

ON THE SUM OF DIGITS OF SOME SEQUENCES OF INTEGERS

JAVIER CILLERUELO, FLORIAN LUCA, JUANJO RUÉ, AND ANA ZUMALACÁRREGUI

ABSTRACT. Let $b \geq 2$ be a fixed positive integer. We show for a wide variety of sequences $\{a_n\}_{n=1}^{\infty}$ that for most n the sum of the digits of a_n in base b is at least $C_b \log(n)$, where C_b is a constant depending on b and on the sequence. We also give an explicit C_b in terms of the asymptotic behavior of the sequence under study. In particular, our approach covers several integer sequences arising from number theory and combinatorics.

1. INTRODUCTION

For a positive integer $b \geq 2$ let us denote by $s_b(m)$ the sum of the digits of the positive integer m when written in base b . Lower bounds for $s_b(m)$ when m runs through the members of a sequence with some interesting combinatorial meaning have been investigated before. For example, it follows from a result of Stewart ([13]; see also [8] for a slightly more general result), that in the case of Fibonacci numbers (namely, the sequence defined by $F_0 := 0$, $F_1 := 1$ and $F_{n+2} := F_{n+1} + F_n$ for all $n \geq 0$) the inequality

$$s_b(F_n) > c_1 \frac{\log n}{\log \log n}$$

holds for all $n \geq 3$ for some positive constant $c_1 := c_1(b)$ depending on b . In [9], it is shown that the inequality

$$s_b(n!) > c_2 \log n$$

holds for all $n \geq 1$, where $c_2 := c_2(b)$ is some positive constant depending on b . In [11], it was shown that if we put $b_n := \binom{2n}{n}$ and $c_n := \frac{1}{n+1} \binom{2n}{n}$ for the middle binomial coefficient and Catalan number, respectively, then both inequalities

$$(1) \quad s_b(b_n) \geq \varepsilon(n) \sqrt{\log n} \quad \text{and} \quad s_b(c_n) > \varepsilon(n) \sqrt{\log n}$$

hold on a set of n of asymptotic density equal to 1, where $\varepsilon(n)$ is any function tending to zero when n tends to infinity. In [12], it was shown that there is some positive constant $c_3 := c_3(b)$ depending on b such that if we put

$$A_n := \sum_{k=0}^n \binom{n}{k}^2 \binom{n+k}{k}^2$$

for the n th Apéry number, then the inequality

$$(2) \quad s_b(A_n) > c_3 \left(\frac{\log n}{\log \log n} \right)^{1/4}$$

holds on a set of n of asymptotic density 1. Some of the above results were superseded by the results from the recent paper [7], where it is shown that if

$\mathbf{r} := (r_0, r_1, \dots, r_m)$ is a fixed vector of nonnegative integers with $r_0 > 0$ and if we put

$$S(n) := \sum_{k=0}^n \binom{n}{k}^{r_0} \binom{n+k}{k}^{r_1} \cdots \binom{n+km}{k}^{r_m} \quad \text{for } n = 0, 1, \dots,$$

then for $\mathbf{r} \neq (1)$ there exists a positive constant $c_4 := c_4(b, \mathbf{r})$ depending on both b and \mathbf{r} such that the inequality

$$(3) \quad s_b(S(n)) > c_4 \frac{\log n}{\log \log n}$$

holds for almost all n . Note that inequality (3) improves (1) for the case of the middle binomial coefficients b_n because $c_n = S(n)$ for $\mathbf{r} = (2)$, as well as inequality (2) for the case of the Apéry numbers A_n because $A_n = S(n)$ for $\mathbf{r} = (2, 2)$.

In [10], it is shown that if $p(n)$ is the partition function of n , then the inequality

$$s_b(p(n)) > \frac{\log n}{7 \log \log n}$$

holds for almost all positive integers n .

The proofs of such results use a variety of methods from number theory, such as elementary methods, sieve methods, linear forms in logarithms and the subspace theorem of Evertse–Schlickewei–Schmidt [2].

In this work we focus on sequences $\{a_n\}_{n=1}^{\infty}$ of positive integers with a certain growth, and show, independently of the combinatorial properties of the sequence, that $s_b(a_n) > C_b \log n$ for almost every element in the sequence. In particular we focus on sequences satisfying the asymptotic behavior

$$a_n = e^{f(n)} (1 + O(n^{-\alpha})), \quad \alpha > 0,$$

and f is a two times differentiable function satisfying $f''(x) \asymp \frac{1}{x}$. Many sequences arising in number theory and combinatorics fit into this scheme. The most basic one, the number of permutations of a set of n elements is clearly a sequence of this kind, since from Stirling's approximation formula we have

$$(4) \quad n! = e^{n \log n - n + \log n + \frac{1}{2} \log 2\pi} (1 + O(n^{-1})).$$

Others, as set partitions (counted by Bell numbers), involutions (number of permutations of n elements with either fixed points or cycles of length 2) and fragmented permutations (unordered collections of permutations; in other words, *fragments* are obtained by breaking a permutation into pieces) have more involved expressions, but they also fit into these hypothesis (see [3] for further details).

In graph enumeration, many important families also follow these asymptotic expressions: the number of labelled trees (Cayley trees) with n vertices is equal to n^{n-1} . More generally, it is shown in [3] that families of labelled trees with degree constraints satisfy asymptotic formulas of the form

$$C_{\mathcal{T}} n^{-3/2} \gamma_{\mathcal{T}}^n \cdot n! (1 + O(n^{-1})) = e^{f_{\mathcal{T}}(n)} (1 + O(n^{-1})),$$

where the subindex \mathcal{T} indicates the considered constraint and the function $f_{\mathcal{T}}$ is given by

$$f_{\mathcal{T}}(n) = n \log n - n - \log n + n \log \gamma_{\mathcal{T}} + \log C_{\mathcal{T}} + \frac{1}{2} \log 2\pi.$$

Very recently, many authors have shown that several families of labelled graphs satisfies similar formulas: Giménez and Noy [5] (see also [6]) proved that the number of labelled planar graphs with n vertices follows an asymptotic formula of the form

$$C_0 n^{-7/2} \gamma^n \cdot n! (1 + O(n^{-1}))$$

where $\gamma \simeq 27.22687$. More generally, as it is shown in [4] (see also [1]), the number of labelled graphs which can be embedded in a surface of genus g satisfies a very similar formula (with the same growth factor). See Table 1 for the exact asymptotic developments of these sequences.

Sequence	Asymptotic
Permutations	$n!$
Involutions	$\frac{1}{2\sqrt{\pi}} n^{-1/2} e^{n/2-1/4} n^{-n/2} \cdot n! (1 + O(n^{-1/5}))$
Bell numbers	$\frac{e^{e^r-1}}{r^n \sqrt{2\pi r(r+1)} e^r} \cdot n! (1 + O(e^{-r/5}))$, $re^r = n + 1$
Fragmented permutations	$\frac{1}{2\sqrt{\pi}} n^{-3/4} e^{-1/2+2\sqrt{n}} \cdot n! (1 + O(n^{-3/4}))$
Cayley trees	$\frac{1}{\sqrt{2\pi}} n^{-3/2} e^n \cdot n! (1 + O(n^{-1}))$
Labelled trees	$C_T n^{-3/2} \gamma_T^n \cdot n! (1 + O(n^{-1}))$
Graphs on surfaces	$C_g n^{5(g-1)/2-1} \gamma^n \cdot n! (1 + O(n^{-1}))$

TABLE 1. Combinatorial families and their enumerative asymptotic behavior.

Our main result states the behavior of $s_b(a_n)$ for sequences of controlled growth described before.

Theorem 1. *Let $\{a_n\}_{n=1}^\infty$ be a sequence of positive integers with asymptotic behavior*

$$(5) \quad a_n = e^{f(n)} (1 + O(n^{-\alpha})), \text{ with } f''(x) \asymp \frac{1}{x},$$

for some $\alpha > 0$ and a two times differentiable function f . For any base $b \geq 2$, the inequality

$$s_b(a_n) > \frac{\beta \log n}{10 \log b}, \quad \beta = \min\{\alpha, 2/3\}$$

holds on a set of positive integers n of asymptotic density equals 1.

It is a straightforward calculation to check that condition (5) holds for all the sequences in Table 1, except for Bell numbers. In this case the asymptotic estimate is given implicitly in terms of r . In Section 3 we analyze this concrete case in detail.

1.1. Notation. We use Landau's symbol O and o as well as the Vinogradov's symbols \ll , \gg and \asymp with their usual meanings. Recall that $A = O(B)$, $A \ll B$ and $B \gg A$ are all equivalent to the fact that the inequality $|A| \leq cB$ holds with some constant c . The constants implied by these symbols in our arguments might depend in the number b . Furthermore, $A \asymp B$ means that both $A \ll B$ and $B \ll A$ hold.

2. PROOF OF THEOREM 1

Consider the following set of positive integers:

$$\mathcal{N}_b(x) := \left\{ n \in [x/2, x) : s_b(a_n) < \frac{\beta \log n}{10 \log b} \right\},$$

where $\beta \leq \alpha$ will be chosen later. We need to show that $\#\mathcal{N}_b(x) = o(x)$ as $x \rightarrow \infty$, therefore the conclusion of Theorem 1 will follow by replacing x by $x/2$, then by $x/4$, and so on, and summing up the resulting estimates.

For $n \in \mathcal{N}_b(x)$, we write

$$(6) \quad a_n = d_{k_1} b^{k_1} + d_{k_2} b^{k_2} + \cdots + d_{k_s} b^{k_s},$$

where $d_{k_1}, \dots, d_{k_s} \in \{1, \dots, b-1\}$ and $k_1 > k_2 > \cdots > k_s$. Observe that for $i = 1, \dots, s$ we have

$$a_n = d_{k_1} b^{k_1} + \cdots + d_{k_i} b^{k_i} (1 + E(i)),$$

where $E(i) = 0$, if $i = s$, and

$$E(i) = \frac{d_{k_{i+1}} b^{k_{i+1}} + \cdots + d_{k_s} b^{k_s}}{d_{k_1} b^{k_1} + \cdots + d_{k_i} b^{k_i}} = O(b^{k_{i+1}-k_1}),$$

if $i < s$. We choose $k(n)$ to be the smallest k_i such that $b^{k_i-k_1} > n^{-\beta}$.

From the definition of $k(n)$, we immediately see that

$$(7) \quad a_n = \left(d_{k_1} b^{k_1} + \cdots + d_{k(n)} b^{k(n)} \right) (1 + O(n^{-\beta})) = b^{k(n)} D(n) (1 + O(n^{-\beta})),$$

where $D(n) = d_{k_1} b^{k_1-k(n)} + d_{k_2} b^{k_2-k(n)} + \cdots + d_{k(n)}$.

Let $\mathcal{D}_b(x)$ be the subset of all possible values for $D(n)$, $n \in \mathcal{N}_b(x)$. Let us find an upper bound for the cardinality of this set. First observe that

$$D(n) < b^{k_1-k(n)+1} \leq b^{(\beta \log n / \log b)+1}.$$

The positive integers $D := D(n)$ bounded by the right hand side of the above inequality have at most $K := \lfloor (\beta \log x / \log b) + 2 \rfloor$ digits in base b . As $n \in \mathcal{N}_b(x)$, the number of nonzero digits of $D(n)$ is bounded by $S := \lfloor (\beta \log x / 10 \log b) \rfloor$, and

$$\begin{aligned} \#\mathcal{D}_b(x) &\leq \sum_{i=0}^S \binom{K}{i} (b-1)^i \leq (S+1) \binom{K}{S} (b-1)^S \leq (S+1) \left(\frac{(b-1)eK}{S} \right)^S \\ &\leq \left(\frac{\beta \log x}{10 \log b} + 1 \right) (10e(b-1) + o(1))^{\frac{\beta \log x}{10 \log b}} = x^{\delta+o(1)} \end{aligned}$$

as $x \rightarrow \infty$, where

$$\delta := \frac{\beta \log(10e(b-1))}{10 \log b}.$$

It can be checked that $\delta < \beta/2$ for all integers $b \geq 2$. Thus, we get that

$$(8) \quad \#\mathcal{D}_b(x) \leq x^{\delta+o(1)} \quad \text{as } x \rightarrow \infty$$

for some $\delta < \beta/2$. Combining the fact that $a_n = e^{f(n)} (1 + n^{-\alpha})$ with relations (6) and (7) we have

$$e^{f(n)} = b^{k(n)} D(n) (1 + O(x^{-\beta})),$$

since $n \in [x/2, x)$ and $\beta \leq \alpha$ by hypothesis. Taking logarithms we get that

$$(9) \quad f(n) = k(n) \log b + \log D(n) + O(x^{-\beta}).$$

We now write

$$\mathcal{N}_b(x) = \bigcup_{D \in \mathcal{D}_b(x)} \mathcal{N}_{b,D}(x),$$

where

$$\mathcal{N}_{b,D}(x) := \{n \in \mathcal{N}_b(x) : D(n) = D\}.$$

Observe that, under this notation, we have

$$\#\mathcal{N}_b(x) \leq \#\mathcal{D}_b(x) \max_{D \in \mathcal{D}_b} \#\mathcal{N}_{b,D}(x)$$

and we must now bound the number of elements lying on each $\mathcal{N}_{b,D}(x)$.

For a fixed $D \in \mathcal{D}_b(x)$ and y depending on x , to be chosen later, we take a look at the elements $n \in \mathcal{N}_{b,D}(x)$. We say that n is *separated* if $[n, n+y] \cap \mathcal{N}_{b,D}(x) = \{n\}$. It is clear that there are at most $x/2y + 1$ elements on $\mathcal{N}_{b,D}(x)$ which are separated.

Let us now count the non-separated elements $n \in \mathcal{N}_{b,D}(x)$. For such an n , there exists $1 \leq m \leq y$ with $n+m \in \mathcal{N}_{b,D}(x)$. Taking the difference of the relations (9) in n , $n+m \in \mathcal{N}_{b,D}(x)$ we get

$$\begin{aligned} (k(n+m) - k(n)) \log b &= (f(n+m) - f(n)) + O(x^{-\beta}) \\ &= m f'(\zeta) + O(x^{-\beta}), \end{aligned}$$

where $\zeta \in [n, n+m]$ is some point whose existence is guaranteed by the Intermediary Value Theorem. It follows from condition (5), which in particular implies $f'(x) \asymp \log x$, that $k(n+m) \neq k(n)$ for large x (as $x/2 < n < x$) in the above estimate. Thus, non-separated elements n in $\mathcal{N}_{b,D}(x)$ are characterized by their values $k(n)$. Denoting by $[x]$ the closest integer to x , for a fixed $m \leq y$, the differences

$$(10) \quad k(m+n) - k(n) = \left\lfloor \frac{m f'(\zeta)}{\log b} \right\rfloor$$

take $O(m)$ integer values, since for two elements $n, n+l \in \mathcal{N}_{b,D}(x)$ we have by condition (5)

$$\frac{m}{\log b} (f'(\zeta_{n+l}) - f'(\zeta_n)) \asymp \frac{ml}{x \log b} = O(m).$$

For a fixed difference in (10), say M , we must be able to count the number elements $n \in \mathcal{N}_{b,D}(x)$ such that

$$k(n+m) - k(n) = M + O(n^{-\beta}),$$

but it follows from the previous argument that

$$\frac{m}{\log b} (f'(\zeta_{n+l}) - f'(\zeta_n)) = O(x^{-\beta})$$

for at most $O(1+x^{1-\beta}/m)$ values of n . Thus, there are $O(y^2 + yx^{1-\beta})$ non-separated elements in $\mathcal{N}_{b,D}(x)$, for an arbitrary $D \in \mathcal{D}_b(x)$. Setting $y := x^{\beta/2}$, we observe that

$$\#\mathcal{N}_{b,D}(x) \ll yx^{1-\beta} + y^2 + 1 \ll x^{1-\beta/2} + x^\beta \ll x^{1-\beta/2},$$

whenever $\beta \leq 2/3$. Thus, if we choose $\beta := \min\{\alpha, 2/3\}$ it follows from estimate (8) that

$$\#\mathcal{N}_b(x) = \sum_{D \in \mathcal{D}_b(x)} \#\mathcal{N}_{b,D}(x) \leq x^{1-\beta/2} \#\mathcal{D}_b(x) < x^{1-\beta/2+\delta+o(1)} = o(x)$$

as $x \rightarrow \infty$, which is what we wanted to prove.

3. THE ANALYSIS IN BELL NUMBERS

The study of Bell numbers needs of a more detailed analysis. We start with the following estimate for $B(n)$ (see formula (41) on page 562 in [3]).

Lemma 2. *Let $r := r(n)$, defined implicitly by*

$$(11) \quad r e^r = n + 1.$$

Then

$$(12) \quad B(n) = \frac{n! e^{e^r - 1}}{r^n \sqrt{2\pi r(r+1)} e^r} \left(1 + O\left(e^{-r/5}\right)\right).$$

The number $r := r(n)$ given in (11) satisfies $r = \log n - \log \log n + o(1)$ as $n \rightarrow \infty$, therefore

$$(13) \quad e^{-r/5} = \left(\frac{\log n}{n}\right)^{1/5} (1 + o(1)) = O\left(n^{-1/6}\right) \quad \text{as } n \rightarrow \infty.$$

Combining Stirling's approximation formula (4) with formula (13) we can rewrite (12) as

$$B(n) = e^{f(n)} \left(1 + O\left(n^{-1/6}\right)\right),$$

where

$$f(n) := n \log n - n - \left(\frac{2n+1}{2}\right) \log r + \frac{1}{2} \log n + e^r - \frac{r}{2} - \frac{1}{2} \log(r+1) - 1.$$

From now on, we will assume that $r := r(x)$ is defined for all real numbers $x \geq 1$ by equation (11), where n is replaced by x . In particular, $r(x)$ has a derivative for real $x > 1$. In fact differentiating relation (11) (with n instead of x) with respect to the variable x , we have

$$r'(x) e^{r(x)} + r(x) r'(x) e^{r(x)} = 1,$$

or equivalently

$$(14) \quad r'(x) e^{r(x)} = \frac{1}{r(x) + 1},$$

and, since $e^{r(x)} = r(x)/(x+1)$,

$$(15) \quad r'(x) = \frac{r(x)}{(x+1)(r(x)+1)}.$$

We need to know the asymptotic behavior of the derivative of $f(x)$. Observe that differentiating by parts we have

$$\begin{aligned} f'(x) &= \frac{d}{dx} \left(x \log x - x - \frac{2x+1}{2} \log r(x) + \frac{1}{2} \log x + e^{r(x)} - \frac{r(x)}{2} - \frac{1}{2} \log(r(x)+1) - 1 \right) \\ &= \log x - \log r(x) - \frac{(2x+1)r'(x)}{2r(x)} + \frac{1}{2x} + r'(x) e^{r(x)} - \frac{r'(x)}{2} - \frac{r'(x)}{2(r(x)+1)} \\ &= \log x - \log r(x) + \frac{1}{2x} - \frac{r(x)^2 + r(x) - 1}{2(x+1)(r(x)+1)^2}, \end{aligned}$$

since, in terms of equations (14) and (15), we note that

$$\begin{aligned} r' e^r - r' \left(\frac{(2x+1)(r+1) + r(r+1) + r}{2r(r+1)} \right) &= \frac{1}{r+1} - \frac{(2x+1)(r+1) + r(r+1) + r}{2(r+1)^2(x+1)} \\ &= -\frac{r^2 + r - 1}{2(x+1)(r+1)^2} = O(x^{-1}). \end{aligned}$$

Finally, it follows from the definition of $r(x)$, that

$$\log\left(\frac{x}{r(x)}\right) = \log\left(\frac{x+1}{r(x)}\right) + \log\left(1 - \frac{1}{(x+1)r(x)}\right) = r(x) + O(x^{-1}),$$

therefore we can conclude that

$$f'(x) = r(x) + O(x^{-1}) \asymp \log x \quad \text{and} \quad f''(x) \asymp \frac{1}{x},$$

and we are under the assumptions of Theorem 1.

Acknowledgements. We thank the referees for their comments and suggestions that definitely improved the quality of the paper. F. L. thanks Professor Arnold Knopfmacher for useful suggestions. This paper was written while F. L. was in sabbatical from the Mathematical Institute UNAM from January 1 to June 30, 2011 and supported by a PASPA fellowship from DGAPA. J. R. is supported by a JAE-DOC grant from the *Junta para la Ampliación de Estudios* (CSIC), jointly financed by the FSE. A. Z. is by a FPU grant from Ministerio de Educación, Ciencia y Deporte, Spain. J. R., A. Z. and J. C. are also supported by the MTM2011-22851 grant, Spain. The authors thank the ICMAT Severo Ochoa project SEV-2011-0087.

REFERENCES

- [1] Bender E., Gao Z., Asymptotic enumeration of labelled graphs with a given genus. *Electr. J. Comb.* 2011, 18(1).
- [2] Evertse J.-H., Schlickewei H.P., Schmidt W.M., Linear equations in variables which lie in a multiplicative group, *Ann. Math.* 155 2002, 1-30.
- [3] Flajolet P., Sedgewick R., *Analytic Combinatorics*, Cambridge Univ. Press, Cambridge, 2009
- [4] Fusy É., Giménez O., Mohar B., Noy M., Asymptotic enumeration and limit laws for graphs of fixed genus, *J. of Comb. Theory, Series A*, 2011, 118(3), 748–777.
- [5] Giménez O., Noy M., Asymptotic enumeration and limit laws of planar graphs, *J. Amer. Math. Soc.* 2009, 22, 309–329.
- [6] Giménez O., Noy M., Rué J., Graph classes with given 3-connected components: asymptotic enumeration and random graphs, to appear at *Random Structures and Algorithms*.
- [7] Knopfmacher A., Luca F., Digit sums of binomial sums, *J. of Number Theory*, 2012, 132, 324–331.
- [8] Luca F., Distinct digits in base b expansions of linear recurrence sequences, *Quaest. Math.*, 2000, 23, 389–404.
- [9] Luca F., The number of nonzero digits of $n!$, *Canad. Math. Bull.*, 2002, 45, 115–118.
- [10] **Luca F., The number of nonzero digits of the partition function, preprint available at...WHERE??**
- [11] Luca F., Shparlinski I. E., On the g -ary expansions of middle binomial coefficients and Catalan numbers, *Rocky Mountain J. Math.*, 2011, 41, 1291–1301.
- [12] Luca F., Shparlinski I. E., On the g -ary expansions of Apéry, Motzkin and Schröder numbers, *Ann. Comb.*, 2010, 14, 507–524.
- [13] Stewart C. L., On the representation of an integer in two different bases, *J. Reine Angew. Math.*, 1980, 19, 63–72.

J. CILLERUELO: INSTITUTO DE CIENCIAS MATEMÁTICAS (CSIC-UAM-UC3M-UCM) AND DEPARTAMENTO DE MATEMÁTICAS UNIVERSIDAD, AUTÓNOMA DE MADRID, 28049 MADRID, SPAIN

E-mail address: `franciscojavier.cilleruelo@uam.es`

F. LUCA: INSTITUTO DE MATEMÁTICAS, UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO, 58089 MORELIA, MICHOACÁN, MEXICO

E-mail address: `fluca@matmor.unam.mx`

J. RUÉ: INSTITUTO DE CIENCIAS MATEMÁTICAS (CSIC-UAM-UC3M-UCM), 28049 MADRID, SPAIN

E-mail address: `juanjo.rue@icmat.es`

A. ZUMALACÁRREGUI: INSTITUTO DE CIENCIAS MATEMÁTICAS (CSIC-UAM-UC3M-UCM) AND DEPARTAMENTO DE MATEMÁTICAS UNIVERSIDAD AUTÓNOMA DE MADRID, 28049 MADRID, SPAIN

E-mail address: `ana.zumalacarregui@uam.es`