

THE DISTRIBUTION OF GAPS BETWEEN SUMMANDS IN GENERALIZED ZECKENDORF DECOMPOSITIONS

AMANDA BOWER, RACHEL INSOF, SHIYU LI, STEVEN J MILLER, AND PHILIP TOSTESON

ABSTRACT. Zeckendorf proved any integer can be decomposed uniquely as a sum of non-adjacent Fibonacci numbers, F_n . Using continued fractions, Lekkerkerker proved the average number of summands of an $m \in [F_n, F_{n+1})$ is essentially $n/(\varphi^2 + 1)$, with φ the golden ratio. This result has been generalized by many, often using Markov processes, to show that for any positive linear recurrence the number of summands in decompositions for integers in $[G_n, G_{n+1})$ converges to a Gaussian distribution. Below we instead follow and further develop the combinatorial approach of Miller and Wang, and show its power in handling related questions. Specifically, we prove the probability of a gap larger than the recurrence length converges to decaying geometrically, and that the distribution of the smaller gaps depends in a computable way on the coefficients of the recurrence. These results hold both for the average over all $m \in [G_n, G_{n+1})$ and almost surely for the gap measure associated to individual m . These techniques also determine the distribution of the longest gap between summands, which we prove is similar to the distribution of the longest gap between heads in tosses of a biased coin. It is a double exponential strongly concentrated about the mean, and is on the order of $\log n$ with computable constants depending on the recurrence.

1. INTRODUCTION

1.1. Background. We explore the distribution between summands in generalized Zeckendorf decompositions. These generalize the standard base b expansions, and have the following desirable properties: (1) existence (every positive integer has a decomposition), (2) uniqueness (there is only one decomposition for each number), and (3) sparseness (many of the possible summands are not used). The latter property suggests that such decompositions can have applications in computer science, where storage costs are a major issue. Zeckendorf [Ze] proved that every positive integer can be written uniquely as a sum of non-adjacent Fibonacci numbers $F_1 = 1, F_2 = 2$ and $F_{n+2} = F_{n+1} + F_n$. The standard proof is by a greedy algorithm, and illustrates the naturalness of the non-adjacency condition.

Lekkerkerker [Lek] proved that for $m \in [F_n, F_{n+1})$, as $n \rightarrow \infty$ the average number of summands needed is $n/(\varphi^2 + 1)$, with $\varphi = \frac{1+\sqrt{5}}{2}$ the golden mean. More generally, we may replace the Fibonacci with other sequences and ask whether or not a decomposition exists with our three desired properties. The following theorem gives a large class of recurrence relations where such a decomposition exists, and gives the analogue of non-adjacency (essentially we cannot use the recurrence relation to reduce our decomposition). See for example [Day, GrTi, Ha, Ho, Ke, Len, MW1, MW2] for some of the history and proofs of results along these lines, [Al, DDKMV] for generalizations to far-difference representations (where we allow signed decompositions), and [DDKMMV] for other generalizations of the notion of a legal decomposition.

Date: April 18, 2015.

2010 Mathematics Subject Classification. 11B39, 11B05 (primary) 65Q30, 60B10 (secondary).

Key words and phrases. Zeckendorf decompositions, positive linear recurrence relations, longest gap.

The fourth named author was partially supported by NSF grants DMS0970067 and DMS1265673, and the remaining authors were partially supported by NSF Grant DMS0850577. It is a pleasure to thank our colleagues from the Williams College 2010, 2011, 2012 and 2013 SMALL REU program for many helpful conversations, Philippe Demontigny and Cameron Miller for discussions on generalizations, and the referees for helpful comments on earlier drafts. Contact information for corresponding author: Steven J. Miller, Department of Mathematics and Statistics, 18 Hoxsey St, Williams College, Williamstown, MA 01267; phone 413-597-3293, email sjm1@williams.edu.

Theorem 1.1 (Generalized Zeckendorf and Lekkerkerker Theorems). *Consider a **positive linear recurrence** $G_{n+1} = c_1 G_n + \dots + c_L G_{n+1-L}$ with non-negative integer coefficients c_i with $c_1, c_L, L \geq 1$, and initial conditions $G_1 = 1$ and $G_{n+1} = c_1 G_n + c_2 G_{n-1} + \dots + c_n G_1 + 1$ for $1 \leq n \leq L$. For each integer $m > 0$ there exists a unique **legal** decomposition $\sum_{i=1}^k a_i G_{N+1-i}$ with $a_1 > 0$, the other $a_i \geq 0$, and one of the following two conditions, which define a legal decomposition, holds.*

- We have $N < L$ and $a_i = c_i$ for $1 \leq i \leq N$.
- There exists an $s \in \{1, \dots, L\}$ such that $a_1 = c_1, a_2 = c_2, \dots, a_{s-1} = c_{s-1}$ and $a_s < c_s, a_{s+1}, \dots, a_{s+\ell} = 0$ 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.

As the decomposition is unique we define $k(m)$, the number of summands in m 's decomposition, by $k(m) := \sum_{i=1}^m a_i$; we often write $m = \sum_{j=1}^{k(m)} G_{r_j}$.

There exist constants $C_{\text{Lek}} > 0$ and d such that as $n \rightarrow \infty$ the average number of summands in a generalized Zeckendorf decomposition of integers in $[G_n, G_{n+1})$ is $C_{\text{Lek}} n + d + o(1)$.

After determining the mean number of summands, the next question is the variance or, more generally, the distribution of the fluctuations about the mean. Using techniques from ergodic theory and number theory the fluctuations about the mean were shown to converge to a Gaussian (see among others [DrGa, DuTh, FGNPT, GTNP, LaTh, Ste]). Using a more combinatorial perspective, Koloğlu, Kopp, Miller and Wang [KKMW, MW1, MW2] reproved these results for the positive linear recurrences studied above, and their proof applies to the far-difference representations as well (see [DDKMV], as well as [CFHMN] for a linear recurrence example where $c_1 = 0$); we use this result and perspective in our study of gaps (see §1.5 for a discussion on alternative viewpoints and other, related problems that one could investigate).

Theorem 1.2 (Gaussian Behavior of Summands in Generalized Zeckendorf Decompositions). *Let $\{G_n\}$ be a positive linear recurrence as in Theorem 1.1. For each $m \in [G_n, G_{n+1})$ let $k(m)$ be the number of summands in m 's generalized Zeckendorf decomposition. As $n \rightarrow \infty$ the distribution of the $k(m)$'s for $m \in [G_n, G_{n+1})$ converges to a Gaussian¹ with explicitly computable mean of order $C_{\text{Lek}} n$ (for some $C_{\text{Lek}} > 0$) and variance of order n .*

1.2. Notation. Before stating our results we first set some notation and recall a needed result. Let $\{G_n\}$ be a positive linear recurrence and take $m \in [G_n, G_{n+1})$. Notice that G_n must be a summand in the decomposition of m , and if the coefficient c_1 in the defining recurrence of G_n is greater than 1 then it is possible to have multiple copies of G_n in m 's decomposition. We write m as

$$m = \sum_{j=1}^{k(m)} G_{r_j} \quad (r_{k(m)} = n). \quad (1.1)$$

For example, if $m = 1G_{701} + 24G_{601} + 2013G_2$, we find 2035 gaps of length 0 (2012 coming from $2013G_2$ and 23 from $24G_{601}$), one gap of length 599 (coming from G_{601} and G_2), and one gap of length 100 (from G_{701} and G_{601}). By Theorem 1.2 the $k(m)$'s converge to being normally distributed with mean of order n and standard deviation of order \sqrt{n} ; in particular, most $k(m)$'s are close, on an absolute scale, to the mean. Below are the statistics we study.

- *Spacing gap measure:* The spacing gap measure of an $m \in [G_n, G_{n+1})$ with $k(m)$ summands is

$$\nu_{m;n}(x) := \frac{1}{k(m) - 1} \sum_{j=2}^{k(m)} \delta(x - (r_j - r_{j-1})), \quad (1.2)$$

¹When we say the number of summands converges to a Gaussian this means that as $n \rightarrow \infty$ the fraction of $m \in [G_n, G_{n+1})$ such that the number of summands in m 's Zeckendorf decomposition is in $[\mu_n - a\sigma_n, \mu_n + b\sigma_n]$ converges to $\frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt$, where μ_n is the mean number of summands for $m \in [G_n, G_{n+1})$ and σ_n^2 is the variance.

where δ is the Dirac delta functional.² We do not include the gap to the first summand, as this is not a gap *between* summands; for almost all m one extra gap is negligible in the limit.

- *Average spacing gap measure:* If $k(m)$ is the number of summands in m 's generalized Zeckendorf decomposition, then it has $k(m) - 1$ gaps. Thus the total number of gaps for all $m \in [G_n, G_{n+1})$ is

$$N_{\text{gaps}}(n) := \sum_{m=G_n}^{G_{n+1}-1} (k(m) - 1), \quad (1.3)$$

and by the Generalized Lekkerkerker Theorem we have

$$N_{\text{gaps}}(n) = C_{\text{Lek}} n (G_{n+1} - G_n) + O(G_{n+1} - G_n). \quad (1.4)$$

We define the average spacing gap measure for all $m \in [G_n, G_{n+1})$ by

$$\begin{aligned} \nu_n(x) &:= \frac{1}{N_{\text{gaps}}(n)} \sum_{m=G_n}^{G_{n+1}-1} \sum_{j=2}^{k(m)} \delta(x - (r_j - r_{j-1})) \\ &= \frac{1}{N_{\text{gaps}}(n)} \sum_{m=G_n}^{G_{n+1}-1} (k(m) - 1) \nu_{m;n}(x). \end{aligned} \quad (1.5)$$

Equivalently, if $P_n(k)$ is the probability of getting a gap of length k among all gaps from the decompositions of all $m \in [G_n, G_{n+1})$, then

$$\nu_n(x) = \sum_{k=0}^{n-1} P_n(k) \delta(x - k). \quad (1.6)$$

- *Limiting average spacing gap measure, limiting gap probabilities:* If the limits exist, we let

$$\nu(x) = \lim_{n \rightarrow \infty} \nu_n(x), \quad P(k) = \lim_{n \rightarrow \infty} P_n(k). \quad (1.7)$$

We prove these limits exist, and determine them.

- *Longest gap:* Given a decomposition $m = G_{r_1} + G_{r_2} + \dots + G_{r_{k(m)}}$ for $m \in [G_n, G_{n+1})$, the longest gap, denoted $L_n(m)$, is the maximum difference between adjacent indices in m 's decomposition. Thus $L_n(m) := \max_{2 \leq j \leq k(m)} |r_j - r_{j-1}|$.

1.3. Results: Gaps in the Bulk.

We can now state our results for gaps in the bulk.

Theorem 1.3 (Average Gap Measure in the Bulk). *Let $\{G_n\}$ be a positive linear recurrence of length L as in Theorem 1.1, and assume each $c_i \geq 1$. Let $\lambda_1 > 1$ denote the largest root (in absolute value) of the characteristic polynomial of the G_n 's, and let a_1 be the leading coefficient in the Generalized Binet expansion (thus $G_n = a_1 \lambda_1^n + o(\lambda_1^n)$; see Appendix A). Let $P_n(k)$ be the probability of having a gap of length k among the decompositions of $m \in [G_n, G_{n+1})$, and let $P(k) = \lim_{n \rightarrow \infty} P_n(k)$. Then*

$$P(k) = \begin{cases} 1 - (\frac{a_1}{C_{\text{Lek}}})(2\lambda_1^{-1} + a_1^{-1} - 3) & \text{if } k = 0 \\ \lambda_1^{-1} (\frac{1}{C_{\text{Lek}}})(\lambda_1(1 - 2a_1) + a_1) & \text{if } k = 1 \\ (\lambda_1 - 1)^2 \left(\frac{a_1}{C_{\text{Lek}}}\right) \lambda_1^{-k} & \text{if } k \geq 2. \end{cases} \quad (1.8)$$

In particular, the probability of having a gap of length $k \geq 2$ decays geometrically, with decay constant the largest root of the characteristic polynomial.

²Thus for any continuous function f we have $\int_{-\infty}^{\infty} f(x) \delta(x - a) dx = f(a)$; we may view $\delta(x - a)$ as representing a unit point mass concentrated at a .

We included the condition $c_i \geq 1$ above to simplify the algebra. An analogue of the above theorem holds for general positive linear recurrences, but the counting becomes more involved and it is not as easy to extract nice closed form expressions. For such recurrences it is clear that there is geometric decay for gaps larger than the recurrence length L , but the behavior for $k < L$ depends greatly on which c_i 's vanish.

We isolate some important examples. For base B decompositions, $P(0) = (B-1)(B-2)/B^2$, and for $k \geq 1$, $P(k) = c_B B^{-k}$, with $c_B = (B-1)(3B-2)/B^2$, while for the standard Zeckendorf decomposition $P(k) = 1/\varphi^k$ for $k \geq 2$, with $\varphi = \frac{1+\sqrt{5}}{2}$ the golden mean.

The proof of Theorem 1.3 falls from a careful counting of the number of times each gap length occurs. For $k \geq 0$ let

$$X_{i,i+k}(n) = \#\{m \in [G_n, G_{n+1}) : G_i, G_{i+k} \text{ in } m\text{'s decomposition, but not } G_{i+q} \text{ for } 0 < q < k\}. \quad (1.9)$$

Note we can deduce the $k = 0$ behavior if we know the answer for each $k \geq 1$. Then

$$P(k) = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^{n-k} X_{i,i+k}(n)}{N_{\text{gaps}}(n)}. \quad (1.10)$$

The denominator is well-understood by Lekkerkerker's Theorem; the proof of Theorem 1.3 follows from a good analysis of $X_{i,i+k}(n)$, which can be deduced from the uniqueness arguments in the generalized Zeckendorf decompositions. We give the proof in §2.

Theorem 1.3 gives the limiting behavior of the *average* of the individual gap measures $\nu_{m;n}(x)$. As $n \rightarrow \infty$, for almost all $m \in [G_n, G_{n+1})$ the *individual* measures are close to the average gap measure.

Theorem 1.4 (Individual Gap Measure in the Bulk). *Let $\{G_n\}$ be a positive linear recurrence as in Theorem 1.1, with the additional assumption that each $c_i \geq 1$. As $n \rightarrow \infty$, the individual gap measures $\nu_{m;n}(x)$ converge almost surely in distribution³ to the limiting gap measure from Theorem 1.3.*

We sketch the main ideas of the proof. Let $\widehat{\nu_{m;n}}(t)$ denote the characteristic function⁴ of $\nu_{m;n}(x)$, and $\widehat{\nu}(t)$ the characteristic function of the average gap distribution from Theorem 1.3. Let $\mathbb{E}_m[\cdots]$ denote the expectation over all $m \in [G_n, G_{n+1})$, with all m equally likely to be chosen. We first show that $\lim_{n \rightarrow \infty} \mathbb{E}_m[\widehat{\nu_{m;n}}(t)]$ equals $\widehat{\nu}(t)$, and then show $\lim_{n \rightarrow \infty} \mathbb{E}_m[(\widehat{\nu_{m;n}}(t) - \widehat{\nu}(t))^2] = 0$. This allows us to invoke Lévy's continuity theorem to obtain convergence in distribution for almost all $m \in [G_n, G_{n+1})$ as $n \rightarrow \infty$. We replace $k(m)$ with its average (and use the Gaussianity results to control the error), and introduce more general indicator functions such as $X_{i,i+g_1,j,j+g_2}(n)$, reducing the proof to a counting problem.

1.4. Results: Longest Gap.

If $G_{n+1} = 2G_n$, then the distribution of the longest gap $L_n(m)$ for $m \in [G_n, G_{n+1})$ is essentially that of the longest run of consecutive tails in n tosses of a fair coin whose final toss is a head. The answer for coins is well-known, both for fair and biased coins (see for example [Sch]). What is particularly remarkable about the coin toss problem is how tightly concentrated the answer is about the mean. For a coin with probability p of heads and $q = 1 - p$ of tails, the expected longest run of heads is

$$\log_{1/p}(nq) - \frac{\gamma}{\log p} - \frac{1}{2} + r_1(n) + \epsilon_1(n) = \frac{\log(nq)}{\log(1/p)} + \frac{\gamma}{\log(1/p)} - \frac{1}{2} + r_1(n) + \epsilon_1(n) \quad (1.11)$$

while the variance is

$$\frac{\pi^2}{6 \log^2 p} + \frac{1}{12} + r_2(n) + \epsilon_2(n), \quad (1.12)$$

³A sequence of random variables R_1, R_2, \dots with corresponding cumulative distribution functions F_1, F_2, \dots converges in distribution to a random variable R with cumulative distribution F if $\lim_{n \rightarrow \infty} F_n(r) = F(r)$ for each r where F is continuous.

⁴The characteristic function of a random variable X is $\mathbb{E}[e^{itX}]$, with a similar definition for a measure. We denote the characteristic function of a measure μ by $\widehat{\mu}$, as it is the Fourier transform of the measure (up to a normalization constant).

where γ is Euler's constant, the $r_i(n)$ are at most .000016, and the $\epsilon_i(n)$ tend to zero as $n \rightarrow \infty$. Very importantly, the variance is bounded independently of n (by essentially 3.5). This implies that there is essentially no fluctuation of the observed longest string of heads. We find similar behavior, both in terms of the logarithmic size of the longest term in our sequence as well as the strong concentration about the average.

Before we can state our results, however, we need to introduce some notation. It is technically more convenient to rewrite the recurrence relation where we only record the **non-zero** coefficients. Thus, in the sections on longest gaps, we write our positive linear recurrence as

$$G_{n+1} = c_{j_1+1}G_{n-j_1} + c_{j_2+1}G_{n-j_2} + \cdots + c_{j_L+1}G_{n-j_L}, \quad (1.13)$$

where $j_1 = 0, j_1 < j_2 < \cdots < j_L$, and all recurrence coefficients not shown are zero.

Theorem 1.5 (Longest gap in generalized Zeckendorf expansions). *Let $\{G_n\}$ be a positive linear recurrence as in Theorem 1.1, and assume the associated polynomials $\mathcal{M}(s) = 1 - c_1s - c_{j_2+1}s^{j_2+1} - \cdots - c_{j_L+1}s^{j_L+1}$ and $\mathcal{R}(s) = c_1 + c_{j_2+1}s^{j_2} + \cdots + (c_{j_L+1} - 1)s^{j_L}$ do not have multiple roots or roots of absolute value 1. Let λ_1 be the largest root of the recurrence for G_n , $\mathcal{G}(s) = -\mathcal{M}(s)/(s - 1/\lambda_1)$ and*

$$P(n, f) := \frac{\#\{m \in [G_n, G_{n+1}) : L_n(m) < f\}}{G_{n+1} - G_n} \quad (1.14)$$

be the cumulative distribution of the longest gap in the Zeckendorf decompositions of $m \in [G_n, G_{n+1})$; note we are computing gaps strictly less than f , and we do not include the gap in the beginning.

(1) *Asymptotically, for any $R_{\min} \in \mathbb{R}$ with $\lambda_1 < R_{\min} < 1$ we have*

$$P(n, f) = \exp\left(-n\lambda_1^{-f} \frac{\lambda_1 \mathcal{R}(\frac{1}{\lambda_1})}{\mathcal{G}(\frac{1}{\lambda_1})}\right) + O\left(nf \left(\frac{R_{\min}}{\lambda_1}\right)^f + n \left(\frac{1}{\lambda_1}\right)^{2f} + f \left(\frac{1}{\lambda_1 R_{\min}}\right)^n\right). \quad (1.15)$$

(2) *Let $K = \lambda_1 \mathcal{R}(1/\lambda_1)/\mathcal{G}(1/\lambda_1)$ and γ be Euler's constant. The mean of the longest gap, μ_n , and the variance of the longest gap, σ_n^2 , are given by*

$$\mu_n = \frac{\log(nK)}{\log \lambda_1} + \frac{\gamma}{\log \lambda_1} - \frac{1}{2} + o(1), \quad \sigma_n^2 = \frac{\pi^2}{6(\log \lambda_1)^2} + o(1). \quad (1.16)$$

The proof proceeds by introducing a generating function for the longest gap distribution, where we obtain the probabilities by analyzing the cumulative distribution function. We use a partial fraction decomposition to extract information from the generating function, and use Rouché's theorem (among others) to deal with the technicalities that arise. Taking into account that we do not consider the initial segment, our theorem applied to $G_{n+1} = 2G_n$ is consistent with tosses of a fair coin.

The fit between numerics and theory is excellent. For example, consider the Fibonacci numbers.⁵ We chose 100 numbers randomly from $[F_n, F_{n+1})$ with $n = 1,000,000$. We observed a mean of 28.51 and a standard deviation of 2.64, which compares very well with the predictions of 28.73 and 2.67. Increasing n to 10,000,000 and looking at 20 randomly chosen numbers yielded a mean of 33.6 and a standard deviation of 2.33, again close to the predictions of 33.52 and 2.665.⁶

⁵Due to costs to store and recall objects from memory, and to use the Binet formula, we found it best to use Binet's formula to find F_n and F_{n+1} , and then use the recurrence to compute backwards.

⁶We saw similar behavior in other recurrences. For $a_{n+1} = 2a_n + 4a_{n-1}$, when $n = 51,200$ (respectively 102,400) the predicted mean was 9.95 (resp. 10.54) and the standard deviation was 1.09; choosing 100 points randomly in the interval yielded a mean of 9.91 (resp. 10.45) and a standard deviation of 1.22 (resp. 1.10).

1.5. Extensions and structure of the paper. Ben-Ari and Miller [B-AM] prove similar theorems as in this paper, based on the observation that the resulting distributions coincide with the distribution of a conditioned Markov chain. This allows them to obtain results on decompositions from analogous results for Markov chains, and gives some insight as to why the gap distribution behaves as it does. One of the main goals of this paper is to highlight and further develop the combinatorial approach and machinery from [KKMW, MW1, MW2], which gives similar results but more elementarily (as well as error terms in some cases); we hope that this vantage will be of use to other researchers studying related problems.

We have chosen to study the distribution of gaps for numbers in $[G_n, G_{n+1})$. We made the decision to localize in this manner as now each element has the same candidate summand set *and* the same largest summand; this is similar in spirit to studying primes p in the interval $[x, 2x)$ so that the weights $\log p$ are essentially constant in the interval. Similar results can easily be derived for the interval $[1, G_n)$ (this is immediate as the G_n 's grow exponentially, and thus most of $[1, G_n)$ is covered by the intervals $[G_{n-\ell-1}, G_{n-\ell})$ for $0 \leq \ell \leq \sqrt{\log n}$, and all these sub-intervals have essentially the same behavior for the statistics of interest). One can also study other cut-offs, and similar behavior is observed for almost all starting points for sub-intervals that are not too small (see [BEDMMTW1, BEDMMTW2]). With a little additional work, one can also extend all our theorems to the general initial segments $[1, N)$ with $N \rightarrow \infty$. For these intervals the Markov perspective of [B-AM] is particularly well-suited, and is cleaner than the combinatorial discussions above. We discuss these extensions in Appendix C.

In §2 we prove Theorem 1.3 for the average gap measure in the bulk, and prove almost sure convergence for the individual measures in §3. We prove Theorem 1.5 in §4, and conclude with some final remarks. For completeness we do all standard calculations in the arXiv version, [BILMT].

2. GAPS IN THE BULK I: AVERAGE BEHAVIOR

In this section we prove Theorem 1.3. Our combinatorial approach begins by computing $X_{i,i+k}(n)$, which allows us to find $P_n(k)$. We can determine $X_{i,i+k}(n)$ by counting the number of choices of the summands $\{G_1, G_2, \dots, G_n\}$ such that G_i, G_{i+k} and G_n are chosen, no summand whose index is between i and $i+k$ is chosen, and all other indices are free to be chosen subject to the requirement that we have a legal decomposition. Let $L_{i,i+k}(n)$ and $R_{i,i+k}(n)$ be the number of ways to choose a valid subset of summands from those before the gap of length k starting at G_i and after the gap (respectively). Since

$$G_{j+1} = c_1 G_j + \dots + c_L G_{j+1-L} \quad (2.1)$$

where $c_i \geq 1$, any time we have a gap of length $k > 1$, the recurrence ‘resets’ itself. We see that $L_{i,i+k}(n)$ and $R_{i,i+k}(n)$ are independent of each other when $k \geq 2$; thus for $k \geq 2$ we have

$$X_{i,i+k}(n) = L_{i,i+k}(n) \cdot R_{i,i+k}(n). \quad (2.2)$$

The behavior for $k \leq 1$ is more delicate due to the dependencies, but follows from a careful counting.

We have the following counting lemma.

Lemma 2.1. *Let $\{G_n\}$ be a positive linear recurrence as in Theorem 1.1 with each $c_i \geq 1$. Consider all $m \in [G_n, G_{n+1})$ with a gap of length $k \geq 2$ starting at G_i for $1 \leq i \leq n-k$. The number of valid choices for subsets of summands before the gap, $L_{i,i+k}(n)$, is*

$$L_{i,i+k}(n) = G_{i+1} - G_i, \quad (2.3)$$

while the number of valid choices for subsets of summands after the gap, $R_{i,i+k}(n)$, is

$$R_{i,i+k}(n) = G_{n-i-k+2} - 2G_{n-i-k+1} + G_{n-i-k}. \quad (2.4)$$

Proof. To count $L_{i,i+k}(n)$, we count the number of ways to have a legal decomposition that must have the summand G_i such that all other summands which are less than G_i are free to be chosen or not. It is very important that $k \geq 2$, as this means the summand at G_{i+k} does not interact with the summands earlier than G_i through the recurrence relation. Thus $L_{i,i+k}(n)$ is the same as the number of legal choices of summands

from $\{G_1, G_2, \dots, G_i\}$ with G_i chosen. As each integer in $[G_i, G_{i+1})$ has a unique legal decomposition with G_i chosen, we see $L_{i,i+k}(n)$ equals the number of elements in this interval, which is just $G_{i+1} - G_i$.

To compute $R_{i,i+k}(n)$, we consider how many ways we can choose summands from $\{G_{i+k}, \dots, G_n\}$ such that G_{i+k} and G_n are chosen and the resulting decomposition is legal; since $k \geq 2$ the summands from G_i and earlier cannot affect our choices here. Thus our problem is equivalent to asking how many legal ways there are to choose summands from $\{G_1, \dots, G_{n-i-k+1}\}$ with $G_1, G_{n-i-k+1}$ both chosen and the rest free. There are many ways to compute this; the simplest is to note that this equals the number of legal choices with $G_{n-i-k+1}$ chosen and where we *may or may not choose* G_1 , minus the number of legal choices with $G_{n-i-k+1}$ chosen where we *do not choose* G_1 . By a similar argument as above, the first count is $G_{n-i-k+2} - G_{n-i-k+1}$ (as it is the number of legal decompositions of a number in $[G_{n-i-k+1}, G_{n-i-k+2})$), while the second is $G_{n-i-k+1} - G_{n-i-k}$. The proof is completed by subtracting. \square

The next lemma counts how many legal decompositions have a gap of length one. The main idea of the proof is to remove the dependencies by breaking into cases and then arguing as above.

Lemma 2.2. *Let $\{G_n\}$ be a positive linear recurrence as in Theorem 1.1 such that $c_i \geq 1$. Consider all $m \in [G_n, G_{n+1})$ with a gap of length 1 starting at G_i for $1 \leq i \leq n-1$. Then*

$$X_{i,i+1}(n) = (G_{n+1} - G_n) - G_{i+1}(G_{n-i} - G_{n-i-1}) - G_i(G_{n-i+1} - 2G_{n-i} + G_{n-i-1}). \quad (2.5)$$

Proof. We cannot count as in Lemma 2.1, since $L_{i,i+1}(n)$ and $R_{i,i+1}$ are no longer independent. Instead, we consider the total number of decompositions in $[G_n, G_{n+1})$ (which is $G_{n+1} - G_n$) and subtract off the three different ways to *not* have a gap of length one starting at G_i for a decomposition: (1) not including G_i and not including G_{i+1} , (2) including G_i but not including G_{i+1} , and finally (3) not including G_i but including G_{i+1} . In each case we can use the methods of Lemma 2.1 since there are no dependency issues. \square

We now prove Theorem 1.3. We use little-oh and big-Oh notation for the lower order terms, which do not matter in the limit.

Proof of Theorem 1.3. There are three cases to consider: $k = 0$, $k = 1$ and $k \geq 2$. When $k \geq 1$ we use the generalized Binet's formula and take limits. When $k = 0$ it is harder to count gaps of length 0 since a decomposition could have multiple gaps of length 0 at G_i ; fortunately we can deduce the number of these gaps by knowing the number of gaps with $k \geq 1$.

As our analysis of gaps of length k had different answers for $k = 1$ and $k \geq 2$, we first consider the case when $k \geq 2$. We need to compute

$$P(k) = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^{n-k} X_{i,i+k}(n)}{N_{\text{gaps}}(n)}. \quad (2.6)$$

By Lemma 2.1,

$$X_{i,i+k}(n) = L_{i,i+k}(n) \cdot R_{i,i+k}(n) = (G_{i+1} - G_i) \cdot (G_{n-i-k+2} - 2G_{n-i-k+1} + G_{n-i-k}), \quad (2.7)$$

and by Binet's formula (Lemma A.1)

$$G_i = a_1 \lambda_1^i + O(i^{L-2} \lambda_2^i) = a_1 \lambda_1^i (1 + O(i^{L-2} (\lambda_2/\lambda_1)^i)). \quad (2.8)$$

We want to use little-oh notation for the error term above; unfortunately the error is not necessarily small if i is close to 0. The error is $o(1)$ if i is at least $\log^2 n$ and is bounded for smaller i . Thus we introduce the notation $o_{i;n}(1)$ for an error that is $o(1)$ for $i \geq \log^2 n$ and bounded otherwise. We have

$$\begin{aligned} X_{i,i+k}(n) &= a_1 \lambda_1^i (\lambda_1 - 1) (1 + o_{i;n}(1)) \cdot a_1 \lambda_1^{n-i-k} (\lambda_1^2 - 2\lambda_1 + 1) (1 + o_{n-i-k;n}(1)) \\ &= a_1^2 \lambda_1^{n-k} (\lambda_1 - 1)^3 (1 + o_{i;n}(1) + o_{n-i-k;n}(1)). \end{aligned} \quad (2.9)$$

As

$$N_{\text{gaps}}(n) = C_{\text{Lek}} n (G_{n+1} - G_n) + O(G_{n+1} - G_n) = C_{\text{Lek}} \cdot n \cdot a_1 \cdot \lambda_1^n (\lambda_1 - 1) + O(\lambda_1^n), \quad (2.10)$$

we find

$$P_n(k) = \frac{\sum_{i=1}^{n-k} X_{i,i+k}(n)}{N_{\text{gaps}}(n)} = (\lambda_1 - 1)^2 \left(\frac{a_1}{C_{\text{Lek}}} \right) \lambda_1^{-k} (1 + o(1)), \quad (2.11)$$

as the sum over $i \leq \log^2 n$ and $i \geq n - k - \log^2 n$ is negligible. By taking the limit, which clearly exists for each n and each $k \geq 2$, we obtain the claimed expression for $P(k)$ for $k \geq 2$.

If $k = 1$ we use Lemma 2.2 to evaluate $X_{i,i+1}(n)$ and use a similar argument as in the $k \geq 2$ case, which gives $P(1)$. When $k = 0$, since probability distributions must sum to one, after some algebra we find

$$P(0) = 1 - \left(P(1) + \sum_{k=2}^{\infty} P(k) \right) = 1 - \left(\frac{a_1}{C_{\text{Lek}}} \right) (2\lambda_1^{-1} + a_1^{-1} - 3), \quad (2.12)$$

which completes the proof. \square

3. GAPS IN THE BULK II: INDIVIDUAL MEASURES

In this section we prove Theorem 1.4. Recall the spacing gap measure of $m \in [G_n, G_{n+1})$ with decomposition given in (1.1) with $k(m)$ summands is defined to be

$$\nu_{m;n}(x) = \frac{1}{k(m) - 1} \sum_{j=2}^{k(m)} \delta(x - (r_j - r_{j-1})). \quad (3.1)$$

We first recall and set some notation.

- $\widehat{\nu_{m;n}}(t)$: The characteristic function of $\nu_{m;n}(x)$.
- $\widehat{\nu}(t)$: The characteristic function of the average gap distribution $\nu(x)$ from Theorem 1.3.
- $\mathbb{E}_m[\cdot]$: The expected value over $m \in [G_n, G_{n+1})$ with the uniform measure; thus if $X : [G_n, G_{n+1}) \rightarrow \mathbb{R}$ then

$$\mathbb{E}_m[X] := \frac{1}{G_{n+1} - G_n} \sum_{m=G_n}^{G_{n+1}-1} X(m). \quad (3.2)$$

- $X_{j_1, j_1+g_1, j_2, j_2+g_2}$: The number of $m \in [G_n, G_{n+1})$ with a gap of length exactly g_1 starting at j_1 and a gap of length exactly g_2 starting at j_2 (we have suppressed the subscript n as it is always understood from context). If g_1 or g_2 is zero then we count with multiplicity. For example, if $g_1 = 0$ and $g_2 = 3$ then an m that has 5 summands at G_{j_1} and has G_{j_2} and G_{j_2+3} (but no summands between these last two) is counted four times. We similarly count with multiplicity if we have X_{j_1, j_1+g_1} .

We sketch the proof. We use Lévy's continuity theorem [FG], which says that if we have a sequence of random variables $\{R_r\}$ (which do not have to be defined on the same probability space) whose characteristic functions $\{\varphi_r\}$ converge pointwise to the characteristic function φ of a random variable R , then the random variables $\{R_r\}$ converge in distribution to R (i.e., the cumulative distribution functions of the $\{R_r\}$ converge to that of $\{R\}$ at all points of continuity). We show given any ϵ there is an N_ϵ such that for all $n \geq N_\epsilon$ the characteristic functions $\widehat{\nu_{m;n}}(t)$ are pointwise within ϵ for almost all $m \in [G_n, G_{n+1})$; we can't have all the characteristic functions close, as some m have very few gaps.

Step 1 is to show that $\mathbb{E}_m[\widehat{\nu_{m;n}}(t)] = \widehat{\nu}(t)$. A key ingredient is to remove the individual normalizations of $\frac{1}{k(m)-1}$, where $k(m)$ is the number of summands in the generalized Zeckendorf decomposition of m ; we can replace these with their average up to a negligible error term because of previous work on the Gaussian behavior of the number of summands. To complete the proof, we prove that most characteristic functions are concentrated near the mean. We do this in step 2 by showing

$$\lim_{n \rightarrow \infty} \mathbb{E}_m \left[(\widehat{\nu_{m;n}}(t) - \widehat{\nu}(t))^2 \right] = 0,$$

which follows by reducing the problem to determining $X_{j_1, j_1+g_1, j_2, j_2+g_2}$.

3.1. Expected Value of Individual Characteristic Functions. The first step towards a proof of Theorem 1.4 is to show that the expected value of the individual characteristic functions of the gap measures converge to the characteristic function of the average gap measure. Convergence in distribution follows from controlling the rate of convergence, which we handle in the next subsection.

Proposition 3.1. *Notation as above, we have*

$$\lim_{n \rightarrow \infty} \mathbb{E}_m[\widehat{\nu_{m;n}}(t)] = \widehat{\nu}(t). \quad (3.3)$$

We need some preliminary results before we can prove this proposition. Notice

$$\widehat{\nu_{m;n}}(t) = \int_0^\infty e^{ixt} \nu_{m;n}(x) dx = \frac{1}{k(m) - 1} \sum_{j=2}^{k(m)} e^{it(r_j - r_{j-1})}, \quad (3.4)$$

where

$$m = G_{r_1} + G_{r_2} + \cdots + G_{r_{k(m)}}. \quad (3.5)$$

Thus we have

$$\mathbb{E}_m[\widehat{\nu_{m;n}}(t)] = \frac{1}{G_{n+1} - G_n} \sum_{m=G_n}^{G_{n+1}-1} \frac{1}{k(m) - 1} \sum_{j=2}^{k(m)} e^{it(r_j - r_{j-1})}. \quad (3.6)$$

The difficulty in evaluating $\mathbb{E}_m[\widehat{\nu_{m;n}}(t)]$ is that we must deal with the presence of the $k(m) - 1$ factors. These vary with m , though weakly because of our Gaussianity result (Theorem 1.2). As the mean is of order n and the standard deviation is of order \sqrt{n} , the $k(m)$ are strongly concentrated about their mean. We first apply standard estimation arguments to show that we may safely replace $k(m)$ with its mean. We have

$$\frac{1}{k(m) - 1} = \frac{1}{C_{\text{Lek}}n + d + o(1)} - \frac{(k(m) - 1) - (C_{\text{Lek}}n + d + o(1))}{(k(m) - 1)(C_{\text{Lek}}n + d + o(1))}. \quad (3.7)$$

We essentially replace $\frac{1}{k(m)-1}$ with $\frac{1}{C_{\text{Lek}}n+d+o(1)}$ at a negligible cost, as the second summand above is extremely small most of the time and of moderate size almost never. We make this claim explicit in the next lemma.

Lemma 3.2. *Let $\{G_n\}$ be a positive linear recurrence as in Theorem 1.1, with each $c_i \geq 1$. Let $m \in [G_n, G_{n+1})$ have decomposition given by (1.1). Then for any fixed $t \geq 1$ we have*

$$\lim_{n \rightarrow \infty} \frac{1}{G_{n+1} - G_n} \sum_{m=G_n}^{G_{n+1}-1} \left(\frac{(k(m) - 1) - (C_{\text{Lek}}n + d + o(1))}{(k(m) - 1)(C_{\text{Lek}}n + d + o(1))} \right) \sum_{j=2}^{k(m)} e^{it(r_j - r_{j-1})} = 0, \quad (3.8)$$

where $C_{\text{Lek}}n + d + o(1)$ is the average number of summands needed in a decomposition for an integer in $[G_n, G_{n+1})$ in Theorem 1.1.

Proof. The distribution of the number of summands in a decomposition for $m \in [G_n, G_{n+1})$ converges to a Gaussian by Theorem 1.2. The average number of summands is $C_{\text{Lek}}n + d + o(1)$ (with $C_{\text{Lek}} > 0$) and the standard deviation is $b\sqrt{n} + o(\sqrt{n})$ for some $b > 0$. The proof is completed by partitioning based on the deviation of $k(m)$ from its expected value. Fix a $\delta > 0$ and let

$$I_n(\delta) := \left\{ m \in [G_n, G_{n+1}) : k(m) \in \left[C_{\text{Lek}}n + d - bn^{1/2+\delta}, C_{\text{Lek}}n + d + bn^{1/2+\delta} \right] \right\}. \quad (3.9)$$

Case 1: Let $m \in [G_n, G_{n+1}) \cap I_n(\delta)$; thus $k(m)$ is very close to $C_{\text{Lek}}n + d + o(1)$. To simplify the expressions below remember we are writing r_j and r_{j-1} for the indices in the decomposition of m ; while

we should really write $r_j(m)$, as the meaning is clear we prefer this more compact notation. Therefore

$$\frac{1}{G_{n+1} - G_n} \sum_{\substack{m=G_n \\ m \in I_n(\delta)}}^{G_{n+1}-1} \left(\frac{(k(m) - 1) - (C_{\text{Lek}}n + d + o(1))}{(k(m) - 1)(C_{\text{Lek}}n + d + o(1))} \right) \sum_{j=2}^{k(m)} e^{it(r_j - r_{j-1})} \ll n^{-1/2+\delta}. \quad (3.10)$$

Case 2: Let $k(m) \notin I_n(\delta)$; thus $k(m)$ is not too close to $C_{\text{Lek}}n + d$. As the distribution of the number of summands needed for a decomposition converges to a Gaussian by Theorem 1.2, for sufficiently large n the probability an $m \in [G_n, G_{n+1})$ has $k(m)$ more than n^δ standard deviations from the mean is essentially

$$2 \int_{bn^{\frac{1}{2}+\delta}}^{\infty} \frac{1}{\sqrt{2\pi b^2 n}} e^{-t^2/2b^2 n} dt \ll e^{-n^{2\delta}/2}. \quad (3.11)$$

Thus for sufficiently large n , the number of $m \in [G_n, G_{n+1})$ such that $k(m) \notin I_n(\delta)$ is essentially $(G_{n+1} - G_n) \cdot e^{-n^{2\delta}/2}$, and we find

$$\frac{1}{G_{n+1} - G_n} \sum_{\substack{m=G_n \\ m \notin I_n(\delta)}}^{G_{n+1}-1} \left(\frac{(k(m) - 1) - (C_{\text{Lek}}n + d + o(1))}{(k(m) - 1)(C_{\text{Lek}}n + d + o(1))} \right) \sum_{j=2}^{k(m)} e^{it(r_j - r_{j-1})} \ll e^{-n^{2\delta}/2}, \quad (3.12)$$

which tends rapidly to zero as $n \rightarrow \infty$. This completes the proof. \square

Remark 3.3. In calculating the variance, we need to approximate $(k(m) - 1)^{-2}$. A similar argument shows that this can be replaced at a negligible cost with $(C_{\text{Lek}}n + d + o(1))^{-2}$; the error in the resulting sums from these replacements is $o(1)$, and thus vanishes in the limit.

Proposition 3.1 now follows. Replace as in the remark above with negligible error by Lemma 3.2. We then pull this factor outside of the m summation, switch orders of summation and send $m \rightarrow \infty$.

3.2. Variance of the Individual Gap Measures. The last ingredient in our proof of Theorem 1.4 is to show that the variance of the characteristic functions of the individual measures tends to zero. We keep the argument as general as possible for as long as possible.

Proposition 3.4. *With the notation as above, we have*

$$\lim_{n \rightarrow \infty} \text{Var}_n(t) := \lim_{n \rightarrow \infty} \mathbb{E}_m[(\widehat{\nu_{m;n}}(t) - \widehat{\nu_n}(t))^2] = 0. \quad (3.13)$$

Proof. Let

$$X_{j_1, j_1+g_1, j_2, j_2+g_2}(n) := \# \left\{ m \in [G_n, G_{n+1}) : G_{j_1}, G_{j_1+g_1}, G_{j_2}, G_{j_2+g_2} \text{ in } m\text{'s decomposition,} \right. \\ \left. \text{but not } G_{j_1+q}, G_{j_2+p} \text{ for } 0 < q < g_1, 0 < p < g_2 \right\}; \quad (3.14)$$

if either g_1 or g_2 is zero then we count m 's with multiplicity equal to the number of gaps of length zero at j_1 or j_2 . Note

$$\text{Var}_n(t) := \mathbb{E}_m[(\widehat{\nu_{m;n}}(t) - \widehat{\nu_n}(t))^2] = \mathbb{E}_m[\widehat{\nu_{m;n}}(t)^2] - \widehat{\nu_n}(t)^2, \quad (3.15)$$

and we know $\lim_n \widehat{\nu_n}(t)^2$ from the proof of Proposition 3.1. We are left with finding $\mathbb{E}_m[\widehat{\nu_{m;n}}(t)^2]$. The algebra is long but standard (see [BILMT] for the calculation); we highlight the main ideas. As $\widehat{\nu_n}(t) = \widehat{\nu}(t) + o(1)$, by the triangle inequality it suffices to show $\lim_n \mathbb{E}_m[\widehat{\nu_{m;n}}(t)^2]$ converges to $\widehat{\nu}(t)^2$. For the limit of the average gap measure, the probability of a gap of length g is $P(g)$, and is given by Theorem 1.3. Thus

$$\widehat{\nu}(t)^2 = \sum_{g_1=0}^{\infty} P(g_1) e^{itg_1} \sum_{g_2=0}^{\infty} P(g_2) e^{itg_2} = \sum_{g_1, g_2} P(g_1) P(g_2) e^{it(g_1+g_2)}. \quad (3.16)$$

The proof follows by showing that $\lim_n \mathbb{E}_m[\widehat{\nu}_{m;n}(t)^2]$ differs from this by $o(1)$. We do this by showing that they are close for each pair (g_1, g_2) , with the difference summable and $o(1)$ over all pairs; the pairwise (almost) agreement follows by using our indicator variables $X_{j_1, j_1+g_1, j_2, j_2+g_2}$ and similar counting arguments as above (though a bit more involved as there are more cases to consider). See Appendix B for the details. \square

3.3. Proof of Theorem 1.4. We now turn to the proof of Theorem 1.4. We have already done the difficult part of the analysis in §3.1 and §3.2. As the proof of convergence follows from standard probability arguments, we just sketch the details below.

Proof of Theorem 3.1. To use Lévy's continuity theorem (see [FG]), we need a sequence of random variables $\{R_r\}$ (which do not have to be defined on the same probability space) whose characteristic functions $\{\varphi_r\}$ converge pointwise to the characteristic function φ of a random variable R . If we have this, then the random variables $\{R_r\}$ converge in distribution to R (i.e., the cumulative distribution functions of the $\{R_r\}$ converge to that of $\{R\}$ at all points of continuity).

For us, R is essentially a geometric decay (it's a pure geometric decay for gaps of length 2 or more), and for each n the R_r 's are the gap measures for each $m \in [G_n, G_{n+1})$. By our results on the convergence of the means $\mathbb{E}_m[\widehat{\nu}_{m;n}(t)]$ to $\widehat{\nu}(t)$ and the variance tending to zero, Chebyshev's inequality implies that given $\epsilon > 0$, for each n almost all m have $\widehat{\nu}_{m;n}(t)$ within ϵ of $\widehat{\nu}(t)$ (we are able to do this for all t simultaneously).

Our set $\{R_r\}$ is thus a collection of gap measures coming from $m \in [G_n, G_{n+1})$. We have shown that as $n \rightarrow \infty$, for almost all $m \in [G_n, G_{n+1})$ we have convergence of these measures to the average gap measure. Thus the conditions of Lévy's continuity theorem are satisfied, completing the proof. \square

4. LONGEST GAP

4.1. Overview. We briefly describe our approach to determining the distribution and limiting behavior of the longest gap in Zeckendorf decompositions. We first find a rational generating function $F(s, f)$, whose coefficients in s give the number of decompositions with longest gap less than f . This allows us to determine the cumulative distribution of the longest gap, which we expand with a partial fraction decomposition. It is here that we need our additional restrictions on the roots of the associated polynomials $\mathcal{M}(s)$ and $\mathcal{R}(s)$. These lead to simpler partial fraction expansions, and minimizes the technical obstructions.

In the process of obtaining this exact expression, we need several technical lemmas about the behavior of the roots of the polynomials in the denominator of our generating functions $F(s, f)$. In particular, in order to obtain estimates for the longest gap for large n , we use Rouché's theorem, and show that the distribution is essentially determined by the behavior of a single root. In turn, this root relates to the largest eigenvalue, λ_1 , of the recurrence relation of the G_i 's. Approximating along these lines, we determine an asymptotic expression for the cumulative distribution function $P(n, f)$, which in the limit is doubly exponential: $P(n, f) = \exp(Cn\lambda_1^{-f}) + o(1)$; here the constant C is a rational function of λ_1 .

The error term in $P(n, f)$ is sufficiently small to allow us to determine asymptotic expressions for the mean and variance of the longest gap. To do this, we sum over a sufficiently large interval (ℓ_n, h_n) containing the mean μ_n , take partial sums, and then use the Euler-Maclaurin formula to smooth out our expression. This yields a particularly nice asymptotic expression for the mean and variance of the longest gap. This result is directly analogous to behavior seen in flipping coins.

4.2. Exact Cumulative Distribution of the Longest Gap. Our first step is to determine the cumulative distribution function of the longest gap. We begin by counting the number of $m \in [G_n, G_{n+1})$ with $L_n(m)$ less than some $f \in \mathbb{N}$, and finding the associated generating function. As the longest gap grows on the order of $\log n$, it suffices to study $f \geq \log \log n$; in other words, in all arguments below we may assume f is much larger than the length of the recurrence relation, and thus we do not need to worry about small numbers.

Lemma 4.1. *Let $f > j_L$. The number of $m \in [G_n, G_{n+1})$ with longest gap less than f is given by the coefficient of s^n in the generating function*

$$F(s, f) = \frac{1 - s^{j_L}}{\mathcal{M}(s) + s^f \mathcal{R}(s)}, \quad (4.1)$$

where

$$\begin{aligned} \mathcal{M}(s) &= 1 - c_1 s - c_{j_2+1} s^{j_2+1} - \dots - c_{j_L+1} s^{j_L+1} \\ \mathcal{R}(s) &= c_1 + c_{j_2+1} s^{j_2} + \dots + (c_{j_L+1} - 1) s^{j_L} \end{aligned} \quad (4.2)$$

and the c_i and j_i are defined as in (1.13).

Before beginning our proof, we fix some notation. Recall that our recurrence relation is written as

$$G_{n+1} = c_{j_1+1} G_{n-j_1} + c_{j_2+1} G_{n-j_2} + \dots + c_{j_L+1} G_{n-j_L} \quad (4.3)$$

with each c_i non-zero and $j_1 = 0$. We write $g_{i-1} = j_i - j_{i-1}$ for the gaps, with the convention $g_0 = 1$.

- A **legal block of length ℓ** is a sequence of non-negative integers $(a_i)_{i=1}^\ell$ where $a_i = c_i$ for $i \leq \ell$ and $a_\ell < c_\ell$. Notice that a legal block a must be of length j_i for some i , in order to satisfy $a_i < c_{j_i}$.
- A **string of zeroes of length ℓ** is a sequence $(b_i)_{i=1}^\ell$ where each $b_i = 0$.
- Denote the **concatenation of two sequences** $a = (a_i)_{i=1}^{\ell_a}$ and $b = (b_i)_{i=1}^{\ell_b}$ by $a \rightarrow b$, where

$$a \rightarrow b = (u_i)_{i=1}^{\ell_a + \ell_b}, \quad (4.4)$$

with $u_i = a_i$ for $i \in [1, \ell_a]$ and $u_i = b_{i-\ell_a}$ for $i \in [\ell_a + 1, \ell_a + \ell_b]$.

- A **legal sequence** is a sequence of non-negative integers $(\rightarrow_{r=1}^k (\eta_r \rightarrow z_r)) \rightarrow T$, where the η_r are legal blocks, the z_r are strings of zeroes, and T is a **terminal block** (a sequence $(c_i)_{i=1}^\ell$ with $\ell \leq j_L$, with L the number of non-zero coefficients in the recurrence relation for the G_i 's; see (4.3)). Informally, a legal sequence consists of k legal blocks, separated by strings of zeroes, and ended by a terminal block. By definition, legal sequences of length n are exactly those sequences that arise as decompositions of $x \in [G_n, G_{n+1})$. We use $|a|$ to denote the length of a sequence a .
- Set $T_f(s) = \mathcal{M}(s) + s^f \mathcal{R}(s)$, with $\mathcal{M}(s)$ and $\mathcal{R}(s)$ as in (4.2). We denote its roots by $\alpha_{i,f}$, with $\alpha_{1,f}$ the smallest root. One of the difficulties in the analysis below is that these roots depend on f , though fortunately the only one that matters is $\alpha_{1,f}$, which exponentially converges to $1/\lambda_1$ (the reciprocal of the largest root of the characteristic polynomial of the recurrence relation for the G_i 's).

Proof of Lemma 4.1. By the Generalized Zeckendorf Theorem, Theorem 1.1, there exists a bijection between $m \in [G_n, G_{n+1})$ and legal decompositions of length n . Accordingly, we count the number of length n legal decomposition with longest gap less than f . As remarked above, we assume f is at least $\log \log n$ for n large, so in particular f is much greater than the length of the recurrence.

A gap of length g in the decomposition corresponds to a string of zeroes of length $g - 1$ contained in the legal sequence. To count the number of decompositions with longest gap less than f , we count the number of legal sequences of length n with all strings of zeroes of length $\leq f - 2$. First we consider legal blocks followed by a string of 0's, or sequences of the form $\eta \rightarrow z$ where η is a legal block and z is a string of 0's.

There are $c_{j_i+1} - 1$ distinct legal blocks that have length $j_i + 1$ and do not end in a zero. Let η be a legal block that does not end in a zero. As $f > j_L$, the only sequences $\eta \rightarrow z$ with strings of zeroes of length at least f are those with $|z| \geq f$. Let $N(r)$ be the number of length r sequences $\eta \rightarrow z$ that contain no string of zeroes of length $\geq f - 1$. Since $|\eta| + |z| = r$, we see that $N(r)$ is given by the generating function

$$\sum_{r=1}^{\infty} N(r) s^r = ((c_1 - 1) s^{j_1+1} + \dots + (c_{j_L+1} - 1) s^{j_L+1}) (1 + s + \dots + s^{f-1}). \quad (4.5)$$

For any $i \in \mathbb{N}$ such that $2 \leq i \leq L$ there is exactly one legal block η that has length $j_i + 1$ and ends in a zero. There are no other legal blocks that end in a zero. Since the last non-zero term in η is then $\eta_{j_{i-1}}$,

the legal block contains a string of $g_{i-1} = j_i - j_{i-1}$ zeroes at the end. Let $M(r)$ be the number of length r sequences of legal blocks ending with a zero, followed by a string of zeroes, with no string of zeroes of length at least f ; we denote this by $\eta \rightarrow z$. As $f > j_L$, the longest string of zeroes of such a block has length $g_i + |z|$. So $\eta \rightarrow z$ contains no strings of zeroes of length at least f if $|z| < f - g_i - 1$. As $|\eta| + |z| = r$, $M(r)$ is given by the generating function

$$\sum_{r=1}^{\infty} M(r)s^r = s^{j_2+1} \left(\frac{1 - s^{f-g_1}}{1-s} \right) + \dots + s^{j_L+1} \left(\frac{1 - s^{f-g_L-1}}{1-s} \right). \quad (4.6)$$

Finally, there is exactly one terminal block of length r for each $r \geq 0$ and $r < j_L$. Thus the number $D(r)$ of length r terminal blocks has the generating function $\sum_{r=1}^{\infty} D(r)s^r = \frac{1-s^{j_L}}{1-s}$.

We now use these generating functions to find the number of legal sequences of length n with k legal blocks and all strings of zeroes of length less than f . Our decomposition based on the number of summands is similar to the analysis done in [KKMW, MW1, MW2]; this is a natural way to split into cases, and provides a manageable route through the combinatorics. That is we fix k and count the number of legal sequences $(\rightarrow_{i=1}^k (\eta_i \rightarrow z_i)) \rightarrow T$ that do not contain a subsequence of zeroes of length at least f ; recall the η_i are legal blocks, z_i are strings of zeroes, and T is terminal. Since the lengths of these separate components must sum to n , the number of such length n sequences is the coefficient of s^n in

$$\left(\sum_{r=1}^{\infty} N(r)s^r + \sum_{r=1}^{\infty} M(r)s^r \right)^k \sum_{r=1}^{\infty} D(r)s^r. \quad (4.7)$$

To find $F(s, f)$, it remains only to sum the above expression over all k . Thus the generating function of the number of length n legal sequences with longest gap $< f$ is

$$\begin{aligned} F(s, f) = & \frac{1-s^{j_L}}{1-s} \sum_{k \geq 0} \left[((c_1 - 1)s^{j_1+1} + \dots + (c_{j_L+1} - 1)s^{j_L+1}) \left(\frac{1-s^f}{1-s} \right) \right. \\ & \left. + s^{j_2+1} \left(\frac{1-s^{f-g_1}}{1-s} \right) + \dots + s^{j_L+1} \left(\frac{1-s^{f-g_L-1}}{1-s} \right) \right]^k. \end{aligned} \quad (4.8)$$

This is a geometric series, so we can evaluate our sum over k and then use the relation $j_{i-1} + g_{i-1} = j_i$ to calculate the desired result. \square

We have found a rational generating function for the cumulative distribution. To analyze it further, we first recall a standard lemma on partial fraction expansion.

Lemma 4.2 (Partial Fraction Expansion). *Let $R(s) = S(s)/T(s)$ be a rational function for $S, T \in \mathbb{C}[x]$ with $\deg(S) < \deg(T)$, and assume T has no multiple roots. Then the coefficient of s^n in $R(s)$'s Taylor expansion around zero is*

$$- \sum_{i=1}^{\deg(T)} \frac{S(\alpha_{i,f})}{\alpha_{i,f} T'(\alpha_{i,f})} \left(\frac{1}{\alpha_{i,f}} \right)^n, \quad (4.9)$$

where $\{\alpha_{i,f}\}$ are the roots of T .

Notice that in order to use this partial fraction expansion lemma, we need to ensure that the denominator of our generating function $F(s, f)$ has no multiple roots. To achieve this, we impose some extra restrictions on our recurrence relation, and obtain the following.

Lemma 4.3. *Let $T_f(s) = \mathcal{M}(s) + s^f \mathcal{R}(s)$, where $\mathcal{M}(s)$ and $\mathcal{R}(s)$ have no multiple roots, and no roots of absolute value 1. Then there exists $\epsilon > 0$ and $F \in \mathbb{N}$ such that for all $f \geq F$ and all roots α of $T_f(s)$ we have $|T'_f(\alpha)| > \epsilon$.*

The proof is standard, and follows from an analysis of the roots of the polynomial $T_f(s)$ as f varies; see Appendix A.2 of [BILMT] for details. Essentially, the behavior of $T_f(s)$ is as we may expect; the roots of $T_f(s) = \mathcal{M}(s) + s^f \mathcal{R}(s)$ with absolute value less than one are close to the roots of $\mathcal{M}(s)$ for large f , and the roots of $T_f(s)$ with absolute value greater than one are close to $\mathcal{R}(s)$ for large f . We also see that the large number of roots of absolute value close to one will have little contribution.

Applying partial fractions, we immediately obtain the following expression for the cumulative distribution function, $P(n, f)$.

Lemma 4.4. *Let $\{\alpha_{i,f}\}_{i=1}^{f+j_L}$ be the roots of $T_f(s)$. The cumulative distribution for the longest gap $P(n, f)$, the probability an $m \in [G_n, G_{n+1})$ has its longest gap in its Zeckendorf decomposition less than f , is*

$$P(n, f) = \frac{1}{G_{n+1} - G_n} \sum_{i=1}^{f+j_L} \frac{1 - \alpha_{i,f}^{j_L}}{\alpha_{i,f} T'_f(\alpha_{i,f})} \left(\frac{1}{\alpha_{i,f}} \right)^n. \quad (4.10)$$

4.3. Asymptotic Expansion for the CDF of the Longest Gap. We need several facts about the roots of the polynomials $T_f(s)$ to use Lemma 4.4. First, from the definition of $\mathcal{M}(s)$, it is immediate that $\mathcal{M}(s)$'s roots are exactly the inverse roots of the characteristic polynomial of the recurrence relation for the G_i . We label the roots of this characteristic polynomial $\{\lambda_i\}_{i=1}^{j_L+1}$. From Binet's formula (see Lemma A.1), $|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_{j_L+1}|$. Furthermore, we know that $\lambda_1 \in \mathbb{R}$ and $\lambda_1 > 1$. In particular, this shows that $\mathcal{M}(s)$ has a single smallest root $1/\lambda_1$, which is real-valued and has absolute value less than 1. In turn, since $T_f(s) = \mathcal{M}(s) + s^f \mathcal{R}(s)$, for large f , $T_f(s)$ has a smallest root that converges to $1/\lambda_1$.

Proposition 4.5. *There exists $F \in \mathbb{N}$ and $R_{\max}, R_{\min} \in \mathbb{R}$ satisfying $1/\lambda_1 < R_{\min} < \min(1, |1/\lambda_2|)$ such that for all $f \geq F$ every root $\alpha_{i,f}$ of $\mathcal{M}(s) + s^f \mathcal{R}(s)$ has $|\alpha_{i,f}| < R_{\max}$, and such that the polynomial $\mathcal{M}(s) + s^f \mathcal{R}(s)$ has exactly one root $\alpha_{1,f}$ with $|\alpha_{1,f}| < R_{\min}$. Furthermore*

$$\alpha_{1,f} = \frac{1}{\lambda_1} + \frac{\mathcal{M}(\alpha_{1,f})}{\mathcal{G}(\alpha_{1,f})} \alpha_{1,f}^f, \quad (4.11)$$

where λ_1 is the largest eigenvalue of the recurrence relation for G_i and $\mathcal{G}(s) := -\mathcal{M}(s)/(s - 1/\lambda_1)$, a polynomial. Moreover, there exists $\delta > 0$ such that $|\mathcal{G}(\alpha_{1,f})| > \delta$ for $f \geq F$.

The proof of Proposition 4.5 is standard and follows by Rouché's theorem; see Appendix A.1 of [BILMT].

The roots $\{\alpha_{i,f}\}_{i=1}^f$ appear in the terms of the sum in Lemma 4.4 as $\alpha_{i,f}^{-n}$, and the smallest root $\alpha_{1,f}$ dominates the sum. Being careful to deal with coefficients, and approximating $\alpha_{1,f}$ by $1/\lambda_1$, we obtain our claimed asymptotic expression for the cumulative distribution function of the longest gap.

Proof of Theorem 1.5(1). From Lemma 4.4, we have

$$P(n, f) = \frac{1}{G_{n+1} - G_n} \sum_{i=1}^{f+j_L} \frac{1 - \alpha_{i,f}^{j_L}}{\alpha_{i,f} T'_f(\alpha_{i,f})} \left(\frac{1}{\alpha_{i,f}} \right)^n. \quad (4.12)$$

By definition, we have that $1/\lambda_1 < R_{\min} < |1/\lambda_2|$. Therefore, by Binet formula (Lemma A.1)

$$G_{n+1} - G_n = C' \lambda_1^n + O((1/R_{\min})^n) \quad (4.13)$$

for some $C' \in \mathbb{R}$ ($C' = (\lambda_1 - 1)a$). Further, for any root $\alpha_{i,f} \neq \alpha_{1,f}$ we have $|\alpha_{i,f}| > R_{\min}$. Also, by Lemma 4.3 there is a bound $B \in \mathbb{R}$ such that $|1/T'_f(\alpha_{i,f})| < B$ for all roots $\alpha_{i,f}$ and for all $f \geq F$.

We see that for $\alpha_{1,f}$, the critical root from before, that

$$P(n, f) = -\frac{(1 - \alpha_{1,f}^{j_L})}{C' \alpha_{1,f} T'_f(\alpha_{1,f})} \left(\frac{1}{\lambda_1 \alpha_{1,f}} \right)^n + O(f (\lambda_1 R_{\min})^{-n}); \quad (4.14)$$

note $\lambda_1 R_{\min} > 1$. Next we use the relation

$$\alpha_{1,f} = \frac{1}{\lambda_1} + \frac{\alpha_{1,f}^f \mathcal{R}(\alpha_{1,f})}{\mathcal{G}(\alpha_{1,f})} \quad (4.15)$$

from Proposition 4.5 to express our formula in terms of λ_1 . Accordingly let $C = \lambda_1/C'$. Since $\alpha_{1,f} < R_{\min} < 1$ and $\mathcal{G}(\alpha_{1,f})$ is bounded away from zero, this shows that $\alpha_{1,f}$ converges to $1/\lambda_1$ exponentially fast. Substituting (4.15) three times and recalling $T_f(s) = \mathcal{M}(s) + s^f \mathcal{R}(s)$ gives

$$P(n, f) = \frac{-C(1 - \lambda_1^{j_L}) + O(\alpha_{1,f}^f)}{\left[\frac{1}{\lambda_1} \mathcal{M}'\left(\frac{1}{\lambda_1}\right) + O(\alpha_{1,f}^f)\right] + O(f \alpha_{1,f}^f)} \left(1 + \lambda_1 \frac{\alpha_{1,f}^f \mathcal{R}(\alpha_{1,f})}{\mathcal{G}(\alpha_{1,f})}\right)^{-n} + O(f(\lambda_1 R_{\min})^{-n}). \quad (4.16)$$

After a lot of algebra (using exponentiation, multiple Taylor expansions and $\alpha_{1,f} < R_{\min} < 1$) we find

$$P(n, f) = -\frac{C(1 - \lambda_1^{j_L})}{\frac{1}{\lambda_1} \mathcal{M}'\left(\frac{1}{\lambda_1}\right)} \exp\left(-\frac{n}{\lambda_1^{f-1}} \frac{\mathcal{R}\left(\frac{1}{\lambda_1}\right)}{\mathcal{G}\left(\frac{1}{\lambda_1}\right)}\right) + O\left(nf \left(\frac{R_{\min}}{\lambda_1}\right)^f + \frac{n}{\lambda_1^{2f}} + \frac{f}{(\lambda_1 R_{\min})^n}\right). \quad (4.17)$$

Further, since we always have $P(n, n+1) = 1$, substituting $f = n$ into (4.17) gives

$$\lim_{n \rightarrow \infty} P(n, n+1) = -\frac{C\lambda_1(1 - \lambda_1^{j_L})}{\mathcal{M}'\left(\frac{1}{\lambda_1}\right)}. \quad (4.18)$$

It follows that $-C\lambda_1(1 - \lambda_1^{j_L})/\mathcal{M}'\left(\frac{1}{\lambda_1}\right) = 1$, completing the proof of Theorem 1.5(1). \square

4.4. Mean and Variance of the Longest Gap. We use our asymptotic expression for the cumulative distribution function to calculate statistics of the longest gap distribution. *Remember that our cumulative distribution is defined for the longest gap being less than a given value.* Thus in the analysis below it is a little easier to first find, not the mean and variance of the random variable X denoting the longest gap, but the mean and the variance of the random variable Y which is one more than the longest gap. These are

$$\mu_{n,Y} = \sum_{g=1}^n g(P(n, g) - P(n, g-1)); \quad \sigma_n^2 = \sum_{g=1}^n g^2(P(n, g) - P(n, g-1)) - \mu_{n,Y}^2. \quad (4.19)$$

Thus our desired mean (for the longest gap) is $\mu_n = \mu_{n,Y} - 1$, and the variance is σ_n^2 .

As our asymptotic expression for $P(n, g)$ is only accurate for values of g on the order of $\log n$ or larger, we replace the sums in (4.19) from 1 to n by sums from ℓ_n to h_n , for suitable choices of ℓ_n and h_n , so that the error from restricting the summation is negligible. This is possible due to the very tight double exponential behavior, which we proved in Theorem 1.5(1). In particular, we have the following proposition.

Proposition 4.6. *Choosing $c, C \in \mathbb{R}$ such that $0 < c < 1/\lambda_1$ and $C > \max(6, 4 \log \lambda_1)$, we let $\ell_n = \lfloor c \log(nK) \rfloor$ and $h_n = \lfloor C \log(nK) \rfloor$ (remember $\lambda_1 > 1$ and K is as in Theorem 1.5(2)). We find that*

$$\begin{aligned} \mu_{n,Y} &= \sum_{g=\ell_n}^{h_n} g(P(n, g) - P(n, g-1)) + o(1) \\ \sigma_n^2 &= -\left(\mu_{n,Y}^2 - \sum_{g=\ell_n}^{h_n} g^2(P(n, g) - P(n, g-1))\right) + o(1). \end{aligned} \quad (4.20)$$

With these values of h_n and ℓ_n , to prove the above proposition only requires the crudest bounds. The analysis is standard, and can be found in Appendix B of [BILMT].

Proof of Theorem 1.5(2). We work simultaneously with $\mu_{n;Y}$ and σ_n^2 . In preparation for approximating our sums with integrals, we first sum by parts so that

$$\begin{aligned}\mu_{n;Y} &= (h_n + 1)P(n, h_n) - \ell_n P(n, \ell_n - 1) - \sum_{g=\ell_n}^{h_n} P(n, g) + o(1) \\ \sigma_n^2 &= - \left(\mu_{n;Y}^2 - (h_n + 1)^2 P(n, h_n) + \ell_n^2 P(n, \ell_n - 1) + \sum_{g=\ell_n}^{h_n} (2g + 1)P(n, g) \right) + o(1).\end{aligned}\quad (4.21)$$

From Theorem 1.5(1), we know that $\ell_n^2 P(n, \ell_n) \rightarrow 0$ and $h_n^2 P(n, h_n) \rightarrow h_n^2$ for large n , and hence

$$\mu_{n;Y} = (h_n + 1) - S_1 + o(1), \quad \sigma_n^2 = -(\mu_{n;Y}^2 - (h_n + 1)^2 + 2S_2 + S_1) + o(1) \quad (4.22)$$

for

$$S_i := \sum_{g=\ell_n}^{h_n} g^{i-1} P(n, g). \quad (4.23)$$

With $K = \lambda_1 \mathcal{R}(1/\lambda_1)/\mathcal{G}(1/\lambda_1)$, our estimates from Theorem 1.5 give us

$$S_i = \sum_{g=\ell_n}^{h_n} g^{i-1} \exp(-nK\lambda_1^{-g}) + O(n^{-\delta}(\log n)^i). \quad (4.24)$$

Now we apply the Euler-Maclaurin formula to S_1 and S_2 , and find

$$S_i = \int_{\ell_n}^{h_n} t^{i-1} \exp(-nK\lambda_1^{-t}) dt + \frac{1}{2} t^{i-1} P(n, t) \Big|_{t=\ell_n}^{h_n} + \text{Error}_{EM}^i + o(1). \quad (4.25)$$

Elementary analysis shows that $\text{Error}_{EM}^1 = o(1)$ and $\text{Error}_{EM}^2 = 1 + o(1)$, and thus the two errors above are negligible (for completeness this standard calculation is done in Appendix C of [BILMT]). The boundary terms approach $1/2$ and $h_n/2$, respectively, since $P(n, h_n) \rightarrow 1$ while $P(n, \ell_n) \rightarrow 0$ so fast that $\ell_n P(n, \ell_n) \rightarrow 0$. We are left with analyzing the two integrals.

Define $w(t) = \exp(-t \log \lambda_1 + \log(nK))$, with $w'(t) = -w(t) \log \lambda_1$ and $t = \frac{\log(nK) - \log w}{\log \lambda_1}$. Writing I_i for the integrals in (4.25), integrating by parts yields and letting $u = w(t)$ gives

$$\begin{aligned}I_1 &= t e^{-w(t)} \Big|_{\ell_n}^{h_n} + \int_{w(\ell_n)}^{w(h_n)} \frac{\log(nK) - \log u}{\log \lambda_1} e^{-u} du, \\ I_2 &= \frac{1}{2} \left(t^2 e^{-w(t)} \Big|_{\ell_n}^{h_n} + \int_{w(\ell_n)}^{w(h_n)} \left(\frac{\log(nK) - \log u}{\log \lambda_1} \right)^2 e^{-u} du \right).\end{aligned}\quad (4.26)$$

We expand the integrals and note that $w(h_n) = 0 + o(1)$ and $w(\ell_n)$ is positive and tends to infinity with n . Then, using the well known identities (see 4.331.1 and 4.335.1 of [GrRy])

$$\int_0^\infty \log(u) e^{-u} du = -\gamma, \quad \int_0^\infty (\log u)^2 e^{-u} du = \gamma^2 + \frac{\pi^2}{6} \quad (4.27)$$

with γ the Euler-Mascheroni constant (note on page xxxii of [GrRy] they set $C = \gamma$), we obtain

$$\begin{aligned}I_1 &= t e^{-w(t)} \Big|_{\ell_n}^{h_n} - \frac{\log(nK) + \gamma}{\log \lambda_1} + o(1) \\ I_2 &= \frac{1}{2} \left(t^2 e^{-w(t)} \Big|_{\ell_n}^{h_n} - \frac{1}{(\log \lambda_1)^2} \left(\log(nK)^2 + 2\gamma \log(nK) + \gamma^2 + \frac{\pi^2}{6} \right) \right) + o(1).\end{aligned}\quad (4.28)$$

Our claimed values for the mean and variance now follow by evaluating the above and substituting. \square

Remark 4.7. We took $h_n = \lfloor C \log(nK) \rfloor$ with $C > \max(6, 4 \log \lambda_1)$. This constant can be replaced with any sufficiently large value; however, we need h_n to be at least this large to facilitate the error analysis arising from truncating the sums.

5. CONCLUDING REMARKS

Building on the combinatorial vantage introduced in [KKMW] and its sequels, we are able to determine the limiting behavior for the distribution of gaps in the bulk, both on average and almost surely for the individual gap measures, as well as mean and variance of the longest gap. A natural future project is to remove some of the assumptions we have made on the recurrence relation. We expect the answers in these cases to be essentially the same, but the resulting algebra will be more involved.

An additional line of investigation is to apply these methods to other decompositions, for example the f -decompositions introduced in [DDKMMV].

Definition 5.1. Given a function $f : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ and a sequence of integers $\{a_n\}$, a sum $m = \sum_{i=0}^k a_{n_i}$ of terms of $\{a_n\}$ is an f -decomposition of m using $\{a_n\}$ if for every a_{n_i} in the f -decomposition, the previous $f(n_i)$ terms $(a_{n_i-f(n_i)}, a_{n_i-f(n_i)+1}, \dots, a_{n_i-1})$ are not in the f -decomposition.

To see that this generalizes the standard Zeckendorf decomposition, take a_n to be the n^{th} Fibonacci number and $f(n) = 1$ for all n . The authors prove for any $f : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ there exists a unique sequence of natural numbers $\{a_n\}$ such that every positive integer has a unique legal f -decomposition in $\{a_n\}$. Interestingly, certain choices of f lead to sequences defined by a recurrence relation with *negative* coefficients in a fundamental way, where there is no equivalent definition using only non-negative coefficients (the Fibonacci can be defined by $F_{n+1} = 2F_n - F_{n-2}$, but they are also given by the standard relation $F_{n+1} = F_n + F_{n-1}$). One example is their b -bin decompositions. We break the natural numbers into bins of length b , and say a decomposition is legal if we never choose two elements from the same bin, nor two adjacent elements from two consecutive bins. This leads to a periodic formula for the associated f . For example, if $b = 3$ our sequence of a_n 's starts 1, 2, 3, 4, 7, 11, 15, 26, 41, 56, 97, 153, and satisfies the recurrence $a_n = 4a_{n-3} - a_{n-6}$, while if $b = 2$ we recover the standard Zeckendorf decomposition involving Fibonacci numbers.

APPENDIX A. GENERALIZED BINET'S FORMULA

This standard generalization of Binet's formula follows from the Perron-Frobenius Theorem for irreducible matrices (though it can be proved directly, which is done in Appendix A of the arXiv version of this paper, [BBGILMT]).

Lemma A.1 (Generalized Binet's Formula). *Consider the positive linear recurrence $G_{n+1} = c_1 G_n + c_2 G_{n-1} + \dots + c_L G_{n+1-L}$ with the c_i 's non-negative integers and $c_1, c_L > 0$. Let $\lambda_1, \dots, \lambda_L$ be the roots of the characteristic polynomial $f(x) := x^L - (c_1 x^{L-1} + c_2 x^{L-2} + \dots + c_{L-1} x + c_L) = 0$, ordered so that $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_L|$. Then $\lambda_1 > |\lambda_2| \geq \dots \geq |\lambda_L|$, $\lambda_1 > 1$ is the unique positive root, and there exist constants such that $G_n = a_1 \lambda_1^n + O(n^{L-2} \lambda_2^n)$. More precisely, if $\lambda_1, \omega_2, \dots, \omega_r$ denote the distinct roots of the characteristic polynomial with multiplicities 1, m_2, \dots, m_r , then there are constants $a_1 > 0, a_{i,j}$ such that $G_n = a_1 \lambda_1^n + \sum_{i=2}^r \sum_{j=1}^{m_i} a_{i,j} n^{j-1} \omega_i^n$.*

APPENDIX B. PROOF OF PROPOSITION 3.4

We now finish the proof of Proposition 3.4, specifically showing that $\lim_{n \rightarrow \infty} \text{Var}_n(t) = 0$. In the calculation below g_1 and g_2 denote two arbitrary gaps that start at the two indices $j_1 \leq j_2$; thus $g_1, g_2 \in \{0, 1, \dots, n-1\}$ and $j_1, j_2 \in \{1, 2, \dots, n\}$. As the number of indices in the proof is growing, we write $\ell_r(m)$ and $\ell_w(m)$ for the summands in m 's decomposition, making explicit the m dependence. In the sum

that follows, we have to separately deal with the case $r = w$. We have

$$\begin{aligned}\mathbb{E}_m[\widehat{\nu_{m;n}}(t)^2] &= \frac{1}{G_{n+1} - G_n} \sum_{m=G_n}^{G_{n+1}-1} \frac{1}{(k(m) - 1)^2} \sum_{r=2}^{k(m)} e^{it(\ell_r(m) - \ell_{r-1}(m))} \sum_{w=2}^{k(m)} e^{it(\ell_w(m) - \ell_{w-1}(m))} \\ &= \frac{1}{(G_{n+1} - G_n)(C_{\text{Lek}}n + d)^2} \left(2 \sum_{\substack{j_1 < j_2 \\ g_1, g_2}} X_{j_1, j_1+g_1, j_2, j_2+g_2}(n) e^{itg_1} e^{itg_2} + \sum_{j_1, g_1} X_{j_1, j_1+g_1}(n) e^{2itg_1} \right) + o(1),\end{aligned}\tag{B.1}$$

where the last line follows by using Remark 3.3 to replace $1/(k(m) - 1)^2$ with its average value up to a negligible error and then doing the same change of variables as before, and the factor of 2 is because we are taking $j_1 < j_2$. As the denominator is of order $n^2(G_{n+1} - G_n)$ while $\sum_{j_1, g_1} X_{j_1, j_1+g_1}(n)$ is of order $n(G_{n+1} - G_n)$, the diagonal term does not contribute in the limit, and the factor of 2 vanishes when we sum over $j_1 < j_2$ (which gives $n^2/2 + O(n)$). Therefore

$$\begin{aligned}\mathbb{E}_m[\widehat{\nu_{m;n}}(t)^2] &= \frac{2}{(G_{n+1} - G_n)(C_{\text{Lek}}n + d)^2} \sum_{\substack{j_1 < j_2 \\ g_1, g_2}} X_{j_1, j_1+g_1, j_2, j_2+g_2}(n) e^{it(g_1+g_2)} + o(1) \\ &= \frac{2}{a_1 \lambda_1^n (\lambda_1 - 1)(C_{\text{Lek}}n + d)^2 (1 + o(1))} \left(o(1) + \sum_{j_1 < j_2} \sum_{g_1, g_2=0}^{n-1} X_{j_1, j_1+g_1, j_2, j_2+g_2}(n) e^{it(g_1+g_2)} \right).\end{aligned}\tag{B.2}$$

There are several different cases to consider for the pair (g_1, g_2) : at least one of them could be 0, at least one of them could be 1, or both exceed 1. The argument is essentially the same in each case; the only difference comes from slight changes in how we count $X_{j_1, j_1+g_1, j_2, j_2+g_2}(n)$. Note that if we restricted ourselves to the Fibonacci numbers the first two cases cannot happen (if we consider only recurrences where all the coefficients are 0 or 1 then the first case cannot happen).

We first consider the case when $g_1 = g_2 = 1$. We chose to do this case in detail as it has some of the counting obstructions, and gives the general flavor. We determine $X_{j_1, j_1+1, j_2, j_2+1}(n)$ by counting the total number of decompositions in $[G_n, G_{n+1})$ which have a gap of length 1 from G_{j_1} to G_{j_1+1} (which we know how to do by Lemma 2.2) and then subtract the three different ways decompositions can have a gap of length 1 from G_{j_1} to G_{j_1+1} *without* having a gap of length 1 at G_{j_2} to G_{j_2+1} : (1) include G_{j_1}, G_{j_1+1} and G_{j_2+1} but do not include G_{j_2} ; (2) include G_{j_1}, G_{j_1+1} but do not include G_{j_2} and G_{j_2+1} ; and (3) include $G_{j_1}, G_{j_1+1}, G_{j_2}$ and G_n but do not include G_{j_2+1} . These three cases can be counted by Lemma 2.1 and similar counting techniques.

Note it is sufficient to analyze these cases under the additional assumption that j_2 is at least $2L$ units from j_1 (where L is the length of the recurrence). The reason is that the denominator has a factor of n^2 ; if j_2 is within a bounded distance of j_1 we only get an n in the numerator, and the contribution is negligible.

There is one last technicality. If any of $j_1, j_2, j_2 - j_1, n - j_1$ or $n - j_2$ is small then expanding a G_γ (where γ is one of these troubling indices) by the generalized Binet formula will not yield an error of size $o(1)$. This is the same issue we had in the proof of Theorem 1.3, and is handled similarly. We introduce the notation $o_{j_1, j_2; n}(1)$, which is $o(1)$ if all of the combinations above are at least $\log^2 n$ away from 0, and bounded otherwise. Again the sum of this over all j_1, j_2 will be lower order. We therefore assume $j_2 \geq j_1 + 2L$. Because of the length of the lines, for formatting reasons we put the error term with the sum over all $j_1 < j_2$

and not over the restricted sums. We find

$$\begin{aligned}
& \sum_{j_1=1}^{n-1} \sum_{j_2=j_1+1}^{n-1} X_{j_1, j_1+1, j_2, j_2+1}(n) + o(n(G_{n+1} - G_n)) \\
&= \sum_{j_1=1}^{n-2L} \sum_{j_2=j_1+2L}^n \left[(G_{n+1} - G_n) - G_{j_1+1}(G_{n-j_1} - G_{n-j_1-1}) - G_{j_1}(G_{n-j_1+1} - 2G_{n-j_1} + G_{n-j_1-1}) \right. \\
&\quad - (G_{n-j+1} - 2G_{n-j} + G_{n-j-1})(G_{j_2} - G_{j_1+1}G_{j_2-j_1-1} - G_{j_1}(G_{j_2-j_1} - G_{j_2-j_1-1})) \\
&\quad - (G_{n-j} - G_{n-j-1})(G_{j_2} - G_{j_1+1}G_{j_2-j_1-1} - G_{j_1}(G_{j_2-j_1} - G_{j_2-j_1-1})) \\
&\quad - (G_{n-j} - G_{n-j-1})((G_{j+1} - G_j) - G_{j_1+1}(G_{j_2-j_1} - G_{j_2-j_1-1})) \\
&\quad \left. - G_{j_1}(G_{j_2-j_1+1} - 2G_{j_2-j_1} + G_{j_2-j_1-1}) \right] \\
&= \sum_{j_1=1}^{n-2L} \sum_{j_2=j_1+2L}^n \left[(a_1 \lambda_1^n (\lambda_1 - 1)(1 - a_1 - a_1 \lambda_1^{-1}(\lambda_1 - 1)(1 + o(1))) \right. \\
&\quad - a_1^2 \lambda_1^{n-1} (\lambda_1 - 1)^2 (1 - a_1 - a_1 \lambda_1^{-1}(\lambda_1 - 1)(1 + o_{j_1, j_2; n}(1))) \\
&\quad - a_1^2 \lambda_1^{n-1} (\lambda_1 - 1)(1 - a_1 - a_1 \lambda_1^{-1}(\lambda_1 - 1)(1 + o_{j_1, j_2; n}(1))) \\
&\quad \left. - a_1^2 \lambda_1^{n-1} (\lambda_1 - 1)^2 (1 - a_1 - a_1 \lambda_1^{-1}(\lambda_1 - 1))(1 + o_{j_1, j_2; n}(1)) \right] \\
&= (1 - a_1 - a_1 \lambda_1^{-1}(\lambda_1 - 1))(a_1 \lambda_1^{n-1} (\lambda_1 - 1)(\lambda_1 - a_1(\lambda_1 - 1) - a_1 - a_1(\lambda_1 - 1)) \\
&\quad \cdot (1 + o(1)) \sum_{j_1=1}^{n-2L} \sum_{j_2=j_1+2L}^n 1 \\
&= \left(\frac{n^2 + O(n)}{2} \right) a_1 \lambda_1^n (\lambda_1 - 1)(1 + o(1))((\lambda_1(1 - 2a_1) + a_1) \lambda_1^{-1})^2. \tag{B.3}
\end{aligned}$$

Notice that as $n \rightarrow \infty$, (B.3) times the coefficient in (B.2) is, up to an error of size $o(1)$,

$$\left(\frac{1}{C_{\text{Lek}}} \lambda_1^{-1} (\lambda_1(1 - 2a_1) + a_1) \right)^2 = P(1)^2,$$

which cancels with corresponding piece in $\widehat{\nu}(t)^2$ in the difference $\mathbb{E}_m[\widehat{\nu}_{m; n}(t)^2] - \widehat{\nu}(t)$.

The other cases for (g_1, g_2) can be handled similarly, and again we find that the contribution equals the corresponding terms from $\widehat{\nu}(t)^2$ in the difference $\mathbb{E}_m[\widehat{\nu}_{m; n}(t)^2] - \widehat{\nu}(t)$. The only complication is we need our error terms to be small enough so that we may sum over all pairs (g_1, g_2) . This is not a problem as our approach allows us to isolate the error term, which is small when summed over all pairs as the sum of $P_n(g)$ is bounded. Therefore, $\lim_{n \rightarrow \infty} \text{Var}_n(t) = 0$, completing the proof.

APPENDIX C. EXTENSION TO INITIAL SEGMENTS (IDDO BEN-ARI AND STEVEN J. MILLER)

Our theorems for intervals $[G_n, G_{n+1})$ generalize to initial segments $[1, N)$. This appendix was inspired by questions of one of the referees on such results, and has influenced the sequel project [B-AM]. There the authors of this appendix adopt a Markov perspective, which simplifies a lot of the technical issues but at the cost of additional machinery. As our purpose here is to just state what is known and quickly highlight the arguments for the extensions, we drop the combinatorial approach (which is very natural for intervals of the form $[G_n, G_{n+1})$, and with work could be made to work here as well; see for example [BEDMMTW1,

BEDMMTW2] for such investigations of sub-intervals of $[G_n, G_{n+1})$ and instead introduce the perspective of [B-AM].

If $N = G_n$ for some n , then the asymptotic results are easy to recover. This follows immediately as the intervals $[G_n, G_{n+1})$ are of exponentially growing length and the dependence on n in all our statistics is at most linear. Thus we can safely combine the results from a small number of intervals (say on the order of $\log n$ or even $\log \log n$) where there is essentially no variation and which asymptotically covers almost all integers in the initial segment $[1, N)$. Similarly we can also immediately do initial segments of the form $[1, N_j)$ with $N_j \rightarrow \infty$ and $N_j/G_{n_j} \rightarrow 1$ as $j \rightarrow \infty$.

The remaining case can be treated by the following class of initial segments. Assume that $\epsilon > 0$ and let $\{N_j\}_{j=2}^\infty$ be an increasing sequence of real numbers, with $N_j \in [(1+\epsilon)G_{j-1}, (1-\epsilon)G_j]$. The uniform measure on the initial segment $[1, N_j)$ is a convex combination of the uniform measures on the intervals $[G_1, G_2), \dots, [G_{j-2}, G_{j-1})$, and the uniform measure on the last subinterval $A_j = \{n \in [G_{j-1}, G_j) : n < N_j\}$. This last measure is the uniform measure on $[G_{j-1}, G_j)$, conditioned on A_j . The respective convex coefficients are proportional to the number of elements in each interval, which, in the case of the last one is the number of elements in the subinterval A_j .

The main problem one needs to address is that the last subinterval, contrary to all others, includes only a portion of the numbers with representation of the corresponding length j , and because of the exponential growth of the intervals and the assumption $N_j \geq (1+\epsilon)G_{j-1}$, this last subinterval is comparable in size to the size of the union of all preceding intervals, so that the associated convex coefficient is not vanishing asymptotically. We just can't ignore this last subinterval, and it's different from all others. Now the uniform measure on A_j is obtained from the uniform measure Q^j on $[G_{j-1}, G_j)$, through conditioning: it is the conditioned measure $Q^j(\cdot | A_j)$. For most statistics appearing in the literature, including the ones studied in this paper, this conditioning has asymptotically vanishing effect, as they are asymptotically independent of A_j under Q^j . To explain this, observe first that determining whether a number in $[G_{j-1}, G_j)$ belongs to the subinterval A_j is essentially determined by a uniformly bounded number (in j) of first digits in its generalized Zeckendorf decomposition, and this is because of the exponential growth of the G_n 's and the fact that $N_j \geq (1+\epsilon)G_{j+1}$. (We write "essentially" because this rule does not apply to all numbers, but rather a proportion tending to 1, and this is good enough.)

In [B-AM] the authors provided a Markov chain interpretation and representation of generalized Zeckendorf decompositions. There the uniform measures on all intervals $[G_{j-1}, G_j)$ are obtained from a single ergodic Markov chain. The key is ergodicity of the chain, which guarantees that all asymptotic statistics are stochastically independent of the first digits. As a rule of thumb, all "scalable" quantities can be treated through this procedure, and this includes the extensively studied results including moments, laws of large numbers and central limit theorems on the number of summands or more generally additive functionals of the Markov chain, as well as law of largest gap, and tails of the gap measures.

REFERENCES

- [AI] 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.
- [B-AM] I. Ben-Ari and S. J. Miller, *A probabilistic approach to generalized Zeckendorf decompositions*, preprint 2014. <http://arxiv-web3.library.cornell.edu/pdf/1405.2379v1>.
- [BEDMMTW1] A. Best, X. Edelsbrunner, P. Dynes, B. McDonald, S. J. Miller and M. Weinstein, *Gaussian Behavior of the Number of Summands in Zeckendorf Decompositions in Small Intervals*, to appear in the Fibonacci Quarterly.
- [BEDMMTW2] A. Best, X. Edelsbrunner, P. Dynes, B. McDonald, S. J. Miller and M. Weinstein, *Gaussian Distribution of the Number of Summands in Generalized Zeckendorf Decompositions in Small Intervals*, joint with Andrew Best, Patrick Dynes, Xixi Edelsbrunner, Brian McDonald, Kimsy Tor, Caroline Turnage-Butterbaugh and Madeleine Weinstein, submitted January 2015 to Integers (reformatted and sent on March 1st).
- [BILMT] A. Bower, R. Insoft, S. Li, S. J. Miller and P. Tosteson, *The distribution of gaps between summands in generalized Zeckendorf decompositions*, arXiv 2014, <http://arxiv.org/pdf/1402.3912v2>.

- [CFHMN] M. Catral, P. Ford, P. Harris, S. J. Miller and D. Nelson, *Generalizing Zeckendorf's Theorem: The Kentucky Sequence*, to appear in the Fibonacci Quarterly. <http://arxiv.org/pdf/1409.0488v1>.
- [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*, preprint 2014. <http://arxiv.org/abs/1309.5599>.
- [DDKMV] P. Demontigny, T. Do, A. Kulkarni, S. J. Miller and U. Varma, *A generalization of Fibonacci far-difference representations and Gaussian behavior*, to appear in the Fibonacci Quarterly. <http://arxiv.org/abs/1309.5600>.
- [DrGa] M. Drmota and J. Gajdosik, *The distribution of the sum-of-digits function*, J. Théor. Nombres Bordeaux **10** (1998), no. 1, 17–32.
- [DuTh] J.-M. Dumont and A. Thomas, *Systemes de numeration et fonctions fractales relatifs aux substitutions*, J. Theoret. Comput. Sci. **65** (1989), 153–169.
- [FGNPT] P. Filippini, P. J. Grabner, I. Nemes, A. Pethő and R. F. Tichy, *Corrigendum to: "Generalized Zeckendorf expansions"*, Appl. Math. Lett., **7** (1994), no. 6, 25–26.
- [FG] B. E. Fristedt and L. F. Gray, *A Modern Approach to Probability Theory*, Birkhäuser, Boston, 1996.
- [GrTi] 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. Pethő, *Generalized Zeckendorf expansions*, Appl. Math. Lett. **7** (1994), no. 2, 25–28.
- [GrRy] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (seventh edition), Academic Press, San Diego, CA, 2007.
- [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.
- [KKMW] M. Koloğlu, G. Kopp, S. J. Miller and Y. Wang, *On the number of summands in Zeckendorf decompositions*, Fibonacci Quarterly **49** (2011), no. 2, 116–130.
- [LaTh] 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.
- [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.
- [MW2] S. J. Miller and Y. Wang, *Gaussian behavior in generalized Zeckendorf decompositions*, to appear in the conference proceedings of the 2011 Combinatorial and Additive Number Theory Conference. <http://arxiv.org/abs/1107.2718>.
- [Sch] M. Schilling, *The longest run of heads*, College Math. J. **21** (1990), no. 3, 196–207.
- [Ste] W. Steiner, *Parry expansions of polynomial sequences*, Integers **2** (2002), Paper A14.
- [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.

E-mail address: amandarg@umd.umich.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF MICHIGAN-DEARBORN, DEARBORN, MI 48128

E-mail address: rinsoft@wellesley.edu

DEPARTMENT OF MATHEMATICS, WELLESLEY COLLEGE, WELLESLEY, MA 02481

E-mail address: jjl2357@berkeley.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, BERKELEY, BERKELEY, CA 94720

E-mail address: sjm1@williams.edu, Steven.Miller.MC.96@aya.yale.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267

E-mail address: ptoste@umich.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109