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 $\#\{p \leqslant x : p \text{ prime}, s_q(p) = k\}$ 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 $\sum_{p\leqslant x} e(\alpha s_q(p))$ (the sum is restricted to p prime), where 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 $q \geqslant 2$ 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 $n = \sum_{j\geqslant 0} \varepsilon_j(n) q^j$ with digits $\varepsilon_j(n) \in \{0, 1, \dots, q-1\}$ then

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

The statistical behaviour of the sum of digits function and, more generally, for q-additive function has been very well studied by several authors. It is, for example, well known (see, for example 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 knows for *higher moments* ([GKPT], see also [Sto77] and [Coq86] for the case $q = 2$). Furthermore, the distribution of the sum-of-digits function can be approximated by a normal distribution

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

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 an uniform estimate of $\#\{n < q^\nu : s_q(n) = k\}$ is provided for any $k \leqslant \mu_q \nu$ and in [FM05] where it is proved that for any fixed $k \geqslant 1$ we have

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

2000 Mathematics Subject Classification Primary: 11A63, 11L03, 11N05, Secondary: 11N60, 11L20, 60F05
Keywords: sum-of-digits function, primes, exponential sums, central limit theorem
The first author is supported by the Austrian Science Foundation FWF, grant S9604, that is part of the National Research Network "Analytic Combinatorics and Probabilistic Number Theory".

uniformly for any $x \geqslant 2$ and any $b : \mathbb{N} \to \mathbb{R}$ such that $|b(\nu)| \leqslant K v^{1/4}$ and $\nu_q \nu + b(\nu) \in \mathbb{N}$ for any $n \geqslant 1$.

The first result concerning the asymptotic behaviour of the sum of digits function restricted to prime numbers is a consequence of the famous theorem by Copeland and Erdős in [CE46] concerning 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 their 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 show in [Shi74] by Shiokawa 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 principal the same if the average is taken over primes $p \leqslant x$. For example, Katai [Kat77] has shown that

$$\sum_{p \leqslant x} |s_q(p) - \mu_q \log_q x|^k \ll x(\log x)^{k/2-1}, k = 1, 2, \cdots ,$$

and [Kat86] that there is a central limit theorem similarly to the above (see also [KM68] for a related result):

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

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

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

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

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

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

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

Such a local version of (2) or (3) was considered by Erdős as "hopelessly difficult" and the first breackthrough 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 exist $\sigma_{q,m} > 0$ such that for every $a \in \mathbb{Z}$ we have

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

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

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

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

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

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

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

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 also 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

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

Such bounds could be used to obtain estimates for *tail distributions*, that is bounds on the numbers

$$\#\{p \leqslant x : s_q(p) \leqslant c_1 \log_q(x)\} \quad \text{resp.} \quad \#\{p \leqslant x : s_q(p) \geqslant c_2 \log_q(x)\}$$

for $0 < c_1 < \mu_q$ and $\mu_q < c_2 < 2\mu_q$. By curiousity 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 of the Proof

The proof of Theorem 1.1 uses two main ingrediences (Propositions 2.1 and 2.2) that we prove in Sections 3 and 4.

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

Proposition 2.1. *For every fixed integer $q \geqslant 2$ there exists a constants $c_1 > 0$ such that*

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

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 adapt the moment method due to Bassily and Kátai [BK95] (see also [KM68] and [Kat77]). The difference to [BK95] is that we provide bounds for the d-th moments (of a certain random variable) that are uniform for all $d \geqslant 1$. Note that the generalization of [BK95] that is provided in [BK96] is not sufficient for our purposes. Therefore we have to adapt all main steps. As usual, $\pi(x; k, q-1)$ denotes the number of primes $p \leqslant x$ with $p \equiv k \bmod q-1$.

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

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

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

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

$$\#\{p \leqslant x : s_q(p) = k\} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\sum_{p \leqslant x} e(\alpha s_q(p)) \right) e(-\alpha k)\, d\alpha \tag{8}$$

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 an integer $\geqslant 1$, 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 $\sum_{p \leqslant x} \mathrm{e}(\alpha s_q(p))$ uniform in terms of α and x.

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

$$\left| \sum_{n \leqslant x} \Lambda(n) e(\alpha s_q(n)) \right| \ll (\log x)^4 x^{1 - c_1 \|(q-1)\alpha\|^2} \tag{9}$$

uniformly for real α.

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

3.1 A combinatorial identity

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

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

where $n_1, \ldots, n_k$ satisfy multiplicative conditions. Vaughan has given an elegant formulation of this method [Vau80], later generalized by Heath-Brown [Hea82].

A drawback of these methods in their original setting is the outcome 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 permits to suppress this difficulty:

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

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

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

Then

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

Proof. This is Lemma 1 of [MR05]. □

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

Lemma 3.2. *Let g be an arithmetic function, $q \geqslant 2$, $0 < \delta < \beta_1 < 1/3$, $1/2 < \beta_2 < 1$. Suppose that uniformly for all complex numbers b_n such that $|b_n| \leqslant 1$, we have*

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

whenever

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

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

$$x^{\beta_1} \leqslant M \leqslant x^{\beta_2} \tag{14}$$

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

Proof. This is Lemma 3 of [MR05]. □

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 $q \geqslant 2$, $x \geqslant 2$, and for every α such that $(q-1)\alpha \in \mathbb{R}\setminus\mathbb{Z}$ we have*

$$\sum_{\frac{M}{q}<m\leqslant M} \max_{\frac{x}{qm}\leqslant t\leqslant \frac{x}{m}} \left| \sum_{t<n\leqslant \frac{x}{m}} \mathrm{e}(\alpha\, s_q(mn)) \right| \ll_q x^{1-\kappa_q(\alpha)} \log x \tag{15}$$

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

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

where $\frac{1}{2} \leqslant \gamma_q(\alpha) < 1$ 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 Proposition 2 of [MR05]. □

3.3 Type II sums

In order to verify (11) we use Lemma 3.2, that is, we will prove the following proposition (which a variant of [MR05, Propositon 1]):

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

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

whenever

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

We note that the constants β_1, β_2, δ, and $\xi_q(\alpha)$ can be stated explicitly in terms of α, compare with (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 Cauchy-Schwarz's inequality and a Van der Corput type inequality in order to *smooth the sums.*

For $q \geqslant 2$ and real α let

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

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

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

By Cauchy-Schwarz's inequality,

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

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

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

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

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

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

$$\left| \sum_{q^{\nu-1}<n\leqslant q^{\nu}} b_n \, \mathrm{e}(f(mn)) \right|^2 \leqslant q^{\nu-\rho} \sum_{|r|<q^{\rho}} \left(1 - \frac{|r|}{q^{\rho}}\right) \Big(\sum_{q^{\nu-1}<n\leqslant q^{\nu}} b_{n+r}\,\overline{b_n}\, \mathrm{e}(f(m(n+r)) - f(mn)) + O(q^{\rho}) \Big),$$

where the term $O(q^{\rho})$ comes from the removal of the condition of summation $q^{\nu-1} < n + r \leqslant q^{\nu}$ which was introduced by Lemma 3.3. Indeed this removal may potentially imply $O(q^{\rho})$ values of n, and each term in the sum is of modulus less or equal to 1, which lead to an error at most $O(q^\rho)$. We separate the cases $r = 0$ and $r \neq 0$, and obtain:

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

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 $\rho \leqslant \nu/2$.

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

For any integer $\lambda \geqslant 0$, 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 integers $\varepsilon_k(n)$ denote the digits of n in basis q. The function f_λ is clearly periodic of period q^λ. This truncated function appears in a different context in [DR05] where Drmota and Rivat study some properties of $f_\lambda(n^2)$ where λ is of order $\log n$. The following lemma is a variant of [MR05, Lemme 5].

Lemma 3.4. *For all integers μ, ν, ρ with $\mu > 0$, $\nu > 0$, $0 \leqslant \rho \leqslant \nu/2$ and for 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 $q^{\mu-1} < m \leqslant q^\mu$, $q^{\nu-1} < n \leqslant q^\nu$ and*

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

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 $0 \leqslant r < q^\rho$. In this case $0 \leqslant mr < q^{\mu+\rho}$. When we compute the sum $mn + mr$, the digits of the product mn of index $\geqslant \mu+\rho$ 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 $\mu+\rho \leqslant j < \mu+2\rho$. Therefore grouping the products mn according to their value a, we obtain

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

where $\tau(a)$ denotes the number of divisors of a and $\chi(a) = 1$ if the digits a_j in basis q of a satisfy $a_j = q-1$ for $\mu+\rho \leqslant j < \mu+2\rho$, and $\chi(a) = 0$ in the opposite case, that is if there exist an index j, with $\mu+\rho \leqslant j < \mu+2\rho$, for which $a_j \neq q-1$. We deduce that

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

For each c fixed we apply Lemma 3.5 below with

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

(by (20) we have $x^{27/82} \leqslant q^{\frac{27}{82}(\mu+\nu)} \leqslant y \leqslant x$), so that we 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 $\mu+\rho \leqslant j < \mu+2\rho$, and we obtain the same upper bound (21). □

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

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

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

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

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

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

□

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

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

where

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

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

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

where $\varphi_q(t)$ is defined in Proposition 3.1,

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

where $\gamma_q(t)$ is defined in Proposition 3.1,

$$\xi_q(\alpha) := \frac{\epsilon_q(\alpha)}{14}, \quad \delta := \frac{\epsilon_q(\alpha)}{28}, \tag{24}$$

$$\beta_1 := \frac{(3-2\epsilon_q(\alpha))\xi_q(\alpha)}{\epsilon_q(\alpha)} + \delta \quad \text{for } q = 2, \tag{25}$$

$$\beta_1 := \frac{(4-2\epsilon_q(\alpha))\xi_q(\alpha)}{\epsilon_q(\alpha)} + \delta \quad \text{for } q \geqslant 3, \tag{26}$$

$$\beta_2 := \frac{1-(5-2\epsilon_q(\alpha))\xi_q(\alpha)}{2-\epsilon_q(\alpha)} - \delta \quad \text{for } q=2, \tag{27}$$

$$\beta_2 := \frac{1-(6-2\epsilon_q(\alpha))\xi_q(\alpha)}{2-\epsilon_q(\alpha)} - \delta \quad \text{for } q\geqslant 3. \tag{28}$$

It is shown in paragraph 7.3 of [MR05] that $0<\delta<\beta_1<1/3$, $1/2<\beta_2<1$ and that for any integers $\mu>0$ and $\nu>0$ verifying

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

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

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

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

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

and

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

Furthermore by Lemma 7 of [MR07] we have

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

so that

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

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 $\rho\leqslant 2c_1\left\|(q-1)\alpha\right\|^2(\mu+\nu)$ so that

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

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 $g(n)=\mathrm{e}(\alpha s_q(n))$ 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

$$x^{\beta_1}\leqslant M\leqslant x^{\beta_2}$$

the estimate

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

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

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

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) is true for any $x \geqslant 2$ 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)) and (32) shows that (11) is true for any $x > x_0$ with some U such that

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

It follows from Lemma 3.1 that for $x > x_0$

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

By (33), the condition $x > x_0$ is equivalent to $\|(q-1)\alpha\| \geqslant c_2(\log x)^{-1/2}$ with $c_2 = \sqrt{\frac{3\log q}{c_1}}$, so that we have proved (9) which ends 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) $x \geqslant 2$ we consider the set of primes

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

Its cardinality is denoted by $\pi(x; k, q-1)$ and it is well known that we have asymptotically

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

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

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

Of coures, $D_j = D_{j,x} = \varepsilon_j$, the j-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

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

that is uniform for $|t| \leqslant (\log x)^\eta$. We just have to set $\alpha = t/(2\pi\sigma_q(\log_q x)^{1/2})$.

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

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

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*

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

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 that $|e^{it} - e^{is}| \leqslant |t - s|$. Consequently

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

This proves the lemma. □

Now we approximate T_x by a sum $\overline{T}_x$ of independent random variables. Let Z_j $(j \geqslant 0)$ be a sequences of independent random variables with range $\{0, 1, \ldots, q-1\}$ and uniform probability distribution

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

We then set

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

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

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

that is also uniform for $|t| \leqslant (\log x)^{\frac{1}{4}}$.

Proof. First note that

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

Now (35) follows by setting

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

and by 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 (despite 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 $\varphi_3(t)$. In what follows we will prove the following bound.

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

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

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

Note that $e^{-c_1 L^\kappa} \ll L^{-1}$. Hence, Proposition 4.1 (together with Lemma 4.1 and Lemma 4.2) immediately imply (34) and, thus, Proposition 2.2.

4.2 Comparision of moments

In what follows we will use the following well known bound on exponential sums over primes.

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

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

where the implied constant is absolute.

Proof. We just have 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 $L^\nu \leqslant j \leqslant L - L^\nu$ and $0 < \Delta < 1/(2q)$ we uniformly have, as $x \to \infty$,

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

where c_3 is a certain positive constant.

Proof. We just have to show that the discrepancy D of the sequence (pq^{-j-1}) where p ranges over all primes $p \leqslant x$ with $p \equiv k \bmod q-1$ is bounded above $D \ll e^{-c_3 L^\nu}$. Of course, (36) follows then immediately.

We use the Erdős-Turán inequality saying that

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

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

First of all recall that

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

Thus, we actually have to estimate exponential sums of the form

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

We represent the rational number in the exponent by

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

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

$$\begin{aligned} D &\ll \frac{1}{H}+\frac{L}{x}\sum_{h=1}^{H}\frac{1}{h}L^2\,x\,q^{-\frac{1}{3}L^\nu} \\ &\ll \frac{1}{H}+L^4 q^{-\frac{1}{3}L^\nu} \\ &\ll e^{-c_3L^\nu}, \end{aligned}$$

where $c_3<\frac{1}{3}\log q$. This completes the proof of the lemma. □

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

Lemma 4.5. *Let $1\leqslant d\leqslant L'$ and $j_1,j_2,\dots,j_d$ and $\ell_1,\ell_2,\dots,\ell_d$ integers with*

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

and

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

Then we have uniformly

$$\begin{aligned} &\frac{1}{\pi(x;k,q-1)}\#\{p\leqslant x : p\equiv k \bmod q-1,\ \epsilon_{j_1}(p)=\ell_1,\dots,\epsilon_{j_d}(p)=\ell_d\} \\ &\qquad = q^{-d}+O\left((4L^\nu)^d e^{-c_4L^\nu}\right), \end{aligned}$$

where c_4 is a certain positive constant.

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

$$\begin{aligned} &\mathbf{Pr}\{D_{j_1,x}=\ell_1,\dots,D_{j_d,x}=\ell_d\} \\ &\qquad = \mathbf{Pr}\{Z_{j_1}=\ell_1,\dots,Z_{j_d}=\ell_d\}+O\left((4L^\nu)^d e^{-c_4L^\nu}\right) \end{aligned} \tag{37}$$

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

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

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

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 $m \neq 0$ by

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

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

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

By definition we have $0 \leqslant f_{\ell,\Delta}(x) \leqslant 1$ 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 $\Delta < 1/(2q)$

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

The third line follows from Lemma 4.4.

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

$$\mathbf{v}_\mathbf{j} = \left(q^{-j_1-1},\ldots,q^{-j_d-1}\right)$$

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

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

Thus, we are led to consider the exponential sum

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

If $\mathbf{m} = (0, \dots, 0)$ 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 we have $m_i = 0$ 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 $j \geqslant j_1$ 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 $p_1,\dots,p_k$ are primes with $p_1 < p_2 < \cdots < p_k$, m' has no prime factors $p_1,\dots,p_k$, and we have $e_i > 0$ and $f_i \geqslant 0$ for $i = 1,\dots,k$. Since $m \not\equiv 0 \bmod q$ there is some i with $f_i < e_i$. Thus, if we write

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

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

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

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

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

Since

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

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

$$\begin{aligned}
&\frac{1}{\pi(x;k,q-1)}\#\{p \leqslant x : p \equiv k \bmod q-1,\ \epsilon_{j_1}(p) = \ell_1,\dots,\epsilon_{j_d}(p) = \ell_d\} \\
&\quad = q^{-d} + O\left(d(e^{-L^\nu} + e^{-c_3L^\nu}\right) + O\left(L^2(4L^\nu)^d e^{-\frac{1}{2}c'L^\nu}\right) \\
&\quad = O\left((4L^\nu)^d e^{-c_4L^\nu}\right)
\end{aligned}$$

for some constant $c_4 > 0$. □

Now we compare centralized moments of T_x and $\overline{T}_x$.

LEMMA 4.6. *We have uniformly for* $1 \leqslant d \leqslant L'$

$$\mathbb{E}\left(\frac{T_x - L'\mu_q}{\sqrt{L'\sigma_q^2}}\right)^d = \mathbb{E}\left(\frac{\overline{T}_x - L'\mu_q}{\sqrt{L'\sigma_q^2}}\right)^d + O\left(\left(\frac{4q}{\sigma_q}\right)^d L^{(\frac{1}{2}+\nu)d} e^{-c_4L^\nu}\right),$$

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

Proof. We expand the following difference

$$\delta_d = \mathbb{E}\left(\sum_{L^\nu \leqslant j \leqslant L-L^\nu} (D_{j,x}-\mu_q)\right)^d - \mathbb{E}\left(\sum_{L^\nu \leqslant j \leqslant L-L^\nu} (Z_j-\mu_q)\right)^d$$

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

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

Of course, this proves the lemma. ∎

4.3 Proof of Proposition 4.1

Finally, we can complete the proof of Proposition 4.1 By Taylor's theorem we have for every integer $D>0$ and real u

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

Consequently we have for any random variables X and Y

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

In particular we will apply that for $X = (T_x - L'\mu_q)/(L'\sigma_q^2)^{1/2}$ 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 $0<\kappa<\nu$ (and assume without loss of generality that D is even) and suppose that $|t| \leqslant L^\eta$ with $0<\eta<\frac{1}{2}\kappa$. Hence, by applying Lemma 4.6 we get

$$\begin{aligned}\sum_{1\leqslant d\leqslant D} \frac{|t|^d}{d!}\left|\mathbb{E}\,|X|^d - \mathbb{E}\,|Y|^d\right| &\ll |t| \sum_{d\leqslant D} \frac{L^{\eta(d-1)}}{d!}\left(\frac{4q}{\sigma_q}\right)^d L^{(\frac12+\nu)d} e^{-c_4 L^\nu}\\ &\ll |t|\, e^{L^\kappa + L^\kappa \log(4q/\sigma_q) + (\frac12+\nu+\eta)L^\kappa \log L - \kappa L^\kappa \log L - c_4 L^\nu}\\ &\ll |t| e^{-(c_4/2)\, L^\nu}\end{aligned}$$

for sufficiently large x.

Finally we have to get some bound for the moments $\mathbb{E}\,|Y|^D$. Following the proof of Lemma 4.2 it follows that the moment generating function of Y is given by

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

uniformly for $|w| \leqslant (\log x)^{\frac14}$. 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 with help 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 given by $w_0 = \sqrt{d}$. Of course this only works if $d = o\left((\log x)^{\frac{1}{2}}\right)$, where we directly get (for even d)

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

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

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

This completes the proof of Proposition 4.1.

5. Proof of Theorems 1.1 and 1.2

5.1 Proof of Theorem 1.1

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

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

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

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

and consequently

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

Thus, Proposition 2.1 also implies the upper bound

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

Further, we have

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

Next we split the integral into two parts:

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

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

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

where

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

The remaining integral can be directly estimated with Proposition 2.1 (resp. with (38)):

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

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

5.2 Proof of Theorem 1.2

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

$$\#\{p<x : s_q(p)=\lfloor \mu_q\log_q p\rfloor\} = \sum_{m<\lfloor \mu_q \log_q x\rfloor} \left(A_m(q^{(m+1)/\mu_q})-A_m(q^{m/\mu_q})\right) + A_{\lfloor \mu_q\log_q x\rfloor}(x) - A_{\lfloor \mu_q\log_q x\rfloor}(q^{\lfloor \mu_q \log_q x\rfloor/\mu_q})$$

Now Theorem 1.1 implies that

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

where

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

Similarly we have

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

Set

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

Then we have

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

Further,

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

and finally

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

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

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

which ends the proof of Theorem 1.2.

Acknowledgement. The authors are grateful to the referee for checking carefully the proofs 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 consecutative digits in the q-ary expansion of some subsequences of integers*, J. Math. Sci. **78** (1996), 11–17.

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

CE46 A. H. COPELAND AND P. ERDŐS, *Note on normal numbers*, Bull. Amer. Math. Soc. **52** (1946), 857–860.

Del75 H. DELANGE Sur la fonction sommatoire de la fonction "Somme de Chiffres", *L'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**,2 (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, *Applications of Fibonacci numbers*, Vol. 5 (St. Andrews, 1992), 263–271, Kluwer Acad. Publ., Dordrecht, 1993.

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, 1991.

Hea82 D. R. HEATH-BROWN, Prime numbers in short intervals and a generalized Vaughan identity, *Can.J.Math.* **34**,6 (déc. 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., *Publicationes Mathematicae 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 Mathematica (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 Arithmetica* **37**,1 (1980), 111–115.

Vin54 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 A. Davenport*, Interscience, London, 1954.

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

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