BENFORD'S LAW

Charles G. Wohl*
Lawrence Berkeley National Laboratory, Berkeley, CA 94720

(Submitted to Physics Today)

Many sets of numbers are alike in this respect: About 30% of the numbers begin with a 1; about 18% begin with a 2; and so on down, until only about 4.6% of the numbers begin with a 9. The first digit is as likely to be a 1 as it is to be a 5, 6, 7, 8, or 9 altogether. The numbers have always been in plain sight, but we don't see the forest for the trees.

> *"That the ten digits do not occur with equal frequency must be evident to any one making much use of logarithmic tables, and noticing how much faster the first pages wear out than the last ones."*
>
> Simon Newcomb, 1881[1]

> *"It has been observed that the first pages of a table of common logarithms show more wear than do the last pages . . . "*
>
> Frank Benford, 1938[2]

> *"The fact that leading digits tend to be small is important to keep in mind; it makes the most obvious techniques of 'average error' estimation for floating-point calculations invalid."*
>
> Donald E. Knuth, 1969[3]

> *"The income tax agencies of several nations and several states, including California, are using detection software based on Benford's Law, as are a score of large companies and accounting businesses."*
>
> *The New York Times*, 1998[4]

Figure 1 shows the distributions of first digits for eight sets of numbers. For example, the Fibonacci numbers F_1 through F_{10} are

$$1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ 55\ ;$$

and thus in Fig. 1(a) these contribute three counts to the 1s bin, two counts to each of the 2s, 3s, and 5s bins, and one count to the 8s bin. In each set in Fig. 1, about 30% of the numbers begin with a 1 and about 5% begin with a 9. Each set follows a first-digits law called Benford's law, shown as a solid histogram in each panel. We discuss the examples after introducing this law.

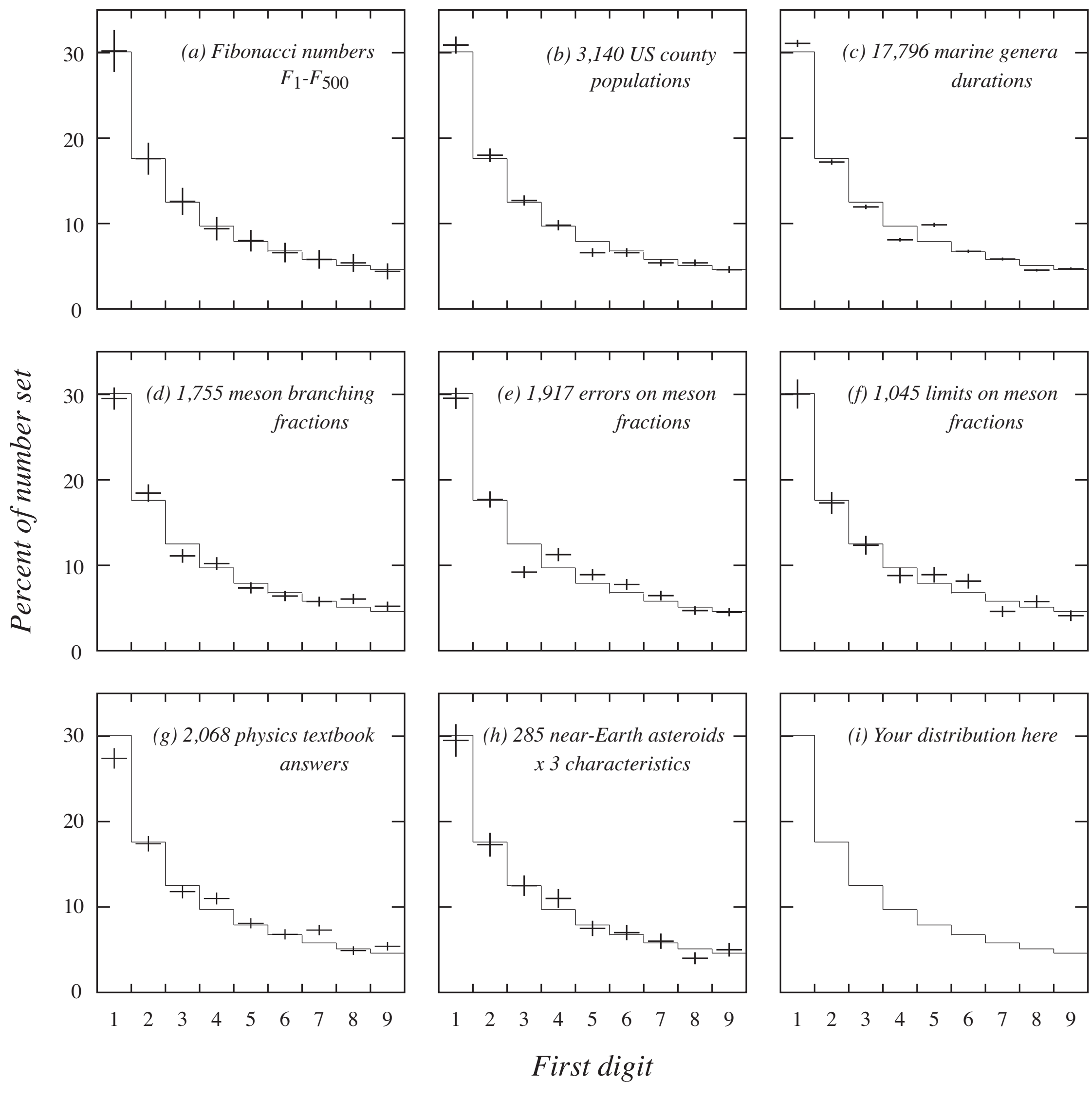

Figure 1. A comparison of eight distributions of first digits with Benford's law. The error bars are plus or minus the square roots of counts per bin.

Benford's law

The first quotation at the beginning of this article is the first sentence of an 1881 paper, "Note on the Frequency of Use of the Different Digits in Natural Numbers,"[1] by Simon Newcomb, a mathematician and astronomer.[5] The obvious inference from the worn early pages of tables of logarithms (an observation unlikely in the era of calculators) is that numbers that begin with a 1 or 2 occur more frequently in calculations than do those that begin with an 8 or 9. After a brief argument (the paper is two short pages, and the argument is neither obvious nor conclusive), Newcomb concludes:

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

Here is what this means: We write (positive) numbers in the scientific form $i.jk\ldots\times 10^n$, where the first digit i is one of 1 through 9. Then

$$i.jk\ldots \times 10^n = 10^m \times 10^n \ , \tag{1}$$

where m is the mantissa, and $0.0 \le m < 1.0$. ("Mantissa" has a different meaning in the context of floating-point notation.) Then

$$\log_{10}(i.jk\ldots \times 10^n) = m + n = n.m \ . \tag{2}$$

Figure 2 shows mantissas, plotted linearly, and the corresponding numbers $i.jk\ldots$ Older readers will be reminded of the C and D scales on slide rules. Newcomb's claim is that the distribution of $i.jk\ldots$ in many sets of numbers is (statistically) uniform along the line in Fig. 2. More numbers start with $i = 1$ than with $i = 9$ because the segment of the line between $i = 1$ and 2 is longer than the segment between 9 and 10.

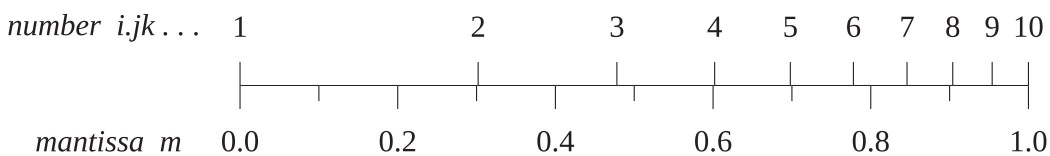

Figure 2. Mantissas, plotted on a linear scale, of the numbers 1 through 10.

According to Newcomb, the probabilities $P(i)$ of the first digits i are

$$\begin{aligned} P(1) &= \log_{10} 2 - \log_{10} 1 &&= \log_{10} 2/1 &&= 0.301 \\ P(2) &= \log_{10} 3 - \log_{10} 2 &&= \log_{10} 3/2 &&= 0.176 \\ &= \quad \cdots \\ P(9) &= \log_{10} 10 - \log_{10} 9 &&= \log_{10} 10/9 &&= 0.046 \ . \end{aligned} \tag{3}$$

These are properly normalized, because the sum of the $P(i)$ is $\log_{10} 10 - \log_{10} 1 = 1$. The full set of first-digit probabilities is:

$$\begin{array}{lccccccccccc} i & = & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ P(i)\ \% & = & 30.1 & 17.6 & 12.5 & 9.7 & 7.9 & 6.7 & 5.8 & 5.1 & 4.6\ . \end{array} \tag{4}$$

The first digit is equally likely to be in the intervals 1-to-2, 2-to-4, 3-to-6, 4-to-8, and 5-to-10.

This is the first-digit law. It is not called Newcomb's law because his paper seems to have gone largely unnoticed. Or perhaps his observation simply slipped into the folklore of numbers, because a 1938 paper, "The Law of Anomalous Numbers,"[2] by Frank Benford, a physicist at General Electric, begins, without attribution, with the second quotation at the beginning of this article. And Benford came to the same conclusion about first-digit probabilities. In addition, however, Benford collected more than 20,000 numbers of various sorts—populations, areas of rivers, specific heats, addresses, numbers in newspaper articles, atomic and molecular weights, and other quantities. To varying degrees, the separate distributions gave "experimental" support for the law, but the agreement was particularly good when all the disparate distributions were summed together. Here are Benford's summary figures:

$$\begin{array}{rccccccccc} \text{numbers } (\%) = & 30.6 & 18.5 & 12.4 & 9.4 & 8.0 & 6.4 & 5.1 & 4.9 & 4.7 \\ & \pm 0.8 & \pm 0.4 & \pm 0.4 & \pm 0.3 & \pm 0.2 & \pm 0.2 & \pm 0.2 & \pm 0.2 & \pm 0.3 \\ \text{the law } (\%) = & 30.1 & 17.6 & 12.5 & 9.7 & 7.9 & 6.7 & 5.8 & 5.1 & 4.6\ . \end{array} \tag{5}$$

Benford's paper did get noticed, and for those 20,000 numbers the law with some justice carries his name.[6]

Benford's law is usually stated, as above, in terms of the probabilities of first digits. However, a probability can be given for any interval x_1 to x_2 in $x \equiv i.jk\ldots$, where $1 \leq x_1 < x_2 < 10$. Then

$$P(x_1 < x < x_2) = \log_{10} x_2 - \log_{10} x_1 = \log_{10}(x_2/x_1)\ . \tag{6}$$

Only the ratio x_2/x_1 matters.

The integrated probability $P(x)$ that $x = i.jk\ldots$ lies between its lower limit 1 and X, where $1 < X < 10$, is, from Eq. (6),

$$P(< X) \equiv P(1 \leq x < X) = \log_{10} X\ . \tag{7}$$

The differential probability $dP(x)$ that $x = i.jk\ldots$ lies between x and $x + dx$ is

$$dP(x) = \rho(x)\, dx = \frac{1}{\log_e 10}\frac{dx}{x} = 0.4343\,\frac{dx}{x}\ , \tag{8}$$

where $\rho(x) = 0.4343/x$ is the probability density. The integral of Eq. 8 from 1 to X gives Eq. 7. Figure 3 shows $P(< X)$ versus X and $\rho(x)$ versus x.

When Benford's law applies, numbers between 1.0 and 1.1 are more probable than those between 1.1 and 1.2, and so on. By adding the probabilities for the intervals 1.0–1.1, 2.0–2.1, …, 9.0–9.1, one finds that the probability that the *second* digit, j, is a 0 (at the second digit, 0 joins the other nine) is 12.0%, whereas the probability that it is a 9 is 8.5%: the second-digit probabilities are much more nearly the same than are the first-digit probabilities. The probabilities for the third digits 0 through 9 are all close to 10%.

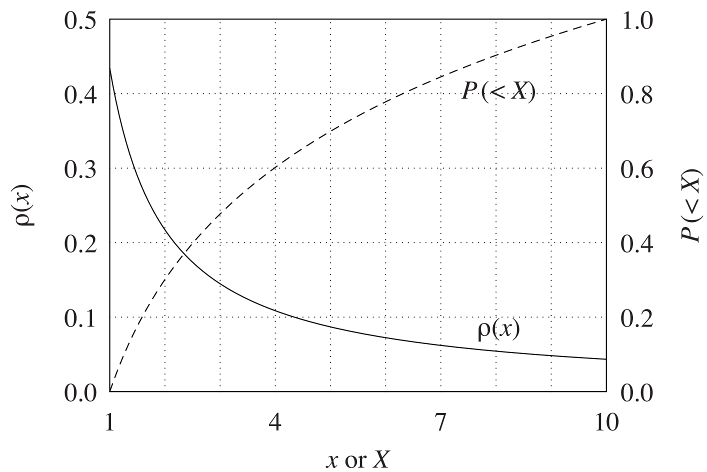

Figure 3. The integrated probability $P(< X)$ versus X (right scale); and the probability density $\rho(x)$ versus x (left scale).

The examples

Distributions of "artificial" numbers such as phone numbers and dates of the month of course do not obey Benford's law, but neither do many distributions of "natural" numbers. A browse through tables in the *Handbook of Chemistry and Physics* yields a disappointingly slim crop of Benford-like distributions. Nevertheless, as the examples in Fig. 1 show, many distributions do agree at least approximately with the law.

Fibonacci numbers—The Fibonacci numbers F_n are defined by $F_0 = 0$, $F_1 = 1$, and thereafter $F_{n+1} = F_n + F_{n-1}$, $n = 1, 2, \ldots$ Figure 1(a) compares the distribution of first digits of F_1 through F_{500} with the Benford distribution.[7] Since the whole set of numbers is determined by the starting values 0 and 1 and a rule, this is not an ordinary statistical distribution. Nevertheless, error bars, the square roots of the numbers of counts N_i, are given to show how nearly perfectly the two distributions agree.

In contrast, the first digits of the prime numbers are much more evenly distributed over 1 through 9.

County populations—Figure 1(b) shows the distribution of first digits of the populations of the 3,140 counties and the like into which the United States is divided.[8] The usual χ^2 sum over bins,

$$\chi^2 = \sum_{i=1}^{9} \left(\frac{N_i - B_i}{\sqrt{N_i}} \right)^2 , \tag{9}$$

where N_i and B_i are the numbers of counts and the Benford predictions, is 11.4, most of it coming from the $i = 5$ bin. The confidence level, with eight degrees of freedom, is $c.l. = 18\%$.

In contrast, the first digits of the areas of the counties are not distributed according to Benford's law.

Genus lifetimes—Figure 1(c) shows the distribution of first digits of 17,796 (!) relatively well-determined durations (in Myr) of marine-animal genera in Earth's history.[9]

(Genus is the grouping above species: *canis* includes dogs, wolves, etc.) The distribution is qualitatively in accord with Benford, but the χ^2 is 146. However, combining the $i = 4$ and 5 bins drops their contribution to the χ^2 from 121 to one. This is too good to be false—some of the "5s" must really be "4s." A breakdown of counts shows that durations between 10 and 100 Myr fall smoothly, but those between 1 and 10 Myr vary erratically, with more in the 5-Myr bin than in any other. An author of the database suggests a possible bias: durations (especially short ones) in much of Earth's history are poorly determined—and five is a "rounder" number than four.[9] Combining the $i = 4$ and 5 bins reduces the χ^2 to 26 ($c.l. = 0.05\%$). It is hard to know whether the data are at fault or the Benford distribution is simply not quite appropriate here.

Meson branching fractions—A meson is a bound state of a quark and an antiquark. The least massive mesons are the π^+, π^0, and π^-, but there are scores of other, more massive mesons. All the mesons are unstable, decaying to lighter particles. More mass means that more decay modes are energetically possible. For example, the π^+ decays 99.99% of the time to $\mu^+\nu_\mu$, but the heaviest known mesons have hundreds of decay modes, for which many but not nearly all the branching fractions have been measured. The Summary Tables of the 2010 *Review of Particle Physics*[10] give 1,755 measured meson branching fractions. Figure 1(d) shows the distribution of first digits of these fractions, compared with Benford's distribution. The χ^2 is 9.0 ($c.l. = 34\%$).

Figure 1(e) shows the distribution of the 1,917 *errors* on the 1,755 measurements of the meson branching fractions (some of the fractions have asymmetric errors, as in $5.7^{+0.3}_{-0.2}$ %). The sweep is clear, but the χ^2 is 35, two-thirds of it coming from the $i = 3$ bin. We return to this example later.

Figure 1(f) shows the distribution of 1,045 experimental *upper limits* on meson branching fractions from the *Review*. A limit measures a sensitivity of an experiment, not a property of a particle. Nevertheless, the χ^2 is 9.6 ($c.l. = 29\%$).

Textbook answers—Figure 1(g) shows the distribution of first digits of 2,068 answers to odd-numbered problems in a physics textbook.[11] The χ^2 is 18.7 ($c.l. = 1.6\%$).

Near-Earth-asteroid characteristics—Figure 1(h) shows the distribution of first digits of three characteristics for each of 285 large near-Earth-orbit asteroids: the estimated diameter (in km), the estimated number of "dynamically distinct potential impacts" with the Earth, and the sum of the probabilities for potential impacts.[12] The χ^2 is 4.3 ($c.l. = 83\%$).

Why Benford? Models

The mathematical literature on Benford's law is substantial. A 1976 paper, whose "purpose is to review all the proposed explanations" of the law, gives 37 references.[13] Nor did that put an end to the matter. Here we shall be content to describe two mathematical models and to mention two other models that lead to Benford's distribution. In the next section, we discuss the invariance properties of the distribution.

Breaking into parts—Figure 1(d) includes all the measured branching fractions of all the established mesons. As noted earlier, light mesons contribute few fractions whereas heavy mesons contribute many. To compare with Fig. 1(d), we make a model in which the branching fractions of a meson with n decay modes result from a purely statistical breakup, like breaking a meter stick at $n-1$ random places. Let the fractions be f_j, with $j=1$ through n, ordered as $f_1 < f_2 < \cdots < f_n$, where $\sum f_j = 1$. The expectation values of the fractions are given in a remarkable expression:[14]

$$\frac{1}{n}\left(\frac{1}{n}+\frac{1}{n-1}+\frac{1}{n-2}+\cdots+\frac{1}{1}\right). \tag{10}$$

The expectation value of f_1 is equal to the first term in the series, $1/n^2$; the expectation value of f_2 is the sum of the first two terms of the series; and so on, until the expectation value of f_n is the sum of all n terms. For $n=2$, $f_1=0.25$ and $f_2=0.75$—a random break in a meter stick is equally likely to be in the central half (giving $f_1>0.25$) or in the end quarters of the stick ($f_1<0.25$). An example with $n=5$ shows the structure of the equations:

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \end{pmatrix} = \begin{pmatrix} 1/5 & 0 & 0 & 0 & 0 \\ 1/5 & 1/4 & 0 & 0 & 0 \\ 1/5 & 1/4 & 1/3 & 0 & 0 \\ 1/5 & 1/4 & 1/3 & 1/2 & 0 \\ 1/5 & 1/4 & 1/3 & 1/2 & 1/1 \end{pmatrix} \begin{pmatrix} 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \end{pmatrix} = \begin{pmatrix} 0.04000 \\ 0.09000 \\ 0.15667 \\ 0.25667 \\ 0.45667 \end{pmatrix}. \tag{11}$$

The sum of the branching fractions is one; in fact, the sum of the elements in every column is one.

Figure 4(a) shows the distribution of first digits of the branching fractions from this statistical model, summed over all the values of n from 2 through 50—altogether $25\times 51-1=1274$ fractions. This is a calculation, not data with errors, but the χ^2 is 2.9. Figure 4(b) shows the modeled distribution for a meson with 500 decay modes. The χ^2 is 3.1.

The expectation values of fractions from a random breaking into parts—either summed over many values of n or for large values of n—obey Benford's law.

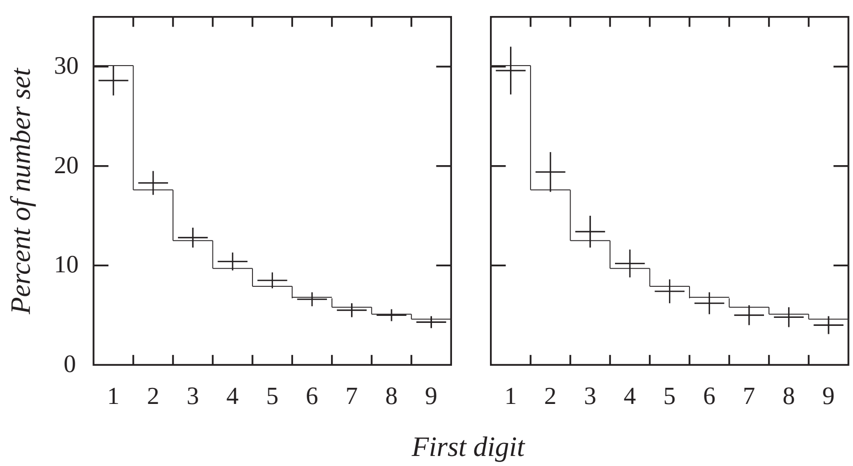

Figure 4. The first-digit distributions for expectation values of fractions resulting from random breaking of a meson or a meter stick: (a) summed over $n=2$ through 50; (b) for $n=500$. See Fig. 1(d).

Sampling an exponential—Let $y(x) = ae^{rx}$ represent the exponential growth ($r > 0$) or decay ($r < 0$) of a quantity y as a function of x; y might be a population of particles or the money in a savings account, and x might be time. Figure 5(a) shows, for a value $r > 0$, a semilog plot of $y(x)$ versus x. The plot is of course a straight line.

The solid horizontal line segment in Fig. 5(a) covers the range of x over which y increases by 900% from 1 to 10. The base of the shaded triangle is the range of x over which y increases by 100% from 1 to 2; the base covers 30.1% of the length of the horizontal line segment. The base of the narrow shaded trapezoid is the range of x over which y increases by 11% from 9 to 10; it covers only 4.6% of the horizontal segment. So if you pick randomly a value of x along the horizontal line segment, the leading digit of y will be 1 or 9 with probabilities 30.1% and 4.6%—just Benford. All this of course repeats as y goes from 10 to 100, 100 to 1000, etc. So if you sample $y(x)$ at random values of x, where the range of x is such that the range of $y(x)$ is some integral number of decades (or—the practical case—the range of $y(x)$ is so many decades that end effects are unimportant), you build up the Benford distribution.

The values of $y(x) = ae^{rx}$, sampled at random values of x (over a large enough range of x), obey Benford's law.

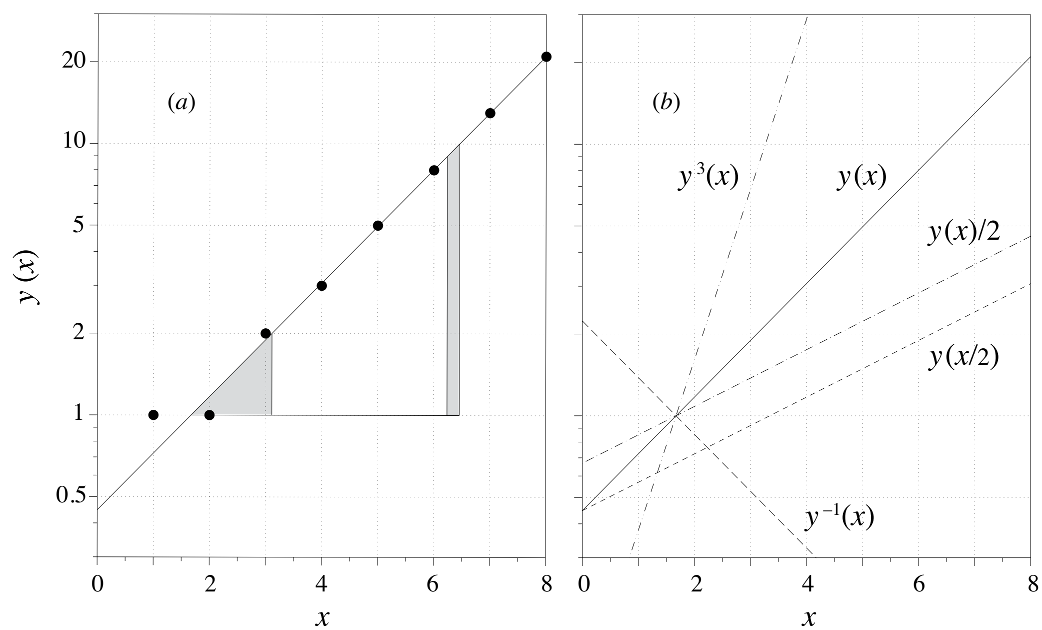

Figure 5. (a) The exponential $y(x) = ae^{rx}$, with $a = 1/\sqrt{5}$ and $e^r = \phi$, the golden ratio. The points are the Fibonacci numbers F_1 through F_{10}. (b) Several transforms of $y(x)$.

Values of $y(x) = ae^{rx}$ at *regular* values of x, say at $x = 0,\ x_1,\ 2x_1,\ \ldots$, are the terms in a geometric series:

$$a\big(1 + e^{rx_1} + e^{2rx_1} + e^{3rx_1} + \cdots\big) = a\big(1 + b + b^2 + b^3 + \cdots\big) \ , \tag{12}$$

where $b \equiv e^{rx_1}$. The distribution of first digits of the terms of a geometric series will obey Benford's law as long as b is not a rational power of 10; but if, say, $b = 10^{7/29}$, then each term in the series repeats, times 10^7, every 29 steps (but even the 29 first digits will approximate the Benford distribution). However, the rational fractions are a "thin" subset of the real numbers; and such mathematical fine points are usually irrelevant for real-world data. (It requires some mathematics to show that the terms fill the mantissa uniformly.)

> *The values of a geometric series, for all but a thin set of values of b, obey Benford's law.*

Example: In Fig. 5(a), $y(x) = ae^{rx}$ is drawn with $a = 1/\sqrt{5}$ and $b = e^r = \phi$, where ϕ is the golden ratio: $\phi = (1+\sqrt{5})/2 = 1.61803\ldots$ The points are the Fibonacci numbers F_n for n (or x) $= 1$ through 10. The line runs through all but the lowest points because the F_n are given by rounding to the nearest integer the terms in a geometric series:[15]

$$\frac{1}{\sqrt{5}}\left(\phi^0 + \phi^1 + \phi^2 + \phi^3 + \cdots\right) . \tag{13}$$

The terms are just the values of $y(x)$ at $x = 0, 1, 2, \ldots$ The first three terms are 0.447, 0.724, and 1.171, poor approximations to 0, 1, and 1; but already $\phi^4/\sqrt{5} = 3.065$ ($F_4 = 3$); and $\phi^{10}/\sqrt{5} = 55.0036$ ($F_{10} = 55$). See Fig. 1(a) again.

Example: The frequencies of the equal-tempered musical scale—12 steps to an octave, each frequency $b = 2^{1/12} = 1.0595\ldots$ times higher than the previous one—are the terms of a geometric series. Thirty percent of the frequencies begin with a one.

Two other models—One of Benford's examples used the first 342 addresses in *American Men of Science*. If all streets are either nine or 99 blocks long, with addresses in the first block being in the 100s, then the first-digit probabilities are all 11%. But any streets shorter than nine or 99 blocks favor the lower first digits. There are many other examples of numberings of finite length, such as of pages and footnotes.

Consider the list of numbers 1 through n. If a number is chosen randomly from the list, what is the probability $P_n(1)$ that it starts with a 1? That of course depends on n: $P_5(1) = 20\%$, $P_{17}(1) = 47\%$, $P_{35}(1) = 31\%$, and so on. The probabilities slowly oscillate as n increases, reaching lows of 11% when $n = 9, 99, 999, \ldots$, and highs of over 50% when $n = 19, 199, \ldots$ Now consider the ensemble of lists having *all* values of n. By repeatedly "smoothing," once can get an ensemble average for $P(1)$ equal to 30.1%,[16] but the process is not unique.[17] Surprisingly, the expression in Eq. 10 makes an appearance in the smoothing.

Another model that produces the Benford distribution—now claimed to be the most fundamental model—involves the selection of random numbers from random distributions.[18] To a greater or lesser extent, several of the distributions in Fig. 1 fall into this class: the county populations; the meson branching fractions; certainly the textbook answers.

Why Benford? Invariances

A wide-ranging law of numbers can scarcely depend on having used British units, or on having ten fingers. Mathematicians have shown that invariance of a distribution under a change of scale or of base leads to Benford's law. (These invariances were already recognized by Newcomb.) Here we merely discuss the invariance properties of a set of numbers that obeys Benford's law.

Figure 5(b) shows the same $y(x)$ as in Fig. 5(a). It also shows:

$$y^{-1}(x) = \frac{1}{a}\, e^{-rx} \; ; \quad \frac{1}{2}\, y(x) = \frac{a}{2}\, e^{rx} \; ; \quad y(x/2) = a e^{rx/2} \; ; \quad y^3 = a^3 e^{3rx} \; .$$

The first of these is the inverse of y; the second and third are y and then x rescaled in units twice as large as before; the fourth is y cubed. All four of these new functions are exponentials—they all graph as straight lines—and therefore all four will, if sampled regularly or randomly along x, give Benford's distribution. At least as arising from an exponential, the distribution is *invariant* under all these transformations of $y(x)$.

String theory—Here are general proofs of the invariance of the Benford distribution under the inversion of numbers and under a change of scale.

Mark the values of the numbers of a set along a long string using a logarithmic scale; see Fig. 6(a). Fibonacci numbers (a few are shown as examples), populations, and the like, would occupy the right (≥ 1) half of the string. Branching fractions and the like would occupy the left half. Another set might scatter along the whole string. Wrap the marked string, with 1 at the bottom, around a circle whose circumference equals the length of the string between 1 and 10; see Fig. 6(b). If the marks on the string populate the circumference of the circle uniformly, the set of numbers obeys Benford's law.

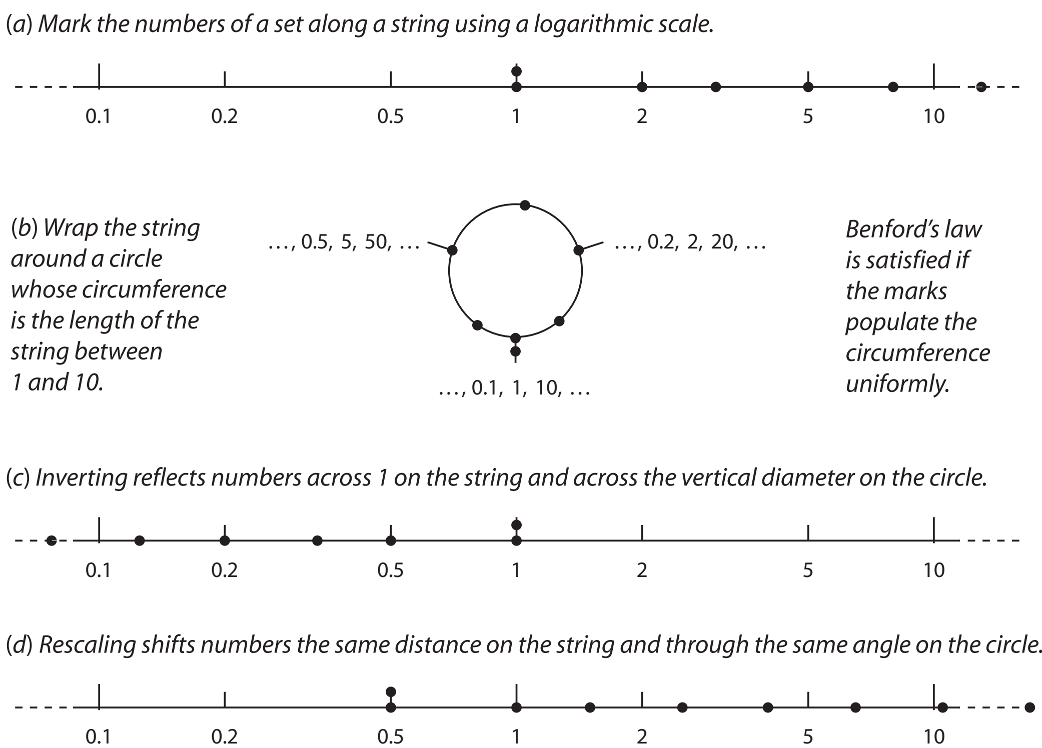

Figure 6. Proofs of inversion and scale invariance of a distribution that obeys Benford's law are reduced to symmetry operations on a circle. The points, for illustration, are a few of the Fibonacci numbers.

Now *invert* all the numbers of the set. A number 2 jumps to 0.5; a number 0.25 jumps to 4, and so on; see Fig. 6(c). In each case, a number and its inverse are equidistant from the center of the string (at 1). Thus an *inversion* of the whole set of numbers just *reflects* the set about the center of the string; and wrapping the string around the circle just populates its circumference the other way around from before. The original and inverted distributions on the circle are related by a reflection across its vertical diameter. If the original distribution of numbers populates the circle uniformly, then so does the inverted distribution.

Now, instead, *rescale* the numbers of the set—perhaps changing from British to SI units. To be concrete, suppose we multiply each of the numbers of a set by $s = 0.5$, which is the same as measuring whatever it is the numbers represent in units twice as large as before. A number 1 goes to 0.5; a number 2 goes to 1; etc; see Fig. 6(d). Each multiplication by 0.5 shifts any point on the string by the same amount. Thus a *rescaling* of the whole set of numbers just shifts the set *uniformly* along the string; and wrapping the string around the circle amounts to rotating the original distribution through some angle around the circle. For $s = 0.5$, this angle is $|\log_{10} 0.5| = 30.1\%$ of $360°$, or $108°$ in the clockwise direction. If the original distribution populates the circle uniformly, then so also does the rescaled distribution.

Scale changes are hardly relevant for distributions of numbers that are dimensionless, such as the Fibonacci numbers, populations, and branching fractions. It is sometimes said that Benford's law only applies to sets whose numbers have dimensions. Five of the distributions in Fig. 1 say otherwise.

To summarize: We would get the Benford distribution from the equal-tempered musical notes whether we used the frequencies f, or their inverses $1/f$, or their wavelengths $\lambda = v/f$, or measured the frequencies in cycles per year, or the wavelengths in Bohr radii.

Changing base—Inversion and scale invariance involve reflections or shifts of the markings for numbers on the string. Base invariance involves keeping the marks in place but changing the circumference of the circle we wrap the string around. Whereas invariance under a change of base is known to lead to Benford's law,[13] a set of numbers in accord with Benford's law is not necessarily base invariant.

The simplest nontrivial change from decimal is to trinary (or ternary). The first few Fibonacci numbers in trinary are

$$0,\ 1,\ 1,\ 2,\ 10,\ 12,\ 22,\ 111,\ \ldots\ ,$$

but the marks on the string stay in place. And now we wrap the string around a circle whose circumference equals the length of the string between (decimal) 1 and 3. This circumference is only $\log_{10} 3 = 47.7\%$ that of the circle in Fig. 6(b). Instead of circling the circle at (decimal) 10, 100, 1000, ..., we now circle it at (decimal) 3, 9, 27, ... The mark at 2 on the circumference would be $\log_3 2 = 63.1\%$ of the way around the circle. If Benford's law is satisfied, about 63% of the numbers will begin with a 1, the rest with a 2.

If the numbers in a set cover only a few orders of magnitude, such as the county populations, then the distribution cannot be invariant under *every* change of base. For suppose we change to a base of 100 million; then the marks would not get around the (large) circle even once. The mathematicians' series and exponentials can run on forever; data sets usually do not.

Another (highly artificial) example is a set of numbers distributed uniformly between 1 and 10 on the logarithmic string. The marks would wrap a bit more than twice around the trinary circle, giving too many 1s.

These counter-examples are, however, quibbles. Any set that obeys Benford's law to base 10 will almost certainly extend over several (decimal) orders of magnitude. If, within statistics, it is not invariant under a change to base 3 or 8 or 16, it probably would not obey the law for base 10 to begin with.

Applications

Benford's law has been shown to describe a set of 477 alpha-decay nuclear half lives,[19] and 3553 more general nuclear half lives.[20] A search for "Benford's law" in the Physical Review family of journals finds five papers, all since 2000. For three of the papers, the law is incidental, but for two it is central. The titles tell the applications: "Stochastic aspects of one-dimensional discrete dynamical systems: Benford's law,"[21] and "Benford's law and complex atomic spectra."[22]

An outlier bin or two in a Benford-like distribution can be a prompt to investigate. The χ^2 for Fig. 1(e), showing the errors on meson branching fractions, is 35, of which 23 comes from the $i = 3$ bin. This (it took me a long time to see) is because the Particle Data Group rounds errors with first digits between 1.0 and 3.55 to two places, but rounds errors between 3.55 and 9.9 to one place.[10] Thus the $i = 3$ bin actually only includes errors between 2.95 and 3.55, and the spillover continues to higher i bins. Properly binned, the χ^2 becomes 10.2 ($c.l. = 25\%$).

When a distribution of numbers is known from past experience to obey Benford's law, then new samples of that distribution, from simulations or experiments or auditing, ought also to obey the law. For example, a simulation of future behavior of a process or activity that in the past has obeyed the law ought to produce results that also obey the law. This has been called "Benford in, Benford out."[23] A failure in this respect is an indication of trouble. A failure to know that data on tax returns, inventories, and the like, often obey Benford's law has got people into trouble with the Law; see the last quote at the start of this article.

Which brings us to Mark Nigrini, whose 1992 Ph.D. thesis, "The Detection of Income Tax Evasion through an Analysis of Digital Distributions," and subsequent work, did much to spur development of these just-mentioned applications. And he has found examples in other areas: My Fig. 1(b) simply repeats for 2000 census data his analysis of 1990 census data.[24] Therefore, it is fitting to close with the remarkable Fig. 7, which uses U.S. Geological Survey records of 457,440 annual stream flow rates (in ft^3/s) measured at 17,822 distinct sites in U.S. rivers and streams, at some sites for more than 100 years.[25] The data are numerous and accurate enough to present in 90 two-place bins, 1.0 though 9.9.

I first learned of Benford's law from two popular articles.[17,23] Any law that unexpectedly describes so many and such varied sets of numbers can be useful in understanding those numbers. Figure 1(i) is an invitation to explore on your own.

I thank R. Rohde for getting me the first-digit counts of the marine-animal durations and for comments on those counts. I thank J.D. Jackson for comments on the draft of this paper.

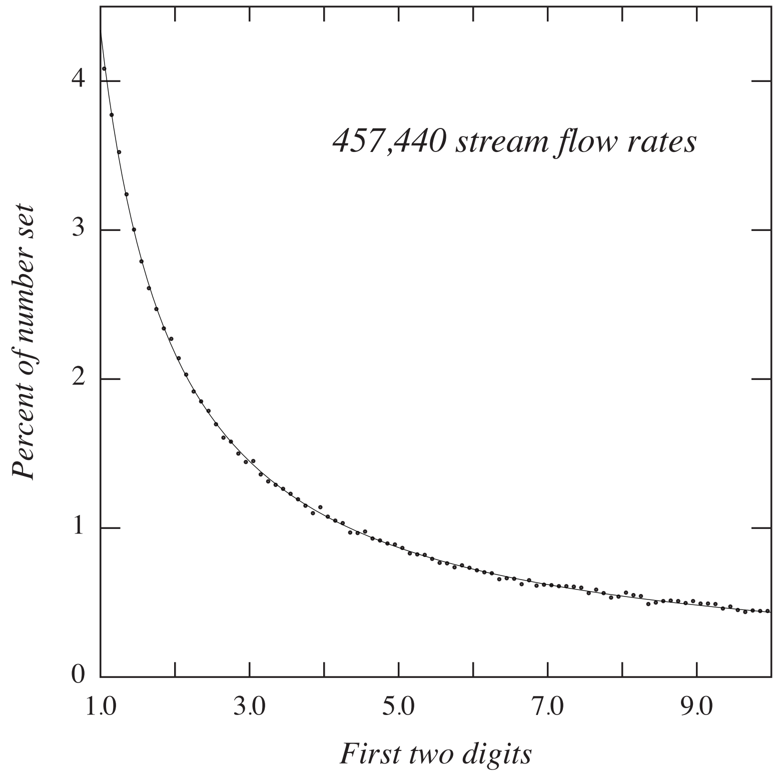

Figure 7. Benford's law and U.S. Geological Survey stream flow rates (in ft^3/s). Redrawn from Ref. 25.

* e-mail: `cgwohl@lbl.gov`

1. S. Newcomb, Am. J. Math. **4**, 39 (1881).
2. F. Benford, Proc. Am. Phil. Soc. **78**, 551 (1938).
3. D.E. Knuth, *The Art of Computer Programming, Vol. 2, Seminumerical Algorithms*, (Addison-Wesley, Reading MA, 1969), p. 220.
4. M.W. Browne, The New York Times, Aug. 4, 1998.
5. See W.E. Carter, M.S. Carter, Phys. Today **62**, 46 (2009), for more about Newcomb.
6. See J.D. Jackson, Am. J. Phys. **76**, 704 (2008), for five examples from physics in which a discovery is not named after its original discoverer. Another example is America.
7. See D. Schweizer, `math.holycross.edu/~davids/fibonacci/fibonacci.html`, for the first 500 Fibonacci numbers. The F_n are of course easily generated, but F_{500} has 105 digits. I don't know who first showed that the F_n satisfy Benford's law.
8. U.S. Census Bureau, *County and City Data Book: 2000.*
9. R.A. Rohde, R.A. Muller, Nature **434**, 208 (2005).
10. K. Nakamura et al., Particle Data Group, *Review of Particle Physics*, J. Phys. **G37**, 075021 (2010).

11. R. Resnick, D. Halliday, K.S. Krane, *Physics*, ext. 4th ed. (John Wiley & Sons, New York, 1992).
12. See `neo.jpl.nasa.gov/risk` for the "Sentry Risk Table," which is frequently updated. My numbers are from 20 May 2010.
13. R.A. Raimu, Am. Math. Mon. **83**, 521 (1976).
14. W.A. Whitworth, *Choice and Chance*, (Hafner, New York and London, 1965), p. 207; a reprint of the 1901 revised edition. D.S. Lemons, Am. J. Phys. **54**, 816 (1986) used a related equation from Whitworth to show analytically that the probabilistic partitioning of a conserved quantity leads to Benford's law.
15. D.E. Knuth, *The Art of Computer Programming, Vol. 1, Fundamental Algorithms*, 2nd ed. (Addison-Wesley, Reading MA, 1973), p. 82.
16. B.J. Flehinger, Am. Math. Mon. **73**, 1056 (1966).
17. R.A. Raimu, Sci. Am. **221**, No. 6, 109 (1969).
18. T.P. Hill, Stat. Sci. **10**, 354 (1996).
19. B. Buck, A.C. Merchant, S.M. Perez, Eur. J. Phys. **14**, 59 (1993).
20. Dongdong Ni, Zhongzhou Ren, Eur. Phys. J. A**38**, 251 (2008).
21. M.A. Snyder, J.H. Curry, A.M. Dougherty, Phys. Rev. E**64**, 026222 (2001).
22. J.-C. Pain, Phys. Rev. E**77**, 012102 (2008).
23. T.P. Hill, Am. Sci. **86**, 358 (1998).
24. M.J. Nigrini, `journalofaccountancy.com/Issues/1999/May/nigrini`.
25. M.J. Nigrini, S.J. Miller, Math. Geol. **39**, 469 (2007).