Benford's Law and Stick Decomposition

by
Joy Jing

Steven J. Miller, Advisor

A thesis submitted in partial fulfillment
of the requirements for the
Degree of Bachelor of Arts with Honors
in Mathematics & Statistics

Williams College
Williamstown, Massachusetts

May 15, 2013

Abstract

Many datasets and real-life functions exhibit a leading digit bias, where the first digit base 10 of a number equals 1 not 11% of the time as we would expect if all digits were equally likely, but closer to 30% of the time. This phenomenon is known as Benford's Law, and has applications ranging from the detection of tax fraud to analyzing the Fibonacci sequence. It is especially applicable in today's world of 'Big Data' and can be used for fraud detection to test data integrity, as most people are unaware of the phenomenon.

The cardinal goal is often determining which datasets follow Benford's Law. We know that the decomposition of a finite stick based on a reiterative cutting pattern determined by a 'nice' probability density function will tend toward Benford's Law. We extend the results of [1] to show that this is also true when the cuts are determined by a finite set of nice probability density functions. We further conjecture that when we apply the same exact cut at every level, as long as that cut is not equal to 0.5, the lengths of the subsegments will converge to a Benford distribution.

Acknowledgements

First and foremost, I want to thank my advisor, Steven Miller, for his constant encouragement and guidance, and his unwavering faith in my ability to plunge head-first into a new field. I also want to express my gratitude to the entire mathematics and statistics department at Williams College, for their incredible tutelage both in and out of the classroom, and for making me the mathematician I am today.

I thank my friends, classmates, and professors, who are the quintessence of Williams and make it the perfect place to pursue any and all interests, and who have made the past four years extraordinary.

Finally, I thank my family for their constant love and support; especially Ing-Miin Hsu who has indelibly shaped who I am today.

Contents

1. Introduction

Imagine a random dataset of numbers. Maybe it's the results from a scientific experiment. Maybe it's the number of likes on Facebook or stock prices on the New York Stock Exchange. Or maybe it's the set of numbers (with repeats) that show up in an issue of Reader's Digest. For each number in the dataset, let us consider just the first digit. For example, if the number is 427,598 we think of it as a '4' and if it is 932 then we think of it as a '9'. Knowing that we have a random, arbitrary data set, how often would we expect this leading digit to be a '1'?

Most respondents to this question guess within the range of $10 - 11\%$, as people often expect a random dataset to be uniformly distributed, and so have an equally likely chance of beginning with a 1, 2, 3, 4, 5, 6, 7, 8, or 9. Hence, they expect a 1 to be the leading digit $\frac{1}{9}$ of the time (or $\frac{1}{10}$ when they forget that 0 cannot be a leading digit). In fact, this is not the case. In many datasets, 1 is the leading digit approximately 30.1% of the time whereas 9 is the leading digit only 4.6% of the time.

This observation is often called **Benford's Law** or more specifically the **First-Digit Phenomenon** [10]. Benford's Law generalizes to include the distribution of all the digits in a given number, but for our purposes we will only focus on the first digit bias. Benford's Law says that the leading significant digits are not uniformly distributed as one might expect, but follow a logarithmic distribution skewed toward the smaller digits. The astronomer-mathematician Simon Newcomb (1835-1909) first noted this pattern when he noticed the increased wear on the first few pages of logarithmic tables [12]. Newcomb reached the following conclusion:

> *The law of probability of the occurrence of numbers is such that all mantissæ of their logarithms are equally probable.* [12]

The phenomenon was popularized by Frank Benford, after whom the law is named. Along with proffering explanations for why digits follow this distribution, he also presented justifications for the significance of studying such a problem. In referring to Newcomb's observation that the first few pages of a logarithm table are more worn, he wrote:

> *Recall that the table is used in the building up of our scientific, engineering, and general factual literature. There may be, in the relative cleanliness of the pages of a logarithm table, data on how we think and how we react when dealing with things that can be described by means of numbers.* [2]

One of Benford's particularly important observations is that while individual datasets may fail to satisfy Benford's law, amalgamating many different sets of data leads to a new sequence whose behavior is typically closer to Benford. Benford's Law is applicable to numerous situations: it is observed in natural systems such as hydrology data [15] and stock prices [9]. It is used in computer science [3, 4, 7] and in accounting to detect tax fraud [13, 14]. A detailed bibliography of the field can be found at [5].

In this paper, we are motivated by a common question in physics:

> *What is the most probable way a conserved quantity can be partitioned into pieces subject to one or more other constraints?* [8]

We will pursue this question of decomposition by investigating a particular cutting algorithm in which we begin with a stick of length L and at every level the stick or its subsequent subsegments will be split by a cut K. We show that if each cut it determined by a finite set of continuous probability density functions that satisfy a 'nice' condition (equation 3.1) on its Mellin transforms, then as we apply $N \to \infty$ levels of cuts, the distribution of the lengths of the subsegments will converge toward a Benford distribution.

We further conjecture that if we apply the same exact cut at every level, as long as that cut is not equal to 0.5, then the distribution will converge to Benford. We discuss the intuition behind this conjecture and provide distribution calculations to several thousand levels.

2. Notation and Terminology

To give a more rigorous definition of Benford's Law, we must first introduce some important notation and terminology.

2.1. **Benford's Law Terminology.**

Definition 2.1. *Any positive number x may be written in* ***scientific notation*** *as $S(x) \cdot 10^k$ where $S(x) \in [1, 10)$ is called the* ***significand*** *and k is an integer (called the* ***exponent****).*

Definition 2.2. *The integer portion of the significand is called the* ***leading digit*** *or* ***first digit****.*

Definition 2.3. *The* ***mantissa***[1] *refers to the fractional part of a logarithm.*

The leading digit will be our primary focus. Since k merely designates the location of the decimal point, for the most part it will be inconsequential. The following example helps to clarify our notation.

Let us take the number $x = 5225020.034$. This can be written in scientific notation as $5.225020034 \cdot 10^6$ which means that the significand is $S(x) = 5.225020034$, the leading digit is 5, and the exponent is $k = 6$. If we take the logarithm base 10 of our number, we find that $\log_{10} 5225020.034 \approx 5.71808796$ which means that the mantissa is approximately .71808796. This can also be thought of as taking the logarithm of the number, modulo 1. So $\log_{10}(x) \bmod 1 = \text{mantissa}$. Since $S(x) \in [1, 10)$, we can also directly compute the mantissa by taking the logarithm of the significand: $\log_{10} 5.225020034 \approx .71808796$.

Studying the leading digits of a dataset has several advantages. It allows us to compare data with different scales; one could be the masses of subatomic particles, whereas the other could be home street addresses of all Williams' students. While the units and magnitudes differ greatly, every number has a unique leading digit and thus we can compare them equitably. We are now ready to precisely state Benford's law.

2.2. **Statement of Benford's Law.**

Definition 2.4. ***Benford's Law for the Leading Digit.*** *A set of numbers satisfies Benford's Law for the Leading Digit if the probability of observing a first digit of d is $\log_{10}\left(\frac{d+1}{d}\right)$.*

[1] In some texts, the term *mantissa* is used to refer to the *significand*. In this paper, we will refer to the mantissa only as the fractional part of a logarithm.

When using real data it is often necessary to modify this definition of Benford's Law to include anything that is approximately close to or a good visual fit for Benford. This is because it is impossible for a finite set of data to perfectly equal the irrational value of $\log_{10}\left(\frac{d+1}{d}\right)$, as d is the leading digit an integer number of times, and the total size of the data must be an integer. Since we are working with mathematical functions, however, we will continue to use the stated definition of Benford's Law.

2.3. **Terms used in Proof.**

Definition 2.5. *For $s \in [1, 10)$, let*

$$\varphi_s(x) = \begin{cases} 1 & \text{if the significand of } x \text{ is at most } s \\ 0 & \text{otherwise.} \end{cases}$$

*We call φ_s the **indicator function** of the event that the significand is at most s.*

Definition 2.6. *The **proportion** $P_N(s)$ of partition pieces $x_1, ..., x_{2^N}$ whose significand is less than or equal to s is defined as follows:*

$$P_N(s) = \frac{\sum_{i=1}^{2^N} \varphi_s(x_i)}{2^N}. \tag{2.1}$$

Definition 2.7. *The **probability density function (PDF)**, also known as the **density** of a continuous random variable, is a function that describes the relative likelihood of that random variable taking on a given value. A probability density function is nonnegative everywhere, and its integral over the entire space is equal to one.*

Definition 2.8. *The **cumulative distribution function (CDF)** describes the probability that a real-valued random variable x with a given PDF will be found at a value less than or equal to x. Essentially, if $f(x)$ gives the density of x, then $\int_0^x f(t)dt$ gives the cumulative distribution.*

Definition 2.9. *The **Mellin transform** $(\mathcal{M}f)(s)$ of a function f is defined by*

$$(\mathcal{M}f)(s) = \int_0^\infty x^{s-1} f(x) dx. \tag{2.2}$$

Definition 2.10. *The **inverse Mellin transform** $(\mathcal{M}^{-1}g)(x)$, is given by*

$$(\mathcal{M}^{-1}g)(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} g(s) x^{-s} ds. \tag{2.3}$$

Definition 2.11. *The **Fourier transform** $\hat{f}$ of a function f is defined as*

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx. \tag{2.4}$$

The Fourier and Mellin transforms are related by a logarithmic change in variables. To see this relationship explicitly, let us set $x = e^{2\pi u}$ and $s = \sigma - i\xi$ and rewrite the definition of the Mellin transform as follows:

$$\begin{aligned}(\mathcal{M}f)(\sigma - i\xi) &= \int_{-\infty}^{\infty} e^{2\pi u(\sigma - i\xi)} f(e^{2\pi u}) 2\pi du \\ &= \int_{-\infty}^{\infty} 2\pi f(e^{2\pi u}) e^{2\pi u\sigma} e^{-2\pi u i\xi} du. \end{aligned} \tag{2.5}$$

If we label $g(u) = 2\pi f(e^{2\pi u}) e^{2\pi u\sigma}$, we see that this is merely the Fourier transform $\hat{g}(u)$. Hence we can transition between the Fourier transform of a function and its Mellin transform easily. Using Poisson Summation, we can relate the Fourier series coefficients of the function to its Fourier transform, and subsequently its Mellin transform as well. Note that $(\mathcal{M}f)(s) = \mathbb{E}[x^{s-1}]$, and hence results on Mellin transforms further translate to results about expected values.

In this paper we extend the results of [1] who show that if we repeatedly apply cuts based on a single 'nice' continuous probability density function, then the resulting decomposition will converge to Beford as we apply more and more cuts. We extend this to the case where we have multiple probability density functions, and show that the same result holds for any finite set of 'nice' density functions.

3. Results

We are concerned with the physical deomposition of matter and the most probable way a conserved quantity will be partitioned into pieces. We model this by considering a stick of length L. We apply cuts in proportion to its length, and then analyze the distribution of the lengths of the substicks produced as we apply infinitely many cuts.

At the first level, we apply some cut $K_1 \in (0,1)$, which will produce two new substicks: one of length LK_1 and the other of length $L(1-K_1)$. At the second level, we will then apply two unrelated cuts K_2 and K_3 to the two substicks, where again $K_2, K_3 \in (0,1)$. Note that K_2 and K_3 need not be equal, although it is a possibility. The first substick will now be split into two substicks of length LK_1K_2 and $LK_1(1-K_2)$ whereas the second substick will be split into substicks of length $L(1-K_1)K_3$ and $L(1-K_1)(1-K_3)$. We can continue this process of decomposition, resulting in 2^N substicks of varying lengths at the N^{th} level.

The K_i values that we apply are determined by a set of probability density functions. At any stage, each cut is determined by a single density function, where the choice of function is arbitrary. The value of each cut relies on its corresponding probability density function, in that drawing from the uniform distribution means that any cut between $(0,1)$ is equally likely to be used.

As we progress through $N \to \infty$ levels of applying this cutting pattern, we show that the distribution of the lengths of the subsegments will converge to Benford.

3.1. Statement of Theorem.

Theorem 3.1. *Let $\{f_1, f_2, ..., f_M\}$ be a finite set of continuous probability density functions on $[0,1]$ such that all functions satisfy the following property:*

$$\lim_{N\to\infty} \sum_{\substack{l=-\infty \\ l\neq 0}}^{\infty} \prod_{n=1}^{N} (\mathcal{M} f_{m_n}) \left(1 - \frac{2\pi i \ell}{\log 10}\right) = 0. \tag{3.1}$$

Given a stick of length L, apply cuts based on the probability density functions $\{f_1, f_2, ..., f_M\}$ where the choice of which PDF to use is random. As we apply $N \to \infty$ cuts, the distribution of the lengths of the remaining subsegments will converge to Benford's Law.

3.2. Outline of Proof.

To simplify notation, we first prove the case when we have only two probability density functions f_1 and f_2, and then generalize our conclusion to apply to any finite set of functions. This means that in Equation (3.1), $f_{m_n}(x)$ represents $f_1(x)$,

$f_1(1-x)$, $f_2(x)$, or $f_2(1-x)$. Since either f_1 or f_2 will be chosen as the density of K the complementary section $1-K$ will have density $f_1(1-x)$ or $f_2(1-x)$.

Given a stick of length L, choose independent identically distributed random variables $K_{1;a}, K_{2;a}, ..., K_{(2^N-1);a}$ with density f_1 and i.i.d. random variables $K_{1;b}, K_{2;b}, ..., K_{(2^N-1);b}$ with density f_2. Divide the stick of length L as follows:

(1) Divide L into $LK_{1;a}$ and $L(1-K_{1;a})$ or divide L into $LK_{1;b}$ and $L(1-K_{1;b})$. Regardless of whether the K_1 is drawn from a density f_1 or f_2, let us label the new lengths $LK_{1;r_1}$ and $L(1-K_{1;r_1})$ where $r_1 \in \{a, b\}$.

(2) Divide $LK_{1;r_1}$ into $LK_{1;r_1}K_{2;a}$ and $LK_{1;r_1}(1-K_{2;a})$, or $LK_{1;r_1}K_{2;b}$ and $LK_{1;r_1}(1-K_{2;b})$. Again, let us call these $LK_{1;r_1}K_{2;r_2}$ and $LK_{1;r_1}(1-K_{2;r_2})$ where $r_2 \in \{a, b\}$.

(3) Divide $L(1-K_{1;r_a})$ into $L(1-K_{1;r_1})K_{3;r_3}$ and $L(1-K_{1;r_1})(1-K_{3;r_3})$ where $r_3 \in \{a, b\}$.

Continue cutting each piece into two in this fashion, by pulling an arbitrary cut from either density function f_1 or f_2, where the probability of applying each cut is determined by the value of the density function. After N iterations we obtain the following set of cuts:

$$\begin{aligned}
x_1 &= LK_{1;r_1}K_{2;r_2}K_{4;r_4}\cdots K_{2^{N-2};r_{(2^N-2)}}K_{2^{N-1};r_{(2^N-1)}} \\
x_2 &= LK_{1;r_1}K_{2;r_2}K_{4;r_4}\cdots K_{2^{N-2};r_{(2^N-2)}}\left(1-K_{2^{N-1};r_{(2^N-1)}}\right) \\
&\vdots \\
x_{2^N-1} &= L\left(1-K_{1;r_1}\right)\left(1-K_{3;r_3}\right)\left(1-K_{7;r_7}\right)\cdots\left(1-K_{2^{N-1};r_{(2^N-2)}}\right)K_{2^N-1;r_{(2^N-1)}} \\
x_{2^N} &= L\left(1-K_{1;r_1}\right)\left(1-K_{3;r_3}\right)\left(1-K_{7;r_7}\right)\cdots\left(1-K_{2^{N-1};r_{(2^N-2)}}\right)\left(1-K_{2^N-1;r_{(2^N-1)}}\right).
\end{aligned}$$

A visualization of this splitting pattern can be seen in Figure 3.1.

The satisfaction of the following two properties guarantees convergence to a Benford distribution [1], so we want to show that they are true for our cutting method:

$$\lim_{N\to\infty} \mathbb{E}[P_N(s)] = \log_{10} s \tag{3.2}$$

$$\lim_{N\to\infty} \mathrm{Var}(P_N(s)) = 0. \tag{3.3}$$

Part one (equation (3.2)) says that as N approaches infinity, the expected value of the amalgamation of the partition pieces will approach Benford. Part two (equation (3.3)) says that as N approaches infinity, the variance around the expected value will tend toward zero,

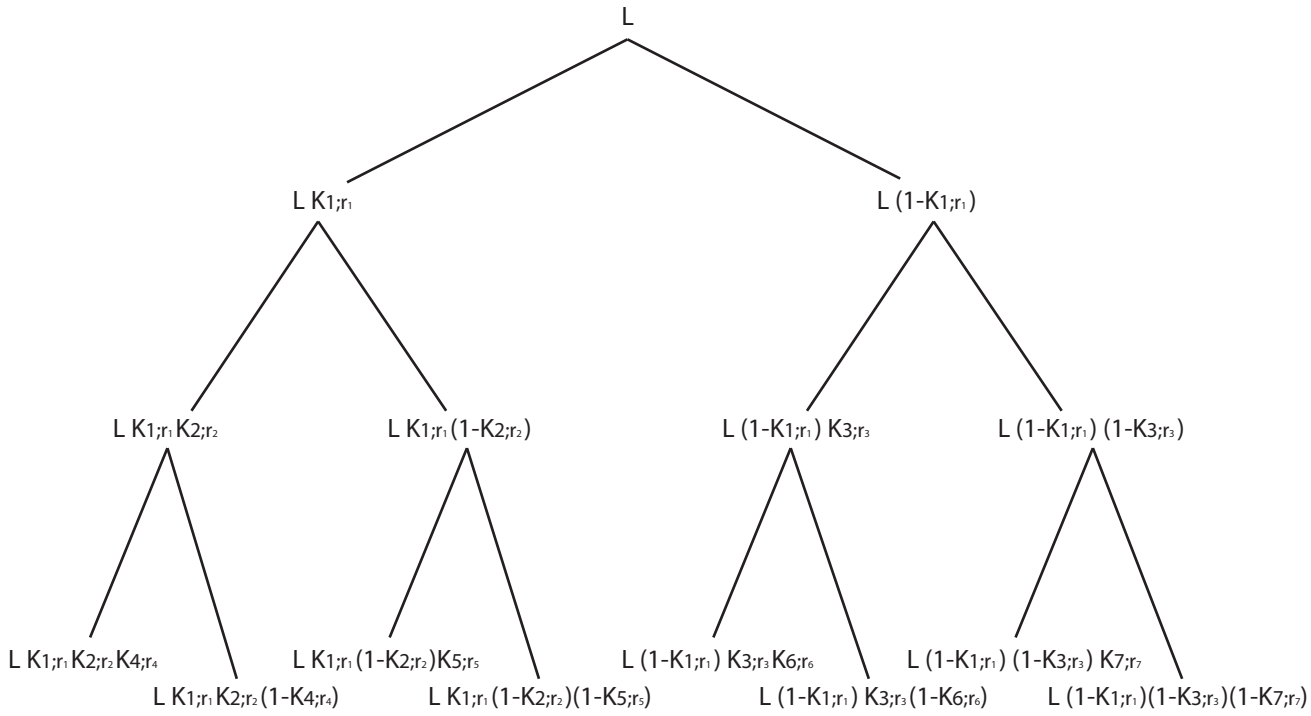


FIGURE 3.1. The decomposition achieved by splitting a stick of length L into pieces using cuts $K_{i;r_i}$ based on densities f_1 and f_2; $N = 3$ levels.

which means that as $N \to \infty$ the decomposition process will in fact tend toward Benford.

3.3. **Proof of Part 1: Benford Behavior.** We begin by proving that the first statement is true; that as we take more and more cuts, the distribution of the cuts tends toward Benford given that both functions $f_1(x)$ and $f_2(x)$ satisfy the condition stated in our theorem. We start by showing that this is equivalent to studying the distribution of a product of N independent random variables.

3.3.1. *Equivalence to a Product of N independent random variables.* By the linearity of the expectation operator, we know that

$$\mathbb{E}[P_N(s)] = \mathbb{E}\left[\frac{\sum_{i=1}^{2^N} \varphi_s(x_i)}{2^N}\right] = \frac{1}{2^N}\sum_{i=1}^{2^N} \mathbb{E}[\varphi_s(x_i)]. \tag{3.4}$$

Since we are merely summing over 2^N expected values and then dividing by 2^N, we are in a sense finding the average of $\mathbb{E}[\varphi_s(x_i)]$. Hence, to complete the proof, we must show that

$$\lim_{N\to\infty} \mathbb{E}[\varphi_s(x_i)] = \log_{10} s. \tag{3.5}$$

By definition, all pieces x_i can be expressed as the product of the starting length L and N independent random variables with values in the range $(0, 1)$. While there are dependencies amongst the x_i's, there are no dependencies among the K_i's by construction.

For a given value of i, there will be some number of cuts drawn from each of the distributions $(K_{f_1}, (1 - K_{f_1}), K_{f_2}, \text{and } (1 - K_{f_2})$). By relabeling if necessary, we can write every subsegment x_i as:

$$\begin{aligned} x_i =& LK_{1;a}K_{2;a}\cdots K_{M_i;a}K_{M_{i+1};b}\cdots K_{M_j;b}(1 - K_{M_{j+1};a}) \\ & \cdots(1 - K_{M_\ell;a})(1 - K_{M_{\ell+1};b})\cdots(1 - K_{N;b}) \end{aligned} \tag{3.6}$$

where $0 \le M_i \le M_j \le M_\ell \le N$. This means that there are:

(1) M_i cuts with factors drawn from K_{f_1}.
(2) $M_j - M_i$ cuts with factors drawn from K_{f_2}.
(3) $M_\ell - M_j$ cuts with factors drawn from $(1 - K_{f_1})$.
(4) $N - M_\ell$ cuts with factors drawn from $(1 - K_{f_2})$.

We can calculate the expected value by multiplying the value of a given cut by the probability of that cut occuring, which is given by the probability density function. Because we are only

concerned as $N \to \infty$, we will integrate over the PDF's. Hence, we have that

$$\mathbb{E}[\varphi_s(x_i)] = \int_{K_1=0}^{1} \int_{K_2=0}^{1} \cdots \int_{K_N=0}^{1}$$

$$\cdot \left(L \prod_{s=1}^{M_i} K_s \prod_{t=M_{i+1}}^{M_j} K_t \prod_{u=M_{j+1}}^{M_\ell} (1-K_u) \prod_{v=M_{\ell+1}}^{N} (1-K_v) \right)$$

$$\cdot \left(\prod_{s=1}^{M_i} f_1(K_s) dK_s \prod_{t=M_i+1}^{M_j} f_2(K_t) dK_t \right)$$

$$\cdot \left(\prod_{u=M_j+1}^{M_\ell} f_1(K_u) dK_u \prod_{v=M_\ell+1}^{N} f_2(K_v) dK_v \right). \tag{3.7}$$

This is equivalent to studying the distribution of a product of N independent random variables and then rescaling the product by L. By the Pigeon-Hole Principle we know that at least one of the following conditions must be met by at least $\frac{N}{4}$ random variables:

(1) Has factors K_s and density $f_1(K_s)$.
(2) Has factors K_t and density $f_2(K_t)$.
(3) Has factors K_u and density $f_1(1-K_u)$.
(4) Has factors K_v and density $f_2(1-K_v)$.

Because at least one of these must occur at least $\frac{N}{4}$ times in the product, as $N \to \infty$ the number of factors $\frac{N}{4} \to \infty$ as well. For this reason, we can use results proved by [6] to show that this converges to Benford. A sketch of the proof that the product of $N \to \infty$ independent random variables that satisfy our 'nice' condition will tend toward Benford is provided in Appendix A.1.

3.3.2. *Use of Mellin Transforms.* The key observation is to note that the **Mellin transform** at $1 - \frac{2\pi i \ell}{\log 10}$ is strictly less than 1 in absolute value for continuous densities if $\ell \neq 0$. This can be seen by trivially inserting absolute values in the definition of the Mellin transform. Explicitly, from Appendix A.1.3 we find that $\mathbb{E}[\varphi_s(x_i)]$ is equal to $\log_{10} s$ plus a rapidly decaying N-dependent error term for all probability densities that satisfy the condition in the statement of the theorem.

We may take the error to be independent of M_i since the formula does not depend on the choice of M_i. Because all error terms are independent and do not depend on M_i, there must be some bound that holds for all decompositions simultaneously. To find this bound, let us note that the Mellin transforms we have (with $\ell \neq 0$) are always less than 1 in absolute value. Thus the error is bounded by the maximum of the error from one of the following four products:

(1) Product of $\frac{N}{4}$ terms with density $f_1(x)$.
(2) Product of $\frac{N}{4}$ terms with density $f_2(x)$.
(3) Product of $\frac{N}{4}$ terms with density $f_1(1-x)$.
(4) Product of $\frac{N}{4}$ terms with density $f_2(1-x)$.

As $N \to \infty, \frac{N}{4} \to \infty$ as well. Since each set of terms is still independent and identically distributed random variables that satisfy equation (3.1), by Theorem 1.1 (proven in [6] and shown in the appendices) we see that

$$\text{Prob}\left(Y_N \bmod 1 \in [a,b]\right) \to b - a.$$

This implies that the base 10 logarithms are uniformly distributed, which means that $\lim_{N\to\infty} \mathbb{E}[P_N(s)] = \log_{10} s$. This by definition proves that the distribution is Benford. □

3.4. **Proof of Part II: Convergence to Benford.** Now we prove part two of the conjecture, namely that the variance tends toward zero:

$$\lim_{N\to\infty} \text{Var}(P_N(s)) = 0.$$

Recall our definition of the indicator function $\varphi_s(x_i)$:

$$\varphi_s(u) = \begin{cases} 1 & \text{if the significand is at most s} \\ 0 & \text{otherwise.} \end{cases}$$

Since $\varphi_s(x_i)$ is either 0 or 1, $\varphi_s(x_i)^2 = \varphi_s(x_i)$. We will use this observation, the definition of variance, and the linearity of the expectation operator to prove that Equation (3.2) is true.

By the definition of variance, we know that

$$\text{Var}(P_N(s)) = \mathbb{E}[P_N(s)^2] - \mathbb{E}[P_N(s)]^2. \tag{3.8}$$

As we showed at the beginning of the proof of part 1, by the linearity of the expectation operator

$$\mathbb{E}[P_N(s)] = \mathbb{E}\left[\frac{\sum_{i=1}^{2^N} \varphi_s(x_i)}{2^N}\right]. \tag{3.9}$$

We can now substitute this into (3.8) to get

$$\mathrm{Var}(P_N(s)) = \mathbb{E}\left[\left(\frac{\sum_{i=1}^{2^N} \varphi_s(x_i)}{2^N}\right)^2\right] - \mathbb{E}[P_N(s)]^2. \tag{3.10}$$

When we square two sums, the resulting sum will have the products of all combinations of i's and j's. We can pull out all cases where $i = j$ and rewrite (3.10) as the following:

$$\mathbb{E}\left[\frac{\sum_{i=1}^{2^N} \varphi_s(x_i)^2}{2^{2N}} + \frac{\sum_{\substack{i,j=1 \\ i\neq j}}^{2^N} \varphi_s(x_i)\varphi_s(x_j)}{2^{2N}}\right] - \mathbb{E}[P_N(s)]^2. \tag{3.11}$$

Because N is an integer, we can pull the denominator $\frac{1}{2^{2N}}$ out in front of the expectation operator. We can also separate the two terms so that we have:

$$\mathrm{Var}(P_N(s)) = \frac{1}{2^{2N}}\mathbb{E}\left[\sum_{i=1}^{2^N} \varphi_s(x_i)^2\right] + \frac{1}{2^{2N}}\mathbb{E}\left[\sum_{\substack{i,j=1 \\ i\neq j}}^{2^N} \varphi_s(x_i)\varphi_s(x_j)\right] - \mathbb{E}[P_N(s)]^2. \tag{3.12}$$

As we said before, since $\varphi_s(x_i) = 0$ or 1, we can make the substitution $\varphi_s(x_i)^2 = \varphi_s(x_i)$. Furthermore, we can include one $\frac{1}{2^N}$ inside the expected value to get $P_N(s)$:

$$\mathrm{Var}(P_N(s)) = \frac{1}{2^N}\mathbb{E}[P_N(s)] + \frac{1}{2^{2N}}\left(\sum_{\substack{i,j=1 \\ i\neq j}}^{2^N} \mathbb{E}[\varphi_s(x_i)\varphi_s(x_j)]\right) - \mathbb{E}[P_N(s)]^2. \tag{3.13}$$

Recall from the proof of part 1 that

$$\mathrm{Prob}(Y_N \bmod 1 \in [a, b]) = b - a + \epsilon_N \tag{3.14}$$

where ϵ_N is some error term that depends on N. We showed that $\epsilon_N \to 0$ as $N \to \infty$ because of our original assumption (3.1). Let us label $o(1)$ an error term that tends to zero as $N \to \infty$. We know from [6] that

$$\mathbb{E}[P_N(s)] = \log_{10} s + o(1). \tag{3.15}$$

We can now substitute this back into (3.13) to get

$$\mathrm{Var}(P_N(s)) = \frac{1}{2^{2N}} \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{2^N} \mathbb{E}[\varphi_s(x_i)\varphi_s(x_j)] \right) + \frac{1}{2^N} \left(\log_{10} s + o(1)\right) - \left(\log_{10} s + o(1)\right)^2. \tag{3.16}$$

As $N \to \infty$, $\frac{1}{2^N} \left(\log_{10} s + o(1)\right)$ and $o(1) \cdot \log_{10} s$ will tend toward zero, so we can lump them in with the error term $o(1)$ to get:

$$\mathrm{Var}(P_N(s)) = \frac{1}{2^{2N}} \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{2^N} \mathbb{E}[\varphi_s(x_i)\varphi_s(x_j)] \right) - \left(\log_{10} s\right)^2 + o(1). \tag{3.17}$$

For the ease of notation, we will expand the double summation and rewrite it as two separate summations over distinct indices. We will also separate out the $(\frac{1}{2^N})^2$ term:

$$\mathrm{Var}(P_N(s)) = \frac{1}{2^N} \frac{1}{2^N} \left(\sum_{\substack{j=1 \\ j \neq i}}^{2^N} \sum_{i=1}^{2^N} \mathbb{E}[\varphi_s(x_i)\varphi_s(x_j)] \right) - \left(\log_{10} s\right)^2 + o(1). \tag{3.18}$$

Since we are in essence taking an average of the expected values, what we want to show is that $\mathbb{E}[\varphi_s(x_i)\varphi_s(x_j)] \approx (\log_{10} s)^2$. We have now reduced the problem to evaluating the cross terms over all $i \neq j$. This is the hardest part of the analysis, and it is not feasible to evaluate the resulting integrals directly. Instead, for each i we will partition the pairs (x_i, x_j) based on how close x_j is to x_i in the tree in Figure 3.1.

3.4.1. *Counting.* Recall that each of the 2^N pieces is a product of the starting length L and N random variables between 0 and 1. Writing any (x_i, x_j) pair in this form, it is clear that they must share some number of these random variables as factors. Let us say that they share M terms. This means that after M stages, the pieces x_i and x_j split such that one piece has a factor K_{M+1} in its product while the other contains the factor $(1 - K_{M+1})$. the

remaining $N - M - 1$ elements in each product are independent from one another. By relabeling the indices, we can thus express any pair without loss of generality as follows:

$$x_i = L \cdot K_1 \cdot K_2 \cdots K_M \cdot K_{M+1} \cdot K_{M+2} \cdots M_N$$
$$x_j = L \cdot K_1 \cdot K_2 \cdots K_M \cdot (1 - K_{M+1}) \cdot \tilde{K}_{M+2} \cdots \tilde{K}_N.$$

Note that this expression does require a relabelling of the indices, as it is not the same as when we first began constructing the x_i tree.

We again denote the probability density functions from which these random variables are drawn as f_1 and f_2. Keeping these definitions in mind, for a fixed pair i, j we can adjust the order of the factors by separating cuts pulled from f_1 and cuts drawn from f_2:

$$\begin{aligned}
\mathbb{E}[\varphi_s(x_i)\varphi_s(x_j) \quad = \quad & \int_{K_1=0}^{1}\int_{K_2=0}^{1}\cdots\int_{K_N=0}^{1}\int_{\tilde{K}_{M+2}=0}^{1}\cdots\int_{\tilde{K}_N=0}^{1} \\
& \cdot \ \varphi_s\left(L\prod_{r=1}^{M}K_r\prod_{r=M+1}^{N}K_r\right) \\
& \cdot \ \varphi_s\left(L\prod_{r=1}^{M}K_r\right)\cdot(1-K_{M+1})\cdot\left(\prod_{r=M+2}^{N}\tilde{K}_r\right) \\
& \cdot \ \prod_{r=1}^{M_1}f_1(K_r)\cdot\prod_{r=M_1}^{M}f_2(K_r) \\
& \cdot \ \prod_{r=M+1}^{N_1}f_1(K_r)\cdot\prod_{r=N_1}^{N}f_2(K_r) \\
& \cdot \ \prod_{r=M+2}^{N_2}f_1(1-\tilde{K}_r)\cdot\prod_{r=N_2}^{N}f_2(1-\tilde{K}_r) \\
& \cdot \ dK_1dK_2\cdots dK_Nd\tilde{K}_{M+2}\cdots d\tilde{K}_N. \qquad (3.19)
\end{aligned}$$

In the expression above, we have already integrated over the remaining $2^N - N - (N - M - 1) = 2^N - 2N + M + 1$ variables. Because none of these variables appear in the cuts of x_i or x_j, their corresponding integrals are 1, so they do not affect the product and we will no longer write them.

The challenge in understanding (3.19) is that many variables occur in both x_i and x_j. The key observation is that most of the time there are many variables occuring in one but not the other, which minimizes the effect of the common variables, allowing us to evaluate $\mathbb{E}[\varphi_s(x_i)\varphi(x_j)]$ at almost independent arguments. We will now make the argument precise.

We can study the behavior of the integral in (3.19) as a function of the significand of the first $M+1$ random variables. More specifically, we define the following functions:

$$\begin{aligned} I_1(L_1) &= \int_{K_{M+1}=0}^{1} \cdots \int_{K_N=0}^{1} \varphi_s\left(L_1 \prod_{r=M+1}^{N_1} K_r \prod_{r=N_1}^{N} K_r\right) \\ &\quad \prod_{r=M+1}^{N_1} f_1(K_r) \prod_{r=N_1}^{N} f_2(K_r) dK_{M+1} dK_{M+2} \cdots dK_N \end{aligned} \tag{3.20}$$

$$\begin{aligned} I_2(L_2) &= \int_{\tilde{K}_{M+1}=0}^{1} \cdots \int_{\tilde{K}_N=0}^{1} \varphi_s\left(L_2 \prod_{r=M+1}^{N_2} \tilde{K}_r \prod_{r=N_2}^{N} \tilde{K}_r\right) \\ &\quad \prod_{r=M+1}^{N_2} f_1(1-\tilde{K}_r) \prod_{r=N_2}^{N} f_2(1-\tilde{K}_r) d\tilde{K}_{M+1} d\tilde{K}_{M+2} \cdots d\tilde{K}_N. \end{aligned} \tag{3.21}$$

These two functions are defined for all $L_1, L_2 \in [1, 10)$. We will show that for any L_1, L_2 we have

$$|I_1(L_1) I_2(L_2) - (\log_{10} s)^2| = o(1). \tag{3.22}$$

Once we have this, all that remains is to integrate $I(L_1)I(L_2)$ over the remaining M variables $(K_1, ..., K_M)$ that both cuts share. We know that this will equal $(\log_{10} s)^2 + o(1)$ because it is the product of random variables, and hence Benford. The rest of the proof follows from counting for a given i, how many j lead to a given M. □

3.4.2. *Proving* $|I(L_1)I(L_2) - (\log_{10} s)^2| = o(1)$. We know by Theorem A.1 (proved by [6] and summarized in Appendix A.1) that as $N \to \infty$ the distribution of leading digits of x_i tends to Benford's Law. Furthermore, the error term is a nice function of the Mellin transforms.

Explicitly, if $y_N = \log_B x_N$ then

$$\begin{aligned} |\text{Prob}(y_n \bmod 1 \quad &\in \quad [a,b]) - (b-a)| \\ &\leq \quad (b-a) \left| \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} \prod_{m=1}^{N} (\mathcal{M} f_{m_n}) \left(1 - \frac{2\pi i \ell}{\log B}\right) \right|. \end{aligned} \tag{3.23}$$

This means that our functions $I_1(L_1)$ and $I_2(L_2)$ are bounded by

$$|I_1(L_1) - \log_{10} s| \leq (b-a) \left| \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} \prod_{m=1}^{N} (\mathcal{M} f_{m_n}) \left(1 - \frac{2\pi i \ell}{\log B}\right) \right|. \tag{3.24}$$

To simplify our computations, let us label

$$\mathcal{M}_n = (b-a) \left| \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} \prod_{m=1}^{N} (\mathcal{M} f_{m_n}) \left(1 - \frac{2\pi i \ell}{\log B}\right) \right|. \tag{3.25}$$

Recall from our original assumption (3.1) that as $N \to \infty$, the limit of the expression within the absolute value converges to zero, and that a and b are merely constants. By the rules of absolute value, we know that

$$0 \quad \leq |I_1(L_1) - \log_{10} s| \leq \mathcal{M}_n$$

$$0 \quad \leq |I_2(L_2) - \log_{10} s| \leq \mathcal{M}_n. \tag{3.26}$$

Using the Triangle Inequality (worked out in detail in Appendix A.2) we get

$$|I_1(L_1) I_2(L_2) - (\log_{10} s)^2| \leq 2\mathcal{M}_n. \tag{3.27}$$

For each of the 2^N choices of i and for each $1 \leq n \leq N$ there are 2^{n-1} choices of j such that x_j has exactly n factors not in common with x_i. We can therefore obtain an upper bound for the sum of the expectation cross terms by summing the bound obtained for $2^{n-1} I_1(L_1) I_2(L_2)$ over all n and all i:

$$\left| \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} \left(\mathbb{E}[\varphi_s(x_i)\varphi_s(x_j)] - (\log_{10} s)^2 \right) \right| \leq \left| \sum_{i=1}^{2^N} \sum_{n=1}^{N} 2^{n-1} \mathcal{M}_n \right|. \tag{3.28}$$

We know that $\lim_{N\to\infty} \mathcal{M}_n = 0$ so no matter how many error terms we add up, when we divide by 2^{2N} the right side will approach zero. Since an absolute value must be ≥ 0, this means that equality holds:

$$\left| \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \left(\mathbb{E}[\varphi_s(x_i)\varphi_s(x_j)] - (\log_{10} s)^2 \right) \right| = 0. \tag{3.29}$$

Recall from Equation (3.17) that

$$\mathrm{Var}(P_N(s)) = \frac{1}{2^{2N}} \left(\sum_{\substack{i,j=1 \\ i\neq j}}^{2^N} \mathbb{E}[\varphi_s(x_i)\varphi_s(x_j)] \right) - (\log_{10} s)^2 + o(1). \tag{3.30}$$

Since N is a positive integer, we know that

$$\frac{1}{2^{2N}} \left(\sum_{\substack{i,j=1 \\ i\neq j}}^{2^N} \mathbb{E}[\varphi_s(x_i)\varphi_s(x_j)] \right) - (\log_{10} s)^2 \leq \left| \sum_{\substack{i,j=1 \\ i\neq j}}^{2^N} \mathbb{E}[\varphi_s(x_i)\varphi_s(x_j)] - (\log_{10} s)^2 \right|, \tag{3.31}$$

where we just showed in (3.29) that the right side is 0. This means that

$$\begin{aligned} \mathrm{Var}(P_N(s)) &\leq 0 + o(1) \\ \mathrm{Var}(P_N(s)) &\leq o(1). \end{aligned} \tag{3.32}$$

By definition, this means that the error tends to zero as $N \to \infty$ which completes the second part of our proof. □

4. Conjecture

We have shown that if we pull cuts from a finite set of nice density functions, the distribution of the lengths of the resulting subsegments will tend toward Benford. This agrees with our intuition, as the mixture of possible cuts is diverse, leading to a sense of randomness which is often associated with Benford's Law. We might expect that when we reduce this randomness, say to the extreme of specifying only one cut to repeatedly apply at every level and for every subsegment, that the distribution of cuts would no longer follow Benford's Law. Consider, if we were to always cut a stick in the middle at 0.5 times its length, then at any given stage, all subsegments will have the same exact length. Clearly, the distribution of the lengths would no longer follow Benford's Law.

Contrary to our intuition, however, we find that even when we apply the exact same cut, as long as that cut is not 0, 0.5, or 1, as we increase the number of levels used, the distribution of lengths will in fact tend toward Benford.

4.1. **Statement of Conjecture.**

Conjecture 4.1. *We are given a stick of length L and a cut $p \in (0,1)$ where $p \neq 0.5$. In the first level $N = 1$, we cut L into two subsegments $L \cdot p$ and $L \cdot (1-p)$. At the second level $N = 2$, we cut both subsegments again using p and $(1-p)$ (resulting in four cuts of length $L \cdot p^2$, $L \cdot p(1-p)$, $L \cdot (1-p)p$, and $L \cdot (1-p)^2$). We continue this cutting process, applying p and $(1-p)$ at every level to each subsegment. As we apply $N \to \infty$ levels of cuts, the distribution of the lengths of the cuts will tend toward Benford's Law.*

Applying the same cut p at every stage will lead to the general decomposition visualized in Figure 4.1. As we can see, at any level N there will be 2^N subsegments. Of those subsegments, $\binom{2^N}{n_d}$ of them will have the value $L \cdot p^{n_d}(1-p)^{N-n_d}$. Given a starting length L, a set cut p, and choice of level N, we can explicitly calculate the frequency of any leading digit and compare it to the Benford probabilities.

4.2. **Discussion of Conjecture.**

Although we are currently unable to prove that this distribution will converge to Benford for any $p \in (0,1)$ where $p \neq 0.5$, through calculation we can see that Conjecture 4.1 seems to hold at high levels of N. In Figure 4.2 we visualize the distribution where $p = 0.75$ and $N = 1000$. The Mathematica code used to perform the calculations can be found in Appendix A.3.

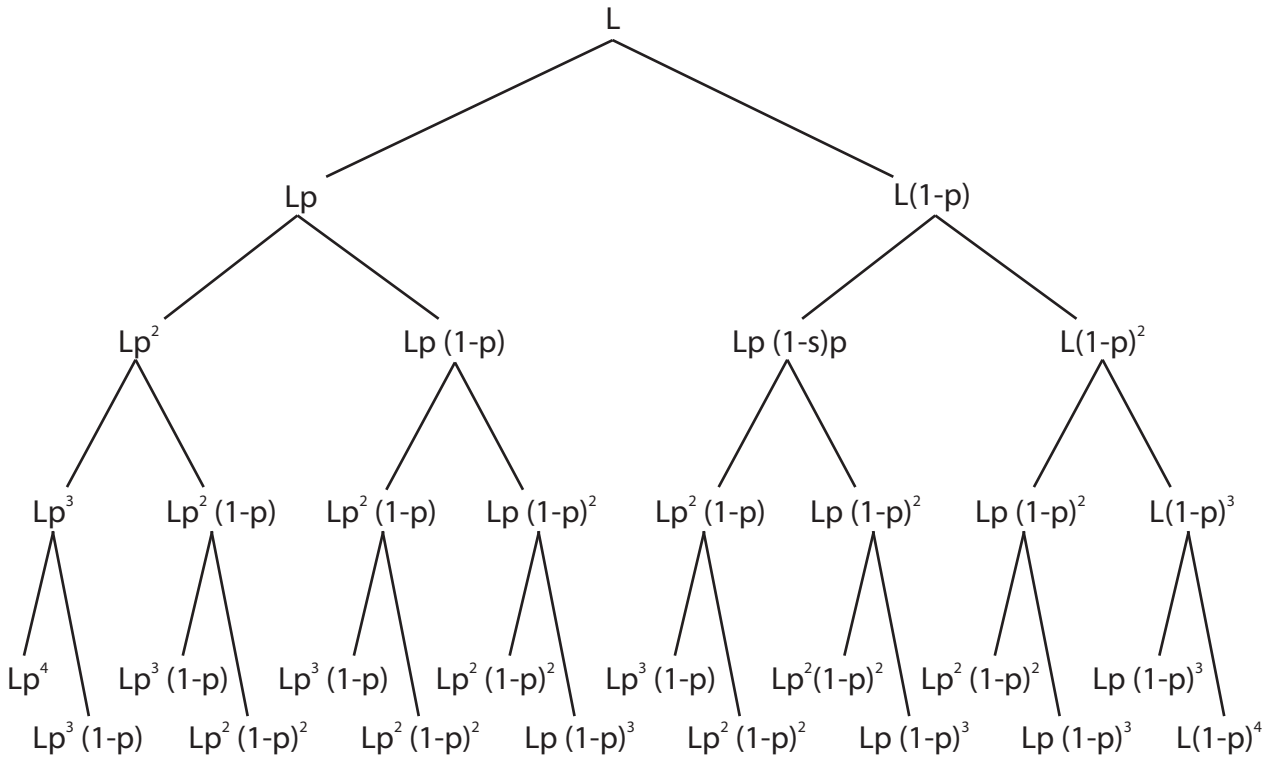


FIGURE 4.1. Lengths of subsegments when we split L into subsegments using p at every level, $N = 4$ levels.

We may expect this tendency to hold only to a certain point, hypothesizing that if p is close enough to 0, 0.5, or 1, that the relationship will stop tending toward Benford. When we first simulate $p = 0.51$ at $N = 1000$ levels, this appears to be the case, as seen in Figure 4.3a. But when we increase N by a factor of 10 and simulate to $N = 10000$ levels, we find that the distribution will eventually tend toward Benford (Figure 4.3b). Since we are only concerned with the distribution as $N \to \infty$, this implies that we did not complete enough levels when we first calculated the distribution, and at higher levels of N the distribution will in fact converge to Benford.

A common mistake in analyzing whether such a process will tend toward Benford is to stop after too few levels, before the pattern becomes evident. This may lead to inconclusive or inaccurate results, so we want to analyze the distribution after as many levels as possible. Ideally, we would compute a trillion levels, or even more. We are limited by computing power, however, as each computation takes time to perform.

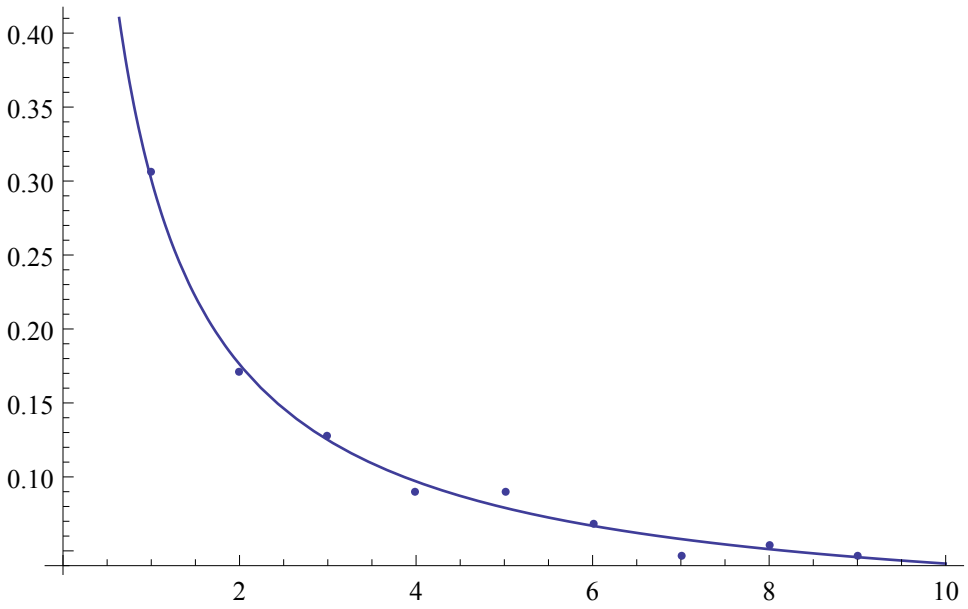
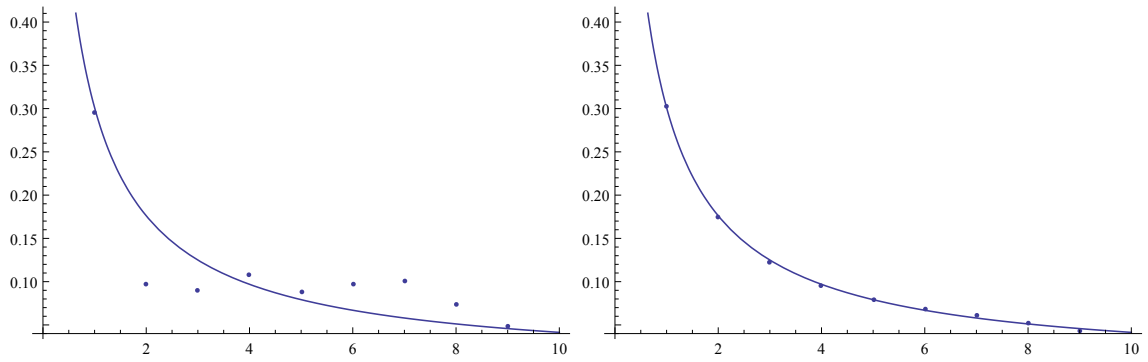


FIGURE 4.2. The distribution of leading digits for a constant cut $p = 0.75$ to $N = 1000$ levels with the Benford probabilities overlaid. Probability of a leading digit occurring as a function of the value of that leading digit.

Leading Digit	**Benford Probability**	**Observed Frequency** ($p = 0.75, N = 1000$)	**Observed Frequency** ($p = 0.51, N = 10000$)
1	0.30103	0.306156	0.302846494679860
2	0.176091	0.170253	0.175495691232183
3	0.124939	0.127292	0.121681641185727
4	0.09691	0.0901434	0.096031886085056
5	0.0791812	0.0898191	0.078481103805340
6	0.0669468	0.0680362	0.068912864854509
7	0.0579919	0.0474461	0.061279883596516
8	0.0511525	0.0536766	0.0513891557265013
9	0.0457575	0.0471777	0.043881278834309

TABLE 1. Table of expected versus observed frequencies of leading digits when applying a constant cut p.

Fortunately, because we know the distribution and values of the subsegments at each level, we need not compute the value of 2^N different cuts but merely $N + 1$ values (Lp^N,



(A) $p = 0.51$ and $N = 1000$.

(B) $p = 0.51$ and $N = 10000$.

FIGURE 4.3. Distribution of leading digits for a constant cut $p = 0.51$ calculated to N levels. Expected Benford distribution overlaid.

$Lp^{N-1}(1-p)$, $Lp^{N-2}(1-p)^2$, $\cdots$, $Lp^2(1-p)^{N-2}$, $Lp(1-p)^{N-1}$, $L(1-p)^N$) and count how many subsegments of each length we will have at the N^{th} stage, namely $\binom{2^N}{n_d}$. But at $N = 1000000$, even a million values requires a nontrivial amount of time to compute, thus limiting our computing ability.

As the examples with $p = 0.51$ and $p = 0.99$ show, even if there is only a little variation to begin with, as we increase the number of levels the distribution becomes close to Benford in at least one point. As $p \to 0$, $p \to 0.5$, or $p \to 1$, it may take longer for the distribution to appear Benford, but it seems that they will ultimately follow Benford's Law.

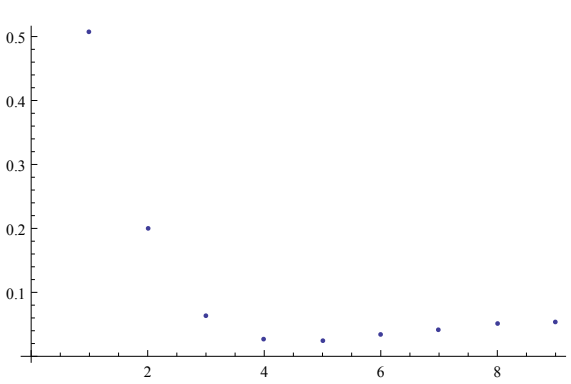
4.2.1. *Potential Periodic Behavior.* The counts may be displaying periodic behavior or there may be an underlying pattern that is non-Benford. Even though the distribution seems to tend toward Benford at first, if we increase N even larger (beyond our computing power) it may tend away from Benford.

4.2.2. *Transformation of leading digit distribution.* To examine this case, we consider the changing distribution of $p = 0.501$ as we increase N from $1000 \to 10000$ and compare it to the changing distribution of $p = 0.51$ as N increases from $100 \to 2200$.

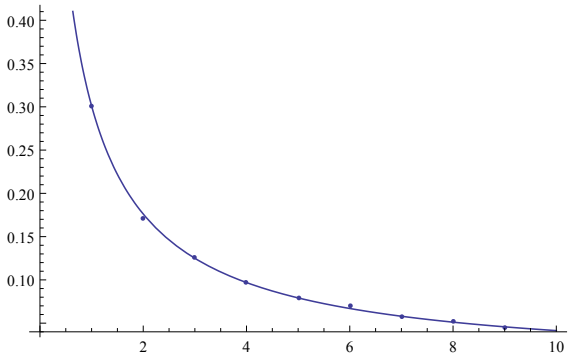
When we simulate the distribution for a value that is closer to 0.5 (such as $p = 0.501$) the distribution seems to remain peaked near the value of 0.5^N where the median of the peak shifts to the left and the base of the peak does not widen (Figure A.1). This implies that $p = 0.501$ may be 'too close' to 0.5, and that it will not ultimately converge to Benford. Yet if we compare it to how the distribution of $p = 0.51$ transforms at lower levels of N (from 100

to 500) we find that the changes in distribution are similar to those that $p = 0.51$ undergo from 1000 to 9000 (Figure A.2). This implies that $p = 0.501$ may still converge to Benford like $p = 0.51$ (Figure A.2), just at a slower rate. Figures depicting these transformations can be referenced in Appendices A.4 and A.5.

Similarly, when we simulate $p = 0.99$ to $N = 10,000$ levels the distribution seems to follow a pattern that is distinctly non-Benford (Figure 4.4a). When we increase our calculation to $N = 50,000$ levels, however, we see that the distribution seems to converge to Benford (Figure 4.4b).



(A) $p = 0.99$ and $N = 10000$.



(B) $p = 0.99$ and $N = 50000$. Benford distribution overlaid.

FIGURE 4.4. Distribution of leading digits for a constant cut $p = 0.99$ calculated to N levels.

If the distribution is in fact displaying periodic behavior, then our conjecture may not be true, although we could calculate the number of levels needed to be 'close' to Benford as a function of p.

Appendix A. Appendices

A.1. **Proof that the product of i.i.d. random variables is Benford.** We sketch the proof from [6] and modified by [1] of the theorem that the product of N 'nice' independent identically distributed random variables as $N \to \infty$ will tend toward Benford. Whereas [6] proved this for chains of random variables, we reiterate the modified version of [1] who showed the same results for products of random variables.

A.1.1. *Terminology.* First, let us set notation and review some pertinent properties of the Mellin transform. Let K_1 and K_2 be two independent random variables from the probability density function f with cumulative distribution function F. Then the **density** of their product is given by

$$\int_{t=0}^{\infty} f\left(\frac{x}{t}\right) f(t)\frac{dt}{t}. \tag{A.1}$$

This can be easily generalized to the product of more factors. To understand why this is the case, we first calculate the probability that $K_1 \cdot K_2 \in \left[0, \frac{x}{t}\right]$ and then differentiate our result with respect to x:

$$\text{Prob}(K_1 \cdot K_2 \in [0, x]) = \int_{t=0}^{\infty} \text{Prob}\left(K_2 \in \left[0, \frac{x}{t}\right]\right) f(t)dt. \tag{A.2}$$

Since $\text{Prob}\left(K_2 \in \left[0, \frac{x}{t}\right]\right)$ is the **cumulative distribution function** $F\left(\frac{x}{t}\right)$ we can rewrite equation (A.2) as

$$\int_{t=0}^{\infty} F\left(\frac{x}{t}\right) f(t)dt. \tag{A.3}$$

Differentiating the cumulative density function merely returns the density of $K_1 \cdot K_2$ as seen in (A.1).

If $g(s)$ is an analytic function for $\Re(s) \in (a, b)$ such that $g(c + iy)$ tends to zero uniformly as $|y| \to \infty$ for any $c \in (a, b)$, then the **inverse Mellin transform**, $(\mathcal{M}^{-1}g)(x)$, is given by

$$(\mathcal{M}^{-1}g)(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} g(s)x^{-s}ds. \tag{A.4}$$

The **Mellin convolution theorem** states that

$$(\mathcal{M}(f_1 \star f_2))(s) = (\mathcal{M}f_1) \cdot (\mathcal{M}f_2)(s), \tag{A.5}$$

which by induction gives

$$(\mathcal{M}(f_1 \star \cdots \star f_N))(s) = (\mathcal{M}f_1)(s) \cdots (\mathcal{M}f_N)(s). \tag{A.6}$$

Note that $f_1 \star \cdots \star f_N$ is the density of the product of N independent random variables.

Now that we have set notation, we sketch the proof that products of independent identically distributed random variables converge to Benford, and isolate out the error term. This proof was first given by [6] and later modified by [1].

A.1.2. *Statement of Theorem.*

Theorem A.1. *Let $K_1, ..., K_N$ be independent random variables with densities f_{m_n}. Assume*

$$\lim_{N\to\infty} \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \prod_{n=1}^{N} (\mathcal{M}f_{m_n}) \left(1 - \frac{2\pi i \ell}{\log B}\right) = 0. \tag{A.7}$$

Then as $N \to \infty$, $x_N = K_1 \cdots K_N$ converges to Benford's law. In particular, if $y_N = \log_B x_N$ then

$$\begin{aligned} |\text{Prob}(y_N \bmod 1 \quad &\in \quad [a,b]) - (b-a)| \\ &\leq \quad (b-a) \cdot \left| \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \prod_{n=1}^{N} (\mathcal{M}f_{m_n}) \left(1 - \frac{2\pi i \ell}{\log B}\right) \right|. \end{aligned} \tag{A.8}$$

A.1.3. *Proof of Theorem.*

Proof. To investigate the distribution of the digits of $x_N = K_1 K_2 \cdots K_N$ (base B) let us first make a logarithmic change of variables, setting $y_N = \log_B x_N$. We now have

$$\text{Prob}(y_N \leq Y) = \text{Prob}(x_N \leq B^Y) = F_N(B^Y), \tag{A.9}$$

where f_N is the density of X_n and F_N is the cumulative distribution function. As we said before, taking the derivative gives the density of y_N which we denote by $g_N(Y)$:

$$g_N(Y) = f_N(B^Y) B^Y \log B. \tag{A.10}$$

A standard method to show that x_N tends to Benford behavior is to show that $y_N \bmod 1$ tends to the uniform distribution on $[0,1]$. The key ingredient to this calculation is **Poisson Summation** [10], which relates the Fourier series coefficients of the periodic summation of a function to values of the function's continuous Fourier transform. Namely,

$$\sum_{j=-\infty}^{\infty} f(j) = \sum_{j=-\infty}^{\infty} \hat{f}(j) \tag{A.11}$$

where $\hat{f}$ is the **Fourier transform** of f, defined as:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx. \tag{A.12}$$

We show that $y_N \bmod 1$ tends to the uniform distribution in the following calculation. Let $h_{N;Y}(t) = g_N(Y + t)$. Then

$$\begin{aligned} \sum_{l=-\infty}^{\infty} g_N(Y+\ell) &= \sum_{l=-\infty}^{\infty} h_{N;Y}(\ell). \\ &\overset{\text{Poiss Sum}}{=} \sum_{l=-\infty}^{\infty} \hat{h}_{N;Y}(\ell) \\ &= \sum_{l=-\infty}^{\infty} e^{2\pi i Y \ell} \hat{g}_N(\ell). \end{aligned} \tag{A.13}$$

Letting $[a, b] \subset [0, 1]$, we see that

$$\begin{aligned} \text{Prob}(y_N \bmod 1 \in [a,b]) &= \sum_{l=-\infty}^{\infty} \int_{a+\ell}^{b+\ell} g_N(Y) dy \\ &= \sum_{l=-\infty}^{\infty} \int_{a+\ell}^{b+\ell} g_N(Y+\ell) dy \\ &= \sum_{l=-\infty}^{\infty} \int_{a+\ell}^{b+\ell} e^{2\pi i Y \ell} \hat{g}_N(\ell) dy \\ &= b - a + \text{Err}\left((b-a) \sum_{l \neq 0} |\hat{g}_N(\ell)| \right) \end{aligned} \tag{A.14}$$

where $\text{Err}(z)$ means an error at most z in absolute value. Note that since g_N is a probability density, $\hat{g}_N(0) = 1$. The proof is completed by showing that the sum over ℓ tends to zero as $N \to \infty$. we thus need to compute $\hat{g}_N(\ell)$. Using equation (A.10) and the substitution

$t = B^Y$ we have the following:

$$\begin{aligned}
\hat{g}_N(\xi) &= \int_{-\infty}^{\infty} g_N(Y) e^{-2\pi i Y \xi} dy \\
&= \int_{-\infty}^{\infty} f_N(B^Y) B^Y \log B \cdot e^{-2\pi i Y \xi} dy \\
&= \int_{-\infty}^{\infty} f_N(t) t^{-2\pi i \xi / \log B} dt \\
&= (\mathcal{M} f_N)\left(1 - \frac{2\pi i \xi}{\log B}\right) \\
&= \prod_{n=1}^{N} (\mathcal{M} f_N)\left(1 - \frac{2\pi i \xi}{\log B}\right). \qquad \text{(A.15)}
\end{aligned}$$

We know by the condition stated in our theorem that every density function f_{m_n} satisfies

$$\prod_{n=1}^{N} (\mathcal{M} f_{m_n})\left(1 - \frac{2\pi i \xi}{\log B}\right) = 0. \qquad \text{(A.16)}$$

We can now substitute (A.16) into (A.14) to show that the error term will tend toward zero, which means that as $N \to \infty$, the following equality will hold:

$$\text{Prob}(y_N \bmod 1 \in [a, b]) = b - a.$$

Thus the product of these independent, identically distributed random variables converges to Benford's law, which concludes our proof [10]. □

A.2. **Triangle Inequality.** Although each individual pair of cuts has a finite distance bounded by the number of separations N, we are concerned with the distribution as N approaches infinity. If our original condition holds, we know that as $N \to \infty$,

$$\sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} \prod_{n=1}^{N} (\mathcal{M} f_{m_n})\left(1 - \frac{2\pi i \ell}{\log B}\right) \to 0.$$

This means that

$$\lim_{N\to\infty} |I_1(L_1) - \log_{10} s)| \leq (b-a) \lim_{N\to\infty} \left| \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} \prod_{n=1}^{N} (\mathcal{M} f_{m_n})\left(1 - \frac{2\pi i \ell}{\log B}\right) \right|$$

$$\lim_{N\to\infty} |I_1(L_1) - \log_{10} s)| \leq 0. \qquad \text{(A.17)}$$

Since absolute values must be non-negative, equality holds:

$$\lim_{N\to\infty} |I_1(L_1) - \log_{10} s)| = 0. \tag{A.18}$$

Similarly, we find that

$$\lim_{N\to\infty} |I_2(L_2) - \log_{10} s)| = 0. \tag{A.19}$$

We know that $|I_1(L_1)|$, $|I_2(L_2)|$, and $|\log_{10} s| \leq 1$, so we can compute the following:

$$\begin{aligned}
& |I_1(L_1)I_2(L_2) - (\log_{10} s)^2| \\
= \; & |I_1(L_1)I_2(L_2) - I_2(L_2)\log_{10} s + I_2(L_2)\log_{10} s - (\log_{10} s)^2| \\
\leq \; & |I_1(L_1)I_2(L_2) - I_2(L_2)\log_{10} s| + |I_2(L_2)\log_{10} s - (\log_{10} s)^2| \\
= \; & |I_2(L_2)||I_1(L_1) - \log_{10} s| + |\log_{10} s||I_2(L_2) - \log_{10} s| \\
\leq \; & |I_1(L_1) - \log_{10} s| + |I_2(L_2) - \log_{10} s|. \qquad \text{(A.20)}
\end{aligned}$$

From equation (A.18) and (A.19) we know that both of these expressions tend to 0. Hence, the entire expression is $= 0$.

A.3. **Mathematica Code for Conjecture Calculation.** The following is the Mathematica code used to calculate cutting a stick of length $L = 1$ into pieces where at any given level we cut at $p \in (0, 1)$ times the length of each subsegment. The function also displays the observed frequency of each leading digit and compares it to its associated probability of occurence under Benford's Law.

```
fd[x_] := Floor[10^Mod[Log[10, 1.0 x], 1]]
BenfordFixedCut[s_, level_] := Module[{},
  data = {};
  For[d = 1, d <= 9, d++, count[d] = 0];
  For[k = 0, k <= level, k++,
    {
      x = fd[(s)^k (1 - s)^(level - k)];
      count[x] = count[x] + Binomial[level, k];
     }];
  (*data = AppendTo[data, fd[(3/4)^k (1/4)^(level-k)]]*)

  dataplot = {};
```

```
  For[d = 1, d <= 9, d++,
    dataplot = AppendTo[dataplot, {d, count[d]/2.^level}]];
    Print[ListPlot[dataplot]];
  For[d = 1, d <= 9, d++,
    Print["d = ", d, " and Benford Prob = ", Log[10, (d + 1.)/d],
    " and observe ", count[d]/2.^level, "."]];
  ]
```

Calculating the distribution when $p = 0.75$ and $N = 1000$ produces the following output (as well as the graph in Figure 4.2):

```
BenfordFixedCut[3/4, 1000]


d = 1 and Benford Prob = 0.30103 and observe 0.306156.
d = 2 and Benford Prob = 0.176091 and observe 0.170253.
d = 3 and Benford Prob = 0.124939 and observe 0.127292.
d = 4 and Benford Prob = 0.09691 and observe 0.0901434.
d = 5 and Benford Prob = 0.0791812 and observe 0.0898191.
d = 6 and Benford Prob = 0.0669468 and observe 0.0680362.
d = 7 and Benford Prob = 0.0579919 and observe 0.0474461.
d = 8 and Benford Prob = 0.0511525 and observe 0.0536766.
d = 9 and Benford Prob = 0.0457575 and observe 0.0471777.
```

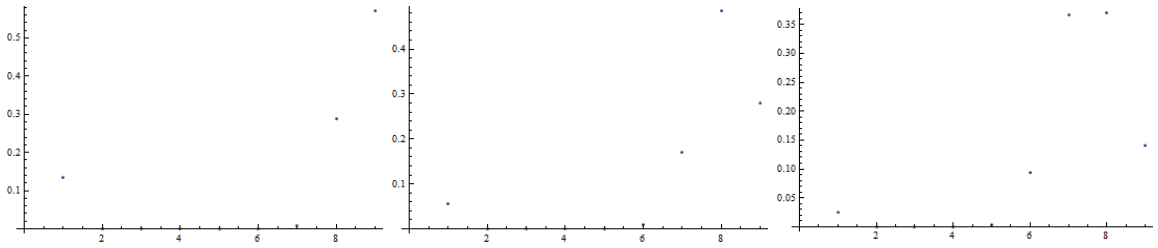
The expected Benford distribution was included by using the command

```
Print[Show[
  Plot[Log[10, (d + 1.)/d], {d, 0, 10}],
  ListPlot[dataplot]
  ]]
```

in place of:

```
Print[ListPlot[dataplot]];.
```

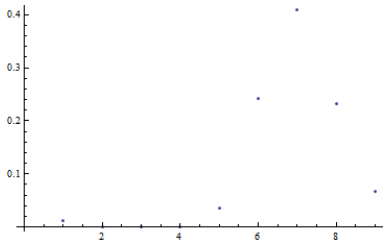
A.4. **Distribution Transformation:** $p = 0.501$.



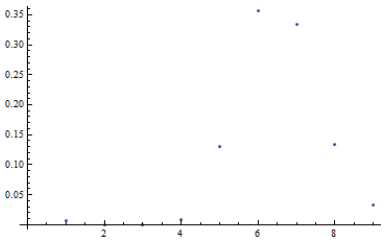
(A) $p = 0.501$, $N = 1000$, and $S(0.5^{1000}) = 9$.

(B) $p = 0.501$, $N = 2000$, and $S(0.5^{2000}) = 8$

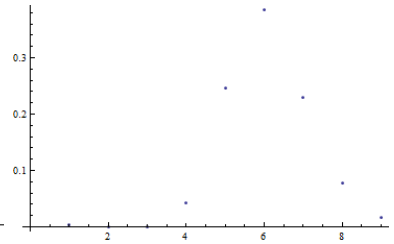
(C) $p = 0.501$, $N = 3000$, and $S(0.5^{3000}) = 8$



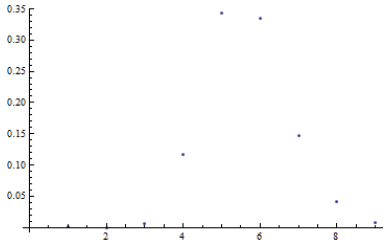
(D) $p = 0.501$, $N = 4000$, and $S(0.5^{4000}) = 7$



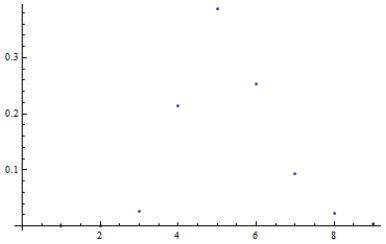
(E) $p = 0.501$, $N = 5000$, and $S(0.5^{5000}) = 7$



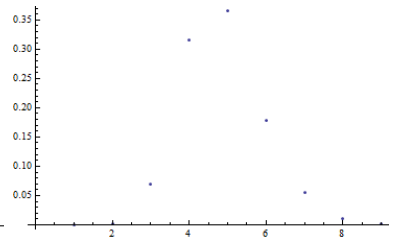
(F) $p = 0.501$, $N = 6000$, and $S(0.5^{6000}) = 6$



(G) $p = 0.501$, $N = 7000$, and $S(0.5^{7000}) = 8$



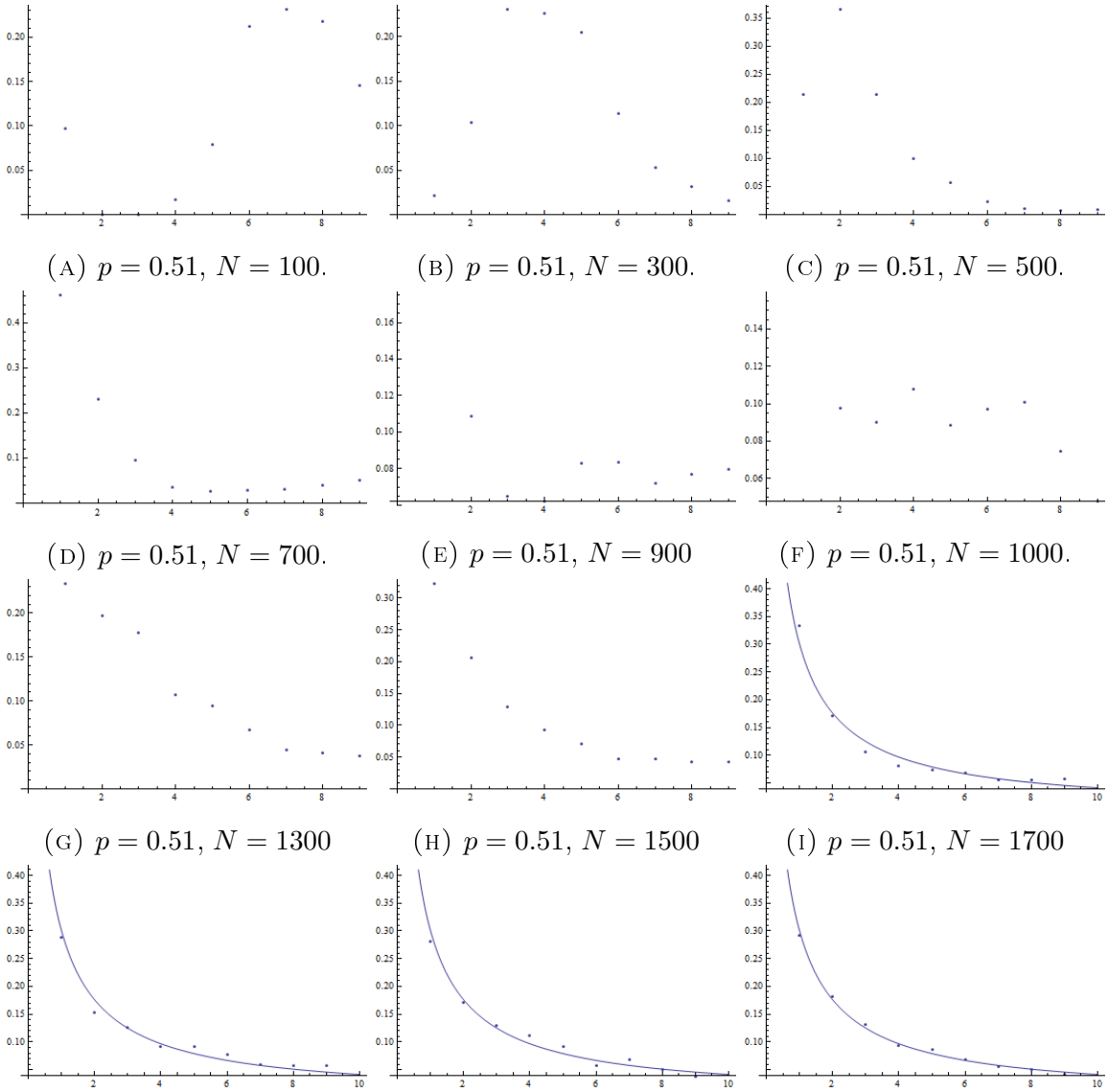
(H) $p = 0.501$, $N = 8000$, and $S(0.5^{8000}) = 5$



(I) $p = 0.501$, $N = 9000$, and $S(0.5^{9000}) = 5$

FIGURE A.1. Distribution of leading digits for a constant cut $p = 0.501$ calculated to N levels.

A.5. **Distribution Transformations: $p = 0.51$ and $p = 0.99$.**



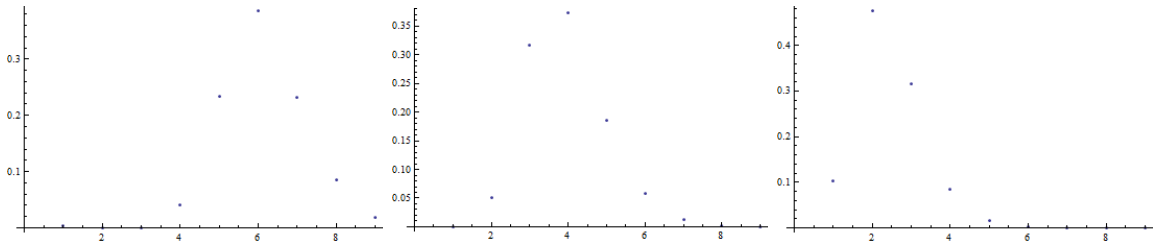
(A) $p = 0.51$, $N = 100$. (B) $p = 0.51$, $N = 300$. (C) $p = 0.51$, $N = 500$.

(D) $p = 0.51$, $N = 700$. (E) $p = 0.51$, $N = 900$ (F) $p = 0.51$, $N = 1000$.

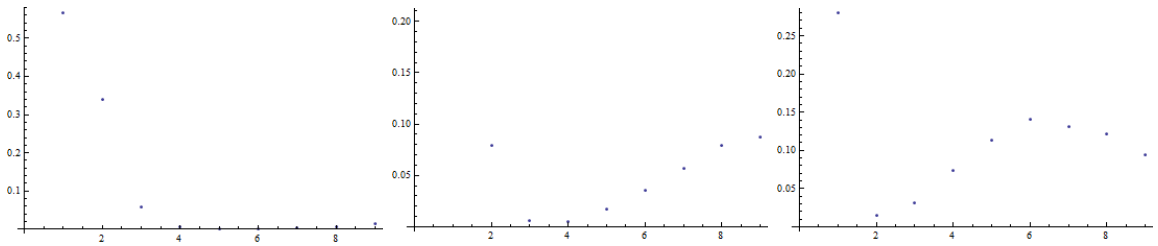
(G) $p = 0.51$, $N = 1300$ (H) $p = 0.51$, $N = 1500$ (I) $p = 0.51$, $N = 1700$

(J) $p = 0.51$, $N = 1900$ (K) $p = 0.51$, $N = 2100$ (L) $p = 0.51$, $N = 2200$

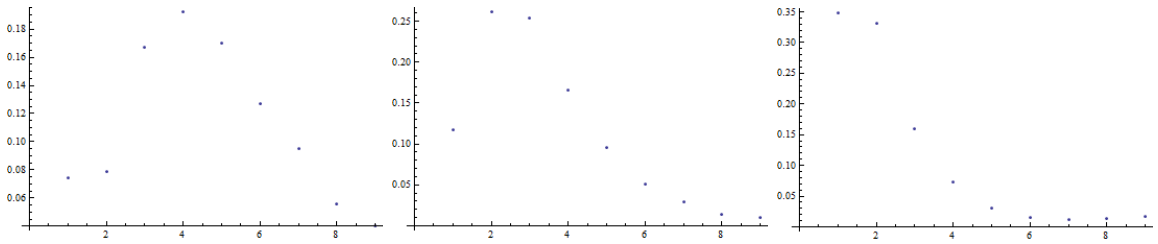
FIGURE A.2. Distribution of leading digits for a constant cut $p = 0.51$ calculated to N levels.



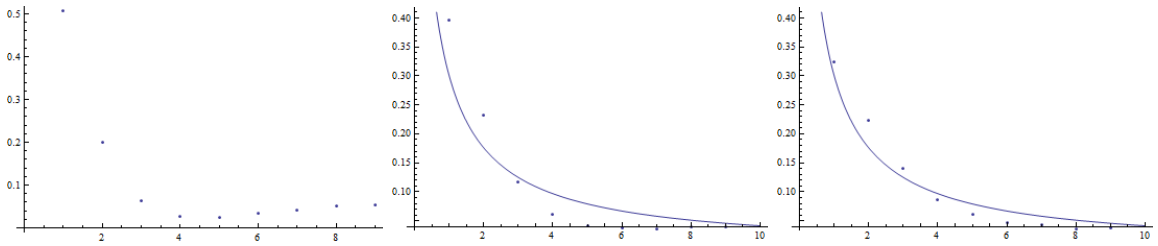
(A) $p = 0.99$ and $N = 1000$. (B) $p = 0.99$ and $N = 2000$. (C) $p = 0.99$ and $N = 3000$.



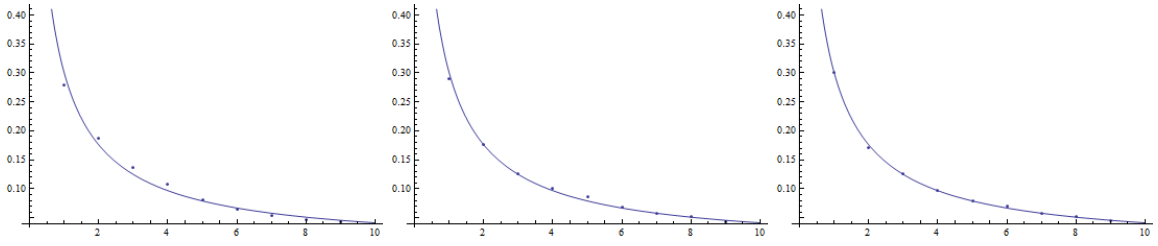
(D) $p = 0.99$ and $N = 4000$. (E) $p = 0.99$ and $N = 5000$. (F) $p = 0.99$ and $N = 6000$.



(G) $p = 0.99$ and $N = 7000$. (H) $p = 0.99$ and $N = 8000$. (I) $p = 0.99$ and $N = 9000$.



(J) $p = 0.99$ and $N = 10000$. (K) $p = 0.99$ and $N = 15000$. (L) $p = 0.99$ and $N = 20000$.



(M) $p = 0.99$ and $N = 30000$. (N) $p = 0.99$ and $N = 40000$. (O) $p = 0.99$ and $N = 50000$.

FIGURE A.3. Distribution of leading digits for a constant cut $p = 0.99$ calculated to N levels.

References

[1] T. Becker, A. Greaves-Tunnell, S. J. Miller, R. Ronan, F. W. Strauch *Benford's Law and Continuous Dependent Random Variables* (preprint 2011) arXiv:1111.0568.

[2] F. Benford, *The Law of Anomalous Numbers*, Proceedings of the American Philosophical Society **78** (1938), no. 4, 551-572. http://www.jstor.org/stable/984802

[3] A. Berger and T. Hill, *Newton's Method Obeys Benford's Law*, The Amer. Math. Monthly **114** (2007), no. 7, 588-601.

[4] R. W. Hamming, *On the distribution of numbers*, Bell Syst. Tech. J. **49** (1970), 1609-1625.

[5] W. Hurlimann, *Benford's Law from 1881 to 2006: a bibliography* http://arxiv.org/abs/math/0607168.

[6] D. Jang, J. U. Kang, A. Kruckman, J. Kudo and S. J. Miller, *Chains of distributions, hierarchical Bayesian models and Benford's Law*, Journal of Algebra, Number Theory: Advances and Applications **1** (2009), no. 1, 37-60, arXiv:0805.4226.

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

[8] D. S. Lemons, *On the numbers of things and the distribution of first digits*, American Journal of Physics **54** (1986), no. 9, 816-817.

[9] E. Ley, *On the peculiar distribution of the U.S. Stock Indices Digits*, The American Statistician **50** (1996), no. 4, 311-313.

[10] S. J. Miller, *Theory and Applications of Benford's Law*, (preprint, 2013).

[11] S. J. Miller and M. J. Nigrini, *The Modulo 1 Central Limit Theorem and Benford's Law for Products* International Journal of Algebra **2** (2008), no. 3, 119-130.

[12] S. Newcomb, *Note on the Frequency of Use of the Different Digits in Natural Numbers*, American Journal of Mathematics **4** (1881), no. 1, 39-40. http://www.jstor.org/stable/2369148

[13] M. Nigrini, *Digital Analysis and the Reduction of Auditor Litigation Risk*. Pages 69-81 in *Proceedings of the 1996 Deloitte & Touche/University of Kansas Symposium on Auditing Problems*, ed. M. Ettredge, University of Kansas, Lawrence, KS, 1996.

[14] M. Nigrini, *The Use of Benford's Law as an Aid in Analytical Procedures*, Auditing: A Journal of Practice & Theory **16** (1997), no. 2, 52-67.

[15] M. Nigrini and S. J. Miller, *Benford's Law applied to hydrology data - results and relevance to other geophysical data*, Mathematical Geology **39** (2007), no. 5, 469-490.