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>
<thead>
<tr>
<th>Digit</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_F )</td>
<td>30</td>
<td>18</td>
<td>13</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>( N_L )</td>
<td>31</td>
<td>16</td>
<td>14</td>
<td>10</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>( N_B )</td>
<td>30.1</td>
<td>17.6</td>
<td>12.5</td>
<td>9.7</td>
<td>7.9</td>
<td>6.7</td>
<td>5.8</td>
<td>5.1</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Table 1
A FURTHER ANALYSIS OF BENFORD'S LAW

\[ G_F = 0.657 \times 10^{-4} \]
\[ G_L = 1.673 \times 10^{-4} \]

\( N_F \): Number of times the digit occurred 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

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

\[
G_L = \sum_{i=1}^{9} \left( \frac{N_L}{100} - \frac{N_B}{100} \right)^2 / 9
\]

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>
<thead>
<tr>
<th>Digit</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_F )</td>
<td>301</td>
<td>177</td>
<td>125</td>
<td>96</td>
<td>80</td>
<td>67</td>
<td>56</td>
<td>53</td>
<td>45</td>
</tr>
<tr>
<td>( N_L )</td>
<td>301</td>
<td>174</td>
<td>127</td>
<td>97</td>
<td>79</td>
<td>66</td>
<td>59</td>
<td>51</td>
<td>46</td>
</tr>
<tr>
<td>( N_B )</td>
<td>301.0</td>
<td>176.1</td>
<td>124.9</td>
<td>96.9</td>
<td>79.1</td>
<td>66.9</td>
<td>58.0</td>
<td>51.2</td>
<td>45.8</td>
</tr>
</tbody>
</table>

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

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

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

<table>
<thead>
<tr>
<th>$i$</th>
<th>$G_F \times 10^4$</th>
<th>$G_L \times 10^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.655</td>
<td>1.673</td>
</tr>
<tr>
<td>200</td>
<td>0.260</td>
<td>0.261</td>
</tr>
<tr>
<td>300</td>
<td>0.139</td>
<td>0.104</td>
</tr>
<tr>
<td>400</td>
<td>0.037</td>
<td>0.031</td>
</tr>
<tr>
<td>500</td>
<td>0.026</td>
<td>0.035</td>
</tr>
<tr>
<td>600</td>
<td>0.025</td>
<td>0.013</td>
</tr>
<tr>
<td>700</td>
<td>0.036</td>
<td>0.028</td>
</tr>
<tr>
<td>800</td>
<td>0.021</td>
<td>0.007</td>
</tr>
<tr>
<td>900</td>
<td>0.012</td>
<td>0.008</td>
</tr>
<tr>
<td>1000</td>
<td>0.011</td>
<td>0.012</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>10474</td>
<td>10696</td>
<td>10495</td>
<td>10476</td>
<td>10431</td>
<td>10516</td>
<td>10433</td>
<td>10576</td>
<td>10350</td>
<td>10369</td>
</tr>
<tr>
<td>L</td>
<td>10393</td>
<td>10690</td>
<td>10783</td>
<td>10519</td>
<td>10699</td>
<td>10278</td>
<td>10507</td>
<td>10624</td>
<td>10285</td>
<td>10420</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Base Digit</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_F )</td>
<td>501</td>
<td>430</td>
<td>389</td>
<td>356</td>
<td>336</td>
<td>314</td>
</tr>
<tr>
<td>( N_L )</td>
<td>502</td>
<td>430</td>
<td>385</td>
<td>355</td>
<td>336</td>
<td>315</td>
</tr>
<tr>
<td>( N_B )</td>
<td>500</td>
<td>430.1</td>
<td>386.9</td>
<td>356.2</td>
<td>333.3</td>
<td>315.5</td>
</tr>
<tr>
<td>2</td>
<td>291</td>
<td>253</td>
<td>227</td>
<td>211</td>
<td>193</td>
<td>187</td>
</tr>
<tr>
<td></td>
<td>292</td>
<td>251</td>
<td>226</td>
<td>207</td>
<td>193</td>
<td>181</td>
</tr>
<tr>
<td></td>
<td>292.5</td>
<td>251.9</td>
<td>226.3</td>
<td>208.4</td>
<td>195.0</td>
<td>184.5</td>
</tr>
<tr>
<td>3</td>
<td>208</td>
<td>178</td>
<td>160</td>
<td>146</td>
<td>140</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>206</td>
<td>180</td>
<td>162</td>
<td>151</td>
<td>139</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>207.5</td>
<td>178.7</td>
<td>160.6</td>
<td>147.8</td>
<td>138.3</td>
<td>130.9</td>
</tr>
<tr>
<td>4</td>
<td>139</td>
<td>123</td>
<td>114</td>
<td>105</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>139</td>
<td>125</td>
<td>113</td>
<td>106</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td></td>
<td>138.6</td>
<td>124.5</td>
<td>114.7</td>
<td>107.3</td>
<td>101.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>101</td>
<td>93</td>
<td>90</td>
<td>83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>102</td>
<td>94</td>
<td>89</td>
<td>82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>101.8</td>
<td>93.7</td>
<td>87.7</td>
<td>83.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>80</td>
<td>73</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>73</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>79.2</td>
<td>74.1</td>
<td>70.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>63</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>64.2</td>
<td>60.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>53.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

G of \( F \) \( \times 10^4 \) | .0114 | .0057 | .0157 | .0198 | .0390 | .0220 |
| Fit \( \times 10^4 \) L   | .0218 | .0075 | .0116 | .0280 | .0233 | .0432 |
A FURTHER ANALYSIS OF BENFORD'S LAW

Dec. 1973

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 con-
cerned, the frequency of occurrence of the digits 1-9 as initial digits is an excellent illustra-
tion 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
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-
45 (5), May 1959, pp. 734-737.
1959, pp. 859-862.
2. J. Arkin and V. E. Hoggatt, Jr., "The Arkin-Hoggatt Game and the Solution of a Classi-