DIRICHLET, SIERPIŃSKI, AND BENFORD

PAUL POLLACK AND AKASH SINGHA ROY

Abstract. Sixty years ago, Sierpiński observed that for any positive integers A and B, and any $g \geq 2$, there are infinitely many primes whose base g-expansion begins with the digits of A and ends with those of B. Sierpiński's short proof rests on the prime number theorem for arithmetic progressions (PNT for APs). We explain how his result can be viewed as a natural intermediary between Dirichlet's theorem on primes in progressions and the PNT for APs. In addition to being of pedagogical interest, this perspective quickly yields a generalization of Sierpiński's result where the initial and terminal digits of p are prescribed in two coprime bases simultaneously; moreover, the proportion (Dirichlet density) of the corresponding primes is determined explicitly. The same quasielementary method shows that the arithmetic functions $\varphi(n)$, $\sigma(n)$, and $d(n)$ obey "Benford's law" in a suitable sense.

1. Introduction

This note centers around the following charming theorem published by Sierpiński in 1959 [18].

Theorem A. *Let $g \geq 2$. Let A and B be positive integers, and assume that $\gcd(B, g) = 1$. Then there are infinitely many primes whose digital expansion in base g begins with the digits of A and ends with the digits of B.*

Sierpiński's proof is short and simple. Let ℓ be the length of the digital expansion of B in base g. It is enough to show that for infinitely many k, the interval $[A \cdot g^k, (A+1) \cdot g^k)$ contains at least one prime from the progression $B \bmod g^\ell$. The prime number theorem for arithmetic progressions [11, Theorem 1, p. 149] tells us that the number of primes $p \leq X$ with $p \equiv B \pmod{g^\ell}$ is $\sim \frac{1}{\varphi(g^\ell)} X/\log X$, as $X \to \infty$. It follows by a short calculation that the number of $p \equiv B \pmod{g^\ell}$ in $[A \cdot g^k, (A+1) \cdot g^k)$ is $\sim \frac{1}{\varphi(g^\ell)} g^k/\log(g^k)$ (as $k \to \infty$). This final quantity tends to infinity with k, so that we obtain primes of the desired form for all large values of k.

In 2006, Harman obtained a remarkable extension of Sierpiński's theorem [9]: Fix integers $g \geq 2$ and $L \geq 0$. *For all large N, there is an N-digit prime in base g having* any *L digits arbitrarily preassigned* (assuming compatibility with a nonzero leading digit and a final digit coprime to g). So for example, for all large odd N one can find an N-digit prime in base 10 whose decimal expansion begins with `271828`, has middle five digits `57721`, and ends with `314159`. Quite recently, Swaenepoel [19] has proved that Harman's result continues to hold even when L grows with N, as long as $L \leq c_g N$, where $c_g > 0$ is a certain constant depending on g. This builds on work of Bourgain [3, 4] (who handled $g = 2$) and sharpens a theorem of Harman and Kátai [10] (who needed L a little smaller than $N^{1/2}$). In the special case where only leading and ending digits are prescribed (i.e., the situation of Theorem A), Thorner and Zaman [21, Corollary 3.2] have shown that (in any base g) any constant $c_g < \frac{5}{12}$ is admissible.

2020 *Mathematics Subject Classification.* Primary 11A63; Secondary 11M06, 11N05.

Out first concern in this article is with a different sort of generalization of Theorem A. As in that result, we consider only leading and ending digits, but we work in two bases simultaneously.

Recall that if $\mathcal{P}$ is a set of primes, the Dirichlet density (or analytic density) of $\mathcal{P}$, with respect to the set of primes, is defined as

$$\lim_{\sigma \downarrow 1} \frac{\sum_{p \in \mathcal{P}} p^{-\sigma}}{\sum_p p^{-\sigma}}.$$

(Here and below, sums on p are always understood to run over primes; so for example, the sum in the denominator runs over all primes p.) As known already to Euler, $\sum_p p^{-\sigma}$ blows up as $\sigma \downarrow 1$ [16, Corollary 2, p. 70]. It follows that a set of primes with positive Dirichlet density is necessarily infinite. One should view the Dirichlet density of $\mathcal{P}$ as measuring, in a certain sense, the proportion of primes belonging to $\mathcal{P}$. The notion of Dirichlet density appears (implicitly) in Dirichlet's 1837 proof that each coprime progression a mod q contains infinitely many primes; what his argument really shows is that the set of such primes has Dirichlet density $1/\varphi(q)$.

Theorem 1. *Let $g, g' \geq 2$ and assume that $\gcd(g, g') = 1$. Let A, B, A', B' be positive integers with $\gcd(B, g) = \gcd(B', g') = 1$. Then there are infinitely many primes which simultaneously have a base g expansion beginning with the digits of A and ending with those of B, as well as a base g' expansion beginning with the digits of A' and ending with those of B'. Moreover, the set of such primes has Dirichlet density*

$$\frac{1}{\varphi(g^\ell g'^{\ell'})} \frac{\log(1+\frac{1}{A})}{\log g} \cdot \frac{\log(1+\frac{1}{A'})}{\log g'},$$

where ℓ and ℓ' are the lengths of the digit expansions of B and B' in bases g and g', respectively.

We do not think Theorem 1 has previously appeared, but we would not be very surprised to be wrong! The basic idea is anything but new. Essentially, we follow Dirichlet's proof of his theorem on primes in prescribed arithmetic progressions (as explained, for instance, in [16, Chapter VI]), with Fourier analysis on $\mathbb{R}/\mathbb{Z}$ supplementing the familiar Fourier analysis on $(\mathbb{Z}/q\mathbb{Z})^\times$. We also need the nonvanishing of Dirichlet L-functions on the entire line $\Re(s) = 1$ and not merely at $s = 1$ itself. Perhaps interestingly, we do not make any direct appeals to any version of the prime number theorem.[1] Theorem 1 could thus be seen as providing a natural motivation for proving nonvanishing of L-functions at $s = 1$. This seems of some pedagogical interest: When teaching an introductory course on analytic number theory, Theorem 1 provides an aesthetically appealing waypoint between Dirichlet's theorem and the prime number theorem.

The proof of Theorem 1 is presented in §3, after we have laid the groundwork in §2. In §4 we conclude by discussing applications of the same method to the leading digit distribution of the functions $\varphi(n)$, $\sigma(n)$, and $d(n)$. Specifically, we will show that all three functions obey "Benford's law", suitably interpreted.

[1]The conventional wisdom in analytic number theory is that the nonvanishing of Dirichlet L-functions on $\Re(s) = 1$ is philosophically equivalent to the PNT for APs. However, as a practical matter, deducing the latter from the former is a nontrivial task. Even the simplest arguments of this kind (such as Newman's approach described in [11, Chapter 5] and [15]) require decidedly more intricate machinations than the Fourier analysis we use here, which is at the same level as the proof of Weyl's equidistribution criterion.

2. Prescribing leading digits of primes, revisited

As a warm up exercise, we give a Dirichlet-style proof of a weak version of Theorem A, namely of the following improvement of a result of Sierpiński (c.f., [17]) dating back to 1951.

Claim. The set of primes beginning with A in base g has Dirichlet density $\frac{\log(1+1/A)}{\log g}$. In particular, there are infinitely many such primes.

The condition that the base g expansion of p begins with the digits of A is equivalent to the requirement that

$$\frac{\log p}{\log g} \in \left[\frac{\log A}{\log g}, \frac{\log(A+1)}{\log g}\right) \pmod 1.$$

("Mod 1" means we identify real numbers differing by an integer; that is, we work in the 1-dimensional torus $\mathbb{R}/\mathbb{Z}$.) To state this equivalence, we adopt the convention that the infinitely many zeros after the "decimal point" (really, "g-imal point") are considered part of the digital expansion when comparing leading digits. Hence, we are allowed to say that the number 37 "begins with the digits of 370000", since $37 = 37.000\ldots$. This convention may seem slightly strange, but in any case it is not so important in this problem: Once $p \geq A$, saying that p begins with the digits of A means what every right-thinking person believes it should, while the finitely many primes $p < A$ are irrelevant to the computation of the Dirichlet density.

We will detect the condition that $\frac{\log p}{\log g}$ belongs to $\left[\frac{\log A}{\log g}, \frac{\log(A+1)}{\log g}\right)$ mod 1 by means of the additive characters $e_k(x) := e^{2\pi i k x}$ ($k \in \mathbb{Z}$), which map $\mathbb{R}/\mathbb{Z}$ to the complex unit circle. By a trigonometric polynomial, we mean a $\mathbb{C}$-linear combination of additive characters, i.e., a function on $\mathbb{R}/\mathbb{Z}$ of the form

$$T(x) := \sum_{k=-K}^{K} c_k e_k(x), \tag{1}$$

where K is a nonnegative integer and the c_k are complex numbers. We make crucial use of the following fundamental theorem of Fejér from basic Fourier analysis (which could also be deduced from the Stone–Weierstrass theorem): *Every continuous function on $\mathbb{R}/\mathbb{Z}$ is a uniform limit of trigonometric polynomials.* (For the Fourier-theoretic proof see, e.g., [22, Corollary 6.14, p. 118].)

Suppose that T is a trigonometric polynomial, written as in (1), and let $\theta := 2\pi/\log g$. For each $\sigma > 1$,

$$\begin{aligned}\sum_p \frac{T(\log p/\log g)}{p^\sigma} &= \sum_{k=-K}^{K} c_k \sum_p \frac{1}{p^{\sigma - ik\theta}} \\ &= c_0 \sum_p \frac{1}{p^\sigma} + \sum_{\substack{-K \leq k \leq K \\ k \neq 0}} c_k \sum_p \frac{1}{p^{\sigma - ik\theta}}.\end{aligned} \tag{2}$$

To understand the rightmost inner sums on p, we recall that there is a branch of $\log \zeta(s)$, holomorphic on $\Re(s) > 1$, given by $Z(s) = \sum_{p^m} \frac{1}{m p^{ms}}$, where the sum is over all prime powers

p^m (see, e.g., p. 74 of [16]). Discarding the contribution from terms with $m \geq 2$, we find that

$$\left| \sum_p \frac{1}{p^{\sigma - ik\theta}} - Z(\sigma - ik\theta) \right| < 1.$$

(See the proof of Corollary 2 on p. 70 of [16].) Since $\zeta(s)$ has no zeros with $\Re(s) = 1$ (p. 120 of [11]), $Z(s)$ continues analytically to $1 - ik\theta$. Hence, $\sum_p 1/p^{\sigma - ik\theta}$ stays bounded as $\sigma \downarrow 1$. Referring back to (2), and noting that $c_0 = \int_0^1 T(x)\,\mathrm{d}x$, we deduce that

$$\lim_{\sigma \downarrow 1} \frac{\sum_p T(\frac{\log p}{\log g}) p^{-\sigma}}{\sum_p p^{-\sigma}} = \int_0^1 T(x)\,\mathrm{d}x.$$

Now we use Fejér's theorem: since each continuous function f is a uniform limit of trigonometric polynomials, this same equality holds with T replaced by any continuous f. Indeed, if $|T - f| \leq \varepsilon$ everywhere, then the triangle inequality shows that replacing T by f on the right-hand side changes the integral by at most ε, and changes the left-hand ratio of sums by at most ε (for each $\sigma > 1$).

Let I be an interval of the form $[\alpha, \beta)$, with $\beta - \alpha \leq 1$. (We will eventually specialize to $I = [\frac{\log A}{\log g}, \frac{\log(A+1)}{\log g})$.) Write $\mathbb{1}_I$ for the characteristic function of I mod 1. For any $\varepsilon > 0$, we can choose continuous functions $\mathbb{1}_I^+, \mathbb{1}_I^-$ from $\mathbb{R}/\mathbb{Z}$ to $[0,1]$ with

$$\mathbb{1}_I^- \leq \mathbb{1}_I \leq \mathbb{1}_I^+,$$

and such that the differences $\mathbb{1}_I^\pm - \mathbb{1}_I$ are supported on $[\alpha - \varepsilon, \alpha + \varepsilon] \cup [\beta - \varepsilon, \beta + \varepsilon]$ mod 1. Then

$$\limsup_{\sigma \downarrow 1} \frac{\sum_p \mathbb{1}_I(\frac{\log p}{\log g}) p^{-\sigma}}{\sum_p p^{-\sigma}} \leq \int_0^1 \mathbb{1}_I^+(x)\,\mathrm{d}x \leq |I| + 4\varepsilon,$$

and similarly

$$\liminf_{\sigma \downarrow 1} \frac{\sum_p \mathbb{1}_I(\frac{\log p}{\log g}) p^{-\sigma}}{\sum_p p^{-\sigma}} \geq \int_0^1 \mathbb{1}_I^-(x)\,\mathrm{d}x \geq |I| - 4\varepsilon.$$

Taking ε arbitrarily small, we conclude that the set of p with $\frac{\log p}{\log g}$ belonging to I mod 1 has Dirichlet density $|I|$.

Finally, taking $I = [\frac{\log A}{\log g}, \frac{\log(A+1)}{\log g})$ completes the proof of our claim.

It seems worth saying a word about the history of this density result. At the end of Chapter VI in *A course in arithmetic*, Serre briefly considers the set P^1 of primes whose decimal expansion begins with the digit 1:

> One sees easily, using the prime number theorem, that P^1 does not have a natural density[2] and on the other hand Bombieri has shown me a proof that the analytic density of P^1 exists (it is equal to $\log_{10} 2 = 0.301029995\ldots$).

[2]The natural density of a set of primes $\mathcal{P}$ is the limit, as $X \to \infty$, of the proportion of primes up to X belonging to $\mathcal{P}$. By an elementary argument with partial summation, whenever the natural density exists, it coincides with the Dirichlet density.

For any string of initial decimal digits, Whitney determined the logarithmic density of the corresponding set of primes in 1972 [23], using a precise estimate for $\sum_{p\le X}\frac{1}{p}$ that follows from (a strong form of) the prime number theorem. His result is equivalent to ours for $g=10$.[3] As far as we know, Bombieri's argument was never published, but a proof of his claim based on the prime number theorem is sketched in Exercise 7.16, p. 244 of [6].

3. Sierpiński's theorem for coprime bases: Proof of Theorem 1

Let us move now to the main event. A prime p satisfies the desired conditions if and only if

$$\frac{\log p}{\log g}\in I:=\left[\frac{\log A}{\log g},\frac{\log(A+1)}{\log g}\right)\pmod 1,\qquad p\equiv B\pmod{g^\ell}$$

and

$$\frac{\log p}{\log g'}\in I':=\left[\frac{\log A'}{\log g'},\frac{\log(A'+1)}{\log g'}\right)\pmod 1,\qquad p\equiv B'\pmod{g'^{\ell'}}.$$

Choose an integer r with $r\equiv B\pmod{g^\ell}, r\equiv B'\pmod{g'^{\ell'}}$, and put $q:=g^\ell g'^{\ell'}$. Then the Dirichlet density of the set under consideration is given by

$$\lim_{\sigma\downarrow 1}\frac{\sum_{p\equiv r\pmod q}\mathbb{1}_I(\frac{\log p}{\log g})\mathbb{1}_{I'}(\frac{\log p}{\log g'})p^{-\sigma}}{\sum_p p^{-\sigma}}.$$

Paralleling our work in §2, we will bound the indicator functions of I and I' by continuous functions and approximate these in turn by trigonometric polynomials. To that end, we consider two arbitrary trigonometric polynomials

$$T(x):=\sum_{-K\le k\le K}c_k e_k(x)\quad\text{and}\quad T'(x):=\sum_{-L\le l\le L}c'_l e_l(x).$$

(Here $'$ does not mean derivative, of course!) Letting $\theta:=2\pi/\log g$ and $\theta':=2\pi/\log g'$, the orthogonality of Dirichlet characters shows that for each $\sigma>1$,

$$\begin{aligned}\sum_{p\equiv r\pmod q}\frac{1}{p^\sigma}T\left(\frac{\log p}{\log g}\right)T'\left(\frac{\log p}{\log g'}\right)&=\frac{1}{\varphi(q)}\sum_\chi\overline{\chi}(r)\frac{\chi(p)}{p^\sigma}\left(\sum_{-K\le k\le K}c_k p^{ik\theta}\right)\left(\sum_{-L\le l\le L}c'_l p^{il\theta'}\right)\\&=\frac{1}{\varphi(q)}\sum_\chi\overline{\chi}(r)\sum_{\substack{-K\le k\le K\\-L\le l\le L}}c_k c'_l\sum_p\frac{\chi(p)}{p^{\sigma-i(k\theta+l\theta')}}\end{aligned}\tag{3}$$

where χ runs over the Dirichlet characters modulo q.

For each χ, let $Z(s,\chi)=\sum_{p^m}\frac{\chi(p^m)}{mp^{ms}}$. Recall that $Z(s,\chi)$ is an analytic logarithm of $L(s,\chi)$ in $\Re(s)>1$ and that $|Z(s,\chi)-\sum_p\chi(p)p^{-s}|<1$ whenever $\Re(s)>1$ (see again pp. 70, 74 of [16]).

Suppose that χ is a nontrivial character mod q. Since $L(s,\chi)$ is analytic and nonvanishing for $\Re(s)\ge 1$ (see p. 145 of [11]), it follows that for any k,l, the quantity $Z(\sigma-i(k\theta+l\theta'),\chi)$ is

[3]The logarithmic density of a set of primes $\mathcal{P}$ is defined as $\lim_{X\to\infty}\frac{\sum_{p\in\mathcal{P},\ p\le X}p^{-1}}{\sum_{p\le X}p^{-1}}$. One can show that the logarithmic and Dirichlet density, with respect to primes, are equivalent: A set of primes with a Dirichlet density necessarily has the same logarithmic density, and vice versa.

bounded as $\sigma \downarrow 1$. Hence, the nontrivial characters χ mod q make bounded contributions to the sum (3). On the other hand, the trivial character χ_0 makes a contribution to (3) of

$$\frac{c_0 c_0'}{\varphi(q)} \sum_{p \nmid q} \frac{1}{p^\sigma} + \frac{1}{\varphi(q)} \sum_{\substack{-K \le k \le K \\ -L \le l \le L \\ (k,l) \neq (0,0)}} c_k c_l' \sum_{p \nmid q} \frac{1}{p^{\sigma - i(k\theta + l\theta')}}. \tag{4}$$

Since g and g' are relatively prime, they are also multiplicatively independent: Whenever $g^e g'^{e'} = 1$ with integers e, e', it must be that $e = e' = 0$. Hence, $\log g$ and $\log g'$ are linearly independent over $\mathbb{Q}$. It follows that $k\theta + l\theta' = 0$ for integers k, l only when $k = l = 0$. Thus, if $(k,l) \neq (0,0)$, then (by the nonvanishing of $\zeta(s)$ on $\Re(s) = 1$) $Z(s, \chi_0)$ continues analytically to $1 - i(k\theta + l\theta')$. We conclude that each of the inner sums in (4) is bounded, as $\sigma \downarrow 1$.

Putting our observations together and noting that $c_0 = \int_0^1 T(x)\,\mathrm{d}x$, $c_0' = \int_0^1 T'(x)\,\mathrm{d}x$, we obtain

$$\lim_{\sigma \downarrow 1} \frac{\sum_{p \equiv r \pmod q} T(\frac{\log p}{\log g}) T'(\frac{\log p}{\log g'}) p^{-\sigma}}{\sum_p p^{-\sigma}} = \frac{1}{\varphi(q)} \int_0^1 T(x)\,\mathrm{d}x \cdot \int_0^1 T'(x)\,\mathrm{d}x.$$

Hereafter, we note that T and T' may be replaced by arbitrary continuous functions f and f' on $\mathbb{R}/\mathbb{Z}$. Indeed, if $|T - f| \le \delta$ and $|T' - f'| \le \delta$, then the triangle inequality shows that $|TT' - ff'| \ll_{f,f'} \delta$, and we can argue with Fejér's theorem as before. To finish matters off, we let $\varepsilon > 0$, and we bound $\mathbb{1}_I \mathbb{1}_{I'}$ from above by $\mathbb{1}_I^+ \mathbb{1}_{I'}^+$ and below by $\mathbb{1}_I^- \mathbb{1}_{I'}^-$ (with the continuous functions $\mathbb{1}_I^\pm, \mathbb{1}_{I'}^\pm$ constructed as in §2). In this way we obtain an upper bound on the lim sup, and a lower bound on the lim inf, of

$$\frac{\sum_{p \equiv r \pmod q} \mathbb{1}_I(\frac{\log p}{\log g}) \mathbb{1}_{I'}(\frac{\log p}{\log g'}) p^{-\sigma}}{\sum_p p^{-\sigma}},$$

as $\sigma \downarrow 1$. Both bounds tend to $\frac{1}{\varphi(q)}|I||I'|$ as $\varepsilon \to 0$. Recalling the definitions of q and I, I' completes the proof of Theorem 1.

Remarks.

(1) As is clear from the proof, if one wishes only to prescribe leading digits then it is not necessary to assume that $\gcd(g, g') = 1$; it suffices to have g and g' multiplicatively independent. It is not necessary to assume either of those conditions to specify leading digits to one base and terminal digits to the other.

(2) It is natural to wonder about the analogue of Theorem 1 for more than two bases. Our proof goes through for bases $g_1, \dots, g_k$ as long as (a) $g_1, \dots, g_k$ are pairwise coprime, and (b) $1/\log g_1, \dots, 1/\log g_k$ are linearly independent over $\mathbb{Q}$. (One can also check that these conditions are necessary for the conclusion of Theorem 1.) We suspect that condition (a) implies (b); in fact, we believe (b) holds whenever $g_1, \dots, g_k$ are multiplicatively independent. This last claim would follow from Schanuel's conjecture that $\mathbb{Q}(z_1, \dots, z_k, \exp(z_1), \dots \exp(z_k))$ has transcendence degree at least k, over $\mathbb{Q}$, whenever $z_1, \dots, z_k$ are $\mathbb{Q}$-linearly independent. However, it seems to be an open problem to establish condition (b) for even a single k-tuple of integers $g_1, \dots, g_k$ with $k > 2$.

(3) It is also possible to generalize our previous results to the case when the bases under consideration are arbitrary real numbers larger than 1. In this case, digital expansions are not necessarily unique; for instance, $\varphi^2 = \varphi + 1$, with φ denoting the golden ratio. Thus, we make a canonical choice of expansion. We define the digit expansion of a positive real number x to a base $\beta > 1$ to be an expression of the form

$$x := \sum_{j=0}^{\infty} a_{k-j}\beta^{k-j}$$

where we have assigned $k := \lfloor \log x / \log \beta \rfloor, x_{k+1} := x$ and for each $j \geq 0$,

$$a_{k-j} := \left\lfloor \frac{x_{k-j+1}}{\beta^{k-j}} \right\rfloor, \quad x_{k-j} := x_{k-j+1} - a_{k-j}\beta^{k-j},$$

so that the a_{k-j} are all nonnegative integers less than β. (It is easily seen that this "greedy" procedure naturally generalizes the familiar notion for positive integers x and integers $\beta > 1$).

Assume β and β' are multiplicatively independent real numbers > 1, and let d and d' be integers with $0 < d < \beta, 0 < d' < \beta'$. Our arguments will show that the set of primes having base β expansion with leading digit d, and (simultaneously) base β' expansion with leading digit d', has Dirichlet density equal to

$$\frac{\log \min\{1 + \frac{1}{d}, \frac{\beta}{d}\}}{\log \beta} \cdot \frac{\log \min\{1 + \frac{1}{d'}, \frac{\beta'}{d'}\}}{\log \beta'}.$$

Here, we have noted that the condition of p beginning with digit d in base β is $\frac{\log p}{\log \beta} \in \left[\frac{\log d}{\log \beta}, \frac{\log(d+1)}{\log \beta}\right)$ mod 1 in the case $d < \lceil \beta \rceil - 1$, whereas it is $\frac{\log p}{\log \beta} \in \left[\frac{\log d}{\log \beta}, 1\right)$ mod 1, in the remaining case $d = \lceil \beta \rceil - 1$. More complicated results could also be worked out for longer strings of leading digits.

4. Leading digits of multiplicative functions

In this section we switch gears to discuss another application of the ideas underlying the proof of Theorem 1.

Recall that set or sequence of positive integers is said obey Benford's law in base g if, for every positive integer A, the proportion of terms starting with the digits of A is given by $\log(1 + \frac{1}{A})/\log g$. (For background on Benford's law, see [2] and [14].) Our work in §2 shows that that the prime numbers are Benford in any base g, if "proportion" means "Dirichlet density relative to the set of primes". Theorem 1 (along with Remark (1) following) implies even more: The prime numbers obey Benford's law in bases g and g' (again, with respect to Dirichlet density), *simultaneously*, as long as g and g' are multiplicatively independent.

In this section, we will be interested not in subsets of primes but in certain sequences indexed by the positive integers. To specify what we mean by Benford's law in this context, we require a precise definition of "proportion" for subsets of $\mathbb{Z}_{>0}$. There are (at least) three reasonable choices. If $\mathcal{E}$ is a set of positive integers, the natural density of $\mathcal{E}$ is defined as

$$\lim_{X \to \infty} \frac{\sum_{n \leq X,\ n \in \mathcal{E}} 1}{\sum_{n \leq X} 1},$$

the logarithmic density of $\mathcal{E}$ is

$$\lim_{X\to\infty} \frac{\sum_{n\le X,\ n\in\mathcal{E}} 1/n}{\sum_{n\le X} 1/n},$$

and the Dirichlet density of $\mathcal{E}$ is[4]

$$\lim_{\sigma\downarrow 1} \frac{\sum_{n\in\mathcal{E}} n^{-\sigma}}{\sum_{n=1}^{\infty} n^{-\sigma}} = \lim_{\sigma\downarrow 1} \left(\zeta(\sigma)^{-1} \sum_{n\in\mathcal{E}} n^{-\sigma} \right).$$

One can prove that the latter two notions are equivalent: Any set with a logarithmic density has the same value as its Dirichlet density, and vice versa.[5] On the other hand, natural density is more stringent: A set with a natural density has that same value as its logarithmic/Dirichlet density, but there are sets with logarithmic/Dirichlet densities that do not possess any natural density. One example is the set of natural numbers with leading decimal digit 1 (this is the natural number analogue of Serre's remark at the end of the last section). For all of this, see, e.g., Chapter III.1 on pp. 413–424 of [20].

Let $\{a_n\}$ be a sequence of positive integers. We say that $\{a_n\}$ is naturally (resp. logarithmically) Benford in base g if, for any positive integer A, the set of n for which a_n begins with the digits of A has natural (resp., logarithmic) density $\log(1+\frac{1}{A})/\log g$. There have been several investigations into the Benford (or non-Benford) nature of common integer sequences. For example, it is a theorem of Diaconis that $\{n!\}$ is naturally Benford in every base g [5]. The sequence of primorials $2, 2\cdot 3, 2\cdot 3\cdot 5, \dots$ is also known to be naturally Benford in every base [12], as is the sequence of values of the classical partition function $p(n)$ (see [1] or [13]). On the other hand, it is straightforward to show that in each base g, the sequence $\{n\}$ of natural numbers is logarithmically Benford but *not* naturally Benford. The same is true for $\{p(n)\}$, where $p(T)$ is a polynomial with integer coefficients taking positive values at $n = 1, 2, 3, \dots$ (see, for instance, the concluding remarks in [13]).

As we now explain, the method of proof of Theorem 1 can be used to establish versions of Benford's law for the classical arithmetic functions $\varphi(n)$ (Euler's totient), $\sigma(n)$ (the sum-of-divisors), and $d(n)$ (the number-of-divisors). Our arguments will be even more elementary than that the proof of Theorem 1, in that no nonvanishing results are required. It suffices to know that $\zeta(s)$ continues analytically to an open set containing $\Re(s) \ge 1$, apart from a pole at $s = 1$.

Theorem 2. *In each base g, both $\{\varphi(n)\}$ and $\{\sigma(n)\}$ are logarithmically Benford. As long as g is not a power of 2, the sequence $\{d(n)\}$ is logarithmically Benford.*

The condition that g is not a power of 2 in the result about $d(n)$ is easily seen to be necessary. More than 60% of integers are squarefree (with respect to natural density, and so also logarithmic density), and for these n, one has $d(n)$ a (usually large) power of 2 (see Theorem 333, p. 355 and Theorem 432, p. 478 in [8]).

[4]These definitions of logarithmic density and Dirichlet density are analogous to, but different than, the definitions of these same terms from earlier in the article. The point is that we are now working relative to the set of all natural numbers, rather than the set of primes.

[5]This lies somewhat deeper than the analogous fact for the corresponding densities relative to the set of primes, which was noted in an earlier footnote.

For the proof of Theorem 2, we work with Dirichlet density rather than the (equivalent) logarithmic density.

Let f be a positive-integer-valued arithmetic function. Proceeding entirely analogously to §2, we find that f is Benford with respect to Dirichlet density (and so also logarithmic density) if, for every nonzero integer k, the quantity

$$\lim_{\sigma\downarrow 1} \frac{\sum_{n=1}^{\infty} e_k(\frac{\log f(n)}{\log g}) n^{-\sigma}}{\sum_{n=1}^{\infty} n^{-\sigma}} = 0. \tag{5}$$

Indeed, in that case we have for any trigonometric polynomial T with constant term c_0 that

$$\lim_{\sigma\downarrow 1} \frac{\sum_{n=1}^{\infty} T(\frac{\log f(n)}{\log g}) n^{-\sigma}}{\sum_{n=1}^{\infty} n^{-\sigma}} = c_0 = \int_0^1 T(x)\mathrm{d}x.$$

Thanks to Fejér, we can replace T here with any continuous function on $\mathbb{R}/\mathbb{Z}$, and the Benford property follows by looking at continuous approximations to the characteristic function of the interval $[\frac{\log A}{\log g}, \frac{\log (A+1)}{\log g})$.

If we assume f is multiplicative then, for every complex s with $\Re(s) > 1$,

$$\sum_{n=1}^{\infty} e_k\left(\frac{\log f(n)}{\log g}\right) n^{-s} = \sum_{n=1}^{\infty} f(n)^{i\theta} n^{-s} = \prod_p \left(1 + f(p)^{i\theta}p^{-s} + f(p^2)^{i\theta}p^{-2s} + \dots\right),$$

where, suppressing the dependence on k, we have set

$$\theta := 2\pi k/\log g.$$

Suppose that $|f(p) - p| \le 1$ for all primes p; this of course holds for both $f = \varphi$ and $f = \sigma$. One would then expect $f(p)^{i\theta} \approx p^{i\theta}$, and indeed

$$f(p)^{i\theta} - p^{i\theta} = p^{i\theta}((f(p)/p)^{i\theta} - 1) = ip^{i\theta}\int_0^{\theta \log (f(p)/p)} \exp(iu)\,\mathrm{d}u,$$

so that

$$|f(p)^{i\theta} - p^{i\theta}| \le |\theta| \cdot |\log(f(p)/p)| \le 2|\theta|/p.$$

The pth term in the product

$$\prod_p \left(1 + f(p)^{i\theta}p^{-s} + f(p^2)^{i\theta}p^{-2s} + \dots\right)\left(1 - p^{i\theta}/p^s\right)$$

has the form $1 + (f(p)^{i\theta} - p^{i\theta})p^{-s} + c_{p^2}p^{-2s} + c_{p^3}p^{-3s} + \dots$, where each $|c_{p^j}| \le 2$. As such, it is $1 + O(p^{-(1+\Re(s))}) + O(p^{-2\cdot\Re(s)})$, thereby establishing that the infinite product in the above display converges to an analytic function on the half-plane $\Re(s) > 1/2$; call this function $G(s)$. Now the representation

$$\sum_{n=1}^{\infty} e_k\left(\frac{\log f(n)}{\log g}\right) n^{-s} = G(s)\zeta(s - i\theta),$$

valid for $\Re(s) > 1$, along with the analyticity of $\zeta(s)$ for $\Re(s) \ge 1$, $s \ne 1$, implies that $\sum_{n=1}^{\infty} e_k(\frac{\log f(n)}{\log g}) n^{-\sigma}$ stays bounded as $\sigma \downarrow 1$. This certainly implies (5), completing the proof

of Theorem 2 for $\varphi(n)$ and $\sigma(n)$. The argument for $f(n) = d(n)$ is similar. In this case,

$$\sum_{n=1}^{\infty} e_k\left(\frac{\log f(n)}{\log g}\right) n^{-s} = H(s)\zeta(s)^{2^{i\theta}} \tag{6}$$

for $\Re(s) > 1$, where

$$H(s) := \prod_p \left(1 + f(p)^{i\theta}p^{-s} + f(p^2)^{i\theta}p^{-2s} + \dots\right)(1 - 1/p^s)^{2^{i\theta}}$$

is analytic for $\Re(s) > 1/2$. Since g is not a power of 2, we know that $2\pi k \log 2/\log g \notin 2\pi\mathbb{Z}$ and so $\Re(2^{i\theta}) = \cos(2\pi k \log 2/\log g) < 1$. So to obtain (5), we may divide both sides of (6) by $\sum_{n=1}^{\infty} n^{-s} = \zeta(s)$ and let s tend down to 1 through real values σ.

Remark. The above arguments do not address the question of whether $\varphi(n), \sigma(n)$ or $d(n)$ is *naturally* Benford. To answer this seems to require deeper methods. In general, a positive integer-valued function $f(n)$ is naturally Benford in base g precisely when $\frac{\log f(n)}{\log g}$ is uniformly distributed mod 1 (see [5, Theorem 1]). By Weyl's criterion, this holds if and only if $f(n)^{i\theta}$ has mean value 0 for each $\theta = 2\pi k/\log g$ ($k \in \mathbb{Z}$, $k \neq 0$). This last criterion is perfectly set up for analysis via Halász's powerful theorem on mean values of multiplicative functions (see §4.3 of [20]). Since both $\varphi(n)^{i\theta}$ and $\sigma(n)^{i\theta}$ "pretend" to be $n^{i\theta}$ (in the suggestive terminology of [7]) whereas $d(n)^{i\theta}$ does not pretend to be n^{it} for any real t (as long as g is not a power of 2), one finds that $\sigma(n)$ and $\varphi(n)$ are not naturally Benford in any base, while $d(n)$ is naturally Benford in every base g not a power of 2.

5. Acknowledgements

We thank the referee for helpful comments that improved the exposition. P.P. is supported by NSF award DMS-2001581.

References

1. T. C. Anderson, L. Rolen, and R. Stoehr, *Benford's law for coefficients of modular forms and partition functions*, Proc. Amer. Math. Soc. **139** (2011), 1533–1541.
2. A. Berger and T. P. Hill, *An introduction to Benford's law*, Princeton University Press, Princeton, NJ, 2015.
3. J. Bourgain, *Prescribing the binary digits of primes*, Israel J. Math. **194** (2013), 935–955.
4. ———, *Prescribing the binary digits of primes, II*, Israel J. Math. **206** (2015), 165–182.
5. P. Diaconis, *The distribution of leading digits and uniform distribution* mod 1, Ann. Probability **5** (1977), 72–81.
6. W. Ellison and F. Ellison, *Prime numbers*, A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York; Hermann, Paris, 1985.
7. A. Granville, *Pretentiousness in analytic number theory*, J. Théor. Nombres Bordeaux **21** (2009), 159–173.
8. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, sixth ed., Oxford University Press, Oxford, 2008.
9. G. Harman, *Primes with preassigned digits*, Acta Arith. **125** (2006), 179–185.
10. G. Harman and I. Kátai, *Primes with preassigned digits. II*, Acta Arith. **133** (2008), 171–184.
11. E. Hlawka, J. Schoissengeier, and R. Taschner, *Geometric and analytic number theory*, Universitext, Springer-Verlag, Berlin, 1991.
12. B. Massé and D. Schneider, *The mantissa distribution of the primorial numbers*, Acta Arith. **163** (2014), 45–58.
13. ———, *Fast growing sequences of numbers and the first digit phenomenon*, Int. J. Number Theory **11** (2015), 705–719.

14. S. J. Miller (ed.), *Benford's Law: theory and applications*, Princeton University Press, Princeton, NJ, 2015.
15. D. J. Newman, *Simple analytic proof of the prime number theorem*, Amer. Math. Monthly **87** (1980), 693–696.
16. J.-P. Serre, *A course in arithmetic*, Graduate Texts in Mathematics, no. 7, Springer-Verlag, New York-Heidelberg, 1973.
17. W. Sierpiński, *Sur l'existence des nombres premiers avec une suite arbitraire de chiffres initiaux*, Matematiche (Catania) **6** (1951), 135–137.
18. _____, *Sur les nombres premiers ayant des chiffres initiaux et finals donnés*, Acta Arith. **5** (1959), 265–266 (1959).
19. C. Swaenepoel, *Prime numbers with a positive proportion of preassigned digits*, Proc. Lond. Math. Soc. (3) **121** (2020), 83–151.
20. G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, third ed., Graduate Studies in Mathematics, vol. 163, American Mathematical Society, Providence, RI, 2015.
21. J. Thorner and A. Zaman, *Refinements to the prime number theorem for arithmetic progressions*, preprint. arXiv:2108.10878 [math.NT], 2021.
22. G. Travaglini, *Number theory, Fourier analysis and geometric discrepancy*, London Mathematical Society Student Texts, vol. 81, Cambridge University Press, Cambridge, 2014.
23. R. E. Whitney, *Initial digits for the sequence of primes*, Amer. Math. Monthly **79** (1972), 150–152.

Department of Mathematics, University of Georgia, Athens, GA 30602

Email address: pollack@uga.edu

ESIC Staff Quarters No.: D2, 143 Sterling Road, Nungambakkam, Chennai 600034, Tamil Nadu, India

Email address: akash01s.roy@gmail.com