



HAL
open science

Sums of the digits in bases 2 and 3, I

Jean-Marc Deshouillers, Laurent Habsieger, Shanta Laishram, Bernard Landreau

► **To cite this version:**

Jean-Marc Deshouillers, Laurent Habsieger, Shanta Laishram, Bernard Landreau. Sums of the digits in bases 2 and 3, I. *Number Theory - Diophantine Problems, Uniform Distribution and Applications*, 2017. hal-02480241

HAL Id: hal-02480241

<https://hal.science/hal-02480241>

Submitted on 15 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sums of the digits in bases 2 and 3, I

Jean-Marc Deshouillers,
Laurent Habsieger, Shanta Laishram, Bernard Landreau

To Robert Tichy, for his 60th birthday

Abstract

Let $b \geq 2$ be an integer and let $s_b(n)$ denote the sum of the digits of the representation of an integer n in base b . For sufficiently large N , one has

$$\text{Card}\{n \leq N : |s_3(n) - s_2(n)| \leq 0.1457232 \log n\} > N^{0.970359}.$$

The proof only uses the separate distributions of the values of $s_2(n)$ and $s_3(n)$.

1 Introduction

For integers $b \geq 2$ and $n \geq 0$, we denote by “the sum of the digits of n in base b ” the quantity

$$s_b(n) = \sum_{j \geq 0} \varepsilon_j, \text{ where } n = \sum_{j \geq 0} \varepsilon_j b^j \text{ with } \forall j : \varepsilon_j \in \{0, 1, \dots, b-1\}.$$

Our attention on the question of the proximity of $s_2(n)$ and $s_3(n)$ comes from the apparently non related question of the distribution of the least non zero digit of $n!$ in base 12 (cf. [1] and [2]).

Computation shows that there are 48 266 671 607 numbers up to 10^{12} for which $s_2(n) = s_3(n)$, but it seems to be unknown whether there are infinitely many integers n for which $s_2(n) = s_3(n)$ or even for which $|s_2(n) - s_3(n)|$ is significantly small.

The first result we mention may well be known but we did not find trace of it in the literature. We recall that a sequence $\mathcal{A} \subset \mathbb{N}$ of integers is said to have asymptotic natural density 1 if

$$\text{Card}\{n \leq N : n \in \mathcal{A}\} = N + o(N).$$

Theorem 1. *For any ψ be a function tending to infinity with its argument. The sequence of natural numbers n for which*

$$\begin{aligned} \left(\frac{1}{\log 3} - \frac{1}{\log 4}\right) \log n - \psi(n)\sqrt{\log n} &\leq s_3(n) - s_2(n) \\ &\leq \left(\frac{1}{\log 3} - \frac{1}{\log 4}\right) \log n + \psi(n)\sqrt{\log n} \end{aligned}$$

has asymptotic natural density 1.

Our main result is that there exist infinitely many n for which $|s_3(n) - s_2(n)|$ is significantly smaller than $\left(\frac{1}{\log 3} - \frac{1}{\log 4}\right) \log n = 0.18889\dots \log n$. More precisely

Theorem 2. *For sufficiently large N , one has*

$$\text{Card}\{n \leq N : |s_3(n) - s_2(n)| \leq 0.1457232 \log n\} > N^{0.970359}. \quad (1)$$

This result is obtained by looking separately at the distributions of $(s_2(n))_n$ and $(s_3(n))_n$, without using any information (nor hypothesis) on their joint distribution, nor any Diophantine argument.

In Section 2, we provide a heuristic approach to Theorems 1 and 2; the actual distribution of $(s_2(n))_n$ and $(s_3(n))_n$ is studied in Section 3. The proof of Theorems 1 and 2 are given in Sections 4 and 5.

2 A heuristic approach

As a warm up for the actual proofs, we sketch a heuristic approach. A positive integer n may be expressed as

$$n = \sum_{j=0}^{J(n)} \varepsilon_j b^j, \text{ with } J(n) = \left\lfloor \frac{\log n}{\log b} \right\rfloor.$$

If we consider an interval of integers around N , the smaller is j the more equidistributed are the ε_j 's, and the smaller is a family $\mathcal{J} = \{j_1 < j_2 < \dots < j_s\}$ the more independent are the ε_j 's for $j \in \mathcal{J}$. Thus a first model for $s_b(n)$ for n around N is to consider a sum of $\left\lfloor \frac{\log N}{\log b} \right\rfloor$ independent random variables uniformly distributed in $\{0, 1, \dots, b-1\}$. Thinking of the central limit theorem, we even consider a continuous model, representing $s_b(n)$, for n around N by a Gaussian random variable $S_{b,N}$ with expectation and dispersion given by

$$\mathbb{E}(S_{b,N}) = \frac{(b-1)\log N}{2\log b} \text{ and } \mathbb{V}(S_{b,N}) = \frac{(b^2-1)\log N}{12\log b}.$$

In particular

$$\mathbb{E}(S_{2,N}) = \frac{\log N}{\log 4} \text{ and } \mathbb{E}(S_{3,N}) = \frac{\log N}{\log 3},$$

and their standard deviations have the order of magnitude $\sqrt{\log N}$.

Towards Theorem 1. If $\psi(N)$ is a function which tends to infinity when N tends to infinity, we have

$$\mathbb{P}\left(\left|S_{3,N} - \frac{\log N}{\log 3}\right| \leq \psi(N)\sqrt{\log N}\right) = 1 + o(1)$$

and

$$\mathbb{P}\left(\left|S_{2,N} - \frac{\log N}{\log 4}\right| \leq \psi(N)\sqrt{\log N}\right) = 1 + o(1),$$

which implies that when N tends to infinity we have almost surely

$$\left|(S_{3,N} - S_{2,N}) - \left(\frac{1}{\log 3} - \frac{1}{\log 4}\right)\log N\right| \leq 2\psi(N)\sqrt{\log N}.$$

Towards Theorem 2. If we wish to deal with a difference $|s_3(n) - s_2(n)| < u \log n$ for some $u < \left(\frac{1}{\log 3} - \frac{1}{\log 4}\right)$ we must, by what we have seen above, consider events of asymptotic probability zero, which means that a heuristic approach must be substantiated by a rigorous proof. Our key remark is that the dispersion of $S_{3,N}$ is larger than that of $S_{2,N}$; this implies the following: the probability that $S_{3,N}$ is at a distance d from its mean is larger than the probability that $S_{2,N}$ is at a distance d from its mean. So,

we have the hope to find some $u < \left(\frac{1}{\log 3} - \frac{1}{\log 4}\right)$ such that the probability that $|S_{2,N} - \mathbb{E}(S_{2,N})| > u \log N$ is smaller than the probability that $S_{3,N}$ is very close to $\mathbb{E}(S_{2,N})$. This will imply that for some ω we have $|S_{3,N}(\omega) - S_{2,N}(\omega)| \leq u \log N$.

3 On the distribution of the values of $s_2(n)$ and $s_3(n)$

In order to prove Theorems 1 and 2 we need

- a central limit theorem for s_2 and s_3 ,
- an upper bound for the tail of the distribution of s_2 ,
- a lower bound for the tail of the distribution of s_3 .

3.1 Central limit theorem for s_b

Proposition 1. *Let ψ be any function tending to infinity with its argument. We have, as N tends to infinity*

$$\text{Card} \left\{ n \leq N : \left| s_b(n) - \frac{(b-1) \log n}{2 \log b} \right| \leq \psi(n) \sqrt{\log n} \right\} = N + o(N).$$

We leave it to the Reader, as an exercise in elementary calculus to show that Proposition 1 is a consequence of (and indeed is equivalent to) the following

Proposition 2. *Let φ be any function tending to infinity with its argument. We have, as the integer L tends to infinity*

$$b^{-L} \text{Card} \left\{ 0 \leq n < b^L : \left| s_b(n) - \frac{(b-1)L}{2} \right| \leq \varphi(L) \sqrt{L} \right\} = 1 + o(1). \quad (2)$$

Proof. We consider L independent random variables X_1, X_2, \dots, X_L which are uniformly distributed in $\{0, 1, \dots, b\}$, and we let $\Sigma_L = X_1 + X_2 + \dots + X_L$

be their sum. For any integer m one has

$$\begin{aligned}
b^{-L} \text{Card} \{0 \leq n < b^L : s_b(n) = m\} &= b^{-L} \sum_{\substack{\ell_0 + \dots + \ell_{b-1} = L \\ \ell_1 + 2\ell_2 + \dots + (b-1)\ell_{b-1} = m}} \frac{L!}{\ell_0! \cdots \ell_{b-1}!} \\
&= \mathbb{P}(\Sigma_L = m). \tag{3}
\end{aligned}$$

Since

$$\mathbb{E}(\Sigma_L) = \frac{(b-1)L}{2} \text{ and } \mathbb{V}(\Sigma_L) = \frac{(b^2-1)L}{12},$$

Proposition 2 is but a reformulation of the central limit theorem applied to the sequence $(\Sigma_L = X_1 + X_2 + \dots + X_L)_L$. \square

3.2 Upper bound for the tail of the distribution of s_2

Proposition 3. *Let $\lambda \in (0, 1)$. For any*

$$\nu > 1 - ((1 - \lambda) \log(1 - \lambda) + (1 + \lambda) \log(1 + \lambda)) / \log 4$$

and any sufficiently large integer H , we have

$$\text{Card}\{n < 2^{2H} : |s_2(n) - H| \geq \lambda H\} \leq 2^{2H\nu}. \tag{4}$$

Proof. When $b = 2$, the distribution of the values of $s_2(n)$ is simply binomial; Equation (3) becomes

$$\text{Card} \{0 \leq n < 2^{2H} : s_2(n) = m\} = \binom{2H}{m}.$$

Using the fact that the sequence $(\binom{2H}{m})_m$ is symmetric and unimodal plus Stirling's formula, we obtain that when $m \leq (1 - \lambda)H$ or $m \geq (1 + \lambda)H$, one has

$$\begin{aligned}
\binom{L}{m} &\leq H^{O(1)} \frac{(2H)^{2H}}{((1 - \lambda)H)^{(1 - \lambda)H} ((1 + \lambda)H)^{(1 + \lambda)H}} \\
&\leq H^{O(1)} \left(\frac{2^2}{(1 - \lambda)^{(1 - \lambda)} (1 + \lambda)^{(1 + \lambda)}} \right)^H \\
&\leq H^{O(1)} \left(2^{1 - ((1 - \lambda) \log(1 - \lambda) + (1 + \lambda) \log(1 + \lambda)) / 2 \log 2} \right)^{2H}.
\end{aligned}$$

Relation (4) comes from the above inequality and the fact that the left hand side of (4) is the sum of at most $2H$ such terms. \square

3.3 Lower bound for the tail of the distribution of s_3

Proposition 4. *Let K be sufficiently large an integer. We have*

$$\text{Card}\{n < 3^L : s_3(n) = \lfloor L \log 3 / \log 4 \rfloor\} \geq 3^{0.9703591L}. \quad (5)$$

Proof. We use (3), and select one term in the sum. We choose

$$l_2 = \lfloor 0.235001143L \rfloor; l_1 = \lfloor L \log 3 / \log 4 \rfloor - 2l_2; l_0 = L - l_1 - l_2.$$

The integers n which have l_0 digits 0, l_1 digits 1 and l_2 digits 2, in base 3, have L digits and thus lie in the interval $[0, 3^L)$, have a sum of digits $l_1 + 2l_2 = \lfloor L \log 3 / \log 4 \rfloor$ and their number is $L! / (l_0! l_1! l_2!)$. A straightforward application of Stirling's formula, similar to the one used in the previous subsection, leads to (5). \square

4 Proof of Theorem 1

Let us consider the two sets

$$\mathcal{A}_N = \left\{ n \leq N : \left| s_2(n) - \frac{\log n}{\log 4} \right| \leq \psi(n) \sqrt{\log n} \right\}$$

and

$$\mathcal{B}_N = \left\{ n \leq N : \left| s_3(n) - \frac{\log n}{\log 3} \right| \leq \psi(n) \sqrt{\log n} \right\}.$$

Since those two sets of integers are included in $[0, N]$, we have, using Proposition 1

$$\text{Card}(\mathcal{A}_N \cap \mathcal{B}_N) = \text{Card}(\mathcal{A}_N) + \text{Card}(\mathcal{B}_N) - \text{Card}(\mathcal{A}_N \cup \mathcal{B}_N) \geq N + o(N).$$

When n belongs to $\mathcal{A}_N \cap \mathcal{B}_N$, it satisfies the double inequality of Theorem 1 (indeed with $2\psi(n)$ instead of $\psi(n)$, which is irrelevant). \square

5 Proof of Theorem 2

Let N be sufficiently large an integer. We let $K = \lfloor \log N / \log 3 \rfloor - 2$ and $H = \lfloor (K - 1) \log 3 / \log 4 \rfloor + 2$. We notice that we have

$$N/27 \leq 3^{K-1} < 3^K < 2^{2H} \leq N. \quad (6)$$

We use Proposition 3 with $\lambda = 0.14572319 \log 4$, which leads to

$$\text{Card}\{n \leq 2^{2H} : |s_2(n) - H| \geq \lambda H\} \leq 2^{0.97039581 \times 2H} \leq N^{0.97039581}. \quad (7)$$

For any $n \in [2 \cdot 3^{K-1}, 3^K)$ we have $s_3(n) = 2 + s_3(n - 2 \cdot 3^{K-1})$ and so it follows from Proposition 4 that we have

$$\begin{aligned} & \text{Card}\{n \in [2 \cdot 3^{K-1}, 3^K) : s_3(n) = H\} \\ &= \text{Card}\{n < 3^{K-1} : s_3(n) = H - 2\} \\ &= \text{Card}\{n < 3^{K-1} : s_3(n) = \lfloor (K - 1) \log 3 / \log 4 \rfloor\} \\ &\geq 3^{0.9703591(K-1)} \geq N^{0.97035905}. \end{aligned}$$

This implies that we have

$$\text{Card}\{n \leq 2^{2H} : s_3(n) = H\} \geq N^{0.970395905}. \quad (8)$$

From (7) and (8), we deduce that for N sufficiently large, we have

$$\text{Card}\{n \leq N : \lfloor s_2(n) - s_3(n) \rfloor \leq 0.1457232 \log n\} \geq N^{0.970359}.$$

□

References

- [1] J-M. Deshouillers, I. Ruzsa, The least non zero digit of $n!$ in base 12, *Pub. Math. Debrecen* 79 (2011), 395-400.
- [2] J-M. Deshouillers, A footnote to *The least non zero digit of $n!$ in base 12*, *Unif. Distrib. Theory*, 7 (2012), 71-73.