$N!$ HAS THE FIRST DIGIT PROPERTY

SHARON KUNOFF

CW Post Campus, Long Island University, Greenvale, NY 11548

(Submitted March 1986)

Observation of extensive collections of numerical data shows that the distribution of first digits is not equally likely. Frank Benford, a General Electric Company physicist hypothesized in 1938 that for any extensive collection of real numbers expressed in decimal form $Pr(j = p) = \log_{10}(1 + 1/p)$ or, equivalently, $Pr(j < p) = \log_{10} p$, where j is the first significant digit and p is an integer $1 \leqslant p \leqslant 9$. Benford presented extensive data to back up his claim. Sequences that have this property are said to obey Benford's law or to have the first digit property.

One can certainly create data which does not obey Benford's law. However, many "natural" collections do behave in this manner. It has been shown that the geometric sequence ar^n is a Benford sequence as long as r is not a rational power of 10, as is any sequence which is asymptotically geometric (see, e.g., [7]). The Fibonacci numbers F_k are asymptotic to $(\sqrt{5}/5)[(1 + \sqrt{5})/2]^k$, so they have the first digit phenomenon.

R. A. Raimi [7] gives an extensive bibliography of work done in the field until 1976. More recently, others have considered the distribution of first digits in specific sequences of mathematical interest using both the natural density

$$n(S) = \lim_{m \to \infty} \frac{(\text{the number of elements in } S < m)}{m}$$

and other density functions (see, e.g., [1], [2], [6]). In this paper, I show that $N!$ obeys Benford's law using the natural density.

Let D_p be the set of all members of R^+ written with standard expansion in terms of some positive integer base b whose most significant digit is an integer $\leqslant p$. Then,

$$D_p = \bigcup_{n=-\infty}^{\infty} [b^n, (p + 1)b^n).$$

This set maps into $E_p = [0, \log_b(p + 1))$ if we take $\log_b D_p \pmod 1$. Using the notation of [4] let (x_n), $n = 1, 2, \ldots$, be a sequence of positive integers in R^+ written in base b and let $((\log_b x_n))$ be the sequence of fractional parts of $(\log_b x_n)$. Note that $b^{(\log_b x_n)}$ has the same first digit as x_n. Let

$$A[S; N; (x_n)]$$

be the number of terms of (x_n), $1 \leqslant n \leqslant N$, for which $x_n \in S$. Then

$$A[D_p; N; (x_n)] = A[E_p; N; ((\log_b x_n))].$$

A sequence (x_n) is said to be uniformly distributed modulo 1 (written u.d. mod 1) if, for every pair of real numbers with $0 \leqslant a \leqslant b \leqslant 1$, we have

$$\lim_{N \to \infty} \frac{A[[a, b); N; ((x_n))]}{N} = b - a.$$

Recall that E_p is simply $[0, (p+1))$ so that if $((\log_b x_n))$ is u.d. mod 1, then (x_n) is Benford under the natural density. Hence, the problem is reduced to considering the sequence $((\log_b x_n))$ for any sequence (x_n), where b is the base in which the sequence is expanded.

For convenience I will consider sequences written in decimal form and will write $\log x$ for $\log_{10} x$.

Theorem: Let $F = \{N! \mid N = 1, 2, 3, \ldots\}$ and let

$$F_k = \{n \mid n \in F \text{ and the first digit of } n \text{ is } K\}.$$

Then $N!$ is Benford; that is,

$$\lim_{m \to \infty} \frac{(\text{the number of elements in } F_k < m)}{(\text{the number of elements in } F < m)} = \log \frac{k+1}{k}.$$

This can be proven utilizing the following theorems from [4]:

(a) If the sequence (x_n), $n = 1, 2, \ldots$, is u.d. mod 1, and if (y_n) is a sequence with the property

$$\lim_{n \to \infty}(x_n - y_n) = a,$$

a real constant, then (y_n) is u.d. mod 1.

(b) The Weyl Criterion: A sequence (x_n), $n = 1, 2, \ldots$, is u.d. mod 1 if and only if

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} = 0 \text{ for all integers } h \neq 0.$$

(c) Let a and b be integers with $a < b$, and let f be twice differentiable on $[a, b)$ with $f'' \geqslant p > 0$ or $f'' \leqslant -p$ for $x \in [a, b)$. Then,

$$\left| \sum_{n=a}^{b} e^{2\pi i f(n)} \right| \leqslant (|f'(b) - f'(a)| + 2)\left(\frac{4}{\sqrt{p}} + 3\right).$$

We observe that

$$\lim_{n \to \infty} \left| \log[(n/e)^n \sqrt{2\pi n}] - \log n! \right| = 0$$

since

$$n! = \sqrt{2\pi n}(n/e)^n e^{r(n)/12n} \text{ with } 1 - 1/(12n+1) < r(n) < 1,$$

so that

$$\lim_{n \to \infty} \log[(\sqrt{2\pi n}(n/e)^n]/n!)] = 0.$$

Thus, if $\sqrt{2\pi n}(n/e)^n$ is Benford, so is $n!$. This is convenient for a statistical analysis because it is much simpler and faster to obtain the first digit of $\sqrt{2\pi n}(n/e)^n$ than that of $n!$ despite the fact that, today, programs are available to compute $n!$ for very large n (see, e.g., [3]). Moreover, using (b) and (c), we can show that $\log(\sqrt{2\pi n}(n/e)^n)$ is u.d. mod 1 so that $(\log n!)$ is also, which means $n!$ is Benford. Define $f(x) = h(\log[\sqrt{2\pi x}(x/e)^x])$. Then

$$f''(x) = h(\log_e 10)^{-1}[(2x-1)/x^2] > h(N \log_e 10)^{-1} \geqslant h/3N \text{ for } 1 \leqslant x \leqslant N.$$

Substituting into Theorem (c) with $p = h/3N$ yields:

$$\left| \sum_{n=1}^{N} e^{2\pi i f(n)} \right| \leqslant \left(\left| \frac{h(1-N)}{6N} + \log N \right| + 2 \right) \left(4\sqrt{\frac{3N}{h}} + 3 \right).$$

Thus,

$$\lim_{N \to \infty} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i f(n)} \right| = 0$$

and $f(n)$ is u.d. mod 1, which implies $\sqrt{2\pi n}(n/e)^n$ is Benford and, therefore, as indicated previously, so is n!.

Another interesting sequence to consider is a^{p_k} where p_k is the k^{th} prime. It has been shown that the primes themselves do have the first digit phenomenon under some non-standard densities (see, e.g., [1]). In a chi-squared analysis at the 95% level for 8 degrees of freedom we would reject the Benford hypothesis if chi-squared is greater than 15.5. Tallying the first digit of the sequence 2^{p_k} for the first 65 primes gives a value of chi-squared of 9.8, while in an analysis of a random sequence of 56 primes less than 10000 a chi-squared value of 12.74 was obtained. Using a Kolmogorov-Smirnov analysis at the 95% level, in the first case, $K = .072$ compared to the table value of .16, and for the random sample, a K value of .14 was obtained, compared to .18 (for table values see, e.g., [5]). These results seem to indicate that 2^{p_k} or, more generally, a^{p_k} may be Benford under other than the natural density. However, this remains an open question.

ACKNOWLEDGMENT

This research was partially supported by a CW Post Campus LIU faculty research grant.

REFERENCES

1. D. Cohen & T. Katz. "Prime Numbers and the First Digit Phenomenon." *J. of Number Theory 18* (1984):261-268.
2. T. Katz & D. Cohen. "The First Digit Property for Exponential Sequences Is Independent of the Underlying Distribution." *The Fibonacci Quarterly 24*, no. 1 (1986):2-7.
3. G. Kiernan. "Computing Large Factorials." *The College Math Journal 16*, no. 5 (1985):403-412.
4. L. Kuipers & H. Niederreiter. *Uniform Distribution of Sequences.* New York: John Wiley & Sons, 1974.
5. B. Lingren. *Statistical Theory.* New York: Macmillan, 1976.
6. J. Peters. "An Equivalent Form of Benford's Law." *The Fibonacci Quarterly 19*, no. 1 (1981):74-76.
7. R.A. Raimi. "The First Digit Problem." *The Amer. Math. Monthly 83*, no. 7 (1976):521-538.
8. B. P. Sarkar. "An Observation on the Significant Digits of the Binomial Coefficients and Factorials." *Sanhka* Ser. B, *35* (1973):269-271.

◆◇◆◇◆