

ON THE DIGITS OF THE SUM OF PROPER DIVISORS

KÜBRA BENLİ, CÉCILE DARTYGE, CHARLOTTE DOMBROWSKY, PAUL POLLACK,
AND LOLA THOMPSON

ABSTRACT. We study several probabilistic questions concerning the digits of $s(n)$, the sum of proper divisors of an integer n . In particular, we show that $s(n)$ obeys Benford’s law with respect to logarithmic density. Moreover, we show that, for every function $k(x) \rightarrow \infty$, almost all integers $n \leq x$ have every decimal digit occurring among the first $k(x)$ digits and the last $k(x)$ digits of $s(n)$.

We also present an upper bound for the number of composite integers n up to x for which $s(n)$ is missing at least one digit in its decimal expansion. This is in contrast with the main result of a recent paper of Benli, Cesana, Dartyge, Dombrowsky, and Thompson, in which the inputs n were not required to be composite. It turns out that the primes make a substantial contribution to the preimage set $s^{-1}(\mathcal{A})$, where $\mathcal{A}$ is a set of integers with missing digits. Our result for composite n shows that the count is much smaller when prime inputs are excluded.

1. INTRODUCTION

Let $s(n)$ denote the sum of proper divisors of a positive integer n . The function $s(n)$ has a long and rich history. Pythagoras first observed that $s(6) = 6$ and $s(28) = 28$. He called such numbers *perfect*. The study of perfect numbers, that is, integers n for which $s(n) = n$, has played an important role in number theory. It has even been remarked [13] that perhaps $s(n)$ is the “first function” to ever have been studied in mathematics.

The goal of this paper is to study the digits of $s(n)$ from a probabilistic standpoint. Observe that, if $n = 2^{5865}$, then $s(n) = 2^{5865} - 1$. The last 10 decimal digits of $s(n)$ are 5046879231, so every decimal digit appears. It is natural to wonder how common this phenomenon is. We show that it is actually very common, provided that we allow the number of “last” decimal digits to be sufficiently large. We also show that the “first” decimal digits have the same property. In fact, we establish these results for an arbitrary base $g \geq 2$.

Theorem 1.1. *Fix $g \geq 2$. If $k(x) \rightarrow \infty$ as $x \rightarrow \infty$, then asymptotically 100% of integers $n \leq x$ are such that $s(n)$ has all g digits in base g appearing among its last (i.e., the rightmost) $k(x)$ digits. Moreover, the same result holds for the first (i.e., the leftmost) $k(x)$ digits.*

This suggests that the base g expansion of $s(n)$ behaves in many respects like that of a random integer of comparable size.

Theorem 1.1 concerns the occurrence of digits in prescribed positions of the decimal expansion of $s(n)$. A complementary question is to determine how the leading digits are distributed. The second goal of this paper is to show that the digits of $s(n)$ obey Benford’s law with respect to logarithmic density.

Fix a base $g \geq 2$. We say that “Benford’s law” (in base g) holds for a sequence of positive real numbers if, for each positive integer D , we have D appearing as a block of leading digits of the terms of the sequence with limiting frequency $P(D) := \log_g(D+1) - \log_g(D) = \log_g(1 + \frac{1}{D})$. For example, if a sequence obeys Benford’s law in base 10, then 1 appears as the leading digit with limiting frequency $P(1) = \log_{10}(2) = 30.1\%$, 7 appears with frequency $P(7) = \log_{10}(8/7) = 5.1\%$, and the leading digits 2026 appear with frequency $\log_{10}(2027/2026) = 0.021\%$. This version of Benford’s law is called “strong

Benford” by Diaconis [3], but we will simply use the term “Benford” here, as it is the only version of Benford’s law to be discussed.

Theorem 1.2. *Fix $g \geq 2$. The function $s(n)$ satisfies Benford’s law with respect to logarithmic density. In other words, for each positive integer D ,*

$$\delta_{\log}\{n > 1 : \text{the leftmost base-}g \text{ digits of } s(n) \text{ are the digits of } D\} = \log_g \left(1 + \frac{1}{D}\right).$$

As we show at the end of Section 3, $s(n)$ does *not* satisfy Benford’s law with respect to natural density.

In Theorem 1.1, we observe that $s(n)$ usually contains all possible digits. Both Theorems 1.1 and 1.2 suggest that the digits of $s(n)$ exhibit a high degree of randomness. We now turn our attention to the values of $s(n)$ whose base- g expansions are unusually restrictive, namely those missing one or more digits. Understanding how frequently such values occur leads us to study the preimages of sparse sets under s .

One surprising fact about $s(n)$ is that it can map sets with asymptotic density zero to sets with positive asymptotic density. The preimages of $s(n)$ can also display surprising behavior. For example, it is possible for a set $\mathcal{A}$ to have positive asymptotic density while $s^{-1}(\mathcal{A})$ has asymptotic density 0. Erdős even showed that there are sets $\mathcal{A}$ with positive asymptotic density for which $s^{-1}(\mathcal{A})$ is empty!

Some behavior of $s(n)$ is still not well-understood. There is a conjecture of Erdős, Granville, Pomerance, and Spiro [6] from 1992 which can be stated as follows:

Conjecture 1.3. *Let $\mathcal{A}$ be a set of integers with asymptotic density zero. Then $s^{-1}(\mathcal{A})$ also has asymptotic density zero.*

We shall refer to this conjecture as the “EGPS Conjecture” throughout this paper. The conjecture remains open, although it has been proven in some special cases. For example, if $\mathcal{A}$ is the set of prime numbers [11], or the set of palindromes in any given base [10], or the set of sums of two squares [17], it has been shown that the EGPS Conjecture holds. In all of these cases, the elements in the set have very specific structural features. It has also been shown in [12] that, without imposing any special structure on the elements of the set, the EGPS Conjecture holds for sets $\mathcal{A}$ with size $O(x^{1/2+\varepsilon})$ where ε is a fixed function tending to 0 as $x \rightarrow \infty$.

In a previous work, Benli, Cesana, Dartyge, Dombrowsky, and Thompson confirmed that the EGPS conjecture holds for sets of integers with missing digits. In particular, they showed:

Theorem 1.4. [1, Theorem 1.8] *Fix $g \geq 2$, $\gamma \in (0, 1)$, and a nonempty set $\mathcal{D} \subsetneq \{0, 1, \dots, g-1\}$. For all sufficiently large x , the number of $n \leq x$ for which $s(n)$ has all of its digits in base g restricted to digits in $\mathcal{D}$ is $O(x \exp(-(\log \log x)^\gamma))$.*

The case $g = 2$, omitted from the statement in [1], also follows by a separate elementary argument. See Appendix A.

Observe that $s(p) = 1$ for all primes p , which means that whenever the set $\mathcal{D}$ contains 1, the size of the preimage set of $\mathcal{A}$ has $\pi(x) \sim x/\log x$ as a lower bound. Thus, the upper bound given by Theorem 1.4 is essentially “best possible” because we cannot replace the constant $\gamma \in (0, 1)$ with any constant strictly greater than 1.

In the present paper, we explore what happens to this upper bound when we exclude the contribution from prime inputs. At first glance, one might expect the prime inputs to play only a negligible role in questions concerning the digits of $s(n)$. Surprisingly, this is not the case. When we consider only composite values of n , we show that the count is considerably smaller:

Theorem 1.5. *Let $g \in \mathbb{N}$, $g \geq 2$ and $a_0 \in \{1, \dots, g-1\}$. There exists a constant $c = c(g) > 0$, depending on g such that*

$$\#\{n \leq x : n \text{ composite, } s(n) \text{ has no } a_0 \text{ in its base-}g \text{ expansion}\} \ll x \exp(-c\sqrt{\log x}).$$

1.1. Notation. We record some notation that will be used throughout the paper. For $n \in \mathbb{N}$, $n \geq 2$, let $P^+(n)$ denote the largest prime factor of n and $P^-(n)$ denote the smallest prime factor of n .

Fix $g \in \mathbb{N}$. Let $\mathcal{D} \subsetneq \{0, \dots, g-1\}$. Then the set of integers with missing digits is given by:

$$\mathcal{W}_{\mathcal{D}} := \left\{ n \in \mathbb{N} : n = \sum_{j=0}^J \varepsilon_j(n) g^j, J \geq 0, \varepsilon_j(n) \in \mathcal{D}, \varepsilon_J \neq 0 \right\}.$$

We denote by $\mathcal{W}_{\mathcal{D}}(g^k)$ the subset of $\mathcal{W}_{\mathcal{D}}$ of integers up to g^k :

$$\mathcal{W}_{\mathcal{D}}(g^k) := \{n \in \mathcal{W}_{\mathcal{D}} : n \leq g^k\}.$$

In Section 3, we will state Benford's law in terms of logarithmic density, which we define as follows:

Definition 1.6. For $A \subseteq \mathbb{N}$, we define the logarithmic density of A as follows:

$$\delta_{\log}(A) = \lim_{x \rightarrow \infty} \frac{1}{\log x} \sum_{\substack{n \leq x \\ n \in A}} \frac{1}{n},$$

provided that the limit exists.

2. DIGITAL HITS

The goal of this section is to prove Theorem 1.1, which has been separated into two parts. The first part is stated below as Theorem 2.2, and the second part is stated as Theorem 2.5.

We will require the following lemma in our first proof:

Lemma 2.1. [11, Lemma 2.1] *Let $x \geq 3$. Let q be a positive integer. Then*

$$\sum_{\substack{n \leq x \\ q \nmid \sigma(n)}} 1 \ll \frac{x}{(\log x)^{1/\varphi(q)}},$$

uniformly in q where φ is Euler's totient function.

Now we are able to prove Theorem 2.2.

Theorem 2.2. *If $k(x) \rightarrow \infty$ as $x \rightarrow \infty$, then asymptotically 100% of integers $n \leq x$ are such that $s(n)$ has all g digits in base g appearing among its last (i.e., the rightmost) $k(x)$ digits.*

Proof. We first note that the case when $g = 2$ is handled in Appendix A, as the statement of the theorem boils down to showing that the number of $n \leq x$ for which all digits of $s(n)$ are 1 is $o(x)$ in this case. Let $g \geq 3$. Let x be large, assume $k = k(x) \rightarrow \infty$ as $x \rightarrow \infty$. Let $\gamma < 1$. Note that $k(x) \leq \frac{\log x}{\log g}$. Without loss of generality, we can assume $k \leq (\log \log \log x)^\gamma$; otherwise, we can simply replace $k(x)$ with $\min\{k(x), (\log \log \log x)^\gamma\}$. Then we would like to show that for any subset $\mathcal{D} \subsetneq \{0, \dots, g-1\}$,

$$\#A := \#\{n \leq x : s(n) \equiv B \pmod{g^k}, \text{ for some } B \in \mathcal{W}_{\mathcal{D}}(g^k - 1)\} = o(x).$$

Following the proof of Theorem 1.8 in [1], for $n \in A$, if $g^k \mid \sigma(n)$, we have

$$n = \sigma(n) - s(n) \equiv -s(n) \equiv -B \pmod{g^k}$$

so that

$$\sum_{\substack{n \in A \\ \sigma(n) \equiv 0 \pmod{g^k}}} 1 \leq \sum_{\substack{n \leq x \\ n \equiv -B \pmod{g^k} \\ B \in \mathcal{W}_{\mathcal{D}}(g^k - 1)}} 1 = \sum_{B \in \mathcal{W}_{\mathcal{D}}(g^k - 1)} \sum_{\substack{n \leq x \\ n \equiv -B \pmod{g^k}}} 1$$

$$\begin{aligned} &\leq |\mathcal{D}|^k \left(\frac{x}{g^k} \right) + |\mathcal{D}|^k \\ &\ll x \exp(-k \log(g/|\mathcal{D}|)) + |\mathcal{D}|^k. \end{aligned}$$

The $|\mathcal{D}|^k$ does not pose a problem since $|\mathcal{D}|^k \leq (g-1)^k \leq \exp(\log(g-1)(\log \log \log x)^\gamma) = (\log \log x)^{o(1)} = o(x)$. Consequently, the last expression in the displayed equation is $o(x)$ if $k(x) \rightarrow \infty$ as $x \rightarrow \infty$ and since $k \ll (\log \log \log x)^\gamma$.

To complete the proof, we need to show that the number of $n \leq x$ with $g^k \nmid \sigma(n)$ is $o(x)$. To accomplish this, we appeal to Lemma 2.1, taking $q = g^k$. Since $k \leq (\log \log \log x)^\gamma$, we then have the number of $n \leq x$ with $g^k \nmid \sigma(n)$ is $\ll x \exp\left(-\frac{\log \log x}{\varphi(g^k)}\right) \leq x \exp\left(-\frac{\log \log x}{g^k}\right) \leq x \exp\left(-\frac{\log \log x}{\exp((\log g)(\log \log \log x)^\gamma)}\right) = o(x)$. $\square$

Again fix $g \geq 2$. We now present the proof that the *first* $k(x)$ base g digits of $s(n)$ include all g digits 100% of the time. We first collect some auxiliary results.

Let $M = \prod_{p \leq z} p^z$, where $z = g^{10k}$. By the Prime Number Theorem, $M \leq 3^{z^2}$. Let $n' \in (0, M]$ be the least positive representative of $n \pmod{M}$ and n_z the z -smooth (or z -friable) part of n' , that is $n_z = \prod_{p^e \parallel n'} p^e$. We have the following.

Lemma 2.3. *Let x be large and $k \in \mathbb{N}$ with $k(x) \rightarrow \infty$ as $x \rightarrow \infty$. Put $z = g^{10k}$. For all but $o(x)$ values of $n \leq x$, one has $0 \leq s(n)/n - s(n_z)/n_z < z^{-1/2}$ and $s(n_z)/n_z \geq 1/k$.*

Proof. First note that if n_z is divisible by p^z for some prime $p \leq z$, then so is n' and since $n \equiv n' \pmod{M}$, so is n . Thus, the number of $n \leq x$ such that n_z is divisible by p^z for some prime $p \leq z$ is at most $x \sum_{p \leq z} 1/p^z \ll x/z^2 = o(x)$.

Thus, we can assume n_z is not divisible by p^z for any prime $p \leq z$. It follows that n_z , the z -smooth part of n' , is also the z -smooth part of n . So, for such n and n_z , we have

$$0 \leq \frac{s(n)}{n} - \frac{s(n_z)}{n_z} = \sum_{\substack{d|n \\ P^+(d) > z}} \frac{1}{d} \leq \sum_{\substack{d|n \\ d > z}} \frac{1}{d}.$$

Averaging over positive integers $l \leq x$ allows us to obtain

$$\sum_{l \leq x} \sum_{\substack{d|l \\ d > z}} \frac{1}{d} \ll \frac{x}{z}.$$

Applying Markov's inequality yields

$$\#\left\{n \leq x : \frac{s(n)}{n} - \frac{s(n_z)}{n_z} \geq z^{-1/2}\right\} \ll z^{1/2} x/z = o(x).$$

Now, if $p \mid n_z$, then $\frac{s(n_z)}{n_z} = \sum_{\substack{d|n_z \\ d > 1}} 1/d \geq 1/p$. So, $\frac{s(n_z)}{n_z} \geq 1/k$ unless n_z is k -rough. But then n is also k -rough, and the number of such $n \leq x$ is $\ll x/\log k = o(x)$. $\square$

Remark. We briefly remark on certain exceptional cases that we can safely ignore in our proof of Theorem 2.5 below, since they contribute only $o(x)$ to the total count. As always, we assume that $k(x) \rightarrow \infty$ as $x \rightarrow \infty$. Then

- If n has r decimal digits, we can assume that $g^r > x/g^{\log \log k}$. Otherwise, if $g^r \leq x/g^{\log \log k}$, then $n \leq g^r \leq x/g^{\log \log k}$. The number of these n is $o(x)$.

- If $s(n)$ has m digits base g then we can assume that $|m - r| < \log \log k$. This will follow from Lemma 2.4, which we prove below.

Lemma 2.4. *Let $g \geq 2$, $g \in \mathbb{N}$. The number of $n \leq x$ with the number m of base g digits of $s(n)$, and the number r of base g digits of n , satisfies $|m - r| \geq \log \log k$ is $o(x)$.*

Proof. If $n \leq x$ has r digits, then $g^{r-1} \leq n < g^r$. And if $s(n)$ has m digits then $g^{m-1} \leq s(n) < g^m$. So, $g^{m-r-1} < \frac{s(n)}{n} < g^{m-r+1}$. If $|m - r| \geq \log \log k$, either $m - r \geq \log \log k$ or $r - m \geq \log \log k$. In the case $m - r \geq \log \log k$, we have $\frac{s(n)}{n} > \frac{1}{g}(\log k)^{\log g}$. On the other hand, if $r - m \geq \log \log k$, then $\frac{s(n)}{n} < \frac{g}{(\log k)^{\log g}}$.

Davenport [4] has shown that for each fixed $u \geq 0$, the density of n with $s(n) \leq un$ exists. Moreover, calling this function $D(u)$, we have that $D(u)$ is continuous, that $D(u) \rightarrow 0$ as $u \rightarrow 0$, and that $D(u) \rightarrow 1$ as $u \rightarrow \infty$. For any fixed $M > 0$, we have that $(\log k)^{\log g}/g$ eventually exceeds M while $g/(\log k)^{\log g}$ is eventually smaller than $1/M$. Hence,

$$\limsup_{x \rightarrow \infty} \frac{1}{x} \left(\# \left\{ n \leq x : \frac{s(n)}{n} > \frac{1}{g}(\log k)^{\log g} \right\} + \# \left\{ n \leq x : \frac{s(n)}{n} < \frac{g}{(\log k)^{\log g}} \right\} \right) \leq (1 - D(M)) + D(1/M).$$

The lemma follows upon sending M to infinity. $\square$

We are now ready to prove the second half of Theorem 1.1.

Theorem 2.5. *Assume $k(x) \rightarrow \infty$ as $x \rightarrow \infty$. Then the number of $n \leq x$ where, in the base g expansion of $s(n)$, not all digits are represented among the first k digits is $o(x)$. In other words, asymptotically 100% of integers $n \leq x$ are such that $s(n)$ has all g digits in base g appearing among its first (i.e., leftmost) $k(x)$ digits.*

Proof. We can assume our function $k = k(x) \rightarrow \infty$ grows very slowly, e.g., $k(x) \leq \log \log \log x$ (otherwise, we can simply replace $k(x)$ with $\min\{k(x), \log \log \log x\}$).

Let $n \leq x$ be such that in the base g expansion of $s(n)$, not all digits are represented among the first k digits. Let $r \in \mathbb{N}$ be such that

$$g^{r-1} \leq n < g^r.$$

By our first remark above, we may assume that

$$g^r > \frac{x}{g^{\log \log k}}.$$

Let $m \in \mathbb{N}$ be such that

$$g^{m-1} \leq s(n) < g^m.$$

By our second remark above, we may assume that

$$|m - r| < \log \log k.$$

Write $s(n) = g^{m-k}L + T$, where $L, T \in \mathbb{N}$ with $0 \leq T < g^{m-k}$. Then $L < g^k$, and the digits of L are the k leading digits of $s(n)$. Note that the number of possibilities for L is $\ll g^{ck}$, for a constant $c < 1$ depending only on g (because L is missing at least one base g digit). This puts $s(n)$ in the interval

$$\begin{aligned} I_L &:= [L \cdot g^{m-k}, (L+1) \cdot g^{m-k}) \\ &:= [t_L, T_L). \end{aligned}$$

We fix r , m , and L , and count the number of corresponding n . It is convenient to sort these n into residue classes modulo $M = \prod_{p \leq z} p^z$, where $z = g^{10k}$. By the Prime Number Theorem, $M \leq 3^{z^2} = x^{o(1)}$,

where the last equality comes from our restriction on the size of k . Let $n' \in (0, M]$ be the least positive representative of $n \pmod{M}$ and n_z be the z -smooth part of n' . Excluding the $o(x)$ exceptional values of n described in Lemma 2.3 above, we have

$$\frac{s(n)}{n} \leq \frac{s(n_z)}{n_z} \left(1 + \frac{n_z}{s(n_z)} z^{-1/2}\right) \leq s(n_z)(1 + kz^{-1/2}).$$

We now fix also $n' \in \mathbb{N} \cap (0, M]$ and consider only nonexceptional integers $n \equiv n' \pmod{M}$ corresponding to our choices of r, m , and L from above.

If $s(n) = n \cdot s(n)/n$ is in one of the intervals I_L from above, then n is in $(n/s(n))I_L$. Notice that $s(n)/n \geq s(n_z)/n_z$, and so $n/s(n) \leq n_z/s(n_z)$. Since T_L is the upper endpoint of I_L , the dilated interval $(n/s(n))I_L$ has upper endpoint at most

$$T_{L'} := (n_z/s(n_z))T_L.$$

Now, we consider the lower endpoint. We have

$$\frac{s(n)}{n} \leq \frac{s(n_z)}{n_z} + z^{-1/2} = \frac{s(n_z)}{n_z} \left(1 + z^{-1/2} \frac{n_z}{s(n_z)}\right) \leq \frac{s(n_z)}{n_z} (1 + z^{-1/3})$$

since $z = g^{10k}$. Hence,

$$\frac{n}{s(n)} \geq \frac{n_z}{s(n_z)} (1 - z^{-1/3}).$$

Thus, if t_L is the lower endpoint of I_L , then $(n/s(n))I_L$ has lower endpoint at least

$$t_{L'} := n_z/s(n_z)(1 - z^{-1/3})t_L.$$

Every nonexceptional $n \equiv n' \pmod{M}$ corresponding to our choices of r, m , and L belongs to the interval $[t_{L'}, T_{L'}]$. The length of this interval is

$$T_{L'} - t_{L'} = \frac{n_z}{s(n_z)}T_L - \frac{n_z}{s(n_z)}(1 - z^{-1/3})t_L = \frac{n_z}{s(n_z)}|I_L| + z^{-1/3} \frac{n_z}{s(n_z)}t_L,$$

and so

$$\begin{aligned} \#\{n \leq x : n \equiv n' \pmod{M}, t_{L'} \leq n < T_{L'}\} &\leq \frac{1}{M} \frac{n_z}{s(n_z)} \left(|I_L| + z^{-1/3}t_L\right) + 1 \\ &\leq 1 + \frac{kg^{m-k}(1 + z^{-1/3}L)}{M}. \end{aligned}$$

We bound the total number of nonexceptional n by summing over parameters previously held fixed. First, we sum over the M possible values of $n' \in (0, M]$. Recalling that $z = g^{10k}$ and $L \ll g^{ck}$ for $c < 1$, the number of n arising, for a fixed choice of r, m , and L , is at most

$$M + 2k \cdot g^{m-k} \leq 3kg^{m-k}.$$

For the last inequality, we used that $M = x^{o(1)}$ while (for large x)

$$g^{m-k} = g^r g^{m-r} g^{-k} \gg \frac{x}{g^{\log \log k}} \cdot \frac{1}{g^{\log \log k}} \cdot \frac{1}{g^k} > \frac{x}{g^{2k}} \geq \frac{x}{(\log \log x)^{O(1)}}.$$

Next, we sum on the possibilities for L ; this gives an upper bound of the form $3k \cdot g^{-c'k} g^m$, for a positive $c' > 0$, bounded away from 0. Furthermore, $m \leq r + \log \log k$. Summing $3k \cdot g^{-c'k} g^m$ on these m , we obtain a quantity bounded above by $g^r g^{-c''k}$, for another small $c'' > 0$. Summing on the possibilities for r gives an upper bound of $xg^{-c'''k}$ for $c''' > 0$. If we combine this upper bound with the previously mentioned $o(x)$ contributions from the exceptions, we still obtain $o(x)$ for our final count.

□

3. BENFORD'S LAW

It was shown by Diaconis [3, Theorem 1] that a sequence $\{a_n\}$ of positive integers obeys Benford's law, in base 10, precisely when $\{\log_{10} a_n\}$ is uniformly distributed modulo 1. Diaconis's result is about decimal expansions and natural density, but the argument works just as well in any base g and for logarithmic density.

Proposition 3.1. *A sequence $a_n > 0$ is base g Benford in the sense of logarithmic density if and only if $\{a_n\}$ is uniformly distributed mod 1, in the sense of logarithmic density. That is, for every interval $I \subseteq [0, 1)$ we have $\delta_{\log}(\{n : \{\log_g a_n\} \in I\}) = |I|$.*

The following criterion (see, for example, [7, p. 209]) will allow us to decide uniform distribution.

Theorem 3.2 (Weyl's criterion for logarithmic densities). *The sequence $\{a_n\}$ is uniformly distributed modulo 1 with respect to logarithmic density if and only if*

$$\lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{n \leq N} \frac{e^{2\pi i \ell a_n}}{n} = 0$$

for all integers $\ell \neq 0$.

We define

$$\{a_k\}_{k \geq 2} := \{\log_g s(k)\}_{k \geq 2}.$$

By Theorem 3.2, in order for $s(n)$ to obey Benford's law, it is enough if $e^{2\pi i \ell \log s(n)/\log g}$ has logarithmic mean 0 (over integers $n \geq 2$) for every fixed integer $\ell \neq 0$.

That certainly holds if $s(n)^{i\alpha}$ has logarithmic mean 0 (over integers $n \geq 2$) for each fixed nonzero real number α . We prove this by appealing to the following weighted version of Halász's Theorem:

Proposition 3.3 (Proposition 1.2.6, [8]). *If f is a multiplicative function with $|f(n)| \leq 1$ for all $n \in \mathbb{N}$, then*

$$\frac{1}{\log x} \left| \sum_{n \leq x} \frac{f(n)}{n} \right| \leq \exp \left(-\frac{1}{2} \sum_{\substack{q \leq x \\ q \text{ prime}}} \frac{1 - \Re(f(q))}{q} \right).$$

Proposition 3.3 tells us that, if f is a multiplicative function of modulus at most 1, then f has logarithmic mean value 0 unless “ f pretends to be 1.” This will come in handy when we prove our main result below.

We will need the following definition:

Definition 3.4. *We say that p^e is the special prime divisor of n if it is the prime power p^e for which p is the least prime divisor of n and $p^e \parallel n$. We denote the special prime divisor of n by $\gamma(n)$.*

Now that we have all of the ingredients, we are ready to prove that $s(n)^{i\alpha}$ has logarithmic mean 0.

Theorem 3.5. *For any fixed nonzero $\alpha \in \mathbb{R}$, we have*

$$\lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{1 < n \leq N} \frac{s(n)^{i\alpha}}{n} = 0.$$

In particular, for any integer base $g \geq 2$, $s(n)$ obeys Benford's law in terms of logarithmic density in base g .

Proof. We cannot directly apply Proposition 3.3 to estimate the mean value of $s(n)^{i\alpha}$, since $s(n)$ is not multiplicative. To work around this, write $s(n) = \sigma(n) - n = \sigma(n)(1 - n/\sigma(n))$, and let

$$u(n) := n/\sigma(n).$$

Since $0 < u(n) < 1$ for all $n > 1$, Newton's binomial theorem gives us that

$$\begin{aligned} s(n)^{i\alpha} &= \sigma(n)^{i\alpha} (1 - u(n))^{i\alpha} \\ &= \sum_{k \geq 0} \binom{i\alpha}{k} \sigma(n)^{i\alpha} (-u(n))^k. \end{aligned}$$

If the sum on k were finite, it would suffice by linearity to show that each term $\sigma(n)^{i\alpha} u(n)^k$ has logarithmic mean value 0. That mean value result will be shown below using Proposition 3.3. However, since this is an infinite sum, we must be more careful. We truncate the binomial series at a finite level, far enough out to make the error small. This would be straightforward if we had $u(n) \leq a$ for some fixed $a < 1$. However, $u(n)$ can be arbitrarily close to 1, and this necessitates partitioning our set of integers n according to their special prime divisors (as in Definition 3.4).

Each $n \in \mathbb{N}$ with special prime divisor $\gamma(n) = p^e$ has the form $n = p^e m$, where the least prime dividing m is greater than p . For each of these n , we have $\sigma(n)/n \geq \sigma(p^e)/p^e$, and so $u(n) = n/\sigma(n) \leq p^e/\sigma(p^e) < 1$. Since $u(n)$ is bounded away from 1, when we look at the series expansion

$$s(n)^{i\alpha} = \sum_{k \geq 0} \binom{i\alpha}{k} \sigma(n)^{i\alpha} (-u(n))^k,$$

we will know where to cut off the righthand side for an acceptable error.

Let B be a fixed real number, which we will later choose to be very large. We split the sum

$$T(N) := \frac{1}{\log N} \sum_{1 < n \leq N} \frac{(s(n))^{i\alpha}}{n}$$

in two sub-sums $T(N) = T_1(N) + T_2(N)$ where

$$T_1(N) := \frac{1}{\log N} \sum_{\substack{1 < n \leq N \\ \gamma(n) \leq B}} \frac{1}{n} \sum_{k \geq 0} \binom{i\alpha}{k} \sigma(n)^{i\alpha} (-u(n))^k$$

and

$$T_2(N) := \frac{1}{\log N} \sum_{\substack{1 < n \leq N \\ \gamma(n) > B}} \frac{1}{n} \sum_{k \geq 0} \binom{i\alpha}{k} \sigma(n)^{i\alpha} (-u(n))^k.$$

3.1. Bounding the sum $T_1(N)$. We define

$$\rho := \max_{p^e \leq B} \frac{p^e}{\sigma(p^e)}.$$

If n has a special prime power divisor $p^e \leq B$, then

$$u(n) = \frac{n}{\sigma(n)} \leq \rho < 1.$$

Hence, the binomial series

$$(1 - u(n))^{i\alpha} = \sum_{k \geq 0} \binom{i\alpha}{k} (-u(n))^k$$

converges uniformly on the set of n with a special prime divisor $\leq B$. We fix $K = K(B, \alpha)$ sufficiently large so that

$$\sup_{0 \leq u \leq \rho} \left| (1-u)^{i\alpha} - \sum_{k=0}^K \binom{i\alpha}{k} (-u)^k \right| < \frac{1}{B}.$$

Since $|\sigma(n)^{i\alpha}| = 1$, we then have uniformly for n with a special prime divisor $\leq B$, that

$$\left| s(n)^{i\alpha} - \sum_{k=0}^K \binom{i\alpha}{k} (-1)^k \sigma(n)^{i\alpha} u(n)^k \right| < \frac{1}{B}.$$

So, we rewrite $T_1(N)$ as

$$\begin{aligned} T_1(N) &= \frac{1}{\log N} \sum_{p^e \leq B} \sum_{\substack{1 < n \leq N \\ \gamma(n) = p^e}} \frac{1}{n} \sum_{0 \leq k \leq K} \binom{i\alpha}{k} \sigma(n)^{i\alpha} (-u(n))^k \\ &\quad + \frac{1}{\log N} \sum_{\substack{1 < n \leq N \\ \gamma(n) \leq B}} \frac{1}{n} \sum_{k > K} \binom{i\alpha}{k} \sigma(n)^{i\alpha} (-u(n))^k \end{aligned}$$

and by our choice of K ,

$$\begin{aligned} \left| \frac{1}{\log N} \sum_{\substack{1 < n \leq N \\ \gamma(n) \leq B}} \frac{1}{n} \sum_{k > K} \binom{i\alpha}{k} \sigma(n)^{i\alpha} (-u(n))^k \right| &\leq \frac{1}{\log N} \sum_{\substack{1 < n \leq N \\ \gamma(n) \leq B}} \frac{1}{n} \left| \sum_{k > K} \sigma(n)^{i\alpha} \binom{i\alpha}{k} (-u(n))^k \right| \\ &\leq \frac{1}{B \log N} \sum_{\substack{1 < n \leq N \\ \gamma(n) \leq B}} \frac{1}{n} \leq \frac{2}{B}, \end{aligned}$$

which will be acceptable to us.

It remains to handle the contribution to $T_1(N)$ from

$$\frac{1}{\log N} \sum_{p^e \leq B} \sum_{\substack{1 < n \leq N \\ \gamma(n) = p^e}} \frac{1}{n} \sum_{0 \leq k \leq K} \binom{i\alpha}{k} \sigma(n)^{i\alpha} (-u(n))^k. \quad (3.1)$$

We change the order of summation in this finite sum and attempt to compute the mean value of one of the terms $\sigma(n)^{i\alpha} (-u(n))^k$ over n with $\gamma(n) \leq B$. We have

$$\begin{aligned} \sum_{\substack{n \leq N \\ \gamma(n) \leq B}} \frac{1}{n} \sigma(n)^{i\alpha} (-u(n))^k &= \sum_{p^e \leq B} \frac{1}{p^e} \sigma(p^e)^{i\alpha} (-u(p^e))^k \sum_{\substack{n \leq N \\ \gamma(n) = p^e \\ n = p^e m}} \frac{1}{m} \mathbf{1}_{q|m, q \text{ prime} \Rightarrow q > p} \sigma(m)^{i\alpha} (u(m))^k \\ &= \sum_{p^e \leq B} \frac{1}{p^e} \sigma(p^e)^{i\alpha} (-u(p^e))^k \sum_{m \leq \frac{N}{p^e}} \frac{1}{m} \mathbf{1}_{q|m, q \text{ prime} \Rightarrow q > p} \sigma(m)^{i\alpha} (u(m))^k. \quad (3.2) \end{aligned}$$

Let $b_p(m)$ be the 1-bounded multiplicative function defined by

$$b_p(m) := \mathbf{1}_{q|m, q \text{ prime} \Rightarrow q > p} \sigma(m)^{i\alpha} (u(m))^k.$$

We will show that each of the functions $b_p(m)$ has logarithmic mean 0. Hence, for each fixed p^e , the inner sum in (3.2) is of size $o(\log N)$. Since B is fixed, the triangle inequality then gives

$$\frac{1}{\log N} \sum_{p^e \leq B} \frac{1}{p^e} \sigma(p^e)^{i\alpha} (-u(p^e))^k \sum_{m \leq \frac{N}{p^e}} \frac{b_p(m)}{m} = o(1).$$

It follows that (3.1) contributes $o(1)$. Putting this together with our earlier estimate, we find that

$$\limsup_{N \rightarrow \infty} |T_1(N)| \leq \frac{2}{B}.$$

Let us see why the b_p have logarithmic mean zero. For this we apply Proposition 3.3. Since $\sum_{q \text{ prime}} \frac{1}{q}$ diverges, it is enough to show that the partial sums of $\sum_{q \leq x, \text{ prime}} \frac{\Re(b_p(q))}{q}$ remain bounded as $x \rightarrow \infty$. In fact, we will prove the stronger result that $\sum_{q \text{ prime}} \frac{\Re(b_p(q))}{q}$ converges.

If $q \leq p$, then $b_p(q) = 0$, while if $q > p$,

$$\Re(b_p(q)) = \Re(\sigma(q)^{i\alpha} (u(q))^k) = \Re\left((q+1)^{i\alpha} \left(\frac{q}{q+1}\right)^k\right) = \left(\frac{q}{q+1}\right)^k \cos(\alpha \log(q+1)).$$

We have the following standard approximations:

$$\left| \left(\frac{q}{q+1}\right)^k - 1 \right| \leq \frac{k}{q+1} \quad \text{and} \quad |\cos(\alpha \log(q+1)) - \cos(\alpha \log q)| \leq |\alpha| |\log(q+1) - \log q| \leq \frac{|\alpha|}{q}.$$

Hence, for $q > p$,

$$\left| \Re(b_p(q)) - \frac{\cos(\alpha \log q)}{q} \right| \leq \frac{k}{q(q+1)} + \frac{|\alpha|}{q^2}.$$

The right hand side converges when summed on primes q . Thus, it suffices to show that

$$\sum_{q \text{ prime}} \frac{\cos(\alpha \log q)}{q} \text{ converges.}$$

But $\frac{\cos(\alpha \log q)}{q} = \Re \frac{1}{q^{1+i\alpha}}$, and it is well known that $\sum_{q \text{ prime}} \frac{1}{q^{1+i\alpha}}$ converges for nonzero real α (see [16, pp. 65-66], where this sum is related to the value of $\log \zeta(1+i\alpha)$).

3.2. Bounding the sum $T_2(N)$. For the n with a special prime divisor $> B$, we can start with the trivial bound

$$\left| \frac{1}{\log N} \sum_{\substack{1 < n \leq N \\ \gamma(n) > B}} \frac{s(n)^{i\alpha}}{n} \right| \leq \frac{1}{\log N} \sum_{\substack{1 < n \leq N \\ \gamma(n) > B}} \frac{1}{n} =: S.$$

Let $E = 10 \log \log B$. We split the sum S into three subsums: $S = S_1 + S_2 + S_3$, where the sum S_1 is over $p > B$, the sum S_2 is over $p \leq B$ with $e \geq E$, and the sum S_3 is over $p < B$ for which $B \leq p^e \leq p^E$.

Now, the integers contributing to S_1 are B -sifted integers. We have

$$S_1 \leq \frac{1}{\log N} \sum_{\substack{n \leq N \\ P^-(n) > B}} \frac{1}{n} \leq \frac{1}{\log N} \prod_{B < p \leq N} (1 - 1/p)^{-1} \ll \frac{1}{\log B}.$$

For S_2 , we have

$$\begin{aligned}
S_2 &\leq \sum_{p \leq B} \sum_{E \leq e \leq \frac{\log N}{\log p}} \sum_{m \leq \frac{N}{p^e}} \frac{1}{p^e m \log N} \\
&\ll \sum_{p \leq B} \sum_{E \leq e \leq \frac{\log N}{\log p}} \frac{1}{p^e} \\
&\ll \sum_{p \leq B} \frac{1}{p^E} \ll \frac{1}{2^E} \ll \frac{1}{\log B},
\end{aligned}$$

recalling that $E = 10 \log \log B$.

It remains to bound the sum S_3 , which corresponds to the prime powers $p^e > B$ with p and e not too large. Writing $n = p^e m$, we have:

$$\begin{aligned}
S_3 &= \frac{1}{\log N} \sum_{p \leq B} \sum_{\frac{\log B}{\log p} < e \leq E} \sum_{\substack{m \leq N p^{-e} \\ P^-(m) > p}} \frac{1}{p^e m} \\
&\ll \sum_{p \leq B} \sum_{\frac{\log B}{\log p} < e \leq E} \frac{1}{p^e \log p} \\
&\ll \sum_{p \leq B} \frac{1}{B \log p} \ll \frac{1}{(\log B)^2}.
\end{aligned}$$

Putting together the estimates for S_1, S_2 , and S_3 , we conclude that $\limsup_{N \rightarrow \infty} |T_2(N)| \ll 1/\log B$. Combining this with our earlier result that $\limsup |T_1(N)| \leq 2/B$, we find that

$$\limsup |T(N)| \ll \frac{1}{B} + \frac{1}{\log B}.$$

Since B can be taken arbitrarily large, $T(N) = o(1)$. This completes the proof of Theorem 3.5. $\square$

Remark. An important step in the proof of Theorem 1.2 is to apply Halász's Theorem (Proposition 3.3) to prove that the functions $g_{\alpha,k} : n \mapsto \sigma(n)^{i\alpha} (n/\sigma(n))^k$ have logarithmic mean 0 when $\alpha \in \mathbb{R} \setminus 0$. This approach does not work in case of natural means because the functions $g_{\alpha,k}$ are $n^{i\alpha}$ -pretentious in the sense of [8].

This suggests that $s(n)$ may not satisfy Benford's law with respect to natural density. In fact, we can prove this.

Proposition 3.6. *For every integer base $g \geq 2$, the sequence $s(n)$ is not Benford with respect to natural density in base g .*

Proof. Fix a base $g \geq 2$. Recall that “Benford in base g ” implies that $\{\log_g s(n)\}$ is uniformly distributed mod 1 with respect to natural density. Roughly speaking, we show that for large N the interval $I := [g^N, g^N + g^{N-100}]$ contains many integers n where $s(n)$ is slightly larger than g^N – more than that should be allowed if Benford's law holds for natural density.

Call $n \in I$ *convenient* if $n = 6m$, where m has no prime factor up to g^{300} . By inclusion-exclusion, the count of convenient n is

$$\sim \frac{g^{N-100}}{6} \prod_{p \leq g^{300}} (1 - 1/p)$$

as $N \rightarrow \infty$. Hence, for large N this count exceeds

$$\frac{g^{N-100}}{6} \cdot \frac{0.5}{\log(g^{300})} = \frac{g^{N-100}}{3600 \log g}.$$

Here, the product on p is estimated using the explicit form of Mertens' theorem appearing as in [14, eq. (3.25)].

We claim that most convenient $n \in I$ have $s(n)$ very close to n , or equivalently, $\sigma(n)$ very close to $2n$. Since $6 \mid n$, it is immediate that $\frac{s(n)}{n} \geq \frac{1}{n} \left(\frac{n}{6} + \frac{n}{3} + \frac{n}{2} \right) = 1$. To bound $\frac{s(n)}{n} = \frac{\sigma(n)}{n} - 1$ from above, we use a first moment argument:

$$\begin{aligned} \sum_{\substack{n \in I \\ n \text{ convenient}}} \left(\frac{\sigma(n)}{n} - 2 \right) &= \sum_{\substack{n \in I \\ n \text{ convenient}}} 2 \left(\frac{\sigma(n/6)}{n/6} - 1 \right) = 2 \sum_{\substack{n \in I \\ n \text{ convenient}}} \sum_{\substack{d \mid \frac{n}{6} \\ d > 1}} \frac{1}{d} \leq 2 \sum_{n \in I} \sum_{\substack{d \mid n \\ d > g^{300}}} \frac{1}{d} \\ &\leq 2 \sum_{n \leq 2g^N} \sum_{\substack{d \mid n \\ d > g^{300}}} \frac{1}{d} \leq 2 \sum_{d > g^{300}} \frac{1}{d} \sum_{\substack{n \leq 2g^N \\ d \mid n}} 1 \leq 4g^N \sum_{d > g^{300}} \frac{1}{d^2} \leq 4g^{N-300}. \end{aligned}$$

Hence, among the convenient $n \in I$, at most $4g^{N-200}$ of them can satisfy $\frac{\sigma(n)}{n} \geq 2 + g^{-100}$. Thus, the number of convenient $n \in I$ with $s(n)/n < 1 + g^{-100}$ is at least

$$\frac{g^{N-100}}{3600 \log g} - 4g^{N-200} = \frac{g^{N-100}}{3600 \log g} \left(1 - \frac{14400 \log g}{g^{100}} \right) > \frac{1}{2} \frac{g^{N-100}}{3600 \log g}. \quad (3.3)$$

Recalling the definition of I , all these convenient $n \in I$ with $s(n)/n < 1 + g^{-100}$ satisfy

$$s(n) \geq n \geq g^N,$$

and also

$$s(n) \leq n(1 + g^{-100}) \leq g^N(1 + g^{-100})^2.$$

Hence,

$$N \leq \log_g s(n) \leq N + 2 \log_g(1 + g^{-100}) \leq N + 4g^{-100}.$$

Therefore, the fractional part $\{\log_g s(n)\}$ belongs to the interval $[0, 4g^{-100}]$.

If Benford's law holds for $s(n)$ with respect to natural density, then the number of $n \in I$ for which $\{\log_g s(n)\}$ belongs to $[0, 4g^{-100}]$ is asymptotically $4g^{-100} \cdot g^{N-100} = 4g^{N-200}$ as $N \rightarrow \infty$. In particular, this number of such $n \in I$ is eventually smaller than $8g^{N-200}$. But that upper bound conflicts with the lower bound for the number of convenient $n \in I$ in (3.3) when $8g^{N-200} < g^{N-100}/(7200 \log g)$, or equivalently,

$$57600 \log g < g^{100}.$$

Since $g^{100}/\log g \geq g^{99} > 2^{99} > 57600$, we indeed have a contradiction. $\square$

4. ELLIPSEPHIC $s(n)$ FOR COMPOSITE n

Throughout this section, let x be large. Put $y = y(x) = \exp(\sqrt{\log x})$. Recall that $\mathcal{W}_{\mathcal{D}}$ denotes the set of integers with no digit a_0 in their base g expansion. Such integers are often called *ellipsephic*.

4.1. Preliminary lemmas. Here we record the lemmas that we will use in our proof. The following is a special case of Lemma 2.3 in [11].

Lemma 4.1. *As $x \rightarrow \infty$, we have*

$$\Psi(x, y) := \#\{n \leq x : P^+(n) \leq y\} \ll x \exp\left(-\frac{\sqrt{\log x}}{2}\right).$$

Next, we will need an upper bound for the count of integers divisible by a large square:

Lemma 4.2. *If $x \geq z \geq 2$, then*

$$\#\{n \leq x : P^+(n) > z, P^+(n)^2 \mid n\} \ll x/z, \text{ as } x \rightarrow \infty.$$

Proof. We have

$$\begin{aligned} \#\{n \leq x, P^+(n) > z, P^+(n)^2 \mid n\} &\leq \sum_{z < p \leq \sqrt{x}} \sum_{\substack{n \leq x \\ p^2 \mid n}} 1 \\ &\leq \sum_{z < p} \frac{x}{p^2} \ll x/z. \end{aligned}$$

□

Finally, we will need to count the integers m in a certain range for which $s(m)$ is divisible by a fixed positive integer q .

Lemma 4.3. *Let $\alpha \in (0, 1)$, $\beta \in (0, \alpha/4)$, and let q be a positive integer with $y^\beta < q \leq y$. For all $X \geq x^\alpha$, we have*

$$\#\{x^\alpha < m \leq X : q \mid s(m)\} \ll \frac{\tau(q)}{\varphi(q)} X \log X,$$

where τ is the divisor-counting function and φ is Euler's totient function.

Proof. Write $m = m_1 P_1$ with $P_1 = P^+(m)$. Let $y_2 = y^2$. If $P_1 \leq y_2$, then all prime factors of m are less than y_2 and thus the number of such $m \leq X$ is less than $\Psi(X, y_2)$ which is $\ll X \exp(-\frac{\log X}{4\sqrt{\log x}}) \leq X \exp(-\alpha \frac{\sqrt{\log x}}{4}) = X y^{-\alpha/4}$ by Lemma 4.1. By Lemma 4.2, we can also suppose that $(P_1, m_1) = 1$. It is thus sufficient to consider the case with $P_1 \nmid m_1$ and $P_1 \geq y_2$. Let A be the corresponding set:

$$A = \{x^\alpha < m_1 P_1 \leq X : q \mid s(m_1 P_1), P^+(m_1) < P_1, P_1 > y_2\}.$$

Then since $s(m) = P_1 s(m_1) + \sigma(m_1)$, we have

$$P_1 s(m_1) \equiv -\sigma(m_1) \pmod{q}.$$

Let $q_1 := \gcd(s(m_1), q)$ and $q_2 = q/q_1$. We fix such a factorization $q_1 q_2$ of q . As q_1 divides both $s(m_1)$ and q , it also divides $\sigma(m_1)$. As a result,

$$q_1 \mid \sigma(m_1) - s(m_1) = m_1.$$

Moreover, $q_2 = q/q_1$ is coprime to $s(m_1)/q_1$, and so the congruence

$$P_1 \frac{s(m_1)}{q_1} \equiv -\frac{\sigma(m_1)}{q_1} \pmod{q_2}$$

puts P_1 in a uniquely determined residue class modulo q_2 . Also, since $y_2 \leq P_1 \leq X/m_1$, so $X/m_1 \geq y_2$. By the Brun-Titchmarsh theorem, for a given m_1 , the number of choices for P_1 is

$$\ll \frac{X/m_1}{\varphi(q_2) \log \frac{X}{m_1 q_2}} \ll \frac{X/m_1}{\varphi(q_2) \log \frac{y_2}{y}}$$

since, by assumption, $q_2 \leq q \leq y$, we have

$$\frac{X}{m_1 q_2} \geq \frac{X}{m_1 y} \geq \frac{y_2}{y} = \frac{y^2}{y} = y.$$

Summing over $m_1 \leq X/y_2$ that are multiples of q_1 , we have (writing $m_1 = q_1 k_1$)

$$\sum_{k_1 \leq X/y_2 q_1} \frac{X/(q_1 k_1)}{\varphi(q_2) \log y} \leq \frac{X}{q_1 \varphi(q_2) \log y} \sum_{k_1 \leq X/y_2 q_1} \frac{1}{k_1} \ll \frac{X \log(X/y_2 q_1)}{q_1 \varphi(q_2) \log y} = \frac{X \log(X/y^2 q_1)}{q_1 \varphi(q_2) \log y}.$$

Now, summing over the choices of q_1 and q_2 yields for $\#A$:

$$\begin{aligned} \#A &\ll \frac{X}{\log y} \sum_{q_1 q_2 = q} \frac{\log(X/y^2 q_1)}{q_1 \varphi(q_2)} \\ &\ll \frac{X \log(X/y^2)}{\log y} \sum_{q_1 q_2 = q} \frac{1}{q_1 \varphi(q_2)}. \end{aligned}$$

As, the sum on q_1, q_2 is a convolution product

$$\sum_{q_1 q_2 = q} \frac{1}{q_1 \varphi(q_2)} = \frac{1}{q} \sum_{q_2 | q} \frac{q_2}{\varphi(q_2)} \leq \frac{\tau(q)}{\varphi(q)},$$

we deduce that

$$\#A \ll \frac{\tau(q)}{\varphi(q)} X \log X.$$

□

4.2. Proof of Theorem 1.5. In the case where $g = 2$, the result trivially holds. Namely, since $a_0 \in \{1, \dots, g-1\}$, we must have $a_0 = 1$ when $g = 2$. Observe that there are no positive integers whose binary expansions contain no digit 1. Since $s(n) > 0$ for all composite values of n , it follows that the count of composite $n \leq x$ for which $s(n)$ has no 1 in its base-2 expansion is 0.

For the remainder of the proof, let $g \geq 3$. We write $n = Pm$ where $P = P^+(n)$. By Lemmas 4.1 and 4.2, we can suppose that $P \geq y$ and $(P, m) = 1$. The last assumption $(P, m) = 1$ implies that

$$s(n) = Ps(m) + \sigma(m). \tag{4.1}$$

First we handle the case $m \leq x^{\alpha_g}$ with $\alpha_g = \frac{1}{2g^2 \log^2 g} \leq 1/4$. We consider

$$E_1 = \# \{mP \leq x : m \leq x^{\alpha_g} \text{ and } s(mP) \text{ has no } a_0 \text{ in its base } g \text{ expansion}\}.$$

We will prove that

$$E_1 \ll x^{1 - \frac{1}{2g^2 \log^2 g}}. \tag{4.2}$$

First we have:

$$E_1 = \sum_{m \leq x^{\alpha_g}} \# \{P \leq x/m : s(mP) \text{ has no } a_0 \text{ in its base } g \text{ expansion}\}.$$

Since $a := s(n) \leq \sigma(n) \leq n\tau(n)$, we have $\sigma(n) \ll n^{1+\varepsilon}$ for all $\varepsilon > 0$. In particular for $mP \leq x$ with $P \leq x$, $m \leq x^{\alpha_g}$ as before, (4.1) implies that $s(Pm) \leq x^{1+\frac{1}{g \log g}}$, for large enough x . If m is fixed and a is given, there is at most one $P \leq x/m$ such that $s(mP) = a$. In other words we have for any $m \leq x^{\alpha_g}$:

$$\begin{aligned} & \#\{P \leq x/m : s(mP) \text{ has no } a_0 \text{ in its base } g \text{ expansion}\} \\ & \leq \#\left\{a \leq x^{1+\frac{1}{g \log g}} \text{ such that } a \text{ has no } a_0 \text{ in its base } g \text{ expansion}\right\}. \end{aligned}$$

Thus

$$E_1 \ll \sum_{m \leq x^{\alpha_g}} \#\{a \leq x^{1+\frac{1}{g \log g}} : a \text{ has no } a_0 \text{ in its base } g \text{ expansion}\}. \quad (4.3)$$

For any $X \geq 2$, the number of $a \leq X$ with no a_0 in base g is $O((g-1)^N)$ where N is the smallest integer such that $X \leq g^N$ (ie. $N = \left\lceil \frac{\log X}{\log g} \right\rceil$). Inserting this into (4.3) allows us to obtain

$$E_1 \leq \sum_{m \leq x^{\alpha_g}} x^{(1+\frac{1}{g \log g}) \frac{\log(g-1)}{\log g}}$$

Note that $\frac{\log(g-1)}{\log g} = 1 + \frac{\log(1-\frac{1}{g})}{\log g}$ which implies

$$\left(1 + \frac{1}{g \log g}\right) \frac{\log(g-1)}{\log g} \leq 1 - \frac{1}{g^2 \log^2 g}.$$

Thus, recalling $\alpha_g = \frac{1}{2g^2 \log^2 g}$, we obtain (4.2)

$$E_1 \ll x^{\alpha_g + 1 - \frac{1}{g^2 \log^2 g}} = x^{1 - \frac{1}{2g^2 \log^2 g}}.$$

We now suppose that $m > x^{\alpha_g}$. Let $k \in \mathbb{Z}$ such that $g^k \leq y < g^{k+1}$. Then for $\mathcal{D} = \{0, 1, \dots, g-1\} \setminus \{a_0\}$,

$$\#\{mP \leq x : m > x^{\alpha_g}, P > y, P \nmid m, s(mP) \text{ with no digit } a_0\} \leq \sum_{\substack{b < g^k \\ b \in \mathcal{W}_{\mathcal{D}}}} \sum_{x^{\alpha_g} < m \leq x/y} \sum_{\substack{y < P \leq x/m \\ P \nmid m \\ s(mP) \equiv b \pmod{g^k}}} 1.$$

Noting (4.1), $s(n) = s(mP) = Ps(m) + \sigma(m) \equiv b \pmod{g^k}$.

Let $d = \gcd(s(m), g^k)$. We first consider the case where the gcd d is small; that is, $d \leq y_1$ with $y_1 = y^{2r\beta}$ and $\beta \leq \alpha_g/4$, where $r = \omega(g)$ with $\omega(g)$ denoting the number of distinct prime divisors of g . Then

$$\frac{Ps(m)}{d} \equiv \frac{b - \sigma(m)}{d} \pmod{\frac{g^k}{d}}.$$

Now, trivially,

$$\#\left\{y < P \leq \frac{x}{m} : \frac{Ps(m)}{d} \equiv \frac{b - \sigma(m)}{d} \pmod{\frac{g^k}{d}}\right\} \leq \frac{x/m}{g^k/d}.$$

Summing over m and d yields

$$\sum_{x^{\alpha_g} < m \leq x/y} \sum_{1 \leq d \leq y_1} \frac{x}{m} \frac{d}{g^k} \leq \frac{x}{g^k} \sum_{x^{\alpha_g} < m \leq x/y} \frac{1}{m} \sum_{1 \leq d \leq y_1} d$$

$$\leq \frac{xy_1^2 \log x/y}{g^k y} \ll \frac{x}{y^{1/2}}.$$

The final inequality follows from the fact that we have chosen k such that $g^k \leq y < g^{k+1}$, thus $1/g^k \leq g/y \ll 1/y$. Finally, summing over b ,

$$\sum_{b \in \mathcal{W}_{\mathcal{D}}(g^k)} \frac{x}{y^{1/2}} \leq \frac{x}{y^{1/2}} (g^k)^{\frac{\log \#\mathcal{D}}{\log g}} \ll x/y^{1/2 - \frac{\log(g-1)}{\log g}}.$$

Now, we consider the contribution from m , P , and b when $(s(m), g^k) > y_1$. Let V denote this contribution:

$$V = \#\{mP \leq x : (s(m), g^k) \geq y_1, P > y, m > x^{\alpha_g}, (P, m) = 1\}.$$

Since $mP \leq x$ and $m > x^{\alpha_g}$, we have $P \leq x^{1-\alpha_g}$.

In order to apply Lemma 4.3 with a convenient modulus q , we use the prime factor decomposition of g :

$$g = p_1^{\gamma_1} \cdots p_r^{\gamma_r},$$

where p_j are primes with $p_i \neq p_j$ if $i \neq j$ and $\gamma_1, \dots, \gamma_r \in \mathbb{N}$. If $\gcd(s(m), g^k) \geq y_1$, then there exists $i \in \{1, \dots, r\}$ such that $\gcd(s(m), p_i^{k\gamma_i}) > y_1^{1/r}$. This implies that $p_i^{\ell_i} | s(m)$ for some $\ell_i \in \mathbb{N}$ and we can choose ℓ_i with $p_i^{\ell_i} \leq y_1^{1/r} < p_i^{\ell_i+1}$. Thus we have

$$V \ll \sum_{y < P < x^{1-\alpha_g}} \#\{x^{\alpha_g} \leq m \leq x/P : p_i^{\ell_i} | s(m)\}.$$

Now, we apply Lemma 4.3 with $q = p_i^{\ell_i} > y_1^{1/(2r)} = y^{\beta_g}$ and $X = x/P$. Since

$$\frac{\tau(p_i^{\ell_i})}{\varphi(p_i^{\ell_i})} = \frac{1 + \ell_i}{p_i^{\ell_i-1}(p_i - 1)} \ll y_1^{-1/r} \log(y_1) \ll y^{-2\beta_g} \log y \ll y^{-\beta_g} \log x,$$

we obtain

$$V \ll \sum_{y < P < x^{1-\alpha_g}} \frac{x(\log x)^2}{Py^{\beta_g}}.$$

Summing over P , and recalling that $y = \exp(\sqrt{\log x})$, we obtain, for some constant $c > 0$,

$$V \ll x \exp(-c\sqrt{\log x}).$$

APPENDIX A. EXTENDING THEOREM 1.4 TO $g = 2$

In this appendix, we prove the omitted case $g = 2$ of Theorem 1.4, thereby completing the proof of the theorem in every base.

Observe that, when $g = 2$, we must have $\mathcal{D} = \{1\}$, since there are no positive integers whose binary expansions contain only 0's. The positive integers whose binary expansions contain only the digit 1 are precisely $2^j - 1$ for $j \geq 1$. Thus, we want to count the values of $n \leq x$ for which

$$s(n) = 2^j - 1.$$

Let

$$K = \left\lceil \frac{2(\log \log x)^\gamma}{\log 2} \right\rceil.$$

If $2^K \mid \sigma(n)$, then

$$n = \sigma(n) - s(n) \equiv 1 - 2^j \pmod{2^K}.$$

As j varies, there are at most K possible residue classes $\pmod{2^K}$ that n can belong to: the classes $1 - 2^j$ for $1 \leq j < K$ and the class $1 \pmod{2^K}$ for all $j \geq K$. Consequently, this case contributes

$$K \left(\frac{x}{2^K} + 1 \right) \ll x e^{-(\log \log x)^\gamma}$$

values of n to the count.

If $2^K \nmid \sigma(n)$, define

$$R(n) := \#\{p : p \text{ odd prime}, \nu_p(n) \text{ odd}\}.$$

Since

$$\sigma(n) = \sigma(2^{\nu_2(n)}) \prod_{\substack{p \mid n \\ p \text{ odd}}} \sigma(p^{\nu_p(n)}),$$

where $\sigma(2^{\nu_2(n)})$ is odd, and $\sigma(p^{\nu_p(n)})$ is even whenever $\nu_p(n)$ is odd. So, each odd prime divisor of n with an odd $\nu_p(n)$, there are $R(n)$ many of them, contributes at least one factor of 2 to $\sigma(n)$. Therefore, we have

$$\nu_2(\sigma(n)) \geq R(n).$$

Since $2^K \nmid \sigma(n)$, we have

$$\nu_2(\sigma(n)) < K,$$

hence $R(n) < K$. Now, we estimate how many integers $n \leq x$ there are with $R(n) < K$. Fix $t \in (0, 1)$, say $t = 1/2$. Then we have

$$\mathbf{1}_{\{R(n) < K\}} \leq t^{R(n)-K}.$$

Indeed, if $R(n) < K$, then $R(n) - K < 0$, and since $0 < t < 1$, we have $t^{R(n)-K} \geq 1$. Note that $(\frac{1}{2})^{R(n)}$ is multiplicative and, by the Selberg-Delange method (see [15, Theorem II.5.3]), it has mean value

$$\sum_{n \leq x} \left(\frac{1}{2} \right)^{R(n)} \ll \frac{x}{\sqrt{\log x}}.$$

Thus, we may conclude that

$$\#\{n \leq x : 2^K \nmid \sigma(n)\} \ll 2^K \frac{x}{\sqrt{\log x}} \ll x e^{-(\log \log x)^\gamma},$$

since $\gamma < 1$.

ACKNOWLEDGEMENTS

We would like to acknowledge ANR grant ANR-20-CE91-0006 ArithRand and the Utrecht University Mathematics Institute for providing funding for us to meet up in Nancy and Utrecht in order to work on this project. A portion of this manuscript was written while the fifth author was on sabbatical at the Max Planck Institute for Mathematics and the Centre de Recherches Mathématiques. She would like to thank both institutions for providing her with a pleasant working environment.

REFERENCES

- [1] K. Benli, G. Cesana, C. Dartyge, C. Dombrowsky, and L. Thompson, *Sums of proper divisors with missing digits*, Research Directions in Number Theory, Springer AWMS, **32** (2024), 93–110.
- [2] A. Berger and T. P. Hill, *The Mathematics of Benford’s Law: A Primer*, Statistical Methods & Applications **30** (2021), 779–795.
- [3] P. W. Diaconis, *The distribution of leading digits and uniform distribution mod 1*, Ann. Probability **5** (1977), no. 1, 72–81; MR0422186
- [4] H. Davenport, *Über numeri abundantes*, S.-Ber. Preuss. Akad. Wiss., math.-nat. Kl. (1933), 830–837.
- [5] P. Erdős, *Some remarks about additive and multiplicative functions*, Bull. Amer. Math. Soc. **52** (1946), 527–537.
- [6] P. Erdős, A. Granville, C. Pomerance, and C. Spiro, *On the normal behavior of the iterates of some arithmetic functions*, Analytic number theory (Allerton Park, IL, 1989), 165–204, Progr. Math. **85**, Birkhäuser Boston, Boston, MA, 1990.
- [7] C. Gong, Y. Gu, J. Lu, P. Pollack, *On the stable reduction of hyperelliptic curves*, Tohoku Math. J. **74** (2022), 195–213.
- [8] A. Granville and K. Soundararajan, *Multiplicative Number Theory: The Pretentious Approach*, preprint, 2014, <https://dms.umontreal.ca/~andrew/PDF/Book.To2.5.pdf>.
- [9] M. Kobayashi, P. Pollack, and C. Pomerance, *On the distribution of sociable numbers*, J. Number Theory **129** (2009), 1990–2009.
- [10] P. Pollack, *Palindromic sums of proper divisors*, Integers **15A** (2015), Paper No. A13, 12 pp.
- [11] P. Pollack, *Some arithmetic properties of the sum of proper divisors and the sum of prime divisors*, Illinois J. Math. **58** (2014), no. 1, 125–147.
- [12] P. Pollack, C. Pomerance, and L. Thompson, *Divisor-sum fibers*, Mathematika **64** (2018), no. 2, 330–342.
- [13] C. Pomerance, *The first function and its iterates*, Connections in Discrete Mathematics: A Celebration of the Work of Ron Graham, pp. 125–138. Cambridge University Press, 2018.
- [14] J. B. Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. **6** (1962), no. 1, 64–94.
- [15] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Third edition, Translated from the 2008 French edition by Patrick D. F. Ion. Graduate Studies in Mathematics, **163**, American Mathematical Society, Providence, RI, 2015, xxiv, 629 pp.
- [16] E. C. Titchmarsh, *The Theory of the Riemann Zeta-Function*, 2nd ed., revised by D. R. Heath-Brown. Oxford Science Publications. Oxford: Clarendon Press, 1986. x + 412 pp. ISBN 0-19-853369-1.
- [17] L. Troupe, *Divisor sums representable as a sum of two squares*, Proc. Amer. Math. Soc. **148** (2020), no. 10, 4189–4202.

BOĞAZICI UNIVERSITY, DEPARTMENT OF MATHEMATICS, BEBEK, 34342, İSTANBUL, TÜRKİYE
Email address: kubra.benli@bogazici.edu.tr

INSTITUT ÉLIE CARTAN DE LORRAINE & INSTITUT UNIVERSITAIRE DE FRANCE, UNIVERSITÉ DE LORRAINE, BP 70239,
 54506 VANDŒUVRE-LÈS-NANCY CEDEX, FRANCE
Email address: cecile.dartyge@univ-lorraine.fr

UNIVERSITÄT BIELEFELD, FAKULTÄT FÜR MATHEMATIK, POSTFACH 100131, 33501 BIELEFELD, GERMANY
Email address: cdombrow@math.uni-bielefeld.de

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF GEORGIA, ATHENS, GA 30602, UNITED STATES
Email address: pollack@uga.edu

MATHEMATICS INSTITUTE, UTRECHT UNIVERSITY, HANS FREUDENTHALGEBOUW, BUDAPESTLAAN 6, 3584 CD UTRECHT,
 THE NETHERLANDS
Email address: l.thompson@uu.nl