

Tommie W. Singleton, Ph.D., CISA, CGEIT, CITP, CPA, is an associate professor of information systems (IS) at the University of Alabama at Birmingham (USA), a Marshall IS Scholar and a director of the Forensic Accounting Program. Prior to obtaining his doctorate in accountancy from the University of Mississippi (USA) in 1995, Singleton was president of a small, value-added dealer of accounting IS using microcomputers. Singleton is also a scholar-in-residence for IT audit and forensic accounting at Carr Riggs Ingram, a large regional public accounting firm in the southeastern US. In 1999, the Alabama Society of CPAs awarded Singleton the 1998–1999 Innovative User of Technology Award. Singleton is the ISACA academic advocate at the University of Alabama at Birmingham. His articles on fraud, IT/IS, IT auditing and IT governance have appeared in numerous publications, including the *ISACA Journal*.



Do you have something to say about this article?

Visit the *Journal* pages of the ISACA web site (www.isaca.org/journal), find the article, and choose the Comments tab to share your thoughts.

Understanding and Applying Benford's Law

There are many tools the IT auditor has to apply to various procedures in an IT audit. Almost all computer-assisted audit tools (CAATs)¹ have a command for Benford's Law.² This article will attempt to describe what Benford's Law is, when it could apply and what constraints to consider before applying it in an IT audit.

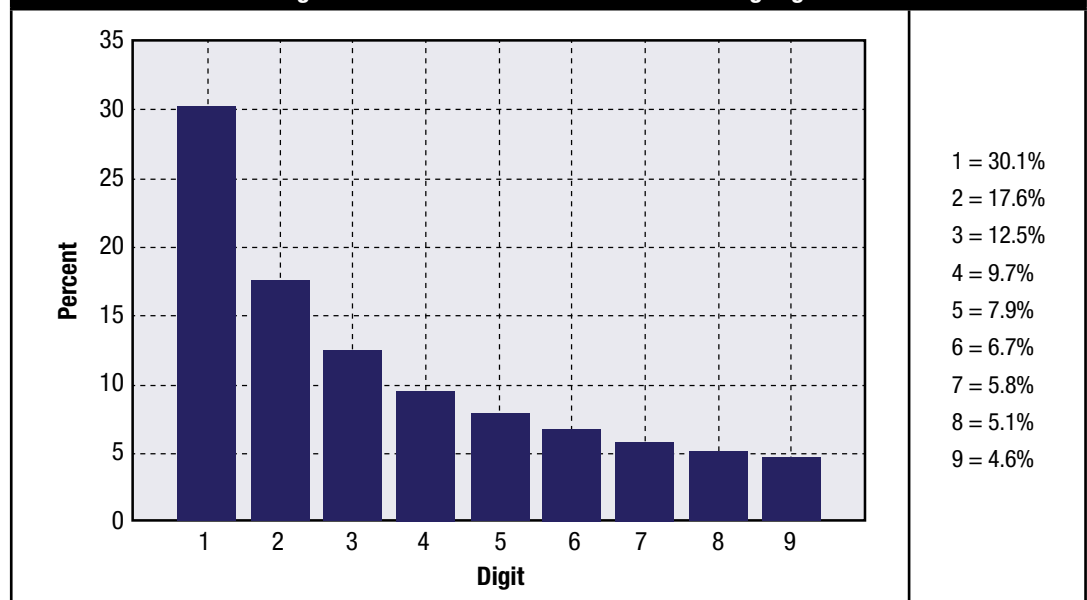
WHAT IS BENFORD'S LAW?

Benford's Law, named for physicist Frank Benford, who worked on the theory in 1938,³ is the mathematical theory of leading digits. Specifically, in data sets, the leading digit(s) is (are) distributed in a specific, nonuniform way. While one might think that the number 1 would appear as the first digit 11 percent of the time (i.e., one of nine possible numbers), it actually appears about 30 percent of the time (see **figure 1**). Nine, on the other hand, is the first digit less than 5 percent of the time. The theory covers the first digit, second digit, first two digits, last digit and other combinations of digits because the theory is based on a logarithm of probability of occurrence of digits.

Benford's Law holds true for a data set that grows exponentially (e.g., doubles, then doubles again in the same time span), but also appears to hold true for many cases in which an exponential growth pattern is not obvious (e.g., constant growth each month in the number of accounting transactions for a particular cycle). It is best applied to data sets that go across multiple orders of magnitude (e.g., populations of towns or cities, income distributions). While it has been shown to apply in a variety of data sets, not all data sets follow this theory.

The theory does not hold true for data sets in which digits are predisposed to begin with a limited set of digits. For instance, Benford's Law will not hold true for data sets of human heights, human weights and intellectual quotient (IQ) scores. Another example would be small insurance claims (e.g., between US \$50 and US \$100). The theory also does not hold true when a data set covers only one or two orders of magnitude.

Figure 1—Benford's Law Distribution Leading Digit



WHAT ARE THE RIGHT CIRCUMSTANCES FOR USING BENFORD'S LAW?

Almost from the beginning, proponents of Benford's Law have suggested that it would be a beneficial tool for fraud detection.

A recent example is Mark Nigrini's research, which showed that Benford's Law could be used as an indicator of accounting and expenses fraud.⁴ One fraudster wrote numerous checks to himself just below US \$100,000 (a policy and procedure threshold), causing digits 7, 8 and 9 to have aberrant percentages of actual occurrence in a Benford's Law analysis. Digital analysis using Benford's Law was also used as evidence of voter fraud in the 2009 Iranian election. In fact, Benford's Law is legally admissible as evidence in the US in criminal cases at the federal, state and local levels. This fact alone substantiates the potential usefulness of using Benford's Law.

Of course the usage of Benford's Law needs to "fit" the audit objective. Some uses are fairly easy to determine for fit. For instance, if the audit objective is to detect fraud in the disbursements cycle, the IT auditor could use Benford's Law to measure the actual occurrence of leading digits in disbursements compared to the digits' probability. Some good examples include thresholds and cutoffs.

For instance, if a bank's policy is to refer loans at or above US \$50,000 to a loan committee, looking just below that approval threshold gives a loan officer the potential to discover loan frauds. If loan fraud was being perpetrated,

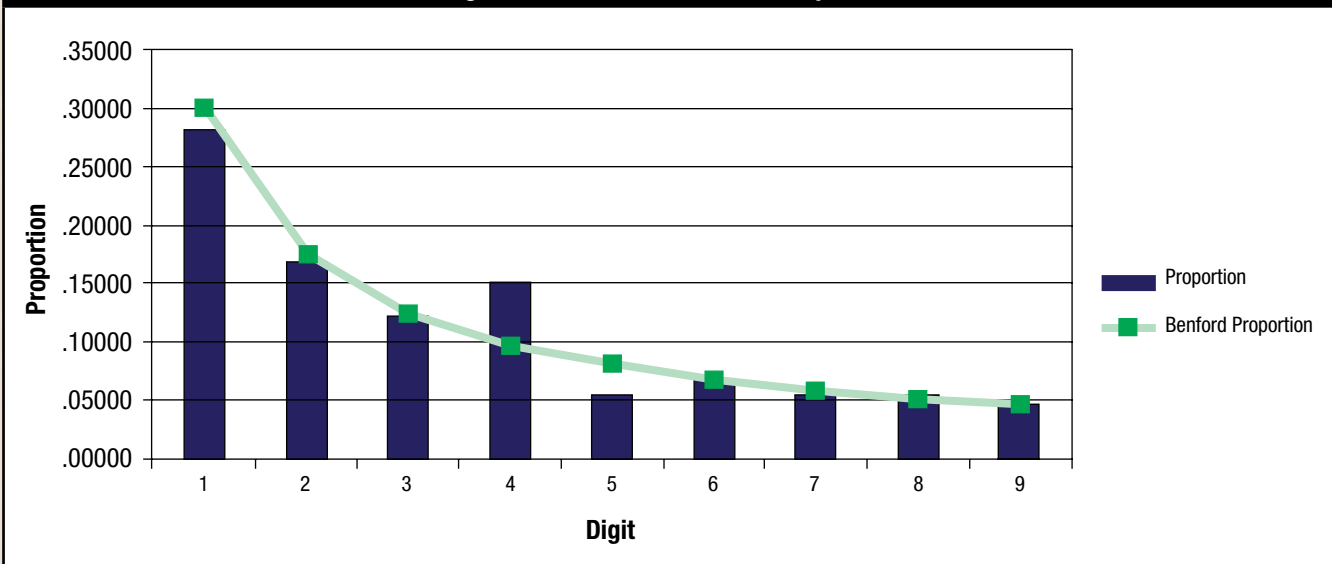
a Benford's Law test of looking at either the leading digit (specifically, the 4) or two leading digits (specifically, 49) has the potential to uncover the fraud. **Figure 2** shows what a Benford's Law test of the leading digit might show as a result in this particular scenario. The line is Benford's Law probabilities and the bars are the actual occurrences. Note that 4 is aberrantly high in occurrence, and 5 is too low, indicating the possible manipulation of the natural occurrence of loans beginning with 5 (US \$50,000 loans) possibly being switched to just under the cutoff or indicating that the suspect could be issuing a lot of \$49,999.99 loans fictitiously to embezzle funds.

Another example might be a cutoff of US \$2,500 for purchases in which a purchase order is required for any purchase at or above this price point. Thus, a Benford's Law test of the two leading digits (specifically, 24) could reveal any anomalies, manipulation or fraud involving this cutoff. It is also useful as a test of controls to see if existing controls for purchase orders are working effectively. It is important to note that since the cutoff amount has two key digits, a two-digit test is needed rather than a single leading digit.

Other objectives are equally applicable, including analysis of:

- Credit card transactions
- Purchase orders

Figure 2—Benford's Law Test/Comparison



- Loan data
- Customer balances
- Journal entries
- Stock prices
- Accounts payable transactions
- Inventory prices
- Customer refunds

Examples of data sets that are not likely to be suitable for Benford's Law include:

- Airline passenger counts per plane
- Telephone numbers
- Data sets with 500 or fewer transactions
- Data generated by formulas (e.g., YYMM#### as an insurance policy number)
- Data restricted by a maximum or minimum number (e.g., hourly wage rate)

As stated previously, the IT auditor will need to determine whether to run a one-digit test or two-digit test. The two-digit test will usually give more granular results, but is also likely to reveal more spikes than a one-digit test. For certain tests, two digits are critical (see the previous example on purchase order cutoff).

Once the test has been run, the IT auditor will need to determine what results deserve more attention or whether the results provide evidence or information related to the audit objective. Generally speaking, the spikes above the Benford's Law line are the numbers of interest (see 4, not 5, in **figure 2**). The IT auditor will want to obtain independent information on why the digit(s) spike(s). The results that show a digit that is lower than probable occurrence are generally ignored, unless the audit objective is in that direction.

WHAT ARE THE CONSTRAINTS IN USING BENFORD'S LAW?

The assumptions regarding the data to be examined by Benford's Law are:⁵

- Numeric data
- Randomly generated numbers:
 - Not restricted by maximums or minimums
 - Not assigned numbers
- Large sets of data
- Magnitude of orders (e.g., numbers migrate up through 10, 100, 1,000, 10,000, etc.) (Other assumptions exist that are unimportant in applying Benford's Law in IT audits.)

The mathematical theory has always been applied to digital analysis, i.e., a logarithmic study of the occurrence of digits by position in a number.

It is important to note that one assumption of Benford's Law is that the numbers in the large data set are randomly generated. For example, hourly wages will have a minimum and possibly some maximum (even if a realistic maximum) that means that the data set is not generated in a completely random fashion, but rather uses a restricted or manipulated set of digits as the potential leading digit. The same is true if there is a formula or structure to the manner in which the number is generated. For example, US telephone numbers are assigned with a specific area code and a limited number of 3-digit prefaces to the last 4 digits (which are the only truly randomly generated numbers in a phone number). Thus, before applying Benford's Law, the IT auditor should ensure that the numbers are randomly generated without any real or artificial restriction of occurrence.

As can be seen, Benford's Law should be applied only to large data sets. For IT auditors, that would be data such as files with hundreds of transactions (e.g., invoices to customers, disbursements, payments received, inventory items). It is inadvisable to use Benford's Law for small-sized data sets, as it would not be reliable in such cases. Thus, some experts recommend data sets of at least 100 records. This author recommends that the data set be 1,000 records or more, or that the IT auditor justify why a lower volume of transactions is suitable to Benford's Law, i.e., show that the smaller size still meets the other constraints and that size will not affect the reliability of results. The orders of magnitude in particular usually take hundreds of transactions. Using fewer than 1,000 can also lead to too many spikes of interest, too many false positives.

The IT auditor should be careful in extracting a sample and then using Benford's Law on the sample. That is especially true for directed samples in which the amount is part of the factor allowing a transaction to be chosen. This is because the sample is not truly a random sample. For example, pulling a sample of all invoices over US \$5,000 leads to a data set that is not random. For small entities, using a data set for the whole month, or a random day of each month, is a better sample for Benford's Law purposes.⁶

CONCLUSION

Benford's Law can recognize the probabilities of highly likely or highly unlikely frequencies of numbers in a data set. The probabilities are based on mathematical logarithms of the occurrence of digits in randomly generated numbers in large data sets. Those who are not aware of this theory and intentionally manipulate numbers (e.g., in a fraud) are susceptible to getting caught by the application of Benford's Law. The IT auditor can also apply Benford's Law in tests of controls and other IT-related tests of data sets. However, the IT auditor needs to remember to make sure that the constraints (mathematical assumptions of the theory) are compatible with the data set to be tested.

ENDNOTES

- ¹ For an article on using Excel formulas and commands to perform Benford's Law, see: Simkin, Mark G.; "Using Spreadsheets and Benford's Law to Test Accounting Data," *ISACA Journal*, Volume 1, 2010.
- ² Sometimes the command is referred to as "digital analysis."
- ³ Actually, Simon Newcomb was the first to posit the leading digits theory in 1881.
- ⁴ Mark J. Nigrini, "I've Got Your Number," *Journal of Accountancy*, May 1999
- ⁵ For more on Benford's Law, especially constraints, see: Hasan, Bassam; "Assessing Data Authenticity with Benford's Law," *Information System Control Journal*, 2002, volume 6.
- ⁶ *Op cit*, Simkin