Benfordness of the Generalized Gamma Distribution

Z. Bi, I. Durmić, and S.J. Miller*

Abstract - Benford's law of digit bias states that in many data sets it is more likely to see lower leading digits than larger; determining why this is the case is one of the most important problems in the subject. Earlier work determined why exponential and more generally Weibull distributions are close to Benford's law. These two families are special cases of a more general distribution, the generalized gamma distribution, which we prove is often close to Benford's law.

Keywords : Benford's law; Benford distribution; 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 20th 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. He deduced that his colleagues had a "bias" towards numbers starting with the digit 1. In particular, the digit 1 was the leading digit roughly 30% of the time, the digit 2 about 18% 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 observe the same bias in many settings. He formulated a law describing this phenomenon, called Benford's law in his honor, as follows.

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

$$F_D = \log_{10} \frac{D + 1}{D},$$

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, and the referee for numerous comments that improved the paper. The second author was supported by the Williams College John & Louise Finnerty Class of 1971 Fund for Applied Mathematical Research.

In addition to being of theoretical interest, Benford's law has applications in detecting fraud, and arises in numerous disciplines; see for example [3, 16, 11].

Many mathematicians have tried to explain the prevalence of Benford's law in the real world. Some have shown that data sets coming from certain probability distributions tend to satisfy Benford's law, and explained the phenomenon mathematically [4, 14]. We adopt the same methodology as used in the aforementioned work to demonstrate and explain the phenomenon that data sets coming from a generalized gamma distribution usually satisfy Benford's law. As the exponential distribution and the Weibull distribution explored in [4, 14] are special cases of the generalized gamma distribution, our results incorporate the results obtained in the earlier research as special cases.

1.1 Benford's law

To begin with, we define Benfordness and restate Benford's law in a self-contained manner.

Definition 1.2 *A set of numbers is said to satisfy Benford's law if the leading digit* $D \in \{1, 2, \dots, , B-1\}$ *occurs with frequency* $F_D = \log_B \frac{D+1}{D}$*, where* $B \geq 2$*.*

A random variable X *or its distribution is said to demonstrate Benfordness if data sets of its realizations satisfy Benford's law.*

Remark 1.3 *Any positive* x *can be written in scientific notation as* $S_B(x) \cdot 10^k$ *with the significand* $S_B(x) \in [1, B)$ *and* k *an integer. A more general version of the law describes the frequencies of different second digits, third digits and so on. One could also give an expression for the probability of a digit occurring in the* n*-th digit of a number, as is described in the first chapter of* [11]*. This is related to the Strong Benford's law, which states that the probability of observing a significand of at most* s *in base* B *is equal to* $\log_B s$*. In an abuse of notation, the distribution of just the leading digit, as well as the distribution of the entire significand, is often referred to as just Benford's law.*

Remark 1.4 *For any finite data set, it is impossible for the leading digit frequencies of a data set to equal the theoretical quantities given by the law exactly (as these probabilities are irrational). As long as they are close, and are expected to converge as the size of the sample tends to infinity, we will still say the data set satisfies Benford's law, and the distribution used to model the data set demonstrates Benfordness.*

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

Example 1.5 *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, since* $\log_{10} 31295192 \approx 7.495477620349604$*, the mantissa is about* 0.495477620349604*.*

We now formally introduce what it means for a random variable to have a Benford distribution.

Definition 1.6 *Given $B \geq 2$, a random variable X has a Benford distribution with base B if* $\operatorname{Prob}(X \leq x) = \log_B x$*, where $x \in [1, B)$.*

If a random variable X has a Benford distribution, then clearly it will demonstrate Benfordness. But random variables with other distributions could also demonstrate Benfordness. Let $X : \Omega \to \mathbb{R}$, be a random variable with cumulative distribution function (cdf) F. If $S_B \circ X$ has a Benford distribution with base B, then one would expect a data set of realizations of X to satisfy Benford's law since

$$\operatorname{Prob}(X \text{ has leading digit } D) \;=\; \operatorname{Prob}(S_B \circ X \in [D, D+1)) \;=\; \log_B \frac{D+1}{D},$$

which is just a direct application of Definition 1.6.

To determine whether a random variable X demonstrates Benfordness or not, we could find the distribution of $S_B \circ X$ directly and compare it to the Benford distribution. Alternatively, the following theorem, which can be found in [5], provides another method, which in some cases is more convenient to use.

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

Proof. For any $s \in [1, B)$, let $u = \log_B s \in [0, 1)$. Let us assume that $S_B \circ X$ has a Benford distribution with base B, then the following holds:

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

Hence it follows that $\log_B X \bmod 1$ has a uniform distribution on $[0, 1)$.
For the other direction, let us now assume that $\log_B X \bmod 1$ has a uniform distribution on $[0, 1)$. Then

$$\begin{aligned}
\operatorname{Prob}(S_B \circ X \in [1, s)) &= \operatorname{Prob}\big(\{X \in [1 \cdot B^k, s \cdot B^k) : k \in \mathbb{Z}\}\big) \\
&= \operatorname{Prob}(\log_B X \bmod 1 \in [0, u)) \\
&= u \\
&= \log_B s,
\end{aligned}$$

which means $S_B \circ X$ has a Benford distribution with base B. □

Theorem 1.7 forms the key foundation of our work. Tto show a positive random variable X demonstrates Benfordness, instead of dealing with $S_B \circ X$, we focus on $\log_B X \bmod 1$ and prove it has (or is close to) a uniform distribution on $[0, 1)$.

Theorem 1.7 also enables us to measure the deviation of a random variable from Benfordness. Assume X and Y are positive random variables and $S_B \circ Y$ has a Benford distribution with base B. 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: $f_{\log_B X \bmod 1}$. For any digit $D \in \{1, 2, \ldots, B-1\}$, the deviation of X from Benfordness is defined and given by

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

Notice that the expression in the last line above no longer depends on D, so it could be used as a general bound for X's deviation from Benfordness. Equation 1 will be used later in the numerical simulation of the probability density function of $\log_B X \bmod 1$. In the next subsection, we present a quick theoretical overview of the generalized gamma distribution.

1.2 The Generalized Gamma Distribution and Its Benfordness

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. Their works demonstrated and explained the phenomenon that data sets coming from these distributions usually are close to Benford's law. Both of these distributions can be seen as "children" of one parent distribution for a particular choice of parameters, which is the generalized gamma distribution.

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 has a generalized gamma distribution with parameters a, d, and p if its cumulative distribution function is of the form*

$$F(x; a, d, p) = \frac{\gamma\left(\frac{d}{p}, \left(\frac{x}{a}\right)^p\right)}{\Gamma\left(\frac{d}{p}\right)}, \quad x > 0;\ a, d, p > 0, \qquad (2)$$

where γ is the lower incomplete gamma function, defined as

$$\gamma(s, x) := \int_0^x t^{s-1} e^{-t} \mathrm{d}t.$$

The corresponding probability density function is

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

One can gain a lot of intuition for the behavior of the distribution by analyzing the graph of the probability density function for different parameters. In particular, we point out that the parameters d and p determine the shape of the probability density function, while the parameter a determines the spread of the probability density function.

We note that, when $d = p$, equation (2) is just the cumulative distribution function 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 relate and compare our results to the results obtained in [4, 14].

Our main results are the following, which we prove in Section 2.

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

$$f_{\log_B X \bmod 1}(u) \;=\; \frac{p\ln B}{\Gamma\left(\frac{d}{p}\right)}\sum_{k=-\infty}^{\infty} e^{-\left(\frac{B^{k+u}}{a}\right)^p}\left(\frac{B^{k+u}}{a}\right)^d, \tag{3}$$

or

$$f_{\log_B X \bmod 1}(u) \;=\; 1+\sum_{k=1}^{\infty}\frac{2}{\Gamma\left(\frac{d}{p}\right)}\operatorname{Re}\left[e^{-2\pi ik\left(u-\frac{\ln a}{\ln B}\right)}\Gamma\left(\frac{d}{p}+\frac{2\pi ki}{p\ln B}\right)\right], \tag{4}$$

where $u \in (0,1)$. Further, the behavior of the probability density function is mainly determined by parameters d and p, for any $m \in \mathbb{Z}$, a and $a \cdot B^m$ result in the same probability density function.

Theorem 1.10 *Let*

$$f^M_{\log_B X \bmod 1}(u) \;=\; 1+\sum_{k=1}^{M}\frac{2}{\Gamma\left(\frac{d}{p}\right)}\operatorname{Re}\left[e^{-2\pi ik\left(u-\frac{\ln a}{\ln B}\right)}\Gamma\left(\frac{d}{p}+\frac{2\pi ki}{p\ln B}\right)\right]$$

be the first M-term (plus the main term 1) partial sum approximation of $f_{\log_B X \bmod 1}(u)$ in equation (4). For any $\epsilon > 0$, if

$$M \;>\; \frac{(d+p)^2(\ln(B))^2}{2\pi^2\epsilon}-1, \tag{5}$$

then $\sup_{u\in(0,1)}|f_{\log_B X \bmod 1}(u)-f^M_{\log_B X \bmod 1}(u)| \le \epsilon$, i.e., we could make the approximation error arbitrarily small by keeping more terms in the partial sum.

In our proof of Theorem 1.9 in the next section, we adopt the same method used by Cuff et.al in [4]. To get equation (4) from equation (3), we need the following result.

Lemma 1.11 (**Poisson summation formula**[1]) *Let f, f' and f'' be continuous functions which eventually decay at least as fast as $x^{-(1+\eta)}$ for some $\eta > 0$, then*

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

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

Before giving the proofs, we present some simulations that show the Benfordness of the generalized gamma distribution.

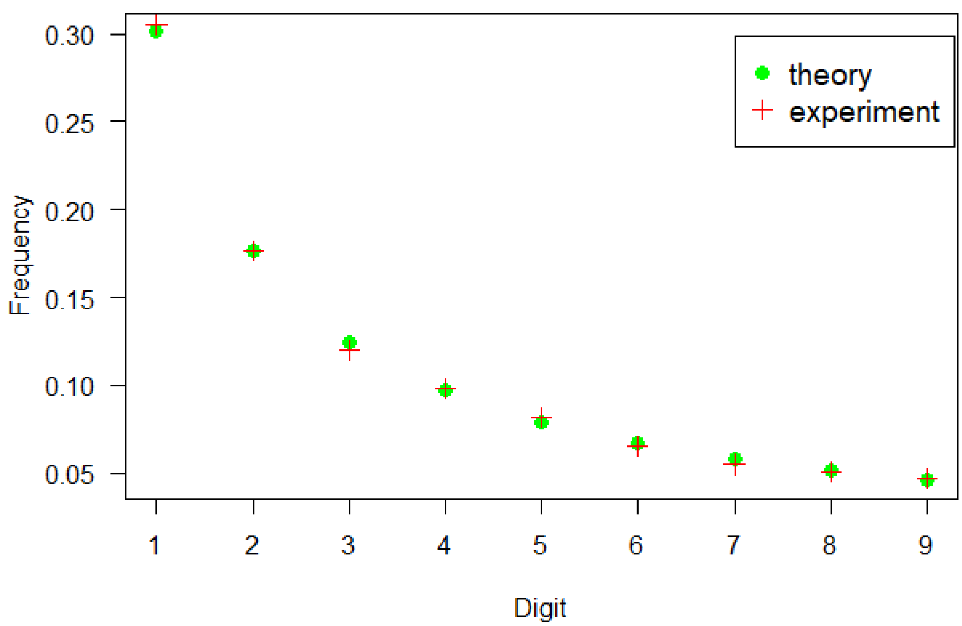

Figure 1: Leading digit frequencies of 10000 samples from a generalized gamma distribution $(B = 10, a = 2, d = 1, p = 1/2)$.

Figure 1 compares the leading digit frequencies of $10,000$ samples from a generalized gamma distribution with parameters $B = 10, a = 2, d = 1$, and $p = 1/2$, with the frequencies predicted by Benford's law. Observe that Benford's law is an excellent fit.

[1]The conditions here are not the weakest needed for the result, but suffice for our purposes.

The Benfordness of the generalized gamma distriubtion can also be demonstrated in another way. Theorem 1.7 guarantees that a positive random variable X demonstrates Benfordness if and only if $\log_B X \bmod 1$ is approximately uniformly distributed over $[0, 1)$. So we could test the Benfordness of a generalized gamma distributed random variable X by inspecting if the distribution of $\log_B X \bmod 1$ is close to a uniform distribution over $[0, 1)$. This can be done through a Kolmogorov-Smirnov test, which could be used to examine whether a given sample comes from a specific distribution. The smaller the test statistic is, the more likely the sample came from the target distribution. We generated data sets from the generalized gamma distribution with different values of parameters d and p, and performed the Kolmogorov-Smirnov tests to compare the transformed data ($\log_B X \bmod 1$) with the uniform distribution on $[0, 1)$. The result is shown in Figure 2. Observe that the test statistics are small in general, indicating the transformed data came from populations with an approximate uniform distribution over $[0, 1)$, and hence the original distributions should demonstrate Benfordness.

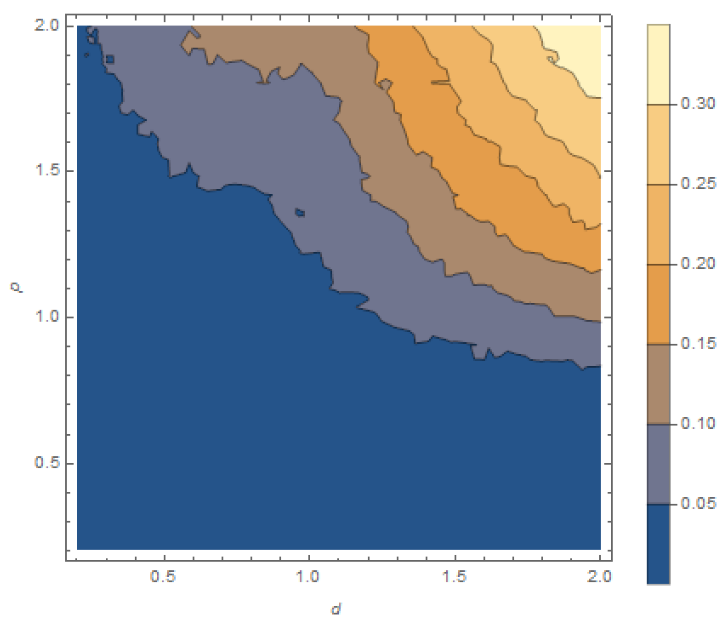

Figure 2: Kolmogorov-Smirnov test results under different values of d and p ($a = 1$).

Given the huge amount of possible combinations of the three parameters of a generalized gamma distribution, it's unrealistic to expect Benfordness to be demonstrated under any given set of parameters. From our experimental experience, the generalized gamma distribution demonstrates very good Benfordness when parameters d and p are small. This could be see in Figure 2.

2 Main Results and Key Observations

Below we prove Theorems 1.9 and 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

$$\begin{aligned} \operatorname{Prob}(\log_B X \bmod 1 \in [0,u]) &= \sum_{k=-\infty}^{\infty} \operatorname{Prob}(\log_B X \in [k, k+u]) \\ &= \sum_{k=-\infty}^{\infty} \operatorname{Prob}(X \in [B^k, B^{k+u}]) \\ &= \frac{1}{\Gamma\left(\frac{d}{p}\right)} \sum_{k=-\infty}^{\infty} \int_{\left(\frac{B^k}{a}\right)^p}^{\left(\frac{B^{k+u}}{a}\right)^p} t^{\frac{d}{p}-1} e^{-t} \mathrm{d}t, \end{aligned} \tag{6}$$

which is the culmulative distribution function of $\log_B X \bmod 1$. To get the probability density function of $\log_B X \bmod 1$, we differentiate the above with respect to u, obtaining

$$\begin{aligned} f_{\log_B X \bmod 1}(u) &= \frac{\mathrm{d}}{\mathrm{d}u} \operatorname{Prob}(\log_B X \bmod 1 \in [0,u]) \\ &= \frac{1}{\Gamma\left(\frac{d}{p}\right)} \frac{\mathrm{d}}{\mathrm{d}u} \sum_{k=-\infty}^{\infty} \int_{\left(\frac{B^k}{a}\right)^p}^{\left(\frac{B^{k+u}}{a}\right)^p} t^{\frac{d}{p}-1} e^{-t} \mathrm{d}t \\ &= \frac{1}{\Gamma\left(\frac{d}{p}\right)} \sum_{k=-\infty}^{\infty} \frac{\mathrm{d}}{\mathrm{d}u} \int_{\left(\frac{B^k}{a}\right)^p}^{\left(\frac{B^{k+u}}{a}\right)^p} t^{\frac{d}{p}-1} e^{-t} \mathrm{d}t, \end{aligned} \tag{7}$$

where $u \in (0,1)$. The justification for doing term-by-term differentiation in the preceding deduction is the presence of e^{-t} term in the integral, which guarantees the uniform convergence of the infinite sum and its derivative over $(0,1)$.

We now continue to work with the integral from (7). Using the Fundamental Theorem of Calculus we find

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}u} \int_{\left(\frac{B^k}{a}\right)^p}^{\left(\frac{B^{k+u}}{a}\right)^p} t^{\frac{d}{p}-1} e^{-t} \mathrm{d}t &= e^{-\left(\frac{B^{k+u}}{a}\right)^p} \left(\left(\frac{B^{k+u}}{a}\right)^p\right)^{\frac{d}{p}-1} p \left(\frac{B^{k+u}}{a}\right)^{p-1} \frac{B^{k+u} \ln B}{a} \\ &= e^{-\left(\frac{B^{k+u}}{a}\right)^p} \left(\frac{B^{k+u}}{a}\right)^d p \ln B, \end{aligned} \tag{8}$$

and then substituting (8) into (7), we derive that the probability density function of $\log_B X \bmod 1$ is

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

Next we apply the Poisson summation formula in Lemma 1.11 to the above result to get the equivalent form (4), which is better since it is divided into a main term, 1, which is what we want as it's the probability density function of a uniform distribution on $[0,1)$, and a residue term given by an infinite series.

For any $u \in (0,1)$, let $z = B^u$, $t = k$. It's easy to check that

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

satisfies the conditions for applying Poisson summation formula. To apply the formula, we first compute the Fourier transform of this function:

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

Now we use the above result to apply Poisson summation to equation (3), which yields equation (4).

$$\begin{aligned}
f_{\log_B X \bmod 1}(u) &= \frac{1}{\Gamma\left(\frac{d}{p}\right)}\sum_{k=-\infty}^{\infty} g(k) \\
&= \frac{1}{\Gamma\left(\frac{d}{p}\right)}\sum_{k=-\infty}^{\infty} \hat{g}(k) \\
&= \frac{1}{\Gamma\left(\frac{d}{p}\right)}\sum_{k=-\infty}^{\infty} \left(\frac{z}{a}\right)^{\frac{2\pi i k}{\ln B}}\Gamma\left(\frac{d}{p}-\frac{2\pi i k}{p\ln B}\right) \\
&= 1+\frac{1}{\Gamma\left(\frac{d}{p}\right)}\sum_{k=1}^{\infty}\left[\left(\frac{z}{a}\right)^{\frac{2\pi i k}{\ln B}}\Gamma\left(\frac{d}{p}-\frac{2\pi i k}{p\ln B}\right)+\right. \\
&\qquad \left.\left(\frac{z}{a}\right)^{-\frac{2\pi i k}{\ln B}}\Gamma\left(\frac{d}{p}+\frac{2\pi i k}{p\ln B}\right)\right] \\
&= 1+\sum_{k=1}^{\infty}\frac{2}{\Gamma\left(\frac{d}{p}\right)}\mathrm{Re}\left[\left(\frac{z}{a}\right)^{-\frac{2\pi i k}{\ln B}}\Gamma\left(\frac{d}{p}+\frac{2\pi k i}{p\ln B}\right)\right] \\
&= 1+\sum_{k=1}^{\infty}\frac{2}{\Gamma\left(\frac{d}{p}\right)}\mathrm{Re}\left[e^{-2\pi i k\left(u-\frac{\ln a}{\ln B}\right)}\Gamma\left(\frac{d}{p}+\frac{2\pi k i}{p\ln B}\right)\right], \qquad (9)
\end{aligned}$$

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

Finally, observe that $e^{2\pi ix} = e^{2\pi i(x+1)}$ for all $x \in \mathbb{R}$, which verifies that for any $m \in \mathbb{Z}$, a and $a \cdot B^m$ result in the same probability density function. $\square$

Remark 2.1 *When reducing the results from equation* (4) *to the Weibull case, we retrieve the same result that was shown in the article by Cuff et.al* [4]*.*

We now turn to estimating the value of the probability density function.

Proof. [Proof of Theorem 1.10] For any $M \geq 1$, $u \in (0,1)$, the approximation error is

$$
\begin{aligned}
|\eta| &= |f_{\log_B X \bmod 1}(u) - f^M_{\log_B X \bmod 1}(u)| \\
&= \left| \sum_{k=M+1}^{\infty} \frac{2}{\Gamma\left(\frac{d}{p}\right)} \operatorname{Re}\left[e^{-2\pi ik\left(u - \frac{\ln a}{\ln B}\right)} \Gamma\left(\frac{d}{p} + \frac{2\pi ki}{p \ln B}\right)\right]\right| \\
&\leq \sum_{k=M+1}^{\infty} \frac{2}{\Gamma\left(\frac{d}{p}\right)} \left| \Gamma\left(\frac{d}{p} + \frac{2\pi ki}{p\ln B}\right)\right|. \qquad (10)
\end{aligned}
$$

The gamma function has the property that[2]

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

and applying this to (10), we get

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

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

$$
\begin{aligned}
|\eta| &\leq \sum_{k=M+1}^{\infty} \frac{[(d+p)\ln B]^2}{2\pi^2 k^2} \\
&\leq \int_{M+1}^{\infty} \frac{[(d+p)\ln B]^2}{2\pi^2 x^2} \mathrm{d}x \\
&= \frac{[(d+p)\ln B]^2}{2\pi^2 (M+1)}. \qquad (11)
\end{aligned}
$$

[2] This could be found in many mathematical handbooks like [1].

Letting $|\eta| < \epsilon$, and using the result from (11), we obtain the lower bound for M as given in (5). $\square$

With the help Theorem 1.10, for a random variable having a generalized gamma distribution with parameters (a, d, p), we could further decompose and bound its deviation from Benfordness $\mathrm{Dev}(D)$ for any digit D, which is given in (1), as:

$$\begin{aligned}
\mathrm{Dev}(D) &= \left|P(X \text{ has leading digit } D) - \log_B \frac{D+1}{D}\right| \\
&\leq \int_0^1 |f(u) - 1| \mathrm{d}u \\
&\leq \int_0^1 |f(u) - f^M(u)| \mathrm{d}u + \int_0^1 |f^M(u) - 1| \mathrm{d}u \\
&\leq \epsilon + \sup_{u \in (0,1)} |f^M(u) - 1|, \qquad (12)
\end{aligned}$$

where f and f^M are exact and approximate probability density functions 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 deviation 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 probability density functions (with approximation error $\epsilon < 0.01$) of $\log_B X \bmod 1$ with different parameters. We could see all curves only deviate slightly from the constant 1, which indicates these probability density functions are pretty close to the probability denisty function of a uniform distribution over $(0, 1)$. We could also see in Figure 3 the scaling invariance property of parameter a.

Figure 4 shows the upper bound of the deviation of a generalized gamma distribution from Benfordness with respect to d and p according to (12). We can see the deviation is small when the parameters d and p are small, but increases as the parameters increase. This is consistent with the result of Kolmogorov-Smirnov test in Figure 2.

3 Conclusion and Future Work

We have shown the Benfordness of the generalized gamma distribution under suitable choices of parameters. When d and p are small, data sets coming from a generalized gamma distribution tend to satisfy Benford's law. It would be interesting to see this result being used in practice. For example, when $p = 1$, the generalized gamma distribution is reduced to a gamma distribution with parameters $\alpha = d$, $\beta = 1/a$, which is wildly used to model real-world data sets, such as the size of insurance claims. For these applications, when the parameter d of the model is small, some data checking methods could be designed accordingly based on Benford's law.

A natural future research avenue is to perform a similar analysis for other families of distributions, in particular dealing with the challenges that arise if the random variables are discrete.

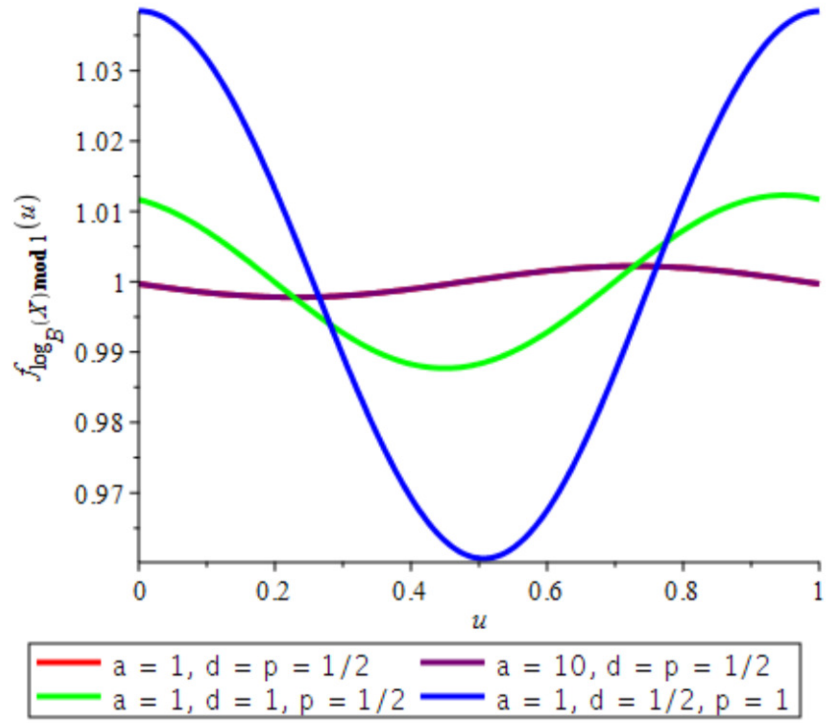

Figure 3: Approximation of probability density functions of $\log_B X$ mod 1 with different parameters.

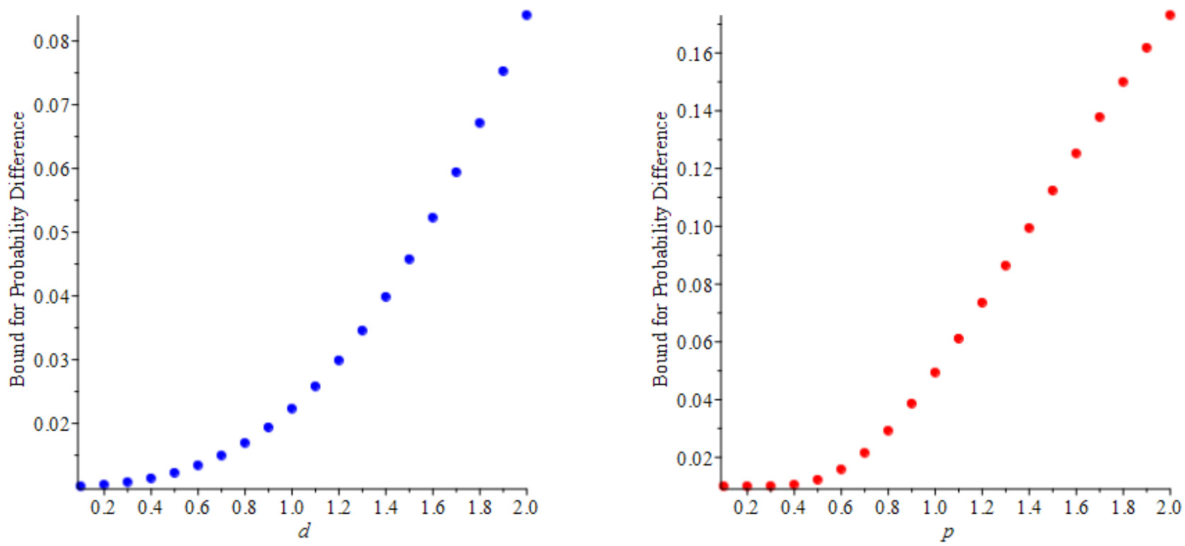

Figure 4: Left: Bound of probability difference (12) with respect to d $(a = 1,\ p = 0.5)$; Right: Bound of probability difference (12) with respect to p $(a = 1,\ d = 0.5)$.

Appendix A 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

```
#source("ggamma.R")
N=10000
a = 2
d = 1
p = 1/2
B <- 10 # B should be an integer greater than 1

#sample <- rggamma(N, a, d, p)

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_t, las=1, xlab="Digit", ylab="Frequency", col="green", xaxt="n",
    pch=19, cex=1.2)
axis(1, at=1:(B-1), labels=1:(B-1))
points(d, freqs, col="red", pch=3, cex=1.2)
legend(B - 3, max(c(freqs, freqs_t)) * 0.98, legend=c("theory", "experiment"),
    col=c("green", "red"), pch=c(19, 3), cex=1.2)
```

Maple code: Plot probability density functions 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 * k + k * 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", thickness = 3, labels=[u, f], legend="a =
    1, d = p = 1/2", labeldirections = ["horizontal", "vertical"]):

B:=10; a:=10; d:=1/2; p:=1/2;
p2 := plot(f_M, u=0..1, color="purple", thickness = 3, labels=[u, f], legend="a
     = 10, d = p = 1/2", labeldirections = ["horizontal", "vertical"]):

B:=10; a:=1; d:=1; p:=1/2;
p3 := plot(f_M, u=0..1, color="green", thickness = 3, labels=[u, f], legend="a
    = 1, d = 1, p = 1/2", labeldirections = ["horizontal", "vertical"]):

B:=10; a:=1; d:=1/2; p:=1;
p4 := plot(f_M, u=0..1, color="blue", thickness = 3, labels=[u, f], legend="a =
     1, d = 1/2, p = 1", labeldirections = ["horizontal", "vertical"]):

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"], labeldirections=["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"], labeldirections=["horizontal", "
    vertical"]);
```

Mathematica code for the Kolmogorov-Smirnov test

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

References

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

[2] F. Benford, The Law of Anomalous Numbers, *Proc. Amer. Philos. Soc.*, **78** (1938), 551–572. Also 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, S.J. Miller, The Weibull distribution and Benford's law, *Involve*, **8** (2015), 859–874. Also 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. Also 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, *Statist. Sci.*, **10** (1995), 354–363. Also available online at the URL: `https://projecteuclid.org/journals/statistical-science/volume-10/issue-4`.

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

[8] A. Kontorovich, S.J. Miller, Benford's law, values of L-functions and the $3x+1$ problem, *Acta Arith.*, **120** (2005), 269–297. Also available online at the URL: `https://arxiv.org/abs/math/0412003`.

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

[10] L.M. Leemis, B.W. Schmeiser, D.L. Evans, Survival Distributions Satisfying Benford's law, *The American Statistician*, **54** (2000), 236–241. Also 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, 2015.

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

[13] S.J. Miller, M. Nigrini, The Modulo 1 Central Limit Theorem and Benford's law for Products, *Int. J. Algebra*, **2** (2008), 119–130. Also available online at the URL: `https://arxiv.org/abs/math/0607686`.

[14] S.J. Miller, M.J. Nigrini, Order Statistics and Benford's law, *Int. J. Math. Math. Sci.*, (2008), 1–13. Also available online at the URL: `https://www.hindawi.com/journals/ijmms/2008/382948/`.

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

[16] M.J. Nigrini, S.J. Miller, Data diagnostics using second order tests of Benford's law, *Auditing: A Journal of Practice and Theory*, **28** (2009), 305–324.

[17] M.J. Nigrini, Benford's law: Applications for Forensic Accounting, Auditing, and Fraud Detection, *Auditing: A Journal of Practice and Theory*, **28** (2009), 305–324.

[18] Z. Rudnick, P. Sarnak, A. Zaharescu, The Distribution of Spacings Between the Fractional Parts of $n^2\alpha$, *Invent. Math.*, **145** (2001), 37–57. Also available online at the URL: `https://arxiv.org/abs/2004.04855`.

[19] R.A. Raimi, The First Digit Problem, *Amer. Math. Monthly*, **83** (1976), 521–538. Also 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, *Ann. Math. Statist.*, **33** (1962), 1187–1192. Also 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: `sjm1@williams.edu, Steven.Miller.MC.96@aya.yale.edu`

Received: February 2, 2022 **Accepted**: May 10, 2022
Communicated by Elizabeth McClellan Ribble