Sums of the digits in bases 2 and 3

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

Sums of the digits in bases 2 and 3

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

November 23, 2016

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

$$\mathrm{Card}\{n \leq N : |s_3(n) - s_2(n)| \leq 0.1457205 \log n\} \; > \; N^{0.970359}.$$

The proof only uses the separate (or marginal) 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

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

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 last non zero digit of $n!$ in base 12 (cf. [2] and [3]).[1]

[1]Indeed, if the last non zero digit of $n!$ in base 12 belongs to $\{1, 2, 5, 7, 10, 11\}$ then $|s_3(n) - s_2(n)| \leq 1$; this seems to occur infinitely many times.

Computation shows that there are 48 266 671 607 positive integers 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.

We do not know the first appearance of the result we quote as Theorem 1; in any case, it is a straightforward application of the fairly general main result of N. L. Bassily and I. Kátai [1]. We recall that a sequence $\mathcal{A} \subset \mathbb{N}$ of integers is said to have asymptotic natural density 1 if

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

Theorem 1. *Let 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 $\left(\frac{1}{\log 3} - \frac{1}{\log 4}\right)\log n = 0.18889... \log n$. More precisely we have the following:

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

$$\mathrm{Card}\{n \leq N : |s_3(n) - s_2(n)| \leq 0.1457205 \log n\} \ > \ N^{0.970359}. \qquad (1)$$

The mere information we use in proving Theorem 2 is the knowledge of the separate (or marginal) distributions of $(s_2(n))_n$ and $(s_3(n))_n$, without using any further information concerning their joint distribution.

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 Theorem 2 is given in Sections 4.

Let us formulate three remarks as a conclusion to this introductory section.

It seems that our present knowledge of the joint distribution of s_2 and s_3 (cf. for exemple C. Stewart [5] for a Diophantine approach or M. Drmota [4] for a probabilistic one) does not permit us to improve on Theorem 2.

Theorem 2 can be extended to any pair of distinct bases, say q_1 and q_2: more than computation, the Authors have deliberately chosen to present an idea to the Dedicatee.

Although we could not prove it, we believe that Theorem 2 represents the limit of our method.

Acknowledgements The authors are indebted to Bernard Bercu for several discussions on the notion of "spacing" between two random variables, a notion to be developped later. They also thank the Referees for their constructive comments. The first, third and fourth authors wish to thank the Indo-French centre CEFIPRA for the support permitting them to collaborate on this project (ref. 5401-A). The first named author acknowledges with thank the support of the French-Austrian project MuDeRa (ANR and FWF).

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

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

If we consider an interval of integers around N, the smaller is j the more equidistributed are the $\varepsilon_j(n)$'s, and the smaller are the elements of a family $\mathcal{J} = \{j_1 < j_2 < \cdots < j_s\}$ the more independent are the $\varepsilon_j(n)$'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 $\{0, 1, \ldots, b-1\}$. 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 variance given by

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

In particular

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

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

Towards Theorem 1. In [1], it is proved that a central limit theorem actually holds for s_b; more precisely, the following proposition is the special case of the first relation in the main Theorem of [1], with $f(n) = s_b(n)$ and $P(X) = X$.

Proposition 1. *For any positive y, as x tend to infinity, one has*

$$\frac{1}{x}\operatorname{Card}\left\{n<x\,:\,|s_b(n)-\mathbb{E}\left(S_{b,n}\right)|<y\left(\mathbb{V}\left(S_{b,n}\right)\right)^{1/2}\right\}\to\frac{1}{\sqrt{2\pi}}\int_{-y}^{y}e^{-t^2/2}dt.$$

Theorem 1 easily follows from Proposition 1: the set under our consideration is the intersection of 2 sets of density 1.

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 variance 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 that 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 Theorem 2 we need

- 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 Upper bound for the tail of the distribution of s_2

Proposition 2. *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

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

Proof. When $b = 2$, the distribution of the values of $s_2(n)$ is simply binomial; we thus get

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

Using the fact that the sequence (in m) $\binom{2H}{m}$ 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

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

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

3.2 Lower bound for the tail of the distribution of s_3

Proposition 3. *Let L be sufficiently large an integer. We have*

$$\mathrm{Card}\{n < 3^L : s_3(n) = \lfloor L\log 3/\log 4\rfloor\} \geq 3^{0.970359238L}. \tag{3}$$

Proof. The positive integer L being given, we write any integer $n \in [0, 3^L)$ in its non necessarily proper representation, as a chain of exactly L characters, $\ell_i(n)$ of them being equal to i, for $i \in \{0,1,2\}$, the sum $\ell_0(n) + \ell_1(n) + \ell_2(n)$ being equal to L, the total number of digits in this representation[2]. One has

$$\mathrm{Card}\left\{0 \leq n < 3^L : s_3(n) = m\right\} = \sum_{\substack{\ell_0+\ell_1+\ell_2=L \\ \ell_1+2\ell_2=m}} \frac{L!}{\ell_0!\ell_1!\ell_2!}. \tag{4}$$

[2]For example, when $L = 5$, the number "sixty" will be represented as 02020. Happy palindromic birthday, Robert!

In order to get a lower bound for the left hand side of (4), it is enough to select one term in its right hand side. We choose

$$l_2 = \lfloor 0.235001144L \rfloor\,;\; l_1 = \lfloor L \log 3/\log 4 \rfloor - 2\,l_2\,;\; l_0 = L - l_1 - l_2.$$

A straightforward application of Stirling's formula, similar to the one used in the previous subsection, leads to (3). □

4 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

$$N/81 \leq 3^{K-1} < 3^K < 2^{2H} \leq N. \tag{5}$$

We use Proposition 2 with $\lambda = 0.14572049 \log 4$, which leads to

$$\mathrm{Card}\{n \leq 2^{2H} : |s_2(n) - H| \geq \lambda H\} \leq 2^{0.970359230 \times 2H} \leq N^{0.970359230}. \tag{6}$$

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 3 that we have

$$\begin{aligned}
&\mathrm{Card}\{n \in [2 \cdot 3^{K-1}, 3^K) : s_3(n) = H\}\\
&= \mathrm{Card}\{n < 3^{K-1}) : s_3(n) = H - 2\}\\
&= \mathrm{Card}\{n < 3^{K-1}) : s_3(n) = \lfloor (K-1)\log 3/\log 4 \rfloor\}\\
&\geq 3^{0.970359238(K-1)} \geq N^{0.970359237}.
\end{aligned}$$

This implies that we have

$$\mathrm{Card}\{n \leq 2^{2H} : s_3(n) = H\} \geq N^{0.970359237}. \tag{7}$$

From (6) and (7), we deduce that for N sufficiently large, we have

$$\mathrm{Card}\{n \leq N : |s_2(n) - s_3(n)| \leq 0.1457205 \log n\} \geq N^{0.970359}.$$

□

References

[1] N. L. Bassily, I. Kátai, Distribution of the values of q-additive functions, on polynomial sequences, Acta Math. Hungar. 68 (1995), 353-361.

[2] J-M. Deshouillers, I. Ruzsa, The least non zero digit of n! in base 12, Pub. Math. Debrecen 79 (2011), 395-400.

[3] J-M. Deshouillers, A footnote to *The least non zero digit of n! in base 12*, Unif. Distrib. Theory, 7 (2012), 71-73.

[4] M. Drmota, The joint distribution of q-additive functions, Acta Arith. 100 (2001), 17-39.

[5] C. Stewart, On the representation of an integer in two different bases, J. Reine Angew. Math. 319 (1980), 6372.

Addresses

Jean-Marc DESHOUILLERS
jean-marc.deshouillers@math.u-bordeaux.fr
Bordeaux INP, Université de Bordeaux, CNRS
Institut Mathématique de Bordeaux, UMR 5251
33405 Talence, France

Laurent HABSIEGER
habsieger@math.univ-lyon1.fr
Université de Lyon, CNRS UMR 5208, Université Claude Bernard Lyon 1,
Institut Camille Jordan
69622 Villeurbanne Cedex, France

Shanta LAISHRAM
shanta@isid.ac.in
Indian Statistical Institute
7 SJS Sansanwal Marg
110016 New Delhi, India

Bernard LANDREAU
bernard.landreau@univ-angers.fr

Université d'Angers, CNRS
LAREMA Laboratoire Angevin de REcherche, UMR 6093, FR 2962
49045 Angers, France

Key words : sums of digits, expansion in bases 2 and 3

AMS 2010 Classification number: 11K16