \mathbf{v_j} = \frac{m}{q^{j+1}} = \frac{p_1^{f_1} \cdots p_k^{f_k} m'}{p_1^{e_1(j+1)} \cdots p_k^{e_k(j+1)} (m')^{j+1}} = \frac{A'}{Q'}$$

where $(A', Q') = 1$, then we certainly have $Q' \geq p_i^{je_i} \geq p_1^j$. Hence, with $c' = (\log p_1)/(\log q)$, we obtain $Q' \geq q^{c'j}$. Finally, since $A/Q = A'/Q' + r/(q-1)$ and $(Q', q-1) = 1$, it follows that $Q \geq Q'$ and, consequently,

$$Q \geq q^{c'j} \geq q^{c'j_1} \geq q^{c'L^\nu}.$$

We now apply Lemma 4.3 (with $K = c'L^\nu$) and obtain

$$S = \frac{\pi(x; k, q-1)}{q^d} + O\biggl(xL^2 e^{-c'L^\nu/2} \sum_{\mathbf{m} \neq \mathbf{0}} |d_{\mathbf{m},\mathbf{l},\Delta}|\biggr).$$

Since

$$\sum_{\mathbf{m} \neq \mathbf{0}} |d_{\mathbf{m},\mathbf{l},\Delta}| \leq (2 + 2\log(1/\Delta))^d,$$

it is possible to choose $\Delta = e^{-L^\nu}$, and so one finally gets

$$\begin{aligned}
&\frac{1}{\pi(x; k, q-1)} \#\{p \leq x : p \equiv k \bmod q-1,\ \epsilon_{j_1}(p) = \ell_1, \ldots, \epsilon_{j_d}(p) = \ell_d\} \\
&\quad = q^{-d} + O(d(e^{-L^\nu} + e^{-c_3 L^\nu})) + O(L^3 (4L^\nu)^d e^{-c'L^\nu/2}) \\
&\quad = q^{-d} + O((4L^\nu)^d e^{-c_4 L^\nu})
\end{aligned}$$

for some constant $c_4 > 0$. □

Next, we shall compare centralized moments of T_x and $\overline{T}_x$.

Lemma 4.6. *We have, uniformly for* $1 \leq d \leq L'$,

$$\mathbb{E}\biggl(\frac{T_x - L'\mu_q}{\sqrt{L'\sigma_q^2}}\biggr)^d = \mathbb{E}\biggl(\frac{\overline{T}_x - L'\mu_q}{\sqrt{L'\sigma_q^2}}\biggr)^d + O\biggl(\biggl(\frac{4q}{\sigma_q}\biggr)^d L^{(1/2+\nu)d} e^{-c_4 L^\nu}\biggr),$$

where $c_4 > 0$ *is the same constant as in Lemma 4.5.*

Proof. We expand the difference

$$\delta_d = \mathbb{E}\biggl(\sum_{L^\nu \leq j \leq L - L^\nu} (D_{j,x} - \mu_q)\biggr)^d - \mathbb{E}\biggl(\sum_{L^\nu \leq j \leq L - L^\nu} (Z_j - \mu_q)\biggr)^d$$

and compare terms with the help of (37). In fact, we have to take $(qL')^d$ terms into account, and thus we get

$$|\delta_d| \ll q^{2d} L^d (4L^\nu)^d e^{-c_4 L^\nu}.$$

Of course, this proves the lemma. □

4.3 Proof of Proposition 4.1

Finally, we are ready to complete the proof of Proposition 4.1. By Taylor's theorem, for every positive integer D and real u we have

$$e^{iu} = \sum_{0 \le d < D} \frac{(iu)^d}{d!} + O\left(\frac{|u|^D}{D!}\right).$$

Consequently, for any random variables X and Y,

$$\mathbb{E}e^{itX} - \mathbb{E}e^{itY} = \sum_{d<D} \frac{(it)^d}{d!}(\mathbb{E}\, X^d - \mathbb{E}\, Y^d) + O\left(\frac{|t|^D}{D!}\big|\mathbb{E}\,|X|^D - \mathbb{E}\,|Y|^D\big| + 2\frac{|t|^D}{D!}\mathbb{E}\,|Y|^D\right).$$

In particular, we will apply the above expansion with $X = (T_x - L'\mu_q)/(L'\sigma_q^2)^{1/2}$ and $Y = (\overline{T}_x - L'\mu_q)/(L'\sigma_q^2)^{1/2}$. Further, we set $D = \lfloor L^\kappa \rfloor$ for some real κ with $0 < \kappa < \nu$ (assuming without loss of generality that D is even) and suppose that $|t| \le L^\eta$ with $0 < \eta < \kappa/2$. Hence, by applying Lemma 4.6, we obtain

$$\begin{aligned}\sum_{1 \le d \le D} \frac{|t|^d}{d!}|\mathbb{E}\, X^d - \mathbb{E}\, Y^d| &\ll |t| \sum_{d \le D} \frac{L^{\eta(d-1)}}{d!}\left(\frac{4q}{\sigma_q}\right)^d L^{(1/2+\nu)d} e^{-c_4 L^\nu} \\ &\ll |t|\, e^{L^\kappa + L^\kappa \log(4q/\sigma_q) + (1/2+\nu+\eta)L^\kappa \log L - \kappa L^\kappa \log L - c_4 L^\nu} \\ &\ll |t| e^{-(c_4/2)\, L^\nu}\end{aligned}$$

for sufficiently large x.

The final step is to get some bound for the moments $\mathbb{E}\,|Y|^D$. Following the proof of Lemma 4.2, the moment generating function of Y is given by

$$\begin{aligned}\sum_{d \ge 0} \mathbb{E}\, Y^d \frac{w^d}{d!} &= \mathbb{E}\, e^{wY} \\ &= \varphi_3(-iw) \\ &= e^{w^2/2}\left(1 + O\left(\frac{w^4}{\log x}\right)\right)\end{aligned}$$

uniformly for $|w| \le (\log x)^{1/4}$. Hence, the moments are given by Cauchy's formula:

$$\mathbb{E}\, Y^d = \frac{d!}{2\pi i}\int_{|w|=w_0} e^{w^2/2}\left(1 + O\left(\frac{w^4}{\log x}\right)\right)\frac{dw}{w^{d+1}}.$$

Asymptotically, these kinds of integrals can be evaluated by means of a saddle-point method, where the saddle point w_0 (of the dominating part of the integrand $e^{w^2/2 - d\log w}$) is $w_0 = \sqrt{d}$. Of course, this works only if $d = o((\log x)^{1/2})$, in which case we obtain directly (for even d) that

$$\mathbb{E}\, Y^d = \frac{d!}{(d/2)!\, 2^{d/2}}\left(1 + O\left(\frac{d^2}{\log x}\right)\right).$$

Thus, for (even) $D = \lfloor L^\kappa \rfloor$ (where $\kappa < \nu < 1/2$) and $|t| \le L^\eta$ (where $\eta < \kappa/2$), we have

$$\begin{aligned}\frac{|t|^D}{D!}\mathbb{E}\,|Y|^D &\ll |t| \frac{L^{\eta(D-1)}}{D^{D/2}e^{-D/2}\sqrt{\pi D}} \\ &\ll |t| e^{\eta L^\kappa \log L - (\kappa L^\kappa \log L)/2 + L^\kappa/2} \\ &\ll |t| e^{-(\kappa/2-\eta)L^\kappa \log L}.\end{aligned}$$

This completes the proof of Proposition 4.1.

5. Proof of Theorems 1.1 and 1.2

5.1 Proof of Theorem 1.1

As a first step, we show that the integral (8) can be reduced to an integral on the interval $[-1/(2(q-1)), 1/(2(q-1))]$, to which we can then apply Propositions 2.1 and 2.2. For this purpose, we set

$$S(\alpha) = \sum_{p \le x} e(\alpha s_q(p)) \quad \text{and} \quad S_k(\alpha) = \sum_{p \le x,\, p \equiv k \bmod q-1} e(\alpha s_q(p)).$$

Since $s_q(n) \equiv n \bmod q-1$, we have

$$S\left(\alpha + \frac{\ell}{q-1}\right) = \sum_{p \le x} e(\alpha s_q(p)) \cdot e\left(\frac{\ell p'}{q-1}\right)$$

and, consequently,

$$\begin{aligned}S_k(\alpha) &= \sum_{p \le x} e(\alpha s_q(p)) \cdot \frac{1}{q-1}\sum_{\ell=0}^{q-2} e\left(\frac{\ell(p-k)}{q-1}\right) \\ &= \frac{1}{q-1}\sum_{\ell=0}^{q-2} e\left(-\frac{\ell k}{q-1}\right) S\left(\alpha + \frac{\ell}{q-1}\right).\end{aligned}$$

Thus, Proposition 2.1 also implies the upper bound

$$S_k(\alpha) \ll (\log x)^3\, x^{1-c_1\|(q-1)\alpha\|^2}. \tag{38}$$

Moreover, we have

$$\begin{aligned}\#\{p \le x : s_q(p) = k\} &= \int_{-1/(2(q-1))}^{1-1/(2(q-1))} S(\alpha) e(-\alpha k)\, d\alpha \\ &= \sum_{\ell=0}^{q-2} \int_{-1/(2(q-1))}^{1/(2(q-1))} S\left(\alpha + \frac{\ell}{q-1}\right) e\left(-\left(\alpha + \frac{\ell}{q-1}\right)k\right) d\alpha \\ &= \int_{-1/(2(q-1))}^{1/(2(q-1))} \sum_{p \le x} e(\alpha(s_q(p) - k)) \cdot \sum_{\ell=0}^{q-2} e\left(\ell\, \frac{p-k}{q-1}\right) d\alpha \\ &= (q-1)\int_{-1/(2(q-1))}^{1/(2(q-1))} \left(\sum_{p \le x,\, p \equiv k \bmod q-1} e(\alpha s_q(p))\right) e(-\alpha k)\, d\alpha \\ &= (q-1)\int_{-1/(2(q-1))}^{1/(2(q-1))} S_k(\alpha)\, e(-\alpha k)\, d\alpha.\end{aligned}$$

 https://doi.org/10.1112/S0010437X08003898

Next, we split the integral into two parts:

$$\int_{-1/(2(q-1))}^{1/(2(q-1))} = \int_{|\alpha|\le(\log x)^{\eta-1/2}} + \int_{(\log x)^{\eta-1/2}<|\alpha|\le 1/(2(q-1))}.$$

The first integral can easily be evaluated with the aid of Proposition 2.2. We use the substitution $\alpha = t/(2\pi\sigma_q\sqrt{\log_q x})$ and obtain

$$\begin{aligned}
&\int_{|\alpha|\le(\log x)^{\eta-1/2}} S_k(\alpha)e(-\alpha k)\,d\alpha \\
&\quad = \pi(x;k,q-1)\int_{|\alpha|\le(\log x)^{\eta-1/2}} e(\alpha(\mu_q\log_q x-k))\,e^{-2\pi^2\alpha^2\sigma_q^2\log_q x}(1+O(\alpha^4\log x))\,d\alpha \\
&\qquad + O\biggl(\pi(x)\int_{|\alpha|\le(\log x)^{\eta-1/2}}|\alpha|\,(\log x)^{\nu}\,d\alpha\biggr) \\
&\quad = \frac{\pi(x;k,q-1)}{2\pi\sigma_q\sqrt{\log_q x}}\int_{-\infty}^{\infty} e^{it\Delta_k-t^2/2}\,dt + O(\pi(x)e^{-2\pi^2\sigma_q^2(\log x)^{2\eta}}) \\
&\qquad + O\biggl(\frac{\pi(x)}{(\log x)^{3/2}}\biggr) + O\biggl(\frac{\pi(x)}{(\log x)^{1-\nu-2\eta}}\biggr) \\
&\quad = \frac{\pi(x;k,q-1)}{\sqrt{2\pi\sigma_q^2\log_q x}}(e^{-\Delta_k^2/2}+O((\log x)^{-1/2+\nu+2\eta})) \\
&\quad = \frac{1}{\varphi(q-1)}\frac{\pi(x)}{\sqrt{2\pi\sigma_q^2\log_q x}}(e^{-\Delta_k^2/2}+O((\log x)^{-1/2+\nu+2\eta}),
\end{aligned}$$

where

$$\Delta_k = \frac{k-\mu_q\log_q x}{\sqrt{\sigma_q^2\log_q x}}.$$

The remaining integral can be estimated directly by using Proposition 2.1 together with (38):

$$\begin{aligned}
\int_{(\log x)^{\eta-1/2}<|\alpha|\le 1/(2(q-1))} S_k(\alpha)\,e(-\alpha k)\,d\alpha &\ll (\log x)^3\,x\,e^{-c_1(q-1)^2(\log x)^{2\eta}} \\
&\ll \frac{\pi(x)}{\log x}.
\end{aligned}$$

Finally, if ε with $0<\varepsilon<1/2$ is given, then we can set $\nu = 2\varepsilon/3$ and $\eta=\varepsilon/6$. Hence $0<\eta<\nu/2$ and $\nu+2\eta=\varepsilon$, and therefore Theorem 1.1 follows immediately.

5.2 Proof of Theorem 1.2

Set $A_m(x) = \#\{p<x : s_q(p)=m\}$. Note that $\lfloor \mu_q\log_q p\rfloor = m$ if and only if $q^{m/\mu_q}\le p < q^{(m+1)/\mu_q}$. Hence,

$$\begin{aligned}
\#\{p<x : s_q(p) = \lfloor\mu_q\log_q p\rfloor\} = &\sum_{m<\lfloor\mu_q\log_q x\rfloor}(A_m(q^{(m+1)/\mu_q})-A_m(q^{m/\mu_q})) \\
&+ A_{\lfloor\mu_q\log_q x\rfloor}(x) - A_{\lfloor\mu_q\log_q x\rfloor}(q^{\lfloor\mu_q\log_q x\rfloor/\mu_q}).
\end{aligned}$$

Now, Theorem 1.1 implies that

$$A_m(q^{m/\mu_q}) = c\frac{q^{m/\mu_q}}{(m/\mu_q)^{3/2}}(1+O(m^{-1/2+\varepsilon})),$$

where

$$c = \frac{q-1}{\varphi(q-1)\log q\sqrt{2\pi\sigma_q^2}}.$$

Similarly, we have

$$A_m(q^{(m+1)/\mu_q}) = c\frac{q^{(m+1)/\mu_q}}{(m/\mu_q)^{3/2}}(1+O(m^{-1/2+\varepsilon})).$$

Set

$$C := \sum_{0\le j<q-1,\,(j,q-1)=1} q^{j/\mu_q}(q^{1/\mu_q}-1) \quad\text{and}\quad \ell_{\max} := \left\lfloor \frac{\mu_q \log_q x}{q-1}\right\rfloor.$$

Then we have

$$\sum_{m<\ell_{\max}(q-1)} (A_m(q^{(m+1)/\mu_q}) - A_m(q^{m/\mu_q})) = \sum_{\ell<\ell_{\max}} c\frac{q^{\ell(q-1)/\mu_q}}{(\ell(q-1)/\mu_q)^{3/2}}\, C\,(1+O(l^{-1/2+\varepsilon}))$$
$$= \frac{c}{(\log_q x)^{3/2}}\, C\, \frac{q^{\ell_{\max}(q-1)/\mu_q}}{q^{(q-1)/\mu_q}-1}(1+O((\log x)^{-1/2+\varepsilon})).$$

Furthermore,

$$\sum_{m=\ell_{\max}(q-1)}^{\lfloor \mu_q \log_q x\rfloor -1} (A_m(q^{(m+1)/\mu_q}) - A_m(q^{m/\mu_q}))$$
$$= \frac{cq^{\ell_{\max}(q-1)/\mu_q}}{(\log_q x)^{3/2}} \sum_{\substack{0\le j<\{(\mu_q\log_q x)/(q-1)\}(q-1)\\ (j,q-1)=1}} q^{j/\mu_q}(q^{1/\mu_q}-1)\,(1+O((\log x)^{-1/2+\varepsilon}))$$

and, finally,

$$A_{\lfloor \mu_q \log_q x\rfloor}(x) - A_{\lfloor \mu_q \log_q x\rfloor}(q^{\lfloor \mu_q \log_q x\rfloor/\mu_q})$$
$$= \frac{c}{(\log_q x)^{3/2}}(q^{\log_q x} - q^{\lfloor \mu_q \log_q x\rfloor/\mu_q})\,(1+O((\log x)^{-1/2+\varepsilon})).$$

Putting these three estimates together, we directly obtain (5) with

$$Q(t) = c\left(C\,\frac{q^{-\{t\}(q-1)/\mu_q}}{q^{(q-1)/\mu_q}-1} + q^{-\{t\}(q-1)/\mu_q} \sum_{\substack{0\le j<(q-1)\{t\}\\ (j,q-1)=1}} q^{j/\mu_q}(q^{1/\mu_q}-1) + 1 - q^{-\{(q-1)t\}/\mu_q}\right),$$

which ends the proof of Theorem 1.2.

Acknowledgement

The authors are grateful to the referee for checking the proofs carefully and suggesting several improvements.

References

BK95 N. L. Bassily and I. Kátai, *Distribution of the values of q-additive functions on polynomial sequences*, Acta Math. Hung. **68** (1995), 353–361.

BK96 N. L. Bassily and I. Kátai, *Distribution of consecutive digits in the q-ary expansion of some subsequences of integers*, J. Math. Sci. **78** (1996), 11–17.

CE46 A. H. Copeland and P. Erdős, *Note on normal numbers*, Bull. Amer. Math. Soc. **52** (1946), 857–860.

Coq86 J. Coquet, *Power sums of digital sums*, J. Number Theory **22** (1986), 161–176.

Del75 H. Delange, *Sur la fonction sommatoire de la fonction "Somme de Chiffres"*, Enseignement Math. **21** (1975), 31–77.

DR05 M. Drmota and J. Rivat, *The sum-of-digits function of squares*, J. London Math. Soc. (2) **72** (2005), 273–292.

FM05 E. Fouvry and C. Mauduit, *Sur les entiers dont la somme des chiffres est moyenne*, J. Number Theory **114** (2005), 135–152.

GKPT P. J. Grabner, P. Kirschenhofer, H. Prodinger and R. F. Tichy, *On the moments of the sum-of-digits function*, in *Applications of Fibonacci numbers, vol. 5 (St. Andrews, 1992).* (Kluwer Academic Publishers, Dordrecht, 1993), 263–271.

GK91 S. Graham and G. Kolesnik, *Van der Corput's method of exponential sums*, London Mathematical Society Lecture Note Series, vol. 126 (Cambridge University Press, Cambridge, 1991).

Hea82 D. R. Heath-Brown, *Prime numbers in short intervals and a generalized Vaughan identity*, Canad. J. Math. **34** (1982), 1365–1377.

Hoh30 G. Hoheisel, *Primzahlprobleme in der analysis*, Sitz. Preuss. Akad. Wiss. **33** (1930), 3–11.

IK04 H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, vol. 53 (American Mathematical Society, Providence, RI, 2004).

Kat67 I. Katai, *On the sum of digits of prime numbers*, Ann. Univ. Sci. Budapest. Eötvös Sect. Math. **10** (1967), 89–93.

Kat77 I. Katai, *On the sum of digits of primes*, Acta Math. Acad. Sci. Hungar. **30** (1977), 169–173.

Kat86 I. Katai, *Distribution of digits of primes in q-ary canonical form*, Acta Math. Acad. Sci. Hungar. **47** (1986), 341–359.

KM68 I. Katai and J. Mogyorodi, *On the distribution of digits*, Publ. Math. Debrecen **15** (1968), 57–68.

MR05 C. Mauduit and J. Rivat, *Sur un problème de Gelfond: la somme des chiffres des nombres premiers*, Ann. Math., to appear.

MR07 C. Mauduit and J. Rivat, *La somme des chiffres des carrés*, Acta Math., to appear.

MS97 C. Mauduit and A. Sárközy, *On the arithmetic structure of the integers whose sum of digits is fixed*, Acta Arith. **81** (1997), 145–173.

Shi74 I. Shiokawa, *On the sum of digits of prime numbers*, Proc. Japan Acad. **50** (1974), 551–554.

Sto77 K. B. Stolarsky, *Power and exponential sums of digital sums related to binomial coefficient parity*, SIAM J. Appl. Math. **32** (1977), 717–730.

Vau80 R. C. Vaughan, *An elementary method in prime number theory*, Acta Arith. **37** (1980), 111–115.

Vin54 I. M. Vinogradov, *The method of trigonometrical sums in the theory of numbers* (Interscience, London, 1954), translated from the Russian, revised and annotated by K .F. Roth and A. Davenport.

Michael Drmota michael.drmota@tuwien.ac.at
Institute of Discrete Mathematics and Geometry, Technische Universität Wien,
Wiedner Hauptstraße 8-10/104, A-1040 Wien, Austria

Christian Mauduit mauduit@iml.univ-mrs.fr
Institut de Mathématiques de Luminy, CNRS UMR 6206, Université de la Méditerranée,
Campus de Luminy, Case 907, 13288 Marseille Cedex 9, France

Joël Rivat rivat@iml.univ-mrs.fr
Institut de Mathématiques de Luminy, CNRS UMR 6206, Université de la Méditerranée,
Campus de Luminy, Case 907, 13288 Marseille Cedex 9, France