AN EXTENSION OF A THEOREM BY CHEO AND YIEN CONCERNING DIGITAL SUMS

Robert E. Kennedy and Curtis N. Cooper
Central Missouri State University, Warrensburg, MO 64093
(Submitted June 1989)

1. Introduction

For a nonnegative integer k, let $s(k)$ denote the digital sum of k. In [1], Cheo and Yien prove that, for a nonnegative integer x,

$$(1.1)\quad \sum_{k=0}^{x-1} s(k) = (4.5)x \log x + O(x).$$

Here $O(f(x))$ is the useful "big-oh" notation and denotes some unspecified function $g(x)$ such that $g(x)/f(x)$ is eventually bounded. We usually write

$$g(x) = O(f(x))$$

and read, "$g(x)$ is big-oh of $f(x)$." For an introduction to this important notation, see [3]. In this paper we determine a formula for

$$(1.2)\quad \sum_{k=0}^{x-1} (s(k))^2.$$

The resulting formula will be used to calculate the mean and variance of the sequence of digital sums. First, in order to facilitate the discussion, we introduce some notation.

2. Notation

For each positive integer x, let $[0, x)$ denote the set of nonnegative integers strictly less than x. In addition, we will let $s([0, x))$ be the sequence

$$(2.1)\quad s(0),\ s(1),\ s(2),\ \ldots,\ s(x-1).$$

That is, we have not only taken into account $s(k)$, but also the frequency of $s(k)$. Then, letting μ and σ^2 be the mean and variance of $s([0, x))$, respectively, we have

$$\mu = \frac{1}{x}\sum_{k=0}^{x-1} s(k)$$

and

$$(2.2)\quad \sigma^2 = \frac{1}{x}\sum_{k=0}^{x-1} (s(k))^2 - \mu^2.$$

If x is a power of 10, then the following lemma gives the exact value of μ and σ^2. Its proof is given in [2] and will not be reproduced here.

Lemma 2.1: Let $x = 10^n$ for a positive integer n. Then

$$\mu = \text{the mean of } s([0, x)) = 4.5n$$

and

$$\sigma^2 = \text{the variance of } s([0, x)) = 8.25n.$$

3. A Formula for (1.2)

The following theorem gives a formula for the expression (1.2).

Theorem 3.1: Let s be the digital sum function and let x be a positive integer. Then

$$\sum_{k=0}^{x-1}(s(k))^2 = 20.25x \log^2 x + O(x \log x). \tag{3.1}$$

Proof: For each positive integer x, define

$$A(x) = \sum_{k=0}^{x-1} s(k) \quad \text{and} \quad B(x) = \sum_{k=0}^{x-1}(s(k))^2.$$

In [1], Cheo and Yien showed that for any positive integer n and for any decimal digit $c(0, 1, 2, 3, \ldots, \text{or } 9)$,

$$A(c \cdot 10^n) = \left(4.5cn + \frac{c(c-1)}{2}\right)10^n. \tag{3.2}$$

Using Lemma 2.1 and formula (2.2) for the variance of $s([0, x))$, we have

$$B(10^n) = 20.25n^2 10^n + 8.25n10^n. \tag{3.3}$$

Therefore, for a positive integer n and a decimal digit c, we can calculate $B(c \cdot 10^n)$. That is,

$$B(c \cdot 10^n) = \sum_{k=0}^{c \cdot 10^n - 1}(s(k))^2.$$

Since for $0 \le k < 10^n$ and $0 \le j \le c - 1$ we have that $s(j \cdot 10^n + k) = j + s(k)$. Thus, the above single sum can be rewritten as a double sum

$$B(c \cdot 10^n) = \sum_{j=0}^{c-1}\sum_{k=0}^{10^n-1}(j + s(k))^2 = \sum_{j=0}^{c-1}\sum_{k=0}^{10^n-1}(j^2 + 2js(k) + (s(k))^2)$$

$$= \sum_{j=0}^{c-1} j^2 10^n + 2\sum_{j=0}^{c-1} j\left(\sum_{k=0}^{10^n-1} s(k)\right) + \sum_{k=0}^{10^n-1} c(s(k))^2.$$

We may now apply (3.2) and (3.3) to obtain

$$B(c \cdot 10^n) = \frac{(c-1)c(2c-1)}{6}10^n + 2\sum_{j=0}^{c-1} j(4.5)n10^n + c(8.25n + 20.25n^2)10^n.$$

Continuing, we have

$$B(c \cdot 10^n) = \frac{(c-1)c(2c-1)}{6}10^n + (c-1)c(4.5n)10^n + c(20.25n^2 + 8.25n)10^n. \tag{3.4}$$

Since

$$\sum_{i=0}^{n-1} i10^i = \frac{1}{9^2}(10^n(9n - 10) + 10) \tag{3.5}$$

and

$$\sum_{i=0}^{n-1} i^2 10^i = \frac{1}{9^3}(10^n(81n^2 - 180n + 110) - 110), \tag{3.6}$$

we can now prove (3.1). Let

$$x = a_n 10^{n-1} + a_{n-1}10^{n-2} + \cdots + a_1 10^0$$

be the decimal representation of the nonnegative integer x. Then

$$B(x) = \sum_{k=0}^{x-1} (s(k))^2.$$

Similarly, as in the determination of $B(c \cdot 10^n)$, this single sum can be written as the following sum of single sums

$$\begin{aligned}
B(x) &= \sum_{k=0}^{a_n 10^{n-1}-1} (s(k))^2 + \sum_{k=0}^{a_{n-1}10^{n-2}-1} (a_n + s(k))^2 \\
&\quad + \cdots + \sum_{k=0}^{a_1 10^0 - 1} (a_n + a_{n-1} + \cdots + a_2 + s(k))^2 \\
&= \sum_{i=1}^{n} B(a_i 10^{i-1}) + 2\sum_{k=1}^{n-1}\left(\sum_{i=k+1}^{n} a_i\right)A(a_k 10^{k-1}) + \sum_{k=1}^{n-1}\left(\sum_{i=k+1}^{n} a_i\right)^2 a_k 10^{k-1}.
\end{aligned}$$

Using (3.4), we have that

$$B(x) = t_1 - t_2 + t_3 + t_4 + t_5 + t_6,$$

where

$$t_1 = 20.25(n-1)^2 x,$$

$$t_2 = 20.25\sum_{i=1}^{n}((n-1)^2 - (i-1)^2)a\ 10^{i-1},$$

$$t_3 = \sum_{i=1}^{n}(4.5a_i^2 + 3.75a_i)(i-1)10^{i-1},$$

$$t_4 = \sum_{i=1}^{n} \frac{(a_i - 1)(a_i)(2a_i - 1)}{6} 10^{i-1},$$

$$t_5 = \sum_{k=1}^{n-1}\left(\sum_{i=k+1}^{n} a_i\right)^2 a_k 10^{k-1},$$

and

$$t_6 = 2\sum_{k=1}^{n-1}\left(\sum_{i=k+1}^{n} a_i\right)A(a_k 10^{k-1}).$$

It can be shown without difficulty that

$$t_3 = O(x \log x), \quad t_4 = O(x),$$

and since the calculation of t_5 and t_6 are similar, only t_6 will be calculated here. Thus,

$$\begin{aligned}
t_6 &= 2\sum_{k=1}^{n-1}\left(\sum_{i=k+1}^{n} a_i\right)A(a_k 10^{k-1}) \le 18\sum_{k=1}^{n-1}(n-k)A(a_k 10^{k-1}) \\
&= 18\sum_{k=1}^{n-1}(n-k)\left(4.5a\ (k-1) + \frac{a_k(a_k-1)}{2}\right)10^{k-1}
\end{aligned}$$

by (3.2). Hence,

$$\begin{aligned}
(3.7)\qquad t_6 &\le 729\sum_{k=1}^{n-1}(-k^2 + (n+1)k - n)10^{k-1} + 729\sum_{k=1}^{n-1}(n-k)10^{k-1} \\
&= 729\left(\frac{-1}{10}\sum_{k=1}^{n-1}k^2 10^k + \frac{n}{10}\sum_{k=1}^{n-1}k10^k\right),
\end{aligned}$$

and using (3.5) and (3.6) we have, after simplification,

$$t_6 \le 9n10^n - 11 \cdot 10^n + 9n + 11,$$

and so

$$t_6 = O(x \log x),$$

since $n = O(\log x)$ and $10^n = O(x)$. Similarly,

$$t_5 = O(x).$$

Thus $t_2 = O(x \log x)$ follows, since

$$\begin{aligned} t_2 &= 20.25 \sum_{i=1}^{n} ((n-1)^2 - (i-1)^2) a_i 10i - 1 \\ &\le (20.25)(n-1)^2(10^n - 1) - (20.25)(9) \sum_{i=1}^{n-1} i^2 10^i, \end{aligned}$$

and so, by (3.6),

$$t_2 \le (20.25)(n-1)^2(10^n - 1) - \frac{1}{4}(10^n(81n^2 - 180n + 110) - 110)$$

and, after simplification, we obtain

$$t_2 = O(x \log x).$$

Therefore,

$$\begin{aligned} B(x) &= 20.25(n-1)\ x - O(x \log x) + O(x \log x) + O(x) \\ &\qquad + O(x) + O(x \log x) \\ &= 20.25(n-1)^2 x + O(x \log x) \\ &= 20.25 x \log^2 x + O(x \log x), \end{aligned}$$

since $n = \log x + O(1)$, and (3.1) has been proven.

Using (1.1), we have an immediate corollary to Theorem 3.1.

Corollary 3.2: Let s be the digital sum function and x be a positive integer. Then

$$\mu = 4.5 \log x + O(1) \tag{3.8}$$

and

$$\sigma^2 = O(\log x). \tag{3.9}$$

Proof: The proof of (3.8) follows immediately from (1.1) by dividing through by x. To prove (3.9), we use (3.8), (3.7), and (2.2) to obtain

$$\sigma^2 = 20.25 \log^2 x + O(\log x) - (4.5 \log x + O(1))^2.$$

This implies that

$$\sigma^2 = O(\log x).$$

4. Conclusion

In [1], Cheo and Yien proved that $O(x)$ is the best possible residual for the relation (1.1). Here, the second term of (3.3) implies that $O(x \log x)$ is the best residual given by (3.1). It should also be mentioned here that we have restricted our discussion to base ten numbers. Cheo and Yien, however, give their results for any positive integer base greater than one. By substituting base 10 by base b, the base 10 digit 9 by $b - 1$, and base 10 logarithms by base b logarithms, the results of this paper can, in like manner, be given without restricting base. Finally, we note that the determination of a formula for

$$\sum_{k=0}^{x-1} (s(k))^n, \text{ for } n \geq 3,$$

appears to be complicated and is left as an open problem.

References

1. P. Cheo & S. Yien. "A Problem on the K-adic Representation of Positive Integers." *Acta Math. Sinica* 5 (1955):433-38.
2. R. Kennedy & C. Cooper. "On the Natural Density of the Niven Numbers." *College Math. Journal* 15 (1984):309-12.
3. D. E. Knuth. *The Art of Computer Programming,* Vol. I. New York: Addison-Wesley Publishing Company, 1969.

*****