# HOW MANY 1'S ARE NEEDED?

**Daniel A. Rawsthorne**
13508 Bartlett St., Rockville, MD 20853
(Submitted January 1987)

Let $f(n)$ denote the number of 1's necessary to express $n$, using the operations + and × (and parentheses). Determining $f(n)$ is an old problem, originally considered by Mahler and Popken in 1953 [3]. We have calculated $f(n)$ for $n \le 3^{10}$ and we present some statistics. (The reason for a power of 3 will be explained later.)

## The Problem

With $f(n)$ defined as above, we have

$$f(n) = \begin{cases} 1, \text{ for } n = 1, \quad \text{and} \\ \min\limits_{\substack{ab=n \text{ or} \\ a+b=n}} \{f(a) + f(b)\}, \text{ for } n > 1. \end{cases} \tag{1}$$

This formula is very time consuming to use for large $n$, but we know of no other way to calculate $f(n)$.

The behavior of $f(n)$ is interesting. Selfridge has shown that $3^k + \Theta 3^{k-1}$ is the largest $n$ for which $f(n) = 3k + \Theta$, for $\Theta = 0, \pm 1$. His proof is by induction, and the induction step is based on the following observation: If $b(m)$ is the largest $n$ for which $f(n) = m$, then $b(m)$ is the largest element of the set

$$\bigcup_{r+s=m} \{b(r) + b(s),\ b(r)b(s)\}. \tag{2}$$

Using (2), it is fairly easy to show that $b(m)$ has the required form.

There are two competing conjectures about the behavior of $f(n)$ for large $n$. It has been conjectured that

$$f(n) < 3(1 + \varepsilon)\log_3 n, \text{ for large } n, \text{ and any } \varepsilon > 0. \tag{3}$$

It has also been conjectured that there is a set $S$ (possibly of positive density) and a positive constant $c$ so that

$$f(n) > 3(1 + c)\log_3 n, \text{ for all } n \text{ in } S. \tag{4}$$

## The Results

We calculated $f(n)$, using equation (1), for $n \le 3^{10}$ (59,049). We chose $3^{10}$ because we would have all $n$ for which $f(n) \le 30$, by Selfridge's results.

We broke the interval into 30 subintervals between the values of the form $3^k + \Theta 3^{k-1}$ for $\Theta = 0, \pm 1$ and we also looked at the sets $S(m) = \{n : f(n) = m\}$, for $m = 1, 2, \ldots, 37$, incomplete beyond $m = 30$.

## Analysis of the thirty subintervals

One typical subinterval is the interval $18 \le n < 27$. The values at the "endpoints" differ by 1. If conjecture (3) were true, we would expect the values of $f(n)/\log_3 n$ in the interval to "flatten out" and approach the values at the endpoints, as $n$ gets large. Table 1 gives the mean and standard deviation of $f(n)/\log_3 n$ in each interval.

While the analysis of such small values of $f(n)$ has very little to do with behavior at large $n$, it is clear that in this range conjecture (4) is strongly supported.

The single worst value of $f(n)/\log_3 n$ encountered was at $n = 1439$, with $f(n) = 26$, and $f(n)\log_3 n = 3.9281$.

**TABLE 1**

Mean and Standard Deviation of $f(n)/\log_3 n$, for $a \le n < b$

| a | b | mean | std dev |
|---|---|---|---|
| 1 | 2 | | 0.0 |
| 2 | 3 | 3.1699 | 0.0 |
| 3 | 4 | 3.0 | 0.0 |
| 4 | 6 | 3.2915 | 0.1216 |
| 6 | 9 | 3.2077 | 0.1340 |
| 9 | 12 | 3.3350 | 0.2716 |
| 12 | 18 | 3.2928 | 0.1382 |
| 18 | 27 | 3.3613 | 0.2273 |
| 27 | 36 | 3.3430 | 0.1754 |
| 36 | 54 | 3.3653 | 0.1607 |
| 54 | 81 | 3.3726 | 0.1748 |
| 81 | 108 | 3.3959 | 0.1630 |
| 108 | 162 | 3.3743 | 0.1307 |
| 162 | 243 | 3.3973 | 0.1473 |
| 243 | 324 | 3.3988 | 0.1395 |
| 324 | 486 | 3.3996 | 0.1327 |
| 486 | 729 | 3.4031 | 0.1290 |
| 729 | 972 | 3.4031 | 0.1194 |
| 972 | 1458 | 3.4037 | 0.1191 |
| 1458 | 2187 | 3.4031 | 0.1130 |
| 2187 | 2916 | 3.4039 | 0.1043 |
| 2916 | 4374 | 3.4017 | 0.1040 |
| 4374 | 6561 | 3.4012 | 0.0995 |
| 6561 | 8748 | 3.4016 | 0.0945 |
| 8748 | 13122 | 3.3996 | 0.0931 |
| 13122 | 19683 | 3.3985 | 0.0893 |
| 19683 | 26244 | 3.3987 | 0.0860 |
| 26244 | 39366 | 3.3965 | 0.0840 |
| 39366 | 59049 | 3.3949 | 0.0806 |

## Analysis of the sets $S(m)$

Let $S(m) = \{n : f(n) = m\}$. In Table 2 we consider the following questions about $S(m)$.

- What is its first element?
- How many elements are in $S(m)$?
- What is its last element?
- What is its average element?

One result about the sets $S(m)$ not captured in Table 2 is: If $b(m)$ and $b_1(m)$ are the largest and second largest elements of $S(m)$, then $b_1(m) = [(8/9)b(m)]$, where $[\cdot]$ is the greatest integer function.

**Outline of proof**: The proof is by induction.

The result is true by inspection for small values of $m$. For large values of $m$ we have an equation similar to (2): $b_1(m)$ is the second-largest member of the set

$$\bigcup_{r+s=m} \{b(r) + b(s),\ b(r)b(s),\ b(r) + b_1(s),\ b(r)b_1(s)\}. \tag{5}$$

For $m \geq 9$, it is easy to show that both $b(m)$ and $(8/9)b(m)$ belong to this set. It only remains to show that there are no elements between these two values, and this can be done by a simple case-by-case examination using the results of Selfridge and the induction hypothesis.

**TABLE 2**
Analysis of the sets $S(m)$

| m | first | last | \|S(m)\| | mean | median |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1.0 | 1.0 |
| 2 | 2 | 2 | 1 | 2.0 | 2.0 |
| 3 | 3 | 3 | 1 | 3.0 | 3.0 |
| 4 | 4 | 4 | 1 | 4.0 | 4.0 |
| 5 | 5 | 6 | 2 | 5.5 | 5.5 |
| 6 | 7 | 9 | 3 | 8.0 | 8.0 |
| 7 | 10 | 12 | 2 | 11.0 | 11.0 |
| 8 | 11 | 18 | 6 | 14.5 | 14.5 |
| 9 | 17 | 27 | 6 | 21.3 | 20.5 |
| 10 | 22 | 36 | 7 | 28.4 | 28.0 |
| 11 | 23 | 54 | 14 | 37.7 | 37.5 |
| 12 | 41 | 81 | 16 | 55.2 | 53.5 |
| 13 | 47 | 108 | 20 | 73.3 | 73.5 |
| 14 | 59 | 162 | 34 | 100.4 | 98.5 |
| 15 | 89 | 243 | 42 | 141.9 | 137.0 |
| 16 | 107 | 324 | 56 | 191.7 | 185.5 |
| 17 | 167 | 486 | 84 | 266.0 | 257.5 |
| 18 | 179 | 729 | 108 | 371.8 | 362.5 |
| 19 | 263 | 972 | 152 | 501.3 | 482.5 |
| 20 | 347 | 1458 | 214 | 701.3 | 675.0 |
| 21 | 467 | 2187 | 295 | 966.1 | 931.0 |
| 22 | 683 | 2916 | 398 | 1335.4 | 1284.5 |
| 23 | 719 | 4374 | 569 | 1842.9 | 1783.0 |
| 24 | 1223 | 6561 | 763 | 2571.0 | 2478.0 |
| 25 | 1438 | 8748 | 1094 | 3513.8 | 3382.5 |
| 26 | 1439 | 13122 | 1475 | 4914.9 | 4734.0 |
| 27 | 2879 | 19683 | 2058 | 6792.4 | 6533.5 |
| 28 | 3767 | 26244 | 2878 | 9378.7 | 9020.0 |
| 29 | 4283 | 39366 | 3929 | 13061.5 | 12534.0 |
| 30 | 6299 | 59049 | 5493 | 18051.5 | 17315.0 |
| 31 | 10079 | 78732 | | | |
| 32 | 11807 | 118098 | | | |
| 33 | 15287 | 177147 | | | |
| 34 | 21599 | 236196 | | | |
| 35 | 33599 | 354294 | | | |
| 36 | 45197 | 531441 | | | |
| 37 | 56039 | 708588 | | | |

## Comments

In his paper [2], Guy relays some questions about the function $f(n)$. We comment on three of these:

*Q:* For what values $a$ and $b$ does $f(2^a3^b) = 2a + 3b$?
*A:* $f(2^a3^b) = 2a + 3b$ for all $2^a3^b < 3^{10}$, at least.

*Q:* If $f(2^a3^b) = 2a + 3b$ and there is a larger $n'$ so that $f(n') = 2a + 3b$ $(a,\ b \geq 0)$, must $n' = 2^a3^b$, for some $r$, $s$?
*A:* No. Two counterexamples are $2^7 < 3^35$, with $f(2^7) = f(3^35) = 14$, and $2^73^2 < 3^55$, with $f(2^73^2) = f(3^55) = 20$.

*Q:* When the value of $f(n)$ is of the form $f(a) = f(b)$, with $a + b = n$, and this minimum is not achieved as a product, is either $a$ or $b$ equal to 1?
*A:* Yes, at least for $n \leq 3^{10}$.

The calculation of $f(n)$ was performed on a Symbolics 3645 LISP machine using equation (1), and we used over 50 hours of CPU time.

## References


1. Richard K. Guy. *Unsolved Problems in Number Theory*, Problem F26. New York: Springer-Verlag, 1981.
2. Richard K. Guy. "Some Suspiciously Simple Sequences." *American Math. Monthly* *93* (1986):186-190.
3. K. Mahler & J. Popken. "On a Maximum Problem in Arithmetic." (Dutch). *Nieuw Arch. Wiskunde* (3) *1* (1953):1-15, MR 14, 852e.


*****