Towards the Gaussianity of Random Zeckendorf Games

Justin Cheigh, Guilherme Zeus Dantas e Moura, Ryan Jeong, Jacob Lehmann Duke, Wyatt Milgrim, Steven J. Miller, and Prakod Ngamlamai

Abstract Zeckendorf proved that any positive integer has a unique decomposition as a sum of non-consecutive Fibonacci numbers, here indexed by $F_1 = 1, F_2 = 2, F_{n+1} = F_n + F_{n-1}$. Motivated by this result, Baird et al. [3] defined the two-player Zeckendorf game, in which two players take turns acting on a multiset of Fibonacci numbers that always sums to N. The game terminates when no possible moves remain, which importantly always happens, and the final player to perform a move wins. Notably, Baird et al. [3] empirically studied the setting of random games, in the sense that the game proceeds by always choosing an available move uniformly at random, and conjecture that as the input $N \to \infty$, the distribution of random game lengths converges to a Gaussian. We study various combinatorial questions concerning the Zeckendorf game. We found that the sum of the number of times certain moves are performed is constant. We prove that the number of shortest games on input N is at least $\prod_{k=1}^{n-2} \mathrm{Cat}(F_k)$, where n denotes the index of

J. Cheigh · J. L. Duke · S. J. Miller (✉)
Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267, USA
e-mail: sjm1@williams.edu; Steven.Miller.MC.96@aya.yale.edu

J. Cheigh
e-mail: jhc5@williams.edu

J. L. Duke
e-mail: jl34@williams.edu

G. Z. Dantas e Moura
Department of Mathematics and Statistics, Haverford College, Haverford, PA 19041, USA
e-mail: gdantasemo@haverford.edu; zeusdanmou@gmail.com

R. Jeong
Department of Mathematics, University of California, Los Angeles, CA 90095, USA
e-mail: rsjeong@sas.upenn.edu

W. Milgrim
Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, USA
e-mail: wmilgrim@math.ucla.edu

P. Ngamlamai
Department of Mathematics, The Ohio State University, Columbus, OH 43210, USA
e-mail: ngamlamai.1@osu.edu

the largest Fibonacci number in the Zeckendorf decomposition of N and $\text{Cat}(F_k)$ is the F_kth Catalan number. The works of Baird, Epstein, Flint, and Miller [3] and Cuzensa et al. [5] determined how to play in order to achieve the shortest and longest possible Zeckendorf game on a given input N, respectively: we improve the current understanding of achievable game lengths by establishing that for any input N, the range of possible game lengths constitutes an interval of natural numbers; in other words, for every input N, every game length between the shortest and longest game lengths can be achieved by some Zeckendorf game. Motivated towards the resolution of the Gaussianity conjecture, we also further the study of probabilistic aspects of random Zeckendorf games. In particular, we study two probability measures on the space of all Zeckendorf games on input N: the uniform measure, and the measure induced by choosing moves uniformly at random at any given configuration. We show under both measures that in the limit $N \to \infty$, both players win with probability $1/2$ when playing under the random game setting. We also find natural partitions of the collection of all Zeckendorf games of a fixed input N, on which we observe weak convergence to a Gaussian in the limit $N \to \infty$. We conclude the work with many open problems.

1 Introduction

The Fibonacci numbers are widely considered to be the most interesting and well-known recursive sequence in mathematics. In this article, we shall index the Fibonacci numbers by $F_1 = 1$, $F_2 = 2$, and for general $n \geq 3$, $F_n = F_{n-1} + F_{n-2}$. Zeckendorf proved the following fundamental theorem: the decomposition in the proceeding theorem is referred to as the **Zeckendorf decomposition** of the positive integer N.

Theorem 1 ([9]) *Every positive integer N can be decomposed uniquely into a sum of distinct, non-consecutive Fibonacci numbers.*

Inspired by this result, the authors of [3] constructed the **two-player Zeckendorf game**.

Definition 1 ([3]) Given input $N \in \mathbb{N}$, the Zeckendorf game is played on a multiset of Fibonacci numbers, initialized at $\mathcal{S} = \{F_1^N\}$. On each turn, a player can act on the multiset by performing one of the following moves if it is available.

1. If we have two consecutive Fibonacci numbers F_{k-1}, F_k for some $k \geq 2$, then we can replace them by F_{k+1}, denoted $F_{k-1} \wedge F_k \to F_{k+1}$.
2. If we have two instances of the same Fibonacci number F_k, then

   a. If $k = 1$, we can play $F_1 \wedge F_1 \to F_2$.
   b. If $k = 2$, we can play $F_2 \wedge F_2 \to F_1 \wedge F_3$.
   c. If $k \geq 3$, we can play $F_k \wedge F_k \to F_{k-2} \wedge F_{k+1}$.

The two players alternate turns until no playable moves remain. The last player to move wins the game.

Observe that the moves of the game are consistent with the Fibonacci recurrence: we either *combine* two consecutive terms, or *split* terms with multiple instances. Perhaps more intuitively, we can understand the game as acting on a row of bins, with bin k corresponding to the Fibonacci number F_k and its height being the multiplicity of F_k in the multiset.

1.1 Prior Work

The article [3] introduces the two-player Zeckendorf game, determine upper and lower bounds on the length of a game on input N (showing in particular that the game always terminates), and shows non-constructively that Player 2 has the winning strategy for all $N \geq 2$. In particular, they provide the following explicit formula for the length of the shortest Zeckendorf game on input N, achieved by only playing combine moves.

Theorem 2 ([3]) *The number of combine moves in any Zeckendorf game on input N is is $N - Z(N)$. Furthermore, any such shortest game terminates in $N - Z(N)$ moves, where $Z(N)$ is the number of terms in the Zeckendorf decomposition of N.*

The works [7] and [5] successively improved the upper bound of [3] on the length of a Zeckendorf game on fixed input N; the former article finds a deterministic game which has longest possible length for input N, while the latter generalizes this paradigm. We frequently make use of the following two results from [5] in our arguments. We note that although the following result provides a strategy to achieve a longest game, finding a convenient closed form for the length of the longest game for non-Fibonacci input N remains open (with the case of Fibonacci input treated by [5]).

Theorem 3 ([5]) *The longest game on any N is achieved by applying split moves or combine 1s (in any order) whenever possible, and, if there is no split or combine 1 move available, combine consecutive indices from smallest to largest.*

Theorem 4 ([5]) *A Zeckendorf game on input N can be played with strictly splitting and combine 1 moves if and only if $N = F_k - 1$ for some $k \geq 2$.*

Finally, we remark that analogous two-player games have been developed for other recurrences: the work [2] extends [3] by defining and studying such games for recursive sequences defined by linear recurrence relations of form $G_n = \sum_{i=1}^{k} cG_{n-i}$ ($c = k - 1 = 1$ yielding the Fibonacci numbers), again giving lower and upper bounds on game lengths (including showing termination) and showing non-constructively that Player 2 has a winning strategy, while [4] similarly studies recurrences of form $a_{n+1} = na_n + a_{n-1}$.

1.2   Notation and Conventions

We let C_1 denote the combine move $F_1 \wedge F_1 \to F_2$, and (for $k \geq 2$) let C_k denote the combine move $F_{k-1} \wedge F_k \to F_{k+1}$. Let S_2 denote the splitting move $2F_2 \to F_1 \wedge F_3$, and (for $k \geq 3$) let S_k denote the splitting move $2F_k \to F_{k-2} \wedge F_{k+1}$. We prefix a particular type of move with M to denote the number of such moves (e.g. MC_1 denotes the number of C_1's in a game).

We let $Z(N)$ denote the number of terms in the Zeckendorf decomposition of N. We loosely refer to the number of instances of F_k as the **height** of bin k, denoted h_k; it will usually be clear from context at which point in the game the quantity h_k refers to. When discussing the height of bin k after a specific number m of moves in the game, we notate this by $h_k(m)$. For $\lambda \in \mathbb{N}$, we shall also occasionally use the shorthand $[\lambda] = \{1, 2, \ldots, \lambda\}$.

In this work, we shall generally work under the assumption that $F_n \leq N < F_{n+1}$ for some $n \in \mathbb{N}$ (i.e., n is the index of the largest Fibonacci number that is no larger than N).[1] While proving Theorem 7, we occasionally refer to moves $C_1, S_2, \ldots, S_{n-1}$ as **Type A moves**, and all other moves (namely, moves C_k for $k \geq 2$) as **Type B moves**. The work [5] also achieved an understanding of precisely when playing strictly Type A moves throughout the whole game is possible.

Finally, the present article furthers the study of random Zeckendorf games. Here, we let Ω_N denotes the (finite) collection of all Zeckendorf games on input N, with $\mathcal{F}_N = 2^{\Omega_N}$ the associated σ-algebra, and express a given Zeckendorf game $\mathcal{G} \in \Omega_N$ as a (finite) sequence of λ moves, written as $\mathcal{G} = (M_1, M_2, \ldots, M_\lambda)$. We study two probability measures to complete the space $(\Omega_N, \mathcal{F}_N)$: the uniform measure μ_N, defined by

$$\mu_N(\mathcal{G}) = \frac{1}{|\Omega_N|} \text{ for all } \mathcal{G} \in \Omega_N$$

and the probability measure $\mathbb{P}_N$ induced by choosing, at every configuration along a given game, uniformly at random among available moves, defined by

$$\mathbb{P}_N(\mathcal{G}) = \prod_{k=1}^{\lambda} \frac{1}{\text{num. playable moves after } (M_1, \ldots, M_{k-1})}$$

for $\mathcal{G} = (M_1, \ldots, M_\lambda) \in \Omega_N$. All of the results we derive in this context apply to both probability spaces.

---

[1] This is why we have elected to deviate from notation traditionally used in papers concerning this game.

1.3 Main Results

The work [5] determines an upper bound on the length of a game on input N. We improve this upper bound using similar techniques as in the work [3].

Theorem 5 *The length of a Zeckendorf game on input N is upper-bounded by*

$$\left\lfloor \varphi^2 N - Z_I(N) - 2Z(N) + (\varphi - 1) \right\rfloor$$

where $Z_1(N)$ represents the index sum of the Zeckendorf decomposition of N. Furthermore, the bound is sharp for infinitely many N.

Much of our work was inspired by the following conjecture (the only one still unresolved in the paper it was introduced in), initially posed by [3], which concerns distributional properties of the length of random Zeckendorf games on input N in the limit $N \to \infty$.

Conjecture 1 ([3, 7]) In the limit $N \to \infty$, the distribution of the number of moves in a random Zeckendorf game on input N converges to a Gaussian, with expectation and variance approximately $0.215N$.[2]

As such, many of our main results have largely arisen from attempting to understand those aspects of Zeckendorf games which may potentially aid in resolving the aforementioned conjecture (and in striving to determine what such aspects are). First, we have the following lower bound on the number of shortest Zeckendorf games of length N. Intuitively, if the distribution of random game lengths were indeed Gaussian, this should be an extreme underestimate compared to the number of ways to achieve other game lengths (shortest games involve the fewest number of decisions, so one might naturally expect that the probability of achieving one via a random game is larger than longer games), yet it still explodes in N.

Theorem 6 *Let $F_n \le N < F_{n+1}$. Then the number of shortest Zeckendorf games with input N is at least $\prod_{k=1}^{n-2} Cat(F_k)$, where $Cat(F_k)$ denotes the F_kth Catalan number.*

The shortest game and longest game were studied in [3] and [5], respectively. It is natural to ask whether every game length between the shortest and longest game length is achievable: we resolve this in the affirmative.

Theorem 7 *For any input N to the Zeckendorf game, let M denote the length of the longest Zeckendorf game with input N. Then for any m satisfying $N - Z(N) \le m \le M$, there exists a Zeckendorf game of length m on input N. In other words, the set of achievable game lengths constitutes an interval in the natural numbers.*

---

[2] The authors of [3] posed this conjecture based on numerical data gathered from 9,999 simulations of a random game with $n = 18$. The authors of [7] gathered further numerical evidence with a sample of 1,000 games with $n = 1, 000, 000$. We ran a brute force enumeration over all possible games for $n \le 18$ and found the distribution of lengths appeared to be Gaussian. This is a slightly different problem than the random game though is closely related.

We also study the winning odds of players in the limit $N \to \infty$ of infinite input when studying random Zeckendorf games, for which one might expect that both players win with probability $1/2$ in the limit if if Conjecture 1 holds as the variance of the conjectured Gaussian grows with N. We establish that this is indeed true by proving a much more general result: we can understand Theorem 8 as saying that in the limit of infinite input, a Z-player random Zeckendorf game is fair, in the sense that all Z players have the same probability of winning.

Theorem 8 *For any integer $Z \geq 1$ and $z \in \{0, 1, \ldots, M - 1\}$, we have that*

$$\lim_{N \to \infty} \mu_N(\text{Game length} \equiv z \bmod Z) = \lim_{N \to \infty} \mathbb{P}_N(\text{Game length} \equiv z \bmod Z) = \frac{1}{Z}.$$

Taking $Z = 2$ in Theorem 8 above yields the following result for the classical two-player Zeckendorf game.

Theorem 9 *For the two-player Zeckendorf game, in the limit $N \to \infty$ under both probability measures μ_N and $\mathbb{P}_N$, Player 1 and Player 2 are equally likely to win. Explicitly,*

$$\lim_{N \to \infty} \mu_N(\text{Player 1 wins}) = \lim_{N \to \infty} \mu_N(\text{Player 2 wins}) = \frac{1}{2},$$

$$\lim_{N \to \infty} \mathbb{P}_N(\text{Player 1 wins}) = \lim_{N \to \infty} \mathbb{P}_N(\text{Player 2 wins}) = \frac{1}{2}.$$

Finally, we establish that there exist natural ways to partition the collection of Zeckendorf games Ω_N on input N so that the distribution of game lengths over the corresponding classes are nearly Gaussian with high probability in the limit $N \to \infty$. The construction of the subsets $\mathcal{R}_N^{\mathcal{P}} \subset \Omega_N$ and $\mathcal{R}_N^{\mathcal{S}} \subset \Omega_N$, and the sets $\mathcal{A}_N(\mathcal{R})$, is elaborated in Propositions 5 and 6.

Theorem 10 *For $\mathcal{R} \in \mathcal{R}_N^{\mathcal{P}}$, let $F_N^{\mathcal{R}}(x) : \mathbb{R} \to [0, 1]$ denote the distribution function corresponding to game lengths in $\mathcal{A}_N(\mathcal{R})$ over the conditional distribution induced by $\mathbb{P}_N$, normalized to have expectation 0 and variance 1. Let $\Phi : \mathbb{R} \to [0, 1]$ denote the distribution function of the standard normal. Then for any $\epsilon > 0$,*

$$\lim_{N \to \infty} \mathbb{P}_N \left(\sup_{x \in \mathbb{R}} \left| F_N^{\mathcal{R}}(x) - \Phi(x) \right| \geq \epsilon \right) = 0.$$

Similarly, for $\mathcal{R} \in \mathcal{R}_N^{\mathcal{S}}$, let $F_N^{\mathcal{R}}(x) : \mathbb{R} \to [0, 1]$ denote the distribution function corresponding to game lengths in $\mathcal{A}_N(\mathcal{R})$ over the conditional distribution induced by $\mathbb{P}_N$, normalized to have expectation 0 and variance 1. Then for any $\epsilon > 0$,

$$\lim_{N \to \infty} \mathbb{P}_N \left(\sup_{x \in \mathbb{R}} \left| F_N^{\mathcal{R}}(x) - \Phi(x) \right| \geq \epsilon \right) = 0.$$

The analogous results hold for the uniform measure μ_N.

2 Structural Results

In this section, we include some straightforward, but fundamental results concerning the nature of the Zeckendorf game; some of these will be invoked in proofs of deeper theorems.

2.1 Combinatorial Observations

We begin by exploring some basic properties of the Zeckendorf game, observable by studying deterministic subroutines of moves. The following simple result mirrors techniques in [3].

Proposition 1 *Consider any decomposition of N into a sum of (possibly non-distinct, non-consecutive) Fibonacci numbers: this decomposition can be achieved via a sequence of combine moves from the starting configuration of the Zeckendorf game.*

Proof We "play the game in reverse": consider the configuration corresponding to this decomposition, and construct a sequence of moves by always taking the game piece not in the first bin (i.e., F_1) and farthest out, and replace it as the result of a combine move. Specifically, if $k \geq 3$, then replace F_k by $\{F_{k-2}, F_{k-1}\}$; if $k = 2$, replace with $2F_1$. Then reverse all the moves to get a Zeckendorf game from the initial state to this state.

The following is an easy consequence of Theorem 3 above, which states that the longest game paradigm from the starting position extends to intermediate game positions on input N which are given by converting $F_n - 1$ instances of $F_1 = 1$ into the Zeckendorf decomposition of $F_n - 1$, where n denotes the index of the largest Fibonacci number in the Zeckendorf decomposition of the input N.

Lemma 1 *Let n denote the index of the largest Fibonacci number in the Zeckendorf decomposition of the input N. A longest Zeckendorf game from an intermediate configuration given by converting $F_n - 1$ instances of 1 into the Zeckendorf decomposition of $F_n - 1$ is given by greedily playing any Type A move whenever possible, and if no such Type A move can be played, play the available Type B move with the smallest index.*

Proof If a game achieved by playing Type A moves whenever possible from this configuration were not maximal (i.e., there existed a Zeckendorf game of strictly larger length), then by initially playing the longest game on input $F_n - 1$ via all Type A moves (possible by Theorem 4), we can play the game exactly according to Theorem 3 but fail to achieve a game of maximal length, contradicting Theorem 3.

Using similar techniques as in [3], we derive the following results in order to improve the upper bound found in [5].

Lemma 2 *Let n be the largest summand in the Zeckendorf decomposition of N, we get that for any $2 \le k \le n - 1$, the following sum is constant:*

$$MS_k + MC_k + MC_{k+1} + \cdots + MC_{n-1}.$$

Proof Consider the following relabeling of the board:

F_1	$\cdots$	F_k	$F_{k+1} - 1$	$F_{k+2} - 2$	$F_{k+3} - 4$	$F_{k+4} - 7$	$\cdots$

where after the kth bin, the value of a bin is equal to one less than the sum of the values of the two bins which precede it. We get that only the moves $S_k, C_k, C_{k+1}, \ldots$ can change the weighted sum of the tokens by the relabeled values, and each of these moves reduce the sum by 1. Since we have a fixed initial sum N and a fixed ending sum depending on the Zeckendorf Decomposition of N, we get that the sum of those moves must be constant regardless of how the game is played.

Lemma 3 *For any Zeckendorf game starting with N tokens,*

$$MC_1 - MS_2 \approx (2 - \varphi)N$$

with approximation error $\le \varphi - 1$.

Proof Similarly, we prove this with a relabeling of the board

2	3	5	$\cdots$	F_{k+1}	$\cdots$

and observing that the sum of token values goes from $2N$ to $\sim \varphi N$ with the sum decreasing by 1 only by performing C_1 and increasing by 1 only by performing S_2. Note that the final sum is equal to shifting each Zeckendorf summands of N forward by one, which is approximated by multiplying each by φ. By Binet's formula, we have $F_k = \frac{\varphi^{k+1} - (-1/\varphi)^{k+1}}{\sqrt{5}}$ due to how we index the Fibonacci sequence. Thus, the error of approximating the summand F_{k+1} with φF_k is

$$\left| \frac{(\varphi^{k+1} - (-1/\varphi)^{k+1})\varphi - (\varphi^{k+2} - (-1/\varphi)^{k+2})}{\sqrt{5}} \right| = \frac{\varphi^2 + 1}{\varphi^{k+2}\sqrt{5}}$$

The largest error happens when $N = F_1 + F_3 + F_5 + \cdots$ with error at most

$$\frac{\varphi^2 + 1}{\varphi^3 \sqrt{5}} \left(\sum_{i=0}^{\infty} \frac{1}{\varphi^{2i}} \right) = \frac{\varphi^2 + 1}{\varphi^3 (1 - 1/\varphi^2) \sqrt{5}} = \frac{1}{\varphi} = \varphi - 1$$

which yields the desired.

As a corollary, we prove Theorem 5

Theorem 5 *The length of a Zeckendorf game on input N is upper-bounded by*

$$\left\lfloor \varphi^2 N - Z_I(N) - 2Z(N) + (\varphi - 1) \right\rfloor$$

where $Z_I(N)$ represents the index sum of the Zeckendorf decomposition of N. Furthermore, the bound is sharp for infinitely many N.

Proof Using the relabeling of the board given by

3	4	5	$\cdots$	$k+2$	$\cdots$

we get that

$$2MC_1 + 2MC_2 + 3MC_3 + 4MC_4 + \cdots = 3N - 2Z(N) - Z_I(N)$$
$$+ MS_3 + MS_4 + \cdots$$

Applying Lemma 3, we get

$$MC_1 + 2MC_2 + 3MC_3 + 4MC_4 + \cdots \leq 3N - 2Z(N) - Z_I(N)$$
$$+(2 - \varphi)N - (\varphi - 1) + MS_2 + MS_3 + MS_4 + \cdots$$

Thus, if we subtract the excess $MC_2 + 2MC_3 + 3MC_4 + \cdots$ from the left hand side, we get the upper bound

$$MC_1 + MC_2 + MC_3 + MC_4 + \cdots \leq (1 + \varphi)N - 2Z(N) - Z_I(N) + (\varphi - 1)$$
$$+ MS_2 + MS_3 + MS_4 + \cdots$$

which we round down as the number of all moves performed is an integer. Note that if we consider $N = F_n - 1$ then we get that there are no C_2, C_3, performed within the longest game. Furthermore, if we consider $N = F_{2n} - 1$, then our approximation error is less than 1 and thus must be sharp after rounding down.

Recall that [3] proved non-constructively that Player 2 always has a winning strategy for any $N \geq 3$: finding such a winning strategy remains open. It is generally believed [8] that the key to such a strategy lies in understanding "parity swaps": distinct sequences of moves of differing length which yield the same effect on the board. The following definition follows from the easy observation that whenever playable, the sequences of moves

$$S_k \rightarrow S_{k-1} \rightarrow \cdots \rightarrow S_{k-\ell+1} \rightarrow C_{k-\ell} \quad C_{k-\ell} \rightarrow S_{k-\ell+1} \rightarrow \cdots \rightarrow S_{k-1} \rightarrow S_k$$

both have the same effect on the board as the move C_k, for some $k \geq 2$.

Definition 2 For any $\ell \geq 0$ and $k \geq 2$, call a sequence of moves of form $S_k \rightarrow S_{k-1} \rightarrow \cdots \rightarrow S_{k-\ell+1} \rightarrow C_{k-\ell}$ an (ℓ, k)-*prefix*, and a sequence of moves of form

$C_{k-\ell} \to S_{k-\ell+1} \to S_{k-\ell+2} \to \cdots \to S_k$ an (ℓ, k)-*suffix*. We call an (ℓ, k)-prefix a *prefix of C_k*, and an (ℓ, k)-suffix a *suffix of C_k*.

It should be emphasized that an (ℓ, k)-prefix or (ℓ, k)-suffix corresponds to an equivalent action as the combine move C_k, for all lengths $\ell \geq 0$ and $k \geq 2$; the combine move is thus "expanded" via a sequence of ℓ splitting moves with contiguous indices. The next result captures the intuition that the variation in game lengths is entirely due to the parity swaps described in Definition 2, namely by describing arbitrary Zeckendorf games via *permutations of suffixes*.

Proposition 2 *Any Zeckendorf game on input N can be achieved by taking a shortest game, expanding combine moves via suffixes, then shifting the splitting moves.*

Proof For an arbitrary Zeckendorf game, greedily take the earliest split move of a game, move it back to be played as early as possible, and compress it into a combine move. Proceed similarly until we only have combine moves, from which we achieve the original game by reversing the compressions and playing split moves later. □

We can interpret the statement of Proposition 2 as saying that we can greedily embed an arbitrary Zeckendorf game on input N into a shortest game on input N in a natural way.

2.2 Shortest Games

Proposition 2 suggests that a study of shortest Zeckendorf games might be fruitful, as any particular Zeckendorf game can be understood as an extension and permutation of a particular shortest game. We first observe the following.

Proposition 3 *Shortest games are exactly those games which strictly use combine moves. Furthermore, such a game exists for any input N, and the multiset of combine moves for any such shortest game is unique.*

Proof A move decreases the number of pieces by at most one, so $N - Z(N)$ lower-bounds the number of moves necessary, achieved exactly by those games using strictly combine moves; such games exist by Proposition 1. To establish uniqueness of the multiset of combine moves for any such game, say $F_n \leq N < F_{n+1}$, and study bin k for $1 \leq k \leq n$. Moves affecting h_k are known precisely: letting the Zeckendorf decomposition of N be denoted $(z_1, z_2, \ldots, z_n)$ (where $z_i \in \{0, 1\}$), this yields the system

$$N - 2MC_1 - MC_2 = z_1$$
$$MC_1 - MC_2 - MC_3 = z_2$$
$$\cdots$$
$$MC_{n-3} - MC_{n-2} - MC_{n-1} = z_{n-2}$$
$$MC_{n-2} - MC_{n-1} = z_{n-1}$$
$$MC_{n-1} = z_n = 1$$

from which it easily follows that this system must have a unique solution.

Proposition 4 *For $n = n(N)$ and any $\delta \in (0, 1)$,*

$$\lim_{N \to \infty} \max_{\mathcal{G} \in \Omega_N} \frac{\#\{C_k \in \mathcal{G}, k > \delta n\}}{\#\{Combine\ moves\ in\ \mathcal{G}\}} = \lim_{N \to \infty} \max_{\mathcal{G} \in \Omega_N} \frac{\#\{C_k \in \mathcal{G}, k > \delta n\}}{N - 1} = 0.$$

Proof For $\mathcal{G} \in \Omega_N$, consider all combine moves C_k with $k > \delta n$; each such combine move corresponds to a particular token "jumped," for which we consider all combine moves with index $\leq \delta n$ that led this token to land in this position, of which there must be at least $\lfloor \delta n/2 \rfloor$. The sets of combine moves with index $\leq \delta n$ corresponding to distinct combine moves with index $> \delta n$ are observed to be disjoint, so

$$\frac{\#\{C_k \in \mathcal{G}, k > \delta n\}}{N - 1} \leq \frac{1}{\lfloor \delta n/2 \rfloor + 1}$$

which vanishes as $N \to \infty$.

Proposition 3 yields the following interesting lower bound[3] on the number of shortest Zeckendorf games with input N.

Theorem 6 *Let $F_n \leq N < F_{n+1}$. Then the number of shortest Zeckendorf games with input N is at least $\prod_{k=1}^{n-2} Cat(F_k)$, where $Cat(F_k)$ denotes the F_kth Catalan number.*

Proof It suffices to study $N = F_n$, since the number of distinct shortest Zeckendorf games is increasing in N: for input $N \neq F_n$, one can first play a shortest Zeckendorf game on input F_n, then proceed by always playing the rightmost available combine move to achieve a shortest game on input N by Proposition 3. (See Lemma 4, where we understand this as incrementing instances of F_1s after playing a shortest game on F_n.) The Zeckendorf decomposition is $(z_1, z_2, \ldots, z_{n-1}, z_n) = (0, 0, \ldots, 0, 1)$, where z_i denotes the number of instances of F_i in the decomposition. By Proposition 3, there exists a unique multiset of combine moves constituting a shortest game: by solving the system above, this multiset is defined by

$$(MC_1, MC_2, \ldots, MC_{n-1}) = (F_{n-2}, F_{n-3}, F_{n-4}, \ldots, F_3, F_2, F_1, F_0)$$

where we let $F_0 = 1$. A permutation of these moves constitutes a game if and only if every move is valid, i.e., no move would force the height of any bin to become negative. Specifically, at any intermediate point in the sequence, the number of C_1's played is no less than the sum of the number of C_2's and C_3's played (bin 2 is nonnegative), the number of C_2's is no less than the number of C_3's and C_4's played (bin 3 is nonnegative), and so on, to the number or C_{n-2}'s being no less than the number of C_{n-1}'s played (bin $n - 1$ is nonnegative). (We need not study bin 1 or bin n, which will necessarily always have nonnegative height.)

---

[3] We suspect the lower bound of Theorem 6 to be somewhat loose, as much is lost when crudely pursuing the interweaving of the Dyck paths $\pi_2, \pi_3, \ldots, \pi_{n-1}$ (see the proof for details).

Let us restrict our attention to the moves C_1, C_2, C_3 (moves affecting the height of bin 2): the number of permutations of these moves such that the number of C_1's played being no less than the sum of the number of C_2's and C_3's performed holds at any point in the game is in bijective correspondence with (up-down) Dyck paths on $j = F_{n-2}$ ($C_1 \leftrightarrow U$; $C_2, C_3 \leftrightarrow D$), the number of which is $\text{Cat}(F_{n-2})$. Similarly, by studying moves affecting bin $k \geq 2$, we achieve bijective correspondences with Dyck paths on $j = F_{n-k}$.

For any choice of Dyck paths $\pi_2, \pi_3, \ldots, \pi_{n-1}$ on $j = F_{n-2}, F_{n-3}, \ldots, F_1$, respectively, there exists a shortest Zeckendorf game on input N where the ordering of the relevant moves is consistent with the bijections described above. To construct such a game, begin by placing $2F_{n-2}$ moves along a line, labeling F_{n-2} of them as C_1 in a manner consistent with π_2 ($C_1 \leftrightarrow U$). Among the $F_{n-2} = F_{n-3} + F_{n-4}$ unlabeled moves, label F_{n-3} of them as C_2 (importantly, including the first D move) in a manner consistent with π_3 (not too many unlabeled moves between consecutive instances of C_2) and the other F_{n-4} as C_3. Now add F_{n-5} instances of C_4 along this line to complete π_3 (include all missing $C_4 \leftrightarrow D$ moves) while respecting π_4 (not too many unlabeled moves between consecutive instances of C_3). Specifically, construct a labeling of the $F_{n-3} = F_{n-4} + F_{n-5}$ D moves in π_3 with F_{n-4} C_3's and F_{n-5} C_4's such that the first D is labeled C_3, and there are not too many unlabeled D moves between consecutive instances of C_3 (with respect to π_4). Insert C_4's to be adjacent to established instances of C_2 and C_3 to be consistent with this labeling.

Continue by similarly adding, for $k \geq 5$, F_{n-k-1} instances of C_k, completing π_{k-1} while respecting π_k, until we add $F_0 = 1$ instance of C_{n-1} such that we complete π_{n-1}. This results in a shortest Zeckendorf game with the ordering of the relevant moves being consistent with $\pi_2, \pi_3, \ldots, \pi_{n-1}$.

3   The Set of Possible Game Lengths Constitute an Interval

In this section, we prove Theorem 7, which we restate below.

Theorem 7 *For any input N to the Zeckendorf game, let M denote the length of the longest Zeckendorf game with input N. Then for any m satisfying $N - Z(N) \leq m \leq M$, there exists a Zeckendorf game of length m on input N. In other words, the set of achievable game lengths constitutes an interval in the natural numbers.*

We begin by establishing some intermediate results that we shall invoke in the proof of the main theorem. In the first lemma, in discussing the position given by the Zeckendorf decomposition of $N - 1$, we refer to the terminal position of the Zeckendorf game when played on input $N - 1$.

Lemma 4 *Consider the Zeckendorf game on input N, satisfying $F_n \leq N < F_{n+1}$, from the position given by the Zeckendorf decomposition on $N - 1$ (as specified above) with an additional instance of 1. There is a unique sequence of moves from this configuration to the Zeckendorf decomposition of N, all of which are combine*

moves. Furthermore, the number of such combine moves performed is bounded by $\lfloor n/2 \rfloor$.

Proof Since no moves can be played from the Zeckendorf decomposition on $N - 1$, any playable move from this position (on input N) is necessarily either C_1 or C_2, possible if and only if $F_1 = 1$ or $F_2 = 2$ is in the Zeckendorf decomposition of $N - 1$, respectively (such cases are disjoint, since the Zeckendorf decomposition on $N - 1$ does not contain consecutive Fibonacci numbers); otherwise, no moves can be played. We study both cases.

- If the Zeckendorf decomposition of $N - 1$ contains an instance of F_1, then after playing C_1, it is easy to see that the only possible move is C_3 (iff the decomposition of $N - 1$ contains an F_3), then C_5 (iff the decomposition of $N - 1$ contains an F_5), and so on, until we exhaust all playable moves.
- If the Zeckendorf decomposition of $N - 1$ contains an instance of F_2, then after playing C_2, it is easy to see that the only possible move is C_4 (iff the decomposition of $N - 1$ contains an F_4), then C_6 (iff the decomposition of $N - 1$ contains an F_6), and so on, until we exhaust all playable moves.

In both cases, it is straightforward to confirm that we cannot play strictly more than $\lfloor n/2 \rfloor$ such combine moves, as otherwise there must be an instance of F_k for k either $n + 1$ or $n + 2$ after completing this sequence of moves, a contradiction on $N < F_{n+1}$. □

We shall also frequently use the following lemma. Intuitively, this states that if we have isolated the Zeckendorf game to a suffix of bins all of height 0 or 1, and this suffix is separated from earlier bins by a bin of height 0 (say bin k), then we can ignore bins $1, \ldots, k$ for the remainder of the game.

Lemma 5 *For some $1 \le k \le n$, let $(x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n)$ denote the heights of bins $1, \ldots, n$ at some point during a Zeckendorf game, with x_i denoting the height of bin i. Assume $x_i \in \{0, 1\}$ for $k \le i \le n$, $x_k = 0$, no playable moves involving bins $1, \ldots, k$ exist, and we play according to Theorem 3. Then heights $x_1, \ldots, x_k$ remain fixed for the rest of the game.*

Proof We prove this on $k = n - j$ by induction on $0 \le j \le n - 2$. The statement trivially holds if $j \in \{0, 1, 2\}$; assuming it for all values less than $j \ge 3$, if $x_{k+1} = 0$ or $x_{k+2} = 0$ we can apply the induction hypothesis to $j - 1$ or $j - 2$, respectively, so assume $x_{k+1} = x_{k+2} = 1$. Take $r \le n$ to be largest possible such that $x_{k+1} = \cdots = x_r = 1$: the game proceeds by playing according to Theorem 3, i.e., by playing the sequence of moves

$$(C_{k+1} \to S_{k+3} \to S_{k+4} \to \cdots \to S_r) \to (C_{k+1} \to S_{k+3} \to \cdots \to S_{r-2}) \to \cdots$$

where the final subsequence of moves, either C_{k+1} or $C_{k+1} \to S_{k+3}$, depends on the parity of $r - k$. It follows immediately by studying the moves involved that $x_1, \ldots, x_k$

remain unchanged during this sequence. Following this sequence, we have $x_r = 0$: invoke the induction hypothesis on $k = r$ (i.e., $n - r < j$) afterwards to complete the proof.

We now proceed with the proof of Theorem 7.

Proof *(Proof of Theorem 7)* We have confirmed this statement for $N \leq F_6 - 1 = 12$ via a computer check. Thus, for $N \geq F_6 = 13$, assume the statement holds for all input sizes at most $N - 1$: we aim to show the result holds for N. We explicitly specify that $F_n \leq N < F_{n+1}$.

Let I_1' denote the interval (by induction hypothesis) of possible Zeckendorf game lengths for input $N - 1$: by Lemma 4, if we include an additional instance of F_1 to the Zeckendorf decomposition of $N - 1$, there is a sequence of combine moves from the resulting configuration to the Zeckendorf decomposition of N. On input N, consider initially playing the Zeckendorf game (to completion) as if the input were $N - 1$, then executing this sequence of moves to terminate the game. Let $I_1 = [L_1, R_1]$ denote the interval I_1' shifted by the length of this sequence: by the preceding description, it follows that every game length in I_1 is achievable for input N. Furthermore, $L_1 = N - Z(N)$, since the game length L_1 as studied above results from playing strictly combine moves (see Lemma 4), which necessarily yields a shortest game by Proposition 3, and for which the length is $N - Z(N)$ by Theorem 2.

By Theorem 4, it is possible to play the Zeckendorf game on input $F_n - 1$ strictly using Type A moves. Let I_2' denote the interval (by induction hypothesis) of possible Zeckendorf game lengths for input $F_n - 1 \leq N - 1$: on input N, consider initially playing the Zeckendorf game (to completion) as if the input were $F_n - 1$, then executing the longest possible sequence of moves from the resulting position to terminate the game. Let $I_2 = [L_2, R_2]$ denote the interval I_2' shifted by the number of moves of this longest sequence. By Theorem 3, R_2 is necessarily the length of the longest Zeckendorf game on input N, since the above approach is consistent with playing Type A moves whenever possible.

Thus, it suffices to show that $L_2 \leq R_1$ to complete the induction and establish the theorem, as this yields that $I_1 \cup I_2 = [L_1, R_2]$ is an interval of achievable game lengths on input N, with the endpoints being the shortest and longest possible game lengths for input N.

The known game with length R_1 corresponds to playing the longest game on input $N - 1$, then performing the unique sequence of combine moves to achieve the Zeckendorf decomposition of N; by Theorems 3 and 4, we can take the first phase of this game (longest game on input $N - 1$) as playing the longest game on input $F_n - 1$, then playing the longest remaining game (on input $N - 1$). The known game with length L_2 corresponds to playing the shortest game on input $F_n - 1$, then playing the longest remaining game until we achieve the Zeckendorf decomposition of N. By Lemma 1, we can take the second phase of this game (playing the longest remaining game) as playing consistent to Theorem 3 until achieving the Zeckendorf decomposition.

We henceforth denote these games with length R_1 and L_2 by $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively, which we depict as follows; "longest on k" indicates that during this phase, we think of the game as being played on input k, and leaving the appropriate number of instances of $F_1 = 1$ in the first bin fixed.

$\mathcal{G}_1 : \big[(\text{Longest on } F_n - 1) \rightarrow (\text{longest remaining on } N - 1)\big] \rightarrow (\text{combine on } N)$

$\mathcal{G}_2 : (\text{Shortest on } F_n - 1) \rightarrow (\text{longest remaining on } N)$

Denoting the difference between the lengths of the longest and shortest games on input $F_n - 1$ by $\ell(n)$, game $\mathcal{G}_1$ took exactly $\ell(n)$ more moves than game $\mathcal{G}_2$ on input $F_n - 1$. However, $\mathcal{G}_2$ may take longer afterwards to finish the game on input N: we aim to show that the discrepancy in the game lengths after initially playing on input $F_n - 1$ is dominated by $\ell(n)$, from which we conclude that $\mathcal{G}_2$ is no longer than $\mathcal{G}_1$, i.e., that $L_2 \leq R_1$. In particular, it certainly suffices to show that the first two segments of $\mathcal{G}_1$ involve at least as many moves as the first two segments of $\mathcal{G}_2$ to establish the result: this is how we shall proceed.

By Theorem 3 and Lemma 1, we can choose how we would like to play a longest game after playing the game on input $F_n - 1$ (specifically, we can fix an ordering on Type A moves which determines what we play when given multiple Type A moves): until the two games diverge,[4] pursue a longest remaining game by always playing the rightmost Type A move whenever a Type A move is playable. Following this ordering on Type A moves, we study the first move on which games $\mathcal{G}_1$ and $\mathcal{G}_2$ deviate. This move is necessarily either C_1 or C_2 in game $\mathcal{G}_2$ (the move must not have been playable in game $\mathcal{G}_1$, and thus must involve bin 1), and by the ordering on Type A moves established before, bins $2, 3, \ldots, n$ are either 0 or 1 when it is played. We perform casework on which move the two games deviate on.

Case 1: Move is C_1. In game $\mathcal{G}_1$, this configuration can be represented by the vector $(1, x_2, x_3, \ldots, x_n)$, where $x_i \in \{0, 1\}$ for $i \geq 2$ denotes the height of the ith bin (the first entry would be a 2 for game $\mathcal{G}_2$). Let us first study the setting $x_2 = 0$, and consider what happens after game $\mathcal{G}_2$ plays C_1. If $x_3 = 0$, then by Lemma 5 applied to $k = 3$, both games are consistent on bins $4, \ldots, n$ so that game $\mathcal{G}_2$ takes one more move than $\mathcal{G}_1$ to finish. Otherwise (i.e., $x_3 = 1$), Lemma 5 on $k = 2$ yields that game $\mathcal{G}_1$ works strictly over bins $3, \ldots, n$, and Lemma 5 on $k = 1$ yields that game $\mathcal{G}_2$ works strictly over the bins $2, \ldots, n$ (i.e., bin 1 becomes irrelevant). Thus, the resulting setting corresponds exactly to Case 2 over the $n - 1$ bins $2, \ldots, n$; here, we have an upper bound of $n - 1$ for the number of additional moves $\mathcal{G}_2$ takes, for a total upper bound of $(n - 1) + 1 = n$ (with the C_1 in $\mathcal{G}_2$) for the number of additional moves $\mathcal{G}_2$ takes.

Thus, assume $x_2 = 1$, and let position $k + 1$, with $k \geq 2$, denote the first index that is 0. Study the length-k prefix $(1, 1, 1, \ldots, 1)$ (in game $\mathcal{G}_1$; first entry is 2 in

---

[4] We shall assume this does happen, as otherwise the lengths of the second segments of $\mathcal{G}_1$ and $\mathcal{G}_2$ are equal, and thus the inequality $L_2 \leq R_1$ is immediate.

game $\mathcal{G}_2$) with no zero entries; we can exactly describe how the two games proceed when playing according to the longest game paradigm of Theorem 3.

– Game $\mathcal{G}_1$: We play the sequence of moves

$$(C_2 \to S_3 \to S_4 \to \cdots \to S_k) \to (C_2 \to S_3 \to \cdots \to S_{k-2}) \to \cdots$$

continuing similarly until one of the first two bins is empty (so we cannot play C_2). In general, shorten the contiguous 1s down two indices and repeat until one of the first two bins is empty (the bin which ends up empty depends on the parity of k).

– Game $\mathcal{G}_2$: We play the sequence of moves

$$(C_1) \to (S_2) \to (S_3 \to C_1) \to (S_4 \to S_2) \to \cdots$$

continuing similarly until we play all splitting moves S_j for $j \leq k$. In general, when first playing S_j, we play the sequence $S_j \to S_{j-2} \to \cdots$ (final move is S_2 or C_1, depending on parity of j), and this continues until position k.

After these subroutines, it is straightforward to confirm that bin k is empty, bins $k + 1, \ldots, n - 1$ are all 0 or 1 with heights agreeing between $\mathcal{G}_1$ and $\mathcal{G}_2$, and no playable moves involving bins $1, \ldots, k$ exist. Thus, both games proceed strictly over bins $k + 1$ to $n - 1$ by Lemma 5, and perform the same sequence of moves; we need only compute the difference in the lengths of these two subroutines on this prefix of length k.

– The length of the sequence of moves in the game of length L_2 is

$$\begin{cases} 2\sum_{j=1}^{(k-1)/2} j + \frac{k+1}{2} & k \text{ odd,} \\ 2\sum_{j=1}^{k/2} j & k \text{ even.} \end{cases}$$

– The length of the sequence of moves in the game of length R_1 is

$$\begin{cases} \sum_{j=1}^{(k-1)/2}(k - 2j + 1) = 2\sum_{j=1}^{(k-1)/2} j & k \text{ odd,} \\ \sum_{j=1}^{k/2}(k - 2j + 1) = \sum_{j=1}^{k/2}(2j - 1) = 2(\sum_{j=1}^{k/2} j) - \frac{k}{2} & k \text{ even.} \end{cases}$$

We can thus study this difference exactly: the difference is given by

$$\begin{cases} \left[2\sum_{j=1}^{(k-1)/2} j + \frac{k+1}{2}\right] - 2\sum_{j=1}^{(k-1)/2} j = \frac{k+1}{2} = \lceil k/2 \rceil & k \text{ odd} \\ 2\sum_{j=1}^{k/2} j - \sum_{j=1}^{k/2}(2j - 1) = \frac{k}{2} & k \text{ even} \end{cases}$$

so in general, the difference is bounded by $\lceil n/2 \rceil \leq n$.

Case 2: Move is C_2. In game $\mathcal{G}_1$, this configuration can be represented by the vector $(0, x_2, x_3, \ldots, x_n)$, where $x_i \in \{0, 1\}$ for $i \geq 2$ denotes the height of the ith bin (the

first entry would be a 1 for game $\mathcal{G}_2$) and $x_2 = 1$ (since otherwise C_2 cannot be played in game $\mathcal{G}_2$). Let us first study the setting $x_3 = 0$, and consider what happens after game $\mathcal{G}_2$ plays C_2. Applying Lemma 5 on $k = 3$ for $\mathcal{G}_1$ and $k = 2$ for $\mathcal{G}_2$ yield that both games work strictly over bins $3, \ldots, n$, and this reduces to the same setting on the suffix of bins $3, \ldots, n$. Say we reduce the problem to a suffix with length reduced by 2 in this manner m times, so we study the case where $x_3 = 1$ over $n - 2m$ bins: by extracting the bound in the following argument (i.e., the $x_3 = 1$ case) for the number of additional moves $\mathcal{G}_2$ takes, this yields a bound of $m + (n - 2m) \leq n$.

Thus, assume $x_3 = 1$, and let position $k + 1$, with $k \geq 3$, denote the first index that is 0. Study the length-k prefix $(0, 1, 1, \ldots, 1)$ (in game $\mathcal{G}_1$; first entry is 1 in game $\mathcal{G}_2$): we can explicitly describe how the two games necessarily proceed when playing according to the longest game paradigm.

- Game $\mathcal{G}_1$: We play the sequence of moves

$$(C_3 \to S_4 \to S_5 \to \cdots \to S_k) \to (C_3 \to S_4 \to \cdots \to S_{k-2}) \to \cdots$$

continuing similarly until either bin 2 or bin 3 is empty (so we cannot play C_3). In general, shorten the contiguous 1s down two indices and repeat until either bin 2 or bin 3 is empty (the bin which ends up empty depends on the parity of k).
- Game $\mathcal{G}_2$: This is exactly the same as game $\mathcal{G}_1$ in Case 1.

After these subroutines, it is straightforward to confirm that bin k is empty, bins $k + 1, \ldots, n - 1$ are all 0 or 1 with heights agreeing between $\mathcal{G}_1$ and $\mathcal{G}_2$, and no playable moves involving bins $1, \ldots, k$ exist. Thus, both games proceed strictly over bins $k + 1$ to $n - 1$ by Lemma 5, and perform the same sequence of moves; we need only compute the difference in the lengths of these two subroutines on this prefix of length k.

The length of the sequence of moves in game $\mathcal{G}_2$ was computed in Case 1, while the length of the sequence of moves in game $\mathcal{G}_1$ is

$$\begin{cases} \sum_{j=1}^{(k-1)/2}(k - 2j) = \sum_{j=1}^{(k-1)/2}(2j - 1) = 2(\sum_{j=1}^{(k-1)/2} j) - \frac{k-1}{2} & k \text{ odd} \\ \sum_{j=1}^{(k-2)/2}(k - 2j) = \sum_{j=1}^{(k-2)/2}(2j - 1) = 2(\sum_{j=1}^{(k-2)/2} j) - \frac{k-2}{2} & k \text{ even.} \end{cases}$$

so the difference is given by

$$\begin{cases} 2\sum_{j=1}^{(k-1)/2} j - \left[2(\sum_{j=1}^{(k-1)/2} j) - \frac{k-1}{2}\right] = \frac{k-1}{2} = \lfloor k/2 \rfloor & k \text{ odd} \\ \left[2(\sum_{j=1}^{k/2} j) - \frac{k}{2}\right] - \left[2(\sum_{j=1}^{(k-2)/2} j) - \frac{k-2}{2}\right] = k - 1 & k \text{ even} \end{cases}$$

so in general, the difference is bounded by $\max\{\lfloor k/2 \rfloor, k - 1\} \leq n - 1 \leq n$.

We observe that in both cases, the difference in the lengths of the second segments of these games is bounded by n. We now show that the difference $\ell(n)$ between the lengths of the longest and shortest games on input $F_n - 1$ is at least n for all $n \geq 6$. One can confirm, by playing a longest game according to Theorem 3 on input $N = F_6 - 1 = 12$, that $\ell(6) = 17 - (12 - Z(12)) = 17 - 9 \geq 6$; it is similarly easy to confirm that $\ell(4), \ell(5) \geq 1$. Since we have $F_{n+1} - 1 = (F_n - 1) + F_{n-1} \geq (F_n - 1) + (F_{n-1} - 1)$, one can pursue a game on input $F_{n+1} - 1$ by first playing a game on input $F_n - 1$, then a game on input $F_{n-1} - 1$, and finally performing some fixed sequence of moves to completion. By combining respective shortest games and longest games on input $F_n - 1$ and input $F_{n-1} - 1$, we observe that $\ell(n + 1) - \ell(n) \geq 1$ for any $n \geq 6$, and thus $\ell(n) \geq n$ for all $n \geq 6$ (recall $\ell(6) \geq 6$).

Therefore, we have that for all $N \geq F_6 = 13$,

$$\begin{aligned} R_1 - L_2 \geq &\ell(n) - \{(\text{len. longest remaining on } N) \\ &- (\text{len. longest remaining on } N - 1)\} \\ \geq &\ell(n) - n \geq \ell(n) - \ell(n) = 0 \end{aligned}$$

from which we conclude that $L_2 \leq R_1$.

4 Winning Odds in the Limit $N \to \infty$

We dedicate this section to establishing Theorem 9, which follows as an immediate corollary of Theorem 8: recall that Theorem 8 is given as follows.

Theorem 8 *For any integer $Z \geq 1$ and $z \in \{0, 1, \ldots, M - 1\}$, we have that*

$$\lim_{N \to \infty} \mu_N(\text{Game length} \equiv z \bmod Z) = \lim_{N \to \infty} \mathbb{P}_N(\text{Game length} \equiv z \bmod Z) = \frac{1}{Z}.$$

The case $Z = 2$ gives Theorem 9.

Theorem 9 *For the two-player Zeckendorf game, in the limit $N \to \infty$ under both probability measures μ_N and $\mathbb{P}_N$, Player 1 and Player 2 are equally likely to win. Explicitly,*

$$\lim_{N \to \infty} \mu_N(\text{Player 1 wins}) = \lim_{N \to \infty} \mu_N(\text{Player 2 wins}) = \frac{1}{2},$$
$$\lim_{N \to \infty} \mathbb{P}_N(\text{Player 1 wins}) = \lim_{N \to \infty} \mathbb{P}_N(\text{Player 2 wins}) = \frac{1}{2}.$$

4.1 Overview

We provide a brief sketch of the proof of Theorem 9. We specifically focus on the number of $(1, k)$-prefixes which occur over the course of a game, recalling that whenever applicable, the sequence of moves (S_k, C_{k-1}) has the same effect on the game as C_k. We partition the collection of games Ω_N into subsets of games which differ only via such $(1, k)$-prefixes, describe the conditional distribution induced by any such subset (under both μ_N and $\mathbb{P}_N$), and argue that in the limit of infinite input, the sizes of the subsets induced by this partition grow so fast that the probability of Player 1 or Player 2 winning can be reduced to the outcome of a binomial random variable with exploding variance.

4.2 Partitioning the Collection of Possible Games

As observed in Definition 2, for $k \geq 2$, the sequence[5] (S_k, C_{k-1}) is a $(1, k)$-prefix. We establish some notation to use later: define $\mathcal{R}_N \in \mathcal{F}_N$ to be the collection[6] of all Zeckendorf games on input N such that (for all $k \geq 2$) any instance of the sequence (S_k, C_{k-1}) is immediately preceded by S_{k+1}; in other words, $\mathcal{R}_N$ is the collection of all Zeckendorf games on input N such that there are no $(1, k)$-prefixes for any $k \geq 2$. We can express this collection as

$$\mathcal{R}_N = \{(M_1, \ldots, M_\lambda) \in \Omega_N : (M_i, M_{i+1}) = (S_k, C_{k-1}) \implies M_{i-1} = S_{k+1}$$
$$\forall \, i \in [\lambda - 1], k \geq 2\}.$$

For a game $\mathcal{R} = (M_1, M_2, \ldots, M_\lambda) \in \mathcal{R}_N$, construct the subset of indices $\mathcal{I}_N(\mathcal{R})$ to denote all combine moves in $\mathcal{R}$, not involved in a sequence (S_k, C_{k-1}) for $k \geq 2$, for which the latter bin has height at least 2 (i.e., combine moves C_k for $k \geq 2$ replaceable with a $(1, k)$-prefix):

$$\mathcal{I}_N(\mathcal{R}) = \{i \in [\lambda] : M_i = C_k, \ M_{i-1} \neq S_{k+1}, \ h_k(i) \geq 2 \text{ for some } k \geq 2\}.$$

Now, construct the *formal* sequence of moves $\mathcal{M}(\mathcal{R}) = (\tilde{M}_1, \tilde{M}_2, \ldots, \tilde{M}_\lambda)$ by replacing M_i by a symbol $\mathcal{E}_k$ (the subscript being the corresponding $k \geq 2$) for all $i \in \mathcal{I}_N(\mathcal{R})$; call $\mathcal{M}(\mathcal{R})$ the **base sequence** of $\mathcal{R} \in \mathcal{R}_N$. Let $\mathcal{A}_N(\mathcal{R}) \in \mathcal{F}_N$ denote the collection of all Zeckendorf games resulting from replacing each instance of $\mathcal{E}_k$ in $\mathcal{M}$ by either C_k or the 1-prefix (S_k, C_{k-1}) (for every $k \geq 2$). We establish the following important result, which makes it clear why we have pursued this construction in the manner that we did.

---

[5] We deviate from notation earlier in the paper and write move sequences as tuples.

[6] We elect to use the notation $\mathcal{R}_N$ as we think of these games on input N as *representatives* of the corresponding classes $\mathcal{A}_N(\mathcal{R})$ that we define later in this discussion.

Lemma 6 *The sets* $\{\mathcal{A}_N(\mathcal{R}) : \mathcal{R} \in \mathcal{R}_N\}$ *partition* Ω_N.

Proof We first show that the sets $\mathcal{A}_N(\mathcal{R})$ are disjoint. Take distinct $\mathcal{R}_1 = (M_1^1, \ldots, M_\lambda^1)$, $\mathcal{R}_2 = (M_1^2, \ldots, M_{\lambda'}^2) \in \mathcal{R}_N$ with base sequences $\mathcal{M}_1(\mathcal{R}_1) = (\tilde{M}_1^1, \tilde{M}_2^1, \ldots, \tilde{M}_\lambda^1)$ and $\mathcal{M}_2(\mathcal{R}_2) = (\tilde{M}_1^2, \tilde{M}_2^2, \ldots, \tilde{M}_{\lambda'}^2)$, respectively. Since $\mathcal{R}_1 \neq \mathcal{R}_2$, take smallest i for which $M_i^1 \neq M_i^2 \implies \tilde{M}_i^1 \neq \tilde{M}_i^2$, and study the construction of any two games $\mathcal{G}_1 \in \mathcal{A}_N(\mathcal{R}_1)$ and $\mathcal{G}_2 \in \mathcal{A}_N(\mathcal{R}_2)$ as described above. If $\mathcal{G}_1$ and $\mathcal{G}_2$ are consistent prior to $\tilde{M}_i^1$ and $\tilde{M}_i^2$ (i.e., for $k \geq 2$, all $\mathcal{E}_k$ are replaced by the same choice of C_k or (S_k, C_{k-1})), which is certainly the only way the two games remain equal up to this point of their construction, then we necessarily produce a difference in the two games on $\tilde{M}_i^1$ and $\tilde{M}_i^2$ if $M_i^1 = \tilde{M}_i^1$ and $M_i^2 = \tilde{M}_i^2$ ($M_i^1 \neq M_i^2$) or $M_i^1 \neq \tilde{M}_i^1$ and $M_i^2 \neq \tilde{M}_i^2$ ($\tilde{M}_i^1 = \mathcal{E}_{k_1}, \tilde{M}_i^2 = \mathcal{E}_{k_2}, k_1 \neq k_2$). If $M_i^1 = \tilde{M}_i^1$ and $M_i^2 \neq \tilde{M}_i^2$ (say $\tilde{M}_i^2 = \mathcal{E}_k$, so necessarily $M_i^2 = C_k$ and $M_{i-1}^2 \neq S_{k+1}$ for some $k \geq 2$), then game $\mathcal{G}_2$ is next filled with either $C_k \neq M_i^1$ ($M_i^1 = M_i^2$ if $C_k = M_i^1$) or (S_k, C_{k-1}). In the latter case, if the two games were equal after this, then necessarily $(M_{i-1}^1, M_i^1, M_{i+1}^1) = (S_{k+1}, S_k, C_{k-1})$ (by definition of $\mathcal{I}_N(\mathcal{R}_1)$ and $\mathcal{R}_N$, M_{i+1}^1 and M_{i-1}^1 follow after establishing $M_i^1 = S_k$), contradicting $M_{i-1}^2 \neq S_{k+1}$ (since $M_{i-1}^1 = M_{i-1}^2$ and $M_i^2 = C_k$, $\tilde{M}_i^2 = \mathcal{E}_k$).

We therefore conclude that $\mathcal{A}_N(\mathcal{R}_1) \cap \mathcal{A}_N(\mathcal{R}_2) = \emptyset$, i.e., the sets $\mathcal{A}_N(\mathcal{R})$ for $\mathcal{R} \in \mathcal{R}_N$ are disjoint; it remains to show that any game $\mathcal{G} \in \Omega_N$ is in some set $\mathcal{A}_N(\mathcal{R})$. For $\mathcal{G} \in \Omega_N$, let $\mathcal{R}$ be the game resulting from replacing every instance of the sequence (M, S_k, C_{k-1}), $M \neq S_{k+1}$ in game $\mathcal{G}$ by the sequence (M, C_k) (for $k \geq 2$). The resulting game $\mathcal{R}$ is such that $(M_i, M_{i+1}) = (S_k, C_{k-1}) \implies M_{i-1} = S_{k+1}$ for all $i \in [\lambda]$ and $k \geq 2$: any sequence $(M_{i-1}, M_i, M_{i+1}) = (M, S_k, C_{k-1})$ with $M \neq S_{k+1}$ in $\mathcal{R}$ necessarily results from having replaced (S_{k-1}, C_{k-2}) for C_{k-1} in game $\mathcal{G}$ (as (S_k, C_{k-1}) would have been replaced by S_{k+1} otherwise), but we know this does not occur by the description above, so we indeed have $\mathcal{R} \in \mathcal{R}_N$. Also, $\mathcal{G} \in \mathcal{A}_N(\mathcal{R})$: we can reverse all the replacements $(S_k, C_{k-1}) \leftrightarrow C_k$ made in achieving $\mathcal{R}$ from $\mathcal{G}$, since the resulting C_k moves correspond to $\mathcal{E}_k$ (for some $k \geq 2$) in the base sequence $\mathcal{M}(\mathcal{R})$ as the preceding move is not S_{k+1}.

Therefore, applying Lemma 6 and the law of total probability,

$$\mu_N(\text{length is } z \bmod Z) = \sum_{\mathcal{R} \in \mathcal{R}_N} \mu_N(\text{length is } z \bmod Z \mid \mathcal{A}_N(\mathcal{R})) \, \mu_N(\mathcal{A}_N(\mathcal{R}))$$

$$\mathbb{P}_N(\text{length is } z \bmod Z) = \sum_{\mathcal{R} \in \mathcal{R}_N} \mathbb{P}_N(\text{length is } z \bmod Z \mid \mathcal{A}_N(\mathcal{R})) \, \mathbb{P}_N(\mathcal{A}_N(\mathcal{R}))$$

$$\tag{1}$$

so we can reduce proving Theorem 8 to establishing that the conditional probabilities for $\mathcal{R} \in \mathcal{R}_N$, with respect to both measures, overwhelmingly tend to $1/2$ in the limit.

4.3 Analysis

Define the random variable $m_N : \Omega_N \to \mathbb{N}$ first on $\mathcal{R} \in \mathcal{R}_N$ by $m_N(\mathcal{R}) = |\mathcal{I}_N(\mathcal{R})|$, denoting the number of terms $\mathcal{E}_k$ (for $k \geq 2$) in the base sequence of $\mathcal{R}$, then lift to arbitrary $\mathcal{G} \in \Omega_N$ by letting $m_N(\mathcal{G}) = m_N(\mathcal{R})$ for the unique $\mathcal{R} \in \mathcal{R}_N$ such that $\mathcal{G} \in \mathcal{A}_N(\mathcal{R})$ (see Lemma 6).[7] Fix $\mathcal{R} \in \mathcal{R}_N$, and observe that on the event $\mathcal{A}_N(\mathcal{R})$, $m_N(\mathcal{G})$ for $\mathcal{G} \in \mathcal{A}_N(\mathcal{R})$ is the fixed constant $|\mathcal{I}_N(\mathcal{R})|$. Now construct corresponding Bernoulli random variables $X_1^{\mathcal{R}}, X_2^{\mathcal{R}}, \ldots, X_{|m_N(\mathcal{R})|}^{\mathcal{R}} : \Omega_N \to \{0, 1\}$ for each of the instances of terms of form $\mathcal{E}_k$ (for $k \geq 2$) in the base sequence of $\mathcal{R}$: here, $X_i^{\mathcal{R}}(\mathcal{G}) = 1$ if and only if $\mathcal{G} \in \mathcal{A}_N(\mathcal{R})$ and $\mathcal{G}$ is achieved by the ith instance of $\mathcal{E}_k$ in the base sequence $\mathcal{R}$ being the $(1, k)$-prefix (S_k, C_{k-1}). Say we replaced this $\mathcal{E}_k$ with (S_k, C_{k-1}), and let n_i denote the number of playable moves in the game available in the turn immediately after playing S_k. We can say more about the random variables $X_i^{\mathcal{R}}$ defined here.

Lemma 7 *Fix $\mathcal{R} \in \mathcal{R}_N$. Define the random variables $X_1^{\mathcal{R}}, X_2^{\mathcal{R}}, \ldots, X_{|m_N(\mathcal{R})|}^{\mathcal{R}}$ as above. When conditioned on $\mathcal{A}_N(\mathcal{R})$, the variables $X_1^{\mathcal{R}}, X_2^{\mathcal{R}}, \ldots, X_{|m_N(\mathcal{R})|}^{\mathcal{R}}$ are independent Bernoulli random variables with each $X_i^{\mathcal{R}}$ having parameter $1/2$ under the measure μ_N, and parameter $p_i = \frac{1}{1+n_i}$ under the measure $\mathbb{P}_N$. Explicitly,*

$$\mu_N\left(X_i^{\mathcal{R}} = 1 \mid \mathcal{A}_N(\mathcal{R})\right) = \frac{1}{2}, \qquad \mathbb{P}_N\left(X_i^{\mathcal{R}} = 1 \mid \mathcal{A}_N(\mathcal{R})\right) = \frac{1}{1 + n_i}.$$

Proof Fix a random variable $X_i^{\mathcal{R}}$, and observe that for every particular setting of all other $\mathcal{E}_k$ terms in $\mathcal{R}$, there exist exactly two games in $\mathcal{A}_N(\mathcal{R})$ faithful to this setting, corresponding to choosing C_k and the $(1, k)$-prefix (S_k, C_{k-1}) for the ith such $\mathcal{E}_k$. It follows immediately that under the uniform measure μ_N, we indeed have $\mu_N\left(X_i^{\mathcal{R}} = 1 \mid \mathcal{A}_N(\mathcal{R})\right) = \frac{1}{2}$, since in particular there exists a bijection between games in the subset $\mathcal{A}_N(\mathcal{R})$ with $X_i^{\mathcal{R}} = 0$ and $X_i^{\mathcal{R}} = 1$ and all games in Ω_N are given equal probability under the measure μ_N.

Under the probability measure $\mathbb{P}_N$, it is straightforward to observe that the game replacing the ith instance of $\mathcal{E}_k$ by the $(1, k)$-prefix (S_k, C_{k-1}) requires an additional decision with probability $\frac{1}{n_i}$ of yielding the desired C_{k-1}, and it thus follows that the parameter p_i of the Bernoulli random variable $X_i^{\mathcal{R}}$ is given by

$$\frac{p_i}{1 - p_i} = \frac{1}{n_i} \implies p_i = \frac{1}{1 + n_i}.$$

To establish independence, it suffices to show that for any subset $S \subseteq [m_N(\mathcal{R})]$, we have the identity

---

[7] In particular, $m_N(\mathcal{R}) = \log_2(\mathcal{A}_N(\mathcal{R}))$.

$$\mathbb{P}_N \left(X_i^{\mathcal{R}} = 1 \text{ iff } i \in S \mid \mathcal{A}_N(\mathcal{R}) \right) = \prod_{i \in S} \mathbb{P}_N \left(X_i^{\mathcal{R}} = 1 \mid \mathcal{A}_N(\mathcal{R}) \right)$$

$$\cdot \prod_{j \notin S} \mathbb{P}_N \left(X_j^{\mathcal{R}} = 0 \mid \mathcal{A}_N(\mathcal{R}) \right)$$

$$= \prod_{i \in S} p_i \cdot \prod_{j \notin S} (1 - p_j).$$

We can relate the conditional probabilities $\{\mathbb{P}_N \left(X_i^{\mathcal{R}} = 1 \text{ iff } i \in S \; \mathcal{A}_N(\mathcal{R}) \right) : S \in [m_N(\mathcal{R})]\}$ whenever $S_1 = S_2 \cup \{j\}$ (for $j \in [m_N(\mathcal{R})]$): all choices for each term $\mathcal{E}_k$ but one are consistent (namely, $X_j^{\mathcal{R}} = 1$ for the numerator in the following), and thus

$$\frac{\mathbb{P}_N \left(X_i^{\mathcal{R}} = 1 \text{ iff } i \in S_1 \mid \mathcal{A}_N(\mathcal{R}) \right)}{\mathbb{P}_N \left(X_i^{\mathcal{R}} = 1 \text{ iff } i \in S_2 \mid \mathcal{A}_N(\mathcal{R}) \right)} = \frac{1}{n_j} = \frac{p_j}{1 - p_j} = \frac{\prod_{i \in S_1} p_i \cdot \prod_{j \notin S_1} (1 - p_j)}{\prod_{i \in S_2} p_i \cdot \prod_{j \notin S_2} (1 - p_j)}.$$

Now, since we have the identity

$$\sum_{S \subseteq [m_N(\mathcal{R})]} \mathbb{P}_N \left(X_i^{\mathcal{R}} = 1 \text{ iff } i \in S \mid \mathcal{A}_N(\mathcal{R}) \right) = 1 = \prod_{i=1}^{m_N(\mathcal{R})} (p_i + (1 - p_i))$$

$$= \sum_{S \subseteq [m_N(\mathcal{R})]} \left(\prod_{i \in S} p_i \cdot \prod_{j \notin S} (1 - p_j) \right)$$

and quotients between summands corresponding to two sets differing by one element are the same, the summands on both sides for any subset $S \subseteq [m_N(\mathcal{R})]$ are necessarily equal. More specifically, letting $p_1 = \mathbb{P}_N \left(X_i^{\mathcal{R}} = 1 \text{ iff } i \in \emptyset \mid \mathcal{A}_N(\mathcal{R}) \right)$ and $p_2 = \prod_{i \in \emptyset} p_i \cdot \prod_{j \notin \emptyset} (1 - p_j)$, by incrementally including elements to some $S \subseteq [m_N(\mathcal{R})]$ we can write the corresponding summands on the left and right hand sides as the same multiple of p_1 and p_2, respectively. This reduces to $p_1 = p_2$, and thus summands corresponding to the same S are equal. Thus, we have the desired identity for any subset $S \subseteq [m_N(\mathcal{R})]$. $\qquad \blacksquare$

Lemma 7 yields the following easy observation.

Corollary 1 *Say $F_n \leq N < F_{n+1}$, fix some $\mathcal{R} \in \mathcal{R}_N$, and define the random variables $X_1^{\mathcal{R}}, X_2^{\mathcal{R}}, \ldots, X_{m_N(\mathcal{R})}^{\mathcal{R}}$ as above. Under the uniform measure μ_N, whenever $\mathcal{R} \in \mathcal{R}_N$ is such that $m_N(\mathcal{R}) > 0$, μ_N (Player 1 wins $\mid \mathcal{A}_N(\mathcal{R})) = 1/2$. Under the probability measure $\mathbb{P}_N$, when conditioned on the event $\mathcal{A}_N(\mathcal{R})$, there are at most $2n$ distinct values of the parameters $p_i = \frac{1}{1+n_i}$ amongst the random variables $X_i^{\mathcal{R}}$, and for any such i, $\frac{1}{2n-2} \leq p_i \leq 1/2$.*

Proof If $m_N(\mathcal{R}) > 0$, then the statement μ_N (Player 1 wins $\mid \mathcal{A}_N(\mathcal{R})) = 1/2$ follows immediately by applying the law of total probability on all settings of the Bernoulli random variables $(X_2^{\mathcal{R}}, \ldots, X_{m_N(\mathcal{R})}^{\mathcal{R}}) \in \{0, 1\}^{m_N(\mathcal{R})-1}$, namely since this further conditioning always yields a conditional probability of $1/2$ (see the proof of

Lemma 7). At any point of a game $\mathcal{G} \in \Omega_N$ satisfying $F_n \leq N < F_{n+1}$, there are at most $2n - 3$ (and thus certainly at most $2n$) playable moves (and thus at most $2n - 3$ distinct values of n_i, and thus p_i): the combine moves $C_1, C_2, \ldots, C_{n-1}$ and the splitting moves $S_2, S_3, \ldots, S_{n-1}$. In particular, since $n_i \in [2n - 3]$ for all $i \in [m_N(\mathcal{R})]$, it follows that $\frac{1}{1+(2n-3)} \leq p_i \leq \frac{1}{1+1}$, i.e., $\frac{1}{2n-2} \leq p_i \leq \frac{1}{2}$.

We will also make use of the following lemma.

Lemma 8 *Say $F_n \leq N < F_{n+1}$.*[8] *For any $c \in (0, \varphi)$ (with φ denoting the golden ratio),*

$$\lim_{N \to \infty} \mu_N \left(m_N(\mathcal{G}) \geq c^n \right) = 1, \qquad \lim_{N \to \infty} \mathbb{P}_N \left(m_N(\mathcal{G}) \geq c^n \right) = 1.$$

Proof Fix $c \in (0, \varphi)$, and consider a game $\mathcal{G} \in \Omega_N$: for the representative $\mathcal{R} \in \mathcal{R}_N$ such that $\mathcal{G} \in \mathcal{A}_N(\mathcal{R})$, every occurrence of the sequence (C_1, C_1, C_2) corresponds to an element of $\mathcal{I}_N(\mathcal{R})$ (specifically the move C_2, as it is not preceded by S_3, and the latter two moves are maintained in $\mathcal{R}$; see the proof of Lemma 6), i.e., $m_N(\mathcal{G}) = |\mathcal{I}_N(\mathcal{R})|$ is at least the number of occurrences of the sequence (C_1, C_1, C_2) in $\mathcal{G}$. Letting $\mathcal{N}(\mathcal{G})$ denote the number of occurrences of the sequence (C_1, C_1, C_2) in the game $\mathcal{G}$, it thus suffices to show that

$$\lim_{N \to \infty} \mu_N \left(\mathcal{N}(\mathcal{G}) \geq c^n \right) = 1, \qquad \lim_{N \to \infty} \mathbb{P}_N \left(\mathcal{N}(\mathcal{G}) \geq c^n \right) = 1.$$

Proceed studying the probability measure $\mathbb{P}_N$; the analysis carries over exactly[9] when $\mathbb{P}_N$ is replaced with μ_N. By Corollary 1, at any move where $h_1 \geq 5$, the probability of achieving the sequence (C_1, C_1, C_2) is at least $\left(\frac{1}{2n} \right)^3 = \frac{1}{8n^3}$, and any sequence of three moves can decrease the height of bin 1 by at most 6 (via three consecutive C_1 moves); observe that the occurrence of C_2 in a sequence (C_1, C_1, C_2) within a game $\mathcal{G}$ necessarily corresponds to an instance of the delimiter $\mathcal{E}_2$ in the base sequence of $\mathcal{G}$ (e.g. see the proof of Lemma 6). Thus, study the sequences of moves in game $\mathcal{G}$ (which necessarily exist by the preceding discussion) given by the triples

$$(M_1, M_2, M_3), (M_4, M_5, M_6), \ldots, (M_{3\lfloor N/6 \rfloor - 2}, M_{3\lfloor N/6 \rfloor - 1}, M_{3\lfloor N/6 \rfloor})$$

each of which independently takes on the value (C_1, C_1, C_2) with probability at least $\frac{1}{8n^3}$. Thus, letting $Y_N \stackrel{d}{=} \mathrm{Bin} \left(\lfloor N/6 \rfloor, p = \frac{1}{8n^3} \right)$, we have that

$$\mathbb{P} \left(Y_N \geq c^n \right) \leq \mathbb{P}_N \left(\mathcal{N}(\mathcal{G}) \geq c^n \right)$$

from which $\mathbb{E}[Y_N] = \frac{\lfloor N/6 \rfloor}{8n^3}$ together with a Chernoff bound with $\delta = \frac{1}{2}$ yields for N large,

---

[8] It is perhaps more appropriate to think of n as a function $n(N)$.

[9] Certainly, the analysis is much looser than necessary when $\mathbb{P}_N$ is replaced with μ_N.

$$\mathbb{P}\left(Y_N \le c^n\right) \le \mathbb{P}\left(Y_N \le \frac{\lfloor N/6 \rfloor}{16n^3}\right) \le \exp\left(-\frac{\lfloor N/6 \rfloor}{32n^3}\right) \xrightarrow{N\to\infty} 0$$

since it is straightforward to verify that $c^n \le \frac{\lfloor N/6 \rfloor}{16n^3}$ for large N (e.g. use Binet).

Corollary 2 *Lemma 8 immediately yields that for any $d \in \mathbb{N}$,*

$$\lim_{N\to\infty} \mu_N\left(m_N(\mathcal{G}) \ge n^d\right) = 1, \qquad \lim_{N\to\infty} \mathbb{P}_N\left(m_N(\mathcal{G}) \ge n^d\right) = 1$$

i.e., with probability approaching 1 as $N \to \infty$, the number of decisions made to achieve $\mathcal{G}$ from the set $\mathcal{A}_N(\mathcal{R})$ containing it is superpolynomial in n. This is more convenient for later.

The proof of Lemma 8 yields the following observations on the likeliest collection $\mathcal{A}_N(\mathcal{R})$ and the number of representative games, which will be referenced again in Sect. 5.

Corollary 3 *The probability of the likeliest collection $\mathcal{A}_N(\mathcal{R})$ vanishes as $N \to \infty$, i.e.,*

$$\lim_{N\to\infty} \max_{\mathcal{G}\in\Omega_N} \mu_N(\mathcal{A}_N(\mathcal{G})) = \lim_{N\to\infty} \max_{\mathcal{G}\in\Omega_N} \mathbb{P}_N(\mathcal{A}_N(\mathcal{G})) = 0$$

and the number of representatives satisfies $|\mathcal{R}_N| \xrightarrow{N\to\infty} \infty$.

Proof Again, proceed on the probability measure $\mathbb{P}_N$, as the analysis carries over exactly for μ_N. The proof of Lemma 2 (taking $d = 1$) determines there are at least $m \ge n$ instances of (C_1, C_1, C_2) in the first $\lfloor N/6 \rfloor$ moves of a random game $\mathcal{G} \in \Omega_N$, with probability tending to 1 in the limit $N \to \infty$: we proceed studying such a game $\mathcal{G}$. Writing $\mathcal{G} \in \mathcal{A}_N(\mathcal{R})$, every latter instance of C_1 in one of these m sequences is in[10] any game in $\mathcal{A}_N(\mathcal{R})$, and the move C_2 could have been played instead of this instance of C_1. For $i = 1, \ldots, m$, denote $\mathcal{B}_i^N(\mathcal{R}) \subseteq \Omega_N$ to be those games consistent with a game in $\mathcal{A}_N(\mathcal{R})$ up to the ith such instance of C_1, but for which the move C_2 is played instead. It is immediately observed that the sets $\mathcal{B}_i^N(\mathcal{R})$ are disjoint and that $\mathbb{P}_N(\mathcal{A}_N(\mathcal{R})) \le \mathbb{P}_N(\mathcal{B}_i^N(\mathcal{R}))$ for each $i = 1, \ldots, m$, so we have

$$n \cdot \mathbb{P}_N(\mathcal{A}_N(\mathcal{R})) \le m \cdot \mathbb{P}_N(\mathcal{A}_N(\mathcal{R})) \le \sum_{i=1}^{m} \mathbb{P}_N(\mathcal{B}_i^N(\mathcal{R})) \le 1$$

$$\implies \mathbb{P}_N(\mathcal{A}_N(\mathcal{R})) \le \frac{1}{n} \xrightarrow{N\to\infty} 0$$

---

[10] This is in the sense that this instance of C_1 is not compressed in achieving $\mathcal{R}$ from $\mathcal{G}$ (see the proof of Lemma 6), and is not replaced with a term of form $\mathcal{E}_k$ in achieving the base sequence $\mathcal{M}(\mathcal{R})$.

so we have $\lim_{N\to\infty} \max_{\mathcal{G}\in\Omega_N} \mathbb{P}_N(\mathcal{A}_N(\mathcal{G})) = 0$, and $|\mathcal{R}_N| \xrightarrow{N\to\infty} \infty$ follows immediately.

Finally, we need the following simple result concerning the behavior of a binomial random variable with sufficiently large variance. Certainly, the case $Z = 2$ in Lemma 9 corresponds to studying the expressions $\mathbb{P}(\mathcal{B} \text{ is odd})$ and $\mathbb{P}(\mathcal{B} \text{ is even})$.

Lemma 9 *Consider a binomial random variable $\mathcal{B} = Bin(m, p)$. For any values of $\epsilon > 0$ and $Z \in \mathbb{N}$, there exists a constant $N(\epsilon, Z)$ such that if $var(\mathcal{B}) = mp(1 - p) \geq N(\epsilon, Z)$ (i.e., if the variance of $\mathcal{B}$ is sufficiently large), then for any $z \in \{0, 1, \ldots, Z - 1\}$,*

$$\left| \mathbb{P}(\mathcal{B} \equiv z \bmod Z) - \frac{1}{Z} \right| \leq \epsilon.$$

Proof The value $\mathbb{P}(\mathcal{B} = k)$ increases on $k \leq \lfloor (m + 1)p \rfloor$ and decreases on $k \geq \lfloor (m + 1)p \rfloor$, and $\max_{0 \leq k \leq m} \mathbb{P}(\mathcal{B} = k) = \mathbb{P}(\mathcal{B} = \lfloor (m + 1)p \rfloor) \xrightarrow{mp(1-p)\to\infty} 0$ (use Stirling); denoting $p_k = \mathbb{P}(\mathcal{B} = k)$, write $p_0 \leq p_1 \leq \cdots \leq p_{\lfloor (m+1)p \rfloor}$ and $p_{\lfloor (m+1)p \rfloor} \geq p_{\lfloor (m+1)p \rfloor+1} \geq \cdots \geq p_m$. Fix $\epsilon > 0$, and choose $N(\epsilon, Z)$ such that $mp(1 - p) \geq N(\epsilon, Z)$ implies $\mathbb{P}(\mathcal{B} = \lfloor (m + 1)p \rfloor) < \frac{\epsilon}{2}$. Certainly, for any distinct values $z_1 < z_2$ in $\{0, 1, \ldots, Z - 1\}$, the number of terms $p_0, p_1, \ldots, p_{\lfloor (m+1)p \rfloor}$ with indices equal to z_1 modulo Z and equal to z_2 modulo Z is either the same or there exists one more such term corresponding to z_1: we can switch the "dominant modulus" by removing the term of largest index equal to either z_1 or z_2 modulo Z, for which the corresponding term $p_k \leq \frac{\epsilon}{2}$ by choice of the constant $N(\epsilon, Z)$. The analogous statement extends to $p_{\lfloor (m+1)p \rfloor}, p_{\lfloor (m+1)p \rfloor+1}, \ldots, p_m$, so it follows that we can write

$$|\mathbb{P}(\mathcal{B} \equiv z_1 \bmod Z) - \mathbb{P}(\mathcal{B} \equiv z_2 \bmod Z)| \leq \epsilon$$

which immediately yields the desired statement.

We are now ready to proceed with the proof of Theorem 8.

Proof *(Proof of Theorem 8)* Fix an integer $Z \geq 2$ corresponding to the number of players in a Z-player Zeckendorf game, and some value $z \in \{0, 1, \ldots, Z - 1\}$. As in the proof of Lemma 8, we strictly concern ourselves with the probability measure $\mathbb{P}_N$, as the analysis carries over exactly for the uniform measure μ_N (again, it is much looser than necessary for μ_N). For the probability measure $\mathbb{P}_N$, since $m_N(\mathcal{G})$ is fixed at $|\mathcal{I}_N(\mathcal{R})| = m_N(\mathcal{R})$ on any set $\mathcal{A}_N(\mathcal{R})$ for $\mathcal{R} \in \mathcal{R}_N$, for $N \geq N_1(\epsilon)$,

$$\sum_{\substack{\mathcal{R}\in\mathcal{R}_N \\ m_N(\mathcal{R})<2n^3}} \mathbb{P}_N(\mathcal{A}_N(\mathcal{R})) \leq \epsilon. \tag{2}$$

Consider a fixed representative $\mathcal{R} \in \mathcal{R}_N$ for which $m_N(\mathcal{R}) \geq 2n^3$: by Corollary 1, there are at most $2n$ distinct values of the parameters $p_i = \frac{1}{1+n_i}$ of the corresponding random variables $X_1^{\mathcal{R}}, X_2^{\mathcal{R}}, \ldots, X_{m_N(\mathcal{R})}^{\mathcal{R}}$, and thus by the pigeonhole principle, the

number of instances of the value of p_i with largest multiplicity is at least $\frac{2n^3}{2n} = n^2 \xrightarrow{N \to \infty} \infty$; henceforth call this p. Let us say there are $m \geq n^2$ instances of this value of p_i: we can study the sum $\mathcal{B}_\mathcal{R} = \sum_{i=1}^{m} Y_i$, where the Y_i correspond to those random variables $X_i^\mathcal{R}$ with this corresponding success probability p. Furthermore, by Lemma 7, the random variables Y_i are independent when conditioned on the event $\mathcal{A}_N(\mathcal{R})$, so $\mathcal{B}_\mathcal{R} \stackrel{d}{=} \mathrm{Bin}(m, p)$ under this conditional distribution. From the bounds on p from Corollary 1, it follows that the variance of this binomial random variable, when conditioned on the event $\mathcal{A}_N(\mathcal{R})$, has the exploding lower bound

$$\mathrm{var}(\mathcal{B}_\mathcal{R}) = mp(1 - p) \geq \frac{n^2}{2(2n - 2)} \xrightarrow{N \to \infty} \infty$$

i.e., the binomial random variable $\mathcal{B}_\mathcal{R}$ has variance exploding in the limit $N \to \infty$ (for any $\mathcal{R} \in \mathcal{R}_N$). By Lemma 9, the random variable $\mathcal{B}_\mathcal{R}$ will take a value equal to z modulo Z with probability approaching $\frac{1}{Z}$ as $N \to \infty$. Thus, studying the quantities given by $\mathbb{P}_N$ (Game length equals z mod $Z \mid \mathcal{A}_N(\mathcal{R})$), we can further condition (upon the conditioning $\mathcal{A}_N(\mathcal{R})$) on all random variables $X_i^\mathcal{R}$ not corresponding to those Y_i constituting a summand in the binomial random variable $\mathcal{B}_\mathcal{R}$ and appeal to the law of total probability:

$\mathbb{P}_N$ (Game length is z mod $Z \mid \mathcal{A}_N(\mathcal{R})$)

$$= \sum_{\substack{\text{assignments } A \\ \text{to } X_i^\mathcal{R} \neq Y_j \forall j}} \mathbb{P}_N \text{ (Game length is } z \text{ mod } Z \mid \mathcal{A}_N(\mathcal{R}), A) \cdot \mathbb{P}_N \left(A \mid \mathcal{A}_N(\mathcal{R}) \right)$$

where each term $\mathbb{P}_N$ (Game length equals z mod $Z \mid \mathcal{A}_N(\mathcal{R}), A$) is understood as the probability that the length of the game is equal to z mod Z when replacing those $\mathcal{E}_k$ which correspond to the random variables Y_i, while leaving all other moves of the game fixed: this is precisely given by the binomial random variable $\mathcal{B}_\mathcal{R}$ added to some fixed length determined by the setting of the $X_i^\mathcal{R}$ which are not of the form Y_j for some j. Letting $\ell(A)$ denote this length for the assignment A, we can thus write

$$\mathbb{P}_N \text{ (Game length is } z \text{ mod } Z \mid \mathcal{A}_N(\mathcal{R}), A) = \mathbb{P}_N \left(\mathcal{B}_\mathcal{R} + \ell(A) \text{ is odd} \mid \mathcal{A}_N(\mathcal{R}), A \right)$$

By Lemma 7, the random variables Y_i constituting the Bernoulli trials in $\mathcal{B}_\mathcal{R}$ are independent from the random variables $X_i^\mathcal{R}$ that were fixed when conditioned on $\mathcal{A}_N(\mathcal{R})$ and the assignment A, and thus we can apply Lemma 9 to deduce that these conditional probabilities are arbitrarily close to $\frac{1}{Z}$ for $N \geq N_2(\epsilon, Z)$ by the preceding discussion (for sufficiently large $N_2(\epsilon, Z) \in \mathbb{N}$). Importantly, this is uniform over all such terms in the sum, in the sense that we can choose $N_2(\epsilon, Z)$ such that we achieve the same guarantee for any such assignment A of binary values to the random variables $X_i^\mathcal{R}$ not constituting $\mathcal{B}_\mathcal{R}$. Thus, for any $\mathcal{R} \in \mathcal{R}_N$ satisfying $m_N(\mathcal{R}) \geq 2n^3$ and $N \geq N_2(\epsilon, Z)$, we have the bound

$$\left| \mathbb{P}_N \text{ (Game length equals } z \text{ mod } Z \mid \mathcal{A}_N(\mathcal{R})) - \frac{1}{Z} \right| \leq \epsilon.$$

Therefore, for $N \geq \max\{N_1(\epsilon), N_2(\epsilon, Z)\}$, we can take Eqs. (1) and (2) to achieve the bound on the probability $\mathbb{P}_N(\text{Game length equals } z \mod Z)$ given by

$$\mathbb{P}_N(\text{Game length is } z \mod Z)$$

$$= \sum_{\mathcal{R} \in \mathcal{R}_N} \mathbb{P}_N(\text{Game length is } z \mod Z \mid \mathcal{A}_N(\mathcal{R})) \cdot \mathbb{P}_N(\mathcal{A}_N(\mathcal{R}))$$

$$\leq \epsilon + \sum_{\substack{\mathcal{R} \in \mathcal{R}_N \\ m_N(\mathcal{R}) \geq 2n^3}} \mathbb{P}_N(\text{Game length is } z \mod Z \mid \mathcal{A}_N(\mathcal{R})) \cdot \mathbb{P}_N(\mathcal{A}_N(\mathcal{R}))$$

$$\leq \epsilon + \sum_{\substack{\mathcal{R} \in \mathcal{R}_N \\ m_N(\mathcal{R}) \geq 2n^3}} \left(\frac{1}{Z} + \epsilon\right) \cdot \mathbb{P}_N(\mathcal{A}_N(\mathcal{R})) + \epsilon \leq \frac{1}{Z} + 2\epsilon$$

and similarly, we have

$$\mathbb{P}_N(\text{Game length is } z \mod Z)$$

$$= \sum_{\mathcal{R} \in \mathcal{R}_N} \mathbb{P}_N(\text{Game length is } z \mod Z \mid \mathcal{A}_N(\mathcal{R})) \cdot \mathbb{P}_N(\mathcal{A}_N(\mathcal{R}))$$

$$\geq \sum_{\substack{\mathcal{R} \in \mathcal{R}_N \\ m_N(\mathcal{R}) \geq 2n^3}} \mathbb{P}_N(\text{Game length is } z \mod Z \mid \mathcal{A}_N(\mathcal{R})) \cdot \mathbb{P}_N(\mathcal{A}_N(\mathcal{R}))$$

$$\geq \sum_{\substack{\mathcal{R} \in \mathcal{R}_N \\ m_N(\mathcal{R}) \geq 2n^3}} \left(\frac{1}{Z} - \epsilon\right) \cdot \mathbb{P}_N(\mathcal{A}_N(\mathcal{R})) \geq \left(\frac{1}{Z} - \epsilon\right)(1 - \epsilon)$$

so we conclude that for $N \geq \max\{N_1(\epsilon), N_2(\epsilon, Z)\}$,

$$\epsilon^2 - \frac{Z+1}{Z}\epsilon \leq \mathbb{P}_N(\text{Game length equals } z \mod Z) - \frac{1}{Z} \leq 2\epsilon$$

which yields the desired limit by sending $\epsilon \downarrow 0$.

As previously mentioned, the $Z = 2$ case of Theorem 8 immediately yields Theorem 9. $\qquad \blacksquare$

4.4 Extending the Partition on Ω_N

Theorem 9 establishes that if both players advance randomly, the two-player Zeckendorf game is fair in the limit $N \to \infty$: a key stage of the proof involved partitioning Ω_N into collections of games where the games in any given collection only differ via certain interchanges of $(1, k)$-prefixes with the corresponding C_k. We can naturally extend this partition of Ω_N (i.e., strictly larger classes in the partition) to encom-

pass arbitrary (ℓ, k)-prefixes or arbitrary (ℓ, k)-suffixes. Section 5 will establish that for these enlarged partitions of Ω_N, we can achieve an analogue of Conjecture 1 for the resulting sets in the partition, in the sense that with high probability in the limit, the corresponding distribution is "nearly Gaussian," in the sense of vanishing Kolmogorov-Smirnov distance (when mean and variance are normalized to be 0 and 1, respectively) with the standard normal.[11]

We first study the generalization for arbitrary (ℓ, k)-prefixes. Define $\mathcal{R}_N^{\mathcal{P}} \in \mathcal{F}_N$ to be the collection of all Zeckendorf games on input N such that any combine move C_k, for $k \geq 2$, cannot be compressed by a $(1, k)$-prefix (and thus by any prefix of C_k). Explicitly,

$$\mathcal{R}_N^{\mathcal{P}} = \{(M_1, \ldots, M_\lambda) \in \Omega_N : M_i = C_k \implies M_{i-1} \neq S_{k+1}, \ \forall i \in [\lambda], k \geq 2\}.$$

For a game $\mathcal{R} = (M_1, M_2, \ldots, M_\lambda) \in \mathcal{R}_N^{\mathcal{P}}$ with number of moves λ, construct the subset of indices $\mathcal{I}_N(\mathcal{R})$ to denote all combine moves in $\mathcal{R}$:

$$\mathcal{I}_N(\mathcal{R}) = \{i \in [\lambda] : M_i = C_k \text{ for some } k \geq 2\}.$$

Now, construct the *formal* sequence of moves $\mathcal{M}(\mathcal{R}) = (\tilde{M}_1, \tilde{M}_2, \ldots, \tilde{M}_\lambda)$ by replacing M_i by a symbol $\mathcal{E}_k^\ell$ for all $i \in \mathcal{I}_N(\mathcal{R})$: the subscript is the corresponding $k \geq 2$, while $\ell \geq 0$ denotes the longest (ℓ, k)-prefix that $M_i = C_k$ can be expanded into; note in particular that $\mathcal{M}(\mathcal{R})$ contains no combine moves. Call $\mathcal{M}(\mathcal{R})$ the **base sequence** of $\mathcal{R} \in \mathcal{R}_N$, and let $\mathcal{A}_N(\mathcal{R}) \in \mathcal{F}_N$ denote the collection of all Zeckendorf games resulting from replacing each instance of $\mathcal{E}_k^\ell$ in $\mathcal{M}$ by an (l, k)-prefix for some $l \leq \ell$. We establish the following analogue of Lemma 6, which yields a strictly broader partition of Ω_N (in the sense that each set in the partition given by Lemma 5 is a union of sets in the partition given by Lemma 6).

Proposition 5 *The sets $\{\mathcal{A}_N(\mathcal{R}) : \mathcal{R} \in \mathcal{R}_N^{\mathcal{P}}\}$ partition Ω_N.*

Proof Observe that a base sequence $\mathcal{M}(\mathcal{R}) = (\tilde{M}_1, \tilde{M}_2, \ldots, \tilde{M}_\lambda)$ uniquely determines $\mathcal{R} \in \mathcal{R}_N^{\mathcal{P}}$ by its explicit split moves moves and the subscripts of each of its symbols $\mathcal{E}_k^\ell$. This is true up to any initial subsequence, in the sense that there exists at most one $\mathcal{R} \in \mathcal{R}_N^{\mathcal{P}}$ which agrees with the moves and symbols $\mathcal{E}_k^\ell$ up to subscript.

We show that the sets $\mathcal{A}_N(\mathcal{R})$ are disjoint. Take distinct $\mathcal{R}_1 = (M_1^1, \ldots, M_\lambda^1)$, $\mathcal{R}_2 = (M_1^2, \ldots, M_{\lambda'}^2) \in \mathcal{R}_N^{\mathcal{P}}$ with base sequences $\mathcal{M}_1(\mathcal{R}_1) = (\tilde{M}_1^1, \tilde{M}_2^1, \ldots, \tilde{M}_\lambda^1)$ and $\mathcal{M}_2(\mathcal{R}_2) = (\tilde{M}_1^2, \tilde{M}_2^2, \ldots, \tilde{M}_{\lambda'}^2)$, respectively. Since $\mathcal{R}_1 \neq \mathcal{R}_2$, take smallest i for which $M_i^1 \neq M_i^2$, and study the construction of any two games $\mathcal{G}_1 \in \mathcal{A}_N(\mathcal{R}_1)$ and $\mathcal{G}_2 \in \mathcal{A}_N(\mathcal{R}_2)$ as described above. If $\mathcal{G}_1$ and $\mathcal{G}_2$ are consistent prior to $\tilde{M}_i^1$ and $\tilde{M}_i^2$ (i.e., for $k \geq 2$, all $\mathcal{E}_k^\ell$ are replaced by the same prefix of C_k), which is the only way the two games remain equal up to this point, then we produce a difference in the two games on $\tilde{M}_i^1$ and $\tilde{M}_i^2$ if $M_i^1 = \tilde{M}_i^1$ and $M_i^2 = \tilde{M}_i^2$ ($M_i^1 \neq M_i^2$) or $M_i^1 \neq \tilde{M}_i^1$

---

[11] In this subsection and Sect. 5, we borrow much of the same notation that was used in establishing Theorem 9. It will be clear from context exactly what objects we are referring to.

and $M_i^2 \neq \tilde{M}_i^2$ ($\tilde{M}_i^1 = \mathcal{E}_{k_1}^{\ell_1}$, $\tilde{M}_i^2 = \mathcal{E}_{k_2}^{\ell_2}$, $k_1 \neq k_2$). If $M_i^1 = \tilde{M}_i^1$ and $M_i^2 \neq \tilde{M}_i^2$ (say $\tilde{M}_i^2 = \mathcal{E}_k^{\ell}$, so $M_i^2 = C_k$), then $\mathcal{G}_2$ is next filled with either $C_k \neq M_i^1$ ($M_i^1 = M_i^2$ if $C_k = M_i^1$) or an (l, k)-prefix for some $l \leq \ell$. In the latter case, if the two games were equal until move $i + l - 1$, necessarily $\mathcal{R}_1$ contains a prefix of length at least 1, contradicting $\mathcal{R}_1 \in \mathcal{R}_N^{\mathcal{P}}$.

We therefore conclude that $\mathcal{A}_N(\mathcal{R}_1) \cap \mathcal{A}_N(\mathcal{R}_2) = \emptyset$, i.e., the sets $\mathcal{A}_N(\mathcal{R})$ for $\mathcal{R} \in \mathcal{R}_N$ are disjoint; it remains to show that any game $\mathcal{G} \in \Omega_N$ is in some set $\mathcal{A}_N(\mathcal{R})$. For $\mathcal{G} \in \Omega_N$, let $\mathcal{R}$ be the game resulting from replacing every combine move C_k with the longest playable (ℓ, k)-prefix. It follows immediately from construction that $\mathcal{R} \in \mathcal{R}_N^{\mathcal{P}}$, and that $\mathcal{G} \in \mathcal{A}_N(\mathcal{R})$.

Similarly, we define the generalization for arbitrary (ℓ, k)-suffixes. Define $\mathcal{R}_N^{\mathcal{S}} \in \mathcal{F}_N$ to be the collection of all Zeckendorf games on input N such that any combine move C_k, for $k \geq 2$, cannot be compressed by a $(1, k)$-suffix (and thus by any suffix of C_k). Explicitly,

$$\mathcal{R}_N^{\mathcal{S}} = \{(M_1, \dots, M_\lambda) \in \Omega_N : M_i = C_k \implies M_{i+1} \neq S_{k+1} \ \forall i \in [\lambda], k \geq 2\}$$

For a game $\mathcal{R} = (M_1, M_2, \dots, M_\lambda) \in \mathcal{R}_N^{\mathcal{S}}$ with number of moves λ, construct the subset of indices $\mathcal{I}_N(\mathcal{R})$ to denote all combine moves in $\mathcal{R}$:

$$\mathcal{I}_N(\mathcal{R}) = \{i \in [\lambda] : M_i = C_k \text{ for some } k \geq 2\}.$$

Now, construct the *formal* sequence of moves $\mathcal{M}(\mathcal{R}) = (\tilde{M}_1, \tilde{M}_2, \dots, \tilde{M}_\lambda)$ by replacing M_i by a symbol $\mathcal{E}_k^{\ell}$ for all $i \in \mathcal{I}_N(\mathcal{R})$: the subscript is the corresponding $k \geq 2$, while ℓ denotes the longest (ℓ, k)-suffix that $M_i = C_k$ can be expanded into. Call $\mathcal{M}(\mathcal{R})$ the **base sequence** of $\mathcal{R} \in \mathcal{R}_N$, and let $\mathcal{A}_N(\mathcal{R}) \in \mathcal{F}_N$ denote the collection of all Zeckendorf games resulting from replacing each instance of $\mathcal{E}_k^{\ell}$ in $\mathcal{M}$ by an (l, k)-suffix for some $l \leq \ell$. We establish the following analogue of Lemma 6.

Proposition 6 *The sets $\{\mathcal{A}_N(\mathcal{R}) : \mathcal{R} \in \mathcal{R}_N^{\mathcal{S}}\}$ partition Ω_N.*

We do not provide the proof of Proposition 6, as it is pursued analogously to the proof of Proposition 5. The easy observation that each set $\mathcal{A}_N(\mathcal{R})$, for $\mathcal{R} \in \mathcal{R}_N^{\mathcal{P}}$ or $\mathcal{R} \in \mathcal{R}_N^{\mathcal{S}}$, is a union of equivalence classes of the corresponding sets studied in the proof of Theorem 9, yields the following trivial extension of Lemma 2, where we promote the notation $\mathcal{A}_N(\mathcal{G}) = \mathcal{A}_N(\mathcal{R})$ for the unique $\mathcal{R}$ satisfying $\mathcal{G} \in \mathcal{A}_N(\mathcal{R})$.

Proposition 7 *Say $F_n \leq N < F_{n+1}$. For any $c \in (0, \varphi)$ (with φ the golden ratio),*

$$\lim_{N \to \infty} \mu_N \left(\log_2 |\mathcal{A}_N(\mathcal{G})| \geq c^n \right) = 1, \qquad \lim_{N \to \infty} \mathbb{P}_N \left(\log_2 |\mathcal{A}_N(\mathcal{G})| \geq c^n \right) = 1,$$

which holds for either of the understandings $\mathcal{R} \in \mathcal{R}_N^{\mathcal{P}}$ and $\mathcal{R} \in \mathcal{R}_N^{\mathcal{S}}$.

Finally, we have the following extension of Corollary 1. We also omit the proof of this result, as it is a straightforward generalization of the proof of the aforementioned result.

Proposition 8 *Say $F_n \leq N < F_{n+1}$, fix some $\mathcal{R} \in \mathcal{R}_N^{\mathcal{P}}$, and define random variables $X_1^{\mathcal{R}}, X_2^{\mathcal{R}}, \ldots, X_{N-1}^{\mathcal{R}}$ corresponding to the length of the expansion corresponding to each combine move for a game in $\mathcal{A}_N(\mathcal{R})$. Under the measures μ_N and $\mathbb{P}_N$ conditioned on $\mathcal{A}_N(\mathcal{R})$, the random variables $X_i^{\mathcal{R}}$ are independent.*

5　Weak Convergence of Random Game Lengths as $N \to \infty$

In pursuit of the resolution of Conjecture 1, it may be productive to ask whether there exist natural subsets of Ω_N on which the distribution of random game lengths converge weakly to a Gaussian in the limit $N \to \infty$ of infinite input. As discussed in Sect. 4.4, the partitions defined in Propositions 5 and 6 enjoy this property in the sense described in Theorem 10, restated below.

Theorem 10 *For $\mathcal{R} \in \mathcal{R}_N^{\mathcal{P}}$, let $F_N^{\mathcal{R}}(x) : \mathbb{R} \to [0, 1]$ denote the distribution function corresponding to game lengths in $\mathcal{A}_N(\mathcal{R})$ over the conditional distribution induced by $\mathbb{P}_N$, normalized to have expectation 0 and variance 1. Let $\Phi : \mathbb{R} \to [0, 1]$ denote the distribution function of the standard normal. Then for any $\epsilon > 0$,*

$$\lim_{N \to \infty} \mathbb{P}_N \left(\sup_{x \in \mathbb{R}} \left| F_N^{\mathcal{R}}(x) - \Phi(x) \right| \geq \epsilon \right) = 0.$$

Similarly, for $\mathcal{R} \in \mathcal{R}_N^{\mathcal{S}}$, let $F_N^{\mathcal{R}}(x) : \mathbb{R} \to [0, 1]$ denote the distribution function corresponding to game lengths in $\mathcal{A}_N(\mathcal{R})$ over the conditional distribution induced by $\mathbb{P}_N$, normalized to have expectation 0 and variance 1. Then for any $\epsilon > 0$,

$$\lim_{N \to \infty} \mathbb{P}_N \left(\sup_{x \in \mathbb{R}} \left| F_N^{\mathcal{R}}(x) - \Phi(x) \right| \geq \epsilon \right) = 0.$$

The analogous results hold for the uniform measure μ_N.

To explicitly define the distributions referenced in Theorem 10, define a probability distribution $\mathbb{P}_N^{\mathcal{R}} : 2^{\mathcal{A}_N(\mathcal{R})} \to [0, 1]$ via, for any $\mathcal{S} \in 2^{\mathcal{A}_N(\mathcal{R})}$,

$$\mathbb{P}_N^{\mathcal{R}}(\mathcal{S}) = \frac{\mathbb{P}_N \left(\mathcal{S} \cap \mathcal{A}_N(\mathcal{R}) \right)}{\mathbb{P}_N \left(\mathcal{A}_N(\mathcal{R}) \right)}$$

Define the random variable $L_N^{\mathcal{R}} : \mathcal{A}_N(\mathcal{R}) \to \mathbb{R}$ on the space $(\mathcal{A}_N(\mathcal{R}), 2^{\mathcal{A}_N(\mathcal{R})}, \mathbb{P}_N^{\mathcal{R}})$ by

$$L_N^{\mathcal{R}} \left((M_1, \ldots, M_\lambda) \right) = \lambda$$

for any game $(M_1, \ldots, M_\lambda) \in \mathcal{A}_N(\mathcal{R})$, i.e., $L_N^{\mathcal{R}}$ studies game lengths in $\mathcal{A}_N(\mathcal{R})$. Then

$$F_{\mathcal{R}}(x) = \mathbb{P}_N^{\mathcal{R}} \left(\frac{L_N^{\mathcal{R}} - \mathbb{E}\left[L_N^{\mathcal{R}}\right]}{\sqrt{\text{var}\left(L_N^{\mathcal{R}}\right)}} \leq x \right). \tag{3}$$

In other words, when we restrict $\mathbb{P}_N$ to the sets $\mathcal{A}_N(\mathcal{R})$ in the natural sense, Theorem 10 states that the distribution of random game lengths enjoys weak convergence to a Gaussian with high probability. We write

$$\mathcal{L}_N^{\mathcal{R}} = \frac{L_N^{\mathcal{R}} - \mathbb{E}\left[L_N^{\mathcal{R}}\right]}{\sqrt{\text{var}\left(L_N^{\mathcal{R}}\right)}} \tag{4}$$

to be the random variable $L_N^{\mathcal{R}}$ normalized to have mean 0 and variance 1. Throughout this section, we proceed on the measure $\mathbb{P}_N$ and the set $\mathcal{R}_N^{\mathcal{P}}$, but the analysis carries over to the uniform measure μ_N and the suffix partition $\mathcal{R}_N^{\mathcal{S}}$.

Proof *(Proof of Theorem 10)*
By Proposition 7, $\lim_{N \to \infty} \mathbb{P}_N \left(\log_2 |\mathcal{A}_N(\mathcal{G})| \geq 1.6^n \right) = 1$ so fix $\mathcal{R} \in \mathcal{R}_N$ such that $|\mathcal{A}_N(\mathcal{R})| > 2^{1.6^n}$, and consider the random variable $\mathcal{L}_N^{\mathcal{R}} : \mathcal{A}_N(\mathcal{R}) \to \mathbb{R}$. Define random variables[12] $X_1^{\mathcal{R}}, \ldots, X_{N-1}^{\mathcal{R}}$ to correspond to each of the combine moves in $\mathcal{R}$, with $X_i^{\mathcal{R}}$ denoting the length of the corresponding prefix for a game in $\mathcal{A}_N(\mathcal{R})$: by Proposition 8, the random variables $X_i^{\mathcal{R}}$ are independent. Expanding Eq. (4),

$$\mathcal{L}_N^{\mathcal{R}} = \frac{L_N^{\mathcal{R}} - \mathbb{E}\left[L_N^{\mathcal{R}}\right]}{\sqrt{\text{var}\left(L_N^{\mathcal{R}}\right)}} = \frac{\sum_{i=1}^{N-1} \left(X_i^{\mathcal{R}} - \mathbb{E}\left[X_i^{\mathcal{R}}\right]\right)}{\sqrt{\text{var}\left(L_N^{\mathcal{R}}\right)}} \tag{5}$$

where the independence (conditioned on $\mathcal{A}_N(\mathcal{R})$) of the random variables $X_i^{\mathcal{R}}$ yields that

$$\text{var}\left(L_N^{\mathcal{R}}\right) = \sum_{i=1}^{N-1} \text{var}\left(X_i^{\mathcal{R}} - \mathbb{E}\left[X_i^{\mathcal{R}}\right]\right)$$

and furthermore, taking the maximum over those summands $X_i^{\mathcal{R}}$ which are non-constant,[13]

---

[12] Define these to be strictly positive: if the ith delimiter is replaced with a combine move, say $X_i^{\mathcal{R}} = 1$.

[13] Indeed, we understand $\mathcal{L}_N^{\mathcal{R}}$ as the sum over the nonconstant random variables amongst those included by the sum in Eq. (5).

$$\max_{1 \le i \le D} \left(\frac{\mathbb{E}\left[\left|X_i^{\mathcal{R}} - \mathbb{E}\left[X_i^{\mathcal{R}}\right]\right|^3\right]}{\mathbb{E}\left[\left(X_i^{\mathcal{R}} - \mathbb{E}\left[X_i^{\mathcal{R}}\right]\right)^2\right] \cdot \mathrm{var}\left(L_N^{\mathcal{R}}\right)} \right) \le \max_{1 \le i \le D} \frac{2n^3}{\mathrm{var}\left(L_N^{\mathcal{R}}\right)} \le \frac{n^3}{\mathrm{var}\left(2^{1.6^n}\right)} \quad (6)$$

and the final expression in Eq. (6) certainly vanishes as $N \to \infty$. Thus, the desired result follows immediately from the case of the Berry-Esseen theorem for independent non-identically distributed summands, which namely yields

$$\sup_{x \in \mathbb{R}} |F_{\mathcal{R}}(x) - \Phi(x)| \le C \cdot \frac{n^3}{\mathrm{var}\left(2^{1.6^n}\right)} \xrightarrow{N \to \infty} 0$$

for some universal constant $C > 0$.

Theorem 10 yields that, when restricted to particular natural subsets of games, we have Gaussianity. The scope of this result is admittedly restricted, especially given the result of Corollary 3.

6 Open Problems

We conclude the work with several potential directions for further inquiry.

6.1 *Other Two-Player Games Based on Recurrences*

Theorem 9 can be interpreted as saying that if two players proceed mindlessly, the Zeckendorf game is fair in the limit of infinite input; Theorem 8 gives that the analogous statement would be true if we were to extend to a Z-player Zeckendorf game. Many papers (such as [1, 2, 4]) have extended the paradigm of the two-player Zeckendorf game to other recurrences: we might ask which of these also enjoy this property. In particular, we pose the following conjecture, concerning the two-player Bergman game [1].

Conjecture 2 In the limit $N \to \infty$ of infinite input, the probabilities of Player 1 and Player 2 winning, under both analogous definitions of random Bergman games, is 1/2.

The core challenges of proving Conjecture 2 are as follows. The principal difference between the two-player Bergman game and the two-player Zeckendorf game is that the move C_1 is now a split move which consumes no tokens. Thus, there are structural differences between the two games which affects the range of achievable game lengths. Lemma 2 also depends on the number of a specific move sequence which may not have an analogue in the Bergman game.

6.2   Towards Gaussianity: Other Approaches

Theorem 10 establishes that certain natural partitions of Ω_N are such that the components become arbitrarily close to being Gaussian (in the sense of vanishing Kolmogorov-Smirnov distance) with arbitrarily high probability: it is unclear how to extend this to the entirety of Ω_N. Proposition 2 remarks that all games in Ω_N can be achieved by permutations of suffixes from a shortest game, which suggests the following question.

Question 1 Can we extend the techniques in Sect. 5 to a partition founded on the greedy embedding of Proposition 2? In particular, could studying the moments of the corresponding components lead to a proof of Conjecture 1?

Another possible direction is to restrict our attention to certain subsets of moves across all of Ω_N, rather than certain subsets of Ω_N. In particular, a sufficiently strong affirmative answer to Question 2 would resolve Conjecture 1.

Question 2 Is the distribution of the number of occurrences of a particular combine or split move asymptotically Gaussian, under either the measure μ_N or $\mathbb{P}_N$?

6.3   Towards Gaussianity: Mixing

We outline one more possible approach towards proving Conjecture 1, which relies on the literature surrounding *mixing central limit theorems*: these are analogues of the central limit theorem concerning sums of dependent random variables, applicable if the dependencies amongst the summands are sufficiently well-behaved. In particular, the main result of [6] states the following.

Theorem 11 ([6]) *Let* $\{X_{N,i} : 1 \le i \le d_N\}$ *be a triangular array of random variables defined on the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, $\bar{X}_{N,d_N} = \frac{1}{d_N} \sum_{i=1}^{d_N} X_{N,i}$, *and* $\alpha_N : \mathbb{N} \to \mathbb{R}$ *by*

$$\alpha_N(k) = \sup_m \sup_{A \in \mathcal{F}_0^m(N), B \in \mathcal{F}_{m+k}^\infty(N)} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \tag{7}$$

where $\mathcal{F}_{m_1}^{m_2}(N) = \sigma\left(X_{N,m_1}, X_{N,m_1+1}, \ldots, X_{N,m_2}\right)$ *denotes the σ-algebra generated by random variables* $X_{N,m_1}, X_{N,m_1+1}, \ldots, X_{N,m_2}$. *If there exists constants* $C_1, C_2 > 0, \delta > 0$ *such that*

$$\mathbb{E}\left[\left|X_{N,i} - \mathbb{E}[X_{N,i}]\right|^{2+\delta}\right] < C_1 \tag{8}$$

and

$$\sum_{k=0}^{\infty} (k+1)^2 \alpha_N^{\frac{\delta}{4+\delta}}(k) < C_2, \tag{9}$$

then

$$\sqrt{d_N} \left(\bar{X}_{N,d_N} - \mathbb{E}\bar{X}_{N,d_N} \right) \xrightarrow[N\to\infty]{\mathcal{D}} N\left(0, var\left(\sqrt{d_N}\bar{X}_{N,d_N} \right) \right). \tag{10}$$

We can define a triangular array of Bernoulli random variables by letting $X_{N,i} = 1$ if and only if the ith move of a Zeckendorf game with input N is a splitting move (in particular, $X_{N,i}$ maps to 0 on the event that the Zeckendorf game on input N terminates prior to move i), with d_N (the number of random variables in the Nth row) the length of the longest game on input N; Conjecture 1 is thus equivalent to establishing that the row sums $\sum_{i=1}^{d_N} X_{N,i}$ weakly converge to a Gaussian, since the number of combine moves for a given input N is constant. This is precisely the statement of Eq. 10. Also, the events A and B in Eq. 7 can be understood as fixed $\{0, 1\}$-realizations of subsets of random variables in $\{X_{N,1}, \ldots, X_{N,m}\}$ and $\{X_{N,m+k}, \ldots, X_{N,d_N}\}$, respectively.

Since Eq. 8 certainly holds under this setup (the constituent random variables are Bernoulli), resolving Conjecture 1 can be reduced to the following statement.

Question 3 For the triangular array $\{X_{N,i} : 1 \le i \le d_N\}$ defined above, does there exist a constant $\delta > 0$ and a constant $C > 0$ such that

$$\sum_{k=0}^{\infty} (k+1)^2 \alpha_N^{\frac{\delta}{4+\delta}}(k) < C \tag{11}$$

for all natural numbers $N \in \mathbb{N}$?

By the preceding discussion, answering Question 3 in the affirmative by establishing 11 immediately yields Conjecture 1. In particular, it would suffice to establish the statement of Question 4 in the affirmative, although it is not immediately clear if this should be true.

Question 4 Does there exist a constant $\epsilon > 0$ and a constant $C > 0$ such that

$$|\mathbb{P}_N(A \cap B) - \mathbb{P}_N(A)\mathbb{P}_N(B)| \le \frac{C}{k^{2+\epsilon}} \tag{12}$$

whenever events A and B correspond to fixed $\{0, 1\}$-realizations of subsets of the Bernoulli random variables in $\{X_{N,1}, \ldots, X_{N,m}\}$ and $\{X_{N,m+k}, \ldots, X_{N,d_N}\}$, respectively?

Of course, the techniques introduced in the main body of this paper are quite elementary, and there may be other promising approaches not included in this section that would contribute towards resolving Conjecture 1.

References

1. Benjamin Baily et al. "The Generalized Bergman Game". In: (Sept. 19, 2021). https://doi.org/10.48550/arXiv.2109.00117. arXiv: 2109.00117 [math]. URL: http://arxiv.org/abs/2109.00117

2. Paul Baird-Smith et al. "The generalized Zeckendorf game". In: *Fibonacci Quart*. 57.5 (2019), pp. 1–14. ISSN: 0015-0517. arXiv: 1809.04883 [math.NT]. URL: https://web.williams.edu/Mathematics/sjmiller/publichtml/math/papers/ZeckGameGeneralFibQ10.pdf.

3. Paul Baird-Smith et al. "The Zeckendorf Game". In: *Combinatorial and Additive Number Theory, New York Number Theory Seminar*. Springer. 2018, pp. 25–38.

4. Elżbieta Bołdyriew et al. "Extending Zeckendorf's Theorem to a Non-constant Recurrence and the Zeckendorf Game on this Non-constant Recurrence Relation'. In: arXiv:2009.12475 (2020).

5. Anna Cusenza et al. "Bounds on Zeckendorf games". In: *Fibonacci Quart*. 60.1 (2022), pp. 57–71. ISSN: 0015-0517.

6. Magnus Ekström. "A general central limit theorem for strong mixing sequences". In: *Statistics & Probability Letters* 94 (2014), pp. 236–238.

7. Ruoci Li et al. "Deterministic Zeckendorf Games". In: arXiv:2006.16457 (2020).

8. Steven J. Miller. "Private communication". In: (2022).

9. Edouard Zeckendorf. "Representations des nombres naturels par une somme de nombres de fibonacci on de nombres de lucas". In: *Bulletin de La Society Royale des Sciences de Liege* (1972), pp. 179–182.