with area k. For $k = 1$, their solution leads to the triangle (5/3, 17/6, 3/2), which is much simpler than the one given above. Other solutions for $k = 1$, suggested by the referee, are (1/2, 13/3, 25/6) and (17/30, 4, 113/30).

We can introduce a second parameter in any solution

$$a = f(k), \ b = g(k), \ c = h(k)$$

by replacing k by km^2, where m is any positive rational, and then dividing by m. Thus we get

$$a' = \frac{1}{m} f(km^2), b' = \frac{1}{m} g(km^2), c' = \frac{1}{m} h(km^2).$$

For example, (26) above leads to

$$\frac{5k^2m^4 - 4km^2 + 4}{m(k^2m^4 - 4)}, \ \frac{km(k^2m^4 - 4km^2 + 20)}{2(k^2m^4 - 4)}, \ \frac{km^2 + 2}{2m}. \tag{27}$$

Of course, this device can be used only once.

DEPARTMENT OF MATHEMATICS, THE PENNSYLVANIA STATE UNIVERSITY, UNIVERSITY PARK, PA. 16802.

THE FIRST DIGIT PROBLEM

RALPH A. RAIMI

1. Introduction and notation. It has been known for a long time that if an extensive collection of numerical data expressed in decimal form is classified according to first significant digit, without regard to position of decimal point, the nine resulting classes are not usually of equal size. Indeed, while a truly random table should show a frequency of 1/9 for the occurrence of a given first digit p ($p = 1, 2, \dots, 9$), many observed tables give a frequency approximately equal to $\log_{10}(p+1)/p$. Thus the initial digit 1 appears about .301 of the time, 2 somewhat less and so on, with 9 occurring as a first digit less than 5 percent of the time. (We do not admit 0 as a possible first digit.)

This particular logarithmic distribution of first digits, while not universal, is so common and yet so surprising at first glance that it has given rise to a varied literature, among the authors of which are mathematicians, statisticians, economists, engineers, physicists and amateurs. The present memoir includes a bibliography as nearly complete as I could collect, deliberately omitting only those references to the problem which make no attempt to add to its understanding. My purpose is to review all the proposed explanations in some rational (but not chronological) order, making plain the hypotheses and results in each case but often suppressing details of proof.

The main bibliography is arranged and numbered chronologically; every simply numbered item is directly and avowedly concerned with the first digit phenomenon. The supplementary bibliography, numbered 1B, 2B, etc., does not refer explicitly to the problem.

A few notations will persist throughout: R stands for the real number system, R^+ is the non-negative part of R, N stands for the set of positive integers. Intervals in R are given as usual, e.g. $[a, b) = \{x \in R : a \leqq x < b\}$. D_p denotes the set of all members of R^+ whose standard decimal expansion begins with an integer $\leqq p$ ($p = 1, 2, \dots, 9$). Thus,

$$D_p = \bigcup_{n=-\infty}^{\infty} [10^n, (p+1)10^n).$$

In some places D_p will be spoken of as a subset of N; in such cases the context will make clear that

$D_p \cap N$ is meant. The mapping $\log_{10}$: $R^+\backslash\{0\} \to R$ will be denoted log, without subscript. E_p will denote $\log D_p$, thus,

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

In the suggestive language of probability theory, the first digit phenomenon is usually expressed: $\text{prob}\{x \in D_p\} = \log(p+1)$. This assertion, whatever it may mean, will be called *Benford's Law* because it has been thought by many writers to have originated with the General Electric Company physicist Frank Benford [2]. Certainly Benford popularized the problem, and he may well have been unaware that the polymathic Simon Newcomb, primarily an astronomer but also sometime editor of *The American Journal of Mathematics*, had also formulated the same law 57 years earlier [1]. There is ample precedent for naming laws and theorems for persons other than their discoverers, else half of analysis would be named after Euler. Besides, even Newcomb implied that the observation giving rise to the Benford law was an old one in his day. One would hate to change the name of the law now only to find later that another change was called for.

2. Empirical and other data. Newcomb [1] opened his paper as follows. "That the ten digits do not occur with equal frequency must be evident to anyone making use of logarithm tables, and noticing how much faster the first pages wear out than the last ones. The first significant figure is oftener 1 than any other digit, and the frequency diminishes up to 9." However, Newcomb gave no actual numerical data.

Benford [2] gives a great deal; he summarizes the counts for each of twenty different tables with lengths ranging from 91 entries (atomic weights) to 5000 entries from a mathematical handbook (n^{-1}, n^5, etc.). These two tables obeyed Benford's law rather badly, in fact, while others of his listings, such as the street addresses of the first 342 persons named in *American Men of Science, 1934*, did better. What came closest of all, however, was the union of all his tables.

$n =$		1	2	3	4	5	6	7	8	9
Benford's Law	$L(n)$	.301	.176	.125	.097	.079	.067	.058	.051	.046
Stigler's Law	$S(n)$	.241	.183	.145	.117	.095	.076	.060	.047	.034
Benford's Data	$B(n)$	.306	.185	.124	.094	.080	.064	.051	.049	.047
Powers of Two	$P(n)$	.30	.17	.13	.10	.07	.07	.06	.06	.05
Electricity	$E(n)$	.316	.167	.116	.087	.085	.064	.057	.050	.057
Vancouver Tel.	$V(n)$	.00	.27	.04	.08	.13	.05	.05	.10	.28
Populations	$PP(n)$	.190	.200	.185	.168	.098	.065	.043	.037	.013

TABLE 1

In Table 1 the first line $L(n) = \log(n+1) - \log n$ is Benford's law, and is compared with the data in the succeeding rows. $S(n)$ is also not empirical, but is the set of predicted frequencies given by George Stigler [5], based on hypotheses to be described later. $B(n)$ is the empirical frequency found by Benford [2] in his ensemble of 20,229 entries. $P(n)$ is the frequency of leading digit n among the first hundred powers of 2, i.e. $2^0, 2, 2^2, \ldots, 2^{99}$; notice that $P(n) = L(n)$ about as nearly as 100 numbers can manage.

The row $E(n)$ was mailed to me by the head of the Electricity Board of Honiara in the (Br.) Solomon Islands. Upon reading of Benford's law in [22], he wrote, he went to the customer records of his 1243 electricity users, whose consumption in October, 1969, ranged from 1 KWH to over 40,000 KWH, and counted first digits for that month's consumption.

The row $V(n)$ I took from two columns of the 1974 Vancouver (Canada) telephone book, about 210 telephone numbers in all. A glance through the rest of the book confirms that no Vancouver telephone numbers begin with the digit 1.

The last row $PP(n)$ is taken from a table in *The World Almanac* (N. Y. Times, New York, 1971), listing the populations of all 'Populated places with population at least 2500' in the United States, using census figures from 1960 and 1970. I used all the entries for five States.

Each listing in Table 1 illustrates some feature of the discussion to follow. Other empirical data are found in [**2**; **5**; **6**; **8**; **24**]. Papers containing numerical data based on mathematical considerations ($P(n)$ is such a set of data) include [**3**; **4**; **5**; **9**; **15**; **16**; **17**; **18**; **24**; **25**; **27**; **28**]. Any reader can collect data of his own, or invent it, as like or as unlike any row of Table 1 as he wishes. In what sense, then, is $L(n)$ a 'Law'?

3. A bit of philosophy. Despite the obvious occurrence of natural tabulations which don't obey Benford's law, many authors (including the present author) have offered explanations of $L(n)$ which are purely mathematical in nature, as though the number system itself, along with the decimal numeration system, dictated these proportions.

Goudsmit and Furry [**3**] write, "It is merely the result of our way of writing numbers," though the sequel by Furry and Hurwitz [**4**] explicitly recognizes otherwise, and Warren Weaver [**11**] says it "is a built-in characteristic of our number system."

This point of view springs from the idea that there is some natural way to calculate a "density" for the set D_p in R^+ or in N, and that this natural way yields $\log(p+1)$. Now it is true, as will be shown in Sections 4 and 5, that certain summability methods will indeed assign these 'correct' densities, but the quoted statements are nonetheless dead wrong. Certainly the compilations giving $V(n)$ and $PP(n)$ in Table 1 are written in "our way" but still violate what Goudsmit and Furry claim is the "mere ... result" of all this. Geometric sequences such as $\{2^n\}$, on the other hand, do have the property that randomly drawn finite samples tend to obey the Benford ratios ($P(n)$ in Fig. 1, for example).

This situation is exactly reversed if instead of looking for a "density" for D_p we look for a density of the even positive integers. Everyone will agree usually that half of N is made up of the evens; certainly all the most popular summability methods assign 1/2 as their density. But now if we look at the same numbers in the *World Almanac* that produced $PP(n)$, or the same columns of the Vancouver telephone directory that produced $V(n)$, we will find that in fact about half of each of them are even numbers; while samples drawn from the sequence $\{2^n\}$ most emphatically behave otherwise.

No purely mathematical argument can be expected, after all, to explain things actually found in the real world, like $B(n)$, without some correspondence between the hypotheses and structure of the mathematical argument on the one hand, and some observed facts and reliable laws of nature on the other. A strict and correct proof that in some precise sense half of N is even, or that the density of D_p is $\log(p+1)$, tells us nothing whatever about nature and $B(n)$ unless that 'precise sense' matches something relevant to the actual origin of $B(n)$.

The density arguments, reviewed in Sections 4 and 5 below, were generally given by their original authors without the least attempt at such justification, and the scale-invariance arguments of Section 6 are likewise philosophically barren. Not until Section 7, (Pinkham's second method) does a *scientific* theory appear, a formula that invokes observation in addition to calculation. The earlier arguments produce Benford's law uncritically and therefore predict its appearance in every possible context, even for $PP(n)$ and $V(n)$ where it fails. The statistical argument of Pinkham produces an approximation to Benford's law, with a criterion relating the closeness of the approximation to some other observable features of the phenomena.

4. Density and summability arguments, discrete model. A first attempt at density is the usual number-theoretic (or Cesaro) method. Call $\alpha_p(n)$ the characteristic function of D_p in N, so that $\alpha_p(n)=1$ if $n\in D_p$ and is otherwise 0. Then put

$$\alpha_p^1(n)=(1/n)\sum_{k=1}^{n}\alpha_p(k). \tag{4.1}$$

If $\lim_n \alpha_p^1(n)$ existed it would be the number-theoretic density of D_p, but the limit does not exist. In

Figure 1 of Section 5 below appears the graph of a function ϕ_1, defined on R^+, which is for all practical purposes the graph of α_1^1 in the sense that for $n \in N$, $|\phi_1(n) - \alpha_1^1(n)| \leqq 1/n$. Figure 1 shows the oscillatory nature of the averaging process for D_1 and much the same thing happens to α_p^1 for the other values of p. Graphs like Figure 1 and Figure 2 appear in **[14]** and **[22]**, and for other values of p in **[2]** and **[11]**.

When α_p^1 is plotted on semilog paper, as in Figure 2 below, one is struck with the near-periodicity of the result and tempted to take an average height of the nth period and call the limit the density of D_p. This is most conveniently done by an integral approximation, so a discussion of the results will be deferred to Section 5.

B. J. Flehinger does something different. Since α_p^1 doesn't converge, she reiterates the Cesaro process, putting

$$\alpha_p^t(n) = (1/n) \sum_{k=1}^{n} \alpha_p^{t-1}(k) \tag{4.2}$$

as $t = 2, 3, \ldots$. In **[14]** she proves that the successive functions oscillate ever more narrowly and that (see **[20]** for a clearer proof) the process converges to $\log(p+1)$ in the following sense:

$$\lim_t \liminf_n \alpha_p^t(n) = \lim_t \limsup_n \alpha_p^t(n) = \log(p+1).$$

The Flehinger limitation method is a regular method, that is, it yields ordinary limits when applied to convergent sequences. There is no shortage of regular methods, and an infinitude of them will *not* yield $\log(p+1)$ as the generalized limit of $\alpha_p(n)$, or even of $\alpha_p^t(n)$ for any given finite t. However, the Flehinger method has the property of being stronger than *all* iterations of the Cesaro methods, and it can be proved **[12B**; Section 6**]** that any *matrix* method having this property must agree with the Flehinger method whenever the latter applies. (Cf. also **[29]** and **[13B]** for relationships between these and yet other summability methods.)

An example of a method stronger than Flehinger's is the logarithmic matrix method $\mathscr{L}$, defined by

$$L_{nj} = (j \log_e n)^{-1} \text{ if } n \leqq j, \text{ and } L_{nj} = 0 \text{ if } n > j.$$

A sequence $\{s_n\}$ is called $\mathscr{L}$-summable if $\Sigma_j L_{nj} s_j$ converges in n. In the present case, then, where s_n is $\alpha_p(n)$, the characteristic function of D_p,

$$\lim_n (\log_e n)^{-1} \sum_{j=1}^{n} \alpha_p(j)/j = \log(p+1).$$

This result was obtained directly by R. L. Duncan **[23]**.

R. E. Whitney **[26]** has shown that the matrix method $\mathscr{L}$ can also be applied to the characteristic function of D_p in the sequence of primes, with the same result. Thus, if $\pi_p(n) = 1$ when the nth prime begins with a digit $\leqq p$, and $\pi_p(n) = 0$ otherwise, then

$$\lim_n (\log_e n)^{-1} \sum_{j=1}^{n} \pi_p(j)/j = \log(p+1).$$

Unlike the case of D_p in N, where one might suspect the Benford law by a careful examination of Figure 1 below, and adjust his choice of summation method accordingly, the same result for D_p in the primes hardly seems intuitive. Of the 1125 primes less than 9999, about 14 percent begin with the digit 1.

J. Cigler, in a personal communication to me in 1969, called my attention to the relevance, in connection with such density arguments, of the notion of equidistributed sequences. Let $\{b_n\}$ be a sequence of real numbers in the interval $[0, 1)$; the sequence is called equidistributed if for each subinterval $[a, b) \subset [0, 1)$ we have

$$\lim_k \frac{1}{k} \sum_{n=1}^{k} \beta(n) = b - a, \tag{4.3}$$

where $\beta(n) = 1$ when $b_n \in [a, b)$ and $\beta(n) = 0$ if $b_n \notin [a, b)$.

Now let $\{a_n\}$ be a sequence in $R^+\backslash\{0\}$, and let $b_n = \log a_n \pmod 1$. Then $a_n \in D_p$ (i.e., a_n has first digit $\leqq p$) if and only if $0 \leqq b_n < \log(p+1)$. If it turns out that b_n is equidistributed on $[0,1)$, then (4.3) holds with $[a,b) = [0, \log(p+1))$, and $\{a_n\}$ may be said to obey Benford's law in the strictest sense, the sense of ordinary density of D_p in the sequence $\{a_n\}$. Cigler proposed calling such a sequence a strong Benford sequence.

The sequence of the positive integers themselves, N, is not a strong Benford sequence, but any geometric sequence $\{ar^n\}$ is, provided r is not a rational power of 10. For, if $a_n = ar^n$, then $b_n = \log a + n \log r \pmod 1$, and it is well known (e.g., [**4**B; p. 390]) that arithmetic sequences with irrational spacing are equidistributed (mod 1). A heuristic 'proof' of this result may be found in [**22**].

The restriction "r is not a rational power of 10" is not entirely needed for $\{ar^n\}$ to have a good approximation to the Benford law. If $r = 10^{p/q}$, then

$$b_n = \log a + \frac{np}{q} \pmod 1$$

is periodic, i.e., $b_{n+q} = b_n$ for all n. The finite range of $\{b_n\}$ is equally spaced and comes as close to equidistribution on $[0,1]$ as the spacing $1/q$ allows. To be precise, we get instead of (4.3) the formula

$$\left| \lim_k \frac{1}{k} \sum_{n=1}^{k} \beta(n) - (b-a) \right| \leqq 1/q, \tag{4.4}$$

and since q can seldom be a small integer, the Benford law is well approximated for all but a finite number of ratios r.

Empirical verification of Benford's law for geometric sequences can be seen strikingly in the first hundred powers of 2 (row $P(n)$ in Table 1), and E. Hafner [**18**] has added some interesting flourishes to compilations of this kind. Benford himself was well aware of this property of geometric sequences, so much so that he made it the philosophic rock on which to base his law. While mere Man counts arithmetically, $1, 2, 3, 4, \ldots$, says Benford, *Nature* counts $e^0, e^x, e^{2x}, e^{3x}, \ldots$, "and builds and functions accordingly." Therefore, Benford's argument goes on, 'naturally' derived data tend to come in mixtures of geometric sequences, which obey Benford's law. He cites numerous examples from science and technology to support this view, all of them variants on "Fechner's Law", a bit of 19th century scientism that says (roughly), "Response is proportional to the logarithm of the stimulus." If Fechner's law made sense and were true, it would still not be apparent why tabular data should favor lists of stimuli over lists of responses. However, for a hilarious demolition of Fechner's law in a typical application, see [3B].

The logarithmic summability method also corresponds to an equidistribution notion. If, as before, $b_n = \log a_n \pmod 1$, Cigler suggests calling $\{a_n\}$ a weak Benford Sequence if $\{b_n\}$ is $(1/n)$-equidistributed, which is to say, if

$$\lim_k (\log_e n)^{-1} \sum_{n=1}^{k} \beta(n)/n = b - a, \tag{4.5}$$

where $\beta(n)$, a and b are as in (4.3) above. What Whitney proved, then, is that a_n = the nth prime defines a weak Benford sequence. Duncan's result in [**23**] now can be read as an alternate statement of the earlier known fact that $\{n\}$ is a weak Benford sequence. In [**5**B; **6**B; **14**B] are accounts of equidistribution (mod 1), and proofs that $\{\sqrt{n}\}$ and $\{Q(n)\}$ for any polynomial Q are also weak Benford sequences.

Benford's observation that "the greatest variations from the logarithmic relation [i.e., Benford's Law] were found in the first digits of mathematical tables from engineering handbooks...", which led him to the conclusion that "the logarithmic law applies particularly to those outlaw numbers that are without known relationship..." was thus a bit hasty, the result of a preference for Cesaro frequency counts over a more sophisticated weighting. Even so, geometric sequences, which do obey his law, can hardly be called 'outlaw... without known relationship'!

Nor are geometric sequences the only kind of strong Benford sequences. D. Singmaster **[30]** calls $\{a_n\}$ asymptotically geometric if there is a geometric sequence $\{ar^n\}$ such that $\lim_n (a_n/ar^n) = 1$. In this case $\log a_n - n \log r$ converges, so that $\log a_n \pmod 1$ is just as equidistributed as $n \log r$. In other words, asymptotically geometric sequences are (unless $\log r$ is rational) strong Benford sequences.

Now Singmaster notices that the Fibonacci sequence is asymptotically geometric with the golden mean $(1+\sqrt{5})/2$ as limiting ratio. Furthermore, $\log[(1+\sqrt{5})/2]$ is irrational, hence the Fibonacci numbers form a strong Benford sequence.

Singmaster's observation follows no less than three earlier papers [**15**; **25**; **28**] giving empirical data supporting the same statement. Yet none of these four authors noticed any of the series of four articles by Brown, Duncan, Kuipers and Shiue [**8B**; **9B**; **10B**; **11B**] which proved that not only the Fibonacci numbers but (almost) all sequences $\{a_n\}$, defined by linear recursion, have the property that $\log a_n \pmod 1$ is equidistributed on $[0,1)$, i.e., $\{a_n\}$ obeys Benford's law. Benford's law, however, was not explicitly mentioned in these papers. The proof given by Brown and Duncan rests precisely on the fact that the sequences in question are asymptotically geometric.

Singmaster does notice that the interleaving of a finite number of asymptotically geometric sequences produces a strong Benford sequence, and proposes the converse as a conjecture.

5. Density and summability arguments, continuous model. Instead of looking at sequences, or subsets of N, one can look at R^+ and ask "What fraction of R^+ is occupied by D_p?" For convenience in this section we shall only consider the domain $[1,\infty)$. If α_p is the characteristic function of D_p in $[1,\infty)$, let

$$\phi_p(x) = \frac{1}{x-1}\int_1^x \alpha_p(t)dt. \tag{5.1}$$

Explicit formulas are easily found. For $p=1$, for example,

$$\begin{aligned} \phi_1(x) &= 1 && \text{on } [1,2] \\ &= 1/(x-1) && \text{on } [2,10] \\ &= 1-8/(x-1) && \text{on } [10,20] \\ &= 11/(x-1) && \text{on } [20,100] \\ &= 1-88/(x-1) && \text{on } [100,200];\ \text{etc.} \end{aligned}$$

The graph of ϕ_1 is given in Figure 1, and in Figure 2 it is given again but with a logarithmically scaled x-axis to show its 'periodicity'. The maxima converge to 5/9 and the minima are 1/9. Similar graphs can be drawn for the other values of p.

George Stigler **[5]** obtains an 'average height' for ϕ_p by integrating over the nth cycle; he computes

$$\frac{1}{10^{n+1}-10^n}\int_{10^n}^{10^{n+1}} \phi_p(t)\,dt \tag{5.2}$$

and takes the limit as n increases. His results are given in Table 1 as Stigler's Law $S(n)$.

Benford, using the same graph, gets prob $D_1 = \log 2$, which is .301 as against Stigler's .241; how? He does it by using Figure 2 instead of Figure 1, i.e., by getting the average height of the nth cycle as actually depicted geometrically in Figure 2 and taking the limit as n increases. Thus Benford takes the limit of

$$\int_n^{n+1} \phi_p(10^x)dx, \tag{5.3}$$

which is not at all the same thing as Stigler's integral (5.2) under the relevant change of variable. (Both

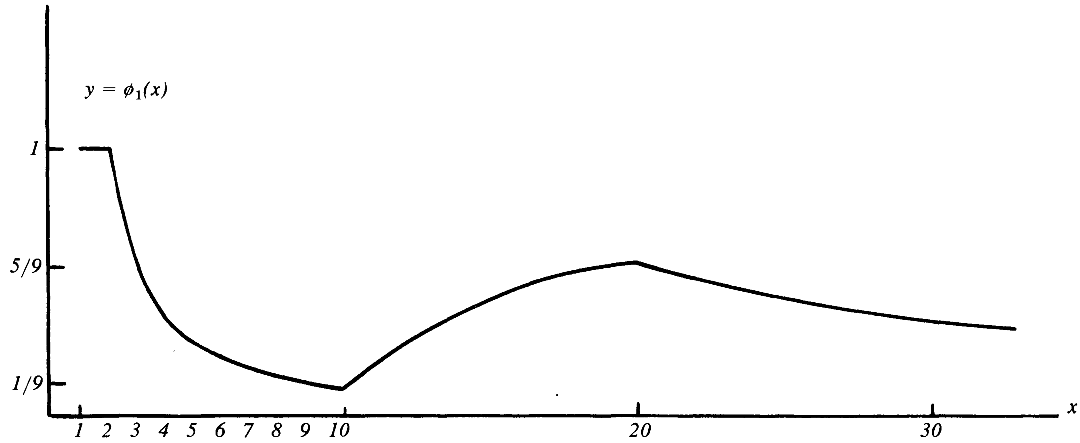

Fig. 1

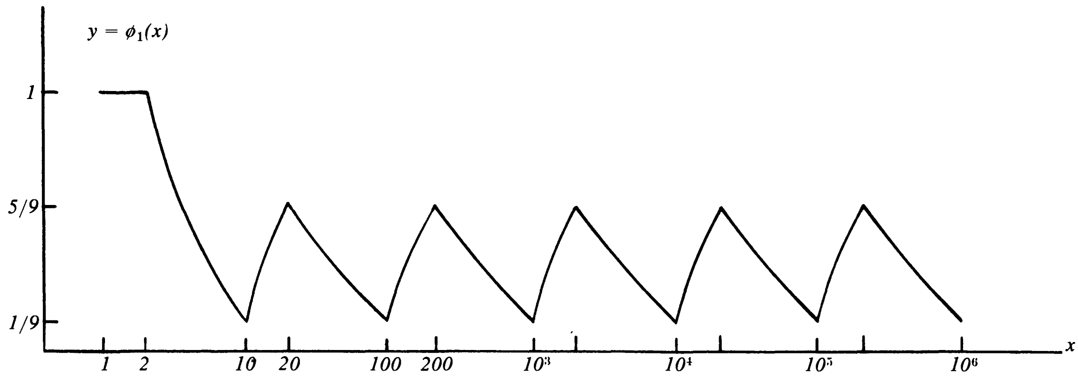

Fig. 2

Stigler's average and Benford's average will assign density 1/2 to the set of even integers, by the way. For this purpose (5.2) and (5.3) should be interpreted as the obviously related discrete sums rather than integrals.)

Warren Weaver, in his popularization "Lady Luck" **[11]**, also draws a graph akin to Figure 2, but for ϕ_4 rather than ϕ_1, to exhibit the probability that a number will begin with 1, 2, 3 or 4. He uses the logarithmic horizontal scale, calling it inessential to his reasoning, as indeed it is since he uses Stigler's method of averaging. But he *thinks* his result is going to be Benford's, apart from errors of approximation, and refers to Benford's as the correct 'theoretical' law.

If Weaver didn't notice the difference between the two methods, certainly Benford didn't; he didn't even consider the 'somewhat distorted' (Weaver's words) horizontal scale of Figure 2 worth mentioning. It took Stigler, an economist, to observe that the two summability methods correspond to two different philosophical hypotheses concerning equiprobability. By implication, Stigler therefore showed something very many of the writers on the problem either ignored or explicitly denied: that mathematics alone cannot justify a first digit law.

6. Probability interpretation of summability methods. From the definition (5.1), $\phi_p(t)$ can be regarded as a conditional probability, the probability that a random variable x is in D_p, given that x is drawn from the rectangular distribution with support $[1, t)$. Let us denote this condition by the phrase "$\max x = t$"; also, let us denote the event "x is drawn from a rectangular distribution with support $[1, t)$, where t is somewhere in the interval $[a, b)$" by $\{\max x \in [a, b)\}$. Now Stigler's integral (5.2) can be approximated by the Riemann sum

$$\sum_{i=1}^{M} \phi_p(t_i)\cdot(1/M), \tag{6.1}$$

where the points $t_0, t_1, \ldots, t_M$ form an equally-spaced partition of $[10^n, 10^{n+1}]$. In probability terms this sum can be rewritten

$$\sum_{i=1}^{M} \operatorname{prob}\{x \in D_p \mid \max x = t_i\}\cdot \operatorname{prob}\{\max x \in [t_{i-1}, t_i)\}, \tag{6.2}$$

provided we assume each of the M disjoint events $\{\max x \in [t_{i-1}, t_i)\}$ has probability $1/M$.

Now for large M, the conditions "$\max x = t_i$" and "$\max x \in [t_{i-1}, t_i)$" are not very different in the first factor of each term of (6.2); therefore, by a sort of Duhamel's Principle, (6.2) may be replaced by

$$\sum_{i=1}^{M} \operatorname{prob}\{x \in D_p \mid \max x \in [t_{i-1}, t_i)\}\cdot \operatorname{prob}\{\max x \in [t_{i-1}, t_i)\} \tag{6.3}$$

as a Riemann approximation to (5.2). By the summation law for conditional probabilities, (6.3) reduces to

$$\operatorname{prob}\{x \in D_p \mid \max x \in [10^n, 10^{n+1})\}, \tag{6.4}$$

which in turn is nothing but $\operatorname{prob}\{x \in D_p\}$, since we are assuming, for the time being, that $\max x \in [10^n, 10^{n+1})$ is in fact the case. Stigler's ratios $S(p)$ of Table 1 come from these probabilities' limiting values as $n \to \infty$.

Stigler's method therefore amounts to imagining that we have a large number of tables of data with the largest entry t_i in each table lying somewhere in the interval $[10^n, 10^{n+1})$ with equal probability, and that the ith table is a sample from the rectangular distribution supported by $[1, t_i)$; also that n is large, and the same for all tables. (Actually, n need not be very large for practical purposes; for $n = 2$ we are already within 1 percent of the values given as $S(p)$.)

Finally, we may drop the restriction that n is the same for all tables by imagining that the tables before us are mixtures from the above situation for various values of n. The hypothesis that the largest entry is equidistributed on $[10^n, 10^{n+1})$ for a mixture of values of n is translated by Stigler into the phrasing "The largest entry in each table is equally likely to begin with the integer 1, 2, ..., or 9." Stigler seems to think that this hypothesis, together with the hypothesis that the conditional distribution of the random entries in a table with a largest given entry is rectangular, produces his results, but I don't see that his phrasing is sufficient.

The corresponding interpretation of Benford's integral (5.3) proceeds from the approximating Riemann sum

$$\sum_{i=1}^{M} \phi_p(10^{t_i})\cdot(1/M), \tag{6.5}$$

where the t_i are equally spaced on $[n, n+1]$. In probability terms, as before, this becomes

$$\sum_{i=1}^{M} \operatorname{prob}\{x \in D_p \mid \max x = 10^{t_i}\}\cdot \operatorname{prob}\{\max x \in [10^{t_{i-1}}, 10^{t_i})\}, \tag{6.6}$$

analogously with (6.2), though now we must consider the events $\{\max x \in [10^{t_{i-1}}, 10^{t_i})\}$ all equal and of value $1/M$. As before, we may replace "$\max x = 10^{t_i}$" by "$\max x \in [10^{t_{i-1}}, 10^{t_i})$" in (6.6) and still have a Riemann–Duhamel approximation to (5.3). The summation law then reduces the new version of (6.6) to $\operatorname{prob}\{x \in D_p \mid \max x \in [10^n, 10^{n+1})\}$, whose limit as $n \to \infty$ gives us Benford's law $L(p)$.

In Stigler's interpretation, then, Benford's universe of tables of data has each table drawn from a rectangular distribution on $[1, t_i)$ where the "largest entry" t_i now has a skewed distribution on some given $[10^n, 10^{n+1})$, skewed in the sense that $\log t_i$ is rectangularly distributed on $[n, n+1)$. (Of course, the same comments about large n, and mixtures of n, apply as before.)

Stigler takes this 'inconsistency' — the combination of a rectangular conditional distribution for random entry with skewed distribution of highest entry — as a flaw in Benford's analysis, but in fact both of these 'urn models' are equally artificial, although it is conceivable they — one or the other — might apply in some particular context like the assignment of street addresses or the populations of cities.

The discrete summability schemes of Section 4 above can also be tortured into probability interpretations, though none of the authors mentioned there (except Diaconis) does so. However, A. Herzel [9] *starts* with an urn model for the question "What is $\text{prob}\{x \in D_p\}$?" in the domain of positive integers N, and ends with summability questions related to Flehinger's and Stigler's.

Herzel imagines M urns, the nth of which contains n balls numbered 1 to n. The conditional probability of choosing a ball whose first digit is $\leqq p$, given that we choose from the nth urn, is of course Flehinger's first average $\alpha_p^1(n)$ ((4.1) above), which is asymptotically $\phi_p(n)$ of Figure 1. The total probability of choosing a ball in D_p then depends on the *a priori* probabilities assigned for choosing among the urns. Herzel gives three schemes: equal probability, probability weighted linearly according to the size of the urn, and probability weighted as the square of the size of the urn.

None of these schemes produces a limit as $M \to \infty$, but Herzel obtains integral approximations and numerical results close to Benford's. Herzel's first scheme is the same as Flehinger's second approximation, and the others have justifications as arbitrary as the first. A vast amount of numerical data is given.

But the mere interpretation of an averaging device as a probability does nothing to answer the real-life question of why observed tables tend (or tend not) to obey Benford's (or Stigler's) law, unless we have some scientific information about the tabular entries enabling us to verify one or another of these 'urn model' hypotheses. It appears that in this form the scientific question is very hard to answer.

7. Another probability interpretation: Scale-invariance. Roger S. Pinkham [10], attributing the basic idea to R. Hamming, put forward an invariance principle attached to another sort of probability model, sufficient to imply Benford's law. If (say) a table of physical constants, or of the surface areas of a set of nations or lakes, is rewritten in another system of units of measurement, ergs for foot-pounds or acres for hectares, the result will be a rescaled table whose every entry is the same multiple of the corresponding entry of the original table. If the first digits of all the tables in the universe obey some fixed distribution law, Stigler's or Benford's or some other, that law must surely be independent of the system of units chosen, since God is not known to favor either the metric system or the English system. In other words, a universal first digit law, if it exists, must be scale-invariant. The simplest way to express scale-invariance, for D_p in R^+ at any rate, is to ask that $\text{prob}\{x \in D_p\} = \text{prob}\{x \in kD_p\}$ for all $k > 0$, whatever meaning we may attach to 'prob'.

Pinkham postulates, then, an 'underlying distribution of all physical constants' with a cumulative distribution function F, so that $F(x) = \text{prob}\{\text{physical constant is} \leqq x\}$. Then since $D_p = \bigcup_{n=-\infty}^{\infty}[10^n, (p+1)10^n)$ and $kD_p = \bigcup_{n=-\infty}^{\infty}[k \cdot 10^n, k(p+1)10^n)$,

$$\text{prob}\{x \in D_p\} = \sum_{n=-\infty}^{\infty} [F((p+1)10^n) - F(10^n)], \text{ and} \tag{7.1}$$

$$\text{prob}\{x \in kD_p\} = \sum_{n=-\infty}^{\infty} [F(k(p+1)10^n) - F(k \cdot 10^n)]. \tag{7.2}$$

Under the reasonable assumption that F is continuous, and that (7.1) = (7.2) for all $k > 0$ and all $p = 1, 2, \ldots, 9$, Pinkham proves that $\text{prob}\{x \in D_p\} = \log(p+1)$ for each p.

All depends on whether there *is* such a distribution function describing the universe of tabular data. As a scientific hypothesis it gives unease. For example, if h is the real number such that

$F(h) = 1/2$, then half the numbers in the universe are less than h, which makes h a most remarkable physical constant. Now what becomes of h when we exercise our freedom to make scale changes?

D. Knuth [20] in fact points out that, science and philosophy aside, mathematics alone can prove no such F exists. If instead of base 10 we used another base b for our numeration system, the set of all numbers in R^+ whose initial digit is $\leqq p$ would be

$$\Delta_p = \bigcup_{n=-\infty}^{\infty} [b^n, (p+1)b^n), \quad \text{and}$$

$$\text{prob } \{x \in \Delta_p\} = \sum_{n=-\infty}^{\infty} [F((p+1)b^n) - F(b^n)],$$

and the same argument used by Pinkham to prove $\text{prob}\{x \in D_p\} = \log(p+1)$ will also produce the result $\text{prob}\{x \in \Delta_p\} = \log_b(p+1)$. Therefore

$$\sum_{n=-\infty}^{\infty} [F(rb^n) - F(b^n)] = \log_b r$$

for all integers r and b with $1 < r < b$. It is not hard to prove no such F can exist.

All this is not to say, however, that the formula $\text{prob}\{x \in \Delta_p\} = \log_b(p+1)$ is wrong. It is in fact the correct generalization of Benford's law to tables written in the base b notation, and it can be derived from any of the summability arguments used above for base 10. As will be seen below, it is as valid as Benford's law, *mutatis mutandis*, wherever Benford's law has validity. In the present context it is used only to show that the Pinkham–Hamming distribution function F cannot exist, and not that deductions from the existence of such an F are necessarily false.

In [19] the present author proposed another sort of probability model with scale-invariance. Instead of using some F as above to give a countably-additive probability measure on R^+, a finitely-additive set function μ is used. Such functions, called Banach measures, exist with the properties: $\mu(R^+) = 1$; $\mu(A) \geqq 0$ for all $A \subset R^+$; $\mu(A \cup B) = \mu(A) + \mu(B)$ when A and B are disjoint; and $\mu(kA) = \mu(A)$ for all $k > 0$ and $A \subset R^+$. The last property is the scale-invariance. All such measures (they exist in profusion) agree on the sets D_p, giving them measure $\log(p+1)$. One can even demand the additional property $\mu([0, h]) = 0$ for all $h \in R^+$, avoiding the philosophically awkward 'midpoint' of all physical constants since the resulting measure is concentrated in the neighborhood of infinity.

There are other ways to improve the same model. We can add the requirement $\mu(A + k) = \mu(A)$ for all $k \geqq 0$ and $A \subset R^+$. This means that an affine change of scale, as from Fahrenheit to Celsius temperatures, will preserve the Benford law. Again, if instead of R^+ for the underlying measure space we wish to use the space of positive rational numbers, or even the set of all rationals with terminating decimal expansions, Banach measures with the corresponding requisite properties can be shown to exist and give the Benford law. Terminating decimals seem particularly appropriate as the model for tabular data, but they cannot of course be used to underlie a countably additive non-atomic probability theory.

Bumby and Ellentuck [21] describe a third sort of model involving a form of scale-invariance, which has its points of mathematical interest but adds little to an understanding of the first-digit problem, except in that it reinforces the notion that every 'reasonable' notion of density should assign value $\log(p+1)$ to D_p (this time as a subset of N).

Yet none of these arguments can finally be convincing, given the knowledge that any child can construct a list of numbers violating Benford's law. It would be perfectly consistent with every known theorem of mathematics to live in a universe whose *World Almanac* failed to contain a single entry beginning with an odd digit.

Besides, what can scale-invariance possibly have to do with the street addresses of the 342 American men of science canvassed by Benford?

8. Statistical explanation: Pinkham's second method. What is to my mind the true explanation is the statistical one given in the second part of Pinkham's paper [**9**]. It is a descriptive argument at bottom; anyone who thinks Isaac Newton didn't really explain the planetary motions, since he merely reduced them to an alternative formulation invoking an unexplained force of gravity, will also complain of what follows on the same philosophical grounds.

Suppose we are looking at a table of data drawn from some distribution $F: R^+ \to [0,1]$, so that $\text{prob}\{x \leqq a\} = F(a)$. We shall suppose F not only continuous but thrice differentiable if need be, for convenience, just as R^+ is for convenience our underlying universe. F is an approximation which can always be built to suit the finite situation at hand. By no means is F to be construed as describing the whole numerical universe, or even the whole book of data before us; it is only to be consistent with our sample.

Then, as in Pinkham's scale-invariance argument but with a different philosophical interpretation,

$$\text{Prob}\{x \in D_p\} = \sum_{k=-\infty}^{\infty} [F((p+1)10^k) - F(10^k)]. \tag{8.1}$$

Let

$$G(x) = F(10^x); \tag{8.2}$$

then

$$\text{Prob}\{x \in D_p\} = \sum_{k=-\infty}^{\infty} [G(k + \log(p+1)) - G(k)]. \tag{8.3}$$

If we define

$$H(x) = \sum_{k=-\infty}^{\infty} [G(k+x) - G(k)] \text{ for all } x \in [0,1], \tag{8.4}$$

then (8.3) may be rewritten

$$\text{Prob}\{x \in D_p\} = H(\log(p+1)) \quad (p = 1, 2, 3, \ldots, 9). \tag{8.5}$$

The statement of Benford's law is now simply

$$H(x) = x \text{ for } x = \log 2, \log 3, \ldots, \log 9; \tag{8.6}$$

and the scientific question becomes: how shall we recognize distributions F whose corresponding H have property (8.6)? And secondly, what reasons do we have to believe that our table of data does or does not derive from a distribution of this sort?

A sufficient condition for the exactness of Benford's law is obviously

$$H(x) = x \text{ for all } x \in [0,1), \tag{8.7}$$

and most of the analysis to be cited centers around (8.7) rather than (8.6). Let us denote the derivative F' by f and G' by g (these are the densities of their respective distribution functions); also H' by h. Then from (8.4) we have, for all $x \in [0,1)$,

$$h(x) = \sum_{k=-\infty}^{\infty} g(x+k); \tag{8.8}$$

and the sufficient condition (8.7) becomes

$$h(x) = 1 \text{ for all } x \in [0,1). \tag{8.9}$$

In terms of random variables, if F is the cumulative distribution function for the random variable x, then G is the cumulative for the variable $\log x$ and H for the variable $\log x \pmod 1$. We are looking for conditions on a variable x which assure that $\log x \pmod 1$ is uniformly (i.e., rectangularly) distributed on $[0,1]$. We can, by the way, include distributions whose densities fail to exist here and there, as the next example will show, and only demand that (8.9) be true at all but a finite number of points.

One sort of distribution satisfying (8.9) precisely comes easily to mind: Suppose F is such that for some fixed $\alpha \geqq 0$, G turns out to be piecewise linear with slope K_n on each interval $(\alpha + n, \alpha + n + 1)$, and such that $\Sigma K_n = 1$. Then $G' = g$ is a step-function such that (8.8) implies (8.9). Benford presciently exhibits this very phenomenon with $\alpha = \log 2$ (Fig. 3, taken from [**2**; p. 561]). G is after all only F plotted on semilog paper, so that Figure 3, though labelled as if it were F for the street addresses of the scientists, is geometrically a graph for G, and would actually be G if the abscissae were labelled with the logarithms of the numbers displayed.

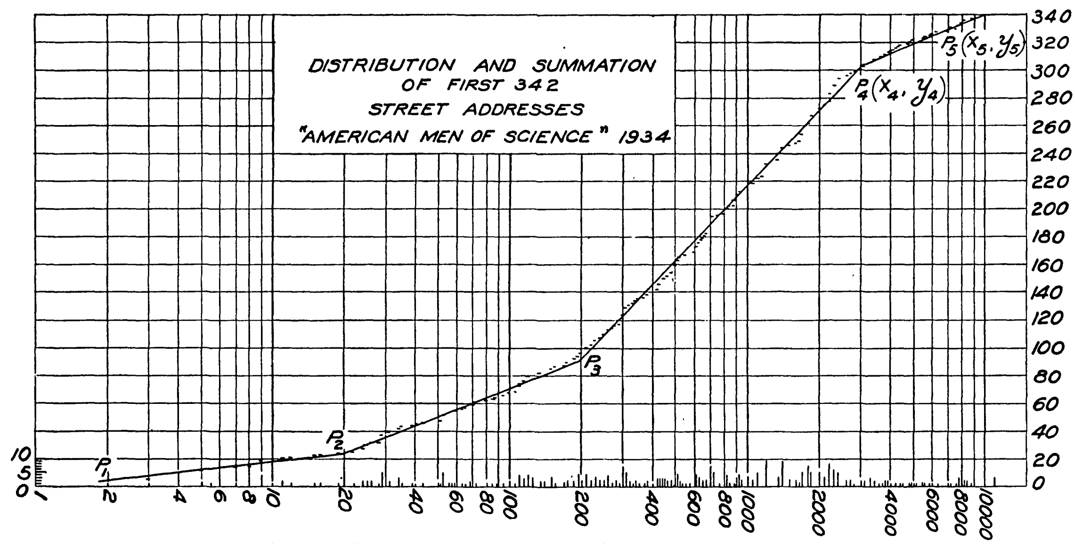

FIG. 3. Distribution and summation of first 342 street addresses, *American Men of Science*, 1934.

Had Benford been looking to verify (8.9) he would have been tempted to move the 'corner' P_4 of Figure 3 to a point above 2000, since the other corners are above 2, 20, 200 and 20,000, more or less (and slope zero after 20,000). This would have been not much of a distortion, and more convincing, than the use to which he did put Figure 3, which was to 'show' that the street addresses fell into a finite number of geometric sequences.

But a miracle like Figure 3 is too much to expect in general. The truly salient feature of Figure 3, and of any other distribution that obeys Benford's law more or less, inheres in its general shape, and not in any piecewise linearity.

Goudsmit and Furry [**3**] and especially Furry and Hurwitz [**4**] actually obtained formulas equivalent to (8.9) and understood perfectly well (though their language is by today's standards obscure) that for any given distribution some test of departure from (8.9) was needed. They actually calculated $\sup\{|h(x) - 1| : x \in [0,1)\}$ for a number of popular distributions, including the Cauchy law which is given below as an example, but they gave no general criteria.

Pinkham's real contribution was to apply some Fourier transform theory to the matter and arrive at the following explicit formulas:

$$H(x) - x = \sum_{\substack{k \neq 0 \\ k=-\infty}}^{\infty} R(k)[1 - \exp(-i2\pi kx)], \text{ where} \tag{8.10}$$

$$R(k) = (4\pi^2 k^2)^{-1} \int_{-\infty}^{\infty} \exp(i2\pi kt)\, dg(t). \tag{8.11}$$

Sufficient conditions for the validity of these formulas are that g be of bounded variation and that $\lim g(t)=0$ as $|t|\to\infty$. For many popular distributions, including the Cauchy, these conditions are satisfied.

Since $|\exp(i2\pi kt)|=1$, $|R(k)|\leqq(4\pi^2k^2)^{-1}\operatorname{var}[g]$, and since

$$|1-\exp(-i2\pi kx)|\leqq 2, \qquad |H(x)-x|\leqq\sum_{k=0} 2\operatorname{var}[g](4\pi^2k^2)^{-1};$$

summing the series in (8.10) gives

$$|H(x)-x|\leqq(1/6)\operatorname{var}[g], \text{ for all } x\in[0,1). \tag{8.12}$$

Pinkham stops with (8.12), but gives no examples to show his bound is practical. Feller [**13**] gets the rougher bound $|H(x)-x|<(x/2)\operatorname{var}[g]$ using the hypothesis that g is unimodal and monotone on each side of its maximum (which is therefore $(1/2)\operatorname{var}[g]$). It is almost intuitively evident that when g is 'wrapped up' mod 1 to give h (by (8.8) above), then (8.9) fails by at most $(1/2)\operatorname{var}[g]$.

But these bounds are not in fact very good. For an example, the Cauchy distribution has density function $f(x)=2a/\pi(x^2+a^2)$ satisfying all the relevant hypotheses. A calculation shows that $\operatorname{var}[g]=1.46$ approximately, so that for this case Pinkham's formula $|H(x)-x|<.25$ assures us only a miserable approximation to Benford's law. In fact, however, Furry and Hurwitz [**4**] have calculated $|H(x)-x|<.056x$ for the Cauchy distribution. In other words, a large sample from the Cauchy distribution will show $.284<$ Frequency of $x\in D_1<.318$, rather close to Benford's .301, while Pinkham's bound only gives $.05<$ Frequency of $x\in D_1<.55$.

At the suggestion of J. H. B. Kemperman I carried Pinkham's analysis one step further. Assuming F (and therefore g) has enough derivatives, one can integrate by parts in (8.11), obtaining

$$R(k)=(8\pi^3k^3)^{-1}\int_{-\infty}^{\infty} e^{i2\pi kt}dg'(t) \tag{8.13}$$

(provided $\lim_{|t|\to\infty} g'(t)=0$, a reasonable assumption). Hence $|R(k)|\leqq(8\pi^3k^3)^{-1}\operatorname{var}[g']$, and combining this with (8.10) yields the estimate

$$|H(x)-x|\leqq(4\pi^3)^{-1}\sum_{k\neq 0} 1/k^3\operatorname{var}[g'], \text{ or} \tag{8.14}$$

$$|H(x)-x|\leqq .0194\operatorname{var}[g'] \text{ approximately.} \tag{8.15}$$

In the case of the Cauchy distribution, to continue the example, it turns out that $\operatorname{var}[g']=2.53$, so that (8.14) gives $|H(x)-x|\leqq .0491$, about one-fifth the discrepancy allowed by Pinkham's formula.

Recently J. H. B. Kemperman, using a method quite different from that of Pinkham, has obtained the slightly better formulas (announced in [**31**]):

$$|H(x)-x|\leqq 1/8\operatorname{var}[g], \text{ and} \tag{8.16}$$

$$|H(x)-x|\leqq 1/64\operatorname{var}[g'], \tag{8.17}$$

with the additional information that 1/8 and 1/16 are the best possible coefficients. Indeed, he is able to describe the distributions which give rise to the worst possible cases. (They are quite jumpy.)

But in general even Kemperman's 'best possible' estimates in (8.16) and (8.17) are not very close to the true value of $|H(x)-x|$. The whole truth is contained in (8.10), and the shape of F (and therefore of g and g') affects the coefficients $R(k)$ as given in (8.11) or (8.13). The integrals here are in general badly majorized by $\operatorname{var}[g]$ or $\operatorname{var}[g']$ for all k, because the integrands are periodic with period $1/k$ with values on the unit circle. For most g the integrals should balance out to 0 or nearly so, rather than anything like $\operatorname{var}[g]$ or $\operatorname{var}[g']$. The literature still lacks a characterization in convenient terms predicting which distributions produce small discrepancies $|H(x)-x|$.

A direct study of (8.10) can throw light on an empirical observation. In Table 1, $PP(n)$ gave the first digit frequencies for populated places with populations $\geqq 2500$. F has a very steep slope at 2500 so that $tf(t)$ has a peak there. Since (from (8.2)) $g(t) = F(10^t) \cdot 10^t \log_e 10$, we see that dg and dg' give a heavy contribution in the neighborhood of the single point log 2500. The integrands in (8.11) and (8.13) therefore have no reason to be small. The Vancouver telephone numbers, deliberately assigned so as to omit the first digit 1, plainly are drawn from an even more artificial distribution, also producing peaks in g and g'.

9. Mixtures of distributions; invariance properties again. Some writers have pointed out that tables which occur naturally often represent distributions which are mixtures or composites of other distributions, and that the mixing process itself improves the approximation to Benford's law.

Furry and Hurwitz [**4**] prove the following: If f is a probability density on R^+, and if we define $f^*f(x) = \int_{-\infty}^{\infty}(1/t)f(x/t)f(t)dt$, then f^*f is also a density. In fact, f^*f corresponds to the procedure of mixing all the rescaled distributions $F(x/t)$ in fractions proportional to $f(t)$, as t runs through R^+. Reiterations of this procedure yield densities f^{*n}; and Furry and Hurwitz show that if h_n is the function h corresponding to f^{*n}, then $\lim_n |h_n(x) - 1| = 0$ uniformly in x. Thus any table in nature that can be construed as coming from high order mixtures of this sort can be expected to obey Benford's law (in the form (8.9)).

The Furry–Hurwitz convolution can also be interpreted as follows: If X_i are identically distributed independent random variables each with probability density function f, then f^{*n} is the probability density for the random variable $Y_n = \prod_{i=1}^n X_i$. Since $\log Y_n = \Sigma_{i=1}^n \log X_i$, the central limit theorem, which applies to sums of this sort under rather non-restrictive conditions, shows that the random variable $\log Y_n \pmod 1$ approaches uniform distribution as n increases. This is an alternate statement of the Furry–Hurwitz result.

Adhikari and Sarkar [**16**; **17**] take X_i as independent random variables uniformly distributed on $[0, 1]$ and prove directly, without reference to the above remarks, that the distribution of $\prod_{i=1}^n X_i$ tends to obey Benford's law as n increases. They do the same for powers in place of products and for certain other multiplicative combinations of the X_i. They also present data from computer simulations to exhibit these phenomena and their rates of convergence. For example, in [**16**] they take 60,000 five-digit random numbers, 6700 ± 100 of which begin with each of the nine possible digits. Machine calculation showed that the 60,000 eighth powers of these numbers obey Benford's law to within 10 percent.

Actually one doesn't have to go this far. The reader can look for himself at the 81 products in the schoolboy notebook multiplication table and see that they are already noticeably closer to Benford's distribution than the rectangularly distributed margins.

Adhikari and Sarkar [**16**; Th. 3] also give a partial converse to the theorems on scale invariance: If X is a random variable such that H, which is the distribution function of $\log X \pmod 1$, obeys (8.7) above, then the random variables $1/X$ and cX for any $c \in R^+ \backslash \{0\}$ have the same property. For, if X has the cumulative distribution function F as above, and $\log X$ the distribution G, an easy calculation shows that $\log X$ has the distribution $G_1(x) = G(x - \log c)$ and $\log 1/X$ has the distribution $G_2(x) = 1 - G(-x)$. Thus G_1 and G_2 have exactly the same dispersion properties as G. In particular, if (8.7) holds for X, it holds for cX and $1/X$.

Among other things, this observation contradicts the many writers who have said vaguely that Benford's law holds better when the distribution F (or its density f) is 'widespread' or 'covers several orders of magnitude'. In fact, a change of scale makes no difference whatever; not only will a table obeying Benford's law continue to do so after a change of scale, but a table that fails to do so will neither improve nor grow worse by rescaling (or taking reciprocals).

One must be careful to interpret the Benford law for this purpose in the sense of (8.7), which is stronger than (8.6). Hilda Geiringer [**7**] was even led to doubt the validity of Benford's law because of this confusion. She observed (as Hamming and Pinkham did some years later) that if Benford's law were true it should be scale-invariant, for the same reasons given by Pinkham. But she then imagined a table of 100 entries, 30 of which were the integer "1", 18 were "2", and so on in the proportions $L(n)$. Doubling every entry, however, would yield a table having very far from these first digit proportions.

The reason for this paradox is that Geiringer's table is a poor sample from any distribution which satisfies (8.7) although it may be a good sample from some crazy distribution satisfying only (8.6). A reasonable betting man confronted with a longish table he knows obeys Benford's law may be confident of its scale and inverse invariance.

This confidence was illustrated in [**22**; p. 118] when in preparing the article, I selected out of thin air a table of only 18 entries obeying Benford's law as well as so short a list can, and found the law so well preserved under the first four scale changes I chose ($c = 3, 6, 9$ and 12) that I published my first effort without the least change.

Finally, R. W. Hamming [**24**] adopts the finite point of view of a computing machine which for multiplicative purposes is indifferent to the position of the decimal point and considers the distribution whose density is given by

$$r(x) = (\log e)/x \quad \text{if } 1/10 \leqq x \leqq 1 \qquad (9.1)$$
$$= 0 \qquad \text{otherwise.}$$

This "reciprocal distribution" satisfies Benford's law, and is in fact the only distribution supported on $[1/10, 1]$ which will satisfy (8.9) precisely.

In floating point arithmetic the product or quotient of two numbers in $[1/10, 1]$ is again placed in $[1/10, 1]$ by a shift of decimal point if necessary. With this definition, and defining the *distance* of an arbitrary probability distribution density f (carried on $[1/10, 1]$) from r as

$$D(f) = \sup\left\{\frac{|f(x) - r(x)|}{r(x)} : x \in [1/10, 1]\right\},$$

Hamming proves the following two theorems:

(a) *If a random variable X has probability density r and if Y is any other random variable with density supported on* $[1/10, 1]$, *then XY, X/Y and Y/X all have the reciprocal distribution*;

(b) *If X and Y are any two random variables with densities supported on* $[1/10, 1]$, *and if X has the probability density f and XY has density g, then $D(g) \leqq D(f)$.*

Equality in (b) is quite rare, and Hamming gives some further information on rates of convergence to r. He also has some theorems on sums, but they have rather artificial hypotheses and do not tend to explain the occurrence of Benford's law as (a) and (b) do for compilations which result from repeated multiplications.

10. Final comments on the literature. Certain items in the Bibliography have not yet been given sufficient attention. The book by Furlan [**6**] I have not myself seen, but Hilda Geiringer's review [**7**] makes it plain that the work is more mystical than mathematical, and offers insight to neither scientist nor mathematician. Apparently Furlan argues, with a superabundance of detail, that Benford's law reflects a profound 'harmonic' truth of nature, related to Benford's notion that nature counts in geometric sequences.

C. Gini [**8**] refers to Furlan's book in a footnote, rather guardedly ("*... un grosso volume, di vari punti di vista molto interessante.*"), but makes no use of it whatever. Gini's paper mainly gives empirical data, census figures supporting Benford's law and so on, but insists that the law cannot be a universal truth or a mere property of the number system; for comparison's sake he tries to fit another law to some of his data.

A. Herzel [**9**], writing at the same time as Gini, also does homage to Furlan — three pages' worth

— but ends up doubting that a universal exponential law of economics or sociology exists. Herzel's paper is mainly devoted to the urn models which in effect try to make N into a probability measure space in which the sets D_p have the right measure. Diaconis [**29**] does much more along these lines.

Varian's note [**27**] is a suggestion that Benford's law be used as a partial test of the honesty or validity of purportedly random scientific data. The suggestion is attractive. In fact, in Table 1 above, one can observe that $PP(9)$ is about 1/3 the size of $PP(8)$, which is very far from the ratio predicted by Benford's law. It is true that the PP table doesn't satisfy the law overall, because of the 2500 threshold, but that really shouldn't affect $PP(9)/PP(8)$. One might therefore be led to suspect that census-takers have a tendency, when faced with a figure just a little below a power of 10, to round it up a little, putting it into the next order of magnitude. This may be wrong in the present case, but social scientists need all the tools of suspicion they can find.

The paper of J. Franel [**2B**] is a belated sequel to a famous argument of H. Poincaré [**1B**] in which the master attempted to show that the third digit of a table of logarithms should be as often odd as even. Franel shows this is not quite so, but that the probability of an odd nth digit converges to 1/2 as n increases.

A good number of the writers cited above have also discovered, quoted or proved (in some sense), as Franel almost did, a second digit law related to Benford's. If S_p is the set of all real numbers in R^+ whose second digit is $\leqq p$ $(p = 0, 1, 2, \ldots, 9)$, the law is

$$\text{Prob}\{x \in S_p\} = \sum_{k=1}^{9} [\log(k + (p+1)/10) - \log k], \tag{10.1}$$

and it follows immediately from any argument that assures (8.7). There are corresponding formulas for third and higher digits, all of which may be summarized in one diagram: The C scale of a slide rule. The probabilities are just those fractions of the C scale occupied by the sets in question.

Another popular generalization repeated very often is the substitution of some base b other than 10 for this entire discussion. When $b = 2$ the first digit problem is of course trivial but otherwise there is nothing new in the generalization; every argument that applies to 10 applies to *b mutatis mutandis.*

Finally, it seems obligatory to describe Simon Newcomb's paper [**1**], in which, by the way, the second digit law is already given and the inessential character of the base 10 already noted. Newcomb notes that any positive number can be written 10^s for some real number s, and that since reducing s modulo 1 does not affect the first digit behavior of 10^s, we may as well assume $0 \leqq s < 1$. All that is needed now is some reasoning sufficient to give Newcomb's conclusion that "The law of probability of the occurrence of numbers is such that all mantissae of their logarithms are equally probable," i.e., that s is uniformly distributed on $[0, 1)$.

Alas, the argument Newcomb does give begs the question, though it has at least the virtue of brevity. At the crucial point he says "... it is evident...". One is forced to conclude that the uniform distribution of s (mod 1) was an inspired guess, akin to Buffon's guess that the angle of his needle, not its cosine or log sin, is what obeys a uniform distribution law (mod π).

One can easily invent a ('biassed'?) needle-tossing machine which violates Buffon's hypothesis, just as any phone company can print a directory violating Benford's law. What remains tantalizing is the notion that there is still some unexplained measure in the universe which says that the probability of such violations is small.

NOTE: References **32** through **37**, and **15B**, were added in proof. Of all these additions, only [**33**] invokes a method or point of view disjoint from those reviewed in this article. I am also in possession of recent preprints or otherwise unpublished memoirs on the first digit problem written by the following authors: Persi Diaconis (Stanford University, Stanford, Calif.), Robert J. Epp (University of California at Los Angeles, Calif.), and Dennis P. Allen, Jr. (Bell Telephone Laboratories, Holmdell, New Jersey 07733).

References

1. Simon Newcomb, Note on the frequency of use of the different digits in natural numbers, Amer. J. Math., 4 (1881) 39–40.

2. Frank Benford, The law of anomalous numbers, Proc. Amer. Phil. Soc., 78 (1938) 551–572.

3. S. A. Goudsmit and W. H. Furry, Significant figures of numbers in statistical tables, Nature, 154 (1944) 800–801.

4. W. H. Furry and Henry Hurwitz, Distributions of numbers and distribution of significant figures, Nature, 155 (1945) 52–53.

5. George J. Stigler, The distribution of leading digits in statistical tables, unpublished but apparently written about 1945. Prof. Stigler's 1975 address was Haskell Hall, University of Chicago, Chicago, Illinois 60637.

6. L. V. Furlan, Das Harmoniegesetz der Statistik: Eine Untersuchung über die metrische Interdependenz der sozialen Erscheinungen, Basel, Switzerland: Verlag für Recht und Gesellschaft A.-G. 1946, pp. xii, 504.

7. Hilda Geiringer, Review of L. V. Furlan's book [6], J. Am. Stat. Assn., 43 (1948) 325–328.

8. C. Gini, Sulla frequenza delle cifre iniziali dei numeri osservati, Bull. Inst. Internat. Stat., (The Hague) 34 (1957) 57–76 (English and French summaries).

9. A. Herzel, Sulla distribuzione delle cifre iniziali dei numeri statistici, atti della XV e XVI Riunione: Societa Italiana di Statistica, Roma (1957), 205–228.

10. Roger S. Pinkham, On the distribution of first significant digits, Ann. Math. Statist., 32 (1961) 1223–1230.

11. Warren Weaver, Lady Luck: The Theory of Probability, Doubleday Anchor Series, New York, 1963. (See pp. 270–277.)

12. I. J. Good, Letter to the Editor, The American Statistician, 19 (1965) 43.

13. William Feller, Introduction to Probability Theory, Vol 2, Wiley, New York, 1966. (See p. 62.)

14. B. J. Flehinger, On the probability that a random number has initial digit A, this MONTHLY, 73 (1966) 1056–1061.

15. L. R. Shenton, Periodicity and density of modified Fibonacci sequences, Fibonacci Quart., 6 (1968) 109–116.

16. A. K. Adhikari and B. P. Sarkar, Distribution of most significant digit in certain functions whose arguments are random variables, Sankhyā Ser. B (Indian J. of Statist.) 30 (1968) 47–58.

17. A. K. Adhikari, Some results on the distribution of the most significant digit, Sankhyā Ser. B 31 (1969) 413–420.

18. Everett M. Hafner, Circular slide roulette, unpublished but dated 1969. Prof. Hafner's 1975 address was Hampshire College, Amherst, Mass., U.S.A.

19. Ralph A. Raimi, On the distribution of first significant figures, this MONTHLY, 76 (1969) 342–348.

20. Donald Knuth, The Art of Computer Programming, Vol. 2, Addison-Wesley, Reading, Mass., 1969. (See pp. 219–229.)

21. R. Bumby and E. Ellentuck, Finitely additive measures and the first digit problem, Fund. Math., 65 (1969) 33–42.

22. Ralph A. Raimi, The peculiar distribution of first digits, Scientific American, 221 (Dec., 1969) 109–120.

23. R. L. Duncan, Note on the initial digit problem, Fibonacci Quart., 7 (1969) 474–475.

24. R. W. Hamming, On the distribution of numbers, Bell System Tech. J., 49 (1970) 1609–1625.

25. J. Wlodarski, Fibonacci and Lucas numbers tend to obey Benford's law, Fibonacci Quart., 9 (1971) 87–88.

26. R. E. Whitney, Initial digits for the sequence of primes, this MONTHLY, 79 (1972) 150–152.

27. H. R. Varian, Benford's law, The American Statistician, 26 (1972) 65–66.

28. W. A. Sentance, A further analysis of Benford's law, Fibonacci Quart., 11 (1973) 490–494.

29. Persi Diaconis, Limits of measures of the integers with applications to random number generators and the distribution of leading digits, Harvard U. Dept. of Statistics Memorandum NS-211, Cambridge, Mass. (USA), 22 March 1973.

30. D. Singmaster, Recurrent sequences and Benford's law, preprint, submitted to Fibonacci Quart., in 1974.

31. J. H. B. Kemperman, Bounds on the discrepancy modulo 1 of a real random variable, Inst. of Math. Stat. Bulletin, 4 (1975) p. 138, abstract No. 75t–47.

32. N. Macon and L. Moser, On the distribution of first digits of powers, Scripta Math., 16 (1950) 290–291.

33. D. Blackwell and M. A. Girshick, Theory of Games and Statistical Decisions, Wiley, New York, 1954 (see p. 74, Problem 2.7.2).

34. B. P. Sarkar, An observation on the significant digits of the binomial coefficients and factorials, Sankhyā Ser. B, 35 (1973) 363–364.

35. Nai-kuan Tsao, On the distributions of significant digits and roundoff errors, Communications of the Association for Computing Machinery, 17 (1974) 269–271.

36. Basil Davis, Some remarks on initial digits, Fibonacci Quart., 14 (1976) 13–14.

37. Daniel I. A. Cohen, An explanation of the first digit phenomenon, J. Combinatorial Theory Ser. A, 20 (1976) to appear.

* * *

1B. H. Poincaré, Calcul des Probabilités, Gauthier-Villars, Paris, 1912, pp. 313–320.
2B. J. Franel, A propos des tables de logarithmes, Festschrift der Naturforschenden Gesellschaft in Zürich Vierteljahrsschrift 62 (1917), 286–295.
3B. Parry Moon, The Scientific Basis of Illuminating Engineering, McGraw-Hill, New York, 1936 (see pp. 420 ff).
4B. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford Univ. Press, New York, 1960.
5B. J. Cigler and G. Helmberg, Neuere Entwicklungen der Theorie der Gleichverteilung, Jber. Deutsch. Math.-Verein., 64 (1961) 1–50.
6B. J. Cigler, Methods of summability and uniform distribution mod 1, Compositio Math., 16 (1964) 44–51.
7B. G. M. Petersen, Regular Matrix Transformations, McGraw-Hill, London, 1966.
8B. R. L. Duncan, An application of uniform distribution to the Fibonacci numbers, Fibonacci Quart., 5 (1967) 137–140.
9B. L. Kuipers, Remark on a paper by R. L. Duncan, etc., Fibonacci Quart., 7 (1969) 465–466, 473.
10B. J. L. Brown and R. L. Duncan, Modulo 1 distribution of the sequence of logarithms of certain recursive sequences, Fibonacci Quart., 8 (1970) 482–486.
11B. L. Kuipers and J.-S. Shiue, Remark on a paper by Duncan and Brown, etc., Fibonacci Quart., 11 (1973) 292–294.
12B. J. P. Duran, Almost-convergence, summability and ergodicity, Canad. J. Math., 26 (1974) 372–387. (See Theorem 6.2.)
13B. Persi Diaconis, Weak and Strong Averages in probability and the theory of numbers, Cambridge, Mass., Harvard University Ph.D. Thesis (Dept. of Statistics) 1974.
14B. L. Kuipers and H. Niederreiter, Uniform Distribution of Sequences, Wiley, New York, 1974.
15B. E. Thorp and R. Whitley, Poincaré's conjecture and the distribution of digits in logarithm tables, Comp. Math., 23 (1971) 233–250.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ROCHESTER, ROCHESTER, N.Y. 14627.

SOME INTERESTING PROPERTIES OF THE VARIATION FUNCTION

FRANK N. HUGGINS

For a function f which is of bounded variation on $[a, b]$, V_f denotes the function defined, for each x in $[a, b]$, by $V_f(x) = V_a^x(f)$, where $V_a^x(f)$ denotes the total variation of f on $[a, x]$. V_f, called the variation function of f on $[a, b]$, possesses some rather interesting properties, particularly the property of inheriting certain properties from its parent function f. While some of the properties of V_f are well known, others are not so well known, and it is the purpose of this paper to present a unified exposition of the properties of the variation function. In order to make this material more readily accessible to the advanced undergraduate or graduate student in mathematics, the author has given detailed proofs of all theorems in this paper. Although expository in nature, this paper contains some new results, namely Theorem 5, and Theorem 6 which is a generalization of a known result.

DEFINITION 1. The statement that the function f is of **bounded variation** on $[a, b]$ means that f is a function whose domain includes $[a, b]$ and there exists a nonnegative number B such that if $\{x_i\}_{i=0}^n$ is any subdivision of $[a, b]$, then

$$\sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \leqq B. \tag{1}$$

The least such number B is called the (total) variation of f on $[a, b]$ and is denoted by $V_a^b(f)$. (Note: $V_a^a(f) = 0$.)