SOME REMARKS ON INITIAL DIGITS

BASIL DAVIS
Indian Statistical Institute, Calcutta–35, India

It can be shown that the distribution of first digits among Fibonacci numbers is as follows: The probability that the first digit of a random Fibonacci number is n is given by

(1) $$P(n) = \log_{10}(1 + 1/n).$$

This property is true for any additive sequence of numbers, the m^{th} term of which is expressed as

(2) $$U_m = U_{m-1} + U_{m-2} \cdots U_{m-k}.$$

For Fibonacci and Lucas sequences $k = 2$. In the general case the ratio U_m/U_{m-1} tends to a limit, say R, as $m \to \infty$. R is related to k as follows:

(3) $$k = \frac{\log (2 - R)^{-1}}{\log R} .$$

Hence an additive sequence tends towards a geometrical progression. As the reader may verify, the initial digits of any geometrical progression of real numbers will obey distribution (1), provided that the ratio is not a rational power of 10 (i.e., $10^{p/q}$, where p and q are integers).

The validity of the above law is tested below for the first 100 Fibonacci and Lucas numbers. The number of incidences of n as initial digit for Fibonacci numbers is given under A and that for Lucas numbers under B. The percentage calculated on the basis of distribution (1) is given under C.

n	A F nos.	B L nos.	C $100 \log_{10}(1 + 1/n)$
1	30	31	30.1
2	18	16	17.6
3	13	14	12.5
4	9	10	9.7
5	8	8	7.9
6	6	5	6.7
7	5	8	5.8
8	7	4	5.1
9	4	4	4.6

The close adherence to the law (1) is evident and the deviations can be explained as due to the finite number of the terms considered.

The distribution of first digits was the topic of a paper published by Benford [1] in 1938. It had been observed that the first few pages of logarithm books were consistently dirtier than the last few, indicating that the users had more occasion to look up numbers with smaller initial digits than larger ones. Benford collected a lot of numbers of the kind that users of logarithms were likely to deal with. They included surface areas of rivers, molecular weights of chemical compounds and such numbers as are found in scientific and statistical tables. Ignoring the decimal point and the magnitude of the numbers, he found that the first digits of these apparently random numbers followed very closely the following distribution:

The probability that the first digit of a random entry is n is given by $P(n) = \log_{10}(1 + 1/n)$. This is known as Benford's Law, and the distribution is identical with (1). Some of the various explanations put forward to explain this Law may be found in the references.

An elementary "explanation" may be provided as follows: Before computers were put into large-scale use (that is when logarithm books had to be used) it was difficult to deal with large and cumbersome numbers. To overcome this difficulty it was necessary that the units in which different quantities were measured were adjusted so as to render the measurements small (though greater than 1). This can be illustrated in the case of measurements on length. In atomic measurements Angstroms and other microscopic units were used to render the very small measurements close to unity. In everyday life units ranging from millimeters and inches to kilometers and miles are still used. In astronomy Astronomical Units, light years and *parsecs* are among the units employed. Similarly for mass, time, area, etc., the units are varied to suit the scale. Hence the numbers found in scientific and statistical tables would tend to be small in magnitude, except when the numbers are less than 1.

Thus one would expect the probability of the occurrence of a number of magnitude x to decrease monotonically with x when $x \geqslant 1$. Considering the simplest distribution $1/x$ as a trial function, one may write for $x \geqslant 1$,

$$f(x)dx = \frac{kdx}{x}, \tag{4}$$

where $f(x)dx$ is the probability of occurrence of a number in the range x to $x + dx$ and k is a constant of proportionality. The form of $f(x)$ in the region $0 \leqslant x < 1$ can be shown to be immaterial in obtaining the result below, provided $f(x)$ is finite throughout that interval.

If a number has initial digit n, it should lie between n and $n + 1$ or $10n$ and $10(n + 1)$ or $100n$ and $100(n + 1)$, etc. Hence the probability of occurrence of a number with initial digit n is

$$P(n) = \int_{n}^{n+1} \frac{kdx}{x} + \int_{10n}^{10(n+1)} \frac{kdx}{x} + \int_{100n}^{100(n+1)} \frac{kdx}{x} + \cdots .$$

Let there be m such integrals (thus putting an upper limit on x which will be removed by letting $m \to \infty$).
Then

$$P(n) = mk[\ln(n+1) - \ln(n)] = mk \ln(1 + 1/n).$$

Let $m \to \infty$ and $k \to 0$ so that mk remains finite. Then normalizing the probability,

$$mk = \frac{1}{\ln 10}.$$

Therefore,

$$P(n) = \log_{10}(1 + 1/n).$$

Thus, Benford's Law is obtained.

REFERENCES

1. F. Benford, "The Law of Anomalous Numbers," *Proc. of the Am. Phil. Soc., 73* (4), Mar. 1938, pp. 551–572.
2. R. A. Raimi, "The Peculiar Distribution of First Digits," *Scientific Am., 221* (6), Dec. 1969, pp. 109–120.
3. R. A. Raimi, "On the Distribution of First Significant Figures," *Amer. Math. Monthly, 76* (4), April 1966. pp. 342–348.
4. B. J. Flehinger, "On the Probability That a Random Integer has Initial Digit *A*," *Amer. Math. Monthly, 73* (10), Dec. 1966, pp. 1056–1061.

****