The question "How many 1's are needed?" revisited

J. Arias de Reyna J. van de Lune

February 11, 2009

Abstract

We present a rigorous and relatively fast method for the computation of the *complexity* of a natural number (sequence A005245), and answer some *old and new* questions related to the question in the title of this note. We also extend the known terms of the related sequence A005520.

Introduction.

The subject of this note was (more or less indirectly) initiated in 1953 by K. Mahler and J. Popken [1]. We begin with a brief description of part of their work: Given a symbol x, consider the set V_n of all formal *sum-products* which can be constructed by using only the symbol x and precisely $n-1$ symbols from $\{+, \times\}$ and an arbitrary number of parentheses "(" and ")".

We have, for example, $V_1 = \{x\}$, $V_2 = \{x + x, x \times x\}$,

$$V_3 = \{x + (x + x), x + (x \times x), x \times (x + x), x \times (x \times x)\}.$$

More generally, for $n \geq 2$,

$$V_n = \bigcup_{k=1}^{n-1} (V_k + V_{n-k}) \cup \bigcup_{k=1}^{n-1} (V_k \times V_{n-k}).$$

Mahler and Popken's question was the following: If x is a positive real number, what is the largest number in V_n ? We restrict ourselves here to the case $x = 1$. Then the answer is [1] $M_n := \max V_n = \max_{1\leq k\leq n} p_{n,k}$ where

$$p_{n,k} = \Big\lfloor \frac{n}{k} \Big\rfloor^{k(\lfloor \frac{n}{k}\rfloor+1)-n} \Big(\Big\lfloor \frac{n}{k} \Big\rfloor + 1\Big)^{n-k\lfloor \frac{n}{k}\rfloor}.$$

This formula was simplified by Selfridge (see Guy [4, p. 189]) to $M_{3m-1} = 2 \cdot 3^{m-1}$, $M_{3m} = 3^m$, $M_{3m+1} = 4 \cdot 3^{m-1}$ for all $m \geq 1$. Clearly $M_1 = 1$.

Our problem is more or less the converse: Write a given natural number n as a *sum-product* as described above, only using the five symbols 1, +, $\times$, (, and). (However, not all these signs need to be used.)

It is clear that this is always possible: $n = 1 + 1 + 1 + \cdots + 1$ (using n 1's). Some further simple examples are

$$5 = 1 + (1 + 1) \times (1 + 1), \quad 6 = (1 + 1) \times (1 + 1 + 1).$$

Our goal will, of course, be to minimize the number of 1's used.

In a sum-product representation of n we will usually write 2 instead of $1+1$, and 3 instead of $1+1+1$. Also, we will replace the symbol $\times$ (times) by a dot $\cdot$ or simply juxtapose. For example, the Fibonacci number F_{25} can then be written with 35 1's as follows

$$F_{25} = 75025 = (1 + 2^2)(1 + 2^2)(1 + 2(1 + 2^2)(1 + 2^2)(2^2 \cdot 3(1 + 2^2)))$$

and $2^{27} - 1$ can be written with 56 as

$$2^{27} - 1 = 134217727 = (1 + 2 \cdot 3)(1 + 2^3 \cdot 3^2)(1 + 2^9 \cdot 3^3(1 + 2 \cdot 3^2)).$$

All these examples are *minimal* in the sense defined in the next section.

1 Definitions and first properties.

Definition 1. *The minimal number of* 1*'s needed to represent* n *as a sum-product will be denoted by* $\|n\|$ *and will be called the* complexity *of* n*.*

It is clear that $\|1\| = 1$ and $\|2\| = 2$, but $\|11\| \neq 2$ ("pasting together" two 1's is not an allowed operation). One may verify directly that $\|3\| = 3$, $\|4\| = 4$, $\|5\| = 5$, $\|6\| = 5$, and by means of our program in Section 2 it may be shown that

$$\|7\| = 6, \quad \|8\| = 6, \quad \|9\| = 6, \quad \|10\| = 7, \quad \|11\| = 8, \quad \|12\| = 7,$$
$$\|13\| = 8, \quad \|14\| = 8, \quad \|15\| = 8, \quad \|16\| = 8, \quad \|17\| = 9, \quad \|18\| = 8.$$

Note that

(a) $\|n\|$ is not monotonic

(b) n may have different minimal representations ($4 = 1 + 1 + 1 + 1 = (1 + 1)(1 + 1)$).

It is clear that we always have

$$\|a + b\| \le \|a\| + \|b\| \quad \text{and} \quad \|a \cdot b\| \le \|a\| + \|b\|$$

so that, for example, $\|2^n\| \le 2n$. Also see Section 4.3.

Some useful bounds on the complexity are known

$$\frac{3}{\log 3} \log n \le \|n\| \le \frac{3}{\log 2} \log n, \qquad n \ge 2.$$

The first can be found in Guy [4] and is essentially due to Selfridge. The second appeared in Arias de Reyna [8] (this inequality can easily be proved. Indeed, just think of the binary expansion of n.) Since it is known that $\|3^k\| = 3k$ for $k \geq 1$, the first inequality cannot be improved. As for the second one:

$$\limsup_{n\to\infty} \|n\|/\log n$$

is not known, but we conjecture that it is considerably $< \frac{3}{\log 2}$ (≈ 4.328). Our most extreme observation is $\|1439\|/\log 1439 \approx 3.575503$.

2 Computing the complexity.

For $n \geq 2$ we may compute $\|n\|$ from

$$\|n\| = \min\Big\{ \min_{1\leq j\leq n/2} \|j\| + \|n-j\| \, , \quad \min_{d|n,\; 2\leq d\leq\sqrt{n}} \|d\| + \|n/d\| \Big\} \, . \tag{1}$$

From this it is clear that, for large n, the computation of

$$\min_{1\leq j\leq n/2} \|j\| + \|n-j\|$$

is quite time consuming, if not eventually prohibitive. Rawsthorne [7, p. 14] wrote *This formula is very time consuming to use for large n, but we know of no other way to calculate $\|n\|$.*

The principal goal of this note is to reduce the number of operations for the computation of $\|n\|$. (We can show that, instead of $\mathcal{O}(n^2)$, our algorithm needs only $\mathcal{O}(n^{1.345})$ operations for the computation of $\|n\|$.)

According to the definition we have to compute

$$P := \min_{1\leq k\leq n/2} \|k\| + \|n-k\| \quad \text{and} \quad T := \min_{d|n,\; 2\leq d\leq\sqrt{n}} \|d\| + \|n/d\|$$

and then set $\|n\| = \min(P, T)$. It is clear that $P \leq \|1\| + \|n-1\|$ so that P is the result of the loop

$P = 1 + \|n-1\|$;
For $k = 2$ to $k = n/2$ do $\quad P = \min(P\,,\ \|k\| + \|n-k\|)$.

Clearly this is cumbersome for large n. It would be very helpful to have a relatively small number **kMax** such that P would just as well be the result of the much shorter loop

$P = 1 + \|n-1\|$;
For $k = 2$ to $k =$ **kMax** do $\quad P = \min(P\,,\ \|k\| + \|n-k\|)$.

Such a relatively small **kMax** may be found indeed by observing that

$$\|m\| \geq \frac{3}{\log 3}\log m \qquad \text{for all } m \geq 1.$$

Indeed, we are through if **kMax** satisfies

$$\|k\| + \|n-k\| \geq \frac{3}{\log 3}\big(\log k + \log(n-k)\big) \geq 1 + \|n-1\| \qquad \text{for } \mathbf{kMax} + 1 \leq k \leq n/2.$$

This only requires to solve a simple quadratic inequality:

$$k^2 - nk + \exp(R) \geq 0 \qquad \text{where} \quad R = \frac{\log 3}{3}(1 + \|n-1\|).$$

It is easily seen that, for $n \geq 7$, we can safely take

$$\mathbf{kMax} = \Big\lfloor \frac{1}{2} + \frac{n}{2}\Big(1 - \sqrt{1 - 4\exp(R - 2\log n)}\Big)\Big\rfloor.$$

It will soon become clear that for large n this **kMax** is very small compared to $n/2$. In our computations covering all $n \leq 905\,000\,000$ we observed that **kMax** ≤ 66 in all cases, with an average value of about 11.57.

However, we can not use this “trick” for the $\times$ part.

Mathematica program to compute `Compl[`n`]:=` $\|n\|$**.**

```
Compl[1] = 1; Compl[2] = 2; Compl[3] = 3; Compl[4] = 4;
Compl[5] = 5; Compl[6] = 5; nDone = 6;
(* Our computed kMax is not real for n<= 6 *)
ComplChamp = 5;
(* = largest value of  C[n]  found so far.*)

n = nDone; While[0 == 0, n += 1;

(*  First we deal with the  PLUS-part.  *)
P = 1 + Compl[n - 1]; R=N[Log[3] P/3];
kMax =  Floor[1/2+n(1-Sqrt[ 1 - 4Exp[R - 2Log[n]]])/2];
For[k = 2 , k <= kMax , k++ ,
P = Min[ P , Compl[k] + Compl[n - k]]];
(* kMax < 2  causes no problem.  *)

(*  Now for the  TIMES-part.  *)
S = Divisors[n];   LSplus1 = Length[S] + 1; T = P;
(* From the PLUS-part we already
                           know  that  Compl[n] <= P  *)
```

```
For[k = 2 , k <= LSplus1/2 , k++ ,
d = S[[k]];  T = Min[ T , Compl[d] + Compl[n/d]]];

Compl[n] = T;   (*  There we are !  *)

(* We output the Champion Compl[n] and the
                         corresponding Compl[n] / Log[n] *)
If[T > ComplChamp,
    ComplChamp = T;
    Print["n = ", n, "    kMax = ", kMax, "    ComplChamp = ",
                 ComplChamp, "    ||n||/Log [n] = ",
                                       N[Compl[n]/Log[n]]]]]
```

A much faster Delphi-Object-Pascal version of this program, run on a Toshiba laptop, computes $\|n\|$ for all $n \leq 905\,000\,000$ in about 2 hours and 40 minutes.

Note. In the range $n \leq 905\,000\,000$ it suffices to take kMax $= 6$. This value (kMax $= 6$) is necessary only for $n = 353\,942\,783$ and $n = 516\,743\,639$. But this is hindsight!

3 Some records.

Definition 2. *The number n is called* highly complex *if $\|k\| < \|n\|$ for all $k < n$.*

P. Fabian (see [10]) has computed the first 58 highly complex numbers. With our new method we have been able to add those with $59 \leq \|n\| \leq 67$ (in boldface at the end of Table 1). There are no others in the range $n \leq 905\,000\,000$. We performed our computations on a Toshiba laptop, 2GB RAM, 3.2 GHz, and could verify Fabian's results within 138 seconds.

We denote by F_m the first number having complexity m (i. e. F_m is the m-th highly complex number). $S(m)$ denotes the set of numbers with complexity m, its first element is F_m, and its maximal element M_m.

TABLE 1
Some data related to Highly Complex numbers

m	F_m	**kMax**	$\|F_m\|/\log F_m$	M_m	$\#S(m)$
1	1			1	1
2	2		2.8853900818	2	1
3	3		2.7307176799	3	1
4	4		2.8853900818	4	1
5	5		3.1066746728	6	2
6	7		3.0833900542	9	3
7	10	2	3.0400613733	12	2
8	11	2	3.3362591314	18	6
9	17	2	3.1766051148	27	6
10	22	2	3.2351545315	36	7

m	F_m	**kMax**	$\Vert F_m\Vert/\log F_m$	M_m	$\#S(m)$
11	23	3	3.5082188779	54	14
12	41	2	3.2313900968	81	16
13	47	3	3.3764939282	108	20
14	59	3	3.4334448653	162	34
15	89	3	3.3417721474	243	42
16	107	3	3.4240500919	324	56
17	167	3	3.3216140197	486	84
18	179	4	3.4699559034	729	108
19	263	4	3.4098124155	972	152
20	347	4	3.4191980703	1458	214
21	467	5	3.4166734517	2187	295
22	683	5	3.3708752513	2916	398
23	719	6	3.4965771927	4374	569
24	1223	5	3.3759727432	6561	763
25	1438	7	3.4383125626	8748	1094
26	1439	10	**3.5755032174**	13122	1475
27	2879	7	3.3897461199	19683	2058
28	3767	8	3.4005202424	26244	2878
29	4283	10	3.4679002280	39366	3929
30	6299	9	3.4292979813	59049	5493
31	10079	8	3.3629090954	78732	7669
32	11807	10	3.4128062668	118098	10501
33	15287	12	3.4250989750	177147	14707
34	21599	12	3.4066763033	236196	20476
35	33599	11	3.3581994945	354294	28226
36	45197	12	3.3585893055	531441	39287
37	56039	14	3.3840009256	708588	54817
38	81647	14	3.3598108962	1062882	75619
39	98999	16	3.3904596729	1594323	105584
40	163259	14	3.3324743393	2125764	146910
41	203999	16	3.3535444722	3188646	203294
42	241883	20	3.3881324998	4782969	283764
43	371447	19	3.3527842988	6377292	394437
44	540539	18	3.3332520048	9565938	547485
45	590399	24	3.3863730003	14348907	763821
46	907199	23	3.3532298662	19131876	1061367
47	1081079	28	3.3828841470	28697814	1476067
48	1851119	23	3.3261034748	43046721	2057708
49	2041199	30	3.3725540867	57395628	2861449
50	3243239	28	3.3350935780	86093442	3982054
51	3840479	34	3.3638703158	129140163	5552628
52	6562079	28	3.3127733211	172186884	7721319
53	8206559	33	3.3290528266	258280326	10758388
54	11696759	33	3.3180085674	387420489	14994291
55	14648759	38	3.3333603679	516560652	20866891
56	22312799	36	3.3095614199	774840978	29079672
57	27494879	42	3.3275907432	1162261467	
58	41746319	40	3.3053853809	1549681956	
59	**52252199**	46	3.3199050612	2324522934	
60	**78331679**	45	3.3009723129	3486784401	
61	**108606959**	46	3.2967188492	4649045868	
62	**142990559**	51	3.3016852310	6973568802	
63	**203098319**	52	3.2933942627	10460353203	
64	**273985919**	55	3.2941149607	13947137604	
65	**382021919**	57	3.2893091281	20920706406	
66	**495437039**	63	3.2965467292	31381059609	
67	**681327359**	66	3.2940742853	41841412812	

4 Some questions solved and proposed.

One of the facts that our extended computation has revealed is that sometimes the minimum in equation (1) is assumed *only by the sums* and with a $j > 1$. In the range $n \leq 905\,000\,000$ there are only two such instances.

The first case is the prime number $p = 353\,942\,783$ (with $j = 6$). Indeed, the representation

$$353\,942\,783 = 2 * 3 + (1 + 2^2 * 3^2) * (2 + 3^4(1 + 2 * 3^{10}))$$

proves that $\|p\| \leq 63$, and one may verify that $\|p\| = 63$ and $\|p - 1\| = 63$, so that

$$\|p\| = \|6\| + \|p - 6\| = 5 + 58 = 63 < 64 = \|p - 1\| + 1.$$

In this case we thus have $\|p\| = \|k\| + \|p - k\|$ with $k = 6$ (and no other choice of k is adequate).

The second example is the number $n = 516\,743\,639$. It is the product of two primes $n = 353 \cdot 1\,463\,863$. We have

$$516\,743\,639 = 2 * 3 + (1 + 2^2 3^6)(2 + 3^{11})$$

so that $\|n\| \leq 63$. Also $\|n - 1\| = 63$, $\|353\| = 19$, $\|1463863\| = 45$, $\|n - 6\| = 58$ and finally $\|n\| = 63$, so that

$$1 + \|n - 1\| = \|353\| + \|1463863\| = 64 > \|6\| + \|n - 6\| = \|n\| = 63.$$

Hence $\|n\| = \|k\| + \|n - k\|$ with $k = 6$ and no other choice of k is adequate, as claimed.

Now we are sufficiently prepared to answer some questions asked by Guy.

4.1 Answering some questions of Guy

Q1: For which values a and b is $\|2^a 3^b\| = 2a + 3b$ *?*

A1: $\|2^a 3^b\| = 2a + 3b$ for all $2^a 3^b \leq 905\,000\,000$. No counter examples are known (to us).

Q2: Is it always true that $\|p\| = 1 + \|p - 1\|$*, if p is prime ?*

A2: No.
The first prime for which this is not true is $p = 353\,942\,783$ with $\|p\| = 63$ and $\|p - 1\| = 63$. This is the only example in the range $n \leq 905\,000\,000$.

Q3: Is it always true that $3 + \|p\| \leq 1 + \|3p - 1\|$*, if p is prime ?*

A3: No.
There are many exceptions: $p = 107, 347, 383, 467, 587, 683, 719, 887, \ldots$

Q4: Is it always true that $\|2p\| = \min\{2+\|p\|,\ 1+\|2p-1\|\}$, *if* p *is prime ?*

A4: Yes for $2p \le 905\,000\,000$.
Putting $L = 2 + \|p\|$ and $R = 1 + \|2p-1\|$, we found in this range

$$\begin{array}{ll} \|2p\| = L\ (< R) & \text{in } 12\,317\,371 \text{ cases} \\ \|2p\| = R\ (< L) & \text{in } \ \ 3\,629\,305 \text{ cases} \\ \|2p\| = L = R & \text{in } \ \ 8\,031\,758 \text{ cases.} \end{array}$$

Note that "$(L < R) + (R < L) + (L = R)$"$= 23\,978\,434 = \pi(905\,000\,000/2)$ where $\pi(\cdot)$ is the prime counting function.

Q5: When the value of $\|n\|$ *is of the form* $\|a\| + \|b\|$, *with* $a + b = n$, *and this minimum is not achieved as a product, is either* a *or* b *equal to* 1 *?*

A5: No.
We have only our two earlier mentioned (counter) examples: The prime $p = 353942783$ and $n = 516743639$ with prime factorization $n = 353 \cdot 1463863$.

We have also searched in the range $n \le 905\,000\,000$ for those cases where the minimum of $\|k\| + \|n-k\|$ is not assumed for $k = 1$. In the cases with $k > 1$ we mostly have $k = 6$, but sometimes we have $k = 8$. In all cases $\|n\| = \|k\| + \|n-k\| = \|n-1\|$. All cases found with $k > 1$ are (those with $k = 8$ in boldface)

21080618, 63241604, 67139098, 116385658, 117448688, 126483083, 152523860, 189724562, 212400458, 229762259, 318689258, 348330652, 353942783, 366873514, 373603732, 379448999, **385159320**, 404764540, 409108300, 460759642, **465722100**, 477258719, 498197068, 511069678, 516743639, 519835084, 538858312, 545438698, 545790940, 546853138, 574842670, **575550972**, 581106238, 590785918, 608504399, 612752632, **612752634**, 613028608, 613175855, 614416318, 636135035, 637198964, 669796594, 673335934, 690342298, **690342300**, 691406048, 692981240, **698494572**, 817595279, 822093928, 833714854, 860101032, 861764920, **865717578**.

Q6: There are two conflicting conjectures:

$$\textit{For large } n, \qquad (3+\varepsilon)\frac{\log n}{\log 3} \qquad \textit{ones suffice ?}$$

There are infinitely many n, *perhaps a set of positive density for which*

$$(3+c)\frac{\log n}{\log 3} \qquad \textit{ones are needed, for some } c > 0 \textit{ ?}$$

A6: To the first question: In view of the values of $\|n\|/\log n$ in Table 1, the answer will most probably be no.

A6: To the second question: Here the answer might very well be yes. If we solve for c in the equation

$$\|n\| = (3+c)\frac{\log n}{\log 3}$$

we get a mean value $\bar{c} > 0.366$ and a standard deviation $\sigma < 0.047$ in the range $2 \le n \le 905\,000\,000$. Also, the frequency of the event $c > 0.3$ is $> 92.5\%$.

Certainly $\liminf_n \|n\|/\log n = 3/\log 3 \approx 2.73072$. Our computations suggest that $\limsup_n \|n\|/\log n \le 3.58$ and that

$$\lim_{N\to\infty} \frac{1}{N-1}\sum_{k=2}^{N} \frac{\|k\|}{\log k} > 3 \quad \text{(possibly even } > 3.06\text{)}.$$

4.2 Some other questions

Note that the sequence $\|n\|$ is not monotonic. It is clear that $\|n-1\| - \|n\| \ge -1$. So, one may pose the question: How large can $\|n-1\| - \|n\|$ be ? We found the first values of n for which this difference is equal to k

Large values of $\|n-1\| - \|n\|$

n	6	12	24	108	720	1440	81648	2041200	612360000
$k = \|n-1\| - \|n\|$	0	1	2	3	4	5	6	7	8

In the range $n \le 905\,000\,000$ there are no larger values of $\|n-1\| - \|n\|$.

Conjecture 1. $\limsup_{n\to\infty} \ (\|n-1\| - \|n\|) \ = \ +\infty.$

Let $n = \prod p_j^{a_j}$ be the standard prime-factorization of n. It is clear that $\|n\| \le \sum_j \|p^{a_j}\|$. So we define a function $\text{AddExc}(n) = \sum_j \|p^{a_j}\| - \|n\|$ (Additive Excess) and ask how large AddExc(n) can be. We found

First n_k with $\text{AddExc}(n_k) = k$

AddExc(n)	1	2	3	4	5	6	7
n	46	253	649	6049	69989	166213	551137
AddExc(n)	8	9	10	11			
n	9064261	68444596	347562415	612220081			

and there are no $n \le 905\,000\,000$ with a larger Additive Excess.

Suppose that in our program for $\|n\|$ we start with $\|1\| = 1$ and $\|2\| = 1+x$ (where x is any given real value). Then the $\|n\|$ will be functions of x. What can be said about the resulting $\|n\|_x$?

Is it true that $\|p^k\| = k\|p\|$? Yes for $p = 3$, but we have some doubts about $p = 2$. See Section 4.3. We conjecture: *False for all other primes.* Examples:

$$\begin{aligned}
\|5^6\| &= \|15\,625\| = 29 < 30 = 6\|5\|;\\
\|7^9\| &= \|40\,353\,607\| = 53 < 54 = 9\|7\|;\\
\|19^6\| &= \|47\,045\,881\| = 53 < 54 = 6\|19\|;\\
\|37^5\| &= \|69\,343\,957\| = 54 < 55 = 5\|37\|.
\end{aligned}$$

Our computations have revealed that for all primes $5 \le p \le 113$ (with the possible exceptions $p = 73$, 97 and 109) it is not true that $\|p^n\| = n\|p\|$ for all $n \ge 1$.

We also wondered how often $\|\prod p^e\| = \sum e\|p\|$. We got the impression that in the long run we have about equally often true and false.

Pegg [10] asks what the smallest number requiring 100 ones is? The points $(\|n\|, \log F_n)$ form approximately a straight line (similarly as the Mahler-Popken-Selfridge points $(m, \log M_m)$). Various least squares fits of the form $A + Bt$ suggest that M_{100} should be situated between

$$11\,857\,300\,000\,000 \quad \text{and} \quad 27\,345\,300\,000\,000.$$

A real challenge for a supercomputer ! The largest number requiring 100 ones is $M_{100} = 7\,412\,080\,755\,407\,364$.

Some other predictions are

$$\begin{aligned}
&F(68) \approx 0.98 \cdot 10^9, &&F(69) \approx 1.35 \cdot 10^9, &&F(70) \approx 1.86 \cdot 10^9,\\
&F(71) \approx 2.56 \cdot 10^9, &&F(72) \approx 3.53 \cdot 10^9, &&F(73) \approx 4.85 \cdot 10^9,\\
&F(74) \approx 6.68 \cdot 10^9, &&F(75) \approx 9.20 \cdot 10^9, &&F(80) \approx 45.54 \cdot 10^9.
\end{aligned}$$

4.3 Is it true that $\|2^k\| = 2k$?

Selfridge asked whether $\|2^k\| = 2k$ for all $k \ge 1$. We have verified this for $1 \le k \le 29$. Nevertheless, we will present an argument suggesting that the answer may very well be no.

Given a natural number n with complexity $\|n\| = a$ we denote by M_a the greatest number with the same complexity, and we will call

$$\mathrm{CR}\,(n) = 1 - \frac{n}{M_a}$$

the *complexity ratio* of n.

We always have $0 \le CR(n) < 1$. In a certain sense the numbers n with a small complexity ratio are *simple* and those with a large complexity ratio are *complex*. To illustrate this we present here some numbers comparable in size but with different complexity ratios and their corresponding minimal representations.

n	$\|n\|$	CR (n)	Minimal Expression
371447	43	0.94	$1{+}2(1{+}2(1{+}2^2(1{+}2^2)(1{+}2(1{+}2^4(1{+}2^4 3^2)))))$
373714	40	0.82	$2(1{+}2^3(1{+}2^2(1{+}2{\cdot}3(1{+}2^2 3^5))))$
377202	39	0.76	$3(1{+}2{\cdot}3)^2(1{+}(1{+}2^2)(1{+}2{\cdot}3^2)3^3)$
377233	38	0.65	$(1{+}2^5 3)(1{+}2^4 3^5)$
360910	37	0.49	$(1{+}2{\cdot}3^3)(1{+}3^8)$
422820	37	0.40	$2^2(1{+}2^4 3^2)3^6$
492075	37	0.31	$(1{+}2^2)^2 3^9$
413343	36	0.22	$(1{+}2{\cdot}3)3^{10}$
531441	36	0	3^{12}

Let S be some (arbitrary but fixed) natural number (this will be the span of n). Choose S not too small. For example, $S = 1\,000\,000$. Let $M = \max\{\|s\| : 1 \leq s \leq S\}$. So, M is also fixed.

Now choose k such that $2^{3k} > S$. Clearly there are infinitely many such k.

Now let n satisfy $2^{3k} - S \leq n < 2^{3k}$, and let $\|n\| = 3a + r$ with $0 \leq r \leq 2$. Then we have

$$\mathrm{CR}\,(n) = 1 - \frac{n}{3^a}, \quad 1 - \frac{n}{4 \cdot 3^{a-1}}, \quad 1 - \frac{n}{2 \cdot 3^a}$$

for $r = 0$, 1 or 2, respectively. Therefore, in all cases we will have

$$\mathrm{CR}\,(n) = 1 - f\frac{n}{3^a}$$

where $f = 1$ for $r = 0$, $f = 3/4$ for $r = 1$ and $f = 1/2$ for $r = 2$.

Now choose a small fixed $p > 0$, ($p = 1/1000$, say).

Let's now consider the inequality

$$\mathrm{CR}\,(n) + p < 1 - \left(\frac{8}{9}\right)^k. \tag{2}$$

For large k this comes very close to the event $\mathrm{CR}\,(n) + p \leq 1$ or $\mathrm{CR}\,(n) \leq 1 - p$. Quite extensive statistics on $\mathrm{CR}\,(n)$ suggest strongly that this event is highly probable (for small $p > 0$). So, we venture to *assume* that we have (2). Observe that this is equivalent to

$$\left(1 - f\frac{n}{3^a}\right) + p \leq 1 - \left(\frac{8}{9}\right)^k.$$

Hence, since $f \leq 1$ (also using previous assumptions)

$$\frac{2^{3k}}{3^a} > \frac{n}{3^a} \geq f\frac{n}{3^a} \geq p + \left(\frac{8}{9}\right)^k = p + \frac{2^{3k}}{3^{2k}}$$

so that $2k > a$ or $2k - a > 0$.

Also observe that

$$p+\left(\frac{8}{9}\right)^k \le 1-\mathrm{CR}\,(n) = f\frac{n}{3^a} \le 3^{2k-a}\frac{n}{3^{2k}} < 3^{2k-a}\frac{2^{3k}}{3^{2k}} = 3^{2k-a}\left(\frac{8}{9}\right)^k$$

so that

$$p+\left(\frac{8}{9}\right)^k < 3^{2k-a}\left(\frac{8}{9}\right)^k \quad \text{or} \quad p\left(\frac{9}{8}\right)^k + 1 < 3^{2k-a}.$$

Now choose k so large that $3^{M+r} \le 3^{M+2} < p\left(\frac{9}{8}\right)^k + 1$, without violating previous assumptions.

Then we clearly have $2k-a > M+r$.

Now we can conclude that

$$\begin{aligned}\|2^{3k}\| = \|n + (2^{3k}-n)\| \le \|n\| + \|2^{3k}-n\| \le 3a + r + \|\text{some } s \le S\| \le \\ \le 3a + r + M < 3a + (2k-a) = 2a + 2k < 2(2k) + 2k = 6k\end{aligned}$$

so that

$$\|2^{3k}\| < 6k = 3k\|2\|.$$

Hence, the answer to Selfridge's question might very well be no.

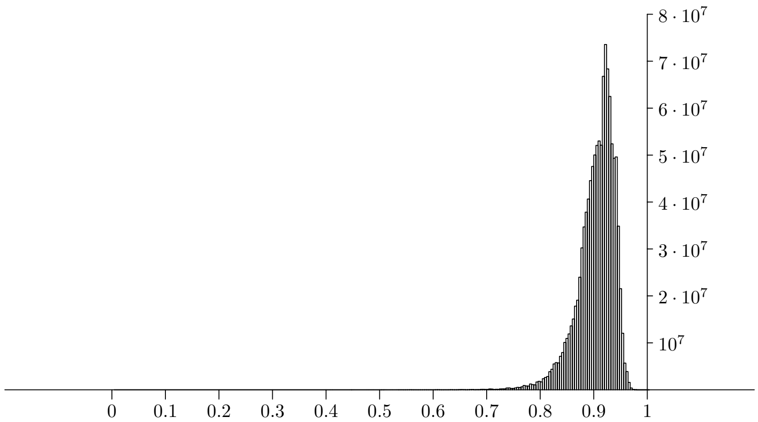

Figure 1: Distribution of $\mathrm{CR}\,(n)$ for $1 \le n \le 905\,000\,000$

References

[1] K. Mahler and J. Popken, *On a maximum problem in arithmetic* (in Dutch), Nieuw Arch. Wiskunde (3) **1** (1953) 1–15.

[2] R. K. Guy, *Unsolved Problems in Number Theory*, Third ed., Springer (2004) p. 399 (Problem F26).

[3] R. K. Guy, *Don't try to solve these problems!*, Amer. Math. Monthly, (1983) 35–41.

[4] R. K. Guy, *What is the least number of ones needed to represent n, using only + and × (and parentheses)?*, Amer. Math. Monthly, **93** (1986) 188–190.

[5] R. K. Guy, *Monthly unsolved problems, 1969–1987*, Amer. Math. Monthly, **94** (1987) 961–970.

[6] R. K. Guy, *Problems come of age*, Amer. Math. Monthly, **96** (1989) 903–909.

[7] D. A. Rawsthorne, *How many 1's are needed?*, Fibonacci Quart., **27** (1989) 14–17.

[8] J. Arias de Reyna, *Complejidad de los números naturales* (in Spanish), Gaceta R. Soc. Mat. Esp., **3** (2000) 230–250.

[9] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, (2008), published electronically at
http://oeis.org/

[10] E. Pegg Jr. *Problem of the week*, 10 April 2001,
http://www.mathpuzzle.com/Aug52001.htm

[11] E. Pegg Jr., *Math Games. Integer Complexity*,
www.maa.org/editorial/mathgames/mathgames_04_12_04.html

[12] Weisstein, E. W., *Integer Complexity. From MathWorld - A Wolfram Web Resource.*
http://mathworld.wolfram.com/IntegerComplexity.html

J. Arias de Reyna, Facultad de Matemáticas, Universidad de Sevilla, Apdo. 1160, 41080-Sevilla, Spain. e-mail arias@us.es.
Supported by grant MTM2006-05622.

J. van de Lune, Langebuorren 49, 9074 CH Hallum, The Netherlands (formerly at CWI, Amsterdam). e-mail j.vandelune@hccnet.nl