»   Solutions to puzzles in my National Review Online *Diary*

➤ June 2011

_____

In my June diary I posed this brain-teaser:

> For this month's puzzle you will need to be in Microsoft Word or some other app — Notepad, for example — that uses standard editing keys.
>
> I shall use the word "keystroke" to mean one of the following:
>
> - a (i.e. you just hit the letter "a" on your keyboard)
> - Ctrl-a (which does "select all" on whatever text is in your document)
> - Ctrl-c (copy selected text to clipboard)
> - Ctrl-v (paste selected text from clipboard to cursor location in document)
>
> Starting from a blank document you execute N keystrokes (as defined). What is the maximum number of a's you could end up with?

_____

Solution

I seek $f(N)$, the maximum number of a's I can produce from N keystrokes (as defined), beginning with a blank page.

For simplicity let's identify each of the four permitted keystrokes by a single symbol. The first I'll just call "a." The Ctrl keys I'll denote by the corresponding uppercase letter in boldface: **A**, **C**, **V**. Repeated strikes on the same key are shown by exponentiation, so for instance "$\mathbf{V}^4$" means "execute the Ctrl-v keystroke four times in succession." So now any sequence of keystrokes has symbolic representation as a string of a's, **A**'s, **C**'s, and **V**'s, any of which might be raised to any power.

Let's also note the following fact about standard editing. (Defined to be: the way things work in Microsoft Word 2010, Windows 7 Notepad, and the Mansfield Software Group's KEDIT for Windows, Version 1.6, in which I am writing this web page.)

> The 4-keystroke sequence $\mathbf{ACV}^2$ selects the entire text, copies it to the clipboard, pastes it to itself (i.e. having no visible effect on the text), then appends a duplicate right after the text, thereby doubling the number of characters on the page. $\mathbf{ACV}^3$ appends *two* duplicates, tripling the number of characters on the page. $\mathbf{ACV}^4$ appends *three* duplicates, quadrupling the number of characters on the page … and so on.

You can easily verify this. There are some slight wrinkles. "Right after" might include some whitespace (a line break in Word under default settings), but this makes no difference to the solution.

OK, here are some theorems:

> *Theorem 1.* When the number of a's on the current page is p, there are only two ways to increase the number of a's:
>
> 1. By executing keystroke a, increasing the number of a's by 1, from p to $p+1$, at a cost of 1 keystroke.
> 2. By executing a keystroke sequence $\mathbf{ACV}^k$, $k \geq 2$, increasing the number of a's by $(k-1)p$, from p to kp, at a cost of $k+2$ keystrokes.

This follows from the above fact about standard editing.

> *Theorem 2.* When $N < 8$, $f(N) = N$.

Obviously $f(1) = 1$. You just have to hit that a key, since currently $p = 0$ and an $\mathbf{ACV}^k$ has nothing to copy.

Executing the minimal $\mathbf{ACV}^2$ copy-paste sequence at this point would increase the number of a's from 1 to 2 while bringing N, my total number of keystrokes, up to 5. Plainly the better option for $n = 5$ is just to hit 5 a's; and the *only* option for $N = 2, 3,$ and 4 similarly. That gets us up to $f(5)$.

For $N = 6$ I have the options a^6, $a^2\mathbf{ACV}^2$, and $a\mathbf{ACV}^2a$, but the latter two only get me 4 and 3 a's respectively, so a^6 is better and $f(6) = 6$.

Likewise for $N = 7$ where I have seven options:

- a^7, yielding 7 a's.
- $a^3\mathbf{ACV}^2$, giving 6 a's.
- $a^2\mathbf{ACV}^3$, also for 6.
- $a^2\mathbf{ACV}^2a$, for 5 a's.
- $a\mathbf{ACV}^4$, for 4.
- $a\mathbf{ACV}^3a$, for 4.
- $a\mathbf{ACV}^2a^2$, for 4.

Clearly a^7 is the way to go, and $f(7) = 7$.

At $N = 8$ the copy-paste option kicks in. With 8 keystrokes allowed there are *ten* options:

- a^8, yielding 8 a's.
- $a^4\mathbf{ACV}^2$, also giving 8 a's.
- $a^3\mathbf{ACV}^3$, for 9.
- $a^3\mathbf{ACV}^2a$, for 7 a's.
- $a^2\mathbf{ACV}^4$, for 8.
- $a^2\mathbf{ACV}^3a$, for 7.
- $a^2\mathbf{ACV}^2a^2$, for 6.
- $a\mathbf{ACV}^4a$, for 5.
- $a\mathbf{ACV}^3a^2$, for 5.
- $a\mathbf{ACV}^2a^3$, for 5.

And the winner is … $a^3\mathbf{ACV}^3$, giving $f(8) = 9$.

Things are getting out of hand, though. Time for some simplification.

> *Theorem 3*. For any number N, the maximum possible number of a's is produced by a sequence of keystrokes looking like this
>
> $$a^{k_0}\mathbf{ACV}^{k_1}\mathbf{ACV}^{k_2}\mathbf{ACV}^{k_3} \dots \mathbf{ACV}^{k_m},$$
>
> where $m \geq 0$ (in the case $m = 0$ we hit nothing but the a key), $k_0 > 0$, the other k's are all ≥ 2, and $k_0 + k_1 + k_2 + \dots + k_m = N - 2m$.

To prove this, try "improving" $f(N)$ by inserting an a between two of the $\mathbf{ACV}^k$ groups, or at the end. To maintain N you must compensate by reducing one of the k's; but any way you do this diminishes the resulting number of a's. Or to look at it another way, a $\mathbf{V}$ keystroke will always get you more a's than an a keystroke!

> *Theorem 4*. The actual number of a's produced by that sequence of keystrokes $a^{k_0}\mathbf{ACV}^{k_1}\mathbf{ACV}^{k_2}\mathbf{ACV}^{k_3} \dots \mathbf{ACV}^{k_m}$ in Theorem 3 is $k_0 \times k_1 \times k_2 \times \dots \times k_m$.
>
> (Note that the number of keystrokes there is $k_0 + k_1 + k_2 + \dots + k_m + 2m$.)

This follows immediately from Theorem 1.

So:

> *Theorem 5*. $f(N) = \max(k_0 \times k_1 \times k_2 \times \dots \times k_m)$, taken over all possible $k_0 + k_1 + k_2 + \dots + k_m = N - 2m$, with $m \geq 0$, $k_0 > 0$, and the other k's all ≥ 2.

That's nice, but doesn't help much in actually computing values of $f(N)$. What, for example, is $f(100)$? We have to consider the cases:

- $m = 0$, $k_0 = 100$, number of a's = 100.
- $m = 1$, $k_0 + k_1 = 98$, number of a's is the maximum of 2×96, 3×95, 4×94, … 49×49 (all possible pairs adding up to 98). That maximum is actually 2,401.
- $m = 2$, $k_0 + k_1 + k_2 = 96$, number of a's is the maximum of $2 \times 2 \times 92$, $2 \times 3 \times 91$, … (all possible triplets adding up to 96). That maximum is actually $32 \times 32 \times 32$, which is 32,768.

- $m = 3$, $k_0 + k_1 + k_2 + k_3 = 94$, number of a's is the maximum of $2 \times 2 \times 2 \times 88$, $2 \times 2 \times 3 \times 87$, … (all possible quadruplets adding up to 94). That maximum is actually $23 \times 23 \times 24 \times 24$, which is 304,704.
- … … …
- … … …
- $m = 24$, $k_0 + k_1 + k_2 + \ldots + k_{24} = 52$. That's as big as m can get under its constraints. Remember the k's are all ≥ 2 by definition. Having all 25 of the k's equal to 2 would be good, except that only adds up to 50. Replace the last two 2's by 3's for a product $2^{23} \times 3^2 = 75{,}497{,}472$.

And we have to figure the maximum of all those 25 maxima!

The following theorem offers some further simplification:

Theorem 6. The maximum value (I'll call it Max) of $k_0 \times k_1 \times k_2 \times \ldots \times k_m$, under the constraint that the k's add up to some fixed total T, occurs when no two k's differ by more than 1.

In other words: For some number u, all the k's are equal either to u or to $u+1$.

In *other* other words, Max $= u^p(u+1)^q$, where $p + q = m$ and, just adding up all the k's, $up + uq + q = T$.

The proof is by *reductio ad absurdum*. Suppose the desired Max occurs when the first i_0 of the k's are all equal to u, the next i_2 are all equal to $(u+\alpha)$, the next i_3 are all equal to $(u+\beta)$, and so on, all the Greek letters being *greater* than 1.

Consider the following set of k's, which also (you might want to check this out) add up to the same total T: the first (i_0-1) of the k's are all equal to u, the next (i_2-1) are all equal to $(u+\alpha)$, the next one of the k's is $(u+1)$, the next one is $(u+\alpha-1)$, then the next i_3 are all equal to $(u+\beta)$, and from here on as before.

The product of all *these* k's is *greater* than Max, because I replaced a u and a $(u+\alpha)$ by a $(u+1)$ and a $(u+\alpha-1)$. The product of the replacees is $u^2+\alpha u$; the product of the replacers is $u^2+\alpha u+\alpha-1$; and α is by definition greater than 1.

Therefore Max isn't the max!

Lather, rinse, repeat. Q.E.D.

That simplifies the computation of $F(100)$ a lot. The last set of bullet points now look like this:

- $m = 0$, $k_0 = 100$, number of a's = 100.
- $m = 1$, $k_0 + k_1 = 98$, number of a's can only be $49 \times 49 = 2{,}401$.
- $m = 2$, $k_0 + k_1 + k_2 = 96$, number of a's can only be $32 \times 32 \times 32$, which is 32,768.
- $m = 3$, $k_0 + k_1 + k_2 + k_3 = 94$, number of a's can only be $23 \times 23 \times 24 \times 24$, which is 304,704.
- … … …
- … … …
- $m = 24$, $k_0 + k_1 + k_2 + \ldots + k_{24} = 52$. That's as big as m can get under its constraints. The maximum product of the 25 k's occurs when 23 of them are equal to 2 and the other two are equal to 3. That gives $2^{23} \times 3^2 = 75{,}497{,}472$.

In the bullet point corresponding to m there, the "average" value of k is the arithmetic mean $(100-2m) / (m+1)$, and the "average" product I end up with in that bullet point is just that number raised to the power of $(m+1)$. A bit of brute calculus will find the maximum possible value of *that*: it occurs when $m = 15.141445\ldots$ Given that we're dealing with "average" & approximate values here, $m = 15$ or 16 does indeed look like our best bet.

It's actually more interesting to write that product in terms of the "average" k, and *then* do the calculus to maximize it. The product is $k^{102/(k+2)}$, which has a maximum at $k = 4.319136\ldots$

Note that this is a constant, independent of N. It's actually $2/W(2/e)$, where $W(x)$ is the Lambert W-function, i.e. the value of y satisfying $x = ye^y$, and e is of course the familiar base of natural logarithms $2.718281\ldots$

This value for maximum k implies that our maximum $f(N)$ will be attained when most of the k's are 4's, with a sprinkling of 5's to bring the sum of the k's up to $N - 2m$.

That's what happens for sufficiently large N. The actual value of $f(100)$ is 17,179,869,184, corresponding to the keystroke sequence $a^4(\mathbf{ACV}^4)^{16}$, i.e. $m = 16$ and all seventeen of the k's are equal to 4. Then $f(101)$ is $4^{16} \times 5$, $f(102)$ is $4^{15} \times 5^2$, $f(103)$ is $4^{14} \times 5^3$, $f(104)$ is $4^{13} \times 5^4$, and $f(105)$ is $4^{12} \times 5^5$. The pattern then repeats in a cycle of six (because you need six keystrokes to execute a new $\mathbf{ACV}^4$ sequence), $f(106)$ being 4^{18}, $f(107)$ being $4^{17} \times 5$, *und so weiter*.

For smaller values of N the pattern breaks up, as usually happens when calculus meets whole numbers. If you try the above bullet-point procedure for $N = 20$, for example, here's what you get:

- $m = 0$, $k_0 = 20$, solution 20.

- $m = 1$, $k_0 + k_1 = 18$, solution $9 \times 9 = 81$.
- $m = 2$, $k_0 + k_1 + k_2 = 16$, solution $5 \times 5 \times 6 = 150$.
- $m = 3$, $k_0 + k_1 + k_2 + k_3 = 14$, solution $3 \times 3 \times 4 \times 4 = 144$.
- $m = 4$, $k_0 + k_1 + k_2 + k_3 + k_4 = 12$, solution $2 \times 2 \times 2 \times 3 \times 3 = 72$.

This irregularity for smaller values of N makes it hard to come up with a single formula. The simplest solution I can state is:

- For $N = 1, 2, 3, 4, 5, 6$, and 7, $f(N) = N$;
- For $N = 8, 12, 13, 19, 20, 26$, and 33, $f(N)$ has the values 9, 25, 30, 125, 150, 625, and 3125 respectively;
- For all other N, $f(N) = 4 \times f(N - 6)$.

If you don't like recursive formulas, readers have come up with explicit ones. The most comprehensive solution came from Dr. Jonathan Rowell:

Using the "floor" function $\lfloor x \rfloor$ (the biggest whole number not bigger than x) and the "ceiling" function $\lceil x \rceil$ (the least whole number not less than x), $f(N)$ is equal to

- N for $N = 1, \ldots, 7$;
- $\lfloor (N - 2) / 2 \rfloor \times \lceil (N - 2) / 2 \rceil$ for $N = 8$;
- $Q^{(C - R)} \times (Q + 1)^R$ for $N = 9, \ldots, 27$, where C is $\lfloor (N + 3) / 6 \rfloor$ and Q and R are, respectively, the quotient and remainder on dividing $(N - 2C - 2)$ by C.
- $4^{(Q' - R')} \times 5^{R'}$ for $N \geq 28$, where Q' and R' are, respectively, the quotient and remainder on dividing $(N + 2)$ by 6.

Jonathan Campbell has basically the same result here — the second of his results, the first one making different assumptions than mine about copy-paste keystroke sequences.

The first 100 values of $f(N)$ are:

$N = 1$ to 10:	1, 2, 3, 4, 5, 6, 7, 9, 12, 16
11 to 20:	20, 25, 30, 36, 48, 64, 80, 100, 125, 150
21 to 30:	192, 256, 320, 400, 500, 625, 768, 1024, 1280, 1600
31 to 40:	2000, 2500, 3125, 4096, 5120, 6400, 8000, 10000, 12500, 16384
41 to 50:	20480, 25600, 32000, 40000, 50000, 65536, 81920, 102400, 128000, 160000
51 to 60:	200000, 262144, 327680, 409600, 512000, 640000, 800000, 1048576, 1310720, 1638400
61 to 70:	2048000, 2560000, 3200000, 4194304, 5242880, 6553600, 8192000, 10240000, 12800000, 16777216
71 to 80:	20971520, 26214400, 32768000, 40960000, 51200000, 67108864, 83886080, 104857600, 131072000, 163840000
81 to 90:	204800000, 268435456, 335544320, 419430400, 524288000, 655360000, 819200000, 1073741824, 1342177280, 1677721600
91 to 100:	2097152000, 2621440000, 3276800000, 4294967296, 5368709120, 6710886400, 8388608000, 10485760000, 13107200000, 17179869184.

There are some more links & cross-references on the splendid OEIS entry, for which I seem to have been the inspiration. (At any rate, when I generated the first 20-odd terms this time last week & checked them against OEIS — always the first port of call when you've generated an integer sequence — I got no match. Today I got that match.)

Many thanks to Jonathan, Dr. Rowell, and all others who helped with this. The friend who set it, by the way, tells me this was one of a number of questions he was "advised to ponder" in preparation for a technical interview at a very large and important software company. He adds:

Since this problem cannot possibly be solved during a 45-minute verbal interview, I assume they pose it simply to test a candidate's ability to comprehend what is, and what is not, a truly challenging intellectual exercise.

Top of this page