Probability Surveys
Vol. 8 (2011) 1–126
ISSN: 1549-5787
DOI: 10.1214/11-PS175

A basic theory of Benford's Law*

Arno Berger[†,‡]

Mathematical and Statistical Sciences
University of Alberta
Edmonton, Alberta T6G 2G1, Canada
e-mail: aberger@math.ualberta.ca

and

Theodore P. Hill[‡]

School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332–0160, USA
e-mail: hill@math.gatech.edu

Abstract: Drawing from a large, diverse body of work, this survey presents a comprehensive and unified introduction to the mathematics underlying the prevalent logarithmic distribution of significant digits and significands, often referred to as *Benford's Law* (BL) or, in a special case, as the *First Digit Law*. The invariance properties that characterize BL are developed in detail. Special attention is given to the emergence of BL in a wide variety of deterministic and random processes. Though mainly expository in nature, the article also provides strengthened versions of, and simplified proofs for, many key results in the literature. Numerous intriguing problems for future research arise naturally.

AMS 2000 subject classifications: Primary 60-01, 11K06, 37M10, 39A60; secondary 37A45, 60F15, 62E10.
Keywords and phrases: Benford's Law, significant digits, uniform distribution mod 1, scale-invariance, base-invariance, sum-invariance, shadowing, difference equation, random probability measure, mixture of distributions.

Received June 2011.

Contents

---
*This is an original survey paper.
†Supported by an NSERC Discovery Grant.
‡The authors are grateful to S. Evans for pointing them towards a proof of Theorem 6.6.

1. Introduction

Benford's Law, or *BL* for short, is the observation that in many collections of numbers, be they e.g. mathematical tables, real-life data, or combinations thereof, the leading significant digits are not uniformly distributed, as might be expected, but are heavily skewed toward the smaller digits. More specifically, BL says that the significant digits in many datasets follow a very particular logarithmic distribution. In its most common formulation, namely the special case of first significant *decimal* (i.e. base-10) digits, BL is also known as the *First-Digit Phenomenon* and reads

$$\text{Prob}\,(D_1 = d_1) = \log_{10}\left(1 + d_1^{-1}\right) \quad \text{for all } d_1 = 1, 2, \ldots, 9\,; \qquad (1.1)$$

here D_1 denotes the first significant decimal digit, e.g.

$$D_1(\sqrt{2}) = D_1(1.414\ldots) = 1\,,$$
$$D_1(\pi^{-1}) = D_1(0.3183\ldots) = 3\,,$$
$$D_1(e^{\pi}) = D_1(23.14\ldots) = 2\,.$$

Thus, for example, (1.1) asserts that

$$\text{Prob}\,(D_1 = 1) = \log_{10} 2 = 0.3010\ldots\,, \quad \text{Prob}\,(D_1 = 2) = \log_{10}\frac{3}{2} = 0.1760\ldots\,,$$

hence the two smallest digits occur with a combined probability close to 50 percent, whereas the two largest digits together have a probability of less than 10 percent,

$$\text{Prob}\,(D_1 = 8) = \log_{10}\frac{9}{8} = 0.05115\ldots\,, \quad \text{Prob}\,(D_1 = 9) = \log_{10}\frac{10}{9} = 0.04575\ldots\,.$$

A crucial part of the content of (1.1), of course, is an appropriate formulation or interpretation of "Prob". In practice, this can take several forms. For sequences of real numbers (x_n), for example, Prob usually refers to the proportion (or relative frequency) of times n for which an event such as $D_1 = 1$ occurs. Thus $\mathrm{Prob}\,(D_1 = 1)$ is the limiting proportion, as $N \to \infty$, of times $n \le N$ that the first significant digit of x_n equals 1. Implicit in this usage of Prob is the assumption that all limiting proportions of interest actually exist. Similarly, for real-valued functions $f : [0, +\infty) \to \mathbb{R}$, $\mathrm{Prob}\,(D_1 = 1)$ refers to the limiting proportion, as $T \to \infty$, of the total length of time $t < T$ for which the first significant digit of $f(t)$ is 1. For a random variable or probability distribution, on the other hand, Prob simply denotes the underlying probability, e.g. if X a random variable then $\mathrm{Prob}\,(D_1(X) = 1)$ is the probability that the first significant digit of X equals 1. Finite datasets of real numbers can also be dealt with this way, with Prob being the empirical distribution of the dataset.

All of these approaches to (1.1) will be studied in detail in subsequent chapters. Fig 1 illustrates several of the possible settings, including simple sequences such as the Fibonacci numbers $(F_n) = (1, 1, 2, 3, 5, 8, 13, \ldots)$, and real-life data from [Ben] as well as recent census statistics; in addition, it previews some of the many scenarios, also to be discussed later, that lead to *exact* conformance with BL. In Fig 1 and throughout, $\#A$ denotes the cardinality (number of elements) of the finite set A.

In a form more complete than (1.1), BL is a statement about the joint distribution of *all* decimal digits: For every positive integer m,

$$\mathrm{Prob}\,\big((D_1, D_2, \ldots, D_m) = (d_1, d_2, \ldots, d_m)\big) = \log_{10}\left(1 + \Big(\sum\nolimits_{j=1}^{m} 10^{m-j} d_j\Big)^{-1}\right) \tag{1.2}$$

holds for all m-tuples $(d_1, d_2, \ldots, d_m)$, where d_1 is an integer in $\{1, 2, \ldots, 9\}$ and for $j \ge 2$, d_j is an integer in $\{0, 1, \ldots, 9\}$; here D_2, D_3, D_4 etc. represent the second, third, forth etc. significant decimal digit, e.g.

$$D_2(\sqrt{2}) = 4\,, \quad D_3(\pi^{-1}) = 8\,, \quad D_4(e^{\pi}) = 4\,.$$

Thus, for example, (1.2) implies that

$$\mathrm{Prob}\,\big((D_1, D_2, D_3) = (3, 1, 4)\big) = \log_{10}\frac{315}{314} = 0.001380\ldots\,.$$

A perhaps surprising corollary of the general form of BL is that the significant digits are *dependent*, and not independent as one might expect [Hi2]. Indeed, from (1.2) it follows for instance that the (unconditional) probability that the second digit equals 1 is

$$\mathrm{Prob}\,(D_2 = 1) = \sum\nolimits_{j=1}^{9} \log_{10}\left(1 + \frac{1}{10j + 1}\right) = \log_{10}\frac{6029312}{4638501} = 0.1138\ldots\,,$$

whereas, given that the first digit equals 1, the (conditional) probability that the second digit equals 1 as well is

$$\mathrm{Prob}\,(D_2 = 1 | D_1 = 1) = \frac{\log_{10} 12 - \log_{10} 11}{\log_{10} 2} = 0.1255\ldots\,.$$

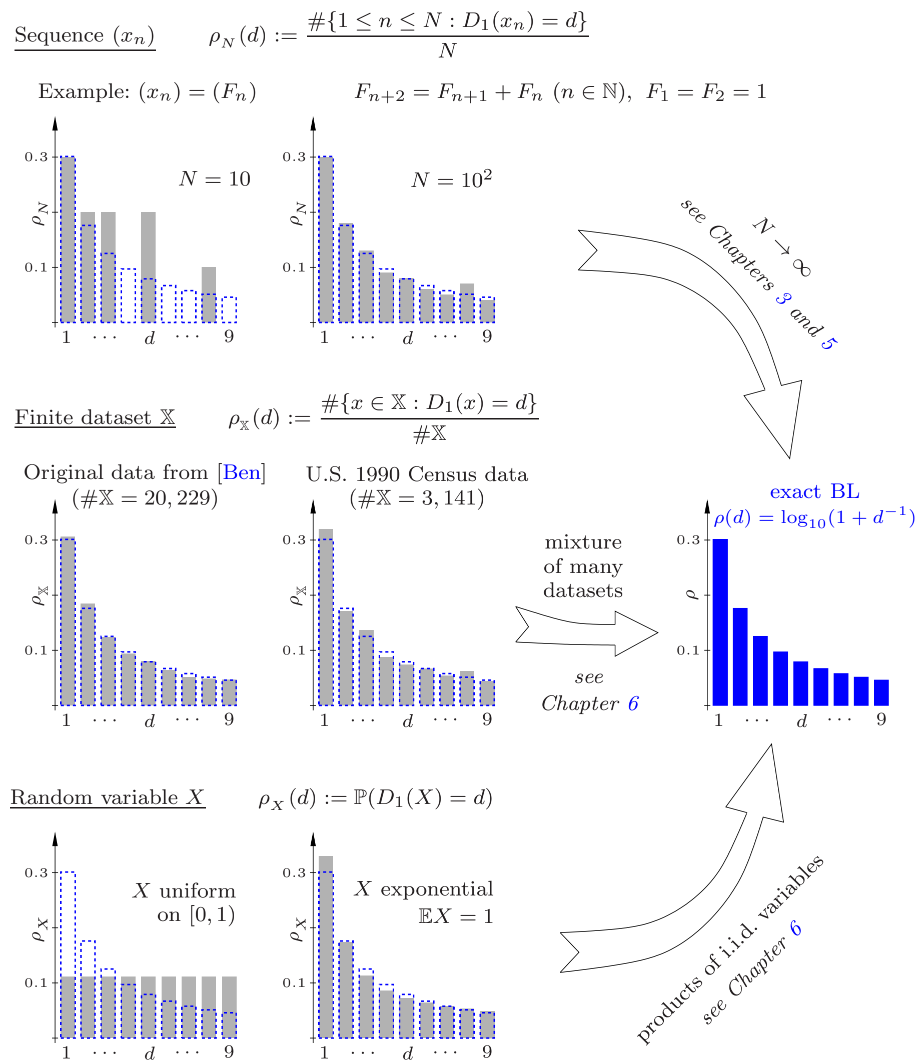

Fig 1. *Different interpretations of (1.1) for sequences, datasets, and random variables, respectively, and scenarios that may lead to exact conformance with BL.*

This dependence among significant digits decreases rapidly, in fact exponentially, as the distance between the digits increases. For example, it follows easily from (1.2) that

$$\text{Prob}\,(D_m = 1 | D_1 = 1) = \text{Prob}\,(D_m = 1) + \mathcal{O}(10^{-m}) \quad \text{as } m \to \infty\,.$$

(Here and throughout, the order symbol $\mathcal{O}$ is used as usual: If (a_n) and (b_n) are sequences of real numbers then $a_n = \mathcal{O}(b_n)$ as $n \to \infty$ simply means that $|a_n| \le c|b_n|$ for all n, with some constant $c > 0$.)

A related consequence of (1.2) is that the distribution of the m-th significant digit approaches the uniform distribution on $\{0, 1, \ldots, 9\}$ exponentially fast also, e.g.

$$\text{Prob}\,(D_m = 1) = \frac{1}{10} + \frac{63}{20\ln 10}10^{-m} + \mathcal{O}(10^{-2m}) \quad \text{as } m \to \infty\,.$$

Apparently first discovered by polymath S. Newcomb [Ne] in the 1880's, (1.1) and (1.2) were rediscovered by physicist F. Benford [Ben] and, Newcomb's article having been forgotten at the time, came to be known as *Benford's* Law. Today, BL appears in a broad spectrum of mathematics, ranging from differential equations to number theory to statistics. Simultaneously, the applications of BL are mushrooming — from diagnostic tests for mathematical models in biology and finance to fraud detection. For instance, the U.S. Internal Revenue Service uses BL to ferret out suspicious tax returns, political scientists use it to identify voter fraud, and engineers to detect altered digital images. As R. Raimi already observed some 35 years ago [Ra1, p.512], "This particular logarithmic distribution of the first digits, while not universal, is so common and yet so surprising at first glance that it has given rise to a varied literature, among the authors of which are mathematicians, statisticians, economists, engineers, physicists, and amateurs." At the time of writing, the online database [BH2] contains more than 600 articles on the subject.

It is the purpose of this article to explain the basic terminology, mathematical concepts and results concerning BL in an elementary and accessible manner. Having read this survey, the reader will find it all the more enjoyable to browse the multifarious literature where a wide range of extensions and refinements as well as applications are discussed.

Note. Throughout this overview of the basic theory of BL, attention will more or less exclusively be restricted to significant *decimal* (i.e. base-10) digits. From now on, therefore, $\log x$ will always denote the logarithm base 10 of x, while $\ln x$ is the natural logarithm of x. For convenience, the convention $\log 0 := 0$ will be adopted. All results stated here only with respect to base 10 carry over easily to arbitrary integer bases $b \geq 2$, and the interested reader may find some pertinent details e.g. in [BBH]. The general form of (1.2) with respect to any such base b is

$$\text{Prob}\left(\left(D_1^{(b)}, D_2^{(b)}, \ldots, D_m^{(b)}\right) = (d_1, d_2, \ldots, d_m)\right) = \log_b\left(1 + \left(\sum\nolimits_{j=1}^{m} b^{m-j} d_j\right)^{-1}\right), \tag{1.3}$$

where $\log_b$ denotes the base-b logarithm and $D_1^{(b)}$, $D_2^{(b)}$, $D_3^{(b)}$ etc. are, respectively, the first, second, third etc. significant digits base b; in particular, therefore, d_1 is an integer in $\{1, 2, \ldots, b-1\}$, and for $j \geq 2$, d_j is an integer in $\{0, 1, \ldots, b-1\}$. Note that in the case $m = 1$ and $b = 2$, (1.3) reduces to $\text{Prob}\left(D_1^{(2)} = 1\right) = 1$, which trivially is true because the first significant digit base 2 of every non-zero number equals 1. ♣

2. Significant digits and the significand

Benford's Law is a statement about the statistical distribution of significant (decimal) digits or, equivalently, about significands viz. fraction parts in floating-point arithmetic. Thus a natural starting point for any study of BL is the formal definition of *significant digits* and the *significand* (*function*).

2.1. Significant digits

Definition 2.1. For every non-zero real number x, the *first significant decimal digit* of x, denoted by $D_1(x)$, is the unique integer $j \in \{1, 2, \ldots, 9\}$ satisfying $10^k j \le |x| < 10^k(j+1)$ for some (necessarily unique) $k \in \mathbb{Z}$.

Similarly, for every $m \ge 2$, $m \in \mathbb{N}$, the *m-th significant decimal digit* of x, denoted by $D_m(x)$, is defined inductively as the unique integer $j \in \{0, 1, \ldots, 9\}$ such that

$$10^k \left(\sum\nolimits_{i=1}^{m-1} D_i(x) 10^{m-i} + j \right) \le |x| < 10^k \left(\sum\nolimits_{i=1}^{m-1} D_i(x) 10^{m-i} + j + 1 \right)$$

for some (necessarily unique) $k \in \mathbb{Z}$; for convenience, $D_m(0) := 0$ for all $m \in \mathbb{N}$.

Note that, by definition, the first significant digit $D_1(x)$ of $x \neq 0$ is never zero, whereas the second, third, etc. significant digits may be any integers in $\{0, 1, \ldots, 9\}$.

Example 2.2.

$$D_1(\sqrt{2}) = D_1(-\sqrt{2}) = D_1(10\sqrt{2}) = 1\,, \quad D_2(\sqrt{2}) = 4\,, \quad D_3(\sqrt{2}) = 1\,;$$
$$D_1(\pi^{-1}) = D_1(10\pi^{-1}) = 3\,, \quad D_2(\pi^{-1}) = 1\,, \quad D_3(\pi^{-1}) = 8\,.$$

2.2. The significand

The *significand* of a real number is its coefficient when it is expressed in floating-point ("scientific notation") form, more precisely

Definition 2.3. The *(decimal) significand function* $S : \mathbb{R} \to [1, 10)$ is defined as follows: If $x \neq 0$ then $S(x) = t$, where t is the unique number in $[1, 10)$ with $|x| = 10^k t$ for some (necessarily unique) $k \in \mathbb{Z}$; if $x = 0$ then, for convenience, $S(0) := 0$.

Observe that, for all $x \in \mathbb{R}$,

$$S(10^k x) = S(x) \quad \text{for every } k \in \mathbb{Z}\,,$$

and also $S\big(S(x)\big) = S(x)$. Explicitly, S is given by

$$S(x) = 10^{\log|x| - \lfloor \log|x| \rfloor} \quad \text{for all } x \neq 0\,;$$

here $\lfloor t \rfloor$ denotes, for any real number t, the largest integer less than or equal to t. (The function $t \mapsto \lfloor t \rfloor$ is often referred to as the "floor function".)

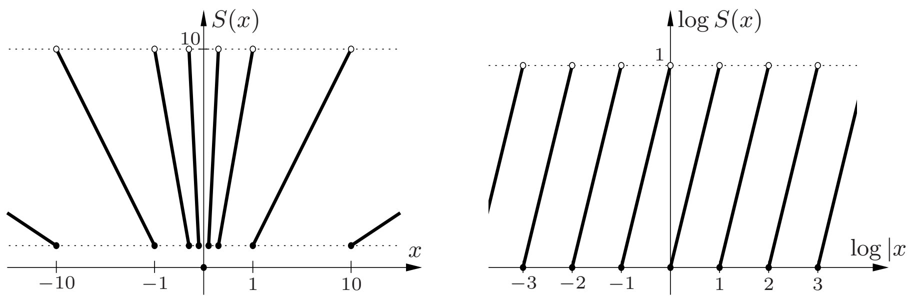

FIG 2. *Graphing the (decimal) significand function S.*

Note. The original word used in American English to describe the coefficient of floating-point numbers in computer hardware seems to have been *mantissa*, and this usage remains common in computing and among computer scientists. However, this use of the word mantissa is discouraged by the IEEE floating-point standard committee and by some professionals such as W. Kahan and D. Knuth because it conflicts with the pre-existing usage of mantissa for the fractional part of a logarithm. In accordance with the IEEE standard, only the term *significand* will be used henceforth. (With the significand as in Definition 2.3, the (traditional) mantissa would simply be $\log S$.) The reader should also note that in some places in the literature, the significand is taken to have values in $[0.1, 1)$ rather than in $[1, 10)$. ♣

Example 2.4.

$$S(\sqrt{2}) = S(10\sqrt{2}) = \sqrt{2} = 1.414\ldots,$$

$$S(\pi^{-1}) = S(10\pi^{-1}) = 10\pi^{-1} = 3.183\ldots.$$

The significand uniquely determines the significant digits, and vice versa. This relationship is recorded in the following proposition which immediately follows from Definitions 2.1 and 2.3.

Proposition 2.5. *For every real number x:*

(i) $S(x) = \sum_{m\in\mathbb{N}} 10^{1-m} D_m(x)$;

(ii) $D_m(x) = \lfloor 10^{m-1} S(x) \rfloor - 10\lfloor 10^{m-2} S(x) \rfloor$ *for every* $m \in \mathbb{N}$.

Thus, Proposition 2.5(i) expresses the significand of a number as an explicit function of the significant digits of that number, and (ii) expresses the significant digits as a function of the significand.

It is important to note that the definition of significand and significant digits per se does not involve any decimal *expansion* of x. However, it is clear from Proposition 2.5(i) that the significant digits provide a decimal expansion of $S(x)$, and that a sequence (d_m) in $\{0, 1, \ldots, 9\}$ is the sequence of significant digits of some positive real number if and only if $d_1 \neq 0$ and $d_m \neq 9$ for infinitely many m.

Example 2.6. It follows from Proposition 2.5, together with Examples 2.2 and 2.4, that

$$S(\sqrt{2}) = D_1(\sqrt{2}) + 10^{-1}D_2(\sqrt{2}) + 10^{-2}D_3(\sqrt{2}) + \ldots = 1.414\ldots = \sqrt{2}\,,$$

as well as

$$\begin{aligned} D_1(\sqrt{2}) &= \lfloor \sqrt{2} \rfloor = 1\,, \\ D_2(\sqrt{2}) &= \lfloor 10\sqrt{2} \rfloor - 10\lfloor \sqrt{2} \rfloor = 4\,, \\ D_3(\sqrt{2}) &= \lfloor 100\sqrt{2} \rfloor - 10\lfloor 10\sqrt{2} \rfloor = 1\,, \text{ etc}\,. \end{aligned}$$

As the significant digits determine the significand, and are in turn determined by it, the informal version (1.2) of BL in the Introduction has an immediate and very concise counterpart in terms of the significand function, namely

$$\text{Prob}\,(S \le t) = \log t \quad \text{for all } 1 \le t < 10\,. \tag{2.1}$$

(Recall that log denotes the base-10 logarithm throughout.) As noted earlier, the formal versions of (1.2) and (2.1) will be developed in detail below.

2.3. The significand σ-algebra

The informal statements (1.1), (1.2) and (2.1) of BL involve *probabilities*. Hence to formulate mathematically precise versions of these statements, it is necessary to re-formulate them in the setting of rigorous probability theory.

The fundamental concept of standard modern probability theory is that of a *probability space* $(\Omega, \mathcal{A}, \mathbb{P})$; here Ω, $\mathcal{A}$ and $\mathbb{P}$ are, respectively, a non-empty set, a σ-algebra on Ω, and a probability measure on $(\Omega, \mathcal{A})$. Recall that a *σ-algebra* $\mathcal{A}$ on Ω is simply a family of subsets of Ω such that $\varnothing \in \mathcal{A}$, and $\mathcal{A}$ is closed under taking complements and countable unions, that is,

$$A \in \mathcal{A} \implies A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}\,,$$

as well as

$$A_n \in \mathcal{A} \text{ for all } n \in \mathbb{N} \implies \bigcup\nolimits_{n\in\mathbb{N}} A_n \in \mathcal{A}\,.$$

Given any collection $\mathcal{E}$ of subsets of Ω, there exists a (unique) smallest σ-algebra on Ω containing $\mathcal{E}$, referred to as the σ-algebra *generated by* $\mathcal{E}$ and denoted by $\sigma(\mathcal{E})$. Perhaps the most important example is the so-called *Borel* σ-algebra $\mathcal{B}$ on $\mathbb{R}$: By definition, $\mathcal{B}$ is the σ-algebra generated by all intervals. If $C \subset \mathbb{R}$ then $\mathcal{B}(C)$ is understood to be the σ-algebra $C \cap \mathcal{B} := \{C \cap B : B \in \mathcal{B}\}$ on C; for brevity, write $\mathcal{B}[a,b)$ instead of $\mathcal{B}\big([a,b)\big)$ and $\mathcal{B}^+$ instead of $\mathcal{B}(\mathbb{R}^+)$, where $\mathbb{R}^+ = \{t \in \mathbb{R} : t > 0\}$.

In general, given any function $f : \Omega \to \mathbb{R}$, recall that, for every $C \subset \mathbb{R}$, the set $f^{-1}(C) \subset \Omega$, called the *pre-image* of C under f, is defined as

$$f^{-1}(C) = \{\omega \in \Omega : f(\omega) \in C\}\,.$$

The σ-algebra on Ω generated by

$$\mathcal{E} = \{f^{-1}(I) : I \subset \mathbb{R} \text{ an interval}\}$$

is also referred to as the σ-algebra generated by f; it will be denoted by $\sigma(f)$. Thus $\sigma(f)$ is the smallest σ-algebra on Ω that contains all sets of the form $\{\omega \in \Omega : a \le f(\omega) \le b\}$, for every $a, b \in \mathbb{R}$. It is easy to check that in fact $\sigma(f) = \{f^{-1}(B) : B \in \mathcal{B}\}$. Similarly, a whole family $\mathcal{F}$ of functions $f : \Omega \to \mathbb{R}$ may be considered, and

$$\sigma(\mathcal{F}) := \sigma\left(\bigcup\nolimits_{f\in\mathcal{F}} \sigma(f)\right) = \sigma\bigl(f^{-1}(I) : I \subset \mathbb{R} \text{ an interval, } f \in \mathcal{F}\bigr)$$

is then simply the smallest σ-algebra on Ω containing all sets $\{\omega \in \Omega : a \le f(\omega) \le b\}$ for all $a, b \in \mathbb{R}$ and all $f \in \mathcal{F}$.

In probability theory, the elements of a σ-algebra $\mathcal{A}$ on Ω are often referred to as *events*, and functions $f : \Omega \to \mathbb{R}$ with $\sigma(f) \subset \mathcal{A}$ are called *random variables*. Probability textbooks typically use symbols X, Y etc., rather than f, g etc., to denote random variables, and this practice will be adhered to here also. Thus, for example, for a Bernoulli random variable X on $(\mathbb{R}, \mathcal{B})$ taking only the values 0 and 1, $\sigma(X)$ is the sub-σ-algebra of $\mathcal{B}$ given by

$$\sigma(X) = \bigl\{\varnothing, \{0\}, \{1\}, \{0,1\}, \mathbb{R}, \mathbb{R}\backslash\{0\}, \mathbb{R}\backslash\{1\}, \mathbb{R}\backslash\{0,1\}\bigr\};$$

here, and throughout, $A\backslash B = A \cap B^c$ is the set of all elements of A that are not in B.

As the third ingredient in the concept of a probability space, a *probability measure* on $(\Omega, \mathcal{A})$ is a function $\mathbb{P} : \mathcal{A} \to [0,1]$ such that $\mathbb{P}(\varnothing) = 0$, $\mathbb{P}(\Omega) = 1$, and

$$\mathbb{P}\left(\bigcup\nolimits_{n\in\mathbb{N}} A_n\right) = \sum\nolimits_{n\in\mathbb{N}} \mathbb{P}(A_n)$$

holds whenever the sets $A_n \in \mathcal{A}$ are disjoint. The obvious probabilistic interpretation of $\mathbb{P}$ is that, for every $A \in \mathcal{A}$, the number $\mathbb{P}(A) \in [0,1]$ is the probability that the event $\{\omega \in A\}$ occurs. Two of the most important examples of probability measures are the *discrete uniform distribution* on a non-empty finite set A, where the probability of any set $B \subset A$ is simply

$$\frac{\#(B \cap A)}{\#A},$$

and its continuous counterpart the *uniform distribution* $\lambda_{a,b}$ with $a < b$, more technically referred to as (*normalized*) *Lebesgue measure* on $[a, b)$, or more precisely on $\bigl([a,b), \mathcal{B}[a,b)\bigr)$, given by

$$\lambda_{a,b}\bigl([c,d]\bigr) := \frac{d-c}{b-a} \quad \text{for every } [c,d] \subset [a,b). \tag{2.2}$$

In advanced analysis courses, it is shown that (2.2) does indeed entail a unique, consistent definition of $\lambda_{a,b}(B)$ for *every* $B \in \mathcal{B}[a,b)$; in particular $\lambda_{a,b}\bigl([a,b)\bigr) =$

1. Another example of a probability measure, on any $(\Omega, \mathcal{A})$, is the *Dirac measure* (or *point mass*) concentrated at some $\omega \in \Omega$, symbolized by δ_ω. In this case, $\delta_\omega(A) = 1$ if $\omega \in A$, and $\delta_\omega(A) = 0$ otherwise. Throughout, unspecified probability measures on $(\Omega, \mathcal{A})$ with $\Omega \subset \mathbb{R}$ and $\mathcal{A} \subset \mathcal{B}$ will typically be denoted by capital Roman letters P, Q etc.

In view of the above, the key step in formulating BL precisely is identifying the appropriate probability space, and hence in particular the correct σ-algebra. As it turns out, in the significant digit framework there is only one natural candidate which, although different from $\mathcal{B}$, is nevertheless both intuitive and easy to describe.

Definition 2.7. The *significand σ-algebra* $\mathcal{S}$ is the σ-algebra on $\mathbb{R}^+$ generated by the significand function S, i.e. $\mathcal{S} = \mathbb{R}^+ \cap \sigma(S)$.

The importance of the σ-algebra $\mathcal{S}$ comes from the fact that for every event $A \in \mathcal{S}$ and every $x > 0$, knowing $S(x)$ is enough to decide whether $x \in A$ or $x \notin A$. Worded slightly more formally, this observation reads as follows.

Lemma 2.8. *For every function $f : \mathbb{R}^+ \to \mathbb{R}$ the following statements are equivalent:*

(i) *f can be described completely in terms of S, that is, $f(x) = \varphi\bigl(S(x)\bigr)$ holds for all $x \in \mathbb{R}^+$, with some function $\varphi : [1, 10) \to \mathbb{R}$ satisfying $\sigma(\varphi) \subset \mathcal{B}[1, 10)$.*

(ii) *$\sigma(f) \subset \mathcal{S}$.*

Proof. First assume (i) and let $I \subset \mathbb{R}$ be any interval. Then $B = \varphi^{-1}(I) \in \mathcal{B}$ and $f^{-1}(I) = S^{-1}\bigl(\varphi^{-1}(I)\bigr) = S^{-1}(B) \in \mathcal{S}$, showing that $\sigma(f) \subset \mathcal{S}$.

Conversely, if $\sigma(f) \subset \mathcal{S}$ then $f(10x) = f(x)$ holds for all $x > 0$. Indeed, assuming by way of contradiction that, say, $f(x_0) < f(10x_0)$ for some $x_0 > 0$, let

$$A := f^{-1}\left(\left[f(x_0) - 1, \frac{f(x_0) + f(10x_0)}{2}\right]\right) \in \sigma(f) \subset \mathcal{S}$$

and note that $x_0 \in A$ while $10x_0 \notin A$. Since $A = S^{-1}(B)$ for some $B \in \mathcal{B}$, this leads to the contradiction that $S(x_0) \in B$ and $S(x_0) = S(10x_0) \notin B$. Hence $f(10x) = f(x)$ for all $x > 0$, and by induction also $f(10^k x) = f(x)$ for all $k \in \mathbb{Z}$. Given $x \in [1, 10)$, pick any $y > 0$ with $S(y) = x$ and define $\varphi(x) := f(y)$. Since any two choices of y differ by a factor 10^k for some $k \in \mathbb{Z}$, $\varphi : [1, 10) \to \mathbb{R}$ is well-defined, and $\varphi\bigl(S(y)\bigr) = f(y)$ holds for all $y > 0$. Moreover, for any interval $I \subset \mathbb{R}$ and $x > 0$, $\varphi(x) \in I$ holds if and only if $x \in \bigcup_{k\in\mathbb{Z}} 10^k f^{-1}(I)$. By assumption, the latter set belongs to $\mathcal{S}$, which in turn shows that $\sigma(\varphi) \subset \mathcal{B}[1, 10)$. □

Informally put, Lemma 2.8 states that the significand σ-algebra $\mathcal{S}$ is the family of all events $A \subset \mathbb{R}^+$ that can be described completely in terms of their significands, or equivalently (by Theorem 2.9 below) in terms of their significant digits. For example, the set A_1 of positive numbers whose first significant digit is 1 and whose third significant digit is not 7, i.e.

$$A_1 = \{x > 0 : D_1(x) = 1, D_3(x) \neq 7\},$$

belongs to $\mathcal{S}$, as does the set A_2 of all $x > 0$ whose significant digits are all 5 or 6, i.e.

$$A_2 = \{x > 0 : D_m(x) \in \{5, 6\} \text{ for all } m \in \mathbb{N}\},$$

or the set A_3 of numbers whose significand is rational,

$$A_3 = \{x > 0 : S(x) \in \mathbb{Q}\}.$$

On the other hand, the interval $[1, 2]$, for instance, does not belong to $\mathcal{S}$. This follows from the next theorem which provides a useful characterization of the significand sets, i.e. the members of the family $\mathcal{S}$. For its formulation, for every $t \in \mathbb{R}$ and every set $C \subset \mathbb{R}$, let $tC := \{tc : c \in C\}$.

Theorem 2.9 ([Hi2])**.** *For every $A \in \mathcal{S}$,*

$$A = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k S(A) \tag{2.3}$$

holds, where $S(A) = \{S(x) : x \in A\} \subset [1, 10)$. Moreover,

$$\mathcal{S} = \mathbb{R}^+ \cap \sigma(D_1, D_2, D_3, \ldots) = \Big\{\bigcup\nolimits_{k\in\mathbb{Z}} 10^k B : B \in \mathcal{B}[1, 10)\Big\}. \tag{2.4}$$

Proof. By definition,

$$\mathcal{S} = \mathbb{R}^+ \cap \sigma(S) = \mathbb{R}^+ \cap \{S^{-1}(B) : B \in \mathcal{B}\} = \mathbb{R}^+ \cap \{S^{-1}(B) : B \in \mathcal{B}[1, 10)\}.$$

Thus, given any $A \in \mathcal{S}$, there exists a set $B \in \mathcal{B}[1, 10)$ with $A = \mathbb{R}^+ \cap S^{-1}(B) = \bigcup_{k\in\mathbb{Z}} 10^k B$. Since $S(A) = B$, it follows that (2.3) holds for all $A \in \mathcal{S}$.

To prove (2.4), first observe that by Proposition 2.5(i) the significand function S is completely determined by the significant digits $D_1, D_2, D_3, \ldots$, so $\sigma(S) \subset \sigma(D_1, D_2, D_3, \ldots)$ and hence $\mathcal{S} \subset \mathbb{R}^+ \cap \sigma(D_1, D_2, D_3, \ldots)$. Conversely, according to Proposition 2.5(ii), every D_m is determined by S, thus $\sigma(D_m) \subset \sigma(S)$ for all $m \in \mathbb{N}$, showing that $\sigma(D_1, D_2, D_3, \ldots) \subset \sigma(S)$ as well. To verify the remaining equality in (2.4), note that for every $A \in \mathcal{S}$, $S(A) \in \mathcal{B}[1, 10)$ and hence $A = \bigcup_{k\in\mathbb{Z}} 10^k B$ for $B = S(A)$, by (2.3). Conversely, every set of the form $\bigcup_{k\in\mathbb{Z}} 10^k B = \mathbb{R}^+ \cap S^{-1}(B)$ with $B \in \mathcal{B}[1, 10)$ obviously belongs to $\mathcal{S}$. $\square$

Note that for every $A \in \mathcal{S}$ there is a *unique* $B \in \mathcal{B}[1, 10)$ such that $A = \bigcup_{k\in\mathbb{Z}} 10^k B$, and (2.3) shows that in fact $B = S(A)$.

Example 2.10. The set A_4 of positive numbers with

$$A_4 = \{10^k : k \in \mathbb{Z}\} = \{\ldots, 0.01, 0.1, 1, 10, 100, \ldots\}$$

belongs to $\mathcal{S}$. This can be seen either by observing that A_4 is the set of positive reals with significand exactly equal to 1, i.e. $A_4 = \mathbb{R}^+ \cap S^{-1}(\{1\})$, or by noting that $A_4 = \{x > 0 : D_1(x) = 1, D_m(x) = 0 \text{ for all } m \geq 2\}$, or by using (2.4) and the fact that $A_4 = \bigcup_{k\in\mathbb{Z}} 10^k\{1\}$ and $\{1\} \in \mathcal{B}[1, 10)$.

Example 2.11. The singleton set $\{1\}$ and the interval $[1,2]$ do not belong to $\mathcal{S}$, since the number 1 cannot be distinguished from the number 10, for instance, using only significant digits. Nor can the interval $[1,2]$ be distinguished from $[10,20]$. Formally, neither of these sets is of the form $\bigcup_{k\in\mathbb{Z}} 10^k B$ for any $B \in \mathcal{B}[1,10)$.

Although the significand function and σ-algebra above were defined in the setting of *real* numbers, the same concepts carry over immediately to the most fundamental setting of all, the set of *positive integers*. In this case, the induced σ-algebra is interesting in its own right.

Example 2.12. The restriction $\mathcal{S}_{\mathbb{N}}$ of $\mathcal{S}$ to subsets of $\mathbb{N}$, i.e. $\mathcal{S}_{\mathbb{N}} = \{\mathbb{N}\cap A : A \in \mathcal{S}\}$ is a σ-algebra on $\mathbb{N}$. A characterization of $\mathcal{S}_{\mathbb{N}}$ analogous to that of $\mathcal{S}$ given in Theorem 2.9 is as follows: Denote by $\mathbb{N}_{\cancel{10}}$ the set of all positive integers not divisible by 10, i.e. $\mathbb{N}_{\cancel{10}} = \mathbb{N}\backslash 10\mathbb{N}$. Then

$$\mathcal{S}_{\mathbb{N}} = \left\{ A \subset \mathbb{N} : A = \bigcup\nolimits_{l\in\mathbb{N}_0} 10^l B \text{ for some } B \subset \mathbb{N}_{\cancel{10}} \right\} .$$

A typical member of $\mathcal{S}_{\mathbb{N}}$ is

$$\{271, 2710, 3141, 27100, 31410, 271000, 314100, \ldots\} .$$

Note that for instance the set $\{31410, 314100, 3141000, \ldots\}$ does *not* belong to $\mathcal{S}_{\mathbb{N}}$ since 31410 is indistinguishable from 3141 in terms of significant digits, so if the former number were to belong to $A \in \mathcal{S}_{\mathbb{N}}$ then the latter would too. Note also that the corresponding significand function on $\mathbb{N}$ still only takes values in $[1,10)$, as before, but may never be an irrational number. In fact, the possible values of S on $\mathbb{N}$ are even more restricted: $S(n) = t$ for some $n \in \mathbb{N}$ if and only if $t \in [1,10)$ and $10^l t \in \mathbb{N}$ for some integer $l \geq 0$.

The next lemma establishes some basic closure properties of the significand σ-algebra that will be essential later in studying characteristic aspects of BL such as scale- and base-invariance. To concisely formulate these properties, for every $C \subset \mathbb{R}^+$ and $n \in \mathbb{N}$, let $C^{1/n} := \{t > 0 : t^n \in C\}$.

Lemma 2.13. *The following properties hold for the significand σ-algebra $\mathcal{S}$:*

(i) $\mathcal{S}$ *is self-similar with respect to multiplication by integer powers of* 10*, i.e.*

$$10^k A = A \quad \text{for every } A \in \mathcal{S} \text{ and } k \in \mathbb{Z} .$$

(ii) $\mathcal{S}$ *is closed under multiplication by a scalar, i.e.*

$$\alpha A \in \mathcal{S} \quad \text{for every } A \in \mathcal{S} \text{ and } \alpha > 0 .$$

(iii) $\mathcal{S}$ *is closed under integral roots, i.e.*

$$A^{1/n} \in \mathcal{S} \quad \text{for every } A \in \mathcal{S} \text{ and } n \in \mathbb{N} .$$

Informally, property (i) says that every significand set remains unchanged when multiplied by an integer power of 10 — reflecting the simple fact that shifting the decimal point keeps all the significant digits, and hence the set itself, unchanged; (ii) asserts that if every element of a set expressible solely in terms of significant digits is multiplied by a positive constant, then the new set is also expressible by significant digits; correspondingly, (iii) states that the collection of square (cubic, fourth etc.) roots of the elements of every significand set is also expressible in terms of its significant digits alone.

Proof. (i) This is obvious from (2.3) since $S(10^k A) = S(A)$ for every k.

(ii) Given $A \in \mathcal{S}$, by (2.4) there exists $B \in \mathcal{B}[1,10)$ such that $A = \bigcup_{k\in\mathbb{Z}} 10^k B$. In view of (i), assume without loss of generality that $1 < \alpha < 10$. Then

$$\alpha A = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k \alpha B = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k \Big(\big(\alpha B \cap [\alpha, 10)\big) \cup \big(\frac{\alpha}{10} B \cap [1,\alpha)\big) \Big) = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k C\,,$$

with $C = \big(\alpha B \cap [\alpha, 10)\big) \cup \big(\frac{\alpha}{10} B \cap [1,\alpha)\big) \in \mathcal{B}[1,10)$, showing that $\alpha A \in \mathcal{S}$.

(iii) Since intervals of the form $[1,t]$ generate $\mathcal{B}[1,10)$, i.e. since $\mathcal{B}[1,10) = \sigma\big(\{[1,t] : 1 < t < 10\}\big)$, it is enough to verify the claim for the special case $A = \bigcup_{k\in\mathbb{Z}} 10^k[1,10^s]$ for every $0 < s < 1$. In this case

$$A^{1/n} = \bigcup\nolimits_{k\in\mathbb{Z}} 10^{k/n}[1,10^{s/n}] = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k \bigcup\nolimits_{j=0}^{n-1} [10^{j/n}, 10^{(j+s)/n}] = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k C\,,$$

with $C = \bigcup_{j=0}^{n-1} [10^{j/n}, 10^{(j+s)/n}] \in \mathcal{B}[1,10)$. Hence $A^{1/n} \in \mathcal{S}$. □

Remark. Note that $\mathcal{S}$ is *not* closed under taking integer powers: If $A \in \mathcal{S}$ and $n \in \mathbb{N}$, then $A^n \in \mathcal{S}$ if and only if

$$S(A)^n = B \cup 10B \cup \ldots \cup 10^{n-1}B \quad \text{for some } B \in \mathcal{B}[1,10)\,.$$

For example, consider

$$A_5 = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k\{1, \sqrt{10}\} \in \mathcal{S}\,,$$

for which $S(A_5)^2 = \{1,10\} = \{1\} \cup 10\{1\}$ and hence $A_5^2 \in \mathcal{S}$, whereas choosing

$$A_6 = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k\{2, \sqrt{10}\}$$

leads to $S(A_6)^2 = \{4,10\}$, and correspondingly $A_6^2 \notin \mathcal{S}$. ♣

Since, by Theorem 2.9, the significand σ-algebra $\mathcal{S}$ is the same as the significant digit σ-algebra $\sigma(D_1, D_2, D_3, \ldots)$, the closure properties established in Lemma 2.13 carry over to sets determined by significant digits. The next example illustrates closure under multiplication by a scalar and integral roots.

Example 2.14. Let A_7 be the set of positive real numbers with first significant digit 1, i.e.

$$A_7 = \{x > 0 : D_1(x) = 1\} = \{x > 0 : 1 \le S(x) < 2\} = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k[1,2)\,.$$

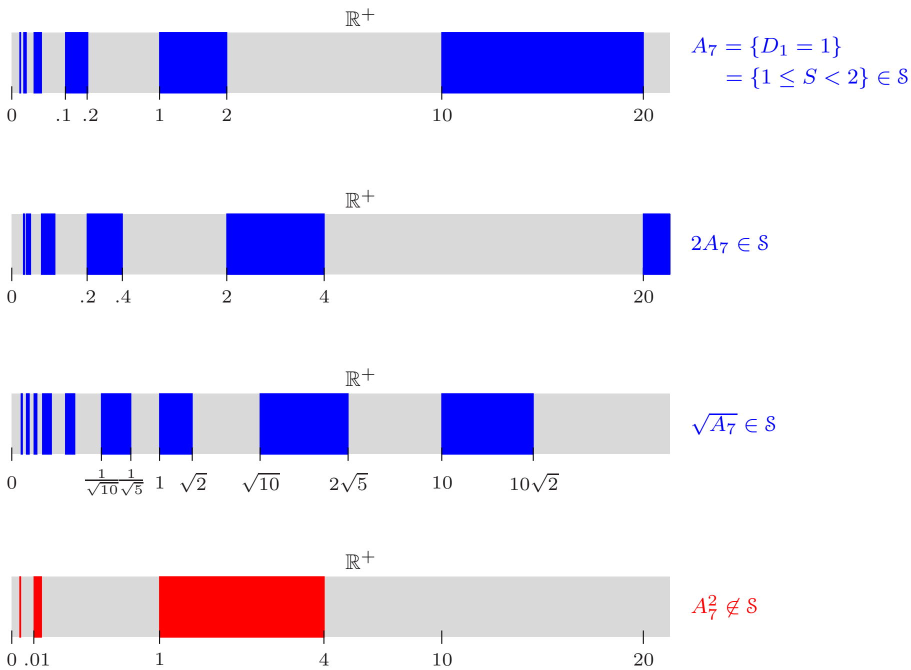

FIG 3. *The σ-algebra $\mathcal{S}$ is closed under multiplication by a scalar and integral roots but not under integer powers (bottom), see Example 2.14.*

Then

$$2A_7 = \bigl\{x > 0 : D_1(x) \in \{2,3\}\bigr\} = \{x > 0 : 2 \le S(x) < 3\} = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k[2,4) \in \mathcal{S}\,,$$

and also

$$A_7^{1/2} = \bigl\{x > 0 : S(x) \in [1,\sqrt{2}) \cup [\sqrt{10},\sqrt{20})\bigr\} = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k\bigl([1,\sqrt{2}) \cup [\sqrt{10}, 2\sqrt{5})\bigr) \in \mathcal{S}\,,$$

whereas on the other hand clearly

$$A_7^2 = \bigcup\nolimits_{k\in\mathbb{Z}} 10^{2k}[1,4) \notin \mathcal{S}\,,$$

since e.g. $[1,4) \subset A_7^2$ but $[10,40) \not\subset A_7^2$; see FIG 3.

Example 2.15. Recall the significand σ-algebra $\mathcal{S}_{\mathbb{N}}$ on the positive integers defined in Example 2.12. Unlike its continuous counterpart $\mathcal{S}$, the family $\mathcal{S}_{\mathbb{N}}$ is not even closed under multiplication by a positive integer, since for example

$$A_8 = \mathbb{N} \cap \{x > 0 : S(x) = 2\} = \{2, 20, 200, \ldots\} \in \mathcal{S}_{\mathbb{N}}\,,$$

but

$$5A_8 = \{10, 100, 1000, \ldots\} \notin \mathcal{S}_{\mathbb{N}} .$$

Of course, this does not rule out that some events determined by significant digits, i.e. some members of $\mathcal{S}_{\mathbb{N}}$, still belong to $\mathcal{S}_{\mathbb{N}}$ after multiplication by an integer. For example, if

$$A_9 = \{n \in \mathbb{N} : D_1(n) = 1\} = \{1, 10, 11, \ldots, 19, 100, 101, \ldots\} \in \mathcal{S}_{\mathbb{N}}$$

then

$$3A_9 = \{3, 30, 33, \ldots, 57, 300, 303, \ldots\} \in \mathcal{S}_{\mathbb{N}} .$$

It is easy to see that, more generally, $\mathcal{S}_{\mathbb{N}}$ is closed under multiplication by $m \in \mathbb{N}$ precisely if $\gcd(m, 10) = 1$, that is, whenever m and 10 have no non-trivial common factor. Moreover, like $\mathcal{S}$, the σ-algebra $\mathcal{S}_{\mathbb{N}}$ is closed under integral roots: If $A = \bigcup_{l \in \mathbb{N}_0} 10^l \widehat{A}$ with $\widehat{A} \subset \mathbb{N}_{\cancel{10}}$ then $A^{1/n} = \bigcup_{l \in \mathbb{N}_0} 10^l \widehat{A}^{1/n} \in \mathcal{S}_{\mathbb{N}}$. With A_9 from above, for instance,

$$\begin{aligned} A_9^{1/2} &= \big\{n \in \mathbb{N} : S(n) \in [1, \sqrt{2}) \cup [\sqrt{10}, \sqrt{20})\big\} \\ &= \{1, 4, 10, 11, 12, 13, 14, 32, 33, \ldots, 44, 100, 101, \ldots\} \in \mathcal{S}_{\mathbb{N}} . \end{aligned}$$

Thus many of the conclusions drawn later for positive real numbers carry over to positive integers in a straightforward way.

The next lemma provides a very convenient framework for studying probabilities on the significand σ-algebra by translating them into probability measures on the classical space of Borel subsets of $[0, 1)$, that is, on $\big([0, 1), \mathcal{B}[0, 1)\big)$. For a proper formulation, observe that for every function $f : \Omega \to \mathbb{R}$ with $\mathcal{A} \supset \sigma(f)$ and every probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$, f and $\mathbb{P}$ together induce a probability measure $f_*\mathbb{P}$ on $(\mathbb{R}, \mathcal{B})$ in a natural way, namely by setting

$$f_*\mathbb{P}(B) = \mathbb{P}\big(f^{-1}(B)\big) \quad \text{for all } B \in \mathcal{B} . \tag{2.5}$$

Other symbols commonly used in textbooks to denote $f_*\mathbb{P}$ include $\mathbb{P} \circ f^{-1}$ and $\mathbb{P}_f$. In the case of a *linear* function f, i.e. for $f(t) \equiv \alpha t$ with some $\alpha \in \mathbb{R}$, instead of $f_*\mathbb{P}$ simply write $\alpha_*\mathbb{P}$. The special case of interest for significands is $(\Omega, \mathcal{A}) = (\mathbb{R}^+, \mathcal{S})$ and $f = \log S$.

Lemma 2.16. *The function $\ell : \mathbb{R}^+ \to [0, 1)$ defined by $\ell(x) = \log S(x)$ establishes a one-to-one and onto correspondence (measure isomorphism) between probability measures on $(\mathbb{R}^+, \mathcal{S})$ and on $\big([0, 1), \mathcal{B}[0, 1)\big)$, respectively.*

Proof. From $\ell^{-1}\big([a, b]\big) = S^{-1}\big([10^a, 10^b]\big)$ for all $0 \le a < b < 1$, it follows that $\sigma(\ell) = \mathbb{R}^+ \cap \sigma(S) = \mathcal{S}$, and hence $\ell_*\mathbb{P}$ according to (2.5) is a well-defined probability measure on $\big([0, 1), \mathcal{B}[0, 1)\big)$.

Conversely, given any probability measure P on $\big([0, 1), \mathcal{B}[0, 1)\big)$ and any A in $\mathcal{S}$, let $B \in \mathcal{B}[0, 1)$ be the unique set for which $A = \bigcup_{k \in \mathbb{Z}} 10^k 10^B$, where $10^B = \{10^s : s \in B\}$, and define

$$\mathbb{P}_P(A) := P(B) .$$

It is readily confirmed that $\ell(A) = B$, $\ell^{-1}(B) = A$, and $\mathbb{P}_P$ is a well-defined probability measure on $(\mathbb{R}^+, \mathcal{S})$. Moreover

$$\ell_* \mathbb{P}_P(B) = \mathbb{P}_P\big(\ell^{-1}(B)\big) = \mathbb{P}_P(A) = P(B) \quad \text{for all } B \in \mathcal{B}[0,1)\,,$$

showing that $\ell_* \mathbb{P}_P = P$, and hence *every* probability measure on $\big([0,1), \mathcal{B}[0,1)\big)$ is of the form $\ell_* \mathbb{P}$ with the appropriate $\mathbb{P}$. On the other hand,

$$\mathbb{P}_{\ell_* \mathbb{P}}(A) = \ell_* \mathbb{P}(B) = \mathbb{P}\big(\ell^{-1}(B)\big) = \mathbb{P}(A) \quad \text{for all } A \in \mathcal{S}\,,$$

and hence the correspondence $\mathbb{P} \mapsto \ell_* \mathbb{P}$ is one-to-one as well. Overall $\mathbb{P} \leftrightarrow \ell_* \mathbb{P}$ is bijective. □

From the proof of Lemma 2.16 it is clear that a bijective correspondence between probability measures on $(\mathbb{R}^+, \mathcal{S})$ and on $\big([0,1), \mathcal{B}[0,1)\big)$, respectively, could have been established in many other ways as well, e.g. by using the function $\widetilde{\ell}(x) = \frac{1}{9}(S(x) - 1)$ instead of ℓ. The special role of ℓ according to that lemma only becomes apparent through its relation to BL. To see this, denote by $\mathbb{B}$ the (unique) probability measure on $(\mathbb{R}^+, \mathcal{S})$ with

$$\mathbb{B}\big(\{x > 0 : S(x) \le t\}\big) = \mathbb{B}\left(\bigcup\nolimits_{k \in \mathbb{Z}} 10^k [1, t]\right) = \log t \quad \text{for all } 1 \le t < 10\,.$$

In view of (2.1), the probability measure $\mathbb{B}$ on $(\mathbb{R}^+, \mathcal{S})$ is the most natural formalization of BL. On the other hand, it will become clear in subsequent chapters that on $\big([0,1), \mathcal{B}[0,1)\big)$ the uniform distribution $\lambda_{0,1}$ has many special properties and hence plays a very distinguished role. The relevance of the specific choice for ℓ in Lemma 2.16, therefore, is that $\ell_* \mathbb{B} = \lambda_{0,1}$. The reader will learn shortly why, for a deeper understanding of BL, the latter relation is very beneficial indeed.

3. The Benford property

In order to translate the informal versions (1.1), (1.2) and (2.1) of BL into more precise statements about various types of mathematical objects, it is necessary to specify exactly what the Benford property means for any one of these objects. For the purpose of the present chapter, the objects of interest fall into three categories: sequences of real numbers, real-valued functions defined on $[0, +\infty)$; and probability distributions and random variables. (Recall also Fig 1.)

3.1. Benford sequences

A sequence $(x_n) = (x_1, x_2, x_3, \ldots)$ of real numbers is a (base-10) *Benford sequence*, or simply *Benford*, if, as $N \to \infty$, the limiting proportion of indices $n \le N$ for which x_n has first significant digit d_1 exists and equals $\log(1 + d_1^{-1})$ for all $d_1 \in \{1, 2, \ldots, 9\}$, and similarly for the limiting proportions of the occurrences of all other finite blocks of initial significant digits. The formal definition is as follows.

Definition 3.1. A sequence (x_n) of real numbers is a *Benford sequence*, or *Benford* for short, if

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : S(x_n) \le t\}}{N} = \log t \quad \text{for all } t \in [1, 10)\,,$$

or equivalently, if for all $m \in \mathbb{N}$, all $d_1 \in \{1, 2, \dots, 9\}$ and all $d_j \in \{0, 1, \dots, 9\}$, $j \ge 2$,

$$\lim_{N\to\infty} \frac{\#\bigl\{1 \le n \le N : D_j(x_n) = d_j \text{ for } j = 1, 2, \dots, m\bigr\}}{N} = \log\left(1 + \left(\sum\nolimits_{j=1}^{m} 10^{m-j} d_j\right)^{-1}\right).$$

As will be shown below, the sequence of powers of 2, namely $(2^n) = (2, 4, 8, \dots)$ is Benford. However, it is not Benford base 2 since the second significant digit base 2 of 2^n is 0 for every n, whereas the generalized version (1.3) of BL requires that $0 < \text{Prob}\bigl(D_2^{(2)} = 0\bigr) = 1 - \text{Prob}\bigl(D_2^{(2)} = 1\bigr) = \log_2 3 - 1 < 1$. Similarly, (3^n), the sequence of powers of 3 is Benford, and so is the sequence of factorials $(n!)$ as well as the sequence (F_n) of Fibonacci numbers. Simple examples of sequences that are not Benford are the positive integers (n), the powers of 10 and the sequence of logarithms $(\log n)$.

The notion of Benford sequence according to Definition 3.1 offers a natural interpretation of Prob in the informal expressions (1.1)–(1.3): A sequence (x_n) is Benford if, when one of the first N entries in (x_n) is chosen (uniformly) at random, the probability that this entry's first significant digit is d approaches the Benford probability $\log(1 + d^{-1})$ as $N \to \infty$, for every $d \in \{1, 2, \dots, 9\}$, and similarly for all other blocks of significant digits.

Example 3.2. Two specific sequences of positive integers will be used repeatedly to illustrate key concepts concerning BL: the Fibonacci numbers and the prime numbers. Both sequences play prominent roles in many areas of mathematics.

(i) As will be seen in Example 4.12, the sequences of Fibonacci numbers $(F_n) = (1, 1, 2, 3, 5, 8, 13, \dots)$, where every entry is simply the sum of its two predecessors, and $F_1 = F_2 = 1$, is Benford. Already the first $N = 10^2$ elements of the sequence conform very well to the first-digit version (1.1) of BL, with Prob being interpreted as relative frequency, see Fig 4. The conformance gets even better if the first $N = 10^4$ elements are considered, see Fig 5.

(ii) In Example 4.11(v), it will become apparent that the sequence of prime numbers $(p_n) = (2, 3, 5, 7, 11, 13, 17, \dots)$ is *not* Benford. Fig 4 shows how, accordingly, the first hundred prime numbers do not conform well to (1.1). Moreover, the conformance gets even worse if the first ten thousand primes are considered (Fig 5).

Fibonacci numbers $F_1, F_2, \ldots, F_{100}$

1	10946	165580141	2504730781961	37889062373143906
1	17711	267914296	4052739537881	61305790721611591
2	28657	433494437	6557470319842	99194853094755497
3	46368	701408733	10610209857723	160500643816367088
5	75025	1134903170	17167680177565	259695496911122585
8	121393	1836311903	27777890035288	420196140727489673
13	196418	2971215073	44945570212853	679891637638612258
21	317811	4807526976	72723460248141	1100087778366101931
34	514229	7778742049	117669030460994	1779979416004714189
55	832040	12586269025	190392490709135	2880067194370816120
89	1346269	20365011074	308061521170129	4660046610375530309
144	2178309	32951280099	498454011879264	7540113804746346429
233	3524578	53316291173	806515533049393	12200160415121876738
377	5702887	86267571272	1304969544928657	19740274219868223167
610	9227465	139583862445	2111485077978050	31940434634990099905
987	14930352	225851433717	3416454622906707	51680708854858323072
1597	24157817	365435296162	5527939700884757	83621143489848422977
2584	39088169	591286729879	8944394323791464	135301852344706746049
4181	63245986	956722026041	14472334024676221	218922995834555169026
6765	165580141	1548008755920	23416728348467685	354224848179261915075

Prime numbers $p_1, p_2, \ldots, p_{100}$

2	31	73	127	179	233	283	353	419	467
3	37	79	131	181	239	293	359	421	479
5	41	83	137	191	241	307	367	431	487
7	43	89	139	193	251	311	373	433	491
11	47	97	149	197	257	313	379	439	499
13	53	101	151	199	263	317	383	443	503
17	59	103	157	211	269	331	389	449	509
19	61	107	163	223	271	337	397	457	521
23	67	109	167	227	277	347	401	461	523
29	71	113	173	229	281	349	409	463	541

FIG 4. *The first one-hundred Fibonacci numbers conform to the first digit law (1.1) quite well (top and bottom), while the first one-hundred prime numbers clearly do not (center and bottom).*

$\#\{D_1 = d\}$

d	1	2	3	4	5	6	7	8	9
Fibonacci	30	18	13	9	8	6	5	7	4
Prime	25	19	19	20	8	2	4	2	1
$10^2 \cdot \log(1+d^{-1})$	30.10	17.60	12.49	9.691	7.918	6.694	5.799	5.115	4.575

Remark. Based on discrete density and summability definitions, many alternative notions of Benford sequences have been proposed, utilizing e.g. reiteration of Cesàro averaging [Fl], and logarithmic density methods. The reader is referred to [Ra1, Ra2] for an excellent summary of those approaches. Those methods, however, do not offer as natural an interpretation of "Prob" as Definition 3.1. On this, Raimi [Ra1, p.529] remarks that "[t]he discrete summability schemes [. . .] can also be tortured into probability interpretations, though none of the authors [. . .] (except Diaconis) does so".

$\#\{D_1 = d\} = N\cdot\rho_N(d)$	d	1	2	3	4	5	6	7	8	9	$10^3 \cdot R$
Fibonacci	$N = 10^2$	30	18	13	9	8	6	5	7	4	18.84
Fibonacci	$N = 10^4$	3011	1762	1250	968	792	668	580	513	456	0.1574(!)
Prime	$N = 10^2$	25	19	19	20	8	2	4	2	1	103.0
Prime	$N = 10^4$	1601	1129	1097	1069	1055	1013	1027	1003	1006	140.9
	$10^4 \cdot \log(1+d^{-1})$	3010.	1760.	1249.	969.1	791.8	669.4	579.9	511.5	457.5	

FIG 5. *Increasing the sample size from $N = 10^2$ to $N = 10^4$ entails an even better conformance with (1.1) for the Fibonacci numbers, as measured by means of the quantity $R = \max_{d=1}^{9} \left|\rho_N(d) - \log(1 + d^{-1})\right|$. For the primes, on the other hand, the rather poor conformance does not improve at all.*

Only the notion according to Definition 3.1 will be used henceforth. However, to get an impression how alternative concepts may relate to Definition 3.1 analytically, denote, for any set $C \subset \mathbb{R}$, by $\mathbb{1}_C$ the *indicator function* of C, i.e. $\mathbb{1}_C : \mathbb{R} \to \{0, 1\}$ with

$$\mathbb{1}_C(t) = \begin{cases} 1 & \text{if } t \in C\,, \\ 0 & \text{otherwise.} \end{cases}$$

Given any sequence (x_n) and any number $t \in [1, 10)$, consider the sequence $\bigl(\mathbb{1}_{[1,t)}(S(x_n))\bigr)$. Clearly, since the latter contains only zeros and ones, it will usually not converge. It may, however, converge in some averaged sense, to a limit that may depend on t. Specifically, (x_n) is Benford if and only if

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} \mathbb{1}_{[1,t)}\bigl(S(x_n)\bigr)}{N} = \log t \quad \text{for all } t \in [1, 10)\,. \tag{3.1}$$

Instead of (3.1), one could more generally consider the convergence of

$$\frac{\sum_{n=1}^{N} a_n \mathbb{1}_{[1,t)}\bigl(S(x_n)\bigr)}{\sum_{n=1}^{N} a_n}\,, \tag{3.2}$$

where the a_n can be virtually any non-negative numbers with $\sum_{n=1}^{N} a_n \to +\infty$ as $N \to \infty$. With this, (3.1) corresponds to the special case $a_n = 1$ for all n. Another popular choice in (3.2), related to the number-theoretic concept of *logarithmic* (or *analytic*) *density* [Se], is $a_n = n^{-1}$ for all n, in which case $(\ln N)^{-1} \sum_{n=1}^{N} a_n \to 1$. Utilizing the latter, a sequence (x_n) of real numbers might be (and has been, see [Ra1]) called *weakly Benford* if

$$\lim_{N\to\infty} \frac{1}{\ln N} \sum_{n=1}^{N} \frac{\mathbb{1}_{[1,t)}\big(S(x_n)\big)}{n} = \log t \quad \text{for all } t \in [1,10)\,.$$

It is easy to check that every Benford sequence is weakly Benford. To see that the converse does not hold in general, take for instance $(x_n) = (n)$. A short calculation confirms that, for every $t \in [1,10)$,

$$\liminf_{N\to\infty} \frac{\sum_{n=1}^{N} \mathbb{1}_{[1,t)}\big(S(n)\big)}{N} = \frac{t-1}{9}\,,$$

whereas

$$\limsup_{N\to\infty} \frac{\sum_{n=1}^{N} \mathbb{1}_{[1,t)}\big(S(n)\big)}{N} = \frac{10}{9} \cdot \frac{t-1}{t}\,,$$

showing that (n) is not Benford. (Recall that the *limit inferior* and *limit superior* of a sequence (a_n), denoted by $\liminf_{n\to\infty} a_n$ and $\limsup_{n\to\infty} a_n$, are defined, respectively, as the smallest and largest accumulation value of (a_n).) On the other hand, (n) turns out to be weakly Benford: Indeed, given N, let $L_N := \lfloor \log N \rfloor$. For any $t \in [1,10)$, it follows from the elementary estimate

$$\begin{aligned} \frac{1}{\ln 10^{L_N+1}} \sum_{i=0}^{L_N-1} \sum_{j=10^i}^{\lfloor 10^i t\rfloor} \frac{1}{j} &\le \frac{1}{\ln N} \sum_{n=1}^{N} \frac{\mathbb{1}_{[1,t)}\big(S(n)\big)}{n} \\ &\le \frac{1}{\ln \lfloor 10^{L_N} t \rfloor} \sum_{i=0}^{L_N} \sum_{j=10^i}^{\lfloor 10^i t\rfloor} \frac{1}{j}\,, \end{aligned}$$

together with

$$\sum_{j=10^i}^{\lfloor 10^i t\rfloor} \frac{1}{j} = 10^{-i} \sum_{j=0}^{\lfloor 10^i t\rfloor - 10^i} \frac{1}{1+10^{-i}j} \quad \to \quad \int_0^{t-1} \frac{\mathrm{d}\tau}{1+\tau} = \ln t\,, \quad \text{as } i \to \infty\,,$$

as well as

$$\lim_{L\to\infty} \frac{\ln 10^{L+1}}{L} = \lim_{L\to\infty} \frac{\ln \lfloor 10^L t \rfloor}{L} = \ln 10$$

and the Cauchy Limit Theorem that

$$\lim_{N\to\infty} \frac{1}{\ln N} \sum_{n=1}^{N} \frac{\mathbb{1}_{[1,t)}\big(S(n)\big)}{n} = \frac{\ln t}{\ln 10} = \log t\,,$$

i.e., (n) is weakly Benford. In a similar manner, the sequence (p_n) can be shown to be weakly Benford without being Benford, see [GG, Wh]. ♣

3.2. Benford functions

BL also appears frequently in real-valued functions such as e.g. those arising as solutions of initial value problems for differential equations (see Section 5.3 below). Thus, the starting point is to define what it means for a function to follow BL.

Recall that a function $f:\mathbb{R}\to\mathbb{R}$ is (*Borel*) *measurable* if $f^{-1}(I)$ is a Borel set, i.e. $f^{-1}(I)\in\mathcal{B}$, for every interval $I\subset\mathbb{R}$. With the terminology introduced in Section 2.3, this is equivalent to saying that $\sigma(f)\subset\mathcal{B}$. Slightly more generally, for any set Ω and any σ-algebra $\mathcal{A}$ on Ω, a function $f:\Omega\to\mathbb{R}$ is (*Borel*) *measurable* if $\sigma(f)\subset\mathcal{A}$. The collection of Borel measurable functions $f:\mathbb{R}\to\mathbb{R}$ contains all functions of practical interest. For example, every piecewise continuous function (meaning that f has at most countably many discontinuities) is measurable. Thus every polynomial, trigonometric and exponential function is measurable, and so is every probability density function of any relevance. In fact, it is a difficult exercise to produce a function that is *not* measurable, or a set $C\subset\mathbb{R}$ that is *not* a member of $\mathcal{B}$, and this can be done only in a non-constructive way. For all practical purposes, therefore, the reader may simply read "set" for "Borel set", and "function" for "Borel measurable function".

Recall that given a set Ω and a σ-algebra $\mathcal{A}$ on Ω, a *measure* μ on $(\Omega,\mathcal{A})$ is a function $\mu:\mathcal{A}\to[0,+\infty]$ that has all the properties of a probability measure, except that $\mu(A)$ may also be bigger than 1, and even infinite. By far the most important example is the so-called *Lebesgue measure* on $(\mathbb{R},\mathcal{B})$, denoted by λ here and throughout. The basic, and in fact defining property of λ is that $\lambda\bigl([a,b]\bigr)=b-a$ for every interval $[a,b]\subset\mathbb{R}$. The relation between the measure λ and the probability measures $\lambda_{a,b}$ considered earlier is such that, for instance,

$$\lambda(B)=\lim\nolimits_{N\to\infty}2N\lambda_{-N,N}\bigl(B\cap[-N,N]\bigr)\quad\text{for every } B\in\mathcal{B}\,.$$

It is customary to also use the symbol λ, often without a subscript etc., to denote the restriction of Lebesgue measure to $\bigl(C,\mathcal{B}(C)\bigr)$ with the Borel set C being clear from the context.

In analogy to the terminology for sequences, a function f is a (base-10) Benford function, or simply Benford, if the limiting proportion of the time $\tau<T$ that the first digit of $f(\tau)$ equals d_1 is exactly $\log(1+d_1^{-1})$, and similarly for the other significant digits, and in fact the significand. The formal definition is as follows.

Definition 3.3. A (Borel measurable) function $f:[0,+\infty)\to\mathbb{R}$ is *Benford* if

$$\lim\nolimits_{T\to+\infty}\frac{\lambda\bigl(\{\tau\in[0,T):S\bigl(f(\tau)\bigr)\le t\}\bigr)}{T}=\log t\quad\text{for all } t\in[1,10)\,,$$

or equivalently, if for all $m\in\mathbb{N}$, all $d_1\in\{1,2,\dots,9\}$ and all $d_j\in\{0,1,\dots,9\}$, $j\ge 2$,

$$\begin{aligned}\lim\nolimits_{T\to+\infty}&\frac{\lambda\bigl(\{\tau\in[0,T):D_j\bigl(f(\tau)\bigr)=d_j\text{ for } j=1,2,\dots,m\}\bigr)}{T}\\&=\log\left(1+\left(\sum\nolimits_{j=1}^{m}10^{m-j}d_j\right)^{-1}\right).\end{aligned}$$

Directly analogous to the probabilistic interpretation of a Benford sequence, the definition of a Benford function given in Definition 3.3 also offers a natural probabilistic interpretation: A function $f:[0,+\infty)\to\mathbb{R}$ is Benford if, when a time τ is chosen (uniformly) at random in $[0, T)$, the probability that the first digit of $f(\tau)$ is d approaches $\log(1+d^{-1})$ as $T \to +\infty$, for every $d \in \{1, 2, \ldots, 9\}$, and similarly for all other blocks of significant digits.

As will be seen in Example 4.5 below, the function $f(t) = e^{\alpha t}$ is Benford whenever $\alpha \neq 0$, but $f(t) = t$ and $f(t) = \sin^2 t$, for instance, are not.

3.3. Benford distributions and random variables

BL appears prominently in a wide variety of statistics and probability settings, such as e.g. in products of independent, identically distributed random variables, mixtures of random samples, and stochastic models like geometric Brownian motion that are of great importance for the stochastic modelling of real-world processes. This section lays the foundations for analyzing the Benford property for probability distributions and random variables. The term *independent, identically distributed* will henceforth be abbreviated *i.i.d.*, in accordance with standard stochastic terminology.

Recall from Section 2.3 that a probability space is a triple $(\Omega, \mathcal{A}, \mathbb{P})$ where Ω is a set, often referred to as the *set of outcomes*, $\mathcal{A}$ is a σ-algebra (the *family of events*), and $\mathbb{P}$ is a probability measure. A (real-valued) *random variable* X on $(\Omega, \mathcal{A}, \mathbb{P})$ is simply a Borel measurable function $X : \Omega \to \mathbb{R}$, and its *distribution* P_X is the probability measure on $(\mathbb{R}, \mathcal{B})$ defined by

$$P_X\big((-\infty, t]\big) = \mathbb{P}(X \leq t) \quad \text{for all } t \in \mathbb{R}\,.$$

Thus with the notation introduced in (2.5), simply $P_X = X_*\mathbb{P}$. The *expectation*, or *expected* (or *mean*) *value* of X is

$$\mathbb{E}X = \int_\Omega X \,\mathrm{d}\mathbb{P} = \int_\mathbb{R} t \,\mathrm{d}P_X(t)\,,$$

provided that this integral exists. More generally, for every measurable function $g : \mathbb{R} \to \mathbb{R}$, the expectation of the random variable $g(X)$ is

$$\mathbb{E}g(X) = \int_\Omega g(X) \,\mathrm{d}\mathbb{P} = \int_\mathbb{R} g(t) \,\mathrm{d}P_X(t)\,.$$

In particular, if $\mathbb{E}X$ exists, then $\operatorname{var} X := \mathbb{E}(X - \mathbb{E}X)^2$ is the *variance* of X.

Any probability measure on $(\mathbb{R}, \mathcal{B})$ will be referred to as a *Borel probability measure* on $\mathbb{R}$. Again, since all subsets of $\mathbb{R}$ of any practical interest are Borel sets, the specifier "Borel" will be suppressed unless there is a potential for confusion, i.e., the reader may read "probability measure on $\mathbb{R}$" for "Borel probability measure on $\mathbb{R}$". Any probability measure P on $\mathbb{R}$ is uniquely determined by its *distribution function* F_P, defined as

$$F_P(t) = P\big((-\infty, t]\big) \quad \text{for all } t \in \mathbb{R}\,.$$

It is easy to check that the function F_P is right-continuous and non-decreasing, with $\lim_{t\to-\infty} F_P(t) = 0$ and $\lim_{t\to+\infty} F_P(t) = 1$. For the sake of notational simplicity, write F_X instead of F_{P_X} for every random variable X. The probability measure P, or any random variable X with $P_X = P$, is *continuous* (or *atomless*) if $P(\{t\}) = 0$ for every $t \in \mathbb{R}$, or equivalently if the function F_P is continuous. It is *absolutely continuous* (*a.c.*) if, for any $B \in \mathcal{B}$, $P(B) = 0$ holds whenever $\lambda(B) = 0$. By the Radon–Nikodym Theorem, this is equivalent to P having a *density*, i.e. to the existence of a measurable function $f_P : \mathbb{R} \to [0, +\infty)$ such that

$$P\big([a,b]\big) = \int_a^b f_P(t)\,\mathrm{d}t \quad \text{for all } [a,b] \subset \mathbb{R}\,. \tag{3.3}$$

Again, for simplicity write f_X instead of f_{P_X} for every a.c. random variable X. Note that (3.3) implies $\int_{-\infty}^{+\infty} f_P(t)\,\mathrm{d}t = 1$. Every a.c. probability measure on $(\mathbb{R}, \mathcal{B})$ is continuous but not vice versa, see e.g. [CT]. Given any probability P on $(\mathbb{R}, \mathcal{B})$, denote $|\cdot|_* P$ simply by $|P|$, that is,

$$|P|(B) = P\big(\{t \in \mathbb{R} : |t| \in B\}\big) \quad \text{for all } B \in \mathcal{B}\,.$$

Clearly, $|P|$ is concentrated on $[0, +\infty)$, i.e. $|P|\big([0, +\infty)\big) = 1$, and

$$F_{|P|}(t) = \begin{cases} 0 & \text{if } t < 0\,, \\ F_P(t) - F_P(-t) + P(\{-t\}) & \text{if } t \geq 0\,; \end{cases}$$

in particular, therefore, if P is continuous or a.c. then so is $|P|$, its density in the latter case being $\big(f_P(t) + f_P(-t)\big) \cdot \mathbb{1}_{[0,+\infty)}$, where f_P is the density of P.

Definition 3.4. A Borel probability measure P on $\mathbb{R}$ is *Benford* if

$$P\big(\{x \in \mathbb{R} : S(x) \leq t\}\big) = S_* P\big(\{0\} \cup [1, t]\big) = \log t \quad \text{for all } t \in [1, 10)\,.$$

A random variable X on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is *Benford* if P_X is Benford, i.e. if

$$\mathbb{P}\big(S(X) \leq t\big) = P_X\big(\{x \in \mathbb{R} : S(x) \leq t\}\big) = \log t \quad \text{for all } t \in [1, 10)\,,$$

or equivalently, if for all $m \in \mathbb{N}$, all $d_1 \in \{1, 2, \ldots, 9\}$ and all $d_j \in \{0, 1, \ldots, 9\}$, $j \geq 2$,

$$\mathbb{P}\big(D_j(X) = d_j \text{ for } j = 1, 2, \ldots, m\big) = \log\left(1 + \left(\sum\nolimits_{j=1}^m 10^{m-j} d_j\right)^{-1}\right).$$

Example 3.5. If X is a Benford random variable on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, then from (1.1) and the numerical values given in Chapter 1,

$$\begin{aligned} &\mathbb{P}(D_1(X) = 1) = \mathbb{P}(1 \leq S(X) < 2) = \log 2 = 0.3010\ldots\,, \\ &\mathbb{P}(D_1(X) = 9) = \log \frac{10}{9} = 0.04575\ldots\,, \\ &\mathbb{P}\big(\big(D_1(X), D_2(X), D_3(X)\big) = (3, 1, 4)\big) = \log \frac{315}{314} = 0.001380\ldots\,. \end{aligned}$$

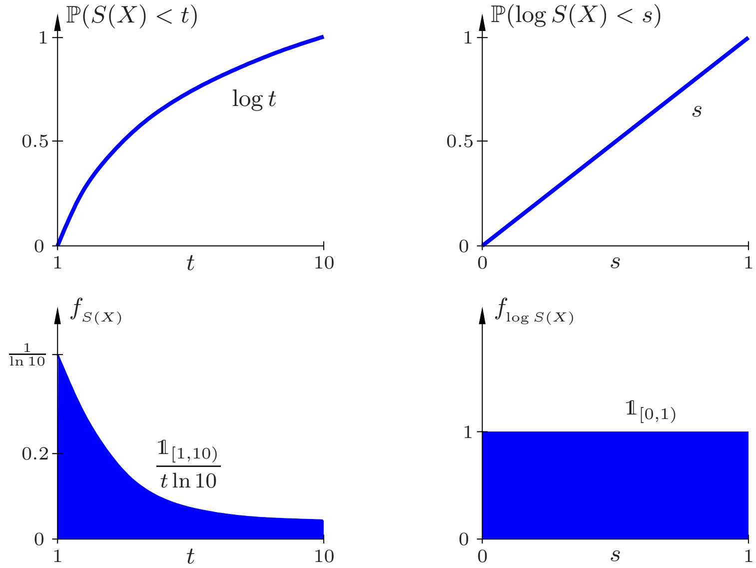

FIG 6. *The distribution functions (top) and densities of $S(X)$ and $\log S(X)$, respectively, for a Benford random variable X.*

As the following example shows, there are many probability measures on the positive real numbers, and correspondingly many positive random variables that are Benford.

Example 3.6. For every integer k, the probability measure P_k with density $f_k(x) = \frac{1}{x \ln 10}$ on $[10^k, 10^{k+1})$ is Benford, and so is e.g. $\frac{1}{2}(P_k + P_{k+1})$. In fact, every convex combination of the $(P_k)_{k\in\mathbb{Z}}$, i.e. every probability measure $\sum_{k\in\mathbb{Z}} q_k P_k$ with $0 \le q_k \le 1$ for all k and $\sum_{k\in\mathbb{Z}} q_k = 1$, is Benford.

As will be seen in Example 6.4 below, if U is a random variable uniformly distributed on $[0, 1)$, then the random variable $X = 10^U$ is Benford, but the random variable $X^{\log 2} = 2^U$ is not.

Definition 3.7. The *Benford distribution* $\mathbb{B}$ is the unique probability measure on $(\mathbb{R}^+, \mathcal{S})$ with

$$\mathbb{B}(S \le t) = \mathbb{B}\left(\bigcup\nolimits_{k\in\mathbb{Z}} 10^k[1, t]\right) = \log t \quad \text{for all } t \in [1, 10)\,,$$

or equivalently, for all $m \in \mathbb{N}$, all $d_1 \in \{1, 2, \dots, 9\}$ and all $d_j \in \{0, 1, \dots, 9\}$, $j \ge 2$,

$$\mathbb{B}\big(D_j = d_j \text{ for } j = 1, 2, \dots, m\big) = \log\left(1 + \left(\sum\nolimits_{j=1}^{m} 10^{m-j} d_j\right)^{-1}\right).$$

The combination of Definitions 3.4 and 3.7 gives

Proposition 3.8. *A Borel probability measure P on $\mathbb{R}$ is Benford if and only if*

$$|P|(A) = \mathbb{B}(A) \quad \text{for all } A \in \mathcal{S}\,.$$

In particular, if $P(\mathbb{R}^+) = 1$ then P is Benford precisely if $P(A) = \mathbb{B}(A)$ for all $A \in \mathcal{S}$.

Note that the Benford distribution $\mathbb{B}$ is a probability distribution on the significant digits, or the significand, of the underlying data, and not on the raw data themselves. That is, $\mathbb{B}$ is a probability measure on the family of sets defined by the base-10 significand, i.e. on $(\mathbb{R}^+, \mathcal{S})$, but not on the bigger $(\mathbb{R}^+, \mathcal{B}^+)$ or the still bigger $(\mathbb{R}, \mathcal{B})$. For example, the probability $\mathbb{B}(\{1\})$ is not defined, simply because the set $\{1\}$ cannot be defined in terms of significant digits or significands alone, and hence does not belong to the domain of $\mathbb{B}$.

Example 3.9. In the framework of the Examples 2.12 and 2.15, it is tempting to call a probability P on $(\mathbb{N}, \mathcal{S}_\mathbb{N})$ a *Benford distribution on* $\mathbb{N}$ if

$$P\big(\{n \in \mathbb{N} : S(n) \le t\}\big) = \log t \quad \text{for all } t \in [1, 10)\,.$$

However, no such probability exists! To see this, for every $n \in \mathbb{N}_{\cancel{10}}$ let $A_n = \bigcup_{l \in \mathbb{N}_0} 10^l \{n\} \in \mathcal{S}_\mathbb{N}$ and note that $\mathbb{N}$ equals the disjoint union of the sets A_n, and $S(A_n) = \{10^{\langle \log n \rangle}\}$; here $\langle \log n \rangle \in [0, 1)$ denotes the fractional part of $\log n$, that is, $\langle \log n \rangle = \log n - \lfloor \log n \rfloor$. With $q_n := P(A_n)$ therefore $\sum_{n \in \mathbb{N}_{\cancel{10}}} q_n = 1$ and $S_* P = \sum_{n \in \mathbb{N}_{\cancel{10}}} q_n \delta_{10^{\langle \log n \rangle}}$. Since the set of discontinuities of $t \mapsto F_{S_* P}(t)$ is $\big\{10^{\langle \log n \rangle} : q_n \ne 0\big\} \ne \varnothing$, it is impossible to have $F_{S_* P}(t) = \log t$ for all $t \in [1, 10)$. Note that, as a consequence, a Borel probability measure P on $\mathbb{R}$ concentrated on $\mathbb{N}$, i.e. with $P(\mathbb{N}) = 1$, cannot be Benford.

On the other hand, given $\varepsilon > 0$ it is not hard to find a probability P_ε on $(\mathbb{N}, \mathcal{S}_\mathbb{N})$ with

$$\big|P_\varepsilon\big(\{n \in \mathbb{N} : S(n) \le t\}\big) - \log t\big| < \varepsilon \quad \text{for all } t \in [1, 10)\,. \tag{3.4}$$

For a concrete example, for any $N \in \mathbb{N}$ consider the probability measure

$$Q_N := c_N^{-1} \sum\nolimits_{j=10^N}^{10^{N+1}-1} j^{-1} \delta_j\,,$$

where $c_N = \sum\nolimits_{j=10^N}^{10^{N+1}-1} j^{-1}$. Note that Q_N may be thought of as a discrete approximation of the Benford probability P_N in Example 3.6. From

$$S_* Q_N = c_N^{-1} \sum\nolimits_{j=10^N}^{10^{N+1}-1} j^{-1} \delta_{S(j)} = c_N^{-1} \sum\nolimits_{j=1}^{10^{N+1}-10^N} \frac{1}{10^N + j - 1} \delta_{1 + 10^{-N}(j-1)}\,,$$

together with the elementary estimate $\ln \frac{M+1}{L} < \sum_{j=L}^M j^{-1} < \ln \frac{M}{L-1}$, valid for all $L, M \in \mathbb{N}$ with $2 \le L < M$, it is straightforward to deduce that, for all $1 \le t < 10$,

$$\big|S_* Q_N\big([1, t]\big) - \log t\big| < -\log(1 - 10^{-N}) = \frac{10^{-N}}{\ln 10} + \mathcal{O}(10^{-2N}) \quad \text{as } N \to \infty\,.$$

Thus (3.4) is guaranteed by taking $P_\varepsilon = Q_N$ with N sufficiently large. A short calculation confirms that it suffices to choose $N > 1 + |\log \varepsilon|$.

Example 3.10. **(i)** If X is distributed according to $U(0,1)$, the uniform distribution on $[0,1)$, i.e. $P_X = \lambda_{0,1}$, then for every $1 \le t < 10$,

$$\mathbb{P}\big(S(X) \le t\big) = \lambda_{0,1}\left(\bigcup\nolimits_{k\in\mathbb{Z}} 10^k[1,t]\right) = \sum\nolimits_{n\in\mathbb{N}} 10^{-n}(t-1) = \frac{t-1}{9} \not\equiv \log t\,,$$

showing that $S(X)$ is uniform on $[1,10)$, and hence $\lambda_{0,1}$ is not Benford.

(ii) If X is distributed according to $\texttt{exp}(1)$, the exponential distribution with mean 1, whose distribution function is given by $F_{\texttt{exp}(1)}(t) = \max(0, 1-e^{-t})$, then

$$\begin{aligned}\mathbb{P}(D_1(X)=1) &= \mathbb{P}\left(X \in \bigcup\nolimits_{k\in\mathbb{Z}} 10^k[1,2)\right) = \sum\nolimits_{k\in\mathbb{Z}}\left(e^{-10^k} - e^{-2\cdot 10^k}\right)\\ &> \left(e^{-1/10} - e^{-2/10}\right) + \left(e^{-1} - e^{-2}\right) + \left(e^{-10} - e^{-20}\right)\\ &= 0.3186\ldots \; > \; \log 2\,,\end{aligned}$$

and hence $\texttt{exp}(1)$ is not Benford either. (See [EL, LSE, MN] for a detailed analysis of the exponential distribution's relation to BL.)

(iii) Let X be distributed according to the $\texttt{Beta}\big(\frac12,\frac12\big)$- or arcsin-distribution, meaning that $\mathbb{P}(X \le s) = \frac{2}{\pi}\arcsin\sqrt{s}$ for all $0 \le s < 1$. It follows that, for every $1 \le t < 10$,

$$\begin{aligned}F_{S(X)}(t) = \mathbb{P}(S(X)\le t) &= \mathbb{P}\left(X \in \bigcup\nolimits_{n\in\mathbb{N}} 10^{-n}[1,t]\right)\\ &= \frac{2}{\pi}\sum\nolimits_{n=1}^{\infty}\left(\arcsin(10^{-n/2}\sqrt{t}) - \arcsin(10^{-n/2})\right)\\ &= \frac{2}{\pi}\sum\nolimits_{l=0}^{\infty}\frac{(2l)!}{2^{2l}(l!)^2(2l+1)}\cdot\frac{t^{l+1/2}-1}{10^{l+1/2}-1}\,,\end{aligned}$$

and hence in particular

$$\begin{aligned}F_{S(X)}(\sqrt{10}) &= \frac{2}{\pi}\sum\nolimits_{l=0}^{\infty}\frac{(2l)!}{2^{2l}(l!)^2(2l+1)}\cdot\frac{1}{10^{l/2+1/4}+1}\\ &< \frac{2}{\pi}\sum\nolimits_{l=0}^{\infty}\frac{(2l)!}{2^{2l}(l!)^2(2l+1)}10^{-(l/2+1/4)}\\ &= \frac{2}{\pi}\arcsin(10^{-1/4}) = 0.3801\ldots \; < \; \frac{2}{5}\,,\end{aligned}$$

which in turn shows that X is not Benford, as $F_{\mathbb{B}}(\sqrt{10}) = \frac12$. Alternatively, $F_{S(X)}$ is easily seen to be strictly *convex* on $[1,10)$ and therefore $F_{S(X)}(t) \equiv \log t$ cannot possibly hold.

4. Characterizations of Benford's Law

The purpose of this chapter is to establish and illustrate four useful characterizations of the Benford property in the context of sequences, functions, distributions and random variables, respectively. These characterizations will be instrumental in demonstrating that certain datasets are, or are not, Benford, and helpful for predicting which empirical data are likely to follow BL closely.

4.1. The uniform distribution characterization

The uniform distribution characterization is undoubtedly the most basic and powerful of all characterizations, mainly because the mathematical theory of uniform distribution mod 1 is very well developed, see e.g. [DT, KN] for authoritative surveys of the theory.

Here and throughout, denote by $\langle t\rangle$ the *fractional part* of any real number t, that is $\langle t\rangle = t - \lfloor t\rfloor$. For example, $\langle\pi\rangle = \langle 3.1415\ldots\rangle = 0.1415\ldots = \pi - 3$. Recall that $\lambda_{0,1}$ denotes Lebesgue measure on $\big([0,1), \mathcal{B}[0,1)\big)$.

Definition 4.1. A sequence (x_n) of real numbers is *uniformly distributed modulo 1*, abbreviated henceforth as *u.d.* mod 1, if

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : \langle x_n\rangle \le s\}}{N} = s \quad \text{for all } s \in [0,1)\,;$$

a (Borel measurable) function $f : [0,+\infty) \to \mathbb{R}$ is *u.d.* mod 1 if

$$\lim_{T\to+\infty} \frac{\lambda\{\tau \in [0,T) : \langle f(\tau)\rangle \le s\}}{T} = s \quad \text{for all } s \in [0,1)\,;$$

a random variable X on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is *u.d.* mod 1 if

$$\mathbb{P}(\langle X\rangle \le s) = s \quad \text{for all } s \in [0,1)\,;$$

and a probability measure P on $(\mathbb{R}, \mathcal{B})$ is *u.d.* mod 1 if

$$P(\{x : \langle x\rangle \le s\}) = P\left(\bigcup\nolimits_{k\in\mathbb{Z}} [k, k+s]\right) = s \quad \text{for all } s \in [0,1)\,.$$

The next simple theorem (cf. [Di]) is one of the main tools in the theory of BL because it allows application of the powerful theory of uniform distribution mod 1. (Recall the convention $\log 0 := 0$.)

Theorem 4.2 (Uniform distribution characterization)**.** *A sequence of real numbers (respectively, a Borel measurable function, a random variable, a Borel probability measure) is Benford if and only if the decimal logarithm of its absolute value is uniformly distributed modulo* 1*.*

Proof. Let X be a random variable and, without loss of generality, assume that $\mathbb{P}(X=0)=0$. Then, for all $s\in[0,1)$,

$$\mathbb{P}(\langle \log|X|\rangle \le s) = \mathbb{P}\left(\log|X| \in \bigcup\nolimits_{k\in\mathbb{Z}}[k,k+s]\right) = \mathbb{P}\left(|X| \in \bigcup\nolimits_{k\in\mathbb{Z}}[10^k,10^{k+s}]\right)$$
$$= \mathbb{P}(S(X)\le 10^s)\,.$$

Hence, by Definitions 3.4 and 4.1, X is Benford if and only if $\mathbb{P}(S(X)\le 10^s) = \log 10^s = s$ for all $s\in[0,1)$, i.e., if and only if $\log|X|$ is u.d. mod 1.

The proofs for sequences, functions, and probability distributions are completely analogous. □

Next, several tools from the basic theory of uniform distribution mod 1 will be recorded that will be useful, via Theorem 4.2, in establishing the Benford property for many sequences, functions, and random variables.

Lemma 4.3. (i) *The sequence (x_n) is u.d.* mod 1 *if and only if the sequence (kx_n+b) is u.d.* mod 1 *for every $k\in\mathbb{Z}\backslash\{0\}$ and every $b\in\mathbb{R}$. Also, (x_n) is u.d.* mod 1 *if and only if (y_n) is u.d.* mod 1 *whenever $\lim_{n\to\infty}|y_n-x_n|=0$.*

(ii) *The function f is u.d.* mod 1 *if and only if $t\mapsto kf(t)+b$ is u.d.* mod 1 *for every non-zero integer k and every $b\in\mathbb{R}$.*

(iii) *The random variable X is u.d.* mod 1 *if and only if $kX+b$ is u.d.* mod 1 *for every non-zero integer k and every $b\in\mathbb{R}$.*

Proof. (i) The "if" part is obvious with $k=1$, $b=0$. For the "only if" part, assume that (x_n) is u.d. mod 1. Note first that

$$\lim\nolimits_{N\to\infty}\frac{\#\{1\le n\le N : \langle x_n\rangle \in C\}}{N} = \lambda_{0,1}(C)$$

holds whenever C is a finite union of intervals. Let $k\in\mathbb{Z}$ be non-zero and observe that, for any $0<s<1$,

$$\bigl\{x : \langle kx\rangle \le s\bigr\} = \begin{cases} \left\{x : \langle x\rangle \in \bigcup_{j=0}^{k-1}\left[\frac{j}{k},\frac{j+s}{k}\right]\right\} & \text{if } k>0\,, \\ \left\{x : \langle x\rangle \in \bigcup_{j=0}^{|k|-1}\left[\frac{j+1-s}{|k|},\frac{j+1}{|k|}\right]\right\} & \text{if } k<0\,. \end{cases}$$

Consequently,

$$\begin{aligned}\lim\nolimits_{N\to\infty}\frac{\#\{1\le n\le N : \langle kx_n\rangle \le s\}}{N} &= \begin{cases} \lambda_{0,1}\left(\bigcup_{j=0}^{k-1}\left[\frac{j}{k},\frac{j+s}{k}\right]\right) & \text{if } k>0\,, \\ \lambda_{0,1}\left(\bigcup_{j=0}^{|k|-1}\left[\frac{j+1-s}{|k|},\frac{j+1}{|k|}\right]\right) & \text{if } k<0\,, \end{cases} \\ &= \begin{cases} k\cdot\frac{s}{k} & \text{if } k>0 \\ |k|\cdot\frac{s}{|k|} & \text{if } k<0 \end{cases} \\ &= s\,,\end{aligned}$$

showing that (kx_n) is u.d. mod 1. Similarly, note that, for any $b, s \in (0,1)$,

$$\{x : \langle x+b\rangle \le s\} = \begin{cases} \{x : \langle x\rangle \in [0, s-b] \cup [1-b, 1)\} & \text{if } s \ge b\,, \\ \{x : \langle x\rangle \in [1-b, 1+s-b]\} & \text{if } s < b\,. \end{cases}$$

Thus, assuming without loss of generality that $0 < b < 1$,

$$\begin{aligned} \lim_{N\to\infty} \frac{\#\{1 \le n \le N : \langle x_n + b\rangle \le s\}}{N} &= \begin{cases} \lambda_{0,1}\big([0, s-b] \cup [1-b, 1)\big) & \text{if } s \ge b \\ \lambda_{0,1}\big([1-b, 1+s-b]\big) & \text{if } s < b \end{cases} \\ &= s\,, \end{aligned}$$

and hence $(x_n + b)$ is also u.d. mod 1. The second assertion is clear from the definition.

The proofs of (ii) and (iii) are completely analogous. $\square$

Example 4.4. **(i)** The sequence $(n\pi) = (\pi, 2\pi, 3\pi, \ldots)$ is u.d. mod 1, by Weyl's Equidistribution Theorem, see Proposition 4.8(i) below. Similarly, the sequence $(x_n) = (n\sqrt{2})$ is u.d. mod 1, whereas $(x_n\sqrt{2}) = (2n) = (2, 4, 6, \ldots)$ clearly is not, as $\langle 2n\rangle = 0$ for all n. Thus the requirement in Lemma 4.3(i) that k be an integer cannot be dropped.

For an analogous example using random variables, let X be uniform on $[0,2)$, that is $P_X = \lambda_{0,2}$. Then X is u.d. mod 1, but $X\sqrt{2}$ is not because

$$\mathbb{P}\big(\langle X\sqrt{2}\rangle \le s\big) = \begin{cases} \frac{3}{2\sqrt{2}} s & \text{if } s \in [0, 2\sqrt{2}-2)\,, \\ \frac{1}{\sqrt{2}} s + \frac{\sqrt{2}-1}{\sqrt{2}} & \text{if } s \in [2\sqrt{2}-2, 1)\,. \end{cases}$$

(ii) The sequence $(\log n)$ is not u.d. mod 1. A straightforward calculation shows that $\big(N^{-1}\#\{1 \le n \le N : \langle \log n\rangle \le s\}\big)_{N\in\mathbb{N}}$ has, for every $s \in [0,1)$,

$$\frac{1}{9}(10^s - 1) \quad \text{and} \quad \frac{10}{9}(1 - 10^{-s})$$

as its limit inferior and limit superior, respectively.

Example 4.5. **(i)** The function $f(t) = at + b$ with real a, b is u.d. mod 1 if and only if $a \ne 0$. Clearly, if $a = 0$ then f is constant and hence not u.d. mod 1. On the other hand, if $a > 0$ then $\langle a\tau + b\rangle \le s$ if and only if $\tau \in \big[\frac{k-b}{a}, \frac{k-b+s}{a}\big]$ for some $k \in \mathbb{Z}$. Note that each of the intervals $\big[\frac{k-b}{a}, \frac{k-b+s}{a}\big]$ has the same length $\frac{s}{a}$. Thus, given $T > 0$ and $s \in [0,1)$,

$$\frac{s}{a}(\lfloor aT\rfloor - 2) \le \lambda\big(\{\tau \in [0,T) : \langle a\tau + b\rangle \le s\}\big) \le \frac{s}{a}(\lfloor aT\rfloor + 2)\,,$$

and since $\lim_{T\to+\infty} \frac{s}{aT}(\lfloor aT\rfloor \pm 2) = s$, the function f is u.d. mod 1. The argument for the case $a < 0$ is similar.

As a consequence, although the function $f(t) = \alpha t$ is not Benford for any α, the function $f(t) = e^{\alpha t}$ is Benford whenever $\alpha \ne 0$, via Theorem 4.2, since $\log f(t) = \alpha t / \ln 10$ is u.d. mod 1, see Fig 7.

(ii) The function $f(t) = \log|at+b|$ is not u.d. mod 1 for any $a, b \in \mathbb{R}$. Indeed, if $a = 0$ then f is constant and hence not u.d. mod 1. On the other hand, for $a \neq 0$ essentially the same calculation as in Example 4.4(ii) above shows that, for every $s \in [0,1)$,

$$\liminf_{T\to+\infty} \frac{\lambda(\{\tau \in [0,T) : \langle \log|a\tau+b| \rangle \le s\})}{T} = \frac{1}{9}(10^s - 1)\,,$$

and

$$\limsup_{T\to+\infty} \frac{\lambda(\{\tau \in [0,T) : \langle \log|a\tau+b| \rangle \le s\})}{T} = \frac{10}{9}(1 - 10^{-s})\,.$$

Again, this implies that $f(t) = at + b$ is not Benford for any a, b.

Similarly, $f(t) = -\log(1+t^2)$ is not u.d. mod 1, and hence $f(t) = (1+t^2)^{-1}$ is not Benford, see Fig 7.

(iii) The function $f(t) = e^t$ is u.d. mod 1. To see this, let $T > 0$ and $N := \lfloor e^T \rfloor$, and recall that $t - \frac{1}{2}t^2 \le \ln(1+t) \le t$ for all $t \ge 0$. Given $0 \le s < 1$, it follows from

$$\lambda\left(\{\tau \in [0,T) : \langle e^\tau \rangle \le s\}\right) = \sum\nolimits_{n=1}^{N-1} \ln\left(1 + \frac{s}{n}\right) + (T - \ln N)$$

that

$$\begin{aligned}\frac{s\sum_{n=1}^{N-1} n^{-1} - \frac{1}{2}s^2 \sum_{n=1}^{N-1} n^{-2}}{\ln(N+1)} &\le \frac{\lambda\left(\{\tau \in [0,T) : \langle e^\tau \rangle \le s\}\right)}{T} \\ &\le \frac{s\sum_{n=1}^{N-1} n^{-1} + \ln(1+N^{-1})}{\ln N}\,,\end{aligned}$$

and hence indeed $\lim_{T\to+\infty} T^{-1}\lambda\left(\{\tau \in [0,T) : \langle e^\tau \rangle \le s\}\right) = s$.

As a consequence, the super-exponential function $f(t) = e^{e^{\alpha t}}$ is also Benford if $\alpha \neq 0$.

(iv) For the function $f(t) = \sin^2 t$, it is straightforward to check that, given any $0 \le s < 1$,

$$\lim_{T\to+\infty} \frac{\lambda(\{\tau \in [0,T) : \langle \sin^2 \tau \rangle \le s\})}{T} = \frac{2}{\pi}\arcsin\sqrt{s}\,.$$

Thus, asymptotically $\langle f \rangle$ is not uniform on $[0,1)$ but rather arcsin-distributed, see Example 3.10(iii).

(v) For the function $f(t) = \log(\sin^2 t)$, it follows from (iv) that the asymptotic distribution of $\langle f \rangle$ has the density

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{2}{\pi}\sum\nolimits_{n=1}^{\infty}\left(\arcsin 10^{(s-n)/2} - \arcsin 10^{-n/2}\right)\right) &= \frac{\ln 10}{\pi}\sum\nolimits_{n=1}^{\infty} \frac{1}{\sqrt{10^{n-s}-1}} \\ &> \frac{\ln 10}{\pi}\cdot\frac{10^{s/2}}{10^{1/2}-1}\,,\end{aligned}$$

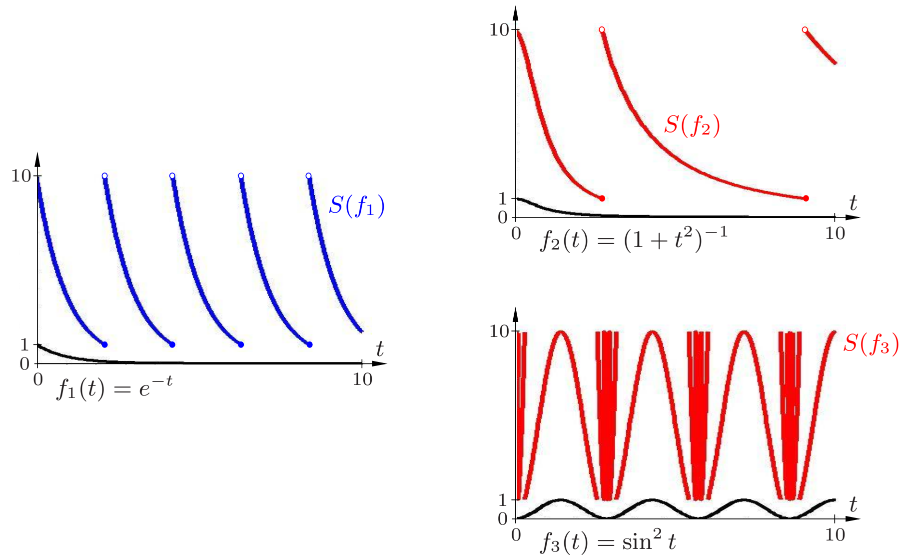

FIG 7. *While the function f_1 is Benford, the functions f_2, f_3 are not, see Example 4.5.*

for $0 \le s < 1$. Thus clearly f is not u.d. mod 1, showing that $t \mapsto \sin^2 t$ is not Benford, see FIG 7.

Example 4.6. **(i)** If the random variable X is uniformly distributed on $[0, 2)$ then it is clearly u.d. mod 1. However, if X is uniform on, say, $[0, \pi)$ then X is not u.d. mod 1.

(ii) No exponential random variable is u.d. mod 1. Specifically, let X be an exponential random variable with mean σ, i.e.

$$F_X(t) = \max(0, 1 - e^{-t/\sigma}), \quad t \in \mathbb{R}.$$

Hence $\operatorname{var} X = \sigma^2$. For every $l \ge 0$,

$$\begin{aligned} \mathbb{P}(l \le X < l + \tfrac{1}{2}) &= F_X(l + \tfrac{1}{2}) - F_X(l) \\ &> F_X(l+1) - F_X(l + \tfrac{1}{2}) = \mathbb{P}(l + \tfrac{1}{2} \le X < l + 1), \end{aligned}$$

and since $\sum_{l=0}^{\infty} \mathbb{P}(l \le X < l+1) = 1$, this implies that

$$\mathbb{P}(\langle X \rangle < \tfrac{1}{2}) = \sum\nolimits_{l=0}^{\infty} \mathbb{P}(l \le X < l + \tfrac{1}{2}) > \tfrac{1}{2},$$

showing that X is not u.d. mod 1. To obtain more explicit information, observe that, for every $0 \le s < 1$,

$$F_{\langle X \rangle}(s) = \mathbb{P}(\langle X \rangle \le s) = \sum\nolimits_{l=0}^{\infty} \big(F_X(l+s) - F_X(l)\big) = \frac{1 - e^{-s/\sigma}}{1 - e^{-1/\sigma}},$$

from which it follows via a straightforward calculation that

$$\max_{0\le s<1}\left|F_{\langle X\rangle}(s)-s\right| = \frac{1}{e^{1/\sigma}-1} - \sigma + \sigma\ln(\sigma e^{1/\sigma}-\sigma) =: R_{\rm ii}(\sigma)\,.$$

Note that $R_{\rm ii}(1) = \ln(e-1) - \frac{e-2}{e-1} = 0.1233\ldots < \frac{1}{8}$. Moreover,

$$R_{\rm ii}(\sigma) = \frac{1}{8\sigma} + \mathcal{O}(\sigma^{-2}) \quad \text{as } \sigma \to +\infty\,,$$

which shows that even though X is not u.d. mod 1, the deviation of $\langle X\rangle$ from uniform is small for large σ. As a consequence, 10^X resembles a Benford random variable ever more closely as $\sigma \to +\infty$.

(iii) If X is a normal random variable then X is not u.d. mod 1, and neither is $|X|$ or $\max(0,X)$. While this is easily checked by a direct calculation as in (ii), it is again illuminating to obtain more quantitative information. To this end, assume that X is a normal variable with mean 0 and variance σ^2. By means of Fourier series [Pi], it can be shown that, for every $0 \le s < 1$,

$$F_{\langle X\rangle}(s) - s = \sum\nolimits_{n=1}^{\infty} \frac{\sin(2\pi n s)}{\pi n} e^{-2\sigma^2\pi^2 n^2}\,.$$

From this, it follows that

$$R_{\rm iii}(\sigma) := \max_{0\le s<1}\left|F_{\langle X\rangle}(s)-s\right| \le \frac{1}{\pi}\sum\nolimits_{n=1}^{\infty} n^{-1} e^{-2\sigma^2\pi^2 n^2}\,,$$

and hence in particular

$$R_{\rm iii}(\sigma) = \frac{e^{-2\sigma^2\pi^2}}{\pi} + \mathcal{O}(e^{-8\sigma^2\pi^2}) \quad \text{as } \sigma \to +\infty\,,$$

showing that $R_{\rm iii}(\sigma)$, the deviation of $\langle X\rangle$ from uniformity, goes to zero very rapidly as $\sigma \to +\infty$. Already for $\sigma = 1$ one finds that $R_{\rm iii}(1) < 8.516 \cdot 10^{-10}$. Thus even though a standard normal random variable X is not u.d. mod 1, the distribution of $\langle X\rangle$ is extremely close to uniform. Consequently, a *log-normal* random variable with large variance is practically indistinguishable from a Benford random variable.

Corollary 4.7. (i) *A sequence* (x_n) *is Benford if and only if, for all* $\alpha \in \mathbb{R}$ *and* $k \in \mathbb{Z}$ *with* $\alpha k \neq 0$*, the sequence* (αx_n^k) *is also Benford.*

(ii) *A function* $f : [0,+\infty) \to \mathbb{R}$ *is Benford if and only if* $1/f$ *is Benford.*

(iii) *A random variable* X *is Benford if and only if* $1/X$ *is Benford.*

The next two statements, recorded here for ease of reference, list several key tools concerning uniform distribution mod 1, which via Theorem 4.2 will be used to determine Benford properties of sequences, functions, and random variables. Conclusion (i) in Proposition 4.8 is Weyl's classical uniform distribution result [KN, Thm.3.3], conclusion (ii) is an immediate consequence of Weyl's criterion [KN, Thm.2.1], conclusion (iii) is [Ber2, Lem.2.8], and conclusion (iv) is [BBH, Lem.2.4.(i)].

Proposition 4.8. *Let (x_n) be a sequence of real numbers.*

(i) *If $\lim_{n\to\infty}(x_{n+1}-x_n)=\theta$ for some irrational θ, then (x_n) is u.d.* mod 1.

(ii) *If (x_n) is periodic, i.e. $x_{n+p}=x_n$ for some $p\in\mathbb{N}$ and all n, then $(n\theta+x_n)$ is u.d.* mod 1 *if and only if θ is irrational.*

(iii) *The sequence (x_n) is u.d.* mod 1 *if and only if $(x_n+\alpha\log n)$ is u.d.* mod 1 *for all $\alpha\in\mathbb{R}$.*

(iv) *If (x_n) is u.d.* mod 1 *and non-decreasing, then $(x_n/\log n)$ is unbounded.*

The converse of (i) is not true in general: (x_n) may be u.d. mod 1 even if $(x_{n+1}-x_n)$ has a rational limit. Also, in (ii) the sequence $(n\theta)$ cannot be replaced by an arbitrary uniformly distributed sequence (θ_n), i.e. (θ_n+x_n) may not be u.d. mod 1 even though (θ_n) is u.d. mod 1 and (x_n) is periodic.

Another very useful result is Koksma's metric theorem [KN, Thm.4.3]. For its formulation, recall that a property of real numbers is said to hold for *almost every* (*a.e.*) $x\in[a,b)$ if there exists a set $N\in\mathcal{B}[a,b)$ with $\lambda_{a,b}(N)=0$ such that the property holds for every $x\notin N$. The probabilistic interpretation of a given property of real numbers holding for a.e. x is that this property holds *almost surely* (*a.s.*), i.e. with probability one, *for every random variable that has a density* (i.e., is absolutely continuous).

Proposition 4.9. *Let f_n be continuously differentiable on $[a,b]$ for all $n\in\mathbb{N}$. If $f'_m-f'_n$ is monotone and $|f'_m(x)-f'_n(x)|\ge\alpha>0$ for all $m\ne n$, where α does not depend on x, m and n, then $\big(f_n(x)\big)$ is u.d.* mod 1 *for almost every $x\in[a,b]$.*

Theorem 4.10 ([BHKR])**.** *If a,b,α,β are real numbers with $a\ne 0$ and $|\alpha|>|\beta|$ then $(\alpha^n a+\beta^n b)$ is Benford if and only if $\log|\alpha|$ is irrational.*

Proof. Since $a\ne 0$ and $|\alpha|>|\beta|$, $\lim_{n\to\infty}\frac{\beta^n b}{\alpha^n a}=0$, and therefore

$$\log|\alpha^n a+\beta^n b|-\log|\alpha^n a|=\log\left|1+\frac{\beta^n b}{\alpha^n a}\right|\to 0\,,$$

showing that $(\log|\alpha^n a+\beta^n b|)$ is u.d. mod 1 if and only if $(\log|\alpha^n a|)=(\log|a|+n\log|\alpha|)$ is. According to Proposition 4.8(i), this is the case whenever $\log|\alpha|$ is irrational. On the other hand, if $\log|\alpha|$ is rational then $\langle\log|a|+n\log|\alpha|\rangle$ attains only finitely many values and hence $(\log|a|+n\log|\alpha|)$ is not u.d. mod 1. An application of Theorem 4.2 therefore completes the proof. $\square$

Example 4.11. (i) By Theorem 4.10 the sequence (2^n) is Benford since $\log 2$ is irrational, but (10^n) is not Benford since $\log 10=1\in\mathbb{Q}$. Similarly, (0.2^n), (3^n), (0.3^n), $\big(0.01\cdot 0.2^n+0.2\cdot 0.01^n\big)$ are Benford, whereas (0.1^n), $\big(\sqrt{10}^{\,n}\big)$, $\big(0.1\cdot 0.02^n+0.02\cdot 0.1^n\big)$ are not.

(ii) The sequence $\big(0.2^n+(-0.2)^n\big)$ is not Benford, since all odd terms are zero, but $\big(0.2^n+(-0.2)^n+0.03^n\big)$ is Benford — although this does not follow directly from Theorem 4.10.

(iii) By Proposition 4.9, the sequence $(x, 2x, 3x, \ldots) = (nx)$ is u.d. mod 1 for almost every real x, but clearly not for *every* x, as for example $x = 1$ shows. Consequently, by Theorem 4.2, (10^{nx}) is Benford for almost all real x, but not e.g. for $x = 1$ or, more generally, whenever x is rational.

(iv) By Proposition 4.8(iv) or Example 4.4(ii), the sequence $(\log n)$ is not u.d. mod 1, so the sequence (n) of positive integers is not Benford, and neither is (αn) for any $\alpha \in \mathbb{R}$, see also FIG 8.

(v) Consider the sequence (p_n) of prime numbers. By the Prime Number Theorem, $p_n = \mathcal{O}(n \log n)$ as $n \to \infty$. Hence it follows from Proposition 4.8(iv) that (p_n) is not Benford, see FIG 8

Example 4.12. Consider the sequence $(F_n) = (1, 1, 2, 3, 5, 8, 13, \ldots)$ of Fibonacci numbers, defined inductively as $F_{n+2} = F_{n+1} + F_n$ for all $n \in \mathbb{N}$, with $F_1 = F_2 = 1$. It is well known (and easy to check) that

$$F_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right) = \frac{1}{\sqrt{5}} \left(\varphi^n - (-\varphi^{-1})^n \right) \quad \text{for all } n \in \mathbb{N}\,,$$

where $\varphi = \frac{1}{2}(1 + \sqrt{5}) \approx 1.618$. Since $\varphi > 1$ and $\log \varphi$ is irrational, (F_n) is Benford, by Theorem 4.10, see also FIG 8. Sequences such as (F_n) which are generated by linear recurrence relations will be studied in detail in Section 5.2.

Theorem 4.13. *Let X, Y be random variables. Then:*

(i) *If X is u.d.* mod 1 *and Y is independent of X, then $X + Y$ is u.d.* mod 1.

(ii) *If $\langle X \rangle$ and $\langle X + \alpha \rangle$ have the same distribution for some irrational α then X is u.d.* mod 1.

(iii) *If (X_n) is an i.i.d. sequence of random variables and X_1 is not purely atomic (i.e. $\mathbb{P}(X_1 \in C) < 1$ for every countable set $C \subset \mathbb{R}$), then*

$$\lim\nolimits_{n\to\infty} \mathbb{P}\left(\left\langle \sum\nolimits_{j=1}^n X_j \right\rangle \le s \right) = s \quad \text{for every } 0 \le s < 1\,, \tag{4.1}$$

that is, $\left\langle \sum_{j=1}^n X_j \right\rangle \to U(0,1)$ in distribution as $n \to \infty$.

Proof. The proof is most transparently done by means of some elementary Fourier analysis. To this end, for any random variable Z with values in $[0, 1)$, or equivalently for the associated probability measure P_Z on $\big([0, 1), \mathcal{B}[0, 1)\big)$, let

$$\begin{aligned} \widehat{P_Z}(k) = \mathbb{E}(e^{2\pi \imath k Z}) &= \int_0^1 e^{2\pi \imath k s} \mathrm{d}P_Z(s) \\ &= \int_0^1 \cos(2\pi k s)\, \mathrm{d}P_Z(s) + \imath \int_0^1 \sin(2\pi k s)\, \mathrm{d}P_Z(s)\,, \quad k \in \mathbb{Z}\,. \end{aligned}$$

The bi-infinite sequence $\big(\widehat{P_Z}(k)\big)_{k\in\mathbb{Z}}$, referred to as the *Fourier* (or *Fourier–Stieltjes*) *coefficients* of Z or P_Z, is a bounded sequence of complex numbers,

$$(F_n) = (1, 1, 2, 3, 5, 8, 13, \dots)$$
$$\big(S(F_n)\big) = (1, 1, 2, 3, 5, 8, 1.3, \dots)$$

$$(2n) = (2, 4, 6, 8, 10, 12, 14, \dots)$$
$$\big(S(2n)\big) = (2, 4, 6, 8, 1, 1.2, 1.4, \dots)$$

$$(p_n) = (2, 3, 5, 7, 11, 13, 17, \dots)$$
$$\big(S(p_n)\big) = (2, 3, 5, 7, 1.1, 1.3, 1.7, \dots)$$

$$(x_n) = (\lfloor 10^{(n+2)/5} \rfloor) = (3, 6, 10, 15, 25, 39, 63, \dots)$$
$$\big(S(x_n)\big) = (3, 6, 1, 1.5, 2.5, 3.9, 6.3, \dots)$$

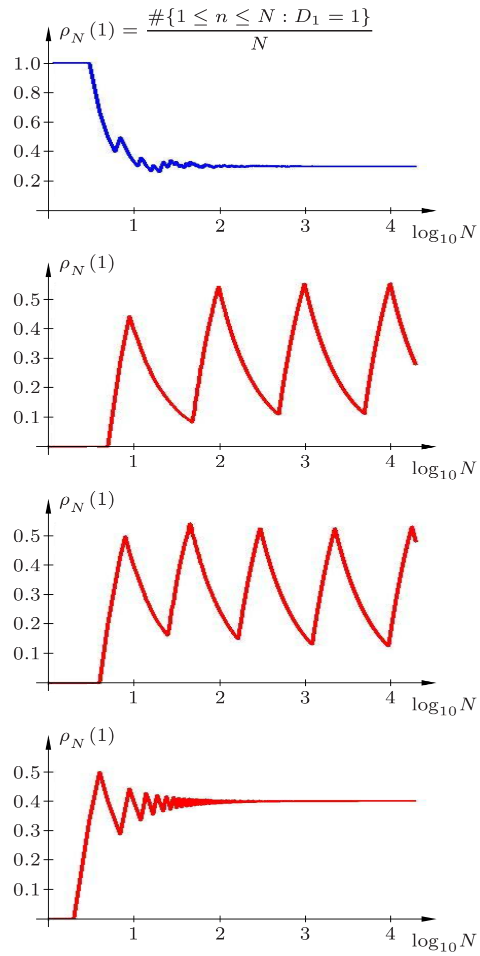

FIG 8. *For a Benford sequence, $\lim_{N\to\infty} \rho_N(1) = \log 2$. Thus if $\big(\rho_N(1)\big)_{N\in\mathbb{N}}$ does not converge (center) or has a different limit (bottom), then the sequence in question is not Benford, see also Example 4.11.*

with $|\widehat{P_Z}(k)| \le 1$ for all $k \in \mathbb{Z}$, and $\widehat{P_Z}(0) = 1$. The three single most important properties of Fourier coefficients are that $\big(\widehat{P_Z}(k)\big)_{k\in\mathbb{Z}}$ uniquely determines P_Z, i.e. $P_{Z_1} = P_{Z_2}$ whenever $\widehat{P_{Z_1}}(k) = \widehat{P_{Z_2}}(k)$ for all $k \in \mathbb{Z}$; that $\widehat{P_{\langle Z_1+Z_2\rangle}}(k) = \widehat{P_{Z_1}}(k) \cdot \widehat{P_{Z_2}}(k)$ for all k, provided that Z_1 and Z_2 are independent; and that $Z_n \to Z$ in distribution if and only if $\lim_{n\to\infty} \widehat{P_{Z_n}}(k) = \widehat{P_Z}(k)$ for every k, see e.g. [CT] for an authoritative discussion of this material. Also note that the sequence of Fourier coefficients is extremely simple if Z is uniform, i.e. for $Z = U(0,1)$, namely

$$\widehat{P_{U(0,1)}}(k) = \widehat{\lambda_{0,1}}(k) = \begin{cases} 1 & \text{if } k = 0\,, \\ 0 & \text{otherwise.} \end{cases}$$

With these preparations, the proof of the theorem is very short indeed.

(i) Since $\widehat{P_{\langle X\rangle}}(k) = 0$ for all $k \neq 0$,

$$\widehat{P_{\langle X+Y\rangle}}(k) = \widehat{P_X}(k) \cdot \widehat{P_Y}(k) = 0\,,$$

which in turn shows that $\langle X+Y\rangle = U(0,1)$, i.e. $X+Y$ is u.d. mod 1.

(ii) Note that if $Z = \alpha$ with probability one then $\widehat{P_Z}(k) = e^{2\pi\imath k\alpha}$ for every $k \in \mathbb{Z}$. Consequently, if $\langle X\rangle$ and $\langle X+\alpha\rangle$ have the same distribution then

$$\widehat{P_{\langle X\rangle}}(k) = \widehat{P_{\langle X+\alpha\rangle}}(k) = e^{2\pi\imath k\alpha}\widehat{P_{\langle X\rangle}}(k)$$

for every $k \in \mathbb{Z}$. If α is irrational then $e^{2\pi\imath k\alpha} \neq 1$ for all $k \neq 0$, implying that $\widehat{P_{\langle X\rangle}}(k) = 0$. Thus $\widehat{P_{\langle X\rangle}} = \widehat{\lambda_{0,1}}$ and hence $P_{\langle X\rangle} = \lambda_{0,1}$, i.e. $\langle X\rangle = U(0,1)$.

(iii) Assume that $X_1, X_2, \ldots$ are independent and all have the same distribution. Then, for every $k \in \mathbb{Z}$ and $n \in \mathbb{N}$,

$$\widehat{P_{\langle X_1+\ldots+X_n\rangle}}(k) = \left(\widehat{P_{\langle X_1\rangle}}(k)\right)^n\,.$$

Recall that $|\widehat{P_{\langle X_1\rangle}}(k)| \le 1$. Thus $\widehat{P_{\langle X_1+\ldots+X_n\rangle}}(k) \to 0$ as $n \to \infty$, and hence $\langle X_1 + \ldots + X_n\rangle \to U(0,1)$ in distribution, unless $|\widehat{P_{\langle X_1\rangle}}(k_0)| = 1$ for some non-zero integer k_0. In the latter case, let $\widehat{P_{\langle X_1\rangle}}(k_0) = e^{2\pi\imath\theta}$ with the appropriate $\theta \in [0,1)$. It follows from

$$\begin{aligned} 0 = 1 - e^{-2\pi\imath\theta}\widehat{P_{\langle X_1\rangle}}(k_0) &= 1 - \widehat{P_{\langle X_1-\theta/k_0\rangle}}(k_0) \\ &= \int_0^1 \bigl(1-\cos(2\pi k_0 s)\bigr)\,\mathrm{d}P_{\langle X_1-\theta/k_0\rangle}(s) \ge 0\,, \end{aligned}$$

that $\cos(2\pi k_0\langle X_1 - \theta/k_0\rangle) = \cos\bigl(2\pi(k_0X-\theta)\bigr) = 1$ with probability one. Hence $\mathbb{P}(k_0X_1 \in \theta + \mathbb{Z}) = 1$, and X_1 is purely atomic. (In fact, X_1 is concentrated on a lattice $\{a + k/|k_0| : k \in \mathbb{Z}\}$ with the appropriate $a > 0$.) $\square$

Example 4.14. **(i)** Let (X_n) be an i.i.d. sequence of Cauchy random variables, i.e.

$$f_{X_1}(t) = \frac{1}{\pi(1+t^2)}\,, \quad t \in \mathbb{R}\,.$$

It is well known, or readily checked by a direct calculation, that $\frac{1}{n}\sum_{j=1}^n X_j$ is again Cauchy. Thus

$$f_{\langle\sum_{j=1}^n X_j\rangle}(s) = \frac{1}{\pi}\sum\nolimits_{k\in\mathbb{Z}} \frac{n}{n^2+(s+k)^2}\,, \quad 0 \le s < 1\,,$$

from which it follows that, uniformly in s,

$$f_{\langle\sum_{j=1}^n X_j\rangle}(s) = \frac{1}{\pi n}\sum\nolimits_{k\in\mathbb{Z}} \frac{1}{1+\bigl((s+k)/n\bigr)^2} \;\to\; \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{\mathrm{d}t}{1+t^2} = 1 \quad \text{as } n \to \infty\,.$$

As asserted by Theorem 4.13, therefore, for every $0 \le s < 1$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left\langle \sum\nolimits_{j=1}^{n} X_j \right\rangle \le s\right) = \lim_{n\to\infty} \int_0^s f_{\langle \sum_{j=1}^{n} X_j\rangle}(\sigma)\,\mathrm{d}\sigma = \int_0^s 1\,\mathrm{d}\sigma = s\,.$$

(ii) Consider an i.i.d. sequence (X_n) where $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = \sqrt{2}) = \frac{1}{2}$. In this case

$$\widehat{P_{X_1}}(k) = \tfrac{1}{2}\left(1 + e^{2\pi\imath k\sqrt{2}}\right) = e^{\pi\imath k\sqrt{2}}\cos(\pi k\sqrt{2})\,, \quad k \in \mathbb{Z}\,.$$

Note that $|\widehat{P_{X_1}}(k)| = |\cos(\pi k\sqrt{2})| < 1$ for all $k \ne 0$. Hence $\widehat{P_{\langle\sum_{j=1}^{n} X_j\rangle}}(k) = \widehat{P_{\langle X_1\rangle}}(k)^n \to 0$ as $n \to \infty$, which in turn shows that (4.1) holds, even though X_1 is purely atomic.

On the other hand, if $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = \frac{1}{2}) = \frac{1}{2}$ then X_1 is also purely atomic, but

$$\mathbb{P}\left(\sum\nolimits_{j=1}^{n} X_j = \tfrac{1}{2}l\right) = 2^{-n}\binom{n}{l} \quad \text{for all } n \in \mathbb{N}, l = 0, 1, \ldots, n\,,$$

and consequently, for every n,

$$\mathbb{P}\left(\left\langle \sum\nolimits_{j=1}^{n} X_j \right\rangle = 0\right) = \sum\nolimits_{l=0,\, l \text{ even}}^{n} 2^{-n}\binom{n}{l} = \frac{1}{2}\,,$$

showing that (4.1) does not hold in this case. Correspondingly, $\widehat{P_{X_1}}(k) = \frac{1}{2}\big(1 + (-1)^k\big)$, and so $\widehat{P_{X_1}}(k) = 1$ whenever k is even.

A careful inspection of the above proof shows that, in the setting of Theorem 4.13(iii), (4.1) holds if and only if $\mathbb{P}(X_1 \in a + \frac{1}{m}\mathbb{Z}) < 1$ for every $a \in \mathbb{R}$ and $m \in \mathbb{N}$. While the "if" part has been proved above, for the "only if" part simply note that if $\mathbb{P}(X_1 \in a + \frac{1}{m}\mathbb{Z}) = 1$ for some $a \in \mathbb{R}$ and $m \in \mathbb{N}$ then $\langle X_1 + \ldots + X_n\rangle$ is, for every $n \in \mathbb{N}$ and possibly up to a rotation, concentrated on the set $\{0, \frac{1}{m}, \ldots, \frac{m-1}{m}\} = \langle \frac{1}{m}\mathbb{Z}\rangle$ and hence does not converge in distribution to $U(0,1)$.

None of the familiar classical probability distributions or random variables, such as e.g. normal, uniform, exponential, beta, binomial, or gamma distributions are Benford. Specifically, *no* uniform distribution is even close to BL, no matter how large its range or how it is centered. This statement can be quantified explicitly as follows.

Proposition 4.15 ([Ber5, BH3])**.** *For every uniformly distributed random variable X,*

$$\max_{0\le s<1} \left|F_{\langle \log X\rangle}(s) - s\right| \ge \frac{-9 + \ln 10 + 9\ln 9 - 9\ln\ln 10}{18\ln 10} = 0.1334\ldots\,,$$

and this bound is sharp.

Similarly, all exponential and normal random variables are uniformly bounded away from BL, as is explained in detail in [BH3]. However, as the following example shows, some distributions do come fairly close to being Benford.

Example 4.16. **(i)** Let X be exponential with mean 1, that is

$$F_X(t) = \max(0, 1 - e^{-t}), \quad t \in \mathbb{R}.$$

An explicit calculation shows that, for every $1 \le t < 10$,

$$\mathbb{P}(S(X) \le t) = \sum\nolimits_{k\in\mathbb{Z}} \bigl(F_X(10^k t) - F_X(10^k)\bigr) = \sum\nolimits_{k\in\mathbb{Z}} \left(e^{-10^k} - e^{-10^k t}\right).$$

Since $\mathbb{P}(S(X) \le t) \not\equiv \log t$, the random variable X is not Benford. Numerically, one finds $\max_{1\le t<10} |\mathbb{P}(S(X) < t) - \log t| < 3.054 \cdot 10^{-2}$, see also Fig 9. Thus even though X is not exactly Benford, it is close to being Benford in the sense that $|\mathbb{P}(S(X) \le t) - \log t|$ is small for all $t \in [1, 10)$.

(ii) Let X be standard normal. Then, for every $t \in [1, 10)$,

$$\mathbb{P}(S(X) \le t) = \sum\nolimits_{k\in\mathbb{Z}} \bigl(\Phi(10^k t) - \Phi(10^k)\bigr),$$

where Φ is the distribution function of X, that is

$$\Phi(t) = F_X(t) = \mathbb{P}(X \le t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}\tau^2} \mathrm{d}\tau, \quad t \in \mathbb{R}.$$

Numerically, one finds $\max_{1\le t\le 10} |\mathbb{P}(S(X) < t) - \log t| < 6.052 \cdot 10^{-2}$. Though larger than in the exponential case, the deviation of X from BL is still rather small.

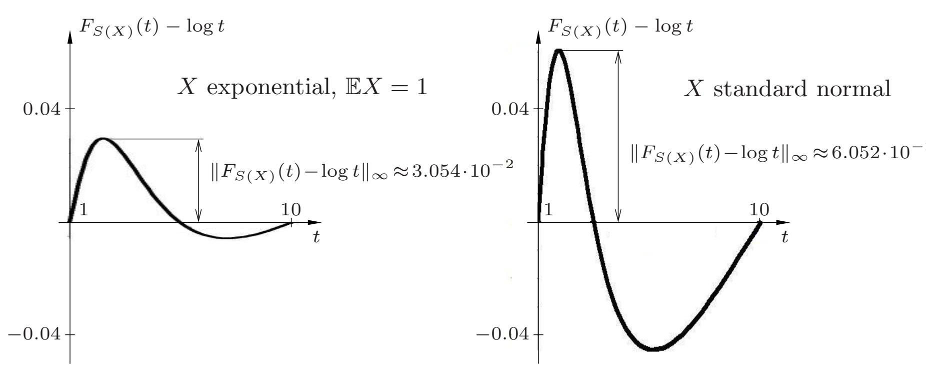

Fig 9. *For standard exponential (left) and normal random variables X, the distribution of $S(X)$ deviates from BL only slightly. Note, however, that* non-*standard normal variables can be far from BL, e.g., if $\mathbb{E}X = 75$ and $\operatorname{var} X = 1$ then $D_1(X) = 7$ with very high probability.*

The next result says that every random variable X with a density is asymptotically uniformly distributed on lattices of intervals as the size of the intervals goes to zero. Equivalently, $\langle nX\rangle$ is asymptotically uniform, as $n \to \infty$. This result has been the basis for several recent fallacious arguments claiming that if a random variable X has a density with a very large "spread" then $\log X$ must also have a density with large spread and thus, by the theorem, must be close to u.d. mod 1, implying in turn that X must be close to Benford (cf. [Fel, Few]). The error in those arguments is that, regardless of which notion of "spread" is used, the variable X may have large spread and at the same time the variable $\log X$ may have small spread; for details, the reader is referred to [BH3].

Theorem 4.17. *If X has a density then*

$$\lim_{n\to\infty} \mathbb{P}(\langle nX\rangle \le s) = s \quad \text{for all } 0 \le s < 1\,, \tag{4.2}$$

that is, $\langle nX\rangle \to U(0,1)$ in distribution as $n \to \infty$.

Proof. Since $\langle nX\rangle = \big\langle n\langle X\rangle\big\rangle$, it can be assumed that X only takes values in $[0,1)$. Let f be the density of X, i.e. $f : [0,1] \to \mathbb{R}$ is a non-negative measurable function with $\mathbb{P}(X \le s) = \int_0^s f(\sigma)\,\mathrm{d}\sigma$ for all $s \in [0,1)$. From

$$\begin{aligned}\mathbb{P}(\langle nX\rangle \le s) &= \mathbb{P}\left(X \in \bigcup\nolimits_{l=0}^{n-1}\left[\frac{l}{n}, \frac{l+s}{n}\right]\right) = \sum\nolimits_{l=0}^{n-1}\int_{l/n}^{(l+s)/n} f(\sigma)\,\mathrm{d}\sigma \\ &= \int_0^s \frac{1}{n}\sum\nolimits_{l=0}^{n-1} f\left(\frac{l+\sigma}{n}\right)\mathrm{d}\sigma\,,\end{aligned}$$

it follows that the density of $\langle nX\rangle$ is given by

$$f_{\langle nX\rangle}(s) = \frac{1}{n}\sum\nolimits_{l=0}^{n-1} f\left(\frac{l+s}{n}\right), \quad 0 \le s < 1\,.$$

Note that if f is continuous, or merely Riemann integrable, then, as $n \to \infty$,

$$f_{\langle nX\rangle}(s) \to \int_0^1 f(\sigma)\,\mathrm{d}\sigma = 1 \quad \text{for all } s \in [0,1)\,.$$

In general, given any $\varepsilon > 0$ there exists a continuous density g_ε such that $\int_0^1 |f(\sigma) - g_\varepsilon(\sigma)|\,\mathrm{d}\sigma < \varepsilon$ and hence

$$\begin{aligned}\int_0^1 |f_{\langle nX\rangle}(\sigma) - 1|\,\mathrm{d}\sigma &\le \int_0^1 \left|\frac{1}{n}\sum\nolimits_{l=0}^{n-1} f\left(\frac{l+\sigma}{n}\right) - \frac{1}{n}\sum\nolimits_{l=0}^{n-1} g_\varepsilon\left(\frac{l+\sigma}{n}\right)\right|\mathrm{d}\sigma \\ &\quad + \int_0^1\left|\frac{1}{n}\sum\nolimits_{l=0}^{n-1} g_\varepsilon\left(\frac{l+\sigma}{n}\right) - 1\right|\mathrm{d}\sigma \\ &\le \int_0^1 |f(\sigma) - g_\varepsilon(\sigma)|\,\mathrm{d}\sigma + \int_0^1\left|\frac{1}{n}\sum\nolimits_{l=0}^{n-1} g_\varepsilon\left(\frac{l+\sigma}{n}\right) - \int_0^1 g(\tau)\,\mathrm{d}\tau\right|\mathrm{d}\sigma\,,\end{aligned}$$

which in turn shows that

$$\limsup\nolimits_{n\to\infty}\int_0^1 |f_{\langle nX\rangle}(\sigma) - 1|\,\mathrm{d}\sigma \le \varepsilon\,,$$

and since $\varepsilon > 0$ was arbitrary, $\int_0^1 |f_{\langle nX\rangle}(\sigma) - 1|\,\mathrm{d}\sigma \to 0$ as $n \to \infty$. From this, the claim follows immediately because, for every $0 \le s < 1$,

$$\big|\mathbb{P}(\langle nX\rangle \le s) - s\big| = \left|\int_0^s (f_{\langle nX\rangle}(\sigma) - 1)\,\mathrm{d}\sigma\right| \le \int_0^1 |f_{\langle nX\rangle}(\sigma) - 1|\,\mathrm{d}\sigma \;\; \to \;\; 0\,.$$

□

Remark. If X does not have a density, then (4.2) may not hold. Trivially, if X is an integer with probability one then $\mathbb{P}(\langle nX\rangle \le s) = 1$ for every n and $0 \le s < 1$. Hence (4.2) fails. For a simple continuous example, let X be uniformly distributed on the classical Cantor middle thirds set. In more probabilistic terms, $X = 2\sum_{j=1}^{\infty} 3^{-j}X_j$ where the X_j are i.i.d. with $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = \frac{1}{2}$. Then $P_X \ne \lambda_{0,1}$ but $\langle 3X\rangle$ has the same distribution as X, and so has $\langle 3^n X\rangle$ for every $n \in \mathbb{N}$. Thus (4.2) fails again.

In fact, using the Fourier analysis tools introduced in the proof of Theorem 4.13, together with the observation that

$$\widehat{P_{\langle nX\rangle}}(k) = \widehat{P_{\langle X\rangle}}(nk) \quad \text{for all } n \in \mathbb{N},\ k \in \mathbb{Z}\,,$$

it is clear that (4.2) holds if and only if X has the property that $\widehat{P_{\langle X\rangle}}(k) \to 0$ as $|k| \to \infty$, i.e. precisely if $P_{\langle X\rangle}$ is a so-called *Rajchman probability*. As Theorem 4.17 shows, a probability on $[0, 1)$ is Rajchman whenever it is a.c. (In advanced calculus, this fact is usually referred to as the *Riemann–Lebesgue Lemma*.) The converse is not true, i.e., there exist Rajchman probabilities on $[0, 1)$ that are not a.c., see [Ly]. ♣

4.2. The scale-invariance characterization

One popular hypothesis often related to BL is that of *scale-invariance*. Informally put, scale-invariance captures the intuitively attractive notion that any universal law should be independent of units. For instance, if a sufficiently large aggregation of data is converted from meters to feet, US$ to € etc., then while the individual numbers change, the statements about the overall distribution of significant digits should not be affected by this change. R. Pinkham [Pi] credits R. Hamming with the idea of scale-invariance, and attempts to prove that the Benford distribution is the only scale-invariant distribution. Pinkham's argument has subsequently been used by numerous authors to explain the appearance of BL in many real-life data, by arguing that the data in question should be invariant under changes of scale and thus must be Benford.

Although this scale-invariance conclusion is correct in the proper setting, see Theorem 4.20 below, Pinkham's argument contains a fatal error. As Knuth [Kn] observes, the error is Pinkham's implicit assumption that there is a scale-invariant Borel probability measure on $\mathbb{R}^+$, when in fact such a probability measure does not exist, cf. [Ra1]. Indeed, the only real-valued random variable X that is *scale-invariant*, i.e., X and αX have the same distribution for all scaling factors $\alpha > 0$, is the random variable that is constant equal to zero, that is $\mathbb{P}(X = 0) = 1$. Clearly, any such random variable is scale-invariant since $X = \alpha X$ with probability one. To see that this is the only scale-invariant random variable, suppose that $\mathbb{P}(|X| > c) = \delta > 0$ for some $c > 0$. Then $\mathbb{P}(|\alpha X| > c) = \mathbb{P}(|X| > c/\alpha) \searrow 0$ as $\alpha \searrow 0$, so for sufficiently small positive α, $\mathbb{P}(|\alpha X| > c) < \delta = \mathbb{P}(|X| > c)$, contradicting scale-invariance. Thus no non-zero random variable is scale-invariant. Note, however, that the measure on $(\mathbb{R}^+, \mathcal{B}^+)$ defined as

$$\mu\big([a,b]\big) := \int_a^b \frac{\mathrm{d}t}{t} = \log \frac{b}{a} \quad \text{for all } [a,b] \subset \mathbb{R}^+ ,$$

is scale invariant because, for every $\alpha > 0$,

$$\alpha_* \mu\big([a,b]\big) = \int_{a/\alpha}^{b/\alpha} \frac{\mathrm{d}t}{t} = \log \frac{b}{a} = \mu\big([a,b]\big) .$$

Obviously, μ is not finite, i.e. $\mu(\mathbb{R}^+) = +\infty$, but is still *$\sigma$-finite*. (Generally, a measure μ on $(\Omega, \mathcal{A})$ is *σ-finite* if $\Omega = \bigcup_{n\in\mathbb{N}} A_n$ for some sequence (A_n) in $\mathcal{A}$, and $\mu(A_n) < +\infty$ for all n.)

In a similar spirit, a sequence (x_n) of real numbers may be called *scale-invariant* if

$$\lim\nolimits_{N\to\infty} \frac{\#\{1 \le n \le N : \alpha x_n \in [a,b]\}}{N} = \lim\nolimits_{N\to\infty} \frac{\#\{1 \le n \le N : x_n \in [a,b]\}}{N}$$

holds for all $\alpha > 0$ and $[a,b] \subset \mathbb{R}$. For example, the sequence

$$\big(2, 2^{-1}, 2, 3, 2^{-1}, 3^{-1}, 2, 3, 4, 2^{-1}, 3^{-1}, 4^{-1}, \ldots, 2, 3, \ldots, n, 2^{-1}, 3^{-1}, \ldots, n^{-1}, 2 \ldots\big)$$

is scale-invariant. As above, it is not hard to see that

$$\lim\nolimits_{N\to\infty} \frac{\#\{1 \le n \le N : x_n \in [a,b]\}}{N} = 0 \quad \text{for all } [a,b] \subset \mathbb{R}\backslash\{0\} ,$$

holds whenever (x_n) is scale-invariant. Most elements of a scale-invariant sequence of real numbers, therefore, are very close to either 0 or $\pm\infty$.

While a positive random variable X cannot be scale-invariant, as shown above, it may nevertheless have *scale-invariant significant digits*. For this, however, X has to be Benford. In fact, Theorem 4.20 below shows that being Benford is (not only necessary but) also sufficient for X to have scale-invariant significant digits. The result will first be stated in terms of probability distributions.

Definition 4.18. Let $\mathcal{A} \supset \mathcal{S}$ be a σ-algebra on $\mathbb{R}^+$. A probability measure P on $(\mathbb{R}^+, \mathcal{A})$ has *scale-invariant significant digits* if

$$P(\alpha A) = P(A) \quad \text{for all } \alpha > 0 \text{ and } A \in \mathcal{S} ,$$

or equivalently if for all $m \in \mathbb{N}$, all $d_1 \in \{1, 2, \ldots, 9\}$ and all $d_j \in \{0, 1, \ldots, 9\}$, $j \ge 2$,

$$P\big(\{x : D_j(\alpha x) = d_j \text{ for } j = 1, 2, \ldots m\}\big) = P\big(\{x : D_j(x) = d_j \text{ for } j = 1, 2, \ldots, m\}\big)$$

holds for every $\alpha > 0$.

Example 4.19. **(i)** The Benford probability measure $\mathbb{B}$ on $(\mathbb{R}^+, \mathcal{S})$ has scale-invariant significant digits. This follows from Theorem 4.20 below but can also be seen from a direct calculation. Indeed, if $A = \bigcup_{k\in\mathbb{Z}} 10^k[a,b]$ with $1 \le a < b < 10$, then, given any $\alpha > 0$,

$$\alpha A = \bigcup_{k\in\mathbb{Z}} 10^{k+\log\alpha}[a,b] = \bigcup_{k\in\mathbb{Z}} 10^{k+\langle\log\alpha\rangle}[a,b] = \bigcup_{k\in\mathbb{Z}} 10^k B\,,$$

where the set B is given by

$$B = \begin{cases} \left[10^{\langle\log\alpha\rangle}a, 10^{\langle\log\alpha\rangle}b\right] & \text{if } 0 \le \langle\log\alpha\rangle < 1 - \log b\,, \\ \left[1, 10^{\langle\log\alpha\rangle-1}b\right] \cup \left[10^{\langle\log\alpha\rangle}a, 10\right) & \text{if } 1-\log b \le \langle\log\alpha\rangle < 1-\log a\,, \\ \left[10^{\langle\log\alpha\rangle-1}a, 10^{\langle\log\alpha\rangle-1}b\right] & \text{if } 1-\log a \le \langle\log\alpha\rangle < 1\,. \end{cases}$$

From this, it follows that

$$\begin{aligned} \mathbb{B}(\alpha A) &= \begin{cases} \log 10^{\langle\log\alpha\rangle}b - \log 10^{\langle\log\alpha\rangle}a \\ \log 10^{\langle\log\alpha\rangle-1}b + 1 - \log 10^{\langle\log\alpha\rangle}a \\ \log 10^{\langle\log\alpha\rangle-1}b - \log 10^{\langle\log\alpha\rangle-1}a \end{cases} \\ &= \log b - \log a \;=\; \mathbb{B}(A)\,, \end{aligned}$$

showing that $\mathbb{B}$ has scale-invariant digits.

(ii) The Dirac probability measure δ_1 concentrated at the constant 1 does not have scale-invariant significant digits, since $\delta_2 = 2_*\delta_1$ yet $\delta_1(D_1 = 1) = 1 \ne 0 = \delta_2(D_1 = 1)$.

(iii) The uniform distribution on $[0,1)$ does not have scale-invariant digits, since if X is distributed according to $\lambda_{0,1}$ then, for example

$$\mathbb{P}(D_1(X) = 1) = \frac{1}{9} < \frac{11}{27} = \mathbb{P}\left(D_1\left(\frac{3}{2}X\right) = 1\right).$$

As mentioned earlier, the Benford distribution is the only probability measure (on the significand σ-algebra) having scale-invariant significant digits.

Theorem 4.20 (Scale-invariance characterization [Hi1])**.** *A probability measure P on $(\mathbb{R}^+, \mathcal{A})$ with $\mathcal{A} \supset \mathcal{S}$ has scale-invariant significant digits if and only if $P(A) = \mathbb{B}(A)$ for every $A \in \mathcal{S}$, i.e., if and only if P is Benford.*

Proof. Fix any probability measure P on $(\mathbb{R}^+, \mathcal{A})$, denote by P_0 its restriction to $(\mathbb{R}^+, \mathcal{S})$, and let $Q := \ell_* P_0$ with ℓ given by Lemma 2.16. According to Lemma 2.16, Q is a probability measure on $\big([0,1), \mathcal{B}[0,1)\big)$. Moreover, under the correspondence established by ℓ,

$$P_0(\alpha A) = P_0(A) \quad \text{for all } \alpha > 0, A \in \mathcal{S} \tag{4.3}$$

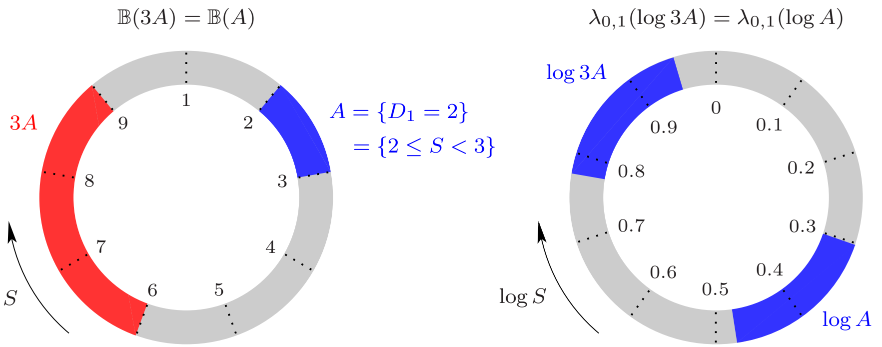

FIG 10. *Visualizing the scale-invariant significant digits of BL.*

is equivalent to

$$Q(\langle t + B\rangle) = Q(B) \quad \text{for all } t \in \mathbb{R}, B \in \mathcal{B}[0,1)\,, \tag{4.4}$$

where $\langle t + B\rangle = \{\langle t + x\rangle : x \in B\}$. Pick a random variable X such that the distribution of X is given by Q. With this, (4.4) simply means that, for every $t \in \mathbb{R}$, the distributions of $\langle X\rangle$ and $\langle t+X\rangle$ coincide. By Theorem 4.13(i) and (ii) this is the case if and only if X is u.d. mod 1, i.e. $Q = \lambda_{0,1}$. (For the "if" part, note that a *constant* random variable is independent from every other random variable.) Hence (4.3) is equivalent to $P_0 = (\ell^{-1})_*\lambda_{0,1} = \mathbb{B}$. $\square$

Example 4.21. For every integer k, let $q_k > 0$ and

$$f_k(t) = \begin{cases} \dfrac{1}{t \ln 10} & \text{if } 10^k \le t < 10^{k+1}\,, \\ 0 & \text{otherwise.} \end{cases}$$

If $\sum_{k\in\mathbb{Z}} q_k = 1$ then, according to Example 3.6, $\sum_{k\in\mathbb{Z}} q_k f_k$ is the density of a Benford probability measure P on $(\mathbb{R}^+, \mathcal{B}^+)$. By Theorem 4.20, P has scale-invariant significant digits. Note that, in full agreement with earlier observations, P is not scale-invariant, as for instance

$$q_k = P\big([10^k, 10^{k+1})\big) = P\big(10^{k-l}[10^l, 10^{l+1})\big) = P\big([10^l, 10^{l+1})\big) = q_l$$

cannot possibly hold for *all* pairs (k, l) of integers.

In analogy to Definition 4.18, a sequence (x_n) of real numbers is said to have *scale-invariant significant digits* if

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : S(\alpha x_n) < t\}}{N} = \lim_{N\to\infty} \frac{\#\{1 \le n \le N : S(x_n) < t\}}{N} \quad \text{for all } \alpha > 0, t \in [1, 10)\,. \tag{4.5}$$

Implicit in (4.5) is the assumption that the limits on either side exit for all t. A similar definition can be considered for real-valued *functions*. To formulate an analog of Theorem 4.20 using this terminology, recall that a set $A \subset \mathbb{N}$ *has density* $\rho \in [0,1]$ if the limit $\lim_{N\to\infty} \#\{1 \le n \le N : n \in A\}/N$ exists and equals ρ. For example, $\rho(\{n : n \text{ even}\}) = \frac{1}{2}$ and $\rho(\{n : n \text{ prime}\}) = 0$, whereas $\{n : D_1(n) = 1\}$ does not have a density.

Theorem 4.22. (i) *For any sequence (x_n) of real numbers, let $\{n : x_n \neq 0\} = \{n_1 < n_2 < \ldots\}$. Then (x_n) has scale-invariant significant digits if and only if $\{n : x_n \neq 0\}$ has a density and either $\rho(\{n : x_n = 0\}) = 1$ or else $(x_{n_j})_{j\in\mathbb{N}}$ is Benford. In particular, if $\rho(\{n : x_n = 0\}) = 0$ then the sequence (x_n) has scale-invariant significant digits if and only if it is Benford.*

(ii) *A (Borel measurable) function $f : [0,+\infty) \to \mathbb{R}$ with $\lambda\big(\{t \ge 0 : f(t) = 0\}\big) < +\infty$ has scale-invariant significant digits if and only if it is Benford. Moreover, f is Benford precisely if αf is Benford for every $\alpha \neq 0$.*

Proof. (i) Assume first that (x_n) has scale-invariant significant digits. According to (4.5),

$$G(s) := \lim_{N\to\infty} \frac{\#\{1 \le n \le N : S(x_n) < 10^s\}}{N}$$

exists for every $0 \le s < 1$. In particular, $\{n : x_n = 0\}$ has a density $G(0)$. For $G(0) = 1$ there is nothing else to show. Thus, assume $G(0) < 1$ from now on, and define a non-decreasing function $H : [0,1) \to \mathbb{R}$ as

$$H(s) = \frac{G(s) - G(0)}{1 - G(0)}, \quad 0 \le s < 1 .$$

Note that

$$\begin{aligned} H(s) &= \lim_{N\to\infty} \frac{\#\{1 \le n \le N : S(x_n) < 10^s, x_n \neq 0\}}{\#\{1 \le n \le N : x_n \neq 0\}} \\ &= \lim_{N\to\infty} \frac{\#\{1 \le j \le N : S(x_{n_j}) < 10^s\}}{N}, \end{aligned}$$

so H takes into account only the non-zero entries in (x_n). Define $h : \mathbb{R} \to \mathbb{R}$ as

$$h(s) = H(\langle s \rangle) - \langle s \rangle \quad \text{for all } s \in \mathbb{R} .$$

Clearly, h is 1-periodic, with $h(0) = 0$ and $|h(s)| \le 1$ for all $s \in \mathbb{R}$. In terms of the function H, the invariance property (4.5) simply reads

$$H(s) = \begin{cases} H(1 + s - \langle \log \alpha \rangle) - H(1 - \langle \log \alpha \rangle) & \text{if } s < \langle \log \alpha \rangle , \\ 1 - H(1 - \langle \log \alpha \rangle) + H(s - \langle \log \alpha \rangle) & \text{if } s \ge \langle \log \alpha \rangle , \end{cases}$$

provided that $\log \alpha \notin \mathbb{Z}$. In terms of h, this is equivalent to

$$h(s) = h(1 + s - \langle \log \alpha \rangle) - h(1 - \langle \log \alpha \rangle) \quad \text{for all } s \in \mathbb{R}, \alpha > 0 . \tag{4.6}$$

As a consequence, $s \mapsto h(1 + s - \langle \log \alpha \rangle) - h(s)$ is constant for every $\alpha > 0$. Since the function h is bounded and 1-periodic, it can be represented (at least in the L^2-sense) by a Fourier series

$$h(s) = \sum\nolimits_{k\in\mathbb{Z}} c_k e^{2\pi\imath ks}\,,$$

from which it follows that

$$\begin{aligned} h(1 + s - \langle \log \alpha \rangle) - h(s) &= \sum\nolimits_{k\in\mathbb{Z}} c_k \left(e^{2\pi\imath k(1+s-\langle \log \alpha \rangle)} - e^{2\pi\imath ks} \right) \\ &= \sum\nolimits_{k\in\mathbb{Z}} c_k \left(e^{-2\pi\imath k\langle \log \alpha \rangle} - 1 \right) e^{2\pi\imath ks}\,. \end{aligned}$$

Pick $\alpha > 0$ such that $\langle \log \alpha \rangle$ is irrational, e.g. $\alpha = 2$. Then $e^{-2\pi\imath k\langle \log \alpha \rangle} \neq 1$ whenever $k \neq 0$, which in turn implies that $c_k = 0$ for all $k \neq 0$, i.e. h is constant almost everywhere. Thus $H(s) = s + c_0$ for a.e. $s \in [0, 1)$, and in fact $H(s) \equiv s$ because H is non-decreasing with $H(0) = 0$. Overall, therefore,

$$\lim\nolimits_{N\to\infty} \frac{\#\{1 \le j \le N : S(x_{n_j}) < 10^s\}}{N} = s \quad \text{for all } s \in [0, 1)\,,$$

showing that (x_{n_j}) is Benford.

Conversely, if $\rho(\{n : x_n = 0\}) = 1$ then (4.5) holds with both sides being equal to 1 for all $t \in [1, 10)$. Assume, therefore, that $\rho(\{n : x_n = 0\}) < 1$ and (x_{n_j}) is Benford. By the above, $h(s) \equiv 0$, so (4.6) and hence also (4.5) hold, i.e., (x_n) has scale-invariant significant digits.

The proof of (ii) is completely analogous, utilizing

$$G(s) := \lim\nolimits_{T\to+\infty} \frac{\lambda\big(\{\tau \in [0, T) : S\big(f(\tau)\big) < 10^s\}\big)}{T}\,, \quad 0 \le s < 1\,.$$

Note that the assumption $\lambda\big(\{t \ge 0 : f(t) = 0\}\big) < +\infty$ implies $G(0) = 0$. $\square$

Example 4.23. Let (x_n) equal either the sequence of Fibonacci or prime numbers. In both cases, $x_n \neq 0$ for all n, and hence by Theorem 4.22(i) (x_n) has scale-invariant significant digits if and only if it is Benford. Thus (F_n) does have scale-invariant significant digits, and (p_n) does not. These facts are illustrated empirically in Fig 11 to 13 which show the relevant data for, respectively, the first 10^2 (Fig 11 and 12) and 10^4 (Fig 13) entries of either sequence, and compare them with the respective expected values for BL.

The next example is an elegant and entertaining application of the ideas underlying Theorems 4.20 and 4.22 to the mathematical theory of games. The game may be easily understood by a schoolchild, yet it has proven a challenge for game theorists not familiar with BL.

original data

1	10946	165580141	2504730781961	37889062373143906
1	17711	267914296	4052739537881	61305790721611591
2	28657	433494437	6557470319842	99194853094755497
3	46368	701408733	10610209857723	160500643816367088
5	75025	1134903170	17167680177565	259695496911122585
8	121393	1836311903	27777890035288	420196140727489673
13	196418	2971215073	44945570212853	679891637638612258
21	317811	4807526976	72723460248141	1100087778366101931
34	514229	7778742049	117669030460994	1779979416004714189
55	832040	12586269025	190392490709135	2880067194370816120
89	1346269	20365011074	308061521170129	4660046610375530309
144	2178309	32951280099	498454011879264	7540113804746346429
233	3524578	53316291173	806515533049393	12200160415121876738
377	5702887	86267571272	1304969544928657	19740274219868223167
610	9227465	139583862445	2111485077978050	31940434634990099905
987	14930352	225851433717	3416454622906707	51680708854858323072
1597	24157817	365435296162	5527939700884757	83621143489848422977
2584	39088169	591286729879	8944394323791464	135301852344706746049
4181	63245986	956722026041	14472334024676221	218922995834555169026
6765	165580141	1548008755920	23416728348467685	354224848179261915075

rescaled by 2

2	21892	331160282	5009461563922	75778124746287812
2	35422	535828592	8105479075762	122611581443223182
4	57314	866988874	13114940639684	198389706189510994
6	92736	1402817466	21220419715446	321001287632734176
10	150050	2269806340	34335360355130	519390993822245170
16	242786	3672623806	55555780070576	840392281454979346
26	392836	5942430146	89891140425706	1359783275277224516
42	635622	9615053952	145446920496282	2200175556732203862
68	1028458	15557484098	235338060921988	3559958832009428378
110	1664080	25172538050	380784981418270	5760134388741632240
178	2692538	40730022148	616123042340258	9320093220751060618
288	4356618	65902560198	996908023758528	15080227609492692858
466	7049156	106632582346	1613031066098786	24400320830243753476
754	11405774	172535142544	2609939089857314	39480548439736446334
1220	18454930	279167724890	4222970155956100	63880869269980199810
1974	29860704	451702867434	6832909245813414	103361417709716646144
3194	48315634	730870592324	11055879401769514	167242286979696845954
5168	78176338	1182573459758	17888788647582928	270603704689413492098
8362	126491972	1913444052082	28944668049352442	437845991669110338052
13530	204668310	3096017511840	46833456696935370	708449696358523830150

rescaled by 7

7	76622	1159060987	17533115473727	265223436612007342
7	123977	1875400072	28369176765167	429140535051281137
14	200599	3034461059	45902292238894	694363971663288479
21	324576	4909861131	74271469004061	1123504506714569616
35	525175	7944322190	120173761242955	1817868478377858095
56	849751	12854183321	194445230247016	2941372985092427711
91	1374926	20798505511	314618991489971	4759241463470285806
147	2224677	33652688832	509064221736987	7700614448562713517
238	3599603	54451194343	823683213226958	12459855912032999323
385	5824280	88103883175	1332747434963945	20160470360595712840
623	9423883	142555077518	2156430648190903	32620326272628712163
1008	15248163	230658960693	3489178083154848	52780796633224425003
1631	24672046	373214038211	5645608731345751	85401122905853137166
2639	39920209	603872998904	9134786814500599	138181919539077562169
4270	64592255	977087037115	14780395545846350	223583042444930699335
6909	104512464	1580960036019	23915182360346949	361764961984008261504
11179	169104719	2558047073134	38695577906193299	585348004428938960839
18088	273617183	4139007109153	62610760266540248	947112966412947222343
29267	442721902	6697054182287	101306338172733547	1532460970841886183182
47355	716339085	10836061291440	163917098439273795	2479573937254833405525

$\#\{D_1 = d\}$	d	1	2	3	4	5	6	7	8	9	$10^3 \cdot R$
	$(F_n)_{n=1}^{100}$	30	18	13	9	8	6	5	7	4	18.84
	$(2F_n)_{n=1}^{100}$	30	19	11	10	8	7	6	5	4	14.93
	$(7F_n)_{n=1}^{100}$	29	19	13	8	8	7	5	4	5	16.91
	$10^2 \cdot \log(1 + d^{-1})$	30.10	17.60	12.49	9.691	7.918	6.694	5.799	5.115	4.575	

FIG 11. *Illustrating the (approximate) scale-invariance of the first one-hundred Fibonacci numbers, cf.* FIG 5.

original data

2	31	73	127	179	233	283	353	419	467
3	37	79	131	181	239	293	359	421	479
5	41	83	137	191	241	307	367	431	487
7	43	89	139	193	251	311	373	433	491
11	47	97	149	197	257	313	379	439	499
13	53	101	151	199	263	317	383	443	503
17	59	103	157	211	269	331	389	449	509
19	61	107	163	223	271	337	397	457	521
23	67	109	167	227	277	347	401	461	523
29	71	113	173	229	281	349	409	463	541

rescaled by 2

4	62	146	254	358	466	566	706	838	934
6	74	158	262	362	478	586	718	842	958
10	82	166	274	382	482	614	734	862	974
14	86	178	278	386	502	622	746	866	982
22	94	194	298	394	514	626	758	878	998
26	106	202	302	398	526	634	766	886	1006
34	118	206	314	422	538	662	778	898	1018
38	122	214	326	446	542	674	794	914	1042
46	134	218	334	454	554	694	802	922	1046
58	142	226	346	458	562	698	818	926	1082

rescaled by 7

14	217	511	889	1253	1631	1981	2471	2933	3269
21	259	553	917	1267	1673	2051	2513	2947	3353
35	287	581	959	1337	1687	2149	2569	3017	3409
49	301	623	973	1351	1757	2177	2611	3031	3437
77	329	679	1043	1379	1799	2191	2653	3073	3493
91	371	707	1057	1393	1841	2219	2681	3101	3521
119	413	721	1099	1477	1883	2317	2723	3143	3563
133	427	749	1141	1561	1897	2359	2779	3199	3647
161	469	763	1169	1589	1939	2429	2807	3227	3661
203	497	791	1211	1603	1967	2443	2863	3241	3787

	d	1	2	3	4	5	6	7	8	9	$10^3 \cdot R$
$\#\{D_1 = d\}$	$(p_n)_{n=1}^{100}$	25	19	19	20	8	2	4	2	1	103.0
	$(2p_n)_{n=1}^{100}$	17	12	13	9	10	10	9	11	9	131.0
	$(7p_n)_{n=1}^{100}$	31	26	22	5	3	2	6	1	4	95.06
	$10^2 \cdot \log(1+d^{-1})$	30.10	17.60	12.49	9.691	7.918	6.694	5.799	5.115	4.575	

FIG 12. *Illustrating the lack of scale-invariance for the first one-hundred prime numbers.*

Example 4.24 ([Mo]). Consider a two-person game where Player A and Player B each independently choose a (real) number greater than or equal to 1, and Player A wins if the product of their two numbers starts with a 1, 2, or 3; otherwise, Player B wins. Using the tools presented in this section, it may easily be seen that there is a strategy for Player A to choose her numbers so that she wins with probability at least $\log 4 \cong 60.2\%$, no matter what strategy Player B uses. Conversely, there is a strategy for Player B so that Player A will win no more than $\log 4$ of the time, no matter what strategy Player A uses.

The idea is simple, using the scale-invariance property of BL discussed above. If Player A chooses her number X randomly according to BL, then since BL is scale-invariant, it follows from Theorem 4.13(i) and Example 4.19(i) that $X \cdot y$ is still Benford no matter what number y Player B chooses, so Player A will win with the probability that a Benford random variable has first significant digit less than 4, i.e. with probability exactly $\log 4$. Conversely, if Player B chooses his number Y according to BL then, using scale-invariance again, $x \cdot Y$ is Benford, so Player A will again win with the probability exactly $\log 4$. In fact, as will now be shown, BL is the only optimal strategy for each player.

	d	1	2	3	4	5	6	7	8	9	$10^3 \cdot R$
Fibonacci	original	3011	1762	1250	968	792	668	580	513	456	0.1574
	$\times 2$	3009	1763	1248	970	792	670	580	511	457	0.2087
	$\times 7$	3009	1762	1249	969	791	668	583	511	458	0.3080
Prime	original	1601	1129	1097	1069	1055	1013	1027	1003	1006	140.9
	$\times 2$	5104	1016	585	573	556	556	541	543	526	209.3
	$\times 7$	1653	1572	1504	1469	1445	1434	584	174	165	135.7
$10^4 \cdot \log(1+d^{-1})$		3010.	1760.	1249.	969.1	791.8	669.4	579.9	511.5	457.5	

FIG 13. *When the sample size is increased from $N = 10^2$ to $N = 10^4$ the Fibonacci numbers are even closer to scale-invariance. For the primes, this is not the case at all, see also* FIG 5.

To prepare for the formal argument, model the *strategy* of Player A, i.e. the way this player chooses a number, by a probability measure P on $(\mathbb{R}^+, \mathcal{B}^+)$. For example, if Player A chooses the same number a all the time, then $P = \delta_a$. (Game theorists refer to this as a *pure* strategy.) Similarly, Q represents the strategy of Player B. Denote by $\mathcal{M}^+$ the set of all probability measures on $(\mathbb{R}^+, \mathcal{B}^+)$ and, given $P, Q \in \mathcal{M}^+$, let $p(P,Q) \in [0,1]$ be the probability that Player A wins, i.e., the product of the chosen numbers begins with 1, 2, or 3, assuming Players A and B choose their numbers independently and according to the strategies P and Q, respectively. It is natural for Player A to try to maximize $\inf_{Q\in\mathcal{M}^+} p(P,Q)$, whereas Player B aims at minimizing $\sup_{P\in\mathcal{M}^+} p(P,Q)$. Which strategies should the players choose, and what probabilities of winning/losing are achievable/unavoidable?

In view of the informal discussion above, it may not come as a complete surprise that these questions ultimately have very simple answers. A little preparatory work is required though. To this end, for every $0 \le s < 1$ and $P \in \mathcal{M}^+$, let

$$G_P(s) := P\big(\{x > 0 : S(x) \le 10^s\}\big)\,,$$

and note that $s \mapsto G_P(s)$ is non-decreasing, right-continuous, with $G_P(0) \ge 0$ as well as $\lim_{s\uparrow 1} G_P(s) = 1$. (With the terminology and notation introduced in Sections 2.3 and 3.3 simply $G_P(s) = F_{S_* P}(10^s)$.) Extend G_P to a (non-decreasing, right-continuous) function $G_P : \mathbb{R} \to \mathbb{R}$ by setting

$$G_P(s) := G_P(\langle s\rangle) + \lfloor s\rfloor \quad \text{for all } s \in \mathbb{R}\,,$$

and let $g_P(s) := G_P(s) - s$. Since

$$g_P(s+1) = G_P(s+1) - (s+1) = G_P(\langle s \rangle) - \langle s \rangle = g_P(s)\,,$$

the function g_P is 1-periodic with $g_P(0) = 0$. Also, g_P is Riemann integrable, and $|g_P(s)| \le 1$ for all $s \in \mathbb{R}$. With these preliminary definitions, observe now that, given any $a > 0$,

$$\begin{aligned} p(P, \delta_a) &= \begin{cases} G_P(\log 4 - \langle \log a \rangle) + 1 - G_P(1 - \langle \log a \rangle) & \text{if } \langle \log a \rangle < \log 4 \\ G_P(1 + \log 4 - \langle \log a \rangle) - G_P(1 - \langle \log a \rangle) & \text{if } \langle \log a \rangle \ge \log 4 \end{cases} \\ &= g_P(1 + \log 4 - \langle \log a \rangle) - g_P(1 - \langle \log a \rangle) + \log 4 \\ &= \log 4 + h_P(\langle \log a \rangle)\,, \end{aligned}$$

where the 1-periodic, Riemann integrable function $h_P : \mathbb{R} \to \mathbb{R}$ is given by

$$h_P(s) = g_P(1 + \log 4 - s) - g_P(1 - s)\,, \quad s \in \mathbb{R}\,.$$

From $\int_0^1 h_P(s)\,\mathrm{d}s = 0$, it follows that $c_P := \inf_{s\in\mathbb{R}} h_P(s) \le 0$. Consequently, if $c_P < 0$ then

$$\inf\nolimits_{Q\in\mathcal{M}^+} p(P, Q) \le \inf\nolimits_{a>0} p(P, \delta_a) = \log 4 + c_P < \log 4\,.$$

On the other hand, if $c_P = 0$ then necessarily $h_P(s) = 0$ for a.e. s and hence, as g_P is right-continuous,

$$g_P(-s + \log 4) = g_P(-s) \quad \text{for all } s \in \mathbb{R}\,.$$

This in turn implies that $g_P(\langle n \log 4 \rangle) = g_P(0)$ for all $n \in \mathbb{N}$. Recall now that g_P has at most countably many discontinuities and that $(\langle n \log 4 \rangle)$ is u.d. mod 1 and hence dense in the interval $[0, 1)$. Thus, if $0 < s_0 < 1$ is a point of continuity of g_P, then choosing a sequence $1 \le n_1 < n_2 < \ldots$ with $\lim_{j\to\infty} \langle n_j \log 4 \rangle = s_0$ shows that

$$g_P(s_0) = \lim\nolimits_{j\to\infty} g_P(\langle n_j \log 4 \rangle) = g_P(0)\,.$$

With the possible exception of at most countably many s therefore, $G_P(s) = s + g_P(0)$ whenever $0 \le s < 1$. But since $s \mapsto G_P(s)$ is non-decreasing with $G_P(s) \ge 0$ and $\lim_{s\uparrow 1} G_P(s) = 1$, $g_P(0) = 0$ and $G_P(s) = s$ must in fact hold for *all* s, i.e.

$$P\big(\{x > 0 : S(x) \le 10^s\}\big) \equiv s\,.$$

In other words, P is Benford. Overall therefore

$$\inf\nolimits_{Q\in\mathcal{M}^+} p(P, Q) \le \log 4 = 0.6020\ldots\,,$$

with equality holding if and only if P is Benford. Thus the unique optimal strategy for Player A is to choose her numbers according to BL.

A completely analogous argument shows that

$$\sup\nolimits_{P\in\mathcal{M}^+} p(P, Q) \ge \log 4\,,$$

with equality holding if and only if Q is Benford. Hence the unique optimal strategy for Player B to minimize the probability of loosing is also to choose numbers obeying BL. Overall,

$$\sup_{P\in\mathcal{M}^+}\inf_{Q\in\mathcal{M}^+} p(P,Q) = \log 4 = \inf_{Q\in\mathcal{M}^+}\sup_{P\in\mathcal{M}^+} p(P,Q)$$

holds, and the value (expected gain) of one game for Player A is given by $\log 4 - (1-\log 4) = 0.2041\ldots > \frac{1}{5}$.

If both players are required to choose *positive integers* then their strategies are probabilities on $(\mathbb{N}, \mathbb{N}\cap\mathcal{B})$. Denote by $\mathcal{M}_{\mathbb{N}}$ the set of all such probabilities. Since $\{\langle \log n\rangle : n \in \mathbb{N}\}$ is dense in $[0,1)$, the above argument shows that

$$\inf_{Q\in\mathcal{M}_{\mathbb{N}}} p(P,Q) < \log 4$$

for every $P \in \mathcal{M}_{\mathbb{N}}$, and similarly

$$\sup_{P\in\mathcal{M}_{\mathbb{N}}} p(P,Q) > \log 4$$

for every $Q \in \mathcal{M}_{\mathbb{N}}$. On the other hand, given $\varepsilon > 0$, it is not hard to find $P_\varepsilon, Q_\varepsilon \in \mathcal{M}_{\mathbb{N}}$ such that

$$\log 4 - \varepsilon < \inf_{Q\in\mathcal{M}_{\mathbb{N}}} p(P_\varepsilon,Q) < \log 4 < \sup_{P\in\mathcal{M}_{\mathbb{N}}} p(P,Q_\varepsilon) < \log 4 + \varepsilon\,.$$

Indeed, it is enough to choose $P_\varepsilon, Q_\varepsilon$ such that these probabilities approximate BL sufficiently well. (Recall Example 3.9 which also showed that no $P \in \mathcal{M}_{\mathbb{N}}$ is Benford.) When played with positive integers only, therefore, the game has no optimal strategy for either player, but there are ε-optimal strategies for every $\varepsilon > 0$, and

$$\sup_{P\in\mathcal{M}_{\mathbb{N}}}\inf_{Q\in\mathcal{M}_{\mathbb{N}}} p(P,Q) = \log 4 = \inf_{Q\in\mathcal{M}_{\mathbb{N}}}\sup_{P\in\mathcal{M}_{\mathbb{N}}} p(P,Q)$$

still holds.

Theorem 4.20 showed that for a probability measure P on $(\mathbb{R}^+, \mathcal{B}^+)$ to have scale-invariant significant digits it is necessary (and sufficient) that P be Benford. In fact, as noted in [Sm], this conclusion already follows from a much weaker assumption: It is enough to require that the probability of a single significant digit remain unchanged under scaling.

Theorem 4.25. *For every random variable X with $\mathbb{P}(X=0)=0$ the following statements are equivalent:*

(i) *X is Benford.*

(ii) *There exists a number $d \in \{1,2,\ldots,9\}$ such that*

$$\mathbb{P}(D_1(\alpha X) = d) = \mathbb{P}(D_1(X)=d) \quad \text{for all } \alpha > 0\,.$$

In particular, (ii) *implies that $\mathbb{P}(D_1(X)=d) = \log(1+d^{-1})$.*

Proof. Assume first that X is Benford. By Theorem 4.20, X has scale-invariant significant digits. Thus for every $\alpha > 0$,

$$\mathbb{P}(D_1(\alpha X) = d) = \log(1 + d^{-1}) = \mathbb{P}(D_1(X) = d) \quad \text{for all } d = 1, 2, \ldots, 9\,.$$

Conversely, assume that (ii) holds. Similarly as in the proof of Theorem 4.22(i), for every $0 \le s < 1$ let

$$G_X(s) := \mathbb{P}(S(X) < 10^s)\,.$$

Hence G_X is non-decreasing and left-continuous, with $G_X(0) = 0$, and

$$\mathbb{P}(D_1(X) = d) = G_X\bigl(\log(1 + d)\bigr) - G_X(\log d)\,.$$

Extend G_X to a (non-decreasing, left-continuous) function $G_X : \mathbb{R} \to \mathbb{R}$ by setting $G_X(s) := G_X(\langle s \rangle) + \lfloor s \rfloor$, and let $g_X(s) := G_X(s) - s$. Hence g_X is 1-periodic, Riemann-integrable, with $g_X(0) = 0$ and $|g_X(s)| \le 1$. Specifically,

$$\mathbb{P}(D_1(X) = d) = g_X\bigl(\log(1 + d)\bigr) - g_X(\log d) + \log(1 + d^{-1})\,,$$

and essentially the same calculation as in Example 4.24 shows that

$$\mathbb{P}(D_1(\alpha X) = d) = g_X(\log(1 + d) - \langle \log \alpha \rangle) - g_X(\log d - \langle \log \alpha \rangle) + \log(1 + d^{-1})\,.$$

With the 1-periodic, Riemann-integrable $h_X : \mathbb{R} \to \mathbb{R}$ given by

$$h_X(s) = g_X(\log(1 + d) - s) - g_X(\log d - s)\,,$$

the assumption that $\mathbb{P}(D_1(\alpha X) = d) = \mathbb{P}(D_1(X) = d)$ for all $\alpha > 0$ simply means that $h_X(s) \equiv h_X(0)$, i.e., h_X is constant, and so is the function $s \mapsto g_X(\log(1 + d) - s) - g_X(\log d - s)$. The same Fourier series argument as in the proof of Theorem 4.22 now applies: From

$$g_X(s) = \sum\nolimits_{k \in \mathbb{Z}} c_k e^{2\pi \imath k s}\,,$$

it follows that

$$\begin{aligned} g_X(\log(1 + d) - s) - g_X(\log d - s) &= \sum\nolimits_{k \in \mathbb{Z}} c_k \Bigl(e^{2\pi \imath k \log(1+d)} - e^{2\pi \imath k \log d}\Bigr) e^{2\pi \imath k s} \\ &= \sum\nolimits_{k \in \mathbb{Z}} c_k e^{2\pi \imath k \log d} \Bigl(e^{2\pi \imath k \log(1+d^{-1})} - 1\Bigr) e^{2\pi \imath k s}\,, \end{aligned}$$

and since $\log(1 + d^{-1})$ is irrational for every $d \in \mathbb{N}$, necessarily $c_k = 0$ for all $k \neq 0$, i.e., g_X is constant almost everywhere, and $G_X(s) = s + c_0$ for a.e. $s \in [0, 1)$. As G_X is non-decreasing with $G_X(0) = 0$, overall, $G_X(s) \equiv s$, which in turn shows that X is Benford. $\square$

Remark. A close inspection of the above proof shows that Theorem 4.25 can still be strengthened in different ways. On the one hand, other significant digits can be considered. For example, the theorem (and its proof also) remain virtually unchanged if in (ii) it is assumed that, for some $m \geq 2$ and some $d \in \{0, 1, \dots, 9\}$,

$$\mathbb{P}(D_m(\alpha X) = d) = \mathbb{P}(D_m(X) = d) \quad \text{for all } \alpha > 0 \, .$$

On the other hand, it is enough to assume in (ii) that, for some $d \in \{1, 2, \dots, 9\}$,

$$\mathbb{P}(D_1(\alpha_n X) = d) = \mathbb{P}(D_1(X) = d) \quad \text{for all } n \in \mathbb{N} \, ,$$

with the sequence (α_n) of positive numbers being such that $\{\langle \log \alpha_n \rangle : n \in \mathbb{N}\}$ is dense in $[0, 1)$. Possible choices for such a sequence include (2^n), (n^2), and the sequence of prime numbers. For example, therefore, X is Benford if and only if

$$\mathbb{P}(D_1(2^n X) = 1) = \mathbb{P}(D_1(X) = 1) \quad \text{for all } n \in \mathbb{N} \, . \qquad \clubsuit$$

Example 4.26 ([Sm]). ("Ones-scaling-test") In view of the last remark, to informally test whether a sample of data comes from a Benford distribution, simply compare the proportion of the sample that has first significant digit 1 with the proportion after the data has been re-scaled, i.e. multiplied by $\alpha, \alpha^2, \alpha^3, \dots$, where $\log \alpha$ is irrational, e.g. $\alpha = 2$. In fact, it is enough to consider only rescalings by α^{n^2}, $n = 1, 2, 3, \dots$. On the other hand, note that merely assuming

$$\mathbb{P}(D_1(2X) = d) = \mathbb{P}(D_1(X) = d) \quad \text{for all } d = 1, 2, \dots, 9 \, , \tag{4.7}$$

is not sufficient to guarantee that X is Benford. Indeed, (4.7) holds for instance if X attains each of the four values $1, 2, 4, 8$ with equal probability $\frac{1}{4}$.

4.3. The base-invariance characterization

One possible drawback to the hypothesis of scale-invariance in some tables is the special role played by the constant 1. For example, consider two physical laws, namely Newton's *lex secunda* $F = ma$ and Einstein's famous $E = mc^2$. Both laws involve universal constants. In Newton's law, the constant is usually made equal to 1 by the choice of units of measurement, and this 1 is then not recorded in most tables of universal constants. On the other hand, the speed of light c in Einstein's equation is typically recorded as a fundamental constant. If a "complete" list of universal physical constants also included the 1s, it seems plausible that this special constant might occur with strictly positive frequency. But that would clearly violate scale-invariance, since then the constant 2, and in fact every other constant as well would occur with this same positive probability, which is impossible.

Instead, suppose it is assumed that any reasonable universal significant-digit law should have *base-invariant significant digits*, that is, the law should be equally valid when rewritten in terms of bases other than 10. In fact, all of the classical arguments supporting BL carry over *mutatis mutandis* [Ra1] to other bases. As will be seen shortly, a hypothesis of base-invariant significant digits characterizes mixtures of BL and a Dirac probability measure concentrated on the special constant 1 which may occur with positive probability.

Just as the only scale-invariant real-valued random variable is 0 with probability one, the only positive random variable X that is *base-invariant*, i.e. $X = 10^Y$ with some random variable Y for which $Y, 2Y, 3Y, \ldots$ all have the same distribution, is the random variable which almost surely equals 1, that is, $\mathbb{P}(X = 1) = 1$. This follows from the fact that all nY have the same distribution for $n = 1, 2, 3\ldots$, and hence $\mathbb{P}(Y = 0) = 1$, as shown in the previous section.

On the other hand, a positive random variable (or sequence, function, distribution) can have *base-invariant significant digits*. The idea behind base-invariance of significant digits is simply this: A base-10 significand event A corresponds to the base-100 event $A^{1/2}$, since the new base $b = 100$ is the square of the original base $b = 10$. As a concrete example, denote by A the set of positive reals with first significant digit 1, i.e.

$$A = \{x > 0 : D_1(x) = 1\} = \{x > 0 : S(x) \in [1, 2)\}\,.$$

It is easy to see that $A^{1/2}$ is the set

$$A^{1/2} = \{x > 0 : S(x) \in [1, \sqrt{2}) \cup [\sqrt{10}, \sqrt{20})\}\,.$$

Consider now the base-100 significand function S_{100}, i.e., for any $x \neq 0$, $S_{100}(x)$ is the unique number in $[1, 100)$ such that $|x| = 100^k S_{100}(x)$ for some, necessarily unique $k \in \mathbb{Z}$. (To emphasize that the usual significand function S is taken relative to base 10, it will be denoted S_{10} throughout this section.) Clearly,

$$A = \{x > 0 : S_{100}(x) \in [1, 2) \cup [10, 20)\}\,.$$

Hence, letting $a = \log 2$,

$$\left\{x > 0 : S_b(x) \in [1, b^{a/2}) \cup [b^{1/2}, b^{(1+a)/2})\right\} = \begin{cases} A^{1/2} & \text{if } b = 10\,, \\ A & \text{if } b = 100\,. \end{cases}$$

Thus, if a distribution P on the significand σ-algebra $\mathcal{S}$ has base-invariant significant digits, then $P(A)$ and $P(A^{1/2})$ should be the same, and similarly for other integral roots (corresponding to other integral powers of the original base $b = 10$). Thus $P(A) = P(A^{1/n})$ should hold for all n. (Recall from Lemma 2.13(iii) that $A^{1/n} \in \mathcal{S}$ for all $A \in \mathcal{S}$ and $n \in \mathbb{N}$, so those probabilities are well-defined.) This motivates the following definition.

Definition 4.27. Let $\mathcal{A} \supset \mathcal{S}$ be a σ-algebra on $\mathbb{R}^+$. A probability measure P on $(\mathbb{R}^+, \mathcal{A})$ has *base-invariant significant digits* if $P(A) = P(A^{1/n})$ holds for all $A \in \mathcal{S}$ and $n \in \mathbb{N}$.

Example 4.28. **(i)** Recall that δ_a denotes the Dirac measure concentrated at the point a, that is, $\delta_a(A) = 1$ if $a \in A$, and $\delta_a(A) = 0$ if $a \notin A$. The probability measure δ_1 clearly has base-invariant significant digits since $1 \in A$ if and only if $1 \in A^{1/n}$. Similarly, δ_{10^k} has base-invariant significant digits for every $k \in \mathbb{Z}$. On the other hand, δ_2 does not have base-invariant significant digits since, with $A = \{x > 0 : S_{10}(x) \in [1, 3)\}$, $\delta_2(A) = 1$ yet $\delta_2(A^{1/2}) = 0$.

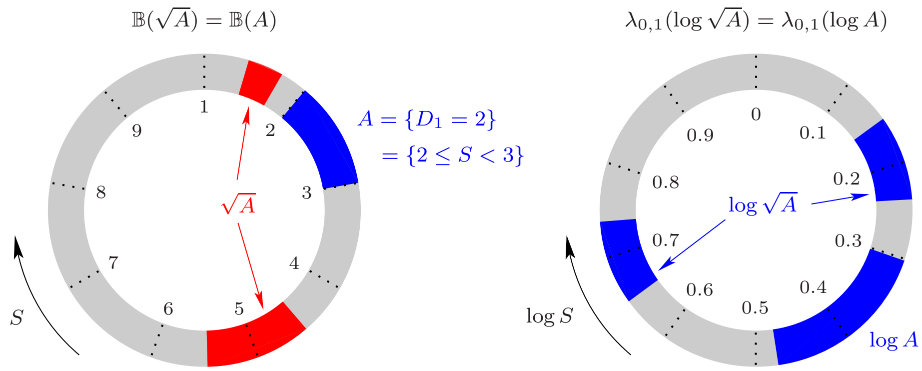

FIG 14. *Visualizing the base-invariant significant digits of BL.*

(ii) It is easy to see that the Benford distribution $\mathbb{B}$ has base-invariant significant digits. Indeed, for any $0 \le s < 1$, let

$$A = \{x > 0 : S_{10}(x) \in [1, 10^s)\} = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k[1, 10^s) \in \mathcal{S}\,.$$

Then, as seen in the proof of Lemma 2.13(iii),

$$A^{1/n} = \bigcup\nolimits_{k\in\mathbb{Z}} 10^k \bigcup\nolimits_{j=0}^{n-1} [10^{j/n}, 10^{(j+s)/n})$$

and therefore

$$\begin{aligned}\mathbb{B}(A^{1/n}) &= \sum\nolimits_{j=0}^{n-1} \left(\log 10^{(j+s)/n} - \log 10^{j/n}\right) = \sum\nolimits_{j=0}^{n-1} \left(\frac{j+s}{n} - \frac{j}{n}\right)\\ &= s \;=\; \mathbb{B}(A)\,.\end{aligned}$$

(iii) The uniform distribution $\lambda_{0,1}$ on $[0, 1)$ does not have base-invariant significant digits. For instance, again taking $A = \{x > 0 : D_1(x) = 1\}$ leads to

$$\begin{aligned}\lambda_{0,1}(A^{1/2}) &= \sum\nolimits_{n\in\mathbb{N}} 10^{-n}(\sqrt{2} - 1 + \sqrt{20} - \sqrt{10}) = \frac{1}{9} + \frac{(\sqrt{5}-1)(2-\sqrt{2})}{9}\\ &> \frac{1}{9} \;=\; \lambda_{0,1}(A)\,.\end{aligned}$$

(iv) The probability measure $\frac{1}{2}\delta_1 + \frac{1}{2}\mathbb{B}$ has base-invariant significant digits since both δ_1 and $\mathbb{B}$ do.

Example 4.29. Completely analogously to the case of scale-invariance, it is possible to introduce a notion of a sequence or function having *base-invariant significant digits* and to formulate an analoge of Theorem 4.22 in the context of Theorem 4.30 below. With this, the sequence (F_n) has base-invariant significant digits, whereas the sequence (p_n) does not. As in Example 4.23, this is illustrated empirically in FIG 15 to 17.

original data

1	10946	165580141	2504730781961	37889062373143906
1	17711	267914296	4052739537881	61305790721611591
2	28657	433494437	6557470319842	99194853094755497
3	46368	701408733	10610209857723	160500643816367088
5	75025	1134903170	17167680177565	259695496911122585
8	121393	1836311903	27777890035288	420196140727489673
13	196418	2971215073	44945570212853	679891637638612258
21	317811	4807526976	72723460248141	1100087778366101931
34	514229	7778742049	117669030460994	1779979416004714189
55	832040	12586269025	190392490709135	2880067194370816120
89	1346269	20365011074	308061521170129	4660046610375530309
144	2178309	32951280099	498454011879264	7540113804746346429
233	3524578	53316291173	806515533049393	12200160415121876738
377	5702887	86267571272	1304969544928657	19740274219868223167
610	9227465	139583862445	2111485077978050	31940434634990099905
987	14930352	225851433717	3416454622906707	51680708854858323072
1597	24157817	365435296162	5527939700884757	83621143489848422977
2584	39088169	591286729879	8944394323791464	135301852344706746049
4181	63245986	956722026041	14472334024676221	218922995834555169026
6765	165580141	1548008755920	23416728348467685	354224848179261915075

base $\mapsto$ base^2

1.000	1.198	2.741	6.273	1.435
1.000	3.136	7.177	1.642	3.758
4.000	8.212	1.879	4.300	9.839
9.000	2.149	4.919	1.125	2.576
2.500	5.628	1.288	2.947	6.744
6.400	1.473	3.372	7.716	1.765
1.690	3.858	8.828	2.020	4.622
4.410	1.010	2.311	5.288	1.210
1.156	2.644	6.050	1.384	3.168
3.025	6.922	1.584	3.624	8.294
7.921	1.812	4.147	9.490	2.171
2.073	4.745	1.085	2.484	5.685
5.428	1.242	2.842	6.504	1.488
1.421	3.252	7.442	1.702	3.896
3.721	8.514	1.948	4.458	1.020
9.741	2.229	5.100	1.167	2.670
2.550	5.836	1.335	3.055	6.992
6.677	1.527	3.496	8.000	1.830
1.748	4.000	9.153	2.094	4.792
4.576	1.047	2.396	5.483	1.254

base $\mapsto$ base^7

1.000	1.882	3.412	6.184	1.120
1.000	5.466	9.907	1.795	3.254
1.280	1.587	2.876	5.213	9.449
2.187	4.608	8.352	1.513	2.743
7.812	1.337	2.424	4.395	7.966
2.097	3.884	7.040	1.276	2.312
6.274	1.127	2.044	3.705	6.715
1.801	3.274	5.935	1.075	1.949
5.252	9.508	1.723	3.123	5.661
1.522	2.760	5.003	9.068	1.643
4.423	8.015	1.452	2.633	4.772
1.283	2.327	4.217	7.644	1.385
3.728	6.756	1.224	2.219	4.023
1.082	1.961	3.555	6.444	1.168
3.142	5.696	1.032	1.871	3.391
9.124	1.653	2.997	5.432	9.846
2.649	4.801	8.703	1.577	2.858
7.692	1.394	2.526	4.579	8.300
2.233	4.047	7.336	1.329	2.410
6.484	1.175	2.130	3.860	6.997

	d	1	2	3	4	5	6	7	8	9	$10^3 \cdot R$
$\#\{D_1 = d\}$	$(F_n)_{n=1}^{100}$	30	18	13	9	8	6	5	7	4	18.84
	$(F_n^2)_{n=1}^{100}$	31	17	12	11	7	8	4	5	5	17.99
	$(F_n^7)_{n=1}^{100}$	31	18	11	9	8	7	6	4	6	14.93
	$10^2 \cdot \log(1 + d^{-1})$	30.10	17.60	12.49	9.691	7.918	6.694	5.799	5.115	4.575	

FIG 15. *Illustrating the (approximate) base-invariance of the first one-hundred Fibonacci numbers. (In the two middle tables, the values of $S(F_n^2)$ and $S(F_n^7)$, respectively, are shown to four correct digits.)*

original data

2	31	73	127	179	233	283	353	419	467
3	37	79	131	181	239	293	359	421	479
5	41	83	137	191	241	307	367	431	487
7	43	89	139	193	251	311	373	433	491
11	47	97	149	197	257	313	379	439	499
13	53	101	151	199	263	317	383	443	503
17	59	103	157	211	269	331	389	449	509
19	61	107	163	223	271	337	397	457	521
23	67	109	167	227	277	347	401	461	523
29	71	113	173	229	281	349	409	463	541

base $\mapsto$ base^2

4.000	9.610	5.329	1.612	3.204	5.428	8.008	1.246	1.755	2.180
9.000	1.369	6.241	1.716	3.276	5.712	8.584	1.288	1.772	2.294
2.500	1.681	6.889	1.876	3.648	5.808	9.424	1.346	1.857	2.371
4.900	1.849	7.921	1.932	3.724	6.300	9.672	1.391	1.874	2.410
1.210	2.209	9.409	2.220	3.880	6.604	9.796	1.436	1.927	2.490
1.690	2.809	1.020	2.280	3.960	6.916	1.004	1.466	1.962	2.530
2.890	3.481	1.060	2.464	4.452	7.236	1.095	1.513	2.016	2.590
3.610	3.721	1.144	2.656	4.972	7.344	1.135	1.576	2.088	2.714
5.290	4.489	1.188	2.788	5.152	7.672	1.204	1.608	2.125	2.735
8.410	5.041	1.276	2.992	5.244	7.896	1.218	1.672	2.143	2.926

base $\mapsto$ base^7

1.280	2.751	1.104	5.328	5.888	3.728	1.453	6.830	2.267	4.844
2.187	9.493	1.920	6.620	6.364	4.454	1.853	7.685	2.344	5.785
7.812	1.947	2.713	9.058	9.273	4.721	2.570	8.967	2.762	6.496
8.235	2.718	4.423	1.002	9.974	6.276	2.813	1.004	2.853	6.879
1.948	5.066	8.079	1.630	1.151	7.405	2.943	1.123	3.142	7.703
6.274	1.174	1.072	1.789	1.235	8.703	3.216	1.208	3.348	8.146
4.103	2.488	1.229	2.351	1.861	1.019	4.353	1.347	3.678	8.851
8.938	3.142	1.605	3.057	2.742	1.073	4.936	1.554	4.163	1.041
3.404	6.060	1.828	3.622	3.105	1.251	6.057	1.667	4.424	1.070
1.724	9.095	2.352	4.637	3.302	1.383	6.306	1.914	4.561	1.356

	d	1	2	3	4	5	6	7	8	9	$10^3 \cdot R$
$\#\{D_1 = d\}$	$(p_n)_{n=1}^{100}$	25	19	19	20	8	2	4	2	1	103.0
	$(p_n^2)_{n=1}^{100}$	35	24	9	5	8	5	5	3	6	63.90
	$(p_n^7)_{n=1}^{100}$	33	15	11	11	4	10	4	7	5	39.18
	$10^2 \cdot \log(1 + d^{-1})$	30.10	17.60	12.49	9.691	7.918	6.694	5.799	5.115	4.575	

FIG 16. *Illustrating the lack of base-invariance for the first one-hundred prime numbers. (In the two middle tables, the values of $S(p_n^2)$ and $S(p_n^7)$, respectively, are shown to four correct digits.)*

The next theorem is the main result for base-invariant significant digits. It shows that convex combinations as in Example 4.28(iv) are the only probability distributions with base-invariant significant digits. To put the argument in perspective, recall that the proof of the scale-invariance theorem (Theorem 4.20) ultimately depended on Theorem 4.13(i,ii) which in turn was proved analytically using Fourier analysis. The situation here is similar: An analytical result (Lemma 4.32 below) identifies all probability measures on $\big([0,1), \mathcal{B}[0,1)\big)$ that are invariant under *every* map $x \mapsto \langle nx \rangle$ on $[0,1)$. Once this tool is available, it is straightforward to prove

Theorem 4.30 (Base-invariance characterization [Hi1])**.** *A probability measure P on $(\mathbb{R}^+, \mathcal{A})$ with $\mathcal{A} \supset \mathcal{S}$ has base-invariant significant digits if and only if, for some $q \in [0,1]$,*

$$P(A) = q\delta_1(A) + (1-q)\mathbb{B}(A) \quad \text{for every } A \in \mathcal{S}\,. \tag{4.8}$$

	d	1	2	3	4	5	6	7	8	9	$10^3 \cdot R$
Fibonacci	original	3011	1762	1250	968	792	668	580	513	456	0.1574
	$b \mapsto b^2$	3012	1760	1248	971	791	672	577	513	456	0.2919
	$b \mapsto b^7$	3011	1762	1248	969	791	671	579	511	458	0.1532
Prime	original	1601	1129	1097	1069	1055	1013	1027	1003	1006	140.9
	$b \mapsto b^2$	2340	1437	1195	1036	944	844	775	745	684	67.02
	$b \mapsto b^7$	3012	1626	1200	987	798	716	609	536	516	36.85
$10^4 \cdot \log(1+d^{-1})$		3010.	1760.	1249.	969.1	791.8	669.4	579.9	511.5	457.5	

FIG 17. *Increasing the sample size from $N = 10^2$ to $N = 10^4$ makes the Fibonacci numbers' leading digits even more closely base-invariant. As in the case of scale-invariance, this is not at all true for the primes, cf.* FIG 13.

Corollary 4.31. *A continuous probability measure P on $\mathbb{R}^+$ has base-invariant significant digits if and only if $P(A) = \mathbb{B}(A)$ for all $A \in \mathcal{S}$, i.e., if and only if P is Benford.*

Recall that $\lambda_{0,1}$ denotes Lebesgue measure on $\big([0,1), \mathcal{B}[0,1)\big)$. For each $n \in \mathbb{N}$, denote the map $x \mapsto \langle nx \rangle$ of $[0,1)$ into itself by T_n. Generally, if $T : [0,1) \to \mathbb{R}$ is measurable, and $T\big([0,1)\big) \subset [0,1)$, a probability measure P on $\big([0,1), \mathcal{B}[0,1)\big)$ is said to be T-*invariant*, or T is P-*preserving*, if $T_*P = P$. Which probability measures are T_n-invariant for all $n \in \mathbb{N}$? A complete answer to this question is provided by

Lemma 4.32. *A probability measure P on $\big([0,1), \mathcal{B}[0,1)\big)$ is T_n-invariant for all $n \in \mathbb{N}$ if and only if $P = q\delta_0 + (1-q)\lambda_{0,1}$ for some $q \in [0,1]$.*

Proof. From the proof of Theorem 4.13 recall the definition of the Fourier coefficients of P,

$$\widehat{P}(k) = \int_0^1 e^{2\pi \imath k s} \mathrm{d}P(s) \,, \quad k \in \mathbb{Z} \,,$$

and observe that

$$\widehat{T_n P}(k) = \widehat{P}(nk) \quad \text{for all } k \in \mathbb{Z}, n \in \mathbb{N} \,.$$

Assume first that $P = q\delta_0 + (1-q)\lambda_{0,1}$ for some $q \in [0,1]$. From $\widehat{\delta_0}(k) \equiv 1$ and

$\widehat{\lambda_{0,1}}(k) = 0$ for all $k \neq 0$, it follows that

$$\widehat{P}(k) = \begin{cases} 1 & \text{if } k = 0\,, \\ q & \text{if } k \neq 0\,. \end{cases}$$

For every $n \in \mathbb{N}$ and $k \in \mathbb{Z}\backslash\{0\}$, therefore, $\widehat{T_n P}(k) = q$, and clearly $\widehat{T_n P}(0) = 1$. Thus $\widehat{T_n P} = \widehat{P}$ and since the Fourier coefficients of P determine P uniquely, $T_{n*}P = P$ for all $n \in \mathbb{N}$.

Conversely, assume that P is T_n-invariant for all $n \in \mathbb{N}$. In this case, $\widehat{P}(n) = \widehat{T_n P}(1) = \widehat{P}(1)$, and similarly $\widehat{P}(-n) = \widehat{T_n P}(-1) = \widehat{P}(-1)$. Since generally $\widehat{P}(-k) = \overline{\widehat{P}(k)}$, there exists $q \in \mathbb{C}$ such that

$$\widehat{P}(k) = \begin{cases} q & \text{if } k > 0\,, \\ 1 & \text{if } k = 0\,, \\ \overline{q} & \text{if } k < 0\,. \end{cases}$$

Also, observe that for every $t \in \mathbb{R}$

$$\lim\nolimits_{n\to\infty} \frac{1}{n} \sum\nolimits_{j=1}^{n} e^{2\pi\imath tj} = \begin{cases} 1 & \text{if } t \in \mathbb{Z}\,, \\ 0 & \text{if } t \notin \mathbb{Z}\,. \end{cases}$$

Using this and the Dominated Convergence Theorem, it follows from

$$P(\{0\}) = \int_0^1 \lim\nolimits_{n\to\infty} \frac{1}{n} \sum\nolimits_{j=1}^{n} e^{2\pi\imath sj} \mathrm{d}P(s) = \lim\nolimits_{n\to\infty} \frac{1}{n} \sum\nolimits_{j=1}^{n} \widehat{P}(j) = q\,,$$

that q is real, and in fact $q \in [0, 1]$. Hence the Fourier coefficients of P are exactly the same as those of $q\delta_0 + (1-q)\lambda_{0,1}$. By uniqueness, therefore, $P = q\delta_0 + (1-q)\lambda_{0,1}$. □

Remark. Note that P is T_{mn}-invariant if it is both T_m- and T_n-invariant. Thus, in Lemma 4.32 it is enough to require that P be T_n-invariant whenever n is a prime number.

It is natural to ask how small the set $\mathbb{M}$ of natural numbers n can be chosen for which T_n-invariance really has to be required in Lemma 4.32. By the observation just made, it can be assumed that $\mathbb{M}$ is closed under multiplication, hence a (multiplicative) semi-group. If $\mathbb{M}$ is *lacunary*, i.e. $\mathbb{M} \subset \{p^m : m \in \mathbb{N}\}$ for some $p \in \mathbb{N}$, then probability measures P satisfying $T_{n*}P = P$ for all $n \in \mathbb{M}$ exist in abundance, and hence an analogue of Lemma 4.32 cannot hold. If, on the other hand, $\mathbb{M}$ is not lacunary, then it is not known in general whether an appropriate analogue of Lemma 4.32 may hold. For example, if $\mathbb{M} = \{2^{m_1}3^{m_2} : m_1, m_2 \in \mathbb{N}_0\}$ then the probability measure $P = \frac{1}{4}\sum_{j=1}^{4} \delta_{j/5}$ is T_n-invariant for every $n \in \mathbb{M}$, but it is a famous open question of H. Furstenberg [Ei] whether any *continuous* probability measure with this property exists — except, of course, for $P = \lambda_{0,1}$. ♣

Proof of Theorem 4.30. As in the proof of Theorem 4.20, fix a probability measure P on $(\mathbb{R}^+, \mathcal{A})$, denote by P_0 its restriction to $(\mathbb{R}^+, \mathcal{S})$, and let $Q = \ell_* P_0$. Observe that P_0 has base-invariant significant digits if and only if Q is T_n-invariant for all $n \in \mathbb{N}$. Indeed, with $0 \le s < 1$ and $A = \{x > 0 : S_{10}(x) < 10^s\}$,

$$\begin{aligned} T_{n*}Q\bigl([0,s)\bigr) &= Q\left(\bigcup\nolimits_{j=0}^{n-1}\Bigl[\frac{j}{n}, \frac{j+s}{n}\Bigr)\right) \\ &= P_0\left(\bigcup\nolimits_{k\in\mathbb{Z}} 10^k \bigcup\nolimits_{j=0}^{n-1}[10^{j/n}, 10^{(j+s)/n})\right) \;=\; P_0(A^{1/n}) \end{aligned}$$

and hence $T_{n*}Q = Q$ for all n precisely if P_0 has base-invariant significant digits. In this case, by Lemma 4.32, $Q = q\delta_0 + (1-q)\lambda_{0,1}$ for some $q \in [0,1]$, which in turn implies that $P_0(A) = q\delta_1(A) + (1-q)\mathbb{B}(A)$ for every $A \in \mathcal{S}$. $\square$

Corollary 4.33. *If a probability measure on $\mathbb{R}^+$ has scale-invariant significant digits then it also has base-invariant significant digits.*

4.4. The sum-invariance characterization

No finite data set can obey BL exactly, since the Benford probabilities of sets with m given significant digits become arbitrarily small as m goes to infinity, and no discrete probability measure with finitely many atoms can take arbitrarily small positive values. But, as first observed by M. Nigrini [Ni], if a table of real data approximately follows BL, then the sum of the significands of all entries in the table with first significant digit 1 is very close to the sum of the significands of all entries with first significant digit 2, and to the sum of the significands of entries with the other possible first significant digits as well. This clearly implies that the table must contain more entries starting with 1 than with 2, more entries starting with 2 than with 3, and so forth. Similarly, the sums of significands of entries with $D_1 = d_1, \ldots, D_m = d_m$ are approximately equal for all tuples $(d_1, \ldots, d_m)$ of a fixed length m. In fact, even the sum-invariance of first or first and second digits yields a distribution close to BL, see Fig 18 and 19. Nigrini conjectured, and partially proved, that this *sum-invariance* property also characterizes BL. Note that it is the *significands* of the data, rather than the data themselves, that are summed up. Simply summing up the raw data will not lead to any meaningful conclusion, as the resulting sums may be dominated by a few very large numbers. It is only through considering significands that the magnitude of the individual numbers becomes irrelevant.

To motivate a precise definition of sum-invariance, note that if (x_n) is Benford then the set $\{x_n : n \in \mathbb{N}\}$ is necessarily infinite, and consequently, for every $d \in \{1, 2, \ldots, 9\}$, the sum $\sum_{n: D_1(x_n)=d} S(x_n)$ is infinite as well. To compare such sums, it is natural to normalise them by considering limiting averages. To this end, for every $m \in \mathbb{N}$, $d_1 \in \{1, 2, \ldots, 9\}$ and $d_j \in \{0, 1, \ldots, 9\}$, $j \ge 2$, define

$$S_{d_1,\ldots,d_m}(x) := \begin{cases} S(x) & \text{if } \bigl(D_1(x), \ldots, D_m(x)\bigr) = (d_1, \ldots, d_m)\,, \\ 0 & \text{otherwise.} \end{cases}$$

d	1	2	3	4	5	6	7	8	9
$N_d = \#\{x_n = d\}$	2520	1260	840	630	504	620	360	315	280
dN_d	2520	2520	2520	2520	2520	2520	2520	2520	2520
N_d/N	0.3535	0.1767	0.1178	0.0884	0.0707	0.0589	0.0505	0.0442	0.0393
$\log(1+d^{-1})$	0.3010	0.1761	0.1240	0.0969	0.0792	0.0669	0.0580	0.0512	0.0458

$N = \sum_d N_d = 7129$

FIG 18. *A (hypothetical) sample $x_1, x_2, \ldots x_N$ containing $N = 7129$ numbers from $\{1, 2, \ldots, 9\}$ and showing* exact *sum-invariance for the first digit. Note that the relative frequencies N_d/N are quite close to the Benford probabilities* $\log(1 + d^{-1})$.

(d_1, d_2)	$N_{d_1,d_2} = \#\{x_n = 10d_1 + d_2\}$	$N_{d_1,d_2}/N$	$\sum_{d_2} N_{d_1,d_2}/N$	$\log(1+d_1^{-1})$
$(1,0)$	6972037522971247716453380893531230355680	0.04258	0.30607	0.30102
$(1,1)$	6338215929973861560412164448664754868800	0.03871		
$(1,2)$	5810031269142706430377817411276025296400	0.03549		
$(1,3)$	5363105786900959781887216071947100273600	0.03276		
$(1,4)$	4980026802122319797466700638236593111200	0.03042		
$(1,5)$	4648025015314165144302253929020820237120	0.02839		
$(1,6)$	4357523451857029822783363058457018972300	0.02661		
$(1,7)$	4101198542924263362619635819724253150400	0.02505		
$(1,8)$	3873354179428470953585211607517350197600	0.02366		
$(1,9)$	3669493433142761956028095207121700187200	0.02241		
$(2,0)$	3486018761485623858226690446765615177840	0.02129		
$\vdots$	$\vdots$	$\vdots$		
$(8,9)$	783375002581039069264424819497891051200	0.00478		
$(9,0)$	774670835885694190717042321503470039520	0.00473	0.04510	0.04575
$(9,1)$	766157969557279968841030867421014324800	0.00468		
$(9,2)$	757830165540353012657976184079481560400	0.00463		
$(9,3)$	749681454082929861984234504680777457600	0.00458		
$(9,4)$	741706119465026352814189456758641527200	0.00453		
$(9,5)$	733898686628552391205619041424340037440	0.00448		
$(9,6)$	726253908642838303797227176409503162050	0.00444		
$(9,7)$	718766754945489455304472257065075294400	0.00439		
$(9,8)$	711432400303188542495242948319513301600	0.00434		
$(9,9)$	704246214441540173379129383184972763200	0.00430		

$N = \sum_{d_1,d_2} N_{d_1,d_2} = 163731975056100444033114230488313094880847 \approx 1.637 \cdot 10^{41}$

$(10d_1 + d_2)N_{d_1,d_2} \equiv 69720375229712477164533808935312303556800 \approx 6.972 \cdot 10^{40}$

FIG 19. *An (even more hypothetical) sample $x_1, x_2, \ldots x_N$ containing $N \approx 1.637 \cdot 10^{41}$ numbers from $\{10, 11, \ldots, 99\}$ and showing* exact *sum-invariance for the first two digits. When compared with the values in* FIG 18*, the relative frequencies $\sum_{d_2} N_{d_1,d_2}/N$ of the first digits are even closer to the Benford values* $\log(1 + d_1^{-1})$.

d	1	2	3	4	5	6	7	8	9
Fibonacci	42.71	43.82	44.75	40.35	43.28	38.67	37.10	59.21	38.58
Prime	37.67	47.68	65.92	89.59	42.17	12.80	29.30	17.20	9.700

$N = 10^2$ Exact sum-invariance: $10^2 \cdot \mathbb{E}S_d = \dfrac{100}{\ln 10} \approx 43.43$ for $d = 1, 2, \ldots, 9$

FIG 20. *Except for $d = 8$, the value of $\sum_{D_1=d} S$ does not vary much with d for the first one-hundred Fibonacci numbers, but it varies wildly for the first one-hundred primes.*

Definition 4.34. A sequence (x_n) of real numbers has *sum-invariant significant digits* if, for every $m \in \mathbb{N}$, the limit

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} S_{d_1,\ldots,d_m}(x_n)}{N}$$

exists and is independent of $d_1, \ldots, d_m$.

In particular, therefore, if (x_n) has sum-invariant significant digits then there exists $c > 0$ such that

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} S_{d_1}(x_n)}{N} = c$$

for all $d_1 = 1, 2, \ldots, 9$.

As will follow from Theorem 4.37 below, the sequence (2^n) and the Fibonacci sequence (F_n) have sum-invariant significant digits. Clearly, (10^n) does not have sum-invariant significant digits since all the first digits are 1, i.e. for all N,

$$\frac{\sum_{n=1}^{N} S_{d_1}(10^n)}{N} = \begin{cases} 1 & \text{if } d_1 = 1\,, \\ 0 & \text{if } d_1 \geq 2. \end{cases}$$

Not too surprisingly, the sequence (p_n) of prime numbers does not have sum-invariant significant digits either, see FIG 20.

The definitions of sum-invariance of significant digits for functions, distributions and random variables are similar, and it is in the context of distributions and random variables that the sum-invariance characterization of BL will be established. Informally, a probability distribution has sum-invariant significant digits if in a collection of numbers with that distribution, the sums of (the significands of) all entries with first significant digit 1 is the same as each of the sums of all entries with the other first significant digits; and the sum of all the entries with, say, first two significant digits 1 and 3, respectively, is the same as the sum of all entries with any other combination of first two significant digits, etc; and similarly for all other finite initial sequences of significant digits. In complete analogy to Definition 4.34, this is put more formally by

Definition 4.35. A random variable X has *sum-invariant significant digits* if, for every $m \in \mathbb{N}$, the value of $\mathbb{E}S_{d_1,\ldots,d_m}(X)$ is independent of $d_1, \ldots, d_m$.

Example 4.36. **(i)** If X is uniformly distributed on $[0,1)$, then X does not have sum-invariant significant digits. This follows from Theorem 4.37 below but can also be seen by a simple direct calculation. Indeed, for every $d_1 \in \{1,2,\ldots,9\}$,

$$\mathbb{E}S_{d_1}(X) = \sum\nolimits_{n\in\mathbb{N}} 10^n \int_{10^{-n}d_1}^{10^{-n}(d_1+1)} t\,\mathrm{d}t = \frac{2d_1+1}{18}\,,$$

which obviously depends on d_1.

(ii) Similarly, if $\mathbb{P}(X=1)=1$ then X does not have sum-invariant significant digits, as

$$\mathbb{E}S_{d_1}(X) = \begin{cases} 1 & \text{if } d_1 = 1\,, \\ 0 & \text{if } d_1 \ge 2\,. \end{cases}$$

(iii) Assume that X is Benford. For every $m \in \mathbb{N}$, $d_1 \in \{1,2,\ldots,9\}$ and $d_j \in \{0,1,\ldots,9\}$, $j \ge 2$,

$$\mathbb{E}S_{d_1,\ldots,d_m}(X) = \int_{d_1+10^{-1}d_2+\ldots+10^{1-m}d_m}^{d_1+10^{-1}d_2+\ldots+10^{1-m}(d_m+1)} t\cdot\frac{1}{t\ln 10}\,\mathrm{d}t = \frac{10^{1-m}}{\ln 10}\,.$$

Thus X has sum-invariant significant digits. Note, however, that even in this example the higher moments of $S_{d_1,\ldots,d_m}(X)$ generally depend on $d_1,\ldots,d_m$, as for instance

$$\mathbb{E}S_{d_1}(X)^2 = \frac{2d_1+1}{2\ln 10}\,, \quad d_1 = 1,2,\ldots,9\,.$$

This example shows that it would be too restrictive to require in Definition 4.35 that the *distribution* of the random variable $S_{d_1,\ldots,d_m}(X)$, rather than its expectation, be independent of $d_1,\ldots,d_m$.

According to Example 4.36(iii) every Benford random variable has sum-invariant significant digits. As hinted at earlier, the converse is also true, i.e., sum-invariant significant digits characterize BL.

Theorem 4.37 (Sum-invariance characterization [Al])**.** *A random variable X with $\mathbb{P}(X=0)=0$ has sum-invariant significant digits if and only if it is Benford.*

Proof. The "if"-part has been verified in Example 4.36(iii). To prove the "only if"-part, assume that X has sum-invariant significant digits. For every $m \in \mathbb{N}$, $d_1 \in \{1,2,\ldots,9\}$ and $d_j \in \{0,1,\ldots,9\}$, $j \ge 2$, let

$$\begin{aligned} J_{d_1,\ldots,d_m} &:= \bigl[d_1+10^{-1}d_2+\ldots+10^{1-m}d_m, d_1+10^{-1}d_2+\ldots+10^{1-m}(d_m+1)\bigr) \\ &= \bigl\{1 \le x < 10 : \bigl(D_1(x), D_2(x), \ldots, D_m(x)\bigr) = (d_1, d_2, \ldots d_m)\bigr\}\,. \end{aligned}$$

With this,

$$S_{d_1,\ldots,d_m}(X) = S(X)\mathbb{1}_{J_{d_1,\ldots,d_m}}\bigl(S(X)\bigr)\,,$$

and by assumption $\mathbb{E}S_{d_1,\ldots,d_m}(X)$ is independent of $d_1,\ldots,d_m$. Note that each of the $9\cdot 10^{m-1}$ intervals $J_{d_1,\ldots,d_m}$ has the same length $\lambda(J_{d_1,\ldots,d_m}) = 10^{1-m}$. Consequently,

$$\mathbb{E}S_{d_1,\ldots,d_m}(X) = \frac{1}{9\cdot 10^{m-1}}\mathbb{E}S(X) = \frac{\lambda(J_{d_1,\ldots,d_m})}{9}\mathbb{E}S(X)\,,$$

and since the family

$$\big\{J_{d_1,\ldots,d_m} : m\in\mathbb{N}, d_1\in\{1,2,\ldots,9\} \text{ and } d_j\in\{0,1,\ldots,9\},\, j\geq 2\big\}$$

generates $\mathcal{B}[1,10)$,

$$\mathbb{E}\big(S(X)\mathbb{1}_{[a,b)}\big(S(X)\big)\big) = \frac{b-a}{9}\,\mathbb{E}S(X) \tag{4.9}$$

holds for every $1\leq a<b<10$. Given any $1<t<10$, consider the sequence of functions (f_n), where $f_n:\mathbb{R}\to\mathbb{R}$ is given by

$$f_n = \sum\nolimits_{j=1}^{n} \frac{n}{n+(t-1)j}\mathbb{1}_{\left[1+(t-1)\frac{j-1}{n},1+(t-1)\frac{j}{n}\right)}\,.$$

Note that $f_n(\tau)\uparrow \frac{\mathbb{1}_{[1,t)}(\tau)}{\tau}$ as $n\to\infty$, uniformly in τ. Hence by the Monotone Convergence Theorem and (4.9),

$$\begin{aligned}
&\mathbb{P}(1\leq S(X)<t) = \mathbb{E}\mathbb{1}_{[1,t)}\big(S(X)\big) = \mathbb{E}\left(S(X)\frac{1}{S(X)}\mathbb{1}_{[1,t)}\big(S(X)\big)\right)\\
&\quad= \lim\nolimits_{n\to\infty}\mathbb{E}\left(S(X)f_n\big(S(X)\big)\right)\\
&\quad= \lim\nolimits_{n\to\infty}\sum\nolimits_{j=1}^{n}\frac{n}{n+(t-1)j}\mathbb{E}\left(S(X)\mathbb{1}_{\left[1+(t-1)\frac{j-1}{n},1+(t-1)\frac{j}{n}\right)}\big(S(X)\big)\right)\\
&\quad= \lim\nolimits_{n\to\infty}\sum\nolimits_{j=1}^{n}\frac{n}{n+(t-1)j}\cdot\frac{t-1}{9n}\,\mathbb{E}S(X)\\
&\quad= \frac{\mathbb{E}S(X)}{9}\lim\nolimits_{n\to\infty}\frac{t-1}{n}\sum\nolimits_{j=1}^{n}\frac{1}{1+(t-1)j/n}\\
&\quad= \frac{\mathbb{E}S(X)}{9}\int_0^1\frac{t-1}{1+(t-1)\sigma}\,\mathrm{d}\sigma\\
&\quad= \frac{\mathbb{E}S(X)}{9}\ln t\,.
\end{aligned}$$

From $\mathbb{P}(1\leq S(X)<10) = \mathbb{P}(X\neq 0) = 1$, it follows that $\mathbb{E}S(X) = \frac{9}{\ln 10}$ and hence

$$\mathbb{P}(S(X)<t) = \frac{\ln t}{\ln 10} = \log t \quad \text{for all } t\in[1,10)\,,$$

i.e., X is Benford. $\square$

Remarks. (i) As shown by Example 4.36(iii) and Theorem 4.37, a random variable X has sum-invariant significant digits if and only if

$$\mathbb{E}S_{d_1,\ldots,d_m}(X) = \frac{10^{1-m}}{\ln 10}\,\mathbb{P}(X \neq 0)$$

holds for all $m \in \mathbb{N}$, $d_1 \in \{1,2,\ldots,9\}$ and $d_j \in \{0,1,\ldots,9\}$, $j \geq 2$.

(ii) Theorem 4.37 provides another informal test for goodness-of-fit to BL: Simply calculate the differences between the sums of the significands of the data corresponding to the same initial sequence of significant digits, see [Ni]. ♣

5. Benford's Law for deterministic processes

The goal of this chapter is to present the basic theory of BL in the context of deterministic processes, such as iterates of maps, powers of matrices, and solutions of differential equations. Except for somewhat artificial examples, processes with linear growth are not Benford, and among the others, there is a clear distinction between those with exponential growth or decay, and those with super-exponential growth or decay. In the exponential case, processes typically are Benford for all starting points in a region, but are not Benford with respect to other bases. In contrast, super-exponential processes typically are Benford for all bases, but have small sets (of measure zero) of exceptional points whose orbits or trajectories are not Benford.

5.1. One-dimensional discrete-time processes

This section presents some elementary facts about BL for one-dimensional discrete-time processes. The focus is first on processes with exponential growth or decay, then on processes with doubly-exponential or more general growth or decay. Finally, some possible applications such as Newton's method, and extensions to nonautonomous and chaotic systems are discussed briefly.

Processes with exponential growth or decay

Many classical integer sequences exhibiting exponential growth are known to be Benford.

Example 5.1. (i) Recall from Examples 4.11(i) and 4.12 that (2^n) and the Fibonacci sequence (F_n) are Benford. Similarly, $(n!)$ is Benford [BBH, Di], see also Fig 21.

(ii) Recall from the remark on p.18 that (n) is not Benford, but weakly Benford in the sense explained there, and the same is true for the sequence of prime numbers.

$\#\{D_1 = d\}$	d	1	2	3	4	5	6	7	8	9
	(2^n)	301	176	125	97	79	69	56	52	45
	$(n!)$	293	176	124	102	69	87	51	51	47
	(F_n)	301	177	125	96	80	67	56	53	45
	$10^3 \cdot \log(1+d^{-1})$	301.0	176.0	124.9	96.91	79.18	66.94	57.99	51.15	45.75

FIG 21. *Empirical frequencies of D_1 for the first 10^3 terms of the sequences (2^n), $(n!)$ and the Fibonacci numbers (F_n), as compared with the Benford probabilities.*

Let $T : C \to C$ be a measurable map that maps $C \subset \mathbb{R}$ into itself, and for every $n \in \mathbb{N}$ denote by T^n the n-fold iterate of T, i.e. $T^1 := T$ and $T^{n+1} := T^n \circ T$; also let T^0 be the identity map id_C on C, that is, $T^0(x) = x$ for all $x \in C$. The *orbit* of $x_0 \in C$ is the sequence

$$O_T(x_0) := \left(T^{n-1}(x_0)\right)_{n\in\mathbb{N}} = \left(x_0, T(x_0), T^2(x_0), \dots\right).$$

Note that this interpretation of the orbit as a *sequence* differs from terminology sometimes used in dynamical systems theory (e.g. [KH]) according to which the orbit of x_0 is the mere *set* $\{T^{n-1}(x_0) : n \in \mathbb{N}\}$.

Example 5.2. (i) If $T(x) = 2x$ then $O_T(x_0) = (x_0, 2x_0, 2^2x_0, \dots) = (2^{n-1}x_0)$ for all x_0. Hence $\lim_{n\to\infty} |x_n| = +\infty$ whenever $x_0 \neq 0$.

(ii) If $T(x) = x^2$ then $O_T(x_0) = (x_0, x_0^2, x_0^{2^2}, \dots) = \left(x_0^{2^{n-1}}\right)$ for all x_0. Here x_n approaches 0 or $+\infty$ depending on whether $|x_0| < 1$ or $|x_0| > 1$. Moreover, $O_T(\pm 1) = (\pm 1, 1, 1, \dots)$.

(iii) If $T(x) = 1 + x^2$ then $O_T(x_0) = (x_0, 1 + x_0^2, 2 + 2x_0^2 + x_0^4, \dots)$. Since $x_n \geq n$ for all x_0 and $n \in \mathbb{N}$, $\lim_{n\to\infty} x_n = +\infty$ for every x_0.

Recall from Example 4.11(i) that (2^n) is Benford, and in fact $(2^n x_0)$ is Benford for every $x_0 \neq 0$, by Theorem 4.22. In other words, Example 5.2(i) says that with $T(x) = 2x$, the orbit $O_T(x_0)$ is Benford whenever $x_0 \neq 0$. The goal of the present sub-section is to extend this observation to a much wider class of maps T. The main result (Theorem 5.8) rests upon three simple lemmas.

Lemma 5.3. *Let $T(x) = ax$ with $a \in \mathbb{R}$. Then $O_T(x_0)$ is Benford for every $x_0 \neq 0$ or for no x_0 at all, depending on whether $\log |a|$ is irrational or rational, respectively.*

Proof. By Theorem 4.10, $O_T(x_0) = (a^{n-1}x_0)$ is Benford for every $x_0 \neq 0$ or none, depending on whether $\log |a|$ is irrational or not. $\square$

Example 5.4. (i) Let $T(x) = 4x$. Since $\log 4$ is irrational, $O_T(x_0) = (4^{n-1}x_0)$ is Benford for every $x_0 \neq 0$; in particular $O_T(4) = (4^n)$ is Benford. Note, however, that (4^n) is not base-2 Benford since $\log_2 4 = 2$ is rational, and correspondingly the second *binary* digit of 4^n is identically equal to zero, whereas for a

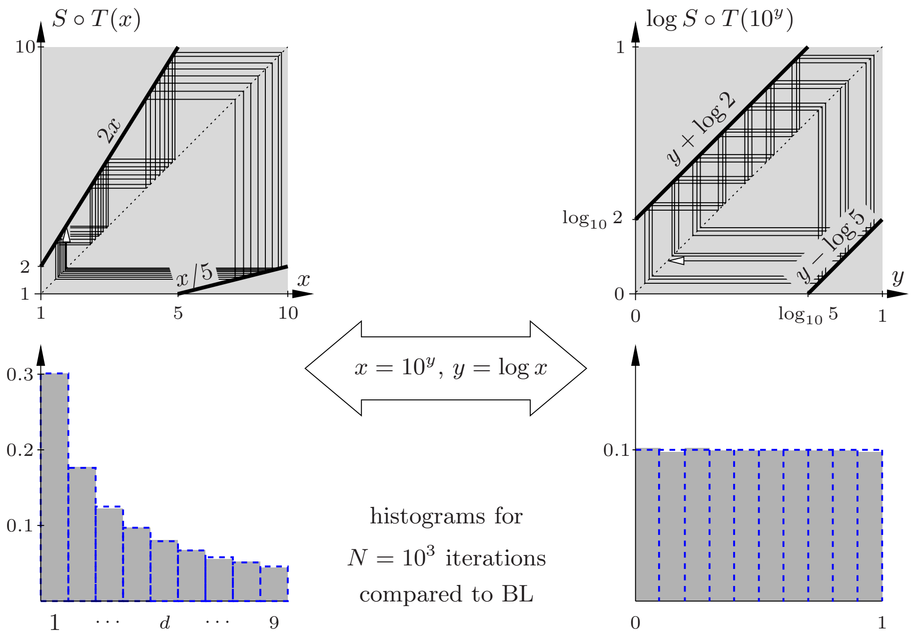

FIG 22. *With $T(x) = 2x$, $O_T(x_0)$ is Benford for all $x_0 \neq 0$.*

2-Benford sequence the second binary digit is zero only with a relative frequency of $\log_2(3/2) \approx 0.5850$.

(ii) Since $\log \pi$ is irrational, every orbit of $T(x) = \pi x$ is Benford, unless $x_0 = 0$. Here $O_T(x_0)$ is actually base-b Benford for every $b \in \mathbb{N}\backslash\{1\}$.

Clearly, the simple proof of Lemma 5.3 works only for maps that are *exactly* linear. The same argument would for instance not work for $T(x) = 2x + e^{-x}$ even though $T(x) \approx 2x$ for large x. To establish the Benford behavior of maps like this, a simple version of *shadowing* will be used. While the argument employed here is elementary, note that in dynamical systems theory, shadowing is a powerful and sophisticated tool, see e.g. [Pa].

To explain the basic idea, fix T as above, i.e. let $T(x) = 2x + e^{-x}$ and note first that $T(x) \geq \max(0, x+1)$ for all x, and hence $\lim_{n\to\infty} T^n(x_0) = +\infty$ for every x_0. While no explicit analytical expression is available for $T^n(x_0)$, it is certainly plausible to expect that, for large n, the orbit $O_T(x_0)$ should resemble an orbit of the linear map $x \mapsto 2x$. Fortunately, this is easily made rigorous. To this end, note that

$$T^n(x_0) = 2^n x_0 + \sum\nolimits_{j=1}^{n} 2^{n-j} e^{-T^{j-1}(x_0)}$$

holds for every $n \in \mathbb{N}$ and $x_0 \in \mathbb{R}$. Since $T^n(x_0) \geq 0$ for all n and x_0, the number

$$\overline{x_0} := x_0 + \sum\nolimits_{j=1}^{\infty} 2^{-j} e^{-T^{j-1}(x_0)} > x_0 + \frac{e^{-x_0}}{2}$$

is well-defined and positive, and a short calculation using the fact that $T^n(x) \geq x + n$ confirms that

$$\begin{aligned}\left|T^n(x_0) - 2^n\overline{x_0}\right| &= \sum\nolimits_{j=n+1}^{\infty} 2^{n-j} e^{-T^{j-1}(x_0)} \\ &\leq \sum\nolimits_{j=1}^{\infty} 2^{-j} e^{-(x_0+j+n-1)} = \frac{e^{1-n-x_0}}{2e-1}\,, \end{aligned} \tag{5.1}$$

and hence $|T^n(x_0) - 2^n\overline{x_0}| \to 0$ exponentially fast as $n \to \infty$. As will be seen shortly, this implies that $O_T(x_0)$ is Benford for all $x_0 \in \mathbb{R}$. Note also that even if $|T^n(x_0) - 2^n y|$ were merely required to remain *bounded* as $n \to \infty$, the only choice for y would still be $y = \overline{x_0}$. Moreover, $\overline{x_0}$ depends continuously on x_0. As the following lemma shows, these observations hold in greater generality.

Lemma 5.5 (Shadowing Lemma)**.** *Let $T : \mathbb{R} \to \mathbb{R}$ be a map, and β a real number with $|\beta| > 1$. If $\sup_{x\in\mathbb{R}} |T(x) - \beta x| < +\infty$ then there exists, for every $x \in \mathbb{R}$, one and only one point $\overline{x}$ such that the sequence $(T^n(x) - \beta^n\overline{x})$ is bounded.*

Proof. Let $\Delta(x) := T(x) - \beta x$ and note that $D := \sup_{x\in\mathbb{R}} |\Delta(x)| < +\infty$ by assumption. With this, for all $x \in \mathbb{R}$ and $n \in \mathbb{N}_0$,

$$T^n(x) = \beta^n x + \sum\nolimits_{j=1}^{n} \beta^{n-j} \Delta \circ T^{j-1}(x)\,.$$

Using this expression, together with the well-defined number

$$\overline{x} := x + \sum\nolimits_{j=1}^{\infty} \beta^{-j} \Delta \circ T^{j-1}(x)\,,$$

it follows that

$$\begin{aligned}|T^n(x) - \beta^n\overline{x}| &= \left|\sum\nolimits_{j=n+1}^{\infty} \beta^{n-j} \Delta \circ T^{j-1}(x)\right| \\ &\leq \sum\nolimits_{j=1}^{\infty} |\beta|^{-j} |\Delta \circ T^{j+n-1}(x)| \leq \frac{D}{|\beta| - 1}\,,\end{aligned}$$

and hence $(T^n(x) - \beta^n\overline{x})$ is bounded. Moreover, the identity

$$T^n(x) - \beta^n y = T^n(x) - \beta^n\overline{x} - \beta^n(y - \overline{x})$$

shows that $(T^n(x) - \beta^n y)$ is bounded only if $y = \overline{x}$. $\square$

Remarks. (i) From the proof of Lemma 5.5 it can be seen that the map $h : x \mapsto \overline{x}$ is continuous whenever T is continuous. In general, h need not be one-to-one. For example, $h(x) = 0$ for every x for which $O_T(x)$ is bounded. Also note that if $\lim_{|x|\to+\infty} |\Delta(x)| = 0$ then $\lim_{|x|\to+\infty} |h(x) - x| = 0$ as well. This is often the case in applications and may be used to improve the bounds on $|T^n(x) - \beta^n\overline{x}|$. For example, for the map $T(x) = 2x + e^{-x}$ considered above, the rough estimate

$$T^n(x_0) \geq 2^n\overline{x_0} - \frac{e^{-x_0}}{2e-1}$$

obtained from (5.1) can be substituted into (5.1) again, leading to the much more accurate

$$|T^n(x_0) - 2^n\overline{x_0}| = \mathcal{O}\left(e^{-2^n\overline{x_0}}\right) \quad \text{as } n \to \infty\,.$$

(ii) Stronger, quantitative versions of the Shadowing Lemma have been established. They are very useful for an analysis of BL in more complicated systems, see e.g. [BBH] or [Ber3]. ♣

Example 5.6. **(i)** Let $T(x) = 2x+1$. For this simple map, T^n can be computed explicitly, and it is illuminating to compare the explicit findings with Lemma 5.5. From

$$T^n(x) = 2^n x + 2^n - 1\,,$$

it is clear that $(T^n(x) - 2^n x)$ is unbounded for every $x \in \mathbb{R}$. However, using $\overline{x} := x + 1$, one obtains

$$T^n(x) - 2^n\overline{x} \equiv -1\,,$$

and hence $(T^n(x) - 2^n\overline{x})$ is bounded.

(ii) Strictly speaking, the map $T(x) = 2x + e^{-x}$ studied above does not meet the assumptions of Lemma 5.5, as $\Delta(x) = e^{-x}$ is not bounded for $x \to -\infty$. The conclusion of the lemma, however, does hold nevertheless because Δ is bounded on $\mathbb{R}^+$ and T maps $\mathbb{R}$ into $\mathbb{R}^+$. Put differently, $\overline{x}$ is well-defined for every $x \in \mathbb{R}$.

(iii) Let $T(x) = 2x - e^{-x}$. Note that T has a unique fixed point x^*, i.e. $T(x^*) = x^*$; numerically, $x^* \approx 0.5671$. Lemma 5.5 applies to T for $x > x^*$. To see this formally, replace $T(x)$ by $x^* + 2(x - x^*)$ whenever $x \le x^*$ and note that this modification of T does not affect $O_T(x_0)$ for $x_0 \ge x^*$. Thus for every $x \ge x^*$ there exists an $\overline{x}$ such that $(T^n(x) - 2^n\overline{x})$ is bounded. Lemma 5.7 below implies that $O_T(x_0)$ is Benford for all $x_0 > x^*$. Clearly, $O_T(x^*) = (x^*, x^*, x^*, \ldots)$ is not Benford. If $x_0 < x^*$ then $T^n(x_0) \to -\infty$ super-exponentially fast. The Benford properties of $O_T(x_0)$ in this case will be analyzed in the next sub-section.

The next lemma enables application of Lemma 5.5 to establish the Benford property for orbits of a wide class of maps.

Lemma 5.7. (i) *Assume that (a_n) and (b_n) are sequences of real numbers with $|a_n| \to +\infty$ and $\sup_{n\in\mathbb{N}} |a_n - b_n| < +\infty$. Then (b_n) is Benford if and only if (a_n) is Benford.*

(ii) *Suppose that the measurable functions $f, g : [0, +\infty) \to \mathbb{R}$ are such that $|f(t)| \to +\infty$ as $t \to +\infty$, and $\sup_{t\ge 0} |f(t) - g(t)| < +\infty$. Then f is Benford if and only if g is Benford.*

Proof. To prove (i), let $c := \sup_{n\in\mathbb{N}} |a_n - b_n| + 1$. By discarding finitely many terms if necessary, it can be assumed that $|a_n|, |b_n| \ge 2c$ for all n. From

$$
\begin{aligned}
-\log\left(1+\frac{c}{|a_n|-c}\right) &\le \log\frac{|b_n|}{|b_n|+c} \\
&\le \log\frac{|b_n|}{|a_n|} \\
&\le \log\frac{|a_n|+c}{|a_n|} \le \log\left(1+\frac{c}{|a_n|-c}\right),
\end{aligned}
$$

it follows that

$$
\big|\log|b_n|-\log|a_n|\big| = \left|\log\frac{|b_n|}{|a_n|}\right| \le \log\left(1+\frac{c}{|a_n|-c}\right) \to 0 \quad \text{as } n\to\infty\,.
$$

Lemma 4.3(i) now shows that $(\log|b_n|)$ is u.d. mod 1 if and only $(\log|a_n|)$ is.

The proof of (ii) is completely analogous. $\square$

Lemmas 5.5 and 5.7 can now easily be combined to produce the desired general result. The theorem is formulated for orbits converging to zero. As explained in the subsequent Example 5.9, a reciprocal version holds for orbits converging to $\pm\infty$.

Theorem 5.8 ([BBH])**.** *Let $T:\mathbb{R}\to\mathbb{R}$ be a C^2-map with $T(0)=0$. Assume that $0<|T'(0)|<1$. Then $O_T(x_0)$ is Benford for all $x_0\neq 0$ sufficiently close to 0 if and only if $\log|T'(0)|$ is irrational. If $\log|T'(0)|$ is rational then $O_T(x_0)$ is not Benford for any x_0 sufficiently close to 0.*

Proof. Let $\alpha := T'(0)$ and observe that there exists a continuous function $f:\mathbb{R}\to\mathbb{R}$ such that $T(x)=\alpha x\big(1-xf(x)\big)$. In particular, $T(x)\neq 0$ for all $x\neq 0$ sufficiently close to 0. Define

$$
\widetilde{T}(x) := T(x^{-1})^{-1} = \frac{x^2}{\alpha\big(x-f(x^{-1})\big)}\,,
$$

and note that

$$
\widetilde{T}(x)-\alpha^{-1}x = \frac{x}{\alpha}\cdot\frac{f(x^{-1})}{x-f(x^{-1})} = \frac{f(x^{-1})}{\alpha} + \frac{f(x^{-1})^2}{\alpha\big(x-f(x^{-1})\big)}\,.
$$

From this it is clear that $\sup_{|x|\ge\xi}|\widetilde{T}(x)-\alpha^{-1}x|$ is finite, provided that ξ is sufficiently large. Hence Lemma 5.5 shows that for every x with $|x|$ sufficiently large, $\big(|\widetilde{T}^n(x)-\alpha^{-n}\overline{x}|\big)$ is bounded with an appropriate $\overline{x}\neq 0$. Lemma 5.7 implies that $O_{\widetilde{T}}(x_0)$ is Benford if and only if $(\alpha^{1-n}\overline{x_0})$ is, which in turn is the case precisely if $\log|\alpha|$ is irrational. The result then follows from noting that, for all $x_0\neq 0$ with $|x_0|$ sufficiently small, $O_T(x_0)=\big(\widetilde{T}^{n-1}(x_0^{-1})^{-1}\big)_{n\in\mathbb{N}}$, and Corollary 4.7(i) which shows that (x_n^{-1}) is Benford whenever (x_n) is. $\square$

Example 5.9. (i) For $T(x)=\frac{1}{2}x+\frac{1}{4}x^2$, the orbit $O_T(x_0)$ is Benford for every $x_0\neq 0$ sufficiently close to 0. A simple graphical analysis shows that $\lim_{n\to\infty}T^n(x)=0$ if and only if $-4<x<2$. Thus for every $x_0\in(-4,2)\backslash\{0\}$,

$O_T(x_0)$ is Benford. Clearly, $O_T(-4) = (-4, 2, 2, \dots)$ and $O_T(2) = (2, 2, 2, \dots)$ are not Benford. For $x_0 < -4$ or $x_0 > 2$, one might try to mimic the proof of Theorem 5.8 and consider

$$\widetilde{T}(x) := T(x^{-1})^{-1} = \frac{4x^2}{1+2x}$$

near $x = 0$. Note that indeed $\widetilde{T}$ is a smooth (C^∞) map near $x = 0$, and $\widetilde{T}(0) = 0$. However, $\widetilde{T}'(0) = 0$ as well, and Theorem 5.8 does not apply. It will follow from the main result of the next subsection (Theorem 5.12) that for almost every point $x_0 \in \mathbb{R}\backslash[-4, 2]$ the orbit $O_T(x_0)$ is Benford. However, $\mathbb{R}\backslash[-4, 2]$ also contains a large set of exceptional points, i.e. points whose orbit is *not* Benford.

(ii) To see that Theorem 5.8 applies to the map $T(x) = 2x + e^{-x}$ considered in Example 5.6(ii), let

$$\widetilde{T}(x) := T(x^{-2})^{-1/2} = \frac{x}{\sqrt{2 + x^2 e^{-1/x^2}}}\,, \quad x \neq 0\,.$$

With $\widetilde{T}(0) := 0$, the map $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ is smooth, and $\widetilde{T}'(0) = \frac{1}{\sqrt{2}}$. Moreover, $\lim_{n\to\infty} \widetilde{T}^n(x) = 0$ for every $x \in \mathbb{R}$. By Theorem 5.8, $O_{\widetilde{T}}(x_0)$ is Benford for every $x_0 \neq 0$, and hence $O_T(x_0)$ is Benford for every $x_0 \neq 0$ as well, because $T^n(x) = \widetilde{T}^n(|x|^{-1/2})^{-2}$ for all n.

(iii) As in (ii), Theorem 5.8 applies to the map $T(x) = 10x + e^{2-x}$. Note that again $\lim_{n\to\infty} T^n(x) = +\infty$ for every $x \in \mathbb{R}$, but since $\log 10$ is rational, no T-orbit is Benford. In fact, it is not hard to see that for every $m \in \mathbb{N}$ and $x \in \mathbb{R}$, the sequence of m-th significant digits of $T^n(x)$, i.e. $\big(D_m(T^n(x))\big)_{n\in\mathbb{N}}$ is eventually constant.

Remark. Theorem 5.8 remains essentially unchanged if the case $|T'(0)| = 1$ is also allowed, the conclusion being that in this case $O_T(x_0)$ is not Benford for any x near 0. However, this extension requires the explicit *assumption* that $x = 0$ be attracting, see [Ber4]. (If $|T'(0)| < 1$ then $x = 0$ is automatically attracting.)

For a simple example, consider the smooth map $T(x) = \sqrt{1 + x^2}$. While $\lim_{n\to\infty} T^n(x) = +\infty$ for every $x \in \mathbb{R}$, it follows from the explicit formula $T^n(x) = \sqrt{n + x^2}$ that $O_T(x_0)$ is not Benford, as $(\log \sqrt{n + x_0^2})$ is not u.d. mod 1, by Proposition 4.8(iv). The extended version of Theorem 5.8 just mentioned easily leads to the same conclusion because

$$\widetilde{T}(x) := T(x^{-1})^{-1} = \frac{x}{\sqrt{x^2+1}}$$

is smooth, with $\widetilde{T}(0) = 0$ and $\widetilde{T}'(0) = 1$, and $x = 0$ is an attracting fixed point for $\widetilde{T}$.

To see that the situation can be more complicated if $|T'(0)| = 1$ yet $x = 0$ is not attracting, fix $\alpha > 1$ and consider the map

$$T_\alpha(x) = \alpha x - \frac{(\alpha-1)x}{1+x^2}\,,$$

for which $T_\alpha(0) = 0$, $T'_\alpha(0) = 1$, and $x = 0$ is *repelling*. As far as the dynamics near $x = 0$ is concerned, all maps T_α are the same. However,

$$\widetilde{T}_\alpha(x) = T_\alpha(x^{-1})^{-1} = x\,\frac{1+x^2}{\alpha + x^2}$$

is smooth with $\widetilde{T}'_\alpha(0) = \alpha^{-1}$. Hence it is clear from Theorem 5.8 that $O_{T_\alpha}(x_0)$ is Benford for all $x_0 \neq 0$ or for none, depending on whether $\log \alpha$ is irrational or not. ♣

Processes with super-exponential growth or decay

As was seen in the previous subsection, for the maps

$$T : x \mapsto \alpha(x + e^{-x})$$

with $\alpha > 1$, either all orbits are Benford (if $\log \alpha$ is irrational) or else none are (if $\log \alpha$ is rational). This all-or-nothing behavior is linked to the exponential growth of orbits since, by the Shadowing Lemma 5.5,

$$T^n(x) = \alpha^n \overline{x} + \mathcal{O}(e^{-n}) \quad \text{as } n \to \infty\,.$$

For an altogether different scenario, consider the smooth map

$$T : x \mapsto \sqrt{30 + 12x^2 + x^4}\,.$$

As before, $\lim_{n\to\infty} T^n(x) = +\infty$ for every $x \in \mathbb{R}$. However, it follows from $T(x)^2 + 6 = (x^2+6)^2$ that

$$T^n(x) = \sqrt{(x^2+6)^{2^n} - 6} = (x^2+6)^{2^{n-1}} + \mathcal{O}\big(6^{-2^{n-1}}\big) \quad \text{as } n \to \infty\,,$$

showing that every T-orbit grows at a *doubly*-exponential rate. Is $O_T(x_0)$ Benford for some or even all $x_0 \in \mathbb{R}$? The main result of this subsection, Theorem 5.12 below, shows that indeed $O_T(x_0)$ is Benford for *most* x_0. While it is difficult to explicitly produce even a single x_0 with this property, it is very easy to see that $O_T(x_0)$ cannot be Benford for *every* x_0. Indeed, taking for example $x_0 = 2$, one obtains

$$O_T(2) = (2, \sqrt{94}, \sqrt{9994}, \sqrt{999994}, \ldots)\,,$$

and it is clear that $D_1\big(T^n(2)\big) = 9$ for every $n \in \mathbb{N}$. Hence $O_T(2)$ is *not* Benford. For another example, choose $x_0 = \sqrt{10^{4/3} - 6} = 3.943\ldots$ for which the sequence of first significant digits is eventually 2-periodic,

$$\big(D_1(T^{n-1}(x_0))\big) = (3, 2, 4, 2, 4, 2, 4, \ldots)\,.$$

As also shown by Theorem 5.12, for maps like T there are always *many* exceptional points.

The following is an analog of Lemma 5.3 in the doubly-exponential setting. Recall that a statement holds for *almost every* x if there is a set of Lebesgue measure zero that contains all x for which the statement does *not* hold.

Lemma 5.10. *Let $T(x) = \alpha x^\beta$ for some $\alpha > 0$ and $\beta > 1$. Then $O_T(x_0)$ is Benford for almost every $x_0 > 0$, but there also exist uncountably many exceptional points, i.e. $x_0 > 0$ for which $O_T(x_0)$ is not Benford.*

Proof. Note first that letting $\widetilde{T}(x) = cT(c^{-1}x)$ for any $c > 0$ implies $O_T(x) = c^{-1}O_{\widetilde{T}}(cx)$, and with $c = \alpha^{(\beta-1)^{-1}}$ one finds $\widetilde{T}(x) = x^\beta$. Without loss of generality, it can therefore be assumed that $\alpha = 1$, i.e. $T(x) = x^\beta$. Define $R : \mathbb{R} \to \mathbb{R}$ as $R(y) = \log T(10^y) = \beta y$. Since $x \mapsto \log x$ establishes a bijective correspondence between both the points and the nullsets in $\mathbb{R}^+$ and $\mathbb{R}$, respectively, all that has to be shown is that $O_R(y)$ is u.d. mod 1 for a.e. $y \in \mathbb{R}$, but also that $O_R(y)$ fails to be u.d. mod 1 for at least uncountably many y. To see the former, let $f_n(y) = R^n(y) = \beta^n y$. Clearly, $f_n'(y) - f_m'(y) = \beta^{n-m}(\beta^m - 1)$ is monotone, and $|f_n' - f_m'| \geq \beta - 1 > 0$ whenever $m \neq n$. By Proposition 4.9, therefore, $O_R(y)$ is u.d. mod 1 for a.e. $y \in \mathbb{R}$.

The statement concerning exceptional points will be proved here only under the additional assumption that β is an *integer*, see [Ber4] for the remaining cases. Given an integer $\beta \geq 2$, let (η_n) be any sequence of 0s and 1s such that $\eta_n \eta_{n+1} = 0$ for all $n \in \mathbb{N}$, that is, (η_n) does not contain two consecutive 1s. With this, consider

$$y_0 := \sum\nolimits_{j=1}^{\infty} \eta_j \beta^{-j}$$

and observe that, for every $n \in \mathbb{N}$,

$$0 \leq \langle \beta^n y_0 \rangle = \sum\nolimits_{j=n+1}^{\infty} \eta_j \beta^{n-j} \leq \frac{1}{\beta} + \frac{1}{\beta^2(\beta-1)} < 1\,,$$

from which it is clear that $(\beta^n y_0)$ is not u.d. mod 1. The proof is completed by noting that there are uncountably many different sequences (η_n), and each sequence defines a different point y_0. □

Example 5.11. Let $T(x) = x^2$. By Lemma 5.10, $O_T(x_0)$ is Benford for almost every but not for *every* $x_0 \in \mathbb{R}$, as for instance $T^n(x) = x^{2^n}$ always has first significant digit $D_1 = 1$ if $x = 10^k$ for some $k \in \mathbb{Z}$.

To study maps like $T(x) = \sqrt{30 + 12x^2 + x^4}$ mentioned above, Lemma 5.10 has to be extended. Note that

$$\widetilde{T}(x) = T(x^{-1})^{-1} = \frac{x^2}{\sqrt{1 + 12x^2 + 30x^4}}\,,$$

so $\widetilde{T}(x) \approx x^2$ near $x = 0$. Again the technique of shadowing can be applied to relate the dynamics of $\widetilde{T}$ to the one of $x \mapsto x^2$ covered by Lemma 5.10. The following is an analog of Theorem 5.8 for the case when T is dominated by power-like terms.

Theorem 5.12 ([BBH]). *Let T be a smooth map with $T(0) = 0$, and assume that $T'(0) = 0$ but $T^{(p)}(0) \neq 0$ for some $p \in \mathbb{N}\backslash\{1\}$. Then $O_T(x_0)$ is Benford for almost every x_0 sufficiently close to 0, but there are also uncountably many exceptional points.*

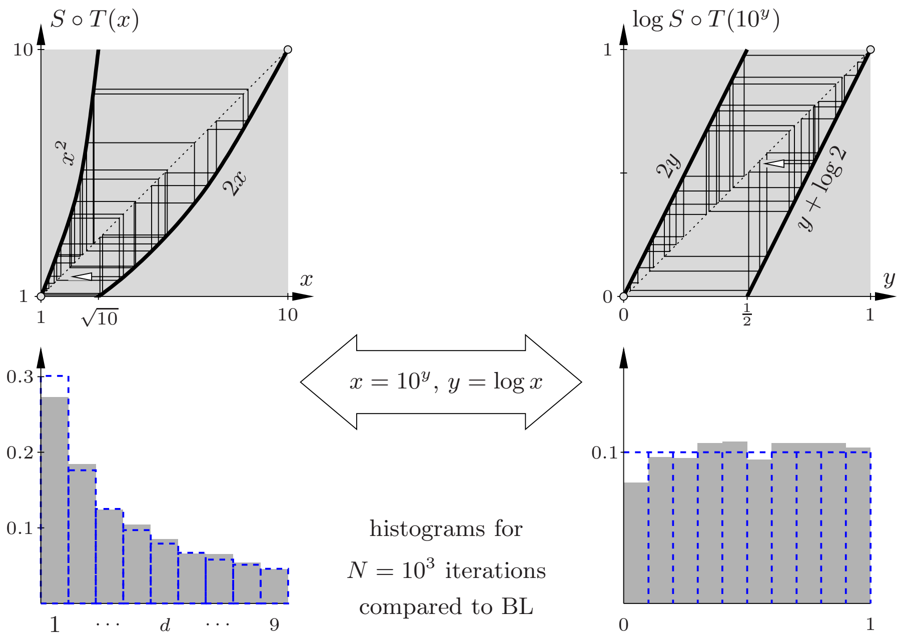

FIG 23. *With $T(x) = x^2$, $O_T(x_0)$ is Benford for almost every, but not every $x_0 \in \mathbb{R}$.*

Proof. Without loss of generality, assume that $p = \min\{j \in \mathbb{N} : T^{(j)}(0) \neq 0\}$. The map T can be written in the form $T(x) = \alpha x^p\big(1 + f(x)\big)$ where f is a C^∞-function with $f(0) = 0$, and $\alpha \neq 0$. As in the proof of Lemma 5.10, it may be assumed that $\alpha = 1$. Let $R(y) = -\log T(10^{-y}) = py - \log\big(1 + f(10^{-y})\big)$, so that $O_T(x_0)$ is Benford if and only if $O_R(-\log x_0)$ is u.d. mod 1. As the proof of Lemma 5.10 has shown, $(p^n y)$ is u.d. mod 1 for a.e. $y \in \mathbb{R}$. Moreover, Lemma 5.5 applies to R, and it can be checked by term-by-term differentiation that the shadowing map

$$h : y \mapsto \overline{y} = y - \sum\nolimits_{j=1}^{\infty} p^{-j} \log\Big(1 + f\big(10^{-R^j(y)}\big)\Big)$$

is a C^∞-diffeomorphism on $[y_0, +\infty)$ for y_0 sufficiently large. For a.e. sufficiently large y, therefore, $O_R(y)$ is u.d. mod 1. As explained earlier, this means that $O_T(x_0)$ is Benford for a.e. x_0 sufficiently close to 0. The existence of exceptional points follows similarly as in the proof of Lemma 5.10. $\square$

Example 5.13. (i) Consider the map $T(x) = \frac{1}{2}(x^2 + x^4)$ and note that $\lim_{n\to\infty} T^n(x) = 0$ if and only if $|x| < 1$. Theorem 5.12 shows that $O_T(x_0)$ is Benford for a.e. $x_0 \in (-1, 1)$. If $|x| > 1$ then $\lim_{n\to\infty} T^n(x) = +\infty$, and the reciprocal version of Theorem 5.12 applies to

$$\widetilde{T}(x) := T(x^{-1})^{-1} = \frac{2x^4}{1 + x^2}$$

near $x = 0$. Overall, therefore, $O_T(x_0)$ is Benford for a.e. $x_0 \in \mathbb{R}$.

(ii) For $T(x) = \sqrt{30 + 12x^2 + x^4}$, Theorem 5.12 applied to

$$\widetilde{T}(x) := T(x^{-1})^{-1} = \frac{x^2}{\sqrt{1 + 12x^2 + 30x^4}}$$

shows that $O_T(x_0)$ is Benford for a.e. $x_0 \in \mathbb{R}$.

(iii) Let $T(x) = 1 + x^2$. Again Theorem 5.12 applied to

$$\widetilde{T}(x) = T(x^{-1})^{-1} = \frac{x^2}{1 + x^2}\,,$$

shows that $O_T(x_0)$ is Benford for a.e. $x_0 \in \mathbb{R}$. As also asserted by that theorem, there are many exceptional points as well. For example, it can be shown that with

$$x_0 = \lim_{n\to\infty} \sqrt{\ldots \sqrt{\sqrt{10^{2^n} - 1} - 1} \ldots} = 9.949\ldots\,,$$

the first significant digit of $T^{n-1}(x_0)$ always equals 9, i.e. $D_1\big(T^{n-1}(x_0)\big) = 9$ for all $n \in \mathbb{N}$. (In fact, x_0 is the only point with this property, see [BBH] for details.)

Remarks. (i) Note that while in Lemma 5.3 and Theorem 5.8 $O_T(x_0)$ is Benford either for all x_0 or for none at all, Lemma 5.10 and Theorem 5.12 guarantee the coexistence of many x_0 for which $O_T(x_0)$ is Benford and many exceptional points. The latter form an uncountable set of Lebesgue measure zero. From a measure-theoretic point of view, therefore, exceptional points are extremely rare. It can be shown, however, that the points x_0 for which $O_T(x_0)$ is Benford form a *set of first category*, i.e. a countable union of nowhere dense sets. In particular, the exceptional points are *dense* in a neighbourhood of $x = 0$. (Recall that a set M is dense in $C \subset \mathbb{R}$ if, given any $c \in C$ and $\varepsilon > 0$, there exists an $m \in M$ with $|m - c| < \varepsilon$.) Thus from a topological point of view, most points are exceptional. This discrepancy between the measure-theoretic and the topological point of view is not uncommon in ergodic theory and may explain why it is difficult to explicitly find even a single point x_0 for which $O_T(x_0)$ is Benford for, say, $T(x) = 1 + x^2$ — despite the fact that Theorem 5.12 guarantees the existence of such points in abundance.

(ii) Theorem 5.12 covers for instance all polynomial or rational functions of degree at least two, for $|x|$ sufficiently large. An example not covered by that theorem is $T(x) = e^x$ or, more precisely, its reciprocal $\widetilde{T}(x) = e^{-1/x}$. In this case, $O_T(x_0)$ grows even faster than doubly-exponential. Theorem 5.21 below shows that nevertheless $O_T(x_0)$ is Benford for a.e. $x_0 \in \mathbb{R}$. Again, there is also a (measure-theoretically small yet topologically large) set of exceptional points.

(iii) In the context of Lemma 5.10 and Theorem 5.12, and in view of (i), many interesting questions may be asked. For instance, $O_T(x_0)$ is Benford for a.e. $x_0 \in \mathbb{R}$ if $T(x) = x^2$. What if $x_0 = 2$, i.e., is $O_T(2) = (2^{2^{n-1}})$ Benford? More generally, let T be any polynomial with integer coefficients and degree at least two. Then $O_T(x_0)$ is Benford for almost all sufficiently large $|x_0|$. Is $O_T(k)$ Benford for some, or even many integers k? In the case of $T(x) = x^2$, this is equivalent to asking whether $(2^n \log |k|)$ is u.d. mod 1 or, in number-theoretic terminology, whether $\log |k|$ is 2-*normal*. At present, 2-normality of common mathematical constants such as $\log 2$, π or e is a well-known open problem, considered to be exceedingly difficult. Similarly, one may ask whether (F_{2^n}) is Benford. Again, this may be a very hard problem, contrasting the simple fact that $(F_{|P(n)|})$ is Benford whenever P is a non-constant polynomial with integer coefficients. ♣

To conclude the present section on one-dimensional processes, a few possible applications and extensions of the results above will be discussed. The presentation is very brief and mostly based on examples; for any details, the interested reader may wish to consult the references mentioned in the text.

An application: Newton's method and related algorithms

In scientific calculations using digital computers and floating point arithmetic, roundoff errors are inevitable, and as Knuth points out in his classic text *The Art of Computer Programming* [Kn, pp.253–255]

> In order to analyze the average behavior of floating-point arithmetic algorithms (and in particular to determine their average running time), we need some statistical information that allows us to determine how often various cases arise . . . [If, for example, the] leading digits tend to be small [that] makes the most obvious techniques of "average error" estimation for floating-point calculations invalid. The relative error due to rounding is usually . . . more than expected.

Thus for the problem of finding numerically the root of a function by means of *Newton's Method* (NM), it is important to study the distribution of significant digits (or significands) of the approximations generated by the method. As will be seen shortly, the differences between successive Newton approximations, and the differences between the successive approximations and the unknown root often exhibit exactly the type of non-uniformity of significant digits alluded to by Knuth — they typically follow BL.

Throughout this subsection, let $f : I \to \mathbb{R}$ be a differentiable function defined on some open interval $I \subset \mathbb{R}$, and denote by N_f the map associated with f by NM, that is

$$N_f(x) := x - \frac{f(x)}{f'(x)} \quad \text{for all } x \in I \text{ with } f'(x) \neq 0.$$

For N_f to be defined wherever f is, set $N_f(x) := x$ if $f'(x) = 0$. Using NM for finding roots of f (i.e. real numbers x^* with $f(x^*) = 0$) amounts to picking an initial point $x_0 \in I$ and iterating N_f. Henceforth, (x_n) will denote the sequence of iterates of N_f starting at x_0, that is $(x_n) = O_{N_f}(x_0)$.

Clearly, if (x_n) converges to x^*, say, and if N_f is continuous at x^*, then $N_f(x^*) = x^*$, so x^* is a fixed point of N_f, and $f(x^*) = 0$. (Note that according to the definition of N_f used here, $N_f(x^*) = x^*$ could also mean that $f'(x^*) = 0$. If, however, $f'(x^*) = 0$ yet $f(x^*) \neq 0$ then N_f is not continuous at x^* unless f is constant.) It is this correspondence between the roots of f and the fixed points of N_f that makes NM work locally. Often, every fixed point x^* of N_f is *attracting*, i.e. $\lim_{n\to\infty} N_f^n(x_0) = x^*$ for all x_0 sufficiently close to x^*. (Observe that if f is *linear* near x^*, i.e. $f(x) = c(x - x^*)$ for some $c \neq 0$, then $N_f(x) = x^*$ for all x near x^*.)

To formulate a result about BL for NM, it will be assumed that $f : I \to \mathbb{R}$ is *real-analytic*. Recall that this means that f can, in a neighbourhood of every point of I, be represented by its Taylor series. Although real-analyticity is a strong assumption indeed, the class of real-analytic functions covers most practically relevant cases, including all polynomials, and all rational, exponential, and trigonometric functions, and compositions thereof.

If $f : I \to \mathbb{R}$ is real-analytic and $x^* \in I$ a root of f, i.e. if $f(x^*) = 0$, then $f(x) = (x - x^*)^m g(x)$ for some $m \in \mathbb{N}$ and some real-analytic $g : I \to \mathbb{R}$ with $g(x^*) \neq 0$. The number m is the *multiplicity* of the root x^*; if $m = 1$ then x^* is referred to as a *simple* root. The following theorem becomes plausible upon observing that $f(x) = (x - x^*)^m g(x)$ implies that N_f is real-analytic in a neighbourhood of x^*, and

$$\begin{aligned} N_f'(x) &= \frac{f(x)f''(x)}{f'(x)^2} \\ &= \frac{m(m-1)g(x)^2 + 2m(x - x^*)g'(x)g(x) + (x - x^*)^2 g''(x)g(x)}{m^2 g(x)^2 + 2m(x - x^*)g'(x)g(x) + (x - x^*)^2 g'(x)^2}, \end{aligned}$$

so that in particular $N_f'(x^*) = 1 - m^{-1}$.

Theorem 5.14 ([BH1])**.** *Let $f : I \to \mathbb{R}$ be real-analytic with $f(x^*) = 0$, and assume that f is not linear.*

(i) *If x^* is a simple root, then $(x_n - x^*)$ and $(x_{n+1} - x_n)$ are both Benford for (Lebesgue) almost every, but not every x_0 in a neighbourhood of x^*.*

(ii) *If x^* is a root of multiplicity at least two, then $(x_n - x^*)$ and $(x_{n+1} - x_n)$ are Benford for all $x_0 \neq x^*$ sufficiently close to x^*.*

The full proof of Theorem 5.14 can be found in [BH1]. It uses the following lemma which may be of independent interest for studying BL in other numerical approximation procedures. Part (i) is an analog of Lemma 5.7, and (ii) and (iii) follow directly from Theorem 5.12 and 5.8, respectively.

Lemma 5.15. *Let $T : I \to I$ be C^∞ with $T(y^*) = y^*$ for some $y^* \in I$.*

(i) *If $T'(y^*) \neq 1$, then for all y_0 such that $\lim_{n\to\infty} T^n(y_0) = y^*$, the sequence $(T^n(y_0) - y^*)$ is Benford precisely when $\big(T^{n+1}(y_0) - T^n(y_0)\big)$ is Benford.*

(ii) *If $T'(y^*) = 0$ but $T^{(p)}(y^*) \neq 0$ for some $p \in \mathbb{N}\backslash\{1\}$, then $(T^n(y_0) - y^*)$ is Benford for (Lebesgue) almost every, but not every y_0 in a neighbourhood of y^*.*

(iii) *If* $0 < |T'(y^*)| < 1$*, then* $(T^n(y_0) - y^*)$ *is Benford for all* $y_0 \neq y^*$ *sufficiently close to* y^* *precisely when* $\log |T'(y^*)|$ *is irrational.*

Example 5.16. **(i)** Let $f(x) = x/(1-x)$ for $x < 1$. Then f has a simple root at $x^* = 0$, and $N_f(x) = x^2$. By Theorem 5.14(i), the sequences (x_n) and $(x_{n+1} - x_n)$ are both Benford sequences for (Lebesgue) almost every x_0 in a neighbourhood of 0.

(ii) Let $f(x) = x^2$. Then f has a double root at $x^* = 0$ and $N_f(x) = x/2$, so by Theorem 5.14(ii), the sequence of iterates (x_n) of N_f as well as $(x_{n+1} - x_n)$ are both Benford for all starting points $x_0 \neq 0$. (They are not, however, 2-Benford.)

Utilizing Lemma 5.15, an analog of Theorem 5.14 can be established for other root-finding algorithms as well.

Example 5.17. Let $f(x) = x + x^3$ and consider the successive approximations (y_n) generated by the Jacobi-Steffensen method,

$$y_{n+1} = y_n - \frac{f(y_n)^2}{f(y_n) - f\big(y_n - f(y_n)\big)}\,, \quad n \in \mathbb{N}_0\,.$$

For almost every, but not every y_0 near 0, (y_n) is Benford. This follows from Lemma 5.15(ii), since $y_n = J_f^n(y_0)$ with the Jacobi-Steffensen transformation

$$J_f^n(x) = -y^5 \frac{1 - y^2}{1 + y^2 - y^4 + y^6}\,,$$

and $J_f(y) \approx -y^5$ near $y = 0$. Alternatively, $J_f = N_{\tilde{f}}$ with the real-analytic function $\tilde{f}(x) = (x + x^3)e^{\frac{1}{4}x^4 - x^2}$, so Theorem 5.14(i) applies directly as well.

If f fails to be real-analytic, then N_f may not be well-behaved analytically. For instance, N_f may have discontinuities even if f is C^∞. Pathologies like this can cause NM to fail for a variety of reasons, of which the reader can gain an impression from [BH1, Sec.4]. Even if N_f is smooth, (x_n) may not be Benford.

Example 5.18. Let f be the C^∞-function

$$f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \neq 0\,, \\ 0 & \text{if } x = 0\,, \end{cases}$$

for which $N_f(x) = x(1 - \frac{1}{2}x^2)$ is C^∞ as well. Note that $\lim_{n\to\infty} N_f^n(x) = 0$ if and only if $|x| < 2$. In this case, however, $O_{N_f}(x)$ is not Benford. This follows from the extended version of Theorem 5.8 mentioned in the remark on p.70 but can also be seen directly. Indeed, let $T(x) = \frac{x}{1+|x|}$ and note that $N_f'(x) > 0$, $T'(x) > 0$ and $|T(x)| \leq |N_f(x)|$ holds whenever $|x| \leq \frac{1}{2}$. From this it follows that

$$|N_f^n(x)| \geq |T^n(x)| = \frac{|x|}{1 + n|x|} \quad \text{for all } n \in \mathbb{N}\,,$$

and consequently $(\log |N_f^n(x)|)$ is not u.d. mod 1 by Proposition 4.8(iv), i.e., $O_{N_f}(x)$ is not Benford. On the other hand, if $|x_0| > 2$ then $\lim_{n\to\infty} |N_f^n(x_0)| = +\infty$, and Theorem 5.12, applied to

$$\widetilde{T}(x) := N_f(x^{-1})^{-1} = -\frac{2x^3}{1-2x^2}$$

near $x = 0$, shows that $O_{N_f}(x_0)$ is Benford for almost every, but not every x_0 in this case.

Theorem 5.14 has important practical implications for estimating roots of a function via NM using floating-point arithmetic. One type of error in scientific computations is overflow (or underflow), which occurs when the running computations exceed the largest (or smallest, in absolute value) floating-point number allowed by the computer. Feldstein and Turner [FT, p.241] show that under "the assumption of the logarithmic distribution of numbers [i.e. BL] floating-point addition and subtraction can result in overflow and underflow with alarming frequency ..." Together with Theorem 5.14, this suggests that special attention should be given to overflow and underflow errors in any computer algorithm used to estimate roots by means of NM.

Another important type of error in scientific computing arises due to round-off. In estimating a root from its Newton approximations, for example, a rule for stopping the algorithm must be specified, such as "stop when $n = 10^6$" or "stop when the differences between successive approximations are less than 10^{-6}". Every stopping rule will result in some round-off error, and Theorem 5.14 shows that this difference is generally Benford. In fact, justified by heuristics and by the extensive empirical evidence of BL in other numerical procedures, analysis of roundoff errors has often been carried out under the *hypothesis* of a statistical logarithmic distribution of significant digits or significands [BB]. Therefore, as Knuth points out, a naive assumption of uniformly distributed significant digits in the calculations tends to underestimate the average relative roundoff error in cases where the actual statistical distribution is skewed toward smaller leading significant digits, as is the case for BL. To obtain a rough idea of the magnitude of this underestimate when the true statistical distribution is BL, let X denote the absolute round-off error at the time of stopping the algorithm, and let Y denote the fraction part of the approximation at the time of stopping. Then the relative error is X/Y, and assuming that X and Y are independent random variables, the average (i.e., expected) relative error is simply $\mathbb{E}X \cdot \mathbb{E}(1/Y)$. As shown in [BH1], the assumption that Y is uniform while its true distribution is BL leads to an average *under*estimation of the relative error by more than one third.

The relevance of BL for scientific computing does not end here. For example, Hamming gives "a number of applications to hardware, software, and general computing which show that this distribution is not merely an amusing curiosity" [Ha, p.1609], and Schatte analyzes the speed of multiplication and division in digital computers when the statistical distribution of floating-point numbers is logarithmic and proves that, for design of computers, "[t]he base $b = 8$ is optimal with respect to [minimizing expected] storage use" [Scha1, p.453].

Extension I: Time-dependent systems

So far, the sequences considered in this chapter have been generated by the iteration of a single map T or, in dynamical systems terminology, by an *autonomous* dynamical system. Autonomous systems constitute a classical and well-studied field. Beyond this field there has been, in the recent past, an increased interest in systems that are *nonautonomous*, i.e. explicitly time-dependent in one way or the other. This development is motivated and driven by important practical applications as well as pure mathematical questions. In this context, it is interesting to study how the results discussed previously extend to systems with the map T explicitly depending on n. In full generality, this is a very wide topic with many open problems, both conceptual and computational. Only a small number of pertinent results (without proofs) and examples will be mentioned here, and the interested reader is referred e.g. to [Ber4] for a fuller account and references as well as to [KM, LS] for an intriguing specific problem.

Throughout, let (T_n) be a sequence of maps that map $\mathbb{R}$ or parts thereof into itself, and for every $n \in \mathbb{N}$ denote by T^n the n-fold composition $T^n := T_n \circ \ldots \circ T_1$; also let T^0 be the identity map on $\mathbb{R}$. Given x_0, it makes sense to consider the sequence $O_T(x_0) := \big(T^{n-1}(x_0)\big)_{n\in\mathbb{N}} = \big(x_0, T_1(x_0), T_2\big(T_1(x_0)\big), \ldots\big)$. As in the autonomous case (which corresponds to T_n being independent of n) the sequence $O_T(x_0)$ is referred to as the (nonautonomous) *orbit* of x_0.

The following is a nonautonomous variant of Theorem 5.8. A proof (of a substantially more general version) can be found in [BBH]. It relies heavily on a nonautonomous version of the Shadowing Lemma.

Theorem 5.19 ([BBH]). *Let $T_j : \mathbb{R} \to \mathbb{R}$ be C^2-maps with $T_j(0) = 0$ and $T_j'(0) \neq 0$ for all $j \in \mathbb{N}$, and set $\alpha_j := T_j'(0)$. Assume that $\sup_j \max_{|x|\le 1} |T_j''(x)|$ and $\sum_{n=1}^{\infty} \prod_{j=1}^{n} |\alpha_j|$ are both finite. If $\lim_{j\to\infty} \log|\alpha_j|$ exists and is irrational, then $O_T(x_0)$ is Benford for all $x_0 \neq 0$ sufficiently close to 0.*

Example 5.20. **(i)** Let $R_j(x) = (2 + j^{-1})x$ for $j = 1, 2, \ldots$. It is easy to see that all assumptions of Theorem 5.19 are met for

$$T_j(x) = R_j(x^{-1})^{-1} = \frac{j}{2j+1}x$$

with $\lim_{j\to\infty} \log|\alpha_j| = -\log 2$. Hence $O_R(x_0)$ is Benford for all $x_0 \neq 0$.

(ii) Let $T_j(x) = F_{j+1}/F_j x$ for all $j \in \mathbb{N}$, where F_j denotes the j-th Fibonacci number. Since $\lim_{j\to\infty} \log(F_{j+1}/F_j) = \log\frac{1+\sqrt{5}}{2}$ is irrational, and by taking reciprocals as in (i), Theorem 5.19 shows that $O_T(x_0)$ is Benford for all $x_0 \neq 0$. In particular, $O_T(F_1) = (F_n)$ is Benford, as was already seen in Example 4.12. Note that the same argument would *not* work to show that $(n!)$ is Benford.

(iii) Consider the family of linear maps $T_j(x) = 10^{-1+\sqrt{j+1}-\sqrt{j}}x$ for $j = 1, 2, \ldots$. Here $\prod_{j=1}^{n} \alpha_j = 10^{-n+\sqrt{n+1}-1}$, so $\sum_{n=1}^{+\infty} \prod_{j=1}^{n} |\alpha_j| < +\infty$. However, Theorem 5.19 does not apply since $\lim_{j\to\infty} \log|\alpha_j| = -1$ is rational. Nevertheless, as $(\sqrt{n})$ is u.d. mod 1 by [KN, Ex.3.9] and

$$\log|T^n(x)| = -n + \sqrt{n+1} - 1 + \log|x|\,,$$

the sequence $O_T(x_0)$ is Benford for every $x_0 \neq 0$.

In situations where most of the maps T_j are power-like or even more strongly expanding, the following generalization of Lemma 5.10 may be useful. (In its fully developed form, the result also extends Theorem 5.12, see [BBH, Thm.5.5] and [Ber3, Thm.3.7].) Again the reader is referred to [Ber4] for a proof.

Theorem 5.21 ([Ber4])**.** *Assume the maps $T_j : \mathbb{R}^+ \to \mathbb{R}^+$ satisfy, for some $\xi > 0$ and all $j \in \mathbb{N}$, the following conditions:*

(i) *$x \mapsto \ln T_j(e^x)$ is convex on $[\xi, +\infty)$;*

(ii) *$xT_j'(x)/T_j(x) \geq \beta_j > 0$ for all $x \geq \xi$.*

If $\liminf_{j\to\infty} \beta_j > 1$ then $O_T(x_0)$ is Benford for almost every sufficiently large x_0, but there are also uncountably many exceptional points.

Example 5.22. (i) To see that Theorem 5.21 does indeed generalize Lemma 5.10, let $T_j(x) = \alpha x^\beta$ for all $j \in \mathbb{N}$. Then $x \mapsto \ln T_j(e^x) = \beta x + \ln\alpha$ clearly is convex, and $xT_j'(x)/T_j(x) = \beta > 1$ for all $x > 0$.

(ii) As mentioned already in (ii) of the remark on p.74, Theorem 5.21 also shows that $O_T(x_0)$ with $T(x) = e^x$ is Benford for almost every, but not every $x_0 \in \mathbb{R}$, as $x \mapsto \ln T(e^x) = e^x$ is convex, and $xT'(x)/T(x) = x$ as well as $T^3(x) > e$ holds for all $x \in \mathbb{R}$. Similarly, the theorem applies to $T(x) = 1 + x^2$.

(iii) For a truly nonautonomous example consider

$$T_j(x) = \begin{cases} x^2 & \text{if } j \text{ is even}\,, \\ 2^x & \text{if } j \text{ is odd}\,, \end{cases} \qquad \text{or} \quad T_j(x) = (j+1)^x\,.$$

In both cases, $O_T(x_0)$ is Benford for almost every, but not every $x_0 \in \mathbb{R}$.

(iv) Finally, it is important to note that Theorem 5.21 may fail if one of its hypotheses is violated even for a single j. For example

$$T_j(x) = \begin{cases} 10 & \text{if } j = 1\,, \\ x^2 & \text{if } j \geq 2\,, \end{cases}$$

satisfies (i) and (ii) for all $j > 1$, but does not satisfy assumption (ii) for $j = 1$. Clearly, $O_T(x_0)$ is not Benford for any $x_0 \in \mathbb{R}$, since $D_1\big(T^n(x_0)\big) = 1$ for all $n \in \mathbb{N}$.

Using slightly more sophisticated tools, Theorem 5.21 can be extended so as to provide the following corollary for polynomial maps.

Corollary 5.23. *Let the maps T_j be polynomials,*

$$T_j(x) = x^{n_j} + a_{j,n_j-1}x^{n_j-1} + \ldots + a_{j,1}x + a_{j,0}\,,$$

with $n_j \in \mathbb{N}\backslash\{1\}$ and $a_{j,l} \in \mathbb{R}$ for all $j \in \mathbb{N}$, $0 \le l < n_j$. If $\sup_{j\in\mathbb{N}} \max_{l=0}^{n_j-1}|a_{j,l}| < +\infty$ then $O_T(x_0)$ is Benford for almost every $x_0 \in \mathbb{R}\backslash[-\xi,\xi]$ with some $\xi \ge 0$. However, $\mathbb{R}\backslash[-\xi,\xi]$ also contains an uncountable dense set of exceptional points.

Example 5.24. Let $T_j(x) = x^j - 1$ for all $j - 1$. Then even though (T_j) do not satisfy the hypothesis (i) of Theorem 5.21, by Corollary 5.23, the orbit $O_T(x_0) = (x_0, x_0 - 1, x_0^2 - 2x_0, \ldots)$ is Benford for almost all $|x_0| \ge 3$, but that region also contains uncountably many points for which $O_T(x_0)$ is not Benford.

Extension II: Chaotic dynamical systems

The dynamical scenarios studied so far for their conformance with BL have all been very simple indeed: In Theorems 5.8, 5.12 and 5.19 $\lim_{n\to\infty} T^n(x) = 0$ holds automatically for all relevant initial values x, whereas $\lim_{n\to\infty} T^n(x) = +\infty$ in Theorem 5.21. While this dynamical simplicity does not necessarily force the behavior of $\big(S\big(T^n(x)\big)\big)$ to be equally simple (recall e.g. Example 5.13(iii)), it makes one wonder what might be observed under more general circumstances. The present subsection presents two simple examples in this regard. Among other things, they illustrate that, as a rule, Benford sequences may be rare in more general dynamical systems.

Example 5.25. Consider the tent-map $T : \mathbb{R} \to \mathbb{R}$ given by $T(x) = 1-|2x-1|$. Using Theorem 5.8, it is not hard to see that $O_T(x_0)$ is Benford whenever x_0 lies outside $[0,1]$. Clearly, $O_T(0) = (0,0,0,\ldots)$ and $O_T(1) = (1,0,0,\ldots)$ are not Benford. As far as BL is concerned, therefore, it remains to analyze $O_T(x_0)$ for $0 < x_0 < 1$. Define two maps $\tau_L, \tau_R : [0,1] \to [0,1]$ as

$$\tau_L(x) = \frac{x}{2}\,, \quad \tau_R(x) = 1 - \frac{x}{2}\,.$$

Then $T \circ \tau_L(x) = T \circ \tau_R(x) = x$ for all $x \in [0,1]$, and τ_L, τ_R can be used for a symbolic description of the dynamics of T. To this end, recall that the set Σ of all sequences consisting of the two symbols L and R, that is $\Sigma = \{L,R\}^{\mathbb{N}}$, is a compact metric space when endowed with the metric

$$d(\omega,\widetilde{\omega}) := \begin{cases} 2^{-\min\{n\,:\,\omega_n \ne \widetilde{\omega}_n\}} & \text{if } \omega \ne \widetilde{\omega}\,, \\ 0 & \text{if } \omega = \widetilde{\omega}\,. \end{cases}$$

Moreover, the (left) shift map σ on Σ, given by $\sigma(\omega) = (\omega_{n+1})$ is a continuous map. With these ingredients, define a map $h : \Sigma \to [0,1]$ as

$$h(\omega) := \lim\nolimits_{n\to\infty} \tau_{\omega_1} \circ \tau_{\omega_2} \circ \ldots \circ \tau_{\omega_n}\left(\tfrac{1}{2}\right)\,.$$

It is easy to see that h is well defined, continuous and onto, and $h \circ \sigma(\omega) = T \circ h(\omega)$ for all $\omega \in \Sigma$. In particular, therefore, $T^{n-1} \circ h(\omega) \in I_{\omega_n}$ holds for all

$\omega \in \Sigma$ and $n \in \mathbb{N}$, where $I_L = \tau_L([0,1]) = [0, \frac{1}{2}]$ and $I_R = \tau_R([0,1]) = [\frac{1}{2}, 1]$. Thus it is reasonable to think of ω as the "symbolic itinerary" of $h(\omega)$ under the iteration of T. (Note that h is not one-to-one, however $\#h^{-1}(\{x\}) = 1$ unless x is a dyadic rational, i.e. unless $2^l x$ is an integer for some $l \in \mathbb{N}_0$.) By means of this symbolic coding, some dynamical properties of T are very easy to understand. For example, the set of x_0 for which $O_T(x_0)$ is periodic is dense in $[0,1]$. To see this simply observe that $h(\omega)$ is periodic (under T) whenever $\omega \in \Sigma$ is periodic (under σ), and periodic sequences are dense in Σ. On the other hand, T is topologically transitive. Informally, this means that there is no non-trivial way of breaking the dynamics of T on $[0,1]$ into non-interacting pieces. In the present example, this property (defined and studied thoroughly e.g. in [KH]) simply means that $O_T(x_0)$ is dense for at least one, but in fact *many* $x_0 \in [0,1]$. Overall, therefore, the map $T : [0,1] \to [0,1]$ is *chaotic* in the sense of [Ber1, Def.2.21]. In particular, it exhibits the hallmark property of chaos, namely *sensitive dependence on initial conditions*. The latter means that, for every $0 < x < 1$ and every $\varepsilon > 0$, a point $\overline{x}$ can be found such that

$$|x - \overline{x}| < \varepsilon \quad \text{yet} \quad \limsup_{n\to\infty} |T^n(x) - T^n(\overline{x})| \geq \frac{1}{2}\,.$$

This follows e.g. from [Ber1, Thm.2.18] but can also be seen directly by noticing that T^n is piecewise linear with slope 2^n.

While the above analysis clearly reveals the complexity of the dynamics of T on $[0,1]$, the reader may still wonder how all this is related to BL. Is $O_T(x_0)$ Benford for many, or even most $x_0 \in [0,1]$? The chaotic nature of T suggests a negative answer. For a more definitive understanding, note that, for every $0 < a < 1$,

$$T_*\lambda_{0,1}\big([0,a]\big) = \lambda_{0,1}\Big(\big[0, \tau_L(a)\big] \cup \big[\tau_R(a), 1\big]\Big) = a = \lambda_{0,1}\big([0,a]\big)\,,$$

showing that $T_*\lambda_{0,1} = \lambda_{0,1}$, i.e. T preserves $\lambda_{0,1}$. In fact, T is known to even be *ergodic* with respect to $\lambda_{0,1}$. As a consequence of the Birkhoff Ergodic Theorem, $O_T(x_0)$ is distributed according to $\lambda_{0,1}$ for Lebesgue almost every $x_0 \in [0,1]$. By Example 3.10(i), for every such x_0 the sequence $\big(S(T^n(x_0))\big)$ is uniformly distributed on $[1,10)$. Thus for a.e. $x_0 \in [0,1]$, the orbit $O_T(x_0)$ is *not* Benford.

It remains to investigate whether $O_T(x_0)$ is Benford for *any* $x_0 \in [0,1]$ at all. To this end first note that while $O_T(x_0)$ is guaranteed to be uniformly distributed for a.e. $x_0 \in [0,1]$, there are plenty of exceptions. In fact, given any sequence $\omega \in \Sigma$ whose asymptotic relative frequencies

$$\lim\nolimits_{N\to\infty} \frac{\#\{1 \leq n \leq N : \omega_n = L\}}{N} \quad \text{and} \quad \lim\nolimits_{N\to\infty} \frac{\#\{1 \leq n \leq N : \omega_n = R\}}{N}$$

do not both equal $\frac{1}{2}$, or perhaps do not even exist at all, the orbit of $h(\omega)$ is not uniformly distributed. For instance, if

$$\omega = \big(\underbrace{L, L, \ldots, L}_{N \text{ times}}, R, R, R, \ldots\big)$$

for some $N \in \mathbb{N}_0$, then $h(\omega) = \frac{2^{1-N}}{3}$, and $T^n\big(h(\omega)\big) = \frac{2}{3}$ for all $n \geq N$. In view of this abundance of exceptional points, one may hope to identify *some* $x_0 \in [0,1]$ for which $O_T(x_0)$ is Benford. Using the symbolic encoding of the dynamics, this can indeed be done as follows: Observe that $T(x) = 2x$ whenever $x \leq \frac{1}{2}$, i.e. whenever $x \in I_L$, in which case

$$\log S\big(T(x)\big) = \langle \log 2 + \log S(x) \rangle \,.$$

Thus if $T^n(x_0) \in I_L$ held for all n, then $O_T(x_0)$ would be Benford. This is impossible since $T^n(x_0) \in I_L$ for all n implies that $x_0 = 0$, and x_0 is a fixed point for T. However, since being Benford is an asymptotic property of $O_T(x_0)$, it is enough for $T^n(x_0) \in I_L$ to hold for *most* n and *along arbitrarily long sequences*. Concretely, let

$$\omega^* = \big(\underbrace{L, L, \ldots, L}_{N_1 \text{ times}}, R, \underbrace{L, L, \ldots, L}_{N_2 \text{ times}}, R, \underbrace{L, L, \ldots, L}_{N_3 \text{ times}}, R, L, \ldots\big)\,, \tag{5.2}$$

where (N_n) is any sequence in $\mathbb{N}$ with $N_n \to \infty$, and set $x^* = h(\omega^*)$. According to (5.2), the orbit $O_T(x^*)$ stays in I_L for the first N_1 steps, then makes a one-step excursion to I_R, then remains in I_L for N_2 steps, etc. It follows from [Ber4, Lem.2.7(i)], but can also be verified directly, that $O_T(x^*)$ is Benford. For a concrete example, choose e.g. $N_n \equiv 2n$, then

$$\omega^* = \big(L, L, R, L, L, L, L, R, L, L, L, L, L, L, R, L, \ldots\big)$$

as well as

$$x^* = h(\omega^*) = \sum\nolimits_{n=1}^{\infty} 2^{1+2n-n^2}(-1)^{n+1} = 0.2422\ldots\,,$$

and $O_T(x^*)$ is Benford. Notice finally that (5.2) provides uncountably many *different* points x^*, and hence the set

$$\{x_0 \in [0,1] : O_T(x_0) \text{ is Benford}\,\}$$

is uncountable; as initial segments of ω^* do not matter, this set is also dense in $[0,1]$. To put this fact into perspective, note that with the points x^* constructed above, $O_T(x^*)$ is actually also Benford base b whenever b is not a power of 2, i.e. whenever $b \notin \{2^n : n \in \mathbb{N}\}$. On the other hand, $O_T(x_0)$ is not Benford base 2, 4, 8 etc. for *any* $x_0 \in \mathbb{R}$, see [Ber4, Ex.2.11].

Example 5.26. The family of quadratic polynomials $Q_\rho : x \mapsto \rho x(1-x)$, with $\rho \in \mathbb{R}$, often referred to as the *logistic family*, plays a pivotal role in dynamical systems theory, see e.g. [Ber1, KH]. Arguably the most prominent member of this family is the map Q_4 which has many features in common with the tent map T from the previous example. Unlike the latter, however, Q_4 is smooth, and it is this smoothness which makes the dynamics of Q_4, or generally the logistic family, a much richer yet also more subtle topic.

To understand the dynamics of Q_4 with regards to BL, note first that near $x = 0$,

$$Q_4(x^{-1})^{-1} = -\frac{x^2}{4(1-x)} = -\frac{x^2}{4} + \mathcal{O}(x^3)\,.$$

Hence Theorem 5.12 applies, showing that $O_{Q_4}(x_0)$ is Benford for almost every, but not every $x_0 \in \mathbb{R}\backslash[0,1]$. As in Example 5.25, it remains to study the dynamics within the interval $[0,1]$. A similar symbolic coding can be applied to demonstrate that on this interval Q_4 is, in any respect, as chaotic as the tent map T. This is somewhat hard to do directly, but it becomes very simple upon introducing the homeomorphism $H : [0,1] \to [0,1]$ with $H(x) = \sin^2(\frac{1}{2}\pi x)$ and noting that, for all $x \in [0,1]$,

$$Q_4 \circ H(x) = \sin^2(\pi x) = H \circ T(x)\,. \tag{5.3}$$

Thus Q_4 and T differ only by a change of coordinates, and all topological properties of T (such as e.g. the existence of a dense set of periodic orbits, and topological transitivity) carry over to Q_4. Together with $T_*\lambda_{0,1} = \lambda_{0,1}$ it follows from (5.3) that

$$Q_{4*}(H_*\lambda_{0,1}) = (Q_4 \circ H)_*\lambda_{0,1} = (H \circ T)_*\lambda_{0,1} = H_*(T_*\lambda_{0,1}) = H_*\lambda_{0,1}\,,$$

hence Q_4 preserves the probability measure $H_*\lambda_{0,1}$, and is in fact ergodic with respect to it. Note that

$$\frac{\mathrm{d}}{\mathrm{d}x} H_*\lambda_{0,1}\big([0,x]\big) = \frac{\mathrm{d}}{\mathrm{d}x}\Big(\lambda_{0,1}\big([0, \frac{2}{\pi}\arcsin\sqrt{x}]\big)\Big) = \frac{1}{\pi\sqrt{x(1-x)}}\,, \quad 0 < x < 1\,,$$

showing that $H_*\lambda_{0,1}$ is simply the arcsin- or $\mathtt{Beta}(\frac{1}{2}, \frac{1}{2})$-distribution, and therefore $H_*\lambda_{0,1}(B) = 0$ if and only if $\lambda_{0,1}(B) = 0$. Again, the Birkhoff Ergodic Theorem implies that $O_{Q_4}(x_0)$ is, for almost every $x_0 \in [0,1]$, distributed according to $H_*\lambda_{0,1}$, and consequently not Benford, see Example 3.10(iii). As in Example 5.25, one may wonder whether $O_{Q_4}(x_0)$ is Benford for any $x_0 \in [0,1]$ at all. Essentially the same argument shows that the answer is, again, positive. With ω^* as in (5.2), the orbit of $H \circ h(\omega^*)$ spends most of its time arbitrarily close to the (unstable) fixed point at $x = 0$, and

$$\log S\big(Q_4(x)\big) = \langle \log 4 + \log S(x) + \log(1-x)\rangle \approx \langle \log 4 + \log S(x)\rangle$$

whenever $x > 0$ is very small. A careful analysis in the spirit of Lemma 4.3(i) then shows that $O_{Q_4}\big(H \circ h(\omega^*)\big)$ is indeed Benford. As in the previous example, it follows that

$$\{x_0 \in [0,1] : O_{Q_4}(x_0) \text{ is Benford}\,\}$$

is uncountable and dense in $[0,1]$.

5.2. Multi-dimensional discrete-time processes

The purpose of this section is to extend the basic results of the previous section to multi-dimensional systems, notably to linear, as well as some non-linear recurrence relations. Recall from Example 4.12 that the Fibonacci sequence (F_n) is Benford. Hence the linear recurrence relation $x_{n+1} = x_n + x_{n-1}$ generates a Benford sequence when started from $x_0 = x_1 = 1$. As will be seen shortly, many, but not all linear recurrence relations generate Benford sequences.

Under a BL perspective, an obvious difficulty when dealing with multi-dimensional systems is the potential for more or less cyclic behavior, either of the orbits themselves or of their significands.

Example 5.27. **(i)** Let the sequence (x_n) be defined recursively as

$$x_{n+1} = x_n - x_{n-1}\,, \quad n = 1, 2, \ldots\,, \tag{5.4}$$

with given $x_0, x_1 \in \mathbb{R}$. By using the matrix $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 1 \end{smallmatrix}\right]$ associated with (5.4), it is straightforward to derive an explicit representation for (x_n),

$$x_n = x_0 \cos\left(\tfrac{1}{3}\pi n\right) + \frac{2x_1 - x_0}{\sqrt{3}} \sin\left(\tfrac{1}{3}\pi n\right)\,, \quad n = 0, 1, \ldots\,.$$

From this it is clear that $x_{n+6} = x_n$ for all n, i.e., (x_n) is 6-periodic. This oscillatory behavior of (x_n) corresponds to the fact that the roots of the characteristic equation $\lambda^2 = \lambda - 1$ associated with (5.4) are $\lambda = e^{\pm \imath\pi/3}$ and hence lie on the unit circle. For no choice of x_0, x_1, therefore, is (x_n) Benford.

(ii) Consider the linear 3-step recursion

$$x_{n+1} = 2x_n + 10x_{n-1} - 20x_{n-2}\,, \quad n = 2, 3, \ldots\,.$$

Again it is easy to confirm that, for any $x_0, x_1, x_2 \in \mathbb{R}$, the value of x_n is given explicitly by

$$x_n = c_1 2^n + c_2 10^{n/2} + c_3(-1)^n 10^{n/2}\,,$$

where

$$c_1 = \frac{10x_0 - x_2}{6}\,, \quad c_{2,3} = \frac{x_2 - 4x_0}{12} \pm \frac{x_2 + 3x_1 - 10x_0}{6\sqrt{10}}\,.$$

Clearly, $\lim_{n\to\infty} |x_n| = +\infty$ unless $x_0 = x_1 = x_2 = 0$, so unlike in (i) the sequence (x_n) is not bounded or oscillatory. However, if $|c_2| \neq |c_3|$ then

$$\log |x_n| = \frac{n}{2} + \log \left| c_1 10^{-n(\frac{1}{2} - \log 2)} + c_2 + (-1)^n c_3 \right| \approx \frac{n}{2} + \log |c_2 + (-1)^n c_3|\,,$$

showing that $\big(S(x_n)\big)$ is asymptotically 2-periodic and hence (x_n) is not Benford. Similarly, if $|c_2| = |c_3| \neq 0$ then $\big(S(x_n)\big)$ is convergent along even (if $c_2 = c_3$) or odd (if $c_2 = -c_3$) indices n, and again (x_n) is not Benford. Only if $c_2 = c_3 = 0$ yet $c_1 \neq 0$, or equivalently if $\frac{1}{4}x_2 = \frac{1}{2}x_1 = x_0 \neq 0$ is (x_n) Benford. Obviously, the oscillatory behavior of $\big(S(x_n)\big)$ in this example is due to the characteristic equation $\lambda^3 = 2\lambda^2 + 10\lambda - 20$ having two roots with the same modulus but opposite signs, namely $\lambda = -\sqrt{10}$ and $\lambda = \sqrt{10}$.

(iii) Let $\gamma = \cos(\pi \log 2) \approx 0.5852$ and consider the sequence (x_n) defined recursively as

$$x_{n+1} = 4\gamma x_n - 4x_{n-1}\,, \quad n = 1, 2\ldots\,, \tag{5.5}$$

with given x_0, $x_1 \in \mathbb{R}$. As before, an explicit formula for x_n is easily derived as

$$x_n = 2^n x_0 \cos(\pi n \log 2) + 2^{n-1} \frac{x_1 - 2\gamma x_0}{\sqrt{1-\gamma^2}} \sin(\pi n \log 2)\,.$$

Although somewhat oscillatory, the sequence (x_n) is clearly unbounded. As will be shown now, however, it is not Benford. While the argument is essentially the same for any $(x_0, x_1) \neq (0,0)$, for convenience let $x_0 = 0$ and $x_1 = 2\sin(\pi \log 2) \approx 1.622$, so that

$$\log|x_n| = \log 2^n |\sin(\pi n \log 2)| = n \log 2 + \log|\sin(\pi n \log 2)|\,, \quad n = 1, 2, \ldots\,.$$

With the (measurable) map $T : [0,1) \to [0,1)$ defined as

$$T(s) = \langle s + \log|\sin(\pi s)|\rangle\,, \quad 0 \leq s < 1\,,$$

therefore simply $\langle \log|x_n|\rangle = T(\langle n \log 2\rangle)$. Recall that $(n \log 2)$ is u.d. mod 1, and hence $(\langle \log|x_n|\rangle)$ is distributed according to the probability measure $T_*\lambda_{0,1}$. Consequently, (x_n) is Benford if and only if $T_*\lambda_{0,1}$ equals $\lambda_{0,1}$. The latter, however, is not the case. While this is clear intuitively, an easy way to see this formally is to observe that T is piecewise smooth and has a unique local maximum at some $0 < s_0 < 1$. (Concretely, $s_0 = 1 - \frac{1}{\pi}\arctan\frac{\pi}{\ln 10} \approx 0.7013$ and $T(s_0) \approx 0.6080$.) Thus if $T_*\lambda_{0,1} = \lambda_{0,1}$, then for all sufficiently small $\varepsilon > 0$,

$$\begin{aligned}\frac{T(s_0) - T(s_0-\varepsilon)}{\varepsilon} &= \frac{\lambda_{0,1}\big(\big[T(s_0-\varepsilon), T(s_0)\big)\big)}{\varepsilon} = \frac{T_*\lambda_{0,1}\big(\big[T(s_0-\varepsilon), T(s_0)\big)\big)}{\varepsilon} \\ &\geq \frac{\lambda_{0,1}\big([s_0-\varepsilon, s_0)\big)}{\varepsilon} = 1\,,\end{aligned}$$

which is impossible since $T'(s_0) = 0$. Hence (x_n) is not Benford. The reason for this can be seen in the fact that, while $\log|\lambda| = \log 2$ is irrational for the characteristic roots $\lambda = 2e^{\pm \imath\pi \log 2}$ associated with (5.5), there obviously is a rational dependence between the two real numbers $\log|\lambda|$ and $\frac{1}{2\pi}\arg\lambda$, namely $\log|\lambda| - 2(\frac{1}{2\pi}\arg\lambda) = 0$.

The above recurrence relations are linear and have constant coefficients. Hence they can be rewritten and analyzed using matrix-vector notation. For instance, in Example 5.27(i)

$$\begin{bmatrix} x_n \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_{n-1} \\ x_n \end{bmatrix}\,,$$

so that, with $A = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 2}$, the sequence (x_n) is simply given by

$$x_n = \begin{bmatrix} 1 & 0 \end{bmatrix} A^n \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}, \quad n = 0, 1, \ldots .$$

It is natural, therefore, to study the Benford property of more general sequences $(x^\top A^n y)$ for any $A \in \mathbb{R}^{d\times d}$ and $x, y \in \mathbb{R}^d$. Linear recurrence relations like the ones in Example 5.27 are then merely special cases. As suggested by that example, in order to guarantee the Benford property for $(x^\top A^n y)$, conditions have to be imposed on A so as to rule out cyclic behavior of orbits or their significands. To prepare for these conditions, denote the real part, imaginary part, complex conjugate, and modulus (absolute value) of $z \in \mathbb{C}$ by $\Re z$, $\Im z$, $\overline{z}$, and $|z|$, respectively. For $z \neq 0$, define $\arg z$ as the unique number in $[-\pi, \pi)$ that satisfies $z = |z|e^{\imath \arg z}$; for notational convenience, let $\arg 0 := 0$. Recall that real or complex numbers $z_1, z_2, \ldots, z_n$ are *rationally independent* (or $\mathbb{Q}$-*independent*) if $\sum_{j=1}^n q_j z_j = 0$ with $q_1, q_2, \ldots, q_n \in \mathbb{Q}$ implies that $q_j = 0$ for all $j = 1, 2, \ldots, n$. A set $Z \subset \mathbb{C}$ is rationally independent if every of its finite subsets is, and *rationally dependent* otherwise.

Let $Z \subset \mathbb{C}$ be any set such that all elements of Z have the same modulus ζ, i.e., Z is contained in the periphery of a circle with radius ζ centered at the origin of the complex plain. Call the set Z *resonant* if either $\#(Z \cap \mathbb{R}) = 2$ or the numbers $1, \log \zeta$ and the elements of $\frac{1}{2\pi} \arg Z$ are rationally dependent, where $\frac{1}{2\pi} \arg Z = \left\{ \frac{1}{2\pi} \arg z : z \in Z \right\} \backslash \{-\frac{1}{2}, 0\}$.

Given $A \in \mathbb{R}^{d\times d}$, recall that the *spectrum* $\sigma(A) \subset \mathbb{C}$ of A is simply the set of all eigenvalues of A. Denote by $\sigma(A)^+$ the "upper half" of the spectrum, i.e., let $\sigma(A)^+ = \{\lambda \in \sigma(A) : \Im \lambda \geq 0\}$. Usage of $\sigma(A)^+$ refers to the fact that non-real eigenvalues of real matrices always occur in conjugate pairs, and hence $\sigma(A)^+$ only contains one of the conjugates.

With the above preparations, what will shortly turn out to be an appropriate condition on A reads as follows.

Definition 5.28. A matrix $A \in \mathbb{R}^{d\times d}$ is *Benford regular* (*base* 10) if $\sigma(A)^+$ contains no resonant set.

Note that in the simplest case, i.e. for $d = 1$, the matrix $A = [a]$ is Benford regular if and only if $\log |a|$ is irrational. Hence Benford regularity may be considered a generalization of this irrationality property. Also note that A is regular (invertible) whenever it is Benford regular.

Example 5.29. None of the matrices associated with the recurrence relations in Example 5.27 is Benford regular. Indeed, in (i), $A = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}$, hence $\sigma(A)^+ = \{e^{\imath\pi/3}\}$, and clearly $\log |e^{\imath\pi/3}| = 0$ is rational. Similarly, in (ii), $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -10 & 10 & 2 \end{bmatrix}$, and $\sigma(A)^+ = \{-\sqrt{10}, 2, \sqrt{10}\}$ contains the resonant set $\{-\sqrt{10}, \sqrt{10}\}$. Finally, for (iii), $A = \begin{bmatrix} 0 & 1 \\ -4 & 4\gamma \end{bmatrix}$, and $\sigma(A)^+ = \{2e^{\imath\pi \log 2}\}$ is resonant.

Example 5.30. Let $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 2}$, with characteristic polynomial $p_A(\lambda) = \lambda^2 - 2\lambda + 2$, and hence $\sigma(A)^+ = \{\sqrt{2}e^{\imath\pi/4}\}$. As $1, \log\sqrt{2}$ and $\frac{1}{2\pi}\cdot\frac{\pi}{4} = \frac{1}{8}$ are rationally dependent, the matrix A is not Benford regular.

Example 5.31. Consider $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 2}$. The characteristic polynomial of A is $p_A(\lambda) = \lambda^2 - \lambda - 1$, and so, with $\varphi = \frac{1}{2}(1+\sqrt{5})$, the eigenvalues of A are φ and $-\varphi^{-1}$. Since p_A is irreducible and has two roots of different absolute value, it follows that $\log\varphi$ is irrational (in fact, even transcendental). Thus A is Benford regular.

With the one-dimensional result (Lemma 5.3), as well as Example 5.27 and Definition 5.28 in mind, it seems realistic to hope that iterating (i.e. taking powers of) any matrix $A \in \mathbb{R}^{d\times d}$ produces many Benford sequences, provided that A is Benford regular. This is indeed the case. To concisely formulate the pertinent result, call a sequence (z_n) of complex numbers *terminating* if $z_n = 0$ for all sufficiently large n.

Theorem 5.32 ([Ber2])**.** *Assume that $A \in \mathbb{R}^{d\times d}$ is Benford regular. Then, for every $x, y \in \mathbb{R}^d$, the sequence $(x^\top A^n y)$ is either Benford or terminating. Also, $(\|A^n x\|)$ is Benford for every $x \neq 0$.*

The proof of Theorem 5.32 will make use of the following variant of [Ber2, Lem.2.9].

Proposition 5.33. *Assume that the real numbers $1, \rho_0, \rho_1, \dots, \rho_m$ are $\mathbb{Q}$-independent. Let (z_n) be a convergent sequence in $\mathbb{C}$, and at least one of the numbers $c_1, c_2, \dots, c_m \in \mathbb{C}$ non-zero. Then (x_n) given by*

$$x_n = n\rho_0 + \log\left|\Re\left(c_1 e^{2\pi\imath n\rho_1} + \dots + c_m e^{2\pi\imath n\rho_m} + z_n\right)\right|$$

is u.d. mod 1.

Proof of Theorem 5.32. Given $A \in \mathbb{R}^{d\times d}$, let $\sigma(A)^+ = \{\lambda_1, \lambda_2, \dots, \lambda_s\}$, where $s \leq d$ and, without loss of generality, $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_s|$. Fix $x, y \in \mathbb{R}^d$ and recall that there exist (possibly non-real) polynomials $p_1, p_2, \dots, p_s$ of degrees at most $d-1$ such that

$$x^\top A^n y = \Re\big(p_1(n)\lambda_1^n + \dots + p_s(n)\lambda_s^n\big), \quad n = 0, 1, \dots . \tag{5.6}$$

(This follows e.g. from the Jordan Normal Form Theorem.) If $(x^\top A^n y)$ is not terminating, then it can be assumed that $p_1 \neq 0$. (Otherwise relabel the p_j and λ_j, and reduce s accordingly.) Now distinguish two cases.

<u>Case 1:</u> $|\lambda_1| > |\lambda_2|$

In this case, λ_1 is a *dominant* eigenvalue. Denote by k the degree of p_1 and let $c := \lim_{n\to\infty} n^{-k} p_1(n)$. Note that c is a non-zero number that is real whenever

λ_1 is real. From

$$\begin{aligned}|x^\top A^n y| &= |\lambda_1|^n n^k \left| \Re\left(n^{-k} p_1(n) \left(\frac{\lambda_1}{|\lambda_1|}\right)^n + n^{-k} p_2(n) \left(\frac{\lambda_2}{|\lambda_1|}\right)^n + \ldots \right.\right. \\ &\qquad\qquad\qquad\qquad \left.\left. + n^{-k} p_s(n) \left(\frac{\lambda_s}{|\lambda_1|}\right)^n \right)\right| \\ &= |\lambda_1|^n n^k |\Re(c e^{\imath n \arg \lambda_1} + z_n)| \,,\end{aligned}$$

with $z_n = (n^{-k}p_1(n) - c)\left(\frac{\lambda_1}{|\lambda_1|}\right)^n + \sum_{j=2}^s n^{-k} p_j(n) \left(\frac{\lambda_j}{|\lambda_1|}\right)^n \to 0$, it follows that

$$\log|x^\top A^n y| = n \log|\lambda_1| + k \log n + \log|\Re(c e^{2\pi \imath n \frac{\arg \lambda_1}{2\pi}} + z_n)| \,.$$

In view of Proposition 4.8(iii), no generality is lost by assuming that $k = 0$. If λ_1 is real then, by Lemma 4.3(i) and the irrationality of $\log|\lambda_1|$, the sequence $(\log|x^\top A^n y|)$ is u.d. mod 1. If λ_1 is not real, then apply Proposition 5.33 with $m = 1$, $\rho_0 = \log|\lambda_1|$, and $\rho_1 = \frac{1}{2\pi}\arg\lambda_1$. In either case, $(x^\top A^n y)$ is Benford.

<u>Case 2:</u> $|\lambda_1| = \ldots = |\lambda_l| > |\lambda_{l+1}|$ for some $l \le s$.

Here several *different* eigenvalues of the same magnitude occur. Let k be the maximal degree of $p_1, p_2, \ldots p_l$ and $c_j := \lim_{n\to\infty} n^{-k} p_j(n)$ for $j = 1, 2, \ldots, l$. Note that if $x^\top A^n y \neq 0$ infinitely often then at least one of the numbers $c_1, c_2, \ldots, c_l$ is non-zero. As before,

$$\begin{aligned}|x^\top A^n y| &= |\lambda_1|^n n^k \left| \Re\left(n^{-k} p_1(n) \left(\frac{\lambda_1}{|\lambda_1|}\right)^n + n^{-k} p_2(n) \left(\frac{\lambda_2}{|\lambda_1|}\right)^n + \ldots \right.\right. \\ &\qquad\qquad\qquad\qquad \left.\left. + n^{-k} p_s(n) \left(\frac{\lambda_s}{|\lambda_1|}\right)^n \right)\right| \\ &= |\lambda_1|^n n^k |\Re(c_1 e^{\imath n \arg \lambda_1} + \ldots + c_l e^{\imath n \arg\lambda_l} + z_n)| \,,\end{aligned}$$

where

$$z_n = \sum\nolimits_{j=1}^{l} (n^{-k} p_j(n) - c_j) \left(\frac{\lambda_j}{|\lambda_1|}\right)^n + \sum\nolimits_{j=l+1}^{s} n^{-k} p_j(n) \left(\frac{\lambda_j}{|\lambda_1|}\right)^n \to 0 \,.$$

Propositions 4.8(iii) and 5.33 with $m = l$ and $\rho_0 = \log|\lambda_1|$, $\rho_1 = \frac{1}{2\pi}\arg\lambda_1, \ldots,$ $\rho_l = \frac{1}{2\pi}\arg\lambda_l$ imply that

$$\log|x^\top A^n y| = n\log|\lambda_1| + k\log n + \log|\Re(c_1 e^{\imath n \arg\lambda_1} + \ldots + c_l e^{\imath n \arg\lambda_l} + z_n)|$$

is u.d. mod 1, hence $(x^\top A^n y)$ is Benford.

The assertion concerning $(\|A^n x\|)$ is proved in a completely analogous manner. □

Example 5.34. According to Example 5.31, the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ is Benford regular. By Theorem 5.32, every solution of the difference equation $x_{n+1} =$

$x_n + x_{n-1}$ is Benford, except for the trivial solution $x_n \equiv 0$ resulting from $x_0 = x_1 = 0$. In particular, therefore, the sequences of Fibonacci and Lucas numbers, $(F_n) = (1, 1, 2, 3, 5, \ldots)$ and $(L_n) = (-1, 2, 1, 3, 4, \ldots)$, generated respectively from the initial values $\begin{bmatrix} x_0 & x_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \end{bmatrix}$ and $\begin{bmatrix} x_0 & x_1 \end{bmatrix} = \begin{bmatrix} -1 & 2 \end{bmatrix}$, are Benford. For the former sequence, this has already been seen in Example 4.12. Note that (F_n^2), for instance, is Benford as well by Corollary 4.7(i), see Fig 24.

Example 5.35. Recall from Example 5.30 that $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ is not Benford regular. Hence Theorem 5.32 does not apply, and the sequence $(x^\top A^n y)$ may, for some $x, y \in \mathbb{R}^2$, be neither Benford nor terminating. Indeed, pick for example $x = y = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$ and note that

$$x^\top A^n y = \begin{bmatrix} 1 & 0 \end{bmatrix} 2^{n/2} \begin{bmatrix} \cos(\frac{1}{4}\pi n) & -\sin(\frac{1}{4}\pi n) \\ \sin(\frac{1}{4}\pi n) & \cos(\frac{1}{4}\pi n) \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2^{n/2} \cos\left(\tfrac{1}{4}\pi n\right) .$$

Hence $(x^\top A^n y)$ is clearly not Benford as $x^\top A^n y = 0$ whenever $n = 2 + 4l$ for some $l \in \mathbb{N}_0$. It will be seen later (in Theorem 5.37) that in the case of a 2×2-matrix A, the Benford regularity of A is actually necessary for *every* sequence of the form $(x^\top A^n y)$ to be either Benford or terminating. Note, however, that this does of course not rule out the possibility that some sequences derived from iterating A may be Benford nevertheless. For a concrete example, fix any $x \neq 0$ and, for each $n \in \mathbb{N}$, denote by E_n the area of the triangle with vertices at $A^n x$, $A^{n-1}x$, and the origin. Then

$$E_n = \frac{1}{2} \left| \det(A^n x, A^{n-1} x) \right| = 2^{n-2} \|x\|^2 , \quad n = 1, 2, \ldots ,$$

so (E_n) is Benford, see Fig 24.

Remark. According to Theorem 5.32, Benford regularity of a matrix A is a simple condition guaranteeing the widespread generation of Benford sequences of the form $(x^\top A^n y)$. Most $d \times d$-matrices are Benford regular, under a topological as well as a measure-theoretic perspective. To put this more formally, let

$$B_d := \{ A \in \mathbb{R}^{d \times d} : A \text{ is Benford regular} \} .$$

While the complement of B_d is dense in $\mathbb{R}^{d \times d}$, it is a topologically small set: $\mathbb{R}^{d \times d} \backslash B_d$ is *of first category*, i.e. a countable union of nowhere dense sets. A (topologically) typical ("generic") $d \times d$-matrix therefore belongs to B_d, i.e. is Benford regular. Similarly, if A is an $\mathbb{R}^{d \times d}$-valued random variable, that is, a random matrix, whose distribution is a.c. with respect to the d^2-dimensional Lebesgue measure on $\mathbb{R}^{d \times d}$, then $\mathbb{P}(A \in B_d) = 1$, i.e., A is Benford regular with probability one. Similar statements hold for instance within the family of *stochastic matrices*, see [BHKR]. ♣

While Benford regularity of A is a property sufficient for all sequences $(x^\top A^n y)$ to be either Benford or terminating, the following example shows that this property is not in general necessary.

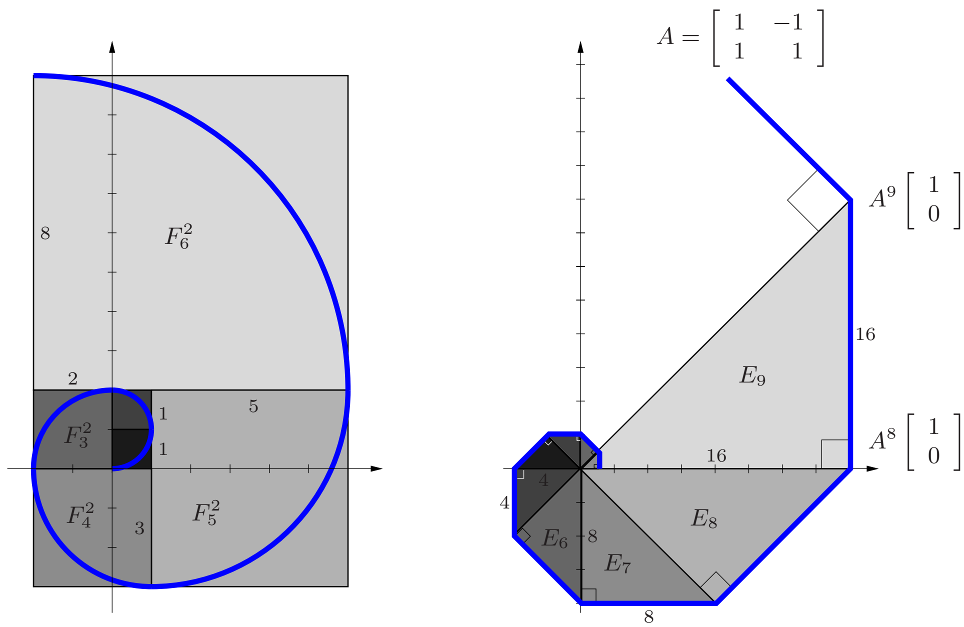

FIG 24. *Two Benford sequences derived from linear* 2*-dimensional systems, see Examples 5.34 and 5.35. Note that the matrix A associated with (E_n) is not Benford regular.*

Example 5.36. Consider the 4×4-matrix

$$A = 10^{\sqrt{2}} \begin{bmatrix} \cos(2\pi\sqrt{3}) & -\sin(2\pi\sqrt{3}) & 0 & 0 \\ \sin(2\pi\sqrt{3}) & \cos(2\pi\sqrt{3}) & 0 & 0 \\ 0 & 0 & \cos(4\pi\sqrt{3}) & -\sin(4\pi\sqrt{3}) \\ 0 & 0 & \sin(4\pi\sqrt{3}) & \cos(4\pi\sqrt{3}) \end{bmatrix},$$

for which $\sigma(A)^+ = \{10^{\sqrt{2}} e^{-2\pi\imath\sqrt{3}}, 10^{\sqrt{2}} e^{4\pi\imath\sqrt{3}}\} =: \{\lambda_1, \lambda_2\}$. Since $2\arg\lambda_1 + \arg\lambda_2 = 0$, the matrix A is not Benford regular. It will now be shown that nevertheless for any $x, y \in \mathbb{R}^4$ the sequence $(x^\top A^n y)$ is either Benford or terminating. Indeed, with $x^\top = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix}$ and $y = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \end{bmatrix}^\top$, a straightforward calculation confirms that

$$x^\top A^n y = 10^{n\sqrt{2}} \Re\Big((x_1 + \imath x_2)(y_1 - \imath y_2) e^{-2\pi\imath n\sqrt{3}} + (x_3 + \imath x_4)(y_3 - \imath y_4) e^{-4\pi\imath n\sqrt{3}} \Big) .$$

Unless $(x_1^2 + x_2^2)(y_1^2 + y_2^2) + (x_3^2 + x_4^2)(y_3^2 + y_4^2) = 0$, therefore, $(x^\top A^n y)$ is not terminating, and

$$\log |x^\top A^n y| = n\sqrt{2} + f(n\sqrt{3}),$$

with the function $f : [0, 1) \to \mathbb{R}$ given by

$$f(s) = \log \left| \Re\big((x_1 + \imath x_2)(y_1 - \imath y_2) e^{-2\pi\imath s} + (x_3 + \imath x_4)(y_3 - \imath y_4) e^{-4\pi\imath s} \big) \right| .$$

Note that f has at most finitely many discontinuities. Moreover, $1, \sqrt{2}, \sqrt{3}$ are $\mathbb{Q}$-independent, and hence [Ber2, Cor.2.6] implies that $(x^\top A^n y)$ is Benford.

The dimension $d = 4$ in Example 5.36 is smallest possible. Indeed, as the following result shows, Benford regularity is (not only sufficient but also) necessary in Theorem 5.32 whenever $d < 4$.

Theorem 5.37. *Assume $d < 4$, and let $A \in \mathbb{R}^{d\times d}$ be invertible. Then the following statements are equivalent:*

(i) *A is Benford regular.*

(ii) *For every $x, y \in \mathbb{R}^d$ the sequence $(x^\top A^n y)$ is either Benford or terminating.*

Proof. As demonstrated by Theorem 5.32, assumption (i) implies (ii) even without any restrictions on d.

Conversely, assume that (ii) holds. Notice that whenever A has a real eigenvalue $\lambda \neq 0$, with a corresponding eigenvector $e_\lambda \neq 0$, then choosing $x = y = e_\lambda$ results in $x^\top A^n y = \lambda^n \|e_\lambda\|^2$. Hence $\log|\lambda|$ must be irrational. For $d = 1$, this shows that A is Benford regular.

Next let $d = 2$. In this case, two different eigenvalues of the same modulus can occur either in the form $\pm\lambda$ with $\lambda > 0$, i.e. as non-zero eigenvalues of opposite sign, or in the form $\lambda = |\lambda| e^{\pm 2\pi\imath\rho}$ with $|\lambda| > 0$ and $0 < \rho < \frac{1}{2}$, i.e. as a pair of conjugate non-real eigenvalues. In the former case, let e_- and e_+ be normalized eigenvectors corresponding to $-\lambda$ and λ, respectively. Note that $1 + e_+^\top e_- > 0$, by the Cauchy–Schwarz inequality. Then

$$(e_+ + e_-)^\top A^n (e_+ + e_-) = \begin{cases} 2\lambda^n (1 + e_+^\top e_-) & \text{if } n \text{ is even}, \\ 0 & \text{if } n \text{ is odd}, \end{cases}$$

showing that $(x^\top A^n y)$ is not Benford for $x = y = e_+ + e_-$. Assuming (ii), therefore, implies that A does not have real eigenvalues of opposite sign. On the other hand, if $\sigma(A)^+ = \{|\lambda| e^{2\pi\imath\rho}\}$ then there exists a regular matrix $P \in \mathbb{R}^{2\times 2}$ such that

$$P^{-1} A P = |\lambda| \begin{bmatrix} \cos(2\pi\rho) & -\sin(2\pi\rho) \\ \sin(2\pi\rho) & \cos(2\pi\rho) \end{bmatrix}.$$

Specifically choosing $x^\top = \begin{bmatrix} 0 & 1 \end{bmatrix} P^{-1}$ and $y = P \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$ yields

$$x^\top A^n y = |\lambda|^n \sin(2\pi n\rho), \quad n = 0, 1, \ldots . \tag{5.7}$$

If $\log|\lambda|$ is rational, say $\log|\lambda| = \frac{p}{q}$, then the sequence

$$\langle q \log|x^\top A^n y| \rangle = \langle q \log|\sin(2\pi n\rho)| \rangle$$

is either periodic (if ρ is rational) or else distributed according to $T_* \lambda_{0,1}$, with $T : [0, 1) \to [0, 1)$ given by $T(s) = \langle q \log|\sin(2\pi s)| \rangle$. As in Example 5.27(iii), it can be shown that $T_* \lambda_{0,1} \neq \lambda_{0,1}$. Thus, as before, rationality of $\log|\lambda|$ is ruled out by assumption (ii). If ρ is rational then $x^\top A^n y = 0$ holds for infinitely many but not all n, and hence $(x^\top A^n y)$ is neither Benford nor terminating. Again, this possibility is ruled out by assumption (ii). To conclude the case $d = 2$, assume that $\log|\lambda|$ and ρ are both irrational, yet 1, $\log|\lambda|$ and ρ are rationally dependent, i.e., there exist integers k_1, k_2, k_3 with $k_2k_3 \neq 0$ such that

$$k_1 + k_2 \log|\lambda| + k_3\rho = 0\,.$$

Without loss of generality, assume $k_3 > 0$. For every $j \in \{1, 2, \ldots, k_3\}$ and $n \in \mathbb{N}_0$ therefore

$$\log|x^\top A^{nk_3+j}y| = (nk_3+j)\log|\lambda| + \log\left|\sin\left(2\pi j\frac{k_1}{k_3} + 2\pi\left(j\frac{k_2}{k_3} + nk_2\right)\log|\lambda|\right)\right|\,,$$

so $(\log|x^\top A^{nk_3+j}y|)$ is distributed according to $T_{j*}\lambda_{0,1}$, with $T_j : [0,1) \to [0,1)$ given by

$$T_j(s) = \left\langle k_3 s + j\log|\lambda| + \log\left|\sin\left(2\pi j\frac{k_1}{k_3} + 2\pi\frac{k_2}{k_3}j\log|\lambda| + 2\pi k_2 s\right)\right|\right\rangle\,,$$

and $(\langle\log|x^\top A^n y|\rangle)$ is distributed according to $\frac{1}{k_3}\sum_{j=1}^{k_3} T_{j*}\lambda_{0,1}$. Again it can be shown that the latter probability measure on $\big([0,1), \mathcal{B}[0,1)\big)$ does *not* equal $\lambda_{0,1}$. Overall, therefore, for $d = 2$ and $\sigma(A)^+ = \{|\lambda|e^{2\pi\imath\rho}\}$, assumption (ii) implies that 1, $\log|\lambda|$, and $\frac{1}{2\pi}\arg\lambda$ are rationally independent. In other words, A is Benford regular.

Finally, consider the case $d = 3$. The only eigenvalue configuration not covered by the preceding arguments is that of three different eigenvalues with the same modulus, i.e. with $|\lambda| > 0$ and $0 < \rho < \frac{1}{2}$ either $\sigma(A)^+ = \{|\lambda|, |\lambda|e^{2\pi\imath\rho}\}$ or $\sigma(A)^+ = \{-|\lambda|, |\lambda|e^{2\pi\imath\rho}\}$. In both cases, there exists a regular matrix $P \in \mathbb{R}^{3\times 3}$ such that

$$P^{-1}AP = |\lambda|\begin{bmatrix} \pm1 & 0 & 0 \\ 0 & \cos(2\pi\rho) & -\sin(2\pi\rho) \\ 0 & \sin(2\pi\rho) & \cos(2\pi\rho) \end{bmatrix}\,,$$

and choosing $x^\top = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} P^{-1}$ and $y = P\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^\top$ again yields (5.7). As before, assumption (i) implies that 1, $\log|\lambda|$, and ρ are rationally independent. ☐

Finally, it is worth noting that even if A is not Benford regular, *many* or even *most* sequences of the form $(x^\top A^n y)$ may nevertheless be Benford.

Example 5.38. Recall from Example 5.30 that $A = \left[\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right]$ is not Benford regular because $\sigma(A)^+ = \{\sqrt{2}e^{\imath\pi/4}\}$ is resonant. However, a short calculation with $x^\top = \begin{bmatrix} x_1 & x_2 \end{bmatrix}$, $y = \begin{bmatrix} y_1 & y_2 \end{bmatrix}^\top$ confirms that

$$x^\top A^n y = 2^{n/2}\|x\|\,\|y\|\cos\left(\tfrac{1}{4}\pi n + \psi\right)\,,\quad n = 0, 1, \ldots\,;$$

here $\psi \in [-\pi, \pi)$ is the angle of a counter-clockwise rotation moving $x/\|x\|$ into $y/\|y\|$. (Note that ψ is unique unless $\|x\|\,\|y\| = 0$ in which case $x^\top A^n y \equiv 0$

anyway.) By virtue of Proposition 4.8(ii), if $\psi \notin [-\pi, \pi) \cap \frac{1}{4}\pi\mathbb{Z}$ then $(\log |x^\top A^n y|)$ is u.d. mod 1. Thus, if ψ is not an integer multiple of $\frac{1}{4}\pi$, or equivalently if

$$\Big((x_1^2 - x_2^2)y_1y_2 - x_1x_2(y_1^2 - y_2^2)\Big)\Big((x_1^2 - x_2^2)(y_1^2 - y_2^2) + 4x_1x_2y_1y_2\Big) \neq 0\,,$$

then $(x^\top A^n y)$ is Benford.

The present section closes with two examples of non-linear systems. The sole purpose of these examples is to hint at possible extensions of the results presented earlier; for more details the interested reader is referred to the references provided.

Example 5.39. Consider the non-linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T : \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} f(x_1) \\ f(x_2) \end{bmatrix},$$

with the bounded continuous function

$$f(t) = \frac{3}{2}|t+2| - 3|t+1| + 3|t-1| - \frac{3}{2}|t-2| = \begin{cases} 0 & \text{if } |t| \ge 2\,, \\ 3t+6 & \text{if } -2 < t < -1\,, \\ -3t & \text{if } -1 \le t < 1\,, \\ 3t-6 & \text{if } 1 \le t < 2\,. \end{cases}$$

Sufficiently far away from the x_1- and x_2-axes, i.e. for $\min\{|x_1|, |x_2|\}$ sufficiently large, the dynamics of T is governed by the matrix $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, and since the latter is Benford regular, one may reasonably expect that $\big(x^\top T^n(y)\big)$ should be Benford. It can be shown that this is indeed the case. More precisely, by means of a multi-dimensional shadowing argument, the following statement can be proved, see [Ber2, Thm.4.1]: Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be of the form $T(x) = Ax + f(x)$ with $A \in \mathbb{R}^{d\times d}$ and a bounded continuous $f : \mathbb{R}^d \to \mathbb{R}^d$. If A is Benford regular and has no eigenvalues inside the unit disc, that is, $|\lambda| > 1$ holds for every eigenvalue λ of A, then the sequence $\big(x^\top T^n(y)\big)$ is Benford whenever it is unbounded. Notice that the provision concerning boundedness is already needed in the present simple example: For instance, if $|\xi| \le \frac{3}{2}$ and $x^\top = \begin{bmatrix} \xi & 0 \end{bmatrix}$ then $\big(T^n(x)\big)$ is eventually 2-periodic and hence $\big(x^\top T^n(x)\big)$ is not Benford.

Example 5.40. Consider the non-linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined as

$$T : \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \begin{bmatrix} 3x_1^3x_2^2 + 4x_1 \\ 5x_1^2x_2^4 - 2x_2^2 + 1 \end{bmatrix}.$$

Unlike in the previous example, the map T is now genuinely non-linear and cannot be considered a perturbation of a linear map. Rather, T may be thought of as a 2-dimensional analogue of the polynomial map $x \mapsto 1 + x^2$. Clearly, if $|x_1|$ or $|x_2|$ is small, then the behavior of $\big(T^n(x)\big)$ may be complicated. For instance, on the x_2-axis, i.e. for $x_1 = 0$, the map T reduces to $x_2 \mapsto 1 - 2x_2^2$ which, up to a change of coordinates, is nothing else but the chaotic map Q_4

studied in Example 5.26. If, however, $|x_1|$ and $|x_2|$ are sufficiently large then a two-dimensional version of Theorem 5.12 asserts that, for (Lebesgue) almost every x, each component of $O_T(x)$ is Benford, see [BS, Thm.16]; at the same time, there is also an abundance of exceptional points [BS, Cor.17].

5.3. Differential equations

By presenting a few results on, and examples of differential equations, i.e. deterministic continuous-time processes, this short section aims at convincing the reader that the emergence of BL is not at all restricted to discrete-time dynamics. Rather, solutions of ordinary or partial differential equations often turn out to be Benford as well. Recall that a (Borel measurable) function $f : [0, +\infty) \to \mathbb{R}$ is Benford if and only if $\log |f|$ is u.d. mod 1.

Consider the *initial value problem* (*IVP*)

$$\dot{x} = F(x), \quad x(0) = x_0, \tag{5.8}$$

where $F : \mathbb{R} \to \mathbb{R}$ is continuously differentiable with $F(0) = 0$, and $x_0 \in \mathbb{R}$. In the simplest case, $F(x) \equiv \alpha x$ with some $\alpha \in \mathbb{R}$. In this case, the unique solution of (5.8) is $x(t) = x_0 e^{\alpha t}$. Unless $\alpha x_0 = 0$, therefore, every solution of (5.8) is Benford, by Example 4.5(i). As in the discrete-time setting, this feature persists for arbitrary C^2-functions F with $F'(0) < 0$. The direct analog of Theorem 5.8 is

Theorem 5.41 ([BBH])**.** *Let $F : \mathbb{R} \to \mathbb{R}$ be C^2 with $F(0) = 0$. Assume that $F'(0) < 0$. Then, for every $x_0 \neq 0$ sufficiently close to 0, the unique solution of* (5.8) *is Benford.*

Proof. Pick $\delta > 0$ so small that $xF(x) < 0$ for all $0 < |x| \leq \delta$. As F is C^2, the IVP (5.8) has a unique local solution whenever $|x_0| \leq \delta$, see [Wa]. Since the interval $[-\delta, \delta]$ is forward invariant, this solution exists for all $t \geq 0$. Fix any x_0 with $0 < |x_0| \leq \delta$ and denote the unique solution of (5.8) as $x = x(t)$. Clearly, $\lim_{t\to+\infty} x(t) = 0$. With $y : [0, +\infty) \to \mathbb{R}$ defined as $y = x^{-1}$ therefore $y(0) = x_0^{-1} =: y_0$ and $\lim_{t\to+\infty} |y(t)| = +\infty$. Let $\alpha := -F'(0) > 0$ and note that there exists a continuous function $g : \mathbb{R} \to \mathbb{R}$ such that $F(x) = -\alpha x + x^2 g(x)$. From

$$\dot{y} = -\frac{\dot{x}}{x^2} = \alpha y - g(y^{-1}),$$

it follows via the variation of constants formula that, for all $t \geq 0$,

$$y(t) = e^{\alpha t} y_0 - \int_0^t e^{\alpha(t-\tau)} g\big(y(\tau)^{-1}\big)\, \mathrm{d}\tau .$$

As $\alpha > 0$ and g is continuous, the number

$$\overline{y_0} := y_0 - \int_0^{+\infty} e^{-\alpha\tau} g\big(y(\tau)^{-1}\big)\, \mathrm{d}\tau$$

is well defined. (Note that $\overline{y_0}$ is simply the continuous-time analogue of the auxiliary point $\overline{x}$ in Lemma 5.5.) Moreover, for all $t > 0$,

$$\begin{aligned}\left|y(t) - e^{\alpha t}\overline{y_0}\right| &= \left|\int_t^{+\infty} e^{\alpha(t-\tau)} g\bigl(y(\tau)^{-1}\bigr)\,\mathrm{d}\tau\right| \\ &\le \int_0^{+\infty} e^{-\alpha\tau}\left|g\bigl(y(t+\tau)^{-1}\bigr)\right|\,\mathrm{d}\tau \le \frac{\|g\|_\infty}{\alpha},\end{aligned}$$

where $\|g\|_\infty = \max_{|x|\le\delta}|g(x)|$, and Lemma 5.7(ii) shows that y is Benford if and only if $t \mapsto e^{\alpha t}\overline{y_0}$ is. An application of Corollary 4.7(ii), together with Example 4.5(i) therefore completes the proof. □

Example 5.42. (i) The function $F(x) = -x + x^4 e^{-x^2}$ satisfies the assumptions of Theorem 5.41. Thus except for the trivial $x = 0$, every solution of $\dot{x} = -x + x^4 e^{-x^2}$ is Benford.

(ii) The function $F(x) = -x^3 + x^4 e^{-x^2}$ is also smooth with $xF(x) < 0$ for all $x \ne 0$. Hence for every $x_0 \in \mathbb{R}$, the IVP (5.8) has a unique solution with $\lim_{t\to+\infty} x(t) = 0$. However, $F'(0) = 0$, and as will be shown now, this prevents x from being Benford. To see this, fix $x_0 \ne 0$ and integrate

$$-\frac{\dot{x}}{x^3} = 1 - xe^{-x^2}$$

from 0 to t to obtain the implicit representation

$$x^2(t) = \frac{x_0^2}{1 + 2tx_0^2 - 2x_0^2\displaystyle\int_0^t x(\tau)e^{-x(\tau)^2}\mathrm{d}\tau}. \tag{5.9}$$

Note that $\lim_{t\to+\infty} x(t) = 0$ implies $\lim_{t\to+\infty}\frac{1}{t}\int_0^t x(\tau)e^{-x(\tau)^2}\mathrm{d}\tau = 0$. Hence it follows from (5.9) that $\lim_{t\to+\infty} 2tx(t)^2 = 1$. Consequently, $t \mapsto |\log x|/\log t$ is bounded as $t \to +\infty$, and (the continuous-time version of) Proposition 4.8(iv) shows that x is not Benford.

Informally, the fact that $F'(0) = 0$ causes the solutions of $\dot{x} = F(x)$ to approach the equilibrium $x = 0$ too slowly in order to be Benford. It is not hard to see that this is true in general: If F is C^2 and $xF(x) < 0$ for all $x \ne 0$ in a neighborhood of 0, and hence $F(0) = 0$, yet $F'(0) = 0$ then, for all $|x_0|$ sufficiently small the solution of (5.8) is not Benford.

(iii) As the previous example showed, for solutions of (5.8) with $F(0) = F'(0) = 0$ to be Benford for all $x_0 \ne 0$ sufficiently close to 0, it is necessary that F not be C^2. (In fact, F must not even be $C^{1+\varepsilon}$ for any $\varepsilon > 0$, see [BBH, Thm.6.7].) For an example of this type, consider

$$F(x) = -\frac{x}{\sqrt{1 + (\log x)^4}}, \quad x \ne 0.$$

With $F(0) := 0$, the function F is C^1 with $F'(0) = 0$, and every non-trivial solution of $\dot{x} = F(x)$ is Benford. To see this, fix $x_0 \neq 0$ and let $y = -\log x$. Then

$$\dot{y} = \frac{1}{\ln 10 \sqrt{1 + y^4}} \, ,$$

from which it is straightforward to deduce that $|y(t) - \sqrt[3]{3t/\ln 10}| \to 0$ as $t \to +\infty$, which in turn shows that y is u.d. mod 1, i.e., x is Benford.

(iv) Theorem 5.41 applies to the smooth function $F(x) = -x + x\log(1+x^2)$. In this case, $\dot{x} = F(x)$ has three equilibria, namely $x = 0$ and $x = \pm 3$, and consequently the solution of (5.8) is Benford whenever $0 < |x_0| < 3$.

To analyze the behavior of solutions outside of $[-3, 3]$, fix $x_0 > 3$ and let $y := \log x - \frac{1}{2}$. Then

$$\dot{y} = \frac{2y}{\ln 10} + \frac{\log(1 + 10^{-1-2y})}{\ln 10} \, ,$$

and hence, for all $t \geq 0$,

$$y(t) = e^{2t/\ln 10} y_0 + \int_0^t e^{2(t-\tau)/\ln 10} \frac{\log(1 + 10^{-1-2y(\tau)})}{\ln 10} \, \mathrm{d}\tau \, .$$

With the usual auxiliary point $\overline{y_0} := y_0 + \int_0^{+\infty} e^{-2\tau)/\ln 10} \frac{\log(1+10^{-1-2y(\tau)})}{\ln 10} \, \mathrm{d}\tau$,

$$\begin{aligned} \left| y(t) - e^{2t/\ln 10}\overline{y_0} \right| &= \left| \int_t^{+\infty} e^{2(t-\tau)/\ln 10} \frac{\log(1 + 10^{-1-2y(\tau)})}{\ln 10} \, \mathrm{d}\tau \right| \\ &\leq \int_0^{+\infty} e^{-2\tau/\ln 10} \frac{\log(1 + 10^{-1-2y(t+\tau)})}{\ln 10} \, \mathrm{d}\tau \\ &\leq \log \sqrt{1 + 10^{-1-2y(t)}} \ \to \ 0 \quad \text{as } t \to +\infty \, . \end{aligned}$$

By the same reasoning as in Example 4.5(iii), the function y is u.d. mod 1. Thus by Theorem 4.2, x is Benford for $|x_0| > 3$ as well. Note that $|x|$ goes to $+\infty$ faster than exponentially in this case, i.e. $\lim_{t\to+\infty} |x(t)e^{-\alpha t}| = +\infty$ for every $\alpha > 0$.

Also, note that the case $|x_0| > 3$ could be rephrased in the setting of Theorem 5.41 as well. Indeed, with $z := x^{-1}$ one finds

$$\dot{z} = z\log(z^2) + z - z\log(1 + z^2) =: \widetilde{F}(z) \, .$$

With $\widetilde{F}(0) := 0$, the function $\widetilde{F} : \mathbb{R} \to \mathbb{R}$ is continuous but not C^1, as $\lim_{z\to 0} \widetilde{F}(z)/z = -\infty$. Thus Theorem 5.41 does not apply. The lack of smoothness of $\widetilde{F}$ corresponds to the fact that solutions of the IVP $\dot{z} = \widetilde{F}(z)$, $z(0) = z_0$, though still unique and globally defined, approach $z = 0$ faster than exponentially whenever $|z_0| < \frac{1}{3}$. For a result in the spirit of Theorem 5.41 that does apply to $\dot{z} = \widetilde{F}(z)$ directly, see [BBH, Thm.6.9].

Just as their discrete-time counterparts, linear differential equations in higher dimensions are also a rich source of Benford behavior. Consider for instance the IVP

$$\ddot{x} - x = 0\,, \quad x(0) = x_0,\, \dot{x}(0) = v_0\,, \tag{5.10}$$

with given numbers $x_0, v_0 \in \mathbb{R}$. The unique solution of (5.10) is

$$x(t) = \frac{x_0 + v_0}{2} e^t + \frac{x_0 - v_0}{2} e^{-t}$$

which clearly is Benford unless $x_0 = v_0 = 0$. Using matrix-vector notation, (5.10) can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix}, \quad \begin{bmatrix} x \\ \dot{x} \end{bmatrix}\Big|_{t=0} = \begin{bmatrix} x_0 \\ v_0 \end{bmatrix}.$$

Much more generally, therefore, consider the linear d-dimensional ordinary differential equation

$$\dot{x} = Ax\,, \tag{5.11}$$

where A is a real $d \times d$-matrix. Recall that every solution of (5.11) is given by $x : t \mapsto e^{tA} x_0$ for some $x_0 \in \mathbb{R}^d$, in fact $x_0 = x(0)$, with the matrix exponential e^{tA} defined as

$$e^{tA} = \sum\nolimits_{l=0}^{\infty} \frac{t^l}{l!} A^l\,.$$

To ensure that every component of x, or that, more generally, for any $x, y \in \mathbb{R}^d$ the function $t \mapsto x^\top e^{tA} y$ is either Benford or trivial, a condition reminiscent of Benford regularity has to be imposed on A.

Definition 5.43. A matrix $A \in \mathbb{R}^{d\times d}$ is *exponentially Benford regular* (*base* 10) if $e^{\tau A}$ is Benford regular for some $\tau > 0$.

Note that in the simplest case, i.e. for $d = 1$, the matrix $A = [a]$ is exponentially Benford regular if and only if $a \neq 0$. Moreover, every exponentially Benford regular matrix is regular. It is easily checked that a matrix A fails to be exponentially Benford regular exactly if there exist $\lambda_1, \lambda_2, \ldots, \lambda_l$ in $\sigma(A)^+$ with $\Re\lambda_1 = \Re\lambda_2 = \ldots = \Re\lambda_l$ such that $\Re\lambda_1/\ln 10$ and the elements of $\{\frac{1}{2\pi}\Im\lambda_1, \frac{1}{2\pi}\Im\lambda_2, \ldots, \frac{1}{2\pi}\Im\lambda_l\}\backslash\frac{1}{2}\mathbb{Z}$ are rationally dependent. Also, it is not hard to see that if A is exponentially Benford regular then the set

$$\{t \in \mathbb{R} : e^{tA} \text{ is not Benford regular}\}$$

actually is at most countable, i.e. finite (possibly empty) or countable. With this, the continuous-time analog of Theorem 5.32 is

Theorem 5.44. *Assume that $A \in \mathbb{R}^{d\times d}$ is exponentially Benford regular. Then, for every $x, y \in \mathbb{R}^d$, the function $t \mapsto x^\top e^{tA} y$ is either Benford or identically equal zero. Also, $t \mapsto \|e^{tA}x\|$ is Benford for every $x \neq 0$.*

Proof. Given $x, y \in \mathbb{R}^d$, define $f : \mathbb{R} \to \mathbb{R}$ according to $f(t) := x^\top e^{tA} y$. As observed above, for almost every $h > 0$ the matrix e^{hA} is Benford regular and, by Theorem 5.32, the sequence $(x^\top (e^{hA})^n y) = \big(f(nh)\big)$ is either terminating or Benford. In the former case, $f = 0$ due to the fact that f is real-analytic. In the latter case, $\big(\log|f(nh)|\big)$ is u.d. mod 1 for almost all $h > 0$, and [KN, Thm.9.6] shows that $\log|f|$ is u.d. mod 1, i.e., f is Benford. The function $t \mapsto \|e^{tA}x\|$ is dealt with similarly. ☐

Example 5.45. **(i)** The matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ associated with (5.10) is exponentially Benford regular, as $\sigma(A)^+ = \{-1, 1\}$, and hence, as seen earlier, the solution of (5.10) is Benford unless $x_0 = v_0 = 0$.

(ii) For $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ recall from Example 5.31 that $\sigma(A)^+ = \{-\varphi^{-1}, \varphi\}$ with $\varphi = \frac{1}{2}(1 + \sqrt{5})$. Hence A is exponentially Benford regular, and every function of the form $t \mapsto x^\top e^{tA} y$ is either Benford or vanishes identically. This is also evident from the explicit formula

$$e^{tA} = \frac{e^{t\varphi}}{2+\varphi} \begin{bmatrix} 1 & \varphi \\ \varphi & 1+\varphi \end{bmatrix} + \frac{e^{-t\varphi^{-1}}}{2+\varphi} \begin{bmatrix} \varphi+1 & -\varphi \\ -\varphi & 1 \end{bmatrix},$$

which shows that the latter is the case if and only if x and y are proportional to $\begin{bmatrix} 1 & \varphi \end{bmatrix}^\top$ and $\begin{bmatrix} -\varphi & 1 \end{bmatrix}^\top$ (or vice versa), i.e. to the two perpendicular eigendirections of A.

(iii) Consider $A = \begin{bmatrix} 1 & -\pi/\ln 10 \\ \pi/\ln 10 & 1 \end{bmatrix}$ with $\sigma(A)^+ = \{1 + \imath\pi/\ln 10\}$. In this case, A fails to be exponentially Benford regular because, with $\lambda = 1 + \imath\pi/\ln 10$,

$$\frac{\Re\lambda}{\ln 10} - 2\frac{\Im\lambda}{2\pi} = 0\,.$$

As a matter of fact, no function $t \mapsto x^\top e^{tA} y$ is Benford. Indeed,

$$e^{tA} = e^t \begin{bmatrix} \cos(\pi t/\ln 10) & -\sin(\pi t/\ln 10) \\ \sin(\pi t/\ln 10) & \cos(\pi t/\ln 10) \end{bmatrix},$$

and picking for instance $x^\top = \begin{bmatrix} 0 & 1 \end{bmatrix}$ and $y = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$ yields

$$\log|x^\top e^{tA} y| = \log\left|e^t \sin\left(\frac{\pi t}{\ln 10}\right)\right| = \frac{t}{\ln 10} + \log\left|\sin\left(\frac{\pi t}{\ln 10}\right)\right| = g\left(\frac{t}{\ln 10}\right),$$

where $g(s) = s + \log|\sin(\pi s)|$. As in Example 5.27(iii), it can be shown that g is *not* u.d. mod 1.

This example suggests that exponential Benford regularity of A may (not only be sufficient but) also be necessary in Theorem 5.44. In analogy to Example 5.36 and Theorem 5.37, one can show that this is indeed true if $d < 4$, but generally false otherwise; details are left to the interested reader.

Finally, it should be mentioned that at present little seems to be known about the Benford property for solutions of *partial* differential equations or more general functional equations such as e.g. delay or integro-differential equations. Quite likely, it will be very hard to decide in any generality whether many, or even most solutions of such systems exhibit the Benford property in one form or another.

Example 5.46. A fundamental example of a partial differential equation is the so-called one-dimensional *heat* (or *diffusion*) *equation*

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}\,, \tag{5.12}$$

a linear second-order equation for $u = u(t, x)$. Physically, (5.12) describes e.g. the diffusion over time of heat in a homogeneous one-dimensional medium. Without further conditions, (5.12) has *many* solutions of which for instance

$$u(t, x) = cx^2 + 2ct\,,$$

with any constant $c \neq 0$, is neither Benford in t ("time") nor in x ("space"), whereas

$$u(t, x) = e^{-c^2 t}\sin(cx)$$

is Benford (or identically zero) in t but not in x, and

$$u(t, x) = e^{c^2 t + cx}$$

is Benford in both t and x. Usually, to specify a unique solution an equation like (5.12) has to be supplemented with initial and/or boundary conditions. A prototypical example of an Initial-boundary Value Problem (IBVP) consists of (5.12) together with

$$\begin{aligned} u(0, x) &= u_0(x) && \text{for all } 0 < x < 1\,, \\ u(t, 0) &= u(t, 1) = 0 && \text{for all } t > 0\,. \end{aligned} \tag{5.13}$$

Physically, the conditions (5.13) may be interpreted as the ends of the medium, at $x = 0$ and $x = 1$, being kept at a reference temperature $u = 0$ while the initial distribution of heat (or temperature etc.) is given by the function $u_0 : [0, 1] \to \mathbb{R}$. It turns out that, under very mild assumptions on u_0, the IBVP consisting of (5.12) and (5.13) has a unique solution which, for any $t > 0$, can be written as

$$u(t, x) = \sum\nolimits_{n=1}^{\infty} u_n e^{-\pi^2 n^2 t}\sin(\pi n x)\,,$$

where $u_n = 2\int_0^1 u_0(s)\sin(\pi n s)\,\mathrm{d}s$. From this it is easy to see that, for every $0 \le x \le 1$, the function $t \mapsto u(t, x)$ either vanishes identically or else is Benford.

Another possible set of initial and boundary data is

$$\begin{aligned} u(0, x) &= u_0(x) && \text{for all } x > 0\,, \\ u(t, 0) &= 0 && \text{for all } t > 0\,, \end{aligned} \tag{5.14}$$

corresponding to a semi-infinite one-dimensional medium kept at zero temperature at its left end $x = 0$, with an initial heat distribution given by the (integrable) function $u_0 : [0, +\infty) \to \mathbb{R}$. Again, (5.12) together with (5.14) has a unique solution, for any $t > 0$ given by

$$u(t, x) = \frac{1}{2\sqrt{\pi t}} \int_0^{+\infty} u_0(y) \left(e^{-(x-y)^2/(4t)} - e^{-(x+y)^2/(4t)} \right) \mathrm{d}y \,.$$

Assuming $\int_0^{+\infty} y|u_0(y)| \,\mathrm{d}y < +\infty$, it is not hard to see that, for every $x \ge 0$,

$$\lim\nolimits_{t\to+\infty} t^{3/2} u(t, x) = \frac{x}{2\sqrt{\pi}} \int_0^{+\infty} y u_0(y) \,\mathrm{d}y \,,$$

and hence, for any fixed $x \ge 0$, the function u is not Benford in time. On the other hand, if for example $u_0(x) = xe^{-x}$ then a short calculation confirms that, for every $t > 0$,

$$\lim\nolimits_{x\to+\infty} \frac{e^x u(t, x)}{x} = e^t \,,$$

showing that u is Benford in space. Similarly, if $u_0(x) = \mathbb{1}_{[0,1)}(x)$ then

$$\lim\nolimits_{x\to+\infty} x e^{(x-1)^2/(4t)} u(t, x) = \sqrt{\frac{t}{\pi}}$$

holds for every $t > 0$, and again u is Benford in space.

6. Benford's Law for random processes

The purpose of this chapter is to show how BL arises naturally in a variety of stochastic settings, including products of independent random variables, mixtures of random samples from different distributions, and iterations of random maps. Perhaps not surprisingly, BL arises in many other important fields of stochastics as well, such as geometric Brownian motion, order statistics, random matrices, Lévy processes, and Bayesian models. The present chapter may also serve as a preparation for the specialized literature on these advanced topics [EL, LSE, MN, Schü1].

6.1. Independent random variables

The analysis of sequences of random variables, notably the special case of (sums or products of) independent, identically distributed (i.i.d.) sequences of random variables, constitutes an important classical topic in probability theory. Within this context, the present section studies general scenarios that lead to BL emerging as an "attracting" distribution. The results nicely complement the observations made in previous chapters.

Recall from Chapter 3 that a (real-valued) random variable X by definition is Benford if $\mathbb{P}(S(X) \le t) = \log t$ for all $t \in [1, 10)$. Also, recall that a sequence

(X_n) of random variables converges *in distribution* to a random variable X, symbolically $X_n \stackrel{\mathcal{D}}{\to} X$, if $\lim_{n\to\infty} \mathbb{P}(X_n \le t) = \mathbb{P}(X \le t)$ holds for every $t \in \mathbb{R}$ for which $\mathbb{P}(X = t) = 0$. By a slight abuse of terminology, say that (X_n) *converges in distribution to BL* if $S(X_n) \stackrel{\mathcal{D}}{\to} S(X)$, where X is a Benford random variable, or equivalently if

$$\lim_{n\to\infty} \mathbb{P}(S(X_n) \le t) = \log t \quad \text{for all } t \in [1, 10)\,.$$

Another important concept is almost sure convergence. Specifically, the sequence (X_n) converges to X *almost surely* (*a.s.*), in symbols $X_n \stackrel{a.s.}{\to} X$, if $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$. It is easy to check that $X_n \stackrel{a.s.}{\to} 1$ implies $X_n \stackrel{\mathcal{D}}{\to} X$. The reverse implication does not hold in general, as is evident from any i.i.d. sequence (X_n) for which X_1 is not constant: In this case, all X_n have the same distribution, so trivially $X_n \stackrel{\mathcal{D}}{\to} X_1$, yet $\mathbb{P}(\,\lim_{n\to\infty} X_n \text{ exists}\,) = 0$.

An especially simple way of generating a sequence of random variables is this: Fix a random variable X, and set $X_n := X^n$ for every $n \in \mathbb{N}$. While the sequence (X_n) thus generated is clearly not i.i.d. unless $X = 0$ a.s. or $X = 1$ a.s., Theorems 4.10 and 4.17 imply

Theorem 6.1. *Assume that the random variable X has a density. Then:*

(i) (X^n) *converges in distribution to BL.*

(ii) *With probability one, (X^n) is Benford.*

Proof. To prove (i), note that the random variable $\log|X|$ has a density as well. Hence, by Theorem 4.17

$$\begin{aligned}\mathbb{P}(S(X_n) \le t) &= \mathbb{P}(\langle \log|X^n|\rangle \le \log t)\\ &= \mathbb{P}(\langle n\log|X|\rangle \le \log t) \;\to\; \log t \quad \text{as } n \to \infty\end{aligned}$$

holds for all $t \in [1, 10)$, i.e. (X_n) converges in distribution to BL.

To see (ii), simply note that $\log|X|$ is irrational with probability one. By Theorem 4.10, therefore, $\mathbb{P}\big((X^n) \text{ is Benford}\big) = 1$. $\square$

Example 6.2. **(i)** Let X be uniformly distributed on $[0, 1)$. For every $n \in \mathbb{N}$,

$$F_{S(X^n)}(t) = \frac{t^{1/n} - 1}{10^{1/n} - 1}\,, \quad 1 \le t < 10\,,$$

and a short calculation, together with the elementary estimate $\frac{e^t - 1 - t}{e^t - 1} < \frac{t}{2}$ for all $t > 0$ shows that

$$\big|F_{S(X^n)}(t) - \log t\big| \le \frac{10^{1/n} - 1 - \frac{\ln 10}{n}}{10^{1/n} - 1} < \frac{\ln 10}{2n} \;\to\; 0 \quad \text{as } n \to \infty\,,$$

and hence (X^n) converges in distribution to BL. Since $\mathbb{P}(\,\log X \text{ is rational}\,) = 0$, the sequence (X^n) is Benford with probability one.

(ii) Assume that $X = 2$ a.s. Thus $P_X = \delta_2$, and X does not have a density. For every n, $S(X^n) = 10^{\langle n \log 2 \rangle}$ with probability one, so (X^n) does not converge in distribution to BL. On the other hand, (X^n) is Benford a.s.

Remarks. (i) In the spirit of Theorem 6.1, several results from Chapter 5 can be extended to a stochastic context. For a prototypical result, consider the map $T : x \mapsto 1 + x^2$ from Example 5.13(iii). If X has a density, then so has $Y := \log |X|$. Recall from the proof of Theorem 5.12 that

$$\log |T^n(X)| - 2^n \overline{Y} \overset{a.s.}{\to} 0 \quad \text{as } n \to \infty \,,$$

with a uniquely defined $\overline{Y} = h(Y)$, and a close inspection of the map h shows that $\overline{Y}$ has a density as well. Hence by Theorems 4.2 and 4.17, $O_T(X) = \big(T^{n-1}(X)\big)$ converges in distribution to BL, whereas Theorem 5.12 implies that $\mathbb{P}(O_T(X) \text{ is Benford}) = 1$.

(ii) For any random variable, it is not hard to see that assertion (ii) in Theorem 6.1 holds whenever (i) does. In the case of an i.i.d. sequence (X_n), the convergence of (X_n) in distribution to BL also implies that (X_n) is Benford for all n, so by independence it is easy to see that (X_n) is Benford with probability one. In general, however, these two properties are independent. For one implication see Example 6.2(ii). For the other implication consider the constant sequence $(X, X, X, \ldots)$ which is evidently not Benford, but if X is a Benford random variable then (X) trivially converges in distribution to BL. ♣

The sequence of random variables considered in Theorem 6.1 is very special in that X^n is the product of n quantities that are identical, and hence dependent *in extremis*. Note that X^n is Benford for all n if and only if X is Benford. This invariance property of BL persists if, unlike the case in Theorem 6.1, products of *independent* factors are considered.

Theorem 6.3. *Let X, Y be two independent random variables with $\mathbb{P}(XY = 0) = 0$. Then:*

(i) *If X is Benford then so is XY.*

(ii) *If $S(X)$ and $S(XY)$ have the same distribution, then either $\log S(Y)$ is rational with probability one, or X is Benford.*

Proof. As in the proof of Theorem 4.13, the argument becomes short and transparent through the usage of Fourier coefficients. Note first that $\log S(XY) = \langle \log S(X) + \log S(Y) \rangle$ and, since the random variables $X_0 := \log S(X)$ and $Y_0 := \log S(Y)$ are independent,

$$\widehat{P_{\log S(XY)}} = \widehat{P_{\langle X_0 + Y_0 \rangle}} = \widehat{P_{X_0}} \cdot \widehat{P_{Y_0}} \,. \tag{6.1}$$

To prove (i), simply recall that X being Benford is equivalent to $P_{X_0} = \lambda_{0,1}$, and hence $\widehat{P_{X_0}}(k) = 0$ for every integer $k \neq 0$. Consequently, $\widehat{P_{\log S(XY)}}(k) = 0$ as well, i.e., XY is Benford.

To see (ii), assume that $S(X)$ and $S(XY)$ have the same distribution. In this case, (6.1) implies that

$$\widehat{P_{X_0}}(k)\big(1-\widehat{P_{Y_0}}(k)\big)=0 \quad \text{for all } k\in\mathbb{Z}\,.$$

If $\widehat{P_{Y_0}}(k)\neq 1$ for all non-zero k, then $\widehat{P_{X_0}}=\widehat{\lambda_{0,1}}$, i.e., X is Benford. Alternatively, if $\widehat{P_{Y_0}}(k_0)=1$ for some $k_0\neq 0$ then, as seen in the proof of Theorem 4.13(iii), $P_{Y_0}(\frac{1}{|k_0|}\mathbb{Z})=1$, hence $|k_0|Y_0=|k_0|\log S(Y)$ is an integer with probability one. □

Example 6.4. Let V, W be independent random variables distributed according to $U(0,1)$. Then $X:=10^V$ and $Y:=W$ are independent and, by Theorem 6.3(i), XY is Benford even though Y is not. If, on the other hand, $X:=10^V$ and $Y:=10^{1-V}$ then X and Y are both Benford, yet XY is not. Hence the independence of X and Y is crucial in Theorem 6.3(i). It is essential in assertion (ii) as well, as can be seen by letting X equal either $10^{\sqrt{2}-1}$ or $10^{2-\sqrt{2}}$ with probability $\frac{1}{2}$ each, and choosing $Y:=X^{-2}$. Then $S(X)$ and $S(XY)=S(X^{-1})$ have the same distribution, but neither X is Benford nor $\log S(Y)$ is rational with probability one.

Corollary 6.5. *Let X be a random variable with $\mathbb{P}(X=0)=0$, and let α be an irrational number. If $S(X)$ and $S(\alpha X)$ have the same distribution, i.e., if X and αX have the same distribution of significant digits, then X is Benford.*

Now consider a *sequence* (X_n) of independent random variables. From Theorem 6.3 it is clear that if the product $\prod_{j=1}^n X_j$ is Benford for all sufficiently large n then one of the factors X_j is necessarily Benford. Clearly, this is a very restrictive assumption, especially in the i.i.d. case, where *all* X_j would have to be Benford. Much greater applicability is achieved by requiring $\prod_{j=1}^n X_j$ to conform to BL only *asymptotically*. As the following theorem, an i.i.d. counterpart of Theorem 6.1, shows, convergence of $\big(\prod_{j=1}^n X_j\big)$ in distribution to BL is a very widespread phenomenon. The result may help explain why BL often appears in mathematical models that describe e.g. the evolution of stock prices by regarding the future price as a product of the current price times a large number of successive, multiplicative changes in price, with the latter being modeled as independent continuous random variables.

Theorem 6.6. *Let (X_n) be an i.i.d. sequence of random variables that are not purely atomic, i.e. $\mathbb{P}(X_1\in C)<1$ for every countable set $C\subset\mathbb{R}$. Then:*

(i) $\big(\prod_{j=1}^n X_j\big)$ *converges in distribution to BL.*

(ii) *With probability one,* $\big(\prod_{j=1}^n X_j\big)$ *is Benford.*

Proof. Let $Y_n=\log|X_n|$. Then (Y_n) is an i.i.d. sequence of random variables that are not purely atomic. By Theorem 4.13(iii), the sequence of $\big\langle\sum_{j=1}^n Y_j\big\rangle=\big\langle\log|\prod_{j=1}^n X_j|\big\rangle$ converges in distribution to $U(0,1)$. Thus $\big(\prod_{j=1}^n X_j\big)$ converges in distribution to BL.

To prove (ii), let Y_0 be u.d. mod 1 and independent of $(Y_n)_{n\in\mathbb{N}}$, and define

$$S_j := \langle Y_0 + Y_1 + \ldots + Y_j \rangle \,, \quad j \in \mathbb{N}_0 \,.$$

Recall from Theorem 4.13(i) that S_j is uniform on $[0,1)$ for every $j \geq 0$. Also note that, by definition, the random variables $Y_{j+1}, Y_{j+2}, \ldots$ are independent of S_j. The following argument is most transparent when formulated in ergodic theory terminology. (For an alternative approach see e.g. [Ro].) To this end, endow

$$\mathbb{T}_\infty := [0,1)^{\mathbb{N}_0} = \{(x_j)_{j\in\mathbb{N}_0} : x_j \in [0,1) \text{ for all } j\}$$

with the σ-algebra

$$\begin{aligned}\mathcal{B}_\infty &:= \bigotimes\nolimits_{j\in\mathbb{N}_0} \mathcal{B}[0,1) \\ &:= \sigma\big(\{B_0 \times B_1 \times \ldots \times B_j \times [0,1) \times [0,1) \times \ldots : j \in \mathbb{N}_0,\, B_0, B_1, \ldots, B_j \in \mathcal{B}[0,1)\}\big)\,.\end{aligned}$$

A probability measure P_∞ is uniquely defined on $(\mathbb{T}_\infty, \mathcal{B}_\infty)$ by setting

$$P_\infty(B_0 \times B_1 \times \ldots \times B_j \times [0,1) \times [0,1) \times \ldots) = \mathbb{P}(S_0 \in B_0, S_1 \in B_1, \ldots, S_j \in B_j)$$

for all $j \in \mathbb{N}_0$ and $B_0, B_1, \ldots, B_j \in \mathcal{B}[0,1)$. The map $\sigma_\infty : \mathbb{T}_\infty \to \mathbb{T}_\infty$ with $\sigma_\infty\big((x_j)\big) = (x_{j+1})$, often referred to as the (*one-sided*) *left shift* on $\mathbb{T}_\infty$ (cf. Example 5.25), is clearly measurable, i.e. $\sigma_\infty^{-1}(A) \in \mathcal{B}_\infty$ for every $A \in \mathcal{B}_\infty$. As a consequence, $(\sigma_\infty)_* P_\infty$ is a well-defined probability measure on $(\mathbb{T}_\infty, \mathcal{B}_\infty)$. In fact, since S_1 is u.d. mod 1 and (Y_n) is an i.i.d. sequence,

$$\begin{aligned}(\sigma_\infty)_* P_\infty&(B_0 \times B_1 \times \ldots \times B_j \times [0,1) \times [0,1) \times \ldots) \\ &= P_\infty([0,1) \times B_0 \times B_1 \times \ldots \times B_j \times [0,1) \times [0,1) \times \ldots) \\ &= \mathbb{P}(S_1 \in B_0, S_2 \in B_1, \ldots, S_{j+1} \in B_j) \\ &= \mathbb{P}(S_0 \in B_0, S_1 \in B_1, \ldots, S_j \in B_j) \\ &= P_\infty(B_0 \times B_1 \times \ldots \times B_j \times [0,1) \times [0,1) \times \ldots)\,,\end{aligned}$$

showing that $(\sigma_\infty)_* P_\infty = P_\infty$, i.e., σ_∞ is P_∞-preserving. (In probabilistic terms, this is equivalent to saying that the random process $(S_j)_{j\in\mathbb{N}_0}$ is *stationary*, see [Sh, Def.V.1.1].) It will now be shown that σ_∞ is even *ergodic* with respect to P_∞. Recall that this simply means that every invariant set $A \in \mathcal{B}_\infty$ has measure zero or one, or, more formally, that $P_\infty(\sigma_\infty^{-1}(A)\Delta A) = 0$ implies $P_\infty(A) \in \{0,1\}$; here the symbol Δ denotes the symmetric difference of two sets, i.e. $A\Delta B = A\backslash B \cup B\backslash A$. Assume, therefore, that $P_\infty(\sigma_\infty^{-1}(A)\Delta A) = 0$ for some $A \in \mathcal{B}_\infty$. Given $\varepsilon > 0$, there exists a number $N \in \mathbb{N}$ and sets $B_0, B_1, \ldots, B_N \in \mathcal{B}[0,1)$ such that

$$P_\infty\big(A \,\Delta\, (B_0 \times B_1 \times \ldots \times B_N \times [0,1) \times [0,1) \times \ldots)\big) < \varepsilon\,.$$

For notational convenience, let $A_\varepsilon := B_0 \times B_1 \times \ldots \times B_N \times [0,1) \times [0,1) \times \ldots \in \mathcal{B}_\infty$, and note that $P_\infty\big(\sigma_\infty^{-j}(A)\Delta\sigma_\infty^{-j}(A_\varepsilon)\big) < \varepsilon$ for all $j \in \mathbb{N}_0$. Recall now from

Theorem 4.13(iii) that, given $S_0, S_1, \dots, S_N$, the random variables S_n converge in distribution to $U(0,1)$. Thus, for all sufficiently large M,

$$\left| P_\infty\big(A_\varepsilon^c \cap \sigma_\infty^{-M}(A_\varepsilon)\big) - P_\infty(A_\varepsilon^c)P_\infty(A_\varepsilon)\right| < \varepsilon \,, \tag{6.2}$$

and similarly $\left|P_\infty\big(A_\varepsilon \cap \sigma_\infty^{-M}(A_\varepsilon^c)\big) - P_\infty(A_\varepsilon)P_\infty(A_\varepsilon^c)\right| < \varepsilon$. (Note that (6.2) may not hold if X_1, and hence also Y_1, is purely atomic, see for instance Example 4.14(ii).) Overall, therefore,

$$\begin{aligned} 2P_\infty(A_\varepsilon)\big(1 - P_\infty(A_\varepsilon)\big) &\le 2\varepsilon + P_\infty\big(A_\varepsilon \Delta \sigma_\infty^{-M}(A_\varepsilon)\big) \\ &\le 2\varepsilon + P_\infty(A_\varepsilon \Delta A) + P_\infty\big(A \Delta \sigma_\infty^{-M}(A)\big) + P_\infty\big(\sigma_\infty^{-M}(A)\Delta\sigma_\infty^{-M}(A_\varepsilon)\big) \\ &< 4\varepsilon \,, \end{aligned}$$

and consequently $P_\infty(A)\big(1 - P_\infty(A)\big) < 3\varepsilon + \varepsilon^2$. Since $\varepsilon > 0$ was arbitrary, $P_\infty(A) \in \{0,1\}$, which in turn shows that σ_∞ is ergodic. (Again, this is equivalent to saying, in probabilistic parlance, that the random process $(S_j)_{j\in\mathbb{N}_0}$ is *ergodic*, see [Sh, Def.V.3.2].) By the Birkhoff Ergodic Theorem, for every (measurable) function $f : [0,1) \to \mathbb{C}$ with $\int_0^1 |f(x)|\,\mathrm{d}x < +\infty$,

$$\frac{1}{n}\sum\nolimits_{j=0}^{n} f(x_j) \;\to\; \int_0^1 f(x)\,\mathrm{d}x \quad \text{as } n \to \infty$$

holds for all $(x_j)_{j\in\mathbb{N}_0} \in \mathbb{T}_\infty$, with the possible exception of a set of P_∞-measure zero. In probabilistic terms, this means that

$$\lim\nolimits_{n\to\infty} \frac{1}{n}\sum\nolimits_{j=0}^{n} f(S_j) = \int_0^1 f(x)\,\mathrm{d}x \quad \text{a.s.} \tag{6.3}$$

Assume from now on that f is actually *continuous* with $\lim_{x\uparrow 1} f(x) = f(0)$, e.g. $f(x) = e^{2\pi\imath x}$. For any such f, as well as any $t \in [0,1)$ and $m \in \mathbb{N}$, denote the set

$$\left\{\omega\in\Omega : \limsup\nolimits_{n\to\infty} \left| \frac{1}{n}\sum\nolimits_{j=1}^{n} f\big(\langle t + Y_1(\omega) + \ldots + Y_j(\omega)\rangle\big) - \int_0^1 f(x)\,\mathrm{d}x \right| < \frac{1}{m}\right\}$$

simply by $\Omega_{f,t,m}$. According to (6.3), $1 = \int_0^1 \mathbb{P}(\Omega_{f,t,m})\,\mathrm{d}t$, and hence $\mathbb{P}(\Omega_{f,t,m}) = 1$ for a.e. $t \in [0,1)$. Since f is uniformly continuous, for every $m \ge 2$ there exists $t_m > 0$ such that $\mathbb{P}(\Omega_{f,t_m,m}) = 1$ and $\Omega_{f,t_m,m} \subset \Omega_{f,0,\lfloor m/2\rfloor}$. From

$$1 = \mathbb{P}\left(\bigcap\nolimits_{m\ge 2} \Omega_{f,t_m,m}\right) \le \mathbb{P}\left(\bigcap\nolimits_{m\ge 2} \Omega_{f,0,\lfloor m/2\rfloor}\right) \le 1\,,$$

it is clear that

$$\lim\nolimits_{n\to\infty} \frac{1}{n}\sum\nolimits_{j=1}^{n} f\big(\langle Y_1 + \ldots + Y_j\rangle\big) = \int_0^1 f(x)\,\mathrm{d}x \quad \text{a.s.} \tag{6.4}$$

As the intersection of countably many sets of full measure has itself full measure, choosing $f(x) = e^{2\pi\imath kx}$, $k \in \mathbb{Z}$ in (6.4) shows that, with probability one,

$$\lim_{n\to\infty} \frac{1}{n} \sum\nolimits_{j=1}^{n} e^{2\pi\imath k(Y_1+\ldots+Y_j)} = \int_0^1 e^{2\pi\imath kx}\mathrm{d}x = 0 \quad \text{for all } k \in \mathbb{Z}, k \neq 0\,. \tag{6.5}$$

By Weyl's criterion [KN, Thm.2.1], (6.5) is equivalent to

$$\mathbb{P}\Big(\big(\sum\nolimits_{j=1}^{n} Y_j\big) \text{ is u.d. mod } 1 \Big) = 1\,.$$

In other words, $(\prod_{j=1}^{n} X_j)$ is Benford with probability one. $\square$

Remarks. (i) As has essentially been observed already in Example 4.14(ii), for Theorem 6.6(i) to hold it is necessary and sufficient that

$$\mathbb{P}(\log |X_1| \in a + \tfrac{1}{m}\mathbb{Z}) < 1 \quad \text{for all } a \in \mathbb{R}, m \in \mathbb{N}\,. \tag{6.6}$$

On the other hand, it is not hard to see that (ii) holds if and only if

$$\mathbb{P}(\log |X_1| \in \tfrac{1}{m}\mathbb{Z}) < 1 \quad \text{for all } m \in \mathbb{N}\,, \tag{6.7}$$

which is a strictly weaker assumption than (6.6). Clearly, if X_1 is not purely atomic then (6.6), and hence also (6.7) hold. However, if e.g. $X_1 = 2$ with probability one then (6.6) does not hold, and correspondingly $\big(\prod_{j=1}^{n} X_j\big) = (2^n)$ does not converge in distribution to BL, whereas (6.7) does hold, and $\big(\prod_{j=1}^{n} X_j\big)$ is Benford with probability one, cf. Example 6.2(ii).

(ii) For more general results in the spirit of Theorem 6.6 the reader is referred to [Schü1, Schü2]. ♣

Example 6.7. (i) Let (X_n) be an i.i.d. sequence with X_1 distributed according to $U(0,a)$, the uniform distribution on $[0,a)$ with $a > 0$. The k-th Fourier coefficient of $P_{\langle \log X_1\rangle}$ is

$$\widehat{P_{\langle \log X_1\rangle}}(k) = e^{2\pi\imath k \log a} \frac{\ln 10}{\ln 10 + 2\pi\imath k}\,, \quad k \in \mathbb{Z}\,,$$

so that, for every $k \neq 0$,

$$\left|\widehat{P_{\langle \log X_1\rangle}}(k)\right| = \frac{\ln 10}{\sqrt{(\ln 10)^2 + 4\pi^2 k^2}} < 1\,.$$

As seen in the proof of Theorem 4.13(iii), this implies that $\big(\prod_{j=1}^{n} X_j\big)$ converges in distribution to BL, a fact apparently first recorded in [AS]. Note also that $\mathbb{E}\log X_1 = \log \frac{a}{e}$. Thus with probability one, $\big(\prod_{j=1}^{n} X_j\big)$ converges to 0 or $+\infty$, depending on whether $a < e$ or $a > e$. In fact, by the Strong Law of Large Numbers [CT],

$$\sqrt[n]{\prod\nolimits_{j=1}^{n} X_j} \overset{a.s.}{\to} \frac{a}{e}$$

holds for every $a > 0$. If $a = e$ then

$$\mathbb{P}\left(\liminf_{n\to\infty} \prod_{j=1}^{n} X_j = 0 \text{ and } \limsup_{n\to\infty} \prod_{j=1}^{n} X_j = +\infty\right) = 1\,,$$

showing that in this case the product $\prod_{j=1}^{n} X_j$ does not converge but rather attains, with probability one, arbitrarily small as well as arbitrarily large positive values. By Theorem 6.6(ii), the sequence $\left(\prod_{j=1}^{n} X_j\right)$ is a.s. Benford, regardless of the value of a.

(ii) Consider an i.i.d. sequence (X_n) with X_1 distributed according to a log-normal distribution such that $\log X_1$ is standard normal. Denote by f_n the density of $\left\langle \log \prod_{j=1}^{n} X_j \right\rangle$. Since $\log \prod_{j=1}^{n} X_j = \sum_{j=1}^{n} \log X_j$ is normal with mean zero and variance n,

$$f_n(s) = \frac{1}{\sqrt{2\pi n}} \sum_{k\in\mathbb{Z}} e^{-(k+s)^2/(2n)}\,, \quad 0 \le s < 1\,,$$

from which it is straightforward to deduce that

$$\lim_{n\to\infty} f_n(s) = 1\,, \quad \text{uniformly in } 0 \le s < 1\,.$$

Consequently, for all $t \in [1, 10)$,

$$\begin{aligned}\mathbb{P}\left(S\left(\prod_{j=1}^{n} X_j\right) \le t\right) &= \mathbb{P}\left(\left\langle \log \prod_{j=1}^{n} X_j \right\rangle \le \log t\right) \\ &= \int_0^{\log t} f_n(s)\,\mathrm{d}s \;\to\; \int_0^{\log t} 1\,\mathrm{d}s \;=\; \log t\,,\end{aligned}$$

i.e., $\left(\prod_{j=1}^{n} X_j\right)$ converges in distribution to BL. By Theorem 6.6(ii) also

$$\mathbb{P}\left(\left(\prod_{j=1}^{n} X_j\right) \text{ is Benford}\right) = 1\,,$$

even though $\mathbb{E} \log \prod_{j=1}^{n} X_j = \sum_{j=1}^{n} \mathbb{E} \log X_j = 0$, and hence, as in the previous example, the sequence $\left(\prod_{j=1}^{n} X_j\right)$ a.s. oscillates forever between 0 and $+\infty$.

Having seen Theorem 6.6, the reader may wonder whether there is an analogous result for *sums* of i.i.d. random variables. After all, the focus in classical probability theory is on sums much more than on products. Unfortunately, the statistical behavior of the significands is much more complex for sums than for products. The main basic reason is that the significand of the sum of two or more numbers depends not only on the significand of each each number (as in the case of products), but also on their *exponents*. For example, observe that

$$S\left(3 \cdot 10^3 + 2 \cdot 10^2\right) = 3.2 \ne 5 = S\left(3 \cdot 10^2 + 2 \cdot 10^2\right),$$

while clearly

$$S\left(3 \cdot 10^3 \times 2 \cdot 10^2\right) = 6 = S\left(3 \cdot 10^2 \times 2 \cdot 10^2\right).$$

Practically, this difficulty is reflected in the fact that for positive real numbers u, v, the value of $\log(u+v)$, relevant for conformance with BL via Theorem 4.2, is not easily expressed in terms of $\log u$ and $\log v$, whereas $\log(uv) = \log u + \log v$.

In view of these difficulties, it is perhaps not surprising that the analog of Theorem 6.6 for sums arrives at a radically different conclusion.

Theorem 6.8. *Let (X_n) be an i.i.d. sequence of random variables with finite variance, that is $\mathbb{E}X_1^2 < +\infty$. Then:*

(i) *Not even a subsequence of $\left(\sum_{j=1}^n X_j\right)$ converges in distribution to BL.*

(ii) *With probability one, $\left(\sum_{j=1}^n X_j\right)$ is not Benford.*

Proof. Assume first that $\mathbb{E}X_1 \neq 0$. By the Strong Law of Large Numbers, $\frac{1}{n}\left|\sum_{j=1}^n X_j\right|$ converges a.s., and hence also in distribution, to the constant $|\mathbb{E}X_1|$. Since

$$\log S\left(\left|\sum\nolimits_{j=1}^n X_j\right|\right) = \left\langle \log\left|\sum\nolimits_{j=1}^n X_j\right|\right\rangle = \left\langle \log\left(\frac{1}{n}\left|\sum\nolimits_{j=1}^n X_j\right|\right) + \log n\right\rangle ,$$

any subsequence of $\left(S\left(\frac{1}{n}\left|\sum_{j=1}^n X_j\right|\right)\right)$ either does not converge in distribution at all or else converges to a constant; in neither case, therefore, is the limit a Benford random variable. Since, with probability one, $\left|\sum_{j=1}^n X_j\right| \to +\infty$, it follows from

$$\log\left|\sum\nolimits_{j=1}^n X_j\right| - \log n = \log\frac{1}{n}\left|\sum\nolimits_{j=1}^n X_j\right| \overset{a.s.}{\to} |\mathbb{E}X_1| ,$$

together with Lemma 4.3(i) and Proposition 4.8(iii) that $\left(\sum_{j=1}^n X_j\right)$ is, with probability one, not Benford.

It remains to consider the case $\mathbb{E}X_1 = 0$. Without loss of generality, it can be assumed that $\mathbb{E}X_1^2 = 1$. By the Central Limit Theorem $\frac{1}{\sqrt{n}}\sum_{j=1}^n X_j$ converges in distribution to the standard normal distribution. Thus for sufficiently large n, and up to a rotation (i.e. an addition mod 1) of $[0,1)$, the distribution of $\langle \log|\sum_{j=1}^n X_j|\rangle$ differs by arbitrarily little from the distribution of $Y := \langle \log|Z|\rangle$, where Z is standard normal. Intuitively, it is clear that $P_Y \neq \lambda_{0,1}$, i.e., Y is *not* uniform on $[0,1)$. To see this more formally, note that

$$F_Y(s) = 2\sum\nolimits_{k\in\mathbb{Z}} \left(\Phi\left(10^{s+k}\right) - \Phi\left(10^k\right)\right) , \quad 0 \le s < 1 , \tag{6.8}$$

with Φ $(= F_Z)$ denoting the standard normal distribution function, see Example 4.16(ii). Thus

$$|F_Y(s) - s| \ge F_Y(s) - s > 2\left(\Phi\left(10^s\right) - \Phi\left(1\right)\right) - s =: R(s) , \quad 0 \le s < 1 ,$$

and since R is concave on $[0,1)$ with $R(0) = 0$ and $R'(0) = \frac{2\ln 10}{\sqrt{2\pi e}} - 1 = 0.1143\ldots > \frac{1}{9}$, it follows that

$$\max\nolimits_{0\le s<1} |F_Y(s) - s| > \max\nolimits_{0\le s<1} R(s) > 0 ,$$

showing that indeed $P_Y \neq \lambda_{0,1}$, and hence $\left(\sum_{j=1}^n X_j\right)$ does not converge in distribution to BL.

The verification of (ii) in the case $\mathbb{E}X_1 = 0$ uses an almost sure version of the Central Limit Theorem, see [LP]. With the random variables X_n defined on some (abstract) probability space $(\Omega, \mathcal{A}, \mathbb{P})$, let $\Omega_1 := \left\{\omega \in \Omega : \left(\sum_{j=1}^n X_j(\omega)\right) \text{ is Benford}\right\}$. By Theorem 4.2 and Proposition 4.8(iii), the sequence $\left(x_n(\omega)\right)$ with

$$x_n(\omega) = \log \frac{1}{\sqrt{n}} \left|\sum\nolimits_{j=1}^n X_j(\omega)\right| , \quad n \in \mathbb{N} ,$$

is u.d. mod 1 for all $\omega \in \Omega_1$. For every interval $[a, b) \subset [0, 1)$, therefore,

$$\frac{1}{\ln N} \sum\nolimits_{n=1}^N \frac{\mathbb{1}_{[a,b)}\left(x_n(\omega)\right)}{n} \to b - a \quad \text{as } N \to \infty .$$

(Recall the remark on p.18.) However, as a consequence of [LP, Thm.2], for every $[a, b) \subset [0, 1)$,

$$\frac{1}{\ln N} \sum\nolimits_{n=1}^N \frac{\mathbb{1}_{[a,b)}\left(x_n\right)}{n} \overset{a.s.}{\to} F_Y(b) - F_Y(a) ,$$

with F_Y given by (6.8). As shown above, $F_Y(s) \not\equiv s$, and therefore $\mathbb{P}(\Omega_1) = 0$. In other words, $\mathbb{P}\left((\sum_{j=1}^n X_j) \text{ is Benford}\right) = 0$. $\square$

Example 6.9. (i) Let (X_n) be an i.i.d. sequence with $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = \frac{1}{2}$. Then $\mathbb{E}X_1 = \mathbb{E}X_1^2 = \frac{1}{2}$, and by Theorem 6.8(i) neither $\left(\sum_{j=1}^n X_j\right)$ nor any of its subsequences converges in distribution to BL. Note that $\sum_{j=1}^n X_j$ is binomial with parameters n and $\frac{1}{2}$, i.e. for all $n \in \mathbb{N}$,

$$\mathbb{P}\left(\sum\nolimits_{j=1}^n X_j = l\right) = 2^{-n} \binom{n}{l} , \quad l = 0, 1, \ldots, n .$$

The Law of the Iterated Logarithm [CT] asserts that

$$\sum\nolimits_{j=1}^n X_j = \frac{n}{2} + Y_n \sqrt{n \ln \ln n} \quad \text{for all } n \geq 3 , \tag{6.9}$$

where the sequence (Y_n) of random variables is bounded, in fact $|Y_n| \leq 1$ a.s. for all n. From (6.9) it is clear that, with probability one, the sequence $\left(\sum_{j=1}^n X_j\right)$ is not Benford.

(ii) Let (X_n) be an i.i.d. sequence of Cauchy random variables. As $\mathbb{E}|X_1|$ is even infinite, Theorem 6.8 does not apply. However, recall from Example 4.14(i) that $\frac{1}{n}\sum_{j=1}^n X_j$ is again Cauchy, and hence the distribution of $\langle \log |(\sum_{j=1}^n X_j)| \rangle$ is but a rotated version of $P_{\langle \log |X_1| \rangle}$, the density of which is given by

$$f_{\langle \log |X_1| \rangle}(s) = \frac{\ln 10}{\pi} \sum\nolimits_{k \in \mathbb{Z}} \frac{1}{\cosh\left((s + k) \ln 10\right)} , \quad 0 \leq s < 1.$$

The density $f_{\langle \log |X_1| \rangle}$ is a smooth function, and

$$f_{\langle \log |X_1| \rangle}(0) = \frac{\ln 10}{\pi} \sum\nolimits_{k \in \mathbb{Z}} \frac{2}{10^k + 10^{-k}} > \frac{\ln 10}{\pi} \left(1 + \frac{40}{101}\right) > 1 + \frac{2}{101}\,,$$

showing that $\left(\log |\sum_{j=1}^n X_j|\right)$ is not u.d. mod 1. Hence the sequence $\left(\sum_{j=1}^n X_j\right)$ does not converge in distribution to BL, and nor does any of its subsequences.

This example shows that the conclusions of Theorem 6.8 may hold, at least in parts, even if the X_n do not have finite first, let alone finite second moments.

Remark. Recall from the remark on p.18 that a sequence failing to be Benford may conform to a weaker form of BL. As seen above, under mild conditions the stochastic sequence $\left(\sum_{j=1}^n X_j\right)$ is not Benford. Under the appropriate assumptions, however, it does obey a weaker form of BL, see [Scha2]. ♣

6.2. Mixtures of distributions

The characterizations of the Benford distribution via scale-, base- and sum-invariance, given in Chapter 4, although perhaps mathematically satisfying, hardly help explain the appearance of BL empirically in real-life data. Application of those theorems requires explaining why the underlying data is scale- or base-invariant in the first place. BL nevertheless does appear in many real-life datasets. Thus the question arises: What do the population data of three thousand U.S. counties according to the 1990 census have in common with the usage of logarithm tables during the 1880s, numerical data from newspaper articles of the 1930's collected by Benford, or universal physical constants examined by Knuth in the 1960's? Why should these data exhibit a logarithmically distributed significand or equivalently, why should they be scale- or base-invariant?

As a matter of fact, most data-sets do not follow BL closely. Benford already observed that while some of his tables conformed to BL reasonably well, many others did not. But, as Raimi [Ra1] points out, "what came closest of all, however, was the union of all his tables." Combine the molecular weight tables with baseball statistics and drainage areas of rivers, and *then* there is a very good fit. Many of the previous explanations of BL have first hypothesized some universal table of constants, such as Raimi's [Ra1] "stock of tabular data in the world's libraries", or Knuth's [Kn] "imagined set of real numbers", and then tried to prove why certain specific sets of real observations were representative of either this mysterious universal table or the set of all real numbers. What seems more natural though is to think of data as coming from many different distributions. This was clearly the case in Benford's original study. After all, he had made an effort "to collect data from as many fields as possible and to include a wide variety of types", noting that "the range of subjects studied and tabulated was as wide as time and energy permitted".

The main goal of this section is to provide a statistical derivation of BL, in the form of a central-limit-like theorem that says that if random samples are taken from different distributions, and the results combined, then — provided the sampling is "unbiased" as to scale or base — the resulting combined samples will converge to the Benford distribution.

Denote by $\mathcal{M}$ the set of all probability measures on $(\mathbb{R}, \mathcal{B})$. Recall that a (real Borel) *random probability measure*, abbreviated henceforth as *r.p.m.*, is a function $P : \Omega \to \mathcal{M}$, defined on some underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$, such that for every $B \in \mathcal{B}$ the function $\omega \mapsto P(\omega)(B)$ is a random variable. Thus, for every $\omega \in \Omega$, $P(\omega)$ is a probability measure on $(\mathbb{R}, \mathcal{B})$, and, given any real numbers a, b and any Borel set B,

$$\{\omega : a \le P(\omega)(B) \le b\} \in \mathcal{A}\,,$$

see e.g. [Ka] for an authoritative account on random probability measures. In more abstract conceptual terms, an r.p.m. can be interpreted as follows: When endowed with the topology of convergence in distribution, the set $\mathcal{M}$ becomes a complete and separable metrizable space. Denote by $\mathcal{B}_{\mathcal{M}}$ its Borel σ-algebra, defined as the smallest σ-algebra containing all open subsets of $\mathcal{M}$. Then $P_*\mathbb{P}$ simply is a probability measure on $(\mathcal{M}, \mathcal{B}_{\mathcal{M}})$.

Example 6.10. **(i)** Let P be an r.p.m. that is $U(0,1)$ with probability $\frac{1}{2}$, and otherwise is $\texttt{exp}(1)$, i.e. exponential with mean 1, hence $\mathbb{P}(X > t) = \min(1, e^{-t})$ for all $t \in \mathbb{R}$, see Example 3.10(i,ii). Thus, for every $\omega \in \Omega$, the probability measure P is either $U(0,1)$ or $\texttt{exp}(1)$, and $\mathbb{P}\bigl(P = U(0,1)\bigr) = \mathbb{P}\bigl(P = \texttt{exp}(1)\bigr) = \frac{1}{2}$. For a practical realization of P simply flip a fair coin — if it comes up heads, $\mathbb{P}(\omega)$ is a $U(0,1)$-distribution, and if it comes up tails, then P is an $\texttt{exp}(1)$-distribution.

(ii) Let X be distributed according to $\texttt{exp}(1)$, and let P be an r.p.m. where, for each $\omega \in \Omega$, $P(\omega)$ is the normal distribution with mean $X(\omega)$ and variance 1. In contrast to the example in (i), here P is continuous, i.e., $\mathbb{P}(P = Q) = 0$ for each probability measure $Q \in \mathcal{M}$.

The following example of an r.p.m. is a variant of a classical construction due to L. Dubins and D. Freedman which, as will be seen below, is an r.p.m. leading to BL.

Example 6.11. Let P be the r.p.m. with support on $[1, 10)$, i.e. $P\bigl([1, 10)\bigr) = 1$ with probability one, defined by its (random) cumulative distribution function F_P, i.e.

$$F_P(t) := F_{P(\omega)}(t) = P(\omega)\bigl([1, t]\bigr)\,, \quad 1 \le t < 10\,,$$

as follows: Set $F_P(1) = 0$ and $F_P(10) = 1$. Next pick $F_P(10^{1/2})$ according to the uniform distribution on $[0, 1)$. Then pick $F_P(10^{1/4})$ and $F_P(10^{3,4})$ independently, uniformly on $\bigl[0, F_P(10^{1/2})\bigr)$ and $\bigl[F_P(10^{1/2}), 1\bigr)$, respectively, and continue in this manner. This construction is known to generate an r.p.m. a.s. [DF, Lem.9.28], and as can easily be seen, is dense in the set of all probability measures on $\bigl([1, 10), \mathcal{B}[1, 10)\bigr)$, i.e., it generates probability measures that are arbitrarily close to any Borel probability measure on $[1, 10)$.

The next definition formalizes the notion of combining data from different distributions. Essentially, it mimics what Benford did in combining baseball statistics with square-root tables and numbers taken from newspapers, etc. This definition is key to everything that follows. It rests upon using an r.p.m. to generate a random sequence of probability distributions, and then successively selecting random samples from each of those distributions.

Definition 6.12. Let m be a positive integer and P an r.p.m. A *sequence of P-random m-samples* is a sequence (X_n) of random variables on $(\Omega, \mathcal{A}, \mathbb{P})$ such that, for all $j \in \mathbb{N}$ and some i.i.d. sequence (P_n) of r.p.m.s with $P_1 = P$, the following two properties hold:

$$\text{Given } P_j = Q, \text{ the random variables } X_{(j-1)m+1}, X_{(j-1)m+2}, \ldots, X_{jm} \text{ are i.i.d. with distribution } Q\,; \tag{6.10}$$

$$\text{The variables } X_{(j-1)m+1}, X_{(j-1)m+2}, \ldots, X_{jm} \text{ are independent of } P_i, X_{(i-1)m+1}, X_{(i-1)m+2}, \ldots, X_{im} \text{ for every } i \neq j\,. \tag{6.11}$$

Thus for any sequence (X_n) of P-random m-samples, for each $\omega \in \Omega$ in the underlying probability space, the first m random variables are a random sample (i.e., i.i.d.) from $P_1(\omega)$, a random probability distribution chosen according to the r.p.m. P; the second m-tuple of random variables is a random sample from $P_2(\omega)$ and so on. Note the two levels of randomness here: First a probability is selected at random, and then a random sample is drawn from this distribution, and this two-tiered process is continued.

Example 6.13. Let P be the r.p.m. in Example 6.10(i), and let $m = 3$. Then a sequence of P-random 3-samples is a sequence (X_n) of random variables such that with probability $\frac{1}{2}$, X_1, X_2, X_3, are i.i.d. and distributed according to $U(0,1)$, and otherwise they are i.i.d. but distributed according to $\texttt{exp}(1)$; the random variables X_4, X_5, X_6 are again equally likely to be i.i.d. $U(0,1)$ or $\texttt{exp}(1)$, and they are independent of X_1, X_2, X_3, etc. Clearly the (X_n) are all identically distributed as they are all generated by exactly the same process. Note, however, that for instance X_1 and X_2 are dependent: Given that $X_1 > 1$, for example, the random variable X_2 is $\texttt{exp}(1)$-distributed with probability one, whereas the unconditional probability that X_2 is $\texttt{exp}(1)$-distributed is only $\frac{1}{2}$.

Remark. If (X_n) is a sequence of P-random m-samples for some m and some r.p.m. P, then the X_n are a.s. identically distributed according to the distribution that is the average (expected) distribution of P (see Proposition 6.15 below), but they are not in general independent (see Example 6.13). On the other hand, given $(P_1, P_2, \ldots)$, the (X_n) are a.s. independent, but clearly are not in general identically distributed. ♣

Although sequences of P-random m-samples have a fairly simple structure, they do not fit into any of the familiar categories of sequences of random variables. For example, they are not in general independent, exchangeable, Markov, martingale, or stationary sequences.

Example 6.14. Assume that the r.p.m. P is, with equal probability, the Dirac measure concentrated at 1 and the probability measure $\frac{1}{2}(\delta_1 + \delta_2)$, respectively, i.e. $\mathbb{P}(P=\delta_1)=\mathbb{P}\big(P=\frac{1}{2}(\delta_1+\delta_2)\big)=\frac{1}{2}$. Let (X_n) be a sequence of P-random 3-samples. Then the random variables $X_1, X_2, \ldots$ are

not independent as

$$\mathbb{P}(X_2=2)=\frac{1}{4} \quad \text{but} \quad \mathbb{P}(X_2=2|X_1=2)=\frac{1}{2}\,;$$

not exchangeable as

$$\mathbb{P}\big((X_1,X_2,X_3,X_4)=(1,1,1,2)\big)=\frac{9}{64}\neq\frac{3}{64}=\mathbb{P}\big((X_1,X_2,X_3,X_4)=(2,2,2,1)\big)\,;$$

not Markov as

$$\mathbb{P}(X_3=1|X_1=X_2=1)=\frac{9}{10}\neq\frac{5}{6}=\mathbb{P}(X_3=1|X_2=1)\,;$$

not martingale as

$$\mathbb{E}(X_2|X_1=2)=\frac{3}{2} \quad \text{but} \quad \mathbb{E}X_2=\frac{5}{4}\,;$$

not stationary as

$$\mathbb{P}\big((X_1,X_2,X_3)=(1,1,1)\big)=\frac{9}{16}\neq\frac{15}{32}=\mathbb{P}\big((X_2,X_3,X_4)=(1,1,1)\big)\,.$$

Recall that, given an r.p.m. P and any Borel set B, the quantity $P(B)$ is a random variable with values between 0 and 1. The following property of the expectation of $P(B)$, as a function of B, is easy to check.

Proposition 6.15. *Let P be an r.p.m. Then $\mathbb{E}P$, defined as*

$$(\mathbb{E}P)(B):=\mathbb{E}P(B)=\int_\Omega P(\omega)(B)\,\mathrm{d}\mathbb{P}(\omega) \quad \text{for all } B\in\mathcal{B}\,,$$

is a probability measure on $(\mathbb{R},\mathcal{B})$.

Example 6.16. (i) Let P be the r.p.m. of Example 6.10(i). Then $\mathbb{E}P$ is the Borel probability measure with density

$$f_{\mathbb{E}P}(t)=\left\{\begin{array}{ll} 0 & \text{if } t<0\,, \\ \frac{1}{2}+\frac{1}{2}e^{-t} & \text{if } 0\le t<1\,, \\ \frac{1}{2}e^{-t} & \text{if } t\ge 1\,, \end{array}\right\}=\frac{1}{2}\mathbb{1}_{[0,1)}(t)+\frac{1}{2}e^{-t}\mathbb{1}_{[0,+\infty)}\,, \quad t\in\mathbb{R}\,.$$

(ii) Consider the r.p.m. P of Example 6.10(ii), that is, $P(\omega)$ is normal with mean $X(\omega)$ and variance 1, where X is distributed according to $\texttt{exp}(1)$. In this case, $\mathbb{E}P$ is also a.c., with density

$$f_{\mathbb{E}P}(t)=\frac{1}{\sqrt{2\pi}}\int_0^{+\infty}e^{-\frac{1}{2}(t-\tau)^2}e^{-\tau}\mathrm{d}\tau=e^{\frac{1}{2}-t}\big(1-\Phi(1-t)\big)\,, \quad t\in\mathbb{R}\,.$$

(iii) Even if P is a.c. only with probability zero, it is possible for $\mathbb{E}P$ to be a.c. As a simple example, let X be $\texttt{exp}(1)$-distributed and $P = \frac{1}{2}(\delta_{-X} + \delta_X)$. Then $\mathbb{P}(P \text{ is purely atomic}) = 1$, yet $\mathbb{E}P$ is the standard Laplace (or double-exponential) distribution; i.e., $\mathbb{E}P$ is a.c. with density

$$f_{\mathbb{E}P}(t) = \frac{e^{-|t|}}{2}, \quad t \in \mathbb{R}.$$

The next lemma shows that the limiting proportion of times that a sequence of P-random m-sample falls in a (Borel) set B is, with probability one, the average $\mathbb{P}$-value of the set B, i.e., the limiting proportion equals $\mathbb{E}P(B)$. Note that this is not simply a direct corollary of the classical Strong Law of Large Numbers as the random variables in the sequence are not in general independent (see Examples 6.13 and 6.14).

Lemma 6.17. *Let P be an r.p.m., and let (X_n) be a sequence of P-random m-samples for some $m \in \mathbb{N}$. Then, for every $B \in \mathcal{B}$,*

$$\frac{\#\{1 \le n \le N : X_n \in B\}}{N} \overset{a.s.}{\to} \mathbb{E}P(B) \quad \text{as } N \to \infty .$$

Proof. Fix $B \in \mathcal{B}$ and $j \in \mathbb{N}$, and let $Y_j = \#\{1 \le i \le m : X_{(j-1)m+i} \in B\}$. It is clear that

$$\lim\nolimits_{N\to\infty} \frac{\#\{1 \le n \le N : X_n \in B\}}{N} = \frac{1}{m} \lim\nolimits_{n\to\infty} \frac{1}{n} \sum\nolimits_{j=1}^{n} Y_j \,, \tag{6.12}$$

whenever the limit on the right exists. By (6.10), given P_j, the random variable Y_j is binomially distributed with parameters m and $\mathbb{E}\big(P_j(B)\big)$, hence a.s.

$$\mathbb{E}Y_j = \mathbb{E}\big(\mathbb{E}(Y_j|P_j)\big) = \mathbb{E}\big(mP_j(B)\big) = m\mathbb{E}P(B) \tag{6.13}$$

since P_j has the same distribution as P. By (6.11), the Y_j are independent. They are also uniformly bounded, as $0 \le Y_j \le m$ for all j, hence $\sum_{j=1}^{\infty} \mathbb{E}Y_j^2/j^2 < +\infty$. Moreover, by (6.13) all Y_j have the same mean value $m\mathbb{E}P(B)$. Thus by [CT, Cor.5.1]

$$\frac{1}{n} \sum\nolimits_{j=1}^{n} Y_j \overset{a.s.}{\to} m\mathbb{E}P(B) \quad \text{as } n \to \infty \,, \tag{6.14}$$

and the conclusion follows by (6.12) and (6.14). ☐

Remark. The assumption that each P_j is sampled exactly m times is not essential: The above argument can easily be modified to show that the same conclusion holds if the j-th r.p.m. is sampled M_j times where (M_j) is a sequence of independent, uniformly bounded $\mathbb{N}$-valued random variables which are also independent of the rest of the process. ♣

The stage is now set to give a statistical limit law (Theorem 6.20 below) that is a central-limit-like theorem for significant digits mentioned above. Roughly speaking, this law says that if probability distributions are selected at random, and random samples are then taken from each of these distributions in such a way that the overall process is scale- or base-neutral, then the significant digit frequencies of the combined sample will converge to the logarithmic distribution. This theorem may help explain and predict the appearance of BL in significant digits in mixtures of tabulated data such as the combined data from Benford's individual datasets, and also his individual dataset of numbers gleaned from newspapers.

In order to draw any conclusions concerning BL for the process of sampling from different distributions, clearly there must be *some* restriction on the underlying r.p.m. that generates the sampling procedure. Otherwise, if the r.p.m. is, say, $U(0,1)$ with probability one, for example, then any resulting sequence of P-random m-samples will be i.i.d. $U(0,1)$, and hence a.s. not Benford, by Example 3.10(i). Similarly, it can easily be checked that sequences of P-random m-samples from the r.p.m.s in Example 6.10 (i) and (ii) will not generate Benford sequences. A natural assumption to make concerning an r.p.m. in this context is that *on average* the r.p.m. is unbiased (i.e. invariant) with respect to changes in scale or base.

Definition 6.18. An r.p.m. P has *scale-unbiased (decimal) significant digits* if, for every significand event A, i.e. for every $A \in \mathcal{S}$, the expected value of $P(A)$ is the same as the expected value $P(\alpha A)$ for every $\alpha > 0$, that is, if

$$\mathbb{E}\big(P(\alpha A)\big) = \mathbb{E}\big(P(A)\big) \quad \text{for all } \alpha > 0, A \in \mathcal{S} .$$

Equivalently, the Borel probability measure $\mathbb{E}P$ has scale-invariant significant digits.

Similarly, P has *base-unbiased significant (decimal) digits* if, for every $A \in \mathcal{S}$ the expected value of $P(A)$ is the same as the expected value of $P(A^{1/n})$ for every $n \in \mathbb{N}$, that is, if

$$\mathbb{E}\big(P(A^{1/n})\big) = \mathbb{E}\big(P(A)\big) \quad \text{for all } n \in \mathbb{N}, A \in \mathcal{S} ,$$

i.e., if $\mathbb{E}P$ has base-invariant significant digits.

An immediate consequence of Theorems 4.20 and 4.30 is

Proposition 6.19. *Let P be an r.p.m. with $\mathbb{E}P(\{0\}) = 0$. Then the following statements are equivalent:*

(i) *P has scale-unbiased significant digits.*

(ii) *$P(\{\pm 10^k : k \in \mathbb{Z}\}) = 0$, or equivalently $S_*P(\{1\}) = 0$ holds with probability one, and P has base-unbiased significant digits.*

(iii) *$\mathbb{E}P(A) = \mathbb{B}(A)$ for all $A \in \mathcal{S}$, i.e., $\mathbb{E}P$ is Benford.*

Random probability measures with scale- or base-unbiased significant digits are easy to construct mathematically (see Example 6.22 below). In real-life examples, however, scale- or base-unbiased significant digits should not be taken for granted. For instance, picking beverage-producing companies in Europe at random, and looking at the metric volumes of samples of m products from each company, is not likely to produce data with scale-unbiased significant digits, since the volumes in this case are probably closely related to liters. Conversion of the data to another unit such as gallons would likely yield a radically different set of first-digit frequencies. On the other hand, if species of mammals in Europe are selected at random and their metric volumes sampled, it seems more likely that the latter process is unrelated to the choice of human units.

The question of base-unbiasedness of significant digits is most interesting when the units in question are universally agreed upon, such as the numbers of things, as opposed to sizes. For example, picking cities at random and looking at the number of leaves of m-samples of trees from those cities is certainly less base-dependent than looking at the number of fingers of m-samples of people from those cities.

As will be seen in the next theorem, scale- or base-unbiasedness of an r.p.m. imply that sequence of P-random samples are Benford a.s. A crucial point in the definition of an r.p.m. P with scale- or base-unbiased significant digits is that it does not require individual realizations of P to have scale- or base-invariant significant digits. In fact, it is often the case (see Benford's original data in [Ben] and Example 6.22 below) that a.s. *none* of the random probabilities has either of these properties, and it is only on average that the sampling process does not favor one scale or base over another. Recall from the notation introduced above that $S_*P(\{1\}) = 0$ is the event $\{\omega \in \Omega : P(\omega)(S = 1) = 0\}$.

Theorem 6.20 ([Hi2])**.** *Let P be an r.p.m. Assume P either has scale-unbiased significant digits, or else has base-unbiased significant digits and $S_*P(\{1\}) = 0$ with probability one. Then, for every $m \in \mathbb{N}$, every sequence (X_n) of P-random m-samples is Benford with probability one, that is, for all $t \in [1, 10)$,*

$$\frac{\#\{1 \le n \le N : S(X_n) < t\}}{N} \overset{a.s.}{\to} \log t \quad \text{as } N \to \infty\,.$$

Proof. Assume first that P has scale-unbiased significant digits, i.e., the probability measure $\mathbb{E}P$ has scale-invariant significant digits. According to Theorem 4.20, $\mathbb{E}P$ is Benford. Consequently, Lemma 6.17 implies that for every sequence (X_n) of P-random m-samples and every $t \in [1, 10)$,

$$\begin{aligned}\frac{\#\{1 \le n \le N : S(X_n) < t\}}{N} &= \frac{\#\left\{1 \le n \le N : X_n \in \bigcup_{k\in\mathbb{Z}} 10^k\big((-t, -1] \cup [1, t)\big)\right\}}{N} \\ &\overset{a.s.}{\to} \mathbb{E}P\left(\bigcup\nolimits_{k\in\mathbb{Z}} 10^k\big((-t, -1] \cup [1, t)\big)\right) = \log t \quad \text{as } N \to \infty\,.\end{aligned}$$

Assume in turn that $S_*P(\{1\}) = 0$ with probability one, and that P has base-unbiased significant digits. Then

$$S_*\mathbb{E}P(\{1\}) = \mathbb{E}P\big(S^{-1}(\{1\})\big) = \int_\Omega S_*P(\omega)(\{1\})\,\mathrm{d}\mathbb{P}(\omega) = 0\,.$$

Hence $q = 0$ holds in (4.8) with P replaced by $\mathbb{E}P$, proving that $\mathbb{E}P$ is Benford, and the remaining argument is the same as before. $\square$

Corollary 6.21. *If an r.p.m. P has scale-unbiased significant digits, then for every $m \in \mathbb{N}$, every sequence (X_n) of P-random m-samples, and every $d \in \{1, 2, \ldots, 9\}$,*

$$\frac{\#\{1 \le n \le N : D_1(X_n) = d\}}{N} \overset{a.s.}{\to} \log(1 + d^{-1}) \quad \text{as } N \to \infty .$$

A main point of Theorem 6.20 is that there are many natural sampling procedures that lead to the same logarithmic distribution. This helps explain how the different empirical evidence of Newcomb, Benford, Knuth and Nigrini all led to the same law. It may also help explain why sampling the numbers from newspaper front pages or almanacs [Ben], or accumulating extensive accounting data [Ni], often tends toward BL, since in each of these cases various distributions are being sampled in a presumably unbiased way. In a newspaper, perhaps the first article contains statistics about population growth, the second article about stock prices, the third about forest acreage. None of these individual distributions itself may be unbiased, but the mixture may well be.

Justification of the hypothesis of scale- or base-unbiasedness of significant digits in practice is akin to justification of the hypothesis of independence (and identical distribution) when applying the Strong Law of Large Numbers or the Central Limit Theorem to real-life processes: Neither hypothesis can be formally proved, yet in many real-life sampling procedures, they appear to be reasonable assumptions.

Many standard constructions of r.p.m. automatically have scale- and base-unbiased significant digits, and thus satisfy BL in the sense of Theorem 6.20.

Example 6.22. Recall the classical Dubins–Freedman construction of an r.p.m. P described in Example 6.11. It follows from [DF, Lem.9.28] that $\mathbb{E}P$ is Benford. Hence P has scale- and base-unbiased significant digits. Note, however, that with probability one P will *not* have scale- or base-invariant significant digits. It is only *on average* that these properties hold but, as demonstrated by Theorem 6.20, this is enough to guarantee that random sampling from P will generate Benford sequences a.s.

In the Dubins–Freedman construction, the fact that $F_P(10^{1/2})$, $F_P(10^{1/4})$, $F_P(10^{3/4})$, etc. are chosen *uniformly* from the appropriate intervals is not crucial: If Q is any probability measure on $(0, 1)$, and the values of $F_P(10^{1/2})$ etc. are chosen independently according to an appropriately scaled version on Q, then, for the r.p.m. thus generated, $\mathbb{E}P$ will still be Benford, provided that Q is *symmetric* about $\frac{1}{2}$, see [DF, Thm.9.29]. As a matter of fact, real-world processes often exhibit this symmetry in a natural way: Many data may be equally well recorded using certain units or their reciprocals, e.g. in miles per gallon vs. gallons per mile, or Benford's "candles per watt" vs. "watts per candle". This suggests that the distribution of $\log S$ should be symmetric about $\frac{1}{2}$.

Data having scale- or base-unbiased significant digits may be produced in many ways other than through random samples. If such data are combined with unbiased random m-samples then the result will again conform to BL in the sense of Theorem 6.20. (Presumably, this is what Benford did when combining mathematical tables with data from newspaper statistics.) For example, consider the sequence (2^n) which may be thought of as the result of a periodic sampling from a (deterministic) geometric process. As (2^n) is Benford, any mixture of this sequence with a sequence of unbiased random m-samples will again be Benford.

Finally, it is important to note that many r.p.m. and sampling processes do not conform to BL, and hence necessarily are scale- and base-biased.

Example 6.23. **(i)** Let P be the constant r.p.m. $P \equiv \delta_1$. Since $\mathbb{E}P = \delta_1$ has base-invariant significant digits, P has base-unbiased significant digits. Nevertheless, for every sequence (X_n) of P-random m-samples, the sequence of first significant digits is constant, namely $D_1(X_n) \equiv 1$.

Similarly, if $P = \lambda_{0,1}$ with probability one, then $\mathbb{E}P = \lambda_{0,1}$ does not have scale- or base-invariant significant digits. Consequently, every sequence of P-random m-samples is an i.i.d. $U(0,1)$-sequence and hence not Benford, by Example 3.10(i).

(ii) The r.p.m. considered in Example 6.10 do not have scale- or base-unbiased significant digits, simply because $\mathbb{E}P$ is not Benford.

(iii) As a another variant of the classical construction in [DF], consider the following way of generating an r.p.m. on $[1, 10)$: First let $X_{1/2}$ be uniformly distributed on $[1, 10)$ and set $F_P(X_{1/2}) = \frac{1}{2}$. Next let $X_{1/4}$ and $X_{3/4}$ be independent and uniformly distributed on $[1, X_{1/2})$ and $[X_{1/2}, 10)$, respectively, and set $F_P(X_{1/4}) = \frac{1}{4}$ and $F_P(X_{3/4}) = \frac{3}{4}$, etc. It follows from [DF, Thm.9.21] that

$$F_{\mathbb{E}P}(t) = \frac{2}{\pi} \arcsin \log t\,, \quad 1 \le t < 10\,,$$

and hence $\mathbb{E}P$ is not Benford. The r.p.m. P thus constructed, therefore, has scale- and base-biased significant digits.

6.3. Random maps

The purpose of this brief concluding section is to illustrate and prove one simple basic theorem that combines the deterministic aspects of BL studied in Chapter 5 with the stochastic considerations of the present chapter. Specifically, it is shown how applying randomly selected maps successively may generate Benford sequences with probability one. Random maps constitute a wide and intensely studied field, and for stronger results than the one discussed here the interested reader is referred e.g. to [Ber3].

For a simple example, first consider the map $T : \mathbb{R} \to \mathbb{R}$ with $T(x) = \sqrt{|x|}$. Since $T^n(x) = |x|^{2^{-n}} \to 1$ as $n \to \infty$ whenever $x \neq 0$, the orbit $O_T(x_0)$ is not Benford for any x_0. More generally, consider the randomized map

$$T(x) = \begin{cases} \sqrt{|x|} & \text{with probability } p\,, \\ x^3 & \text{with probability } 1-p\,, \end{cases} \tag{6.15}$$

and assume that, at each step, the iteration of T is independent of the entire past process. If $p = 1$, this is simply the map studied before, and hence for every $x_0 \in \mathbb{R}$, the orbit $O_T(x_0)$ is not Benford. On the other hand, if $p = 0$ then Theorem 5.12 implies that, for almost every $x_0 \in \mathbb{R}$, $O_T(x_0)$ is Benford. It is plausible to expect that the latter situation persists for small $p > 0$. As the following theorem shows, this is indeed that case even when the non-Benford map $\sqrt{|x|}$ occurs more than half of the time: If

$$p < \frac{\log 3}{\log 2 + \log 3} = 0.6131\ldots , \tag{6.16}$$

then, for a.e. $x_0 \in \mathbb{R}$, the (random) orbit $O_T(x_0)$ is Benford with probability one. To concisely formulate this result, recall that for any (deterministic or random) sequence (T_n) of maps mapping $\mathbb{R}$ or parts thereof into itself, the orbit $O_T(x_0)$ of $x_0 \in \mathbb{R}$ simply denotes the sequence $\big(T_{n-1} \circ \ldots \circ T_1(x_0)\big)_{n\in\mathbb{N}}$.

Theorem 6.24 ([Ber3])**.** *Let (β_n) be an i.i.d. sequence of positive random variables, and assume that $\log \beta_1$ has finite variance, i.e. $\mathbb{E}(\log\beta_1)^2 < +\infty$. For the sequence (T_n) of random maps given by $T_n : x \mapsto x^{\beta_n}$ and a.e. $x_0 \in \mathbb{R}$, the orbit $O_T(x_0)$ is Benford with probability one or zero, depending on whether $\mathbb{E}\log\beta_1 > 0$ or $\mathbb{E}\log\beta_1 \le 0$.*

Proof. For every $x \in \mathbb{R}$ and $n \in \mathbb{N}$,

$$\log\big(T_n \circ \ldots \circ T_1(x)\big) = \Big(\prod\nolimits_{j=1}^{n} \beta_j\Big) \log|x| = 10^{B_n} \log|x| ,$$

where $B_n = \sum_{j=1}^{n} \log\beta_j$. Assume first that $\mathbb{E}\log\beta_1 > 0$. In this case, $\frac{B_n}{n} \overset{a.s.}{\to} \log\beta_1$ as $n \to \infty$, and it can be deduced from [KN, Thm.4.2] that, with probability one, the sequence $(10^{B_n} y)$ is u.d. for a.e. $y \in \mathbb{R}$. Since $x \mapsto \log|x|$ maps the family of (Lebesgue) nullsets into itself, with probability one $O_T(x_0)$ is Benford for a.e. $x_0 \in \mathbb{R}$. More formally, with $(\Omega, \mathcal{A}, \mathbb{P})$ denoting the underlying probability space, there exists $\Omega_1 \in \mathcal{A}$ with $\mathbb{P}(\Omega_1) = 1$ such that for every $\omega \in \Omega_1$ the sequence $O_T(x_0)$ is Benford for all $x_0 \in \mathbb{R}\backslash B_\omega$, where $B_\omega \in \mathcal{B}$ with $\lambda(B_\omega) = 0$. Denote by $N \subset \mathbb{R} \times \Omega$ the set of all (x_0, ω) for which $O_T(x_0)$ is *not* Benford, and let

$$\begin{aligned} N_x &= \{\omega \in \Omega : (x, \omega) \in N\} , \quad x \in \mathbb{R} , \\ N^\omega &= \{x \in \mathbb{R} : (x, \omega) \in N\} , \quad \omega \in \Omega . \end{aligned}$$

Then $N_x \in \mathcal{A}$ and $N^\omega \in \mathcal{B}$ for all $x \in \mathbb{R}$ and $\omega \in \Omega$, respectively, and $\lambda(N^\omega) = 0$ for all $\omega \in \Omega_1$. By Fubini's Theorem,

$$0 = \int_\Omega \lambda(N^\omega)\, \mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{R}\times\Omega} \mathbb{1}_N \, \mathrm{d}(\lambda \times \mathbb{P}) = \int_{\mathbb{R}} \mathbb{P}(N_x)\, \mathrm{d}\lambda(x) ,$$

showing that $\mathbb{P}(N_x) = 0$ for a.e. $x \in \mathbb{R}$. Equivalently $\mathbb{P}(O_T(x_0) \text{ is Benford}) = 1$ holds for a.e. $x_0 \in \mathbb{R}$.

Next assume that $\mathbb{E}\log\beta_1 < 0$. Then $T_n \circ \ldots \circ T_1(x) \overset{a.s.}{\to} 1$ as $n \to \infty$ for every $x \neq 0$, and hence $O_T(x)$ is not Benford. (Note, however, that $(T_n \circ \ldots \circ T_1(x) - 1)$ may be Benford in this case.)

Finally, it remains to consider the case $\mathbb{E}\log\beta_1 = 0$. It follows from the Law of the Iterated Logarithm that, for every $t \in \mathbb{R}$,

$$\limsup_{N\to\infty} \frac{\#\{1 \le n \le N : B_n \le t\}}{N} \ge \frac{1}{2} \quad \text{with probability one}.$$

Clearly, this implies $\mathbb{P}(O_T(x_0) \text{ is Benford}) = 0$ for every $x_0 \in \mathbb{R}$. □

Example 6.25. **(i)** For the random map given by (6.15),

$$\mathbb{P}\left(\beta = \frac{1}{2}\right) = p = 1 - \mathbb{P}(\beta = 3),$$

and the condition $\mathbb{E}\log\beta = -p\log 2 + (1-p)\log 3 > 0$ is equivalent to (6.16). Note that $\mathbb{E}\log\beta > 0$ is not generally equivalent to the equally plausible (yet incorrect) condition $\mathbb{E}\beta > 1$. In the present example, the latter reduces to $p < \frac{4}{5}$.

(ii) Consider the sequence (T_n) of random maps $T_n : x \mapsto |x|^{10^{2n+\gamma_n}}$ where (γ_n) is an i.i.d. sequence of Cauchy random variables. Since $\mathbb{E}|\gamma_1| = +\infty$, Theorem 6.24 does not apply. However, $B_n = n(n+1) + \sum_{j=1}^n \gamma_j$, and [CT, Thm.5.22] shows that $\frac{B_n}{n^2} \overset{a.s.}{\to} 1$ as $n \to \infty$. The latter is enough to deduce from [KN, Thm.4.2] that $(10^{B_n} y)$ is u.d. mod 1 for a.e. $y \in \mathbb{R}$. The same argument as in the above proof shows that $\mathbb{P}(O_T(x_0) \text{ is Benford}) = 1$ for a.e. $x_0 \in \mathbb{R}$. Thus the conclusions of Theorem 6.24 may hold under weaker assumptions.

(iii) Statements in the spirit of Theorem 6.24 are true also for more general random maps, not just monomials [Ber3].

List of symbols

$\mathbb{N}, \mathbb{N}_0, \mathbb{Z}, \mathbb{Q}, \mathbb{R}^+, \mathbb{R}, \mathbb{C}$	set of positive integer, non-negative integer, integer, rational, positive real, real, complex numbers
(F_n)	sequence of Fibonacci numbers, $(F_n) = (1, 1, 2, 3, 5, 8, 13, \ldots)$
(p_n)	sequence of prime numbers, $(p_n) = (2, 3, 5, 7, 11, 13, 17, \ldots)$
$\lfloor x \rfloor$	largest integer not larger than $x \in \mathbb{R}$
$\langle x \rangle$	fractional part of $x \in \mathbb{R}$, i.e. $\langle x \rangle = x - \lfloor x \rfloor$
$\Re z$, $\Im z$	real, imaginary part of $z \in \mathbb{C}$
$\overline{z}$, $\lvert z \rvert$	conjugate, absolute value (modulus) of $z \in \mathbb{C}$
C^l	set of all l times continuously differentiable functions, $l \in \mathbb{N}_0$
C^∞	set of all smooth (i.e. infinitely differentiable) functions, i.e. $C^\infty = \bigcap_{l \ge 0} C^l$

S	significand function (Definition 2.3)
D_1, D_2, D_3 etc.	first, second, third etc. significant decimal digit (Definition 2.1)
$D_m^{(b)}$	m-th significant digit base b
$\log x$	base 10 logarithm of $x \in \mathbb{R}^+$
$\ln x$	natural logarithm of $x \in \mathbb{R}^+$
$\#A$	cardinality (number of elements) of the finite set A
$\mathcal{O}$	order symbol; $a_n = \mathcal{O}(b_n)$ as $n \to \infty$ provided that $\lvert a_n\rvert \le c\lvert b_n\rvert$ for some $c > 0$ and all n
$(\Omega, \mathcal{A}, \mathbb{P})$	probability space
A^c	complement of A in some ambient space Ω clear from the context, i.e. $A^c = \{\omega \in \Omega : \omega \notin A\}$
$A\backslash B$	set of elements of A not in B, i.e. $A\backslash B = A \cap B^c$
$A \Delta B$	symmetric difference of A and B, i.e. $A\Delta B = A\backslash B \cup B\backslash A$
$\sigma(f)$	σ-algebra generated by the function $f : \Omega \to \mathbb{R}$
$f_*\mathbb{P}$	probability measure on $\mathbb{R}$ induced by $\mathbb{P}$ and the measurable function $f : \Omega \to \mathbb{R}$, via $f_*\mathbb{P}(\bullet) := \mathbb{P}\big(f^{-1}(\bullet)\big)$
δ_a	Dirac probability measure concentrated at $a \in \Omega$
$\mathcal{B}$	Borel σ-algebra on $\mathbb{R}$ or parts thereof
λ	Lebesgue measure on $(\mathbb{R}, \mathcal{B})$ or parts thereof
$\mathcal{S}$	significand σ-algebra (Definition 2.7)
$\mathbb{1}_A$	indicator function of the set A
$\lambda_{a,b}$	normalized Lebesgue measure (uniform distribution) on $\big([a, b), \mathcal{B}[a, b)\big)$
i.i.d.	independent, identically distributed (sequence or family of random variables)
a.e.	(Lebesgue) almost every
a.s.	almost surely, i.e. with probability one
u.d. mod 1	uniformly distributed modulo one (Definition 4.1)
$X, Y, \ldots$	(real-valued) random variable $\Omega \to \mathbb{R}$
$\mathbb{E}X$	expected (or mean) value of the random variable X
$\operatorname{var} X$	variance of the random variable with $\mathbb{E}\lvert X\rvert < +\infty$; $\operatorname{var} X = \mathbb{E}(X - \mathbb{E}X)^2$
P	probability measure on $(\mathbb{R}, \mathcal{B})$, possibly random
P_X	distribution of the random variable X

F_P, F_X	distribution function of P, X
$\mathbb{B}$	Benford distribution on $(\mathbb{R}^+, \mathcal{S})$
$O_T(x_0)$	orbit of x_0 under the map T, possibly nonautonomous
N_f	Newton map associated with differentiable function f
$\sigma(A)$	spectrum (set of eigenvalues) of $d \times d$-matrix A
$X_n \stackrel{\mathcal{D}}{\to} X$	(X_n) converges in distribution to X
$X_n \stackrel{a.s.}{\to} X$	(X_n) converges to X almost surely
$\mathbb{E}P$	expectation of r.p.m. P (Proposition 6.15)
$\Box$	end of *Proof*
$\clubsuit$	end of *Note* and *Remark(s)*

References

[AS] Adhikari, A.K. and Sarkar, B.P. (1968), Distributions of most significant digit in certain functions whose arguments are random variables, *Sankhya-The Indian Journal of Statistics Series B* **30**, 47–58. MR0236969

[Al] Allaart, P.C. (1997), An invariant-sum characterization of Benford's law, *J. Appl. Probab.* **34**, 288–291. MR1429075

[BB] Barlow, J.L. and Bareiss, E.H. (1985), On Roundoff Error Distributions in Floating Point and Logarithmic Arithmetic, *Computing* **34**, 325–347. MR0804633

[Ben] Benford, F. (1938), The law of anomalous numbers, *Proc. Amer. Philosophical Soc.* **78**, 551–572.

[Ber1] Berger, A. (2001), *Chaos and Chance*, deGruyter, Berlin. MR1868729

[Ber2] Berger, A. (2005), Multi-dimensional dynamical systems and Benford's Law, *Discrete Contin. Dyn. Syst.* **13**, 219–237. MR2128801

[Ber3] Berger, A. (2005), Benford's Law in power-like dynamical systems, *Stoch. Dyn.* **5**, 587–607. MR2185507

[Ber4] Berger, A. (2010), Some dynamical properties of Benford sequences, to appear in *J. Difference Equ. Appl.*

[Ber5] Berger, A. (2010), Large spread does not imply Benford's law, *preprint*.

[BBH] Berger, A., Bunimovich, L. and Hill, T.P. (2005), Onde-dimensional dynamical systems and Benford's Law, *Trans. Amer. Math. Soc.* **357**, 197–219. MR2098092

[BH1] Berger, A. and Hill, T.P. (2007), Newton's method obeys Benford's law, *Amer. Math. Monthly* **114**, 588–601. MR2341322

[BH2] Berger, A. and Hill, T.P. (2009), *Benford Online Bibliography*; accessed May 15, 2011 at `http://www.benfordonline.net`.

[BH3] Berger, A. and Hill, T.P. (2011), Benford's Law strikes back: No simple explanation in sight for mathematical gem, *Math. Intelligencer* **33**, 85–91.

[BHKR] Berger, A., Hill, T.P., Kaynar, B. and Ridder, A. (2011), Finite-state Markov Chains Obey Benford's Law, to appear in *SIAM J. Matrix Analysis.*

[BS] Berger, A. and Siegmund, S. (2007), On the distribution of mantissae in nonautonomous difference equations, *J. Difference Equ. Appl.* **13**, 829–845. MR2343033

[CT] Chow, Y.S. and Teicher, H. (1997), *Probability Theory. Independence, Interchangeability, Martingales* (3rd ed.), Springer. MR1476912

[Di] Diaconis, P. (1977), The Distribution of Leading Digits and Uniform Distribution Mod 1, *Ann. Probab.* **5**, 72–81. MR0422186

[DT] Drmota, M. and Tichy, R.F. (1997), *Sequences, Discrepancies and Applications*, Springer. MR1470456

[DF] Dubins, L. and Freedman, D. (1967), Random distribution functions, *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66)* Vol. II: Contributions to Probability Theory, Part 1, 183–214, Univ. California Press, Berkeley, Calif. MR0214109

[Ei] Einsiedler, M. (2009), What is measure rigidity? *Notices Amer. Math. Soc.* **56** 600–601. MR2509063

[EL] Engel, H. and Leuenberger, C. (2003), Benford's law for exponential random variables, *Statist. Probab. Lett.* **63**, 361–365. MR1996184

[FT] Feldstein, A. and Turner, P. (1986), Overflow, Underflow, and Severe Loss of Significance in Floating-Point Addition and Subtraction, *IMA J. Numer. Anal.* **6**, 241–251. MR0967665

[Fel] Feller, W. (1966), *An Introduction to Probability Theory and Its Applications* vol 2, 2nd ed., J. Wiley, New York.

[Few] Fewster, R. (2009), A simple Explanation of Benford's Law, *Amer. Statist.* **63**(1), 20–25. MR2655700

[Fl] Flehinger, B.J. (1966), On the Probability that a Random Integer has Initial Digit A, *Amer. Math. Monthly* **73**, 1056–1061. MR0204395

[GG] Giuliano Antonioni, R. and Grekos, G. (2005), Regular sets and conditional density: an extension of Benford's law, *Colloq. Math.* **103**, 173–192. MR2197847

[Ha] Hamming, R. (1970), On the distribution of numbers, *Bell Syst. Tech. J.* **49**(8), 1609–1625. MR0267809

[Hi1] Hill, T.P. (1995), Base-Invariance Implies Benford's Law, *Proc. Amer. Math. Soc.* **123**(3), 887–895. MR1233974

[Hi2] Hill, T.P. (1995), A Statistical Derivation of the Significant-Digit Law, *Statis. Sci.* **10**(4), 354–363. MR1421567

[Ka] Kallenberg, O. (1997), *Foundations of modern probability*, Springer. MR1464694

[KH] Katok, A. and Hasselblatt, B. (1995), Introduction to the Modern Theory of Dynamical Systems, Cambridge University Press, Cambridge. MR1326374

[Kn] Knuth, D. (1997), *The Art of Computer Programming*, pp 253-264, vol. 2, 3rd ed, Addison-Wesley, Reading, MA. MR0378456

[KM] Kontorovich, A.V. and Miller, S.J. (2005), Benford's Law, Values of L-functions and the 3x+1 Problem, *Acta Arithm.* **120**(3), 269–297. MR2188844

[KN] Kupiers, L. and Niederreiter, H. (1974), *Uniform distribution of sequences*, John Wiley & Sons, New York. MR0419394

[LP] Lacey, M. and Phillip, W. (1990), A note on the almost sure central limit theorem, *Statist. Probab. Lett.* **63**, 361–365. MR1045184

[LS] Lagarias, J.C. and Soundararajan, K. (2006), Benford's law for the $3x + 1$ function, *J. London Math. Soc.* **74**, 289–303. MR2269630

[LSE] Leemis, L.M., Schmeiser, B.W. and Evans, D.L. (2000), Survival Distributions Satisfying Benford's Law, *Amer. Statist.* **54**(4), 236–241. MR1803620

[Ly] Lyons, R. (1995), Seventy years of Rajchman measures, *J. Fourier Anal. Appl.*, Kahane Special Issue, 363–377. MR1364897

[MN] Miller, S.J. and Nigrini, M.J. (2008), Order Statistics and Benford's Law, to appear in: *Int. J. Math. Math. Sci.* MR2461421

[Mo] Morrison, K.E. (2010), The Multiplication Game, *Math. Mag.* **83**, 100–110. MR2649322

[Ne] Newcomb, S. (1881), Note on the frequency of use of the different digits in natural numbers, *Amer. J. Math.* **9**, 201–205.

[Ni] Nigrini, M.J. (1992), *The Detection of Income Tax Evasion Through an Analysis of Digital Frequencies*, PhD thesis, University of Cincinnati, OH, USA.

[Pa] Palmer, K. (2000), *Shadowing in dynamical systems*, Kluwer. MR1885537

[Pi] Pinkham, R. (1961), On the Distribution of First Significant Digits, *Ann. Math. Statist.* **32**(4), 1223–1230. MR0131303

[Ra1] Raimi, R. (1976), The First Digit Problem, *Amer. Math. Monthly* **83**(7), 521–538. MR0410850

[Ra2] Raimi, R. (1985), The First Digit Phenomenon Again, *Proc. Amer. Philosophical Soc.* **129**, 211–219.

[Ro] Robbins, H. (1953), On the equidistribution of sums of independent random variables, *Proc. Amer. Math. Soc.* **4**, 786–799. MR0056869

[Scha1] Schatte, P. (1988), On random variables with logarithmic mantissa distribution relative to several bases, *Elektron. Informationsverarbeit. Kybernetik* **17**, 293–295. MR0653759

[Scha2] Schatte, P. (1988), On a law of the iterated logarithm for sums mod 1 with application to Benford's law, *Probab. Theory Related Fields* **77**, 167–178. MR0927235

[Schü1] Schürger, K. (2008), Extensions of Black-Scholes processes and Benford's law, *Stochastic Process. Appl.* **118**, 1219–1243. MR2428715

[Schü2] Schürger, K. (2011), Lévy Processes and Benford's Law, *preprint.*
[Se] Serre, J.P. (1973), *A course in arithmetic*, Springer. MR0344216
[Sh] Shiryayev, A.N. (1984), *Probability*, Springer. MR0737192
[Sm] Smith, S.W. (1997), Explaining Benford's Law, Chapter 34 in: *The Scientist and Engineer's Guide to Digital Signal Processing.* Republished in softcover by Newnes, 2002
[Wa] Walter, W. (1998), *Ordinary Differential Equations*, Springer. MR1629775
[Wh] Whitney, R.E. (1972), Initial digits for the sequence of primes, *Amer. Math. Monthly* **79**(2), 150–152. MR0304337