

involve

a journal of mathematics

The Weibull distribution and Benford's law

Victoria Cuff, Allison Lewis and Steven J. Miller



The Weibull distribution and Benford's law

Victoria Cuff, Allison Lewis and Steven J. Miller

(Communicated by John C. Wierman)

Benford's law states that many data sets have a bias towards lower leading digits (about 30% are 1s). It has numerous applications, from designing efficient computers to detecting tax, voter and image fraud. It's important to know which common probability distributions are almost Benford. We show that the Weibull distribution, for many values of its parameters, is close to Benford's law, quantifying the deviations. As the Weibull distribution arises in many problems, especially survival analysis, our results provide additional arguments for the prevalence of Benford behavior. The proof is by Poisson summation, a powerful technique to attack such problems.

1. Introduction to and applications of Benford's law

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 significand¹ 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

$$\text{Prob}(\text{first digit is } d \text{ base } B) = \log_B \left(\frac{d+1}{d} \right). \quad (1-1)$$

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

$$\text{Prob}(1 \leq S_B \leq s) = \log_B s. \quad (1-2)$$

MSC2010: primary 60F05, 11K06; secondary 60E10, 42A16, 62E15, 62P99.

Keywords: Benford's law, Weibull distribution, digit bias, Poisson summation.

Cuff and Lewis were supported by NSF Grant DMS0850577 and Williams College; Miller was supported by NSF Grant DMS0970067 and DMS1265673. This work was done at the 2010 SMALL REU at Williams College, and a summary of it will appear in Chapter Three of *The Theory and Applications of Benford's Law*, to be published by Princeton University Press and edited by Miller. Since this work was written many of these results have been independently derived and applied to Internet traffic; see the work of Arshadi and Jahangir [2014].

¹The 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.

In particular, in 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 Newcomb [1881], who noticed that the earlier pages in the logarithmic books were more worn than other pages. Fifty years later, Benford [1938] 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 [Hill 1998; Raimi 1976], or go to [Berger and Hill 2012] or [Miller 2015] for additional reading.

One of the most fascinating aspects of Benford's law is the large and diverse list of fields studying it (auditing, computer science, dynamical systems, engineering, number theory, and statistics, to list a few). There are numerous applications, especially in fraud and data integrity. Two of the more famous are detecting tax and voter fraud [Cho and Gaines 2007; Mebane 2006; Nigrini 1996; 1997], but there are also applications in many other fields, ranging from round-off errors in computer science [Knuth 1997] to detecting image fraud and compression in engineering [Abdallah et al. 2015]. Already Benford's law has led to a variety of tests, either to detect fraud (in everything from corporate returns to medical studies) or to test data integrity; see, for example, [Judge and Schechter 2009; Nigrini 1997; Miller and Nigrini 2009].

In the next section we discuss attempts to explain the prevalence of Benford's law; unfortunately, some of these approaches are flawed, and have been incorrectly used for decades. Our purpose in this article is to highlight techniques from Fourier analysis that may not be widely known to the diverse group of researchers and aficionados in the field, emphasizing how Poisson summation provides a clean and correct way to quantify deviations from Benford's law for a variety of phenomena. Our main result is to quantify how close Weibull distributions are to Benford (we state these in Theorem 4.1 in Section 4, after first reviewing the needed prerequisites in Section 3; the proof is given in Section 5). For certain values of the scale and shape parameter, these distributions are almost Benford; this is quite important, as many survival distributions are modeled by Weibull distributions, and thus Benford tests are applicable.

2. Explanations of Benford's law

There have been numerous attempts to pass from observing the prevalence of Benford's law to explaining its occurrence in different and diverse systems. Such knowledge gives us a deeper understanding of which natural data sets should follow Benford's law. One of the earliest and most popular is due to Feller [1966], and has been the subject of many articles and papers since (a very good, recent description of this approach is given in [Fewster 2009]). It suggests that Benford behavior arises when a probability distribution is spread out over several orders of

magnitude. Unfortunately, while some distributions satisfying this condition are close to Benford, others are not, and the method is sadly fundamentally flawed. See [Berger and Hill 2010; 2011b; Hill 2011] for detailed critiques of this method. The first rigorous explanation of Benford's law is due to Hill [1995] through scale invariance and measure theory (essentially, the distribution of leading digits should be invariant if we change scale); see also [Berger and Hill 2011a].

Rather than trying to prove why so many different phenomena are almost Benford, another approach is to study specific, important instances. In particular, there is an extensive literature on the leading digits of random variables and products of random variables of specific distributions (see for example [Miller and Nigrini 2008a]). While these arguments cannot be as general, the systems described arise in many important applications, making the importance of these researches clear.

The starting point of this work is the paper by Leemis, Schmeiser, and Evans [Leemis et al. 2000], who champion this viewpoint. They ran numerical simulations on a variety of parametric survival distributions to examine conformity to Benford's law. Among these distributions was the Weibull distribution, whose density is

$$f(x; \alpha, \gamma) = \begin{cases} (\gamma/\alpha)(x/\alpha)^{(\gamma-1)} \exp(-(x/\alpha)^\gamma) & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (2-1)$$

where $\alpha, \gamma > 0$. Note that α adjusts the scale of the data and only γ affects the shape of the distribution.² Special cases of the Weibull distribution include the exponential distribution ($\gamma = 1$) and the Rayleigh distribution ($\gamma = 2$). The most common use of the Weibull distribution is in survival analysis, where a random variable X modeled by the Weibull distribution 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, along with many others; see [Carroll 2003; Corzo and Bracho 2008; Fry 2004; McShane et al. 2008; Mikolaj 1972; Miller 2007; Terawaki et al. 2006; Weibull 1951; Yiannoutsos 2009; Zhao et al. 2011] to name just a few. As the extensiveness of this list indicates, many data sets follow a Weibull distribution, and thus if we are going test for fraud or data integrity, it is essential to quantify how close these distributions are to Benford. Our goal in this work is to provide proofs of the observations of Leemis, Schmeiser, and Evans [Leemis et al. 2000] that Weibull distributions are often close to Benford, emphasizing the ideas behind the method, as these are applicable to a variety of other problems (see, for example, [Jang et al. 2009; Kontorovich and Miller 2005; Miller and Nigrini 2008b]).

²One could introduce another parameter, β , which would represent a translation of the data. Doing so replaces x with $x - \beta$, and the condition $x \geq 0$ becomes $x \geq \beta$. In this paper we concentrate on the case $\beta = 0$.

3. Mathematical preliminaries

Our analysis generalizes the work of [Miller and Nigrini 2008b], where the exponential case was studied in detail (see also [Dümbgen and Leuenberger 2008] 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 . We first review some needed definitions, then describe why it is so useful to study the logarithms modulo 1, and conclude with a quick review of Poisson summation.

- (1) The gamma function $\Gamma(s)$ generalizes the factorial function; for n a nonnegative integer, we have $\Gamma(n+1) = n!$, and for $\Re(s) > 0$, we have

$$\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} dx$$

(we will need to evaluate the gamma function at complex arguments in our analysis); here $\Re(z)$ denotes the real part of z . See [Whittaker and Watson 1996] for an introduction and proofs of needed properties.

- (2) We say a is congruent to b modulo 1 if $a - b$ is an integer; we denote this by $a = b \pmod{1}$.
- (3) A sequence $\{a_n\}_{n=1}^\infty \subset [0, 1]$ is equidistributed if

$$\lim_{N \rightarrow \infty} \frac{\#\{n : n \leq N, a_n \in [a, b]\}}{N} = b - a$$

for all $[a, b] \subset [0, 1]$. Similarly a continuous random variable on $[0, \infty)$ whose probability density function is p is equidistributed modulo 1 if

$$\lim_{T \rightarrow \infty} \frac{\int_0^T \chi_{a,b}(x) p(x) dx}{\int_0^T p(x) dx} = b - a$$

for any $[a, b] \subset [0, 1]$, where $\chi_{a,b}(x) = 1$ for $x \pmod{1} \in [a, b]$ and 0 otherwise.

- (4) If f is an integrable function (so $\int_{-\infty}^\infty |f(x)| dx < \infty$) then its Fourier transform, denoted \hat{f} , is given by

$$\hat{f}(y) = \int_{-\infty}^\infty f(x) e^{-2\pi i xy} dx, \quad \text{where } e^{iu} = \cos u + i \sin u.$$

Note if X is a random variable with density f then this is a rescaled version of its characteristic function, $\mathbb{E}[e^{itX}]$.

- (5) Let $\eta > 0$. We say f decays like $x^{-(1+\eta)}$ if there are constants $x_0, C_\eta > 0$ such that $|f(x)| \leq C_\eta |x|^{-(1+\eta)}$ for all $|x| > x_0$.

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 [Diaconis 1977; Miller and Nigrini 2008b; Miller and Takloo-Bighash 2006] 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 $\lim_{N \rightarrow \infty} \#\{n \leq N : y_n \in [a, b]\} / N = b - a$. Taking $[a, b] = [0, \log_B s]$ implies that as $N \rightarrow \infty$, 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 $\log_B s$, the Benford probability.

Given a random variable X , let F_B denote the cumulative distribution function of $\log_B X \bmod 1$. The above discussion shows that Benford's law is equivalent to $F_B(z) = z$, or our original random variable X is Benford if $F'_B(z) = 1$. This suggests that 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.

Fourier analysis is ideally suited for these computations. The reason is that in general one cannot throw away part of a mathematical expression and still maintain equality. For example, note $\sqrt{(x \bmod 1) + (y \bmod 1)}$ is neither equal to nor congruent modulo 1 to $\sqrt{x + y}$; however, $e^{2\pi i x}$ does equal $e^{2\pi i(x \bmod 1)}$. By using the complex exponentials, it is harmless to drop modulo 1 restrictions. As these restrictions naturally arise in investigating the first digit, it is natural to attack the problem with Fourier techniques.

The last ingredient we need is Poisson summation. We don't state it in its most general form, as the following weak version typically suffices for Benford investigations due to the smoothness of the underlying densities. See [Miller and Takloo-Bighash 2006] or [Stein and Shakarchi 2003] for a proof.

Theorem 3.1 (Poisson summation). *Let f, f' and f'' be continuous functions which decay like $x^{-(1+\eta)}$ for some $\eta > 0$. Then*

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \hat{f}(n).$$

Our assumptions about f imply that \hat{f} decays rapidly. The power of Poisson summation is that it typically allows us to exchange a slowly converging sum with a rapidly converging sum. In many applications only the $n = 0$ term matters; if f is a probability density then it integrates to 1, and hence $\hat{f}(0) = 1$. For us, this is important as it implies a sum over nonzero n can measure a deviation.

For example, consider the density of a normal random variable Y with mean 0 and variance $N/2\pi$; this example is very important in showing Brownian motions

and many products of independent random variables become Benford (see [Miller and Takloo-Bighash 2006; Miller and Nigrini 2008a]). If we want to see how often $Y \bmod 1$ is in an interval $[a, b] \subset [0, 1]$, we need to study $\text{Prob}(Y \bmod 1 \in [a, b]) = \sum_{n=-\infty}^{\infty} \text{Prob}(Y \in [a + n, b + n])$. We *sketch* how Poisson summation enters, and provide full details when we prove our main result. The latter probabilities are integrals of the density over the intervals $[a + n, b + n]$, and if N is large each of these is approximately $b - a$ times the density at n . By Poisson summation, summing the density over n is the same as summing the Fourier transform at n :

$$\sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{N}} e^{-\pi n^2/N} = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 N}.$$

Note the sharp contrast between the two sums. For the first sum, all n with $|n| \leq \sqrt{N}$ contribute the same order of magnitude, while for the second sum, the $n = 0$ term contributes 1 and the next term is immensely smaller (by a factor of $e^{-\pi N}$). This example illustrates how Poisson summation allows us to replace a slowly decaying sum of a density with a rapidly decaying one.

4. Main results

Our main result is the following extension of results for the exponential distribution, which measures the deviation of the logarithm modulo 1 of the Weibull distribution and the uniform distribution. It’s thus not surprising that for γ close to 1, the digits are close to Benford, as $\gamma = 1$ corresponds to the exponential distribution. The main contribution below is quantifying how the fit worsens as γ grows. The larger γ is, the worse the fit. The effect of α is easier to explain. As the result of replacing α by αB is simply to rescale our random variable by a factor of B , the significand is unaffected. Thus it suffices to study α in the window $[1, B)$, but γ may be any real value.

Theorem 4.1. *Let $Z_{\alpha,\gamma}$ be a random variable whose density is Weibull with parameters $\alpha, \gamma > 0$ arbitrary. For $z \in [0, 1]$, let $F_B(z)$ be the cumulative distribution function of $\log_B Z_{\alpha,\gamma} \bmod 1$; thus $F_B(z) := \text{Prob}(\log_B Z_{\alpha,\gamma} \bmod 1 \in [0, z])$. Then the density of $\log_B Z_{\alpha,\gamma} \bmod 1$, $F'_B(z)$, is given by*

$$F'_B(z) = 1 + 2 \sum_{m=1}^{\infty} \Re \left(e^{-2\pi i m(z - \log \alpha / \log B)} \Gamma \left(1 + \frac{2\pi i m}{\gamma \log B} \right) \right). \tag{4-1}$$

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

From the fundamental equivalence, a straightforward integration immediately translates (4-1) into quantifying differences in the distribution of leading digits of Weibull random variables and Benford’s law. Specifically, the probability of a first

digit of d is obtained by integrating $F'_B(z)$ from $\log_B d$ to $\log_B(d + 1)$. The main term comes from the constant 1, and is $\log_B((d + 1)/d)$, the Benford probability; we discuss the size of the error in Theorem 4.2.

The above theorem is proved in the next section. As in [Miller and Nigrini 2008b], the proof involves applying Poisson summation to the derivative of the cumulative distribution function of the logarithms modulo 1, which as discussed in the previous section 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.

As the deviations of $F'_B(z)$ from being identically 1 measure the deviations from Benford behavior, it is important to have good estimates for the sum over m in (4-1). The bounds below have not been optimized, but instead have been chosen to simplify the algebra in the proofs (given in the Appendix). Thus we assume k below is at least 6, which is essentially equivalent to only investigating the case where the error ϵ is required to be of at most modest size (which is reasonable, as a series expansion with a large error is useless).

Theorem 4.2. *Let $F'_B(z)$ be as in (4-1).*

- (1) *For $M \geq (\gamma \log B \log 2)/4\pi^2$, the error from dropping the $m \geq M$ terms in $F'_B(z)$ is at most*

$$\frac{2\sqrt{2}(\pi^2 + \gamma \log B)\sqrt{\gamma \log B}}{\pi^3} M e^{-\pi^2 M/(\gamma \log B)}.$$

- (2) *In order to have an error of at most ϵ in evaluating $F'_B(z)$, it suffices to take the first M terms, where $M = (k + \ln k + 1/2)/a$, with $k = \max(6, -\ln(a\epsilon/C))$, $a = \pi^2/(\gamma \log B)$, and*

$$C = \frac{2\sqrt{2}(\pi^2 + \gamma \log B)\sqrt{\gamma \log B}}{\pi^3}.$$

For further analysis, we compared our series expansion for the derivative to the uniform distribution through a Kolmogorov–Smirnov test; see Figure 1 for a contour plot of the discrepancy. This statistic measures the absolute value of the greatest difference in cumulative distribution functions of two densities. Thus the larger the value, the further apart they are. 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 ([Dümbgen and Leuenberger 2008; Leemis et al. 2000; Miller and Nigrini 2008b]).

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

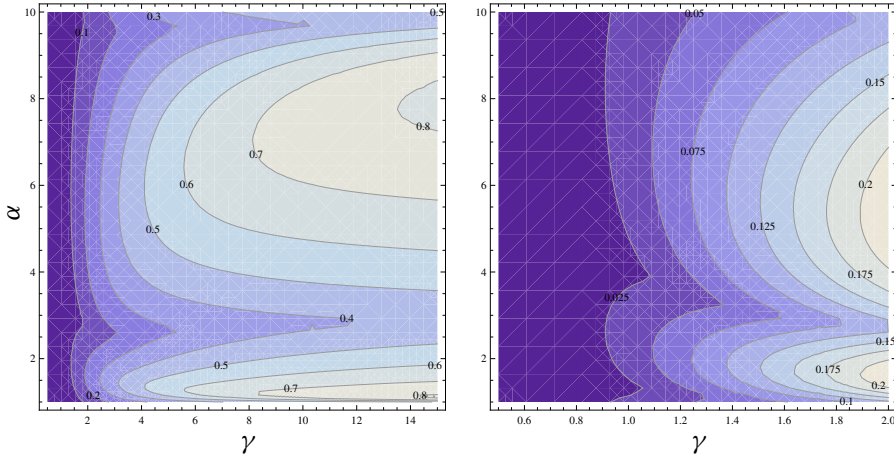


Figure 1. Kolmogorov–Smirnov test. Left: $\gamma \in [0, 15]$. Right: $\gamma \in [.5, 2]$. As γ (the shape parameter on the horizontal axis) increases, the Weibull distribution is no longer a good fit compared to the uniform. Note that α (the scale parameter on the vertical axis) has less of an effect on the overall conformance.

also test the differences between (4-1) and the uniform distribution; see Figure 2. The L_1 -norm of $f - g$ is $\int_0^1 |f(t) - g(t)| dt$, which puts equal weights on the all 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 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 - and L_2 -norms are independent of α .

The combination of the Kolmogorov–Smirnov 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 distribution no longer conforms to the expected leading digit probabilities. The scale parameter α does have a small

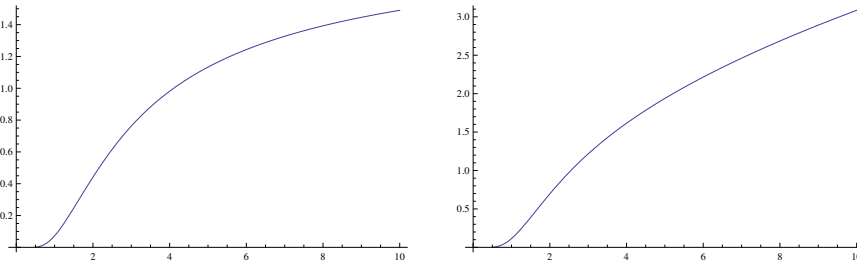


Figure 2. Left: L_1 -norm of $F'_B(z) - 1$ for $\gamma \in [0.5, 10]$. Right: L_2 -norm of $F'_B(z) - 1$ for $\gamma \in [0.5, 10]$.

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

5. Proof of main result

To prove Theorem 4.1, we study the distribution of $\log_B Z_{\alpha,\gamma} \bmod 1$ when $Z_{\alpha,\gamma}$ has the Weibull distribution with parameters α and γ . The analysis is aided by the fact that the cumulative distribution function for a Weibull random variable has a nice closed form expression; for $Z_{\alpha,\gamma}$, the cumulative distribution function is $F_{\alpha,\gamma}(x) = 1 - \exp(-(x/\alpha)^\gamma)$. Let $[a, b] \subset [0, 1]$. Then

$$\begin{aligned} \text{Prob}(\log_B Z_{\alpha,\gamma} \bmod 1 \in [a, b]) &= \sum_{k=-\infty}^{\infty} \text{Prob}(\log_B Z_{\alpha,\gamma} \bmod 1 \in [a+k, b+k]) \\ &= \sum_{k=-\infty}^{\infty} \text{Prob}(Z_{\alpha,\gamma} \in [B^{a+k}, B^{b+k}]) \\ &= \sum_{k=-\infty}^{\infty} \exp\left(-\left(\frac{B^{a+k}}{\alpha}\right)^\gamma\right) - \exp\left(-\left(\frac{B^{b+k}}{\alpha}\right)^\gamma\right). \end{aligned} \tag{5-1}$$

Proof of Theorem 4.1. It suffices to investigate (5-1) 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 $\log_B Z_{\alpha,\gamma} \bmod 1$ for $z \in [0, 1]$, which we denote by $F_B(z)$:

$$\begin{aligned} F_B(z) &:= \text{Prob}(\log_B Z_{\alpha,\gamma} \bmod 1 \in [0, z]) \\ &= \sum_{k=-\infty}^{\infty} \exp\left(-\left(\frac{B^k}{\alpha}\right)^\gamma\right) - \exp\left(-\left(\frac{B^{z+k}}{\alpha}\right)^\gamma\right). \end{aligned} \tag{5-2}$$

This series expansion is rapidly converging, and the closeness of $Z_{\alpha,\gamma}$ to the Benford distribution is equivalent to the rapidly converging series in (5-2) for $F_B(z)$ being close to z for all z .

A natural way to investigate the closeness of $F_B(z)$ to z is to compare $F'(z)$ to 1. As in [Miller and Nigrini 2008b], 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 $e^{-(B^k/\alpha)^\gamma} - e^{-(B^{z+k}/\alpha)^\gamma}$. We then can analyze the obtained Fourier transform by applying Poisson summation (Theorem 3.1).

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

$$\begin{aligned}
 F'_B(z) &= \sum_{k=-\infty}^{\infty} \frac{1}{\alpha} \exp\left(-\left(\frac{B^{z+k}}{\alpha}\right)^\gamma\right) B^{z+k} \left(\frac{B^{z+k}}{\alpha}\right)^{\gamma-1} \gamma \log B \\
 &= \sum_{k=-\infty}^{\infty} \frac{1}{\alpha} \exp\left(-\left(\frac{\zeta B^k}{\alpha}\right)^\gamma\right) \zeta B^k \left(\frac{\zeta B^k}{\alpha}\right)^{\gamma-1} \gamma \log B,
 \end{aligned} \tag{5-3}$$

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

We introduce

$$H(t) = \frac{1}{\alpha} \exp\left(-\left(\frac{\zeta B^t}{\alpha}\right)^\gamma\right) \zeta B^t \left(\frac{\zeta B^t}{\alpha}\right)^{\gamma-1} \gamma \log B,$$

where $\zeta \geq 1$ as $\zeta = B^z$ with $z \geq 0$. Since $H(t)$ is decaying rapidly we may apply Poisson summation; thus

$$\sum_{k=-\infty}^{\infty} H(k) = \sum_{k=-\infty}^{\infty} \hat{H}(k), \tag{5-4}$$

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

$$\begin{aligned}
 F'_B(z) &= \sum_{k=-\infty}^{\infty} H(k) = \sum_{k=-\infty}^{\infty} \hat{H}(k) \\
 &= \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\alpha} \exp\left(-\left(\frac{\zeta B^t}{\alpha}\right)^\gamma\right) \zeta B^t \left(\frac{\zeta B^t}{\alpha}\right)^{\gamma-1} \gamma e^{-2\pi i t k} \log B dt.
 \end{aligned} \tag{5-5}$$

We change variables again, setting $w = (\zeta B^t/\alpha)^\gamma$, which implies

$$t = \log_B\left(\frac{\alpha w^{1/\gamma}}{\zeta}\right) \quad \text{and} \quad dw = \frac{1}{\alpha} \left(\frac{\zeta B^t}{\alpha}\right)^{\gamma-1} \zeta B^t \gamma \log B dt, \tag{5-6}$$

so that

$$\begin{aligned}
 F'_B(z) &= \sum_{k=-\infty}^{\infty} \int_0^\infty e^{-w} \exp\left(-2\pi i k \log_B\left(\frac{\alpha w^{1/\gamma}}{\zeta}\right)\right) dw \\
 &= \sum_{k=-\infty}^{\infty} \left(\frac{\alpha}{\zeta}\right)^{-2\pi i k / \log B} \int_0^\infty e^{-w} w^{-2\pi i k / (\gamma \log B)} dw \\
 &= \sum_{k=-\infty}^{\infty} \left(\frac{\alpha}{\zeta}\right)^{-2\pi i k / \log B} \Gamma\left(1 - \frac{2\pi i k}{(\gamma \log B)}\right),
 \end{aligned} \tag{5-7}$$

where we used the definition of the gamma function in the last line. As $\Gamma(1) = 1$, we have

$$F'_B(z) = 1 + \sum_{m=1}^{\infty} \left(\frac{\zeta}{\alpha}\right)^{2\pi im/\log B} \Gamma\left(1 - \frac{2\pi im}{\gamma \log B}\right) + \left(\frac{\zeta}{\alpha}\right)^{-2\pi im/\log B} \Gamma\left(1 + \frac{2\pi im}{\gamma \log B}\right). \tag{5-8}$$

As in [Miller and Nigrini 2008b], the above series expansion is rapidly convergent. As $\zeta = B^z$, we have

$$\left(\frac{\zeta}{\alpha}\right)^{2\pi im/\log B} = \cos\left(2\pi mz - 2\pi m\left(\frac{\log \alpha}{\log B}\right)\right) + i \sin\left(2\pi mz - 2\pi m\left(\frac{\log \alpha}{\log B}\right)\right), \tag{5-9}$$

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

Using properties of the gamma function, we are able to improve (5-8). If $y \in \mathbb{R}$ then $\Gamma(1 - iy) = \overline{\Gamma(1 + iy)}$ (where the bar denotes complex conjugation). Thus the m -th summand in (5-8) is the sum of a number and its complex conjugate, which is simply twice the real part. We use the following standard relationship (see, for example, [Abramowitz and Stegun 1964]):

$$|\Gamma(1 + ix)|^2 = \frac{\pi x}{\sinh(\pi x)} = \frac{2\pi x}{e^{\pi x} - e^{-\pi x}}. \tag{5-10}$$

Writing the summands in (5-8) as

$$2\Re\left(e^{-2\pi im(z - \log \alpha / \log B)} \Gamma\left(1 + \frac{2\pi im}{\gamma \log B}\right)\right),$$

(5-8) becomes

$$F'_B(z) = 1 + 2 \sum_{m=1}^{\infty} \Re\left(e^{-2\pi im(z - \log \alpha / \log B)} \Gamma\left(1 + \frac{2\pi im}{\gamma \log B}\right)\right). \tag{5-11}$$

Finally, 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, with the main term corresponding to Benford behavior and the other terms measuring the deviation.

Appendix: Proofs of bounding estimates

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 4.2(1). We must bound the truncation error

$$\mathcal{E}_B(z) := \Re \sum_{m=M}^{\infty} e^{-2\pi im(z - \log \alpha / \log B)} \Gamma\left(1 + \frac{2\pi im}{\gamma \log B}\right), \quad (\text{A-1})$$

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, $u = 2\pi m / (\gamma \log B)$. As u increases there is more oscillation and therefore more cancellation, resulting in a smaller value for our integral. Since $|e^{i\theta}| = 1$, if we take absolute values inside the sum, we have $|e^{-2\pi im(z - \log \alpha / \log B)}| = 1$, and thus we may ignore this term in computing an upper bound.

Using standard properties of the gamma function, we have

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

This yields

$$|\mathcal{E}_B(z)| \leq \sum_{m=M}^{\infty} 1 \left(\frac{4\pi^2 m}{\gamma \log B} \frac{1}{e^{2\pi^2 m / (\gamma \log B)} - e^{-2\pi^2 m / (\gamma \log B)}} \right)^{1/2}. \quad (\text{A-3})$$

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 $m \geq (\gamma \log B \log 2) / 4\pi^2$, which we may do as we assumed M exceeds this value and $m \geq M$. We substitute this bound into (A-2), and replace \sqrt{m} with m to simplify the resulting integral:

$$|\mathcal{E}_B(z)| \leq \sum_{m=M}^{\infty} \left(\frac{4\pi^2 m}{\gamma \log B} \right)^{1/2} \frac{\sqrt{2}}{e^{\pi^2 m / (\gamma \log B)}} \leq \frac{2\sqrt{2}\pi}{\sqrt{\gamma \log B}} \int_M^{\infty} m e^{-\pi^2 m / (\gamma \log B)} dm. \quad (\text{A-4})$$

Letting $a = \pi^2 / (\gamma \log B)$, integrating by parts gives

$$|\mathcal{E}_B(z)| \leq \frac{2\sqrt{2}\pi}{\sqrt{\gamma \log B}} \frac{1}{a^2} (aM e^{-aM} + e^{-aM}) \leq \frac{2\sqrt{2}\pi}{\sqrt{\gamma \log B}} \frac{a+1}{a^2} M e^{-aM} \quad (\text{A-5})$$

(since $M \geq 1$, $aM + 1 \leq (a+1)M$), which after some algebra simplifies to

$$|\mathcal{E}_B(z)| \leq \frac{2\sqrt{2}(\pi^2 + \gamma \log B) \sqrt{\gamma \log B}}{\pi^3} M e^{-\pi^2 M / (\gamma \log B)}, \quad (\text{A-6})$$

which is the error listed in Theorem 4.2(1). \square

Proof of Theorem 4.2(2). Given the estimation of the error term from above, we now ask the related question: given an $\epsilon > 0$, how large must M be so that the first M terms give $F'_B(z)$ accurately to within ϵ of the true value? Let

$$C = \frac{2\sqrt{2}(\pi^2 + \gamma \log B) \sqrt{\gamma \log B}}{\pi^3}$$

and $a = \pi^2 / (\gamma \log B)$. We must choose M so that $CM e^{-aM} \leq \epsilon$, or equivalently

$$\frac{C}{a} a M e^{-aM} \leq \epsilon. \tag{A-7}$$

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

$$u e^{-u} \leq a\epsilon / C. \tag{A-8}$$

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

$$u e^{-u} \leq e^{-k} \quad \text{is equivalent to} \quad \frac{k + x}{e^x} \leq 1. \tag{A-9}$$

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

$$\frac{k + x}{e^x} \leq 1 \quad \text{is equivalent to} \quad \frac{k + \ln k + \frac{1}{2}}{k e^{1/2}} \leq 1. \tag{A-10}$$

From here, we want to determine the value of k such that $\ln k \leq \frac{1}{2}k$, as this ensures the needed inequality above holds. 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 $k \geq 6$; this holds for ϵ sufficiently small. For $k \geq 6$, we have

$$k + \ln k + \frac{1}{2} \leq k + \frac{1}{2}k + \frac{1}{12}k = \frac{19}{12}k \approx 1.5833k, \tag{A-11}$$

but

$$k e^{1/2} \approx 1.64872k. \tag{A-12}$$

Therefore a correct cutoff value for M , in order to have an error of at most ϵ , is

$$M = \frac{k + \ln k + \frac{1}{2}}{a}, \tag{A-13}$$

where

$$k = \max\left(k, -\ln \frac{a\epsilon}{C}\right), \quad a = \frac{\pi^2}{\gamma \log B}, \quad C = \frac{2\sqrt{2}(\pi^2 + \gamma \log B) \sqrt{\gamma \log B}}{\pi^3}. \tag{A-14}$$

□

References

- [Abdallah et al. 2015] C. T. Abdallah, G. 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*, edited by S. J. Miller, Princeton University Press, 2015.
- [Abramowitz and Stegun 1964] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, National Bureau of Standards Applied Mathematics Series **55**, U.S. Government Printing Office, Washington, DC, 1964. Reprinted by Dover, New York, 1974. MR 29 #4914 Zbl 0171.38503
- [Arshadi and Jahangir 2014] L. Arshadi and A. H. Jahangir, “Benford’s law behavior of internet traffic”, *J. Netw. Comput. Appl.* **40** (2014), 194–205.
- [Benford 1938] F. Benford, “The law of anomalous numbers”, *Proc. Amer. Phil. Soc.* **78**:4 (1938), 551–572. Zbl 0018.26502
- [Berger and Hill 2010] A. Berger and T. P. Hill, “Fundamental flaws in Feller’s classical derivation of Benford’s law”, preprint, 2010. arXiv 1005.2598
- [Berger and Hill 2011a] A. Berger and T. P. Hill, “A basic theory of Benford’s law”, *Probab. Surv.* **8** (2011), 1–126. MR 2012h:37015 Zbl 1245.60016
- [Berger and Hill 2011b] A. Berger and T. P. Hill, “Benford’s law strikes back: no simple explanation in sight for mathematical gem”, *Math. Intelligencer* **33**:1 (2011), 85–91. MR 2012j:62006 Zbl 1221.60010
- [Berger and Hill 2012] A. Berger and T. P. Hill, “Benford online bibliography”, 2012, available at <http://www.benfordonline.net>.
- [Carroll 2003] K. J. Carroll, “On the use and utility of the Weibull model in the analysis of survival data”, *Control. Clin. Trials* **24**:6 (2003), 682–701.
- [Cho and Gaines 2007] W. K. T. Cho and B. J. Gaines, “Breaking the (Benford) law: statistical fraud detection in campaign finance”, *Amer. Stat.* **61**:3 (2007), 218–223. MR 2393725
- [Corzo and Bracho 2008] O. Corzo and N. Bracho, “Application of Weibull distribution model to describe the vacuum pulse osmotic dehydration of sardine sheets”, *LWT Food Sci. Technol.* **41**:6 (2008), 1108–1115.
- [Diaconis 1977] P. Diaconis, “The distribution of leading digits and uniform distribution mod 1”, *Ann. Probability* **5**:1 (1977), 72–81. MR 54 #10178 Zbl 0364.10025
- [Dümbgen and Leuenberger 2008] L. Dümbgen and C. Leuenberger, “Explicit bounds for the approximation error in Benford’s law”, *Electron. Commun. Probab.* **13** (2008), 99–112. MR 2009b:60056 Zbl 1189.60044
- [Feller 1966] W. Feller, *An introduction to probability theory and its applications*, vol. 2, 2nd ed., Wiley, New York, 1966. MR 35 #1048 Zbl 0138.10207
- [Fewster 2009] R. M. Fewster, “A simple explanation of Benford’s law”, *Amer. Stat.* **63**:1 (2009), 26–32. MR 2655700
- [Fry 2004] S. Fry, “How political rhetoric contributes to the stability of coercive rule: a Weibull model of post-abuse government survival”, 2004. Paper presented at the annual meeting of the International Studies Association (Montreal, 2004).
- [Hill 1995] T. P. Hill, “A statistical derivation of the significant-digit law”, *Statist. Sci.* **10**:4 (1995), 354–363. MR 98a:60021 Zbl 0955.60509
- [Hill 1998] T. P. Hill, “The first digit phenomenon”, *Amer. Sci.* **86**:4 (1998), 358–363.
- [Hill 2011] T. P. Hill, “Benford’s law blunders”, *Amer. Stat.* **65**:2 (2011), 141.

- [Jang et al. 2009] D. Jang, J. U. Kang, A. Kruckman, J. Kudo, and S. J. Miller, "Chains of distributions, hierarchical Bayesian models and Benford's law", *J. Alg. Number Theory Adv. Appl.* **1**:1 (2009), 37–60. Zbl 1180.11026 arXiv 0805.4226
- [Judge and Schechter 2009] G. Judge and L. Schechter, "Detecting problems in survey data using Benford's law", *J. Hum. Resour.* **44**:1 (2009), 1–24.
- [Knuth 1997] D. E. Knuth, *The art of computer programming, 2: Seminumerical algorithms*, 3rd ed., Addison-Wesley, Reading, MA, 1997. MR 83i:68003 Zbl 0895.65001
- [Kontorovich and Miller 2005] A. V. Kontorovich and S. J. Miller, "Benford's law, values of L -functions and the $3x + 1$ problem", *Acta Arith.* **120**:3 (2005), 269–297. MR 2007c:11085 Zbl 1139.11033
- [Leemis et al. 2000] L. M. Leemis, B. W. Schmeiser, and D. L. Evans, "Survival distributions satisfying Benford's law", *Amer. Stat.* **54**:4 (2000), 236–241. MR 1803620
- [McShane et al. 2008] B. McShane, M. Adrian, E. T. Bradlow, and P. S. Fader, "Count models based on Weibull interarrival times", *J. Bus. Econom. Statist.* **26**:3 (2008), 369–378. MR 2009h:60095
- [Mebane 2006] W. R. Mebane, Jr., "Election forensics: the second-digit Benford's law test and recent American presidential elections", 2006, available at <http://www.umich.edu/wmebane/fraud06.pdf>. Paper presented at the Election Fraud Conference (Salt Lake City, UT, 2006).
- [Mikolaj 1972] P. G. Mikolaj, "Environmental applications of the Weibull distribution function: oil pollution", *Science* **176**:4038 (1972), 1019–1021.
- [Miller 2007] S. J. Miller, "A derivation of the Pythagorean won-loss formula in baseball", *Chance* **20**:1 (2007), 40–48. MR 2361359
- [Miller 2015] S. J. Miller (editor), *Benford's law: theory and applications*, Princeton University Press, 2015. Zbl 06446729
- [Miller and Nigrini 2008a] S. J. Miller and M. J. Nigrini, "The modulo 1 central limit theorem and Benford's law for products", *Int. J. Algebra* **2**:3 (2008), 119–130. MR 2009e:60053 Zbl 1148.60008
- [Miller and Nigrini 2008b] S. J. Miller and M. J. Nigrini, "Order statistics and Benford's law", *Int. J. Math. Math. Sci.* **2008** (2008), Article ID 382948. MR 2010c:62168 Zbl 05534756 arXiv math/0601344
- [Miller and Nigrini 2009] S. J. Miller and M. J. Nigrini, "Data diagnostics using second order tests of Benford's Law", *Audit. J. Pract. Theory* **28**:2 (2009), 305–324.
- [Miller and Takloo-Bighash 2006] S. J. Miller and R. Takloo-Bighash, *An invitation to modern number theory*, Princeton University Press, 2006. MR 2006k:11002 Zbl 1155.11001
- [Newcomb 1881] S. Newcomb, "Note on the frequency of use of the different digits in natural numbers", *Amer. J. Math.* **4**:1-4 (1881), 39–40. MR 1505286 JFM 13.0161.01
- [Nigrini 1996] M. J. Nigrini, "Digital analysis and the reduction of auditor litigation risk", pp. 68–81 in *Auditing symposium XIII: proceedings of the 1996 Deloitte & Touche/University of Kansas Symposium on Auditing Problems* (Lawrence, KS, 1996), edited by M. L. Ettredge, Division of Accounting and Information Systems, School of Business, University of Kansas, Lawrence, KS, 1996.
- [Nigrini 1997] M. J. Nigrini, "The use of Benford's law as an aid in analytical procedures", *Audit. J. Pract. Theory* **16**:2 (1997), 52–67.
- [Raimi 1976] R. A. Raimi, "The first digit problem", *Amer. Math. Monthly* **83**:7 (1976), 521–538. MR 53 #14593 Zbl 0349.60014
- [Stein and Shakarchi 2003] E. M. Stein and R. Shakarchi, *Fourier analysis: an introduction*, Princeton Lectures in Analysis **1**, Princeton University Press, 2003. MR 2004a:42001 Zbl 1026.42001

- [Terawaki et al. 2006] 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”, *J. Dairy Sci.* **89**:10 (2006), 4058–4065.
- [Weibull 1951] W. Weibull, “A statistical distribution function of wide applicability”, *J. Appl. Mech.* **18** (1951), 293–297. Zbl 0042.37903
- [Whittaker and Watson 1996] E. T. Whittaker and G. N. Watson, *A course of modern analysis: an introduction to the general theory of infinite processes and of analytic functions, with an account of the principal transcendental functions*, Reprint of the 4th ed., Cambridge University Press, 1996. MR 97k:01072 Zbl 0951.30002
- [Yiannoutsos 2009] C. T. Yiannoutsos, “Modeling AIDS survival after initiation of antiretroviral treatment by Weibull models with changepoints”, *J. Int. AIDS Soc.* **12**:9 (2009).
- [Zhao et al. 2011] Y. Zhao, A. H. Lee, K. K. W. Yau, and G. J. McLachlan, “Assessing the adequacy of Weibull survival models: a simulated envelope approach”, *J. Appl. Stat.* **38**:10 (2011), 2089–2097. MR 2843245

Received: 2014-07-31

Revised: 2014-10-19

Accepted: 2014-12-01

vcuff@g.clemson.edu

*Department of Mathematics, Clemson University,
Clemson, SC 29634, United States*

allewis2@ncsu.edu

*Department of Mathematics, North Carolina State University,
Raleigh, NC 27695, United States*

sjm1@williams.edu

*Department of Mathematics and Statistics, Williams College,
Williamstown, MA 01267, United States*

MANAGING EDITOR

Kenneth S. Berenhaut, Wake Forest University, USA, berenhks@wfu.edu

BOARD OF EDITORS

| | | | |
|----------------------|---|-------------------------|---|
| Colin Adams | Williams College, USA colin.c.adams@williams.edu | David Larson | Texas A&M University, USA larson@math.tamu.edu |
| John V. Baxley | Wake Forest University, NC, USA baxley@wfu.edu | Suzanne Lenhart | University of Tennessee, USA lenhart@math.utk.edu |
| Arthur T. Benjamin | Harvey Mudd College, USA benjamin@hmc.edu | Chi-Kwong Li | College of William and Mary, USA ckli@math.wm.edu |
| Martin Bohner | Missouri U of Science and Technology, USA bohner@mst.edu | Robert B. Lund | Clemson University, USA lund@clemson.edu |
| Nigel Boston | University of Wisconsin, USA boston@math.wisc.edu | Gaven J. Martin | Massey University, New Zealand g.j.martin@massey.ac.nz |
| Amarjit S. Budhiraja | U of North Carolina, Chapel Hill, USA budhiraj@email.unc.edu | Mary Meyer | Colorado State University, USA meyer@stat.colostate.edu |
| Pietro Cerone | La Trobe University, Australia P.Cerone@latrobe.edu.au | Emil Minchev | Ruse, Bulgaria eminchev@hotmail.com |
| Scott Chapman | Sam Houston State University, USA scott.chapman@shsu.edu | Frank Morgan | Williams College, USA frank.morgan@williams.edu |
| Joshua N. Cooper | University of South Carolina, USA cooper@math.sc.edu | Mohammad Sal Moselehian | Ferdowsi University of Mashhad, Iran moslehian@ferdowsi.um.ac.ir |
| Jem N. Corcoran | University of Colorado, USA corcoran@colorado.edu | Zuhair Nashed | University of Central Florida, USA znashed@mail.ucf.edu |
| Toka Diagana | Howard University, USA tdiagana@howard.edu | Ken Ono | Emory University, USA ono@mathcs.emory.edu |
| Michael Dorff | Brigham Young University, USA mdorff@math.byu.edu | Timothy E. O'Brien | Loyola University Chicago, USA tobrie1@luc.edu |
| Sever S. Dragomir | Victoria University, Australia sever@matilda.vu.edu.au | Joseph O'Rourke | Smith College, USA orourke@cs.smith.edu |
| Behrouz Emamizadeh | The Petroleum Institute, UAE bemamizadeh@pi.ac.ae | Yuval Peres | Microsoft Research, USA peres@microsoft.com |
| Joel Foisy | SUNY Potsdam foisyjs@potsdam.edu | Y.-F. S. Pétermann | Université de Genève, Switzerland petermann@math.unige.ch |
| Errin W. Fulp | Wake Forest University, USA fulp@wfu.edu | Robert J. Plemmons | Wake Forest University, USA rplemmons@wfu.edu |
| Joseph Gallian | University of Minnesota Duluth, USA jgallian@d.umn.edu | Carl B. Pomerance | Dartmouth College, USA carl.pomerance@dartmouth.edu |
| Stephan R. Garcia | Pomona College, USA stephan.garcia@pomona.edu | Vadim Ponomarenko | San Diego State University, USA vadim@sciences.sdsu.edu |
| Anant Godbole | East Tennessee State University, USA godbole@etsu.edu | Bjorn Poonen | UC Berkeley, USA poonen@math.berkeley.edu |
| Ron Gould | Emory University, USA rg@mathcs.emory.edu | James Propp | U Mass Lowell, USA jpropp@cs.uml.edu |
| Andrew Granville | Université Montréal, Canada andrew@dms.umontreal.ca | József H. Przytycki | George Washington University, USA przytyck@gwu.edu |
| Jerrold Griggs | University of South Carolina, USA griggs@math.sc.edu | Richard Rebarber | University of Nebraska, USA rrebarbe@math.unl.edu |
| Sat Gupta | U of North Carolina, Greensboro, USA sngupta@uncg.edu | Robert W. Robinson | University of Georgia, USA rwr@cs.uga.edu |
| Jim Haglund | University of Pennsylvania, USA jhaglund@math.upenn.edu | Filip Saidak | U of North Carolina, Greensboro, USA f_saidak@uncg.edu |
| Johnny Henderson | Baylor University, USA johnny_henderson@baylor.edu | James A. Sellers | Penn State University, USA sellersj@math.psu.edu |
| Jim Hoste | Pitzer College jhoste@pitzer.edu | Andrew J. Sterge | Honorary Editor andy@ajsterge.com |
| Natalia Hritonenko | Prairie View A&M University, USA nahritonenko@pvamu.edu | Ann Trenk | Wellesley College, USA atrenk@wellesley.edu |
| Glenn H. Hurlbert | Arizona State University, USA hurlbert@asu.edu | Ravi Vakil | Stanford University, USA vakil@math.stanford.edu |
| Charles R. Johnson | College of William and Mary, USA crjohnso@math.wm.edu | Antonia Vecchio | Consiglio Nazionale delle Ricerche, Italy antonia.vecchio@cnr.it |
| K. B. Kulasekera | Clemson University, USA kk@ces.clemson.edu | Ram U. Verma | University of Toledo, USA verma99@msn.com |
| Gerry Ladas | University of Rhode Island, USA gladas@math.uri.edu | John C. Wierman | Johns Hopkins University, USA wierman@jhu.edu |
| | | Michael E. Zieve | University of Michigan, USA zieve@umich.edu |

PRODUCTION

Silvio Levy, Scientific Editor


Cover: Alex Scorpan

See inside back cover or msp.org/involve for submission instructions. The subscription price for 2015 is US \$140/year for the electronic version, and \$190/year (+\$35, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFLOW[®] from Mathematical Sciences Publishers.

PUBLISHED BY

 **mathematical sciences publishers**
nonprofit scientific publishing

<http://msp.org/>

© 2015 Mathematical Sciences Publishers

involve

2015

vol. 8

no. 5

| | |
|--|-----|
| A simplification of grid equivalence NANCY SCHERICH | 721 |
| A permutation test for three-dimensional rotation data DANIEL BERO AND MELISSA BINGHAM | 735 |
| Power values of the product of the Euler function and the sum of divisors function LUIS ELESBAN SANTOS CRUZ AND FLORIAN LUCA | 745 |
| On the cardinality of infinite symmetric groups MATT GETZEN | 749 |
| Adjacency matrices of zero-divisor graphs of integers modulo n MATTHEW YOUNG | 753 |
| Expected maximum vertex valence in pairs of polygonal triangulations TIMOTHY CHU AND SEAN CLEARY | 763 |
| Generalizations of Pappus' centroid theorem via Stokes' theorem COLE ADAMS, STEPHEN LOVETT AND MATTHEW McMILLAN | 771 |
| A numerical investigation of level sets of extremal Sobolev functions STEFAN JUHNKE AND JESSE RATZKIN | 787 |
| Coalitions and cliques in the school choice problem SINAN AKSOY, ADAM AZZAM, CHAYA COPPERSMITH, JULIE GLASS, GIZEM KARAALI, XUEYING ZHAO AND XINJING ZHU | 801 |
| The chromatic polynomials of signed Petersen graphs MATTHIAS BECK, ERIKA MEZA, BRYAN NEVAREZ, ALANA SHINE AND MICHAEL YOUNG | 825 |
| Domino tilings of Aztec diamonds, Baxter permutations, and snow leopard permutations BENJAMIN CAFFREY, ERIC S. EGGE, GREGORY MICHEL, KAILEE RUBIN AND JONATHAN VER STEEGH | 833 |
| The Weibull distribution and Benford's law VICTORIA CUFF, ALLISON LEWIS AND STEVEN J. MILLER | 859 |
| Differentiation properties of the perimeter-to-area ratio for finitely many overlapped unit squares PAUL D. HUMKE, CAMERON MARCOTT, BJORN MELLEM AND COLE STIEGLER | 875 |
| On the Levi graph of point-line configurations JESSICA HAUSCHILD, JAZMIN ORTIZ AND OSCAR VEGA | 893 |



1944-4176(2015)8:5;1-2