AN EQUIVALENT FORM OF BENFORD'S LAW

JAMES V. PETERS

C. W. Post Center of L.I. University, Greenvale, NY 11548

Benford's law states that the probability of a positive integer having 1st digit d is given by

$$Pr(j = d) = \log_{10}(1 + 1/d). \tag{1}$$

In terms of the cumulative probability distribution, (1) is restated as

$$Pr(j < d) = \log_{10}d.$$

This result was first noted by Benford [1] in 1938 and has since been extended to counting bases other than 10 as well as to certain subsets, called Benford sequences, of the positive integers. Geometric progressions or, more generally, integer solutions of finite difference equations are examples of Benford sequences that have received considerable attention in the literature, e.g., [2]. This interest is due, in part, to the fact that the Fibonacci and Lucas numbers are obtained as solutions of the finite difference equation

$$x_{n+2} = x_{n+1} + x_n.$$

We refer the reader to [3] for an extensive bibliography concerning this and other aspects of the 1st-digit problem.

Since the consideration of varying counting bases will be of concern to us here, we introduce the following notation. We write $Pr(j < d)_b$ for the probability of $j < d$ when numbers are represented as digits in base $b \geq 2$. In this notation, Benford's law states that

$$Pr(j < d)_b = \log_b d, \text{ for } d \leq b. \tag{2}$$

The purpose of this paper is to establish that, for the set of positive integers, (2) is equivalent to the following "monotonicity statement":

<u>If $b \leq b'$, then $Pr(j < d)_b \geq Pr(j < d)_{b'}$.</u>

While this statement still makes sense for $b < d \leq b'$, we confine our attention to $d \leq b$. In so doing, it follows immediately that the monotonicity statement is implied by Benford's law as given in (2).

To reverse the above implication for the positive integers, we need two lemmas. Both of these results could be established via the functional equation

$$Pr(j < a) + Pr(j < c) = Pr(j < ac),$$

which is valid whenever the positive integers a and c as well as their product divide b. Instead of this approach, we present arguments based on a counting machine that randomly generates numbers in varying counting bases. The idea is as follows. It is clear that in binary ($b = 2$) the 1st digit must be 1. Consequently, if we represent numbers in oct 1 ($b = 8$) where each digit is denoted by a string of three binary symbols, then the 1st digit is determined by simply ascertaining the length of the binary representation modulo 3. Since the possible lengths (mod 3) of the binary representation of a randomly chosen number are equally likely, we obtain some probabilities. More generally, we have the following.

<u>*Lemma 1*</u>: Let $m, n \geq 0$, $a \geq 2$ denote integers. If randomly chosen positive integers are represented in base $b = a^n$, then

$$Pr(j < a^m) = m/n, \; m \leq n. \tag{3}$$

Proof: We denote by $b_1 b_2 \ldots b_k$ the random number as represented in base b. Thus, $0 \leq b_i < b$ for $i = 1, 2, \ldots, k$ and $b_1 \neq 0$. Rewrite each b_i as a_{1i} a_{2i} $\ldots$ a_{ni}, where the a_i's represent digits in base a. This yields a string of nk digits each of which is less than a. Removing the 0 digits occurring at the beginning of this, we obtain the base a representation of the random number. Suppose this base a representation contains x digits. We solve the congruence relation $x \equiv y \pmod{n}$ where $0 \leq y < n$. If $y = 0$, the 1st digit j (in base b) satisfies $a^{n-1} \leq j < a^n = b$. For any other value of y, the 1st digit satisfies $a^{y-1} \leq j < a^y$. Since each value of y is equally likely, we obtain

$$Pr(a^{y-1} \leq j < a^y) = Pr(a^{n-1} \leq j < a^n) = 1/n. \tag{4}$$

Equation (3) follows immediately from (4). This completes the proof.

By a simple variation of the combinatoric argument used in the proof of Lemma 1, we next obtain a result that permits the comparison of the distribution of the 1st digit with respect to two different bases.

Lemma 2: Using the notation introduced above, we have

$$Pr(j < d)_b = mPr(j < d)_{b^m}.$$

Proof: A random number represented by k digits in base b^m is rewritten as a string of km digits in base b. As in Lemma 1, we delete all consecutive zeros from the left-hand side of the km digits. This yields a base b representation of the number. For $j < d$, in base b, there are m equally likely possible values for the position of j in the base b^m representation. Since the position of j is independent of its value, we conclude that the probability of $j < d$ in base b^m is $1/m$ times the corresponding probability in base b. This is equivalent to the statement of Lemma 2 and completes the proof.

To deduce Benford's law from the lemmas, we proceed as follows. According to Lemma 2,

$$Pr(j < d)_b = mPr(j < d)_{b^m}. \tag{5}$$

The monotonicity statement and Lemma 1 yield the inequality

$$\frac{1}{n} = Pr(j < d)_{d^n} \geq Pr(j < d)_{b^m} \geq Pr(j < d)_{d^{n+1}} = \frac{1}{n+1} \tag{6}$$

whenever

$$d^n \leq b^m \leq d^{n+1}. \tag{7}$$

By the euclidean algorithm, (7) is always satisfied by some $n \geq 0$ for any given values of $b > d > 1$ and $m \geq 0$. Combining (5) and (6), we obtain

$$\frac{m}{n} \geq Pr(j < d)_b \geq \frac{m}{n+1}.$$

Now let $m \to \infty$ and choose n so as to maintain the validity of (7). Taking logarithms in (7), this implies that

$$\frac{m}{n+1} \leq \log_b d \leq \frac{m}{n}.$$

To show that $m/n \to \log_b d$ as $m \to \infty$, we simply note that

$$\frac{m}{n} - \frac{m}{n+1} = \frac{1}{n}\left(\frac{m}{n+1}\right) \leq \frac{1}{n} \log_b d \to 0.$$

This establishes (2).

The proofs presented here rely heavily upon properties of the set of positive integers which are not shared by other Benford sequences. As such, it is worth commenting on the more general situation. By definition, any Benford sequence satisfies (2) and, as noted above, this implies the monotonicity statement. The lemmas are also valid although the proofs given above are not. To give a more interesting example, consider the geometric progression $\{a^k\}$ which constitutes a Benford sequence in base b if and only if $a \neq b^{p/q}$ (p, q integers). Setting $a = 3$ and $b' = 9$, we obtain a subset of the positive integers which is not a Benford sequence. Moreover, $Pr(j < 4)_9 = 1$ for the geometric progression $\{3^k\}$. Since $\{3^k\}$ is a Benford sequence in base $b = 8$, we may apply Lemma 1 with $a = 2$, $m = 2$, $n = 3$ to yield $Pr(j < 4)_8 = 2/3$. A comparison of the above probabilities for $b = 8$ and $b' = 9$ shows that the monotonicity statement is false for this example.

REFERENCES

1. F. Benford. "The Law of Anomalous Numbers." *Proc. Amer. Phil. Soc.* 78 (1938): 551-557.
2. J. L. Brown & R. L. Duncan. "Modulo One Uniform Distribution of the Sequence of Logarithms of Certain Recursive Sequences." *The Fibonacci Quarterly* 8 (1970):482-486.
3. R. A. Raimi. "The First Digit Problem." *Amer. Math. Monthly* 83 (1976):521-538.

*****

A NEW TYPE MAGIC LATIN 3-CUBE OF ORDER TEN

JOSEPH ARKIN

197 Old Nyack Turnpike, Spring Valley, NY 10977

K. SINGH

University of New Brunswick, Fredericton, N.B. E3B 5A3, Canada

A Latin 3-cube of order n is an $n \times n \times n$ cube (n rows, n columns, and n files) in which the numbers 0, 1, 2, ..., $n - 1$ are entered so that each number occurs exactly once in each row, column, and file. A magic Latin 3-cube of order n is an arrangement of n^3 integers in three orthogonal Latin 3-cubes, each of order n (where every ordered triple 000, 001, ..., $n-1$, $n-1$, $n-1$ occurs) such that the sum of the entries in every row, every column, and every file, in each of the four major diagonals (diameters) and in each of the n^2 broken major diagonals is the same; namely, $\frac{1}{2}n(n^3 + 1)$. We shall list the cubes in terms of n squares of order n that form its different levels from the top square 0 down through (inclusively) square 1, square 2, ..., square $n - 1$. We define a broken major diagonal as a path (route) which begins in square 0 and goes through the n different levels (square 0, square 1, ..., square $n - 1$) of the cube and passes through precisely one cell in each of the n squares in such a way that no two cells the broken major diagonal traverses are ever in the same file.

The sum of the entries in the n cells that make up a broken major diagonal equals $\frac{1}{2}n(n^3 + 1)$. A complete system consists of n^2 broken major diagonals, where each broken major diagonal emanates from a cell in square 0, and thus the n^2 broken major diagonals traverse each of the n^3 cells of the cube in n^2 distinct routes. The cube is initially constructed as a Latin 3-cube in which the numbers are expressed in the scale of n (0, 1, 2, ..., $n - 1$). However, after adding 1 throughout and converting the numbers to base 10, we have the n^3 numbers 1, 2, ..., n^3 where the sum of the entries in every row, every column, and every file in each of the four major diagonals, and in each of the n^2 broken major diagonals is the same; namely, $\frac{1}{2}n(n^3 + 1)$.

In this paper, for the first time in mathematics, we construct a magic Latin 3-cube of order ten. In this case, the sum of the numbers in every row,