Chapter Two

A Short Introduction to the Mathematical Theory of Benford's Law

Arno Berger and T. P. Hill[1]

This chapter is an abbreviated version of [BerH4], which can be consulted for additional details. Many of the results presented here, notably those in Sections 2.5 and 2.6, can be strengthened considerably; the interested reader may want to consult [BerH5] in this regard.

2.1 INTRODUCTION

Benford's Law, or **BL** for short, is the observation that in many collections of numbers, be they 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. The reader may find many formulations and applications of BL in the online database [BerH2].

More specifically, BL says that the significant digits in many data sets 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

$$\text{Prob}\,(D_1 = d_1) \;=\; \log_{10}\left(1 + d_1^{-1}\right) \quad \text{for all } d_1 = 1, 2, \dots, 9\,; \tag{2.1}$$

here D_1 denotes the first significant decimal digit [Ben, New]. For example, (2.1) asserts that

$$\begin{aligned} \text{Prob}\,(D_1 = 1) \;&=\; \log_{10} 2 \;=\; 0.3010\ldots, \\ \text{Prob}\,(D_1 = 9) \;&=\; \log_{10}\frac{10}{9} \;=\; 0.04575\ldots. \end{aligned} \tag{2.2}$$

In a form more complete than (2.1), BL is a statement about the joint distribution of *all* decimal digits: For every positive integer m,

$$\begin{aligned} &\text{Prob}\,\big((D_1, D_2, \dots, D_m) \;=\; (d_1, d_2, \dots, d_m)\big) \\ &\quad=\; \log_{10}\left(1 + \left(\sum_{j=1}^{m} 10^{m-j} d_j\right)^{-1}\right) \end{aligned} \tag{2.3}$$

[1]Department of Mathematical & Statistical Sciences, University of Alberta, and School of Mathematics, Georgia Institute of Technology.

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 $j \geq 2$, d_j is an integer in $\{0, 1, \ldots, 9\}$; here D_2, D_3, D_4, etc. represent the second, third, fourth, etc. significant decimal digit. Thus, for example, (2.3) implies that

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

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 in this chapter, 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*.

2.2 SIGNIFICANT DIGITS AND THE SIGNIFICAND

Since Benford's Law is a statement about the statistical distribution of significant (decimal) digits, a natural starting point for any study of BL is the formal definition of *significant digits* and the *significand* (*function*).

2.2.1 Significant Digits

Definition 2.2.1 (First significant decimal digit)**.** *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* $10^k j \leq |x| < 10^k (j+1)$ *for some (necessarily unique)* $k \in \mathbb{Z}$.

Similarly, for every $m \geq 2$, $m \in \mathbb{N}$*, 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*

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

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 2.2.2. *Since* $\sqrt{2} \approx 1.414$ *and* $1/\pi \approx 0.3183$*,*

$$D_1(\sqrt{2}) \;=\; D_1(-\sqrt{2}) \;=\; D_1(10\sqrt{2}) \;=\; 1\,, \quad D_2(\sqrt{2}) \;=\; 4\,, \quad D_3(\sqrt{2}) \;=\; 1\,;$$

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

2.2.2 The Significand

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

Definition 2.2.3. *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,* $S(0) := 0$.

Example 2.2.4.

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

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

The significand uniquely determines the significant digits, and vice versa. This relationship is recorded in the next proposition which immediately follows from Definitions 2.2.1 and 2.2.3. Here and throughout the **floor function**, $\lfloor x \rfloor$, denotes the largest integer not larger than x.

Proposition 2.2.5. *For every real number* x*,*

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

Since the significant digits determine the significand, and are in turn determined by it, the informal version (2.3) of BL in the introduction has an immediate and very concise counterpart in terms of the significand function, namely

$$\text{Prob}\,(S \leq t) \;=\; \log t \quad \text{for all } 1 \leq t < 10\,. \tag{2.4}$$

2.2.3 The Significand σ-Algebra

The informal statements (2.1), (2.3), and (2.4) of BL involve *probabilities*. 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 is both intuitive and easy to describe.

Definition 2.2.6. *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, where $\sigma(f)$ denotes the σ-algebra generated by f, i.e., the smallest σ-algebra containing all sets of the form $\{x : a \leq f(x) \leq b\}$, and $\mathcal{B}(I)$ denotes the real **Borel σ-algebra** restricted to an interval I. If $I = \mathbb{R}$ or $I = \mathbb{R}^+ = \{t \in \mathbb{R} : t > 0\}$ then, for convenience, instead of $\mathcal{B}(I)$ simply write $\mathcal{B}$ and $\mathcal{B}^+$, respectively. Also, here and throughout, for every set $C \subset \mathbb{R}$ and $t \in \mathbb{R}$, let $tC := \{tc : c \in C\}$.

Lemma 2.2.7. *For every function* $f : \mathbb{R}^+ \to \mathbb{R}$ *the following statements are equivalent:*

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

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

Proof. Routine. □

Theorem 2.2.8 ([Hi4])**.** *For every $A \in \mathcal{S}$,*

$$A = \bigcup_{k\in\mathbb{Z}} 10^k S(A) \tag{2.5}$$

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

$$\mathcal{S} = \mathbb{R}^+ \cap \sigma(D_1, D_2, D_3, \ldots) = \left\{ \bigcup_{k\in\mathbb{Z}} 10^k B : B \in \mathcal{B}[1, 10) \right\}. \tag{2.6}$$

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 (2.5) holds for all $A \in \mathcal{S}$.

To prove (2.6), first observe that by Proposition 2.2.5(1) 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 2.2.5(2), 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 (2.6), 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 (2.5). 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 $\mathcal{S}$. □

Note that for every $A \in \mathcal{S}$ there is a *unique* $B \in \mathcal{B}[1, 10)$, the Borel subsets of $[1, 10)$, such that $A = \bigcup_{k\in\mathbb{Z}} 10^k B$, and (2.5) shows that in fact $B = S(A)$.

Example 2.2.9. *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 (2.6) and the fact that $A_4 = \bigcup_{k\in\mathbb{Z}} 10^k\{1\}$ and $\{1\} \in \mathcal{B}[1, 10)$.

Example 2.2.10. *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 $B \in \mathcal{B}[1, 10)$.*

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

1. *$\mathcal{S}$ is self-similar with respect to multiplication by integer powers of* 10*, i.e.,*
$$10^k A \;=\; A \quad \text{for every } A \in \mathcal{S} \text{ and } k \in \mathbb{Z}\,.$$
2. *$\mathcal{S}$ is closed under multiplication by a scalar, i.e.,*
$$\alpha A \in \mathcal{S} \quad \text{for every } A \in \mathcal{S} \text{ and } \alpha > 0\,.$$
3. *$\mathcal{S}$ is closed under integral roots, i.e.,*
$$A^{1/n} \in \mathcal{S} \quad \text{for every } A \in \mathcal{S} \text{ and } n \in \mathbb{N}\,.$$

Proof. (1) This is obvious from (2.5) since $S(10^k A) = S(A)$ for every k.

(2) Given $A \in \mathcal{S}$, by (2.6) there exists $B \in \mathcal{B}[1,10)$ such that $A = \bigcup_{k\in\mathbb{Z}} 10^k B$. In view of (1), assume without loss of generality that $1 < \alpha < 10$. Then

$$\begin{aligned}
\alpha A \;&=\; \bigcup_{k\in\mathbb{Z}} 10^k \alpha B \\
&=\; \bigcup_{k\in\mathbb{Z}} 10^k \left(\big(\alpha B \cap [\alpha, 10)\big) \cup \big(\frac{\alpha}{10} B \cap [1,\alpha)\big) \right) \\
&=\; \bigcup_{k\in\mathbb{Z}} 10^k C,
\end{aligned}$$

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 $\alpha A \in \mathcal{S}$.

(3) 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

$$\begin{aligned}
A^{1/n} \;&=\; \bigcup_{k\in\mathbb{Z}} 10^{k/n} [1, 10^{s/n}] \\
&=\; \bigcup_{k\in\mathbb{Z}} 10^k \bigcup_{j=0}^{n-1} [10^{j/n}, 10^{(j+s)/n}] \\
&=\; \bigcup_{k\in\mathbb{Z}} 10^k C,
\end{aligned}$$

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

Since, by Theorem 2.2.8, the significand σ-algebra $\mathcal{S}$ is the same as the significant digit σ-algebra $\sigma(D_1, D_2, D_3, \cdots)$, the closure properties established in Lemma 2.2.11 carry over to sets determined by significant digits. The next example illustrates closure under multiplication by a scalar and integral roots, and that $\mathcal{S}$ is *not* closed under taking integer powers.

Example 2.2.12. *Let A_5 be the set of positive real numbers with first significant digit* 1*, i.e.,*

$$A_5 \;=\; \{x > 0 : D_1(x) = 1\} \;=\; \{x > 0 : 1 \le S(x) < 2\} \;=\; \bigcup_{k\in\mathbb{Z}} 10^k [1,2)\,.$$

Then

$$\begin{aligned} 2A_5 &= \big\{x > 0 : D_1(x) \in \{2,3\}\big\} = \{x > 0 : 2 \le S(x) < 3\} \\ &= \bigcup_{k\in\mathbb{Z}} 10^k[2,4) \in \mathcal{S}, \end{aligned} \tag{2.7}$$

and also

$$\begin{aligned} A_5^{1/2} &= \big\{x > 0 : S(x) \in [1,\sqrt{2}) \cup [\sqrt{10},\sqrt{20})\big\} \\ &= \bigcup_{k\in\mathbb{Z}} 10^k\big([1,\sqrt{2}) \cup [\sqrt{10}, 2\sqrt{5})\big) \in \mathcal{S}, \end{aligned}$$

whereas on the other hand clearly

$$A_5^2 = \bigcup_{k\in\mathbb{Z}} 10^{2k}[1,4) \notin \mathcal{S},$$

since e.g. $[1,4) \subset A_5^2$ *but* $[10,40) \not\subset A_5^2$.

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 $\big([0,1), \mathcal{B}[0,1)\big)$.

Lemma 2.2.13. *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* $\big([0,1), \mathcal{B}[0,1)\big)$.

Proof. Routine. □

2.3 THE BENFORD PROPERTY

In order to translate the informal versions (2.1), (2.3), and (2.4) 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 section, the objects of interest fall into three categories: sequences of real numbers; real-valued functions defined on $[0,+\infty)$; and probability distributions associated with random variables. Accordingly, denote by $\#M$ the cardinality of a finite set M, and let λ symbolize Lebesgue measure on $(\mathbb{R}, \mathcal{B})$ (or parts thereof).

2.3.1 Benford Sequences

Definition 2.3.1. *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* $d_1 \in \{1, 2, \cdots, 9\}$ *and all* $d_j \in \{0, 1, \cdots, 9\}$*,* $j \geq 2$*,*

$$\lim_{N\to\infty} \frac{\#\{1 \leq n \leq N : D_j(x_n) \ = \ d_j \text{ for } j = 1, 2, \ldots\}}{N} = \log\left(1 + \left(\sum_{j=1}^{m} 10^{m-j} d_j\right)^{-1}\right).$$

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. As will be seen in Example 2.4.12, the sequence $(F_n) = (1, 1, 2, 3, 5, 8, 13, \ldots)$ of Fibonacci numbers, where every entry is simply the sum of its two predecessors, and $F_1 = F_2 = 1$, is Benford. In Example 2.4.11(v), it will be shown that the sequence $(p_n) = (2, 3, 5, 7, 11, 13, 17, \ldots)$ of prime numbers is *not* Benford.

2.3.2 Benford Functions

BL also appears frequently in real-valued functions such as those arising as solutions of initial value problems for differential equations (see Section 2.5.3 below). Thus, the starting point is to define what it means for a function to follow BL.

Definition 2.3.2. *A (Borel measurable) function* $f : [0, +\infty) \to \mathbb{R}$ *is* ***Benford*** *if*

$$\lim_{T\to+\infty} \frac{\lambda\big(\{\tau \in [0, T) : S\big(f(\tau)\big) \leq t\}\big)}{T} = \log t \quad \text{for all } t \in [1, 10),$$

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

$$\lim_{T\to+\infty} \frac{\lambda\big(\{\tau \in [0, T) : D_j\big(f(\tau)\big) \ = \ d_j \text{ for } j = 1, 2, \ldots, m\}\big)}{T} = \log\left(1 + \left(\sum_{j=1}^{m} 10^{m-j} d_j\right)^{-1}\right).$$

As will be seen 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.

2.3.3 Benford Distributions and Random Variables

This section lays the foundations for analyzing the Benford property for probability distributions and random variables.

Definition 2.3.3. *A Borel probability measure* P *on* $\mathbb{R}$ *is* ***Benford*** *if*

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

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

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

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

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

Example 2.3.4. *If X is a Benford random variable on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, then*

$$\mathbb{P}(D_1(X) = 1) \;=\; \mathbb{P}(1 \le S(X) < 2) \;=\; \log 2 \;=\; 0.3010\dots\,,$$

$$\mathbb{P}(D_1(X) = 9) \;=\; \log\frac{10}{9} \;=\; 0.04575\dots\,,$$

$$\mathbb{P}\big((D_1(X), D_2(X), D_3(X)) \;=\; (3, 1, 4)\big) \;=\; \log\frac{315}{314} \;=\; 0.001380\dots\,.$$

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

Example 2.3.5. *For every integer k, the probability measure P_k with density $f_k(x) = 1/(x \ln 10)$ on $[10^k, 10^{k+1})$ is Benford, and so is $\frac{1}{2}(P_k + P_{k+1})$. 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 2.6.4 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 2.3.6 (Benford distribution). *The* Benford distribution $\mathbb{B}$ *is the unique probability measure on $(\mathbb{R}^+, \mathcal{S})$ with*

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

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

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

The combination of Definitions 2.3.3 and 2.3.6 gives

Proposition 2.3.7. *A Borel probability measure P on $\mathbb{R}^+$ is Benford if and only if*

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

(For probability measures on all of $\mathbb{R}$, an analogous result holds via Definition 2.3.3; cf. [BerH4].)

Example 2.3.8. **(i)** *If X is distributed according to $U(0,1)$, the uniform distribution on $[0,1)$, then for every $1 \le t < 10$,*

$$\mathbb{P}\big(S(X) \le t\big) \;=\; \sum_{n\in\mathbb{N}} 10^{-n}(t-1) \;=\; \frac{t-1}{9} \not\equiv \log t\,,$$

showing that $S(X)$ is uniform on $[1,10)$, and hence 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 $F_{\texttt{exp}(1)}(t) = \mathbb{P}(\texttt{exp}(1) \le t) = \max(0, 1-e^{-t})$, then*

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

and hence $\texttt{exp}(1)$ is not Benford either. (See [EngLeu, MiNi2] for a detailed analysis of the exponential distribution's relation to BL.)

2.4 CHARACTERIZATIONS OF BENFORD'S LAW

The purpose of this section 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 data sets are, or are not, Benford, and helpful for predicting which empirical data are likely to follow BL closely.

2.4.1 The Uniform Distribution Characterization

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 $\lambda_{a,b}$, for any $a<b$, denotes (normalized) Lebesgue measure on $\big([a,b), \mathcal{B}[a,b)\big)$.

Definition 2.4.1. *A sequence (x_n) of real numbers is* ***uniformly distributed modulo one****, abbreviated henceforth as* ***u.d.*** $\mathrm{mod} 1$*, 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.*** $\mathrm{mod} 1$ *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.*** $\mathrm{mod} 1$ *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.*** $\bmod 1$ *if*

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

The next simple theorem (cf. [Dia, MiT-B]) is one of the main tools in the theory of BL because it allows application of the powerful theory of uniform distribution mod 1, as developed e.g. in [KuiNi]. (Recall the convention $\log 0 := 0$.)

Theorem 2.4.2 (Uniform distribution characterization)**.** *A sequence of real numbers (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 $s \in [0,1)$,

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

Hence, by Definitions 2.3.3 and 2.4.1, 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 2.4.2, in establishing the Benford property for many sequences, functions, and random variables; for proofs, see [BerH4].

Lemma 2.4.3.

1. *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 non-zero integer* k *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* $\lim_{n\to\infty} |y_n - x_n| = 0$.
2. *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* $b \in \mathbb{R}$.
3. *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* $b \in \mathbb{R}$.

Example 2.4.4. **(i)** *The sequence* $(n\pi) = (\pi, 2\pi, 3\pi, \ldots)$ *is* u.d. mod 1*, by Weyl's Equidistribution Theorem; see Proposition 2.4.8(1) 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 2.4.3(i) that* k *be an integer cannot be removed.*

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

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

as its limit inferior and limit superior, respectively.

Example 2.4.5. **(i)** *The function* $f(t) = at + b$ *with real* a, b *is u.d.* mod 1 *if and only if* $a \neq 0$. *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* $\alpha \neq 0$, *via Theorem 2.4.2, since* $\log f(t) = \alpha t / \ln 10$ *is u.d.* mod 1.

(ii) *The function* $f(t) = \log|at+b|$ *is not u.d.* mod 1 *for any* $a, b \in \mathbb{R}$. *Similarly,* $f(t) = -\log(1+t^2)$ *is not u.d.* mod 1, *and hence* $f(t) = (1+t^2)^{-1}$ *is not Benford.*

(iii) *The function* $f(t) = e^t$ *is u.d.* mod 1. *As a consequence, the* ***superexponential function*** $f(t) = e^{e^{\alpha t}}$ *is also Benford if* $\alpha \neq 0$.

Example 2.4.6. **(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 *(cf. [BerH3, BerH4, LeScEv, MiNi2]).*

(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, it is 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 [Pin], it can be shown that*

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

where $F_{\langle X \rangle}(s) = \mathbb{P}(\langle X \rangle \le s)$. *In particular* $\Delta(\sigma) = (e^{-2\sigma^2\pi^2})/\pi + \mathcal{O}(e^{-8\sigma^2\pi^2})$ *as* σ *tends to infinity, showing that* $\Delta(\sigma)$, *the deviation of* $\langle X \rangle$ *from uniformity, goes to zero very rapidly as* $\sigma \to +\infty$. *Already for* $\sigma = 1$ *one finds that* $\Delta(1) < 8.516 \cdot 10^{-10}$. *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 2.4.7.

1. *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.*
2. *A function* $f : [0, +\infty) \to \mathbb{R}$ *is Benford if and only if* $1/f$ *is Benford.*
3. *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 2.4.2 will be used to determine Benford properties of sequences, functions, and random variables. Conclusion (1) in Proposition 2.4.8 is Weyl's classical uniform distribution result [KuiNi, Thm.3.3], conclusion (2) is an immediate consequence of Weyl's criterion [KuiNi, Thm.2.1], conclusion (3) is [Ber2, Lem.2.8], and conclusion (4) is [BerBH, Lem.2.4.(i)].

Proposition 2.4.8. *Let (x_n) be a sequence of real numbers.*

1. *If $\lim_{n\to\infty}(x_{n+1} - x_n) = \theta$ for some irrational θ, then (x_n) is u.d.* mod 1.
2. *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.*
3. *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 $\alpha \in \mathbb{R}$.*
4. *If (x_n) is u.d.* mod 1 *and non-decreasing, then $(x_n / \log n)$ is unbounded.*

Another very useful result is **Koksma's metric theorem** [KuiNi, 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.*), which means that with probability one for every random variable that has a density (i.e., is absolutely continuous).

Proposition 2.4.9. *Let f_n be continuously differentiable on $[a, b]$ for all $n \in \mathbb{N}$. If $f'_m - f'_n$ is monotone and $|f'_m(x) - f'_n(x)| \geq \alpha > 0$ for all $m \neq n$, 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 2.4.10 ([BerHKR])**.** *If a, b, α, β are real numbers with $a \neq 0$ and $|\alpha| > |\beta|$ then $(\alpha^n a + \beta^n b)$ is Benford if and only if $\log |\alpha|$ is irrational.*

Proof. Since $a \neq 0$ and $|\alpha| > |\beta|$, $\lim_{n\to\infty} \dfrac{\beta^n b}{\alpha^n a} = 0$, 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 2.4.8(1), 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 2.4.2 therefore completes the proof. □

Example 2.4.11. **(i)** *By Theorem 2.4.10 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 2.4.10.*

(iii) *By Proposition 2.4.9, 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 2.4.2, (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 2.4.8(4) or Example 2.4.4(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* $\alpha \in \mathbb{R}$.

(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 2.4.8(4) that* (p_n) *is not Benford.*

Example 2.4.12. *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*

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

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 2.4.10. Sequences such as* (F_n) *which are generated by linear recurrence relations will be studied in detail in Section 2.5.2.*

Theorem 2.4.13. *Let* X, Y *be random variables. Then*

1. *if* X *is u.d.* mod 1 *and* Y *is independent of* X*, then* $X + Y$ *is u.d.* mod 1*;*

2. *if* $\langle X \rangle$ *and* $\langle X + \alpha \rangle$ *have the same distribution for some irrational* α *then* X *is u.d.* mod 1*;*

3. *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*

$$\lim_{n\to\infty} \mathbb{P}\left(\left\langle \sum_{j=1}^{n} X_j \right\rangle \le s\right) = s \quad \text{for every } 0 \le s < 1\,, \tag{2.8}$$

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

Proof. Elementary Fourier analysis; see [BerH4, Thm.4.13]. □

None of the familiar classical probability distributions or random variables, such as 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 where it is centered. This statement can be quantified explicitly as follows.

Proposition 2.4.14 ([BerH3])**.** *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 [BerH3]. However, some distributions, such as the exponential distribution with mean 1, and the standard normal distribution, do come fairly close to being Benford.

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 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. 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 [BerH3].

Theorem 2.4.15. *If X has a density then*

$$\lim_{n\to\infty} \mathbb{P}(\langle nX \rangle \le s) \;=\; s \quad \text{for all } 0 \le s < 1\,, \tag{2.9}$$

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

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

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

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

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, for any $\varepsilon > 0$ there exists a continuous density g_ε with $\int_0^1 |f(\sigma) - g_\varepsilon(\sigma)|\,\mathrm{d}\sigma$

$< \varepsilon$ and hence

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

which in turn shows that

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

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 $0 \le s < 1$,

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

□

2.4.2 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 dollars to euros, etc., then while the individual numbers change, the statements about the overall distribution of significant digits should not be affected by this change.

While a positive random variable X cannot be scale invariant, it may nevertheless have *scale-invariant significant digits*. For this, however, X has to be Benford. In fact, Theorem 2.4.18 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. For every function $f : \Omega \to \mathbb{R}$ with $\mathcal{A} \supset \sigma(f)$ and every probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$, let $f_*\mathbb{P}$ denote the probability measure on $(\mathbb{R}, \mathcal{B})$ defined according to

$$f_*\mathbb{P}(B) \;=\; \mathbb{P}\big(f^{-1}(B)\big) \quad \text{for all } B \in \mathcal{B}\,. \tag{2.10}$$

Definition 2.4.16. *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* $d_1 \in \{1, 2, \dots, 9\}$ *and all* $d_j \in \{0, 1, \dots, 9\}$*,* $j \geq 2$*,*

$$P\big(\{x : D_j(\alpha x) \ = \ d_j \text{ for } j = 1, 2, \dots, m\}\big) \\ = \ P\big(\{x : D_j(x) = d_j \text{ for } j = 1, 2, \dots, m\}\big) \qquad (2.11)$$

holds for every $\alpha > 0$*.*

Example 2.4.17. **(i)** *The Benford probability measure* $\mathbb{B}$ *on* $(\mathbb{R}^+, \mathcal{S})$ *has scale-invariant significant digits. This follows from Theorem 2.4.18 below.*

(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* $\delta_1(D_1 = 1) = 1 \neq 0 = \delta_2(D_1 = 1)$.

(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 2.4.18 (Scale-invariance characterization [Hi3])**.** *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 2.2.13. According to Lemma 2.2.13, Q is a probability measure on $\big([0, 1), \mathcal{B}[0, 1)\big)$. Moreover, under the correspondence established by ℓ,

$$P_0(\alpha A) \ = \ P_0(A) \quad \text{for all } \alpha > 0, A \in \mathcal{S} \qquad (2.12)$$

is equivalent to

$$Q(\langle t + B\rangle) \ = \ Q(B) \quad \text{for all } t \in \mathbb{R}, B \in \mathcal{B}[0, 1)\,, \qquad (2.13)$$

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, (2.13) simply means that, for every $t \in \mathbb{R}$, the distributions of $\langle X\rangle$ and $\langle t + X\rangle$ coincide. By Theorem 2.4.13(1) and (2) 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 random variable.) Hence (2.12) is equivalent to $P_0 = (\ell^{-1})_* \lambda_{0,1} = \mathbb{B}$. □

The next example is an elegant and entertaining application of the ideas underlying Theorem 2.4.18 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.

Example 2.4.19 ([Morr])**.** *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 2.4.13(1) and Example 2.4.17(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* $\log 4$.

Theorem 2.4.18 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 2.4.20. *For every random variable* X *with* $\mathbb{P}(X = 0) = 0$ *the following statements are equivalent:*

1. X *is Benford.*
2. *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, (2) implies that $\mathbb{P}(D_1(X) = d) = \log(1 + d^{-1})$.

Example 2.4.21 ("Ones-scaling test" [Sm])**.** *In view of the last theorem, 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 rescaled, i.e., multiplied by* $\alpha, \alpha^2, \alpha^3, \ldots$*, where* $\log \alpha$ *is irrational, e.g.* $\alpha = 2$.

2.4.3 The Base-Invariance Characterization

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 $a = \log 2$,

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

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 2.2.11(3) 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 2.4.22. *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 2.4.23. **(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 $\delta_2(A^{1/2}) = 0$.*

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

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

Then, as seen in the proof of Lemma 2.2.11(3),

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

and therefore

$$\mathbb{B}(A^{1/n}) = \sum_{j=0}^{n-1} \left(\log 10^{(j+s)/n} - \log 10^{j/n} \right) = \sum_{j=0}^{n-1} \left(\frac{j+s}{n} - \frac{j}{n} \right) = s = \mathbb{B}(A) .$$

(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*

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

The next theorem is the main result for base-invariant significant digits.

Theorem 2.4.24 (Base-invariance characterization [Hi3]**).** *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 $q \in [0,1]$,*

$$P(A) \;=\; q\delta_1(A) + (1-q)\mathbb{B}(A) \quad \text{for every } A \in \mathcal{S}\,. \tag{2.14}$$

Corollary 2.4.25. *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 every $n \in \mathbb{N}$, 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 2.4.26. *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. 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 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 $t \in \mathbb{R}$,

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

Using this and the Dominated Convergence Theorem, it follows from

$$P(\{0\}) \;=\; \int_0^1 \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} e^{2\pi\imath s j}\,\mathrm{d}P(s) = \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\widehat{P}(j) = q\,,$$

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, $P = q\delta_0 + (1-q)\lambda_{0,1}$. □

Proof. As in the proof of Theorem 2.4.18, 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 $A = \{x > 0 : S_{10}(x) < 10^s\}$,

$$\begin{aligned} T_{n*}Q\big([0,s)\big) \;&=\; Q\left(\bigcup_{j=0}^{n-1}\left[\frac{j}{n}, \frac{j+s}{n}\right)\right) \\ &=\; P_0\left(\bigcup_{k\in\mathbb{Z}} 10^k \bigcup_{j=0}^{n-1}[10^{j/n}, 10^{(j+s)/n})\right) \;=\; P_0(A^{1/n}) \quad (2.15)\end{aligned}$$

and hence $T_{n*}Q = Q$ for all n precisely if P_0 has base-invariant significant digits. In this case, by Lemma 2.4.26, $Q = q\delta_0 + (1-q)\lambda_{0,1}$ 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 $A \in \mathcal{S}$. □

Corollary 2.4.27. *If a probability measure on $\mathbb{R}^+$ has scale-invariant significant digits then it also has base-invariant significant digits.*

2.4.4 The Sum-Invariance Characterization

As first observed by M. Nigrini [Nig1], 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. This motivates the following definition.

Definition 2.4.28. *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 $d_1, \ldots, d_m$.

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 stated.

Definition 2.4.29. *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 2.4.30. **(i)** *If X is uniformly distributed on $[0, 1)$, then X does not have sum-invariant significant digits. This follows from Theorem 2.4.31 below.*

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

$$\mathbb{E}S_d(X) = \begin{cases} 1 & \text{if } d = 1\,, \\ 0 & \text{if } d \geq 2\,. \end{cases}$$

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

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

Thus X has sum-invariant significant digits.

According to Example 2.4.30(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 2.4.31 (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. See [Al] or [BerH4, Thm.4.37]. □

2.5 BENFORD'S LAW FOR DETERMINISTIC PROCESSES

The goal of this section 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 superexponential 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, superexponential processes typically are Benford for all bases, but have small sets (of measure zero) of exceptional points whose orbits or trajectories are not Benford.

2.5.1 One-Dimensional Discrete-Time Processes

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

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

Example 2.5.1. **(i)** *If* $T(x) = 2x$ *then* $O_T(x_0) = (x_0, 2x_0, 2^2x_0, \ldots) = (2^{n-1}x_0)$ *for all* x_0*. Hence* $\lim_{n\to\infty} |x_n| = +\infty$ *whenever* $x_0 \neq 0$.

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

(iii) *If* $T(x) = 1 + x^2$ *then* $O_T(x_0) = (x_0, 1 + x_0^2, 2 + 2x_0^2 + x_0^4, \ldots)$*. 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 2.4.11(i) that (2^n) is Benford, and in fact $(2^n x_0)$ is Benford for every $x_0 \neq 0$. In other words, Example 2.5.1(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 subsection is to extend this observation to a much wider class of maps T. The main result (Theorem 2.5.5) rests upon three lemmas.

Lemma 2.5.2. *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 2.4.10, $O_T(x_0) = (a^{n-1}x_0)$ is Benford for every $x_0 \neq 0$ or none, depending on whether $\log |a|$ is irrational or not. □

Clearly, the simple proof of Lemma 2.5.2 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.

Lemma 2.5.3 (**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. See [BerBH]. □

The next lemma enables application of Lemma 2.5.3 to establish the Benford property for orbits of a wide class of maps.

Lemma 2.5.4.

1. *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.*

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

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

$$-\log\left(1+\frac{c}{|a_n|-c}\right) \leq \log\frac{|b_n|}{|b_n|+c} \leq \log\frac{|b_n|}{|a_n|} \leq \log\frac{|a_n|+c}{|a_n|} \leq \log\left(1+\frac{c}{|a_n|-c}\right),$$

it follows that

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

Lemma 2.4.3(1) now shows that $(\log|b_n|)$ is u.d. mod 1 if and only $(\log|a_n|)$ is. The proof of (2) is completely analogous. □

Lemmas 2.5.3 and 2.5.4 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 2.5.6, a reciprocal version holds for orbits converging to $\pm\infty$.

Theorem 2.5.5 ([BerBH]). *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 $\sup_{|x|\geq\xi}|\widetilde{T}(x)-\alpha^{-1}x|$ is finite, provided that ξ is sufficiently large. Hence Lemma 2.5.3 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 2.5.4 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 2.4.7(1) which shows that (x_n^{-1}) is Benford whenever (x_n) is. □

Example 2.5.6. **(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, $O_T(-4)=(-4,2,2,\ldots)$ and $O_T(2)=(2,2,2,\ldots)$ are not Benford.*

(ii) *To see that Theorem 2.5.5 applies to the map $T(x) = 2x + e^{-x}$, 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 2.5.5, $O_{\widetilde{T}}(x_0)$ 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.

Processes with Superexponential Growth or Decay

The following is an analog of Lemma 2.5.2 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 2.5.7. *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 2.4.9, 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

$$y_0 := \sum_{j=1}^{\infty} \eta_j \beta^{-j}$$

and observe that, for every $n \in \mathbb{N}$,

$$0 \leq \langle \beta^n y_0 \rangle = \sum_{j=n+1}^{\infty} \eta_j \beta^{n-j} \leq \frac{1}{\beta} + \frac{1}{\beta^2(\beta - 1)} < 1,$$

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. □

The following is an analog of Theorem 2.5.5 for the case when T is dominated by power-like terms.

Theorem 2.5.8 ([BerBH])**.** *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.*

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 2.5.7, 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 2.5.7 has shown, $(p^n y)$ is u.d. mod 1 for a.e. $y \in \mathbb{R}$. Moreover, Lemma 2.5.3 applies to R, and it can be checked by term-by-term differentiation that the shadowing map

$$h : y \mapsto \overline{y} \; = \; y - \sum_{j=1}^{\infty} p^{-j} \log\Big(1 + f\big(10^{-R^j(y)}\big)\Big)$$

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 2.5.7. □

Example 2.5.9. **(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 2.5.8 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 Theorem 2.5.8 applies to the reciprocal version $\widetilde{T}$ of T, namely*

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

near $x = 0$. Overall, therefore, $O_T(x_0)$ is Benford for a.e. $x_0 \in \mathbb{R}$.

(ii) *Let $T(x) = 1 + x^2$. Again Theorem 2.5.8 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. $x_0 \in \mathbb{R}$.

An Application: Newton's Method and Related Algorithms

In scientific calculations using digital computers and floating point arithmetic, round-off errors are inevitable, thus, for the problem of finding numerically the root of a function by means of **Newton's Method**, it is important to study the distribution of significant digits (or significands) of the approximations generated by the method.

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 Newton's Method, 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 $f'(x) = 0$.

If $f : I \to \mathbb{R}$ is real-analytic and $x^* \in I$ is 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.

Theorem 2.5.10 ([BerH1]). *Let $f : I \to \mathbb{R}$ be real-analytic with $f(x^*) = 0$, and assume that f is not linear. Then*

1. *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 neighborhood of x^*;*
2. *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^*.*

Here (x_n) denotes the sequence of iterates of N_f starting at x_0, that is, $(x_n) = O_{N_f}(x_0)$.

The full proof of Theorem 2.5.10 can be found in [BerH1]. It uses the following lemma which may be of independent interest for studying BL in other numerical approximation procedures. Part (1) is an analog of Lemma 2.5.4, and (2) and (3) follow directly from Theorems 2.5.8 and 2.5.5, respectively.

Lemma 2.5.11. *Let $T : I \to I$ be C^∞ with $T(y^*) = y^*$ for some $y^* \in I$.*

1. *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.*
2. *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 neighborhood of y^*.*
3. *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 2.5.12. **(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 2.5.10(1), the sequences (x_n) and $(x_{n+1} - x_n)$ are both Benford sequences for (Lebesgue) almost every x_0 in a neighborhood 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 2.5.10(2), 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 2.5.11, an analog of Theorem 2.5.10 can be established for other root-finding algorithms as well (see [BerH1]).

Time-Dependent Systems

So far, the sequences considered in this section have been generated by the iteration of a single map T. Beyond this setting there has been, in the recent past, an increased interest in systems that are **non-autonomous**, i.e., explicitly time dependent in one way or the other.

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 $T^n := T_n \circ \cdots \circ T_1$; also let T^0 be the identity map on $\mathbb{R}$. Given x_0, it makes sense to consider the sequence $O_T(x_0) := \big(T^{n-1}(x_0)\big)_{n\in\mathbb{N}} = \big(x_0, T_1(x_0), T_2\big(T_1(x_0)\big), \ldots\big)$.

The following is a non-autonomous variant of Theorem 2.5.5. A proof (of a substantially more general version) can be found in [BerBH]. It relies heavily on a non-autonomous version of the Shadowing Lemma, Lemma 2.5.3.

Theorem 2.5.13 ([BerBH]). *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 2.5.14. **(i)** *Let $R_j(x) = (2 + j^{-1})x$ for $j = 1, 2, \ldots$. It is easy to see that all assumptions of Theorem 2.5.13 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 jth 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 2.5.13 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 2.4.12. Note that the same argument would* not *work to show that $(n!)$ is Benford.*

In situations where most of the maps T_j are power-like or even more strongly expanding, the following generalization of Lemma 2.5.7 may be useful. (In its fully developed form, the result also extends Theorem 2.5.8; see [BerBH, Thm.5.5] and [Ber3, Thm.3.7].) Again the reader is referred to [Ber4] for a proof.

Theorem 2.5.15 ([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:*

1. *$x \mapsto \ln T_j(e^x)$ is convex on $[\xi, +\infty)$;*
2. *$xT_j'(x)/T_j(x) \ge \beta_j > 0$ for all $x \ge \xi$.*

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 2.5.16. **(i)** *To see that Theorem 2.5.15 does indeed generalize Lemma 2.5.7, 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) *Theorem 2.5.15 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 non-autonomous example consider*

$$T_j(x) \ = \ \begin{cases} x^2 & \text{if } j \text{ is even}\,, \\ 2^x & \text{if } j \text{ is odd}\,, \end{cases} \qquad \textit{or} \quad T_j(x) = (j+1)^x\,.$$

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 2.5.15 may fail if one of its hypotheses is violated even for a single j. For example,*

$$T_j(x) \ = \ \begin{cases} 10 & \text{if } j = 1\,, \\ x^2 & \text{if } j \ge 2\,, \end{cases}$$

satisfies (1) and (2) for all $j > 1$, but does not satisfy assumption (2) for $j = 1$. Clearly, $O_T(x_0)$ is not Benford for any $x_0 \in \mathbb{R}$, since $D_1\big(T^n(x_0)\big) \equiv 1$ for all $n \in \mathbb{N}$.

2.5.2 Multidimensional Discrete-Time Processes

The purpose of this subsection is to extend the basic results of the previous section to multidimensional systems, notably to linear, as well as some non-linear recurrence relations. Recall from Example 2.4.12 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.

Example 2.5.17. **(i)** *Let the sequence (x_n) be defined recursively as*

$$x_{n+1} = x_n - x_{n-1}\,, \quad n = 1, 2, \ldots\,, \tag{2.16}$$

with given x_0, $x_1 \in \mathbb{R}$. By using the matrix $\begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}$ associated with (2.16)*, 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. For no choice of x_0, x_1, therefore, is (x_n) Benford.

(ii) *Consider the linear 3-step recursion*

$$x_{n+1} \ = \ 2x_n + 10x_{n-1} - 20x_{n-2}\,, \quad n = 2, 3, \ldots\,. \tag{2.17}$$

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.

The above recurrence relations (2.16) and (2.17) are linear and have constant coefficients. Hence they can be rewritten and analyzed using matrix–vector notation. For instance, in Example 2.5.17(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 2.5.17 are then merely special cases.

Recall complex numbers $z_1, z_2, \ldots, z_m$ are **rationally independent** if $\sum_{j=1}^m q_j z_j = 0$ with rational $q_1, q_2, \ldots, q_m$ implies that $q_j = 0$ for all $j = 1, 2, \ldots, m$. 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 $\frac{1}{2\pi} \arg Z = \left\{ \frac{1}{2\pi} \arg z : z \in Z \right\} \setminus \{-\frac{1}{2}, 0\}$.

Definition 2.5.18. *A matrix $A \in \mathbb{R}^{d\times d}$ is **Benford regular (base** 10**)** if $\sigma(A)^+$ (the subset of the spectrum of A with non-negative imaginary components) 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 2.5.19. *None of the matrices associated with the recurrence relations in Example 2.5.17 are 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 $\{-\sqrt{10}, \sqrt{10}\}$.*

Example 2.5.20. *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 2.5.21. *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 2.5.2), as well as Example 2.5.17 and Definition 2.5.18 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 2.5.22 ([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 $x \neq 0$.*

Proof. Apply Theorem 2.4.2 and the following proposition, a variant of [Ber2, Lem.2.9]. □

Proposition 2.5.23. *Assume that the real numbers $1, \rho_0, \rho_1, \ldots, \rho_m$ are rationally independent. Let (z_n) be a convergent sequence in $\mathbb{C}$, and at least one of the numbers $c_1, c_2, \ldots, c_m \in \mathbb{C}$ non-zero. Then (x_n) given by*

$$x_n \;=\; n\rho_0 + \log\left|\Re\left(c_1 e^{2\pi\imath n\rho_1} + \cdots + c_m e^{2\pi\imath n\rho_m} + z_n\right)\right|$$

is u.d. mod 1.

Example 2.5.24. *According to Example 2.5.21, the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ is Benford regular. By Theorem 2.5.22, 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 2.4.12. Note that (F_n^2), for instance, is Benford as well by Corollary 2.4.7(1).*

Example 2.5.25. *Recall from Example 2.5.20 that $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ is not Benford regular. Hence Theorem 2.5.22 does not apply, and $(x^\top A^n y)$ 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 for $n = 0, 1, \ldots$,*

$$x^\top A^n y \;=\; \begin{bmatrix} 1 & 0 \end{bmatrix} 2^{n/2} \begin{bmatrix} \cos(\frac{1}{4}\pi n) & -\sin(\frac{1}{4}\pi n) \\ \sin(\frac{1}{4}\pi n) & \cos(\frac{1}{4}\pi n) \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \;=\; 2^{n/2}\cos\left(\tfrac{1}{4}\pi n\right)$$

is clearly not Benford as $x^\top A^n y = 0$ whenever $n = 2 + 4l$ for some $l \in \mathbb{N}_0$.

The present section closes with an example of a non-linear system. The sole purpose is to hint at possible extensions of the results presented earlier; for more details the interested reader is referred to [Ber2].

Example 2.5.26. *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

$$f(t) \;=\; \frac{3}{2}|t+2| - 3|t+1| + 3|t-1| - \frac{3}{2}|t-2| \;=\; \begin{cases} 0 & \text{if } |t| \geq 2\,, \\ 3t+6 & \text{if } -2 < t < -1\,, \\ -3t & \text{if } -1 \leq t < 1\,, \\ 3t-6 & \text{if } 1 \leq t < 2\,. \end{cases}$$

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. This is indeed the case.

2.5.3 Differential Equations

By presenting a few results on, and examples of, differential equations, i.e., deterministic continuous-time processes, this 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**)

$$\dot{x} = F(x)\,, \quad x(0) = x_0\,, \tag{2.18}$$

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 (2.18) is $x(t) = x_0 e^{\alpha t}$. Unless $\alpha x_0 = 0$, therefore, every solution of (2.18) is Benford. As in the discrete-time setting, this feature persists for arbitrary C^2-functions F with $F'(0) < 0$. The direct analog of Theorem 2.5.5 is

Theorem 2.5.27 ([BerBH])**.** *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* (2.18) *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 (2.18) has a unique local solution whenever $|x_0| \leq \delta$; see [Walt]. 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 (2.18) 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. Moreover, for all $t > 0$,

$$\begin{aligned} \big|y(t) - e^{\alpha t}\overline{y_0}\big| \;&=\; \left|\int_t^{+\infty} e^{\alpha(t-\tau)} g\big(y(\tau)^{-1}\big)\,\mathrm{d}\tau\right| \\ &\leq \int_0^{+\infty} e^{-\alpha\tau}\,\big|g\big(y(t+\tau)^{-1}\big)\big|\,\mathrm{d}\tau \leq \frac{\|g\|_\infty}{\alpha}, \end{aligned}$$

where $\|g\|_\infty = \max_{|x|\leq\delta}|g(x)|$, and Lemma 2.5.4(2) shows that y is Benford if and only if $t \mapsto e^{\alpha t}\overline{y_0}$ is. An application of Corollary 2.4.7(2) therefore completes the proof. $\square$

Example 2.5.28. **(i)** *The function $F(x) = -x + x^4 e^{-x^2}$ satisfies the assumptions of Theorem 2.5.27. Thus except for the trivial solution $x = 0$, 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 $x \neq 0$. Hence for every $x_0 \in \mathbb{R}$, the IVP* (2.18) *has a unique solution with $\lim_{t\to+\infty} x(t) = 0$. However, $F'(0) = 0$, and it is not hard to see that this causes x to approach 0 rather slowly. In fact, $\lim_{t\to+\infty} 2tx(t)^2 = 1$ whenever $x_0 \neq 0$, and this prevents x from being Benford.*

Similar results follow for the linear d-dimensional ordinary differential equations $\dot{x} = Ax$, where A is a real $d \times d$-matrix; see [Ber2].

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 2.5.29. *A fundamental example of a partial differential equation is the so-called one-dimensional* ***heat*** *(or* ***diffusion****)* ***equation***

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}\,, \tag{2.19}$$

a linear second-order equation for $u = u(t,x)$. Physically, (2.19) *describes e.g. the diffusion over time of heat in a homogeneous one-dimensional medium. Without further conditions,* (2.19) *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,

$$u(t,x) \;=\; \frac{1}{\sqrt{t}} e^{-x^2/(4t)} \quad (t > 0)$$

is Benford in x but not in t, and

$$u(t,x) \;=\; e^{c^2 t + cx}$$

is Benford in both t and x.

2.6 BENFORD'S LAW FOR RANDOM PROCESSES

The purpose of this section 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, random matrices, Lévy processes, and Bayesian models. The present section may also serve as a preparation for the specialized literature on these advanced topics [EngLeu, JaKKKM, LeScEv, MiNi1, MiNi2, Schür2].

2.6.1 Independent Random Variables

Recall that a sequence (X_n) of random variables **converges in distribution** to a random variable X, symbolically $X_n \overset{\mathcal{D}}{\to} X$, 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 $S(X_n) \overset{\mathcal{D}}{\to} S(X)$, where X is a Benford random variable, or equivalently if

$$\lim_{n\to\infty} \mathbb{P}(S(X_n) \le t) \;=\; \log t \quad \text{for all } t \in [1, 10)\,.$$

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 2.4.10 and 2.4.15 imply

Theorem 2.6.1. *Assume that the random variable X has a density. Then*

1. *X^n converges in distribution to BL;*
2. *with probability one, (X^n) is Benford.*

Proof. To prove (1), note that the random variable $\log|X|$ has a density as well. Hence, by Theorem 2.4.15,

$$\mathbb{P}(S(X_n) \le t) \;=\; \mathbb{P}(\langle \log|X^n| \rangle \le \log t) \;=\; \mathbb{P}(\langle n \log|X| \rangle \le \log t) \;\to\; \log t$$

as $n \to \infty$ holds for all $t \in [1, 10)$, i.e., (X_n) converges in distribution to BL.

To see (2), simply note that $\log|X|$ is irrational with probability one. By Theorem 2.4.10, therefore, $\mathbb{P}\big((X^n) \text{ is Benford}\big) = 1$. $\square$

Example 2.6.2. **(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 $\dfrac{e^t-1-t}{e^t-1} < \dfrac{t}{2}$ for all $t>0$ shows that

$$\bigl|F_{S(X^n)}(t) - \log t\bigr| \le \frac{10^{1/n}-1-\frac{\ln 10}{n}}{10^{1/n}-1} < \frac{\ln 10}{2n} \;\to\; 0 \quad \text{as } n \to \infty\,,$$

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.*

The sequence of random variables considered in Theorem 2.6.1 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 2.6.1, products of *independent* factors are considered.

Theorem 2.6.3. *Let X, Y be two independent random variables with $\mathbb{P}(XY = 0) = 0$. Then*

1. *if X is Benford then so is XY;*
2. *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 Lemma 2.4.26, 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,

$$\widehat{P_{\log S(XY)}} \;=\; \widehat{P_{\langle X_0+Y_0\rangle}} \;=\; \widehat{P_{X_0}} \cdot \widehat{P_{Y_0}}\,. \tag{2.20}$$

To prove (1), 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 (2), assume that $S(X)$ and $S(XY)$ have the same distribution. In this case, (2.20) implies that

$$\widehat{P_{X_0}}(k)\bigl(1 - \widehat{P_{Y_0}}(k)\bigr) \;=\; 0 \quad \text{for all } k \in \mathbb{Z}\,.$$

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 $P_{Y_0}(\frac{1}{|k_0|}\mathbb{Z}) = 1$, hence $|k_0|Y_0 = |k_0|\log S(Y)$ is an integer with probability one. $\square$

Example 2.6.4. *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 2.6.3(1), 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 2.6.3(1). It is essential in assertion (2) 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.*

Theorem 2.6.5. *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*

1. *$\left(\prod_{j=1}^n X_j\right)$ converges in distribution to BL;*
2. *with probability one, $\left(\prod_{j=1}^n X_j\right)$ 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 2.4.13(3), the sequence of $\left\langle \sum_{j=1}^n Y_j \right\rangle = \left\langle \log |\prod_{j=1}^n X_j| \right\rangle$ converges in distribution to $U(0,1)$. Thus $\left(\prod_{j=1}^n X_j\right)$ converges in distribution to BL.

To prove (2), let Y_0 be u.d. mod 1 and independent of $(Y_n)_{n\in\mathbb{N}}$, and define

$$S_j := \langle Y_0 + Y_1 + \cdots + Y_j \rangle, \quad j \in \mathbb{N}_0 .$$

Recall from Theorem 2.4.13(1) that S_j is u.d. mod 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. 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

$$\begin{aligned} \mathcal{B}_\infty &:= \sigma\big(\{B_0 \times B_1 \times \cdots \times B_j \times [0,1) \times [0,1) \times \cdots : j \in \mathbb{N}_0, \\ &\qquad B_0, B_1, \ldots, B_j \in \mathcal{B}[0,1)\}\big) \qquad (2.21) \\ &= \bigotimes_{j\in\mathbb{N}_0} \mathcal{B}[0,1). \end{aligned}$$

A probability measure P_∞ is uniquely defined on $(\mathbb{T}_\infty, \mathcal{B}_\infty)$ by setting

$$\begin{aligned} &P_\infty(B_0 \times B_1 \times \cdots \times B_j \times [0,1) \times [0,1) \times \cdots) \\ &\quad = \mathbb{P}(S_0 \in B_0, S_1 \in B_1, \ldots, S_j \in B_j) \end{aligned}$$

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 $\mathbb{T}_\infty$, 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,

$$\begin{aligned} &(\sigma_\infty)_* P_\infty(B_0 \times B_1 \times \cdots \times B_j \times [0,1) \times [0,1) \times \cdots) \\ &\quad = P_\infty([0,1) \times B_0 \times B_1 \times \cdots \times B_j \times [0,1) \times [0,1) \times \cdots) \\ &\quad = \mathbb{P}(S_1 \in B_0, S_2 \in B_1, \ldots, S_{j+1} \in B_j) \\ &\quad = \mathbb{P}(S_0 \in B_0, S_1 \in B_1, \ldots, S_j \in B_j) \\ &\quad = P_\infty(B_0 \times B_1 \times \cdots \times B_j \times [0,1) \times [0,1) \times \cdots), \end{aligned}$$

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 [Shi, 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

$$P_\infty\big(A\,\Delta\,(B_0 \times B_1 \times \cdots \times B_N \times [0,1) \times [0,1) \times \cdots)\big) < \varepsilon\,.$$

For notational convenience, let $A_\varepsilon := B_0 \times B_1 \times \cdots \times B_N \times [0,1) \times [0,1) \times \cdots \in \mathcal{B}_\infty$, 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 2.4.13(3) that, given $S_0, S_1, \ldots, S_N$, the random variables S_n converge in distribution to $U(0,1)$. Thus, for all sufficiently large M,

$$\begin{aligned}
&\big|P_\infty\big(A_\varepsilon^c \cap \sigma_\infty^{-M}(A_\varepsilon)\big) - P_\infty(A_\varepsilon^c)P_\infty\big(\sigma_\infty^{-M}(A_\varepsilon)\big)\big| \\
&\quad = \big|P_\infty\big(A_\varepsilon^c \cap \sigma_\infty^{-M}(A_\varepsilon)\big) - P_\infty(A_\varepsilon^c)P_\infty(A_\varepsilon)\big| < \varepsilon,
\end{aligned}$$

and similarly $\big|P_\infty\big(A_\varepsilon \cap \sigma_\infty^{-M}(A_\varepsilon^c)\big) - P_\infty(A_\varepsilon)P_\infty(A_\varepsilon^c)\big| < \varepsilon$. (Note that (2.22) may not hold if X_1, and hence also Y_1, is purely atomic.) Overall, therefore,

$$\begin{aligned}
2P_\infty(A_\varepsilon)\big(1 - P_\infty(A_\varepsilon)\big) &\le 2\varepsilon + P_\infty\big(A_\varepsilon\Delta\sigma_\infty^{-M}(A_\varepsilon)\big) \\
&\le 2\varepsilon + P_\infty(A_\varepsilon\Delta A) + P_\infty\big(A\Delta\sigma_\infty^{-M}(A)\big) \\
&\qquad + P_\infty\big(\sigma_\infty^{-M}(A)\Delta\sigma_\infty^{-M}(A_\varepsilon)\big) \qquad (2.22) \\
&< 4\varepsilon,
\end{aligned}$$

and consequently $P_\infty(A)\big(1 - P_\infty(A)\big) < 4\varepsilon + \varepsilon^2$. 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 [Shi, Def.V.3.2].) By the **Birkhoff Ergodic Theorem** (e.g. [Ber1]), for every (measurable) function $f : [0,1) \to \mathbb{C}$ with $\int_0^1 |f(x)|\,\mathrm{d}x < +\infty$,

$$\frac{1}{n}\sum_{j=0}^{n} f(x_j) \;\to\; \int_0^1 f(x)\,\mathrm{d}x \quad \text{as } n \to \infty$$

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

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n} f(S_j) \;=\; \int_0^1 f(x)\,\mathrm{d}x \quad \text{a.s.} \qquad (2.23)$$

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 $m \in \mathbb{N}$, let

$$\Omega_{f,t,m} := \left\{\omega \in \Omega : \limsup_{n\to\infty}\left|\frac{1}{n}\sum_{j=1}^{n} f\big(\langle t + Y_1(\omega) + \cdots + Y_j(\omega)\rangle\big) - \int_0^1 f(x)\,\mathrm{d}x\right| < \frac{1}{m}\right\}.$$

According to (2.23), $1 = \int_0^1 \mathbb{P}(\Omega_{f,t,m})\,\mathrm{d}t$, and hence $\mathbb{P}(\Omega_{f,t,m}) = 1$ for a.e. $t \in [0,1)$. Since f is uniformly continuous, for every $m \geq 2$ 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

$$1 \;=\; \mathbb{P}\left(\bigcap_{m\geq 2} \Omega_{f,t_m,m}\right) \;\leq\; \mathbb{P}\left(\bigcap_{m\geq 2} \Omega_{f,0,\lfloor m/2\rfloor}\right) \;\leq\; 1\,,$$

it is clear that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} f\big(\langle Y_1 + \cdots + Y_j\rangle\big) \;=\; \int_0^1 f(x)\,\mathrm{d}x \quad \text{a.s.} \tag{2.24}$$

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 (2.24) shows that, with probability one,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} e^{2\pi\imath k(Y_1+\cdots+Y_j)} \;=\; \int_0^1 e^{2\pi\imath kx}\mathrm{d}x = 0 \quad \text{for all } k \in \mathbb{Z}, k \neq 0\,. \tag{2.25}$$

By Weyl's criterion [KuiNi, Thm.2.1], (2.25) is equivalent to

$$\mathbb{P}\left(\left(\sum_{j=1}^{n} Y_j\right) \text{ is u.d. mod } 1\right) \;=\; 1\,.$$

In other words, $(\prod_{j=1}^{n} X_j)$ is Benford with probability one. □

Example 2.6.6. **(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 kth 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 2.4.13(3), this implies that $\left(\prod_{j=1}^{n} X_j\right)$ converges in distribution to BL, a fact apparently first recorded in [AdhSa]. Note also that $\mathbb{E}\log X_1 = \log\frac{a}{e}$. Thus with probability one, $\left(\prod_{j=1}^{n} X_j\right)$ converges to 0 or $+\infty$, depending on whether $a < e$ or $a > e$. In fact, by the ***Strong Law of Large Numbers*** *[ChT],*

$$\sqrt[n]{\prod_{j=1}^{n} X_j} \;\overset{a.s.}{\to}\; \frac{a}{e}$$

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 2.6.5(2), 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* $\langle \log \prod_{j=1}^n X_j \rangle$. *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 2.6.5(2) 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 2.6.5, 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 number (as in the case of products), but also on their *exponents*. For example, observe that

$$S\big(3 \cdot 10^3 + 2 \cdot 10^2\big) = 3.2 \ne 5 = S\big(3 \cdot 10^2 + 2 \cdot 10^2\big),$$

while clearly

$$S\big(3 \cdot 10^3 \times 2 \cdot 10^2\big) = 6 = S\big(3 \cdot 10^2 \times 2 \cdot 10^2\big).$$

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 2.4.2, 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 2.6.5 for sums arrives at a radically different conclusion.

Theorem 2.6.7. *Let (X_n) be an i.i.d. sequence of random variables with finite variance, that is, $\mathbb{E}X_1^2 < +\infty$. Then*

1. *not even a subsequence of $\left(\sum_{j=1}^n X_j\right)$ converges in distribution to BL;*
2. *with probability one, $\left(\sum_{j=1}^n X_j\right)$ is not Benford.*

Proof. See [BerH4, Thm.6.8]. □

Example 2.6.8. *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 2.6.7(1) 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 $n \in \mathbb{N}$,*

$$\mathbb{P}\left(\sum_{j=1}^n X_j = l\right) = 2^{-n}\binom{n}{l}, \quad l = 0, 1, \ldots, n.$$

The law of the iterated logarithm [ChT] asserts that

$$\sum_{j=1}^n X_j = \frac{n}{2} + Y_n\sqrt{n \ln\ln n} \quad \text{for all } n \geq 3, \tag{2.26}$$

where the sequence (Y_n) of random variables is bounded; in fact $|Y_n| \leq 1$ a.s. for all n. From (2.26) it is clear that, with probability one, the sequence $\left(\sum_{j=1}^n X_j\right)$ is not Benford.

2.6.2 Mixtures of Distributions

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,

$$\{\omega : a \leq P(\omega)(B) \leq b\} \in \mathcal{A};$$

see e.g. [Ka] for an authoritative account on random probability measures.

Example 2.6.9. **(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 2.3.8(i,ii). Thus, for every $\omega \in \Omega$, the probability measure $P(\omega)$ is either $U(0,1)$ or $\texttt{exp}(1)$, and $\mathbb{P}\big(P(\omega) = U(0,1)\big) = \mathbb{P}\big(P(\omega) = \texttt{exp}(1)\big) = \frac{1}{2}$. 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 $P(\omega)$ 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 2.6.10. *Let P be the r.p.m. with support on $[1, 10)$, i.e., $P\big([1, 10)\big) = 1$ with probability one, defined by its (random) cumulative distribution function F_P, i.e.,*

$$F_P(t) \ := \ F_{P(\omega)}(t) = P(\omega)\big([1, t]\big)\,, \quad 1 \le t < 10\,,$$

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 $\big[0, F_P(10^{1/2})\big)$ and $\big[F_P(10^{1/2}), 1\big)$, respectively, and continue in this manner. This construction is known to generate an r.p.m. a.s. [DuFr, Lem.9.28], and as can easily be seen, is dense in the set of all probability measures on $\big([1, 10), \mathcal{B}[1, 10)\big)$, 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 2.6.11. *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:*

Given that $P_j = Q$, the random variables $X_{(j-1)m+1}, X_{(j-1)m+2}, \ldots, X_{jm}$ are i.i.d. with distribution Q. (2.27)

The random variables $X_{(j-1)m+1}, X_{(j-1)m+2}, \ldots, X_{jm}$ are independent of $P_i, X_{(i-1)m+1}, X_{(i-1)m+2}, \ldots, X_{im}$ for every $i \ne j$. (2.28)

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 2.6.12. *Let P be the r.p.m. in Example 2.6.9(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 $\frac{1}{2}$.*

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. (See [Hi4]).

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 2.6.13. *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 2.6.14. *Let P be the r.p.m. of Example 2.6.9(i). Then $\mathbb{E}P$ is the Borel probability measure with density*

$$f_{\mathbb{E}P}(t) = \left\{ \begin{array}{ll} 0 & \text{if } t < 0\,, \\ \frac{1}{2} + \frac{1}{2}e^{-t} & \text{if } 0 \le t < 1\,, \\ \frac{1}{2}e^{-t} & \text{if } t \ge 1\,, \end{array} \right\} = \frac{1}{2}\mathbb{1}_{[0,1)}(t) + \frac{1}{2}e^{-t}\mathbb{1}_{[0,+\infty)}\,, \quad t \in \mathbb{R}\,.$$

The next lemma shows that the limiting proportion of times that a sequence of P-random m-samples 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.

Lemma 2.6.15. *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

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : X_n \in B\}}{N} \;=\; \frac{1}{m} \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} Y_j\,, \tag{2.29}$$

whenever the limit on the right exists. By (2.27), 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.

$$\mathbb{E}Y_j \;=\; \mathbb{E}\big(\mathbb{E}(Y_j|P_j)\big) \;=\; \mathbb{E}\big(mP_j(B)\big) \;=\; m\mathbb{E}P(B) \tag{2.30}$$

since P_j has the same distribution as P. By (2.28), the Y_j are independent. They are also uniformly bounded, as $0 \le Y_j \le m$ for all j, and hence $\sum_{j=1}^{\infty} \mathbb{E}Y_j^2/j^2 < +\infty$. Moreover, by (2.30) all Y_j have the same mean value $m\mathbb{E}P(B)$. Thus by [ChT, Cor.5.1]

$$\frac{1}{n}\sum_{j=1}^{n} Y_j \overset{a.s.}{\to} m\mathbb{E}P(B) \quad \text{as } n \to \infty\,, \tag{2.31}$$

and the conclusion follows by (2.29) and (2.31). □

The stage is now set to give a statistical limit law (Theorem 2.6.18 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 data sets, and also his individual data set of numbers gleaned from newspapers.

Definition 2.6.16. *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 2.4.18 and 2.4.24 is

Proposition 2.6.17. *Let P be an r.p.m. with $\mathbb{E}P(\{0\}) = 0$. Then the following statements are equivalent:*

1. *P has scale-unbiased significant digits.*

2. *$P(\{\pm 10^k : k \in \mathbb{Z}\}) = 0$, or equivalently $S_*P(\{1\}) = 0$, holds with probability one, and P has base-unbiased significant digits.*

3. *$\mathbb{E}P(A) = \mathbb{B}(A)$ for all $A \in \mathcal{S}$, i.e., $\mathbb{E}P$ is Benford.*

As will be seen in the next theorem, scale- or base-unbiasedness of an r.p.m. implies that a 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 2.6.20 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 2.6.18 ([Hi4]). *Let P be an r.p.m. Assume that 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 2.4.18, $\mathbb{E}P$ is Benford. Consequently, Lemma 2.6.15 implies that for every sequence (X_n) of P-random m-samples and every $t \in [1, 10)$,

$$\begin{aligned}
&\frac{\#\{1 \le n \le N : S(X_n) < t\}}{N} \\
&\quad = \frac{\#\left\{1 \le n \le N : X_n \in \bigcup_{k\in\mathbb{Z}} 10^k\big((-t,-1] \cup [1,t)\big)\right\}}{N} \\
&\quad \overset{a.s.}{\to} \mathbb{E}P\left(\bigcup_{k\in\mathbb{Z}} 10^k\big((-t,-1] \cup [1,t)\big)\right) = \log t
\end{aligned}$$

as $N \to \infty$. 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 (2.14) with P replaced by $\mathbb{E}P$, proving that $\mathbb{E}P$ is Benford, and the remaining argument is the same as before. □

Corollary 2.6.19. *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 $d \in \{1, 2, \ldots, 9\}$,*

$$\frac{\#\{1 \le n \le N : D_1(X_n) = d\}}{N} \overset{a.s.}{\to} \log(1 + d^{-1}) \quad \text{as } N \to \infty\,.$$

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 of the standard constructions of r.p.m. automatically have scale- and base-unbiased significant digits, and thus satisfy BL in the sense of Theorem 2.6.18.

Example 2.6.20. *Recall the classical Dubins–Freedman construction of an r.p.m. P described in Example 2.6.10. It follows from [DuFr, 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 2.6.18, this is enough to guarantee that random sampling from P will generate Benford sequences a.s.*

2.6.3 Random Maps

The purpose of this brief concluding section is to illustrate one basic theorem that combines the deterministic aspects of BL studied in Section 2.5 with the stochastic considerations of the present section. 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

$$T(x) \;=\; \begin{cases} \sqrt{|x|} & \text{with probability } p\,, \\ x^3 & \text{with probability } 1-p\,, \end{cases} \tag{2.32}$$

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 2.5.8 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

$$p \;<\; \frac{\log 3}{\log 2 + \log 3} \;=\; 0.6131\ldots\,, \tag{2.33}$$

then, for a.e. $x_0 \in \mathbb{R}$, the (random) orbit $O_T(x_0)$ is Benford with probability one. To concisely formulate the result from which this follows, 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 $\big(T_{n-1} \circ \cdots \circ T_1(x_0)\big)_{n\in\mathbb{N}}$.

Theorem 2.6.21 ([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* $\mathbb{E} \log \beta_1 \leq 0$.

Proof. See [Ber3]. □

Statements in the spirit of Theorem 2.6.21 are true also for more general random maps, not just monomials [Ber3].