Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 53 (6) (2024), 1724 – 1741
DOI : 10.15672/hujms.1316712

Research Article

Statistical analysis of fitting Pareto and Weibull distributions with Benford's Law: theoretical approach and empirical evidence

Jelena Stanojević *, Dragana Radojičić, Vesna Rajić, Tatjana Rakonjac-Antić

Faculty of Economics and Business, University of Belgrade, Belgrade, Serbia

Abstract

This paper studies the fundamental properties of Benford's Law which investigates the distribution of the first digits' appearance within datasets. The purpose and the usefulness of the research developed within the paper are to identify additional distributions, beyond those already investigated, that conform to the Benford distribution. As a main contribution, we state and prove with the new approach that the Pareto distribution and appropriate constant times Weibull density function, under some parameter constraint, obey Benford's Law. Further, with the statistical tests and simulation method, we quantify how the fit varies as the parameters of the Pareto distribution change. As Benford's Law is one of the main used approaches for detecting data manipulations and frauds in practice, we use that methodology to consider eventual manipulations in a set of data from the financial reports of three private hospitals operating in Serbia. Moreover, we present the conformity of the Weibull distribution to Benford's Law through the analysis of real-world data, where in the Weibull distribution demonstrates a good fit, even proof of that conformity is a known result in the literature. By demonstrating the adherence of Benford's characteristics to the Pareto and Weibull distributions, commonly employed for modeling in various fields, those findings can be utilized in many practical studies.

Mathematics Subject Classification (2020). 62P05, 62P30, 62P12, 60E10

Keywords. Data manipulations, Benford's Law, Pareto distribution, Weibull distribution, Healthcare simulation study, Statistical tests

*Corresponding Author.
Email addresses: jelena.stanojevic@ekof.bg.ac.rs (J. Stanojević), dragana.radojicic@ekof.bg.ac.rs (D. Radojicić), vesna.rajic@ekof.bg.ac.rs (V. Rajić), tatjana.rakonjac@ekof.bg.ac.rs (T. Rakonjac-Antić)
Received: 19.06.2023; Accepted: 25.09.2024

1. Introduction

Benford's Law is a complex mathematical method that predicts the distribution of digits for a wide variety of datasets including financial reports, stock prices and survey results. The Law is in general, one complex mathematical method and tool with the application for the detection of irregularities in a large dataset, i.e., it is a tool for determining whether investigated real-life statements contain errors or fraud. It actually describes the relative frequency distribution for leading digits in datasets. The Law claims that leading digits in data derived from measurements do not follow uniform distribution, which is in contradiction with our intuition. Actually, it claims that many numerical datasets follow the trend that leading digits 1-9 appear with decreasing logarithmic distribution, where digit 1 appears with the largest frequency, almost 30% and digit 9 with a frequency of 4.58%.

Many statistical techniques become increasingly important when mass data (Big Data) needs to be analyzed. For example, Bayerstadler et al. [3] introduces a predictive model for detecting fraud and abuse in health insurance which utilizes a manually reviewed claims dataset and employs a multinomial Bayesian latent variable approach and the estimation of model parameters is based on a Markov Chain Monte Carlo (MCMC) algorithm using Bayesian shrinkage techniques. Fraud detection in motor insurance was applied in [5], with a Bayesian model with latent variables.

On the other side, many authors have studied and applied Benford's Law for fraud and some error detection. Initially, the phenomenon of the relative frequency distribution for leading digits of numbers in datasets was investigated by Newcomb [27]. Almost 60 years later, Benford came to a similar conclusion and formally defined it in [4]. Application in the area of accounting was presented in [42]. Nigrini [28] wrote about the application of the Law in the detection of fraud, in the area of accounting, auditing and taxation. The Law has been subjected to an investigation by many researchers after Nigrini's initial paper in the area of forensic accounting, for example, see [10,22,29,31,33,38,40] and many others. Further, Hill [18] pointed out that Benford's Law is a phenomenon that is empirically provable and suggested a strict proof of the Law based on mathematical theory. He defined the probability frame of the Law and also proved that a combination of two distributions can give Benford's distribution, even if they do not follow Benford's distribution. The first time Benford's Law was used to detect errors in economic data by Varian [41], with application in economic forecasts. After that, the method was used by Michalski and Stoltz to detect errors in macroeconomic data [25]. Application of Benford's Law and possibilities for its use in international and governmental macroeconomic statistics can be found in [16,20,32]. A guide for detecting errors in transaction data was given in [30,31]. Investigation of Benford's Law in the field of online social networks was given in [11,14,15]. Also, the Law was used to check the reported number of COVID-19 cases. In [2,43], the Law was used to detect whether the countries manipulated COVID-19 data during the pandemic.

Due to the unexpected outcomes obtained from conventional statistical tests and procedures when evaluating small data samples, certain authors have suggested a novel test to examine whether the data conforms to Benford's Law distribution, for example, see [26]. Hill [19] asks the question: which common distribution functions obey Benford's Law. Leemis et al. [23] used numerical simulations and considered several distributions with different parameters. Approximation of the exponential distribution function with Benford's Law is shown by Engel and Leuenberger [12]. In [7,9], authors studied with the numerical simulations that the Weibull and the Inverse Gamma distributions are close to the Law. Particularly, in [7], authors successfully use the approach of complex Fourier transformation and Poisson summation technique to argue that the Weibull distribution is close to the Law. Note that our study reaches similar conclusions utilizing a new theoretical approach. Scott and Fasli [37] showed that the Log-normal distribution, as widely used in applications for modeling natural phenomena, also conforms to the Law and Rodriguez [36] showed that with the numerical simulation. Balado and Silvestre [1] examines general expressions for the joint distributions of the k most significant digits of continuous random variables. Additionally, it presents the general convergence Law for the distribution of the j-th significant digit, with a particular focus on Pareto random variables. However, our approach differs significantly from the aforementioned works, as we focus on studying the density functions and their expressions to theoretically demonstrate conformity with Benford's Law.

The aim of this paper is to investigate whether some other distributions than those examined so far, follow Benford's Law. We are especially interested in distributions that are used in the modeling of many practical problems. We consider real datasets from the financial report of three private hospitals operating in Serbia during the COVID-19 pandemic. The new theoretical results are presented within two theorems, employing a completely novel approach not previously used in the literature. The newly proposed method merits attention, as its primary contribution is the avoidance of complex techniques. As an auxiliary result, we represent that the transformed uniform variable follows Benford's distribution. As a first main result, we prove that the Pareto distribution with some constraint of the shape parameter a follows Benford's Law. As a second main result, we prove that the Weibull distribution with the same constraint is approximately Benford's distribution. Furthermore, we perform simulation studies to confirm the obtained theoretical results. Moreover, we intend to apply this methodology in the healthcare sector, for the case of three private hospitals operating in Serbia. In particular, individually for each hospital's dataset, we apply two statistical tests to assess the adherence of the data to Benford's Law, which aims to evaluate potential manipulation and fraud.

The paper is organized as follows: Section 2 gives some mathematical preliminaries, including some basic definitions and theorems and an analysis of Benford's Law. In Section 3, we present the main theoretical result that the Pareto and Weibull distributions, with some constraint of the parameter, follow Benford's Law. In Section 4, we conduct a simulation study with real data. Finally, in the last section, we provide some final remarks and directions for future research.

2. Preliminaries and analytic of Benford's Law

One useful consequence of Benford's Law is the fact that a path of appearance of some digits may indicate not only fraud, but it is also possible errors in data or bias in data presentation, [28]. Note that Benford's Law has the primary purpose of showing that data may not be accurate and if there are no coincidences with Benford's Law it is necessary to use deeper analysis to detect potential errors in data. Also, if there are coincidental digits with Benford's Law that does not immediately mean that there is no fraud. It is important to remark that positive and negative values have the same treatment in testing with Benford's Law, as also values with decimal numbers. Note that there are limitations to Benford's Law applicability, such as with small samples, coded data, perfectly uniform distributions, psychologically rounded numbers and mathematical sequences, such as the square roots and reciprocals of consecutive positive integers [34]. For further limitations of the Benford's Law applicability see [10, 31].

We first list standard definitions and the main theorem, which gives us conditions under which random variable follows Benford's Law, generally in the base B. We focus on the random variable with positive support since the Pareto distribution random variable takes values in $[b, +\infty)$, where $b > 0$ is the scale parameter of the distribution and we consider the Weibull distribution on positive support.

For every real number x there exists the integer part $\lfloor x \rfloor$ and the fractional part $\langle x \rangle$ of x, where x can be expressed uniquely as $x = \lfloor x \rfloor + \langle x \rangle$. It is satisfied that $\lfloor x \rfloor = \max\{k \in$

$Z : k \leq x\}$ and $\langle x \rangle = x - \lfloor x \rfloor$. For example, $\lfloor 3 \rfloor = 3$ and $\langle 3 \rangle = 0$, while $\lfloor 10e \rfloor = 27$ and $\langle 10e \rangle = 0.182...$

The main notation that concerns Benford's Law is the leading significant digits and more generally, the significand of a number. The next two definitions are given in the [13].

Definition 2.1. (Significand, [13]) For any positive number $x > 0$ and base B, x is represented as $x = S_B(x) \cdot B^{k(x)}$, where $S_B(x) \in [1, B)$ is called the significand of x and integer $k(x)$ (necessarily unique) represents the exponent. For negative number x, $S_B(x) = S_B(-x)$ and for convenience, $S_B(0) = 0$.

We focus on a general case and then consider the decimal number system, i.e., $B = 10$.

Definition 2.2. (Benford's Law, [13]) A (real-valued) random variable X follows Benford's Law in base B if and only if for all $t \in [1, B)$,

$$P\{S_B(X) \leq t\} = \log_B(t),$$

in particular,

$$P\{FSD = d\} = \log_B\left(\frac{d+1}{d}\right) = \log_B\left(1 + \frac{1}{d}\right), \quad d \in \{1, 2, ..., B-1\},$$

where FSD stands for first significant digit of X, that is the first (leftmost) digit of $S_B(X)$.

The next important result is given within the following theorem (Theorem 2.3 in [13]) and we use it to prove one of our main results in the paper, namely that the Pareto distribution under parameter constraint follows Benford's Law.

Theorem 2.3. *([13]) A random variable $X > 0$, follows Benford's Law in base B if and only if the random variable $Y = \langle \log_B(X) \rangle$, the fraction part of $\log_B(X)$ is uniformly distributed in $[0, 1]$.*

Proof. Let $X > 0$ be a random variable. For all $y \in [0, 1)$, the next equations are satisfied,

$$\begin{aligned} P\{Y \leq y\} &= P\{\langle \log_B(X) \rangle \leq y\} \\ &= P\{\log_B(X) - \lfloor \log_B(X) \rfloor \leq y\} \\ &= P\{\log_B(X) \in \bigcup_{k \in Z} [k, k+y]\} \\ &= P\{X \in \bigcup_{k \in Z} [B^k, B^{k+y}]\} \\ &= P\{S_B(X) \leq B^y\}. \end{aligned} \tag{2.1}$$

By Definition 2.2 a random variable X is Benford if and only if

$$P\{S_B(X) \leq B^y\} = \log_B(B^y) = y$$

for all $y \in [0, 1)$ and by above calculations that is equivalent with the condition that $Y = \langle \log_B(X) \rangle$ is uniformly distributed in $[0, 1]$ and the proof is completed. □

Further, we focus on the decimal number system, i.e., $B = 10$. For the random variable X which follows Benford's Law, the probability of an occurrence of the first digit is obtained by the next formula [31],

$$P\{D_1(X) = d_1\} = \log\left(1 + \frac{1}{d_1}\right), \quad d_1 \in \{1, 2, ..., 9\}, \tag{2.2}$$

where $D_1(X)$ is a random variable representing the first digit of $S(X)$ (where $S(\cdot) = S_{10}(\cdot)$) and d_1 is a digit on the first position in the number that takes value in the set of all possible outcomes.

In Equation (2.3) the probabilities of occurrence for the higher-order digits up to the last digit are derived, whereas higher-order digits appear with an equal probability of 0.1 which is identical to a uniform distribution,

$$P\{D_k(X) = d_k\} = \sum_{d_1=1}^{9} \sum_{d_2=0}^{9} \dots \sum_{d_{k-1}=0}^{9} \log\Big(1 + \frac{1}{\sum_{i=1}^{k} 10^{k-i} d_i}\Big), \tag{2.3}$$

where $D_k(X)$ is a random variable representing the digit on the k-th position of the random variable and d_k is a digit on that position, $d_k \in \{0, 1, ..., 9\}$, see [21].

It is possible to extend Benford's Law on the first k digits in the number. The appropriate formula is summarized within the next corollary that can be found in [33].

Corollary 2.4. *A random variable X follows Benford's Law if and only if*

$$P\{D_1(X) = d_1, D_2(X) = d_2, ..., D_k(X) = d_k\} = \log\Big(1 + \frac{1}{\sum_{i=1}^{k} 10^{k-i} d_i}\Big), \tag{2.4}$$

for all $k \in N$, all $d_1 \in \{1, 2, ..., 9\}$ and all $d_i \in \{0, 1, ..., 9\}$, $i \geq 2$.

3. Main results

As the auxiliary result in this paper, we give the next theorem, which idea we used to prove that appropriate constant times the Weibull density function, under one parameter constraint, can be approximated with the Benford distribution. The first part of that result was given in [17] and the second part can be found as an example in the literature, but we give the proof here.

Theorem 3.1. *If $Y = 10^X$, where X has uniform distribution in $[0, 1]$, then Y is Benford random variable and it is satisfied $f_Y(y) = \frac{1}{y \ln 10}$, where $f_Y(y)$ is appropriate density function for Y.*

Proof. Let $Y = 10^X$, where X has the uniform distribution in $[0, 1]$. From Definition 2.1, for $y \in [1, 10)$ it is satisfied $y = S(y) \cdot 10^{k(y)} = S(y) \cdot 10^0 = S(y)$, then $S(Y) = Y$. We may conclude:

$$P\{S(Y) \leq t\} = P\{Y \leq t\} = P\{10^X \leq t\} = P\{X \leq \log t\} = \log t, \ t \in [1, 10),$$

and $Y = 10^X$ has a Benford distribution. Further, we have to prove the second part of the theorem. For transformation $y = 10^x$, we have $x = h(y) = \log y$, for $y \in [1, 10)$ and $f_Y(y)dy = f_X(h(y))|h'(y)|dy = f_X(x)dx$. It is satisfied:

$$f_Y(y) = f_X(x)\frac{dx}{dy},$$

and with $x = \log y = \frac{\ln y}{\ln 10}$, it can be calculated that

$$\frac{dx}{dy} = \frac{1}{y \ln 10},$$

and $f_Y(y) = \frac{1}{y \ln 10}$, what is the claim of the theorem. □

The first main result in this paper claims that the Pareto distribution with two parameters a and b, follows Benford's Law when the shape parameter a is sufficiently small. We derive an expression of the distribution of the fraction of log transformation of a random variable with the Pareto distribution in any number system B and show that it is uniformly distributed in $[0, 1]$. It means that the distribution follows Benford's Law, what is the claim of Theorem 2.3.

Theorem 3.2. *Let X be a random variable that has the Pareto distribution with the shape parameter a and the location parameter b. Let $F_B(y)$ be the cdf of $Y = \langle \log_B(X) \rangle$ for $y \in [0,1]$. Then $F_B^{'}(y)$ can be expressed as*

$$F_B^{'}(y) = \frac{b^a \cdot a \ln B \cdot B^{-a(k_1-1+y)}}{B^a - 1}, \tag{3.1}$$

which approaches 1 for sufficiently small parameter a, where $k_1 = \lfloor \frac{\ln b}{\ln B} \rfloor + 1$.

Proof. Let X be a random variable that has the Pareto distribution, where $F_X(x) = 1 - b^a x^{-a}$, $x \geq b$, $a, b > 0$ and let a be a sufficiently small parameter. Further, we will find the cdf of $\langle \log_B(X) \rangle$ and show that its first derivative is almost 1. By further calculations we obtain,

$$\begin{aligned} F_B(y) &= P\{\langle \log_B(X) \rangle \in [0, y]\} \\ &= \sum_{k=-\infty}^{+\infty} P\{\log_B(X) \in [k, k+y]\} \\ &= \sum_{k=-\infty}^{+\infty} P\{X \in [B^k, B^{k+y}]\} \\ &= \sum_{k=k_1}^{+\infty} (F_X(B^{k+y}) - F_X(B^k)), \end{aligned} \tag{3.2}$$

where $B^k \geq b$, $B > 1$ and then $k \geq \dfrac{\ln b}{\ln B}$ and $k_1 = \lfloor \frac{\ln b}{\ln B} \rfloor + 1$. After some calculations, we may obtain,

$$\begin{aligned} F_B(y) &= b^a(1 - B^{-ay}) \sum_{k=k_1}^{+\infty} (B^{-a})^k \\ &= b^a(1 - B^{-ay}) \frac{B^{-ak_1}}{1 - B^{-a}} \\ &= b^a B^{-ak_1}(1 - B^{-ay}) \frac{B^a}{B^a - 1} \\ &= \frac{b^a B^{-a(k_1-1)}(1 - B^{-ay})}{B^a - 1} \end{aligned} \tag{3.3}$$

and

$$\begin{aligned} F_B^{'}(y) &= \frac{b^a B^{-a(k_1-1)} \cdot a \ln B \cdot B^{-ay}}{B^a - 1} \\ &= \frac{b^a \cdot a \ln B \cdot B^{-a(k_1-1+y)}}{B^a - 1}. \end{aligned} \tag{3.4}$$

When parameter a is sufficiently small, then term b^a approaches to 1, $B^{-a(k_1-1+y)}$ approaches to 1, $a \ln B = \ln(1 + (B^a - 1))$ approaches to $B^a - 1$ and $F_B^{'}(y)$ approaches to 1 and we have proved the theorem. $\square$

With the previous theorem, we proved that the Pareto distribution follows Benford's Law. Furthermore, motivated with Theorem 3.1 the second main result in this paper is given in the next theorem and it claims that appropriate constant times the Weibull density function, under one parameter constraint, can be approximate with the Benford distribution.

Theorem 3.3. *Let X be a random variable which has two-parameter Weibull distribution with parameters a and b ($X \sim W_2(a, b)$), where a is a shape parameter, b is a scale parameter and $a, b > 0$. For sufficiently small parameter a, $c \cdot f_X(x)$ can be approximate with the Benford distribution, where $c = \frac{e}{a \ln 10}$ and $f_X(x)$ is a density function of X.*

Proof. Let X be a random variable that has two-parameter Weibull distribution, $X \sim W_2(a, b)$. Following the Weibull distribution property (page 74 in [8]) we have

$$X = b \cdot (-\ln Y)^{1/a},$$

where Y is uniformly distributed in $(0, 1)$. Under transformation of the random variables, we obtain

$$f_Y(y) = f_X(h(y))|h^{'}(y)|,$$

where $x = h(y) = b \cdot (-\ln y)^{1/a}$. Further,

$$\begin{aligned} 1 &= f_X(x) \cdot |b \cdot \frac{1}{a}(-\ln y)^{\frac{1}{a}-1} \cdot \Big(-\frac{1}{y}\Big)| \\ &= f_X(x) \cdot |b \cdot \frac{1}{a} \cdot \frac{\frac{x}{b}}{\left(\frac{x}{b}\right)^a} \cdot \Big(-e^{\left(\frac{x}{b}\right)^a}\Big)|, \end{aligned}$$

and from the previous equation, we have

$$f_X(x) = \frac{a}{x} \cdot \left(\frac{x}{b}\right)^a \cdot e^{-\left(\frac{x}{b}\right)^a}.$$

For sufficiently small parameter a, term $\left(\frac{x}{b}\right)^a$ approaches to 1 and $e^{-\left(\frac{x}{b}\right)^a}$ approaches to $\frac{1}{e}$, then we obtain that $f_X(x)$ approaches to $\frac{a}{x} \cdot 1 \cdot \frac{1}{e}$ and also $\frac{e}{a \ln 10} \cdot f_X(x)$ approaches to $\frac{1}{x \ln 10}$, and with this the statement of the theorem is proved.

□

4. A simulation study and real data application

In order to assess the fitting of the observed datasets to Benford's Law we perform two following tests: Z-test and Chi-square test.

For the Z-test, to obtain the statistical significance of the deviations in the expected and the observed proportion, we calculate the Z-statistic for the appropriate digit. The formula is given with the next equation, [31]:

$$Z_i = \frac{|p_{oi} - p_i| - \frac{1}{2n}}{\sqrt{\frac{p_i(1-p_i)}{n}}}, \tag{4.1}$$

where Z_i is Z-statistic for the digit i ($i = 1, ..., 9$), p_{oi} is the observed frequency proportion of the digit i, p_i is the expected frequency proportion of the digit i according to Benford's Law, n is the number of observations of the examined variable, the term $\frac{1}{2n}$ is Yates' correction factor and it is used when it is smaller than the absolute difference $|p_{oi} - p_i|$ in the numerator.

For this test, the null hypothesis claims that the observed proportion does not statistically differ from the expected proportion based on Benford's Law. We will give all calculations with a 5% level of significance.

The second test which we applied is the Chi-square test. While the Z-test tests each digit separately (what is its lack), this test is conducted over all digits at the same time (simultaneously) (for example, see [6, 10, 39]). This test explores whether the observed frequency statistically differs from the expected frequency based on Benford's Law and the null hypothesis is that the distribution of all digits confirms the expected distribution under Benford's Law. If the test rejects the null hypothesis then this is a signal for data manipulation and it is recommended to look deeper into the data.

When we consider the first position, the Chi-square statistic is determined with the next formula:

$$\chi^2 = \sum_{i=1}^{9} \frac{\left(O_i - E_i\right)^2}{E_i}, \tag{4.2}$$

where O_i is the observed frequency of the digit i, E_i is the expected frequency of the digit i implied by the Benford's Law ($E_i = np_i$), n is the number of observations for the examined variable. Statistic χ^2 has eight degrees of freedom, i.e., $\chi^2 = \chi^2_{(8)}$. When we consider this test we will give all calculations with a 5% level of significance.

4.1. A simulation study with the Pareto distribution

To empirically validate the theoretically derived results, we generate random samples from the Pareto distribution, keeping the scale parameter b fixed and varying the shape parameter a (ensuring a remains sufficiently small, i.e., parameter a close to zero, but not equal to zero).

Table 1 presents the results of the Z-test conducted on the data that follows the Pareto distribution with parameters $a = 0.03$ and b = 1.5. The differences (between observed and Benford's Law frequencies) which are presented in the table are not statistically significant for each digit (Z-statistic is less than 1.96).

Table 1. Result of Z-test for the first digit in the simulation of the Pareto distribution when $a = 0.03$ and $b = 1.5$.

Digit	Count	Frequency	Benford's Law	Absolute difference	Z-Statistic
1	28	0.35	0.301	0.049	0.9554
2	13	0.1625	0.1761	0.0136	0.3193
3	9	0.1125	0.1249	0.0124	0.3355
4	12	0.15	0.0969	0.0531	1.60549
5	4	0.05	0.0792	0.0292	0.9671
6	3	0.0375	0.067	0.0295	1.0553
7	4	0.05	0.058	0.008	0.3061
8	3	0.0375	0.0512	0.0137	0.5559
9	4	0.05	0.0458	0.0042	0.1797
$\sum$	80	1	1		

Source: Authors' calculations.

Testing with the Chi-square test, results for the Pareto distribution when $a = 0.03$ and $b = 1.5$ are presented in Table 2. According to this test, the observed test statistic is equal to 5.4612, while the critical value is equal to 15.51 (with $\alpha = 0.05$ and 8 degrees of freedom), the null hypothesis can not be rejected, etc. it can be concluded that the first digit of the Pareto distribution follows Benford's Law.

Table 2. Result of Chi-square test for the first digit in the simulation of the Pareto distribution when $a = 0.03$ and $b = 1.5$.

Digit	Obs. frq. O_i	Benford's Law	Exp. freq. E_i	$(O_i - E_i)^2/E_i$
1	28	0.301	24.08	0.6381
2	13	0.1761	14.088	0.084
3	9	0.1249	9.992	0.0985
4	12	0.0969	7.752	2.3278
5	4	0.0792	6.336	0.8612
6	3	0.067	5.36	1.0391
7	4	0.058	4.64	0.0883
8	3	0.0512	4.096	0.2932
9	4	0.0458	3.664	0.0308
$\sum$	80	1	80	5.4612

Source: Authors' calculations.

Testing with Z-test, results for the Pareto distribution when $a = 0.1$ and $b = 1.5$ are presented in Table 3. The differences (between observed and Benford's Law frequencies) which are presented in the table are not statistically significant at all digits (Z-statistic is less than 1.96). That result suggests that attention may be given to other tests.

Testing with the Chi-square test, results for the Pareto distribution when $a = 0.1$ and $b = 1.5$ are presented in Table 4. According to this test, the observed test statistic is equal to 6.95, while the critical value is equal to 15.51 (with $\alpha = 0.05$ and 8 degrees of freedom), the null hypothesis can not be rejected, etc. it can be concluded that the first digit of the Pareto distribution follows Benford's Law.

Table 3. Result of Z-test for the first digit in the simulation of the Pareto distribution when $a = 0.1$ and $b = 1.5$.

Digit	Count	Frequency	Benford's Law	Absolute difference	Z-Statistic
1	19	0.2375	0.301	0.0635	1.2382
2	11	0.1375	0.1761	0.0386	0.9064
3	14	0.175	0.1249	0.0501	1.3554
4	6	0.075	0.0969	0.0219	0.662
5	9	0.1125	0.0792	0.0333	1.1029
6	7	0.0875	0.067	0.0205	0.7333
7	4	0.05	0.058	0.008	0.306
8	4	0.05	0.0512	0.0012	0.0487
9	6	0.075	0.0458	0.0292	1.2493
$\sum$	80	1	1		

Source: Authors' calculations.

Table 4. Result of Chi-square test for the first digit in the simulation of the Pareto distribution when $a = 0.1$ and $b = 1.5$.

Digit	Obs. freq. O_i	Benford's Law	Exp. freq. E_i	$(O_i - E_i)^2/E_i$
1	19	0.301	24.08	1.07169
2	11	0.1761	14.088	0.6768
3	14	0.1249	9.992	1.60769
4	6	0.075	7.752	0.3959
5	9	0.0792	6.336	1.12
6	7	0.067	5.36	0.5
7	4	0.058	4.64	0.088
8	4	0.0512	4.096	0.00225
9	6	0.0458	3.664	1.489
$\sum$	80	1	80	6.9539

Source: Authors' calculations.

Testing with Z-test, results for the Pareto distribution when $a = 0.01$ and $b = 1.5$ are presented in Table 5. The differences (between observed and Benford's Law frequencies) which are presented in the table are not statistically significant at all digits (Z-statistic is less than 1.96). That result suggests that attention may be given to other tests.

Testing with the Chi-square test, results for the Pareto distribution when $a = 0.01$ and $b = 1.5$ are presented in Table 6. According to this test, the observed test statistic is equal to 11.32, while the critical value is equal to 15.51 (with $\alpha = 0.05$ and 8 degrees of freedom), the null hypothesis can not be rejected, etc. it can be concluded that the first digit of the Pareto distribution follows Benford's Law.

Table 5. Result of Z-test for the first digit in the simulation of the Pareto distribution when $a = 0.01$ and $b = 1.5$.

Digit	Count	Frequency	Benford's Law	Absolute difference	Z-Statistic
1	21	0.2625	0.301	0.0385	0.7507
2	14	0.175	0.1761	0.0011	0.0258
3	7	0.0875	0.1249	0.0374	1.0118
4	11	0.1375	0.0969	0.0406	1.2275
5	11	0.1375	0.0792	0.0583	1.9309
6	3	0.0375	0.067	0.0295	1.0553
7	4	0.05	0.058	0.008	0.3061
8	2	0.025	0.0512	0.0262	1.0632
9	7	0.0875	0.0458	0.0417	1.7841
$\sum$	80	1	1		

Source: Authors' calculations.

Table 6. Result of Chi-square test for the first digit in the simulation of the Pareto distribution when $a = 0.01$ and $b = 1.5$.

Digit	Obs. freq. O_i	Benford's Law	Exp. freq. E_i	$(O_i - E_i)^2/E_i$
1	21	0.301	24.08	0.3939
2	14	0.1761	14.088	0.0005
3	7	0.1249	9.992	0.8959
4	11	0.0969	7.752	1.3608
5	11	0.0792	6.336	3.4332
6	3	0.067	5.36	1.039
7	4	0.058	4.64	0.08827
8	2	0.0512	4.096	1.0725
9	7	0.0458	3.3664	3.037
$\sum$	80	1	80	11.3218

Source: Authors' calculations.

From the previous analysis, with simulations of Pareto distribution for different values of shape parameter a (sufficiently small) and fixed scale parameter b, we confirmed our theoretical result from the previous section, that Pareto distribution follows Benford's Law, using a Chi-square test. The whole approach can be applied to any other parameter settings, with sufficiently small parameter a.

4.2. Real data application

Motivated by the current problems in the health system under the influence of the COVID-19 pandemic, we decided to focus on the field of health insurance. As a commonly employed method for identifying potential anomalies in datasets, Benford's Law is utilized to investigate irregularities within the financial statements of three private hospitals in the

Republic of Serbia. The basic right that can be exercised by insured persons based on health insurance is the right to comprehensive health care, [35]. As a consequence of the COVID-19 pandemic, the costs of health care have increased significantly, which requires more intensive supervision and control, first of all, of financial reports. The balance sheet is a financial report that contains a comparative view of assets, i.e., assets at the disposal of the hospital and liabilities, i.e., origin of funds. The income statement is a financial report that shows the hospital's income and expenses in order to determine the business results (profit or loss) in a specific period of time [24]. In particular, we decided to include in the analysis the balance sheets and profit and loss statements from 2018 - 2020 of three private hospitals operating in Serbia: Hospital 1, Hospital 2 and Hospital 3 (including the periods before the declaration of the COVID-19 pandemic and after the declaration of the pandemic). Data are available on the website of the Serbian Business Registers Agency (SBRA)[†]. The characteristic of the observed data is reflected in the fact that financial reports are subject to data manipulations and frauds by the reporters in order to achieve additional subsidies or financial incentives, especially in a period of crisis such as Covid time, which is also considered in the paper. Therefore, motivated by the desire to determine possible irregularities in the financial reports of the observed hospitals, we proposed the application of Benford's Law to the given data, as a model and method known in the literature.

If the data follows a two-parameter Weibull distribution with sufficiently small shape parameter, then we can apply Theorem 3.3.

We confirmed that all datasets fit the Weibull distribution well, with the corresponding parameter estimates: 0.44956 for shape parameter and 36419.0 for scale parameter for Hospital 1 and appropriate Kolmogorov Smirnov (K-S) statistic is 0.08346, while critical value is 0.17; 0.44511 for shape parameter and 23688.0 for scale parameter for Hospital 2 and appropriate K-S statistic is 0.08883, while critical value is 0.19429; 0.51084 for shape parameter and 33887.0 for scale parameter for Hospital 3 and appropriate K-S statistic is 0.12943, while critical value is 0.14839.

Within the context of Theorem 3.3, we theoretically demonstrated that the Weibull distribution, when subjected to a certain parameter constraint (for a sufficiently small shape parameter a), closely approximates the Benford distribution. We investigate whether the financial statement datasets, which follows a Weibull distribution with the estimated aforementioned parameter a values, aligns with Benford's Law.

From the financial report of Hospital 1, we obtain the following results. Testing with Z-test, results for Hospital 1 report are presented in Table 7. The differences (between observed and Benford's Law frequencies) which are presented in the table are not statistically significant at most digits (Z-statistic is less than 1.96), except for digits three and seven. Within the 64 observations in the considered sample, 2 observations start with the digit three, which is 3.12% of all positions and 9 observations start with the digit seven, which is 14.06% of all positions. That result suggests that more attention must be given to these two digits.

Testing with Chi-square test, results for Hospital 1 report are presented in Table 8. According to this test, the observed test statistic is equal to 15.66, while the critical value is equal to 15.51, with $\alpha = 0.05$ and 8 degrees of freedom (p-value is 0.0474), the null hypothesis can be rejected, etc. it can be concluded that first digit distribution of this report does not follow Benford's Law.

[†]www.apr.gov.rs

Table 7. Result of Z-test for the first digit from the financial report of Hospital 1.

Digit	Count	Freq.	Benford's Law	Absolute difference	Z-Statistic
1	23	0.3594	0.301	0.0584	0.8818
2	11	0.1719	0.1761	0.0042	0.0887
3	2	0.0312	0.1249	0.0937	2.0771*
4	3	0.0469	0.0969	0.0500	1.1416
5	7	0.1094	0.0792	0.0302	0.6625
6	4	0.0625	0.067	0.0045	0.1440
7	9	0.1406	0.058	0.0826	2.5605*
8	2	0.0312	0.0512	0.0200	0.4405
9	3	0.0469	0.0458	0.0011	0.0411
$\sum$	64	1	1		

Source: Authors' calculations.

Table 8. Result of Chi-square test for the first digit from the financial report of Hospital 1.

Digit	Obs. freq. O_i	Benford's Law	Exp. freq. E_i	$(O_i - E_i)^2/E_i$
1	23	0.301	19.264	0.7245
2	11	0.1761	11.2704	0.0065
3	2	0.1249	7.9936	4.4940
4	3	0.0969	6.2016	1.6528
5	7	0.0792	5.0688	0.7358
6	4	0.067	4.288	0.0193
7	9	0.058	3.712	7.5331
8	2	0.0512	3.2768	0.4975
9	3	0.0458	2.9312	0.0016
$\sum$	64	1		15.6651

Source: Authors' calculations.

From the financial report of Hospital 2, we obtain the following results.

Testing with Z-test, results for Hospital 2 report are presented in Table 9. The differences (between observed and Benford's Law frequencies) which are presented in the table are not statistically significant at most digits (Z-statistic is less than 1.96), except for the digit seven. Within the 49 observations in the considered sample, 7 observations start with the digit seven, which is 14.29% of all positions. That result suggests that more attention must be given to this digit.

Testing with the Chi-square test, results for the Hospital 2 report are presented in Table 10. According to this test, the observed test statistic is equal to 13.16, while the critical value is equal to 15.51, with $\alpha = 0.05$ and 8 degrees of freedom (p-value is 0.1064), the

Table 9. Result of Z-test for the first digit from the financial report of Hospital 2.

Digit	Count	Freq.	Benford's Law	Absolute difference	Z-Statistic
1	17	0.3469	0.301	0.0459	0.5453
2	8	0.1633	0.1761	0.0128	0.0483
3	4	0.0816	0.1249	0.0433	0.7001
4	7	0.1429	0.0969	0.0460	0.8460
5	1	0.0204	0.0792	0.0588	1.2594
6	1	0.0204	0.067	0.0466	1.0188
7	7	0.1429	0.058	0.0849	2.2357*
8	1	0.0204	0.0512	0.0308	0.6539
9	3	0.0612	0.0458	0.0154	0.1748
$\sum$	49	1	1		

Source: Authors' calculations.

null hypothesis can not be rejected, i.e., it can be concluded that first digit distribution of this report follows Benford's Law.

Table 10. Result of Chi-square test for the first digit from the financial report of Hospital 2.

Digit	Obs.freq. O_i	Benford's Law	Exp. freq. E_i	$(O_i - E_i)^2/E_i$
1	17	0.301	14.749	0.3435
2	8	0.1761	8.6289	0.0458
3	4	0.1249	6.1201	0.7344
4	7	0.0969	4.7481	1.0680
5	1	0.0792	3.8808	2.1385
6	1	0.067	3.283	1.5876
7	7	0.058	2.842	6.0834
8	1	0.0512	2.5088	0.9074
9	3	0.0458	2.2442	0.2545
$\sum$	49	1		13.1631

Source: Authors' calculations.

From the financial report of Hospital 3, we obtain the following results. Testing with Z-test, results for Hospital 3 report are presented in Table 11. Within the 84 observations in the considered sample, the differences (between observed and Benford's Law frequencies) which are presented in the table are not statistically significant at all digits (Z-statistic is less than 1.96).

Testing with the Chi-square test, results for Hospital 3 are presented in Table 12. According to this test, the observed test statistic is equal to 3.49, while the critical value is equal to 15.51, with $\alpha = 0.05$ and 8 degrees of freedom (p-value is 0.8998), the null hypothesis

Table 11. Result of Z-test for the first digit from the financial report of Hospital 3.

Digit	Count	Freq.	Benford's Law	Absolute difference	Z-Statistic
1	27	0.3214	0.301	0.0204	0.2892
2	13	0.1548	0.1761	0.0213	0.3702
3	6	0.0714	0.1249	0.0535	1.3173
4	10	0.1190	0.0969	0.0221	0.5018
5	8	0.0952	0.0792	0.0160	0.3423
6	5	0.0595	0.067	0.0075	0.0559
7	5	0.0595	0.058	0.0015	0.0597
8	5	0.0595	0.0512	0.0083	0.0986
9	5	0.0595	0.0458	0.0137	0.3407
$\sum$	84	1	1		

Source: Authors' calculations.

can not be rejected, etc. it can be concluded that first digit distribution of this report follows Benford's Law.

Table 12. Result of Chi-square test for the first digit from the financial report of Hospital 3.

Digit	Obs. freq. O_i	Benford's Law	Exp. freq. E_i	$(O_i - E_i)^2/E_i$
1	27	0.301	25.284	0.1165
2	13	0.1761	14.7924	0.2172
3	6	0.1249	10.4916	1.9229
4	10	0.0969	8.1396	0.4252
5	8	0.0792	6.6528	0.2728
6	5	0.067	5.628	0.0701
7	5	0.058	4.872	0.0034
8	5	0.0512	4.3008	0.1137
9	5	0.0458	3.8472	0.3454
$\sum$	84	1		3.4872

Source: Authors' calculations.

From previous analysis and results presented in the tables, we may conclude that for two of the three considered hospitals Chi-square tests show confirmation with the Benford distribution. We recommend the application of Benford's Law in the field of health as a first step in detecting fraud and irregularities in financial reporting.

5. Conclusions

Our objective was to identify additional distributions, beyond those already studied, that adhere to the Benford distribution. We enhance the theoretical studies of Benford's

Law by presenting, for the first time, its conformity with two distributions. Furthermore, we illustrate their behavior under parameter variations through simulations and statistical testing. In this paper, we gave theoretical results that the Pareto distribution and appropriate constant times Weibull density function, under some parameter constraint when shape parameter a is sufficiently small, obey Benford's Law. The simulation method was used to illustrate the alignment of the Pareto distribution with Benford's Law. On the other side, based on the data from the balance sheet and income statement of the three analyzed private hospitals, which fit well with the Weibull distribution, we confirmed that the Weibull distribution also obeys the Benford distribution, what is already known results from the literature. Certainly, even if we had obtained a different result, it could only be interpreted as a call for caution and undertaking further investigations. We provided simulation studies or empirical evidence supporting the primary findings in the paper, demonstrating that Pareto and Weibull distributions adhere to Benford's Law under certain parameter constraints.

We also verify the importance of the application of Benford's Law in practical problems, through widely considered literature and therein results and we confirmed that Benford's analysis can be employed as a useful tool to detect possible fraud and suspect reports for further analysis, with testing first (what we had done in this paper), second or k-th digit and also k digits simultaneously (what is out of the scope of this paper, but we gave the main formulas). Moreover, through the simulation or real data analysis of data following the Pareto and Weibull distribution, we further corroborated the results by subjecting them to statistical tests, i.e., Z-test and Chi-test to confirm that those distributions obey Benford's Law.

In future research, more distributions could be considered whether they obey the Benford distribution, by applying analog theorems as the main results of this paper and also it is possible to apply theoretical results of fitting Benford distribution on the observed dataset with the new test methods. Highlighting the conformity of Benford's characteristics to the Pareto and Weibull distributions, which are widely utilized in various modeling fields, these findings can be applied in numerous practical studies.

Acknowledgements

We are thankful for the helpful and clear suggestions from the editor and reviewers. These comments improved our paper, especially in its structure and clarity.

Author contributions. All the co-authors have contributed equally in all aspects of the preparation of this submission.

Conflict of interest statement. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding. This work is not supported by any funding.

Data availability. The data utilized in the research presented in this article are accessible at www.apr.gov.rs.

References

[1] F. Balado and G.C. Silvestre, *General distributions of number representation elements*, Probab. Eng. Inf. Sci., 1-23, 2024.

[2] V. Balashov, Y. Yuxing and Z. Xiaodi, *Who Manipulates Data During Pandemics? Evidence from Newcomb-Benford Law*, Evidence from Newcomb-Benford Law, arXiv preprint arXiv:2007.14841, 2020.

[3] A. Bayerstadler, L. van Dijk and F. Winter, *Bayesian multinomial latent variable modeling for fraud and abuse detection in health insurance*, Insur.: Math. Econ. **71**, 244-252, 2016.

[4] F. Benford, *The law of anomalous numbers*, Proc. Am. Philos. Soc., 551-572, 1938.

[5] L. Bermúdez, J.M. Pérez, M. Ayuso, E. Gómez and F.J. Vazquez, *A Bayesian dichotomous model with asymmetric link for fraud in insurance*, Insur.: Math. Econ. **42** (2), 779-786, 2008.

[6] C. Carslaw, *Anomalies in income numbers: Evidence of goal oriented behavior*, Account. Rev., 321-327, 1998

[7] V. Cuff, A. Lewis and S.J. Miller, *The Weibull distribution and Benford's Law*, Involve, a Journal of Mathematics **8** (5), 859-874, 2015.

[8] D. Djoric, J. Malisic, V. Jevremovic and E. Nikolic-Djoric, *Atlas raspodela*, Gradjevinski fakultet, Beograd, 2007.

[9] R.F. Durst, C. Huynh, A. Lott, S.J. Miller, E.A. Palsson, W. Touw and G. Vriend, *The Inverse Gamma Distribution and Benford's Law*, arXiv preprint arXiv:1609.04106, 2016.

[10] C. Durtschi, W. Hillison and C. Pacini, *The effective use of Benford's Law to assist in detecting fraud in accounting data*, Journal of Forensic Accounting **5** (1), 17-34, 2004

[11] T. El Sehity, E. Hoelzl and E. Kirchler, *Price developments after a nominal shock: Benford's Law and psychological pricing after the euro introduction*, Int. J. Res. Mark. **22** (4), 471-480, 2005.

[12] H.A. Engel and C. Leuenberger, *Benford's Law for exponential random variables*, Stat. Probab. Lett. **63** (4), 361-365, 2003.

[13] G. Fang and Q. Chen, *Several common probability distributions obey Benford's Law*, Phys. A: Stat. Mech. Appl. **540**, 123129, 2020.

[14] D.E. Giles, *Benford's Law and naturally occurring prices in certain eBay auctions*, Appl. Econ. Lett. **14** (3), 157-161, 2007.

[15] J. Golbeck, *Benford's Law applies to online social networks*, PLoS One **10** (8), 2015.

[16] J. Gonzalez-Garcia and G.C. Pastor, *Benford's Law and macroeconomic data quality*, International Monetary Fund, 2009.

[17] R.C. Hill, *Statistical relationship involving Benford's Law, the Lognormal distribution, and the Summation theorem*, arXiv:1710.0129, viXra.org.

[18] T.P. Hill, *Base-invariance implies Benford's Law*, Proc. Am. Math. Soc. **123** (3), 887-895, 1995.

[19] T.P. Hill, *A statistical derivation of the significant-digit law*, Stat. Sci., 354-363, 1995.

[20] R. Hindls and S. Hronova, *Benford's Law and possibilities for its use in governmental statistics*, Statistika: Statistics & Economy Journal **95** (2), 2015.

[21] H. Jovšić and B. Žmuk, *Assessing the quality of Covid-19 data: evidence from Newcomb-Benford Law*, Facta Universitatis, Series: Economics and Organization, 135-156, 2021.

[22] Z. Krakar and M. Žgela, *Application of Benford's Law in payment systems auditing*, Journal of Information and Organizational Sciences **33** (1), 39-51, 2009.

[23] L.M. Leemis, B.W. Schmeiser and D.L. Evans, *Survival distributions satisfying Benford's Law*, Am. Stat. **54** (4), 236-241, 2000.

[24] M. Meiryani, G. Soepriyanto, D. Wahyuningtias and K. Dewi, *Accounting Perspective in Hospital*, International Journal of Online and Biomedical Engineering **16** (8), 2020.

[25] T. Michalski and G. Stoltz, *Do countries falsify economic data strategically? Some evidence that they might*, Rev. Econ. Stat. **95** (2), 591-616, 2013.

[26] J. Moreno-Montoya, *Benford's Law with small sample sizes: A new exact test useful in health sciences during epidemics*, Revista de la Universidad Industrial de Santander. Salud **52** (2), 161-163, 2020.

[27] S. Newcomb, *Note on the frequency of use of the different digits in natural numbers*, Am. J. Math. **4** (1), 39-40, 1881.

[28] M.J. Nigrini, *A taxpayer compliance application of Benford's Law*, The Journal of the American Taxation Association **18** (1), 72, 1996.

[29] M.J. Nigrini and L.J. Mittermaier, *The use of Benford's Law as an aid in analytical procedures*, Auditing **16** (2), 52, 1997.

[30] M.J. Nigrini and S.J. Miller, *Data diagnostics using second-order tests of Benford's Law*, Auditing: A Journal of Practice & Theory **28** (2), 305-324, 2009.

[31] M.J. Nigrini, *Benford's Law: Applications for forensic accounting, auditing, and fraud detection*, John Wiley & Sons, 2012.

[32] J. Nye and C. Moul, *The political economy of numbers: on the application of Benford's Law to international macroeconomic statistics*, The BE Journal of Macroeconomics **7** (1), 2007.

[33] M. Papić, N. Vudrić and K. Jerin, *Benfordov zakon i njegova primjena u forenzičkom računovodstvu*, Zbornik sveučilišta Libertas **1** (1-2), 153-172, 2017.

[34] R.A. Raimi, *The first digit problem*, Am. Math. Mon. **83** (7), 521-538, 1976.

[35] T. Rakonjac-Antić, *Penzijsko i zdravstveno osiguranje*, Publishing Centre, Faculty of Economics and Business, University of Belgrade, 2018.

[36] R.J. Rodriguez, *Reducing false alarms in the detection of human influence on data*, J. Account. Audit. Finance **19** (2), 141-158, 2004.

[37] P.D. Scott and M. Fasli, *Benford's Law: An empirical investigation and a novel explanation*, Unpublished manuscript, 2001.

[38] S. Slijepčević and B. Blašković, *Statistical detection of fraud in the reporting of Croatian public companies*, Financial theory and practice **38** (1), 81-96, 2014.

[39] J.K. Thomas, *Unusual patterns in reported earnings*, Account. Rev., 773-787, 1989.

[40] I. Tota, A. Aliaj and J. Lamçja, *The use of Benford's Law as a tool for detecting fraud in accounting data*, Interdisplinary Journal of Research and Development **3** (1), 73-77, 2016.

[41] H.R. Varian, *Benford's Law*, Am. Stat. **26** (3), 1972.

[42] M.S. Warshavsky, *Applying Benford's Law in financial forensic investigations*, National Litigation Consultant's Review **10** (2), 1-4, 2010.

[43] J. Zhang, *Testing case number of coronavirus disease 2019 in China with Newcomb-Benford Law*, arXiv preprint arXiv:2002.05695, 2020.