Generalized stick fragmentation and Benford's Law

*Xinyu Fang[1], *Maxwell Sun[2], Amanda Verga[3]
Mentor: Steven J. Miller

[1]fxinyu@umich.edu [2]mrsun@mit.edu [3]amanda.verga@trincoll.edu

SMALL REU, Williams College

Young Mathematicians Conference
Ohio State University

August 15, 2023

Table of Contents

1. Introduction: Benford's Law
2. Our Problem: Stick Breaking
3. Results
4. Further Directions

Table of Contents

1. Introduction: Benford's Law
2. Our Problem: Stick Breaking
3. Results
4. Further Directions

Benford's Law

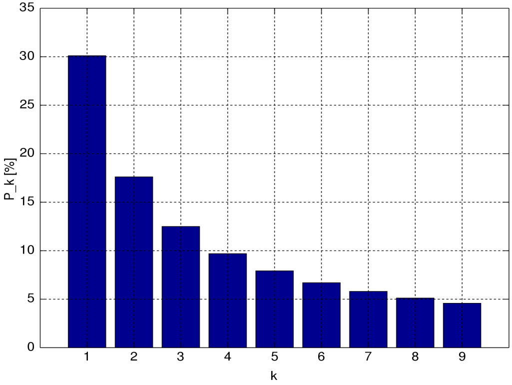
Definition (Benford's Law)

A data set is said to satisfy **Benford's Law base** B (where $B > 1$) if the probability of observing a value with first digit d is $\log_B\left(\frac{d+1}{d}\right)$.

Definition (Benford's Law)

A data set is said to satisfy **Benford's Law base** B (where $B > 1$) if the probability of observing a value with first digit d is $\log_B\left(\frac{d+1}{d}\right)$.

For example, when $B = 10$ (figure from Wikipedia):



Benford's Law

Definition (Benford's Law)

A data set is said to satisfy **Benford's Law base** B (where $B > 1$) if the probability of observing a value with first digit d is $\log_B\left(\frac{d+1}{d}\right)$.

Benford's Law

Definition (Benford's Law)

A data set is said to satisfy **Benford's Law base** B (where $B > 1$) if the probability of observing a value with first digit d is $\log_B\left(\frac{d+1}{d}\right)$.

Examples:

Benford's Law

Definition (Benford's Law)

A data set is said to satisfy **Benford's Law base** B (where $B > 1$) if the probability of observing a value with first digit d is $\log_B\left(\frac{d+1}{d}\right)$.

Examples:

- special sequences and functions (e.g., $n!$ and the Fibonaccis)
- iterations of the $3x + 1$ map [[1]]
- financial data (fraud detection)
- **products of random variables**

Strong Benford

Definition (Significand, Mantissa)

For $x > 0$, the **significand** of x base B is $S_B(x) \in [1, B)$ such that

$$x \ = \ S_B(x) \cdot B^k$$

for some integer k.

Strong Benford

Definition (Significand, Mantissa)

For $x > 0$, the **significand** of x base B is $S_B(x) \in [1, B)$ such that

$$x \ = \ S_B(x) \cdot B^k$$

for some integer k. The **mantissa** of x base B is

$$M_B(x) \ = \ \log_B(S_B(x)) \ = \ \log_B(x) \pmod 1.$$

Strong Benford

Definition (Significand, Mantissa)

For $x > 0$, the **significand** of x base B is $S_B(x) \in [1, B)$ such that

$$x \;=\; S_B(x) \cdot B^k$$

for some integer k. The **mantissa** of x base B is

$$M_B(x) \;=\; \log_B(S_B(x)) \;=\; \log_B(x) \pmod 1.$$

Definition (Strong Benford)

A data set $\{x_n\}$ is **strong Benford base** B if $\{M_B(x_n)\}$ is distributed uniformly in $[0, 1]$.

Strong Benford

Definition (Significand, Mantissa)

For $x > 0$, the **significand** of x base B is $S_B(x) \in [1, B)$ such that

$$x \;=\; S_B(x) \cdot B^k$$

for some integer k. The **mantissa** of x base B is

$$M_B(x) \;=\; \log_B(S_B(x)) \;=\; \log_B(x) \pmod 1.$$

Definition (Strong Benford)

A data set $\{x_n\}$ is **strong Benford base** B if $\{M_B(x_n)\}$ is distributed uniformly in $[0, 1]$. In other words, if

$$\mathbb{P}(M_B(x_n) \in [a, b]) = b - a$$

for all $[a, b] \subseteq [0, 1]$.

Table of Contents

1. Introduction: Benford's Law
2. Our Problem: Stick Breaking
3. Results
4. Further Directions

Basic Stick Breaking Model

Start with a stick of length L. Choose a random point on the stick to break it in two, and repeat the process on each new stick obtained.

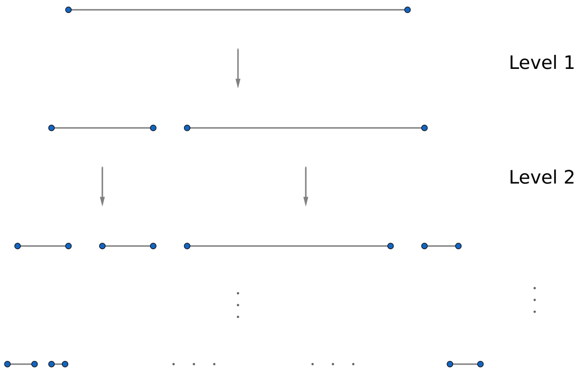


Figure 1: Illustration of stick breaking

Motivation from Physics

This process and its variations may be of interest to nuclear physicists for modelling particle decay ([2]).

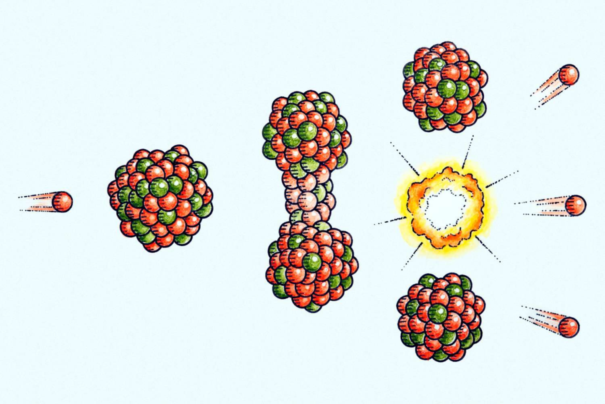


Figure 2: Random Stick Breaking Shares Similarities with Nuclear Fission

Figure source: https://www.thoughtco.com/nuclear-fission-definition-and-examples-4065372

Previous Results: Unrestricted Continuous Breaking

Theorem ([3])

Fix some distribution $\mathcal{D}$ on $(0, 1)$ satisfying the Mellin transform condition[a].

Start with a stick of length L, and break it in two with ratio sampled from $\mathcal{D}$. If we repeat this on both fragments for N levels, then the final collection of stick lengths converges to strong Benford as $N \to \infty$.

[a] Precisely, this means that

$$\lim_{n\to\infty} \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \prod_{m=1}^{n} \mathcal{M}f_{\mathcal{D}}\left(1 - \frac{2\pi i\ell}{\log B}\right) = 0.$$

Previous Results: Discrete One-Side Breaking

Theorem ([3])

Start with a stick of integer length L. Choose an integer $X \in \{1, \cdots, L\}$ uniformly, and break off a fragment of length X. Repeat this process on the remaining stick $L - X$, until no more such breaking can be done. The final collection converges to strong Benford as $L \to \infty$.

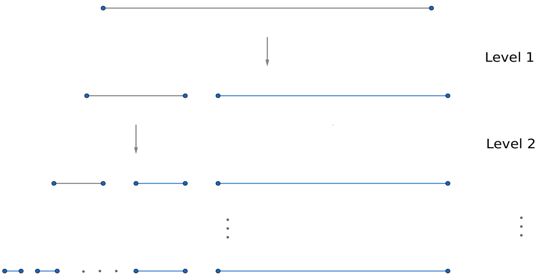


Figure 3: Illustration of discrete one-side breaking

Our Generalization: Discrete Breaking with Stopping Set

What if we break on both sides with extra *stopping conditions*?

Our Generalization: Discrete Breaking with Stopping Set

What if we break on both sides with extra *stopping conditions*?

More precisely:

Our Generalization: Discrete Breaking with Stopping Set

What if we break on both sides with extra *stopping conditions*?

More precisely:

- Define $\mathfrak{S} \subseteq \mathbb{Z}_+$ called *the stopping set*. Assume $1 \in \mathfrak{S}$.

Our Generalization: Discrete Breaking with Stopping Set

What if we break on both sides with extra *stopping conditions*?

More precisely:

- Define $\mathfrak{S} \subseteq \mathbb{Z}_+$ called *the stopping set*. Assume $1 \in \mathfrak{S}$.
- Declare a stick "dead" if its length falls into $\mathfrak{S}$ and do not break it further.

Our Generalization: Discrete Breaking with Stopping Set

What if we break on both sides with extra *stopping conditions*?

More precisely:

- Define $\mathfrak{S} \subseteq \mathbb{Z}_+$ called *the stopping set*. Assume $1 \in \mathfrak{S}$.
- Declare a stick "dead" if its length falls into $\mathfrak{S}$ and do not break it further.
- Continue until all sticks are dead.

Our Generalization: Discrete Breaking with Stopping Set

What if we break on both sides with extra *stopping conditions*?

More precisely:

- Define $\mathfrak{S} \subseteq \mathbb{Z}_+$ called *the stopping set*. Assume $1 \in \mathfrak{S}$.
- Declare a stick "dead" if its length falls into $\mathfrak{S}$ and do not break it further.
- Continue until all sticks are dead.

Question

Which sets $\mathfrak{S}$ would lead to strong Benford behavior as $L \to \infty$?

Table of Contents

1. Introduction: Benford's Law
2. Our Problem: Stick Breaking
3. Results
4. Further Directions

Result 1: Stop at Evens

Theorem (F.-Miller-S.-Verga, 2023)

Start with a stick of odd integer length L. Let the stopping set be $\mathfrak{S} = \{1\} \cup \{2m : m \in \mathbb{Z}_+\}$. Then the distribution of lengths of all dead sticks at the end approaches strong Benfordness as $L \to \infty$.

Result 1: Stop at Evens

Theorem (F.-Miller-S.-Verga, 2023)

Start with a stick of odd integer length L. Let the stopping set be $\mathfrak{S} = \{1\} \cup \{2m : m \in \mathbb{Z}_+\}$. Then the distribution of lengths of all dead sticks at the end approaches strong Benfordness as $L \to \infty$.

Question

Can we generalize this to other sets defined by residue classes mod *n*?

Simulation Results: Stop At Odds, Many Trials

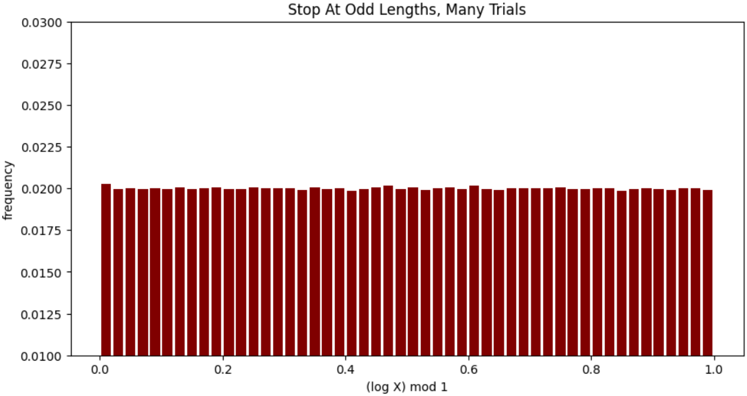


Figure 4: Histogram for $M_{10}(X)$, $L \approx 10^{1000}$, $R = 1000$ [1]

[1] R is the number of trials run with the same starting length L. The figure depicts the aggregated distribution of ending sticks from these trials.

Simulation Results: $n = 3$, stop at 1 residue class

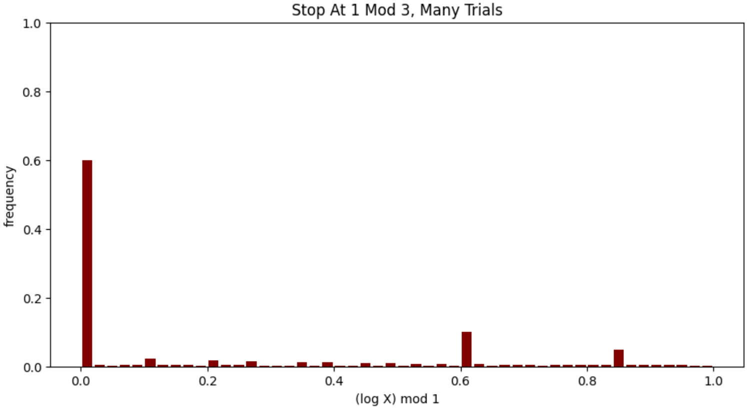


Figure 5: Histogram for $M_{10}(X)$, $L \approx 8 \cdot 10^{11}$, $R = 1000$

Simulation Results: $n = 3$, stop at 2 residue classes

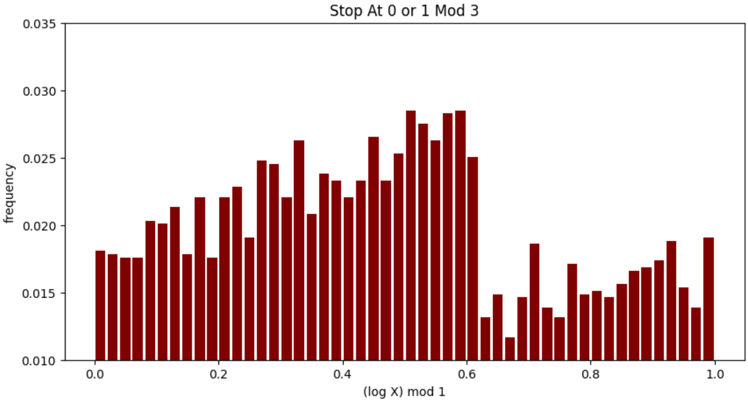


Figure 6: Histogram for $M_{10}(X)$, $L \approx 4 \cdot 10^{502}$, $R = 1000$

Simulation Results: $n = 4$, stop at 2 residue classes

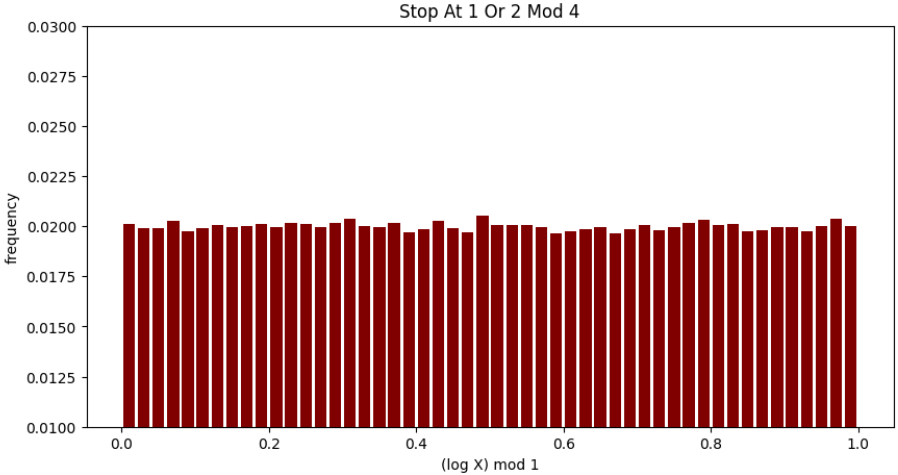


Figure 7: Histogram for $M_{10}(X)$, $L \approx 4 \cdot 10^{502}$, $R = 1000$

Result 2: Stop at Half of the Residues

After lots of trials up to $n = 10$...

Result 2: Stop at Half of the Residues

After lots of trials up to $n = 10$...

Theorem (F.-Miller-S.-Verga, 2023)

Fix an even modulus $n \geq 2$ *and a subset* $S \subset \{0, \ldots, n-1\}$ *of size* $n/2$*. Let the stopping set be*

$$\mathfrak{S} \;:=\; \{1\} \cup \{m \in \mathbb{Z}_+ : m = qn + r,\ \ r \in S, q \in \mathbb{Z}\}.$$

Result 2: Stop at Half of the Residues

After lots of trials up to $n = 10$...

Theorem (F.-Miller-S.-Verga, 2023)

Fix an even modulus $n \geq 2$ and a subset $S \subset \{0, \dots, n-1\}$ of size $n/2$. Let the stopping set be

$$\mathfrak{S} \ := \ \{1\} \cup \{m \in \mathbb{Z}_+ : m = qn + r,\ \ r \in S, q \in \mathbb{Z}\}.$$

If we start with R identical sticks of positive integer length $L \notin \mathfrak{S}$, then the collection of ending stick lengths converges to strong Benford behavior given that $R > (\log L)^3$ as $L \to \infty$.

Idea of Proof

1. Approximate the discrete process with a continuous analogue.

Idea of Proof

1. Approximate the discrete process with a continuous analogue.
2. Show that the continuous analogue results in strong Benford behavior. (Easier!) [Key Input]

Idea of Proof

1. Approximate the discrete process with a continuous analogue.
2. Show that the continuous analogue results in strong Benford behavior. (Easier!) [Key Input]
3. Deduce that the discrete process also results in strong Benford behavior by showing they are "close" enough. [Key Lemma]

Idea of Proof

1. Approximate the discrete process with a continuous analogue.
2. Show that the continuous analogue results in strong Benford behavior. (Easier!) [Key Input]
3. Deduce that the discrete process also results in strong Benford behavior by showing they are "close" enough. [Key Lemma]

Our contribution:

1. Generalize and adapt the "continuous approximation" strategy
2. Prove the Key Input
3. Give a new proof of the Key Lemma

Key Input: Continuous Fragmentation with Stopping Ratio

The continuous analogue is as follows.

Key Input: Continuous Fragmentation with Stopping Ratio

The continuous analogue is as follows.

- Start from R active sticks of length $L > 0$.

Key Input: Continuous Fragmentation with Stopping Ratio

The continuous analogue is as follows.

- Start from R active sticks of length $L > 0$.
- Each new stick dies with probability $1/2$. We allow the dying of sticks to be non-independent!

Key Input: Continuous Fragmentation with Stopping Ratio

The continuous analogue is as follows.

- Start from R active sticks of length $L > 0$.
- Each new stick dies with probability $1/2$. We allow the dying of sticks to be non-independent!
- Repeat until all sticks die.

Key Input: Continuous Fragmentation with Stopping Ratio

The continuous analogue is as follows.

- Start from R active sticks of length $L > 0$.
- Each new stick dies with probability $1/2$. We allow the dying of sticks to be non-independent!
- Repeat until all sticks die.

Theorem (F.-Miller-S.-Verga, 2023)

The above process ends in finitely many levels with probability 1, and the collection of ending stick lengths almost surely converges to strong Benford behavior as $R \to \infty$.

When $|S| < n/2$: Non-Benford!

Theorem (F.-Miller-S.-Verga, 2023)

If $|S| < n/2$, then as $R \to \infty$ and $L \to \infty$, the collection of mantissas of ending stick lengths does not converge to any continuous distribution on $[0, 1]$. *In particular, it does not converge to strong Benford behavior.*

When $|S| < n/2$: Non-Benford!

Theorem (F.-Miller-S.-Verga, 2023)

If $|S| < n/2$, then as $R \to \infty$ and $L \to \infty$, the collection of mantissas of ending stick lengths does not converge to any continuous distribution on $[0, 1]$. *In particular, it does not converge to strong Benford behavior.*

Table of Contents

1. Introduction: Benford's Law
2. Our Problem: Stick Breaking
3. Results
4. Further Directions

Future Work: $|S| > n/2$?

Conjecture

When $|S| > n/2$, the result does not converge to strong Benford.

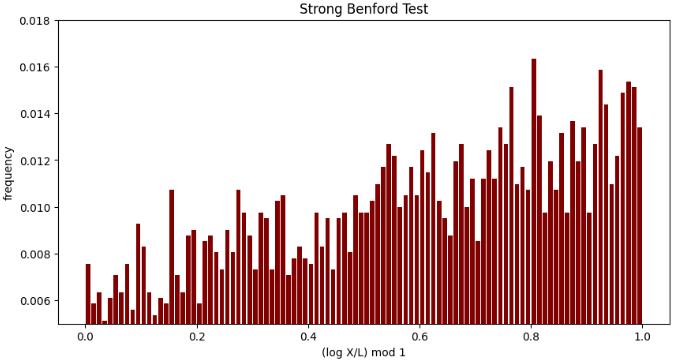


Figure 8: Stop at 8 Residue Classes Mod 12, $L = 82 \cdot 10^{12000}$, $R = 1000$

Future Work: General Number of Parts?

Future Work: General Number of Parts?

In fact, we have a more general version of the Key Input.

Theorem (F.-Miller-S.-Verga, 2023)

Fix some $k \geq 2$. Consider the continuous breaking process in which we start from R sticks of length L, break each stick into k pieces, and let a new stick die with probability $1 - 1/k$. The process ends in finitely many levels with probability 1, and the collection of ending stick lengths almost surely converges to strong Benford behavior as $R \to \infty$.

Future Work: General Number of Parts?

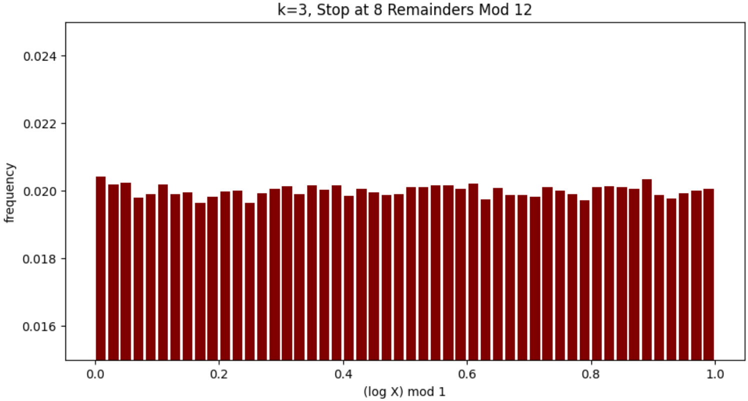
Conjecture

Fix $k \geq 2$. If we break each active stick into k pieces and stop at $(k-1)n/k$ residue classes modulo n, where n is a multiple of k, then the result converges to strong Benford.

Future Work: General Number of Parts?

Conjecture

Fix $k \geq 2$. If we break each active stick into k pieces and stop at $(k-1)n/k$ residue classes modulo n, where n is a multiple of k, then the result converges to strong Benford.



Thank you!

Work supported by NSF Grants DMS2241623 and DMS1947438, Williams College, and University of Michigan.

(1) A. V. Kontorovich and S. J. Miller, *Acta Arithmetica*, 2005, **120**, 269–297.

(2) J.-C. Pain, *Physical Review E*, 2008, **77**, DOI: 10.1103/physreve.77.012102.

(3) T. Becker, D. Burt, T. C. Corcoran, A. Greaves-Tunnell, J. R. Iafrate, J. Jing, S. J. Miller, J. D. Porfilio, R. Ronan, J. Samranvedhya, F. W. Strauch and B. Talbut, *Ann. Physics*, 2018, **388**, 350–381.