Some Thoughts on Benford's Law

Steven J. Miller*

November 11, 2004

Abstract

For many systems, there is a bias in the distribution of the first digits. For example, if one looks at the first digit of 2^n in base 10, as n ranges over the positive integers, one observes 1 about 30% of the time (and not $\frac{1}{9} \approx .11\%$ of the time as one might expect). This bias is known as Benford's Law, and occurs in a variety of phenomena. In fact, the IRS uses Benford's Law to check the tax returns of large corporations!

We will show that if $y_n = \log_b x_n$ is equidistributed mod 1, then x_n is Benford base b. This is sufficient to prove that Recurrence Relations (with distinct roots $\lambda_1, \dots, \lambda_k$ such that $|\lambda_1| \geq \cdots \geq |\lambda_k|$ and $|\lambda_1| \neq 1$) are Benford. In particular, this will imply that the Fibonacci numbers, which satisfy the Recurrence Relation $a_n = a_{n-1} + a_{n-2}$, are Benford.

In these notes we develop most of the techniques needed to prove these results. The only fact which we must assume is that if $\alpha \notin \mathbb{Q}$, then $n\alpha \bmod 1$ is equidistributed.

The first section introduces Benford's Law, and highlights the method of proof. The second section investigates Recurrence Relations. The final section is drawn from *An Invitation to Modern Number Theory*, by Steven J. Miller and Ramin Takloo-Bighash, and connects Benford's Law to the $3x + 1$ problem (as well as providing some numerical investigations and explanation of statistical inference). The material in this section assumes prior knowledge of probability theory.

For a nice introduction to Benford's Law, the reader should see [Hi1]; for an exposition on Benford's Law and Recurrence Relations, see [BrDu, NS].

Contents

*E-mail: sjmiller@math.brown.edu

1 Benford's Law

While looking through tables of logarithms in the late 1800s, Newcomb noticed a surprising fact: certain pages were significantly more worn out than others. People were looking up numbers whose logarithm started with 1 significantly more frequently than other digits. In 1938, Benford observed the same digit bias in a variety of phenomenon. See [Hi1] for a description and history, [Hi2, BBH, KonMi] for recent results, and [Knu] for connections between Benford's law and rounding errors in computer calculations.

We say a sequence of positive numbers $\{x_n\}$ is **Benford (base b)** if the probability of observing the first digit of x_n (in base b) is j is $\log_b\left(1 + \frac{1}{j}\right)$.

More precisely, we would have

$$\lim_{N\to\infty} \frac{\#\{n \le N : \text{first digit of } x_n \text{ is } j\}}{N} \;=\; \log_b\left(1 + \frac{1}{j}\right). \tag{1}$$

Note that $j \in \{1, \dots, b-1\}$. This is a division of probability, as one of the $b - 1$ events must occur, and the total probability is

$$\sum_{j=1}^{b-1} \log_b\left(1+\frac{1}{j}\right) = \log_b \prod_{j=1}^{b-1}\left(1+\frac{1}{j}\right) = \log_b \prod_{j=1}^{b-1} \frac{j+1}{j} = 1 = \log_b b = 1. \tag{2}$$

Note it is possible to be Benford to some bases but not others. As $\log_{10} 2 \approx .3$, this means that about 30% of the time the first digit is a 1. This is a very strong digit bias; if all digits (1 through 9) were equally likely, than the probability of the first digit being 1 would be $\frac{1}{9} \approx .11$.

A common way to prove a sequence is Benford is to show its logarithms (modulo 1) are equidistributed. Recall

Definition 1.1 (Equidistributed)**.** *A sequence* $\{y_n\}_{n=1}^{\infty}$, $y_n \in [0,1]$, *is equidistributed in* $[0,1]$ *if*

$$\lim_{N\to\infty} \frac{\#\{n : |n| \leq N, y_n \in [a,b]\}}{2N+1} = \lim_{N\to\infty} \frac{\sum_{n=-N}^{N} \chi_{(a,b)}(y_n)}{2N+1} = b-a \tag{3}$$

for all $(a,b) \subset [0,1]$.

The following theorem will be central to our presentation, and will be proved in §1.2:

Theorem 1.2. *If* $y_n = \log_b x_n$ *equidistributed mod* b*, then* x_n *is Benford (base* b*).*

1.1 Preliminaries

We need the following simple fact:

Lemma 1.3. *If* $u \equiv v$ *mod* 1*, then the first digits of* b^u *and* b^v *are the same in base* b*.*

Proof. (of Lemma 1.3): As $u \equiv v \bmod 1$, without loss of generality we may write $v = u + m$, $m \in \mathbb{Z}$. If

$$b^u = u_k b^k + u_{k-1} b^{k-1} + \cdots + u_0, \tag{4}$$

then

$$\begin{aligned} b^v &= b^{u+m} \\ &= b^u \cdot b^m \\ &= (u_k b^k + u_{k-1} b^{k-1} + \cdots + u_0) b^m \\ &= u_k b^{k+m} + \cdots + u_0 b^m. \end{aligned} \tag{5}$$

Thus, the first digits of each are u_0, proving the claim. □

The utility of the above lemma is that it in order to study the first digit of b^y (in base b), it suffices to study $y \bmod 1$.

1.2 Equidistribution and Benford

Proof (of Theorem 1.2): Assume $y_n = \log_b x_n$ is equidistributed mod 1. Consider the unit interval $[0,1)$. For $j \in \{1,\ldots,b\}$, define p_j by

$$b^{p_j} \;=\; j; \tag{6}$$

equivalently, we have

$$p_j \;=\; \log_b j. \tag{7}$$

For $j \in \{1,\ldots,b-1\}$, let

$$I_j \;=\; [p_j, p_{j+1}) \;\subset [0,1). \tag{8}$$

Claim 1.4. *If y mod $1 \in I_j$, then b^y has first digit j.*

The proof is immediate. By Lemma 1.3, it is sufficient to prove this for $y \in I_j$, which we now assume. Then

$$y \;\in\; [p_j, p_{j+1}) \;\;\text{implies that}\;\; b^{p_j} \;\le y \;< b^{p_{j+1}}. \tag{9}$$

From the definitions of the p_j, it follows that

$$j \;\le\; y \;<\; j+1, \tag{10}$$

proving the claim.

Thus, the measure of the subset of $[0,1)$ which, when we exponentiate by b has first digit j, is simply the length of I_j. This is

$$|I_j| \;=\; p_{j+1} - p_j \;=\; \log_b \frac{j+1}{j} \;=\; \log_b\left(1+\frac{1}{j}\right), \tag{11}$$

the Benford (base b) probabilities.

Returning to the proof of Theorem 1.2, we see that the intervals I_j have length $\log_b\left(1+\frac{1}{j}\right)$. As y_n is equidistributed mod 1, in the limit the percent of time $y_n \in I_j$ is equal to $|I_j|$, ie, $\log b_b\left(1+\frac{1}{j}\right)$.

Now $x_n = b^{y_n}$. Each y_n is equivalent to some $\widetilde{y_n}$ mod 1, and by Lemma 1.3, b^{y_n} and $b^{\widetilde{y_n}}$ have the same first digit.

Thus, in the limit, the probability that the first digit of x_n is j (base b) is just $\log_b\left(1+\frac{1}{j}\right)$, proving the theorem. □

2 Recurrence Relations and Benford's Law

2.1 Recurrence Preliminaries

We consider Recurrence Relations of the following form:

$$a_n \;=\; c_1 a_{n-1} + \cdots + c_k a_{n-k}, \tag{12}$$

where $c_1, \dots, c_k, k$ are fixed integers. It is well known that we may explicitly write a_n in Binet form:

$$a_n \;=\; u_1\lambda_1^n + \cdots + u_n\lambda_k^n, \tag{13}$$

where we have ordered the eigenvalues such that $|\lambda_1| \geq \cdots \geq |\lambda_k|$.

We give a quick sketch in a special case when $k = 2$; the reader can generalize the arguments. Assume $a_n = c_1 a_{n-1} + c_2 a_{n-2}$. Let us guess that $a_n = r^n$ for some r. If this were true, then

$$r^n \;=\; c_1 r^{n-1} + c_2 r^{n-2}. \tag{14}$$

After a little algebra, this leads us to the equation

$$r^2 - c_1 r - c_2 \;=\; 0. \tag{15}$$

There are two solutions to that, say r_1 and r_2. A little algebra shows that any solution a_n is of the form

$$a_n \;=\; u_1 r_1^n + u_2 r_2^n, \tag{16}$$

for some $u_1, u_2 \in \mathbb{C}$. If we are given initial conditions (say the values of a_0 and a_1), we can then solve for α_1, α_2; if the two roots are the same.

Remark 2.1. *We call the equation $r^2 - c_1 r - c_2$ the characteristic polynomial. Technically, we need to assume its roots are distinct; if there are repeated roots, the solution must be modified. Below, we always assume we have Recurrence Relations where the roots are distinct.*

For example, for the Fibonacci numbers $k = 2$, $c_1 = c_2 = 1$, $u_1 = -u_2 = \frac{1}{\sqrt{5}}$, and $\lambda_1 = \frac{1+\sqrt{5}}{2}$, $\lambda_2 = \frac{1-\sqrt{5}}{2}$.

If $|\lambda_1| = 1$, we do not expect the first digit of a_n to be Benford (base b). For example, if we consider

$$a_n \;=\; 2a_{n-1} - a_{n-2} \tag{17}$$

with initial values $a_0 = a_1 = 1$, every $a_n = 1$! If we instead take $a_0 = 0$, $a_1 = 1$, we get $a_n = n$.

2.2 Geometric Series are Benford

Let $\{x\} = x - [x]$ denote the fractional part of x, where $[x]$ as always is the greatest integer at most x. Recall the following:

Theorem 2.2. *Let $\alpha \notin \mathbb{Q}$. Then the fractional parts of $n\alpha$ are equidistributed mod* 1.

For a proof, see [HW].

From this and Theorem 1.2, it immediately follows that Geometric Series (series where $x_n = r^n$) are Benford (modulo a certain irrationality condition on r):

Theorem 2.3. *Let $x_n = ar^n$, $\log_b r \notin \mathbb{Q}$. Then x_n is Benford (base b).*

Proof: Let $y_n = \log_b x_n = n \log_b r + \log_b a$. As $\log_b r \notin \mathbb{Q}$, the fractional parts of y_n are equidistributed. Exponentiating by b, we obtain that x_n is Benford (base b) by Theorem 1.2.

2.3 Recurrence Relations are Benford

We first introduce some notation, and then show recurrence relations are Benford.

Definition 2.4 (Big-Oh, Little-Oh)**.** *If F and G are two real functions with $G(x) > 0$ for x large, we write*

$$F(x) \ = \ O(G(x)) \tag{18}$$

if there exist $M, x_0 > 0$ such that $|F(x)| \le MG(x)$ for all $x > x_0$. If

$$\lim_{x\to+\infty} \frac{F(x)}{G(x)} \ = \ 0, \tag{19}$$

we write $F(x) = o(G(x))$ and say F is little-oh of G.

An alternative notation for $F(x) = O(g(x))$ is $F(x) \ll G(x)$. If the constant depends on parameters α, β but not on parameters a, b, we sometimes write $F(x) \ll_{\alpha,\beta} G(x)$.

Exercise 2.5. *Prove for any $r, \epsilon > 0$, as $x \to \infty$ we have $x^r = O(e^x)$ and $\log x = O(x^\epsilon)$.*

Theorem 2.6. *Let a_n be a Recurrence Relation as before, with $|\lambda_1| \neq 1$ (note $|\lambda_1|$ is the largest absolute value of the eigenvalues). If $\log_b |\lambda_1| \notin \mathbb{Q}$, then a_n is Benford (base b).*

Proof: for notational simplicity, we assume $\lambda_1 > 0$, $\lambda_1 > |\lambda_2|$, and $u_1 > 0$. We will comment at the end on how to handle the more general case.

As always, let $y_n = \log_b x_n$. By Theorem 1.2, it is sufficient to show y_n is equidistributed mod 1. We have

$$\begin{aligned} x_n &= u_1\lambda_1^n + \cdots + u_n\lambda_k^n \\ x_n &= u_1\lambda_1^n \left[1 + O\left(\frac{ku\lambda_2^n}{\lambda_1^n}\right)\right], \end{aligned} \tag{20}$$

where $u = \max_i |u_i| + 1$ (so $ku > 1$ and the big-Oh constant is 1). Choose a small ϵ and an n_0 such that

1. $|\lambda_2| < \lambda_1^{1-\epsilon}$;
2. for all $n > n_0$, $\frac{(ku)^{\frac{1}{n}}}{\lambda_1^\epsilon} < 1$, and note $\frac{ku}{\lambda_1^{n\epsilon}} = \left(\frac{(ku)^{\frac{1}{n}}}{\lambda_1^\epsilon}\right)^n$.

As $ku > 1$, $(ku)^{\frac{1}{n}}$ is monotonically decreasing to 1. Note $\epsilon > 0$ if $\lambda_1 > 1$ and $\epsilon < 0$ if $\lambda_1 < 1$. Letting

$$\beta \ = \ \frac{(ku)^{\frac{1}{n_0}}}{\lambda_1^\epsilon} \frac{|\lambda_2|}{\lambda_1^{1-\epsilon}} \ < \ 1, \tag{21}$$

we find that the error term above is bounded by β^n for $n > n_0$, which tends to 0. Therefore

$$\begin{aligned} y_n &= \log_b x_n \\ &= \log_b(u_1 \lambda_1^n) + O\left(\log_b(1+\beta^n)\right) \\ &= n \log_b \lambda_1 + \log_b u_1 + O(\beta^n), \end{aligned} \tag{22}$$

where the big-Oh constant is 1 (actually, the constant is slightly greater than 1, but for notational ease we will use 1 below). As $\log_b \lambda_1 \notin \mathbb{Q}$, the fractional parts of $n \log_b \lambda_1$ are equidistributed mod 1. Therefore, so are the shifts obtained by adding the fixed constant $\log_b u_1$.

We need only show that the error term $O(\beta^n)$ is negligible. It is possible for the error term to change the first digit; for example, if we had 999999999999999 (or 10000000000), then if the error term contributes 2 (or -2), we would change the first digit (base 10).

However, for n sufficiently large, the error term will change a vanishingly small number of first digits.

Say $n \log_b \lambda_1 + \log_b u_1$ exponentiates (base b) to first digit j, $j \in \{1, \dots, b-1\}$. This means

$$n \log_b \lambda_1 + \log_b u_1 \ \in \ I_j \ = \ [p_{j-1}, p_j). \tag{23}$$

The error term is at most β^n. Thus, y_n will have exponentiate to a different first digit than $n \log_b \lambda_1 + \log_b u_1$ only if one of the following holds:

1. $n \log_b \lambda_1 + \log_b u_1$ is within β^n of p_j, and adding the error term pushes us to or past p_j;

2. $n \log_b \lambda_1 + \log_b u_1$ is within β^n of p_{j-1}, and adding the error term pushes us before p_{j-1}.

The first set is contained in $[p_j - \beta^n, p_j)$, of length β^n. The second is contained in $[p_{j-1}, p_{j-1} + \beta^n)$, also of length β^n.

Thus, the length of the interval where $n \log_b \lambda_1 + \log_b u_1$ and y_n could exponentiate (base b) to different first digits is of size $2\beta^n$. If we choose N sufficiently large, than for all $n > N$, we can make these lengths arbitrarily small.

Thus, as $n \log_b \lambda_1 + \log_b u_1$ is equidistributed mod 1, we can control the size of the subsets of $[0, 1)$ where $n \log_b \lambda_1 + \log_b u_1$ and y_n disagree. The Benford behavior (base b) of x_n now follows (in the limit, of course).

2.4 Weakening of Recurrence Constraints (Sketch)

We now show that we can weaken most of the Recurrence Relation assumptions, namely

1. $\lambda_1 > 0$,

2. $\lambda_1 > |\lambda_2|$,

3. $u_1 > 0$.

It is possible that $|\lambda_1| = |\lambda_2| = \cdots = |\lambda_i|$. If so (including signs), we can combine these terms to give

$$u_1\lambda_1^n + \cdots + u_i\lambda_i^n \;=\; u_*\lambda_1^n + u_\#(-\lambda_1)^n. \tag{24}$$

Of course, if the different eigenvalues of modulus λ_1 range over more than $\pm\lambda_1$, one replaces the sum above with the obvious generalization.

The proof will proceed similarly if the $\lambda_1, \dots, \lambda_i$ are real-valued (simply split into even and odd powers of n, and $2\log_b \lambda_1 \notin \mathbb{Q}$ (in the odd case, we get an extra translation by a multiple of $\log_b \lambda_1$). Note this shows how to handle the negative sign constraint (for we do not want to take logarithms of negative numbers, hence we break our sequence into two sequences). Similarly, if u_1 (or the net effect from eigenvalues of modulus $|\lambda_1|$) is negative, we consider $-x_n$, and show that satisfies Benford (base b).

3 Applications of Probability: Benford's Law and Hypothesis Testing

The Gauss-Kuzmin Theorem tells us that the probability that the millionth digit of a randomly chosen continued fraction expansion is k is approximately $q_k = \log_2\left(1 + \frac{1}{k(k+2)}\right)$. What if we choose N algebraic numbers, say the cube roots of N consecutive primes: how often do we expect to observe the millionth digit equal to k? If we believe that algebraic numbers (other than rationals and quadratic irrationals) satisfy the Gauss-Kuzmin Theorem, we expect to observe $q_k N$ digits equal to k, and probably fluctuations on the order of $\sqrt{N}$. If we observe M digits equal to k, how confident are we (as a function of M and N, of course) that the digits are distributed according to the Gauss-Kuzmin Theorem? This leads us to the subject of **hypothesis testing**: if we assume some process has probability p of success, and we observe M successes in N trials, does this provide support for or against the hypothesis that the probability of success is p?

We develop some of the theory of hypothesis testing by studying a concrete problem, the distribution of the first digit of certain sequences. In many problems (for example, 2^n base 10), the distribution of the first digit is given by Benford's Law, described below. We first investigate situations where we can easily prove the sequences are Benford, and then discuss how to analyze data in harder cases where the proofs aren't as clear. The error analysis is, of course, the same we would use to investigate whether or not the digits of the continued fraction expansions of algebraic numbers satisfy the Gauss-Kuzmin Theorem. In the process of investigating Benford's Law, we encounter equidistributed sequences, logarithmic probabilities (similar to the Gauss-Kuzmin probabilities), and Poisson Summation, as well as many of the common problems in statistical testing (such as non-independent events and multiple comparisons).

3.1 Benford's Law

While looking through tables of logarithms in the late 1800s, Newcomb noticed a surprising fact: certain pages were significantly more worn out than others. People were looking up numbers whose logarithm started with 1 significantly more frequently than other digits. In 1938, Benford observed the same digit bias in a variety of phenomenon. See [Hi1] for a description and history, [Hi2, BBH, KonMi] for recent results, and [Knu] for connections between Benford's law and rounding errors in computer calculations.

A sequence of positive numbers $\{x_n\}$ is **Benford (base** b**)** if the probability of observing the first digit of x_n in base b is j is $\log_b\left(1+\frac{1}{j}\right)$. More precisely,

$$\lim_{N\to\infty} \frac{\#\{n \le N : \text{first digit of } x_n \text{ in base } b \text{ is } j\}}{N} = \log_b\left(1+\frac{1}{j}\right). \tag{25}$$

Note that $j \in \{1,\dots,b-1\}$. This is a probability distribution as one of the $b-1$ events must occur, and the total probability is

$$\sum_{j=1}^{b-1} \log_b\left(1+\frac{1}{j}\right) \;=\; \log_b \prod_{j=1}^{b-1}\left(1+\frac{1}{j}\right) \;=\; \log_b \prod_{j=1}^{b-1} \frac{j+1}{j} \;=\; \log_b b \;=\; 1. \tag{26}$$

It is possible to be Benford to some bases but not others; we show the first digit of 2^n is Benford base 10, but clearly it is not Benford base 2 as the first digit is always 1. For many processes, we obtain a sequence of points, and the distribution of the first digits are Benford. For example, consider the $3x+1$ **problem**. Let a_0 be any positive integer, and consider the sequence where

$$a_{n+1} \;=\; \begin{cases} 3a_n+1 & \text{if } a_n \text{ is odd} \\ a_n/2 & \text{if } a_n \text{ is even.} \end{cases} \tag{27}$$

For example, if $a_0 = 13$, we have

$$\begin{aligned} 13 \;&\longrightarrow\; 40 \longrightarrow 20 \longrightarrow 10 \longrightarrow 5 \longrightarrow 16 \longrightarrow 8 \longrightarrow 4 \longrightarrow 2 \longrightarrow 1 \\ &\longrightarrow\; 4 \longrightarrow 2 \longrightarrow 1 \longrightarrow 4 \longrightarrow 2 \longrightarrow 1 \cdots \end{aligned} \tag{28}$$

An alternate definition is to remove as many powers of two as possible in one step. Thus,

$$a_{n+1} \;=\; \frac{3a_n+1}{2^k}, \tag{29}$$

where k is the largest power of 2 dividing $3a_n+1$. It is conjectured that for *any* a_0, eventually the sequence becomes $4 \to 2 \to 1 \to 4\cdots$ (or in the alternate definition $1 \to 1 \to 1\cdots$). While this is known for all $a_0 \le 2^{60}$, the problem has resisted numerous attempts at proofs (Kakutani has described the problem as a conspiracy to slow down mathematical research because of all the time spent on it). See [Lag] for an excellent survey of the problem. How do the first digits behave (for a_0 large)? Do numerical simulations support the claim that this process is Benford? Does it matter which definition we use?

3.2 Benford's Law and Equidistributed Sequences

As we can write any positive x as b^u for some u, the following lemma shows that it suffices to investigate $u \bmod 1$:

Lemma 3.1. *The first digits of b^u and b^v are the same in base b if and only if $u \equiv v$ mod 1.*

Proof. We prove one direction as the other is similar. If $u \equiv v \bmod 1$, we may write $v = u + m$, $m \in \mathbb{Z}$. If

$$b^u \;=\; u_k b^k + u_{k-1} b^{k-1} + \cdots + u_0, \tag{30}$$

then

$$\begin{aligned} b^v &= b^{u+m} \\ &= b^u \cdot b^m \\ &= (u_k b^k + u_{k-1} b^{k-1} + \cdots + u_0) b^m \\ &= u_k b^{k+m} + \cdots + u_0 b^m. \end{aligned} \tag{31}$$

Thus the first digits of each are u_k, proving the claim. □

Exercise 3.2. *Prove the other direction of the if and only if.*

Consider the unit interval $[0, 1)$. For $j \in \{1, \dots, b\}$, define p_j by

$$b^{p_j} = j \quad \text{or equivalently} \quad p_j \;=\; \log_b j. \tag{32}$$

For $j \in \{1, \dots, b-1\}$, let

$$I_j^{(b)} \;=\; [p_j, p_{j+1}) \;\subset [0, 1). \tag{33}$$

Lemma 3.3. *The first digit of b^y base b is j if and only if y mod $1 \in I_j^{(b)}$.*

Proof. By Lemma 3.1, we may assume $y \in [0, 1)$. Then $y \in I_j^{(b)} = [p_j, p_{j+1})$ if and only if $b^{p_j} \le y < b^{p_{j+1}}$, which from the definition of p_j is equivalent to $j \le b^y < j + 1$, proving the claim. □

The following theorem shows that the exponentials of equidistributed sequences (see definition 1.1) are Benford.

Theorem 3.4. *If $y_n = \log_b x_n$ is equidistributed mod 1 then x_n is Benford (base b).*

Proof. By Lemma 3.3,

$$\{n \le N : y_n \bmod 1 \in [\log_b j, \log_b (j+1))\} \;=\; \{n \le N : \text{first digit of } x_n \text{ in base } b \text{ is } j\}. \tag{34}$$

Therefore

$$\lim_{N \to \infty} \frac{\#\{n \le N : y_n \bmod 1 \in [\log_b j, \log_b (j+1))\}}{N} \;=\; \lim_{N \to \infty} \frac{\#\{n \le N : \text{first digit of } x_n \text{ in base } b \text{ is } j\}}{N}. \tag{35}$$

If y_n is equidistributed, then the left side of (35) is $\log_b \left(1 + \frac{1}{j}\right)$ which implies x_n is Benford base b. □

Remark 3.5. *One can extend the definition of Benford from statements concerning the distribution of the first digit to the distribution of the first k digits. With such an extension, Theorem 3.4 becomes $y_n = \log_b x_n \bmod 1$ is equidistributed if and only if x_n is Benford base b. For details, see [KonMi].*

Let $\{x\} = x - [x]$ denote the fractional part of x, where $[x]$ as always is the greatest integer at most x. It is known that for $\alpha \notin \mathbb{Q}$ the fractional parts of $n\alpha$ are equidistributed mod 1 (for a proof, see [HW]). From this and Theorem 3.4, it immediately follows that geometric series are Benford (modulo the irrationality condition):

Theorem 3.6. *Let $x_n = ar^n$, $\log_b r \notin \mathbb{Q}$. Then x_n is Benford (base b).*

Proof. Let $y_n = \log_b x_n = n \log_b r + \log_b a$. As $\log_b r \notin \mathbb{Q}$, the fractional parts of y_n are equidistributed. Exponentiating by b, we obtain that x_n is Benford (base b) by Theorem 3.4. □

Theorem 3.6 implies that 2^n is Benford base 10, but not surprisingly that it is not Benford base 2.

Exercise 3.7. *Do the first digits of e^n follow Benford's Law? What about $e^n + e^{-n}$?*

3.3 Recurrence Relations and Benford's Law

We show many recurrence relations are Benford. The interested reader should see [BD, NS] for more on the subject.

Exercise 3.8 (Recurrence Relations)**.** *Let $\alpha_0, \dots, \alpha_{k-1}$ be fixed integers and consider the recurrence relation (of order k)*

$$x_{n+k} \ = \ \alpha_{k-1}x_{n+k-1} + \alpha_{k-2}x_{n+k-2} + \cdots + \alpha_1 x_{n+1} + \alpha_0 x_n. \tag{36}$$

Show once k values of x_m are specified, all values of x are determined. Let

$$f(r) \ = \ r^{k+1} - \alpha_{k-1} r^{k-1} - \cdots - \alpha_0; \tag{37}$$

we call this the characteristic polynomial of the recurrence relation. Show if $f(\rho) = 0$, then $x_n = c\rho^n$ satisfies the recurrence relation for any $c \in \mathbb{C}$. If $f(r)$ has k distinct roots $r_1, \dots, r_k$, show that any solution of the recurrence equation can be represented as

$$x_n \ = \ c_1 r_1^n + \cdots + c_k r_k^n \tag{38}$$

for some $c_i \in \mathbb{C}$. The Initial Value Problem is when k values of x_n are specified; using linear algebra, this determines the values of $c_1, \dots, c_k$. Investigate the cases where the characteristic polynomial has repeated roots. For more on recursive relations, see §3.3 and [GKP], §7.3.

Exercise 3.9. *Solve the Fibonacci recurrence relation $F_{n+2} = F_{n+1} + F_n$, given $F_0 = F_1 = 1$. Show F_n grows exponentially, i.e. F_n is of size r^n for some $r > 1$. What is r? Let $r_n = \frac{F_{n+1}}{F_n}$. Show that the even terms r_{2m} are increasing and the odd terms r_{2m+1} are decreasing. Investigate $\lim_{n\to\infty} r_n$ for the Fibonacci numbers. Show r_n converges to the golden mean, $\frac{1+\sqrt{5}}{2}$.*

Exercise 3.10 (Binet's Formula)**.** *For F_n as in the previous exercise, prove*

$$F_{n-1} \;=\; \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]. \tag{39}$$

This formula should be surprising at first: F_n is an integer, but the expression on the right involves irrational numbers and division by 2. More generally, for which positive integers m is

$$\frac{1}{\sqrt{m}}\left[\left(\frac{1+\sqrt{m}}{2}\right)^n - \left(\frac{1-\sqrt{m}}{2}\right)^n\right] \tag{40}$$

an integer for any positive integer n?

Exercise 3.11. *Let $x = [a_0, \dots, a_n]$ be a simple continued fraction. For $m \geq 1$, show $q_m \geq F_m$; therefore, the q_ms grow exponentially. Find a number $c > 1$ such that for* any *simple continued fraction, $q_n \geq c^n$.*

3.3.1 Recurrence Preliminaries

We consider recurrence relations of length k:

$$a_{n+k} \;=\; c_1 a_{n+k-1} + \cdots + c_k a_n, \tag{41}$$

where $c_1, \dots, c_k$ are fixed real numbers. If the equation

$$r^k - c_1 r^{k-1} - c_2 r^{k-2} - \cdots - c_{k-1} r - c_k \;=\; 0 \tag{42}$$

has k distinct roots $\lambda_1, \dots, \lambda_k$, there exist k numbers $u_1, \dots, u_k$ such that

$$a_n \;=\; u_1 \lambda_1^n + \cdots + u_k \lambda_k^n, \tag{43}$$

where we have ordered the roots such that $|\lambda_1| \geq \cdots \geq |\lambda_k|$.

For the Fibonacci numbers $k = 2$, $c_1 = c_2 = 1$, $u_1 = -u_2 = \frac{1}{\sqrt{5}}$, and $\lambda_1 = \frac{1+\sqrt{5}}{2}$, $\lambda_2 = \frac{1-\sqrt{5}}{2}$ (see exercise 3.10). If $|\lambda_1| = 1$, we do not expect the first digit of a_n to be Benford (base b). For example, if we consider

$$a_n \;=\; 2a_{n-1} - a_{n-2} \tag{44}$$

with initial values $a_0 = a_1 = 1$, every $a_n = 1$! If we instead take $a_0 = 0$, $a_1 = 1$, we get $a_n = n$.

3.3.2 Recurrence Relations are Benford

Theorem 3.12. *Let a_n satisfy a recurrence relation of length k with k distinct real roots. Assume $|\lambda_1| \neq 1$ with $|\lambda_1|$ the largest absolute value of the roots. If $\log_b |\lambda_1| \notin \mathbb{Q}$, then a_n is Benford (base b).*

Proof. For simplicity we assume $\lambda_1 > 0$, $\lambda_1 > |\lambda_2|$, and $u_1 > 0$. Again let $y_n = \log_b x_n$. By Theorem 3.4, it suffices to show y_n is equidistributed mod 1. We have

$$\begin{aligned} x_n &= u_1\lambda_1^n + \cdots + u_n\lambda_k^n \\ x_n &= u_1\lambda_1^n \left[1 + O\left(\frac{ku\lambda_2^n}{\lambda_1^n}\right)\right], \end{aligned} \tag{45}$$

where $u = \max_i |u_i| + 1$ (so $ku > 1$ and the big-Oh constant is 1). As $\lambda_1 > |\lambda_2|$, we "borrow" some of the growth from λ_1^n; this is a very common technique. Choose a small ϵ and an n_0 such that

1. $|\lambda_2| < \lambda_1^{1-\epsilon}$;
2. for all $n > n_0$, $\frac{(ku)^{1/n}}{\lambda_1^\epsilon} < 1$, which then implies $\frac{ku}{\lambda_1^{n\epsilon}} = \left(\frac{(ku)^{1/n}}{\lambda_1^\epsilon}\right)^n$.

As $ku > 1$, $(ku)^{1/n}$ is decreasing to 1 as n tends to infinity. Note $\epsilon > 0$ if $\lambda_1 > 1$ and $\epsilon < 0$ if $\lambda_1 < 1$. Letting

$$\beta = \frac{(ku)^{1/n_0}}{\lambda_1^\epsilon}\frac{|\lambda_2|}{\lambda_1^{1-\epsilon}} < 1, \tag{46}$$

we find that the error term above is bounded by β^n for $n > n_0$, which tends to 0. Therefore

$$\begin{aligned} y_n &= \log_b x_n \\ &= \log_b(u_1\lambda_1^n) + O\left(\log_b(1+\beta^n)\right) \\ &= n\log_b \lambda_1 + \log_b u_1 + O(\beta^n), \end{aligned} \tag{47}$$

where the big-Oh constant is bounded by C say. As $\log_b \lambda_1 \notin \mathbb{Q}$, the fractional parts of $n \log_b \lambda_1$ are equidistributed mod 1, and hence so are the shifts obtained by adding the fixed constant $\log_b u_1$.

We need only show that the error term $O(\beta^n)$ is negligible. It is possible for the error term to change the first digit; for example, if we had 999999 (or 1000000), then if the error term contributes 2 (or -2), we would change the first digit (base 10). However, for n sufficiently large, the error term will change a vanishingly small number of first digits. Say $n\log_b \lambda_1 + \log_b u_1$ exponentiates (base b) to first digit j, $j \in \{1, \dots, b-1\}$. This means

$$n\log_b \lambda_1 + \log_b u_1 \in I_j^{(b)} = [p_{j-1}, p_j). \tag{48}$$

The error term is at most $C\beta^n$ and y_n exponentiates to a different first digit than $n\log_b \lambda_1 + \log_b u_1$ only if one of the following holds:

1. $n\log_b \lambda_1 + \log_b u_1$ is within $C\beta^n$ of p_j, and adding the error term pushes us to or past p_j;
2. $n\log_b \lambda_1 + \log_b u_1$ is within $C\beta^n$ of p_{j-1}, and adding the error term pushes us before p_{j-1}.

The first set is contained in $[p_j - C\beta^n, p_j)$, of length $C\beta^n$. The second is contained in $[p_{j-1}, p_{j-1} + C\beta^n)$, also of length $C\beta^n$. Thus the length of the interval where $n \log_b \lambda_1 + \log_b u_1$ and y_n could exponentiate base b to different first digits is of size $2C\beta^n$. If we choose N sufficiently large, than for all $n > N$, we can make these lengths arbitrarily small. As $n \log_b \lambda_1 + \log_b u_1$ is equidistributed mod 1, we can control the size of the subsets of $[0, 1)$ where $n \log_b \lambda_1 + \log_b u_1$ and y_n disagree. The Benford behavior (base b) of x_n now follows (in the limit, of course). □

Exercise 3.13. *Weaken the conditions of Theorem 3.12 as much as possible. What if several eigenvalues equal λ_1? What does a general solution to* (41) *look like now? What if λ_1 is negative? Can anything be said if there are complex roots?*

3.4 Random Walks and Benford

Consider the following (colorful) problem: A drunk starts off at time zero at a lamppost. Each minute he stumbles with probability p one unit to the right and with probability $q = 1 - p$ one unit to the left. Where do we expect the drunk to be after N tosses? This is known as a **Random Walk**. By the Central Limit Theorem, his distribution after N tosses is well approximated by a Gaussian with mean $1 \cdot p + (-1) \cdot (1 - p) = 2p - 1$ and variance $p(1 - p)N$. For more details on Random Walks, see [Re].

For us, a **Geometric Brownian Motion** is a process such that its logarithm is a Random Walk. We show below that the first digits of Geometric Brownian Motions are Benford. In [KonSi] the $3x + 1$ problem is shown to be an example of Geometric Brownian Motion. For heuristic purposes we use the first definition of the $3x + 1$ map, though the proof is for the alternate definition. We have two operators: T_3 and T_2, with $T_3(x) = 3x + 1$ and $T_2(x) = \frac{x}{2}$. If a_n is odd, $3a_n + 1$ is even, so T_3 must always be followed by T_2. Thus, we have really have two operators T_2 and $T_{3/2}$, with $T_{3/2}(x) = \frac{3x+1}{2}$. If we assume each operator is equally likely, half the time we go from $x \to \frac{3}{2}x + 1$, and half the time to $\frac{1}{2}x$.

If we take logarithms, $\log x$ goes to $\log \frac{3}{2}x = \log x + \log \frac{3}{2}$ half the time and $\log \frac{1}{2}x = \log x + \log \frac{1}{2}$ the other half. Hence on average we send $\log x \to \log x + \frac{1}{2} \log \frac{3}{4}$. As $\log \frac{3}{4} < 0$, on average our sequence is decreasing (which agrees with the conjecture that eventually we reach $4 \to 2 \to 1$). Thus we might expect our sequence to look like $\log x_k = \log x + \frac{k}{2} \log \frac{3}{4}$. As $\log \frac{3}{4} \notin \mathbb{Q}$, its multiples are equidistributed mod 1, and thus when we exponentiate we expect to see Benford behavior. Note, of course, that this is simply a heuristic, suggesting we might see Benford's Law.

While we can consider Random Walks or Brownian Motion with non-zero means, for simplicity below we assume the means are zero. Thus, in the example above, $p = \frac{1}{2}$.

3.4.1 Needed Gaussian Integral

Consider a sequence of Gaussians with mean 0 and variance σ^2, with $\sigma^2 \to \infty$. The following lemma shows that for any $\delta > 0$, as $\sigma \to \infty$ almost all of the probability is in the interval $[-\sigma^{1+\delta}, \sigma^{1+\delta}]$. We will use this lemma to show that it is enough to investigate Gaussians in the range $[-\sigma^{1+\delta}, \sigma^{1+\delta}]$.

Lemma 3.14.

$$\frac{2}{\sqrt{2\pi\sigma^2}}\int_{\sigma^{1+\delta}}^{\infty} e^{-x^2/2\sigma^2}dx \ \ll\ e^{-\sigma^{2\delta}/2}. \tag{49}$$

Proof. Change the variable of integration to $w = \frac{x}{\sigma\sqrt{2}}$. Denoting the above integral by I, we find

$$I \ = \ \frac{2}{\sqrt{2\pi\sigma^2}}\int_{\sigma^\delta/\sqrt{2}}^{\infty} e^{-w^2}\cdot\sigma\sqrt{2}dw \ = \ \frac{2}{\sqrt{\pi}}\int_{\sigma^\delta/\sqrt{2}}^{\infty} e^{-w^2}dw. \tag{50}$$

The integrand is monotonically decreasing. For $w \in \left[\frac{\sigma^\delta}{\sqrt{2}}, \frac{\sigma^\delta}{\sqrt{2}}+1\right]$, the integrand is bounded by substituting in the left endpoint, and the region of integration is of length 1. Thus,

$$\begin{aligned} I &< 1\cdot\frac{2}{\sqrt{\pi}}e^{-\sigma^{2\delta}/2} + \frac{2}{\sqrt{\pi}}\int_{\frac{\sigma^\delta}{\sqrt{2}}+1}^{\infty} e^{-w^2}dw \\ &= \frac{2}{\sqrt{\pi}}e^{-\sigma^{2\delta}/2} + \frac{2}{\sqrt{\pi}}\int_{\frac{\sigma^\delta}{\sqrt{2}}}^{\infty} e^{-(u+1)^2}du \\ &= \frac{2}{\sqrt{\pi}}e^{-\sigma^{2\delta}/2} + \frac{2}{\sqrt{\pi}}\int_{\frac{\sigma^\delta}{\sqrt{2}}}^{\infty} e^{-u^2}e^{-2u}e^{-1}du \\ &< \frac{2}{\sqrt{\pi}}e^{-\sigma^{2\delta}/2} + \frac{2}{e\sqrt{\pi}}e^{-\sigma^{2\delta}/2}\int_{\frac{\sigma^\delta}{\sqrt{2}}}^{\infty} e^{-2u}du \\ &< \frac{2(e+1)}{\sqrt{\pi}}e^{-\sigma^{2\delta}/2} \\ &< 4e^{-\sigma^{2\delta}/2}. \end{aligned} \tag{51}$$

□

3.4.2 Geometric Brownian Motions are Benford

We investigate the distribution of digits of processes that are Geometric Brownian Motions. By Theorem 3.4, it suffices to show that the Geometric Brownian Motion converges to being equidistributed mod 1. Explicitly, we have the following: after N iterations, by the Central Limit Theorem the expected value converges to a Gaussian with mean 0 and variance proportional to $\sqrt{N}$. We must show that the Gaussian with growing variance is equidistributed mod 1.

For convenience we assume the mean is 0 and the variance is $N/2\pi$. This corresponds to a fair coin where for each head (tail) we move $\frac{1}{\sqrt{4\pi}}$ units to the right (left). By the Central Limit Theorem the probability of being x units to the right of the origin after N tosses is asymptotic to

$$p_N(x) \ = \ \frac{e^{-\pi x^2/N}}{\sqrt{N}}. \tag{52}$$

For ease of exposition, we assume that rather than being asymptotic to a Gaussian, the distribution is a Gaussian. For our example of flipping a coin, this cannot be true. If every minute we flip a coin and record the outcome, after N minutes there are 2^N possible outcomes, a finite number. To each of these we attach a a number equal to the excess of heads to tails. There are technical difficulties in working with discrete probability distributions; thus we study instead continuous processes such that at time N the probability of observing x is given by a Gaussian with mean 0 and variance $N/2\pi$. For complete details see [KonMi].

Theorem 3.15. *As* $N \to \infty$, $p_N(x) = \frac{e^{-\pi x^2/N}}{\sqrt{N}}$ *becomes equidistributed mod* 1.

Proof. We want the probability that for $x \in \mathbb{R}$, $x \bmod 1 \in [a,b] \subset [0,1)$. This is

$$\frac{1}{\sqrt{N}} \sum_{n\in\mathbb{Z}} \int_{x=a}^{b} e^{-\pi(x+n)^2/N} dx. \tag{53}$$

We need to show the above converges to $b-a$ as $N \to \infty$. For $x \in [a,b]$, standard calculus (Taylor series expansions) gives

$$e^{-\pi(x+n)^2/N} \ = \ e^{-\pi n^2/N} + O\left(\frac{\max(1,|n|)}{N} e^{-n^2/N}\right). \tag{54}$$

We claim that in (53), it is sufficient to restrict the summation to $|n| \le N^{5/4}$. The proof is immediate from Lemma 3.14: we increase the integration by expanding to $x \in [0,1]$, and then trivially estimate. Thus, up to negligible terms, all the contribution is from $|n| \le N^{5/4}$.

The Poisson Summation formula states that

$$\frac{1}{\sqrt{N}} \sum_{n\in\mathbb{Z}} e^{-\pi n^2/N} \ = \ \sum_{n\in\mathbb{Z}} e^{-\pi n^2 N}. \tag{55}$$

The beauty of this formula is it converts one infinite sum with *slow* decay to another sum with *rapid* decay. The exponential terms on the left of (55) are all of size 1 for $n \le \sqrt{N}$, and don't become small until $n \gg \sqrt{N}$ (for instance, once $n > \sqrt{N}\log N$, the exponential terms are small for large N); however, almost all of the contribution on the right comes from $n = 0$. The power of Poisson Summation is it often allows us to approximate well long sums with short sums. We therefore have

$$\begin{aligned}
\frac{1}{\sqrt{N}} \sum_{|n|\le N^{5/4}} \int_{x=a}^{b} e^{-\pi(x+n)^2/N} dx \ &= \ \frac{1}{\sqrt{N}} \sum_{|n|\le N^{5/4}} \int_{x=a}^{b} \left[e^{-\pi n^2/N} + O\left(\frac{\max(1,|n|)}{N} e^{-n^2/N}\right)\right] \\
&= \ \frac{b-a}{\sqrt{N}} \sum_{|n|\le N^{5/4}} e^{-\pi n^2/N} + O\left(\frac{1}{N} \sum_{n=0}^{N^{5/4}} \frac{n+1}{\sqrt{N}} e^{-\pi(n/\sqrt{N})^2}\right) \\
&= \ \frac{b-a}{\sqrt{N}} \sum_{|n|\le N^{5/4}} e^{-\pi n^2/N} + O\left(\frac{1}{N} \int_{w=0}^{N^{3/4}} (w+1) e^{-\pi w^2} \sqrt{N} dw\right) \\
&= \ \frac{b-a}{\sqrt{N}} \sum_{|n|\le N^{5/4}} e^{-\pi n^2/N} + O\left(N^{-1/2}\right).
\end{aligned} \tag{56}$$

By Lemma 3.14 we can expand all sums to $n \in \mathbb{Z}$ in (56) with negligible error. We now apply Poisson Summation and find that up to lower order terms,

$$\frac{1}{\sqrt{N}} \sum_{n\in\mathbb{Z}} \int_{x=a}^{b} e^{-\pi(x+n)^2/N} dx \;\approx\; (b-a) \cdot \sum_{n\in\mathbb{Z}} e^{-\pi n^2 N}. \tag{57}$$

For $n = 0$ the right hand side of (57) is $b - a$. For all other n, we trivially estimate the sum:

$$\sum_{n\neq 0} e^{-\pi n^2 N} \;\leq\; 2\sum_{n\geq 1} e^{-\pi n N} \;\leq\; \frac{2e^{-\pi N}}{1-e^{-\pi N}}, \tag{58}$$

which is less than $4e^{-\pi N}$ for N sufficiently large. □

We can interpret the above arguments as follows: for each N, consider a Gaussian $p_N(x)$ with mean 0 and variance $N/2\pi$. As $N \to \infty$ for each x (which occurs with probability $p_N(x)$) the first digit of 10^x converges to the Benford base 10 probabilities.

Remark 3.16. *The above framework is very general and applicable to a variety of problems. In [KonMi] it is shown that these arguments can be used to prove Benford behavior in discrete systems such as the $3x+1$ problem as well as continuous systems such as the absolute values of the Riemann Zeta Function (and any "good" L-function) near the critical line! For these number theory results, the crucial ingredients are Selberg's result (near the critical line, $\log|\zeta(s+it)|$ for $t \in [T, 2T]$ converges to a Gaussian with variance tending to infinity in T) and estimates by Hejhal on the rate of convergence.*

3.5 Statistical Inference

Often we have reason to believe that some process occurs with probability p of success and $q = 1 - p$ of failure. For example, consider the $3x + 1$ problem. Choose a large a_0 and look at the first digit of the a_ns. There is reason to believe the distribution of the first digits is given by Benford's Law for most a_0 as $a_0 \to \infty$. We describe how to test this and similar hypotheses. We content ourselves with describing one simple test; the interested reader should consult a statistics textbook (for example, [BD, LF, MoMc]) for the general theory and additional applications.

3.5.1 Null and Alternative Hypotheses

Suppose we think some population has a parameter with a certain value. If the population is small, it is possible to investigate every element; in general this is not possible.

For example, say the parameter is how often the millionth decimal or continued fraction digit is 1 in two populations: all rational numbers in $[0, 1)$ with denominator at most 5, and all real numbers in $[0, 1)$. In the first, there are only 10 numbers, and it is easy to check them all. In the second, as there are infinitely many numbers, it is impossible to numerically investigate each. What we do in practice is we sample a large number of elements (say N elements) in $[0, 1)$, and calculate the average value of the parameter for this sample.

We thus have two **populations**, the **underlying population** (in the second case, all numbers in $[0, 1)$), and the **sample population** (in this case, the N sampled elements).

Our goal is to test whether or not the underlying population's parameter has a given value, say p. To this end, we want to compare the sample population's value to p. The **null hypothesis**, denoted H_0, is the claim that there is no difference between the sample population's value and the underlying population's value; the **alternative hypothesis**, denoted H_a, is the claim that there is a difference between the sample population's value and the underlying population's value.

When we analyze the data from the sample population, either we reject the null hypothesis, or we fail to reject the null hypothesis. It is important to note that we *never* prove the null or alternative hypothesis is true or false. We are always rejecting or failing to reject the null hypothesis, we are never accepting it. If we flip a coin 100 times and observe all heads, this does not mean the coin isn't fair: it is possible the coin is fair but we had a very unusual sample (though, of course, it is extremely unlikely).

We now discuss how to test the null hypothesis. Our main tool is the Central Limit Theorem.

3.5.2 Bernoulli Trials and the Central Limit Theorem

Assume we have some process where we expect a probability p of observing a given value. For example, if we choose numbers uniformly in $[0, 1)$ and look at the millionth decimal digit, we believe that the probability this digit is 1 is $\frac{1}{10}$. If we look at the continued fraction expansion, by the Gauss-Kuzmin Theorem the probability that the millionth digit is 1 is approximately $\log_2 \frac{4}{3}$. What if we restrict to algebraic numbers? What is the probability the millionth digit (decimal or continued fraction expansion) equals 1?

In general, once we formalize our conjecture, we test it by choosing N elements from the population independently at random. Consider the claim that a process has probability p of success. We have N independent Bernoulli trials. The null hypothesis is the claim that p percent of the sample are a success. Let s_N be the number of successes; if the null hypothesis is correct, by the Central Limit Theorem we expect s_N to have a Gaussian distribution with mean pN and variance pqN. This means that if we were to look at many samples with N elements, on average each sample would have $pN \pm O(\sqrt{pqN})$ successes. We calculate the probability of observing a difference $|s_N - pN|$ as large or larger than a. This is given by the area under the Gaussian with mean pN and variance pqN:

$$\frac{1}{\sqrt{2\pi pqN}} \int_{|s-pN| \geq a} e^{-(s-pN)^2/2pqN} ds. \tag{59}$$

If this integral is small, it is extremely unlikely that we choose N independent trials from a process with probability p of success and we reject the null hypothesis; if the integral is large, we do not reject the null hypothesis, and we have support for our claim that the underlying process does have probability p of success.

Unfortunately, the Gaussian is a difficult function to integrate, and we would need to tabulate these integrals for *every* different pair of mean and variance. It is easier, therefore, to renormalize and look at a new statistic which should also be Gaussian, but with mean 0 and variance 1. The advantage is that we need only tabulate *one* special Gaussian, the standard normal.

Let $z = \frac{s_N - pN}{\sqrt{pqN}}$. This is known as the **z-statistic**. If s_N's distribution is a Gaussian with mean pN and variance pqN, note z will be a Gaussian with mean 0 and variance 1.

Exercise 3.17. *Prove the above statement about the distribution of z.*

Let

$$I(a) \ = \ \frac{1}{\sqrt{2\pi}} \int_{|z| \geq a} e^{-z^2/2} dz, \tag{60}$$

the area under the standard normal (mean 0, standard deviation 1) that is at least a units from the mean. We consider different **confidence intervals**. If we were to randomly choose a number z from such a Gaussian, what is the probability (as a function of a) that z is at most a units from the mean? Approximately 68% of the time $|z| \leq 1$ ($I(1) \approx .32$), approximately 95% of the time $z \leq 1.96$ ($I(1.96) \approx .05$), and approximately 99% of the time $|z| \leq 2.57$ ($I(2.57) = .01$). In other words, there is only about a 1% probability of observing $|z| \geq 2.57$. If $|z| \geq 2.57$, we have strong evidence against the hypothesis that the process occurs with probability p, and we would be reasonably confident in rejecting the null hypothesis; of course, it is possible we were unlucky and obtained an unrepresentative set of data (but it is extremely unlikely that this occurred; in fact, the probability is at most 1%).

Remark 3.18. *For a Gaussian with mean μ and standard deviation σ, the probability $|x - \mu| \leq \sigma$ is approximately $.68$, and so on.*

To test the claim that some process occurs with probability p, we observe N independent trials, calculate the z-statistic, and see how likely it is to observe $|z|$ that large or larger. We give two examples below.

3.5.3 Digits of the $3x + 1$ Problem

Consider again the $3x + 1$ problem. Choose a large integer a_0, and look at the iterates: $a_1, a_2, a_3, \dots$. We study how often the first digit of terms in the sequence equal $d \in \{1, \dots, 9\}$. We can regard the first digit of a term as a Bernoulli trial with a success (or 1) if the first digit is d and a failure (or 0) otherwise. If the distribution of digits is governed by Benford's Law, the theoretical prediction is that the percent of the first digits that equal d is $p = \log_{10}(\frac{d+1}{d})$. Assume there are N terms in our sequence (before we hit the pattern $4 \to 2 \to 1 \to 4 \cdots$), and say M of them have first digit d. For what M does this experiment provide support that the digits follow Benford's Law?

Exercise 3.19. *The terms in the sequence generated by a_0 are not independent, as a_{n+1} is determined by a_n. Show that if the first digit of a_n is 2, the first digit of a_{n+1} cannot be a 2.*

The above exercise shows that the first digit of the terms *cannot* be considered independent Bernoulli trials. As the sequence is completely determined by the first term, this is not surprising. If we look at an enormous number of terms, however, these effects "should" average out. Another possible experiment is to look at the first digit of the millionth term for N different a_0s.

Let $a_0 = 333\dots333$ be the integer that is 10,000 threes. There are 177,857 terms in the sequence before we hit $4 \to 2 \to 1$. The following data comparing the number of first digits equal to d to the Benford predictions is from [Min]:

digit	observed	predicted	variance	z-statistic	I(z)
1	53425	53540	193.45	-0.596	0.45
2	31256	31310	160.64	-0.393	0.31
3	22257	22220	139.45	0.257	0.21
4	17294	17230	124.76	0.464	0.36
5	14187	14080	113.88	0.914	0.63
6	11957	11900	105.40	0.475	0.36
7	10267	10310	98.57	-0.480	0.37
8	9117	9090	92.91	0.206	0.16
9	8097	8130	88.12	-0.469	0.36

As the values of the z-statistics are all small (well below 1.96 and 2.57), the above table provides evidence that the first digits in the $3x+1$ problem follow Benford's Law, and we would not reject the null hypothesis for any of the digits. If we had obtained large z-statistics, say 4, we would reject the null hypothesis and doubt that the distribution of digits follow Benford's Law.

Remark 3.20 (Important)**.** *One must be very careful when analyzing all the digits. Once we know how many digits are $1, \dots, 8$, then the number of 9s is forced: these are not 9 independent tests. Our point here is not to write a treatise on statistical inference, but merely highlight some of the tools and concepts.*

Additionally, if we have many different experiments, then "unlikely" events should happen. For example, if we have 100 different experiments we would not be surprised to see an outcome which only has a 1% chance of occurring (see exercise 3.21). Thus, if there are many experiments, the confidence intervals need to be adjusted. One common method is the Bonferroni adjustment method for multiple comparisons. See [BD, MoMc].

Exercise 3.21. *Assume for each trial there is a 95% chance of observing the percent of first digits equal to 1 is in $[\log_{10} 2 - \sigma, \log_{10} 2 + \sigma]$ (for some σ). If we have 10 independent trials, what is the probability that* all *the observed percents are in this interval? If we have 14 independent trials?*

Remark 3.22. *How does one calculate with 10,000 digit numbers? Such large numbers are greater than the standard number classes (int, long, double) of many computer programming languages. The solution is to represent numbers as arrays. To go from a_n to $3a_n + 1$, we multiply the array by 3, carrying as needed, and then add 1; we leave space-holding zeros at the start of the array. For example,*

$$3 \cdot [0, \dots, 0, 0, 5, 6, 7] \ = \ [0, \dots, 0, 1, 7, 0, 1]. \tag{61}$$

We need only do simple operations on the array. For example, $3 \cdot 7 = 21$, so the first entry of the product array is 1 and we carry the 2 for the next multiplication. We must also compute $a_n/2$ if a_n is even. Note this *is the same as $5a_n$ divided by 10. The advantage of this approach is that it is easy to calculate $5a_n$, and as a_n is even, the last digit of $5a_n$ is zero, hence array division by 10 is trivial.*

Exercise 3.23. *Consider the first digits of the $3x + 1$ problem in base 6. Choose a large integer a_0, and look at the iterates $a_1, a_2, a_3, \dots$. Show that as $a_0 \to \infty$, the distribution of digits is Benford base 6.*

Exercise 3.24 (Recommended)**.** *Here is another variant of the $3x + 1$ problem:*

$$a_{n+1} \;=\; \begin{cases} 3a_n + 1 & \text{if } a_n \text{ is odd} \\ a_n/2^k & \text{if } a_n \text{ is even and } 2^k || a_n. \end{cases} \tag{62}$$

By $2^k||a_n$ we mean 2^k divides a_n, but 2^{k+1} does not. Consider the distribution of first digits of this sequence. What is the null hypothesis? Do the data support the null hypothesis, or the alternative hypothesis? Do you think these numbers also satisfy Benford's Law? What if instead we define

$$a_{n+1} \;=\; \frac{3a_n + 1}{2^k}, \quad 2^k||a_n. \tag{63}$$

3.5.4 Digits of Continued Fractions

Let us test the hypothesis that the digits of algebraic numbers are given by the Gauss-Kuzmin Theorem. Let us look at how often the 1000^{th} digit equals 1. By Kuzmin, this should be approximately $\log_2 \frac{4}{3}$. Let p_n be the n^{th} prime. In the continued fraction expansions of $\sqrt[3]{p_n}$ for $n \in \{100000, 199999\}$, exactly 41,565 have 1000^{th} digit equal to 1. Assuming we have a Bernoulli process with probability of success (a digit of 1) of $p = \log_2 \frac{4}{3}$, the z-statistic is .393. As the z-statistic is small (95% of the time we expect to observe $|z| \le 1.96$), we do not reject the null hypothesis, and we have obtained evidence supporting the claim that the probability that the 1000^{th} digit is 1 is given by the Gauss-Kuzmin Theorem.

3.6 Summary

We have chosen to motivate our presentation of statistical inference by investigating the first digits of the $3x+1$ problem, but of course the methods apply to a variety of problems. Our main tool is the Central Limit Theorem: if we have a process with probability p ($q = 1 - p$) of success (failure), then in N independent trials we expect about pN successes, with fluctuations of size $\sqrt{pqN}$. To test whether or not the underlying probability is p we formed the z-statistic: $\frac{\#\text{S}-pN}{\sqrt{pqN}}$, where #S is the number of successes observed in the N trials.

If the process really does have probability p of success, then by the Central Limit Theorem the distribution of #S is Gaussian with mean pN and standard deviation $\sqrt{pqN}$, and we then expect the z-statistic to be of size 1. If, however, the underlying process occurs not with probability p but p', then we expect #S to be a Gaussian with mean $p'N$ and standard deviation $\sqrt{p'q'N}$. We now expect the z-statistic to be of size $\frac{(p'-p)N}{\sqrt{p'q'N}}$. This is of size $\sqrt{N}$, much larger than 1.

We see the z-statistic is very sensitive to $p' - p$: if p' is differs from p, for large N we quickly observe large values of z. Note, of course, that statistical tests can only provide compelling evidence in favor or against a hypothesis, never a proof.

A Probability Review

A.1 Bernoulli Distribution

Recall the binomial coefficient $\binom{N}{r} = \frac{N!}{r!(N-r)!}$ is the number of ways to choose r objects from N objects when order does not matter. Consider n independent repetitions of a process with only two possible outcomes. We typically call one outcome **success** and the other **failure**, the event a **Bernoulli trial**, and a collection of independent Bernoulli trials a **Bernoulli process**. In each Bernoulli trial, let there be probability p of success and $q = 1 - p$ of failure. Often, we represent a success with 1 and a failure with 0.

Exercise A.1. *Consider a Bernoulli trial with random variable x equal to 1 for a success and 0 for a failure. Show $\overline{x} = p$, $\sigma_x^2 = pq$, and $\sigma_x = \sqrt{pq}$. Note x is also an indicator random variable.*

Let y_N be the number of successes in N trials. Clearly the possible values of y_N are $\{0, 1, \dots, N\}$. We analyze $p_N(k) = \text{Prob}(y_N(\omega) = k)$. Here the sample space Ω is all possible sequences of N trials, and the random variable $y_N : \Omega \to \mathbb{R}$ is given by $y_N(\omega)$ equals the number of successes in ω.

If $k \in \{0, 1, \dots, N\}$, we need k successes and $N - k$ failures. We don't care what order we have them (i.e., if $k = 4$ and $N = 6$ then $SSFSSF$ and $FSSSSF$ both contribute equally). Each such string of k successes and $N - k$ failures has probability of $p^k \cdot (1-p)^{N-k}$. There are $\binom{N}{k}$ such strings, which implies $p_N(k) = \binom{N}{k} p^k \cdot (1-p)^{N-k}$ if $k \in \{0, 1, \dots, N\}$ and 0 otherwise.

By clever algebraic manipulations, one can directly evaluate the mean $\overline{y_N}$ and the variance $\sigma_{y_N}^2$; however, standard lemmas allow one to calculate both quantities immediately, once one knows the mean and variance for a single occurrence (see Exercise A.1).

Lemma A.2. *For a Bernoulli process with N trials, each having probability p of success, the expected number of successes is $\overline{y_N} = Np$, and the variance is $\sigma_{y_N}^2 = Npq$.*

Lemma A.2 states the expected number of successes is of size Np, and the fluctuations about Np are of size $\sigma_{y_N}^2 = \sqrt{Npq}$. Thus, if $p = \frac{1}{2}$ and $N = 10^6$, we expect 500,000 successes, with fluctuations on the order of 500. Note how much smaller the fluctuations about the mean are than the mean itself (the mean is of size N, the fluctuations of size $\sqrt{N}$). This is an example of a general phenomenon, which we describe in greater detail in §A.3.

Exercise A.3. *Prove Lemma A.2. Prove the variance is largest when $p = q = \frac{1}{2}$.*

Consider the following problem: Let $\Omega = \{S, FS, FFS, \dots\}$ and let z be the number of trials before the first success. What is $\overline{z}$ and σ_z^2?

First we determine the **Bernoulli distribution** $p(k) = \text{Prob}(z(\omega) = k)$, the probability that the first success occurs after k trials. Clearly this probability is non-zero only for k a positive integer, in which case the string of results must be $k-1$ failures followed by 1 success. Therefore

$$p(k) \;=\; \begin{cases} (1-p)^{k-1}\cdot p & \text{if } k \in \{1,2,\dots\} \\ 0 & \text{otherwise} \end{cases} \tag{64}$$

To determine the mean $\overline{z}$ we must evaluate

$$\overline{z} \;=\; \sum_{k=1}^{\infty} k(1-p)^{k-1}p \;=\; p\sum_{k=1}^{\infty} kq^{k-1}, \quad 0 < q = 1-p < 1. \tag{65}$$

Consider the geometric series

$$f(q) \;=\; \sum_{k=0}^{\infty} q^k \;=\; \frac{1}{1-q}. \tag{66}$$

A careful analysis shows we can differentiate term by term if $-1 \le q < 1$. Then

$$f'(q) \;=\; \sum_{k=0}^{\infty} kq^{k-1} \;=\; \frac{1}{(1-q)^2}. \tag{67}$$

Recalling $q = 1-p$ and substituting yields

$$\overline{z} \;=\; p\sum_{k=1}^{\infty} kq^{k-1} \;=\; \frac{p}{(1-(1-p))^2} \;=\; \frac{1}{p}. \tag{68}$$

Remark A.4. *Differentiating under the summation sign is a powerful tool in Probability Theory, and is a common technique for proving such identities.*

Exercise A.5. *Calculate σ_z^2.* Hint: *differentiate $f(q)$ twice.*

A.2 Random Sampling

We introduce the notion of **random sampling**. Consider a countable set $\Omega \subset \mathbb{R}$ and a probability function p on Ω; we can extend p to all of $\mathbb{R}$ by setting $p(r) = 0$ if $r \notin \Omega$. Using the probability function p, we can choose elements from $\mathbb{R}$ **at random**. Explicitly, the probability that we choose $\omega \in \Omega$ is $p(\omega)$.

For example, let $\Omega = \{1,2,3,4,5,6\}$ with each event having probability $\frac{1}{6}$ (the rolls of a fair die). If we were to roll a fair die N times (for N large), we observe a particular sequence of outcomes. It is natural to assume the rolls are independent of each other. Let x_i denote the outcome of the i^{th} roll. The x_is all have the same distribution (arising from p). We call the x_i **i.i.d.r.v.** (independent identically distributed random variables), and we say the x_i are a **sample** from the probability distribution p. We say we **randomly sample (with respect to p)** $\mathbb{R}$. Often we simply say we have **randomly chosen N numbers**.

A common problem is to sample some mathematical or physical process, and use the observations to make inferences about the underlying system. For example, we may be given a coin without being told what its probabilities for heads and tails are. We can attempt to infer the probability p of a head by tossing the coin many times, and recoding the outcomes. Let x_i be the outcome of the i^{th} toss (1 for head, 0 for tail). After N tosses we expect to see about Np heads; however, we observe some number, say s_N. Given that we observe s_N heads after N tosses, what is our best guess for p? We guess $p = \frac{s_N}{N}$. It is extremely unlikely that our guess is exactly right. This leads us to a related question: given that we observe s_N heads, can we give a small interval about our best guess where we are extremely confident the true value p lies? The solution is given by the Central Limit Theorem (§A.3).

Exercise A.6. *For the above example, if p is irrational, show the best guess can never be correct.*

One can generalize the above to include the important case where p is a continuous distribution. For example, say we wish to investigate the digits of numbers in $[0,1]$. It is natural to put the uniform distribution on this interval, and choose numbers at random relative to this distribution; we say we choose N numbers randomly with respect to the uniform distribution on $[0,1]$, or simply we choose N numbers uniformly from $[0,1]$. Two natural problems are to consider the n^{th} digit in the base 10 expansion and the n^{th} digit in the continued fraction expansion. By observing many choices, we hope to infer knowledge about how these digits are distributed. The first problem is theoretically straightforward. It is not hard to calculate the probability that the n^{th} digit is d; it is just $\frac{1}{10}$. The probabilities of the digits of continued fractions are significantly harder (unlike decimal expansions, *any* positive integer can occur as a digit).

Exercise A.7 (Important for Computational Investigations)**.** *For any continuous distribution, the probability we chose a number in $[a,b]$ is $\int_a^b p(x)dx$. If we were to choose N numbers, N large, then we expect approximately $N\int_a^b p(x)dx$ to be in $[a,b]$. Often computers have built in random number generators for certain continuous distributions, such as the standard Gaussian or the uniform, but not for less common ones. Show if one can randomly choose numbers from the uniform distribution, one can use this to randomly choose from any distribution.* Hint: *use $C_p(x) = \int_{-\infty}^x p(x)dx$, the **Cumulative Distribution Function** of p; it is the probability of observing a number at most x.*

Remark A.8. *The observant reader may notice a problem with sampling from a continuous distribution: the probability of choosing any particular real number is zero, but some number is chosen! One explanation is that, fundamentally, we cannot choose numbers from a continuous probability distribution. For example, if we use computers to choose our numbers, all computers can do is a finite number of manipulations of 0s and 1s; thus, they can only choose numbers from a countable (actually finite) set. The other interpretation of the probability of any $r \in \mathbb{R}$ is zero is that, while at each stage some number is chosen, no number is ever chosen twice. Thus, in some sense, any number we explicitly write down is "special".*

For our investigations, we approximate continuous distributions by discrete distributions with many outcomes. From a practical point of view, this suffices for many experiments; however, one should note that while theoretically we can write statements such as "choose a real number uniformly from $[0,1]$", we can never actually do this.

A.3 The Central Limit Theorem

We close our introduction to probability with a statement of *the* main theorem about the behavior of a sum of independent events. We give a proof in an important special case in §A.3.2.

A.3.1 Statement of the Central Limit Theorem

Let x_i ($i \in \{1, \dots, N\}$) be i.i.d.r.v. as in §A.2, all sampled from the same probability distribution p with mean $\mathbb{E}[p] = \mu$ and variance $\sigma_p^2 = \sigma^2$ (so $\mathbb{E}[x_i] = \mu$ and $\sigma_{x_i}^2 = \sigma^2$ for all i). Let $s_N = \sum_{i=1}^N x_i$. We are interested in the distribution of s_N as $N \to \infty$. As each x_i has expected value $\mathbb{E}[x] = \mu$, $\mathbb{E}[s_N] = N\mu$. We now consider a more refined question: how is s_N distributed about $N\mu$? The Central Limit Theorem answers this, and tells us what the correct scale is to study the fluctuations about μ.

Theorem A.9 (Central Limit Theorem)**.** *As $N \to \infty$,*

$$\textit{Prob}(s_N \in [\alpha, \beta]) \ \sim \ \frac{1}{\sqrt{2\pi\sigma^2 N}} \int_\alpha^\beta e^{-(t-\mu N)^2/2\sigma^2 N} dt. \tag{69}$$

In other words, the distribution of s_N converges to a Gaussian with mean μN and variance $\sigma^2 N$. We may re-write this as

$$\textit{Prob}\left(\frac{s_N - \mu N}{\sqrt{\sigma^2 N}} \in [a, b]\right) \ \longrightarrow \ \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt. \tag{70}$$

Here $z_n = \frac{s_N - \mu N}{\sqrt{\sigma^2 N}}$ converges to a Gaussian with mean 0 and variance 1.

The probability density $e^{-t^2/2}/\sqrt{2\pi}$ is the standard Gaussian. It is *the* universal curve of probability. Note how robust the Central Limit Theorem is: it does not depend on fine properties of the x_j, just that they all have the same distributions. Sometimes it is important to know how rapidly z_N is converging to the Gaussian; see [Fe].

Exercise A.10. *The Central Limit Theorem gives us the correct scale to study fluctuations. For example, say we toss a fair coin N times (hence $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{4}$). We expect s_N to be about $\frac{N}{2}$. Find values of a and b such that the probability of $s_N - N\mu \in [a\sqrt{N}/2, b\sqrt{N}/2]$ converges to 95% (99%). For large N, show for any fixed $\delta > 0$ that the probability of $s_N - N\mu \in [aN^{\frac{1}{2}+\delta}/2, bN^{\frac{1}{2}+\delta}/2]$ tends to zero. Thus, we expect to observe half of the tosses as heads, and we expect deviations from one-half to be of size $2/\sqrt{N}$.*

One common application of the Central Limit Theorem is to test whether or not we are sampling the x_i independently from a fixed probability distribution with mean μ. Choose N numbers randomly from what we expect has mean μ. We form s_N as before and investigate $\frac{s_N - \mu}{\sqrt{N}}$. As $s_N = \sum_{i=1}^N x_i$, we expect s_N to be of size N. If the x_i are not drawn from a distribution with mean μ, then $s_N - N\mu$ will also be of size N. Thus, $\frac{s_N - N\mu}{\sqrt{N}}$ will be of size $\sqrt{N}$ if the x_i are not drawn from something with mean μ. If, however, the x_i are from sampling a distribution with mean μ, the Central Limit Theorem states that $\frac{s_N - N\mu}{\sqrt{N}}$ will be of size 1. See Chapter 3 for more details.

Finally, we note that the Central Limit Theorem is an example of the **Philosophy of Square-root Cancellation**: the sum is of size N, but the deviations are of size $\sqrt{N}$.

A.3.2 Proof for Bernoulli Processes

We sketch the proof of the Central Limit Theorem for Bernoulli Processes where the probability of success is $p = \frac{1}{2}$. Consider the random variable x that is 1 with probability $\frac{1}{2}$ and -1 with probability $\frac{1}{2}$ (for example, tosses of a fair coin; the advantage of making a tail -1 is that the mean is zero). Note the mean of x is $\overline{x} = 0$, the standard deviation is $\sigma_x = \frac{1}{2}$ and the variance is $\sigma_x^2 = \frac{1}{4}$.

Let $x_1, \dots, x_{2N}$ be independent identically distributed random variables, distributed as x (it simplifies the expressions to consider an even number of tosses). Consider $s_{2N} = x_1 + \cdots + x_{2N}$. Its mean is zero and its variance is $2N/4$, and we expect fluctuations of size $\sqrt{2N/4}$. We show that for N large, the distribution of s_{2N} is approximately normal. We need

Lemma A.11 (Stirling's Formula)**.** *For n large,*

$$n! \;=\; n^n e^{-n}\sqrt{2\pi n}\,(1 + O(1/n))\,. \tag{71}$$

For a proof, see [WW]. We show the above is a reasonable approximation.

$$\log n! \;=\; \sum_{k=1}^{n} \log k \;\approx\; \int_1^n \log t dt \;=\; (t\log t - t)|_1^n. \tag{72}$$

Thus, $\log n! \approx n \log n - n$, or $n! \approx n^n e^{-n}$.

We now consider the distribution of s_{2N}. The probability that $s_{2N} = 2k$ is just $\binom{2N}{N+k}(\frac{1}{2})^{N+k}(\frac{1}{2})^{N-k}$. This is because for $s_{2N} = 2k$, we need $2k$ more 1s (heads) than -1s (tails), and the number of heads and tails adds to $2N$. Thus we have $N+k$ heads and $N-k$ tails. There are 2^{2N} strings of 1s and -1s, $\binom{2N}{N+k}$ have exactly $N+k$ heads and $N-k$ tails, and the probability of each string is $(\frac{1}{2})^{2N}$. We have written $(\frac{1}{2})^{N+k}(\frac{1}{2})^{N-k}$ to show how to handle the more general case when there is a probability p of heads and $1-p$ of tails.

We use Stirling's Formula to approximate $\binom{2N}{N+k}$. After elementary algebra we find

$$\begin{aligned} \binom{2N}{N+k} \quad &\approx \quad \frac{2^{2N}}{(N+k)^{N+k}(N-k)^{N-k}}\sqrt{\frac{N}{\pi(N+k)(N-k)}} \\ &= \quad \frac{2^{2N}}{\sqrt{\pi N}}\frac{1}{(1+\frac{k}{N})^{N+\frac{1}{2}+k}(1-\frac{k}{N})^{N+\frac{1}{2}-k}}. \end{aligned} \tag{73}$$

Using $(1+w)^N \approx e^w$, after some more algebra we find

$$\binom{2N}{N+k} \quad \approx \quad \frac{2^{2N}}{\sqrt{\pi N}} e^{2k^2/N}. \tag{74}$$

Thus

$$\text{Prob}(s_{2N} = 2k) \;=\; \binom{2N}{N+k}\frac{1}{2^{2N}} \quad\approx\quad \frac{1}{\sqrt{2\pi\cdot(2N/4)}}\, e^{-k^2/2(2N/4)}. \tag{75}$$

The distribution of s_{2N} looks like a Gaussian with mean 0 and variance $2N/4$. While we can only observe an integer number of heads, for N enormous the Gaussian is very slowly varying and hence approximately constant from $2k$ to $2k+2$.

Exercise A.12. *Generalize the above arguments to handle the case when $p \neq \frac{1}{2}$.*

References

[BBH] A. Berger, Leonid A. Bunimovich and T. Hill, *One-dimensional dynamical systems and Benford's Law*, accepted for publication in Transactions of the American Mathematical Society.

[BD] P. Bickel and K. Doksum, *Mathematical statistics: basic ideas and selected topics*, Holden-Day, 1977.

[BrDu] J. Brown and R. Duncan, *Modulo one uniform distribution of the sequence of logarithms of certain recursive sequences*, Fibonacci Quarterly **8**, 1970, 482-486.

[Du] R. Durrett, *Probability: Theory and Examples*, Duxbury Press, second edition, 1996.

[Fe] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II. Second edition. John Wiley & Sons, Inc., New York-London-Sydney 1971.

[GKP] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A foundation for computer science*, Addison-Wesley Publishing Company, 1988.

[HW] G. H. Hardy and E. Wright, *An Introduction to the Theory of Numbers*, fifth edition, Oxford Science Publications, Clarendon Press, Oxford, 1995.

[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.

[Kel] D. Kelley, *Introduction to Probability*, Macmillian Publishing Company, 1994.

[Knu] D. Knuth, *The Art of Computer Programming, Vol. 2*, Addison-Wesley, second edition, 1981.

[KonMi] A. Kontorovich and S. J. Miller, *Poisson Summation, Benford's Law and values of L-functions*, preprint.

[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.

[Lag] J. Lagarias, *The $3x + 1$ Problem and its Generalizations*,

[LF] R. Larson and B. Farber, *Elementary Statistics: Picturing the World*, Prentice Hall, Inc., 2003.

[Min] S. Minteer, *Analysis of Benford's Law Applied to $3x + 1$ Problem*, Number Theory Working Group, The Ohio State University, 2004.

[MoMc] D. Moore and G. McCabe, *Introduction to the practice of statistics*, W. H. Freeman and Co., 2003.

[NS] K. Nagasaka and J. S. Shiue, *Benford's law for linear recurrence sequences*, Tsukuba J. Math. **11**, 1987, 341-351.

[NZM] I. Niven, H. Zuckerman, and H. Montgomery, *An Introduction to the Theory of Numbers*, John Wiley & Sons, Inc., fifth edition, 1991.

[Re] F. Reif, *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill.

[WW] E. Whittaker and G. Watson, *A Course of Modern Analysis*, Cambridge University Press; 4th edition, 1996.

[We] E. Weisstein, *MathWorld–A Wolfram Web Resource*, http://mathworld.wolfram.com/