Some inner structures of look-and-say sentences

In memoriam John Conway

1. Abstract

Our main observations are:

(a) Every look-and-say sequence has a typical, fractal-like structure if we write the first, second, third, etc. lines of a certain look-and-say sequence under each other.

(b) All of the look-and-say sentences' structures seem to be fundamentally similar.

(c) Examining every look-and-say sentences up to one million; and then up to one billion randomly choosing one million numbers, roughly the 80 percent of every sentences' first column (in short: FC) begins with the "n, 1, 1, 3" pattern where "n" is the first digit of the given number (e.g. 1 if the starting number is 1 or if the starting number is 119). 1,1, and 3 are the starting digits of the second, third and fourth rows. Here "n" is called "head" and the infinitely repeating pattern is the "tail".

(d) Sooner or later the FC of every look-and-say sequence takes the 1,1,3 tail and, from that point, this pattern is repeating to the infinity.

(e) In the case of a second-generation FC (where the inputs, instead of the look-and-say sequences, the results of the comparisons of two look-and-say sequences), the tail is either 0,0,0 or 0,2,2.

(f) It is plausible that every tail is 3 digit long and there are only three tail versions.

(g) Comparing the digits of two look-and-say sequences in their overlapping parts, two main types can be identified: Where the number of the non-equal digits (difference, or, in short, DIFF) of the same rows in the same position is almost zero; and where the digits of the two look-and-say sequences mainly different.

(h) A DIFF converges to a certain value if the tail of the comparison of two look-and-say sequences is 0,0,0. Otherwise, it converges to infinity.

(i) The DIFFs of some different pairs are identical, thus either there are some pairs where this equality exists, or every pairs' DIFFs belongs to a group where the DIFFs are the same.

2. The FC structure of the look-and-say sequences

Examining those look-and-say sequences where the first line (=starting number) is 1; and where it is 333 as examples, it is easy to see the similarities between their structures (see **Figure 1**, and **Figure 2**. All illustrations below are in the Appendix). According to our conjecture, every look-and-say sequence has a similar structure–and similarly, the structure of the FCs is similar:

starting number	**head**	**tail**	**FC**
1	1,1,2	1,1,3	=1,1,2 1,1,3,1,1,3,1,1,3…
333	3,3,2	1,1,3	=3,3,2 1,1,3,1,1,3,1,1,3…

Notice that in every FC, the tail is 1,1,3 and this pattern is repeating infinitely.

The FC of the 1 and the 333 look-and-say sequences are not identical. In the case of the second sequence, the head of the FC is 3,3,2 to the beginning of the tail.

See the Appendix about the lists of the first digits in the vertical columns of the first 50 look-and-say sentences (**Figure 3**).

3. Some head statistics:

The most usual head form is a sole number between 2 and 9 following by the tail. For the first million look-and-say sequence, the proportion of this head form is about 80 percent.

n	**occurrences**
1	0
2	99100
3	99100
4	99910

5	99991
6	100000
7	100000
8	100000
9	100000

Altogether: 798101 (79.8%)

Examining 1 million random samples up to 100 million, this is 78.9%; and up to one billion for 1 million random samples, it is 78.9%. Thus, the occurence of this type of look-and-say sequence seems to be roughly the same for larger numbers. Similarly, these statistics seem to support the conjecture that there are slightly more look-and-say sentences beginning with 6,7,8 or 9 than with 2,3,4,5 (while there isn't a look-and-say sentence beginning with a sole 1 as head followed by the tail).

4. Overlapping proportions in look-and-say sequences:

4.1 Identical heads

Comparing the overlapping parts of the look-and-say sequences, there are pairs where the difference is minimal, e.g. in the case of 23 – 240777 pair (**Figure 4.**), the DIFF (=proportion of non-zero and zero items) is 6.812778046504023e-06 at the 40th row's last digit (**Figure 5**.). This DIFFs, that are calculated for the rows of the comparisons independently, converge to a certain value in every case. The next table contains some, randomly chosen examples where the heads are identical. In the FC column, the first digit is the head; the second three digits are the tails of the results of the comparisons of two look-and say sequences:

pair	heads	DIFF	FC	
23 - 240777	2 -2	6.812778046504023e-06	0	0,0,0
60 - 6253	6 - 6	6.812778046504023e-06	0	0,0,0
60 - 621855	6 - 6	6.812778046504023e-06	0	0,0,0
2759 - 240777	2 - 2	0.30437448478366025	0	0,0,0
60 - 637	6 - 6	0.3509348210054711	0	0,0,0
60 - 6153	6 - 6	0.6069190102463484	0	0,0,0
3 - 35	3 - 3	1.6744604316546763	0	0,0,0

Every FC is equal if the heads in the two look-and-say sequences used for comparison are equal, and according to our actual knowledge (that is based on a very limited number of observations), the values of the DIFFs are somewhere between 6.8e-06 and 1.7.
It did not escape our attention that the DIFFS of some different pairs seem to be identical, and it is possible that the DIFFs (or, at least, some of them) can be organized into groups.
E.g. 23 – 240777; 60 – 6253; 60 – 621855. This seems to be at least partly true for the pairs with either partly different or different heads, too (see below).

4.2 Partly different heads:

If the head is partly different, the values of the DIFFS are somewhere between 2.7e-05 - 6.3:

pair	heads	DIFF	FC	
69 - 208	6 - 2	2.7251297843059776e-05	4	0,0,0
23 - 69	2- 6	2.7251297843059776e-05	4	0,0,0
23 - 65157	2- 6	2.7251297843059776e-05	4	0,0,0
69 - 345	6 - 3	3.960791366906475	3	0,0,0
334 - 449	32 - 42	1.1836959613196814	1	0,0,0
334 - 770	32 - 72	1.1836959613196814	4	0,0,0
208 - 345	2- 3	6.24306269270298	1	0,0,0

4.3 Different heads:

If the heads of the compared look-and-say sequences are different, the DIFFs will be (usually) significantly larger than 0. If the tail is 0,2,2 then not only one, but three consecutive values are shown in the DIFF column. For some details, see **Figure 8** and **Figure 9**.

pair	**head**	**DIFF**	**FC**
2- 9	2 - 9	1.3625648921529888e-05	7 0,0,0
1 - 9	1,1,2 - 9	1.3625648921529888e-05	6 0,0,0
1 - 2	1,1,2 - 2	22466.0, 0, 0	6,1,0,2,1 0,2,2
1- 3	1,1,2 - 3	22466.0, 0, 0	2,0,1,2 0,2,2
79 - 110	7 – 1,2,1,1,2	0, 35717.0, 0	6,1,0,2,1 0,2,2
341 - 449	3 – 4,2	0, 86170.0, 0	1,1 0,2,2
208 - 449	2 – 4,2	0, 86821.0, 0	2,1 0,2,2

Some FCs' tails instead of 0,0,0 is 0,2,2 and the values of the DIFFs alternating between 0 and a certain, changing value in most cases . Notice that the DIFF is small if the tail is 0,0,0. IN other words, the prerequisite of a large DIFF is a tail in 0,2,2 form.

5. Second-generation FC structure

In the above, we compared the overlapping parts of two look-and-say sequences. If we use the results of these comparisons (e.g. 208 - 345 and 334 – 770) as inputs, then we can study these new results, too.

Some examples:

pairs	**heads**	**DIFF**	**FC**
1 - 9 and 1- 69	1,1,2 - 9 and 1,1,2 - 9	0.0	3 0,0,0
208 - 345 and 334 - 770	2 – 3; 32 - 72	4.661253854059609	3 0,0,0
1 - 69 AND 79 - 110	1,112 – 6 and 7 – 1,2,1,1,2	0, 0, 34916.0	1,1,1,0,1,2 0,2,2

Thus, it seems to be plausible to state that the second generation FC structure is similar to the first generation's structure. The tail is either 0,0,0 or 0,2,2 and according to our conjecture, the same is true for a third, fourth, etc. generation FC structures as well. The two main tail versions (1,1,3 vs 0,0,0/0,2,2) are mutually exclusive.

FC generation	tail: 1,1,3	tail: 0,0,0	0,2,2
1st	+	-	-
2nd	-	+	+
3rd	-	+	+
4th	-	+	+

Sources:

https://mathworld.wolfram.com/LookandSaySequence.html
https://oeis.org/A005150

Programs written by me.

Zoltán Galántai
April 18, 2020

Appendix

Figure 1: look-and-say sequence's inner structure (starting number: 1):

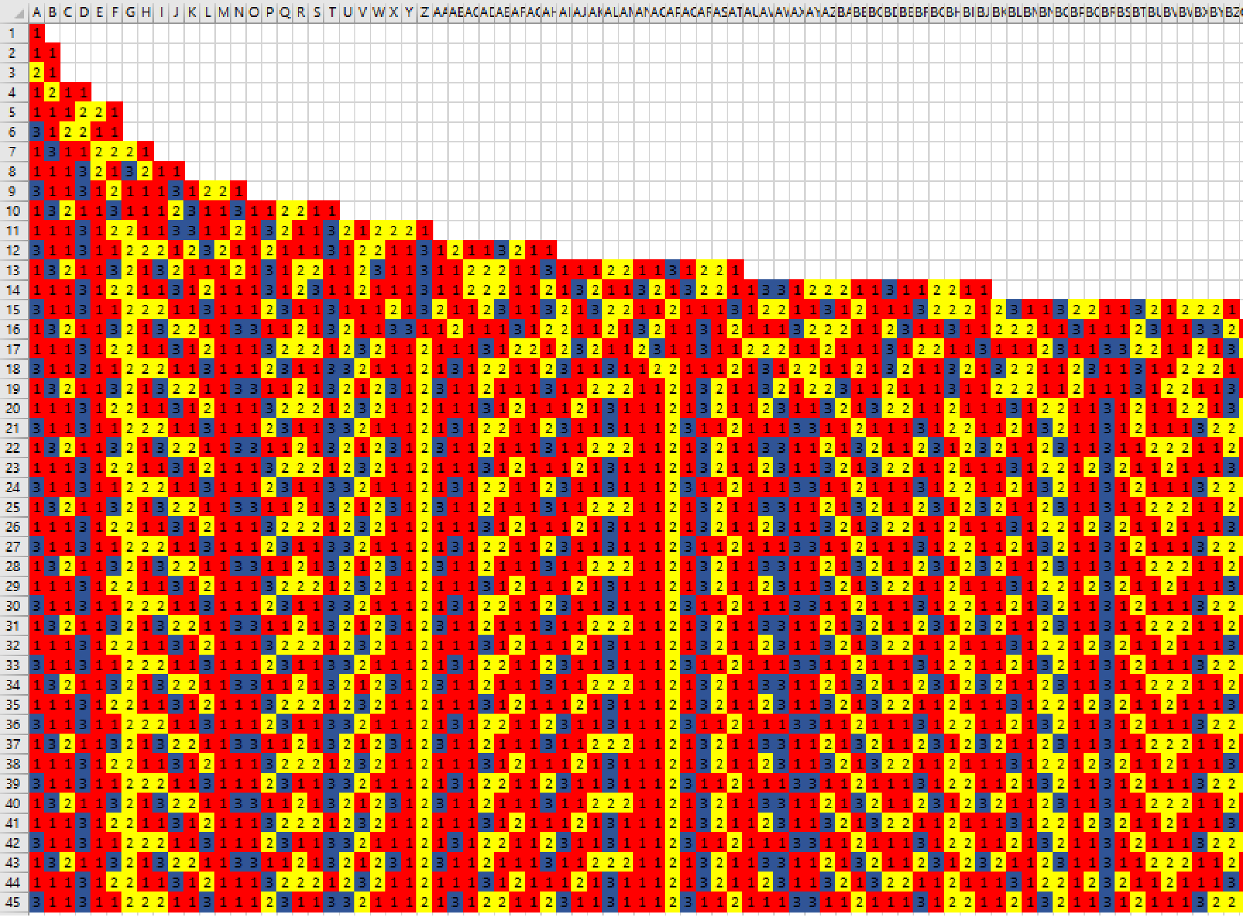


Figure 2: look-and-say sequence (starting number: 333)

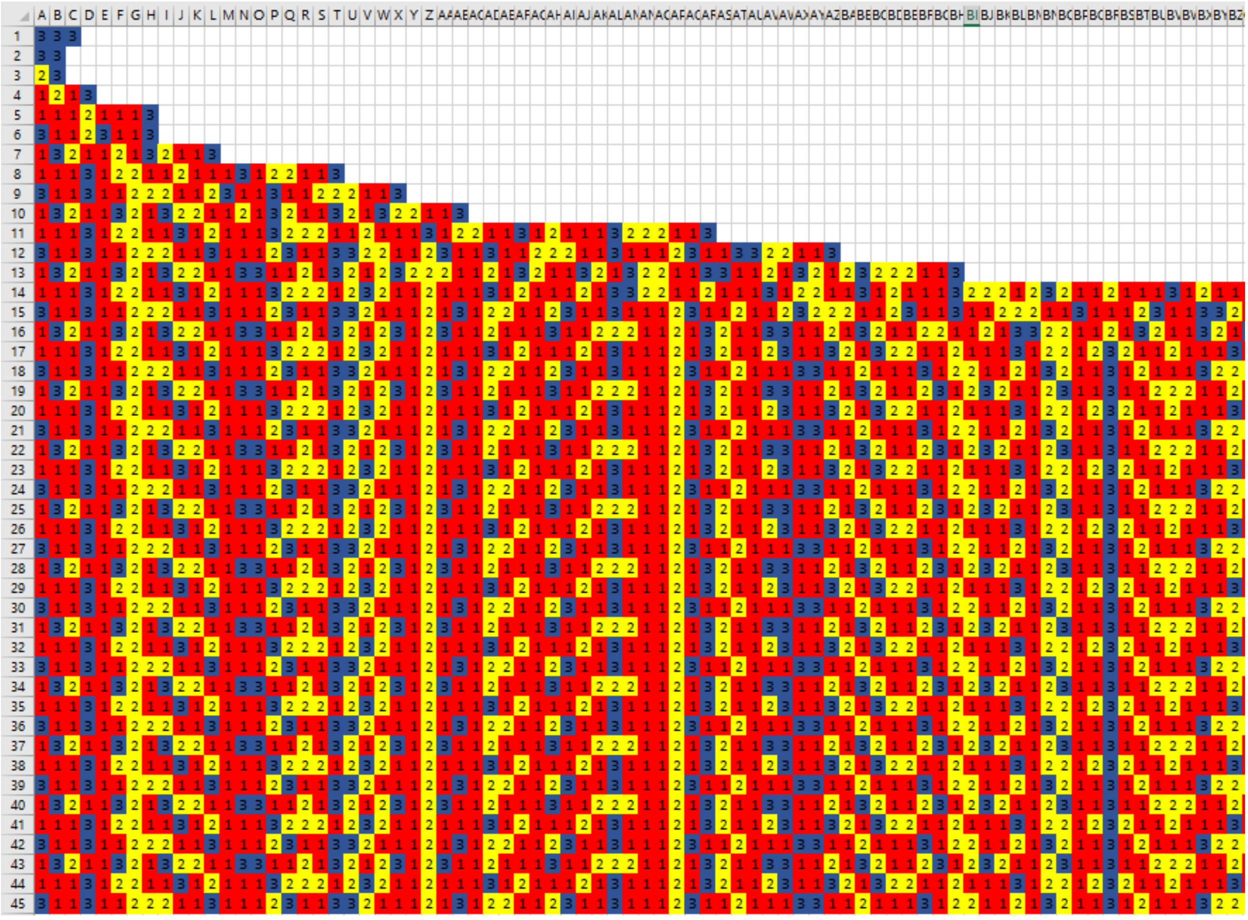


Figure 3: first digits in the first vertical columns of the look-and-say sequences up to 50:

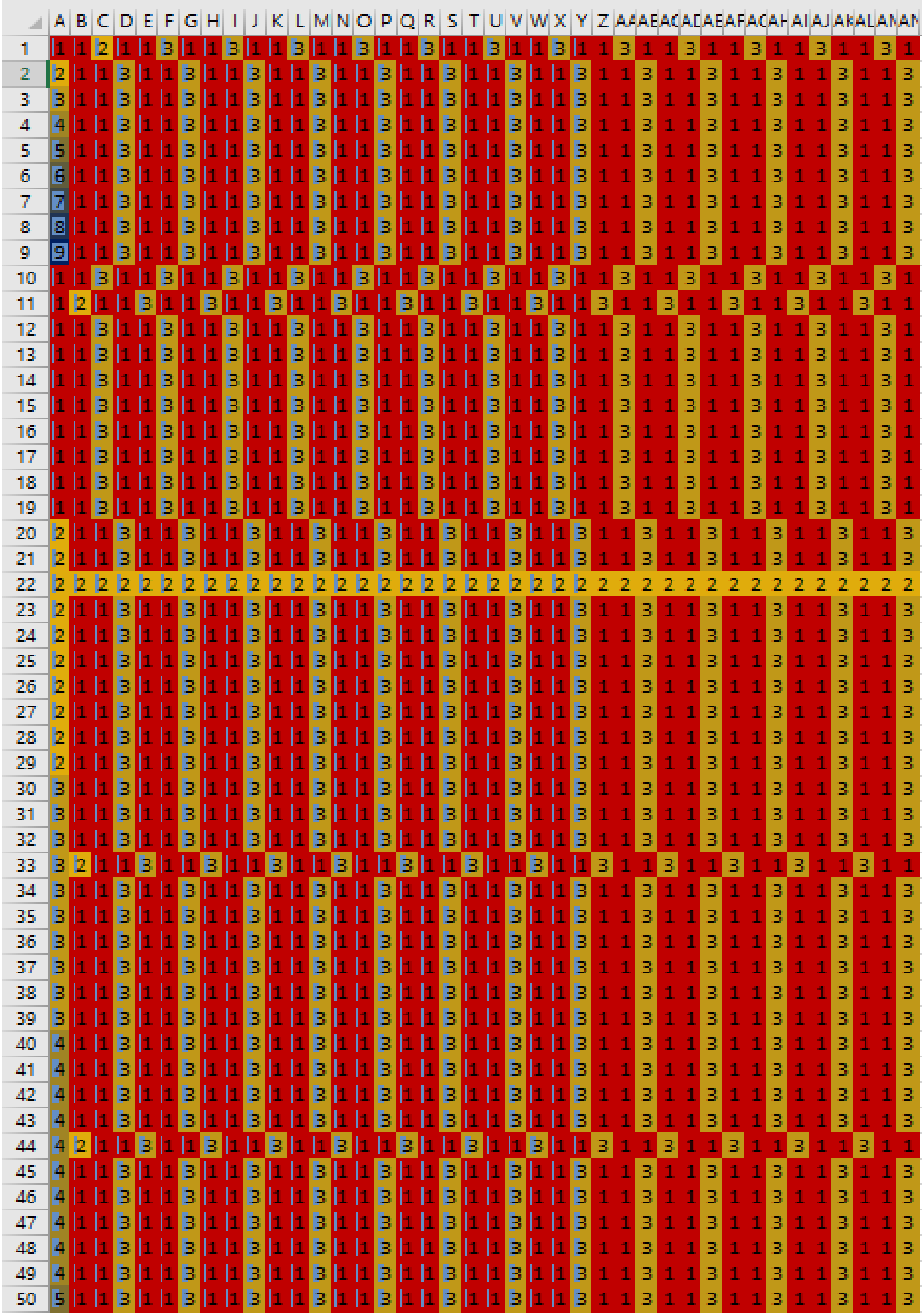


Figure 4: Overlapping parts of 23 and 240777: look-and-say sequences with identical heads

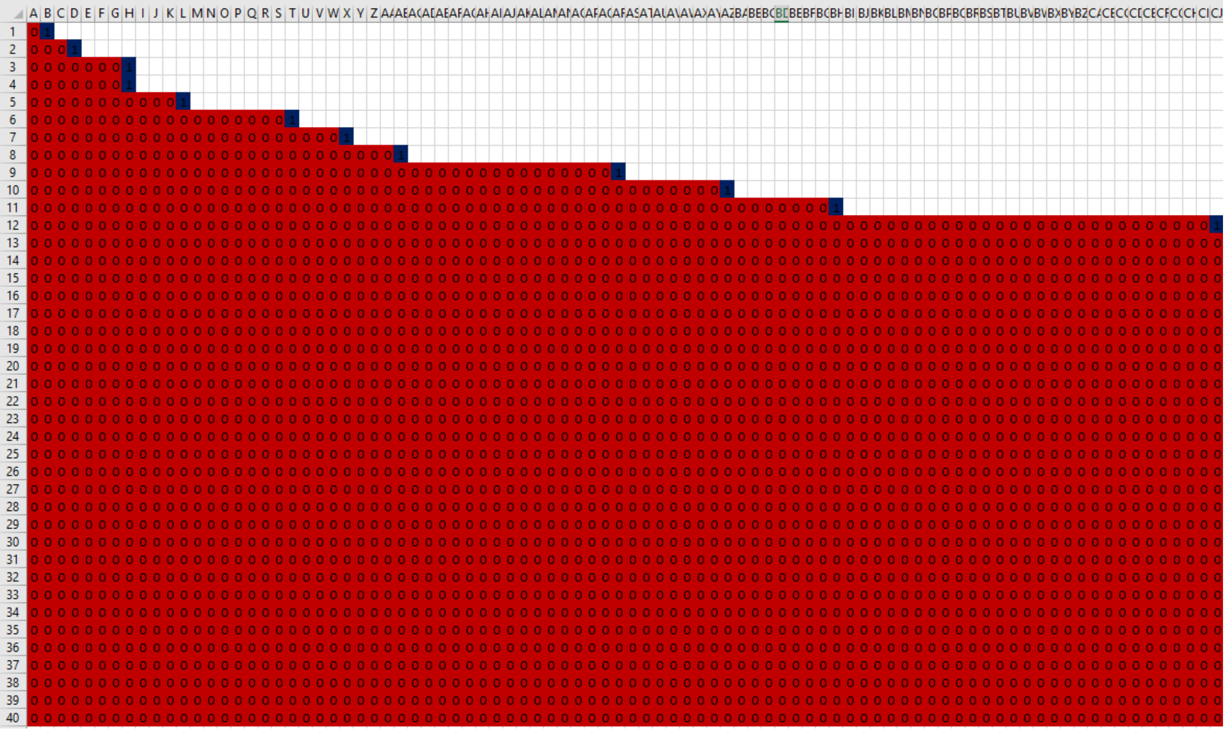
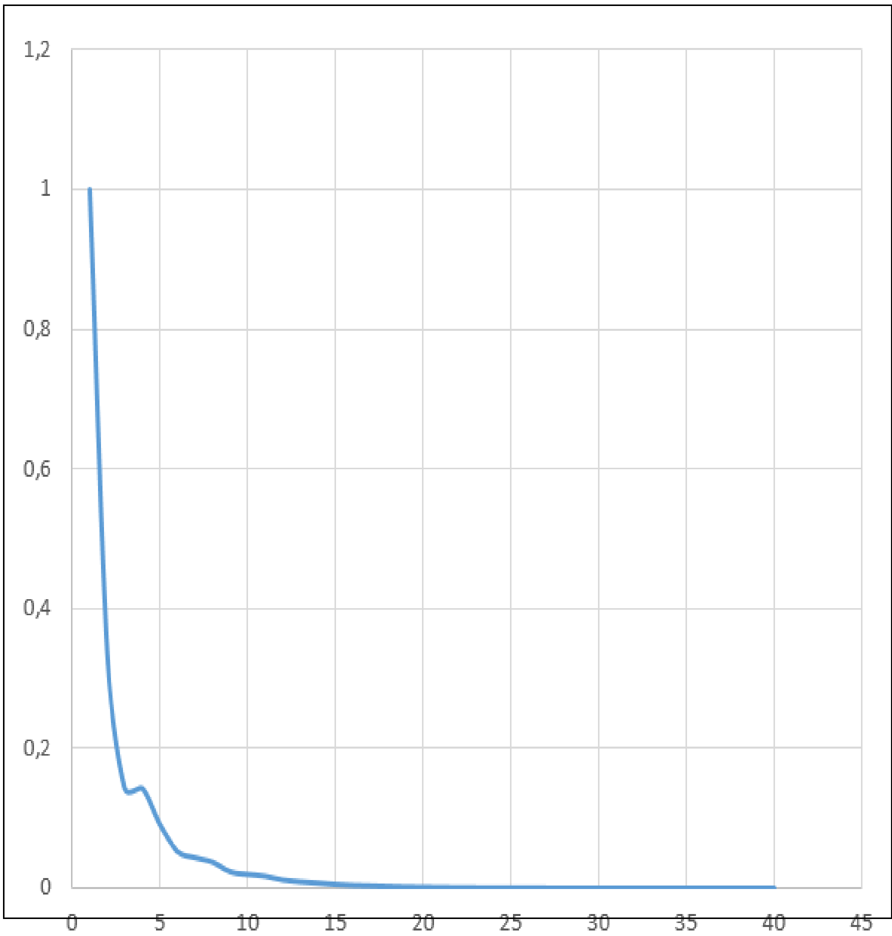


Figure 5: DIFF of 23 and 240777 look-and-say sequences



1.0
0.3333333333333333
0.14285714285714285
0.14285714285714285
0.09090909090909091
0.05263157894736842
0.043478260869565216
0.037037037037037035
0.023255813953488372
0.0196078431372549
0.01694915254237288
0.011494252873563218
0.009009009009009009
0.007194244604316547
0.005128205128205128
0.003861003861003861
0.0030959752321981426
0.002320185614849188
0.0017152658662092624
0.0013986013986013986
0.0010649627263045794
0.0007968127490039841
0.0006317119393556538
0.0004764173415912339
0.0003614022406938923
0.0002797202797202797
0.00021427040925648167
0.00016278691193228064
0.00012583364791745311
9.687106461300009e-05
7.370826269624825e-05
5.6911957202208187e-05
4.3784754148605455e-05
3.33522329319948e-05
2.5662740267405754e-05
1.9694731659281143e-05
1.5078635081952381e-05
1.1557620517087941e-05
8.888651858172672e-06
6.812778046504023e-06

Figure 6. Overlapping parts of 69 and 208: look-and-say sequences with partly different heads

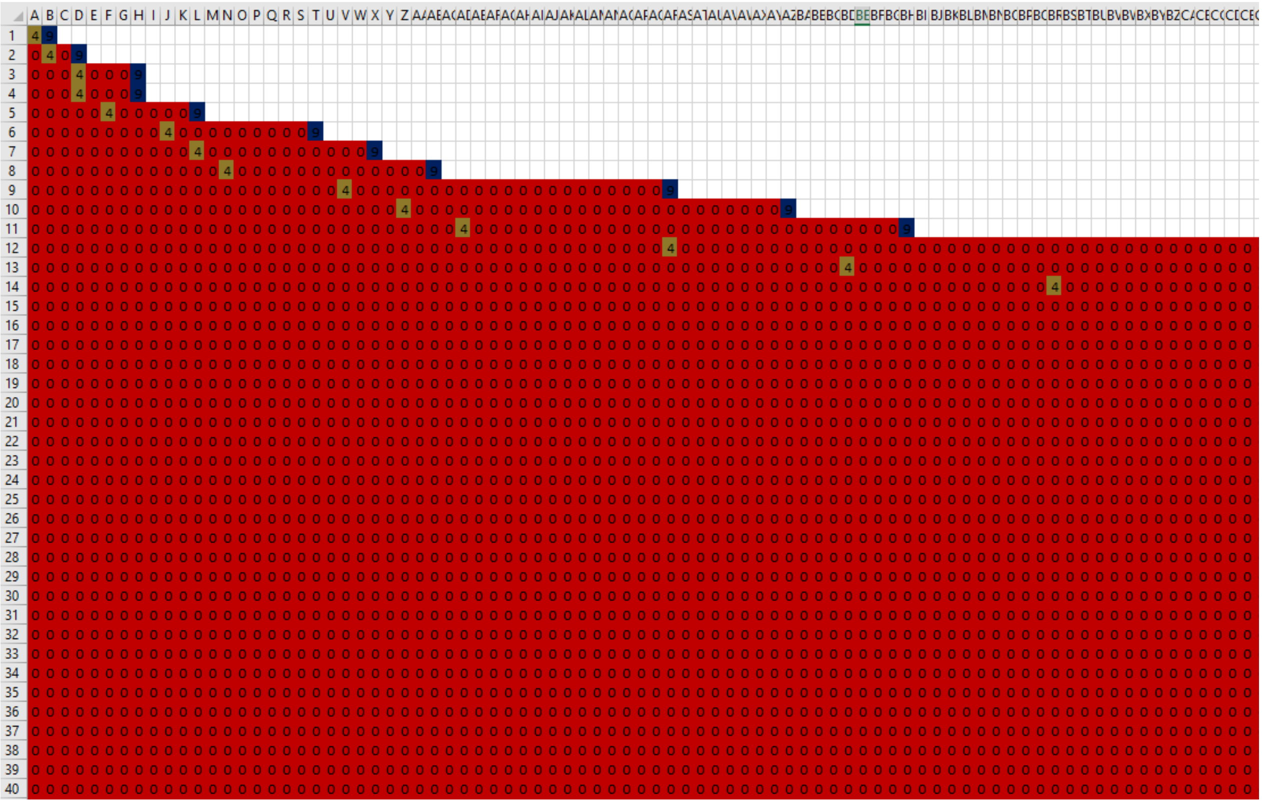
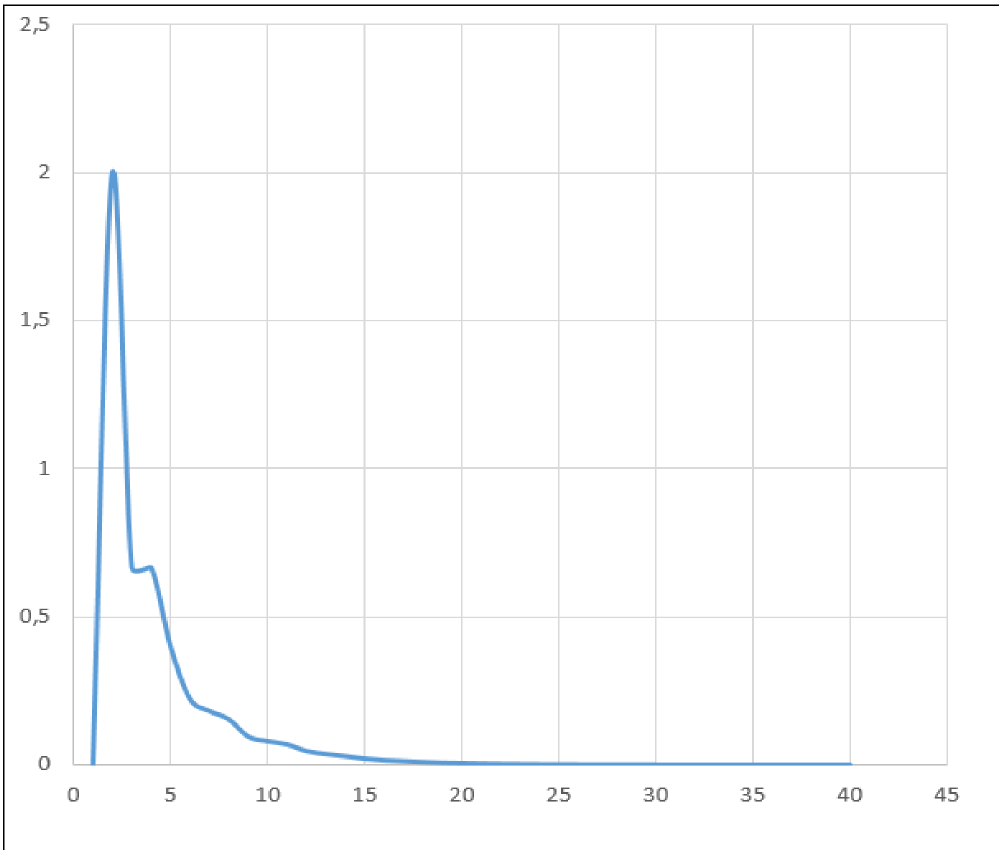


Figure 7. DIFF of 69 and 208: look-and-say sequences



0.0
2.0
0.6666666666666666
0.6666666666666666
0.4
0.2222222222222222
0.18181818181818182
0.15384615384615385
0.09523809523809523
0.08
0.06896551724137931
0.046511627906976744
0.03636363636363636
0.028985507246376812
0.020618556701030927
0.015503875968992248
0.012422360248447204
0.009302325581395349
0.006872852233676976
0.0056022408963585435
0.0042643923240938165
0.003189792663476874
0.0025284450063211127
0.0019065776930409914
0.0014461315979754157
0.001119194180190263
0.0008572653236176596
0.0006512536633018561
0.0005033979360684622
0.00038752179810114316
0.0002948547840188707
0.00022766078542970974
0.00017514668534897976
0.00013341338136215063
0.00010265359544218037
7.878047819750266e-05
6.03154498024669e-05
4.623101638889531e-05
3.555492346802723e-05
2.7251297843059776e-05

Figure 8. Comparing look-and say sequences beginning with 79 and 110:

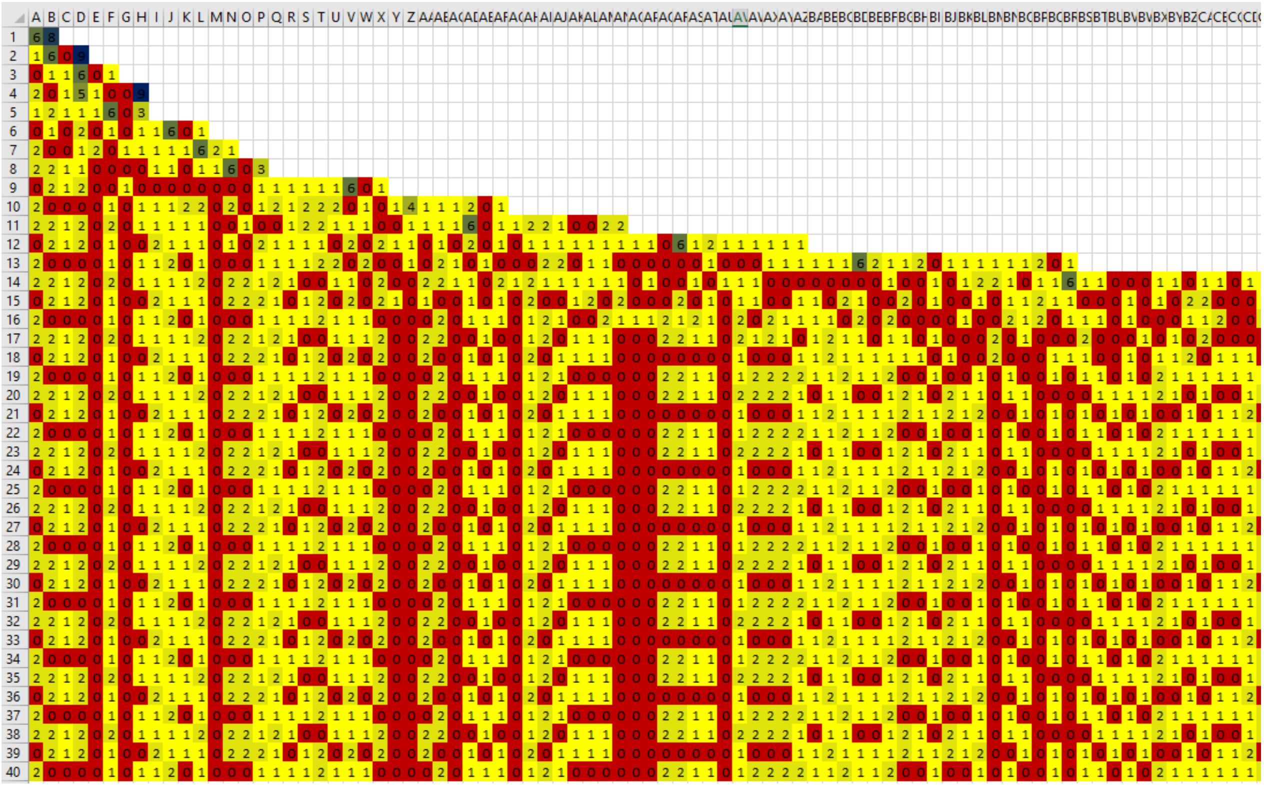


Figure 9. DIFF of look-and say sequences beginning with 79 and 110: