Journal of Integer Sequences, Vol. 15 (2012),
Article 12.2.2

Counting Primes whose Sum of Digits is Prime

Glyn Harman
Department of Mathematics
Royal Holloway, University of London
Egham
Surrey TW20 0EX
United Kingdom
g.harman@rhul.ac.uk

Abstract

Motivated by recent work of Drmota, Mauduit and Rivat, we discuss the possibility of counting the number of primes up to x whose sum of digits is also prime. We show that, although this is not possible unless we assume a hypothesis on the distribution of primes stronger than what is implied by the Riemann hypothesis, we can establish a Mertens-type result. That is, we obtain a formula for the number of such primes p up to x weighted with a factor $1/p$. Indeed, we can iterate the method and count primes with the sum of digits a prime whose sum of digits is a prime, and so on.

1 Introduction

Thanks to the revolutionary work of Mauduit and Rivat [7], in particular as developed in [2], our knowledge of the distribution of primes with constraints on their sum of digits has undergone a substantial expansion in recent years. The question of counting the number of primes up to x, whose sum of digits in base b is also prime, would have been considered far too difficult ten years ago in view of the difficulty in restricting the sum of digits function when its variable is prime. In this paper we discuss this question, showing that it is still impossibly difficult at present, but this is now a result of our lack of knowledge concerning the distribution of primes in short intervals, and no longer caused by the problems of restricting the sum of digits function. However, we are able to prove a Mertens-type result on counting primes whose sum of digits is prime. Throughout this paper the letters p and q are reserved for prime variables. In 1874 Mertens proved that

$$\sum_{p<X} \frac{1}{p} = \log\log X + A + O((\log x)^{-1}),$$

for a certain constant A (see [1, p. 56]). In this paper we will obtain analogues of this result for primes whose sum of digits written in base b ($b \geq 2$) are also prime. For brevity we shall write this set as $\mathcal{P}_b$. If

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

for the sum of digits function in base b. In fact we shall be able to iterate our working and establish Mertens-type theorems for primes belonging to sets defined as follows:

$$\mathcal{P}_b^{(1)} = \mathcal{P}_b, \qquad \mathcal{P}_b^{(j)} = \{p \in \mathcal{P}_b : \sigma_b(p) \in \mathcal{P}_b^{(j-1)}\},\ j = 2, 3 \ldots .$$

We write $\log_b x$ for the logarithm to base b, $\log x$ for the natural logarithm, and define $L_j(x)$ inductively by

$$L_1(x) = \max(1, \log x), \qquad L_j(x) = L_1(L_{j-1}(x)) \ \text{ for } j \geq 2.$$

Our starting point will be a remarkable theorem of Drmota, Mauduit and Rivat [2]. Before stating the result precisely we need the following notation:

$$s_b = \frac{b^2-1}{12}, \quad \mu_b = \tfrac{1}{2}(b-1), \quad S(k,x) = \left|\{p \leq x : \sigma_b(p) = k\}\right| .$$

As usual we write $\phi(n)$ for Euler's totient function, and $\pi(x)$ for the number of primes up to x. We recall the prime number theorem in the form

$$\pi(x) = \frac{x}{\log x} + O\left(\frac{x}{(\log x)^2}\right) . \tag{1}$$

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

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

where $\epsilon > 0$ is arbitrary but fixed.

Since $\sigma_b(p) \equiv p \,(\operatorname{mod} b-1)$, for all but the finitely many primes p satisfying $(p, b-1) > 1$ we have $(\sigma_b(p), b-1) = 1$, and this leads to the $(b-1)/\phi(b-1)$ factor appearing frequently. From the above result we can deduce some very interesting corollaries. Clearly, (2) shows that every sufficiently large integer is the sum of digits of a prime number. So, in particular, there are infinitely many primes whose sum of digits is also prime. Also, every sufficiently large prime is the sum of digits of another prime which in turn is the sum of digits of another prime, and so on. We use Theorem 1 to obtain a Mertens-type result as follows.

Theorem 2. *Let $b \geq 2$ be an integer and $X \geq 2$. Then we have*

$$\sum_{\substack{p<X \\ p \in \mathcal{P}_b}} \frac{1}{p} = \frac{b-1}{\phi(b-1)} L_3(X) + O(1). \tag{3}$$

In fact we shall use Theorem 2 as the first step in an inductive argument to establish the following more general result for the sets $\mathcal{P}_b^{(j)}$.

Theorem 3. *Let $b, j \geq 2$ be integers and $X \geq 2$. Then we have*

$$\sum_{\substack{p<X \\ p \in \mathcal{P}_b^{(j)}}} \frac{1}{p} = \left(\frac{b-1}{\phi(b-1)}\right)^j L_{j+2}(X) + O(1). \tag{4}$$

As a final application of Theorem 1 we leave the following for the reader to prove.

Theorem 4. *Write $\mathcal{P}_b^*$ for the set of primes in $\mathcal{P}_b$ which have a prime number of digits. Then*

$$\sum_{\substack{p<X \\ p \in \mathcal{P}_b^*}} \frac{L_2(p)}{p} = \frac{b-1}{\phi(b-1)} L_3(X) + O(1).$$

2 The problem of the unweighted sum

In this section we briefly consider the sum

$$\pi_b^*(X) = \sum_{\substack{p<X \\ p \in \mathcal{P}_b}} 1\,.$$

First we note the following result of Kátai [5].

Lemma 5. *For every positive integer k we have*

$$\sum_{p<x} |\sigma_b(p) - \mu_b \log_b x|^k \ll_k x(\log x)^{k/2-1}.$$

Now, even assuming the Riemann hypothesis and the pair correlation conjecture [3] we cannot prove that every interval $\mathcal{I}_y = [y - y^{\frac{1}{2}}(\log y)^{\frac{1}{3}}, y + y^{\frac{1}{2}}(\log y)^{\frac{1}{3}}]$ contains a prime number. Suppose that $\mathcal{I}_y$ contains no primes, and choose X so that $y = \mu_b \log_b X$. Then Theorem 1 cannot supply an asymptotic formula for $\pi_b^*(X)$, and by Lemma 5 with $k = 3A$

$$\pi_b^*(X) \ll_A \frac{X}{XL_1(X)(L_2(X))^A}$$

for every positive integer A. This shows that no asymptotic formula is possible unless a stronger result than can be obtained on the Riemann hypothesis is assumed, for clearly $\pi_b^*(X)$ should be of size $X(L_1(X)L_2(X))^{-1}$. We state here a hypothetical result to illustrate what we mean.

Theorem 6. *Suppose that there exists a positive function $f(x) \to 0$ such that for all large x every interval of the form $[x, x + x^{\frac{1}{2}} f(x)]$ contains $f(x)x^{\frac{1}{2}}(1 + o(1))(\log x)^{-1}$ primes. Then*

$$\pi_b^*(X) \sim \frac{b-1}{\phi(b-1)} \frac{X}{L_1(X)L_2(X)}. \tag{5}$$

The proof of this result is straightforward. By Lemma 5 we need only count primes whose sum of digits is q with

$$|q - \mu_b \log_b x| < (\log_b x)^{\frac{2}{3}}. \tag{6}$$

Applying Theorem 1 we need to estimate

$$\sum_{(6)} \exp\left(-\frac{(q - \mu_b \log_b x)^2}{2s_b \log_b x}\right).$$

Breaking the summation range into shorter ranges of the form $[y, y + y^{\frac{1}{2}} f(y)]$ and then comparing a sum with an integral we obtain

$$\frac{1 + o(1)}{\log(\mu_b \log_b x)} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2s_b \log_b x}\right) dt = \frac{1 + o(1)}{\log(\mu_b \log_b x)} \sqrt{2s_b \pi \log_b x},$$

which quickly establishes the result.

We remark that the hypothesis of Theorem 6 is "almost always" true unconditionally in the following sense. The method of Heath-Brown [4] shows that, given $\epsilon > 0$, there exists $\eta > 0$ such that for all integers $y \in [Y, 2Y)$, with at most $\ll Y^{\frac{3}{4}+\epsilon}$ exceptions, every subinterval $[z, z + z^{\frac{1}{2}-\eta}]$ of $[y - Y^{\frac{1}{2}+\epsilon}, y + Y^{\frac{1}{2}+\epsilon}]$ contains $z^{\frac{1}{2}-\eta}(1 + O((\log z)^{-1}))(\log z)^{-1}$ primes. For these non-exceptional y and with X chosen to satisfy $y = \mu \log_b X$ we then have (5). Indeed, we can obtain the expected asymptotic formula for X on quite long ranges. We note that we are unable to use the stronger results obtainable using a sieve method [6] since we require an asymptotic formula for the number of primes in an interval rather than upper or lower bounds.

3 Proof of Theorem 2

Lemma 5, together with (2), shows that $\sigma_b(p)$ is concentrated near its mean value $\mu_b \log_b p$, and so enables us to restrict our attention to certain $p \in \mathcal{P}_b$ with negligible error. We write

$$M(m) = \exp\left((\log b) m^{\frac{7}{8}}\right), \qquad \mu(m) = \mu_b m^{\frac{7}{8}},$$
$$\epsilon(m) = \frac{M(m+1) - M(m)}{M(m)}, \qquad \delta(m) = (m+1)^{\frac{7}{8}} - m^{\frac{7}{8}},$$

and put $\mathcal{M}(m) = [M(m), M(m+1))$. We note for future reference that

$$\epsilon(m) = O\left(m^{-\frac{1}{8}}\right) \quad \text{and} \quad \epsilon(m) = \delta(m)\left(\log b + O\left(m^{-\frac{1}{8}}\right)\right). \tag{7}$$

Let V be the largest integer with $M(V) < X$. Also, write

$$\mathcal{P}(m) = \{p \in \mathcal{P}_b : \sigma_b(p) = q, |q - \mu(m)| < q^{\frac{2}{3}}\}.$$

We remark that the exponents $\frac{7}{8}, \frac{2}{3}$ above are chosen as convenient fixed numbers between $\frac{1}{2}$ and 1 (which must satisfy certain inequalities to enable certain sums to converge later) and have no other special significance. From Lemma 5 we can deduce that

$$\sum_{m=1}^{\infty} \sum_{\substack{p \in \mathcal{M}(m) \\ p \in \mathcal{P}_b, p \notin \mathcal{P}(m)}} \frac{1}{p}$$

converges. Hence

$$S(X) := \sum_{\substack{p<X \\ p \in \mathcal{P}_b}} \frac{1}{p} = \sum_{m \leq V} \frac{1}{M(m)} \sum_{\substack{p \in \mathcal{M}(m) \\ p \in \mathcal{P}(m)}} 1 + O(1).$$

Now write $Q = \log_b X$. Then

$$S(X) = \sum_{q<Q} \sum_{\substack{m \leq V \\ |q-\mu(m)|<q^{\frac{2}{3}}}} \frac{1}{M(m)} \sum_{\substack{p \in \mathcal{M}(m) \\ \sigma_b(p)=q}} 1 + O(1).$$

After applications of Theorem 1 with $\epsilon = 1/14$ and using (1) this becomes

$$\sum_{q<Q} \sum_{\substack{m \leq V \\ |q-\mu(m)|<q^{\frac{2}{3}}}} \frac{b-1}{\phi(b-1)} \frac{\epsilon(m)}{(\log b) m^{\frac{7}{8}} \sqrt{2\pi s_b m^{\frac{7}{8}}}} \left(\exp\left(-\frac{(q-\mu(m))^2}{2 s_b m^{\frac{7}{8}}} \right) + O(m^{-\frac{1}{4}}) \right). \tag{8}$$

Now the $O(m^{-\frac{1}{4}})$ term in the above leads to an error (using $m = (q/\mu_b)^{\frac{8}{7}} + O(q^{\frac{8}{7}-\frac{1}{3}})$)

$$\ll \sum_{q<Q} q^{\frac{8}{7}-\frac{1}{3}} q^{-\frac{1}{7}} q^{-1} q^{-\frac{1}{2}} q^{-\frac{2}{7}} = O(1),$$

and so we can concentrate on the main term in (8). By (7) we can replace $\epsilon(m)$ with $\delta(m) \log b$ and only incur an error which is $O(m^{-\frac{1}{8}})$ times the main term. We can ignore this error as it will prove to be negligible. By using $m^{\frac{7}{8}} = q/\mu_b + O(q^{\frac{2}{3}})$ we can reduce the sum to

$$\frac{b-1}{\phi(b-1)} \frac{1}{\sqrt{2\pi s_b}} \sum_{q<Q} (\mu_b/q)^{\frac{3}{2}} \Sigma_q + O(1),$$

with

$$\Sigma_q = \sum_{\substack{m \leq V \\ |q-\mu(m)|<q^{\frac{2}{3}}}} \delta(m) \exp\left(-\frac{\mu_b (q-\mu(m))^2}{2 s_b q} \right).$$

We can identify this as a Riemann sum of an integral and so obtain

$$\begin{aligned}\Sigma_q &= \int_{q/\mu_b - q^{\frac{2}{3}}}^{q/\mu_b + q^{\frac{2}{3}}} \exp\left(-\frac{\mu_b(q-\mu_b x)^2}{2s_b q}\right) dx + O(1)\\ &= \frac{\sqrt{2s_b q\pi}}{\mu_b^{\frac{3}{2}}} + O(1).\end{aligned}$$

We have thus shown that

$$S(X) = \frac{b-1}{\phi(b-1)} \sum_{q<Q} \frac{1}{q} + O(1) = \frac{b-1}{\phi(b-1)} L_3(X) + O(1) \tag{9}$$

as required to complete the proof.

4 Proof of Theorem 3

We prove Theorem 3 by induction; Theorem 2 is the case $j = 1$. Now assume that the case j has been proved. If we follow the working in §3, then in place of (9) we obtain

$$\sum_{\substack{p<X\\ p\in\mathcal{P}_b^{(j+1)}}} \frac{1}{p} = \frac{b-1}{\phi(b-1)} \sum_{\substack{p<Y\\ p\in\mathcal{P}_b^{(j)}}} \frac{1}{p} + O(1)$$

where $Y = \log_b X$. By the inductive hypothesis we therefore have

$$\sum_{\substack{p<X\\ p\in\mathcal{P}_b^{(j+1)}}} \frac{1}{p} = \left(\frac{b-1}{\phi(b-1)}\right)^{j+1} L_{j+3}(X) + O(1)$$

as required to complete the proof.

5 Acknowledgement

I would like to thank the referee for his suggestions and Christian Elsholtz whose e-mail conversation sparked my interest in this topic.

References

[1] H. Davenport, *Multiplicative Number Theory*, 2nd edition, revised by H. L. Montgomery, Springer, 1980.

[2] M. Drmota, C. Mauduit and J. Rivat, Primes with an average sum of digits, *Compositio Math.* **145** (2009), 271–292.

[3] D. Goldston and D. R. Heath-Brown, A note on the difference between consecutive primes, *Math. Ann.* **266** (1984), 317–320.

[4] D. R. Heath-Brown, The differences between consecutive primes, III, *J. London Math. Soc.* (2), **20**, 177–178.

[5] I. Kátai, On the sum of digits of primes, *Acta Math. Acad. Sci. Hungar.* **30** (1977), 169–173.

[6] K. Matomäki, Large differences between consecutive primes, *Quart. J. Math.* **58** (2007), 489–517.

[7] C. Mauduit and J. Rivat, Sur un problème de Gelfond: la somme des chiffres des nombres premiers, *Ann. of Math.* **171** (2010), 1591–1646.

2010 *Mathematics Subject Classification*: Primary 11A63; Secondary 11N05, 11N60.
Keywords: prime number, sum of digits, Mertens' theorem.

(Concerned with sequences A046704, A070027, and A109981.)

Received September 16 2011; revised version received December 29 2011. Published in *Journal of Integer Sequences*, December 30 2011.

Return to Journal of Integer Sequences home page.