ASYMPTOTIC FLUCTUATIONS OF THE STANDARD RANDOM WALK

XIAOTONG YANG

Contents

Date: Started April 12, 2022. Last revised April 11, 2023. Advisor: Liviu I. Nicolaescu.

Notation and conventions

- We will denote by $|S|$ of $\#S$ the cardinality of a finite set S.
- We denote by $\mathbb{N}$ the set of natural numbers, $\mathbb{N} = \{1, 2, \dots\}$ and we set $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.
- For any set X we denote by 2^X the collection of all the subsets of X
- For any set S, contained in some ambient space X, we denote by I_S the *indicator function* of S

$$I_S : X \to \{0, 1\}, \;\; I_S(x) = \begin{cases} 1, & x \in S, \\ 0, & x \notin S. \end{cases}$$

- We will denote by $\mathbb{E}\big[\, X \,\big]$ the expectation of a random variable X any by $\mathbb{P}\big[\, S \,\big]$ the probability of an event S.
- $f(t) = O\big(\, g(t) \,\big)$ as $t \to t_0$ if $\exists C > 0$ such that $\big|\, f(t) \,\big| \le C \big|\, g(t) \,\big|$ for t close to t_0.
- $f(t) = o\big(\, g(t) \,\big)$ as $t \to t_0$ if

$$\lim_{t \to t_0} \frac{f(t)}{g(t)} = 0.$$

- $f(t) \sim g(t)$ as $t \to t_0$ if

$$\lim_{t \to t_0} \frac{f(t)}{g(t)} = 1.$$

Introduction

The standard random walk is easy to describe. Flip a fair coin with faces labelled ± 1. If the face 1 shows up take a step of size 1 in the positive direction (along the x-axis) while if the face -1 shows up take a step of size 1 in the negative direction. We denote by S_n your location after n steps. The random walk is formally the sequence of random variables $S_0, S_1, S_2, \dots,$. Throughout we assume that the walk starts at the origin of the x-axis, i.e., $S_0 = 0$. If we think that we flip the coin once per unit of time, we can also interpret n as measuring time.

This simple description makes the standard random walk amenable to combinatorial methods of investigation. The goal of this thesis is to describe a few less advertised asymptotic results concerning the fluctuations of the standard random walk. In the process we will reveal a few counterintuitive results that, surprisingly, occur in more general situations.

The random walk obviously displays the Markov property: its future behavior is independent of the past, given the present. For example, if the walk returns to 0 after a number of coin flips, then the future behavior is as if we start the walk anew, and the history of how the walk wandered back to 0 is irrelevant. Note that since the step sizes are the odd integers ± 1, the location S_n has the same parity as the number n of flips. Hence the returns to the origin can only occur at even moments of time.

A key character in our story is the epoch T_0 of the the first return to the origin

$$T_0 = \min \left\{ n \in \mathbb{N};\ \ S_n = 0 \right\}.$$

This is a random variables that takes only even positive integral values. It could also be infinite. We determine its distribution in Lemma 1.5. This computation shows that the probability that $T_0 = \infty$ is 0. This proves that the random walk is recurrent: if we continue flipping coins indefinitely the random walk will keep returning to 0, over, and over. again.

We base our computation of the distribution of T_0 on a remarkable combinatorial result, Lemma 1.4: the probability that during the first $2n$ flips the random walk never returns to the origin is equal to the probability after $2n$ flips is back at the origin, maybe not for the first time. We denote by u_{2n} the probability that the random walk returns to the origin after $2n$ flips.

The proof of Lemma 1.4 is based on a very ingenious technique usually referred to as the *reflection principle.*

Our first asymptotic result concerns the last return to 0,

$$R_{2n} = \max \left\{ k;\, 0 \leq 2k \leq 2n;\ \ S_{2k} = 0 \right\}.$$

Its distribution can be easily computed using Lemma 1.4 and, using Stirling approximation we are able to prove our first asymptotic result.

$$\lim_{n \to \infty} \mathbb{P}\left[\frac{R_{2n}}{2n} \leq x \right] = \frac{2}{\pi} \arcsin \sqrt{x}.$$

Traditionally, this result is stated as saying the the random variable $\frac{R_{2n}}{2n}$ converges in distribution to the arcsine distribution.

The graph of the probability density of the arcsine distribution is depicted in Figure 2. It shows that the last return to 0 is more likely to occur either very early or very wait in the sequence of $2n$ flips.

It is convenient to think of the random walk in terms of a fair game of chance between two players A and B.

Suppose we have a fair coin with faces labelled ± 1. We flip the coin and, if the face 1 shows up, player A gets \$1 from B. If -1 shows up, A gives \$1 to B. The quantity S_n is then the amount A won/lost after n coin flips. If $S_n > 0$, then A is in the lead, i.e., his fortune increased S_n dollars, while if $S_n < 0$, then B is in the lead by $|S_n|$ dollars, or, equivalently A lost $|S_n|$ dollars to B. We set

$$A_n := \#\big\{k;\ \ 1 \leq k \leq n,\ \ \max(S_{k-1}, S_k) > 0\,\big\}.$$

The quantity $\frac{A_n}{n}$ indicates the amount of time player A is in the lead during a game of duration n.

We present a result of Chung and Feller (Theorem 2.3) stating that that the random variable A_{2n} has the same distribution as R_{2n}. In particular, this shows that the fraction of time $\frac{A_{2n}}{2n}$ player A is in the lead also converges in distribution to the arcsine distribution. Given the shape (Figure 2) of the arcsine distribution this is a bit counterintuitive. This result states that in the long run it is very improbably that A will be in the lead half of the time as one would expect from a fair gain. Instead, it is more likely that he will be in the lead very little or very long. This is so much more surprising given another result that we prove, Theorem 2.13. This states that if we know that after $2n$ games ($n \gg 1$) no player won or lost anything (S_{2n}), then the fraction of times player A was in the lead is no longer arcsine distributed, but uniformly distributed!

It turns out that the arcsine distribution lurks behind another random quantity. Denote by M_n the first moment in a string of n coin flips when payer A reaches the highest fortune during this game of duration n, i.e.,

$$M_n := \min\big\{\,0 \leq m \leq n;\ \ S_m = \max\big(\,S_0, S_1, \dots, S_m\,\big)\,\big\}.$$

In Theorem 2.12 we show that the fraction of time it takes player A to reach its maximum fortune during a long game also approaches the arcsine distribution.

We conclude with a section dedicated to the Central Limit Theorem (or de Moivre's formulæ) satisfied by S_n. Again, our approach is elementary, based on Stirling's formula.

1. The standard random walk

1.1. The basics.

The standard random walk is a sequence of random sums

$$S_n = X_1 + \cdots + X_n,\ \ n \in \mathbb{N},$$

where X_n are independent Rademacher variables, i.e., random variables that take only the values ± 1 with equal probabilities. A random walk of length/duration n is a finite sequence of random sums

$$S_1 = X_1, \dots, S_k = X_1 + \cdots + S_k, \dots, S_n = X_1 + \cdots + X_n.$$

Intuitively, think that at each epoch $n \in \mathbb{N}$ we flip a fair coin with faces ± 1. The random variable X_n describes the result of the n-th flip. At time 0 we are located at the origin of the real axis. After each flip with take a step of size 1 in the positive/negative direction according to the sign of the flip. S_n is the location after n flips. A random walk of length n can be identified with a random element of $\mathscr{W}_n = \{-1, 1\}^n$.

Equivalently, we can think of the standard walk in terms of a sequence of fair games between two players A and B. After the n-th flip of the coin player P get \$1 from B if $X_n = 1$, or B gets \$1 from A, if $X_n = -1$. Then S_n denotes the amount of money A has won from B if $S_n \geq 0$ or lost to B is $S_n < 0$.

Proposition 1.1. *For each n, the location S_n has the same parity as n, and*

$$\mathbb{P}\big[\, S_n = s \,\big] = \frac{1}{2^n}\binom{n}{\frac{n+s}{2}}. \tag{1.1}$$

Proof. We follow the approach in [2, Sec. 3.2]. Since $X_k \equiv 1 \bmod 2$ we deduce $S_n = X_1 + \cdots + X_n \equiv n \bmod 2$. There are 2^n walks of duration n. The final location S_n is determined by the number p of positive steps, or the number of times we flipped 1 during the first n steps. We flipped -1's $q = n - p$ times. Then the location y at epoch n is

$$s = p - (n - p) = 2p - n.$$

We deduce that $p = \frac{n+s}{2}$. Thus, of the 2^n paths of duration n, exactly $\binom{n}{\frac{n+s}{2}}$ have final location y. □

Note that if the random walk returns to the origin at an epoch n, then n must be even, $n = 2k$. We set

$$u_{2k} := \mathbb{P}\big[\, S_{2k} = 0 \,\big].$$

The equality (1.1) implies

$$u_{2k} = \frac{1}{2^{2k}}\binom{2k}{k}. \tag{1.2}$$

1.2. **André's reflection principle and the ballot theorem.** As it is customary, we will refer to the points in $\mathbb{R}^2$ with integral coordinates as *lattice points*. Let us define a *zig-zag of length n* to be a sequence of lattice points

$$\boldsymbol{z} = (P_0, P_1, \dots, P_n), \;\; P_k = (x_k, y_k),$$

such that

$$x_k - x_{k-1} = 1, \;\; y_k - y_{k-1} = \pm 1, \;\; \forall k.$$

We connect successive points P_{k-1}, P_k by a line segment so we can visualize a zig-zag as a piecewise linear curve with edges of slopes ± 1; see Figure 1.

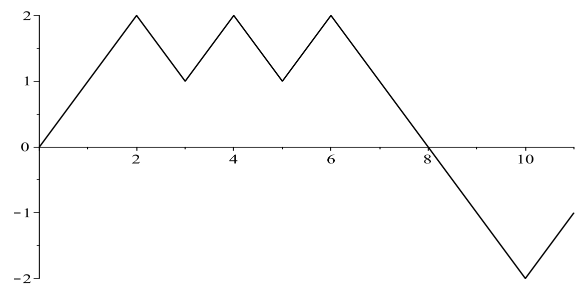

FIGURE 1. *A zig-zag describing a random walk started at* $S_0 = 0$

We will refer to the points P_k as the vertices of the zig-zag. The point P_0 is called the initial point and the point P_n the final point of the zig-zag. We denote by $\mathscr{Z}$ the collection of zig-zags.

For each $P = (x, y) \in \mathbb{R}^2$ we denote by $\bar{P}$ its reflection in the x-axis, $\bar{P} = (x, -y)$. To a zig-zag

$$\boldsymbol{z} = (P_0, P_1, \dots, P_n)$$

we can associate its reflection in the x-axis

$$\bar{\boldsymbol{z}} = \big(\bar{P}_0, \bar{P}_1, \dots, \bar{P}_n\big).$$

Clearly $\bar{\boldsymbol{z}}$ is also a zig-zag.

To a random walk of duration n we associate the zig-zag from $(0, 0)$ to (n, S_n) with vertex points (k, S_k), $0 \leq k \leq n$.

Suppose that $A = (a_1, a_2)$, $B = (b_1, b_2)$ are lattice points. Denote by $\mathscr{Z}(A, B)$ the set of zig-zags from A to B. We set

$$n = n(A, B) = a_2 - a_1 > 0, \;\; s = s(A, B) = b_2 - b_1.$$

Note that

$$\mathscr{Z}(A, B) \neq \emptyset \iff \big|\, s \,\big| \leq n.$$

We denote by $N(A, B)$ the number of zig-zags from A to B, i.e., the cardinality of $\mathscr{Z}(A, B)$. When $A = (0, 0)$ and $B = (n, s)$ we will use the simpler notation

$$N_{n,s} := N(A, B).$$

We denote by $p_{n,s}$ the probability that a zigzag of length n started at the origin ends at the point (n, s). Note that

$$p_{n,s} = \frac{N_{n,s}}{2^n}.$$

Note that there exists a bijection between zig-zags from $(0, 0)$ to (n, s) and random walks of duration n such that $S_n = s$. The argument in the proof of Proposition 1.1 shows that

$$N(A, B) = N_{n,s} = \binom{n}{(n+s)/2}, \;\; n = n(A, B), \;\; s = s(A, B). \tag{1.3}$$

Note that if $A, B, C \in \mathbb{Z}^2$ are such that $\mathscr{Z}(A,B), \mathscr{Z}(B,C) \neq \emptyset$, then we have a *concatenation* operation

$$* : \mathscr{Z}(A,B) \times \mathscr{Z}(B,C) \to \mathscr{Z}(A,C)$$

that associates to a zig-zag $\boldsymbol{z}$ from A to B and a zig-zag $\boldsymbol{z}'$ from B to C the zig-zag $\boldsymbol{z} * \boldsymbol{z}'$ obtained by gluing the end point of $\boldsymbol{z}$ with the initial point of z'

Theorem 1.2 (André's Reflection Principle)**.** *Let $A = (a_1, a_1)$, $B = (b_1, b_2)$ be two lattice points situated on the same side of the x-axis, but not on the x axis. Denote by $\mathscr{Z}^x(A,B)$ the set of zig-zags from A to B that touch the x axis. Then*

$$\#\mathscr{Z}^x(A,B) = \#\mathscr{Z}\big(\bar{A}, B\big),$$

where $\bar{A}$ denotes the reflection of A in the x-axis.

Proof. The proof is inspired by [2, Sec. 3.2]. We will construct a bijection

$$\Phi : \mathscr{Z}^x(A,) \to \mathscr{Z}(\bar{A}, B).$$

Given a zig-zag $\boldsymbol{z} \in \mathscr{Z}^x(A,B)$ we denote by $P_{\boldsymbol{z}}$ the first point on the x-axis touched by $\boldsymbol{z}$. The point $P_{\boldsymbol{z}}$ splits the zig-zag $\boldsymbol{z}$ into two zigzags: a zig-zag z_1 from A to $P_{\boldsymbol{z}}$ and the rest, a zig-zag z_2 from $P_{\boldsymbol{z}}$ to B such that $\boldsymbol{z}_1 * \boldsymbol{z}_2$. Define

$$\Phi(\boldsymbol{z}) := \bar{\boldsymbol{z}}_1 * \boldsymbol{z}_2$$

We claim that this map is a bijection.

Injectivity. Suppose that $\boldsymbol{z}, \boldsymbol{z}' \in \mathscr{Z}^x(A,B)$ such that $\Phi(\boldsymbol{z}) = \Phi(\boldsymbol{z}')$. Note first that $P_{\boldsymbol{z}}$ is the first point where the zig-zag $\Phi(\boldsymbol{z})$ touches the x-axis. This means that $P_{\boldsymbol{z}'} = P_{\boldsymbol{z}}$ We deduce that

$$\bar{\boldsymbol{z}}_1 = \bar{\boldsymbol{z}}_1', \;\; \bar{\boldsymbol{z}}_2 = \bar{\boldsymbol{z}}_2' \implies \boldsymbol{z} = \boldsymbol{z}'.$$

Surjectivity. Let $\zeta \in \mathscr{Z}(\bar{A}, B)$. Denote by P_ζ the first point on the x-axis touched by ζ. We split similarly

$$\zeta = \zeta_1 * \zeta_2$$

where ζ_1 is the part of ζ from $\bar{A}$ to P_ζ. Observe that

$$\zeta = \Phi(\bar{\zeta}_1 * \zeta_2).$$

□

Theorem 1.3 (Ballot theorem)**.** *Let $(n,s) \in \mathbb{N}^2$. Then, the number $B_{n,s}$ of zig-zags of length n from $(0,0)$ to (n,s) that do not touch or cross the x-axis equals*

$$B_{n,s} = N_{n-1,s-1} - N_{n-1,s+1} = \frac{s}{n}\binom{n}{\frac{n+s}{2}} = \frac{s}{n} N_{n,s}. \tag{1.4}$$

Proof. We follow the strategy in [2, Sec. 3.2]. Observe that $S_1 = \pm 1$. Denote by $\mathscr{Z}_{n,s}$ the set of zig-zags from $O = (0,0)$ to $B = (n,s)$. Denote $\mathscr{Z}^{\pm}_{n,s}$ the set of zig-zags $\boldsymbol{z} \in \mathscr{Z}_{n,s}$ that go through $A_\pm := (1, \pm 1)$.

$$\#\mathscr{Z}_{n,s} = \#\mathscr{Z}^+_{n,s} + \#\mathscr{Z}^-_{n,s}.$$

The number $B_{n,s}$ of zig-zags O to B that do not touch or cross the x-axis is equal to the number of zig-zags in $\#\mathscr{Z}^+_{n,s}$ that do not touch or cross the x-axis. This is exactly the number of zig-zags from A_+ to B that do not touch the x axis. From the Reflection Principle we deduce that

$$\begin{aligned} B_{n,s} &= \#\mathscr{Z}(A_+, B) - \#\mathscr{Z}^x(A_+, B) = \#\mathscr{Z}(A_+, B) - \#\mathscr{Z}(\bar{A}_+, B) \\ &= N(A_+, B) - N(\bar{A}_+, B). \end{aligned}$$

Note that

$$n(A_+, B) = n - 1 = n(\bar{A}_+, B), \;\; s(A_+, B) = s - 1, \;\; s(\bar{A}_+, B) = s + 1.$$

Using (1.3) we deduce

$$N(A, B) = \binom{n-1}{\frac{n+s}{2} - 1}, \;\; N(\bar{A}_+, B) = \binom{n-1}{\frac{n+s}{2}}.$$

Thus

$$\begin{aligned} B_{n,s} &= N_{n-1,s-1} - N_{n-1,s+1} = \binom{n-1}{\frac{n+s}{2} - 1} - \binom{n-1}{\frac{n+s}{2}} \\ &= \frac{(n-1)!}{(\frac{n+s}{2} - 1)!(n - \frac{n+s}{2})!} - \frac{(n-1)!}{(\frac{n+s}{2})!(n - \frac{n+s}{2} - 1)!} \\ &= \frac{(n-1)!}{(\frac{n+s}{2})!(n - \frac{n+s}{2})!}\left(\frac{n+s}{2} - \big(n - \frac{n+s}{2}\big)\right) = \frac{s}{n}\binom{n}{\frac{n+s}{2}} = \frac{s}{n} N_{n,s}. \end{aligned}$$

□

1.3. Returns to the origin.

Lemma 1.4. *The probability that a zig-zag of length $2n$ never returns to the origin equals the probability that a zig-zag of length $2n$ returns to the origin at epoch $2n$. i.e.*

$$\mathbb{P}\big[\, S_1 \neq 0, S_2 \neq 0, \dots, S_{2n} \neq 0 \,\big] = \mathbb{P}\big[\, S_{2n} = 0 \,\big] = u_{2n}. \tag{1.5}$$

Proof. We argue as in [2, Sec. 3.3]. We have

$$\begin{aligned} &\mathbb{P}\big[\, S_1 \neq 0, S_2 \neq 0, \dots, S_{2n} \neq 0 \,\big] \\ &= \mathbb{P}\big[\, S_1 > 0, \dots, S_{2n} > 0 \,\big] + \mathbb{P}\big[\, S_1 < 0, \dots, S_{2n} < 0 \,\big]. \end{aligned}$$

By symmetry, the lemma is equivalent to

$$\mathbb{P}\big[\, S_1 > 0, \cdots, S_{2n} > 0 \,\big] = \frac{1}{2} u_{2n}.$$

Observe that

$$\mathbb{P}\big[\, S_1 > 0, \dots, S_{2n} > 0 \,\big] = \sum_{a=1}^{n} \mathbb{P}\big[\, S_1 > 0, \dots, S_{2n-1} > 0, S_{2n} = 2a \,\big].$$

The number of zig-zags that satisfy $S_1 > 0, \dots, S_{2n-1} > 0, S_{2n} = 2a$ is equivalent to number $B_{2n,2a}$ of zig-zags of length n from $(0,0)$ to $(2n, 2a)$ that do not touch or cross the x-axis.

We have

$$\begin{aligned}\mathbb{P}\big[\, S_1 > 0, \dots, S_{2n-1} > 0, S_{2n} = 2a \,\big] &= \frac{B_{2n,2a}}{2^n} \\ &\overset{(1.4)}{=} \frac{N_{2n-1,2a-1} - N_{2n-1,2a+1}}{2^n}.\end{aligned} \tag{1.6}$$

Hence

$$\sum_{a=1}^{n} \mathbb{P}\big[\, S_1 > 0, \dots, S_{2n-1} > 0, S_{2n} = 2a \,\big] = \sum_{a=1}^{n} \frac{N_{2n-1,2a-1} - N_{2n-1,2a+1}}{2^n}$$

$$= \frac{N_{2n-1,1} - N_{2n-1,2n+1}}{2^n} = \frac{N_{2n-1,1} - 0}{2^n} = \frac{1}{2}\mathbb{P}\big[\, S_{2n-1} = 1 \,\big].$$

Therefore,

$$\mathbb{P}\big[\, S_1 > 0, S_2 > 0, \cdots, S_{2n} > 0 \,\big] = \frac{1}{2}\mathbb{P}\big[\, S_{2n-1} = 1 \,\big].$$

Since $S_{2n-1} = 1$ and $S_{2n-1} = -1$ are both one flip away from reaching $S_{2n} = 0$ we deduce

$$u_{2n} = \mathbb{P}\big[\, S_{2n} = 0 \,\big] = \frac{1}{2}\mathbb{P}\big[\, S_{2n-1} = 1 \,\big] + \frac{1}{2}\mathbb{P}\big[\, S_{2n-1} = -1 \,\big].$$

By symmetry,

$$\mathbb{P}\big[\, S_{2n-1} = 1 \,\big] = \mathbb{P}\big[\, S_{2n-1} = -1 \,\big]$$

so that

$$\mathbb{P}\big[\, S_{2n-1} = 1 \,\big] = u_{2n}.$$

We conclude that

$$\mathbb{P}\big[\, S_1 > 0, S_2 > 0, \cdots, S_{2n} > 0 \,\big] = \frac{1}{2}\mathbb{P}\big[\, S_{2n-1} = 1 \,\big] = \frac{1}{2}u_{2n}, \tag{1.7}$$

and thus

$$\mathbb{P}\big[\, S_1 \neq 0, S_2 \neq 0, \cdots, S_{2n} \neq 0 \,\big] = u_{2n}.$$

□

Let

$$T_0 := \min\big\{\, r \in \mathbb{N};\ \ S_r = 0 \,\big\}.$$

Thus, T_0 is the time of the first return to the origin. Note that T_0 is necessarily an even number, possibly infinite. We denote by f the probability distribution of the random variable T_0. Thus

$$f_r = \mathbb{P}\big[\, T_0 = r \,\big].$$

As observed, f_r is zero if r is odd.

Lemma 1.5. *The probability that the first return at the origin happens at time $2n$ is*

$$f_{2n} = \frac{1}{2n-1}u_{2n} = \frac{1}{(2n-1)2^n}\binom{2n}{n} \tag{1.8}$$

Proof. We follow the proof in [2, Sec. 3.3]. By (1.5),

$$\begin{aligned} f_{2n} &= \mathbb{P}\big[\, S_1 \neq 0, S_2 \neq 0, \cdots, S_{2n-2} \neq 0 \,\big] - \mathbb{P}\big[\, S_1 \neq 0, S_2 \neq 0, \cdots, S_{2n} \neq 0 \,\big] \\ &= u_{2n-2} - u_{2n} = \frac{1}{2^{2n-2}}\binom{2n-2}{n-1} - \frac{1}{2^{2n}}\binom{2n}{n} \\ &= \frac{1}{2^{2n}}\binom{2n-2}{n-1}\left(4 - \frac{2n(2n-1)}{n^2}\right) = \frac{1}{2^{2n}}\binom{2n-2}{n-1}\frac{2n}{n^2} \\ &= \frac{1}{2^{2n}(2n-1)}\binom{2n}{n} = \frac{1}{2n-1}u_{2n}. \end{aligned}$$

□

Lemma 1.6. *The probability that a zig-zag of infinite length never returns to the origin*

$$\lim_{n\to\infty} \mathbb{P}\big[\, S_1 \neq 0, \dots, S_{2n} \neq 0 \,\big] = 0. \tag{1.9}$$

Proof.

$$\begin{aligned} \mathbb{P}\big[\, S_1 \neq 0, \dots, S_{2n} \neq 0 \,\big] &\overset{(1.5)}{=} u_{2n} = \frac{1}{2^{2n}}\binom{2n}{n} = \frac{1}{2^{2n}}\frac{(2n)!}{n!n!} \\ &= \frac{(2n)!}{(2n)!!(2n)!!}] = \frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2n-1}{2n} \end{aligned}$$

Since

$$\frac{m}{m+1} < \frac{m+1}{m+2}, \quad \forall m \in \mathbb{N},$$

we deduce,

$$\left(\frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2n-1}{2n}\right)^2 < \left(\frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2n-1}{2n}\right)\cdot\left(\frac{2}{3}\cdot\frac{4}{5}\cdots\frac{2n}{2n+1}\right) = \frac{1}{2n+1}.$$

Therefore,

$$\lim_{n\to\infty}\frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2n-1}{2n} = 0.$$

□

2. Arcsine Laws

2.1. **Time of last return to the origin.** Consider the standard random walk $(S_n)_{n\geq 0}$. For $n \in \mathbb{N}$ we set

$$R_{2n} := \max\big\{\, 0 \leq k \leq n; \ \ S_{2k} = 0 \big\},$$

and

$$\alpha_{k,n-k} := \mathbb{P}\big[\, R_{2n} = 2k \,\big].$$

Lemma 2.1. *The probability that the* last *return to the origin of a zig-zag of length 2n happens at epoch 2k.*

$$\alpha_{k,n-k} = \mathbb{P}\big[\, S_{2k} = 0, S_{2k+1} \neq 0, \dots, S_{2n} \neq 0 \,\big] = u_{2k}u_{2n-2k} \tag{2.1}$$

Proof. The proof is from [2, Sec. 3.4].

$$\begin{aligned}\alpha_{k,n-k} &= \mathbb{P}\big[\, S_{2k}=0\,\big]\mathbb{P}\big[\, S_{2k+1}\neq 0,\dots,S_{2n}\neq 0\,\big]\\ &= u_{2k}\mathbb{P}\big[\, S_{2k+1}\neq 0,\dots,S_{2n}\neq 0\,\big]\end{aligned}$$

By (1.5),

$$\mathbb{P}\big[\, S_{2k+1}\neq 0,\dots,S_{2n}\neq 0\,\big] = u_{2n-2k}.$$

Thus,

$$\alpha_{k,n-k} = u_{2k}u_{2n-2k}.$$

□

Theorem 2.2 (Arcsine Law for Last Visits)**.** *For any* $0\leq a<b\leq 1$ *we have*

$$\lim_{n\to\infty}\mathbb{P}\Big[\, a\leq \frac{R_{2n}}{2n}\leq b\,\Big] = \frac{2}{\pi}\arcsin\sqrt{b}-\frac{2}{\pi}\arcsin\sqrt{a}. \tag{2.2}$$

i.e.,

$$\lim_{n\to\infty}\sum_{an\leq k\leq bn}\alpha_{k,n-k} = \frac{2}{\pi}\arcsin\sqrt{b}-\frac{2}{\pi}\arcsin\sqrt{a}. \tag{2.3}$$

Proof. We follow the approach in [2, Sec. 3.4]. Set $Q_n := \frac{R_{2n}}{2n}$ We will carry the proof in three steps.

Step 1. Assume $0<a<b<1$. We deduce from Stirling's Formula (A.2) that as $m\to\infty$ we have

$$\begin{aligned} u_{2m} &= \frac{1}{2^{2m}}\binom{2m}{m} = \frac{1}{2^{2m}}\frac{(2m)!}{(m!)^2}\\ &= \frac{1}{\sqrt{\pi m}}\big(\,1+O(1/m)\,\big) = \frac{1}{\sqrt{\pi m}}+O\big(\,1/m^{3/2}\,\big).\end{aligned} \tag{2.4}$$

Fix $0<a<b<1$. Let k be even integer in $\big[\,na,nb\,\big], 0\leq a\leq b\leq 1$. For n sufficiently large,

$$\frac{1}{k}=O(1/n),\quad \frac{1}{n-k}=O(1/n),$$

$$\begin{aligned}\alpha_{k,n-k} &= u_{2k}u_{2n-2k}\\ &= \left(\frac{1}{\sqrt{\pi k}}+O(1/k^{3/2})\right)\cdot\left(\frac{1}{\sqrt{\pi(n-k)}}+O(1/(n-k)^{3/2})\right)\\ &= \frac{1}{\pi\sqrt{k(n-k)}}+O(1/n^2).\end{aligned} \tag{2.5}$$

$$\mathbb{P}\big[\,2\lceil na\rceil\leq R_{2n}\leq 2\lfloor nb\rfloor\,\big] = \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}u_{2k}u_{2n-2k}$$

$$= \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} \Big(\frac{1}{\pi\sqrt{k(n-k)}} + O(1/n^2) \Big)$$

(the above sum consists of $O(n)$ terms)

$$= \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} \frac{1}{\pi\sqrt{k(n-k)}} \;+\; O(1/n) = \sum_{a\leq \frac{k}{n}\leq b} \frac{\frac{1}{n}}{\pi\sqrt{\frac{k}{n}(1-\frac{k}{n})}} + O(1/n).$$

The last sum is a Riemann sum approximation of the integral

$$\int_a^b \frac{1}{\pi\sqrt{x(1-x)}}dx.$$

Thus, if $0 < a < b < 1$ then, as $n \to \infty$,

$$\mathbb{P}\Big[\, a \leq Q_n \leq b\,\Big] \sim \int_a^b \frac{1}{\pi\sqrt{x(1-x)}}dx = \frac{2}{\pi}\arcsin\sqrt{b} - \frac{2}{\pi}\arcsin\sqrt{a}.$$

Step 2. We will prove that (2.2) holds for $a = 0$ and $b = 1/2$, i.e.,

$$\lim_{n\to\infty} \sum_{0\leq k\leq n/2} \alpha_{k,n-k} = \frac{2}{\pi}\arcsin\sqrt{\frac{1}{2}} = \frac{1}{2}.$$

Set

$$x_n := \sum_{0\leq k\leq n/2} \alpha_{k,n-k}, \;\; y_n = \sum_{0\leq k< n/2} \alpha_{k,n-k}.$$

We have to prove that $x_n \to \frac{1}{2}$ as $n \to \infty$. Note that

$$z_n = x_n - y_n = \begin{cases} 0, & n \text{ odd},\\ \alpha_{n/2,n/2}, & n \text{ even}. \end{cases}.$$

If $n = 2m$, then

$$\alpha_{n/2,n/2} = u_{2m}^2 = \frac{1}{\pi m} + O(1/m^2) \;\text{ as } m \to \infty.$$

Hence $z_n \to 0$ as $n \to \infty$. Now observe that $\alpha_{n,n-k} = \alpha_{n-k,k}$ so that

$$y_n = \sum_{0\leq k<n/2} \alpha_{k,n-k} = \sum_{0\leq k<n/2} \alpha_{n-k,k}$$

(set $j = n - k$ so that $k = n - j$)

$$= \sum_{n/2<j\leq n} \alpha_{j,n-j}.$$

On the other hand

$$1 = \sum_{k=0}^{n} \mathbb{P}\big[\, R_{2n} = 2k\big] = \sum_{j=0}^{n} \alpha_{j,n-j}$$

$$= \sum_{0\leq j\leq n/2}^{n} \alpha_{j,n-j} + \sum_{n/2<j\leq n} \alpha_{j,n-j} = x_n + y_n = 2x_n - z_n$$

so that

$$x_n = \frac{1+z_n}{2} \to \frac{1}{2} \text{ as } n \to \infty.$$

Step 3. We will show that results in **Step 1** and **Step 2** allow us to conclude (2.2) is valid in the general case.

First, we show that for $a = \frac{1}{2}$ and $b = 1$, i.e.,

$$\lim_{n\to\infty} \sum_{\frac{n}{2}\le k\le n} \alpha_{k,n-k} = \frac{2}{\pi}\arcsin\sqrt{1} - \frac{2}{\pi}\arcsin\sqrt{\frac{1}{2}} = \frac{1}{2}.$$

Set

$$x'_n := \sum_{\frac{n}{2}\le k\le n} \alpha_{k,n-k},$$

$$= \sum_{0\le k<\frac{n}{2}}^{n} \alpha_{k,n-k} + \sum_{\frac{n}{2}\le k\le n} \alpha_{k,n-k} = y_n + x'_n = x_n - z_n + x'_n = 1$$

Because $x_n \to \frac{1}{2}$ and $z_n \to 0$ as $n \to \infty$, $1 - x_n + z_n = x'_n \to \frac{1}{2}$ as $n \to \infty$.

Now we consider the general case $0 \le a < b \le 1$.

When $a = 0, b = 1$,

$$\sum_{an\le k\le bn} \alpha_{k,n-k} = 1.$$

When $a = 0$, $b > \frac{1}{2}$,

$$\begin{aligned}\sum_{0\le k\le bn} \alpha_{k,n-k} &= \sum_{0\le k<\frac{n}{2}} \alpha_{k,n-k} + \sum_{\frac{n}{2}\le k\le bn} \alpha_{k,n-k} \\ &= \sum_{0\le k\le\frac{n}{2}} \alpha_{k,n-k} + \sum_{\frac{n}{2}\le k\le bn} \alpha_{k,n-k} \\ &= \frac{1}{2} + \frac{2}{\pi}\arcsin\sqrt{b} - \frac{2}{\pi}\arcsin\sqrt{\frac{1}{2}} = \frac{2}{\pi}\arcsin\sqrt{b}.\end{aligned}$$

When $a = 0$, $b < \frac{1}{2}$,

$$\sum_{0\le k\le bn} \alpha_{k,n-k} = \sum_{0\le k\le\frac{n}{2}} \alpha_{k,n-k} - \sum_{bn<k\le\frac{n}{2}} \alpha_{k,n-k}.$$

For $x \in (0,1)$ fixed and $k = xn$ we deduce from (2.5)

$$\lim_{n\to\infty} \alpha_{k,n-k} = 0.$$

Therefore,

$$\sum_{bn<k\le\frac{n}{2}} \alpha_{k,n-k} = \sum_{bn\le k\le\frac{n}{2}} \alpha_{k,n-k}.$$

Thus,

$$\sum_{0\le k\le bn}\alpha_{k,n-k}=\sum_{0\le k\le \frac{n}{2}}\alpha_{k,n-k}-\sum_{bn\le k\le \frac{n}{2}}\alpha_{k,n-k}$$
$$=\frac{1}{2}-\frac{2}{\pi}\arcsin\sqrt{\frac{1}{2}}+\frac{2}{\pi}\arcsin\sqrt{b}=\frac{2}{\pi}\arcsin\sqrt{b}.$$

When $a<\frac{1}{2}, b=1$,

$$\sum_{an\le k\le n}\alpha_{k,n-k}=\sum_{an\le k\le \frac{n}{2}}\alpha_{k,n-k}+\sum_{\frac{n}{2}<k\le n}\alpha_{k,n-k}$$
$$=\sum_{an\le k\le n}\alpha_{k,n-k}=\sum_{an\le k\le \frac{n}{2}}\alpha_{k,n-k}+\sum_{\frac{n}{2}\le k\le n}\alpha_{k,n-k}$$
$$\frac{2}{\pi}\arcsin\sqrt{\frac{1}{2}}-\frac{2}{\pi}\arcsin\sqrt{b}+\frac{1}{2}=1-\frac{2}{\pi}\arcsin\sqrt{b}.$$

When $a>\frac{1}{2}, b=1$,

$$\sum_{an\le k\le n}\alpha_{k,n-k}=\sum_{\frac{n}{2}\le k\le n}\alpha_{k,n-k}-\sum_{\frac{n}{2}\le k<an}\alpha_{k,n-k}$$
$$=\sum_{\frac{n}{2}\le k\le n}\alpha_{k,n-k}-\sum_{\frac{n}{2}\le k\le an}\alpha_{k,n-k}$$
$$=\frac{1}{2}-\frac{2}{\pi}\arcsin\sqrt{a}+\frac{2}{\pi}\arcsin\sqrt{\frac{1}{2}}=1-\frac{2}{\pi}\arcsin\sqrt{a}.$$

Therefore, for $0\le a<b\le 1$,

$$\lim_{n\to\infty}\sum_{an\le k\le bn}\alpha_{k,n-k}=\frac{2}{\pi}\arcsin\sqrt{b}-\frac{2}{\pi}\arcsin\sqrt{a}.$$

□

2.2. **Lead time.** For each zigzag of length n started at the origin we denote by A_n the number of its edges above the x-axis. Equivalently

$$A_n:=\#\big\{k;\ \ 1\le k\le n,\ \ \max(S_{k-1},S_k)>0\,\big\}. \tag{2.6}$$

The next result is due to K. L. Chung and W. Feller, [1].

Theorem 2.3. *For any $0\le k\le n$ we have*

$$\mathbb{P}\big[\,A_{2n}=2k\,\big]=\mathbb{P}\big[\,R_{2n}=2k\big]=u_{2k}u_{2n-2k}.$$

Proof. We follow the proof given in [2, Sec. 3.4]. We set $\beta_{2k,2n}:=\mathbb{P}\big[\,A_{2n}=2k\,\big]$. By symmetry,

$$\beta_{0,2n}=\beta_{2n,2n}.$$

On the other hand, all the edges of a zig-zag are above the x-axis iff $S_1\ge 0,\dots,S_{2n}\ge 0$ so

$$\beta_{2n,2n}=\mathbb{P}\big[\,S_1\ge 0,S_2\ge 0,\dots,S_{2n}\ge 0\,\big].$$

Let us observe that

$$\mathbb{P}\big[\, S_1 \geq 0, S_2 \geq 0, \dots, S_{2n} \geq 0 \,\big] = u_{2n}. \tag{2.7}$$

To see this, extend the zig-zag to $(-1,-1)$ and relabel this point as the origin. We obtain a new, longer, zigzag $(0, S_0'), (1, S_1'), \dots, (2n+1, S_{2n+1}')$,

$$S_0' = 0, \;\; S_1' = S_0 + 1 = 1, \;\; S_2' = S_1 + 1, \;\; , S_{k+1}' = S_k + 1, \dots .$$

Note that S_{2n}' is even so $S_{2n}' > 0 \implies S_{2n+1}' > 0$. Hence

$$\begin{aligned}\mathbb{P}\big[\, S_1 \geq 0, S_2 \geq 0, \dots, S_{2n} \geq 0 \,\big] &= 2\mathbb{P}\big[\, S_1' = 1, S_2' > 0, \dots, S_{2n+1}' > 0 \,\big] \\ &= 2\mathbb{P}\big[\, S_1' > 0, S_2' > 0, \dots, S_{2n}' > 0 \,\big] \overset{(1.7)}{=} u_{2n}.\end{aligned}$$

This proves (2.7).

We condition on the time T_0 of the first return to the origin. We have

$$\beta_{2k,2n} = \mathbb{P}\big[\, A_{2n} = 2k \,\big]$$

$$= \sum_{r=1}^{k} \mathbb{P}\big[\, A_{2n} = 2k \,|\, S_{2r-1} > 0, T_0 = 2r \,\big] \mathbb{P}\big[\, S_{2r-1} > 0, T_0 = 2r \,\big]$$

$$+ \sum_{r=1}^{n-k} \mathbb{P}\big[\, A_{2n} = 2k \,|\, S_{2r-1} < 0, T_0 = 2r \,\big] \mathbb{P}\big[\, S_{2r-1} < 0, T_0 = 2r \,\big].$$

Note that

$$\mathbb{P}\big[\, S_{2r-1} > 0, T_0 = 2r \,\big] = \mathbb{P}\big[\, S_{2r-1} < 0, T_0 = 2r \,\big] = \frac{1}{2}\mathbb{P}\big[\, T_0 = 2r \,\big] = \frac{1}{2} f_{2r}.$$

Also

$$\mathbb{P}\big[\, A_{2n} = 2k \,|\, S_{2r-1} > 0, T_0 = 2r \,\big] = \beta_{2k-2r, 2n-2r},$$

$$\mathbb{P}\big[\, A_{2n} = 2k \,|\, S_{2r-1} < 0, T_0 = 2r \,\big] = \beta_{2k, 2n-2r}.$$

Hence

$$\beta_{2k,2n} = \frac{1}{2} \sum_{r=0}^{k} \beta_{2k-2r, 2n-2r} f_{2r} + \frac{1}{2} \sum_{r=0}^{n-k} \beta_{2k, 2n-2r} f_{2r}.$$

We will prove by induction on n that

$$P(n) : \forall 0 \leq k \leq n, \;\; \beta_{2k,2n} = \alpha_{k,n-k} = u_{2k} u_{2n-2k}.$$

$P(1)$: When $n = 1$, $\beta_{2,2} = u_2 u_0 = \frac{1}{2}$, the proposition holds.

Assume $P(n)$ holds for all $n \leq m-1$. Then, for $n = m$,

$$\beta_{2k,2m} = \frac{1}{2} \sum_{r=1}^{k} \beta_{2k-2r, 2m-2r} f_{2r} + \frac{1}{2} \sum_{r=1}^{m-k} \beta_{2k, 2m-2r} f_{2r}$$

Clearly, $m - r \leq m - 1$.

Therefore, $\beta_{2k-2r,2m-2r} = u_{2k}u_{2m-2k}$ and $\beta_{2k,2m-2r} = u_{2k}u_{2m-2k-2r}$. Hence,

$$\beta_{2k,2m} = \frac{1}{2}\sum_{r=1}^{k} u_{2k-2r}u_{2m-2k}f_{2r} + \frac{1}{2}\sum_{r=1}^{m-k} u_{2k}u_{2m-2k-2r}f_{2r}$$
$$= \frac{1}{2}u_{2m-2k}\sum_{r=1}^{k} u_{2k-2r}f_{2r} + \frac{1}{2}u_{2k}\sum_{r=1}^{m-k} u_{2m-2k-2r}f_{2r}.$$

Observe that

$$\sum_{r=1}^{k} u_{2k-2r}f_{2r} = \sum_{r=1}^{k}\mathbb{P}\big[\, S_{2k-2r} = 0\,\big]\mathbb{P}\big[\, T_0 = 2r\,\big] = \mathbb{P}\big[\, S_{2k} = 0\,\big] = u_{2k}$$

and similarly,

$$\sum_{r=1}^{m-k} u_{2m-2k-2r}f_{2r} = \sum_{r=1}^{k}\mathbb{P}\big[\, S_{2m-2k-2r} = 0\,\big]\mathbb{P}\big[\, T_0 = 2r\,\big]$$
$$= \mathbb{P}\big[\, S_{2m-2k} = 0\,\big] = u_{2m-2k}.$$

Thus,

$$\beta_{2k,2m} = \frac{1}{2}u_{2m-2k}u_{2k} + \frac{1}{2}u_{2k}u_{2m-2k} = u_{2k}u_{2m-2k} = \alpha_{k,m-k}.$$

Therefore, we have proved that $\beta_{2k,2n} = \alpha_{k,n-k}$.

□

The next result was first proved by P. Lévy for Brownian motion.

Corollary 2.4 (Arcsine law for the leading time)**.** *In particular, for any* $x \in [0,1]$, *we have*

$$\lim_{n\to\infty}\mathbb{P}\Big[\,\frac{A_{2n}}{2n} \leq x\,\Big] = \frac{2}{\pi}\arcsin\sqrt{x}.$$

Proof. As $n \to \infty$

$$\lim_{x\to\infty}\mathbb{P}\Big[\,\frac{A_{2n}}{2n} \leq x\,\Big] = \lim_{n\to\infty}\sum_{k<xn} u_{2k}u_{2n-2k} \overset{(2.3)}{=} \frac{2}{\pi}\arcsin\sqrt{x}.$$

□

Remark 2.5. The probability density of the arcsine distribution is $\frac{1}{\pi\sqrt{x(1-x)}}$. This density is depicted in Figure 2. A random variable with this distribution is more likely to have values close to 0 or 1 then near the midpoint of $[0,1]$.

To understand the significance of Corollary 2.4 it is convenient to think of the random walk in terms of a fair game of chance between two players A and B.

Suppose we have a fair coin with faces labelled ± 1. We flip a coin and if the face 1 shows up player A gets \$ 1 from B, and if -1 shows up, A gives \$ 1 to B. The quantity S_n is then the amount A won/lost after n coin flips. If $S_n > 0$, then A is in the lead, i.e.,

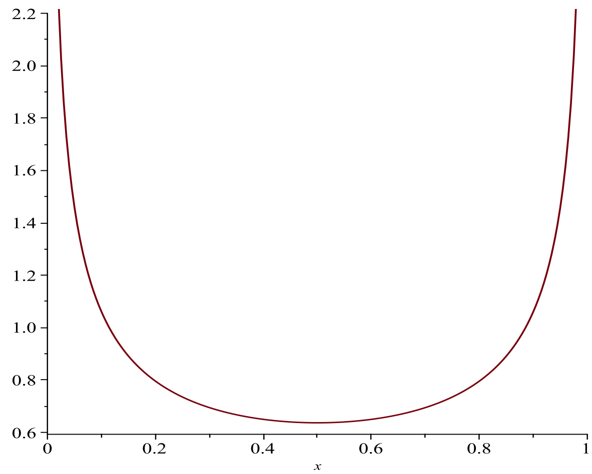

FIGURE 2. *The graph of* $\frac{1}{\pi\sqrt{x(1-x)}}$.

his fortune increased S_n dollars, while if $S_n < 0$, then B is in the lead by $|S_n|$ dollars, or, equivalently A lost $|S_n|$ dollars to B. The quantity $\frac{A_n}{n}$ indicates the amount of time player A is in the lead during a game of duration n

Since the game is fair, intuition might suggest that the fractiion of time A is in the lead ought to be roughly equal to so that $\frac{A_n}{n}$ should be close to $\frac{1}{2}$. Figure 2 shows that this is not the case. For example, for n large

$$\mathbb{P}\big[\,0.49 \leq A_n/n \leq 0.51\,\big] \approx \int_{0.49}^{0.51} \frac{1}{\pi\sqrt{x(1-x)}}dx$$

$$= \frac{2}{\pi}\big(\arcsin\sqrt{0.51} - \arcsin\sqrt{0.49}\big) \approx 0.012.$$

On the other hand

$$\mathbb{P}\big[\,A_n/n \leq 0.1\,\big] \approx \int_{0}^{0.1} \frac{1}{\pi\sqrt{x(1-x)}}dx$$

$$= \frac{2}{\pi}\arcsin\sqrt{0.1} \approx 0.204$$

By symmetry, we deduce that the probability that one of the players is in the lead most of the time is ≈ 0.408! This is about 40-times more likely than them changing lead often. □

2.3. **Maximum and first passage.** Let S_n^* be the maximum of a random walk, i.e.

$$S_n^* = \max\big\{S_0, S_1, \dots, S_n\big\}.$$

Lemma 2.6. *Given $r > 0$, the probability that a zig-zag of length n from the origin ends at $(n,k), k \leq r$ and $S_n^* \geq r$ equals*

$$p_{n,2r-k} = \mathbb{P}\big[\, S_n = 2r - k \,\big] = \begin{cases} 0, & n \not\equiv k \bmod 2, \\ \frac{1}{2n}\binom{n}{r+\frac{n-k}{2}}, & n \equiv k \bmod 2. \end{cases}$$

Proof. We use Feller's approach from [2, Sec. 3.7]. Think of the horizontal line $x = r$ as the new x-axis. Because $k \leq r$ and $0 < r$, the beginning and end lattice points are on the same side of the new x-axis. Then, by André's reflection principle, the number of paths from $(0,0)$ to (n,k) that touches or crosses $x = r$ equals the number of paths from $(0,0)$ to the reflection of (n,k) in $x = r$ which is $(n, 2r-k)$.

Thus,

$$\mathbb{P}\big[\, S_n = k,\ S_n^* \geq r \,\big] = \frac{\#\mathscr{Z}\big((0,0),(n,2r-k)\big)}{2^n} = p_{n,2r-k}.$$

□

Therefore, we can derive that the probability that a zig-zag of length n from the origin ends at $(n,k), k \leq r$ and S_n^* equals

$$\mathbb{P}\big[\, S_n = k, S_n^* \geq r \,\big] - \mathbb{P}\big[\, S_n = k, S_n^* \geq r+1 \,\big]$$
$$= p_{n,2r-k} - p_{n,2r+2-k}.$$

Theorem 2.7. *Given $r \geq 0$, the probability that a zig-zag of length n from the origin satisfies $S_n^* = r$ equals* $\max\big\{p_{n,r}, p_{n,r+1}\big\}$.

Proof. The proof below is from [2, Sec. 3.7].

$$\mathbb{P}\big[\, S_n^* = r \,\big] = \sum_{k=-n}^{r} \mathbb{P}\big[\, S_n = k;\ \ k \leq r, S_n^* = r \,\big]$$

$$= \sum_{k=-n}^{r} \big(p_{n,2r-k} - p_{n,2r+2-k} \big) = p_{n,r+1} + p_{n,r}$$

$$= \begin{cases} p_{n,r}, & n \equiv r \bmod 2, \\ p_{n,r+1}, & n \equiv r+1 \bmod 2. \end{cases}$$

□

Consider the standard random walk $(S_n)_{n\geq 0}$. For $n \in \mathbb{N}$ and $r \geq 0$ we denote by H_r the hitting time of the location r, i.e., random variable

$$H_r := \min\big\{\, 0 \leq k \leq n;\ \ S_k = r \big\}.$$

We set

$$\gamma_{r,n} := \mathbb{P}\big[\, H_r = n \,\big].$$

Theorem 2.8. *For a zig-zag of length n, the probability that the first passage through r occurs at n equals $\gamma_{r,n} = \frac{1}{2}\big(\, p_{n-1,r-1} - p_{n-1,r+1} \,\big)$.*

Proof. The proof is from [2, Sec. 3.7]. Note that $S_n = r$ is the first passage through $r \Rightarrow S_{n-1} = r - 1$ and $S_n^* = r - 1$.

$$\gamma_{r,n} = \frac{1}{2}\mathbb{P}\big[\, S_{n-1} = r-1; S_n^* = r-1 \,\big]$$
$$= \frac{1}{2}\Big(\, \mathbb{P}\big[\, S_{n-1} = r-1;\;\; S_n^* \geq r-1 \,\big] - \mathbb{P}\big[\, S_{n-1} = r-1;\;\; S_n^* \geq r \,\big] \,\Big)$$
$$= \frac{1}{2}\big(\, p_{n-1,r-1} - p_{n-1,r+1} \,\big).$$

□

Theorem 2.9. *The probability that the r-th return to the origin happens at n equals $\gamma_{r,n-r}$, i.e., the probability that the first passage through r happens at epoch n.*

Proof. The following proof is inspired by [2, Sec. 3.7]. Denote by $\mathscr{W}_{n,r}$ the set of walks $\boldsymbol{x} = X_1, \dots, X_n$ of length n

- return to 0 exactly r times,
- and the last return occurs at epoch n, i.e., $S_n = 0$.

The theorem is equivalent to the equality

$$\#\mathscr{W}_{n,r} = 2^n \gamma_{r,n-r}. \tag{2.8}$$

We denote by $\mathscr{W}_{n,r}^-$ the subset of $\mathscr{W}_{n,r}$ consisting of nonpositive walks, i.e.,

$$S_k = X_1 + \cdots + X_k \leq 0, \;\; \forall k = 1, \dots, n.$$

We first show that

$$\#\mathscr{W}_{n,r}^- = 2^{n-r} \gamma_{r,n-r}. \tag{2.9}$$

We denote by $\mathscr{W}_{n,r}^*$ the set of walks of length n that end with a first passage through r. Note that

$$\#\mathscr{W}_{n-r,r}^* = 2^{n-r} \gamma_{r,n-r}.$$

We will prove the equality (2.9) by constructing a bijection $\mathscr{W}_{n,r}^- \to \mathscr{W}_{n-r,r}^*$.

Let $\boldsymbol{x} = \big(\, X_1, \dots, X_n \,\big) \in \mathscr{W}_{n,r}^-$. Denote by

$$0 < k_1 < k_2 < \cdots < k_r = n$$

the r epochs of return to the origin of the path $\boldsymbol{x}$, i.e.,

$$S_{k_1}, S_{k_2}, \dots, S_{k_r} = 0.$$

Set $k_0 = 0$. Because the walk is entirely nonpositive, every step after the walk reaches the origin must be in the negative direction,

$$X_{k_0+1} = X_{k_2+1} = \cdots = X_{k_{r-1}+1} = -1.$$

Consider a new walk $\Psi(\boldsymbol{x})$ of length $(n-r)$ obtained from the random walk $(X_1,\dots,X_n)$ by skipping the steps X_{k_i+1}, $i=0,\dots,r-1$. For $j=1,\dots,r$, the first moment the walk $\Psi(\boldsymbol{x})$ reaches the location j is k_j-j. This shows that $\Psi(\boldsymbol{x})\in\mathscr{W}^*_{n-r,r}$. We will prove that the resulting map $\Psi:\mathscr{W}^-_{n,r}\to\mathscr{W}^*_{n-r,r}$ is a bijection.

The walk $\boldsymbol{x}$ is uniquely determined by the epochs $k_1,\dots,k_r$ and the (walk) segments

$$\sigma_j(\boldsymbol{x})=X_{k_{j-1}+2},\dots,X_{k_j},\;\; j=1,\dots r.$$

The walk $\Psi(\boldsymbol{x})$ is obtained from the succesive concatenation of the segments $\sigma_1(\boldsymbol{x}),\dots,\sigma_r(\boldsymbol{x})$.

Observe that each of the walks $\sigma_j(\boldsymbol{x})$ is a walk from 0 to 1 that reaches one only at the last step. Note that the first moment $\Psi(\boldsymbol{x})$ reaches location j is after the completion of the segments $\sigma_1,\dots,\sigma_j$, that is at epoch k_j-j. Thus the epochs k_j are uniquely determined by $\Psi(\boldsymbol{x})$. The segment $\sigma_j(\boldsymbol{x})$ corresponds to the segment of the walk $\Psi(\boldsymbol{x})$ from the first epoch it reaches location $j-1$ to the first epoch it reaches location j. Hence $\boldsymbol{x}$ is uniquely determined by $\Psi(\boldsymbol{x})$ proving that Ψ is injective.

Suppose now that $\boldsymbol{y}=(Y_1,\dots,Y_{n-r})$ is a walk in $\mathscr{W}^*_{n-r,r}$. It determines r epochs t_j, $j=1,\dots,r$, where t_j is the first epoch the walk $\boldsymbol{y}$ reaches location j. We set $t_0=0$. These epochs cut out r walk segments

$$u_j(\boldsymbol{y})=Y_{t_{j-1}+1},\dots,Y_{t_j}.$$

Then $\boldsymbol{y}=\Psi(\boldsymbol{x})$, where $\boldsymbol{x}\in\mathscr{W}^-_{n,r}$ is the walk

$$\boldsymbol{x}=-1,u_1(\boldsymbol{y}),-1,u_2(\boldsymbol{y}),\dots,,-1,u_r(\boldsymbol{y}).$$

Hence Ψ is a bijective.

The equality (2.8) is equivalent to

$$\mathscr{W}_{n,r}=2^n\gamma_{r,n-r}=2^r\#\mathscr{W}^-n,r \tag{2.10}$$

We will construct a bijection

$$\Phi:\mathscr{W}_{n,r}\to\{-1,1\}^r\times\mathscr{W}^-_{n,r}.$$

Let $\boldsymbol{x}=\big(X_1,\dots,X_n\big)\in\mathscr{W}^-_{n,r}$. Set $k_0=0$ and nenote by

$$0<k_1<k_2<\cdots<k_r=n$$

the r epochs of return to the origin of the path $\boldsymbol{x}$. These epochs cut out r walk segments

$$v_j(\boldsymbol{x})=X_{k_{j-1}+1},\dots,X_{k_j},\;\; j=1,\dots,k$$

Set $\epsilon_j=\epsilon_j(\boldsymbol{x}):=X_{k_j+1}\in\{-1,1\}$. The zigzag associated to $v_k(\boldsymbol{x})$ stays above the x-axis if $\epsilon_j=1$ and below the x-axis if $\epsilon_j=-1$. We denote by $\bar{\boldsymbol{x}}$ the walk

$$-\epsilon_1v_1(\boldsymbol{x}),\dots,-\epsilon_rv_r(\boldsymbol{x}),$$

where for any walk $\boldsymbol{y}=Y_1,\dots,Y_m$ and $\epsilon=\pm1$ we set $\epsilon\boldsymbol{y}:=\epsilon Y_1,\dots,\epsilon Y_m$. Clearly $\bar{x}x\in\mathscr{W}^-n,r$. We have thus produced a map

$$\Phi:\mathscr{W}_{n,r}\to\{-1,1\}^r\times\mathscr{W}^-_{n,r},\;\;\Phi(\boldsymbol{x})=\big(\epsilon_1(\boldsymbol{x}),\dots,\epsilon_r(\boldsymbol{x}),\;\bar{x}x\big).$$

It is a bijection with inverse

$$\{-1,1\}^r\times\mathscr{W}^-_{n,r}\ni\big(\epsilon_1,\dots,\epsilon_r,\;\boldsymbol{y}\big)\mapsto-\epsilon_1v_1(\boldsymbol{y}),\dots,-\epsilon_rv_r(\boldsymbol{y})\in\mathscr{W}_{n,r}.$$

□

2.4. **Duality and the position of maxima.** To a given a random walk of finite lengh n

$$S_n = X_1 + \cdots + X_n,$$

we can bijectively associate a dual random walk of the same length

$$S'_n = X'_1 + \cdots + X'_n.$$

The dual zigzag is obtained by rotating the original zigzag around its right endpoint $(n, S_n)$by 180 degrees, changing the orientation of the x-axis and then relabeling (n, S_n) as the new origin. Formally we set

$$X'_1 = X_n, X'_2 = X_{n-1}, \ldots, X'_n = X_1,$$

i.e., $\forall k = 0, 1, 2, \ldots, n$,

$$X'_k = X_{n+1-k}; \;\; X_k = X'_{n+1-k}.$$

Lemma 2.10. *For a zig-zag of length n, let $r = \left\lfloor \frac{n}{2} \right\rfloor$, the probability that the final epoch n is a first passage through a point equals u_{2r}, i.e.,*

$$\begin{gathered}\mathbb{P}\big[\, S_n > \max\{S_0, \ldots, S_{n-1}\}\,\big] \\ = \mathbb{P}\big[\, S_n < \min\big\{S_0, \ldots, S_{n-1}\big\}\,\big] = \frac{1}{2}u_{2r}.\end{gathered}$$

Proof. We use approach in [2, Sec. 3.8]. Given a random walk of length n, the right endpoint (n, S_n) is the first passage through a point if and only if the corresponding zigzag reaches a maximum or a minimum at that epoch.

From the definition of duality we see that the right endpoint (n, S_n) is the only maximum in the original path iff the origin is the only minimum in the dual.

When n is even, i.e. $n = 2r$,

If S_{2r} is the maximum, consider

$$\begin{gathered}\mathbb{P}\big[\, S_{2r} > \max\{S_0, \ldots, S_{2r-1}\}\,\big] \\ = \mathbb{P}\big[\, S_{2r} - S_{2r-1} > 0, S_{2r} - S_{2r-2} > 0, \ldots, S_{2r} - S_1 > 0, S_{2r} - S_0 > 0\,\big] \\ = \mathbb{P}\big[\, X_r > 0, X_r + X_{r-1} > 0, \ldots, X_r + X_{r-1} + \cdots + X_1 > 0\,\big] \\ = \mathbb{P}\big[\, X'_1 > 0, X'_1 + X'_2 > 0, \ldots, X'_1 + X'_2 + \cdots + X'_r > 0\,\big] \\ = \mathbb{P}\big[\, S'_1 > 0, S'_2 > 0, \ldots, S'_{2r} > 0\,\big] \stackrel{(1.7)}{=} \frac{1}{2}u_{2r}.\end{gathered}$$

If S_{2r} is the minimum,

$$\begin{gathered}\mathbb{P}\big[\, S_{2r} < \min\big\{S_0, \ldots, S_{2r-1}\big\}\,\big] \\ = \mathbb{P}\big[\, S_{2r} - S_{2r-1} < 0, S_{2r} - S_{2r-2} < 0, \ldots, S_{2r} - S_1 > 0, S_{2r} - S_0 < 0\,\big] \\ = \mathbb{P}\big[\, X_r < 0, X_r + X_{r-1} < 0, \ldots, X_r + X_{r-1} + \cdots + X_1 < 0\,\big] \\ = \mathbb{P}\big[\, X'_1 < 0, X'_1 + X'_2 < 0, \ldots, X'_1 + X'_2 + \cdots + X'_r < 0\,\big]\end{gathered}$$

$$= \mathbb{P}\big[\, S_1' < 0, S_2' < 0, \ldots, S_{2r}' < 0 \,\big] \overset{(1.7)}{=} \frac{1}{2} u_{2r}.$$

Similarly, when n is even, i.e., $n = 2r + 1$

If S_{2r+1} is the maximum,

$$\mathbb{P}\big[\, S_{2r+1} > \max\{S_0, \ldots, S_{2r}\} \,\big]$$
$$= \mathbb{P}\big[\, S_1' > 0, S_2' > 0, \ldots, S_{2r+1}' > 0 \,\big] \overset{(1.7)}{=} \frac{1}{2} u_{2r}.$$

If S_{2r+1} is the minimum,

$$\mathbb{P}\big[\, S_{2r+1} < \min\big\{S_0, \ldots, S_{2r}\big\} \,\big]$$
$$= \mathbb{P}\big[\, S_1' < 0, S_2' < 0, \ldots, S_{2r+1}' < 0 \,\big] \overset{(1.7)}{=} \frac{1}{2} u_{2r}.$$

Therefore, given $r = \lfloor \frac{n}{2} \rfloor$,

$$\begin{aligned} &\mathbb{P}\big[\, S_n > \max\{S_0, \ldots, S_{n-1}\} \,\big] \\ &= \mathbb{P}\big[\, S_n < \min\big\{S_0, \ldots, S_{n-1}\big\} \,\big] = \frac{1}{2} u_{2r}. \end{aligned} \tag{2.11}$$

□

Theorem 2.11 (Distribution of maxima)**.** *The probability that the first visit to the right endpoint of a zig-zag of length $2n$ happens at $2k$*

$$\mathbb{P}\big[\, S_{2k} = S_{2n} > \max\{S_0, \ldots, S_{2k-1}\} \,\big] = u_{2k} u_{2n-2k}.$$

Proof. The following proof is from [2, Sec. 3.8].

$$\begin{aligned} &\mathbb{P}\big[\, S_{2k} = S_{2n} > \max\{S_0, \ldots, S_{2k-1}\} \,\big] \\ &= \mathbb{P}\big[\, S_1 \neq S_{2k}, \ldots, S_{2k-1} \neq S_{2k} \,\big] \mathbb{P}\big[\, S_{2k} = S_{2n} \,\big] \\ &= \Big\{ \mathbb{P}\big[\, S_{2k} > \max\{S_0, \ldots, S_{2k-1}\} \,\big] + \mathbb{P}\big[\, S_{2k} < \min\big\{S_0, \ldots, S_{2k-1}\big\} \,\big] \Big\} \cdot u_{2n-2k} \\ &= (\frac{1}{2} u_{2k} + \frac{1}{2} u_{2k}) u_{2n-2k} = u_{2k} u_{2n-2k}. \end{aligned}$$

□

Consider the standard random walk $(S_{2n})_{n \geq 0}$. For $n \in \mathbb{N}$ and $r \geq 0$ we denote by $M = M_{2n}$ the first moment when the maximum S_{2n}^* is attained. More precisely,

$$M_{2n} := \min\big\{\, 0 \leq m \leq 2n; \;\; S_m = \max\{S_0, \ldots, S_{2n}\} \big\}.$$

Theorem 2.12 (Location of Maximum)**.** *The probability that the maximum of a random walk of length $2n$ first occurs at $2r$ or $2r + 1$,*

$$\mathbb{P}\big[\, M_{2n} = 2r \,\big] = \mathbb{P}\big[\, M_{2n} = 2r + 1 \,\big] = \frac{1}{2} u_{2r} u_{2n-2r}. \tag{2.12}$$

In particular

$$\lim_{n \to \infty} \mathbb{P}\Big[\, \frac{M_{2n}}{2n} \leq x \,\Big] = \frac{2}{\pi} \arcsin\big(\sqrt{x} \,\big). \tag{2.13}$$

Proof. We follow the approach in [2, Sec. 3.8]. For simplicity we set $M = M_{2n}$. When $M = 2r$,

$$
\begin{aligned}
\mathbb{P}\big[\,M = 2r\,\big] &= \mathbb{P}\big[\,S_{2r} > \max\{S_0,\dots,S_{2r-1}\}; S_{2r} \geq \max\{S_{2r},\dots,S_{2n}\}\,\big]\\
&= \mathbb{P}\big[\,S_{2r} > \max\{S_0,\dots,S_{2r-1}\}\,\big]\mathbb{P}\big[\,S_{2r} \geq \max\{S_{2r},\dots,S_{2n}\}\,\big]\\
&\stackrel{(2.11)}{=} \frac{1}{2}u_{2r}\mathbb{P}\big[\,S_{2r} \geq S_{2r}, S_{2r} \geq S_{2r+1},\dots,S_{2r} \geq S_{2n}\,\big]\\
&= \frac{1}{2}u_{2r}\mathbb{P}\big[\,S_{2r} - S_{2r+1} \geq 0,\dots,S_{2r} - S_{2n} \geq 0\,\big]\\
&= \frac{1}{2}u_{2r}\mathbb{P}\big[\,X_{2r+1} \leq 0,\dots,X_{2r+1} + \cdots + X_{2n} \leq 0\,\big]\\
&= \frac{1}{2}u_{2r}\mathbb{P}\big[\,S_1 \leq 0,\dots,S_{2n-2r} \leq 0\,\big] \stackrel{(2.7)}{=} \frac{1}{2}u_{2r}u_{2n-2r}.
\end{aligned}
$$

When $M = 2r + 1$,

$$
\begin{aligned}
\mathbb{P}\big[\,M = 2r+1\,\big] &= \mathbb{P}\big[\,S_{2r+1} > \max\{S_0,\dots,S_{2r}\}; S_{2r+1} \geq \max\{S_{2r+1},\dots,S_{2n}\}\,\big]\\
&= \mathbb{P}\big[\,S_{2r+1} > \max\{S_0,\dots,S_{2r}\}\,\big]\mathbb{P}\big[\,S_{2r+1} \geq \max\{S_{2r+1},\dots,S_{2n}\}\,\big]\\
&\stackrel{(2.11)}{=} \frac{1}{2}u_{2r}\mathbb{P}\big[\,S_{2r+1} \geq S_{2r+1}, S_{2r+1} \geq S_{2r+2},\dots,S_{2r+1} \geq S_{2n}\,\big]\\
&= \frac{1}{2}u_{2r}\mathbb{P}\big[\,S_{2r+1} - S_{2r+2} \geq 0,\dots,S_{2r+1} - S_{2n} \geq 0\,\big]\\
&= \frac{1}{2}u_{2r}\mathbb{P}\big[\,X_{2r+2} \leq 0,\dots,X_{2r+2} + \cdots + X_{2n} \leq 0\,\big]\\
&= \frac{1}{2}u_{2r}\mathbb{P}\big[\,S_1 \leq 0,\dots,S_{2n-2r-1} \leq 0\,\big].
\end{aligned}
$$

Note that $S_{2n-2r+1}$ is odd so $S_{2n-2r-1} \leq 0 \implies S_{2n-2r} \leq 0$. Hence

$$
\mathbb{P}\big[\,M = 2r\,\big] = \frac{1}{2}u_{2r}\mathbb{P}\big[\,S_1 \leq 0,\dots,S_{2n-2r} \leq 0\,\big] \stackrel{(2.7)}{=} \frac{1}{2}u_{2r}u_{2n-2r}.
$$

This proves (2.12).

To prove (2.13) note that (2.12) can be rewritten as

$$
\mathbb{P}\big[\,M_{2n} = m\,\big] = \frac{1}{2}u_{2k}u_{2n-2k}, \quad k = \lfloor m/2 \rfloor.
$$

We introduce a slightly modified variable

$$
\widehat{M}_{2n} := 2\big\lfloor\,M_{2n}/2\,\big\rfloor.
$$

This new variables takes only even values and

$$
\mathbb{P}\big[\,\widehat{M}_{2n} = 2k\,\big] = u_{2k}u_{2n-2k}
$$

Hence $\frac{\widehat{M_{2n}}}{2n}$ converges in distribution to the arcsine distribution. The claim (2.13) follows by observing that for $x \in (0,1)$

$$
\Big|\,\mathbb{P}\Big[\,\frac{M_{2n}}{2n} \leq x\,\Big] - \mathbb{P}\Big[\,\frac{\widehat{M}_{2n}}{2n} \leq x\,\Big]\,\Big| \leq r_{2n}(x) := \frac{1}{2}u_{2k(x)}u_{2n-2k(x)}, \quad k(x) = \lfloor nx \rfloor.
$$

The estimate (2.5) implies that

$$\lim_{n\to\infty} r_{2n}(x) = \lim_{n\to\infty} \frac{1}{2\pi\sqrt{\lfloor nx\rfloor\big(\,n-\lfloor nx\rfloor\,\big)}} = 0, \quad \forall x \in (0,1).$$

□

2.5. **Equidistribution of conditioned leading time.** Recall that in (2.6) we defined A_n the number of edges above the x-axis of a zig-zag of length n started at the origin. Theorem 2.3 shows that the ratio $\frac{A_{2n}}{2n}$ is biased towards its extreme values 0 and 1. Our next result shows that this bias disappears if we know additionally that the random walk returned to 0 after $2n$ steps.

Theorem 2.13. *The probability that a zig-zag of length $2n$ satisfy $A_{2n} = 2k$, $0 \le k \le n$ is independent of k. More precisely*

$$\mathbb{P}\big[\,A_{2n} = 2k \,|\, S_{2n} = 0\,\big] = \frac{1}{n+1}$$

which is independent of k. Thus, the random variable A_{2n}, conditioned by $\{S_{2n} = 0\}$ is uniformly distributed and thsus

$$\lim_{n\to\infty} \mathbb{P}\left[\,\frac{A_{2n}}{2n} \le x \,\middle|\, S_{2n} = 0\,\right] = x.$$

Proof. We follow the method in [2, Sec. 3.9]. When $k = 0$ or $k = n$, by symmetry,

$$\mathbb{P}\big[\,A_{2n} = 0 \,|\, S_{2n} = 0\,\big] = \mathbb{P}\big[\,A_{2n} = 2n \,|\, S_{2n} = 0\,\big].$$

All the edges are above the x-axis iff $S_1 \ge 0, \dots, S_{2n-1} \ge 0$ so

$$\begin{aligned}\mathbb{P}\big[\,A_{2n} = 2n \,|\, S_{2n} = 0\,\big] &= \mathbb{P}\big[\,S_1 \ge 0, S_2 \ge 0, \dots, S_{2n-1} \ge 0 \,|\, S_{2n} = 0\,\big] \\ &= \frac{\mathbb{P}\big[\,S_1 \ge 0, S_2 \ge 0, \dots, S_{2n-1} \ge 0, S_{2n} = 0\,\big]}{\mathbb{P}\big[\,S_{2n} = 0\,\big]} \\ &= \frac{\mathbb{P}\big[\,S_1 \ge 0, S_2 \ge 0, \dots, S_{2n-1} \ge 0, S_{2n} = 0\,\big]}{u_{2n}}.\end{aligned}$$

Now we consider the number of zig-zags that satisfy

$$S_1 \ge 0, S_2 \ge 0, \dots, S_{2n-1} \ge 0, S_{2n} = 0.$$

Clearly $S_1 = 1$, so the number of zig-zags from the origin to $(2n, 0)$ that have all edges above the x-axis equals to the number of random walks from $(1,1)$ to $(2n,0)$ that $S_1, \dots, S_{2n} \ne -1$. Denote $\mathscr{Z}^{-1}(A,B)$ the set of zig-zags from A to B that touch the horizontal line $y = -1$. By Ballot Theorem,

$$\#\mathscr{Z}^{-1}((1,1),(2n,0)) = \#\mathscr{Z}((1,-3),(2n,0)).$$

Therefore,

$$\begin{aligned}&\#\{S_1 \ge 0, S_2 \ge 0, \dots, S_{2n-1} \ge 0, S_{2n} = 0\} \\ &= \#\mathscr{Z}((1,1),(2n,0)) - \#\mathscr{Z}^{-1}((1,1),(2n,0)) \\ &= \#\mathscr{Z}((1,1),(2n,0)) - \#\mathscr{Z}((1,-3),(2n,0))\end{aligned}$$

$$= \binom{2n-1}{n-1} - \binom{2n-1}{n+1} = \frac{1}{n+1}\binom{2n}{n} = \frac{2^{2n}u_{2n}}{n+1}.$$

Thus,

$$\mathbb{P}\big[\, S_1 \geq 0, S_2 \geq 0, \dots, S_{2n-1} \geq 0, S_{2n} = 0 \,\big] = \frac{\frac{2^{2n}u_{2n}}{n+1}}{2^{2n}} = \frac{u_{2n}}{n+1}.$$

Therefore,

$$\begin{gathered} \eta_{0,2n} = \eta_{2n,2n} = \mathbb{P}\big[\, A_{2n} = 2n,\ S_{2n} = 0 \,\big] = \frac{u_{2n}}{n+1}, \\ \mathbb{P}\big[\, A_{2n} = 2n \,|\, S_{2n} = 0 \,\big] = \frac{\frac{u_{2n}}{n+1}}{u_{2n}} = \frac{1}{n+1}. \end{gathered} \tag{2.14}$$

When $0 < k < n$, let

$$\eta_{2k,2n} := \mathbb{P}\big[\, A_{2n} = 2k, S_{2n} = 0 \,\big].$$

We condition on the time of the first return to origin T_0. Then

$$\eta_{2k,2n} = \mathbb{P}\big[\, A_{2n} = 2k, S_{2n} = 0 \,\big]$$

$$= \sum_{r=1}^{k} \mathbb{P}\big[\, A_{2n} = 2k, S_{2n} = 0 \,|\, S_{2r-1} > 0, T_0 = 2r \,\big]\mathbb{P}\big[\, S_{2r-1} > 0, T_0 = 2r \,\big]$$
$$+ \sum_{r=1}^{n-k} \mathbb{P}\big[\, A_{2n} = 2k, S_{2n} = 0 \,|\, S_{2r-1} < 0, T_0 = 2r \,\big]\mathbb{P}\big[\, S_{2r-1} < 0, T_0 = 2r \,\big].$$

Note that

$$\begin{gathered} \mathbb{P}\big[\, S_{2r-1} > 0, T_0 = 2r \,\big] = \mathbb{P}\big[\, S_{2r-1} < 0, T_0 = 2r \,\big] \\ = \frac{1}{2}\mathbb{P}\big[\, T_0 = 2r \,\big] = \frac{1}{2} f_{2r} \stackrel{(1.8)}{=} \frac{1}{2}\frac{1}{2r-1}u_{2r}. \end{gathered}$$

Also

$$\begin{gathered} \mathbb{P}\big[\, A_{2n} = 2k, S_{2n} = 0 \,|\, S_{2r-1} > 0, T_0 = 2r \,\big] = \eta_{2k-2r,2n-2r}. \\ \mathbb{P}\big[\, A_{2n} = 2k, S_{2n} = 0 \,|\, S_{2r-1} < 0, T_0 = 2r \,\big] = \eta_{2k,2n-2r}, \end{gathered}$$

Hence

$$\eta_{2k,2n} = \frac{1}{2}\sum_{r=0}^{k} \eta_{2k-2r,2n-2r} f_{2r} + \frac{1}{2}\sum_{r=0}^{k} \eta_{2k,2n-2r} f_{2r}.$$

We will prove by induction on n that

$$\forall 1 \leq k \leq n-1,\ \ \eta_{2k,2n} = \frac{u_{2n}}{n+1}. \tag{2.15}$$

When $n = 1$, $\eta_{2,2} = \frac{u_2}{1+1} = 1$, the proposition holds.

Assume that (2.15) holds for all $n \leq m-1$. Then, for $n = m$,

$$\eta_{2k,2m} = \frac{1}{2}\sum_{r=1}^{k} \eta_{2k-2r,2m-2r} f_{2r} + \frac{1}{2}\sum_{r=1}^{m-k} \eta_{2k,2m-2r} f_{2r}.$$

From the induction assumption

$$\eta_{2k-2r,2m-2r} = \eta_{0,2m-2r} = \frac{u_{2m-2r}}{r+1}, \quad \eta_{2k,2m-2r} = \eta_{0,2m-2r} = \frac{u_{2m-2r}}{r+1}.$$

Hence

$$\eta_{2k,2m} = \sum_{r=1}^{k} \frac{u_{2m-2r}}{r+1} f_{2r}.$$

Clearly, $m - r \leq m - 1$.

Therefore,

$$\eta_{2k-2r,2m-2r} = \eta_{2k,2m-2r} = \frac{u_{2m-2r}}{m-r+1}.$$

Hence,

$$\eta_{2k,2m} = \frac{1}{2}\sum_{r=1}^{k} \frac{u_{2m-2r}}{m-r+1} f_{2r} + \frac{1}{2}\sum_{r=1}^{m-k} \frac{u_{2m-2r}}{m-r+1} f_{2r}.$$

And,

$$\begin{aligned}\frac{u_{2m-2r}}{m-r+1} f_{2r} &= \frac{u_{2m-2r}}{m-r+1} \frac{1}{2r-1} \frac{1}{2^{2r}} \binom{2r}{r} \\ &= \frac{1}{2^{2r}} \frac{u_{2m-2r}}{m-r+1} \binom{2r-2}{r-1} \frac{2r}{r^2} = \frac{1}{2^{2r}} \frac{2u_{2m-2r}u_{2r-2}}{r(m-r+1)}.\end{aligned}$$

Thus,

$$\eta_{2k,2m} = \sum_{r=1}^{k} \frac{1}{2^{2r}} \frac{u_{2m-2r}u_{2r-2}}{r(m-r+1)} + \sum_{r=1}^{m-k} \frac{1}{2^{2r}} \frac{u_{2m-2r}u_{2r-2}}{r(m-r+1)}.$$

Setting $i := m - r + 1$, we get

$$\begin{aligned}\eta_{2k,2m} &= \sum_{r=1}^{k} \frac{1}{2^{2r}} \frac{u_{2m-2r}u_{2r-2}}{r(m-r+1)} + \sum_{i=k+1}^{m} \frac{1}{2^{2r}} \frac{u_{2i-2}u_{2m-2i}}{(m+1-i)i} \\ &= \sum_{r=1}^{m} \frac{1}{2^{2r}} \frac{u_{2m-2r}u_{2r-2}}{r(m-r+1)}.\end{aligned}$$

The above term is obviously independent of k. We denote it by z_{2m} and we deduce

$$\underbrace{\sum_{k=1}^{m-1} \eta_{2k,2m}}_{(n-1)\text{ terms}} + \mathbb{P}\big[\, A_{2n} = 0, S_{2n} = 0 \,\big] + \mathbb{P}\big[\, A_{2n} = 2n, S_{2n} = 0 \,\big] = u_{2n}$$

$$\overset{(2.14)}{=} (n-1)z_{2n} + \frac{2u_{2n}}{m+1} = u_{2n}$$

$$\implies \eta_{2k,2m} = z_{2m} = \frac{u_{2m}}{m+1}, \quad \forall k = 0, 1, 2, \ldots, n.$$

Therefore, $\forall 0 \leq k \leq n$,

$$\mathbb{P}\big[\, A_{2n} = 2k \,|\, S_{2n} = 0 \,\big] = \frac{1}{n+1}.$$

□

Remark 2.14. Above we have proved that

$$\mathbb{P}\big[\, A_{2n} = 2k, S_{2n} = 0 \,\big] = \frac{1}{2^{2n}(n+1)} \binom{2n}{n}.$$

The fraction

$$C_n := \frac{1}{n+1} \binom{2n}{n}$$

is in fact a positive *integer* called the n-th Catalan number; see [7, Sec. 1.4]. □

Remark 2.15. For $p \in \mathbb{N}$ we denote by $T_{0,p}$ the moment of the first return to the origin

$$T_{0,1} = T_0 = \min \big\{ \, n \in \mathbb{N}; \;\; S_n = 0 \, \big\}, \;\; T_{0,p+1} = \min \big\{ n > T_{0,p}; \;\; S_n = 0 \, \big\}$$

The condition S_{2n} in the above theorem can by rephrased as $\exists p \in \mathbb{N}$ such that $2n = R_p$. However, we have no information about p.

A surprising result of P. Lévy [4, Cor.2, p. 303-304] shows that equidistribution is lost if the ambiguity about p is removed. More precisely Lévy proved that as $p \to \infty$ fraction of lead time $\frac{A_n}{n}$ conditioned by $n = R_p$ converges again to the arcsine distribution,

$$\lim_{p \to \infty} \mathbb{P}\Big[\, \frac{A_{T_{0,p}}}{T_{0,p}} \leq x \,\Big] = \frac{2}{\pi} \arcsin\big(\sqrt{x} \, \big).$$

To quote Chung and Feller [1] "*these results should serve as a warning to statististicians who might assume that fluctuation phenomena always follow the bell shaped pattern*". □

3. Central Limit Theorem

3.1. **Central Limit Theorem: the local version.** Let T_n be a binomial distribution corresponding to n independent trials with success probability $\frac{1}{2}$. Define $p_n(k)$ as the probability that T_n equals k, i.e.

$$p_n(k) = \mathbb{P}\big[\, T_n = k \,\big] = \frac{1}{2^n} \binom{n}{k}$$

Theorem 3.1 (Local Limit Theorem)**.** *For any function $\phi : \mathbb{N} \to (0, \infty)$ such that $\phi(n) = o\big(n^{\frac{2}{3}} \big)$ as $n \to \infty$, we have*

$$\sup_{\{k : \, |k - \frac{n}{2}| \leq \phi(n)\}} \left| \frac{p_n(k)}{\frac{2}{\sqrt{2\pi n}} e^{-\frac{(2k-n)^2}{2n}}} - 1 \right| \to 0. \tag{3.1}$$

Proof. We follow the approach in [6, Sec. 1.6]. Set

$$K_n := \big\{ \, k \in \mathbb{N} : \;\; \big| \, k - n/2 \, \big| \leq \phi(n) \, \big\}.$$

In the sequence we will denote by $\mathscr{S}$ the set of sequences of functions

$$\epsilon(n) : K_n \to [0, \infty), \;\; k \mapsto \epsilon(n, k),$$

such that

$$\lim_{n \to \infty} \sup_{k \in K_n} \epsilon(n, k) = 0.$$

Define

$$d(n) : K_n \to [0,\infty), \quad d(n,k) = \left| \frac{p_n(k)}{\frac{2}{\sqrt{2\pi n}} e^{-\frac{(2k-n)^2}{2n}}} - 1 \right|.$$

We have to show that

$$d(n) \in \mathscr{S}. \tag{3.2}$$

The proof is based on Stirling's formula (A.2)

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \big(1 + O(1/n)\big) \text{ as } n \to \infty.$$

$$p_n(k) = \frac{1}{2^n}\binom{n}{k} = \frac{1}{2^n}\frac{n!}{k!(n-k)!}.$$

$$= \frac{1}{2^n}\frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \big(1+O(1/n)\big)}{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k \big(1+O(1/k)\big)\sqrt{2\pi(n-k)}\left(\frac{n-k}{e}\right)^{n-k}\Big(1+O\big(1/(n-k)\big)\Big)}$$

$$= \frac{1}{2^n}\frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k \sqrt{2\pi(n-k)}\left(\frac{n-k}{e}\right)^{n-k}}\frac{1+O(1/n)}{\big(1+O(1/k)\big)\Big(1+O\big(1/(n-k)\big)\Big)}.$$

Observe that if $k \in K_n$ then

$$\frac{n}{2} - \phi(n) \le k, \; n-k \le \frac{n}{2} + \phi(n).$$

Hence

$$\frac{1}{k}, \quad \frac{1}{n-k} = O(1/n) \text{ when } k \in K_n \text{ and } n \to \infty.$$

Hence, when $k \in K_n$ and $n \to \infty$, we have

$$\frac{1+O(1/n)}{\big(1+O(1/k)\big)\Big(1+O\big(1/(n-k)\big)\Big)} = 1 + \epsilon_1(n), \quad \epsilon_1(n) \in \mathscr{S}.$$

We deduce

$$p_n(k) = \frac{1}{2^n}\frac{\sqrt{2\pi n}n^n}{\sqrt{2\pi k}k^k\sqrt{2\pi(n-k)}(n-k)^{n-k}}\big(1+\epsilon_1(n)\big)$$

$$= \frac{1}{2^n}\frac{\sqrt{n}n^n}{\sqrt{2\pi k}k^k\sqrt{n-k}n^{n-k}(1-\frac{k}{n})^{n-k}}\big(1+\epsilon_1(n)\big)$$

$$= \frac{1}{2^n}\frac{\sqrt{n}}{\sqrt{2\pi k}\sqrt{n(1-\frac{k}{n})}(\frac{k}{n})^k(1-\frac{k}{n})^{n-k}}\big(1+\epsilon_1(n)\big)$$

$$= \frac{1}{2^n}\frac{1}{\sqrt{2\pi n\frac{k}{n}(1-\frac{k}{n})}}\frac{1}{(\frac{k}{n})^k(1-\frac{k}{n})^{n-k}}\big(1+\epsilon_1(n)\big)$$

Set

$$\hat{p} := \frac{k}{n}$$

Then

$$p_n(k) = \frac{1}{2^n}\frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}}\left(\frac{1}{\hat{p}}\right)^k\left(\frac{1}{1-\hat{p}}\right)^{n-k}\big(1+\epsilon_1(n)\big)$$

$(2^{-n} = \big(e^{\log 1/2}\big)^n)$

$$= \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}} \exp\Big\{ k \log\Big(\frac{1}{2\hat{p}}\Big) + (n-k)\log\Big(\frac{1}{2(1-\hat{p})}\Big)\Big\}\big(1+\epsilon_1(n)\big)$$

$(\hat{p} = k/n)$

$$= \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}} \exp\Big\{ -\big(n\hat{p}\log 2\hat{p} + n(1-\hat{p})\log\big(2(1-\hat{p})\big)\Big\}\big(1+\epsilon_1(n)\big).$$

Set

$$H(x) := x\log\big(2x\big) + (1-x)\log\big(2(1-x)\big).$$

Then

$$p_n(k) = \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}} \exp\{-nH(\hat{p})\}\big(1+\epsilon_1(n)\big). \tag{3.3}$$

We want to replace $H(x)$ with the Taylor expansion of $H(x)$ around $x_0 = \frac{1}{2}$. For $0 < x < 1$,

$$H'(x) = \log\Big(\frac{x}{1-x}\Big), \quad H''(x) = \frac{1}{x(1-x)},$$

$$H'''(x) = -\frac{1}{x^2} + \frac{1}{(1-x)^2}.$$

We have

$$H\Big(\frac{1}{2}\Big) = \frac{1}{2}\log 1 + \frac{1}{2}\log 1 = 0, \quad H'\Big(\frac{1}{2}\Big) = \log 1 = 0, \quad H''\Big(\frac{1}{2}\Big) = 4,$$

$$H(\hat{p}) = H\Big(\frac{1}{2}\Big) + H'\Big(\frac{1}{2}\Big)\Big(\hat{p} - \frac{1}{2}\Big) + \frac{H''(\frac{1}{2})}{2}\Big(\hat{p}-\frac{1}{2}\Big)^2 + O\Big(\Big(\hat{p}-\frac{1}{2}\Big)^3\Big)$$

$$= 2\Big(\hat{p}-\frac{1}{2}\Big)^2 + O\Big(\Big((\hat{p}-\frac{1}{2}\Big)^3\Big).$$

Note that

$$\hat{p} - \frac{1}{2} = \frac{1}{n}\Big(k - \frac{n}{2}\Big),$$

so that, if $k \in K_n$ we have

$$\Big|\hat{p} - \frac{1}{2}\Big| \le \frac{\phi(n)}{n}.$$

Therefore,

$$n\Big|\hat{p} - \frac{1}{2}\Big|^3 \le \frac{\phi(n)^3}{n^2} \to 0,$$

so that

$$n\Big|\hat{p} - \frac{1}{2}\Big|^3 = \epsilon_2(n), \quad \epsilon_2(n) \in \mathscr{S}. \tag{3.4}$$

Then (3.3) is equivalent to

$$p_n(k) = \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}} \exp\Big\{-2n\Big(\hat{p}-\frac{1}{2}\Big)^2 + nO\Big(\Big(\hat{p}-\frac{1}{2}\Big)^3\Big)\Big\}\big(1+\epsilon_1(n)\big)$$

$(\hat{p} = k/n)$

$$= \frac{2}{\sqrt{2\pi n}} \exp\Big\{ -2n\Big(\frac{k}{n} - \frac{1}{2}\Big)^2 + O\Big(n\Big(\hat{p} - \frac{1}{2}\Big)^3 \Big) \Big\} \frac{1+\epsilon_1(n)}{2\sqrt{\hat{p}(1-\hat{p})}}$$

$$= \frac{2}{\sqrt{2\pi n}} \exp\Big\{ -\frac{(2k-n)^2}{2n} \Big\} \frac{1+\epsilon_1(n)}{2\sqrt{\hat{p}(1-\hat{p})}} \exp\Big\{ O\Big(n\Big(\hat{p} - \frac{1}{2}\Big)^3 \Big) \Big\}.$$

Using (3.4) we deduce that

$$\exp\Big\{ O\Big(n\Big(\hat{p} - \frac{1}{2}\Big)^3 \Big) \Big\} = 1 + \epsilon_3(n), \;\; \epsilon_3(n) \in \mathscr{S}.$$

Note also that

$$\Big| \hat{p} - \frac{1}{2} \Big|, \;\; \Big| \big(1 - \hat{p} \big) - \frac{1}{2} \Big| \leq \frac{\phi(n)}{n}, \;\; \forall k \in K_n.$$

Hence

$$\frac{1+\epsilon_1(n)}{2\sqrt{\hat{p}(1-\hat{p})}} = 1 + \epsilon_4(n), \;\; \epsilon_4(n) \in \mathscr{S}.$$

We deduce that for any $k \in K_n$

$$p_n(k) = \frac{2}{\sqrt{2\pi n}} e^{-\frac{(2k-n)^2}{2n}} \big(1 + \epsilon_3(n) \big) \big(1 + \epsilon_4(n) \big). \tag{3.5}$$

This proves (3.2). □

3.2. **The Central Limit Theorem: the global version.** The expectation, variance and standard deviation of T_n can be calculated respectively as

$$t_n := \mathbb{E}\big[T_n \big] = \frac{n}{2}, \;\; v_n := \mathrm{Var}\big[T_n \big] = \frac{n}{4}, \;\; \sigma_n := \sigma[T_n] = \frac{\sqrt{n}}{2}.$$

Theorem 3.2 (De Moivre-Laplace Integral Theorem)**.** *Given $a, b \in [-\infty, \infty]$, Let*

$$\mathbb{P}\Big\{ a < \frac{T_n - t_n}{\sigma_n} \leq b \Big\} = p_n(a, b].$$

As $n \to \infty$,

$$p_n(a, b] - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \to 0. \tag{3.6}$$

Proof. The proof below is inspired by [3, Sec. 7.4] and [6, Sec. 1.6]. We will carry the proof in several steps

Step 1. Assume that $a, b \in \mathbb{R}$. Let $T = \max(|a|, |b|)$ so

$$(a, b] \subset [-T, T].$$

Define A_n as the interval

$$(a\sigma_n, b\sigma_n] = \Big(a\frac{\sqrt{n}}{2}, b\frac{\sqrt{n}}{2} \Big].$$

Denote I_{A_n} the indicator function of A_n, i.e.,

$$I_{A_n} : \mathbb{R} \to \{0,1\}, \quad I_S(x) = \begin{cases} 1, & x \in A_n, \\ 0, & x \notin A_n. \end{cases}$$

Then,

$$p_n(a,b] = \mathbb{P}\Big[\, T_n - \frac{n}{2} \in A_n \,\Big] = \sum_{k=0}^{n} I_{A_n}\Big(\, k - \frac{n}{2} \,\Big) \cdot p_n(k)$$

Note that

$$I_{A_n}\Big(\, k - \frac{n}{2} \,\Big) \implies (k - n/2) \in A_n \implies \big|\, k - n/2 \,\big| \le T\sqrt{n} =: \phi(n).$$

Note that

$$\phi\big(\, n \,\big) = O\big(\, n^{1/2} \,\big) \;\text{ as } n \to \infty.$$

By (3.5), as $n \to \infty$ we have

$$\begin{aligned} \mathbb{P}[T_n - \frac{n}{2} \in A_n] &= \frac{2}{\sqrt{2\pi n}} \sum_{k=0}^{n} \Big[\, I_{A_n}\Big(\, k - \frac{n}{2} \,\Big) e^{-\frac{(2k-n)^2}{2n}} \big(\, 1 + \epsilon_3(n) \,\big)\big(\, 1 + \epsilon_4(n) \,\big) \,\Big] \\ &= \frac{1}{\sqrt{2\pi}\sigma_n} \sum_{k=0}^{n} \Big[\, I_{A_n}\Big(\, k - \frac{n}{2} \,\Big) e^{-\frac{(2k-n)^2}{2n}} \big(\, 1 + \epsilon_3(n) \,\big)\big(\, 1 + \epsilon_4(n) \,\big) \,\Big]. \end{aligned}$$

Define

$$1 + \epsilon_5(n) = \big(\, 1 + \epsilon_3(n) \,\big)\big(\, 1 + \epsilon_4(n) \,\big).$$

Then,

$$\begin{aligned} \mathbb{P}[T_n - \frac{n}{2} \in A_n] &= \frac{1}{\sqrt{2\pi}\sigma_n} \sum_{k=0}^{n} \Big[\, I_{A_n}\Big(\, k - \frac{n}{2} \,\Big) e^{-\frac{(2k-n)^2}{2n}} \big(\, 1 + \epsilon_5(n) \,\big) \,\Big]. \qquad (3.7) \\ &= \mathbb{P}\Big[\, T_n \in (n/2 + a\sqrt{n}/2, n/2 + b\sqrt{n}/2] \,\Big]. \end{aligned}$$

Define

$$t(k) := \frac{k - t_n}{\sigma_n} = \frac{k - n/2}{\frac{\sqrt{n}}{2}},$$

where the set of points $t(k)$ also fills the real line. Because we restricted k such that $\big|\, k - n/2 \,\big| \le \phi(n), \phi(n) = o\big(\, n^{\frac{2}{3}} \,\big)$, equivalently we also have

$$t(k) = \frac{k - n/2}{\frac{\sqrt{n}}{2}} \le \psi(n), \quad \psi(n) = O(1).$$

Therefore,

$$\mathbb{P}\big[\, T_n = k \,\big] = \mathbb{P}\Big[\, T_n = \frac{n}{2} + t(k)\frac{\sqrt{n}}{2} \,\Big] = p_n\Big(\, \frac{n}{2} + t(k)\frac{\sqrt{n}}{2} \,\Big)$$

Notice that

$$k - n/2 = t(k)\sigma_n, \quad \frac{1}{\sigma_n} = t(k+1) - t(k), \quad -\frac{(2k-n)^2}{2n} = -\frac{t(k)^2}{2},$$

Then, by (3.7),

$$\mathbb{P}\big[\,T_n - \frac{n}{2} \in A_n\,\big] = \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{n} \Big[\,I_{A_n}\big(\,t(k)\sigma_n\,\big)\frac{1}{\sigma_n}e^{-\frac{t(k)^2}{2}}\big(\,1+\epsilon_5(n)\,\big)\,\Big]$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{n} \Big[\,I_{A_n}\big(\,t(k)\sigma_n\,\big)\big(\,t(k+1)-t(k)\,\big)e^{-\frac{t(k)^2}{2}}\big(\,1+\epsilon_5(n)\,\big)\,\Big]$$

$\big(\,I_{A_n}(x\sigma_n) = I_{(a,b]}(x) = f(x)\,\big)$

$$= \underbrace{\frac{1}{\sqrt{2\pi}} \sum_{k=0}^{n} I_{(a,b]}\big(\,t(k)\,\big)\big(\,t(k+1)-t(k)\,\big)e^{-\frac{t(k)^2}{2}}}_{=:T_1(n)}$$

$$+ \underbrace{\frac{1}{\sqrt{2\pi}} \sum_{k=0}^{n} I_{(a,b]}\big(\,t(k)\,\big)\big(\,t(k+1)-t(k)\,\big)e^{-\frac{t(k)^2}{2}}\epsilon_5(n)}_{=:T_2(n)}$$

By the properties of Riemann sums, as $n \to \infty$,

$$T_1(n) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}dx.$$

Now we evaluate $T_2(n)$ as follows

$$T_2(n) = \epsilon_5(b)T_1(n) \to 0 \;\text{ as } n \to \infty.$$

Step 2. Assume $a = -\infty$ and $0 \le b < \infty$. Then

$$p_n\big(\,(-\infty, b]\,\big) = p_n\big(\,(-\infty, 0]\,\big) + p_n\big(\,(0, b]\,\big).$$

Arguing as in **Step 2** in the proof of Theorem 2.2. we deduce that

$$\lim_{n\to\infty} p_n\big(\,(-\infty, 0]\,\big) = \frac{1}{2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 e^{-x^2/2}dx.$$

From Step 1 we deduce

$$p_n\big(\,(0, b]\,\big) \to \frac{1}{\sqrt{2\pi}} \int_0^b e^{-\frac{x^2}{2}}dx.$$

Hence (3.6) is true when $a = -\infty$ and $b \ge 0$.

Step 3. Assume $a = -\infty$ and $b < 0$. We have

$$p_n\big(\,(-\infty, b]\,\big) = p_n\big(\,(-\infty, 0]\,\big) - p_n\big(\,(b, 0]\,\big).$$

and we conclude as before..

Step 4. Assume $b = \infty$. We conclude as in Step 2 and 3.

□

3.3. **Random walks again.** Recall that T_n is a binomial distribution corresponding to n independent trials with success probability $\frac{1}{2}$. Hence

$$T_n = Y_1 + \cdots + Y_n \;\; n \in \mathbb{N},$$

where Y_n are Bernoulli random variables with probability $\frac{1}{2}$. Also notice that the random variables $X_i = 2Y_i - 1$ take only the values ± 1 with equal probability so that we can describe the standard random walk S_n as

$$S_n = X_1 + \cdots + X_n = 2T_n - n, \;\; n \in \mathbb{N},$$

By (3.6), we have as $n \to \infty$,

$$\begin{aligned}\mathbb{P}\big[\, a\sqrt{n} < S_n \leq b\sqrt{n}\,\big] &= \mathbb{P}\Big[\, a\frac{\sqrt{n}}{2} + \frac{n}{2} < T_n \leq b\frac{\sqrt{n}}{2} + \frac{n}{2}\,\Big] \\ &= \mathbb{P}\Big[\, a < \frac{T_n - t_n}{\sigma_n} \leq b\,\Big] \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx.\end{aligned}$$

Appendix A. Stirling's Formula

We present a proof of Stirling's formula from [2, Sec. 2.9]

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}. \tag{A.1}$$

Proof. Since $\log x$ is an increasing function, by taking the Riemann Sum using right and left endpoints respectively,

$$\int_{k-1}^{k} \log x dx < 1 \log k < \int_k^{k+1} \log x dx.$$

Take the sum of $k = 1, 2, 3, \cdots, n$,

$$\int_0^n \log x dx < \log n! < \int_1^{n+1} \log x dx$$

$$n \log n - n < \log n! < (n+1)\log(n+1) - n$$

Clearly

$$n \log n - n < \left(n + \frac{1}{2}\right)\log n - n < (n+1)\log(n+1) - n.$$

We want to estimate the error

$$d_n := \log n! - \left(n + \frac{1}{2}\right)\log n + n.$$

We have

$$\begin{aligned} d_n - d_{n+1} &= \log n! - \left(n + \frac{1}{2}\right)\log n + n - \log(n+1)! + \left(n + 1 + \frac{1}{2}\right)\log(n+1) - n - 1 \\ &= \log \frac{n!(n+1)^{n+1+\frac{1}{2}}}{n^{n+\frac{1}{2}}(n+1)!} - 1 = \left(n + \frac{1}{2}\right)\log\left(\frac{n+1}{n}\right) - 1 \\ &= \left(\frac{2n+1}{2}\right)\log\left(\frac{1 + \frac{1}{2n+1}}{1 - \frac{1}{2n+1}}\right) - 1 \end{aligned}$$

Let $t = \frac{1}{2n+1}$,

$$d_n - d_{n+1} = \frac{2n+1}{2} \log \frac{1+t}{1-t} - 1$$

Take the Taylor series of $\log \frac{1+t}{1-t}$ at 1,

$$\begin{aligned} d_n - d_{n+1} &= \frac{2n+1}{2}(2t + \frac{2}{3}t^3 + \frac{2}{5}t^5 + \cdots) - 1 \\ &= \frac{1}{3(2n+1)^2} + \frac{1}{5(2n+1)^4} + \cdots \end{aligned}$$

Thus,

$$\begin{aligned} 0 < d_n - d_{n+1} < \frac{1}{3(2n+1)^2} + \frac{1}{3(2n+1)^4} + \cdots &= \frac{\frac{1}{3(2n+1)^2}}{1-(2n+1)^{-2}} \\ &= \frac{1}{12}\left(\frac{1}{n} - \frac{1}{n+1}\right) \end{aligned}$$

Therefore, the sequence d_n is Cauchy thus convergent. We set $C := \lim_{x\to\infty} d_n$ and we deduce

$$d_n = \log n! - (n + \frac{1}{2}) \log n + n \sim C, \quad n! =\sim e^C n^{n+\frac{1}{2}} e^{-n}$$

We shall prove that $C = \log\big(\sqrt{2\pi}\,\big)$.

By the Wallis' formula [5, Ex. 9.50],

$$\lim_{n\to\infty} \frac{2 \cdot 2 \cdot 4 \cdot 4 \cdot 6 \cdot 6 \cdots 2n \cdot 2n}{1 \cdot 1 \cdot 3 \cdot 3 \cdot 5 \cdot 5 \cdots (2n-1) \cdot (2n-1) \cdot (2n+1)} = \frac{\pi}{2}$$

therefore,

$$\frac{2 \cdot 4 \cdot 6 \cdots 2n}{1 \cdot 3 \cdot 5 \cdots (2n-1) \cdot \sqrt{2n}} \sim \sqrt{\frac{\pi}{2}}.$$

$$\frac{2 \cdot 4 \cdot 6 \cdots 2n}{1 \cdot 3 \cdot 5 \cdots (2n-1) \cdot \sqrt{2n}} = \frac{(2 \cdot 4 \cdot 6 \cdots 2n)^2}{(2n)!\sqrt{2n}} = \frac{(2^n n!)^2}{(2n)!\sqrt{2n}}$$

Because $n! \sim e^C n^{n+\frac{1}{2}} e^{-n}$,

$$\frac{(2^n n!)^2}{(2n)!\sqrt{2n}} \sim \frac{(2^n e^C n^{n+\frac{1}{2}} e^{-n})^2}{e^C 2n^{2n+\frac{1}{2}} e^{-2n} \sqrt{2n}} \sim \sqrt{\frac{\pi}{2}}.$$

Therefore, $e^C = \sqrt{2\pi}$.

□

The error of the approximation of $n!$ using Stirling's Formula can be calculated as the following:

$$\frac{1}{12n+1} - \frac{1}{12(n+1)+1} < \frac{1}{3(2n+1)^2} < d_n - d_{n+1} < \frac{1}{12}\Big(\frac{1}{n} - \frac{1}{n+1}\Big).$$

Thus $d_n - \frac{1}{12n}$ is increasing, and $d_n - \frac{1}{12n+1}$ is decreasing, therefore

$$C + \frac{1}{12n+1} < d_n < C + \frac{1}{12n}, \quad C = \log\big(\sqrt{2\pi}\,\big),$$

$$\frac{\sqrt{2\pi}\cdot n^{n+\frac{1}{2}}\cdot e^{\frac{1}{12n+1}}}{e^n} < n! < \frac{\sqrt{2\pi}\cdot n^{n+\frac{1}{2}}\cdot e^{\frac{1}{12n}}}{e^n}.$$

As $n\to\infty$ we have

$$e^{\frac{1}{12n}},\ \ e^{\frac{1}{12n+1}} = \big(1+O(1/n)\big).$$

Hence

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n\big(1+O(1/n)\big) \ \text{ as } n\to\infty. \tag{A.2}$$

References

[1] K. L. Chung, W. Feller: *On fluctuations in coin tossing*, Proc. Nat. Acad. Sci. USA, **35**(1949), 605-608.

[2] W. Feller: *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd Edition, John Wiley & Sons, 1968.

[3] E. Lesigne: *Heads or Tails. An Introduction to Limit Theorems in Probability*, Student Math. Library, vol.28, Amer. Math. Soc., 2005.

[4] P. Lévy: *Sur certains processus stochastiques homogènes*, Compositio Math. **7**(1939), 283-339.

[5] L. I. Nicolaescu: *Introduction to Real Analysis*, World Scientific, 2020.

[6] A. N. Shiryaev: *Probability 1*, 3rd Edition, Springer Verlag, 2016.

[7] R. Stanley: *Catalan numbers*, Cambridge University Press, 2015.

Department of Mathematics, University of Notre Dame, Notre Dame, IN 46556, USA

Email address: `xyang7@nd.edu`