Saint Mary's College of California

Department of Mathematics

Senior Essay

Benford's Law

Author:
Aaron Lenci

Mentor:
Professor Charles Hamaker

Supervisor:
Professor Kathy Porter

May 7, 2017

History:

Benford's Law dates back to 1881, when Simon Newcomb (1835-1909), an astronomer who was in the U.S. Navy, recognized "that the ten digits do not occur with equal frequency" after making use of logarithmic tables so often and noticed "how much faster the first pages wear out than the last ones" [5]. He saw that the first digit was most often a number 1 and that the frequencies decreased monotonically to 9. Then about fifty-seven years later, Frank Benford (1883-1948), a physicist at General Electric, rediscovered Newcomb's work and provided over 20,000 entries from twenty different tables that supported Newcomb's findings. The tables included data of "catchment areas of 335 rivers, specific heats of 1,389 chemical compounds, American League Baseball statistics, and numbers gleaned from front pages of newspapers and *Reader's Digest* articles" (Benford Book). However, evidence has been found that shows Benford manipulated his data to gain a better fit for each table. In any case, Benford's unmanipulated data is still very close to following the law. Benford's findings nevertheless gained a lot of attention which made Newcomb's article overlooked and as a consequence, the law became known as *Benford's law* [2].

Introduction: To understand Benford's Law, consider the populations of $3,142$ United States counties based on 2016 census data. How many county populations would you expect to start with the digit 1? What about 2? Or what about 9? You might reasonably expect that all nine possible leading digits occur equally often as the first digit. This leads us to the concept of Benford's Law. *Benford's Law* addresses special cases of data sets where the leading digit of a number is not equally likely to be any of the nine possible digits 1, 2, 3, . . ., 9, but is 1 more than 30% of the time, and is 9 less than 5% of the time, with the probabilities decreasing monotonically in between. More precisely, the exact values of Benford's Law are given by Table 1. As well, the correct terms for the leading digit of a number are given by the following two definitions:

Percentage of Times the Natural Numbers 1 to 9 are Used as First Digits in Numbers, as Determined by 20,229 Observations

Group	Title	First Digit									Count
		1	2	3	4	5	6	7	8	9	
A	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
B	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
C	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
D	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
E	Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
F	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
G	H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
H	Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
I	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
J	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
K	$n^{-1}, \sqrt{n}, \cdots$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
L	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
M	*Digest*	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
N	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
O	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
P	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Q	Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
R	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
S	$n^1, n^2 \cdots n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
T	Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average.......		30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Probable Error		± 0.8	± 0.4	± 0.4	± 0.3	± 0.2	± 0.2	± 0.2	± 0.2	± 0.3	—

Figure 1: Tables Found by Frank Benford [2]

Def. First Significant Decimal Digit. [2] The *First Significant Decimal Digit* of a positive real number x is the first non-zero digit appearing in the decimal expansion of x. More formally, for every non-zero real number x, the *first significant digit* of x, denoted by $D_1(x)$, is the unique integer $d \in \{1, 2, ..., 9\}$ satisfying $10^k d \leq |x| < 10^k(d+1)$ for some $k \in \mathbb{Z}$. Throughout the paper, the first significant digit will be denoted by d.

Example. The first significant digit of 2017, 2.017, and .02017 is 2, i.e.,

$$D_1(2017) = D_1(2.017) = D_1(.02017) = 2$$

Def. Significand. [2] The *Significand* of a real number is its coefficient when expressed in scientific notation.

More formally, the *significand function* $S : \mathbb{R} \to [1, 10)$ is defined as follows: If $x \neq 0$ then $S(x) = t$, where t is the unique number in $[1, 10)$ with $|x| = 10^k t$ for some $k \in \mathbb{Z}$; if $x = 0$ then, for convenience, $S(0) := 0$.

$$S(x) = 10^{\log|x| - \lfloor \log|x| \rfloor} \text{ for all } x \neq 0$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to x.

Example. 2017 expressed in scientific notation is $2.017 \cdot 10^3$. So its significand is 2.107, i.e.,

$$S(2017) = 2.017$$

The significand is useful because it will get rid of the ambiguity that results from looking at the first significant digits. For example, if we look at numbers like 19.99... and 1.99..., the first significant digit is clearly 1. However, 19.99... is really close to 20. So the significand of 19.99... = 20 and 1.99... = 2. So this clearly changes our first digit to a 2 instead of 1 which can greatly alter the data sets if many numbers are very close to the next significant digit.

First Digit	1	2	3	4	5	6	7	8	9
Percent of Occurence	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6

Table 1: Probability of each first digit according to Benford's Law

First Digit	1	2	3	4	5	6	7	8	9
Number of Counties	958	586	385	301	232	204	164	168	144
Percent of Counties	30.49	18.65	12.25	9.58	7.38	6.49	5.22	5.35	4.58

Table 2: Percent of U.S. Counties according to the first digit of their population size (Significant Digit) [1]

So with that brief overview of Benford's Law, let us look back at the data for the $3,142$ counties, which is given by Table 2. Clearly, the first digits do not appear equally among all nine possibilities. With that said, the first digits do not follow Benford's Law perfectly since 8 occurs more often than 7 among other minor differences. However, the data set is still very close to Benford's Law. Then looking at the significand for the county populations, we get Figure 3. For the signicand, we only rounded county populations up to the next significant digit if the decimal was greater than or equal to 0.995. So we can see that in Figure 3, the first digit of 1 became closer to Benford's Law, but consequently, 2 is now further from Benford's Law. There were other minor changes as well, but none were greater than the change between 1 and 2. However, we still come out to a distribution that follows very closely to Benford's Law and the exact number of counties for the significand is given by Table 3.

To grasp why Benford's Law exists, let us start with a simple explanation before moving into a more mathematical explanation. A way to make sense of Benford's Law is to think about the growth of money starting at $\$1,000$. At an annual 20% increase, it will take 5

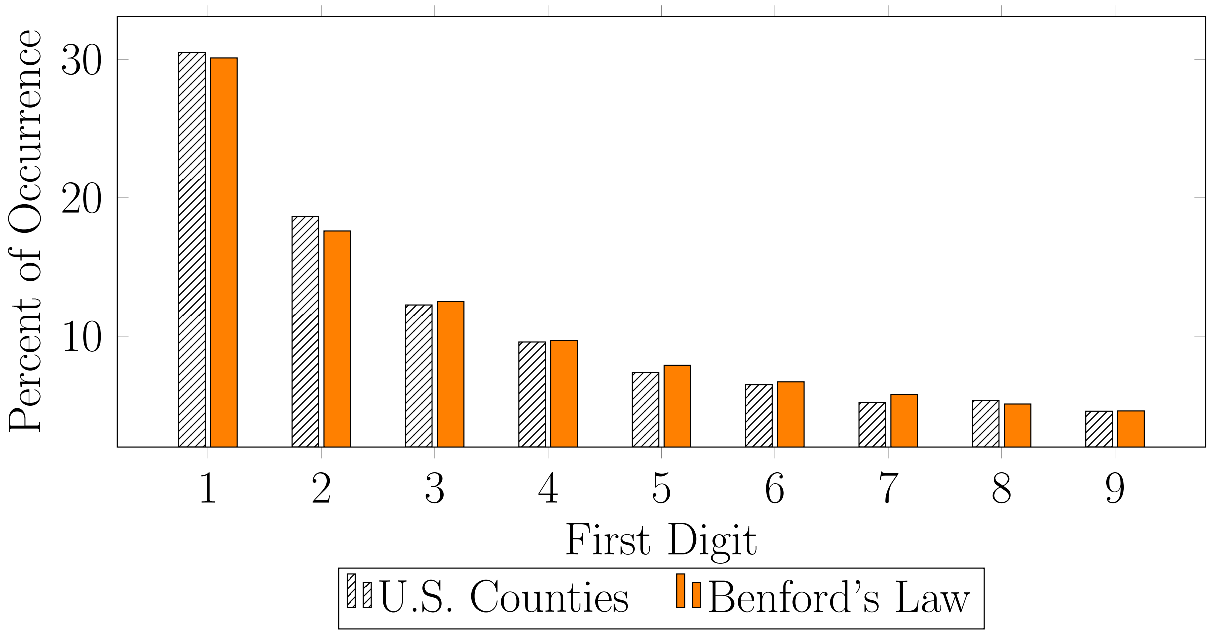

Figure 2: U.S. County Populations vs. Benford's Law

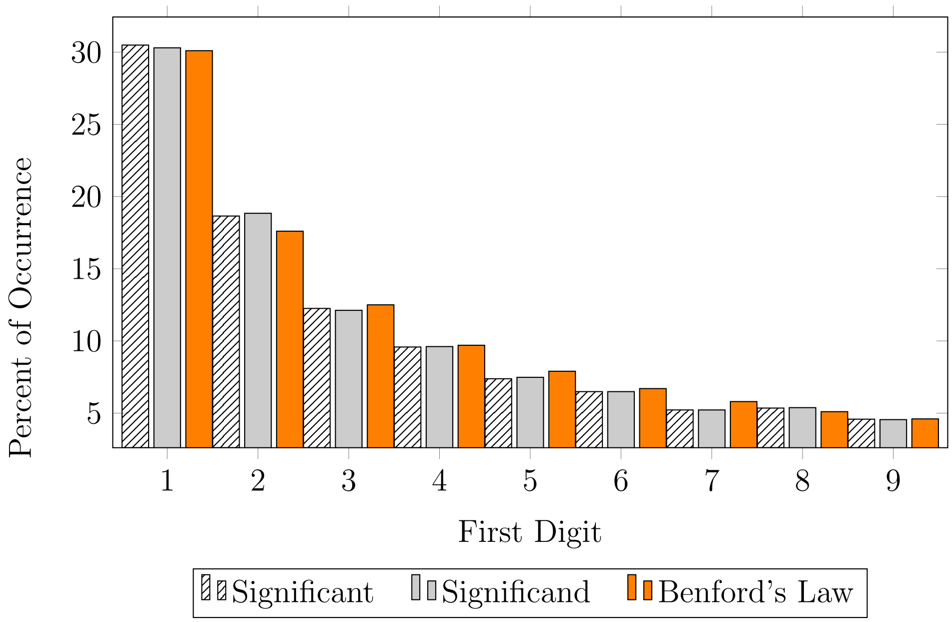

Figure 3: U.S. County Populations with Significand vs. Benford's Law

First Digit	1	2	3	4	5	6	7	8	9
Number of Counties	952	592	381	302	235	204	163	169	144
Percent of Counties	30.3	18.84	12.12	9.61	7.48	6.49	5.22	5.38	4.55

Table 3: Percent of U.S. Counties according to the first digit of their population size (Significand) [1]

years to increase from $\$1,000$ to $\$2,000$. So during the whole 5 years, the first digit is always 1. Then as the money continues to increase at 20%, it will take about 2.5 more years to get $\$3,000$ and about another 1.67 more years after that to get $\$4,000$. This trend continues until it takes about 7 months to go from $\$9,000$ to $\$10,000$. Then from here, the trend cycles back around to the beginning where it will take another 5 years to get to $\$20,000$. Thus, the first digit is going to be 1 the majority of the time since we spend 5 years at 1 during one cycle, then it will be 2 a slightly less amount of time, and so on until 9 appears for the shortest amount of time.

Understanding Benford's Law:

Now with a general understanding of Benford's Law, lets start breaking down why Benford's Law exist and when it applies. To begin with, we state the formal definition of Benford's Law.

Def. Benford's Law. [2] *Benford's Law*, also known as the *First-Digit* or *Significant-Digit Law*, is a case where the first significant digit is not equally likely to be any of the nine possible digits 1, 2, 3, . . ., 9, but is 1 more than 30% of the time, and is 9 less than 5% of the time, with the probabilities decreasing monotonically in between and satisfies the following for the first significant digit D_1, where the exact values are given on Table 1,

$$\text{Prob}[D_1 = d] = \log_{10}(1 + \tfrac{1}{d}) \qquad \text{for all } d = 1, 2, ..., 9.$$

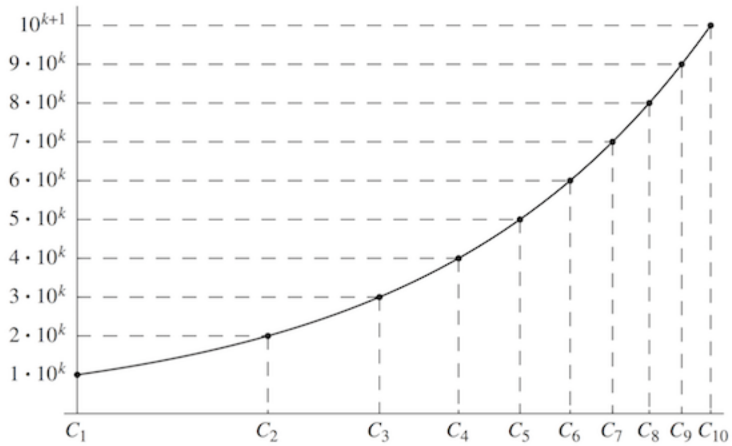

Figure 4: $P(d) = d \cdot 10^k$ [4]

Now, to arrive at the probability function for Benford's Law we will look at the function $P(n) = ar^n$ for $n \in \mathbb{Z}$. Figure 4 depicts the curve and the points $(d, P(d))$ and we would expect that the proportion of points that would land in the interval $[C_d, C_{d+1})$ would be about $\log_{10}(d+1) - \log_{10} d = \log_{10}(1 + \frac{1}{d})$ since each interval is $[d \cdot 10^k, (d+1) \cdot 10^k]$, i.e., the probability of the first digit according to Benford's Law. The first theorem that we will prove is a good model for Benford's Law if we are looking at a specific data set over a long period of time, i.e., population of a county over a long period of time.

Theorem. 1. [4] Suppose that $1 \leq a < 10$, $1 < r < 10$, and that r is rational. Then

$$\lim_{n \to \infty} \frac{1}{n} |\{k \leq n : D(ar^k) = d\}| = \log_{10}(d+1) - \log_{10} d = \log_{10}(1 + \frac{1}{d}),$$

for $d = 1, 2, ..., 9$.

Proof. [4] Let $P(n) = ar^n$ such that $n \in \mathbb{N}$. Suppose that $1 \leq a < 10$, $1 < r < 10$ and that r is rational. Let $g(x) := \log_{10}(x) \bmod 1$ for $x \in [1, \infty)$. Then g is a function defined by $g : [0, \infty) \to [0, 1)$. Notice that $\log_{10}(x) \bmod 1$ results in the decimal portion of $\log_{10}(x)$, e.g., $1.23 \equiv .23 \bmod 1$.

Then for $x = ar^k$, $g(ar^k) = (\log_{10} a + k \log_{10} r) \bmod 1$ for all k by the properties of logarithms. First, for simplicity, assume $a = 1$, so $g(r^k) = k \log_{10} r \bmod 1$ for all k. Observe that $D(x) = d$ if and only if $d \cdot 10^j \leq x < (d+1) \cdot 10^j$ for some integer j and first significant digit d of a number x. This statement is true since we can look at Figure 4 and notice that each number must fall within the interval $[d \cdot 10^j, (d+1)10^j]$. Then applying log to all terms in the inequality, $D(x) = d$ if and only if $\log_{10} d \leq \log_{10} x \bmod 1 < \log_{10}(d+1)$. Hence,

$$D(x) = d \text{ if and only if } \log_{10} d \leq g(x) < \log_{10}(d+1) \tag{1}$$

$$\Rightarrow D(r^k) = d \text{ if and only if } \log_{10} d \leq k \log_{10} r \bmod 1 < \log_{10}(d+1).$$

Then by applying the Uniform Distribution Theorem (see appendix [1]) to (1), we need $x = \log_{10} r$ to be irrational which is true by the following theorem (Appendix [1]).

Theorem. [6] For any positive rational r, $\log_{10} r$ is irrational unless $r = 10^n$ for $n \in \mathbb{Z}$.

Since we are assuming that r is rational and $1 < r < 10$, r will not be an integer power of 10. Hence, $x = \log_{10} r$ is irrational. We can then apply the Uniform Distribution Theorem to the interval $I = [\log_{10} d, \log_{10}(d+1)]$ since $g(x)$ must be within I by (1). Since, Length$(I) = \log_{10}(d+1) - \log_{10} d$, we obtain,

$$\lim_{n\to\infty} \frac{1}{n} |\{k \leq n : \log_{10} d \leq k \log_{10} r \bmod 1 < \log_{10}(d+1)\}| = \log_{10}(d+1) - \log_{10} d = \log_{10}(1 + \frac{1}{d})$$

for $d = 1, 2, ..., 9$.

Note: The proof still holds if $a > 1$ because the Uniform Distribution Thoerem still holds if $kx \bmod 1$ is replaced by $(a + kx) \bmod 1$ since $(a + kx) \bmod 1$ is just $kx \bmod 1$ shifted by a. $\square$

The next theorem provides a better model for Benford's Law since it allows us to look at several data sets that can have different initial values and different growth rates. To understand this, we can think about Theorem 1 as looking at only one county and its population change over a long period of time. Now, we are looking at data for many counties with different initial populations and growth rates. However, before we start the proof, we must define a measurable function and induced measures.

Def. Measurable Function. [2] A function $f : X \to \mathbb{R}$ is *measurable* if, for every real number a, the set

$$\{x \in X : f(x) > a\}$$

is measurable. Or simply, a measurable function is a function where we assign probabilities to its values.

As well, a measurable space is a space where we can find the probabilities of the sets within the space.

Def. Induced Measures. [2] If ϕ is a measurable function from a measurable space $(X, \mathcal{S})$ to a measurable space $(Y, \mathcal{T})$ where $\mathcal{S} \subset X$ and $\mathcal{T} \subset Y$, then every measure v on X induces a measure on $v \circ \phi^{-1}$ on Y defined by

$$v \circ \phi^{-1}(E) = v(\phi^{-1}(E)) \text{ for measurable sets } E \subset Y.$$

Observe that

$$\textstyle\int_Y f(y) d(v \circ \phi^{-1})(y) = \int_X f(\phi(x)) dv(x) \qquad (10)$$

for bounded measurable functions f on Y.

An induced measure will allow us to make a change of variables, as in integration by substitution, while allowing us to still find the probability on a measurable space.

Theorem. 2. [4] Let a be selected at random from the interval $[1, 10)$ and assume a is uniformly distributed, meaning the values of a are in the interval I with probability, length$(I)/9$. Let r be selected uniformly from a fixed interval $[s, u) \subset [1, 10)$, meaning for each $I \subset [s, u)$, r is selected with probability, length$(I)/(u - s)$. Then,

$$\lim_{n\to\infty} \Pr[D(ar^n) = d] = \log_{10}(d + 1) - \log_{10} d = \log_{10}(1 + \frac{1}{d}),$$

Proof. [4] Let r be selected at random from a fixed interval $[s, u) \in [1, 10)$. Let U be the uniform probability measure on $[s, u)$ such that

$$U(I) = \frac{\text{length}(I)}{u-s} \quad \text{where} \quad I \subset [s, u).$$

Let a be selected at random and independently from r using any probability measure μ on $[1, 10)$, i.e., $\Pr[a \in E] = \mu(E)$. Then like in Theorem 1, we will use the function $g : [1, 10) \to [0, 1)$ that is defined by $g(x) = \log_{10} x$. Then it follows that $D(ar^n) = d$ if and only if $d \cdot 10^k \leq ar^n < (d + 1)10^k$ for some $k \in \mathbb{Z}_0^+$. Hence we have,

$$d \cdot 10^k \leq ar^n < (d + 1)10^k$$

$$\implies \log_{10}(d \cdot 10^k) \leq \log_{10}(ar^n) < \log_{10}((d + 1) \cdot 10^k)$$

$$\implies \log_{10}(d) \leq g(a) + ng(r) < \log_{10}(d + 1)$$

Then we can find the probability of the first digit by,

$$\Pr[D(ar^n) = d] = \int_1^{10} \int_1^{10} \chi_{I_d}(g(a) + ng(r))d\mu(a)dU(r), \tag{11}$$

where χ_{I_d} is an indicator function of the interval $I_d = [\log_{10} d, \log_{10}(d+1))$. Now we need to show that the integral converges to $\log_{10}(d+1)-\log_{10} d = \lambda(I_d)$ where λ is a Lebsegue measure on $[0, 1)$. Then there exists bounded measurable function h on $[1, 10)$, $dU(r) = h(r)dr$, i.e. h is the density function for U. So it is true that,

$$\int_1^{10} f(r)dU(r) = \int_1^{10} f(r)h(r)dr \tag{12}$$

where f is a bounded measurable function on $[1, 10)$. Now let P be the induced measure $U \circ g^{-1}$ on $[0, 1)$ and Q be the induced measure $\mu \circ g^{-1}$ on $[0, 1)$. Then from (10) and (11) where $X = [1, 10)$ and $Y = [0, 1)$

$$\Pr[D(ar^n) = d] = \int_0^1 \int_0^1 \chi_{I_d}(b + nx) dQ(b) dP(x).$$

Then for each n, let P_n be the induced measure $P \circ \phi_n^{-1}$ on $[0, 1)$ such that $\phi_n(x) = nx$ mod 1 for $x \in [0, 1)$. We can use (10) again along with $X = Y = [0, 1)$ to obtain a change of variables from X to Y,

$$\Pr[D(ar^n) = d] = \int_0^1 \int_0^1 \chi_{I_d}(b + \phi_n(x)) dQ(b) dP(x)$$

$$= \int_0^1 \int_0^1 \chi_{I_d}(b + y) dQ(b) dP_n(y).$$

Then by [7.19], we find that this integral is the convolution $Q * P_n$. Hence,

$$\Pr[D(ar^n) = d] = \int_0^1 \chi_{I_d} d(Q * P_n) = (Q * P_n)(I_d).$$

Now we want to show that

$$\lim_{n \to \infty} (Q * P_n)(I_d) = \lambda(I_d).$$

Then similar to the proof for the Uniform Distribution Theorem, we must show

$$\lim_{n \to \infty} \widehat{Q * P_n}(m) = 0 \qquad \text{where } m \neq 0.$$

We will use the Convolution Theorem which states,

Convolution Theorem. [7] Let $f(m)$ and $g(m)$ be arbitrary functions of m with Fourier transforms. Then $\widehat{f * g}(m) = \widehat{f}(m)\widehat{g}(m)$ for all m.

Hence, we have that $\widehat{Q * P_n}(m) = \widehat{Q}(m)\widehat{P_n}(m)$ for all m. So we must show

$$\lim_{n \to \infty} \widehat{P_n}(m) = 0 \qquad \text{where } m \neq 0.$$

Since $\chi_m(\phi_n(x)) = e^{2\pi i m\phi_n(x)} = e^{2\pi i mnx} = \chi_{mn}(x)$, we get

$$\widehat{P_n}(m) = \int_0^1 \overline{\chi_m(x)} dP_n(x) = \int_0^1 \overline{\chi_m(\phi_n(x))} dP(x) = \int_0^1 \overline{\chi_{mn}(x)} dP(x) = \widehat{P}(mn)$$

Then like (12), for $dP = Hd\lambda$ where H is a bounded measurable function on $[0, 1)$,

$$\widehat{H} = \int_0^1 f(x)dP(x) = \int_0^1 f(x)H(x)dx$$

for bounded measurable functions f on $[0, 1)$. Then we can find that $H(x) = \log_e(10)h(10^x)10^x$ since $P = U \circ g^{-1} \implies dP = d(U \circ g^{-1}) = h'(g^{-1})(g^{-1})' = h(10^x)10^x \ln(10)$. Note that $\ln(10) = \log_e(10)$. Since H is integrable,

$$\lim_{n\to\infty} \widehat{H}(n) = 0$$

by the Riemann-Lebesgue Lemma.

Riemann-Lebesgue Lemma. [3] Let f be a Lebesgue function, then

$$\lim_{n\to\infty} \widehat{f}(n) = 0.$$

Therefore, for $m \neq 0$,

$$\lim_{n\to\infty} \widehat{P}_n(m) = \lim_{n\to\infty} \widehat{P}(mn) = \lim_{n\to\infty} \widehat{H}(mn) = 0.$$

□

Now, Theorem 2 can be modeled even better, and more realistically, if we allow the growth rates to vary over time instead of being fixed rates as in the previous two models. So Theorem 3 contains various r_i's to account for different growth rates over a long period of time.

Theorem 3. [4] Let a be selected at random from $[1, 10)$ and suppose each r_i is selected uniformly from some fixed interval $[s, u) \in [1, 10)$. Assume each r_i is selected independently. Then,

$$\lim_{n\to\infty} \Pr[D(ar_1r_2\cdots r_n) = d] = \log_{10}(d+1) - \log_{10} d = \log_{10}(1 + \frac{1}{d}).$$

Now the proof for Theorem 3 follows similarly to Theorem 2, however, in Theorem 3, each r_i will be selected at random from $[1, 10)$ using the uniform probability U and then the a's and each r_i are selected independently of each other. As well, when looking at the convolution, we will look at $Q * P^n$ where P^n is the convolution power $P^n = \underbrace{P * P * \cdots * P}_{\text{n times}}$. Finally, to show that

$$\lim_{n\to\infty} \widehat{P^n}(m) = 0 \qquad \text{for all integers } m \neq 0,$$

we will use the following Lemma.

Lemma 2. [4] For a probability measure P on $G = [0, 1)$, we have

$$P^n \to \lambda \text{ in the } weak - * \text{ (or vague) topology}$$

provided that the support of P is not a subset of any coset of a finite subgroup of G.

For the complete proof, please see [4].

Sequences

Def. Benford Sequence: [2] A sequence $\{x_n\}_{n=1}^{\infty}$ of real numbers is a *Benford Sequence*, or *Benford* for short, if

$$\lim_{N\to\infty} \frac{|\{1 \leq n \leq N : S(x_n) \leq t\}|}{N} = \log t \quad \text{for all } t \in [1, 10).$$

By this definition, we can look at a sequence of real numbers and determine if it is a Benford sequence by taking n to infinity and see if the number of elements in the sequence with first significant digit d equals $\log(1 + d^{-1})$. More simply, a Benford sequence is any sequence of numbers where the first significant digits of the elements of the sequence satisfy Benford's Law.

Fibonacci Numbers:

A famous sequence of numbers known to obey Benford's Law is the Fibonacci Numbers. The Fibonacci Numbers are a sequence of numbers where every number in the sequence is the sum of the previous two numbers with the exception of the first two numbers which are both 1. More precisely,

$$f_{n+1} = f_n + f_{n-1} \qquad \text{where } f_1 = f_2 = 1 \text{ and } n \in \mathbb{N}.$$

To prove that the Fibonacci Numbers are Benford, we must show that the sequence is exponential and then by Theorem 1, we will have that the Fibonacci Numbers are Benford.

Proof. Define the function f by

$$f_{n+1} = f_n + f_{n-1} \qquad \text{where } n \in \mathbb{N}.$$

$$f_0 = 1 \text{ and } f_1 = 1$$

Now writing the function in matrix form we get,

$$\begin{bmatrix} f_{n+1} \\ f_n \end{bmatrix} = \begin{bmatrix} f_n + f_{n-1} \\ f_n \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}}_{A} \begin{bmatrix} f_n \\ f_{n-1} \end{bmatrix} \qquad \text{or} \qquad \overline{x}_{n+1} = A\overline{x}_n$$

Now if $\overline{x}_1$ satisfies $A\overline{x}_1 = \lambda\overline{x}_1$, we have

$$\overline{x}_2 = \lambda\overline{x}_1 \Longrightarrow A\overline{x}_1 = \overline{x}_2$$

Then it is true that

$$\overline{x}_3 = A\overline{x}_2 = A(\lambda\overline{x}_1) = \lambda(A\overline{x}_1) = \lambda(\lambda\overline{x}_1) = \lambda^2\overline{x}_1.$$

Then if we continue this process for $\overline{x}_4, \overline{x}_5, ..., \overline{x}_n$ we find that in general,

$$\overline{x}_{n+1} = \lambda^n \overline{x}_1 \qquad (*)$$

Now we can find the eigenvalue (λ) and eigenvector ($\overline{v}$) satisfying $A\overline{v} = \lambda\overline{v}$. Then if we know λ, we can find $\overline{v}$ by solving for $(A - \lambda I)\overline{v} = 0$ where I is the identity matrix. Then the eigenvalue must satisfy

$$\det(A - \lambda I) = 0.$$

So we have that,

$$\det\left(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{bmatrix}\right) = (1-\lambda)(-\lambda) - 1(1) = \lambda^2 - \lambda - 1 = 0.$$

Then our two solutions are,

$$\lambda = \phi = \tfrac{1+\sqrt{5}}{2} \qquad \text{and} \qquad \lambda = \psi = \tfrac{1-\sqrt{5}}{2}.$$

Notice that $\phi + \psi = 1$ and $\phi \cdot \psi = -1$. So solving $\overline{v}$ where $\lambda = \phi$,

$$(A - \lambda I)\overline{v} = \left(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} - \phi \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 1-\phi & 1 \\ 1 & -\phi \end{bmatrix} \overline{v} = 0.$$

Then through Gaussian elimination,

$$\begin{bmatrix} 1-\phi & 1 \\ 1 & -\phi \end{bmatrix} R_2 \leftrightarrow R_1 : \begin{bmatrix} 1 & -\phi \\ 1-\phi & 1 \end{bmatrix} R_2 - (1-\phi)R_1 : \begin{bmatrix} 1 & -\phi \\ 0 & 1+\phi-\phi^2 \end{bmatrix} = \begin{bmatrix} 1 & -\phi \\ 0 & 0 \end{bmatrix}$$

since $1 + \phi - \phi^2 = 1 + \frac{1+\sqrt{5}}{2} - \left(\frac{1+\sqrt{5}}{2}\right)^2 = \frac{4+2+2\sqrt{5}-1-2\sqrt{5}-5}{4} = 0$.

So we have that x_2 is free, meaning that x_2 does not correspond to a pivot point in the matrix, and $x_1 - \phi x_2 = 0$. Then if $x_2 = 1 \Longrightarrow x_1 = \phi$. This gives,

$$\overline{v}_\phi = \begin{bmatrix} \phi \\ 1 \end{bmatrix} \text{ such that } \lambda = \phi.$$

Similarly we can find,

$$\overline{v}_\psi = \begin{bmatrix} \psi \\ 1 \end{bmatrix} \text{ such that } \lambda = \psi.$$

Since $\{\overline{v}_\phi, \overline{v}_\psi\}$ are linearly independent, we can write x_1 in the following form,

$$\overline{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = a \begin{bmatrix} \phi \\ 1 \end{bmatrix} + b \begin{bmatrix} \psi \\ 1 \end{bmatrix} \qquad \text{or} \qquad \begin{bmatrix} \phi & \psi \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Now solving the matrix using Gaussian elimination,

$$\left[\begin{array}{cc|c} \phi & \psi & 1 \\ 1 & 1 & 1 \end{array}\right] R_2 - \tfrac{1}{\phi} R_1 : \left[\begin{array}{cc|c} \phi & \psi & 1 \\ 0 & 1 - \frac{\psi}{\phi} & 1 - \frac{1}{\phi} \end{array}\right]$$

$$\implies (1 - \tfrac{\psi}{\phi}) b = 1 - \tfrac{1}{\phi} \implies (\phi - \psi) b = \phi - 1$$

$$\implies (\tfrac{1+\sqrt{5}}{2} - \tfrac{1-\sqrt{5}}{2}) b = \tfrac{1+\sqrt{5}}{2} - 1 \implies \sqrt{5} b = \tfrac{-1+\sqrt{5}}{2} \implies b = \tfrac{-1}{\sqrt{5}} \psi.$$

Since $a + b = 1$, $a = 1 - (-\frac{1}{\sqrt{5}}\psi) = \frac{\sqrt{5}+\psi}{\sqrt{5}} = \frac{\frac{2\sqrt{5}}{2} + \frac{1-\sqrt{5}}{2}}{\sqrt{5}} = \frac{1}{\sqrt{5}} \frac{1+\sqrt{5}}{2} = \frac{\phi}{\sqrt{5}}$.

Hence,

$$\overline{x}_1 = \begin{bmatrix} f_1 \\ f_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \tfrac{\phi}{\sqrt{5}} \begin{bmatrix} \phi \\ 1 \end{bmatrix} + \tfrac{-\psi}{\sqrt{5}} \begin{bmatrix} \psi \\ 1 \end{bmatrix} = \tfrac{1}{\sqrt{5}} \begin{bmatrix} \phi^2 \\ \phi \end{bmatrix} - \tfrac{1}{\sqrt{5}} \begin{bmatrix} \psi^2 \\ \psi \end{bmatrix}.$$

Then by $(*)$,

$$\overline{x}_2 = A\overline{x}_1 = \tfrac{1}{\sqrt{5}}(A \begin{bmatrix} \phi^2 \\ \phi \end{bmatrix} - A \begin{bmatrix} \psi^2 \\ \psi \end{bmatrix}) = \tfrac{1}{\sqrt{5}}(\begin{bmatrix} \phi^3 \\ \phi^2 \end{bmatrix} - \begin{bmatrix} \psi^3 \\ \psi^2 \end{bmatrix}) = \begin{bmatrix} f_2 \\ f_1 \end{bmatrix}.$$

Thus, $f_2 = \frac{\phi^3 - \psi^3}{\sqrt{5}}$ and if we continue the process for $\overline{x}_3, \overline{x}_4, ..., \overline{x}_n$, we will find that in general,

$$\overline{x}_n = \tfrac{1}{\sqrt{5}}(\begin{bmatrix} \phi^{n+1} \\ \phi^n \end{bmatrix} - \begin{bmatrix} \psi^{n+1} \\ \psi^n \end{bmatrix}).$$

This means that the nth number in the Fibonacci sequence is given by,

$$f_n = \frac{\phi^{n+1} - \psi^{n+1}}{\sqrt{5}}.$$

Note that $\psi = \frac{1-\sqrt{5}}{2} \approx -.618$. Then

$$\lim_{n \to \infty} \psi^n = 0$$

and for n large,

$$f_n \approx \frac{\phi^{n+1}}{\sqrt{5}}.$$

Thus, the Fibonacci sequence is asymptotically exponential. Therefore, by Theorem 1, Fibonacci Numbers are Benford.

$\square$

Accounting Fraud:

The most used application of Benford's Law is in the field of forensic auditing. Benford's Law is used to detect data fabrication, meaning that someone creates their own data, and data falsification, meaning that someone manipulates the data. The main idea of using the law within auditing is that when certain financial data sets have been known to closely follow Benford's Law, then the manipulated or falsified data can sometimes be recognized by comparing the leading digits of the data sets with the law. This application of Benford's Law was discovered by Mark Nigrini, a professor at West Virginia University. He found that verified IRS tax data followed similarly with Benford's Law while many data sets that are fraudulent are not Benford. This has lead to the acceptance that true data is difficult to create so standard goodness-of-fit tests, like the chi-squared test are tests, that raise red flags. However, just because a data set does not follow Benford's Law does not prove that the data is fraudulent; rather the data becomes suspect. Then from here, forensic accountants can apply non-Benford tests and monitor the activity of the suspect. As well, data such as tax returns, many forensic accountants expect to find deviations from the true data since some people try to change their data to get lower taxes. This has lead to the development of specialized one-sided goodness-of-fit tests to detect false data. The first time Benford's Law was reportedly successful at detecting fraud was by the Brooklyn District Attorney's office in New York in 1995 where the chief financial investigator successfully prosecuted seven companies for theft by using goodness-of-fit tests to the first digit. This has led to newly developed tests on the first digit that are now used by most American state revenue services, the IRS, and foreign tax agencies to provided red flags for falsified data. [2]

Facts About Benford's Law:

Some basic facts about Benford data sets are that a data set or a sequence will follow Benford's Law if the set or sequence is exponential. This results is from Theorem 1 and an easy way to understand why is to look at the graph that pertains to the function since we can see that the majority of time is spent in the range of numbers that start with a 1. As well, any data set or sequence that follows Benford's Law, we can multiply the entire set or sequence by 10 and still have a set that follows Benford's Law since multiplying by 10 would not change the first digit. Now any easy way to determine if a set is not Benford, is if the data set is centralized around one number without a large standard deviation. A simple example of this is adult human heights since humans tend to be around five and a half feet tall and will only vary from around four and a half feet to seven feet. So we can see here that the 1's, 2's, 3's, 8's, and 9's will not even be accounted for. Now another fact about Benford sets is scale invariance. What this means is that if we took the 60 tallest structures in the world, they would would follow Benford's Law no matter if we measured the structures in feet, inches, or meters. However, if we also look at the scale invariance with human heights, the heights will not follow Benford's Law if we measured the structures in feet, inches, or meters. [2]

Appendix

[1] **Uniform Distribution Theorem Proof.**

Uniform Distribution Theorem. [4] For any irrational x in $[0, 1)$, the sequence $\{kx \bmod 1\}$ is uniformly distributed. That is, for any interval I in $[0, 1)$, we have

$$\lim_{n\to\infty} \frac{1}{n} |\{k \leq n : kx \bmod 1 \text{ is in } I\}| = \text{length}(I).$$

Before we prove the Uniform Distribution Theorem, we must describe a Lebesgue measure, Fourier-Stieltjes Transforms, and state a lemma that will be used.

Def. Lebesgue Measure. [2] A *Lebesgue measure*, denoted by λ, is a subset of an interval where the probability can be assigned for the subset.

More simply, a Lebesgue measure extends the concept that the probability of an interval should be proportional to its length.

Def. Fourier-Stieltjes transform. [8] Let f be a strictly positive, measurable function on the interval $(-\infty, \infty)$. Then there exists a monotone increasing, real-valued bounded function $\alpha(t)$ such that

$$f(x) = \int_{-\infty}^{\infty} e^{itx} d\alpha(t)$$

for "almost all" x. If $\alpha(t)$ is nondecreasing and bounded and $f(x)$ is defined as above, then $f(x)$ is called the *Fourier-Stieltjes transform* of $\alpha(t)$, and is both continuous and positive definite.

For a measure μ on an interval $G = [0, 1)$, a Fourier-Stieltjes Transform $\widehat{\mu}$ is defined by $\widehat{\mu}(m) = \int_G \overline{\chi}_m d\mu$ such that $m \in \mathbb{Z}$ and χ_m is defined by $\chi_m = e^{2\pi imx}$ for $x \in G$.

Lemma 1. [4] Given a sequence (μ_n) of probability measures on $G = [0, 1)$ and a probability measure μ on G, statements (a)-(d) are equivalent.

(a) $\mu_n \to \mu$ in the weak-$*$ (or vague) topology, i.e., $\lim_{n\to\infty} \int_G f d\mu_n = \int_G f d\mu$ for all continuous functions on G.

(b) If $F \subseteq G$ is closed, then $\limsup_n \mu_n(F) \leq \mu(F)$.

(c) If $V \subseteq G$ is open, then $\liminf_n \mu_n(V) \geq \mu(V)$.

(d) $\lim_{n\to\infty} \widehat{\mu_n}(m) = \widehat{\mu}(m)$ for all integers m.

Also, if (a)-(d) are true, we have:

(e) If $\mu(\{x\}) = 0$ for all x in $[0,1)$, then

$$\lim_{n\to\infty} \mu_n(I) = \mu(I) \text{ for each interval } I \in G.$$

Proof. **Lemma 1**

[4] (a)-(c) are elementary facts of all metric spaces which is a set where the distances between all members are known. Then we have that (a)$\Longrightarrow$(d) since χ_m is continuous on G which gives $\lim_{n\to\infty} \int_G \overline{\chi_m} d\mu = \lim_{n\to\infty} \int_G \overline{\chi_m} d\mu_n$. Then for (e), suppose that there exist an a and a b such that $(a,b) \subseteq I \subseteq [a,b]$. Then from (b) and (c),

$$\begin{aligned} &\lim_{n\to\infty} \sup \mu_n(I) \leq \lim_{n\to\infty} \sup \mu_n([a,b]) \leq \mu([a,b]) \\ &= \mu((a,b)) \leq \lim_{n\to\infty} \inf \mu_n((a,b)) \leq \lim_{n\to\infty} \mu_n(I) \\ \Longrightarrow &\lim_{n\to\infty} \sup \mu_n(I) \leq \lim_{n\to\infty} \mu_n(I) \Longrightarrow \lim_{n\to\infty} \mu_n(I) = \mu(I). \end{aligned}$$

□

Note, for the conclusion of the the Uniform Distribution Theorem,

$$\lim_{n\to\infty} \frac{1}{n} |\{k \leq n : kx \bmod 1 \text{ is in } I\}| = \text{length}(I)$$

can be rewritten as

$$\lim_{n\to\infty} P_n(I) = \lambda(I) \qquad \text{for each interval } I \in G \text{ such that } G = [0,1).$$

Now we can begin the proof for the Uniform Distribution Theorem.

Proof. **Uniform Distribution Theorem**

[4] Let P_n be the probability measure $\frac{1}{n}\sum_{k=1}^{n}\delta_{kx}$ on the interval $G = [1, 0)$ such that δ_{kx} is a point mass at $kx \mod 1$. Let λ be the Lebesgue measure on $[0, 1)$ and so $\widehat{\lambda}(0) = \widehat{P_n}(0) = 1$ and $\widehat{\lambda}(m) = 0$ for $m \neq 0$ since,

$$\widehat{\lambda}(0) = \int_0^1 e^{2\pi i(0)x} d\lambda = \int_0^1 1 d\lambda = 1$$

$$\widehat{P}_n(0) = \int_0^1 e^{2\pi i(0)x} dP_n = \int_0^1 1 dP_n = 1$$

$$\widehat{\lambda}(m) = \int_0^1 e^{2\pi i(m)x} d\lambda = \frac{e^{2\pi imx}}{2\pi im}\bigg|_0^1 = \frac{e^{2\pi im(1)} - e^0}{2\pi im} = 0.$$

Notice that the real part of $e^{2\pi im(1)}$ is always 1 and the imaginary is always 0 because $m \in \mathbb{Z}$.

Then by property (d) of Lemma 1 we must show that

$$\lim_{n\to\infty} \widehat{P_n}(m) = 0 = \widehat{\lambda}(m) \qquad \text{for } m \neq 0.$$

Since

$$\widehat{\delta_{kx}}(m) = \int_0^1 e^{2\pi imkx} d\widehat{\delta_{kx}} = e^{-2\pi imkx},$$

we have that

$$\lim_{n\to\infty} \widehat{P}_n(m) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} e^{-2\pi imkx} \qquad (1)$$

And since $mx \mod 1$ is the fractional portion of a number, mx is not an integer because x is irrational and $e^{-2\pi imkx} \neq 1$. Now we can rewrite (1) in the following form,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \alpha^k \qquad \text{where } \alpha = e^{-2\pi i m x}$$

Then clearly this sum is a geometric series that becomes $\alpha\ \frac{\alpha^n-1}{\alpha-1}$

This sum is bounded since $\alpha \neq 1$ and $|\alpha| = 1$. Note that $|re^{i\theta}| = 1$ when $r = 1$. So with $M = |\frac{\alpha}{\alpha-1}|$,

$$\left|\alpha\frac{\alpha^n-1}{\alpha-1}\right| = \left|\frac{\alpha}{\alpha-1}\right||\alpha^n - 1| = M|\alpha^n - 1| \leq M(|\alpha^n| + |-1|) = M(1+1) = 2M.$$

Thus,

$$\left|\widehat{P}_n(m)\right| < \tfrac{1}{n} \cdot 2M \to 0 \text{ as } n \to \infty$$

and so

$$\lim_{n\to\infty} P_n(I) = \lambda(I) \qquad \text{for each interval } I \in G$$

$\square$

References

[1] “American FactFinder.” United States Census Bureau . N.p., 05 Oct. 2010. Web. 15 Feb. 2017.

[2] Berger, Arno, and Theodore P. Hill. An Introduction to Benford’s Law. PRINCETON; OXFORD, Princeton University Press, 2015, www.jstor.org/stable/j.ctt1dr35m0.

[3] Katznelson, Yitzhak. “An Introduction To Harmonic Analysis.” (2004): n. pag. Cambridge University Press. Web. 12 Apr. 2017.

[4] Kenneth A. Ross. “Benford’s Law, A Growth Industry.” The American Mathematical Monthly, vol. 118, no. 7, 2011, pp. 571-583., www.jstor.org/stable/10.4169/amer.math.monthly.118.07.571.

[5] Newcomb, Simon. “Note on the Frequency of Use of the Different Digits in Natural Numbers.” American Journal of Mathematics 4.4 (1881): 39-40. Mathematical Association of America. Web. 17 Feb. 2017.

[6] Niven, Ivan. “The Carus Mathematical Monographs.” Irrational Numbers (1956): 1-157. Mathematical Association of America. Web. 1 Mar. 2017.

[7] Weisstein, Eric W. “Convolution Theorem.” From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/ConvolutionTheorem.html

[8] Weisstein, Eric W. “Fourier-Stieltjes Transform.” From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Fourier-StieltjesTransform.html