Fraud Detection in Supply Chain using Benford Distribution

Tej Narain Prasad Varma [1], Dr. DA Khan [2]
Department of Computer Applications, NIT, Jamshedpur , India

ABSTRACT

The complexity of supply chains network data allows fraudsters to commit the fraud beyond the scope of internal controls. Detecting fraud by analyzing the large amounts of data is a complicated task for detecting or auditing agencies. The careful application of Benford analysis leads to identify abnormally mismatch data and in depth analysis of those data helps those agencies to perform their task more effectively, efficiently and economically within a short span of time to detect and prevent fraudulent transactions. This article demonstrates an effective approach of locating fraudulent on a data-set of supply chain network by applying statistical test on Benford distribution with help of excel sheet.

Key Words: Benford Distribution, Fraud, Supply Chain Management

Corresponding Author: Tej Narain Prasad Varma

INTRODUCTION

Supply chain management (SCM) is an integrated and complex network concept that refers to the sum of all the processes starting from the procurement of the raw material from the manufacturer/producer and ending with delivery of the end-product to the consumer (Silver, Pyke and Peterson 1998) as far back as the 1950's. Due to increased complexity of data, uncertainty risk in supply chains are growing [Christopher and Peck; Hillman and Keltz, 2007] that lead to an increasing vulnerability of fraudulent activities. The internal controls arranged for in the business process implemented in the SCM impede possible fraud, but they cannot avoid fraud altogether. Because now a days, fraud is a million dollar business and it is increasing every year. "*45% of companies worldwide have fallen victim to economic crime in 2004 and 2005. The average damage to the companies from tangible frauds (i.e. asset misappropriation, false pretences and counterfeiting) was US$ 1.7 million.*" according to the 'Global economic crime survey 2005' of Price Waterhouse Coopers. Journal headlines and news topics indicate the same trend of increasing fraudulent behavior. Given these numbers, it is remarkable that 34% of these frauds are detected by chance. Fraud can be detected in many ways or at least one tries to detect it in many ways. There are many different approaches to proactive forensic data analysis. All approaches are based on a basic set of principles and knowledge.

The Benford analysis is a powerful and relatively simple digital analysis technique that specifies the probabilistic distribution of digits for many commonly occurring phenomena for pointing out expected patterns data to create a new fraud discovery approach with help of excel sheet. Benford's distributions as an indicators of anomalous behaviour that are strong indicators of fraud. This article demonstrates an effective method of searching hidden pattern or error or fabricated data in a very large-scale auditable data-set from Supply Chain Network by applying statistical test on Benford distribution. The identification of abnormally behaved data and making in-depth analysis of those data will always lead to identify and prevent fraudulent transactions. The tool and technique explained in this article helps the audit or detective agencies to identify objectionable data in order to perform their task more effectively, efficiently and economically within a short span of time on excel sheets without applying any data analysis software (easy to understand and use). In a blind test of our approach, using data this method successfully identified actual fraudsters among the test data. In a recent article in the German Economic Review, Todter (2009) describes the potential of 'Benford's law as an indicator of fraud in economics' In this paper we demonstrate, the use of Benford's first significant digit (FSD) law as identifier of anomalous behavior of data and drill down of this data set help to detect the fraud in Supply Chain.

LITERATURE

The principles of Benford law were first published in American Journal of Mathematics during 1881. Simon Newcomb published an article in the said American journal and concluded that more numbers exist, which begin with number one than with other numbers. Dr. Frank Benford (1938) became convinced that more numbers have small leading digits, like 1 or 2, than large leading digits. Durtschi et al. (2004) review the types of accounting data that are likely to conform to Benford's Law and the conditions under which a ''Benford Analysis'' is likely to be useful. Benford's Law as a test of data authenticity has not been limited to internal audit and the attestation functions. Hoyle et al. (2002) apply Benford's Law to biological findings and Nigrini and Miller (2007) apply Benford's Law to earth science data. The mathematical theory supporting Benford's Law is still evolving. Examples include Berger et al.(2005), Kontorovich and Miller (2005), Berger and Hill (2007), Miller and Nigrini (2008) and Jang et al. (2009). Recent mathematical papers have shown interesting new cases where Benford's Law holds true, yet the tests used by auditors in practice and in published studies are the same tests advocated in Nigrini and Mittermaier (1997).A set of numbers that closely conforms to Benford's Law is called a ''Benford Set'' in Nigrini (2000, 12). A bibliography by Hurlimann (2006) lists 350 publications on Benford's law between 1881 and 2006, of which 166 appeared between 2000 and 2006. Today, Benford's law is successfully applied in many areas, from optimizing computer algorithms (Gent and Walsh, 2001) to testing eBay auctions (Giles, 2007). In business and administration it gains increasing importance as a form of 'doping control' for datasets. While sample size is important in applying the Law to empirical data, the minimum size necessary to conduct digital analysis has not been established—except that it must be large (Nigrini, 1996a; Hill, 1995). Sample sizes small as 100 have been tested with little success. Sample sizes around 500 have provided mixed results, while those above1000 have provided the best results when used with appropriate data-types (Hales etal., 2008).we examine extending the use of Benford's Law to Fraud Detection in supply chain management for vendor invoice data set of an organisation.

LITERATURE REVIEW SUMMARY

This study tends to examining how a new analytical tool, Benford's Law, can be useful to detect the fraudulent activities in Supply Chain Management. The literature is likely to confirm the Benford's Law with the conditions. Benford's Law as a test of data authenticity has not been limited to internal audit and the attestation functions but can be examined as Fraud indicator in SCM by using the 'trust-but-verify' approach that is advocated in the practitioner literature.

BENFORD'S LAW

Benford researched this assertion by studying many lists of data, such as the areas of rivers, and the atomic weights of the elements. These empirical studies led Benford to propose that in many real world applications the first digits D follow the probability distribution:

$$P(D = d_i) = \log(1 + 1/d_i) \qquad d_i \in \{1, 2, ..., 9\} \qquad (1)$$

The conditional probability of the second digit d2 (0, 1, 2, . . . , 9) is P(D2=d2/D1=d1) =log[1 + 1/(10d1 + d2)]. The unconditional probability of second digit d2 is obtained by summing over all first digits. Table 1 shows the distribution of first and second digits. In the final columns their expectation [E(d)] and variance [Var(d)] are reported. The probabilities of first and second digits decline monotonically.

Table 1 Benford distribution of first and second digits

Digit (D)	0	1	2	3	4	5	6	7	8	9	E(d)	Var(d)
Probability of 1st digit P(D)	-	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046	3.440	6.057
Probability of 2nd digit P(D)	0.120	0.114	0.109	0.104	0.100	0.097	0.093	0.090	0.088	0.085	4.187	8.254

A derivation of this formula is given in Cohen (1976). Hill (1995) extends this to the general law given by

$$P=(D_1 \cdot \cdot \cdot D_i = d_1 \cdot \cdot \cdot d_i) = \log_{10}(1 + 1/(d_1 \dots d_i)) \qquad (2)$$

Using the above formulas, we observed that approximately 30.1% of the numbers have first digit one, while only 4.6% of the numbers have first digit nine. For the second digit, this gap is not so evident, though. Indeed, approximately 12% of the numbers have second digit zero, 11.4% have second digit one and 8.5% have second digit nine. Actually, once we increase the position of the digit in a number, the gap converges to zero.

BENFORD'S DISTRIBUTION AS SUPPLY CHAIN FRAUD INDICATOR

DATA

We took a sample data from procurement cycle of Supply Chain Network i.e; invoice data to the vendors in two sets i) all vendors data of particular department of a particular organization and ii) data set of individual vendor. Although the values of individual invoice (payment) data are proprietary of the organisation.

METHODOLOGY

To test the applicability and feasibility of Benford's Law digital analysis in supply chain management, an invoiced data sample of vendors was used. Firstly we took the first data set in a excel sheet i.e; all vendors payment data of particular department and extract the first digit of this data with their frequency. We compared the frequencies of first digit with the Benford law predetermines the frequencies of first digit as mentioned in the Table 1.We observed that this data set follow the Benford probability distribution which is shown in figure 1.To test our hypotheses, whether deviations are statistically significant, we used Chi-square test of significance as shown in table 2 and found data confirm the Benford Distribution.

Then we performed the same exercise on data sets of each individual vendors and we found that there was violation of Benford probability distribution against one vendor data as shown in figure 2 and this patterns strongly reject the Benford's Law as per Chi-square test of significance, which is shown in table 3. This test indicates strong suspension against this vendor. Finally we drill down the data, where the difference between observed frequency and expected frequency percentage was high and low and then we analysed such data for fraud detection.

RESULTS AND DISCUSSION

Analysis of data set no 1

From a review the data of the table 2, it appears the data generally conform to Benford's law, although quantities with an FSD are much more common than suggested by Benford's law. Visual inspection of Figure 1 makes it very tempting to argue that the invoice first digit distribution approaches this distribution. The distribution of invoice first digit percentages with those percentages predicted by the Benford's Law is almost identical. In addition to merely visually reviewing the data, we use a Chi- square goodness-of-fit test to check the extent to which the data conforms to Benford's law. To measure the statistical significance of this comparison, we use the chi-square test. The results of this test of comparison of actuals to predicted are shown in Table 2.

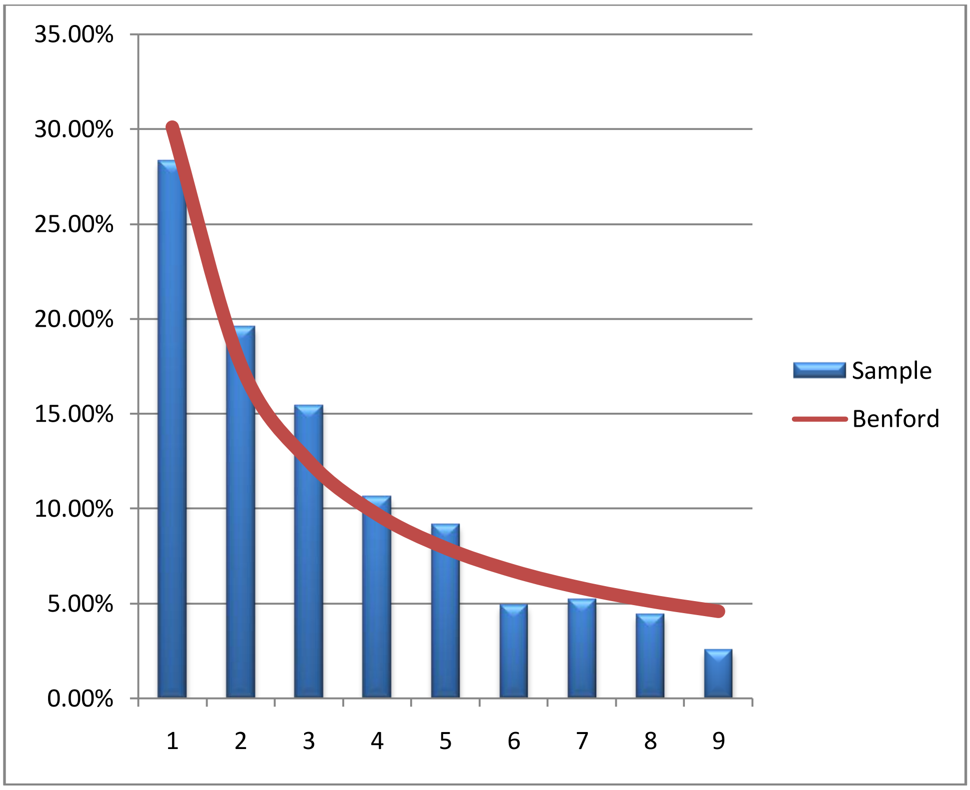

Figure 1 . Sample frequency distribution %s (y-axis) of the first digit(x-axis) of all individual invoice counts shown as line shows the expected frequencies from Benford's Law [in Eq. (1)] and the bar shows the observed frequencies from sample.

Table 2: Chi-Square Test for Actual Frequency of Invoice First Digits

First Digit	Observed %	Benford %	Difference	Squared Difference	Squared Difference / Predicted
1	28.30	30.10	-1.80	3.24	0.107641
2	19.55	17.61	1.94	3.766368	0.213887
3	15.43	12.49	2.94	8.644834	0.691926
4	10.61	9.69	0.92	0.846273	0.087326
5	9.13	7.92	1.21	1.473088	0.18604
6	4.89	6.69	-1.81	3.266041	0.487856
7	5.21	5.80	-0.59	0.348326	0.060065
8	4.37	5.12	-0.74	0.550953	0.107708
9	2.51	4.58	-2.07	4.275427	0.934366
				Total	**2.876815**

The eight degrees of freedom of observed frequency in each in the empirical data and the frequency expected by Benford. The 10 percent, 5 percent, and 1 percent critical values for are 13.36, 15.51, and 20.09.This total is smaller than the critical value for confidence level, so we at least initially perceive that we are more than confident that Benford's Law correctly describes the frequency of our first invoice digits in this. Therefore, it seems that this pattern is strong enough to correlate with Benford's Law.

Analysis of data set no 2

Then we analysed the individual vendor's payment data and found the violation of Benford's Law against one vendor which is shown in Figure 2, against digit 9. To measure the statistical significance of this comparison, we used the chi-square test. The results of this test of comparison of actuals to predicted are shown in Table 3

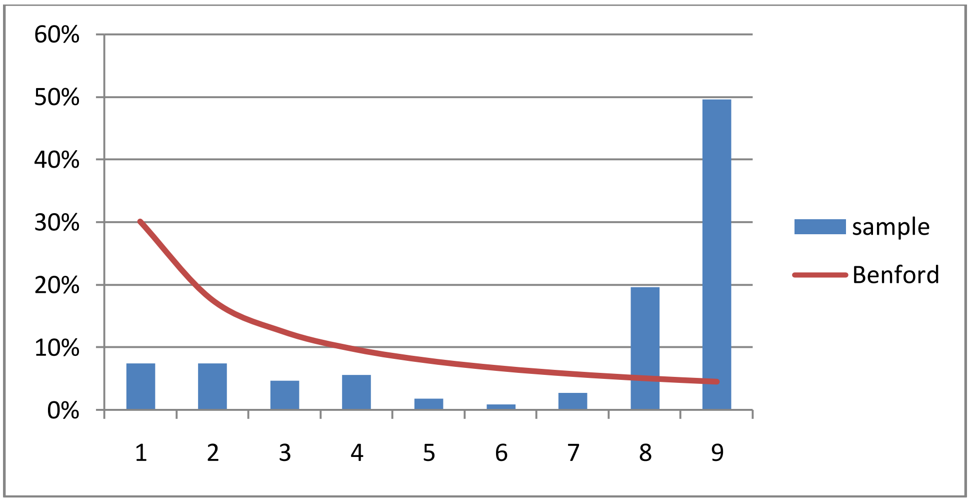

Figure 2. Sample frequency distribution %s (y-axis) of the first digit(x-axis) of all individual invoice counts shown as line shows the expected frequencies from Benford's Law [in Eq. (1)] and the bar shows the observed frequencies from sample. The frequency of the digit 1 is one since this plot concerns the first digit.

Table 2: Chi-Square Test for Actual Frequency of Invoice First Digits

First Digit	Observed %	Benford %	Difference	Squared Difference	Squared Difference / Predicted
1	7.48	30.10	-22.63	511.95	17.01
2	7.48	17.61	-10.13	102.67	5.83
3	4.67	12.49	-7.82	61.17	4.90
4	5.61	9.69	-4.08	16.68	1.72
5	1.87	7.92	-6.05	36.59	4.62
6	0.93	6.69	-5.76	33.18	4.96
7	2.80	5.80	-3.00	8.97	1.55
8	19.63	5.12	14.51	210.57	41.16
9	49.53	4.58	44.96	2021.13	441.70
				Total	**523.45**

This total is very high than the critical value at confidence level and it seems that this pattern strongly reject Benford's Law. By drilling down the data, where the difference between observed frequency and expected frequency percentage was high (first digit 9), we found the strong red flags of procurement frauds. Mostly payment for this was done between Rs. 90000 to Rs. 99999 to avoid the higher approval authorities limit i.e. one lakh. It was done by splitting the purchase order as well as framing the repeat orders for doing such manipulated payment to the vendor .On further analysis we observed the another red flags of this fraudulent activities as lack of clarity in job specification, selection of vendor without capability assessment, wrong inputs/ incomplete data in negotiation sheet to highlight capability of vendor etc.

CONCLUSION

There is no effective audit tool available as on date for identification of all types of mistakes/frauds/irregularities. In the fraudulent transactions scenario of Supply Chain, the various detection techniques for fraud can be seen as a problem of classification of legitimate transactions from the fraudulent transactions. Data analytics can be applied to detecting existing frauds hidden within procurement cycle data, it can inform fraud risk assessments and identify process and control weaknesses, and it supports the detection of non-obvious frauds. Our study shows that the typical fraud detection in supply chain by Benford Distribution is simple and easy for analysing distribution laws which helps the auditor or detecting agencies to identify the abnormal transactions and assists him to perform his task more effectively, efficiently and economically within a short span of time. The process of breaking the data and applying the Benford analysis will no doubt reduce the risk of fraud in supply chain.

REFERENCE

[1] Benford, Frank. "The Law of Anomalous Numbers." *The American Philosophical Society*78.4 (1938): 551-72. *JSTOR*. Web. http://www.jstor.org (July 15,2012)
[2] Berger, Arno, and Theodore P. Hill. 2007. "Newton's Method Obeys Benford's Law." American Mathematical Monthly 114(7):588-60I.

[3] Dr. L. Kailasam. " Benford Distribution - An Effective Audit Tool. " *The Chartered Accountant.* November 2011 : 716-720
[4] Durtschi, Cindy, William Hillison, and Carl Pacini. "The Effective Use of Benford's Law to Assist in Detective Fraud in Accounting Data." V.1524-5586 (2004): 17-34. *UIC*. Journal of Forensic Accounting. Web http://www.uic.edu/ (August 1,2012)
[5] Newcomb, Simon. "Note on the Frequency of Use of the Different Digits in Natural Numbers." *American Journal of Mathematics* 4.1 (1881): 39-40. *JSTOR*. Web http://www.jstor.org/ (August 1,2012)
[6]Nigrini, Mark J. *Datas 2009 for Excel 2007 Program Details*. Web. http://www.nigrini.com/datas_software.htm (July 12,2012)
[7] Roukema, Bouewjin. "Benford's Law Anomalies in the 2009 Iranian Presidential Election. " *Annals of Applied Statistics* (2009): 1-14. Annals of Supplied Statistics,4316 June 2000. Web. http://arxiv.org/PS_cache/arxiv/pdf/0906/0906.2789v1.pdf (July 12,2012)
[8] Saville, AD. "Using Benford's Law to Detect Data Error and Fraud: An Examination of Companies Listed on the Johannesburg Stock Exchange." *SAJEMS* 9.3 (2006): 341-54. *Scribd*. Scribd. Web.http://www.scribd.com/doc/47789223/Saville-Using-282006-29
[9] Tam Cho, Wendy K. and Gaines, Brian J. (2007) "Breaking the (Benford's) Law: Statistical Fraud
Detection in Campaign Finance," *American Statistician*, vol. 61, Issue 3, August, p. 218-223
Thomas, Jacob K. "Unusual Patterns in Reported Earnings." *The Accounting Review* LXIV.4 (1989): 773-87. *JSTOR*. Web.http://www.jstor.org (July 15,2012)
[9] Varian, H.R. (1972) "Benford's Law," *American Statistician*, vol. 26, pp. 65-66
[10] Weiss, Neil A. (1999) Introductory Statistics, Fifth Edition, Addison Wesley, Reading, MA, p.A-16
Wikipedia (2012), "Weber-Fechner law", http://en.wikipedia.org/wiki/Weber-Fechner_Law (July 12,2012)