Benford Behavior of Generalized Zeckendorf Decompositions

Andrew Best, Patrick Dynes, Xixi Edelsbrunner, Brian McDonald, Steven J. Miller, Kimsy Tor, Caroline Turnage-Butterbaugh and Madeleine Weinstein

Abstract We prove connections between Zeckendorf decompositions and Benford's law. Recall that if we define the Fibonacci numbers by $F_1 = 1, F_2 = 2$ and $F_{n+1} = F_n + F_{n-1}$, every positive integer can be written uniquely as a sum of non-adjacent elements of this sequence; this is called the Zeckendorf decomposition, and similar unique decompositions exist for sequences arising from recurrence relations of the form $G_{n+1} = c_1 G_n + \cdots + c_L G_{n+1-L}$ with c_i positive and some other restrictions. Additionally, a set $S \subset \mathbb{Z}$ is said to satisfy Benford's law base 10 if the density of the elements in S with leading digit d is $\log_{10}(1+\frac{1}{d})$; in other words, smaller leading digits are more likely to occur. We prove that as $n \to \infty$ for a randomly selected integer m in $[0, G_{n+1})$ the distribution of the leading digits of the summands in its generalized Zeckendorf decomposition converges to Benford's law almost surely. Our results hold more generally: one obtains similar theorems to those regarding the distribution of leading digits when considering how often values in sets with density are attained in the summands in the decompositions.

Keywords: Zeckendorf decompositions, Fibonacci numbers, positive linear recurrence relations, Benford's law.

MSC 2010: 11B39, 11B05, 60F05 (primary) 11K06, 65Q30, 62E20 (secondary)

Department of Mathematics and Statistics, Williams College, ajb51@williams.edu, Department of Mathematical Sciences, Clemson University, pdynes@clemson.edu, Department of Mathematics and Statistics, Williams College, xe1@williams.edu, Department of Mathematics, University of Rochester, bmcdon11@u.rochester.edu, Department of Mathematics and Statistics, Williams College, sjm1@williams.edu, Department of Mathematics, Manhattan College, ktor.student@manhattan.edu, Department of Mathematics, North Dakota State University, cturnagebutterbaugh@gmail.com, and Department of Mathematics, Harvey Mudd College, mweinstein@hmc.edu.

1 Introduction

Zeckendorf's theorem states that every positive integer m can be written uniquely as a sum of nonconsecutive Fibonacci numbers, where the Fibonacci numbers are defined by $F_{n+1} = F_n + F_{n-1}$ with $F_1 = 1$ and $F_2 = 2$ (we must re-index the Fibonaccis, as if we included 0 or had two 1's we clearly could not have uniqueness). Such a sum is called the Zeckendorf decomposition of m, and each number in the sum is called a summand. Zeckendorf decompositions have been generalized to many other sequences, specifically those arising from positive linear recurrences. More generally, we can consider a positive linear recurrence sequence given by

$$G_{n+1} \;=\; c_1 G_n + \cdots + c_L G_{n+1-L}, \tag{1}$$

with c_i nonnegative, L, c_1 and c_L positive, as well as rules to specify the first L terms of the sequence and a generalization of the non-adjacency constraint to what is a legal decomposition. Unique decompositions exist both here and for other sequences; see [Al, Day, DDKMMV, DDKMV, DG, DT, FGNPT, GT, GTNP, Ke, KKMW, Len, MW1, MW2, Ste1, Ste2, Ze] for a sample of the vast literature on this topic.

Our purpose is to connect generalized Zeckendorf decompositions and Benford's law. In fact, what we show is more general, and the connection with Benford's law follows as a corollary. Still, Benford's law was the motivation for our investigation, so we discuss its history. First discovered by Simon Newcomb [New] in the 1880s, it was rediscovered by Benford [Ben] approximately fifty years later, who noticed that the distributions of the leading digits of numbers in many data sets were not uniform. In fact, there was a strong bias towards lower values. For example, the leading digit 1 appeared about 30% of the time and the leading digit 9 under 5% of the time. Data sets with such leading digit distributions are said to follow Benford's law. More precisely, the probability of a first digit base B of d is $\log_B(1+1/d)$, or more generally the probability that the significand[1] is at most s is $\log_B(s)$. Benford's law appears in astoundingly many data sets, from physical constants to census information to financial and behavioral data, and has a variety of applications (two of the most interesting being its use to detect accounting or voting fraud). This digit bias is in fact quite natural once one realizes that a data set will follow Benford's law if its logarithms modulo 1 are equidistributed.[2] See [BHi, Hi1, Hi2, MT-B, Rai] for more on the theory of Benford's law, as well as the edited volume [M] for a compilation of articles on its theory and applications.

Before exploring whether or not the summands in Zeckendorf decompositions obey Benford's law, it's natural to ask the question about the sequence of Fibonacci numbers themselves. The answer is yes, and follows almost immediately from Binet's formula,

[1] If $x > 0$ and $B > 1$ we may uniquely write x as $S_B(x) \cdot B^{k_B(x)}$, where $S_B(x) \in [1, B)$ is the significand of x and $k_B(x)$ is an integer.

[2] Given a data set $\{x_n\}$, let $y_n = \log_{10} x_n \bmod 1$. If $\{y_n\}$ is equidistributed modulo 1 then in the limit the percentage of the time it is in $[\alpha, \beta] \subset [0,1]$ is just $\beta - \alpha$. For example, to restrict to significands of d take $\alpha = \log_{10} d$ and $\beta = \log_{10}(d+1)$.

$$F_n \;=\; \frac{5+\sqrt{5}}{10}\left(\frac{1+\sqrt{5}}{2}\right)^n + \frac{5-\sqrt{5}}{10}\left(\frac{1-\sqrt{5}}{2}\right)^n \tag{2}$$

(note this is slightly different than the standard expression for Binet's formula as we have re-indexed our sequence so that the Fibonaccis begin 1, 2, 3, 5). The proof is completed by showing the logarithms modulo 1 are equidistributed, which is immediate from the irrationality of $\log_{10}(\frac{1+\sqrt{5}}{2})$ and Kronecker's theorem (if α is irrational then $n\alpha$ is equidistributed modulo 1) and simple book-keeping to bound the error of the secondary piece; see [DG, MT-B, Was] for details.

Instead of studying Benfordness of summands in Zeckendorf decompositions, we could instead look at other properties of the summands, such as how often we have an even number or how often they are a square modulo B for some fixed B. So long as our sequence has a positive density, our arguments will be applicable.[3] We quickly review this notion. Given a set of positive integers $\mathscr{G} = \{G_n\}_{n=1}^{\infty}$ and a subset $S \subset \mathscr{G}$, we let $q(S,n)$ be the fraction of elements of $\mathscr{G}$ with index at most n that are in S:

$$q(S,n) \;:=\; \frac{\#\{G_i \in S \;:\; 1 \le i \le n\}}{n}. \tag{3}$$

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

$$q(S) \;:=\; \lim_{n\to\infty} q(S,n), \tag{4}$$

and for brevity often say the sequence S has **density** $q(S)$.

In an earlier work we proved that if a set S has a positive density $q(S)$ in the Fibonaccis, then so too do its summands in the Zeckendorf decompositions, and in particular Zeckendorf decompositions using Fibonacci numbers follow Benford's law [BDEMMTTW]. Our main result below is generalizing these results to the case of a **positive linear recurrence sequence**, which is a sequence of positive integers $\{G_n\}_{n=1}^{\infty}$ and a set of non-negative coefficients $c_1,\dots,c_L$ with $L, c_1, c_L > 0$,

$$G_{n+1} \;=\; c_1 G_n + c_1 G_{n-1} + \cdots + c_L G_{n+1-L}, \tag{5}$$

and prescribed positive initial terms $G_1, G_2, \dots, G_L$.

Theorem 1.1 *Fix a positive linear recurrence sequence $\{G_n\}$. Let $S \subseteq \{G_n\}_{n=1}^{\infty}$ be a set with positive density $q(S)$, and fix an $\varepsilon > 0$. As $n \to \infty$, for an integer m selected uniformly at random from $[0, G_{n+1})$ the proportion of the summands in m's Zeckendorf decomposition which belong to S is within ε of $q(S)$ with probability $1 + o(1)$.*

We define some concepts needed to prove Theorem 1.1 in §2, in particular the notion of a super-legal decomposition. We derive some needed properties of these

[3] For example, in the limit one-third of the Fibonacci numbers are even. To see this we look at the sequence modulo 2 and find it is $1,0,1,1,0,1,1,0,1,\dots$; it is thus periodic with period 3 and one-third of the numbers are even.

decompositions, and then prove our main result in §3 by showing related random variables (the number of summands, and the number of summands in our set with positive density in our recurrence sequence) are strongly concentrated.

Acknowledgements: This research was conducted as part of the 2014 SMALL REU program at Williams College and was supported by NSF grants DMS 1347804 and DMS 1265673, Williams College, and the Clare Boothe Luce Program of the Henry Luce Foundation. It is a pleasure to thank them for their support, and the participants at SMALL and at the 16th International Conference on Fibonacci Numbers and their Applications for helpful discussions.

2 Legal and Super-Legal Decompositions

For the rest of the paper any positive linear recurrence sequence $\{G_n\}_{n=1}^{\infty}$ ***satisfies*** (5) ***with*** $c_i \geq 0$ ***and*** $L, c_1, c_L \geq 1$**.**

Let $\{G_n\}_{n=1}^{\infty}$ be a positive linear recurrence sequence. Its the characteristic polynomial is

$$f(\lambda) \;=\; c_0 + c_1\lambda + \cdots + c_{L-1}\lambda^{L-1} + c_L\lambda^L, \tag{6}$$

with roots $\lambda_1, \ldots, \lambda_L$. By the Generalized Binet Formula (for a proof see, for example, Appendix A of [BBGILMT]) we have λ_1 is the unique positive root, $\lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_L|$, and there exists an $A > 0$ such that

$$G_n \;=\; A\lambda_1^n + O(n^{L-2}\lambda_2^n). \tag{7}$$

We introduce a few important terms needed to state our results.

Definition 2.1 *Let* $\{G_n\}$ *be a positive linear recurrence sequence. Given non-negative integers* $a_1, \ldots, a_n$*, the sum* $\sum_{i=1}^{n} a_i G_{n+1-i}$ *is a* ***legal*** *Zeckendorf decomposition if one of the following conditions holds.*

1. *We have* $n < L$ *and* $a_i = c_i$ *for* $1 \leq i \leq n$*.*
2. *There exists an* $s \in \{1, \ldots, L\}$ *such that*

$$a_1 = c_1, \quad a_2 = c_2, \quad \cdots \quad , \quad a_{s-1} = c_{s-1}, \quad \textit{and} \quad a_s < c_s, \tag{8}$$

$a_{s+1}, \ldots, a_{s+\ell} = 0$ for some $\ell \geq 0$, and $\{b_i\}_{i=1}^{n-s-\ell}$ with $b_i = a_{s+\ell+i}$ is either legal or empty.

Definition 2.2 *Let* $\{G_n\}$ *be a positive linear recurrence sequence. Given non-negative integers* $a_1, \ldots, a_n$*, the sum* $\sum_{i=1}^{n} a_i G_{n+1-i}$ *is a* ***super-legal*** *Zeckendorf decomposition if there exists an* $s \in \{1, \ldots, L\}$ *such that*

$$a_1 = c_1, \quad a_2 = c_2, \quad \cdots \quad , \quad a_{s-1} = c_{s-1}, \quad \textit{and} \quad a_s < c_s, \tag{9}$$

$a_{s+1}, \ldots, a_{s+\ell} = 0$ for some $\ell \geq 0$, and $\{b_i\}_{i=1}^{n-s-\ell}$ with $b_i = a_{s+\ell+i}$ is either super-legal or empty.

In other words, a decomposition is super-legal if it satisfies condition (2) of Definition 2.1.

Definition 2.3 *Let $\{G_n\}$ be a positive linear recurrence sequence, and assume that the sum $\sum_{i=1}^{n} a_i G_{n+1-i}$ is a legal Zeckendorf decomposition. We call each string described by one of the conditions of Definition 2.1 (not counting the additional 0's) a **block**, and call the number of terms in each block its **length**.*

We note that every super-legal Zeckendorf decomposition is legal and that a concatenation of super-legal Zeckendorf decompositions makes a super-legal Zeckendorf decomposition.

Example 2.4 *The recurrence $G_{n+1} = G_n + 2G_{n-1} + 3G_{n-2}$ with $G_1 = 1,\ G_2 = 2,\ G_3 = 5$ produces the sequence $1, 2, 5, 12, 28, 67, 159, 377, \ldots$. The decomposition of 858 is*

$$858 \ = \ 377 + 2(159) + 2(67) + 28 + 1 \ = \ G_8 + 2G_7 + 2G_6 + G_5 + G_1. \tag{10}$$

This example gives coefficients $(1,2,2,1,0,0,0,1)$, so the blocks of 1274 are $(1,2,2),(1,0)$, and (1), with lengths 3, 2, and 1 respectively. Note that even though the definition of a block excludes the additional 0's (i.e., the $a_{s+1} = a_{s+2} = \cdots = a_{s+\ell} = 0$ from the Definition 2.1), it is still permissible for a block to end with a 0. The decomposition for 1274 is legal but not super-legal, since the final block (1) satisfies condition (1) but not condition (2) from Definition 2.1.

Example 2.5 *An example of a super-legal decomposition using the recurrence from Example 2.4 is*

$$860 \ = \ 377 + 2(159) + 2(67) + 28 + 2 + 1 \ = \ G_8 + 2G_7 + 2G_6 + G_5 + G_2 + G_1, \tag{11}$$

which gives coefficients $(1,2,2,1,0,0,1,1)$. In this case, the final block is $(1,1)$, which satisfies the condition of Definition 2.2.

Given two legal decompositions, we do not necessarily obtain a new legal sequence by concatenating the coefficients. However, if we require that the leading block be super-legal, we do obtain a new legal decomposition by concatenation. With the help of a few lemmas which help us count the number of super-legal decompositions, we can circumvent this obstruction.

Lemma 2.6 *Let $\{G_n\}$ be a positive linear recurrence sequence with relation given by (5), and let H_n be the number of super-legal decompositions using only $G_1, G_2, \ldots, G_n$. We have*

$$H_{n+1} \;=\; c_1 H_n + c_2 H_{n-1} + \cdots + c_L H_{n+1-L}. \tag{12}$$

Proof. Note that $H_{n+1} - H_n$ is the number of super-legal decompositions with largest element G_{n+1}. We count how many such decompositions there are by summing over the possible lengths of the leading block. Say the leading block is of length j with $1 < j \leq L$. Then the leading block is $(c_1, c_2, \ldots, c_{j-1}, a_j)$, where a_j is chosen from $\{0, 1, \ldots, c_j - 1\}$. Therefore there are c_j ways of choosing this leading block. Because we require G_{n+1} to be included in the decomposition, if $j = 1$ there are $c_1 - 1$ ways of choosing this leading block, since the leading coefficient must be nonzero. For any choice of leading block of length j, there are H_{n+1-j} ways of choosing the remaining coefficients. Therefore, we find that

$$H_{n+1} - H_n \;=\; \sum_{j=1}^{L} c_j H_{n+1-j} - H_n, \tag{13}$$

completing the proof.

Lemma 2.7 *Let $\{G_n\}$ be a positive linear recurrence sequence, and let H_n be the number of super-legal decompositions using only $G_1, G_2, \ldots, G_n$. Then $\lim_{n\to\infty} H_n/G_n$ exists and is positive.*

Proof. Since H_n is generated by the same recursion as G_n, it has the same characteristic polynomial, which then has the same roots. Therefore for some $B \geq 0$ we have

$$H_n \;=\; B\lambda_1^n + O(n^{L-2}\lambda_2^n). \tag{14}$$

Thus $\lim_{n\to\infty} H_n/G_n = B/A$ and it suffices to show that $B > 0$. Note that we always have $H_j > 0$, so we must have

$$\alpha \;:=\; \min_{1\leq j\leq L} \frac{H_j}{G_j} \;>\; 0. \tag{15}$$

It follows by induction on n that $H_n \geq \alpha G_n$ for all n. Thus we conclude that $B > 0$, as desired.

3 Density Theorem

To prove the main result as stated in Theorem 1.1, we compute expected values and variances of the relevant random variables. An essential part of the ensuing analysis is the following estimate on the probability that $a_j = k$ for a fixed k, and showing that it has little dependence on j. We prove the theorem via casework based on the structure of the blocks in the decomposition of m. Namely, in the case that a_j is in the rth position of a block of length ℓ, the two subcases are $r = \ell$ (that is, a_j is the last element in the block) or $r < \ell$ (that is, a_j is not the last element in the block).

This is why the notion of a super-legal decomposition is useful; if we want to know whether the legal decomposition $(a_1, a_2, \ldots, a_n)$ has a block that terminates at a_r, this is equivalent to whether $(a_1, a_2, \ldots, a_r)$ forms a super-legal decomposition. So, we first prove some useful lemmas and then collect our results to prove Theorem 1.1.

Lemma 3.1 *Let $\{G_n\}$ be a positive linear recurrence sequence, and choose an integer m uniformly at random from $[0, G_{n+1})$, with legal decomposition*

$$m \;=\; \sum_{j=1}^{n} a_j G_{n+1-j}. \tag{16}$$

Note that this defines random variables $A_1, \ldots, A_n$ taking on values $a_1, \ldots, a_n$.

Let $p_{j,k}(n) := \mathrm{Prob}\,(A_j = k)$. Then, for $\log n < j < n - \log n$, we have

$$p_{j,k}(n) \;=\; p_k(n)(1 + o(1)), \tag{17}$$

where $p_k(n)$ is computable and does not depend on j.

Proof. We divide the argument into cases based on the length of the block containing a_j, as well as the position a_j takes in this block. Suppose that a_j is in the rth place in a block of length ℓ. In order to have $a_j = k$, we must either have $r < \ell$ and $k = c_r$, or $r = \ell$ and $k < c_r$.

In the former case, there are c_ℓ ways to choose the terms in the block containing a_j, due to the c_ℓ choices there are for the final term, and everything else is fixed. There are H_{j-r} ways to choose the coefficients for the terms greater than those in the block containing a_j, and $G_{n-j-\ell+r+1}$ ways to choose the smaller terms.

We now consider the latter case, where $r = \ell$ and $k < c_r$. There is now only one possibility for the coefficients in the block containing a_j, but the rest of the argument remains the same as in the previous case. Therefore, by Lemma 2.7 we find that

$$\begin{aligned} N_{j,k,\ell,r}(n) \;&:=\; \#\{m \in \mathbb{Z} \cap [0, G_{n+1}) : a_j = k,\ a_j \text{ in } r^{th}\text{position in block of length } \ell\} \\ &=\; \begin{cases} c_\ell G_{n-j-\ell+r+1} H_{j-r} & \text{if } r < \ell,\ k = c_r, \\ G_{n-j-\ell+r+1} H_{j-r} & \text{if } r = \ell,\ k < c_r, \\ 0 & \text{otherwise} \end{cases} \\ &=\; N_{k,\ell,r}(n)(1 + o(1)), \end{aligned} \tag{18}$$

where

$$N_{k,\ell,r}(n) \;:=\; \begin{cases} c_\ell AB\lambda_1^{n-\ell+1} & \text{if } r < \ell,\ k = c_r, \\ AB\lambda_1^{n-\ell+1} & \text{if } r = \ell,\ k < c_r, \\ 0 & \text{otherwise,} \end{cases} \tag{19}$$

and $N_{k,\ell,r}(n)$ does not depend on j; these formulas follow from applications of the Generalized Binet Formula to the sequences for the G_n's and H_n's. We conclude the proof by noting that

$$p_{j,k}(n) \;=\; \frac{1}{G_{n+1}}\sum_{\ell=1}^{L}\sum_{r=1}^{\ell} N_{j,k,\ell,r}(n) \;=\; \left(\frac{1}{G_{n+1}}\sum_{\ell=1}^{L}\sum_{r=1}^{\ell} N_{k,\ell,r}(n)\right)\cdot(1+o(1)), \quad (20)$$

where we used (18) to replace $N_{j,k,\ell,r}(n)$. The claim now follows by defining

$$p_k(n) \;:=\; \frac{1}{G_{n+1}}\sum_{\ell=1}^{L}\sum_{r=1}^{\ell} N_{k,\ell,r}(n) \quad (21)$$

and noting that its size is independent of j. More is true, as the Generalized Binet Formula again gives us that G_{n+1} is a constant times λ_1^{n+1} (up to lower order terms), and similarly the sum for $p_k(n)$ is a multiple of λ_1^{n+1} plus lower order terms.

We also use the following result, which follows immediately from Theorems 1.2 and 1.3 in [MW1] (see also [MW2] for a survey on the subject).

Lemma 3.2 *Let $\{G_n\}$ be a positive linear recurrence sequence, with $s(m)$ the number of summands in the decomposition of m. That is, for $m = \sum_{j=1}^{n} a_j G_{n+1-j}$, let $s(m) := \sum_{j=1}^{n} a_j$. Let $X_n(m)$ be the random variable defined by $X_n(m) = s(m)$, where m is chosen uniformly at random from $[0, G_{n+1})$. Then*

$$\mathbb{E}[X_n(m)] \;=\; nC + o(n) \quad \textit{and} \quad \mathrm{Var}[X_n(m)] \;=\; o(n^2). \quad (22)$$

We define another random variable similarly.

Lemma 3.3 *Let $\{G_n\}$ be a positive linear recurrence sequence, and let $S \subseteq \{G_n\}$ be a set with positive density $q(S)$ in $\{G_n\}$. For m chosen uniformly at random in $[0, G_{n+1})$, let*

$$Y_n(m) \;:=\; \sum_{j\in T_n} a_j, \quad (23)$$

where $T_n = \{j \le n | G_{n+1-j} \in S\}$.

Then, for some constant $C > 0$, we have

$$\mathbb{E}[Y_n(m)] \;=\; dnC + o(n), \quad \mathrm{Var}[Y_n(m)] \;=\; o(n^2). \quad (24)$$

Proof. We first compute the expected value. We have

$$\begin{aligned}
\mathbb{E}[Y_n(m)] &= \mathbb{E}\left[\sum_{j\in T_n} a_j\right] = \sum_{j\in T_n}\mathbb{E}[a_j] = \sum_{j\in T_n}\sum_{k=1}^{\infty} kp_{j,k}(n) \\
&= O(\log n) + \sum_{j\in T_n}\sum_{k=1}^{\infty} kp_k(n)(1+o(1)). \\
&= O(\log n) + dn(1+o(1))\sum_{k=1}^{\infty} kp_k(n) \\
&= O(\log n) + d(1+o(1))\sum_{j=1}^{n}\sum_{k=1}^{\infty} kp_k(n) \\
&= O(\log n) + d(1+o(1))\sum_{j=1}^{n}\sum_{k=1}^{\infty} kp_{j,k}(n) \\
&= O(\log n) + d(1+o(1))\sum_{j=1}^{n}\mathbb{E}[a_j] \\
&= O(\log n) + \mathbb{E}[X_n(m)]d(1+o(1)) \\
&= dnC + o(n). \qquad (25)
\end{aligned}$$

Note that the above sums are actually finite, since $p_{j,k} = p_k = 0$ for sufficiently large k. The $\log n$ term appears since Lemma 3.1 only allows us to say $p_{j,k} = p_k(1+o(1))$ when $\log n < j < n - \log n$.

We now must consider the variance. First note that if $i + \log n < j$, then letting

$$q_{i,r}(n) := \text{Prob}\left(\text{the block containing } a_i \text{ ends at } a_{i+r} | a_i = k\right), \qquad (26)$$

we have

$$\begin{aligned}
\text{Prob}\left(a_j = \ell | a_i = k\right) &= \sum_{r=1}^{L-1} q_{i,r}(n) p_{j-i-r,\ell}(n) \\
&= (1+o(1))p_\ell(n)\sum_{r=1}^{L-1} q_{i,r}(n) \\
&= p_\ell(n)(1+o(1)). \qquad (27)
\end{aligned}$$

Thus, we compute

$$\begin{aligned}
\mathbb{E}[Y_n(m)^2] &= \mathbb{E}\left[\sum_{i,j\in T_n} a_i a_j\right] = \sum_{i,j\in T_n} \mathbb{E}[a_i a_j] \\
&= \sum_{i,j\in T_n}\sum_{k,\ell=1}^{\infty} k\ell p_{i,k}(n)\text{Prob}\,(a_j=\ell|a_i=k) \\
&= O(n\log n) + 2 \sum_{\substack{i,j\in T_n \\ 2\log n<i+\log n<j<n-\log n}} \sum_{k,\ell=1}^{\infty} k\ell p_{i,k}(n)\text{Prob}\,(a_j=\ell|a_i=k) \\
&\le O(n\log n) + 2 \sum_{\substack{i,j\in T_n \\ 2\log n<i+\log n<j<n-\log n}} \sum_{k,\ell=1}^{\infty} k\ell p_k(n)p_\ell(n)(1+o(1)) \\
&= O(n\log n) + (1+o(1))d^2n^2 \sum_{k,\ell=1}^{\infty} k\ell p_k(n)p_\ell(n) \\
&= O(n\log n) + (1+o(1))d^2n^2 \left(\sum_{k=1}^{\infty} kp_k(n)\right)^2 \\
&= O(n\log n) + d^2n^2C^2(1+o(1)) = d^2n^2C^2 + o(n^2). \qquad (28)
\end{aligned}$$

Therefore

$$\text{Var}[Y_n(m)] = \mathbb{E}[Y_n(m)^2] - \mathbb{E}[Y_n(m)]^2 = o(n^2), \qquad (29)$$

completing the proof.

We are now ready to prove our main result. The idea of the proof is that the results above strongly concentrate $Y_n(m)$ and $X_n(m)$.

Proof (Proof of Theorem 1.1). Note that the proportion of the summands in m's Zeckendorf decomposition which belong to S is $\frac{Y_n(m)}{X_n(m)}$, where $X_n(m), Y_n(m)$ are defined as in the previous lemmas. Therefore it suffices to show that for any $\varepsilon > 0$, with probability $1+o(1)$ we have

$$\left|\frac{Y_n(m)}{X_n(m)} - d\right| < \varepsilon. \qquad (30)$$

By Chebyshev's inequality, letting $g(n) = n^{1/2}\text{Var}[X_n(m)]^{-1/4}$, we obtain

$$\text{Prob}\left(|X_n(m) - \mathbb{E}[X_n(m)]| > \frac{\mathbb{E}[X_n(m)]}{g(n)}\right) \le \frac{\text{Var}[X_n(m)]g(n)^2}{\mathbb{E}[X_n(m)]^2} = o(1). \qquad (31)$$

Letting

$$e_1(n) := \frac{1}{nC}\left(\frac{\mathbb{E}[X_n(m)]}{g(n)} + |\mathbb{E}[X_n(m)] - nC|\right), \qquad (32)$$

we have with probability $1+o(1)$ that

$$nC(1-e_1(n)) \;\le\; X_n(m) \;\le\; Cn(1+e_1(n)). \tag{33}$$

Note that $e_1(n) = o(1)$. A similar argument for $Y_n(m)$ shows that there exists some $e_2(n) = o(1)$ such that with probability $1+o(1)$ we have

$$dnC(1-e_2(n)) \;\le\; Y_n(m) \;\le\; dnC(1+e_2(n)). \tag{34}$$

Therefore, we have that

$$\frac{Y_n(m)}{X_n(m)} \;\le\; \frac{dnC(1+e_2(n))}{nC(1-e_1(n))} \;<\; d+\varepsilon, \tag{35}$$

with probability $1+o(1)$, and we can similarly obtain

$$\frac{Y_n(m)}{X_n(m)} \;>\; d-\varepsilon. \tag{36}$$

Thus we conclude that with probability $1+o(1)$

$$\left|\frac{Y_n(m)}{X_n(m)} - d\right| \;<\; \varepsilon, \tag{37}$$

completing the proof.

The Benfordness of the summands follows from Theorem 1.1. Let S be the set of numbers in $\{G_n\}_{n=1}^{\infty}$ that start with a given digit. Since G_n is a positive linear recurrence sequence, the density of S in $\{G_n\}_{n=1}^{\infty}$ will follow Benford's law in base B, provided that $\log_B \lambda$ is irrational, where λ is the characteristic polynomial of $\{G_n\}^{\infty}$. If we have a Zeckendorf decomposition with summands from $\{G_n\}_{n=1}^{\infty}$, the proportion of those summands which are in S will also follow Benford's law. We can state this more precisely as follows.

Corollary 3.4 *Fix a positive linear recurrence sequence $\{G_n\}$. Let $S_d \subseteq \{G_n\}_{n=1}^{\infty}$ be a set of numbers with a given first digit d. Then S has Benford density (base B) $q(S_d) = log_B(1+\frac{1}{d})$. Fix an $\varepsilon > 0$. As $n \to \infty$, for an integer m selected uniformly at random from $[0, G_{n+1})$ the proportion of the summands in m's Zeckendorf decomposition which belong to S_d is within ε of $q(S_d)$ with probability $1+o(1)$.*

4 Conclusion and Future Work

We were able to handle the behavior of Zeckendorf decompositions in fairly general settings by cleverly separating any decomposition into manageable blocks. The key step was the notion of a super-legal decomposition, which simplified the combinatorial analysis of the generalized Zeckendorf decompositions significantly. This allowed us to prove not just Benford behavior for the leading digits, but also similar results for other sequences with positive density.

We obtained results for a large class of linear recurrences by considering only the main term of Binet's formula for each linear recurrence. In future work we plan on revisiting these problems for other sequences. Obvious candidates include far-difference representations [Al, DDKMV], f-decompositions [DDKMMV], and recurrences with leading term zero (some of which do not have unique decompositions) [CFHMN1, CFHMN2].

References

Al. H. Alpert, *Differences of multiple Fibonacci numbers*, Integers: Electronic Journal of Combinatorial Number Theory **9** (2009), 745–749.

BBGILMT. O. Beckwith, A. Bower, L. Gaudet, R. Insoft, S. Li, S. J. Miller and P. Tosteson, *The Average Gap Distribution for Generalized Zeckendorf Decompositions*, The Fibonacci Quarterly **51** (2013), 13–27.

Ben. F. Benford, *The Law of Anomalous Numbers*, Proceedings of the American Philosophical Society **78** (1938), 551–572. http://www.jstor.org/stable/984802.

BHi. A. Berger and T. Hill, *An Introduction to Benford's Law*, Princeton University Press, 2015.

BDEMMTTW. A. Best, P. Dynes, X. Edelsbrunner, B. McDonald, S. J. Miller, K. Tor, C. Turnage-Butterbaugh, M. Weinstein, *Benford Behavior of Zeckendorf Decompositions*, Fibonacci Quarterly **52** (2014), no. 5, 35–46.

BD. 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.

CFHMN1. M. Catral, P. Ford, P. Harris, S. J. Miller and D. Nelson, *Generalizing Zeckendorf's Theorem: The Kentucky Sequence*, Fibonacci Quarterly **52** (2014), no. 5, 68–90.

CFHMN2. M. Catral, P. Ford, P. Harris, S. J. Miller and D. Nelson, *The Generacci Recurrences and Zeckendorf's Theorem*, preprint.

Day. D. E. Daykin, *Representation of Natural Numbers as Sums of Generalized Fibonacci Numbers*, J. London Mathematical Society **35** (1960), 143–160.

DDKMMV. P. Demontigny, T. Do, A. Kulkarni, S. J. Miller, D. Moon and U. Varma, *Generalizing Zeckendorf's Theorem to f-decompositions*, Journal of Number Theory **141** (2014), 136–158.

DDKMV. P. Demontigny, T. Do, A. Kulkarni, S. J. Miller and U. Varma, *A Generalization of Fibonacci Far-Difference Representations and Gaussian Behavior*, Fibonacci Quarterly **52** (2014), no. 3, 247–273.

DG. M. Drmota and J. Gajdosik, *The distribution of the sum-of-digits function*, J. Théor. Nombrés Bordeaux **10** (1998), no. 1, 17–32.

DT. J. M. Dumont and A Thomas, *Gaussian asymptotic properties of the sum-of-digits function*, J. Number Theory **62** (1997), no. 1, 19–38.

FGNPT. P. Filipponi, P. J. Grabner, I. Nemes, A. Pethö, and R. F. Tichy, *Corrigendum to: "Generalized Zeckendorf expansions"*, Appl. Math. Lett., **7** (1994), no. 6, 25–26.

GT. P. J. Grabner and R. F. Tichy, *Contributions to digit expansions with respect to linear recurrences*, J. Number Theory **36** (1990), no. 2, 160–169.

GTNP. P. J. Grabner, R. F. Tichy, I. Nemes, and A. Pethö, *Generalized Zeckendorf expansions*, Appl. Math. Lett. **7** (1994), no. 2, 25–28.

Hi1. T. Hill, *The first-digit phenomenon*, American Scientist **86** (1996), 358–363.

Hi2. T. Hill, *A statistical derivation of the significant-digit law*, Statistical Science **10** (1996), 354–363.

Ke. T. J. Keller, *Generalizations of Zeckendorf's theorem*, Fibonacci Quarterly **10** (1972), no. 1 (special issue on representations), pages 95–102.

KKMW. M. Kologlu, G. Kopp, S. J. Miller and Y. Wang, *On the Number of Summands in Zeckendorf Decompositions*, Fibonacci Quarterly **49** (2011), no. 2, 116–130.

Len. T. Lengyel, *A Counting Based Proof of the Generalized Zeckendorf's Theorem*, Fibonacci Quarterly **44** (2006), no. 4, 324–325.

M. S. J. Miller (editor), *Benford's Law: Theory and Applications*, Princeton University Press, 2015.

MT-B. S. J. Miller and R. Takloo-Bighash, *An Invitation to Modern Number Theory*, Princeton University Press, Princeton, NJ, 2006.

MW1. S. J. Miller and Y. Wang, *From Fibonacci numbers to Central Limit Type Theorems*, Journal of Combinatorial Theory, Series A **119** (2012), no. 7, 1398–1413. Expanded arXiv version: http://arxiv.org/pdf/1008.3202.pdf.

MW2. S. J. Miller and Y. Wang, *Gaussian Behavior in Generalized Zeckendorf Decompositions*, Combinatorial and Additive Number Theory, CANT 2011 and 2012 (Melvyn B. Nathanson, editor), Springer Proceedings in Mathematics & Statistics (2014), 159–173. http://arxiv.org/pdf/1107.2718v1.

New. S. Newcomb, *Note on the Frequency of Use of the Different Digits in Natural Numbers*, Amer. J. Math. **4** (1881), 39–40.

Rai. R. A. Raimi, *The first digit problem*, Amer. Math. Monthly **83** (1976), no. 7, 521–538.

Ste1. W. Steiner, *Parry expansions of polynomial sequences*, Integers **2** (2002), Paper A14.

Ste2. W. Steiner, *The Joint Distribution of Greedy and Lazy Fibonacci Expansions*, Fibonacci Quarterly **43** (2005), 60–69.

Was. L. Washington, *Benford's law for Fibonacci and Lucas numbers*, Fibonacci Quarterly **19** (1981), no. 2, 175–177. http://www.fq.math.ca/Scanned/19-2/washington.pdf.

Ze. E. Zeckendorf, *Représentation des nombres naturels par une somme des nombres de Fibonacci ou de nombres de Lucas*, Bulletin de la Société Royale des Sciences de Liége **41** (1972), pages 179–182.