19. Path Properties

We conclude our investigation of random walk with a look at the trajectories of simple random walk on $\mathbb{Z}$. Here we think of a random walk $S_1, S_2, \ldots$ as being represented by the polygonal curve in $\mathbb{R}^2$ having vertices $(1, S_1), (2, S_2), \ldots$ where successive vertices (n, S_n) and $(n+1, S_{n+1})$ are connected by a line segment.

We will call any polygonal curve which is a possible realization of simple random walk a *path*.

Now any path from $(0,0)$ to (n,x) consists of a steps in the positive direction and b steps in the negative direction where a and b satisfy

$$\begin{aligned} a+b &= n \\ a-b &= x \end{aligned},$$

hence $a = \dfrac{n+x}{2}$ and $b = \dfrac{n-x}{2}$.

(Note that if (n,x) is a vertex on a possible trajectory of SRW, then $n \in \mathbb{N}_0$ and $x \in \mathbb{Z}$ must have the same parity and satisfy $|x| \le n$.)

As each such path is uniquely determined by the locations of the positive steps, the total number is

$$N_{n,x} = \binom{n}{a} = \binom{n}{\frac{n+x}{2}}.$$

Our first result is an enumeration of the paths beginning and ending above the x-axis which hit 0 at some point.

Lemma 19.1 (Reflection principle)**.** *For any $q, t, n \in \mathbb{N}$, the number of paths from $(0,q)$ to (n,t) that are 0 at some point is equal to the number of paths from $(0,-q)$ to (n,t).*

Proof. (Draw picture)

Suppose that $(0, r_0), (1, r_1), \ldots, (n, r_n)$ is a path from $(0,q)$ to (n,t) which is 0 at some point.

Let $K = \min\{k : r_k = 0\}$ be the first time the path touches the x-axis. Then, setting $r_i' = -r_i$ for $0 \le i < K$ and $r_i' = r_i$ for $K \le i \le n$, we see that $(0, r_0'), (1, r_1'), \ldots, (n, r_n')$ is a path from $(0,-q)$ to (n,t).

Conversely, if $(0, s_0), (1, s_1), \ldots, (n, s_n)$ is a path from $(0,-q)$ to (n,t), then it must cross the x-axis at some point. Let $L = \min\{l : s_l = 0\}$ be the first time this happens. Then $(0, s_0'), (1, s_1'), \ldots, (n, s_n')$ defined by $s_j' = -s_j$ for $0 \le j < L$ and $s_j' = s_j$ for $L \le j \le n$ is a path from $(0,q)$ to (n,t) which is 0 at time L.

Thus the set of paths from $(0,q)$ to (n,t) which are 0 at some point is in $1-1$ correspondence with the set of paths from $(0,-q)$ to (n,t), and the theorem is proved. □

A consequence of this simple observation is

Theorem 19.1 (The Ballot Theorem)**.** *Suppose that in an election candidate A gets α votes and candidate B gets β votes where $\alpha > \beta$. The probability that candidate A is always in the lead when the votes are counted one by one is $\dfrac{\alpha-\beta}{\alpha+\beta}$.*

Proof. Let $x = \alpha - \beta$ and $n = \alpha + \beta$. The number of favorable outcomes is equal to the number of paths from $(1,1)$ to (n,x) which are never 0. (Think of a vote for A as a positive step and a vote for B as a negative step, keeping in mind that the first vote counted must be for A.)

Shifting by one time step shows that this is equal to the number of paths from $(0,1)$ to $(n-1,x)$ which are never zero.

The number of paths from $(0,1)$ to $(n-1,x)$ that hit 0 is equal to the number of paths from $(0,-1)$ to $(n-1,x)$, so subtracting this from the total number of paths gives the number of favorable outcomes.

Shifting in the vertical direction to get paths starting at zero shows that the number in question is

$$N_{n-1,x-1} - N_{n-1,x+1} = \binom{n-1}{\frac{(n-1)+(x-1)}{2}} - \binom{n-1}{\frac{(n-1)+(x+1)}{2}} = \binom{n-1}{\alpha-1} - \binom{n-1}{\alpha}$$
$$= \frac{(n-1)!}{(\alpha-1)!(n-\alpha)!} - \frac{(n-1)!}{\alpha!(n-\alpha-1)!} = \frac{(n-1)!\,(\alpha-(n-\alpha))}{\alpha!(n-\alpha)!} = \frac{2\alpha-n}{n}\binom{n}{\alpha}.$$

Since the total number of sequences of α $A's$ and β $B's$ is $\binom{n}{\alpha}$, the probability that A always leads is

$$\frac{\frac{2\alpha-n}{n}\binom{n}{\alpha}}{\binom{n}{\alpha}} = \frac{2\alpha-n}{n} = \frac{2\alpha-(\alpha+\beta)}{\alpha+\beta} = \frac{\alpha-\beta}{\alpha+\beta}.$$

□

This kind of reasoning involved in the preceding arguments is useful in computing the distribution of the hitting time of $\{0\}$ for simple random walk.

Lemma 19.2. *For simple random walk on* $\mathbb{Z}$,

$$P(S_1 \neq 0, ..., S_{2n} \neq 0) = P(S_{2n} = 0)$$

Proof.

If $S_1 \neq 0, ..., S_{2n} \neq 0$, then either $S_1, ..., S_{2n} > 0$ or $S_1, ..., S_{2n} < 0$ (since simple random walk cannot skip over integers) and the two events are equally likely by symmetry.
As such, it suffices to show that

$$P(S_1 > 0, ..., S_{2n} > 0) = \frac{1}{2}P(S_{2n} = 0).$$

Breaking up the event $\{S_1, ..., S_{2n} > 0\}$ according to the value of S_{2n} (which is necessarily an even number less than or equal to $2n$) gives

$$P(S_1 > 0, ..., S_{2n} > 0) = \sum_{r=1}^{n} P(S_1 > 0, ..., S_{2n-1} > 0, S_{2n} = 2r)$$

Now each path of length $2n$ has probability 2^{-2n} of being realized, and the number of paths from $(0,0)$ to $(2n, 2r)$ which are never zero at positive times is $N_{2n-1,2r-1} - N_{2n-1,2r+1}$ by the argument in the proof of the Ballot theorem.

Accordingly,

$$P(S_1 > 0, ..., S_{2n} > 0) = \sum_{r=1}^{n} P(S_1 > 0, ..., S_{2n-1} > 0, S_{2n} = 2r)$$
$$= \frac{1}{2^{2n}} \sum_{r=1}^{n} (N_{2n-1,2r-1} - N_{2n-1,2r+1})$$
$$= \frac{1}{2^{2n}} \left[(N_{2n-1,1} - N_{2n-1,3}) + (N_{2n-1,3} - N_{2n-1,5}) + ... + (N_{2n-1,2n-1} - N_{2n-1,2n+1})\right]$$
$$= \frac{N_{2n-1,1} - N_{2n-1,2n+1}}{2^n} = \frac{N_{2n-1,1}}{2^n}.$$

where the final equality is because you can't get to $2n+1$ in $2n-1$ steps of size 1.

To complete the proof, we observe that

$$\begin{aligned}P\left(S_{2n}=0\right) &= P\left(S_{2n}=0, S_{2n-1}=1\right)+P\left(S_{2n}=0, S_{2n-1}=-1\right) \\ &= 2P\left(S_{2n}=0, S_{2n-1}=1\right) \\ &= 2P(S_{2n-1}=1, X_{2n}=-1) \\ &= 2P(X_{2n}=-1)P(S_{2n-1}=1) \\ &= P(S_{2n-1}=1)=\frac{N_{2n-1,1}}{2^{n-1}},\end{aligned}$$

hence

$$P(S_1>0,...,S_{2n}>0)=\frac{N_{2n-1,1}}{2^n}=\frac{1}{2}P\left(S_{2n}=0\right). \qquad \square$$

Lemma 19.2 and our previous computations for simple random walk show that the distribution function of $\alpha=\inf\{n\in\mathbb{N}:S_n=0\}$ is given by

$$P(\alpha\le 2n)=1-P(S_1\ne 0,...,S_{2n}\ne 0)=1-P(S_{2n}=0)=1-\frac{1}{2^{2n}}\binom{2n}{n}\approx\frac{\sqrt{\pi n}-1}{\sqrt{\pi n}},$$
$$P(\alpha\le 2n+1)=P(\alpha\le 2n)$$

for $n=1,2...$

Our final set of results concern the so-called arcsine laws, which show that certain suitably normalized random walk statistics have limiting distributions that can be described using the arcsine function.

These theorems are typically stated in terms of Brownian motion, which arises as a scaling limit of random walk.

We first consider the arcsine law associated with

$$L_{2n}=\max\{0\le m\le 2n:S_m=0\},$$

the last visit to zero in time $2n$.

We begin with the following simple lemma.

Lemma 19.3. *Let $u_{2m}=P(S_{2m}=0)$. Then $P(L_{2n}=2k)=u_{2k}u_{2n-2k}$ for $k=0,1,...,n$.*

Proof. Using Lemma 19.2, we compute

$$\begin{aligned}P\left(L_{2n}=2k\right) &= P\left(S_{2k}=0, S_{2k+1}\ne 0,...,S_{2n}\ne 0\right) \\ &= P\left(S_{2k}=0, X_{2k+1}\ne 0,...,X_{2k+1}+...+X_{2n}\ne 0\right) \\ &= P\left(S_{2k}=0\right)P\left(X_{2k+1}\ne 0,...,X_{2k+1}+...+X_{2n}\ne 0\right) \\ &= P(S_{2k}=0)P(S_1\ne 0,...,S_{2n-2k}\ne 0)=u_{2k}u_{2n-2k}.\end{aligned} \qquad \square$$

From here, it is a small step to deduce the limit law.

Theorem 19.2. *For* $0 < a < b < 1$,

$$P\left(a \le \frac{L_{2n}}{2n} \le b\right) \to \int_a^b \frac{1}{\pi\sqrt{x(1-x)}}dx.$$

Proof. We first note that

$$nP(L_{2n} = 2k) = nu_{2k}u_{2(n-k)} \approx \frac{n}{\sqrt{\pi k}\sqrt{\pi(n-k)}} = \frac{1}{\pi}\frac{1}{\sqrt{\frac{nk-k^2}{n^2}}} = \frac{1}{\pi\sqrt{\frac{k}{n}(1-\frac{k}{n})}},$$

so if $\frac{k}{n} \to x$, then

$$nP(L_{2n} = 2k) = \left(\frac{nP(L_{2n} = 2k)}{\frac{1}{\pi\sqrt{\frac{k}{n}(1-\frac{k}{n})}}} \cdot \frac{1}{\pi\sqrt{\frac{k}{n}(1-\frac{k}{n})}}\right) \to \frac{1}{\pi\sqrt{x(1-x)}}.$$

Now define a_n and b_n so that $2na_n$ is the smallest even integer greater than or equal to $2na$ and $2nb_n$ is the largest even integer less than or equal to $2nb$.

Setting $f_n(x) = nP(L_{2n} = 2k)$ for $\frac{k}{n} \le x < \frac{(k+1)}{n}$, we see that

$$P\left(a \le \frac{L_{2n}}{2n} \le b\right) = P\left(2na_n \le L_{2n} \le 2nb_n\right) = \sum_{k=na_n}^{nb_n} nP(L_{2n} = 2k) \cdot \frac{1}{n} = \int_{a_n}^{b_n+\frac{1}{n}} f_n(x)dx.$$

Moreover, our work above shows that $f_n(x) \to f(x) = \frac{1}{\pi\sqrt{x(1-x)}}1_{(0,1)}(x)$ uniformly on compact sets, so

$$\sup_{a_n \le x \le b_n+\frac{1}{n}} f_n(x) \to \sup_{a \le x \le b} f(x) < \infty$$

for any $0 < a < b < 1$, thus we can apply the bounded convergence theorem to conclude

$$P\left(a \le \frac{L_{2n}}{2n} \le b\right) = \int_{a_n}^{b_n+\frac{1}{n}} f_n(x)dx \to \int_a^b f(x)dx. \qquad \square$$

To see the reason for the name, observe that the substitution $y = \sqrt{x}$ yields

$$\int_a^b \frac{1}{\pi\sqrt{x(1-x)}}dx = \frac{2}{\pi}\int_{\sqrt{a}}^{\sqrt{b}} \frac{dy}{\sqrt{1-y^2}} = \frac{2}{\pi}\left(\sin^{-1}\left(\sqrt{b}\right) - \sin^{-1}\left(\sqrt{a}\right)\right).$$

Note that $P\left(\frac{L_{2n}}{2n} \le \frac{1}{2}\right) \to \frac{2}{\pi}\sin^{-1}\left(\frac{1}{\sqrt{2}}\right) = \frac{1}{2}$. (This symmetry is also apparent in the mass function derived in Lemma 19.3.)

An amusing consequence is that if two people were to bet \$1 on a coin flip every day of the year, then with probability approximately $\frac{1}{2}$, one player would remain consistently ahead from July 1st onwards. In other words, if you were to play this game against me, then the probability that you would be in the lead for the first two days is about the same as the probability that you would be in the lead for the entire second half of the year!

Finally, note that the proof of Theorem 19.2 shows that any statistic T_n satisfying $P(T_{2n} = 2k) = u_{2k}u_{2n-2k}$ for $k = 0, 1, ..., n$ obeys the asymptotic arcsine law

$$P\left(a \le \frac{T_{2n}}{2n} \le b\right) \to \int_a^b \frac{1}{\pi\sqrt{x(1-x)}}dx, \quad 0 < a < b < 1.$$

Theorem 19.3. *Let π_{2n} be the number of line segments $(k-1, S_{k-1})$ to (k, S_k) that lie above the x-axis and let $u_m = P(S_m = 0)$. Then*

$$P(\pi_{2n} = 2k) = u_{2k}u_{2n-2k}.$$

Proof. Write $\beta_{2k,2n} = P(\pi_{2n} = 2k)$. We will proceed by (strong) induction.

When $n = 1$, it is clear that

$$\beta_{0,2} = \beta_{2,2} = \frac{1}{2} = u_0u_2.$$

(After two steps, the walk has either been always nonpositive or nonnegative, each being equally likely, and of the four equiprobable two-step paths, half end up at 0.)

For a general n, the proof of Lemma 19.2 shows that

$$\begin{aligned}\frac{1}{2}u_{2n}u_0 = \frac{1}{2}u_{2n} &= P(S_1 > 0, ..., S_{2n} > 0)\\ &= P(S_1 = 1, S_2 - S_1 \geq 0, ..., S_{2n} - S_1 \geq 0)\\ &= \frac{1}{2}P(S_1 \geq 0, ..., S_{2n-1} \geq 0)\\ &= \frac{1}{2}P(S_1 \geq 0, ..., S_{2n} \geq 0) = \frac{1}{2}\beta_{2n,2n}\end{aligned}$$

where the penultimate equality is due to the fact that $S_{2n-1} \geq 0$ implies $S_{2n-1} \geq 1$.

This proves the result for $k = n$, and since $\beta_{0,2n} = \beta_{2n,2n}$ (replacing S_n with $-S_n$ shows that always nonnegative is as likely as always nonpositive), we see that the result also holds for $k = 0$.

Suppose now that $1 \leq k \leq n-1$. Let R be the time of the first return to 0 (so that $R = 2m$ with $0 < m < n$) and write $f_{2m} = P(R = 2m)$. Breaking things up according to whether the first excursion was on the positive or negative side gives

$$\begin{aligned}\beta_{2k,2n} &= \frac{1}{2}\sum_{m=1}^{k} f_{2m}\beta_{2k-2m,2n-2m} + \frac{1}{2}\sum_{m=1}^{n-k} f_{2m}\beta_{2k,2n-2m}\\ &= \frac{1}{2}\sum_{m=1}^{k} f_{2m}u_{2k-2m}u_{2n-2k} + \frac{1}{2}\sum_{m=1}^{n-k} f_{2m}u_{2k}u_{2n-2m-2k}\\ &= \frac{1}{2}u_{2n-2k}\sum_{m=1}^{k} f_{2m}u_{2k-2m} + \frac{1}{2}u_{2k}\sum_{m=1}^{n-k} f_{2m}u_{2n-2m-2k}\end{aligned}$$

where the second equality used the inductive hypothesis.

Since

$$u_{2k} = \sum_{m=1}^{k} f_{2m}u_{2k-2m}, \qquad u_{2n-2k} = \sum_{m=1}^{n-k} f_{2m}u_{2n-2k-2m}$$

(by considering the time of the first return to 0), we have

$$\beta_{2k,2n} = \frac{1}{2}u_{2n-2k}u_{2k} + \frac{1}{2}u_{2k}u_{2n-2k} = u_{2k}u_{2n-2k}$$

and the result follows from the principle of induction. □

Since $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$ has a minimum at $x = \frac{1}{2}$ and goes to ∞ as $x \to 0, 1$, Theorem 19.3 shows that an equal division of steps above and below the axis is least likely, and completely one-sided divisions have the greatest probability.