# INDIVIDUAL GAP MEASURES FROM GENERALIZED ZECKENDORF DECOMPOSITIONS

ROBERT DORWARD, PARI L. FORD, EVA FOURAKIS, PAMELA E. HARRIS, STEVEN J. MILLER, EYVINDUR A. PALSSON, AND HANNAH PAUGH

ABSTRACT. Zeckendorf's theorem states that every positive integer can be uniquely decomposed as a sum of nonconsecutive Fibonacci numbers. The distribution of the number of summands converges to a Gaussian, and the individual measures on gaps between summands for $m \in [F_n, F_{n+1})$ converge to geometric decay for almost all $m$ as $n \to \infty$. While similar results are known for many other recurrences, previous work focused on proving Gaussianity for the number of summands or the average gap measure. We derive general conditions, which are easily checked, that yield geometric decay in the individual gap measures of generalized Zeckendorf decompositions attached to many linear recurrence relations.

## 1. INTRODUCTION

Zeckendorf [Ze] proved the remarkable property that every positive integer can be uniquely written as a sum of non-consecutive Fibonacci numbers $\{F_n\}_{n=1}^{\infty}$, where $F_1 = 1$, $F_2 = 2$ and $F_{n+1} = F_n + F_{n-1}$, and that this property is equivalent to this definition of the Fibonaccis (if we used the more common $F_0 = 0$, $F_1 = 1$ and $F_2 = 1$ then clearly such decompositions are not always unique). Zeckendorf's theorem has been generalized to other sequences, see among others [Al, Day, DDKMMV, DDKMV, DG, GT, GTNP, Ste1, Ste2]. Many authors proved that sequences $\{a_n\}$ defined by suitable linear recurrences lead to unique decompositions, with the number of summands of $m \in [a_n, a_{n+1})$ converging to a Gaussian (see for example [LT, MW]) and the average gap measure converging to geometric decay (see [BBGILMT, BILMT]). It is significantly easier to focus on the average gap measures rather than the individual gap measures associated to each $m$; in this note we isolate a general set of conditions which suffice to prove these individual measures converge almost surely to geometric decay.

We work in great generality so the arguments below will apply to numerous sequences. We assume we have a strictly increasing integer sequence $\{b_n\}$ and a decomposition rule that leads to unique decomposition. Fix constants $c_1, d_1, c_2, d_2$ such that $I_n := [b_{c_1 n + d_1}, b_{c_2 n + d_2})$ is a well-defined interval for all $n > 0$. Let $\delta(x - a)$ denote the Dirac delta functional (assigning a mass of 1 to $x = a$ and 0 otherwise), $k(z)$ be the number of summands in $z$'s decomposition

$(z = b_{\ell_1} + \cdots + b_{\ell_{k(z)}})$, and the total number of gaps for all $z \in I_n$ is

$$N_{\text{gaps}}(n) := \sum_{z=b_{c_1 n+d_1}}^{b_{c_2 n+d_2}-1} (k(z) - 1). \tag{1.1}$$

*We always assume the summands are ordered from least to greatest, so $\ell_1 \le \ell_2 \le \cdots \le \ell_{k(z)}$.*

- *Spacing gap measure:* We define the spacing gap measure of a $z \in I_n$ by

$$\nu_{z,n}(x) := \frac{1}{k(z) - 1} \sum_{j=2}^{k(z)} \delta(x - (\ell_j - \ell_{j-1})) \tag{1.2}$$

   if $k(z) > 0$ and $\delta(0)$ if $k(z) = 0$; note that the number of $z \in I_n$ with $k(z) = 0$ is

- *Average spacing gap measure:* The average spacing gap measure for all $z \in I_n$ is

$$\nu_n(x) := \frac{1}{N_{\text{gaps}}(n)} \sum_{z=b_{c_1 n+d_1}}^{b_{c_2 n+d_2}-1} \sum_{j=2}^{k(z)} \delta\left(x - (\ell_j - \ell_{j-1})\right) = \frac{1}{N_{\text{gaps}}(n)} \sum_{z=b_{c_1 n+d_1}}^{b_{c_2 n+d_2}-1} (k(z) - 1)\, \nu_{z,n}(x). \tag{1.3}$$

   Letting $P_n(g)$ denote the probability of a gap of length $g$ among all gaps from the decompositions of all $m \in I_n$, we have

$$\nu_n(x) = \sum_{g=0}^{c_2 n+d_2-1} P_n(g)\delta(x - g). \tag{1.4}$$

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

$$\nu(x) := \lim_{n\to\infty} \nu_n(x), \quad P(k) := \lim_{n\to\infty} P_n(k). \tag{1.5}$$

- *Indicator function for two gaps:* For $g_1, g_2 \ge 0$

$$X_{j_1,j_1+g_1,j_2,j_2+g_2}(n) := \#\left\{ z \in I_n : \begin{smallmatrix} b_{j_1}, b_{j_1+g_1}, b_{j_2}, b_{j_2+g_2} \text{ in } z\text{'s decomposition,} \\ \text{but not } b_{j_1+q}, b_{j_2+p} \text{ for } 0<q<g_1, 0<p<g_2 \end{smallmatrix} \right\}. \tag{1.6}$$

We generalize the work in [BILMT]. The authors there concentrated on a specific class of recurrences; our arguments are general and apply to any sequence where certain conditions can be verified. In addition to holding for the oft studied positive linear recurrences, they hold for new systems such as the $m$-gonal numbers of [DFFHMPP]. Our result below also holds for some sequences without unique decomposition (see [CFHMN]); in the cases we have studied to date there is a natural decomposition, frequently constructed from a greedy algorithm, which always exists and to which these results hold.

In the theorem and proofs below, we use big-Oh and $\ll$ notation interchangeably; thus $f(x) = O(g(x))$ and $f(x) \ll g(x)$ both mean there exists a constant $C$ and an $x_0$ such that for all $x \ge x_0$ we have $|f(x)| \le Cg(x)$. By $f(x) = o(1)$ we mean $\lim_{x\to\infty} f(x) = 0$. We give the proof when the number of summands converge to being normally distributed as that is what happens in all cases of interest, and remark on the proof in general. We also assume that very few numbers have exactly one summand in their decomposition (such a case causes a small book-keeping issue, as we cannot have a $k(z) - 1$ in the denominator in this case as that would vanish); in practice this is not a problem as in our cases of interest our sequence (and thus $|I_n|$) grows exponentially and the number of $z \in I_n$ with just one summand grows linearly.

**Theorem 1.1.** *For $z \in I_n$, the individual gap measures $\nu_{z,n}(x)$ converge almost surely in distribution to the average gap measure $\nu(x)$ if the following hold.*

(1) *The sequence $\{b_n\}$ is a strictly increasing sequence of integers, there is a decomposition rule such that every positive integer has a unique decomposition in terms of the $b_n$'s (with the summands indexed from smallest to largest), and $k(z) \ll n$ for all $z \in I_n$. Further, if $S_n$ is the set of $z \in I_n$ which have exactly one summand in their decomposition, then $|S_n|/|I_n| = o(1)$.*

(2) *The number of summands for decompositions of $z \in I_n$ converge pointwise to a Gaussian with mean $\mu_n = c_{\text{mean}}n + O(1)$ and variance $\sigma_n^2 = c_{\text{var}}n + O(1)$, for constants $c_{\text{mean}}, c_{\text{var}} > 0$ independent of $n$ (or, if we do not have pointwise convergence but still have convergence to normal distributions, there exists a $\delta \in (0, 1/2)$ such that the probability of being at least $n^\delta$ standard deviations from the mean decays faster than $1/n^{1+\epsilon}$ for some $\epsilon > 0$).*

(3) *We have the following, with $\lim_{n \to \infty} \sum_{g_1, g_2} \text{error}(n, g_1, g_2) = 0$:*

$$\frac{2}{|I_n|\mu_n^2} \sum_{j_1 < j_2} X_{j_1, j_1+g_1, j_2, j_2+g_2}(n) \;=\; P(g_1)P(g_2) + \text{error}(n, g_1, g_2). \tag{1.7}$$

(4) *The limits in Equation (1.5) exist.*

## 2. Proof of Theorem 1.1

We need the following definitions. We use the standard normalization that the characteristic function of a random variable $X$ with density $f_X$ is

$$\widehat{f_X}(t) \;:=\; \mathbb{E}[e^{itX}] \;=\; \int_{-\infty}^{\infty} e^{itx} f_X(x)dx. \tag{2.1}$$

- $\widehat{\nu_{z,n}}(t)$: The characteristic function of $\nu_{z,n}(x)$.
- $\widehat{\nu}(t)$: The characteristic function of the limiting average gap distribution $\nu(x)$.
- $\mathbb{E}_z[\ldots]$: The expected value over $z \in I_n$ with the uniform measure:

$$\mathbb{E}_z[X] \;:=\; \frac{1}{|I_n|} \sum_{z=b_{c_1 n+d_1}}^{b_{c_2 n+d_2}-1} X(z). \tag{2.2}$$

- *Indicator function for one gap:* For $g \geq 0$ let

$$X_{i,i+g}(n) \;=\; \#\{z \in I_n : b_i, b_{i+g} \text{ in } z\text{'s decomposition, but not } G_{i+q} \text{ for } 0 < q < g\}. \tag{2.3}$$

**Proposition 2.1.** *We have*

$$\lim_{n \to \infty} \mathbb{E}_z[\widehat{\nu_{z;n}}(t)] \;=\; \widehat{\nu}(t). \tag{2.4}$$

First notice that when $k(z) > 1$

$$\widehat{\nu_{z,n}}(t) \;:=\; \int_0^{\infty} e^{ixt} \nu_{z,n}(t)dx \;=\; \frac{1}{k(z)-1} \sum_{j=2}^{k(z)} e^{it(\ell_j - \ell_{j-1})}, \tag{2.5}$$

3

where $z = b_{\ell_1} + \cdots + b_{\ell_{k(z)}}$. Therefore

$$\mathbb{E}_z[\widehat{\nu_{z,n}}(t)] = \frac{1}{|I_n|} \sum_{z=b_{c_1 n + d_1}}^{b_{c_2 n + d_2} - 1} \frac{1}{k(z) - 1} \sum_{j=2}^{k(z)} e^{it(\ell_j - \ell_{j-1})}. \tag{2.6}$$

**Lemma 2.2.** *We have*

$$\lim_{n \to \infty} \frac{1}{|I_n|} \sum_{z=b_{c_1 n + d_1}}^{b_{c_2 n + d_2} - 1} \left( \frac{(k(z) - 1) - \mu_n}{(k(z) - 1)\mu_n} \right) \sum_{j=2}^{k(z)} e^{it(\ell_j - \ell_{j-1})} = 0, \tag{2.7}$$

*where we adopt the convention that for any $z \in I_n$ with $k(z) = 1$ the summand is just 1 (as the corresponding spacing measure is just a delta spike at 0).*

*Proof.* We break into cases based on how far away $k(z)$ is from the mean. For $0 < \delta < 1/2$

$$I_n(\delta) := \{z \in I_n : k(z) \in [\mu_n - (c_{\text{var}} n)^{1/2} n^\delta, \mu_n + (c_{\text{var}} n)^{1/2} n^\delta]\}. \tag{2.8}$$

**Case 1**: Let $z \in I_n(\delta)$. Thus $k(z)$ is close to $\mu_n$. As $k(z) \ll n$

$$\frac{1}{|I_n|} \sum_{\substack{z=b_{c_1 n + d_1} \\ z \in I_n(\delta)}}^{b_{c_2 n + d_2} - 1} \left( \frac{(k(z) - 1) - \mu_n}{(k(z) - 1)\mu_n} \right) \sum_{j=2}^{k(z)} e^{it(\ell_j - \ell_{j-1})} \ll \frac{1}{|I_n|} \sum_{z=b_{c_1 n + d_1}}^{b_{c_2 n + d_2} - 1} \frac{n^{1/2 + \delta}}{n^2} \sum_{j=2}^{k(z)} e^{it(\ell_j - \ell_{j-1})}$$

$$\ll \frac{|I_n| n}{n^{3/2 - \delta} |I_n|} = n^{-1/2 + \delta}, \tag{2.9}$$

where the last line follows because $k(z) \ll n$.

**Case 2**: Let $z \in I_n \setminus I_n(\delta)$. By assumption the number of $z \in I_n \setminus I_n(\delta)$ where $k(z) = 1$ divided by $I_n$ is $o(1)$, and thus trivial estimation shows the contribution of such $z$ in (2.7) is $o(1)$. Henceforth we assume $k(z) > 1$.

We first deal with the case when the number of summands converges pointwise to a Gaussian with mean $\mu_n = c_{\text{mean}} n + O(1)$ and variance $\sigma_n^2 = c_{\text{var}} n + O(1)$. For sufficiently large $n$ we have $\frac{c_{\text{var}}}{2\pi} n \le \sigma_n^2 \le 2 c_{\text{var}} n$, and thus the probability that $z \in I_n$ is in $I_n \setminus I_n(\delta)$ is bounded by

$$2016 \cdot 2 \int_{c_{\text{var}}^{1/2} n^{1/2 + \delta}}^{\infty} (2\pi c_{\text{var}} n / 2\pi)^{-1/2} \exp\left(-t^2 / 2 \cdot 2 c_{\text{var}} n\right) dt \ll n^{-1/2} e^{-n^{2\delta}/4} \tag{2.10}$$

(we do not need to be too careful here as this is a significantly lower order term; thus we have bounded the probability by 2016 times the integral of a Gaussian, where we increased the variance in the exponential and decreased it in the normalization constant in order to increase the integral). Therefore the number of integers $z \in I_n \setminus I_n(\delta)$ is bounded by $2016 |I_n| e^{-n^{2\delta}/4}$. Thus, remembering that we can assume $k(z) > 1$ so there is no division by zero below,

$$\frac{1}{|I_n|} \sum_{\substack{z=b_{c_1 n + d_1} \\ z \notin I_n(\delta)}}^{b_{c_2 n + d_2} - 1} \left( \frac{(k(z) - 1) - \mu_n}{(k(z) - 1)\mu_n} \right) \sum_{j=2}^{k(z)} e^{it(\ell_j - \ell_{j-1})} \ll \frac{1}{|I_n|} \cdot |I_n| e^{-n^{2\delta}/4} \cdot n = n e^{-n^{2\delta}/4}, \tag{2.11}$$

4

which tends to zero as $n \to \infty$ and proves the claim.

If the convergence were not pointwise but instead we had the probability of being more than $n^{\delta}$ standard deviations away decaying faster than $1/n^{1+\epsilon}$ for some $\epsilon > 0$, the only change is that now the left hand side of (2.11) is $O(1/n^{\epsilon})$, which still tends to zero as $n \to \infty$ and thus the rest of the argument is unchanged. $\qquad\square$

Through a similar argument we have

**Lemma 2.3.**

$$\lim_{n \to \infty} \frac{1}{|I_n|} \sum_{z=b_{c_1 n + d_1}}^{b_{c_2 n + d_2} - 1} \left( \frac{(k(z) - 1)^2 - \mu_n^2}{(k(z) - 1)^2 \mu_n^2} \right) \left( \sum_{j=2}^{k(z)} e^{it(\ell_j - \ell_{j-1})} \right)^2 = 0. \qquad (2.12)$$

We can now conclude the proof of Proposition 2.1 as follows.

*Proof of Proposition 2.1.* By Lemma 2.2, we replace $\frac{1}{k(z)-1}$ with $\frac{1}{\mu_n}$ with negligible error:

$$
\begin{aligned}
\mathbb{E}_z[\widehat{\nu_{z,n}}(t)] &= \frac{1}{|I_n|} \sum_{z=b_{c_1 n + d_1}}^{b_{c_2 n + d_2} - 1} \frac{1}{k(z) - 1} \sum_{j=2}^{k(z)} e^{it(\ell_j - \ell_{j-1})} = \frac{1}{|I_n| \mu_n} \sum_{z=b_{c_1 n + d_1}}^{b_{c_2 n + d_2} - 1} \sum_{j=2}^{k(z)} e^{it(\ell_j - \ell_{j-1})} + o(1) \\
&= \frac{1}{|I_n| \mu_n} \sum_{g=0}^{c_2 n + d_2 - 1} \sum_{j=1}^{c_2 n + d_2 - g} X_{j,j+g}(n) e^{itg} + o(1) = \sum_{g=0}^{c_2 n + d_2 - 1} P_n(g) e^{itg} + o(1),
\end{aligned}
$$
$$(2.13)$$

with the last equality follows by definition. Then

$$\lim_{n \to \infty} \mathbb{E}_z[\widehat{\nu_{z,n}}(t)] = \lim_{n \to \infty} \left( \sum_{g=0}^{c_2 n + d_2 - 1} P_n(g) e^{itg} + o(1) \right) = \sum_{g=0}^{\infty} P(g) e^{itg} = \widehat{\nu}(t), \qquad (2.14)$$

which completes the proof. $\qquad\square$

**Proposition 2.4.** *We have*

$$\lim_{n \to \infty} \mathrm{Var}_n(t) := \lim_{n \to \infty} \mathbb{E}_z[(\widehat{\nu_{z,n}}(t) - \widehat{\nu}_n(t))^2] = 0. \qquad (2.15)$$

*Proof.* Note that

$$\mathrm{Var}_n(t) := \lim_{n \to \infty} \mathbb{E}_z[(\widehat{\nu_{z,n}}(t) - \widehat{\nu}_n(t))^2] = \mathbb{E}_z[\widehat{\nu_{z,n}}(t)^2] - \widehat{\nu}_n(t)^2. \qquad (2.16)$$

We show that $\lim_{n \to \infty} \mathbb{E}_z[\widehat{\nu_{z,n}}(t)^2]$ differs from

$$\widehat{\nu}(t)^2 = \sum_{g_1=0}^{\infty} P(g_1) e^{itg_1} \sum_{g_2=0}^{\infty} P(g_2) e^{itg_2} = \sum_{g_1, g_2} P(g_1) P(g_2) e^{it(g_1 + g_2)} \qquad (2.17)$$

5

by $o(1)$. Let $g_1$ and $g_2$ be two arbitrary gaps starting at the indices $j_1 \leq j_2$. We have

$$\mathbb{E}_z[\widehat{\nu_{z,n}}(t)^2] = \frac{1}{|I_n|} \sum_{z=b_{c_1 n + d_1}}^{b_{c_2 n + d_2} - 1} \frac{1}{(k(z)-1))^2} \sum_{r=2}^{k(z)} e^{it(\ell_r(z) - \ell_{r-1}(z))} \sum_{w=2}^{k(z)} e^{it(\ell_w(z) - \ell_{w-1}(z))}$$

$$= \frac{1}{|I_n|} \frac{1}{\mu_n} \left( 2 \sum_{\substack{j_1 < j_2 \\ g_1, g_2}} X_{j_1, j_1 + g_1, j_2, j_2 + g_2}(n) e^{itg_1} e^{itg_2} + \sum_{j_1, g_1} X_{j_1, j_1 + g_1}(n) e^{2itg_1} \right) + o(1). (2.18)$$

The last line follows by Lemma 2.3 (the 2 is from $j_1 < j_2$). The diagonal term doesn't contribute to the limit as the denominator is of order $n^2 |I_n|$ and $\sum_{j_1, g_1} X_{j_1, j_1 + g_1}(n) e^{2itg_1}$ is of order $n|I_n|$. Using the second condition from Theorem 1.1 gives $\lim_{n \to \infty} \mathrm{Var}_n(t) = 0$. $\square$

*Proof of Theorem 1.1.* Lévy's Criterion (see [FG], page 361) states that if a sequence of random variables $\{R_n\}$ whose characteristic functions $\{\phi_n\}$ converge pointwise to $\phi$, where $\phi$ is the characteristic function of some random variable $R$, then the random variables $R_n$ converge to $R$ in distribution. In our case, Propositions 2.1 and 2.4 along with Chebyshev's Theorem (bounding the probability of being $s$ standard deviations from the mean by $1/s^2$) ensure that for any $\varepsilon > 0$, almost all of the characteristic functions $\widehat{\nu_{z,n}}(t)$ are within $\varepsilon$ of $\widehat{\nu}(t)$. Thus we can take a subset of $z \in I_n$ where the individual gap measure of each $z$ converge to the average measure as $n$ tends to infinity and almost all $z \in I_n$ are chosen. $\square$

## References

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

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

[BILMT]    A. Bower, R. Insoft, S. Li, S. J. Miller and P. Tosteson, *Distribution of Gaps between Summands in Generalized Zeckendorf Decompositions* (with an appendix on *Extensions to Initial Segments* with Iddo Ben-Ari), Journal of Combinatorial Theory, Series A **135** (2015), 130–160.

[CFHMN]    M. Catral, P. Ford, P. E. Harris, S. J. Miller and D. Nelson, *New Behavior in Legal Decompositions Arising from Non-positive Linear Recurrences*, preprint.

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

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

[DDKMV]    P. Demontigny, T. Do, A. Kulkarni, S. J. Miller and U. Varma, *A Generalization of Fibonacci Far-Difference Representations and Gaussian Behavior*, to appear in the Fibonacci Quarterly.
           http://arxiv.org/pdf/1309.5600v2.

[DFFHMPP]  R. Dorward, P. Ford, E. Fourakis, P. E. Harris, S. J. Miller, E. Palsson and H. Paugh, *A Generalization of Zeckendorf's Theorem via Circumscribed m-gons*, preprint.

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

[FG]       B. E. Fristedt and L. F. Gray, *A modern approach to probability theory*, Birkhäuser, Boston (1996).

[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.

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

[MW]        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.

[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 Liége **41** (1972), pages 179–182.

*E-mail address*: rdorward@oberlin.edu

Dept. of Mathematics, Oberlin College, Oberlin, OH 44074

*E-mail address*: fordpl@bethanylb.edu

Dept. of Mathematics and Physics, Bethany College, Lindsborg, KS 67456

*E-mail address*: erf1@williams.edu

Dept. of Mathematics and Statistics, Williams College, Williamstown, MA 01267

*E-mail address*: pamela.harris@usma.edu

Dept. of Mathematical Sciences, United States Military Academy, West Point, NY 10996

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

Dept. of Mathematics and Statistics, Williams College, Williamstown, MA 01267

*E-mail address*: eap2@williams.edu

Dept. of Mathematics and Statistics, Williams College, Williamstown, MA 01267

*E-mail address*: hannah.paugh@usma.edu

Dept. of Mathematical Sciences, United States Military Academy, West Point, NY 10996