Benford's Law in Forensic Analysis of Registered Turnover

Edin Glogić

College for finance and accounting FINra Tuzla and
University "VITEZ" Vitez, Faculty of Business Economics
edin.glogic@gmail.com

Zoran Jasak

College for finance and accounting FINra Tuzla
jasak_z@bih.net.ba

Abstract

Forensic accounting in scientific sense is the part of accounting that assumes the practice of scientific techniques and methods in conducting investigations and detecting criminal activities in financial statements, business statements and companies. One such tool in detecting anomalies in accounting records is the Benford's Law, which gives the expected pattern of digit frequencies in numeric data sets according to their position in numbers. Because of this property, Benford's law has become a significant forensic tool for the detection of anomalies, especially in financial business. One of the most important sources is account turnover data in the observed period, which has a debt and credit side. A classic way of analyzing these liabilities is to merge them and then look for a pattern of leading digits. In such approach, it is not possible to properly determine the source of anomalies, which are a guide to deeper analysis. For such purposes, a variant of the Hosmer-Lemeshow test is designed.

Keywords: Benford's Law, Hosmer-Lemeshow Test, Forensic Accounting, Forensic Analysis.

Paper type: Non-research article

Received: May 04, 2021

Accepted: May 30, 2021

Citation: Glogić, E., Jasak, Z. (2021), "Benford's Law in Forensic Analysis of Registered Turnover", Journal of Forensic Accounting Profession, Vol. 1, No. 1, pp. 50-60.

DOI: <https://doi.org/10.2478/jfap-2021-0004>

1. Introduction

In 1881 Simon Newcomb argued in an article titled "Note on the Frequency of use of different digits in natural numbers" (Newcomb, 1881), that "*the ten digits do not occur with equal frequency must be evident to any one making much use of logarithmic tables. The first significant figure is oftener 1 than any other digit, and the frequency diminishes up to 9*", and he noticed that the first pages wear out faster than the last ones. From this, he concluded that people are more likely to use numbers that start with smaller digits. This phenomenon was rediscovered in 1938 by Frank Albert Benford, who published an article titled "The Law of Anomalous Numbers" (Benford, 1938). He acquired 20 different sets of natural numbers that were included in over 20,000 examples, from news pages to home addresses. He gave a mathematical formulation:

$$p(d) = P[D = d] = \log_{10} \left(1 + \frac{1}{d} \right) \quad (1)$$

where $d \in \{1, 2, \dots, 9\}$ is a leading digit, the first non-zero digit from left. Here $p(d)$ is an abbreviation. The following table gives probabilities for individual digits.

Table 1: Probabilities for Individual Digits

Digit (d)	1	2	3	4	5
$p(d)$	0.30103	0.17609	0.12494	0.09691	0.07918
Digit (d)	6	7	8	9	
$p(d)$	0.06695	0.05799	0.05115	0.04576	

Source: Author's processing

This is extended to groups of leading and non-leading digits. The theory has been tested on a large number of different statistical sets and it has been proven to be almost always valid, along with laws derived for digits and groups of digits on other positions. It has been observed that the larger the data mixture, the closer is the distribution of digit to the logarithmic. This is now called "Benford's Law". Also names "Newcomb-Benford phenomenon", "Digit analysis" and some other can be found in literature. Huge number of natural phenomena is in almost perfect agreement with this law, in biology, medicine, human sciences, engineering and many others. Also, there is almost no aspect of scientific or human activity for which this phenomenon is not found to be valid for. The central term for Benford's law is significant, which for number x is defined as:

$$S(x) = \frac{x}{10^{\lfloor \log_{10} x \rfloor}} \quad (2)$$

This transformation is analog to the operation of moving the decimal point in a way to get numbers with integer part consisting of one non-zero digit. For example, if $x = 273,4094$ then $S(x) = 2,734094$. Similarly, when $x = 0,003821$ then $S(x) = 3,821$. The first significant digit is now obtained as an integral part of the significand. Analogously the first two, three, ... digits, as well as digits or groups of digits to the right of the leading position can be obtained. The most part of theory for Benford's is based on significands, including base, scale and sum invariance properties. Base invariance means that the sum law holds if number are presented in any other base $B > 1$. Scale invariance means that probabilities are not changed if we multiply all numbers in sample with some non-zero value. Sum invariance means that sums of significands with the same leading digits are the same in expectation.

There are numerous methods for detection of anomalies, all of them are based on simple logic: if we have the law we need tool to detect any deviation from this law. The main aim of this paper is to represent methodology to detect deviation of this law in accounting data. Method presented in this paper is specific because data are treated in their source way of grouping on debit and credit, as standard accounting system. This allows to auditor to conduct analysis without possibility to compromise data in any way. According to this, method presented in this paper is of practical importance in analysis of financial data. This aspect is especially important in forensic accounting, in which we need to make as less as possible of transformations which can in any way to compromise source properties and relations among data. Time-line is very important aspect of accounting data.

After the 1980s, this law became a powerful tool for detecting anomalies in financial data, thanks to growing power of computing techniques, primarily on personal computers. The first known example was an analysis of Bill Clinton's tax payments made by Mark Nigrini. Regardless of the conclusion, the case showed enormous potential and an expansion of applications soon followed. The result, among other things, are well-known large financial affairs (Enron, Parmalat, Societe General, ...) detected by the use of Benford Law. In the U.S. every auditor must know how to use this tool, embedded in known program packages called Digit Analysis, or similar ones.

In Section 2 short review of literature is presented. In Section 3 methodology is presented. In Section 4 are results of one practical example. In Section 5 are conclusions.

2. Methodology

Selecting tools to analyze digit dynamics is very important aspect of every analysis. Another important aspect is selection of the analysis methodology. One of the first sources in the analysis of financial data is traffic in a given period. Their specificity is that they have a debt and credit side. Each of these two groups can be the basis for analysis, but this approach carries the risk of incomplete picture by losing some details, as time-line in the first place. As strange as that may seem, there is no basis to consider the two groups to be independent from each other. It is common practice for these two parts to be merged and treated as a unique sample, without visible separation. Such an approach has a serious drawback. It may be that the test shows no anomalies in the distribution of numbers if they exist in both groups. If the test points to anomalies, the analyst must make an additional effort to detect which side the source is on.

To overcome this, a version of the Hosmer-Lemeshow test (Hosmer, 2001) that overcomes these problems and ensures a clear indication of the source of the possible deviation has been designed. Theoretical basis and example of the application of this test is in sequel. The Hosmer-Lemeshow test was originally proposed to assess the consent of the logistics regression model in case the population is divided on two disjunctive subpopulations, partitions. Theoretical basis is text of D. Hosmer and S. Lemeshow [5]. Corner point of this method is statistics C , obtained by calculating Pearson's chi – square statistic of sample and expected frequencies in the form of $g \times 2$ table. The formula that calculates C is:

$$C = \sum_{k=1}^g \frac{(O_k - n'_k \pi_k)^2}{n'_k \pi_k (1 - \pi_k)} \quad (3)$$

Here g is the number of groups, n'_k is the number of objects in the k -th group,

$$O_k = \sum_{j=1}^{c_k} y_j \quad (4)$$

number of realizations within c_k group and:

$$\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n'_k} \quad (5)$$

is an average estimated (theoretical) probability.

The basic assumptions for this test are:

1. The sample is divided into two separate subpopulations corresponding to cases of presence or absence of a property;
2. The probabilities of the property pattern, a unique combination of predictive variables values, are π_k and $1 - \pi_k$ both for the presence and absence of some properties respectively; their sum is 1 for k-th decile;
3. Estimates of expected frequencies are $m_j \hat{\pi}_j$ and $m_j(1 - \hat{\pi}_j)$ respectively, for cells corresponding to $y=0$ and $y=1$ cases;
4. The sums of sample and expected frequencies for k-th decile are the same:

$$O_{1k} + O_{0k} = E_{1k} + E_{0k} = N_g \quad (6)$$

Here O_{1k} the O_{0k} , E_{1k} , E_{0k} , N_g stand for sample and expected values, and the number of observations in the g-th group respectively. The central problem of this test is creating groups of values. The authors proposed two strategies, in line with the original idea of the test.

3. Relation to Benford's law

In accordance with Benford's law, sample is divided into $G=9$ groups, corresponding to the leading digits $d \in \{1,2,3,4,5,6,7,8,9\}$. Some data is divided into two partitions (subsamples), marked with $Y=1$ for debit and $Y=0$ for the credit side. For the test, we need the following:

- O_{jd} : Number of observations in d group partition $j \in \{0,1\}$. Accordingly:

$$N_j = \sum_{d=1}^9 O_{jd} \quad (7)$$

- $O_d = O_{0d} + O_{1d}$: number of observations in group d. Accordingly, it is:

$$N = N_0 + N_1 \quad (8)$$

- Benford's (theoretical) probabilities in group d:

$$b_d = P[D = d] = \log \left(1 + \frac{1}{d} \right) \quad (9)$$

- $E_d = E_{0d} + E_{1d}$: the expected number of elements in group d, calculated by formula:

$$\begin{aligned} E_{jd} &= E(Y = j|d) = b_d \cdot N_j \\ \Rightarrow E_d &= E_{0d} + E_{1d} = N \cdot b_d \end{aligned} \quad (10)$$

$$N_j = \sum_{d=1}^9 E_{jd} = \sum_{d=1}^9 O_{jd} \quad (11)$$

The specifics of this case are:

- The probabilities in group d for both partitions are $b_d < 1$;
- The sum of expected values in group d does not have to be equal to the sum of sample cases:

$$E_{1d} + E_{0d} \neq O_{1d} + O_{0d} \quad (12)$$

The corresponding statistic is:

$$H = \sum_{g=1}^G \left[\frac{(O_{1g} - E_{1g})^2}{E_{1g}} + \frac{(O_{0g} - E_{0g})^2}{E_{0g}} \right] \quad (13)$$

For the purpose of the application from the Benford's Law perspective, this formula has the two following forms:

$$\begin{aligned} H_1 &= \left(\sum_{d=1}^9 \frac{O_{1d}^2}{N_1 \cdot b_d} - N_1 \right) + \left(\sum_{d=1}^9 \frac{O_{0d}^2}{N_0 \cdot b_d} - N_0 \right) \\ H_1 &= \left(\sum_{d=1}^9 \frac{O_{1d}^2}{E_1} - N_1 \right) + \left(\sum_{d=1}^9 \frac{O_{0d}^2}{E_0} - N_0 \right) \end{aligned} \quad (14)$$

4. Research results, application of Benford's law in Forensic accounting

For the purpose of method demonstration, a sample in size of 33,563 was taken, of which 21,244 were debts and 12,319 credits. Table 2 shows leading digit frequencies for all three categories.

Table 2: Leading Digit Frequencies

Digits	Exemplary frequencies			Simulated frequencies		
	All Items	Debt	Demanding	All items	Debt	Demanding
1	9.169	5.328	3.841	10.104	6.300	3.804
2	6.104	3.830	2.274	5.910	3.850	2.060
3	4.197	2.772	1.425	4.193	2.610	1.583
4	4.048	2.830	1.218	3.253	2.070	1.183
5	3.019	1.976	1.043	2.658	1.680	978
6	2.366	1.594	772	2.247	1.433	814
7	1.870	1.186	684	1.946	1.232	714
8	1.536	926	610	1.717	1.092	625
9	1.254	802	452	1.536	977	559
Total	33.563	21.244	12.319	33.564	21.244	12.320

Source: Author's calculation

Graphical presentation of this data is shown below.

Graph 1: Frequencies of leading digits on the analyzed sample

Source: Author's calculation

The first thing that can be noticed is a significant difference in frequencies for debit and credit items taken separately and together. The frequency of digit 4 is much higher in separate data in relation to the pooled sample.

The value of the chi-square statistic for the pooled sample is $\chi^2 = 416.5090$, which is significantly above the critical value $\chi^2_{7,0.05} = 14.0671$.

The value of the Hosmer-Lemeshow test statistic in this case is:

$$\begin{aligned}
 H_{11} &= \sum_{d=1}^9 \frac{O_{1d}^2}{E_1} = 21\,845,64568 \\
 H_{10} &= \sum_{d=1}^9 \frac{O_{0d}^2}{E_0} = 12\,370,9340
 \end{aligned} \tag{15}$$

$$H = H_{11} + H_{10} - N = 21845,64568 + 12370,9340 - 33564 = 652,57968$$

If we make a simulation of frequencies like in Table 2, in a way that the total number of data per partition is equal to the initial data, we get:

$$\begin{aligned}
 H_{11} &= \sum_{d=1}^9 \frac{O_{1d}^2}{E_1} = 21.249,52932 \\
 H_{10} &= \sum_{d=1}^9 \frac{O_{0d}^2}{E_0} = 12.329,54479
 \end{aligned} \tag{16}$$

$$H = H_{11} + H_{10} - N = 21.249,52932 + 12.329,54479 - 33564 = 15.07411$$

Since a critical value is $\chi_{7,0.05}^2 = 14.0671$, we need to reject the null hypothesis on conformity with Benford's Law. On the other side it is:

$$\begin{aligned}
 H_{11} &= \sum_{d=1}^9 \frac{O_{1d}^2}{E_1} - N_1 = 5.52932 \\
 H_{10} &= \sum_{d=1}^9 \frac{O_{0d}^2}{E_0} - N_0 = 9.54479
 \end{aligned} \tag{17}$$

Based on these two values, there would be no basis for rejecting the hypothesis on conformity of frequencies with Benford's Law, although that conformity actually does not exist. In other words, we would have two contrasting conclusions based on the same sample.

5. Conclusion

In this paper, the authors presented basics on the Benford's Law, together with practical aspect of the application of Benford's Law within their empirical research. By analyzing the distribution of leading digits of registered turnover data compared to Benford's distribution, its application in forensic accounting is practically presented. The presentation of the Benford's Law application as a forensic tool is presented in the empirical research carried out on a sample

in size 33,563 of registered traffic, of which 21,244 debit and 12,319 credit items. The results of the empirical investigation of registered traffic items showed certain deviations from Benford's Law, which directs the researcher to pay attention to such anomalies and to approach a more detailed analysis.

Application of this method can point to some bad account practices, like repeating or similar transactions, frequent use of small amounts and some other aspects, which have source in some internal procedures or pattern of behavior.

Benford's Law helps to a forensic accountant (auditor) to detect an anomaly (criminal action, fraud), i.e. irregularities caused by some manipulation of data in business records and possible fraud. Based on the above mentioned, Benford's Law may become the standard control tool for initial verification. Application of this law multiple, which leaves space for further research in the future.

Reference

- Konheim, Alan G. (1965). Mantissa Distributions. *Mathematics of Computation*, 19(89), 143–144. Retrieved from URL: www.jstor.org/stable/2004111 (Accessed 13.4.2021.)
- Berger, A.; Hill, T. P. (2011). A basic theory of Benford's Law, *Probability Surveys*, 8, 1-126, DOI: 10.1214/11-PS175
- Iudica, F. (2012). *Benford's Law: Mathematical Properties and Forensic Accounting Applications [Bachelor's Degree Thesis]*, LUISS Guido Carli, Facoltà di Impresa e Management, Cattedra di Matematica, Anno Accademico 2011/2012, 56 p.
- Benford, F. A. (1938). The Law of Anomalous Numbers, *Proceedings of the American Philosophical Society*, 78(4), 551-572.
- Hosmer. D., Lemeshow S. (2000). *Applied Logistic Regression* (2nd Edition). NY: John Wiley & Sons, p. 148
- Csiszar, I., Shields, P. C. (2004). Information Theory and Statistics: A Tutorial. *Foundations and Trends™ in Communications and Information Theory*, 1(4), 417-528.
- Jasak Z. (2016). Benford's Law and Hosmer-Lemeshow test. *Journal of Mathematical Sciences: Advances and Applications*, 41, 57-73.
- Jasak Z. (2015). Benford's law and arithmetic sequences. *Journal of Mathematical Sciences: Advances and Applications*, 32, 1 – 16.
- Hoey, J. (2012). *The Two-Way Likelihood Ratio (G) Test and Comparison to Two-Way χ^2 Test*, arXiv:1206.4881v2 [stat.ME].
- Shao, L., Ma BQ. (2010). First-Digit Law in Nonextensive Statistics. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 82(4). arXiv:1010.2699v1 [physics.data-an].
- Nigrini, M. (2011). *Forensic Analytics - Methods and Techniques for Forensic Accounting*

Investigations. Hoboken, N.J.: John Wiley & Sons, 144 – 146.

Sharpe, M. J. (2006). Limit Laws and Mantissa Distributions. *Probability and Mathematical Statistics*, 24(1), 175-185.

Harremoës, P., Tusnady, G. (2012). *Information Divergence is more χ^2 -distributed than the χ^2 -statistics*. arXiv:1202.1125v2 [math.ST].

Wong, S. C. Y. (2003). *Testing Benford's Law with the First Two Significant Digits*. [Masterr's Thesis]. University of Victoria, Department of Mathematics and Statistics. Retrieved from URL: <https://studylib.net/doc/18033613/testing-benford-s-law-with-the-first-two-significant-digits> (Accessed 15.4.2021.)

Newcomb, S. (1881). Note on the Frequency of Use of the Different Digits in Natural Numbers, *American Journal of Mathematics*, 4 (1), 39-40.

Miller, S. J., Nigrini, M. J. (2008). Order Statistics and Benford's Law, Research article, Hindawi Publishing Corporation, *International Journal of Mathematics i Mathematical Sciences*, 2008, Article ID 382948, doi:10.1155/2008/382948.

Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence, *Computational linguistics*, 19(1), 61-74.

Hill, T. P. (1995). A Statistical Derivation of the Significant-Digit Law, *Statistical Science*, 10(4), 354-363.

Kozlov, V. V. (2005). Vesovye Srednie, ravnomernoe raspredelyenie and strogaya ergodichnost, *Uspjehi matematicheskikh nauk*, tom 60, ed. 6(366), Nov. - Dec. 2005.

Hürlimann, W. (2009). Generalizing Benford's Law Using Power Laws: Application to Integer Sequences, *Journal of Mathematics i Mathematical Science*, 2009, Article ID 970284, 10 p., doi:10.1155/2009/970284

Sažetak

Forenzičko računovodstvo u naučnom smislu predstavlja dio računovodstva koji podrazumijeva primjenu naučnih tehnika i metoda u provođenju istraga i otkrivanju kriminalnih radnji u finansijskim izvještajima, poslovnim iskazima i kompanijama. Upravo jedan takav „alat“ u otkrivanju anomalija u računovodstvenim evidencijama je i Benfordov zakon. Benfordov zakon daje očekivani obrazac frekvencija cifara u skupovima numeričkih podataka u skladu sa njihovom pozicijom u brojevima. Zbog ove osobine Benfordov zakon je postao značajan forenzički „alat“ za detekciju anomalija, posebno u finansijskom poslovanju. Jedan od najznačajnijih izvora su podaci o prometu računa u posmatranom periodu, koji ima dugovnu i potražnu stranu. Klasičan način analize je da se ove stavke objedine i zatim se traži obrazac vodećih cifara. U takvom pristupu nije moguće na pravilan način odrediti izvor anomalija, koji su vodič za dublju analizu. Za takve potrebe je dizajnirana varijanta Hosmer-Lemeshow testa.

Ključne riječi: Benfordov zakon, Hosmer-Lemeshow test, Forenzičko računovodstvo, Forenzička analiza