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 \( y_1 < y_2 < y_3 < \ldots \), where \( y_1 \) is a given positive integer and \( y_{n+1} = y_n + B(y_n) \) for \( n = 1, 2, \ldots \). 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 \( y_1 < y_2 < y_3 < \ldots \) is called a "digitaddition series" if
\[
y_{n+1} = y_n + B(y_n), \quad n = 1, 2, \ldots
\]

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
\[
S(n) = S(n; y_1) = \sum_{m=1}^{n} y_m.
\]

We prove
\[
S(n) \sim (n^2/4)(\log n)/(\log 2),
\]
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
\[
\sum_{j>(T/2)+\lambda} \binom{T}{j} < 2T \exp(-2\lambda^2/T);
\]
see [6, p. 17] or [5].

Received by the editors July 25, 1975.


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.

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

Since

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

Theorem 1 can be deduced easily from the following result.

Theorem 2.

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

In particular, $y_m \sim \frac{m}{2} \log m$.

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

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

The trivial bound $B(x) \leq 1 + \lfloor \log x \rfloor$, where $\lfloor z \rfloor$ denotes the greatest integer in $z$, yields

(3.2) $y_{m+1} \leq y_1 + m + \log(y_1, y_2, \ldots, y_m)$.

The trivial bound $B(x) \leq x$, together with (1.1), yields $y_m \leq 2^my_1$. Thus, from (3.2), we find that

(3.3) $y_{m+1} \leq m^2$

for $m$ sufficiently large, say $m > M$. By (3.4) and (3.2) again, we obtain

(3.4) $y_{m+1} \leq y_1 + m + \log(y_1, \ldots, y_M) + \log(m!)^2 \leq 3m \log m$

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

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

(3.5) $\lfloor t/\log t \rfloor = m$.

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

(3.6) $1 \leq y_i \leq y_m \leq 3t$ for $1 \leq i \leq m$.

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

(3.7) $u = T/2 + \lambda$.

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

(3.8) $B(y) \geq u$.

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

\[ \lambda = (T/2)^{1/2} \left\{ \log(\log^2 t) \right\}^{1/2}. \]

Thus

\[ s < 6t/\log^2 t \]

and from (3.1) we have

\[
y_m < y_1 + u \{ m - 1 - s \} + Ts
\]

\[ = y_1 + \left\{ \frac{\log t}{2} + O \left( \{\log t \log \log t\}^{1/2} \right) \right\} \left\{ \frac{t}{\log t} + O \left( \frac{t}{\log^2 t} \right) \right\}.
\]

We conclude that

\[ y_m < t/2 + O \left( t(\log t)^{-1/2}(\log \log t)^{1/2} \right). \]

From (3.5) it is easy to obtain

\[ m \log m < t < m \log m + O \left( m \log \log m \right). \]

Hence

\[ y_m < (m/2) \log m + O \left( m(\log m \log \log m)^{1/2} \right). \]

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

\[ u = T/2 - \lambda \]

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

\[ B(y) < u. \]

Then (note that \( J = (T, T) \)) we have

\[ s \leq \sum_{j < u} \left( \begin{array}{c} T \\ j \end{array} \right) < 6t \exp\left\{ -2\lambda^2 / T \right\}. \]

By choosing \( \lambda \) exactly as before, we obtain

\[
y_m \geq u \{ m - 1 - s \}
\]

\[ = \left\{ \frac{\log t}{2} + O \left( \{\log t \log \log t\}^{1/2} \right) \right\} \left\{ \frac{t}{\log t} + O \left( \frac{t}{\log^2 t} \right) \right\}.
\]

We conclude from (3.19) and (3.14) that

\[ y_m \geq t/2 + O \left( t(\log t)^{-1/2}(\log \log t)^{1/2} \right) \]

and
This completes the proof.

4. Remarks. Theorem 2 cannot be improved to

\begin{equation}
    y_m = \frac{m}{2} \log m + O\left( \frac{\log m \log \log m}{\log \log \log m} \right).
\end{equation}

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

\begin{equation}
    y_{m+1} - 2y_m + y_{m-1} \leq -\log m + 4 \log \log m
\end{equation}

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 $\Omega(m^{1-\epsilon})$ for any $\epsilon > 0$. This was pointed out by Paul Erdős; the main idea of its demonstration which follows is also due to Professor Erdős.

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

\begin{equation}
    y_{n+1} = y_n + B(y_n) + E(y_n)
\end{equation}

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

$$
    k = \left[ n^{-12n^{(1-\epsilon)}} \right] \quad \text{and} \quad m = \left[ n^{-12n^{+1}}(1 + n^{-0.1}) \right].
$$

A direct application of Theorem 2 yields

\begin{equation}
    2^n < y_m < y_{1.1m} < 2^{n+1}.
\end{equation}

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

\begin{equation}
    z_{h+1} = z_h + B(z_h) + 1 \quad (h \geq 1).
\end{equation}

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

\begin{equation}
    y_{m+k} - y_m = ((m + k)/2)\log(m + k) - (m/2)\log m + O(m^{1-\epsilon})
\end{equation}

\begin{equation}
    > (k/2)\log m + O(m^{1-\epsilon})
\end{equation}

\begin{equation}
    = \frac{1}{2} 2^{n^{(1-\epsilon)}} + O(2^{n^{(1-\epsilon)}}n^{-1+\epsilon}).
\end{equation}

But by the theorem itself,
\[ y_{m+k} - y_m = z_k = \left(\frac{k}{2}\right)\log k + O\left(k(\log k)^{3/4}\right) \]
\[ = \left(\frac{1 - \varepsilon}{2}\right)2^{n(1-\varepsilon)} + O\left(2^{n(1-\varepsilon)n^{-1/4}}\right), \]
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
\[ g(n) = B(n + B(n)) - B(n), \]
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 \( q \geq 3 \). If \( a \geq 1 \), 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 \( n = 2^{m_1} + \cdots + 2^{m_r} + 2^r - 2^q \) where \( m_1 > m_2 > \cdots > m_r > r \).

**References**


**Department of Mathematics, University of Illinois, Urbana, Illinois 61801**