Iurie BOREICO, Daniel EL-BAZ et Thomas STOLL

On a conjecture of Dekking : The sum of digits of even numbers

<http://jtnb.cedram.org/item?id=JTNB_2014__26_1_17_0>
On a conjecture of Dekking : The sum of digits of even numbers

par Iurie BOREICO, Daniel EL-BAZ et Thomas STOLL

Résumé. A propos d’une conjecture de Dekking : la somme des chiffres des nombres pairs
Soient \( q \geq 2 \) et \( s_q \) la fonction somme des chiffres en base \( q \). Pour \( j = 0, 1, \ldots, q - 1 \) on considère
\[
\# \{ 0 \leq n < N : s_q(2n) \equiv j \pmod{q} \}.
\]
En 1983, F. M. Dekking a conjecturé que cette quantité est strictement supérieure à \( N/q \) et, respectivement, strictement inférieure à \( N/q \) pour une infinité de \( N \), affirmant ce faisant l’absence d’un phénomène de dérive (ou phénomène de Newman). Dans cet article, nous démontrons sa conjecture.

Abstract. Let \( q \geq 2 \) and denote by \( s_q \) the sum-of-digits function in base \( q \). For \( j = 0, 1, \ldots, q - 1 \) consider
\[
\# \{ 0 \leq n < N : s_q(2n) \equiv j \pmod{q} \}.
\]
In 1983, F. M. Dekking conjectured that this quantity is greater than \( N/q \) and, respectively, less than \( N/q \) for infinitely many \( N \), thereby claiming an absence of a drift (or Newman) phenomenon. In this paper we prove his conjecture.

1. Introduction

Let \( q \geq 2 \) and denote by \( s_q : \mathbb{N} \rightarrow \mathbb{N} \) the sum-of-digits function in the \( q \)-ary digital representation of integers. In his influential paper from 1968, Gelfond [5] proved the following result.\(^1\)

Theorem 1.1. Let \( q, d, m \geq 2 \) and \( a, j \) be integers with \( 0 \leq a < d \) and \( 0 \leq j < m \). If \( (m, q - 1) = 1 \) then
\begin{equation}
\# \{ 0 \leq n < N : n \equiv a \pmod{d}, \ s_q(n) \equiv j \pmod{m} \} = \frac{N}{dm} + g(N),
\end{equation}
where \( g(N) = O_q(N^\lambda) \) with \( \lambda = \frac{1}{2 \log q} \log \frac{q \sin(\pi/2m)}{\sin(\pi/2mq)} < 1. \)

\(^1\)As usual, we write \( f(N) = O(1) \) if \( |f(N)| < C \) for some absolute constant \( C \), and \( f(N) = O_q(1) \) if the implied constant depends on \( q \).
Shevelev [8, 9] recently determined the optimal exponent \( \lambda \) in the error term in Gelfond’s asymptotic formula when \( q = m = 2 \), and Shparlinski [10] showed that in this case it can be arbitrarily small for sufficiently large primes \( d \).

The oscillatory behaviour of the error term \( g(N) \) in (1.1) is still not completely understood. The story can be said to have originated with the observation by Moser in the 1960s that for the quintuple of parameters

\[(1.2) \quad (q, a, d, j, m) \equiv (2, 0, 3, 0, 2)\]

the error term seems to have constant positive sign, \( i.e. \), \( g(N) > 0 \) for all \( N \geq 1 \). In 1969, Newman [7] (with a much more precise result by Coquet [2]) proved this observation and there is at present a large number of articles which establish so-called Newman phenomena, Newman-like phenomena or drifting phenomena for general classes of quintuples \((q, a, d, j, m)\) extending (1.2). The two main techniques come from a direct inspection of the recurrence relations using the \( q \)-additivity of the sum-of-digits function, and from the determination of the maximal and minimal value of a related fractal function which is continuous but nowhere differentiable [6, 2, 11]. We refer the reader to the monograph of Allouche and Shallit [1] and the article of Drmota and Stoll [4] for a list of references. Characterizing all \((q, a, d, j, m)\) for which one has a Newman-like phenomenon is still wide open.

The aim of the present article is to prove a related conjecture Dekking (see [3, “Final Remark”, p. 32-11]) made in 1983 at the Séminaire de Théorie des Nombres de Bordeaux concerning a non-drifting phenomenon, that is, a situation where the error \( g(N) \) is oscillating in sign (as \( N \to \infty \)). To our knowledge, this conjecture has not yet been addressed in the literature, and we will provide a self-contained proof here.

**Conjecture (Dekking, 1983):** Let \( q \geq 2 \) and \( 0 \leq j < q \) and set

\[(q, a, d, j, m) \equiv (q, 0, 2, j, q).\]

Then \( g(N) < 0 \) and \( g(N) > 0 \) infinitely often.

Dekking was mostly interested in finding the optimal error term in (1.1) (or, as he puts it, the typical exponent of the error term) and obtained various results for the cases \( q = 2 \), \( d \) arbitrary, and \( d = 2 \), \( q \) arbitrary. As for the conjecture, he proved the case of \( q = 3 \), \( j = 0, 1, 2 \) via an argument with a geometrical flavour.

Our main result is as follows.

**Theorem 1.2.** Let \( q \geq 2 \), \( 0 \leq j < q \) and set

\[(q, a, d, j, m) \equiv (q, a, d, j, q).\]
On a conjecture of Dekking

(i) If \( d \mid q \), then \( g(N) = O(1) \) and \( g(N) \) changes signs infinitely often as \( N \to \infty \).

(ii) If \( d \mid q - 1 \), then \( g(N) \) can take arbitrarily large positive values as well as arbitrarily large negative values as \( N \to \infty \).

In the case of \( d = 2 \) this proves Dekking’s conjecture and covers all bases \( q \geq 2 \).

2. Proof of Theorem 1.2

For an integer \( n \geq 0 \), we write

\[ n = (\varepsilon_k, \varepsilon_{k-1}, \ldots, \varepsilon_0)_q \]

to refer to its \( q \)-ary digital expansion \( n = \sum_{i=0}^{k} \varepsilon_i q^i \). Let \( U(n) = \{ z \in \mathbb{C} \mid z^n = 1 \} \) denote the set of the \( n \)th roots of unity. We will make use of the following well-known formula from discrete Fourier analysis.

**Proposition 2.1.** Let \( f(x) = \sum_{k=0}^{\infty} a_k x^k \in \mathbb{C}[x], \ n \geq 1, \ l \geq 0 \) and set \( \omega_n = e^{2\pi i/n} \). Then

\[
\sum_{k \equiv l \pmod{n}} a_k x^k = \frac{1}{n} \sum_{s=0}^{n-1} \omega_n^{-ls} f(\omega_n^s x).
\]

**Proof.** The coefficient of \( x^j \) in \( \frac{1}{n} \sum_{s=0}^{n-1} \omega_n^{-ls} f(\omega_n^s x) \) is \( \frac{1}{n} \sum_{s=0}^{n-1} a_j \omega_n^{s(j-l)} \), that is \( a_j \) if \( j \equiv l \pmod{n} \) and 0 otherwise. \( \square \)

We deal with (i) \( d \mid q \) and (ii) \( d \mid q - 1 \) in Theorem 1.2 separately in the two subsequent sections.

2.1. The case \( d \mid q \). For \( d = 2, q \) even, Dekking remarked and left to the readers of his article (see [3, Remark before Proposition 5, p.32-08]) that the typical exponent \( \lambda \) equals 0, i.e., \( g(N) = O(1) \). This is due to the fact that when \( q \) is even then the parity of an integer is completely encoded in the last digit of its base \( q \) expansion. A similar situation applies when \( d \mid q \). In order to find the oscillatory behaviour of \( g(N) \), we calculate \( g(N) \) explicitly.

Define

\[ f_j(n) = c_j(n) - \frac{1}{q}, \]

where

\[ c_j(n) = \begin{cases} 1 & \text{if } s_q(n) \equiv j \pmod{q}; \\ 0 & \text{otherwise}. \end{cases} \]

Consider

\[
D_j(N) = \sum_{0 \leq n < N \atop n \equiv a \pmod{d}} f_j(n), \tag{2.1}
\]
thus

\[ g(N) = D_j(N) - \frac{N}{dq} + \frac{1}{q} \left[ \frac{N-a}{d} \right]. \]

We want to find infinitely many values of \( N \) such that \( g(N) > 0 \), respectively, \( g(N) < 0 \). Since an integer in base \( q \) (with \( q \) divisible by \( d \)) gives remainder \( a \) mod \( d \) if and only if its last digit in base \( q \) gives remainder \( a \) mod \( d \), we get for \( N = (\varepsilon_k, \ldots, \varepsilon_0)_q \),

\[
D_j(N) = \sum_{r=2}^{k} \sum_{\delta=0}^{\varepsilon_r-1} \sum_{\substack{0 \leq i_0, i_1, \ldots, i_{r-1} \leq q-1 \atop i_0 \equiv a \ (\text{mod } d)} } f_j((\varepsilon_k, \ldots, \varepsilon_{r+1}, \delta, i_{r-1}, \ldots, i_0)_q) \\
+ \sum_{\delta=0}^{\varepsilon_2-1} \sum_{\substack{q-1 \leq i_0 \leq q-1 \atop i_0 \equiv a \ (\text{mod } d)} } f_j((\varepsilon_k, \ldots, \varepsilon_2, \delta, i_0)_q) \\
+ \sum_{\substack{\delta=0 \atop \delta \equiv a \ (\text{mod } d)}}^{\varepsilon_0-1} f_j((\varepsilon_k, \ldots, \varepsilon_1, \delta)_q).
\]

For \( r \geq 2 \) we get

\[
\sum_{\substack{0 \leq i_0, i_1, \ldots, i_{r-1} \leq q-1 \atop i_0 \equiv a \ (\text{mod } d)}} f_j((\varepsilon_k, \ldots, \varepsilon_{r+1}, \delta, i_{r-1}, \ldots, i_0)_q) = D_{j-\varepsilon_k-\ldots-\varepsilon_{r+1}-\delta}(q^r) = 0.
\]

Set \( \alpha = j - s_q(N) + \varepsilon_1 + \varepsilon_0 \) and \( \beta = j - s_q(N) + \varepsilon_0 \). For the other two terms we then get by a direct calculation,

\[
(2.3) \sum_{\delta=0}^{\varepsilon_2-1} \sum_{\substack{0 \leq i_0 < q \atop i_0 \equiv a \ (\text{mod } d)}} f_j((\varepsilon_k, \ldots, \varepsilon_2, \delta, i_0)_q) = -\frac{\varepsilon_1}{d} + \sum_{\delta=0}^{\varepsilon_1-1} \sum_{\substack{0 \leq i_0 < q \atop i_0 \equiv a - \delta \ (\text{mod } d)}} 1
\]

and

\[
(2.4) \sum_{\substack{\delta=0 \atop \delta \equiv a \ (\text{mod } d)}}^{\varepsilon_0-1} f_j((\varepsilon_k, \ldots, \varepsilon_1, \delta)_q) = -\frac{1}{q} \left\lfloor \frac{\varepsilon_0 - a}{d} \right\rfloor + \sum_{\substack{0 \leq \delta < \varepsilon_0 \atop \delta \equiv a \ (\text{mod } d) \atop \delta \equiv \beta \ (\text{mod } q)}} 1.
\]

From (2.2), (2.3) and (2.4) it is straightforward to find sequences of positive integers \( N \) with \( g(N) > 0 \), respectively \( g(N) < 0 \). In fact, if \( a \neq 0 \) we can take all \( N \) with \( \varepsilon_1 = 0, \varepsilon_0 = a \) to get \( g(N) = -\frac{a}{dq} < 0 \). For \( a = 0 \) we take all \( N \) with \( \varepsilon_1 = 1, \varepsilon_0 = a \) and \( s_q(N) \neq j+1 \ (\text{mod } d) \) to get \( g(N) = -1/d < 0 \). On the other hand, if \( a + 1 < q \) we may take all \( N \) with \( \varepsilon_1 = 0, \varepsilon_0 = a + 1 \) to find \( g(N) = 1 + \frac{1}{d} - \frac{a+1}{qd} - \frac{1}{q} > 0 \). If \( a + 1 = q \) (which again implies
On a conjecture of Dekking

21

2.2. The case $d \mid q - 1$. In what follows, set

$$E_{a,j}(k) = \# \{ 0 \leq n < q^k : n \equiv a \pmod{d}, \ s_q(n) \equiv j \pmod{q} \},$$

where $a, j$ are fixed integers with $0 \leq a < d$, $0 \leq j < q$ and $k \geq 1$. Consider the generating polynomial in two variables

$$P(x, y) = \prod_{i=0}^{k-1} (1 + xy^{q^i} + x^2y^{2q^i} + \ldots + x^{q-1}y^{(q-1)q^i}),$$

which encodes the digits of integers less than $q^k$ in base $q$. Denote by $E_{a,j}(k)$

$$E_{a,j}(k) - \frac{q^{k-1}}{d} = \frac{1}{dq} \sum_{\omega \in U(q)} \sum_{\varepsilon \in U(d), \omega \varepsilon \neq 1} \omega^{-j} \varepsilon^{-a} \left( \frac{1 - \varepsilon}{1 - \omega \varepsilon} \right)^k.$$
then we conclude that $\omega$ and $\omega'$ have the same argument so $\omega = \omega'$, and then $\varepsilon = \varepsilon'$. This means that there are no cancellations in (2.5).

Write

$$R = \max \left\{ \frac{1 - \varepsilon}{1 - \omega \varepsilon} : \omega \in U(q), \varepsilon \in U(d), \omega \varepsilon \neq 1 \right\}$$

and let $r_1, r_2, \ldots, r_h$ be all of the numbers $(1 - \varepsilon)/(1 - \omega \varepsilon)$ whose absolute value equals $R$.

The set $U(d)$ divides the unit circle into $d \geq 2$ equal parts, so it always contains an element $\varepsilon_0$ in the open half-plane $\text{Re}(\varepsilon) < 0$. Similarly, $U(q)$ must contain an element $\omega_0$ in the closed half-plane $\text{Re}(\varepsilon_0 \omega) \geq 0$. Then $|1 - \varepsilon_0| > \sqrt{2}$ while $|1 - \omega_0 \varepsilon_0| \leq \sqrt{2}$, thus

$$\left| \frac{1 - \varepsilon_0}{1 - \omega_0 \varepsilon_0} \right| > 1.$$

Note also that $\omega_0 \varepsilon_0 \neq 1$ as $(d, q) = 1$ and $\varepsilon_0 \neq 1$.

It follows that $R > 1$, which in particular implies that the value 1 is not among these $r_i$. Then, as $k \to \infty$,

$$\sum_{\omega \in U(q), \varepsilon \in U(d), \omega \varepsilon \neq 1} \omega^{-j} \varepsilon^{-a} \left( \frac{1 - \varepsilon}{1 - \omega \varepsilon} \right)^k \sim R^k \sum_{i=1}^{h} c_i \left( \frac{r_i}{R} \right)^k,$$

for certain $c_i \in \mathbb{C}$ which are not all zero. As the $r_i$ all have arguments equal to rational multiples of $\pi$, the $r_i/R$, $i = 1, \ldots, h$, are roots of unity. Therefore there exists an integer $M \geq 1$ such that $(r_i/R)^M = 1$ for all $i$. Write

$$c'(k) = \sum_{i=1}^{h} c_i \left( \frac{r_i}{R} \right)^k.$$

Since $E_{a,j}(k)$ is real and $c'(k + M) = c'(k)$ for all $k$ we must have that $c'(k) \in \mathbb{R}$ for all $k$. Moreover,

$$\sum_{k=0}^{M-1} c'(k) = \sum_{i=1}^{h} c_i \sum_{k=0}^{M-1} \left( \frac{r_i}{R} \right)^k = 0,$$

since $r_i$ is not real for all $i$. Thus, among all the $c'(k)$ there is at least one positive and at least one negative value. Let $-c'_1 = c'(k_1) < 0$ be the smallest negative value and $c_2 = c'(k_2) > 0$ be the largest positive value among them. Then, as $k \to \infty$,

$$E_{a,j}(k) - \frac{q^{k-1}}{d} \sim -\frac{c'_1}{dq} R^k < 0, \quad \text{for } k \equiv k_1 \pmod{M}$$

and

$$E_{a,j}(k) - \frac{q^{k-1}}{d} \sim \frac{c'_2}{dq} R^k > 0, \quad \text{for } k \equiv k_2 \pmod{M}.$$

This completes the proof.
3. Acknowledgment

We would like to express our gratitude to J.-P. Allouche, F. M. Dekking and J. Shallit for related correspondence on this problem. T. Stoll was supported by the Agence Nationale de la Recherche, grant ANR-10-BLAN 0103 MUNUM.

References


