

BENFORD'S LAW AND CONTINUOUS DEPENDENT RANDOM VARIABLES

THEALEXA BECKER, TAYLOR C. CORCORAN, ALEC GREAVES-TUNNELL,
JOSEPH R. IAFRATE, JOY JING, STEVEN J. MILLER, JACLYN D. PORFILIO, RYAN RONAN,
JIRAPAT SAMRANVEDHYA, AND FREDERICK W. STRAUCH

ABSTRACT. Many mathematical, man-made and natural systems exhibit a leading-digit bias, where a first digit (base 10) of 1 occurs not 11% of the time, as one would expect if all digits were equally likely, but rather 30%. This phenomenon is known as Benford's Law. Analyzing which datasets adhere to Benford's Law and how quickly Benford behavior sets in are the two most important problems in the field. Most previous work studied systems of independent random variables, and relied on the independence in their analyses.

Inspired by natural processes such as particle decay, we study the dependent random variables that emerge from models of decomposition of conserved quantities. We prove that in many instances the distribution of lengths of the resulting pieces converges to Benford behavior as the number of divisions grow, and give several conjectures for other fragmentation processes. The main difficulty is that the resulting random variables are dependent, which we handle by a careful analysis of the dependencies and tools from Fourier analysis to obtain quantified convergence rates. Our method can be applied to many other systems; as an example, we show that the $n!$ entries in the determinant expansions of $n \times n$ matrices with entries independently drawn from nice random variables converges to Benford's Law.

CONTENTS

1. Introduction	2
1.1. Background	2
1.2. Models and Notation	3
1.3. Results	6
2. Proof of Theorem 1.5: Unrestricted Decomposition	8
2.1. Expected Value	9
2.2. Variance	10
3. Proof of Theorem 1.7: Restricted Decomposition	12
4. Proof of Theorem 1.9: Fixed Proportion Decomposition	14
4.1. Case I: $y \in \mathbb{Q}$.	14
4.2. Case II: $y \notin \mathbb{Q}$ has finite irrationality exponent	15

Date: December 27, 2013.

2010 Mathematics Subject Classification. 60A10, 11K06 (primary), (secondary) 60E10.

Key words and phrases. Benford's Law, Fourier transform, Mellin transform, dependent random variables, fragmentation.

The 1st, 2nd, 3rd, 4th, 7th, 8th and 9th named authors were supported by Williams College and NSF grant DMS0850577, the 6th named author was partially supported by NSF grants DMS0970067 and DMS1265673, and the 10th named author was partially supported by NSF grant PHY1005571 and a Research Corporation Cottrell College Science Award. We thank Don Lemons, Corey Manack, Charles Wohl, and the participants of the Williams SMALL REU (2011 and 2013) for helpful discussions, and the referee for many suggestions which improved this paper.

4.3. Case III: $y \notin \mathbb{Q}$ has infinite irrationality exponent	19
5. Proof of Theorem 1.10: Determinant Expansion	19
6. Future Work	20
Appendix A. Non-Benford Behavior	22
Appendix B. Notes on Lemons' Work	24
References	25

1. INTRODUCTION

1.1. **Background.** In 1881, American astronomer Simon Newcomb [New] noticed that the earlier pages of logarithm tables, those which corresponded to numbers with leading digit 1, were more worn than other pages. He proposed that the leading digits of certain systems are logarithmically, rather than uniformly, distributed. In 1938, Newcomb's leading digit phenomenon was popularized by physicist Frank Benford, who examined the distribution of leading digits in datasets ranging from street addresses to molecular weights. The digit bias investigated by these scientists is now known as Benford's Law.

Formally, a dataset is said to follow Benford's Law base B if the probability of observing a leading digit d base B is $\log_B(\frac{d+1}{d})$; thus we would have a leading digit of 1 base 10 approximately 30% of the time, and a leading digit of 9 less than 5% of the time. More generally, we can consider all the digits of a number. Specifically, given any $x > 0$ we can write it as

$$x = S_B(x) \cdot 10^{k_B(x)}, \quad (1.1)$$

where $S_B(x) \in [1, B)$ is the significand of x and $k_B(x)$ is an integer; note two numbers have the same leading digits if their significands agree. Benford's Law is now the statement that $\text{Prob}(S_B(x) \leq s) = \log_B(s)$.

Benford's Law arises in applied mathematics [BH1], auditing [DrNi, MN3, Nig1, Nig2, Nig3, NigMi], biology [CLTF], computer science [Knu], dynamical systems [Ber1, Ber2, BBH, BHKR, Rod], economics [Tö], geology [NM], number theory [ARS, KonMi, LS], physics [PTTV], signal processing [PHA], statistics [MN2, CLM] and voting fraud detection [Meb], to name just a few. See [BH2, Hu] for extensive bibliographies and [BH3, BH4, BH5, Dia, Hi1, Hi2, JKKKM, JR, MN1, Pin, Rai] for general surveys and explanations of the Law's prevalence, as well as the book edited by Miller [Mil], which develops much of the theory and discusses at great length applications in many fields.

One of the most important questions in the subject, as well as one of the hardest, is to determine which processes lead to Benford behavior. Many researchers [Adh, AS, Bh, JKKKM, LéV1, LéV2, MN1, Rob, Sa, Sc1, Sc2, Sc3, ST] observed that sums, products and in general arithmetic operations on random variables often lead to a new random variable whose behavior is closer to satisfying Benford's law than the inputs, though this is not always true (see [BH5]). Many of the proofs use techniques from measure theory and Fourier analysis, though in some special cases it is possible to obtain closed form expressions for the densities, which can be analyzed directly. In certain circumstances these results can be interpreted through the lens of a central limit theorem law; as we only care about the logarithms modulo 1, the Benfordness follows from convergence of this associated density to the uniform distribution.

A crucial input in many of the above papers is that the random variables are *independent*. In this paper we explore situations where there are dependencies. The dependencies we investigate are different than many others in the literature. For example, one can look at dynamical systems and iterates or powers of a given random variable, where once the initial seed is chosen the resulting process is deterministic; see for example [AS, BBH, Dia]. In our systems instead of having just one random variable we have a large number of independent random variables N generating an enormous number of dependent random variables M (often $M = 2^N$, though in one of our examples involving matrices we have $M = N!$).

Our introduction to the subject came from reading an article of Lemons [Lem] (though see the next subsection for other related problems), who studied the decomposition of a conserved quantity; for example, what happens during certain types of particle decay. As the sum of the piece sizes must equal the original length, the resulting summands are clearly dependent. While it is straightforward to show whether or not individual pieces are Benford, the difficulty is in handling all the pieces simultaneously. We comment in greater detail about Lemons' work in Appendix B.

In the next subsection we describe some of the systems we study. In analyzing these problems we develop a technique to handle certain dependencies among random variables, which we then show is applicable in other systems as well.

1.2. Models and Notation. We prove Benford behavior for three variations of a stick decomposition, quantifying when possible their rate of convergence to Benford's Law, and discuss conjectures about other possible decomposition processes. As remarked, the techniques we develop to show Benford behavior for these problems are applicable to a variety of dependent systems. As an example, we also prove that the leading digits of the $n!$ terms in the determinant expansion of a matrix whose entries are independent, identically distributed 'nice' random variables follow a Benford distribution as n tends to infinity.

There is an extensive literature on decomposition problems; we briefly comment on some other systems that have been successfully analyzed and place our work in context. Kakutani [Ka] considered the following deterministic process. Let $Q_0 = \{0, 1\}$ and given $Q_k = \{x_0 = 0, x_1, \dots, x_k = 1\}$ (where the x_i 's are in increasing order) and an $\alpha \in (0, 1)$, construct Q_{k+1} by adding points $x_i + \alpha(x_{i+1} - x_i)$ in each subinterval $[x_i, x_{i+1}]$ where $x_{i+1} - x_i = \max_{1 \leq \ell \leq k-1} |x_{\ell+1} - x_\ell|$. He proved that as $k \rightarrow \infty$, the points of Q_k become uniformly distributed, which implies that this process is non-Benford. This process has been generalized; see for example [AF, Ca, Lo, PvZ, Sl, vZ] and the references therein, and especially the book [Bert]. See also [Kol] for processes related to particle decomposition, [CaVo, Ol] for 2-dimensional examples, and [IV] for a fractal setting.

Most of this paper is devoted to the following decomposition process, whose first few levels are shown in Figure 1. Begin with a stick of length L , and a density function f on $(0, 1)$; all cuts will be drawn from this density. Cut the stick at proportion p_1 . This is the first level, and results in two sticks. We now cut the left fragment at proportion p_2 and the right at proportion p_3 . This process continues for N iterations. Thus if we start with one stick of length L , after one iteration we have sticks of length Lp_1 and $L(1 - p_1)$, after two iterations we have sticks of length Lp_1p_2 , $Lp_1(1 - p_2)$, $L(1 - p_1)p_3$, and $L(1 - p_1)(1 - p_3)$, and so on. Iterating this process N times, we are left with 2^N sticks.

We analyze whether the lengths of the resulting pieces follow Benford's Law for different choices of f , as well as modifications of the fragmentation procedure. This process builds on earlier work in the field, which we discuss after describing our systems.

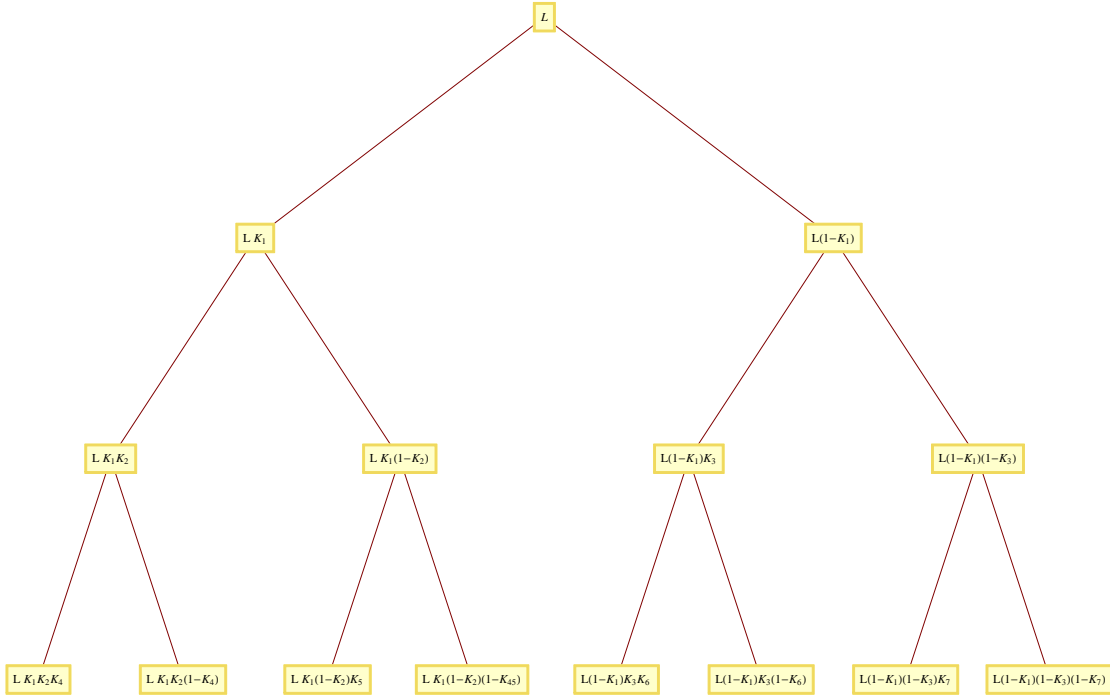


FIGURE 1. Unrestricted Decomposition: Breaking L into pieces, $N = 3$.

- (1) **Unrestricted Decomposition Model:** As described above, each proportion is drawn from a distribution f and all pieces decompose.
- (2) **Restricted Decomposition Model:** Proportions are chosen as in Case (1), but only one of the two resulting pieces from each iteration decomposes further.
- (3) **Fixed Proportion Decomposition:** All pieces decompose, but a fixed proportion p is chosen prior to the decomposition process and is used for all sticks during every iteration.

In addition to similarities with the work mentioned above, the last problem is a similar to a fragmentation tree model investigated by Janson and Neininger [JN]. Phrasing their work in our language, they randomly chose and fixed b probabilities p_1, \dots, p_b and then at each stage each piece of length x split into b pieces of length p_1x, \dots, p_bx , unless x is below a critical threshold in which case the piece is never decomposed further. They were interested in the number of pieces after their deterministic process ended, whereas we are interested in the distribution of the leading digits of the lengths. While it is possible to apply some of their results to attack our third model, the problem can be attacked directly. The situation here is similar to other problems in the field. For example, Miller and Nigrini [MN1] prove that certain products of random variables become Benford. While it is possible to prove this by invoking a central limit theorem type argument, it is not necessary as we do not need to know the distribution of the product completely, but rather we only need to know the distribution of its logarithm modulo 1. Further, by not using the central limit theorem

they are able to handle more general distributions; in particular, they can handle random variables with infinite variance.

Before we can state our results, we first introduce some notation which is needed to determine which f lead to Benford behavior.

Definition 1.1 (Mellin Transform, $\mathcal{M}_f(s)$). *Let $f(x)$ be a continuous real-valued function on $[0, \infty)$.¹ We define its Mellin transform², $\mathcal{M}_f(s)$, by*

$$\mathcal{M}_f(s) := \int_0^\infty f(x)x^s \frac{dx}{x}. \tag{1.2}$$

We next define the significant indicator function; while we work base 10, analogous definitions hold for other bases.

Definition 1.2 (Significant indicator function, φ_s). *For $s \in [1, 10)$, let*

$$\varphi_s(x) := \begin{cases} 1 & \text{if the significant of } x \text{ is at most } s \\ 0 & \text{otherwise;} \end{cases} \tag{1.3}$$

thus φ_s is the indicator function of the event of a significant at most s .

In all proofs, we label the set of stick lengths resulting from the decomposition process by $\{X_i\}$. Note that a given stick length can occur multiple times, so each element of the set $\{X_i\}$ has associated to it a frequency.

Definition 1.3 (Stick length proportions, P_N). *Given stick lengths $\{X_i\}$, the proportion whose significant is at most s , $P_N(s)$, is*

$$P_N(s) := \frac{\sum_i \varphi_s(X_i)}{\#\{X_i\}}, \tag{1.4}$$

In the Fixed Proportion Decomposition Model, we are able to quantify the rate of convergence if $\log_{10} \frac{1-p}{p}$ has finite irrationality exponent.

Definition 1.4 (Irrationality exponent). *A number α has irrationality exponent κ if κ is the supremum of all γ with*

$$\underline{\lim}_{q \rightarrow \infty} q^{\gamma+1} \min_p \left| \alpha - \frac{p}{q} \right| = 0. \tag{1.5}$$

By Roth's theorem, every algebraic irrational has irrationality exponent 1. See for example [HS, Ro] for more details.

Finally, we occasionally use big-Oh and little-oh notation. We write $f(x) = O(g(x))$ (or equivalently $f(x) \ll g(x)$) if there exists an x_0 and a $C > 0$ such that, for all $x \geq x_0$, $|f(x)| \leq Cg(x)$, while $f(x) = o(g(x))$ means $\lim_{x \rightarrow \infty} f(x)/g(x) = 0$.

¹As our random variables are proportions, for us f is always a probability density with support on the unit interval.

²Note $\mathcal{M}_f(s) = \mathbb{E}[x^{s-1}]$, and thus results about expected values translate to results on Mellin transforms; as f is a density $\mathcal{M}_f(1) = 1$. Letting $x = e^{2\pi u}$ and $s = \sigma - i\xi$ gives $\mathcal{M}_f(\sigma - i\xi) = 2\pi \int_{-\infty}^\infty (f(e^{2\pi u})e^{2\pi\sigma u}) e^{-2\pi i u \xi} du$, which is the Fourier transform of $g(u) = 2\pi f(e^{2\pi u})e^{2\pi\sigma u}$. The Mellin and Fourier transforms as thus related; this logarithmic change of variables explains why both enter into Benford's Law problems. We can therefore obtain proofs of Mellin transform properties by mimicking the proofs of the corresponding statements for the Fourier transform; see [SS1, SS2].

1.3. Results. We state our results for the fragmentation models of §1.2. While a common way of proving a sequence is Benford base B is to show that its logarithms base B are equidistributed³ (see, for example, [Dia, MT-B]), as we are using the Mellin transform and not the Fourier transform we instead often directly analyze the significant function. To show that the first digits of $\{X_i\}$ follow a Benford distribution, it suffices to show that

- (1) $\lim_{N \rightarrow \infty} \mathbb{E}[P_n(s)] = \log_{10}(s)$, and
- (2) $\lim_{N \rightarrow \infty} \text{Var}(P_n(s)) = 0$.

Viewing $P_n(s)$ as the cumulative distribution function of the process, the above shows that we have convergence in distribution⁴ to the Benford cumulative distribution function (see [GS]).

For ease of exposition and proof we often concentrate on the uniform distribution case, and remark on generalizations. In our proofs the key technical condition is that the densities satisfy (1.6) below. This is a very weak condition if the densities are fixed, essentially making sure we stay away from random variables where the logarithm modulo 1 of the densities are supported on translates of subgroups of the unit interval. If we allow the densities to vary at each stage, it is still a very weak condition but it is possible to construct a sequence of densities so that, while each one satisfies the condition, the sequence does not. We give an example in Appendix A; see [MN1] for more details.

Theorem 1.5 (Unrestricted Decomposition Model). *Fix a continuous probability density f on $(0, 1)$ such that*

$$\lim_{N \rightarrow \infty} \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} \prod_{m=1}^N \mathcal{M}_h \left(1 - \frac{2\pi i \ell}{\log 10} \right) = 0, \quad (1.6)$$

where $h(x)$ is either $f(x)$ or $f(1-x)$ (the density of $1-p$ if p has density f). Given a stick of length L , independently choose cut proportions $p_1, p_2, \dots, p_{2^N-1}$ from the unit interval according to the probability density f . After N iterations we have

$$\begin{aligned} X_1 &= L p_1 p_2 p_4 \cdots p_{2^{N-2}} p_{2^{N-1}} \\ X_2 &= L p_1 p_2 p_4 \cdots p_{2^{N-2}} (1 - p_{2^{N-1}}) \\ &\vdots \\ X_{2^{N-1}} &= L (1 - p_1) (1 - p_3) (1 - p_7) \cdots (1 - p_{2^{N-1}-1}) p_{2^{N-1}} \\ X_{2^N} &= L (1 - p_1) (1 - p_3) (1 - p_7) \cdots (1 - p_{2^{N-1}-1}) (1 - p_{2^{N-1}}), \end{aligned} \quad (1.7)$$

and

$$P_N(s) := \frac{\sum_{i=1}^{2^N} \varphi_s(X_i)}{2^N} \quad (1.8)$$

is the fraction of partition pieces X_1, \dots, X_{2^N} whose significant is less than or equal to s (see (1.3) for the definition of φ_s). Then

- (1) $\lim_{N \rightarrow \infty} \mathbb{E}[P_N(s)] = \log_{10} s$,

³A sequence $\{x_n\}$ is equidistributed modulo 1 if for any $(a, b) \subset (0, 1)$ we have $\lim_{N \rightarrow \infty} \frac{1}{N} \cdot \#\{n \leq N : x_n \in (a, b)\} = b - a$.

⁴A sequence of random variables R_1, R_2, \dots with corresponding cumulative distribution functions F_1, F_2, \dots converges in distribution to a random variable R with cumulative distribution F if $\lim_{n \rightarrow \infty} F_n(r) = F(r)$ for each r where F is continuous.

$$(2) \lim_{N \rightarrow \infty} \text{Var}(P_N(s)) = 0.$$

Thus as $N \rightarrow \infty$, the significands of the resulting stick lengths converge in distribution to Benford's Law.

Remark 1.6. *Theorem 1.5 can be greatly generalized. We assumed for simplicity that at each stage each piece must split into exactly two pieces. A simple modification of the proof shows Benford behavior is also attained in the limit if at each stage each piece independently splits into $0, 1, 2, \dots$ or k pieces with probabilities $q_0, q_1, q_2, \dots, q_k$ (so long as $q_0 < 1$). Furthermore, we do not need to use the same density for each cut, but can draw from a finite set of densities that satisfy the Mellin transform condition. Interestingly, we can construct a counter-example if we are allowed to take infinitely many distinct densities satisfying the Mellin condition; we give one in Appendix A.⁵*

Theorem 1.7 (Restricted Decomposition Model). *Consider the following decomposition process: Begin with a stick of length L , and cut this stick at a proportion p_1 chosen uniformly at random from $(0, 1)$. This results in two sticks, one length Lp_1 and one of length $L(1 - p_1)$. Do not decompose the stick of length $L(1 - p_1)$ further, but cut the other stick at proportion p_2 also chosen uniformly from the unit interval. The resulting sticks will be of lengths p_1p_2 and $p_1(1 - p_2)$. Again do not decompose the latter stick any further. Recursively repeat this process $N-1$ times, leaving N sticks:*

$$\begin{aligned} X_1 &= L(1 - p_1) \\ X_2 &= Lp_1(1 - p_2) \\ &\vdots \\ X_{N-1} &= Lp_1p_2 \cdots p_{N-2}(1 - p_{N-1}). \\ X_N &= Lp_1p_2 \cdots p_{N-1}. \end{aligned} \tag{1.9}$$

The distribution of the leading digits of these resulting N sticks converges in distribution to Benford's Law.

Remark 1.8. *We may replace the uniform distribution with any nice distribution that satisfies the Mellin transform condition of (1.6).*

Theorem 1.9 (Fixed Proportion Decomposition Model). *Choose any $p \in (0, 1)$. In Stage 1, cut a given stick at proportion p to create two pieces. In Stage 2, cut each resulting piece into two pieces at the same proportion p . Continue this process N times, generating 2^N sticks with $N + 1$ distinct lengths (assuming $p \neq 1/2$) given by*

$$\begin{aligned} x_1 &= Lp^N \\ x_2 &= Lp^{N-1}(1 - p) \\ x_3 &= Lp^{N-2}(1 - p)^2 \\ &\vdots \\ x_N &= Lp(1 - p)^{N-1} \\ x_{N+1} &= L(1 - p)^N, \end{aligned} \tag{1.10}$$

⁵The reason we need to be careful is that, while typically products of independent random variables converge to Benford behavior, there are pathological choices where this fails (see Example 2.4 of [MN1]).

where the frequency of x_n is $\binom{N}{n}/2^N$. Choose y so that $10^y = (1-p)/p$, which is the ratio of adjacent lengths (i.e., x_{i+1}/x_i). The decomposition process results in stick lengths that converge in distribution to Benford's Law if and only if $y \notin \mathbb{Q}$. If y has finite irrationality exponent, the convergence rate can be quantified in terms of the exponent, and there is a power savings.

We briefly comment on the proofs. We proceed by quantifying the dependencies between the various fragments, and showing that the number of pairs that are highly dependent is small. This technique is applicable to a variety of other systems, and we give another example below. These dependencies introduce complications which prevent us from proving our claims by directly invoking standard theorems on the Benfordness of products. For example, we cannot use the well-known fact that powers of an irrational number r are Benford to prove Theorem 1.9 because we must also take into account how many pieces we have of each fragment (equivalently, how many times we have r^m as a function of m). We provide arguments in greater detail than is needed for the proofs so that, if someone wished to isolate out rates of convergence, that could be done with little additional effort. While optimizing the errors is straightforward, doing so clutters the proof and can have computations very specific to the system studied, and thus we have chosen not to isolate the best possible error bounds in order to keep the exposition as simple as possible.

We end with the promised example of another system where our techniques are applicable. The proof, given in §5, utilizes the same techniques as that of the stick decomposition. We again have a system with a large number of independent random variables, n , leading to an enormous number of dependent random variables, $n!$.

Theorem 1.10. *Let A be an $n \times n$ matrix with independent, identically distributed entries a_{ij} drawn from a distribution X with density f . The distribution of the significands of the $n!$ terms in the determinant expansion of A converge in distribution to Benford's Law if (1.6) holds with $h = f$.*

2. PROOF OF THEOREM 1.5: UNRESTRICTED DECOMPOSITION

A crucial input in this proof is a quantified convergence of products of independent random variables to Benford behavior, with the error term depending on the Mellin transform. We use Theorem 1.1 of [JKKKM] (and its generalization, given in Remark 2.3 there); for the convenience of the reader we quickly review this result and its proof in Appendix A of [B-]. The dependencies of the pieces is a major obstruction; we surmount this by breaking the pairs into groups depending on how dependent they are (specifically, how many cut proportions they share).

Remark 2.1. *The key condition in Theorem 1.5, 1.6, is extremely weak and is met by most distributions. For example, if f is the uniform density on $(0, 1)$ then*

$$\mathcal{M}_f \left(1 - \frac{2\pi i \ell}{\log 10} \right) = \left(1 - \frac{2\pi i \ell}{\log 10} \right)^{-1}, \quad (2.1)$$

which implies

$$\lim_{N \rightarrow \infty} \left| \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} \prod_{m=1}^N \mathcal{M}_f \left(1 - \frac{2\pi i \ell}{\log 10} \right) \right| \leq 2 \lim_{N \rightarrow \infty} \sum_{\ell=1}^{\infty} \left| 1 - \frac{2\pi i \ell}{\log 10} \right|^{-N} = 0 \quad (2.2)$$

(we wrote the condition as $\prod_{m=1}^N \mathcal{M}_f$ instead of \mathcal{M}_f^N to highlight where the changes would surface if we allowed different densities for different cuts). While this condition is weak, it is absolutely necessary to ensure convergence to Benford behavior; see Appendix A.

To prove convergence in distribution to Benford's Law, we first prove in §2.1 that $\mathbb{E}[P_N(s)] = \log_{10} s$, and then in §2.2 prove that $\text{Var}(P_N(s)) \rightarrow 0$; as remarked earlier these two results yield the desired convergence. The proof of the mean is significantly easier than the proof of the variance as expectation is linear, and thus there are no issues from the dependencies in the first calculation, but there are in the second.

2.1. Expected Value.

Proof of Theorem 1.5 (Expected Value). By linearity of expectation,

$$\mathbb{E}[P_N(s)] = \mathbb{E}\left[\frac{\sum_{i=1}^{2^N} \varphi_s(X_i)}{2^N}\right] = \frac{1}{2^N} \sum_{i=1}^{2^N} \mathbb{E}[\varphi_s(X_i)]. \quad (2.3)$$

We recall that all pieces can be expressed as the product of the starting length L and cutting proportions p_i . While there are dependencies among the lengths X_i , there are no dependencies among the p_i 's. A given stick length X_i is determined by some number of factors k of p_i and $N - k$ factors of $1 - p_i$ (where p_i is a cutting proportion between 0 and 1 drawn from a distribution with density f). By relabeling if necessary, we may assume

$$X_i = L p_1 p_2 \cdots p_k (1 - p_{k+1}) \cdots (1 - p_N); \quad (2.4)$$

the first k proportions are drawn from a distribution with density $f(x)$ and the last $N - k$ from a distribution with density $f(1 - x)$.

The proof is completed by showing $\lim_{N \rightarrow \infty} \mathbb{E}[\varphi_s(X_i)] = \log_{10} s$. We have

$$\begin{aligned} \mathbb{E}[\varphi_s(X_i)] &= \int_{p_1=0}^1 \int_{p_2=0}^1 \cdots \int_{p_N=0}^1 \varphi_s \left(L \prod_{r=1}^k p_r \prod_{m=k+1}^N (1 - p_m) \right) \\ &\quad \cdot \prod_{r=1}^k f(p_r) \prod_{m=k+1}^N f(1 - p_m) dp_1 dp_2 \cdots dp_N. \end{aligned} \quad (2.5)$$

This is equivalent to studying the distribution of a product of N independent random variables (chosen from one of two densities) and then rescaling the result by L . The convergence of $L \prod_{r=1}^k p_r \prod_{m=k+1}^N (1 - p_m)$ to Benford follows from [JKKKM] (the key theorem is summarized for the reader's convenience in Appendix A in [B-], the expanded arXiv version of this paper). We find $\mathbb{E}[\varphi_s(X_i)]$ equals $\log_{10} s$ plus a rapidly decaying N -dependent error term. This is because the Mellin transforms (with $\ell \neq 0$) are always less than 1 in absolute value. Thus the error is bounded by the maximum of the error from a product with $N/2$ terms with density $f(x)$ or a product with $N/2$ terms with density $f(1 - x)$ (where the existence of $N/2$ such terms follows from the pigeonhole principle). Thus $\lim_{N \rightarrow \infty} \mathbb{E}[P_N(s)] = \log_{10} s$, completing the proof. \square

Remark 2.2. For specific choices of f we can obtain precise bounds on the error. For example, if each cut is chosen uniformly on $(0, 1)$, then the densities of the distributions of the p_i 's and the $(1 - p_i)$'s are the same. By [MN1] or Corollary A.2 of [B-],

$$\mathbb{E}[\varphi_s(X_i)] - \log_{10} s \ll \frac{1}{2.9^N}, \quad (2.6)$$

and thus

$$\mathbb{E}[P_N(s)] - \log_{10} s \ll \frac{1}{2^N} \sum_{i=1}^{2^N} \frac{1}{2.9^N} = \frac{1}{2.9^N}. \quad (2.7)$$

2.2. Variance.

Proof of Theorem 1.5 (Variance). For ease of exposition we assume all the cuts are drawn from the uniform distribution on $(0, 1)$. To facilitate the minor changes needed for the general case, we argue as generally as possible for as long as possible.

We begin by noting that since $\varphi_s(X_i)$ is either 0 or 1, $\varphi_s(X_i)^2 = \varphi_s(X_i)$. From this observation, the definition of variance and the linearity of the expectation, we have

$$\begin{aligned} \text{Var}(P_N(s)) &= \mathbb{E}[P_N(s)^2] - \mathbb{E}[P_N(s)]^2 \\ &= \mathbb{E} \left[\left(\frac{\sum_{i=1}^{2^N} \varphi_s(X_i)}{2^N} \right)^2 \right] - \mathbb{E}[P_N(s)]^2 \\ &= \mathbb{E} \left[\frac{\sum_{i=1}^{2^N} \varphi_s(X_i)^2}{2^{2N}} + \sum_{\substack{i,j=1 \\ i \neq j}}^{2^N} \frac{\varphi_s(X_i)\varphi_s(X_j)}{2^{2N}} \right] - \mathbb{E}[P_N(s)]^2 \\ &= \frac{1}{2^N} \mathbb{E}[P_N(s)] + \frac{1}{2^{2N}} \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{2^N} \mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)] \right) - \mathbb{E}[P_N(s)]^2. \end{aligned} \quad (2.8)$$

From §2.1, $\mathbb{E}[P_N(s)] = \log_{10} s + o(1)$. Thus

$$\text{Var}(P_N(s)) = \frac{1}{2^{2N}} \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{2^N} \mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)] \right) - \log_{10}^2 s + o(1). \quad (2.9)$$

The problem is now reduced 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 partition the pairs (X_i, X_j) based on how ‘close’ X_j is to X_i in our tree (see Figure 1). We do this as follows. Recall that each of the 2^N pieces is a product of the starting length L and N cutting proportions. Note X_i and X_j must share some number of these proportions, say k terms. Then one piece has the factor p_{k+1} in its product, while the other contains the factor $(1 - p_{k+1})$. The remaining $N - k - 1$ elements in each product are independent from each other. After re-labeling, we can thus express any (X_i, X_j) pair as

$$\begin{aligned} X_i &= L \cdot p_1 \cdot p_2 \cdots p_k \cdot p_{k+1} \cdot p_{k+2} \cdots p_N \\ X_j &= L \cdot p_1 \cdot p_2 \cdots p_k \cdot (1 - p_{k+1}) \cdot \tilde{p}_{k+2} \cdots \tilde{p}_N. \end{aligned} \quad (2.10)$$

With these definitions in mind, we have

$$\begin{aligned}
 \mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)] &= \int_{p_1=0}^1 \int_{p_2=0}^1 \cdots \int_{p_N=0}^1 \int_{\tilde{p}_{k+2}=0}^1 \cdots \int_{\tilde{p}_N=0}^1 \varphi_s \left(L \prod_{r=1}^{k+1} p_r \prod_{r=k+2}^N p_r \right) \\
 &\quad \cdot \varphi_s \left(L \prod_{r=1}^k p_r \cdot (1 - p_{k+1}) \cdot \prod_{r=k+2}^N \tilde{p}_r \right) \\
 &\quad \cdot \prod_{r=1}^N f(p_r) \prod_{r=k+2}^N f(1 - \tilde{p}_r) dp_1 dp_2 \cdots dp_N d\tilde{p}_{k+2} \cdots d\tilde{p}_N. \quad (2.11)
 \end{aligned}$$

The difficulty in understanding (2.11) is that many variables occur in both $\varphi_s(X_i)$ and $\varphi_s(X_j)$. The key observation is that most of the time there are many variables occurring in one but not the other, which minimizes the effects of the common variables and essentially leads to evaluating φ_s at almost independent arguments. We make this precise below, keeping track of the errors. Define

$$L_1 := L \left(\prod_{r=1}^k p_r \right) p_{k+1}, \quad L_2 := L \left(\prod_{r=1}^k p_r \right) (1 - p_{k+1}), \quad (2.12)$$

and consider the following integrals:

$$\begin{aligned}
 I(L_1) &:= \int_{p_{k+2}=0}^1 \cdots \int_{p_N=0}^1 \varphi_s \left(L_1 \prod_{r=k+2}^N p_r \right) \prod_{r=k+2}^N f(p_r) dp_{k+2} dp_{k+3} \cdots dp_N \\
 J(L_2) &:= \int_{\tilde{p}_{k+2}=0}^1 \cdots \int_{\tilde{p}_N=0}^1 \varphi_s \left(L_2 \prod_{r=k+2}^N \tilde{p}_r \right) \prod_{r=k+2}^N f(\tilde{p}_r) d\tilde{p}_{k+2} d\tilde{p}_{k+3} \cdots d\tilde{p}_N. \quad (2.13)
 \end{aligned}$$

We show that, for any L_1, L_2 , we have $|I(L_1)J(L_2) - (\log_{10} s)^2| = o(1)$. Once we have this, then all that remains is to integrate $I(L_1)J(L_2)$ over the remaining $k+1$ variables. The rest of the proof follows from counting, for a given X_i , how many X_j 's lead to a given k .

It is at this point where we require the assumption about $f(x)$ from the statement of the theorem, namely that $f(x)$ and $f(1-x)$ satisfy (1.6). For illustrative purposes, we assume that each cut p is drawn from a uniform distribution, meaning $f(x)$ and $f(1-x)$ are the probability density functions associated with the uniform distribution on $(0, 1)$. The argument can readily be generalized to other distributions; we choose to highlight the uniform case as it is simpler, important, and we can obtain a very explicit, good bound on the error.

Both $I(L_1)$ and $J(L_2)$ involve integrals over $N - k - 1$ variables; we set $n := N - k - 1$. For the case of a uniform distribution, equation (3.7) of [JKKKM] (or see Corollary A.2 in [B-]) gives for $n \geq 4$ that⁶

$$|I(L_1) - \log_{10} s| < \left(\frac{1}{2.9^n} + \frac{\zeta(n) - 1}{2.7^n} \right) 2 \log_{10} s, \quad (2.14)$$

where $\zeta(s)$ is the Riemann zeta function, which for $\text{Re}(s) > 1$ equals $\sum_{n=1}^{\infty} 1/n^s$. Note that for all choices of L_1 , $I(L_1) \in [0, 1)$, and for $n \leq 4$ we may simply bound the difference by

⁶Our situation is slightly different as we multiply the product by L_1 ; however, all this does is translate the distribution of the logarithms by a fixed amount, and hence the error bounds are preserved.

1. It is also important to note that for $n > 1$, $\zeta(n) - 1$ is $O(1/2^n)$, and thus the error term decays very rapidly.

A similar bound exists for $J(L_2)$, and we can choose a constant C such that

$$|I(L_1) - \log_{10} s| \leq \frac{C}{2.9^n}, \quad |J(L_2) - \log_{10} s| \leq \frac{C}{2.9^n} \quad (2.15)$$

for all n, L_1, L_2 . Because of this rapid decay, by the triangle inequality it follows that

$$|I(L_1) \cdot J(L_2) - (\log_{10} s)^2| \leq \frac{2C}{2.9^n}. \quad (2.16)$$

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 $2^{n-1}I(L_1) \cdot J(L_2)$ over all n and all i :

$$\left| \sum_{\substack{i,j=1 \\ i \neq j}}^{2^N} (\mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)] - \log_{10}^2 s) \right| \leq \sum_{i=1}^{2^N} \sum_{n=1}^N 2^{n-1} \frac{2C}{2.9^n} \leq 2^N \cdot 4C. \quad (2.17)$$

Substituting this into equation (2.9) yields

$$\text{Var}(P_N(s)) \leq \frac{4C}{2^N} + o(1). \quad (2.18)$$

Since the variance must be non-negative by definition, it follows that $\lim_{N \rightarrow \infty} \text{Var}(P_N(s)) = 0$, completing the proof if each cut is drawn from a uniform distribution. The more general case follows analogously, appealing to [MN1] (or Theorem A.1 of [B-]). \square

3. PROOF OF THEOREM 1.7: RESTRICTED DECOMPOSITION

As the proof is similar to that of Theorem 1.5, we concentrate on the differences.

Proof of Theorem 1.7. We may assume that $L = 1$ as scaling does not affect Benford behavior. In the analysis below we may ignore the contributions of X_i for $i \in [1, \log N]$ and all pairs of X_i, X_j such that X_i and X_j do not differ by at least $\log N$ proportions. Removing these terms does not affect whether or not the resulting stick lengths converge to Benford (because $\frac{\log N}{N} \rightarrow 0$ when N gets large), but does eliminate strong dependencies or cases with very few products, both of which complicate our analysis.

As before, to prove that the stick lengths tend to Benford as $N \rightarrow \infty$ we show that $\mathbb{E}[P_s(N)] \rightarrow \log_{10} s$ and $\text{Var}(P_s(N)) \rightarrow 0$. The first follows identically as in §2.1:

$$\begin{aligned} \mathbb{E}[\varphi_s(x_i)] &= \int_{p_1=0}^1 \int_{p_2=0}^1 \cdots \int_{p_i=0}^1 \varphi \left((1-p_i) \prod_{l=1}^{i-1} p_l \right) \\ &\quad f(1-p_i) \prod_{l=1}^{i-1} f(p_l) dp_1 dp_2 \cdots dp_i \end{aligned} \quad (3.1)$$

tends to $\log_{10} s + o(1)$ as $n \rightarrow \infty$.

For the variance, we now have N and not 2^N pieces, and find

$$\text{Var}(P_N(s)) = \frac{\log_{10} s}{N^2} + \frac{1}{N^2} \left(\sum_{\substack{i,j=1 \\ i \neq j}}^N \mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)] \right) - \log_{10}^2 s + o(1). \quad (3.2)$$

Note that we may replace the above sum with twice the sum over $i < j$. Further, let \mathcal{A} be the set

$$\mathcal{A} := \{(i, j) : \log N \leq i \leq j - \log N \leq N - \log N\}. \quad (3.3)$$

We may replace the sum in (3.2) with twice the sum over pairs in \mathcal{A} , as the contribution from the other pairs is $o(1)$. The analysis is thus reduced to bounding

$$2 \sum_{(i,j) \in \mathcal{A}} \mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)], \quad (3.4)$$

where

$$\begin{aligned} X_i &= p_i \prod_{r=1}^{i-1} (1 - p_r) \\ X_j &= p_j \prod_{r=1}^{i-1} (1 - p_r) \prod_{r=i}^{j-1} (1 - p_r); \end{aligned} \quad (3.5)$$

we write X_i and X_j in this manner to highlight the terms they have in common. Letting $f(x)$ be the density for the uniform distribution on the unit interval, we have

$$\begin{aligned} \mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)] &= \int_{p_1=0}^1 \cdots \int_{p_i=0}^1 \cdots \int_{p_j=0}^1 \varphi_s \left(p_i \prod_{r=1}^{i-1} (1 - p_r) \right) \\ &\quad \cdot \varphi_s \left(p_j \prod_{r=1}^{i-1} (1 - p_r) \prod_{r=i}^{j-1} (1 - p_r) \right) \cdot \prod_{r=1}^j f(p_r) dp_1 \cdots dp_j. \end{aligned} \quad (3.6)$$

The analysis of this integral is similar to that in the previous section. Let

$$\mathcal{L}_i := \prod_{r=1}^{i-1} (1 - p_r), \quad \mathcal{L}_j = \prod_{r=i+1}^{j-1} (1 - p_r). \quad (3.7)$$

That is, \mathcal{L}_i consists of the terms shared by X_i and X_j , and \mathcal{L}_j is the product of the terms only in X_j . We are left with showing that the integral

$$\int_{p_1=0}^{\infty} \cdots \int_{p_j=0}^{\infty} \varphi_s(\mathcal{L}_i p_i) \varphi_s(\mathcal{L}_i (1 - p_i) \mathcal{L}_j p_j) \cdot \prod_{r=1}^j f(p_r) dp_1 \cdots dp_j \quad (3.8)$$

is close to $\log_{10}^2 s$.

We highlight the differences from the previous section. The complication is that here \mathcal{L}_1 appears in both arguments, while before it only occurred once. This is why we restricted our pairs (i, j) to lie in \mathcal{A} . Since we assume $i \geq \log N$, there are a lot of terms in the product of \mathcal{L}_1 , and by the results of [JKKKM] (or see Appendix A of [B-]) the distribution of \mathcal{L}_1 converges to Benford. Similarly, there are at least $\log N$ new terms in the product for \mathcal{L}_2 , and thus $\mathcal{L}_i(1 - p_i)\mathcal{L}_j p_j$ converges to Benford. An analysis similar to that in §2 shows that the

integral is close to $\log_{10}^2 s$ as desired. The proof is completed by noting that the cardinality of \mathcal{A} is $N^2/2 + O(N \log N)$. Substituting our results into (3.2), we see the variance tends to 0. Thus the distribution of the leading digits converges in distribution to Benford's Law. \square

4. PROOF OF THEOREM 1.9: FIXED PROPORTION DECOMPOSITION

Recall that we are studying the distribution of the stick lengths that result from cutting a stick at a fixed proportion p . We define y by $10^y := \frac{1-p}{p}$, the ratio between adjacent piece lengths. The resulting behavior is controlled by the rationality of y .

4.1. **Case I:** $y \in \mathbb{Q}$. Let $y = r/q$. Here $r \in \mathbb{Z}, q \in \mathbb{N}$ and $\gcd(r, q) = 1$. Let $S_{10}(x_j)$ denote the first digit of x_j . As

$$x_{j+q} = \left(\frac{1-p}{p}\right)^q x_j = (10^y)^q x_j = 10^r x_j, \quad (4.1)$$

it follows that

$$S_{10}(x_{j+q}) = S_{10}(x_j). \quad (4.2)$$

Thus the significand of x_j repeats every q indices.⁷ We now show that the q different classes of leading digits occur equally often as $N \rightarrow \infty$.

To do this, we use the multisection formula. Given a power series $f(x) = \sum_{k=0}^{\infty} a_k x^k$, we can take the multisection $\sum_{\ell=0}^{\infty} a_{\ell q+j} x^{\ell q+j}$, where j and q are integers with $0 \leq j < q$. The multisection itself is a power series that picks out every q^{th} term from the original series, starting from the j^{th} term. We have a closed expression for the multisection in terms of the original function (for a proof of this formula, see [Che]):

$$\sum_{\ell=0}^{\infty} a_{\ell q+j} x^{\ell q+j} = \frac{1}{q} \sum_{s=0}^{q-1} \omega^{-js} f(\omega^s x), \quad (4.3)$$

where $\omega = e^{2\pi i/q}$ is the primitive q^{th} root of unity. We apply this to $f(x) = (1+x)^N = \sum_{k=0}^N \binom{N}{k} x^k$. To extract the sum of equally spaced binomial coefficients, we take the multisection of the binomial theorem with $x = 1$:

$$\sum_{\ell} \binom{N}{\ell q+j} = \frac{2^N}{q} \sum_{s=0}^{q-1} \left(\cos \frac{\pi s}{q}\right)^N \cos \frac{\pi(N-2j)s}{q}; \quad (4.4)$$

note in the algebraic simplifications we took the real part of $\omega^{(N-2j)s/2}$, which is permissible as the left hand side is real and therefore the imaginary part sums to zero.

⁷We are interested in determining the frequency with which each leading digit occurs. It is possible that two sticks x_j and x_i are not a multiple of q indices apart but still have the same leading digit. Thus summing the frequency of every q^{th} length tells us that for each digit d the probability of a first digit d is a/q for some $a \in \mathbb{N}$.

All terms with index $j \bmod q$ share the same leading digit. Therefore the probability of observing a term with index $j \bmod q$ is given by

$$\begin{aligned} \frac{1}{2^N} \left[\binom{N}{j} + \binom{N}{j+q} + \binom{N}{j+2q} + \cdots \right] &= \frac{1}{q} \sum_{s=0}^{q-1} \left(\cos \frac{\pi s}{q} \right)^N \cos \frac{\pi(N-2j)s}{q} \\ &= \frac{1}{q} \left(1 + \sum_{s=1}^{q-1} \left(\cos \frac{\pi s}{q} \right)^N \cos \frac{\pi(N-2j)s}{q} \right) \\ &= \frac{1}{q} \left(1 + \text{Err} \left[(q-1) \left(\cos \frac{\pi}{q} \right)^N \right] \right), \quad (4.5) \end{aligned}$$

where $\text{Err}[X]$ indicates an absolute error of size at most X . When $q = 1$, the term inside the Err vanishes. For $q \in \mathbb{N}$, $q > 1$, $\cos(\pi/q) \in [0, 1)$; as that value is raised to the N^{th} power, it approaches 0 exponentially fast. As $N \rightarrow \infty$, the term inside the Err disappears, leaving us $1/q$. Hence the probability of observing a particular leading digit converges to a multiple of $1/q$, which is a rational number. On the other hand, the probability from the Benford distribution is $\log_{10}(1 + 1/d)$ which is an irrational number. Therefore the described cutting process does not result in perfect Benford behavior. \square

Remark 4.1. *Instead of using the multisection formula, we could use the monotonicity (as we move towards the middle) to show that the different classes of $j \bmod q$ have approximately the same probability by adding or removing the first and/or last term in the sequence, which changes which class dominates the other. We chose this approach as the multisection formula is useful in the proof of Theorem 1.9 when the irrationality exponent of y is finite.*

4.2. Case II: $y \notin \mathbb{Q}$ has finite irrationality exponent. We prove the leading digits of the 2^N stick lengths are Benford by showing that the logarithms of the piece lengths are equidistributed modulo 1 (Benford's Law then follows by simple exponentiation; see [Dia, MT-B]). The frequency of the lengths x_j follow a binomial distribution with mean $N/2$ and standard deviation $\sqrt{N}/2$. As the argument is long we briefly outline it. First we show that the contributions from the tails of the binomial distribution are small. We then break the binomial distribution into intervals that are a power of N smaller than the standard deviation, and show both that the probability density function does not change much in each interval and that the logarithms of the lengths in each interval are equidistributed modulo 1.

Specifically, choose a $\delta \in (0, 1/2)$; the actual value depends on optimizing various errors. Note that $N^\delta \ll \sqrt{N}/2$, the standard deviation. Let

$$x_\ell := \frac{N}{2} + \ell N^\delta, \quad x_{\ell,i} = \frac{N}{2} + \ell N^\delta + i, \quad I_\ell := \{x_\ell, x_\ell + 1, \dots, x_\ell + N^\delta - 1\}. \quad (4.6)$$

There are $N/N^\delta = N^{1-\delta}$ such intervals. By symmetry, it suffices to just study the right half of the binomial.

4.2.1. Truncation. Instead of considering the entire binomial distribution, for any $\epsilon > 0$ we may truncate the distribution and examine only the portion that is within N^ϵ standard deviations of the mean. Recall that we are only considering the right half of the binomial as well.

For $\epsilon > 0$, Chebyshev's Inequality⁸ gives that the proportion of the density that is beyond N^ϵ standard deviations of the mean is

$$\text{Prob} \left(\left| x - \frac{N}{2} \right| \geq N^\epsilon N^{1/2}/2 \right) \leq \frac{1}{N^{2\epsilon}}. \quad (4.7)$$

As N tends to infinity this probability becomes negligible, and thus we are justified in only considering the portion of the binomial from $\frac{N}{2} - N^{\frac{1}{2}+\epsilon}$ to $\frac{N}{2} + N^{\frac{1}{2}+\epsilon}$. Thus ℓ ranges from $-N^{\frac{1}{2}-\delta+\epsilon}$ to $N^{\frac{1}{2}-\delta+\epsilon}$.

4.2.2. Roughly Equal Probability Within Intervals. Let $x_\ell = N/2 + \ell N^\delta$. Consider the difference in the binomial coefficients of adjacent intervals, which is related to the difference in probabilities by a factor of $1/2^N$. Note that this is a bound for the maximum change in probabilities in an interval of length N^δ away from the tails of the distribution. For future summation, we want to relate the difference to a small multiple of either endpoint probability; it is this restriction that necessitated the truncation from the previous subsection. Without loss of generality we may assume $\ell \geq 0$ and we find

$$\begin{aligned} \binom{N}{x_\ell} - \binom{N}{x_{\ell+1}} &= \binom{N}{\frac{N}{2} + \ell N^\delta} - \binom{N}{\frac{N}{2} + (\ell + 1)N^\delta} \\ &= \frac{N!}{\left(\frac{N}{2} + \ell N^\delta\right)! \left(\frac{N}{2} - \ell N^\delta\right)!} - \frac{N!}{\left(\frac{N}{2} + (\ell + 1)N^\delta\right)! \left(\frac{N}{2} - (\ell + 1)N^\delta\right)!} \\ &= \frac{N! \left(\frac{N}{2} + (\ell + 1)N^\delta\right)! \left(\frac{N}{2} - (\ell + 1)N^\delta\right)! - N! \left(\frac{N}{2} + \ell N^\delta\right)! \left(\frac{N}{2} - \ell N^\delta\right)!}{\left(\frac{N}{2} + \ell N^\delta\right)! \left(\frac{N}{2} - \ell N^\delta\right)! \left(\frac{N}{2} + (\ell + 1)N^\delta\right)! \left(\frac{N}{2} - (\ell + 1)N^\delta\right)!} \\ &= \binom{N}{\frac{N}{2} + \ell N^\delta} \left[1 - \frac{\left(\frac{N}{2} + \ell N^\delta\right)! \left(\frac{N}{2} - \ell N^\delta\right)!}{\left(\frac{N}{2} + (\ell + 1)N^\delta\right)! \left(\frac{N}{2} - (\ell + 1)N^\delta\right)!} \right] \\ &= \binom{N}{x_\ell} \left[1 - \frac{\left(\frac{N}{2} + \ell N^\delta\right)! \left(\frac{N}{2} - \ell N^\delta\right)!}{\left(\frac{N}{2} + (\ell + 1)N^\delta\right)! \left(\frac{N}{2} - (\ell + 1)N^\delta\right)!} \right]. \end{aligned} \quad (4.8)$$

Notice here that the difference in binomial coefficients is in terms of the probability at the left endpoint of the interval, which allows us to express the difference in probabilities relative to the probability within an interval. Let $\alpha_{\ell;N} = \frac{\left(\frac{N}{2} + \ell N^\delta\right)! \left(\frac{N}{2} - \ell N^\delta\right)!}{\left(\frac{N}{2} + (\ell + 1)N^\delta\right)! \left(\frac{N}{2} - (\ell + 1)N^\delta\right)!}$. We show that $1 - \alpha_{\ell;N} \rightarrow 0$, which implies the probabilities do not change significantly over an interval. We have

⁸While we could get better bounds by appealing to the Central Limit Theorem, Chebyshev's Inequality suffices.

$$\begin{aligned}
 \alpha_{\ell;N} &= \frac{(\frac{N}{2} + \ell N^\delta)! (\frac{N}{2} - \ell N^\delta)!}{(\frac{N}{2} + (\ell + 1)N^\delta)! (\frac{N}{2} - (\ell + 1)N^\delta)!} \\
 &\geq \frac{(\frac{N}{2} - (\ell + 1)N^\delta)^{N^\delta}}{(\frac{N}{2} + (\ell + 1)N^\delta)^{N^\delta}} \\
 &= \left(\frac{1 - \frac{2(\ell+1)}{N^{1-\delta}}}{1 + \frac{2(\ell+1)}{N^{1-\delta}}} \right)^{N^\delta} \\
 \log \alpha_{\ell;N} &\geq N^\delta \left[\log \left(1 - \frac{2(\ell + 1)}{N^{1-\delta}} \right) - \log \left(1 + \frac{2(\ell + 1)}{N^{1-\delta}} \right) \right]. \tag{4.9}
 \end{aligned}$$

From Taylor expanding we know $\log(1 + u) = -u + u^2/2 - O(u^3)$. Thus letting $u = 2(\ell + 1)/N^{1-\delta}$ (which is much less than 1 for N large as $\ell \leq N^{\frac{1}{2}-\delta+\epsilon}$) in the difference of logarithms above we see the linear terms reinforce and the quadratic terms cancel, and thus the error is of size $O(u^3) = O(\ell^3/N^{3-3\delta}) = O(N^{3\epsilon-3/2})$. Therefore

$$\begin{aligned}
 \log \alpha_{\ell;N} &\geq N^\delta \left[-\frac{4(\ell + 1)}{N^{1-\delta}} + O(N^{3\epsilon-3/2}) \right] \\
 \alpha_{\ell;N} &\geq e^{-4(\ell+1)N^{2\delta-1} + O(N^{\delta+3\epsilon-3/2})} \\
 &\geq 1 - 4(\ell + 1)N^{2\delta-1} + O(N^{\delta+3\epsilon-3/2}) \\
 1 - \alpha_{\ell;N} &\leq 4(\ell + 1)N^{2\delta-1} + O(N^{\delta+3\epsilon-3/2}). \tag{4.10}
 \end{aligned}$$

Since we have truncated and $\ell \leq N^{\frac{1}{2}-\delta+\epsilon}$, this implies $(\ell + 1)N^{2\delta-1} \ll N^{\delta+\epsilon-\frac{1}{2}}$, which tends to zero if $\delta < 1/2 - \epsilon$. Substituting (4.10) into (4.8) yields

$$\begin{aligned}
 \binom{N}{x_\ell} - \binom{N}{x_{\ell+1}} &= \binom{N}{\frac{N}{2} + \ell N^\delta} (1 - \alpha_{\ell;N}) \\
 &\leq \binom{N}{x_\ell} (4(\ell + 1)N^{2\delta-1} + O(N^{\delta+3\epsilon-3/2})). \tag{4.11}
 \end{aligned}$$

Since $\ell \leq N^{1/2-\delta+\epsilon}$, it follows that

$$\left| \binom{N}{x_\ell} - \binom{N}{x_{\ell+1}} \right| \ll \binom{N}{x_\ell} \left(N^{-\frac{1}{2}+\delta+\epsilon} + O(N^{\delta+3\epsilon-3/2}) \right). \tag{4.12}$$

As $\delta < 1/2 - \epsilon$, $O(N^{\delta+3\epsilon-3/2})$ is dominated by $N^{-\frac{1}{2}+\delta+\epsilon}$ since ϵ is small. We have proved

Lemma 4.2. *Let $\delta \in (0, 1/2 - \epsilon)$ and $\ell \leq N^{\frac{1}{2}-\delta+\epsilon}$. Then for any $i \in \{0, 1, \dots, N^\delta\}$ we have*

$$\left| \binom{N}{x_\ell} - \binom{N}{x_{\ell,i}} \right| \ll \binom{N}{x_\ell} N^{-\frac{1}{2}+\delta+\epsilon}. \tag{4.13}$$

4.2.3. Equidistribution. We first do a general analysis of equidistribution of a sequence related to our original one, and then show how this proves our main result.

Given an interval

$$I_\ell := \left\{ x_{\ell,i} : \frac{N}{2} + \ell N^\delta, \dots, \frac{N}{2} + (\ell + 1)N^\delta - 1 \right\}, \tag{4.14}$$

we prove that $\log(x_{\ell,i})$ becomes equidistributed modulo 1 as $N \rightarrow \infty$. Fix an $(a, b) \subset (0, 1)$. Let $J_\ell(a, b) \subset \{0, 1, \dots, N^\delta\}$ be the set of all $i \in I_\ell$ such that $\log(x_{\ell,i}) \bmod 1 \in (a, b)$; we want to show its measure is $(b - a)N^\delta$ plus a small error. As the $x_{\ell,i}$ form a geometric progression with common ratio $r = \frac{1-p}{p} = 10^y$, their logarithms are equidistributed modulo 1 if and only if $\log r = y$ is irrational (see [Dia, MT-B]). Moreover, we can quantify the rate of equidistribution if the irrationality exponent κ of y is finite. From [KN] we obtain a power savings:

$$|J_\ell(a, b)| = (b - a)N^\delta + O\left(N^{\delta(1-\frac{1}{\kappa}+\epsilon')}\right); \quad (4.15)$$

see [KonMi] for other examples of systems (such as the $3x + 1$ map) where the irrationality exponent controls Benford behavior.

We now combine this quantified equidistribution with Lemma 4.2 and find (we divide by 2^N later, which converts these sums to probabilities)

$$\begin{aligned} \sum_{i \in J_\ell(a, b)} \binom{N}{x_{\ell,i}} &= \sum_{i \in J_\ell(a, b)} \left[\binom{N}{x_\ell} + O\left(\binom{N}{x_\ell} N^{-\frac{1}{2}+\delta+\epsilon}\right) \right] \\ &= \binom{N}{x_\ell} \left[\left(\sum_{i \in J_\ell(a, b)} 1 \right) + O\left(N^{-\frac{1}{2}+\delta+\epsilon} N^\delta\right) \right] \\ &= \binom{N}{x_\ell} \left[\left((b - a)N^\delta + O\left(N^{\delta(1-\frac{1}{\kappa}+\epsilon')}\right) \right) + O\left(N^{-\frac{1}{2}+2\delta+\epsilon}\right) \right] \\ &= (b - a)N^\delta \binom{N}{x_\ell} + N^\delta \binom{N}{x_\ell} \cdot O\left(N^{-\frac{1}{2}+\delta+\epsilon} + N^{-\delta(\frac{1}{\kappa}-\epsilon')}\right). \end{aligned} \quad (4.16)$$

Notice the error term above is a power smaller than the main term. If we show the sum over ℓ of the main term is $1 + o(1)$, then the sum over ℓ of the error term is $o(1)$ and does not contribute in the limit (it will contribute $N^{-\eta}$ for some $\eta > 0$).

As $x_\ell = \frac{N}{2} + \ell N^\delta$, we use the multisection formula (see (4.5)) with $q = N^\delta$, and find

$$\sum_{\ell=-N/N^\delta}^{N/N^\delta} \binom{N}{x_\ell} = \frac{2^N}{N^\delta} \left(1 + \text{Err} \left[N^\delta \left(\cos \frac{\pi}{N^\delta} \right)^N \right] \right) = \frac{2^N}{N^\delta} + O\left(2^N \cdot e^{-3N^{1-2\delta}}\right), \quad (4.17)$$

where $\text{Err}[X]$ indicates an absolute error of size at most X and the simplification of the error comes from Taylor expanding the cosine and standard analysis:

$$\begin{aligned} \log \left(\cos(\pi/N^\delta) \right)^N &= N \log \left(1 - \frac{1}{2!} \frac{\pi^2}{N^{2\delta}} + O(N^{-4\delta}) \right) \\ &= N \left(-\frac{1}{2!} \frac{\pi^2}{N^{2\delta}} + O(N^{-4\delta}) \right) = -\frac{\pi^2}{2} N^{1-2\delta} + O(N^{1-4\delta}) \\ \cos(\pi/N^\delta) &\leq e^{-3N^{1-2\delta}} \end{aligned} \quad (4.18)$$

for N large.

Remember, though, that we are only supposed to sum over $|\ell| \leq N^{\frac{1}{2}-\delta+\epsilon}$. The contribution from the larger $|\ell|$, however, was shown to be at most $O(N^{-2\epsilon})$ in §4.2.1, and thus we find

$$\frac{1}{2^N} \sum_{\ell=-N^{\frac{1}{2}-\delta+\epsilon}}^{N^{\frac{1}{2}-\delta+\epsilon}} (b-a)N^\delta \binom{N}{x_\ell} = 1 + O\left(N^\delta e^{-3N^{1-2\delta}} + N^{-2\epsilon}\right). \quad (4.19)$$

As this is of size 1, the lower order terms in (4.16) do not contribute to the main term (their contribution is smaller by a power of N).

We can now complete the proof of Theorem 1.9 when $y \notin \mathbb{Q}$ has finite irrationality exponent. Convergence in distribution to Benford is equivalent to showing, for any $(a, b) \subset (0, 1)$, that

$$\sum_{\ell=-N/N^\delta}^{N/N^\delta} \sum_{\substack{i=0 \\ S_{10}(x_{\ell,i}) \in (a,b)}}^{N_\delta-1} \frac{\binom{N}{x_{\ell,i}}}{2^N} = b - a + o(1); \quad (4.20)$$

however, we just showed that. Furthermore, our analysis gives a power savings for the error, and thus we may replace the $o(1)$ with $N^{-\eta}$ for some computable $\eta > 0$ which is a function of ϵ, ϵ' and δ . This completes the proof of this case of Theorem 1.9. \square

Remark 4.3. *A more careful analysis allows one to optimize the error. We want $\delta(-\frac{1}{\kappa} + \epsilon') = -\frac{1}{2} + \delta + \epsilon$, and thus we should take $\delta = (\frac{1}{2} - \epsilon)/(1 + \frac{1}{\kappa} - \epsilon')$. Of course, if we are going to optimize the error we want a significantly better estimate for the probability in the tail. This is easily done by invoking the Central Limit Theorem instead of Chebyshev's Inequality.*

4.3. Case III: $y \notin \mathbb{Q}$ has infinite irrationality exponent. While almost all numbers have irrationality exponent at most 2, the argument in Case II does not cover all possible y (for example, if y is a Liouville number such as $\sum_n 10^{-n!}$). We can easily adapt our proof to cover this case, at the cost of losing a power savings in our error term. As y is still irrational, we still have equidistribution for the logarithms of the segment lengths modulo 1; the difference is now we must replace $O\left(N^{\delta(1-\frac{1}{\kappa}+\epsilon')}\right)$ with $o(N^\delta)$. The rest of the argument proceeds identically, and we obtain in the end an error of $o(1)$ instead of $O(N^{-\eta})$.

5. PROOF OF THEOREM 1.10: DETERMINANT EXPANSION

The techniques introduced to prove that the continuous stick decomposition processes result in Benford behavior can be applied to a variety of dependent systems. To show this, we prove that the $n!$ terms of a matrix determinant expansion follow Benford's Law in the limit as long as the matrix entries are independent, identically distributed nice random variables. As the proof is similar to our previous results, we content ourselves with sketching the arguments, highlighting where the differences occur.

Consider an $n \times n$ matrix A with independent identically distributed entries a_{pq} drawn from a continuous real valued density $f(x)$. Without loss of generality, we may assume that all entries of A are non-negative. For $1 \leq i \leq n!$, let $X_{i,n}$ be the i^{th} term in the determinant expansion of A . Thus $X_{i,n} = \prod_{p=1}^n a_{p\sigma_i(p)}$ where the σ_i 's are the $n!$ permutation functions on $\{1, \dots, n\}$.

We prove that the distribution of the significands of the sequence $\{X_{i,n}\}_{p=1}^{n!}$ converges in distribution to Benford's Law when the entries of A are drawn from a distribution f that satisfies (1.6) (with $h = f$). Recall that it suffices to show

- (1) $\lim_{N \rightarrow \infty} \mathbb{E}[P_n(s)] = \log_{10}(s)$, and
(2) $\lim_{N \rightarrow \infty} \text{Var}(P_n(s)) = 0$.

We first quantify the degree of dependencies, and then sketch the proofs of the mean and variance. Fix $i \in \{1, \dots, n!\}$ and consider the number of terms $X_{j,n}$ that share exactly k entries of A with $X_{i,n}$. Equivalently: If we permute the numbers $1, 2, \dots, n$, how likely is it that exactly k are returned to their original starting position? This is the well known problème des rencontres (the special case of $k = 0$ is counting derangements), and in the limit the probability distribution converges to that of a Poisson random variable with parameter 1 (and thus the mean and variance tend to 1; see [HSW]). Therefore, if $K_{i,j}$ denotes the number of terms $X_{i,n}$ and $X_{j,n}$ share, the probability that $K_{i,j} > \log N$ is $o(1)$.

The determination of the mean follows as before. By linearity of expectation we have

$$\mathbb{E}[P_n(s)] = \frac{1}{n!} \sum_{i=1}^{n!} \mathbb{E}[\varphi_s(X_{i,n})]. \quad (5.1)$$

It suffices to show that $\lim_{n \rightarrow \infty} \mathbb{E}[\varphi_s(X_{i,n})] = \log_{10} s$. We have

$$\begin{aligned} \mathbb{E}[\varphi_s(X_{i,n})] &= \int_{a_{1\sigma_i(1)}} \int_{a_{2\sigma_i(2)}} \cdots \int_{a_{n\sigma_i(n)}} \varphi_s \left(\prod_{p=1}^n a_{p\sigma_i(p)} \right) \\ &\quad \cdot \prod_{p=1}^n f(a_{p\sigma_i(p)}) da_{1\sigma_i(1)} da_{2\sigma_i(2)} \cdots da_{n\sigma_i(n)}, \end{aligned} \quad (5.2)$$

where $a_{p\sigma_i(p)}$ are the entries of A . As these random variables are independent and $f(x)$ satisfies (1.6), the convergence to Benford follows from [JKKKM] (or see Appendix A of [B-]).

To complete the proof of convergence to Benford, we need to control the variance of $P_n(s)$. Arguing as before gives

$$\text{Var}(P_n(s)) = \frac{\log_{10}(s)}{n!} + \frac{1}{(n!)^2} \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{n!} \mathbb{E}[\varphi_s(X_{i,n})\varphi_s(X_{j,n})] \right) - \log_{10}^2 s + o(1). \quad (5.3)$$

We then mimic the proof from §2.2. There we used that, for a fixed i , the number of the 2^N pairs (i, j) with n factors not in common was 2^{n-1} ; in our case we use $K_{i,j}$ is approximately Poisson distributed to show that, with probability $1 + o(1)$ there are at least $\log N$ different factors. The rest of the proof proceeds similarly.

6. FUTURE WORK

This paper studies continuous decomposition models in which a stick is broken at a proportion p . We propose several variations of a discrete decomposition model in which a stick breaks into pieces of integer length, which we hope to return to in a future paper.

Consider the following additive decomposition process. Begin with a stick of length L and uniformly at random choose an integer length $c \in [1, L - 1]$ at which to cut the stick. This results in two pieces, one of length c and one of length $L - c$. Continue this cutting process on both sticks, and only stop decomposing a stick if its length corresponds to a term in a

given sequence $\{a_n\}$. As one cannot decompose a stick of length one into two integer pieces, sticks of length one stop decomposing as well.

Conjecture 6.1. *The stick lengths that result from this decomposition process follow Benford's Law for many choices of $\{a_n\}$. In particular, if either (i) $\{a_n\} = \{2n\}$ or (ii) $\{a_n\}$ is the set of all prime numbers then the resulting stick lengths are Benford.*

More generally one can investigate processes where the stick lengths continue decomposing if they are in certain residue classes to a fixed modulus, or instead of stopping at the primes one could stop at a sequence with a similar density, such as $\lfloor n \log n \rfloor$. It is an interesting question to see, for a regularly spaced sequence, the relationship between the density of the sequence and the fit to Benford's law. While one must of course be very careful with any numerical exploration, as many processes are almost Benford (see for example the results on exponent random variables in [LSE, MN2]), Theorem 1.7 is essentially a continuous analogue of this problem when $a_n = 2n$, and thus provides compelling support.

Figure 2(a) shows a histogram of chi-square values from numerical simulations with a χ^2_8 distribution overlaid (as we have 9 first digits, there are 8 degrees of freedom). A stick of length approximately 10^4 was decomposed according to the above cutting process, where $\{a_n\} = 2n$. A chi-square value⁹ was calculated by comparing the frequencies of each digit obtained from the simulation to the frequencies predicted by Benford's Law. If the simulated data is a good fit to Benford's Law, the χ^2 values should follow a χ^2 distribution. Examining the plot below shows that our numerical simulations support Conjecture 6.1(i). Similarly, Figure 2(b), which shows the chi-square values obtained when $\{a_n\}$ is the set of all primes, supports Conjecture 6.1(ii).

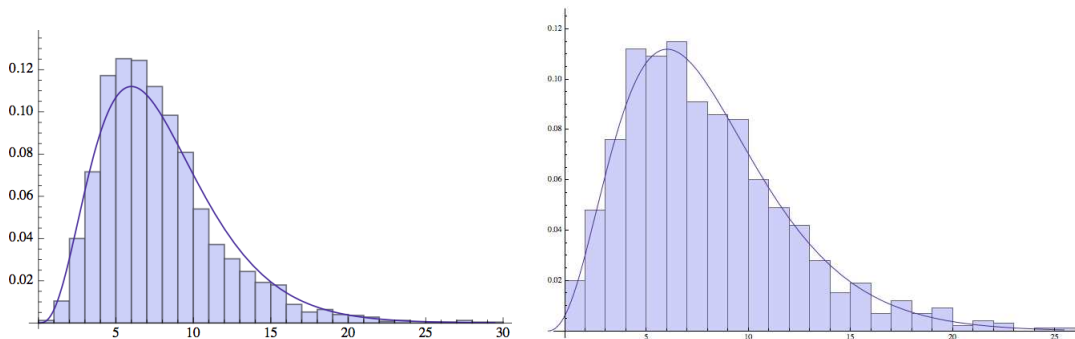


FIGURE 2. (a) $\{a_n = 2n\}$: 2500 chi-square values of stick lengths to Benford behavior after decomposition of sticks of length approximately 10^7 . (b) $\{a_n = p_n\}$ (the set of primes): 1000 chi-square values of stick lengths to Benford behavior after decomposition of sticks of length approximately 10^{500} .

Conjecture 6.2. *Fix a monotonically increasing sequence $\{a_n\}$. Consider a decomposition process where a stick whose length is in the sequence does not decompose further. The process is not Benford if $\{a_n\}$ is any of the following: $\{n^2\}$, $\{2^n\}$, $\{F_n\}$ where F_n is the n^{th} Fibonacci number.*

⁹If X_i are the simulation frequencies of each digit $i = 1, 2, \dots, 9$ and Y_i are the frequencies of each digit as predicted by Benford's Law, where $i = 1, \dots, 9$, then the chi-squared value we calculated is given by $\sum_{i=1}^9 (X_i N - Y_i N)^2 / NY_i$, where N is the number of pieces.

More generally, we do not expect Benford behavior if the sequence is too sparse, which is the case for polynomial growth in n (when the exponent exceeds 1) or, even worse, geometric growth.

Figure 3 features plots of the observed digit frequencies vs the Benford probabilities. These plots also give the total number of fragmented pieces, the number of pieces whose lengths belong to the stopping sequence, and the chi-squared value. Notice that the stopping sequences in Conjecture 6.2 result in stick lengths with far too high a probability of being of length 1. This fact leads us to believe that the sequences are not “dense” enough to result in Benford behavior.

In addition to proofs of the conjectures, future work could include further exploration of the relationship between stopping sequence density and Benford behavior.

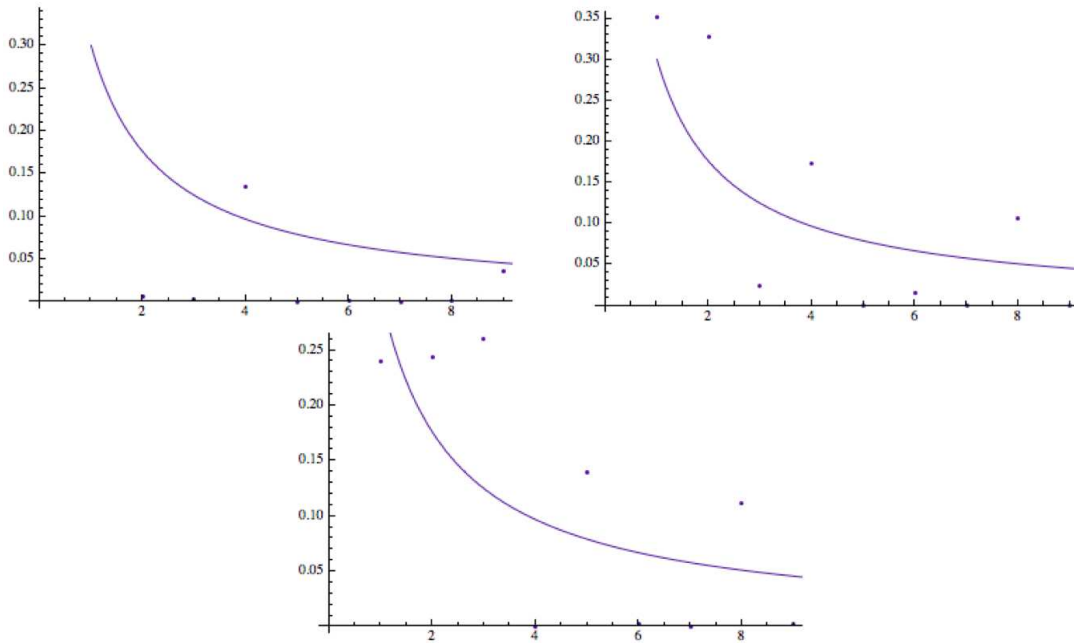


FIGURE 3. (a) $\{a_n = n^2\}$: Distribution of stick lengths after decomposing sticks of starting length approximately 10^6 . (b) $\{a_n = 2^n\}$: Distribution of stick lengths after decomposing 1000 sticks of starting length approximately 10^{500} . (c) $\{a_n = F_n\}$, the Fibonacci Numbers: Distribution of stick lengths after decomposing 500 sticks of starting length approximately 10^{500} .

APPENDIX A. NON-BENFORD BEHAVIOR

We consider the following generalization of the Unrestricted Decomposition Model of Theorem 1.5, where now the proportions at level n are drawn from a random variable with density ϕ_n . If there are only finitely many densities and if the Mellin transform condition is satisfied, then the sequence of stick lengths converges to Benford; if there are infinitely many possibilities, however, then it is possible to obtain non-Benford behavior.

Specifically, we give an example of a sequence of distributions \mathcal{D} with the following property: If all cut proportions are drawn from any one distribution in \mathcal{D} then the resulting stick

lengths converge to Benford's Law, but if all cut proportions in the n^{th} level are chosen from the n^{th} distribution in \mathcal{D} , the stick lengths do not exhibit Benford behavior.

It is technically easier to work with the densities of the logarithms of the cut proportions. Fix $\delta > 0$ and choose a sequence of ϵ_n 's so that they monotonically decrease to zero and $\epsilon_n < \min\left(\sqrt{\frac{3}{20\pi^2(n+1)^2}}, \frac{1}{2} \frac{\delta}{2^n}, \frac{\log \frac{1}{2}}{2}, \frac{1 - \log \frac{1}{2}}{2}\right)$; these values and the meaning of these constants will be made clear during the construction. Consider the sequence of distributions with the following densities

$$\phi_n(x) = \begin{cases} \frac{1}{2\epsilon_n} & \text{if } |x - \log \frac{1}{2}| < \epsilon_n \\ 0 & \text{otherwise;} \end{cases} \quad (\text{A.1})$$

these will be the densities of the logarithms of the cut proportions.

If the logarithm of the cut proportion is drawn solely from the distribution ϕ_k for a fixed k , then as the number of iterations of this cut process tends to infinity the resulting stick lengths converge to Benford's Law. This follows immediately from Theorem 1.5 as the associated densities of the cut proportions satisfy the Mellin condition (1.6).

We show that if the logarithms of the cut proportions for the n^{th} iteration of the decomposition process is drawn from ϕ_n then the resulting stick lengths do not converge to Benford behavior for certain δ . What we do is first show that the distribution of one stick piece (say the resulting piece from always choosing the left cut at each stage) is non-Benford, and then we prove that the ratio of the lengths of any two final pieces is approximately 1. The latter claim is reasonable as our cut proportions are becoming tightly concentrated around 1/2; if they were all exactly 1/2 then all final pieces would be the same length.

Instead of studying the Mellin transforms of the cut proportions we can study the Fourier coefficients of the logarithms of the proportions (again, this is not surprising as the two transforms are related by a logarithmic change of variables). For a function ϕ on $[0, 1]$, its n^{th} Fourier coefficient is

$$\widehat{\phi}(n) := \int_0^1 \phi(x) e^{-2\pi i n x} dx. \quad (\text{A.2})$$

Miller and Nigrini [MN1] proved that if R_1, \dots, R_M are continuous independent random variables with g_m the density of $\log_{10} S_{10}(|R_m|)$, then the product $R_1 \cdots R_M$ converges to Benford's law if and only if for each non-zero n we have $\lim_{M \rightarrow \infty} \widehat{g}_1(n) \cdots \widehat{g}_M(n) = 0$.

We show that if we take our densities to be the ϕ_n 's that the limit of the product of the Fourier coefficients at 1 does not tend to 0. We have

$$\widehat{\phi}_n(1) = \frac{1}{2\epsilon_n} \int_{\frac{1}{2}-\epsilon_n}^{\frac{1}{2}+\epsilon_n} e^{-2\pi i x} dx = -1 + \frac{2\pi^2 \epsilon_n^2}{3} + O(\epsilon_n^4). \quad (\text{A.3})$$

As we are assuming $\epsilon_n < \sqrt{\frac{3}{20\pi^2(n+1)^2}}$, we find for n large that

$$\left| \widehat{\phi}_n(1) \right| \geq \frac{n^2 + 2n}{(n+1)^2}. \quad (\text{A.4})$$

As

$$\lim_{N \rightarrow \infty} \prod_{n=1}^N \frac{n^2 + 2n}{(n+1)^2} = \frac{1}{2}, \quad (\text{A.5})$$

we see that

$$\lim_{N \rightarrow \infty} \prod_{n=1}^N \left| \widehat{\phi}_1(1) \cdots \widehat{\phi}_N(1) \right| > 0. \quad (\text{A.6})$$

This argument shows that the product of N random variables, where the logarithm of the n^{th} variable is drawn from the distribution ϕ_n , is not Benford. This product is analogous to the length of one of the 2^N sticks that are created after N iterations of our cut process. To show that the entire collection of stick lengths does not follow a Benford distribution, we argue that all of the lengths are virtually identical because the cutting proportion tends to $1/2$ (specifically, that the ratio of any two lengths is approximately 1).

Let us denote the lengths of the sticks left after the N^{th} iteration of our cutting procedure by $X_{N,i}$, where $i = 1, \dots, 2^N$. Each length is a product of N random variables; the n^{th} term in the product is either p_n (or $1 - p_n$), where the logarithm of p_n is drawn from the distribution ϕ_n . Proving the ratio of any two lengths is approximately 1 is the same as showing that $\log(X_{N,i}/X_{N,j})$ is approximately 0. If we can show the largest ratio has a logarithm close to 0 then we are done. The largest (and similarly smallest) ratio comes when $X_{N,i}$ and $X_{N,j}$ have no terms in common, as then we can choose one to have the largest possible cut at each stage and the other the smallest possible cut. Thus $X_{N,i}$ always has the largest possible cut; as the largest logarithm at the n^{th} stage is $\epsilon_n + \log \frac{1}{2}$, its proportion at the n^{th} level is $e^{\epsilon_n + \log \frac{1}{2}}$. Similarly $X_{N,j}$ always has the smallest cut, which at the n^{th} level is $e^{-\epsilon_n + \log \frac{1}{2}}$. Therefore

$$\log \frac{X_{N,i}}{X_{N,j}} = \log \prod_{n=1}^N \frac{e^{\epsilon_n + \log \frac{1}{2}}}{e^{-\epsilon_n + \log \frac{1}{2}}} = \sum_{n=1}^N 2\epsilon_n; \quad (\text{A.7})$$

as we chose $\epsilon_n < \frac{1}{2} \frac{\delta}{2^n}$, the maximum ratio between two pieces is at most δ . By choosing δ sufficiently small we can ensure that all the pieces have approximately the same significands (for example, if $\delta < 10^{-2014}$ then we cannot get all possible first digits).

APPENDIX B. NOTES ON LEMONS' WORK

In his 1986 paper, *On the Number of Things and the Distribution of First Digits*, Lemons [Lem] models a particle decomposition process and offers it as evidence for the prevalence of Benford behavior, arguing that many sets that exhibit Benford behavior are merely the breaking down of some conserved quantity. However, Lemons is not completely mathematically rigorous in his analysis of the model (which he states in the paper), and glosses over several important technical points. We briefly mention some issues, such as concerns about the constituent pieces in the model as well as how the initial piece decomposes. We discuss our resolutions of these issues as well as their impacts on the behavior of the system.

The first issue in Lemons' model concerns the constituent pieces. He assumes the set of possible piece sizes is bounded above and below and is drawn from a finite set, eventually specializing to the case where the sizes are in a simple arithmetic progression (corresponding to a uniform spacing), and then taking a limit to assume the pieces are drawn from a continuous range. In this paper, we allow our piece lengths to be drawn continuously from intervals at the outset, and not just in the limit. This removes some, but by no means all, of the technical complications. One must always be careful in replacing discrete systems with continuous ones, especially as there can be number-theoretic restrictions on which discrete systems have a solution. Modeling any conserved quantity is already quite hard with the restriction that the sum of all parts must total to the original starting amount; if the pieces

are forced to be integers then certain number theoretic issues arise. For example, imagine our pieces are of length 2, 4 or 6, so we are trying to solve $2x_1 + 4x_2 + 6x_3 = n$. There are no solutions if $n = 2013$, but there are 84,840 if n is 2012. By considering a continuous system from the start, we avoid these Diophantine complications. We hope to return to the corresponding discrete model in a sequel paper.

A second issue missing from Lemons argument is how the conserved quantity, the number of pieces, and the piece sizes should be related. The continuous limit thus requires consideration of three quantities. Without further specification of their relative scaling, the power law distribution (which leads to Benford behavior) is but one possible outcome. Statistical models of the fragmentation of a conserved quantity based on integer partitioning have been constructed [ChaMek, LeeMek, Mek]. These models can lead to a power law distribution but only for special weightings for the different partitions. Whether this distribution can be obtained from equally weighted partitions (as used in Lemons argument) is an important question, to which we hope to return.

A related issue is that it is unclear how the initial piece breaks up. The process is not described explicitly, and it is unclear how likely some pieces are relative to others. Finally, while he advances heuristics to determine the means of various quantities, there is no analysis of variances or correlations. This means that, though it may seem unlikely, the averaged behavior could be close to Benford while most partitions would be far from Benford.¹⁰

These are important issues, and their resolution and model choice impacts the behavior of the system. We mention a result from Miller-Nigrini [MN2], where they prove that while order statistics are close to Benford's Law (base 10), they do not converge in general to Benford behavior. In particular, this means that if we choose N points randomly on a stick of length L and use those points to partition our rod, the distribution of the resulting piece sizes *will not* be Benford. Motivated by this result and Lemons' paper, we instead considered a model where at stage N we have 2^N sticks of varying lengths, and each stick is broken into two smaller sticks by making a cut on it at some proportion. Each cut proportion is chosen from the unit interval according to a density f . Dependencies clearly exist within this system as the lengths of final sticks must sum to the length of the starting stick.

REFERENCES

- [Adh] A. K. Adhikari, *Some results on the distribution of the most significant digit*, Sankhyā: The Indian Journal of Statistics, Series B **31** (1969), 413–420.
- [AS] A. K. Adhikari and B. P. Sarkar, *Distribution of most significant digit in certain functions whose arguments are random variables*, Sankhyā: The Indian Journal of Statistics, Series B **30** (1968), 47–58.
- [AF] R. L. Adler and L. Flatto, *Uniform distribution of Kakutani's interval splitting procedure*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **38** (1977), no. 4, 253–259.
- [ARS] T. Anderson, L. Rolin and R. Stoehr, *Benford's Law for Coefficients of Modular Forms and Partition Functions*, Proc. Am. Math. Soc. **139** (2011), no. 5, 1533–1541.
- [B–] T. Becker, T. C. Corcoran, A. Greaves-Tunnell, J. R. Iafate, J. Jing, S. J. Miller, J. D. Porfilio, R. Ronan, J. Samranvedhya and F. Strauch, *Benford's Law and Continuous Dependent Random Variables*, arXiv version. <http://arxiv.org/pdf/1309.5603>.
- [Ben] F. Benford, *The law of anomalous numbers*, Proceedings of the American Philosophical Society **78** (1938), 551–572.

¹⁰Imagine we toss a coin one million times, always getting either all heads or all tails. Let's say these two outcomes are equally likely. If we were to perform this process trillions of times, the total number of heads and tails would be close to each other; however, no individual experiment would be close to 50%.

- [Ber1] A. Berger, *Multi-dimensional dynamical systems and Benford's Law*, Discrete Contin. Dyn. Syst. **13** (2005), 219–237.
- [Ber2] A. Berger, *Benford's Law in power-like dynamical systems* **5** (2005), Stoch. Dyn., 587–607.
- [BBH] A. Berger, Leonid A. Bunimovich and T. Hill, *One-dimensional dynamical systems and Benford's Law*, Trans. Amer. Math. Soc. **357** (2005), no. 1, 197–219.
- [BH1] A. Berger and T. P. Hill, *Newton's method obeys Benford's Law*, American Mathematical Monthly **114** (2007), 588–601.
- [BH2] A. Berger, and T. P. Hill, *Benford Online Bibliography*, <http://www.benfordonline.net>.
- [BH3] A. Berger and T. P. Hill, *Fundamental flaws in Feller's classical derivation of Benford's Law*, Preprint, University of Alberta, 2010.
- [BH4] A. Berger and T. P. Hill, *Benford's Law strikes back: No simple explanation in sight for mathematical gem*, The Mathematical Intelligencer **33** (2011), 85–91.
- [BH5] A. Berger and T. P. Hill, *A Basic Theory of Benford's Law*, Probab. Surv. **8** (2011), 1–126.
- [BHKR] A. Berger, T. P. Hill, B. Kaynar and A. Ridder, *Finite-state Markov Chains Obey Benford's Law*, to appear in SIAM J. Matrix Analysis.
- [Bert] J. Bertoin, *Random fragmentation and coagulation processes*, Cambridge Studies in Advanced Mathematics **102**, Cambridge University Press, Cambridge, 2006.
- [Bh] R. N. Bhattacharya, *Speed of convergence of the n -fold convolution of a probability measure on a compact group*, Z. Wahrscheinlichkeitstheorie verw. Geb. **25** (1972), 1–10.
- [BrDu] M. D. Brennan and R. Durrett, *Splitting intervals*, Ann. Probab. **14** (1986), no. 3, 1024–1036.
- [Ca] I. Carbone, *Discrepancy of LS-sequences of partitions and points*, Ann. Mat. Pura Appl. **191** (2012), 819–844.
- [CaVo] I. Carbone and A. Volčič, *Kakutani's splitting procedure in higher dimension*, Rend. Istit. Mat. Univ. Trieste **39** (2007), 119–126.
- [ChaMek] K. C. Chase and A. Z. Mekjian, *Nuclear fragmentation and its parallels*, Physical Review C **49** (1994), 2164.
- [Che] Hongwei Chen (2010). *On the Summation of First Subseries in Closed Form*. International Journal of Mathematical Education in Science and Technology 41:4, 538–547.
- [CLTF] E. Costasa, V. López-Rodasa, F. J. Torob, A. Flores-Moya, *The number of cells in colonies of the cyanobacterium Microcystis aeruginosa satisfies Benford's Law*, Aquatic Botany **89** (2008), no. 3, 341–343.
- [CLM] V. Cuff, A. Lewis and S. J. Miller, *The Weibull distribution and Benford's Law*, preprint.
- [Dia] P. Diaconis, *The distribution of leading digits and uniform distribution mod 1*, Ann. Probab. **5** (1979), 72–81.
- [DrNi] P. D. Drake and M. J. Nigrini, *Computer assisted analytical procedures using Benford's Law*, Journal of Accounting Education **18** (2000), 127–146.
- [GS] G. R. Grimmett and D. R. Stirzaker, *Probability and random processes* (2009), 2nd edition, Clarendon Press, Oxford.
- [HSW] D. Hanson, K. Seyffarth and J. H. Weston, *Matchings, Derangements, Rencontres*, Mathematics Magazine **56** (1983), no. 4, 224–229.
- [Hi1] T. Hill, *The first-digit phenomenon*, American Scientists **86** (1996), 358–363.
- [Hi2] T. Hill, *A statistical derivation of the significant-digit law*, Statistical Science **10** (1996), 354–363.
- [HS] M. Hindry and J. Silverman, *Diophantine geometry: An introduction*, Graduate Texts in Mathematics **201**, Springer, New York, 2000.
- [Hu] W. Hurlimann, *Benford's Law from 1881 to 2006*, <http://arxiv.org/pdf/math/0607168>.
- [IV] M. Infusino and A. Volčič, *Uniform distribution on fractals*, Unif. Distrib. Theory **4** (2009), no. 2, 47–58.
- [JKKKM] 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, volume 1, number 1 (March 2009), 37–60.
- [JN] S. Janson and R. Neininger, *The size of random fragmentation trees*, Probab. Theory Related Fields **142** (2008), no. 3–4, 399–442.
- [JR] E. Janvresse and T. de la Rue, *From uniform distribution to Benford's Law*, Journal of Applied Probability **41** (2004) no. 4, 1203–1210.

- [Ka] S. Kakutani, *A problem of equidistribution on the unit interval* $[0, 1]$, Measure theory (Proc. Conf., Oberwolfach, 1975), Lecture Notes in Math., Vol. 541, Springer, Berlin, 1976, pp. 369–375.
- [Knu] D. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, 3rd edition, Addison-Wesley, MA, 1997.
- [Kol] A. N. Kolmogoroff, *Über das logarithmisch normale Verteilungsgesetz der Dimensionen der Teilchen bei Zerstückelung*, C. R. (Doklady) Acad. Sci. URSS (N. S.) **31** (1941), 99–101.
- [KonMi] A. Kontorovich and S. J. Miller, *Benford's Law, values of L-functions and the $3x + 1$ problem*, Acta Arith. **120** (2005), 269–297.
- [KN] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, John Wiley & Sons, 1974.
- [LS] J. Lagarias and K. Soundararajan, *Benford's Law for the $3x + 1$ Function*, J. London Math. Soc. **74** (2006), series 2, no. 2, 289–303.
- [LeeMek] S. J. Lee and A. Z. Mekjian, *Canonical studies of the cluster distribution, dynamical evolution, and critical temperature in nuclear multifragmentation processes*, Physical Review C **45** (1992), 1284.
- [LSE] L. M. Leemis, B. W. Schmeiser and D. L. Evans, *Survival Distributions Satisfying Benford's Law*, The American Statistician **54** (2000), no. 4, 236–241.
- [Lem] D. S. Lemons, *On the number of things and the distribution of first digits*, Am. J. Phys. **54** (1986), no. 9, 816–817.
- [Lév1] P. Lévy, *Théorie de l'Addition des Variables Aléatoires*, Gauthier-Villars, Paris, 1937.
- [Lév2] P. Lévy, *L'addition des variables aléatoires définies sur une circonférence*, Bull. de la S. M. F. **67** (1939), 1–41.
- [Lo] J. C. Lootgieter, *Sur la répartition des suites de Kakutani. I*, Ann. Inst. H. Poincaré Sect. B (N.S.) **13** (1977), no. 4, 385–410.
- [Meb] W. Mebane, *Detecting Attempted Election Theft: Vote Counts, Voting Machines and Benford's Law*, Prepared for delivery at the 2006 Annual Meeting of the Midwest Political Science Association, April 20-23, Palmer House, Chicago. <http://www.umich.edu/~wmebane/mw06.pdf>.
- [Mek] A. Z. Mekjian, *Model of a fragmentation process and its power-law behavior*, Physical Review Letters **64** (1990), 2125.
- [Mil] S. J. Miller (editor), *The Theory and Applications of Benford's Law*, Princeton University Press, to appear.
- [MN1] S. J. Miller and M. Nigrini, *The Modulo 1 Central Limit Theorem and Benford's Law for Products*, International Journal of Algebra **2** (2008), no. 3, 119–130.
- [MN2] S. J. Miller and M. Nigrini, *Order statistics and Benford's Law*, International Journal of Mathematics and Mathematical Sciences, Volume 2008 (2008), Article ID 382948, 19 pages. doi:10.1155/2008/382948
- [MN3] S. J. Miller and M. Nigrini, *Data diagnostics using second order tests of Benford's Law*, Auditing: A Journal of Practice and Theory **28** (2009), no. 2, 305–324. doi: 10.2308/aud.2009.28.2.305
- [MT-B] S. J. Miller and R. Takloo-Bighash, *An Invitation to Modern Number Theory*, Princeton University Press, Princeton, NJ, 2006.
- [New] S. Newcomb, *Note on the frequency of use of the different digits in natural numbers*, Amer. J. Math. **4** (1881), 39–40.
- [Nig1] M. Nigrini, *The Detection of Income Tax Evasion Through an Analysis of Digital Frequencies*, PhD thesis, University of Cincinnati, OH, USA, 1992.
- [Nig2] M. Nigrini, *Using Digital Frequencies to Detect Fraud*, The White Paper **8** (1994), no. 2, 3–6.
- [Nig3] 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.
- [NM] 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.
- [NigMi] M. Nigrini and L. Mittermaier, *The Use of Benford's Law as an Aid in Analytical Procedures*, Auditing: A Journal of Practice & Theory **16** (1997), no. 2, 52–67.
- [Ol] J. Olli, *Division point measures resulting from triangle subdivisions*, Geom. Dedicata **158** (2012), 69–86.

- [PHA] F. Perez-Gonzalez, G. L. Heileman, and C.T. Abdallah, *Benford's Law in Image Processing*, Proceedings of the IEEE International Conference on Image Processing (ICIP), San Antonio, TX, Volume 1, 2007, pp. 405–408.
- [PTTV] L. Pietronero, E. Tosatti, V. Tosatti and A. Vespignani, *Explaining the Uneven Distribution of Numbers in Nature: the Laws of Benford and Zipf*, Physica A: Statistical Mechanics and its Applications, 2001.
- [Pin] R. Pinkham, *On the Distribution of First Significant Digits*, The Annals of Mathematical Statistics **32**, no. 4 (1961), 1223–1230.
- [PvZ] R. Pyke and W. R. van Zwet, *Weak convergence results for the Kakutani interval splitting procedure*, Ann. Probab. **32** (2004), no. 1A, 380–423.
- [Rai] R. A. Raimi, *The first digit problem*, Amer. Math. Monthly **83** (1976), no. 7, 521–538.
- [Rob] H. Robbins, *On the equidistribution of sums of independent random variables*, Proc. Amer. Math. Soc. **4** (1953), 786–799.
- [Rod] R. Rodriguez, *First significant digit patterns from mixtures of uniform digits*, The American Statistician, 2004.
- [Ro] K. Roth, *Rational approximations to algebraic numbers*, Mathematika **2** (1955), 1–20.
- [Sa] H. Sakamoto, *On the distributions of the product and the quotient of the independent and uniformly distributed random variables*, Tôhoku Math. J. **49** (1943), 243–260.
- [Sc1] P. Schatte, *On sums modulo 2π of independent random variables*, Math. Nachr. **110** (1983), 243–261.
- [Sc2] P. Schatte, *On the asymptotic uniform distribution of sums reduced mod 1*, Math. Nachr. **115** (1984), 275–281.
- [Sc3] P. Schatte, *On the asymptotic logarithmic distribution of the floating-point mantissas of sums*, Math. Nachr. **127** (1986), 7–20.
- [Sl] E. Slud, *Entropy and maximal spacings for random partitions*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **41** (1977/78), no. 4, 341–352.
- [Son] J. Sondow, *Irrationality Measures, Irrationality Bases, and a Theorem of Jarník*. <http://arxiv.org/ftp/math/papers/0406/0406300.pdf>.
- [SS1] E. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press, Princeton, NJ, 2003.
- [SS2] E. Stein and R. Shakarchi, *Complex Analysis*, Princeton University Press, Princeton, NJ, 2003.
- [ST] M. D. Springer and W. E. Thompson, *The distribution of products of independent random variables*, SIAM J. Appl. Math. **14** (1966) 511–526.
- [Tö] K-H Tödter, *Benford's Law as an Indicator of Fraud in Economics*, German Economic Review **10** (2009), 339–351.
- [vZ] W. R. van Zwet, *A proof of Kakutani's conjecture on random subdivision of longest intervals*, Ann. Probability **6** (1978), no. 1, 133–137.

E-mail address: tbecker@smith.edu

SMITH COLLEGE, NORTHAMPTON, MA 01063

E-mail address: taylorc3@email.arizona.edu

THE UNIVERSITY OF ARIZONA, TUCSON, AZ 85721

E-mail address: ahg1@williams.edu

WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267

E-mail address: Joseph.R.Iafrate@williams.edu

WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267

E-mail address: Joy.Jing@williams.edu

WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267

E-mail address: sjm1@williams.edu, Steven.Miller.MC.96@aya.yale.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267

E-mail address: Jaclyn.D.Porfilio@williams.edu

WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267

E-mail address: ryan.p.ronan@gmail.com

COOPER UNION, NEW YORK, NY 10003

E-mail address: Jirapat.Samranvedhya@williams.edu

WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267

E-mail address: fws1@williams.edu

DEPARTMENT OF PHYSICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267