BENFORD'S LAW, FAMILIES OF DISTRIBUTIONS AND A TEST BASIS

JOHN MORROW†

This Draft: October 9, 2010
First Draft: August 6, 2006

ABSTRACT. The distribution of first significant digits known as Benford's Law has been used to test for erroneous and fraudulent data. By testing for conformance with the Law, applied researchers have pinpointed anomalous data using a standard hypothesis testing approach. While novel, there are two weaknesses in this methodology. First, test values used in practice are too conservative once Benford specific values are derived. The new test values of this paper are more powerful and I investigate their small sample properties. Second, testing requires the Null hypothesis of Benford's Law to hold, which often does not for real data. I therefore present a simple method by which all continuous distributions may be transformed to satisfy Benford with arbitrary precision and induce *scale invariance,* one of the properties underlying Benford's Law in the literature. This allows application of Benford tests to arbitrary samples, a hurdle to empirical work. I additionally derive a rate of convergence to Benford's Law. Finally, the theoretical results are applied to commonly used distributions to exhibit when the Law holds within distributional families. The results yield improved tests for Benford's Law applicable to a broader class of data and contribute to understanding occurrences of the Law.

KEY WORDS: Benfords Law; data quality; fraud detection
JEL Codes: C10, C24, C46
AMS Classification: 62E20, 62F12

Acknowledgement. I thank William Brock, George Judge, Thomas Kurtz and Laura Schechter for helpful comments, guidance and encouragement. This paper has benefited from discussion with Yasushi Asako, Swati Dhingra, Ching-Yang Lin and Mian Zhu.

† UW-Madison, Contact: morrow1@wisc.edu. A utility for testing Benford's Law is available from www.checkyourdata.com. Comments appreciated.

1. Introduction

Benford's Law states that for commonly observed empirical data, regularities should occur in the First Significant Digits (FSDs) of the data. The FSD of a number x is the leading digit of x in the base 10 numbering, for instance

$$\text{FSD of } \pi = 3 \text{ since } \pi = \underbrace{3}_{FSD}.14159\ldots$$

In its strong form, Benford's law says for the FSDs $\{1, \ldots, 9\}$, the frequency observed of each digit $d \in \{1, \ldots, 9\}$ should be approximately $\log_{10}(1 + 1/d)$. Many papers have detailed occurrences of Benford's Law (see Benford (1938); Berger and Hill (2007); Giles (2007)). A few papers have also categorized properties characterizing distributions satisfying Benford's Law (see Boyle (1994); Hill (1995b); Allaart (1997)), or found distribution families which satisfy it for particular parameter values (see Leemis et al. (2000); Scott and Fasli (2001)). Unfortunately, no *general principle* has been found to explain the Benford phenomenon in data, nor provide general criteria as to when to expect Benford's Law to hold.

Benford's Law has also been used to test for fraud and error present in a variety of contexts. Examples using Benford's law for fraud and error detection include tax fraud Nigrini (1996), reliability of survey data Judge and Schechter (2009), environmental law compliance Marchi and Hamilton (2006) and campaign finance Cho and Gaines (2007). This paper first focuses on the testing issues that arise when assessing conformance with Benford's Law, then contributes towards general characterizations of the Law, in particular providing a rate of convergence to the law under appropriate transformation.

Testing for Benford's Law has recently been performed on a variety of data sets, in the broad context of detecting fraud. This paper focuses on two testing issues. The first is the suitability of existing tests which have been used in the literature. Such tests are too conservative and consequently Section 2 derives new asymptotically valid test values which allow for more powerful tests and evaluates small sample values of the tests. Measures of fit have also been used as "rules of thumb" to check concordance with Benford's Law. Section 2 also provides a new interpretation for such measures and derives critical values for hypothesis testing. The second testing issue is the application of tests on data which inherently do not satisfy the law (for a discussion, see Durtschi et al. 2004). Clearly, rejection of tests for Benford on data which inherently fails the law will not help uncover fraud or error. Section 3 develops a result that the transformation of a random variable to a sufficiently high power satisfies Benford within arbitrary precision, allowing application of the above tests to any sample. Section 4 answers how quickly a random variable converges to Benford, provides a discussion of the main results, applies them to distribution families of interest and concludes.

2. Testing and Benford's Law

One of the most popular applications of Benford's Law is fraud detection and testing of data quality. A few tests have been constructed, and new tests recently proposed, but at present it appears that properties of the estimators themselves are not well understood. In fact, asymptotic results indicate that the test values used in some recently published papers can be made more powerful at the significance levels used (for example Cho and Gaines 2007; Giles 2007). In addition, such tests appear rather *ad hoc* and the power of such tests appears to be almost wholly unexamined. I now discuss tests in use, provide asymptotically valid test values, and explore their small sample properties finding that tests I provide are very good for $N \geq 80$.

2.1. **Popular Tests in Use.** Pearson's χ^2 test is a natural candidate for testing whether an observed sample satisfies Benford's Law, however, due to its low power for even moderately small sample sizes it is often unsuitable. Consequently, other tests have been devised, and commonly used tests for conformance with Benford's Law include the Kolmogorov-Smirnov test and the Kuiper test. More recently Leemis et al. (2000) have introduced the statistic m (max)

$$m \equiv \max_{d \in \{1,\ldots,9\}} \left|\Pr\left(X \text{ has FSD} = d\right) - \log_{10}\left(1 + 1/d\right)\right|$$

Similarly, Cho and Gaines (2007) propose the d (distance) statistic.

$$d \equiv \left[\sum_{d \in \{1,\ldots,9\}} \left[\Pr\left(X \text{ has FSD} = d\right) - \log_{10}\left(1 + 1/d\right)\right]^2\right]^{1/2}$$

In both cases the sample analogue of $\Pr\left(X \text{ has FSD} = d\right)$ is used for evaluation, although no test values are known for these statistics.

2.2. **Issues with current tests in use: Kolmogorov-Smirnov and Kuiper.** The χ^2, Kolmogorov-Smirnov (D_N) and Kuiper (V_N) tests for a sample of size N appear to be the most common tests in use. In fact, latter two have a "correction factor" introduced by Stephens (1970) which when applied to such tests produce fairly accurate test statistics regardless of sample size. Denote these tests with the correction factor applied as D_N^* and V_N^*, respectively. For instance, for the modified Kuiper test V_N^* presented in Stephens, a 99% confidence set is produced by all samples $\{X_i\}$ such that $V_N^* < 2.001$. However, such tests are based on the null hypothesis of a continuous distribution, and are generally conservative for testing discrete distributions as discussed by Noether (1963). A simple example where the sample is drawn from a Bernoulli distribution with $p = 1/2$ (fair coin tosses) in the supplemental appendix shows that a V_N^* test at 99% significance generates a .99994% critical region. Thus test values derived for continuous distributions can be *extremely* conservative in rejecting the null.

The Stephens (1970) test values for the modified Kuiper (D_N^*) and Kolmogorov-Smirnov (V_N^*) tests at commonly used significance levels are reported in the first column of Table 1. New asymptotically valid test values under the specific null hypothesis that Benford's Law holds are in the second column of Table 1. These test values are derived from an application of the CLT to a multivariate Bernoulli variable that corresponds to a random variable which exactly satisfies Benford's Law. Inspection shows that in fact the test values based on the assumption of a continuous underlying distribution are too high, and thus too conservative. One appropriate test is that of Conover (1972), but is sufficiently involved and computationally expensive that practitioners have adopted the above tests. Furthermore, the test statistics as in Table 1 allow easy computation of the relevant test as well as allowing evaluatation of published literature.

Table 1. Continuous vs Benford Specific Test Values

	Continuous			Benford Specific		
Test Statistic	$\alpha = .10$	$\alpha = .05$	$\alpha = .01$	$\alpha = .10$	$\alpha = .05$	$\alpha = .01$
Kuiper Test (V_N^*)	1.620	1.747	2.001	1.191	1.321	1.579
KS Test (D_N^*)	1.224	1.358	1.628	1.012	1.148	1.420

Pulling in an example from the Benford literature, Giles (2007) looks for deviations from Benford's Law in certain eBay auctions to detect for collusion by buyers or interference by sellers. Giles uses the Kuiper Test for continuous distributions ($N = 1161$) as in Table 1 with a test value of 1.5919 and cannot reject conformance to Benford at any level. However, we see that the Benford specific tests reject conformance to Benford at $\alpha = .01$. Marchi and Hamilton (2006) examine discrepancies in air pollution reporting by testing for conformance to Benford using the Kolmogorov-Smirnov test. In this case, the authors explicitly point out potential problems with their test values, and the results would have changed if they had used an $\alpha = .01$ test level.

2.3. **The m and d tests and critical values.** As far as the m and d tests are concerned, no test values have been reported for use which address the above issues. In order to derive asymptotic test statistics, define the modified test statistics m_N^* and d_N^* given in Equations (2.1-2.2), where N is the number of observations.

$$m_N^* \equiv \sqrt{N} \cdot \max_{d \in \{1,\ldots,9\}} \left| \Pr\left(X \text{ has FSD} = d\right) - \log_{10}\left(1 + 1/d\right) \right| \tag{2.1}$$

$$d_N^* \equiv \sqrt{N} \cdot \left[\sum_{d \in \{1,\ldots,9\}} \left[\Pr\left(X \text{ has FSD} = d\right) - \log_{10}\left(1 + 1/d\right)\right]^2 \right]^{1/2} \tag{2.2}$$

The reason for the appearance of the $\sqrt{N}$ term is as follows. The true FSD frequencies $\Pr\left(X \text{ has FSD} = d\right)$ correspond to Bernoulli parameters as do the Benford $\log_{10}\left(1 + 1/d\right)$ terms. Letting $\mathbf{1}_{FSD=d}(X)$ be the indicator that X has a FSD equal to d, the random vector

$$\mathbf{T}_N \equiv \left[\ \overline{\mathbf{1}_{FSD=1}(X)} - \log_{10}\left(1 + 1/1\right) \quad \ldots \quad \overline{\mathbf{1}_{FSD=8}(X)} - \log_{10}\left(1 + 1/8\right) \ \right]$$

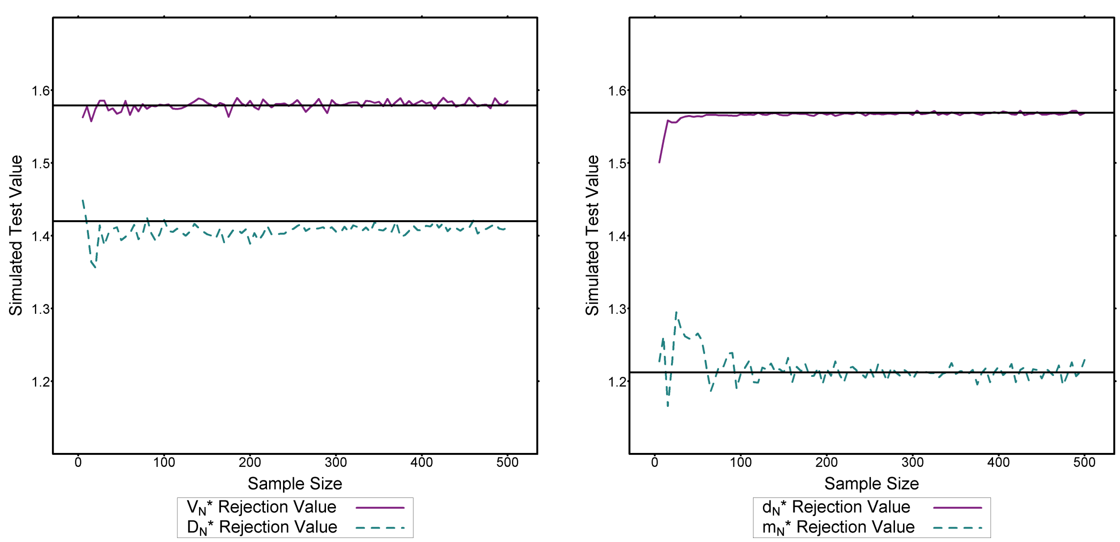
is iid and by the CLT, $\sqrt{N}\mathbf{T}_N$ converges in distribution to a $N(0, \mathbf{\Sigma})$ random variable. Both m_N^* and d_N^* can be formed as continuous mappings of $\sqrt{N}\mathbf{T}_N$ in which the $\sqrt{N}$ term can be slipped outside since the functions max and $\left(\sum x_i^2\right)^{1/2}$ are homogeneous. The end result is both m_N^* and d_N^* converge in distribution to a continuous function of a $N(0, \mathbf{\Sigma})$ variable, where $\mathbf{\Sigma}$ can be computed from $\mathbf{T}_N$. Rejecting the null that Benford's Law holds when m_N^* and d_N^* are large provides a consistent test statistic (e.g. Lemma 14.15 of van der Vaart 2000). Rejection regions for common test levels are provided in Table 2. The new d^* test values confirm the conclusions of Cho and Gaines (2007) who test political contribution data, broadly finding that the data does not fit Benford's Law.

TABLE 2. m^* and d^* Test Values

Test Statistic	Asymptotic Test Level $\alpha = .10$	$\alpha = .05$	$\alpha = .01$
Max Test (m_N^*)	0.851	0.967	1.212
Distance Test (d_N^*)	1.212	1.330	1.569

2.4. **Test Performance for Small Samples.** Naturally, the question arises of how good the critical values reported in Tables 1 and 2 are in practice for small sample sizes. For sample sizes $N \leq 500$ I have numerically computed the appropriate test values for a level $\alpha = .01$ test for all four statistics as shown in Figure 1, based on 10^6 draws for each sample size. The Figure contains numerically obtained test values in sample size increments of $N = 5$, and horizontally superimposed are the asymptotic test values for each test. The small N performance is fairly good in that the simulated test statistics are very close to the asymptotic values, especially for $N \geq 80$. This shows that the critical regions in Table 2 are reasonable for small as well as large N.

FIGURE 1. m_N^* and d_N^* Test Values for Small Samples



(A) Kuiper and KS Tests

(B) Max and Distance Tests

In conclusion, this section has given more powerful tests for the Kolmogorov-Smirnov and Kuiper statistics as well as valid test statistics for the m and d statistics used in the Benford literature. However, when these tests are used for error or fraud detection, they are based on the Null Hypothesis that in the absence of fraud or error, Benford's Law is satisfied. We address the ramifications of this Hypothesis in the next section.

3. Ensuring Conformity to Benford's Law

The general approach of using Benford's Law for fraud detection is to compare FSD frequencies in sample data with the Law, as for the tests discussed above. Of course, whether Benford's Law holds for a particular sample depends upon the underlying distribution. Therefore testing for Benford is restricted by the underlying properties of data. One of the major obstacles in using this approach is that often the distribution one would like to test does not remotely satisfy Benford's Law, regardless of data quality (see Table 3). The results in this section ameliorate this issue by developing a transformation (Theorem 1) that may be applied to data that induce compliance with Benford's Law. The implications of Theorem 1 are further developed in the next Section.

Before applying tests based on Benford's Law to a random variable X, one should first expect that X approximately satisfies Benford. This idea is formalized in the following Definition.

Definition. A random variable X $\epsilon-$satisfies Benford's Law if for all FSDs d

$$|\Pr(X \text{ has FSD} = d) - \log_{10}(1 + 1/d)| < \epsilon$$

Before applying the tests in Section 2 it is necessary to ensure that the sample $\epsilon-$satisfies Benford's Law. This is best illustrated with an example. Consider a sample S composed of two sub-samples, S_H and S_C and hypothesize S_H comes from an "Honest" data source while S_C comes from "Cheaters." The underlying assumption for fraud detection is that S_H is closer to satisfying Benford than S_C. But to apply the tests of Section 2, a minimum requirement is that S_H is approximately Benford, one option being that X $\epsilon-$satisfies Benford's Law. If the sample S could be transformed to satisfy the Law so that S_H satisfies the Law while S_C fails, the transformation would be a basis for detecting anomalies in S_C. The main result shown in this Section, Theorem 1, provides such a means of transforming S.

Theorem 1 (Exponential-Scale Families)**.** *Let X be a random variable with continuous pdf and fix $\epsilon > 0$. There is an α^* such that for all $\alpha \geq \alpha^*$:*

$$(X/\sigma)^{\alpha} \quad \epsilon - \textit{satisfies Benford's Law for all } \sigma$$

In light of the above discussion if one is fairly confident about the distribution of X (say, using a Kernel Density Estimate), one strategy is to apply Theorem 1 to transform X to $\epsilon-$satisfy Benford's Law and then perform tests. Methods for computing sufficiently large α follow from the intermediate results in this Section. To be concrete, suppose we have a random sample $\{X_i\}$ and we feel confident that $(X - \mu)/\sigma \sim N(0, 1)$, perhaps by estimating μ and σ from the sample. There are several values of μ and σ where we should not expect that the sample will obey Benford's Law. However, fix any $\epsilon > 0$ and from Theorem 1 we know there is an $\alpha(\epsilon)$ such that for $Y \sim (X - \mu)^{\alpha(\epsilon)}/\sigma^{\alpha(\epsilon)}$, the FSD frequencies observed in Y should be within ϵ of Benford's Law. A sufficiently large $\alpha(\epsilon)$ may be calculated from the distribution of X using the techniques below. Accordingly, m_N^* and d_N^* calculated with Y in place of X should be close to zero, allowing for detection of anomalous observations. This Section proceeds with intermediate steps leading up to a proof of Theorem 1.

3.1. ***Approximation by step functions.*** The following definition has an important relationship with Benford's Law, as will be shown shortly.

Definition. Let Y be a random variable with pdf f. Fix $\epsilon > 0$ then Y can be $\epsilon-$*approximated by integer step functions*, denoted $Y \in I(\epsilon)$ if there exist $\{c_i\}$ s.t.

$$\left| \int_A f(y)dy - \int_A \sum c_i \mathbf{1}_{[i,i+1)}(y)dy \right| \leq \epsilon \quad \text{for all measurable } A$$

For example, by taking $c_i \equiv 0$ for any random variable Y, $Y \in I(1)$. Although the definition of $I(\epsilon)$ is simple, any continuous random variable X for which $\log_{10} X \in I(\epsilon)$ "approximately" satisfies Benford's Law. The formal statement of this fact is Lemma 1.

Lemma 1. *Suppose X is a positive random variable with continuous pdf and let $Y \sim \log_{10} X$. If $Y \in I(\epsilon)$ then X $\epsilon-$satisfies Benford's Law.*

Proof. See Appendix. □

This lemma provides a check of whether a random variable X ϵ-satisfies Benford's law by checking whether $\log_{10} X \in I(\epsilon)$. Since Lemma 1 will be the workhorse throughout the rest of the paper, some remarks on its hypotheses are in order. First, the assumption of a continuous pdf is fairly mild and examination of the proofs shows it can be weakened, but this assumption will be maintained for brevity. Second, the restriction to positive random variables is really not an imposition since the First Significant Digits of X are identical to those of $|X|$.

3.2. **Characterization of $I(\epsilon)$.** The simplicity of the definition of $I(\epsilon)$ allows for a precise characterization of the least ϵ s.t. $X \in I(\epsilon)$. By definition, $X \in I(\epsilon)$ requires that

$$\sup_{A \text{ measurable}} \left| \int_A f(y)dy - \int_A \sum c_i \mathbf{1}_{[i,i+1)}(y)dy \right| \leq \epsilon \tag{3.1}$$

In solving for the best choice of $\{c_i\}$ it suffices to consider each interval $[i, i+1]$ individually. Surprisingly, the solution to these individual problems is quite simple in that the optimal c_i turn out to be the gross estimates $c_i \equiv \int_{[i,i+1]} f(x)dx$. These c_i are optimal because of the "maxi-min" nature of Equation (3.1): the best c_i must minimize integrals of the form $\left|\int_A [f(y) - c_i]_- dy\right|$ and $\left|\int_A [f(y) - c_i]_+ dy\right|$. Following this idea leads to a proof of Lemma 2.

Lemma 2. *Suppose* $\int |f(x)|\, dx < \infty$. *Then* $c^* \equiv \int_{[0,1]} f(y)dy$ *solves*

$$\min_c \sup_{A \text{ measurable}} \left| \int_{[0,1]\cap A} [f(x) - c]dx \right|$$

and the minimum attained is $\frac{1}{2}\int_{[0,1]} |f(x) - c^*|\, dx$.

Proof. See Appendix. □

A first consequence of Lemma 2 is that for random variables X_k with pdfs of the form $f(x) = k\mathbf{1}_{[0,\frac{1}{k}]}$, $X_k \in I(1 - \frac{1}{k})$ so considering large k, nothing can be said about $X \in I(\epsilon)$ for $\epsilon < 1$ without more information about the distribution of X. A second consequence of Lemma 2 is that choosing the optimal $\{c_i\}$ allows computation of the least ϵ such that $X \in I(\epsilon)$ directly. This characterizes the sets $I(\epsilon)$ completely, a consequence stated as Theorem 2.

Theorem 2. *Let* X *be a random variable with pdf* f. *The least* ϵ *s.t.* $X \in I(\epsilon)$ *is given by*

$$\epsilon = \frac{1}{2} \sum_i \int_{[i,i+1]} |f(x) - \int_{[i,i+1]} f(t)dt|dx \tag{3.2}$$

Proof. Application of Lemma 2 on each interval $[i, i+1]$. □

Paired with Lemma 1 this forms a method to test for conformance with Benford's Law within a parametric family using analytic methods: take any random variable

X with parameters θ, find the pdf of $\log_{10} X$, say g, and solve Equation (3.2) for g. Intuitively, for parameters θ where g is fairly "flat," $\int_{[i,i+1]} |g(x) - \int_{[i,i+1]} g(t)dt|dx$ is fairly small. Lemma 1 implies that X will $\epsilon-$satisfy Benford's Law for such θ, an implication expanded on in the next section. These results provide precise analytical tools to find parameters θ for X which will induce Benford's Law.

3.3. **Location-Scale Families and $I(\epsilon)$.** By virtue of the fact $Y \in I(\epsilon)$ means Y can be approximated by integer step functions, integer shifts and scaling of Y preserve the ability to approximate Y by integer step functions. In particular for integers a, b, let $Z \equiv aY + b$ and then Z can be approximated by translating the $\{c_i\}$ used to approximate Y. The new approximation will guarantee $Z = aY + b \in I(\epsilon)$. Since this holds for all integers a and b, $I(\epsilon)$ is invariant under such transformations as summarized in Lemma 3.

Lemma 3. *$Y \in I(\epsilon)$ iff $aY + b \in I(\epsilon)$ for all integers a, b with $a \neq 0$.*

Proof. See Supplemental Appendix. □

The last step towards proving Theorem 1 is a method of transforming any random variable within its mean-scale family so that the transformed variable is in $I(\epsilon)$ for arbitrary ϵ. This result is given in Theorem 3 and is followed by a sketch of the proof.

Theorem 3 (Mean-Scale Approximation)**.** *Let Y be a random variable with continuous pdf. For each $\epsilon > 0$ there exists a $\sigma(\epsilon)$ s.t. $\sigma \leq \sigma(\epsilon)$ implies $(Y - \mu)/\sigma \in I(\epsilon)$ for all μ.*

Proof. See Appendix. □

The basic idea of the proof is as follows. To show that $Y/\sigma \in I(\epsilon)$ consider σ as a transformation that flattens out the pdf of Y/σ as $\sigma \longrightarrow 0$. Once Y/σ is sufficiently flattened out, approximate its pdf via constants $\{c_i\}$ which correspond to appropriately chosen elements of a Riemann sum, giving an ϵ approximation to the pdf. In order to show $(Y-\mu)/\sigma = Y/\sigma - \mu/\sigma \in I(\epsilon)$ appeal to Lemma 3 to argue that without loss of generality $\mu/\sigma \in [0,1]$. Finally, show that smoothing Y further by dropping σ to $\sigma/2$ is enough that the improved approximation absorbs the μ/σ term.

3.4. **Proof of Theorem 1.** With the above results, it is a simple step to get to the main result of the section, Theorem 1. Let X be a positive random variable with continuous pdf. Fix ϵ and note

$$\log_{10}(X/\sigma)^{\alpha} = (\log_{10} X - \log_{10}\sigma)/(1/\alpha)$$

so from Theorem 3 for all sufficiently large α, $\log_{10}(X/\sigma)^{\alpha} \in I(\epsilon)$ for all $\sigma > 0$. The result then follows from an application of Lemma 1. If X is not positive a similar argument applies to $|X|$.

4. Discussion: Exponential-Scale Families

This section discusses additional implications of Theorem 1, restated here for ease of reference:

Theorem. *Let X be a random variable with continuous pdf and fix $\epsilon > 0$. There is an α^* such that for all $\alpha \geq \alpha^*$, $(X/\sigma)^{\alpha}$ $\epsilon-$ satisfies Benford's Law for all σ.*

Another way of stating this result is that the exponential transformation $g(x) = x^{\alpha}$ induces conformity to Benford's Law for all sufficiently large α. More surprising is that this transformation simultaneously induces approximate *scale invariance,* in that $(X/\sigma)^{\alpha}$ satisfies Benford's Law for any scaling parameter σ. Scale invariance is one of the fundamental properties that distributions satisfying Benford's Law should have (see Raimi 1976; Hill 1995a for formal definitions and results). Earlier work has detailed experimental evidence of high exponents of random variables to conform to Benford's Law independent of scale (For instance Scott and Fasli (2001) find the Log-Normal distribution satisfies the Law for $\sigma \gtrsim 1.2$).

Raising a random variable Y to the power α has the effect of leveling out the pdf of $\log_{10} Y^\alpha$. Looking back to Theorem 2, this has the effect of scaling the $\int_{[i,i+1]} |f(x) - \int_{[i,i+1]} f(t)dt|dx$ terms in Equation (3.2) to $\int_{[i,i+1]} |f(x/\alpha)/\alpha - \int_{[i,i+1]} f(t/\alpha)/\alpha dt|dx$ thereby improving the approximation. More generally, any transformation g which has this effect on $\log_{10} Y$ will eventually make $g(Y)$ $\epsilon-$satisfy Benford's Law. However, the particular transformation $g(x) = x^\alpha$ is of interest due to its simplicity and relevance for commonly modeled distributions. FSD frequencies of common distributions are contrasted with the same distributions raised to the tenth power in Table 3.

Table 3. FSD Frequencies

	First Significant Digit								
	1	2	3	4	5	6	7	8	9
Benford's Law	**.301**	**.176**	**.125**	**.097**	**.079**	**.067**	**.058**	**.051**	**.046**
Normal(0,1)	.359	.129	.087	.081	.077	.073	.069	.064	.060
Uniform(0,1)	.111	.111	.111	.111	.111	.111	.111	.111	.111
Log-Normal(0,1)	.308	.170	.119	.094	.079	.068	.060	.053	.048
Exponential(1)	.330	.174	.113	.086	.072	.064	.058	.053	.049
Pareto(1,1)	.556	.185	.093	.056	.037	.026	.020	.015	.012
Normal(0,1)^{10}	.301	.176	.125	.097	.079	.067	.058	.051	.046
Uniform(0,1)^{10}	.277	.171	.126	.100	.084	.072	.063	.056	.051
Log-Normal(0,1)^{10}	.301	.176	.125	.097	.079	.067	.058	.051	.046
Exponential(1)^{10}	.301	.176	.125	.097	.079	.067	.058	.051	.046
Pareto(1,1)^{10}	.326	.180	.123	.093	.075	.062	.053	.046	.041

Sample Size of 10^7 using the default pseudo-random generator in R.

Table 3 shows a striking convergence of FSDs to Benford's Law following the transformation of being raised to the tenth power. Table 4 highlights the conformance to Benford's Law induced by the transformation x^{10}. The Max Deviation column of Table 4 lists the maximum FSD frequency deviation from the Benford prediction for each row, showing that even the Uniform$(0,1)^{10}$ distribution obeys Benford's Law reasonably well. The Theorem 2 Bound column lists the Upper Bound on deviation from Benford's Law given by Theorem 2. Although this bound is not terribly good for the first column of distributions in Table 3, they become reasonable for the second column after the transformation x^{10} is applied.

TABLE 4. Conformance with Benford's Law (Sample Size: 10^7)

Distribution	Max Deviation	Theorem 2 Bound	**Distribution**	Max Deviation	Theorem 2 Bound
Normal(0,1)	.058	.673	Normal$(0,1)^{10}$	.000	.056
Uniform(0,1)	.190	.538	Uniform$(0,1)^{10}$	.024	.058
Log-Normal(0,1)	.007	.547	Log-Normal$(0,1)^{10}$	.000	.046
Exponential(1)	.029	.520	Exponential$(1)^{10}$	.000	.042
Pareto(1,1)	.255	.538	Pareto$(1,1)^{10}$	.025	.058

We have just seen that the transformation $g(x) = x^\alpha$ ensures reasonable conformance to Benford's Law for $\alpha = 10$. More generally, how fast do random variables conform to Benford's Law as α increases? Here I first show that under mild conditions, a rate of convergence of $O(1/\log_{10}\alpha)$ to Benford's Law can be guaranteed. This means for a random variable X^α, the maximum FSD deviation ϵ from the Law is $\leq C/\log_{10}\alpha$ for some constant C determined by X.

I then consider families of distributions which are closed under the transformation $g(x) = x^\alpha$, in other words if X is the initial random variable then X^α is again in the distributional family. These considerations allow us to connect conformance to Benford's Law with parameter values for some common distributions.

4.1. **A Rate of Convergence to Benford's Law.** This paper has shown that as α increases, X^α tends to satisfy Benford's Law. However, for statistical testing of Benford's Law, we need to pick α so that X^α satisfies the Law within, say $\epsilon = .01$. How large does α need to be? In other words, if $\epsilon(\alpha)$ denotes the least ϵ such that

X^α $\epsilon-$satisfies Benford's Law, how fast does $\epsilon(\alpha)$ decrease? The answer is provided by the following result.

Theorem 4. *Let X be a random variable with a differentiable pdf f. Let $\epsilon(\alpha)$ denote the least ϵ such that X^α $\epsilon-$satisfies Benford's Law. $\epsilon(\alpha)$ is $O(1/\log_{10}\alpha)$ provided that*

(1) $\mathrm{E}\,|\log_{10} X| < \infty$
(2) $\sup_x \left|\frac{d}{dx} x f(x)\right| < \infty$

In addition, $\epsilon(\alpha)$ is $o(1/\log_{10}\alpha)$ when $\mathrm{E}\,|\log_{10} X|^2 < \infty$.

Proof. See Appendix. □

This theorem shows that if $\epsilon(\alpha)$ is the maximum deviation of X^α from Benford's Law, then $\epsilon(\alpha) \leq C/\log_{10}\alpha$ for some constant C determined by X. The constant may be determined from the proof for a given X, but as the Tables above illustrate, actual conformance to Benford's Law is often better than guaranteed. In practice, direct numerical calculation of how well X^α conforms to the Law is a superior method when one needs to know the exact level of conformance. However, the result does provide a useful stopping point for numerical algorithms by bounding α.

4.2. **Particular Families.** Motivated by the convergence results above, it is a natural question to ask which families of distributions will satisfy Benford's law for particular parameter values. From Theorem 1, a natural way to start looking is to find families of a variable X where X^s is again within the family. Three such common families are the Log-Normal, Weibull, and Pareto distributions. The effect of a transformation of $X \longrightarrow (X/\nu)^s$ within these families are summarized in Table 5. Theorem 1 implies that the transformed variables $(X/\nu)^s$ will $\epsilon-$satisfy Benford's Law for sufficiently large s and any ν. Table 5 shows it is no coincidence that the Log-Normal and Pareto families appear in the Table and the literature on scaling laws. If such distributions commonly occur in data, since for particular parameter values Theorem 1 applies, Benford's Law will be commonly observed in samples drawn from these distributions as well.

TABLE 5. Families Closed under Powers

Distribution	Functional Form	$(X/\nu)^s$ Parameters	Var(X)
Log-Normal(μ, σ)	$\left(x\sigma\sqrt{2\pi}\right)^{-1} \exp\{-(\ln x - \mu)^2/2\sigma^2\}$	$(s\mu - \ln v, s\sigma)$	$(\exp\{\sigma^2\} - 1)\exp\{2\mu + \sigma^2\}$
Weibull(k, λ)	$(k/\lambda)\,(x/\lambda)^{k-1} \exp\left\{-(x/\lambda)^k\right\}$	$(k/s, \lambda^s/\nu)$	$\lambda^2[\Gamma(1+2/k) - \Gamma(1+1/k)^2]$
Pareto(k, b)	$kb^k x^{-(k+1)}\mathbf{1}_{[b,\infty)}(x)$	$(k/s, b^2/\nu)$	$b^2k/[(k-1)^2(k-2)]$

For example, according to Table 5, if X is distributed Log-Normal(μ, σ^2) then $(X/\nu)^s$ is distributed Log-Normal$(s\mu - \ln v, s^2\sigma^2)$. Appealing to Theorem 1, $(X/\nu)^s$ $\epsilon-$satisfies Benford's Law for sufficiently large s, or equivalently, the Log-Normal distribution $\epsilon-$satisfies Benford's Law for sufficiently large σ^2. Consequently, for each distribution in Table 5 and $\epsilon > 0$ there is a region in the parameter space where the distribution will ϵ-satisfy Benford's Law. Referring to the Variance column in Table 5 this is roughly when the variance or shape parameter is sufficiently large. This formally confirms observations by Leemis et al. (2000) that increases in the shape parameter increase compliance with Benford's Law.

4.3. **Conclusion.** This paper derives new test values and improves upon existing tests for evaluating compliance with Benford's Law. Also provided are new results which broaden the range of data to which such tests can be applied through a simple transformation. This transformation also induces scale invariance with respect to compliance with Benford's Law which frees tests from dependence of choice of measurement units. An upper bound on the rate of convergence to Benford's Law is also provided. Methods in this paper may therefore be used to characterize precisely which particular members of a family of distributions satisfy Benford's Law, and have particularly clean implications for the Log-Normal, Weibull, and Pareto families. Finally, it is my hope that the methods of this paper might be applied when considering generalized classes of FSD distributions (Rodriguez 2004; Hurlimann 2006; Grendar et al. 2007) which are other promising avenues for relating limited distributional information to data quality.

References

Allaart, P. C. (1997), "An Invariant-Sum Characterization of Benford's Law," *Journal of Applied Probability*, 34, 288–291.

Benford, F. (1938), "The Law of Anomalous Numbers," *Proceedings of the American Philosophical Society*, 78, 551–572.

Berger, A. and Hill, T. P. (2007), "Newton's Method Obeys Benford's Law," *American Mathematical Monthly*, 114, 588–601.

Boyle, J. (1994), "An Application of Fourier Series to the Most Significant Digit Problem," *The American Mathematical Monthly*, 101, 879–886.

Cho, W. K. T. and Gaines, B. J. (2007), "Breaking the (Benford) law: Statistical fraud detection in campaign finance," *The American statistician*, 61, 218–223.

Conover, W. J. (1972), "A Kolmogorov Goodness-of-Fit Test for Discontinuous Distributions," *Journal of the American Statistical Association*, 67, 591–596.

Durtschi, C., Hillison, W., and Pacini, C. (2004), "The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data," *Journal of Forensic Accounting*, 5, 17–34.

Giles, D. E. (2007), "Benford's law and naturally occurring prices in certain ebaY auctions," *Applied Economics Letters*, 14, 157–161.

Grendar, M., Judge, G., and Schechter, L. (2007), "An empirical non-parametric likelihood family of data-based Benford-like distributions," *Physica A: Statistical Mechanics and its Applications*, 380, 429–438.

Hill, T. P. (1995a), "A Statistical Derivation of the Significant-Digit Law," *Statistical Science*, 10, 354–363.

— (1995b), “Base-Invariance Implies Benford’s Law,” *Proceedings of the American Mathematical Society*, 123, 887–895.

Hurlimann, W. (2006), “Generalizing Benford’s law using power laws: application to integer sequences,” *Arxiv preprint math.ST/0607166*.

Judge, G. and Schechter, L. (2009), “Detecting Problems in Survey Data using Benford’s Law,” *Journal of Human Resources*, 44, 1–24.

Leemis, L. M., Schmeiser, B. W., and Evans, D. L. (2000), “Survival Distributions Satisfying Benford’s Law,” *The American Statistician*, 54, 236–241.

Marchi, S. and Hamilton, J. T. (2006), “Assessing the Accuracy of Self-Reported Data: an Evaluation of the Toxics Release Inventory,” *Journal of Risk and Uncertainty*, 32, 57–76.

Nigrini, M. (1996), “A taxpayer compliance application of Benford’s law,” *Journal of the American Taxation Association*, 18, 72–91.

Noether, G. E. (1963), “Note on the Kolmogorov statistic in the discrete case,” *Metrika*, 7, 115–116.

Raimi, R. A. (1976), “The First Digit Problem,” *The American Mathematical Monthly*, 83, 521–538.

Rodriguez, R. J. (2004), “First Significant Digit Patterns from Mixtures of Uniform Distributions.” *The American Statistician*, 58, 64–72.

Scott, P. D. and Fasli, M. (2001), “Benford’s Law: An Empirical Investigation and a Novel Explanation,” Tech. rep., CSM Technical Report 349, Department of Computer Science, University Essex.

Stephens, M. A. (1970), “Use of the Kolmogorov-Smirnov, Cramer-Von Mises and Related Statistics Without Extensive Tables,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 32, 115–122.

van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge University Press.

Appendix A. Proofs

It is useful to partition $(0, \infty)$ into sets $\{A_{d,k}\}$ related to First Significant Digits.

Definition. For real k define the d^{th} FSD set of order k, $A_{d,k}$ by

$$A_{d,k} \equiv [d \cdot 10^k,\ (d+1) \cdot 10^k)$$

Clearly for any $x > 0$ the FSD of x is d iff there exists an integer k s.t. $x \in A_{d,k}$, so that x has FSD equal to d iff $x \in A_d$ where $A_d \equiv \bigcup_{k \text{ integer}} A_{d,k}$. In particular

$$\log_{10} A_{d,k} = [\log_{10} d \cdot 10^k,\ \log_{10}(d+1) \cdot 10^k) = [k + \log_{10} d,\ k + \log_{10}(d+1))$$

so that (where $|\cdot|$ denotes Lebesgue measure when appropriate) $|\log_{10} A_{d,k}| = \log_{10}(1 + 1/d)$ for any k. Carrying over the results to a general base b presents no overwhelming difficulties. However, as the literature has focused on applications using base 10 I stick to base 10 avoiding the extra notational baggage.

A.1. **Proofs for the Main Text.**

Lemma. *Suppose X is a positive random variable with continuous pdf and let $Y \sim \log_{10} X$. If $Y \in I(\epsilon)$ then X $\epsilon-$satisfies Benford's Law.*

Proof. Let f denote the pdf of Y, and by definition of $A_{k,d}$ and A_d we have that

$$\Pr(X \text{ has FSD} = d) = \Pr(Y \in \log_{10} A_d) = \sum_{k=-\infty}^{\infty} \int_{\log_{10} A_{d,k}} f(y)dy \tag{A.1}$$

By assumption $Y \in I(\epsilon)$ so there exist constants $\{c_i\}$ such that for each FSD d,

$$\begin{aligned} \epsilon &\geq \left| \sum_{k=-\infty}^{\infty} \int_{\log_{10} A_{d,k}} f(y)dy - \int_{\log_{10} A_d} \sum c_i \mathbf{1}_{[i,i+1)}(y)dy \right| \\ &= \left| \Pr(X \text{ has FSD} = d) - \sum_{k=-\infty}^{\infty} \int_{\log_{10} A_{d,k}} \sum c_i \mathbf{1}_{[i,i+1)}(y)dy \right| \qquad \text{(A.2)} \end{aligned}$$

where the second line follows from Equation (A.1). Since $\log_{10} d < 1$ we know that $[k + \log_{10} d, k + \log_{10} d + 1) \cap [i, i+1) = \emptyset$ unless $k = i$ so letting $\mathbf{1}_A$ denote the set indicator function,

$$\mathbf{1}_{[k+\log_{10} d, k+\log_{10} d+1]}(y) \sum c_i \mathbf{1}_{[i,i+1)}(y) = c_k \mathbf{1}_{\log_{10} A_{d,k}}(y) \tag{A.3}$$

Using Equation (A.3), we have

(A.4)
$$\sum_{k=-\infty}^{\infty} \int_{\log_{10} A_{d,k}} \sum c_i \mathbf{1}_{[i,i+1)}(y) dy = \sum_{k=-\infty}^{\infty} \int_{\log_{10} A_{d,k}} c_k dy = [\sum_{k=-\infty}^{\infty} c_k] \log_{10}(1 + 1/d)$$

Pairing Equations (A.4) with Equation (A.2) we have that

$$\epsilon \geq |\Pr(X \text{ has FSD} = d) - [\sum_{k=-\infty}^{\infty} c_k] \log_{10}(1 + 1/d)| \tag{A.5}$$

Finally from Lemma 2 we may assume WLOG that $c_i = \int_{[i,i+1]} f(x)dx$ so that $\sum c_k = 1$, giving the desired inequalities. $\square$

Lemma. *Suppose* $\int |f(x)| \, dx < \infty$. *Then* $c^* \equiv \int_{[0,1]} f(y)dy$ *solves*

$$\min_c \sup_{A \text{ measurable}} \left| \int_{[0,1] \cap A} [f(x) - c] dx \right|$$

and the minimum attained is $\frac{1}{2} \int_{[0,1]} |f(x) - c^*| \, dx$.

Proof. This holds for the same reason that the median is a minimum absolute distance estimator. See the supplemental appendix for details. $\square$

A useful bound on the minimum $\frac{1}{2} \int_{[0,1]} \left| f(x) - \int_{[0,1]} f(y)dy \right| dx$ in the last Lemma is the following:

Lemma 4. *Let Y be a random variable with continuous pdf f.*

$$\frac{1}{2}\int_{[0,1]}\left|f(x)-\int_{[0,1]}f(y)dy\right|dx \leq \min\left\{\int_{[0,1]}f(y)dy, \frac{1}{2}\sup_{y\in[0,1]}f(y)-\frac{1}{2}\inf_{y\in[0,1]}f(y)\right\}$$

Proof. The last Lemma showed that

$$\frac{1}{2}\int_{[0,1]}\left|f(x)-\int_{[0,1]}f(y)dy\right|dx = \min_{c}\sup_{A}\left|\int_{[0,1]\cap A}[f(x)-c]dx\right|$$

where A is any measurable set, so clearly for $c=0$ we have $\frac{1}{2}\int_{[0,1]}\left|f(x)-\int_{[0,1]}f(y)dy\right|dx \leq \int_{[0,1]}f(y)dy$. Alternatively, consider estimating $c^* \equiv \int_{[0,1]}f(y)dy$ by $\hat{c} \equiv \frac{1}{2}\sup_{y\in[0,1]}f(y)+\frac{1}{2}\inf_{y\in[0,1]}f(y)$. In this case, $|f(x)-\hat{c}| \leq \frac{1}{2}\sup_{y\in[0,1]}f(y)-\frac{1}{2}\inf_{y\in[0,1]}f(y)$ so

$$\sup_{A}\left|\int_{[0,1]\cap A}[f(x)-\hat{c}]dx\right| \leq \sup_{A}\int_{[0,1]\cap A}|f(x)-\hat{c}|\,dx \leq \frac{1}{2}\sup_{y\in[0,1]}f(y)-\frac{1}{2}\inf_{y\in[0,1]}f(y)$$

Putting the two bounds together gives the result. □

Theorem (Mean-Scale Approximation)**.** *Let Y be a random variable with continuous pdf. For each $\epsilon>0$ there exists a $\sigma(\epsilon)$ s.t. $\sigma\leq\sigma(\epsilon)$ implies $(Y-\mu)/\sigma\in I(\epsilon)$ for all μ.*

Proof. I first show $rY\in I(\epsilon)$ for sufficiently large r. Fix $\epsilon>0$ and denote the pdf of Y as f. For any fixed r, the pdf of rY is $f(x/r)/r$ so from Lemma 2, it is sufficient to show that

$$\sum_{k}\frac{1}{2}\int_{[k,k+1]}\left|f(x/r)/r-\int_{[k,k+1]}f(y/r)/rdy\right|dx\leq\epsilon$$

Since $\lim_{n\longrightarrow\infty}\Pr(|Y|\leq n)=1$ there exists an N s.t. $\Pr(|Y|\geq N-2)<\epsilon/2$. Now from Lemma 4 we know that

$$\sum_{|k|\geq rN-1}\frac{1}{2}\int_{[k,k+1]}\left|f(x/r)/r-\int_{[k,k+1]}f(y/r)/rdy\right|dx\leq\sum_{|k|\geq rN-1}\int_{[k,k+1]}f(y/r)/rdy=$$

$$\sum_{|k|\geq rN-1} \int_{[k/r,(k+1)/r]} f(y)dy \leq \sum_{|k|\geq N-2} \int_{[k,k+1]} f(y)dy < \epsilon/2$$

So to show $rY \in I(\epsilon)$ it is sufficient that for all sufficiently large r,

$$\sum_{|k|\leq rN} \frac{1}{2} \int_{[k,k+1]} \left| f(x/r)/r - \int_{[k,k+1]} f(y/r)/r dy \right| dx < \epsilon/2$$

Again from Lemma 4 we know

$$\text{(A.6)} \qquad \sum_{|k|\leq rN} \frac{1}{2r} \int_{[k,k+1]} \left| f(x/r) - \int_{[k,k+1]} f(y/r)dy \right| dx \leq \sum_{|k|\leq rN} \frac{1}{2r} \left[\sup_{y\in[k,k+1]} f(y/r) - \inf_{y\in[k,k+1]} f(y/r) \right]$$

Since f is uniformly continuous on $[-N, N]$ compact, $\exists \delta \in (0, 1)$ s.t.

$$\text{(A.7)} \qquad \sup_{y\in B(x,\delta)} f(y) - \inf_{y\in B(x,\delta)} f(y) < \epsilon/2N \quad \forall x \in [-N, N]$$

where $B(x, \delta)$ denotes a closed ball of radius δ around x. Equation (A.6) implies for all $r \geq 1/\delta$,

$$\sup_{y\in B(x,1)} f(y/r) - \inf_{y\in B(x,1)} f(y/r) < \epsilon/2N \quad \forall x \in [-N, N]$$

combining this with Equation (A.6), we have

$$\sum_{|k|\leq rN} \frac{1}{2r} \left[\sup_{y\in[k,k+1]} f(y/r) - \inf_{y\in[k,k+1]} f(y/r) \right] \leq \frac{2rN}{2r} \frac{\epsilon}{2N} = \frac{\epsilon}{2}$$

and we conclude $rY \in I(\epsilon)$ for all $r \geq 1/\delta$.

I now show that for sufficiently large r, $r(Y - \mu) \in I(\epsilon)$ for all μ. From Lemma 3 for any particular r it is sufficient to consider only $r\mu \in [0, 1)$ and since $r \geq 1$, WLOG $\mu \in [0,1)$. The proof proceeds as above, but now we must show that

$$\sum_{|k| \leq rN} \frac{1}{2} \int_{[k,k+1]} \left| f(x/r+\mu)/r - \int_{[k,k+1]} f(y/r+\mu)/r dy \right| dx < \epsilon/2$$

Following the proof exactly, simply choose $\tilde{\delta} \equiv \delta/2$ so that Equation (A.7) holds and for all $r \geq 1/\tilde{\delta}$ we have

$$\sup_{y \in B(x,2)} f(y/r) - \inf_{y \in B(x,2)} f(y/r) < \epsilon/2N \quad \forall x \in [-N,N]$$

This implies for all $\mu \in (-1,1)$ that

$$\sup_{y \in B(x,1)} f(y/r+\mu) - \inf_{y \in B(x,1)} f(y/r+\mu) < \epsilon/2N \quad \forall x \in [-N,N]$$

which when substituted into the proof above gives the result. □

Theorem. *Let X be a random variable with a differentiable pdf f. Let $\epsilon(\alpha)$ denote the least ϵ such that X^α $\epsilon-$satisfies Benford's Law. $\epsilon(\alpha)$ is $O(1/\log_{10}\alpha)$ provided*

(1) $\mathrm{E}\left|\log_{10} X\right| < \infty$

(2) $\sup_x \left|\frac{d}{dx} x f(x)\right| < \infty$

In addition, $\epsilon(\alpha)$ is $o(1/\log_{10}\alpha)$ when $\mathrm{E}\left|\log_{10} X\right|^2 < \infty$.

Proof. WLOG assume X is positive. Let Y_α be the random variable defined by $Y_\alpha \equiv \log_{10} X^\alpha$ so by Lemma 1, $\epsilon(\alpha)$ is bounded above by $\bar{\epsilon}(\alpha)$, where $\bar{\epsilon}(\alpha) \equiv \inf\{\epsilon : \ Y_\alpha \in I(\epsilon)\}$. Letting g_α denote the pdf of Y_α, Lemma 4 shows that $\bar{\epsilon}(\alpha)$ is bounded above by the following equation

$$\bar{\epsilon}(\alpha) \leq \sum_i \min\left\{ \int_{[i,i+1]} g_\alpha(y) dy, \sup_{y \in [i,i+1]} g_\alpha(y)/2 - \inf_{y \in [i,i+1]} g_\alpha(y)/2 \right\} \tag{A.8}$$

The first expression in the min of this is expression is exactly $\int_{[i,i+1]} g_\alpha(y) dy = \Pr\left(Y_\alpha = \log_{10} X^\alpha \in [i,i+1]\right)$. For the second expression, fix i and consider the change of variable

$$\sup_{y\in[i,i+1]} g_\alpha(y) = \sup_{y\in[i,i+1]} \frac{d}{dy}\Pr\left(\log_{10} X^\alpha \le y\right) \quad = \sup_{y\in[i,i+1]} \frac{d}{dy}\Pr\left(X \le 10^{y/\alpha}\right)$$
$$= \sup_{y\in[i,i+1]} \ln 10\cdot 10^{y/\alpha} f(10^{y/\alpha})/\alpha = \sup_{y\in[10^{i/\alpha},10^{i+1/\alpha}]} \ln 10 \cdot y f(y)/\alpha$$

Similar reasoning holds for the inf term. Since by assumption $M \equiv \sup\left|\frac{d}{dx} x f(x)\right| < \infty$, the mean value theorem implies

$$\sup_{y\in[a,b]} y f(y) - \inf_{y\in[a,b]} y f(y) \le M(b-a)$$

and therefore

$$\sup_{y\in[i,i+1]} g_\alpha(y) - \inf_{y\in[i,i+1]} g_\alpha(y) = \sup_{y\in[10^{i/\alpha},10^{i+1/\alpha}]} \ln 10\cdot y f(y)/\alpha - \inf_{y\in[10^{i/\alpha},10^{i+1/\alpha}]} \ln 10\cdot y f(y)/\alpha$$
$$\le M \ln 10 \cdot \left(10^{i+1/\alpha} - 10^{i/\alpha}\right)/\alpha$$

Substitution of these expressions into Equation (A.8) yields

$$\bar{\epsilon}(\alpha) \le \sum_i \min\left\{\Pr\left(\log_{10} X^\alpha \in [i, i+1]\right), M\ln 10\cdot\left(10^{i+1/\alpha} - 10^{i/\alpha}\right)/\alpha\right\}$$

Now for any positive real number k we have

$$\bar{\epsilon}(\alpha) \le \sum_{|i|\ge k} \Pr\left(\log_{10} X^\alpha \in [i, i+1]\right) + \sum_{i<k+1} M\ln 10\cdot\left(10^{i+1/\alpha} - 10^{i/\alpha}\right)/\alpha$$
$$\le \Pr\left(\left|\log_{10} X^\alpha\right| \ge k\right) + M\ln 10\cdot 10^{(k+1)/\alpha}/\alpha \tag{A.9}$$

A Chebyshev type inequality shows that

$$\Pr\left(\left|\log_{10} X^\alpha\right| \ge k\right) = \Pr\left(\left|\log_{10} X\right| \ge k/\alpha\right) \le \alpha \mathrm{E}\left|\log_{10} X\right|/k$$

Using this bound in Equation (A.9) yields the following bound on $\overline{\epsilon}(\alpha)$:

$$\overline{\epsilon}(\alpha) \leq \alpha \mathrm{E} \left|\log_{10} X\right| / k + M \ln 10 \cdot 10^{(k+1)/\alpha} / \alpha$$

Consider the choice $k = \alpha \log_{10} \alpha / 2$ so that

$$\overline{\epsilon}(\alpha) \leq 2\mathrm{E} \left|\log_{10} X\right| / \log_{10} \alpha + 10^{1/\alpha} M \ln 10 \cdot \alpha^{-1/2}$$

Clearly then $\lim_{\alpha \longrightarrow \infty} \overline{\epsilon}(\alpha) \log_{10} \alpha \leq 2\mathrm{E} \left|\log_{10} X\right| < \infty$ so $\epsilon(\alpha) \leq \overline{\epsilon}(\alpha)$ is $O(1/\log_{10} \alpha)$. Apply a similar Chebyshev type inequality when $\mathrm{E} \left|\log_{10} X\right|^2 < \infty$ for the same choice of k shows $\epsilon(\alpha)$ is $O(1/(\log_{10} \alpha)^2)$ and therefore $o(1/\log_{10} \alpha)$. $\square$