Benfordness of the Generalized Gamma Distribution

Z. Bi, I. Durmić, and S.J. Miller*

February 2, 2022

Abstract - The generalized gamma distribution shows up in many problems related to engineering, hydrology as well as survival analysis. Earlier work has been done that estimated the deviation of the exponential and the Weibull distribution from Benford's Law. We give a mathematical explanation for the Benfordness of the generalized gamma distribution and present a measure for the deviation of the generalized gamma distribution from the Benford distribution.

Keywords : Benford's Law; Generalized Gamma Distribution; Digit Bias; Poisson Summation

Mathematics Subject Classification (2020) **:** 11K06; 62P99; 42A16; 33B15

1 Introduction and an Overview of the Theory

At the dawn of the 20^{th} century, the astronomer and mathematician Simon Newcomb observed that the logarithmic books at his workplace showed a lot of wear and tear at the early pages, but the more he progressed through the book, the less usage could be observed. Newcomb deduced that his colleagues had a "bias" towards numbers starting with the digit 1. In particular, the digit 1 shows up as the first digit roughly 30% of the time, the digit 2 about 9% of the time, and so on. While he did come up with a mathematical model for this interesting relationship, his work stayed mostly unnoticed.

It took another 57 years after Newcomb's discovery for physicist Frank Benford to make the exact same observation as Newcomb: the first pages of logarithmic tables were used far more than others. He formulated this law as follows.

Definition 1.1 [2] *The frequency of first digits follows closely the logarithmic relation:*

$$F_d = \log_{10} \frac{d+1}{d}, \tag{1}$$

where d represents the leading digit, and F_d represents the frequency of the digit d.

*This work was done as part of the Benford Summer 2021 Group led by Professor Steven J. Miller. We would like to thank everyone involved with the Polymath Jr. REU for enabling students to do meaningful research even during a pandemic. The second author was supported by the Williams College John & Louise Finnerty Class of 1971 Fund for Applied Mathematical Research.

Nowadays, Benford's Law is used in detecting many different forms of fraud, and its prevalence in the world fascinates not only mathematicians, but many other scientists as well (to learn more about Benford's Law and its many applications, we recommend [3, 16, 11] to name a few).

Many mathematicians have tried to explain the prevalence of Benford's law in the real world. Some have shown that samples coming from certain probability distributions tend to demonstrate Benfordness [4, 14]. We adopt the same methodology in this article and prove that data coming from a generalized gamma distribution is likely to be close to Benford's law, and provide an explicit formula to bound the deviation.

Remark 1.2 *It is worthwhile to make a quick comment about Benfordness. For any finite set, it is impossible to have a perfect fit so we analyze whether or not such a data set is close to Benford which is enough for most applications. If we let the size of the data set tend to infinity, then there is a chance it will converge to Benford.*

1.1 Benford's Law

While the formulation of Benford's Law, as presented in Definition 1.1, does have its merits, we want to move away from the idea of frequency and data sets to develop a more probabilistic formualation of the theory. We first present a more complete definition of the Benford distribution for base B.

Definition 1.3 *A set of numbers is said to satisfy Benford's law if the leading digit* $d \in \{1, 2, ..., B-1\}$ *occurs with frequency* $F_d = \log_B \frac{d+1}{d}$, *where* $B \geq 2$.

Remark 1.4 *Note that this is the most common way of stating Benford's law. A more general version of the law also describes the frequencies of different second digits, third digits and so on. One could also give an expression for the probability of a digit occuring in the* n^{th} *spot of a number, as is described in the first chapter of* [11]. *This is related to the so-called Strong Benford's Law which states that the probability of observing a significand of at most* x *in base* B *is equal to* $\log_B x$. *The distribution of just the first digit, as well as the distribution of the entire significand, is often referred to as just Benford's law. For the purposes of our paper, this difference is inconsequential.*

We now move on to the idea of base and scientific notation. Given a base $B \geq 2$, any nonzero real number r can be uniquely expressed in the form $r = aB^n$, where $|a| \in [1, B), n \in \mathbb{Z}$. This is usually referred to as **scientific notation**, and it motivates the definition of the **significand**.

Definition 1.5 *Given a base* $B \geq 2$, *we define the significand as the mapping* $S_B : \mathbb{R}_{\neq 0} \to [1, B)$, *where* $S_B(x)$ *is the significand of any input* $x = aB^n$ *written in scientific notation with* $|a| \in [1, B)$ *and* $n \in \mathbb{Z}$. *It follows that* $S_B(x) = |a|$ *is the* ***significand*** *and* n *is the* ***exponent***.

One also studies the **mantissa**, which is the fractional part of the logarithm.

Example 1.1 As an example, let $x = 31295192$. When we write this in scientific notation using base 10, it becomes $x = 3.1295192 \cdot 10^7$ and it follows that $S_B(3.1295192 \cdot 10^7) = 3.1295192$ is the significand, and 7 is the exponent. Furthermore, observe that $\log_{10} 31295192 \approx 7.495477620349604$ so the mantissa is about 0.495477620349604.

We now formally introduce the notion of what it means for a random variable to have the Benford distribution [1].

Definition 1.6 *A random variable X is Benford base B if* $\text{Prob}(X \leq x) = \log_B x$, *where $x \in [1, B)$, $B \geq 2$.*

Let $X : \Omega \to \mathbb{R}$, be a random variable with some particular cumulative distribution function (cdf) F. If the random variable $S_B \circ X$ is Benford (or close to Benford), then one would expect a data set coming from a population with distribution F to satisfy Benford's law since

$$\text{Prob}(X \text{ has leading digit } d) = \text{Prob}(S_B \circ X \in [d, d+1)) = \log_B \frac{d+1}{d}, \tag{2}$$

which is just a direct application of Definition 1.6.

We could directly find the distribution of $S_B \circ X$ and compare it to the Benford distribution. Alternatively, the following well-known theorem, which can be found in [5], provides an indirect method, which in some cases is more convenient.

Theorem 1.7 *Given a base $B \geq 2$ and a non-negative random variable X, $S_B \circ X$ is Benford if and only if* $\log_B X \bmod 1$ *has a uniform* $[0, 1)$ *distribution.*

Proof. For any $s \in [1, B)$, let $u = \log_B s \in [0, 1)$. Let us assume that $S_B \circ X$ is Benford, so the following holds:

$$\begin{aligned} \text{Prob}(\log_B X \bmod 1 \in [0, u)) &= \text{Prob}\left(\{X \in [1 \cdot B^k, s \cdot B^k) : k \in \mathbb{Z}\}\right) \\ &= \text{Prob}(S_B \circ X \in [1, s)) \\ &= \log_B s \\ &= u. \end{aligned} \tag{3}$$

Hence it follows that $\log_B X \bmod 1$ has uniform $[0, 1)$ distribution.
Let us now assume that $\log_B X \bmod 1$ has uniform $[0, 1)$ distribution. Then:

$$\begin{aligned} \text{Prob}(S_B \circ X \in [1, s)) &= \text{Prob}\left(\{X \in [1 \cdot B^k, s \cdot B^k) : k \in \mathbb{Z}\}\right) \\ &= \text{Prob}(\log_B X \bmod 1 \in [0, u)) \\ &= u \\ &= \log_B s. \end{aligned} \tag{4}$$

[1] Note that, throughout this paper, we denote that a random variable X follows the Benford distribution base B by stating that X is Benford base B

□

Theorem 1.7 forms the key foundation of our work since it reduces our problem substantially by allowing us to focus on a logarithmically rescaled random variable mod1. We use this result to explore the relationship between the generalized gamma distribution and Benford's law.

It is exactly Theorem 1.7 which enables us to develop a measure for the deviation of a random variable from the Benford distribution. This deviation, which we denote as D, quantifies the deviation of the leading digits from their corresponding Benford counterparts. We derive an estimate for this deviation.

Assume X is a nonnegative random variable and that $S_B \circ Y$ is Benford. We now transform the random variable X via the transformation noted in Theorem 1.7 to $\log_B X \bmod 1$, and denote its probability density function (pdf) as: $f_{\log_B X \bmod 1}$. From here it follows that

$$\begin{aligned}
D &= \left|\text{Prob}\left(X \text{ has leading digit } d\right) - \log_B \frac{d+1}{d}\right| \\
&= \left|\text{Prob}\left(S_B \circ X \in [d, d+1)\right) - \text{Prob}\left(S_B \circ Y \in [d, d+1)\right)\right| \\
&= \Big|\text{Prob}\left(\{X \in [d \cdot B^k, (d+1) \cdot B^k) : k \in \mathbb{Z}\}\right) - \\
&\qquad \text{Prob}\left(\{Y \in [d \cdot B^k, (d+1) \cdot B^k) : k \in \mathbb{Z}\}\right)\Big| \\
&= \Big|\text{Prob}\left(\log_B X \bmod 1 \in [\log_B d, \log_B d + 1)\right) - \\
&\qquad \text{Prob}\left(\log_B Y \bmod 1 \in [\log_B d, \log_B d + 1)\right)\Big| \\
&= \left|\int_{\log_B d}^{\log_B d+1} f_{\log_B X \bmod 1}(u) - 1 \; du\right| \\
&\leq \int_0^1 \left|f_{\log_B X \bmod 1}(u) - 1\right| du. \qquad (5)
\end{aligned}$$

This expression allows us to bound the deviation of any random variable from the Benford distribution. We now present a quick theoretical overview of the generalized gamma distribution.

1.2 The Generalized Gamma Distribution and Its Connection to Benford

The work done by Miller and Nigrini in [14], as well as the paper by Leemis, Schmeiser, and Evans [10], explored the exponential distribution and how it relates to Benford's Law, whereas Cuff et.al in [4] explored a similar form of a relationship between the Weibull distribution and Benford's Law. Both of these distributions can be seen as "children" of one parent distribution for a particular choice of parameters.

For the purposes of this paper, we use the following definition of the generalized gamma distribution, as presented in [20].

Definition 1.8 *A random variable X follows the generalized gamma distribution with*

parameters a, d, and p if its cumulative distribution function (cdf) is of the form

$$F(x;a,d,p) \;=\; \frac{\gamma\left(\frac{d}{p},\left(\frac{x}{a}\right)^p\right)}{\Gamma\left(\frac{d}{p}\right)}, \quad x>0;\ a,d,p>0, \tag{6}$$

where γ is the lower incomplete gamma function, defined as

$$\gamma(s,x) \;:=\; \int_0^x t^{s-1}e^{-t}dt. \tag{7}$$

The corresponding probability density function (pdf) is

$$f(x;a,d,p) \;=\; \frac{\left(\frac{p}{a^d}\right)x^{d-1}e^{-(x/a)^p}}{\Gamma\left(\frac{d}{p}\right)}. \tag{8}$$

We note that, when $d = p$, equation (6) is just the cdf of a Weibull distribution, and that is further reduced to the exponential distribution for the special case of $d = p = 1$. This is a very useful observation because it enables us to directly compare our results with the work completed in [4, 14].

Having covered the relevant background material, our goal now is to show the following key results, that we prove in Section 2.

Theorem 1.9 *If X is a random variable having the generalized gamma distribution with parameters a, d, p, then the pdf of $\log_B X \bmod 1$ is*

$$f_{\log_B X \bmod 1}(u) \;=\; \frac{p\ln B}{\Gamma\left(\frac{d}{p}\right)}\sum_{k=-\infty}^{+\infty} e^{-\left(\frac{B^{k+u}}{a}\right)^p}\left(\frac{B^{k+u}}{a}\right)^d, \tag{9}$$

or

$$f_{\log_B X \bmod 1}(u) \;=\; 1+\sum_{k=1}^{+\infty}\frac{2}{\Gamma\left(\frac{d}{p}\right)}\operatorname{Re}\left[e^{2\pi ui-\frac{2\pi i\ln a}{\ln B}}\Gamma\left(\frac{d}{p}-\frac{2\pi ki}{p\ln B}\right)\right]. \tag{10}$$

where $u\in(0,1)$. Further, the scaling parameter a has limited effect on the pdf, for any $m\in\mathbb{Z}$, a and $a\cdot B^m$ result in the same pdf.

Theorem 1.10 *Given $\epsilon>0$, and the second form of the pdf of $\log_B X \bmod 1$, i.e. equation (10) in Theorem 1.9, we have*

$$f_{\log_B X \bmod 1}(u) \;=\; 1+\sum_{k=1}^{+\infty}\frac{2}{\Gamma\left(\frac{d}{p}\right)}\operatorname{Re}\left[e^{2\pi ui-\frac{2\pi i\ln a}{\ln B}}\Gamma\left(\frac{d}{p}-\frac{2\pi ki}{p\ln B}\right)\right].$$

To approximate this function with main term 1 *and first* M*-term partial sum of the residue*

$$f^{M}_{\log_B X \bmod 1}(u) \quad = \quad 1 + \sum_{k=1}^{M} \frac{2}{\Gamma\left(\frac{d}{p}\right)} \operatorname{Re}\left[e^{2\pi u i - \frac{2\pi i \ln a}{\ln B}} \Gamma\left(\frac{d}{p} - \frac{2\pi k i}{p \ln B}\right)\right], \tag{11}$$

to make sure the approximation error at any point $u \in (0,1)$ *is bounded by* ϵ*,* M *should satisfy*

$$M \quad > \quad \frac{(d+p)^2 (\ln(B))^2}{2\pi^2 \epsilon} - 1. \tag{12}$$

One can gain a lot of intuition for the behavior of the distribution by analysing the graph of the pdf for different parameters. In particular, the parameters d and p determine the shape of the pdf, while the parameter a determines the spread of the pdf.

We also present some simulations, along with figures, that show how close the generalized gamma distribution comes to Benford's Law. The purpose of these figures is to show us that the deviation from Benford should be relatively low.

Figure 1 compares the first-digit frequencies of 10000 samples from a generalized gamma distribution, using the parameters: $B = 10, a = 2, d = 1, p = \frac{1}{2}$, with the frequencies predicted by Benford's law. Observe that Benford's law provides us with a remarkably good fit.

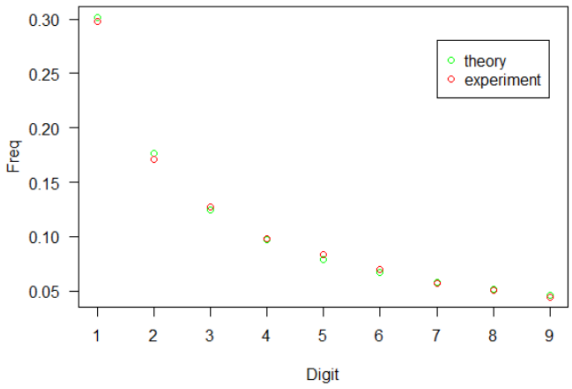

Figure 1: First-digit frequencies of 10000 samples from a generalized gamma distribution $\left(B = 10, a = 2, d = 1, p = \frac{1}{2}\right)$.

The Benfordness of samples coming from the generalized gamma distribtion can also be observed in a different way. We already stated in Theorem 1.7 that a random variable X is Benford if and only if $\log_B X \bmod 1$ is uniform $[0,1)$. Since our claim is that X is close to Benford if it follows the generalized gamma distribution, then $\log_B X \bmod 1$ should have to be close to the uniform $[0,1)$ distribution. We explore this further using a Kolmogorov-Smirnov test.

The Kolmogorov-Smirnov test is used to examine whether or not a sample comes from a population with a specific distribution. The smaller the test statistic is, the

more likely the sample came from the target distribution. We generated samples from the generalized gamma distribution with different values of parameters d and p, and performed a Kolmogorov-Smirnov test to compare the transformed data ($\log_B X \bmod 1$) with the uniform $[0, 1)$ distribution. The result is shown in Figure 2. Observe that the test statistic is pretty small, indicating the (transformed) data came from a population with an approximately uniform $[0, 1)$ distribution, and hence the original distribution is close to Benford. The match is better when d and p are small, which is reasonable considering that we know how close the exponential and Weibull distributions come to the Benford distribution.

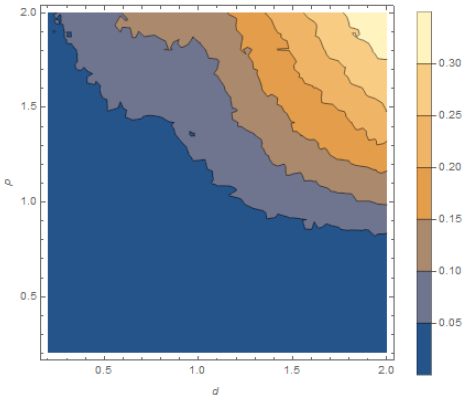

Figure 2: Kolmogorov-Smirnov test results under different values of d and p.

2 Main Results and Key Observations

In this section, we prove the results of Theorem 1.9 and Theorem 1.10, and justify the observations shown in Figures 1 and 2.

Proof. [Proof of Theorem 1.9] Given $u \in [0, 1)$, we have

$$\begin{aligned} \text{Prob}(\log_B X \bmod 1 \in [0, u]) &= \sum_{k=-\infty}^{\infty} \text{Prob}(\log_B X \in [k, k+u]) \\ &= \sum_{k=-\infty}^{\infty} \text{Prob}(X \in [B^k, B^{k+u}]) \\ &= \frac{1}{\Gamma\left(\frac{d}{p}\right)} \sum_{k=-\infty}^{\infty} \int_{\left(\frac{B^k}{a}\right)^p}^{\left(\frac{B^{k+u}}{a}\right)^p} t^{\frac{d}{p}-1} e^{-t} dt. \end{aligned} \tag{13}$$

By using some results from analysis, the following can be verified.

1. The resulting function (13) converges for all $u \in [0, 1)$.

2. The resulting function (13) is differentiable for all $u \in (0,1)$, and we can differentiate it term by term.

Appendix A provides the interested reader with more details of the proof as well as the mathematical machinery used in the paper.

We now work with the integral from (13) to get the probability density function. Using the Fundamental Theorem of Calculus we find

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}u}\int_{\left(\frac{B^k}{a}\right)^p}^{\left(\frac{B^{k+u}}{a}\right)^p} t^{\frac{d}{p}-1}e^{-t}dt &= e^{-\left(\frac{B^{k+u}}{a}\right)^p}\left(\left(\frac{B^{k+u}}{a}\right)^p\right)^{\frac{d}{p}-1} p\left(\frac{B^{k+u}}{a}\right)^{p-1}\frac{B^{k+u}\ln B}{a} \\ &= e^{-\left(\frac{B^{k+u}}{a}\right)^p}\left(\frac{B^{k+u}}{a}\right)^d p\ln B, \end{aligned} \tag{14}$$

and then plugging (14) back into (13), we get that the pdf of $\log_B X \bmod 1$ is

$$f_{\log_B X \bmod 1}(u) = \frac{p\ln B}{\Gamma\left(\frac{d}{p}\right)}\sum_{k=-\infty}^{+\infty} e^{-\left(\frac{B^{k+u}}{a}\right)^p}\left(\frac{B^{k+u}}{a}\right)^d. \tag{15}$$

Next we apply Poisson summation to (9) to get the equivalent form (10), which is better since it is divided into a main term, 1, which is what we want, and a residue term given by an infinite series. See Appendix A or [4] for details about Poisson summation.

For any $u \in (0,1)$, let $z = B^u$, $t = k$. We claim that

$$g(t) = p\ln B e^{-\left(\frac{B^t z}{a}\right)^p}\left(\frac{B^t z}{a}\right) \tag{16}$$

satisfies the conditions for applying Poisson summation. The details of why this is true can be found in Appendix A, but the important insight is that we have the Fourier transform of this function:

$$\begin{aligned}\hat{g}(f) &= \int_{-\infty}^{+\infty} p\ln B e^{-\left(\frac{B^t z}{a}\right)^p}\left(\frac{B^t z}{a}\right)^d e^{-2\pi i t f}dt \\ &= \int_0^{\infty} e^{-\omega}\omega^{\frac{d}{p}-1}\left(\frac{a\omega^{\frac{1}{p}}}{z}\right)^{-\frac{2\pi i f}{\ln B}} d\omega \quad , \text{ where } \omega = \left(\frac{B^t z}{a}\right)^p \\ &= \left(\frac{z}{a}\right)^{\frac{2\pi i f}{\ln B}}\Gamma\left(\frac{d}{p}-\frac{2\pi i f}{p\ln B}\right). \end{aligned} \tag{17}$$

Now we use the above result to apply Poisson summation to equation (9), which gives us equation (10).

$$
\begin{aligned}
f_{\log_B X \bmod 1}(u) &= \frac{1}{\Gamma\left(\frac{d}{p}\right)} \sum_{k=-\infty}^{\infty} g(k) \\
&= \frac{1}{\Gamma\left(\frac{d}{p}\right)} \sum_{k=-\infty}^{\infty} \hat{g}(k) \\
&= \frac{1}{\Gamma\left(\frac{d}{p}\right)} \sum_{k=-\infty}^{\infty} \left(\frac{z}{a}\right)^{\frac{2\pi i k}{\ln B}} \Gamma\left(\frac{d}{p} - \frac{2\pi i k}{p \ln B}\right) \\
&= 1 + \sum_{k=1}^{\infty} \left[\left(\frac{z}{a}\right)^{\frac{2\pi i k}{\ln B}} \Gamma\left(\frac{d}{p} - \frac{2\pi i k}{p \ln B}\right) + \left(\frac{z}{a}\right)^{-\frac{2\pi i k}{\ln B}} \Gamma\left(\frac{d}{p} + \frac{2\pi i k}{p \ln B}\right)\right] \\
&= 1 + \sum_{k=1}^{+\infty} \frac{2}{\Gamma\left(\frac{d}{p}\right)} \operatorname{Re}\left[\left(\frac{z}{a}\right)^{-\frac{2\pi i k}{\ln B}} \Gamma\left(\frac{d}{p} - \frac{2\pi k i}{p \ln B}\right)\right] \\
&= 1 + \sum_{k=1}^{+\infty} \frac{2}{\Gamma\left(\frac{d}{p}\right)} \operatorname{Re}\left[e^{2\pi u i - \frac{2\pi i \ln a}{\ln B}} \Gamma\left(\frac{d}{p} - \frac{2\pi k i}{p \ln B}\right)\right], \qquad (18)
\end{aligned}
$$

where for the second last equality, we used the property $\Gamma(\overline{z}) = \overline{\Gamma(z)}$.

Finally, observe that $e^{2\pi i x} = e^{2\pi i(x+1)}$ for all $x \in \mathbb{R}$, which verifies the scaling invariance of a. □

Remark 2.1 *It is worth noting that, when we reduce the results from equation* (9) *and equation* (10) *to the Weibull case, we retrieve the same results that were shown in the article by Cuff et.al* [4].

The following result, expressed as Theorem 1.10, enables us to estimate the value of the pdf numerically.

Proof. [Proof of Theorem 1.10] For any $M \geq 1$, the approximation error is

$$
\begin{aligned}
|\eta| &= \left| \sum_{k=M+1}^{+\infty} \frac{2}{\Gamma\left(\frac{d}{p}\right)} \operatorname{Re}\left[e^{2\pi u i - \frac{2\pi i \ln a}{\ln B}} \Gamma\left(\frac{d}{p} - \frac{2\pi k i}{p \ln B}\right)\right] \right| \\
&\leq \sum_{k=M+1}^{+\infty} \frac{2}{\Gamma\left(\frac{d}{p}\right)} \left| \Gamma\left(\frac{d}{p} - \frac{2\pi k i}{p \ln B}\right) \right|, \qquad (19)
\end{aligned}
$$

where $u \in (0, 1)$.

The gamma function has the property that

$$
|\Gamma(a + bi)|^2 = |\Gamma(a)|^2 \prod_{k=0}^{\infty} \frac{1}{1 + \frac{b^2}{(a+k)^2}}, \qquad (20)
$$

and applying this to (19), we get

$$\begin{aligned}
\left|\Gamma\left(\frac{d}{p}-\frac{2\pi ki}{p\ln B}\right))\right|^2 &= \left[\Gamma\left(\frac{d}{p}\right)\right]^2\prod_{l=0}^{\infty}\frac{1}{1+\frac{\left(\frac{2\pi k}{p\ln B}\right)^2}{\left(\frac{d}{p}+l\right)^2}} \\
&\le \left[\Gamma\left(\frac{d}{p}\right)\right]^2\prod_{l=0}^{1}\frac{1}{1+\frac{\left(\frac{2\pi k}{p\ln B}\right)^2}{\left(\frac{d}{p}+l\right)^2}} \\
&\le \left[\Gamma\left(\frac{d}{p}\right)\right]^2\frac{[(d+p)\ln B]^4}{(2\pi k)^4}, \qquad (21)
\end{aligned}$$

where the first inequality is because all terms in the product are positive numbers less than or equal to 1. Finally we have

$$\begin{aligned}
|\eta| &\le \sum_{k=M+1}^{+\infty}\frac{[(d+p)\ln B]^2}{2\pi^2k^2} \\
&\le \int_{M+1}^{\infty}\frac{[(d+p)\ln B]^2}{2\pi^2x^2}dx \\
&= \frac{[(d+p)\ln B]^2}{2\pi^2(M+1)}. \qquad (22)
\end{aligned}$$

Letting $|\eta| < \epsilon$, and using the result from (22), we get the lower bound for M as in (12). □

With the help of (5) and Theorem 1.10, for a random variable following the generalized gamma distribution with parameters (a, d, p), we have the following for its deviation:

$$\begin{aligned}
D &= \left|P(X \text{ has leading digit } d) - \log_B\frac{d+1}{d}\right| \\
&\le \int_0^1 |f(u)-1|du \\
&\le \int_0^1 |f(u)-f_M(u)|du + \int_0^1 |f_M(u)-1|du \\
&\le \epsilon + \sup_{u\in(0,1)}|f_M(u)-1|, \qquad (23)
\end{aligned}$$

where f and f_M are exact and approximate pdfs of $\log_B X \bmod 1$. Since we can control ϵ (which then determines M), and $\sup_{u\in(0,1)}|f_M(u)-1|$ can be evaluated (at least) numerically, we can get an upper bound for the difference of $S_B \circ X$ from the Benford distribution for any given parameters a, d and p.

Figure 3 shows the graphs of some approximate pdfs (with approximation error < 0.01) of $\log_B X \bmod 1$ with different parameters, which are pretty close to the constant function 1. The term 1 in (10) plays a major role in the function.

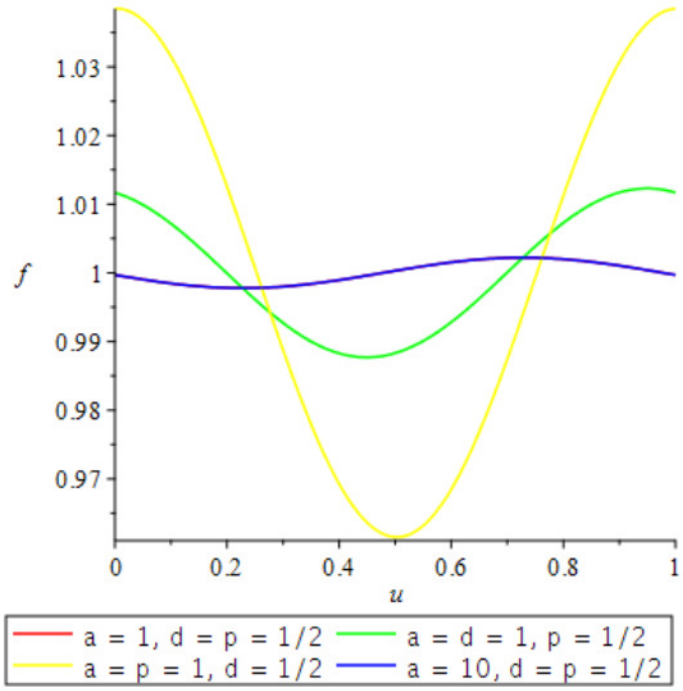

Figure 3: Approximation of pdfs of $\log_B X \bmod 1$ with different parameters.

Figure 4 shows the upper bound of D with respect to d and p according to (23). We see X is very close to Benford when the parameters d and p are small, which is consistent with the Kolmogorov-Smirnov test we showed in Figure 2.

3 Conclusion and Future Work

We have shown that the generalized gamma distribution for the right choice of parameters, meaning a relatively small p and d, conforms well to Benford's Law for the leading digit. It would be interesting to see where we can find an application of this result, considering how common the generalized gamma distribution is in nature.

A possible research avenue would be to perform a similar analysis for other families of distributions.

Appendix A Details of Proofs in Section 2

In the proof of Theorem 1.9, we claimed the following.

1. The cdf of $\log_B X \bmod 1$ (13) converges pointwise for all $u \in [0, 1)$, it is differentiable for all $u \in (0, 1)$, and we can differentiate it term by term.

2. Poisson summation can be applied to the pdf of $\log_B X \bmod 1$ as stated in (15).

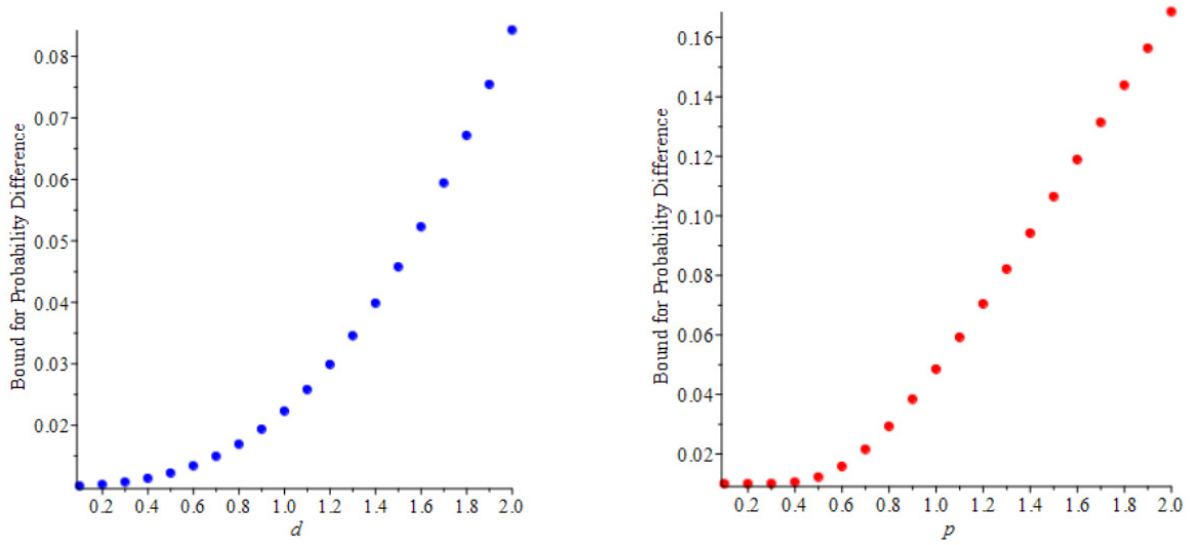

Figure 4: Left: Bound of probability difference (23) with respect to d $(a = 1,\ p = 0.5)$; Right: Bound of probability difference (23) with respect to p $(a = 1,\ d = 0.5)$.

We use the following results to prove these facts. Many of these proofs are standard and the arguments below are provided in order to give the reader a quick overview of the key ideas.

Theorem A.1 (**Weierstrass M-test**) *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of real valued functions on a set X, suppose each $|f_n|$ is bounded by $M_n \geq 0$, then if $\sum_{n=1}^{\infty} M_n$ converges, $\sum_{n=1}^{\infty} f_n$ converges uniformly.*

Theorem A.2 *If $\{f_n\}_{n=1}^{\infty}$ is a sequence of C^1 functions on (a, b), if both $\sum_{n=1}^{\infty} f_n$ and $\sum_{n=1}^{\infty} f_n'$ converge uniformly, then $\sum_{n=1}^{\infty} f_n$ is differentiable and $(\sum_{n=1}^{\infty} f_n)' = \sum_{n=1}^{\infty} f_n'$.*

Theorem A.3 (**Poisson Summation**) *Let f, f' and f'' be continuous functions which eventually decay at least as fast as $x^{-(1+\eta)}$ for some $\eta > 0$, then*

$$\sum_{n=-\infty}^{+\infty} f(n) \quad = \quad \sum_{n=-\infty}^{+\infty} \hat{f}(n), \tag{24}$$

where $\hat{f}(y) = \int_{-\infty}^{+\infty} f(x)e^{-2\pi xyi}dx$ is the Fourier transformation of f.

Our second claim above is a direct result of Theorem A.3. For the first claim to be true, we need to check (13) satisfies the conditions in Theorem A.2.

1. The term in the sum of (13),

$$\int_{\left(\frac{B^k}{a}\right)^p}^{\left(\frac{B^{k+u}}{a}\right)^p} t^{\frac{d}{p}-1}e^{-t}dt \tag{25}$$

is C^1 in $(0,1)$.

2. $F_{\log_B X \bmod 1}$ in (13) converges uniformly in $(0,1)$.

3. $f_{\log_B X \bmod 1}$ in (15) converges uniformly in $(0,1)$.

1) is easy to check. The main reason for the uniform convergence of $F_{\log_B X \bmod 1}$ and $f_{\log_B X \bmod 1}$ is the fast-decay e^{-u} like term in both of them. Next we give a sketch for part of the proof for this. We will show that if $d \geq p$, $F_{\log_B X \bmod 1}$ converges uniformly in $(0,1)$. All other cases could be checked similarly.

Proof. (*Sketch*) For any $u \in (0,1)$, if $d \geq p$, it's easy to check

$$t^{\frac{d}{p}-1}e^{-t} \;\leq\; \left(\frac{B^{k+1}}{a}\right)^{d-p} e^{-\left(\frac{B^k}{a}\right)^p},$$

for all $t \in \left[\left(\frac{B^k}{a}\right)^p, \left(\frac{B^{k+u}}{a}\right)^p\right]$, so we have

$$\int_{\left(\frac{B^k}{a}\right)^p}^{\left(\frac{B^{k+u}}{a}\right)^p} t^{\frac{d}{p}-1}e^{-t}dt \;\leq\; \left(\frac{B^{k+1}}{a}\right)^{d-p} e^{-\left(\frac{B^k}{a}\right)^p} \left(\frac{B^{k+1}}{a}\right)^{d} \quad\leq\quad \left(\frac{B^{k+1}}{a}\right)^{d} e^{-\left(\frac{B^k}{a}\right)^p}.$$

Apply this result to the sum in (13), we get

$$\begin{aligned}
\Gamma\left(\frac{d}{p}\right) F_{\log_B X \bmod 1} &= \sum_{k=-\infty}^{\infty} \int_{\left(\frac{B^k}{a}\right)^p}^{\left(\frac{B^{k+u}}{a}\right)^p} t^{\frac{d}{p}-1}e^{-t}dt \\
&\leq \sum_{k=-\infty}^{\infty} \left(\frac{B^{k+1}}{a}\right)^{d} e^{-\left(\frac{B^k}{a}\right)^p} \\
&\leq \sum_{k=0}^{\infty} \left(\frac{B^{k+1}}{a}\right)^{d} e^{-\left(\frac{B^k}{a}\right)^p} + \sum_{k=1}^{\infty} \left(\frac{B}{aB^k}\right)^{d} e^{-\left(\frac{1}{aB^k}\right)^p}. \qquad (26)
\end{aligned}$$

The two sums in (26) are convergent by the integration test, then by Weierstrass M-test, i.e. Theorem A.1, we know the original sum in (13) converges uniformly. □

Appendix B Simulation Code

R script: Sample from a Generalized Gamma Distribution and compare the first-digit frequencies of the data with values predicted by Benford's law

```
N=10000
a = 2
d = 1/2
p = 1/2
B <- 10 # B should be an integer greater than 1
```

```
sample <- as.vector(qgamma(runif(N), shape=d/p, scale=a^p)^(1/p))

for (i in 1:N) {
  while (sample[i] < 1 | sample[i] >= B) {
    if (sample[i] < 1) {
      sample[i] <- sample[i] * B
    } else {
      sample[i] <- sample[i] / B
    }
  }
  sample[i] <- trunc(sample[i])
}

freqs <- as.numeric(table(sample))

error <- 0.0
freqs_t <- vector("list", B - 1)
freqs_t <- unlist(freqs_t)

for (i in 1:(B-1)) {
  freqs_t[i] <- logb((i + 1) / i, base=B)
  error <- error + (freqs[i] / N - freqs_t[i])^2
}

freqs <- freqs / N
d <- 1:(B-1)

plot(d, freqs, las=1, xlab="Digit", ylab="Freq", col="red", xaxt="n")
axis(1, at=1:(B-1), labels=1:(B-1))
points(d, freqs_t, col="green")
legend(B - 3, max(c(freqs, freqs_t)) * 0.95,
legend=c("theory", "experiment"),
col=c("green", "red"), pch=c(21, 21))
```

Maple code: Plot pdfs and calculate probability deviation bound

```
restart;
with(plots):

g_k:=2/GAMMA(d/p)*GAMMA(d/p-2*Pi*I*k/p/ln(B))
*exp(2*Pi*I*u- 2*Pi*I*ln(a)/ln(B));
f_M := 1 + sum(Re(g_k), k=1..M);
```

```
M := ceil(((d + p)*ln(B))^2/2/Pi/Pi/e - 1);

# set pointwise error bound
e := 0.01;

# set parameters and plot
B:=10; a:=1; d:=1/2; p:=1/2;
p1 := plot(f_M, u=0..1, color="red",
labels=[u, f], legend="a = 1, d = p = 1/2"):

B:=10; a:=1; d:=1; p:=1/2;
p2 := plot(f_M, u=0..1, color="green",
labels=[u, f], legend="a = d = 1, p = 1/2"):

B:=10; a:=1; d:=1/2; p:=1;
p3 := plot(f_M, u=0..1, color="yellow",
labels=[u, f], legend="a = p = 1, d = 1/2"):

B:=10; a:=10; d:=1/2; p:=1/2;
p4 := plot(f_M, u=0..1, color="blue",
labels=[u, f], legend="a = 10, d = p = 1/2"):

display(p1, p2, p3, p4, legendstyle = [font=["HELVETICA", 12]
, location=bottom]);

# calculate bounds for probability difference under different parameters
# d = p = 0.5, a changes from 1 to 10
# unassign('a'); d := 0.5; p := 0.5;
# points:={seq([a, Optimization[Maximize]
(abs(f_M - 1),u = 0..1)[1] + e], a=1..10)};
# pointplot(points, symbol=solidcircle, symbolsize = 15,
color =orange, labels=["a", "Bound for Probability Difference"],
labeldirections=[ "horizontal", "vertical"]);

# a = 1, p = 0.5, d changes from 0.1 to 2
unassign('d'); a := 1; p := 0.5;
ds := seq(n/10, n=1..20);
points:={seq([d, Optimization[Maximize]
(abs(f_M - 1),u = 0..1)[1] + e], d=ds)};
pointplot(points, symbol=solidcircle, symbolsize = 15,
color =blue, labels=[d, "Bound for Probability Difference"],
label directions=[ "horizontal", "vertical"]);
```

```
# a = 1, d = 0.5, p changes from 0.1 to 2
unassign('p'); a := 1; d := 0.5;
ps := seq(n/10, n=1..20);
points:={seq([p, Optimization[Maximize]
(abs(f_M - 1),u = 0..1)[1] + e], p=ps)};
pointplot(points, symbol=solidcircle, symbolsize = 15,
color =red, labels=[p, "Bound for Probability Difference"],
label directions=[ "horizontal", "vertical"]);
```

Mathematica code for the Kolmogorov-Smirnov test

```
Clear[Diff]
Diff[a_, d_, p_, B_] := KolmogorovSmirnovTest[Mod[Log[B,
Random Variate[GammaDistribution[d, a, p, 0], 10^4]], 1],
Uniform Distribution[], "TestStatistic"]
ContourPlot[Diff[1, d, p, 10], {d, 0.2, 2}, {p, 0.2, 2},
Frame Label -> Automatic, PlotLegends -> Automatic]
```

References

[1] M. Abromovich, I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, tenth printing, National Bureau of Standards, Applied Mathematics Series **55**, 1972.

[2] F. Benford, The Law of Anomalous Numbers, *Proceedings of the American Philosophical Society*, **78** (1938), 551-572. Available online at the URL: `https://www.inzichten.nl/contact/images/benford_law.pdf`.

[3] A. Berger, T.P.Hill *An Introduction to Benford's Law*, Princeton University Press, 2015

[4] V. Cuff, A. Lewis and S. J. Miller, The Weibull distribution and Benford's law, *Involve, a Journal of Mathematics*, **8-5** (2015), 859–874. Available online at the URL: `https://arxiv.org/abs/1402.5854`.

[5] P. Diaconis, The distribution of leading digits and uniform distribution mod 1, *Ann. Probab.*, **5** (1979), 72–81. Available online at the URL: `https://projecteuclid.org/journals/annals-of-probability/volume-5/issue-1`.

[6] T. P. Hill, A Statistical Derivation of the Significant-Digit Law, *Statistical Science*, **10** (1995), no. 4, 354-363. Available online at the URL: `https://projecteuclid.org/journals/statistical-science/volume-10/issue-4`.

[7] T. P. Hill, The first-digit phenomenon, *American Scientists*, **86** (1996), 358–363. Available online at the URL: `https://www.jstor.org/stable/27857060?seq=1#metadata_info_tab_contents`.

[8] A. Kontorovich and S. J. Miller, Benford's Law, values of L-functions and the $3x+1$ problem, *Acta Arithmetica*, **120** (2005), no. 3, 269–297. Available online at the URL: `https://arxiv.org/abs/math/0412003`.

[9] A. Y. Khinchin, *Continued Fractions*, Third Edition, The University of Chicago Press, Chicago 1964.

[10] L. M. Leemis, B. W. Schmeiser and D. L. Evans, Survival Distributions Satisfying Benford's Law, *The American Statistician*, **54** (2000), no. 3. Available online at the URL: `https://www.jstor.org/stable/2685773?seq=1#metadata_info_tab_contents`.

[11] S. J. Miller, *Benford's Law: Theory and Applications*, Princeton University Press, Princeton, NJ, 2015.

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

[13] 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. Available online at the URL: `https://arxiv.org/abs/math/0607686`.

[14] S. J. Miller and M. J. Nigrini, Order Statistics and Benford's Law, *International Journal of Mathematics and Mathematical Sciences*, (2008), 1-13. Available online at the URL: `https://www.hindawi.com/journals/ijmms/2008/382948/`.

[15] S. Newcomb, Note on the frequency of use of the different digits in natural numbers, *Amer. J. Math.*, **4** (1881), 39-40. Available online at the URL: `https://www.jstor.org/stable/i340583`.

[16] M. J. Nigrini and S. J. Miller, *Data diagnostics using second order tests of Benford's Law*, John Wiley&Sons, Inc., Hoboken, New Jersey, 2012

[17] M. J. Nigrini, Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection, *Auditing: A Journal of Practice and Theory*, **28** (2009), no. 2, 305–324.

[18] Z. Rudnick, P. Sarnak, and A. Zaharescu, The Distribution of Spacings Between the Fractional Parts of $n^2\alpha$, *Invent. Math.*, **145** (2001), no. 1, 37–57. Available online at the URL: `https://arxiv.org/abs/2004.04855`.

[19] R. A. Raimi, The First Digit Problem, *The American Mathematical Monthly*, **83:7** (1976), no. 7, 521-538. Available online at the URL: https://www.jstor.org/stable/2319349?seq=1#metadata_info_tab_contents.

[20] E. W. Stacy, A Generalization of the Gamma Distribution, *The Annals of Mathematical Statistics*, **33** (1962), no. 3, 1187-1192. Available online at the URL: https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-33/issue-3.

Zelong Bi
University of New South Wales
Sydney NSW, 2052
E-mail: zelong.bi@student.unsw.edu.au

Irfan Durmić
Williams College
Williamstown, MA 01267
E-mail: Irfan.Durmic@williams.edu

Steven J. Miller
Williams College
Williamstown, MA 01267
E-mail: Steven.J.Miller@williams.edu

Received: April 31, 2017 **Accepted**: June 31, 2017
Communicated by Some Editor