Benford's Law / *A Formula for Predicting Probability*

Beating the Odds (and the Frauds) With the Number 1

By Malcolm W. Browne
New York Times Service

NEW YORK — Theodore Hill asks his mathematics students at the Georgia Institute of Technology to go home and either flip a coin 200 times and record the results or merely pretend to flip a coin and fake 200 results. The following day he runs his eye over the homework data, and to the students' amazement, he easily fingers nearly all those who faked their tosses.

"The truth is," he said in an interview, "most people don't know the real odds of such an exercise, so they can't fake data convincingly."

There is more to this than a classroom trick.

Mr. Hill is one of a growing number of statisticians, accountants and mathematicians who are convinced that an astonishing mathematical theorem known as Benford's Law is a powerful and relatively simple tool for pointing suspicion at frauds, embezzlers, tax evaders, sloppy accountants and even computer bugs.

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.

Benford's Law is named for the late Frank Benford, a physicist at the General Electric Co. In 1938 he noticed that pages of logarithms corresponding to numbers starting with the numeral 1 were much dirtier and more worn than other pages.

(A logarithm is an exponent. Any number can be expressed as the fractional exponent — the logarithm — of some base number, such as 10. Published tables permit users to look up logarithms corresponding to numbers or numbers corresponding to logarithms.)

Logarithm tables (and the slide rules derived from them) are not much used for routine calculating any more; electronic calculators and computers are simpler and faster. But logarithms remain important in many scientific and technical applications, and they were a key element in Mr. Benford's discovery.

Mr. Benford concluded that it was unlikely that physicists and engineers had some special preference for logarithms starting with 1.

He therefore embarked on a mathematical analysis of 20,229 sets of numbers, including such wildly disparate categories as the areas of rivers, baseball statistics, numbers in magazine articles and the street addresses of the first 342 people listed in the book "American Men of Science."

All these seemingly unrelated sets of numbers followed the same first-digit probability pattern as the worn pages of logarithm tables suggested. In all cases, the number 1 turned up as the first digit about 30 percent of the time, more often than any other.

Mr. Benford derived a formula to explain this. If absolute certainty is defined as 1 and absolute impossibility as 0, then the probability of any number "d" from 1 through 9 being the first digit is log to the base 10 of (1+1/d). This formula predicts the frequencies of numbers found in many categories of statistics.

Probability predictions are often surprising. In the case of the coin-tossing experiment, Mr. Hill wrote in the current issue of the magazine American Scientist, a "quite involved calculation" revealed a surprising probability. It showed, he said, that the overwhelming odds are that at some point in a series of 200 tosses, either heads or tails will come up six or more times in a row.

Most fakers do not know this and avoid guessing long runs of heads or tails, which they mistakenly believe to be improbable. At just a glance, Mr. Hill can see whether a student's 200 coin-toss results contain a run of six heads or tails; if they do not, the student is branded a fake.

Even more astonishing are the effects of Benford's Law on number sequences. Intuitively, most people assume that in a string of numbers sampled randomly from some body of data, the first nonzero digit could be any number from 1 through 9. All nine numbers would be regarded as equally probable.

But, as Mr. Benford discovered, in a huge assortment of number sequences — random samples from a day's stock quotations, a tournament's tennis scores, the numbers on the front page of The New York Times, the populations of towns, electricity bills in the Solomon Islands, the molecular weights of compounds, the half-lives of radioactive atoms and much more — this is not so.

Given a string of at least four numbers sampled from one or more of these sets of data, the chance that the first digit will be 1 is not one in nine, as many people would imagine; according to Benford's Law, it is 30.1 percent, or nearly one in three. The chance that the first number in the string will be 2 is only 17.6 percent, and the probabilities that successive numbers will be the first digit decline smoothly up to 9, which has only a 4.6 percent chance.

A strange feature of these probabilities is that they are "scale invariant" and "base invariant." For example, it does not matter whether the numbers are based on the dollar prices of stocks or their prices in yen or marks, nor does it matter if the numbers are in terms of stocks per dollar; provided there are enough numbers in the sample, the first digit of the sequence is more likely to be 1 than any other.

The larger and more varied the sampling of numbers from different data sets, mathematicians have found, the more closely the distribution of numbers approaches what Benford's Law predicted.

One of the experts putting this discovery to practical use is Mark Nigrini, an accounting consultant affiliated with the University of Kansas who this month is joining the faculty of Southern Methodist University in Dallas.

Mr. Nigrini gained recognition a few years ago by applying a system he devised based on Benford's Law to some fraud cases in Brooklyn, New York. The idea underlying his system is that if the numbers in a set of data like a tax return more or less match the frequencies and ratios predicted by Benford's Law, the data are probably honest.

But if a graph of such numbers is markedly different from the one predicted by Benford's Law, he said, "I think I'd call someone in for a detailed audit."

Some of the tests based on Benford's Law are so complex that they require a computer to carry out. Others are surprisingly simple; just finding too few ones and too many sixes in a sequence of data to be consistent with Benford's Law is sometimes enough to arouse suspicion of fraud.

Robert Burton, the chief financial investigator for the Brooklyn district attorney, recalled in an interview that he had read an article by Mr. Nigrini that fascinated him.

"He had done his Ph.D. dissertation on the potential use of Benford's Law to detect tax evasion, and I got in touch with him in what turned out to be a mutually beneficial relationship," Mr. Burton said. "Our office had handled seven cases of admitted fraud, and we used them as a test of Nigrini's computer program. It correctly spotted all seven cases as involving probable fraud."

BUT the fit of number sets with Benford's Law is not infallible. "You can't use it to improve your chances in a lottery," Mr. Nigrini said. "In a lottery someone simply pulls a series of balls out of a jar, or something like that. The balls are not really numbers; they are labeled with numbers, but they could just as easily be labeled with the names of animals. The numbers they represent are uniformly distributed, every number has an equal chance, and Benford's Law does not apply to uniform distributions."

Another problem Mr. Nigrini acknowledges is that some of his tests may turn up too many false positives. Various anomalies having nothing to do with fraud can appear for innocent reasons.

For example, the double digit 24 often turns up in analyses of corporate accounting, biasing the data, causing it to diverge from Benford's Law patterns and sometimes arousing suspicion wrongly, Mr. Nigrini said.

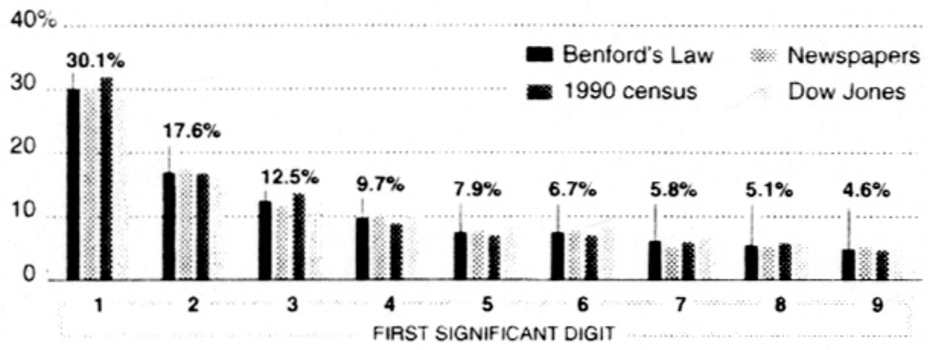
"But the cause is not real fraud, just a little shaving," he said. "People who travel on business often have to submit receipts for any meal costing $25 or more, so they put in lots of claims for $24.90, just under the limit. That's why we see so many 24s."

A Singular Phenomenon

Frank Benford's 1938 theorem dealing with the prevalence of the number 1 as an initial digit is being used to spot fraud.

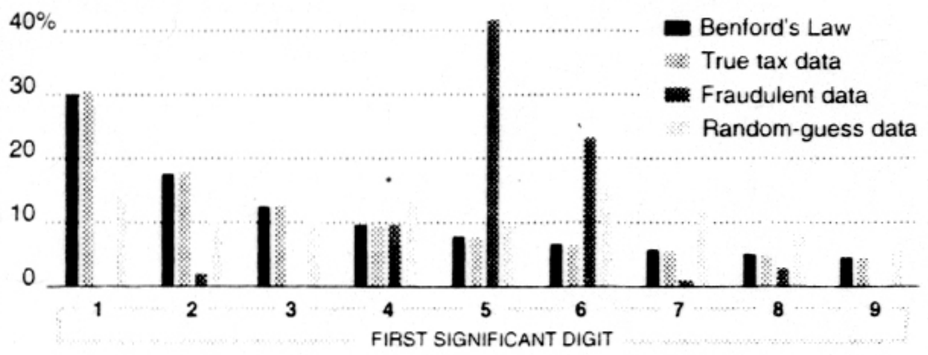
AN UNEXPECTED PATTERN

Very different sets of data follow Benford's Law within 2 percent, including numbers appearing on front pages of newspapers (collected by Mr. Benford), 3,141 county populations in the 1990 census (by Mr. Mark Nigrini) and the Dow Jones Industrial Average from 1990-93 (by Eduardo Ley).



DETECTING FRAUD

The first digits of true tax data taken from the lines of 169,662 tax forms follow Benford's Law closely. But fraudulent data taken from a 1995 Brooklyn study of cash disbursement and payroll in business do not follow Benford's Law. Likewise, data taken from 743 freshmen's responses to a request to write down randomly a six-digit number do not follow the law.



Source American Scientist

The New York Times