

Fraud detection and num

Declan Lavelle explains how using Benford's Law can detect fraud.

IF WE WERE TO TAKE ALL THE NUMBERS appearing in this month's edition of *The Actuary* how many do you think would have a '1' as their first digit? Somewhat surprisingly, the answer should turn out to be about 30%. This is a result of a not widely known result in number theory known as Benford's Law.

Frank Benford was a physicist working with the General Electric Company in the US. In 1938 Benford noticed that the book of logarithms he used in his daily work was more finger-marked and worn for the pages corresponding to the lower digits than for the higher ones. This led him to believe that numbers with lower-valued initial digits were more commonplace in nature than a purely random distribution would imply. In fact, Benford had rediscovered a phenomenon first noted by Simon Newcomb, a mathematician and astronomer, in the late-19th century. However, unlike Newcomb, Benford went on to test an enormous set of data ranging from sports statistics and lengths of rivers to stock prices and populations of cities. Benford then published his findings in a paper entitled 'The law of anomalous numbers'.

When all the data had been compiled, over 22,000 numbers, Benford found that the frequency of the initial digits of the numbers conformed very closely to the rule $P(n) = \text{Log}_{10}(1+1/n)$, where $P(n)$ is the proportion of numbers with n as their first digit. For example, a proportion of 30.1% ($= \text{Log}_{10}(1+1/1)$) began with the digit '1'.

The full set of digit frequencies is given below

n	$P(n)$
1	30.1%
2	17.6%
3	12.5%
4	9.7%
5	7.9%
6	6.7%
7	5.8%
8	5.1%
9	4.6%

It is worth noting that Benford's Law can be extended to the second and subsequent digits of the numbers in a suitable dataset, and can also be applied to other number

bases. In base b , the frequency of digit d is $\text{Log}_b(1+1/d)$ for each $d < b$.

Why does it hold?

It is easiest to show why the law may hold by way of example. Consider a share priced at exactly £1, which is growing at a constant rate of 1% per month. The price will reach £10 after 232 months. However, the price will have been between £1 and £1.99 for 70 months (30.2%) of the time, while it will have spent only 11 months (4.7%) between £9.00 and £9.99. World data relating to share prices and market capitalisations are known to fit the Benford pattern, which is not surprising given the example above.

More rigorous mathematical work has been done. Dr Ted Hill, a mathematician from the Georgia Institute proved the law in 1996. Roughly speaking, Hill's work says that if probability distributions are selected at random, and random samples are then taken from each of these distributions, then the digit frequencies of the combined sample will converge to the Benford pattern.

Note, however, that the law will not apply to all datasets. In particular, for the law to hold, the numbers must not be restricted by artificial maximum or minimum values. If the data contains values from any given order of magnitude (in base 10), then it must contain all the values from that order of magnitude. For example, if the value 5,134 is included then all data from 1,000 to 9,999 must be included. Also, the numbers must not be invented or assigned. For example, telephone numbers will not satisfy the law as they are forced to have a fixed number of digits. The numbers must also come from a large enough sample to be statistically credible.

What has this got to do with fraud?

Probably the most high-profile application of Benford's law has been developed in the last ten to 15 years in the field of 'digital analysis'. An expert in the field is Dr Mark Nigrini who studied Benford's Law to earn his PhD. Nigrini has used the law to analyse patterns to detect fraud in accounting data.

A typical case is the analysis of expense payments. An employee with a limit of, say, £100 for unapproved expenses might try to fool the system by putting in a large number of fraudulent expense claims of, say, £95. A



Declan Lavelle is a non-life consultant with HLD Actuarial Consultants in Dublin

ber theory

Benford analysis will show the actual frequency of expenses with leading digit '9' will exceed the expected 4%–5% level.

The application of Benford's Law to accounting data forms part of the field of forensic accounting, and Nigrini's analysis techniques have been used successfully by the tax authorities of several US states and by the large international auditing firms.

Other aspects of Bedford's Law

Digital frequency analysis is not the only interesting aspect of Benford's Law. Benford's Law is very closely related to Zipf's Law and the Pareto distribution. Examples of areas where applications are being considered include the following.

- The storage of data on computers: if world data is not uniformly distributed by initial digit, there may well be more efficient ways to store data than those currently employed.
- Detection of irregularities in clinical trials and election results.
- The consistency of Benford's law with population statistics allows it to be used as a common sense check on the output of demographic models.
- It has been proposed that computer-generated images can be differentiated from 'real' images by applying Benford's law.
- It has been suggested that the length of time a customer maintains a relationship with a supplier will follow Benford's law. If this is true, then it has some significant consequences for the management of customer relationships. It may well be worth applying resources to maintaining long-serving customers rather than attempting to gain new ones, because the expected future lifetime of a loyal customer could exceed that of a newly acquired one.

Can actuaries make use of Benford's Law?

I leave this as a question for anyone interested in further investigation. Some suggestions are the detection of insurance fraud by examining claims data and the checking for mistakes in model output, and verification of model suitability.

Benford's Law may also be used for simple data diagnostics, such as the analysis of data to assess whether values are clumped around some trigger amounts. For example,

Figure 1 Data rounded near £300k

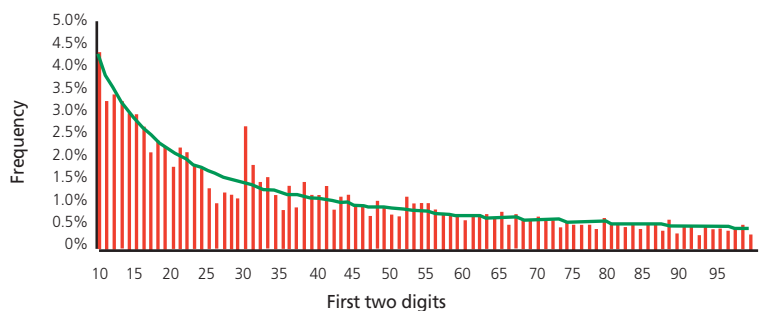
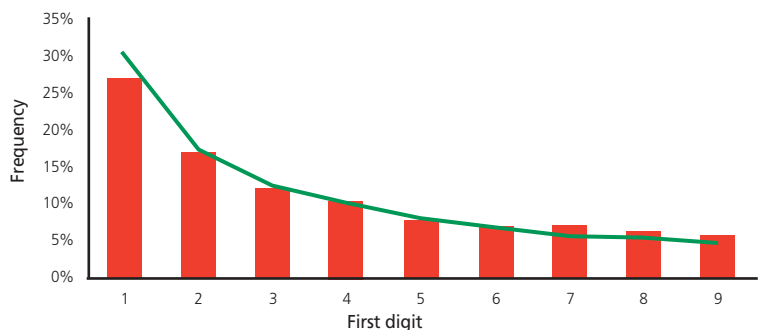


Figure 2 Data truncated to maximum five digits



claims settled by courts could tend to sit just below the limit allowable in those courts. See figure 1 for an example of a two-digit analysis applied to data where all values between £250k and £310k are amended to £290k.

Another example is the detection of flaws in data. In figure 2, data has been truncated to show only the five rightmost digits, not unknown in some older computer applications! The graph shows lower than expected frequency for low valued initial digits, and higher than expected frequency for higher-valued initial digits.

Two further uses are in extreme value theory or heavy-tailed distribution theory and in the analysis of insurance policy lapse rates and their implications for the management of customer relationships. □

Readers who would like to know more about Benford's Law are encouraged to visit Dr Nigrini's website, www.nigrini.com, for more information.