A FURTHER ANALYSIS OF BENFORD'S LAW

W. A. SENTANCE
The City University, London, England

In a recent paper [1] J. Wlodarski noted the interesting fact that Benford's "Law of anomalous numbers" was obeyed very closely by the first 100 Fibonacci numbers and the first 100 Lucas numbers. The same paper ended with the suggestion, taken up by the present author, that many more than the first 100 Fibonacci and Lucas numbers should be used for the purpose of analyzing Benford's Law more closely.

In a list of random numbers, one would normally expect to find that the distribution of the initial digit would have an approximately equal spread over the nine integers 1 to 9. However, it is an observed fact that in many tabulations the digit 1 occurs almost three times more often than any of the other eight digits. It was this that led Frank Benford in 1938 to enunciate his "law of anomalous numbers" that the probability of a random decimal number beginning with digit p is $\log(p + 1) - \log(p)$ where the logarithms are expressed to the base 10. [2]

Using a computer, it has been possible to extend the study to cover the first 1000 Fibonacci and the first 1000 Lucas numbers. Such a study would be perhaps unfeasible and certainly very tedious without the aid of a computer since F_{100} has 209 digits. Normally, numbers are held within the computer to an accuracy of so many digits, usually within the range of 10 to 20, and any arithmetic performed on such numbers will only be correct to this accuracy. However, by assigning one computer word for each digit of any particular number, we are able to store exactly large integer numbers. It is a relatively easy matter to simulate the operation of addition between any two such numbers. Addition is the only operation we need since the two sequences in which we are interested are defined by the additive recurrence formula

$$A_{n+1} = A_n + A_{n-1}$$

different initial conditions giving rise to the Fibonacci and Lucas sequences. To give some idea of the time involved, the additions which were needed to produce F_{1000}, took approximately 18 seconds. Such a method has other distinct advantages besides its great speed and ease as is shown later in this paper.

Table 1

Digit	1	2	3	4	5	6	7	8	9
N_F	30	18	13	9	8	6	5	7	4
N_L	31	16	14	10	8	5	8	4	4
N_B	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6

$$G_F = 0.657 \times 10^{-4}$$

$$G_L = 1.673 \times 10^{-4}$$

N_F : Number of times the digit occured as initial digit in the Fibonacci sequence.

N_L : Same as N_F but for the Lucas sequence.

N_B : Expected value, given by Benford's Law, of the digit to be the initial digit.

Table 1 reproduces the figures from [1] for the distribution of the initial digits of the first 100 Fibonacci numbers and the first 100 Lucas numbers, together with the expected value given by Benford's Law. In order to effect a comparison with later results, we have calculated "goodness of fit" constants G_F and G_L where

$$G_F = \sum_{i=1}^{9} \left(\frac{N_F}{100} - \frac{N_B}{100} \right)^2 \Big/ 9$$

$$G_L = \sum_{i=1}^{9} \left(\frac{N_L}{100} - \frac{N_B}{100} \right)^2 \Big/ 9 \quad .$$

Table 2 is exactly the same as Table 1 except that it gives the results pertaining to the first 1000 Fibonacci numbers and the first 1000 Lucas numbers, again with "goodness of fit" constants. It is readily seen that the behaviour exhibited by the small set of numbers has been propagated by the large set of numbers. The goodness of fit constant is in both cases considerably reduced indicating that the distribution of initial digits is approximating more closely to that predicted by Benford's Law as more numbers in the respective sequence are taken into account.

Table 2

Digit	1	2	3	4	5	6	7	8	9
N_F	301	177	125	96	80	67	56	53	45
N_L	301	174	127	97	79	66	59	51	46
N_B	301.0	176.1	124.9	96.9	79.1	66.9	58.0	51.2	45.8

Note: N_B correct only to 1D. Accurate values used in calculating G_F and G_L

$$G_F = \sum_{i=1}^{9} \left(\frac{N_F}{1000} - \frac{N_B}{1000} \right)^2 \Big/ 9 = 0.0114 \times 10^{-4}$$

$$G_L = \sum_{i=1}^{9} \left(\frac{N_L}{1000} - \frac{N_B}{1000} \right)^2 \Big/ 9 = 0.0118 \times 10^{-4}$$

The point could be made at this stage that the reduction in the values of G_F and G_L is purely fortuitous and that the author was fortunate in finding that G_F and G_L for the first 1000 numbers of each sequence were considerably smaller than for the first 100 numbers. To counteract this argument we give in Table 3, the values of G_F and G_L for the first i of the Fibonacci numbers and for the first i of the Lucas numbers where i takes the values 100 to 1000 in steps of 100. Although there are fluctuations in these values they do exhibit in general a downward trend.

Table 3

i	$G_F \times 10^4$	$G_L \times 10^4$
100	0.656	1.673
200	0.260	.261
300	0.139	.104
400	0.037	.031
500	0.026	.035
600	0.025	.013
700	0.036	.028
800	0.021	.007
900	0.012	.008
1000	0.011	.012

Again one may try to explain this strange distribution by the hypothesis that for these two sequences of numbers, the frequency of occurrence of each of the digits 1 to 9 throughout the numbers follows this pattern. However, Table 4 shows this not to be the case.

Table 4

	0	1	2	3	4	5	6	7	8	9
F	10474	10696	10495	10476	10431	10516	10433	10576	10350	10369
L	10393	10690	10783	10519	10699	10278	10507	10524	10285	10420

For the Fibonacci sequence the total number of digits in the first 1000 numbers is 104818. Assuming that each digit is distributed randomly then we expect each digit to occur with the same frequency. In this case the expectation for each digit is 10481.8. It is seen that the actual occurrence for each digit is very close to this expected value. Similar remarks apply to the Lucas sequence, too. The digit 1 therefore does not have an overall distribution different to any of the other digits.

This paper ends with a proposal to extend Benford's Law so that it now reads:

"The probability that a random number expressed in the number base b begins with digit p is log (p + 1) - log p, where the logarithms are to the base b."

Benford's Law is a particular case of this with b equal to 10. The idea behind such a proposal is that if it is true then it means that the distribution of initial digits seems to be some function inherent within the number system itself.

The method we have used to implement the addition of large integers is capable of being adapted to give results expressed with respect to any number base. Table 5 reproduces the

Table 5

Digit \ Base		4	5	6	7	8	9
1	N_F	501	430	389	356	336	314
	N_L	502	430	385	355	336	318
	N_B	500	430.1	386.9	356.2	333.3	315.5
2		291	253	227	211	193	187
		292	251	226	207	193	181
		292.5	251.9	226.3	208.4	195.0	184.5
3		208	178	160	146	140	132
		206	180	162	151	139	134
		207.5	178.7	160.6	147.8	138.3	130.9
4			139	123	114	105	99
			139	125	113	106	108
			138.6	124.5	114.7	107.3	101.6
5				101	93	90	83
				102	94	89	82
				101.8	93.7	87.7	83.0
6					80	73	69
					80	73	70
					79.2	74.1	70.2
7						63	62
						64	60
						64.2	60.8
8							54
							52
							53.6
G of Fit $\times 10^4$	F	.0114	.0057	.0157	.0198	.0390	.0220
	L	.0218	.0075	.0116	.0280	.0233	.0432

computer results for the first 1000 Fibonacci and Lucas numbers using bases 4 to 9 inclusive together with the theoretical expectation based on the extension to Benford's Law. Again we include a goodness-of-fit constant.

It can be seen that the distribution of initial digits in the other number bases closely resembles that predicted by this extension of Benford's Law.

In conclusion then, as far as the sequences of Fibonacci and Lucas numbers are concerned, the frequency of occurrence of the digits 1-9 as initial digits is an excellent illustration of Benford's Law. The distribution would seem to approach that given by Benford as more and more numbers are taken into account. If we choose to express them in any other base, then there is a very strong indication that the initial digits occur in a distribution given by the extension to Benford's Law proposed earlier in this paper.

REFERENCES

1. J. Wlodarski, "Fibonacci and Lucas Numbers Tend to Obey Benford's Law," Fibonacci Quarterly, Vol. 9, No. 1 (1971), pp. 87-88.
2. F. Benford, Proceedings of the American Philosophical Society, Vol. 78, No. 4, pp. 551-572 (March 31, 1938).

[Continued from page 489.]

REFERENCES

1. A. R. C. Bose and S. S. Shrikande, "On the Falsity of Euler's Conjecture About the Non-Existence of Two Orthogonal Latin Squares of Order 4t + 2," Proc. Nat'l. Acad. Sci., 45 (5), May 1959, pp. 734-737.

   B. E. T. Parker, "Orthogonal Latin Squares," Proc. Nat'l. Acad. Sci., 45 (6), June, 1959, pp. 859-862.
2. J. Arkin and V. E. Hoggatt, Jr., "The Arkin-Hoggatt Game and the Solution of a Classical Problem," Jour. Recreational Math., Vol. 6, No. 2 (Spring, 1973), pp. 120-122.
3. J. Arkin, "The First Solution of the Classical Eulerian Magic Cube Problem of Order Ten," Fibonacci Quarterly, Vol. 11, No. 2 (April, 1973), pp. 174-178.