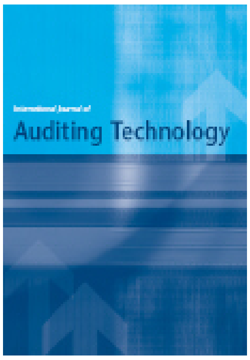


International Journal of Auditing Technology

ISSN online: 1757-8760 - ISSN print: 1757-8752
https://www.inderscience.com/ijaudit

Benford test based on logarithmic property

Javad Kazemitabar, Jalil Kazemitabar

DOI: 10.1504/IJAUDIT.2022.10047587

Article History:

Received:	18 December 2021
Accepted:	04 March 2022
Published online:	09 March 2023

Copyright © 2022 Inderscience Enterprises Ltd.

Benford test based on logarithmic property

Javad Kazemitabar*

Department of Electrical and Computer Engineering,
Babol Noshirvani University of Technology,
Shariati Street, Babol,
Mazandaran, Iran
Email: j.kazemitabar@nit.ac.ir
*Corresponding author

Jalil Kazemitabar

Statistics Department,
University of California Los Angeles,
Los Angeles, CA 90095, USA
Email: sjalilk@g.ucla.edu

Abstract: The Benford-like behaviour is commonly seen in practical data. That is, the digit frequency more or less has the distribution first introduced by Benford. Despite common belief, however, most datasets do not conform perfectly to Benford's law; they fail famous Benford tests in the literature, or as Nigrini puts it, these tests are too powerful for checking the conformity. We propose a new approach on measuring the deviation of datasets from Benford distribution to determine possible abnormality. We show that the conventional digit frequency tests do not fully absorb the 'significant digit' property. We discuss barriers on the way of auditors in using digit tests mainly when the number of samples is too small or too large. We then propose our method using the logarithmic basis of Benford's law which states the mantissa of the logarithm of all practical numbers should be uniformly distributed. We then test several goodness-of-fit techniques that compare the sample data's mantissa distribution with that of the uniform distribution between zero and one. Our experiment on sample datasets show that Kolmogrov-Smirnov test for uniformity works best for small, medium size and even large records.

Keywords: Benford's law; mantissa; logarithmic test; goodness of fit.

Reference to this paper should be made as follows: Kazemitabar, J. and Kazemitabar, J. (2022) 'Benford test based on logarithmic property', *Int. J. Auditing Technology*, Vol. 4, No. 4, pp.279–291.

Biographical notes: Javad Kazemitabar received his BSc from the Sharif University of Technology in 2003 and his MSc and PhD from the University of California Irvine in 2005 and 2008 respectively. His research interests are in fraud detection using statistical techniques. He is currently an Assistant Professor in the Babol Noshirvani University of Technology.

Copyright © 2022 Inderscience Enterprises Ltd.

Jalil Kazemitabar received his PhD in Statistics from the University of California Los Angeles, in 2020. He is currently working as a statistician in Apple Inc., USA.

1 Introduction

Benford's law was the result of a simple discovery; that certain pages of logarithmic tables have more tear and ware compared to others (Newcomb, 1881). There has been a lot of effort ever since in verifying and explaining the 'significant digit' phenomenon (Raimi, 1976; Pinkham, 1961; Washington, 1981; Sarkar, 1973). Frank Benford's explanation of this phenomenon was rather philosophical; that "the 1, 2, 3, … scale is not the natural scale […] Nature counts $e^0, e^x, e^{2x}, e^{3x}, ...$ and builds and functions accordingly" (Benford, 1938). While the statement was not exactly mathematical, it more or less served as a foundation for Theodore Hill's proof of the law (Hill, 1995). Hill showed through rigorous derivations that "if distributions are selected at random (in any 'unbiased' way), and random samples are then taken from each of these distributions, the significant digits of the combined sample will converge to the logarithmic (Benford) distribution". The logarithmic (Benford) distribution here refers to Newcomb's finding on practical numbers for which the mantissa of the logarithm is uniformly distributed. In other words, despite the common belief, the 'significant digit' phenomenon is more than mere digit frequency property that was pointed out by Benford. The digit frequency property mentions that in a set of practical numbers the percentage of the leftmost digit, roughly follows the last column in Figure 1. It further states that the distribution of the combination of the first two digits resembles the inner matrix of Figure 1. Benford provided a method of calculating these distributions for any set of combined first digits, e.g., first two digits, first three digits and etc. However, the digit frequency property which is the building block of most conventional Benford tests used in fraud detection industry is only a by-product of the logarithmic property. The logarithmic property that was found by Newcomb and elaborated by Hill states that if you take the logarithm of a set of practical numbers the fractional part of the log values will be uniformly distributed. A test of abnormality, i.e., Benford test, should naturally be defined based on the amount of deviation of this distribution from ideal (Kazemitabar and Kazemitabar, 2019). In other words, one should measure how much the distribution of the fractional part of the log values of a dataset differs from uniform. This difference would then determine how abnormal the dataset is. Conventional Benford tests or more formally digit tests, are only partially using this information as a measure of abnormality. Moreover, these tests are limited in scenarios where the sample size is either too large or too small. According to Nigrini, using digit frequency test on data samples sized less than 300 is not recommended (Nigrini, 2011). Nigrini further adds that for a dataset to conform to Benford's law properly (in the digit frequency sense), the size should be at least 1,000. The rationale is simple; the proportion of each digit has a precision that increases by size of the data. For example, in a list of size 1,000 we may have about 100 numbers (or 10%) that have 4 as their first digit. If we are going to estimate the probability of having 4 as the first digit with an acceptable percision, we roughly need that much number. On the other hand, when the list is larger than 5,000, the test suffers from excess power. In other words, most practical datasets of size larger than 5,000 will not pass Benford digit test. This is because common measures used such as Z-test and chi-square tighten their criterion as the size increases (Nigrini, 2011). As a matter of fact, auditing softwares such as IDEA (https://idea.caseware.com) only allow lists of size less than 2,500. The details of the excess power phenomenon can be found in Nigrini's work (Nigrini, 2011). Thus, we propose a new method that takes into account the full information provided by the logarithmic distribution property. We test our method on several practical datasets and show that our method no longer suffers from sample size problem.

Figure 1 Ideal digit frequency for the first two digits according to Benford's law

First significant digit \ Second significant digit	0	1	2	3	4	5	6	7	8	9	≈ Σ
1	4.14	3.78	3.48	3.22	3.00	2.80	2.63	2.48	2.35	2.23	30.10
2	2.12	2.02	1.93	1.85	1.77	1.70	1.64	1.58	1.52	1.47	17.61
3	1.42	1.38	1.34	1.30	1.26	1.22	1.19	1.16	1.13	1.10	12.49
4	1.07	1.05	1.02	1.00	.98	.95	.93	.91	.90	.88	9.69
5	.86	.84	.83	.81	.80	.78	.77	.76	.74	.73	7.92
6	.72	.71	.69	.68	.67	.66	.65	.64	.63	.62	6.69
7	.62	.61	.60	.59	.58	.58	.57	.56	.55	.55	5.80
8	.54	.53	.53	.52	.51	.51	.50	.50	.49	.49	5.12
9	.48	.47	.47	.46	.46	.45	.45	.45	.44	.44	4.58
≈ Σ	11.97	11.39	10.88	10.43	10.03	9.67	9.34	9.04	8.76	8.50	

Source: The International Statistical Institute (2013)

2 Remember: real data will never conform perfectly to Benford's law

This is what the R package, *Benford Analysis* (https://cran.r-project.org/web/packages/benford.analysis/index.html) prints out each time you test a dataset for conformity to Benford's law. Despite what most people might think, real data does not perfectly satisfy the digit test. In Figures 4–9, we ran several conventional Benford tests over real datasets. Even after taking into account the size recommendations[1] the result shows complete failure for one and two-digit tests. Some authors noted this phenomenon and have proposed their own customised tests (Nigrini, 2011; Amiram et al., 2015). Nigrini proposed mean absolute deviation (MAD) test that is not a function of the sample size. As he himself admits, however, one cannot mathematically provide conformity thresholds for this test. Therefore, Nigirini relied on his auditing experience and compiled a list of thresholds for different types of datasets. For example, for taxpayer balance if MAD test result is .0006, the conformity is acceptable, whereas for election data same threshold is within the rejected region. We believe that the method to be proposed in this paper that relies on the original 'significant digit' concept has the potential to come up with a criterion that can work for different classes of datasets.

Figure 2 Mean absolute deviation (MAD) test thresholds for different auditing fields

	A	B	C	D	E	F
1	Data#	Data Description	FTD	>= $10	Records	Notes
2	2	Streamflow Data, Mathematical Geology	0.0001	0	457,440	Near perfect natural data
3	20	Seismic signals, January 2010	0.0005	0	160,090,478	Assumed perfect
4	16	Taxpayer, Balance Due 1989	0.0006	0	93,582	Expected to conform
5	18	Ledger dump, Positive amounts	0.0008	0	151,202	Assumed good
6	3	Conglomerate, invoices in source currency	0.0009	1	706,106	No fraud or errors suspected
7	4	Conglomerate, invoices in US dollars	0.0009	1	704,390	No fraud or errors suspected
8	19	Ledger dump, Negative amounts	0.0009	0	136,987	Assumed good
9	23	Utility, KWH credits	0.0009	0	86,279	Surprisingly Benford
10	8	Streamflow Data, 1998-2002	0.0011	0	37,879	Expected perfection here
11	9	Invoice, Software company 1996	0.0012	1	36,515	This is a typical financial data set
12	13	Invoice, North Carolina	0.0012	1	247,811	No fraud or errors suspected
13	24	Census, 2000, county populations	0.0014	0	3,141	Assumed good
14	25	Invoices, Transport company	0.0014	0	198,955	Assumed good
15	5	Census, 1990, county populations	0.0015	0	3,141	Assumed good
16	12	Federal Govt., purchasing cards	0.0015	1	81,842	Should be a good fit
17	17	Conglomerate, travel reimbursements	0.0015	1	160,057	Sime issues expected
18	11	Utility, purchasing cards	0.0019	1	44,614	Should be a good fit
19	22	Utility, KWH billed	0.0019	0	10,669,357	Clear non-Benford pattern
20	10	Conglomerate, legal fees	0.0023	0	14,667	This is quite a weak fit
21	1	California Accounts Payable	0.0024	1	177,763	Data seems to have much by way of fraud or errors
22	14	Internet site, balances	0.0025	1	40,060	Clear issues in this data
23	15	Invoices, Texas	0.0026	1	1,887,958	Some issues
24	21	Tobacco, Sales numbers	0.0030	0	34,716	Highly questionable data
25	7	California, Special Election Governor	0.0033	1	3,112	Still a weak fit
26	6	California, Special Election Governor	0.0101	0	6,384	Weak fit to Benford
27						
28				1 = numbers less than 10 were deleted		

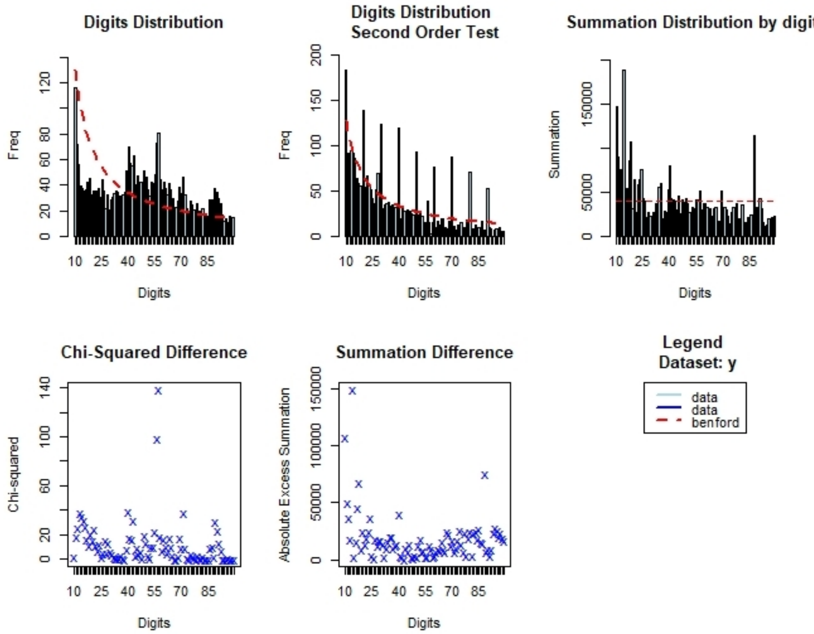
Source: Nigrini (2011)

Figure 3 MAD test thresholds obtained thru auditing experience for different digit tests

Digits	Range	Conclusion
First Digits	0.000 to 0.006	Close conformity
	0.006 to 0.012	Acceptable conformity
	0.012 to 0.015	Marginally acceptable conformity
	Above 0.015	Nonconformity
Second Digits	0.000 to 0.008	Close conformity
	0.008 to 0.010	Acceptable conformity
	0.010 to 0.012	Marginally acceptable conformity
	Above 0.012	Nonconformity
First-Two Digits	0.0000 to 0.0012	Close conformity
	0.0012 to 0.0018	Acceptable conformity
	0.0018 to 0.0022	Marginally acceptable conformity
	Above 0.0022	Nonconformity
First-Three Digits	0.00000 to 0.00036	Close conformity
	0.00036 to 0.00044	Acceptable conformity
	0.00044 to 0.00050	Marginally acceptable conformity
	Above 0.00050	Nonconformity

Source: Nigrini (2011)

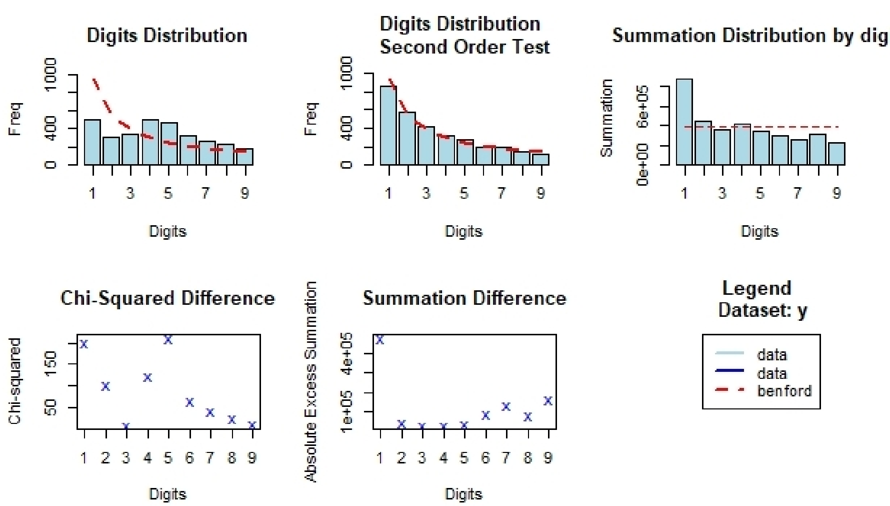
Figure 4 First-two digit test results using R's Benford analysis package for census data (see online version for colours)



Note: The data does not conform to Benford distribution based on Pearson's chi-squared test.

Source: Available at https://www.nigrini.com/ForensicAnalytics.htm

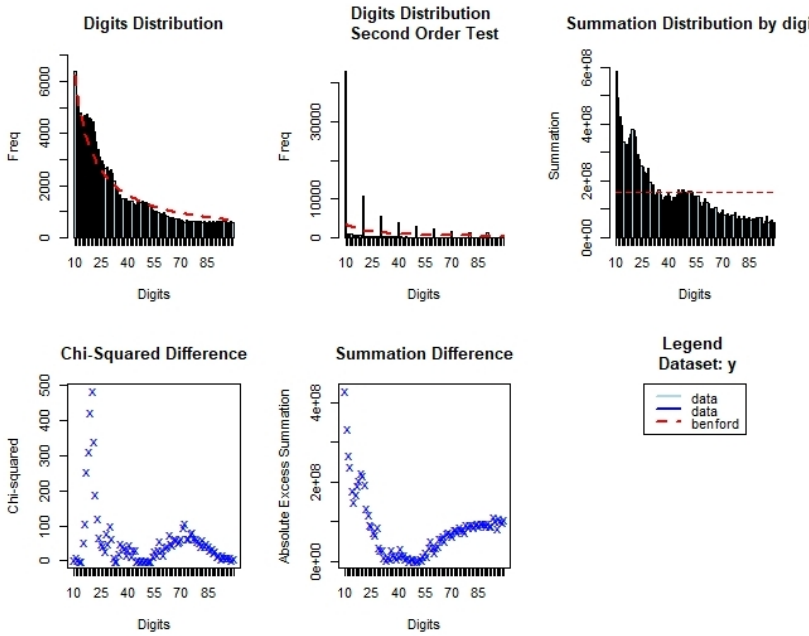
Figure 5 First digit test results using R's Benford analysis package for census data (see online version for colours)



Note: The data does not conform to Benford distribution based on Pearson's Chi-squared test.

Source: Available at https://www.nigrini.com/ForensicAnalytics.htm

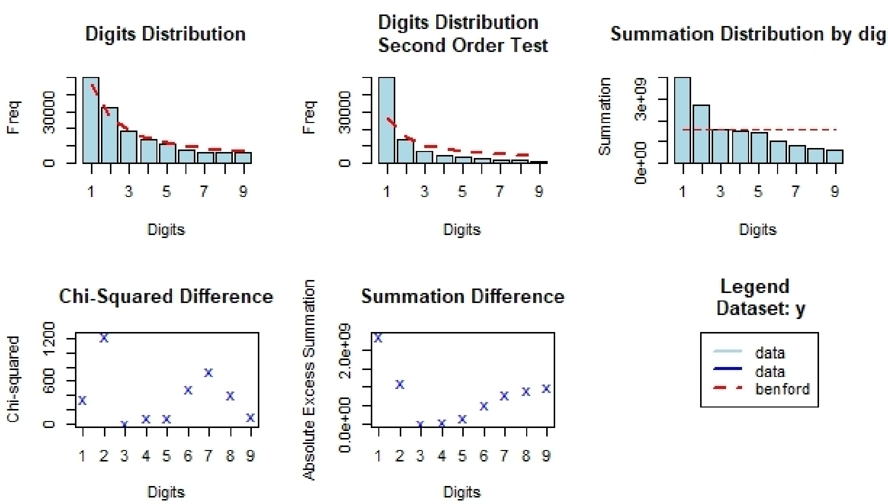
Figure 6 First-two digit test results using R's Benford analysis package for income data (see online version for colours)



Note: The data does not conform to Benford distribution based on Pearson's chi-squared test.

Source: Available at https://www.nigrini.com/ForensicAnalytics.htm

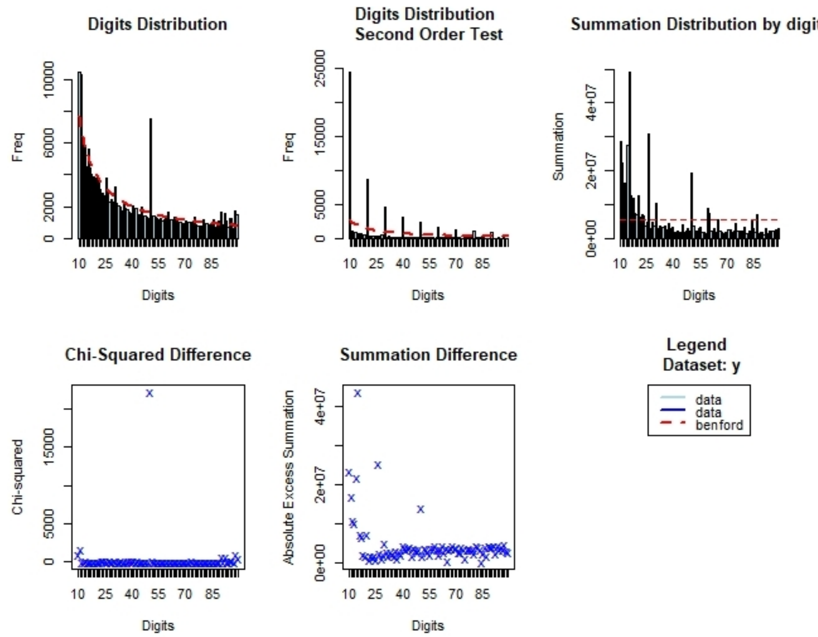
Figure 7 First digit test results using R's Benford analysis package for income data (see online version for colours)



Note: The data does not conform to Benford distribution based on Pearson's chi-squared test.

Source: Available at https://www.nigrini.com/ForensicAnalytics.htm

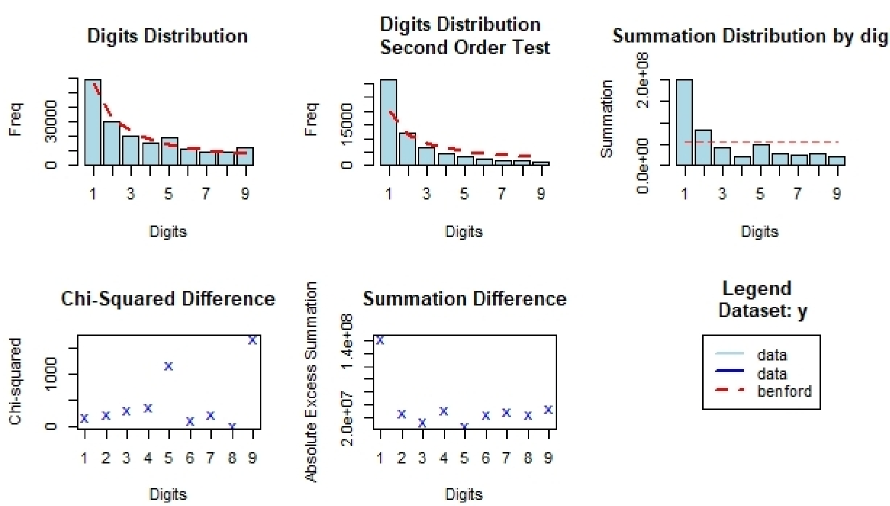
Figure 8 First-two digit test results using R's Benford analysis package for invoice data (see online version for colours)



Note: The data does not conform to Benford distribution based on Pearson's chi-squared test.

Source: Available at https://www.nigrini.com/ForensicAnalytics.htm

Figure 9 First digit test results using R's Benford analysis package for income data (see online version for colours)



Note: The data does not conform to Benford distribution based on Pearson's chi-squared test.

Source: Available at https://www.nigrini.com/ForensicAnalytics.htm

3 Benford test revisited

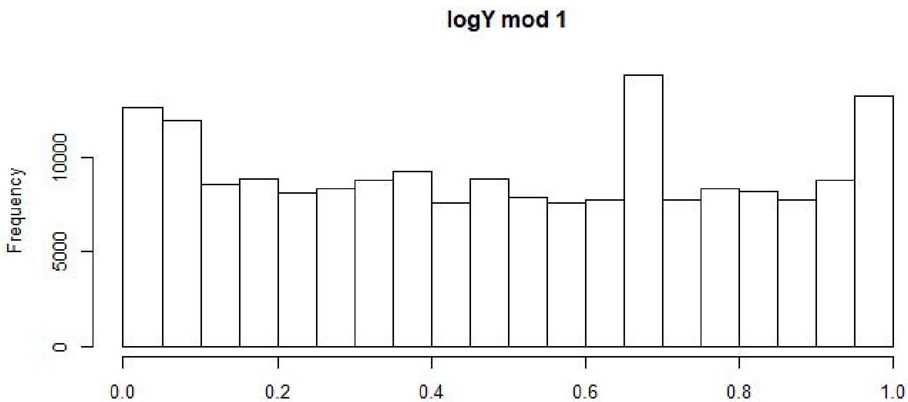
Take X as an arbitrary positively valued r.v. Let ϕ be a function that maps X to a new r.v. defined as follows

$$\phi(X) = \log_{10} X \mod 1 \tag{1}$$

where mod denotes fractional part or mantissa of a number. It is known that a ‘Benford like’ r.v. X generates a uniform $\phi(X)$ (Hill, 1995), or

$$\phi(X) \sim \text{Uniform}[0, 1) \tag{2}$$

Figure 10 Mantissa of the logarithm of Nigrini invoice dataset resembles uniform



Imagine now, we are given a dataset to investigate its abnormality. We should calculate $\phi(x)$ for all the records in that dataset and compare them with $\text{Uniform}[0, 1)$. If it fits well we declare conformity to Benford’s law, and vice versa. There are several goodness-of-fit techniques that compare a dataset with a given distribution (D’Agostino and Stephens, 1986). Specifically, there are goodness-of-fit techniques designed to measure a dataset’s fitness with uniform distribution. Fortunately, an R package titled ‘uniftest’ has been developed by Ruslan Pusev and Maxim Melnik of Saint Petersborg University (https://cran.r-project.org/web/packages/uniftest/index.html) that implements many of the famous goodness-of-fit tests for uniform distribution. For the sake of completeness, we also added simple regression test, a.k.a quantile-quantile, as directed by D’Agostino and Stephens (1986) to test uniformity. In what follows we briefly explain all these goodness-of-fit techniques.

3.1 Sherman test for uniformity

The statistic behind the Sherman test (Sherman, 1950) is as follows:

$$B_n = \frac{1}{2} \sum_{i=1}^{n+1} |X_{(i)} - X_{(i-1)} - \frac{1}{n+1}|, \tag{3}$$

where $X_{(0)} = 0,\ X_{(n+1)} = 1$.

3.2 Quesenberry-Miller test for uniformity

The statistic behind the Quesenberry-Miller test (Quesenberry and Miller, 1977) is as follows:

$$B_n = \sum_{i=1}^{n+1} (X_{(i)} - X_{(i-1)})^2 + \sum_{i=1}^{n} (X_{(i)} - X_{(i-1)})(X_{(i+1)} - X_{(i)}), \tag{4}$$

where $X_{(0)} = 0,\ X_{(n+1)} = 1$. The p-value is computed by Monte Carlo simulation.

3.3 Hegazy-Green test for uniformity

The statistic behind the Hegazy-Green test (Hegazy and Green, 1975) is as follows:

$$T_p = \frac{1}{n} \sum_{i=1}^{n} \left| X_{(i)} - \frac{i}{n+1} \right|^p. \tag{5}$$

3.4 Frosini test for uniformity

The statistic behind the Frosini test for uniformity is as follows:

$$B_n = \frac{1}{\sqrt{n}} \sum \left| X_{(i)} - \frac{i - 0.5}{n} \right|. \tag{6}$$

3.5 Neyman-Barton test for uniformity

The statistic behind the Neyman-Barton test (Neyman, 1937) is as follows:

$$N_k = \sum_{j=1}^{k} \left(\frac{1}{\sqrt{n}} \sum \pi_j(x_i) \right)^2, \tag{7}$$

where $\pi_j(x_i)$ are Legendre polynomials orthogonal on the interval [0, 1].

3.6 Kolmogorov-Smirnov test for uniformity

The statistic behind the Kolmogorov-Smirnov test (Kolmogorov, 1933) is as follows:

$$D^+ = \max_i \left(x_i - \frac{i}{n+1} \right), D^- = \max_i \left(\frac{i}{n+1} - x_i \right), D = \max(D^+, D^-). \tag{8}$$

3.7 Regression test for uniformity (Q-Q)

The regression test calculates $n(1 - R(x, s))$ where n is the sample size, s is the sample vector, x is the ideal distribution sample and $R()$ is the simple correlation operation. According to D'Agostino and Stephens (1986), for significance level of 0.05 the calculated output shall not exceed 1.774. The threshold holds for all datasets with greater than 80 records.

3.8 Anderson-Darling test

The Anderson-Darling test assesses whether a sample comes from a specified distribution. It makes use of the fact that, when given a hypothesised underlying distribution and assuming the data does arise from this distribution, the cumulative distribution function (CDF) follows a uniform distribution. The formula for the test statistic A to assess if data $\{x_1 < \cdots < x_n\}$ comes from a CDF F is

$$A^2 = -n - S\,, \tag{9}$$

where

$$S = \sum_{i=1}^{n} \frac{2i-1}{n} \left[\ln(F(x_i)) + \ln\left(1 - F(x_{n+1-i})\right)\right]. \tag{10}$$

The test statistic can then be compared against the critical values of the theoretical distribution.

4 Results

In order to run any of these tests on a dataset, one needs to first calculate the mantissa of logarithm of the records. We performed these tests on several datasets provided by Nigrini (https://www.nigrini.com/ForensicAnalytics.htm) the results of which can be found in Tables 1–3.

Table 1 p-value output of several uniformity goodness-of-fit tests applied to famous datasets

	Nigrini invoice	*Nigrini income*	*Census (area)*	*Census (population)*	*Benford set*
Anderson-Darling	0.0194	3.979835e-09	1.909004e-07	0.1998424	0.6485023
Sherman	2.2e-16[a]	2.2e-16	2.2e-16	0.501	0.88
Quesenberry-Miller	2.2e-16	2.2e-16	2.2e-16	0.0805	0.879
Hegazy-Green	2.2e-16	2.2e-16	2.2e-16	0.171	0.2105
Frosini	2.2e-16	2.2e-16	2.2e-16	0.1555	0.208
Neyman-Barton	2.2e-16	2.2e-16	2.2e-16	0.07	0.3625
Kolmogorov-Smirnov	0.1765	0.5715	0.5925	0.3345	0.789

Note: [a]software's smallest value. Actual number may be less.

Table 2 Quantile-quantile (regression) uniformity test applied to famous datasets based on $n(1 - R(x, s))$

Nigrini invoice	*Nigrini income*	*Census (area)*	*Census (population)*	*Benford set*
179.8132	1,025.082	225.8024	2.17	0.7796801

Note: Except for the synthetic Benford set, all other datasets fail to show conformity to Benford distribution in Q-Q sense.

Based on the results in Tables 1, 2 and 3, it is our suggestion to use Kolmogrov Smirnov goodness of fit to test the uniformity of the mantissa of the logarithm. We say this because all reasonable datasets pass this test. This is a sign that the test is not too powerful. Moreover, the test is not very loose. For example, small-sized samples of Nigrini invoice and Nigrini income datasets barely pass the 0.05 threshold. One should note, however, that the way we use Kolmogrov Smirnov goodness of fit is different from conventional Benford tests. In conventional Benford tests, the auditor compares goodness of fit of a discrete set of distributions with the Benford distribution, i.e., the expected frequency of 9 or 10 digits. However, in our method we will be comparing the set of data with a continuous distribution, i.e., the uniform distribution in the interval of [0, 1).

Table 3 (p-value, D-statistic) output of Kolmogrov-Smirnov uniformity test applied to different sample sizes of famous datasets

	Nigrini invoice	*Nigrini income*	*Census (area)*	*Census (population)*	*Benford set*
50	(0.096, 0.947)	(0.082, 0.953)	(0.807, 0.816)	(0.743, 0.831)	(0.3285, 0.901)
200	(0.129, 0.971)	(0.093, 0.976)	(0.38, 0.949)	(0.4575, 0.943)	(0.5595, 0.934)
2,000	(0.055, 0.994)	(0.511, 0.980)	(0.72, 0.974)	(0.608, 0.977)	(0.8425, 0.969)
20,000	(0.129, 0.997)	(0.3235, 0.995)	N/A[a]	N/A[a]	(0.8275, 0.990)
150,000	(0.1765, 0.998)	(0.5715, 0.997)	N/A[a]	N/A[a]	(0.752, 0.996)

Note: [a]dataset size limit reached.

5 Discussion

The proposed method has several advantages over conventional digit tests. We take the list as a whole rather than breaking it to different bins (each bin associated with a digit from 0 to 9). As a result, even in low sample size, we could obtain a good precision. Remember, one needs a good precision for each bin in the conventional method. However, because we are looking at the complete set, we do not face scarcity even when the list is of size 100. In other words, our method works for low sample size. Due to the same reason, we are saved from excess power phenomenon when the list is large. This is because, we are looking at a continuous distribution rather than a discrete one where every member of the list is a data point. When the list is large so are the data points to be compared. This is while in the conventional digit test, the number of data points to be compared, i.e., bins, are fixed (9 or 10 for single digits). List members are mapped to a bin and when the list size is large the population of each bin increases hence the excess power problem occurs.

6 Conclusions

Despite common belief, most datasets do not conform perfectly to Benford's law; they fail famous Benford tests in the literature, or as Nigrini puts it, these tests are too powerful for checking the conformity. In this paper we proposed a new approach on measuring the deviation of datasets from Benford distribution to determine possible abnormality. We showed that the conventional digit frequency tests do not fully absorb the 'significant digit' property. We discussed barriers on the way of auditors in using digit tests mainly when the number of samples is too small or too large. We then proposed our method using the logarithmic basis of Benford's law which states the mantissa of the logarithm of all practical numbers should be uniformly distributed. We then test several goodness-of-fit techniques that compare the sample data's mantissa distribution with that of the uniform distribution between zero and one. Our experiment on sample datasets showed that Kolmogrov-Smirnov test for uniformity works best for small, medium size and even large records.

References

Amiram, D., Bozanic, Z. and Rouen, E. (2015) 'Financial statement errors: evidence from the distributional properties of financial statement numbers', *Review of Accounting Studies*, Vol. 20, No. 4, pp.1540–1593.

Anderson, T.W. and Darling, D.A. (1954) 'A test of goodness of fit', *Journal of the American Statistical Association*, Vol. 49, No. 268, pp.765–769.

Benford, F. () 'The law of anomalous numbers', *Proc. Am. Philos. Soc.*, March, Vol. 78, No. 4, pp.551–572.

D'Agostino, R.B. and Stephens, M.A. (1986) *Goodness of Fit Techniques*, CRC Press Books, New York, USA.

Hegazy, Y.A.S. and Green, J.R. (1975) 'Some new goodness-of-fit tests using order statistics', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 24, No. 3, pp.299–308.

Hill, T.P. (1995) 'A statistical derivation of the significant-digit law', *Statistical Science*, Vol. 10, No. 4, pp.354–363.

Kazemitabar, J. and Kazemitabar, J. (2019) 'Measuring the conformity of distributions to Benford's law', *Communications in Statistics – Theory and Methods*, DOI: 10.1080/03610926.2019.1590599.

Kolmogorov, A. (1933) 'Sulla determinazione empirica di una legge di distribuzione', *G. Ist. Ital. Attuari*, Vol. 4, No. 1, pp.83–91.

Newcomb, S. (1881) 'Note on the frequency of use of the different digits in natural numbers', *American Journal of Mathematics*, Vol. 4, Nos. 1–4, pp.39–40.

Neyman, J. (1937) ''Smooth' test for goodness-of-fit', *Scand. Aktuarietidsrift*, Vol. 20, No. 3, pp.149–199.

Nigrini, M.J. (2011) *Forensic Analytics, Methods and Techniques for Forensic Accounting Investigations*, Wiley, New York, USA.

Pinkham, R.S. (1961) 'On the distribution of first significant digits', *Ann. Math. Statist.*, Vol. 32, No. 4, pp.1223–1230.

Quesenberry, C.P. and Miller, F.L. (1977) 'Power studies of some tests for uniformity', *J. Stat. Comput. Simul.*, Vol. 5, No. 3, pp.169–191.

Raimi, R.A. (1976) 'The first digit problem', *American Mathematical Monthly*, August–September, Vol. 83, No. 7, pp.521–538.

Sarkar, P.B. (1973) 'An observation on the significant digits of binomial coefficients and factorials', *Sankhya B*, Vol. 35, No. 3, pp.363–364.

Sherman, B. (1950) 'A random variable related to the spacing of sample values', *Ann. Math. Stat.*, Vol. 21, No. 3, pp.339–361.

The International Statistical Institute (2013) *Proceedings of the 59th World Statistics Congress of the International Statistical Institute*, August, pp.3881–3886.

Washington, L.C. (1981) 'Benford's law for Fibonacci and Lucas numbers', *The Fibonacci Quarterly*, Vol. 19, No. 2, pp.175–177.

Websites

https://cran.r-project.org/web/packages/benford.analysis/index.html (accessed 25 March 2022).

https://cran.r-project.org/web/packages/uniftest/index.html (accessed 25 March 2022).

https://idea.caseware.com (accessed 25 March 2022).

https://www.nigrini.com/ForensicAnalytics.htm (accessed 25 March 2022).

Notes

1 Nigrini recommends datasets of size 1,000–4,000 for good match.