

Naked-Eye Quantum Mechanics: Practical Applications of Benford's Law for Integer Quantities

by Dean Brooks

ABSTRACT

Benford's Law (1938) predicts that digit frequencies for many scientific, engineering, and business data sets will follow $P(dd) = \log(1 + 1/dd)$ for digits dd . This law has been used by auditors since 1989 to detect errors and fraud in data sets. Benford also postulated a separate law for integer quantities. This little-known variant of the law is shown to be substantially correct, despite an error by Benford in its derivation. The integer variant is then shown to be extraordinarily common in everyday life, correctly predicting the distribution of footnotes per page in textbooks, sizes of groups walking in public parks and visiting restaurants, fatality counts in air crashes, repeat visits to service businesses, and purchase quantities for goods.

The practical value of the integer variant of Benford's Law is illustrated using cases from the author's consulting experience, as a limit toward which a distribution will tend. A potential proof for a universal distribution law for integer quantities (an 'ontic' distribution) combining Benford's Law, Zipf's Law, and Pareto's Law is outlined. The philosophical implications of 'naked-eye quantum mechanics' are briefly considered.

The possibility of a general law predicting digit frequencies was first addressed more than a century ago by astronomer Simon Newcomb. Newcomb observed in a brief 1881 paper for the *American Journal of Mathematics* that standard logarithm tables tended to be much more soiled and worn on their front pages, where the conver-

sion values for figures with low initial digits were found. He reasoned that this could only occur if the lower initial digits 1-3 occurred with much more regularity in typical computations, and proposed an empirical distribution that was itself logarithmic: $P(dd) = \log(1 + 1/dd)$, where P represents the relative frequency of a digit combination ' dd '. The formula as written is intended to apply to any number of digits, such as '1' or '467'. Newcomb also determined the probability of the ten second digits, independent of the first digits, e.g. $P(b) = \log[(ab+1)/ab] / \log[(a+1)/a]$, where ' a ' represents the first digit and ' b ' the second.

In 1938, Frank Benford (an engineer for General Electric) independently reported on the same law, deriving the same formula. He collected 20,229 observations from sources such as county populations, tables of engineering constants, baseball scores, and the street addresses from *American Men of Science*, and showed that although the digit distributions for any one particular source might diverge from the expected value, a set formed from diverse sources conformed very well. His paper, "The Law of Anomalous Numbers," chanced to appear in the same issue of the *Proceedings of the American Philosophical Society* as a critical paper on electron scattering by Hans Bethe; this association brought Benford's findings to a much larger audience than Newcomb's earlier work, and thus the law came to bear his name.

Editor's note: This paper was originally submitted to the Statistics Education seminar of the 2002 American Statistical Association's national convention, held in New York City in August 2002. Because of schedule conflicts, it was not presented at the ASA seminar. It appears here in a slightly lengthier form.

Between 1938 and the present, just over 100 papers have been published on aspects of Benford's Law (see Raimi, 1976, for a survey). Papers written prior to the 1990s were predominantly focused on explaining or deriving the law in theoretical terms. Several attempts were made to prove that the law was a consequence of our numbering system, and not 'real'. However, it had been clear even to Newcomb that the law implied an underlying logarithmic distribution of the actual sizes of events and entities, with many small items and progressively fewer large ones.

In a key development, Roger Pinkham argued (Pinkham, 1961) that for any digit-distribution law to hold consistently, it would have to be scale-invariant. A data set expressed in varying units of measurement, such as pounds versus kilograms, should conform equally well regardless of the units used. Furthermore, the number of units on successively larger value ranges should remain constant: if there are N units on the range x to kx, then there should also be N units found on the range kx to k²x. Pinkham demonstrated that the only law which satisfies both these conditions is the one Benford and Newcomb found.

In the late 1980s, attention turned to practical applications, particularly in auditing. Large sets of inventory data or financial transactions had already been shown to conform to the law. During the 1990s experimental work by Nigrini and others showed that if erroneous or fraudulent entries were included in an otherwise conforming data set, they could be expected to create deviating 'spikes' on a digit-frequency graph. This detection technique was the subject of *Digital Analysis Using Benford's Law: Tests and Statistics for Auditors* (2000), written by Mark J. Nigrini, Ph.D. and edited by the present author. The practice has become well known among auditors.

This balance of this paper reexamines Benford's original essay in its treatment of the various 'digital orders', in which a significant error occurred. Benford supplied a more general version of the formula to deal with situ-

ations where values were rounded or truncated. It seemed likely to him that the relative frequencies or probabilities P for initial digits 1-9 would be different if second digits were non-existent, rather than simply set aside in the frequency calculations. Benford called his result "the general equation for the Law of Anomalous Numbers", where r is the number of digits allowed:

$$P_1^r = \left[\log_e \frac{10(2 \cdot 10^{r-1} - 1)}{10^r - 1} + \frac{8}{10^r} \right] \frac{1}{N}$$

$$P_a^r = \left[\log_e \frac{(a+1)10^{r-1} - 1}{a \cdot 10^{r-1} - 1} - \frac{1}{10^r} \right] \frac{1}{N} \quad a \neq 1$$

It can be seen that this equation is approximated by P(dd) = log(1+1/dd) for high orders of r, allowing the use of the simpler formula in typical cases. For the lowest order of r, representing numbers that were truncated or rounded to one digit, Benford computed the theoretical frequencies as shown in Table 1 below.

A key assumption underlying all of Benford's work on anomalous digits is that any actual quantity in nature can be approximated by a continuous function, capable of being manipulated using integral calculus. As Benford put it, "The justification for using a continuous form is that the things we use the number system to represent are nearly always perfectly continuous functions, and the number, say 9, given to any phenomenon will be used in some degree for all the infinite sizes of phenomena between 8 and 10 when we confine ourselves to single digit numbers." Benford further observed that the spacing of observations tended to be geometric, and argued that there was "no necessity or implication of limits at either the upper or lower regions of the series."

These were curious statements. Taken literally, they seem to imply several things, none of which can be true. First, that the distributions of discrete entities such as oranges, cows or trees, can be precisely represented by continuous functions. Justifying this procedure, Benford wrote that "The summation of area under the curve . . . is taken as the probability for using a 9 for phenomena in

Digit	1	2	3	4	5	6	7	8	9	All
Freq	0.393	0.257	0.133	0.081	0.053	0.036	0.024	0.015	0.008	1.000

Table 1. Benford's predicted digit frequencies for r = 1.

this region. This is about equivalent to knowing accurately the size of all phenomena in this region and deciding to call everything between 8.5 and 9.5 by the number 9." We can concede that if we are measuring the span of a thundercloud as it moves, some sort of rounding in the dimensions seems legitimate. But in reality, one cannot typically observe 0.75 of a cow alongside 1.38 cows, much less can we decide pragmatically to call their sum two cows.

Furthermore, limits exist in reality that are not accounted for here. Observed values at the left-hand limit of the distribution are assumed to decline asymptotically to some nonzero value, while their probability of occurrence increases geometrically. But if it is legitimate to round all values between 8.5 and 9.5, should we not also round down all values between 0 and 0.5, to zero? Zero is eliminated as a leading digit in the continuous version of the law, because all the observed quantities are known with high precision. But if we are rounding—or counting entities that are integer-like in nature—then zero should re-enter the picture. To put this question another way, is it more reasonable in forming an empirical law of distribution to anticipate finding herds of ever-smaller microscopic cows (or fractions of cows) in the millions and then billions, or do we reach a point where we simply expect zero cows?

Benford's right-hand limit is equally doubtful. The assumption that observed sizes of phenomena will form a

scale-invariant geometric series of unlimited range has never been seriously tested in the literature, and seems impossible on the face of it. Every data set derived from actual entities or measurements has a largest member. We tend to believe, following modern cosmology as well as arguments from Aristotle, that the universe is finite—but even if the universe as a whole were somehow infinite, nothing on Earth is.

Revisiting the first digital order

In presenting the simpler form of his digits law, $P(dd) = \log(1+1/dd)$, Benford made sure to validate his results against a wide variety of data. The abstract reasoning which he applied to create his 'general equation' was not validated in the same way. Instead, Benford focused on a test using exactly one type of data: the distribution of footnotes in engineering reference books.

Benford explained his choice this way: "The frequencies of the single digits 1 to 9 vary enough from the frequencies of the limiting order to allow a statistical test if a source of digits used singly can be found. The footnotes so commonly used in technical literature are an excellent source, consisting of units that are indicated by numbers, letters or symbols."

By examining the *Standard Handbook for Electrical Engineers* and several other reference works, Benford assembled a table of relative frequencies of occurrence for pages with 1-9 footnotes each.¹

	1	2	3	4	5	6	7	8	9	Total pages
Book 1	55.1	22.7	12.3	5.0	2.4	1.7	0.3	0.3	0.3	586
Book 2	56.3	22.1	6.6	6.1	5.0	2.2	1.1	0.6	0.0	181
Book 3	52.8	23.6	8.5	5.5	4.0	3.2	0.8	0.0	1.6	127
Book 4	37.2	25.7	12.1	9.5	4.8	5.2	2.6	2.2	0.9	230
Book 5	29.7	26.6	14.6	11.0	8.0	5.9	1.8	1.0	1.4	287
Book 6	19.5	17.4	17.7	11.9	11.3	9.2	6.1	5.8	1.1	293
Book 7	33.0	27.5	11.8	10.7	4.3	5.9	2.8	2.4	1.6	254
Book 8	56.8	23.2	6.7	7.6	2.4	1.4	0.5	1.4	0.0	211
Book 9	49.5	22.3	13.7	6.9	2.3	1.5	1.5	1.5	0.8	394
Book 10	41.7	25.2	13.4	9.1	4.7	3.2	1.7	0.5	0.5	405
Average	43.2	23.6	11.8	8.3	4.9	3.9	1.9	1.6	0.8	2968

Table 2. Benford's percentage frequencies of footnote counts per page in various engineering texts¹

Benford then observed that the agreement with theory was as good as the computed probable errors, and so concluded that the 'general equation' was correct. Benford's claim appears to have gone largely unchallenged by subsequent writers, and in practical applications, the simpler version of his equation is invariably used. Progress in justifying and applying the law has been effectively confined to those quantities whose distributions approximate continuous functions of high order r , far from either left-hand or right-hand limits.

For many quantities like money or weight, the limits may not be of great practical importance, and units can indeed be varied to almost any degree we please. This is why corporate databases tend to conform closely to Benford's simple formula, and why digital analysis aimed at detecting anomalies has worked so well. But for 'quantized' items such as oranges, cows and trees, the universe dictates one particular, non-optional unit of measure. The left-hand limit becomes zero, the right-hand limit becomes significant, and scale invariance can no longer be regarded as an essential condition.

Quantization, zero, and the Fibonacci proportion

Benford derived his 'general equation' using a log-linear graphical method, into which a value of zero cannot be introduced. To allow for the existence of zero as the left-hand limit requires a new rationale for estimating the relative probabilities.

Here we will resort to an intuitive analogy. The relative probabilities of the digits (shown in the table) seem suspiciously geometric in nature. The likelihood of observing 1 cow (say) is 1.8 times as great as the likelihood of 2 cows . . . and the likelihood of observing 2 cows is 1.7 times as great as 3 cows, and so on. These values vary quite symmetrically above and below a median of 1.57 and a mean of 1.64. If we leave aside continuous function space, and think of each integer quantity as being

somewhat like a quantum energy level, perhaps these probabilities are not merely similar, but identical.

Furthermore, these ratios cluster closely around the famous Fibonacci proportion or 'golden mean' of 1.618034... The correspondence is so close, in fact, that we would need a very large sample size before we could argue on strictly empirical grounds which is the true underlying pattern (see Figure 1 on the next page). Table 3 shows the proportions that would result if each successively higher integer occurred 0.618034 times as often as the previous one, compared with Benford's. Although much more can be said about the merits of using the Fibonacci proportion in this way, here we will simply let it stand—as a promising hypothesis in much the same spirit as the original approximation proposed by Newcomb and Benford.

We immediately note an opportunity to test our theory: in computing his footnote frequencies, Benford helpfully included the total number of pages he examined, for six of the ten books in his sample. If the fixed-probability model is correct, then we would expect to find 1.618 times as many zeroes (pages with no footnotes) as ones in those books. For now we will neglect the number of pages with 10 footnotes or more, as likely being very small.

Benford observed 499 pages with one footnote, and 1,402 pages with one footnote or more, out of a total of 2,166 pages. The ratio of zeros to ones is approximately $(2,166-1,402)/499$, or around 1.53. The fit is as good for the zeros as for any other digit; therefore our 'quantum' model, generalized to cover zeros, works on Benford's original data.

In fact, there is no doubt that this represents a real, and common, distribution pattern. The author has observed the same relative frequencies in dozens of other situations, and has had success in using the 'quantum' curve

Digit	1	2	3	4	5	6	7	8	9
Benford	0.39319	0.25760	0.13266	0.08152	0.05348	0.03575	0.02352	0.01456	0.00772
Ratio		1.53	1.94	1.63	1.52	1.50	1.52	1.62	1.89
Fibonacci	0.38706	0.23922	0.14784	0.09137	0.05647	0.0349	0.02157	0.01333	0.00824
Ratio		1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62

Table 3. Comparison of Benford's expected frequencies with frequencies based on uniform geometric series using Fibonacci

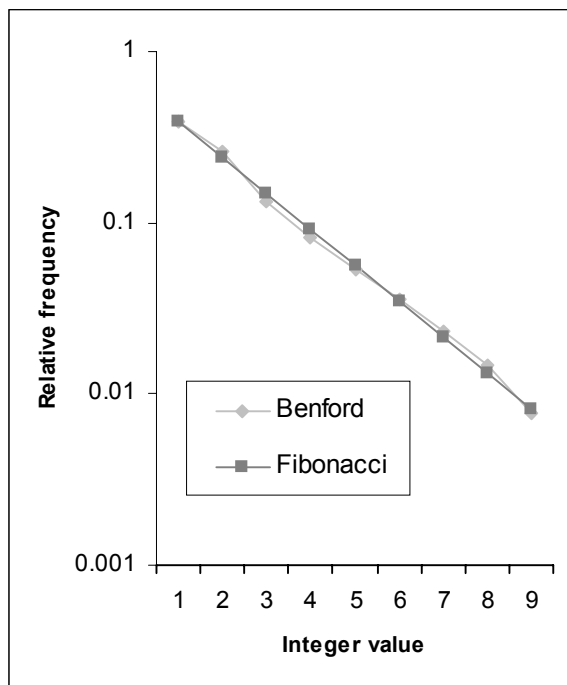


Figure 1. Graphic comparison of Benford's curve to Fibonacci

as a practical forecasting tool. For example, copies of business software sold to different customers: on average, 1.618 times as many customers buy one copy as buy two, and 1.618 times as many buy two as buy three, and so on. This same pattern applies to return visits to service businesses such as hairdressers, and to subscription renewals, and to many other aspects of economic life. It also applies to, for example, the distribution of field goals per player during NCAA basketball: even the proportion of zeros, representing at least one attempt by that player but no goals scored, is as predicted.

When we examine transportation accident statistics, we bring into focus one more striking consequence of this distribution. Assuming that we consider only accidents that have both fatalities and survivors (ignoring accidents that are 100 percent fatal, or 100 percent nonfatal), the distribution of fatalities and the distribution of survivors *each* follow the ontic distribution, with one survivor or fatality being 1.62 times as likely to occur as two, and so on. This pattern emerges regardless of the particular transportation type: survival statistics for lost ships, crashing planes, and even submarines destroyed in combat all follow the law. The law is consistent even when considering a single aircraft type, such as every crash involving a Douglas DC-3, or every accident that occurred in the month of March.

As a separate constraint, the proportion of fatalities to survivors in such cases averages very close to 1:1. This latter constraint may turn out to be an application of Zipf's Law, where the largest group (all passengers prior to the crash) averages twice the size of the second-largest group (all passengers still alive after the crash).

Equally intriguing are those cases in which the distribution falls short, failing to fill in to the limiting values. For example, in the early years of television, very few households had more than one TV. This type of shortfall tends to indicate some practical barrier, in this case most likely high price, or the inability to use more than one of the item (e.g. few households own more than one set of encyclopedias). Further investigation can often determine the specific cause and in many cases action can be taken, altering the distribution: thus in recent decades the price of televisions has fallen such that many if not most households now have several.

That this curve represents a genuine limit is borne out by the rarity of cases where the distribution of low integer values rises above Benford's expected line. Nonetheless it is clearly not a strict physical limit, merely a statistical one; it can be violated, temporarily or in unusual circumstances.

An 'ontic' distribution

If this 'quantum' pattern were to persist unchanged to indefinitely large values, not only for the integers 1 to 9 but the integers 1 to 1 billion, some new paradoxes would soon emerge. Cities would be non-existent, as the relative probability of a grouping of N individuals is given by $0.618034^{(N+1)}$. But we already know from diverse observations that this pattern does not persist to indefinitely large values; for larger values, the distribution must transition to a curve approximating Benford's Law.

This puzzle introduces us to a much wider problem, that of reconciling two competing forms of distribution, Pareto's Law and Zipf's Law, with these two variants of Benford's Law. This problem is discussed in depth in successive issues of *Frequencies: The Journal of Size Law Applications*. The author has determined that a single distribution, called 'ontic' from the Greek 'ontos' for entity, can account for the patterns observed by Pareto, Zipf, and Benford.

The critical point is that we must abandon Benford's assumption, as ratified by Pinkham, that the density of

items on successively larger ranges is constant. In the ontic distribution, for each doubling of the range, the number of items on the new range falls to 0.618 of the previous range: thus if there are 1,000 corporations with sales between \$1 million and \$2 million, there should be approximately 618 corporations with sales between \$2 million and \$4 million. The same would apply to counts of employees or personal computers. This is confirmed by *Fortune* 500 listings from recent decades.

Interestingly, the expected proportions of the initial digits 1-9 under an ontic distribution are close enough to the standard Benford distribution that for most practical purposes the deviation would not be noticed. There ought to be a tendency for ontically distributed data to have a slight but statistically significant excess of '1's or '2's, and too few '8's and '9's, by comparison to Benford's Law—and this is precisely what is observed in many actual data sets (see for example several in Nigrini, 2000, or the U.S. population data from Benford, 1938). Such a mild deviation in slope does not interfere with the anomaly-detection techniques used by auditors.

The slope of the ontic distribution is essentially the same as that found by Pareto in his late nineteenth-century studies of income and land distributions among European elites. Pareto's law of elites revealed a gradual decline in numbers, so that if the number of observations on a given range is N , the number found after extending the upper limit of the range by a factor of k should be $k^{-0.7}$ times N . Inserting $k=2$ and $1.618N$ as the observed result, we find that the implied exponent in the ontic distribution is -0.69424 , very close indeed to Pareto's average value of -0.7 .

This ontic distribution strongly suggests a fractal or self-similar ordering. In the general case, if there are N items in the range x to $2x$, then there will be approximately 1.618 times N items in the range from $2x$ to infinity (meaning to the highest value in the set).

This follows as a consequence of summing the infinite series of contributions from successively larger ranges:

	x-2x:	1.000	
compared to	2x-4x:	0.618	1.618 total
plus	4x-8x:	0.382	
plus	8x-16x:	0.236	
plus	16x-32x:	0.146	
	...	...	

In the left-hand limit, where we deal with zeros or small integer values, the pattern is modified to the 'quantum' version found above, in part because the 'range' between successive integers cannot be occupied. In the right-hand limit, where we deal with the largest values, the distribution is typically ontic but occasionally approximates (under certain conditions) Zipf's Law. ²

Even where the ontic distribution in its most literal form is not evident, the underlying ordering principle still may be, and can have many strange and interesting side consequences. The Fibonacci proportion shows up in many contexts in which a range of possible values is evenly divided into two halves. For example:

- Approximately 62 percent of the individual names in North American phone directories begin with the letters A-M, while 38 percent begin with N-Z. The same is true of the corporate names in the *Fortune* 1000.
- About 62 percent of professional hockey players have birthdays between January 1 and June 30 (which gives them several months' height and weight advantage in junior hockey compared with boys born later in the calendar year). Among professional soccer players (who play a different season), 62 percent have birthdays between July and December.
- On four-lane roadways, 62 percent of traffic tends to travel in the curb lanes, and 38 percent in the faster median lanes.
- Initial assassination attempts against public figures tend to succeed about 62 percent of the time. A successful assassin tends either to be killed in the act, or captured and subsequently put to death, about 62 percent of the time as well.
- In Stanley Milgrim's famous psychology experiment, in which volunteer lab assistants were urged by an authority figure to inflict intense (simulated) pain on a protesting subject, on average 62 percent of the volunteers followed instructions and completed the procedure. A 62 percent "yes" or "no" vote is very common in surveys with exactly two answer choices.

Philosophical overview

There is no firm consensus at present on the precise causes behind Benford's Law, or Zipf's Law, or Pareto's Law. This is particularly evident when we consider that Pareto's Law applies to tangible entities such as land or money, while Zipf's Law originally applied to frequencies of occurrence for words in a language, and Benford's Law applies to practically any number, including constants from engineering handbooks.

One commonly voiced suggestion is that these are all in effect "meta-laws," the subtle and aggregate consequences of many individual constraints and influences. Theodore Hill has described Benford's Law (Hill, 1998) as arising from "a distribution of distributions", meaning that although any particular data set may not conform to it, as the diversity of data grows, the Benford digit pattern inevitably emerges.³

The ontic distribution has the effect of reorganizing these various meta-laws into two broader ones, at the cost of losing some of the generality of Benford's original digit law. It applies to observed quantities of distinct entities, not necessarily to abstract ratios or scientific constants or all numbers in general:

1st Law of Ontic Distributions

Where both the size and number of observations are free to vary, the sums of all observations in successive ranges x to kx and kx to k^2x will tend to be equal.

2nd Law of Ontic Distributions

Where both the size and number of observations are free to vary, the number of observations in successive ranges x to $2x$, and $2x$ to the largest observation in the set (much larger than $2x$), will tend to the ratio 1:1.618...

The 1st Law replaces Pinkham's scale invariance principle with an "indifference principle": given a sufficient diversity of conditions and influences, the distribution of observations tends to be such that the same total amount of substance (cows, oranges, dollars, fatalities) can be found on any two adjacent ranges of similar proportion.

The 2nd Law reconciles Benford's Law with Pareto's Law and Zipf's Law by treating the distribution of the sizes of observations as a self-similar, fractal ordering in which

the proportion of units on a given range is a constant fraction of the number of units above that range. Under Benford's Law, the proportion of units found upon successive doublings of the range should remain a constant 1:1; under Zipf's Law it falls to 1:0.5; but in the most general case (covering Pareto's observations and a great diversity of real data) it is actually 1:0.618.

Notice the symmetric deviations from these two laws when we examine the left-hand and right-hand limits.

- In the left-hand limit, the 2nd Law still can be adhered to exactly, if we allow for the quantization of the ranges: there tend to be 1.618 times as many observations above a given integer, as at that integer. However, the 1st Law can not be adhered to consistently because an integers-only range like 1 to 2, or 2 to 4, is not a good approximation for a freely varying range of values.
- In the right-hand limit, the 1st Law can be adhered to nearly exactly, given that the sizes of the individual observations can vary. However, if the total number of observations is held fixed, as for example when we are ranking population sizes for cities and the number of cities is therefore fixed, the distribution deviates away from the 2nd Law to conform with Zipf's Law.

Ultimately, the ontic distribution is likely to prove more commonly observed and perhaps just as significant in its practical applications as the bell-shaped normal distribution. The many measures that it pertains to (sales, land ownership, accidental deaths, behavioral choices) may make it just as vital to everyday life as the more well-established uses of the Gaussian distribution, such as statistical process control in manufacturing. In the long run we will have to adjust our intuition about what is 'normal' to what nature shows us.

Educational opportunities

This paper has been submitted under the heading of statistics education. In addition to its practical forecasting value and philosophical significance, this distribution also has unique educational potential. Math education for high school students and university undergraduates does not always stress the tangible, practical connection of mathematics to everyday reality, and even gifted beginners find it difficult to produce original work.

However, in this case the scope for original discovery is great, even for amateurs. There are literally hundreds of potential applications, many only requiring some patient observation and record-keeping, or re-examination of existing data sets. Examples include:

- Sizes of bird and animal flocks
- Sizes of icebergs
- Numbers of disaster casualties (earthquake, flood, fire)
- Sizes of groups entering restaurants
- Sizes of stock market transactions
- Sizes of computing tasks
- Sizes of oil and chemical spills
- Sizes of insurance claims

Those interested in reporting their observations are invited to submit to *Frequencies: The Journal of Size Law Applications*, at www.ekaros.ca.

References

"The Law of Anomalous Numbers," Frank Benford, *Proceedings of the American Philosophical Society*, 1938

"The First Digit Phenomenon," Theodore P. Hill, *American Scientist*, Jul-Aug 1998

Digital Analysis Using Benford's Law: Tests & Statistics for Auditors, Mark J. Nigrini, Global Audit Publications, 2000

"On the distribution of first significant digits," Roger Pinkham, *Annals of Mathematical Statistics*, 1961

"The First Digit Problem," Ralph A. Raimi, *American Mathematical Monthly*, Aug-Sep 1976

Acknowledgments

I want to thank Mark Nigrini for the unique opportunity of editing his book and the insights into the background of Benford's Law that our conversations afforded me; Milo Schield of Augsburg College, for inviting me to the 2002 ASA session on Statistical Education and encouraging my work on these questions; and BethAnn Burton, freelance researcher, for her patient support and enthusiasm in collecting data and assisting in my many experiments.

Notes

1. This table of footnote frequencies by Benford contains several minor inconsistencies that I cannot account for as being wrong additions or similar kinds of error. In private notes shared during the editing of his book, Mark Nigrini observed that Benford's main table of results (Benford, 1938) errs by what appears to be occasional improper rounding in the direction of greater conformity to predicted values. The inconsistencies in this case do not seem to be of that kind, but like those previously observed by Nigrini, they are not significant statistically.
2. Zipf's Law is usually stated as a ranking rule: the Nth-ranked item in a distribution with largest member size C will have average size C/N. The second-largest item is 1/2 the size of the largest, the third-largest is 1/3rd, and so on. This implies that item counts decline by 50 percent with each doubling of the range, as items ranked 51st through 100th will occupy half the range of items ranked 26th through 50th, etc.
3. Hill rephrased the significant-digit law this way: "If distributions are selected at random (in any 'unbiased' way) and random samples are taken from each of these distributions, then the significant-digit frequencies of the combined sample will converge to Benford's distribution, even though the individual distributions selected may not closely follow the law."



About this publication

This is a free technical supplement to:

*Frequencies:
The Journal of Size Law Applications*

For copies of *Frequencies* or for more information

e-mail us at: frequencies@ekaros.ca
visit our website: www.ekaros.ca/frequencies
or write to us care of:

Ekaros Analytical Inc.
Box 408
125A - 1030 Denman Street
Vancouver, B.C. Canada
V6G 2M6