Chilean Journal of Statistics
Vol. 17, No. 1, June/July 2026, 68–87
DOI: 10.32372/ChJS.17-01-04

Forensic Statistics
Research Article

When the Benford law might fail: A corrected first-digit test for power-law distributions

Francisco Bustillos[1] and Víctor Leiva[2,*]

[1]Independent Researcher, Santiago, Chile

[2]School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

(Received: 20 May 2026 · Accepted in final form: 20 June 2026)

Abstract

Goodness-of-fit tests for the Benford law can falsely flag authentic data when their first digits arise from an inverse power law. In this article, we show that, for any finite power-law exponent, the chi-squared test rejects such data with probability tending to one as the sample size increases, which establishes a model-misspecification problem. To address this problem, we develop a corrected method based on the power-law first-digit statistical distribution. In addition, our method jointly estimates the power-law exponent and scale phase by using the minimum chi-squared technique and measures the remaining deviation from the fitted first-digit baseline. Under regularity conditions, the associated corrected statistic follows an asymptotic chi-squared distribution with six degrees of freedom. Furthermore, a bootstrap equivalence decision is used to evaluate compatibility with the fitted power-law first-digit family. In reproducible simulations and public datasets, the corrected method distinguishes data compatible with a unit-scale power-law first-digit distribution from the selected fabricated, retracted, and bounded-support benchmarks. The developed method is intended for quantities spanning several orders of magnitude.

Keywords: Benford model · Chi-squared test · Data fabrication · First-digit test · Forensic statistics · Goodness of fit · Inverse power law

Mathematics Subject Classification: Primary 62F03 · Secondary 11K06

1. Introduction

Testing whether the first digits of a dataset conform to the Benford law has become a standard technique for detecting data fabrication and manipulation in finance, public administration, and science [1]. The appeal is clear, because a chi-squared test provides an automated anomaly signal. However, users of the test encounter a persistent problem that has no clean solution in the existing literature. Large datasets that are perfectly authentic are routinely flagged as suspicious. The flag appears not because the data were manipulated, but because the data follow an inverse power law, a model for which the baseline Benford law is demonstrably wrong. Then, two questions can be asked. The first is when such false positives are mathematically inevitable. The second is how to correct for them rigorously.

*Corresponding author. Email: victorleivasanchez@gmail.com (V. Leiva)

In 1938, the physicist Frank Benford [2] observed a rule indicating that the first nonzero digit of numbers appearing in tables of physical constants, river areas, and other datasets follows a non-uniform statistical distribution. Throughout this article, the term first digit refers to the first nonzero digit of a positive number. The corresponding first-digit distribution assigns to each digit from 1 to 9 the probability that it appears as the first digit of an observation. According to the Benford rule, the digit 1 occurs approximately thirty percent of the time, whereas the digit 9 occurs less than five percent of the time. This rule is known as the Benford law. Hill's scale-invariance characterization [3] shows that the Benford law is the unique first-digit distribution that remains invariant under changes in measurement scale. Hill's statistical derivation also explains why mixtures of data drawn from different distributions may converge to the Benford law [3]. The rule has been widely used as a screening tool in accounting, auditing, and fraud detection [4, 5], and has also been investigated in election forensics [6, 7].

The main contribution of the present article is an inferential framework built on the generalized Benford family introduced in [8]. This one-parameter family describes the first-digit probabilities induced by a power law and coincides, under an equivalent reparameterization, with the unit-scale model considered in the present article. To emphasize its generating mechanism, we refer to it here as the unit-scale power-law first-digit distribution. The present article extends this family by explicitly incorporating the scale phase induced by the multiplicative constant and develops four inferential components based on the resulting model.

The first component is a false-positive result showing that, for every finite power-law exponent, the traditional chi-squared test for the Benford law rejects authentic data generated from the power law with probability tending to one as the sample size increases. The result also provides an exponent-dependent reference sample size at which the departure from the Benford law reaches the usual rejection threshold.

The second is a corrected chi-squared test calibrated with the appropriate degrees of freedom after estimating the parameters of the unit-scale power-law first-digit distribution.

The third is a joint minimum-chi-squared estimator of the power-law exponent and scale phase of the unit-scale power-law first-digit distribution induced by the multiplicative constant.

The fourth is a contamination model in which observations generated from the unit-scale power-law first-digit distribution are mixed with Benford-conforming observations. Under this model, the structure index, defined as a Pearson-type measure of the discrepancy between the first-digit distribution and the Benford law, decreases quadratically with the proportion of the power-law component retained in the mixture. Therefore, conditional on the specified contamination mechanism, the index can be interpreted as a measure of retained power-law structure.

The remainder of the article is organized as follows. Section 2 reviews the Benford law, chi-squared test and power-law model. In Section 3, we derive the unit-scale power-law first-digit distribution, the false positive theorem, and the corrected test. In Section 4, simulation studies are reported to evaluate the corrected test, whereas Section 5 applies the test to real data. In Section 6, we close with some concluding remarks. The proofs of theorems, lemmas, and propositions are in Appendix.

2. Background

2.1 *The Benford law and chi-squared test*

Let Y be a positive random variable and $D \in \{1, \dots, 9\}$ denote the first nonzero digit in the decimal representation of Y.

Under the Benford law, the probability mass function of D is given by

$$\mathbb{P}(D = d) = \mathbb{P}(d) = \log_{10}\left(1 + \frac{1}{d}\right), \quad d \in \{1, \ldots, 9\}. \tag{1}$$

The distribution formulated in (1) is known as the Benford law. The hypotheses for assessing conformity with this law, $\mathbb{P}(d)$ namely, are defined as

$$\mathrm{H}_0\colon \mathbb{P}(D = d) = \mathbb{P}(d), \forall\, d \in \{1, \ldots, 9\}, \quad \text{versus} \quad \mathrm{H}_1\colon \mathbb{P}(D = d) \neq \mathbb{P}(d), \exists\, d \in \{1, \ldots, 9\}. \tag{2}$$

For a dataset of n positive real numbers, let $P_{\mathrm{obs}}(d) = n_d/n$, where n_d denotes the observed number of first digits equal to d. The Pearson chi-squared statistic for the Benford hypotheses stated in (2) is formulated as

$$\chi^2 = n \sum_{d=1}^{9} \frac{\big(P_{\mathrm{obs}}(d) - \mathbb{P}(d)\big)^2}{\mathbb{P}(d)}, \tag{3}$$

which is asymptotically chi-squared distributed with eight degrees of freedom, denoted by $\chi^2(8)$, under the null hypothesis H_0. The decision rule rejects the Benford hypothesis H_0 at a significance level α when the calculated value of χ^2 defined in (3) exceeds the $(1-\alpha)\times 100$th percentile of its corresponding null distribution [9]. Hereafter, $\chi^2_q(\nu)$ denotes the upper-tail critical value satisfying $\mathbb{P}(\chi^2 > \chi^2_q(\nu)) = q$.

2.2 *Power-law model*

Many scientific relationships can be represented approximately by an inverse power law of the form $Y = CX^{-k}$, where X and Y are positive quantities, $C > 0$ is a scale constant, and $k > 0$ is the power-law exponent. Inverse-square relationships with $k = 2$ arise in gravitation and electrostatics, inverse-cube relationships with $k = 3$ arise in ideal dipole fields, and reciprocal relationships with $k = 1$ may arise in some kinetic settings. Such a relationship does not, by itself, determine the distribution of Y, because this distribution also depends on the distribution, support, and sampling mechanism of X.

Power-law behavior also arises asymptotically in the extremes of heavy-tailed distributions. Non-Gaussian stable models exhibit asymptotic power-law decay and have been used to describe financial data [10]. Although a stable distribution is not generally equivalent to the inverse power-law distribution considered here, its tails exhibit power-law decay and therefore provide an additional setting in which first-digit behavior may be relevant.

In a related but distinct formulation, power functions also arise in proportionate-effect models, where the magnitude of each new impulse depends on the current level of an accumulated process through a function such as $g(z) = z^{\delta}$. This setting has been used in environmental and medical modeling [11, 12]. Although this stochastic accumulation process differs from the inverse power-law relationship $Y = CX^{-k}$ considered here, both formulations describe scale-dependent effects governed by a power function.

For the probabilistic formulation of the present article, let Y be a continuous positive random variable, with observed value y, whose PDF is proportional to $y^{-(1+1/k)}$ over its support. This PDF is induced, for example, by the transformation $Y = CX^{-k}$ when X is uniformly distributed over a suitable positive interval. The corresponding first-digit analysis is conducted under the support conditions stated in the following section.

A generalized Benford law showing that the first-digit probabilities induced by a power law depart from those of the Benford law or distribution was stated in [8]. A size-dependent first-digit law for the prime-counting sequence was studied in [13], whereas the analysis was extended to multidigit distributions in [14]. The present article quantifies the consequences of this departure and develops a corrected test.

The continuous distribution is the primary model considered here. The first-digit results hold for every $k > 0$, subject to the support conditions stated in the following section. Throughout this article, $\mathrm{frac}(x) = x - \lfloor x \rfloor$ denotes the fractional part of a real number x. The effect of the multiplicative constant C on the first digits is represented by the scale phase $\varphi = \mathrm{frac}(\log_{10}(C)) \in [0, 1)$, as shown in Proposition 3.4.

A deterministic counterpart of the continuous distribution considered in the present article is given by the sequence $Y_j = Cj^{-k}$, for $j \in \{1, \ldots, J\}$. Its empirical first-digit behavior depends on the exponent, scale phase, and endpoint phase. Because the inferential results developed in the following section are formulated for independent first-digit counts, this deterministic sequence is not used as the data-generating model for the corrected test.

3. Theoretical and methodological results

3.1 *The structure index*

The structure index is an important indicator in the context of our work and is defined next.

Definition 3.1 (Structure index) For a dataset of n positive real numbers with observed first-digit relative frequencies $P_{\mathrm{obs}}(d) = n_d/n$, the structure index is given by

$$\psi = \sum_{d=1}^{9} \frac{\left(P_{\mathrm{obs}}(d) - \mathbb{P}(d)\right)^2}{\mathbb{P}(d)}, \tag{4}$$

where $\mathbb{P}(d)$ is stated in (1).

Note that the product $n\psi$ based on the expression presented in (4) equals the Pearson statistic defined in (3). When n is fixed, ψ is proportional to the statistic χ^2 established in (3) and inherits its distributional properties. The value $\psi = 0$ marks perfect Benford conformity. For large n, even a small deviation becomes statistically significant.

As $\chi^2 = n\psi$, the critical values of the Pearson chi-squared test can be expressed directly on the scale of the structure index. At significance level 0.01, the Benford null hypothesis is rejected when a calculated value of ψ, $\psi_n > 20.09/n$, where 20.09 is the upper one-percent critical value of the $\chi^2(8)$ distribution. Similarly, a value $\psi_n < 2.73/n$ falls in the lower five-percent tail of this distribution and may indicate anomalous over-conformity with the Benford law. For $n = 100$, the corresponding thresholds are approximately 0.20 and 0.027, respectively.

3.2 *Unit-scale power-law first-digit distribution*

The starting point is that, as the integer number indexed by r ranges over $\{1, \ldots, L\}$, the fractional parts, $\mathrm{frac}(\log_{10}(r))$ say, are not uniformly distributed on $[0, 1)$. The argument partitions the range into orders of magnitude and counts the integers on the logarithmic scale, as formalized in the following lemma.

Lemma 3.2 (Logarithmic PDF) Let $L = 10^M$ and consider the empirical distribution of the fractional parts of $\log_{10}$ of the integer numbers indexed by $r \in \{1, \ldots, L\}$, $\mathrm{frac}(\log_{10}(r))$ namely, for $r \in \{1, \ldots, L\}$. Then, as $M \to \infty$, this empirical distribution converges to a continuous distribution on $[0, 1)$ whose PDF is defined as

$$f(t) = \frac{\ln(10)}{9}\, 10^t, \quad t \in [0, 1). \tag{5}$$

The PDF stated in (5) is increasing, with ratio $f(1^-)/f(0) = 10$. More integer numbers lie between 500 and 1000 than between 100 and 500 per unit of $\log_{10}$, so that the folded version of this PDF produces the unit-scale power-law first-digit distribution stated in the following theorem.

THEOREM 3.3 (Unit-scale power-law first-digit distribution) Let Y have a PDF proportional to $y^{-(1+1/k)}$ on $[10^A, 10^B]$, where $A < B$ are integer numbers, and D denote the first digit of Y. Then, if $\mathbb{P}(d;k)$ is defined in (6), the probability that $D = d$ is stated as

$$\mathbb{P}(d;k) = \frac{10^{1/k}}{10^{1/k}-1}\left(d^{-1/k} - (d+1)^{-1/k}\right), \quad d \in \{1,\ldots,9\}, k > 0. \tag{6}$$

Remark 1 (Relation to the generalized Benford law) The unit-scale power-law first-digit family stated in (6) is the generalized Benford law introduced in [8], expressed under a different parameterization. Under the reparameterization $\beta = -1/k$, the first-digit probabilities are established as

$$\mathbb{P}(d;\beta) = \frac{(d+1)^\beta - d^\beta}{10^\beta - 1}, \quad d \in \{1,\ldots,9\}, \beta < 0.$$

Theorem 3.3 states the explicit condition that the support endpoints be aligned with integer orders of magnitude, under which these probabilities follow exactly from the continuous distribution considered here.

3.3 *Scale phase*

Power-law relationships may include a multiplicative constant. Because the unit-scale power-law first-digit distribution is not scale invariant, the effect of this constant must be incorporated through its phase, as established in the following proposition.

PROPOSITION 3.4 (Scale-phase family) Let Y_0 follow the continuous unit-scale power-law distribution of Theorem 3.3 and let $Y = C\,Y_0$, for $C > 0$. Let $D \in \{1,\ldots,9\}$ denote the first digit of Y and state the scale phase as $\varphi = \mathrm{frac}(\log_{10}(C)) \in [0,1)$. Let G represent the cumulative distribution function (CDF) of $\mathrm{frac}(\log_{10}(Y_0))$ given by

$$G(y;k) = \begin{cases} 0, & y < 0; \\ \dfrac{10^{-y/k}-1}{10^{-1/k}-1}, & 0 \le y \le 1; \\ 1, & y > 1. \end{cases}$$

Define the extension of G to the real line by means of

$$G_{\mathrm{ext}}(y;k) = \lfloor y \rfloor + G(\mathrm{frac}(y);k) = \lfloor y \rfloor + \frac{10^{-\,\mathrm{frac}(y)/k}-1}{10^{-1/k}-1}, \quad y \in \mathbb{R}.$$

This extension satisfies $G_{\mathrm{ext}}(y+1;k) = G_{\mathrm{ext}}(y;k)+1$. Then, the probability mass function of D is formulated as

$$\mathbb{P}(d;k,\varphi) = G_{\mathrm{ext}}(\log_{10}(d+1)-\varphi;k) - G_{\mathrm{ext}}(\log_{10}(d)-\varphi;k), \quad d \in \{1,\ldots,9\}. \tag{7}$$

At $\varphi = 0$, the unit-scale family is recovered, that is, $\mathbb{P}(d;k,0) = \mathbb{P}(d;k)$. For $\varphi \neq 0$, the first-digit probabilities may differ from their unit-scale values. Therefore, the unit-scale power-law first-digit family is not scale invariant.

Because the first-digit probabilities depend on φ, the procedure developed in this article estimates the scale phase jointly with the power-law exponent when fitting the unit-scale power-law first-digit distribution.

3.4 *Limiting behavior and deviation*

As the exponent grows, the power-law distribution collapses onto the Benford baseline, as established in the following corollary.

Corollary 3.5 (Benford law as the limiting case) Let $\mathbb{P}(d)$ be defined in (1) and $\mathbb{P}(d;k)$ in (6), with $d \in \{1,\dots,9\}$. Then, we have that $\lim_{k\to\infty} \mathbb{P}(d;k) = \mathbb{P}(d) = \log_{10}(1+1/d)$.

Corollary 3.5 establishes convergence to the Benford law as the exponent tends to infinity. The following proposition strengthens this limiting result by showing that the Pearson-type deviation from the Benford law, quantified by $\psi(k)$, decreases strictly as k increases.

Proposition 3.6 (Monotone deviation) Let $\mathbb{P}(d)$ be defined in (1), $\mathbb{P}(d;k)$ in (6), for $k > 0$, and $\psi(k) = \sum_{d=1}^{9}(\mathbb{P}(d;k) - \mathbb{P}(d))^2/\mathbb{P}(d)$. Then, $\psi(k)$ is strictly decreasing in k.

3.5 *The false-positive theorem*

Proposition 3.6 shows that the deviation $\psi(k)$ from the Benford law is strictly positive for every finite $k > 0$. Therefore, although this deviation decreases as k increases, the unit-scale power-law first-digit distribution does not coincide exactly with the Benford law at any finite exponent. The consequence for the traditional Pearson chi-squared test is formalized in the following theorem.

Theorem 3.7 (False positives from traditional Benford testing) Let $\mathbb{P}(d)$ be defined in (1), $\mathbb{P}(d;k)$ in (6), and $D_1,\dots,D_n$ be independent and identically distributed first digits satisfying $\mathbb{P}(D_i = d) = \mathbb{P}(d;k)$, for $i \in \{1,\dots,n\}$, $d \in \{1,\dots,9\}$, and some fixed $k > 0$. Let χ^2 denote the Pearson chi-squared statistic for testing the Benford law. Then, we have that

$$\frac{\chi^2}{n} \xrightarrow{\mathbb{P}} \psi(k) = \sum_{d=1}^{9} \frac{(\mathbb{P}(d;k) - \mathbb{P}(d))^2}{\mathbb{P}(d)} > 0,$$

where $\xrightarrow{\mathbb{P}}$ denotes convergence in probability to. Consequently, for every fixed significance level $\alpha \in (0,1)$, $\mathbb{P}(\chi^2 > \chi^2_\alpha(8)) \longrightarrow 1$, as $n \to \infty$.

Here, a false positive refers to an anomalous flag for data correctly generated from the power-law first-digit model, rather than to a type-I error under the Benford null hypothesis.

Theorem 3.7 establishes that rejection becomes asymptotically inevitable, but it does not provide a guaranteed finite-sample rejection threshold. Since the leading deterministic term of the statistic is $n\psi(k)$, an exponent-specific reference sample size can be obtained by equating this term to the five-percent critical value of the chi-squared distribution with eight degrees of freedom. Accordingly, define

$$n_{\min}(k) = \left\lceil \frac{\chi^2_{0.05}(8)}{\psi(k)} \right\rceil = \left\lceil \frac{15.51}{\psi(k)} \right\rceil. \tag{8}$$

The quantity $n_{\min}(k)$ is a reference scale indicating the sample size at which the leading term $n\psi(k)$ first reaches the usual five-percent critical value. It is not a sample size at which rejection is guaranteed. By Proposition 3.6 and Corollary 3.5, this reference sample size is finite for every finite $k > 0$, is nondecreasing in k, and diverges as $k \to \infty$, as formalized in the following corollary.

Corollary 3.8 (Behavior of the reference sample size) Under the assumptions of Theorem 3.7, $n_{\min}(k)$ defined in (8) is finite for every finite $k > 0$, is nondecreasing in k, and satisfies $n_{\min}(k) \longrightarrow \infty$ as $k \to \infty$.

Corollary 3.8 shows that larger exponents require increasingly large samples for the leading departure from the Benford law to reach the same fixed critical value. Table 1 illustrates this reference scale for selected exponents. These values are numerical examples and must not be interpreted as guaranteed rejection thresholds.

Table 1. Reference sample size $n_{\min}(k)$ at which the leading term $n\psi(k)$ equals the five-percent chi-squared critical value under the unit-scale power-law first-digit model.

k	$n_{\min}(k)$	Illustrative case
1	43	Reciprocal relationship
2	154	Gravitational inverse-square relationship
3	338	Ideal dipole inverse-cube relationship
4	595	Higher-order inverse power law
5	924	Higher-order inverse power law

3.6 *The corrected test*

The false-positive result of Theorem 3.7 arises because the traditional Pearson statistic uses the Benford law as its reference distribution, even when the first digits are generated by a power-law model. Therefore, the natural correction is to replace the Benford probabilities with the probabilities of the unit-scale power-law first-digit distribution, as established in the following theorem.

Proposition 3.9 (Corrected test) Let $\mathbb{P}(d; k, \varphi)$ be stated in (7), $(O_1, \ldots, O_9) \sim$ Multinomial$(n; \mathbb{P}(1; k, \varphi), \ldots, \mathbb{P}(9; k, \varphi))$, and define

$$\chi^2_{\text{corrected}}(k, \varphi) = \sum_{d=1}^{9} \frac{(O_d - n\mathbb{P}(d; k, \varphi))^2}{n\mathbb{P}(d; k, \varphi)}, \tag{9}$$

where O_d denotes the number of observations whose first digit equals d. If (k, φ) is known, then $\chi^2_{\text{corrected}}(k, \varphi) \xrightarrow{\mathbb{D}} \chi(8)^2$, where $\xrightarrow{\mathbb{D}}$ denotes convergence in distribution to. If both parameters k and φ are estimated, $\widehat{k}$ and $\widehat{\varphi}$ say, from the same first-digit counts, then $\chi^2_{\text{corrected}}(\widehat{k}_n, \widehat{\varphi}_n) \xrightarrow{\mathbb{D}} \chi^2(6)$, assuming that the true parameter is an interior point of the parameter space, the cell probabilities are strictly positive, and the Jacobian of the probability vector with respect to (k, φ) has rank two. If $\varphi = \varphi_0$ is fixed and only k is estimated under the analogous one-parameter regularity conditions, then $\chi^2_{\text{corrected}}(\widehat{k}_n, \varphi = \varphi_0) \xrightarrow{\mathbb{D}} \chi^2(7)$.

The reductions from eight to six or seven degrees of freedom follow from the regular minimum-chi-squared theory for grouped data [15, 16]. These limiting distributions require independent multinomial first-digit counts and correct specification of the first-digit distribution. They do not directly apply to data with dependence, repeated measurements, rounding, truncation, or clustering. Such settings require a calibration or resampling scheme adapted to their sampling structure.

Before considering the joint estimation of the exponent and scale phase, it is helpful to establish that the exponent is identifiable in the unit-scale subfamily, as shown in the following proposition.

Proposition 3.10 (Identifiability of the power-law exponent) Let $\mathbb{P}(d; k)$ be defined in (6) and $k, k' > 0$. If $\mathbb{P}(d; k) = \mathbb{P}(d; k')$ for every $d \in \{1, \ldots, 9\}$, then $k = k'$.

Thus, two distinct exponents cannot generate the same unit-scale first-digit probability vector. The following theorem gives an asymptotic consequence of the identifiability result established in Proposition 3.10 by showing that a Pearson statistic based on an incorrect fixed exponent diverges linearly with the sample size.

Theorem 3.11 (Asymptotic separation of distinct exponents) Let $\mathbb{P}(d;k)$ be defined in (6), for $k, k' > 0$, with $k \neq k'$, and define

$$\Delta(k,k') = \sum_{d=1}^{9} \frac{(\mathbb{P}(d;k') - \mathbb{P}(d;k))^2}{\mathbb{P}(d;k)},$$

where $\Delta(k,k') > 0$. Let $D_1,\ldots,D_n$ be independent and identically distributed first-digit random variables satisfying $\mathbb{P}(D_i = d) = \mathbb{P}(d;k')$, for $i \in \{1,\ldots,n\}$ and $d \in \{1,\ldots,9\}$. Let O_d be stated as in Proposition 3.9 and define the Pearson statistic computed against the fixed baseline associated with k as

$$\chi^2_n(k) = \sum_{d=1}^{9} \frac{(O_d - n\mathbb{P}(d;k))^2}{n\mathbb{P}(d;k)}.$$

Then, $\chi^2_n(k)/n \xrightarrow{\mathbb{P}} \Delta(k,k')$. Consequently, the probability of rejecting the fixed baseline $\mathbb{P}(d;k)$ tends to one as $n \to \infty$.

Proposition 3.10 and Theorem 3.11 concern the unit-scale family. In the general model, the exponent and scale phase must be estimated jointly from the first-digit frequencies.

To implement this estimation, let $K > 1$ be a fixed finite upper bound and restrict the exponent to $k \in [1, K]$. The lower bound $k \geq 1$ is a modeling restriction adopted to prevent combinations of small exponents and scale phases from fitting certain bounded-support or single-peaked first-digit patterns. It is not required for the definition of the continuous power-law family, which is valid for every $k > 0$.

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denote the circular phase space, where the slash denotes the quotient-space operation under which real numbers that differ by an integer represent the same phase. Equivalently, phases are interpreted modulo one. The following theorem establishes the conditional consistency of the joint estimator of the exponent and scale phase over this parameter space.

Theorem 3.12 (Conditional consistency of the exponent and scale-phase estimator) Let $\mathbb{P}(d;k,\varphi)$ be stated in (7) and $D_1,\ldots,D_n$ be independent and identically distributed first-digit random variables satisfying $\mathbb{P}(D_i = d) = \mathbb{P}(d;k_0,\varphi_0)$, for $i \in \{1,\ldots,n\}$ and $d \in \{1,\ldots,9\}$, where $(k_0,\varphi_0) \in [1,K] \times \mathbb{T}$. Let $P_n(d)$ denote the empirical first-digit frequency of digit d and let

$$(\widehat{k}_n, \widehat{\varphi}_n) \in \underset{(k,\varphi)\in[1,K]\times\mathbb{T}}{\arg\min} \left\{ \sum_{d=1}^{9} \frac{(P_n(d) - \mathbb{P}(d;k,\varphi))^2}{\mathbb{P}(d;k,\varphi)} \right\}. \tag{10}$$

Define the criterion stated as

$$Q_0(k,\varphi) = \sum_{d=1}^{9} \frac{(\mathbb{P}(d;k_0,\varphi_0) - \mathbb{P}(d;k,\varphi))^2}{\mathbb{P}(d;k,\varphi)}.$$

If (k_0,φ_0) is the unique minimizer of Q_0 over $[1,K] \times \mathbb{T}$, then $(\widehat{k}_n, \widehat{\varphi}_n) \xrightarrow{\mathbb{P}} (k_0,\varphi_0)$.

Joint injectivity of the map $(k,\varphi) \longmapsto \mathbb{P}(d;k,\varphi)$ is not established analytically in this article. Therefore, the consistency result is conditional on the unique-minimizer assumption stated in Theorem 3.12. Computationally, the phase space $\mathbb{T}$ is represented by $[0, 1)$, with phases interpreted modulo one.

An estimate satisfying $\widehat{k}_n < K$ corresponds to a fitted finite-exponent member in the interior of the optimization range. A fit attaining $\widehat{k}_n = K$ is reported as an upper-bound fit, not as an estimate equal to infinity. Such a fit may have first-digit probabilities numerically close to the Benford limit, in which the scale phase becomes weakly identified.

The corrected index measures the Pearson-type discrepancy between the observed first-digit frequencies and the jointly fitted power-law first-digit probability. Since $P_{\text{obs}}(d) = P_n(d)$, define

$$\psi_{\text{corrected}} = \frac{1}{n}\chi^2_{\text{corrected}}(\widehat{k}_n, \widehat{\varphi}_n) = \sum_{d=1}^{9} \frac{\left(P_{\text{obs}}(d) - \mathbb{P}(d; \widehat{k}_n, \widehat{\varphi}_n)\right)^2}{\mathbb{P}(d; \widehat{k}_n, \widehat{\varphi}_n)},$$

where $\chi^2_{\text{corrected}}$ is defined in (9).

For a first-digit probability vector $\boldsymbol{p} = (p_1, \ldots, p_9)$, the corresponding divergence is given by

$$\psi_\star = \inf_{(k,\varphi)\in[1,K]\times\mathbb{T}} \left\{ \sum_{d=1}^{9} \frac{(p_d - \mathbb{P}(d; k, \varphi))^2}{\mathbb{P}(d; k, \varphi)} \right\}.$$

Thus, $\psi_{\text{corrected}}$ is the empirical counterpart of $\psi_\star$. The value $\psi_\star = 0$ indicates exact membership in the fitted first-digit family, whereas a positive value measures the remaining departure after optimizing over the exponent and scale phase.

Because the statistic $n\psi_{\text{corrected}}$ is consistent against every fixed positive departure from the fitted family, a point-null goodness-of-fit test eventually detects any nonzero discrepancy as the sample size increases. Therefore, the practical question is whether the divergence is less than a prespecified tolerance $\delta > 0$, rather than whether it is exactly zero. To address this question, multinomial-bootstrap samples of size n are drawn from the empirical first-digit frequencies. The exponent and scale phase are re-estimated in every bootstrap sample, and the corrected index is recomputed. Let $u_{1-\alpha}$ denote the resulting one-sided upper $(1-\alpha)$ bootstrap confidence bound for $\psi_\star$. The following proposition formalizes the resulting bootstrap equivalence decision.

Proposition 3.13 (Bootstrap equivalence decision) Fix an equivalence margin $\delta > 0$ and a significance level $\alpha \in (0, 1)$. Consider the hypotheses H_0: $\psi_\star \geq \delta$ versus H_1: $\psi_\star < \delta$. Suppose that $u_{1-\alpha}$ is an asymptotically valid one-sided upper confidence bound for $\psi_\star$. Then, the decision rule that declares compatibility with the fitted power-law first-digit distribution when $u_{1-\alpha} < \delta$ defines a one-sided bootstrap equivalence test with asymptotic level at most α.

The finite-sample level, coverage, and power of this decision rule, as well as its behavior at the lower exponent boundary and at the upper optimization bound, are evaluated by simulation in Subsection 4.4. The multinomial bootstrap is calibrated for independent first-digit counts and does not by itself account for dependence or clustering.

3.7 *Contamination through a Benford mixture*

Consider a mixture of a power-law first-digit component and a Benford component, interpreted here as a contamination model in which the Benford component dilutes the power-law first-digit structure. The following theorem establishes that the structure index of the mixture equals the structure index of the power-law component multiplied by the squared mixture weight.

Theorem 3.14 (Quadratic scaling under contamination) Let $\mathbb{P}(d)$ be defined in (1), $\mathbb{P}(d; k, \varphi)$ in (7), $P_{\text{mix}}(d) = \rho\,\mathbb{P}(d; k, \varphi) + (1-\rho)\mathbb{P}(d)$, for $k > 0$, $\varphi \in [0, 1)$, $\rho \in [0, 1]$, and define

$$\psi(k, \varphi) = \sum_{d=1}^{9} \frac{(\mathbb{P}(d; k, \varphi) - \mathbb{P}(d))^2}{\mathbb{P}(d)}.$$

Then, the structure index of the mixture relative to the Benford law is given by $\psi_{\text{mix}} = \rho^2\psi(k, \varphi)$.

For an independent multinomial sample of size n collected from P_{mix}, let χ_n^2 denote the chi-squared statistic for testing the Benford law. Thus, $\chi_n^2/n \xrightarrow{\mathbb{P}} \rho^2\psi(k,\varphi)$ and

$$\mathbb{E}(\chi_n^2) = n\rho^2\psi(k,\varphi) + \sum_{d=1}^{9} \frac{P_{\text{mix}}(d)(1-P_{\text{mix}}(d))}{\mathbb{P}(d)}.$$

When $\rho = 0$, the second term equals eight, as under the Benford null hypothesis.

Under the specific mixture model stated in Theorem 3.14, the parameter ρ is the mixture weight of the power-law first-digit component. The following corollary uses the quadratic relation established in that theorem to obtain a consistent estimator of ρ^2 and the associated threshold property.

Corollary 3.15 (Mixture-weight criterion) Under the mixture model of Theorem 3.14, with (k,φ) known and $\psi(k,\varphi) > 0$, we have that $\widehat{\rho}^2 = (\chi_n^2/n)/\psi(k,\varphi)$ is a consistent estimator of ρ^2. Moreover, we get

$$\mathbb{P}\left(\chi_n^2 < \frac{n\psi(k,\varphi)}{4}\right) \longrightarrow \begin{cases} 1, & \rho < 1/2; \\ 0, & \rho > 1/2. \end{cases}$$

The interpretation of $\widehat{\rho}_n^2$ as an estimator of the squared mixture weight, and of the threshold event above as evidence that $\rho < 1/2$, is conditional on the specified two-component mixture model and does not apply to arbitrary forms of contamination.

4. Simulation studies

4.1 *Computational framework*

All computations were carried out in Python version 3.12 with the `numpy` and `scipy` libraries on a desktop workstation. The probabilities stated in (6), the structure index, and the minimum-chi-squared estimator defined in (10) are implemented in the openly available scripts of the repository listed in the data-availability statement provided by the authors, which reproduce the analytical reference values, the reference sizes $n_{\min}(k)$, contamination scaling, simulated baselines under a fixed seed, deterministic Fibonacci case, and bootstrap calibration of Table 3. The real-data cases of Section 5 are drawn from the public repositories listed in the data-availability statement also provided by the authors, each cited with its source and any retraction or correction notice.

The bound $u_{0.95}$ of Table 5 is computed by the multinomial bootstrap of Proposition 3.13. First, $B = 600$ resamples of the observed digit counts are drawn as multinomial draws of size n from the observed frequencies. Second, on each resample the pair $(\widehat{k},\widehat{\varphi})$ is re-estimated by using the formulation stated in (10) and $\psi_{\text{corrected}}$ is recomputed, so that the joint estimation is repeated inside the resampling. Third, $u_{0.95}$ is the upper 95th percentile of the resampled $\psi_{\text{corrected}}$, and the equivalence verdict is consistent when $u_{0.95} < \delta = 0.10$. This resampling scheme is used for numerical calibration of the equivalence decision, which is also applied to the boundary and saturation cases, for which the interior chi-squared asymptotics do not apply. Because the resampling is independent multinomial, it does not by itself absorb dependence such as clustering, which is left to the caveats of Proposition 3.9.

4.2 *Analytical reference values*

The closed form presented in (6) yields exact reference values for the deviation of authentic data generated from the power law. Table 2 lists, for $n = 1000$, the leading-digit probability $\mathbb{P}(1;k)$, the structure index $\psi(k)$, and the product $n\psi(k)$. The deviation decreases monotonically in k, in agreement with Proposition 3.6, and vanishes in the Benford limit of Corollary 3.5, that is, as $k \to \infty$ and $\mathbb{P}(d;k) \to \mathbb{P}(d)$.

Table 2. Analytical reference values for data generated by an inverse power law, with $n = 1000$, where the Benford limit corresponds to $k \to \infty$.

k	$\mathbb{P}(1;k)$	$\psi(k)$	$n\psi(k)$	Example
1	0.556	0.363	363.1	Chemical kinetics
2	0.428	0.101	101.0	Gravitation
3	0.385	0.046	46.0	Magnetic dipoles
4	0.364	0.026	26.1	Higher-order effects
5	0.351	0.017	16.8	Complex decay
∞	0.301	0	0	Benford limit

For the gravitational exponent $k = 2$, the digit one is overrepresented relative to the Benford value 0.301 by a factor of 1.42, and the product $n\psi(k) = 101$ far exceeds the critical value 15.51. Thus, a traditional test on a few hundred authentic measurements rejects the Benford hypothesis with near certainty, which is the content of Theorem 3.7.

4.3 *Power-law and fabricated baselines*

Two controlled baselines isolate the behavior of the two tests. A power-law baseline draws $n = 2000$ first digits from the theoretical distribution $\mathbb{P}(d;k)$, with $k = 2$, and represents authentic data generated from a known power law. A fabricated baseline draws $n = 2000$ values with a flat first-digit distribution, and mimics manufactured data with no natural structure. The index of the power-law baseline is 0.107, which the rule reads as borderline, while the joint fit gives a corrected index of 0.002, well inside the equivalence margin. The fabricated baseline gives a corrected index of 0.245, far outside it. Therefore, the two baselines separate cleanly under the corrected test and not under the traditional test. The behavior on public datasets is reported next.

4.4 *Calibration of the equivalence decision*

The inferential properties of the equivalence decision of Proposition 3.13 are assessed by simulation, since its level and coverage are asymptotic claims. For each target population, the corrected index $\psi_{\text{corrected}}$ is computed by the joint minimum-chi-squared fit, and the decision "consistent" when $u_{0.95} < \delta = 0.10$ is run on $R = 400$ independent samples of size $n = 2000$, with $B = 400$ bootstrap resamples per decision, at a significance level $\alpha = 0.05$ and a fixed seed. Table 3 reports the results. For the level rows, the displayed values of (k, φ) identify the reference power-law first-digit distribution. The true generating distributions are perturbed nine-cell probability vectors numerically calibrated so that their minimum corrected divergence satisfies $\psi_{\text{corrected}} = \delta = 0.10$. The exact probability vectors and the calibration algorithm are provided in the script `boot_calibration.py`.

At the equivalence boundary $\psi_{\text{corrected}} = \delta$, the empirical type-I error rate is at most 0.025 across the three shapes considered, indicating conservative behavior relative to the nominal level 0.05 in these simulations. At $\psi_{\text{corrected}} = 0.15$, which lies within the null region, the procedure rarely declares consistency. For data generated from the power-law first-digit distribution, with $\psi_{\text{corrected}} \approx 0$, the procedure declares consistency in all replications, yielding empirical power equal to one under this simulation design. The boundary case $\widehat{k} = 1$ and the saturation case $\widehat{k} = K$ are also examined numerically, since the regular interior chi-squared asymptotics do not apply in these cases. The corresponding bootstrap upper bounds and empirical coverage results are reported in Table 3.

Table 3. Calibration of the bootstrap equivalence decision at $\delta = 0.10$, $\alpha = 0.05$, and $n = 2000$.

Case	Population	Rate
Level ($\psi_{\text{corrected}} = \delta$)	$k = 2,\ \varphi = 0.3$	0.025
	$k = 1.5,\ \varphi = 0.6$	0.013
	$k = 3,\ \varphi = 0$	0.018
Interior null ($\psi_{\text{corrected}} = 0.15$)	perturbed $k = 2$	0.000
Power ($\psi_{\text{corrected}} \approx 0$)	$k = 2,\ \varphi = 0.3$	1.000
	$k = 4,\ \varphi = 0.7$	1.000
Boundary	$k = 1,\ \varphi = 0.6$	1.000
Benford limit	Benford probabilities	1.000
Coverage of $u_{0.95}$	$k = 2,\ \varphi = 0.3$	1.000
	$k = 1.5,\ \varphi = 0.6$	1.000

5. Application to real data

5.1 *Decision rules*

The corrected index $\psi_{\text{corrected}}$ is read as an effect size, the chi-squared distance to the best-fitting power law after the exponent and scale phase are removed, whereas the formal verdict is the equivalence decision of Proposition 3.13 at margin $\delta = 0.10$ and significance level $\alpha = 0.05$. The data are consistent with the power-law first-digit distribution when the upper bootstrap bound $u_{0.95}$ on $\psi_{\text{corrected}}$ falls below δ, and anomalous otherwise. This equivalence reading is used because the corrected statistic, being a consistent goodness-of-fit statistic, rejects any large sample under a point null hypothesis. The thresholds of Table 4 are retained only as effect-size guides, and an anomalous label marks incompatibility with the fitted baseline, not misconduct.

The margin $\delta = 0.10$ is fixed a priori, as a moderate effect-size tolerance on the corrected index, and not tuned to the benchmark. This margin sits inside the borderline band $[0.05, 0.15)$ of the corrected-index guide of Table 4 and matches, on the traditional scale, the consistent-to-borderline cutoff of that same guide, a conventional tolerance chosen independently of the observed bootstrap bounds.

Once δ is fixed, the data leave a wide gap on either side. Across the authentic datasets, the upper bound $u_{0.95}$ never exceeds 0.068 (World Bank data), and it is stable under resampling. Over twelve independent bootstrap seeds, the bound stays in $[0.056, 0.068]$ for World Bank and $[0.058, 0.064]$ for the physical constants, the two authentic cases nearest the boundary, and neither verdict flips. The flagged cases sit far above, the retracted data at $u_{0.95} = 0.147$ and the bounded-support catalogue near 3. Therefore, the classification is the same for any margin in the interval $(0.068, 0.147)$, and so the verdict is insensitive to the exact value of δ. This insensitivity is observed after fixing δ and it does not justify the choice of the margin. Accordingly, the benchmark is read as calibrated exploratory evidence that the corrected index separates authentic from fabricated data generated from a power law, not as a confirmatory performance claim tuned to the data.

Table 4. Effect-size guides for the index and corrected index, where the corrected verdict is the equivalence decision of Proposition 3.13 at $\delta = 0.10$.

Test	Threshold	Effect-size reading
Traditional	$\psi < 0.10$	Consistent
	$0.10 \leq \psi < 0.30$	Borderline
	$\psi \geq 0.30$	Anomalous
Corrected	$\psi_{\text{corrected}} < 0.05$	Consistent
	$0.05 \leq \psi_{\text{corrected}} < 0.15$	Borderline
	$\psi_{\text{corrected}} \geq 0.15$	Anomalous

Table 5. Traditional and corrected first-digit diagnostics for simulated and public datasets.

Dataset	Source	n	ψ	$\widehat{k}$	$\widehat{\varphi}$	$\psi_{\text{corrected}}$	$u_{0.95}$	Verdict
Simulated power law $k = 2$	seeded	2000	0.107	1.91	0.99	0.002	0.013	Consistent
Simulated uniform	seeded	2000	0.391	1.69	0.66	0.245	0.279	Anomalous
Astronomical catalogue	HYG	109400	0.018	3.10	0.17	0.002	0.003	Consistent
Election returns	Dataverse	20272	0.001	31.19	0.95	0.000	0.001	Consistent
Fibonacci scaled	OEIS	500	0.000	K	—	0.000	0.022	Consistent
Financial statements	SEC EDGAR	3236586	0.002	15.01	0.65	0.001	0.001	Consistent
Physical constants	CODATA	355	0.036	4.89	0.93	0.021	0.063	Consistent
Pruitt (retracted)	Dryad	10269	0.373	1.38	0.75	0.130	0.147	Anomalous
Seismic catalogue	USGS	500	7.466	1.00	0.60	2.707	2.980	Anomalous
World Bank	World Bank	264	0.015	12.77	0.55	0.013	0.060	Consistent

Note: The entry $\widehat{k} = K$ denotes a fit at the numerical upper bound of the optimization range and not an estimate equal to infinity.

5.2 *Application to public datasets*

The corrected test is applied to datasets obtained from documented public queries, so that every value below is regenerated by the supplied scripts. Table 5 reports, for each dataset, the index ψ, jointly estimated exponent $\widehat{k}$ and scale phase $\widehat{\varphi}$, corrected index $\psi_{\text{corrected}}$, its upper bootstrap bound $u_{0.95}$, and equivalence verdict of Proposition 3.13 at $\delta = 0.10$. Two simulated baselines anchor the comparison: (i) authentic data generated from the power law with $k = 2$ give an index of 0.107 that the rule reads as borderline, while the corrected index falls to 0.002, which is a false positive and its remedy in one line; and (ii) fabricated uniform data stay far from any power-law baseline, at $\psi_{\text{corrected}} = 0.245$.

The equivalence decision separates the two cases cleanly. Every authentic dataset, spanning astronomy, economics, finance, mathematics, physics, and political science, is declared consistent with the power-law first-digit distribution. Its corrected index is small, from 0.000 to 0.021, and its upper bootstrap bound lies below the margin $\delta = 0.10$. The three benchmark cases declared incompatible with the fitted distribution are the fabricated uniformly distributed data, the retracted dataset presented in [17], with $\psi_{\text{corrected}} = 0.130$, and the seismic catalogue [18], with $\psi_{\text{corrected}} = 2.71$. The seismic catalogue has bounded support and therefore lies outside the intended domain of applicability discussed below. The retracted case has a corrected index inside the borderline band of the effect-size guide. However, its upper bound $u_{0.95} = 0.147$ exceeds δ, and the equivalence decision flags it, a reminder that the guide is descriptive and the verdict is the inferential decision. The equivalence reading is what makes this decision meaningful at large n. This is so because the corrected statistic is a consistent goodness-of-fit statistic and no empirical dataset is exactly a power law. A point-null p-value would reject even the county returns, the astronomical catalogue and the corporate statements, whose $n\psi_{\text{corrected}}$ runs into the hundreds or thousands despite a corrected index below 0.02. The equivalence test instead concludes that the first-digit distribution of these authentic data lies within δ of the fitted power-law first-digit family, which is the behavior intended for the method under the reported multinomial-bootstrap calibration, rather than a classification read from a fixed cutoff.

5.3 *Domain of applicability*

The method targets quantities whose magnitudes span several orders of magnitude, where the leading digit carries information. Two natural boundaries of this domain are worth stating, each pointing to a complementary tool. Bounded-magnitude data, such as a seismic catalogue whose magnitudes occupy a short range, are governed by that range, and a bounded-support diagnostic is the appropriate instrument and is a natural companion to the present test.

Fixed-scale survey responses on a Likert scale, such as the retracted moral-virtue data presented in [19], are single-digit by construction and are addressed by duplicate- and digit-pattern tests designed for that setting. Recognizing these boundaries sharpens the test by directing each type of data to the diagnostic appropriate for its data structure. Extending the empirical evaluation to further retracted datasets, such as the handshake data presented in [20], corrected in 2024 [21], is a natural continuation as their raw records become available.

6. Conclusions

The central contribution of this article is a corrected first-digit method based on the unit-scale power-law first-digit family. The closed-form distribution itself coincides with the generalized Benford law [8]. The method adds a false-positive result for the Pearson chi-squared test based on the Benford law under the specified first-digit model, joint minimum-chi-squared estimation of the exponent and scale phase, and an equivalence decision based on the first-digit distance. Under independent first digits generated from a finite-exponent member of the power law, the chi-squared statistic diverges linearly with the sample size. The corrected procedure instead fits $\mathbb{P}(d; k, \varphi)$ and evaluates compatibility with that fitted first-digit family. The simulation and application studies illustrate its behavior for the selected benchmarks, but compatibility of first-digit frequencies does not by itself establish a complete power-law mechanism or data authenticity. Under the specific two-component mixture considered in Theorem 3.14, the structure index satisfies $\psi_{\text{mix}} = \rho^2 \psi_{k,\varphi}$.

The proposed method is intended for positive quantities whose magnitudes span several orders of magnitude and whose first-digit probabilities can reasonably be compared with the fitted power-law first-digit family. Bounded-magnitude data and fixed-scale responses require diagnostics adapted to their supports and measurement structures. When the best-fitting member of the family is numerically close to the Benford limit, the corrected index approaches the traditional one. In general, however, data outside the power-law model may still have a finite best-fitting exponent. Therefore, the procedure must be interpreted as a diagnostic of first-digit compatibility rather than as a test of the complete distributional mechanism or of authenticity.

A natural extension is a joint diagnostic combining the first-digit correction with second-digit and bounded-support information. The method could also be extended to the Birnbaum–Saunders family [22, 23]. In particular, future work may characterize the first-digit probabilities induced by its shape and scale parameters and investigate whether parameter-adjusted first-digit tests are required for that family.

Appendix: proofs

Proof of Lemma 3.2 Let $L = 10^M$ and define the CDF of the fractional parts, $\text{frac}(\log_{10}(r))$ say, for $r \in \{1, \dots, L\}$, by

$$F(t; M) = \frac{1}{10^M} \sum_{r=1}^{10^M} \mathbb{1}\big(\text{frac}(\log_{10}(r)) \leq t\big), \quad t \in [0, 1),$$

where $\mathbb{1}(B)$ is the indicator function of the set B. Partition $\{1, \dots, 10^M\}$ into the orders of magnitude $[10^j, 10^{j+1})$, for $j \in \{0, \dots, M-1\}$. Within the jth order of magnitude, the condition $\text{frac}(\log_{10}(r)) \leq t$ is equivalent to $10^j \leq r \leq 10^{j+t}$.

Consequently, we have that

$$F(t; M) = \frac{1}{10^M}\left(1 + \sum_{j=0}^{M-1}\left(\lfloor 10^{j+t}\rfloor - 10^j + 1\right)\right),$$

where the initial term one accounts for the endpoint $r = 10^M$. Since the effects of the floor function and the endpoints vanish as $M \to \infty$, it follows that

$$\lim_{M\to\infty} F(t; M) = \lim_{M\to\infty} \frac{1}{10^M}\sum_{j=0}^{M-1}\left(10^{j+t} - 10^j\right) = \frac{10^t - 1}{9}.$$

Therefore, the limiting CDF is stated as

$$F(t) = \frac{10^t - 1}{9}, \quad t \in [0, 1),$$

whose PDF is given by

$$f(t) = \frac{\mathrm{d}}{\mathrm{d}t}F(t) = \frac{\ln(10)}{9}10^t, \quad t \in [0, 1).$$

□

Proof of Theorem 3.3 Let $W = \log_{10}(Y)$. The change of variables sends the PDF to one proportional to $10^{-(1+1/k)W}\, 10^W = 10^{-W/k}$ on $[A, B]$. Let $U = \mathrm{frac}(W)$. The first digit of Y equals d when $U \in [\log_{10}(d), \log_{10}(d + 1))$. Equivalently, defining $S = 1 - U$, this event is, up to endpoints of probability zero, $S \in [1 - \log_{10}(d + 1),\, 1 - \log_{10}(d))$. Because A and $B - A$ are integers, partition $[A, B]$ into the unit orders of magnitude $[A + j, A + j + 1)$, for $j \in \{0, \ldots, B - A - 1\}$. On order of magnitude j, the PDF carries the factor $10^{-(A+j)/k}$, constant in the fractional coordinate. Summing over j, this factor multiplies the mass on every cell and the total mass alike, and so it cancels on normalization, independently of A and B. Therefore, the reflected fractional part S has a PDF formulated as

$$g(s; k) = \frac{\ln(10)}{k}\,\frac{10^{s/k}}{(10^{1/k} - 1)}, \quad s \in [0, 1). \tag{11}$$

Integrating g over the cell of digit d gives the expression stated as

$$\mathbb{P}(d; k) = \frac{10^{1/k}}{10^{1/k} - 1}\left(d^{-1/k} - (d + 1)^{-1/k}\right).$$

A telescoping sum confirms that the nine probabilities add to one. □

Proof of Proposition 3.4 Let $M = \mathrm{frac}(\log_{10}(Y_0))$. Under the model of Theorem 3.3, the CDF of M on $[0, 1)$ is given by $G(y; k) = (10^{-y/k} - 1)/(10^{-1/k} - 1)$. Multiplication by C adds $\log_{10}(C)$ to the logarithm. Therefore, the logarithmic fractional part of $Y = C\,Y_0$ is $M_\varphi = \mathrm{frac}(M + \varphi)$ and $\varphi = \mathrm{frac}(\log_{10}(C))$. The first digit of Y equals d when $M_\varphi \in [\log_{10}(d), \log_{10}(d + 1))$, or, equivalently, when M lies in the corresponding arc $[\log_{10}(d) - \varphi,\, \log_{10}(d + 1) - \varphi)$ interpreted modulo one. Measuring this arc through G_{ext} gives the expression stated in (7). Summing over $d \in \{1, \ldots, 9\}$ yields total probability one, and setting $\varphi = 0$ recovers $\mathbb{P}(d; k)$. □

Proof of Corollary 3.5 Set $\varepsilon = 1/k \to 0^+$. A Taylor expansion gives $d^{-\varepsilon} - (d+1)^{-\varepsilon} = \varepsilon \ln(1+1/d) + O(\varepsilon^2)$, while $10^\varepsilon/(10^\varepsilon - 1) \sim 1/(\varepsilon \ln(10))$, where here $\sim$ denotes asymptotic equivalence as $1/k \to 0$. Therefore, the product tends to $\log_{10}(1 + 1/d)$. □

Proof of Proposition 3.6 Write $\theta = \ln(10)/k > 0$. Then, the PDF stated in (11) can be written as

$$f(s;\theta) = \frac{\theta \exp(\theta s)}{\exp(\theta) - 1}, \qquad s \in [0, 1).$$

The digit cells are $[a_d, b_d) = [1 - \log_{10}(d+1),\ 1 - \log_{10}(d))$, and their Benford probabilities are $w_d = b_d - a_d = \mathbb{P}(d)$. Define $v_d(\theta) = \mathbb{P}(d;k)/w_d$. Then, $\sum_{d=1}^{9} w_d v_d = 1$.

Let D_B be an auxiliary random digit satisfying $\mathbb{P}(D_B = d) = w_d$, and define $V_B = v_{D_B}$. Since $\mathbb{E}(V_B) = 1$, we have

$$\psi(k) = \sum_{d=1}^{9} w_d (v_d - 1)^2 = \mathrm{Var}(V_B).$$

Furthermore, $\partial_\theta f(s;\theta) = f(s;\theta)(s - \mu_\theta)$ and $\mu_\theta = \int_0^1 s f(s;\theta)\,\mathrm{d}s$. Consequently, $v_d'(\theta) = v_d(\theta)(\bar{s}_d - \mu_\theta)$, where $\bar{s}_d = \int_{a_d}^{b_d} s f(s;\theta)\,\mathrm{d}s / \int_{a_d}^{b_d} f(s;\theta)\,\mathrm{d}s$.

Let D_k be an auxiliary random digit satisfying $\mathbb{P}(D_k = d) = \mathbb{P}(d;k)$ and define $V_k = v_{D_k}$, $S_k = \bar{s}_{D_k}$. Because

$$\mathbb{E}(S_k) = \sum_{d=1}^{9} \mathbb{P}(d;k)\bar{s}_d = \mu_\theta,$$

differentiation gives

$$\frac{\mathrm{d}\psi(k)}{\mathrm{d}\theta} = 2\ \mathrm{Cov}(V_k, S_k).$$

Both v_d and $\bar{s}_d$ are strictly ordered in the same direction across the digit cells. Therefore, the weighted Chebyshev sum inequality gives $\mathrm{Cov}(V_k, S_k) > 0$. Thus, $\mathrm{d}\psi(k)/\mathrm{d}\theta > 0$. Since $\theta = \ln(10)/k$ is strictly decreasing in k, it follows that $\psi(k)$ is strictly decreasing in k. □

Proof of Theorem 3.7 Let O_d denote the observed count of digit d. By the law of large numbers, $O_d/n \xrightarrow{\mathbb{P}} \mathbb{P}(d;k)$, for $d \in \{1, \ldots, 9\}$. Therefore,

$$\frac{\chi^2}{n} = \sum_{d=1}^{9} \frac{(O_d/n - \mathbb{P}(d))^2}{\mathbb{P}(d)} \xrightarrow{\mathbb{P}} \sum_{d=1}^{9} \frac{(\mathbb{P}(d;k) - \mathbb{P}(d))^2}{\mathbb{P}(d)} = \psi(k).$$

For every finite k, Proposition 3.6 and Corollary 3.5 imply that $\psi(k) > 0$. Hence, χ^2 diverges linearly in n, and its rejection probability tends to one for any fixed critical value. □

Proof of Corollary 3.8 The statement is immediate from the formulation presented in (8), with $\psi(k)$ defined in Proposition 3.6. As an instance, $n_{\min}(2) = \lceil 15.51/0.101 \rceil = 154$. □

Proof of Proposition 3.9 Under the multinomial model stated in the proposition, the law of large numbers gives $O_d/n \xrightarrow{\mathbb{P}} \mathbb{P}(d;k,\varphi)$, for $d \in \{1,\ldots,9\}$. When (k,φ) is known, the standard Pearson chi-squared theorem yields $\chi^2_{\text{corrected}} \xrightarrow{\mathbb{D}} \chi^2(8)$. When both parameters are estimated from the same counts by minimum chi-squared, the regular theory for grouped data reduces the degrees of freedom by two. Under the interiority, local-identification, full-rank, and positivity conditions stated in the proposition, $\chi^2_{\text{corrected}} \xrightarrow{\mathbb{D}} \chi^2(6)$. When $\varphi = \varphi_0$ is fixed and only k is estimated, the corresponding regular limit is $\chi^2(7)$. □

Proof of Proposition 3.10 Suppose $\mathbb{P}(d;k) = \mathbb{P}(d;k')$. The function $h(s) = g(s;k) - g(s;k')$ integrates to zero over each of the nine digit intervals. When $k \neq k'$, the two exponentials in h have distinct growth rates. Hence, the ratio $g(s,k)/g(s,k')$ is monotone, and h changes sign at most once on $[0,1)$. At most one digit interval can straddle that sign change. On every other interval, h keeps a constant sign, and a vanishing integral then forces $h = 0$ there. Because h is analytic, vanishing on an open interval forces $h = 0$ throughout, which contradicts $k \neq k'$. □

Proof of Theorem 3.11 By Proposition 3.10, $k \neq k'$ implies that $\mathbb{P}(d;k) \neq \mathbb{P}(d;k')$, and hence $\Delta(k,k') > 0$. The law of large numbers gives $\chi^2(k)n \xrightarrow{\mathbb{P}} \Delta(k,k')$. Therefore, the statistic grows linearly in n, and the rejection probability tends to one. □

Proof of Theorem 3.12 By the law of large numbers, $P_n(d) \xrightarrow{\mathbb{P}} \mathbb{P}(d;k_0,\varphi_0)$, for $d \in \{1,\ldots,9\}$. Because the number of cells is finite and the cell probabilities are continuous and strictly positive on the compact parameter space $[1,K] \times \mathbb{T}$, the sample minimum-chi-squared criterion converges uniformly in probability to its counterpart. Under the unique-minimizer assumption stated in the theorem, the criterion has the unique minimizer (k_0,φ_0). The argmin consistency theorem therefore gives $(\widehat{k}_n,\widehat{\varphi}_n) \xrightarrow{\mathbb{P}} (k_0,\varphi_0)$. As k increases, the family approaches the Benford limit and the phase becomes weakly identified. Fits attaining $k = K$ are consequently treated as upper-bound fits. □

Proof of Proposition 3.13 Conditional on the validity of the one-sided bootstrap upper bound, the event $u_{1-\alpha} < \delta$ places the entire one-sided confidence region below the equivalence margin and therefore makes the decision to reject H_0: $\psi_\star \geq \delta$. □

Proof of Theorem 3.14 From the definition of the mixture, $P_{\text{mix}}(d) - \mathbb{P}(d) = \rho(\mathbb{P}(d;k,\varphi) - \mathbb{P}(d))$. Substitution into the definition of the structure index gives $\psi_{\text{mix}} = \rho^2\psi_{k,\varphi}$. The probability limit follows from the law of large numbers. For the expectation, the multinomial marginal variance is $\mathrm{Var}(O_d) = nP_{\text{mix}}(d)\,(1 - P_{\text{mix}}(d))$. Substitution into the Pearson statistic yields the stated expectation. When $\rho = 0$, $P_{\text{mix}}(d) = \mathbb{P}(d)$ and the bounded term reduces to $\sum_{d=1}^{9}(1 - \mathbb{P}(d)) = 8$. □

Proof of Corollary 3.15 By Theorem 3.14, $\chi^2_n/n \xrightarrow{\mathbb{P}} \rho^2\psi_{k,\varphi}$. Therefore, $\widehat{\rho}^2 = (\chi^2_n/n)/(\psi_{k,\varphi}) \xrightarrow{\mathbb{P}} \rho^2$. The event $\chi^2_n < n\psi_{k,\varphi}/4$ is equivalent to $\widehat{\rho}^2 < 1/4$. Its probability consequently tends to one when $\rho < 1/2$ and to zero when $\rho > 1/2$. □

Statements

Acknowledgments

The authors thank the reviewers for their valuable comments, which helped improve the presentation and content of this article, as well as the maintainers of the public data repositories used in our validation studies.

Author contributions

Conceptualization: F. Bustillos, V. Leiva. Data curation: F. Bustillos. Formal analysis: F. Bustillos, V. Leiva. Investigation: F. Bustillos, V. Leiva. Methodology: F. Bustillos, V. Leiva. Software: F. Bustillos. Supervision: V. Leiva. Validation: F. Bustillos, V. Leiva. Visualization: F. Bustillos, V. Leiva. Writing of the original draft: F. Bustillos. Writing, reviewing, and editing: V. Leiva. All authors have read and agreed to the published version of the article.

Conflicts of interest

The authors declare no conflict of interest.

Data and code availability

All code, documented queries, cleaning rules, data dictionary, and archive checksums are openly available at github.com/bustillosmorales-lgtm/benford-powerlaw-firstdigit. The scripts `benford_powerlaw.py`, `joint_phi_tost.py`, `recompute_table.py` and `fetch_real_data.py` reproduce the analytical values, the reference sizes $n_{\min}(k)$, the contamination scaling, the seeded simulated baselines, the Fibonacci case [24], and the bootstrap calibration of Table 3, under fixed random seeds. The numerical evaluations of the closed-form expressions were checked to machine precision by the script `verification.py`. The script `identif_proof.py` provides numerical diagnostics for the joint parameterization, including evaluations of the Jacobian rank and searches for distinct parameter pairs producing the same first-digit probabilities over the investigated grid. These computations support, but do not prove, the unique-minimizer condition used in Theorem 3.12. The level, power and coverage of the equivalence decision reported in Table 3 are produced by `boot_calibration.py`. The public sources, each queried as documented in the repository, are the USGS ComCat archive [18], NIST CODATA 2022 set [25], MIT Election Data and Science Lab [26], SEC EDGAR financial statement data sets [27], HYG stellar database [28], World Bank development indicators [29], and Dryad deposit of [17]. Because several of these archives are maintained by third parties and are versioned or updated over time, the repository records the retrieval date and an SHA-256 checksum for each downloaded archive, so that the exact snapshots underlying Table 5 can be identified even when a live query returns a newer version. The retracted cases are the Gino moral-virtue data [19] and handshake data presented in [20], corrected in 2024 [21].

Declaration on the use of artificial intelligence (AI) technologies

The authors declare that no generative AI was used in the technical contents of this article.

Funding

This work received no external funding.

Open access statement

References

[1] Nigrini, M.J., 2012. Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection. Wiley, Hoboken, NJ, USA. DOI: 10.1002/9781119203094.

[2] Benford, F., 1938. The law of anomalous numbers. Proceedings of the American Philosophical Society, 78, 551–572. URL: www.jstor.org/stable/984802.

[3] Hill, T.P., 1995. A statistical derivation of the significant-digit law. Statistical Science, 10, 354–363. DOI: 10.1214/ss/1177009869.

[4] Nigrini, M.J. and Mittermaier, L.J., 1997. The use of Benford's law as an aid in analytical procedures. Auditing: A Journal of Practice and Theory, 16, 52–67. URL: www.econbiz.de/10006996665.

[5] Durtschi, C., Hillison, W.A., and Pacini, C.J., 2004. The effective use of Benford's law to assist in detecting fraud in accounting data. Journal of Forensic Accounting, 5, 17–34. URL: digitalcommons.usf.edu/fac_publications/939.

[6] Pericchi, L.R. and Torres, D., 2011. Quick anomaly detection by the Newcomb–Benford law, with applications to electoral processes data from the USA, Puerto Rico and Venezuela. Statistical Science, 26, 502–516. DOI: 10.1214/09-STS296.

[7] Deckert, J., Myagkov, M., and Ordeshook, P.C., 2011. Benford's law and the detection of election fraud. Political Analysis, 19, 245–268. DOI: 10.1093/pan/mpr014.

[8] Pietronero, L., Tosatti, E., Tosatti, V., and Vespignani, A., 2001. Explaining the uneven distribution of numbers in nature: the laws of Benford and Zipf. Physica A, 293, 297–304. DOI: 10.1016/S0378-4371(00)00633-6.

[9] Berger, A. and Hill, T.P., 2015. An Introduction to Benford's Law. Princeton University Press, Princeton, NJ, USA. DOI: 10.1515/9781400866588.

[10] Coulibaly, B.D., Chaibi, G., and El Khomssi, M., 2024. Parameter estimation of the alpha-stable distribution and applications to financial data. Chilean Journal of Statistics, 15, 60–80. DOI: 10.32372/ChJS.15-01-04.

[11] Leiva, V., Marchant, C., Ruggeri, F., and Saulo, H., 2015. A criterion for environmental assessment using Birnbaum–Saunders attribute control charts. Environmetrics, 26, 463–476. DOI: 10.1002/env.2349.

[12] Leão, J., Leiva, V., Saulo, H., and Tomazella, V., 2018. Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. Statistics in Medicine, 37, 4421–4440. DOI: 10.1002/sim.7929.

[13] Luque, B. and Lacasa, L., 2009. The first-digit frequencies of prime numbers and Riemann zeta zeros. Proceedings of the Royal Society A, 465, 2197–2216. DOI: 10.1098/rspa.2009.0126.

[14] Barabesi, L. and Pratelli, L., 2020. On the generalized Benford law. Statistics and Probability Letters, 160, 108702. DOI: 10.1016/j.spl.2020.108702.

[15] Fisher, R.A., 1924. The conditions under which chi-squared measures the discrepancy between observation and hypothesis. Journal of the Royal Statistical Society, 87, 442–450. URL: www.jstor.org/stable/2341149.

[16] Cramér, H., 1946. Mathematical Methods of Statistics. Princeton University Press, Princeton, NJ, USA. DOI: 10.1515/9781400883868.

[17] Laskowski, K.L., Montiglio, P.O., and Pruitt, J.N., 2016. Individual and group performance suffers from social niche disruption. The American Naturalist, 187, 776–785. DOI: 10.1086/686220. Data DOI: 10.5061/dryad.33f0n. Retraction DOI: 10.1086/708066.

[18] U.S. Geological Survey, 2024. Advanced National Seismic System Comprehensive Catalog (ComCat). URL: earthquake.usgs.gov/data/comcat.

[19] Gino, F., Kouchaki, M., and Galinsky, A.D., 2015. The moral virtue of authenticity: How inauthenticity produces feelings of immorality and impurity. Psychological Science, 26, 983–996. DOI: 10.1177/0956797615575277. Retraction DOI: 10.1177/09567976231187596.

[20] Schroeder, J., Risen, J.L., Gino, F., and Norton, M.I., 2019. Handshaking promotes deal-making by signaling cooperative intent. Journal of Personality and Social Psychology, 116, 743–768. DOI: 10.1037/pspi0000157.

[21] Schroeder, J., Risen, J.L., Gino, F., and Norton, M.I., 2024. Correction to "Handshaking promotes deal-making by signaling cooperative intent". Journal of Personality and Social Psychology, 126, 212. DOI: 10.1037/pspi0000460.

[22] Santana, L., Vilca, F., and Leiva, V., 2011. Influence analysis in skew-Birnbaum–Saunders regression models and applications. Journal of Applied Statistics, 38, 1633–1649. DOI: 10.1080/02664763.2010.515679.

[23] Saulo, H., Leão, J., Leiva, V., and Aykroyd, R.G., 2019. Birnbaum–Saunders autoregressive conditional duration models applied to high-frequency financial data. Statistical Papers, 60, 1605–1629. DOI: 10.1007/s00362-017-0888-6.
[24] OEIS Foundation, 2024. Sequence A000045 (Fibonacci numbers). The On-Line Encyclopedia of Integer Sequences. URL: oeis.org/A000045.
[25] Mohr, P.J., Newell, D.B., Taylor, B.N., and Tiesinga, E., 2025. CODATA recommended values of the fundamental physical constants: 2022. Reviews of Modern Physics, 97, 025002. DOI: 10.1103/RevModPhys.97.025002.
[26] MIT Election Data and Science Lab, 2018. County Presidential Election Returns 2000–2020. Harvard Dataverse, V13. DOI: 10.7910/DVN/VOQCHQ.
[27] U.S. Securities and Exchange Commission, 2026. Financial Statement Data Sets, 2026 first quarter. URL: www.sec.gov/data-research/sec-markets-data/financial-statement-data-sets.
[28] Astronomy Nexus, 2024. HYG Stellar Database, version 4.1. URL: github.com/astronexus/HYG-Database/tree/main/hyg.
[29] World Bank, 2024. World Development Indicators, series SP.POP.TOTL (total population), 2022. URL: data.worldbank.org/indicator/SP.POP.TOTL.