Department of Agricultural & Resource Economics, UCB

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

Joanne Lee [*]   Wendy K. Cho [†]

George G. Judge [‡]

[*]University of California, Berkeley

[†]University of Illinois, Urbana Champaign

[‡]University of California, Berkeley and Giannini Foundation

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

Abstract

In 1881, Newcomb conjectured that the first significant digits (FSDs) of numbers in statistical tables would follow a logarithmic distribution with the digit "1" occurring most often. However, because Newcomb's proposal was not presented with a theoretical basis, it was not given much attention. Fifty-seven years later, Benford argued for the same principle and showed it was relevant to a large range of data sets, and the logarithmic FSD distribution became known as "Benford's Law." In the mid-1940s, Stigler claimed Benford's Law contained a theoretical inconsistency and supplied an alternative derivation for the distribution of FSDs. In this paper, we examine the theoretical basis of the Stigler distribution and extend his reasoning by incorporating FSD first moment information and information-theoretic methods.

Stigler's Approach to Recovering the Distribution of First Significant Digits in Natural Data Sets

Joanne Lee[*]

Wendy K. Tam Cho

George G. Judge

Draft: January 19, 2009

Abstract

In 1881, Newcomb conjectured that the first significant digits (FSDs) of numbers in statistical tables would follow a logarithmic distribution with the digit "1" occurring most often. However, because Newcomb's proposal was not presented with a theoretical basis, it was not given much attention. Fifty-seven years later, Benford argued for the same principle and showed it was relevant to a large range of data sets, and the logarithmic FSD distribution became known as "Benford's Law." In the mid-1940s, Stigler claimed Benford's Law contained a theoretical inconsistency and supplied an alternative derivation for the distribution of FSDs. In this paper, we examine the theoretical basis of the Stigler distribution and extend his reasoning by incorporating FSD first moment information and information-theoretic methods.

Keywords: Benford's Law, Stigler's Law, Power Law, Maximum Entropy, Distance Measures

---

[*]Joanne Lee is a doctoral student in the Graduate School at University of California, Berkeley, CA 94720. Email: jlee@are.berkeley.edu. Wendy K. Tam Cho is Associate Professor in the Departments of Political Science and Statistics and Senior Research Scientist at the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign. Email: wendycho@uiuc.edu. George Judge is Professor in the Graduate School at University of California, Berkeley, CA 94720. Email: judge@are.berkeley.edu.

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, "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):

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

where $P(d)$ is the frequency of the digit d as a first significant digit. The resulting monotonically decreasing 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 began to test Newcomb's hypothesis empirically 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). This FSD phenomenon was then named "Benford's Law" after its popularizer rather than its discoverer (Raimi, 1976). Benford's Law has been shown to approximately apply to a surprisingly large number of data sets, including populations of cities, street addresses of the first 348 persons named in *American Men of Science* (1934), electricity usage, word frequency, ebaY bids, census statistics, campaign donations, and the daily returns to the Dow Jones Industrial Average (Benford, 1938; Raimi, 1976; Zipf, 1949; Hill, 1995; Giles, 2006; Ley, 1996; Cho and Gaines, 2007).

It was another 57 years before Hill (1995), using a base-invariance argument, became the first to rigorously prove Benford's Law. Prior to Hill, others had only suggested possible explanations for the phenomenon. For instance, 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. However, as Raimi (1976) noted, Benford's mixture scheme would be arbitrary and approximate. If Benford's argument were true, 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 related to describing distributions of first significant digits. Minimally, George Stigler, a future Nobel Laureate in Economics, claimed that the specific mixture of uniform distributions with non-uniformly distributed maximum values is an inconsistency. This obser-

vation 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.

More recently, power-law and information-theoretic methods have been proposed as being more intuitively appealing and generalizable ways of determining similar FSD distributions (Grendar, Judge and Schechter, 2006). Pietronero et al. (2001) suggest that Benford's Law is a special case of Zipf's Law, which claims that all rankings of natural processes by size follow power laws. For example, word frequencies have such a distribution—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). In this fashion, probability of occurrence is inversely proportional to its rank. The argument that Zipf's Law is a generalization of Benford's Law is based on the scale-invariant nature of both laws' respective applications, but since Zipf's Law is simply an empirical observation of a family of distributions, the claim that Zipf's Law justifies Benford's Law is intriguing but hardly rigorous.

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 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 frequency of d as a leading significant digit is

$$F_d = \frac{d \ln(d) - (d+1) \ln(d+1) + 3.55843}{9}. \tag{2}$$

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 $[10^2, 10^2 + (2-1)(10^2)) = [100, 200)$. Hence, the lower limit of the proportion of FSDs that are d is

$$\frac{10^r/9}{10^r + (d-1)10^r}. \tag{3}$$

The expectation P_1 of the proportion of FSDs that are d in the interval $[10^r, d10^r)$ is

$$P_1 = \frac{1}{(d-1)10^r} \int_0^{(d-1)10^r} \frac{10^r/9}{10^r + n} dn = \frac{1}{9(d-1)} \ln d. \tag{4}$$

FSDs in the next 10^r numbers are all d, so the proportion of d becomes

$$\frac{10^r/9 + 10^r}{10^r + (d-1)10^r + 10^r}. \tag{5}$$

The expectation P_2 of the proportion of FSDs that are d in the interval $[d10^r, (d+1)10^r)$ is then

$$P_2 = \frac{1}{10^r} \int_0^{10^r} \frac{10^r/9 + n}{10^r + (d-1)10^r + n} dn = 1 - \left(d - \frac{1}{9}\right) \ln \frac{d+1}{d}. \tag{6}$$

Finally, during the last $(9-d)10^r$ numbers of the rth cycle, there are no numbers with FSDs that are d, so the proportion of FSDs that are d will decrease to approximately $\frac{1}{9}$. Hence, the average proportion in the interval $[(d+1)10^r, 10^{r+1})$ will be

$$P_3 = \frac{1}{(9-d)10^r} \int_0^{(9-d)10^r} \frac{10^r/9 + 10^r}{(d+1)10^r + n} dn = \frac{10}{9(9-d)} \ln \left(\frac{10}{d+1}\right). \tag{7}$$

Thus, by Stigler's proposed alternative to Benford's Law, the average proportion of d over the rth cycle is

$$\frac{(d-1)P_1 + P_2 + (9-d)P_3}{9} = \frac{1}{9} \left[d \ln(d) - (d+1) \ln(d+1) + 3.55843\right], \tag{8}$$

where the constant 3.55843 is the mean, $m = \sum_{i=1}^9 d_i p_i$, of the Stigler FSD distribution. Consequently,

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

Solving for m gives us

$$m = \frac{\sum_{i=1}^9 i^2 \ln(d_i) - d_i(d_i + 1) \ln(d_i + 1)}{9 - \sum_{i=1}^9 d_i}. \tag{10}$$

The resulting frequencies from Benford's Law and Stigler's alternative are presented in Table 1.

Table 1: Comparison of Benford and Stigler Distributions

FSD	Benford's Law	Stigler's Law
1	0.301	0.242
2	0.176	0.183
3	0.125	0.146
4	0.097	0.118
5	0.079	0.095
6	0.067	0.077
7	0.058	0.061
8	0.051	0.047
9	0.046	0.034

Stigler claims that the difference between his alternative and Benford's Law arises because of the hidden assumptions made by Benford about the frequencies of the largest numbers in given statistical tables. Benford assumed that smaller numbers with corresponding smaller FSDs occurred more often as bounds for statistical tables. In contrast, Stigler's assumption is that the largest entries in statistical tables were equally likely to begin with $d = 1, 2, \ldots, 9$. 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, 1945; Raimi, 1976; Rodriguez, 2004).[1] 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 (Stigler, 1945). Because no logarithmic FSD distribution holds generally for all natural data sets, Stigler's Law and Benford's Law can be viewed as members of a family of monotonically decreasing distributions of FSDs.

3   Connections to the Power Law and Zipf's 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, when either the underlying data distribution, $P(D)$, or its FSD counterpart, $P(d)$, is multiplied by a constant s, an identical outcome results (Mandelbrot, 1982). Pietronero et al. (2001) note that scale invariance leads to the functional relation

$$P(sD) = P(D^*) = K(p)P(D), \tag{11}$$

---

[1]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.

and that the general solution to (11) has the power law nature

$$P(D^*) = P(D^{*-\alpha}) = s^{-\alpha} D^{-\alpha}. \tag{12}$$

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), we can write for $P(d)$ that, for $\alpha \neq 1$,

$$P(D^*) = \int_{\alpha}^{\alpha+1} D^{-\alpha} \, dD = \frac{1}{1-\alpha}[(d+1)^{1-\alpha} - d^{1-\alpha}]. \tag{13}$$

For $d = 1$,

$$P(D^*) = \int_{d}^{d+1} D \, dD = \int_{d}^{d+1} d\,(\log D) = \log\left(\frac{d+1}{d}\right). \tag{14}$$

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. For values of $\alpha < 1$, the FSD distribution is tilted toward a uniform FSD distribution. Pietronero et al. (2001) call this family of power laws a generalized Benford law.

Zipf's Law is an instance of a rank order statistic, is scale invariant, and is applicable to a large range of phenomena, including income distributions, city sizes, and word frequency (Pietronero et al., 2001; Raimi, 1976; Zipf, 1949). We are particularly interested in the connection between Benford's Law and Zipf's Law. Following Pietronero et al. (2001) in analyzing the rank-order properties of a set of numbers extracted from a general distribution, $P(N) \sim N^{\alpha}$, if a maximum number, N_{max}, corresponds to the rank $k = 1$ and the rank N_k is given by all the numbers between N_k and N_{max}, then

$$k = \int_{N(k)}^{N_{\text{max}}} P(N) \, dN \sim N(k)^{1-\alpha}. \tag{15}$$

Inverting (15) gives us

$$N(k) \sim k^{\frac{1}{1-\alpha}}, \tag{16}$$

which highlights a link between the Benford ($\alpha = 1$) and Zipf's Laws. $\frac{1}{1-\alpha}$. Benford's and Zipf's Laws are examples of scale-invariant distributions, but not of the same type.

4 Problem Reformulation and Solution

In the previous section, we discussed the Benford, Stigler, and power law approaches to determining the distribution of FSDs and investigated the connection of these approaches to one another. We now discuss how information theoretic methods also produce similar distributions. One feature of information theoretic

methods that is absent in these other approaches is the 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 might be helpful.

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:

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

Assuming that the only information that exists is this first-moment information, we are faced with the inverse problem 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. From an information recovery standpoint, the resulting inverse problem is ill-posed. Consequently, there exist an infinite number of possible discrete probability distributions with $\bar{d} \in [1, 9]$.

Based only on the information $\sum_{j=1}^{9} d_j p_j = \bar{d}$, $\sum_{j=1}^{9} p_j = 1$, and $0 \leq p_j \leq 1$, 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 use sample based information recovery methods without having to choose a parametric family of probability densities, on which to base the FSD function. In other words, we seek a way to reduce the infinite dimensional nonparametric problem to a finite dimensional one.

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):

$$I(\mathbf{p}, \mathbf{q}, \gamma) = \frac{1}{\gamma(1 + \gamma)} \sum_{j=1}^{9} \left(p_j \left[\left(\frac{p_j}{q_j} \right)^{\gamma} - 1 \right] \right), \tag{18}$$

where γ is an arbitrary and unspecified parameter.

In the context of recovering the unknown FSD distribution, use of the CR criterion (18) suggests we seek, given $\mathbf{q}$, a solution to the following extremum problem:

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} \left[I(\mathbf{p}, \mathbf{q}, \gamma) \,\middle|\, \sum_{j=1}^{9} p_j d_j = \bar{d}, \sum_{j=1}^{9} p_j = 1, p_j \geq 0 \right]. \tag{19}$$

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, Judge and Miller (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 (1957*a,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. Instead, the solution must be numerically determined via a computer optimization algorithm.

4.2 ME formulation

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

$$H(\mathbf{p}) = -\sum_{j=1}^{9} p_j \ln(p_j), \tag{20}$$

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

$$\bar{d} = \sum_{j=1}^{9} p_j d_j, \tag{21}$$

and the condition that the probabilities must sum to one

$$\sum_{j=1}^{9} p_j = 1. \tag{22}$$

Table 2: Estimated Maximum Entropy (ME) Distributions (with a uniform reference distribution) for the Digit Problem and their Correlation with Stigler's Distribution

FSD Mean	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{p}_5$	$\hat{p}_6$	$\hat{p}_7$	$\hat{p}_8$	$\hat{p}_9$	$H(\hat{\mathbf{p}})$	Corr
2.0	0.496	0.251	0.126	0.064	0.032	0.016	0.008	0.004	0.002	1.38	0.943
3.0	0.306	0.217	0.153	0.108	0.077	0.054	0.038	0.027	0.019	1.88	0.994
3.44	0.250	0.194	0.150	0.117	0.090	0.070	0.054	0.042	0.033	2.01	0.999
3.55	0.238	0.188	0.149	0.118	0.093	0.074	0.058	0.046	0.036	2.03	0.999
4.0	0.191	0.163	0.140	0.120	0.103	0.088	0.075	0.065	0.055	2.12	0.995
4.5	0.148	0.137	0.127	0.118	0.109	0.101	0.094	0.087	0.081	2.18	0.985
5.0	0.111	0.111	0.111	0.111	0.111	0.111	0.111	0.111	0.111	2.20	0.000
5.5	0.081	0.087	0.094	0.101	0.109	0.118	0.127	0.137	0.148	2.18	-0.941

The Lagrangian for the extremum problem is

$$L = -\sum_{j=1}^{9} p_j \ln(p_j) + \lambda \left(\bar{d} - \sum_{j=1}^{9} p_j d_j \right) + \eta \left(1 - \sum_{j=1}^{9} p_j \right). \tag{23}$$

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

$$\hat{p}_i = \frac{\exp(-d_i \hat{\lambda})}{9 \sum_{j=1}^{9} \exp(-d_j \hat{\lambda})}, \tag{24}$$

for the jth outcome. In this context, the $\hat{p}_i$ are exponentially FSD and the chosen FSD distribution is the one that happens in the most likely way (multiplicity). We note again that $p(\lambda)$ is a member of a canonical exponential family with mean

$$\bar{d} = \sum_{j=1}^{9} p_j(\lambda) d_j, \tag{25}$$

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

$$I(\lambda) = \sum_{j=1}^{9} p_j(\lambda) d_j^2 - \left(\sum_{j=1}^{9} p_j(\lambda) d_j \right)^2 = \mathrm{Var}(d). \tag{26}$$

4.3   Some Mean-Related ME Distributions

Using the ME uniform reference distribution formulation and solution, the resulting distributions for a range of FSD means (including the Stigler mean of 3.55 and Benford mean of 3.44) are presented in Table 2. Table 3 shows similar results when the reference distribution is the Stigler distribution. From the table 2, we can see that 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 Benford FSD mean, 3.44, yields a monotonically decreasing ME distribution

Table 3: The Estimated Maximum Entropy (ME) Distributions (with a Stigler FSD reference distribution) for the Digit Problem and their Correlation with Stigler's Distribution

FSD Mean	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{p}_5$	$\hat{p}_6$	$\hat{p}_7$	$\hat{p}_8$	$\hat{p}_9$	Corr
2.0	0.503	0.243	0.124	0.064	0.033	0.017	0.009	0.004	0.002	0.939
3.0	0.312	0.211	0.150	0.108	0.078	0.056	0.040	0.027	0.018	0.994
3.44	0.255	0.188	0.147	0.116	0.092	0.073	0.056	0.043	0.030	0.999
3.55	0.243	0.183	0.146	0.118	0.094	0.076	0.060	0.046	0.033	0.999
4.0	0.194	0.159	0.137	0.120	0.104	0.091	0.078	0.065	0.051	0.996
4.5	0.150	0.133	0.124	0.117	0.111	0.105	0.098	0.088	0.075	0.973
5.0	0.112	0.107	0.108	0.111	0.113	0.116	0.116	0.113	0.103	-0.059
5.5	0.081	0.084	0.091	0.101	0.111	0.123	0.133	0.140	0.138	-0.948

similar to the Benford logarithmic distribution. Its correlation with the Stigler distribution is 0.96. Under ME, the exponential null hypotheses that result have two especially appealing properties. First, the result is achieved while adhering to the principles of Occam's razor (minimizing underlying assumptions). The second appealing aspect of this criterion choice is maximum multiplicity. In the absence of assumptions, which distribution among the possible distributions is the best choice? The answer must be the one that occurs most frequently, i.e. the choice with maximum multiplicity, which is the result of this information theoretic method.

5 Discussion

Benford's Law has been shown to be applicable to a large set of seemingly unrelated phenomenas from the area of rivers to stock market prices 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, Hillison and Pacini, 2004). Some appear to be related to Stigler's Law. Others follow the outlines of the Power Law and Zipf's Law. Each law appears to fit certain contexts well, but may not apply to other data 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 is provided 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.

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 and Brian J. Gaines. 2007. "Breaking the (Benford) Law: Statistical Fraud Detection and Campaign Finance." *The American Statistician* 61(3):218–223.

Cressie, Noel and Timoth R C Read. 1984. "Multinomial goodness of fit tests." *Journal of the Royal Statistical Society, Series B* 46:440–464.

Durtschi, Cindy, William Hillison and Carl Pacini. 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, George Judge and Douglas Miller. 1996. *Maximum Entropy Econometrics*. Wiley, New York, NY.

Grendar, Marian, George Judge and Laura Schechter. 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. 1957*a*. "Information Theory and Statistical Mechanics." *Physical Review* 106(4):620–630.

Jaynes, E T. 1957*b*. "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 U.S. Stock Indexes' Digits." *The American Statistician* 50(4):311–313.

Mandelbrot, B. B. 1982. *The Fractal Geometry of Nature*. New York: Freeman.

Mittelhammer, Ron, George G Judge and Douglas J Miller. 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., E. Tosatti, V. Tosatti and A. Vespignani. 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.

Raimi, Ralph A. 1976. "The First Digit Problem." *The American Mathematical Monthly* 83(7):521–538.

Read, Timothy R C and Noel Cressie. 1988. *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer-Verlag.

Rodriguez, Ricardo J. 2004. "First Significant Digit Patterns from Mixtures of Uniform Digits." *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.

A 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,

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

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

$$
\begin{aligned}
P(FSD = k) &= \int_{10^r}^{k10^r} \frac{10^r}{9b} dF(b) + \int_{k10^r}^{(k+1)10^r} \left(\frac{10^r}{9b} + \frac{b - k10^r}{b} \right) dF(b) + \int_{(k+1)10^r}^{10^{r+1}} \frac{10^{r+1}}{9b} dF(b) \\
&= \frac{1}{9 \cdot 10^r} \left(\frac{10^r}{9} \int_{10^r}^{(d+1)10^r} \frac{db}{b} + \int_{d10^r}^{(d+1)10^r} db - d10^r \int_{d10^r}^{(d+1)10^r} \frac{db}{b} + \frac{10^{r+1}}{9} \int_{(k+1)10^r}^{10^{r+1}} \frac{db}{b} \right) \\
&= \frac{1}{9 \cdot 10^r} \left[\frac{10^r}{9} \ln(k + 1) + 10^r - k10^r \ln \left(\frac{k + 1}{k} \right) + \frac{10^{r+1}}{9} \ln \left(\frac{10}{k + 1} \right) \right] \\
&= \frac{1}{9} \left(1 + \frac{10}{9} \ln(10) + k \ln(k) - (k + 1) \ln(k + 1) \right) \tag{A-2}
\end{aligned}
$$