AN EXPLICIT EXPRESSION FOR LARGE DIGIT SUMS IN BASE B EXPANSIONS

Curtis Cooper and Robert E. Kennedy
Department of Mathematics, Central Missouri State University
Warrensburg, MO 64093-5045

Abstract. Let n and b be positive integers. We will present a formula for the number of large digits ($\lceil b/2 \rceil$ or more) in the base b representation of the sequence of positive integers less than n.

AMS Classification number: 11A63

1. Introduction. Let $s_b(i)$ denote the sum of the digits in the base b representation of the nonnegative integer i and $L_b(i)$ denote the number of large digits ($\lceil b/2 \rceil$ or more) in the base b representation of the nonnegative integer i. Bush [1] showed that

$$\sum_{n<x} s_b(n) \sim \frac{b-1}{2} x \log_b x.$$

Here, $\log_b x$ denotes the base b logarithm of x. Mirsky [4], and later Cheo and Yien [2], proved that

$$\sum_{n<x} s_b(n) = \frac{b-1}{2} x \log_b x + O(x).$$

Trollope [5] discovered the following result. Let $g(x)$ be periodic of period one and defined on $[0, 1]$ by

$$g(x) = \begin{cases} \frac{1}{2}x, & 0 \le x \le \frac{1}{2} \\ \frac{1}{2}(1-x), & \frac{1}{2} < x \le 1, \end{cases}$$

and let

$$f(x) = \sum_{i=0}^{\infty} \frac{1}{2^i} g(2^i x).$$

Now, if $n = 2^m(1+x)$, $0 \le x < 1$, then

$$\sum_{i<n} s_2(i) = \frac{1}{2} n \log_2 n - E_2(n),$$

where

$$E_2(n) = 2^{m-1}\Big(2f(x) + (1+x)\log_2(1+x) - 2x\Big).$$

We will discuss some similar results for L_b.

2. Main Term and Big-Oh Term. Our first result will be to give a main term and big-Oh term for a sum involving L_b. We will parallel a proof of Mirsky [4].

Theorem 1.

$$\sum_{n<x} L_b(n) = \frac{\lfloor \frac{b}{2} \rfloor}{b} x \log_b x + O(x).$$

Proof. We begin by observing that if $k \geq 0$ and $0 \leq d \leq b-1$, then the representation of n in base b contains the term db^k if and only if n can be expressed in the form

$$n = mb^{k+1} + \mu,$$

where $m \geq 0$ and $db^k \leq \mu < (d+1)b^k$. Hence $f(x,k,d)$, the number of positive integers not exceeding x whose representation in base b contains the term db^k, is given by

$$\begin{aligned} f(x,k,d) &= \sum_{\substack{mb^{k+1}+\mu \leq x \\ m \geq 0; db^k \leq \mu < (d+1)b^k}} 1 \\ &= \sum_{db^k \leq \mu < (d+1)b^k} \left(\frac{x}{b^{k+1}} + O(1) \right) = \frac{x}{b} + O(b^k). \end{aligned}$$

But clearly

$$\sum_{n<x} L_b(n) = \sum_{\substack{0 \leq k \leq \log_b x \\ 0 \leq d \leq b-1}} \left\lfloor \frac{d}{\lceil \frac{b}{2} \rceil} \right\rfloor f(x,k,d).$$

Therefore, by the formula for $f(x,k,d)$,

$$\begin{aligned} \sum_{n<x} L_b(n) &= \sum_{\substack{0 \leq k \leq \log_b x \\ 0 \leq d \leq b-1}} \left\lfloor \frac{d}{\lceil \frac{b}{2} \rceil} \right\rfloor \left(\frac{x}{b} + O(b^k) \right) \\ &= \frac{\lfloor \frac{b}{2} \rfloor}{b} x \log_b x + \left\lfloor \frac{b}{2} \right\rfloor O\left(\frac{1 - b^{\log_b x + 1}}{1-b} \right) \\ &= \frac{\lfloor \frac{b}{2} \rfloor}{b} x \log_b x + O(x). \end{aligned}$$

To show that this is best possible, we consider the sequence $x_N = b^{N+1} + b^N$. We have that

$$\begin{aligned}\sum_{n<x_N} L_b(n) &= \frac{\lfloor \frac{b}{2} \rfloor}{b}(N+1)b^{N+1} + \frac{\lfloor \frac{b}{2} \rfloor}{b} N b^N \\ &= (b^{N+1} + b^N)\frac{\lfloor \frac{b}{2} \rfloor}{b} N + b^{N+1}\frac{\lfloor \frac{b}{2} \rfloor}{b}\end{aligned}$$

and

$$\begin{aligned}&\frac{\lfloor \frac{b}{2} \rfloor}{b} x_N \log_b x_N \\ &= \frac{\lfloor \frac{b}{2} \rfloor}{b}(b^{N+1} + b^N)N + \frac{\lfloor \frac{b}{2} \rfloor}{b}(b^{N+1} + b^N)\log_b(b+1).\end{aligned}$$

Thus,

$$\begin{aligned}&\sum_{n<x_N} L_b(n) - \frac{\lfloor \frac{b}{2} \rfloor}{b} x_N \log_b x_N \\ &= b^{N+1}\frac{\lfloor \frac{b}{2} \rfloor}{b} - \frac{\lfloor \frac{b}{2} \rfloor}{b}(b^{N+1} + b^N)\log_b(b+1) + b^N\frac{\lfloor \frac{b}{2} \rfloor}{b} - b^N\frac{\lfloor \frac{b}{2} \rfloor}{b} \\ &= \frac{\lfloor \frac{b}{2} \rfloor}{b}(1 - \log_b(b+1))(b^{N+1} + b^N) - b^N\frac{\lfloor \frac{b}{2} \rfloor}{b}\frac{b+1}{b+1} \\ &= \frac{\lfloor \frac{b}{2} \rfloor}{b}(b^{N+1} + b^N)\left(1 - \log_b(b+1) - \frac{1}{b+1}\right).\end{aligned}$$

Hence,

$$\sum_{n<x_N} L_b(n) = \frac{\lfloor \frac{b}{2} \rfloor}{b} x_N \log_b x_N + c x_N,$$

where

$$c = \frac{\lfloor \frac{b}{2} \rfloor}{b}\left(1 - \log_b(b+1) - \frac{1}{b+1}\right).$$

3. Notation and Basic Results. We next present some notation. The first idea is due to Kenneth Iverson, the creator of the programming language APL, and is discussed in [3]. Suppose that k is an integer and $P(k)$ is some statement about k which is either true or false. Then

$$[P(k)] = \begin{cases} 1, & P(k) \text{ is true} \\ 0, & P(k) \text{ is false.} \end{cases}$$

Second, to make some of the results easier to state, we will use the notation

$$L(n) = \sum_{i<n} L_b(i).$$

Our first result is a statement about the number of base b large digits in the sequence of positive integers up to a digit times a power of b. The proof of this formula follows from a straightforward counting argument and will be omitted.

Lemma 1. Let d be a nonzero digit and m a nonnegative integer. Then

$$L(d \cdot b^m) = [d > \lceil b/2 \rceil] \cdot (d - \lceil b/2 \rceil) b^m + \lfloor b/2 \rfloor d m b^{m-1}.$$

Next, let n be a positive integer with base b representation

$$n = \sum_{k=0}^{m} d_k b^k.$$

Also, let

$$n_i = \sum_{k=0}^{i} d_k b^k, \quad \text{for} \quad i \geq 0; \quad n_{-1} = 0.$$

Now, we make the important observation that if $n = d_m b^m + n_{m-1}$, then

$$L(d_m b^m + n_{m-1}) = L(d_m b^m) + [d_m \geq \lceil b/2 \rceil] \cdot n_{m-1} + L(n_{m-1}).$$

Hence, using mathematical induction on the number of digits in n, the above equation, and Lemma 1, we have the following more general result.

Lemma 2. Let n be a positive integer with base b representation

$$n = \sum_{k=0}^{m} d_k b^k$$

and define

$$n_i = \sum_{k=0}^{i} d_k b^k, \quad \text{for} \quad i \geq 0; \quad n_{-1} = 0.$$

Then

$$L(n) = \lfloor b/2 \rfloor \sum_{k=0}^{m} d_k k b^{k-1} + \sum_{k=0}^{m} [d_k > \lceil b/2 \rceil] \cdot (d_k - \lceil b/2 \rceil) b^k + \sum_{k=0}^{m} [d_k \geq \lceil b/2 \rceil] \cdot n_{k-1}.$$

4. The Remainder Term. The next step in analyzing $L(n)$ is to study it from a different perspective. Let

$$R(n) = \frac{\lfloor \frac{b}{2} \rfloor}{b} mn - \lfloor b/2 \rfloor \sum_{k=0}^{m} d_k k b^{k-1}$$
$$- \sum_{k=0}^{m-1} [d_k > \lceil b/2 \rceil] \cdot (d_k - \lceil b/2 \rceil) b^k - \sum_{k=0}^{m-1} [d_k \geq \lceil b/2 \rceil] \cdot n_{k-1}.$$

The next lemma will state some properties of R, which will be extremely useful throughout the rest of the paper.

<u>Lemma 3</u>.

(a) For any positive integer n, $R(bn) = bR(n)$.

(b) Let n be a positive integer with base b representation

$$n = \sum_{k=0}^{m} d_k b^k.$$

Then

$$R(n+1) - R(n) = \frac{\lfloor \frac{b}{2} \rfloor}{b} m + [d_m \geq \lceil b/2 \rceil] - L_b(n).$$

(c) Let m be a nonnegative integer, d a digit, and p an integer such that $0 \leq p < (b-1) \cdot b^m$. Then

$$R(b^{m+1} + bp + d) - d \cdot R(b^m + p + 1) - (b - d) \cdot R(b^m + p)$$
$$= \frac{\lfloor \frac{b}{2} \rfloor}{b} d - [d > \lceil b/2 \rceil](d - \lceil b/2 \rceil).$$

<u>Proof</u>. The proof of (a) involves a fairly easy, but tedious, derivation using the definition of $R(n)$. In passing, we note that using Lemma 3(a) and the fact that

$$L(n) = \frac{\lfloor \frac{b}{2} \rfloor}{b} mn + [d_m > \lceil b/2 \rceil] \cdot (d_m - \lceil b/2 \rceil) b^m$$
$$+ [d_m \geq \lceil b/2 \rceil] \cdot n_{m-1} - R(n),$$

it is immediate that

$$L(bn) = bL(n) + \lfloor b/2 \rfloor n$$

for all $n \geq 1$.

The proof of (b) follows from scrutinizing three cases. The first case is when n and $n+1$ have a different number of digits. Thus, $n = b^{m+1} - 1$. The second case is when n and $n+1$ have the same number of digits but have a different first digit. Thus $n = d \cdot b^m - 1$ for $d = 2, 3, \ldots, b-1$. The third case is the rest of the story, i.e., when n and $n+1$ have the same number of digits and the same first digit. In every case, we have that

$$R(n+1) - R(n) = \frac{\lfloor \frac{b}{2} \rfloor}{b} m + [d_m \geq \lceil b/2 \rceil] - L_b(n).$$

The proof of (c) is a little more involved. Using Lemma 3(b) twice, Lemma 3(a) once, and the assumption that the base b representation of $b^m + p$ is

$$b^m + p = \sum_{k=0}^{m} d_k b^k,$$

we have the following sequence of equalities.

$$\begin{aligned}
&R(b^{m+1} + bp + d) - dR(b^m + p + 1) - (b-d)R(b^m + p) \\
&= R(b^{m+1} + bp + d) - d\big(R(b^m + p + 1) - R(b^m + p)\big) - bR(b^m + p) \\
&= R(b^{m+1} + bp + d) - R(b^{m+1} + bp) - d\left(\frac{\lfloor \frac{b}{2} \rfloor}{b} m + [d_m \geq \lceil b/2 \rceil] - L_b(b^m + p)\right) \\
&= \sum_{k=0}^{d-1} \big(R(b^{m+1} + bp + k + 1) - R(b^{m+1} + bp + k)\big) - d\frac{\lfloor \frac{b}{2} \rfloor}{b} m \\
&- [d_m \geq \lceil b/2 \rceil]d + dL_b(b^m + p) \\
&= \sum_{k=0}^{d-1} \left(\frac{\lfloor \frac{b}{2} \rfloor}{b}(m+1) + [d_m \geq \lceil b/2 \rceil] - L_b(b^{m+1} + bp + k)\right) - d\frac{\lfloor \frac{b}{2} \rfloor}{b} m \\
&- [d_m \geq \lceil b/2 \rceil]d + dL_b(b^m + p) \\
&= d\frac{\lfloor \frac{b}{2} \rfloor}{b} m + d\frac{\lfloor \frac{b}{2} \rfloor}{b} + d[d_m \geq \lceil b/2 \rceil] - \sum_{k=0}^{d-1} L_b(b^{m+1} + bp + k) - d\frac{\lfloor \frac{b}{2} \rfloor}{b} m \\
&- [d_m \geq \lceil b/2 \rceil]d + dL_b(b^m + p) \\
&= d\frac{\lfloor \frac{b}{2} \rfloor}{b} - [d > \lceil b/2 \rceil](d - \lceil b/2 \rceil).
\end{aligned}$$

This completes the proof of (c) and the proof of Lemma 3.

5. Some Functions. Let m be a nonnegative integer and p be an integer such that $0 \le p < (b-1)b^m$. Define the function $\phi(x)$ by

$$\phi\left(\frac{p}{(b-1)b^m}\right) = \frac{R(b^m + p)}{b^m}.$$

Note that by Lemma 3(a), $\phi(x)$ is uniquely defined. For if x has any other representation, i.e.

$$\frac{p'}{(b-1)b^{m'}},$$

then

$$p' = b^{m'-m}p.$$

If we assume, without loss of generality, that $m' > m$, then $m' - m$ is a positive integer. Therefore,

$$\frac{R(b^{m'} + p')}{b^{m'}} = \frac{R(b^m + p)}{b^m}.$$

The function $\phi(x)$ is defined only on a subset of $[0, 1]$. We now consider the problem of extending this function continuously to $[0, 1]$. Here, we solve this problem by considering the limit of a sequence of "polygonal" functions which identify with $\phi(x)$ on the rationals of the form

$$\frac{p}{(b-1)b^m}.$$

These polygonal functions are defined in the following way. Let m be a non-negative integer. $f_m(x)$ is defined on $[0, 1]$ to be the function whose graph is the polygon joining the points

$$\left\{(0,0), \left(\frac{1}{(b-1)b^m}, \phi\left(\frac{1}{(b-1)b^m}\right)\right), \right.$$
$$\left. \cdots, \left(\frac{p}{(b-1)b^m}, \phi\left(\frac{p}{(b-1)b^m}\right)\right), \cdots, (1,0)\right\}.$$

Then, the definition of $\{f_m(x)\}$ is extended to the reals by $f_m(x \pm 1) = f_m(x)$. From the definition, $f_0 = 0$. In addition, $f_1(x)$ is equal to the auxiliary function $g(x)$, which is defined on the $[0, 1/(b-1)]$ by

$$g(x) = \left[0 \le x \le \frac{\lceil b/2 \rceil}{(b-1)b}\right] \frac{(b-1)\lfloor b/2 \rfloor}{b} x$$
$$+ \left[\frac{\lceil b/2 \rceil}{(b-1)b} < x \le \frac{1}{b-1}\right] \frac{(b-1)\lceil b/2 \rceil}{b} \left(\frac{1}{b-1} - x\right)$$

and extended to the reals by $g(x \pm 1/(b-1)) = g(x)$. Furthermore, using Lemma 3(c), it follows that for any nonnegative integer m and all real x,

$$f_{m+1}(x) - f_m(x) = \frac{1}{b^m} g(b^m x).$$

Repeated iterations of this equation yields

$$f_{m+1}(x) = \sum_{i=0}^{m} \frac{1}{b^i} g(b^i x).$$

Since $g(x)$ is bounded, the sequence $\{f_m(x)\}$ converges uniformly for all x. Hence, the limiting function

$$f(x) = \sum_{i=0}^{\infty} \frac{1}{b^i} g(b^i x)$$

is a continuous extension of $\phi(x)$.

6. The Main Result.

<u>Theorem 2</u>. Let n be a positive integer with base b representation

$$n = \sum_{k=0}^{m} d_k b^k.$$

Next, let $n = b^m + p$ and

$$x = \frac{p}{(b-1)b^m}.$$

Finally, let $g(x)$ be periodic of period $1/(b-1)$ and defined by

$$g(x) = \left[0 \le x \le \frac{\lceil b/2 \rceil}{(b-1)b}\right] \frac{(b-1)\lfloor b/2 \rfloor}{b} x + \left[\frac{\lceil b/2 \rceil}{(b-1)b} < x \le \frac{1}{b-1}\right] \frac{(b-1)\lceil b/2 \rceil}{b} \left(\frac{1}{b-1} - x\right)$$

and let

$$f(x) = \sum_{i=0}^{\infty} \frac{1}{b^i} g(b^i x).$$

Then

$$\sum_{i<n} L_b(i) = \frac{\lfloor \frac{b}{2} \rfloor}{b} n \log_b n - E_b(n),$$

where

$$E_b(n) = b^{m-1}\Big(bf(x) + \lfloor b/2\rfloor(1+(b-1)x)\log_b(1+(b-1)x)$$
$$- [d_m \geq \lceil b/2\rceil] \cdot b(1-d_m+(b-1)x) - [d_m > \lceil b/2\rceil] \cdot b(d_m - \lceil b/2\rceil)\Big).$$

Proof.

$$\begin{aligned}L(n) &= \frac{\lfloor\frac{b}{2}\rfloor}{b}mn + [d_m > \lceil b/2\rceil] \cdot (d_m - \lceil b/2\rceil)b^m \\ &+ [d_m \geq \lceil b/2\rceil] \cdot n_{m-1} - R(n) \\ &= \frac{\lfloor\frac{b}{2}\rfloor}{b}mn + [d_m > \lceil b/2\rceil] \cdot (d_m - \lceil b/2\rceil)b^m \\ &+ [d_m \geq \lceil b/2\rceil] \cdot n_{m-1} - b^m f\left(\frac{p}{(b-1)b^m}\right).\end{aligned}$$

Next, since

$$n = b^m + p = b^m\left(1 + \frac{p}{b^m}\right),$$

$$m = \log_b n - \log_b\left(1 + \frac{p}{b^m}\right).$$

Also,

$$n_{m-1} = n - d_m b^m = b^m + p - d_m b^m = b^m\left(1 + \frac{p}{b^m} - d_m\right).$$

Substituting for these three quantities, setting

$$x = \frac{p}{(b-1)b^m},$$

and simplifying, we have

$$L(n) = \frac{\lfloor\frac{b}{2}\rfloor}{b}n\log_b n - \frac{\lfloor\frac{b}{2}\rfloor}{b}n\log_b(1+(b-1)x) - b^m f(x)$$
$$+ [d_m \geq \lceil b/2\rceil] \cdot b^m(1-d_m+(b-1)x) + [d_m > \lceil b/2\rceil] \cdot (d_m - \lceil b/2\rceil)b^m.$$

Thus, we have the result.

7. Questions. Some open questions remain. One problem is to study the function f in Theorem 2. Another problem that is noteworthy is concerning Trollope's result. Trollope's original result is in base 2. What would Trollope's result be for base b?

References

1. L. E. Bush, "An Asymptotic Formula for the Average Sum of Digits of Integers," *American Mathematical Monthly*, 47 (1940), 154–156.
2. P. Cheo and S. Yien, "A Problem on the K-adic Representation of Positive Integers," *Acta Math. Sinica,* 5 (1955), 433–438.
3. R. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, Reading, Massachusetts, Addison-Wesley Publishing Company, 1989, 24–25, 33.
4. L. Mirsky, "A Theorem on Representations of Integers in the Scale of r," *Scripta. Math.*, 15 (1949), 11–12.
5. J. R. Trollope, "An Explicit Expression for Binary Digital Sums," *Mathematics Magazine*, 41 (1968), 21–25.