PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 59, Number 1, August 1976

THE SUM OF A DIGITADDITION SERIES

KENNETH B. STOLARSKY

ABSTRACT. Let $B(x)$ be the number of ones in the binary expansion of x. A "digitaddition series" is a sequence $y_1 < y_2 < y_3 < \dots$, where y_1 is a given positive integer and $y_{n+1} = y_n + B(y_n)$ for $n = 1, 2, \dots$. Various questions involving the y_m are studied; in particular, the asymptotic result $y_m \sim (m \log m)/(2 \log 2)$ is proved.

1. **Introduction.** For positive integers x, let $B(x)$ denote the sum of the digits in the binary expansion of x. For example, the binary expansion of 13 is 1101, so $B(13) = 3$. A sequence of integers $y_1 < y_2 < y_3 < \dots$ is called a "digitaddition series" if

$$y_{n+1} = y_n + B(y_n), \qquad n = 1, 2, \dots. \tag{1.1}$$

Such series have been studied by Kaprekar [7], [11]–[14] and others [1]–[10], [15]–[18]. Much attention [7], [10]–[14], [17]–[18] has been given to *self-numbers*, the integers that are not of the form $x + B(x)$. However, the asymptotics of digitaddition series seem to have been neglected. M. Gardner [7] points out (for the corresponding problem in base ten) that no simple formula seems to be known for the sum

$$S(n) = S(n; y_1) = \sum_{m=1}^{n} y_m. \tag{1.2}$$

We prove

$$S(n) \sim (n^2/4)(\log n)/(\log 2), \tag{1.3}$$

and in fact a bit more. We remark that the right side of (1.3) is independent of y_1. Here $f(n) \sim g(n)$ has the usual meaning, that $\lim f(n)/g(n) \to 1$ as $n \to \infty$.

We first show that the sequence y_m grows "slowly" by obtaining a crude upper bound for y_m. Next, we note that if x is a "typical" integer, then $B(x)$ is approximately $(\log_2 x)/2$. Thus, since the sequence y_m grows "slowly", most of its terms must be "typical" integers, and hence y_m is approximately $\sum_{x=1}^{m} (\log_2 x)/2 \sim (m \log_2 m)/2$. To carry out the details we use the inequality

$$\sum_{j > (T/2) + \lambda} \binom{T}{j} < 2^T \exp(-2\lambda^2/T); \tag{1.4}$$

see [6, p. 17] or [5].

Received by the editors July 25, 1975.
AMS (MOS) subject classifications (1970). Primary 10A35, 10A40, 10L99.
Key words and phrases. Binary expansion, digitaddition series, self-number.

2. The results. Henceforth, log t shall denote the logarithm of t to the base 2.

THEOREM 1.

$$S(n) = (n^2/4)\log n + O\{n^2(\log n \log\log n)^{1/2}\}. \tag{2.1}$$

Since

$$\int x \ln x \, dx = (x^2/2)\ln x - (x^2/4) + C, \tag{2.2}$$

Theorem 1 can be deduced easily from the following result.

THEOREM 2.

$$y_m = (m/2)\log m + O\{m(\log m \log\log m)^{1/2}\}. \tag{2.3}$$

In particular, $y_m \sim (m/2)\log m$.

3. **The proof.** We first obtain a crude upper bound on y_m. Iteration of (1.1) yields

$$y_{m+1} = y_1 + \sum_{k=1}^{m} B(y_k). \tag{3.1}$$

The trivial bound $B(x) \leqslant 1 + [\log x]$, where $[z]$ denotes the greatest integer in z, yields

$$y_{m+1} \leqslant y_1 + m + \log(y_1 y_2 \cdots y_m). \tag{3.2}$$

The trivial bound $B(x) \leqslant x$, together with (1.1), yields $y_m \leqslant 2^m y_1$. Thus, from (3.2), we find that

$$y_{m+1} \leqslant m^2 \tag{3.3}$$

for m sufficiently large, say $m > M$. By (3.3) and (3.2) again, we obtain

$$y_{m+1} \leqslant y_1 + m + \log(y_1 \cdots y_M) + \log(m!)^2 \leqslant 3m \log m \tag{3.4}$$

for m sufficiently large, say $m \geqslant m_0$.

We now refine this upper bound. Choose t real so that

$$[t/\log t] = m. \tag{3.5}$$

Then for $m \geqslant m_0$ we have from (3.4) and (3.5) that

$$1 \leqslant y_i \leqslant y_m \leqslant 3t \quad \text{for } 1 \leqslant i \leqslant m. \tag{3.6}$$

Next, set $T = 1 + [\log 3t]$ and let λ be a positive real number. Define $u = u(\lambda)$ by

$$u = T/2 + \lambda. \tag{3.7}$$

Let $s = s(t, \lambda)$ denote the number of integers y such that $1 \leqslant y \leqslant 3t$ and

$$B(y) \geqslant u. \tag{3.8}$$

The number of y such that $0 \leqslant y \leqslant 3t$ and $B(y) = j$ is at most $\binom{T}{j}$, so by (1.4) we have

$$s \leqslant \sum_{j > u} \binom{T}{j} < 6t \exp\{-2\lambda^2/T\}. \tag{3.9}$$

Now choose

$$\lambda = (T/2)^{1/2}\{\log(\log^2 t)\}^{1/2}. \tag{3.10}$$

Thus

$$s < 6t/\log^2 t \tag{3.11}$$

and from (3.1) we have

$$\begin{aligned} y_m &\leqslant y_1 + u\{m - 1 - s\} + Ts \\ &= y_1 + \left\{ \frac{\log t}{2} + O(\{\log t \log\log t\}^{1/2}) \right\} \\ &\quad \cdot \left\{ \frac{t}{\log t} + O\left(\frac{t}{\log^2 t} \right) \right\} + O\left(\frac{t}{\log t} \right). \end{aligned} \tag{3.12}$$

We conclude that

$$y_m \leqslant t/2 + O\left(t(\log t)^{-1/2}(\log\log t)^{1/2}\right). \tag{3.13}$$

From (3.5) it is easy to obtain

$$m \log m \leqslant t \leqslant m \log m + O(m \log\log m). \tag{3.14}$$

Hence

$$y_m \leqslant (m/2)\log m + O\left(m(\log m \log\log m)^{1/2}\right). \tag{3.15}$$

We now use the same method to obtain a lower bound for y_m. This time define u by

$$u = T/2 - \lambda \tag{3.16}$$

and let $s = s(t, \lambda)$ be the number of integers y such that $1 \leqslant y \leqslant 3t$ and

$$B(y) \leqslant u. \tag{3.17}$$

Then (note that $\binom{T}{j} = \binom{T}{T-j}$) we have

$$s \leqslant \sum_{j \leqslant u} \binom{T}{j} < 6t \exp\{-2\lambda^2/T\}. \tag{3.18}$$

By choosing λ exactly as before, we obtain

$$\begin{aligned} y_m &\geqslant u\{m - 1 - s\} \\ &= \left\{ \frac{\log t}{2} + O(\{\log t \log\log t\}^{1/2}) \right\} \left\{ \frac{t}{\log t} + O\left(\frac{t}{\log^2 t} \right) \right\}. \end{aligned} \tag{3.19}$$

We conclude from (3.19) and (3.14) that

$$y_m \geqslant t/2 + O\left(t(\log t)^{-1/2}(\log\log t)^{1/2}\right) \tag{3.20}$$

and

$$y_m \geqslant (m/2)\log m + O\big(m(\log m \log\log m)^{1/2}\big). \tag{3.21}$$

This completes the proof.

4. **Remarks.** Theorem 2 *cannot* be improved to

$$y_m = \frac{m}{2}\log m + O\left(\frac{\log m}{\log\log m}\right). \tag{4.1}$$

We also remark that the second difference of y_m is unbounded from below. In fact, the inequality

$$y_{m+1} - 2y_m + y_{m-1} \leqslant -\log m + 4\log\log m \tag{4.2}$$

holds infinitely often. Both of these assertions are easy consequences of the fact that when the digitaddition series goes past $2^n - 1$, the number of ones in the binary representations of the y_m drops precipitously. We omit the details. Much more than the negation of (4.1) is proved below.

Some open questions: (1) Is $|y_m - (m/2)\log m|/m$ unbounded? (2) Is $B(y_{m+1}) - B(y_m)$ unbounded from *above* as $m \to \infty$? (3) Does the second difference of a digitaddition sequence attain every integer value infinitely often? It is also of interest to determine whether the answers to these questions depend on the choice of y_1. It is conceivable [2], [3], [8] that for any two digitaddition sequences $y_1 < y_2 < \ldots$ and $y_1' < y_2' < \ldots$ there exists an integer k depending only on y_1 and y_1' such that $y'_{n+k} = y_n$ for n sufficiently large.

In connection with question (1) we remark that the error term of Theorem 2 is in fact $\Omega(m^{1-\varepsilon})$ for any $\varepsilon > 0$. This was pointed out by Paul Erdös; the main idea of its demonstration which follows is also due to Professor Erdös.

The proof of Theorem 2 is valid, with no essential change, for any recursion of the form

$$y_{n+1} = y_n + B(y_n) + E(y_n) \tag{4.4}$$

provided $E(x) = O[(\log x \log\log x)^{1/2}]$. We only need this fact for $E(x) \equiv 1$. For $\varepsilon > 0$ and n large, define

$$k = \left[n^{-1}2^{n(1-\varepsilon)}\right] \quad \text{and} \quad m = \left[n^{-1}2^{n+1}(1 + n^{-0.1})\right].$$

A direct application of Theorem 2 yields

$$2^n < y_m < y_{1.1m} < 2^{n+1}. \tag{4.5}$$

Thus for $h < .1m$ we have that $y_{m+h} = 2^n + z_h$ where $y_m = 2^n + z_0$ and

$$z_{h+1} = z_h + B(z_h) + 1 \qquad (h \geqslant 1). \tag{4.6}$$

Assume that Theorem 2 is valid with an error term $O(m^{1-\varepsilon})$. Then

$$\begin{aligned} y_{m+k} - y_m &= ((m+k)/2)\log(m+k) - (m/2)\log m + O(m^{1-\varepsilon}) \\ &> (k/2)\log m + O(m^{1-\varepsilon}) \qquad (4.7) \\ &= \tfrac{1}{2}2^{n(1-\varepsilon)} + O(2^{n(1-\varepsilon)}n^{-1+\varepsilon}). \end{aligned}$$

But by the theorem itself,

$$(4.8)\qquad \begin{aligned} y_{m+k} - y_m &= z_k = (k/2)\log k + O\left(k(\log k)^{3/4}\right) \\ &= ((1-\varepsilon)/2)2^{n(1-\varepsilon)} + O\left(2^{n(1-\varepsilon)}n^{-1/4}\right), \end{aligned}$$

and this contradicts (4.7).

In connection with question (3), we remark that if $y_1 = n$, then the sequence of second differences begins with $g(n)$, where

$$(4.3)\qquad g(n) = B(n + B(n)) - B(n),$$

and that we have the following

PROPOSITION. *Given an integer a, there are infinitely many positive integers n such that $g(n) = a$.*

PROOF. If $a = 0$ let $n = 2^q + 2$ where $q \geqslant 3$. If $a \geqslant 1$, set $p = 2^a - 1$ and $n = 2^{m_1} + \cdots + 2^{m_{p-1}} + 2^p$ where $m_1 > m_2 > \cdots > m_{p-1} > p$. If $a < 0$ set $q = |a| + 1$, $p = 2^q - q$, $r = 2q$, and $n = 2^{m_1} + \cdots + 2^{m_p} + 2^r - 2^q$ where $m_1 > m_2 > \cdots > m_p > r$.

REFERENCES

1. N. Agronomof, *Problem* 4421, Interméd. Math. **21** (1914), 147.
2. A. Auric, Interméd. Math. **22** (1915), 215.
3. H. Brocard, Interméd. Math. **22** (1915), 134–135.
4. A. Cunningham and others, *Problem* 12074, Math. Quest. Educ. Times (2) **21** (1912), 52–53.
5. H. Chernoff, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Ann. Math. Statist. **23** (1952), 493–507. MR **15**, 241.
6. P. Erdös and J. Spencer, *Probabilistic methods in combinatorics*, Academic Press, New York; Akad. Kiadó, Budapest, 1974.
7. M. Gardner, *Mathematical games*, Sci. Amer. **232** (1975), 113–114.
8. A. Gérardin, Sphinx-Oedipe **1** (1906), 19, 47–48.
9. R. Goormaghtigh, *A bit of numerology*, Scripta Math. **15** (1949), 91.
10. V. S. Joshi, *A note on self-numbers*, Math. Student **39** (1971), 327–328 (1972). MR **48** #8371.
11. D. R. Kaprekar, *Cycles of recurring decimals*. Vols. I, II, Published by the author, 311 Devlali Camp, Devlali, India, 1953. MR **14**, 247; **15**, 506.
12. ———, *Self-numbers*, Scripta Math. **22** (1956), 80–81.
13. ———, *Puzzles of the self-numbers*, Published by the author, 311 Devlali Camp, Devlali, India, 1959. MR **20** #6381.
14. ———, *The mathematics of the new self-numbers*, Published by the author, 311 Devlali Camp, Devlali, India, 1963.
15. A. Makowski, *On Kaprekar's "junction numbers"*, Math. Student **34** (1966), **77** (1967), MR **36** #6340.
16. A. Narasinga Rao, *On a technique for obtaining numbers with a multiplicity of generators*, Math. Student **34** (1966), 79–84 (1967). MR **37** #5147.
17. B. Recamán S., *Solution to problem* E 2408 [1973, 434], Amer. Math. Monthly **81** (1974), 407.
18. A. M. Vaidya, *On Kaprekar's tests for self numbers*, Math. Student **37** (1969), 212–214. MR **42** #186.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, URBANA, ILLINOIS 61801