Contents

4 Random walks

4a Reflection

Consider the one-dimensional simple random walk: $S_n = X_1 + \dots + X_n$ (where X_k are independent random signs, as in 1a), and let $M_n = \max(S_0, \dots, S_n)$. We know the distribution of S_n: $\mathbb{P}\big(S_n = m\big) = 2^{-n}\binom{n}{\frac{n+m}{2}}$ for $m = -n, -n+2, \dots, n$. Interestingly, we can calculate the distribution of M_n, and moreover, the joint distribution of S_n and M_n.

4a1 Proposition. For every $m \ge 0$,

$$\mathbb{P}\big(M_n = m\big) = \mathbb{P}\big(S_n = m\big) + \mathbb{P}\big(S_n = m+1\big) = \\ = 2^{-n} \cdot \begin{cases} \binom{n}{\frac{n}{2} \pm \frac{m}{2}} & \text{for } m+n \text{ even,} \\ \binom{n}{\frac{n}{2} \pm \frac{m+1}{2}} & \text{for } m+n \text{ odd.} \end{cases}$$

4a2 Lemma. $\mathbb{E}\big(f(S_n - m)1\!\!1_{M_n \ge m}\big) = 0$ for all $m \ge 0$ and every odd (antisymmetric) function f.[1]

In other words, the conditional distribution (if defined) is symmetric around m.

Proof. For $m = 0$: trivial. For $m > 0$: define "first hit" events

$$A_k = \{S_1 < m, \dots, S_{k-1} < m, S_k = m\} \quad \text{for } k = 1, \dots, n\,;$$

clearly, $A_1 \uplus \dots \uplus A_n = \{M_n \ge m\}$; it is sufficient to prove that $\mathbb{E}\big(f(S_n - m)1\!\!1_{A_k}\big) = 0$ for all k.

In terms of the corresponding sets $B_k \subset \mathbb{R}^k$ defined by

$$B_k = \{(x_1, \dots, x_k) : x_1 < m, x_1 + x_2 < m, \dots, x_1 + \dots + x_{k-1} < m, x_1 + \dots + x_k = m\}$$

[1] That is, $\forall x \;\; f(-x) = -f(x)$.

we have

$$\mathbb{E}\left(f(S_n-m)\mathbb{1}_{A_k}\right) = 2^{-n} \sum_{x_1,\dots,x_n=\pm1} f(x_1+\dots+x_n-m)\mathbb{1}_{B_k}(x_1,\dots,x_k) =$$
$$= 2^{-n} \sum_{x_1,\dots,x_k=\pm1} \mathbb{1}_{B_k}(x_1,\dots,x_k) \sum_{x_{k+1},\dots,x_n=\pm1} f(m+x_{k+1}+\dots+x_n-m) = 0\,.$$

□

4a3 Corollary. $\mathbb{E}\left(f(S_n-m)\mathbb{1}_{M_n<m}\right) = \mathbb{E}\left(f(S_n-m)\right)$ for $m \ge 0$ and odd functions f.

4a4 Lemma. $\mathbb{P}\big(M_n < m\big) = \mathbb{P}\big(S_n < m\big) - \mathbb{P}\big(S_n > m\big)$ for all $m \ge 0$.

Proof. Applying 4a3 to $f = \operatorname{sgn}$ and noting that $S_n \le M_n$ we get $-\mathbb{P}\big(M_n < m\big) = \mathbb{P}\big(S_n - m > 0\big) - \mathbb{P}\big(S_n - m < 0\big)$. □

Proof of 4a1.

$$\begin{aligned}\mathbb{P}\big(M_n = m\big) &= \mathbb{P}\big(M_m < m+1\big) - \mathbb{P}\big(M_n < m\big) = \\ &= \mathbb{P}\big(S_n < m+1\big) - \mathbb{P}\big(S_n > m+1\big) - \mathbb{P}\big(S_n < m\big) + \mathbb{P}\big(S_n > m\big) = \\ &= \mathbb{P}\big(S_n = m\big) + \mathbb{P}\big(S_n = m+1\big)\,.\end{aligned}$$

□

4a5 Proposition. For every s, m such that $m \ge 0$ and $m \ge s$,

$$\mathbb{P}\big(S_n = s, M_n = m\big) = \mathbb{P}\big(S_n = 2m - s\big) - \mathbb{P}\big(S_n = 2m - s + 2\big)\,.$$

4a6 Lemma. $\mathbb{P}\big(S_n = m-c, M_n < m\big) = \mathbb{P}\big(S_n = m-c\big) - \mathbb{P}\big(S_n = m+c\big)$ for all $m \ge 0$ and $c \ge 0$.

Proof. For $c = 0$: trivial. For $c > 0$: apply 4a3 to $f(c) = -1$, $f(-c) = 1$, $f(\cdot) = 0$ otherwise. □

In other words,

$$\mathbb{P}\big(S_n = s, M_n < m\big) = \mathbb{P}\big(S_n = s\big) - \mathbb{P}\big(S_n = 2m - s\big)$$

for all $m \ge 0$ and $s \le m$.

Proof of 4a5.

$$\begin{aligned}\mathbb{P}\big(S_n = s, M_n = m\big) &= \mathbb{P}\big(S_n = s, M_n < m+1\big) - \mathbb{P}\big(S_n = s, M_n < m\big) = \\ &= \Big(\mathbb{P}\big(S_n = s\big) - \mathbb{P}\big(S_n = 2(m+1) - s\big)\Big) - \Big(\mathbb{P}\big(S_n = s\big) - \mathbb{P}\big(S_n = 2m - s\big)\Big) = \\ &= \mathbb{P}\big(S_n = 2m - s\big) - \mathbb{P}\big(S_n = 2m - s + 2\big)\,.\end{aligned}$$

□

4a7 Proposition. [1]

For every a, b such that $a > b \ge 0$,

$$\mathbb{P}\big(S_1 > 0, \dots, S_{a+b} > 0 \,\big|\, S_{a+b} = a - b\big) = \frac{a-b}{a+b}\,.$$

The latter is well-known as 'the ballot theorem' (1878): "Suppose that in an election candidate A gets a votes and candidate B gets b votes where $b < a$. Then the (conditional) probability that throughout the counting A always beats B is $(a-b)/(a+b)$."

4a8 Lemma. $\mathbb{P}\big(S_1 < 0, \dots, S_n < 0; S_n = -c\big) = \frac12 \mathbb{P}\big(S_{n-1} = c-1\big) - \frac12 \mathbb{P}\big(S_{n-1} = c+1\big)$ for $c \ge 0$.

Proof.

$$\begin{aligned}
&\mathbb{P}\big(S_1 < 0, \dots, S_n < 0; S_n = -c\big) = \\
&\quad \mathbb{P}\big(S_1 = -1; S_2 - S_1 \le 0, \dots, S_n - S_1 \le 0; S_n - S_1 = -c+1\big) = \\
&\tfrac12 \mathbb{P}\big(S_1 \le 0, \dots, S_{n-1} \le 0; S_{n-1} = -c+1\big) = \tfrac12 \mathbb{P}\big(M_{n-1} < 1; S_{n-1} = -c+1\big) = \\
&\quad = \tfrac12 \Big(\mathbb{P}\big(S_{n-1} = -c+1\big) - \mathbb{P}\big(S_{n-1} = 2\cdot 1 - (-c+1)\big)\Big)\,,
\end{aligned}$$

since $(S_2 - S_1, \dots, S_n - S_1) \sim (S_1, \dots, S_{n-1})$. ☐

In other words, $\mathbb{P}\big(S_1 > 0, \dots, S_n > 0; S_n = s\big) = \frac12 \mathbb{P}\big(S_{n-1} = s-1\big) - \frac12 \mathbb{P}\big(S_{n-1} = s+1\big)$ for all $s \ge 0$.

Proof of 4a7. Denoting $n = a+b$ and $s = a-b$ we have

$$\begin{aligned}
\mathbb{P}\big(S_1 > 0, \dots, S_{a+b} > 0; S_{a+b} = a-b\big) &= \mathbb{P}\big(S_1 > 0, \dots, S_n > 0; S_n = s\big) = \\
&= \frac12 \mathbb{P}\big(S_{n-1} = s-1\big) - \frac12 \mathbb{P}\big(S_{n-1} = s+1\big)\,;
\end{aligned}$$

$$\begin{aligned}
&\mathbb{P}\big(S_1 > 0, \dots, S_{a+b} > 0 \,\big|\, S_{a+b} = a-b\big) = \frac{\mathbb{P}\big(S_{n-1} = s-1\big) - \mathbb{P}\big(S_{n-1} = s+1\big)}{2\mathbb{P}\big(S_n = s\big)} \\
&\quad = \frac{2^{-(n-1)}\binom{n-1}{\frac{n-1}{2} \pm \frac{s-1}{2}} - 2^{-(n-1)}\binom{n-1}{\frac{n-1}{2} \pm \frac{s+1}{2}}}{2 \cdot 2^{-n}\binom{n}{\frac n2 \pm \frac s2}} = \\
&\frac{\frac{n-s}{2}!\,\frac{n+s}{2}!}{n!}\bigg(\frac{(n-1)!}{\frac{n-s}{2}!\,(\frac{n+s}{2}-1)!} - \frac{(n-1)!}{(\frac{n-s}{2}-1)!\,\frac{n+s}{2}!}\bigg) = \frac1n\Big(\frac{n+s}{2} - \frac{n-s}{2}\Big) = \frac sn = \frac{a-b}{a+b}\,.
\end{aligned}$$

☐

[1] [KS, Sect. 6.2, Lemma 6.6], [D, Sect. 3.3].

Here is another use of reflection. Let us say that k is a point of increase if

$$\begin{aligned} S_l < S_k \quad &\text{for } l = 0, \dots, k-1\,, \\ S_l \ge S_k \quad &\text{for } l = k+1, \dots, n\,. \end{aligned}$$

4a9 Proposition. The expected number of points of increase is equal to 1.

However, it is well-known that for large n the walk typically has no points of increase. A paradox! What do you think? A clue: I tried 1000 paths of length $n = 100$ and got the following empirical distribution for the number of points of increase:

value	0	1	2	3	4	5	6	7	8	9	10	11	12	14	19	21
occurs	722	63	45	41	34	24	20	9	14	8	7	1	4	4	2	2

Proof of 4a9. Consider events

A_k : k is a point of increase, that is,

$$S_0 < S_k, \dots, S_{k-1} < S_k,\ S_{k+1} \ge S_k, \dots, S_n \ge S_k;$$

B_k : k is the first maximizer, that is,

$$S_0 < S_k, \dots, S_{k-1} < S_k,\ S_{k+1} \le S_k, \dots, S_n \le S_k\,.$$

We have $\mathbb{P}\big(A_k\big) = \mathbb{P}\big(B_k\big)$ for each k, since $(x_1, \dots, x_n) \in A_k$ if and only if $(x_1, \dots, x_k, -x_{k+1}, \dots, -x_n) \in B_k$. The expected number of points of increase $\sum \mathbb{P}\big(A_k\big)$ is equal to $\sum \mathbb{P}\big(B_k\big) = 1$ (exactly one first maximizer). □

4b Recurrence

The two-dimensional simple random walk is $S_n = X_1 + \dots + X_n$ where X_k are independent identically distributed two-dimensional random vectors taking on the four values $(1, 0)$, $(-1, 0)$, $(0, 1)$, $(0, -1)$ with equal probabilities (0.25). (Note that the first coordinate is *not* a one-dimensional simple random walk.) The d-dimensional simple random walk is defined similarly.

4b1 Theorem. [1] (Polya) The simple d-dimensional random walk returns to the origin (almost surely) infinitely many times if $1 \le d \le 2$ (recurrence), but only finitely many times if $d \ge 3$ (transience).

[1] [D, Sect. 3.2, Th. (2.3)]; [KS, Sect. 6.1, Th. 6.5].

'A drunk man will find his way home but a drunk bird may get lost forever' (Kakutani).

The proof uses Propositions 4b2 and 4b3.

Denote by $p_n^{(d)}$ the probability of the event $S_n = 0$ for the d-dimensional simple random walk $(S_0, \dots, S_n)$. Clearly, $p_n^{(d)} = 0$ for odd n.

4b2 Proposition. [1]

$$
\begin{aligned}
p_{2n}^{(1)} &= 2^{-2n} \binom{2n}{n} ; \\
p_{2n}^{(2)} &= (p_{2n}^{(1)})^2 = 4^{-2n} \binom{2n}{n}^2 ; \\
p_{2n}^{(3)} &= 6^{-2n} \binom{2n}{n} \sum_{k+l+m=n} \binom{n}{k,l,m}^2 .
\end{aligned}
$$

Note that $p_{2n}^{(3)} \ne (p_{2n}^{(1)})^3$.

A d-dimensional random walk (general, not just simple) is $S_n = X_1 + \dots + X_n$ where X_k are independent identically distributed d-dimensional random vectors (their common distribution being arbitrary).

4b3 Proposition. [2] The following three conditions are equivalent for every d-dimensional random walk $(S_n)_n$:

(a) $S_n = 0$ for at least one $n \ge 1$, almost surely;

(b) $S_n = 0$ for infinitely many n, almost surely;

(c) $\sum_{n=1}^\infty \mathbb{P}\big(S_n = 0\big) = \infty$.

***Proof of 4b1**, assuming 4b2 and 4b3.* Case $d = 1$: by 1a13, $p_{2n}^{(1)} \sim \frac{2}{\sqrt{2\pi \cdot 2n}}$. Thus, $\sum p_{2n}^{(1)} = \infty$. Use 4b3.

Case $d = 2$: by 4b2 and the above, $p_{2n}^{(2)} = \big(p_{2n}^{(1)}\big)^2 \sim \frac{4}{2\pi \cdot 2n}$. Still, a divergent series.

Case $d = 3$. First, by 4b3 it is sufficient to prove that the series converges. To this end it is sufficient to prove that

$$
\sum_{k+l+m=n} \binom{n}{k,l,m}^2 \le \text{const} \cdot \frac{3^{2n}}{n} ,
$$

since $p_{2n}^{(3)} = p_{2n}^{(1)} \cdot 3^{-2n} \sum_{k+l+m=n} \binom{n}{k,l,m}^2$ by 4b2, and $\sum \frac1n p_{2n}^{(1)} < \infty$.

[1] [D, Sect. 3.2].

[2] [D, Sect. 3.2, Th. (2.2)]; [KS, Sect. 6.1, Lemma 6.4].

Second, it is sufficient to prove that

$$\max_{k+l+m=n} \binom{n}{k,l,m} \le \text{const} \cdot \frac{3^n}{n} \,,$$

since $\sum_{k+l+m=n} \binom{n}{k,l,m} = 3^n$, and $\sum \binom{n}{k,l,m}^2 \le \max \binom{n}{k,l,m} \cdot \sum \binom{n}{k,l,m}$.

Third, we may assume $n \in 3\mathbb{Z}$, since the maximum is increasing in n; indeed, $\binom{n+1}{k+1,l,m} \ge \binom{n}{k,l,m}$.

The maximum is reached at $k = l = m = n/3$ only (think, why). It remains to prove that

$$\binom{n}{n/3, n/3, n/3} \le \text{const} \cdot \frac{3^n}{n} \quad \text{for } n \in 3\mathbb{Z} \,,$$

which follows easily from the Stirling formula (check it).

Case $d > 3$. We take the 3-dimensional projection of the d-dimensional walk, discard adjacent equal points, and get the 3-dimensional simple random walk; eventually it leaves the origin forever.[1] □

Proof of 4b2. Case $d = 1$: we choose n positions for -1 among the given $2n$ positions ($\binom{2n}{n}$ possibilities).

Case $d = 2$: let $S_k = (S'_k, S''_k)$, then $S'_k - S''_k$ and $S'_k + S''_k$ are independent 1-dimensional simple random walks.

Case $d = 3$: we should have a sum like this:

$$-e_2 + e_3 + e_3 + e_1 - e_3 + e_2 - e_1 - e_3 = 0 \,;$$

we choose the signs first ($\binom{2n}{n}$ possibilities); then, among the n minus terms, we choose some k positions for e_1, l positions for e_2 and m positions for e_3 ($\binom{n}{k,l,m}$ possibilities), and the same among the n plus terms (also $\binom{n}{k,l,m}$ possibilities). □

By the way, you may try to do it otherwise: first, choose $2k$ positions for $\pm e_1$, $2l$ positions for $\pm e_2$ and $2m$ positions for $\pm e_3$, and then choose the signs… Try it also for $d = 2$…

Toward 4b3

Given a random walk (S_n) (general, not just simple; n-dimensional), we define $\tau_1, \tau_2, \dots : \Omega \to \{1, 2, \dots\} \cup \{\infty\}$:

$$\tau_1 = \inf\{n > 0 : S_n = 0\} \,; \quad \tau_2 = \inf\{n > \tau_1 : S_n = 0\} \,; \text{ and so on.}$$

[1] In fact, $p^{(d)}_{2n} \sim \text{const}(d)/n^{d/2}$.

Can we say that random variables $\tau_{n+1} - \tau_n$ are independent, identically distributed? Not quite; it may happen that $\tau_n = \infty$, then necessarily $\tau_{n+1} = \infty$, and $\tau_{n+1} - \tau_n$ is not defined. But still,
(4b4)
$$\mathbb{P}\big(\tau_1 = t_1, \tau_2 - \tau_1 = t_2, \dots, \tau_n - \tau_{n-1} = t_n\big) = \mathbb{P}\big(\tau_1 = t_1\big) \dots \mathbb{P}\big(\tau_1 = t_n\big)$$
for all n and all $t_1, \dots, t_n \in \{1, 2, \dots\}$. (Infinity disallowed!)

Proof of (4b4) *for* $n = 2$. Consider sets (here $s_i = x_1 + \dots + x_i$)
$$A = \{(x_1, \dots, x_{k+l}) : s_1 \ne 0, \dots, s_{k-1} \ne 0,\ s_k = 0,\ s_{k+1} \ne 0, \dots, s_{k+l-1} \ne 0,\ s_{k+l} = 0\};$$
$$B = \{(x_1, \dots, x_k) : s_1 \ne 0, \dots, s_{k-1} \ne 0,\ s_k = 0\};$$
$$C = \{(x_1, \dots, x_l) : s_1 \ne 0, \dots, s_{l-1} \ne 0,\ s_l = 0\}.$$
We have $A = B \times C$;
$$\mathbb{P}\big(\tau_1 = k,\ \tau_2 = k+l\big) = \int \mathbb{1}_A \mathrm{d}\mu^{k+l} = \int_{\mathbb{R}^{k+l}} \mathbb{1}_A(x_1, \dots, x_{k+l})\, \mu(\mathrm{d}x_1) \dots \mu(\mathrm{d}x_{k+l}) =$$
$$= \int_{\mathbb{R}^{k+l}} \mathbb{1}_B(x_1, \dots, x_k) \mathbb{1}_C(x_{k+1}, \dots, x_{k+l})\, \mu(\mathrm{d}x_1) \dots \mu(\mathrm{d}x_{k+l}) =$$
$$\bigg(\int_{\mathbb{R}^k} \mathbb{1}_B(x_1, \dots, x_k)\, \mu(\mathrm{d}x_1) \dots \mu(\mathrm{d}x_k) \bigg) \bigg(\int_{\mathbb{R}^l} \mathbb{1}_C(x_{k+1}, \dots, x_{k+l})\, \mu(\mathrm{d}x_{k+1}) \dots \mu(\mathrm{d}x_{k+l}) \bigg)$$
$$= \bigg(\int \mathbb{1}_B \, \mathrm{d}\mu^k \bigg) \bigg(\int \mathbb{1}_C \, \mathrm{d}\mu^l \bigg) = \mathbb{P}\big(\tau_1 = k\big) \mathbb{P}\big(\tau_1 = l\big).$$
□

The proof for any n is similar.

Thus,
$$\mathbb{P}\big(\tau_2 < \infty\big) = \sum_{k,l} \mathbb{P}\big(\tau_1 = k, \tau_2 = k + l\big) = \sum_{k,l} \mathbb{P}\big(\tau_1 = k\big) \mathbb{P}\big(\tau_1 = l\big) =$$
$$= \Big(\sum_k \mathbb{P}\big(\tau_1 = k\big) \Big)^2 = \big(\mathbb{P}\big(\tau_1 < \infty\big)\big)^2;$$
similarly,

(4b5)
$$\mathbb{P}\big(\tau_n < \infty\big) = \big(\mathbb{P}\big(\tau_1 < \infty\big)\big)^n.$$

Proof of 4b3. We reformulate the conditions in terms of τ_n: (a) $\mathbb{P}\big(\tau_1 < \infty\big) = 1$; (b) $\mathbb{P}\big(\tau_n < \infty\big) = 1$ for all n; (c) $\mathbb{E} \sup\{n : \tau_n < \infty\} = \infty$. Trivially, (b) implies both (a) and (c). By (4b5), (a) implies (b). Finally, (c) implies (a), since $\max\{n : \tau_n < \infty\}$ cannot be distributed geometrically and have infinite expectation. □