Probability in the Engineering and Informational Sciences (2024), **38:**3 594–616
doi:10.1017/S0269964823000207

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

General distributions of number representation elements

Félix Balado and Guénolé C. M. Silvestre

School of Computer Science, University College Dublin, Dublin, Ireland
Corresponding author: F. Balado; Email: felix@ucd.ie

Keywords: Benford's law; continued fraction coefficients models; Pareto distribution; significant digits models

Abstract
We provide general expressions for the joint distributions of the k most significant b-ary digits and of the k leading continued fraction (CF) coefficients of outcomes of arbitrary continuous random variables. Our analysis highlights the connections between the two problems. In particular, we give the general convergence law of the distribution of the jth significant digit, which is the counterpart of the general convergence law of the distribution of the jth CF coefficient (Gauss-Kuz'min law). We also particularise our general results for Benford and Pareto random variables. The former particularisation allows us to show the central role played by Benford variables in the asymptotics of the general expressions, among several other results, including the analogue of Benford's law for CFs. The particularisation for Pareto variables—which include Benford variables as a special case—is especially relevant in the context of pervasive scale-invariant phenomena, where Pareto variables occur much more frequently than Benford variables. This suggests that the Pareto expressions that we produce have wider applicability than their Benford counterparts in modelling most significant digits and leading CF coefficients of real data. Our results may find practical application in all areas where Benford's law has been previously used.

1. Introduction

Benford's law [7]—originally enunciated by Newcomb [24] after spotting a pattern of use in logarithmic tables—states that the most significant digits in a set of numbers written in decimal form often follow the distribution:

$$\Pr(A = a) = \log_{10}\left(1 + \frac{1}{a}\right), \tag{1.1}$$

where A is the discrete random variable that models the first non-null decimal digit of a number, and $a \in \{1, 2, \ldots, 9\}$, i.e., its possible values. At first sight, the uniform distribution might appear to be more reasonable than (1.1). However Benford demonstrated that different sets of numbers originating in miscellaneous sources conform quite accurately to (1.1). The discovery of (1.1) took place with the understanding that it could be extended to further include the second, third, etc., most significant digits, and it was also clear from the start that essentially the same law could model significant digits of numbers written in any base—not only base 10.

Research into Benford's law and its applications has experienced a significant boost in recent years—the online bibliography on the subject set up by Berger *et al.* [10] is witness to this. An important factor behind this boost is the fact that Benford's probabilistic model of significant digits has found practical use in areas such as forensic analysis, data quality and consistency analysis, generation of synthetic data, floating point operations error analysis, etc., stemming from fields such as accountancy, economics, signal processing, computer science, natural sciences, and beyond (see [21]). A second—but more

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

interesting—factor behind the aforementioned boost is the desire to explain the recurrent emergence of Benford's law [9, 16], which has led to the realisation that, while relatively common, it does not apply in numerous circumstances. Therefore it would be helpful to have models more general than (1.1) and its immediate extensions. This would widen the application range of significant digits distributions, and perhaps also help to shed light on the origins of Benford's law itself. In particular, even though scale invariance has been cited as an underlying reason for the emergence of Benford's law ([30]), the fact is that this law often does not apply to scale invariant phenomena [15, 16]—noticeably, to outcomes of the ubiquitous Pareto distribution. Lastly, another motivation for further research is the fact that the positional numeral systems behind significant digits models such as (1.1) have an alternative: since real numbers can also be represented using continued fraction (CF) expansions, leading CF coefficients can play a role analogous to that of significant digits in the aforementioned applications. For example, as $0.1875_{10} = 0.14_8 = 1/(5+1/3)$, the two leading CF coefficients of this number (5 and 3) could replace its two most significant decimal digits (1 and 8), or its two most significant octal digits (1 and 4), in the aforementioned practical applications.

Our main goal in this paper is to provide a general treatment of the related problems of modelling probabilistically the most significant digits and the leading CF coefficients of outcomes of arbitrary continuous random variables, and to evince the parallelisms between the two problems. Ever since the observations made by Newcomb [24] and Benford [7], studies of the distribution of most significant digits have largely focused on Benford variables (i.e., on situations in which Benford's law holds)—for an overview, see the introduction by Berger and Hill [8] and the book by Miller [21]. Some generalisations for non-Benford scenarios have been pursued by Pietronero *et al.* [29] and by Barabesi and Pratelli [6], among others [13, 15, 28, 33], but a truly general approach to modelling significant digits has never been presented. A lot less attention has been devoted to modelling leading CF coefficients. Except for an approximation due to Blachman [11], most of the work in this area has been pioneered by Miller and Takloo-Bighash [22]. In any case, all existing results for leading CF coefficients models are solely for the particular case in which the fractional part of the data represented by means of CFs is uniformly distributed. No general approach has been investigated in this problem either.

This paper is organised as follows. In Section 2, we give the general expression for the joint distribution of the k most significant b-ary digits of outcomes drawn from an arbitrary positive-valued distribution. One application of our analysis is a proof of the general asymptotic distribution of the jth most significant b-ary digit, which, as we will discuss, is the near exact counterpart of the Gauss-Kuz'min law [20] for the general asymptotic distribution of the jth continued fraction coefficient. Our approach to modelling the k most significant b-ary digits through a single variable—rather than through k separate variables as in previous works—leads to further contributions[1] in the particularisation of our general results in Section 4. Therein, we produce a new closed-form expression for the distribution of the jth significant b-ary digit of a Benford variable, and we give a short new proof of the asymptotic sum-invariance property of these variables. We also show that Benford's distribution is just a particular case of a more general distribution based on Pareto variables, which must have wider applicability in the pervasive realm of scale-invariant data—a fact first pointed out by Pietronero *et al.* [29] and then expanded upon by Barabesi and Pratelli [6], who however only gave the model for a special case.

In Section 3, we give the general expression for the joint distribution of the k leading CF coefficients of outcomes drawn from an arbitrary distribution. This is shown to be explicitly analogous to modelling the k most significant b-ary digits of the data when the CF coefficients correspond to the logarithm base b of the data. It follows that modelling leading CF coefficients is a realistic practical alternative to modelling most significant digits. Most of the results in Section 3 are novel, and so are their particularisations in Section 4—in particular the analogue of Benford's law for continued fractions—except for the special cases previously given by Miller and Takloo-Bighash [22].

Additionally, we show in Section 4.1.3, the central role played by the particular results for Benford variables in the asymptotics of the general expressions—both when modelling significant digits

[1] Some preliminary results in this paper previously appeared in [4].

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

and leading CF coefficients—and we demonstrate this in the particular case of Pareto variables in Section 4.2.3.

Finally, we empirically verify all our theoretical results in Section 5, using both Monte Carlo experiments and real datasets.

Notation and preliminaries. Calligraphic letters are sets, and $|\mathcal{V}|$ is the cardinality of set $\mathcal{V}$. Boldface Roman letters are row vectors. Random variables (r.v.'s) are denoted by capital Roman letters, or by functions of these. The cumulative distribution function (cdf) of r.v. Z is $F_Z(z) = \Pr(Z \leq z)$, where $z \in \mathbb{R}$. The expectation of Z is denoted by $\mathrm{E}(Z)$. If Z is continuous with support $\mathcal{Z}$, its probability density function (pdf) is $f_Z(z)$, where $z \in \mathcal{Z}$. A r.v. Z which is uniformly distributed between a and b is denoted by $Z \sim U(a, b)$. The probability mass function (pmf) of a discrete r.v. Z with support $\mathcal{Z}$ is denoted by $\Pr(Z = z)$, where $z \in \mathcal{Z}$. The unit-step function is defined as $u(z) = 1$ if $z \geq 0$, and $u(z) = 0$ otherwise. The fractional part of $z \in \mathbb{R}$ is $\{z\} = z - \lfloor z \rfloor$. Curly braces are also used to list the elements of a discrete set, and the meaning of $\{z\}$ (i.e., either a fractional part or a one-element set) is clear from the context. We exclude zero from the set of natural numbers $\mathbb{N}$. We use Knuth's notation for the rising factorial powers of $z \in \mathbb{R}$, i.e., $z^{\overline{m}} = \Pi_{i=0}^{m-1}(z+i) = \Gamma(z+m)/\Gamma(z)$ [14].

Throughout the manuscript, X denotes a positive continuous r.v. We also define the associated r.v.:

$$Y = \log_b X,$$

for an arbitrary $b \in \mathbb{N}\backslash\{1\}$. The fractional part of Y, i.e., $\{Y\}$, will play a particularly relevant role in our analysis. The cdf of $\{Y\}$ is obtained from the cdf of Y as follows:

$$F_{\{Y\}}(y) = \sum_{i \in \mathbb{Z}} F_Y(y+i) - F_Y(i), \tag{1.2}$$

for $y \in [0, 1)$. Because $\{Y\}$ is a fractional part, it always holds that $F_{\{Y\}}(y) = 0$ for $y \leq 0$ and $F_{\{Y\}}(y) = 1$ for $y \geq 1$. Also, $F_{\{Y\}}(y)$ is a continuous function of y because X is a continuous r.v. and because cdf's are bounded.

2. General probability distribution of the k most significant b-ary digits

In this section, we will obtain the general expression for the joint probability distribution of the k most significant digits of a positive real number written in a positional base b numeral system, where $b \in \mathbb{N}\backslash\{1\}$. Let us first define,

$$\mathcal{A} = \{0, 1, \ldots, b-1\}.$$

The b-ary representation of $x \in \mathbb{R}^+$ is formed by the unique digits $a_i \in \mathcal{A}$ such that $x = \sum_{i \in \mathbb{Z}} a_i b^i$—unicity requires ruling out representations where there exists j such that $a_i = b-1$ for all $i < j$. If we now let $n = \lfloor \log_b x \rfloor$, the most significant b-ary digit of x is $a_n \in \mathcal{A}\backslash\{0\}$. This is because from the definition of n we have $n \leq \log_b x < n+1$, or, equivalently, $b^n \leq x < b^{n+1}$, which implies $a_n > 0$ and $a_j = 0$ for $j > n$. Using n, the k most significant b-ary digits of x can be jointly inferred as follows:

$$a = \lfloor x\, b^{-n+k-1} \rfloor = \lfloor b^{\{\log_b x\}+k-1} \rfloor. \tag{2.1}$$

By using $0 \leq \{\log_b x\} < 1$ in (2.1) we can verify that a belongs to the following set of integers:

$$\mathcal{A}_{(k)} = \{b^{k-1}, \ldots, b^k - 1\}, \tag{2.2}$$

whose cardinality is $|\mathcal{A}_{(k)}| = b^k - b^{k-1}$. We propose to call a in (2.1) the kth *integer significand* of x. Coining a new term is needed for clarity, as for some authors the significand of x is the integer

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

$\lfloor x\, b^{-n} \rfloor = a_n$ or even $\lfloor x\, b^{-n+k-1} \rfloor$ itself [26, pages 7 and 321], but for some others the significand of x is the real $x\, b^{-n} \in [1, b)$ [8, 21]—which is sometimes also called the normalised significand. In any case, the advantages of consistently working with the kth integer significand will become clear throughout this paper. To give an example of (2.1) and (2.2), say that $b = 10$ and $x = 0.00456678$. In this case $n = \lfloor \log_{10} x \rfloor = -3$, and so, for instance, the second integer significand of x is $a = \lfloor x\, 10^4 \rfloor = \lfloor 10^{1.65961} \rfloor = 45 \in \mathcal{A}_{(2)} = \{10, 11, 12, \ldots, 98, 99\}$.

Theorem 2.1. (General distribution of the k most significant b-ary digits). *If $A_{(k)}$ denotes the discrete r.v. that models the k most significant b-ary digits (i.e., the kth integer significand) of a positive continuous r.v. X, then*

$$\Pr(A_{(k)} = a) = F_{\{Y\}}\big(\log_b(a+1) - k + 1\big) - F_{\{Y\}}\big(\log_b a - k + 1\big), \tag{2.3}$$

where $a \in \mathcal{A}_{(k)}$ and $Y = \log_b X$.

Proof. Seeing (2.1), the r.v. we are interested in is defined as:

$$A_{(k)} = \lfloor b^{\{\log_b X\}+k-1} \rfloor. \tag{2.4}$$

From this definition, $A_{(k)} = a$ when $a \le b^{\{\log_b X\}+k-1} < a + 1$, or, equivalently, when

$$\log_b a - k + 1 \le \{\log_b X\} < \log_b(a+1) - k + 1. \tag{2.5}$$

Using (2.5) and the cdf of $\{Y\} = \{\log_b X\}$ we get (2.3). □

Remark 2.2. It is straightforward to verify that the pmf (2.3) adds up to one over its support, as:

$$\sum_{a \in \mathcal{A}_{(k)}} \Pr(A_{(k)} = a) = F_{\{Y\}}(1) - F_{\{Y\}}(0), \tag{2.6}$$

due to the cancellation of all consecutive summands in the telescoping sum on the LHS of (2.6) except for the two shown on the RHS. Because $F_{\{Y\}}(y)$ is the cdf of a r.v. with support $[0, 1)$, the RHS of (2.6) must equal one.

2.1. Distribution of the jth most significant b-ary digit

Next, let us denote by $A_{[j]}$ the r.v. that models the jth most significant b-ary digit of X. This variable can be obtained from the variable $A_{(j)}$ that models the jth integer significand as follows:

$$A_{[j]} = A_{(j)} \pmod b. \tag{2.7}$$

Obviously, $A_{[1]} = A_{(1)}$. From (2.7), the pmf of $A_{[j]}$ for $j \ge 2$ is:

$$\Pr(A_{[j]} = a) = \sum_{r \in \mathcal{A}_{(j-1)}} \Pr(A_{(j)} = rb + a), \tag{2.8}$$

where $a \in \mathcal{A}$.

Remark 2.3. Observe that (2.3) is also the joint pmf of $A_{[1]}, \ldots, A_{[k]}$. To see this we just have to write $a = \sum_{j=1}^{k} a_j b^{-j+k}$, which implies that $\Pr(A_{[1]} = a_1, \ldots, A_{[k]} = a_k) = \Pr(A_{(k)} = a)$. Under

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

this view, (2.8) is simply a marginalisation of (2.3). However, the derivation of (2.3) is simpler using the kth integer significand variable $A_{(k)}$ than using $A_{[1]}, \ldots, A_{[k]}$. Further examples of the advantages of working with kth integer significands are the following theorem and the results in Sections 4.1.1 and 4.2.1.

Theorem 2.4. (General asymptotic distribution of the jth most significant b-ary digit). *For any positive continuous random variable X, it holds that:*

$$\lim_{j\to\infty} \Pr(A_{[j]} = a) = b^{-1}. \tag{2.9}$$

Proof. For any $\epsilon > 0$ there exists $j_{\min}$ such that for all $j \geq j_{\min}$

$$\log_b(rb + a) - \log_b(rb) < \epsilon, \tag{2.10}$$

for all $a \in \mathcal{A}$ and $r \in \mathcal{A}_{(j-1)}$. Specifically, $j_{\min} = \lceil 1 + \log_b(\frac{b-1}{b^\epsilon - 1})\rceil + 1$. Therefore, inequality (2.10) and the continuity of $F_{\{Y\}}(y)$ in (2.3) imply that $\lim_{j\to\infty} \Pr(A_{(j)} = rb + a) - \Pr(A_{(j)} = rb) = 0$ for all $a \in \mathcal{A}$. Thus, from (2.8), we have that (2.9) holds. □

Remark 2.5. Theorem 2.4 tells us that the common intuition about digits being uniformly distributed is correct, but only for digits that lie away from the most significant one. In general, the larger b is, the faster the convergence of $A_{[j]}$ to a uniform discrete r.v. This is because $j_{\min}$ is nonincreasing on b for a given ϵ. Informally, the theorem can be argued as follows: since $rb \geq b^{j-1}$ in (2.8), for large j we have that $rb \gg a$, and therefore $rb + a \approx rb$. Consequently, when j is large $\Pr(A_{(j)} = rb + a) \approx \Pr(A_{(j)} = rb)$ because of the continuity of the cdf of $\{Y\}$. In such case, (2.8) is approximately constant over $a \in \mathcal{A}$, and so we have that $A_{[j]}$ is asymptotically uniformly distributed.

To conclude this remark we should point out that, albeit out of the scope of this paper, Theorem 2.4 provides a practical way to generate uniformly distributed discrete numbers from arbitrarily distributed continuous ones—even if the pdf is unknown. This is one of the goals of cryptographic key derivation functions (see for instance [17]).

3. General probability distribution of the k leading CF coefficients

Continued fraction expansions are an alternative to positional base b numeral systems for the representation of real numbers. In this section, we will obtain the general expression for the joint probability distribution of the k leading coefficients in the simple CF of a real number. Let $y_0 = y \in \mathbb{R}$ and define the recursion $y_j = \{y_{j-1}\}^{-1}$ based on the decomposition $y_{j-1} = \lfloor y_{j-1}\rfloor + \{y_{j-1}\}$. By letting $a_j = \lfloor y_j\rfloor$ we can express y as the following simple CF:

$$y = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}. \tag{3.1}$$

A CF is termed *simple* (or regular) when all the numerators of the nested fractions equal 1. For typographical convenience, we will write the CF representation of y in (3.1) using the standard notation:

$$y = [a_0; a_1, a_2, \ldots].$$

From the construction of the simple CF, we have that $a_0 \in \mathbb{Z}$, whereas $a_j \in \mathbb{N}$ for $j \geq 1$. The recursion stops if $\{y_j\} = 0$ for some j, which only occurs when $y \in \mathbb{Q}$; otherwise the CF is infinite (for an in-depth

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

introduction to CFs see [18]). Our goal in this section is to model probabilistically the a_j coefficients for $j \geq 1$. To this end, we will assume that y is drawn from a continuous r.v. Y. Because $\mathbb{Q}$ is of measure zero then $\Pr(Y \in \mathbb{Q}) = 0$, and so we may assume that the CF of y drawn from Y is almost surely infinite—and thus that the a_j coefficients are unique. In practical terms, this means that, letting $Y_0 = Y$, we can define the continuous r.v.'s:

$$Y_j = \{Y_{j-1}\}^{-1},$$

with support $(1, \infty)$ for all $j \geq 1$. Therefore, the CF coefficients we are interested in are modelled by the discrete r.v.'s:

$$A_j = \lfloor Y_j \rfloor, \tag{3.2}$$

with support $\mathbb{N}$, for all $j \geq 1$.

Additional notation. In order to streamline the presentation in this section we define the k-vector:

$$\mathbf{A}_k = [A_1, \ldots, A_k],$$

comprising the first k r.v.'s defined by (3.2), i.e., the r.v.'s modelling the k leading CF coefficients of Y. A realisation of $\mathbf{A}_k$ is $\mathbf{a}_k = [a_1, \ldots, a_k] \in \mathbb{N}^k$. Also, $\mathbf{e}_k$ denotes a unit k-vector with a one at the kth position and zeros everywhere else, i.e., $\mathbf{e}_k = [0, \ldots, 0, 1]$. A vector symbol placed within square brackets denotes a finite CF; for example, $[\mathbf{a}_k]$ denotes $[a_1; a_2, \ldots, a_k]$. Observe that we can write $[\mathbf{a}_k]^{-1} = [0; \mathbf{a}_k] = [0; a_1, a_2, \ldots, a_k]$. Finally, the subvector of consecutive entries of $\mathbf{a}_k$ between its mth entry and its last entry is denoted by $\mathbf{a}_k^{m:k} = [a_m, \ldots, a_k]$. When $m > 1$ the amount $[\mathbf{a}_k^{m:k}]$ is called a remainder of $[\mathbf{a}_k]$ [18].

As a preliminary step, we will next prove a lemma which we will then use as a stepping stone in the derivation of the joint pmf of $\mathbf{A}_k$ in Theorem 3.3.

Lemma 3.1. *The following two sets of inequalities hold for the* $(j-1)$*th order convergent* $[\mathbf{a}_j] = [a_1; a_2, \ldots, a_j]$ *of the infinite simple CF* $[a_1; a_2, a_3, \ldots]$ *with* $a_i \in \mathbb{N}$*:*

$$(-1)^{j-1}[\mathbf{a}_j] < (-1)^{j-1}[\mathbf{a}_j + \mathbf{e}_j], \tag{3.3}$$

and

$$(-1)^{j-1}[\mathbf{a}_{j-1} + \mathbf{e}_{j-1}] \leq (-1)^{j-1}[\mathbf{a}_j] < (-1)^{j-1}[\mathbf{a}_{j+1} + \mathbf{e}_{j+1}], \tag{3.4}$$

where the lower bound in (3.4) *requires* $j > 1$.

Proof. Consider the function $\varphi(\mathbf{a}_j) = [\mathbf{a}_j]$. Taking a_j momentarily to be a continuous variable, we can obtain the partial derivative of $\varphi(\mathbf{a}_j)$ with respect to a_j. For $j \geq 2$, applying the chain rule $j-1$ times yields:

$$\frac{\partial \varphi(\mathbf{a}_j)}{\partial a_j} = (-1)^{j-1} \prod_{r=2}^{j} [\mathbf{a}_j^{r:j}]^{-2}, \tag{3.5}$$

whereas when $j = 1$ we have that $d\varphi(\mathbf{a}_1)/da_1 = 1$. As the product indexed by r is positive, the sign of (3.5) only depends on $(-1)^{j-1}$.

Consequently, if j is odd then $\varphi(\mathbf{a}_j)$ is strictly increasing on a_j, and if j is even then $\varphi(\mathbf{a}_j)$ is strictly decreasing on a_j. Thus, when j is odd $[\mathbf{a}_j] < [\mathbf{a}_j + \mathbf{e}_j]$ and when j is even $[\mathbf{a}_j] > [\mathbf{a}_j + \mathbf{e}_j]$, which proves inequality (3.3).

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

Let us now prove the two inequalities in (3.4) assuming first that j is odd. Considering again (3.5), the upper bound can be obtained by seeing that $[\mathbf{a}_j] < [\mathbf{a}_{j-1}, a_j + \epsilon]$ for $\epsilon > 0$, and then choosing $\epsilon = 1/(a_{j+1} + 1)$. The lower bound, which requires $j > 1$, is due to $[\mathbf{a}_j] \geq [\mathbf{a}_{j-1}, 1] = [\mathbf{a}_{j-1} + \mathbf{e}_{j-1}]$. To conclude the proof, when j is even the two inequalities we have just discussed are reversed due to the change of sign in (3.5). □

Remark 3.2. Lemma 3.1 is closely related to the fact that the jth order convergent of a CF is smaller or larger than the CF it approximates depending on the parity of j [18, Theorem 4].

Theorem 3.3. (General distribution of the k leading CF coefficients). *For any continuous r.v. Y represented by the simple continued fraction $Y = [A_0; A_1, A_2, \dots]$ it holds that*

$$\Pr(\mathbf{A}_k = \mathbf{a}_k) = (-1)^k \big(F_{\{Y\}}([0; \mathbf{a}_k + \mathbf{e}_k]) - F_{\{Y\}}([0; \mathbf{a}_k])\big), \tag{3.6}$$

where $\mathbf{a}_k \in \mathbb{N}^k$.

Proof. For all $j \geq 2$, if $a_m \leq Y_m < a_m + 1$ for all $m = 1, \dots, j-1$ then we have that $Y_1 = [\mathbf{a}_{j-1}, Y_j]$. In these conditions it holds that $\{a_j \leq Y_j < a_j + 1\} = \{(-1)^{j-1}[\mathbf{a}_j] \leq (-1)^{j-1} Y_1 < (-1)^{j-1}[\mathbf{a}_j + \mathbf{e}_j]\}$, where the reason for the alternating signs is (3.3). Therefore

$$\begin{aligned} \Pr(\mathbf{A}_k = \mathbf{a}_k) &= \Pr(A_1 = a_1, \dots, A_k = a_k) \\ &= \Pr(\cap_{j=1}^k \{a_j \leq Y_j < a_j + 1\}) \\ &= \Pr(\cap_{j=1}^k \{(-1)^{j-1}[\mathbf{a}_j] \leq (-1)^{j-1} Y_1 < (-1)^{j-1}[\mathbf{a}_j + \mathbf{e}_j]\}). \end{aligned} \tag{3.7}$$

From (3.4), we have that the lower bounds on Y_1 in (3.7) are related as:

$$[\mathbf{a}_1] < [\mathbf{a}_2 + \mathbf{e}_2] \leq [\mathbf{a}_3] < [\mathbf{a}_4 + \mathbf{e}_4] \leq [\mathbf{a}_5] \cdots, \tag{3.8}$$

whereas the upper bounds on Y_1 in the same expression are related as:

$$[\mathbf{a}_1 + \mathbf{e}_1] \geq [\mathbf{a}_2] > [\mathbf{a}_3 + \mathbf{e}_3] \geq [\mathbf{a}_4] > [\mathbf{a}_5 + \mathbf{e}_5] \cdots. \tag{3.9}$$

Hence, except for possible equality constraints (which are anyway immaterial in probability computations with continuous random variables), the intersection of the k events in (3.7) equals the kth event, and thus

$$\begin{aligned} \Pr(\mathbf{A}_k = \mathbf{a}_k) &= \Pr((-1)^{k-1}[\mathbf{a}_k] < (-1)^{k-1} Y_1 < (-1)^{k-1}[\mathbf{a}_k + \mathbf{e}_k]) \\ &= (-1)^{k-1} \big(F_{Y_1}([\mathbf{a}_k + \mathbf{e}_k]) - F_{Y_1}([\mathbf{a}_k])\big). \end{aligned} \tag{3.10}$$

Finally, using

$$F_{Y_1}(y) = \Pr(Y_1 \leq y) = \Pr(\{Y\} \geq y^{-1}) = 1 - F_{\{Y\}}(y^{-1}),$$

in (3.10) we get (3.6). □

Remark 3.4. Observe that if we choose $Y = \log_b X$, then both (3.6) and (2.3) depend solely on the same variable $\{Y\}$, which is the reason why we have used the notation Y rather than X in this section. With this choice of Y, the general expression (3.6) models the k leading CF coefficients of $\log_b X$ (with the exception of A_0), and becomes analogous to the general expression (2.3) that models the k most

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

significant b-ary digits of X. The reason why we have left A_0 out of the joint distribution (3.6) is because, unlike the rest of variables (i.e., A_j for $j \geq 1$), it cannot be put as a sole function of Y_1. Moreover, it is not possible to model A_0 in one important practical scenario—see Section 4.1.2.

We can also verify that the joint pmf (3.6) adds up to one over its support, namely $\mathbb{N}^k$. Let us first add the joint pmf of $\mathbf{A}_k$ over $a_k \in \mathbb{N}$ assuming $k > 1$. As this infinite sum is a telescoping series, in the computation of the partial sum $S_k^{(n)} = \sum_{a_k=1}^{n} \Pr(\mathbf{A}_k = \mathbf{a}_k)$ all consecutive terms but two are cancelled, and so

$$S_k^{(n)} = (-1)^k \big(F_{\{Y\}}([0; \mathbf{a}_{k-1}, n+1]) - F_{\{Y\}}([0; \mathbf{a}_{k-1}, 1])\big).$$

Now, as $\lim_{n\to\infty}[0; \mathbf{a}_{k-1}, n+1] = [0; \mathbf{a}_{k-1}]$ and $[0; \mathbf{a}_{k-1}, 1] = [0; \mathbf{a}_{k-1} + \mathbf{e}_{k-1}]$, we then have that

$$\begin{aligned}\lim_{n\to\infty} S_k^{(n)} &= (-1)^{k-1}\big(F_{\{Y\}}([0; \mathbf{a}_{k-1} + \mathbf{e}_{k-1}]) - F_{\{Y\}}([0; \mathbf{a}_{k-1}])\big) \\ &= \Pr(\mathbf{A}_{k-1} = \mathbf{a}_{k-1}).\end{aligned}$$

The continuity of the cdf $F_{\{Y\}}(y)$ allows writing $\lim_{n\to\infty} F_{\{Y\}}(g(n)) = F_{\{Y\}}(\lim_{n\to\infty} g(n))$, which justifies the limit above. In view of this result, it only remains to verify that the pmf of $\mathbf{A}_1 = A_1$ adds up to one. The partial sum up to n is:

$$S_1^{(n)} = F_{\{Y\}}(1) - F_{\{Y\}}(1/(n+1)),$$

and therefore $\lim_{n\to\infty} S_1^{(n)} = 1$ for the same reason that makes (2.6) equal to one. Incidentally, observe that it would have been rather more difficult to verify the fact that (3.6) adds up to one by summing out the random variables in $\mathbf{A}_k$ in an order different than the decreasing order $A_k, A_{k-1}, \ldots, A_1$ that we have used above.

3.1. Distribution of the jth CF coefficient

Just like in Section 2.1, we can marginalise the joint pmf of $\mathbf{A}_j$ to obtain the distribution of the jth CF coefficient A_j of Y. Although we already know that $A_1 = \mathbf{A}_1$, the main obstacle to explicitly getting the distribution of A_j for $j > 1$ is that in this case marginalisation involves $j-1$ infinite series, rather than a single finite sum as in (2.8). In general, it is difficult to carry out the required summations in closed form. Moreover, the order of evaluation of these series may influence the feasibility of achieving a closed-form expression, which is connected to the comment in the very last sentence of the previous paragraph.

However, under the sole assumption that $\{Y\}$ is a continuous r.v. with support $[0, 1)$, the Gauss-Kuz'min theorem furnishes the general asymptotic distribution of A_j [18, Theorem 34]:

$$\lim_{j\to\infty} \Pr(A_j = a) = \log_2\left(1 + \frac{1}{a(a+2)}\right), \tag{3.11}$$

where convergence is exponentially fast on j.

Remark 3.5. Observe that Theorem 2.4, which gives the general asymptotic behaviour of $A_{[j]}$ (the jth most significant b-ary digit), is the near exact counterpart of the Gauss-Kuz'min theorem (3.11), which gives the general asymptotic behaviour of A_j (the jth CF coefficient). Apart from the convergence speed, the only essential difference is the requirement that the support of $\{Y\}$ be precisely $[0, 1)$ in (3.11) [18, Theorem 33], whereas this condition is not required in (2.9), i.e., the support of $\{Y\}$ may be a subset of $[0, 1)$ in Theorem 2.4.

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

4. Particular cases

In this section, we will particularise the general expressions in Sections 2 and 3 for two especially relevant distributions of X. As it is clear from (2.3) and (3.6), we just need the cdf $F_{\{Y\}}(y)$ of the r.v. $\{Y\} = \{\log_b X\}$ in order to achieve our goal.

4.1. Benford Variables

We consider in this section a r.v. X for which $\{Y\} \sim U(0,1)$. We call such a r.v. a *Benford variable*, although we must note that some authors call it a *strong* Benford variable instead. At any rate, this is the archetypal case in which a model of the k most significant b-ary digits has been widely used and discussed—i.e., Benford's law [7]. The cdf of $\{Y\}$ for a Benford variable X is simply:

$$F_{\{Y\}}(y) = y, \tag{4.1}$$

for $y \in [0,1)$.

4.1.1. Most significant b-ary digits of X

For a Benford variable, applying (4.1) to (2.3) yields

$$\Pr(A_{(k)} = a) = \log_b\left(1 + \frac{1}{a}\right), \tag{4.2}$$

which is the well-known Benford distribution for the k most significant b-ary digits. This distribution has almost always been expressed in previous works as the joint pmf of $A_{[1]}, \ldots, A_{[k]}$ rather than as the pmf of the kth integer significand $A_{(k)}$ (see for example [8]). As evinced in Theorems 2.1 and 2.4, and as it will become clear in the remainder of this section, working with the kth integer significant is not just an esthetical notation choice—although, conveniently, it does make for simpler expressions.

Let us obtain next the pmf of $A_{[j]}$ when $j \geq 2$ (i.e., the distribution of the jth most significant b-ary digit). From (2.8) and (4.2) we have that:

$$\Pr(A_{[j]} = a) = \sum_{r \in \mathcal{A}_{(j-1)}} \log_b\left(1 + \frac{1}{rb + a}\right) \tag{4.3}$$

$$= \log_b\left(\prod_{r \in \mathcal{A}_{(j-1)}} \frac{(a+1)b^{-1} + r}{ab^{-1} + r}\right) \tag{4.4}$$

$$= \log_b\left(\frac{\Gamma\left((a+1)b^{-1} + b^{j-1}\right)\Gamma\left(ab^{-1} + b^{j-2}\right)}{\Gamma\left((a+1)b^{-1} + b^{j-2}\right)\Gamma\left(ab^{-1} + b^{j-1}\right)}\right). \tag{4.5}$$

The last equality is due to the fact that the argument of the logarithm in (4.4) can be expressed as a fraction whose numerator and denominator are the rising factorial powers $((a+1)b^{-1} + b^{j-2})^{\overline{|\mathcal{A}_{(j-1)}|}}$ and $(ab^{-1} + b^{j-2})^{\overline{|\mathcal{A}_{(j-1)}|}}$, respectively.

We can also explicitly restate the general result in Theorem 2.4 for a Benford variable by relying on (4.5). Invoking the continuity of the logarithm and using $\lim_{z\to\infty} z^{w-v}\Gamma(v+z)/\Gamma(w+z) = 1$ [1]

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

in (4.5) twice—with $z = b^{j-1}$ and $z = b^{j-2}$, respectively—yields

$$\lim_{j\to\infty} \Pr(A_{[j]} = a) = \log_b \lim_{j\to\infty} \frac{b^{(j-1)b^{-1}}}{b^{(j-2)b^{-1}}} = b^{-1},$$

a fact that was originally pointed out by Benford [7] through the marginalisation of (4.2).

Remark 4.1. The closed-form analytic expression (4.5) for the pmf of $A_{[j]}$ deserves some comments, as it appears that it was never given in studies of Benford's distribution previous to [4]: only the equivalent of (4.3) was previously published. This is another sensible reason for working with the pmf of the jth integer significand variable $A_{(j)}$ instead of the joint pmf of $A_{[1]}, \ldots, A_{[j]}$. The former approach makes the obtention of closed-form distributions for $A_{[j]}$ more feasible: if we use $A_{(j)}$ we just have to evaluate one single sum [i.e., (2.8)], whereas if we use $A_{[1]}, \ldots, A_{[j]}$ we have to evaluate $j-1$ separate sums—which obscures the result. This appears to be the reason why previous works never produced (4.5).

Asymptotic sum-invariance property. A further advantage of working with the kth integer significand $A_{(k)}$ is that it allows for an uncomplicated statement and proof of the asymptotic sum-invariance property of a Benford variable [8, 26]. In the literature, this property has been alternatively called the "sum-invariance property" [8] or the "summation theorem" [26]. Here we prefer to stress the fact that its validity is only asymptotic: the sum-invariance property is an approximation when one considers a finite number k of most significant digits of X (i.e., the kth integer significand $A_{(k)}$, which in fact originally motivated the definition of the property by Nigrini [25]).

Theorem 4.2. (Asymptotic sum-invariance property of Benford variables). *If X is a Benford variable, then it holds that:*

$$\lim_{\substack{k\to\infty,\\ a\in\mathcal{A}_{(k)}}} a \Pr(A_{(k)} = a) = (\ln b)^{-1}. \tag{4.6}$$

Proof. We just need to see that $\lim_{k\to\infty, a\in\mathcal{A}_{(k)}} a \log_b \left(1 + \frac{1}{a}\right) = \lim_{v\to\infty} v \log_b \left(1 + \frac{1}{v}\right)$ due to (2.2). The proof is completed using either l'Hôpital's theorem, or the continuity of the logarithmic function and the definition of Euler's number as $\lim_{v\to\infty}(1 + 1/v)^v$. □

Remark 4.3. Informally, Theorem 4.2 tells us that in a large set of outcomes from a Benford variable the sum of $a \in \mathcal{A}_{(k)}$ over all those outcomes whose kth integer significand is equal to a is roughly invariant over a when k is large (i.e., Nigrini's empirical observation). The convergence speed to limit (4.6) is exponential on k—the faster the larger b is. Figure 1 shows that the sum-invariance property holds approximately when $k = 3$ already, for $b = 10$. Theorem 4.2 also implies that

$$\lim_{k\to\infty} \frac{\mathrm{E}(A_{(k)})}{|\mathcal{A}_{(k)}|} = (\ln b)^{-1}.$$

The corresponding approximation $\mathrm{E}(A_{(k)}) \approx (b^k - b^{k-1})(\ln b)^{-1}$ improves with k, but it never achieves strict equality for finite k—in fact, $\mathrm{E}(A_{(k)}) < (b^k - b^{k-1})(\ln b)^{-1}$ for all k.

4.1.2. Leading CF coefficients of $\log_b X$

For a Benford variable the application of (4.1) to (3.6) yields

$$\Pr(\mathbf{A}_k = \mathbf{a}_k) = (-1)^k \big([0; \mathbf{a}_k + \mathbf{e}_k] - [0; \mathbf{a}_k]\big). \tag{4.7}$$

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

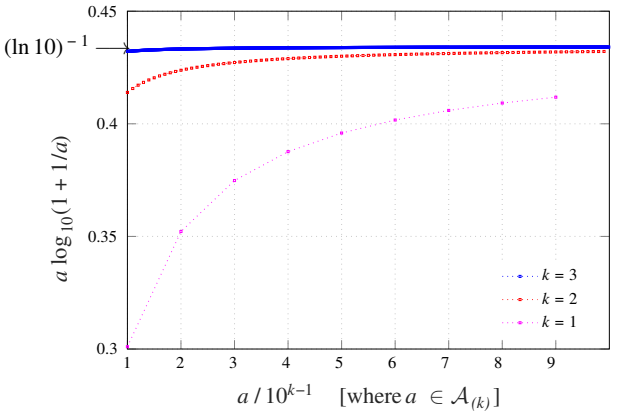

Figure 1. *Illustration of the asymptotic sum-invariance property of a Benford variable for* $b = 10$.

According to our discussion in Remark 3.4, this distribution of the k leading CF coefficients of $\log_b X$ is the counterpart of Benford's distribution of the k most significant b-ary digits of X. Therefore (4.7) is the analogue of Benford's law for CFs. In particular, any real dataset that complies with (4.2) will also comply with (4.7).

By transforming the subtraction of fractions into a single fraction, (4.7) can also be written as the inverse of the product of $[\mathbf{a}_k]$, $[\mathbf{a}_k + \mathbf{e}_k]$ and all of their remainders, i.e.,

$$\Pr(\mathbf{A}_k = \mathbf{a}_k) = \prod_{j=1}^{k} \left[0; \mathbf{a}_k^{j:k}\right] \left[0; \mathbf{a}_k^{j:k} + \mathbf{e}_{k-j+1}\right], \tag{4.8}$$

which, apart from showing at a glance that (4.7) cannot be negative, may find application in log-likelihood computations. The equivalent of expression (4.8) was previously given by Miller and Takloo-Bighash [22, Lemma 10.1.8] in their exploration of the distribution of CF coefficients—called digits by these authors. However, unlike our result above, the version of (4.8) given by Miller and Takloo-Bighash is not explicit, as it is presented in terms of CF convergents. Therefore, (4.8) or (4.7) are clearly more useful when it comes to practical applications—furthermore, we show in Section 5 the empirical accuracy of our expressions using both synthetic and real data, something that was not attempted by Miller and Takloo-Bighash. Lastly, Blachman also gave the following explicit approximation for uniform $\{Y\}$ [11, equation (9)]:

$$\Pr(\mathbf{A}_k = \mathbf{a}_k) \approx \left| \log_2 \left(\frac{1 + [0; \mathbf{a}_k]}{1 + [0; \mathbf{a}_k + \mathbf{e}_k]} \right) \right|. \tag{4.9}$$

The reader should be cautioned that this expression appears in [11] with an equal sign, although the author unequivocally produced it as an approximation. Using $\ln(1+z) \approx z$, which is accurate for $|z| \ll 1$, we can see that (4.9) is roughly off by a factor of $(\ln 2)^{-1}$ with respect to the exact expression (4.7).

Let us now look at the marginals, that is to say, the distributions of individual A_j coefficients. When $k = 1$ expression (4.7) gives the distribution of $A_1 = \mathbf{A}_1$ straightaway:

$$\Pr(A_1 = a) = a^{-1} - (a+1)^{-1}. \tag{4.10}$$

This pmf, also previously given by Miller and Takloo-Bighash [22, page 232], can be rewritten as $\Pr(A_1 = a) = a^{-1}(a+1)^{-1}$, which is the form that (4.8) takes in this particular case. In passing, observe that $\mathrm{E}(A_1) = \infty$ because of the divergence of the harmonic series. It is also instructive to particularise

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

Blachman's approximation (4.9) for A_1 [11, equation (10)]: this particularisation renders the asymptotic Gauss-Kuz'min law (3.11) instead of the exact pmf (4.10).

Keeping in mind our discussion at the start of Section 3.1, the Benford case is probably unusual in the fact that we can also obtain the distribution of A_2 in closed form by marginalising (4.7) for $k=2$. Summing $\Pr(\mathbf{A}_2 = \mathbf{a}_2) = 1/(a_1 + (a_2+1)^{-1}) - 1/(a_1 + a_2^{-1})$ over $a_1 \in \mathbb{N}$, and using the digamma function defined as $\psi(1+z) = -\gamma + \sum_{n=1}^{\infty} z/(n(n+z))$ [1]—which is applicable because the range of validity $z \notin \mathbb{Z}^-$ of this definition always holds here—one finds that:

$$\Pr(A_2 = a) = \psi\left(1 + a^{-1}\right) - \psi\left(1 + (1+a)^{-1}\right). \tag{4.11}$$

It does not seem possible to obtain a closed-form exact expression for the distribution of a single CF coefficient A_j when $j > 2$ in the Benford case. However it is possible to explicitly produce the Gauss-Kuz'min law (3.11) from (4.7) by pursuing an approximation of $\Pr(A_j = a_j)$ for all $j \geq 2$. To see this, consider first the sum:

$$\sum_{x=1}^{\infty} \left(\frac{1}{x+b} - \frac{1}{x+c}\right) = \psi(1+c) - \psi(1+b), \tag{4.12}$$

for some $b, c > 0$, which is just a generalisation of (4.11), and its integral approximation:

$$\int_1^{\infty} \left(\frac{1}{x+b} - \frac{1}{x+c}\right) dx = \ln(1+c) - \ln(1+b), \tag{4.13}$$

which attests to the intimate connection between the digamma function and the natural logarithm [3, see Exercise 8.2.20 and equation (8.51)]. Now, in the marginalisation of (4.7) that leads to $\Pr(A_j = a_j)$ the summation on a_1 is of the form (4.12), and so we may approximate it by the integral (4.13):

$$\Pr(A_j = a_j) = (-1)^j \sum_{a_{j-1}=1}^{\infty} \cdots \sum_{a_1=1}^{\infty} \left([0; \mathbf{a}_j + \mathbf{e}_j] - [0; \mathbf{a}_j]\right) \tag{4.14}$$

$$\approx (-1)^j \sum_{a_{j-1}=1}^{\infty} \cdots \sum_{a_2=1}^{\infty} \int_1^{\infty} \left([0; \mathbf{a}_j + \mathbf{e}_j] - [0; \mathbf{a}_j]\right) da_1$$

$$= (-1)^j \sum_{a_{j-1}=1}^{\infty} \cdots \sum_{a_2=1}^{\infty} (-1)\left(\ln(1 + [0; \mathbf{a}_j^{2:j} + \mathbf{e}_{j-1}]) - \ln(1 + [0; \mathbf{a}_j^{2:j}])\right). \tag{4.15}$$

As $[0; \mathbf{a}_j^{i:j}] \ll 1$ nearly always for $\mathbf{a}_j^{i:j} = [a_i, \cdots, a_j] \in \mathbb{N}^{j-i+1}$, then we may use $\ln(1+z) \approx z$ in (4.15) to obtain:

$$\Pr(A_j = a_j) \approx (-1)^{j+1} \sum_{a_{j-1}=1}^{\infty} \cdots \sum_{a_2=1}^{\infty} \left([0; \mathbf{a}_j^{2:j} + \mathbf{e}_{j-1}]) - [0; \mathbf{a}_j^{2:j}]\right). \tag{4.16}$$

Remarkably, (4.16) has the exact same form as (4.14)—but with one less infinite summation. Therefore we can keep sequentially applying the same approximation procedure described above to the summations on $a_2, a_3, \ldots, a_{j-1}$. In the final summation on a_{j-1} we do not need the approximation used in (4.16) anymore, and thus in the last step we have that:

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

$$\begin{aligned}\Pr(A_j = a_j) &\approx (-1)^{2j-2} \int_1^\infty \left([0;\mathbf{a}_j^{j-1:j} + \mathbf{e}_2]) - [0;\mathbf{a}_j^{j-1:j}]\right) da_{j-1} \\ &= (-1)^{2j-2} \int_1^\infty \left(\frac{1}{a_{j-1} + \dfrac{1}{a_j + 1}} - \frac{1}{a_{j-1} + \dfrac{1}{a_j}} \right) da_{j-1} \\ &= (-1)^{2j-1} \left(\ln\left(1 + \frac{1}{a_j + 1}\right) - \ln\left(1 + \frac{1}{a_j}\right) \right) \qquad (4.17) \\ &= \ln \frac{(a_j+1)^2}{a_j(a_j+2)}. \qquad (4.18)\end{aligned}$$

Due to the successive approximations (4.18) is not necessarily a pmf, and so we need to normalise it. The normalisation factor is:

$$\sum_{a_j=1}^{\infty} \ln \frac{(a_j+1)^2}{a_j(a_j+2)} = \ln \prod_{a_j=1}^{\infty} \frac{(a_j+1)^2}{a_j(a_j+2)} = \ln \frac{2 \cdot \cancel{2}}{1 \cdot \cancel{3}} \cdot \frac{\cancel{3} \cdot \cancel{3}}{\cancel{2} \cdot \cancel{4}} \cdot \frac{\cancel{4} \cdot \cancel{4}}{\cancel{3} \cdot \cancel{5}} \cdots = \ln 2.$$

Applying this factor to (4.18), we finally obtain:

$$\Pr(A_j = a_j) \approx \log_2 \frac{(a_j+1)^2}{a_j(a_j+2)} = \log_2 \left(1 + \frac{1}{a_j(a_j+2)}\right), \qquad (4.19)$$

for $j \geq 2$, i.e., the Gauss-Kuz'min law (3.11).

Remark 4.4. Like in the verification of the general joint pmf (3.6) in Remark 3.4, the right order of evaluation of the marginalisation sums has been again key for us to be able to produce approximation (4.19). Also, had we used $\ln(1+z) \approx z$ one last time in (4.17) then we would have arrived at (4.10) instead of at the Gauss-Kuz'min law as the final approximation. This shows that the pmf of the first CF coefficient and the asymptotic law are close already, which was also mentioned by Miller and Takloo-Bighash [22, Exercise 10.1.1]. Since the convergence of the distribution of A_j to the asymptotic distribution is exponentially fast on j, it is unsurprising that the pmf (4.11) of the second CF coefficient turns out to be even closer to (3.11), as suggested by approximation (4.19)—see empirical validation in Section 5.

Although beyond the goals of this paper, it should be possible to refine the approximation procedure that we have given to get (4.18) in order to to explicitly obtain the exponential rate of convergence to the Gauss-Kuz'min law, by exploiting the expansion of the digamma function in terms of the natural logarithm and an error term series [3, equation (8.51)].

Finally, see that although A_0 is not included in the joint pmf (4.7), this variable cannot be modelled anyway when the only information that we have about X is its "Benfordness".

4.1.3. Benford variables and the asymptotics of the general analysis

To conclude Section 4.1, we examine the role played by the particular analysis for a Benford variable [i.e., (4.2) and (4.7)] in the general analysis [i.e., Theorems 2.1 and 3.3] when k is large. Let us start by looking at the asymptotics of (2.3). For any $\epsilon > 0$ there exists $k_{\min}$ such that $\log_b(1 + a^{-1}) < \epsilon$ for all $k \geq k_{\min}$ and $a \in \mathcal{A}_{(k)}$. Explicitly, this minimum index is $k_{\min} = \lceil -\log_b(b^\epsilon - 1) + 1 \rceil + 1$. This inequality

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

and the continuity of $F_{\{Y\}}(y)$ allow us to approximate (2.3) for large k using the pdf of $\{Y\}$ as:

$$\Pr(A_{(k)} = a) \approx f_{\{Y\}}(\log_b a - k + 1)\ \log_b\left(1 + \frac{1}{a}\right). \tag{4.20}$$

We now turn our attention to the asymptotics of (3.6). Similarly as above, for any $\epsilon > 0$ (3.8) and (3.9) guarantee that there exists $k_{\min}$ such that $(-1)^k([0;\mathbf{a}_k + \mathbf{e}_k] - [0;\mathbf{a}_k]) < \epsilon$ for all $k \geq k_{\min}$. Invoking again the continuity of $F_{\{Y\}}(y)$ we can approximate (3.6) for large k using again the pdf of $\{Y\}$ as:

$$\Pr(\mathbf{A}_k = \mathbf{a}_k) \approx f_{\{Y\}}([0;\mathbf{a}_k])\ (-1)^k\big([0;\mathbf{a}_k + \mathbf{e}_k] - [0;\mathbf{a}_k]\big). \tag{4.21}$$

The key point that we wish to make here is that the Benford expressions (4.2) and (4.7) appear as factors in the asymptotic approximations (4.20) and (4.21) of the general expressions (2.3) and (3.6), respectively. This shows the special place that Benford variables take in the problems of modelling significant digits and leading CF coefficients. Of course, for Benford X the pdf of $\{Y\}$ is $f_{\{Y\}}(y) = 1$ for $y \in [0, 1)$, and so in this case approximations (4.20) and (4.21) coincide with their exact counterparts.

4.2. Pareto variables

In this section, we let X be a Pareto r.v. with minimum value x_{m} and shape parameter s, whose pdf is

$$f_X(x) = s\, x_{\mathrm{m}}^s\, x^{-(s+1)}, \quad 0 < x_{\mathrm{m}} \leq x,\ s > 0.$$

The main motivation for considering the Pareto distribution is its pervasiveness in natural phenomena, which is reflected in the fact that Pareto variables are able to model a wealth of scale-invariant datasets. According to Nair *et al.* [23], heavy-tailed distributions are just as prominent as the Gaussian distribution, if not more. This is a consequence of the Central Limit Theorem (CLT) *not* yielding Gaussian distributions—but heavy-tailed ones—in common scenarios where the variance of the random variables being added is infinite (or does not exist). Furthermore, heavy-tailed distributions appear when the CLT is applied to the logarithm of variables emerging from multiplicative processes. In this context, the relevance of the Pareto distribution owes to the fact that the tails of many heavy-tailed distributions follow the Pareto law. The fact that the Pareto distribution is the only one that fulfills exactly the relaxed scale-invariance criterion:

$$f_X(x) = \alpha^{s+1} f_X(\alpha\, x), \tag{4.22}$$

for any scaling factor $\alpha > 0$, where $s > 0$, will also be seen to be relevant to our discussion.

Let us firstly obtain the cdf of $\{Y\}$ in this case. The cdf of a Pareto r.v. X is $F_X(x) = 1 - x_{\mathrm{m}}^s x^{-s}$ for $x \geq x_{\mathrm{m}}$, and thus the cdf of $Y = \log_b X$ is $F_Y(y) = F_X(b^y) = 1 - x_{\mathrm{m}}^s b^{-sy}$ for $y \geq \log_b x_{\mathrm{m}}$. Letting

$$\rho = \{\log_b x_{\mathrm{m}}\},$$

and using (1.2), we have that the cdf of $\{Y\}$ for a Pareto r.v. X is:

$$F_{\{Y\}}(y) = b^{s(\rho-1)}\ \frac{1 - b^{-sy}}{1 - b^{-s}} + u(y - \rho)\left(1 - b^{-s(y-\rho)}\right), \tag{4.23}$$

for $y \in [0, 1)$, where $u(\cdot)$ is the unit-step function.

Remark 4.5. By application of l'Hôpital's rule, it can be verified that (4.23) tends to (4.1) as $s \downarrow 0$, and so a Pareto variable becomes asymptotically Benford as its shape parameter s vanishes—for any value

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

of ρ. Because (2.3) and (3.6) only depend on $\{Y\}$, the distributions that we will produce in this section generalise their counterparts in the previous section [i.e. (4.2), (4.5) and (4.7) are particular cases of (4.24), (4.25) and (4.27), respectively, when $s \downarrow 0$]. The fact that Benford variables can appear as a particular case of Pareto variables is a likely reason for the sporadic emergence of Benford's distribution (4.2) in scale-invariant scenarios. Finally, observe that, asymptotically as $s \downarrow 0$, the relaxed scaled invariance property (4.22) becomes *strict*, i.e., $f_X(x) = \alpha f_X(\alpha x)$. Strict scale invariance is, in turn, a property that can drive the appearance of Benford's distribution [4].

An interesting line of research beyond the scope of this paper would entail pursuing analytical insights about the probability distribution of the s parameter itself in scale-invariant scenarios. If this distribution could be found, perhaps under constraints yet to be specified, it would determine the frequency of emergence of Benford variables in those scenarios. In any case, it can be empirically verified that scale-invariant datasets are far more often Paretian rather than just Benfordian (see some examples in Figure 6). Thus, the expressions that we will give in this section may have wider practical applicability than the ones in Section 4.1 in the context of scale-invariant datasets—with the caveat that two parameters (s and ρ or x_{m}) must be estimated when using the Pareto distribution results.

4.2.1. Most significant b-ary digits of X

Combining (2.3) and (4.23), and Letting:

$$\xi = \rho + k - 1,$$

yields the Paretian generalisation of (4.2):

$$\begin{aligned}\Pr(A_{(k)} = a) = {} & \frac{b^{s(\xi-1)}}{1-b^{-s}}\left(a^{-s} - (a+1)^{-s}\right) \\ & + u\left(a+1-b^{\xi}\right)\left(1-b^{s\,\xi}(a+1)^{-s}\right) - u\left(a-b^{\xi}\right)\left(1-b^{s\,\xi}\,a^{-s}\right).\end{aligned} \tag{4.24}$$

Let us obtain the distribution of $A_{\lfloor j \rfloor}$ for $j \geq 2$ next. For this single purpose we make two definitions: $\eta_v = \lceil b^{\xi-1} - vb^{-1} \rceil$ and

$$\tau_s(v) = \begin{cases} -\psi(v), & s = 1 \\ \zeta(s,v), & s \neq 1 \end{cases},$$

where $\psi(\cdot)$ is again the digamma function and $\zeta(s,v) = \sum_{n=0}^{\infty}(n+v)^{-s}$ is Hurwitz's zeta function [2]. Now, combining (2.8) and (4.24) and using the two previous definitions it is tedious but straightforward to show that the Paretian generalisation of (4.5) is:

$$\begin{aligned}\Pr(A_{\lfloor j \rfloor} = a) = {} & \frac{b^{s(\xi-2)}}{1-b^{-s}}\left(\tau_s(ab^{-1}+b^{j-2}) - \tau_s((a+1)b^{-1}+b^{j-2})\right) \\ & - \frac{b^{s(\xi-1)}}{1-b^{-s}}\left(\tau_s(ab^{-1}+b^{j-1}) - \tau_s((a+1)b^{-1}+b^{j-1})\right) \\ & + b^{s(\xi-1)}\left(\tau_s(ab^{-1}+\eta_a) - \tau_s((a+1)b^{-1}+\eta_{a+1})\right) \\ & + \eta_a - \eta_{a+1}.\end{aligned} \tag{4.25}$$

Remark 4.6. Like in Section 4.1.1, we have been able to obtain a closed-form expression for the pmf of $A_{\lfloor j \rfloor}$ thanks to the use of the jth integer significand. Of particular interest is the distribution of $A_{(k)}$ (4.24), which had only been published before our own work [4] for the special case in which the fractional part

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

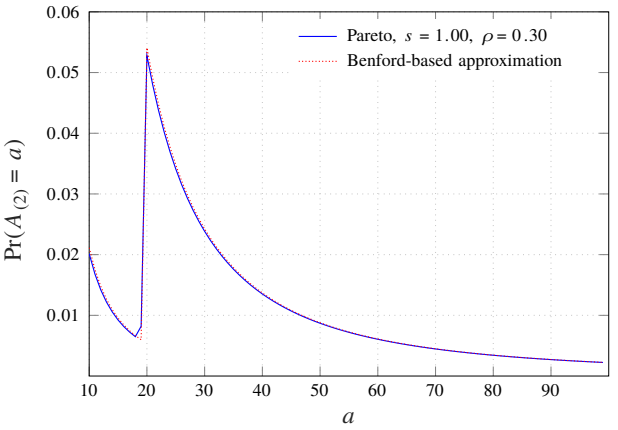

Figure 2. *Theoretical distribution of the two most significant decimal digits of Pareto X (4.24) versus theoretical Benford-based asymptotic approximation (4.20). The lines join adjacent probability mass points for clarity.*

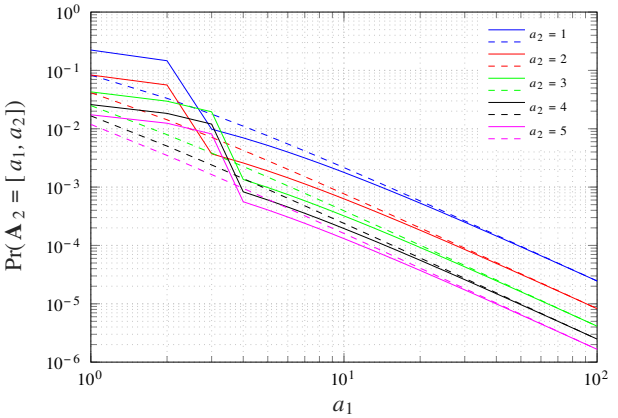

Figure 3. *Theoretical joint pmf of the first two CF coefficients of* $\log_{10} X$ *for Pareto X with* $s = 1$ *and* $\rho = 0.3$ *[solid lines, (4.27)] versus theoretical Benford-based asymptotic approximation [dashed lines, (4.21)]. The lines join adjacent probability mass points corresponding to equal* a_2 *for clarity.*

of the minimum of the Pareto distribution is zero, i.e., $\rho = 0$, and thus $\xi = k - 1$. In this case (4.24) becomes

$$\Pr(A_{(k)} = a) = \frac{a^{-s} - (a+1)^{-s}}{b^{-s(k-1)} - b^{-sk}}. \tag{4.26}$$

The case $k = 1$ of (4.26) was first given by Pietronero *et al.* [29] in the course of their investigation on the generalisation of Benford's distribution to scale-invariant phenomena. Barabesi and Pratelli [6] then extended Pietronero *et al.*'s result and obtained (4.26) itself. A related expression was also produced by Tseng *et al.* [33] for the most significant decimal digit in a special case of the double Pareto distribution. As we will empirically verify in Section 5, the fact that (4.24) can handle the general case $\rho > 0$ is not a minor detail, but a major factor in terms of that expression being able to model real data that cannot be modelled by (4.26) alone.

Interestingly, (4.26) was first identified as a new distribution only a few years ago by Kozubowski *et al.* [19], who called it the *discrete truncated Pareto* (DTP) distribution. Kozubowski *et al.*

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

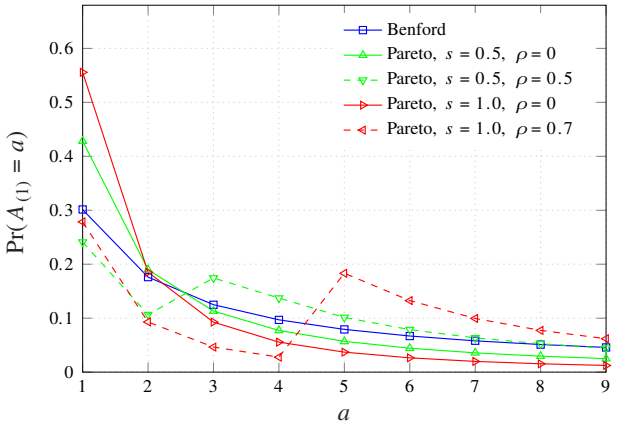

Figure 4. *Distributions of the most significant decimal digit of X. The theoretical pmf's (solid and dashed lines) are* (4.2) *and* (4.24)*, and the empirical frequencies (symbols) correspond to* $p = 10^7$ *pseudorandom outcomes in each case.*

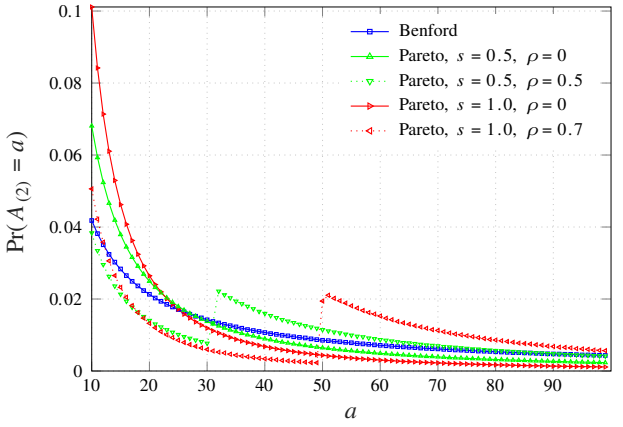

Figure 5. *Distributions of the two most significant decimal digits of X. The theoretical pmf's (solid and dotted lines) are* (4.2) *and* (4.24)*, and the empirical frequencies (symbols) correspond to* $p = 10^7$ *pseudorandom outcomes in each case.*

also noticed that the DTP distribution generalises Benford's distribution, but they landed on this fact solely because of the mathematical form of (4.26). In fact, their practical motivation was far removed from the distribution of most significant digits: it was a biological problem involving the distribution of diet breadth in Lepidoptera. Another striking fact is that Kozubowski *et al.* arrived at the DTP distribution through the quantisation of a *truncated* Pareto variable, instead of through the discretisation of the fractional part of the logarithm of a *standard* Pareto variable—i.e., the procedure that we have followed to get to (4.26), which is the ultimate reason why the DTP distribution is connected with Benford's distribution. A Pareto variable must surely be the only choice for which two such remarkably different procedures yield the very same outcome. The reason for this serendipitous coincidence is that the complementary cdf of the variable to be quantised or discretised, respectively, turns out to be a negative exponential function in both cases. To end this remark, Kozubowski *et al.* rightly point out that that the shape parameter s in (4.26) can be taken to be negative in terms of its validity as a pmf. However observe that s must be strictly positive for (4.26) to have physical meaning in terms of modelling a distribution of most significant digits.

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

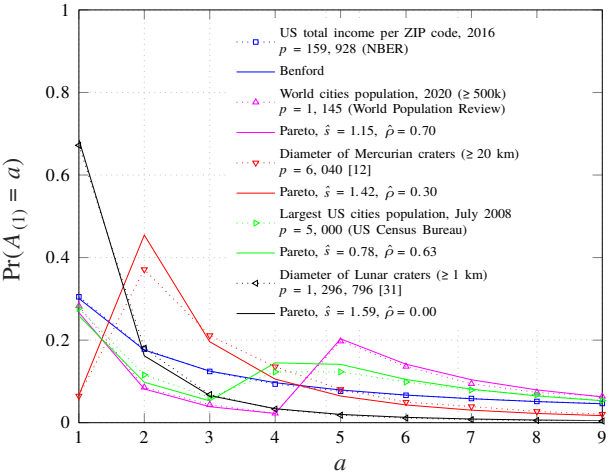

Figure 6. *Distributions of the most significant decimal digit of X for real Benfordian and Paretian datasets. The theoretical pmf's (solid lines) are* (4.2) *and* (4.24)*, and the empirical frequencies (symbols) are joined through dotted lines for clarity.*

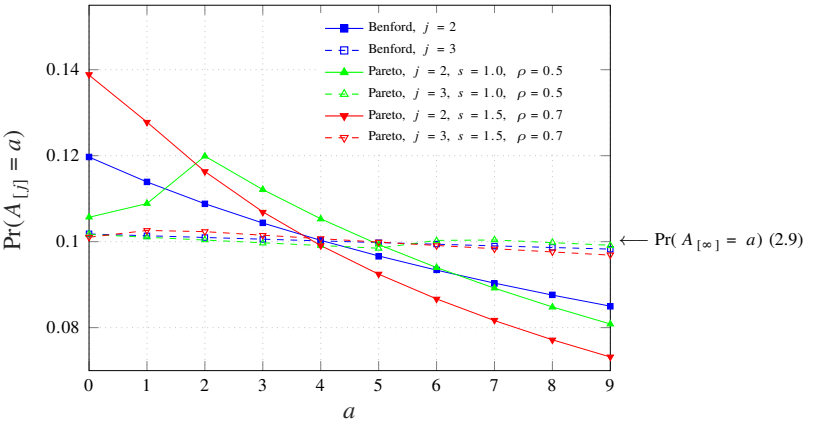

Figure 7. *Distributions of the* j*th most significant decimal digit of X. The theoretical pmf's (solid and dashed lines) are* (4.5) *and* (4.25)*, and the empirical frequencies (symbols) correspond to* $p = 5 \times 10^7$ *pseudorandom outcomes in each case.*

4.2.2. Leading CF coefficients of $\log_b X$

Applying (4.23) to (3.6) yields the Paretian generalisation of (4.7):

$$\begin{aligned}\Pr(\mathbf{A}_k = \mathbf{a}_k) = (-1)^k \Big(& b^{s(\rho-1)} \, \frac{b^{-s[0;\mathbf{a}_k]} - b^{-s[0;\mathbf{a}_k+\mathbf{e}_k]}}{1-b^{-s}} \\ & + u\big([0;\mathbf{a}_k+\mathbf{e}_k]-\rho\big)\,\big(1-b^{-s([0;\mathbf{a}_k+\mathbf{e}_k]-\rho)}\big) \\ & - u\big([0;\mathbf{a}_k]-\rho\big)\,\big(1-b^{-s([0;\mathbf{a}_k]-\rho)}\big)\Big). \end{aligned} \tag{4.27}$$

The special case of (4.27) for $\rho = 0$ yields the counterpart of the DTP distribution (4.26) in the CF setting:

$$\Pr(\mathbf{A}_k = \mathbf{a}_k) = (-1)^k \left(\frac{b^{-s[0;\mathbf{a}_k]} - b^{-s[0;\mathbf{a}_k+\mathbf{e}_k]}}{1-b^{-s}} \right). \tag{4.28}$$

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

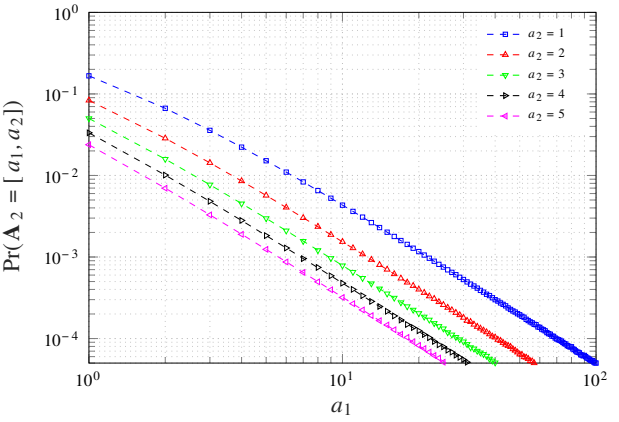

Figure 8. *Joint distribution of the first two CF coefficients of* $\log_{10} X$ *for Benford X. The theoretical joint pmf (dashed lines) is* (4.7) *and the empirical frequencies (symbols) correspond to* $p = 10^8$ *pseudorandom outcomes.*

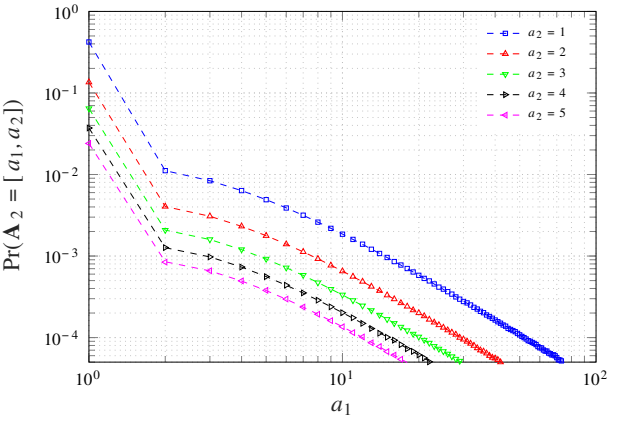

Figure 9. *Joint distribution of the first two CF coefficients of* $\log_{10} X$ *for Pareto X,* $s = 1.5$, $\rho = 0.48$. *The theoretical joint pmf (dashed lines) is* (4.27) *and the empirical frequencies (symbols) correspond to* $p = 10^8$ *pseudorandom outcomes.*

Expression (4.27) is clearly not amenable to analytic marginalisation beyond $A_1 = \mathbf{A}_1$. An interesting particular case of A_1 is given by specialising (4.28) for $k = 1$:

$$\Pr(A_1 = a) = \frac{b^{-\frac{s}{a+1}} - b^{-\frac{s}{a}}}{1 - b^{-s}}. \tag{4.29}$$

Recalling Remark 4.5, (4.29) tends to (4.10) as $s \downarrow 0$.

4.2.3. Comparison with Benford-based asymptotic approximations

Now that we have particularised (2.3) and (3.6) for a non-Benfordian variable, we are in a position to examine how well the Paretian expressions (4.24) and (4.27) are approximated by the Benford-based asymptotic expressions discussed in Section 4.1.3. In order to evaluate approximations (4.20) and (4.21) we need the pdf of $\{Y\}$, $f_{\{Y\}}(y)$. This is obtained by differentiating (4.23), which yields $f_{\{Y\}}(y) = s(\ln b)\, b^{-s(y-\rho)} \left(b^{-s}/(1 - b^{-s}) + u(y - \rho)\right)$ for $y \in [0, 1)$. We now compare in Figures 2 and 3 the exact Paretian expressions and their Benford-based approximations for some arbitrary values of s and

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

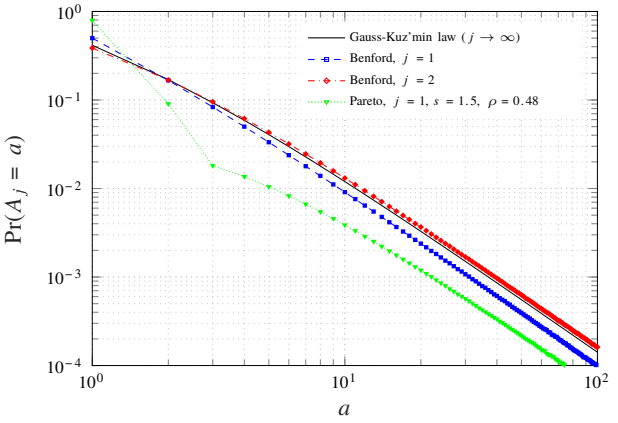

Figure 10. *Distributions of the* j*th CF coefficient of* $\log_{10} X$*. The theoretical pmf's (solid, dashed, dash-dotted and dotted lines) are* (3.11)*,* (4.10)*,* (4.11) *and* (4.27) *[with* $k=1$*]. The empirical frequencies (symbols) correspond to* $p = 10^8$ *pseudorandom outcomes in each of the cases.*

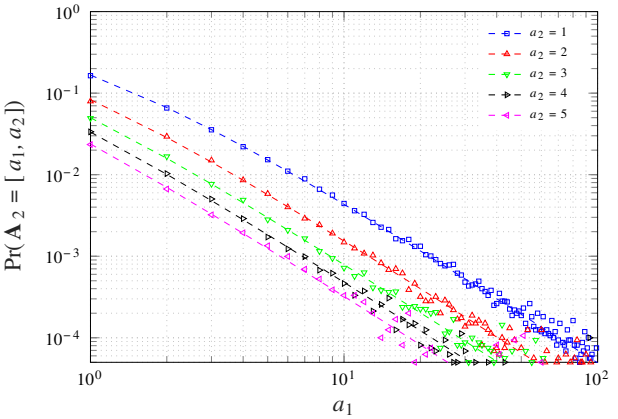

Figure 11. *Joint distribution of the first two CF coefficients of* $\log_{10} X$ *for a real Benfordian dataset (US total income per ZIP code, National Bureau of Economic Research, 2016,* $p = 159,928$*). The theoretical joint pmf (dashed lines) is* (4.7)*, whereas the symbols represent empirical frequencies.*

ρ and for $k=2$, which we have chosen because the visualisation of a joint distribution is not simple for $k>2$ in the CF case. Even though the Benford-based approximations were obtained assuming k to be large, we can see that they are close to the exact expressions already. This is markedly true for the most significant digits model in Figure 2, where the approximation is accurate for nearly all $a \in \mathcal{A}_{(2)}$ already—in fact, it can be verified that the asymptotic approximation is also acceptable for $k=1$ in this case. Regarding the CF coefficients model, the asymptotic approximation becomes accurate when $a_1 + a_2 \gtrapprox 100$ for $k=2$.

5. Empirical tests

In this section the theoretical expressions given in Sections 4.1 and 4.2 are verified. In all plots, lines represent theoretical probabilities (joining adjacent discrete probability mass points) except otherwise indicated, whereas symbols (squares or triangles) represent empirical frequencies obtained using some dataset $\{x_1, x_2, \ldots, x_p\}$. For simplicity, we use the maximum likelihood estimators $\hat{x}_{\mathrm{m}} = \min_i x_i$ and

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

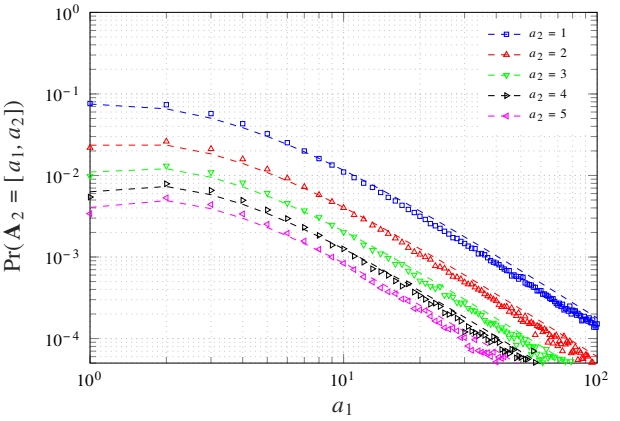

Figure 12. *Joint distribution of the first two CF coefficients of* $\log_{10} X$ *for a real Paretian dataset (diameter of Lunar craters* ≥ 1 *km,* $p = 1,296,796$ [31]*). The theoretical joint pmf (dashed lines) is* (4.27)*, driven by* $\hat{s} = 1.59$ *and* $\hat{\rho} = 0.00$*, whereas the symbols represent empirical frequencies.*

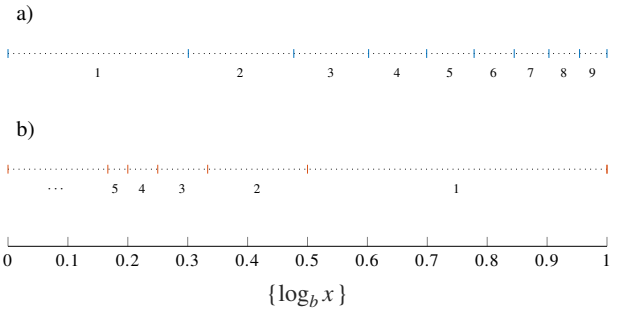

Figure 13. *Mappings between the range of* $\{\log_b x\}$ *and the supports of a)* $A_{(1)}$ *and b)* A_1*.*

$\hat{s} = \left(\frac{1}{p}\sum_i \ln(x_i/\hat{x}_\mathrm{m})\right)^{-1}$ to drive the Paretian expressions with real datasets, but the reader should be aware that better estimation approaches are possible (see for instance [23]).

We start with distributions of the most significant digits of X. Figures 4 and 5 present the distributions of the most significant decimal digit $A_{(1)}$ and of the two most significant decimal digits $A_{(2)}$, respectively. The Benford results, which are a particular case of (4.24), are well known. The Paretian cases with $\rho = 0$ are covered by (4.26), as previously shown by Barabesi and Pratelli [6]. However, it is essential to use the general expression (4.24) when $\rho > 0$. In this case the pmf's of Paretian significant digits do not behave anymore in a monotonically decreasing way (i.e., like Benford's pmf) but rather feature a peak midway along the support of $A_{(k)}$. Therefore modelling real Paretian datasets requires being able to take a general value of ρ into account, as there is no special reason why ρ should be zero in practice—observe some examples in Figure 6. In particular, observe the pronounced peak when using the dataset in [12].

Next, Figure 7 shows distributions of the jth most significant decimal digit $A_{[j]}$. Again, peaks can sometimes be seen in the distributions when $\rho > 0$, but in general these are less pronounced than in the distribution of $A_{(k)}$ due to (2.9). An illustration of the asymptotic uniform behaviour proved in Theorem 2.4 is the fact that $A_{[3]}$ is very close already to $A_{[\infty]}$ for all three distributions of X considered in Figure 7.

We move on next to distributions of the leading CF coefficients of $\log_{10} X$. Figures 8 and 9 verify the validity of the joint distributions (4.7) and (4.27) of the two leading CF coefficients $\mathbf{A}_2$ when X is Benford and Pareto, respectively. For clarity, only the first five values of a_2 are shown. In Figure 10,

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

we show the distributions of the marginals A_1 and A_2 for Benford X [i.e., (4.10) and (4.11)], and we compare them to the Gauss-Kuz'min law (3.11) which is the distribution of A_∞. As we can see, A_j converges really fast to A_∞: A_1 is close to it already, as remarked by Miller and Takloo-Bighash [22], but A_2 is noticeably closer, as expected from approximation (4.19). Figure 10 also depicts the distribution of A_1 for Paretian X, using (4.27). As we know, for Pareto X the distribution of A_j must also converge exponentially fast to the Gauss-Kuz'min law, although this is not graphically illustrated in Figure 10 due to the lack of a theoretical expression for A_2 in this case. Finally, we show in Figures 11 and 12 how (4.7) and (4.27) correctly model real Benfordian and Paretian datasets, respectively.

6. Conclusions

We have provided a general theoretical analysis of the distributions of the most significant digits and the leading CF coefficients of the outcomes of an arbitrary random variable, which highlights the connections between the two subjects. Empirical verification for two relevant particularisations of our general expressions (for Benford and Pareto variables, respectively) also supports the accuracy of our results in practice. Our analysis reveals novel facts—especially, but not only, concerning modelling CF coefficients—and provides simpler proofs and new closed-form expressions for already known ones. In particular, we have shown that the use of what we have proposed to call kth integer significands considerably simplifies modelling significant digits, allowing for uncomplicated finite and asymptotic analyses. We have also shown the parallelism between the general asymptotics of the probabilistic models for the jth significant b-ary digit and of the jth CF coefficient—i.e., between (2.9) and the Gauss-Kuz'min law (3.11)—and the role played by the Benford variables in the asymptotics of the general analyses.

Since power-law distributions are found in an extensive array of systems, our results may enable the use of significant digits models with non-Benford data in all areas where Benford's law has been previously applied. For example, consider the economic data in [32] or the hydrological data in [27]. It should also be possible to devise new hypothesis tests based on our models for leading CF coefficients, along the lines of the work of Barabesi *et al.* [5], both for Benford and non-Benford data.

We must finally note that the fundamental connection between significant digits and leading CF coefficients models is also the source of their main limitation when used in hypothesis tests: both models arise from discretisations of the fractional part of the logarithmic data, i.e., of the continuous random variable $\{\log_b X\}$ (see comparison in Figure 13 for $k = 1$) and therefore they necessarily discard potentially valuable information. This limitation becomes less important as the number k of significant digits, or leading CF coefficients, increases. Nonetheless there is probably room to investigate whether discretisation strategies completely unrelated to either significant digits or leading CF coefficients (e.g., uniform binnings of $\{\log_b X\}$) can work better than both in hypothesis tests.

Competing interest. The authors declare no conflict of interest.

References

[1] Abramowitz, M. & Stegun, I.A. (1972). *Handbook of Mathematical Functions With formulas, graphs, and Mathematical tables*, New York: U.S. Government Printing Office.
[2] Apostol, T.M. (1976). *Introduction to Analytic Number theory*, New York, USA: Springer-Verlag.
[3] Arfken, G.B. & Weber, H.J. (2005). 6th ed. *Mathematical Methods for physicists*, Elsevier.
[4] Balado, F. & Silvestre, G.C. (2021). Benford's law: Hammering a square peg into a round hole? In *29th European Signal Processing Conference (EUSIPCO)*, Dublin, Ireland, EURASIP, pp. 796–800.
[5] Barabesi, L., Cerasa, A., Cerioli, A. & Perrotta, D. (2022). *On characterizations and tests of Benford's law*. *Journal of the American Statistical Association* 117(540): 1887–1903.
[6] Barabesi, L. & Pratelli, L. (2020). On the generalized Benford law. *Statistics & Probability Letters* 160: 108702.
[7] Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society* 78(4): 551–572.
[8] Berger, A. & Hill, T.P. (2011). A basic theory of Benford's law. *Probability Surveys* 8: 1–126.
[9] Berger, A. & Hill, T.P. (2011). Benford's law strikes back: no simple explanation in sight for mathematical gem. *Math Intelligencer* 33: 85–91.

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press

[10] Berger, A., Hill, T.P. & Rogers, E. (2009). Benford online bibliography. last accessed July 2023, http://www.benfordonline.net.
[11] Blachman, N. (1984). The continued fraction as an information source. *IEEE Transactions on Information Theory* 30(4): 671–674.
[12] Fassett, C.I., Kadish, S.J., Head, J.W., Solomon, S.C. & Strom, R.G. (2011). The global population of large craters on Mercury and comparison with the Moon. *Geophysics Research Letters* 38(10).
[13] Fu, D., Shi, Y.Q. & Su, W. (2007). A generalized Benford's law for JPEG coefficients and its applications in image forensics. In *Proceedings of SPIE: Security, Steganography and Watermarking of Multimedia Contents IX*, San José, USA. Vol. 6505.
[14] Graham, R.L., Knuth, D.E. & Patashnik, O. (1994). *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. Harlow, UK: Addison-Wesley.
[15] Grendar, M., Judge, G. & Schechter, L. (2007). An empirical non-parametric likelihood family of data-based Benford-like distributions. *Physica A: Statistical Mechanics and its Applications* 380: 429–438.
[16] Hill, T.P. (1995). A statistical derivation of the significant-digit law. *Statistical Science* 10(4): 354–363.
[17] Katz, J. & Lindell, Y. (2021). *Introduction to Modern cryptography*, 3rd ed. Boca Ratón, USA: CRC Press.
[18] Khinchin, A.Y. (1961). *Continued fractions*, 3rd ed. Chicago, IL: The University of Chicago Press.
[19] Kozubowski, T.J., Panorska, A.K. & Forister, M.L. (2015). A discrete truncated Pareto distribution. *Statistical Methodology* 26: 135–150.
[20] Kuz'min, R.O. (1928). *Sur un problème de Gauss, Atti del Congresso Internazionale dei Matematici*, Bologna, Italy, pp. 83–89.
[21] S.J. Miller, ed. (2015). *Benford's law: Theory and applications*. Princeton, NJ: Princeton University Press
[22] Miller, S.J. & Takloo-Bighash, R. (2006). *An Invitation to Modern Number theory*. Princeton, NJ: Princeton University Press.
[23] Nair, J., Wierman, A. & Zwart, B. (2022). *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. Cambridge, UK: Cambridge University Press.
[24] Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics* 4(1): 39–40.
[25] Nigrini, M.J. (1992). *The detection of income tax evasion through an analysis of digital frequencies*, Ph.D. Thesis, Ohio, USA: University of Cincinnati.
[26] Nigrini, M.J. (2012). *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. New Jersey, USA: Wiley & Sons.
[27] Nigrini, M.J. & Miller, S.J. (2007). Benford's law applied to hydrology data—results and relevance to other geophysical data. *Mathematical Geology* 39(5): 469–490.
[28] Pérez-González, F., Heileman, G.L. & Abdallah, C.T. (2007). A generalization of Benford's law and its application to images. In *European Control Conference (ECC)*, Kos, Greece, EURASIP, pp. 3613–3619.
[29] Pietronero, L., Tosatti, E., Tosatti, V. & Vespignani, A. (2001). Explaining the uneven distribution of numbers in nature: the laws of Benford and Zipf. *Physica A: Statistical Mechanics and its Applications* 293(1): 297–304.
[30] Pinkham, R. (1961). On the distribution of first significant digits. *Annals of Mathematical Statistics* 32(4): 1223–1230.
[31] Robbins, S.J. (2018). A new global database of Lunar impact craters >1–2 km: 1. Crater locations and sizes, comparisons with published databases, and global analysis. *Journal of Geophysical Research: Planets* 124(4): 871–892.
[32] Rodriguez, R. (2004). First significant digit patterns from mixtures of uniform digits. *American Statistician* 58(1): 64–71.
[33] Tseng, H., Huang, W. & Huang, D. (2017). Modified Benford's law for two-exponent distributions. *Scientometrics* 110(3): 1403–1413.

Cite this article: Balado F. and Silvestre GCM. (2024). General distributions of number representation elements. *Probability in the Engineering and Informational Sciences* 38(3): 594–616. https://doi.org/10.1017/S0269964823000207

https://doi.org/10.1017/S0269964823000207 Published online by Cambridge University Press