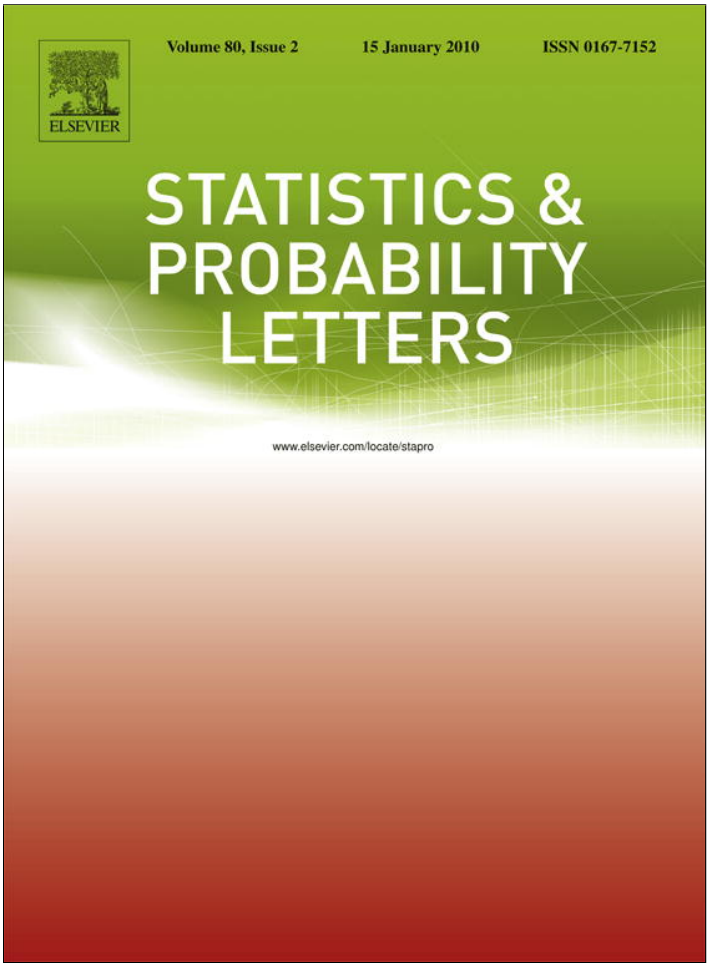
Volume 80, Issue 2
15 January 2010
ISSN 0167-7152
ELSEVIER
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Statistics and Probability Letters 80 (2010) 82–88

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Stigler's approach to recovering the distribution of first significant digits in natural data sets

Joanne Lee [a], Wendy K. Tam Cho [b,c,d,*], George G. Judge [a]

[a] Graduate School at University of California, Berkeley, CA 94720, United States
[b] Department of Political Science, University of Illinois at Urbana-Champaign, United States
[c] Department of Statistics, University of Illinois at Urbana-Champaign, United States
[d] The National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, United States

ARTICLE INFO

Article history:
Received 21 January 2009
Received in revised form 7 September 2009
Accepted 20 September 2009
Available online 1 October 2009

JEL classification:
C10
C24

AMS classification:
primary 62E20

ABSTRACT

Benford's Law can be seen as one of the many first significant digit (FSD) distributions in a family of monotonically decreasing distributions. We examine the interrelationship between Benford and other monotonically decreasing distributions such as those arising from Stigler, Zipf, and the power laws. We examine the theoretical basis of the Stigler distribution and extend his reasoning by incorporating FSD first-moment information into information-theoretic methods. We present information-theoretic methods as a way to describe, connect, and unify these related distributions and thereby extend the reach of Benford's Law and FSD research more generally.

1. Introduction

In 1881, astronomer and mathematician Simon Newcomb noticed that the first several pages of logarithm tables were more worn than subsequent pages. This observation led him to the counter-intuitive conjecture that, in a "natural" data set, the digit "1" would occur most frequently and "9" would occur least frequently (Newcomb, 1881). Newcomb stated, "the law of probability of the occurrence of numbers is such that all mantissæ of their logarithms are equally probable," and suggested the following expression for the empirical distribution of first significant digits (FSD):

$$P(d) = \log_{10}\left(\frac{1+d}{d}\right) \quad \text{for } d = 1, \ldots, 9, \tag{1}$$

where $P(d)$ is the relative frequency of the digit d as a first significant digit. The resulting monotonically decreasing relative frequency values for $d = 1, 2, \ldots, 9$ are (0.301, 0.176, 0.125, 0.097, 0.079, 0.067, 0.058, 0.051, 0.046). Perhaps because Newcomb did not proffer a theoretical explanation or an empirical verification of the phenomenon, his conjecture did not garner much immediate attention.

1.1. Benford's Law

Fifty-seven years later, Frank Benford set out to empirically verify Newcomb's hypothesis by demonstrating that 20,229 observations compiled from seemingly unrelated sets of numbers provided a good fit to the distribution first laid out by Newcomb (Benford, 1938). These diverse data sets included the populations of cities, street addresses, American League baseball statistics, numbers appearing in *Reader's Digest*, and the area of rivers, among others. Benford's empirical display of

* Corresponding address: 240 Computing Applications Building, 605 E. Springfield Ave, Champaign, IL 61820, United States.
E-mail addresses: joanne.lee@berkeley.edu (J. Lee), wendycho@illinois.edu (W.K.T. Cho), gjudge@berkeley.edu (G.G. Judge).

0167-7152/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2009.09.015

this FSD pattern led to the naming the phenomenon as "Benford's Law," after its popularizer rather than its discoverer. Subsequently, Benford's Law has been shown to approximately apply to a large number of other data sets, including electricity usage, word frequency, ebaY bids, census statistics, and campaign donations (Raimi, 1976; Zipf, 1949; Hill, 1995; Giles, forthcoming; Cho and Gaines, 2007). Fascinatingly, Pinkham (1961) has additionally demonstrated that Benford's Law is scale invariant—whether the unit is dollars, yen, inches, meters, or hectares has no bearing on the fit to Benford's Law.

These findings are intriguing and have led many to wonder why numbers might follow Benford's Law. Benford suggested that the law held when data came from a mixture of uniform distributions that were more likely to have relatively small upper bounds.[1] Raimi (1976) posited that Benford's mixture scheme is arbitrary and approximate because it implies that a variety of other "laws" could also be created by mixing different distributions, causing one to wonder why mixtures of *uniform* distributions would be especially relevant to describing distributions of first significant digits. George Stigler, a future Nobel Laureate in Economics, claimed that the specific mixture of uniform distributions with non-uniformly distributed maximum values is, minimally, an inconsistency. This observation led Stigler (1945) to propose an alternative FSD distribution that was less skewed toward the lower digits and was derived without the use of such assumptions.

Indeed, despite the empirical verification that a large number of unrelated data sets follow Benford's Law, the literature has also recognized that many data sets deviate from the "law of anomalous numbers" (Durtschi et al., 2004). Curiously, even when the FSDs in data sets deviate from the logarithmic pattern, the relative frequency of digits appear to still favor the lower digits and declines monotonically in a manner akin to Benford's Law, implying perhaps that a generalized form of Benford's Law might be more widely applicable. Recently, power law and information-theoretic methods have been proposed as being more intuitively appealing and generalizable ways of determining similar FSD distributions (Grendar et al., 2006). Pietronero et al. (2001) suggest that Benford's Law is a special case of a power law.

The purpose of this paper is to review the basis of Stigler's FSD solution and to present a data-based, information-theoretic approach to recovering Stigler-like FSD distributions. The structure of the paper is as follows. Section 2 describes Stigler's proposed alternative approach and compares it to that of Benford. Section 3 introduces the power law concept and uses it to exhibit the fact that the relative frequency of a first significant digit decays as a power law of its rank in terms of appearance. Section 4 demonstrates how Cressie–Read minimum divergence-distance measures create Benford-like distributions based on the first moment of given data. Finally, Section 5 discusses implications for the use of these scale-invariant methods.

2. Stigler's FSD concept

Stigler (1945) reviewed the Newcomb–Benford FSD phenomenon and proposed that the average relative frequency of a leading significant digit, d, is

$$F_d = \frac{d\ \ln(d) - (d+1)\ \ln(d+1) + \left(1 + \frac{10}{9}\ln(10)\right)}{9}. \tag{2}$$

He arrived at this conclusion by first assuming that the largest entry in the given statistical table is equally likely to begin with $d = 1, 2, \ldots, 9$, and that all other entries in the table are randomly selected from the uniform distribution of numbers smaller than the largest entry. Defining the rth cycle of numbers as being the interval $[10^r, 10^{r+1}]$ for some real number r, Stigler finds the distribution of FSDs for the highest entry in a cycle of numbers from the table and then averages the probabilities over all highest entries. Since table entries are from a uniform distribution, any digit d should have, at the end of the $(r-1)$st cycle, occurred $(10^r - 1)/9$ times as an FSD out of $10^r - 1$ numbers, approximately $10^r/9$ and 10^r, respectively. For example, at the end of the first cycle, i.e., [10,100), the digit "2" has occurred as an FSD $(10^2 - 1)/9 = 11$ times out of $10^2 - 1 = 99$ numbers, including those from all previous cycles. After the $(r-1)$st cycle, d does not appear as an FSD for the next $(d-1)10^r$ numbers, e.g., "2" does not arise as an FSD in the interval $[10^2, 10^2 + (2-1)(10^2)) = [100, 200)$. Following this logic, we see that

$$p_i = \frac{d_i \ln d_i - (d_i + 1)\ln(d_i + 1) + m}{9}, \tag{3}$$

where m is defined as

$$m = \frac{\sum_{i=1}^{9} i^2 \ln(d_i) - d_i(d_i+1)\ln(d_i+1)}{9 - \sum_{i=1}^{9} d_i}. \tag{4}$$

The resulting frequencies from Stigler's derivation are presented in Table 1.[2] The frequencies from Benford's Law are presented for comparison. While the relative frequencies differ, the sets of frequencies are similar in their monotonically

[1] Rodriguez (2003) noted that Benford's Law can be obtained without mixtures of distributions, but is obtainable from data drawn from a lognormal distribution with a relatively high variance parameter.

[2] While there have been many empirical examples of Benford's Law, there have been relatively few empirical examples of Stigler's Law. One exception derives from Ley (1996), who presented stock market data as an example of Benford's Law at work. Rodriguez (2004) later showed that Stigler's Law provided a better fit to these data.

Table 1
Comparison of the Benford and Stigler distributions.

FSD	Stigler's Law	Benford's Law
1	0.241	0.301
2	0.183	0.176
3	0.146	0.125
4	0.117	0.097
5	0.095	0.079
6	0.077	0.067
7	0.061	0.058
8	0.047	0.051
9	0.034	0.046

decreasing pattern. Because no logarithmic FSD distribution holds generally for all natural data sets, Stigler's Law and Benford's Law might be viewed as members of a family of monotonically decreasing distributions of FSDs.

Stigler claims that the difference between his alternative and Benford's Law arises from the hidden assumptions Benford made about the relative frequencies of the largest numbers in statistical tables. Benford assumed that smaller numbers with corresponding smaller FSDs occurred more often as bounds for statistical tables. In particular, given a mixture of uniform distributions $U[0, b)$, the density of the upper bound b is assumed to be proportional to $\frac{1}{b}$. Stigler argued that this assumption was unnecessary in deriving a logarithmic rule, since it neither expanded the scope of the law nor contributed to the theoretical basis for modeling a distribution of first significant digits. In contrast, Stigler's assumption is that the largest entries in statistical tables were equally likely to begin with $d = 1, 2, \ldots, 9$ (Stigler, 1945).[3]

3. Connections to the power law and Zipf's Law

3.1. Power law

In Section 1, we noted the suggested role of scale invariance that underlies the uneven distributions in data outcomes in economics, linguistics, and many other natural phenomena. Scale invariance occurs if the outcome does not change when either the underlying data distribution, $Prob(D) = P(D)$, or its FSD counterpart, $P(d)$, is multiplied by a constant s (Mandelbrot, 1982). Pietronero et al. (2001) note that scale invariance leads to the functional relation

$$P(sD) = P(D^*) = K(p)P(D), \tag{5}$$

and that the general solution to (5) has the power law nature

$$P(D^*) = P(D^{*-\alpha}) = s^{-\alpha}D^{-\alpha}, \tag{6}$$

where the exponent α is a constant.

For these types of distributions, we can, in Stigler-like fashion, compute the probability of the first digit by noting that we have the same (uniform) relative probability for the integers $d = 1, 2, \ldots, 9$, for each cycle. Following Pietronero et al. (2001), for $\alpha \neq 1$,

$$P(D^*) = \int_d^{d+1} D^{-\alpha}\,\mathrm{d}D = \frac{1}{1-\alpha}[(d+1)^{1-\alpha} - d^{1-\alpha}]. \tag{7}$$

For $\alpha = 1$,

$$P(D^*) = \int_d^{d+1} D^{-1}\,\mathrm{d}D = \int_d^{d+1} d\,(\log D) = \log\left(\frac{d+1}{d}\right). \tag{8}$$

This expresses Benford's Law as determined from the underlying data distribution. Consequently, in a power law context when $\alpha = 1$, we have a uniform FSD in logarithmic space. For values of $\alpha > 1$, the FSD distribution is more tilted than Benford so that the first digit 1 is even more frequent. For values of $\alpha < 1$, the FSD distribution tends toward a uniform FSD distribution. The case of $\alpha = 1$ seems to appear frequently in nature, as evidenced by the large number of data sets that exhibit the Benford pattern. Clearly, however, α may take on other values as well, and so, this is one way in which we might view this family of power laws as a generalized Benford law (Pietronero et al., 2001).

3.2. Zipf's Law

Zipf's Law characterizes a rank order statistic and bears similarities to Benford's Law, most notably in that now familiar monotonically decreasing distribution of relative frequencies. In addition, Zipf's Law is scale invariant and is applicable to

[3] An alternative method of deriving Stigler's FSD rule based on the idea of mixing uniform distributions is given in Rodriguez (2004) and is provided in the Appendix for interested readers.

a large range of phenomena, including income distributions, city sizes, and word frequency (Zipf, 1949; Raimi, 1976). For instance, Zipf's Law links word frequency, W, to its rank order k,

$$W(k) = \frac{A}{k}, \tag{9}$$

where A is the frequency of the most common word (i.e. the one with rank $k = 1$). The second most common or frequent word would have rank $k = 2$ and frequency $W(2) = A/2$, and so forth so that the most frequent word occurs approximately twice as often as the second most frequent word, which occurs twice as often as the fourth most frequent word (Zipf, 1949).

Following Pietronero et al. (2001), consider the rank order properties of a set of $\mathcal{N}$ numbers extracted from a general distribution, $P(N) \sim N^{\alpha}$. Let $N_{\max}$ be the largest value in the set $\mathcal{N}$, a finite value that corresponds to the rank $k = 1$. The rank k for any number in the set is then

$$k = \mathcal{N} \int_{N(k)}^{N_{\max}} P(N)\,\mathrm{d}N \sim N(k)^{1-\alpha}. \tag{10}$$

Inverting (10) gives us

$$N(k) \sim k^{\frac{1}{1-\alpha}}, \tag{11}$$

which highlights a link between Benford's Law and Zipf's Law. Benford's Law ($\alpha = 1$) does not lead to a Zipf-type law because $N_{\max}$ diverges. However, any power law distribution where $\alpha > 1$ leads to a generalized Zipf's Law with exponent $1/(1-\alpha)$, and Benford's Law can be seen as a part of the family of power laws.

4. Problem reformulation and solution

In the previous section, we discussed the Benford, Stigler, power, and Zipf's Law approaches to determining the distribution of FSDs and investigated their interrelationship. We now discuss how information-theoretic methods produce similar distributions, and highlight their unique ability to easily adapt the specific distribution to moment information from any particular data set. Since phenomena often have unique traits, a distribution that is adaptable to data peculiarities is desirable if such individual idiosyncrasies might affect the particularities of the monotonically decreasing distribution.

In the context of recovering the FSD distribution from a sequence of positive real numbers, assume for the discrete random variable d_i for $i = 1, 2, \ldots, 9$, that at each trial, one of nine digits is observed with probability p_i. Suppose after n trials, we are given first-moment information in the form of the average value of the FSD:

$$\sum_{j=1}^{9} d_j p_j = \bar{d}. \tag{12}$$

Assuming that the only information that exists is this first-moment information, our inverse problem consists of identifying an FSD distribution that reflects the best predictions of the unknown probabilities, $p_1, p_2, \ldots, p_9$. It is readily apparent that there is one data point and nine unknowns, resulting in an ill-posed inverse problem where there exist an infinite number of possible discrete probability distributions with $\bar{d} \in [1, 9]$. Based only on the mean, $\sum_{j=1}^{9} d_j p_j = \bar{d}$, and two constraints on probabilities, $\sum_{j=1}^{9} p_j = 1$, and $0 \le p_j \le 1$, the problem does not have a unique solution. A function must be inferred from insufficient information when only a feasible set of solutions is specified. In such a situation, it is useful to have an approach that allows the investigator to adapt sample-based information recovery methods without having to commit the FSD function to a particular parametric family of probability densities. The goal is to reduce the infinite dimensional non-parametric problem to one that is finite dimensional. Ideally, we do so without imposing more assumptions than are necessary.

4.1. An information-theoretic approach

One way to solve this ill-posed inverse problem for the unknown p_j without making a large number of assumptions or introducing additional information is to formulate it as an extremum problem. This type of extremum problem is in many ways analogous to allocating probabilities in a contingency table where p_j and q_j are, respectively, the observed and expected probabilities of a given event. A solution is achieved by minimizing the divergence between the two sets of probabilities. That is, we are optimizing a goodness-of-fit (pseudo-distance measure) criterion subject to data-moment constraint(s). One attractive set of divergence measures is the Cressie–Read (CR) power divergence family of statistics (Cressie and Read, 1984; Read and Cressie, 1988; Baggerly, 1998):

$$I(\mathbf{p}, \mathbf{q}, \gamma) = \frac{1}{\gamma(1+\gamma)} \sum_{j=1}^{9} \left(p_j \left[\left(\frac{p_j}{q_j} \right)^{\gamma} - 1 \right] \right), \tag{13}$$

where γ is an arbitrary and unspecified parameter.

In the context of recovering the unknown FSD distribution, use of the CR criterion (13) suggests we seek, given $\mathbf{q}$, a solution to the following extremum problem:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \left[I(\mathbf{p}, \mathbf{q}, \gamma) \mid \sum_{j=1}^{9} p_j d_j = \bar{d}, \sum_{j=1}^{9} p_j = 1, p_j \geq 0 \right]. \tag{14}$$

In the limit, as γ ranges from -1 to 1, two main variants of $I(\mathbf{p}, \mathbf{q}, \gamma)$ have received explicit attention in the literature (see Mittelhammer et al. (2000)). Assuming for expository purposes that the reference distribution is discrete uniform, i.e. $q_j = 1/9\ \forall j$, then $I(\mathbf{p}, \mathbf{q}, \gamma)$ converges to an estimation criterion equivalent to the Owen (2001) empirical likelihood (EL) criterion $\sum_{j=1}^{9} \ln(p_j)$, when $\gamma \to -1$. The EL criterion assigns discrete mass across the nine possible FSD outcomes, and in the sense of objective function analogies, it is closest to the classical maximum-likelihood approach. In fact, it results in a maximum non-parametric likelihood alternative. The second prominent case for the CR statistic corresponds to letting $\gamma \to 0$ and leads to the criterion $-\sum_{j=1}^{9} p_j \ln(p_j)$, which is the maximum entropy (ME) or the Shannon (1948) and Jaynes (1957a,b) entropy function.

The ME criterion distance measure is equivalent to the Kullback–Leibler (KL) information criterion (Kullback, 1959), and finds the feasible $\hat{\mathbf{p}}$ that define the minimum value of all possible expected log-likelihood ratios consistent with, in our case, the FSD mean. Solutions for these distance measures cannot be written in closed form, but are instead, determined numerically through optimization algorithms.

4.2. Maximum entropy formulation

If we use the CR ($\gamma = 0$) criterion for the first digit case, we would select the ME probabilities that maximize

$$H(\mathbf{p}) = -\sum_{j=1}^{9} p_j \ln(p_j), \tag{15}$$

subject to the mean $\bar{d}$, where

$$\bar{d} = \sum_{j=1}^{9} p_j d_j, \tag{16}$$

and the condition that the probabilities must sum to one

$$\sum_{j=1}^{9} p_j = 1. \tag{17}$$

The Lagrangian for the extremum problem is

$$L = -\sum_{j=1}^{9} p_j \ln(p_j) + \lambda \left(\bar{d} - \sum_{j=1}^{9} p_j d_j \right) + \eta \left(1 - \sum_{j=1}^{9} p_j \right). \tag{18}$$

Since H is strictly concave, there is a unique interior solution. Solving the first-order conditions yields the ME exponential result

$$\hat{p}_i = \frac{\exp(-d_i \hat{\lambda})}{9 \sum_{j=1}^{9} \exp(-d_j \hat{\lambda})}, \tag{19}$$

for the ith outcome. In this context, the $\hat{p}_i$ are exponentially FSD and the chosen FSD distribution is the one that has the greatest combinatorial multiplicity. We note again that $p(\lambda)$ is a member of a canonical exponential family with mean

$$\bar{d} = \sum_{j=1}^{9} p_j(\lambda) d_j, \tag{20}$$

and Fisher's information measure for λ (see Golan et al., 1996, p. 26)

$$I(\lambda) = \sum_{j=1}^{9} p_j(\lambda) d_j^2 - \left(\sum_{j=1}^{9} p_j(\lambda) d_j \right)^2 = \mathrm{Var}(d). \tag{21}$$

Table 2
The estimated maximum entropy (ME) distributions (with a uniform reference distribution) for the digit problem.

FSD mean	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{p}_5$	$\hat{p}_6$	$\hat{p}_7$	$\hat{p}_8$	$\hat{p}_9$	$H(\hat{\mathbf{p}})$
2.0	0.496	0.251	0.126	0.064	0.032	0.016	0.008	0.004	0.002	1.38
3.0	0.306	0.217	0.153	0.108	0.077	0.054	0.038	0.027	0.019	1.88
3.55	0.238	0.188	0.149	0.118	0.093	0.074	0.058	0.046	0.036	2.03
4.0	0.191	0.163	0.140	0.120	0.103	0.088	0.075	0.065	0.055	2.12
4.5	0.148	0.137	0.127	0.118	0.109	0.101	0.094	0.087	0.081	2.18
5.0	0.111	0.111	0.111	0.111	0.111	0.111	0.111	0.111	0.111	2.20
5.5	0.081	0.087	0.094	0.101	0.109	0.118	0.127	0.137	0.148	2.18

4.3. Some mean-related ME distributions

Using a uniform reference distribution, the resulting ME distributions for a range of FSD means (including the Stigler mean of 3.55) are presented in Table 2.[4] As we can see from the table, when the FSD mean is 5, the ME solution is a uniform distribution and results in the maximum entropy value for $H(\hat{\mathbf{p}})$. The Stigler FSD mean, 3.55, yields a monotonically decreasing ME distribution consistent with the Stigler distribution and a correlation with the Stigler distribution that approaches 1.0. Several other monotonically decreasing distributions resulting from a variety of mean values are also shown.

In Table 2, consider any mean. There are an infinite number of solutions or sets of probabilities, $p_1, p_2, \ldots, p_9$, that are consistent with any particular mean value. That is, there are many combinations of first significant digits that will yield a particular mean. In order to choose among these possible solutions, we have employed the maximum entropy principle. The entropy value, $H(\hat{p})$, that we maximize can be understood as a numerical measure ranging from zero (completely informative) to its maximum value (completely uninformative). These monotonically decreasing maximum entropy distributions with their corresponding mean values are shown in Table 2.

Under ME, the exponential null hypotheses that result have especially appealing properties. We have minimized the number of underlying assumptions required to arrive at a solution. In addition, this criterion choice yields a solution distribution with maximum combinatorial multiplicity. This property is desirable because, in the absence of assumptions, the chosen distribution among the possible distributions should logically be the one that occurs most frequently, i.e. the choice with maximum multiplicity. Lastly, the chosen distribution is as close to the uniform distribution as the data will permit.

5. Discussion

Benford's Law has been shown to be applicable to a large set of seemingly unrelated phenomena from the area of rivers to campaign donations to census statistics. Indeed, the boundaries of this set are far ranging. At the same time, not all data sets follow Benford's Law (Durtschi et al., 2004). Some appear to be related to Stigler's Law. Others follow the outlines of the power law or Zipf's Law. Each law appears to apply to some data contexts, but none apply to all contexts. As we have shown, these various laws are related and can be viewed as members of a family of monotonically decreasing distributions.

In this paper, we have provided a basis for describing, connecting, and unifying this family of distributions. We have also highlighted how first significant digits can be examined in a data-adaptive context. As a data set's FSD mean changes, our information-theoretic methods suggest alternative null hypotheses for the digit proportions. These methods also supply a basis for realizing an exponential family of FSD distributions and relating it to a particular underlying data set distribution. In so doing, our results extend the range of Benford's Law to data contexts that initially seem to violate Benford's Law.

Appendix

A.1. Mixing uniform distributions

From Section 2, we know that the probability of an FSD being d depends on which of three distinct ranges within the rth cycle we are examining. Noting Stigler's assumption of uniformly distributed upper bounds in a given data set, we obtain the density function of the upper bound b,

$$f(b) = \frac{1}{9 \times 10^r} \tag{A.1}$$

[4] For comparison and curiosity, results when the reference distribution is the Stigler distribution are presented in the Appendix.

and integrate over the three regions to find Stigler's Law for $d \in 1, 2, \ldots, 9$,

$$\begin{aligned}
P(FSD = d) &= \int_{10^r}^{d10^r} \frac{10^r}{9b} \mathrm{d}F(b) + \int_{d10^r}^{(d+1)10^r} \left(\frac{10^r}{9b} + \frac{b - d10^r}{b} \right) \mathrm{d}F(b) + \int_{(d+1)10^r}^{10^{r+1}} \frac{10^{r+1}}{9b} \mathrm{d}F(b) \\
&= \frac{1}{9 \times 10^r} \left(\frac{10^r}{9} \int_{10^r}^{(d+1)10^r} \frac{\mathrm{d}b}{b} + \int_{d10^r}^{(d+1)10^r} db - d10^r \int_{d10^r}^{(d+1)10^r} \frac{\mathrm{d}b}{b} + \frac{10^{r+1}}{9} \int_{(d+1)10^r}^{10^{r+1}} \frac{\mathrm{d}b}{b} \right) \\
&= \frac{1}{9 \times 10^r} \left[\frac{10^r}{9} \ln(d+1) + 10^r - d10^r \ln\left(\frac{d+1}{d} \right) + \frac{10^{r+1}}{9} \ln\left(\frac{10}{d+1} \right) \right] \\
&= \frac{1}{9} \left(1 + \frac{10}{9} \ln(10) + d \ln(d) - (d+1) \ln(d+1) \right). \qquad \text{(A.2)}
\end{aligned}$$

A.2. Estimates with Stigler FSD reference distribution

See Table A.1.

Table A.1
The estimated maximum entropy (ME) distributions (with a Stigler FSD reference distribution) for the digit problem.

FSD mean	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{p}_5$	$\hat{p}_6$	$\hat{p}_7$	$\hat{p}_8$	$\hat{p}_9$	$H(\hat{\mathbf{p}})$
2.0	0.503	0.244	0.124	0.064	0.033	0.017	0.009	0.004	0.002	1.38
3.0	0.312	0.211	0.150	0.108	0.078	0.056	0.040	0.027	0.018	1.88
3.55	0.241	0.183	0.146	0.117	0.095	0.077	0.061	0.047	0.034	2.03
4.0	0.194	0.159	0.137	0.119	0.104	0.091	0.078	0.065	0.051	2.12
4.5	0.150	0.133	0.124	0.117	0.111	0.105	0.098	0.088	0.075	2.18
5.0	0.112	0.107	0.109	0.110	0.113	0.116	0.116	0.113	0.103	2.20
5.5	0.081	0.084	0.091	0.100	0.111	0.123	0.133	0.140	0.138	2.18

References

Baggerly, K., 1998. Empirical likelihood as a goodness of fit measure. Biometrika 85 (3), 535–547.
Benford, Frank, 1938. The law of anomalous numbers. Proceedings of the American Philosophical Society 78 (4), 551–572.
Cho, Wendy K. Tam, Gaines, Brian J., 2007. Breaking the (Benford) Law: Statistical Fraud detection and campaign finance. The American Statistician 61 (3), 218–223.
Cressie, Noel, Read, Timoth R.C., 1984. Multinomial goodness of fit tests. Journal of the Royal Statistical Society, Series B 46, 440–464.
Durtschi, Cindy, Hillison, William, Pacini, Carl, 2004. The effective use of Benford's Law to assist in detecting Fraud in accounting data. Journal of Forensic Accounting 5 (1), 17–34.
Giles, David E., 2006. Benford's Law and naturally occurring prices in certain ebaY auctions. Applied Economics Letters (forthcoming).
Golan, Amos, Judge, George, Miller, Douglas, 1996. Maximum Entropy Econometrics. Wiley, New York, NY.
Grendar, Marian, Judge, George, Schechter, Laura, 2006. An Empirical Non-Parametric Likelihood Family of Data-Based Benford-Like Distributions. Working Paper.
Hill, T.P., 1995. Base-invariance implies Benford's Law. Proceedings of the American Mathematical Society 123, 887–895.
Jaynes, E.T., 1957a. Information theory and statistical mechanics. Physical Review 106 (4), 620–630.
Jaynes, E.T., 1957b. Information theory and statistical mechanics II. Physical Review 108 (4), 171–190.
Kullback, Solomon, 1959. Information Theory and Statistics. J. Wiley and Sons, New York.
Ley, Eduardo, 1996. On the peculiar distribution of the US stock indexes' digits. The American Statistician 50 (4), 311–313.
Mandelbrot, B.B., 1982. The Fractal Geometry of Nature. Freeman, New York.
Mittelhammer, Ron, Judge, George G., Miller, Douglas J., 2000. Econometric Foundations. Cambridge University Press, New York, NY.
Newcomb, Simon, 1881. Note on the frequency of use of the different digits in natural numbers. American Journal of Mathematics 4 (1/4), 39–40.
Owen, Art B., 2001. Empirical Likelihood. Chapman & Hall/CRC.
Pietronero, L., Tosatti, E., Tosatti, V., Vespignani, A., 2001. Explaining the uneven distribution of numbers in nature: The Laws of Benford and Zipf. Physica A: Statistical Mechanics and its Applications 293 (1–2), 297–304.
Pinkham, Roger S., 1961. On the distribution of first significant digits. The Annals of Mathematical Statistics 32, 1223–1230.
Raimi, Ralph A., 1976. The first digit problem. The American Mathematical Monthly 83 (7), 521–538.
Read, Timothy R.C., Cressie, Noel, 1988. Goodness-of-Fit Statistics for Discrete Multivariate Data. Springer-Verlag.
Rodriguez, Ricardo J., 2003. Reducing false alarms in the detection of human influence on data. Journal of Accounting, Auditing & Finance 19 (2), 141–158.
Rodriguez, Ricardo J., 2004. First significant digit patterns from mixtures of uniform distributions. The American Statistician 58 (1), 64–71.
Shannon, Claude E., 1948. A mathematical theory of communication. Bell System Technical Journal 27, 379–423.
Stigler, George J., 1945. The distribution of leading digits in statistical tables. Written 1945–1946. Stigler's 1975 address was Haskell Hall. University of Chicago, Chicago, Illinois 60637.
Zipf, G.K., 1949. Human Behavior and the Principle of Least Effort. Addison-Wesley Press, Inc.