Some Results on Benford's Law and Ulam Sequences

Nicholas Álvarez, Andrew Hwang, Aaron Kriegman

Mentor: Jayadev Athreya

September 28, 2017

Abstract

In this paper, we give an overview on the ubiquitous nature of Benford's Law and cover principles of equidistribution to answer questions concerning under which conditions Benford's Law is satisfied. The paper expands to cover Benford's law for different bases, exponential sequences, recursive sequences, regular sequences, and certain Ulam sequences. By gaining a deeper insight of the criteria for sequences to satisfy Beford's Law, we were able to draw out conclusions about the possible linear growth of certain Ulam Sequences, particularly the standard (1,2) Ulam sequence. Furthermore, we establish an undiscovered, greater structure found within Ulam sequences, which sheds insight on the possibility of a formula to calculate all Ulam numbers.

1 Introduction

The mysterious phenomena that characterizes the essence of Benfords Law lies in an observation of the frequency distribution of the leading digits in both very abstract and many real-life sets of numerical data (the leading digit of a number is its leftmost digit). It is natural to seek the criteria a set or sequence must satisfy in order to obey this intriguing frequency distribution. Many distinctly defined sets and sequences satisfy Benfords Law, and our group was able to find infinitely many more examples of sequences that satisfy this law.

As we discover more of these criteria, we begin to uncover structures in seemingly unnatural sets. By running extensive code, we can take sets of numerical data and calculate the frequency of leading digits and observe whether these sets satisfy Benfords Law for an arbitrarily large number of terms. Using this method, we were able to establish structures found within particular sequences with an uncanny, ubiquitous nature: the Ulam sequences.

A sequence formed by appending the smallest number with a unique representation as a sum of two earlier numbers is an Ulam sequence. Despite the simplicity these sequences display at first, they have many unexpected properties which have thus far been observed but entirely evaded explanation or proof. Aside from the insight gained by manipulating Ulam sequences into sequences that satisfy Benfords Law, we made previously undiscovered observations on Ulam sequences which build upon the existing observations. Previous results have shown that a certain class of numbers are not in the Ulam sequence, but our observations provide a more precise explanation as to why and a possible way to prove it. We crafted our observation of the data into a conjecture which is stronger than existing conjectures, provides more insight into the nature of these sequences, and should be easier to prove.

Notation

The following notation/definitions will be used throughout the paper:

- We define a function $S_{(a,b)} : \mathbb{N} \to \mathbb{N}_0$ for positive integers a and b as the number of ways to write a natural number as the sum of two distinct terms of the (a, b) Ulam sequence.
- $S_{(1,2)}(n)$ will be simply denoted $S(n)$.
- Let $\mathbb{U}(a, b)$ denote the set of all Ulam numbers in the (a, b) Ulam Sequence.
- $\mathbb{U}(1, 2)$ will be simply denoted $\mathbb{U}$
- $\log(n)$ denotes the logarithm of n in base 10
- $\{n\}$ denotes the fractional part of $n \in \mathbb{R}$
- A sequence is denoted as (a_n) instead of the traditional $\{a_n\}$ (because we reference the fractional part of expressions frequently, this avoids confusion).
- For $b \in \mathbb{N}$, we define a function L^b as the following:

$$\begin{aligned} L^b : \quad & \mathbb{R}^+ \to \{1, 2, \dots, b-1\} \\ & x \mapsto \ell_b \end{aligned}$$

  where ℓ_b is the unique number in $\{1, 2, \dots, b-1\}$ s.t. $x = \ell_b \cdot b^n$ for some $n \in \mathbb{Z}$.

- $L^{10}(x)$ will simply be denoted $L(x)$.
- The cardinality of a countable set S will be denoted $\#S$.

2 Ulam Sequences

The (a, b) **Ulam sequence** (u_i) is defined by $u_1 = a$, $u_2 = b$, with the general term u_n for $n > 2$ given by the least integer uniquely expressible as the sum of two distinct earlier terms. The numbers in the sequence are called "u-numbers" or "Ulam numbers." In general, when referring to the Ulam sequence, we are referring to the $(1, 2)$ Ulam sequence. When speaking of other Ulam sequences, we will appropriately call them by their full name: the (a, b) Ulam sequence.

2.1 Known Findings of Ulam Sequences

In [1], Steinerberger found a hidden signal persisting within the Ulam numbers. Dilating the sequence by an irrational α where $\alpha \approx 2.571447...$, Steinberger found that $\cos(2.571447u_n) < 0$ for all $u_n \neq 2, 3, 47, 69$. The structure Steinenberger observed has been best described by Phillip Gibbs in the following conjecture:

Conjecture 1 (Gibbs [1]). *Let $\lambda = \frac{2\pi}{\alpha}$. For any Ulam sequence a_n there is a natural wavelength $\lambda \geq 2 \in \mathbb{R}$ such that if r_n is the residual of a_n mod λ in the interval $[0, \lambda)$ then for any $\epsilon > 0$ there are only a finite number of elements in the Ulam sequence such that $r_n < \frac{\lambda}{3} - \epsilon$ and $r_n > \frac{2\lambda}{3} + \epsilon$.*

2.2 Our Findings

We propose a stronger conjecture:

Conjecture 2. *For any (a, b) Ulam sequence (u_n) there is a natural wavelength $\lambda \geq 2 \in \mathbb{R}$ and differentiable periodic function $c : \mathbb{R} \to \mathbb{R}$ with period λ such that*

$$\limsup_{n\to\infty} |S_{(a,b)}(n) - nc(n)|$$

exists and is finite. Moreover, $c(x) = 0$ if and only if $\frac{\lambda}{3} \leq r_x \leq \frac{2\lambda}{3}$ where r_x is the residual of x $\pmod{\lambda}$.

Theorem 1. *Conjecture 2 $\implies$ Conjecture 1*

Proof. Let

$$\mathcal{C} := \left\{ x \in \mathbb{R} : \quad (x \mod \lambda) \in \left[\frac{\lambda}{3}, \frac{2\lambda}{3}\right] \right\}$$

and let $\mathcal{W} = \mathbb{R} - \mathcal{C}$, where $\mathcal{C}$ stands for center and $\mathcal{W}$ stands for wings. For any $r \in \mathcal{W}$, we can consider what happens to $\xi(n\lambda + r)$ as $n \to \infty$ where

$$\xi(x) := xc(x)$$

and $n \in \mathbb{Z}$. Notice $c(n\lambda + r)$ is constant, but $n\lambda + r \to \infty$ as $n \to \infty$, so $\xi(n\lambda + r) \to \infty$. Let

$$L = \limsup_{n\to\infty} |S_{(a,b)}(n) - nc(n)|$$

and consider some $N \in \mathbb{R}^+$ such that $\forall n \geq N$, we have

$$\xi(n\lambda + r') > L + 1$$

where r' is an arbitrary element of a small open neighborhood around r completely contained in $\mathcal{W}$. We can consider the asymptotic behavior of this whole neighborhood because $\frac{d}{dx}c(x)$ is bounded (although continuity may be sufficient for this step). So if $m = n\lambda + r'$ for some $m \in \mathbb{Z}$, then $S_{(a,b)}(m) > 1$, and therefore m is not an Ulam number.
Just outside of $\mathcal{C}$ we can see the outliers Gibbs described. As r approaches $\frac{\lambda}{3}$ from the left or $\frac{2\lambda}{3}$ from the right, $c(x)$ approaches 0.

Within $\mathcal{C}$, we can have arbitrarily many Ulam numbers as long as $L \geq 1$. □

Conjecture 2 is likely easier to prove because it explains more about why these patterns exist. Conjecture 1 doesn't specify whether we see few Ulam numbers $n \in \mathcal{W}$ because they are "misses" ($S_{(a,b)}(n) = 0$) or "hits" ($S_{(a,b)}(n) \geq 2$), whereas Conjecture 2 not only implies they are hits, but gives insight into the magnitude of $S_{(a,b)}(n)$. We discovered this structure when we plotted $S(n)$ for the first 10000 numbers against $n \pmod{\lambda}$ (Figure 1).

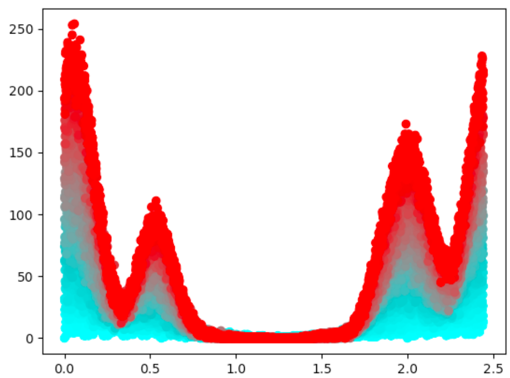

Figure 1: A scatter plot of $S(n)$ against $n \pmod{\lambda}$ for $0 \leq n < 10000$.

The points are colored on a gradient, so the blue is the early behavior and the red is the behavior around 10000.
Notice that the points approach a curve which is slowly rising. It is important that every point in W approaches this curve, and there are no rare exceptions where we sudddenly have an Ulam number at some random point such as $\frac{\lambda}{6}$.

Previously it has been observed that any irregular Ulam sequence is not equidistributed in $\mathcal{C}$, and in the case of the $(1,2)$ Ulam sequence, we see two peaks. We have a very good reason for this behavior. If we plot the Ulam numbers against their residues modulo λ, we can see some symmetries (Figure 2):

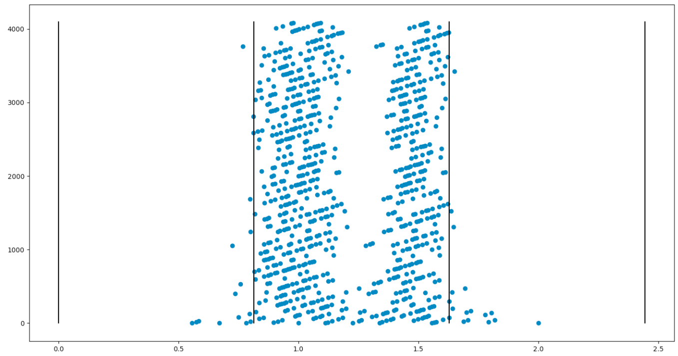

Figure 2: The Ulam numbers divided by λ plotted against their residues modulo λ

It appears the the left tower is the right tower shifted over. In fact it is. Most of the terms in the right tower will give a term in the left tower when added to 2. You'll notice the left tower is thicker. This is because the terms on the left side of the left tower are made by adding an outlier from the right side other than 2 to a term from the right tower. Terms in the right tower are created similarly, by adding a left outlier to a term from the left tower. These are merely patterns that we have observed in our data, but their formalization thus far is pure conjecture. We can see a sharper pattern in the $(2,3)$ Ulam sequence (Figure 3).

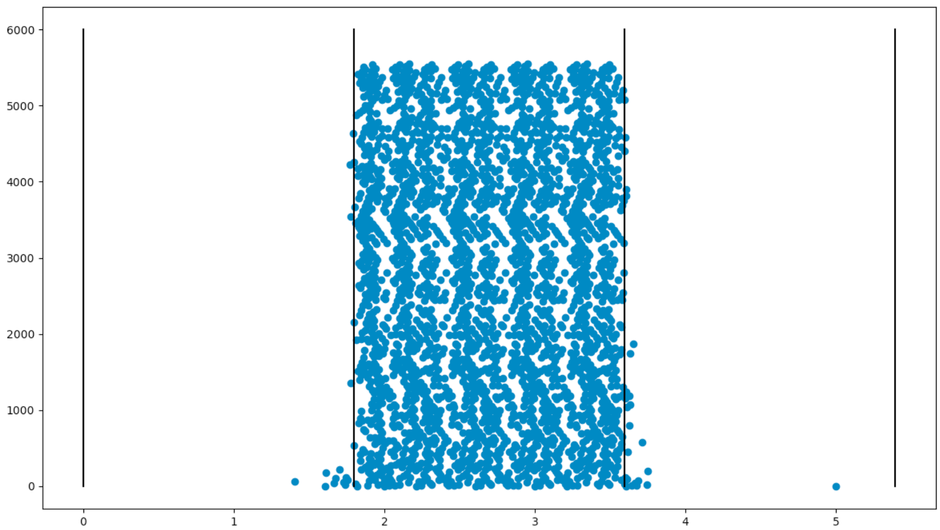

Figure 3: The $(2,3)$ Ulam numbers divided by λ plotted against their residues modulo λ

We see four and a half towers, each of which is a rough translation of the one to the right. There is a good reason for the smaller distance between the towers. While the furthest outlier for the $(1, 2)$ Ulam sequence, 2, is in the middle of $\left[\frac{2\lambda}{3}, \lambda\right]$, the outlier for the $(2, 3)$ case, 5, is very close to the end of the interval. Therefore, adding five leads to a relatively small shift left modulo λ.

3 A Brief Overview of Benford's Law and Equidistribution

Understanding the intuition behind a sequence satisfying Benford's Law becomes clear upon the understanding of the two following equidistribution theorems and Lemma 1.

3.1 Equidistribution

Theorem 2 (Weyl's Equidistribution theorem [3])**.** *Let α be irrational, $a, b \in R$, and $0 \leq a \leq b \leq 1$. Then*

$$\lim_{N \to \infty} \frac{\#\{0 \leq n \leq N : a \leq \{n\alpha\} \leq b\}}{N} = b - a$$

that is, the fractional parts of multiples of α are equidistributed in $[0, 1]$.

Theorem 3 (Difference Theorem [3])**.** *If a sequence $(x_n)_{n \geq 1}$ has the property*

$$\lim_{k \to \infty} (x_{k+1} - x_k) = \alpha,$$

where α is an irrational number, then the sequence (x_n) is equidistributed modulo 1.

3.2 Benford's Law

Law 1 (Benford)**.** *For $1 \leq d \leq 9$, the frequency, f_d, of the leading digit d in a sequence $\{|a_n|\}$ is given by*

$$f_d = \lim_{N \to \infty} \frac{\#\{0 \leq n \leq N : L(a_n) = d\}}{N} = \log(d+1) - \log(d) = \log\left(1 + \frac{1}{d}\right).$$

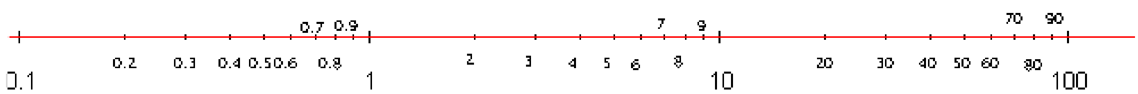

Figure 4: Logarithmic Scale

Consider the sequence

$$1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384...$$

Looking at the sequence of leading digits, $1, 2, 4, 8, 1, 3, 6, 1, 2, 5, 1, 2, 4, 8, 1, ...$ it appears that some digits appear more frequently as leading digits than others. We investigate this property in order to become more familiar with proving that a sequence satisfies Benford's Law. When proving that a sequence satisfies Benford's Law, we generally require a manipulation of that sequence to equidistribute under some interval (usually [0,1)).

Example 1. *The sequence $(2^n)_{n\geq 1}$ satisfies Benford's Law*

To prove this, we must first prove the following lemma:

Lemma 1. $L(N) = L\left(10^{\{\log N\}}\right)$ *for any positive integer* N

Proof. Consider a positive integer N. In order to find $L(N)$, we write N as

$$N = 10^{\log N} = 10^{\lfloor \log N \rfloor + \{\log N\}}$$

where $\{k\}$ denotes the fractional part of any integer k. Because multiplication by $10^{\lfloor \log N \rfloor}$ doesn't change the leading digit of an integer, we see that $L(N) = L\left(10^{\{\log N\}}\right)$. □

Since we know that for any $N \in \mathbb{Z}^+$, $L(N) = L\left(10^{\{\log N\}}\right)$, we can better understand the proof of Example 1 as we see how $n \log 2$ equidistributes over an interval:

Proof. Let $L(2^n) = k$, then

$$k \cdot 10^p \leq 2^n \leq (k+1) \cdot 10^p.$$

Taking the logarithm base 10, we find that $\{n \log(2)\} \in [\log(k), \log(k+1)]$.
By Weyl's equidistribution theorem, the fractional part of $n \log(2)$ equidistributes over the interval $[0, 1)$. Thus, the number of times this map falls between the interval $[a, b]$ is $b - a$, meaning that the proportion of powers of 2 that start with k equals

$$\log(k+1) - \log(k) = \log\left(1 + \frac{1}{k}\right).$$

□

Lemma 2. *If a sequence $(x_n)_{n\geq 1}$ has the property*

$$\lim_{k\to\infty} \{\log(x_k) - \log(x_{k-1})\} = \alpha,$$

where α is a positive irrational number, then it satisfies Benford's Law.

Proof. If

$$\lim_{k\to\infty} \{\log(x_k) - \log(x_{k-1})\} = \alpha$$

where α is irrational, then by Theorem 3, the sequence $(\{\log(x_n)\})_{n\geq 1}$ is equidistributed modulo 1. Because the fractional part of $\log x_n$ equidistributes over the interval $[0,1)$, by Theorem 2 we know the number of times the map falls between the interval $[a,b]$ is $b-a$, meaning that the proportion of (x_n) with leading digit d is

$$\log(d+1) - \log(d) = \log\left(1 + \frac{1}{d}\right)$$

Therefore, (x_n) satisfies Benford's Law. The idea of Lemma 1 is used again to justify the last step of this proof. □

4 Extension to Other Bases

Law 2 (Extended Benford's)**.** *For $1 \leq d \leq b-1$, the frequency of the leading digit d in a sequence $\{|a_n|\}$ in base b is given by*

$$f_d = \lim_{N\to\infty} \frac{\#\{0 \leq n \leq N : L^b(a_n) = d\}}{N} = \log(d+1) - \log(d) = \log\left(1 + \frac{1}{d}\right).$$

Lemma 3. *If a sequence $(x_n)_{n\geq 1}$ has the property such that*

$$\lim_{k\to\infty} \{\log_b(x_{k+1}) - \log_b(x_k)\} = \alpha$$

where α is a positive irrational base b, then (x_n) satisfies Benford's Law.

Proof. If α is irrational, then by Theorem 3, the sequence $\{\log_b(x_n)\}_{n\geq 1}$ is equidistributed modulo 1_b. Because the fractional part of $\log_b(x_n)$ eqidstributes over the interval $[0,1)$, by Theorem 2 we have that the number of times the map falls between the interval [a,b] is $b-a$, implying that the proportion of (x_n) with leading digit d is

$$\log_b(d+1) - \log_b(d) = \log_b\left(1 + \frac{1}{d}\right).$$

□

This tool implies that nearly all structures that follow Benford's Law in one base follow Benford's law in another.

For example, an exponential series of the form x^n satisfies Benford's law in nearly all bases. By Lemma 3, for example, we have that for any $k \in \mathbb{N}$,

$$\{\log_b(x^k) - \log_b(x^{k-1})\} = \{\log_b(x)\},$$

implying that $\log_b(x)$'s irrationality is the only criterion necessary for (x_n) to satisfy Benford's Law.

5 Recursive Sequences

To see the relationship between recursive sequences and Benford's Law, first, let's look at the following example:

Example 2. *The Fibonnaci sequence* $1, 1, 2, 3, 5, 8, 13, 21, ...$ *satisfies Benford's Law.*

Proof. Let (F_n) denote the Fibonacci sequence where F_k is defined as the k^{th} Fibonacci number. Because

$$\lim_{n\to\infty} \frac{F_n}{F_{n-1}} = \phi,$$

we can approximate F_n as $F_n \approx \phi F_{n-1}$. Now consider

$$\begin{aligned}\lim_{n\to\infty} \{\log(F_n) - \log(F_{n-1})\} &\approx \lim_{n\to\infty} \{\log(\phi F_{n-1}) - \log(F_{n-1})\} \\ &\approx \{\log(\phi)\}\end{aligned}$$

Although $\log(F_n) - \log(F_{n-1})$ is not exactly ϕ, it is seemingly close enough to ϕ, so we can apply Lemma 2 to show that the Fibonacci sequence satisfies Benford's Law. □

Looking beyond this notorious sequence, we wanted to extend our results to prove a more general relation between recursive sequences and Benford's Law.

Now consider the sequence $(x_n)_{n\geq 1}$ satisfying the following linear recursion

$$x_{n+m} = a_{m-1}x_{n+m-1} + a_{m-2}x_{n+m-2} + ... + a_0 x_n \tag{1}$$

for $n \geq 1$ and additionally, $x_i =$ a constant c_i $\forall i \in \{1, 2, ..., m\}$. In order to prove that (x_n) satisfies Benford's Law, we consider the linear recurrences case by case:

First, we consider the case where the characteristic polynomial of the linear recurrence only has one root. Then, we will expand our considerations to characteristic polynomials with s roots with a finite multiplicity:

Case 1 (One Root)**.** *If the characteristic polynomial has only one root* $r \neq \pm 10^l, l \in \mathbb{Z}$*, then the linear recurrence satisfies Benford's Law.*

Proof. We can create a characteristic polynomial to generate an equation giving x_n for any n. In general,

$$x_n = r^{n-1} \cdot \sum_{k=0}^{m-1} b_k n^k$$

where b_i is constant $\forall i \in \{1, 2, ..., m-1\}$. By substitution,

$$\begin{aligned}\lim_{n\to\infty} \frac{x_{n+1}}{x_n} &= \lim_{n\to\infty} \frac{(b_0 + b_1(n+1) + ... + b_{m-1}(n+1)^{m-1})r^n}{(b_0 + b_1 n + ... + b_{m-1}n^{m-1})r^{n-1}} \\ &= \lim_{n\to\infty} \frac{(b_0 + b_1(n+1) + ... + b_{m-1}(n+1)^{m-1})r}{(b_0 + b_1 n + ... + b_{m-1}n^{m-1})} \\ &= r\end{aligned} \tag{2}$$

Hence,

$$\lim_{n\to\infty} \{\log|x_{n+1}| - \log|x_n|\} = \{\log|r|\}.$$

Since $\log|r|$ is irrational, the linear recurrence satisfies Benford's Law. □

Case 2 (Multiple Roots)**.** *If the characteristic polynomial has distinct roots $r_1, r_2, ...r_s$ with multiplicities $y_1, y_2, ...y_s$ respectively, then the linear recurrence will obey Benford's Law.*

Proof. Without loss of generality, suppose $|r_1| \geq |r_i|$ for $i \in \{2, 3, ..., s\}$. Note that x_n can be represented as

$$x_n = \sum_{k=1}^{s} P_k(n-1) \cdot r_k^{n-1}$$

where P_i is a polynomial with $\deg(P_i) \leq y_i - 1$.

Thus,

$$\begin{aligned}\lim_{n\to\infty} \frac{x_{n+1}}{x_n} &= \lim_{n\to\infty} \frac{P_1(n)r_1^n + P_2(n)r_2^n + ... + P_s(n)r_s^n}{P_1(n-1)r_1^{n-1} + P_2(n-1)r_2^{n-1} + ... + P_s(n-1)r_s^{n-1}} \\ &= \lim_{n\to\infty} \frac{r_1^n(P_1(n) + P_2(n)(r_2/r_1)^n + ... + P_s(n)(r_s/r_1)^n)}{r_1^{n-1}(P_1(n-1) + P_2(n-1)(r_2/r_1)^{n-1} + ... + P_s(n-1)(r_s/r_1)^{n-1})} \\ &= r_1\end{aligned} \tag{3}$$

Since

$$\lim_{n\to\infty} \log|x_{n+1}| - \log|x_n| = \log|r_1|$$

the linear recurrence will satisfy Benfords' Law as long as $\log|r_1|$ is irrational and $P_1(n-1) \neq 0$.

However, if $r_1 = -r_2$ and n is odd, then we can write

$$x_n = (P_1(n-1) + P_2(n-2))r_1^{n-1} + \sum_{k=3}^{s} P_k(n-1) \cdot r_k^{n-1}$$

and if n is even,

$$x_n = (P_1(n-1) - P_2(n-2))r_1^{n-1} + \sum_{k=3}^{s} P_k(n-1) \cdot r_k^{n-1}.$$

By a similar approach, we have that

$$\lim_{n\to\infty}\left|\frac{x_{2n+2}}{x_{2n}}\right| = r_1^2$$

and

$$\lim_{n\to\infty}\left|\frac{x_{2n+3}}{x_{2n+1}}\right| = r_1^2.$$

If $P_1(n-1) + P_2(n-1) \neq 0$ and $r_1 \neq \pm 10^l$ for $l \in \mathbb{Z}$, the sequence will obey Benford's Law. By Lemma 2, the sequences $(x_{2n})_{n\geq 1}$ and $(x_{2n+1})_{n\geq 1}$ satisfy Benford's Law since

$$\lim_{n\to\infty}\left\{\log\left|\frac{x_{2n+2}}{x_{2n}}\right|\right\} = \{2\log(r_1)\}$$

and

$$\lim_{n\to\infty}\left\{\log\left|\frac{x_{2n+3}}{x_{2n+1}}\right|\right\} = \{2\log(r_1)\}.$$

Taken together, the sequence (x_n) will satisfy Benford's law as well.

□

Now that we've proven these two cases to be true, we have a strong corollary:

Corollary 1. *All linear recurrences (x_n) satisfy Benford's Law excluding those which have $r = 10^l$ for some $l \in \mathbb{Z}$, where r is either the only root of the characteristic polynomial of (x_n) or it is the root of the characteristic polynomial of (x_n) with the largest magnitude.*

Proof. This directly follows from Case 1 and Case 2 □

Notice that linear recursive sequences will satisfy Benford's Law for any other base b as long as $r \neq b^n, n \in \mathbb{Z}$ where r is either the only root of the characteristic polynomial of (x_n) or it is the root of the characteristic polynomial of (x_n) with the largest magnitude.

6 Regular Sequences

Definition 1 (Regular Sequence)**.** *A sequence (x_n) is said to be regular if after a finite number of terms, the differences between successive terms in the sequence becomes periodic.*

An example of a regular sequence is

$$(x_n) := 1, 5, 7, 4, 3, 8, 10, 14, 22, 24, 28, 36, 38, 42, 50, ...$$

The sequence of differences between successive terms of (x_n) is

$$4, 2, -3, -1, 5, 2, 4, 8, 2, 4, 8, 2, 4, 8, ...$$

Eventually the differences between terms become periodic and the differences between terms rotates through $2, 4$, and 8. Since regular sequences demonstrate linear growth, most of them seem to not satisfy Benford's Law; however, we can still prove something interesting about them related to Benford's Law:

Lemma 4. *$(a^{x_n})_{n\geq 1}$ satisfies Benford's Law, where (x_n) is a regular sequence and a is a natural number $\neq 10^k$ for $k \in \mathbb{Z}$.*

Proof. If we can show that

$$\lim_{k\to\infty} \{\log(a^{x_k}) - \log(a^{x_{k-1}})\} = \alpha$$

where α is a positive irrational number, then Lemma 2 shows that (a^{x_n}) satisfies Benford's Law. However,

$$\begin{aligned}\{\log(a^{x_k}) - \log(a^{x_{k-1}})\} &= \left\{\log\left(\frac{a^{x_k}}{a^{x_{k-1}}}\right)\right\} \\ &= \{(x_k - x_{k-1})\log(a)\}\end{aligned}$$

and it becomes evident that the differences between $\log(a^{x_k})$ and $\log(a^{x_{k-1}})$ does not seem to approach a constant irrational number, since $x_k - x_{k-1}$ is not constant $\forall k$. Because of this, we take a new approach to the proof using the fact that (a^{x_n}) is regular.

Since (a^{x_n}) is a regular sequence, there exists some sufficiently large natural number d where for $k \geq 1$,

$$x_{d+mk+1} - x_{d+mk} = c_1,$$

$$x_{d+mk+2} - x_{d+mk+1} = c_2,$$

$$\vdots$$

$$x_{d+mk+m-1} - x_{d+mk+m-2} = c_{m-1}$$

where all c_i are constants and m is the length of the periodic differences (in the example above, when the differences between terms rotated between 2,4, and 8, the length of the periodic differences was 3). Now consider the m distinct sequences $(x_{d+mk})_{k\geq 0}, (x_{d+mk+1})_{k\geq 0}, ..., (x_{d+mk+m-1})_{k\geq 0}$. By showing that the subsequences $(a^{x_{d+mk}})$, $(a^{x_{d+mk+1}})$, $\ldots$, $(a^{x_{d+mk+m-1}})$ all satisfy Benford's Law, we have that (a^{x_n}) satisfies Benford's Law, simply because an infinite number of terms in the sequence obey Benford's Law, while finitely many may not (these are the terms before x_{d+mk}).

We know that for any $i \in \{0, 1, \ldots, m-1\}$, the difference between any two successive terms in the sequence x_{d+mk+i} is

$$s = \sum_{i=1}^{m-1} c_i$$

Hence, for any $j \in \mathbb{N}$,

$$\begin{aligned}\{\log(a^{x_{d+m(k+j)+i}}) - \log(a^{x_{d+m(k+(j-1))+i}})\} &= \{((x_{d+m(k+j)+i}) - (x_{d+m(k+(j-1))+i}))\log(a)\} \\ &= \{s\log(a)\}.\end{aligned}$$

Since a is not a power of 10 by the conditions of the lemma, $\{s\log(a)\}$ is irrational, and thus,

$$(a^{x_{d+mk}}), (a^{x_{d+mk+1}}), \ldots, (a^{x_{d+mk+m-1}})$$

all satisfy Benford's Law. Therefore $(a^{x_n})_{n\geq 1}$ satisfies Benford's Law. □

Similar arguments used in the proof of this result could be used to prove the following conjecture relating Ulam sequences to Benford's Law:

Conjecture 3. *$(a^{u_n})_{n\geq 1}$ satisfies Benford's Law, where (u_n) is the (1,2) Ulam sequence and a is a natural number $\neq 10^k$ for $k \in \mathbb{Z}$.*

In Lemma 4, we proved that sequences of a similar form satisfy Benfords Law. Since we know that some Ulam sequences are regular, we know that sequences of the form (a^{a_k}) where (a_k) is a regular Ulam sequence satisfy Benfords Law.

To prove Conjecture 3, we use a similar approach to the proof of Lemma 4, where we partitioned a sequence into subsequences, and showed that the base a raised to each of those subsequences will satisfy Benford's Law, and then pieced all of the subsequences together to prove that the sequence as a whole will satisfy Benford's Law:

Recall that

$$\mathcal{C} := \left\{x \in \mathbb{R} : \quad (x \mod \lambda) \in \left[\frac{\lambda}{3}, \frac{2\lambda}{3}\right]\right\}.$$

Note that we can partition $\mathcal{C}$ into equal, disjoint intervals:

$$\mathcal{C} = \bigcup_{1\leq j\leq n} I_j$$

where the length of all $I_j = \epsilon$ for some arbitrarily small ϵ. All Ulam numbers in a subinterval have approximately the same residual modulo λ since ϵ is very small. Thus, we may assume that all Ulam numbers in the same subinterval have the same residual modulo λ.

For any subinterval I_k, we define a new sequence $v(I_k, r)$ where $v(I_k, i)$ is the i^{th} smallest Ulam number in I_k $(1 \leq k \leq n)$. According to Lemma 2, if we show that

$$\lim_{i\to\infty} \left\{ \log\left(a^{v(I_k,i)}\right) - \log\left(a^{v(I_k,i-1)}\right) \right\} = \alpha$$

where α is a positive irrational number, then the sequence $\left(a^{v(I_k,r)}\right)$ satisfies Benford's Law. However, for any i,

$$\begin{aligned} \left\{ \log\left(a^{v(I_k,i)}\right) - \log\left(a^{v(I_k,i-1)}\right) \right\} &= \left\{ \log\left(\frac{a^{v(I_k,i)}}{a^{v(I_k,i-1)}} \right) \right\} \\ &= \{((v(I_k,i) - v(I_k,i-1)) \log(a)\} \\ &\approx \{m\lambda \log(a)\} \end{aligned}$$

for some $m \in \mathbb{N}_0$. Using Lemma 2 to prove that $\left(a^{v(I_k,r)}\right)$ satisfies Benford's Law is more tricky than we expected. If we are able to study the properties of m, we can still salvage our proof to show that $\left(a^{v(I_k,r)}\right)$ satisfies Benford's Law. We found that m happens to only be equal to the numerators of continued fractions that approximate λ. For example, $\frac{22}{9} \approx \lambda$, and so m happens to equal 22 frequently. If we can show that m takes on a certain value an infinite number of times, then the sequence will satisfy Benford's Law. Lastly, if we can show that $\left(a^{v(I_k,r)}\right)$ satisfies Benford's Law for almost all k, then we will have proven that $(a^{u_n})_{n\geq 1}$ satisfies Benford's Law, by showing that an infinite number of terms in the sequence (a^{u_n}) satisfy Benford's Law, while finitely many may not.

7 Conclusions and Future Directions

7.1 Ulam Sequences

Conjecture 2 about Ulam sequences suggests that they have some underlying periodic structures that are yet to be formalized. Naturally, the concept of periodic functions leads our group to believe that perhaps the function $S_{(a,b)} : \mathbb{N} \to \mathbb{N}_0$ can be accurately approximated using some Fourier Analysis. The two obstacles we must overcome in order to use Fourier Analysis are: the fact that the graph of $S_{(a,b)}$ is a scatter plot, rather than a definite curve that can be approximated, and the fact that as we graph $S_{(a,b)}$ at increasing values, these new points are simply laid on top of previous points, and therefore we must have a growth factor k such that

$$\limsup_{n\to\infty} |S_{(a,b)}(n) - kc(n)|$$

exists and is finite, where $c(n)$ is the differentiable period function we are seeking. As seen in our conjecture, we assume k to be equal to n; however, it is possible that the growth factor could be $\sqrt{n}$.

Because of how $S_{(a,b)}$ is defined, we know that $\mathbb{U}(a,b) = \{n \in \mathbb{N} : S_{(a,b)}(n) = 1\}$. If we do find such a function c that satisfies our conjecture, it is possible to find all the elements in $\mathbb{U}(a,b)$.

7.2 Periodic Ulam Sequences

Certain classes of Ulam sequences have been shown to be regular, but the proofs are clunky, complicated, and yet conceptually simple. These proofs should not be hard to extend to other classes of Ulam sequences, but they are solely due to computational complexity. For these reasons the problem seems to lend itself excellently to computer proof. Conceivably, it would not be too hard to write a program which, given the seeds a and b of a given Ulam sequence, will halt if the sequence is regular and not halt if it is irregular.

7.3 Benford's Law

We were able to prove that a wide variety of differently structured sequences satisfy Benford's Law. The fact that so many diverse sequences satisfy this strange elegantly expresses the beauty of Benford's Law, and how its essence lies in deep mathematics, rather than just applied mathematics. An important lemma not mentioned in our paper (it has a trivial proof) states that sequences that demonstrate linear growth do not satisfy Benford's Law. It has been conjectured previously that some linear growth structure can be found in the Ulam sequence, so perhaps a way to show this is to prove that the regular Ulam sequence doesn't satisfy Benford's Law. Similarly, if we can prove that the (1,2) Ulam sequence has a linear growth structure, then we can immediately prove that the Ulam sequence does not satisfy Benford's Law.

8 Acknowledgements

We would like to thank Larsen Linov, our counselor/mentor for our project during PROMYS, Dr. Dan Jerison, who was an effective overall supervisor who gave meaningful feedback on our project, Professor Jayade Athreya, who was our official research mentor for Siemens, and lastly, the PROMYS program for facilitating the collaboration for this paper.

References

[1] S. Steinerberger, A Hidden Signal in the Ulam Sequence, https://arxiv.org/ abs/1507.00267.

[2] P. Gibbs, A conjecture for Ulam Sequences (2015).

[3] A. Jamain, Benford's Law, *Imperial College of London Department of Mathematics.*