BENFORD'S LAW AND CONTINUOUS DEPENDENT RANDOM VARIABLES

THEALEXA BECKER, ALEC GREAVES-TUNNELL, STEVEN J. MILLER, RYAN RONAN, AND FREDERICK W. STRAUCH

ABSTRACT. Many systems exhibit a digit bias. For example, the first digit base 10 of the Fibonacci numbers, or of 2^n, equals 1 not 10% or 11% of the time, as one would expect if all digits were equally likely, but about 30% of the time. This phenomenon, known as Benford's Law, has many applications, ranging from detecting tax fraud for the IRS to analyzing round-off errors in computer science.

The central question is determining which data sets follow Benford's law. Inspired by natural processes such as particle decay, our work examines models for the 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. The main difficulty is that the resulting random variables are dependent, which we handle by a careful analysis of the dependencies and tools from Fourier analysis to obtain quantified convergence rates.

CONTENTS

1. INTRODUCTION

Date: November 6, 2011.

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

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

The first, second and fourth named authors were supported by Williams College and NSF grant DMS0850577, the third named author was partially supported by NSF grant DMS0970067, and the fifth named author was partially supported by NSF grant PHY1005571 and a Research Corporation Cottrell College Science Award. We thank Don Lemons and Charles Wohl for helpful discussions.

1.1. **Background and Problem.** The distribution of leading digits of numbers in data sets has a fascinating history, with numerous applications arising in very diverse fields. Base 10, the probability of observing a first digit of d is frequently not 1/9, as one would expect if all digits are equally likely, but rather $\log_{10}\left(\frac{d+1}{d}\right)$, leading to almost 30% of the leading digits being a 1. Though this bias was first observed by Newcomb [New] in the 1880s when he noticed that certain pages in tables of logarithms were more worn than others, the subject gained popularity with Benford's [Ben] 1938 paper, where he studied sets ranging from mathematical functions to street addresses of 'famous' people. Today, Benford's law can be found in applied mathematics [BH1], auditing [DrNi, MN3, Nig1, Nig2, Nig3, NigMi], biology [CLTF], computer science [Knu], dynamical systems [Ber1, Ber2, BBH, BHKR, Rod], economics [Tö], geology [NM], number theory [ARS, KonMi, LS], physics [PTTV], signal processing [PHA], statistics [MN2, CLM] and voting fraud [Meb], to name just a few. See [BH2, Hu] for extensive bibliographies, and [BH3, BH4, BH5, Dia, JKKKM, JR, MN1, Pin, Rai] for general surveys and explanations of the law's prevalence.

One of the most important questions in the subject, as well as one of the hardest, is to determine which processes lead to Benford behavior. Many researchers [Adh, AS, Bh, JKKKM, Lév1, Lév2, MN1, Rob, Sa, Sc1, Sc2, Sc3, ST] observed that sums, products and in general arithmetic operations on random variables leads to more Benford behavior. Many of the proofs use techniques from measure theory and Fourier analysis, though in some special cases it is possible to obtain closed form expressions for the densities, which can be analyzed directly.

A crucial input in the above arguments is that the random variables are independent. In this paper, we explore situations where there are dependencies. Our motivating example is due to Lemons [Lem], who proposed studying 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 number, the resulting summands are clearly dependent. While it is frequently easy to show that individual pieces are Benford, the difficulty is in handling all the pieces simultaneously.

Lemons models this 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.

The first 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 $n = 2011$, but there are 84,840 if n is 2012. By considering a continuous system from the start, we avoid these Diophantine complications. We will return to the corresponding discrete model in a sequel paper [BGMRS].

Another 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.[1]

These are important issues, and their resolution and model choice impacts the behavior of the system. We mention a result from Miller-Nigrini [MN2], where they prove that while order statistics are close to Benford's law (base 10), they do not converge in general to Benford behavior. In particular, this means that if we choose N points randomly on a stick of length L and use those points to partition our rod, the distribution of the resulting piece sizes *will not* be Benford. Motivated by this result and Lemons' paper, we instead consider 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. Each cut is chosen from some random variable K on $(0, 1)$ representing the fraction of the stick's length at which we make the cut. Explicitly, if we start with one stick of length L, after one iteration we have sticks of length LK_1 and $L(1 - K_1)$, and after two iterations we have sticks of length LK_1K_2, $LK_1(1 - K_2)$, $L(1 - K_1)K_3$, and $L(1 - K_1)(1 - K_3)$. Iterating this process N times, we are left with 2^N sticks coming from $2^N - 1$ random variables $K_1, \dots, K_{2^N-1}$, with lengths ranging from $X_1 = LK_1K_2K_8 \cdots K_{2^{N-1}}$ to $X_{2^N} = L(1-K_1)(1-K_3)(1-K_7)\cdots(1-K_{2^N-1})$ (see Figure 1). Clearly the X_i's are not independent, as $\sum_{i=1}^{2^N} X_i = L$.

Fix an N and perform the division on $M \gg N$ sticks of length L, recording the observed lengths. If we let M tend to infinity while N is fixed, and then let N go to infinity, we can easily show the distribution of all the lengths converges to Benford's law. The interesting case is to show that we also get Benford's law (or, more precisely, with high probability we are very close to Benford's law) if we take $M = 1$; in other words, we have a large number of divisions of one stick.

1.2. **Notation.**

To quantify the first digits distribution of the set $\{X_i\}_{i=1}^{2^N}$, we introduce some notation. More generally, instead of studying just the first digits we could look at the significand (called the mantissa in some sources). Recall any positive number x can be written as $S_{10}(x)10^{k(x)}$, where $S_{10}(x) \in [1, 10)$ is the significand. Two numbers have the same leading digits if and only if their significands are equal.

A sequence $\{y_n\}$ is equidistributed modulo 1 if for any $[a, b] \subset [0, 1]$ we have $\lim_{N\to\infty} \frac{1}{N} \cdot \#\{n \le N : x_n \in [a, b]\} = b - a$. Benford's law (in a stronger form for the distribution of the significands) is equivalent to the base-10 logarithms of the

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

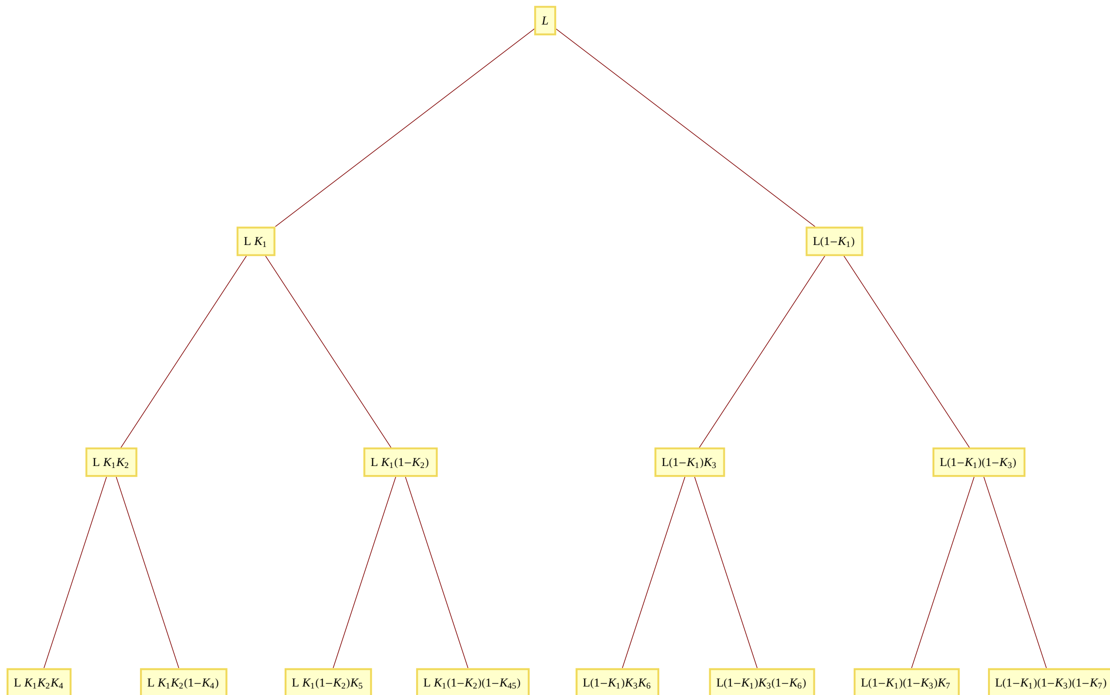

FIGURE 1. Breaking L into pieces, $N = 3$.

elements in the data sets being equidistributed modulo 1 (see [Dia, MT-B] for a proof). To have a first digit of d, we need the significand in $[\log_{10} d, \log_{10}(d+1))$; note the length of this interval is $\log_{10}(d+1) - \log_{10}(d) = \log_{10} \frac{d+1}{d}$, and we recover Benford's law for the distribution of the first digit.

For $s \in [1, 10)$, let

$$\varphi_s(u) \ := \ \begin{cases} 1 & \text{if the significand of } u \text{ is at most } s \\ 0 & \text{otherwise;} \end{cases} \tag{1.1}$$

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

While in the proofs we concentrate on the special case where each cut is chosen uniformly on its interval, we can consider the more general process of the cuts being identically distributed but non-uniform, or even the case where different cuts are taken from different distributions. We need to be a bit careful; while typically products of independent random variables converge to Benford behavior, there are pathological choices where this fails (see Example 2.4 of [MN1]). It is convenient to phrase the needed conditions in terms of the Mellin transform.

Let $f(x)$ be a continuous real-valued function on $[0, \infty)$ (though in our applications, as our random variables are cuts represented as a percentage of an interval, f will be a probability density defined on $[0,1]$). We define its Mellin transform, $(\mathcal{M}f)(s)$, by

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

Note $(\mathcal{M}f)(s) = \mathbb{E}[x^{s-1}]$, and thus results about expected values translate to results on Mellin transforms; for example, since f is a density we have $(\mathcal{M}f)(1) = 1$. If we let $x = e^{2\pi u}$ and $s = \sigma - i\xi$, then $(\mathcal{M}f)(\sigma - i\xi) = 2\pi \int_{-\infty}^{\infty} (f(e^{2\pi u})e^{2\pi\sigma u})\, e^{-2\pi i u\xi} du$, which is the Fourier transform of $g(u) = 2\pi f(e^{2\pi u})e^{2\pi\sigma u}$. The Mellin and Fourier transforms as thus related; in fact, it is this logarithmic change of variables which explains why both enter into Benford's law problems. For proofs of the Mellin transform properties one can therefore just mimic the proofs of the corresponding statements for the Fourier transform; good references are [SS1, SS2].

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

1.3. **Results.** Our main result is the following.

Theorem 1.1 (Limiting behavior of decompositions)**.** *Fix a continuous probability density f on $[0,1]$ such that*

$$\lim_{N\to\infty} \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \prod_{m=1}^{N} (\mathcal{M}h)\left(1 - \frac{2\pi i\ell}{\log 10}\right) = 0, \tag{1.3}$$

where $h(x)$ is either $f(x)$ or $f(1-x)$ (the density of $1-K$ if K has density f). Given a stick of length L, choose independent identically distributed random variables $K_1, K_2, \dots, K_{2^N-1}$ with density f and divide the stick as follows:

- *Divide L into LK_1 and $L(1-K_1)$.*
- *Divide LK_1 into LK_1K_2 and $LK_1(1-K_2)$, and $L(1-K_1)$ into $L(1-K_1)K_3$, $L(1-K_1)(1-K_3)$.*
- *Continue cutting each piece in two, obtaining after N iterations*

$$\begin{aligned} X_1 &= LK_1K_2K_4\cdots K_{2^{N-2}}K_{2^{N-1}} \\ X_2 &= LK_1K_2K_4\cdots L_{2^{N-2}}(1-K_{2^{N-1}}) \\ &\vdots \\ X_{2^N-1} &= L(1-K_1)(1-K_3)(1-K_7)\cdots(1-K_{2^{N-1}})K_{2^N-1} \\ X_{2^N} &= L(1-K_1)(1-K_3)(1-K_7)\cdots(1-K_{2^{N-1}})K_{2^N-1}. \end{aligned} \tag{1.4}$$

Define

$$P_N(s) := \frac{\sum_{i=1}^{2^N} \varphi_s(X_i)}{2^N}, \tag{1.5}$$

which is the fraction of partition pieces $X_1, \dots, X_{2^N}$ whose significand is less than or equal to s (see (1.1) *for the definition of ϕ_s). Then*

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

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

Thus as $N \to \infty$, the decomposition process is Benford.

Note that Part 1 of this theorem states that in the limit of having more and more copies of this process and the number of levels tending to infinity, the amalgamation converges to Benford's Law; Part 2 says we may consider just one process so long as the number of levels tends to infinity. The proof uses results from probability, analysis, and combinatorics, and is given in §2. A crucial input 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 (and its generalization, given in Remark 2.3) of [JKKKM]; for the convenience of the reader we quickly review this result and its proof in Appendix A. Unlike Part 1, the dependencies of the pieces is a major obstruction; we surmount this by breaking the pairs into groups depending on how dependent they are (specifically, how many cuts they share).

Remark 1.2. *We briefly remark on the key condition in the theorem, namely that the density satisfies* (1.3)*. This is an extremely weak condition, and is met by most distributions. For example, if f is the uniform density $\chi_{[0,1]}$ on $[0,1]$, then*

$$(\mathcal{M}\chi_{[0,1]})\left(1-\frac{2\pi i\ell}{\log 10}\right) \;=\; \left(1-\frac{2\pi i\ell}{\log 10}\right)^{-1} \tag{1.6}$$

(this is also true replacing $\chi_{[0,1]}(x)$ with $\chi_{[0,1]}(1-x)$, as these densities are the same on $[0,1]$), which implies

$$\lim_{N\to\infty}\left|\sum_{\substack{\ell=-\infty\\ \ell\neq 0}}^{\infty}\prod_{m=1}^{N}(\mathcal{M}f)\left(1-\frac{2\pi i\ell}{\log 10}\right)\right| \;=\; 2\lim_{N\to\infty}\sum_{\ell=1}^{\infty}\left|1-\frac{2\pi i\ell}{\log 10}\right|^{-N} \;=\; 0. \tag{1.7}$$

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. This can be done by using methods from Jang, Kang, Kruckman, Kudo and Miller [JKKKM]*, which we discuss in Appendix A.*

Remark 1.3. *In Theorem 1.1 we assumed for simplicity that at each stage each piece must split into exactly two pieces. Modifying the proof, one can show Benford behavior is also attained in the limit if at each stage each piece independently splits into $0, 1, 2, \dots$ or k pieces with probabilities $p_0, p_1, p_2, \dots, p_k$ (the case above is $p_2 = 1$).*

2. Proof of Main Result

We now turn to the proof of Theorem 1.1. We first prove that the first digit distribution of the given decomposition model is Benford on average and in the limit (Part 1 of the theorem). We then prove that this average first-digit distribution will, in fact, almost surely be the outcome of the model in the limit (Part 2 of the theorem).

2.1. **Proof of Part 1.**

Proof of Theorem 1.1(1). By the linearity of the expectation operator, we have

$$\mathbb{E}[P_N(s)] \;=\; \mathbb{E}\left[\frac{\sum_{i=1}^{2^N}\varphi_s(X_i)}{2^N}\right] \;=\; \frac{1}{2^N}\sum_{i=1}^{2^N}\mathbb{E}[\varphi_s(X_i)]. \tag{2.1}$$

We recall that all pieces X_i can be expressed as the product of the starting length L and N independent random variables taking on values in $(0,1)$. While there are dependencies among the X_i's, there are no dependencies among the K_i's. For a given i, there will be some number, say M_i, of cuts with factors drawn from K, and $N-M_i$ cuts with factors drawn from $1-K$ (where K is a random variable on $[0,1]$ with density f). By relabeling if necessary, we may assume

$$X_i \;=\; LK_1K_2\cdots K_{M_i}(1-K_{M_{i+1}})\cdots(1-K_N); \tag{2.2}$$

the first M_i random variables have density $f(x)$ and the last $N-M_i$ have $f(1-x)$.

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

$$\begin{aligned}\mathbb{E}[\varphi_s(X_i)] \quad &= \quad \int_{k_1=0}^{1}\int_{k_2=0}^{1}\cdots\int_{k_N=0}^{1}\varphi_s\left(L\prod_{r=1}^{M_i}k_r\prod_{\rho=M_i+1}^{N}(1-k_\rho)\right)\\ &\qquad \cdot\prod_{r=1}^{M_i}f(k_r)dk_r\prod_{\rho=M_i+1}^{N}f(1-k_\rho)dk_\rho.\end{aligned} \tag{2.3}$$

This is equivalent to studying the distribution of a product of N independent random variables and then rescaling the result by L. By the Pidgeon-hole Principle, we have at least $N/2$ random variables with factors k_r and density $f(k_r)$ or at least $N/2$ with factors $1-k_\rho$ and density $f(1-k_\rho)$. The convergence to Benford now follows from results of Jang, Kang, Kruckman, Kudo and Miller [JKKKM] (which are summarized for the reader's convenience in Appendix A). The key observation is to note that the Mellin transform at $1-\frac{2\pi i\ell}{\log 10}$ is strictly less than 1 in absolute value for continuous densities if $\ell \neq 0$ (which is seen by trivially inserting absolute values in the definition of the Mellin transform).

Explicitly, from Appendix A we find $\mathbb{E}[\phi_s(X_i)]$ equals $\log_{10} s$ plus a rapidly decaying N-dependent error term for all probability densities that satisfy the condition in the statement of the Theorem, which includes the uniform distribution. We may take the error to be independent of M_i (in other words, we can obtain a bound that holds for all decompositions simultaneously). We can do this by noting the Mellin transforms we have (with $\ell \neq 0$) are always less than 1 in absolute value. Thus the error is bounded by the maximum of the error from a product with $N/2$ terms with density $f(x)$ or a product with $N/2$ terms with density $f(1-x)$. Thus $\lim_{N\to\infty}\mathbb{E}[P_N(s)] = \log_{10} s$, completing the proof. □

Remark 2.1. *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 density of K_i and $1-K_i$ is the same. By Corollary A.2,*

$$\mathbb{E}[\phi_s(X_i)] - \log_{10} s \;\ll\; \frac{1}{2.9^N}, \tag{2.4}$$

and thus

$$\mathbb{E}[P_N(s)] - \log_{10} s \ \ll\ \frac{1}{2^N}\sum_{i=1}^{2^N}\frac{1}{2.9^N} \ =\ \frac{1}{2.9^N} \tag{2.5}$$

(we may use 10 for the big-Oh constant above).

2.2. **Proof of Part 2.**

Proof of Theorem 1.1(2). For ease of exposition, we assume all the cuts are drawn from the uniform distribution on $(0, 1)$. To facilitate doing 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, and the definition of variance and the linearity of the expectation operator, we have

$$\begin{aligned}
\mathrm{Var}\,(P_N(s)) &= \mathbb{E}[P_N(s)^2] - \mathbb{E}[P_N(s)]^2 \\
&= \mathbb{E}\left[\left(\frac{\sum_{i=1}^{2^N}\varphi_s(X_i)}{2^N}\right)^2\right] - \mathbb{E}[P_N(s)]^2 \\
&= \mathbb{E}\left[\frac{\sum_{i=1}^{2^N}\varphi_s(X_i)^2}{2^{2N}} + \sum_{\substack{i,j=1\\ i\neq j}}^{2^N}\frac{\varphi_s(X_i)\varphi_s(X_j)}{2^{2N}}\right] - \mathbb{E}[P_N(s)]^2 \\
&= \frac{1}{2^N}\mathbb{E}\left[P_N(s)\right] + \frac{1}{2^{2N}}\left(\sum_{\substack{i,j=1\\ i\neq j}}^{2^N}\mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)]\right) - \mathbb{E}[P_N(s)]^2.
\end{aligned} \tag{2.6}$$

From Theorem 1.1(1), $\mathbb{E}[P_N(s)] = \log_{10} s + o(1)$; here $o(1)$ means an error that tends to zero as $N \to \infty$. Thus

$$\mathrm{Var}\,(P_N(s)) \ =\ \frac{1}{2^{2N}}\left(\sum_{\substack{i,j=1\\ i\neq j}}^{2^N}\mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)]\right) - \log_{10}^2 s + o(1). \tag{2.7}$$

The problem is now reduced to evaluating the cross terms over all $i \neq j$. This is the hardest part of the analysis, and it is not feasible to evaluate the resulting integrals directly. Instead, for each i we partition the pairs (X_i, X_j) based on how 'close' X_j is to X_i in our tree (see Figure 1). We do this as follows. Recall that each of the 2^N pieces is a product of the starting length L and N random variables between 0 and 1. Writing any (X_i, X_j) pair in this form, it is clear that they must share some number of these random variables, say M terms. After M stages, the pieces X_i and X_j split, such that one piece has a factor K_{M+1} in its product, while the other contains the factor $(1 - K_{M+1})$. The remaining $N - M - 1$ elements in each product are independent from one another. After re-labeling, we can thus express any X_i, X_j pair, without loss of generality[2], as:

$$\begin{aligned} X_i &= L \cdot K_1 \cdot K_2 \cdots K_M \cdot K_{M+1} \cdot K_{M+2} \cdots K_N \\ X_j &= L \cdot K_1 \cdot K_2 \cdots K_M \cdot (1 - K_{M+1}) \cdot \tilde{K}_{M+2} \cdots \tilde{K}_N. \end{aligned} \tag{2.8}$$

With these definitions in mind, and again denoting the probability density function from which these random variables are drawn as $f(k)$ and $f(1-k)$, for a fixed i, j pair we have

$$\begin{aligned} \mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)] &= \int_{k_1=0}^{1} \int_{k_2=0}^{1} \cdots \int_{k_N=0}^{1} \int_{\tilde{k}_{M+2}=0}^{1} \cdots \int_{\tilde{k}_N=0}^{1} \varphi_s\left(L \prod_{r=1}^{M+1} k_r \prod_{r=M+2}^{N} k_r\right) \\ &\quad \cdot \varphi_s\left(L \prod_{r=1}^{M} k_r \cdot (1 - k_{M+1}) \cdot \prod_{r=M+2}^{N} \tilde{k}_r\right) \\ &\quad \cdot \prod_{r=1}^{N} f(k_r) \prod_{r=M+2}^{N} f(1-\tilde{k}_r) dk_1 dk_2 \cdots dk_N d\tilde{k}_{M+2} \cdots d\tilde{k}_N. \end{aligned} \tag{2.9}$$

In the statement above we have integrated over the remaining $2^N - N - (N - M - 1) = 2^N - 2N + M + 1$ variables; since those variables do not appear in any of the cuts in X_i or X_j, their corresponding integrals are 1.

The difficulty in understanding (2.9) 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.

We can study the behavior of the integral in (2.9) as a function of the significand of the first $M+1$ random variables. More specifically, we define the functions

$$\begin{aligned} I_1(L_1) &:= \int_{k_{M+1}=0}^{1} \cdots \int_{k_N=0}^{1} \varphi_s\left(L_1 \prod_{r=M+1}^{N} k_r\right) \prod_{r=M+1}^{N} f(k_r) dk_{M+1} dk_{M+2} \cdots dk_N \\ I_2(L_2) &:= \int_{\tilde{k}_{M+1}=0}^{1} \cdots \int_{\tilde{k}_N=0}^{1} \varphi_s\left(L_2 \prod_{r=M+1}^{N} \tilde{k}_r\right) \prod_{r=M+1}^{N} f(1-\tilde{k}_r) d\tilde{k}_{M+1} d\tilde{k}_{M+2} \cdots d\tilde{k}_N, \end{aligned} \tag{2.10}$$

which are defined for all $L_1, L_2 \in [1, 10)$. We will show that, for any L_1, L_2, we have $|I(L_1)I(L_2) - (\log_{10} s)^2| = o(1)$. Once we have this, then all that remains is to integrate $I(L_1)I(L_2)$ over the remaining M variables ($k_1, \dots, k_M$), which will be $(\log_{10} s)^2 + o(1)$. The rest of the proof follows from counting, for a given i, how many j lead to a given M.

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 (1.3). For illustrative purposes, we assume that each cut K is drawn from a uniform distribution, meaning $f(x)$ and

[2]Looking at Figure 1, we see that the labeling given below cannot be right (for example, K_2 and K_3 are in different branches); however, it is convenient to relabel the indices of the random variables.

$f(1-x)$ are the probability density function 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 $I(L_2)$ involve integrals over $N-M+1$ variables; we set $n := N-M-1$. For the case of a uniform distribution, equation (3.7) of [JKKKM] (or see Corollary A.2) gives for $n \geq 4$ that [3]

$$|I_1(L_1) - \log_{10} s| \ < \ \left(\frac{1}{2.9^n} + \frac{\zeta(n)-1}{2.7^n}\right) 2\log_{10} s, \tag{2.11}$$

Note that for all choices of L_1, $I_1(L_1) \in [0,1]$, 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 $O\left(1/2^n\right)$, and thus the error term decays very rapidly.

Replacing L_1 with L_2 yields a similar bound for $I_2(L_2)$. As such, we can choose a constant C such that

$$\begin{aligned} |I_1(L_1) \ - \ \log_{10} s| \ &\leq \ \frac{C}{2.9^n} \\ |I_2(L_2) \ - \ \log_{10} s| \ &\leq \ \frac{C}{2.9^n} \end{aligned} \tag{2.12}$$

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

$$\left|I_1(L_1)\cdot I_2(L_2) - (\log_{10} s)^2\right| \ \leq \ \frac{2C}{2.9^n}. \tag{2.13}$$

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

$$\left|\sum_{\substack{i,j=1 \\ i\neq j}}^{2^N} \left(\mathbb{E}[\varphi_s(X_i)\varphi_s(X_j)] - \log_{10}^2 s\right)\right| \ \ \leq \ \ \sum_{i=1}^{2^N}\sum_{n=1}^{N} 2^{n-1}\frac{2C}{2.9^n} \ \leq \ 2^N\cdot 4C. \tag{2.14}$$

Substituting this into equation (2.7) yields

$$\mathrm{Var}\left(P_N(s)\right) \ \ \leq \ \ \frac{4C}{2^N} + o(1). \tag{2.15}$$

Since the variance must be non-negative by definition, it follows that $\lim_{N\to\infty} \mathrm{Var}\left(P_N(s)\right) = 0$, completing the proof if each cut is drawn from a uniform distribution. The more general case follows analogously, appealing to Theorem A.1 with non-identically distributed random variables (though there are only two densities). □

[3]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.

Appendix A. Convergence of Products to Benford

As this paper crucially builds on the fact that many products of independent random variables are Benford, we quickly sketch a proof of this result. The arguments below are a condensed version of [JKKKM]. The main change is that in [JKKKM] the results were stated for chains of random variables, whereas here we give the equivalent formulation for products.

A.1. **Preliminaries.** Before giving the proof, we review some notation and needed properties of the Mellin transform. The density of the random variable $\Xi_1 \cdot \Xi_2$ (the product of two independent random variables with density f and cumulative distribution function F) is

$$\int_0^\infty f\left(\frac{x}{t}\right) f(t) \frac{dt}{t} \tag{A.1}$$

(the generalization to more products is straightforward). To see this, we first calculate the probability that $\Xi_1 \cdot \Xi_2 \in [0, x]$ and then differentiate with respect to x. Thus

$$\begin{aligned} \text{Prob}(\Xi_1 \cdot \Xi_2 \in [0,x]) &= \int_{t=0}^\infty \text{Prob}\left(\Xi_2 \in \left[0, \frac{x}{t}\right]\right) f(t) dt \\ &= \int_{t=0}^\infty F\left(\frac{x}{t}\right) f(t) dt. \end{aligned} \tag{A.2}$$

Differentiating gives the density of $\Xi_1 \cdot \Xi_2$, which equals

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

the factor of $1/t$ actually facilitates the upcoming analysis.

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

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

(provided that the integral converges absolutely). If we set $g(s) = (\mathcal{M}f)(s)$ then $f(x) = (\mathcal{M}^{-1} g)(x)$. We define the convolution of two functions f_1 and f_2 by

$$(f_1 \star f_2)(x) = \int_0^\infty f_2\left(\frac{x}{t}\right) f_1(t) \frac{dt}{t} = \int_0^\infty f_1\left(\frac{x}{t}\right) f_2(t) \frac{dt}{t}. \tag{A.5}$$

The Mellin convolution theorem states that

$$(\mathcal{M}(f_1 \star f_2))(s) = (\mathcal{M}f_1)(s) \cdot (\mathcal{M}f_2)(s), \tag{A.6}$$

which by induction gives

$$(\mathcal{M}(f_1 \star \cdots \star f_N))(s) = (\mathcal{M}f_N)(s) \cdots (\mathcal{M}f_N)(s); \tag{A.7}$$

note $f_1 \star \cdots \star f_N$ is the density of the product of N random variables.

A.2. **Proof of Product Result.** We now sketch the proof that products of independent, identically distributed random variables converge to Benford, and isolate out the error term. The description below is modified from [JKKKM] with permission.

Theorem A.1 (Jang, Kang, Kruckman, Kudo and Miller [JKKKM])**.** *Let* $\Xi_1, \dots, \Xi_N$ *be independent random variables with densities* f_{Ξ_m}*. Assume*

$$\lim_{N\to\infty} \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \prod_{m=1}^{N} (\mathcal{M} f_{\Xi_m}) \left(1 - \frac{2\pi i \ell}{\log B}\right) \ = \ 0. \tag{A.8}$$

Then as $n \to \infty$*,* $X_N = \Xi_1 \cdots \Xi_N$ *converges to Benford's law. In particular, if* $Y_N = \log_B X_N$ *then*

$$\begin{aligned} &|\mathrm{Prob}(Y_N \bmod 1 \in [a,b]) - (b-a)| \\ &\leq \ (b-a) \cdot \left| \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \prod_{m=1}^{N} (\mathcal{M} f_{\Xi_m}) \left(1 - \frac{2\pi i \ell}{\log B}\right) \right|. \end{aligned} \tag{A.9}$$

Proof. To investigate the distribution of the digits of $X_N = \Xi_1 \cdots \Xi_N$ (base B) it's convenient to make a logarithmic change of variables, setting $Y_N = \log_B X_N$. We have

$$\mathrm{Prob}(Y_N \leq y) \ = \ \mathrm{Prob}(X_N \leq B^y) \ = \ F_N(B^y), \tag{A.10}$$

where f_N is the density of X_N and F_N is the cumulative distribution function. Taking the derivative gives the density of Y_N, which we denote by $g_N(y)$:

$$g_N(y) \ = \ f_N(B^y) B^y \log B. \tag{A.11}$$

A standard method to show X_N tends to Benford behavior is to show that $Y_N \bmod 1$ tends to the uniform distribution on $[0,1]$ (see for example [Dia, MT-B]). This can be seen from the following calculation. The key ingredient is Poisson Summation. Let $h_{N,y}(t) = g_N(y+t)$. Then

$$\sum_{\ell=-\infty}^{\infty} g_N(y+\ell) \ = \ \sum_{\ell=-\infty}^{\infty} h_{N,y}(\ell) \ = \ \sum_{\ell=-\infty}^{\infty} \widehat{h}_{N,y}(\ell) \ = \ \sum_{\ell=-\infty}^{\infty} e^{2\pi i y \ell} \widehat{g}_N(\ell), \tag{A.12}$$

where $\widehat{f}$ denotes the Fourier transform of f:

$$\widehat{f}(\xi) \ = \ \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx. \tag{A.13}$$

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

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

where $\text{Err}(z)$ means an error at most z in absolute value. Note that since g_N is a probability density, $\widehat{g}_N(0) = 1$. The proof is completed by showing that the sum over ℓ tends to zero as $n \to \infty$. We thus need to compute $\widehat{g}_N(\ell)$:

$$\begin{aligned}\widehat{g}_N(\xi) &= \int_{-\infty}^{\infty} g_N(y)e^{-2\pi i y \xi}dy \\ &= \int_{-\infty}^{\infty} f_N(B^y)B^y \log B \cdot e^{-2\pi i y \xi}dy \\ &= \int_0^{\infty} f_N(t)t^{-2\pi i \xi/\log B}dt \\ &= (\mathcal{M}f_N)\left(1 - \frac{2\pi i \xi}{\log B}\right) \\ &= \prod_{m=1}^{N} (\mathcal{M}f_{\Xi_m})\left(1 - \frac{2\pi i \xi}{\log B}\right). \quad \text{(A.15)}\end{aligned}$$

Substituting completes the proof. □

We isolate the result in the special case that all cuts are drawn from the uniform distribution on $(0, 1)$. The error term below depends on the value of the Riemann zeta function $\zeta(s)$,

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \quad (\text{Re}(s) > 1), \quad \text{(A.16)}$$

at positive integers. As $\zeta(N) - 1 \ll 1/2^N$, the error term below is essentially $1/2.9^N$ for N large.

Corollary A.2 (Products of Independent Uniform Random Variables)**.** *Let* $\Xi_1, \dots, \Xi_N$ *be* N *independent random variables that are uniformly distributed on* $(0, 1)$*, and let* $\text{Sig}_N(s)$ *be the probability that the significand of* $\Xi_1 \cdots \Xi_N$ *(base 10) is at most* s*. For* $N \geq 4$ *we have*

$$|\text{Sig}_N(s) - \log_{10} s| \leq \left(\frac{1}{2.9^N} + \frac{\zeta(N) - 1}{2.7^N}\right) 2 \log_{10} s. \quad \text{(A.17)}$$

REFERENCES

[Adh] A. K. Adhikari, *Some results on the distribution of the most significant digit*, Sankhyā: The Indian Journal of Statistics, Series B **31** (1969), 413–420.

[AS] A. K. Adhikari and B. P. Sarkar, *Distribution of most significant digit in certain functions whose arguments are random variables*, Sankhyā: The Indian Journal of Statistics, Series B **30** (1968), 47–58.

[ARS] T. Anderson, L. Rolen and R. Stoehr, *Benford's Law for Coefficients of Modular Forms and Partition Functions*, Proc. Am. Math. Soc. **139** (2011), no. 5, 1533–1541.

[BGMRS] T. Becker, A. Greaves-Tunnell, S. J. Miller, R. Ronan and F. Strauch, *Benford's Law and Continuous Dependent Random Variables*, in preparation.

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

[Ber1] A. Berger, *Multi-dimensional dynamical systems and Benford's Law*, Discrete Contin. Dyn. Syst. **13** (2005), 219–237.

[Ber2] A. Berger, *Benford's Law in power-like dynamical systems* **5** (2005), Stoch. Dyn., 587–607.

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

[BH1] A. Berger and T. P. Hill, *Newton's method obeys Benford's law*, American Mathematical Monthly **114** (2007), 588–601.

[BH2] A. Berger, and T. P. Hill, *Benford Online Bibliography*, `http://www.benfordonline.net`.

[BH3] A. Berger and T. P. Hill, *Fundamental flaws in Feller's classical derivation of Benford's law*, Preprint, University of Alberta, 2010.

[BH4] A. Berger and T. P. Hill, *Benford's Law strikes back: No simple explanation in sight for mathematical gem*, The Mathematical Intelligencer **33** (2011), 85–91.

[BH5] A. Berger and T. P. Hill, *A Basic Theory of Benford's Law*, Probab. Surv. **8** (2011), 1–126.

[BHKR] A. Berger, T. P. Hill, B. Kaynar and A. Ridder, *Finite-state Markov Chains Obey Benford's Law*, to appear in SIAM J. Matrix Analysis.

[Bh] R. N. Bhattacharya, *Speed of convergence of the n-fold convolution of a probability measure on a compact group*, Z. Wahrscheinlichkeitstheorie verw. Geb. **25** (1972), 1–10.

[CLTF] E. Costasa, V. López-Rodasa, F. J. Torob, A. Flores-Moya, *The number of cells in colonies of the cyanobacterium Microcystis aeruginosa satisfies Benford's law*, Aquatic Botany **89** (2008), no. 3, 341–343.

[CLM] V. Cuff, A. Lewis and S. J. Miller, *The Weibull distribution and Benford's law*, preprint.

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

[DrNi] P. D. Drake and M. J. Nigrini, *Computer assisted analytical procedures using Benford's Law*, Journal of Accounting Education **18** (2000), 127–146.

[Fe] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd edition, Vol. II, John Wiley & Sons, New York, 1971.

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

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

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

[Hu] W. Hurlimann, *Benford's law from 1881 to 2006*, `http://arxiv.org/pdf/math/0607168`.

[JKKKM] D. Jang, J. U. Kang, A. Kruckman, J. Kudo and S. J. Miller, *Chains of distributions, hierarchical Bayesian models and Benford's Law*, to appear in the Journal of Algebra, Number Theory: Advances and Applications.

[JR] E. Janvresse and T. de la Rue, *From uniform distribution to Benford's law*, Journal of Applied Probability **41** (2004) no. 4, 1203–1210.

[Knu] D. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, 3rd edition, Addison-Wesley, MA, 1997.

[KonMi] A. Kontorovich and S. J. Miller, *Benford's law, values of L-functions and the $3x+1$ problem*, Acta Arith. **120** (2005), 269–297.

[LS] J. Lagarias and K. Soundararajan, *Benford's Law for the $3x+1$ Function*, J. London Math. Soc. **74** (2006), series 2, no. 2, 289–303.

[Lem] D. S. Lemons, *On the number of things and the distribution of first digits*, Am. J. Phys. **54** (1986), no. 9, 816–817.

[Lév1] P. Lévy, *Théorie de l'Addition des Variables Aléatoires*, Gauthier-Villars, Paris, 1937.

[Lév2] P. Lévy, *L'addition des variables aléatoires définies sur une circonférence*, Bull. de la S. M. F. **67** (1939), 1–41.

[Meb] W. Mebane, *Detecting Attempted Election Theft: Vote Counts, Voting Machines and Benford's Law*, Prepared for delivery at the 2006 Annual Meeting of the Midwest Political Science Association, April 20-23, Palmer House, Chicago. `http://www.umich.edu/ wmebane/mw06.pdf`.

[MN1] S. J. Miller and M. Nigrini, *The Modulo 1 Central Limit Theorem and Benford's Law for Products*, International Journal of Algebra **2** (2008), no. 3, 119–130.

[MN2] S. J. Miller and M. Nigrini, *Order statistics and Benford's law*, International Journal of Mathematics and Mathematical Sciences, Volume 2008 (2008), Article ID 382948, 19 pages. doi:10.1155/2008/382948

[MN3] S. J. Miller and M. Nigrini, *Data diagnostics using second order tests of Benford's Law*, Auditing: A Journal of Practice and Theory **28** (2009), no. 2, 305–324. doi: 10.2308/aud.2009.28.2.305

[MT-B] S. J. Miller and R. Takloo-Bighash, *An Invitation to Modern Number Theory*, Princeton University Press, Princeton, NJ, 2006.

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

[Nig1] M. Nigrini, *The Detection of Income Tax Evasion Through an Analysis of Digital Frequencies*, PhD thesis, University of Cincinnati, OH, USA, 1992.

[Nig2] M. Nigrini, *Using Digital Frequencies to Detect Fraud*, The White Paper **8** (1994), no. 2, 3–6.

[Nig3] M. Nigrini, *Digital Analysis and the Reduction of Auditor Litigation Risk*. Pages 69–81 in *Proceedings of the 1996 Deloitte & Touche / University of Kansas Symposium on Auditing Problems*, ed. M. Ettredge, University of Kansas, Lawrence, KS, 1996.

[Nig4] M. Nigrini, *A taxpayer compliance application of Benford's law*, Journal of the American Taxation Association, 1996.

[Nig5] M. Nigrini, *The Use of Benford's Law as an Aid in Analytical Procedures*, Auditing: A Journal of Practice & Theory, **16** (1997), no. 2, 52–67.

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

[NigMi] M. Nigrini and L. Mittermaier, *The Use of Benford's law as an Aid in Analytical Procedures*, Auditing: A Journal of Practice & Theory **16** (1997), no. 2, 52–67.

[PHA] F. Perez-Gonzalez, G. L. Heileman, and C.T. Abdallah, *Benford's Law in Image Processing*, Proceedings of the IEEE International Conference on Image Processing (ICIP), San Antonio, TX, Volume 1 2007, pp. 405–408.

[PTTV] L. Pietronero, E. Tosatti, V. Tosatti and A. Vespignani, *Explaining the Uneven Distribution of Numbers in Nature: the Laws of Benford and Zipf*, Physica A: Statistical Mechanics and its Applications, 2001.

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

[Rai] R. A. Raimi, *The first digit problem*, Amer. Math. Monthly **83** (1976), no. 7, 521–538.

[Rob] H. Robbins, *On the equidistribution of sums of independent random variables*, Proc. Amer. Math. Soc. **4** (1953), 786–799.

[Rod] R. Rodriguez, *First significant digit patterns from mixtures of uniform digits*, The American Statistician, 2004.

[Sa] H. Sakamoto, *On the distributions of the product and the quotient of the independent and uniformly distributed random variables*, Tôhoku Math. J. **49** (1943), 243–260.

[Sc1] P. Schatte, *On sums modulo* 2π *of independent random variables*, Math. Nachr. **110** (1983), 243–261.

[Sc2] P. Schatte, *On the asymptotic uniform distribution of sums reduced mod 1*, Math. Nachr. **115** (1984), 275–281.

[Sc3] P. Schatte, *On the asymptotic logarithmic distribution of the floating-point mantissas of sums*, Math. Nachr. **127** (1986), 7–20.

[SS1] E. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press, Princeton, NJ, 2003.

[SS2] E. Stein and R. Shakarchi, *Complex Analysis*, Princeton University Press, Princeton, NJ, 2003.

[ST] M. D. Springer and W. E. Thompson, *The distribution of products of independent random variables*, SIAM J. Appl. Math. **14** (1966) 511–526.

[Tö] K-H Tödter, *Benford's Law as an Indicator of Fraud in Economics*, German Economic Review **10** (2009), 339–351.

E-mail address: tbecker@smith.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, SMITH COLLEGE, NORTHAMPTON, MA 01063

E-mail address: ahg1@williams.edu

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

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

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

E-mail address: ronan2@cooper.edu

DEPARTMENT OF ELECTRICAL ENGINEERING, COOPER UNION, NEW YORK, NY 10003

E-mail address: fws1@williams.edu

DEPARTMENT OF PHYSICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267