# Randomness in Number Theory

Distribution of Leading Digits of Arithmetic Sequences

Zhaodong Cai, Matthew Faust, Yuan Zhang
Project Leader: Junxian Li
Faculty Mentor: A.J. Hildebrand

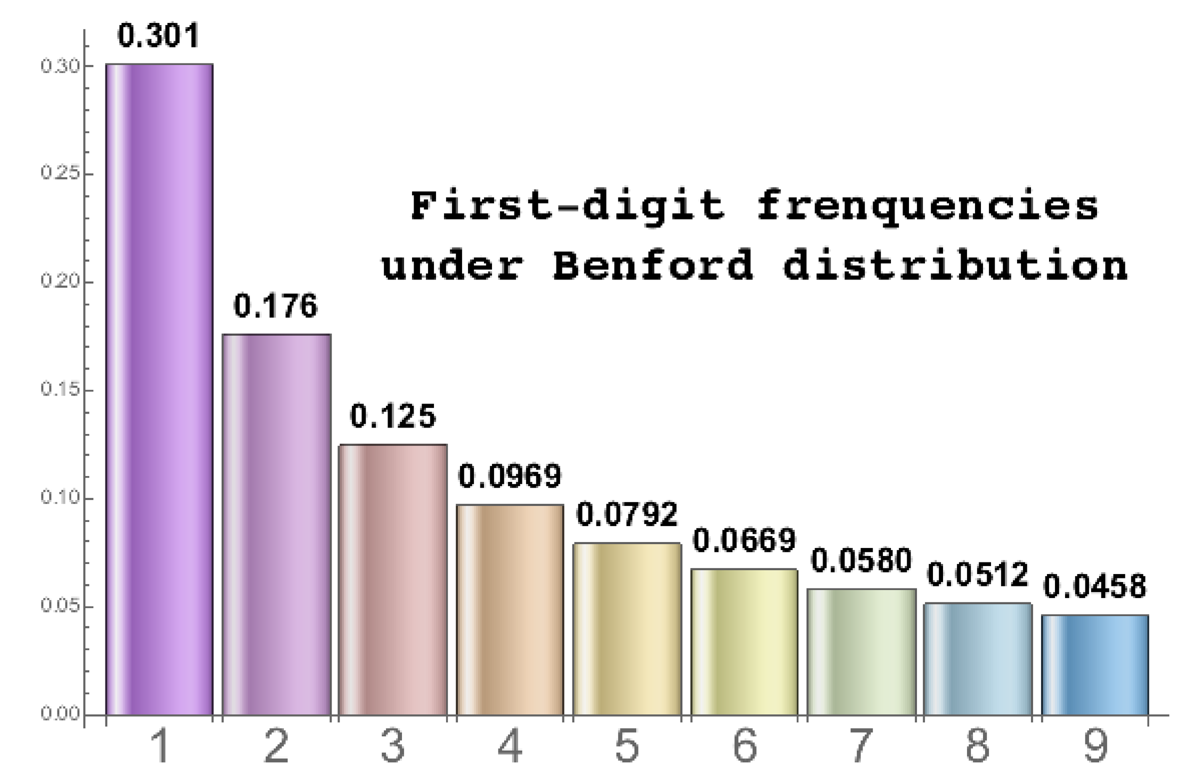
## Benford's Law

A sequence $\{x_n\}$ is said to be **Benford distributed** if it satisfies

$$P(\text{first digit } d) = \log_{10}\left(1 + \frac{1}{d}\right)$$

**Real world examples satisfying (approx.) Benford's Law:**

- Populations of US cities
- Areas of countries
- Physical constants
- File sizes in Linux file system




## Benford's Law in Mathematics

**Leading Digits of** $\{2^n\}$: 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, 32768, 65536, 131072, 262144, 524288, 1048576.

The table below shows the actual counts and Benford predictions ($10^9 \log_{10}(1 + \frac{1}{d})$) for the leading digits of the **first billion terms in** $\{2^n\}$.

| First Digit | Actual count | Benford Prediction | Error |
|---|---|---|---|
| 1 | 301029995 | 301029995.66 | −0.66 |
| 2 | 176091267 | 176091259.06 | +7.94 |
| 3 | 124938729 | 124938736.61 | −7.61 |
| 4 | 96910014 | 96910013.01 | +0.99 |
| 5 | 79181253 | 79181246.05 | +6.95 |
| 6 | 66946788 | 66946789.63 | −1.63 |
| 7 | 57991941 | 57991946.98 | −5.98 |
| 8 | 51152528 | 51152522.45 | +5.55 |
| 9 | 45757485 | 45757490.56 | −5.56 |

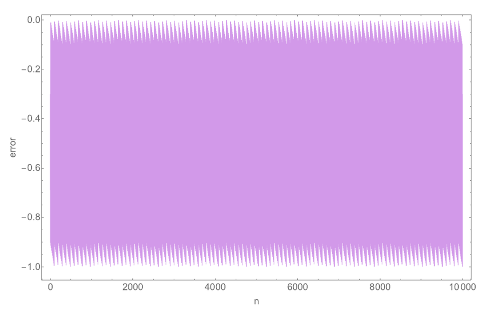
## Behavior and Distribution of Benford Errors

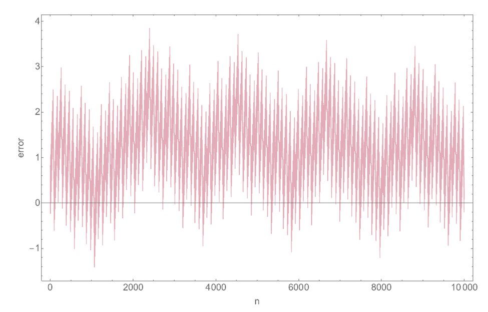
**Definition:** The **Benford error** of a sequence $\{a_n\}$ for digit $d$ up to $N$ is

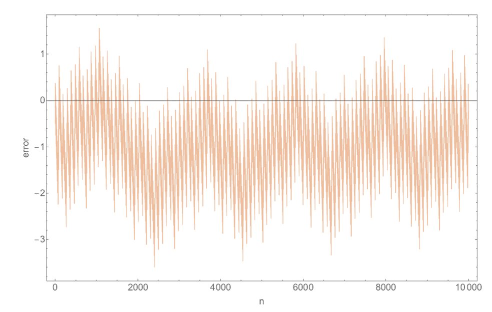
$$E_d(\{a_n\}, N) = \#\{n \le N : a_n \text{ has first digit } d\} - N \cdot \log_{10}(1 + 1/d).$$

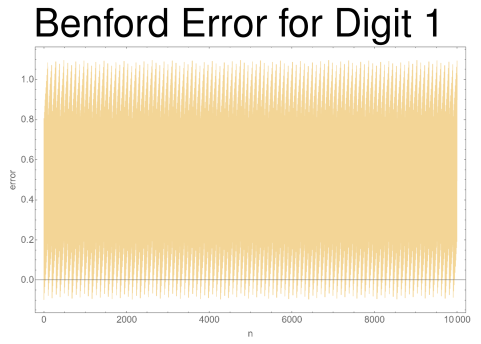
## Benford Errors for the Sequence $\{2^n\}$

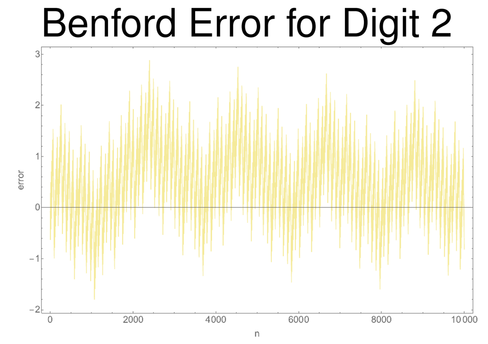
The graphs below are the plots of $E_d(\{2^n\}, N)$ for $1 \le N \le 10000$, and $d = 1, 2, \ldots, 9$. Observe the fractal and almost periodic features for digits $2, 3, 5, 6, 7, 8, 9$.


Benford Error for Digit 1

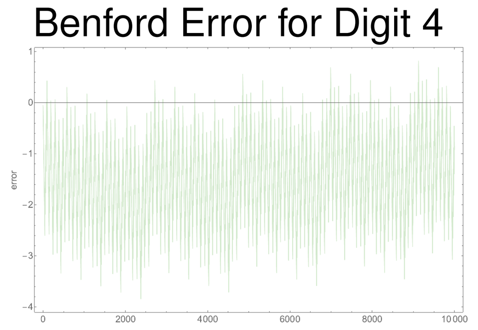

Benford Error for Digit 2

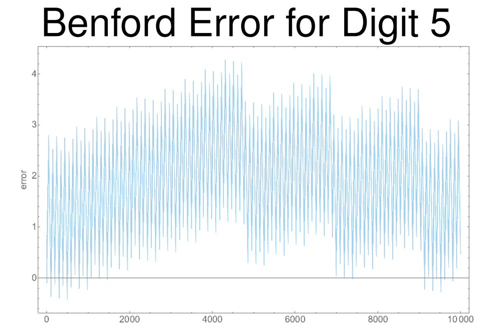

Benford Error for Digit 3

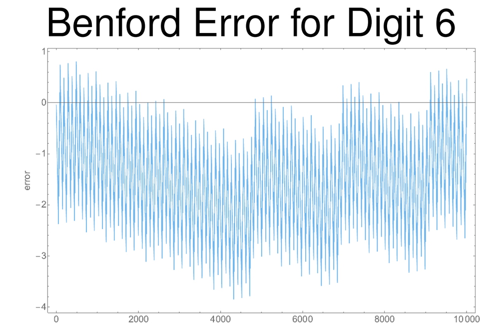

Benford Error for Digit 4


Benford Error for Digit 5


Benford Error for Digit 6

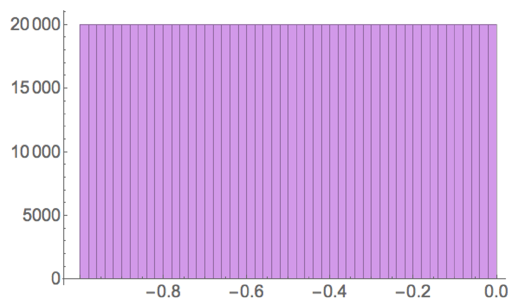

Benford Error for Digit 7

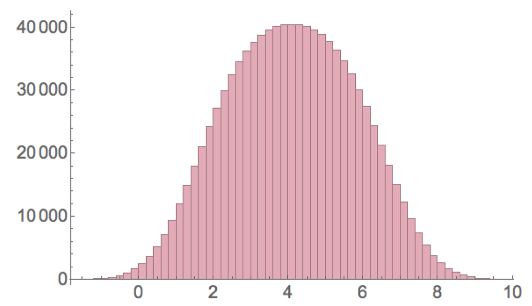

Benford Error for Digit 8

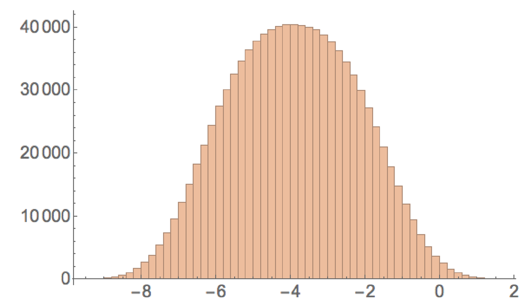

Benford Error for Digit 9

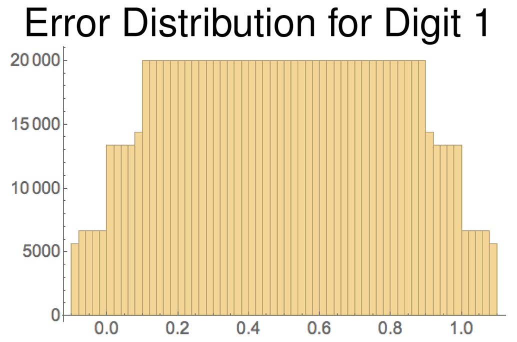
## Distribution of Benford Errors for the Sequence $\{2^n\}$

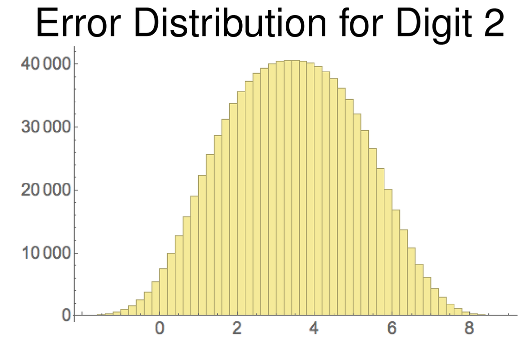
The graphs below are the histograms of $E_d(\{2^n\}, N)$ for $1 \le N \le 10^9$, and $d = 1, 2, \ldots, 9$. Observe the normal shape of the distribution for digits $2, 3, 5, 6, 7, 8, 9$.

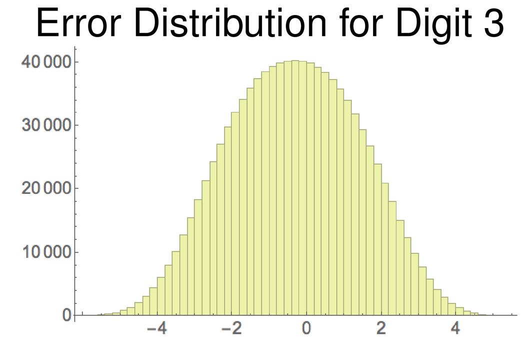

Error Distribution for Digit 1

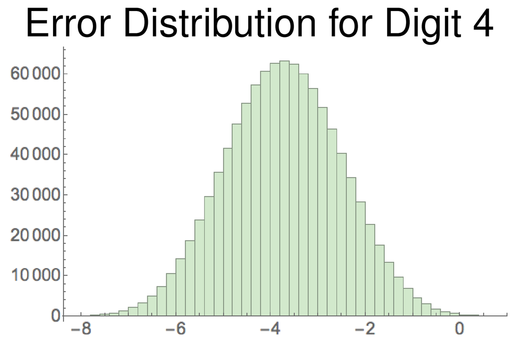

Error Distribution for Digit 2

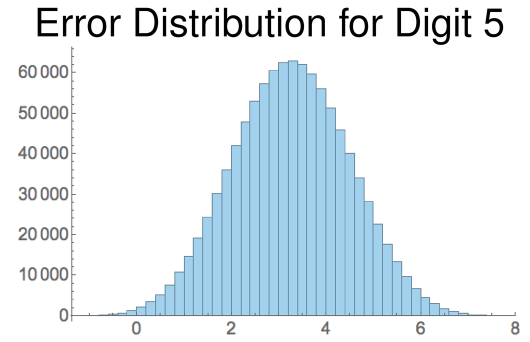

Error Distribution for Digit 3

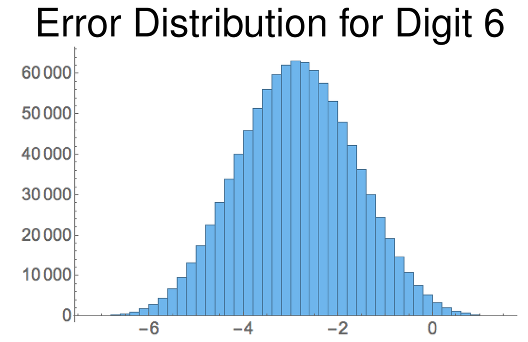

Error Distribution for Digit 4


Error Distribution for Digit 5


Error Distribution for Digit 6


Error Distribution for Digit 7


Error Distribution for Digit 8


Error Distribution for Digit 9

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

## Experimental Results

- As $N \to \infty$, the Benford error $E_1(\{2^n\}, N)$ for digit 1 is uniformly distributed on $(-1, 0)$.
- For all other digits $d \ne 1$, there exist infinitely many $N$ such that $|E_d(\{2^n\}, N)| > 1$.
- For all digits $d \ne 1, 4$ as $N \to \infty$, the Benford error $E_d(\{2^n\}, N)$ is unbounded and normally distributed.
- For all other sequences of the form $\{a^n\}$, and all $d = 1, \ldots, 9$, there exist infinitely many $N$ such that $|E_d(\{a^n\}, N)| > 1$.

## Theoretical Results

- For the sequence $\{2^n\}$ and digit 1, the Benford error is always between $-1$ and 0, i.e., $E_1(\{2^n\}, N) \in (-1, 0)$. In particular, the number of terms with digit 1 up to $N$ is exactly $\lfloor NP(1) \rfloor$, the Benford prediction.
- The distribution of $E_1(\{2^n\}, N)$ approaches the uniform distribution on $(-1, 0)$, as $N$ goes to infinity.

## Conjectures

- For $d \ne 1, 4$, $E_d(\{2^n\}, N)$ obeys normal distribution.
- $E_d(\{2^n\}, N)$ is almost periodic, with the almost periods being the denominators of convergents of $\log_{10} 2$.
- There exist sequences of the form $\{a^n\}$, whose limiting distribution of Benford Errors is neither normal nor uniform.

## Future Directions

- Investigate the distribution of Benford Error for sequences such as
  - $\{a^{f(n)}\}$, where $f$ is a polynomial
  - $\{n!\}$
  - $\{2^{p_n} - 1\}$, where $p_n$ is the $n$-th prime
- For sequences of the form $\{a^n\}$, find an explicit connection between continued fractions of $\{\log_{10} a\}$ and the distribution of the Benford Error.

## References

- J. Beck, Probabilistic Diophantine Approximation, Springer (2014)
- F. Benford, The Law of Anomalous Numbers, Proc. Amer. Philos. Soc., 78(4) (1938), pp. 551- 572.
- P. Diaconis, The Distribution of Leading Digits and Uniform Distribution Mod 1, Annals of Probability, Vol. 5, No. 1 (1977), pp. 72-81.
- T.P. Hill, The Significant-digit Phenomenon, Amer. Math. Monthly, Vol 102, No.4 (1995), pp. 322-327.
- L. Kuipers, L. Niederreiter, Uniform Distribution of Sequences, Dover Publications (2006).

*Support for this project was provided by the Illinois Geometry Lab and the Department of Mathematics at the University of Illinois at Urbana-Champaign.*