Lecture 2: Random Walks 1, Reflection and Reversal

We recall that a random walk is defined by a sequence of i.i.d elements $X_1, X_2, ..$ of $\mathbb{Z}$. S_0 is the initial position (which is the origin unless explicitly stated). The position S_t after t steps is $S_0 + \sum_{i=1}^{t} X_i$.

We now show that these walks are homogeneous, and that they are memoryless (i.e. they have the Markov property).

Lemma 0.1. *Any random walk is homogeneous; that is*

$$\mathbb{P}(S_n = j \mid S_0 = x) = \mathbb{P}(S_{n+k} = j + y \mid S_k = x + y).$$

Proof. If LHS and RHS denote the left hand respectively right hand sides of the equation, we have

$$\text{LHS} = \mathbb{P}(\sum_{i=1}^{n} X_i = j - x) = \mathbb{P}(\sum_{i=k+1}^{n+k} X_i = j - x) = \text{RHS}.$$

□

Lemma 0.2. *Any random walk has the* Markov *property; that is*

$$\mathbb{P}(S_{m+n} \mid S_0, S_1, \ldots, S_m) = \mathbb{P}(S_{m+n} \mid S_m).$$

Proof. For any set of integers $j, i_0, i_1, \ldots, i_m$, we have that

$$\begin{aligned} \mathbb{P}(S_{m+n} &= j \mid S_0 = i_0, S_1 = i_1, \ldots, S_m = i_m) \\ &= \mathbb{P}(\textstyle\sum_{t=m+1}^{m+n} X_t = j - i_m) = \mathbb{P}(S_{m+n} = j \mid S_m = i_m). \end{aligned}$$

□

In this lecture we restrict our attention to walks in which $X_i = 1$ with probability p and $X_i = -1$ with probability $q = 1 - p$. We say that the ith step is a backward step if $X_i = -1$, and a forward step if $X_i = 1$. Suppose that $S_0 = u$ and $S_n = v$. Letting b denote the number of backward steps, and f the number of forward steps, we have that $f + b = n$, and $f - b = v - u$, so $f = \frac{1}{2}(n + v - u)$ and $b = \frac{1}{2}(n - v + u)$. Therefore

$$\mathbb{P}(S_n = v) = \binom{n}{\frac{1}{2}(n + v - u)} p^{\frac{1}{2}(n+v-u)} q^{\frac{1}{2}(n-v+u)},$$

since there are exactly $\binom{n}{f}$ paths of length n having f forward steps.

We are interested in the probability that a walk starting at $S_0 = 0$ stays positive for the first n steps, given that $S_n = v$, where $1 \le v \le n$. If $S_n = v$, there are $\frac{1}{2}(n - v)$ backward steps, so the conditional probability of the first step being a backward step is $\frac{1}{2}\frac{n-v}{n}$. As shown in Figure 1, there is a one-to-one correspondence between paths that go positive in the first step, revisit 0, and end at v, and paths that go negative in the first step and end at v. It follows

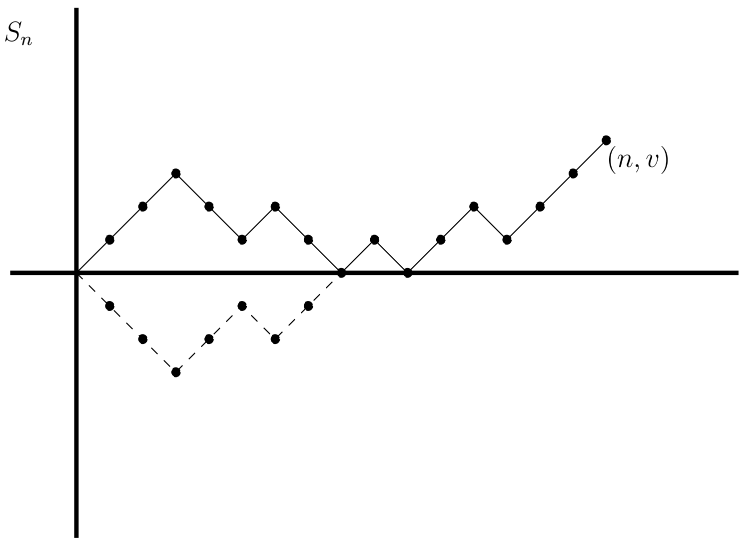


Figure 1: The reflection principle

that the probability that the walk revisits 0 before the nth step, given that it ends at v, is $\frac{n-v}{n}$, and hence the probability that the walk does not revisit 0 given that it ends at v is $\frac{v}{n}$.

We record the one-to-one correspondence described above as a theorem. Let $N_n(u, v)$ be the number of paths from $(0, u)$ to (n, v), and $N_n^0(u, v)$ the number of paths from $(0, u)$ to (n, v) which intersect the x-axis. Then we have the following theorem.

Theorem 0.3 (The Reflection Principle)**.** *If* $u, v > 0$ *then* $N_n^0(u, v) = N_n(-u, v)$.

□

Theorem 0.4 (The Ballot Theorem)**.** *If* $v > 0$ *then the number of paths from* $(0, 0)$*, to* (n, v) *which do not revisit the* x*-axis equals* $\frac{v}{n} N_n(0, v)$.

Proof. Let N denote the number of paths in question, and $\pi = \mathbb{P}($ do not revisit $0 \mid S_n = v)$. Then $\pi = \frac{N}{N_n(0,v)}$, but also $\pi = \frac{v}{n}$. Hence

$$N = \pi N_n(0, v) = \frac{v}{n} N_n(0, v).$$

□

By the same methods, one can show that for $v < 0$ the number of such paths is $\frac{|v|}{n} N_n(0, v)$.

Corollary 0.5. *The probability that the first return to* 0 *takes place at time* $2n$*, given that* $S_{2n} = 0$*, is* $\frac{1}{2n-1}$.

Proof. By the Ballot Theorem, the probability that the walk has not revisited 0 given that $S_{2n-1} = 1$ is $\frac{1}{2n-1}$, as is the probability of not revisiting 0 given that $S_{2n-1} = -1$. In the first case, we go to 0 with probability q in the last step, and with probability p in the second case. Thus the probability we're interested in equals

$$q\frac{1}{2n-1} + p\frac{1}{2n-1} = \frac{1}{2n-1}.$$

□

Theorem 0.6. *If $S_0 = 0$ then, for $n \geq 1$,*

$$\mathbb{P}(S_1 S_2 \cdots S_n \neq 0,\ S_n = v) = \frac{|v|}{n}\mathbb{P}(S_n = v),$$

and therefore

$$\mathbb{P}(S_1 S_2 \cdots S_n \neq 0) = \frac{1}{n}\mathbb{E}|S_n|.$$

Proof. Suppose that $S_0 = 0$ and $S_n = v(> 0)$. The event in question occurs if and only if the path of the random walk does not revisit the x-axis in the interval $[1, n]$. The number of such paths is, by the ballot theorem, $\frac{v}{n}N_n(0, v)$, and each such path has $\frac{1}{2}(n + v)$ rightward steps and $\frac{1}{2}(n - v)$ leftward steps. Therefore

$$\mathbb{P}(S_1 S_2 \cdots S_n \neq 0,\ S_n = v) = \frac{v}{n}N_n(0, v)p^{\frac{1}{2}(n+v)}q^{\frac{1}{2}(n-v)} = \frac{v}{n}\mathbb{P}(S_n = v)$$

as required. A similar calculation is valid for $v < 0$. □

We now introduce a second important tool based on symmetry: reversal. If the steps of the original walk are

$$\{0, S_1, S_2, \ldots, S_n\} = \left\{0, X_1, X_1 + X_2, \ldots, \sum X_i\right\}$$

then the *reverse* walk is defined by

$$\{0, T_1, T_2, \ldots, T_n\} = \left\{0, X_n, X_n + X_{n-1}, \ldots, \sum X_i\right\}.$$

Note that since the X_i are i.i.d. random variables, both walks have the same distribution, even if $p \neq \frac{1}{2}$. Also observe that both walks start at 0 and end at $\sum X_i$. We now use reversal to derive the Hitting Time Theorem from Theorem 0.6.

Theorem 0.7. *The probability $f_v(n)$ that a random walk takes the value v for the first time at step n, having started at 0, is*

$$f_v(n) = \frac{|v|}{n}\mathbb{P}(S_n = v) \text{ for } n \geq 1.$$

Proof. A random walk starting at $S_0 = 0$ satisfies $S_n = v(> 0)$ and $S_1 S_2 \cdots S_n > 0$ if and only if the reverse walk satisfies $T_n = v$ and the first visit to v takes place at time n. Thus

$$f_v(n) = \frac{v}{n}\mathbb{P}(S_n = v) \text{ for } v > 0.$$

A similar argument is valid for $v < 0$. □

We remark that the Hitting Time Theorem and its reverse can actually be expressed in terms of conditional expectations.

We write $M_n = \max\{S_i : 0 \le i \le n\}$ for the maximum value up to time n, and shall suppose $S_0 = 0$, so that $M_n \ge 0$. Clearly $M_n \ge S_n$, and the first part of the next theorem is therefore trivial.

Theorem 0.8. *Suppose that $S_0 = 0$ and $p = \frac{1}{2}$. Then, for $r \ge 1$,*

$$\mathbb{P}(M_n \ge r, S_n = v) = \begin{cases} \mathbb{P}(S_n = v) & \text{if } v \ge r, \\ \mathbb{P}(S_n = 2r - v) & \text{if } v < r. \end{cases}$$

It follows that, for $r \ge 1$,

$$\begin{aligned} \mathbb{P}(M_n \ge r) &= \mathbb{P}(S_n \ge r) + \textstyle\sum_{v=-\infty}^{r-1} \mathbb{P}(S_n = 2r - v) \\ &= \mathbb{P}(S_n = r) + \textstyle\sum_{c=r+1}^{\infty} 2\mathbb{P}(S_n = c), \end{aligned}$$

and thus

$$\mathbb{P}(M_n \ge r) = 2\mathbb{P}(S_n \ge r + 1) + \mathbb{P}(S_n = r),$$

which is easily expressed in terms of the binomial distribution.

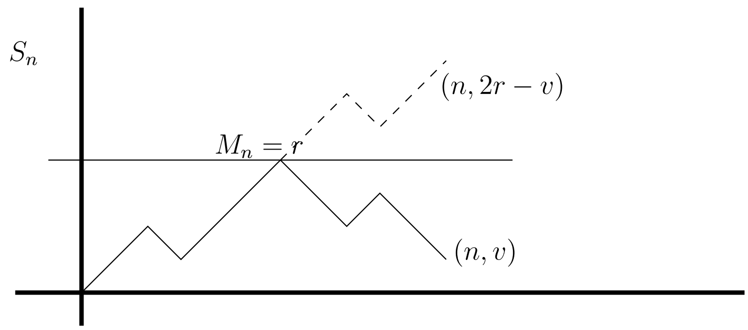
Proof. We may assume that $r \ge 1$ and $v < r$. Let $N_n^r(0, v)$ be the number of paths from $(0, 0)$ to (n, v) which include some point having height r, which is to say some point (i, r) with $0 < i < n$; for such a path π, let (i_π, r) be the earliest such point. We reflect the segment of the path with $i_\pi \le x \le n$ in the line $y = r$ to obtain a path π' joining $(0, 0)$ to $(n, 2r - v)$. Any such path π' is obtained thus from a unique path π, and therefore $N_n^r(0, v) = N_n(0, 2r - v)$. It follows as required that

$$\begin{aligned} \mathbb{P}(M_n \ge r, S_n = v) &= N_n^r(0, v)(\tfrac{1}{2})^n \\ &= N_n(0, 2r - v)(\tfrac{1}{2})^n \\ &= \mathbb{P}(S_n = 2r - v). \end{aligned}$$

□

Corollary 0.9. *Suppose that $S_0 = 0$. Then, for $r \ge 1$,*

$$\mathbb{P}(M_n = r) = \mathbb{P}(S_n = r) + \mathbb{P}(S_n = r + 1).$$



Proof.

$$\begin{aligned}\mathbb{P}(M_n = r) &= \mathbb{P}(M_n \geq r) - \mathbb{P}(M_n \geq r+1) \\ &= 2\mathbb{P}(S_n \geq r+1) + \mathbb{P}(S_n = r) - 2\mathbb{P}(S_n \geq r+2) - \mathbb{P}(S_n = r+1) \\ &= \mathbb{P}(S_n = r) + \mathbb{P}(S_n = r+1).\end{aligned}$$

□

Theorem 0.10. *If $p = \frac{1}{2}$ and $S_0 = 0$, the mean number μ_b of visits to $b \neq 0$ before the walk returns to its starting point is 1.*

Proof. Let X denote the number of visits to b before returning to 0. Then

$$\mu_b = \mathbb{E}(X) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k).$$

Conditioning on the event $\{X \geq 1\}$ (which has non-zero probability) we have

$$\begin{aligned}\mathbb{P}(X \geq 2) &= \mathbb{P}(X \geq 2,\ X \geq 1) \\ &= \mathbb{P}(X \geq 2 \,|\, X \geq 1)\mathbb{P}(X \geq 1) \\ &= (1 - \mathbb{P}(X \geq 1))\mathbb{P}(X \geq 1)\end{aligned}$$

since by symmetry $\mathbb{P}(X \geq 2 \,|\, X \geq 1) = 1 - \mathbb{P}(X \geq 1)$. It follows by induction that, for $k \geq 1$,

$$\mathbb{P}(X \geq k) = (1 - \mathbb{P}(X \geq 1))^{k-1}\mathbb{P}(X \geq 1).$$

Thus

$$\begin{aligned}\mu_b &= \mathbb{E}(X) = \sum \mathbb{P}(X \geq k) \\ &= \mathbb{P}(X \geq 1) \textstyle\sum_{k \geq 1} (1 - \mathbb{P}(X \geq 1))^{k-1} \\ &= 1.\end{aligned}$$

□

Lemma 0.11. *For the symmetric simple random walk, we have*

$$\mathbb{P}(S_{2n} = 0) = \mathbb{P}(S_1 S_2 \cdots S_{2n} \neq 0).$$

Proof. We have

$$\begin{aligned}\mathbb{P}(S_1S_2\cdots S_{2n}>0) &= \mathbb{P}(S_1=1,S_2\geq 1,\ldots,S_{2n}\geq 1)\\ &= \tfrac{1}{2}\mathbb{P}(S_1\geq 0,\ldots,S_{2n-1}\geq 0).\end{aligned}$$

Note that $2n-1$ is odd, so $S_{2n-1}\geq 0$ implies $S_{2n}\geq 0$, and hence

$$\mathbb{P}(S_1S_2\cdots S_{2n}>0)=\frac{1}{2}\mathbb{P}(S_1\geq 0,\ldots,S_{2n}\geq 0).$$

Reflect the entire walk about the x-axis to find

$$\begin{aligned}\mathbb{P}(S_1\geq 0,\ldots,S_{2n}\geq 0) &= \mathbb{P}(M_{2n}=0)\\ &= \mathbb{P}(S_{2n}=0)+\mathbb{P}(S_{2n}=1).\end{aligned}$$

But $\mathbb{P}(S_{2n}=1)=0$ as $2n$ is even, and so

$$\mathbb{P}(S_1S_2\cdots S_{2n}>0)=\frac{1}{2}\mathbb{P}(S_{2n}=0).$$

By symmetry, $\mathbb{P}(S_1S_2\cdots S_{2n}<0)=\frac{1}{2}\mathbb{P}(S_{2n}=0)$ also, and thus

$$\begin{aligned}\mathbb{P}(S_1S_2\cdots S_{2n}\neq 0) &= \mathbb{P}(S_1S_2\cdots S_{2n}<0)+\mathbb{P}(S_1S_2\cdots S_{2n}>0)\\ &= \mathbb{P}(S_{2n}=0).\end{aligned}$$

as claimed. □

Using this lemma, we can prove the following theorem.

Theorem 0.12 (Arc sine law for the last return to the origin.)**.** *Suppose that* $S_0=0$ *and* $p=\frac{1}{2}$*. Then the probability that the last visit to* 0 *occurred at time* $2k$ *is*

$$\mathbb{P}(S_{2k}=0)\mathbb{P}(S_{2n-2k}=0).$$

Proof. The probability in question is

$$\begin{aligned}&\mathbb{P}(S_{2k}=0)\mathbb{P}(S_{2k+1}S_{2k+2}\cdots S_{2n}\neq 0\mid S_{2k}=0)\\ =\;&\mathbb{P}(S_{2k}=0)\mathbb{P}(S_1S_2\cdots S_{2n-2k}\neq 0)\end{aligned}$$

Applying the lemma to the second factor, we find that the probability of the last return to the origin occurring at time $2k$ equals

$$\mathbb{P}(S_{2k}=0)\mathbb{P}(S_{2n-2k}=0),$$

as claimed. □

Theorem 0.13 (Arc sine law for sojourn time)**.** *Let* R_l *be the number of the first* l *arcs which are to the right of the origin. Then*

$$\mathbb{P}(R_{2n}=2k)=\mathbb{P}(S_{2k}=0)\mathbb{P}(S_{2n-2k}).$$

Proof. Obviously,

$$\mathbb{P}(S_{2k}=0)=\sum_{j=1}^{k}\mathbb{P}(2j \text{ is the first return})\mathbb{P}(S_{2k-2j}=0).$$

So

$$\mathbb{P}(S_{2k}=0)\mathbb{P}(S_{2n-2k}=0)=\sum_{j=1}^{k}\mathbb{P}(2j \text{ is the first return})\mathbb{P}(S_{2k-2j}=0)\mathbb{P}(S_{2n-2k}=0).$$

Symmetrically

$$\mathbb{P}(S_{2k}=0)\mathbb{P}(S_{2n-2k}=0)=\sum_{j=1}^{n-k}\mathbb{P}(2j \text{ is the first return})\mathbb{P}(S_{2n-2k-2j}=0)\mathbb{P}(S_{2k}=0).$$

Thus

$$\begin{aligned}\mathbb{P}(S_{2k}=0)\mathbb{P}(S_{2n-2k}=0) &= \tfrac{1}{2}\textstyle\sum_{j=1}^{k}\mathbb{P}(2j \text{ is the first return})\mathbb{P}(S_{2k-2j}=0)\mathbb{P}(S_{2n-2k}=0)\\ &+ \tfrac{1}{2}\textstyle\sum_{j=1}^{n-k}\mathbb{P}(2j \text{ is the first return})\mathbb{P}(S_{2n-2k-2j}=0)\mathbb{P}(S_{2k}=0).\end{aligned}$$

By induction on n we have:

$$\begin{aligned}\mathbb{P}(S_{2k}=0)\mathbb{P}(S_{2n-2k}=0) &= \textstyle\sum_{j=1}^{k}\tfrac{1}{2}\mathbb{P}(2j \text{ is the first return})\mathbb{P}(R_{2n-2j}=2k-2j)\\ &+ \tfrac{1}{2}\textstyle\sum_{j=1}^{n-k}\tfrac{1}{2}\mathbb{P}(2j \text{ is the first return})\mathbb{P}(R_{2n-2j}=2k).\end{aligned}$$

By reflection, the probability that $X_1=1$ given that $2j$ is the first return to the origin is $\frac{1}{2}$. So we obtain:

$$\begin{aligned}\mathbb{P}(S_{2k}=0)\mathbb{P}(S_{2n-2k}=0) &= \textstyle\sum_{j=1}^{k}\mathbb{P}(2j \text{ is the first return}\cap X_1=1\cap R_{2n}=2k)\\ &+ \textstyle\sum_{j=1}^{n-k}\mathbb{P}(2j \text{ is the first return}\cap X_1=-1\cap R_{2n}=2k).\end{aligned}$$

The desired result follows. □