Benford Ulam Paper

Nicholas Alvarez, Andrew Hwang, Aaron Kriegman
Counselor: Larsen Linov

August 12, 2017

Abstract

We give an overview on the ubiquitous nature of Benford's Law. We cover principles of equidistribution and answer questions concerning under which conditions Benford's Law is satisfied. The paper expands to cover Benford's law for different bases, for exponential sequences, recursive sequences, and certain Ulam sequences. Furthermore, we establish greater structure found within the Ulam sequence.

Notation

The following notations/definitions will be used throughout this 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 $S(n)$.
- $\log(n)$ 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).
- We define a function in base b as the following:

$$L^b : \quad \mathbb{R}^+ \to \{1, 2, \ldots, b-1\}$$
$$x \mapsto \ell_b$$

  where ℓ_b is the unique number in $\{1, 2, \ldots, b-1\}$ 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 $\#S$.

1 Ulam sequence

$$1, 2, 3, 4, 6, 8, 11, 13, 16, 18, ...$$

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 denoted as "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.

1.1 Known Findings of Ulam Sequences

In a paper by Stefan Steinerberger, 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])**.** *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 $r_n > \frac{2\lambda}{3} + \epsilon$.*

1.2 Our Findings

We propose a stronger conjecture:

Conjecture 2 (Alvarez-Hwang-Kriegman)**.** *For any Ulam sequence (a_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*

$$\limsup_{n\to\infty} |S(n) - nc(n)|$$

exists and is finite, and $c(x) = 0$ 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 $(xc)(n\lambda + r)$ as $n \to \infty$ where

$$(xc)(y) := yc(y)$$

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

$$L = \limsup_{n\to\infty} |S(n) - nc(n)|$$

and consider some N such that $\forall n \geq N$, we have

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

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 $S(m) > 1$, 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 $L \geq 1$.

□

This conjecture 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 ($S(n) = 0$) or hits ($S(n) \geq 2$), whereas Conjecture 2 not only implies they are hits, but gives insight into the magnitude of $S(n)$. We discovered this structure when we plotted $S(n)$ for the first 10000 numbers against $n \pmod{\lambda}$ (Figure 1).

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 numbers, we see two peaks. We have a very good reason for

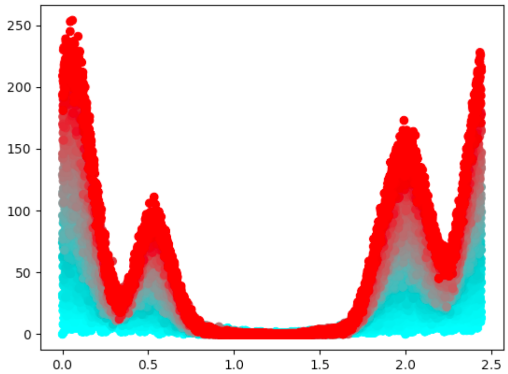

Figure 1: A scatter plot of $S(n)$ against $n \pmod{\lambda}$ for $0 \le n < 10000$.

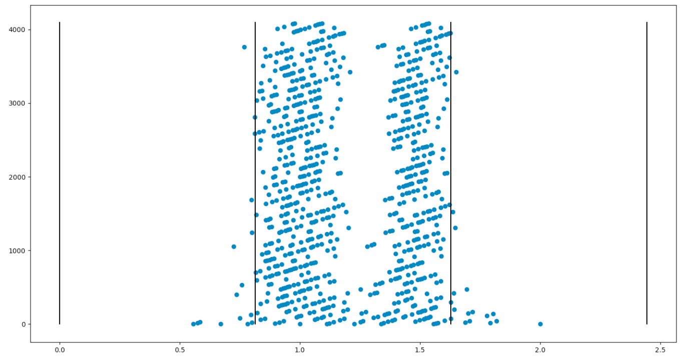

Figure 2: The Ulam numbers divided by λ plotted against their residues modulo λ

this behavior. If we plot the Ulam numbers against their residues modulo λ, we can see some symmetries:

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).

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 amounts to a relatively small shift left modulo λ.

2 Benford's Law and Equidistribution

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 attempt to investigate this property.

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

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

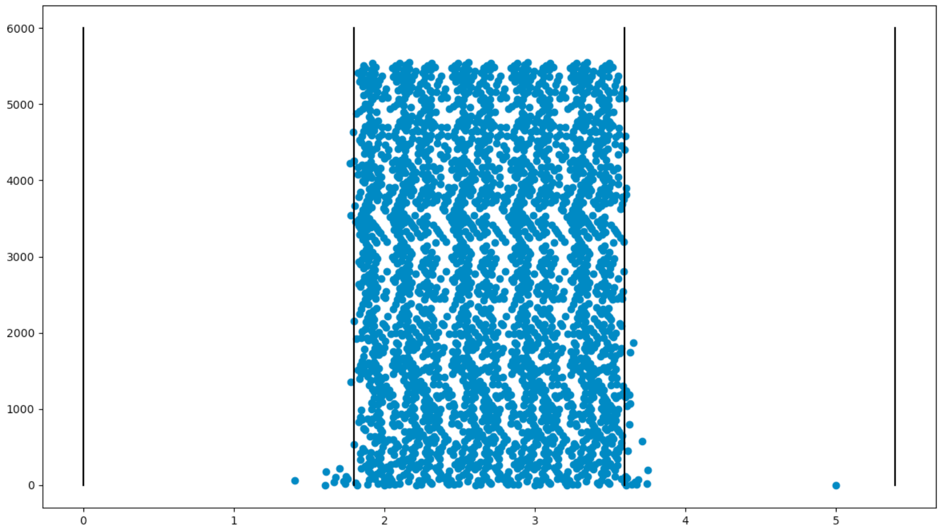

Figure 3: The $(2, 3)$ Ulam numbers divided by λ plotted against their residues modulo λ

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 $L(N) = L\left(10^{\{\log N\}}\right)$. $\square$

2.1 Equidistribution

Theorem 2 (Weyl's equidistribution theorem). *Let α be irrational, 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). *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, then the sequence (x_n) is equidistributed modulo 1.

2.2 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*

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

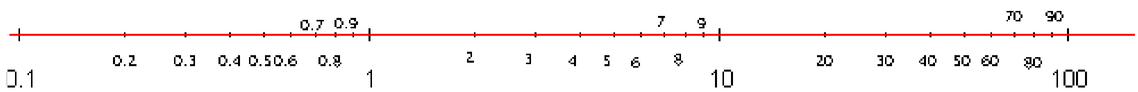

Figure 4: Logarithmic Scale

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

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 which 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 an 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. □

3 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 Weyl's we have that 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

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

□

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 7, 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 the only criteria necessary for Benford's Law to be satsified is for $\log_b(x)$ to be irrational.
Similarly, linear recursive sequences of the form shown in (1) also satisfy Benford's Law. Adopting the same approach in (2) and (3), we see that linear recursive sequences will hold true in another base as long as $r \neq b^n, n \in \mathbb{Z}$ or $r_1 \neq b^n, n \in \mathbb{Z}$. The same holds true for regular sequences.

4 Recursive Sequences

Example 2. *The Fibonnaci sequence* $1, 1, 2, 3, 5, 8, 13, 21, \ldots$ *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. □

Theorem 4. *Nearly all linear recurrences satisfy Benford's Law.*

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

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

for $n \geq 1$ and additionally, $x_i =$ a constant c_i $\forall i \in \{1, 2, \ldots, m\}$. To prove that the following linear recursion 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:

Lemma 4. *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 $\forall i \in \{1, 2, \ldots, m-1\}$. By substitution,

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

so

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

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

Lemma 5. *If the characteristic polynomial has distinct roots* $r_1, r_2, \ldots r_s$ *with multiplicity* $y_1, y_2, \ldots y_s$ *respectively, then the linear recurrence will follow Benford's Law.*

Proof. Without a loss of generality, suppose $|r_1| > |r_i|$ for $i \in \{2, 3, \ldots, s\}$. 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,

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

Since

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

our 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, we can then 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,

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

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

Therefore, nearly all linear recurrences satisfy Benford's Law. □

Definition 1. *A sequence (x_n) is defined as being regular if eventually 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 differences between terms of (x_n) is

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

Eventually the differences between terms becomes 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 in regards to Benford's Law:

Lemma 6. *$(a^{x_n})_{n\geq 1}$ 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 tells us 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 difference between $\log(a^{x_k})$ and $\log(a^{x_{k-1}})$ doesn't 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, we know that 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 before, when the differences between terms rotated between 2,4, and 8, the length of the periodic differences is 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, then (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

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

This means that for any $j \in \mathbb{N}$,

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

Because 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, and therefore $(a^{x_n})_{n\geq 1}$ satisfies Benford's Law. □

References

S. Steinerberger, A Hidden Signal in the Ulam Sequence, https://arxiv.org/ abs/1507.00267.
P Gibbs, A conjecture for Ulam Sequences. viXra:1508.0045