ON THE DISTRIBUTION OF FIRST SIGNIFICANT DIGITS[1]

By Roger S. Pinkham

Rutgers—The State University

Introduction. It has been noticed by astute observers that well used tables of logarithms are invariably dirtier at the front than at the back. Upon reflection one is led to inquire whether there are more physical constants with low order first significant digits than high. Actual counts by Benford [2] show that not only is this the case but that it seems to be an empirical truth that whenever one has a large body of physical data, Farmer's Almanac, Census Reports, Chemical Rubber Handbook, etc., the proportion of these data with first significant digit n or less is approximately $\log_{10}(n + 1)$. Any reader formerly unaware of this "peculiarity" will find an actual sampling experiment wondrously tantalizing. Thus, for example, approximately 0.7 of the physical constants in the Chemical Rubber Handbook begin with 4 or less ($\log_{10}(4 + 1) = 0.699$). This is to be contrasted with the widespread intuitive evaluation $\frac{4}{9}$ths.

At least two books call attention to this peculiarity, Furlan [6] and Wallis [18], but to my knowledge there are only five published papers on the subject, Benford [2], Furry et al [7], [9], Gini [8], and Herzel [11]. The first consists of excellent empirical verifications and a discussion of the implied distribution of 2nd, 3rd, $\cdots$ significant digits. The second and third put forth the thesis that the distribution of significant digits should not depend markedly on the underlying distribution, and the authors present numerical evaluations for a range of underlying distributions in support of their contention. The fourth maintains that explanation is to be sought in empiric considerations. The fifth considers three different urn models; each yields a distribution of initial digits which the author compares with $\log_{10}(n + 1)$.

This paper is a theoretical discussion of why and to what extent this so called "abnormal law" must hold. The flavor of the results is, I think, conveyed in the following remarks.

(i) The only distribution for first significant digits which is invariant under scale change of the underlying distribution is $\log_{10}(n + 1)$. Contrary to suspicion this is a non-trivial mathematical result, for the variable n is discrete.

(ii) Suppose one has a horizontal circular disc of unit circumference which is pivoted at the center. Let the disc be given a random angular displacement θ where $-\infty < \theta < \infty$. If the final position of the disc mod one is called φ, i.e.,

$$\varphi \equiv \theta \bmod(1), \qquad 0 \leqq \varphi < 1,$$

Received June 23, 1960; revised June 12, 1961.

[1] This work was supported by the Office of Naval Research under Contract Nonr 404 (16). Reproduction in whole or in part is permitted for any purpose of the United States Government.

then φ is a random variable whose probability structure is determined entirely by that of θ. In fact if

$$\Pr(x \leqq \theta < x + dx) = g(x)\,dx,$$

and

$$\Pr(y \leqq \varphi < y + dy) = f(y)\,dy,$$

then

$$f(y) = \sum_{m=-\infty}^{\infty} g(y + m).$$

Now it is intuitively obvious that for a wide range of possible distributions of θ the distribution of φ should be approximately uniform i.e.,

$$f(y) \approx 1, \qquad 0 \leqq y \leqq 1.$$

This and related properties of distributions wrapped around a circle have been known for some time, Dvoretsky [4], Lévy [14], Robbins [15], and put to various uses, Aitchison [1], Brown [3], Horton and Smith [12], Tocher [17].

The logarithmic law of left-most significant digits is a consequence of the above property of random variables mod one. One can see this as follows. Let $F(x)$ be the cumulative distribution function for the population of physical constants (taken non-negative for convenience). Define $D(x)$ by

$$D(x) = \sum_{m=-\infty}^{\infty} [F(x10^m) - F(10^m)], \qquad x > 0.$$

$D(n)$ for $n = 2, 3, \cdots, 10$ gives the proportion of the population with first significant digit $n - 1$ or less. The logarithmic "law" states that $D(n)$ should be approximately $\log_{10}(n)$. Thus one suspects that

$$\log_{10}(x) \approx \sum_{m=-\infty}^{\infty} [F(x10^m) - F(10^m)].$$

A change of variables will make clear the connexion with the spinning disc. Let

$$y = \log_{10}(x) \text{ and } G(y) = F(10^y).$$

One then has

$$y \approx \sum_{m=-\infty}^{\infty} [G(y + m) - G(m)],$$

or, taking derivatives,

$$1 \approx \sum_{m=-\infty}^{\infty} g(y + m).$$

This latter approximate equality is the one mentioned before in connexion with random variables mod one.

Section 1 gives the mathematical support for contention (i) while Section 2 provides a mathematical basis for the approximation alluded to in (ii).

After the mathematical work of Section 2 had been completed I discovered the basic mathematical idea without the detail in a discussion by I. J. Good of a paper by Tocher [17].

1. An invariance principle. The population of known physical constants changes daily, but the collection of such constants can be regarded as a large sample from an unknown underlying distribution of all physical constants. It is this underlying distribution in which interest will center.

Such mental constructs are familiar in the natural sciences. Thus most physical objects are regarded as having a density even though they are "known" to have a granular structure at the atomic level. Such entities are of course outside the compass of mathematics per se.

Consider the population of all physical constants and the derived distribution of first significant digits. Suppose all the physical constants were multiplied by some fixed number. What would happen to the distribution of first significant digits? One feels, I think, that it would be the same as before. This invariance property is enough, as is shown below, to characterize the distribution completely. $\text{Log}_{10}(n + 1)$ emerges as the necessary cumulative. The basic mathematical fact is that a certain derived functional equation has one and only one solution.

Suppose $F(x)$ is the cumulative distribution function for the population of all physical constants (assumed non-negative) in accordance with their size. Then

$$D(x) = \sum_{m=-\infty}^{\infty} [F(x10^m) - F(10^m)], \qquad x > 0, \tag{1}$$

is a well defined function for positive x; $D(n)$ for $n = 2, \cdots, 9, 10$ gives the proportion with first significant digit $n - 1$ or less, since all numbers between 10^m and $n \times 10^m$ begin with $n - 1$ or less.

If all the physical constants are multiplied by a positive constant c, then the resulting cumulative is $F(x/c)$. The postulated invariance yields

$$D(n) = \sum_{m=-\infty}^{\infty} \left[F\left(\frac{n}{c} 10^m\right) - F\left(\frac{10^m}{c}\right) \right]$$

or

$$D(n) = D(n/c) - D(1/c), \qquad c > 0; n = 2, \cdots, 10. \tag{2}$$

If the relation (2) held for arbitrary positive real n rather than $n = 2, \cdots, 10$ one could, assuming continuity, immediately deduce $D(n) = \log_{10} n$. We now show this conclusion to be justified under even weaker conditions than are implicit in Equation (2).

THEOREM 1. *If*

1. $D(2) + D(x) = D(2x), \qquad x > 0;$
2. $D(10) + D(x) = D(10x), \qquad x > 0;$
3. $D(x)$ *is continuous;*
4. $D(10) = 1;$

then $D(x) = \log_{10}(x)$, $x > 0$.

PROOF. Let $H(x) = D(10^x)$. Then conditions 1 and 2 become

$$H(\log n) + H(y) = H(\log n + y), \qquad -\infty < y < \infty, n = 2, 10. \tag{3}$$

Thus $H(N \log n) = NH(\log n)$ if N integral, and one has $H(N) = N$ since $H(1) = 1$. From the theory of continued fractions one knows, Hardy and Wright [10],

$$\log 2 = (p_m/q_m) + o(1/q_m) \qquad (m \to \infty)$$

with p_m, q_m integers. I.e.,

$$q_m \log 2 = p_m + o(1) \qquad (m \to \infty).$$

Hence by hypothesis 2

$$q_m H(\log 2) = p_m + o(1) \qquad (m \to \infty).$$

Therefore $H(\log 2) = \log 2$. Suppose a irrational, and let $[x]$ denote the largest integer not exceeding x. Then it is well known, Kac([13], p. 41), that the sequence

$$a_n = na - [na], \qquad n = 1, 2, \cdots,$$

is uniformly distributed on [0, 1]. Thus there exists a subsequence $a_{n'}$ converging to any fixed $h(0 \leqq h < 1)$. Take a to be log 2.

$$H(a_{n'}) = n'H(\log 2) - [n'\log 2] = a_{n'}.$$

Letting n' tend to infinity yields, by the assumed continuity, $H(h) = h$. Since $y = [y] + y - [y]$, $H(y) = y$, $(-\infty < y < \infty)$, and $D(x) = \log_{10}x$.

It is reasonable to consider $F(x)$ continuous from which it follows that the $D(x)$ of (1) is continuous and thence by the Theorem 1 that $D(x) = \log_{10}(x)$.

2. An approximation. Drop from consideration any invariance postulate. Consider to what degree $\log_{10}x$ provides an approximation to

$$\sum_{m=-\infty}^{\infty} [F(x10^m) - F(10^m)], \qquad 1 \leqq x \leqq 10.$$

($F(\cdot)$ has the same significance as before.) Let $G(y) = F(10^y)$. Then one may as well consider how x approximates

$$J(x) = \sum_{m=-\infty}^{\infty} [G(x + m) - G(m)], \qquad 0 \leqq x \leqq 1. \tag{4}$$

It is reasonable to take some canonical representation of $G(x)$ and hope that the sum $J(x)$ can be evaluated explicitly. A statistician immediately thinks of Fourier transforms since characteristic functions always exist for distributions.

A trouble immediately appears. In trying to evaluate the sum

$$J(x) = \sum_{m=-\infty}^{\infty} [G(x + m) - G(m)],$$

one immediately stubs against sums of the form

$$\sum_{m=0}^{\infty} e^{imu},$$

which of course do not converge. To overcome this difficulty one may introduce a "convergence factor" and subsequently sneak it out again at the end.

Thus, define $J(x \mid t)$ by

$$J(x \mid t) = \sum_{m=-\infty}^{\infty} [G(x + m) - G(m)]t^{|m|}, \qquad 0 < t < 1, \tag{5}$$

and $W(u)$ by $W(u) = \int_{-\infty}^{\infty} e^{iut}\, dG(t)$. Then $W(u)$ exists for all $-\infty < u < \infty$, and

$$G(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1 - e^{-ixu}}{iu} W(u)\, du.$$

Suppose

$$W(u) = O(u^{-h}), \qquad h > 0, \qquad (|u| \to \infty)$$

Then, by merely summing geometric series after switching the order of summation and integration, one has

$$J(x \mid t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1 - e^{-ixu}}{iu} W(u)P(u, t)\, du;$$

$P(u, t)$ is the Poisson kernel given by

$$P(u, t) = \frac{1 - t^2}{1 + t^2 - 2t \cos u}. \tag{6}$$

The interchange of limits is justified by the assumed order condition.

The Poisson kernel when properly normalized is a frequency function uncommon in statistical circles. Thus

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} P(u, t)\, du = 1 \quad \text{and} \quad P(-u, t) = P(u, t).$$

Furthermore, the variance of the distribution goes to zero as t tends to one. Hence

$$\lim_{t \to 1} \frac{1}{2\pi} \int_{-\pi}^{\pi} P(u, t)Q(u)\, du = Q(0),$$

if Q is continuous at the origin.

Now to return to $J(x \mid t)$. Splitting the integral up into integrals over contiguous intervals of length 2π, and utilizing the periodicity of $P(u, t)$ yields

$$J(x \mid t) = \sum_{k=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1 - e^{-ix(u+2\pi k)}}{iu + i2\pi k} W(u + 2\pi k) P(u, t)\, du.$$

A second appeal to the assumed order conditions on $W(u)$ allows one to take limits as $t \to 1$ term by term. Thus, since W is continuous,

$$J(x \mid 1-) = x + \sum_{x \neq 0} \frac{1 - e^{-i2\pi kx}}{i2\pi k} W(2\pi k).$$

But by Abel's theorem, Titchmarch [16], $J(x) = J(x \mid 1 -)$. Hence finally

$$J(x) - x = \sum_{k \neq 0} \frac{1 - e^{-i2\pi kx}}{i2\pi k} W(2\pi k). \tag{7}$$

In the case $G(x)$ is symmetric about zero, viz. $G(x) = 1 - G(-x)$, one has

$$J(x) - x = \sum_{k=1}^{\infty} W(2\pi k) \frac{\sin (2\pi kx)}{\pi k}. \tag{8}$$

It is now clear that the quality of the approximation is in general high and does not depend on the fine structure of G and hence of F. For only W in the neighborhood of the origin is liable to inflate the sum appreciably and this depends primarily on the nature of G at infinity.

If, for example, $G(y)$ is Gaussian with mean zero and variance σ^2, then $W(u) = \exp(-\frac{1}{2}\sigma^2 u^2)$, and it is very clear that as σ increases and the tails lift the approximation improves markedly. This is in excellent accord with one's intuition.

An explicit bound on $J(x) - x$ may be obtained by noticing that $|1 - e^{-i2\pi kx}| \leqq 2$, and hence

$$|J(x) - x| \leqq \sum_{k \neq 0} \frac{1}{\pi k} |W(2\pi k)|.$$

If G has a density g, then

$$W(2\pi k) = \frac{1}{2\pi k i} \int_{-\infty}^{\infty} e^{i2\pi kx}\, dg.$$

Thus

$$|W(2\pi k)| \leqq \frac{1}{2\pi |k|} \int_{-\infty}^{\infty} |dg| = \frac{1}{2\pi |k|} V[g],$$

where $V[g]$ is the variation of g on $(-\infty, \infty)$. Whence

$$|J(x) - x| \leqq \frac{V[g]}{2\pi^2} \sum_{k \neq 0} \frac{1}{k^2} = \frac{1}{6} V[g].$$

We summarize in the following

THEOREM 2. *If*

1. $$W(u) = \int_{-\infty}^{\infty} e^{iux}\, dG(x);$$

2. $$W(u) = O(|u|^{-h}), \qquad h > 0, |u| \to \infty;$$

then

$$\sum_{m=-\infty}^{\infty} [G(x + m) - G(m)] = \sum_{k=-\infty}^{\infty} \frac{1 - e^{-i2\pi kx}}{i2\pi k} W(2\pi k).$$

COROLLARY. *If*

1. $$W(u) = \int_{-\infty}^{\infty} e^{iux} g(x)\, dx;$$

2. $$V[g] < \infty;$$

3. $$J(x) = \sum_{m=-\infty}^{\infty} [G(x + m) - G(m)];$$

then

$$|J(x) - x| \leqq \tfrac{1}{6} V[g].$$

3. Remarks. In inventory problems one is often concerned with non-negative random variables X_i, $i = 1, 2, \cdots$ which are independent, identically distributed, and possess a mean much smaller than some number K. One is interested in the first time $S_n = X_1 + \cdots + X_n$, $n = 1, 2, \cdots$, exceeds K. Let this time be a random variable T. If the time axis is split up into contiguous intervals (periods) of length P, much smaller than the mean and variance of T, then it is often assumed that the time during the period at which the first exceedance occurs has an approximately uniform distribution. Time is here being measured from the beginning of the period. This is intuitively very appealing. Suppose T has cumulative $G(t)$ and that the problem is scaled such that $P = 1$. Then intuition says $J(x) - x$ is "small", where

$$J(x) = \sum_{n=0}^{\infty} [G(x + n) - G(n)].$$

Previous results make it clear why this is in fact so.

A close connexion exists between $J(x) - x$ being small and Poincaré's observation on finely divided roulette wheels. Suppose the disc mentioned in the introduction is divided up into $2n$ contiguous intervals alternately of length ρ and β. Let $\rho/(\rho + \beta)$ and $\beta/(\rho + \beta)$ be independent of n. The segments of length ρ are called red, the others black. Fréchet [5] shows, for arbitrary distributions of θ, that the probability of obtaining red approaches $\rho/(\rho + \beta)$ as n tends to infinity and thus similarly for black. Here the quality of the approximation is improved by shrinking the fundamental unit relative to the variance of the underlying distribution rather than increasing the variance relative to the fundamental unit.

These considerations have an obvious import for the generation of pseudo-random numbers both by electronic computers and by special purpose machines.

The foregoing results bear on questions of round-off in computing machines. Since $d(uv) = u\,dv + v\,du$ the error resulting from multiplying two rounded numbers will be governed primarily by the first significant digits of the two numbers being multiplied. Now the distribution of first significant digits, favoring as it does low order digits, tends to produce less error than would be the case if first significant digits were uniform as has sometimes been assumed.

Acknowledgments. F. Mosteller and W. Kruskal provided the references to previously published material, and I am most grateful. R. Hamming was first to call my attention to the logarithmic law. He also was first to suggest that only the logarithm would satisfy the invariance principle. To one who invariably provides delightful ideas I am thankful indeed.

REFERENCES

[1] J. Atichison, "A statistical theory of remnants," *J. Roy. Stat. Soc.*, Vol. 21 (1959), pp. 158–168.

[2] Frank Benford, "The law of anomalous numbers," *Proc. Amer. Philos. Soc.*, Vol. 78, No. 4 (1938), pp. 551–572.

[3] G. W. Brown, "History of RAND's random digits," *Nat. Bur. Stds., App. Math. Series*, Vol. 12 (1951), pp. 31–32.

[4] A. Dvoretsky and J. Wolfowitz, "Sums of random integers reduced modulo m," *Duke Math. J.*, Vol. 18 (1951), pp. 501–507.

[5] Maurice Fréchet, *Calcul dés Probabilitiés*, Vol. 2, Gauthier-Villars, Paris, 1950.

[6] L. V. Furlan, *Das Harmoniegesetz der Statistik*, Basel, 1946.

[7] W. H. Furry and H. Hurwitz, "Distribution of numbers and distribution of significant figures," *Nature*, Vol. 155 (1945), pp. 52–53.

[8] Corrado Gini, "Sulla frequenza delle cifre iniziali dei numeri osservati," *Bull. Inst. Internat. Stat.*, 29th session, Vol. 35, 2è me Livraison, Rio De Janeiro (1957), pp. 57–76.

[9] S. A. Goudsmit and W. H. Furry, "Significant figures of numbers in statistical tables," *Nature*, Vol. 154 (1944), pp. 800–801.

[10] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford Press, Oxford, 1954.

[11] Amato Herzel, "Sulla distribuzione delle cifre iniziali dei numeri statistici," Atti della XV e XVI Riunione, *Societa Italiana di Statistica*, Rome, 1957.

[12] H. Burke Horton and R. Tynes Smith III, "A direct method for producing random digits in any number system," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 82–90.

[13] M. Kac, *Independence in Probability: Analysis and Theory of Numbers*, John Wiley and Sons, New York, 1959.

[14] P. Lévy, "L'addition des variables aléatoires definies sur une circonférence," *Bull. Soc. Math. France*, Vol. 67 (1939), pp. 1–41.

[15] Herbert Robbins, "On the equidistribution of sums of independent random variables," *Proc. Amer. Math. Soc.*, Vol. 4 (1953), pp. 786–799.

[16] E. C. Titchmarsh, *The Theory of Functions*, Oxford University Press, Oxford, 1932.

[17] K. D. Tocher, "The application of automatic computers to sampling experiments," *J. Roy. Stat. Soc.*, Ser. B, Vol. 16 (1954), pp. 39–61.

[18] W. Allen Wallis and Harry V. Roberts, *Statistics: A New Approach*, The Free Press, Glencoe, Illinois, 1957.