SIAM J. MATRIX ANAL. APPL.
Vol. 32, No. 3, pp. 665–684
© 2011 Society for Industrial and Applied Mathematics

FINITE-STATE MARKOV CHAINS OBEY BENFORD'S LAW*

ARNO BERGER[†], THEODORE P. HILL[‡], BAHAR KAYNAR[§], AND AD RIDDER[¶]

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 transition probability 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 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 probability matrix is chosen in an absolutely continuous manner, then the resulting Markov chain is Benford with probability one. Concrete examples illustrate the various cases that arise, and the theory is complemented with simulations and potential applications.

Key words. Markov chain, Benford's law, uniform distribution modulo 1, significant digits, significand, n-step transition probabilities, stationary distribution

AMS subject classifications. 11J71, 15B51, 60J22, 65C40

DOI. 10.1137/100789890

1. Introduction.

Benford's law (BL) is the widely known logarithmic probability distribution on significant digits. Its most familiar form is the special case of *leading* significant digits (base 10), namely,

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

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

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

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,

*Received by the editors March 23, 2010; accepted for publication (in revised form) by D. A. Bini March 29, 2011; published electronically July 21, 2011.
http://www.siam.org/journals/simax/32-3/78989.html

[†]Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, Canada (berger@ualberta.ca). Corresponding author. This author's research was supported by an NSERC Discovery Grant.

[‡]School of Mathematics, Georgia Institute of Technology, Atlanta (hill@math.gatech.edu). This author's research was supported by grants 400-06-044 and 040-11-091 from the Netherlands Organization for Scientific Research (NWO).

[§]Department of Econometrics and Operations Research, Vrije Universiteit, Amsterdam, The Netherlands (bkaynar@feweb.vu.nl). This author's research was supported by grants 400-06-044 and 040-11-091 from the NWO.

[¶]Department of Econometrics and Operations Research, Vrije Universiteit, Amsterdam, The Netherlands (arrider@feweb.vu.nl).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

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 given by (1).

D_1	(2^n)	$(n!)$	(F_n)	Benford
1	0.301	0.293	0.301	0.30103
2	0.176	0.176	0.177	0.17609
3	0.125	0.124	0.125	0.12494
4	0.097	0.102	0.096	0.09691
5	0.079	0.069	0.080	0.07918
6	0.069	0.087	0.067	0.06695
7	0.056	0.051	0.056	0.05799
8	0.052	0.051	0.053	0.05115
9	0.045	0.047	0.045	0.04576

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

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

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

Formally, for every $n \in \mathbb{N}$, $n \geq 2$, the number $D_n(x)$, the nth *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

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

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

The formal probability framework for BL is described in [13], [14]. The sample space is $\mathbb{R}^+$, and the σ-algebra of events is generated by the (decimal) *significand* (or *mantissa*) *function* S: $\mathbb{R}^+ \to [1, 10)$, where $S(x)$ is the unique number such that $x = 10^k S(x)$ 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

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

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 distributions 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.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

BL is now known to hold in great generality, e.g., for classical integer sequences such as (2^n), $(n!)$ and the Fibonacci numbers (F_n), iterations of linearly dominated or power-like maps, solutions of ordinary differential equations, products of independent random variables, random mixtures of data, and random maps (e.g., see [1], [4], [5], [9], [14]). Table 1 compares the empirical frequencies of D_1 for the first 1000 terms of the sequences (2^n), $(n!)$, and (F_n). This illustrates what it means to follow BL and also foreshadows the discussion in section 5.

The main contribution of this article is 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 12 and 17, respectively, show that 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. Several potential applications of the results are discussed, 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 sets 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 $\Re e z$, $\Im m z$, $\bar{z}$, and $|z|$, respectively. For $z \neq 0$, the argument $\arg z$ is the unique number in $(-\pi, \pi]$ that satisfies $z = |z| e^{\imath \arg z}$. 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 $(a(1), a(2), a(3), \dots)$ is denoted by $(a(n))$. Thus, for example, $(\alpha^n) = (\alpha^1, \alpha^2, \alpha^3, \dots)$ and $(P^{n+1} - P^n) = (P^2 - P^1, P^3 - P^2, P^4 - P^3, \dots)$. Boldface symbols indicate random(ized) quantities; e.g., $\boldsymbol{X}$ denotes a random variable or vector and $\boldsymbol{P}$ a random transition probability matrix.

Definition 1. *A sequence (x_n) of real numbers is Benford ("follows BL") if*

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

The central theme 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., [18]), recent results for BL in one- and multidimensional dynamical systems ([1], [2]), and the classical Perron–Frobenius theory for Markov chains; see, e.g., [6], [24]. Here and throughout, the term *uniformly distributed modulo* 1 is abbreviated as *u.d.* mod 1. The relationship between uniform distribution and BL is clarified by the following proposition.

Proposition 2 (see [9]). *A sequence (x_n) of real numbers is Benford if and only if $(\log |x_n|)$ is u.d.* mod 1.

An immediate consequence of Proposition 2 is the following useful fact.

Proposition 3 (see [1], [2]).

(i) *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.*

(ii) *If (x_n) is 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.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Propositions 2 and 3 are fundamental tools for analyzing BL in the setting of multidimensional dynamical systems. While the results in [2] do not apply directly to the Markov chain setting, the first part of the theory established below nevertheless relies heavily on those ideas, thereby adapting them to the case of row-stochastic matrices.

Example 4.

(i) The sequences (2^n), (0.2^n), (3^n), (0.3^n), $(0.01 \cdot 0.2^n + 0.2 \cdot 0.01^n)$ are Benford, whereas (10^n), (0.1^n), $(\sqrt{10}^n)$, $(0.1 \cdot 0.02^n + 0.02 \cdot 0.1^n)$ are not.

(ii) The sequence $(0.2^n + (-0.2)^n)$ is not Benford, since all odd terms are zero, but $(0.2^n + (-0.2)^n + 0.03^n)$ is Benford—although this does not follow directly from Proposition 3(i).

For every integer $d > 1$, denote the set of all row-stochastic matrices of size $d \times d$ by $\mathcal{P}_d$, and let $P \in \mathcal{P}_d$ be the transition probability matrix of a Markov chain. All Markov chains (or their associated matrices P) considered hereafter are assumed to be finite-state (with $d > 1$ states), irreducible, and aperiodic. Let $\lambda_1, \ldots, \lambda_s$, $s \leq d$, be the *distinct* (possibly nonreal) eigenvalues of 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 $\sigma(P)^+ = \{\lambda \in \sigma(P)\colon \Im m\lambda \geq 0\}$ forms the "upper half" of the spectrum. The usage of $\sigma(P)^+$ refers to the fact that nonreal eigenvalues of real matrices always occur in conjugate pairs, so the set $\sigma(P)^+$ includes only one of the conjugates. Without loss of generality, throughout this work the eigenvalues in $\sigma(P)$ are labeled such that

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_s|.$$

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 proposition summarizes some of the special properties of transition probability matrices for ease of reference; see, e.g., [23] for details.

PROPOSITION 5. *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 $n \in \mathbb{N}$,*

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

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 $k_\ell^{(i,j)} < d$.

The analysis is especially straightforward if all eigenvalues are simple, i.e., if $\#\sigma(P) = d$. In this case, for every $n \in \mathbb{N}$,

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

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 of this paper.

DEFINITION 6. *A Markov chain, or its associated transition probability matrix P, is* Benford *if each component of $(P^n - P^*)$ and $(P^{n+1} - P^n)$ is either Benford or eventually zero.*

Remark 7. The proof of Theorem 12 below will make it clear that requiring only the components of $(P^n - P^*)$ to be either Benford or eventually zero would result in an

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

equivalent definition. The more symmetric wording in Definition 6 merely reflects the fact that $(P^n - P^*)$, although of obvious theoretical importance, may not always be available in practice, whereas $(P^{n+1} - P^n)$ is.

The following examples illustrate the notions of Benford and non-Benford Markov chains. As will be observed later, the moduli of the eigenvalues as well as a specific rational relationship between them play a crucial role, similar to the results in [2].

Example 8 (a Benford Markov chain). Let

$$P = \begin{bmatrix} 0.9 & 0.0 & 0.1 \\ 0.6 & 0.3 & 0.1 \\ 0.1 & 0.0 & 0.9 \end{bmatrix}.$$

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, which leads 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 Proposition 3(i) all other components follow BL. Hence, the Markov chain defined by the transition probability matrix P is Benford.

Example 9 (a non-Benford Markov chain). Let

$$P = \begin{bmatrix} 0.0 & 0.1 & 0.9 \\ 0.1 & 0.3 & 0.6 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}.$$

The eigenvalues of P are $\lambda_1 = 1$, $\lambda_2 = 0.2$, and $\lambda_3 = -0.1$. The three eigenvalues are distinct, and

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

as well as

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

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

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 $(P^{n+1} - P^n)$ 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 $(P^{n+1} - P^n)$ 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 $\sum_{j=1}^k q_j x_j = 0$, where $q_1, \ldots, q_k \in \mathbb{Q}$ implies that $q_j = 0$ for all $j = 1, \ldots, k$; otherwise $x_1, \ldots, x_k$ are *rationally dependent.*

DEFINITION 10. *A stochastic matrix P is nonresonant if every nonempty subset $\Lambda = \{\lambda_{i_1}, \ldots, \lambda_{i_k}\}$ of $\sigma(P)^+ \setminus \{\lambda_1\}$ with $|\lambda_{i_1}| = \cdots = |\lambda_{i_k}| = L$ satisfies $\#(\Lambda \cap \mathbb{R}) \leq 1$, and the numbers 1, $\log L$ and the elements of $\frac{1}{2\pi} \arg \Lambda$ are rationally independent, where*

$$\frac{1}{2\pi} \arg \Lambda := \left\{ \frac{1}{2\pi} \arg \lambda_{i_1}, \ldots, \frac{1}{2\pi} \arg \lambda_{i_k} \right\} \Big\backslash \left\{ 0, \frac{1}{2} \right\}.$$

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 logarithm of the modulus of every eigenvalue other than $\lambda_1 = 1$ is irrational; in particular, every nonresonant P is invertible. The matrix in Example 8, for instance, is nonresonant. Theorem 12 below asserts that nonresonance is sufficient for P to be Benford. There is a close correspondence between Definition 10 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 10, whereas every stochastic matrix has 10-resonant spectrum. Also, it is worth pointing out that prior to [2] solutions of certain *linear recurrence relations* have been studied for their conformance with BL, e.g., second-order relations with a pair of nonreal eigenvalues in [22] and relations with integer coefficients and various restrictions on the eigenvalues in [19]. (See also [15] and the references therein.) Note that none of these results apply to stochastic matrices, i.e., in the setting of Definition 10.

Example 11 (examples of resonant matrices).

(i) Two real eigenvalues of opposite sign. Let

$$P = \begin{bmatrix} 0.6 & 0.4 & 0.0 \\ 0.8 & 0.0 & 0.2 \\ 0.0 & 0.6 & 0.4 \end{bmatrix}.$$

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 $\Lambda = \{\sqrt{0.2}, -\sqrt{0.2}\}$, clearly $\#(\Lambda \cap \mathbb{R}) = 2$; hence P is resonant. From

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

it is clear that P is not Benford either.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

(ii) Eigenvalues with rational logarithms. Let

$$P = \begin{bmatrix} 0.0 & 0.1 & 0.9 \\ 0.5 & 0.1 & 0.4 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}.$$

The eigenvalues are $\lambda_1 = 1$ and $\lambda_{2,3} = -0.25 \pm 0.05\imath\sqrt{15}$. Since $\log|\lambda_{2,3}| = -0.5$ is rational, the matrix P is resonant.

(iii) Eigenvalues with rational argument. Let

$$P = \begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.7 & 0.2 \end{bmatrix}.$$

The eigenvalues are $\lambda_1 = 1$ and $\lambda_{2,3} = \pm 0.2\imath$. Note that $\log|0.2\imath| = -1 + \log 2$ is irrational, but $\frac{1}{2\pi}\arg(0.2\imath) = \frac{1}{4}$ is rational. Thus P is resonant.

(iv) Eigenvalues leading to rational dependencies within $\{1, \log L\} \cup \frac{1}{2\pi}\arg\Lambda$. Let

$$P = \begin{bmatrix} 0.2 & 0.1 & 0.0 & 0.0 & 0.1 & 0.0 & 0.6 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.2 & 0.0 & 0.4 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.2 & 0.3 \\ 0.0 & 0.2 & 0.3 & 0.0 & 0.2 & 0.0 & 0.3 \\ 0.1 & 0.2 & 0.1 & 0.1 & 0.0 & 0.1 & 0.4 \\ 0.2 & 0.0 & 0.2 & 0.1 & 0.1 & 0.0 & 0.4 \\ 0.1 & 0.2 & 0.2 & 0.0 & 0.0 & 0.0 & 0.5 \end{bmatrix}.$$

The characteristic polynomial ψ_P of P factors as

$$\psi_P(\lambda) = (\lambda - 1)(\lambda^2 + 0.1\lambda - 0.01)(\lambda^2 - 0.01(2 - \imath))(\lambda^2 - 0.01(2 + \imath)),$$

which implies that

$$\sigma(P)^+ \setminus \{\lambda_1\} = \frac{1}{20}\{-(\sqrt{5}+1), \sqrt{5}-1, -2\sqrt{2-\imath}, 2\sqrt{2+\imath}\}.$$

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

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

Since

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

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

The first main theoretical result of this paper is as follows.

THEOREM 12. *Every nonresonant irreducible and aperiodic finite-state Markov chain is Benford.*

The proof of Theorem 12 makes use of the following lemma.

LEMMA 13. *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 nonzero. Then, for every $\alpha \in \mathbb{R}$, the sequence*

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

is u.d. mod 1, *where*

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

Proof. The proof follows directly as in the proof of [2, Lemma 2.9], which considers $\log|\Re e\xi_n|$ in (5). $\square$

Proof of Theorem 12. By Proposition 5(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. For notational convenience, for every $n \in \mathbb{N}$ denote the component (i, j) of $P^n - P^*$ by p_n. If p_n as given by (3) is not equal to zero for all but finitely many n, let $\sigma \in \{1, \ldots, s\}$ be the minimal index such that $C_\sigma^{(i,j)} \neq 0$. As in [2, p. 224], to analyze (3), distinguish two cases.

Case 1. $|\lambda_\sigma| > |\lambda_{\sigma+1}|$. In this case λ_σ 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

$$p_n = \sum_{\ell=\sigma}^{d} \lambda_\ell^n C_\ell^{(i,j)} = |\lambda_\sigma|^n n^{k_\sigma^{(i,j)}} \sum_{\ell=\sigma}^{d} \left(\frac{\lambda_\ell}{|\lambda_\sigma|}\right)^n \frac{C_\ell^{(i,j)}}{n^{k_\sigma^{(i,j)}}} = |\lambda_\sigma|^n n^{k_\sigma^{(i,j)}} \left(c_\sigma^{(i,j)} \left(\frac{\lambda_\sigma}{|\lambda_\sigma|}\right)^n + \zeta_{i,j}(n)\right),$$

where, since $k_\sigma^{(i,j)}$ is the degree of $C_\sigma^{(i,j)}$,

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

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

$$\log|p_n| = n \log|\lambda_\sigma| + k_\sigma^{(i,j)} \log n + \log|c_\sigma^{(i,j)}| + \eta_n,$$

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

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

$$p_n = |\lambda|^n n^{k^{(i,j)}} \left(c_\sigma^{(i,j)} \left(\frac{\lambda_\sigma}{|\lambda|}\right)^n + \cdots + c_\tau^{(i,j)} \left(\frac{\lambda_\tau}{|\lambda|}\right)^n + \zeta_{i,j}(n)\right),$$

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$$\log|p_n| = n\log|\lambda| + k^{(i,j)}\log n + \log\left|c_\sigma^{(i,j)}\left(\frac{\lambda_\sigma}{|\lambda|}\right)^n + \cdots + c_\tau^{(i,j)}\left(\frac{\lambda_\tau}{|\lambda|}\right)^n + \zeta_{i,j}(n)\right|.$$

Write λ_ℓ as $\lambda_\ell = |\lambda|e^{\imath \arg\lambda_\ell}$ for $\ell = \sigma, \ldots, \tau$, and hence

$$\log|p_n| = n\log|\lambda| + k^{(i,j)}\log n + \log|c_\sigma^{(i,j)}e^{\imath n\arg\lambda_\sigma} + \cdots + c_\tau^{(i,j)}e^{\imath n\arg\lambda_\tau} + \zeta_{i,j}(n)|.$$

Since P is nonresonant, Lemma 13 applies with $m = \tau - \sigma + 1$ and $\rho_0 = \log|\lambda|$, $\rho_1 = \frac{1}{2\pi}\arg\lambda_\sigma, \ldots, \rho_m = \frac{1}{2\pi}\arg\lambda_\tau$. Thus (p_n) is Benford. $\square$

Example 14 (the general two-dimensional case). Let

$$P = \begin{bmatrix} 1-x & x \\ y & 1-y \end{bmatrix}$$

with $x, y \in (0,1)$. By Feller [12, 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

$$P^* = \frac{1}{x+y}\begin{bmatrix} y & x \\ y & x \end{bmatrix}.$$

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 12, the Markov chain with transition probability matrix P is Benford whenever $\log|1-x-y|$ is irrational. On the other hand, by Proposition 3(i) 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, as the next example shows.

Example 15 (a resonant Benford Markov chain). Let

$$P = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.4 & 0.3 & 0.3 \\ 0.6 & 0.1 & 0.3 \end{bmatrix}.$$

The eigenvalues are $\lambda_1 = 1$ and $\lambda_{2,3} = \pm 0.2\imath$. With $\Lambda = \{0.2\imath\}$, therefore $\frac{1}{2\pi}\arg\Lambda = \{\frac{1}{4}\} \subset \mathbb{Q}$; hence P is resonant. However, spectral decomposition shows that $B_3 = \bar{B}_2$, i.e., B_2, B_3 are conjugates, and each component of B_2 has nonzero real *and* imaginary part. Thus for every $(i,j) \in \{1,2,3\}^2$,

$$|(P^n - P^*)^{(i,j)}| = |2\Re e(0.2\imath)^n B_2^{(i,j)}| = \begin{cases} 2\times\cdot\ 0.2^n|\Re e B_2^{(i,j)}| & \text{if } n \text{ is even,} \\ 2\times\cdot\ 0.2^n|Jm B_2^{(i,j)}| & \text{if } n \text{ is odd,} \end{cases}$$

and $(P^n - P^*)^{(i,j)}$ is Benford. Similarly, since each component of $5B_2 - \imath B_2$ has nonzero real and imaginary part, $(P^{n+1} - P^n)^{(i,j)}$ is Benford as well.

Remarks on general Markov chains.

(i) Theorem 12 cannot be applied to Markov chains that fail to be irreducible. However, every finite-state Markov chain can be decomposed into classes of

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

recurrent and transient states. Hence, the transition probability 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 probability matrices of the r disjoint recurrent classes and $B^{(1)}, B^{(2)}, \ldots, B^{(r)}$ denote the transition probability matrices 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 12 can be applied separately to the 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 $(L_n^{(j)})$ and depend on the (nonautonomous) summation of the powers of A. Their Benford properties are beyond the scope of this paper; see, e.g., [5].

(ii) For an irreducible Markov chain that is not aperiodic but rather periodic with period $p > 1$, Definition 6 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

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

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, Corollary 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.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

(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 $n \in \mathbb{N}$,

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

and an application of Stirling's formula shows that $(P^n)^{(i,j)}$ is Benford if and only if $\log(4\rho(1-\rho))$ 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 and in an absolutely continuous (a.c.) 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 16. *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 $\mathcal{P}_d$. A random Markov chain $\boldsymbol{P}\colon \Omega \to \mathcal{P}_d$ is a.c. if its distribution on $\mathcal{P}_d$ is a.c. w.r.t. $Leb_{\mathcal{P}_d}$, the normalized $d(d-1)$-dimensional volume on $\mathcal{P}_d \subset \mathbb{R}^{d\times d}$, that is, if $\mathbb{P}(\boldsymbol{P} \in A) = 0$ holds for $A \subset \mathcal{P}_d$ whenever $Leb_{\mathcal{P}_d}(A) = 0$.*

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

Theorem 17. *Every a.c. random Markov chain is Benford with probability one.*

Before giving a full proof for Theorem 17, the special case of a random two-state chain will be examined to show how the absolute continuity of $\boldsymbol{P}$ allows the application of Theorem 12. The case $d = 2$ is especially transparent since the eigenvalues are explicitly given by simple expressions, unlike for the general case where the eigenvalues are only known implicitly and the implicit function theorem has to be resorted to.

Example 18. Consider the random two-state Markov chain

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

and assume that the joint distribution of $(\boldsymbol{X},\, \boldsymbol{Y})$ on $[0,1]^2$ is a.c. (Equivalently, $\boldsymbol{P}$ is a.c. on $\mathcal{P}_2$.) 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 the random variable $\boldsymbol{Z} = 1 - \boldsymbol{X} - \boldsymbol{Y}$, by Example 14. Since $\boldsymbol{P}$ is

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

a.c., $\boldsymbol{Z}$ is a.c. as well, 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 12.

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 $f\colon U \to \mathbb{C}$ is *real-analytic* whenever it can, in the neighborhood of every point in its domain U (a connected 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 zero-locus of f is a nullset unless f vanishes identically on U. Although this is probably a well-known fact, no specific reference is known to the authors. Since this fact plays a crucial role in the proof of Theorem 17 below, a proof is included for the reader's convenience. With $\mathrm{Leb}_{\mathbb{R}^\ell}$ denoting the ℓ-dimensional volume on $\mathbb{R}^\ell$, it reads as follows.

LEMMA 19. *Let $f\colon U \to \mathbb{C}$ be real-analytic and $N_f := \{x \in U\colon f(x) = 0\}$. Then either $Leb_{\mathbb{R}^\ell}(N_f) = 0$ or $N_f = U$.*

Proof. Assume $N_f \neq U$. If $\ell = 1$, then N_f is at most countable [21, Theorem 10.18], and hence $\mathrm{Leb}_{\mathbb{R}^1}(N_f) = 0$. For $\ell \geq 2$, proceed by induction: Given any set $C \subset \mathbb{R}^\ell$ and $\gamma \in \mathbb{R}$, define $C_\gamma := \{(x_2, \ldots, x_\ell)\colon (\gamma, x_2, \ldots, x_\ell) \in C\} \subset \mathbb{R}^{\ell-1}$. By Fubini's theorem,

$$\begin{aligned}\mathrm{Leb}_{\mathbb{R}^\ell}(N_f) &= \int_{N_f} \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_\ell = \int_{\mathbb{R}} \left(\int_{(N_f)_{x_1}} \mathrm{d}x_2 \ldots \mathrm{d}x_\ell \right) \mathrm{d}x_1 \\ &= \int_{\mathbb{R}} \mathrm{Leb}_{\mathbb{R}^{\ell-1}}((N_f)_{x_1}) \mathrm{d}x_1.\end{aligned}$$

Notice that $(N_f)_{x_1} = N_{\tilde{f}}$, where $\tilde{f}\colon U_{x_1} \to \mathbb{C}$ is the real-analytic function with $\tilde{f}(x_2, \ldots, x_\ell) = f(x_1, x_2, \ldots, x_\ell)$. If $\mathrm{Leb}_{\mathbb{R}^{\ell-1}}(N_{\tilde{f}}) > 0$, then, by the induction assumption, $\tilde{f}$ must vanish identically on some connected component V of U_{x_1}. (Note that U_{x_1} may not be connected.) Fix any $(x_2, \ldots, x_\ell) \in V$. Since $N_f \neq U$, it can be assumed that $\gamma \mapsto f(\gamma, x_2, \ldots, x_\ell)$ does not vanish identically. It follows that $\mathrm{Leb}_{\mathbb{R}^{\ell-1}}((N_f)_{x_1}) > 0$ for at most countably many x_1, and hence $Leb_{\mathbb{R}^\ell}(N_f) = 0$.. $\square$

Remark 20. As the proof of Lemma 19 shows, $\mathrm{Leb}_{\mathbb{R}^\ell}$ can be replaced by any product of ℓ atomless measures on $\mathbb{R}$ (and $\mathrm{Leb}_{\mathbb{R}^\ell}$ then simply corresponds to the special case of each factor being $\mathrm{Leb}_{\mathbb{R}^1}$).

The proof of Theorem 17 will be based on several preliminary facts. First, given $a = (a_1, \ldots, a_d) \in \mathbb{C}^d$, let $p_a\colon \mathbb{C} \to \mathbb{C}$ denote the polynomial

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

By the fundamental theorem of algebra, p_a has exactly d zeroes (counted with multiplicities). If p_a has a multiple zero, then a universal polynomial relation must necessarily be satisfied by a. For instance, if $d = 2$ and p_a has a double zero, then $Q_2(a) = 0$, where $Q_2(a) = -a_1^2 + 4a_2$. The generalization to arbitrary d is classical (see, e.g., [8, Lemma 3.3.4]); for a proper formulation recall that the degree of a polynomial $\sum_j c_j x_1^{n_{1,j}} x_2^{n_{2,j}} \ldots x_d^{n_{d,j}}$ in d variables is defined as $\max\{n_{1,j} \times + \cdots + n_{d,j\times}\colon c_j \neq 0\}$.

PROPOSITION 21. *For every integer $d > 1$, there exists a nontrivial polynomial Q_d in d variables, with integer coefficients and of degree $2d - 2$, with the following property: Whenever p_a has a multiple zero, i.e., $p_a(z_0) = p_a{}'(z_0) = 0$ for some $z_0 \in \mathbb{C}$, then $Q_d(a) := Q_d(a_1, \ldots, a_d) = 0$.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

This fact will now be used to show that if a stochastic matrix P_0 is invertible and has distinct nonzero eigenvalues, then all stochastic matrices P sufficiently close to P_0 also are invertible and have distinct nonzero 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., $B_\varepsilon(P_0) = \{P \in \mathcal{P}_d\colon |P^{(i,j)} - P_0^{(i,j)}| < \varepsilon \text{ for all } 1 \le i, j \le d\}$. Note that with this topology, $\mathcal{P}_d$ is compact.

LEMMA 22. *Suppose $P_0 \in \mathcal{P}_d$ is invertible and has d distinct nonzero eigenvalues. Then there exist $\varepsilon > 0$ and $d-1$ nonconstant real-analytic functions $\lambda_2, \ldots, \lambda_d\colon B_\varepsilon(P_0) \to \mathbb{C}$ such that, for every $P \in B_\varepsilon(P_0)$,*

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

(ii) *$\lambda_i(P) \neq \overline{\lambda_j(P)}$ whenever $i \neq j$, unless $\lambda_i = \bar{\lambda}_j$ 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 nonzero eigenvalues. Thus the characteristic polynomial ψ_P of P has $d-1$ distinct nonzero 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 [17, 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 exist $\varepsilon \le \delta$ and a real-analytic function $g\colon B_\varepsilon(P_0) \to \mathbb{C}$ with $g(P_0) = z_0$ such that $\psi_P(g(P)) = 0$ for all $P \in B_\varepsilon(P_0)$. Overall, there exist $\varepsilon > 0$ and $d-1$ real-analytic functions $\lambda_i\colon B_\varepsilon(P_0) \to \mathbb{C}$ satisfying (i); note that $\lambda_1 \equiv 1$ by Proposition 5. 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 $2 \le i \le d$ and all $P \in B_\varepsilon(P_0)$. In this case, the real-analytic function $P \mapsto \psi_P(\lambda_i(P_0))$ 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 $\psi_{I_{d\times d}}(\lambda_i(P_0)) = (\lambda_i(P_0) - 1)^d \neq 0$. Consequently, none of the functions $\lambda_2, \ldots, \lambda_d\colon B_\varepsilon(P_0) \to \mathbb{C}$ 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 $\lambda_i(P_1) \in \mathbb{C} \setminus \mathbb{R}$, since if $\lambda_i(P_1)$ were real, then $\lambda_i(P_1) = \lambda_j(P_1)$, which is impossible because the eigenvalues are distinct. Since all matrices in $\mathcal{P}_d$ are real, their nonreal 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 λ_j coincide locally near P_1 and therefore, by real-analyticity, on all of $B_\varepsilon(P_0)$. ∎

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

LEMMA 23. *If the random Markov chain $\boldsymbol{P}$ is a.c., then, with probability one,*

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

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

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

Proof.

(i) Since $\boldsymbol{P}$ is a.c., with probability one $\boldsymbol{P}^{(i,j)} \in (0,1)$ 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 real-analytic on $\mathcal{P}_d$ and clearly not constant, as $\mathcal{P}_d$ contains both $I_{d\times d}$ and the matrix whose components all equal $\frac{1}{d}$. By Lemma 19, $\mathrm{Leb}_{\mathcal{P}_d}(\{P \in \mathcal{P}_d\colon \det P = 0\}) = 0$, and this in turn implies that $\mathbb{P}(\det \boldsymbol{P} = 0) = 0$.

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

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

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, $q(P) := Q_d(q_1(P), \ldots, q_d(P))$ defines a nonconstant real-analytic (in fact, polynomial) map $q\colon \mathcal{P}_d \to \mathbb{R}$, and since z_0 is a multiple eigenvalue of P if and only if $\psi_P(z_0) = \psi_P{}'(z_0) = 0$, Proposition 21 implies that

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

As before, $\mathbb{P}(q(\boldsymbol{P}) = 0) = 0$ by Lemma 19, 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 $\Phi_\rho\colon \mathbb{R}^2 \to \mathbb{R}$ by $\Phi_\rho(x) := (x_1^2 + x_2^2 - 10^{2\rho})^2$, and also $\Theta\colon \mathbb{R}^4 \to \mathbb{R}$ as $\Theta(x) := (x_1^2 + x_2^2 - x_3^2 - x_4^2)^2$. By (i) and (ii), $\boldsymbol{P}$ almost surely satisfies the hypotheses of Lemma 22, so let P_0, ε, and $\lambda_2, \ldots, \lambda_d$ be as in Lemma 22, and define real-analytic functions $\Phi_{\rho,i}$ and $\Theta_{i,j}$ on $B_\varepsilon(P_0)$ as

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

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

$$\begin{aligned}\Theta_{i,j}(P) &:= \Theta(\Re e\lambda_i(P), \Im m\lambda_i(P), \Re e\lambda_j(P), \Im m\lambda_j(P)) \\ &= (|\lambda_i(P)|^2 - |\lambda_j(P)|^2)^2.\end{aligned}$$

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

$$F_\rho(P) := \prod_{i=2}^d \Phi_{\rho,i}(P) \cdot \prod_{2\le i<j\colon \lambda_i \ne \bar{\lambda}_j} \Theta_{i,j}(P).$$

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

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

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

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

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

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

Since F_ρ is real-analytic and nonconstant, $\{P \in B_\varepsilon(P_0)\colon F_\rho(P) = 0\}$ is a $\mathrm{Leb}_{\mathcal{P}_d}$-nullset for every $\rho \in \mathbb{Q}$, and so is $\bigcup_{\rho \in \mathbb{Q}} \{P \in B_\varepsilon(P_0) : F_\rho(P) = 0\}$. Analogously to (i) and (ii), therefore, $\mathbb{P}(\boldsymbol{P} \text{ is resonant}) = 0$. ∎

Proof of Theorem 17. Let the random $d \times d$-matrix $\boldsymbol{P}$ be a.c. By Lemma 23, $\boldsymbol{P}$ is almost surely irreducible, aperiodic, and nonresonant. By Theorem 12, this implies that $\boldsymbol{P}$ is Benford with probability one. ∎

COROLLARY 24. *If the transition probabilities (i.e., the rows) of a random d-state Markov chain* $\boldsymbol{P}$ *are independent and a.c. on the standard d-simplex, then* $\boldsymbol{P}$ *is Benford with probability one.*

Remark 25.

(i) It is clear that without absolute continuity, Lemma 23 and Theorem 17 may fail. For example, for the conclusion of Lemma 23 to hold it is not enough to assume that the distribution of $\boldsymbol{P}$ on $\mathcal{P}_d$ is atomless; i.e., $\mathbb{P}(\boldsymbol{P} = P) = 0$ for every $P \in \mathcal{P}_d$. 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 23(i) and (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 $[0, 1]$. With this, the distribution of $\boldsymbol{P}$ on $\mathcal{P}_3$ is atomless yet concentrated on a $\mathrm{Leb}_{\mathcal{P}_3}$-nullset. The eigenvalues of $\boldsymbol{P}$ are

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

Note that $|\lambda_3| \le 0.05 < \lambda_2$. Clearly, $\boldsymbol{P}$ is resonant with probability one, and Lemma 23(iii) fails. Even more important perhaps, Theorem 17 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) A careful inspection of the above arguments shows that Lemma 23 and Theorem 17 hold whenever the distribution of $\boldsymbol{P}$ on $\mathcal{P}_d \subset \mathbb{R}^{d\times d} \equiv \mathbb{R}^{d^2}$ is such that $\mathbb{P}(f(\boldsymbol{P}) = 0) = 0$ for every real-analytic function $f\colon \mathbb{R}^{d^2} \to \mathbb{C}$ that does not vanish identically on $\mathcal{P}_d$. Evidently, this property of $\boldsymbol{P}$ holds automatically if $\boldsymbol{P}$ is a.c.

(iii) With hardly any effort, the tools employed in the proof of Lemmas 22 and 23 also yield a topological analogue of Theorem 17: 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. Some computational implications. For Markov chains with small state space, i.e., for small values of d, the limiting matrix P^* is easy to compute explicitly.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

In this case, an understanding of the behavior of significant digits in $(P^n - P^*)$ and $(P^{n+1} - P^n)$, however valuable in its own right, may be of limited practical use. On the other hand, for important applications that involve a very large state space, e.g., in computer science or theoretical biology [20], P^* may be very costly or practically impossible to determine explicitly. In this case, P^* typically is approximated using a variety of numerical algorithms. As detailed below, it is in situations like these that a proper understanding of the typical behavior of significant digits may be crucial also from a practical or computational point of view.

As a concrete example, a *Markov chain Monte Carlo* (MCMC) method will be considered. MCMC is a popular and powerful tool for generating samples from an arbitrary distribution [6, Chapter 7]. One of the most important advantages of MCMC is that it requires only specification of the target distribution up to a normalization constant, the determination of which often constitutes a challenging problem in itself. Historically, MCMC was motivated by computational problems in statistical physics that led to the idea of generating a Markov chain whose limiting distribution is equal to the target distribution. The most prominent MCMC algorithms are the Metropolis–Hastings and the Gibbs sampling algorithms, which both originated from the following Metropolis algorithm.

Assume that a random variable $\boldsymbol{X}$ is to be generated that takes values in $E = \{1, \ldots, m\}$, according to the target distribution $\{\pi_i\}$, where

$$\pi_i = \frac{b_i}{B}, \qquad i = 1, \ldots, m,$$

with all b_i positive, m large, and the normalization constant $B = \sum_{i=1}^m b_i$ difficult to calculate. The Metropolis algorithm constructs a Markov chain $(\boldsymbol{X}_n)_{n\in\mathbb{N}_0}$ on $\{1, \ldots, m\}$ whose evolution relies on an appropriately chosen stochastic matrix $Q = (q_{ij})$ in the following way:

(i) Given $\boldsymbol{X}_n = i$, generate a random variable $\boldsymbol{Y}$ which satisfies $\mathbb{P}(\boldsymbol{Y} = j) = q_{ij}$ for all $j = 1, \ldots, m$ and is independent of $\boldsymbol{X}_0, \ldots, \boldsymbol{X}_{n-1}$.

(ii) Given $\boldsymbol{Y} = j$, let $\alpha_{ij} := \min\left\{\frac{b_j q_{ji}}{b_i q_{ij}}, 1\right\}$ and choose

$$\boldsymbol{X}_{n+1} = \begin{cases} j & \text{with probability } \alpha_{ij}, \\ i & \text{with probability } 1 - \alpha_{ij}. \end{cases}$$

The Markov chain $(\boldsymbol{X}_n)$ thus defined has the transition probability matrix

$$P^{(i,j)} = \begin{cases} q_{ij}\alpha_{ij} & \text{if } j \neq i, \\ 1 - \sum_{k\neq i} q_{ik}\alpha_{ik} & \text{if } j = i. \end{cases}$$

To illustrate this through a simple specific example, consider the case of the matrix Q having identical rows, i.e., $q_{ij} = p_j$ for all $i, j \in E$, where $p = (p_i)_{i\in E}$ is a strictly positive probability distribution on E. With $w_i := \pi_i / p_i$, the off-diagonal elements of P are

$$P^{(i,j)} = p_j \min\left\{1, \frac{w_j}{w_i}\right\} \quad \forall i, j \in E\colon i \neq j.$$

Assume that the states of E are labeled in such a way that $w_1 \geq w_2 \geq \cdots \geq w_m$. In this case, the eigenvalues of P are easily seen to be $\lambda_1 = 1$ and

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$$\lambda_j = \sum_{\ell=j}^{m} \pi_\ell \left(\frac{1}{w_\ell} - \frac{1}{w_{j-1}} \right) \geq 0, \qquad j = 2, \ldots, m.$$

Consequently, if $\log \lambda_j$ is irrational for all $j = 2, \ldots, m$, then the Markov chain $(\boldsymbol{X}_n)$ is Benford, by Theorem 12, and so is the Metropolis algorithm.

Why is it important to know whether an algorithm often, or even typically, generates Benford distributed data? A most compelling reason has been put forth by Knuth in his classic text *The Art of Computer Programming* [16, 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 of the components of $(P^n - P^*)$ and $(P^{n+1} - P^n)$. Theorem 17 above shows that these components 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 [11, 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 17, this suggests that special attention should be given to overflow and underflow errors in any numerical 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 all components in $(P^{n+1} - P^n)$ are less than 10^{-10}," will result in some error, and Theorem 17 shows that this difference is generally Benford. In fact, justified by heuristics and by the extensive empirical evidence of BL in other numerical calculations, the analysis of roundoff errors has often been carried out under the *hypothesis* of a logarithmic statistical distribution (cf. [11, 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 is skewed toward smaller leading significant digits, as is the case for BL. To obtain a rough idea of the magnitude of this underestimate when the true statistical distribution is BL, let $\boldsymbol{X}$ 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]).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

In view of the relevance of BL for large-scale computations involving Markov chains, it is important to also note that the speed of convergence to BL can vary considerably. This is apparent already from Table 1: The digit distributions of the sequences (2^n) and (F_n) converge to BL faster than $(n!)$. A possible explanation for this is suggested in [18, Chapter 2, Theorem 3.4], where explicit bounds for rates of convergence to uniformity are found for $(\alpha n \bmod 1$. These bounds depend on the continued fraction expansion of the irrational number α—the smaller the coefficients in this expansion, the faster the convergence. Both $\log 2$ and $\log \frac{1+\sqrt{5}}{2}$ appear to have very few large coefficients in their continued fraction expansion. Via Proposition 2, this translates into relatively fast convergence to BL.

Similarly, the speed of convergence to BL for Markov chains may vary considerably, and the result for $(\alpha n \bmod 1)$ mentioned in the previous paragraph suggests that this speed is determined by the continued fraction expansion of the logarithm of the moduli of the eigenvalues of P as well as of the elements of the sets $\frac{1}{2\pi} \arg \Lambda$. The next example illustrates this. The reader should keep in mind that relatively little is known at present about the precise speed of convergence to BL (or uniformity) in higher-dimensional systems; see, e.g., [10].

Example 26 (different speeds of convergence to BL for Markov chains).

(i) Let

$$P = \begin{bmatrix} 0.25 & 0.35 & 0.40 \\ 0.30 & 0.45 & 0.25 \\ 0.65 & 0.15 & 0.20 \end{bmatrix}.$$

The eigenvalues of P are $\lambda_1 = 1$ and $\lambda_{2,3} = -\frac{1}{20} \mp \frac{1}{20}\sqrt{21}$. Since $\log|\lambda_2|$ and $\log|\lambda_3|$ are irrational and different, P is nonresonant, and Theorem 12 implies that the Markov chain defined by P is Benford. Since $|\lambda_2| > |\lambda_3|$, for the speed of convergence to BL it is important how well $\log|\lambda_2|$ is approximated by rational numbers. The first 50 coefficients of the continued fraction expansion of $\log|\lambda_2|$,

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

do not exceed 86 and are mostly small numbers not showing rapid growth at all. A comparatively rapid initial approach to BL is therefore expected. This is confirmed experimentally by Figure 1, which shows, as a function of n, the L_1-distance between the empirical frequencies for the significant digits of $(P^{n+1} - P^n)^{(2,1)}$ and the Benford probabilities; the behavior of $(P^n - P^*)^{(2,1)}$ is very similar, as is in fact the behavior of all other components.

(ii) Let

$$P = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.4 \\ 0.4 & 0.0 & 0.6 \end{bmatrix},$$

with eigenvalues $\lambda_1 = 1$ and $\lambda_{2,3} = \frac{7}{20} \pm \frac{1}{20}\imath\sqrt{3}$. The behavior of the significant digits is governed by the two irrational numbers $\log|\lambda_2|$ and $\frac{1}{2\pi} \arg \lambda_2$. For instance,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

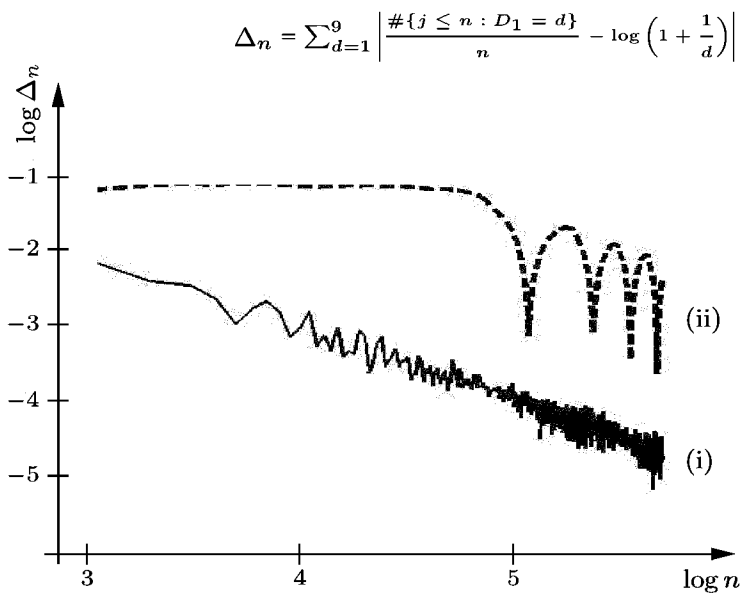

FIG. 1. *Plotting the L_1-distance Δ_n between the empirical frequencies and the Benford probabilities for the first significant digits of $(P^{n+1} - P^n)^{(2,1)}$, with the transition probability matrices P from Example* 26(i) *and* (ii), *respectively.*

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

When compared with (i), the repeated early large coefficients in the continued fraction expansion of $\frac{1}{2\pi} \arg \lambda_2$ suggest a somewhat slower initial approach to BL. Again, this is confirmed experimentally by Figure 1.

Finally, observe that Theorems 12 and 17 should make it possible to adapt the current plethora of BL-based goodness-of-fit statistical tests, e.g., for detecting fraud [7], to the problem of detecting whether or not a sequence of realizations of a finite-state stochastic process originates from a Markov chain, i.e., whether or not the process is Markov. By Theorem 17, conformance with BL for the differences $(P^{n+1} - P^n)$ is typical in finite-state Markov chains, so a standard (e.g., chi-square) 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.

Acknowledgments. A. Berger is indebted to J. Kuttler for a helpful discussion about resultants. T. P. Hill wishes to express his gratitude to the Vrije Universiteit Amsterdam and to the Econometrics Department, especially Professor H. Tijms, for their hospitality during several visits when part of this research was conducted. All authors are grateful to the editor and referees whose thoughtful suggestions helped improve the exposition.

REFERENCES

[1] A. BERGER, L. A. BUNIMOVICH, AND T. P. HILL, *One-dimensional dynamical systems and Benford's law*, Trans. Amer. Math. Soc., 357 (2005), pp. 197–219.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[2] A. Berger, *Multi-dimensional dynamical systems and Benford's law*, Discrete Contin. Dyn. Syst. Ser. A, 13 (2005), pp. 219–237.
[3] A. Berger, *Some dynamical properties of Benford sequences*, J. Difference Equ. Appl.,17 (2011), pp 137–159.
[4] A. Berger and T. P. Hill, *Newton's method obeys Benford's law*, Amer. Math. Monthly, 114 (2007), pp. 588–601.
[5] A. Berger and S. Siegmund, *On the distribution of mantissae in nonautonomous difference equations*, J. Difference Equ. Appl., 13 (2008), pp. 829–845.
[6] P. Bremaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer, New York, 1999.
[7] C. Carslaw, *Anomalies in income numbers. Evidence of goal oriented behavior*, The Accounting review, 63 (1988), pp. 321–327.
[8] H. Cohen, *A Course in Computational Algebraic Number Theory*, Springer, New York, 1996.
[9] P. Diaconis, *The distribution of leading digits and uniform distribution mod* 1, Ann. Probab., 5 (1979), pp. 72–81.
[10] M. Drmota and R. F. Tichy, *Sequences, Discrepancy and Applications*, Springer, Berlin, 1997.
[11] A. Feldstein and P. Turner, *Overflow, underflow, and severe loss of significance in floating-point addition and subtraction*, IMA J. Numer. Anal., 6 (1986), pp. 241–251.
[12] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol.* 1, 3rd ed., Wiley, New York, 1950.
[13] T. P. Hill, *Base invariance implies Benford's law*, Proc. Amer. Math. Soc., 123 (1995), pp. 887–895.
[14] T. P. Hill, *A statistical derivation of the significant-digit law*, Statist. Sci., 10 (1995), pp. 354–363.
[15] S. Kanemitsu, K. Nagasaka, G. Rauzy, and J. Shiue, *On Benford's Law: The First Digit Problem*, Lecture Notes in Mathematics 1299, eds., S. Watanabe and Y. V. Prokhorov, Springer, New York, 1988, pp. 158–169.
[16] D. Knuth, *The Art of Computer Programming, Vol.* 2, 3rd ed., Addison-Wesley, Reading, MA, 1997.
[17] S. Krantz and H. Parks, *A Primer of Real Analytic Functions*, Birkhäuser, Basel, 2002.
[18] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Wiley, New York, 1974.
[19] K. Nagasaka and J. Shiue, *Benford's law for linear recurrence sequences*, Tsukuba J. Math., 11 (1987), pp. 341–351.
[20] J. R. Norris, *Markov Chains*, Cambridge University Press, Cambridge, UK, 1997.
[21] W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill, New York, 1987.
[22] P. Schatte, *On the uniform distribution of certain sequences and Benford's law*, Math. Nachr., 136 (1988), pp. 271–273.
[23] E. Seneta, *Non-negative Matrices and Markov Chains*, 2nd ed., Springer, New York, 1981.
[24] D. Serre, *Matrices. Theory and Applications*, Springer, New York, 2002.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.