But Seriously:

Real Results from Joke Sequences

Paul Michael Kielstra

November 7, 2017

1 Introduction

The Look and Say Sequence, otherwise known as the See what you Say Sequence, goes like this: $1, 11, 21, 1211, 111221, 312211....$ Its main use in mathematics before now has been to gain a form of petty revenge on colleagues. Alice will give the sequence to Bob, asking innocently if he might be able to help with her latest research project. Bob will go away and come back hours later, admitting to Alice that he cannot guess at the generating rule for the sequence, whereupon Alice will reveal that all you have to do is read out the numbers. You start with 1, and read that as “One one”, or 11. This, you read as “Two ones”, or 21, which is in turn read as “One two, one one” or 1211 and so on. Bob leaves, feeling stupid that he failed to guess at such an apparently simple idea, and the cause of mathematics is not advanced at all.

In this paper, I offer a remedy in the form of a number of small proofs which hint at much deeper structures within the sequence. My aim is to turn it from a curiosity into a rigorous object, with a well-formalized theory and links to other areas of mathematics, that can stand up to further analysis. The OEIS mentions very few papers written about this sequence (A005150); it is entirely possible that mine is the first to attempt to analyze it in such a way.

2 Notation and definitions

The Look and Say Sequence is singularly poorly designed to be defined by normal mathematical notation. I have therefore invented some of my own. The notation ++ means “Concatenate”, so $21{+}{+}17 = 2117$. For convenience, we will denote the n-th number in the sequence by L_n, and the n-th number in the sequence in base b by ${}^bL_n$. We define the Look and Say Sequence algorithmically. To generate L_{n+1} from L_n, do the following:

1. Split L_n, into the smallest possible number of strings that each consist of repetitions of one digit only. For example, we split 111221 into 111, 22, and 1.

2. Replace every string of length k containing only the digit a with the string $k{+}{+}a$. For example, we replace 111 with 31 as it contains only the digit 1 and is of length 3.

3. Recombine the strings in the original order. For example, we recombine the given strings to get $31{+}{+}22{+}{+}11 = 312211$, which is the number that comes after 111221 in the sequence.

Note that this means that the substring 1111 will give rise to 41, not 2121. Should a substring exist that could have been written more succinctly, it will be called *malformed*. Obviously, no malformed substrings should ever arise in any Look and Say Sequence number. The *classical sequence* (my notation) arises from letting $L_0 = 1$.

3 Preliminary analysis

Theorem 1 *No member of the classical sequence will contain a digit greater than* 3.

If L_{n+1} contains a digit greater than 3, L_n must contain a string consisting of more than 3 repetitions of a single digit. Consider the smallest n for which this is true. Then L_n must contain a substring string $kkkk...$, which can be read in a number of ways. If it, or part of it, denotes kk or more of something, then L_{n-1} also contained a string of more than 3 repetitions of a single digit as any two-digit number must be bigger than 3. Thus, it must denote either some k, k of k, and k of something, or k of k and k of k. In either case it is malformed. We now have a contradiction as no malformed strings should ever exist. Note that this also implies that no member of the classical sequence will contain more than 3 consecutive instances of a single digit.

Theorem 2 *The classical sequence is exactly the same in any base greater than or equal to* 4*.*

This follows immediately from the last theorem. If only three symbols – $1, 2, 3$ – are required to form the classical sequence, then it is formed in exactly the same way if we carry out the algorithm in any base that contains three or more non-zero symbols.

This leaves the question of what to do in bases 1, 2, and 3, which form the next large chunk of this paper.

4 The unary classical sequence

Theorem 3 *If the classical sequence is generated in base* 1*, it gives rise to the positive integers.*

We prove this by induction. For the base case, ${}^1L_0 = 1$ by the definition of the classical sequence. Now, assume that ${}^1L_n = n$. Expressed in base 1, n is a string of length n containing only the digit 1. Thus, to get ${}^1L_{n+1}$, we replace this string with $n + +1$. But, in base 1, this is just a string of length $n + 1$ containing only the digit 1, which is also the expression of $n + 1$ in base 1. This shows that if ${}^1L_n = n$, then ${}^1L_{n+1} = n + 1$ and completes both the inductive step and the proof.

This does to a large extent preclude interesting and novel analysis of the unary classical sequence since the sequence $1, 2, 3, 4...$ has already been studied fairly exhaustively. With this in mind, we go on to the far more complicated binary classical sequence.

5 The binary classical sequence

The binary classical sequence is generated similarly to the unary one. We start with 1 and 11, as usual. Then, instead of generating 21, we instead have 101 as 2 is 10 in binary. The next number in the sequence is thus 111011, and so on. We may prove a number of results about this sequence.

Theorem 4 ${}^2L_n$ *will never contain a string of more than two zeros.*

To prove this, we consider whether there exists a ${}^2L_{n-1}$ which would give rise to a string of more than two zeros. Clearly, such a string must be generated by a longer string of zeros – the string "00000000" would give rise to 8++0 or, in binary, 100++0 = 1000. However, such a string could only exist as part of 100000000, since no number used in the generation of the Look and Say Sequence is given a leading zero. The string 100000000 could come either from a string of 2^8 zeros, or from a string of 2^9 1s (which gives us 1000000001). In either case, only an even longer string could have given rise to our eight zeros. Thus, the previous number in the sequence contains such a long sequence, which could only have come from an even longer sequence, and so on. Therefore, the length of ${}^2L_1$ must be greater than 1. However, ${}^2L_1 = 1$, so this is clearly absurd and ${}^2L_n$ will never contain more than two zeros in a row.

Theorem 5 *${}^2L_n$ will never contain a string of more than four ones.*

Consider the smallest n for which ${}^2L_n$ contains a string of five or more ones. We must now ask where this string came from. If it denotes nothing more than a longer string of ones, clearly ${}^2L_{n-1}$ contains a string of five or more ones and there is a contradiction. Should we split the string to denote 111 or more 1s, followed by 1 or more zeros, then ${}^2L_{n-1}$ contains 11 or more ones and we have the exact same contradiction. Should we split the string to denote 11 or fewer ones followed by 11 or more zeros, we have a sequence of three or more zeros in ${}^2L_{n-1}$, which is in contradiction with the previous theorem. Thus, no matter which way we split the string, we get a contradiction and no number in the classical sequence will ever contain a string of more than four ones.

Theorem 6 *The number of zeros in ${}^2L_{n-1}$ will always be less than or equal to the number of zeros in ${}^2L_n$.*

Consider the two previous theorems. For the number of zeros to decrease as n increases, we must have a string containing a certain number of zeros which gives rise to a string containing fewer. Since the only strings containing zeros are 0 and 00, which generate the strings 10 and 10++0 = 100 respectively, the number of zeros in successive numbers in the binary classical sequence will never decrease.

Theorem 7 *${}^2L_{n-1}$ is always shorter than ${}^2L_n$.*

To prove that the length of the binary classical sequence is strictly increasing, we consider all possible substrings. As shown in the proof of the previous theorem, substrings consisting of zeros always increase in length, going from, for example, 0 to 10. The only possible substrings consisting of ones are 1, 11, 111, and 1111, by theorem 5. These go to 11, 101, 111, and 1001 respectively, none of which are shorter than the substrings that created them. Therefore, any time ${}^2L_n$ contains a zero, it will be shorter ${}^2L_{n+1}$. By theorem 6, if ${}^2L_n$ contains a 0, so does ${}^2L_{n+1}$. Since ${}^2L_3 = 101$ contains a zero, this theorem holds true for all ${}^2L_n$ where $n \geq 2$ by induction. Simple calculation proves the rest.

In summary, the binary classsical sequence is strictly increasing in length as well as in value, while the number of zeros is increasing. This provides, among other things, a fairly elegant proof that the series created by summing the terms of the sequence does not converge. The next stop on our tour is the "trinary classical sequence", or the classical sequence derived in base 3.

6 The trinary classical sequence

Theorem 8 *${}^3L_n$ will never contain a string of more than three of the same digit.*

Consider the smallest n for which ${}^3L_n$ contains more than three of the same digit. The string $kkkk$ could come from a few different things: $kkkk$ of something else, k of k and then k of k, as kk of k and then k of something else, or as some k, k of k, and then k of something else. All of these possibilities are malformed or necessitate that ${}^3L_{n-1}$ also contained more than 3 ks in a row. Either way, we have a contradiction and no ${}^3L_n$ should contain more than 3 instances of a single digit in a row.

Theorem 9 *${}^3L_n$ will never contain more than one zero in a row.*

Consider the smallest n for which ${}^3L_n$ contains a string of two or more zeros. This must be part of 100... or 200..., since substrings with leading zeros are malformed. Now, the string 100... either implies that ${}^3L_n$ contained three zeros, or it implies that it contained nine or more of something else. Either way, we have a contradiction. A similar argument works for 200....

Theorem 10 *The number of zeros in ${}^3L_{n-1}$ will always be less than or equal to the number of zeros in ${}^3L_n$.*

Similarly to the equivalent proof for the binary sequence, we consider strings of zeros that can arise within the trinary classical sequence. There is only one, 0, which gives rise to the sequence 10. As these have the same number of zeros, it is clear that the number of zeros can never decrease from ${}^3L_{n-1}$ to ${}^3L_n$.

Theorem 11 *${}^3L_{n-1}$ is always shorter than or the same length as ${}^3L_n$.*

Consider all possible substrings of ${}^3L_{n-1}$. Those of length 1 and 2 give rise to a string of length 2 (for example, 1 goes to 11, which itself goes to 21), while those of length 3 give rise to a string of length 3 (222 gives rise to 102 since 3 is 10 in base 3). Thus, no string ever gives rise to a string shorter than itself.

7 The quatenary classical sequence

The most important theorem of this section has been proven earlier: theorem 1. However, it is still worth considering this base separately in order to establish two results, which the reader will undoubtedly notice have parallels in all other bases.

Theorem 12 *L_n will never contain strings of more than two $3s$.*

Consider the smallest n for which L_n contains the string 333.... Because, as a consequence of theorem 1, L_n cannot contain a string of length 4 or more, the string 333..., which contains a 33, implies that L_{n-1} also contains three threes. This is a contradiction.

Theorem 13 *L_{n-1} is always shorter than or the same length as L_n.*

Split up L_{n-1} according to the first step in the algorithm. By theorem 1, we now have strings of length 1, 2, or 3 containing the digits 1, 2, or 3. Strings of length 1 are replaced in L_n by strings of length 2, as are strings of length 2. For example, 2 is replaced by 12 and 11 is replaced by 21. Therefore, if we are to replace a string with something shorter, the original string is of length 3 and will be replaced by a string of length 2. Now, consider what can follow a string of length 3. Clearly, a string of length 3, as, for example, the string 333222 seems to imply that the previous number either contained a certain number of 3s followed by three threes, which should not be written as $x333$ but instead as $(x+3)$++3, or three twos and two twos, which should not be written as 3222 but as 52. The problem is just the same if the two strings of three digits are separated by an even number of other digits. For example, the string 3332211333 could be read as "three threes, three twos, two ones, one three, two threes" or "x threes, three threes...", both of which are wrong. Therefore, the only possibility is for a string of length 3 to be followed by zero or more strings of length 2 and then a string of length 1. Taken together, this entire ensemble does not change in length – the string of length 3 decreases to one of length 2, the string of length 1 increases to one of length 2 and the strings of length 2 each remain the same length. The only alternative is for the sequence to end with a string of length 3 and then a series of strings of length 2. In this case, because the length of every member of the sequence except the first is even, there must be a string of length 1 somewhere before that is not yet matched up with a string of length 3 in this way. Note that it is impossible to match up two strings of length 3 as shown earlier in this paragraph.

8 Tabulation of results

It is now interesting to create a table of maximum lengths of strings of certain digits for the classical sequence in certain bases. The upper bounds have been established earlier; examples of strings that reach these upper bounds can be easily found through direct calculation.

	Base 1	Base 2	Base 3	Base ≥ 4
Maximum 0s	0	2	1	0
Maximum 1s	∞	4	3	3
Maximum 2s	0	0	3	3
Maximum 3s	0	0	0	2

We note that the number of consecutive digits decreases as the base increases. This is due to the greater information density of a single symbol in a higher base: more symbols would be needed to encode a given number in lower bases.

9 Conclusion

This paper has provided a rigorous study of the Look and Say Sequence and its behaviour in different bases. We have observed several common threads: the length of the sequence is always increasing; there is a hard limit on the length of substrings containing only one distinct digit; and the total number of occurrences of particular digits within any number in the sequence will obey certain rules. These are by no means the only results obtainable from this sequence – I have truncated this paper due to respect for my readers' valuable time rather than a lack of ideas. I believe, however, that the most important innovation of this paper is to place the Look and Say Sequence firmly within the bounds of serious professional Mathematics. Previously, the most profound result about it was due to John Conway, a man known for his mathematical iconoclasm and strongly idiosyncratic sense of "fun." Should you have had any fun reading this, it was unintentional and I sincerely apologise.