The Application of Benford’s Law in Fraud Detection:
A Systematic Methodology

Nirosh Kuruppu

1 Assistant Professor, Department of Accounting, Sultan Qaboos University, Muscat, Oman

Correspondence: Nirosh Kuruppu, Department of Accounting, Sultan Qaboos University, PO Box 20, PC 123, Muscat, Oman.

Received: July 1, 2019 Accepted: September 2, 2019 Online Published: September 16, 2019
doi:10.5539/ibr.v12n10p1 URL: https://doi.org/10.5539/ibr.v12n10p1

Abstract

Benford’s Law relies on a recently proven mathematical distribution about the frequencies of naturally occurring numbers that can be efficiently applied to the detection of financial fraud. Despite the value of Benford’s Law for detecting fraud, most financial professionals are often unaware of its existence and how to best utilise the method for fraud detection. The purpose of this paper is therefore to present a systematic methodology for incorporating Benford’s Law for detecting and flagging potentially fraudulent financial transactions, that can be further investigated. This paper describes the development of Benford’s Law and demonstrates how it can be implemented systematically through a spreadsheet program to detect potential fraud. Given that the cost of financial fraud is significant with firms losing up to a tenth of their revenues, the methodology presented in this paper for implementing Benford’s Law can be a valuable tool for auditors and other financial professionals for detecting fraud.

Keywords: auditing, benford’s law, forensic accounting, fraud detection

1. Introduction

A remarkable yet recent development in the field of accounting and finance over the past decade has been the recognition of the value of Benford’s Law in assisting accountants, auditors, regulators and other interested parties in detecting fraud in accounting numbers. The ramifications of Benford’s Law for detecting fraud is significant, given the high costs and steep increase in fraud in recent times. For instance, PwC’s 2018 Global Economic Crime and Fraud Survey indicate that 49% of organisations globally reported that they have been a victim of fraud and economic crime, while 64% of respondents said losses due directly to their most disruptive fraud could reach US$1 million (Lavion, 2018). The Accountancy Daily also reports that the global impact of fraud is close $3 trillion dollars (Sweet, 2018). Another recent report entitled “The Financial Cost of Fraud 2018” by Crowe Clark Whitehill, a member of an international accountancy firm reports that organisations lose up to ten percent of annual revenues to fraud (Gee, 2018).

A common technique engaged by corporate managers to commit fraudulent financial reporting is by the creation of fictitious accounting entries, such as the recording bogus sales transactions. In a similar vein, employees may attempt at the misappropriation of assets by submitting false expense claims (Pomykacz, Olmsted, & Tantinan, 2017; Amouzegar & Moshirvaziri, 2018). The success of both these schemes depend on being able to blend the fictitious entries among real transactions in such a manner that it becomes difficult for auditors to detect the fictious entries from the real ones (Simkin, 2010; Nigrini, 2019).

However, a novel technique based on Benford’s Law has the ability to “see through” these bogus accounting entries and thereby alert auditors to potential areas of fraud so that these may be further investigated (Singleton, 2011; Collins, 2017). Benford’s Law relies on the fact that naturally occurring number distributions are vastly dissimilar to artificially created number distributions such as when sales transactions are creatively created to overstate revenue (Nigrini, 2019). Naturally occurring distributions can be observed in many accounting transactions such as the occurrence of individual sales transactions or depreciation expenses, when they are not constrained by set minimums and maximum limits (Pomykacz, Olmsted, & Tantinan, 2017). Consequently, naturally occurring distributions in accounting and finance are ideally suited to the application of Benford’s Law, which can indicate the presence of artificial numbers suggesting the likelihood of possible fraud (Singleton, 2011; Kruger & Yadavalli, 2017).
It is therefore not surprising that Generalised Audit Software available on the market has also recognised the value of Benford’s Law as a technique for detecting fraud by incorporating them within audit software. These include, for instance, the well-known IDEA™ and ACL™ audit software (Oregon Audit, 2016; Pomykacz et al., 2017). Furthermore, courts of law have accepted the use of Benford’s Law in cases involving fraud, further lending credibility to its use in auditing and forensic accounting (Pomykacz et al., 2017; Nigrini, 2019).

One of the key advantages in applying Benford’s Law is its relative efficiency, in light of the tight time pressures under which audits are typically completed. Furthermore, Benford’s Law can be easily implemented without specialist audit software through the commonly available spreadsheet programs such as Microsoft Excel, Apple Numbers or the opensource Libre Office (Kyd, 2017; Pomykacz et al., 2017). This is a significant benefit as generalised audit software incorporating Benford’s Law is expensive and often beyond the reach of many smaller and medium sized organisations. Benford’s law allows auditors and accountants to get an impression about the likelihood that a particular collection of transactions needs further investigation in a matter of minutes (Nigrini, 2018). Given the high costs of fraud experienced by organisations, effective and efficient new methodologies such as Benford’s Law analysis that can enable and enhance auditors’ ability to detect fraud is invaluable (Durtschi, Hillison, & Pacini, 2004; Collins, 2017).

This paper has a twofold objective given the above motivations for applying the Benford’s law in accounting and auditing: firstly, it explains the background and validity of utilising Benford’s Law as a technique for detecting anomalies in accounting numbers. Secondly, the practical application of Benford’s Law is demonstrated using Microsoft Excel through a structured methodology that can be implemented by accounting professionals to swiftly identify number patterns that do not fit the expected distribution given by Benford’s Law. The presented methodology will allow auditors and other accounting professionals to quickly and efficiently flag values that should be scrutinised further, forming the starting point in fraud investigation and analysis.

This paper is organised as follows. The next section presents the historical development of Benford’s Law, with a discussion on how it became recognised as a useful tool for fraud detection in accounting and auditing by reference to prior research. Following this, the application of Benford’s Law is demonstrated using Microsoft Excel so that the process can be easily replicated by accounting professionals on any number distribution to detect fraud. Benford’s Law is demonstrated on both a natural number distribution and a fictitious number distribution. Both the compliance with and the deviation from Benford’s distribution is demonstrated. The analysis and interpretation of the results of Benford’s Law is then discussed, with the final section concluding with opportunities for further research in the application of Benford’s Law as a form of Digital Analysis.

2. The Development of Benford’s Law

The distribution frequencies of naturally occurring numbers that led to Benford’s Law was first observed by Simon Newcomb, an American astronomer in 1881 (Bhattacharya, Xu, & Kumar, 2011; Singleton, 2011). He observed that books of logarithms commonly used at the time to aid in computations showed more wear in the first pages than the last ones. Newcomb supposed that more computations were done with the number 1 for instance, compared to the number 9, and came up with a simple yet remarkable formula that explained this pattern in the frequency of naturally occurring numbers (Newcomb, 1881; Matthews, 1999; Singleton, 2011; Nigrini, 2019).

Newcomb’s formula was able to effectively summarise the distribution frequencies of naturally occurring numbers. The formula took the form:

\[
P(D_t = d) = \log_{10} \left(1 + \frac{1}{d}\right)
\]

where \(d = 1, 2, 3, \ldots 9\), and \(D_t\) stands for the proportion of numbers starting with the digit \(d\).

However, without a rigorous analysis of the observed phenomenon, Newcomb’s discovery was soon forgotten (Moolman, 2017). This is despite having published his observations in the American Journal of Mathematics (Hill, 1998; Matthews, 1999; Moolman, 2017).

Nearly 60 years later on in 1938, the phenomenon first noted by Newcomb was independently “rediscovered” by Frank Benford, who was a physicist with the General Electric Company in the United States (Singleton, 2011). Benford also observed that the first pages of his logarithm tables books were more worn out compared to the last pages of the book (Benford, 1938). Since the first pages of logarithm tables books gave the logs of lower level numerals such as 1 and 2, Benford postulated that he was looking up the logarithms of lower level numbers with greater frequency simply because there were more numbers with lower level numerals in the world (Benford, 1938; Nigrini, 1999). However, unlike Newcomb, Benford attempted to investigate whether his postulate was true. He did this by examining data from many diverse sources such as geographic, scientific and demographic...
data. The data he examined contained figures from 20 separate lists with some 20,229 individual observations (Benford, 1938; Silva, Travassos, & Costa, 2017).

What Benford found out was intriguing. He observed that the leading digit number frequency distributions of diverse databases such as the populations of cities, death rates and the drainage of rivers were distributed logarithmically and not arithmetically (Silva et al., 2017). The leading digit is the first numeral in a number. For instance, the population of New Zealand is 4,746,889 million. Thus, the leading or first digit of New Zealand’s population is the number 4. The logarithmic distribution of the leading digits of numbers means that smaller numbers such as the number 1 or 2 occur more frequency in nature than, for example, the number 8 or 9. This finding goes against the intuitive expectation that one would naturally expect to observe each numeral occurring in the same proportion of around 11% (which is one out of nine possible numbers). Consequently, this also means that larger digits such as the numeral 9 are less likely to occur compared to earlier numerals in nature, and certainly does not occur 11% of the time as one would normally expect (Kruger & Yadavalli, 2017).

Figure 1 graphically presents the frequency of the occurrence of first digits as discovered by Benford, expressed in percentages. It can be clearly seen that lower level numbers occur with much greater frequency than higher level numbers.

![Benford's Law: Percentage Distribution of Leading Digits](image)

**Figure 1. Leading digits and their expected occurrence according to Benford’s Law**

It can be clearly seen from Figure 1 above that the leading numeral 1 occurs significantly more than for instance, the numeral 8 or 9. The number 1 occurs about 30.1% of the time as the first digit in a number, compared to digit 9, which occurs less than 5% of the time.

Despite the early beginning to the Benford’s law, the first robust proof as to why it holds was put forward much later. Among these initial attempts, Boyle (1994) demonstrated that dataset elements that has been processed by mathematical operators such as multiplication or division also follow the Benford’s Law. This was postulated as the reason why accounting numbers generally appear to follow Benford’s distribution (Boyle, 1994). For example, many accounting numbers are the result of the output of some mathematical operation, such as when cost is multiplied by the number of units purchased to result in total cost (Boyle, 1994; Durtschi et al., 2004). The number of units purchased forms a separate distribution to the cost of the units which forms another number distribution (Durtschi et al., 2004).

Subsequently, Hill (1995) in his seminal study demonstrated that Benford’s Law is an empirically observable
phenomena similar to the normal distribution in statistics. He showed its application using stock market data, census data and also on accounting data. Hill (1995) demonstrated that numbers following Benford’s Law are second generation distributions, which essentially mean that they are combinations of other distributions. The idea is that if distributions are selected at random, and then samples are drawn from these distributions, then the significant digit frequencies of the combined samples will converge to Benford’s distribution (Hill, 1998). This will hold even if the individual distributions may not follow Benford’s Law (Hill, 1998; Durtschi et al., 2004).

Hill (1995) was the seminal work that provided a theoretical basis underpinning Benford’s Law, as well as providing evidence from its application to both stock market and accounting data.

A number of studies have also demonstrated the practical efficacy of Benford’s Law in detecting accounting fraud. These include Carslaw (1998), who investigated and found that reported earnings by New Zealand listed companies did not confirm to the expected Benford’s distribution. Thomas (1989) extended the former study to U.S. firms. He found that a similar deviation also exists for U.S. firms, although less than reported by Carslaw (1998).

Later studies by Nigrini (1996, 1999) and (Watrin, Struffert, & Ullmann, 2008) investigated the use of Benford’s law in detecting accounting fraud, particularly tax fraud and demonstrated the potential for utilising Benford’s Law in this context. Tax returns whose first digits do not follow Benford’s Law can be identified for closer scrutiny and possible audit (Cho and Gaines, 2007; p. 9). They go on to conclude that “…the ability to ferret through millions of tax returns quickly to identify a suspicious set is an accomplishment to be hailed”.

The usefulness of the technique in Auditing was also investigated by Nigrini and Mittermaier (1997). The former study suggested the use of Benford’s Law in auditing as an analytical procedure that can assist auditors in determining the nature and extent of other audit procedures to perform during the audit. Weinberg and Busta (1998) also showed the practical value of a Benford law based analytical review procedure in auditing. In addition, the practical efficacy of the method for audits has been shown by Silva et al., (2016), who analysed 210,899 contracts issued by sixty management units in Brazil, and found discrepancies according to Benford’s law. They concluded that the procedure can improve the accuracy of sampling procedures in continuous auditing.

Despite the fact that the general benefits of Benford’s Law for detecting financial fraud has been known since the 1990s, its possible application in accounting as an efficient method for flagging fraudulent transactions is a relatively recent phenomenon (Wells, 2013; Moolman, 2017). The statement that “Benford’s law is widely applicable but not widely known…” Cho and Gaines (2007; p. 2), is a fitting description of the extent of its usage among accountants, auditors and forensic accountants. However, it is also encouraging to note that the awareness of the potential applications for Benford’s law is accelerating (Wells, 2013; Pomykacz et al., 2017; Nigrini, 2019).

Given the high cost associated with fraud, a dependable and efficient methodology that accounting professionals can use to apply Benford’s Law to flag questionable accounts and transactions will be considerably useful (Kruger & Yadavalli, 2017; IDEA, 2018). The purpose of this paper is therefore to present a structured methodology for applying Benford’s law using Microsoft Excel, that can be rapidly applied by accounting professionals to flag accounts and transactions that may require further investigation. The next section details this methodology.

3. The Application and Interpretation of Benford’s Law

A realistic dataset is important to demonstrate and validate the application of Benford’s Law in practice. This paper therefore makes use of two real world opensource accounting datasets made available by the IBM Watson Analytics platform to demonstrate both the compliance to, and deviation from, Benford’s Law.

The first IBM dataset (herein Dataset A) is a relatively large dataset containing sales data from 21 different countries with 87,894 individual transactions. This dataset is in the form of a Microsoft Excel file. The spreadsheet columns and the variables names are summarised in Panel A of Table 1 below. The second IBM dataset (herein Dataset B), is also a relatively large dataset containing the invoice amounts of 2,466 accounts receivables customers. Panel B of Table 1 summarises the spreadsheet columns and their respective variable names. The data series that is to be investigated for compliance with Benford’s Law is shown in bold italics.
Table 1. Attributes in the IBM Analytics datasets

<table>
<thead>
<tr>
<th>Panel A: Sales data</th>
<th>Panel B: Account receivables data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column</td>
<td>Variable Name</td>
</tr>
<tr>
<td>A</td>
<td>Retailer country</td>
</tr>
<tr>
<td>B</td>
<td>Order method</td>
</tr>
<tr>
<td>C</td>
<td>Retailer type</td>
</tr>
<tr>
<td>D</td>
<td>Product line</td>
</tr>
<tr>
<td>E</td>
<td>Product type</td>
</tr>
<tr>
<td>F</td>
<td>Product</td>
</tr>
<tr>
<td>G</td>
<td>Year</td>
</tr>
<tr>
<td>H</td>
<td>Quarter</td>
</tr>
<tr>
<td>I</td>
<td>Revenue</td>
</tr>
<tr>
<td>J</td>
<td>Quantity</td>
</tr>
<tr>
<td>K</td>
<td>Gross margin</td>
</tr>
<tr>
<td>L</td>
<td></td>
</tr>
</tbody>
</table>

A partial screenshot extract of the Sales dataset (Dataset A) is shown below in Figure 2. The objective for an auditor or financial professional is to assess whether the 87,894 revenue transactions reported in Column I are not fraudulent.

![Figure 2. Partial extract of Dataset A - Sales data](image)

Using the above datasets, an efficient method for applying Benford’s Law is delineated in 5 concise steps. These steps will ensure that the technique is applied efficiently to assess whether the number distribution under examination complies with the expected Benford’s distribution or not, thus flagging it for more detailed examination. Recall that the objective of the auditor to determine whether the revenues reported in Column I of the above spreadsheet is not fabricated as a way of committing fraudulent financial reporting. The steps to be followed by the auditor for applying Benford’s Law to achieve this objective are as follows.

**Step 1: Import the data**

The first step in applying Benford’s Law is to import the data distribution to be examined into Excel. It is important to ensure that the data is clean so that the applied formulas function without errors. For this demonstration, the file WA_Sales_Products.xlsx from IBM is imported into Excel, which contains 87,894 individual sales transactions. This far exceeds the minimum number recommended to apply the Benford’s distribution of 100 values (Simkin,
Step 2: Extract the leading digit

The objective is to assess if the distribution of the 87,894 revenue numbers comply with Benford’s expected distribution. As discussed earlier in the paper, the typical Benford’s Law method examines the frequencies of leading or first digits in a given distribution to see if it corresponds to the expected distribution given by Benford’s Law.

Therefore, to apply Benford’s Law to Column I containing the revenue figures, it is first necessary to extract the leading or first digit of these numbers into a separate column. This can be quickly achieved with Excel’s LEFT function. This function extracts a given number of characters from a string and takes the syntax form:

=LEFT (text, [num_chars]). Here:

- text is the location from where to extract the given number of characters. This can also be indicated by reference to the cell location containing the number from which to extract the leading digit.
- num_chars is an optional input, which stands for the number of characters to extract from the left. If none is specified, the default of 1 character will be extracted.

A column is added to the dataset to contain the extracted first digit. As shown in Figure 2 above, this added column can be observed in Column L, and is named “LEFT DIGIT” for easy identification. The LEFT formula is entered into the cell L2, and copied all the way down to extract the leading digits. In this instance, the formula will take the form: =LEFT(I2,1).

The above formula will take the value of the first revenue transaction which is in cell I2, extract the first digit in the number, and place it in cell L2. The formula in L2 is copied down to extract the remaining first digits of the revenue numbers. Once this step is completed, Column L will fill the range L2:L87895 as shown below in Figure 3:

![Figure 3](image)

**Figure 3.** Extract showing the extracted lead digits from Column I

Step 3: Determine the actual and expected leading number distributions

Once the leading digit has been extracted as above, the next step is to compare the frequencies of each of the leading digits to the frequencies expected by Benford’s Law. An efficient way to achieve this objective is to create a table on the same sheet where the leading digits have been extracted. The structure of this table is shown by the screenshot extract in Figure 4.
Once the table structure is set up as above, Columns O, P, and Q needs to be populated. Note that Column N for the leading digits and Column R for Benford’s distribution has been manually populated. Column O is the actual count of the leading digits from Column L of the spreadsheet (see Figure 2), and Column P is the percentage occurrence of each leading digit. Column Q gives the expected leading digit frequencies for the sales transactions according to the expected Benford’s frequencies in Column R.

These three columns as derived as follows. Recall that the lead digits of the sales transactions are in Column L, from cells L2:L87895 given above in Step 2.

- **Actual count in Column O:** it is necessary to count each of the digits occurring in the range L2:L87895. This is straightforward using the COUNTIF formula, which takes the form =COUNTIF (range, criteria), where:
  - the range - The range of cells to count.
  - the criteria – determines what is counted. In this instance, the leading digit.

Therefore, the formula is implemented by moving the cursor to O2 and entering the formula =COUNTIF(L2:L87895,1) to give the total count of the digit 1 as 26,113. The formula needs to be copied down to populate the count as appropriate.

- **Sample percentage in Column P:** this is simply derived by dividing the total count for digit 1 in cell O2 by the total count of all the digits (87,894) in Cell O11. The formula is initially entered in cell P2, taking the form =O2/$O$11, which also copied down. An absolute reference is used for the cell containing total count to fix its cell reference.

- **Benford’s expected count in Column Q:** the expected count column contains Benford’s frequencies for the total number of revenue transactions which appears in cell O10 (87,894). This is found by entering the formula =$O$11*R2 in cell Q2 and copying down. This gives, for instance, Benford’s Law expected count of 26,456 for transactions starting with leading digit 1, and 15,469 transactions starting with leading digit 2, and so on.

Once completed, the table of leading number frequencies will be as shown above in Figure 4.

**Step 4: Chart the actual count to the expected count of leading digits:** the objective now is to visualise if the number distribution of interest deviates from Benford’s expected distribution, indicating potential fraud that needs to be further investigated. This is effortlessly achieved by using Excel’s Combo Chart function. Ensuring that a cell within the table shown in Figure 4 is selected, go to the Insert Menu and select Combo charts. This will give a basic combo chart with both the actual count and Benford’s expected count plotted as shown below in Figure 5. The bars represent the actual observed counts of the leading digits, while the line represent Benford’s expected count.
Figure 5. Chart of actual leading digits to Benford’s expected distribution – Dataset A

**Step 5: Interpret the chart:** an accountant or auditor looking to interpret the chart needs to evaluate whether the bars representing the actual count and the expected distribution from Benford’s Law are consistent. Figure 5 clearly indicates that the actual and observed distributions are as expected given that they flow together. There are no anomalies to suggest that the actual counts of the leading digits are unnatural numbers, which indicates the possibility of potential fraud that needs to be investigated.

To also illustrate a scenario where the actual number distribution does not follow Benford’s expected distribution, a different opensource dataset on a distribution of 2,466 accounts receivables data from the IBM Watson Analytics platform was used (Dataset B). This dataset was subjected to the same five steps as shown above. Step 4 of the process leads to the following chart shown in Figure 6 below:

Figure 6. Chart of actual leading digits to Benford’s expected distribution – Dataset B
In figure 6, it can be clearly seen that the accounts receivables number distribution does not follow the expected distribution of Benford’s Law. It can be seen that the first leading digits (1-3) are well below the expected counts, whereas leading digits 4-8 are well above the counts expected from Benford’s distribution. The only exception is the occurrence of the leading digit 9. In this instance, an auditor will identify these findings as requiring further investigation due to the fact that the actual number distribution appears to have been manipulated in some way, indicating a higher risk of fraud. To an auditor this is useful information that will allow him or her to reassess the levels of inherent risk and control risk for the client, and therefore plan the audit such that sufficient and appropriate evidence is collected to keep audit risk at acceptable levels.

4. Summary and Implications

This paper discussed the usefulness and application of Benford’s Law for detecting abnormal number patterns in light of the high costs associated with corporate fraud. The fact that fabricated number distributions that underlie fraudulent financial reporting can be easily flagged by Benford’s Law has been recognised by both courts of law and also by firms developing generalised audit software. However, specialist audit software is prohibitively expensive and beyond the reach of smaller accounting firms and organisations. Consequently, a methodology that can be used to apply Benford’s Law using the commonly available spreadsheet packages is invaluable to professionals in accounting and finance. This paper demonstrated a systematic methodology that can be used to apply Benford’s Law using Microsoft Excel to relatively large real-world datasets, containing thousands of individual transactions. Both the identification of normal number patterns confirming to Benford’s Law and unnatural number patterns highlighting the possibility of fraud was explained. The methodology delineated in this paper can be efficiently applied to financial data in a matter of minutes, which can be especially useful to auditors and forensic accountants who often have to work under considerable time pressure. It can be used by auditors as an analytical procedure for getting an understanding of clients’ transactions and consequently flagging potential areas that require more thorough audit work commensurate with the identified risk. An interesting area for further research is to ascertain auditors’ perceptions about the practical utility and efficacy of applying Benford’s Law for detecting fraud compared to other available analytical procedures.

References


**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).