Submitted exclusively to the London Mathematical Society
doi:10.1112/0000/000000

Primes whose sum of digits is prime and metric number theory

Glyn Harman

Abstract

It is shown that almost all real x contain infinitely many primes in their decimal expansions (to any base) whose sum of digits is also prime, generalising a previous result by the author. To do this, the earlier method in metric number theory is combined with recent work by Drmota, Mauduit and Rivat on primes with prescribed sum of digits.

1. *Introduction*

In [**8**] (see also [**9**, Chapters 6 & 8] for the more general context of mixing multiplicative and metrical questions) we gave a complete solution to the following question.

Given an increasing sequence of positive reals a_n, where $a_n \to \infty$ as $n \to \infty$, are there infinitely many primes in the sequences $[\alpha a_n], [\alpha^{a_n}], [a_n^\alpha]$ for almost all real $\alpha > 1$ (in the sense of Lebesgue measure), where $[\ ]$ denotes integer part?

In particular, writing p to denote a prime here and throughout this paper, we showed that if $b \geq 2$ is an integer then, for almost all $\alpha > 0$, we have

$$\lim_{N\to\infty} \frac{1}{\log N} |\{[\alpha b^n] = p; n \leq N\}| = \frac{1}{\log b}. \qquad (1)$$

Thus a "typical" real number has infinitely many primes in its decimal expansion. While [**8**] was being written some interesting developments were taking place regarding our understanding of the sum of digits function restricted to certain sets [**4**, **5**], culminating in the very recent work by Drmota, Mauduit and Rivat [**3**] and by Mauduit and Rivat [**13**]. It is the purpose of this paper to develop the result given by (1) so that only those primes whose sum of digits written in base b ($b \geq 2$) are also prime are counted. For brevity we shall write this set as $\mathcal{P}_b$. The main result we shall prove is then as follows.

Theorem 1. *Let $b \geq 2$ be an integer. Then, for almost all $\alpha > 0$, we have*

$$\limsup_{N\to\infty} \frac{1}{\log\log N} |\{[\alpha b^n] \in \mathcal{P}_b; n \leq N\}| \geq \frac{b-1}{\phi(b-1)\log b}. \qquad (2)$$

and

$$\liminf_{N\to\infty} \frac{1}{\log\log N} |\{[\alpha b^n] \in \mathcal{P}_b; n \leq N\}| \leq \frac{b-1}{\phi(b-1)\log b}. \qquad (3)$$

Remarks. It is slightly disappointing only to be able to give results for the lim sup and lim inf rather than a true asymptotic formula. This arises from weaknesses in our 'overlap estimates' (see Section 6) where we can only obtain upper bounds of the correct order of magnitude rather than asymptotic formulae. Since a countable union of sets with measure zero still has measure zero, for almost all $\alpha > 0$ (2) and (3) hold for *every* $b \geq 2$. A natural question to ask is whether the sequence b^n could be replaced by some increasing sequence of reals d_n

---

2000 *Mathematics Subject Classification* 11K16 (primary), 11N36 (secondary) .

say. We leave it to the reader to verify that the proof could be modified (with suitable changes to the form of (2) and (3)) to accommodate a sequence d_n provided that there exist constants $c_2 > c_1 > 1$ such that

$$c_1 \leq \frac{d_{n+1}}{d_n} \leq c_2 \text{ for all sufficiently large } n.$$

It is plausible that one could obtain a result counting primes whose sum of digits also belongs to $\mathcal{P}_b$ (and iterated further a finite number of times). Certainly some of our auxiliary results can be adapted to this situation, but the work in Section 6 becomes rather complicated at this point.

2. *The sum of digits function*

We state here the results we need on the sum of digits function in base b. An important early work on this topic, which provided the inspiration for later developments, is [**6**]. We write $\log x$ for the natural logarithm modified to mean $\log(\max(2, x))$, and $\log_b x$ is the logarithm to base b. Henceforth we suppose that b is fixed and so we can leave the dependence of various functions on b implicit for simplicity when there will be no confusion. We can likewise write $\mathcal{P}$ instead of $\mathcal{P}_b$. We also put $\mu = (b-1)/2$. If

$$n = \sum_{k=0}^{r} a_k b^k, \quad 0 \leq a_k \leq b-1, \quad \text{then we put } \sigma(n) = \sum_{k=0}^{r} a_k$$

for the sum of digits function in base b. An important feature of this function is that it is concentrated near its mean value $\mu \log_b n$, and in [**3**] (see also [**11**]) it is shown that this still holds true when n takes only prime values. Before stating the result precisely we need the following notation:

$$s = \frac{b^2-1}{12}, \quad S(k,x) = |\{p \leq x : \sigma(p) = k\}| .$$

We also write, as is customary, $\pi(x)$ for the number of primes up to x and recall the Prime Number Theorem in the form

$$\pi(x) = \frac{x}{\log x} + O\left(\frac{x}{(\log x)^2}\right) . \tag{4}$$

Theorem 2 (Drmota/Mauduit/Rivat). *We have, uniformly for all integers $k \geq 0$ with $(k, b-1) = 1$,*

$$S(k,x) = \frac{b-1}{\phi(b-1)} \frac{\pi(x)}{\sqrt{2\pi s \log_b x}} \left(\exp\left(-\frac{(k-\mu\log_b x)^2}{2s\log_b x}\right) + O\left((\log x)^{-\frac{1}{2}+\epsilon}\right)\right) , \tag{5}$$

where $\epsilon > 0$ is arbitrary but fixed, and $\phi(n)$ is Euler's totient function.

Remarks. Of course, for all but the finitely many primes p satisfying $(p, b-1) > 1$ we have $(\sigma(p), b-1) = 1$, and this leads to the $(b-1)/\phi(b-1)$ factor appearing throughout this paper.

It immediately follows from this theorem that there are infinitely many primes whose sum of digits is also prime. A moment's thought then reveals that given any positive integer t there are infinitely many primes p whose sum of digits is prime and there are t different truncations of p having the same property. Moving on we can then deduce that there are uncountably many α (which form a dense set in the set of positive reals) such that for infinitely many n we have $[b^n\alpha] \in \mathcal{P}$. However, the form of the asymptotic formula (5) alerts us to the influence of irregularities in the distribution of primes in short intervals. Since, given the current state of knowledge (see [**1**]), we do not know that every interval of the form $[z, z + z^{\frac{1}{2}}]$ contains primes (let alone asymptotically the right number of primes) we may have intervals of x for which Theorem 2 gives no non-trivial formula for k taking any prime value. It should be noted that we would have this difficulty even if we knew the Riemann Hypothesis were true. Happily we need to deal only with a weighted sum that enables us to sidestep this difficulty in our proof. In any case, we could have appealed to results on sums of differences between primes (see [**12**], for example) to obtain an "almost-all" type result.

Another immediate consequence of the above theorem is that there are infinitely many primes with $\sigma(p)$ and $\sigma(\sigma(p))$ both prime also, and this can be iterated as often as we wish. We shall show elsewhere how it is possible to obtain a Mertens' type theorem for primes restricted in this way.

In the course of our working we shall need an upper bound for the number of primes in an interval whose sum of digits is also prime. In order to apply a sieve method we shall need the following result which is a strengthened version of the result in [**13**] obtained by applying one of the main theorems in [**3**].

THEOREM 3. *There exists a constant $\theta = \theta(b) > 0$ with the following property. Suppose that $N, D \geq 2$. Then we have*

$$\sum_{\substack{d \leq D \\ (d,b-1)=1}} \left| \sum_{\substack{p \leq N \\ \sigma(p) \equiv 0 \, (\mathrm{mod}\, d)}} 1 - \frac{\pi(N)}{d} \right| \ll ND(\log N)^3 \exp\left(-\frac{\theta \log N}{D^2}\right). \qquad (6)$$

We explain how this follows from the work in [**3**] in Section 7 below.

3. *Outline of the argument*

We first suppose that $\alpha \in [0, 1)$. The argument for any other interval of unit length follows *mutatis mutandis*. We then restrict our attention to those primes whose sum of digits is sufficiently close to the expected mean value. To do this we write $\nu = \mu\sigma(p)^{\frac{2}{3}}$ in the following, and put

$$\mathcal{P}(n) = \{p \in \mathcal{P} : |\sigma(p) - \mu n| < \nu, p < b^n\}.$$

We can safely neglect all other $p \in \mathcal{P}$ since, by the work of Kátai [**11**], for any $A \geq 1$ we have

$$\sum_{\substack{p \in \mathcal{P}, p < b^n \\ p \notin \mathcal{P}(n)}} 1 \ll_A \frac{b^n}{n^A}.$$

We write

$$\mathcal{B}_n = \bigcup_{p \in \mathcal{P}(n)} \left[\frac{p}{b^n}, \frac{p+1}{b^n}\right).$$

Let $\lambda(\mathcal{A})$ denote the Lebesgue measure of a real set $\mathcal{A}$. Write

$$F_N(\alpha) = |\{n \leq N : \alpha \in \mathcal{B}_n\}|, \quad V(N) = \sum_{n \leq N} \lambda(\mathcal{B}_n).$$

Our starting point for the proof of Theorem 1 is then the following which can be found as [**10**, Theorem 3].

Lemma 1. *Suppose that $V(\infty)$ diverges. Suppose also that there is a positive constant c_3 such that if $\mathcal{I}$ is any subinterval of $[0,1)$ then*

$$\lim_{N\to\infty} \frac{1}{V(N)} \sum_{n=1}^{N} \lambda(\mathcal{I} \cap \mathcal{B}_n) = \lambda(\mathcal{I}) \tag{7}$$

and

$$\limsup_{N\to\infty} \left(\sum_{n=1}^{N} \lambda(\mathcal{I} \cap \mathcal{B}_n) \right)^2 \left(\sum_{1\le j,k\le N} \lambda(\mathcal{B}_j \cap \mathcal{B}_k \cap \mathcal{I}) \right)^{-1} \ge c_3 \lambda(\mathcal{I}). \tag{8}$$

Then we have, for almost all real $\alpha \in [0,1)$,

$$\limsup_{N\to\infty} \frac{F_N(\alpha)}{V(N)} \ge 1 \quad \text{and} \quad \liminf_{N\to\infty} \frac{F_N(\alpha)}{V(N)} \le 1.$$

To complete the proof of Theorem 1 we thus need to establish (7) and (8) with the correct function $V(N)$.

4. *Results from Multiplicative Number Theory*

We require the following simple upper bound for primes in short intervals which follows from the more general Brun-Titchmarsh inequality in the form given by Montgomery and Vaughan [**14**, (1.12)].

Lemma 2. *We have, for all $x > 0, y > 1$, that*

$$\pi(x+y) - \pi(x) \le \frac{2y}{\log y}.$$

We also need to apply a simple upper bound sieve; the result we need may be deduced from [**7**, p.20 & p.144-5].

Lemma 3. *Let $D \ge 2, f \in \mathbb{N}$. Then there exists a sequence of reals λ_d with $|\lambda_d| \le 1$, and a function $\rho(n)$ such that*

$$\rho(n) \ge \begin{cases} 1, & \text{if } n > D^{\frac{1}{2}} \text{ is a prime;} \\ 0, & \text{otherwise;} \end{cases}$$

and there is an absolute constant K with

$$\rho(n) \le \frac{K}{\log D} \frac{f}{\phi(f)} + \sum_{\substack{d\le D \\ (d,f)=1}} \lambda_d \left(\sum_{d|n} 1 - \frac{1}{d} \right). \tag{9}$$

5. *Proof of* (7)

We now let q also denote a prime variable. Suppose $\mathcal{I} = [g,h) \subseteq [0,1)$ (the proof for other intervals follows similarly) and write

$$\mathcal{P}(n,g,h) = \{ p \in \mathcal{P} : |\sigma(p) - \mu n| < \nu, g b^n \le p < h b^n \}, \quad M = \mu \left(N + N^{\frac{2}{3}} \right).$$

Then

$$\sum_{n=1}^{N} \lambda(\mathcal{I} \cap \mathcal{B}_n) = \sum_{n \le N} b^{-n} |\mathcal{P}(n, g, h)| \; + \; O(1).$$

Here the $O(1)$ term comes from 'edge effects', that is the difference between $p < hb^n$ and $p + 1 < hb^n$. We then open out the inner term on the right hand side above and reverse the order of summation to give

$$\sum_{q \le M} \sum_{\substack{n \\ |n - q/\mu| < \nu/\mu}} \frac{1}{b^n} \sum_{\substack{gb^n \le p < hb^n \\ \sigma(p) = q}} 1 \; + \; O(1).$$

In the above ν is now a function of q: $\nu = \mu q^{\frac{2}{3}}$. Applications of (5) estimate the above sum to be

$$\frac{b-1}{\phi(b-1)} \sum_{q \le M} \sum_{\substack{n \le N \\ |n - q/\mu| < \nu/\mu}} \frac{1}{b^n} \frac{\pi(hb^n) - \pi(gb^n)}{\sqrt{2\pi s n}} \left(\exp\left(-\frac{(q - \mu n)^2}{2sn} \right) + O(n^{-\frac{1}{2} + \epsilon}) \right).$$

Here we have absorbed into the $O(n^{-\frac{1}{2}+\epsilon})$ expression those terms which arise from changing $\log_b(hb^n)$ or $\log_b(gb^n)$ into $\log_b(b^n) = n$. After an appeal to (4) (and absorbing similar errors arising from g, h as before) this becomes

$$(h - g) \frac{b-1}{\phi(b-1)} \sum_{q \le M} \sum_{\substack{n \le N \\ |n - q/\mu| < \nu/\mu}} \frac{1}{n\sqrt{2\pi s n} \log b} \left(\exp\left(-\frac{(q - \mu n)^2}{2sn} \right) + O(n^{-\frac{1}{2} + \epsilon}) \right).$$

We now concentrate on the inner sum over n above. We write η for an error term of size $O(q^{-\frac{1}{3}})$. We may replace the $n^{-3/2}$ term with $(q/\mu)^{-3/2}(1 + \eta)$. Taking $\epsilon = \frac{1}{6}$ the term $O(n^{-\frac{1}{2}+\epsilon})$ also becomes η. By the Mean Value Theorem it is quickly seen that

$$\exp\left(-\frac{(q - \mu n)^2}{2sn} \right) = \exp\left(-\frac{\mu (q - \mu n)^2}{2sq} \right) + \eta,$$

where we have noted that if $|q - \mu n| > n^{5/9}$ then the exponential term itself is $O(\eta)$. Since

$$\sum_{q \le M} \frac{\nu \eta}{q^{\frac{3}{2}}} \ll \sum_{q \le M} q^{-7/6},$$

which converges, we are left to estimate

$$\sum_{q \le M} \frac{\mu^{\frac{3}{2}}}{q\sqrt{2\pi s q} \log b} \sum_{\substack{n \le N \\ |n - q/\mu| < \nu/\mu}} \exp\left(-\frac{(q - \mu n)^2}{2sq} \right).$$

The inner sum can be replaced by an integral with a suitably small error:

$$\begin{aligned} \int_{(q-\nu)/\mu}^{(q+\nu)/\mu} \exp\left(-\frac{(q - \mu x)^2}{2sq} \right) dx (1 + \eta) &= \frac{\sqrt{2sq}}{\mu^{\frac{3}{2}}} \int_{-\infty}^{\infty} \exp(-y^2) \, dy (1 + \eta) \\ &= \frac{\sqrt{2\pi s q}}{\mu^{\frac{3}{2}}} (1 + \eta) \end{aligned}$$

We note that by Mertens' Theorem [**2**, p.56]

$$\sum_{q \le M} \frac{1}{q} = \log \log M + O(1) = \log \log N + O(1).$$

Assembling all our information so far, we thus have

$$\sum_{n=1}^{N} \lambda(\mathcal{I} \cap \mathcal{B}_n) = \frac{b-1}{\phi(b-1)} \frac{(h-g)\log\log N}{\log b} + O(1).$$

In particular, we have established (7) with

$$V(N) = \frac{b-1}{\phi(b-1)} \frac{\log\log N}{\log b} + O(1).$$

6. *Proof of* (8)

First we need an upper bound result corresponding to the work of the previous section without the averaging over n.

Lemma 4. *For $0 \le g, h < 1$ and $n \ge 2$ we have*

$$\sum_{\substack{gb^n < p < hb^n \\ p \in \mathcal{P}}} b^{-n} \le \frac{b-1}{\phi(b-1)} \frac{6(h-g)}{n \log n \log b} + O(n^{-\frac{7}{6}}).$$

Proof. We have, by two applications of Theorem 2,

$$\sum_{\substack{gb^n < p < hb^n \\ p \in \mathcal{P}}} b^{-n} = \sum_{\substack{q \\ |q - n\mu| < \nu}} \frac{b-1}{\phi(b-1)} \frac{h-g}{n\sqrt{2\pi s n}\log b} \left(\exp\left(-\frac{(q-\mu n)^2}{2sn} \right) + O(n^{-\frac{1}{3}}) \right).$$

We divide the summation range over q into subranges of length $n^{\frac{1}{3}}$. By Lemma 2 the number of primes in such an interval does not exceed $6n^{\frac{1}{3}}/(\log n)$. The proof may then be completed in a similar manner to the working of the last section. □

Write, for $m < n$,

$$\mathcal{A}(m,n) = \{(p,q) \in \mathcal{P}^2 : b^{n-m}p - 1 < q < (p+1)b^{n-m}, gb^m < p < hb^m\}.$$

Then

$$\sum_{1 \le j,k \le N} \lambda(\mathcal{B}_j \cap \mathcal{B}_k \cap \mathcal{I}) \le V(N)(h-g) + O(1) + 2 \sum_{1 \le m < n \le N} b^{-n} |\mathcal{A}(m,n)|.$$

We split the range of summation over n above into two parts:
(i) $m < n \le m \log m$;
(ii) $n > m \log m$.
For case (i) we apply Lemma 2 and so make no use of the restriction $q \in \mathcal{P}$. We thus obtain an upper bound for this portion of the sum which is

$$\begin{aligned} &\le 2 \sum_{(i)} \sum_{\substack{gb^m < p < hb^m \\ p \in \mathcal{P}}} b^{-n} \frac{b^{n-m}+1}{(n-m)\log b} \\ &\le 12 \frac{b-1}{\phi(b-1)} \sum_{(i)} \frac{h-g}{(\log b)^2 (m \log m)(n-m)} + O(1) \end{aligned}$$

after an application of Lemma 4. A simple calculation shows that this term is $\ll (h-g)V(N)^2$.

For case (ii) we require Lemma 3 where we take $D = (n/m)^{\frac{1}{4}}, f = b$. The first term on the right hand side of (9) is then

$$\begin{aligned}
&\ll \sum_{(ii)} \sum_{\substack{gb^m<p<hb^m \\ p\in\mathcal{P}}} b^{-n}\frac{b^{n-m}+1}{n\log(n/m)} \\
&\ll (h-g)\sum_{(ii)} \frac{1+o(1)}{(m\log m)(n\log(n/m))},
\end{aligned}$$

with an application of Lemma 4. This term is again clearly $\ll (h-g)V(N)^2$. The remaining term in (9) leads to the expression

$$S^* = \sum_{(ii)} \sum_{gb^m<p<hb^m} b^{-n} \sum_{b^{n-m}p-1<q<(p+1)b^{n-m}} \sum_{\substack{d\le D \\ (d,b-1)=1}} \lambda_d \left(\sum_{d|\sigma(q)} 1 - \frac{1}{d}\right).$$

Applications of (6) estimate this to be

$$\ll \sum_{(ii)} \sum_{gb^m<p<hb^m} b^{-n}n^3b^nD\exp\left(-\frac{\theta n}{D^2}\right).$$

We note that $\exp(-\theta nD^{-2}) \le \exp(-\theta n^{\frac{1}{2}}m^{\frac{1}{2}}) \ll b^{-m}n^{-5}$ since $n > m\log m$. It follows that $S^* \ll (h-g)V(N)$ which is a suitable upper bound.

Combining all the above results we obtain

$$\sum_{1\le j,k\le N} \lambda(\mathcal{B}_j \cap \mathcal{B}_k \cap \mathcal{I}) \ll (h-g)V(N)^2$$

where the implied constant only depends on b. This, with (7), establishes (8) and so completes the proof.

7. *Proof of* (6)

Our remaining task is to explain how (6) can be deduced from the work in [**3**]. The following is Proposition 2.1 from that paper. Here we write $e(\alpha) = \exp(2\pi i\alpha)$ and $\|\beta\| = \min_{n\in\mathbb{Z}} |\beta - n|$.

PROPOSITION 1. *For $b \ge 2, x \ge 2$ and $\alpha \in \mathbb{R}$ there exists a real number θ depending only on b such that*

$$\sum_{p\le x} e(\alpha\sigma(p)) \ll x(\log x)^3 \exp\left(-\theta(\log x)\|(b-1)\alpha\|^2\right)$$

where the implied constant depends only on b.

The proof may now be easily completed since

$$\begin{aligned}
\sum_{\substack{p\le N \\ \sigma(p)\equiv 0 \pmod d}} 1 &= \frac{1}{d}\sum_{\ell=0}^{d-1}\sum_{p\le N} e\left(\frac{\ell}{d}\sigma(p)\right) \\
&= \frac{\pi(N)}{d} + \frac{1}{d}\sum_{\ell=1}^{d-1} O\left(N(\log N)^3\exp\left(-\theta\log N\|(b-1)\ell/d\|^2\right)\right) \\
&= \frac{\pi(N)}{d} + O\left(N(\log N)^3\exp\left(-\frac{\theta\log N}{d^2}\right)\right),
\end{aligned}$$

where we used $(d, b-1) = 1$ to deduce that $\|(b-1)\ell/d\|^2 \geq d^{-2}$.

Acknowledgements. The author would like to express his thanks to the referee for pointing out some typos in the original paper, and to Christian Elsholtz for his e-mail correspondence which triggered this research.

References

1. R. C. Baker, G. Harman and J. Pintz, 'The difference betweeen consecutive primes, II', *Proc. London Math. Soc.* (3) 83 (2001) 532–562.
2. H. Davenport, *Multiplicative Number Theory*, 2nd edn, revised by H.L. Montgomery (Springer, New York 1980).
3. M. Drmota, C. Mauduit and J. Rivat, 'Primes with an average sum of digits', *Compositio Mathematica* 145 (2009) 271–292.
4. E. Fouvry and C. Mauduit, 'Méthodes de crible et fonctions sommes des chiffres', *Acta Arith.* 77 (1996) 339–351.
5. E. Fouvry and C. Mauduit, 'Sommes des chiffres et nombres presque premiers', *Math. Ann.* 305 (1996) 571–599.
6. A.O. Gelfond, 'Sur les nombres qui ont des propriétés additives et multiplicatives donées', *Acta Arith.* 13 (1968) 259–261.
7. G.R.H.Greaves, *Sieves in Number Theory* (Springer-Verlag Berlin Heidelberg 2001).
8. G. Harman, 'Metrical theorems on prime values of the integer parts of real sequences', *Proc. London Math. Soc.* (3) 75 (1997) 481–496.
9. G. Harman, *Metric Number Theory* LMS Monographs New Series 18 (Oxford 1998).
10. G. Harman, 'Variants of the Second Borel-Cantelli Lemma and their Applications in Metric Number Theory', in *Number Theory* 121-140, Trends Math., Birkhauser, Basel, 2000.
11. I. Kátai, 'On the sum of digits of primes', *Acta Math. Acad. Sci. Hungar.* 30 (1977) 169–173.
12. K. Matomäki, 'Large differences between consecutive primes', *Quart. J. Math.* 58 (2007) 489–517.
13. C. Mauduit and J. Rivat, 'Sur un problème de Gelfond: la somme des chiffres des nombres premiers', *Ann. of Math.* 171 (2010) 1591–1646.
14. H. L. Montgomery and R. C. Vaughan, 'The large sieve', *Mathematika* 20 (1973) 119–134.

Department of Mathematics,
Royal Holloway, University of London,
EGHAM,
Surrey TW20 0EX

g.harman@rhul.ac.uk