GAUSSIAN DISTRIBUTION OF THE NUMBER OF SUMMANDS IN GENERALIZED ZECKENDORF DECOMPOSITIONS IN SMALL INTERVALS

Andrew Best

Department of Mathematics, Ohio State University, Columbus, Ohio
`best.221@osu.edu`

Patrick Dynes

Dept. of Mathematical Sciences, Clemson University, Clemson, South Carolina
`pdynes@clemson.edu`

Xixi Edelsbrunner

Dept. of Mathematics and Statistics, Williams College, Williamstown, Mass.
`xe1@williams.edu`

Brian McDonald

Department of Mathematics, University of Rochester, Rochester, New York
`sjm1@williams.edu`

Steven J. Miller

Dept. of Mathematics and Statistics, Williams College, Williamstown, Mass.
`bmcdon11@u.rochester.edu`

Kimsy Tor

Faculté de mathématiques, Pierre and Marie Curie University, Paris, France
`kimsy.tor@gmail.com`

Caroline Turnage-Butterbaugh

Mathematics Department, Duke University, Durham, North Carolina
`cturnagebutterbaugh@gmail.com`

Madeleine Weinstein

Department of Mathematics, Harvey Mudd College, Claremont, California
`mweinstein@g.hmc.edu`

Abstract

Zeckendorf's theorem states that every positive integer can be written uniquely as a sum of non-consecutive Fibonacci numbers F_n, with initial terms $F_1 = 1, F_2 = 2$. Previous work proved that as $n \to \infty$ the distribution of the number of summands in the Zeckendorf decompositions of $m \in [F_n, F_{n+1})$, appropriately normalized, converges to the standard normal. The proofs crucially used the fact that all integers

1

in $[F_n, F_{n+1})$ share the same potential summands, and hold for more general positive linear recurrence sequences $\{G_n\}$.

We generalize these results to subintervals of $[G_n, G_{n+1})$ as $n \to \infty$ for certain sequences. The analysis is significantly more involved here as different integers have different sets of potential summands. Explicitly, fix an integer sequence $\alpha(n) \to \infty$. As $n \to \infty$, for almost all $m \in [G_n, G_{n+1})$ the distribution of the number of summands in the generalized Zeckendorf decompositions of integers in the subintervals $[m, m + G_{\alpha(n)})$, appropriately normalized, converges to the standard normal. The proof follows by showing that, with probability tending to 1, m has at least one appropriately located large gap between indices in its decomposition. We then use a correspondence between this interval and $[0, G_{\alpha(n)})$ to obtain the result, since the summands are known to have Gaussian behavior in the latter interval.

1. Introduction

1.1. Background

Let $\{F_n\}$ denote the Fibonacci numbers, normalized so that $F_1 = 1$, $F_2 = 2$, and $F_{n+1} = F_n + F_{n-1}$. One of the more interesting, equivalent definitions of the Fibonacci numbers is that they are the unique sequence of positive integers such that every positive number has a unique legal decomposition as a sum of non-adjacent terms.[1] This equivalence is known as Zeckendorf's theorem [29]. Once we know a decomposition exists, a natural question to ask is how the number of summands varies. The first result along these lines is due to Lekkerkerker [20], who proved that the average number of summands needed in the Zeckendorf decomposition of an integer $m \in [F_n, F_{n+1})$ is $\frac{n}{\varphi^2 + 1} + O(1)$, where $\varphi = \frac{1 + \sqrt{5}}{2}$, the golden mean, is the largest root of the Fibonacci recurrence. More is true, and many authors have shown that the distribution of summands of $m \in [F_n, F_{n+1})$ converges to a Gaussian. These results have been extended to a variety of other sequences. There are several different methods of proof, from continued fractions to combinatorial perspectives to Markov processes; see [4, 9, 11, 12, 13, 14, 15, 17, 21, 18, 19, 24, 25, 27, 28] for a sampling of results and methods along these lines, [1, 6, 7, 8, 10, 11] for generalizations to other types of representations, and [2, 5] for related questions on the distribution of gaps between summands.

The analysis in much of the previous work was carried out for $m \in [F_n, F_{n+1})$ (or, for more general sequences $\{G_n\}$, for $m \in [G_n, G_{n+1})$). The advantage of such a localization[2] is that each m has the same candidate set of summands and is of roughly the same size. The purpose of this work is to explore some of the above questions on a significantly smaller scale and determine when and how often we

---

[1] The requirement of uniqueness of decomposition forces us to start the sequence this way.

[2] As the sequence $\{F_n\}$ is exponentially growing, it is easy to pass from m in this interval to $m \in [0, F_n)$.

obtain Gaussian behavior. Note that we cannot expect such behavior to hold for all sub-intervals of $[F_n, F_{n+1})$, even if we require the size to grow with n. To see this, consider the interval

$$[F_{2n} + F_n + F_{n-2} + \cdots + F_{\lfloor n^{1/4} \rfloor}, \ F_{2n} + F_{n+1} + F_{\lfloor n^{1/4} \rfloor}). \tag{1}$$

The integers in the above interval that are less than $F_{2n} + F_{n+1}$ have on the order of $n/2$ summands, while those that are larger have at most on the order of $n^{1/4}$ summands. Thus the behavior cannot be Gaussian.[3]

In [3], we proved Gaussian behavior for the number of summands in the Zeckendorf decomposition for almost all small subintervals of $[F_n, F_{n+1})$; in this work we generalize to other sequences. Henceforth $\{G_n\}$ will denote a positive linear recurrence sequence.

Definition 1. A sequence $\{G_n\}_{n=1}^{\infty}$ of positive integers is a *Positive Linear Recurrence Sequence (PLRS)* if the following properties hold.

1. *Recurrence relation:* There are non-negative integers $L, c_1, \ldots, c_L$ such that

$$G_{n+1} \ = \ c_1 G_n + \cdots + c_L G_{n+1-L}, \tag{2}$$

   with L, c_1 and c_L positive.

2. *Initial conditions:* $G_1 = 1$, and for $1 \le n < L$ we have

$$G_{n+1} \ = \ c_1 G_n + c_2 G_{n-1} + \cdots + c_n G_1 + 1. \tag{3}$$

A decomposition $\sum_{i=1}^{m} a_i G_{m+1-i}$ of a positive integer N (and the sequence $\{a_i\}_{i=1}^{m}$) is *legal* if $a_1 > 0$, the other $a_i \ge 0$, and one of the following two conditions holds:

- Condition 1: We have $m < L$ and $a_i = c_i$ for $1 \le i \le m$.

- Condition 2: There exists $s \in \{0, \ldots, L\}$ such that

$$a_1 \ = \ c_1, \ a_2 \ = \ c_2, \ \cdots, \ a_{s-1} \ = \ c_{s-1} \text{ and } a_s < c_s, \tag{4}$$

   $a_{s+1}, \ldots, a_{s+\ell} \ = \ 0$ for some $\ell \ge 0$, and $\{b_i\}_{i=1}^{m-s-\ell}$ (with $b_i = a_{s+\ell+i}$) is legal.

If $\sum_{i=1}^{m} a_i G_{m+1-i}$ is a legal decomposition of N, we define the *number of summands* (of this decomposition of N) to be $a_1 + \cdots + a_m$.

---

[3]Though in this situation it would be interesting to investigate separately the behavior on both sides.

Informally, a legal decomposition is one where we cannot use the recurrence relation to replace a linear combination of summands with another summand, and the coefficient of each summand is appropriately bounded. For example, if $G_{n+1} = 2G_n + 3G_{n-1} + G_{n-2}$, then $G_5 + 2G_4 + 3G_3 + G_1$ is legal, while $G_5 + 2G_4 + 3G_3 + G_2$ is not (we can replace $2G_4 + 3G_3 + G_2$ with G_5), nor is $7G_5 + 2G_2$ (as the coefficient of G_5 is too large). Note the Fibonacci numbers correspond to the special case of $L = 2$ and $c_1 = c_2 = 1$.

Earlier work (see, for example, [24, 25]) proved that if $\{G_n\}$ is a PLRS then we again have unique legal decompositions and Gaussian behavior for the number of summands from $m \in [G_n, G_{n+1})$.

1.2. Main Result

Fix an increasing positive integer valued function $\alpha(n)$ with

$$\lim_{n \to \infty} \alpha(n) = \lim_{n \to \infty} (n - \alpha(n)) = \infty. \tag{5}$$

Our main result, given in the following theorem, extends the Gaussian behavior of the number of summands in Zeckendorf decompositions to smaller intervals. Note that requiring m to be in $[G_n, G_{n+1})$ is not a significant restriction because given any m, there is always an n such that this holds.

Theorem 1 (Gaussianity on small intervals). *Let $\{G_n\}$ be a positive linear recurrence sequence with recurrence*

$$G_{n+1} = c_1 G_n + c_2 G_{n-1} + \cdots + c_L G_{n+1-L}, \tag{6}$$

where we additionally assume $c_1 \geq c_2 \geq \cdots \geq c_L \geq 1$. Choose $\alpha(n)$ satisfying (5). The distribution of the number of summands in the decompositions of integers in the interval $[m, m + G_{\alpha(n)})$ converges to a Gaussian distribution when appropriately normalized for almost all $m \in [G_n, G_{n+1})$.

In §2 we derive some useful properties of Zeckendorf decompositions, which we use in §3 to prove Theorem 1. The reason for our extra condition on the c_i's surfaces in Lemma 1, where this constraint forces a truncated legal decomposition to remain legal, though we conjecture that Theorem 1 holds more generally. Recently the fifth named author and some of his colleagues in [8] developed a new method which bypasses similar technical difficulties in related problems, and a future project is to see if those ideas would allow a lessening of some of the conditions for these problems.

2. Preliminaries

In order to prove Theorem 1, we establish a correspondence between the decompositions of integers in the interval $[m, m+G_{\alpha(n)})$ and those in $[0, G_{\alpha(n)})$. Throughout this paper, by $[a, b]$ we mean the set $\{N \in \mathbb{N} : a \leq N \leq b\}$. We also introduce the following notation. Fix a non-decreasing positive function $q(n)$, taking on even values, such that

$$q(n) \; < \; n - \alpha(n), \quad q(n) \; = \; o\left(\sqrt{\alpha(n)}\right), \quad \lim_{n \to \infty} q(n) \; = \; \infty; \qquad (7)$$

the reason for the second condition is to allow us to appeal to known convergence results for a related system, as $q(n)$ will be significantly less than the standard deviation in that setting.

For $m \in [G_n, G_{n+1})$ with decomposition

$$m \; = \; \sum_{j=1}^{n} a_j G_j, \qquad (8)$$

define

$$
\begin{aligned}
C_1(m) \; &:= \; (a_1, a_2, \ldots, a_{\alpha(n)}), \\
C_2(m) \; &:= \; (a_{\alpha(n)+1}, \ldots, a_{\alpha(n)+q(n)}), \\
C_3(m) \; &:= \; (a_{\alpha(n)+q(n)+1}, \ldots, a_n).
\end{aligned} \qquad (9)
$$

Let $s(m)$ be the number of summands in the decomposition of m. That is, let

$$s(m) \; := \; \sum_{j=1}^{n} a_j. \qquad (10)$$

Similarly, let $s_1(m)$, $s_2(m)$, and $s_3(m)$ be the number of summands contributed by $C_1(m)$, $C_2(m)$, and $C_3(m)$, respectively.

Lemma 1. *Let* $x \in [m, m + G_{\alpha(n)})$. *If there are at least* $3L$ *consecutive 0's in* $C_2(m)$, *then* $C_3(x)$ *is constant, and hence* $s_3(x)$ *is constant as well.*

Proof. Assume there are at least $3L$ consecutive 0's in $C_2(m)$. Then for some $k \in [\alpha(n) + 3L, \alpha(n) + q(n))$, we have $a_{k-3L+1} = a_{k-3L+2} = \cdots = a_k = 0$. Let m' denote the integer obtained by truncating the decomposition of m at $a_{k-3L}G_{k-3L}$. Then $m' < G_{k-3L+1}$. Since $G_{\alpha(n)} \leq G_{k-3L}$, it follows that for any $h < G_{\alpha(n)}$ we have

$$
\begin{aligned}
m' + h \; &< \; G_{k-3L+1} + G_{k-3L} \\
&\leq \; c_1 G_{k-3L+1} + c_2 G_{k-3L} + \cdots + c_L G_{k-2L+2} \; = \; G_{k-3L+2}, \qquad (11)
\end{aligned}
$$

and thus the decomposition of $m'+h$ has largest summand no greater than G_{k-3L+2}. Then since $3L - 2 \geq L$, the Zeckendorf decomposition of $m + h$ is obtained simply by concatenating the decompositions for $m - m'$ and $m' + h$. Hence $C_3(m + h) = C_3(m - m') = C_3(m)$. □

With this lemma, we see that the distribution of the number of summands involved in the decomposition of $x \in [m, m + G_{\alpha(n)})$ depends (up to a shift) only on what happens in $C_1(x)$ and $C_2(x)$, *provided* that there is a gap between summands of length at least $3L$ somewhere in $C_2(m)$. In light of this stipulation, we will show the following items in order to prove our main theorem.

- With high probability, m is of the desired form (i.e., there is a gap between summands of length at least $3L$ in $C_2(m)$).

- When m is of the desired form, the distribution of the number of summands involved in $C_1(x)$ for $x \in [m, m + G_{\alpha(n)})$ converges to Gaussian when appropriately normalized.

- The summands involved in $C_2(x)$ produce a negligible error term (i.e., there are significantly fewer summands from $C_2(x)$ than there are from $C_1(m)$).

We address the first point with the following lemma.

Lemma 2. *As $n \to \infty$, with probability $1 + o(1)$ there are at least $3L$ consecutive 0's in $C_2(m)$ if m is chosen uniformly at random from the integers in $[G_n, G_{n+1})$.*

In [5], Bower, Insoft, Li, Miller and Tosteson analyze the largest gap between summands in legal decompositions, assuming only that the c_i's are non-negative (a less restrictive condition than we have). They prove that for $m \in [G_n, G_{n+1})$ as $n \to \infty$ with probability $1 + o(1)$ there is a gap of length $C \log n$ for some constant $C > 0$. This almost proves what we need; the only difficulty is we need the gap to be in $C_2(m)$. Instead of modifying their technical approach, in Appendix 4 we derive a difference equation which proves that as $n \to \infty$ with probability 1 we have a gap of any fixed finite size; while we prove that such a gap exists somewhere, the argument there is trivially modified to apply to just $C_2(m)$ (as if we have no summands in that region then the result is trivially true!), and completes the proof of the above lemma.

Until this point we have not used the condition that $c_1 \geq c_2 \geq \cdots \geq c_L$. The following lemma illustrates the necessity of this condition.

Lemma 3. *If $c_1 \geq c_2 \geq \cdots \geq c_L$, then*

$$\sum_{j=1}^{n} a_j G_{n+1-j} \text{ is legal implies that } \sum_{j=2}^{n} a_j G_{n+1-j} \text{ is legal.} \qquad (12)$$

Proof. If the first sum is legal, then for some s we have $a_1 = c_1$, $a_2 = c_2$, $\cdots$, $a_s < c_s$. Then we necessarily have $a_2 \leq c_1$, $a_3 \leq c_2$, $\cdots$ $a'_s < c_{s'-1}$ for some $s' \leq s$. We can repeat this process, noting that a block ends at a_s. Therefore, the decomposition is legal. $\qquad\square$

Corollary 1. *If $c_1 \geq c_2 \geq \cdots \geq c_L$, then legal decompositions are chosen by the greedy algorithm.*

Proof. Suppose not, so that $\sum_{j=1}^{n} a_j G_{n+1-j}$ is a legal decomposition with

$$\sum_{j=k+1}^{n} a_j G_{n+1-j} \;\geq\; G_{n+1-k}. \tag{13}$$

By an iterated application of the previous lemma, the above sum is a legal decomposition, and it is known that legal decompositions of integers in $[G_n, G_{n+1})$ all have largest term G_n, which gives a contradiction. $\qquad\square$

Assuming m is of the desired form (meaning there is a gap of length at least $3L$ in its decomposition), we now consider the distribution of $s(x)$ for $x \in [m, m+G_{\alpha(n)})$.

Lemma 4. *If m has at least $3L$ consecutive 0's in $C_2(m)$, then for all $x \in [m, m + G_{\alpha(n)})$, we have*

$$0 \;\leq\; s(x) - s_3(m) - s(t(x)) \;<\; Kq(n), \tag{14}$$

where $K := \max_j c_j$, and $t(x)$ denotes the bijection

$$t : [m, m + G_{\alpha(n)}) \rightarrow [0, G_{\alpha(n)}) \tag{15}$$

given by

$$t(m+h) \;:=\; \begin{cases} m_0 + h, & \text{if } m_0 + h < G_{\alpha(n)} \\ m_0 + h - G_{\alpha(n)} & \text{if } m_0 + h \geq G_{\alpha(n)}, \end{cases} \tag{16}$$

with m_0 the sum of the terms in the decomposition of m truncated at $a_{\alpha(n)-1} G_{\alpha(n)-1}$.

Proof. First, note that the number of summands in the decomposition of x with indices $i \in [\alpha(n), \alpha(n) + q(n))$ must be less than $Kq(n)$. Next, we verify that t from above in fact is a bijection, for which it suffices to show that it is an injection. Note that if $t(m + h_1) = t(m + h_2)$, then since $|h_1 - h_2| < G_{\alpha(n)}$, we must have $m_0 + h_1 = m_0 + h_2$, so we may conclude that t is injective, hence bijective since the domain and codomain are finite. For any $x \in [m, m + G_{\alpha(n)})$, the decompositions of $t(x)$ and x agree for the terms with index less than $\alpha(n)$ by virtue of Corollary 1. Furthermore, the decompositions of x and m agree for terms with index greater than $\alpha(n) + q(n)$. Therefore, the number of summands in the decomposition of x with indices $i \in [\alpha(n), \alpha(n) + q(n))$ is equal to $s(x) - s_3(m) - s(t(x))$. Combining this with our initial observation, the lemma now follows. $\qquad\square$

As a result of this lemma, the distribution of $s(x)$ over the integers in $[m, m+G_\alpha)$ is a shift of its distribution over $[0, G_{\alpha(n)})$, up to an error bounded by $q(n)$. With this fact, we are now ready to prove the main theorem.

3. Proof of Theorem 1

We now prove our main result. We assume below that $\{G_n\}$ is a PLRS with $c_1 \geq c_2 \geq \cdots \geq c_L \geq 1$, that $\alpha(n)$ obeys (5), and that $q(n)$ obeys (7).

Proof of Theorem 1. For a fixed $m \in [G_n, G_{n+1})$ with at least $3L$ consecutive 0's somewhere in $C_2(m)$ (which is true for almost all m as $n \to \infty$), we define random variables X_n and Y_n by

$$X_n := s(H_1), \quad Y_n := s(H_2), \tag{17}$$

where H_1 is chosen uniformly at random from $[m, m + G_{\alpha(n)})$ and $H_2 = H_1 - m$ (and thus is chosen uniformly at random from $[0, G_{\alpha(n)})$). Let

$$X'_n := \frac{1}{\sigma_x(n)}(X_n - \mathbb{E}[X_n]), \tag{18}$$

and

$$Y'_n := \frac{1}{\sigma_y(n)}(Y_n - \mathbb{E}[Y_n]), \tag{19}$$

where $\sigma_x(n)$ and $\sigma_y(n)$ are the standard deviations of X_n and Y_n, respectively, so that X_n and Y_n are normalized with mean 0 and variance 1. It is known that Y'_n converges to a Gaussian distribution with mean and variance of order $\alpha(n)$ (see, for example, [24]), and we claim that X'_n converges to a Gaussian distribution as well.

Note that $X_n = Y_n + C + e(n)$, where C is a constant and $e(n)$ is an error term with $|e(n)| < Kq(n)$. From the assumption that $q(n) = o\left(\sqrt{\alpha(n)}\right)$ we have $q(n)/\sigma_y(n) \to 0$. Therefore, we have

$$
\begin{aligned}
\mathrm{Var}[X_n] &= \mathrm{Var}[Y_n] + \mathrm{Var}[e(n)] + 2\mathrm{Cov}[Y_n, e(n)] \\
&\leq \mathrm{Var}[Y_n] + 4K^2 q(n)^2 + 2\mathbb{E}[(Y_n - \mathbb{E}[Y_n])(e(n) - \mathbb{E}[e(n)])] \\
&\leq \mathrm{Var}[Y_n] + 4K^2 q(n)^2 + 2\mathbb{E}[|Y_n - \mathbb{E}[Y_n]| \cdot |e(n) - \mathbb{E}[e(n)]|] \\
&\leq \mathrm{Var}[Y_n] + 4K^2 q(n)^2 + 4Kq(n)\mathbb{E}[|Y_n - \mathbb{E}[Y_n]|] \\
&\leq \mathrm{Var}[Y_n] + 4K^2 q(n)^2 + 4Kq(n)\mathbb{E}[(Y_n - \mathbb{E}[Y_n])^2]^{1/2} \\
&= \mathrm{Var}[Y_n] + 4K^2 q(n)^2 + 4Kq(n) + 4Kq(n)\sigma_y(n) \\
&= \mathrm{Var}[Y_n](1 + o(1)). \tag{20}
\end{aligned}
$$

Hence $\sigma_x(n) = \sigma_y(n)(1 + o(1))$. Note that $|X_n - Y_n - C| < Kq(n)$, so $|X'_n\sigma_x(n) - Y'_n\sigma_y(n)| = O(q(n))$, and thus

$$|X'_n - Y'_n| \;=\; \left| X'_n \frac{\sigma_x(n)}{\sigma_y(n)} - Y'_n \right|(1 + o(1)) \;=\; O\left(\frac{q(n)}{\sigma_y(n)}\right) \;=\; o(1). \qquad (21)$$

Let A_n and B_n be the cumulative distribution functions for X'_n and Y'_n, respectively. By Lemma 4 combined with the above bound on $|X'_n - Y'_n|$, we have

$$B_n(x - o(1)) \;\leq\; A_n(x) \;\leq\; B_n(1 + o(1)). \qquad (22)$$

Since $\{B_n\}_n$ converges pointwise to the cumulative distribution function for a Gaussian distribution, say $B(x)$, and it is easy to show that if $\beta_n \to 0$, then $B_n(x + \beta_n) \to B(x)$ since B is continuous and B_n is monotone non-decreasing, it follows that $\{A_n\}_n$ also converges pointwise to $B(x)$. This completes the proof. $\square$

4. Conclusion and Future Work

By finding a correspondence between generalized Zeckendorf decompositions in the interval $[m, m + G_{\alpha(n)})$ and in the interval $[0, G_{\alpha(n)})$, we are able to prove convergence to Gaussian behavior on many sub-intervals. The key step is to show that almost surely an integer m chosen uniformly at random from $[G_n, G_{n+1})$ permits the construction of a bijection onto the interval $[0, G_{\alpha(n)})$. Our results then follow from previous work on the Gaussian behavior of the number of generalized Zeckendorf summands in this interval.

In the future, we plan to extend our results to more general sequences. The first natural candidate is to remove the assumption on the c_i's among the positive linear recurrence sequences we study. Other interesting topics include the signed decompositions (or far difference representations) where both positive and negative summands are allowed (see [1, 11]), the f-decompositions of [10], and some other recurrences where the leading coefficient is zero (which in some cases leads to a loss of unique decompositions), such as [6, 7, 8].

References

[1] H. Alpert, Differences of multiple Fibonacci numbers, *Integers* **9** (2009), 745-749.

[2] O. Beckwith, A. Bower, L. Gaudet, R. Insoft, S. Li, S. J. Miller and P. Tosteson, The average gap distribution for generalized Zeckendorf decompositions, *Fibonacci Quart.* **51** (2013), 13-27.

[3] A. Best, P. Dynes, X. Edelsbrunner, B. McDonald, S. J. Miller, K. Tor, C. Turnage-Butterbaugh and M. Weinstein, Gaussian behavior of the number of summands in Zeckendorf decompositions in small intervals, *Fibonacci Quart.* **52** (2014), no. 5, 47–53.

INTEGERS: 16 (2016)

[4] I. Ben-Ari and S. J. Miller, A Probabilistic Approach to Generalized Zeckendorf Decompositions, *preprint*.

[5] A. Bower, R. Insoft, S. Li, S. J. Miller and P. Tosteson, Distribution of gaps between summands in generalized Zeckendorf decompositions (with an appendix on *Extensions to Initial Segments* with Iddo Ben-Ari), *Combin. Theory Ser. A* **135** (2015), 130–160.

[6] M. Catral, P. Ford, P. E. Harris, S. J. Miller and D. Nelson, Generalizing Zeckendorf's theorem: the Kentucky sequence, *Proceedings of the Sixteenth International Conference on Fibonacci Numbers and Their Applications, Fibonacci Quart.* **52** (2014), Number 5, 69–91.

[7] M. Catral, P. Ford, P. E. Harris, S. J. Miller and D. Nelson, Legal decompositions drising from non-positive linear recurrences, *preprint*.

[8] M. Catral, P. Ford, P. E. Harris, S. J. Miller and D. Nelson, New behavior in legal decompositions arising from non-positive linear recurrences, preprint.

[9] D. E. Daykin, Representation of natural numbers as sums of generalized Fibonacci numbers, *J. Lond. Math. Soc.* **35** (1960), 143–160.

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

[11] P. Demontigny, T. Do, A. Kulkarni, S. J. Miller and U. Varma, A generalization of Fibonacci far-difference representations and Gaussian behavior, *Fibonacci Quart.* **52** (2014), no. 3, 247–273.

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

[13] 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.

[14] 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.

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

[16] V. E. Hoggatt, Generalized Zeckendorf theorem, *Fibonacci Quart.* **10** (1972), no. 1 (special issue on representations), 89–93.

[17] T. J. Keller, Generalizations of Zeckendorf's theorem, *Fibonacci Quart.* **10** (1972), no. 1 (special issue on representations), 95–102.

[18] M. Lamberger and J. M. Thuswaldner, Distribution properties of digital expansions arising from linear recurrences, *Math. Slovaca* **53** (2003), no. 1, 1–20.

[19] T. Lengyel, A counting based proof of the generalized Zeckendorf's theorem, *Fibonacci Quart.* **44** (2006), no. 4, 324–325.

[20] C. G. Lekkerkerker, Voorstelling van natuurlyke getallen door een som van getallen van Fibonacci, *Bull. Belg. Math. Soc. Simon Stevin* **29** (1951-1952), 190–195.

[21] M. Koloğlu, G. Kopp, S. J. Miller and Y. Wang, On the number of summands in Zeckendorf decompositions, *Fibonacci Quart.* **49** (2011), no. 2, 116–130.

[22] T. Koshy, *Fibonacci and Lucas Numbers with Applications*, Wiley-Interscience, New York, 2001.

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

[24] S. J. Miller and Y. Wang, From Fibonacci numbers to central limit type theorems, *Combin. Theory Ser. A* **119** (2012), no. 7, 1398–1413.

[25] 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.

[26] M. Nathanson, *Additive Number Theory: The Classical Bases*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1996.

[27] W. Steiner, Parry expansions of polynomial sequences, *Integers* **2** (2002), Article A14.

[28] W. Steiner, The Joint Distribution of Greedy and Lazy Fibonacci Expansions, *Fibonacci Quart.* **43** (2005), 60–69.

[29] E. Zeckendorf, Représentation des nombres naturels par une somme des nombres de Fibonacci ou de nombres de Lucas, *Bull. Soc. Math. France* **41** (1972), 179–182.

Appendix: Elementary Proof of Moderate Gaps

A crucial ingredient in our proof is that almost surely the Zeckendorf decomposition of an $m \in [G_n, G_{n+1})$ has a gap of length Z or more for some fixed $Z > 3L$. This follows immediately from the work of Beckwith, Bower, Gaudet, Insoft, Li, Miller and Tosteson [2] (see also [4]), who showed that almost surely the longest gap is of the order $\log n$. It is possible to elementarily prove this result by deriving a recurrence relation for these probabilities and analyzing its growth rate directly, which we do below both in the hopes that it might be of use in related problems, and also to keep the argument elementary.

Theorem 2. *Let $\{G_n\}$ be a PLRS with recurrence relation with positive integer coefficients:*

$$G_{n+1} \;=\; c_1 G_n + \cdots + c_L G_{n-L+1}, \quad c_i, L \geq 1. \tag{23}$$

Fix an integer $Z > L$ and let H_{n+1} be the number of integers $m \in [0, G_{n+1})$ such that m's legal decomposition does not have a gap between summands of length Z or greater.[4] *Then H_n/G_n converges to zero exponentially fast.*

Proof. We first find a recurrence relation for $\{H_n\}$. Consider n much larger than $L + Z$, and let $m \in [0, G_{n+1})$ be arbitrary. We count how many m have a legal decomposition with no gap between summands of length Z or more by looking at the possible beginning strings of m's decomposition. Specifically, we look at how often $G_n, G_{n-1}, \ldots, G_{n-L+1}$ occur. As the analysis is trivial when $L = 1$ we assume $L \geq 2$ below.

---

[4]Note that we only care about a gap of length at least Z between adjacent summands; thus our decomposition may miss many elements before the first chosen summand. It is also easier to first count in $[0, G_{n+1})$ and later shift to $[G_n, G_{n+1})$ by subtraction.

- *Case n: At most $c_1 - 1$ summands of G_n:* We may have 0, 1, ..., $c_1 - 1$ or c_1 occurrences of G_n. If we have zero G_n's, then the number of possible completions of the decomposition such that all gaps are less than Z is, by definition, H_n. If we have 1, ..., $c_1 - 1$ summands of G_n then the number of completions is $H_n - H_{n-Z}$ (we must subtract H_{n-Z} as we have a summand, and we cannot have Z or more non-chosen summands now). Thus the contribution from this case to H_{n+1} is

$$H_n + (c_1 - 1)(H_n - H_{n-Z}) \;=\; c_1(H_n - H_{n-Z}) + H_{n-Z}. \qquad (24)$$

- *Case $n - 1$: At most $c_2 - 1$ summands of G_{n-1}:* To be in this case, we first have c_1 summands of G_n. If we have no G_{n-1} terms in our decomposition, then we have $H_{n-1} - H_{n-1-(Z-1)} = H_{n-1} - H_{n-Z}$ ways to complete the decomposition (as we have zero summands, we must be careful and avoid not taking any of the next $Z - 1$ summands). If we have 1, ..., $c_2 - 1$ copies of G_{n-1} then there are $H_{n-1} - H_{n-1-Z}$ ways to complete the decomposition as desired. Thus this case contributes to H_{n+1}

$$(H_{n-1} - H_{n-Z}) + (c_2 - 1)(H_{n-1} - H_{n-1-Z})$$
$$= \; c_2(H_{n-1} - H_{n-1-Z}) + (H_{n-1-Z} - H_{n-Z}). \qquad (25)$$

- *Case $n - \ell$: At most $c_\ell - 1$ summands of $G_{n-\ell+1}$:* Similar to the earlier cases, if we have zero copies of $G_{n-\ell+1}$ then there are $H_{n-\ell} - H_{n-\ell-(Z-1)} = H_{n-\ell} - H_{n-\ell+1-Z}$ ways to complete the decomposition, while if there are 1, ..., $c_\ell - 1$ copies of $G_{n-\ell+1}$ there are $H_{n-\ell} - H_{n-\ell-Z}$ possibilities. Thus the total contribution to H_{n+1} is

$$(H_{n-\ell} - H_{n-\ell+1-Z}) + (c_\ell - 1)(H_{n-\ell} - H_{n-\ell-Z})$$
$$= \; c_\ell(H_{n-\ell} - H_{n-\ell-Z}) + (H_{n-\ell-Z} - H_{n-\ell+1-Z}). \qquad (26)$$

  If $\ell < L$ we continue to the next case with now c_ℓ copies of $G_{n-\ell+1}$, while if $\ell = L$ then the process terminates, as the recurrence relation and our definition of legality means we would replace this beginning string with $c_1 G_n + \cdots + c_L G_{n-L+1}$ with G_{n+1}, contradicting both the fact that we have a legal decomposition and that our number is less than G_{n+1}.

Combining the above, we obtain the recurrence for $\{H_n\}$ (notice that we have a telescoping sum)

$$H_{n+1} \;=\; c_1(H_n - H_{n-Z}) + \cdots + c_L(H_{n-L+1} - H_{n-L+1-Z}) + H_{n-L+1-Z}. \quad (27)$$

Setting $\widetilde{H}_{n+1} = H_{n+1} - H_n$, we see $\widetilde{H}_{n+1}$ counts the number of integers in $[G_n, G_{n+1})$ with no gaps of length Z or more, and using (27) with n and $n-1$ gives us (27) with

each H_k replaced with $\widetilde{H}_k$. As $\{\widetilde{H}_n\}$ is a strictly increasing sequence, the largest eigenvalue of its recurrence must be greater than 1. From the Generalized Binet Formula expansion for H_n (see, for example, Theorem A.1 of [2]), there is some constant $\alpha \in (0, 1)$ such that $H_{n-\ell-Z} \geq \alpha H_{n-\ell}$. Letting

$$\widetilde{c}_\ell = \begin{cases} c_\ell(1-\alpha) & \text{if } 1 \leq \ell < L \\ c_L & \text{if } \ell = L, \end{cases} \tag{28}$$

we find

$$\widetilde{H}_{n+1} < \widetilde{c}_1 \widetilde{H}_1 + \cdots + \widetilde{c}_L \widetilde{H}_{n-L+1}. \tag{29}$$

Notice this is almost the same recurrence as that of G_n, the only difference being that at least one of the coefficients $\widetilde{c}_\ell$ is smaller than c_ℓ. To prove $\{\widetilde{H}_n\}$ grows exponentially slower than $\{G_n\}$, we instead study the sequence $\{\widehat{H}_n\}$ which satisfies the recurrence

$$\widehat{H}_{n+1} = \widetilde{c}_1 \widehat{H}_n + \cdots + \widetilde{c}_L \widehat{H}_{n-L+1}, \tag{30}$$

as $\widetilde{H}_n \leq \widehat{H}_n$.

We use many standard properties of the Generalized Binet Formula expansions below; see Theorem A.1 of [2] for statements and proofs. The solution to a linear recurrence of fixed, finite length is of the form

$$\beta_{1,r_1} n^{r_1} \lambda_1^n + \cdots + \beta_{1,0} \lambda_1^n + \cdots + \beta_{k,r_k} n^{r_k} \lambda_k^n + \cdots + \beta_{k,0} \lambda_k^n. \tag{31}$$

As the c_i's are positive, there is a unique positive root $\lambda > 1$ for $\{G_n\}$, and all other roots are less than 1 in absolute value. Thus $G_n = \beta\lambda^n + O(|n^r \lambda_2|^n)$ for some $0 < |\lambda_2| < \lambda$ (if $\beta = 0$ then G_n is exponentially decaying). As all the coefficients of the recurrence for $\{\widehat{H}_n\}$ are positive, there is also a unique root of largest absolute value.

We claim $\omega < \lambda$. Clearly $\omega \leq \lambda$ as this sequence grows slower; if they were equal then ω would be a root of both characteristic polynomials, and we would find

$$\omega^L = \widetilde{c}_1 \omega^{L-1} + \cdots + \widetilde{c}_L < c_1 \omega^{L-1} + \cdots + c_L = \omega^L, \tag{32}$$

a contradiction. Thus $\widehat{H}_n$ grows exponentially slower than G_n, which completes the proof. $\qquad\square$