IN PRACTICE

The promises and pitfalls of Benford's law

If a data set deviates from Benford's law, is that evidence that the figures within are fraudulent? Perhaps not, says **William Goodman**. To those hoping for an automatic fraud detector, he argues that important statistical requirements are not necessarily being met

William Goodman is a professor at the University of Ontario Institute of Technology, where he teaches business statistics, forecasting and risk management. He has conducted extensive research into the distribution of incident outcome severities (magnitudes) in diverse applications — from industrial and radiological health and safety, to medication-administration errors, to evaluating educational innovations

Numbers are easily manipulated. Accounting scandals, overrated mortgage derivatives, billion-dollar pyramid schemes and falsified deficit figures have shown this to be true. Such manipulations are easy to fall for – but statistical analysis can offer some defence. Since the 1990s, a phenomenon known as Benford's law has been held aloft as a guard against fraud – as a way to check whether data sets are free from interference.

Benford's law tells us something about the frequency of leading digits in natural data sets – that is, how many numbers beginning with 1s, 2s, 3s, etc. we should expect to see. The law was first discovered by astronomer Simon Newcomb[1] in 1881. However, in keeping with Stigler's law of eponymy – which states that no scientific discovery is named after its original discoverer – the law was named after Frank Benford,[2] a physicist at General Electric, who rediscovered it some 50 years later.

Newcomb and Benford hit upon the law while flicking through books containing tables of logarithms, which were used to assist in performing mathematical operations, such as multiplication or calculations with exponents, in the days before calculators. One might expect the numbers being sought in the log table to be scattered throughout the book – no more likely to be on the first pages (numbers starting with 1 or 2) than near the back (starting with 8 or 9). But Newcomb and Benford both discovered that pages nearer the front were more soiled and worn than those at the back.

For Newcomb, the finding inspired a short, mathematically oriented paper, which few picked up on. Benford's later paper, on the other hand, has (in the current idiom) "gone viral", particularly in the last 20 years. It includes data from 20 data sets that he collected from diverse sources, which seem to support his model. Benford's data suggest that, in natural data sets, numbers with 1 as their leading digit should crop up around 30% of the time, 2 around 17% of the time, and so on, up to 9, which should appear as a leading digit only 4% of the time.

Mark Nigrini was perhaps the first to suggest that Benford's law could be used for fraud detection in financial, electoral or other data sets. The basic assumption here is that when the distribution of leading digits in a data set diverges noticeably from that expected under Benford's law, this anomaly might be an indication of fraud or manipulation.

For the past two decades, Nigrini has identified and discussed numerous possible cases where this holds true, including data from the well-known Enron accounting scandal.[3] Benford's law also informed a 2011 paper in which researchers reported an "abnormal" distribution of numbers in Greece's economic reports to European authorities over several years,[4] apparently confirming the European Commission's independent allegations of data manipulation.[5]

This and similar work has promise; but there has been a worrying tendency in recent years for Benford's law to be seen as an automatic fraud detector and that any divergent data sets are in some way fraudulent. This view has been supported by some uncritical reporting in both academic and popular media. Combining this with the inclusion of Benford's law-based utilities in commercial auditing software,[6] there is a very real risk that mechanically produced results are being misinterpreted.

To correct this imbalance, this article cautions that the statistical requirements for using Benford's law for formal hypothesis testing ("fraud detecting") are not necessarily being met. Examples of approaches that might better acknowledge the method's limits are then presented.

Benford's basics

Figure 1 illustrates the key observation for Benford's law, in this case regarding prices paid for racehorses at auction. You might expect about equal numbers of prices to begin with digit 1, as compared to digit 9 or some other digit. But we see that several numbers start with 1, and only one number starts with 9.

Benford's law predicts this very pattern: for many data sets, proportionally more numbers

115 000	5 000	12 000	12 500	15 000
20 000	46 000	8 000	15 000	5 000
6 750	28 000	36 000	10 000	5 000
3 000	2 000	19 000	9 000	40 000
3 250	27 500	7 000	4 500	75 000

FIGURE 1 Prices paid, in dollars, for racehorses at an auction (subset from larger sample)

Main image: Sky_Blue/iStock/Thinkstock

TABLE 1 Proportions of first digits expected by Benford's law, compared to proportions in the racehorse prices

First digit	BL-expected proportions of specific first digits in the data set	Actual counts of specific first digits in the data set	Actual proportions of specific first digits in the data set
1	0.301030	7	0.280000
2	0.176091	4	0.160000
3	0.124939	3	0.120000
4	0.096910	3	0.120000
5	0.079181	3	0.120000
6	0.066947	1	0.040000
7	0.057992	2	0.080000
8	0.051153	1	0.040000
9	0.045757	1	0.040000

start with 1 than 2, and more with 2 than 3, and so on. The expected proportions of numbers that start with each digit are shown in Table 1. The racehorse prices from Figure 1 conform approximately to these expectations, as also shown in Figure 2.

The expected proportions of numbers in a data set having particular first digits can be calculated as:

$$\text{Prob (first digit)} = \log_b (1 + 1/d)$$

where d is the particular first digit of interest (e.g., 5), and b is the base for the number system in use (base 10 for our number system). A key property of Benford's law is that, in instances where it is observed, the same pattern usually persists if all the numbers are changed to a different base (such as base 8) or are converted to different units of measurement. In extensions of Benford's law, expected proportions can also be calculated for digits in second or other digit positions in a data set, or even for digit combinations. Beyond the second digit, however, the Benford pattern essentially disappears.

And yet, even when we stick to first digits only, it has been empirically shown that many data sets – fraud-free ones included – do not have first-digit distributions that adhere to Benford's law. Benford's 20 data sets mostly did. The box plots in Figure 3, based on these data sets, suggest a range for the proportions of numbers that start with specific digits. In aggregate, all the first-digit proportions appear to centre around the expected values, however the outliers (dots) suggest individual exceptions. That said, we are not told what Benford's data selection criteria were, or if he cleaned the data. He also includes mixtures from unrelated sources, such as numbers happening to appear in newspaper stories when he was writing. The included numbers may have ranged from sports scores to war casualties or stock prices.

What, then, are the requirements for a data set to be compatible with Benford's law – to be "Benford-suitable"? The following guidelines have been suggested:

(a) *Sufficient sample size*. In a small sample, of say 20 numbers, even two numbers more than expected starting with digit 1 would represent a seemingly noticeable (10%) divergence, yet statistically valid conclusions could not be reached about the population.

(b) *Large span of number values*. Suppose that first digits for the population of cash receipts inherently follow Benford's law, but for some reason a sample's values range only from \$20 to \$500. First digit 1s would be underrepresented (only observable as first digits in the 100s range), whereas first digit 2s could be observed in the 20s *and* 200s. More orders of magnitude could minimise such problems.

(c) *Positively ("right-") skewed distributions of numbers*. Data sets that conform to Benford's law often have combinatory or multiplicative origins; such as expense receipt amounts (derived from (price) × (quantity)), or values for hurricane-damage insurance claims (derived from (strength of hurricane) × (insured amounts for affected properties)). Numbers generated like this tend to have logarithmic-type distributions, with extended "right tails" of large values, and are, in turn, more likely to exhibit Benford patterns.

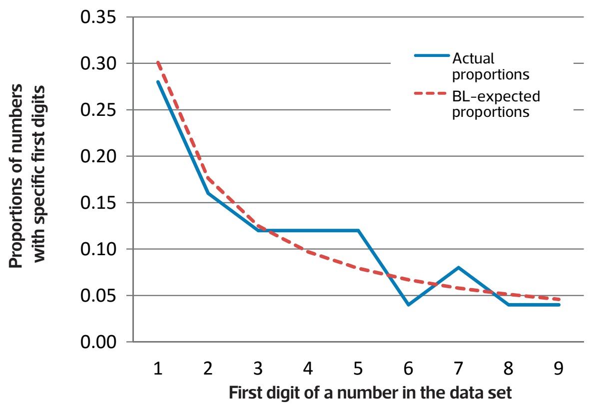

FIGURE 2 Actual versus Benford-expected distributions of specific first digits for the racehorse prices

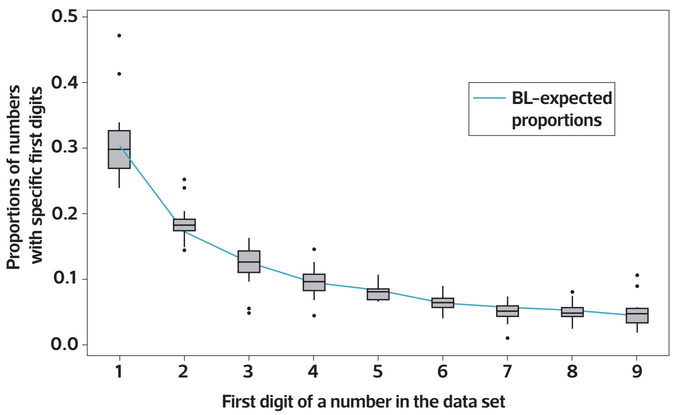

FIGURE 3 Distributions of first-digit proportions for Benford's 20 data sets

This article has been updated to correct a wording error in the description for calculating "d*" on page 41. Apologies for the error. Note that the correct version of the formula was used for all the article's calculations and figures.

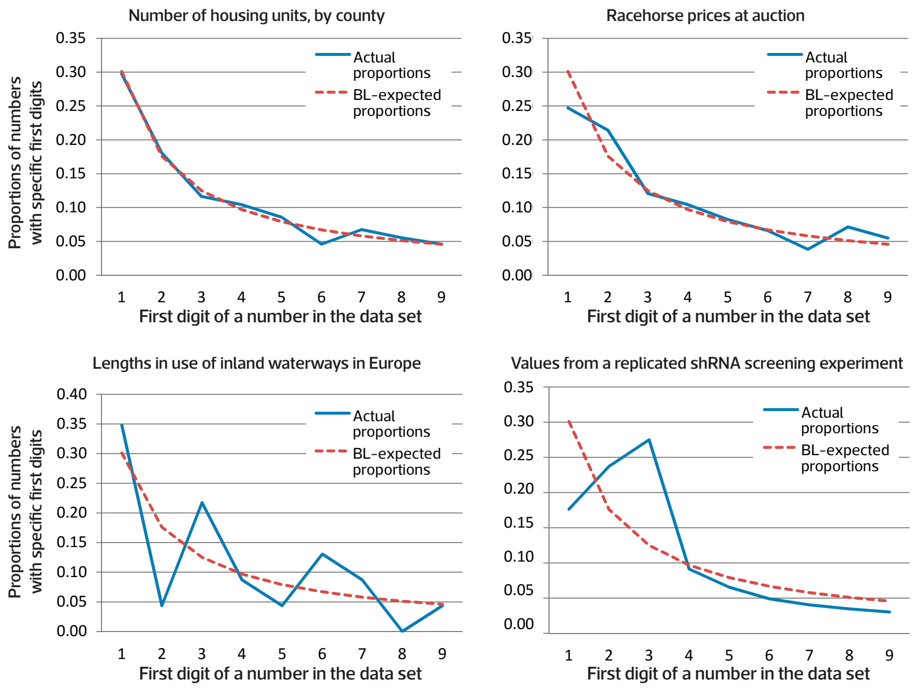

FIGURE 4 Distributions of first-digit proportions for four Benford-suitable data sets

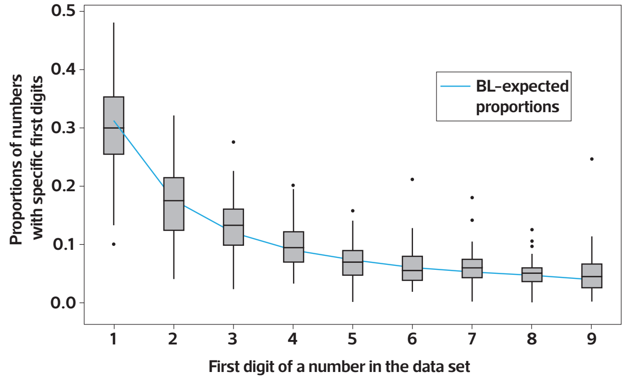

FIGURE 5 Distributions of first-digit proportions for 40 sampled Benford-suitable data sets

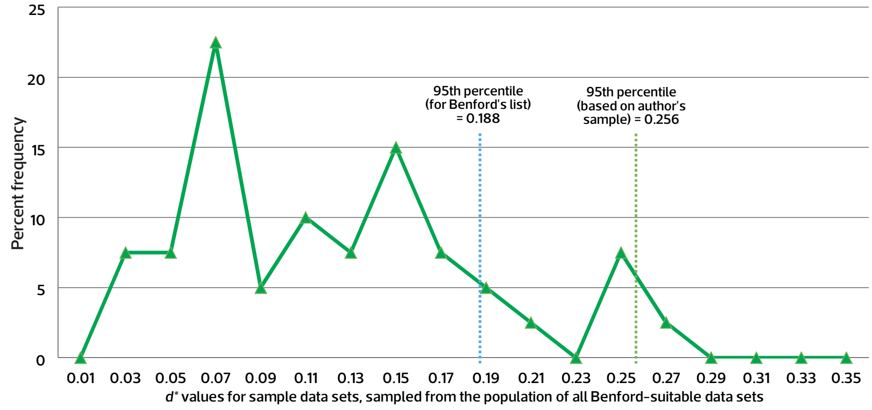

FIGURE 6 Sampling distribution for d^*, a measure of non-conformance with Benford's law

(d) *Not human-assigned numbers.* Numbers that are merely assigned, such as arbitrarily assigned telephone numbers, or "bonus points" awarded in preset amounts of \$5 or \$10, tend not to exhibit Benford patterns.

Most proposed explanations for the Benford patterns relate to the properties of numbers themselves and of various mathematical sequences.[7,8] Nevertheless, even the widely acknowledged authority, the mathematician T. P. Hill, concedes that the law has not been precisely derived, nor is it understood why some Benford-suitable data sets still do not conform to it. Indeed, Figure 4 (from cases collected by the author) illustrates that some Benford-suitable data sets fit better than others.

Fraud detector?

Given the above, it seems to be something of a stretch to say that Benford's law is in any way an automatic fraud detector, which would depend on the assumption that *all* non-corrupt Benford-suitable data sets conform to Benford's expected distribution. That premise being unsupported, some try to combine (a) an empirical observation about "convergence" for Benford distributions with (b) an assumption that (a)'s finding can be interpreted as analogous to the central limit theorem in statistics (see box).

Point (a), above, can be granted as a replicated finding: regardless of variation among individual data sets, if you aggregate many non-corrupt, Benford-suitable data sets, there appears to be collectively a kind of convergence of first-digit proportions towards those values expected by Benford's law. Figure 3 showed this for Benford's collected cases; and Figure 5 shows the same for 40 Benford-suitable data sets collected by the author. In each case, among all samples taken, the distributions of specific first-digit proportions tend to centre where Benford's law expects.

The observed convergence in (a), however, does not support the premise that all Benford-suitable data sets conform. The analogy to the central limit theorem in (b) is suggestive – but (b) does not fully follow from (a). At best, it may be acceptable to state that all non-corrupt Benford-suitable data sets fall within the sampling distribution from an infinite population of all Benford-suitable data sets, and that this sampling distribution is centred on the converged-to line of first digit proportions illustrated in Figure 5.

The Central Limit Theorem

Under suitable conditions, when sampling for the mean from a population, the means of all samples one might take are likely to be normally (bell-curve) distributed around the true population's mean value μ. Presuming that all values in any given sample are selected randomly and independently of each other, it would be expected that many values larger than μ in the sample would balance out other values smaller than μ, so the sample's overall mean (centre) would be roughly close to μ.

This modified premise, however, does not tell us the magnitude or distribution shape of the expected error for the converged-to model that is posited. Articles extolling the virtues of Benford's law as a fraud detector rarely acknowledge this oversight.

The missing error term

Without an error term it is too imprecise to say simply that a data set "does not conform". Figures 5 and 6 give strong evidence that data sets can be clean and Benford-suitable yet quite varied with respect to first-digit distributions. By how much does a data set have to differ from the exact Benford-expected values to "not conform"?

Many writers imply an answer to the question of what counts as "not (sufficiently) conforming". They use the language of statistical hypothesis testing, whereby a data set does not conform to Benford's law if its first-digit distribution differs "significantly" (by some statistical test) from the values expected. Yet such tests require that the error term (i.e., how much error would be typical and non-problematic) be made explicit.

In preparing this article, we have attempted to measure that error, using 40 collected data sets (described in full at significancemagazine.com/benfordslaw). Figure 6 illustrates how samples taken from the infinite population of Benford-suitable data sets might vary for a particular measure of a data set's overall non-conformance to Benford's law (other measures could also have been used).

The measure used for a data set's non-conformance to Benford's law, d^*, was introduced by Cho and Gaines.[9] For a data set which totally conforms to Benford's law, $d^* = 0.0$; for a data set that is as non-conforming as possible, $d^* = 1.0$. d^* is simple to calculate. For each potential first digit of numbers in the data set, find the difference between the actual and the Benford-expected proportions for that digit, and square that difference. Next, add the squared differences for all the nine first digits and take the square root of that sum, then divide the result by the largest possible value for the numerator (1.03606). A measure called chi-square (χ^2) is more commonly used; but given how it is calculated, it is misleadingly sensitive to the size of the data sets being tested.

We see empirically in Figure 6 that none of the samples conform to Benford's law completely, whereas d^* values up to 0.25 (one-quarter of the way to total non-conformance) are not unusual. Perhaps $d^* > 0.25$ might flag that a sample is "non-conforming" to Benford's expectations – but even for Benford's set of examples, d^* values up to 0.19 are not uncommon. Note that many published tests of Benford's law, which are based on *post-hoc* examinations of public records and claim to have found suspect non-conformance, actually report levels of variation that are well within the above range of ordinary variation.

> **Without an error term it is too imprecise to say that a data set "does not conform" to Benford's law. By how much does it have to differ from expected values to "not conform"?**

How does this happen? Conventional tests and software that can be used to statistically test for non-conformance build in assumptions about error terms, appropriate for those tests. For example, the chi-square model presumes that the error term is consistent with what is called a Poisson distribution. Others examine first-digit proportions, one digit at a time, to test their individual conformances to Benford's law. The corresponding test model, generally used, presumes a binomial distribution of error, often approximated by z-scores. The point is that these conventional *a priori* error assumptions are not consistent with the actual, empirically observed error for Benford's law convergence, and this impacts the validity of findings.

A cautious way forward

Despite the concerns expressed here, Benford's law may still find applications if findings are expressed more carefully, and with acknowledgement of the known (or unknown) error term. In the press-cited study of Greece's economic statistics, mentioned at the outset, its authors propose a ranking of Benford's law test outcomes to, in effect, merely suggest possible places to start for conventional audit procedures.[4] In a new, proposed application area, one author suggests using Benford's law as part of attempts to resolve disputed text authorships – if the texts include many number values for currency and other quantities – by comparing how (not just if) different number sets vary from Benford's pattern.[10]

The key is that, in itself, Benford's law is not a hypothesis test, and does not ground such tests without considerable qualification. Sometimes, it may flag things that could be found another way, such as "Why were so many expense approvals in the $9000 range?" (Perhaps because one's signing authority stopped at $10 000.) But the hope is that no one is ever formally accused, let alone convicted, based on Benford's law, without independent, and carefully thought-out, evidence. ■

References

1. Newcomb, S. (1881) Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, **4**(1), 39–40.

2. Benford, F. (1938) The law of anomalous numbers. *Proceedings of the American Philosophical Society*, **78**(4), 551–572.

3. Nigrini, M. J. (2012) *Benford's Law.* Hoboken, NJ: John Wiley & Sons.

4. Rauch, B., Göttsche, M., Brähler, G. and Engel, S. (2011) Fact and fiction in EU-governmental economic data. *German Economic Review*, **12**(3), 243–255.

5. Harford, T. (2011) Look out for No. 1. *Financial Times*, 9 September. on.ft.com/24p8tx1

6. ACL Services (2009) About Benford analysis. *ACL User Guide*. bit.ly/1rEUYvt

7. Hill, T. P. (1995) A statistical derivation of the significant-digit law. *Statistical Science* **10**(4), 354–363.

8. Miller, S. (2015) *Benford's Law: Theory and Applications.* Princeton, NJ: Princeton University Press.

9. Cho, W. K. T. and Gaines, B. J. (2007) Breaking the (Benford's) law: Statistical fraud detection in campaign finance. *American Statistician*, **61**(3), 218–223.

10. Zenkov, A. V. (2015) Deviation from Benford's law and identification of author peculiarities in texts. *Computer Research and Modeling*, **7**(1), 197–201.