A CLOSER LOOK AT THE FINE STRUCTURE OF THE RANDOM WALK ON $(0,1)$

MARINA SKYERS AND LEE J. STANLEY

ABSTRACT. Let S_n be the random walk on $(0,1)$. The S_n have been the subject of intense study; their definition is immediately intuitive. Nevertheless, they are quite disorderly and this disorder is mirrored by the fact that, pointwise, $\left(\frac{S_n}{\sqrt{n}} | n \in \mathbb{N}^+\right)$ behaves quite badly. In this paper we provide our results on the fine structure of the random walk that give insight into this behavior.

1. Introduction

Let S_n be the random walk on $(0,1)$. The random walk has been the subject of intense study (see the work of Erdös and Revesz [3] and Shi and Toth [6]). Indeed, the definition of the S_n is immediately accessible and intuitive and each S_n is readily representable as the sum of an i.i.d. family (of size n) of irreducibly simpler random variables. While immediately intuitive, the S_n are quite disorderly. This disorder is mirrored by the fact that, for almost all x, $\{S_n(x)/\sqrt{n}\}$ diverges ([1], [2], [4], and [5]).

CONGRESSUS NUMERANTIUM 232 (2019). PP.265-279

Definition 1. Let λ will denote Lebesgue measure on $[0,1]$ (or on one of the variants with either end-point or both excluded). As usual, a *probability space* is a triple $(\Omega, \mathscr{S}, P)$, where Ω is the set of points, $\mathscr{S}$ is the σ-algebra of Borel subsets of Ω, and $P : \mathscr{S} \to [0,1]$ is the (σ-additive) probability measure. In this paper we will have $\Omega = [0,1)$, $\mathscr{S}$ will be the σ-algebra of Borel subsets of Ω, and P will be the restriction of Lebesgue measure to the Borel sets.

We will use $\operatorname{card}(x)$ to denote the cardinality of the set x. C will denote Cantor space, $\{0,1\}^{\mathbb{N}^+}$. We also take C' to be the set of those $x \in C$ such that $x^{-1}[\{0\}]$ is infinite, i.e., such that for infinitely many i, $x_i = 0$. Thus, as usual, C' is canonically identified with the half-open unit interval, $[0,1)$, by identifying x with $\sum_{i=1}^{\infty} \frac{x_i}{2^i} \in [0,1)$.

Definition 2. For $x \in C'$, and $n \in \mathbb{N}^+$, by $S_n(x)$ we mean $\sum_{i=1}^{n} (-1)^{1+x_i}$.

Note that, obviously, S_n depends only on the first n coordinates of x. We exploit this to regard S_n as having domain $\{0,1\}^n$ when it suits our purposes to do so:

$$S_n(\boldsymbol{r}) := S_n(x) \text{ for any } x \in C' \text{ such that } x \supseteq \boldsymbol{r}.$$

We now define the set of *extreme sequences* as follows.

Definition 3. For such $x \in C'$, and $k \in \mathbb{N}^+$, we set $x \in \mathfrak{X}_{k,n}$ if and only if $|S_n(x)| > k\sqrt{n}$ (x is extreme at level n). We also take $\mathfrak{X}_k$ to be $\bigcup_{n\in\mathbb{N}^+} \mathfrak{X}_{k,n}$, and we define $\mathfrak{Y}_{k,n}$ to be $\mathfrak{X}_{k,n} \setminus \bigcup_{m=1}^{n-1} \mathfrak{X}_{k,m}$. For $k \in \mathbb{N}^+$, we set $U_k := \{n \in \mathbb{N}^+ | \mathfrak{Y}_{k,n} \neq \emptyset\}$.

The $n \in U_k$ are the levels at which some $x \in C'$ first become extremes. We sometimes refer to the $n \in U_k$ as "successes". Note that clearly $\mathfrak{X}_k = \bigsqcup_{n\in\mathbb{N}^+} \mathfrak{Y}_{k,n}$. It is immediate that if $x \in C' \cap \mathfrak{X}_k$, then there is a unique $n^* = n_k^*(x)$ such that $x \in \mathfrak{Y}_{k,n^*}$, and that this n^* is the least n such that $x \in \mathfrak{X}_{k,n}$.

Our main result, Theorem 19, gives a purely arithmetic condition on n that characterizes membership in U_k. On the way to this result, we will see that U_k is infinite and we take $(u_{k,j} | j \geq 1)$ to be its increasing enumeration. We provide a complete analysis of U_k, concentrating on the "gaps", i.e., the $u_{k,j+1} - u_{k,j}$.

Proposition 21 shows that far enough out, the gaps of four disappear and the gaps of two predominate. The number of gaps of two in between each gap of three depends on the interplay between the greatest integer function $[k\sqrt{n}]$ and S_n. Combined with our results about the gain/loss $\lambda\left(\mathfrak{X}_{k,u_j+1}\right) - \lambda\left(\mathfrak{X}_{k,u_j}\right)$ [7], this analysis would provide a new proof of the almost everywhere failure of pointwise convergence for $\frac{S_n}{\sqrt{n}}$.

In this paper we provide our results on the fine structure of the random walk that give insight into the disorderly behavior of the S_n. It turns out that much of the fine structure of the random walk on $(0,1)$ depends on how often the sequence $\{\lfloor k\sqrt{n} \rfloor\}_{n=1}^{\infty}$ increases and the way the fractional part of $k\sqrt{n}$ is distributed, for a fixed positive integer k. This work lays the groundwork for further developments where k is a function of n.

2. Extreme Sequences

For $x \in C'$, and $n \in \mathbb{N}^+$, we set $\mathrm{maj}_n(x) := 1$ if and only if $S_n(x) \geq 0$; otherwise, $\mathrm{maj}_n(x) := -1$. We set $\min_n(x) := -\mathrm{maj}_n(x)$. The $\mathrm{maj}_n(x)$ and $\min_n(x)$ are the majority and minority values in $S_n(x)$, with ties (when $S_n(x) = 0$) being in favor of 1 as the majority. We set

$$\mathrm{Maj}_n(x) := \left\{ i \in [1,n] \cap \mathbb{N} \,\middle|\, (-1)^{1+x_i} = \mathrm{maj}_n(x) \right\}$$

and similarly for $\mathrm{Min}_n(x)$ and $\min_n(x)$, and we set $M_n(x) := \mathrm{card}(\mathrm{Maj}_n(x))$, $m_n(x) := \mathrm{card}(\mathrm{Min}_n(x))$. Thus, for example, for such x and n, $\mathrm{Maj}_n(x)$ is the set of coordinates, i, with $1 \leq i \leq n$, where the majority value in $S_n(x)$ occurs. Also, note that

$n = M_n(x) + m_n(x)$, $|S_n(x)| = M_n(x) - m_n(x) = n - 2m_n(x) = 2M_n(x) - n$, and so $S_n(x)$ always has the same parity as n.

Lemma 4. *We have the following three simple observations.*

(a) *If* $n \le k^2$ *then* $\mathfrak{X}_{k,n} = \emptyset$.

(b) $\mathfrak{X}_{k,k^2+1} = \mathfrak{Y}_{k,k^2+1} = \{x | x \restriction [1, k^2+1] \cap \mathbb{N} \text{ is constant}\}$.

(c) $\lambda(\mathfrak{X}_{k,k^2+1}) = \lambda(\mathfrak{Y}_{k,k^2+1}) = 2 \cdot \frac{1}{2^{k^2+1}} = \frac{1}{2^{k^2}}$.

Proof. Immediate from the definitions. □

Before going farther, it is worth noting that if $\mathfrak{Y}_{k,t} = \emptyset$, (i.e., if $t \notin U_k$), then $\bigcup_{i \in [1,t) \cap \mathbb{N}} \mathfrak{X}_{k,i} = \bigcup_{i \in [1,t] \cap \mathbb{N}} \mathfrak{X}_{k,i}$. A key step is the formulation of the following notion.

Definition 5. For $k, n \in \mathbb{N}^+$, we define $\sigma_{k,n} := \lfloor k\sqrt{n} \rfloor + 2$, if $\lfloor k\sqrt{n} \rfloor$ has the same parity as n, and $\sigma_{k,n} := \lfloor k\sqrt{n} \rfloor + 1$, otherwise. We also set $m_{k,n} := \frac{1}{2}(n - \sigma_{k,n})$ and $c_{k,n} := \mathrm{card}(U_k \cap [1, n))$.

Lemma 6. *If* $x \in \mathfrak{Y}_{k,n}$ *then* $|S_n(x)| = \sigma_{k,n}$.

Proof. $\sigma_{k,n}$ is the smallest integer greater than $k\sqrt{n}$ which has the same parity as n. Also, $|S_n(x)|$ always has the same parity as n. So $|S_n(x)| > k\sqrt{n}$ if and only if $|S_n(x)| \ge \sigma_{k,n}$. Let $x \in \mathfrak{Y}_{k,n}$. Then $x \in \mathfrak{X}_{k,n}$, so $|S_n(x)| \ge \sigma_{k,n}$. Assume, towards a contradiction, that $|S_n(x)| > \sigma_{k,n}$. Then $|S_n(x)| \ge \sigma_{k,n} + 2$, since they have the same parity. Since $|S_n(x)| \le |S_{n-1}(x)| + 1$, we have

$$\begin{aligned} |S_{n-1}(x)| \ge |S_n(x)| - 1 \ge \sigma_{k,n} + 2 - 1 &= \sigma_{k,n} + 1 \\ &> k\sqrt{n} \\ &\ge k\sqrt{n-1}. \end{aligned}$$

So $|S_{n-1}(x)| > k\sqrt{n-1}$ and $x \in \mathfrak{X}_{k,n-1}$, a contradiction, since $x \in \mathfrak{Y}_{k,n}$. This proves $|S_n(x)| = \sigma_{k,n}$. □

Lemma 6 implies that for $n \in U_k$, $m_{k,n} = m_n(x)$ for all $x \in \mathfrak{Y}_{k,n}$ and we can let $m_n := m_n(x)$ for any $x \in \mathfrak{Y}_{k,n}$. And so for $n \in U_k$, $m_n = m_{k,n}$.

The next Definition and Lemma are purely arithmetical:

Definition 7. We define the following notions.

(a) *The greatest integer jumps at n* if and only if $\lfloor k\sqrt{n+1} \rfloor = \lfloor k\sqrt{n} \rfloor + 1$.

(b) *The parity situation is the same at n* if and only if $\lfloor k\sqrt{n} \rfloor \equiv n \pmod 2$ and *the parity situation is different* otherwise.

Lemma 8. *The relation between Definition 7 (a) and (b) is as follows: if the greatest integer does not jump at n, then, in passing from n to $n+1$, the parity situation changes. If the greatest integer jumps at n, then, in passing from n to $n+1$, the parity situation does not change. Further, $\sigma_{k,n}$ depends on whether the greatest integer jumps at n as described below. In particular, $|\sigma_{k,n+1} - \sigma_{k,n}| = 1$.*

Proof. There are four possible cases. In Cases 1 and 2, we assume that the greatest integer does not jump at n, and we consider the possibilities for the parity situation. In Cases 3 and 4, we assume the greatest integer does jump at n and the parity situation is as in Cases 1 and 2, respectively.

So first assume that $\lfloor k\sqrt{n+1} \rfloor = \lfloor k\sqrt{n} \rfloor$. The four possible cases are as follows.

Case 1. Suppose $\lfloor k\sqrt{n} \rfloor \equiv n \pmod 2$. Then $\lfloor k\sqrt{n} \rfloor = \lfloor k\sqrt{n+1} \rfloor \not\equiv n+1 \pmod 2$, i.e., the parity situation changes. Note also that in this case, $\sigma_{k,n} = \lfloor k\sqrt{n} \rfloor + 2$, $\sigma_{k,n+1} = \lfloor k\sqrt{n+1} \rfloor + 1 = \lfloor k\sqrt{n} \rfloor + 1$, i.e., $\sigma_{k,n+1} = \sigma_{k,n} - 1$.

Case 2. Suppose $\lfloor k\sqrt{n} \rfloor \not\equiv n \pmod 2$. Then $\lfloor k\sqrt{n} \rfloor = \lfloor k\sqrt{n+1} \rfloor \equiv n+1 \pmod 2$, so, here too, the parity situation changes. Also, here $\sigma_{k,n} = \lfloor k\sqrt{n} \rfloor + 1$, and $\sigma_{k,n+1} = \lfloor k\sqrt{n+1} \rfloor + 2 = \lfloor k\sqrt{n} \rfloor + 2$, i.e., $\sigma_{k,n+1} = \sigma_{k,n} + 1$.

So, when the greatest integer does not jump, the parity situation changes. Now assume that the greatest integer does jump, i.e., $\lfloor k\sqrt{n+1}\rfloor = \lfloor k\sqrt{n}\rfloor + 1$, and, in addition:

Case 3. Suppose $\lfloor k\sqrt{n}\rfloor \equiv n \pmod 2$. Then $\lfloor k\sqrt{n}\rfloor + 1 = \lfloor k\sqrt{n+1}\rfloor \equiv n+1 \pmod 2$ so the parity situation does not change. Also, here, $\sigma_{k,n} = \lfloor k\sqrt{n}\rfloor + 2$ and also $\sigma_{k,n+1} = \lfloor k\sqrt{n+1}\rfloor + 2 = (k\sqrt{n}+1) + 2$, i.e., $\sigma_{k,n+1} = \sigma_{k,n} + 1$.

Case 4. Suppose $\lfloor k\sqrt{n}\rfloor \not\equiv n \pmod 2$. Then $\lfloor k\sqrt{n}\rfloor + 1 = \lfloor k\sqrt{n+1}\rfloor \not\equiv n+1 \pmod 2$, so, here too, the parity situation does not change. Also, here, $\sigma_{k,n} = \lfloor k\sqrt{n}\rfloor + 1$ and $\sigma_{k,n+1} = \lfloor k\sqrt{n+1}\rfloor + 1 = (k\sqrt{n}+1) + 1$, i.e., $\sigma_{k,n+1} = \sigma_{k,n} + 1$. □

Lemma 9. *If $x \in \mathfrak{Y}_{k,n}$, then $|S_n(x)| > |S_l(x)|$ for all $l < n$. Therefore if $x \in \mathfrak{Y}_{k,n}$, then $x_n, x_{n-1} = maj_n(x)$.*

Proof. If $x \in \mathfrak{X}_{k,n}$, $l < n$ and $|S_n(x)| \le |S_l(x)|$, then $|S_l(x)| \ge |S_n(x)| > \lfloor k\sqrt{n}\rfloor \ge \left\lfloor k\sqrt{l}\right\rfloor$, and so $x \in \mathfrak{X}_{k,l}$. Suppose $x \in \mathfrak{Y}_{k,n}$ and $n \in \mathrm{Min}_n(x)$. Note that the case where $S_{n-1}(x) = 0$ and the majority value at n switches cannot arise, since $n \in U_k \Rightarrow n \ge k^2+1 \Rightarrow |S_n(x)| \ge k^2+1$. Then since $|S_n(x)| \ge 2$, $|S_n(x)| = |S_{n-1}(x)| - 1$, which we just saw was impossible. If $n-1 \in \mathrm{min}_n(x)$, then, since $n-1 \in \mathrm{maj}_n(x)$, we have that $S_n(x) = S_{n-2}(x)$, which, again we saw was impossible. Thus we have a contradiction in both cases which proves the second sentence of the lemma. □

3. Structural Properties

Proposition 10. *Suppose $n \in U_k$. Then there is $1 \le j^* \le 4$ such that $n + j^* \in U_k$.*

Proof. Suppose $x \in \mathfrak{Y}_{k,n}$. We will construct a modification, x^*, of x, which is in $\mathfrak{X}_{k,n+4}$ and not in $\mathfrak{X}_{k,n}$, as follows.

Step 1. $x^*_i = x_i$ for $i < n$ or $i > n+4$.

Step 2. $x^*_n = 1 - x_n$.

Step 3. x^*_{n+j} is such that $(-1)^{1+x_{n+j}} = \text{maj}_n(x)$, $1 \le j \le 4$.

Then Step 1 implies x^* satisfies all the side conditions below level n. Step 1 and Step 2 imply $|S_n(x^*)| = |S_n(x)| - 2$, so $x^* \notin \mathfrak{X}_{k,n}$. Note that, by construction, $|S_{n+4}(x^*)| = |S_n(x)| + 2$. Recall that $\mathfrak{Y}_{k,n} \neq \phi \Rightarrow n > k^2 \Rightarrow \sqrt{n} > k$. We have that

$$\begin{aligned} |S_{n+4}(x^*)|^2 = |S_n(x)|^2 + 4|S_n(x)| + 4 &> k^2n + 4k\sqrt{n} + 4 \\ &> k^2n + 4k^2 + 4 \\ &> k^2n + 4k^2 \\ &= k^2(n+4), \end{aligned}$$

so $|S_{n+4}(x^*)| > k\sqrt{n+4}$, and therefore, $x^* \in \mathfrak{X}_{k,n+4}$.

We will not show $x^* \notin (\mathfrak{X}_{k,n+1} \cup \mathfrak{X}_{k,n+2} \cup \mathfrak{X}_{k,n+3})$; indeed this may be false. Rather, we have shown there is $1 \le j \le 4$ such that $x^* \in \mathfrak{X}_{k,n+j}$. Let j^* be the least such j. Then $x^* \in \mathfrak{Y}_{k,n+j^*}$. So there is $u \in U_k$ such that $n < u \le n+4$. □

Corollary 11. *Thus, U_k is infinite.*

If we let $(u_{k,j} | j \ge 1)$ be the increasing enumeration of U_k, then Proposition 10 can be restated as $u_{k,j+1} - u_{k,j} \le 4$ for all $j \in \mathbb{N}^+$. It is also worth noting that for all $j \ge 1$, $c_{k,u_{k,j}} = j - 1$.

When we are taking k to be fixed, we will lighten the notation by using u_j in place of $u_{k,j}$.
Conventionally, we set $u_{k,0} = u_0 = 0$ for all $k \in \mathbb{N}^+$.

Lemma 12. *If $t \in U_k$, then $\sigma_{k,t} = \lfloor k\sqrt{t} \rfloor + 1$.*

Proof. Let $y \in \mathfrak{Y}_{k,t}$. Then $y \notin \mathfrak{X}_{k,t-1}$, so $|S_{t-1}(y)| \le \lfloor k\sqrt{t-1} \rfloor$. Suppose, towards a contradiction, that $\sigma_{k,t} = \lfloor k\sqrt{t} \rfloor + 2$. Since $y \in \mathfrak{Y}_{k,t}$, $|S_t(y)| = \sigma_{k,t} = \lfloor k\sqrt{t} \rfloor + 2$. Also, Lemma 9 shows that $|S_{t-1}(y)| + 1 = |S_t(y)|$. Then $|S_{t-1}(y)| + 1 = \lfloor k\sqrt{t} \rfloor + 2$, and so $|S_{t-1}(y)| = \lfloor k\sqrt{t} \rfloor + 1 \ge \lfloor k\sqrt{t-1} \rfloor + 1$. Then $|S_{t-1}(y)| > k\sqrt{t-1}$, a contradiction since $y \notin \mathfrak{X}_{k,t-1}$. □

Lemma 13. *If $n \in U_k$ then $m_n = c_{k,n}$ ($=$ card$(U_k \cap [1,n))$). That is, the minority count is the number of previous successes.*

Proof. Suppose $n \in U_k$, so $n = u_t$, for some t. Then, for all $x \in \mathfrak{Y}_{k,u_t}$, $|S_{u_t}(x)| = u_t - 2m_{u_t}$.

Claim: $m_{u_t} = t-1$ ($= c_{k,u_t} = \text{card}(U_k \cap [1,u_t))$).

We prove, by induction on s, that the equation of the Claim holds, with s in place of t. The basis is $m_{u_1} = 0$, which is true since $u_1 = k^2+1$. Assume $m_{u_s} = s-1$. We will show $m_{u_{s+1}} = m_{u_s} + 1$. Then $m_{u_{s+1}} = (s-1)+1 = s = (s+1)-1$. Let $x \in \mathfrak{Y}_{k,u_s}$ and construct x^*, j^* as in Proposition 10. Then $u_s + j^* = u_{s+1}$ and, by construction of x^*, $m_{u_s+j^*}(x^*) = m_{u_s}(x)+1$. So $m_{u_{s+1}} = m_{u_{s+1}}(x^*) = m_{u_s}(x)+1 = m_{u_s}+1$. □

Lemma 14. *If $n \in U_k$, then $n+1 \notin U_k$ (so there are never two consecutive successes).*

Proof. Towards a contradiction, suppose $n, n+1 \in U_k$ and let $x \in \mathfrak{Y}_{k,n+1}$. Then, by Lemma 9, we have $x_{n+1} = x_n = \text{maj}_{n+1}(x) = \text{maj}_n$

Note that $c_{k,n+1} = c_{k,n}+1$ (since $n \in U_k$). But $m_n(x) = m_{n+1}(x)$, by the previous paragraph, and this contradicts Lemma 13. □

The next Lemma gives the converse of Lemma 13.

Lemma 15. *If $m_{k,t} = c_{k,t}$, then $t \in U_k$.*

Proof. Recall $m_{k,n} = \frac{1}{2}(n - \sigma_{k,n})$. Suppose $m_{k,t} = c_{k,t}$. We construct a $y \in \mathfrak{Y}_{k,t}$ by setting

$$y_i = \begin{cases} 1 & \text{if } i \in U_k \cap [1,t) \\ 0 & \text{otherwise.} \end{cases}$$

Appealing to the construction of y_i and Lemma 14, it is an easy induction to see that for all $j \geq k^2$,

$S_n(y) \leq -k^2+1$, and so in particular, for all $j < t$, $\text{maj}_t(y_i) = 0$.

Then $m_t(y) = \text{card}(U_k \cap [1,t)) = m_{k,t} = \frac{1}{2}(t - \sigma_{k,t})$. So $|S_t(y)| = t - 2m_t(y) = t - t + \sigma_{k,t} = \sigma_{k,t} > k\sqrt{t}$, so $y \in \mathfrak{X}_{k,t}$. Now we show that for all $1 \leq r < t$, $y \notin \mathfrak{X}_{k,r}$. Towards a contradiction, suppose otherwise.

Consider the smallest r such that $y \in \mathfrak{X}_{k,r}$. Then also $y \in \mathfrak{Y}_{k,r}$. Then $r \in U_k$, so $c_{k,r} = m_{k,r} = m_r(y)$. By construction, $m_r(y) = \operatorname{card}(U_k \cap [1,r]) = c_{k,r} + 1$ since $r \in U_k$. But then we have $c_{k,r} = m_r(y) = c_{k,r} + 1$, a contradiction. So there is no such r. So $y \in \mathfrak{Y}_{k,t}$. □

Note that if $m_{k,t} < c_{k,t}$, then $t \notin U_k$. This is since if $t \in U_k$, then $m_{k,t} = c_{k,t}$ by Lemma 13.

Lemma 16. *We have that $m_{k,n} \leq c_{k,n}$ for all $n \geq 1$.*

Proof. By induction. Our induction hypothesis is that

$$m_{k,n} \leq c_{k,n}.$$

$c_{k,n}$ is non-decreasing and increases by 1 from n to $n+1$ if and only if $n \in U_k$. Also $\sigma_{k,n}$ either increases or decreases by 1. Therefore, $m_{k,n}$ is also non-decreasing, increases by 1 from n to $n+1$ if $\sigma_{k,n+1} = \sigma_{k,n} - 1$, and does not change if $\sigma_{k,n+1} = \sigma_{k,n} + 1$. We show that $m_{k,n+1} \leq c_{k,n+1}$.

Case 1. Suppose that $m_{k,n} < c_{k,n}$. Then $m_{k,n+1} \leq c_{k,n+1}$.

Case 2. Suppose that $m_{k,n} = c_{k,n}$. Then $n \in U_k$, so $c_{k,n+1} = c_{k,n} + 1$. Thus $m_{k,n+1} \leq m_{k,n} + 1 \leq c_{k,n} + 1 = c_{k,n+1}$. □

Lemma 17. *For all $n \geq 1$, $c_{k,n} \leq m_{k,n} + 1$.*

Proof. We prove that for all $t \geq u_1$, $m_{k,t} \leq c_{k,t} \leq m_{k,t} + 1$. Lemma 16 gives the first inequality, so we are only concerned with the second. The proof is by induction on t; the basis is that for $t = u_1$, we have $c_{k,u_1} = 0 = m_{k,u_1}$. For the induction step, the crucial point is that the $m_{k,n}$'s are non-decreasing. So, suppose that $c_{k,t} \leq m_{k,t} + 1$. Towards a contradiction, assume that $c_{k,t+1} > m_{k,t+1} + 1$. We consider two cases.

Case 1. Suppose $t \notin U_k$. Then $c_{k,t} > m_{k,t}$, and so, by the induction hypothesis, $c_{k,t} = m_{k,t} + 1$. Also, since $t \notin U_k$, $c_{k,t+1} = c_{k,t}$, so $m_{k,t} + 1 = c_{k,t+1} > m_{k,t+1} + 1$, i.e, $m_{k,t} > m_{k,t+1}$, a contradiction.

Case 2. Suppose $t \in U_k$. Then $c_{k,t} = m_{k,t}$, and $m_{k,t} + 1 = c_{k,t} + 1 = c_{k,t+1} > m_{k,t+1} + 1$, and the contradiction is as before. □

The next corollary follows immediately from Lemma 13.

Corollary 18. *For all j, there is exactly one n such that $\mathfrak{Y}_{k,n} \neq \emptyset$ and $card(U_k \cap [1, n)) = j - 1 = m_n$ (namely $n = u_j$).*

4. Main Results

We can now state our major result.

Theorem 19. *$n + 1 \in U_k$ if and only if $\left(\lfloor k\sqrt{n+1}\rfloor = \lfloor k\sqrt{n}\rfloor \text{ and } n \equiv \lfloor k\sqrt{n}\rfloor \ (mod\ 2)\right)$.*

Proof. ($\Rightarrow$) By the proof of Lemma 8, $\left(\lfloor k\sqrt{n+1}\rfloor = \lfloor k\sqrt{n}\rfloor\right.$ and $\left.n \equiv \lfloor k\sqrt{n}\rfloor \pmod 2\right)$ if and only if $\sigma_{k,n+1} = \sigma_{k,n} - 1$, and, if $n + 1 \in U_k$, then $\sigma_{k,n+1} = \sigma_{k,n} - 1$, by Lemma 8 and the Claim of Lemma 14.

($\Leftarrow$) Suppose $\left(\lfloor k\sqrt{n+1}\rfloor = \lfloor k\sqrt{n}\rfloor\right.$ and $\left.n \equiv \lfloor k\sqrt{n}\rfloor \pmod 2\right)$, i.e., $\sigma_{k,n+1} = \sigma_{k,n} - 1$. Then $\sigma_{k,n} = \lfloor k\sqrt{n}\rfloor + 2$, $\sigma_{k,n+1} = \lfloor k\sqrt{n}\rfloor + 1$. Since $\sigma_{k,n} \neq \lfloor k\sqrt{n}\rfloor + 1$, by Lemma 12, $n \notin U_k$. Note that $m_{k,n+1} = \frac{n+1-\sigma_{k,n+1}}{2} = \frac{n+1-\sigma_{k,n}+1}{2} = \frac{n-\sigma_{k,n}}{2} + 1 = m_{k,n} + 1$. Since $n \notin U_k$, $c_{k,n} = c_{k,n+1}$. Also, by Lemmas 15, 16, 17, $m_{k,n} + 1 = c_{k,n}$. Thus $m_{k,n+1} = m_{k,n} + 1 = c_{k,n} = c_{k,n+1}$, and so, by Lemma 15, again, $n + 1 \in U_k$. □

As already noted, Theorem 19 gives a purely arithmetical condition on n equivalent to $n + 1 \in U_k$. As is clear from the proof, the Theorem can be reformulated as $n + 1 \in U_k$ if and only if $\sigma_{k,n+1} < \sigma_{k,n}$.

Corollary 20. *Suppose $n \in U_k$.*

(a) $n+2 \in U_k$ *if and only if* $\lfloor k\sqrt{n}\rfloor = \lfloor k\sqrt{n+2}\rfloor$.

(b) *If* $n + 2 \notin U_k$ *then*

(i) $n+3 \in U_k$ *if and only if* $\lfloor k\sqrt{n+3}\rfloor = \lfloor k\sqrt{n}\rfloor + 1$,

(ii) $n+4 \in U_k$ *if and only if* $\lfloor k\sqrt{n+3}\rfloor = \lfloor k\sqrt{n}\rfloor + 2$.

Proof. For (a), suppose $n \in U_k$.

($\Rightarrow$) If also $n+2 \in U_k$, then we are in the case where $\sigma_{k,n+2} = \sigma_{k,n+1} - 1$ and $\lfloor k\sqrt{n+2} \rfloor = \lfloor k\sqrt{n+1} \rfloor \not\equiv n+2 \pmod 2$. Since $n \in U_k$, we also have $\sigma_{k,n} = \sigma_{k,n-1} - 1$ and $\lfloor k\sqrt{n} \rfloor \not\equiv n \pmod 2$. If $\lfloor k\sqrt{n} \rfloor \neq \lfloor k\sqrt{n+1} \rfloor$ then we would have $\lfloor k\sqrt{n+1} \rfloor \equiv n \equiv n+2 \pmod 2$, a contradiction. So $\lfloor k\sqrt{n} \rfloor = \lfloor k\sqrt{n+1} \rfloor = \lfloor k\sqrt{n+2} \rfloor$.

($\Leftarrow$) Let $\lfloor k\sqrt{n} \rfloor = \lfloor k\sqrt{n+2} \rfloor$. Since $n \in U_k$, we are in the case where $\sigma_{k,n} = \sigma_{k,n-1} - 1$ and $\lfloor k\sqrt{n} \rfloor \not\equiv n \pmod 2$. Then $\lfloor k\sqrt{n+2} \rfloor \not\equiv n \pmod 2$ and $\sigma_{k,n+2} = \lfloor k\sqrt{n+2} \rfloor + 1 = \sigma_{k,n}$. Since there are never two consecutive successes,

$\operatorname{card}(U_k \cap [1, n+2)) = \operatorname{card}(U_k \cap [1, n)) + 1 = \frac{n-\sigma_{k,n}}{2} + 1 = \frac{n+2-\sigma_{k,n+2}}{2}$, so $n+2 \in U_k$.

For (b), suppose $n \in U_k$ but $n+2 \notin U_k$. Since $n \in U_k$, $\lfloor k\sqrt{n} \rfloor \not\equiv n \bmod 2$ and $\sigma_{k,n} = \lfloor k\sqrt{n} \rfloor + 1$.

For (i): ($\Rightarrow$) Suppose $n+3 \in U_k$. Then, by Theorem 19, $\lfloor k\sqrt{n+2} \rfloor = \lfloor k\sqrt{n+3} \rfloor$. So the greatest integer does not jump at $n+2$. Since $n+2 \notin U_k$, by (a), above, $\lfloor k\sqrt{n} \rfloor \neq \lfloor k\sqrt{n+2} \rfloor$, so there is a jump at n or $n+1$. Since there is no jump at $n+2$, thus there is a jump only at n or $n+1$, i.e., $\lfloor k\sqrt{n+3} \rfloor = \lfloor k\sqrt{n} \rfloor + 1$.

($\Leftarrow$) First suppose there is a jump only at $n+1$, i.e., $\lfloor k\sqrt{n+3} \rfloor = \lfloor k\sqrt{n+2} \rfloor = \lfloor k\sqrt{n+1} \rfloor + 1 = \lfloor k\sqrt{n} \rfloor + 1 \not\equiv n+1 \pmod 2$. Then we have $\lfloor k\sqrt{n+3} \rfloor = \lfloor k\sqrt{n+2} \rfloor \equiv n+2 \pmod 2$, so by Theorem 19, $n+3 \in U_k$.

Now suppose there is a jump only at n, i.e., $\lfloor k\sqrt{n+3} \rfloor = \lfloor k\sqrt{n+2} \rfloor = \lfloor k\sqrt{n+1} \rfloor = \lfloor k\sqrt{n} \rfloor + 1 \not\equiv n+1 \pmod 2$. Then, just as above, we have $\lfloor k\sqrt{n+3} \rfloor = \lfloor k\sqrt{n+2} \rfloor \equiv n+2 \pmod 2$, and so $n+3 \in U_k$.

For (ii): ($\Rightarrow$) Suppose $n+4 \in U_k$. Then, by Theorem 19, $\lfloor k\sqrt{n+4} \rfloor = \lfloor k\sqrt{n+3} \rfloor \equiv n+3 \pmod 2$. So the greatest integer does not jump at $n+3$. Since $n \in U_k$ and $n+2 \notin U_k$, $\lfloor k\sqrt{n+2} \rfloor \neq \lfloor k\sqrt{n} \rfloor \not\equiv n \pmod 2$, so there is a jump at n or $n+1$, and $\lfloor k\sqrt{n+2} \rfloor \equiv n \equiv n+2 \pmod 2$. Then, since $\lfloor k\sqrt{n+3} \rfloor \equiv n+3 \pmod 2$, $\lfloor k\sqrt{n+2} \rfloor \neq \lfloor k\sqrt{n+3} \rfloor$, i.e., the greatest integer jumps at $n+2$. We cannot have two consecutive jumps, so we conclude there are jumps at n and at $n+2$, i.e., $\lfloor k\sqrt{n+3}\rfloor = \lfloor k\sqrt{n}\rfloor + 2$.

($\Leftarrow$) Suppose $\lfloor k\sqrt{n+4}\rfloor = \lfloor k\sqrt{n+3}\rfloor = \lfloor k\sqrt{n+2}\rfloor + 1 = \lfloor k\sqrt{n+1}\rfloor + 1 = \lfloor k\sqrt{n}\rfloor + 2 \not\equiv n+2 \pmod 2$. Then $\lfloor k\sqrt{n+4}\rfloor = \lfloor k\sqrt{n+3}\rfloor \equiv n+3 \pmod 2$, and, by Theorem 19, $n+4 \in U_k$. □

Proposition 21. *Suppose $j, l \in \mathbb{N}^+$. Then*

(a) *for sufficiently large j, $u_{j+1} - u_j \le 3$,*
(b) $\lim_{l\to\infty} \frac{u_l}{l} = 2$.

Proof. For (a), a fairly tight lower bound is $u_j \ge \frac{9k^2}{4}$. In order to have $u_{j+1} - u_j = 4$ (a gap of four) for some $j \in \mathbb{N}^+$, letting $n = u_j$, there must be a jump at n and $n+2$; we must have $\lfloor k\sqrt{n}\rfloor + 2 = \lfloor k\sqrt{n+3}\rfloor$. We consider the necessary conditions so that the least l such that $\lfloor k\sqrt{n+l}\rfloor \ge \lfloor k\sqrt{n}\rfloor + 2$ is greater than or equal to four.

Note that $k\sqrt{n+l} = k\sqrt{n}\cdot\sqrt{1+\frac{l}{n}}$. The series expansion at 0 of $\sqrt{1+x}$, for $|x| < 1$, is 1+ the alternating series

$$\frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16} - \frac{5x^4}{128} + \frac{7x^5}{256} + O\left(x^6\right).$$

Since it is alternating, the sum is less than $1 + \frac{x}{2}$. Putting $\frac{3}{n}$ for x, we have $\sqrt{1+\frac{3}{n}} \le 1 + \frac{3}{2n}$. Then $k\sqrt{n+3} \le k\sqrt{n}\left(1+\frac{3}{2n}\right) = k\sqrt{n} + \frac{3k}{2\sqrt{n}}$. If there is a gap of four, i.e., $\lfloor k\sqrt{n}\rfloor + 2 = \lfloor k\sqrt{n+3}\rfloor$, then we will have

$$\lfloor k\sqrt{n}\rfloor + 2 \le \lfloor k\sqrt{n+3}\rfloor \le k\sqrt{n+3} \le k\sqrt{n} + \frac{3k}{2\sqrt{n}}.$$

For a contradiction, it is sufficient to have $\frac{3k}{2\sqrt{n}} \le 1$. Thus for $n \ge \frac{9}{4}k^2$ there are no more gaps of four, so for $n = u_j$ with j sufficiently large, $u_{j+1} - u_j \le 3$.

For (b), for $\frac{9k^2}{4} < j < l$, let

$$T_2(j,l) := \{i \mid j < i \le l \text{ and } u_i - u_{i-1} = 2\},$$

$$T_3(j,l) := \{i \mid j < i \le l \text{ and } u_i - u_{i-1} = 3\}.$$

Then $T_2(j,l) \cup T_3(j,l) = (j,l] \cap \mathbb{N}$. Also $u_l - u_j = 2\operatorname{card}(T_2(j,l)) + 3\operatorname{card}(T_3(j,l))$.

Claim: For any $\varepsilon > 0$ there is a j such that for all $l > j$, $\frac{\mathrm{card}(T_3(j,l))}{\mathrm{card}(T_2(j,l))} \leq \varepsilon$.

To prove the claim, fix $\varepsilon > 0$ and let $d = \frac{2}{\varepsilon} + 3$ (in fact $d \geq \frac{2}{\varepsilon} + 3$ is enough). Choose j sufficiently large so that $u_j > \frac{9k}{4}$, $u_j - u_{j-1} = 3$ and $\frac{k}{2\sqrt{u_j}} < \frac{1}{d+2}$. Let $j < l$, let $c = \mathrm{card}\,(T_3\,(j, l))$ and let $(t_i | 1 \leq i \leq c)$ be the increasing enumeration of $T_3\,(j, l)$. Also, let $t_0 = j$. Note that by Lemma 20, for all $0 \leq i < c$, $k\sqrt{u_{t_{i+1}} - 1} - k\sqrt{u_{t_i} - 3} > 1$. Therefore, by the Mean Value Theorem, for all such i, $d+2 < (u_{t_{i+1}} - 1) - (u_{t_i} - 3)$, i.e., $d < u_{t_{i+1}} - u_{t_i}$. But

$$\begin{aligned} 3c + 2\mathrm{card}\,(T_2\,(j, l)) &= u_l - u_j \\ &\geq \sum_{i=0}^{c-1} \left(u_{t_{i+1}} - u_{t_i}\right) \\ &> cd. \end{aligned}$$

So

$$\frac{c}{\mathrm{card}\,(T_2\,(j, l))} < \frac{2}{d-3} \leq \varepsilon,$$

as required.

Temporarily fixing $\varepsilon > 0$, fix a j as in the Claim. For large enough $l > j$, $\mathrm{card}\,(T_2\,(j, l))$ is large enough so that $\frac{u_j}{\mathrm{card}(T_2(j,l))} < \varepsilon$. We have

$$\begin{aligned} u_l &= u_j + u_l - u_j \\ &= u_j + 2\text{card}\left(T_2\left(j,l\right)\right) + 3\text{card}\left(T_3\left(j,l\right)\right) \\ &= \frac{u_j}{\text{card}\left(T_2\left(j,l\right)\right)}\text{card}\left(T_2\left(j,l\right)\right) + 2\text{card}\left(T_2\left(j,l\right)\right) \\ &\quad +3\frac{\text{card}\left(T_3\left(j,l\right)\right)}{\text{card}\left(T_2\left(j,l\right)\right)}\text{card}\left(T_2\left(j,l\right)\right) \\ &= \left(\frac{u_j}{\text{card}\left(T_2\left(j,l\right)\right)} + 2 + 3\frac{\text{card}\left(T_3\left(j,l\right)\right)}{\text{card}\left(T_2\left(j,l\right)\right)}\right)\text{card}\left(T_2\left(j,l\right)\right) \\ &< \left(\varepsilon + 2 + 3\varepsilon\right)\text{card}\left(T_2\left(j,l\right)\right) \end{aligned}$$

So $u_l < (4\varepsilon + 2)\,\text{card}\left(T_2\left(j,l\right)\right)$, and since $l > l - j > \text{card}\left(T_2\left(j,l\right)\right)$, $\frac{u_l}{l} < \frac{u_l}{\text{card}(T_2(j,l))} < 4\varepsilon + 2$. So for sufficiently large l, $\frac{u_l}{l} < 4\varepsilon + 2$. This is true for any ε, so $\lim_{l\to\infty} \frac{u_l}{l} = 2$. □

Numerical calculation has shown that (even for $k = 3$, for example), gaps of four do occur and in fact, early on, gaps of three and four predominate, but the gaps of four disappear fairly quickly, and eventually, the gaps of two predominate.

References

[1] Richard M. Dudley. *Real Analysis and Probability.* Wadsworth & Brooks/Cole Mathematics Series (1989).

[2] Erdos, On the Law of the Iterated Logarithm, \textit{Annals of Mathematics}, \textbf{43} (1942) 419-436.

[3] P. Erdös and P. Revesz. *On the favourite points of a random walk.* Mathematical Structures, Computational Mathematics, Mathematical Modelling 2, (1984) 152-157.

[4] W. Feller. *The General Form of the So-Called Law of the Iterated Logarithm.* Transactions of the American Mathematical Society, **54** (1943) 373-402

[5] A. Khintchine. *Über dyadische Brüche.* Math. Zeit., **18** (1923) 109-116

[6] Z. Shi and B. Tóth. *Favourite sites of simple random walk.* Periodica Mathematica Hungarica, **41** (2000) 237–249.

[7] M. Skyers and L. Stanley, *Measuring the fine structure of the random walk on (0,1), gain and loss*, preprint (2018).

Department of Mathematics, Penn State Brandywine, Media, Pennsylvania
Email address: mus61@psu.edu

Department of Mathematics, Lehigh University, Bethlehem, Pennsylvania
Email address: ljs4@lehigh.edu