Contents

4 Random walks  
4a Reflection ............................................... 40  
4b Recurrence ............................................... 43

4 Random walks

4a Reflection

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

4a1 Proposition. For every \( m \geq 0 \),
\[
P(M_n = m) = P(S_n = m) + P(S_n = m + 1) = 2^{-n} \begin{cases} \left( \frac{n}{m} \right) & \text{for } m + n \text{ even}, \\ \left( \frac{n}{m+1} \right) & \text{for } m + n \text{ odd}. \end{cases}
\]

4a2 Lemma. \( \mathbb{E}(f(S_n - m) \mathbb{1}_{M_n \geq m}) = 0 \) for all \( m \geq 0 \) and every odd (antisymmetric) function \( f \).

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

Proof. For \( m = 0 \): trivial. For \( m > 0 \): define “first hit” events
\[
A_k = \{ S_1 < m, \ldots, S_{k-1} < m, S_k = m \} \quad \text{for } k = 1, \ldots, n;
\]
clearly, \( A_1 \cup \cdots \cup A_n = \{ M_n \geq m \} \); it is sufficient to prove that
\[
\mathbb{E}(f(S_n - m) \mathbb{1}_{A_k}) = 0 \text{ for all } k.
\]
In terms of the corresponding sets \( B_k \subset \mathbb{R}^k \) defined by
\[
B_k = \{ (x_1, \ldots, x_k) : x_1 < m, x_1 + x_2 < m, \ldots, x_1 + \cdots + x_{k-1} < m, x_1 + \cdots + x_k = m \}
\]
\[\text{That is, } \forall x \ f(-x) = -f(x).\]
we have
\[
\mathbb{E} (f(S_n - m) \mathbb{1}_{A_k}) = 2^{-n} \sum_{x_1, \ldots, x_n = \pm 1} f(x_1 + \cdots + x_n - m) \mathbb{1}_{B_k}(x_1, \ldots, x_k) = \\
= 2^{-n} \sum_{x_1, \ldots, x_k = \pm 1} \mathbb{1}_{B_k}(x_1, \ldots, x_k) \sum_{x_{k+1}, \ldots, x_n = \pm 1} f(m + x_{k+1} + \cdots + x_n - m) = 0.
\]

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

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

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

Proof of 4a1
\[
\mathbb{P}(M_n = m) = \mathbb{P}(M_n < m + 1) - \mathbb{P}(M_n < m) = \\
= \mathbb{P}(S_n < m + 1) - \mathbb{P}(S_n > m + 1) - \mathbb{P}(S_n < m) + \mathbb{P}(S_n > m) = \\
= \mathbb{P}(S_n = m) + \mathbb{P}(S_n = m + 1).
\]

4a5 Proposition. For every \( s, m \) such that \( m \geq 0 \) and \( m \geq s \),
\[
\mathbb{P}(S_n = s, M_n = m) = \mathbb{P}(S_n = 2m - s) - \mathbb{P}(S_n = 2m - s + 2).
\]

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

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

In other words,
\[
\mathbb{P}(S_n = s, M_n < m) = \mathbb{P}(S_n = s) - \mathbb{P}(S_n = 2m - s)
\]
for all \( m \geq 0 \) and \( s \leq m \).

Proof of 4a5
\[
\mathbb{P}(S_n = s, M_n = m) = \mathbb{P}(S_n = s, M_n < m + 1) - \mathbb{P}(S_n = s, M_n < m) = \\
= (\mathbb{P}(S_n = s) - \mathbb{P}(S_n = 2(m+1) - s)) - (\mathbb{P}(S_n = s) - \mathbb{P}(S_n = 2m - s)) = \\
= \mathbb{P}(S_n = 2m - s) - \mathbb{P}(S_n = 2m - s + 2).
\]
4a7 Proposition. \(^1\)

For every \(a, b\) such that \(a > b \geq 0\),

\[
P(S_1 > 0, \ldots, S_{a+b} > 0 \mid S_{a+b} = a - b) = \frac{a - b}{a + b}.
\]

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. \(P(S_1 < 0, \ldots, S_n < 0; S_n = -c) = \frac{1}{2}P(S_{n-1} = c - 1) - \frac{1}{2}P(S_{n-1} = c + 1)\) for \(c \geq 0\).

Proof.

\[
P(S_1 < 0, \ldots, S_n < 0; S_n = -c) = P(S_1 = -1; S_2 - S_1 \leq 0, \ldots, S_n - S_1 \leq 0; S_n - S_1 = -c + 1) = \frac{1}{2}P(S_1 \leq 0, \ldots, S_{n-1} \leq 0; S_{n-1} = -c + 1) = \frac{1}{2}(P(S_{n-1} = -c + 1) - P(S_{n-1} = 2 \cdot 1 - (-c + 1))),
\]

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

In other words, \(P(S_1 > 0, \ldots, S_n > 0; S_n = s) = \frac{1}{2}P(S_{n-1} = s - 1) - \frac{1}{2}P(S_{n-1} = s + 1)\) for all \(s \geq 0\).

Proof of 4a7 Denoting \(n = a + b\) and \(s = a - b\) we have

\[
P(S_1 > 0, \ldots, S_{a+b} > 0; S_{a+b} = a - b) = P(S_1 > 0, \ldots, S_n > 0; S_n = s) = \frac{1}{2}P(S_{n-1} = s - 1) - \frac{1}{2}P(S_{n-1} = s + 1);
\]

\[
P(S_1 > 0, \ldots, S_{a+b} > 0 \mid S_{a+b} = a-b) = \frac{P(S_{n-1} = s - 1) - P(S_{n-1} = s + 1)}{2P(S_n = s)} = \frac{2^{-(n-1)}(\frac{n-1}{2}+\frac{1}{2}) - 2^{-(n-1)}(\frac{n-1}{2}+\frac{1}{2})}{2 \cdot 2^{n}(\frac{n}{2}+\frac{1}{2})} = \frac{n-s}{2} \cdot \frac{n+s}{n!} \left( \frac{(n-1)!}{2^{n-s}(\frac{n-s}{2}-1)!} - \frac{(n-1)!}{2^{n-s}(\frac{n-s}{2}-1)!} \right) = 1 \left( n \left( \frac{n+s}{2} - \frac{n-s}{2} \right) \right) = \frac{s}{n} = \frac{a-b}{a+b}.
\]

\(^1\)[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

\[
S_l < S_k \quad \text{for } l = 0, \ldots, k - 1, \\
S_l \geq S_k \quad \text{for } l = k + 1, \ldots, n.
\]

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:

<table>
<thead>
<tr>
<th>value</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>19</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>occurs</td>
<td>722</td>
<td>63</td>
<td>45</td>
<td>41</td>
<td>34</td>
<td>24</td>
<td>20</td>
<td>9</td>
<td>14</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Proof of 4a9 Consider events

$$A_k : \text{ } k \text{ is a point of increase, that is,}$$

$$S_0 < S_k, \ldots, S_{k-1} < S_k, S_{k+1} \geq S_k, \ldots, S_n \geq S_k;$$

$$B_k : \text{ } k \text{ is the first maximizer, that is,}$$

$$S_0 < S_k, \ldots, S_{k-1} < S_k, S_{k+1} \leq S_k, \ldots, S_n \leq S_k.$$

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

4b Recurrence

The two-dimensional simple random walk is $S_n = X_1 + \cdots + X_n$ 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. (Polya) The simple $d$-dimensional random walk returns to the origin (almost surely) infinitely many times if $1 \leq d \leq 2$ (recurrence), but only finitely many times if $d \geq 3$ (transience).

\footnote{[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 $(S_0, \ldots, S_n)$. Clearly, $p_n^d = 0$ for odd $n$.

4b2 Proposition. 1

\[ p_{2n}^{(1)} = 2^{-2n} \binom{2n}{n}; \]
\[ p_{2n}^{(2)} = (p_{2n}^{(1)})^2 = 4^{-2n} \binom{2n}{n}; \]
\[ p_{2n}^{(3)} = 6^{-2n} \binom{2n}{n} \sum_{k+l+m=n} \binom{n}{k,l,m}^2. \]

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

A $d$-dimensional random walk (general, not just simple) is $S_n = X_1 + \cdots + X_n$ 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 $n \geq 1$, almost surely;
(b) $S_n = 0$ for infinitely many $n$, almost surely;
(c) $\sum_{n=1}^\infty P(S_n = 0) = \infty$.

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

Thus, $\sum p_{2n}^{(1)} = \infty$. Use 4b3.

Case $d = 2$: by 4b2 and the above, $p_{2n}^{(2)} = (p_{2n}^{(1)})^2 \sim \frac{1}{2\pi n} \sqrt{\frac{2}{n}}$. 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

\[ \sum_{k+l+m=n} \binom{n}{k,l,m}^2 \leq \text{const} \cdot \frac{3^{2n}}{n}, \]

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

---

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
\[
\max_{k+l+m=n} \binom{n}{k,l,m} \leq \text{const} \cdot \frac{3^n}{n},
\]
since \(\sum_{k+l+m=n} \binom{n}{k,l,m} = 3^n\), and \(\sum (\binom{n}{k,l,m})^2 \leq \max (\binom{n}{k,l,m}) \cdot \sum (\binom{n}{k,l,m})\).

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

The maximum is reached at \(k = l = m = n/3\) only (think, why). It remains to prove that
\[
\binom{n}{n/3, n/3, n/3} \leq \text{const} \cdot \frac{3^n}{n} \quad \text{for } n \in 3\mathbb{Z},
\]
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 \(2n\) possibilities (\(\binom{2n}{n}\) 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).\(^\square\)

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, \ldots : \Omega \to \{1, 2, \ldots \} \cup \{\infty\}:

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

\(^1\)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,

\[ \mathbb{P}(\tau_1 = t_1, \tau_2 - \tau_1 = t_2, \ldots, \tau_n - \tau_{n-1} = t_n) = \mathbb{P}(\tau_1 = t_1) \ldots \mathbb{P}(\tau_1 = t_n) \]

for all $n$ and all $t_1, \ldots, t_n \in \{1, 2, \ldots\}$. (Infinity disallowed!)

**Proof of (4b4) for $n = 2$.** Consider sets (here $s_i = x_1 + \cdots + x_i$)

\[ A = \{(x_1, \ldots, x_{k+l}) : s_1 \neq 0, \ldots, s_{k-1} \neq 0, s_k = 0, s_{k+1} \neq 0, \ldots, s_{k+l-1} \neq 0, s_{k+l} = 0\}; \]

\[ B = \{(x_1, \ldots, x_k) : s_1 \neq 0, \ldots, s_{k-1} \neq 0, s_k = 0\}; \]

\[ C = \{(x_1, \ldots, x_l) : s_1 \neq 0, \ldots, s_{l-1} \neq 0, s_l = 0\}. \]

We have $A = B \times C$;

\[ \mathbb{P}(\tau_1 = k, \tau_2 = k+l) = \int \mathbb{I}_A \, d\mu^{k+l} = \int_{\mathbb{R}^{k+l}} \mathbb{I}_A(x_1, \ldots, x_{k+l}) \, \mu(dx_1) \ldots \mu(dx_{k+l}) = \]

\[ = \int_{\mathbb{R}^{k+l}} \mathbb{I}_B(x_1, \ldots, x_k) \mathbb{I}_C(x_{k+1}, \ldots, x_{k+l}) \, \mu(dx_1) \ldots \mu(dx_{k+l}) = \]

\[ = \left( \int_{\mathbb{R}^{k}} \mathbb{I}_B \, d\mu^k \right) \left( \int_{\mathbb{R}^{l}} \mathbb{I}_C \, d\mu^l \right) = \mathbb{P}(\tau_1 = k) \mathbb{P}(\tau_1 = l). \]

\[ \square \]

The proof for any $n$ is similar.

Thus,

\[ \mathbb{P}(\tau_2 < \infty) = \sum_{k,l} \mathbb{P}(\tau_1 = k, \tau_2 = k+l) = \sum_{k,l} \mathbb{P}(\tau_1 = k) \mathbb{P}(\tau_1 = l) = \]

\[ = \left( \sum_k \mathbb{P}(\tau_1 = k) \right)^2 = \left( \mathbb{P}(\tau_1 < \infty) \right)^2; \]

similarly,

\[ \mathbb{P}(\tau_n < \infty) = \left( \mathbb{P}(\tau_1 < \infty) \right)^n. \]

**Proof of (4b3).** We reformulate the conditions in terms of $\tau_n$: (a) $\mathbb{P}(\tau_1 < \infty) > 0$; (b) $\mathbb{P}(\tau_n < \infty) = 1$ 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 $\mathbb{E} \sup \{n : \tau_n < \infty\}$ cannot be distributed geometrically and have infinite expectation. 

\[ \square \]