International Journal of Algebra, Vol. 2, 2008, no. 3, 119 - 130

The Modulo 1 Central Limit Theorem and Benford's Law for Products

Steven J. Miller[1]

Department of Mathematics
Brown University, Providence, RI 02912, USA
sjmiller@math.brown.edu

Mark J. Nigrini

Department of Business Administration and Accounting
Saint Michael's College, Colchester, VT 05439, USA
mnigrini@smcvt.edu

Abstract. Using elementary results from Fourier analysis, we provide an alternate proof of a necessary and sufficient condition for the sum of M independent continuous random variables modulo 1 to converge to the uniform distribution in $L^1([0,1])$, and discuss generalizations to discrete random variables. A consequence is that if $X_1, \dots, X_M$ are independent continuous random variables with densities $f_1, \dots, f_M$, for any base B as $M \to \infty$ for many choices of the densities the distribution of the digits of $X_1 \cdots X_M$ converges to Benford's law base B. The rate of convergence can be quantified in terms of the Fourier coefficients of the densities, and provides an explanation for the prevalence of Benford behavior in many diverse systems. To highlight the difference in behavior between identically and non-identically distributed random variables, we construct a sequence of densities $\{f_i\}$ with the following properties: (1) for each i, if every X_k is independently chosen with density f_i then the sum converges to the uniform distribution; (2) if the X_k's are independent but non-identical, with X_k having distribution f_k, then the sum does not converge to the uniform distribution.

Mathematics Subject Classification: 60F05, 60F25, 11K06 (primary), 42A10, 42A61, 62E15 (secondary)

Keywords: Central Limit Theorem, sums modulo 1, Fourier series, L^1-convergence, Benford's Law

[1]We thank Brian Cole, Ted Hill, Jill Pipher and Sergei Treil for discussions on an earlier draft. The first author was partially supported by NSF grant DMS-0600848.

1. Introduction

We investigate necessary and sufficient conditions for the distribution of a sum of random variables modulo 1 to converge to the uniform distribution. This topic has been fruitfully studied by many previous researchers. Our purpose here is to provide an elementary proof of prior results, and explicitly connect this problem to related problems in the Benford's Law literature concerning the distribution of the leading digits of products of random variables. As this question has motivated much of the research on this topic, we briefly describe that problem and its history, and then state our results.

For any base B we may uniquely write a positive $x \in \mathbb{R}$ as $x = M_B(x) \cdot B^k$, where $k \in \mathbb{Z}$ and $M_B(x)$ (called the mantissa) is in $[1, B)$. A sequence of positive numbers $\{a_n\}$ is said to be **Benford base** B (or to satisfy Benford's Law base B) if the probability of observing the base-B mantissa of a_n of at most s is $\log_B s$. More precisely,

$$\lim_{N\to\infty} \frac{\#\{n \le N : 1 \le M_B(a_n) \le s\}}{N} \ = \ \log_B s. \tag{1.1.1}$$

Benford behavior for continuous systems is defined analogously. Thus base 10 the probability of observing a first digit of j is $\log_{10}(j+1) - \log_{10}(j)$, implying that about 30% of the time the first digit is a 1.

Benford's Law was first observed by Newcomb in the 1880s, who noticed that pages of numbers starting with 1 in logarithm tables were significantly more worn than those starting with 9. In 1938 Benford [Ben] observed the same digit bias in 20 different lists with over 20,000 numbers in all. See [Hi1, Rai] for a description and history. Many diverse systems have been shown to satisfy Benford's law, ranging from recurrence relations [BrDu] to $n!$ and $\binom{n}{k}$ $(0 \le k \le n)$ [Dia] to iterates of power, exponential and rational maps [BBH, Hi2] to values of L-functions near the critical line and characteristic polynomials of random matrix ensembles [KoMi] to iterates of the $3x+1$ Map [KoMi, LS] to differences of order statistics [MN]. There are numerous applications of Benford's Law. It is observed in natural systems ranging from hydrology data [NM] to stock prices [Ley], and is used in computer science in analyzing round-off errors (see page 255 of [Knu] and [BH]), in determining the optimal way to store numbers[2] [Ha], and in accounting to detect tax fraud [Nig1, Nig2]. See [Hu] for a detailed bibliography of the field.

In this paper we consider the distribution of digits of products of independent random variables, $X_1 \cdots X_M$, and the related questions about probability densities of random variables modulo 1. Many authors [Sa, ST, AS, Adh, Ha, Tu] have observed that the product (and more generally, any nice arithmetic operation) of two random variables is often closer to satisfying Benford's law than

[2]If the data is distributed according to Benford's Law base 2, the probability of having to shift the result of multiplying two numbers if the mantissas are written as $0.x_1x_2x_3\cdots$ is about .38; if they are written as $x_1.x_2x_3\cdots$ the probability is about .62.

the input random variables; further, that as the number of terms increases, the resulting expression seems to approach Benford's Law.

Many of the previous works are concerned with determining exact formulas for the distribution of $X_1 \cdots X_M$; however, to understand the distribution of the digits all we need is to understand $\log_B |X_1 \cdots X_M| \bmod 1$. This leads to the equivalent problem of studying sums of random variables modulo 1. This formulation is now ideally suited for Fourier analysis. The main result is a variant of the Central Limit Theorem, which in this context states that for "nice" random variables, as $M \to \infty$ the sum[3] of M independent random variables modulo 1 tends to the uniform distribution; by simple exponentiation this is equivalent to Benford's Law for the product (see [Dia]). To emphasize the similarity to the standard Central Limit Theorem and the fact that our sums are modulo 1, we refer to such results as Modulo 1 Central Limit Theorems. Many authors [Bh, Bo, Ho, JR, Lev, Lo, Ro, Sc1, Sc2, Sc3] have analyzed this problem in various settings and generalizations, obtaining sufficient conditions on the random variables (often identically distributed) as well as estimates on the rate of convergence.

Our main result is a proof, using only elementary results from Fourier analysis, of a necessary and sufficient condition for a sum modulo 1 to converge to the uniform distribution in $L^1([0,1])$. We also give a specific example to emphasize the different behavior possible when the random variables are not identically distributed. We let $\widehat{g_m}(n)$ denote the n^{th} Fourier coefficient of a probability density g_m on $[0,1]$:

$$\widehat{g_m}(n) \ = \ \int_0^1 g_m(x) e^{-2\pi i n x} dx. \tag{1.1.2}$$

Theorem 1.1. *(The Modulo 1 Central Limit Theorem for Independent Continuous Random Variables) Let $\{Y_m\}$ be independent continuous random variables on $[0,1)$, not necessarily identically distributed, with densities $\{g_m\}$. A necessary and sufficient condition for the sum $Y_1 + \cdots + Y_M$ modulo 1 to converge to the uniform distribution as $M \to \infty$ in $L^1([0,1])$ is that for each $n \neq 0$ we have $\lim_{M\to\infty} \widehat{g_1}(n) \cdots \widehat{g_M}(n) = 0$.*

As Benford's Law is equivalent to the associated base B logarithm being equidistributed modulo 1 (see [Dia]), from Theorem 1.1 we immediately obtain the following result on the distribution of digits of a product.

Theorem 1.2. *Let $X_1, \dots, X_M$ be independent continuous random variables, and let $g_{B,m}$ be the density of $\log_B M_B(|X_m|)$. A necessary and sufficient condition for the distribution of the digits of $X_1 \cdots X_M$ to converge to Benford's*

[3]That is, we study sums of the form $Y_1 + \cdots + Y_M$. For the standard Central Limit Theorem one studies $\frac{\sum_m Y_m - \mathbb{E}[\sum_m Y_m]}{\mathrm{StDev}(\sum_m Y_m)}$. We subtract the mean and divide by the standard deviation to obtain a quantity which will be finite as $M \to \infty$; however, sums modulo 1 are a priori finite, and thus their unscaled value is of interest.

Law (base B) as $M \to \infty$ in $L^1([0,1])$ is for each $n \neq 0$ that $\lim_{M\to\infty} \widehat{g_{B,1}}(n) \cdots \widehat{g_{B,M}}(n) = 0$.

As other authors have noticed, the importance of results such as Theorem 1.2 is that they give an explanation of why so many data sets follow Benford's Law (or at least a close approximation to it). Specifically, if we can consider the observed values of a system to be the product of many independent processes with reasonable densities, then the distribution of the digits of the resulting product will be close to Benford's Law.

We briefly compare our approach with other proofs of results such as Theorem 1.1 (where the random variables are often taken as identically distributed). If the random variables are identically distributed with density g, our condition reduces to $|\widehat{g}(n)| < 1$ for $n \neq 0$. For a probability distribution, $|\widehat{g}(n)| = 1$ for $n \neq 0$ if and only if there exists $\alpha \in R$ such that all the mass is contained in the set $\{\alpha, \alpha + \frac{1}{n}, \dots, \alpha + \frac{n-1}{n}\}$. (As we are assuming our random variables are continuous and not discrete, the corresponding densities are in $L^1([0,1])$ and this condition is not met; in Theorem 1.3 we discuss generalizations to discrete random variables.) In other words, the sum of identically distributed random variables modulo 1 converges to the uniform distribution if and only if the support of the distribution is not contained in a coset of a finite subgroup of the circle group $[0,1)$. Interestingly, Levy [Lev] proved this just one year after Benford's paper [Ben], though his paper does not study digits. Levy's result has been generalized to other compact groups, with estimates on the rate of convergence [Bh]. Stromberg [Str] proved that[4] *the n-fold convolution of a regular probability measure on a compact Hausdorff group G converges to the normalized Haar measure in the weak-star topology if and only if the support of the distribution is not contained in a coset of a proper normal closed subgroup of G.*

Our arguments in the proof of Theorem 1.1 may be generalized to independent discrete random variables, at the cost of replacing L^1-convergence with weak convergence. Below $\delta_\alpha(x)$ denotes a unit point mass at α.

Theorem 1.3. *(Modulo 1 Central Limit Theorem for Certain Independent Discrete Random Variables) Let $\{Y_m\}$ be independent discrete random variables on $[0,1)$, not necessarily identically distributed, with densities*

$$g_m(x) \ = \ \sum_{k=1}^{r_m} w_{k,m} \delta_{\alpha_{k,m}}(x), \quad w_{k,m} > 0, \quad \sum_{k=1}^{r_m} w_{k,m} \ = \ 1. \tag{1.1.3}$$

Assume that there is a finite *set $A \subset [0,1)$ such that all $\alpha_{k,m} \in A$. A necessary and sufficient condition for the sum $Y_1 + \cdots + Y_M$ modulo 1 to converge weakly to the uniform distribution as $M \to \infty$ is that for each $n \neq 0$ we have* $\lim_{M\to\infty} \widehat{g_1}(n) \cdots \widehat{g_M}(n) = 0$.

[4] The following formulation is taken almost verbatim from the first paragraph of [Bh].

In §2 we prove Theorem 1.1 using only elementary facts from Fourier analysis, showing our condition is a consequence of Lebesgue's Theorem (on L^1-convergence of the Fejér series) and a standard approximation argument. We give an example of distinct densities $\{f_i\}$ with the following properties: (1) for each i, if every X_k is independently chosen with density f_i then the sum converges to the uniform distribution; (2) if the X_k's are independent but non-identical, with X_k having distribution f_k, then the sum does not converge to the uniform distribution. This example illustrates the difference in behavior when the random variables are not identically distributed: to obtain uniform behavior for the sum it does not suffice for each random variable to satisfy Levy or Stromberg's condition (the distribution is not concentrated on a coset of a finite subgroup of $[0,1)$). We conclude in §3 by sketching the proof of Theorem 1.3, and in Appendix A we comment on alternate techniques to prove results such as Theorem 1.2 (in particular, why our arguments are more general than applying the standard Central Limit Theorem to $\log_B|X_1| + \cdots + \log_B|X_M|$ to analyze the distribution of digits of $|X_1 \cdots X_N|$).

2. Analysis of Sums of Continuous Random Variables

We recall some standard facts from Fourier analysis (see for example [SS]). The convolution of two functions in $L^1([0,1])$ is

$$(f * g)(x) \;=\; \int_0^1 f(y)g(x-y)dy \;=\; \int_0^1 f(x-y)g(y)dy. \tag{2.2.1}$$

Convolution is commutative and associative, and the n^{th} Fourier coefficient of a convolution is the product of the two n^{th} Fourier coefficients.

Let g_1 and g_2 be two probability densities in $L^1([0,1])$. If Z_i is a random variable on $[0,1)$ with density g_i, then the density of $Z_1 + Z_2 \bmod 1$ is the convolution of g_1 with g_2.

Definition 2.1 (Fejér kernel, Fejér series)**.** *Let $f \in L^1([0,1])$. The N^{th} Fejér kernel is*

$$F_N(x) \;=\; \sum_{n=-N}^{N}\left(1 - \frac{|n|}{N}\right)e^{2\pi i n x}, \tag{2.2.2}$$

and the N^{th} Fejér series of f is

$$T_N f(x) \;=\; (f * F_N)(x) \;=\; \sum_{n=-N}^{N}\left(1 - \frac{|n|}{N}\right)\widehat{f}(n)e^{2\pi i n x}. \tag{2.2.3}$$

The Fejér kernels are an approximation to the identity (they are non-negative, integrate to 1, and for any $\delta \in (0,1/2)$ we have $\lim_{N\to\infty}\int_\delta^{1-\delta} F_N(x)dx = 0$).

Theorem 2.2 (Lebesgue's Theorem)**.** *Let $f \in L^1([0,1])$. As $N \to \infty$, $T_N f$ converges to f in $L^1([0,1])$.*

Lemma 2.3. *Let $f, g \in L^1([0,1])$. Then $T_N(f * g) = (T_N f) * g$.*

Proof. The proof follows immediately from the commutative and associative properties of convolution. □

We can now prove Theorem 1.1.

Proof of Theorem 1.1. We first show our condition is sufficient. The density of the sum modulo 1 is $h_M = g_1 * \cdots * g_M$. It suffices to show that, for any $\epsilon > 0$,

$$\lim_{M\to\infty} \int_0^1 |h_M(x) - 1| dx \ < \ \epsilon. \tag{2.2.4}$$

Using Lebesgue's Theorem (Theorem 2.2), choose N sufficiently large so that

$$\int_0^1 |h_1(x) - T_N h_1(x)| dx \ < \ \frac{\epsilon}{2}. \tag{2.2.5}$$

While N was chosen so that (2.2.5) holds with h_1, in fact this N works for *all* h_M (with the same ϵ). This follows by induction. The base case is immediate (this is just our choice of N). Assume now that (2.2.5) holds with h_1 replaced by h_M; we must show it holds with h_1 replaced by $h_{M+1} = h_M * g_{M+1}$. By Lemma 2.3 we have

$$T_N h_{M+1} \ = \ T_N(h_M * g_{M+1}) \ = \ (T_N h_M) * g_{M+1}. \tag{2.2.6}$$

This implies

$$\begin{aligned}
&\int_0^1 |h_{M+1}(x) - T_N h_{M+1}(x)| dx \\
&\quad = \int_0^1 |(h_M * g_{M+1})(x) - (T_N h_M) * g_{M+1}(x)| dx \\
&\quad = \int_0^1 \left| \int_0^1 (h_M(y) - T_N h_M(y)) \cdot g_{M+1}(x-y) \right| dy dx \\
&\quad \le \int_0^1 \int_0^1 |h_M(y) - T_N h_M(y)| \cdot g_{M+1}(x-y) dx dy \\
&\quad = \int_0^1 |h_M(y) - T_N h_M(y)| dy \cdot 1 \ < \ \frac{\epsilon}{2};
\end{aligned} \tag{2.2.7}$$

the interchange of integration above is justified by the absolute value being integrable in the product measure, and the x-integral is 1 as g_{M+1} is a probability density.

To show h_M converges to the uniform distribution in $L^1([0,1])$, we must show $\lim_{M\to\infty} \int_0^1 |h_M(x) - 1| dx = 0$. Let N and ϵ be as above. By the triangle inequality we have

$$\int_0^1 |h_M(x) - 1| dx \quad \le \quad \int_0^1 |h_M(x) - T_N h_M(x)| dx + \int_0^1 |T_N h_M(x) - 1| dx. \tag{2.2.8}$$

From our choices of N and ϵ, $\int_0^1 |h_M(x) - T_N h_M(x)|dx < \epsilon/2$; thus we need only show $\int_0^1 |T_N h_M(x) - 1|dx < \epsilon/2$ to complete the proof. As $\widehat{h_M}(0) = 1$,

$$\begin{aligned}\int_0^1 |T_N h_M(x) - 1|dx &= \int_0^1 \left| \sum_{\substack{n=-N \\ n\neq 0}}^{N} \left(1 - \frac{|n|}{N}\right) \widehat{h_M}(n) e^{2\pi i n x} \right| dx \\ &\leq \sum_{\substack{n=-N \\ n\neq 0}}^{N} \left(1 - \frac{|n|}{N}\right) |\widehat{h_M}(n)|. \end{aligned} \tag{2.2.9}$$

However, $\widehat{h_M}(n) = \widehat{g_1}(n) \cdots \widehat{g_M}(n)$, and by assumption tends to zero as $M \to \infty$ (as each $\widehat{g_m}(n)$ is at most 1 in absolute value, for each n the absolute value of the product is non-increasing in M). For fixed N and ϵ, we may choose M sufficiently large so that $|\widehat{h_M}(n)| < \epsilon/4N$ whenever $n \neq 0$ and $|n| \leq N$. Thus

$$\int_0^1 |T_N h_M(x) - 1|dx \ < \ 2N \cdot \frac{\epsilon}{4N} \ = \ \frac{\epsilon}{2}, \tag{2.2.10}$$

which implies

$$\int_0^1 |h_M(x) - 1|dx \ < \ \epsilon \tag{2.2.11}$$

for M sufficiently large. As ϵ is arbitrary, this completes the proof of the sufficiency; we now prove this condition is necessary.

Assume for some $n_0 \neq 0$ that $\lim_{M\to\infty} |\widehat{h}_M(n_0)| \neq 0$ (where as always $h_M = g_1 * \cdots g_M$). As the g_m are probability densities, $|\widehat{g}_m(n)| \leq 1$; thus the sequence $\{|\widehat{h}_M(n)|\}_{M=1}^{\infty}$ is non-increasing for each n, and hence by assumption converges to some number $c_n \in (0, 1]$.

Let $E_M(x) = h_M(x) - 1$; note $\widehat{E}_M(n) = \widehat{h}_M(n)$ for $n \neq 0$. To show h_M does not converge to the uniform distribution on $[0, 1]$, it suffices to show that E_M does not converge almost everywhere to the zero function on $[0, 1]$. Let n_0 be as above. We have

$$\left|\widehat{h}_M(n_0)\right| \ = \ \left|\widehat{E}_M(n_0)\right| \ = \ \left|\int_0^1 E_M(x) e^{2\pi i n_0 x} dx\right| \ \geq \ c_{n_0} \ > \ 0. \tag{2.2.12}$$

Therefore at least one of the following integrals is at least $c_{n_0}/2$:

$$\begin{aligned} &\int_{\substack{x\in[0,1] \\ \mathrm{Re}(E_M(x))\geq 0}} \mathrm{Re}\,(E_M(x))\, dx, && \int_{\substack{x\in[0,1] \\ \mathrm{Re}(E_M(x))\leq 0}} \mathrm{Re}\,(-E_M(x))\, dx \\ &\int_{\substack{x\in[0,1] \\ \mathrm{Im}(E_M(x))\geq 0}} \mathrm{Im}\,(E_M(x))\, dx, && \int_{\substack{x\in[0,1] \\ \mathrm{Im}(E_M(x))\leq 0}} \mathrm{Im}\,(-E_M(x))\, dx, \end{aligned} \tag{2.2.13}$$

and h_M cannot converge to the zero function in $L^1([0, 1])$; further, we obtain an estimate on the L^1-distance between the uniform distribution and h_M. $\square$

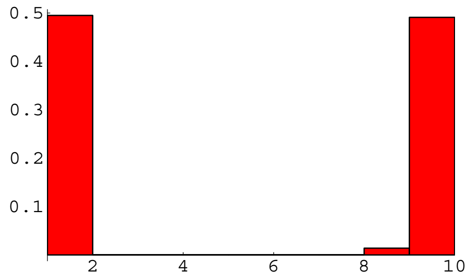

FIGURE 1. Distribution of digits (base 10) of 1000 products $X_1 \cdots X_{1000}$, where $g_{10,m} = \phi_{11^m}$.

The behavior is non-Benford if the conditions of Theorem 1.2 are violated. It is enough to show that we can find a sequence of densities $g_{B,m}$ such that $\lim_{M\to\infty} \prod_{m=1}^{M} \widehat{g_{B,m}}(1) \neq 0$. We are reduced to searching for an infinite product that is non-zero; we also need each term to be at most 1, as the Fourier coefficients of a probability density are dominated by 1. A standard example is $\prod_m c_m$, where $c_m = \frac{m^2+2m}{(m+1)^2}$; the limit of this product is 1/2. Thus as long as $\widehat{g_{B,m}}(1) \geq \frac{m^2+2m}{(m+1)^2}$, the conclusion of Theorem 1.2 will not hold for the products of the associated random variables; analogous reasoning yields a sum of independent random variables modulo 1 which does not converge to the uniform distribution.

Example 2.4 (Non-Benford Behavior of Products)**.** *Consider*

$$\phi_m \;=\; \begin{cases} m & \text{if } |x - \frac{1}{8}| \leq \frac{1}{2m} \\ 0 & \text{otherwise;} \end{cases} \tag{2.2.14}$$

ϕ_m is non-negative and integrates to 1. As $m \to \infty$ we have $|\widehat{\phi_m}(1)| \to 1$ because the density becomes concentrated at $1/8$ (direct calculation gives $\widehat{\phi_m}(1) = e^{2\pi i/8} + O(m^{-2})$). Let $X_1, \dots, X_M$ be independent random variables where the associated densities $g_{B,m}$ of $\log_B M(|X_m|)$ are ϕ_{11^m}. The behavior is non-Benford (see Figure 1). Note, however, that if each X_m had the common distribution ϕ_i for any fixed i, then in the limit the product will satisfy Benford's law.

Remark 2.5. *Generalizations of Theorem 1.1 hold for more general sums of random variables. Instead of $Y_1 + \cdots + Y_M$ we may study $\eta_1 Y_1 + \cdots + \eta_M Y_M$, where each η_m is a random variable taking values in $\{-1, 1\}$; the proof follows from the observation that if Y_m has density $g_m(y)$ then $-Y_m$ has density $g_m(1-y)$.*

3. Analysis of Sums of Discrete Random Variables

Many results from Fourier analysis do not apply if the random variables are discrete; Lebesgue's Theorem cannot be correct for a point mass as the density is concentrated on a set of measure zero. Let $\delta_\alpha(x)$ be a unit point mass[5] at α. Its Fourier coefficients are $\widehat{\delta_\alpha}(n) = e^{-2\pi i n\alpha}$, and simple algebra shows that its Fejér series is

$$F_N\delta_\alpha(x) \;=\; \frac{e^{-2\pi i(N-1)(x-\alpha)}(e^{2\pi i N(x-\alpha)}-1)^2}{(e^{2\pi i(x-\alpha)}-1)^2\; N}. \tag{3.3.1}$$

For $x \neq \alpha$, $\lim_{N\to\infty} F_N\delta_\alpha(x) = \delta_\alpha(x) = 0$; moreover, for x near α we have $|F_N\delta_\alpha(x)| \sim N$. Instead of convergence in $L^1([0,1])$ we have weak convergence: for any Schwartz function ϕ,

$$\lim_{N\to\infty}\int_0^1 F_N\delta_\alpha(x)\phi(x)dx \;=\; \int_0^1 \delta_\alpha(x)\phi(x)dx \;=\; \phi(\alpha). \tag{3.3.2}$$

Sketch of the proof of Theorem 1.3. We argue as in Theorem 1.1. Note Lemma 2.3 holds if f and g are sums of point masses. Instead of using Lebesgue's Theorem, we use weak convergence: given an $\epsilon > 0$ and a Schwartz function $\phi(x)$, by weak convergence there is an N such that

$$\left|\int_0^1 (h_1(x) - T_N h_1(x))\,\phi(x)dx\right| \;<\; \frac{\epsilon}{2}. \tag{3.3.3}$$

This is the generalization of (2.2.5). Further, we may assume (3.3.3) holds with $\phi(x)$ replaced with $\phi_{\alpha_{k,m}}(x) = \phi(x+\alpha_{k,m})$ for any $\alpha_{k,m} \in A$. *This is only true because A is finite*; while $N = N(\phi)$ depends on ϕ, as there are only finitely many test functions $\phi_{\alpha_{k,m}}$ we may take $N = \max N(\phi_{\alpha_{k,m}})$. A similar analysis as before shows (3.3.3) also holds with h_1 replaced by h_M. The key step in the induction is

$$\begin{aligned}
&\int_0^1 (h_M(y) - T_N h_M(y))\, g_{M+1}(x-y)\phi(x)dxdy \\
&= \int_0^1 (h_M(y) - T_N h_M(y)) \sum_{k=1}^{r_{M+1}} w_{k,M+1}\phi(y+\alpha_{k,M+1})dy \\
&= \sum_{k=1}^{r_{M+1}} w_{k,M+1}\int_0^1 (h_M(y) - T_N h_M(y))\,\phi_{\alpha_{k,M+1}}(y)dy,
\end{aligned} \tag{3.3.4}$$

which, as the $w_{k,M+1}$ sum to 1, is less than $\epsilon/2$ in absolute value. Arguing as in Theorem 1.1 completes the proof. □

Appendix A. Comparison with Alternate Techniques

We discuss an alternate proof of Theorem 1.2, applying the standard Central Limit Theorem to the sum $\log_B|X_1| + \cdots + \log_B|X_M|$ and noting that as the variance of a Gaussian increases to infinity, the Gaussian becomes uniformly distributed modulo 1. A significant drawback of a proof by the Central Limit

[5] Thus $\delta_\alpha(x)$ is a Dirac delta functional; if $\phi(x)$ is a Schwartz function then $\int_0^1 \delta_\alpha(x)\phi(x)dx$ is defined to be $\phi(\alpha)$.

Theorem is the requirement (at a minimum) that the variance of each $\log_B |X_m|$ be finite. This is a very weak condition, and in fact many random variables X with infinite variance (such as Pareto or modified Cauchy distributions) do have $\log_B |X|$ having finite variance; however, there are distributions where $\log_B |X|$ has infinite variance.

To a density f on $[0,\infty)$ we associate the density of the mantissa, f_B. Explicitly, the probability that X has first digit (base B) in $[1,s)$ is just

$$\int_1^s f_B(t)dt \ = \ \sum_{m=-\infty}^{\infty} \int_{1\cdot B^m \leq x \leq s\cdot B^m} f(x)dx. \tag{A.A.1}$$

Let X be the random variable with density

$$f_\alpha(x) \ = \ \begin{cases} \alpha/(x\log^{\alpha+1} x)^{-1} & \text{if } x \geq e \\ 0 & \text{otherwise.} \end{cases} \tag{A.A.2}$$

This is a probability distribution for $\alpha > 0$, and is a modification of a Pareto distribution; see [Mi] for some applications and properties of this distribution. We study the distribution of the digits base e; analogous results hold for other bases. The density of $Y = \log X$ is $g(y) = \alpha y^{-(\alpha+1)}$ for $y \geq 1$ and 0 otherwise. For $\alpha \in (0,2]$ the random variable Y has infinite variance, and thus we cannot prove the Benford behavior of products through the Central Limit Theorem; however, we can show the random variable X does satisfy the conditions of Theorem 1.2.

Let $F_{e,\alpha}$ be the cumulative distribution function of the digits (base e) associated to the density f_α of (A.A.2), and let $f_{e,\alpha}$ be the corresponding density of $F_{e,\alpha}$. We assume $\alpha > 1$ below to ensure convergence. By (A.A.1) we have

$$F_{e,\alpha}(s) \ = \ \int_1^s f_{e,\alpha}(t)dt \ = \ \sum_{m=0}^{\infty} \int_{1\cdot e^m \leq x \leq s\cdot e^m} f_\alpha(x)dx, \tag{A.A.3}$$

with $s \in [1,e)$. A simple integration gives

$$F_{e,\alpha}(s) \ = \ -\sum_{m=0}^{\infty} \frac{1}{\log^\alpha(s\cdot e^m)} \ + \ \sum_{m=0}^{\infty} \frac{1}{m^\alpha}; \tag{A.A.4}$$

note the second sum converges if $\alpha > 1$. The derivative of the first infinite sum in the expansion of $F_{e,\alpha}(s)$ is the sum of the derivatives of the individual summands, which follows from the rapid decay of the summands (see, for example, Corollary 7.3 of [La]). Differentiating the cumulative distribution function $F_{e,\alpha}$ gives the density

$$f_{e,\alpha}(s) \ = \ \alpha \sum_{m=0}^{\infty} \frac{1}{s\log^{\alpha+1}(s\cdot e^m)}, \quad s \in [1,e). \tag{A.A.5}$$

As $\alpha > 1$, for $m \neq 0$ the m^{th} summand is bounded by $m^{-(\alpha+1)}$. Thus the series for $f_{e,\alpha}(s)$ converges and is uniformly bounded for all s. A simple analysis shows that the conditions of Theorem 1.2 are satisfied for $\alpha \in (1,2]$.

The reason the Central Limit Theorem fails for densities such as that in (A.A.2) is that it tries to provide *too much* information. The Central Limit Theorem tries to give us the limiting distribution of $\log_B |X_1 \cdots X_M| = \log_B |X_1| + \cdots + \log_B |X_M|$; however, as we are only interested in the distribution of the digits of $X_1 \cdots X_M$, this is more information than we need.

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.

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

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

[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.

[BH] A. Berger and T. Hill, *Newton's method obeys Benford's law*, The Amer. Math. Monthly **114** (2007), no. 7, 588-601.

[Bo] J. Boyle, *An application of Fourier series to the most significant digit problem* Amer. Math. Monthly **101** (1994), 879–886.

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

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

[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 Scientist **86** (1996), 358–363.

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

[Ho] P. J. Holewijn, *On the uniform distribuiton of sequences of random variables*, Z. Wahrscheinlichkeitstheorie verw. Geb. **14** (1969), 89–92.

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

[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.

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

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

[LS] J. Lagarias and K. Soundararajan, *Benford's Law for the* $3x+1$ *Function*, preprint.

[La] S. Lang, *Undergraduate Analysis*, 2nd edition, Springer-Verlag, New York, 1997.

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

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

[Lo] R. M. Loynes, *Some results in the probabilistic theory of asympototic uniform distributions modulo* 1, Z. Wahrscheinlichkeitstheorie verw. Geb. **26** (1973), 33–41.

[Mi] S. J. Miller, *When the Cramér-Rao Inequality provides no information*, to appear in Communications in Information and Systems.

[MN] S. J. Miller and M. Nigrini, *Differences between Independent Variables and Almost Benford Behavior*, preprint. `http://arxiv.org/abs/math/0601344`

[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, *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.

[Nig2] 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.

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

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

[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.

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

[Str] K. Stromberg, *Probabilities on a compact group*, Trans. Amer. Math. Soc. **94** (1960), 295–309.

[Tu] P. R. Turner, *The distribution of leading significant digits*, IMA J. Numer. Anal. **2** (1982), no. 4, 407–412.

Received: September 6, 2007