Slovak University of Technology in Bratislava
Faculty of Mechanical Engineering
17th Conference on Applied Mathematics **APLIMAT 2018**
Proceedings

VERIFYING THE AUTHENTICITY OF THE ACCOUNTING DATA USING THE BENFORD'S ALGORITHM

KOŠŤÁLEK Josef (CZ), KOŤÁTKOVÁ STRÁNSKÁ Pavla (CZ)

Abstract. This article describes examples of such algorithms, which is basing on the knowledge of American physics Benford. Benford's law describes the behavior of empirical values, when the size of the relative frequency digits forms a numeric or financial data. The aim of the article is a description of the model that is capable of decomposition on single digit figures. The digits are calculating relative to frequency and those are then comparing to the probabilistic model respectful of Benford's law.

Keywords: analysis of the accounting data, Benford's distribution, Benford's law, first digit distribution

Mathematics Subject Classification: Primary 62E99; Secondary 62P99, 62P05.

1 Literature review

The initial scientific studies dealing with the application the Benford's law in the area of accounting and taxation issues come from authors [3,11], and the collective [4,14]. They followed the articles on mathematics [5,12]. At present, the scientific studies focus on the application of the law in the area of Benford macro-economic data, and government economic data, especially in studies of [12] will be used in the article named On the Application of Benford's Law's International Macroeconomic Statistics published in the magazine of the German Economics. Finally, it is necessary to mention also the credibility of scientific results, for example the regression coefficients. In scientific study [13] dealt with the place of convergence test data to evaluate the Benford distribution of national according to the width of the data from the Benford distribution of variances, respectively, the average result of χ^2 test in the field of national accounting. The results of the scientific research of the 27 member States of the EU have shown that the Czech Republic showed the smallest diameters of the χ^2 test. The research focused on the 27 Member States. The resulting findings have been that of all the EU Member States showed a significant deviation from the Benford distribution, namely Greece and Ireland.

1.1 Theoretical basis of Benford´s law

The model uses Benford's law of frequency distributions of numbers in the first place of the numeric data. The likelihood that the figure will begin with the digit 1 is not as high as the probability that will begin with the digit 9. This probability according to [1,5,6,7,8] (see Tab. 1. and formula (1)).

Digits	1	2	3	4	5	6	7	8	9
Probability (%)	30.10	17.61	12.49	9.69	7.92	6.70	5.80	5.11	4.58

Tab. 1. Benford's probability distribution of first digits.

The values (of Tab. 1.) are calculated using (1), in [2,8].

$$P(\%) = log\left(1 + \frac{1}{d_1}\right) \times 100 \text{, where } d_1 \in \{1, 2, 3, \ldots, 8, 9\} \quad (1)$$

In addition to the likelihood of the digit in the first place can be formula (1) extend to formula (2), which calculates the probability of occurrence of the first two digits. We calculate so eg. What is the probability that a numeric value begins with the value 11 or 36.

$$P(\%) = log\left(1 + \frac{1}{10 \times d_1 + d_2}\right) \times 100 \text{, where} d_1 \in \{1, 2, 3, \ldots, 8, 9\} \text{ a } d_2 \in \{1, 2, 3, \ldots, 8, 9\} \quad (2)$$

These are similar s for determining the probability of the first three digits, etc., but here the differences in probabilities are no longer so much contrast. In formula (3) it used to calculate the probability of the occurrence of digits in second place. The probability that it will be the second, for example the number 3 is the sum of the probabilities of scenarios: 13, 23, 33... 83, 93.

$$P(\%) = \sum_{d_1=1}^{9} log\left(1 + \frac{1}{10 \times d_1 + d_2}\right) \times 100 \text{, } d_1 \in \{1, 2, 3, \ldots, 8, 9\} \text{ a } d_2 \in \{1, 2, 3, \ldots, 8, 9\} \quad (3)$$

The use of these patterns has its limits. Since the probability is necessary to ensure the most figures. This rule States that the basic statistical (4). To prove the measurability and Benford's distribution recommended at least 200 figures, in [2,4,5,6].

$$\lim \frac{n_i}{n} \rightarrow p \quad (4)$$

The input condition is that the figures include at least three digits.

1.2 Principle of functioning of the model

The model created in SW MS Excel in which accounting data entered. Then automatically verify the accuracy of data according to Benford's laws. The first digit validates of the frequency, the frequency of the first two digits and the frequency of the second digit. The Benford's distribution verified with of statistical tests: χ^2 test and Kolmogorov-Smirnov test when the specified confidence level "α", which can specified as a parameter in the model. [9]

Created by the program must first take the figure and divide it into individual digits, in order to calculate the frequency of these digits. If the input entered is a negative number, the absolute value created from it. If the specified entry has only one number, it is automatically skip. Decimal places deleted. First, it is necessary to eliminate incorrect values from the input data. The following algorithm is in a form (5).

$$IF(\textit{size of the input sample of data}; S\ VYHLEDAT\ (\textit{serial number of the input data}; \textit{search area}; \textit{column to select rounded data}; \textit{untruth}); \textit{empty cell}) \quad (5)$$

After the elimination of incorrect values from the input data are calculated the frequency digits. It shows (see Fig. 1.) the result of the frequency digits in first place compared with theoretical – Benford´s values (according to formula (1)).

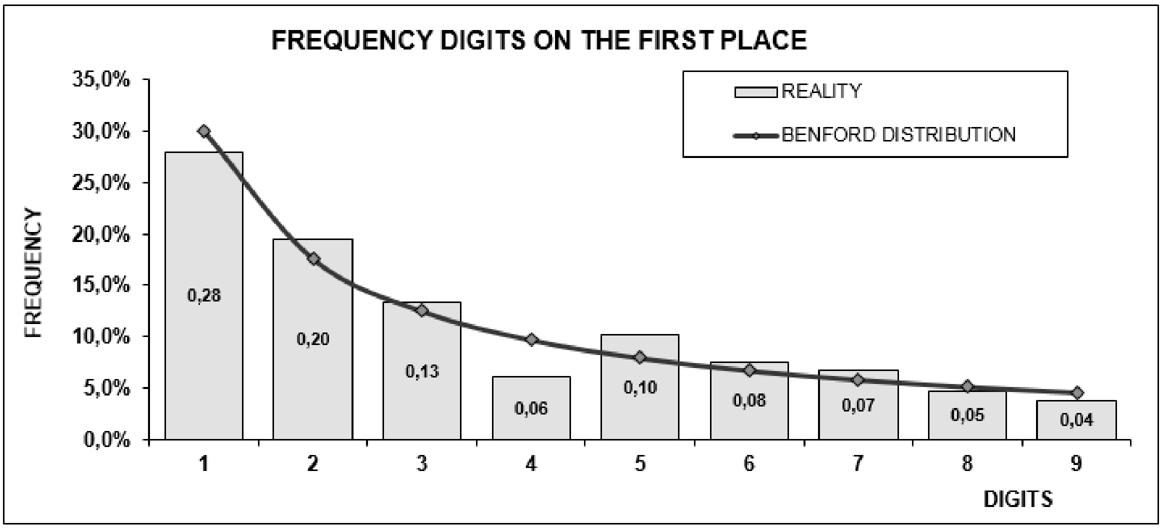

Fig. 1. Frequency digits on the first place.

Further statistical tests performed to verify the conformity of the Benford´s frequency distributions with frequency digits of empirical data used for testing the first digit χ^2 test. When testing the second digit and the first two digits is used Kolmogorov-Smirnov test, due to the existence of small probability and failure to comply with conditions for χ^2 test, and that the theoretical frequency must have a minimum value "5", in [5,8, 9].

All calculations and statistical tests created automatically in the model (see Fig. 2.). Just insert the data and test result appears with "OK" or "NOK".

The tests could also do with the frequencies of digits in third and fourth place, etc. Why is it usually does not do is that while the first digits are the differences of frequency (likely) very contrasting (see Tab. 1. or Fig. 2.) (frequency number 1 is 30.1%, the frequency of the number 9 is 4.6%) and the additional digits, these contrasts gradually reduced.

If there is a small difference of real numbers, and frequency numbers Benford´s breakdown (model), is a testing criterion of small value. Then the reality matches the model and the statistical test to apply. If the difference is greater than the critical value of the statistical test does not apply and there is no consensus.

The validity of the test is according to the parameter α (level of significance). The normal value is 0.05. The critical value is calculating according to statistical functions CHISQ.INV.

2 Result of the model

The model created in MS Excel, easily accessible for everyone. The data we need to verify is enough to insert into the green entry column. Automatic calculations of three statistical tests to verify compliance with the Benford´s distribution performed automatically.

In Fig. 2., you can see a part of the input of the values, and the parameter α is entered, eg 0.05. If the test is positive, "OK" appears in the green box, if the test is negative, "NOK" appears in the red box.

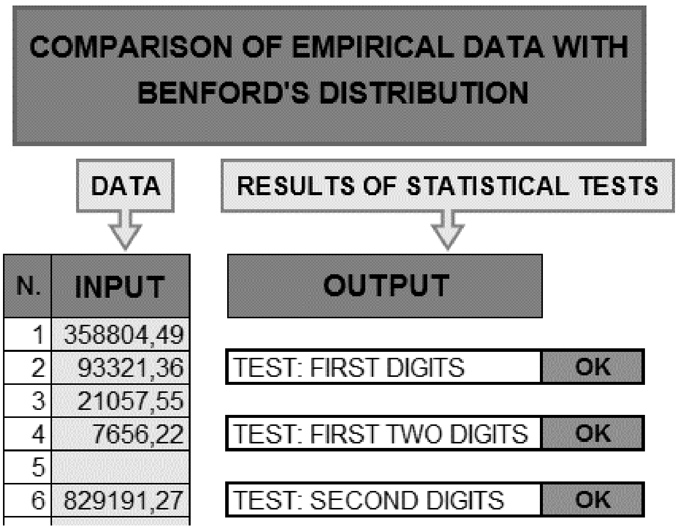

Fig. 2. Demonstration of work in the model.

The real data entered in to the model. Then there was a validation when α (level of significance) of 0.05. The result indicates that the frequency digits on the first two points correspond to the Benford’s laws. Therefore, there is no reason to believe that the data tampered with and falsified.

3 Conclusion

The accounts of any company gets special characteristics, in terms of frequency distributions of digits. Quantity of digits that occurs in the accounting respects the Benford's law. Data input assumption is realistic data. Therefore, you cannot take advantage of Benford's law for the so-called “random digits”. As mentioned above, Benford's law used to detect falsified data, in particular falsification of data in accounting. If the handling of a digits in accounting (small, medium or large business), then cease to satisfy Benford's law. In the case that occurs to manipulate the digits, it is with a certain probability followed the premise of the so-called statistical distribution. The statistical breakdown achieved by efforts to ensure the same likelihood of digit 1 or 2 or ... 9 at the beginning of the accounting data. This is a big mistake. With real data, the occurrence of the digit “1” at the beginning of the accounting data according to the Benford's law about 30%, rather than about 11% in the case of compliance with the so-called statistical distribution.

As has already been mentioned earlier in the text, the data file contained over 300 figures from the accounts of the real business. The results confirmed the validity of the Benford's distribution. It is to demonstrate the functioning this law. It also includes a confirmation of the fact that, in the accounting of this business not to manipulate the data.

The described model can very quickly and easily, whether the data are in some way suspicious. It used here to describe some properties of the Benford´s laws of the empirical data. This model is well suited especially for the verification of accounting data as first degree of control that identifies which of the accounts is in order and where values are suspicious and possibly rigged. This selection of the accounting unit is then to be exam such as dale using the audit of the accounts.

References

[1] BARROW, D. J. Benford’s very strange law. *Lecture on Gresham College* [online]. 2011.25. August 2011. [vid. 18. Mai 2017]. Available at: https://www.youtube.com/watch?v=4iz4EHriYz0&feature=related.

[2] BERGER, A., HILL, P. T. A Basic Theory of Benford’s Law. *Probability Surveys,* vol. 8, no. 1. 2011, ISSN 1549-5787. DOI: 10.1214/11-PS175 php?id=696&article.

[3] CARSLAW, C. A. P. Anomalies in Income Numbers: Evidence of Goal Oriented Behaviour. *Accounting Review*, vol. 63, no. 2, 1988: pp. 321–327, ISSN 0001-4826.

[4] DURTSCHI, C., HILLISON, W., PACINI, C. The effective use of Benford’s law to assist in detecting fraud in accounting data. *Journal of Forensic accounting*, vol. 5, no. 1, 2004: pp.17–34. ISSN 1524-5586.

[5] HILL, T. P. The First Digit Phenomenon. *American Scientist*, vol. 86, no. 1, 1998: pp. 358–363, ISSN 0003-0996.

[6] HILL, T. P. A Statistical Derivation of the SignificantDigit Law. *Statistical Science*, vol. 10, no. 4, 1995b: pp. 354–363, ISSN 0883-4237.

[7] HILL, T. P. Base-Invariance Implies Benford's Law. *Proceedings of the American Mathematical Society*, vol. 123, no. 3, 1995b: pp. 887–895, ISSN 0894-0347.
[8] CHO, W. K. T., GAINES, B. J. Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance. *The American Statistician*, vol. 61, no. 3, 2007: pp. 218-223, ISSN 0003-1305. DOI 10.1198/000313007X223496.
[9] NIGRINI, M. *Forensic Analytics: Methods and Techniques for Forensic Accounting Investigation.* NY: John Wiley&Sons Inc., 2011. ISBN 9781118087633. 352 s.
[10] NIGRINI, M. J. *Benford's law: applications for forensic accounting, auditing, and fraud detection*. NY: John Wiley&Sons Inc., 2012. ISBN 978-1-118-15285-0. 352 s.
[11] NIGRINI, M., MITTERMAIER, L. Audit Sampling Using Benford's Law: A Review of the Literature With Some New Perspectives. *Journal of Emerging Technologies in Accounting*. DOI 10.2308/jeta-51783.
[12] NYE, J., MOUL, C. The Political Economy of Numbers: On the Application of Benford's Law to International Macroeconomic Statistics. *Journal of Macroeconomics,* vol. 7, no. 1, 2007: pp. 1–14, ISSN 0164-0704.
[13] RAUCH, B., GÖTTSCHE, M., BRÄHLER, G. AND ENGEL, S. Fact and Fiction in EU-Governmental Economic Data. *German Economic Review*, vol. 12, no. 1, 2011: pp. 243–255, ISSN 1468-0475. DOI 10.1111/j.1468-0475.2011.00542.x.
[14] WATRIN, CH. Benford's Law: An Instrument for Selectin Tax Audit Targets?. *Review of Managerial Science*, vol. 2, no. 3, 2008: pp. 219–237, ISSN 1863-6683. DOI 0.1007/s11846-008-0019-9 1863-6683.

Current address

Košťálek, Josef, Ing.
Department of Management and Economics
Faculty of Mechanical Engineering
Czech Technical University in Prague
Karlovo náměstí 13, 121 35 Prague 2, Czech Republic
E-mail: josef.kostalek@fs.cvut.cz

Koťátková Stránská Pavla, Ing., PhD.
Department of Economic studies
Masaryk Institute of Advanced Studies, MIAS School of Business Institution
Czech Technical University in Prague
Kolejní 2637/2A, 160 00 Praha 6, Czech Republic
E-mail: pavla.kotatkova.stranska@cvut.cz