BENFORD'S LAW, VALUES OF L-FUNCTIONS AND THE $3x + 1$ PROBLEM

ALEX V. KONTOROVICH AND STEVEN J. MILLER

ABSTRACT. We show the leading digits of a variety of systems satisfying certain conditions follow Benford's Law. For each system proving this involves two main ingredients. One is a structure theorem of the limiting distribution, specific to the system. The other is a general technique of applying Poisson Summation to the limiting distribution. We show the distribution of values of L-functions near the central line and (in some sense) the iterates of the $3x + 1$ Problem are Benford.

1. INTRODUCTION

While looking through tables of logarithms in the late 1800s, Newcomb [New] noticed a surprising fact: certain pages were significantly more worn than others. People were referencing numbers whose logarithm started with 1 more frequently than other digits. In 1938 Benford [Ben] observed the same digit bias in a wide variety of phenomena.

Instead of observing one-ninth (about 11%) of entries having a leading digit of 1, as one would expect if the digits $1, 2, \dots, 9$ were equally likely, over 30% of the entries had leading digit 1 and about 70% had leading digit less than 5. Since $\log_{10} 2 \approx 0.301$ and $\log_{10} 5 \approx 0.699$, one may speculate that the probability of observing a digit less than k is $\log_{10} k$, meaning that the probability of seeing a particular digit j is $\log_{10}(j+1) - \log_{10} j = \log_{10}\left(1 + \frac{1}{j}\right)$. This logarithmic phenomenon became known as Benford's Law after his paper containing extensive empirical evidence of this distribution in diverse data sets gained popularity. See [Hi1] for a description and history, [Hi2, BBH, Dia] for some results, and page 255 of [Knu] for connections between Benford's law and rounding errors in computer calculations.

In [BBH] it was proved that many dynamical systems are Benford, including most power, exponential and rational functions, linearly-dominated

Date: August 22, 2005.

2000 *Mathematics Subject Classification.* 11K06, 60A10, 11B83, 11M06 (primary), 60F05, 11J86, 60J65, 46F12 (secondary).

Key words and phrases. Benford's Law, Poisson Summation, L-Function, $3x + 1$, Equidistribution, Irrationality Type.

systems, and non-autonomous dynamical systems. This adds to the ever-growing family of systems known or believed to satisfy Benford's Law, such as physical constants, stock market indices, tax returns, sums and products of random variables, the factorial function and Fibonacci numbers, just to name a few.

We introduce two new additions to the family, the Riemann zeta function (and other L-functions) and the $3x+1$ Problem (and other (d, g, h)-Maps), though we prove the theorems in sufficient generality to include other systems. Roughly, the distribution of digits of values of L-functions near the critical line and the ratio of observed versus predicted values of iterates of the $3x+1$ Map tend to Benford's Law. For exact statements of the results, see Theorem 4.4 and Corollary 4.5 for L-functions and Theorem 5.3 for the $3x+1$ Problem. While the best error terms just miss proving Benford behavior for L-functions on the critical line, we show that the values of the characteristic polynomials of unitary matrices are Benford in Appendix A; as these characteristic polynomials are believed to model the values of L-function, this and our theoretical results naturally lead to the conjecture that values of L-functions on the critical line are Benford.

A standard method of proving Benford behavior is to show the logarithms of the values become equidistributed modulo 1; Benford behavior then follows by exponentiation. There are two needed inputs. For both systems the main term of the distribution of the logarithms is a Gaussian, which can be shown to be equidistributed modulo 1 by Poisson summation. The second ingredient is to control the errors in the convergence of the distribution of the logarithms to Gaussians. For L-functions this is accomplished by Hejhal's refinement of the error terms (his result follows from an analysis of high moments of integrals of $\log |L(s, f)|$), and for the $3x+1$ Problem it involves an analysis of the discrepancy of the sequence $k \log_B 2 \bmod 1$ (which follows from $\log_B 2$ is of finite type; see below).

The reader should be aware that the standard notations from number theory and probability theory sometimes conflict; for example, σ is used to denote the real part of a point in the complex plane as well as the standard deviation of a distribution. We try and follow common custom as much as possible. We denote the Fourier transform (or characteristic function) of f by $\widehat{f}(y) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x y} dx$. Recall $g(T) = o(1)$ means $g(T) \to 0$ as $T \to \infty$, and $g(T) \ll h(T)$ or $g(T) = O(h(T))$ means there is some constant C such that for all T sufficiently large, $|g(T)| \le C h(T)$. Our proof of the Benford behavior of the $3x+1$ problem uses the (irrationality) type of $\log_B 2$ to control the errors; a number α is of type κ if κ is the supremum of all γ

with

$$\underline{\lim}_{q\to\infty} q^{\gamma+1} \min_p \left| \alpha - \frac{p}{q} \right| \;=\; 0. \tag{1.1}$$

By Roth's theorem, every algebraic irrational is of type 1. See for example [HS, Ro] for more details.

2. Benford's Law

To study leading digits we use the mantissa function, a generalization of scientific notation. Fix a base $B > 1$ and for a real number $x > 0$ define the mantissa function, $M_B(x)$, from the unique representation of x by

$$x \;=\; M_B(x) \cdot B^k, \text{ with } k \in \mathbb{Z} \text{ and } M_B(x) \in [1, B). \tag{2.1}$$

We extend the domain of mantissa to all of $\mathbb{C}$ via

$$M_B(x) \;=\; \begin{cases} 0 & \text{if } x = 0 \\ M_B(|x|) & \text{if } x \neq 0. \end{cases} \tag{2.2}$$

We study the mantissa of many different types of processes (discrete, continuous and mixed), and it is convenient to be able to use the same language for all. Take an ordered total space Ω, for example $\mathbb{N}$ or $\mathbb{R}^+$, and a (weak notion of) measure μ on Ω such as the counting measure or Lebesgue measure. For a subset $A \subset \Omega$ and an element $T \in \Omega$, denote by $A_T = \{\omega \in A : \omega \leq T\}$ the truncated set. We define the probability of A via density in Ω:

Definition 2.1. $\mathbb{P}(A) = \lim_{T\to\infty} \frac{\mu(A_T)}{\mu(\Omega_T)}$, *provided the limit exists.*

For $A \subset \mathbb{N}$ and μ the counting measure, $\mathbb{P}(A) = \lim_{T\to\infty} \frac{\#\{n\in A:\ n\leq T\}}{T}$, while if $A \subset \mathbb{R}^+$ and μ is Lebesgue measure then $\mathbb{P}(A) = \lim_{T\to\infty} \frac{\mu(0\leq t\leq T:\ t\in A)}{T}$. In Appendix A we extend our notion of probability to a slightly more general setting, but this will do for now.

For a sequence of real numbers indexed by Ω, $\overrightarrow{X} = \{x_\omega\}_{\omega\in\Omega}$, and a fixed $s \in [1, B)$, consider the pre-image of mantissa, $\{\omega \in \Omega : 1 \leq M_B(x_\omega) \leq s\}$; we abbreviate this by $\{1 \leq M_B(\overrightarrow{X}) \leq s\}$.

Definition 2.2. *A sequence $\overrightarrow{X}$ is said to be **Benford** (base B) if for all $s \in [1, B)$,*

$$\mathbb{P}\left\{1 \leq M_B(\overrightarrow{X}) \leq s\right\} \;=\; \log_B s. \tag{2.3}$$

Definition 2.2 is applicable to the values of a function f, and we say f is Benford base B if

$$\lim_{T\to\infty} \frac{\mu\left(0 \leq t \leq T : 1 \leq M_B\left(f(t)\right) \leq\ s\right)}{T} = \log_B s. \tag{2.4}$$

We describe an equivalent condition for Benford behavior which is based on equidistribution. Recall

Definition 2.3. *A set $A \subset \mathbb{R}$ is equidistributed modulo* 1 *if for any* $[a, b] \subset [0, 1]$ *we have*

$$\lim_{T\to\infty} \frac{\mu\left(\{x \in A_T : x \bmod 1 \in [a, b]\}\right)}{\mu\left(A_T\right)} \;=\; b - a. \tag{2.5}$$

The following two statements are immediate:

Lemma 2.4. *We have $u \equiv v \bmod 1$ if and only if the mantissa of B^u and B^v are the same, base B.*

Lemma 2.5. *We have $y \bmod 1 \in [0, \log_B s]$ if and only if B^y has mantissa in $[1, s]$.*

The following result is a standard way to prove Benford behavior:

Theorem 2.6. *Let $\overrightarrow{Y_B} = \log_B |\overrightarrow{X}|$, so pointwise $y_{\omega,B} = \log_B |x_\omega|$, and set $\log_B 0 = 0$. Then $\overrightarrow{Y_B}$ is equidistributed modulo 1 if and only if $\overrightarrow{X}$ is Benford base B.*

Proof. By Lemma 2.5, the set $\{\overrightarrow{Y_B} \bmod 1 \in [0, \log_B s]\}$ is the same as the set $\{M_B(\overrightarrow{X}) \in [1, s]\}$. Hence $\overrightarrow{Y_B}$ is equidistributed modulo 1 if and only if

$$\log_B s \;=\; \mathbb{P}\left\{\overrightarrow{Y_B} \bmod 1 \in [0, \log_B s]\right\} = \;\mathbb{P}\left\{M_B(\overrightarrow{X}) \in [1, s]\right\} \tag{2.6}$$

if and only if $\overrightarrow{X}$ is Benford base B. □

Theorem 2.6 reduces investigations of Benford's Law to equidistribution modulo 1, which we analyze below.

Remark 2.7. *The limit in Definition 2.1, often called the natural density, will exist for the sets in which we are interested, but need not exist in general. For example, if A is the set of positive integers with first digit 1, then $\frac{\#\{n\in A:\ n\le T\}}{T}$ oscillates between its $\liminf$ of $\frac{1}{9}$ and its $\limsup$ of $\frac{5}{9}$. One can study such sets by using instead the analytic density*

$$\mathbb{P}_{an}\left(A\right) \;=\; \lim_{s\to 1^+} \frac{\sum_{n\in A} n^{-s}}{\zeta(s)}, \tag{2.7}$$

where $\zeta(s)$ is the Riemann Zeta Function (see §4). A straightforward argument using analytic density gives Benford-type probabilities. In particular, Bombieri (see [Se]*, page 76) has noted that the analytic density of primes with first digit 1 is $\log_{10} 2$, and this can easily be generalized to Benford behavior for any first digit.*

3. Poisson Summation and Equidistribution modulo 1

We investigate systems $\overrightarrow{X_T}$ converging to a system $\overrightarrow{X}$ with associated logarithmic processes $\overrightarrow{Y_{T,B}}$. For example, take some function $g : \mathbb{R} \to \mathbb{C}$ and let $\overrightarrow{X} = \{g(t)\}_{t\in\mathbb{R}}$. Then $\overrightarrow{X_T} = \{g(t)\}_{0\le t\le T}$ are truncations of $\overrightarrow{X}$, with log-process $\overrightarrow{Y_{T,B}} = \{\log_B |g(t)|\}_{0\le t\le T}$. When there is no ambiguity we drop the dependence on B and write just $\overrightarrow{Y_T}$ for $\overrightarrow{Y_{T,B}}$.

Let $f(x)$ be a fixed probability density with cumulative distribution function $F(x) = \int_{-\infty}^{x} f(t)\,dt$. In our applications the probability densities of $\overrightarrow{Y_{T,B}}$ are approximately a spread version of f such as $f_T(x) = \frac{1}{T} f\left(\frac{x}{T}\right)$. There is, however, an error term, and the log-process $\overrightarrow{Y_{T,B}}$ has a cumulative distribution function given by

$$\begin{aligned} F_T(x) &= \mathbb{P}\left\{\overrightarrow{Y_{T,B}} \le x\right\} \\ &= \int_{-\infty}^{x} \frac{1}{T} f\left(\frac{t}{T}\right) dt + E_T(x) \\ &= F\left(\frac{x}{T}\right) + E_T(x), \end{aligned} \tag{3.1}$$

where E_T is an error term. Our goal is to show that, under certain conditions, the error term is negligible and $f_T(x)$ spreads to make $\overrightarrow{Y_{T,B}}$ equidistributed modulo 1 as $T \to \infty$. This will imply that $\overrightarrow{X}$ is Benford base B.

In our investigations we need the density f, cumulative distribution function F_T and errors E_T to satisfy certain conditions in order to control the error terms.

Definition 3.1 (Benford-good)**.** *Systems $\overrightarrow{Y_{T,B}}$ with cumulative distribution functions F_T are **Benford-good** if the F_T satisfy* (3.1)*, the probability density f satisfies sufficient conditions for Poisson Summation ($\sum_n f(n) = \sum_n \widehat{f}(n)$), and there is a monotone increasing function $h(T)$ with $\lim_{T\to\infty} h(T) = \infty$ such that f and E_T satisfy*

Condition 1. *Small tails:*

$$F_T(\infty) - F_T(Th(T)) \;=\; o(1), \qquad F_T(-Th(T)) - F_T(-\infty) \;=\; o(1). \tag{3.2}$$

Condition 2. *Rapid decay of the characteristic function:*

$$S(T) \;=\; \sum_{\substack{k\in\mathbb{Z}\\ k\neq 0}} \left| \frac{\widehat{f}(Tk)}{k} \right| \;=\; o(1). \tag{3.3}$$

Condition 3. *Small truncated translated error: for all* $0 \le a < b \le 1$,

$$\mathcal{E}_T(a,b) \ = \ \sum_{|k| \le Th(T)} \left[E_T(b+k) - E_T(a+k)\right] \ = \ o(1). \tag{3.4}$$

In all our applications f will be a Gaussian, in which case the Poisson Summation Formula holds. See for example [Da] (pages 14 and 63).

Condition 1 asserts that essentially all of the mass lies in $[-Th(T), Th(T)]$. In applications T will be the standard deviation, and this will follow from Central Limit type convergence.

Condition 2 is quite weak and is satisfied in all cases of interest. For example, if f is differentiable and f' is integrable (as is the case if f is the Gaussian density), then $|\widehat{f}(y)| \le \frac{1}{|y|} \int |f'(x)| dx = O\left(\frac{1}{|y|}\right)$, which suffices to show $S(T) = o(1)$.

Condition 3 is the most difficult to prove for a system, and to our knowledge has not previously been analyzed in full detail. It is well known (see [Fe]) that there are some processes (for example, Bernoulli trials) with standard deviation of size T where the best attainable estimate is $E_T(x) = O\left(\frac{1}{T}\right)$. Errors this large lead to $\mathcal{E}_T(a,b) = O(1)$.

We now see why these conditions suffice. For $[a,b] \subset [0,1)$, let $P_T(a,b)$ denote the probability that $\overrightarrow{Y_{T,B}} \bmod 1 \in [a,b]$. To prove $\overrightarrow{Y_{T,B}}$ becomes equidistributed modulo 1, we must show that $P_T[a,b] \to b-a$. We would like to argue as follows:

$$\begin{aligned} P_T[a,b] \ &= \ \mathbb{P}\left\{\overrightarrow{Y_{T,B}} \bmod 1 \in [a,b]\right\} \\ &= \ \sum_{k\in\mathbb{Z}} \mathbb{P}\left\{\overrightarrow{Y_{T,B}} \in [a+k, b+k]\right\} \\ &= \ \sum_{k\in\mathbb{Z}} \left(F_T\left(b+k\right) - F_T\left(a+k\right)\right) \\ &= \ \sum_{k\in\mathbb{Z}} \left[\int_a^b \frac{1}{T} f\left(\frac{x+k}{T}\right) dx + E_T(b+k) - E_T(a+k)\right] \\ &= \ \sum_{k\in\mathbb{Z}} \left[\int_a^b \frac{1}{T} f\left(\frac{x+k}{T}\right) dx\right] + \sum_{k\in\mathbb{Z}} \left[E_T(b+k) - E_T(a+k)\right]. \end{aligned} \tag{3.5}$$

While the main term can be handled by a straightforward application of Poisson Summation, the best pointwise bounds for the error term are not summable over all $k \in \mathbb{Z}$. This is why Condition 1 is necessary, so that we may restrict the summation.

Theorem 3.2. *Assume log-processes* $\overrightarrow{Y_{T,B}}$ *are **Benford-good**. Then* $\overrightarrow{Y_{T,B}} \to \overrightarrow{Y_B}$*, where* $\overrightarrow{Y_B}$ *is equidistributed modulo 1.*

Proof. As the Fourier transform converts translation to multiplication, if $g_x(u) = f\left(\frac{u+x}{T}\right)$ then a straightforward calculation shows that $\widehat{g_x}(w) = e^{2\pi i x w} T \widehat{f}(Tw)$ for any fixed x. Our assumptions on f allow us to apply Poisson Summation to g, and we find

$$\sum_{k\in\mathbb{Z}} f\left(\frac{x+k}{T}\right) \ = \ \sum_{k\in\mathbb{Z}} g_x(k) \ = \ \sum_{k\in\mathbb{Z}} \widehat{g_x}(k) \ = \ T\sum_{k\in\mathbb{Z}} e^{2\pi i x k}\widehat{f}(Tk). \quad (3.6)$$

Let $[a,b] \subset [0,1]$. By Condition 1 and (3.1),

$$\begin{aligned} P_T(a,b) \ &= \ \sum_{|k|\le Th(T)} \left(F_T(b+k) - F_T(a+k)\right) \\ &\qquad + O\left(F_T(\infty) - F_T(Th(T))\right) \\ &\qquad + O\left(F_T(-Th(T)) - F_T(-\infty)\right) \\ &= \ \sum_{|k|\le Th(T)} \left[\frac{1}{T}\int_a^b f\left(\frac{x+k}{T}\right)dx + E_T(b+k) - E_T(a+k)\right] + o(1) \\ &= \ \sum_{|k|\le Th(T)} \frac{1}{T}\int_a^b f\left(\frac{x+k}{T}\right)dx + \mathcal{E}_T(a,b) + o(1). \quad (3.7) \end{aligned}$$

By Condition 3, $\mathcal{E}_T(a,b) = o(1)$; as f is integrable we may return the sum to all $k \in \mathbb{Z}$ at a cost of $o(1)$. The interchange of summation and integration below is justified from the decay properties of f. To see this, simply insert absolute values in the arguments. Therefore using (3.6),

$$\begin{aligned} P_T[a,b] \ &= \ \frac{1}{T}\sum_{k\in\mathbb{Z}}\int_a^b f\left(\frac{x+k}{T}\right)dx + o(1) \\ &= \ \frac{1}{T}\int_a^b \left(\sum_{k\in\mathbb{Z}} g_x(k)\right)dx + o(1) \\ &= \ \frac{1}{T}\int_a^b \left(\sum_{k\in\mathbb{Z}} \widehat{g_x}(k)\right)dx + o(1) \\ &= \ \sum_{k\in\mathbb{Z}} \widehat{f}(Tk)\int_a^b e^{2\pi i x k}dx + o(1) \\ &= \ \widehat{f}(0)(b-a) + \sum_{k\ne 0}\widehat{f}(Tk)\frac{e^{2\pi i b k} - e^{2\pi i a k}}{2\pi i k} + o(1). \quad (3.8) \end{aligned}$$

As f is a probability density, $\widehat{f}(0) = 1$, and by Condition 2 the sum in (3.8) is $o(1)$. Therefore

$$P_T(a,b) \;=\; b - a + o(1), \tag{3.9}$$

which completes the proof. □

As an immediate consequence, we have:

Theorem 3.3. *Let $\overrightarrow{X_T}$ (the truncation of $\overrightarrow{X}$) have corresponding log-process $\overrightarrow{Y_{T,B}}$. Assume the $\overrightarrow{Y_{T,B}}$ are **Benford-good**. Then $\overrightarrow{X}$ is Benford base B.*

Proof. This follows immediately from Theorems 3.2 and 2.6. □

An immediate application of Theorem 3.3 is to processes where the distribution of the logarithms is exactly a spreading Gaussian (i.e., there are no errors to sum). We describe such a situation below.

Recall a *Brownian motion* (or Wiener process) is a continuous process with independent, normally distributed increments. So if W is a Brownian motion, then $W_t - W_s$ is a random variable having the Gaussian distribution with mean zero and variance $t-s$, and is independent of the random variable $W_s - W_u$ provided $u < s < t$.

A standard realization of Brownian motion is as the scaled limit of a random walk. Let $x_1, x_2, x_3, \dots$ be independent Bernoulli trials (taking the values $+1$ and -1 with equal probability) and let $S_n = \sum_{i=1}^n x_i$ denote the partial sum. Then the normalized process

$$W_t^{(n)} \;=\; \frac{1}{\sqrt{n}}\, S_{nt} \tag{3.10}$$

(extended to a continuous process by linear interpolation) converges as $n \to \infty$ to the Wiener process. See [Bi] or Chapter 2.4 of [KaSh] for further details.

A *geometric Brownian motion* is simply a process Y such that the process $\log Y$ is a Brownian motion. It was known to Benford that stock market indices empirically demonstrated this digit bias, and for almost as long these indices have been modelled by geometric Brownian motion. Thus Theorem 3.3 implies the well-known observation that

Corollary 3.4. *A geometric Brownian motion is Benford.*

4. Values of L-Functions

Consider the Riemann Zeta function

$$\zeta(s) \;=\; \sum_{n=1}^{\infty} \frac{1}{n^s} \;=\; \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}. \tag{4.1}$$

Initially defined for $\mathrm{Re}(s) > 1$, $\zeta(s)$ has a meromorphic continuation to all of $\mathbb{C}$. More generally, one can study an L-function

$$L(s,f) \ = \ \sum_{n=1}^{\infty} \frac{a_f(n)}{n^s} \ = \ \prod_{p \text{ prime}} \prod_{j=1}^{d} \left(1 - \frac{\alpha_{f,d}(p)}{p^s}\right)^{-1}, \qquad (4.2)$$

where the coefficients $a_f(n)$ have arithmetic significance. Common examples include Dirichlet L-functions (where $a_f(n) = \chi(n)$ for a Dirichlet character χ) and elliptic curve L-functions (where $a_f(p)$ is related to the number of points on the elliptic curve modulo p).

All the L-functions we study satisfy (after suitable renormalization) a functional equation relating their value at s to their value at $1-s$. The region $0 \le \mathrm{Re}(s) \le 1$ is called the critical strip, and $\mathrm{Re}(s) = \frac{1}{2}$ the critical line. The behavior of L-functions in the critical strip, especially on the critical line, is of great interest in number theory. The Generalized Riemann Hypothesis (GRH) asserts that the zeros of any "nice" L-function are on the critical line. The location of the zeros of $\zeta(s)$ is intimately connected with the error estimates in the Prime Number Theorem. The Riemann Zeta function can be expressed as the moment of the maximum of a Brownian Excursion, and the distribution of the zeros (respectively, values) of L-functions is believed to be connected to that of eigenvalues (respectively, values of characteristic polynomials) of random matrix ensembles. See [BPY, Con, KaSa, KeSn] for excellent surveys.

We investigate the leading digits of L-functions near the critical line, and show that the distribution of the digits of their absolute values is Benford (see Theorem 4.4 for the precise statement). The starting point of our investigations of values of the Riemann zeta function along the critical line $s = \frac{1}{2} + it$ is the log-normal law (see [Lau, Sel1]):

$$\lim_{T\to\infty} \frac{\mu\left(\left\{0 \le t \le T : \log\left|\zeta\left(\frac{1}{2}+\mathrm{i}t\right)\right| \ \le \ y\sqrt{\frac{1}{2}\log\log T}\right\}\right)}{T}$$
$$= \ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2} du. \qquad (4.3)$$

Thus the density of values of $\log\left|\zeta\left(\frac{1}{2}+\mathrm{i}t\right)\right|$ for $t \in [0,T]$ are well approximated by a Gaussian with mean zero and standard deviation

$$\psi_T \ = \ \sqrt{\tfrac{1}{2}\log\log T} + O(\log\log\log T). \qquad (4.4)$$

Such results are often used to investigate small values of $\left|\zeta\left(\frac{1}{2}+\mathrm{i}t\right)\right|$ and gaps between zeros. As such, the known error terms are too crude for our purposes. In particular, one has (trivially modifying (4.21) of [Hej] or (8) of [Iv]) that

$$\frac{\mu\left(\left\{t \in [T,2T] : a \leq \log\left|\zeta\left(\frac{1}{2}+\mathrm{i}t\right)\right| \leq b\right\}\right)}{T}$$
$$= \frac{1}{\sqrt{2\pi\psi_T^2}} \int_a^b e^{-u^2/2\psi_T^2} du + O\left(\frac{\log^2 \psi_T}{\psi_T}\right). \tag{4.5}$$

The main term is Gaussian with increasing variance, precisely what we require for equidistribution modulo 1. The error term, however, is too large for pointwise evaluation (as we have of the order $\psi_T \log \psi_T$ intervals $[a + n, b + n]$).

Better pointwise error estimates are obtained for many L-functions in [Hej]. These estimates are good enough for us to see Benford behavior as $T \to \infty$ near the line $\mathrm{Re}(s) = \frac{1}{2}$. Explicitly, consider an L-function (or a linear combination of L-functions, though for simplicity of exposition we confine ourselves to the case of one L-function) satisfying

Definition 4.1 (Good L-Function)**.** *We say an L-function is good if it satisfies the following properties:*

(1) *Euler product:*

$$L(s,f) \ = \ \sum_{n=1}^{\infty} \frac{a_f(n)}{n^s} \ = \ \prod_{p \text{ prime}} \prod_{j=1}^{d} \left(1 - \alpha_{f,j}(p)p^{-s}\right)^{-1}. \tag{4.6}$$

(2) *$L(s,f)$ has a meromorphic continuation to $\mathbb{C}$, is of finite order, and has at most finitely many poles (all on the line $\mathrm{Re}(s) = 1$).*

(3) *Functional equation:*

$$e^{\mathrm{i}\omega} G(s) L(s,f) \ = \ e^{-\mathrm{i}\omega} \overline{G(1-\overline{s}) L(1-\overline{s})}, \tag{4.7}$$

where $\omega \in \mathbb{R}$ and

$$G(s) \ = \ Q^s \prod_{i=1}^{h} \Gamma(\lambda_i s + \mu_i) \tag{4.8}$$

with $Q, \lambda_i > 0$ and $\mathrm{Re}(\mu_i) \geq 0$.

(4) *For some $\aleph > 0$, $c \in \mathbb{C}$, $x \geq 2$ we have*

$$\sum_{p \leq x} \frac{|a_f(p)|^2}{p} \ = \ \aleph \log\log x + c + O\left(\frac{1}{\log x}\right). \tag{4.9}$$

(5) *The $\alpha_{f,j}(p)$ are (Ramanujan-Petersson) tempered: $|\alpha_{f,j}(p)| \leq 1$.*

(6) *If $N(\sigma, T)$ is the number of zeros ρ of $L(s)$ with $\mathrm{Re}(\rho) \geq \sigma$ and $\mathrm{Im}(\rho) \in [0,T]$, then for some $\beta > 0$ we have*

$$N(\sigma,T) \ = \ O\left(T^{1-\beta\ \sigma-\frac{1}{2}}\ \log T\right). \tag{4.10}$$

Remark 4.2. *There are many families of L-functions which satisfy the above six conditions. The last two are the most difficult conditions to verify, as in all cases where these are known the first four conditions can be shown to be satisfied. The last two conditions are established for many L-functions (for example, see* [Sel1] *for $\zeta(s)$ and* [Luo] *for holomorphic Hecke cuspidal forms of full level and even weight $k > 0$; see Chapter 10* [IK] *for more on the subject), and is an immediate consequence of GRH.*

We quote a version of the log-normal law with better error terms (see (4.20) from [Hej] with a trivial change of variables in the Gaussian integral); for the convenience of the reader we list where the various parameters in Hejhal's result are defined. The error terms will be pointwise summable, and allow us to prove Benford behavior.

Theorem 4.3 (Hejhal). *Let $L(s,f)$ be a good L-function as in Definition 4.1, and*

- *fix $\delta \in (0,1)$ (*[Hej]*, Lemmas 2 and 3, page 556), $g \in (0,1]$ (*[Hej]*, Lemma 3, page 556) and $\kappa \in (1,3]$ (*[Hej]*, page 560 and (4.18) on page 562);*
- *choose $\sigma \geq \frac{1}{2} + \frac{g}{\log y}$ (*[Hej]*, page 563) and $\frac{1}{2} \leq \sigma \leq \frac{1}{2} + \frac{1}{\log^\delta T}$ (*[Hej]*, page 562);*
- *the variance $\psi(\sigma,T)$ (see* [Hej]*, Lemma 1, page 566) satisfies*

$$\psi(\sigma,T) \;=\; \aleph \log\left[\min\left(\log T, \frac{1}{\sigma-\frac{1}{2}}\right)\right] + O(1); \tag{4.11}$$

- *choose $N = \lfloor \psi(\sigma,T)^\kappa \rfloor$ and $y = T^{1/2N}$ (*[Hej]*, (4.18), page 565).*

Then we have

$$\frac{\mu\left(\{t \in [T,2T] : a \leq \log|L(\sigma+\mathrm{i}t,f)| \leq b\}\right)}{T} = \frac{1}{\sqrt{\psi(\sigma,T)}}\int_a^b e^{-\pi u^2/\psi(\sigma,T)}du$$
$$+\ O\left(\frac{1}{\psi(\sigma,T)}\min\left(1, \frac{|b-a|}{\sqrt{\psi(\sigma,T)}}\right) + \psi(\sigma,T)^{-\kappa/2} + y^{(1/3)(1-2\sigma)}\right), \tag{4.12}$$

the implied constant depends only on β (Condition (6) of Definition 4.1), f, δ, g and κ.

For our purposes, a satisfactory choice is to take $\sigma = \frac{1}{2} + \frac{1}{\log^\delta T}$ and $\kappa > 2$. Then $\psi(\sigma,T) = \aleph \log\log T + O(1)$ and

$$\begin{aligned} y^{(1/3)(1-2\sigma)} &= T^{\frac{1}{\log^\delta T}\frac{-1}{3(\aleph\log\log T + O(1))^\kappa}} = \exp\left(-\frac{\log^{1-\delta} T}{3(\aleph \log\log T + O(1))^\kappa}\right) \\ &\ll \frac{(\log\log T)^\kappa}{\log^{1-\delta} T}. \end{aligned} \tag{4.13}$$

We now show, in a certain sense, the values of $|L(s,f)|$ are Benford. While any modest cancellation would yield the following result on the critical line, due to our error terms for each interval $[T,2T]$ we must stay slightly to the right of $\mathrm{Re}(s)=\frac{1}{2}$.

Theorem 4.4. *Let $L(s,f)$ be a good L-function as in Definition 4.1; for example we may take $\zeta(s)$. If the GRH and Ramanujan conjectures hold we may take any cuspidal automorphic L-function; see also Remark 4.2. Fix a $\delta\in(0,1)$. For each T, let $\sigma_T=\frac{1}{2}+\frac{1}{\log^\delta T}$. Then*

$$\lim_{T\to\infty}\frac{\mu\left\{t\in[T,2T]:\ 1\le M_B\left(|L(\sigma_T+\mathrm{i}t,f)|\right)\le\tau\right\}}{T} \;=\; \log_B\tau. \tag{4.14}$$

Thus the values of the L-function satisfy Benford's Law in the limit (with the limit taken as described above) for any base B.

Proof. We first prove the claim for base e, and then comment on the changes needed for a general base B. Unfortunately the notation from number theory slightly conflicts with the standard notation from probability theory of §3. By Theorem 2.6, it suffices to show that

$$\lim_{T\to\infty}\frac{\mu\left\{t\in[T,2T]:\ a\le\log|L(\sigma_T+\mathrm{i}t,f)|\le b\right\}}{T} \;=\; b-a. \tag{4.15}$$

Let $\psi_T=\psi(\sigma_T,T)$ be the variance of the Gaussian in (4.12), which tends to infinity with T. The standard deviation is thus $\sqrt{\psi_T}$, and corresponds to what we called T in §3. Let $\eta(x)$ be the standard normal (mean zero, variance one; η plays the role of f from §3 – as it is standard to denote L-functions by $L(s,f)$, we use η here and in §5), and set $\eta_{\sqrt{\psi_T}}(x)=\frac{1}{\sqrt{\psi_T}}\,\eta\left(\frac{x}{\sqrt{\psi_T}}\right)$. Note $\eta_{\sqrt{\psi_T}}(x)$ is the density of a normal with mean zero and variance ψ_T. By (4.12) we have

$$F_T(x) \;=\; \int_{-\infty}^{x}\eta_{\sqrt{\psi_T}}(x)dx+E_T(x), \tag{4.16}$$

where $E_T(x)=O(\psi_T^{-1})$. We must show the logarithms of the absolute values of the L-function are Benford-good. As η is a Gaussian it satisfies the conditions for the Poisson Summation Formula, and the log-process $\overrightarrow{Y_T}=\log|L(\sigma_T+\mathrm{i}t,f)|$ satisfies (3.1). Thus to apply Theorem 3.3 it suffices to show η, F_T and $\mathcal{E}_T$ satisfy Conditions 1 through 3 for some monotone increasing function $h(\psi_T)$ with $\lim_{T\to\infty}h(\psi_T)=\infty$. We take $h(\psi_T)=\sqrt{\log\psi_T}$.

Condition 1 is immediately verified. To show $F_{\sqrt{\psi_T}}(\infty)-F_{\sqrt{\psi_T}}(\sqrt{\psi_T}h(\psi_T)) = o(1)$ we use (4.12) to conclude the contribution from the error is $o(1)$, and then note that the integral of the Gaussian with standard deviation $\sqrt{\psi_T}$

past $\sqrt{\psi_T \log \psi_T}$ is small (as η is the density of the standard normal, this integral is dominated by

$$\frac{1}{\sqrt{2\pi}} \int_{|x| \geq \sqrt{\log \psi_T}} \eta(x)dx, \tag{4.17}$$

which is $o(1)$). Identical arguments show $F_{\sqrt{\psi_T}}(-\sqrt{\psi_T}h(\psi_T)) - F_{\sqrt{\psi_T}}(-\infty) = o(1)$. As we are integrating a sizable distance past the standard deviation, it is easy to see that the contribution from the Gaussian is small. We do not need the full strength of the bounds in (4.12); the bounds from (4.5) suffice to control the errors.

Condition 2 follows from the trivial fact that η' is integrable. We now show Condition 3 holds. Here the bounds from (4.5) just fail. Using those bounds and summing over $|k| \leq \sqrt{\psi_T}h(\psi_T)$ would yield an error of size $O\left(\sqrt{\psi_T}h(\psi_T) \cdot \frac{\log^2 \sqrt{\psi_T}}{\sqrt{\psi_T}}\right) = O\left(\log^{2.5} \psi_T\right)$. We instead use (4.12), and find for $[a,b] \subset [0,1]$ that

$$\begin{aligned} \mathcal{E}_T(a,b) &= \sum_{|k| \leq \sqrt{\psi_T}h(\psi_T)} [E_T(b+k) - E_T(a+k)] \\ &= \sum_{|k| \leq \sqrt{\psi_T \log \psi_T}} O\left(\frac{1}{\psi_T} \min\left(1, \frac{|b-a|}{\sqrt{\psi_T}}\right) + \psi_T^{-\kappa/2} + y^{(1/3)(1-2\sigma)}\right) \\ &= O\left(\frac{\sqrt{\log \psi_T}}{\sqrt{\psi_T}} + \psi_T^{\frac{1}{2}-\frac{\kappa}{2}}\sqrt{\log \psi_T} + \sqrt{\psi_T \log \psi_T}\frac{(\log\log T)^\kappa}{\log^{1-\delta} T}\right) \\ &= o(1) \end{aligned} \tag{4.18}$$

because $\kappa > 1$, $\delta < 1$ and $\psi_T \ll \log\log T$.

As all the conditions of Theorem 3.2 are satisfied, we can conclude that

$$P_{\sqrt{\psi_T}}(a,b) \ = \ b - a + o(1). \tag{4.19}$$

We have shown that tending to infinity in this manner, the distribution corresponding to $\log|L(\sigma_T + it, f)|$ converges to being equidistributed modulo 1, which by Theorem 3.3 implies the values of $|L(\sigma_T + it, f)|$ are Benford base e (as always, along the specified path converging to the critical line).

For a general base B, note $\log_B x = \frac{\log x}{\log B}$. The effect of changing base is that $\log_B |L(\sigma_T + \mathrm{i}t, f)|$ converges to a Gaussian with mean zero and variance $\frac{1}{\log B} \cdot \sqrt{\psi(\sigma_T, T)}$ (instead of mean zero and variance $\sqrt{\psi(\sigma_T, T)}$). The argument now proceeds as before. □

Corollary 4.5. *Theorem 4.4 is valid if instead of intervals* $[T, 2T]$ *we consider intervals* $[0, T]$.

Proof. Let $\alpha(T) = (\log\log\log T)^{\log 2}$. We consider the intervals $I_0 = [0, T/\alpha(T)]$ and

$$I_i \;=\; \left[2^{i-1}T/\alpha(T),\ 2^i T/\alpha(T)\right], \quad i \in \{1, 2, \dots, \log\log\log\log T\}. \qquad (4.20)$$

We may ignore I_0 as it has length $o(T)$. For each interval I_i, $i \geq 1$, we use (4.12) and argue as before. We may keep the same values of $\beta, \delta, g, \kappa, \sigma_T$ as before. T and y change, which implies $\psi_T = \psi(\sigma_T, T)$ changes; however, the leading term of ψ_T is still $\aleph \log\log T$, and $y^{(1/3)(1-2\sigma)}$ again leads to negligible contributions. As there are only $\log\log\log\log T$ intervals, we may safely add all the errors. □

Remark 4.6. *If we stay a fixed distance off the critical line, we do not expect Benford behavior. This is because for a fixed $\sigma > \frac{1}{2}$, for $\zeta(s)$ we have a distribution function G_σ such that*

$$\lim_{T\to\infty} \frac{\mu\{t \in [0,T] : \log|\zeta(\sigma + \mathrm{i}t)| \in [a,b]\}}{T} \;=\; \int_a^b G_\sigma(u)du. \qquad (4.21)$$

Unlike the log-normal law (4.5)*, where the variance increases with T, note here there is no increasing variance for fixed σ (though of course the variance depends on σ); see* [BJ, JW] *for proofs. Thus to see Benford behavior it is essential that as T increases our distance to the critical line decreases.*

For investigations on the critical line, one can easily show Benford's Law holds for a truncation of the series expansion of $\log|L(\frac{1}{2} + \mathrm{i}t, f)|$, where the truncation depends on the height T. See (4.12) of [Hej] for the relevant version of the log-normal law (which has a significantly better error term than (4.12)). Similarly, one can prove statements along these lines for the real and imaginary parts of L-functions.

Numerical investigations also support the conjectured Benford behavior. In Figure 1 we plot the percent of first digits of $\left|\zeta\left(\frac{1}{2} + \mathrm{i}t\right)\right|$ versus the Benford probabilities for $t = \frac{k}{4}$, $k \in \{0, 1, \dots, 65535\}$, and note the Benford behavior quickly sets in. Of course, we believe that this is strong evidence for Benford behavior exactly on the critical line, but as they stand, our error terms are too big and our cancellation too small to demonstrate this statement.

It is believed that values of characteristic polynomials of random matrix ensembles model values of L-functions on the critical line. In Theorem A.2 of Appendix A we show that the digit distribution of the values of these characteristic polynomials converge to the Benford probabilities (as the size of the matrices tend to infinity), providing additional support for the conjecture that L-functions are Benford on the critical line.

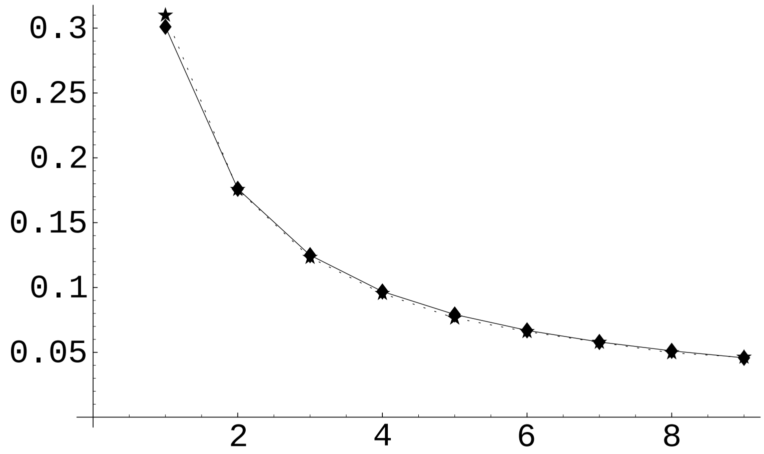

FIGURE 1. Distribution of Digits of $|\zeta(s)|$ versus Benford Probabilities

5. THE $3x+1$ PROBLEM

People working on the Syracuse-Kakutani-Hasse-Ulam-Hailstorm-Collatz-$(3x+1)$-Problem (there have been a few) often refer to two striking anecdotes. One is Erdös' comment that "Mathematics is not yet ready for such problems." The other is Kakutani's communication to Lagarias: "For about a month everybody at Yale worked on it, with no result. A similar phenomenon happened when I mentioned it at the University of Chicago. A joke was made that this problem was part of a conspiracy to slow down mathematical research in the U.S." Coxeter has offered \$50 for its solution, Erdös \$500, and Thwaites, £1000. The problem has been connected to holomorphic solutions to functional equations, a Fatou set having no wandering domain, Diophantine approximation of $\log_2 3$, the distribution mod 1 of $\left\{\left(\frac{3}{2}\right)^k\right\}_{k=1}^{\infty}$, ergodic theory on $\mathbb{Z}_2$, undecidable algorithms, and geometric Brownian motion, to name a few (see [Lag1, Lag2]). We now relate the $(3x+1)$-Problem to Benford's Law.

5.1. **The Structure Theorem.** If x is a positive odd integer then $3x+1$ is even, so we can find an integer $k \geq 1$ such that $2^k \parallel (3x+1)$, i.e. so that

$$y \;=\; \frac{3x+1}{2^k} \tag{5.1}$$

is also odd. In this way, we get the $(3x+1)$-Map

$$M : x \longmapsto y. \tag{5.2}$$

We call the value of k that arises in the definition of y the k**-value** of x. Notice that y is odd and relatively prime to 3, so the natural domain for iterating M is the set Π of positive integers prime to 2 and 3. Write $\Pi = 6\mathbb{N} + E$, where $E = \{1, 5\}$ is the set of possible congruence classes modulo 6. The total space is $\Omega = \Pi$, not $\mathbb{N}$ or $\mathbb{R}$, and the measure is the appropriate counting measure.

For every integer $x \in \Pi$ with $0 < x < 2^{60}$, computers have verified that enough iterations of the $(3x+1)$-Map eventually send x to the unique fixed point, 1. The natural conjecture asks if the same statement holds for all $x \in \Pi$:

Conjecture 5.1 ($(3x+1)$-Conjecture)**.** *For every $x \in \Pi$, there is an integer n such that $M^n(x) = 1$.*

Suppose we apply M a total of m times, calling $x_0 = x$ and $x_i = M^i(x)$, $i \in \{1, 2, \dots, m\}$. For each x_{i-1} there is a k-value, say k_i, such that

$$x_i \ = \ M(x_{i-1}) \ = \ \frac{3x_{i-1} + 1}{2^{k_i}}, \qquad i \ \in \ \{1, 2, \dots, m\}. \tag{5.3}$$

We store this information in an ordered m-tuple $(k_1, k_2, \dots, k_m)$, called the m**-path** of x. Let γ_m denote the map sending x to its m-path,

$$\gamma_m : x \mapsto (k_1, k_2, \dots, k_m). \tag{5.4}$$

The natural question is whether given an m-tuple of positive integers $(k_1, k_2, \dots, k_m)$, there is an integer x whose m-path is precisely this m-tuple. If so, we would like to classify the set of all such x. In other words, we want to study the inverse map γ_m^{-1}.

The answer is given by the Structure Theorem, proved in [KonSi]: for each m-tuple $(k_1, k_2, \dots, k_m)$, not only does there exist an x having this m-path, but this path is enjoyed by two full arithmetic progressions, $x \in \{a_1 n + b_1, a_2 n + b_2\}_{n=0}^{\infty}$, and we can solve explicitly for a_i and b_i. In fact, $a_1 = a_2 = 6 \cdot 2^{k_1 + k_2 + \cdots + k_m}$, and $b_i < a_i$ (so the progressions are full; we do not miss any terms at the beginning). Moreover, the two progressions fall into the two possible equivalence classes modulo 6; i.e., $\{b_1 \bmod 6, b_2 \bmod 6\} = \{1, 5\}$. The structure theorem is the key ingredient in analyzing the limiting distributions. These will satisfy the conditions of our main theorem (Theorem 3.3), and yield Benford's Law.

Recall (Definition 2.1) that we define the probability of a subset $A \subset \Pi$ by

$$\mathbb{P}(A) \ = \ \lim_{T \to \infty} \frac{|A_T|}{|\Pi_T|}, \tag{5.5}$$

provided the limit exists. We say a random variable ξ has geometric distribution with parameter $\frac{1}{2}$ (for brevity, **geometrically distributed**) if $\mathbb{P}(\xi = n) = \frac{1}{2^n}$ for $n = 1, 2, \dots$. A consequence of the structure theorem is that

$$\mathbb{P}\left(x : \gamma_m(x) = (k_1, \dots, k_m)\right) \;=\; \frac{1}{2^{k_1+\cdots+k_m}} \;=\; \prod_{i=1}^{m} \frac{1}{2^{k_i}}. \tag{5.6}$$

Both the expectation and variance of a geometrically distributed random variable is 2. For a seed x_0 let $x_m = M^m(x_0)$ be the m^{th} iterate. A natural quantity to investigate is $\frac{x_m}{\left(\frac{3}{4}\right)^m x_0}$, where $\left(\frac{3}{4}\right)^m x_0$ is the expected value of x_m.

Theorem 5.2 ([KonSi])**.** *The k-values are independent geometrically distributed random variables. Further, for any $a \in \mathbb{R}$*

$$\mathbb{P}\left(\frac{\log_2\left[\frac{x_m}{\left(\frac{3}{4}\right)^m x_0}\right]}{\sqrt{2m}} \le a\right) \;=\; \mathbb{P}\left(\frac{S_m - 2m}{\sqrt{2m}} \le a\right), \tag{5.7}$$

where S_m is the sum of m geometrically distributed (with parameter $\frac{1}{2}$) i.i.d.r.v. By the Central Limit Theorem, the right hand side converges to a Gaussian integral as $m \to \infty$. The paths are also independent, and so the $(3x+1)$*-Paths are those of a geometric Brownian motion with drift* $\log \frac{3}{4}$.

We remind the reader that a Brownian motion (and hence a geometric Brownian motion) can be realized as the limit of a random walk; the same phenomenon occurs here. The drift corresponds to the fact that the expected value is $\left(\frac{3}{4}\right)^m x_0$, rather than just x_0.

It is worth remarking that a consequence of the drift being $\log \frac{3}{4}$ (which is negative) is that it is natural to expect that typical trajectories return to the origin. This statement extends completely to (d, g, h)-Maps discussed in Appendix B. Theorem 5.2 is immediately applicable to investigations in base two (which is uninteresting as all first digits are 1). To study the $3x+1$ Problem in base B, one simply multiplies by $\frac{1}{\log_2 B}$, as $\frac{\log_2 x}{\log_2 B} = \log_2 B$. This replaces $S_m - 2m$ with $\frac{S_m - 2m}{\log_2 B}$ or $(S_m - 2m)\log_B 2$.

5.2. **A Tale of Two Limits.** The $(3x+1)$-system, $\overrightarrow{X_T} = \{x_i\}_{0 \le i \le T}$, is probably not Benford for *any* starting seed x_0 as we expect all of the terms to eventually be 1. If we stop the sequence after hitting 1 and consider the proportion of terms having a given leading digit j, this is a rational number, whereas $\log_{10} j$ is not. Of course, this rational number should be close to $\log_{10} j$, although it is easy to find arbitrarily large numbers decaying to 1 after even one iteration of the $(3x+1)$-map.

One sense in which Benford behavior can be proved is the same as the sense in which $(3x+1)$-paths are those of a geometric Brownian motion. We use the structure theorem to prove

Theorem 5.3. *Let B be any real number such that $\log_B 2$ is irrational of type $\kappa < \infty$; for example, one may take any integer B which is not a perfect power of 2 (see (1.1) for a definition of type κ and Theorem B.1 for a proof of the irrationality type of such integers). Then for any $[a,b] \subset [0,1]$,*

$$\lim_{m\to\infty} \mathbb{P}\left(\log_B \left[\frac{x_m}{\left(\frac{3}{4}\right)^m x_0}\right] \mod 1 \in [a,b]\right) \ = \ b-a. \tag{5.8}$$

As $\left(\frac{3}{4}\right)^m x_0$ is the expected value of x_m, this implies the distribution of the ratio of the actual versus predicted value after m iterates obeys Benford's Law (base B). If $B = 2^n$ for some integer n, in the limit $\log_B \left[\frac{x_m}{\left(\frac{3}{4}\right)^m x_0}\right]$ mod 1 takes on the values $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}$ with equal probability, leading to a non-Benford digit bias depending only on n.

Notice that since probability is defined through density, this is really two highly non-interchangeable limits:

$$\begin{aligned}
&\lim_{m\to\infty} \mathbb{P}\left(\log_B \left[\frac{x_m}{\left(\frac{3}{4}\right)^m x_0}\right] \bmod 1 \in [a,b]\right) \\
&= \lim_{m\to\infty}\lim_{T\to\infty} \frac{\#\left\{x_0 \in \Pi_T : \log_B \left[\frac{x_m}{\left(\frac{3}{4}\right)^m x_0}\right] \bmod 1 \in [a,b]\right\}}{\#\Pi_T}.
\end{aligned} \tag{5.9}$$

Though this is completely natural, it is worth remarking for the sake of precision. Of course, a good starting seed (one with a long life-span) should give a close approximation of Benford behavior, just as it will also be a generic Brownian sample path; this is supported by numerical investigations (see §5.4).

Let $\xi_1, \xi_2, \dots$ be independent geometrically distributed random variables with $\mathbb{P}(\xi_i = n) = \frac{1}{2^n}$, $n = 1, 2, \dots$, and $\mathbb{E}(\xi_i) = 2$, $\mathrm{Var}(\xi_i) = 2$. Let $S_m = \sum_{i=1}^m \xi_i$. Let $\zeta_i = \xi_i - 2$, $\overline{S}_m = \sum_{i=1}^m \zeta_i = S_m - 2m$. We know the distribution of $\log_B \left[\frac{x_m}{\left(\frac{3}{4}\right)^m x_0}\right]$ is the same as that of $(S_m - 2m)\log_B 2 = \overline{S}_m \log_B 2$. The proof is complicated by the fact that the sum of m geometrically distributed random variables itself has a binomial distribution, supported on the integers. This gives a lattice distribution for which we cannot obtain sufficient bounds on the error, even by performing an Edgeworth expansion and estimating the rate of convergence in the Central Limit Theorem. The problem is that the error in missing a lattice point is of size $\frac{1}{\sqrt{m}}$, and we need to sum $\sqrt{m}h(m)$ terms (for some $h(m) \to \infty$). We are able to surmount this obstacle by an error analysis of the rate of convergence to equidistribution of $k \log_B 2$ mod 1.

5.3. **Proof of Theorem 5.3.** To prove Theorem 5.3 we first collect some needed results. The proof is similar in spirit to Theorem 3.3, with the needed results playing a similar role as the three conditions; however, the discreteness of the $3x+1$ problem leads to some interesting technical complications, and it is easier to give a similar but independent proof than to adjust notation and show Conditions 1 through 3 are satisfied.

In the statements below, $[a,b]$ is an arbitrary sub-interval of $[0,1]$. By the Central Limit Theorem, the distribution of $\overline{S}_m$ (although it only takes integer values) is approximately a Gaussian with standard deviation of size $\sqrt{m}$. Let $c \in \left(0, \frac{1}{2}\right)$ and set $M = m^c$. Let

$$I_\ell \ = \ \{\ell M, \ell M + 1, \dots, (\ell+1)M - 1\} \tag{5.10}$$

and $C = \log_B 2$ be an irrational number of type κ (see (1.1)). Soundararajan informed us that one does not need $\log_B 2$ to be of finite type for our applications. For integer B, if $B^p - 2^q > 0$ then it is at least 1, and one obtains $o(M)$ instead of $O(M^\delta)$ in (5.15); the advantage of using finite type is we obtain sharper estimates on the rate of convergence, as well as being able to handle non-integral bases B.

Let $\eta(x)$ denote the density of the standard normal:

$$\eta(x) \ = \ \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}. \tag{5.11}$$

We collect some results needed for the proof of Theorem 5.3:

- By the Central Limit Theorem (see [Fe], Chapter XV), for any $k \in \mathbb{Z}$ we have

$$\begin{aligned} \mathrm{Prob}(C \cdot \overline{S}_m = C \cdot k) \ &= \ \mathrm{Prob}\left(\frac{\overline{S}_m}{\sqrt{m}} = \frac{k}{\sqrt{m}}\right) \\ &= \ \frac{1}{\sqrt{m}}\, \eta\left(\frac{k}{\sqrt{m}}\right) + o\left(\frac{1}{\sqrt{m}}\right). \end{aligned} \tag{5.12}$$

  We may write $o\left(\frac{1}{\sqrt{m}}\right)$ as $O\left(\frac{1}{\sqrt{m}g(m)}\right)$ for some monotone increasing $g(m)$ which tends to infinity. We use this to approximate the probability of $\overline{S}_m = k$. For future use, choose any monotone $h(m)$ tending to infinity such that $h(m) = o\left(g(m)\right)$, $h(m) = o\left(m^{1/2005}\right)$ and $\frac{h(m)M}{\sqrt{m}} = o\left(m^{-1/2005}\right)$. As $M = m^c$ with $c < \frac{1}{2}$, if c is sufficiently small then such an h exists.

- Let $k_1, k_2 \in I_\ell$. Then
$$\left|\frac{1}{\sqrt{m}}\,\eta\left(\frac{k_1}{\sqrt{m}}\right) - \frac{1}{\sqrt{m}}\,\eta\left(\frac{k_2}{\sqrt{m}}\right)\right| \leq \frac{1}{\sqrt{m}}\,e^{-\ell^2M^2/2m} \cdot \left(1-\exp\left(-\frac{2\ell M^2+M^2}{2m}\right)\right). \tag{5.13}$$
In practice this implies that for the ℓ we must study, there is negligible variation in the Gaussian for $k \in I_\ell$.
- By Poisson Summation (see page 63 of [Da]),
$$\frac{1}{\sigma}\sum_{n=-\infty}^{\infty} e^{-n^2\pi/\sigma^2} \ = \ \sum_{n=-\infty}^{\infty} e^{-n^2\pi\sigma^2}, \qquad \sigma > 0. \tag{5.14}$$
We often take $\sigma^2 = \frac{2m}{\pi M^2}$, and use this to calculate the main term (as $\sigma \to \infty$, both sides of (5.14) tend to 1).
- For any $\epsilon > 0$, letting $\delta = 1+\epsilon-\frac{1}{\kappa} < 1$ we have
$$\#\{k \in I_\ell : kC \bmod 1 \in [a,b]\} \ = \ M(b-a) + O(M^\delta). \tag{5.15}$$
The quantification of the equidistribution of $kC \bmod 1$ is the key ingredient in proving Benford behavior base B (with $C = \log_B 2$). The rate of equidistribution, given the finiteness of the irrationality type of C, follows from the Erdös-Turan Theorem. As this is the key argument in our analysis, we provide a sketch of the proof in Appendix B; see Theorem 3.3 on page 124 of [KN] for complete details (while the proof given only applies for I_0, a trivial translation yields the claim for any I_ℓ).

Proof of Theorem 5.3. We must show that as $m \to \infty$, for any $[a,b] \subset [0,1]$,
$$P_m(a,b) \ = \ \text{Prob}(C\overline{S}_m \bmod 1 \in [a,b]) \tag{5.16}$$
tends to $b-a$. We have
$$P_m(a,b) \ = \ \sum_{|\ell| \leq \frac{\sqrt{m}h(m)}{M}} \text{Prob}(\overline{S}_m = k \in I_\ell : kC \bmod 1 \in [a,b]) + \sum_{|\ell| > \frac{\sqrt{m}h(m)}{M}} \text{Prob}(\overline{S}_m = k \in I_\ell : kC \bmod 1 \in [a,b]). \tag{5.17}$$
The second sum in (5.17) is bounded by
$$\text{Prob}\left(\overline{S}_m = k : |k| \geq \frac{\sqrt{m}h(m)}{M}\right). \tag{5.18}$$

By the Central Limit Theorem, (5.18) is $o(1)$. Alternatively, using the techniques below (with $[a,b]=[0,1]$), one can show $\text{Prob}\left(|\overline{S}_m| \le \frac{\sqrt{m}h(m)}{M}\right) = 1+o(1)$, which implies (5.18) is $o(1)$. As we are not summing (5.18), it is okay to have an error here of size $\frac{1}{\sqrt{m}}$ (and errors of approximately this size arise if we add or subtract a lattice point). Therefore

$$\begin{aligned} P_m(a,b) &= \sum_{|\ell| \le \frac{\sqrt{m}h(m)}{M}} \text{Prob}(\overline{S}_m = k \in I_\ell : kC \bmod 1 \in [a,b]) + o(1) \\ &= \sum_{|\ell| \le \frac{\sqrt{m}h(m)}{M}} P_{m,\ell}(a,b) + o(1). \end{aligned} \tag{5.19}$$

The proof is completed by showing the above is $b-a+o(1)$. Consider an interval I_ℓ. By (5.15), the number of $k \in I_\ell$ such that $kC \bmod 1 \in [a,b]$ is $(b-a)M + O(M^\delta)$, $\delta < 1$. By (5.12), the probability of each such k is $\frac{1}{\sqrt{m}}\,\eta\left(\frac{k}{\sqrt{m}}\right) + O\left(\frac{1}{\sqrt{m}g(m)}\right)$. We now use (5.13) to bound the error from evaluating all the $\eta\left(\frac{k}{\sqrt{m}}\right)$ at $k=\ell M$ and find

$$\begin{aligned} P_{m,\ell}(a,b) = \frac{(b-a)M}{\sqrt{m}} &\left[\eta\left(\frac{\ell M}{\sqrt{m}}\right) + O\left(e^{-\ell^2M^2/2m}\right)\cdot\left(1-\exp\left(-\frac{2\ell M^2+M^2}{2m}\right)\right)\right] \\ &+ O\left(M\cdot\frac{1}{\sqrt{m}g(m)}\right) + O\left(M^\delta\cdot\frac{1}{\sqrt{m}}\,\eta\left(\frac{\ell M}{\sqrt{m}}\right)\right); \end{aligned} \tag{5.20}$$

summing over all $|\ell| \le \frac{\sqrt{m}h(m)}{M}$ gives $P_m(a,b)+o(1)$. This gives four sums, which we must show are $b-a+o(1)$.

The sums over $|\ell| \le \frac{\sqrt{m}h(m)}{M}$ of the first and fourth pieces of (5.20) are handled by Poisson Summation. We have for the first piece that

$$\begin{aligned} &\sum_{|\ell| \le \frac{\sqrt{m}h(m)}{M}} \frac{(b-a)M}{\sqrt{m}}\,\eta\left(\frac{\ell M}{\sqrt{m}}\right) \\ &\quad = \sum_{\ell=-\infty}^{\infty} \frac{(b-a)M}{\sqrt{m}}\,\eta\left(\frac{\ell M}{\sqrt{m}}\right) - \sum_{|\ell| > \frac{\sqrt{m}h(m)}{M}} \frac{(b-a)M}{\sqrt{m}}\,\eta\left(\frac{\ell M}{\sqrt{m}}\right). \end{aligned} \tag{5.21}$$

As $h(m) \to \infty$, the second sum in (5.21) is bounded by

$$\int_{|x| \ge \frac{\sqrt{m}h(m)}{M}} \frac{1}{\sqrt{2\pi m/M^2}}\,e^{-x^2/2(m/M^2)}dx \;=\; \frac{1}{\sqrt{2\pi}}\int_{|u|\ge h(m)} e^{-u^2/2}du \;=\; o(1). \tag{5.22}$$

Using (5.14) with $\sigma^2 = \frac{2m}{\pi M^2}$ gives

$$\begin{aligned}
\sum_{|\ell| \le \frac{\sqrt{m}h(m)}{M}} \frac{(b-a)M}{\sqrt{m}} \eta\left(\frac{\ell M}{\sqrt{m}}\right) &= (b-a) \sum_{\ell=-\infty}^{\infty} \frac{1}{\sqrt{2\pi m/M^2}}\ e^{-\ell^2/2(m/M^2)} + o(1) \\
&= (b-a) \sum_{\ell=-\infty}^{\infty} e^{-\ell^2 \cdot 2\pi^2 m/M^2} + o(1) \\
&= b-a+O\left(\frac{e^{-2\pi^2 m/M^2}}{1-e^{-2\pi^2 m/M^2}}\right) + o(1) \qquad (5.23)
\end{aligned}$$

as the final sum over $\ell \neq 0$ is bounded by a geometric series and $M = m^c$ with $c < \frac{1}{2}$. Thus the first piece from (5.20) gives $b - a + o(1)$.

As the Gaussian is a monotone function (for $x \geq 0$ or $x \leq 0$), a similar argument shows the sum over $|\ell| \leq \frac{\sqrt{m}h(m)}{M}$ of the fourth piece of (5.20) contributes $O(M^{\delta-1}) + o(1)$. It is here that we use $C\overline{S}_m$ is a very special equidistributed sequence modulo 1, namely it is of the form kC mod 1. This allows us to control the discrepancy (how many $k \in I_\ell$ give kC mod $1 \in [a, b]$).

We must now sum over $|\ell| \leq \frac{\sqrt{m}h(m)}{M}$ the second and third pieces of (5.20). For the second piece, we have

$$\sum_{|\ell| \le \frac{\sqrt{m}h(m)}{M}} \frac{M}{\sqrt{m}}\ e^{-\ell^2 M^2/2m} \left[1 - \exp\left(-\frac{2\ell M^2 + M^2}{2m}\right)\right]. \qquad (5.24)$$

As $|\ell| \leq \frac{\sqrt{m}h(m)}{M}$ and $M = m^c$ with $c < \frac{1}{2}$, we have

$$\frac{2\ell M^2 + M^2}{2m} \ll \frac{h(m)M}{\sqrt{m}}. \qquad (5.25)$$

Recall we chose $h(m)$ and c such that $\frac{h(m)M}{\sqrt{m}} = o\left(m^{-1/2005}\right)$. Therefore

$$1 - \exp\left(-\frac{2\ell M^2 + M^2}{2m}\right) \ll m^{-1/2005}. \qquad (5.26)$$

As we chose $h(m)$ such that $h(m) = o\left(m^{1/2005}\right)$, the sum in (5.24) is

$$\ll \frac{\sqrt{m}h(m)}{M} \cdot \frac{M}{\sqrt{m}} \frac{1}{m^{1/2005}} = \frac{h(m)}{m^{1/2005}} = o(1), \qquad (5.27)$$

proving the second piece in (5.20) is negligible.

We are left with the sum over $|\ell| \leq \frac{\sqrt{m}h(m)}{M}$ of the third piece in (5.20). Its contribution is

$$O\left(\frac{\sqrt{m}h(m)}{M} \cdot \frac{M}{\sqrt{m}g(m)}\right) = O\left(\frac{h(m)}{g(m)}\right) = o(1). \qquad (5.28)$$

Collecting the evaluations of the sums of the four pieces in (5.20), we see that

$$P_m(a,b) \ = \ b-a+o(1), \tag{5.29}$$

which completes the proof of Theorem 5.3 if $B \neq 2^n$ (and thus proves Benford behavior base 10 because, by Theorem B.1, $\log_{10} 2$ has finite irrationality type).

Consider now the case when $B = 2^n$. As S_m takes on integer values, the possible values modulo 1 for $(S_m - 2m)\log_B 2$ are $\{0, \frac{1}{n}, \dots, \frac{n-1}{n}\}$. An identical argument shows each of these values is equally likely; by determining which intervals $[\log_B d, \log_B(d+1))$ they lie in, one can determine the (non-Benford) digit bias in this case. See also §5.4. □

In Appendix C a generalization of the $3x+1$ map is discussed; for such systems, one can easily prove the analogue of Theorem 5.3.

5.4. **Numerical Investigations.** Theorem 5.3 implies that the first digit of $\frac{x_m}{\left(\frac{3}{4}\right)^m x_0}$ will not be Benford in a base $B = 2^n$. As S_m takes on integer values, $(S_m - 2m)\log_B 2$ is equally likely to be any of $0, \frac{1}{n}, \dots, \frac{n-1}{n}$. We considered $100,000$ seeds congruent to 1 modulo 6, starting at $419,753,999,998,525$. We can rapidly analyze the behavior of such large numbers by representing each number as an array and then performing the required operations (multiplication by 3, addition by 1 and division by 2) digit by digit. Taking $m = 10$, we analyzed the first digits for $B = 4, 8$ and 16. We have (theoretical predictions in parentheses)

First Digit	1	2	3	4	5, 6, 7
Base 4	50.2% (50.0%)	49.8% (50.0%)	0%	0%	N/A
Base 8	33.1% (33.3%)	33.6% (33.3%)	0%	33.3% (33.3%)	all 0%

In base 16 we only observe digits 1, 2, 4 and 8; all should occur 25% of the time; we observe them with frequencies 25.0%, 25.0%, 25.3% and 24.8%. In base 10, we observe

First Digit	1	2	3	4	5	6	7	8	9
Observed	29.8%	17.9%	12.1%	10.0%	8.5%	9.8%	2.4%	8.7%	0.9%
Benford	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%

The difficulty in performing these experiments is that our theory is that of two limits, $T \to \infty$ and then $m \to \infty$. We want to choose large seeds x_0 (at least large enough so that after m applications of the $3x+1$ map we haven't hit 1); however, that requires us to examine (at least on a log scale) a large number of x_0. Taking larger starting values (say of the order 10^{100}) makes it impractical to study enough consecutive seeds. In these cases, to approximate the limit as $T \to \infty$ it is best to choose $100,000$ seeds from a variety of starting values and average.

While we cannot yet prove that the iterates of a generic fixed seed are Benford, we expect this to be so. The table below records the percent of first digits equal to j base 10 for a 100,000 random digit number under the $3x+1$ map (as the $3x+1$ map involves simple digit operations, we may represent numbers as arrays, and the computations are quite fast). We performed two experiments: in the first we removed the highest power of 2 in each iteration ($799,992$ iterates), while in the second we had $M(x) = 3x+1$ for x odd and $\frac{x}{2}$ for x even ($2,402,282$ iterates). In both, the observed probabilities are extremely close to the Benford predictions (for each digit, the corresponding z-statistics range from about -2 to 2).

First Digit	Benford Probability	Removing 2	z-statistic	Not Removing 2	z-statistic
1	0.3010	0.3021	2.00	0.3012	0.63
2	0.1761	0.1752	-2.10	0.1763	0.98
3	0.1249	0.1242	-1.97	0.1248	-0.69
4	0.0969	0.0967	-0.50	0.0967	-1.14
5	0.0792	0.0792	0.03	0.0792	-0.06
6	0.0670	0.0671	0.56	0.0667	-1.32
7	0.0580	0.0582	0.68	0.0581	0.89
8	0.0512	0.0513	0.79	0.0510	-0.77
9	0.0458	0.0460	0.99	0.0459	1.02

We calculated the χ^2 values for both experiments: it is 12.38 in the first ($M(x) = \frac{3x+1}{2^k}$) and 6.60 in the second ($M(x) = 3x+1$ for x odd and $\frac{x}{2}$ otherwise). As for 8 degrees of freedom, $\alpha = .05$ corresponds to a χ^2 value of 15.51, and $\alpha = .01$ corresponds to 20.09, we do not reject the null hypothesis and our experiments support the claim that the iterates of both maps obey Benford's law.

6. Conclusion and Future Work

The idea of using Poisson Summation to show certain systems are Benford is not new (see for example [Pin] or page 63 of [Fe]); the difficulty is in bounding the error terms. Our purpose here is to codify a certain natural set of conditions where the Poisson Summation can be executed, and show that interesting systems do satisfy these conditions; a natural future project is to determine additional systems that can be so analyzed. One of the original goals of the project was to prove that the first digits of the terms x_m in the $3x+1$ Problem are Benford. While the techniques of this paper are close to handling this, the structure theorem at our disposal makes $\frac{x_m}{\left(\frac{3}{4}\right)^m x_0}$ the natural quantity to investigate (although numerical investigations strongly support the claim that for any generic seed, the iterates of the $3x+1$ map are Benford); however, we have not fully exploited the structure theorem and the geometric Brownian motion, and hope to return to analyzing the first digit of x_m at a later time. Since the submission of this paper Lagarias and Soundarajan [LS] have shown (with the $3x+1$ map defined by $T(x) = 3x+1$ if x is odd and $\frac{x}{2}$ if x is even)

Theorem 6.1 (Lagarias-Soundararajan)**.** *Let $B \geq 2$ be a fixed integer base. For each $N \geq 1$ and each $X \geq 2^N$, for all but at most $c(B)N^{-1/36}X$ initial seeds (where $c(B)$ is a positive constant depending only on B) the distribution of the first N iterates of the $3x+1$ map are within $2N^{-1/36}$ of the Benford probabilities.*

Similarly, additional analysis of the error terms in the expansions and integrations of L-functions may lead to proving Benford behavior on the critical line, and not just near it, although our results on values of L-functions near the critical line as well as the digits of values of characteristic polynomials of random matrix ensembles support the conjectured Benford behavior.

Acknowledgements

We thank Arno Berger, Keith Conrad, Ted Hill, Ioannis Karatzas, Jeff Lagarias, James Mailhot, Jeff Miller, Michael Rosen, Yakov Sinai and Kannan Soundararajan for many enlightening conversations, Dean Eiger and Stewart Minteer for running some of the $3x+1$ calculations, Klaus Schuerger for pointing out some typos in an earlier draft, and the referee for many valuable comments (especially suggesting we study characteristic polynomials of unitary matrices). Both authors would also like to thank the 2003 Hawaii International Conference on Statistics and The Ohio State University for their hospitality, where much of this work was written up.

Appendix A. Values of Characteristic Polynomials

Consider the random matrix ensemble of $N \times N$ unitary matrices U (with eigenvalues $e^{\mathrm{i}\theta_n}$) with respect to Haar measure; the probability density of U is

$$p_N(U) \ = \ \frac{1}{(2\pi)^N N!} \prod_{1 \leq j < m \leq N} \left| e^{\mathrm{i}\theta_j} - e^{\mathrm{i}\theta_m} \right|. \tag{A.1}$$

Let

$$Z(U,\theta) \;=\; \det(I - Ue^{-\mathrm{i}\theta}) \;=\; \prod_{n=1}^{N}\left(1 - e^{\mathrm{i}(\theta_n-\theta)}\right) \tag{A.2}$$

be the characteristic polynomial of U. The values of characteristic polynomials have been shown to be a good model for the values of L-functions. Of interest to us are the results in [KeSn], where an analogue of the log-normal law of L-functions (Theorem 4.3) is shown for random matrix ensembles: as $N \to \infty$ the average of the absolute value of the characteristic polynomials of unitary matrices is Gaussian. Specifically, let $\rho_N(x)$ be the probability density for $\log|Z(U,\theta)|$ averaged with respect to Haar measure (Equation (36) of [KeSn]), and set

$$\widetilde{\rho}_N(x) \;=\; \sqrt{Q_2(N)}\;\rho_N(\sqrt{Q_2(N)}\;x). \tag{A.3}$$

Here $Q_2(N)$ is the variance, and by Equation (11) of [KeSn] satisfies

$$Q_2(N) \;=\; \frac{\log N}{2} + \frac{\gamma+1}{2} + \frac{1}{24N^2} + O(N^{-4}). \tag{A.4}$$

Equation (53) of [KeSn] (and the comment immediately after it) yield

Theorem A.1 (Keating-Snaith)**.** *With $\widetilde{\rho}_N$ as above,*

$$\widetilde{\rho}_N(x)dx \;=\; \frac{1}{\sqrt{2\pi}}\;e^{-x^2/2}dx + O\left((\log N)^{-3/2}dx\right). \tag{A.5}$$

In terms of ρ_N, from (A.3) we immediately deduce that

$$\rho_N(x)dx \;=\; \frac{1}{\sqrt{2\pi Q_2(N)}}\;e^{-x^2/2Q_2(N)}dx + O\left(Q_2(N)^{-2}dx\right); \tag{A.6}$$

note the pointwise errors are of size one over the square of the variance. It is easy to show the conditions of Theorem 3.2 are satisfied. These errors are significantly smaller than the number theory analogues, in part due to the additional averaging (the formulas here are for averages with respect to Haar measure, whereas in number theory we studied one specific L-function). We thus have

Theorem A.2. *As $N \to \infty$, the distribution of digits of the absolute values of the characteristic polynomials of $N \times N$ unitary matrices (with respect to Haar measure) converges to the Benford probabilities.*

Proof. As the main term is given by a Gaussian, the only difficulty is in verifying Conditions 1 and 3. In our current setting, $\sqrt{Q_2(N)}$ is playing the role of T. Let $h(N) = \log Q_2(N)$. As

$$\int_{-\sqrt{Q_2(N)}h(N)}^{\sqrt{Q_2(N)}h(N)} \frac{1}{\sqrt{2\pi Q_2(N)}}\;e^{-x^2/2Q_2(N)}dx \;=\; 1 + o(1), \tag{A.7}$$

Condition 1 is satisfied. For Condition 3, note $E_T(b+k)-E_t(a+k)$ becomes $O\left(Q_2(N)^{-2}\right)$, and thus

$$\sum_{|k|\le\sqrt{Q_2(N)}h(N)} [E_N(b+k)-E_N(a+k)] \ll \sqrt{Q_2(N)}h(N)Q_2(N)^{-2} \ll \frac{\log Q_2(N)}{Q_2(N)^{3/2}}. \quad \text{(A.8)}$$

□

Remark A.3. *While we believe the distribution of digits of L-functions on the critical line is Benford, our results (Theorem 4.4 and Corollary 4.5) apply to values just off the critical line. Theorem A.2 may thus be interpreted as providing additional support to the conjectured Benford behavior of L-functions on the critical line.*

Remark A.4. *In our earlier investigations of Benford behavior, we used either the counting measure (first N terms of a sequence) or Lebesgue measure (values of the function at arguments $t \in [0,N]$), with $N \to \infty$. We have an extra averaging here. We are not looking at the characteristic polynomials of a sequence of unitary matrices U_N (where U_N is $N \times N$). Instead for each N we use Haar measure on $N \times N$ unitary matrices to average the values of the characteristic polynomials, and then send $N \to \infty$. The averaged characteristic polynomials play an analogous role to our L-functions from before.*

Appendix B. Irrationality type of $\log_B 2$ and Equidistribution

Theorem B.1. *Let B be a positive integer not of the form 2^n for an integer n. Then $\log_B 2$ is of finite type.*

Proof. By (1.1), we must show for some finite $\kappa > 0$ that

$$\left|\log_B 2 - \frac{p}{q}\right| \gg \frac{1}{q^\kappa}. \quad \text{(B.1)}$$

As

$$\left|\frac{\log 2}{\log B} - \frac{p}{q}\right| = \frac{|q\log 2 - p\log B|}{|q|\log B}, \quad \text{(B.2)}$$

it suffices to show $|q\log 2 - p\log B| \gg q^{-\kappa'}$. This follows immediately from Theorem 2 of [Ba], which implies that if α_j and β_j are algebraic integers of heights at most $A_j(\ge 4)$ and $B(\ge 4)$, then if $\Lambda = \beta_1\log\alpha_1+\cdots+\beta_n\log\alpha_n \neq 0$, $|\Lambda| > B^{-C\Omega\log\Omega'}$, where d is the degree of the extension of $\mathbb{Q}$ generated by the α_j and β_j, $C = (16nd)^{200n}$, $\Omega = \log\alpha_1\cdots\log\alpha_n$ and $\Omega' = \Omega/\log\alpha_n$. We take B to be maximum of $\beta_1 = q$ and $\beta_2 = -p$. (As stated we need $\alpha_1,\alpha_2 \ge 4$; we replace $q\log 2 - p\log B$ with $\frac{1}{2}(q\log 4 - p\log B^2)$). In our case

$d = 1, n = 2, \alpha_1 = 4, \alpha_2 = B^2$. As B is not a power of 2, $q \log 4 - p \log B^2 \neq 0$ unless $p, q = 0$. In particular,

$$\left| \log_B 2 - \frac{p}{q} \right| \gg \frac{1}{q^{1+C\Omega \log \Omega'}}. \tag{B.3}$$

For $B = 10$ we may take $\kappa = 2.3942 \times 10^{602}$ (though almost surely a lower number would suffice). □

We show the connection between the irrationality type of α and equidistribution of $n\alpha$ mod 1; see Theorem 3.3 on page 124 of [KN] for complete details. Define the discrepancy of a sequence x_n $(n \leq N)$ by

$$D_N = \frac{1}{N} \sup_{[a,b] \subset [0,1]} |N(b-a) - \#\{n \leq N : x_n \bmod 1 \in [a,b]\}|. \tag{B.4}$$

The Erdös-Turan Theorem (see [KN], page 112) states that there exists a C such that for all m,

$$D_N \leq C \left(\frac{1}{m} + \sum_{h=1}^{m} \frac{1}{h} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} \right| \right). \tag{B.5}$$

If $x_n = n\alpha$, then the sum on n above is bounded by $\min\left(N, \frac{1}{|\sin \pi h\alpha|}\right) \leq \min\left(N, \frac{1}{2||h\alpha||}\right)$, where $||x||$ is the distance from x to the nearest integer. If α is of finite type, this leads to $\sum_{h=1}^{m} \frac{1}{h||h\alpha||}$. For α of type κ, this sum is of size $m^{\kappa-1+\epsilon}$, and the claimed equidistribution rate follows from taking $m = \lfloor N^{1/\kappa} \rfloor$.

Appendix C. (d, g, h)-Maps

The Benford behavior of $3x + 1$ also occurs in (d, g, h)-Maps, defined as follows. Consider positive coprime integers d and g, with $g > d \geq 2$, and a periodic function $h(x)$ satisfying:

(1) $h(x + d) = h(x)$,
(2) $x + h(x) \equiv 0 \bmod d$,
(3) $0 < |h(x)| < g$.

The map M is defined by the formula

$$M(x) = \frac{gx + h(gx)}{d^k}, \tag{C.1}$$

where k is uniquely chosen so that the result is not divisible by d. Property (2) of h guarantees $k \geq 1$. The natural domain of this map is the set Π of positive integers not divisible by d and g. Let E be the set of integers between 1 and dg that divide neither d nor g, so we can write $\Pi = dg\mathbb{Z}^+ + E$. The size of E can easily be calculated: $|E| = (d-1)(g-1)$. In the same way as before, we have m-paths, which are the values of k that appear in iterations of M, and we again denote them by $\gamma_m(x)$.

The $3x+1$ Problem corresponds to $g=3$, $d=2$, and $h(1)=1$, the $3x-1$ Problem corresponds to $g=3$, $d=2$, and $h(1)=-1$, the $5x+1$ Problem corresponds to $g=5$, $d=2$, and $h(1)=1$, and so on. Similar to Theorem 5.2, one can show

Theorem C.1 ([KonSi])**.** *The (d,g,h)-Paths are those of a geometric Brownian motion with drift $\log g - \frac{d}{d-1}\log d$.*

We expect paths to decay for negative drift and escape to infinity for positive drift. All results on Benford's Law for the $(3x+1)$-Problem, in particular Theorem 5.3, generalize trivially to all (d,g,h)-Maps, with the (irrationality) type of $\log_B d$ the generalization of the (irrationality) type of $\log_B 2$; note Theorem B.1 is easily modified to analyze $\log_B d$.

References

[BPY] P. Baine, J. Pitman and M. Yor, *Probability Laws Related to the Jacobi Theta and Riemann Zeta Functions, and Brownian Excursions*, Bulletin of the AMS, **Vol 38**, Number 4, June 2001, 435-465.

[Ba] A. Baker, *The Theory of Linear Forms in Logarithms*, Transcendence Theory: Advances and Applications (editors A. Baker and D. W. Masser), Academic Press, 1977.

[Ben] F. Benford, *The law of anomalous numbers*, Proceedings of the American Philosophical Society **78** (1938), 551-572.

[BBH] A. Berger, Leonid A. Bunimovich and T. Hill, *One-dimensional dynamical systems and Benford's Law*, Trans. Amer. Math. Soc. **357** (2005), no. 1, 197-219.

[Bi] P. Billingsley, *Prime numbers and Brownian motion*, Amer. Math. Monthly **80** (1973), 1099-1115.

[BJ] H. Bohr and B. Jessen, *On the distribution of the values of the Riemann zeta-function*, Amer. J. Math. **58** (1936), 35-44.

[Con] J. B. Conrey, *The Riemann Hypothesis*, Notices of the AMS, March 2003, 341-353.

[Da] H. Davenport, *Multiplicative Number Theory, 2nd edition*, Graduate Texts in Mathematics **74**, Springer-Verlag, New York, 1980, revised by H. Montgomery.

[Dia] P. Diaconis, *The distribution of leading digits and uniform distribution mod 1*, Ann. Probab. **5** (1979), 72-81.

[Fe] W. Feller, *An Introduction to Probability Theory and its Applications, Vol. II*, second edition, John Wiley & Sons, Inc., 1971.

[Hej] D. Hejhal, *On a result of Selberg concerning zeros of linear combinations of L-functions*, International Math. Res. Notices **11** (2000), 551-557.

[Hi1] T. Hill, *The first-digit phenomenon*, American Scientists **86** (1996), 358-363.

[Hi2] T. Hill, *A statistical derivation of the significant-digit law*, Statistical Science **10** (1996), 354-363.

[HS] M. Hindry and J. Silverman, *Diophantine geometry: An introduction*, Graduate Texts in Mathematics **201**, Springer, New York, 2000.

[IK] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, AMS, Providence, RI, 2004.

[Iv] A. Ivić, *On small values of the Riemann Zeta-Function on the critical line and gaps between zeros*, Lietuvos Matematikos Rinkinys **42** (2002), 31-45.

[JW] B. Jessen and A. Winter, *Distribution functions and the Riemann zeta function*, Transactions of the AMS **38** (1935), 48-88.

[KaSh] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics **113**, Springer 1991.

[KaSa] N. Katz and P. Sarnak, *Zeros of zeta functions and symmetries*, Bull. AMS **36** (1999), 1-26.

[KeSn] J. P. Keating and N. C. Snaith, *Random matrix theory and* $\zeta(1/2 + it)$, Comm. Math. Phys. **214** (2000), no. 1, 57-89.

[KonSi] A. Kontorovich and Ya. G. Sinai, *Structure theorem for* (d, g, h)*-maps*, Bull. Braz. Math. Soc. (N.S.) **33** (2002), no. 2, 213-224.

[Knu] D. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, Addison-Wesley, third edition, 1997.

[KN] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, John Wiley & Sons, 1974.

[Lag1] J. Lagarias, *The* $3x + 1$ *Problem and its Generalizations*, Organic mathematics (Burnaby, BC, 1995), 305-334, CMS Conf. Proc., 20, Amer. Math. Soc., Providence, RI, 1997.

[Lag2] J. Lagarias, *The 3x+1 problem: An annotated bibliography*, http://arxiv.org/abs/math/0309224.

[LS] J. Lagarias and K. Soundararajan, *Benford's Law for the* $3x + 1$ *Function*, preprint.

[Lau] A. Laurinčikas, *Limit Theorems for the Riemann zeta-function on the critical line II*, Lietuvos Mat. Rinkinys **27** (1987), 489-500.

[Luo] W. Luo, *Zeros of Hecke L-functions associated with cusp forms*, Acta Arith. **71** (1995), no. 2, 139-158.

[New] S. Newcomb, *Note on the frequency of use of the different digits in natural numbers*, Amer. J. Math. **4** (1881), 39-40.

[Pin] R. Pinkham, *On the Distribution of First Significant Digits*, The Annals of Mathematical Statistics **32**, no. 4 (1961), 1223-1230.

[Ro] K. Roth, *Rational approximations to algebraic numbers*, Mathematika **2** (1955), 1-20.

[Sel1] A. Selberg, *Contributions to the theory of the Riemann zeta-function*, Arch. Math. Naturvid. **48** (1946), no. 5, 89-155.

[Sel2] A. Selberg, *Old and new conjectures and results about a class of Dirichlet series*, in Proceedngs of the Amalfi Conference on Analytic Number Theory, (eds. E. Bombierie et al.), Università di Salerno, Salerno 1992, 367-385.

[Se] J. P. Serre, *A Course in Arithmetic*, Springer-Verlag, 1996.

[Sin] Ya. G. Sinai, *Statistical* $(3x + 1)$ *problem*, Comm. Pure Appl. Math. **56** (2003), no. 7, 1016-1028.

E-mail address: alexk@math.columbia.edu

Department of Mathematics, Columbia, New York, NY, 10027

E-mail address: sjmiller@math.ohio-state.edu

Department of Mathematics, The Ohio State University, Columbus, OH 43210

Current address: Department of Mathematics, Brown University, Providence, RI 02912

E-mail address: `sjmiller@math.brown.edu`