ON THE NUMBER OF NIVEN NUMBERS UP TO x

Jean-Marie DeKoninck[1]

Département de Mathématiques et de statistique, Université Laval, Québec G1K 7P4, Canada
e-mail: jmdk@mat.ulaval.ca

Nicolas Doyon

Département de Mathématiques et de statistique, Université Laval, Québec G1K 7P4, Canada

(Submitted June 2001)

1. INTRODUCTION

A positive integer n is said to be a *Niven number* (or a Harshad number) if it is divisible by the sum of its (decimal) digits. For instance, 153 is a Niven number since 9 divides 153, while 154 is not.

Let $N(x)$ denote the number of Niven numbers $\leq x$. Using a computer, one can obtain the following table:

x	$N(x)$
10	10
100	33
1000	213

x	$N(x)$
10^4	1538
10^5	11872
10^6	95428

x	$N(x)$
10^7	806095
10^8	6954793
10^9	61574510

It has been established by R.E. Kennedy & C.N. Cooper [4] that the set of Niven numbers is of zero density, and later by I. Vardi [5] that, given any $\varepsilon > 0$

$$N(x) \ll \frac{x}{(\log x)^{1/2-\varepsilon}}. \tag{1}$$

We have not found in the literature any lower bound for $N(x)$, although I. Vardi [5] has obtained that there exists a positive constant α such that

$$N(x) > \alpha \frac{x}{(\log x)^{11/2}} \tag{2}$$

for infinitely many integers x, namely for all sufficiently large x of the form $x = 10^{10k+n+2}$, k and n being positive integers satisfying $10^n = 45k+10$. Even though inequality (2) most likely holds for all sufficiently large x, it has not yet been proved. More recent results concerning Niven numbers have been obtained (see for instance H.G. Grundman [3] and T. Cai [1]).

Our goal is to provide a non trivial lower bound for $N(x)$ and also to improve on (1). Hence we shall prove the following result.

Theorem: *Given any $\varepsilon > 0$, then*

$$x^{1-\varepsilon} \ll N(x) \ll \frac{x \log\log x}{\log x}. \tag{3}$$

We shall further give a heuristic argument which would lead to an asymptotic formula for $N(x)$, namely $N(x) \sim c\frac{x}{\log x}$, where

[1] Research supported in part by a grant from CRSNG.

$$c = \frac{14}{27}\log 10 \approx 1.1939. \tag{4}$$

2. THE LOWER BOUND FOR $N(x)$

We shall establish that given any $\varepsilon > 0$, there exists a positive real number $x_0 = x_0(\varepsilon)$ such that

$$N(x) > x^{1-\varepsilon} \quad \text{for all } x \geq x_0. \tag{5}$$

Before we start the proof of this result, we introduce some notation and establish two lemmas.

Given a positive integer $n = [d_1, d_2, \dots, d_k]$, where $d_1, d_2, \dots, d_k$ are the (decimal) digits of n, we set $s(n) = \sum_{i=1}^{k} d_i$. Hence n is a Niven number if $s(n)|n$. For convenience we set $s(0) = 0$.

Further let H stand for the set of positive integers h for which there exist two non negative integers a and b such that $h = 2^a \cdot 10^b$. Hence

$$H = \{1, 2, 4, 8, 10, 16, 20, 32, 40, 64, 80, 100, 128, 160, 200, 256, 320, 400, 512, 640, \dots\}.$$

Now given a positive integer n, define $h(n)$ as the largest integer $h \in H$ such that $h \leq n$. For instance $h(23) = 20$ and $h(189) = 160$.

Lemma 1: *Given $\varepsilon > 0$, there exists a positive integer n_0 such that $\frac{n}{h(n)} < 1+\varepsilon$ for all $n \geq n_0$.*

Proof: Let $\varepsilon > 0$ and assume that $n \geq 2$. First observe that

$$\frac{n}{h(n)} < 1 + \varepsilon \Longleftrightarrow \log n - \log h(n) < \log(1+\varepsilon) := \varepsilon_1,$$

say. It follows from classical results on approximation of real numbers by rational ones that there exist two positive integers p and q such that

$$0 < \delta := p \log 10 - q \log 2 < \varepsilon_1. \tag{6}$$

For each integer $n \geq 2$, define

$$r := \left[\frac{\log n}{\log 2}\right] \quad \text{and} \quad t := \left[\frac{\log n - r\log 2}{\delta}\right]. \tag{7}$$

From (6) and (7), it follows that

$$\log n - (r\log 2 + t(p\log 10 - q\log 2)) < \delta < \varepsilon_1,$$

that is

$$\frac{n}{2^{r-qt} \cdot 10^{tp}} < 1 + \varepsilon.$$

In order to complete the proof of Lemma 1, it remains to establish that $2^{r-qt} \cdot 10^{tp} \in H$, that is that $r - qt \geq 0$. But it follows from (7) that

$$t \leq \frac{\log n - r \log 2}{\delta} \leq \frac{\log n}{\delta} - \frac{\log 2}{\delta}\left(\frac{\log n}{\log 2} - 1\right) = \frac{\log 2}{\delta},$$

so that

$$r - qt \geq r - \frac{q \log 2}{\delta} = \left[\frac{\log n}{\log 2}\right] - \frac{q \log 2}{\delta} > \frac{\log n}{\log 2} - \frac{q \log 2}{\delta} - 1,$$

a quantity which will certainly be positive if n is chosen to satisfy

$$\frac{\log n}{\log 2} \geq \frac{q \log 2}{\delta} + 1,$$

that is

$$n \geq n_0 := \left[2^{(q \log 2)/\delta + 1}\right] + 1.$$

Noting that q and δ depend only on ε, the proof of Lemma 1 is complete.

Given two non negative integers r and y, let

$$M(r, y) := \#\{0 \leq n < 10^r : s(n) = y\}. \tag{8}$$

For instance $M(2, 9) = 10$. Since the average value of $s(n)$ for $n = 0, 1, 2, \ldots, 10^r - 1$ is $\frac{9}{2}r$, one should expect that, given a positive integer r, the expression $M(r, y)$ attains its maximal value at $y = [\frac{9}{2}r]$. This motivates the following result.

Lemma 2: *Given any positive integer r, one has*

$$M(r, [4.5r]) \geq \frac{10^r}{9r + 1}.$$

Proof: As n runs through the integers $0, 1, 2, 3, \ldots, 10^r - 1$, it is clear that $s(n)$ takes on $9r+1$ distinct values, namely $0, 1, 2, 3, \ldots, 9r$. This implies that there exists a number $y = y(r)$ such that $M(r, y) \geq \frac{10^r}{9r+1}$. By showing that the function $M(r, y)$ takes on its maximal value when $y = [4.5r]$, the proof of Lemma 2 will be complete. We first prove:

(a) If r is even, $M(r, 4.5r + y) = M(r, 4.5r - y)$ for $0 \leq y \leq 4.5r$; if r is odd, $M(r, 4.5r + y + 0.5) = M(r, 4.5r - y - 0.5)$ for $0 \leq y < 4.5r$;

(b) if $y < 4.5r$, then $M(r, y) \leq M(r, y + 1)$.

To prove (a), let

$$z = \begin{cases} 4.5r + y & \text{if } r \text{ is even,} \\ 4.5r + y + 0.5 & \text{if } r \text{ is odd,} \end{cases} \tag{9}$$

and consider the set K of non negative integers $k < 10^r$ such that $s(k) = z$ and the set L of non negative integers $\ell < 10^r$ such that $s(\ell) = 9r - z$. Observe that the function $\sigma : K \to L$ defined by

$$\sigma(k) = \sigma([d_1, d_2, \ldots, d_r]) = [9 - d_1, 9 - d_2, \ldots, 9 - d_r]$$

is one-to-one. Note that here, for convenience, if n has t digits, $t < r$, we assume that n begins with a string of $r - t$ zeros, thus allowing it to have r digits. It follows from this that $|K| = |L|$ and therefore that

$$M(r, z) = M(r, 9r - z). \tag{10}$$

Combining (9) and (10) establishes (a).

To prove (b), we proceed by induction on r. Since $M(1, y) = 1$ for $0 \le y \le 9$, it follows that (b) holds for $r = 1$.

Now given any integer $r \ge 2$, it is clear that

$$M(r, y) = \sum_{i=0}^{9} M(r - 1, y - i),$$

from which it follows immediately that

$$M(r, y + 1) - M(r, y) = M(r - 1, y + 1) - M(r - 1, y - 9). \tag{11}$$

Hence to prove (b) we only need to show that the right hand side of (11) is non negative. Assuming that y is an integer smaller than $4.5r$, we have that $y \le 4.5r - 0.5 = 4.5(r - 1) + 4$ and hence $y = 4.5(r - 1) + 4 - j$ for some real number $j \ge 0$ (actually an integer or half an integer). Using (a) and the induction argument, it follows that $M(r - 1, y + 1) - M(r - 1, y - 9) \ge 0$ holds if $|4.5(r - 1) - (y + 1)| \le |4.5(r - 1) - (y - 9)|$. Replacing y by $4.5(r - 1) + 4 - j$, we obtain that this last inequality is equivalent to $|j - 5| \le |j + 5|$, which clearly holds for any real number $j \ge 0$, thus proving (b) and completing the proof of Lemma 2.

We are now ready to establish the lower bound (5). In fact, we shall prove that given any $\varepsilon > 0$, there exists an integer r_0 such that

$$N\left(10^{r(1+\varepsilon)}\right) > 10^{r(1-\varepsilon)} \quad \text{for all integers } r \ge r_0. \tag{12}$$

To see that this statement is equivalent to (5), it is sufficient to choose $x_0 > 10^{r_0(1+\varepsilon)}$. Indeed, by doing so, if $x \ge x_0$, then

$$10^{r(1+\varepsilon)} \le x \le 10^{(r+1)(1+\varepsilon)} \quad \text{for a certain } r \ge r_0,$$

in which case

$$N(x) \ge N\left(10^{r(1+\varepsilon)}\right) > 10^{r(1-\varepsilon)},$$

and since $x \le 10^{(r+1)(1+\varepsilon)}$, we have

$$x^{\frac{r(1-\varepsilon)}{(r+1)(1+\varepsilon)}} \le 10^{r(1-\varepsilon)} < N(x),$$

that is

$$x^{1-\varepsilon_1} \leq 10^{r(1-\varepsilon)} < N(x),$$

for some $\varepsilon_1 = \varepsilon_1(r, \varepsilon)$ which tends to 0 as $\varepsilon \to 0$ and $r \to \infty$.

It is therefore sufficient to prove the existence of a positive integer r_0 for which (12) holds.

First for each integer $r \geq 1$, define the non negative integers $a(r)$ and $b(r)$ implicitly by

$$2^{a(r)} \cdot 10^{b(r)} = h([4.5r]). \tag{13}$$

We shall now construct a set of integers n satisfying certain conditions. First we limit ourselves to those integers n such that $s(n) = 2^{a(r)} \cdot 10^{b(r)}$. Such integers n are divisible by $s(n)$ if and only if their last $a(r) + b(r)$ digits form a number divisible by $2^{a(r)} \cdot 10^{b(r)}$. Hence we further restrict our set of integers n to those for which the (fixed) number v formed by the last $a(r) + b(r)$ digits of n is a multiple of $s(n)$.

Finally for the first digit of n, we choose an integer d, $1 \leq d \leq 9$, in such a manner that

$$2^{a(r)} \cdot 10^{b(r)} - s(v) - d \equiv 0 \pmod 9. \tag{14}$$

Thus let n be written as the concatenation of the digits of d, u and v, which we write as $n = [d, u, v]$, where u is yet to be determined. Clearly such an integer n shall be a Niven number if $d + s(u) + s(v) = s(n) = 2^{a(r)} \cdot 10^{b(r)}$, that is if $s(u) = 2^{a(r)} \cdot 10^{b(r)} - d - s(v)$. We shall now choose u among those integers having at most $\beta := \frac{2^{a(r)} \cdot 10^{b(r)} - d - s(v)}{4.5}$ digits. Note that β is an integer because of condition (14).

Now Lemma 2 guarantees that there are at least $\frac{10^\beta}{9\beta+1}$ possible choices for u.

Let us now find upper and lower bounds for β in terms of r.

On one hand, we have

$$\beta = \frac{h([4.5r]) - d - s(v)}{4.5} < \frac{h([4.5r])}{4.5} \leq r. \tag{15}$$

On the other hand, recalling (13), we have $s(v) < 9(a(r) + b(r)) < 9\frac{\log h([4.5r])}{\log 2}$, and therefore

$$\beta = \frac{h([4.5r]) - d - s(v)}{4.5} > \frac{h([4.5r]) - 9 - 9\frac{\log h([4.5r])}{\log 2}}{4.5}. \tag{16}$$

Using Lemma 1, we have that, if r is large enough, $h([4.5r]) > 4.5r(1 - \varepsilon/2)$. Hence it follows from (16) that

$$\beta > \frac{4.5r(1 - \varepsilon/2) - 9 - 9\frac{\log h([4.5r])}{\log 2}}{4.5} > r(1 - \varepsilon), \tag{17}$$

provided r is sufficiently large, say $r \geq r_1$.

Again using (13), we have that

$$a(r) + b(r) + 1 < \frac{\log(h[4.5r])}{\log 2} + 1.$$

Since $h(n) \leq n$, and choosing r sufficiently large, say $r \geq r_2$, it follows from this last inequality that

$$a(r) + b(r) + 1 < \frac{\log(4.5r)}{\log 2} + 1 < r\varepsilon \quad (r \geq r_2).$$

Combining this inequality with (15), we have that

$$\beta + a(r) + b(r) + 1 < r(1+\varepsilon) \quad (r \geq r_2). \tag{18}$$

Hence, because n has $\beta + a(r) + b(r) + 1$ digits, it follows from (18) that

$$n < 10^{r(1+\varepsilon)} \quad (r \geq r_2) \tag{19}$$

Since, as we saw above, there are at least $\frac{10^\beta}{9\beta+1}$ ways of choosing u, we may conclude from (19) that there exist at least $\frac{10^\beta}{9\beta+1}$ Niven numbers smaller than $10^{r(1+\varepsilon)}$, that is

$$N\left(10^{r(1+\varepsilon)}\right) > \frac{10^\beta}{9\beta+1} > \frac{10^{r(1-\varepsilon)}}{9r(1-\varepsilon)+1} > 10^{r(1-2\varepsilon)},$$

for r sufficiently large, say $r \geq r_3$, where we used (17) and the fact that $\frac{10^\beta}{9\beta+1}$ increases with β.

From this, (12) follows with $r_0 = \max(r_1, r_2, r_3)$, and thus the lower bound (5).

3. THE UPPER BOUND FOR $N(x)$

We shall establish that

$$N(x) < 330 \cdot \log 10 \cdot \frac{x}{\log x} + \frac{495}{2} \cdot \log 10 \cdot \frac{x}{\log x} \log\left(\frac{5\log x + 5\log 10}{\log 10}\right), \tag{20}$$

from which the upper bound of our Theorem will follow immediately.

To establish (20), we first prove that for any positive integer r,

$$N(10^r) < \frac{99 \cdot \log(5r)}{4r} \cdot 10^r + \frac{33}{r} \cdot 10^r. \tag{21}$$

Clearly (20) follows from (21) by choosing $r = \left[\frac{\log x}{\log 10}\right] + 1$.

In order to prove (21), we first write

$$N(10^r) = A(r) + B(r) + 1,$$

where

$$A(r) = \#\{1 \le n < 10^r : s(n)|n \text{ and } |s(n) - 4.5r| > 0.5r\}$$

and

$$B(r) = \#\{1 \le n < 10^r : s(n)|n \text{ and } 4r \le s(n) \le 5r\}$$

To estimate $A(r)$, we use the idea introduced by Kennedy & Cooper [4] of considering the value $s(n)$, in the range $0, 1, 2, \ldots, 10^r - 1$ as a random variable of mean $\mu = 4.5r$ and variance $\sigma^2 = 8.25r$. This is justified by considering each digit of n as an independant variable taking each of the values $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ with a probability equal to $\frac{1}{10}$. Thus, according to Chebyshev's inequality (see for instance Galambos [2], p. 23), we have

$$P(|s(n) - \mu| > k) < \frac{\sigma^2}{k^2}, \text{ that is } P(|s(n) - 4.5r| > 0.5r) < \frac{8.25r}{(0.5r)^2} = \frac{33}{r}.$$

Now multiplying out this probability by the length of the interval $[1, 10^r - 1]$, we obtain the estimate

$$A(r) < \frac{33 \cdot 10^r}{r}. \tag{22}$$

The estimation of $B(r)$ requires a little bit more effort.

If we denote by $\alpha = \alpha(s(n))$ the number of digits of $s(n)$, then, since $4r \le s(n) \le 5r$, we have

$$\left[\frac{\log 4r}{\log 10}\right] + 1 \le \alpha \le \left[\frac{\log 5r}{\log 10}\right] + 1. \tag{23}$$

We shall write each integer n counted in $B(r)$ as the concatenation $n = [c, d]$, where $d = d(n)$ is the number formed by the last α digits of n and $c = c(n)$ is the number formed by the first $r - \alpha$ digits of n. Here, again for convenience, we allow c and thus n to begin with a string of 0's. Using this notation, it is clear that $s(n) = s(c) + s(d)$ which means that $s(c) = s(n) - s(d)$. From this, follows the double inequality

$$s(n) - 9\alpha \le s(c) \le s(n).$$

Hence, for any fixed value of $s(n)$, say $a = s(n)$, the number of distinct ways of choosing c is at most

$$\sum_{s(c)=a-9\alpha}^{a} M(r - \alpha, s(c)), \tag{24}$$

where $M(r,y)$ was defined in (8).

For fixed values of $s(n)$ and c, we now count the number of distinct ways of choosing d so that $s(n)|n$. This number is clearly no larger than the number of multiples of $s(n)$ located in the interval $I := [c \cdot 10^\alpha, (c+1) \cdot 10^\alpha]$. Since the length of this interval is 10^α, it follows that I contains at most $L := \left[\frac{10^\alpha}{s(n)} + 1\right]$ multiples of $s(n)$. Since α represents the number of digits of $s(n)$, it is clear that $L \leq 10 + 1 = 11$.

We have thus established that for fixed values of $s(n)$ and c, we have at most 11 different ways of choosing d.

It follows from this that for a fixed value a of $s(n) \in [4r, 5r]$, the number of "c, d combinations" yielding a positive integer $n < 10^r$ such that $s(n)|n$, that is $a|n$, is at most 11 times the quantity (24), that is

$$11 \sum_{s(c)=a-9\alpha}^{a} M(r-\alpha, s(c)). \tag{25}$$

Summing this last quantity in the range $4r \leq a \leq 5r$, we obtain that

$$B(r) \leq 11 \sum_{a=4r}^{5r} \sum_{s(c)=a-9\alpha}^{a} M(r-\alpha, s(c)).$$

Observing that in this double summation, $s(c)$ takes its values in the interval $[4r - 9\alpha, 5r]$ and that $s(c)$ takes each integer value belonging to this interval at most 9α times, we obtain that

$$B(r) \leq 11 \cdot 9\alpha \sum_{s(c)=4r-9\alpha}^{5r} M(r-\alpha, s(c)).$$

By widening our summation bounds and using (23), we have that

$$B(r) \leq 99\alpha \sum_{y=0}^{9r} M(r-\alpha, y) = 99\alpha \cdot 10^{r-\alpha} < 99\left(\frac{\log 5r}{\log 10} + 1\right) \cdot 10^{r-\alpha}.$$

Since by (23), $\alpha > \frac{\log 4r}{\log 10}$, we finally obtain that

$$B(r) \leq \frac{99 \cdot \log(4r) \cdot 10^r}{4r}. \tag{26}$$

Recalling that $N(10^r) = A(r) + B(r) + 1$, (21) follows immediately from (22) and (26), thus completing the proof of the upper bound, and thus of our Theorem.

Remarks:

1. We treated both $r - \alpha$ and $4r - 9\alpha$ as non negative integers without justification. Since it is sufficient to check that $4r > 9\alpha$ and since $\alpha \leq \frac{\log 5r + \log 10}{\log 10}$, it is enough to verify that $4r > \frac{9 \log 5r + 9 \log 10}{\log 10}$, which holds for all integers $r \geq 6$. For each $r \leq 5$, (21) is easily verified by direct computation.
2. Although we used probability theory, there was no breach in rigor. Indeed, this is because it is a fact that for $n < 10^r$, the i^{th} digit of n, for each $i = 1, 2, \ldots, r$ (allowing, as we did above, each number n to begin with a string of 0's so that is has r digits), takes on each integer value in [0,9] exactly one time out of ten.

4. THE SEARCH FOR THE ASYMPTOTIC BEHAVIOUR OF $N(x)$

By examining the table in §1, it is difficult to imagine if $N(x)$ is asymptotic to some expression of the form $x/L(x)$, where $L(x)$ is some slowly oscillating function such as $\log x$.

Nevertheless we believe that, as $x \to \infty$

$$N(x) = (c + o(1)) \frac{x}{\log x}. \tag{27}$$

where c is given in (4). We base our conjecture on a heuristic argument.

Here is how it goes. First we make the reasonable assumption that the probability that $s(n)|n$ is $1/s(n)$, provided that $s(n)$ is not a multiple of 3. On the other hand, since $3|s(n)$ if and only if $3|n$, we assume that, if $3 \parallel s(n)$, then the probability that $s(n)|n$ is $3/s(n)$. In a like manner, we shall assume that, if $9|s(n)$, then $s(n)|n$ with a probability of $9/s(n)$.

Hence using conditional probability, we may write that

$$\begin{aligned} P(s(n)|n) = {} & P(s(n)|n \text{ assuming that } 3 \not| (n)) \cdot P(3 \not| s(n)) \\ & + P(s(n)|n \text{ assuming that } 3 \parallel s(n)) \cdot P(3 \parallel s(n)) \\ & + P(s(n)|n \text{ assuming that } 9|s(n)) \cdot P(9|s(n)) \\ = {} & \frac{1}{s(n)} \cdot \frac{2}{3} + \frac{3}{s(n)} \cdot \frac{2}{9} + \frac{9}{s(n)} \cdot \frac{1}{9} = \frac{7}{3} \cdot \frac{1}{s(n)}. \end{aligned} \tag{28}$$

As we saw above, the expected value of $s(n)$ for $n \in [0, 10^r - 1]$ is $\frac{9}{2}r$. Combining this observation with (28), we obtain that if n is chosen at random in the interval $[0, 10^r - 1]$, then

$$P(s(n)|n) = \frac{7}{3} \cdot \frac{1}{9r/2} = \frac{14}{27r}.$$

Multiplying this probability by the length of the interval $[0, 10^r - 1]$, it follows that we can expect $\frac{14 \cdot 10^r}{27 \cdot r}$ Niven numbers in the interval $[0, 10^r - 1]$.

Therefore, given a large number x, if we let $r = \left[\frac{\log x}{\log 10}\right]$, we immediately obtain (27).

REFERENCES

[1] T. Cai. "On 2-Niven Numbers and 3-Niven Numbers." *The Fibonacci Quarterly* **34** (1996): 118-120.

[2] J. Galambos. *Advanced Probability Theory*. Marcel Dekker, 1988.

[3] H.G. Grundman. "Sequences of Consecutive n-Niven Numbers." *The Fibonacci Quarterly* **32** (1994): 174-175.

[4] R. E. Kennedy & C.N. Cooper. "On the Natural Density of the Niven Numbers." *The College Math Journal* **15** (1984): 309-312.

[5] I. Vardi. *Computational Recreations in Mathematics.* Addison-Wesley Publishing Co., Redwood City, CA 1991.

AMS Classification Numbers: 11A63, 11A25

✠ ✠ ✠