Stochastic aspects of one-dimensional discrete dynamical systems: Benford's law

Mark A. Snyder, James H. Curry,* and Anne M. Dougherty
Department of Applied Mathematics, University of Colorado, UCB 526, Boulder, Colorado 80309-0526
(Received 20 February 2001; published 26 July 2001)

Benford's law owes its discovery to the "Grubby Pages Hypothesis," a 19th century observation made by Simon Newcomb that the beginning pages of logarithm books were grubbier than the last few pages, implying that scientists referenced the values toward the front of the books more frequently. If a data set satisfies Benford's law, then it's significant digits will have a logarithmic distribution, which favors smaller significant digits. In this article we demonstrate two ways of creating discrete one-dimensional dynamical systems that satisfy Benford's law. We also develop a numerical simulation methodology that we use to study dynamical systems when analytical results are not readily available.

DOI: 10.1103/PhysRevE.64.026222 PACS number(s): 05.45.−a, 05.10.Gg, 05.10.Ln

I. INTRODUCTION

The "Grubby Pages Hypothesis" [1] led to the conjecture that the significant digits of many data sets describing the physical world are not uniformly distributed, but distributed in a way that favors smaller significant digits and smaller combinations of significant digits (e.g., the first significant digit will be 4 more frequently than 5 and the first two significant digits will be 89 more often than they will be 91). Benford's law makes this conjecture more precise by providing a probability mass function for combinations of significant digits. While the law may seem obscure, many authors have connected Benford's law to data sets like stock prices [2], tax data [3], and census statistics [4]. Since some accounting data seems to satisfy Benford's law, it is now being used to detect accounting fraud [3]. Recently an empirical study suggested that some dynamical systems satisfy Benford's law [5]. This is important, because if dynamical systems are used to model physical systems, and the physical data satisfy Benford's law, then the dynamical system models should too.

This article focuses on the long run behavior and digital frequencies of discrete one-dimensional dynamical systems (see Ref. [5] for the continuous case) and how well such systems satisfy Benford's law. We use notation and tools similar to those developed in Refs. [6] and [7]. We start with a space S and a transformation τ where τ describes the dynamics of the system and $\tau: \quad S\rightarrow S$. Given an initial vector of points, $\mathbf{x_0}$, distributed in S according to some probability density function f, the vector of points given by $\mathbf{x_1}=\tau(\mathbf{x_0})$ has a probability density function that depends on both τ and f. Successive iterates are given by $\mathbf{x_{n+1}}=\tau(\mathbf{x_n})$. We develop methods to determine if a dynamical system satisfies Benford's law, we introduce a method to modify dynamical systems such as the logistic map so that they satisfy Benford's law, and we show how to construct new dynamical systems that satisfy Benford's law.

II. BENFORD'S LAW

In order to define Benford's law, we first define the functions d_i, which extract the first i significant digits from a random variable. Given a random variable X, $d_i(X)$ is a discrete random variable. We denote the probability that X equals k as $P[X=k]$. A random variable X satisfies Benford's law if for $i=1,2,3,\ldots,$

$$P[d_i(X)=k]=\log_{10}(1+1/k)$$
$$\text{for} \quad k=10^{i-1},10^{i-1}+1,\ldots,10^i-1. \tag{1}$$

Using a random variable to introduce Benford's law is appropriate here because random variables and dynamical systems can both be easily associated with densities. This definition of Benford's law is equivalent to other definitions in the literature (see Ref. [8]).

Definition 1. If the entries in a random vector $\mathbf{x}$ *are distributed according to the density p, then we denote the probability density function for the random vector* $\{\log_{10}(|\mathbf{x}|)\}$ *by* Δp, *where* $\{\log_{10}(|\mathbf{x}|)\}$ *is the fractional part of* $\log_{10}(|\mathbf{x}|)$.

Theorem 1. A random vector $\mathbf{x}$, *distributed according to some density p, satisfies Benford's law if and only if* $\Delta p=1$.

Proof. See [9], Theorem 1. ■

III. INVARIANT DENSITY FUNCTIONS

We represent dynamical systems by (S, p, τ), where S is the interval the dynamical system is defined on, p is the invariant density, and τ is the function that defines the dynamics of the system. Implicit in this representation is that p is the only invariant probability density associated with τ. Every dynamical system mentioned in this article has exactly one associated invariant density. In general, if τ is smooth, locally invertible, aperiodic, piecewise expanding, and satisfies the Markov property, then it has one invariant measure, so consequently it has a single invariant density (Ref. [6], Theorem 6.1.1). If one of the proceeding conditions is not satisfied, one can still show the existence of an invariant density function for τ by using results from Ref. [6], Chap. 5. Also, using Ref. [6], Theorem 8.2.1 an upper bound can be placed on the number of distinct invariant density functions for τ.

The Frobenius-Perron operator is defined as

$$P_\tau f=\frac{d}{dx}\int_{\tau^{-1}(-\infty,x]} f(\lambda)d\lambda. \tag{2}$$

*Author to whom correspondence should be addressed.

The probability density functions generated by successive applications of τ are denoted by $f,\ P_\tau f,\ P_\tau^2 f,\ P_\tau^3 f,\ \dots, P_\tau^n f, \dots$ where $P_\tau^n f$ denotes the nth application of the Frobenius–Perron operator. This notation reinforces the fact that each new probability density function depends on both τ and f.

The dynamical systems mentioned in this paper all have one invariant density, so $P_\tau^n f$ will converge to a single unique invariant probability density function, denoted by p, which is independent of the initial density f.

IV. BENFORD DYNAMICAL SYSTEMS

Definition 2. A dynamical system (S,τ,p) satisfies Benford's law if $\Delta p = 1$.

We propose Definition 2 because the iterates of (S,τ,p), given by $\mathbf{x_{n+1}} = \tau(\mathbf{x_n})$, will converge to the distribution specified by p, which satisfies Benford's law if and only if $\Delta p = 1$.

The reciprocal density is

$$p_R(x) = \begin{cases} \dfrac{1}{x \ln(b/a)} & \text{if} \quad 0 < a \leqslant x \leqslant b < \infty, \\ 0 \quad \text{else.} \end{cases} \tag{3}$$

If $b/a = 10^j$, where j is a natural number, we say the Benford ratio is satisfied.

Theorem 2. A dynamical system satisfies Benford's law if its invariant ergodic probability density function is the reciprocal density and the Benford ratio is satisfied.

Proof. If the elements in the random vector $\mathbf{x}$ are distributed according to the density p_R, then the elements in $\mathbf{y} = \log_{10}(\mathbf{x})$ are distributed uniformly on $[\log_{10} a,\ \log_{10} b]$. The elements of $\mathbf{y}$ are distributed uniformly on $[\log_{10} a,\ j + \log_{10} a]$ because of the condition $b/a = 10^j$. Since the elements of $\mathbf{y}$ are distributed uniformly on an interval where the upper and lower limits differ by an integer, the fractional part of the elements of $\mathbf{y}$ must be distributed uniformly on the interval $[0,1]$, so $\Delta p_R = 1$. ■

See Ref. [10] for more densities that satisfy Benford's law.

V. THE TENT AND LOGISTIC MAPS

The transformation that produces the tent map is

$$\tau_T(x) = \begin{cases} 2x & \text{if} \quad 0 \leqslant x \leqslant 1/2, \\ 2 - 2x & \text{if} \quad 1/2 < x \leqslant 1. \end{cases} \tag{4}$$

The tent map is the dynamical system defined by $([0,1], \tau_T, p_T = 1)$. For $0 \leqslant x \leqslant 1$ we calculate

$$\Delta p_T = \sum_{k=1}^{\infty} \ln(10) 10^{x-k} = \frac{\ln(10) 10^x}{9} \neq 1, \tag{5}$$

which implies that the tent map does not satisfy Benford's law. The logistic map is the dynamical system defined by $([0,1],\quad \tau_L = 4x(1-x),\quad p_L = 1/(\pi\sqrt{(1-x)x}))$. For $0 \leqslant x \leqslant 1$ we calculate

$$\Delta p_L = \sum_{k=1}^{\infty} \frac{\ln(10)\sqrt{10^{x-k}}}{\pi\sqrt{1-10^{x-k}}} \neq 1, \tag{6}$$

so the logistic map does not satisfy Benford's law either.

VI. MODIFYING THE LOGISTIC MAP

Here we use homeomorphisms to alter the iterates of the logistic map so that they satisfy Benford's law.

Definition 3. Let (S,τ,p) and (S^*,τ^*,p^*) be dynamical systems. The two dynamical systems are topologically conjugate if there exists a homeomorphisms h such that $S^* = h(S)$, $\tau = h^{-1} \circ \tau^* \circ h$, and

$$p^*(x) = \frac{d}{dx}\int_{-\infty}^{h^{-1}(x)} p(\lambda)\, d\lambda. \tag{7}$$

Topologically conjugation preserves both measure theoretic and topological properties of τ. See [6], Sec. 3.6 for a concise treatment of the above ideas.

A well-known result (see Ref. [11]) is that the logistic and tent maps are related according to $\tau_T = g \circ \tau_L \circ g^{-1}$, where

$$g(x) = \int_0^x p_L(\lambda)\, d\lambda = 2\sin^{-1}(\sqrt{x})/\pi \quad \text{for} \quad 0 \leqslant x \leqslant 1 \tag{8}$$

is the homeomorphism connecting the two maps. We now use the homeomorphism

$$h(x) = \left(\int_a^x p_R(\lambda)\, d\lambda\right)^{-1} = a\left(\frac{b}{a}\right)^x \quad \text{for} \quad 0 \leqslant x \leqslant 1 \tag{9}$$

to modify the tent map so that it satisfies Benford's law. The transformation

$$\tau_B(x) = h \circ \tau_T \circ h^{-1} = \begin{cases} a\left(\dfrac{b}{a}\right)^{2[\ln(x/a)/\ln(b/a)]} & \text{if} \quad a \leqslant x \leqslant a\left(\dfrac{b}{a}\right)^{1/2}, \\ a\left(\dfrac{b}{a}\right)^{2-2[\ln(x/a)/\ln(b/a)]} & \text{if} \quad a\left(\dfrac{b}{a}\right)^{1/2} \leqslant x \leqslant b \end{cases} \tag{10}$$

(see Fig. 1) defines a dynamical system on $[a,b]$ where $0 < a < b < \infty$. The invariant density of this dynamical system is

$$\frac{d}{dx}\int_0^{h^{-1}(x)} p_T(\lambda)\, d\lambda = \frac{1}{x \ln(b/a)} \quad \text{if} \quad a \leqslant x \leqslant b, \tag{11}$$

which is just the reciprocal density. We know from Theorem 2 that the dynamical system defined by $([a,b], \tau_B, p_R)$ satisfies Benford's law when a and b satisfy the Benford ratio. The following conjugacy diagram illustrates the process of converting the logistic map into a system that satisfies Benford's law (see Ref. [10] for more details and examples).

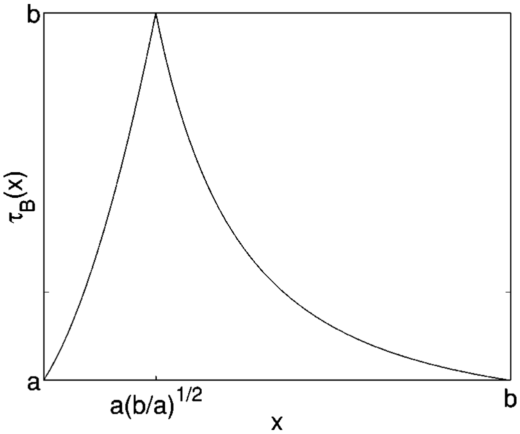

FIG. 1. Plot of τ_B.

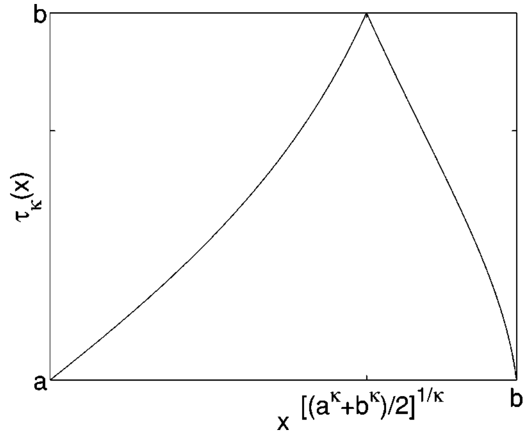

FIG. 2. Plot of τ_k for $\kappa=2$.

Logistic map $[0,1]\overset{\tau_L}{\rightarrow}[0,1]$

$g\downarrow \quad \downarrow g$

Tent map $[0,1]\overset{\tau_T}{\rightarrow}[0,1]$

$h\downarrow \quad \downarrow h$

Benford system $[a,b]\overset{\tau_B}{\rightarrow}[a,b]$

Since g and h are homeomorphisms there is a one-to-one correspondence between the dynamics of the logistic map and the map given in (10). The method we used to transform the logistic map is easily generalized. In Ref. [10] we have applied it to systems from Refs. [6,12,13].

VII. CONSTRUCTING DYNAMICAL SYSTEMS

Methods from Ref. [6], Chap. 12 allow us to choose an invariant density p_I and then construct dynamical systems defined by ($I=[-1,1]$, τ_I, p_I). Here we derive a dynamical system ($[a,b]$, τ_κ, p_R), where p_R is the reciprocal density, that is topologically conjugate to (I, τ_I, p_I). The homeomorphism we will use to establish the conjugacy between the two systems is $\phi(x)=\alpha x^\kappa+\beta$, where $0<\kappa<\infty$ and α and β are chosen so that $\phi([a,b])=I$. The system on $[a,b]$ will have invariant density p_R if we set

$$p_I=\frac{d}{dx}\int_a^{\phi^{-1}(x)} p_R(\lambda)\,d\lambda \quad \text{if} \quad -1\leqslant x\leqslant 1. \tag{12}$$

After finding τ_I, we set $\tau_\kappa=\phi^{-1}\circ\tau_I\circ\phi$ (see Fig. 2). With some manipulation (see Ref. [10] for full details) we find

$$\tau_\kappa(x)=\begin{cases} bx(a^\kappa+b^\kappa-x^\kappa)^{-1/\kappa} & \text{if} \quad 0<a\leqslant x\leqslant([a^\kappa+b^\kappa]/2)^{1/\kappa},\\ b(a^\kappa+b^\kappa-x^\kappa)^{1/\kappa}/x & \text{if} \quad ([a^\kappa+b^\kappa]/2)^{1/\kappa}\leqslant x\leqslant b<\infty. \end{cases} \tag{13}$$

It is easily shown that $P_{\tau_\kappa}(p_R)=p_R$. When the Benford ratio is satisfied, the dynamical system ($[a,b],\tau_\kappa,p_R$) satisfies Benford's law.

VIII. NUMERICAL SIMULATIONS

The invariant densities of most dynamical systems are unknown. When a dynamical system's invariant density is unknown, we cannot use Definition 2 to prove that it satisfies (or nearly satisfies) Benford's law. Another approach to measure how close a dynamical system is to satisfying Benford's law is simulation.

In what follows we use a two step process to simulate dynamical systems. First we generate $\mathbf{x_0}$, a vector of initial conditions by using pseudorandom numbers from a density defined on S, where each pseudorandom number corresponds to the initial condition of one orbit. Then we iterate $\mathbf{x_{n+1}}=\tau(\mathbf{x_n})$. The vector $\mathbf{x_n}$ is completely determined by τ and $\mathbf{x_0}$, but for large n the density of $\mathbf{x_n}$ is invariant and independent of the density of $\mathbf{x_0}$. In the test simulations we use 10^4, 10^5, and 10^6 orbits and examine the orbits after 100 applications of τ. Empirical evidence shows that by iteration 100, the density of $\mathbf{x_n}$ is nearly independent of the density of $\mathbf{x_0}$.

We test the null hypothesis, $\mathcal{H}_0$, that the first significant digits of the orbits of a dynamical system satisfy the predictions of Benford's law. From each simulation, under the assumptions of the null hypothesis, a test statistic with a chi-square distribution and 8 degrees of freedom may be constructed by comparing the simulation frequencies of the first significant digits with the frequencies predicted by Benford's law (see Ref. [10]). We run m simulations (where m

≥ 500) to obtain m independent identically distributed observations. From these observations we use a Kolmogorov–Smirnov (KS) goodness of fit test to determine whether we accept or reject $\mathcal{H}_0$ (see Ref. [10]). A goodness of fit test is a hypothesis test that allows us to compare our empirical distribution function with the chi-square distribution. The KS statistic, D_m, is the largest vertical distance between the two distribution functions. Obviously a large value of D_m indicates a difference between the empirical and chi-square distributions. The form of our hypothesis test is to reject $\mathcal{H}_0$ at a significance level α if $D_m > d_{m,1-\alpha}$, where $d_{m,1-\alpha}$ is a constant depending on the desired significance level and the number of observations. A good approximation to this test (see Ref. [14], Chap. 6 or Ref. [10] for a table of p values) is to reject $\mathcal{H}_0$ if the test statistic C_m is greater than a constant, $c_{1-\alpha}$, where $C_m = (\sqrt{m} + 0.12 + 0.11/\sqrt{m})D_m$. This alternative test requires only one table of critical values, where the original KS test requires a different table of critical values for each value of m.

A. Interpretation of the simulation results

A type I error is made if $\mathcal{H}_0$ is rejected when $\mathcal{H}_0$ is true. The probability of a type I error is α, which is the level of the test. A type II error is made if $\mathcal{H}_0$ is accepted when the alternative hypothesis is true. The probability of a type II error increases when α decreases. The p value is the smallest level of significance for which the observed data indicate the null hypothesis should be rejected. The smaller the p value is, the more compelling the evidence is that the null hypothesis should be rejected.

If $\mathcal{H}_0$ is rejected, this implies that with high probability the simulated dynamical system does not satisfy Benford's law. Simulation can show that a dynamical system does not satisfy Benford's law since showing that (1) does not hold for even one value of i, shows that Benford's law is not satisfied.

If $\mathcal{H}_0$ is accepted, this only implies that the first significant digits of the simulated dynamical system closely match the frequencies predicted by Benford's law. Benford's law applies to all significant digits, so simulation is not a practical tool for showing that a dynamical system satisfies Benford's law in the sense of Definition 2. However, simulation can show that the first few significant digits of a dynamical system are very close to the predictions of Benford's law.

B. Simulation of the reciprocal system

We know that the reciprocal system, the dynamical system defined by $([a,b],\tau_\kappa,p_R)$, satisfies Benford's law when a and b satisfy the Benford ratio. Our simulation results for $([a,b],\tau_\kappa,p_R)$, where $\kappa=1$, are summarized in Table I (the "Benford" column indicates if the dynamical system being simulated satisfies Benford's law). Our test statistic is C_m; larger values of C_m correspond to an increased probability that the dynamical system under test does not satisfy Benford's law (see Ref. [10] for full details).

We know that $([1,10],\tau_\kappa,p_R)$ satisfies Benford's law, so it's no surprise that our two simulations of this system strongly favor the acceptance of $\mathcal{H}_0$. These simulations give us a good idea of how small we can expect the test statistic, C_m, to be when Benford's law is satisfied.

TABLE I. Reciprocal system $([1,b],\tau_\kappa,p_R)$.

b	Orbits	m	C_m	p value	Benford	$\mathcal{H}_0$
10	10^5	1000	0.8879	$0.15 < p$	yes	accept
10	10^6	500	1.0304	$0.15 < p$	yes	accept
10.1	10^6	500	2.2778	$p \leq 0.01$	no	reject
10.01	10^6	500	1.3980	$0.025 < p < 0.05$	no	reject

In the simulations of $([1,10.1],\tau_\kappa,p_R)$ and $([1,10.01],\tau_\kappa,p_R)$ $\mathcal{H}_0$ was rejected. These results indicate that our simulations method is extremely effective at detecting small deviations from Benford's law.

C. Simulation of Renyi's example

Renyi's example is a dynamical system given by $([0,\infty),\tau_H,p_H)$ where

$$\tau_H(x) = \begin{cases} \dfrac{x}{1-x} & \text{if} \quad 0<x<1, \\ x-1 & \text{if} \quad \text{if} \;\; 1 \leq x. \end{cases} \tag{14}$$

The value of p_H is unknown, so we can learn something new about the digital frequencies of Renyi's example with a simulation study. Our simulation results (Table II) give very strong evidence that the first significant digits of the orbits generated by Renyi's example very nearly or exactly match the predictions of Benford's law.

Notice that for the Renyi's example simulation with 10^6 orbits, the C_m statistic is less than the statistic generated by simulating dynamical systems that satisfy Benford's law (see results for $([a,b],\tau_\kappa,p_R)$).

We have not shown that Renyi's example satisfies Benford's law. What we have shown is that that the first significant digits of the orbits generated by Renyi's example are very close to the predictions of Benford's law; the simulations of Renyi's example generated data that was closer to the predictions of Benford's law than dynamical systems that we know satisfy Benford's law.

D. Dynamics on [0,1]

In Sec. VII we constructed the dynamical system $([a,b], \tau_\kappa, p_R)$ where $0<a<b<\infty$ and $0<\kappa<\infty$. If we consider this dynamical system on $[0,1]$ an obvious problem is that p_R is not defined when $a=0$. Because of this, we write $([0,1], \tau_\kappa, p_U)$ where p_U is an unknown density.

TABLE II. Renyi's example.

Orbits	m	C_m	p value	$\mathcal{H}_0$
10^4	1000	0.9881	$0.15 < p$	accept
10^5	1000	1.0890	$0.15 < p$	accept
10^6	500	0.62303	$0.15 < p$	accept

Since p_U is unknown we use a first digit numerical simulation using 10^6 orbits to test ($[0,1]$, τ_κ, p_U) satisfies Benford's law. Our simulation produced a p value very near 0, so the probability that ($[0,1]$, τ_κ, p_U) does not satisfy Benford's law is very close to 1.

IX. CONCLUSION

This article provides a positive answer to the question, ''Do dynamical systems satisfy Benford's law?'' Because of the many physical data sets that satisfy Benford's law we believe that the methods we have developed for determining how closely a dynamical system satisfies Benford's law may have important applications in validating mathematical models. If the observed data generated by a physical process satisfy Benford's law, then results generated by a model of that physical process should also satisfy Benford's law. The analytical approach from Sec. V may be used to determine with certainty if a model has the correct significant digit distribution, or when analytical results are impossible to obtain, the method from Sec. IX, may be used to build numerical evidence that a mathematical model produces the appropriate distribution of significant digits.

A survey of the Benford's law literature (see Refs. [1–4,8,9]) reveals a surprising prevalence of physical data sets and common sequences (e.g., the Fibonacci sequence or $1!,2!,3!,\ldots$) that satisfy Benford's law. Because of the ubiquity of Benford data, we believe that the development of the mathematics associated with Benford's law is important.

ACKNOWLEDGMENT

M.A.S. was supported by the NSF through Grant No. DMS-9810751.

[1] S. Newcomb, Am. J. Math. **4**, 39 (1881).
[2] T. P. Hill, *The First Digit Phenomenon* (American Scientist), Vol. 86.
[3] M. Nigrini, J. Am. Tax. Assoc. **18**, 72 (1996).
[4] R. A. Raimi, Am. Math. Monthly **83**, 521 (1976).
[5] C. L. Tolle, J. L. Budzien, and R. A. LaViolette, Chaos **10**, 331 (2000).
[6] A. Boyarsky and P. Gora, *Laws of Chaos: Invariant Measures and Dynamical Systems in One Dimension* (Birkhäuser, Boston, 1997).
[7] T. Y. Li and J. A. Yorke, Trans. Am. Math. Soc. **235**, 183 (1978).
[8] T. P. Hill, Stat. Sci. **10**, 354 (1995).
[9] P. Diaconis, Ann. Prob. **2**, 72 (1977).
[10] M. A. Snyder, J. H. Curry, and A. M. Dougherty (unpublished).
[11] D. Ruelle, *Chaotic Evolution and Strange Attractors* (Cambridge University Press, Cambridge, 1989).
[12] L. Billings and E. M. Bolt, Chaos **12**, 365 (2001).
[13] L. Billings, J. H. Curry, and V. Robins, Am. Math. Soc., Contemporary Math. **252**, 63 (1999).
[14] A. M. Law and D. W. Kelton, *Simulation Modeling and Analysis* (McGraw-Hill Higher Education, New York, 2000).