Chapter 1
Gaussian Behavior in Generalized Zeckendorf Decompositions

Steven J. Miller and Yinghui Wang

Abstract A beautiful theorem of Zeckendorf states that every integer can be written uniquely as a sum of non-consecutive Fibonacci numbers $\{F_n\}_{n=1}^{\infty}$; Lekkerkerker proved that the average number of summands for integers in $[F_n, F_{n+1})$ is $n/(\varphi^2 + 1)$, with φ the golden mean. This has been generalized to certain classes of linear recurrence relations, where using techniques from number theory and ergodic theory the fluctuations about the mean are shown to be Gaussian. We discuss an alternative proof that is more combinatorial, and comment on generalizations to related decompositions and problems which have not been handled by these other methods. For example, every integer can be written uniquely as a sum of the $\pm F_n$'s, such that every two terms of the same (opposite) sign differ in index by at least 4 (3). The distribution of the numbers of positive and negative summands converges to a bivariate normal with computable, negative correlation, namely $-(21 - 2\varphi)/(29 + 2\varphi) \approx -0.551058$. Another consequence of this perspective is proving that the distribution of gaps between summands in the Zeckendorf decomposition converges to a geometric random variable.

Department of Mathematics and Statistics, Williams College, sjm1@williams.edu (Steven.Miller.MC.96@aya.yale.edu) and Department of Mathematics, Columbia University, yw2450@columbia.edu

1.1 Introduction

1.1.1 History

The Fibonacci numbers have fascinated professional mathematicians and amateurs for centuries. They have a wealth of properties and interesting relationships; see for example [Kos]. This article is concerned with how Fibonacci numbers arise in decompositions. A beautiful theorem of Zeckendorf [Ze] states that every positive integer may be written uniquely as a sum of non-adjacent Fibonacci numbers; of course, to ensure that the decomposition is unique we need to use the normalization $F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5$ and in general $F_{n+1} = F_n + F_{n-1}$.

The standard proof is by induction, and is also constructive. Simply take the largest Fibonacci number that is at most our number x; say this is F_n. If $F_n + F_{n-1} \le x$ then we could take F_{n-1} in our decomposition, and then replace F_n and F_{n-1} with F_{n+1} by the recurrence relation, which contradicts the maximality of F_n. Thus $F_n + F_{n-1} > x$ and F_{n-1} cannot be used. By induction, we can write $x - F_n$ as a sum of non-adjacent Fibonacci numbers; further, since $x - F_n < F_{n-1}$ clearly F_{n-1} is not in our decomposition. Adding F_n to the decomposition of $x - F_n$ yields our desired expansion. To prove uniqueness, we again proceed by induction. Arranging the Fibonacci summands in decreasing order, if $x = F_{n_1} + \cdots + F_{n_k} = F_{m_1} + \cdots + F_{m_\ell}$, we must have $F_{n_1} = F_{m_1}$; if not, the two sums cannot be equal.[1] The claim now follows by induction as $x - F_{n_1} < x$.

There are many questions one may ask about the Zeckendorf decomposition. The first is to understand the average number of summands needed. Lekkerkerker [Lek] proved that for $x \in [F_n, F_{n+1})$, as $n \to \infty$ the average number of summands needed is $n/(\varphi^2 + 1)$, with $\varphi = \frac{1+\sqrt{5}}{2}$ the golden mean. These results can be generalized to other related decompositions; see [BCCSW, Day, GT, Ha, Ho, Ke, Len] for some of the history and results along these lines. After determining the mean, the next question is the variance or, more generally, the distribution of the fluctuations about the mean. Using techniques from number theory and ergodic theory, the fluctuations about the mean have been shown [DG, FGNPT, GTNP, LT, Ste1] to converge to a Gaussian.

In this chapter we discuss a new approach to proving these results. Koloğlu, Kopp, Miller and Wang [KKMW] adopt a combinatorial vantage to attack this and related problems. They reprove the Gaussian behavior for the Zeckendorf decomposition by writing down an explicit formula for the number of x where there are exactly k summands in the decomposition, and then using Stirling's formula to analyze the resulting binomial coefficients. This method is extended in Miller-Wang [MW] to other decompositions in terms of elements satisfying special recurrence properties. We describe the main results in §1.1.2 and then sketch the proofs in the

---

[1] If $F_{m_1} > F_{n_1}$, then the largest the n-sum can be is $F_{n_1} + F_{n_1-2} + F_{n_1-4} + \cdots + \delta$, where δ is either 1 or 2. Adding 1 or 2 to this and using the recurrence relation gives F_{n_1+1}, and thus the expressions are unequal.

Fibonacci case in §1.2. The key idea is recasting the problem from number theory to combinatorics, specifically to the number of integer partitions satisfying certain constraints. We describe this idea in detail, and outline the resulting algebra (complete details are available in [KKMW]). Unfortunately, while the combinatorial framework applies to generalizations of Zeckendorf decompositions, the proof in general is significantly harder. This is due to the fact that for the Fibonacci numbers, we have explicit formulas for the probability a number in $[F_n, F_{n+1})$ has exactly $k+1$ summands in its Zeckendorf decomposition; it is just $\binom{n-1-k}{k}/F_{n-1}$. All our results follow from a careful analysis of the behavior as $n \to \infty$, which can be accomplished with Stirling's formula. In the general case, the resulting expressions are not as tractable, and the combinatorial approach must be supplemented. We sketch the main ingredients in the analysis in §1.3, and refer the reader to [MW] for full proofs. We end in §1.4 with some open questions as well as recent successes in applying these techniques to determine the distribution of gaps between summands.

1.1.2 Main Results

Before stating the main results, we first set some notation. The sequences defined below are the generalizations of the Fibonacci numbers that can be handled using the techniques of [MW] (as well as the other methods).

Definition 1.1.1 *A sequence $\{H_n\}_{n=1}^{\infty}$ 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*

$$H_{n+1} = c_1 H_n + \cdots + c_L H_{n+1-L}, \qquad (1.1)$$

   with L, c_1 and c_L positive.
2. Initial conditions: *$H_1 = 1$, and for $1 \le n < L$ we have*

$$H_{n+1} = c_1 H_n + c_2 H_{n-1} + \cdots + c_n H_1 + 1. \qquad (1.2)$$

*A decomposition $\sum_{i=1}^{m} a_i H_{m+1-i}$ 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*

$$a_1 = c_1, \ a_2 = c_2, \ \cdots, \ a_{s-1} = c_{s-1} \text{ and } a_s < c_s, \qquad (1.3)$$

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

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

Informally, in a legal decomposition we can't use the recurrence relation to replace a linear combination of summands with another summand, and the coefficient of each summand is appropriately bounded. Other authors [DG, Ste1] use the phrase G-ary decomposition for a legal decomposition, and sum-of-digits function for the number of summands. For example, if $H_{n+1} = 2H_n + 3H_{n-1} + H_{n-2}$, then $H_5 + 2H_4 + 3H_3 + H_1$ is legal, while $H_5 + 2H_4 + 3H_3 + H_2$ is not (we can replace $2H_4 + 3H_3 + H_2$ with H_5), nor is $7H_5 + 2H_2$ (as the coefficient of H_5 is too large). Note the Fibonacci numbers are just the special case of $L = 2$ and $c_1 = c_2 = 1$.

The following probabilistic language is convenient for stating some of our main results.

Definition 1.1.2 (Associated Probability Space to a PLRS) *Let $\{H_n\}$ be a Positive Linear Recurrence Sequence. For each n, consider the discrete outcome space*

$$\Omega_n = \{H_n, H_n + 1, H_n + 2, \cdots, H_{n+1} - 1\} \tag{1.4}$$

with probability measure

$$\mathbb{P}_n(A) = \sum_{\substack{\omega \in A \\ \omega \in \Omega_n}} \frac{1}{H_{n+1} - H_n}, \quad a \subset \Omega_n; \tag{1.5}$$

in other words, each of the $H_{n+1} - H_n$ numbers is weighted equally. We define the random variable K_n by setting $K_n(\omega)$ equal to the number of summands of $\omega \in \Omega_n$ in its legal decomposition. Implicit in this definition is that each integer has a unique legal decomposition; we prove this in Theorem 1, and thus K_n is well-defined.

We denote the cardinality of Ω_n by

$$\Delta_n = H_{n+1} - H_n, \tag{1.6}$$

and we set $p_{n,k}$ equal to the number of elements in $[H_n, H_{n+1})$ whose generalized Zeckendorf decomposition has exactly k summands; thus

$$p_{n,k} = \Delta_n \cdot \text{Prob}(K_n = k). \tag{1.7}$$

The following special case suggests why the limiting behavior is Gaussian. Consider the PLRS given by $L = 1$, $H_{n+1} = 10H_n$. Thus our PLRS is just the geometric series $1, 10, 100, \ldots$, and a legal decomposition of N is just its decimal expansion. Clearly every positive integer has a unique legal decomposition. Further, the distribution of the number of summands converges to a Gaussian by the Central Limit Theorem, as we essentially have the sum of $n - 1$ independent, identically distributed discrete uniform random variables. To see this, write $N = a_1 10^n + \cdots + a_{n+1} 1$. We are interested in the large n behavior of $a_1 + \cdots + a_{n+1}$ as we vary over x in $[10^n, 10^{n+1})$. For large n the contribution of a_1 is immaterial, and the remaining a_i's can be understood by considering the sum of n independent, identically distributed discrete uniform random variables on $\{0, \ldots, 9\}$ (which have mean 4.5 and standard deviation approximately 2.87). Denoting these random

variables by A_i, by the Central Limit Theorem $A_2 + \cdots + A_{n+1}$ converges to being normally distributed with mean $4.5n$ and standard deviation $\sqrt{33n/4}$.

We first review previous results and methods (which are also proved using our methods in [MW]), and then describe our new perspective and extensions. See [**?**, Ha, Ho, Ke, Len] for more on generalized Zeckendorf decompositions, [GT] for a proof of Theorems 1 and 2, and [DG, FGNPT, GTNP, LT, Ste1] for a proof and some generalizations of Theorem 3.

The first result is that a PLRS leads to a unique, generalized Zeckendorf decomposition.

Theorem 1 (Generalized Zeckendorf's Theorem for PLRS). *Let* $\{H_n\}_{n=1}^{\infty}$ *be a* Positive Linear Recurrence Sequence. *Then*

(a) *There is a unique legal decomposition for each integer* $N \geq 0$.

(b) *There is a bijection between the set* $\mathscr{S}_n$ *of integers in* $[H_n, H_{n+1})$ *and the set* $\mathscr{D}_n$ *of legal decompositions* $\sum_{i=1}^{n} a_i H_{n+1-i}$.

In addition to being of interest in its own right, the Generalized Lekkerkerker Theorem below is needed as input to prove the Gausian behavior of the number of summands.

Theorem 2 (Generalized Lekkerkerker's Theorem for PLRS). *Let* $\{H_n\}_{n=1}^{\infty}$ *be a* Positive Linear Recurrence Sequence, *let* K_n *be the random variable of Definition 1.1.2 and denote its mean by* μ_n. *Then there exist constants* $C > 0$, d *and* $\gamma_1 \in (0,1)$ *depending only on L and the* c_i'*s in the recurrence relation of the* H_n'*s such that*

$$\mu_n \ = \ Cn + d + o(\gamma_1^n). \tag{1.8}$$

The main result in the subject is

Theorem 3 (Gaussian Behavior for PLRS). *Let* $\{H_n\}_{n=1}^{\infty}$ *be a* Positive Linear Recurrence Sequence *and let* K_n *be the random variable of Definition 1.1.2. As* $n \to \infty$, *the distribution of* K_n *converges to a Gaussian.*

Unlike previous proofs, which required all summands to be positive, our method allows us to handle negative summands as well, and thus we can prove Gaussian behavior in many other situations. We state one particularly interesting situation.

Definition 1. We call a sum of the $\pm F_n$'s a *far-difference representation* if every two terms of the same sign differ in index by at least 4, and every two terms of opposite sign differ in index by at least 3.

Albert [Al] proved the analogue of Zeckendorf's Theorem for the far-difference representation. It is convenient to set

$$S_n \ = \ \begin{cases} \sum_{0 < n-4i \leq n} F_{n-4i} \ = \ F_n + F_{n-4} + F_{n-8} + \cdots & \text{if } n > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{1.9}$$

Theorem 4 (Generalized Zeckendorf's Theorem for Far-Difference Representations [Al]). *Every integer has a unique far-difference representation. For each $N \in (S_{n-1} = F_n - S_{n-3} - 1, S_n]$, the first term in its far-difference representation is F_n, and the unique far-difference representation of 0 is the empty representation.*

The far-difference representations have both positive and negative summands, which opens up the fascinating question of how the number of each are related. It turns out they are correlated, though the total number of summands and the difference in the number of positive and negative summands are not. Specifically,

Theorem 5 (Generalized Lekkerkerker's Theorem and Gaussian Behavior for Far-Difference Representations [MW]). *Let $\mathcal{K}_n$ and $\mathcal{L}_n$ be the corresponding random variables denoting the number of positive summands and the number of negative summands in the far-difference representation for integers in $(S_{n-1}, S_n]$. As n tends to infinity, $\mathbb{E}[\mathcal{K}_n] = \frac{1}{10}n + \frac{371 - 113\sqrt{5}}{40} + o(1)$, and is $\frac{\sqrt{5}+1}{4} = \frac{\varphi}{2}$ greater than $\mathbb{E}[\mathcal{L}_n]$; the variance of both is of size $\frac{15+21\sqrt{5}}{1000}n$; the joint density of $\mathcal{K}_n$ and $\mathcal{L}_n$ is a bivariate Gaussian with negative correlation $\frac{10\sqrt{5}-121}{179} = -\frac{21-2\varphi}{29+2\varphi} \approx -0.551$; and $\mathcal{K}_n + \mathcal{L}_n$ and $\mathcal{K}_n - \mathcal{L}_n$ are independent.*

1.2 The Fibonacci Case

As the framework of the Fibonacci case is the basis for the general case, and as the Fibonacci case can be handled elementarily, we describe it in some detail. The key idea is to change our perspective and apply the solution to a standard problem in combinatorics, the stars and bars problem. It is also known as the cookie problem in some circles (see for instance Chapter 1 of [MT-B]), or Waring's problem with first powers (see [Na]).

Lemma 1.2.1 *The number of ways to partition C identical objects into P distinct sets is $\binom{C+P-1}{P-1}$.*

Here it is very important that the objects being partitioned are indistinguishable; all that matters is how many objects are given to a specified set, not which objects are given.

Proof. Imagine instead $C + P - 1$ objects. There are $\binom{C+P-1}{P-1}$ ways to choose $P - 1$ of the C. Each of these choices corresponds to a partition of C objects into P sets where order does not count and the objects are indistinguishable. Specifically, for a given choice all of the remaining C items are divided among the P sets, where everything up to the first of the $P - 1$ chosen objects goes to the first set, the objects between the first and second chosen element goes into the second set, and so on.

We may recast the above as saying the number of solutions to $x_1 + \cdots + x_P = C$, with each $x_i \geq 0$, is $\binom{C+P-1}{P-1}$. One of the advantages of this approach is that it is very easy to add in lower restrictions. For example, to find the number of solutions to $y_1 + \cdots + y_P = C$ with $y_i \geq n_i$ for some choice of non-negative integers (with of course $n_1 + \cdots + n_P \leq C$), simply let $y_i = x_i + n_i$. Now each $x_i \geq 0$ is integer valued, and the number of solutions to the equation in the y_i's is the number of solutions to $x_1 + \cdots + x_P = C - (n_1 + \cdots + n_P)$, or $\binom{C-(n_1+\cdots+n_P)+P-1}{P-1}$.

We use the above perspective to analyze the Zeckendorf decomposition of an $x \in [F_n, F_{n+1})$. The number of such x is $F_{n+1} - F_n$, which by the Fibonacci recurrence relation is just F_{n-1}. We count the number of $x \in [F_n, F_{n+1})$ with exactly $k + 1$ summands in their Zeckendorf decomposition, denoting this quantity by $N_n(k)$; we choose to record the number of summands as $k + 1$ as each x must have at least one summand, namely F_n. By the standard proof of Zeckendorf's theorem, we know each x has at most one such valid decomposition. The existence of a Zeckendorf decomposition follows by showing $\sum_{k=0}^{n} P_n(k) = F_{n-1}$ (since no number has two valid decompositions, it is enough to know that the number of valid decompositions equals the number of integers in the interval $[F_n, F_{n+1})$).

Lemma 1.2.2 *Let $P_n(k) = N_n(k)/F_{n-1}$, which is the probability an $x \in [F_n, F_{n+1})$ has exactly $k + 1$ summands in its Zeckendorf decomposition. Then $P_n(k) = \binom{n-1-k}{k} / F_{n-1}$.*

Proof. If x has exactly $k + 1$ summands, then $x = F_{i_1} + F_{i_2} + \cdots + F_{i_k} + F_{i_{k+1}}$, where $F_{i_{k+1}} = F_n$, $1 \leq i_1 < i_2 < i_3 < \cdots < i_k < i_{k+1}$, and $d_j := i_j - i_{j-1} \geq 2$ for $2 \leq j \leq k+1$ (and $d_1 := i_1 - 0 \geq 1$). We can recast this in terms of the cookie problem above. Clearly $d_1 + d_2 + \cdots + d_{k+1} = n$ (we have a telescoping series and the last summand in the decomposition is F_n). Let $d_1 = x_1 + 1$ and $d_j = x_j + 2$ for $2 \leq j \leq k+1$. Then the number of x that have exactly $k + 1$ summands is the number of tuples $(d_1, \ldots, d_{k+1})$ with $d_1 + d_2 + \cdots + d_{k+1} = n$, or equivalently the number of tuples $(x_1, \ldots, x_{k+1})$ with $x_1 + x_2 + \cdots + x_{k+1} = n - (2k+1)$. By our combinatorial result, Lemma 1.2.1, taking $C = n - (2k+1)$ and $P = k+1$ we see the number of such tuples is just

$$N_n(k) = \binom{C+P-1}{P-1} = \binom{n-(2k+1)+(k+1-1)}{k+1-1} = \binom{n-1-k}{k}; \quad (1.10)$$

the lemma now follows.

Now that we have an explicit formula for $N_n(k)$ and $P_n(k)$, *all* the claims follow for the Fibonacci case. We quickly provide a sketch; see [KKMW] for details.

1. Zeckendorf expansion (Theorem 1): As each $x \in [F_n, F_{n+1})$ has at most one Zeckendorf decomposition, the claim follows by counting the number of valid Zeckendorf expansions and seeing that this equals F_{n-1}. This number is just $\sum_{k=0}^{n} \binom{n-1-k}{k}$. The summands vanish if $k \geq \lfloor (n-1)/2 \rfloor$ (the binomial coefficients are extended so that $\binom{n}{\ell} = 0$ if $\ell > n$ or $\ell < 0$). We claim

$$\sum_{k=0}^{n} \binom{n-1-k}{k} = F_{n-1}. \tag{1.11}$$

We proceed by induction. The base case is clear, and the general case follows from using the standard identity that $\binom{m}{\ell} + \binom{m}{\ell+1} = \binom{m+1}{\ell+1}$. Specifically,

$$
\begin{aligned}
\sum_{k=0}^{n} \binom{n+1-1-k}{k} &= \sum_{k=0}^{n} \left[\binom{n-1-k}{k-1} + \binom{n-1-k}{k} \right] \\
&= \sum_{k=1}^{n} \binom{n-2-(k-1)}{k-1} + \sum_{k=0}^{n} \binom{n-1-k}{k} \\
&= \sum_{k=0}^{n-1} \binom{n-2-k}{k} + \sum_{k=0}^{n} \binom{n-1-k}{k} \\
&= F_{n-2} + F_{n-1}
\end{aligned}
\tag{1.12}
$$

by the inductive assumption; noting $F_{n-2} + F_{n-1} = F_n$ completes the proof.

2. Lekkerkerker's Theorem (Theorem 2): The claim follows by computing the expected value, which is

$$\sum_{k=0}^{n} (k+1) P_n(k) = 1 + \frac{1}{F_{n-1}} \sum_{k=0}^{n} k \binom{n-1-k}{k}. \tag{1.13}$$

Let

$$\mathscr{E}(n) = \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} k \binom{n-1-k}{k}. \tag{1.14}$$

Straightforward algebra shows

$$\mathscr{E}(n) = (n-2) F_{n-3} - \mathscr{E}(n-2). \tag{1.15}$$

To see this, note

$$
\begin{aligned}
\mathscr{E}(n) &= \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} k \binom{n-1-k}{k} = \sum_{k=1}^{\lfloor \frac{n-1}{2} \rfloor} k \frac{(n-1-k)!}{k!(n-1-2k)!} \\
&= \sum_{k=1}^{\lfloor \frac{n-1}{2} \rfloor} (n-2-(k-1)) \frac{(n-3-(k-1)!}{(k-1)!(n-3-2(k-1))!} \\
&= \sum_{\ell=0}^{\lfloor \frac{n-3}{2} \rfloor} (n-2-\ell) \binom{n-3-\ell}{\ell} \\
&= (n-2) \sum_{\ell=0}^{\lfloor \frac{n-3}{2} \rfloor} \binom{n-3-\ell}{\ell} - \sum_{\ell=0}^{\lfloor \frac{n-3}{2} \rfloor} \ell \binom{n-3-\ell}{\ell} \\
&= (n-2)F_{n-3} - \mathscr{E}(n-2), \quad\quad\quad (1.16)
\end{aligned}
$$

which proves the claim. Note that we used (1.11) to replace the sum of binomial coefficients with a Fibonacci number.

We study the telescoping sum

$$
\sum_{\ell=0}^{\lfloor \frac{n-3}{2} \rfloor} (-1)^{\ell} \left(\mathscr{E}(n-2\ell) + \mathscr{E}(n-2(\ell+1)) \right) \quad\quad\quad (1.17)
$$

(which is essentially $\mathscr{E}(n)$). Using (1.15) yields

$$
\sum_{\ell=0}^{\lfloor \frac{n-3}{2} \rfloor} (-1)^{\ell} (n-3-2\ell) F_{n-3-2\ell} + O(F_{n-2}). \quad\quad\quad (1.18)
$$

While we could evaluate the last sum exactly, trivially estimating it suffices to obtain the main term, which will give us Lekkerkerker's Theorem.

Binet's formula[2] states that $F_n = \frac{\varphi}{\sqrt{5}} \cdot \varphi^n - \frac{1-\varphi}{\sqrt{5}} \cdot (1-\varphi)^n$, with $\varphi = \frac{1+\sqrt{5}}{2}$ is the golden mean. We use this to convert the sum into a weighted geometric series (where each factor is multiplied by a simple polynomial):

$$
\begin{aligned}
\mathscr{E}(n) &= \frac{\varphi}{\sqrt{5}} \sum_{\ell=0}^{\lfloor \frac{n-3}{2} \rfloor} (n-3-2\ell)(-1)^{\ell} \varphi^{n-3-2\ell} + O(F_{n-2}) \\
&= \frac{\varphi^{n-2}}{\sqrt{5}} \left[(n-3) \sum_{\ell=0}^{\lfloor \frac{n-3}{2} \rfloor} (-\varphi^{-2})^{\ell} - 2 \sum_{\ell=0}^{\lfloor \frac{n-3}{2} \rfloor} \ell(-\varphi^{-2})^{\ell} \right] + O(F_{n-2}). (1.19)
\end{aligned}
$$

---

[2] It is worth noting that while one can establish Binet's formula by substituting and checking , it can also be derived via generating functions; it is no coincidence that in the general proof generating functions will play a central role.

We use the geometric series formula to evaluate the first term, and differentiating identities[3] on the second. After some algebra we obtain

$$\mathscr{E}(n) \;=\; \frac{n\varphi^{n-2}}{\sqrt{5}(1+\varphi^{-2})} + O(F_{n-2}) \;=\; \frac{n\varphi^{n}}{\sqrt{5}(\varphi^2+1)} + O(F_{n-2}). \qquad (1.20)$$

As $F_{n-1} = \frac{\varphi^n}{\sqrt{5}} + O(1)$ for large n, we finally obtain

$$\mathscr{E}(n) \;=\; \frac{nF_{n-1}}{\varphi^2+1} + O(F_{n-2}). \qquad (1.21)$$

We note a more careful analysis is possible, and such a computation leads to an exact form for the mean.

3. Gaussian Behavior (Theorem 3): As we have the density function $P_n(k) = \binom{n-1-k}{k}/F_{n-1}$, one way to prove the Gaussian behavior is to show that as $n \to \infty$ the function $P_n(k)$ converges to a normal distribution. This may be accomplished via Stirling's formula (see [KKMW] for the computation). Another possible approach would be to generalize the proof of Lekkerkerker's theorem to compute all moments, and then appeal to the Moment Method; however, as the combinatorics become harder as the moment increases, the Stirling approach is more tractable.

1.3 Positive Linear Recurrence Sequences

We discuss the main ideas in proving the Gaussian behavior for the generalized Zeckendorf decompositions for Positive Linear Recurrence Sequences $\{H_n\}$. In particular, we discuss the obstructions that arise in trying to use the argument from the previous section, and describe the techniques that handle them. The computations become quite long and technical; we refer the reader to [MW] for these details, and content ourselves with describing the method below.

The first step is to prove that all integers have a unique, legal representation involving the Positive Linear Recurrence Sequence. The proof is essentially just careful book-keeping.

The next step is to determine the size of a general term H_n in our sequence. The most important use of this is to count the number of integers we have in a given window, which is used to normalize our counts to probabilities. In the Fibonacci case, the number of integers in the interval $[F_n, F_{n+1})$ is just $F_{n+1} - F_n = F_{n-1}$; we then used Binet's formula to approximate well the size of F_{n-1}. This all generalizes immediately as we have a linear recurrence relation.

---

[3] As $\sum_{k=0}^{m} x^k = (1 - x^{m+1})/(1-x)$, applying $x\frac{d}{dx}$ to both sides gives $\sum_{k=0}^{m} k x^k = x(1 - (m+1)x^m + mx^{m+1})/(1-x)^2$.

We first sketch another proof of the results from §1.2, and then comment on how to generalize these arguments. While we do not discuss the proof of Theorem 5 in detail, it too can be handled by this method (see [MW]).

1.3.1 New Approach: Case of Fibonacci Numbers

We change notation slightly for the rest of this chapter in order to match the notation of [MW]. Let $p_{n,k} := \#\{N \in [F_n, F_{n+1})$: the Zeckendorf decomposition of N has exactly k summands$\}$. This double sequence satisfies a nice recurrence relation. For any $N \in [F_{n+1}, F_{n+2})$, we have $N = F_{n+1} + F_t + \cdots$ where $t \leq n - 1$ (we are not allowed to have adjacent Fibonacci numbers in our decomposition). Imagine N has exactly $k + 1$ summands in its decomposition. It must have F_{n+1} and it cannot have F_n, and needs exactly k more non-adjacent summands from F_1 to F_{n-1}. There are $p_{n-1,k}$ ways to have k non-adjacent summands with F_{n-1} included, $p_{n-2,k}$ ways to have k non-adjacent summands without F_{n-1} but with F_{n-2} included, and so on. We thus obtain the following formula for $p_{n+1,k+1}$:

$$p_{n+1,k+1} \;=\; p_{n-1,k} + p_{n-2,k} + \cdots . \tag{1.22}$$

Similarly (replacing $n + 1$ with n) we find

$$p_{n,k+1} \;=\; p_{n-2,k} + p_{n-3,k} + \cdots . \tag{1.23}$$

Subtracting, we find

$$p_{n+1,k+1} \;=\; p_{n,k+1} + p_{n-1,k}. \tag{1.24}$$

Our goal is to extract information about the $p_{n,k}$. A powerful approach is to use generating functions. The generating function in this case is

$$\sum_{n,k>0} p_{n,k} x^k y^n \;=\; \frac{y}{1 - y - xy^2}. \tag{1.25}$$

Using partial fractions, we find

$$\frac{y}{1 - y - xy^2} \;=\; -\frac{y}{y_1(x) - y_2(x)} \left(\frac{1}{y - y_1(x)} - \frac{1}{y - y_2(x)} \right), \tag{1.26}$$

where $y_1(x)$ and $y_2(x)$ are the roots of $1 - y - xy^2 = 0$ and the coefficient of y^n is $g(x) = \sum_{k>0} p_{n,k} x^k$. Note the roots are readily computed via the quadratic formula; this is not true for the general case, and is the source of much of the technicalities.

As in the introduction, let K_n be the corresponding random variable associated with k. Using the Method of Moments, it suffices to prove the moments of K_n converge to the moments of a Gaussian to prove Gaussian behavior. We do so through differentiating identities. With $g(x)$ as above, differentiating once and setting $x = 1$

yields

$$g(1) \;=\; \sum_{k>0} p_{n,k} \;=\; F_{n+1} - F_n, \tag{1.27}$$

which is just the number of elements in our interval $[F_n, F_{n+1})$. Differentiating again gives

$$g'(x) \;=\; \sum_{k>0} k p_{n,k} x^{k-1}. \tag{1.28}$$

If we take $x = 1$ we essentially obtain the mean of K_n; we need to divide by F_{n-1} as the $p_{n,k}$'s are counts and not probabilities. As $g(1) = F_{n-1}$, we find

$$g'(1) \;=\; g(1) E[K_n]. \tag{1.29}$$

We continue, and find

$$\bigl(x g'(x)\bigr)' \;=\; \sum_{k>0} k^2 p_{n,k} x^{k-1}, \tag{1.30}$$

which leads to

$$\bigl(x g'(x)\bigr)' \Big|_{x=1} \;=\; g(1) E[K_n^2], \tag{1.31}$$

and then

$$\Bigl(x \bigl(x g'(x)\bigr)'\Bigr)' \Big|_{x=1} \;=\; g(1) E[K_n^3], \tag{1.32}$$

and so on.

Similar results hold for the centralized random variable $K_n' = K_n - E[K_n]$. Miller and Wang [MW] prove that $E[(K_n')^{2m}]/(\mathrm{SD}(K_n'))^{2m} \to (2m-1)!!$ (with $\mathrm{SD}(K_n')$ the standard deviation of K_n') and $E[(K_n')^{2m-1}]/(\mathrm{SD}(K_n'))^{2m-1} \to 0$, which yields the Gaussian behavior of K_n.

1.3.2 New Approach: General Case

We generalize the arguments from §1.3.1 to the case of Zeckendorf expansions arising from Positive Linear Recurrence Sequences $\{H_n\}$. We now set $p_{n,k} = \#\{N \in [H_n, H_{n+1})$: the generalized Zeckendorf decomposition of N has exactly k summands$\}$. We can find a recurrence relation as before. For the Fibonacci numbers, we found

$$p_{n+1,k+1} \;=\; p_{n,k+1} + p_{n,k}. \tag{1.33}$$

In the general case, we have

$$p_{n+1,k} \;=\; \sum_{m=0}^{L-1} \sum_{j=s_m}^{s_{m+1}-1} p_{n-m,k-j}, \tag{1.34}$$

where $s_0 = 0, s_m = c_1 + c_2 + \cdots + c_m$ (with these c's the same c's as in the definition of our recurrence sequence).

The generating function in the Fibonacci case was relatively straightforward, being $\frac{y}{1-y-xy^2}$. Now our generating function equals

$$\frac{\sum_{n\leq L} p_{n,k}x^k y^n - \sum_{m=0}^{L-1}\sum_{j=s_m}^{s_{m+1}-1} x^j y^{m+1} \sum_{n<L-m} p_{n,k}x^k y^n}{1 - \sum_{m=0}^{L-1}\sum_{j=s_m}^{s_{m+1}-1} x^j y^{m+1}}. \qquad (1.35)$$

We again perform a partial fraction expansion in order to glean information about the coefficients. Instead of

$$-\frac{y}{y_1(x) - y_2(x)}\left(\frac{1}{y - y_1(x)} - \frac{1}{y - y_2(x)}\right), \qquad (1.36)$$

now we have

$$-\frac{1}{\sum_{j=s_{L-1}}^{s_L-1} x^j} \sum_{i=1}^{L} \frac{B(x,y)}{(y - y_i(x))\prod_{j\neq i}(y_j(x) - y_i(x))}, \qquad (1.37)$$

where

$$B(x,y) = \sum_{n\leq L} p_{n,k}x^k y^n - \sum_{m=0}^{L-1}\sum_{j=s_m}^{s_{m+1}-1} x^j y^{m+1} \sum_{n<L-m} p_{n,k}x^k y^n \qquad (1.38)$$

and the $y_i(x)$'s are the roots of $A(y) = 1 - \sum_{m=0}^{L-1}\sum_{j=s_m}^{s_{m+1}-1} x^j y^{m+1} = 0$. One of the major difficulties of the proof is the analysis of the roots $y_i(x)$. Unlike the Fibonacci case, where it was easy to write down simple expressions for these via the quadratic formula, in general the arguments become quite involved. If however the coefficients in the recurrence relation are non-increasing then the proofs are easy; see Appendix C of [MW]. The general cases (see Appendix A of [MW]) is more complicated, involving continuity and the range of the $|y_i(x)|$'s. The main idea is to first show that there exists $x > 0$ such that $A(y)$ has no multiple roots and then prove that there are only finitely many $x > 0$ such that $A(y)$ has multiple roots.

The coefficient of y^n is $g(x) = \sum_{k>0} p_{n,k}x^k$. We use the method of differentiating identities as before to get the moments, and find that K_n converges to a Gaussian as $n \to \infty$.

The proof of Theorem 5 proceeds similarly. We state the two key elements, the recurrence relation and the generating function, without proof, and refer the reader to [MW] for the details. Let $p_{n,k,\ell}$ be the number of far-difference representations of integers in $(S_{n-1}, S_n]$ with k positive summands and ℓ negative summands. The recurrence relation is

$$p_{n,k,\ell} = p_{n-1,k,\ell} + p_{n-4,k-1,\ell} + p_{n-3,\ell,k-1}, \; n \geq 5, \qquad (1.39)$$

and the generating function is

$$\hat{G}(x,y,z) \;=\; \sum_{n>0,k>0,\ell\geq 0} p_{n,k,\ell} x^k y^\ell z^n. \qquad\qquad (1.40)$$

1.4 Conclusion and Future Research

The combinatorial approach has extended previous work, allowing us to prove Gaussian behavior for the number of summands for a large class of expansions in terms of solutions to linear recurrence relations. This is just the first of many questions one can ask. Others, which we hope to return to at a later date, include:

1. Are there similar results for linearly recursive sequences with arbitrary integer coefficients (i.e., negative coefficients are allowed in the defining relation)?

2. Lekkerkerker's theorem, and the Gaussian extension, are for the behavior in intervals $[F_n, F_{n+1})$. Do the limits exist if we consider other intervals, say $[F_n + g_1(F_n), F_n + g_2(F_n))$ for some functions g_1 and g_2? If yes, what must be true about the growth rates of g_1 and g_2?

3. For the generalized recurrence relations, what happens if instead of looking at $\sum_{i=1}^n a_i$ we study $\sum_{i=1}^n \min(1, a_i)$? In other words, we only care about how many distinct H_i's occur in the decomposition.

4. What can we say about the distribution of the largest gap between summands in the Zeckendorf decomposition? Appropriately normalized, how does the distribution of gaps between the summands behave? What is the distribution of the largest gap? How often is there a gap of 2?

Recently there has been significant progress on the last question by Beckwith and Miller [BM]. They prove the following two results.

Theorem 1.4.1 (Base B Gap Distribution) *For base B decompositions, as $n \to \infty$ the probability of a gap of length 0 between summands for numbers in $[B^n, B^{n+1})$ tends to $\frac{(B-1)(B-2)}{B^2}$, and for gaps of length $k \geq 1$ to $\frac{(B-1)(3B-2)}{B^2} B^{-k}$.*

Theorem 1.4.2 (Zeckendorf Gap Distribution) *For Zeckendorf decompositions, for integers in $[F_n, F_{n+1})$ the probability of a gap of length $k \geq 2$ tends to $\frac{\varphi(\varphi-1)}{\varphi^k}$ for $k \geq 2$, with $\varphi = \frac{1+\sqrt{5}}{2}$ the golden mean.*

The proofs use the combinatorial vantage, and should be generalizable to other recurrences.

References

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

[BM]   O. Beckwith and S. J. Miller, *The Distribution of Gaps between Summands in Zeckendorf Decompositions*, preprint.

[BCCSW]   E. Burger, D. C. Clyde, C. H. Colbert, G. H. Shin and Z. Wang, *A Generalization of a Theorem of Lekkerkerker to Ostrowski's Decomposition of Natural Numbers*, preprint, 2011.

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

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

[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.

[FS]   P. Flajolet and R. Sedgewick, Analytic combinatorics, Cambridge University Press, Cambridge, 2009.

[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.

[Ha]   N. Hamlin, *Representing Positive Integers as a Sum of Linear Recurrence Sequences*, Abstracts of Talks, Fourteenth International Conference on Fibonacci Numbers and Their Applications (2010), pages 2–3.

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

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

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

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

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

[KKMW]   M. Koloğlu, G. Kopp, S. J. Miller and Y. Wang, *On the number of summands in Zeckendorf decompositions*, to appear in the Fibonacci Quartely.

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

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

[MW]   S. J. Miller and Y. Wang, *From Fibonacci Numbers to Central Limit Type Theorems*, preprint, 2010. http://arxiv.org/abs/1008.3202

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

[RS]   R. Stanley, *Enumerative Combinatorics*, Cambridge University Press, Cambridge, UK, 2000–2001.

[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.

[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 Liège **41** (1972), pages 179–182.