Finite-state Markov Chains Obey Benford's Law

Bahar Kaynar
Vrije Universiteit, Amsterdam, The Netherlands

Arno Berger
University of Alberta, Edmonton, Canada

Theodore P. Hill
Georgia Institute of Technology, Atlanta, USA

Ad Ridder
Vrije Universiteit, Amsterdam, The Netherlands

March 5, 2010

Abstract

A sequence of real numbers (x_n) is Benford if the significands, i.e. the fraction parts in the floating-point representation of (x_n), are distributed logarithmically. Similarly, a discrete-time irreducible and aperiodic finite-state Markov chain with probability transition matrix P and limiting matrix P^* is Benford if every component of both sequences of matrices $(P^n - P^*)$ and $(P^{n+1} - P^n)$ is Benford or eventually zero. Using recent tools that established Benford behavior both for Newton's method and for finite-dimensional linear maps, via the classical theories of uniform distribution modulo 1 and Perron-Frobenius, this paper derives a simple sufficient condition ("nonresonance") guaranteeing that P, or the Markov chain associated with it, is Benford. This result in turn is used to show that almost all Markov chains are Benford, in the sense that if the transition probabilities are chosen independently and continuously, then the resulting Markov chain is Benford with probability one. Concrete examples illustrate the various cases that arise, and the theory is complemented with several simulations and potential applications.

Keywords: Markov chain, Benford's Law, uniform distribution modulo 1, significant digits, significand, n-step transition probabilities, stationary distribution.

Kaynar and Hill were supported by grants 400-06-044 and 040-11-091 from The Netherlands Organization for Scientific Research (NWO); Berger was supported by an NSERC Discovery Grant.

Corresponding author's address: Ad Ridder, Department of Econometrics and Operations Research, Vrije Universiteit Amsterdam, The Netherlands; email: `aridder@feweb.vu.nl`

1 Introduction

Benford's Law (BL) is the widely-known logarithmic probability distribution on significant digits (or equivalently, on significands), and its most familiar form is the special case of first significant digits (base 10), namely,

$$\mathbb{P}\left(D_1 = d_1\right) = \log_{10}\left(1 + \frac{1}{d_1}\right), \quad \forall d_1 \in \{1, 2, \ldots, 9\}, \tag{1}$$

where for each $x \in \mathbb{R}^+$, the number $D_1(x)$ is the *first significant digit* (base 10) of x, i.e. the unique integer $d \in \{1, 2, \ldots, 9\}$ satisfying $10^k d \le x < 10^k(d+1)$ for some, necessarily unique, $k \in \mathbb{Z}$. Thus, for example, $D_1(30122) = D_1(0.030122) = D_1(3.0122) = 3$, and (1) implies that

$$\mathbb{P}(D_1 = 1) = \log_{10} 2 \cong 0.301\,, \quad \mathbb{P}(D_1 = 2) = \log_{10}(3/2) \cong 0.176\,, \quad \text{etc.},$$

see also Table 1 below.

In a form more complete than (1), BL is a statement about joint distributions of the first n significant digits (base 10) for any $n \in \mathbb{N}$, namely,

$$\begin{aligned} &\mathbb{P}\big((D_1, D_2, D_3, \ldots, D_n) = (d_1, d_2, d_3, \ldots, d_n)\big) \\ &\qquad = \log_{10}\left(\sum\nolimits_{j=1}^{n} 10^{n-j} d_j + 1\right) - \log_{10}\left(\sum\nolimits_{j=1}^{n} 10^{n-j} d_j\right) \\ &\qquad = \log_{10}\left(1 + \frac{1}{\sum_{j=1}^{n} 10^{n-j} d_j}\right), \end{aligned} \tag{2}$$

where $d_1 \in \{1, 2, \ldots, 9\}$ and $d_j \in \{0, 1, 2, \ldots, 9\}$ for $j \ge 2$, and D_2, D_3, etc. represent the second, third, etc. *significant digit functions* (base 10). Thus, for example, $D_2(30122) = D_2(0.030122) = D_2(3.0122) = 0$, and a special case of (2) is

$$\mathbb{P}\big((D_1, D_2, D_3) = (3, 0, 1)\big) = \log_{10} 302 - \log_{10} 301 = \log_{10}\left(1 + \frac{1}{301}\right) \cong 0.00144\,.$$

Formally, for every $n \in \mathbb{N}$, $n \ge 2$, the number $D_n(x)$, the *n-th significant digit* (base 10) of $x \in \mathbb{R}^+$, is defined inductively as the unique integer $d \in \{0, 1, 2, \ldots, 9\}$ such that

$$10^k \left(d + \sum\nolimits_{j=1}^{n-1} 10^{n-j} D_j(x)\right) \le x < 10^k \left(d + 1 + \sum\nolimits_{j=1}^{n-1} 10^{n-j} D_j(x)\right)$$

for some (unique) $k \in \mathbb{Z}$.

The formal probability framework for the significant-digit law is described in [12, 13]. The sample space is the set of positive reals, and the σ-algebra of events is the σ-algebra generated by the (decimal) *significand* (or *mantissa*) *function* $S : \mathbb{R}^+ \to [1, 10)$, where $S(x)$ is the unique number $s \in [1, 10)$ such that $x = 10^k s$ for some $k \in \mathbb{Z}$. Equivalently, the significand events are the sets in the σ-algebra generated by the significant digit functions D_1, D_2, D_3, etc. The probability measure on this sample space associated with BL is

$$\mathbb{P}(S \le t) = \log_{10} t\,, \quad \forall t \in [1, 10)\,.$$

It is easy to see that the significant digit functions D_1 and D_2, D_3, etc. are well-defined $\{1, 2, \ldots, 9\}$- and $\{0, 1, 2, \ldots, 9\}$-valued random variables, respectively, on this probability space with probability mass functions as given in (1) and (2).

Note. Throughout this article, all results are restricted to decimal (base 10) significant digits, and accordingly log always denotes the base 10 logarithm. For notational convenience, $D_n(0) := 0$ for all $n \in \mathbb{N}$. The results carry over easily to arbitrary bases $b \in \mathbb{N} \setminus \{1\}$, as is evident from [2], where the essential difference is replacing $\log_{10}$ by $\log_b$, and the decimal significant digits by the base-b significant digits.

Benford's Law is now known to hold in great generality, e.g. for classical combinatorial sequences such as (2^n), $(n!)$ and the Fibonacci numbers (F_n); iterations of linearly- or nonlinearly-dominated functions; solutions of ordinary differential equations; products of independent random variables; random mixtures of data; and random maps (e.g., see [1, 4, 5, 8, 13]). Table 1 compares the empirical frequencies of D_1 for the first 1000 terms of the sequences (2^n), $(n!)$ and (F_n). These empirical frequencies illustrate what it means to follow BL and also foreshadow the simulations in Section 5.

The main contribution of this article is to adapt recent results on BL in the multi-dimensional setting ([2]) in order to establish BL in finite-dimensional, time-homogeneous Markov chains, and to suggest several applications including error analysis in numerical simulations of n-step transition matrices.

Concretely, given the transition matrix P of a finite-state Markov chain (i.e., P is a row-stochastic matrix), a common problem is to estimate the limit $P^* = \lim_{n\to\infty} P^n$. The two main theoretical results below, Theorems A and B, respectively, show that

D_1	$(n!)$	(2^n)	(F_n)	Benford
1	0.293	0.292	0.301	0.30103
2	0.176	0.180	0.176	0.17609
3	0.124	0.126	0.126	0.12493
4	0.102	0.098	0.096	0.09691
5	0.087	0.081	0.079	0.07918
6	0.069	0.068	0.067	0.06694
7	0.051	0.057	0.057	0.05799
8	0.051	0.053	0.053	0.05115
9	0.047	0.045	0.045	0.04575

Table 1: *Empirical frequencies of* D_1 *for the first* 1000 *terms of the sequences* (2^n), $(n!)$ *and the Fibonacci numbers* (F_n), *as compared with the Benford probabilities.*

under a natural condition ("nonresonance") every component of the sequence of matrices $(P^n - P^*)$ and $(P^{n+1} - P^n)$ obeys BL, and that this behavior is typical, i.e., it occurs for almost all Markov chains. Simulations are provided for illustration, followed by several potential applications including the estimation of roundoff errors incurred when estimating P^* from P^n, and possible (partial negative) statistical tests to decide whether data comes from a finite-state Markov process.

2 Benford Markov chains and main tools

The set of natural, integer, rational, positive real, real and complex numbers are symbolized by $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}^+, \mathbb{R}$ and $\mathbb{C}$, respectively. The real part, imaginary part, complex conjugate and absolute value (modulus) of a number $z \in \mathbb{C}$ is denoted by $\mathfrak{Re} z, \mathfrak{Im} z, \bar{z}$ and $|z|$, respectively. For $z \neq 0$, the argument $\arg z$ is the unique number in $(-\pi, \pi]$ that satisfies $z = |z| e^{i \arg z}$. For ease of notation, $\arg 0 := 0$ and $\log 0 := 0$. The cardinality of the finite set A is $\#A$. Throughout this article, the sequence $\big(a(1), a(2), a(3), \ldots\big)$ is denoted by $\big(a(n)\big)$. Thus, for example, $(\alpha^n) = (\alpha^1, \alpha^2, \alpha^3, \ldots)$ and $\big(P^{n+1} - P^n\big) = \big(P^2 - P^1, P^3 - P^2, P^4 - P^3, \ldots\big)$. Boldface symbols indicate randomized quantities, e.g. $\boldsymbol{X}$ denotes a random variable or vector and $\boldsymbol{P}$ a random transition probability matrix.

Definition 2.1. A sequence (x_n) of real numbers is *Benford* ("*follows BL*") if

$$\lim_{n\to\infty} \frac{\#\{j \le n : S(|x_j|) \le t\}}{n} = \log t\,, \quad \forall t \in [1,10)\,.$$

The main subject of this paper is the Benford behavior of finite-state Markov chains. The theory uses three main tools: the classical theory of *uniform distribution modulo 1*, see e.g. [16]; recent results for BL in one- and multi-dimensional dynamical systems ([1, 2]); and the classical Perron-Frobenius theory for Markov chains, see e.g. [6, 19]. The first lemma records the relationship between uniform distribution theory and BL, and the second lemma is an application establishing BL for certain basic sequences that will be used repeatedly below. Here and throughout, the term *uniformly distributed modulo 1* is abbreviated as *u.d.* mod 1.

Lemma 2.2 ([8]). *A sequence (x_n) of real numbers is Benford if and only if* $(\log |x_n|)$ *is u.d.* mod 1*.*

An immediate application of Lemma 2.2 is the following useful lemma.

Lemma 2.3 ([1]). *Let (x_n) be Benford. Then for all $\alpha \in \mathbb{R}$ and $k \in \mathbb{Z}$ with $\alpha k \neq 0$, the sequence (αx_n^k) is also Benford.*

Lemmas 2.2 and 2.3 are fundamental tools for analyzing BL in the setting of multi-dimensional dynamical systems ([2]), and although those results do not apply directly to the Markov chain setting, the first part of the theory established below relies heavily on those ideas specialized to the case of row-stochastic matrices.

The next lemma follows easily from known results. It is included here since these observations play a central role in determining whether a Markov chain is Benford, as illustrated in the three examples following the lemma. Stronger conclusions are possible, as suggested in Example 2.5(iii) below, but are not needed here.

Lemma 2.4. *Let a, b, α, β be real numbers with $a \neq 0$ and $|\alpha| > |\beta|$. Then $(a\alpha^n + b\beta^n)$ is Benford if and only if* $\log |\alpha|$ *is irrational.*

Proof. Since $|\alpha| > |\beta|$, the significands of α^n dominate those of β^n asymptotically, so the conclusion follows from Lemma 2.2, Lemma 2.3 and Weyl's classical theorem that iterations of an irrational rotation on the circle are uniformly distributed. □

Example 2.5.

(i) The sequences (2^n), (0.2^n), (3^n), (0.3^n) are Benford, whereas (10^n), (0.1^n), $\left(\sqrt{10}^n\right)$ are not Benford.

(ii) The sequence $\left(0.01\cdot 0.2^n + 0.2\cdot 0.01^n\right)$ is Benford, whereas $\left(0.1\cdot 0.02^n + 0.02\cdot 0.1^n\right)$ is not Benford.

(iii) The sequence $\left(0.2^n + (-0.2)^n\right)$ is not Benford, since all odd terms are zero, but $\left(0.2^n + (-0.2)^n + 0.03^n\right)$ is Benford — although this does not follow directly from Lemma 2.4.

Notation. For every integer $d > 1$, the set of all row-stochastic matrices of size $d \times d$ is denoted by $\mathcal{P}_d$.

Now, let $P \in \mathcal{P}_d$ be the transition probability matrix of a Markov chain. All Markov chains (or their associated matrices P) considered in this work are assumed to be finite-state (with $d > 1$ states), irreducible and aperiodic. Let $\lambda_1, \ldots, \lambda_s$, $s \le d$, be the *distinct* (possibly non-real) eigenvalues of the stochastic matrix P, with corresponding spectrum $\sigma(P) = \{\lambda_1, \ldots, \lambda_s\}$, i.e., $\sigma(P)$ is the set of all distinct eigenvalues. Accordingly, the set $\sigma(P)^+ = \{\lambda \in \sigma(P) : \mathfrak{Im}\lambda \ge 0\}$ forms the "upper half" of the spectrum. The usage of $\sigma(P)^+$ refers to the fact that non-real eigenvalues of real matrices always occur in conjugate pairs, so the set $\sigma(P)^+$ only includes one of the conjugates. Without loss of generality, throughout this work it is also assumed that the eigenvalues in $\sigma(P)$ are labeled such that

$$|\lambda_1| \ge |\lambda_2| \ge \ldots \ge |\lambda_s| .$$

Furthermore, the column vectors $u_1, \ldots, u_s$ and $v_1, \ldots, v_s$ denote associated sequences of left and right eigenvectors, respectively. The third main tool in this paper is the classical Perron-Frobenius theory of Markov chains, and the following lemma summarizes some of the special properties of transition matrices for ease of reference.

Lemma 2.6. *Suppose $P \in \mathcal{P}_d$ is irreducible and aperiodic. Then $\lambda_1 = 1 > |\lambda_\ell|$ for all $\ell = 2, \ldots, s$, and there exists a $P^* \in \mathcal{P}_d$ such that*

(i) $\lim_{n\to\infty} P^n = P^*$*;*

(ii) *for every* $n \in \mathbb{N}$,

$$P^n - P^* = \lambda_2^n C_2 + \ldots + \lambda_s^n C_s\,, \tag{3}$$

where each C_ℓ *is a* $d \times d$*-matrix whose components* $C_\ell^{(i,j)}$ *are polynomials in* n *with complex coefficients and degrees* $k_\ell^{(i,j)} < d$.

Proof. Immediate from the Perron-Frobenius theorem, see e.g. [18]. □

The second dominant eigenvalue λ_2 plays an important role whenever $C_2^{(i,j)} \neq 0$. The analysis is especially straightforward if all eigenvalues are simple, i.e., if $\#\sigma(P) = d$. In this case, for every $n \in \mathbb{N}$,

$$P^n - P^* = \sum\nolimits_{\ell=2}^{d} \lambda_\ell^n B_\ell \quad \text{and} \quad P^{n+1} - P^n = \sum\nolimits_{\ell=2}^{d} \lambda_\ell^n(\lambda_\ell - 1)B_\ell \tag{4}$$

holds with the $d-1$ matrices $B_\ell = v_\ell u_\ell^\top / v_\ell^\top u_\ell \in \mathbb{C}^{d\times d}$. Next is the key definition in this paper.

Definition 2.7. A Markov chain, or its associated transition probability matrix P, is *Benford* if each component of $(P^n - P^*)$ and $\left(P^{n+1} - P^n\right)$ is either Benford or eventually zero.

The following examples illustrate the notions of Benford and non-Benford Markov chains.

Example 2.8. (Examples of Benford Markov chains)

(i) Let $d = 2$ and $P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$. By [10, p. 432], $P^* = \frac{1}{7}\begin{bmatrix} 4 & 3 \\ 4 & 3 \end{bmatrix}$, and

$$P^n - P^* = \frac{0.3^n}{7}\begin{bmatrix} 3 & -3 \\ -4 & 4 \end{bmatrix} \quad \text{and} \quad P^{n+1} - P^n = 0.3^n \begin{bmatrix} -0.3 & 0.3 \\ 0.4 & -0.4 \end{bmatrix}$$

holds for all $n \in \mathbb{N}$. In both sequences every component is a multiple of (0.3^n), and hence Benford by Lemma 2.4 since $\log 0.3$ is irrational. The two-dimensional case will be discussed in more generality in Examples 3.5 and 4.2.

(ii) Let $d = 3$ and $P = \begin{bmatrix} 0.9 & 0.0 & 0.1 \\ 0.6 & 0.3 & 0.1 \\ 0.1 & 0.0 & 0.9 \end{bmatrix}$. It is easy to check via spectral decomposition (e.g. [6]) that the eigenvalues of P are $\lambda_1 = 1$, $\lambda_2 = 0.8$ and $\lambda_3 = 0.3$, and $P^* = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 \end{bmatrix}$. The three eigenvalues are distinct, leading to

$$P^n - P^* = 0.8^n \begin{bmatrix} 0.5 & 0 & -0.5 \\ 0.5 & 0 & -0.5 \\ -0.5 & 0 & 0.5 \end{bmatrix} + 0.3^n \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

as well as

$$P^{n+1} - P^n = 0.8^n \begin{bmatrix} -0.1 & 0 & 0.1 \\ -0.1 & 0 & 0.1 \\ 0.1 & 0 & -0.1 \end{bmatrix} + 0.3^n \begin{bmatrix} 0 & 0 & 0 \\ 0.7 & -0.7 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

As can be seen directly, in both cases the components $(1,2)$ and $(3,2)$ are zero for all n, whereas by Lemma 2.4 all other components follow BL. Hence, the Markov chain defined by the transition probability matrix P is Benford.

As will be observed later, the moduli of the eigenvalues as well as a specific rational relationship between them play a crucial role in the analysis of BL in Markov chains, similar to the results in [2].

Example 2.9. (Examples of non-Benford Markov chains)

(i) Let $d = 2$ and $P = \begin{bmatrix} 0.2 & 0.8 \\ 0.1 & 0.9 \end{bmatrix}$, hence $P^* = \frac{1}{9}\begin{bmatrix} 1 & 8 \\ 1 & 8 \end{bmatrix}$ and, for every $n \in \mathbb{N}$,

$$P^n - P^* = \frac{0.1^n}{9}\begin{bmatrix} 8 & -8 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad P^{n+1} - P^n = 0.1^n \begin{bmatrix} -0.8 & 0.8 \\ 0.1 & -0.1 \end{bmatrix}.$$

Since $\log 0.1$ is rational, Lemma 2.4 implies that no component of $(P^n - P^*)$ or $\left(P^{n+1} - P^n\right)$ is Benford. For example, $D_1\left(|(P^n - P^*)^{(1,1)}|\right) = 8$ for all $n \in \mathbb{N}$.

(ii) Let $d = 3$ and $P = \begin{bmatrix} 0.0 & 0.1 & 0.9 \\ 0.1 & 0.3 & 0.6 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$. The eigenvalues of P are $\lambda_1 = 1$, $\lambda_2 = 0.2$ and $\lambda_3 = -0.1$. Since these three eigenvalues are distinct, again by spectral decomposition,

$$P^n - P^* = \frac{0.2^n}{8}\begin{bmatrix} 0 & -1 & 1 \\ 0 & 7 & -7 \\ 0 & -1 & 1 \end{bmatrix} + \frac{(-0.1)^n}{11}\begin{bmatrix} 10 & 0 & -10 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix},$$

as well as

$$P^{n+1} - P^n = 0.2^n \begin{bmatrix} 0 & 0.1 & -0.1 \\ 0 & -0.7 & 0.7 \\ 0 & 0.1 & -0.1 \end{bmatrix} + (-0.1)^n \begin{bmatrix} -1 & 0 & 1 \\ 0.1 & 0 & -0.1 \\ 0.1 & 0 & -0.1 \end{bmatrix}.$$

The first column of B_2 is zero, hence for that column the relevant eigenvalue is $\lambda_3 = -0.1$. Since $\log 0.1$ is rational, no component in the first column of either sequence $(P^n - P^*)$ and $\left(P^{n+1} - P^n\right)$ follows BL, i.e., P is not Benford.

3 Sufficient condition that a Markov chain is Benford

To analyze the behavior of the sequences $(P^n - P^*)$ and $\left(P^{n+1} - P^n\right)$ associated with a Markov chain, a nonresonance condition on P will be helpful. Recall that real numbers $x_1, \ldots, x_k$ are *rationally independent* (or $\mathbb{Q}$-*independent*) if $\sum_{j=1}^{k} q_j x_j = 0$ with $q_1, \ldots, q_k \in \mathbb{Q}$ implies that $q_j = 0$ for all $j = 1, \ldots, k$; otherwise $x_1, \ldots, x_k$ are *rationally dependent*.

Definition 3.1. A stochastic matrix P is *nonresonant* if every nonempty subset $\Lambda_0 = \{\lambda_{i_1}, \ldots, \lambda_{i_k}\} \subset \sigma(P)^+ \setminus \{\lambda_1\}$ with $|\lambda_{i_1}| = \ldots = |\lambda_{i_k}| = L_0$ satisfies $\#(\Lambda_0 \cap \mathbb{R}) \leq 1$, and the numbers 1, $\log L_0$ and the elements of $\frac{1}{2\pi} \arg \Lambda_0$ are rationally independent, where

$$\tfrac{1}{2\pi} \arg \Lambda_0 := \left\{ \tfrac{1}{2\pi} \arg \lambda_{i_1}, \ldots, \tfrac{1}{2\pi} \arg \lambda_{i_k} \right\} \setminus \left\{0, \tfrac{1}{2}\right\} .$$

A Markov chain is nonresonant whenever its transition probability matrix is. A stochastic matrix or Markov chain is *resonant* if it is not nonresonant.

Notice that for P to be nonresonant, it is required specifically that the logarithms of the moduli of all the eigenvalues other than $\lambda_1 = 1$ are irrational; in particular, P has to be invertible. Theorem A below establishes that nonresonance is sufficient for P to be Benford. There is a close correspondence between Definition 3.1 of a nonresonant matrix and the notion of a matrix not having 10-*resonant spectrum*, as introduced in [2]. The main difference is that the eigenvalue $\lambda_1 = 1$ is excluded in Definition 3.1, whereas every stochastic matrix has 10-resonant spectrum.

Example 3.2. (Examples of nonresonant matrices)

(i) Both transition matrices in Example 2.8 are nonresonant.

(ii) Let $d = 5$ and $P = \begin{bmatrix} 0.0 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.0 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.0 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.0 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.0 \end{bmatrix}$. The eigenvalues of P

are $\lambda_1 = 1$ and $\lambda_2 = -0.25$ (with multiplicity four), so $\Lambda_0 = \{-0.25\}$, with $L_0 = 0.25$ and $\frac{1}{2\pi} \arg \Lambda_0 = \emptyset$. Since $\log 0.25$ is irrational, P is nonresonant.

Example 3.3. (Examples of resonant matrices)

(i) Two real eigenvalues of opposite sign: Let $d = 3$ and $P = \begin{bmatrix} 0.6 & 0.4 & 0.0 \\ 0.8 & 0.0 & 0.2 \\ 0.0 & 0.6 & 0.4 \end{bmatrix}$.
The eigenvalues of P are $\lambda_1 = 1$ and $\lambda_{2,3} = \pm\sqrt{0.2}$. Notice that $\log|\lambda_2| = \log|\lambda_3| = -\frac{1}{2}\log 5$ is irrational. With $\Lambda_0 = \{\sqrt{0.2}, -\sqrt{0.2}\}$ clearly $\#(\Lambda_0 \cap \mathbb{R}) = 2$, hence P is resonant. The spectral decomposition (4) yields

$$(P^n - P^*)^{(1,1)} = 0.2\lambda_2^n + 0.2\lambda_3^n = \begin{cases} 0.4\left(\sqrt{0.2}\right)^n & \text{if } n \text{ is even,} \\ 0 & \text{if } n \text{ is odd,} \end{cases}$$

showing that P is not Benford either.

(ii) Eigenvalues with rational logarithms: Let $d = 3$ and $P = \begin{bmatrix} 0.0 & 0.1 & 0.9 \\ 0.5 & 0.1 & 0.4 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$.
The eigenvalues are $\lambda_1 = 1$ and $\lambda_{2,3} = -0.25 \pm 0.05i\sqrt{15}$. Since $\log|\lambda_{2,3}| = -0.5$ is rational, the matrix P is resonant.

(iii) Eigenvalues with rational argument: Let $d = 3$ and $P = \begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.7 & 0.2 \end{bmatrix}$.
The eigenvalues are $\lambda_1 = 1$ and $\lambda_{2,3} = \pm 0.2i$. Note that $\log|0.2i| = -1 + \log 2$ is irrational, but $\frac{1}{2\pi}\arg(0.2i) = \frac{1}{4}$ is rational. Thus P is resonant. Spectral decomposition gives $B_1^{(2,2)} = B_2^{(2,2)} = \frac{1}{4}$, hence

$$(P^n - P^*)^{(2,2)} = \tfrac{1}{4}\left((0.2i)^n + (-0.2i)^n\right) = \begin{cases} \frac{1}{2} \cdot (-1)^{n/2} \cdot 0.2^n & \text{if } n \text{ is even,} \\ 0 & \text{if } n \text{ is odd,} \end{cases}$$

which in turn shows that P is not Benford.

(iv) Eigenvalues leading to rational dependencies within $\{1, \log L_0\} \cup \frac{1}{2\pi}\arg\Lambda_0$: Let

$d = 7$ and $P = \begin{bmatrix} 0.2 & 0.1 & 0.0 & 0.0 & 0.1 & 0.0 & 0.6 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.2 & 0.0 & 0.4 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.2 & 0.3 \\ 0.0 & 0.2 & 0.3 & 0.0 & 0.2 & 0.0 & 0.3 \\ 0.1 & 0.2 & 0.1 & 0.1 & 0.0 & 0.1 & 0.4 \\ 0.2 & 0.0 & 0.2 & 0.1 & 0.1 & 0.0 & 0.4 \\ 0.1 & 0.2 & 0.2 & 0.0 & 0.0 & 0.0 & 0.5 \end{bmatrix}$. The characteristic polynomial ψ_P of P factors as

$$\psi_P(\lambda) = (\lambda - 1)\ \left(\lambda^2 + 0.1\lambda - 0.01\right)\ \left(\lambda^2 - 0.01(2-i)\right)\ \left(\lambda^2 - 0.01(2+i)\right)\ .$$

The roots of the second factor are $-\frac{1}{20}\left(1 \pm \sqrt{5}\right)$; the third factor has roots

$$\pm\tfrac{1}{10}\sqrt{2-i} = \pm\tfrac{1}{20}\left(\sqrt{4+2\sqrt{5}} - i\sqrt{-4+2\sqrt{5}}\right)\ ,$$

and the fourth factor has roots

$$\pm\tfrac{1}{10}\sqrt{2+i} = \pm\tfrac{1}{20}\left(\sqrt{4+2\sqrt{5}} + i\sqrt{-4+2\sqrt{5}}\right)\ .$$

Thus, the dominated positive spectrum is

$$\sigma(P)^+ \setminus \{\lambda_1\} = \tfrac{1}{20}\left\{-(\sqrt{5}+1),\ \sqrt{5}-1,\ -2\sqrt{2-i},\ 2\sqrt{2+i}\right\}.$$

Clearly, the logarithms of the absolute values of the two real eigenvalues are irrational. The four non-real eigenvalues all have the same modulus $L_0 = \frac{1}{10}5^{1/4}$ (different from the two real eigenvalues), and $\log L_0 = -1 + \frac{1}{4}\log 5$ is irrational. Let $\Lambda_0 = \frac{1}{10}\left\{-\sqrt{2-i},\ \sqrt{2+i}\right\}$. Notice that $\arg(2\mp i) = \mp\arctan\frac{1}{2}$, so

$$\tfrac{1}{2\pi}\arg\Lambda_0 = \left\{\tfrac{1}{2} - \tfrac{1}{4\pi}\arctan\tfrac{1}{2},\ \tfrac{1}{4\pi}\arctan\tfrac{1}{2}\right\} =: \{x_3, x_4\}\,.$$

Since

$$-1\cdot 1 + 0\cdot\log L_0 + 2\cdot x_3 + 2\cdot x_4 = 0\,,$$

the elements of $\{1, \log L_0\} \cup \frac{1}{2\pi}\Lambda_0$ are $\mathbb{Q}$-dependent, and hence P is resonant.

The first main theoretical result of this paper is

Theorem A. *Every nonresonant irreducible and aperiodic finite-state Markov chain is Benford.*

The proof of Theorem A makes use of the following

Lemma 3.4. *Let $m \in \mathbb{N}$ and assume that $1, \rho_0, \rho_1, \ldots, \rho_m$ are $\mathbb{Q}$-independent, (z_n) is a convergent sequence in $\mathbb{C}$, and at least one of the $2m$ numbers $c_1, \ldots, c_{2m} \in \mathbb{C}$ is non-zero. Then, for every $\alpha \in \mathbb{R}$, the sequence*

$$\left(n\rho_0 + \alpha\log n + \log|\xi_n|\right) \tag{5}$$

is u.d. mod 1*, where*

$$\xi_n := c_1 e^{2\pi i n\rho_1} + c_2 e^{-2\pi i n\rho_1} + \ldots + c_{2d-1}e^{2\pi i n\rho_m} + c_{2d}e^{-2\pi i n\rho_m} + z_n.$$

Proof. Follows directly as in the proof of [2, Lemma 2.9] which considers $\log|\mathfrak{Re}\xi_n|$ in (5). $\square$

Proof of Theorem A. By Lemma 2.6(i), $\lim_{n\to\infty} P^n = P^*$ exists for the Markov chain defined by P. Fix $(i,j) \in \{1,\ldots,d\}^2$. As the analysis of $(P^{n+1}-P^n)^{(i,j)}$ is completely analogous, only $(P^n - P^*)^{(i,j)}$ will be considered here. If $(P^n - P^*)^{(i,j)}$ as given by (3) is not equal to zero for all but finitely many n, let $s_{i,j} \in \{1,\ldots,s\}$ be the minimal index such that $C_{s_{i,j}}^{(i,j)} \neq 0$. As in [2, p.224], to analyze (3), distinguish two cases.

Case 1: $|\lambda_{s_{i,j}}| > |\lambda_{s_{i,j}+1}|$.
In this case $\lambda_{s_{i,j}}$ is a *dominant* eigenvalue, and it is real since otherwise its conjugate would be an eigenvalue with the same modulus. Equation (3) can be written as

$$\begin{aligned}(P^n - P^*)^{(i,j)} &= \sum\nolimits_{\ell=s_{i,j}}^{d} \lambda_\ell^n C_\ell^{(i,j)} = |\lambda_{s_{i,j}}|^n \, n^{k_{s_{i,j}}^{(i,j)}} \sum\nolimits_{\ell=s_{i,j}}^{d} \left(\frac{\lambda_\ell}{|\lambda_{s_{i,j}}|}\right)^n \frac{C_\ell^{(i,j)}}{n^{k_{s_{i,j}}^{(i,j)}}} \\ &= |\lambda_{s_{i,j}}|^n \, n^{k_{s_{i,j}}^{(i,j)}} \left(c_{s_{i,j}}^{(i,j)} \left(\frac{\lambda_{s_{i,j}}}{|\lambda_{s_{i,j}}|}\right)^n + \zeta_{i,j}(n)\right),\end{aligned}$$

where

$$c_{s_{i,j}}^{(i,j)} := \lim\nolimits_{n\to\infty} n^{-k_{s_{i,j}}^{(i,j)}} C_{s_{i,j}}^{(i,j)} \neq 0\,,$$

and $\zeta_{i,j}(n) \to 0$ as $n \to \infty$ because $\lambda_{s_{i,j}}$ is a dominating eigenvalue. Therefore,

$$\log\left|(P^n - P^*)^{(i,j)}\right| = n \log|\lambda_{s_{i,j}}| + k_{s_{i,j}}^{(i,j)} \log n + \log|c_{s_{i,j}}^{(i,j)}| + \eta_n\,,$$

with $\eta_n = \log\left|1 + \zeta_{i,j}(n) e^{-in\arg\lambda_{s_{i,j}}} / c_{s_{i,j}}^{(i,j)}\right|$. Since $\eta_n \to 0$ and $\log|\lambda_{s_{i,j}}|$ is irrational, the sequence $(P^n - P^*)^{(i,j)}$ is Benford by Lemma 2.2 and the fact that $(x_n + \alpha\log n + \beta)$ is u.d. mod 1 whenever (x_n) is (e.g. [2, Lem. 2.8]).

Case 2: $|\lambda_{s_{i,j}}| = |\lambda_{s_{i,j}+1}| = \ldots = |\lambda_{t_{i,j}}| =: |\lambda_{i,j}|$ for some $t_{i,j} > s_{i,j}$.
Here several *different* eigenvalues of the same magnitude occur, such as e.g. conjugate pairs of non-real eigenvalues. Let $k^{(i,j)}$ be the maximal degree of the polynomials $C_\ell^{(i,j)}$, $\ell = s_{i,j},\ldots,t_{i,j}$. As in Case 1, express (3) as

$$(P^n - P^*)^{(i,j)} = |\lambda_{i,j}|^n n^{k^{(i,j)}} \left(c_{s_{i,j}}^{(i,j)} \left(\frac{\lambda_{s_{i,j}}}{|\lambda_{s_{i,j}}|}\right)^n + \ldots + c_{t_{i,j}}^{(i,j)} \left(\frac{\lambda_{t_{i,j}}}{|\lambda_{t_{i,j}}|}\right)^n + \zeta_{i,j}(n)\right),$$

where $c_\ell^{(i,j)} := \lim_{n\to\infty} n^{-k^{(i,j)}} C_\ell^{(i,j)} \in \mathbb{C}$ for $\ell = s_{i,j}, \ldots, t_{i,j}$, with $c_\ell^{(i,j)} \neq 0$ for at least one ℓ, and $\zeta_{i,j}(n) \to 0$ as $n \to \infty$. Consequently,

$$\log \left|(P^n - P^*)^{(i,j)}\right| = n \log|\lambda_{i,j}| + k^{(i,j)} \log n \\ + \log \left| c_{s_{i,j}}^{(i,j)} \left(\frac{\lambda_{s_{i,j}}}{|\lambda_{s_{i,j}}|}\right)^n + \ldots + c_{t_{i,j}}^{(i,j)} \left(\frac{\lambda_{t_{i,j}}}{|\lambda_{t_{i,j}}|}\right)^n + \zeta_{i,j}(n) \right| .$$

Write λ_ℓ as $\lambda_\ell = |\lambda_{i,j}| e^{i \arg \lambda_\ell}$ for $\ell = s_{i,j}, \ldots, t_{i,j}$, and hence

$$\log \left|(P^n - P^*)^{(i,j)}\right| = n \log|\lambda_{i,j}| + k^{(i,j)} \log n \\ + \log \left| c_{s_{i,j}}^{(i,j)} e^{in \arg \lambda_{s_{i,j}}} + \ldots + c_{t_{i,j}}^{(i,j)} e^{in \arg \lambda_{t_{i,j}}} + \zeta_{i,j}(n) \right| .$$

Since P is nonresonant, Lemma 3.4 applies with $m = t_{i,j} - s_{i,j} + 1$ and $\rho_0 = \log|\lambda_{i,j}|$, $\rho_1 = \frac{1}{2\pi} \arg \lambda_{s_{i,j}}, \ldots, \rho_m = \frac{1}{2\pi} \arg \lambda_{t_{i,j}}$. Thus $(P^n - P^*)^{(i,j)}$ is Benford. $\square$

Example 3.5. (The general two-dimensional case)
Let $d = 2$ and $P = \begin{bmatrix} 1-x & x \\ y & 1-y \end{bmatrix}$ with $x, y \in (0,1)$. By Feller [10, p. 432],

$$P^n = \frac{1}{x+y} \begin{bmatrix} y & x \\ y & x \end{bmatrix} + \frac{(1-x-y)^n}{x+y} \begin{bmatrix} x & -x \\ -y & y \end{bmatrix} , \tag{6}$$

from which it is clear that $\lambda_1 = 1$, $\lambda_2 = 1 - x - y$, and $P^* = \dfrac{1}{x+y} \begin{bmatrix} y & x \\ y & x \end{bmatrix}$. It follows from (6) that each component of $(P^n - P^*)$ and $(P^{n+1} - P^n)$ is a multiple of (λ_2^n). By Theorem A, the Markov chain with transition probability matrix P is Benford whenever $\log|1-x-y|$ is irrational. On the other hand, by Lemma 2.4 P is not Benford if $\log|1-x-y| \in \mathbb{Q}$. Thus for $d = 2$, nonresonance is (not only sufficient but also) necessary for P to be Benford. For $d \geq 3$, this is no longer true, see Example 3.7 below.

Example 3.6. (The general three-dimensional case)
Let $d = 3$ and $P = \begin{bmatrix} x_1 & x_2 & 1-x_1-x_2 \\ y_1 & y_2 & 1-y_1-y_2 \\ z_1 & z_2 & 1-z_1-z_2 \end{bmatrix}$, where $x_1, x_2, y_1, y_2, z_1, z_2 \in (0,1)$ are such that $x_1 + x_2, y_1 + y_2, z_1 + z_2$ all lie between 0 and 1. Solving the characteristic equation yields the eigenvalues $\lambda_1 = 1$ and $\lambda_{2,3} = a \pm \sqrt{a^2 - b}$, with

$$a = \tfrac{1}{2}(x_1 + y_2 - z_1 - z_2) \quad \text{and} \quad b = x_1 y_2 - x_1 z_2 + y_1 z_2 - x_2 y_1 + x_2 z_1 - y_2 z_1 .$$

Furthermore, using

$$c = 1 - y_2 + z_1 - y_2 z_1 + x_2(-y_1 + z_1) + x_1(-1 + y_2 - z_2) + z_2 + y_1 z_2 \neq 0\,,$$

one finds that

$$P^* = \frac{1}{c}\begin{bmatrix} z_1 - y_2 z_1 + y_1 z_2 & x_2 z_1 + z_2 - x_1 z_2 & 1 - x_1 - x_2 y_1 - y_2 + x_1 y_2 \\ z_1 - y_2 z_1 + y_1 z_2 & x_2 z_1 + z_2 - x_1 z_2 & 1 - x_1 - x_2 y_1 - y_2 + x_1 y_2 \\ z_1 - y_2 z_1 + y_1 z_2 & x_2 z_1 + z_2 - x_1 z_2 & 1 - x_1 - x_2 y_1 - y_2 + x_1 y_2 \end{bmatrix}.$$

If $a^2 \neq b$, then $P^n - P^* = \lambda_2^n B_2 + \lambda_3^n B_3$, where B_ℓ for $\ell = 2, 3$ are as in (4). There are two cases to consider:

(i) $\underline{a^2 > b.}$

In this case, $\lambda_{2,3}$ are real, and the dominant eigenvalue must be identified. If $a > 0$, then $|\lambda_2| > |\lambda_3|$, hence λ_2 is dominant. If $B_2^{(i,j)} \neq 0$ for all $(i, j) \in \{1, 2, 3\}^2$, then the Markov chain defined by P is Benford if $\log|\lambda_2|$ is irrational. In case there also exists (i, j) with $B_2^{(i,j)} = 0$ yet $B_3^{(i,j)} \neq 0$, then for P to be Benford $\log|\lambda_3|$ has to be irrational as well. For $a < 0$ the roles of λ_2 and λ_3 have to be interchanged. If $a = 0$, then P is resonant but may still be Benford, see Example 3.7(ii).

(ii) $\underline{a^2 < b.}$

Here $\lambda_{2,3}$ are conjugate and non-real, with $|\lambda_2| = |\lambda_3| = \sqrt{b}$. Thus P is nonresonant if and only if the numbers $1, \frac{1}{2}\log b, \frac{1}{2\pi}\arctan\sqrt{b/a^2 - 1}$ are $\mathbb{Q}$-independent.

Finally, if $a^2 = b$ then $\lambda_2 = \lambda_3 = a$, so P is Benford whenever $\log|a|$ is irrational.

The next example shows that for a Markov chain to be Benford, nonresonance is not necessary in general.

Example 3.7. (Markov chains that are resonant yet Benford)

(i) Eigenvalues with rational argument: Let $d = 3$ and $P = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.4 & 0.3 & 0.3 \\ 0.6 & 0.1 & 0.3 \end{bmatrix}$.

The eigenvalues are $\lambda_1 = 1$ and $\lambda_{2,3} = \pm 0.2i$. With $\Lambda_0 = \{0.2i\}$ therefore $\frac{1}{2\pi}\arg\Lambda_0 = \{\frac{1}{4}\} \subset \mathbb{Q}$, so P is resonant. However, spectral decomposition shows that $B_3 = \overline{B_2}$, i.e., B_2, B_3 are conjugates, and each component of B_2 has non-zero real *and* imaginary part. Thus for every $(i,j) \in \{1,2,3\}^2$,

$$\left|(P^n - P^*)^{(i,j)}\right| = \left|2\mathfrak{Re}(0.2i)^n B_2^{(i,j)}\right| = \begin{cases} 2 \cdot 0.2^n \left|\mathfrak{Re} B_2^{(i,j)}\right| & \text{if } n \text{ is even,} \\ 2 \cdot 0.2^n \left|\mathfrak{Im} B_2^{(i,j)}\right| & \text{if } n \text{ is odd,} \end{cases}$$

and $(P^n - P^*)^{(i,j)}$ is Benford.

(ii) Two real eigenvalues of opposite sign: Let $d = 3$ and $P = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.7 & 0.2 & 0.1 \\ 0.4 & 0.2 & 0.4 \end{bmatrix}$. The eigenvalues are $\lambda_1 = 1$ and $\lambda_{2,3} = \pm 0.3$. It can be checked that each component of $B_2 \pm B_3$ is non-zero. Thus for every $(i,j) \in \{1,2,3\}^2$,

$$(P^n - P^*)^{(i,j)} = 0.3^n \left(B_2^{(i,j)} + (-1)^n B_3^{(i,j)} \right),$$

which is Benford because $\log 0.3 \notin \mathbb{Q}$.

Remarks on general Markov chains:

(i) Theorem A can not be applied to Markov chains that fail to be irreducible. However, every finite-state Markov chain can be decomposed into classes of recurrent and transient states. Hence, the transition matrix P can be block-partitioned as

$$P = \begin{bmatrix} P_1 & 0 & \cdots & 0 & 0 \\ 0 & P_2 & 0 & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & P_r & 0 \\ B^{(1)} & B^{(2)} & \cdots & B^{(r)} & A \end{bmatrix},$$

where $P_1, P_2, \ldots, P_r$ are the transition matrices of the r disjoint recurrent classes, and $B^{(1)}, B^{(2)}, \ldots, B^{(r)}$ denote the transition probabilities from the collection of transient states into each recurrent class. As $n \to \infty$,

$$P^n = \begin{bmatrix} P_1^n & 0 & \cdots & 0 & 0 \\ 0 & P_2^n & 0 & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & P_r^n & 0 \\ L_n^{(1)} & L_n^{(2)} & \cdots & L_n^{(r)} & A^n \end{bmatrix} \to \begin{bmatrix} P_1^* & 0 & \cdots & 0 & 0 \\ 0 & P_2^* & 0 & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & P_r^* & 0 \\ SB^{(1)}P_1^* & SB^{(2)}P_2^* & \cdots & SB^{(r)}P_r^* & 0 \end{bmatrix},$$

where $L_n^{(j)} = \sum_{\ell=0}^{n-1} A^\ell B^{(j)} P_j^{n-\ell-1}$ for $j = 1, 2, \ldots, r$, and $S = \sum_{k=0}^{\infty} A^k$. Theorem A can be applied separately to the transition matrices P_j associated with the recurrent classes. Consequently, if $P_1, P_2, \ldots, P_r$ are Benford, then the corresponding components of P are also Benford. Additionally, if A is nonresonant, then that part follows BL as well. The only remaining parts are formed by the sequences $\big(L_n^{(j)}\big)$ and depend on the (nonautonomous) summation of the powers of A. Their Benford properties are beyond the scope of this paper.

(ii) For an irreducible Markov chain that is not aperiodic, but rather periodic with period $p > 1$, Definition 2.7 still makes sense, provided that P^* is understood as the unique row-stochastic matrix with $P^*P = P^*$. However, such a chain cannot be Benford since for every $(i,j) \in \{1,\ldots,d\}^2$ one can choose $k \in \{0,\ldots,p-1\}$ such that

$$|(P^n - P^*)^{(i,j)}| = (P^*)^{(i,j)} > 0\,, \quad \forall n \in \mathbb{N}\backslash(k + p\mathbb{N})\,.$$

Similarly, each component of $(P^{n+1} - P^n)$ equals zero at least $(p-2)/p$ of the time and thus cannot be Benford either whenever $p \geq 3$. The distribution of significands of $(P^{n+1} - P^n)^{(i,j)}$ observed in this situation is a convex combination of BL and a pure point mass, see [5, Cor. 6]. Only in the case $p = 2$ is it possible for each component of $(P^{n+1} - P^n)$ to be either Benford or eventually zero.

(iii) Although this paper deals with finite-state Markov chains only, it is worth noting that chains with *infinitely* many states may also obey BL in one way or the other. For a very simple example, let $0 < \rho < 1$ and consider the homogeneous random walk on $\mathbb{Z}$ with

$$P^{(i,j)} = \begin{cases} \rho^2 & \text{if } j = i-1\,, \\ 2\rho(1-\rho) & \text{if } j = i\,, \\ (1-\rho)^2 & \text{if } j = i+1\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Clearly, this Markov chain is irreducible and aperiodic. It is (null-)recurrent if $\rho = \frac{1}{2}$, and transient otherwise. For all $(i,j) \in \mathbb{Z}^2$ and $n \in \mathbb{N}$,

$$(P^n)^{(i,j)} = \begin{pmatrix} 2n \\ n+i-j \end{pmatrix} \rho^{n+i-j}(1-\rho)^{n-i+j}\,,$$

and an application of Stirling's formula shows that $(P^n)^{(i,j)}$ is Benford if and only if $\log\big(4\rho(1-\rho)\big)$ is irrational. For all but countably many ρ, therefore, $(P^n)^{(i,j)}$ is Benford for every (i,j). Note that one of the excluded values is $\rho = \frac{1}{2}$, i.e. the recurrent case. For *recurrent* chains virtually every imaginable behavior of significant digits or significands can be manufactured by means of advanced ergodic theory tools, see [3] and the references therein.

4 Almost all Markov chains are Benford

The second main theoretical objective of this paper is to show that Benford behavior is typical in finite-state Markov chains. Indeed, if the transition probabilities of the chain are chosen at random, independently and in any continuous manner, then the chain almost always, i.e. with probability one, obeys BL. To formulate this more precisely, the following terminology will be used.

Definition 4.1. A *random* (*d-state*) *Markov chain* is a random $d \times d$-matrix $\boldsymbol{P}$, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $\mathcal{P}_d$, i.e., each row $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d$ of $\boldsymbol{P}$ is a random vector taking values in the standard d-simplex

$$\Delta_d := \left\{ (x_1, \ldots, x_d) \in \mathbb{R}^d : x_j \geq 0 \text{ for all } 1 \leq j \leq d, \text{ and } \sum\nolimits_{j=1}^{d} x_j = 1 \right\}.$$

A random vector $\boldsymbol{X} : \Omega \to \Delta_d$ is *continuous* if its distribution on Δ_d is continuous w.r.t. the (normalised) Lebesgue measure on Δ_d, that is, if $\mathbb{P}(\boldsymbol{X} \in A) = 0$ whenever $A \subset \Delta_d$ is a nullset.

With this terminology, it is the purpose of the present section to illustrate and prove

Theorem B. *If the transition probabilities (i.e. the rows) of a random Markov chain* $\boldsymbol{P}$ *are independent and continuous, then* $\boldsymbol{P}$ *is Benford with probability one.*

Before giving a full proof for Theorem B, the special case of a random two-state chain will be examined to show how independence and continuity together allow the application of Theorem A. The case $d = 2$ is especially transparent since the eigenvalue functions are simple and explicit, unlike for the general case where the eigenvalues are only known implicitly, and the Implicit Function Theorem has to be resorted to.

Example 4.2. Consider the random two-state Markov chain

$$\boldsymbol{P} = \begin{bmatrix} 1 - \boldsymbol{X} & \boldsymbol{X} \\ \boldsymbol{Y} & 1 - \boldsymbol{Y} \end{bmatrix},$$

where the random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are i.i.d. (absolutely) continuous random variables on the unit interval $[0,1]$. Since $\boldsymbol{X}$ and $\boldsymbol{Y}$ are continuous, each of the four entries of $\boldsymbol{P}$ is strictly positive with probability one, so the chain is irreducible and aperiodic with probability one. Since $\boldsymbol{P}$ is random, the second-largest eigenvalue is a random variable $\boldsymbol{Z}$ which, by Example 3.5, satisfies $\boldsymbol{Z} = 1 - \boldsymbol{X} - \boldsymbol{Y}$. Since $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent and continuous, $\boldsymbol{Z}$ is also continuous, and hence the probability that $\boldsymbol{Z}$ is in any given countable set is zero. But this implies that the probability of $\log|\boldsymbol{Z}|$ being rational is zero, which in turn shows that with probability one, $\boldsymbol{P}$ is nonresonant, and hence Benford, by Theorem A.

Similarly to the analysis of Newton's method in [4], a key property in the present Markov chain setting is the *real-analyticity* of certain functions, notably the eigenvalue functions. Recall that a function $f : U \to \mathbb{C}$ is *real-analytic* whenever it can, in the neighborhood of every point in its domain U (an open subset of $\mathbb{R}^\ell$ for some $\ell \geq 1$), be written as a convergent power series. Clearly, every real-analytic function is C^∞, i.e. has derivatives of all orders. An important property of real-analytic functions not shared by arbitrary $\mathbb{C}$-valued C^∞-functions defined on U is that the set $\{x \in U : f(x) = 0\}$ is a nullset unless f vanishes identically on U.

The proof of Theorem B will be based on several preliminary results. First, given $a = (a_1, \ldots, a_d) \in \mathbb{C}^d$, let $p_a : \mathbb{C} \to \mathbb{C}$ denote the polynomial

$$p_a(z) = z^d + a_1 z^{d-1} + \ldots + a_{d-1} z + a_d \,.$$

By the Fundamental Theorem of Algebra, p_a has exactly d zeroes (counted with multiplicities). If p_a and p_a', or more generally, if p_a and p_b with $a \neq b$ have a common zero then a universal polynomial relation must necessarily be satisfied by a and b. Only a special case of this elementary fact is required here, and since no reference is known to the authors, a proof is included for completeness.

Lemma 4.3. *For every integer $d > 1$, there exists a non-trivial polynomial Q_d in $2d-1$ variables with the following property: Whenever $a = (a_1, \ldots, a_d) \in \mathbb{C}^d$, $b = (b_1, \ldots, b_{d-1}) \in \mathbb{C}^{d-1}$, and $p_a(z_0) = p_b(z_0) = 0$ for some $z_0 \in \mathbb{C}$, then $Q_d(a,b) := Q_d(a_1, \ldots, a_d, b_1, \ldots, b_{d-1}) = 0$.*

Proof. For $d = 2$, let $Q_2(a,b) := a_1 b_1 - a_2 - b_1^2$ for all $a = (a_1, a_2) \in \mathbb{C}^2$ and $b = b_1 \in \mathbb{C}$. To see that Q_2 has the desired property, note that if $p_a(z_0) = 0 = p_b(z_0)$, then

$z_0^2 + a_1 z_0 + a_2 = 0$ and $z_0 = -b_1$, hence $Q_2(a,b) = 0$. Assume now that Q_d has already been constructed. For every $a \in \mathbb{C}^{d+1}$ and $b \in \mathbb{C}^d$ let $\rho = a_2 - b_2 - (a_1 - b_1)b_1 \in \mathbb{C}$, as well as

$$c = \big(a_3 - b_3 - (a_1 - b_1)b_2, \ldots, a_d - b_d - (a_1 - b_1)b_d, a_{d+1} - (a_1 - b_1)b_d\big) \in \mathbb{C}^{d-1},$$

and define

$$Q_{d+1}(a,b) := \rho^{1+\deg Q_d} Q_d\left(b, \frac{c}{\rho}\right),$$

where $\deg\left(\sum_j c_j x_1^{n_{1,j}} x_2^{n_{2,j}} \ldots x_\ell^{n_{\ell,j}}\right) := \max\{n_{1,j} + \ldots + n_{\ell,j} : c_j \neq 0\}$. Clearly, Q_{d+1} is a polynomial in $2d+1$ variables, and $Q_{d+1} \neq 0$. If $p_a(z_0) = p_b(z_0) = 0$ for some $z_0 \in \mathbb{C}$, then

$$\begin{aligned} 0 &= p_a(z_0) - \big(z_0 + (a_1 - b_1)\big)p_b(z_0) \\ &= \sum\nolimits_{j=1}^{d-1} (a_{j+1} - b_{j+1} - (a_1 - b_1)b_i)\, z_0^{d-j} + a_{d+1} - (a_1 - b_1)b_d\,. \end{aligned} \tag{7}$$

If $\rho = 0$, then clearly $Q_{d+1}(a,b) = 0$. Otherwise, it is easy to check that (7) implies $p_{c/\rho}(z_0) = 0$, in which case $Q_d(b, c/\rho) = 0$, by assumption. In either case, therefore, $Q_{d+1}(a,b) = 0$. □

Corollary 4.4. *For every integer $d > 1$, there exists a non-trivial polynomial Q_d^* in d variables such that $Q_d^*(a) = 0$ whenever $p_a(z_0) = p_a'(z_0) = 0$ for some $z_0 \in \mathbb{C}$.*

Proof. Take $Q_d^* = Q_d(a,b)$ with $b = \big(\frac{d-1}{d}a_1, \frac{d-2}{d}a_2, \ldots, \frac{2}{d}a_{d-2}, \frac{1}{d}a_{d-1}\big)$. □

This corollary will now be used to show that if a stochastic matrix P_0 is invertible and has distinct non-zero eigenvalues, then all stochastic matrices P sufficiently close to P_0 also are invertible and have distinct non-zero eigenvalues. In fact, these eigenvalues are real-analytic functions of P. To formulate this efficiently, for every $P_0 \in \mathcal{P}_d$ and $\varepsilon > 0$ denote by $B_\varepsilon(P_0)$ the open ball with radius ε centered at P_0, i.e. $B_\varepsilon(P_0) = \big\{P \in \mathcal{P}_d : |P^{(i,j)} - P_0^{(i,j)}| < \varepsilon \text{ for all } 1 \le i,j \le d\big\}$.

Lemma 4.5. *Suppose $P_0 \in \mathcal{P}_d$ is invertible and has d distinct non-zero eigenvalues. Then there exists $\varepsilon > 0$ and and $d-1$ non-constant real-analytic functions $\lambda_2, \ldots, \lambda_d : B_\varepsilon(P_0) \to \mathbb{C}$ such that, for every $P \in B_\varepsilon(P_0)$,*

(i) *$1, \lambda_2(P), \ldots, \lambda_d(P)$ are the eigenvalues of P, and $\lambda_2(P) \cdot \ldots \cdot \lambda_d(P) \neq 0$;*

(ii) $\lambda_i(P) \neq \overline{\lambda_j(P)}$ *whenever* $i \neq j$, *unless* $\lambda_i = \overline{\lambda_j}$ *on* $B_\varepsilon(P_0)$.

Proof. Note first that by the continuity of $(P, z) \mapsto \det(zI_{d\times d} - P) = \psi_P(z)$, there exists $\delta > 0$ such that every $P \in B_\delta(P_0)$ is invertible and has distinct non-zero eigenvalues. Thus the characteristic polynomial ψ_P of P has $d-1$ distinct non-zero roots different from 1. Let z_0 be one of those roots. Since z_0 is a simple root, $\psi'_{P_0}(z_0) \neq 0$, so by the Implicit Function Theorem [15, Theorem 2.3.5], z_0 depends real-analytically on the coefficients of ψ_P which themselves are real-analytic (in fact polynomial) functions of the entries of P. More formally, there exists $\varepsilon \leq \delta$ and a real-analytic function $g : B_\varepsilon(P_0) \to \mathbb{C}$ with $g(P_0) = z_0$ such that $\psi_P\big(g(P)\big) = 0$ for all $P \in B_\varepsilon(P_0)$. Overall, there exists $\varepsilon > 0$ and $d-1$ real-analytic functions $\lambda_i : B_\varepsilon(P_0) \to \mathbb{C}$ satisfying (i); note that $\lambda_1 \equiv 1$ by Lemma 2.6. To see that $\lambda_2, \ldots, \lambda_d$ are not constant on $B_\varepsilon(P_0)$, suppose by way of contradiction that $\lambda_i(P) = \lambda_i(P_0) \neq 1$ for some $2 \leq i \leq d$ and all $P \in B_\varepsilon(P_0)$. In this case, the real-analytic function $P \mapsto \psi_P\big(\lambda_i(P_0)\big)$ vanishes identically on $B_\varepsilon(P_0)$, and hence on all of $\mathcal{P}_d$. Since $I_{d\times d} \in \mathcal{P}_d$, this obviously contradicts $\psi_{I_{d\times d}}\big(\lambda_i(P_0)\big) = \big(\lambda_i(P_0) - 1\big)^d \neq 0$. Consequently, none of the functions $\lambda_2, \ldots, \lambda_d : B_\varepsilon(P_0) \to \mathbb{C}$ is constant.

To show (ii), assume that $\lambda_i(P_1) = \overline{\lambda_j(P_1)}$ for some $i \neq j$ and $P_1 \in B_\varepsilon(P_0)$. Thus $\lambda_i(P_1) \in \mathbb{C}\backslash\mathbb{R}$, since if $\lambda_i(P_1)$ were real, then $\lambda_i(P_1) = \lambda_j(P_1)$, which is impossible since the eigenvalues are distinct. Since all matrices in $\mathcal{P}_d$ are *real*, their non-real eigenvalues occur in conjugate pairs. Hence, for all P sufficiently close to P_1, the number $\overline{\lambda_j(P)}$ is an eigenvalue of P which, by continuity, can only be $\lambda_i(P)$. Consequently, λ_i and $\overline{\lambda_j}$ coincide locally near P_1 and therefore, by real-analyticity, on all of $B_\varepsilon(P_0)$. $\square$

By means of the above auxiliary results, several almost sure properties of random Markov chains can be identified.

Lemma 4.6. *If the rows of the random Markov chain* $\boldsymbol{P}$ *are independent and continuous then, with probability one,*

(i) $\boldsymbol{P}$ *is irreducible, aperiodic, and invertible;*

(ii) $\boldsymbol{P}$ *has* d *distinct non-zero eigenvalues; and*

(iii) $\boldsymbol{P}$ *is nonresonant.*

Proof. Fix $\boldsymbol{P}$ and assume its rows $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d$ are independent and continuous.

(i) Since each $\boldsymbol{X}_i$ is continuous, $\mathbb{P}(\boldsymbol{X}_i \in A) = 0$ for every Lebesgue nullset $A \subset \Delta_d$, so in particular $\mathbb{P}(\boldsymbol{X}_{i,j} \in \{0,1\}) = 0$ for all i and j. With probability one, therefore, $\boldsymbol{P}^{(i,j)} \in (0,1)$ for all i and j, and $\boldsymbol{P}$ is irreducible and aperiodic. To see that $\boldsymbol{P}$ is almost surely invertible, note that $P \mapsto \det P$ is a non-constant, real-analytic function on $\mathcal{P}_d$. With $N = \big\{(x_1, \ldots, x_d) \in \Delta_d \times \ldots \times \Delta_d : \det(x_1, \ldots, x_d) = 0\big\}$,

$$\begin{aligned}\mathbb{P}(\det \boldsymbol{P} = 0) &= \int_N \mathrm{d}\mathbb{P}(x_1, \ldots, x_d) = \int \cdots \int_N \mathrm{d}\mathbb{P}(x_1) \ldots \mathrm{d}\mathbb{P}(x_d) \\ &= \int \cdots \int \left(\int_N \mathrm{d}\mathbb{P}(x_1) \right) \mathrm{d}\mathbb{P}(x_2) \ldots \mathrm{d}\mathbb{P}(x_d) = 0\,,\end{aligned}$$

where the second equality follows from the independence of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d$, the third from Fubini's theorem, and the fourth from the continuity of the $\boldsymbol{X}_i$.

(ii) There exist d non-constant polynomial functions $q_1, \ldots, q_d : \mathcal{P}_d \to \mathbb{R}$ such that

$$\psi_P(z) = \det(zI_{d\times d} - P) = z^d + q_1(P)z^{d-1} + \ldots + q_{d-1}(P)z + q_d(P)$$

holds for all $P \in \mathcal{P}_d$ and $z \in \mathbb{C}$; for example, $q_1(P) = -\sum_{i=1}^d P^{(i,i)}$ and $q_d(P) = (-1)^d \det P$. Consequently, $q(P) := Q_d^*\big(q_1(P), \ldots, q_d(P)\big)$ defines a non-constant real-analytic (in fact, polynomial) map $q : \mathcal{P}_d \to \mathbb{R}$, and since z_0 is a multiple eigenvalue of P if and only if $\psi_P(z_0) = \psi_P'(z_0) = 0$, Corollary 4.4 implies that

$$\big\{P \in \mathcal{P}_d : P \text{ has multiple eigenvalues}\big\} \subset \big\{P \in \mathcal{P}_d : q(P) = 0\big\}\,.$$

As before, by Fubini's Theorem $\mathbb{P}(q(\boldsymbol{P}) = 0) = 0$, showing that with probability one all eigenvalues of $\boldsymbol{P}$ are simple.

(iii) For every $\rho \in \mathbb{Q}$ define the real-analytic auxiliary function $\Phi_\rho : \mathbb{R}^2 \to \mathbb{R}$ by $\Phi_\rho(x) := (x_1^2 + x_2^2 - 10^{2\rho})^2$, and also $\Theta : \mathbb{R}^4 \to \mathbb{R}$ as $\Theta(x) := \big(x_1^2 + x_2^2 - x_3^2 - x_4^2\big)^2$. By (i) and (ii), $\boldsymbol{P}$ almost surely satisfies the hypotheses of Lemma 4.5, so let P_0, ε, and $\lambda_2, \ldots, \lambda_d$ be as in Lemma 4.5, and define real-analytic functions $\Phi_{\rho,i}$ and $\Theta_{i,j}$ on $B_\varepsilon(P_0)$ as

$$\Phi_{\rho,i}(P) := \Phi_\rho\big(\mathfrak{Re}\lambda_i(P), \mathfrak{Im}\lambda_i(P)\big) = \big(|\lambda_i(P)|^2 - 10^{2\rho}\big)^2\,, \quad \forall i : 2 \le i \le d\,,$$

and, for all $2 \le i, j \le d$,

$$\Theta_{i,j}(P) := \Theta\big(\mathfrak{Re}\lambda_i(P), \mathfrak{Im}\lambda_i(P), \mathfrak{Re}\lambda_j(P), \mathfrak{Im}\lambda_j(P)\big) = \big(|\lambda_i(P)|^2 - |\lambda_j(P)|^2\big)^2\,.$$

Finally, let $F_\rho : B_\varepsilon(P_0) \to \mathbb{R}$ be defined as

$$F_\rho(P) := \prod_{i=2}^{d} \Phi_{\rho,i}(P) \cdot \prod_{2\le i<j:\lambda_i \neq \overline{\lambda_j}} \Theta_{i,j}(P) \,.$$

The definition of F_ρ becomes transparent upon noticing that $F_\rho(P) = 0$ for some $\rho \in \mathbb{Q}$ whenever P is invertible and resonant. Next, it will be shown that F_ρ does not vanish identically on $B_\varepsilon(P_0)$. To see this, note first that if $P \in B_\varepsilon(P_0)$, then also $(1-\delta)P + \delta I_{d\times d} \in B_\varepsilon(P_0)$ for all sufficiently small $\delta > 0$. Moreover, if $\Phi_{\rho,i}(P) = 0$ for some $i = 2, \ldots, d$, then

$$\begin{aligned}\Phi_{\rho,i}\big((1-\delta)P + \delta I_{d\times d}\big) &= \Big(\big((1-\delta)\mathfrak{Re}\lambda_i(P) + \delta\big)^2 + (1-\delta)^2\mathfrak{Im}\lambda_i(P)^2 - 10^{2\rho}\Big)^2 \\ &= \delta^2\Big((2-\delta)\left(\mathfrak{Re}\lambda_i(P) - |\lambda_i(P)|^2\right) + \delta\big((1-\mathfrak{Re}\lambda_i(P)\big)\Big)^2 > 0\,,\end{aligned}$$

provided that $\delta > 0$ is small enough. (Recall that $1 - \mathfrak{Re}\lambda_i(P) > 0$ whenever $P \in B_\varepsilon(P_0)$.) Similarly, if $\Theta_{i,j}(P) = 0$ for some $2 \le i < j \le d$ with $\lambda_i \neq \overline{\lambda_j}$ and $\lambda_i(P) \neq 0$, then a short calculation confirms that, for all $\delta > 0$ sufficiently small,

$$\Theta_{i,j}\big((1-\delta)P + \delta I_{d\times d}\big) = \delta^2(1-\delta)^2 \frac{|\lambda_i(P) - \lambda_j(P)|^2 |\lambda_i(P) - \overline{\lambda_j(P)}|^2}{|\lambda_i(P)|^2} > 0\,.$$

Overall, F_ρ does not vanish identically on $B_\varepsilon(P_0)$. As every $P \in B_\varepsilon(P_0)$ is invertible,

$$\big\{P \in B_\varepsilon(P_0) : P \text{ is resonant}\,\big\} \subset \bigcup_{\rho\in\mathbb{Q}} \big\{P \in B_\varepsilon(P_0) : F_\rho(P) = 0\big\}\,.$$

Since F_ρ is real-analytic and non-constant, $\big\{P \in B_\varepsilon(P_0) : F_\rho(P) = 0\big\}$ is a nullset for every $\rho \in \mathbb{Q}$, and so is $\bigcup_{\rho\in\mathbb{Q}}\big\{P \in B_\varepsilon(P_0) : F_\rho(P) = 0\big\}$. Analogously to (i) and (ii), therefore, $\mathbb{P}(\boldsymbol{P} \text{ is resonant}\,) = 0$. □

Proof of Theorem B. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d$ denote the random transition probabilities (row vectors) of the random $d\times d$-matrix $\boldsymbol{P}$. If $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d$ are independent and continuous, then by Lemma 4.6, $\boldsymbol{P}$ is almost surely irreducible, aperiodic, and nonresonant. By Theorem A, this implies that $\boldsymbol{P}$ is Benford with probability one. □

Remark 4.7. (i) It is clear that without independence, or without continuity, Lemma 4.6 and Theorem B are generally false. For example, for the conclusion of Lemma 4.6 to hold it is not enough to assume that the distribution on Δ_d of each row of $\boldsymbol{P}$ is atomless. As very simple examples show, under this weaker assumption, $\boldsymbol{P}$ may, with positive probability, be reducible and have multiple or zero eigenvalues. Even if Lemma 4.6 (i,ii) hold with probability one, $\boldsymbol{P}$ may still be resonant and not Benford. To see this, consider the random three-state Markov chain

$$\boldsymbol{P} = \frac{1}{40}\begin{bmatrix} \boldsymbol{X}+4 & \boldsymbol{X} & 36-2\boldsymbol{X} \\ \boldsymbol{Y} & \boldsymbol{Y}+4 & 36-2\boldsymbol{Y} \\ \boldsymbol{Z}+2 & \boldsymbol{Z}+2 & 36-2\boldsymbol{Z} \end{bmatrix},$$

where $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ are independent and uniformly distributed on $[0, 1]$. The eigenvalues of $\boldsymbol{P}$ are

$$\lambda_1 = 1\,, \quad \lambda_2 = 0.1\,, \quad \lambda_3 = \tfrac{1}{40}(\boldsymbol{X}+\boldsymbol{Y}-2\boldsymbol{Z})\,.$$

Note that $|\lambda_3| \leq 0.05 < \lambda_2$. Clearly, $\boldsymbol{P}$ is resonant with probability one, and Lemma 4.6(iii) fails. Perhaps even more importantly, Theorem B fails as well since, as spectral decomposition shows, $B_2 \neq 0$ with probability one and hence $\mathbb{P}(\boldsymbol{P} \text{ is Benford}) = 0$.

(ii) With hardly any effort, the tools employed in the proof of Lemmas 4.5 and 4.6 also yield a topological analogue of Theorem B: Within the compact metric space $\mathcal{P}_d$, the matrices that are irreducible, aperiodic, invertible and nonresonant form a *residual* set, that is, a set whose complement is the countable union of nowhere dense sets. Being Benford, therefore, is a typical property for $P \in \mathcal{P}_d$ not only under a probabilistic perspective but under a topological perspective as well.

5 Simulations

In this section, numerical simulations will illustrate the theoretical results of previous sections, and based on these simulations the *rate of convergence* towards BL will be discussed. Since it is not possible to observe the empirical frequencies of infinite sequences, $(P^n - P^*)$ and $(P^{n+1} - P^n)$ are simulated up to a predefined value of n, such as $n = 1000$ or $n = 10000$, and the empirical distributions of first significant digits of each component are compared to the Benford probabilities. For some Markov chains, simulations up to $n = 1000$ yield empirical frequencies very close to BL, whereas for others even $n = 10000$ does not give a good approximation, although theoretically all chains considered here follow BL. Thus, convergence rates towards BL may differ significantly.

Example 5.1.
From Table 1, it is clear that the sequences (2^n), $(n!)$, (F_n) give different empirical frequencies for the simulation up to $n = 1000$. Compared to the other two, (F_n) gives empirical frequencies much closer to BL.

Similarly, rates of convergence can be discussed for Markov chains. The important question is what property is creating the difference in convergence rates. Theorem B shows that every homogeneous Markov chain chosen independently and continuously is Benford with probability one. Besides irreducibility and aperiodicity, nonresonance is crucial. Irreducibility and aperiodicity do not determine the rate of convergence. This leaves nonresonance as the only source for different rates of convergence. According to Definition 3.1, nonresonance is based on the rational independence of 1, $\log L_0$ and the elements of $\frac{1}{2\pi}\arg \Lambda_0$, provided that $\Lambda_0 \neq \emptyset$. Thus, it is natural to expect this rational independence to be reflected in some quantitative manner in the rate of convergence towards BL.

It is well known that there are infinitely many rational approximations for a given accuracy to any irrational number. Let x be an irrational number. Given any $\varepsilon > 0$, there exist infinitely many pairs $(p, q) \in \mathbb{Z} \times \mathbb{N}$ with $\gcd(p, q) = 1$ and

$$\left| x - \frac{p}{q} \right| < \varepsilon .$$

One way to obtain rational approximations of irrational numbers is provided by the method of continued fractions. Every irrational real number x is represented uniquely by its continued fraction expansion

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}} ,$$

also denoted as $x = [a_0; a_1, a_2, a_3, \ldots]$, where $a_0 \in \mathbb{Z}$ and $a_n \in \mathbb{N}$ for $n \geq 1$ are referred to as the *partial quotients* of x. By [11, Theorem 149], if p_n and q_n are defined iteratively as

$$p_0 = a_0 , \quad p_1 = a_1 a_0 + 1 , \qquad p_n = a_n p_{n-1} + p_{n-2} , \quad \forall n \geq 2 ,$$
$$q_0 = 1 , \quad q_1 = a_1 , \qquad q_n = a_n q_{n-1} + q_{n-2} , \quad \forall n \geq 2 ,$$

then, for all $n \in \mathbb{N}$,

$$\frac{p_n}{q_n} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_n}}}} =: [a_0; a_1, \ldots, a_n]\,;$$

the rational numbers p_n/q_n are called the *convergents* of the continued fraction of x. Leaving aside trivial exceptions, best rational approximations to an irrational x are of the form p_n/q_n, and

$$\left| x - \frac{p_n}{q_n} \right| < \frac{1}{a_{n+1} q_n^2}\,, \quad \forall n \geq 2. \tag{8}$$

It is clear from (8) that p_n/q_n yields a particularly good approximation of x when a_{n+1} is large. Hence x can be rapidly approximated if its continued fraction expansion contains a sequence of rapidly increasing partial quotients. On the other hand, if (a_n) does not grow fast (or at all), then it is difficult to approximate x by a rational number with small error, see [11, 16] for details. For example, [16, Ch. 2, Theorem 3.4] asserts that if (a_n) is bounded for some x then the distribution mod 1 of (nx) approaches the uniform distribution rather quickly. Thus irrationals which are hard to approximate by rational numbers, due to a small upper bound on, or slow growth of (a_n), are also the ones for which one expects to see fast convergence to Benford probabilities. Specifically, for the golden ratio $\frac{1+\sqrt{5}}{2} = [1; 1, 1, 1, \ldots]$, every a_n has the smallest possible value. Since $\left| \log F_n - n \log \frac{1+\sqrt{5}}{2} \right| \to 0$ as $n \to \infty$, this may explain why the convergence to BL is faster for the Fibonacci sequence than for the other two sequences in Example 5.1. (See [17] for further insights on BL for continued fractions.)

It is important to note that (a_n) is unbounded for almost every x, [11, Theorem 196]. Hence, in most simulations it is not possible to observe convergence as fast as for the Fibonacci sequence. However, to highlight the difference in rates of convergence and irrationality, two examples are studied. The first 50 partial quotients are given for every relevant irrational number that arises.

Example 5.2. (Markov chain showing *fast* convergence)
Let $d = 3$ and $P = \begin{bmatrix} 0.25 & 0.35 & 0.40 \\ 0.30 & 0.45 & 0.25 \\ 0.65 & 0.15 & 0.20 \end{bmatrix}$. The eigenvalues of P are $\lambda_1 = 1$ and

$\lambda_{2,3} = -\frac{1}{20} \mp \frac{1}{20}\sqrt{21}$, hence $\sigma(P)^+ \setminus \{\lambda_1\} = \left\{-\frac{1}{20} - \frac{1}{20}\sqrt{21}, -\frac{1}{20} + \frac{1}{20}\sqrt{21}\right\}$. Since $\log|\lambda_2|$ and $\log|\lambda_3|$ are irrational and different, P is nonresonant. Thus Theorem A implies that the Markov chain defined by P is Benford.

Table 2 shows the empirical frequencies of significant digits for the first 1000 and 10000 terms of $(P^n - P^*)$, respectively; the behavior of $(P^{n+1} - P^n)$ is very similar.

$(1,1)$	$(2,1)$	$(3,1)$	$(1,2)$	$(2,2)$	$(3,2)$	$(1,3)$	$(2,3)$	$(3,3)$	Benford
0.300	0.301	0.300	0.303	0.303	0.299	0.300	0.306	0.300	0.30103
0.175	0.177	0.177	0.176	0.174	0.176	0.178	0.174	0.175	0.17609
0.126	0.124	0.123	0.125	0.125	0.125	0.124	0.124	0.127	0.12493
0.098	0.096	0.100	0.096	0.096	0.097	0.096	0.098	0.097	0.09691
0.078	0.081	0.079	0.080	0.080	0.079	0.079	0.078	0.077	0.07918
0.068	0.067	0.065	0.068	0.067	0.066	0.068	0.067	0.069	0.06694
0.058	0.059	0.059	0.056	0.057	0.060	0.059	0.057	0.058	0.05799
0.050	0.050	0.051	0.051	0.052	0.052	0.050	0.050	0.052	0.05115
0.047	0.045	0.046	0.045	0.046	0.046	0.046	0.046	0.045	0.04575
0.3008	0.3009	0.3009	0.3011	0.3012	0.3008	0.3011	0.3017	0.3010	0.30103
0.1761	0.1762	0.1764	0.1762	0.1758	0.1762	0.1763	0.1759	0.1760	0.17609
0.1249	0.1250	0.1247	0.1248	0.1251	0.1249	0.1249	0.1249	0.1250	0.12493
0.0971	0.0968	0.0972	0.0969	0.0968	0.0970	0.0968	0.0969	0.0970	0.09691
0.0792	0.0793	0.0791	0.0792	0.0793	0.0790	0.0790	0.0790	0.0789	0.07918
0.0668	0.0669	0.0666	0.0670	0.0670	0.0668	0.0672	0.0671	0.0673	0.06694
0.0582	0.0582	0.0582	0.0580	0.0578	0.0582	0.0580	0.0577	0.0579	0.05799
0.0510	0.0509	0.0512	0.0510	0.0512	0.0514	0.0510	0.0512	0.0513	0.05115
0.0459	0.0458	0.0457	0.0458	0.0458	0.0457	0.0457	0.0456	0.0456	0.04575

Table 2: *Comparing empirical frequencies for the first significant digits with Benford probabilities for the first* 1000 *(top half) and* 10000 *(bottom half) terms of the sequences* $(P^n - P^*)^{(i,j)}$*, where* P *is the transition probability matrix in Example 5.2.*

Since $|\lambda_2| > |\lambda_3|$, all that matters is how well

$$\log|\lambda_2| = [-1;2,4,8,1,5,1,6,3,1,2,2,1,1,2,1,1,2,1,66,5,1,1,2,1,3,\\ 1,2,1,1,3,1,3,2,3,2,7,3,86,1,1,1,1,1,26,3,1,5,3,1,5,\ldots]$$

is approximated by rational numbers. From the above, $a_n \leq 86$ for all $1 \leq n \leq 50$, and a rapid increase of quotients is not observed. This continued fraction expansion should be compared to the ones in the example below.

Example 5.3. (Markov chain showing *slow* convergence)

Let $d = 3$ and $P = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.4 \\ 0.4 & 0.0 & 0.6 \end{bmatrix}$, with eigenvalues $\lambda_1 = 1$ and $\lambda_{2,3} = \frac{7}{20} \pm \frac{1}{20}\sqrt{3}\, i$. Thus $\sigma(P)^+ \setminus \{\lambda_1\} = \{\frac{7}{20} + \frac{1}{20}\sqrt{3}\ i\} =: \Lambda_0$, and the behavior of significant digits is governed by the two irrational numbers

$$\log|\lambda_2| = [-1;1,1,3,1,7,1,15,1,2,1,1,7,1,6,2,1,3,1,1,2,4,1,1,2,3,\\ 8,1,2,1,1,2,1,2,1,7,1,1,2,1,33,1,2,1,2,1,1,11,1,24,8,\ldots],$$

$$\tfrac{1}{2\pi}\arg\lambda_2 = [0;25,1,9,3,168,2,1,1,32,1,6,3,1,9,1,1,92,2,13,2,1,1,10,2,5,\\ 1,3,1,1,1,1,3,1,2,7,1,5,1,1,4,1,3,14,3,10,1,1,3,1,3,\ldots].$$

Note that $\max_{n=1}^{50} a_n = 33$ for $\log|\lambda_2|$, whereas $\max_{n=1}^{50} a_n = 168$ for $\frac{1}{2\pi}\arg\lambda_2$. When compared with Example 5.2, the repeated early high values in the continued fraction expansion of $\frac{1}{2\pi}\arg\lambda_2$ suggest a somewhat slower convergence to BL. As shown in Table 3, this slower convergence is clearly recognizable in simulations of $(P^n - P^*)$; again the behavior of $(P^{n+1} - P^n)$ is very similar.

6 Applications

In scientific calculations using digital computers and floating point arithmetic, round-off errors are inevitable, and as Knuth points out in his classic text *The Art of Computer Programming* [14, pp. 253–255],

> In order to analyze the average behavior of floating-point arithmetic algorithms (and in particular to determine their average running time), we need some statistical information that allows us to determine how often various cases arise ... [If, for example, the] leading digits tend to be small [that] makes the most obvious techniques of average error estimation for floating-point calculations invalid. The relative error due to rounding is usually ... more than expected.

Thus for the problem of numerical estimation of P^* from P^n, it is important to study the distribution of significant digits (or, equivalently, the fraction parts of floating-point numbers) of the components of $(P^n - P^*)$ and $(P^{n+1} - P^n)$.

(1,1)	(2,1)	(3,1)	(1,2)	(2,2)	(3,2)	(1,3)	(2,3)	(3,3)	Benford
0.302	0.313	0.311	0.327	0.290	0.286	0.293	0.298	0.297	0.30103
0.176	0.169	0.170	0.152	0.178	0.181	0.192	0.181	0.184	0.17609
0.127	0.137	0.136	0.137	0.110	0.114	0.103	0.122	0.122	0.12493
0.096	0.081	0.085	0.087	0.101	0.101	0.123	0.105	0.102	0.09691
0.075	0.079	0.080	0.086	0.093	0.091	0.061	0.071	0.074	0.07918
0.074	0.080	0.084	0.072	0.055	0.056	0.061	0.063	0.069	0.06694
0.072	0.049	0.048	0.046	0.055	0.054	0.061	0.083	0.074	0.05799
0.039	0.047	0.043	0.046	0.056	0.055	0.070	0.038	0.041	0.05115
0.039	0.045	0.043	0.047	0.062	0.062	0.036	0.039	0.037	0.04575
0.2998	0.3150	0.3158	0.3167	0.2910	0.2922	0.2938	0.2982	0.2981	0.30103
0.1798	0.1620	0.1610	0.1570	0.1865	0.1867	0.1877	0.1816	0.1821	0.17609
0.1312	0.1397	0.1399	0.1354	0.1069	0.1079	0.1090	0.1232	0.1236	0.12493
0.0943	0.0828	0.0837	0.0859	0.1002	0.0983	0.1192	0.1033	0.1027	0.09691
0.0716	0.0825	0.0825	0.0965	0.0877	0.0887	0.0640	0.0702	0.0698	0.07918
0.0753	0.0789	0.0782	0.0610	0.0570	0.0561	0.0600	0.0682	0.0694	0.06694
0.0665	0.0476	0.0478	0.0496	0.0550	0.0546	0.0618	0.0748	0.0741	0.05799
0.0416	0.0458	0.0462	0.0478	0.0575	0.0570	0.0680	0.0412	0.0409	0.05115
0.0399	0.0457	0.0449	0.0501	0.0582	0.0585	0.0365	0.0393	0.0393	0.04575

Table 3: *Comparing empirical frequencies for the first significant digits with Benford probabilities for the first* 1000 *(top half) and* 10000 *(bottom half) terms of the sequences* $(P^n - P^*)^{(i,j)}$*, where* P *is the transition probability matrix in Example 5.3.*

Theorem B above shows that the components of both $(P^n - P^*)$ and $(P^{n+1} - P^n)$ typically exhibit exactly the type of nonuniformity of significant digits alluded to by Knuth: Not only do the first few significant digits of the differences between the components of the successive n-step transition matrices P^n and the limiting distribution P^*, as well as the differences between P^{n+1} and P^n tend to be small but, much more specifically, they typically follow BL.

This prevalence of BL has important practical implications for estimating P^* from P^n using floating-point arithmetic. One type of error in scientific calculations is overflow (or underflow), which occurs when the running calculations exceed the largest (or smallest, in absolute value) floating-point number allowed by the computer. Feldstein and Turner show that [9, p. 241], "[u]nder the assumption of the logarithmic distribution of numbers (i.e., BL) floating-point addition and subtraction can result in overflow and underflow with alarming frequency . . . ". Together with Theorem B, this suggests that special attention should be given to overflow and underflow errors in any computer algorithm used to estimate P^* from P^n.

Another important type of error in scientific computing is due to roundoff. In estimating P^* from P^n, for example, every stopping rule, such as "stop when n=1000" or "stop when the components in $(P^{n+1} - P^n)$ are less than 10^{-10}", will result in some error, and Theorem B shows that this difference is generally Benford. In fact, justified by heuristics and by the extensive empirical evidence of BL in other numerical calculations, analysis of roundoff errors has often been carried out under the *hypothesis* of a logarithmic statistical distribution (cf. [9, p. 326]). Therefore, as Knuth pointed out, a naive assumption of uniformly distributed significands in the calculations tends to underestimate the average relative roundoff error in cases where the actual statistical distribution of fraction parts is skewed toward smaller leading significant digits, as is the case in BL. To obtain a rough idea of the magnitude of this underestimate when the true statistical distribution is BL, let $\boldsymbol{X}$ denote the absolute roundoff error at the time of stopping the algorithm, and let $\boldsymbol{Y}$ denote the fraction part of the approximation at the time of stopping. Then the relative error is $\boldsymbol{X}/\boldsymbol{Y}$, and assuming that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent random variables, the average (i.e., expected) relative error is simply $\mathbb{E}\boldsymbol{X} \cdot \mathbb{E}(1/\boldsymbol{Y})$. Thus if $\boldsymbol{Y}$ is assumed to be uniformly distributed on $[1, 10)$, ignoring the fact that $\boldsymbol{Y}$ is Benford creates an average underestimation of the relative error by *more than one third* (cf. [4]).

As one potential application of Theorems A and B, it should be possible to adapt the current plethora of BL-based goodness-of-fit statistical tests for detecting fraud (e.g. [7]), to the problem of detecting whether or not a sequence of realizations of a finite-state process originates from a Markov chain, i.e., whether or not the process is Markov. By Theorem B, conformance with BL for the differences $P^{n+1} - P^n$ is typical in finite-state Markov chains, so a standard (e.g. chi-squared) goodness-of-fit to BL of the empirical estimates of the differences between P^{n+1} and P^n may help detect non-Markov behavior.

Acknowledgements

T.P. Hill wishes to express his gratitude to the Vrije Universiteit Amsterdam, and to the Econometrics Department, especially Professor Henk Tijms, for their hospitality

during several visits when part of this research was conducted.

References

[1] Berger, A., Bunimovich, L.A. and Hill, T. 2005. One-dimensional Dynamical Systems and Benford's Law, *Transactions of the American Mathematical Society* 357, pp. 197–219.

[2] Berger, A. 2005. Multi-dimensional Dynamical Systems and Benford's Law, *Discrete and Continuous Dynamical Systems* 13, pp. 219–237.

[3] Berger, A. 2010. Some dynamical properties of Benford sequences, to appear in *Journal of Difference Equations and Applications.*

[4] Berger, A. and Hill, T. 2007. Newton's Method obeys Benford's Law, *The American Mathematical Monthly* 114, pp. 588–602.

[5] Berger, A. and Siegmund, S. 2008. On the distribution of mantissae in nonautonomous difference equations, *Journal of Difference Equations and Applications* 13, pp. 829–845.

[6] Bremaud, P. 1999. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, Springer.

[7] Carslaw, C. 1988. Anomalies in income numbers. Evidence of goal oriented behavior, *The Accounting Review* 63, pp. 321–327.

[8] Diaconis, P. 1979. The distribution of leading digits and uniform distribution mod 1, *Annals of Probability* 5, pp. 72–81.

[9] Feldstein, A. and Turner, P. 1986. Overflow, underflow, and severe loss of significance in floating-point addition and subtraction, *IMA Journal of Numerical Analysis* 6, pp. 241–251.

[10] Feller, W. 1950. *An Introduction to Probability Theory and Its Applications*, Volume 1, 3rd ed., Wiley.

[11] Hardy, G.H. and Wright E. M. 1960. *An introduction to the theory of numbers*, 4th ed., Oxford University Press.

[12] Hill, T. 1995. Base invariance implies Benford's Law, *Proceedings of the American Mathematical Society* 123, pp. 887–895.

[13] Hill, T. 1996. A statistical derivation of the significant-digit law, *Statistical Science* 10, pp. 354–363.

[14] Knuth, D. 1997. *The Art of Computer Programming*, vol. 2, 3rd ed., Addison-Wesley, Reading, MA.

[15] Krantz, S. and Parks, H. 2002. *A Primer of Real Analytic Functions*, Birkhäuser.

[16] Kuipers, L. and Niederreiter, H. 1974. *Uniform Distribution of Sequences*, Wiley, New York.

[17] Schatte, P. 1990. On Benford's law for continued fractions, *Mathematische Nachrichten*, 148, pp. 139–144.

[18] Seneta, E. 1981. *Non-negative Matrices and Markov Chains*, 2nd ed., Springer, New York.

[19] Serre, D. 2002. *Matrices. Theory and Applications*, Springer, New York.