On the distribution in residue classes of integers with a fixed sum of digits

by

Christian Mauduit
(Institut de Mathématiques de Luminy, UPR 9016 CNRS
163, Avenue de Luminy, Case 907,
F-13288 Marseille Cedex 9, France)

Carl Pomerance
(Fundamental Mathematics Research
Bell Labs - Lucent Technologies
Murray Hill, NJ 07974, USA)

and

András Sárközy *
(Eötvös Loránd University
Department of Algebra and Number Theory
H-1117 Budapest, Pázmány Péter sétány 1/C, Hungary)

For our friend Jean-Louis Nicolas on his sixtieth birthday

*Research partially supported by the Hungarian National Foundation for Scientific Research, Grant No. T029759, and "Balaton" French-Hungarian exchange program F-18/00.

1 INTRODUCTION

Let g be an integer with $g \geq 2$. Let $S(n) = S_g(n)$ be the sum of the base-g "digits" of the natural number n. That is, if

$$(1.1) \qquad n = \sum_{j=0}^{J} a_j g^j, \quad 0 \leq a_j \leq g-1, \; a_J \geq 1,$$

then

$$S_g(n) = \sum_{j=0}^{J} a_j.$$

Also we let $S_g(0) = 0$. The function $S_g(n)$ evidently satisfies

$$(1.2) \qquad S_g(ig^\mu + j) = S_g(ig^\mu) + S_g(j) = S_g(i) + S_g(j) \;\text{ for }\; 0 \leq j < g^\mu;$$

the first equation representing a property called g-additivity.

Other than the familiar "rule of nines" in the case $g = 10$, which generalizes to the congruence

$$(1.3) \qquad n \equiv S_g(n) \pmod{g-1},$$

it is natural to conjecture that n and $S_g(n)$ are in some sense independent events as far as their distribution in residue classes. For example, Gelfond [7] has such a result when the moduli are fixed.

For a number $N \geq 1$, and integers m, h with $m \geq 1$, let

$$\begin{aligned} V_k(N) &= \#\{0 \leq n < N : S_g(n) = k\}, \\ V_k(N; m, h) &= \#\{0 \leq n < N : S_g(n) = k, \, n \equiv h \pmod{m}\}. \end{aligned}$$

In [9] the first and third authors, using the saddle point method, showed that uniformly in wide ranges, if $(m, g(g-1)) = 1$, then $V_k(N; m, h) \sim \frac{1}{m} V_k(N)$. It is our goal in this paper to study $V_k(N; m, h)$ with no coprimality condition on the modulus m. We are able to give a result that is uniformly valid in wide ranges and we use this result to solve some problems in elementary number theory.

In a popular lecture in 1977 at Miami University in Ohio, USA, Ivan Niven gave an example of how an easy child's puzzle might be thought of by a professional mathematician. The puzzle: find a whole number larger than 10 and less than 20 which is a multiple of the sum of its (base-10) digits. Niven suggested that a mathematician might ask instead for an asymptotic formula for the number of integers $n < N$ with $S_{10}(n)|n$, and to generalize to other bases. Thus was born the concept of a "Niven number". A base-g Niven number is a positive integer n with $S_g(n)|n$. Let $A_g(N)$ be the number of base-g Niven numbers $n < N$. In [2], Cooper and Kennedy show that $A_{10}(N) = o(N)$. (Other, related papers are [3]–[6], [8].) It is easy to see that

$$A_g(N) \;=\; \sum_{k\geq 1} V_k(N; k, 0),$$

so that it is clear that an understanding of the expressions $V_k(N; m, h)$ could be of help in the estimation of $A_g(N)$. In fact, our main theorem allows us to give an asymptotic formula for $A_g(N)$.

In 1976, Olivier [10] gave an asymptotic formula for the distribution of integers n with $(n, S_g(n)) = q$, where q is an arbitrary, but fixed positive integer. Our main theorem allows us to extend his result to nearly a best-possible range for q (namely, beyond this range, the asymptotic formula of Olivier cannot hold).

We also discuss some other applications, and some open problems.

First in Section 2 we will recall the result from [9] dealing with $V_k(N; m, h)$, and we show how the condition $(m, g(g-1)) = 1$ can be relaxed to $(m, g) = 1$ (with now a different main term). In Section 3 we give some lemmas that will be useful in relaxing the condition $(m, g) = 1$ to all m, and useful in some of the applications. In Section 4 we prove our main result on the distribution in residue classes of the numbers n with $S_g(n) = k$. Applications to Niven numbers, the problem of Olivier that we mentioned, and further applications and problems are discussed in Section 5.

Throughout this paper we we write $e(\alpha) = e^{2\pi i\alpha}$. We denote by $\mathbb{R}, \mathbb{Z}$ and $\mathbb{N}$ the sets of real numbers, integers, and positive integers. For $x \in \mathbb{R}$, we let $\lfloor x \rfloor$ be the greatest integer that does not exceed x, we let $\lceil x \rceil$ be the least integer which is not less than x, and we let $\|x\|$ be the distance of x from the nearest integer, that is the minimum of $\lceil x \rceil - x$ and $x - \lfloor x \rfloor$. All implicit constants, as well as the numbers $\ell_0, \ell_1, c_0, c_1, \ldots$, depend at most on the choice of g. The integer $g \geq 2$ is arbitrary, but considered as fixed throughout the paper. (It is probably not hard to cast our results with an explicit dependence on g, but we have not done so here.)

2 EARLIER RESULTS AND THE CONDITION $(m, g-1) = 1$

For N a real number at least 1, define

$$\nu = \nu(N) = \lfloor \log_g N \rfloor = \left\lfloor \frac{\log N}{\log g} \right\rfloor, \tag{2.1}$$

so that $g^\nu \leq N < g^{\nu+1}$. Set

$$\mu = \mu(N) = \frac{g-1}{2}\nu. \tag{2.2}$$

In [9] first $V_k(N)$ was estimated under various conditions on k and N. In particular, it was proved (Corollary 2 in [9]) that

LEMMA 1. *For $N \to \infty$ and*

$$\Delta = |\mu - k| = o(\nu), \tag{2.3}$$

we have

$$V_k(g^\nu) = 6^{1/2}\pi^{-1/2}(g^2-1)^{-1/2} g^\nu \nu^{-1/2} \exp\left(-\frac{6}{g^2-1} \cdot \frac{\Delta^2}{\nu} + O\big(\Delta^3\nu^{-2} + \nu^{-1/2}\big)\right).$$

One of the main results in [9] is that if $(m, g(g-1)) = 1$, $\ell := \min(k, (g-1)\nu - k)$ is large, and

$$m < \exp(c_0 \ell^{1/2}), \tag{2.4}$$

then $V_k(N)$ is well-distributed in the modulo m residue classes.

THEOREM A. *There exist positive constants ℓ_0, c_1, c_2 (all depending on g only) such that if $N, k, m \in \mathbb{N}$, $m \geq 2$,*

$$\big(m, g(g-1)\big) = 1, \tag{2.5}$$

$h \in \mathbb{Z}$, $\ell > \ell_0$ and (2.4) holds, then

$$\left| V_k(N; m, h) - \frac{1}{m} V_k(N) \right| < c_1 \frac{1}{m} V_k(N) \exp\left(-c_2 \frac{\ell}{\log m} \right).$$

(Indeed, this is Theorem 2 in [9].) The proof uses the saddle point method, and the following lemma (Lemma 2 in [9]) plays a crucial role in the proof:

LEMMA 2. *If $g, m, \varrho \in \mathbb{N}$, $m, g \geq 2$,*

$$\big(m, (g-1)g\big) = 1, \tag{2.6}$$

$$1 \leq j \leq m-1, \;\; \varrho \geq 2 \frac{\log m}{\log g} + 8 \quad \textit{and} \quad \beta \in \mathbb{R},$$

then

$$\sum_{u=0}^{\varrho-1} \big\| \beta + g^u \frac{j}{m} \big\|^2 \geq \frac{(g-1)^2}{128 g^4} \cdot \frac{\varrho}{\log m}.$$

Note that, as pointed out in the first paragraph in Section 5 of [9], the condition (2.6) can be replaced by

$$(m, g) = 1 \;\; \text{and} \;\; (g-1)\frac{j}{m} \notin \mathbb{Z}. \tag{2.7}$$

Using this idea, we now give a self-contained proof of the following strengthening of Lemma 2:

LEMMA 2'. *If the hypotheses of Lemma 2 hold except with (2.7) replacing (2.6), we have*

$$\sum_{u=0}^{\varrho-1} \big\| \beta + g^u \frac{j}{m} \big\|^2 \geq \frac{(g-1)^2}{20 g^4} \cdot \frac{\varrho}{\log m}.$$

PROOF. We first show that if k is an integer with $(g-1)k/m \notin \mathbb{Z}$ then there is an integer $n \in [0, \lceil \log_g m \rceil - 1]$ with

$$\|g^n(g-1)k/m\| \geq \frac{g-1}{g^2}. \tag{2.8}$$

Let $\alpha = \|(g-1)k/m\|$, so that $\alpha \geq 1/m$. Let n_1 be the least integer with $g^{n_1}\alpha \geq 1$, so that $1 \leq n_1 \leq \lceil \log_g m \rceil$. We have $g^{-1} \leq g^{n_1-1}\alpha < 1$. Say $g^{n_1-1}\alpha \leq 1-g^{-1}$. Then we have $\|g^{n_1-1}(g-1)k/m\| = \|g^{n_1-1}\alpha\| \geq g^{-1}$, so that we may take $n = n_1 - 1$. Thus, we may assume that $1 - g^{-1} < g^{n_1-1}\alpha < 1$. But $1 - g^{-1} \geq 1/2$, so we have $n_1 - 1 \geq 1$ and $g^{-1} - g^{-2} < g^{n_1-2}\alpha < g^{-1}$. Hence in this case we may take $n = n_1 - 2$. Thus, we have (2.8).

A consequence of (2.8) is that if k is an integer with $(g-1)k/m \notin \mathbb{Z}$, then

$$\sum_{n=0}^{\lceil \log_g m \rceil} \Big\|\beta + g^n \frac{k}{m}\Big\|^2 \geq \frac{(g-1)^2}{2g^4}. \tag{2.9}$$

Indeed, (2.9) follows from (2.8) and the inequality

$$\|\beta + g^n k/m\|^2 + \|\beta + g^{n+1}k/m\|^2 \geq \frac{1}{2}\|g^n(g-1)k/m\|^2.$$

To complete the proof of Lemma 2', let $b = \lceil \log_g m \rceil + 1$ and let $q = \lfloor (\varrho - 1)/b \rfloor$. An elementary calculation, using the hypothesis $\varrho \geq 2\log_g m + 8$, shows that

$$q \geq \frac{1}{2}\,\frac{\varrho}{\lceil \log_g m \rceil + 1} > \frac{\varrho}{10\log m}.$$

Thus,

$$\sum_{u=0}^{\varrho-1} \Big\|\beta + g^u \frac{j}{m}\Big\|^2 \geq \sum_{i=1}^{q} \sum_{u=(i-1)b}^{ib-1} \Big\|\beta + g^u \frac{j}{m}\Big\|^2 \geq q\frac{(g-1)^2}{2g^4} > \frac{(g-1)^2}{20g^4} \cdot \frac{\varrho}{\log m},$$

where the next-to-last inequality follows by applying (2.9) to the inner sum with $k = g^{(i-1)b}j$. This completes the proof of Lemma 2'.

Replacing Lemma 2 in the proof of Theorem A in [9] by Lemma 2', we can extend Theorem A to the case when $(m, g-1) = 1$ is not assumed:

THEOREM B. *There exist positive constants ℓ_1, c_3, c_4 (all depending on g only) such that if $N > 1$ is a real number, m is a positive integer with*

$$(m, g) = 1,$$

k, h, ℓ are integers such that $\ell > \ell_1$ and (2.4) holds, then, writing

$$d = (m, g-1), \tag{2.10}$$

we have

$$\left| V_k(N; m, h) - \frac{d}{m} V_k(N) \right| < c_3 \frac{1}{m} V_k(N) \exp\Big(- c_4 \frac{\ell}{\log m} \Big) \quad \text{for } k \equiv h \pmod{d} \tag{2.11}$$

and

$$V_k(N; m, h) = 0 \quad \text{for } k \not\equiv h \pmod{d}. \tag{2.12}$$

PROOF. The theorem is trivial if $m = 1$, so assume $m \geq 2$. Note too that (2.12) is trivial, since if $S_g(n) = k$, (1.3) implies that $n \equiv k \pmod{d}$. The non-trivial part of Theorem B, i.e., (2.11), can be proved along the same lines as Theorem A in [9], only the computation of the main term becomes slightly more complicated. Thus we will present this computation here, and we will omit the rest of the proof.

As in [9], we may restrict ourselves to the case $0 < k \leq \frac{g-1}{2}\nu$, and we use the saddle point method, which leads to the definition of the parameter r as the unique solution of the equation

$$\frac{r + 2r^2 + \cdots + (g-1)r^{g-1}}{1 + r + \cdots + r^{g-1}} = \frac{k}{\nu}$$

with $0 < r \leq 1$. Next we consider the generating function

$$G(z, \gamma) = \sum_{n=1}^{N} z^{S(n)} e(n\gamma)$$

(where $z \in \mathbb{C}, \gamma \in \mathbb{R}$) so that

$$\frac{1}{m}\sum_{j=0}^{m-1} e\left(-\frac{hj}{m}\right) G\left(z, \frac{j}{m}\right) = \sum_{\substack{1\le n\le N \\ n\equiv h \pmod m}} z^{S(n)}.$$

Thus taking $z = re(\beta)$ we have

$$V_k(N; m, h) = r^{-k}\int_0^1 e(-k\beta) \sum_{\substack{1\le n\le N \\ n\equiv h \pmod m}} (re(\beta))^{S(n)} d\beta$$

$$= \frac{1}{m} r^{-k} \sum_{j=0}^{m-1} \int_0^1 e\left(-k\beta - \frac{hj}{m}\right) G\left(re(\beta), \frac{j}{m}\right) d\beta. \tag{2.13}$$

In [9], assuming $(m, g-1) = 1$, there was a single main term: the one with $j = 0$. Now $(m, g-1) = 1$ is not assumed, and thus all the terms with j satisfying $(g-1)j/m \in \mathbb{Z}$ contribute to the main term. If $(g-1)j/m \in \mathbb{Z}$, then $\frac{j}{m}$ can be written as $\frac{a}{d}$ with $0 \le a < d$. Thus, the main term in (2.13) is

$$\frac{1}{m} r^{-k} \int_0^1 \sum_{a=0}^{d-1} e\left(-k\beta - h\frac{a}{d}\right) G\left(re(\beta), \frac{a}{d}\right) d\beta$$

$$= \frac{1}{m} r^{-k} \int_0^1 e(-k\beta) \sum_{n=1}^{N} r^{S(n)} e(S(n)\beta) \sum_{a=0}^{d-1} e\left((n-h)\frac{a}{d}\right) d\beta$$

$$= \frac{d}{m} r^{-k} \int_0^1 e(-k\beta) \sum_{\substack{n\le N \\ n\equiv h \pmod d}} r^{S(n)} e(S(n)\beta) d\beta = \frac{d}{m} V_k(N) \tag{2.14}$$

since now $k \equiv h \pmod d$ is assumed.

It follows from (2.13) and (2.14) that for $k \equiv h \pmod d$,

$$\left| V_k(N; m, h) - \frac{d}{m} V_k(N) \right| \le \frac{1}{m} r^{-k} \sum_{\substack{0\le j<m \\ (g-1)j/m \notin \mathbb{Z}}} \int_0^1 \left| G\left(re(\beta), \frac{j}{m}\right)\right| d\beta.$$

This upper bound can be estimated further in exactly the same way as in [9] except that now we use Lemma 2' in place of Lemma 2, thus we leave the details to the reader.

3 FURTHER LEMMAS

LEMMA 3. *For each positive integer* ν*, the sequence* $V_0(g^\nu), V_1(g^\nu), \ldots, V_{(g-1)\nu}(g^\nu)$ *is unimodal, with peak value* $V_{\lfloor (g-1)\nu/2\rfloor}(g^\nu)$*.*

PROOF. We have $V_k(g^\nu) = V_{(g-1)\nu-k}(g^\nu)$, so that it suffices to prove that if $1 \le k < \lfloor (g-1)\nu/2\rfloor$, then $V_k(g^\nu) \le V_{k+1}(g^\nu)$. This assertion is obvious for $\nu = 1$ since each $V_k(g) = 1$. Assume the lemma holds for $\nu - 1$. Clearly for $i = 0, 1, \ldots, g-2$, we have $V_k(g^\nu; g, i) = V_{k+1}(g^\nu; g, i+1)$, since if $n < g^\nu$, $n \equiv i \pmod g$, $S_g(n) = k$, then $n+1 < g^\nu$, $n+1 \equiv i+1 \pmod g$, $S_g(n+1) = k+1$. So, it suffices to prove that

$$V_k(g^\nu; g, g-1) \le V_{k+1}(g^\nu; g, 0), \tag{3.1}$$

where $g - 1 \le k < \lfloor (g-1)\nu/2\rfloor$. We have

$$\begin{aligned} &V_k(g^\nu; g, g-1) = V_{k-(g-1)}(g^{\nu-1}),\\ &V_{k+1}(g^\nu; g, 0) = V_{k+1}(g^{\nu-1}) = V_{(g-1)(\nu-1)-(k+1)}(g^{\nu-1}). \end{aligned}$$

If $k+1 \le (g-1)(\nu-1)/2$, the induction hypothesis implies that $V_{k-(g-1)}(g^{\nu-1}) \le V_{k+1}(g^{\nu-1})$, so that (3.1) holds. So assume $k+1 > (g-1)(\nu-1)/2$. Then

$$k - (g-1) < (g-1)(\nu-2)/2 \le (g-1)(\nu-1) - (k+1) < (g-1)(\nu-1)/2,$$

so that the induction hypothesis implies that

$$V_{k-(g-1)}(g^{\nu-1}) \le V_{(g-1)(\nu-1)-(k+1)}(g^{\nu-1}),$$

and again (3.1) holds. This completes the proof of the lemma.

Remark. It is likely that some argument similar to the one just given can show that for any N, the sequence $V_k(N)$ is unimodal in the variable k.

LEMMA 4. *There are positive constants* c_5, c_6*, depending at most on* g*, such that if* $N > N_0(g)$ *and* $\lambda > 0$*, then, with* μ, ν *as in (2.1), (2.2),*

$$\sum_{|k-\mu| \ge \lambda \nu^{1/2}} V_k(N) \le \max\left\{ c_5 N \exp\left(-\frac{6}{g^2-1}\lambda^2\right),\ c_5 N^{1-c_6/\log\log N} \right\}.$$

PROOF. Clearly $V_k(N)$ is increasing in N, so that we have $V_k(N) \le V_k(g^{\nu+1})$. If $|k - \frac{1}{2}(g-1)(\nu+1)| = o(\nu)$, Lemma 1 implies that (for a number c depending at most on g)

$$
\begin{aligned}
V_k(g^{\nu+1}) &\le cN\nu^{-1/2} \exp\left(-\frac{6}{g^2-1}\frac{(k-\frac{1}{2}(g-1)(\nu+1))^2}{\nu}\right) \\
&\le 2cN\nu^{-1/2} \exp\left(-\frac{6}{g^2-1}\frac{(k-\mu)^2}{\nu}\right).
\end{aligned} \tag{3.2}
$$

If $|k-\mu| \ge \nu/(\log\nu)^{1/2}$, then (3.2) and Lemma 3 imply that

$$V_k(g^{\nu+1}) \le c_5 N^{1-c_7/\log\log N}.$$

Thus,

$$\sum_{|k-\mu|\ge \nu/(\log\nu)^{1/2}} V_k(g^{\nu+1}) \le c_5 N^{1-c_6/\log\log N}, \tag{3.3}$$

where $0 < c_6 < c_7$. Using

$$\sum_{|k-\mu|\ge\lambda\nu^{1/2}} \exp\left(-\frac{6}{g^2-1}\frac{(k-\mu)^2}{\nu}\right) = O\left(\nu^{1/2}\exp\left(-\frac{6}{g^2-1}\lambda^2\right)\right),$$

the lemma now follows from (3.2) and (3.3).

Remark. By using Bernstein's inequality (see, e.g., [11, Ch. 7]) it is possible to obtain an upper bound of the form $N\exp(-c\lambda^2)$ in essentially the entire range.

LEMMA 5. *For each real number $N > 1$ and integer $g \ge 2$, let ν be as in (2.1). For each integer k satisfying*

$$\Delta := \left|\frac{g-1}{2}\nu - k\right| \le \nu^{5/8}, \tag{3.4}$$

we have

$$V_k(N) = 6^{1/2}\pi^{-1/2}(g^2-1)^{-1/2}N\nu^{-1/2}\exp\left(-\frac{6}{g^2-1}\frac{\Delta^2}{\nu} + O(\nu^{-1/8})\right). \tag{3.5}$$

PROOF. Write

$$H = \left\lceil \frac{\log g}{g^2 - 1} \nu^{1/4} \right\rceil = O\left((\log N)^{1/4}\right), \tag{3.6}$$

and let

$$\nu_0 = \nu(N/g^H) = \nu - H. \tag{3.7}$$

Set $Q = \lfloor N/g^{\nu_0} \rfloor$. Then clearly we have

$$V_k(N) = \sum_{i=0}^{Q-1} \left(V_k((i+1)g^{\nu_0}) - V_k(ig^{\nu_0})\right) \; + \; V_k(N) - V_k(Qg^{\nu_0}). \tag{3.8}$$

For $ig^{\nu_0} < n \le (i+1)g^{\nu_0}$, write n in the form

$$n = ig^{\nu_0} + j \quad \text{with } 0 \le j < g^{\nu_0}.$$

By the g-additivity (1.2), for this n we have

$$S_g(n) = S_g(i) + S_g(j).$$

Since $i \le Q - 1 < g^H$, we have

$$S_g(i) \le (g-1)H = O\left((\log N)^{1/4}\right). \tag{3.9}$$

Thus the general term in the sum in (3.8) can be rewritten as

$$V_k((i+1)g^{\nu_0}) - V_k(ig^{\nu_0}) = V_{k_i'}(g^{\nu_0}) \tag{3.10}$$

with

$$k_i' = k - S_g(i).$$

Further, by (3.9) we have

$$k - k_i' = O\left((\log N)^{1/4}\right). \tag{3.11}$$

Now we will use Lemma 1 with k_i', ν_0, and $\Delta_0 = |\frac{g-1}{2}\nu_0 - k_i'|$ in place of k, ν and Δ, respectively. Note then by (3.6) and (3.11) we have

$$|\Delta_0 - \Delta| \le \frac{g-1}{2} H + k - k_i' = O\left((\log N)^{1/4}\right) = O\left(\nu_0^{1/4}\right) \tag{3.12}$$

so that (2.3) holds and thus, indeed, Lemma 1 can be applied. Since (3.4), (3.6), (3.7) and (3.12) imply that $\Delta_0^3 \nu_0^{-2} = O(\nu_0^{-1/8})$, we obtain

$$V_{k_i'}(g^{\nu_0}) = 6^{1/2}\pi^{-1/2}(g^2-1)^{-1/2} g^{\nu_0} \nu_0^{-1/2} \exp\left(-\frac{6}{g^2-1} \frac{\Delta_0^2}{\nu_0} + O\left(\nu_0^{-1/8}\right)\right). \tag{3.13}$$

By (3.4), (3.6), (3.7) and (3.12), here we have

$$\begin{aligned}
\nu_0^{-1/2} &= \nu^{-1/2}\left(1 + O(\nu^{-3/4})\right) \\
\nu_0^{-1} &= \nu^{-1}\left(1 + O(\nu^{-3/4})\right) \\
\Delta_0^2 &= \Delta^2 + O\left(\nu^{1/4}\Delta + \nu^{1/2}\right) = \Delta^2 + O\left(\nu^{7/8}\right),
\end{aligned}$$

so that (3.13) implies that

$$V_{k_i'}(g^{\nu_0}) = 6^{1/2}\pi^{-1/2}(g^2-1)^{-1/2} g^{\nu_0} \nu^{-1/2} \exp\left(-\frac{6}{g^2-1} \frac{\Delta^2}{\nu} + O(\nu^{-1/8})\right). \tag{3.14}$$

By (3.8), (3.10), (3.14) and the definition of Q, we have

$$\begin{aligned}
V_k(N) =& \lfloor N g^{-\nu_0} \rfloor 6^{1/2}\pi^{-1/2}(g^2-1)^{-1/2} g^{\nu_0} \nu^{-1/2} \exp\left(-\frac{6}{g^2-1} \frac{\Delta^2}{\nu} + O(\nu^{-1/8})\right) \\
&+ V_k(N) - V_k(Q g^{\nu_0}) \\
=& 6^{1/2}\pi^{-1/2}(g^2-1)^{-1/2} N \nu^{-1/2} \exp\left(-\frac{6}{g^2-1} \frac{\Delta^2}{\nu} + O(\nu^{-1/8})\right) \\
&+ O\left(g^{\nu_0} + N - Q g^{\nu_0}\right).
\end{aligned} \tag{3.15}$$

The expression in the last O-term in (3.15) is at most

$$g^{\nu_0} + N - Q g^{\nu_0} < 2 g^{\nu_0} \le 2N/g^H \le 2N \exp\left(-\frac{1}{g^2-1} \nu^{1/4}\right),$$

using (3.6). Thus, (3.5) follows by putting this estimate into (3.15) and using (3.4). This completes the proof of Lemma 5.

LEMMA 6. *For integers k, k_1 satisfying the hypotheses of Lemma 5, with $|k - k_1| \le (\log N)^{1/4}$, we have*

$$V_{k_1}(N) = V_k(N)(1 + O\left((\log N)^{-1/8}\right)).$$

PROOF. If Δ corresponds to k in (3.4) and Δ_1 corresponds to k_1, then $|\Delta - \Delta_1| \le (\log N)^{1/4}$. Then, by (3.4), $\Delta^2/\nu = \Delta_1^2/\nu + O(\nu^{-1/8})$. Thus, Lemma 6 follows from (3.5).

4 RELAXATION OF $(m, g) = 1$

We now prove the following extension of Theorem B to the general case. Unfortunately, the error estimate and the range are not as good; this appears to be more an artifact of the proof than the truth.

THEOREM C *There is a positive constant c_8 (depending on g only) such that if $N \in \mathbb{R}$, $N > 1$, $k, m \in \mathbb{N}$, (3.4) holds, $h \in \mathbb{Z}$, $m < 2^{(\log N)^{1/4}}$ and d is as in* (2.10), *then*

(4.1)
$$\left| V_k(N; m, h) - \frac{d}{m} V_k(N) \right| < c_8 \frac{1}{m} V_k(N)/(\log N)^{1/8}$$
$$\text{for} \quad k \equiv h \pmod{d}$$

and

(4.2)
$$V_k(N; m, h) = 0 \qquad \text{for} \quad k \not\equiv h \pmod{d}.$$

PROOF. Clearly (4.2) holds if $k \not\equiv h \pmod{d}$, so henceforth we shall assume that $k \equiv h \pmod{d}$.

Write $m = m_1 m_2$ where each prime factor of m_1 divides g and no prime factor of m_2 divides g. Let x be the least integer with $m_1 | g^x$. Clearly x is at most the largest exponent in the canonical factorization of m_1 into powers of primes, so that

$$x \leq \log m_1 / \log 2 \leq \log m / \log 2 < (\log N)^{1/4},$$

by our hypothesis. If n is an integer in $[1, N)$, write

$$n = n_2 g^x + n_1,$$

where n_1 is a nonnegative integer smaller than g^x. Further, by (1.2), $S_g(n) = k$ if and only if

$$S_g(n_2) = k'(n_1) := k - S_g(n_1).$$

By the bound on x we have

$$S_g(n_1) \le (g-1)x < (g-1)(\log N)^{1/4}.$$

We have that $n \equiv h \pmod{m}$ if and only if

$$n \equiv h \pmod{m_1} \ \text{ and } \ n \equiv h \pmod{m_2}$$

if and only if

$$n_1 \equiv h \pmod{m_1} \ \text{ and } \ n_2 \equiv h'(n_1) \pmod{m_2},$$

where $h'(n_1)$ is the least nonnegative residue of $(h-n_1)g^{-x} \pmod{m_2}$, with g^{-x} being the multiplicative inverse of g^x modulo m_2. Thus,

$$V_k(N;m,h) = \sum_{0\le n_1<g^x,\, n_1\equiv h \pmod{m_1}} V_{k'(n_1)}((N-n_1)/g^x; m_2, h'(n_1)) \ . \tag{4.3}$$

Let $\nu'(n_1) = \nu((N-n_1)/g^x)$. Then $\nu'(n_1) = \nu - x + O(1)$.

We now apply Theorem B to the individual terms on the right side of (4.3). By construction, we have $(g, m_2) = 1$, so that the only hypothesis that needs to be checked is that (2.4) holds. Since $m_2 \le m < 2^{(\log N)^{1/4}}$, it suffices to show that $\ell' := \min(k'(n_1), (g-1)\nu'(n_1) - k'(n_1)) > c_0^{-2}(\log 2)^2(\log N)^{1/2}$. However $|\ell' - \ell| = O((\log N)^{1/4})$, where $\ell = \min(k, (g-1)\nu - k)$, since $\nu - \nu'(n_1) = x + O(1)$ with $0 \le x < (\log N)^{1/4}$ and since $k - k'(n_1) = S_g(n_1) < (g-1)(\log N)^{1/4}$. Hence (3.4) implies we have (2.4). Thus, we may apply Theorem B.

Since we are assuming that $k \equiv h \pmod{d}$, and since

$$\begin{aligned} k'(n_1) &= k - S_g(n_1) \equiv k - n_1 \pmod{g-1}, \\ h'(n_1) &\equiv (h-n_1)g^{-x} \equiv h - n_1 \pmod{(g-1, m_2)}, \end{aligned}$$

we have that $k'(n_1) \equiv h'(n_1) \pmod{(g-1, m_2)}$. So, from (4.3) and Theorem B, we have

$$\begin{aligned} V_k(N;m,h) = \frac{(g-1,m_2)}{m_2} &\sum_{0\le n_1<g^x,\, n_1\equiv h \pmod{m_1}} V_{k'(n_1)}((N-n_1)/g^x) \\ &+ O\left(\frac{1}{m_2} \sum_{0\le n_1<g^x,\, n_1\equiv h \pmod{m_1}} V_{k'(n_1)}(N/g^x)\exp(-c_3 \log N/\log m)\right). \end{aligned}$$

We have, by Lemmas 5 and 6,

$$\begin{aligned}
&\sum_{0\le n_1<g^x,\, n_1\equiv h \pmod{m_1}} V_{k'(n_1)}((N-n_1)/g^x)\\
&\qquad = \sum_{0\le n_1<g^x,\, n_1\equiv h \pmod{m_1}} V_{k'(n_1)}(N/g^x) \ + \ O(g^x/m_1)\\
&\qquad = \frac{1}{g^x}\sum_{0\le n_1<g^x,\, n_1\equiv h \pmod{m_1}} V_{k'(n_1)}(N)\left(1+O\big((\log N)^{-1/8}\big)\right)\\
&\qquad = \frac{1}{g^x}\sum_{0\le n_1<g^x,\, n_1\equiv h \pmod{m_1}} V_{k}(N)\left(1+O\big((\log N)^{-1/8}\big)\right)\\
&\qquad = \frac{1}{m_1}V_k(N)\left(1+O\big((\log N)^{-1/8}\big)\right).
\end{aligned}$$

Using this calculation in the prior one, we obtain

$$V_k(N;m,h) = \frac{(g-1,m_2)}{m_1m_2}V_k(N)\left(1+O\big(\exp(-c_3\log N/\log m)+(\log N)^{-1/8}\big)\right),$$

and since $(g-1,m_2)=(g-1,m)$ and $m_1m_2=m$, we have the theorem.

5 APPLICATIONS AND PROBLEMS

In this section we give several applications of Theorem C, as well as some additional problems.

Let

$$D(g) = \frac{2\log g}{g-1}\prod_{p^\alpha\|g-1}\left(1+\alpha(1-p^{-1})\right),$$

for each integer $g\ge 2$. We shall prove the following result, which gives an asymptotic formula for the distribution of the g-Niven numbers defined in Section 1.

THEOREM D. *If g is an integer at least 2, then for each number $N>1$,*

$$A_g(N) \ = \ \#\{0<n<N : S_g(n)|n\} \ = \ D(g)\frac{N}{\log N} + O\left(\frac{N}{(\log N)^{9/8}}\right).$$

PROOF. Recall the notation from (2.1), (2.2) and let

$$Z = \nu^{5/8}.$$

We have

$$A_g(N) = \sum_{n<N,\, S_g(n)|n} 1 = \sum_{k=1}^{(g-1)\nu} \sum_{n<N,\, S_g(n)=k,\, k|n} 1$$
$$= \sum_{k=1}^{(g-1)\nu} V_k(N;k,0)$$
$$= \sum_{|k-\mu|\le Z} V_k(N;k,0) \; + \; \sum_{|k-\mu|>Z} V_k(N;k,0). \tag{5.1}$$

It follows from Lemma 4 that

$$\sum_{|k-\mu|>Z} V_k(N;k,0) \le \sum_{|k-\mu|>Z} V_k(N) = O\left(N\exp(-6\nu^{1/4}/(g^2-1))\right). \tag{5.2}$$

Using Theorem C we have that

$$\sum_{|k-\mu|\le Z} V_k(N;k,0) = \sum_{|k-\mu|\le Z} \frac{(k,g-1)}{k} V_k(N) \; + \; O\left(\sum_{|k-\mu|\le Z} \frac{1}{k} V_k(N)(\log N)^{-1/8}\right).$$

For $|k-\mu|\le Z$, we have $1/k = 1/\mu + O(Z/\mu^2) = 1/\mu + O((\log N)^{-11/8})$, so that

$$\sum_{|k-\mu|\le Z} V_k(N;k,0) = \frac{1}{\mu}\sum_{|k-\mu|\le Z}(k,g-1)V_k(N) \; + \; O\left(\frac{N}{\mu(\log N)^{1/8}}\right). \tag{5.3}$$

(A better error estimate may be had at this point, but it will later be swamped, so we have used the trivial estimate $V_k(N) < N$.) Further, the function $f(k) = (k, g-1)$ is periodic with period $g-1$, and, as a simple calculation shows, the average value of $f(k)$ is

$$C(g) := \prod_{p^\alpha \| g-1} \left(1 + \alpha(1-p^{-1})\right).$$

Thus, by Lemma 6, we have

$$\sum_{|k-\mu|\le Z}(k,g-1)V_k(N) = C(g)\sum_{|k-\mu|\le Z} V_k(N)\, \left(1 + O((\log N)^{-1/8})\right). \tag{5.4}$$

Further,

$$\sum_{|k-\mu|\le Z} V_k(N) = N + O\left(\sum_{|k-\mu|>Z} V_k(N)\right),$$

so that using (5.2), we have

$$\sum_{|k-\mu|\le Z} (k, g-1)V_k(N) \;=\; C(g)N\left(1 + O((\log N)^{-1/8})\right).$$

Using this in (5.3), we have

$$\begin{aligned}\sum_{|k-\mu|\le Z} V_k(N; k, 0) \;&=\; C(g)\frac{N}{\mu} \;+\; O\left(\frac{N}{\mu(\log N)^{1/8}}\right)\\ &=\; D(g)\frac{N}{\log N} \;+\; O\left(\frac{N}{(\log N)^{9/8}}\right).\end{aligned}$$

With (5.1) and (5.2), this calculation completes the proof of Theorem D.

Now we consider a problem of Olivier [10]. He showed that for each fixed positive integer q,

$$\#\{0 < n < N : \, (n, S_g(n)) = q\} \;=\; a_q N + O(N/(\log N)^{1/8+o(1)}),$$

where

$$a_q = 6\pi^{-2}(q, g-1)q^{-2} \prod_{p|(g-1)/(q,g-1)} \left(1 + p^{-1}\right)^{-1},$$

the letter p in the product running over primes. Using Theorem C, the lemmas, and

$$\#\{n \le N : \, (n, S_g(n)) = q\} \;=\; \sum_{m\ge 1}\sum_{l|m} \mu(l)V_{mq}(N; ql, 0),$$

(here, μ is the Möbius function), it is routine to show that uniformly for $q \le \nu^{1/2}/(\log \nu)^2$,

$$\#\{n \le N : \, (n, S_g(n)) = q\} \;\sim\; a_q N,$$

as $N \to \infty$ (and explicit error estimates may be worked out as well). We have not optimized the exponent on $\log \nu$ in the range for q, but it is fairly easy to see that the range $q \le \nu^{1/2}/(\log \nu)^2$ is close to best possible. For example, if $q \approx c\nu^{1/2}$ where c is a large constant (depending at most on g), then a value of q with a multiple quite close to μ will give quite different behavior from a value of q whose multiple closest to μ is about $q/2$ away from μ.

In [3] Cooper and Kennedy generalized the problem of the Niven numbers by considering arithmetic functions $f\colon \mathbb{N} \to \mathbb{N}$, and then estimating the number of integers n with $n \le N$, $f(n)|n$. Later Erdős and Pomerance [4] extended and sharpened their results. There are various other digital-sum problems that our methods in this paper can handle. E.g., it follows easily from Theorem C, in the same way that Theorem D is proved, that for any fixed positive integers g, t with $g \ge 2$, an asymptotic formula for the number of integers n with

$$(S_g(n))^t \mid n, \qquad n \le N \ , \tag{5.5}$$

is attainable.

One might like to extend the problem by studying numbers n which are Niven numbers simultaneously with respect to several distinct bases $g_1, g_2, \dots, g_t$:

$$S_{g_1}(n) \mid n,\ S_{g_2}(n) \mid n, \dots, S_{g_t}(n) \mid n \ . \tag{5.6}$$

It is easy to see that for any $t \in \mathbb{N}$, (5.6) has infinitely many solutions in $g_1, \dots, g_t, n$. Indeed, consider a number $n \in \mathbb{N}$ with $(t+1)! \mid n$, and for $1 \le i \le t$ set $g_i = \dfrac{n}{i+1}$. Then the representation of n in the number system to base g_i is

$$n = (i+1)g_i + 0$$

so that

$$S_{g_i}(n) = (i+1) \mid (t+1)! \mid n \qquad \text{for } i = 1, 2, \dots, t.$$

To exclude this trivial example, one might like to add the restriction

$$g_1, g_2, \dots, g_t < n^{1/2}. \tag{5.7}$$

Still further simple constructions can be given:

PROPOSITION. *Let $g, t, k \in \mathbb{N}$, $g \ge 2$,*

$$(k, g) = 1 \tag{5.8}$$

and let $h_1 < h_2 < \cdots < h_k$ be positive integers with

$$g^{h_j} \equiv 1 \pmod{k} \qquad (\textit{for } j = 1, 2, \dots, k). \tag{5.9}$$

(By (5.8), there are infinitely many $h_j \in \mathbb{N}$ with this property.) Set

$$n = \sum_{j=1}^{k} g^{t! h_j}$$

and

$$g_i = g^i \qquad \text{for } i = 1, 2, \ldots, t.$$

Then the number n is a Niven number with respect to each of the bases $g_1, g_2, \ldots, g_t$ (and for $t > 3$ (5.7) also holds).

PROOF. For each of $i = 1, 2, \ldots, t$, the number n is the sum of k distinct powers of $g_i = g^i$, and thus we have

$$S_{g_i}(n) = k \qquad (\text{for } i = 1, 2, \ldots, t).$$

Moreover, it follows from (5.9) that we have

$$n = \sum_{j=1}^{k} g^{t! h_j} \equiv \sum_{j=1}^{k} 1 \equiv 0 \pmod{k}$$

so that, indeed, $S_{g_i}(n) \mid n$ for each of $i = 1, 2, \ldots, t$.

Note that each of the constructions above gives only a few simultaneous Niven numbers. It seems to be a much more difficult problem to give asymptotics for the number of solutions up to N. Two further problems that we have not been able to settle:

PROBLEM 1. Is it true that for any $t \in \mathbb{N}$ there are infinitely many $g_1, g_2, \ldots, g_t$, n satisfying (5.6), (5.7) and

$$(g_i, g_j) = 1 \qquad \text{for } 1 \le i < j \le t\ ?$$

PROBLEM 2. Is it true that if $g_1, g_2 \in \mathbb{N}$ and $g_1, g_2 \ge 2$ then there are infinitely many positive integers n which are Niven numbers simultaneously to both bases g_1, g_2? (Note that if g_1, g_2 are multiplicatively dependent, i.e., $\dfrac{\log g_1}{\log g_2}$ is rational, then by the proposition above, the answer is affirmative.) In particular, are there infinitely many numbers which are Niven numbers simultaneously to both bases $g_1 = 2$ and $g_2 = 3$?

More fundamentally, we may consider possible generalizations of Theorems A, B, and C to simultaneous bases. The papers [1], [12], [13], and [14] are perhaps relevant here.

ADDED IN PROOF. It has recently come to our attention that J.-M. De Koninck, N. Doyon, and I. Kátai, in "On the counting function for the Niven numbers," Acta Arith. **106** (2003), 265–275, have achieved results very similar to ours. In particular, their Theorem 2 is similar to our Theorem C, and their Theorem 1 is similar to our Theorem D, but with a weaker error estimate.

References

[1] J. Bésineau, Sur un problème de Gel'fond relatif à la fonction "somme des chiffres", C. R. Acad. Sci. Paris Sér. A-B 272 (1971), A453–A456.

[2] C. N. Cooper and R. E. Kennedy, On the natural density of the Niven numbers, College Math. J. 15 (1984), 309–312.

[3] C. N. Cooper and R. E. Kennedy, On an asymptotic formula for the Niven numbers, Internat. J. Math. Math. Sci. 8 (1985), 537–543.

[4] C. N. Cooper and R. E. Kennedy, A partial asymptotic formula for the Niven numbers, Fibonacci Quart. 26 (1988), 163–168.

[5] C. N. Cooper and R. E. Kennedy, Chebyshev's inequality and natural density, Amer. Math. Monthly 96 (1989), 118–124.

[6] P. Erdős and C. Pomerance, On a theorem of Besicovitch: values of arithmetic functions that divide their arguments, Indian J. Math. 32 (1990), 279–287.

[7] A. O. Gel'fond, Sur les nombres qui ont des propriétés additives et multiplicatives données, Acta Arith. 13 (1968) 259–265.

[8] R. E. Kennedy, T. Goodman and C. Best, Mathematical discovery and Niven numbers, MATYC J. 14 (1980), 21–25.

[9] C. Mauduit and A. Sárközy, On the arithmetic structure of the integers whose sum of digits is fixed, Acta Arith. 81 (1997), 145–173.

[10] M. Olivier, Fonctions g-additives et formule asymptotique pour la propriété $(n, f(n)) = q$, Acta Arith. 31 (1976), 361–384.

[11] A. Rényi, Probability Theory, North–Holland Series in Applied Mathematics and Mechanics, Vol. 10, North–Holland Publishing Co., New York, 1970.

[12] H. G. Senge and E. G. Straus, PV-numbers and sets of multiplicity, Proc. Washington State U. Conf. Number Theory, Dept. Math., Washington State U., Pullman, WA, 1971, pp. 55–67.

[13] H. G. Senge and E. G. Straus, PV-numbers and sets of multiplicity, Period. Math. Hungar. 3 (1973), 93–100.

[14] C. L. Stewart, On the representation of an integer in two different bases, J. Reine Angew. Math. 319 (1980), 63–72.