The level of distribution of the sum-of-digits function in arithmetic progressions

Nathan Toumi
Université de Lorraine, IECL

Introduction

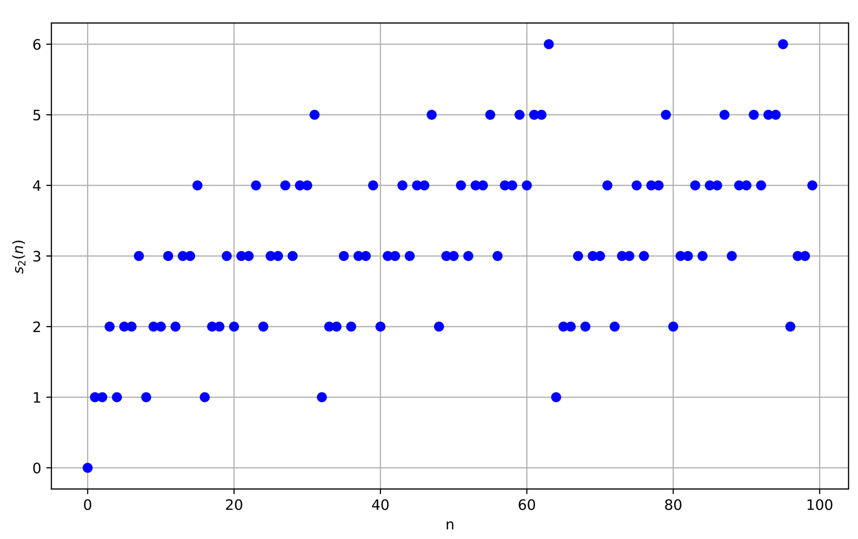
Let $q \geq 2$. Let $b \geq 2$ such that $(b, q-1) = 1$. Every integer n can be uniquely written as

$$n = \sum_{k=0}^{+\infty} n_k q^k,$$

where $n_k \in \{0, \dots, q-1\}$ for all $k \geq 0$ are the digits which are zero starting from some finite rank. The sum-of-digits function s_q is defined for all $n \in \mathbb{N}$ by

$$s_q(n) = s_q\left(\sum_{k=0}^{+\infty} n_k q^k\right) := \sum_{k=0}^{+\infty} n_k.$$

The following graph gives the behaviour of $s_2(n)$ for $0 \leq n < 100$.



A natural question that arises is how the sum-of-digits function behaves in arithmetic progressions.

Gelfond's works

Let $q \geq 2$, $b \geq 2$. Gelfond **[3]** proved in 1967/68 that the sum-of-digits function s_q is well-distributed in arithmetic progressions. More precisely, with

$$A_q(y, z; m, r) = \left|\{y \leq n < z \,:\, s_q(n) \equiv a \pmod{b},\ \ n \equiv r \pmod{m}\}\right|,$$

he showed that if $(b, q-1) = 1$ then, for all $a \in \{0, \dots, b-1\}$ and all $r \in \{0, \dots, m-1\}$, we have

$$\left|A_q(0, N; m, r) - \frac{N}{bm}\right| = O(N^{\lambda}), \qquad N \to +\infty,$$

where

$$\lambda = \frac{1}{2\log(q)} \log \frac{q \sin(\pi/2b)}{\sin(\pi/2bq)} < 1.$$

This result is non-trivial only if $m < N^{1-\lambda}$.

Estimates on average

For larger values of m, in many applications, a result on average on m is sufficient. Estimates of the type

$$\sum_{1 \leq m \leq x^{\theta}} \max_{\substack{y,z \geq 0 \\ z-y \leq x}} \max_{0 \leq r < m} \left|A_q(y, z; m, r) - \frac{z-y}{bm}\right| \ll x/(\log x)^B, \qquad x \to +\infty,$$

for $0 < \theta = \theta_q < 1$ as large as possible, are of great importance. In this direction, Fouvry–Mauduit **[1,2]** proved that $\theta_2 = 0.55711$ is admissible and it is possible to have some exponents θ_q with $\theta_q \to 1$ for $q \to \infty$.

In 2020, Spiegelhofer **[6]** improved this result in the case of the Thue–Morse sequence $\mathbf{t} = ((-1)^{s_2(n)})_{n \in \mathbb{N}}$ which corresponds to the case $q = b = 2$. He proved that the level of distribution of the Thue–Morse sequence equals 1, meaning that $\theta_2 = 1 - \varepsilon$ with $\varepsilon > 0$ arbitrarily small is an admissible exponent for $\mathbf{t}$. The error term is of the form $x^{1-\eta}$ with $0 < \eta = \eta(\varepsilon) < 1$ in place of $x/(\log x)^B$.

Generalization in base q

In 2014, Martin, Mauduit, and Rivat **[5]** determined an estimate for sums of type II.

Theorem (Martin/Mauduit/Rivat, 2014). *Let $\alpha \in \mathbb{R}/\mathbb{Z}$. Let $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ be sequences of complex numbers such that for all $n \geq 1$, we have $|a_n| \leq 1$ and $|b_n| \leq 1$. Let $x \geq 2$, $0 < \varepsilon \leq 1/2$, $x^{\varepsilon} \leq M, N \leq x$ and $MN \leq x$. We set*

$$\Theta_q := \left(1 - \frac{1}{q}\right)\left(1 - \sqrt{1 - \frac{2q-1}{3q(q-1)}}\right),$$

$$\eta_q := \max\left(\frac{1}{2} - \frac{\log(4 - 2\sqrt{2})}{2\log 2}, \frac{1}{2} + \frac{\log\left(1 - \Theta_q\right))}{4\log 2}\right)$$

and let $\gamma_q \in \mathbb{R}$ be such that

$$q^{\gamma_q} := 2 \max_{t \in \mathbb{R}} \sqrt{\left|\frac{\sin\left(q(\alpha - qt)\pi\right) \sin\left(q(\alpha - t)\pi\right)}{\sin\left((\alpha - qt)\pi\right) \sin\left((\alpha - t)\pi\right)}\right|}.$$

Finally, set

$$\xi_{q,\varepsilon} := \min\left(\frac{\varepsilon}{6}, \frac{1}{20}\right) \min\left(\frac{1}{2} - \eta_q, 2(1 - \gamma_q)\right). \tag{1}$$

Then

$$\sum_{M < m \leq 2M} \sum_{N < n \leq 2N} a_m b_n \mathrm{e}(\alpha s_q(mn)) \ll x^{1 - \xi_{q,\varepsilon}} \log(x),$$

where for all $x \in \mathbb{R}$,

$$\mathrm{e}(x) = \exp(2i\pi x).$$

Spiegelhofer showed that Theorem allows to get the estimate:

$$\sum_{1 \leq m \leq D} \left| \sum_{\substack{0 \leq n \leq x \\ n \equiv 0 \pmod{m}}} (-1)^{s_2(n)} \right| \leq C x^{1 - \xi_{2,\varepsilon}},$$

for $0 < \varepsilon < 1/2$ and for $D = x^{\varepsilon}$. This is a weaker version of the main result of Spiegelhofer **[6]** but the constant $\xi_{2,\varepsilon}$ is explicit.

The main result

We succeeded in **[7]** to generalize the result of Spiegelhofer **[6]** to base-q. Moreover, we propose an explicit value for the exponent η.

Theorem 1 (T., 2024+)

Let $0 < \varepsilon < 1/2$. There exist a constant $C = C(\varepsilon) > 0$ and an exponent $\eta = \eta(\varepsilon) > 0$ such that

$$\sum_{1 \leq m \leq x^{1-\varepsilon}} \max_{\substack{y,z \\ 0 \leq y < z \\ z - y \leq x}} \max_{r \geq 0} \left|A_q(y, z; m, r) - \frac{z-y}{bm}\right| \leq C x^{1-\eta}.$$

An admissible value for η is given by

$$\eta(\varepsilon) = \frac{\varepsilon^3 \min\left(1/4, 3\log_q\left(P^-(q)\right)\right)}{7200 \times 8^{1/\varepsilon}\left(\log(4q/\varepsilon) + 5b\log(q)/\varepsilon\right)} \times \exp\left(-\frac{5}{\varepsilon}\left(\log(4q/\varepsilon) + \frac{5b\log(q)}{\varepsilon}\right)\right). \tag{2}$$

A straightforward Python program allows to plot the graph of the logarithm of the quantity defined in (2), seen as a function of ε.

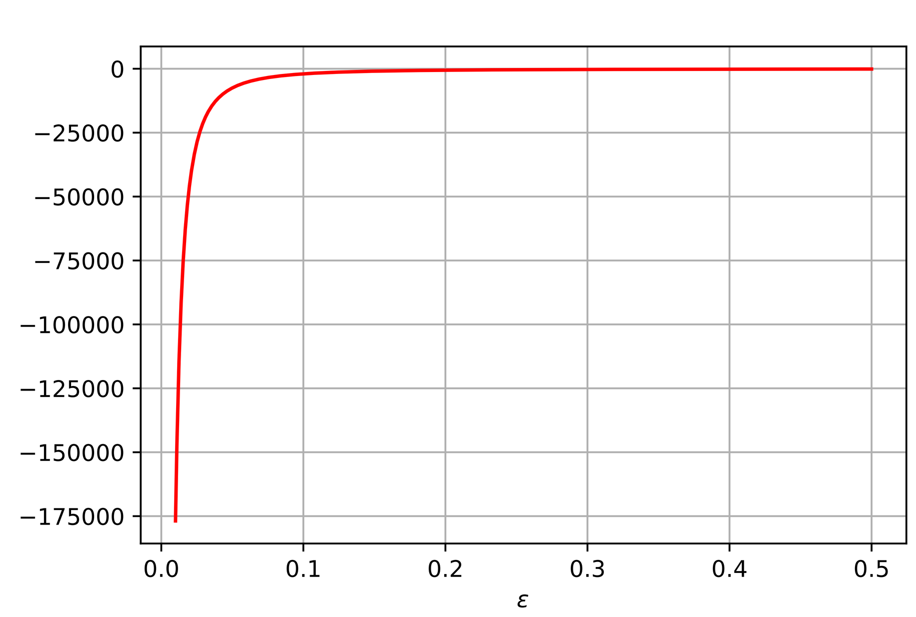


FIGURE 1- The graph of the function $\varepsilon \mapsto \log \eta(\varepsilon)$ in $]0.01, 0.5[$ for $q = b = 2$.

The main problem in the generalization

The idea of Spiegelhofer **[6]** is to apply the van der Corput inequality (and variations thereof) for an exponential sum a certain number of times in order to remove digits to reduce the estimate of the exponential sums to estimates of Gowers norms for the Thue–Morse sequence. In order to follow the path of Spiegelhofer, we proved an estimate for the Gowers norm for $(t_q(n))_{n \geq 1}$ defined by

$$t_q(n) = \mathrm{e}\left(\frac{\ell}{b} s_q(n)\right).$$

The proof of the following theorem is inspired by arguments given by Koniezcny **[4]**.

Theorem 2 (T., 2024) *Let $k \geq 3$ and set*

$$K = \left\lfloor \frac{\log(k)}{\log(q)} \right\rfloor + 1.$$

We define

$$\eta_0 := \frac{1}{\log(q)(K + (k+1)b)q^{(k+1)(K + b(k+1))}}.$$

Then, as $\rho \to +\infty$,

$$\frac{1}{q^{(k+1)\rho}} \sum_{\substack{0 \leq n < q^{\rho} \\ 0 \leq h_0, \dots, h_{k-1} < q^{\rho}}} \mathrm{e}\left(\frac{\ell}{b} \sum_{\mathbf{w} = (w_0, \dots, w_{k-1}) \in \{0,1\}^k} (-1)^{s_2(w)} s_q(n + \mathbf{w} \cdot \mathbf{h})\right) \ll q^{-\eta_0 \rho}.$$

References

[1] E. Fouvry and C. Mauduit, Méthodes de crible et fonctions sommes des chiffres. *Acta Arith.* **77** (1996), no.4, 339–351.
[2] E. Fouvry and C. Mauduit, Sums of digits and almost primes, *Math. Ann.* **305** (1996), no. 3, 571–599.
[3] A. Gelfond, Sur les nombres qui ont des propriétés additives et multiplicatives données, *Acta Arith.* **13** (1968), no.3, 259–265.
[4] J. Konieczny, Gowers norms for the Thue–Morse and Rudin–Shapiro sequences. *Ann. Inst. Fourier (Grenoble)*, **69** (2019), no.4, 1897–1913.
[5] B. Martin, C. Mauduit, and J. Rivat, Théorème des nombres premiers pour les fonctions digitales. *Acta Arith.* **165** (2014), no.1, 11–45.
[6] L. Spiegelhofer, The level of distribution of the Thue–Morse sequence, *Compos. Math.* **156** (2020), no.12, 2560–2587.
[7] N. Toumi, The level of distribution of the sum-of-digits function in arithmetic progressions, in preparation (2024).