Randomness in Number Theory: Leading Digits Distribution

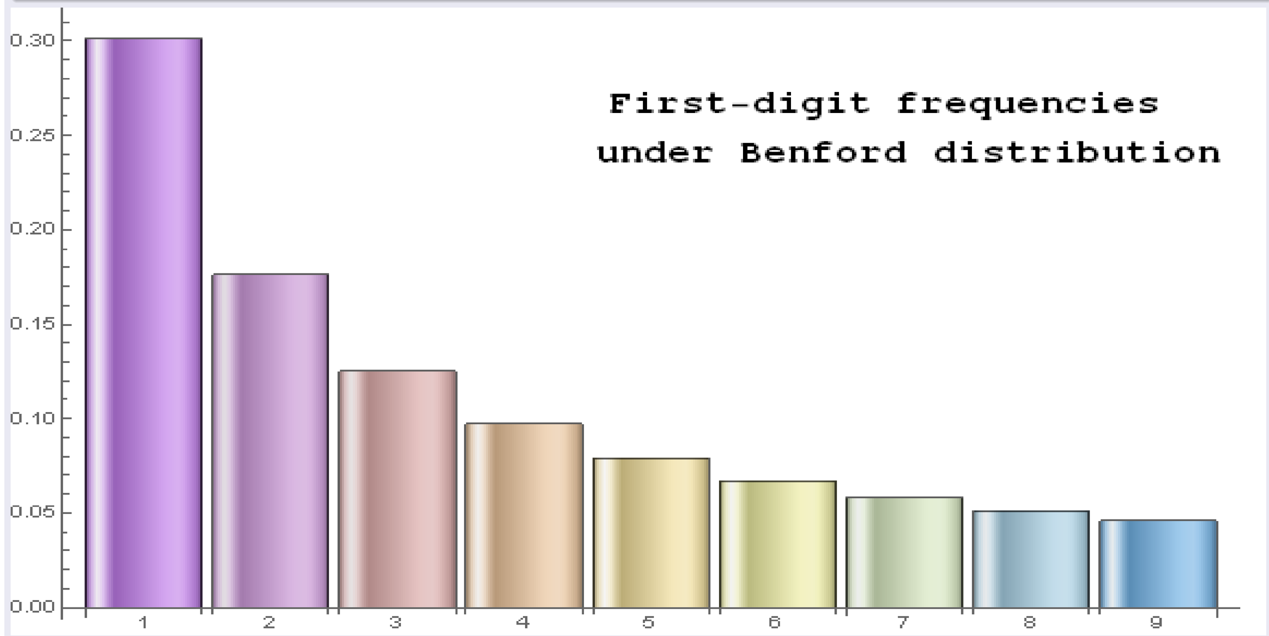
Zhaodong Cai, Yewen Fan, Kevin Kwan, Yuda Wang, Junxian Li (Project Leader), A.J. Hildebrand (Faculty Mentor)

ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN   Illinois Geometry Lab

IGL Open House, Dec 10, 2015

Benford's Law

Benford's Law *A sequence $\{x_n\}$ is said to be Benford distributed if the probability of having first digit d is $\log_{10}(1+\frac{1}{d})$ for any $d = 1, 2, \ldots, 9$.*



- **Mathematical Examples** a^n, $n!$, n^n, F_n, $p(n)$ (partition function) are Benford distributed (Diaconis (1977), Hill (1995)).
- **Real World Examples** Populations of US cities, areas of countries, physical constants, file sizes in Linux file system and numbers in US tax returns are approximately Benford distributed.

Uniform Distribution Modulo 1

- **Definition** $\{x_n\}$ is said to be uniformly distributed modulo 1 if
$$\lim_{N\to\infty} \frac{1}{N} \#\{1 \le n \le N : x_n \bmod 1 \in [a,b]\} = b - a$$
for any $0 \le a \le b \le 1$.
- **Examples** $n\alpha$, $n\log(n)$, $n!$, αn^k, where $\alpha \notin \mathbb{Q}$, are uniformly distributed modulo 1.
- **Connection with Benford Distribution** A sequence $\{a_n\}$ is Benford-distributed if $\{\log(a_n)\}$ is uniformly distributed modulo 1. (Kuipers and Niederreiter (2006))
- **Discrepancy** The discrepancy of the sequence $\{x_n\}$, $n = 1 \ldots N$ is defined to be
$$D_N = \sup_{0\le a\le b\le 1} \left| \frac{1}{N} \#\{1 \le n \le N : x_n \bmod 1 \in [a,b]\} - (b-a) \right|$$

References

- 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,The Annals of Probability, Vol. 5, No. 1 (1977), pp.72-81.
- T.P. Hill, The Significant-digit Phenomenon, The Amer. Math. Monthly, Vol 102, No.4 (1995), pp.322-327.
- L. Kuipers, L. Niederreiter, Uniform Distribution of Sequences, Dover Publications (2006), pp.7-23, 121-126.

Main Results and Conjectures

Global Randomness

Definition (Benford Deviation)

We define the *Benford deviation* $\Delta(N)$ of a sequence $\{a_n\}$ as
$$\Delta(N) = \Delta(N, \{a_n\}) = \sum_{i=1}^{9} |P_i(N) - \hat{P}_i|,$$
where $P_i(N)$ is the observed frequency of digit i among the first N terms of sequence $\{a_n\}$ and $\hat{P}_i = \log_{10}(1+\frac{1}{i})$ is the Benford frequency.

Theorem (Benford Deviation and Discrepancy)

The Benford deviation of a sequence $\{a_n\}$ is bounded by the discrepancy of the sequence $\{\log_{10} a_n\}$:
$$\Delta(N, \{a_n\}) \ll D_N(\{\log_{10} a_n\}).$$

Corollary (Upper Bound for Geometric Sequences)

If $a_n = a^n$, where $\log_{10} a$ is an algebraic irrational number, then the Benford deviation satisfies
$$\Delta(N) \ll \frac{\log N}{N}.$$

Conjecture (Lower Bound for Geometric Sequences)

If $a_n = a^n$, then
$$\Delta(N) \ne o\left(\frac{\log N}{N}\right).$$

Local Randomness

Theorem (First-digit Coupon Collector Waiting Times)

(i) *If $a_n = a^n(1+o(1))$ where $\log_{10} a \notin \mathbb{Q}$, the sequence has bounded first-digit coupon collector waiting times.*

(ii) *If $\frac{a_{n+1}}{a_n} = n^k(1+o(1))$ for some $k > 0$, the sequence has unbounded first-digit coupon collector waiting times.*

Examples:
- $\{2^n\}$ and $\{F_n\}$ have bounded waiting times.
- $\{n!\}$, $\{n^n\}$, and $\{p(n)\}$, where p is the partition function, have unbounded waiting times.

Theorem (Non-Periodicity)

If $a_n = a^n(1+o(1))$, where $\log_{10} a \notin \mathbb{Q}$, the leading digit of the sequence is not periodic.

Definition (Local Benford Distribution)

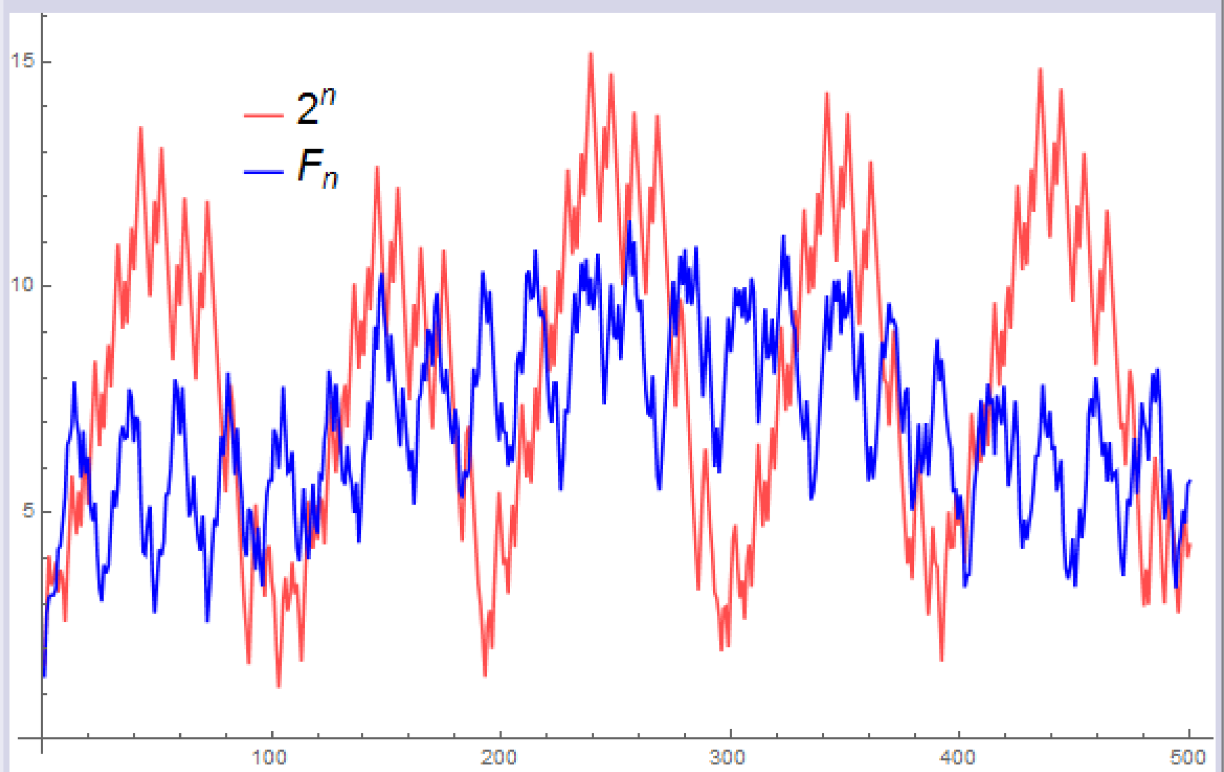
A sequence is called locally Benford-distributed to degree k if the leading digits of $(a_{n+1}, \ldots, a_{n+k})$ have the same asymptotic distribution as k independent Benford distributions.

Conjecture (Local Randomness Conjecture)

If $a_n = a^{p(n)}$, where $\log_{10} a \notin \mathbb{Q}$ and $p(n)$ is a polynomial of degree d, then a_n is locally Benford distributed to any degree $k \le d$, but not to degree $k+1$.

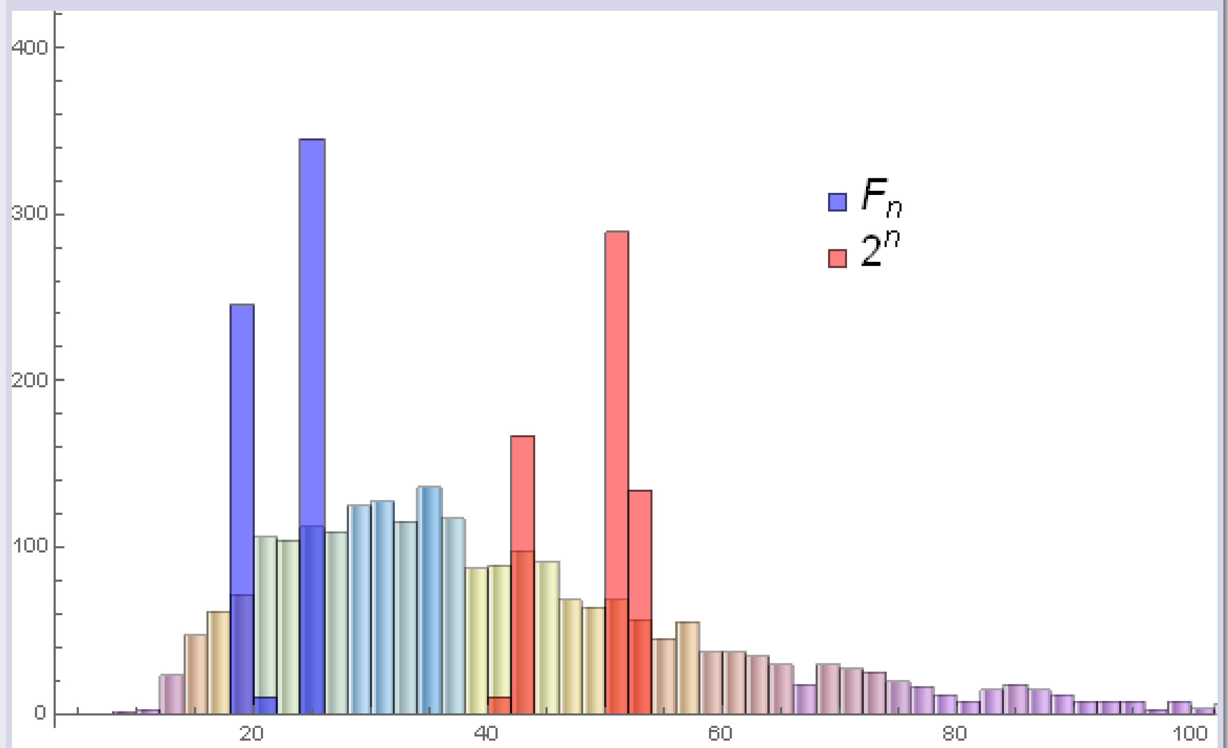
Further Observations

Fractal Behavior



For sequences $a_n = a^n$, where $\log_{10} a \notin \mathbb{Q}$, the Benford deviation $\Delta(N)$ shows a fractal-type behavior. The graph above shows $\Delta(N)$ for the case 2^n (red plot) and F_n (blue plot)

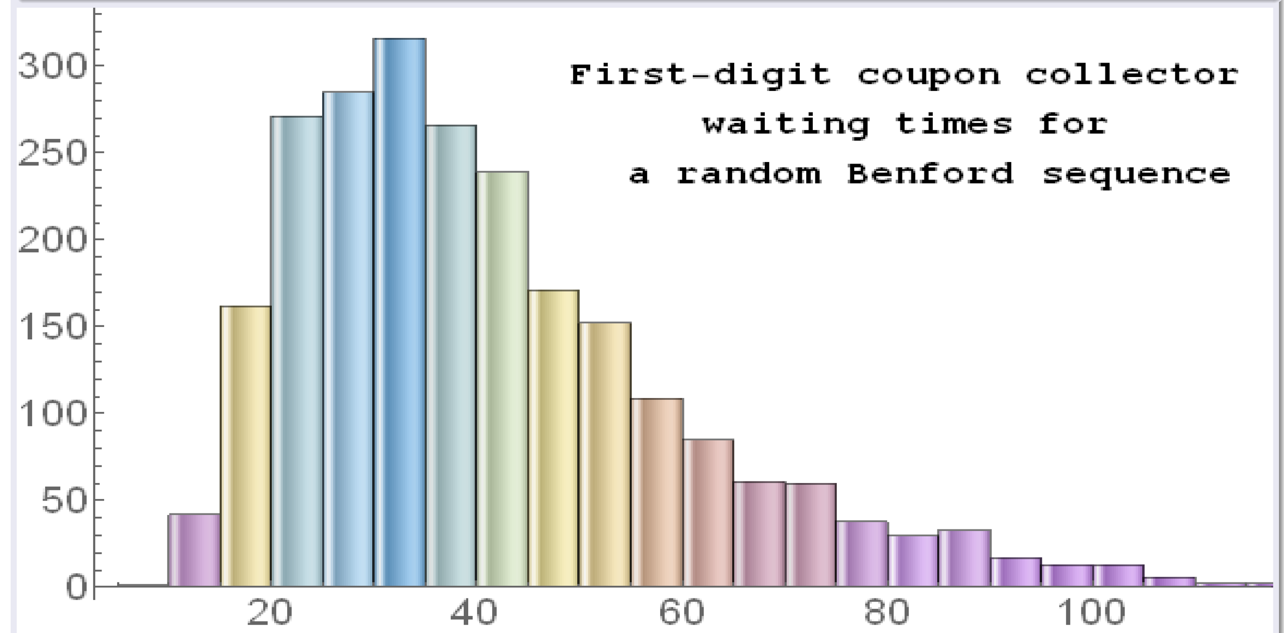
Waiting Time Distribution



For sequences $a_n = a^n$, where $\log_{10} a \notin \mathbb{Q}$, the distribution of coupon waiting times is very different from that of random simulations of Benford sequences. The graph above shows this distribution for the sequences F_n (in blue), 2^n (in red), and a random simulation of Benford sequences (in pastel).

The Coupon Collector Problem

Coupon Collector Problem *Given n coupons, each equally likely, how many coupons does one have to draw with replacement in order to obtain a complete collection of coupons (Dawkins, 1991; Schelling, 1954)?*



- **Coupon Collector Waiting Times** The sequence W_1, W_2, ... where W_1 is the number of draws needed to obtain a complete collection of coupons, W_2 the number of additional draws needed to obtain a second complete collection of coupons, and so on.
- **Expected Waiting Time** In the classical coupon collector problem, the expected waiting time is
$$E(W_1) = n \sum_{k=1}^{n} \frac{1}{k} \sim n \log n$$
as $n \to \infty$.
- **First-digit Coupon Collector Waiting Times** The coupon collector waiting times in the case when the coupons are digits $1, 2, \ldots, 9$, drawn according to the Benford distribution. The distribution of the first-digit coupon collector waiting times is shown in the above histogram.
- **Coupon Collector Randomness Test** A randomness test for a sequence of digits that compares the actual distribution of first-digit coupon waiting times with the theoretical one (Greenwood, 1955; Knuth, 1997).

References

- B. Dawkins, Siobhan's Problem: The Coupon Collector Revisited, The American Statistician, Vol. 45, No. 1 (Feb., 1991), pp. 76-82.
- R.E.Greenwood, Coupon Collector's Test for Random Digits, Math Tables Aids Compt. 9 (1955), pp.1-5.
- D.E. Knuth, Art of Computer Programming, Vol 2: Seminumerical Algorithms (3rd Ed.), Addison-Wesley Professional (1997), Sec. 3.3.
- H.V. Schelling, Coupon Collecting for Unequal Probabilities, Amer. Math. Monthly, 61 (1954), pp. 306-311.