Statistical Science
1995, Vol. 10, No. 4, 354–363

A Statistical Derivation of the Significant-Digit Law

Theodore P. Hill

Abstract. The history, empirical evidence and classical explanations of the significant-digit (or Benford's) law are reviewed, followed by a summary of recent invariant-measure characterizations. Then a new statistical derivation of the law in the form of a CLT-like theorem for significant digits is presented. If distributions are selected at random (in any "unbiased" way) and random samples are then taken from each of these distributions, the significant digits of the combined sample will converge to the logarithmic (Benford) distribution. This helps explain and predict the appearance of the significant-digit phenomenon in many different empirical contexts and helps justify its recent application to computer design, mathematical modelling and detection of fraud in accounting data.

Key words and phrases: First-digit law, Benford's law, significant-digit law, scale invariance, base invariance, random distributions, random probability measures, random k-samples, mantissa, logarithmic law, mantissa sigma algebra.

THE SIGNIFICANT-DIGIT LAW

The significant-digit law of statistical folklore is the empirical observation that in many naturally occurring tables of numerical data, the leading significant digits are not uniformly distributed as might be expected, but instead follow a particular logarithmic distribution. The first known written reference is an 1881 two-page article in the *American Journal of Mathematics* by the astronomer/mathematician Simon Newcomb, who stated:

> The law of probability of the occurrence of numbers is such that all mantissae of their logarithms are equally likely.

[Recall that the *mantissa* (base 10) of a positive real number x is the unique number r in $[1/10, 1)$ with $x = r \times 10^n$ for some integer n; e.g., the mantissas of 314 and 0.0314 are both 0.314.]

This law implies that a number has leading significant digit 1 with probability $\log_{10} 2 \cong 0.301$, leading significant digit 2 with probability $\log_{10}(3/2) \cong 0.176$ and so on monotonically down to probability 0.046 for leading digit 9. The exact laws for the first two significant digits (also given by Newcomb) are

$$\text{Prob(first significant digit} = d) = \log_{10}(1 + d^{-1}), \quad d = 1, 2, \ldots, 9, \tag{1}$$

and

$$\text{Prob(second significant digit} = d) = \sum_{k=1}^{9} \log_{10}(1 + (10k + d)^{-1}), \quad d = 0, 1, 2, \ldots, 9. \tag{2}$$

The general form of the law,

$$\text{Prob(mantissa} \le t/10), = \log_{10} t, \quad t \in [1, 10) \tag{3}$$

even specifies the *joint distribution* of the significant digits. Letting D_1, $D_2, \ldots$ denote the (base 10) *significant-digit functions* [e.g., $D_1(0.0314) = 3$, $D_2(0.0314) = 1$, $D_3(0.0314) = 4$], the general law (3) takes the following form:

(4) *General significant-digit law.* For all positive integers k, all $d_1 \in \{1, 2, \ldots, 9\}$ and all $d_j \in \{0, 1, \ldots, 9\}$, $j = 2, \ldots, k$,

$$\text{Prob}(D_1 = d_1, \ldots, D_k = d_k) = \log_{10}\left[1 + \left(\sum_{i=1}^{k} d_i \times 10^{k-i}\right)^{-1}\right].$$

In particular, $\text{Prob}(D_1 = 3, D_2 = 1, D_3 = 4) = \log_{10}(1 + (314)^{-1}) \cong 0.0014$.

Theodore P. Hill is Professor of Mathematics at the School of Mathematics and Center for Applied Probability, Georgia Institute of Technology, Atlanta, Georgia 30332-0160.

A perhaps surprising corollary of the general law (4) is that (cf. Hill, 1995b)

the significant digits are dependent

and not independent as one might expect.

From (2) it follows that the (unconditional) probability that the second digit is 2 is $\cong 0.109$, but by (4) the (conditional) probability that the second digit is 2, *given* that the first digit is 1, is $\cong 0.115$. This dependence among significant digits decreases rapidly as the distance between the digits increases, and it follows easily from the general law (4) that the distribution of the nth significant digit approaches the uniform distribution on $\{0, 1, \ldots, 9\}$ exponentially fast as $n \to \infty$. This article will concentrate on decimal (base-10) representations and significant digits; the corresponding analog of (3) for other bases $b > 1$ is simply Prob (mantissa (base b) $\leq t/b) = \log_b t$ for all $t \in [1, b)$.]

EMPIRICAL EVIDENCE

Of course, many tables of numerical data do *not* follow this logarithmic distribution—lists of telephone numbers in a given region typically begin with the same few digits—and even "neutral" data such as square-root tables of integers are not good fits. However, a surprisingly diverse collection of empirical data does seem to obey the significant-digit law.

Newcomb (1881) noticed "how much faster the first pages [of logarithmic tables] wear out than the last ones," and after several short heuristics, concluded the equiprobable-mantissae law. Some 57 years later the physicist Frank Benford rediscovered the law and supported it with over 20,000 entries from 20 different tables including such diverse data as areas of 335 rivers, specific heats of 1389 chemical compounds, American League baseball statistics and numbers gleaned from *Reader's Digest* articles and front pages of newspapers. Although Diaconis and Freedman (1979, page 363) offer convincing evidence that Benford manipulated round-off errors to obtain a better fit to the logarithmic law, even the unmanipulated data are a remarkably good fit. Newcomb's article having been overlooked, the law also became known as Benford's law.

Since Benford's popularization of the law, an abundance of additional empirical evidence has appeared. In physics, for example, Knuth (1969) and Burke and Kincanon (1991) observed that of the most commonly used physical constants (e.g., the constants such as speed of light and force of gravity listed on the inside cover of an introductory physics textbook), about 30% have leading significant digit 1. Becker (1982) observed that the decimal parts of failure (hazard) rates often have a logarithmic distribution, and Buck, Merchant and Perez (1993), in studying the values of the 477 radioactive half-lives of unhindered α decays which have been accumulated throughout the present century and which vary over many orders of magnitude, found that the frequency of occurrence of the first digits of both measured and calculated values of the half-lives is in "good agreement" with Benford's law.

In scientific calculations the assumption of logarithmically distributed mantissae "is widely used and well established" (Feldstein and Turner, 1986, page 241), and as early as a quarter-century ago, Hamming (1970, page 1609) called the appearance of the logarithmic distribution in floating-point numbers "well-known." Benford-like input is often a common assumption for extensive numerical calculations (Knuth, 1969), but Benford-like *output* is also observed even when the input has random (non-Benford) distributions. Adhikari and Sarkar (1968) observed experimentally "that when random numbers or their reciprocals are raised to higher and higher powers, they have log distribution of most significant digit in the limit." Schatte (1988, page 443) reports that "In the course of a sufficiently long computation in floating-point arithmetic, the occurring mantissas have nearly logarithmic distribution."

Extensive evidence of the significant-digit law has also surfaced in accounting data. Varian (1972) studied land usage in 777 tracts in the San Francisco Bay area and concluded "As can be seen, both the input data and the forecasts are in fairly good accord with Benford's Law." Nigrini and Wood (1995) show that the 1990 census populations of the 3141 counties in the United States "follow Benford's Law very closely," and Nigrini (1996) calculated that the digital frequencies of income tax data reported to the Internal Revenue Service of interest received and interest paid is an extremely good fit to Benford. Ley (1995) found "that the series of one-day returns on the Dow-Jones Industrial Average Index (DJIA) and the Standard and Poor's Index (S&P) reasonably agrees with Benford's law."

All these statistics aside, the author also highly recommends that the justifiably skeptical reader perform a simple experiment, such as randomly selecting numerical data from front pages of several local newspapers, "or a Farmer's Almanack" as Knuth (1969) suggests.

CLASSICAL EXPLANATIONS

Since the empirical significant-digit law (4) does not specify a well-defined statistical experiment or sample space, most attempts to prove the law have been purely mathematical (deterministic) in nature, attempting to show that the law "is a built-in characteristic of our number system," as Weaver (1963) called it. The idea was to prove first that the set of real numbers satisfies (4), and then suggest that this explains the empirical statistical evidence.

A common starting point has been to try to establish (4) for the positive integers $\mathbb{N}$, beginning with the prototypical set $\{D_1 = 1\} = \{1, 10, 11, 12, 13, 14, \ldots, 19, 100, 101, \ldots\}$, the set of positive integers with leading significant digit 1. The source of difficulty and much of the fascination of the problem is that this set $\{D_1 = 1\}$ does not have a *natural density* among the integers, that is,

$$\lim_{n\to\infty} \frac{1}{n} \{D_1 = 1\} \cap \{1, 2, \ldots, n\}$$

does not exist, unlike the sets of even integers or primes which have natural densities 1/2 and 0, respectively. It is easy to see that the empirical density of $\{D_1 = 1\}$ oscillates repeatedly between 1/9 and 5/9, and thus it is theoretically possible to assign any number in [1/9, 5/9] as the "probability" of this set. Flehinger (1966) used a reiterated-averaging technique to define a generalized density which assigns the "correct" Benford value $\log_{10} 2$ to $\{D_1 = 1\}$, Cohen (1976) showed that "any generalization of natural density which applies to the [significant digit sets] and which satisfies one additional condition must assign the value $\log_{10} 2$ to $[\{D_1 = 1\}]$" and Jech (1992) found necessary and sufficient conditions for a finitely-additive set function to be the log function. None of these solutions, however, resulted in a true (countably additive) *probability*, the difficulty being exactly the same as that in the foundational problem of "picking an integer at random" (cf. de Finetti, 1972, pages 86 and 98–99), namely, if each singleton integer occurs with equal probability, then countable additivity implies that the whole space must have probability zero or infinity.

These discrete-summability arguments have been extended via various integration schemes, Fourier analysis and Banach measures to continuous densities on the positive reals, where $\{D_1 = 1\}$ is now the set of positive numbers with first significant digit 1, that is,

$$\{D_1 = 1\} = \bigcup_{n=-\infty}^{\infty} [1, 2) \times 10^n. \tag{5}$$

One popular assumption in this context has been that of *scale invariance*, which corresponds to the intuitively attractive idea that any universal law should be independent of units (e.g., metric or English). The problem here, however, as Knuth (1969) observed, is that *there is no scale-invariant Borel probability measure on the positive reals* since then the probability of the set (0, 1) would equal that of $(0, s)$ for all s, which again would contradict countable additivity. (Raimi, 1976, has a good review of many of these arguments.) Just as with the density proofs for the integers, none of these methods yielded either a true probabilistic law or any statistical insights.

Attempts to prove the law based on various urn schemes for picking significant digits at random have been equally unsuccessful in general, although in some restricted settings log-limit laws have been established. Adhikari and Sarkar (1968) proved that powers of a uniform (0, 1) random variable satisfy Benford's law in the limit, Cohen and Katz (1984) showed that a prime chosen at random with respect to the zeta distribution satisfies the logarithmic significant-digit law and Schatte (1988) established convergence to Benford's law for sums and products of certain nonlattice i.i.d. variables.

THE NATURAL PROBABILITY SPACE

The task of putting the significant-digit law into a proper countably additive probability framework is actually rather easy. Since the conclusion of the law (4) is simply a statement about the significant-digit functions (random variables) $D_1, D_2, \ldots$, let the sample space be $\mathbb{R}^+$, the set of positive reals, and let the sigma algebra of events simply be the σ-field generated by $\{D_1, D_2, \ldots\}$ [or equivalently, generated by the single function $x \mapsto \text{mantissa}(x)$]. It is easily seen that this σ-algebra, which will be denoted $\mathscr{M}$ and will be called the (decimal) *mantissa σ-algebra*, is a sub-σ-field of the Borels and that in fact

$$S \in \mathscr{M} \quad \Leftrightarrow \quad S = \bigcup_{n=-\infty}^{\infty} B \times 10^n \quad \text{for some Borel } B \subseteq [1, 10), \tag{6}$$

which is just the obvious generalization of the representation (5) for $\{D_1 = 1\}$.

The mantissa σ-algebra $\mathscr{M}$, although quite simple, has several interesting properties:

(7)
(i) every nonempty set in $\mathscr{M}$ is infinite with accumulation points at 0 and at $+\infty$;
(ii) $\mathscr{M}$ is closed under scalar multiplication ($s > 0$, $S \in \mathscr{M} \Rightarrow sS \in \mathscr{M}$);
(iii) $\mathscr{M}$ is closed under integral roots ($m \in \mathbb{N}, S \in \mathscr{M} \Rightarrow S^{1/m} \in \mathscr{M}$), but not powers;
(iv) $\mathscr{M}$ is self-similar in the sense that if $S \in \mathscr{M}$, then $10^m S = S$ for every integer m

(where aS and S^a denote the sets $\{as: s \in S\}$ and $\{s^a: s \in S\}$, respectively).

Property (i) implies that finite intervals such as $[1, 2)$ are *not* in $\mathscr{M}$ (i.e., are not expressible in terms of the significant digits alone; e.g., significant digits alone cannot distinguish between the numbers 2 and 20) and thus the countable-additivity contradictions associated with scale invariance disappear. Properties (i), (ii) and (iv) follow easily by (6), but (iii) warrants a closer inspection. The square root of a set in $\mathscr{M}$ may consist of two "parts," and similarly for higher roots. For example, if

$$S = \{D_1 = 1\} = \bigcup_{n=-\infty}^{\infty} [1, 2) \times 10^n,$$

then

$$S^{1/2} = \bigcup_{n=-\infty}^{\infty} [1, \sqrt{2}) \times 10^n \cup \bigcup_{n=-\infty}^{\infty} [\sqrt{10}, \sqrt{20}) \times 10^n \in \mathscr{M},$$

but

$$S^2 = \bigcup_{n=-\infty}^{\infty} [1, 4) \times 10^{2n} \notin \mathscr{M},$$

since it has gaps which are too large and thus cannot be written in terms of $\{D_1, D_2, \ldots\}$. Just as property (ii) is the key to the hypothesis of scale invariance, property (iv) is the key to a hypothesis of base invariance, which will be described below.

(Although the space $\mathbb{R}^+$ is emphasized above, the analogous mantissa σ-algebra on the positive integers $\mathbb{N}$ is essentially the same and as such removes the countable-additivity density problem on $\mathbb{N}$ since nonempty finite sets are not in the domain of the probability function.)

SCALE AND BASE INVARIANCE

With the proper measurability structure now identified, a rigorous notion of scale invariance is easy to state. Recall (7) (ii) that $\mathscr{M}$ is closed under scalar multiplication.

DEFINITION 1. A probability measure P on $(\mathbb{R}^+, \mathscr{M})$ is *scale invariant* if $P(S) = P(sS)$ for all $s > 0$ and all $S \in \mathscr{M}$.

In fact, scale invariance characterizes the general significant-digit law (4).

THEOREM 1 (Hill, 1995a). *A probability measure P on $(\mathbb{R}^+, \mathscr{M})$ is scale invariant if and only if*

$$P\left(\bigcup_{n=-\infty}^{\infty} [1, t) \times 10^n\right) = \log_{10} t \quad \text{for all } t \in [1, 10). \tag{8}$$

One possible drawback to a hypothesis of scale invariance in tables of "universal constants," however, is the special role played by the constant 1. For example, consider the two physical laws $f = ma$ and $e = mC^2$. Both laws involve universal constants, but the force equation constant 1 is not recorded in most tables, whereas the speed of light constant C is. If a "complete" list of universal physical constants also included the 1's, it seems plausible that this special constant might occur with strictly positive frequency. However, that would violate scale invariance, since then the constant 2 (and all other constants) would occur with this same positive probability.

Instead, suppose it is assumed that any reasonable universal significant-digit law should be *base invariant*, that is, should be equally valid when rewritten in terms of bases other than 10. In fact, all of the classical arguments supporting Benford's law carry over mutatis mutandis (Raimi, 1976, page 536) to other bases. As will be seen shortly, the *hypothesis* of base invariance characterizes mixtures of Benford's law and a Dirac probability measure on the special constant 1, which may occur with positive probability.

To motivate the definition of base invariance, consider the set $\{D_1 = 1\}$ of positive numbers with leading significant digit 1 (base 10). This same set of numbers can also [cf. (5)] be written as

$$\{D_1 = 1\} = \bigcup_{n=-\infty}^{\infty} [1, 2) \times 100^n \cup \bigcup_{n=-\infty}^{\infty} [10, 20) \times 100^n,$$

that is, $\{D_1 = 1\}$ is also the set of positive numbers whose leading significant digit (base 100) is in the set $\{1, 10, 11, \ldots, 19\}$. In general, every set of real numbers S (base 10) in $\mathscr{M}$ is *exactly the same set* as the set of real numbers $S^{1/2}$ (base 100) in $\mathscr{M}$. Thus if a probability is base invariant, the measure of any given set of real numbers (in the mantissa σ-algebra $\mathscr{M}$) should be the same for all bases and, in particular, for bases which are powers of the original base. This suggests the following natural definition [recall that $\mathscr{M}$ is also closed under integral roots, property (7)(iii)].

DEFINITION 2. A probability measure P on $(\mathbb{R}^+, \mathscr{M})$ is *base invariant* if $P(S) = P(S^{1/n})$ for all positive integers n and all $S \in \mathscr{M}$.

Next, observe that the set of numbers

$$\begin{aligned} S_1 &= \{D_1 = 1, D_j = 0 \text{ for all } j > 1\} \\ &= \{\ldots, 0.01, 0.1, 1, 10, 100, \ldots\} \\ &= \bigcup_{n=-\infty}^{\infty} \{1\} \times 10^n \in \mathscr{M} \end{aligned}$$

has [by (6)] no nonempty $\mathscr{M}$-measurable subsets, so the Dirac delta measure δ_1 of this set is well defined. [Here $\delta_1(S) = 1$ if $S \supseteq S_1$ and $= 0$ otherwise, for all $S \in \mathscr{M}$.] Letting P_L denote the logarithmic probability distribution on $(\mathbb{R}^+, \mathscr{M})$ given in (8), a complete characterization for base-invariant significant-digit probability measures can now be given.

THEOREM 2 (Hill, 1995a). *A probability measure P on $(\mathbb{R}^+, \mathscr{M})$ is base invariant if and only if*

$$P = qP_L + (1 - q)\delta_1 \quad \textit{for some } q \in [0, 1].$$

From Theorems 1 and 2 it is easily see that *scale invariance implies base invariance*, but not conversely (e.g., δ_1 is clearly base but not scale invariant).

The proof of Theorem 1 follows easily from the fact that scale invariance corresponds to invariance under irrational rotations $x \to (x + s) \pmod 1$ on the circle, and the unique invariant probability measure under this transformation is well known to be the uniform (Lebesgue) measure, which in turn corresponds to the log mantissa distribution. Proof of Theorem 2 is slightly more complicated, since base invariance corresponds to invariance under multiplication $x \to nx \pmod 1$. The key tool used here (Hill, 1995a, Proposition 4.1) is that a Borel probability Q on $[0, 1)$ is invariant under the mappings $nx \pmod 1$ for all n if and only if Q is a convex combination of uniform measure and point mass at 0. [A number of basic questions concerning invariance under multiplication are still open, such as Furstenberg's 25-year-old conjecture that the uniform distribution on $[0, 1)$ is the only atomless probability distribution invariant under both $2x \pmod 1$ and $3x \pmod 1$.]

RANDOM SAMPLES FROM RANDOM DISTRIBUTIONS

Theorems 1 and 2 may be clean mathematically, but they hardly help explain the appearance of Benford's law empirically. What do 1990 census populations of U.S. counties have in common with 1880 users of logarithm tables, numerical data from front-page newspaper articles of the 1930s collected by Benford or universal physical constants examined by Knuth in the 1960s? Why should these tables be logarithmic or, equivalently, scale or base invariant? Many tables are not of this form, including even Benford's individual tables (as he noted), but as Raimi (1969) pointed out, "what came closest of all, however, was the union of all his tables." Combine the molecular weight tables with baseball statistics and areas of rivers, and *then* there is a good fit. Many of the previous explanations of Benford's law have hypothesized some universal table of constants, Raimi's (1985, page 217) "stock of tabular data in the world's libraries" or Knuth's (1969) "some imagined set of real numbers," and tried to prove why certain specific sets of real observations were representative of either this mystical universal table or the set of all real numbers.

What seems more natural is to think of data as coming from *many different distributions*, as was clearly the case in Benford's (1938) study in his "effort to collect data from as many fields as possible and to include a wide variety of types" (page 552); "the range of subjects studied and tabulated was as wide as time and energy permitted" (page 554).

Recall that a (real Borel) *random probability measure* (r.p.m.) $\mathbb{M}$ is a random vector [on an underlying probability space $(\Omega, \mathscr{F}, \mathbf{P})$] taking values which are Borel probability measures on $\mathbb{R}$ and which is regular in the sense that for each Borel set $B \subset \mathbb{R}$, $\mathbb{M}(B)$ is a random variable (cf. Kallenberg, 1983).

DEFINITION 3. The *expected distribution measure* of a r.p.m. $\mathbb{M}$ is the probability measure $\mathbf{E}\mathbb{M}$ (on the Borel subsets of $\mathbb{R}$) defined by

$$(9) \quad (\mathbf{E}\mathbb{M})(B) = E(\mathbb{M}(B)) \quad \text{for all Borel } B \subset \mathbb{R}$$

[where here and throughout, $E(\cdot)$ denotes expectation with respect to $\mathbf{P}$ on the underlying probability space].

For example, if $\mathbb{M}$ is a random probability which is $U[0, 1]$ with probability 1/2 and otherwise is an exponential distribution with mean 1, then $\mathbf{E}\mathbb{M}$ is simply the continuous distribution with density $f(x) = (1 + e^{-x})/2$ for $0 \le x \le 1$ and $= e^{-x}/2$ for $x > 1$.

The next definition plays a central role in this section and formalizes the concept of the following natural process which mimics Benford's data-collection procedure: pick a distribution at random and take a sample of size k from this distribution; then pick a second distribution at random and take a sample of size k from this second distribution and so forth.

DEFINITION 4. For an r.p.m. $\mathbb{M}$ and positive integer k, a *sequence of* $\mathbb{M}$*-random* k*-samples* is a sequence of random variables $X_1, X_2, \ldots$ on $(\Omega, \mathcal{F}, \mathbf{P})$ so that for some i.i.d. sequence $\mathbb{M}_1, \mathbb{M}_2, \mathbb{M}_3, \ldots$ of r.p.m.'s with the same distribution as $\mathbb{M}$ and for each $j = 1, 2, \ldots,$

(10) given $\mathbb{M}_j = P$, the random variables $X_{(j-1)k+1}, \ldots, X_{jk}$ are i.i.d. with d.f. P;

and

(11) $X_{(j-1)k+1}, \ldots, X_{jk}$ are independent of $\{\mathbb{M}_i, X_{(i-1)k+1}, \ldots, X_{ik}\}$ for all $i \ne j$.

The following lemma shows the somewhat curious structure of such sequences.

LEMMA 1. *Let* $X_1, X_2, \ldots$ *be a sequence of* $\mathbb{M}$*-random* k*-samples for some* k *and some r.p.m.* $\mathbb{M}$. *Then*:

(i) *the* $\{X_n\}$ *are a.s. identically distributed with distribution* $\mathbf{E}\mathbb{M}$, *but are not in general independent*;

(ii) *given* $\{\mathbb{M}_1, \mathbb{M}_2, \ldots\}$, *the* $\{X_n\}$ *are a.s. independent, but are not in general identically distributed.*

PROOF. The first part of (ii) follows easily by (10) and (11); the second part follows since whenever $\mathbb{M}_i \ne \mathbb{M}_j$, X_{ik} will not have the same distribution as X_{jk}. The first part of (i) follows by conditioning on $\mathbb{M}_j$:

$$\begin{aligned}\mathbf{P}(X_j \in B) &= E[\mathbb{M}_j(B)] \\ &= E[\mathbb{M}(B)] \quad \text{for all Borel } B \subset \mathbb{R},\end{aligned}$$

where the last equality follows since $\mathbb{M}_j$ has the same distribution as $\mathbb{M}$. The second part of (i) follows from the fact that i.i.d. samples from a distribution may give information about the distribution, as seen in the next example. □

In general, sequences of $\mathbb{M}$-random k-samples are not independent, not exchangeable, not Markov, not martingale and not stationary sequences.

EXAMPLE. Let $\mathbb{M}$ be a random measure which is the Dirac probability measure $\delta(1)$ at 1 with probability 1/2, and which is $(\delta(1) + \delta(2))/2$ otherwise, and let $k = 3$. Then $\mathbf{P}(X_2 = 2) = 1/4$, but $\mathbf{P}(X_2 = 2 \mid X_1 = 2) = 1/2$, so X_1, X_2 are not independent. Since

$$\begin{aligned}&\mathbf{P}((X_1, X_2, X_3, X_4) = (1, 1, 1, 2)) \\ &\quad = 9/64 > 3/64 = \mathbf{P}((X_1, X_2, X_3, X_4) \\ &\quad = (2, 1, 1, 1)),\end{aligned}$$

the $\{X_n\}$ are not exchangeable; since

$$\begin{aligned}&\mathbf{P}(X_3 = 1 \mid X_1 = X_2 = 1) \\ &\quad = 9/10 > 5/6 = \mathbf{P}(X_3 = 1 \mid X_2 = 1),\end{aligned}$$

the $\{X_n\}$ are not Markov; since

$$E(X_2 \mid X_1 = 2) = 3/2,$$

the $\{X_n\}$ are not a martingale; and since

$$\begin{aligned}&\mathbf{P}((X_1, X_2, X_3) = (1, 1, 1)) \\ &\quad = 9/16 > 15/32 = \mathbf{P}((X_2, X_3, X_4) = (1, 1, 1)),\end{aligned}$$

the $\{X_n\}$ are not stationary.

The next lemma is simply the statement of the intuitively plausible fact that the empirical distribution of $\mathbb{M}$-random k-samples converges to the expected distribution of $\mathbb{M}$; that this is not completely trivial follows from the independence–identically distributed dichotomy stated in Lemma 1. If $k = 1$, it is just the Bernoulli case of the strong law of large numbers.

LEMMA 2. *Let* $\mathbb{M}$ *be a r.p.m., and let* $X_1, X_2 \ldots$ *be a sequence of* $\mathbb{M}$*-random* k*-samples for some* k. *Then*

$$\begin{aligned}&\lim_{n\to\infty} \frac{\#\{i \le n: X_i \in B\}}{n} \\ &\quad = E[\mathbb{M}(B)] \quad \textit{a.s. for all Borel } B \subset \mathbb{R}.\end{aligned}$$

PROOF. Fix B and $j \in \mathbb{N}$, and let

$$Y_j = \#\{m, 1 \le m \le k: X_{(j-1)k+m} \in B\}.$$

Clearly,

$$\lim_{n\to\infty} \frac{\#\{i \le n: X_i \in B\}}{n} = \lim_{n\to\infty} \frac{\sum_{j=1}^{m} Y_j}{km} \tag{12}$$

(if the limit exists)

By (10), given $\mathbb{M}_j$, Y_j is binomially distributed with parameters k and $E[\mathbb{M}_j(B)]$, so by (9),

$$\begin{aligned} EY_j &= E(E(Y_j \mid \mathbb{M}_j)) \\ &= kE[\mathbb{M}(B)] \quad \text{a.s. for all } j, \end{aligned} \tag{13}$$

since $\mathbb{M}_j$ has the same distribution as $\mathbb{M}$.

By (11), the $\{Y_j\}$ are independent. Since they have [via (13)] identical means $kE[\mathbb{M}(B)]$ and are uniformly bounded [so $\sum^{\infty}(\mathrm{Var}(Y_j)/j^2) < \infty$], it follows (cf. Loève, 1977, page 250) that

$$\lim_{m\to\infty} \frac{\sum_{j=1}^{m} Y_j}{m} = kE[\mathbb{M}(B)] \quad \text{a.s.}, \tag{14}$$

and the conclusion follows by (12) and (14). □

An even shorter proof can be based on the observation that the variables X_i, X_{k+i}, $X_{2k+i}, \ldots$ are i.i.d. for all $1 \le i \le k$, but the argument given above can be easily modified to show that the assumption that each $\mathbb{M}_j$ is sampled exactly k times is not essential; if the jth r.p.m. is sampled K_j times, where the $\{K_j\}$ are independent uniformly bounded $\mathbb{N}$-valued random variables (which are also independent of the rest of the process), then the same conclusion holds.

A NEW STATISTICAL DERIVATION

The stage is now set to give a new statistical limit law (Theorem 3 below) which is a central-limit-like theorem for significant digits. Roughly speaking, this law says that if probability distributions are selected at random and random samples are then taken from each of these distributions in any way so that the overall process is scale (or base) neutral, then the significant-digit frequencies of the combined sample will converge to the logarithmic distribution. This theorem helps explain and predict the appearance of the logarithmic distribution in significant digits of tabulated data.

DEFINITION 5. A sequence of random variables $X_1, X_2, \ldots$ has *scale-neutral mantissa frequency* if

$$n^{-1}|\#\{i \le n: X_i \in S\} - \#\{i \le n: X_i \in sS\}| \to 0 \quad \text{a.s.}$$

for all $s > 0$ and all $S \in \mathcal{M}$, and has *base-neutral mantissa frequency* if

$$n^{-1}|\#\{i \le n: X_i \in S\} - \#\{i \le n : X_i \in S^{1/m}\}| \to 0 \quad \text{a.s.}$$

for all $m \in \mathbb{N}$ and $S \in \mathcal{M}$.

For example, if $\{X_n\}$, $\{Y_n\}$, and $\{Z_n\}$ are the sequences of (constant) random variables defined by $X_n \equiv 1$, $Y_n \equiv 2$ and $Z_n = 2^n$, then $\{X_n\}$ has base- but not scale-neutral mantissa frequency, $\{Y_n\}$ has neither and (by Theorem 1 above and Theorem 1 of Diaconis, 1977) $\{Z_n\}$ has both.

Mathematical examples of scale-neutral and scale-biased processes are easy to construct, as will be described below. For a real-life example, pick a beverage-producing company in continental Europe at random and look at the metric volumes of a sample of k of its products; then pick a second company and so forth. Since product volumes in this case are probably closely related to liters, this (random k-sample) process is most likely *not* scale neutral and conversion to another unit such as gallons would probably yield a radically different set of first-digit frequencies. On the other hand, if species of mammals in Europe are selected at random and their metric volumes sampled, it seems less likely that this second process is related to the choice of units.

Similarly, base-neutral and base-biased processes are also easy to construct mathematically. The question of base-neutrality is most interesting when the units in question are universally agreed upon, such as the numbers of things. For real-life examples, picking cities at random and looking at the number of fingers of k-samples of people from those cities is certainly heavily base-10 dependent (that is where base 10 originated), whereas picking cities at random and looking at the number of leaves of k-samples of trees from those cities is probably less base dependent. As will be seen in the next theorem, scale and base neutrality of random k-samples are essentially equivalent to scale and base *unbiasedness* of the underlying r.p.m. $\mathbb{M}$.

DEFINITION 6. An r.p.m. $\mathbb{M}$ is *scale unbiased* if its expected distribution $\mathbf{E}\mathbb{M}$ is scale invariant on $(\mathbb{R}^+, \mathcal{M})$ and is *base unbiased* if $\mathbf{E}\mathbb{M}$ is base invariant on $(\mathbb{R}^+, \mathcal{M})$. [Recall that $\mathcal{M}$ is a sub-σ-algebra of the Borels, so every Borel probability on $\mathbb{R}$ (such as $\mathbf{E}\mathbb{M}$) induces a unique probability on $(\mathbb{R}^+, \mathcal{M})$.]

A crucial point here is that the definition of scale and base unbiased does not require that *individual* realizations of $\mathbb{M}$ be scale or base invariant; in fact it is often the case [see Benford's (1938) data and example below] that *none* of the realizations is scale invariant, but only that the sampling process *on the average* does not favor one scale over another.

Now for the main new statistical result: here $\mathbb{M}(t)$ denotes the random variable $\mathbb{M}(D_t)$, where $D_t = \bigcup_{n=-\infty}^{\infty}[1, t)\times 10^n$ is the set of positive numbers with mantissae in $[1/10, t/10)$. [Thus in light of the representation (6), $\mathbb{M}(t)$ may be viewed as the random

cumulative distribution function for the mantissae of the r.p.m. $\mathbb{M}$.]

THEOREM 3 (Log-limit law for significant digits). *Let $\mathbb{M}$ be an r.p.m. on $(\mathbb{R}^+, \mathscr{M})$. The following are equivalent*:

(i) $\mathbb{M}$ *is scale unbiased*;
(ii) $\mathbb{M}$ *is base unbiased and* $\mathbf{E}\mathbb{M}$ *is atomless*;
(iii) $E[\mathbb{M}(t)] = \log_{10} t$ *for all* $t \in [1, 10)$;
(iv) *every* $\mathbb{M}$*-random k-sample has scale-neutral mantissa frequency*;
(v) $\mathbf{E}\mathbb{M}$ *is atomless, and every* $\mathbb{M}$*-random k-sample has base-neutral mantissa frequency*;
(vi) *for every* $\mathbb{M}$*-random k-sample* $X_1, X_2, \ldots,$

$$n^{-1}\#\{i \le n: \mathit{mantissa}(X_i) \in [1/10, t/10)\} \to \log_{10} t \text{ a.s. for all } t \in [1, 10).$$

PROOF. (i) $\Leftrightarrow$ (iii). Immediate by Definitions 1 and 6 and Theorem 1.

(ii) $\Leftrightarrow$ (iii). It follows easily from (6) that the Borel probability $\mathbf{E}\mathbb{M}$ is atomless if and only if it is atomless on $\mathscr{M}$. That (ii) is equivalent to (iii) then follows easily by Definitions 2 and 6 and Theorem 2.

(iii) $\Leftrightarrow$ (iv). By Lemma 2,

$$A_n := n^{-1}|\#\{i \le n: X_i \in S\}| \to E[\mathbb{M}(S)] \quad \text{a.s.},$$

and

$$B_n := n^{-1}|\#\{i \le n: X_i \in sS\}| \to E[\mathbb{M}(sS)] \quad \text{a.s.},$$

so $|A_n - B_n| \to 0$ a.s. if and only if $\mathbf{E}\mathbb{M}(S) = \mathbf{E}\mathbb{M}(sS)$, which by Definition 1 and Theorem 1 is equivalent to (iii).

(iii) $\Leftrightarrow$ (v). Similar, using Lemma 2, Definition 2 and Theorem 2.

(iii) $\Leftrightarrow$ (vi). Immediate by Lemma 2. □

One of the points of Theorem 3 is that there are *many* (natural) sampling procedures which lead to the log distribution, helping explain how the different empirical evidence of Newcomb, Benford, Knuth and Nigrini all led to the same law. This may also help explain why sampling the numbers from newspaper front pages (Benford, 1938, page 556), or almanacs or extensive accounting data often tends toward the log distribution, since in each of these cases various distributions are being sampled in a presumably unbiased way. Perhaps the first article in the newspaper has statistics about population growth, the second article about stock prices and the third about forest acreage. None of these individual distributions itself may be unbiased, but the *mixture* may well be.

Justification of the hypothesis of scale or base unbiasedness is akin to justification of the hypothesis of independence (and identical distribution) in applying the strong law of large numbers or central limit theorem to real-life processes: neither hypothesis can be *proved*, yet in many real-life sampling procedures, they appear to be reasonable assumptions. Conversely, Theorem 3 suggests a straightforward test for unbiasedness of data—simply test goodness-of-fit to the logarithmic distribution.

Many standard constructions of r.p.m.'s are automatically scale and base neutral, and thus satisfy the log-limit significant-digit law. Consider the problem of generating a random variable X (or r.p.m.) on $[1, 10)$. If the units chosen are desired to be just as likely stock per dollars as dollars per stock [or Benford's (1938) "candles per watt" versus "watts per candle"], then the distribution generated should be reciprocal invariant, so for example its $\log_{10}$ should be symmetric about 1/2. So first set $F(1) = 0$ and $F(10^-) = 1$; next pick $F(\sqrt{10})$ randomly [according to, say, uniform measure on $(0, 1)$] since $\sqrt{10}$ is the reciprocal-invariant point $t = 10/t$; then pick $F(10^{1/4})$ and $F(10^{3/4})$, independently and uniformly on $(0, F(\sqrt{10}))$ and $(F(\sqrt{10}), 1)$, respectively, and continue in this manner. This classical construction of Dubins and Freedman (1967, Lemma 9.28) is known to generate an r.p.m. a.s. whose expected distribution $\mathbf{E}\mathbb{M}$ is the logarithmic probability P_L of (8), and hence by Theorem 3 is scale and base unbiased, even though *with probability* 1 *every distribution generated this way will be both scale and base biased*. On the average, this r.p.m. is unbiased, so the log-limit significant-digit law will apply to all $\mathbb{M}$-random k-samples. [The construction described above using uniform measure is not crucial. Any base measure on $(0, 1)$ symmetric about 1/2 will have the same property (Dubins and Freedman, 1967, Theorem 9.29).]

Also, many significant-digit data sets *other* than random k-samples have scale- or base-neutral mantissa frequency, in which case combining such data together with unbiased random k-samples (as did Benford, perhaps, in combining data from mathematical tables with that from newspaper statistics) will still result in convergence to the logarithmic distribution. For example, if certain data represents (deterministic) periodic sampling of a geometric process (e.g., $X_n = 2^n$), then by Theorem 1 of Diaconis (1977), this deterministic process is a strong Benford sequence, which implies that its limiting frequency (separately or averaged with unbiased random k-samples) will satisfy (4).

An interesting open problem is to determine which common distributions (or mixtures thereof)

satisfy Benford's law, that is, are scale or base invariant or which have mantissas with logarithmic distributions. For example, the standard Cauchy distribution is close to satisfying Benford's law (cf. Raimi, 1976) and the standard Gaussian is not, but perhaps certain natural mixtures of some common distributions are.

Of course there are many r.p.m.'s and sampling processes which do *not* satisfy the log-limit law (and hence are necessarily both scale and base biased), such as the (a.s.) constant uniform distribution on $[1, 10)$ or (for some reason not yet well understood by the author) the r.p.m. constructed via Dubins–Freedman with base probability uniform measure on the *horizontal* bisector of the rectangle, which has expected log distribution a renormalized arcsin distribution (Dubins and Freedman, 1967, Theorem 9.21).

APPLICATIONS

The statistical log-limit significant-digit law Theorem 3 may help justify some of the recent applications of Benford's law, several of which will now be described.

In scientific calculating, if the distribution of input data into a central processing station is known, then this information can be used to design a computer which is optimal (in any of a number of ways) *with respect to that distribution*. Thus if the computer users are like the log-table users of Newcomb or the taxpayers of Nigrini's study, their data represent an unbiased (as to units, base, reciprocity, ...) random mixture of various distributions, in which case it will (by Theorem 3) necessarily follow Benford's law. Once a specific input distribution has been identified, in this case the logarithmic distribution, then that information can be exploited to improve computer design. Feldstein and Turner (1986) show that

> under the assumption of the logarithmic distribution of numbers, floating-point addition and subtraction can result in overflow or underflow with alarming frequency... and lead to the suggestion of a long word format which will reduce the risks to acceptable levels.

Schatte (1988) concludes that under assumption of logarithmic input, base $b = 2^3$ is optimal with respect to minimizing storage space. Knuth (1969) after having "established the logarithmic law for integers by direct calculation," leaves as an exercise (page 228) determining the desirability of hexadecimal versus binary with respect to different objectives. Barlow and Bareiss (1985)

> conclude that the logarithmic computer has smaller error confidence intervals for roundoff errors than a floating point computer with the same computer word size and approximately the same number range.

A second modern application of Benford's law is to mathematical modelling, where goodness-of-fit against the logarithmic distribution has been suggested (cf. Varian, 1972) as a *test of reasonableness* of output of a proposed model, a sort of "Benford-in–Benford-out" criteria. In Nigrini and Wood's (1995) census tabulations, for example, the 1990 census populations of the counties in the United States follow the significant-digit logarithmic law very closely, so it seems reasonable that mathematical models for *predicting* future populations of the counties should also be a close fit to Benford. If not, perhaps a different model should be considered.

As one final example, Nigrini has amassed a vast collection of U.S. tax and accounting data including 91,022 observations of IRS-reported interest income (Nigrini, 1996), and share volumes (at the rate of 200–350 million per day) on the New York Stock Exchange (Nigrini, 1995), and in most of these cases the logarithmic distribution is an excellent fit (perhaps exactly because each is an unbiased mixture of data from different distributions). He postulates that Benford is often a reasonable distribution to expect for the significant digits of large accounting data sets and has proposed a *goodness-of-fit test against Benford to detect fraud.* In an article in the *Wall Street Journal* in July 1995 (Berton, 1995) it was announced that the District Attorney's office in Brooklyn, New York, using Nigrini's Benford goodness-of-fit tests, has detected and charged groups at seven New York companies with fraud. The Dutch IRS has expressed interest in using this Benford test to detect income tax fraud, and Nigrini has submitted proposals to the U.S. IRS.

ACKNOWLEDGMENTS

The author is grateful to the Free University of Amsterdam and especially Professor Piet Holewijn for their support and hospitality during the summer of 1995, and also is grateful to Pieter Allaart, David Gilat, Ralph Raimi and Peter Schatte for a number of suggestions, to Klaas van Harn for several corrections and valuable advice concerning notation and

to an anonymous Associate Editor for excellent ideas for improving the exposition. This research was partially supported by NSF Grant DMS-95-03375.

REFERENCES

ADHIKARI, A. and SARKAR, B. (1968). Distribution of most significant digit in certain functions whose arguments are random variables. *Sankhyā Ser. B* **30** 47–58.

BARLOW, J. and BAREISS, E. (1985). On roundoff error distributions in floating point and logarithmic arithmetic. *Computing* **34** 325–347.

BECKER, P. (1982). Patterns in listings of failure-rate and MTTF values and listings of other data. *IEEE Transactions on Reliability* **R-31** 132–134.

BENFORD, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society* **78** 551–572.

BERTON, L. (1995). He's got their number: scholar uses math to foil financial fraud. *Wall Street Journal*, July 10.

BUCK, B., MERCHANT, A. and PEREZ, S. (1993). An illustration of Benford's first digit law using alpha decay half lives. *European J. Phys.* **14** 59–63.

BURKE, J. and KINCANON, E. (1991). Benford's law and physical constants: the distribution of initial digits. *Amer. J. Phys.* **59** 952.

COHEN, D. (1976). An explanation of the first digit phenomenon. *J. Combin. Theory Ser. A* **20** 367–370.

COHEN, D. and KATZ, T. (1984). Prime numbers and the first digit phenomenon. *J. Number Theory* **18** 261–268.

DE FINETTI, B. (1972). *Probability, Induction and Statistics*. Wiley, New York.

DIACONIS, P. (1977). The distribution of leading digits and uniform distribution mod 1. *Ann. Probab.* **5** 72–81.

DIACONIS, P. and FREEDMAN, D. (1979). On rounding percentages. *J. Amer. Statist. Assoc.* **74** 359–364.

DUBINS, L. and FREEDMAN, D. (1967). Random distribution functions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* 183–214. Univ. California Press, Berkeley.

FELDSTEIN, A. and TURNER, P. (1986). Overflow, underflow, and severe loss of significance in floating-point addition and subtraction. *IMA J. Numer. Anal.* **6** 241–251.

FLEHINGER, B. (1966). On the probability that a random number has initial digit *A*. *Amer. Math. Monthly* **73** 1056–1061.

HAMMING, R. (1970). On the distribution of numbers. *Bell System Technical Journal* **49** 1609–1625.

HILL, T. (1995a). Base-invariance implies Benford's law. *Proc. Amer. Math. Soc.* **123** 887–895.

HILL, T. (1995b). The significant-digit phenomenon. *Amer. Math. Monthly* **102** 322–327.

JECH, T. (1992). The logarithmic distribution of leading digits and finitely additive measures. *Discrete Math.* **108** 53–57.

KALLENBERG, O. (1983). *Random Measures*. Academic Press, New York.

KNUTH, D. (1969). *The Art of Computer Programming* **2** 219–229. Addison-Wesley, Reading, MA.

LEY, E. (1995). On the peculiar distribution of the U.S. stock indices digits. *Amer. Statist.* To appear.

LOÈVE, M. (1977). *Probability Theory* **1**, 4th ed. Springer, New York.

NEWCOMB, S. (1881). Note on the frequency of use of the different digits in natural numbers. *Amer. J. Math.* **4** 39–40.

NIGRINI, M. (1995). Private communication.

NIGRINI, M. J. (1996). A taxpayer compliance application of Benford's law. *Journal of the American Taxation Association* **18** 72–91.

NIGRINI, M. and WOOD, W. (1995). Assessing the integrity of tabulated demographic data. Preprint, Univ. Cincinnati and St. Mary's Univ.

RAIMI, R. (1969). The peculiar distribution of first digits. *Scientific American* December 109–119.

RAIMI, R. (1976). The first digit problem. *Amer. Math. Monthly* **102** 322–327.

RAIMI, R. (1985). The first digit phenomenon again. *Proceedings of the American Philosophical Society* **129** 211–219.

SCHATTE, P. (1988). On mantissa distributions in computing and Benford's law. *J. Inform. Process. Cybernet.* **24** 443–455.

VARIAN, H. (1972). Benford's law. *Amer. Statist.* **23** 65–66.

WEAVER, W. (1963). *Lady Luck: The Theory of Probability* 270–277. Doubleday, New York.