Benfordness of Zeckendorf Decompositions

Patrick Dynes, Xixi Edelsbunner, Kimsy Tor
(Joint with Andrew Best, Brian McDonald, Madeleine Weinstein)

Advisors: Caroline Turnage-Butterbaugh and Steven J. Miller

Williams College, SMALL REU 2014

The Sixteenth International Conference on Fibonacci Numbers and Their Applications
Rochester, New York

July 25, 2014

Table of Contents

Elementary Facts

Fibonacci Recurrence

$$F_{n+1} = F_n + F_{n-1}$$
$$F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, \ldots$$

Elementary Facts

Fibonacci Recurrence

$$F_{n+1} = F_n + F_{n-1}$$
$$F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, \ldots$$

Binet's Formula

$$F_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}$$

A Beautiful Theorem

Theorem (Zeckendorf - 1939)

Every positive integer can be written uniquely as the sum of non-consecutive Fibonacci numbers.

A Beautiful Theorem

Theorem (Zeckendorf - 1939)

Every positive integer can be written uniquely as the sum of non-consecutive Fibonacci numbers.

Example

$$\begin{aligned} 12 &= 8 + 3 + 1 \\ &= F_5 + F_3 + F_1 \\ 2016 &= 1597 + 377 + 34 + 8 \\ &= F_{16} + F_{13} + F_8 + F_5 \end{aligned}$$

Counting the Summands

Theorem (Lekkerkerker - 1952)

The average number of Fibonacci summands in the Zeckendorf decomposition for integers in $[F_n, F_{n+1})$ tends to

$$\frac{n}{\varphi^2 + 1} \approx .276n \qquad (1)$$

where $\varphi = \frac{1+\sqrt{5}}{2}$ is the golden mean.

An Inquiry

Question

For a given set of numbers, how often would you expect a leading digit of 1, 2, 3, 4, . . . , 9 to occur?

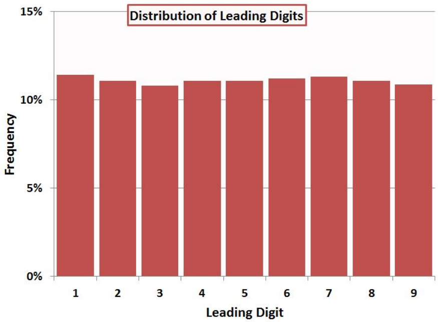
Perhaps, we expect something more uniform like this?

An Inquiry

Question

For a given set of numbers, how often would you expect a leading digit of 1, 2, 3, 4, . . . , 9 to occur?

Perhaps, we expect something more uniform like this?

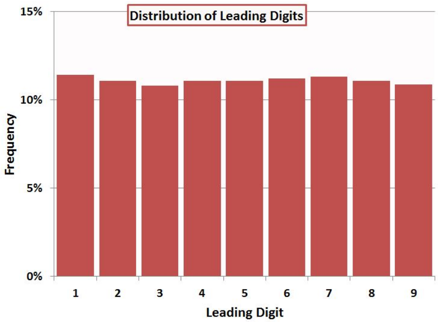


An Inquiry

Question

For a given set of numbers, how often would you expect a leading digit of 1, 2, 3, 4, . . . , 9 to occur?

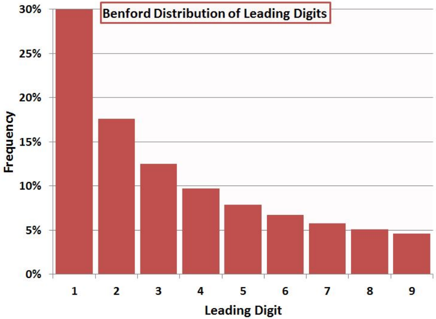
Perhaps, we expect something more uniform like this?



Not Quite!

The Benford Distribution

In fact, it's more like this!



Note

A leading digit 1 occurs with 30% frequency, while a leading digit 9 occurs with only 4.5 % frequency.

History of Benford's Law

- Benford's law is named after physicist Frank Benford in 1938.
- Although it was discovered earlier by Simon Newcomb in 1881.

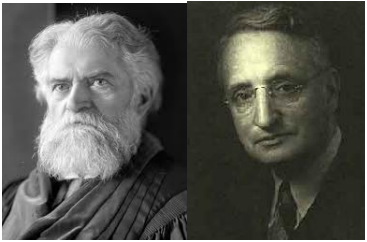


Figure: Newcomb and Benford

Benford's law

Definition (Benford's Law in Arbitrary Base)

A dataset is said to follow Benford's Law (base B) if the probability of observing a first digit of d is

$$\log_B\left(1+\frac{1}{d}\right)$$

Example

$$\mathbb{P}(d=1)=\log_{10}\left(1+\frac{1}{1}\right)\approx 0.301$$

$$\mathbb{P}(d=9)=\log_{10}\left(1+\frac{1}{9}\right)\approx 0.045$$

First Digit Bias

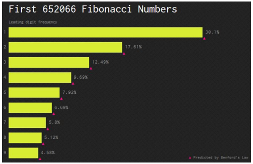


Figure: Fibonacci numbers

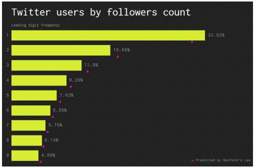


Figure: # Twitter followers

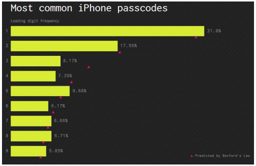


Figure: iPhone passwords

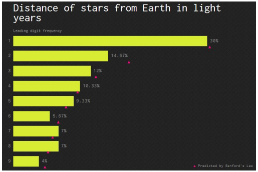


Figure: Distance of stars from Earth

Benford's Law: Applications

- $3x + 1$ problem.
- Analyzing round-off errors.
- Determining the optimal way to store numbers.
- Detecting tax fraud and data integrity.

Known Results

- Fibonacci numbers are Benford.
- Not every recurrence relation is Benford.

Example

$$a_{n+2} = 2a_{n+1} - a_n, \text{ with } a_1 = a_2 = 1$$

Sequence: $\{1, 1, 1, 1, 1, 1, 1, 1, 1, \dots\}$ is definitely not Benford.

New Results: SMALL 2014

- **Benfordness in Interval**: The distribution of the summands in the Zeckendorf decompositions, averaged over the entire interval $[F_n, F_{n+1})$.
- **Random Decomposition**: If we choose each Fibonacci number with probability q, disallowing the choice of two consecutive Fibonacci numbers, the resulting sequence follows Benford's law.
- **Benfordness of Decomposition**: If we pick a random integer M in the interval $[0, F_{n+1})$, then as $n \to \infty$, its Zeckendorf decomposition will follow Benford's Law with high probabiltiy.

Table of Contents

Theorem 1 (SMALL 2014): Benfordness in Interval

The distribution of the summands in the Zeckendorf decompositions, averaged over the entire interval $[F_n, F_{n+1})$, follows Benford's Law.

Theorem 1 (SMALL 2014): Benfordness in Interval

The distribution of the summands in the Zeckendorf decompositions, averaged over the entire interval $[F_n, F_{n+1})$, follows Benford's Law.

Example

Looking at the interval $[F_5, F_6) = [8, 13)$

$$\begin{aligned} 8 &= 8 && = F_5 \\ 9 &= 8 + 1 && = F_5 + F_1 \\ 10 &= 8 + 2 && = F_5 + F_2 \\ 11 &= 8 + 3 && = F_5 + F_3 \\ 12 &= 8 + 3 + 1 && = F_5 + F_3 + F_1 \end{aligned}$$

Proof of Theorem 1

Let S be a subset of the Fibonacci numbers. Let $q(S,n)$ be the density of S over the Fibonacci numbers in the interval $[1, F_n]$. That is

$$q(S,n) = \frac{\#\{F_j \in S \mid 1 \leq j \leq n\}}{n}$$

In the case that $\lim_{n\to\infty} q(S,n)$ exists, we define the asymptotic density $q(S)$ as

$$q(S) = \lim_{n\to\infty} q(S,n)$$

Let S_d be the subset of the Fibonacci numbers which share a fixed digit d where $1 \leq d < B$. It is a well-known result that

Theorem

$$q(S_d) = \lim_{n\to\infty} q(S_d, n) = \log_B\left(1 + \frac{1}{d}\right)$$

i.e. the Fibonacci numbers are Benford!

Consider integers in the interval $I_n := [F_n, F_{n+1})$. Let $X(I_n)$ be a random variable, which takes Fibonacci numbers in $[F_1, F_n)$, weighted by how often they occur in decompositions in I_n. Then,

$$P\{X(I_n) = F_k\} := \begin{cases} \frac{F_{k-1}F_{n-k-2}}{\mu_n F_{n-1}}, & \text{if } 1 \leq k \leq n-2 \\ \frac{1}{\mu_n}, & \text{if k = n} \\ 0, & \text{otherwise} \end{cases}$$

where μ_n is the average number of summands in Zeckendorf decompositions of integers in the interval $[F_n, F_{n+1})$.

We compute an approximation for $P\{X(I_n) = F_k\}$ when $1 \leq k \leq n-2$.

$$P\{X(I_n) = F_k\} = \frac{1}{\mu_n \phi \sqrt{5}} + O\left(\phi^{-2k} + \phi^{-2n+2k}\right)$$

Then, for a fixed integer r, we compute the approximation

$$\sum_{r<k<n-r} P\{X(I_n) = F_k\} = 1 - r \cdot O\left(\frac{1}{n}\right)$$

Noting that the values of r for which the sum is vacuous do not hurt our estimates.

Set $r := \left\lfloor \frac{\log n}{\log \phi} \right\rfloor$. With these estimates, we may now compute the density of S over the Zeckendorf summands in the interval $I_n = [F_n, F_{n+1})$

$$P\{X(I_n) \in S\} = \frac{nq(S)}{\mu_n \phi \sqrt{5}} + o(1)$$

In the limit, we have

$$\lim_{n \to \infty} P\{X(I_n) \in S\} = q(S)$$

Remark

- Stronger result than Benfordness of Zeckendorf summands.
- Global property of the Fibonacci numbers can be carried over locally into the Zeckendorf summands.
- If we have a subset of the Fibonacci numbers S with asymptotic density $q(S)$, then the density of the set S over the Zeckendorf summands will converge to this asymptotic density.

Table of Contents

Theorem 2 (SMALL 2014): Random Decomposition

If we choose each Fibonacci number with probability q, disallowing the choice of two consecutive Fibonacci numbers, the resulting sequence follows Benford's law.

Example: n = 10

$$F_1 + F_2 + F_3 + F_4 + F_5 + F_6 + F_7 + F_8 + F_9 + F_{10}$$
$$= 2 + 8 + 21 + 89$$
$$= 120$$

Choosing a Random Decomposition

Select a random subset A of the Fibonaccis in the following way. Given some $q \in (0, 1)$. Let $A_0 := \emptyset$.

For $n \geq 1$, if $F_{n-1} \in A_{n-1}$, let $A_n := A_{n-1}$. Otherwise, let A_n equal $A_{n-1} \cup \{F_n\}$ with probability q and A_{n-1} with probability $1 - q$.

Let $A := \bigcup_n A_n$.

Goal

We will prove that, with probability 1, A is Benford.

Goal

We will prove that, with probability 1, A is Benford.

Stronger claim: For any subset S of the Fibonaccis with density d in the Fibonaccis, with probability 1, $S \cap A$ will have density d in A.

The probability that $F_k \in A$

The probability that $F_k \in A$ is

Lemma

$$p_k = \frac{q}{1+q} + O(q^k).$$

Expected Value of X_n

Define $X_n := \#A_n$. Using elementary techniques, we get

Lemma

$$E[X_n] = \frac{nq}{1+q} + O(1)$$

Expected Value of X_n

$X_n := \#A_n$. Let

$$x_k := \begin{cases} 1 & \text{if} F_k \in A \\ 0 & \text{if} F_k \notin A. \end{cases}$$

Expected Value of X_n

$X_n \ := \ \#A_n$. Let

$$x_k \ := \ \begin{cases} 1 & \text{if} F_k \in A \\ 0 & \text{if} F_k \notin A. \end{cases}$$

And note that $X_n := \sum_{k=1}^{n} x_k$. Then

$$\begin{aligned} E[X_n] \ &= \ \sum_{k=1}^{n} E[x_k] \\ &= \ \sum_{k=1}^{n} p_k \\ &= \ \frac{nq}{1+q} + O(1). \end{aligned}$$

Variance of X_n

$X_n := \#A_n$. Via standard calculations, we get

Lemma

$$Var[X_n] = O(n).$$

By Chebyshev's inequality, we deduce that

Corollary

$$\begin{aligned} X_n &= E[X_n] + o(n) \\ &= \frac{nq}{1+q} + o(n) \end{aligned}$$

with probability $1 + o(1)$.

Expected Value of Y_n

Define $Y_{n,S} := \#A_n \cap S$. Using standard techniques, we get

Lemma

$$E[Y_n] = \frac{nqd}{1+q} + o(n).$$

Variance of $Y_{n,S}$

$$Y_{n,S} := \#A_n \cap S$$

Lemma

$$Var[Y_{n,S}] = o(n^2)$$

Corollary

$$Y_{n,S} = \frac{nqd}{1+q} + o(n)$$

with probability $1 + o(1)$

Lemma

$$\lim_{n\to\infty} \frac{Y_{n,S}}{X_n} = \lim_{n\to\infty} \frac{\#A_n \cap S}{\#A_n} = \lim_{n\to\infty} \frac{\frac{nqd}{1+q} + o(n)}{\frac{nq}{1+q} + o(n)} = d$$

with probability 1.

But by definition, this means that $A \cap S$ has density d in A. Therefore, our claim is proven.

Table of Contents

Theorem 3 (SMALL 2014): Benfordness of Decomposition

If we pick a random integer M in the interval $[0, F_{n+1})$, then as $n \to \infty$, its Zeckendorf decomposition will follow Benford's Law with high probabiltiy.

Proof of Theorem 3

Let M be an integer in $[0, F_{n+1})$ with decomposition $M = F_{a_1} + F_{a_2} + \cdots + F_{a_\ell}$. Then the probability that $M = \sum_{F_k \in A_n} F_k$

is

$$p_M = \begin{cases} q^\ell (1-q)^{n-2\ell} \text{if} & a_\ell \leq n \\ q^\ell (1-q)^{n-2\ell+1} \text{if} & a_\ell = n \end{cases}$$

Choosing $q = \frac{1}{\varphi^2}$, the previous formula simplifies to

Lemma

$$p_M = \begin{cases} \varphi^{-n} \text{if} & M \in [0, F_n) \\ \varphi^{-n-1} \text{if} & M \in [F_n, F_{n+1}) \end{cases}$$

We can prove that when selecting integers from $[0, F_{n+1})$ uniformly at random, for any $\varepsilon > 0$ the proportion of the summands of M which are in S will be within ε of S with probability $1 + o(1)$.

Table of Contents

References

- S. J. Miller (editor), *The Theory and Applications of Benford's Law*, Princeton University Press, to appear.
- L. C. Washington, *Benford's Law for Fibonacci and Lucas Numbers*, The Fibonacci Quarterly, 19.2, (1981), 175-177. http://www.fq.math.ca/Scanned/19-2/washington.pdf
- M. Kologlu, G. Kopp, S. J. Miller and Y. Wang, *On the Number of Summands in Zeckendorf Secompositions*, Fibonacci Quarterly 49 (2011), no. 2, 116-130. http://www.fq.math.ca/Papers/49-2/KologluKoppMillerWang.pdf
- J. Brown and R. Duncan, *Modulo One Uniform Distribution of the Sequence of Logarithms of Certain Recursive Sequences*, Fibonacci Quarterly **8** (1970) 482–486. , http://www.mathstat.dal.ca/FQ/Scanned/8-5/brown.pdf

Acknowledgements

We would like to thank our advisors, Caroline Turnage-Butterbaugh, Steven J. Miller and our co-authors Andrew Best, Brian McDonald and Madeleine Weinstein.

We would also like to thank the Williams College SMALL summer research program.

This research was funded by NSF grant DMS1347804, Williams College and the Clare Boothe Luce Program of the Henry Luce Foundation.

Contact

PATRICK DYNES, DEPARTMENT OF MATHEMATICAL SCIENCES
CLEMSON UNIVERSITY, CLEMSON, SC
Email address: pdynes@clemson.edu

XIXI EDELSBRUNNER, DEPARTMENT OF MATHEMATICS AND STATISTICS
WILLIAMS COLLEGE, WILLIAMSTOWN, MA
Email address: xe1@williams.edu

KIMSY TOR, DEPARTMENT OF MATHEMATICS
MANHATTAN COLLEGE, RIVERDALE, NY
Email address: ktor.student@manhattan.edu

Appendix

We know that the probability of choosing $M = F_{a_1} + F_{a_2} + \cdots + F_{a_\ell}$ is

p_M equals $q^\ell(1-q)^{n-2\ell}$ if $a_\ell \leq n$ and $q^\ell(1-q)^{n-2\ell+1}$ if $a_\ell = n$.

Choosing $q = \frac{1}{\varphi^2}$ gives

p_M equals φ^{-n} if $M \in [0, F_n)$ and φ^{-n-1} if $M \in [F_n, F_{n+1})$.

Let $e_n(x) = \left|\frac{\#D_M \cap S}{D_M} - d\right|$, where D_M is the decomposition of M. Let $E = \{M \in [0, F_{n+1}) : e(x) \geq \varepsilon\}$.
Let $B_n = \sum_{F_k \in A_n} F_k$, so that $p(M) = P(B_n = M)$. For sufficiently large n, we have $F_{n+1} \geq \frac{\phi^{n+1}}{\sqrt{5}}$. Now let M_n be a random variable selected uniformly at random from the integers in $[0, F_{n+1})$. Then

$$P(M_n \in E_n) = \frac{\#E_n}{F_{n+1}} = \sum_{M \in E_n} \frac{1}{F_{n+1}} \tag{2}$$

$$\leq \sqrt{5} \sum_{M \in E_n} \phi^{-n-1} \leq \sqrt{5} \sum_{M \in E_n} p(M) \tag{3}$$

$$= \sqrt{5} P(B_n \in E_n) = o(1) \tag{4}$$

Therefore, we have proved that when selecting integers from $[0, F_{n+1})$ uniformly at random, for any $\varepsilon > 0$ the proportion of the summands of x which are in S will be within ϵ of S with probability $1 + o(1)$.