THE WEIBULL DISTRIBUTION AND BENFORD'S LAW

VICTORIA CUFF, ALLISON LEWIS, AND STEVEN J. MILLER

Abstract. Benford's law states that many data sets have a bias towards lower leading digits, with a first digit of 1 about 30.1% of the time and a 9 only 4.6%. There are numerous applications, ranging from designing efficient computers to detecting tax, voter and image fraud. An important, open problem is to determine which common probability distributions are close to Benford's law. We show that the Weibull distribution, for many values of its parameters, is close to Benford's law, and derive an explicit formula measuring the deviations. As the Weibull distribution arises in many problems, especially survival analysis, our results provide additional arguments for the prevalence of Benford behavior.

1. Introduction

For any positive number x and base B, we can represent x in scientific notation as $x = S_B(x) \cdot B^{k(x)}$, where $S_B(x) \in [1, B)$ is called the significand[1] of x and the integer $k(x)$ represents the exponent. Benford's Law of Leading Digits proposes a distribution for the significands which holds for many data sets, and states that the proportion of values beginning with digit d is approximately

$$\text{Prob(first digit is } d \text{ base } B) \ = \ \log_B \left(\frac{d+1}{d}\right); \tag{1.1}$$

more generally, the proportion with significand at most s base B is

$$\text{Prob}(1 \le S_B \le s) \ = \ \log_B s. \tag{1.2}$$

In particular, base 10 the probability that the first digit is a 1 is about 30.1% (and not the 11% one would expect if each digit from 1 to 9 were equally likely).

This leading digit irregularity was first discovered by Simon Newcomb in 1881 [Ne], who noticed that the earlier pages in the logarithmic books were more worn than other pages. In 1938 Frank Benford [Ben] observed the same digit bias in a variety of data sets. Benford studied the distribution of the first digits of 20 sets of data with over 20,000 total observations, including river lengths, populations, and mathematical sequences. For a full history and description of the law, see [Hi1, Rai].

There are numerous applications of Benford's law. Two of the more famous are detecting tax and voter fraud [Me, Nig1, Nig2], but there are also applications in many other fields, ranging from round-off errors in computer science [Knu] to detecting image fraud and compression in engineering [AHMP-GQ].

It is an interesting question to determine which probability distributions lead to Benford behavior. Such knowledge gives us a deeper understanding of which natural data sets should follow Benford's law (other approaches to understanding the universality of the law range from scale invariance [Hi2] to products of random variables [MiNi1]). Leemis, Schmeiser, and Evans [LSE] ran numerical simulations on a variety of parametric survival distributions to examine

Key words and phrases. Benford's Law, Weibull Distribution, Poisson Summation.

The first and second named authors were supported by NSF Grant DMS0850577 and Williams College; the third named author was supported by NSF Grant DMS0970067.

[1]The significand is sometimes called the mantissa; however, such usage is discouraged by the IEEE and others as mantissa is used for the fractional part of the logarithm, a quantity which is also important in studying Benford's law.

conformity to Benford's Law. Among these distributions was the Weibull distribution, whose density is

$$f(x;\alpha,\beta,\gamma) \;=\; \begin{cases} \frac{\gamma}{\alpha}\left(\frac{x-\beta}{\alpha}\right)^{(\gamma-1)} \exp\left(-\left(\frac{x-\beta}{\alpha}\right)^{\gamma}\right) & \text{if } x \geq \beta \\ 0 & \text{otherwise,} \end{cases} \tag{1.3}$$

where $\alpha, \gamma > 0$. Note that α adjusts the scale of the data, β translates the input, and only γ affects the shape of the distribution. Special cases of the Weibull include the exponential distribution ($\gamma = 1$) and the Rayleigh distribution ($\gamma = 2$). The most common use of the Weibull is in survival analysis, where a random variable X (modeled by the Weibull) represents the "time-to-failure", resulting in a distribution where the failure rate is modeled relative to a power of time. The Weibull distribution arises in problems in such diverse fields as food contents, engineering, medical data, politics, pollution and sabermetrics [An, Ca, CB, Cr, Fr, MABF, Mik, Mi, TKD, We] to name just a few.

Our analysis generalizes the work of Miller and Nigrini [MiNi2], where the exponential case was studied in detail (see also [DL] for another approach to analyzing exponential random variables). The main ingredients come from Fourier analysis, in particular applying Poisson summation to the derivative of the cumulative distribution function of the logarithms modulo 1, F_B. One of the most common ways to prove a system is Benford is to show that its logarithms modulo 1 are equidistributed; we quickly sketch the proof of this equivalence (see also [Dia, MiNi2, MT-B] for details). If $y_n = \log_B x_n \bmod 1$ (thus y_n is the fractional part of the logarithm of x_n), then the significands of B^{y_n} and $x_n = B^{\log_B x_n}$ are equal, as these two numbers differ by a factor of B^k for some integer k. If now $\{y_n\}$ is equidistributed modulo 1, then by definition for any $[a,b] \subset [0,1]$ we have $\lim_{N\to\infty} \#\{n \leq N : y_n \in [a,b]\}/N = b - a$. Taking $[a,b] = [0, \log_B s]$ implies that as $N \to \infty$ the probability that $y_n \in [0, \log_B s]$ tends to $\log_B s$, which by exponentiating implies that the probability that the significand of x_n is in $[1,s]$ tends to $\log_B s$. Thus Benford's Law is equivalent to $F_B(z) = z$, implying that our random variable is Benford if $F_B'(z) = 1$. Therefore, a natural way to investigate deviations from Benford behavior is to compare the deviation of $F_B'(z)$ from 1, which would represent a uniform distribution.

We use the following notation for the various error terms:

Let $\mathcal{E}(x)$ denote an error of at most x in absolute value; thus $f(z) = g(z) + \mathcal{E}(x)$ means $|f(z) - g(z)| \leq x$.

Our main result is the following.

Theorem 1.1. *Let $Z_{\alpha,0,\gamma}$ be a random variable whose density is a Weibull with parameters $\beta = 0$ and $\alpha, \gamma > 0$ arbitrary. For $z \in [0,1]$, let $F_B(z)$ be the cumulative distribution function of $\log_B Z_{\alpha,0,\gamma} \bmod 1$; thus $F_B(z) := \mathrm{Prob}(\log_B Z_{\alpha,0,\gamma} \bmod 1 \in [0,z)$. Then*

(1) *The density of $Z_{\alpha,0,\gamma}$, $F_B'(z)$, is given by*

$$\begin{aligned} F_B'(z) \;&=\; 1 + 2\sum_{m=1}^{\infty} \mathrm{Re}\left[e^{-2\pi i m\left(z - \frac{\log\alpha}{\log B}\right)} \cdot \Gamma\left(1 + \frac{2\pi i m}{\gamma \log B}\right)\right] \\ &=\; 1 + 2\sum_{m=1}^{M-1} \mathrm{Re}\left[e^{-2\pi i m\left(z - \frac{\log\alpha}{\log B}\right)} \cdot \Gamma\left(1 + \frac{2\pi i m}{\gamma \log B}\right)\right] \\ &\quad\; + \mathcal{E}\left(\frac{1}{\pi^3} 2\sqrt{2} M (40 + \pi^2)\sqrt{\gamma \log B} \cdot e^{-\pi^2 M/\gamma \log B}\right). \end{aligned} \tag{1.4}$$

In particular, the densities of $\log_B Z_{\alpha,0,\gamma} \bmod 1$ and $\log_B Z_{\alpha B,0,\gamma} \bmod 1$ are equal, and thus it suffices to consider only α in an interval of the form $[a, aB)$ for any $a > 0$.

(2) *For $M \geq \frac{\gamma \log B \log 2}{4\pi^2}$, the error from keeping the first M terms is*

$$|\mathcal{E}| \;\leq\; \frac{1}{\pi^3} 2\sqrt{2} M (40 + \pi^2)\sqrt{\gamma \log B} \cdot e^{-\pi^2 M/\gamma \log B}. \tag{1.5}$$

(3) *In order to have an error of at most ϵ in evaluating $F'_B(z)$, it suffices to take the first M terms, where*

$$M \quad = \quad \frac{k + \ln k + \frac{1}{2}}{a}, \tag{1.6}$$

with $k \geq 6$ and

$$k \ = \ -\ln\left(\frac{a\epsilon}{C}\right), \quad a \ = \ \frac{\pi^2}{\gamma \log B}, \quad C \ = \ \frac{2\sqrt{2}(40+\pi^2)\sqrt{\gamma \log B}}{\pi^3}. \tag{1.7}$$

The above theorem is proved in the next section. Again, as in [MiNi2], the proof involves applying Poisson Summation to the derivative of the cumulative distribution function of the the logarithms modulo 1, which is a natural way to compare deviations from the resulting distribution and the uniform distribution. The key idea is that if a data set satisfies Benford's Law, then the distribution of its logarithms will be uniform. Our series expansions are obtained by applying properties of the Gamma function.

For further analysis, our series expansion for the derivative was compared to the uniform distribution through a Kolmogorov-Smirnov test; see Figure 1 for a contour plot of the discrepency. Note the good fit observed between the two distributions when $\gamma = 1$ (representing the Exponential distribution), which has already been proven to be a close fit to the Benford distribution [DL, LSE, MiNi2].

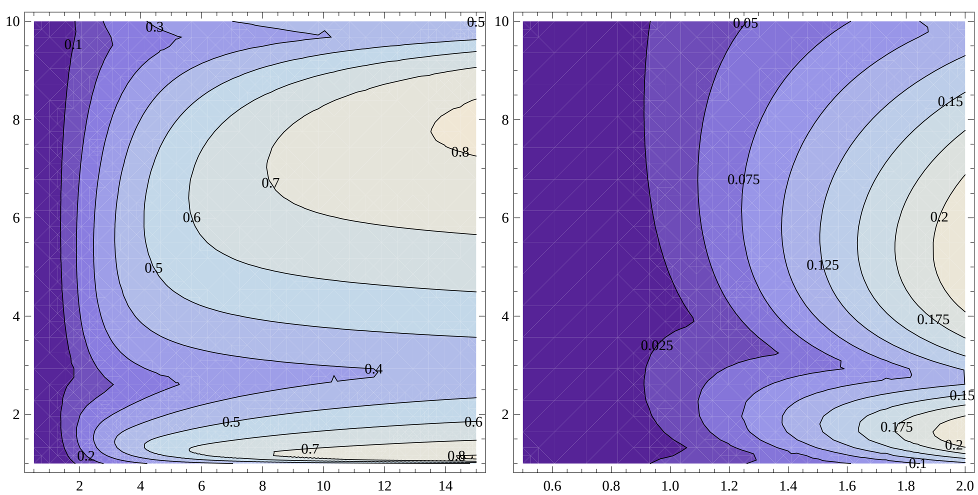

FIGURE 1. Kolmogorov–Smirnov Test: Left: $\gamma \in [0, 15]$, Right: $\gamma \in [0, 2]$. As γ (the shape parameter on the x-axis) increases, the Weibull distribution is no longer a good fit compared to the uniform. Note that α (the scale parameter on the y-axis) has less of an effect on the overall conformance.

The Kolmogorov–Smirnov metric gives a good comparison because it allows us to compare the distributions in terms of both parameters, γ and α. We also look at two other measures of closeness, the L_1-norm and the L_2-norm, both of which also test the differences between (1.4) and the uniform distribution; see Figures 2 and 3. The L_1-norm of $f - g$ is $\int_0^1 |f(t) - g(t)|dt$, which puts equal weights on the deviations, while the L_2-norm is given by $\int_0^1 |f(t) - g(t)|^2 dt$, which unlike the L_1-norm puts more weight on larger differences.

The combination of the Kolmogorov-Smirnoff tests and the L_1 and L_2 norms show us that the Weibull distribution almost exhibits Benford behavior when γ is modest; as γ increases the Weibull no longer conforms to the expected leading digit probabilities. The scale parameter α does have a small effect on the conformance as well, but not nearly to the same extreme as the shape parameter, γ. Fortunately in many applications the scale parameter γ is not too large

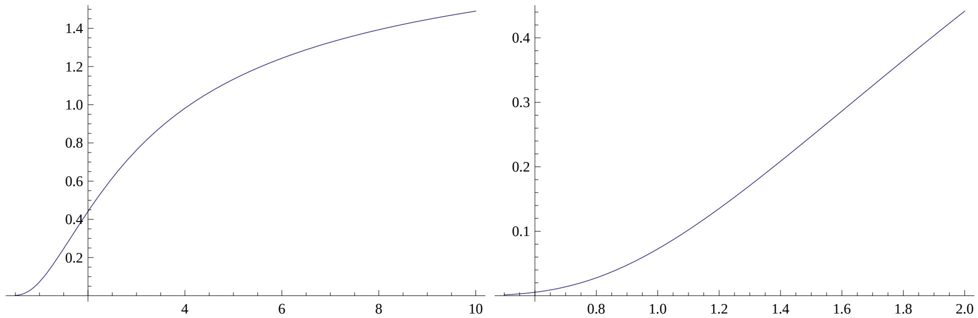

FIGURE 2. L_1-norm of $F_B'(z) - 1$: Left: $\gamma \in [0.5, 10]$, Right: $\gamma \in [0.5, 2]$. The closer γ is to zero the better the fit. As γ increases the cumulative Weibull distribution is no longer a good fit compared to 1. The L_1-norm is independent of α.

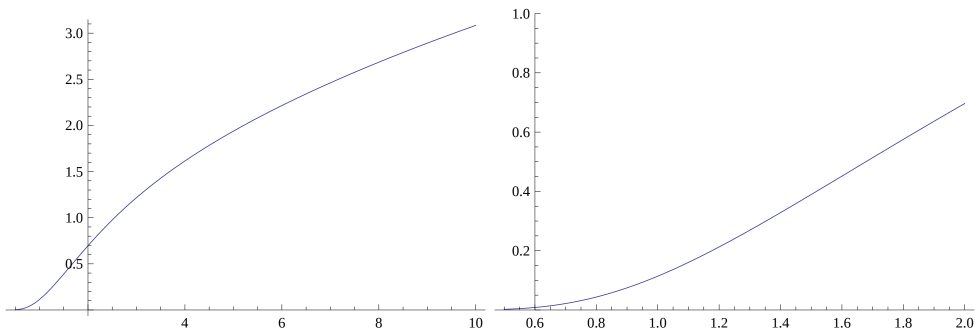

FIGURE 3. L_2-norm of $F_B'(z) - 1$: Left: $\gamma \in [0.5, 10]$, Right: $\gamma \in [0.5, 2]$. The closer γ is to zero the better the fit. As γ increases the cumulative Weibull distribution is no longer a good fit compared to 1. The L_2-norm is independent of α.

(it is frequently less than 2 in the Weibull references cited earlier), and thus our work provides additional support for the prevalence of Benford behavior.

2. PROOF OF THEOREM 1.1

To prove Theorem 1.1, it suffices to study the distribution of $\log_B Z_{\alpha,0,\gamma} \bmod 1$ when $Z_{\alpha,0,\gamma}$ has the standard Weibull distribution; see (1.3). The analysis is aided by the fact that the cumulative distribution function for a Weibull random variable has a nice closed form expression; for $Z_{\alpha,0,\gamma}$ the cumulative distribution function is $\mathcal{F}_{\alpha,0,\gamma}(x) = 1 - \exp(-(x/a)^c)$. Let $[a, b] \subset [0, 1]$. Then,

$$\begin{aligned}
\text{Prob}(\log_B Z_{\alpha,0,\gamma} \bmod 1 \in [a,b]) \quad &= \quad \sum_{k=-\infty}^{\infty} \text{Prob}(\log_B Z_{\alpha,0,\gamma} \bmod 1 \in [a+k, b+k]) \\
&= \quad \sum_{k=-\infty}^{\infty} \text{Prob}(Z_{\alpha,0,\gamma} \in [B^{a+k}, B^{b+k}]) \\
&= \quad \sum_{k=-\infty}^{\infty} \left(\exp\left(-\left(\frac{B^{a+k}}{\alpha} \right)^{\gamma} \right) - \exp\left(-\left(\frac{B^{b+k}}{\alpha} \right)^{\gamma} \right) \right).
\end{aligned} \tag{2.1}$$

2.1. **Proof of the Series Expansion.** We first prove the claimed series expansion in Theorem 1.1.

Proof of Theorem 1.1(1). It suffices to investigate (2.1) in the special case when $a = 0$ and $b = z$, since for any other interval $[a, b]$ we may determine its probability by subtracting the probability of $[0, a]$ from $[0, b]$. Thus, we study the cumulative distribution function of $\log_B Z_{\alpha,0,\gamma} \bmod 1$ for $z \in [0, 1]$:

$$\begin{aligned} F_B(z) &:= \mathrm{Prob}(\log_B Z_{\alpha,0,\gamma} \bmod 1 \in [0, z)) \\ &= \sum_{k=-\infty}^{\infty} \left(\exp\left(-\left(\frac{B^k}{\alpha}\right)^{\gamma} \right) - \exp\left(-\left(\frac{B^{b+k}}{\alpha}\right)^{\gamma} \right) \right). \end{aligned} \tag{2.2}$$

This series expansion is rapidly converging, and Benford behavior for $Z_{\alpha,0,\gamma}$ is equivalent to the rapidly converging series in (2.2) equalling z for all z.

As stated earlier, Benford behavior is equivalent to $F_B(z) = z$ for all $z \in [0, 1]$, thus a natural way to test for equivalence is to compare $F'(z)$ to 1. As in [MiNi2], studying the derivative $F'_B(z)$ is an easier way to approach this problem, because we obtain a simpler Fourier transform than the Fourier transform of $e^{-\left(\frac{B^{a+k}}{\alpha}\right)^{\gamma}} - e^{-\left(\frac{B^{b+k}}{\alpha}\right)^{\gamma}}$. We then can analyze the obtained Fourier transform by applying Poisson Summation.

Similarly as in [MiNi2] we use the fact that the derivative of the infinite sum $F_B(z)$ is the sum of the derivatives of the individual summands. This is justified by the rapid decay of summands, yielding

$$\begin{aligned} F'_B(z) &= \sum_{k=-\infty}^{\infty} \frac{1}{\alpha} \cdot \left[\exp\left(-\left(\frac{B^{b+k}}{\alpha}\right)^{\gamma} \right) B^{b+k} \left(\frac{B^{b+k}}{\alpha}\right)^{\gamma-1} \gamma \log B \right] \\ &= \sum_{k=-\infty}^{\infty} \frac{1}{\alpha} \cdot \left[\exp\left(-\left(\frac{\zeta B^{k}}{\alpha}\right)^{\gamma} \right) \zeta B^{k} \left(\frac{\zeta B^{k}}{\alpha}\right)^{\gamma-1} \gamma \log B \right], \end{aligned} \tag{2.3}$$

where for $z \in [0, 1]$, we use the change of variables $\zeta = B^z$.

We introduce $H(t) = \frac{1}{\alpha} \cdot \exp\left(-\left(\frac{\zeta B^t}{\alpha}\right)^{\gamma} \right) \zeta B^t \left(\frac{\zeta B^t}{\alpha}\right)^{\gamma-1} \gamma \log B$, where $\zeta \geq 1$ as $\zeta = B^z$ with $z \geq 0$. Since $H(t)$ is decaying rapidly we may apply Poisson Summation (see [MT-B, SS]), thus

$$\sum_{k=-\infty}^{\infty} H(k) \;=\; \sum_{k=-\infty}^{\infty} \widehat{H}(k), \tag{2.4}$$

where $\widehat{H}$ is the Fourier Transform of $H : \widehat{H}(u) = \int_{-\infty}^{\infty} H(t) e^{-2\pi i t u} dt$. Therefore

$$\begin{aligned} F'_B(z) &= \sum_{k=-\infty}^{\infty} H(k) \\ &= \sum_{k=-\infty}^{\infty} \widehat{H}(k) \\ &= \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\alpha} \cdot \exp\left(-\left(\frac{\zeta B^{k}}{\alpha}\right)^{\gamma} \right) \zeta B^{k} \left(\frac{\zeta B^{k}}{\alpha}\right)^{\gamma-1} \gamma \log B \cdot e^{-2\pi i t k} dt. \end{aligned} \tag{2.5}$$

We change variables again, setting

$$w \;=\; \left(\frac{\zeta B^t}{\alpha}\right)^{\gamma} \quad \text{or} \quad t \;=\; \log_B \left(\frac{\alpha w^{1/\gamma}}{\zeta}\right), \tag{2.6}$$

so that

$$dw \;=\; \frac{1}{\alpha} \left(\frac{\zeta B^t}{\alpha}\right)^{\gamma-1} \cdot \zeta B^t \gamma \log B \; dt. \tag{2.7}$$

Then

$$\begin{aligned}
F_B'(z) &= \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-w} \cdot \exp\left(-2\pi ik \cdot \log_B\left(\frac{\alpha w^{1/\gamma}}{\zeta}\right)\right) dw \\
&= \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-w} \cdot \exp\left(\log\left(\frac{\alpha w^{1/\gamma}}{\zeta}\right)\right)^{-2\pi ik/\log B} dw \\
&= \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-w} \cdot \left(\frac{\alpha w^{1/\gamma}}{\zeta}\right)^{-2\pi ik/\log(B)} dw \\
&= \sum_{k=-\infty}^{\infty} \left(\frac{\alpha}{\zeta}\right)^{-2\pi ik/\log B} \int_{-\infty}^{\infty} e^{-w} \cdot w^{-2\pi ik/\gamma\log B} dw \\
&= \sum_{k=-\infty}^{\infty} \left(\frac{\alpha}{\zeta}\right)^{-2\pi ik/\log B} \Gamma\left(1 - \frac{2\pi ik}{\gamma \log B}\right), \qquad (2.8)
\end{aligned}$$

where we have used the definition of the Γ-function:

$$\Gamma(s) \ = \ \int_0^{\infty} e^{-u} u^{s-1} du, \quad \mathrm{Re}(s) > 0. \qquad (2.9)$$

As $\Gamma(1) = 1$, we have

$$F_B'(z) \ = \ 1 + \sum_{m=1}^{\infty} \left[\left(\frac{\zeta}{\alpha}\right)^{\frac{2\pi im}{\log B}} \Gamma\left(1 - \frac{2\pi im}{\gamma \log B}\right) + \left(\frac{\zeta}{\alpha}\right)^{\frac{-2\pi im}{\log B}} \Gamma\left(1 + \frac{2\pi im}{\gamma \log B}\right)\right] \qquad (2.10)$$

As in the [MiNi2], the above series expansion is rapidly convergent. As $\zeta = B^z$ we have

$$\left(\frac{\zeta}{\alpha}\right)^{2\pi im/\log B} = \cos\left[2\pi mz - 2\pi mb\left(\frac{\log \alpha}{\log B}\right)\right] + i \sin\left[2\pi mz - 2\pi mb\left(\frac{\log \alpha}{\log B}\right)\right], \qquad (2.11)$$

which gives a Fourier series expansion for $F'(z)$ with coefficients arising from special values of the Γ-function.

Using properties of the Γ-function we are able to improve (2.10). If $y \in \mathbb{R}$ then from (2.9) we have $\Gamma(1 - iy) = \overline{\Gamma(1 + iy)}$ (where the bar denotes complex conjugation). Thus the m^{th} summand in (2.10) is the sum of a number and its complex conjugate, which is simply twice the real part. We use the following relationship:

$$|\Gamma(1 + ix)|^2 \ = \ \frac{\pi x}{\sinh(\pi x)} \ = \ \frac{2\pi x}{e^{\pi x} - e^{-\pi x}}. \qquad (2.12)$$

Writing the summands in (2.10) as $2\mathrm{Re}\left[e^{-2\pi im\left(z - \frac{\log \alpha}{\log B}\right)} \cdot \Gamma\left(1 + \frac{2\pi im}{\gamma \log B}\right)\right]$, (2.10) then becomes

$$\begin{aligned}
F_B'(z) &= 1 + 2\sum_{m=1}^{M-1} \mathrm{Re}\left[e^{-2\pi im\left(z - \frac{\log \alpha}{\log B}\right)} \cdot \Gamma\left(1 + \frac{2\pi im}{\gamma \log B}\right)\right] \\
&\quad + 2\sum_{m=M}^{\infty} \left[e^{-2\pi im\left(z - \frac{\log \alpha}{\log B}\right)} \cdot \Gamma\left(1 + \frac{2\pi im}{\gamma \log B}\right)\right]. \qquad (2.13)
\end{aligned}$$

In the exponential argument above there is no change in replacing α with αB, as this changes the argument by $2\pi i$. Thus it suffices to consider $\alpha \in [a, aB)$ for any $a > 0$. □

This proof demonstrates the power of using Poisson summation in Benford's law problems, as it allows us to convert a slowly convergent series expansion into a rapidly converging one.

2.2. **Bounding the Error.** We first estimate the contribution to $F_B'(z)$ from the tail, say from the terms with $m \geq M$. We do not attempt to derive the sharpest bounds possible, but rather highlight the method in a general enough case to provide useful estimates.

Proof of Theorem 1.1(2). We must bound

$$\mathrm{Re} \sum_{m=M}^{\infty} e^{-2\pi i m\left(z - \frac{\log \alpha}{\log B}\right)} \cdot \Gamma\left(1 + \frac{2\pi i m}{\gamma \log B}\right) \tag{2.14}$$

where $\Gamma(1+iu) = \int_0^\infty e^{-x} x^{iu} dx = \int_0^\infty e^{-x} e^{iu\log x} dx$. Note that in our case, $u = \frac{2\pi m}{\gamma \log B}$. As u increases there is more oscillation and therefore more cancelation, resulting in a smaller value for our integral. Since $|e^{i\theta}| = 1$, if we take absolute values inside the integrand we have $|e^{-2\pi i m\left(z - \frac{\log \alpha}{\log B}\right)}| = 1$, and thus we may ignore this term in computing an upper bound.

Using standard properties of the Gamma function, we have

$$|\Gamma(1+ix)|^2 \quad = \quad \frac{\pi x}{\sinh(\pi x)} \quad = \quad \frac{2\pi x}{e^{\pi x} - e^{-\pi x}}, \quad \text{where } x = \frac{2\pi m}{\gamma \log B}. \tag{2.15}$$

This yields

$$|\mathcal{E}| \quad \leq \quad \sum_{m=M}^{\infty} 1 \cdot \left(\frac{4\pi^2 m}{\gamma \log B} \cdot \frac{1}{e^{2\pi^2 m/\gamma \log B} - e^{-2\pi^2 m/\gamma \log B}}\right)^{1/2}. \tag{2.16}$$

Let $u = e^{2\pi^2 m/\gamma \log B}$. We overestimate our error term by removing the difference of the exponentials in the denominator. Simple algebra shows that for $\frac{1}{u - \frac{1}{u}} \leq \frac{2}{u}$ we need $u \geq \sqrt{2}$. For us this means $e^{2\pi^2 m/\gamma \log B} \geq \sqrt{2}$, allowing us to simplify the denominator if $m \geq \frac{\gamma \log B \log 2}{4\pi^2}$.

We substitute our new value for m into (2.15):

$$\begin{aligned} |\mathcal{E}| \quad &\leq \quad \sum_{m=M}^{\infty} \left(\frac{4\pi^2 m}{\gamma \log B}\right)^{1/2} \cdot \frac{\sqrt{2}}{e^{\pi^2 m/\gamma \log B}} \\ &\leq \quad \frac{2\sqrt{2}\pi}{\sqrt{\gamma \log B}} \sum_{m=M}^{\infty} \frac{\sqrt{m}}{e^{\pi^2 m/\gamma \log B}} \\ &\leq \quad \frac{2\sqrt{2}\pi}{\sqrt{\gamma \log B}} \int_M^\infty m e^{-\pi^2 m/\gamma \log B} dm. \end{aligned} \tag{2.17}$$

After integrating by parts, we have the following (with the only restriction being $m \geq \frac{\gamma \log B \log 2}{4\pi^2} \geq M$):

$$|\mathcal{E}| \;\leq\; \frac{1}{\pi^3} 2\sqrt{2} M (40 + \pi^2) \sqrt{\gamma \log B} \cdot e^{-\pi^2 M/\gamma \log B}. \tag{2.18}$$

Equation (2.18) provides the error for a given truncation, and is the error listed in (1.4). □

Proof of Theorem 1.1(3). Given the estimation of the error term from above, now ask the related question of, given an $\epsilon > 0$, how large must M be so that the first M terms give the $F_B'(z)$ accurately to within ϵ of the true value. Let $C = \frac{2\sqrt{2}(40+\pi^2)\sqrt{\gamma \log B}}{\pi^3}$ and $a = \frac{\pi^2}{\gamma \log B}$. We must choose M so that $CMe^{-aM} \leq \epsilon$, or equivalently

$$\frac{C}{a}\, aMe^{-aM} \;\leq\; \epsilon. \tag{2.19}$$

As this is a transcendental equation in M, we do not expect a nice closed form solution, but we can obtain a closed form expression for a bound on M; for any specific choices of C and a we can easily numerically approximate M. We let $u = aM$, giving

$$ue^{-u} \;\leq\; a\epsilon/C. \tag{2.20}$$

With a further change of variables, we let $k = -\ln(a\epsilon/C)$ and then expand u as $u = k + x$ (as the solution should be close to k). We find

$$\begin{aligned} u \cdot e^{-u} &\leq e^{-k} \\ (k+x)\, e^{-(k+x)} &\leq e^{-k} \\ \frac{k+x}{e^x} &\leq 1. \end{aligned} \tag{2.21}$$

We try $x = \ln k + \frac{1}{2}$:

$$\begin{aligned} \frac{k+x}{e^x} &\leq 1 \\ \frac{k+\ln k+\frac{1}{2}}{e^{\ln k+\frac{1}{2}}} &\leq 1 \\ \frac{k+\ln k+\frac{1}{2}}{k\cdot e^{1/2}} &\leq 1. \end{aligned} \tag{2.22}$$

From here, we want to determine the value of k such that $\ln k \leq \frac{1}{2}k$. Exponentiating, we need $k^2 \leq e^k$. As $e^k \geq k^3/3!$ for k positive, it suffices to choose k so that $k^2 \leq k^3/6$, or $k \geq 6$. For $k \geq 6$, we have

$$k + \ln k + \frac{1}{2} \ \leq\ k + \frac{1}{2}k + \frac{1}{12}k \ =\ \frac{19}{12}k \ \approx\ 1.5833, \tag{2.23}$$

but

$$e^{k/2} \ \approx\ 1.648222k. \tag{2.24}$$

Therefore we can conclude that the correct cutoff value for M, in order to have an error of at most ϵ, is

$$M \quad = \quad \frac{k+\ln k+\frac{1}{2}}{a}, \tag{2.25}$$

where $k \geq 6$ and

$$k \ =\ -\ln\left(\frac{a\epsilon}{C}\right), \quad a \ =\ \frac{\pi^2}{\gamma \log B}, \quad C \ =\ \frac{2\sqrt{2}(40+\pi^2)\sqrt{\gamma\log B}}{\pi^3}. \tag{2.26}$$

$\square$

References

[AHMP-GQ] C. T. Abdallah, Gregory L. Heileman, S. J. Miller, F. Pérez-González and T. Quach, *Application of Benford's Law to Images*, in Theory and Applications of Benford's Law (editors Steven J. Miller, Arno Berger and Ted Hill), Princeton University Press, to appear.

[An] Anonymous, *Weibull Survival Probability Distribution*, last modified May 11, 2009. `http://www.wepapers.com/Papers/30999/Weibull_Survival_Probability_Distribution`.

[Ben] F. Benford, *The Law of Anomalous Numbers*, Proceedings of the American Philosophical Society **78** (1938), 551-572.

[Ca] K. J. Carroll, *On the use and utility of the Weibull model in the analysis of survival data*, Controlled Clinical Trials **24** (2003), no. 6, 682–701.

[CB] O. Corzoa and N. Brachob, *Application of Weibull distribution model to describe the vacuum pulse osmotic dehydration of sardine sheets*, LWT - Food Science and Technology **41** (2008), no. 6, 1108–1115.

[Cr] T. S. Creasy, *A method of extracting Weibull survival model parameters from filament bundle load/strain data*, Composites Science and Technology **60** (2000), no. 6, 825–832.

[DL] L. Dümbgen and C. Leuenberger, *Explicit bounds for the approximation error in Benford's law*, Electronic Communications in Probability **13** (2008), 99–112.

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

[Fr] S. Fry, *How political rhetoric contributes to the stability of coercive rule: A Weibull model of post-abuse government survival.* Paper presented at the annual meeting of the International Studies Association, Le Centre Sheraton Hotel, Montreal, Quebec, Canada, Mar 17, 2004.

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

[Hi2] T. P. Hill, *A Statistical Derivation of the Significant-Digit Law*, Statistical Science **10** (1995), no. 4, 354-363.

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

[LSE] L. M. Leemis, B. W. Schmeiser and D. L. Evans, *Survival Distributions Satisfying Benford's Law*, The American Statistician **54** (2000), no. 3.

[MABF] B. McShane, M. Adrian, E. T. Bradlow and P. S. Fader, *Count Models Based on Weibull Interarrival Times*, Journal of Business and Economic Statistics **26** (2006), no. 3, 369–378.

[Me] W. Mebane, *Election Forensics: The Second-Digit Benford's Law Test and Recent American Presidential Elections*, Election Fraud Conference, Salt Lake city, Utah, September 29-30, 2006. `http://www.umich.edu/~wmebane/fraud06.pdf`.

[Mik] P. G. Mikolaj, *Environmental Applications of the Weibull Distribution Function: Oil Pollution*, Science 2 **176** (1972), no. 4038, 1019–1021.

[Mi] S. J. Miller, *A derivation of the Pythagorean Won-Loss Formula in baseball*, Chance Magazine **20** (2007), no. 1, 40–48.

[MiNi1] 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.

[MiNi2] S. J. Miller and M. J. Nigrini, *Order Statistics and Benford's Law*, International Journal of Mathematics and Mathematical Sciences, (2008), 1-13.

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

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

[Nig1] M. J. Nigrini, *Digital Analysis and the Reduction of Auditor Litigation Risk*. Pages 68-81 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. J. 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.

[Rai] R. A. Raimi, *The First Digit Problem*, The American Mathematical Monthly, **83:7** (1976), no. 7, 521-538.

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

[TKD] Y. Terawaki, T. Katsumi and V. Ducrocq, *Development of a survival model with piecewise Weibull baselines for the analysis of length of productive life of Holstein cows in Japan*, Journal of Dairy Science **89** (2006), no. 10, 4058–4065.

[We] W. Weibull, *A statistical distribution function of wide applicability*, J. Appl. Mech. **18** (1951), 293–297.

[Yi] C. T. Yiannoutsos, *Modeling AIDS survival after initiation of antiretroviral treatment by Weibull models with changepoints*, J Int AIDS Soc. **12** (2009), no. 9, doi:10.1186/1758-2652-12-9.

[ZLYM] Y. Zhao, A. H. Lee, K. K. W. Yau, and G. J. McLachlan, *Assessing the adequacy of Weibull survival models: a simulated envelope approach*, Journal of Applied Statistics (2011), to appear.

E-mail address: `vcuff@clemson.edu`

Department of Mathematics, Clemson University, Clemson, SC 29634

E-mail address: `lewisa11@up.edu`

Department of Mathematics, University of Portland, Portland, OR 97203

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

Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267