Triage in Forensic Accounting using Zipf's Law

Adeola Odueke & George R. S. Weir[1]

Department of Computer and Information Sciences,
University of Strathclyde, Glasgow G1 1 XH, UK
george.weir@strath.ac.uk

Abstract. In forensic accounting, use of Benford's law has long been acknowledged as a technique for identifying anomalous numerical data. Zipf's law has received considerably less attention in this domain despite the fact that it is not limited to analysis of numerical datasets. The present paper outlines the context of fraud detection and then describes an experiment that contrasted Benford's law and Zipf's law as highlighters for data anomaly, with a view to enhancing current techniques in fraud detection. Results from tests on two datasets using each technique showed similarities in the samples characterized as 'fraudulent' from which we propose that, when combined with its extended realm of data applicability, Zipf's law has significant potential as an aid to fraud detection as a supplement to other analysis techniques. In particular, this approach could be employed as a component in forensic accounting triage in order to enhance the detection rate of fraud and assist in fraud prevention.

Keywords: Forensic accounting. Benford's Law, Zipf's Law, triage.

1 Fraud Categories

There are many ways that financial frauds can be committed. Some can be likened to misrepresentations affecting a companies' financial statement with the aim of deceiving members of the public or investors or asset misappropriation. Others can be attributed to corruption, theft (ghost employees, proceeds skimming, theft of company assets) or tax evasion.

The major ways a financial statement can be misrepresented include:

- Overstatement of Sales Revenue: Sales are usually the largest item in a profit and loss (P&L) account of most organisations and by overstating the sales of an organisation, perpetrators try to create a 'selling more' illusion.
- Understatement of Expenditure: For any given level of sales revenue, an understatement of expenditure is translated into higher profits or lower losses. This portrays the organisation as having cost less in the running of the business than is actually true.

[1] Corresponding author.

- Overstatement of assets: A business valuation can be done in different ways which can be generalised as earning-based and asset-based. Overstating an organisation's assets creates an illusion that the organisation has more strength and capabilities than it truly does.
- Understatement of liabilities: This act is aimed at creating an illusion where the organisation appears to owe less than it truly does.
- Presenting it better: This is a way of misclassifying accounts to aid the reader's perception of the true position of the organisation

1.2 Control as a Deterrent

Controls can help to prevent or detect fraud and the effectiveness of controls is vitally dependent on the technical ability and integrity of personnel. Controls can be classified into two categories namely internal control (established by management) and external control (established by governing bodies). Most internal controls are detective in nature, i.e. performed only after a transaction is completed and are therefore deterrent and not preventive of fraud [1]. Examples of internal controls include bank reconciliations and external audits. According to Wells [2], fraud prevention is more appealing than detection as it is often difficult to recover fraud losses once they are detected. Of course, some losses might never be recovered and the impact of fraud may go beyond direct monetary loss to include harm to reputation and loss of investor confidence.

1.3 The Financial Detective

Forensic accounting can be regarded as the counterpart of crime scene investigation (CSI), with the distinction that it involves less risk of being shot at and virtually no interaction with deceased people. Instead, the focus is on detecting and exposing complex financial frauds.

Asset control, fund control and fraud prevention and detection were the main responsibilities of accountant's until the turn of the 20th century. As accrual basis accounting became common, reporting issues became a top priority [2]. As a result, there came the need for a fraud auditor. According to Crumbley [3], the definition of forensic accounting has changed gradually from merely testifying in court to more of an investigative accountant: a financial detective with a suspicious mind. Forensic accountants usually employ a mix of accounting, computer technology, law, ethics and criminology to search for accounting irregularities [4]. The forensic accountant takes a proactive, sceptical approach to examining the books of a company and makes no assumption of management integrity by bringing to the evaluation process less concern and more interest in exposing any possibility of fraud when compared to an auditor whose main objective is to ascertain whether the reports of a company conforms to generally accepted accounting principles [4].

1.4 Impact of Technology

The ubiquity of today's IT means that it is often regarded as a way to gain competitive advantage. Management frequently finds new ways to utilize IT toward operational efficiency or as an aid in decision making. This penetration of IT can pose a major problem to the auditing profession which has led to the introduction of guidance into audits conducted in an IT environment. Various authoritative bodies, such as the American Institute of Certified Public Accountants (AICPA), the International Federation of Accountants (IFAC) and the Information Systems Audit and Control Association (ISACA), have issued standards in this area [5].

2 Fraud technology

Kimin et al. [6] wrote on the importance of a tool to extract and analyse accounting data in order to detect any financial fraud which is to be used by a forensic accountant to improve detection rate. The writers also went further to propose the use of data mining and statistical analysis techniques and proceeded to design a tool which according to them could extract data and do the analysis of the extracted data at once. The designed tool was divided into four modules which included target input (selects a target to extract information and ability to decide storing process of extracted information), extraction, analysis (use of advanced analysis functions like statistical analysis to perform functions such as sorting, adding, grouping, searching etc. on extracted data) and finally the result (displays extracted or analysed result unto a screen either graphically, in text format or imported into spreadsheets). However this approach only took into consideration the accounting system in South Korea and hence cannot be generalised as according to the paper South Korea is still lacking fundamental analysis on computerized data.

The role of knowledge discovery techniques such as data mining was also reported as a Reactive Fraud Management procedure by Phua et al. [7]. They described it as being used to implement and execute algorithmic processing and complex calculations over stored transactional data. According to the writers, instances of fraudulent actions are identified either against pre-defined fraud pattern libraries or as irregular behaviour against the accounts previous behavioural history. They also identified the challenges of reactive fraud management solutions which is the fact that they suffer due to their dependence upon labelled training data sets for assembly of the needed behavioural models against which to assess new data instances. Furthermore, the maintenance of reactive solutions was also said to present further difficulties as new data instances have to be labelled and models continually retrained for detection of the most recent fraud threats from unlabelled incoming transaction requests. Hence a significant delay is therefore incurred as an adequate number of labelled fraud cases are recognized and labelled appropriately for addition to the priming data set, during which fraud instances will go undetected and contribute to a significant financial loss [8].

The use of advanced quantitative techniques which includes sophisticated methods derived from statistics and artificial intelligence, like Neural Networks and regression analysis have also been used in areas of fraud detection.

Kaminski et al. [9], in their study on the fraud detection capabilities of ratio analysis reported about the limited ability of the technique in detecting financial statement fraud. Grove & Basilico [10] studied key ratios and posit that ratios which examined daily sales in receivables, asset quality, gross margin, sales increase and changes in accruals were more likely to detect issues of fraudulent financial reporting. In a related work, Liou [11] wrote on the possibility of financial variables effectiveness at detecting both bankruptcy and fraud. According to the writer, ratios that do an examination of capital turnover, inventory, accounts receivable, and other indicators of asset composition and liabilities (total liabilities / total assets, fixed assets / total assets, and working capital / total assets) offer better insights on fraud.

The use of discovery sampling was reported in a study by Bierstaker et al. [12]. According to the writers, 'discovery sampling is a statistical means of estimating the percentage of a population that possesses a particular characteristic or attribute', this is based on an expected error rate of zero. It is often employed when the accountant needs to know whether a population contains any error indicative of fraud. In addition, knowledge discovery techniques have also been widely applied in fraud detection for data analysis and training of supervised learning algorithms to support the extraction of fraudulent account behaviour within static data sets [8].

2.1 Digital Analysis and Fraud detection

Other researchers [13, 14 & 15] tried addressing this problem using an analytical tool called the Digital Analysis Tests and Statistics (DATAS) which is based on the theory of numbers or Benford's law. The theory posits that there are expected frequencies or occurrences of digits in a list of numbers. It works by (DATAS) identifying 'process inefficiencies, errors, and fraud by searching for abnormal:

- Digit and number patterns
- Round number occurrences
- Duplications of numbers [13, p296]

In 1881, Simon Newcomb a Canadian - American astronomer and Mathematician observed that the first few pages with low digits of logarithm tables appear worn from use when compared to the subsequent pages with higher digits. This observation led him to formulate the principle that given any data set, the percentage of numbers with leading digit "1" will be far more than any other leading digit. According to Durtschi et al. [16], there existed no theoretical explanation for the theory postulated by Newcomb which led to the article going unnoticed. Several years later, Newcomb's principle was rediscovered by Frank Benford a physicist who also observed the rate of wear on logarithm tables in 1938. He assumed that since he made use of the first pages of the logarithm table which contained low digits more often than the later pages, this according to him indicates that there exists more numbers with low digits in the world and hence came to the same conclusion as Newcomb [17]. Benford was

of the opinion that According to Benford, 'this indicates that more used numbers begin with digit 1 than with digit 9' [18, p.551].

The law was tested by Benford on various real data sets such as death rates, areas of rivers, newspaper items, a complete count of an issue of the reader's digest, etc. and all were in agreement of the law [18]. Simon Newcomb calculated the probability that a number n would assume a certain position in a number as;

$$P_{(d)} = log_{10}(1 + \frac{1}{d})$$

Where d is any number from 1 to 9 and p is the probability [16]. The above formula was also used by Benford to calculate the frequency of the first digit [18]. According to Benford, the frequency Fb of a second-place digit b following first-place digit a is

$$F_b = \log\left(\frac{ab+1}{ab}\right) / log\,\frac{a+1}{a} \quad [18, p.555]$$

When applied to accounting, auditors would analyze an entire account to see if it follows the Benford's distribution [16].

Consider the market value of a firm that is growing at 10% per year. When the total assets are $1million, the first digit of total assets is *1*. The first digit will continue to be *1* until total assets reach $2million. This will require a 100% increase (from *1* to *2*), at a growth rate of 10% per year, will take about 7.3 years (with compounding). At $5million the first digit will be *5*. Growing at 10% per year, the total assets will rise from $5million to $6million in about 1.9 years, significantly less time than assets took to grow from $1 million to $2 million. At $9 million, the first digit will be *9* until total assets reach $100million, or about 1.1 years at 10%. Once total assets are $100million the first digit will again be *1*, until total assets again grow by another 100%. The persistence of a *1* as a first digit will occur with any phenomenon that has a constant (or even an erratic) growth rate [17]. Invariably, the likelihood of occurrence of smaller digits is much more persistent than larger digits. It follows that a distribution of any financial data would obey the Benford rule.

Boyle [19] showed a proof of datasets conforming to Benfords's law. Several scholars have written about the applicability of Benford's law to physics (speed of light & force of gravity), radioactive half-lives of unhindered alpha decays [20]. Hill applied Benford's law to stock market data, census statistics and certain accounting data. The use of Benford's law which is one of the most widely used power laws for fraud detection was also demonstrated by [17, 16 & 12]. They considered the effectiveness of the theory in detecting fraud and also the limitations to the types of fraud that could be detected by Benford's distribution and those that couldn't be detected by such analysis. According to Durtschi et al. [16], the result of the analysis as in any other statistical test is derived by comparing the actual number of item observed to the expected (Benford's law) and calculating the deviation.

Bierstaker et al. [12] also went further to state the type of fraud that couldn't be detected by data analysis as a result of the data sets under examination not being

appropriate for the intended analysis technique. According to [12], digital analysis will not detect frauds such as contract rigging, defective deliveries, or defective shipments. Also, duplicate addresses or bank accounts cannot be uncovered. Silvio [21] cited in [22] also wrote about the limitations of Benford's distribution on amounts which are closely and mostly identical.

Durtschi et al. [16] also considered the likely limitations of analysis based upon Benford's law and suggest that datasets which are comprised of assigned numbers (i.e. numbers which are influenced by human thought) are less likely to be detected by Benford's analysis. According to the authors, this might result in the generation of too many cases to review. Hence as effective as the theory has been said to be at facilitating auditors review on overwhelming volumes of data and transactions, its reliability is questionable and also care has to be taken in the interpretation of the result of the test. As a result there is the need for some filter to narrow the datasets to a manageable size and hence reduce some noisy data.

Zipf's law, proposed by George Kingsley Zipf in 1938, states that in a collection of any natural language, 'the frequency of any word is inversely proportional to its rank in the frequency table' [23 & 24]. If f1 is the most common term in a collection f2 is the next most common etc, then the collection frequency cf_i of the i_{th} most common term is proportional to 1/i.

$$cfi \propto \frac{1}{i}$$ (where i is the assigned rank 1,2,3,4,…,n)

Equivalently, Zipf's law can be rewritten as $cfi = ci^k$

or as $logcfi = logc + klogi$ (where and c is a constant)

From the above equation, when i = 1, it implies the most frequent word occurs 1time, when i=2, the next most frequent word occurs ½ the times of the 1st, when i=3, the next most frequent then occurs 1/3 of the first and so on. This implies that the frequency decreases rapidly with the rank.

Under review as a potential tool for fraud detection with the possibility of filtering dataset to a manageable size is the use of Zipf's law [22]. According to these authors, the basic concept of Zipf's law is based on the frequency of the word occurrence in an article being inversely proportional to its rank which helps to manage word significance. This was demonstrated in their [22] research whose main objective was to introduce an innovative fraud detection mechanism on the basis of Zipf's law. They stated that the main purpose of the technique was to assist auditors in managing the volumes of data sets and transactions with the aim of identifying and possibly reduction of the noisy data. It is believed by numerous scholars that Benford's law is a special case of Zipf's law [23]. According to Terence [24] Zipf's law can go beyond application to numeric attributes to verify the frequency of string or date fields.

In their study, Pietronero et al [23] explored an innovative analytical procedure based on Zipf's law to verify the frequency of string or date attributes, and to further confirm its ability to detect fraud records. One of the major benefits of Zipfs law was also demonstrated in the work of Situngkir & Surya [25] as its ability to verify diverse attributes other than numeric attribute which is the case in Benford's law. This

technique is expected to be able to improve the performance of analytical procedures for auditors [22].

Benford's law which is the technique being used today by many software vendors has some identified limitation which has already been discussed above. An example can be seen in the area of pricing psychology where a product could have a price of £199 just to create a make believe that it is below £200 in order to get buyers attention. Benford's law would characterize a couple of such amount as fraudulent based on the expected percentage of the probability of a 9 occurring as the second digit. As a result, there arise the need for a better technique to identify various classes of fraud in order to reduce the rate of false negatives and false positives. Zipf's law would be used with the aim of addressing the identified limitations of Benford's law. Based on certain features of the underlying technology in the Zipf's implementation, it is expected to work both as a preventive and detective mechanism for fraud.

3 Experimental comparison

Using the Python programming language, we implemented the analytical functions based upon Zipf's law. In order to apply these comparisons on sample data, we worked with the open source software Picalo (developed by Conan Albrecht[2] with the aim of providing a tool for fraud examiners, auditors and data analysts which would assist in detecting anomalies during the process of their data analysis). Picalo, which was also written in Python, has a range of features including:

- Ability to load data into the Picalo repository in various formats and from different sources (spreadsheets, emails and database).
- Offers various analysis options, one of which is the detectlet wizard which gives non-programmers the opportunity to run analysis routines developed by others.
- Contains a library of analytical functions.

In addition, a major reason behind the choice of Picalo as the analysis is that it comes ready loaded with a detectlet based upon Benford's law. Coupled with our Zipf's law detectlet, we could use Picalo to compare the performance of these two techniques in highlighting anomalous data items.

Each test would involve the analysis of a dataset using Benford's law followed by analysis of the same dataset using Zipf's law. The result from these tests was compared to determine the similarities and contrasts.

The first step is the computation of actual frequency of occurrence (fi) for each pattern. This would involve the sorting of the patterns in order of their frequency next we assign a corresponding rank Ri to each pattern starting from 1 upwards i.e. the highest frequency will be assigned rank 1, followed by the next highest which will be

[2] http://www.picalo.org/

assigned a rank of 2 and so on. Subsequently, we determine the expected frequency of occurrence of the analysed attributes as is predicted by Zipf's Law. The purpose of this stage is to distinguish abnormal data from the observed value. Zipf's law states that 'the frequency of a word is inversely proportional to its statistical rank r' such that

$$p(r) \approx \frac{1}{r \ln(1.78R)}$$

Where r is the rank and R is the count of all words.

3.1 Anomaly Detection

The detection of anomalies is derived from the difference between the actual and the theoretical frequency. From the result at this 'triage' stage, further statistical tests such as z-test and the confidence interval would be required to enhance the search for fraudulent data. Zipf's approach, implemented as a detectlet, is compared to the existing Benford's approach on Picalo, with the two analytical techniques being tested on two different datasets. The results obtained from these tests are then compared.

3.2 Datasets

The datasets used for our comparative analysis are provided as part of the Picalo distribution. We used the procurement dataset and the invoice dataset. Although these datasets have been created artificially and are not drawn from a real world setting, this does not affect our purpose. At this stage, we wish only to consider how the Zipf's law analytical approach to anomaly detection compares to the Benford's law approach. With the procurement dataset, the aim is to check the data on employee purchases in order to highlight any that deviate from 'expectation' (where expectation is defined by the respective analytical function).

3.3 Test 1

The guide to performing the Benford's analysis in Picalo was obtained from the Picalo Workbook[3]. From the displayed output it was observed that the first twenty account numbers which shows significant deviation from expected (with the lowest percentages) are 2038, 2058, 2082, 2065, 2083, 2088, 2028, 2047, 2054, 2018, 2000, 2071, 2016, 2099, 2003, 2059, 2026, 2050, 2021, 2005 and 2082. From the result obtained, one could conclude that the identified numbers from their low Benford's expected number indicates that the numbers might have been generated because the numbers do not match Benford's Law (i.e. not real transactions). According to

[3] http://www.picalo.org/download/PicaloWorkbook.pdf

Benford's law, the highest percentage that can be obtained by any row is 0.30. Hence the lower the percentage, the higher the chance the numbers are made up.

The implemented Zipf's approach was used to analyse the same dataset. The technique involved to analyse the data for the possibility of fraud involves expressing the expense on an account as a function of its rank. This requires the count and sorting of account number in decreasing expense on the account. The unique accounts are then ranked in order of the corresponding amount from highest to lowest. Thereby, account number 2082 with sum of 118499 will be ranked 1 and so on down the line in decreasing order.

From the displayed result of the analysis, we also selected the first twenty records which showed significant deviation. The figures presented were: 2082, 2088, 2050, 2040, 2025, 2005, 2012, 2003, 2071, 2000, 2026, 2018, 2058, 2099, 2072, 2080, 2054, 2079, 2052 and 2016.

The distinct number of accounts in the dataset is 100. Comparing the two results obtained above, we observe that out of the first 20 records obtained from each analysis technique 13 records appear in both results (although their order of occurrence differs). It was also observed that the similarity between the two approaches in terms of accounts showing some level of deviation recognised by each approach increases as one further extends the selected sample (e.g., selecting 25 records, 17 identical records were obtained). This suggests that the 'agreement' between the techniques persists for those accounts with lesser 'strength of anomaly' (and may not be sufficient to suggest the occurrence of fraud on those accounts).

3.4 Test 2

The aim of the test on this dataset is to check for contractor IDs that deviate from the expectation (and might signify a ghost contractor). Applying the Benford's analytical approach, it was observed that the first twenty Contractor IDs which show significant deviation from expected (with the lowest percentages) are 5177, 8017, 5005, 5075, 5081, 5062, 5008, 5127, 5155, 5171, 5056, 5001, 5160, 5242, 5052, 5110, 5036, 5237, 5180 and 5218. From the result obtained, one may conclude that the identified numbers from their low Benford's expected number indicates that the invoices from the above Contractor Ids might have been artificially (rather than naturally) generated.

When we applied the Zipf's law technique to this dataset, the first twenty records which showed significant deviation were: 5255, 5032, 5018, 5237, 5171, 5008, 5213, 5233, 5145, 5098, 5062, 5102, 5067, 5087, 5194, 5250, 5005, 5160, 5110 and 5155.

For this dataset, the distinct number of contractors is 64. Comparing the two sets of results (those from Benford's approach and those from Zipf's approach), we observe that of the first 20 records in each, only 8 records appear in both results though their order of occurrence differs. Notably, the similarity between the two approaches in terms of accounts showing some level of deviation increases as one further extends the selected sample (e.g., when considering 25 results, 14 identical records were obtained).

4 Conclusion

In this paper we introduced key concepts in financial fraud detection, with a specific focus on the application of digital analysis techniques. To shed light on the prospects of using Zipf's law as a means of highlighting anomalies in sample data, we undertook an experimental comparison of this analytical approach with the more commonly acknowledged Benford's law. This was implemented via the open source software application called Picalo and used two datasets that are provided with this software tool. On the basis of these tests, the implemented approach (Zipf's law) and the existing approach (Benford's law) show some level of similarity in their results. This can be taken as some indication that Zipf's law has applicability in fraud detection, in like fashion to Benford's law.

In order to achieve further discrimination between these two approaches, we would require additional datasets from 'natural' contexts for which the anomalous values were already known (through reference to the real world sources for the figures). Thereby, we may independently evaluate the efficacy of each analytical method.

Pending such extensions to this work, we may only conclude that Zipf's law has potential in conjunction with other techniques for anomaly detection, to serve as a means of highlighting 'suspicious' data values. While its effectiveness in contrast with Benford's law remains unproven, the fact that Zipf's law is not limited to numeric data may yet give it an edge that takes its prospective anti-fraud use beyond the numerically bound Benford's law. For the present, both Benford's law and Zipf's law may be used as companion tools toward triage in forensic accounting. In conjunction with other fraud detection techniques, they enable auditors to build a level of confidence in their result.

References

1. Huntington, I.K., (1992). Fraud: Prevention and Detection, United Kingdom: Butterworth & Co (publishers) Ltd.
2. Wells, J.T. (2004). New approaches to fraud deterrence, Journal of Accountancy, Vol. 197, pp.72-76.
3. Crumbley, D.L. (2001). Forensic Accounting: Older Than You Think. Journal of Forensic Accounting. 2: pp.181–202.
4. Crumbley, D.L. & Apostolou, N. (2002). Forensic Accounting: A New Growth Area in Accounting. Ohio CPA Journal, July /September: p.16.
5. Yang, D. C & Guan, L. (2004). The evolution of IT auditing and internal control standards in financial statement audits. The case of the United States. Managerial Auditing Journal, 19 (4), pp.544-555.
6. Kimin S, Jaemin C, Yong-seok C, Dong-chan L, Sangjin L. (2009). Research about extracting and analyzing accounting data of company to detect financial fraud, in IEEE International Conference on Intelligence and Security Informatics, ISI '09, pp.200-202.
7. Phua, C., Lee, V., Smith, K., & Gayler R. (2005). A comprehensive survey of data mining-based fraud detection research. Available at http://www.bsys.monash.edu.au/people/cphua/.
8. Edge, M. E. & Sampaio, P. R. F. (2009). A survey of signature based methods for financial fraud detection, in Computers & Security 28(6): pp.381-394.

9. Kaminski, K.A., Wetzel, T.S., Guan, L (2004). Can financial ratios detect fraudulent financial reporting? in Managerial Auditing Journal, Vol. 19 (1), pp.15-28.
10. Grove, H., and Basilico, E., (2008). Fraudulent financial reporting detection: Key ratios plus corporate governance factors. International Studies of Management and Organization 38(3): pp.10-42.
11. Liou, F-M., (2008). Fraudulent financial reporting detection and business failure prediction models: A comparison. Managerial Auditing Journal 23(7): pp.650-652.
12. Bierstaker, J.L., Brody, R.G. & Pacini, C. (2006), Accountants' Perceptions regarding Fraud Detection and Prevention Methods, Managerial Auditing Journal, 21(5), pp.520-535.
13. Lanza, R., (2000). Using digital analysis to detect fraud, in Journal of Forensic Accounting, I(2), pp.291-296.
14. Richard, B. (2000). Using Digital Analysis To Detect Fraud: Review of the DATAS Statistical Analysis Tool. Journal of Forensic Accounting, 1, pp.291.
15. James, A, (2007). Digital Analysis: A Better Way to Detect Fraud. The Journal of Corporate Accounting & Finance, DOI 10.1002/jcaf.20305, pp.27 – 36.
16. Durtschi, C., Hillison, W., & Carl, P. (2004). The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. Journal of Forensic Accounting. Fraud Examiners Manual. 2010. Austin, TX: Association of Certified Fraud Examiners (ACFE).
17. Nigrini, M. J. (1999). I've got your number: How a mathematical phenomenon can help CPAs uncover fraud and other irregularities. Journal of Accountancy, May 1999. Available at: http://www.journalofaccountancy.com/Issues/1999/May/nigrini.htm.
18. Benford, F. (1938). The Law of Anomalous Numbers. Proceedings of the American Philosophical Society 78(4): pp.551-572.
19. Boyle, J. (1994). An Application of Fourier Series to the Most Significant Digit Problem. The American Mathematical Monthly 101(9): pp.879-886.
20. Hill, T. P., (1996). "A Statistical Derivation of the Significant - Digit law". Statistical Science. 10(4): pp.1-21.
21. Silvio, C., (2004). Using digital analysis to detect fraud. Proceedings of 10th Annual ACFE Canadian Fraud Conference.
22. Huang, S.-M., D. C. Yen, et al. (2008). An investigation of Zipf's Law for fraud detection (DSS#06-10-1826R(2)). Decision Support Systems 46(1): pp.70-83.
23. Linkage. (1999). Zipf's law. http://linkage.rockefeller.edu/wli/zipf/index_ru.html.
24. Black, P.E. (2009). Zipf's law, in Dictionary of Algorithms and Data Structures, P. E. Black, (Ed.), U.S. National Institute of Standards and Technology. Available at: http://www.itl.nist.gov/div897/sqg/dads/HTML/zipfslaw.html
23. Pietronero, L., E. Tosatti, et al. (2001). Explaining the uneven distribution of numbers in nature: the laws of Benford and Zipf. Physica A: Statistical Mechanics and its Applications 293(1-2): pp.297-304.
24. Terence, T., (2009). Benford's Law, Zipf's law and the Pareto distribution. Available at: http://terrytao.wordpress.com/2009/07/03/benfords-law-zipfs-law-and-the-pareto-distribution/.
25. Situngkir, H. & Surya, Y. (2005). What can we see from investment simulation based on generalized (m, 2)-Zipf law? Bandung Fe Institute Working Paper No.WPE2005, Social Science Research Network.