Check for
updates

Benford's law and continuous dependent random variables

Thealexa Becker [a], David Burt [b], Taylor C. Corcoran [c],
Alec Greaves-Tunnell [d], Joseph R. Iafrate [e], Joy Jing [b],
Steven J. Miller [f,*], Jaclyn D. Porfilio [b], Ryan Ronan [g],
Jirapat Samranvedhya [h], Frederick W. Strauch [i], Blaine Talbut [j]

[a] Smith College, Northampton, MA 01063, United States
[b] Williams College, Williamstown, MA 01267, United States
[c] UCLA, Los Angeles, CA 90095, United States
[d] Department of Statistics, University of Washington, Seattle, WA 98195, United States
[e] University of Michigan, Ann Arbor, MI 48109, United States
[f] Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267, United States
[g] CUNY Ph.D. Program in Mathematics, The Graduate Center, New York, NY 10016, United States
[h] NC State, Raleigh, NC 27695, United States
[i] Department of Physics, Williams College, Williamstown, MA 01267, United States
[j] University of California Los Angeles, Los Angeles, CA 90095, United States

ARTICLE INFO

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. We

* Corresponding author.

E-mail addresses: thealexa.g.becker@gmail.com (T. Becker), david.r.burt94@gmail.com (D. Burt), taylor.corcoran.1@anderson.ucla.edu (T.C. Corcoran), alecgt@uw.edu (A. Greaves-Tunnell), joseph.r.iafrate@hotmail.com (J.R. Iafrate), Joy.Jing1@gmail.com (J. Jing), sjm1@williams.edu, Steven.Miller.MC.96@aya.yale.edu (S.J. Miller), jpor2121@gmail.com (J.D. Porfilio), rronan@gradcenter.cuny.edu (R. Ronan), smrnvdhy@gmail.com (J. Samranvedhya), fws1@williams.edu (F.W. Strauch), blainetalbut@math.ucla.edu (B. Talbut).

handle this by using tools from Fourier analysis and irrationality exponents to obtain quantified convergence rates as well as introducing and developing techniques to measure and control the dependencies. The construction of these tools is one of the major motivations of this work, as our approach can be applied to many other dependent 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.

1. Introduction

1.1. Background

In 1881, American astronomer Simon Newcomb [1] 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 [2], 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 the absolute value of a number. Specifically, given any $x > 0$ we can write it as

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

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 $\mathrm{Prob}(S_B(x) \leq s) = \log_B(s)$.

Benford's Law arises in applied mathematics [3], auditing [4–9], biology [10,11], computer science [12], dynamical systems [13–17], economics [18,19], geology [20], number theory [21–23], physics and astrophysics [24–31], signal processing [32], statistics [33,34] and voting fraud detection [35], to name just a few. See [36,37] for extensive bibliographies and [38–49] for general surveys and explanations of the Law's prevalence, as well as the book edited by Miller [50], 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 [45,47,51–61] 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 [40]). Many of the proofs use techniques from measure theory and Fourier analysis, and 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 (see for example [47] or Chapter 3 of [50]).

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, previous work studied 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 [15,22,23,42,52]. 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 $M = N!$). This work is the compilation of many projects over the past several years of efforts to build a counting approach to control the dependencies. To show the utility of this new perspective we successfully apply it in many systems.

Our introduction to the subject of this paper came from reading an article of Lemons [62] (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. 1-Dimensional decomposition models and notation

The techniques we develop to show Benford behavior for problems with dependent random variables are applicable to many systems. In the interest of space, we will describe in detail here just three variations of a stick decomposition, and later discuss conjectures about other possible decomposition processes and some results in higher dimensions. As an example of the power of this approach 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 [63] considered the following deterministic process. Let $Q_0 = \{0, 1\}$ and given $Q_k = \{x_0 = 0, x_1, \ldots, x_k = 1\}$ (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 $x_{i+1} - x_i = \max_{1 \le \ell \le k-1} |x_{\ell+1} - x_\ell|$. He proved that as $k \to \infty$, the points of Q_k become uniformly distributed, which implies that this process is non-Benford. This process has been generalized; see for example [64–69] and the references therein, and especially the book [70]. See also [71] for processes related to particle decomposition, [72,73] for 2-dimensional examples, [74] for a fractal setting, and [75] for some results on discrete fragmentation also inspired by Lemons' work.

Most of this paper is devoted to the following decomposition process (and generalizations) whose first few levels are shown in Fig. 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.

(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 similar to a fragmentation tree model investigated by Janson and Neininger [76]. Phrasing their work in our language, they randomly chose and fixed b probabilities $p_1, \ldots, p_b$ and then at each stage each piece of length x split into b pieces of length $p_1 x, \ldots, p_b x$, unless x is below a critical threshold in which
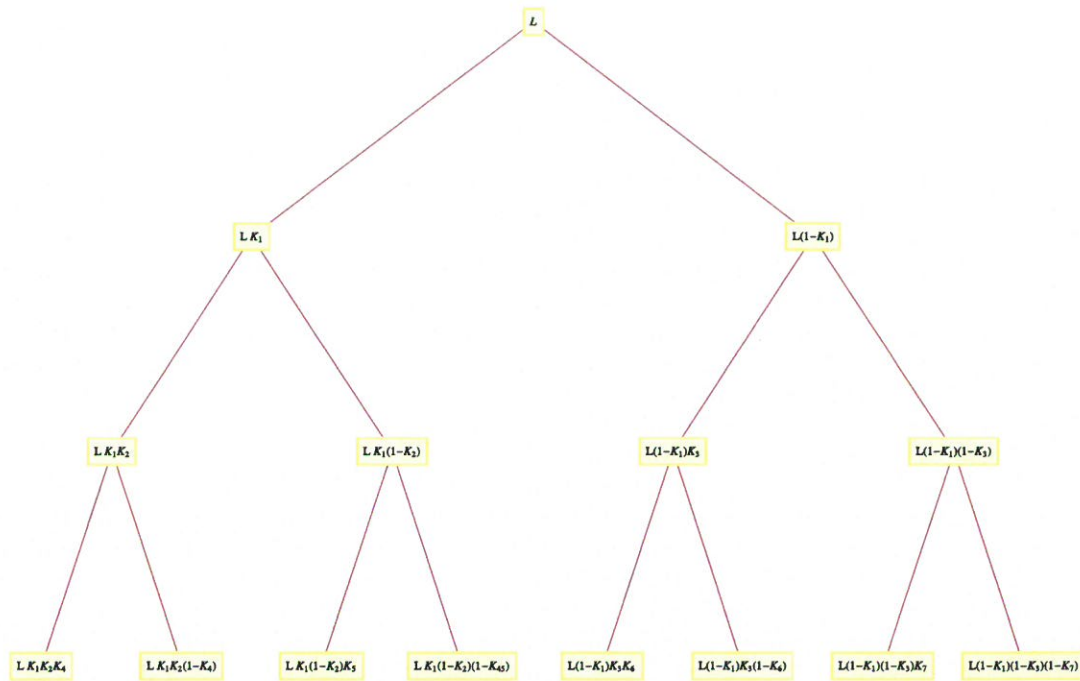
Fig. 1. Unrestricted decomposition: Breaking L into pieces, $N = 3$.

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 [47] 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 leads 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 transform,[2] $\mathcal{M}_f(s)$, by

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

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

---

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

[2] Note $\mathcal{M}_f(s) = \mathbb{E}[x^{s-1}]$, 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 $\mathcal{M}_f(\sigma - i\xi) = 2\pi \int_{-\infty}^\infty \left(f(e^{2\pi u})e^{2\pi \sigma u}\right) e^{-2\pi i u \xi} du$, 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 [77,78].

Definition 1.2 (*Significand Indicator Function, φ_s*)**.** For $s \in [1, 10)$, let

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

thus φ_s is the indicator function of the event of a significand 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 significand 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

$$\varliminf_{q \to \infty} q^{\gamma+1} \min_p \left| \alpha - \frac{p}{q} \right| = 0. \tag{1.5}$$

By Roth's theorem, every algebraic irrational has irrationality exponent 1. See for example [79–81] 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 exist 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 $\lim_{x \to \infty} f(x)/g(x) = 0$.

1.3. Results

We state our results for the fragmentation models of Section 1.2 and some generalizations. While a common way of proving a sequence is Benford base B is to show that its logarithms base B is equidistributed[3] (see, for example, [42,80]), as we are using the Mellin transform and not the Fourier transform we instead often directly analyze the significand function. To show that the first digits of $\{X_i\}$ follow a Benford distribution, it suffices to show that

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

Viewing $P_n(s)$ as the cumulative distribution function of the process, the above shows that we have convergence in distribution[4] to the Benford cumulative distribution function (see [82]).

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). 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 is 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 [47] for more details.

---

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

[4] A 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 $\lim_{n \to \infty} F_n(r) = F(r)$ for each r where F is continuous.

1.3.1. 1-Dimensional results

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

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

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 $p_1, p_2, \ldots, p_{2^N-1}$ from the unit interval according to the probability density f. After N iterations we have

$$
\begin{aligned}
X_1 &= Lp_1p_2p_4 \cdots p_{2^N-2}p_{2^N-1} \\
X_2 &= Lp_1p_2p_4 \cdots p_{2^N-2}(1 - p_{2^N-1}) \\
&\vdots \\
X_{2^N-1} &= L(1 - p_1)(1 - p_3)(1 - p_7) \cdots (1 - p_{2^{N-1}-1})p_{2^N-1} \\
X_{2^N} &= L(1 - p_1)(1 - p_3)(1 - p_7) \cdots (1 - p_{2^{N-1}-1})(1 - p_{2^N-1}),
\end{aligned}
\tag{1.7}
$$

and

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

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

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

Thus as $N \to \infty$, 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, ... or *k* pieces with probabilities $q_0, q_1, q_2, \ldots, q_k$ (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 (the Pigeonhole Principle ensures that we get arbitrarily large number of copies of at least one, which suffices for the needed decay); we give one in Appendix A.[5]

The discrete analogue of the unrestricted model also results in Benford behavior asymptotically.

Theorem 1.7. *Consider the following fragmentation process. Start with a rod of integer length $\ell = \ell_0$, in iteration k select an integer $X_k \in [1, \ell_k]$ with uniform probability and fracture the rod so that you are left with a piece of length X_k and a piece of length $\ell_k - X_k =: \ell_{i+1}$. Continue this process until $\ell_k = 0$. Then if $\{X_\ell\}$ denotes the (random) set of the X_k, as $\ell \to \infty$ the distribution of $\{X_\ell\}$ converges to (strong) Benford behavior with probability 1.*

---

[5] The 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 [47]).

Remark 1.8. The techniques used in the proof of Theorem 1.7 generalize naturally to a wide class of integer-valued probability functions. In particular similar proofs should work for density functions that produce a large number of fragments with high probability and can be well approximated by a continuous process satisfying the Mellin Transform property of the previous section.

Theorem 1.9 (*Restricted 1-Dimensional Decomposition Model*). *Start 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 of 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 $p_1 p_2$ 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:*

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

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

Remark 1.10. We may replace the uniform distribution with any nice distribution that satisfies the Mellin transform condition of (1.6).

Theorem 1.11 (*Fixed Proportion 1-Dimensional 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*

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

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.

1.3.2. 2-Dimensional results

The next two results are indicative of what can be done for fragmentation problems in higher dimensions.

Fix a continuous probability density f on a $45° - 45° - 90°$ triangle $T_{45-45-90}$. Any other triangle T is the image of $T_{45-45-90}$ under some affine transformation $\vec{v} \mapsto A\vec{v} + \vec{b}$. Because affine transformations preserve ratios of areas, this affine transformation carries f to a continuous probability distribution f_T on T. Consider the following two-dimensional decomposition process. Beginning with any triangle T_0, select a point in the interior of T_0 according to the distribution f_{T_0} and connect this point to each of the three vertices to obtain three sub-triangles. Now independently select a point in the interior of each of the three sub-triangles $T_{1,i}$ according to $f_{T,i}$ ($i \in \{1, 2, 3\}$) and repeat this process until there are 3^N triangles. Let $\mathcal{X}_N$ denote the set of areas of the sub-triangles that result from N iterations of the decomposition of T_0 described above.

Theorem 1.12 (*Unrestricted Decomposition of Triangles*). *Let $f^{(1)}(x) = \int_{(x,y) \in T_{45-45-90}} f(x,y) dy, f^{(2)}(y) = \int_{(x,y) \in T_{45-45-90}} f(x,y) dx$. If, for all choices of $f_m \in \{f^{(1)}, f^{(2)}, 1 - f^{(1)} - f^{(2)}\}$,*

$$\lim_{N \to \infty} \sum_{\substack{\ell = -\infty \\ \ell \neq 0}}^{\infty} \prod_{m=1}^{N} \mathcal{M} f_m \left(1 - \frac{2\pi i \ell}{\log 10} \right) = 0, \tag{1.11}$$

then the significands of the areas in $\mathcal{X}$ converge in distribution to Benford's law.

Remark 1.13. We could modify the condition of (1.11) slightly. As the absolute values of the Mellin transforms are at most 1 (as we have probability densities), notice that as each f_m is drawn from the same set of three functions that in the product at least one of these three functions appears at least $N/3$ times. Thus we could replace the arbitrary product of f_m's with identical f_m's.

Remark 1.14. The condition (1.11) is again quite weak, and holds if f has bounded moments. Note further that there is nothing special about our use of a $45° - 45° - 90°$ triangle except that it allows us to express the Mellin condition relatively simply.

Affine mappings do not exist between arbitrary quadrilaterals, as affine mappings preserve parallel lines. However, there are continuous mappings from arbitrary convex quadrilaterals to squares that preserve the ratio of areas; the construction of Gromov after Knothe provides such a mapping with other nice properties [83].

Theorem 1.15 (*Unrestricted Decomposition of Quadrilaterals*). *Start with the unit square, a continuous probability density f on $(0, 1)$, and a continuous probability density g on $(0, 1) \times (0, 1)$. We independently select a point on each side according to f. Call these A, B, C, D. Let Φ be a homeomorphism between the quadrilateral ABCD and the unit square with the property that Φ preserves ratios of areas. We then choose a point E in the interior of ABCD according to $g \circ \Phi$. We now connect E to each of A, B, C, D in order to decompose the square into four convex quadrilaterals. We then perform the same decomposition independently on each of these quadrilaterals, repeating this process until we obtain 4^N quadrilaterals. Suppose f and g have bounded moments. Let $\mathcal{X}_N$ denote the set of areas of the sub-quadrilaterals that result from N iterations of the decomposition of the unit square described above. Then, as $N \to \infty$, the significands of the areas in $\mathcal{X}_N$ converge in distribution to Benford's law.*

Remark 1.16. A more general version of Theorem 1.15 is true, with f and g satisfying a complicated Mellin condition that is implied by finite moments.

1.4. Sketch of proofs

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 new 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 1.11 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 extract 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 Section 7, 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 1.17. *Let A be an n × 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 [45] (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 [84] (the expanded arXiv version of this paper). The dependencies of the pieces are major obstructions and prevent us from simply quoting this result; 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, Eq. (1.6), is extremely weak and is met by most distributions. For example, if f is the uniform density on $(0, 1)$ then

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

which implies

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

(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 Section 2.1 that $\mathbb{E}[P_N(s)] = \log_{10} s$, and then in Section 2.2 prove that $\text{Var}(P_N(s)) \to 0$; 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. The key contribution of this work is quantifying how often certain dependencies can arise, which leads to a tractable analysis.

2.1. Expected value

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

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

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

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

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 $\lim_{N\to\infty}\mathbb{E}[\varphi_s(X_i)] = \log_{10}s$. We have

$$
\mathbb{E}[\varphi_s(X_i)] = \int_{p_1=0}^{1}\int_{p_2=0}^{1}\cdots\int_{p_N=0}^{1}\varphi_s\left(L\prod_{r=1}^{k}p_r\prod_{m=k+1}^{N}(1-p_m)\right)
$$

$$
\cdot\prod_{r=1}^{k}f(p_r)\prod_{m=k+1}^{N}f(1-p_m)\,dp_1 dp_2\cdots dp_N. \tag{2.5}
$$

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 $L\prod_{r=1}^{k}p_r\prod_{m=k+1}^{N}(1-p_m) = X_i$ to Benford follows from [45] (the key theorem is summarized for the reader's convenience in Appendix A in [84], 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 $\lim_{N\to\infty}\mathbb{E}[P_N(s)] = \log_{10}s$, 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 [47] or Corollary A.2 of [84],

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

and thus

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

2.2. Variance

The argument below introduces our technique to handle dependencies among random variables.

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

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

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

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

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 Fig. 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

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

Note that if X_i, our fixed piece, has some factors of $1 - p_i$ we can incorporate those at the cost of replacing some $f(p_i)$ with $f(1 - p_i)$, and the argument would proceed similarly; in the special case of the uniform distribution these are the same.

With these definitions in mind, we have

$$\mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)] \;=\; \int_{p_1=0}^{1}\int_{p_2=0}^{1}\cdots\int_{p_N=0}^{1}\int_{\tilde{p}_{k+2}=0}^{1}\cdots\int_{\tilde{p}_N=0}^{1} \varphi_s\left(L\prod_{r=1}^{k+1}p_r\prod_{r=k+2}^{N}p_r\right)$$
$$\cdot\; \varphi_s\left(L\prod_{r=1}^{k}p_r\cdot(1-p_{k+1})\cdot\prod_{r=k+2}^{N}\tilde{p}_r\right)$$
$$\cdot\; \prod_{r=1}^{N}f(p_r)\prod_{r=k+2}^{N}f(1-\tilde{p}_r)\,dp_1dp_2\cdots dp_N\,d\tilde{p}_{k+2}\cdots d\tilde{p}_N. \tag{2.11}$$

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

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

and consider the following integrals:

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

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, Eq. (3.7) of [45] (or see Corollary A.2 in [84]) gives for $n \geq 4$ that[6]

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

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 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

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

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

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

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:

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

Substituting this into Eq. (2.9) yields

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

Since the variance must be non-negative by definition, it follows that $\lim_{N \to \infty} \text{Var}\,(P_N(s)) = 0$, completing the proof if each cut is drawn from a uniform distribution. The more general case follows analogously, appealing to [47] (or Theorem A.1 of [84]). $\square$

3. Proof of Theorem 1.7: Discrete decomposition

The main idea of the proof is to show that fragments generated for sufficiently large rods can be well approximated by a corresponding continuous fragmentation process in a way that preserves Benford behavior. We formalize this notion in the following lemma.

Lemma 3.1. *Suppose that the random integer on $[1, \ell_k]$ is generated first by selecting a random real number $c_k \in [0, 1]$, then rounding it up to the nearest multiple of $1/\ell_k$. Let $\mathcal{Q}$ denote the continuous process in which we start with a piece of length ℓ, and fracture it in each iteration at c_k.*

Let X_k denote the kth fragment generated $\mathcal{P}$, and Y_k denote the kth fragment generated by $\mathcal{Q}$. Then

$$Y_k \; \leq \; X_k \left(1 + \prod_{j=1}^{k} \left(1 + \frac{1}{\ell_j} \right) \right) + O(1), \tag{3.1}$$

where ℓ_j denote the remaining length of the fragment in process $\mathcal{P}$ after j breaks.

In practice, we use the following corollary of this lemma.

---

[6] Our 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.

Corollary 3.2. *Suppose $\ell_{k-1}, X_k > \log^2(\ell)$ and $g(\ell) = o(\sqrt{\log(\ell)})$ where $g(\ell)$ tends to infinity with ℓ. Then for k such that $k < g(\ell)\log(\ell)$ we have*

$$X_k = Y_k(1 + o(1)). \tag{3.2}$$

We first show the corollary contingent on Lemma 3.1.

Proof of Corollary 3.2. Using that the ℓ_k are monotonically decreasing, we can tightly bound how close X_k and Y_k are:

$$Y_k \leq X_k \ll Y_k \prod_{j=1}^{k} \left(1 + \frac{1}{\ell_j}\right) + O(1). \tag{3.3}$$

Here we bound the terms uniformly above by the largest term in our assumptions, raised to the maximum number of fragments allowed in our hypothesis giving,

$$X_k \ll Y_k \left(1 + \frac{1}{\log^2(\ell)}\right)^{g(\ell)\log(\ell)} + O(1)$$
$$\ll Y_k e^{\frac{g(\ell)}{\log(\ell)}} + O(1),$$

where the second line follows by the limit definition of e since we are asymptotic in ℓ. Taking the limit in ℓ and combining error terms gives the desired bound. $\quad\square$

We now turn to the lemma.

Proof of Lemma 3.1. Let h_k denote the length of the fragments in process $\mathcal{Q}$ after k breaks, c_k denote the continuous value on chosen on $[1, h_k]$ and d_k denote the rounded version of c_k used in process $\mathcal{P}$. Then we have $X_{k+1} = \ell_k(1 - d_k)$, $Y_{k+1} = h_k(1 - c_k)$ and $d_k < c_k < d_k + \frac{1}{\ell_{k-1}}$.

To prove the lemma, it suffices to show that $\ell_k \leq h_k \prod_{j=1}^{k-1}(1 + \frac{1}{\ell_j})$, and then absorb the error from the final cut into the $O(1)$ term. Note

$$h_k \leq \ell_k = \ell \prod_{i=1}^{k} d_i \leq \ell \prod_{i=1}^{k} \left(c_i + \frac{1}{\ell_{i-1}}\right). \tag{3.4}$$

Then since $\ell_{k+1} = d_k \ell_k < c_k \ell_k$, factoring out c_i from each term yields

$$\ell \prod_{i=1}^{k} \left(c_i + \frac{1}{\ell_{i-1}}\right) \leq \ell \prod_{i=1}^{k} c_i \prod_{i=1}^{k} \left(1 + \frac{1}{\ell_i}\right) \leq h_k \prod_{i=1}^{k} \left(1 + \frac{1}{\ell_i}\right). \quad\square \tag{3.5}$$

We now note that if we can show that almost all of the fragments satisfy the properties laid out in Corollary 3.2, this will complete the proof of Theorem 1.7, as the pieces generated by $\mathcal{Q}$ are strong Benford distributed by Theorem 1.5.

Lemma 3.3. *Suppose $\{Y_\ell\}_\ell = \{Y_1, \ldots, Y_{k_\ell}\}$ is strong Benford as ℓ tends to infinity. Then any set $\{X_\ell\}_\ell = \{X_1, \ldots, X_{k_\ell}\}$ such that $X_i = Y_i(1 + o(1))$ is strong Benford as ℓ tends to infinity.*

Proof. We fix a digit j and show that the distribution of the jth digit is Benford. Let $D_j(a)$ denote the jth digit of a. Then since $X_i = Y_i(1 + o(\ell))$, there exists some function of ℓ tending to infinity with ℓ such $D_j(X_i) = D_j(Y_i)$, unless $D_{j-k}(Y_i) = 9$ or $D_{j-k}(X_i) = 9$ for all $k \leq f(\ell)$. Since the Y_i are strong Benford and are therefore Benford in each $(j - k)$th digit, this occurs a vanishing percentage of the time. $\quad\square$

We thus turn our attention to estimating the number of fragments generated by a piece of length ℓ. The two following results give the necessary approximations.

Lemma 3.4. *Let F_ℓ denote the number of fragments generated by a piece of length ℓ. Then as $\ell \to \infty$,*

$$\mathbb{P}\left((\log(\log(\ell)))^2 < F_\ell < \log(\ell)g(\ell)\right) = 1 - o(1). \tag{3.6}$$

Corollary 3.5. *Define ℓ_k to be the length of the rod after k iterations of the process. Consider $Y' := \{y_k : \ell_k \gg \log^3(\ell)\}$. Then with probability tending to 1,*

$$\lim_{\ell \to \infty} \frac{|Y'|}{|Y|} = 1. \tag{3.7}$$

We show how Corollary 3.5 follows from 3.4.

Proof of Corollary 3.5. From the upper bound on F_ℓ, $|X \setminus X'| \ll \log(\log^3(\ell))g(\log^3(\ell))$. Thus employing our lower bound on Y directly,

$$\frac{|X \setminus X'|}{|X|} \ll \frac{\log(\log(\ell))g(\log^3(\ell))}{(\log(\log(\ell)))^2} \ll \frac{g(\log^3(\ell))}{\log(\log(\ell))}. \tag{3.8}$$

Taking g such that $g(u) = o(\log(u))$ completes the proof. $\square$

All that remains is to prove Lemma 3.4.

Proof of Lemma 3.4. We first prove the upper bound. We do this by finding the expected value of F_ℓ and applying Markov's inequality. We take the inductive hypothesis that $\mathbb{E}[F_1] = \sum_{j \le \ell} \frac{1}{j}$, with the base case $\mathbb{E}[F_1] = 1$ being clear. Since in the first break we get a single piece of length $\ell - s$ for some $s > 0$, we have the recurrence relation

$$\mathbb{E}[F_\ell] = \frac{1}{\ell} + \frac{1}{\ell} \sum_{s < \ell} (1 + \mathbb{E}[F_s]). \tag{3.9}$$

Therefore, by the inductive hypothesis, we have that

$$\mathbb{E}[F_\ell] = \frac{1}{\ell} + \frac{1}{\ell} \sum_{s=1}^{\ell-1} \left(1 + \sum_{i=1}^{s} \frac{1}{i}\right). \tag{3.10}$$

Each summand of the form $1/i$ appears in $\ell - i$ of the sums, so that

$$\frac{1}{\ell} + \frac{1}{\ell} \sum_{s=1}^{\ell-1} \left(1 + \sum_{i=1}^{s} \frac{1}{i}\right) = \frac{1}{\ell} + \frac{1}{\ell} \sum_{i=1}^{\ell-1} \frac{i + (\ell - i)}{i} = \sum_{i=1}^{\ell} \frac{1}{i} \sim \log(\ell) + O(1). \tag{3.11}$$

By Markov's inequality, $P(F_\ell > \log(\ell)g(\ell)) = O\left(\frac{1}{g(\ell)}\right)$.

We now prove the lower bound. The probability any of $\log(\log(\ell))^2$ breaks of a piece of length greater than $\ell^{1/2}$ at $1/(\log(\ell))$ of its original length is $o(1)$, since

$$\lim_{\ell \to \infty} \left(1 - \frac{1}{\log(\sqrt{(\ell)})}\right)^{\log(\log(\ell))^2} = \lim_{\ell \to \infty} e^{\frac{2\log(\log(\ell))^2}{\log(\ell)}} = 1. \tag{3.12}$$

For ℓ sufficiently large, if $\log(\log(\ell))^2$ breaks happen, none of which cut a piece down by a factor of $1/\log(\ell)$, then the piece is of length greater than $\ell^{1/2}$. This follows from,

$$\ell \cdot \left(\frac{1}{\log(\ell)}\right)^{\log(\log(\ell))^2} \gg \ell^{1/2}. \quad \square \tag{3.13}$$

Proof of Theorem 1.7. The theorem follows by considering the subset X', defined earlier as the set of all X_k such that $\ell_k \ge \log^3(\ell)$, which is strong Benford by Lemma 3.3 and Corollary 3.2. Since with probability tending to one this X' is almost all of X (that is as $\ell \to \infty$, we have $\frac{|X'|}{|X|} \to 1$), X also exhibits strong Benford behavior asymptotically in ℓ with probability tending toward 1. $\square$

4. Proof of Theorem 1.9: Restricted 1-dimensional decomposition

As the proof is similar to that of Theorem 1.5, we just highlight the differences below.

Proof of Theorem 1.9. 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 $\log N / N \to 0$ as $N \to \infty$), 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 $N \to \infty$ we show that $\mathbb{E}[P_s(N)] \to \log_{10} s$ and $\mathrm{Var}\,(P_s(N)) \to 0$. The first follows identically as in Section 2.1. We have

$$\mathbb{E}[\varphi_s(x_i)] = \int_{p_1=0}^{1} \int_{p_2=0}^{1} \cdots \int_{p_i=0}^{1} \varphi\left((1-p_i)\prod_{l=1}^{i-1} p_l\right) f(1-p_i)\prod_{l=1}^{i-1} f(p_l)dp_1 dp_2 \cdots dp_i \qquad (4.1)$$

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

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

$$\mathrm{Var}\,(P_N(s)) = \frac{\log_{10} s}{N^2} + \frac{1}{N^2}\left(\sum_{\substack{i,j=1 \\ i \neq j}}^{N} \mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)]\right) - \log_{10}^2 s + o(1). \qquad (4.2)$$

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

$$\mathcal{A} := \{(i, j) : \log N \leq i \leq j - \log N \leq N - \log N\}. \qquad (4.3)$$

We may replace the sum in (4.2) 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

$$2 \sum_{(i,j)\in\mathcal{A}} \mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)], \qquad (4.4)$$

where

$$X_i = p_i \prod_{r=1}^{i-1}(1-p_r)$$

$$X_j = p_j \prod_{r=1}^{i-1}(1-p_r)\prod_{r=i}^{j-1}(1-p_r); \qquad (4.5)$$

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

$$\mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)] = \int_{p_1=0}^{1} \cdots \int_{p_i=0}^{1} \cdots \int_{p_j=0}^{1} \varphi_s\left(p_i\prod_{r=1}^{i-1}(1-p_r)\right)$$

$$\cdot \varphi_s\left(p_j\prod_{r=1}^{i-1}(1-p_r)\prod_{r=i}^{j-1}(1-p_r)\right) \cdot \prod_{r=1}^{j} f(p_r)dp_1 \cdots dp_j. \qquad (4.6)$$

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

$$\mathcal{L}_i := \prod_{r=1}^{i-1}(1-p_r), \qquad \mathcal{L}_j = \prod_{r=i+1}^{j-1}(1-p_r). \qquad (4.7)$$

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

$$\int_{p_1=0}^{\infty} \cdots \int_{p_j=0}^{\infty} \varphi_s\left(\mathcal{L}_i p_i\right)\varphi_s\left(\mathcal{L}_i(1-p_i)\mathcal{L}_j p_j\right) \cdot \prod_{r=1}^{j} f(p_r)dp_1 \cdots dp_j \qquad (4.8)$$
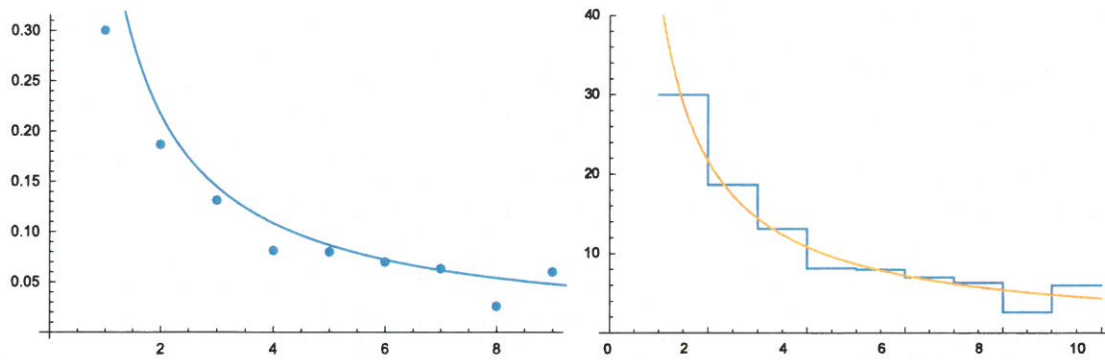
is close to $\log_{10}^2 s$.

Fig. 2. Irrational case: $p = 3/11$, 1000 levels; $y = \log_{10}(8/3) \notin \mathbb{Q}$.
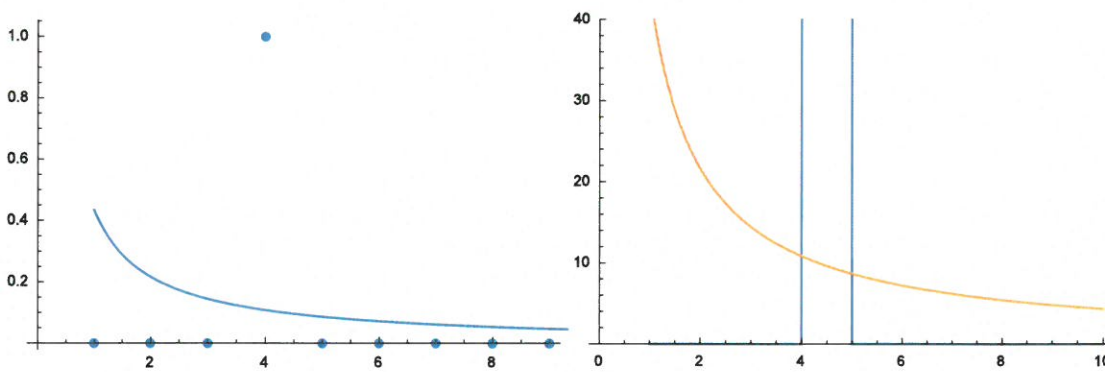

Fig. 3. Rational case: $p = 1/11$, 1000 levels; $y = 1 \in \mathbb{Q}$.

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 [45] (or see Appendix A of [84]) 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 Section 2 shows that the integral is close to $\log^2_{10}s$ 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 (4.2), we see the variance tends to 0. Thus the distribution of the leading digits converges in distribution to Benford's Law. $\square$

5. Proof of Theorem 1.11: Fixed 1-dimensional 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. We see this clearly in the three examples in Figs. 2 through 4, where we show observed behavior plotted against Benford behavior.

5.1. 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

$$x_{j+q} = \left(\frac{1-p}{p}\right)^q x_j = (10^y)^q x_j = 10^r x_j, \tag{5.1}$$

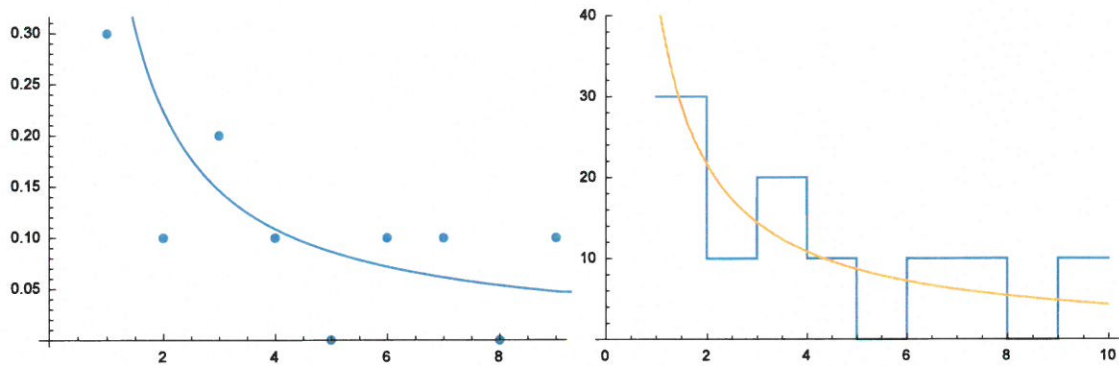**Fig. 4.** Rational case: $p = 1/(1 + 10^{33/10})$, 1000 levels; $y = 33/10 \in \mathbb{Q}$.

it follows that

$$S_{10}(x_{j+q}) = S_{10}(x_j). \tag{5.2}$$

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 $N \to \infty$.

To do this, we use the multisection formula. Given a power series $g(x) = \sum_{k=0}^{\infty} a_k x^k$, we can take the multisection $\sum_{\ell=0}^{\infty} a_{\ell q+j} x^{\ell q+j}$, where j and q are integers with $0 \le j < q$. The multisection itself is a power series that picks out every qth term from the original series, starting from the jth term. We have a closed expression for the multisection in terms of the original function (see [85] for a proof of this formula):

$$\sum_{\ell=0}^{\infty} a_{\ell q+j} x^{\ell q+j} = \frac{1}{q} \sum_{s=0}^{q-1} \omega^{-js} g(\omega^s x), \tag{5.3}$$

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

$$\sum_{\ell} \binom{N}{\ell q + j} = \frac{2^N}{q} \sum_{s=0}^{q-1} \left(\cos \frac{\pi s}{q} \right)^N \cos \frac{\pi(N - 2j)s}{q}; \tag{5.4}$$

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

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

$$\frac{1}{2^N} \left[\binom{N}{j} + \binom{N}{j+q} + \binom{N}{j+2q} + \cdots \right] = \frac{1}{q} \sum_{s=0}^{q-1} \left(\cos \frac{\pi s}{q} \right)^N \cos \frac{\pi(N - 2j)s}{q}$$

$$= \frac{1}{q} \left(1 + \sum_{s=1}^{q-1} \left(\cos \frac{\pi s}{q} \right)^N \cos \frac{\pi(N - 2j)s}{q} \right)$$

$$= \frac{1}{q} \left(1 + \mathrm{Err} \left[(q - 1) \left(\cos \frac{\pi}{q} \right)^N \right] \right), \tag{5.5}$$

---

[7] We 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 qth length tells us that for each digit d the probability of a first digit d is a/q for some $a \in \mathbb{N}$.

where 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 Nth power, it approaches 0 exponentially fast. As $N \to \infty$, 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. □

Remark 5.1. Instead of using the multisection formula, we could use the monotonicity (as we move toward or away from 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 1.11 when the irrationality exponent of y is finite.

5.2. 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 [42,80]). The frequency of the lengths x_j follows 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

$$x_\ell := \frac{N}{2} + \ell N^\delta, \quad x_{\ell,i} = \frac{N}{2} + \ell N^\delta + i, \quad I_\ell := \{x_\ell, x_\ell + 1, \ldots, x_\ell + N^\delta - 1\}. \quad (5.6)$$

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

5.2.1. Truncation

Instead of considering the entire binomial distribution, for any $\epsilon > 0$ we show that 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 Inequality[8] gives that the proportion of the density that is beyond N^ϵ standard deviations of the mean is

$$\mathrm{Prob}\left(\left|x - \frac{N}{2}\right| \geq N^\epsilon \cdot \sqrt{N}/2\right) \leq \frac{1}{N^{2\epsilon}}. \quad (5.7)$$

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 $N^{\frac{1}{2}-\delta+\epsilon}$.

5.2.2. 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;

---

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

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

$$
\begin{aligned}
\binom{N}{x_\ell} - \binom{N}{x_{\ell+1}} &= \binom{N}{\frac{N}{2} + \ell N^\delta} - \binom{N}{\frac{N}{2} + (\ell+1)N^\delta} \\
&= \frac{N!}{(\frac{N}{2} + \ell N^\delta)!(\frac{N}{2} - \ell N^\delta)!} - \frac{N!}{(\frac{N}{2} + (\ell+1)N^\delta)!(\frac{N}{2} - (\ell+1)N^\delta)!} \\
&= \frac{N!(\frac{N}{2} + (\ell+1)N^\delta)!(\frac{N}{2} - (\ell+1)N^\delta)! - N!(\frac{N}{2} + \ell N^\delta)!(\frac{N}{2} - \ell N^\delta)!}{(\frac{N}{2} + \ell N^\delta)!(\frac{N}{2} - \ell N^\delta)!(\frac{N}{2} + (\ell+1)N^\delta)!(\frac{N}{2} - (\ell+1)N^\delta)!} \\
&= \binom{N}{\frac{N}{2} + \ell N^\delta}\left[1 - \frac{(\frac{N}{2} + \ell N^\delta)!(\frac{N}{2} - \ell N^\delta)!}{(\frac{N}{2} + (\ell+1)N^\delta)!(\frac{N}{2} - (\ell+1)N^\delta)!}\right] \\
&= \binom{N}{x_\ell}\left[1 - \frac{(\frac{N}{2} + \ell N^\delta)!(\frac{N}{2} - \ell N^\delta)!}{(\frac{N}{2} + (\ell+1)N^\delta)!(\frac{N}{2} - (\ell+1)N^\delta)!}\right]. \quad (5.8)
\end{aligned}
$$

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 $\alpha_{\ell;N} = \frac{(\frac{N}{2} + \ell N^\delta)!(\frac{N}{2} - \ell N^\delta)!}{(\frac{N}{2} + (\ell+1)N^\delta)!(\frac{N}{2} - (\ell+1)N^\delta)!}$. We show that $1 - \alpha_{\ell;N} \to 0$, which implies the probabilities do not change significantly over an interval. We have

$$
\begin{aligned}
\alpha_{\ell;N} &= \frac{(\frac{N}{2} + \ell N^\delta)!(\frac{N}{2} - \ell N^\delta)!}{(\frac{N}{2} + (\ell+1)N^\delta)!(\frac{N}{2} - (\ell+1)N^\delta)!} \\
&\geq \frac{(\frac{N}{2} - (\ell+1)N^\delta)^{N^\delta}}{(\frac{N}{2} + (\ell+1)N^\delta)^{N^\delta}} \\
&= \left(\frac{1 - \frac{2(\ell+1)}{N^{1-\delta}}}{1 + \frac{2(\ell+1)}{N^{1-\delta}}}\right)^{N^\delta} \\
\log \alpha_{\ell;N} &\geq N^\delta \left[\log\left(1 - \frac{2(\ell+1)}{N^{1-\delta}}\right) - \log\left(1 + \frac{2(\ell+1)}{N^{1-\delta}}\right)\right]. \quad (5.9)
\end{aligned}
$$

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

$$
\begin{aligned}
\log \alpha_{\ell;N} &\geq N^\delta \left[-\frac{4(\ell+1)}{N^{1-\delta}} + O(N^{3\epsilon - 3/2})\right] \\
\alpha_{\ell;N} &\geq e^{-4(\ell+1)N^{2\delta-1} + O(N^{\delta+3\epsilon-3/2})} \\
&\geq 1 - 4(\ell+1)N^{2\delta-1} + O(N^{\delta+3\epsilon-3/2}) \\
1 - \alpha_{\ell;N} &\leq 4(\ell+1)N^{2\delta-1} + O(N^{\delta+3\epsilon-3/2}). \quad (5.10)
\end{aligned}
$$

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 (5.10) into (5.8) yields

$$
\begin{aligned}
\binom{N}{x_\ell} - \binom{N}{x_{\ell+1}} &= \binom{N}{\frac{N}{2} + \ell N^\delta}(1 - \alpha_{\ell;N}) \\
&\leq \binom{N}{x_\ell}\left(4(\ell+1)N^{2\delta-1} + O(N^{\delta+3\epsilon-3/2})\right). \quad (5.11)
\end{aligned}
$$

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

$$\left| \binom{N}{x_\ell} - \binom{N}{x_{\ell+1}} \right| \ll \binom{N}{x_\ell} \left(N^{-\frac{1}{2}+\delta+\epsilon} + O(N^{\delta+3\epsilon-3/2}) \right). \tag{5.12}$$

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

$$\left| \binom{N}{x_\ell} - \binom{N}{x_{\ell,i}} \right| \ll \binom{N}{x_\ell} N^{-\frac{1}{2}+\delta+\epsilon}. \tag{5.13}$$

5.2.3. 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

$$I_\ell := \left\{ x_{\ell,i} : \frac{N}{2} + \ell N^\delta, \ldots, \frac{N}{2} + (\ell+1)N^\delta - 1 \right\}, \tag{5.14}$$

we prove that $\log(x_{\ell,i})$ becomes equidistributed modulo 1 as $N \to \infty$. Fix an $(a, b) \subset (0, 1)$. Let $J_\ell(a, b) \subset \{0, 1, \ldots, N^\delta\}$ 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 [42,80]). Moreover, we can quantify the rate of equidistribution if the irrationality exponent κ of y is finite. From [86] we obtain a power savings:

$$|J_\ell(a, b)| = (b - a)N^\delta + O\left(N^{\delta(1-\frac{1}{\kappa}+\epsilon')} \right); \tag{5.15}$$

see [22] for other examples of systems (such as the $3x + 1$ map) where the irrationality exponent controls Benford behavior. The key idea is to keep approximating general sums with simpler ones that are tractable, with manageable errors at each step.

We now combine this quantified equidistribution with Lemma 5.2 and find (we divide by 2^N later, which converts these sums to probabilities)

$$
\begin{aligned}
\sum_{i \in J_\ell(a,b)} \binom{N}{x_{\ell,i}} &= \sum_{i \in J_\ell(a,b)} \left[\binom{N}{x_\ell} + O\left(\binom{N}{x_\ell} N^{-\frac{1}{2}+\delta+\epsilon} \right) \right] \\
&= \binom{N}{x_\ell} \left[\left(\sum_{i \in J_\ell(a,b)} 1 \right) + O\left(N^{-\frac{1}{2}+\delta+\epsilon} N^\delta \right) \right] \\
&= \binom{N}{x_\ell} \left[\left((b-a)N^\delta + O\left(N^{\delta(1-\frac{1}{\kappa}+\epsilon')} \right) \right) + O\left(N^{-\frac{1}{2}+2\delta+\epsilon} \right) \right] \\
&= (b-a)N^\delta \binom{N}{x_\ell} + N^\delta \binom{N}{x_\ell} \cdot O\left(N^{-\frac{1}{2}+\delta+\epsilon} + N^{-\delta(\frac{1}{\kappa}-\epsilon')} \right).
\end{aligned}
\tag{5.16}
$$

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 (5.5)) with $q = N^\delta$, and find

$$\sum_{\ell=-N/N^\delta}^{N/N^\delta} \binom{N}{x_\ell} = \frac{2^N}{N^\delta} \left(1 + \mathrm{Err}\left[N^\delta \left(\cos \frac{\pi}{N^\delta} \right)^N \right] \right) = \frac{2^N}{N^\delta} + O\left(2^N \cdot e^{-3N^{1-2\delta}} \right), \tag{5.17}$$

where 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:

$$\log\left(\cos(\pi/N^\delta)\right)^N = N\log\left(1 - \frac{1}{2!}\frac{\pi^2}{N^{2\delta}} + O(N^{-4\delta})\right)$$

$$= N\left(-\frac{1}{2!}\frac{\pi^2}{N^{2\delta}} + O(N^{-4\delta})\right) = -\frac{\pi^2}{2}N^{1-2\delta} + O(N^{1-4\delta})$$

$$\cos(\pi/N^\delta) \le e^{-3N^{1-2\delta}} \tag{5.18}$$

for N large.

Remember, though, that we are only supposed to sum over $|\ell| \le N^{\frac{1}{2}-\delta+\epsilon}$. The contribution from the larger $|\ell|$, however, was shown to be at most $O(N^{-2\epsilon})$ in Section 5.2.1, and thus we find

$$\frac{1}{2^N}\sum_{\ell=-N^{\frac{1}{2}-\delta+\epsilon}}^{N^{\frac{1}{2}-\delta+\epsilon}} (b-a)N^\delta\binom{N}{x_\ell} = 1 + O\left(N^\delta e^{-3N^{1-2\delta}} + N^{-2\epsilon}\right). \tag{5.19}$$

As this is of size 1, the lower order terms in (5.16) do not contribute to the main term (their contribution is smaller by a power of N).

We can now complete the proof of Theorem 1.11 when $y \notin \mathbb{Q}$ has finite irrationality exponent. Convergence in distribution to Benford's law is equivalent to showing, for any $(a, b) \subset (0, 1)$, that

$$\sum_{\ell=-N/N^\delta}^{N/N^\delta} \sum_{\substack{i=0 \\ S_{10}(x_{\ell,i})\in(a,b)}}^{N_\delta-1} \frac{\binom{N}{x_{\ell,i}}}{2^N} = b - a + o(1); \tag{5.20}$$

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 1.11. $\square$

Remark 5.3. A more careful analysis allows one to optimize the error. We want $\delta\left(-\frac{1}{\kappa} + \epsilon'\right) = -\frac{1}{2} + \delta + \epsilon$, 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.

5.3. 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 $O(N^{-\eta})$.

6. Two-dimensional fragmentation

In this section we prove Theorems 1.12 and 1.15, extending our results to two dimensions.

6.1. Decomposition of triangles

Proof of Theorem 1.12. Considering at each stage of the decomposition the proportions $p^{(1)}, p^{(2)}, p^{(3)}$ of the resultant sub-triangles to their parent, we obtain a sequence of random variables $\{(p_i^{(1)}, p_i^{(2)},$
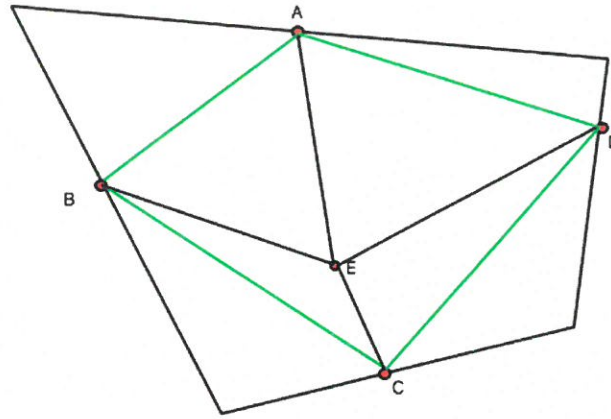
Fig. 5. Points A, B, C and D are chosen by taking random draws from a continuous distribution f and placing the point that fraction of the way along the interval. A point is then selected from the unit square independently, and E is chosen based on the mapping of this point under a fixed, continuous measure preserving map from the unit square to the green region. This corresponds to a special case of the one-dimensional fragmentation and combination process described in Lemma 6.1, in which the rod is first split into lengths corresponding to the area of the region enclosed in green and the sum of the outer regions. The inner piece is split into four triangular pieces, which we can consider as a split of the rod into four pieces with length corresponding to the areas of the triangles, and similarly for the outer piece. The inner and outer pieces are then combined in pairs to reform the quadrilaterals outlined in black.

$p_i^{(3)}$)$\}_{i \in \mathbb{N}}$. Since affine transformations preserve ratios of areas, the probability density g on $(0, 1)^3$ of the proportions $p^{(1)}, p^{(2)}, p^{(3)}$ is constant with respect to the shape of the triangle and depends only on f. We may therefore reframe this two-dimensional process in terms of our one-dimensional rod decomposition, where at each step the rod is broken into three pieces with proportions $p^{(1)}, p^{(2)}, p^{(3)}$ according to the distribution g.

We give our initial triangle $T_{45-45-90}$ vertices at $(0, 0), (0, 1), (1, 0)$. If we select the point (x, y) in the interior of our triangle, the formula area $=$ base $\times$ height gives that the sub-triangles have areas $\frac{1}{2}x, \frac{1}{2}y, \frac{1}{2}(1 - x - y)$. The proportions of the areas of the sub-triangles to the original triangle are therefore $x, y, 1 - x - y$ (in particular, if f is uniform, then so is g.) These have distribution $f^{(1)}, f^{(2)}, 1 - f^{(1)} - f^{(2)}$, for $f^{(1)}, f^{(2)}$ as in the statement of this theorem, and so Theorem 1.12 now follows directly from Theorem 1.5. $\square$

6.2. Decomposition of quadrilaterals

We now consider our model for decomposing convex quadrilaterals. We will describe the construction process in detail below, illustrating it in Fig. 5.

We first set some notation, and then state and prove a lemma which allows us to reduce the 2-dimensional fragmentation to 1-dimensional problems that our techniques can handle.

Let

$$S^k := \{\mathbf{x} \in (0, 1)^k : x_1 + \cdots + x_k = 1\}, \tag{6.1}$$

and for any probability distribution f on S^k let $\tilde{f}$ be its first marginal, i.e.,

$$\tilde{f}(x) = \int_{y_2=0}^1 \cdots \int_{y_k=0}^1 f(x, y_2, \ldots, y_k) dy_2 \cdots dy_k. \tag{6.2}$$

A key lemma is a variant of the unrestricted fragmentation process in which pieces are allowed to rejoin after breaks. We state and prove it in greater generality than is needed for our construction, as the additional freedom may be of independent interest and is no harder to state and prove.

Lemma 6.1. *Fix a probability density π on $(0, 1)$, a probability density f on S^k, and a one-parameter probability density $G(\theta)$ on S^k depending on a single random variable θ such that (6.3) holds. Given a stick*

of length L, we perform the following iterated process.

1. *Break the stick at a proportion p chosen according to π.*
2. *Break one of the pieces into k sub-pieces $X_1, \ldots, X_k$ with proportions chosen according to f.*
3. *Independently break the other piece into k sub-pieces $Y_1, \ldots, Y_k$ with proportions according to G(p).*
4. *Combine each piece X_i with the piece Y_i, so that we have k pieces of lengths $X_1 + Y_1, \ldots, X_k + Y_k$.*
5. *Repeat this process independently on each of the k resulting sub-pieces until there are k^N pieces $\mathcal{X}$.*

If

$$\lim_{N \to \infty} \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} \left(\int_0^\infty \int_0^1 \int_0^1 \int_0^1 \pi(p)\tilde{f}(y)\tilde{G}(p;z)\tilde{f}\left(\frac{x-pz}{(1-p)y}\right) x^{-2\pi i \ell/\log 10} dz\,dy\,dp\,dx \right)^N = 0, \quad (6.3)$$

then as $N \to \infty$ the significands of the lengths in $\mathcal{X}$ converge in distribution to Benford's law.

Proof of Lemma 6.1. For readability of the proof, we assume that our distributions $f, G(\theta)$ are symmetric; the general case follows similarly. Let φ_s denote the significand indicator function, so that

$$\varphi_s(x) := \begin{cases} 1 & \text{if the significand of } x \text{ is at most } s \\ 0 & \text{otherwise.} \end{cases} \quad (6.4)$$

Let $P_N(s)$ denote the proportion of pieces in $\mathcal{X}$ the significand of whose length is at most s. That is,

$$P_N(s) := \frac{1}{|\mathcal{X}|} \sum_{X \in \mathcal{X}} \varphi_s(X). \quad (6.5)$$

We now show that

$$\lim_{N \to \infty} \mathbb{E}[P_N(s)] = \log_{10}(s) \quad (6.6)$$

and

$$\lim_{N \to \infty} \text{Var}(P_N(s)) = 0, \quad (6.7)$$

which together suffice to prove our theorem. The proof of (6.6) proceeds as in the proof of Theorem 1.5 except that we apply Theorem 1.1 of [45] with $f_{D(\theta)} = 1/(\theta + p/4)$ as opposed to $f_{D(\theta)} = 1/\theta$. We therefore concentrate our attention on showing (2.9).

We have

$$\begin{aligned} \text{Var}(P_N(s)) &= \mathbb{E}[P_N^2(s)] - \mathbb{E}[P_N(s)]^2 \\ &= \frac{1}{4^{2N}} \sum_{X,Y \in \mathcal{X}} \mathbb{E}[\varphi_s(X)\varphi_s(Y)] - \mathbb{E}[P_N(s)]^2 \\ &= \frac{1}{4^N} E[P_N(s)] + \frac{1}{4^{2N}} \sum_{X \neq Y \in \mathcal{X}} \mathbb{E}[\varphi_s(X)\varphi_s(Y)] - E[P_N(s)]^2 \\ &= \frac{1}{4^{2N}} \sum_{X \neq Y \in \mathcal{X}} \mathbb{E}[\varphi_s(X)\varphi_s(Y)] - \log_{10}(s) + o(1). \end{aligned} \quad (6.8)$$

Now, consider two arbitrary pieces $X \neq Y \in \mathcal{X}$. Call one iteration of the process 1–5 a "cut", each piece in $\mathcal{X}$ being the result of a sequence of N cuts. We can express X, Y each as a product of terms of the form

$$p_i \alpha_i + (1 - p_i)\beta_i \quad (6.9)$$

representing the ith cut, where p_i is the proportion chosen in step 1 of our iterated process, α_i is the length of one of the pieces X_j obtained in step 2, and β_i is the length of one of the pieces Y_i obtained

in step 3. Since the sequences of cuts that produced X and Y may be the same up to some $k < N$, we express X, Y more precisely by

$$
\begin{aligned}
X &= L \left(p_1\alpha_1 + (1-p_1)\beta_1\right) \cdots \left(p_{k-1}\alpha_{k-1} + (1-p_{k-1})\beta_{k-1}\right) \left(p_k\alpha_k + (1-p_k)\beta_k\right) \\
&\qquad \left(p_{k+1}\alpha_{k+1} + (1-p_{k+1})\beta_{k+1}\right) \cdots \left(p_N\alpha_N + (1-p_N)\beta_N\right) \\
&=: La_1 \cdots a_{k-1} \left(p_k\alpha_k + (1-\beta_k)\beta_k\right) a_{k+1} \cdots a_N \\
&=: L_1 a_{k+1} \cdots a_N
\end{aligned}
$$

$$
\begin{aligned}
Y &= L \left(p_1\alpha_1 + (1-p_1)\beta_1\right) \cdots \left(p_{k-1}\alpha_{k-1} + (1-p_{k-1})\beta_{k-1}\right) \left(p_k\alpha'_k + (1-p_k)\beta'_k\right) \\
&\qquad \left(p_{k+1}\tilde{\alpha}_{k+1} + (1-p_{k+1})\tilde{\beta}_{k+1}\right) \cdots \left(p_N\tilde{\alpha}_N + (1-p_N)\tilde{\beta}_N\right) \\
&=: La_1 \cdots a_{k-1} \left(p_k\alpha'_k + (1-p_k)\beta'_k + \frac{p}{4}\right) \tilde{a}_{k+1} \cdots \tilde{a}_N \\
&=: L_2 \tilde{a}_{k+1} \cdots \tilde{a}_N
\end{aligned}
\tag{6.10}
$$

where $a_i, \tilde{a}_i$ are independent random variables. We seek to describe their distribution.

Each $\alpha_i, \tilde{\alpha}_i$ is some component of an independent random variable chosen according to f; by symmetry, we may suppose it to be the first coordinate. Thus $\alpha_i, \tilde{\alpha}_i$ are independent random variables with distribution $\tilde{f}$. Likewise, the $\beta_i, \tilde{\beta}_i$ are independent variables chosen according to $\tilde{G}(p_i)$. However, α'_i and β'_i are not independent. The distribution of the $a_i, \tilde{a}_i$ has density

$$
h(x) = \int_0^1 \phi(p) \int_0^1 \tilde{f}(y) \int_0^1 \tilde{G}(p; z) \tilde{f}\left(\frac{x - pz}{(1-p)y}\right) dz\,dy\,dp.
\tag{6.11}
$$

As in Section 2, we write

$$
\begin{aligned}
I(L_1) &= \int_{a_{k+1}=0}^1 \cdots \int_{a_N=0}^1 \varphi_s \left(L_1 \prod_{i=k+1}^N a_i\right) \prod_{i=k+1}^N h(a_i)\,da_{k+1} \cdots da_N \\
I(J_2) &= \int_{\tilde{a}_{k+1}=0}^1 \cdots \int_{\tilde{a}_N=0}^1 \varphi_s \left(L_2 \prod_{i=k+1}^N \tilde{a}_i\right) \prod_{i=k+1}^N h(\tilde{a}_i)\,d\tilde{a}_{k+1} \cdots d\tilde{a}_N
\end{aligned}
\tag{6.12}
$$

so that

$$
\mathbb{E}[\varphi_s(X)\varphi(Y)] = \int_{a_1=0}^1 \cdots \int_{a_{k-1}=0}^1 \int_{\alpha_k=0}^1 \int_{\beta_k=0}^1 I(L_1)J(L_2)\,da_1 \cdots da_{k-1}\,d\alpha_k\,d\beta_k.
\tag{6.13}
$$

It suffices to show that, for any L_1, L_2, we have $|I(L_1)J(L_2) - (\log_{10}s)^2| = o(1)$ in N. This is proven in Section 2 if h satisfies the Mellin condition (6.3). In other words if

$$
\lim_{N \to \infty} \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} \left(\int_0^{\infty} \int_0^1 \int_0^1 \int_0^1 \pi(p)\tilde{f}(y)\tilde{G}(p; z)\tilde{f}\left(\frac{x - pz}{(1-p)y}\right) x^{-2\pi i\ell/\log 10}\,dz\,dy\,dp\,dx \right)^N = 0,
\tag{6.14}
$$

which is precisely (6.3) on π, f, G. $\quad\square$

We can now proceed to the proof of Benford behavior in the quadrilateral decomposition model.

Proof of Theorem 1.15. Unlike in the triangular case, affine transformations do not map between arbitrary quadrilaterals, since affine transformations must preserve parallelism. However, there are numerous continuous mappings from arbitrary convex quadrilaterals to squares that preserve the ratio of areas; the construction of Gromov after Knothe provides such a mapping with other nice properties [83]. The idea of the mapping is to send slices of the first quadrilateral to slices of the second in a continuous way; see [87] for details on the specific construction. Therefore, given a probability density f on the square, we obtain a unique probability density f' on any convex quadrilateral, and if f is uniform then so is f'.

We now formulate our decomposition process in precise terms. We begin with the unit square, a probability density f on $(0, 1)$, and a probability density g on $(0, 1)^2$. We independently select a point on each side according to f. Call these A, B, C, D. We then choose a point E in the interior of the quadrilateral $ABCD$ according to the composition of g with a mapping that preserves ratios of areas. We now connect E to each of A, B, C, D in order to decompose the square into four convex quadrilaterals. We then perform the same decomposition independently on each of these quadrilaterals, repeating this process until we obtain 4^N quadrilaterals.

Again, we consider the ratios of the areas of the sub-quadrilaterals to their parent, and use the resulting distribution to reframe this two-dimensional process as a decomposition of the rod. Because of the existence of continuous mappings that preserve ratios of areas, it is enough to consider the decomposition of the unit square. Each sub-quadrilateral can be thought of as having two triangular components: one outside $ABCD$ and one inside it. We will determine the distribution of the proportions of the sub-quadrilaterals by conditioning on the area of the inner quadrilateral $ABCD$. In terms of the decomposition of the rod, this corresponds to the following case of Lemma 6.1. We split the rod into two pieces, one corresponding to the inner quadrilateral $ABCD$ and one corresponding to the outer area. We then split the former piece according to the areas of the inner triangles, and the outer piece according to the areas of the outer triangles, and then recombine the corresponding sub-pieces.

If (a, b, c, d) are the proportions of the points on each side, then the outer triangles have areas $\frac{1}{2}a(1-b)$, $\frac{1}{2}b(1-c)$, $\frac{1}{2}c(1-d)$, $\frac{1}{2}d(1-a)$ and the quadrilateral $ABCD$ has area $\mathcal{A} = 1 - \frac{1}{2}(a(1-b) + b(1-c) + c(1-d) + d(1-a))$. Let π denote the density of the probability distribution of $\mathcal{A}$. Let $h_1(\mathcal{A})$ denote the density of the distribution of the areas of the outer triangles conditional on $\mathcal{A}$. Note that if f is continuous, then so are π and h_1. Let h_2 denote the density of the distribution of the ratios of the areas of the inner triangles to $\mathcal{A}$. Pulling back using any continuous area preserving mapping, we may determine the distribution of h_2 by considering the case when $ABCD$ is the unit square. In this case, by selecting the point (x, y) we produce triangles of area $\frac{1}{2}xy$, $\frac{1}{2}x(1-y)$, $\frac{1}{2}(1-x)y$, $\frac{1}{2}(1-x)(1-y)$. The areas of the inner triangles in $ABCD$ are obtained by multiplying through by $\mathcal{A}$. Therefore, if f and g are both continuous, we see that h_2 is continuous.

Theorem 1.15 follows from applying Lemma 6.1 with this π above corresponding to the π of the lemma and h_1 and h_2 corresponding to G and f, respectively.   □

7. Proof of Theorem 1.17: 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. See [88] for additional examples of Benford behavior in matrix ensembles.

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 ith 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 $\{1, \ldots, n\}$.

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

We first quantify the degree of dependencies, and then sketch the proofs of the mean and variance. Fix $i \in \{1, \ldots, n!\}$ 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 $1, 2, \ldots, n$, how likely is it that exactly k are returned to their original starting position? This is the well-known probléme 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 [89]). 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

$$\mathbb{E}[P_n(s)] = \frac{1}{n!} \sum_{i=1}^{n!} \mathbb{E}[\varphi_s(X_{i,n})]. \tag{7.1}$$

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

$$\mathbb{E}[\varphi_s(X_{i,n})] = \int_{a_{1\sigma_i(1)}} \int_{a_{2\sigma_i(2)}} \cdots \int_{a_{n\sigma_i(n)}} \varphi_s \left(\prod_{p=1}^{n} a_{p\sigma_i(p)} \right)$$
$$\cdot \prod_{p=1}^{n} f(a_{p\sigma_i(p)}) \, da_{1\sigma_i(1)} da_{2\sigma_i(2)} \cdots da_{n\sigma_i(n)}, \tag{7.2}$$

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 [45] (or see Appendix A of [84]).

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

$$\mathrm{Var}\,(P_n(s)) = \frac{\log_{10}(s)}{n!} + \frac{1}{(n!)^2} \left(\sum_{\substack{i,j=1 \\ i\neq j}}^{n!} \mathbb{E}[\varphi_s(X_{i,n})\varphi_s(X_{j,n})] \right) - \log_{10}^2 s + o(1). \tag{7.3}$$

We then mimic the proof from Section 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.

8. Future work

Many of our results concern 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 8.1. *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 [33,90]), Theorem 1.9 is essentially a continuous analogue of this problem when $a_n = 2n$, and thus provides compelling support.
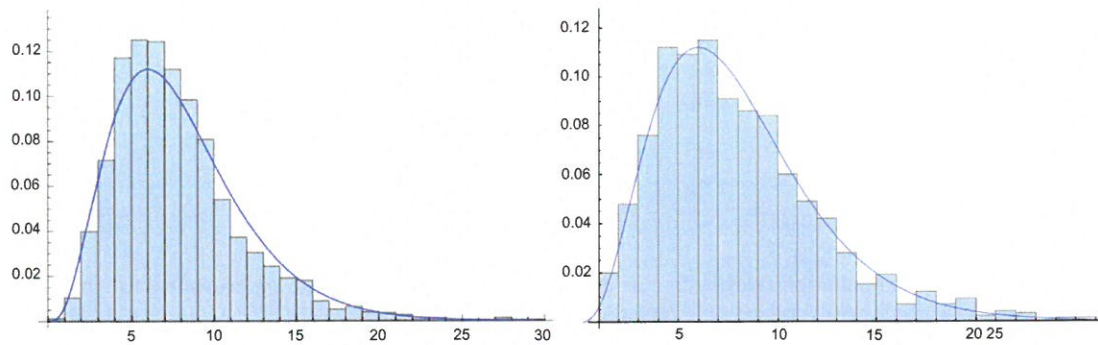
Fig. 6. (a) $\{a_n = 2n\}$: 2500 chi-square values of stick lengths to Benford behavior after decomposition of sticks of length approximately 10^7. (b) $\{a_n = p_n\}$ (the set of primes): 1000 chi-square values of stick lengths to Benford behavior after decomposition of sticks of length approximately 10^{500}.

Fig. 6(a) shows a histogram of chi-square values from numerical simulations with a χ_8^2 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 value[9] 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 8.1(i). Similarly, Fig. 6(b), which shows the chi-square values obtained when $\{a_n\}$ is the set of all primes, supports Conjecture 8.1(ii).

Conjecture 8.2. *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 nth Fibonacci number.*

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.

Fig. 7 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 8.2 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.

Acknowledgments

---

[9] If X_i are the simulation frequencies of each digit $i = 1, 2, \ldots, 9$ and Y_i are the frequencies of each digit as predicted by Benford's Law, where $i = 1, \ldots, 9$, then the chi-squared value we calculated is given by $\sum_{i=1}^{9}(X_i N - Y_i N)^2 / N Y_i$, where N is the number of pieces.
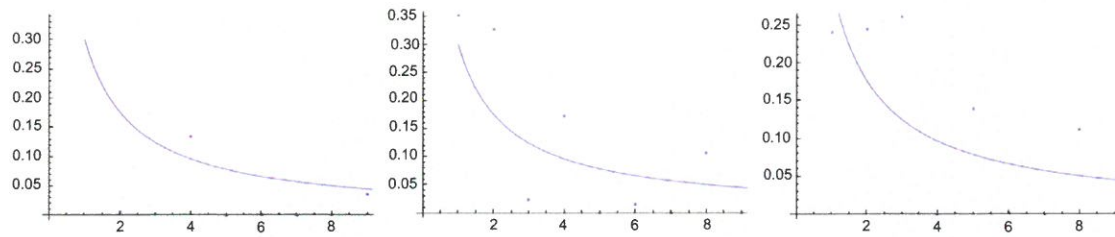
Fig. 7. (a) $\{a_n = n^2\}$: Distribution of stick lengths after decomposing sticks of initial length approximately 10^6. (b) $\{a_n = 2^n\}$: Distribution of stick lengths after decomposing 1000 sticks of initial length approximately 10^{500}. (c) $\{a_n = F_n\}$, the Fibonacci Numbers: Distribution of stick lengths after decomposing 500 sticks of initial 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 nth level are chosen from the nth 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

$$\epsilon_n \; < \; \min\left(\sqrt{\frac{3}{20\pi^2(n+1)^2}}, \frac{1}{2}\frac{\delta}{2^n}, \frac{\log\frac{1}{2}}{2}, \frac{1 - \log\frac{1}{2}}{2}\right); \tag{A.1}$$

these values and the meaning of these constants will be made clear during the construction. Consider the sequence of distributions with the following densities

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

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 nth iteration of the decomposition process are 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 nth Fourier coefficient is

$$\widehat{\phi}(n) \; := \; \int_0^1 \phi(x)e^{-2\pi inx}dx. \tag{A.3}$$

Miller and Nigrini [47] proved that if $R_1, \ldots, R_M$ 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 $\lim_{M \to 0} \widehat{g_1}(n) \cdots \widehat{g_M}(0) = 0$.

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

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

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

$$\left|\widehat{\phi_n}(1)\right| \geq \frac{n^2 + 2n}{(n+1)^2}. \tag{A.5}$$

As

$$\lim_{N \to \infty} \prod_{n=1}^{N} \frac{n^2 + 2n}{(n+1)^2} = \frac{1}{2}, \tag{A.6}$$

we see that

$$\lim_{N \to \infty} \prod_{n=1}^{N} \left|\widehat{\phi_1}(1) \cdots \widehat{\phi_N}(1)\right| > 0. \tag{A.7}$$

This argument shows that the product of N random variables, where the logarithm of the nth 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 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 Nth iteration of our cutting procedure by $X_{N,i}$, where $i = 1, \ldots, 2^N$. Each length is a product of N random variables; the nth 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 nth stage is $\epsilon_n + \log \frac{1}{2}$, its proportion at the nth level is $e^{\epsilon_n + \log \frac{1}{2}}$. Similarly $X_{N,j}$ always has the smallest cut, which at the nth level is $e^{-\epsilon_n + \log \frac{1}{2}}$. Therefore

$$\log \frac{X_{N,i}}{X_{N,j}} = \log \prod_{n=1}^{N} \frac{e^{\epsilon_n + \log \frac{1}{2}}}{e^{-\epsilon_n + \log \frac{1}{2}}} = \sum_{n=1}^{N} 2\epsilon_n; \tag{A.8}$$

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 [62] 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 $n = 2017$, but there are 85,345 if n is 2018. By considering a continuous system from the start, we avoid these Diophantine complications; see [75] for an analysis of the discrete setting.

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 [91–93]. 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 impact the behavior of the system. We mention a result from Miller–Nigrini [33], 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

[1] S. Newcomb, Amer. J. Math. 4 (1881) 39–40.
[2] F. Benford, Proc. Amer. Philos. Soc. 78 (1938) 551–572.
[3] A. Berger, T.P. Hill, Amer. Math. Monthly 114 (2007) 588–601.
[4] P.D. Drake, M.J. Nigrini, J. Acc. Educ. 18 (2000) 127–146.
[5] S.J. Miller, M. Nigrini, Auditing: J. Pract. Theory 28 (2) (2009) 305–324. http://dx.doi.org/10.2308/aud.2009.28.2.305.
[6] M. Nigrini, The Detection of Income Tax Evasion Through an Analysis of Digital Frequencies, (Ph.D. thesis), University of Cincinnati, OH, USA, 1992.
[7] M. Nigrini, White Paper 8 (2) (1994) 3–6.
[8] M. Nigrini, Proceedings of the 1996 Deloitte & Touche / University of Kansas Symposium on Auditing Problems, University of Kansas, Lawrence, KS, 1996, p. 6981.
[9] M. Nigrini, L. Mittermaier, Auditing: J. Prac. Theory 16 (2) (1997) 52–67.

---

[10] Imagine we toss a coin one million times, always getting either all heads or all tails. Let us 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%.

[10] E. Costasa, V. López-Rodasa, F.J. Torob, A. Flores-Moya, Aquatic Botany 89 (3) (2008) 341–343.
[11] M. Moret, V. de Senna, M. Santana, G. Zebende, Internat. J. Modern Phys. C 20 (12) (2009) 1981–1988.
[12] D. Knuth, The Art of Computer Programming, Volume 2: Seminumerical Algorithms, third ed., Addison-Wesley, MA, 1997.
[13] A. Berger, Discrete Contin. Dyn. Syst. 13 (2005) 219–237.
[14] A. Berger, Stoch. Dyn. 5 (2005) 587–607.
[15] A. Berger, Leonid A. Bunimovich, T. Hill, Trans. Amer. Math. Soc. 357 (1) (2005) 197–219.
[16] A. Berger, T.P. Hill, B. Kaynar, A. Ridder, SIAM J. Matrix Anal. Appl. 32 (3) 665–684.
[17] R. Rodriguez, Amer. Statist. (2004).
[18] K.-H. Tödter, German Econom. Rev. 10 (2009) 339–351.
[19] S.B. Villas-Boas, Q. Fu, G. Judge, Physica A (2017) 711–719.
[20] M. Nigrini, S.J. Miller, Math. Geol. 39 (5) (2007) 469–490.
[21] T. Anderson, L. Rolen, R. Stoehr, Proc. Amer. Math. Soc. 139 (5) (2011) 1533–1541.
[22] A. Kontorovich, S.J. Miller, Acta Arith. 120 (2005) 269–297.
[23] J. Lagarias, K. Soundararajan, J. London Math. Soc. Ser. 2 74 (2) (2006) 289–303.
[24] B. Buck, A. Merchant, S. Perez, Eur. J. Phys. 14 (1993) 59–63.
[25] I. Eliazar, Ann. Physics 380 (2017) 168–187.
[26] M. Moret, V. de Senna, M. Pereira, G. Zebende, Internat. J. Modern Phys. C 17 (11) (2006) 1597–1604.
[27] D. Ni, Z. Ren, Eur. Phys. J. A 38 (2008) 251–255.
[28] D. Ni, L. Wei, Z. Ren, Commun. Theor. Phys. 51 (2009) 713–716.
[29] L. Pietronero, E. Tosatti, V. Tosatti, A. Vespignani, Physica A (2001).
[30] L. Shao, B. Ma, Modern Phys. Lett. A 24 (40) (2009) 3275–3282.
[31] L. Shao, B. Ma, Astropart. Phys. 33 (2010) 255–262.
[32] F. Perez-Gonzalez, G.L. Heileman, C.T. Abdallah, Proceedings of the IEEE International Conference on Image Processing, ICIP, Vol. 1, San Antonio, TX, 2007, pp. 405–408.
[33] S.J. Miller, M. Nigrini, Int. J. Math. Math. Sci. 2008 (2008) 382948. http://dx.doi.org/10.1155/2008/382948. 19 pages.
[34] V. Cuff, A. Lewis, S.J. Miller, The Weibull distribution and Benfords Law preprint.
[35] W. Mebane, 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.
[36] A. Berger, T.P. Hill, *Benford Online Bibliography*, http://www.benfordonline.net.
[37] W. Hurlimann, Benfords Law from 1881 to 2006, http://arxiv.org/pdf/math/0607168.
[38] A. Berger, T.P. Hill, Fundamental Flaws in Fellers Classical Derivation of Benfords Law, University of Alberta, 2010 Preprint.
[39] A. Berger, T.P. Hill, Math. Intell. 33 (2011) 85–91.
[40] A. Berger, T.P. Hill, Probab. Surv. 8 (2011) 1–126.
[41] A. Berger, T.P. Hill, An Introduction To BenfordS Law, Princeton University Press, Princeton, 2015.
[42] P. Diaconis, Ann. Probab. 5 (1979) 72–81.
[43] T. Hill, Amer. Sci. 86 (1996) 358–363.
[44] T. Hill, Statist. Sci. 10 (1996) 354–363.
[45] D. Jang, J.U. Kang, A. Kruckman, J. Kudo, S.J. Miller, J. Algebra, Number Theory: Adv. Appl. 1 (1) (2009) 37–60.
[46] E. Janvresse, T. de la Rue, J. Appl. Probab. 41 (4) (2004) 1203–1210.
[47] S.J. Miller, M. Nigrini, Internat. J. Algebra 2 (3) (2008) 119–130.
[48] R. Pinkham, Ann. Math. Stat. 32 (4) (1961) 1223–1230.
[49] R.A. Raimi, Amer. Math. Monthly 83 (7) (1976) 521–538.
[50] S.J. Miller (Ed.), Theory and Applications of BenfordS Law, Princeton University Press, 2015, p. 438.
[51] A.K. Adhikari, Sankhyā 31 (1969) 413–420.
[52] A.K. Adhikari, B.P. Sarkar, Sankhyā 30 (1968) 47–58.
[53] R.N. Bhattacharya, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 25 (1972) 1–10.
[54] P. Lévy, Théorie de laddition des Variables Aléatoires, Gauthier-Villars, Paris, 1937.
[55] P. Lévy, Bull. De La S. M. F. 67 (1939) 1–41.
[56] H. Robbins, Proc. Amer. Math. Soc. 4 (1953) 786–799.
[57] H. Sakamoto, Tôhoku Math. J. 49 (1943) 243–260.
[58] P. Schatte, Math. Nachr. 110 (1983) 243–261.
[59] P. Schatte, Math. Nachr. 115 (1984) 275–281.
[60] P. Schatte, Math. Nachr. 127 (1986) 7–20.
[61] M.D. Springer, W.E. Thompson, SIAM J. Appl. Math. 14 (1966) 511–526.
[62] D.S. Lemons, Amer. J. Phys. 54 (9) (1986) 816–817.
[63] S. Kakutani, Measure Theory (Proc. Conf. Oberwolfach, 1975), in: Lecture Notes in Math., vol. 541, Springer, Berlin, 1976, pp. 369–375.
[64] R.L. Adler, L. Flatto, Z. Wahrscheinlichkeitstheor. Verwandte Geb 38 (4) (1977) 253–259.
[65] I. Carbone, Ann. Mat. Pura Appl. 191 (2012) 819–844.
[66] J.C. Lootgieter, Ann. Inst. Henri Poincare A 13 (4) (1977) 385–410.
[67] R. Pyke, W.R. van Zwet, Ann. Probab. 32 (1A) (2004) 380–423.
[68] E. Slud, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 41 (4) (1977/78) 341–352.
[69] W.R. van Zwet, Ann. Probab. 6 (1) (1978) 133–137.
[70] J. Bertoin, Cambridge Studies in Advanced Mathematics, Vol. 102, Cambridge University Press, Cambridge, 2006.
[71] A.N. Kolmogoroff, C. R. (Doklady) Acad. Sci. URSS (N. S.) 31 (1941) 99–101.

[72] I. Carbone, A. Volčič, Rend. Istit. Mat. Univ. Trieste 39 (2007) 119–126.
[73] J. Olli, Geom. Dedicata 158 (2012) 69–86.
[74] M. Infusino, A. Volčič, Unif. Distrib. Theory 4 (2) (2009) 47–58.
[75] J. Iagrate, S.J. Miller, F. Strauch, Phys. Rev. E 91 (6) (2015) 062138 (6 pages.
[76] S. Janson, R. Neininger, Probab. Theory Related Fields 142 (3–4) (2008) 399–442.
[77] E. Stein, R. Shakarchi, Fourier Analysis: An Introduction, Princeton University Press, Princeton, NJ, 2003.
[78] E. Stein, R. Shakarchi, Complex Analysis, Princeton University Press, Princeton, NJ, 2003.
[79] M. Hindry, J. Silverman, Graduate Texts in Mathematics, Vol. 201, Springer, New York, 2000.
[80] S.J. Miller, R. Takloo-Bighash, An Invitation To Modern Number Theory, Princeton University Press, Princeton, NJ, 2006.
[81] K. Roth, Mathematika 2 (1955) 1–20.
[82] G.R. Grimmett, D.R. Stirzaker, Probability and Random Processes, second ed., Clarendon Press, Oxford, 2009.
[83] M. Gromov, Appendix I To Asymptotic Theory of Finite Dimensional Normed Spaces By Vitali D Milman and Gideon Schechtman, in: Lecture Notes in Mathematics, vol. 1200, Springer-Verlag, New York, 1986.
[84] T. Becker, D. Burt, T.C. Corcoran, A. Greaves-Tunnell, J.R. Iafrate, J. Jing, S.J. Miller, J.D. Porfilio, R. Ronan, J. Samranvedhya, F. Strauch, B. Talbut, Benfords Law and Continuous Dependent Random Variables, arXiv version. http://arxiv.org/pdf/1309.5603.
[85] H. Chen, Int. J. Math. Educ. Sci. Technol. 41 (4) (2010) 538–547.
[86] L. Kuipers, H. Niederreiter, Uniform Distribution of Sequences, John Wiley & Sons, 1974.
[87] J. Brothers, F. Morgan, Michigan Math. J. 41 (3) (1994) 419–431.
[88] C. Manack, S.J. Miller, Res. Number Theory (2015) 1:22. http://dx.doi.org/10.1007/s40993-015-0024-4.
[89] D. Hanson, K. Seyffarth, J.H. Weston, Math. Mag. 56 (4) (1983) 224–229.
[90] L.M. Leemis, B.W. Schmeiser, D.L. Evans, Amer. Statist. 54 (4) (2000) 236–241.
[91] K.C. Chase, A.Z. Mekjian, Phys. Rev. C 49 (1994) 2164.
[92] S.J. Lee, A.Z. Mekjian, Phys. Rev. C 45 (1992) 1284.
[93] A.Z. Mekjian, Phys. Rev. Lett. 64 (1990) 2125.