BOUNDS FOR THE NUMBER OF LARGE DIGITS IN THE POSITIVE INTEGERS NOT EXCEEDING n

Curtis Cooper
Department of Mathematics and Computer Science
Central Missouri State University
Warrensburg, MO 64093
U.S.A.
email: cnc8851@cmsu2.cmsu.edu

Abstract.

Let $l(m)$ denote the number of large digits (5 or bigger) in the base 10 representation of the positive integer m and let

$$L(n) = \sum_{1 \le m < n} l(m),$$

where n is a positive integer. We will show that if n is a positive integer, then

$$\frac{1}{2} n \log n - 0.389n \le L(n) \le \frac{1}{2} n \log n.$$

Here log denotes the base 10 logarithm.

1. Introduction and Result.

Large digits are base 10 digits which are greater than or equal to 5, i.e., 5, 6, 7, 8, and 9. The number of large digits in a positive integer is the number of large digits in its decimal representation. For example, there are no large digits in 34 and there are 2 large digits in 84512. To present our results concerning large digits, we need to use APL notation, which was discussed in [3, p. 24] and originally introduced by Kenneth Iverson. The basic idea is to enclose a true-or-false statement in brackets, and to say that the result is 1 if the statement is true, 0 if the statement is false. For example,

$$[p \text{ prime}] = \begin{cases} 1, & \text{if } p \text{ is a prime number;} \\ 0, & \text{if } p \text{ is not a prime number.} \end{cases}$$

A positive integer m can be uniquely represented in base 10 form as

$$m = \sum_{i=0}^{k} d_i 10^i,$$

where the d_i's are decimal digits for $i = 0, \dots, k$. Here, we assume that $d_k \neq 0$. Now, if m and n are positive integers, define

$$l(m) = \sum_{i=0}^{k} [d_i \geq 5] \qquad \text{and} \qquad L(n) = \sum_{1 \leq m < n} l(m).$$

Now, let log denote the base 10 logarithm. It was shown by Cooper and Kennedy in [1, pp. 26–27] that

$$L(n) = \frac{1}{2} n \log n + O(n).$$

Here, O denotes the big-Oh notation. In what follows we will prove the following theorem which gives us a better idea about the big-Oh term.

<u>Theorem 1</u>. Let n be a positive integer. Then

$$\frac{1}{2} n \log n - 0.389n \leq L(n) \leq \frac{1}{2} n \log n.$$

2. Proof of the Right Inequality of Theorem 1.

To begin the proof of Theorem 1, we need some lemmas which can be found in [1, p. 28].

<u>Lemma 2</u>. Let d be a nonzero decimal digit, k be a nonnegative integer, and m be a nonnegative integer less than 10^k. Then

$$L(d \cdot 10^k) = \frac{1}{2} dk 10^k + [d \geq 5](d - 5)10^k$$

and

$$L(d \cdot 10^k + m) = L(d \cdot 10^k) + [d \geq 5]m + L(m).$$

The following lemma is another one that we will need.

Lemma 3. Let d be a nonzero decimal digit, k be a nonnegative integer, and m be a nonnegative integer less than 10^k. Then

$$[d \geq 5](d-5) \leq \frac{1}{2} d \log d$$

and

$$\frac{2[d \geq 5]\big((d-5)10^k + m\big)}{d \cdot 10^k + m} \leq \log\left(d + \frac{m}{10^k}\right).$$

Proof. The first inequality of Lemma 3 follows by checking all the values of d from 1 to 9. The second inequality of Lemma 3 is true for $d = 1, 2, 3, 4$ since the left-hand side of the inequality is 0. To establish the inequality for $d = 5, 6, 7, 8, 9$ we can simplify the inequality. Therefore, if we can show

$$\frac{2(d \cdot 10^k + m) - 10^{k+1}}{d \cdot 10^k + m} \leq \log\left(d + \frac{m}{10^k}\right)$$

then we are done. Dividing both numerator and denominator of the left-hand side of this last inequality by 10^k, we are done if we can show that

$$\frac{2(d + \frac{m}{10^k}) - 10}{d + \frac{m}{10^k}} \leq \log\left(d + \frac{m}{10^k}\right).$$

If we generalize this result, we would be done if we could show that

$$\frac{2x - 10}{x} \leq \log x$$

for $5 \leq x < 10$. Simplifying the above inequality, we would be done if we could show that

$$2(x - 10) + 10 \leq x \log x$$

for $5 \leq x < 10$. Letting $x \in [5, 10)$, and applying the mean-value theorem to the function

$$f(x) = x \log x$$

we have that

$$\frac{f(x) - f(10)}{x - 10} = f'(\zeta)$$

for $\zeta \in (x, 10)$. But

$$f'(x) = \log e + \log x$$

so

$$\frac{f(x) - f(10)}{x - 10} = \log e + \log \zeta \leq 2.$$

But since $x - 10 < 0$ we have that

$$f(x) - f(10) \geq 2(x - 10).$$

Thus,

$$x \log x \geq 10 + 2(x - 10).$$

This completes the proof.

We are now ready to prove the right inequality of Theorem 1. Let $n = d \cdot 10^k + m$, where d is a nonzero decimal digit, k be a nonnegative integer, and m be a nonnegative integer less than 10^k. We will prove that

$$L(d \cdot 10^k + m) \leq \frac{1}{2}(d \cdot 10^k + m) \log(d \cdot 10^k + m)$$

by induction on the number of nonzero decimal digits in $d \cdot 10^k$ and m.

For the base step, we assume that $d \cdot 10^k$ and m have exactly one nonzero decimal digit, i.e., $m = 0$. Then, by Lemma 2

$$L(d \cdot 10^k + 0) = L(d \cdot 10^k) = \frac{1}{2} dk10^k + [d \geq 5](d - 5)10^k$$

and

$$\begin{aligned} &\frac{1}{2}(d \cdot 10^k + 0) \log(d \cdot 10^k + 0) = \frac{1}{2}(d \cdot 10^k) \log(d \cdot 10^k) \\ =&\frac{1}{2}(d \cdot 10^k)(k + \log d) = \frac{1}{2} dk10^k + \frac{1}{2} d \log d \cdot 10^k. \end{aligned}$$

To prove

$$L(d \cdot 10^k + 0) \leq \frac{1}{2}(d \cdot 10^k + 0) \log(d \cdot 10^k + 0)$$

we need to show that

$$[d \geq 5](d-5) \leq \frac{1}{2} d \log d.$$

But this is the first part of Lemma 3. This completes the base step of the induction.

For the induction step, we assume that our result is true for $d \cdot 10^k$ and m having fewer that j nonzero decimal digits, i.e., if $d \cdot 10^k$ and m have fewer than j nonzero decimal digits then

$$L(d \cdot 10^k + m) \leq \frac{1}{2}(d \cdot 10^k + m) \log(d \cdot 10^k + m).$$

Now suppose $d \cdot 10^k$ and m have j nonzero decimal digits. But, by Lemma 2

$$\begin{aligned} L(d \cdot 10^k + m) &= L(d \cdot 10^k) + [d \geq 5]m + L(m) \\ &= \frac{1}{2} dk10^k + [d \geq 5](d-5)10^k + [d \geq 5]m + L(m) \\ &= \frac{1}{2} dk10^k + L(m) + [d \geq 5](d-5)10^k + [d \geq 5]m \end{aligned}$$

and by simplifying

$$\begin{aligned} &\frac{1}{2}(d \cdot 10^k + m) \log(d \cdot 10^k + m) = \frac{1}{2}(d \cdot 10^k + m) \log 10^k \left(d + \frac{m}{10^k}\right) \\ &= \frac{1}{2}(d \cdot 10^k + m)\left(k + \log\left(d + \frac{m}{10^k}\right)\right) \\ &= \frac{1}{2} dk10^k + \frac{1}{2} km + \frac{1}{2}(d \cdot 10^k + m) \log\left(d + \frac{m}{10^k}\right). \end{aligned}$$

The first terms of the above 2 expressions match, i.e.,

$$\frac{1}{2} dk10^k.$$

Since m has less than j nonzero decimal digits, by the induction hypothesis,

$$L(m) \leq \frac{1}{2} m \log m \leq \frac{1}{2} km.$$

Finally, by the second part of Lemma 3

$$[d \geq 5](d-5)10^k + [d \geq 5]m \leq \frac{1}{2}(d \cdot 10^k + m) \log\left(d + \frac{m}{10^k}\right).$$

This completes the proof of the right inequality of Theorem 1.

3. Proof of the Left Inequality of Theorem 1.

To start the proof of the left inequality of Theorem 1, we need the following lemmas.

Lemma 4. Let n and k be a positive integers, m be a nonnegative integer less than 10^k, and d be a decimal digit. Then

$$L(10n) = 10L(n) + 5n,$$

$$L(10^k n + m) = L(10^k n) + L(m) + ml(n),$$

$$\text{and } L(10d) = 5d + [d \geq 5](d-5)10.$$

Proof. The first formula of Lemma 4 can be found in [1, pp. 28–29]. The second formula is a generalization of the second formula of Lemma 2. The proof is similar to the proof of the second formula of Lemma 2. The proof of the third formula follows by examining $L(0)$, $L(10)$, $L(20)$, $\ldots$, $L(90)$.

Lemma 5. Let m be a positive integer and d be a decimal digit. Then

$$\frac{L(10m) - \frac{1}{2}(10m+10)\log(10m+10)}{10m} \leq \frac{L(10m+d) - \frac{1}{2}(10m+d)\log(10m+d)}{10m+d}$$

Proof. To show the above inequality, it suffices to show that

$$(10m+d)L(10m) - \frac{1}{2}(10m+d)(10m+10)\log(10m+10) \leq 10mL(10m+d) - \frac{1}{2}10m(10m+d)\log(10m+d).$$

But since

$$L(10m+d) = L(10m) + L(d) + dl(m)$$

by the second formula in Lemma 4 and moving some terms around, it suffices to show that

$$\begin{aligned}&10mL(10m) + dL(10m) + \frac{1}{2}10m(10m+d)\log(10m+d)\\&\le 10mL(10m) + 10mL(d) + 10mdl(m) + \frac{1}{2}(10m+d)(10m+10)\log(10m+10).\end{aligned}$$

Cancelling the first term and using the right inequality of Theorem 1, i.e.,

$$L(10m) \le \frac{1}{2}10m\log 10m,$$

we are done if we show the stronger inequality that

$$\begin{aligned}&\frac{1}{2}d\cdot 10m\log 10m + \frac{1}{2}10m(10m+d)\log(10m+d)\\&\le 10mL(d) + 10mdl(m) + \frac{1}{2}(10m+d)(10m+10)\log(10m+10)\\&= 10mL(d) + 10mdl(m) + \frac{1}{2}(10m+d)\cdot 10m\log(10m+10)\\&\qquad\qquad + \frac{1}{2}(10m+d)\cdot 10\log(10m+10).\end{aligned}$$

But,

$$\begin{aligned}\frac{1}{2}d\cdot 10m\log 10m &\le \frac{1}{2}10\cdot(10m+d)\log(10m+10)\\&= \frac{1}{2}(10m+d)\cdot 10\log(10m+10)\\ \text{and }\frac{1}{2}10m(10m+d)\log(10m+d) &\le \frac{1}{2}10m(10m+d)\log(10m+10)\end{aligned}$$

so the above inequality is true. This completes the proof.

Lemma 6. Let m be a positive integer and d be a decimal digit. Then

$$\begin{aligned}&\frac{L(10m) - \frac{1}{2}(10m+10)\log(10m+10)}{10m}\\&\le \frac{L(100m+10d) - \frac{1}{2}(100m+10d+10)\log(100m+10d+10)}{100m+10d}.\end{aligned}$$

Proof. First of all, by the first formula of Lemma 4, we have that

$$\begin{aligned}
&\frac{L(10m) - \frac{1}{2}(10m+10)\log(10m+10)}{10m} \\
&= \frac{10L(10m) - \frac{1}{2}(100m+100)\log(10m+10)}{100m} \\
&= \frac{L(100m) - \frac{1}{2}10\cdot 10m - \frac{1}{2}(100m+100)\log(10m+10)}{100m}.
\end{aligned}$$

Second by simplifying, we have that

$$\begin{aligned}
&\frac{L(100m+10d) - \frac{1}{2}(100m+10d+10)\log(100m+10d+10)}{100m+10d} \\
&= \frac{L(100m+10d) - \frac{1}{2}(100m+10d+10)(1+\log(10m+d+1))}{100m+10d} \\
&= \frac{L(100m+10d) - \frac{1}{2}(100m+10d+10) - \frac{1}{2}(100m+10d+10)\log(10m+d+1)}{100m+10d}.
\end{aligned}$$

To prove our result, we must show that

$$\begin{aligned}
&(100m+10d)\left(L(100m) - \frac{1}{2}10\cdot 10m - \frac{1}{2}(100m+100)\log(10m+10)\right) \\
&\le 100m\left(L(100m+10d) - \frac{1}{2}(100m+10d+10) - \frac{1}{2}(100m+10d+10)\log(10m+d+1)\right).
\end{aligned}$$

Now since

$$L(100m+10d) = L(100m) + 10dl(m) + L(10d)$$

by the second formula of Lemma 4, we must show that

$$\begin{aligned}
&(100m+10d)\left(L(100m) - \frac{1}{2}10\cdot 10m - \frac{1}{2}(100m+100)\log(10m+10)\right) \\
&\le 100m\Bigg(L(100m) + 10dl(m) + L(10d) \\
&\quad - \frac{1}{2}(100m+10d+10) - \frac{1}{2}(100m+10d+10)\log(10m+d+1)\Bigg).
\end{aligned}$$

Multiplying terms and juggling terms so that every term is positive, we must show that

$$\begin{aligned}
&100mL(100m) + 10dL(100m) + \frac{1}{2}100m(100m + 10d + 10)\\
&\quad + \frac{1}{2}100m(100m + 10d + 10)\log(10m + d + 1)\\
&\le 100mL(100m) + 10d \cdot 100ml(m) + 100mL(10d)\\
&\quad + \frac{1}{2}100m(100m + 10d) + \frac{1}{2}(100m + 10d)(100m + 100)\log(10m + 10).
\end{aligned}$$

Canceling the terms

$$100mL(100m) \quad \text{and} \quad \frac{1}{2}100m(100m + 10d)$$

we must show that

$$\begin{aligned}
&10dL(100m) + \frac{1}{2}10 \cdot 100m + \frac{1}{2}100m(100m + 10d + 10)\log(10m + d + 1)\\
&\le 10d \cdot 100ml(m) + 100m \cdot L(10d) + \frac{1}{2}(100m + 10d)(100m + 100)\log(10m + 10).
\end{aligned}$$

But by the right inequality of Theorem 1,

$$\begin{aligned}
L(100m) &\le \frac{1}{2}100m\log 100m\\
&= \frac{1}{2}100m(1 + \log 10m)\\
&= \frac{1}{2}100m + \frac{1}{2}100m\log 10m.
\end{aligned}$$

So if we can show the stronger inequality that

$$\begin{aligned}
&\frac{1}{2}10d \cdot 100m + \frac{1}{2}10d \cdot 100m\log 10m\\
&\quad + \frac{1}{2}10 \cdot 100m + \frac{1}{2}100m(100m + 10d + 10)\log(10m + d + 1)\\
&\le 10d \cdot 100ml(m) + 100mL(10d) + \frac{1}{2}(100m + 10d)(100m + 100)\log(10m + 10)
\end{aligned}$$

then we can go back and we are done. Now for any decimal digit d

$$L(10d) = 5d + [d \ge 5](d - 5)10$$

by the third formula of Lemma 4. Therefore combining terms on the left side and using this result on the right side, we are done if we can show that

$$100m(5d+5)+\frac{1}{2}10d\cdot 100m\log 10m+\frac{1}{2}100m(100m+10d+10)\log(10m+d+1)$$
$$\leq 100m(5d+[d\geq 5](d-5)10)+10d\cdot 100ml(m)$$
$$+\frac{1}{2}(100m+10d)(100m+100)\log(10m+10).$$

Reworking the right-hand side of the above inequality, we must show that

$$100m(5d+5)+\frac{1}{2}10d\cdot 100m\log 10m+\frac{1}{2}100m(100m+10d+10)\log(10m+d+1)$$
$$\leq 100m(5d+[d\geq 5](d-5)10)+10d\cdot 100ml(m)$$
$$+\frac{1}{2}100m(100m+100)\log(10m+10)+\frac{1}{2}10d(100m+100)\log(10m+10).$$

Again, reworking the right-hand side, we must show that

$$100m(5d+5)+\frac{1}{2}10d\cdot 100m\log 10m+\frac{1}{2}100m(100m+10d+10)\log(10m+d+1)$$
$$\leq 100m(5d+[d\geq 5](d-5)10)+10d\cdot 100ml(m)$$
$$+\frac{1}{2}100m(100m+10d+10)\log(10m+10)$$
$$+\frac{1}{2}100m(100-10d-10)\log(10m+10)+\frac{1}{2}10d(100m+100)\log(10m+10).$$

For $d=6,7,8,9$,

$$100m(5d+5)\leq 100m(5d+[d\geq 5](d-5)10),$$
$$\frac{1}{2}10d\cdot 100m\log 10m\leq\frac{1}{2}10d(100m+100)\log(10m+10),\ \text{and}$$
$$\frac{1}{2}100m(100m+10d+10)\log(10m+d+1)\leq\frac{1}{2}100m(100m+10d+10)\log(10m+10)$$

and the other terms on the right-hand side of the inequality are 0 or positive so our inequality is true. For $d=0,1,2,3,4,5$,

$$100m\cdot 5d=100m\cdot 5d,$$
$$100m\cdot 5=\frac{1}{2}10\cdot 100m\leq\frac{1}{2}100m(100-10d-10)\log(10m+10),$$
$$\frac{1}{2}10d\cdot 100m\log 10m\leq\frac{1}{2}10d(100m+100)\log(10m+10),\ \text{and}$$
$$\frac{1}{2}100m(100m+10d+10)\log(10m+d+1)\leq\frac{1}{2}100m(100m+10d+10)\log(10m+10)$$

and the other terms on the right-hand side of the inequality are 0 or positive so our inequality is true. This completes the proof of Lemma 6.

We are now ready to prove the left inequality of Theorem 1. Define

$$S = \{100, 101, 102, 103, \ldots, 999\}.$$

Let

$$\mu = \min_{n \in S} \frac{L(10n) - \frac{1}{2}(10n+10)\log(10n+10)}{10n}.$$

By some calculations, we have that

$$\mu \approx -.3884483\ldots \geq -.389.$$

In addition, this minimum value occurs at $n = 455$. Now we wish to show that $-.389$ is smaller than any value of

$$R(n) = \frac{L(n) - \frac{1}{2}n\log n}{n},$$

for any positive integer n. To do this we need the following lemma, which can be found in [1, pp. 28–29].

<u>Lemma 7</u>. Let n be a positive integer. Then

$$R(10n) = R(n).$$

Now if $n < 1000$, then $R(n) = R(10n) = R(100n) = R(1000n)$ so it is enough to consider only $n \geq 1000$. Write $n = 10m + d$, where d is a decimal digit. Also, write m in its decimal representation, i.e.,

$$m = \sum_{i=0}^{k} d_i 10^i.$$

Finally, for $0 \leq j \leq k$, let

$$m_j = \sum_{i=j}^{k} d_i 10^{i-j}.$$

Note that $m_0 = m$ and for $j = k, k-1, k-2, \ldots, 2, 1$

$$m_{j-1} = 10m_j + d_{j-1}.$$

Now we have the following sequence of inequalities.

$$\begin{aligned}
&-.389 \leq \mu \leq \frac{L(10m_{k-2}) - \frac{1}{2}(10m_{k-2}+10)\log(10m_{k-2}+10)}{10m_{k-2}} \\
&\leq \frac{L(100m_{k-2}+10d_{k-3}) - \frac{1}{2}(100m_{k-2}+10d_{k-3}+10)\log(100m_{k-2}+10d_{k-3}+10)}{100m_{k-2}+10d_{k-3}} \\
&= \frac{L(10m_{k-3}) - \frac{1}{2}(10m_{k-3}+10)\log(10m_{k-3}+10)}{10m_{k-3}} \\
&\leq \cdots \\
&\leq \frac{L(10m_1) - \frac{1}{2}(10m_1+10)\log(10m_1+10)}{10m_1} \\
&\leq \frac{L(100m_1+10d_0) - \frac{1}{2}(100m_1+10d_0+10)\log(100m_1+10d_0+10)}{100m_1+10d_0} \\
&= \frac{L(10m_0) - \frac{1}{2}(10m_0+10)\log(10m_0+10)}{10m_0} \\
&= \frac{L(10m) - \frac{1}{2}(10m+10)\log(10m+10)}{10m}.
\end{aligned}$$

The second inequality follows from the definition of μ. The next set of inequalities follow by applying Lemma 6 successively with m_{k-2}, d_{k-3}; m_{k-3}, d_{k-4}; $\ldots$; m_1, d_0. Now applying Lemma 5 with m and d we get

$$\frac{L(10m) - \frac{1}{2}(10m+10)\log(10m+10)}{10m} \leq \frac{L(10m+d) - \frac{1}{2}(10m+d)\log(10m+d)}{10m+d}.$$

Since $n = 10m + d$ was arbitrary, we have the left inequality of Theorem 1.

4. Further Questions.

The result in Theorem 1 for the main term and error term of $L(n)$ was done for base 10. Can Theorem 1 be extended to determine the main term and error term of $L(n)$ in base k, where $k > 1$ is an integer? Here, $L(n)$ is the number of large digits in base k. Next, in [2, pp. 37–41] it was shown that

$$\sum_{1 \leq m < n} l(m)^2 = \frac{1}{4} n \log^2 n + O(n \log n)?$$

Can we find bounds for the big-Oh term in the above formula? Finally, we conjecture that the sequence

$$n = 5, 45, 455, 4545, 45455, 454545, \ldots$$

gives the smallest value for

$$\frac{L(n) - \frac{1}{2} n \log n}{n}$$

for $1, 2, 3, 4, 5, 6, \ldots$ digit numbers. Also, we ask to find the limit of those values above, if it exist?

References

1. C. Cooper and R. E. Kennedy, "An Explicit Expression for Large Digit Sums in Base b Expansions," *Journal of Institute of Mathematics & Computer Science,* 11.1 (1998), 25–32.
2. C. Cooper and R. E. Kennedy, "Finding Moments of the Large Digit Function Using Delange's Method," *Journal of Institute of Mathematics & Computer Science,* 13.1 (2000), 35–41.
3. R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics,* 2nd ed., Addison-Wesley Publishing Company, Reading, Massachusetts, 1994.

AMS Classification Numbers: 11A63.