Simple random walk

Sven Erick Alm

9 April 2002
(revised 8 March 2006)
(translated to English 28 March 2006)

Contents

1 Introduction

A *random walk* is a stochastic sequence $\{S_n\}$, with $S_0 = 0$, defined by

$$S_n = \sum_{k=1}^{n} X_k,$$

where $\{X_k\}$ are independent and identically distributed random variables (i.i.d.).

The random walk is *simple* if $X_k = \pm 1$, with $P(X_k = 1) = p$ and $P(X_k = -1) = 1-p = q$. Imagine a particle performing a random walk on the integer points of the real line, where it in each step moves to one of its neighboring points; see Figure 1.

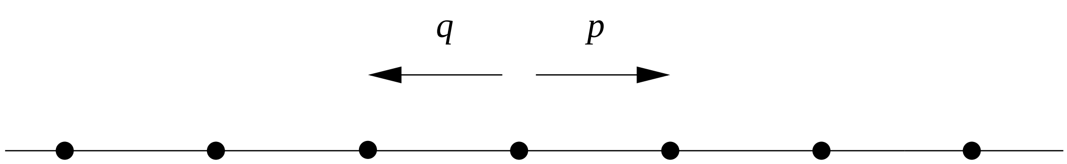

Figure 1: Simple random walk

Remark 1. You can also study random walks in higher dimensions. In two dimensions, each point has 4 neighbors and in three dimensions there are 6 neighbors. □

A simple random walk is *symmetric* if the particle has the same probability for each of the neighbors.

General random walks are treated in Chapter 7 in Ross' book. Here we will only study simple random walks, mainly in one dimension.

We are interested in answering the following questions:

- What is the probability that the particle will ever reach the point a?
  (The case $a = 1$ is often called "The monkey at the cliff".)
- What time does it take to reach a?
- What is the probability of reaching $a > 0$ before $-b < 0$? ("The gambler's ruin")
- If the particle after n steps is at $a > 0$, what is the probability that
  - it has been on the positive side since the first step?
  - it has never been on the negative side?

  ("The Ballot problem")
- How far away from 0 will the particle get in n steps?

When analyzing random walks, one can use a number of general methods, such as

- conditioning,
- generating functions,
- difference equations,

- the theory for Markov chains,
- the theory for branching processes,
- martingales,

but also some more specialized, such as

- counting paths,
- mirroring,
- time reversal.

2 The monkey at the cliff

A monkey is standing one step from the edge of a cliff and takes repeated independent steps; forward, with probability p, or backward, with probability q.

2.1 Passage probabilities

What is the probability that the monkey, sooner or later, will fall off the cliff?

Call this probability P_1. Then

$$P_1 = P(\text{a random walk particle will ever reach } x = 1).$$

We can also study, for $k > 0$,

$$P_k = P(\text{a random walk particle will ever reach } x = k),$$

corresponding to the monkey starting k steps from the edge.
By independence (and the strong Markov property) we get

$$P_k = P_1^k.$$

To determine P_1, condition on the first step.

$$P_1 = p \cdot 1 + q \cdot P_2 = p + q \cdot P_1^2,$$

so that

$$P_1^2 - \frac{1}{q}P_1 + \frac{p}{q} = 0,$$

with solutions

$$\frac{1}{2q} \pm \sqrt{\frac{1}{4q^2} - \frac{p}{q}} = \frac{1}{2q} \pm \frac{\sqrt{1-4pq}}{2q}.$$

Observe that

$$1 = p + q = (p+q)^2 = p^2 + q^2 + 2pq,$$

so that

$$1 - 4pq = p^2 + q^2 - 2pq = (p-q)^2$$

and thus

$$\sqrt{1-4pq} = |p-q|.$$

The solutions can thus be written

$$\frac{1 \pm (p-q)}{2q} = \begin{cases} 1, \\ \frac{p}{q}. \end{cases} \tag{1}$$

If $p > q$ the solution $\frac{p}{q} > 1$ is rejected, so that, for $p \geq q$ (i.e. $p \geq 1/2$), we get $P_1 = 1$, and thus $P_k = 1$ for $k \geq 1$.

It is somewhat more difficult to see that for $p < q$ the correct solution is $P_1 = p/q < 1$, which gives $P_k = (p/q)^k$.

This can be shown using generating functions, e.g. by using the theory for branching processes, where the extinction probability is the smallest positive root to the equation $g(s) = s$; see Problem 2.

A more direct way is to study

$$P_k(n) = P(\text{to reach } x = k \text{ in the } n \text{ first steps}).$$

Again conditioning on the first step gives

$$P_1(n) = p + q \cdot P_2(n-1) \leq p + q \cdot P_1^2(n-1).$$

Since $P_1(1) = p \leq p/q$, we can, by induction, show that $P_1(n) \leq p/q$ for all $n \geq 1$, if we can show that $P_1(n) \leq p/q$ implies that also $P_1(n+1) \leq p/q$.
Suppose that $P_1(n) \leq p/q$. Then

$$P_1(n+1) \leq p + q \cdot P_1^2(n) \leq p + q \cdot (p/q)^2 = p + \frac{p^2}{q} = \frac{p}{q}.$$

Since

$$P_1 = \lim_{n \to \infty} P_1(n) \leq \frac{p}{q},$$

we can, for $p < q$, reject the solution $P_1 = 1$.

We have thus proved

Theorem 1. With P_k defined as above, for $k \geq 1$,

$$P_k = \begin{cases} 1 & \text{if } p \geq q, \\ (\frac{p}{q})^k & \text{if } p < q. \end{cases}$$

Remark 2. This implies that a symmetric random walk, with probability 1, will visit **all** points on the line! □

Problem 1. Let $p < q$. Determine the distribution of $Y = \max_{n \geq 0} S_n$. Compute $E(Y)$.
Hint: Study $P(Y \geq a)$. □

Problem 2. Let $p < q$. Show that P_1 is the extinction probability of a branching process with reproduction distribution $p_0 = p$, $p_2 = q$. □

2.2 Passage times

What time does it take until the monkey falls over the edge?

Let $T_{jk} =$ "the time it takes to go from $x = j$ to $x = k$", so that
$T_{0k} =$ "the time it takes the particle to reach $x = k$ for the first time (when starting in $x = 0$)", and further let

$$E_k = E(T_{0k}),$$

if the expectation exists. If it doeas, we must have, for $k > 0$,

$$E_k = k \cdot E_1.$$

Conditioning gives

$$E_1 = 1 + p \cdot 0 + q \cdot E_2 = 1 + 2q \cdot E_1.$$

Separate the cases $p < q$, $p = q$ and $p > q$.

For $p < q$ we have, by Theorem 1, that $P(T_{01} = \infty) = 1 - P_1 > 0$, which implies that $E_1 = +\infty$.

For $p = q = 1/2$ we get, if E_1 is supposed to be finite, that

$$E_1 = 1 + E_1,$$

that is a contradiction, so that also in this case $E_1 = +\infty$.

Finally, for $p > q$, we get

$$E_1 = \frac{1}{1 - 2q} = \frac{1}{p - q} < \infty.$$

We have proved

Theorem 2. For $k \geq 1$,

$$E_k = \begin{cases} +\infty & \text{if } p \leq q, \\ \frac{k}{p-q} & \text{if } p > q. \end{cases}$$

Using Theorems 1 and 2 we can also study returns to the starting point. Let

$$\begin{aligned} P_0 &= P(\text{that the particle ever returns to the starting point}), \\ T_{00} &= \text{“the time until the first return”}, \\ E_0 &= E(T_{00}). \end{aligned}$$

Theorem 3. $E_0 = +\infty$ for all p and

$$P_0 = \begin{cases} 1 & \text{if } p = q, \\ 1 - |p - q| & \text{if } p \neq q. \end{cases}$$

Proof: With notation as before, we get by conditioning that

$$P_0 = p \cdot P_{-1} + q \cdot P_1.$$

If $p = q$, by Theorem 1, we get $P_{-1} = P_1 = 1$, so that also $P_0 = 1$. If $p \neq q$, either $P_{-1} < 1$ or $P_1 < 1$, so that $P_0 < 1$.
In the case $p < q$, $P_{-1} = 1$ and $P_1 = p/q$, so that

$$P_0 = p + q \cdot \frac{p}{q} = 2p = 1 - (q - p) < 1.$$

In the case $p > q$, instead $P_{-1} = q/p$ and $P_1 = 1$, so that

$$P_0 = p \cdot \frac{q}{p} + q = 2q = 1 - (p - q) < 1.$$

For $p \neq q$, we thus have $P_0 = 1 - |p - q| < 1$ and consequently $P(T_{00} = \infty) > 0$, so that $E_0 = +\infty$.
If $p = q$ we get

$$E_0 = 1 + \frac{1}{2}E_{-1} + \frac{1}{2}E_1 = +\infty,$$

by Theorem 2. □

Remark 3. The symmetric random walk will therefore, with probability 1, return to 0. This holds after each return, so that

$$P(S_n = 0 \text{ i.o.}) = 1.$$

Although the walk will return infinitely often, the expected time between returns is infinite! The law of large numbers can therefore not be interpreted as saying that the particle usually is close to 0. In fact, the particle is rarely close to 0 and a large proportion of the time is spent far away from 0, even in a symmetric random walk! See Section 4.5. □

Remark 4. It can be shown that the symmetric random walk in two dimensions also returns to the origin with probability 1, while in three dimensions the probability is ≈ 0.35. □

Problem 3. Let $p \neq q$. Determine the distribution for $Y =$ # returns to 0. Compute $E(Y)$. □

3 The gambler's ruin

The monkey at the cliff can be interpreted as placing an absorbing barrier at $x = 1$ (or $x = k$). By studying a random walk with two absorbing barriers, one on each side of the staring point, we can solve *The Gambler's ruin:*

Two players, A and B, play a game with independent rounds where, in each round, one of the players wins 1 krona from his opponent; A with probability p and B with probability $q = 1 - p$. A starts the game with a kronor and B with b kronor. The game ends when one of the players is ruined.

3.1 Absorption probabilities

What are the player's ruin probabilities?

This corresponds to a random walk where the particle starts at 0 and is absorbed in the states b and $-a$, or, equivalently, starts in a and is absorbed in 0 and $a + b$.
Let

$$A_k = P(\text{A wins when he has } k \text{ kronor}).$$

Then, $A_0 = 0$, $A_{a+b} = 1$ and we seek A_a. Condition on the outcome of the first round!

$$A_k = p \cdot A_{k+1} + q \cdot A_{k-1}. \tag{2}$$

This homogeneous difference equation can be solved by determining the zeroes of the characteristic polynomial

$$z = p \cdot z^2 + q \Leftrightarrow z^2 - \frac{1}{p} \cdot z + \frac{q}{p} = 0,$$

with solutions $z_1 = 1$ and $z_2 = q/p$. (Compare with (1).) This gives, for $p \neq q$, the following general solution to (2)

$$A_k = C_1 \cdot 1^k + C_2 \cdot (\frac{q}{p})^k,$$

where the constants C_1 and C_2 are determined by the boundary conditions

$$A_0 = 0 \Rightarrow C_1 + C_2 = 0,$$
$$A_{a+b} = 1 \Rightarrow C_1 + C_2(\frac{q}{p})^{a+b} = 1,$$

so that

$$C_1 = \frac{-1}{(\frac{q}{p})^{a+b} - 1},$$
$$C_2 = \frac{1}{(\frac{q}{p})^{a+b} - 1},$$

and

$$A_k = \frac{-1}{(\frac{q}{p})^{a+b} - 1} + \frac{1}{(\frac{q}{p})^{a+b} - 1}(\frac{q}{p})^k = \frac{(\frac{q}{p})^k - 1}{(\frac{q}{p})^{a+b} - 1},$$
$$A_a = \frac{(\frac{q}{p})^a - 1}{(\frac{q}{p})^{a+b} - 1}.$$

For $p = q$, we get the difference equation

$$A_k = \frac{1}{2}A_{k+1} + \frac{1}{2}A_{k-1}.$$

The characteristic polynomial $z^2 - 2z + 1$ has the double root $z_1 = z_2 = 1$, so that we need one more solution. $A_k = k$ works, so that

$$A_k = C_1 \cdot 1^k + C_2 \cdot k.$$

$$A_0 = 0 \Rightarrow C_1 = 0,$$
$$A_{a+b} = 1 \Rightarrow C_2 = \frac{1}{a+b},$$

so that

$$A_k = \frac{k}{a+b},$$
$$A_a = \frac{a}{a+b}.$$

Above we have shown

Theorem 4. The probability that A ruins B (the particle is absorbed in $x = b$) is

$$A_a = \begin{cases} \dfrac{(\frac{q}{p})^a - 1}{(\frac{q}{p})^{a+b} - 1} & \text{if } p \neq \frac{1}{2}, \\ \dfrac{a}{a+b} & \text{if } p = \frac{1}{2}. \end{cases}$$

Remark 5. $a = +\infty, b = 1$ corresponds to *The monkey at the cliff.* $P_1 = P(\text{the monkey falls over the edge}) = \lim_{a\to\infty} P(\text{A wins})$. For $p = q$ we get

$$P_1 = \lim_{a\to\infty} \frac{a}{a+1} = 1$$

and for $p \neq q$

$$P_1 = \lim_{a\to\infty} \frac{(\frac{q}{p})^a - 1}{(\frac{q}{p})^{a+1} - 1} = \begin{cases} 1 & \text{if } p > q, \\ \frac{p}{q} & \text{if } p < q. \end{cases}$$

□

Example 1. Consider a symmetric random walk on the points $0, 1, \ldots, n$, on the circumference of a circle; see Figure 2. The random walk starts at 0. It is easily seen that, with probability 1, all points will be visited. *(Why?)*

What is the probability that the point k ($k = 1, \ldots, n$) is the last to be visited?

Before k is visited one of $k-1$ and $k+1$ must be visited. Consider the time point when this happens for the first time. Because of symmetry, we can assume that it is $k-1$ that is visited. That k is the last point visited means that $k+1$ must be visited before k and this can only occur if the random walk passes clockwise from $k-1$ to $k+1$ before it visits k. The probability for this is the same as the ruin probability for a player with $n-1$ kronor against an opponent with 1 krona, i.e. $1/n$. We then have shown the surprising result that

$$P(k \text{ is the last point visited }) = \frac{1}{n} \text{ for } k = 1, \ldots, n.$$

□

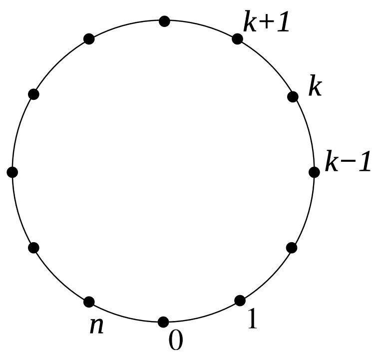

Figure 2: Random walk on a circle

Example 2. Consider a symmetric random walk starting at 0 and a position $a > 0$.
Let $Y_a =$ "# visits in a before the random walk returns to 0". For a to be visited at all, the first step must be to the right, so that $P(Y_a > 0) = \frac{1}{2} \cdot P(Y_a > 0|S_1 = 1)$. This conditional probability is the chance to win for a player with 1 krona against an opponent with $a-1$ kronor, i.e. $P(Y_a > 0|S_1 = 1) = \frac{1}{a}$, so that $P(Y_a > 0) = \frac{1}{2a}$. Similarly, we can compute $P(Y_a = 1|Y_a > 0)$. Consider the random walk when a is visited, which we know will happen if $Y_a > 0$. For this to be the last visit in a, before the next visit to 0, the first step must be to the left, so that $P(Y_a = 1|Y_a > 0) = \frac{1}{2} \cdot P(0 \text{ is reached before } a \text{ starting in } a-1) = \frac{1}{2} \cdot \frac{1}{a} = \frac{1}{2a}$.

The same situation appears at every visit in a, so that $(Y_a|Y_a > 0)$ is ffg($1/2a$) (Geometric distribution starting at 1) and $P(Y_a > 0) = 1/2a$. This gives

$$E(Y_a) = \frac{1}{2a} \cdot 2a = 1$$

for all $a > 0$! (Due to symmetry, this also holds for negative a.)

Above we have shown something very surprising:

Between two visits to 0, the symmetric random walk will make on average 1 visit to **all** *other points!*

For this to be possible, we must have $E_0 = \infty$, which also was shown in Theorem 3. □

Problem 4. Use Theorem 1 to prove that, for all p, a and b, the game will finish with probability 1. □

Problem 5. Suppose that you have 10 kronor and your opponent has 100 kronor. You get the choice to play with stakes of 1, 2, 5 or 10 kronor per round. How would you choose, and what are your chances of winning, if your probability to win a single round is
a) $p = 0.5$, b) $p = 0.4$, c) $p = 0.8$. □

3.2 Absorption times

How long will it take before someone is ruined?

Let

$$Y_k = \text{“\# remaining rounds when A has } k \text{ kronor”},$$
$$E_k = E(Y_k).$$

Conditioning gives

$$E_k = 1 + p \cdot E_{k+1} + q \cdot E_{k-1}, \tag{3}$$

with

$$E_0 = E_{a+b} = 0.$$

Equation (3) is a non-homogeneous difference equation. To solve this we need both to solve the corresponding homogeneous equation, as above, and also to find a particular solution to the inhomogeneous equation.

We start with the symmetric case ($p = q = 1/2$). The difference equation is

$$E_k = 1 + \frac{1}{2} \cdot E_{k+1} + \frac{1}{2} \cdot E_{k-1},$$

with homogeneous solution $A + B \cdot k$. A particular solution is $-k^2$, so that the general solution is

$$E_k = A + B \cdot k - k^2.$$

The boundary conditions are $A = 0$ and $B = a + b$, so that

$$E_k = k \cdot (a + b - k),$$
$$E_a = a \cdot b.$$

In the asymmetric case, the homogeneous solution is, as before, $A + B \cdot (q/p)^k$ and a particular solution is given by $k/(q - p)$, which gives the general solution

$$E_k = \frac{k}{q - p} + A + B \cdot (\frac{q}{p})^k. \tag{4}$$

By using the boundary conditions, we can determine

$$A = \frac{a-b}{(q-p)((\frac{q}{p})^{a+b}-1)},$$
$$B = -A.$$

Inserting this into (4) gives

Theorem 5. The expected number of rounds until one of the players is ruined is

$$E_a = \begin{cases} a \cdot b & \text{if } p = q = \frac{1}{2}, \\ \dfrac{a}{q-p} - \dfrac{a+b}{q-p} \cdot \dfrac{(\frac{q}{p})^a - 1}{(\frac{q}{p})^{a+b}-1} & \text{if } p \neq q. \end{cases}$$

Example 3. In a fair game with $a + b = 100$, we get

a	b	A_a	E_a
1	99	1/100	99
2	98	1/50	196
5	95	1/20	475
10	90	1/10	900
20	80	1/5	1600
25	75	1/4	1875
50	50	1/2	2500

□

Problem 6. Compute the corresponding table as in Example 3 when
a) $p = 0.6$, b) $p = 0.2$. □

Problem 7. For which value of a is E_a maximal/minimal when $a + b = 100$ and
a) $p = 0.4$, b) $p = 0.8$. □

Problem 8. Consider the random walk in Example 1 and let E_k =“# steps until $x = k$ is visited for the first time”. Calculate E_k for $k = 1, \ldots, n$, when
a) $n = 2$, b) $n = 3$, c) $n = 4$. □

Problem 9. In Example 1, let $P(\text{counterclockwise}) = p$ and $P(\text{clockwise}) = q = 1-p$. Show that, for all $0 \leq p \leq 1$, with probability 1, all points will be visited. □

3.3 Reflecting barriers

So far we have studied absorbing barriers. One can also consider *reflecting barriers*. That a is a reflecting barrier means that as soon as the particle reaches $x = a$, in the next step it returns with probability 1 to the previous position. With two reflecting barriers in a and b, $a < b$, there are only $b-a+1$ possible positions, and all will be visited infinitely many times (if $0 < p < 1$).

Example 4. Let 0 be a reflecting barrier, i.e. $P(S_{n+1} = 1 | S_n = 0) = 1$. Consider a symmetric random walk with $S_0 = 0$ and let

$$E_{j,k} = E(\text{the time it takes to go from } j \text{ to } k).$$

Then,

$$\begin{aligned}
E_{0,1} &= 1, \\
E_{1,2} &= 1 + \frac{1}{2} \cdot E_{0,2} = 1 + \frac{1}{2} \cdot (E_{0,1} + E_{1,2}) = \frac{3}{2} + \frac{1}{2} \cdot E_{1,2},
\end{aligned}$$

so that $E_{1,2} = 3$ and $E_{0,2} = 1 + 3 = 4$. In the same way we see that

$$E_{2,3} = 1 + \frac{1}{2} \cdot E_{1,3} = 1 + \frac{1}{2} \cdot (E_{1,2} + E_{2,3}) = \frac{5}{2} + \frac{1}{2} \cdot E_{2,3},$$

so that $E_{2,3} = 5$ and $E_{0,3} = 4 + 5 = 9$. Obviously, for $k = 1, 2, 3$,

$$\begin{aligned}
E_{k-1,k} &= 2k - 1, \\
E_{0,k} &= k^2.
\end{aligned}$$

suppose that this holds for $k \leq n$. Then

$$\begin{aligned}
E_{n,n+1} &= 1 + \frac{1}{2} \cdot E_{n-1,n+1} = 1 + \frac{1}{2} \cdot (E_{n-1,n} + E_{n,n+1}) = 1 + \frac{2n-1}{2} + \frac{1}{2} \cdot E_{n,n+1} \\
&= \frac{2n+1}{2} + \frac{1}{2} \cdot E_{n,n+1} = 2n + 1, \\
E_{0,n+1} &= E_{0,n} + E_{n,n+1} = n^2 + 2n + 1 = (n+1)^2.
\end{aligned}$$

The induction shows that, for $n \geq 1$,

$$\begin{aligned}
E_{n-1,n} &= 2n - 1, \\
E_{0,n} &= n^2.
\end{aligned}$$

□

Problem 10. Determine recursive expressions for $E_{n-1,n}$ and $E_{0,n}$ in the general case ($0 < p < 1$). Show that $E_{0,n} < \infty$ for all $n \geq 1$. What is $E_{0,2}$ expressed in p? □

4 Counting paths

Define the events

$$\begin{aligned}
F_n &= \text{“the particle is at 0 after } n \text{ steps”}, \\
G_n &= \text{“particle returns to 0 for the first time after } n \text{ steps”}
\end{aligned}$$

and the corresponding probabilities

$$\begin{aligned}
f_n &= P(F_n), \\
g_n &= P(G_n).
\end{aligned}$$

Observe that the particle can only return after an even number of steps, so that

$$f_{2n+1} = g_{2n+1} = 0.$$

For n even, the particle must have taken equally many steps to the right and to the left, so that

$$f_{2n} = \binom{2n}{n} p^n q^n.$$

One way of determining g_n is to count paths. This is facilitated if the random walk is illustrated by plotting the pairs (n, S_n); see Figure 3.

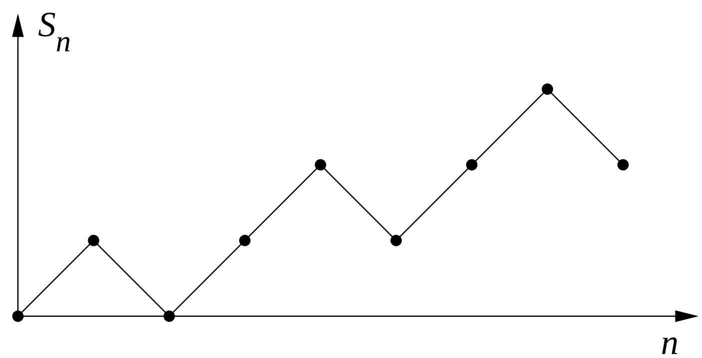

Figure 3: Plotted random walk

Remark 6. Here time is given on the x-axis and the movement of the particle to the right or to the left corresponds to steps up or down in the figure. □

Observe that each path that contains h steps to the right (up) and v steps to the left (down) has probability $p^h \cdot q^v$, so that it often suffices to count the number of such paths in order to determine the requested probabilities.

For example, if we want to determine

$$f_n(a, b) = P(\text{the particle goes from } a \text{ to } b \text{ in } n \text{ steps}),$$

we must have

$$h + v = n \quad \text{and} \quad h - v = b - a,$$

i.e.

$$h = \frac{n + b - a}{2} \quad \text{and} \quad v = \frac{n + a - b}{2}, \tag{5}$$

so that

$$f_n(a, b) = \binom{n}{h} p^h \, q^v,$$

if we define

$$\binom{n}{h} = 0$$

when h is not an integer.

Problem 11. Use Stirling's formula

$$n! \sim \sqrt{2\pi n}\, n^n\, e^{-n}$$

to show that, for $p = q = 1/2$,

a) $f_{2n} \to 0$ when $n \to \infty$, b) $\displaystyle\sum_{n=1}^{\infty} f_{2n} = \infty$, c) $f_{2n} \sim \dfrac{1}{\sqrt{\pi n}}$.

Hint: CLT can be useful! □

Problem 12. What is f_{2n} for a symmetric two-dimensional random walk? □

4.1 Mirroring

To determine g_n introduce the notation

$$\begin{aligned}
N_n(a,b) &= \text{\# paths from } a \text{ to } b \text{ in } n \text{ steps},\\
N_n^{\neq 0}(a,b) &= \text{\# paths from } a \text{ to } b \text{ in } n \text{ steps that do not visit 0},\\
N_n^{0}(a,b) &= \text{\# paths from } a \text{ to } b \text{ in } n \text{ steps that visit 0}.
\end{aligned}$$

Observe that, with h and v as in (5), the following holds

$$\begin{aligned}
N_n(a,b) &= \binom{n}{h},\\
N_n(a,b) &= N_n^{0}(a,b) + N_n^{\neq 0}(a,b),\\
N_n^{\neq 0}(a,b) &= 0 \quad \text{if } a \text{ and } b \text{ have different signs},\\
N_n^{0}(a,b) &= N_n(a,b) \quad \text{if } a \text{ and } b \text{ have different signs}.
\end{aligned}$$

To compute

$$g_{2n} = N_{2n}^{\neq 0}(0,0)\cdot p^n \cdot q^n$$

it is sufficient to determine $N_{2n}^{\neq 0}(0,0)$. Note that

$$N_{2n}^{\neq 0}(0,0) = N_{2n-1}^{\neq 0}(1,0) + N_{2n-1}^{\neq 0}(-1,0) = 2\cdot N_{2n-1}^{\neq 0}(1,0) = 2\cdot N_{2n-2}^{\neq 0}(1,1).$$

To determine $N_{2n-2}^{\neq 0}(1,1)$ we can use the following elegant mirror argument; see Figure 4. *Every path from 1 to 1 that visits 0 can, by mirroring the beginning of the path, up to the first 0-visit, be transformed to a path from -1 to 1 with the same number of steps.*
All paths from -1 to 1 must visit 0!

The above argument gives

$$\begin{aligned}
N_{2n-2}^{0}(1,1) &= N_{2n-2}^{0}(-1,1) = N_{2n-2}(-1,1),\\
N_{2n-2}^{\neq 0}(1,1) &= N_{2n-2}(1,1) - N_{2n-2}^{0}(1,1) = N_{2n-2}(1,1) - N_{2n-2}(-1,1)\\
&= \binom{2n-2}{n-1} - \binom{2n-2}{n} = \binom{2n-2}{n-1}\left(1 - \frac{n-1}{n}\right) = \frac{1}{n}\binom{2n-2}{n-1}.
\end{aligned}$$

This gives

$$N_{2n}^{\neq 0}(0,0) = 2\cdot N_{2n-2}^{\neq 0}(1,1) = 2\frac{1}{n}\binom{2n-2}{n-1} = \frac{1}{2n-1}\binom{2n}{n},$$

so that we have shown

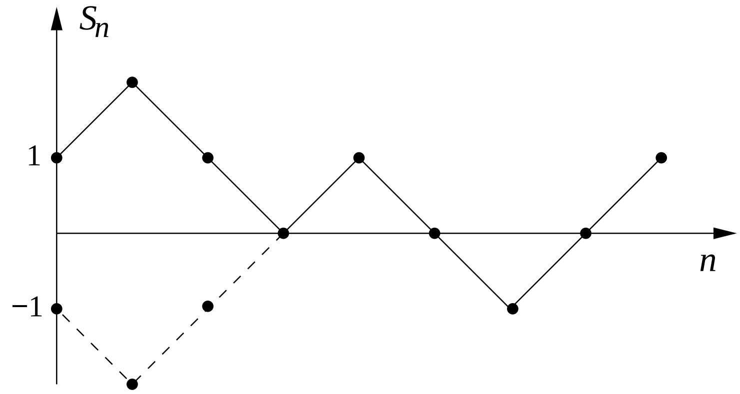

Figure 4: Mirroring

Theorem 6. For $n > 0$,

$$g_{2n} = \frac{1}{2n-1} f_{2n}.$$

The mirroring principle holds in greater generality than what we have used.

Theorem 7. (The mirroring theorem) For $a > 0$ and $b > 0$,

$$\begin{aligned} N_n^0(a,b) &= N_n(-a,b), \\ N_n^{\neq 0}(a,b) &= N_n(a,b) - N_n(-a,b). \end{aligned}$$

Proof: Shown as above. □

Remark 7. The mirroring theorem holds even if n has the wrong parity in relation to a and b as then all the expressions in the theorem equal 0. □

4.2 The Ballot problem

If $a = 0$ in the mirroring theorem we get

Theorem 8. (The Ballot theorem) For $b > 0$,

$$N_n^{\neq 0}(0,b) = \frac{b}{n} \cdot N_n(0,b).$$

Proof: Let $h = \frac{n+b}{2}$ and $v = \frac{n-b}{2}$ be the number of steps to the right and left that are required to go from 0 to b in n steps. Then, $N_n(0,b) = \binom{n}{h}$ and

$$\begin{aligned} N_n^{\neq 0}(0,b) &= N_{n-1}^{\neq 0}(1,b) = N_{n-1}(1,b) - N_{n-1}(-1,b) \\ &= \binom{n-1}{h-1} - \binom{n-1}{h} = \binom{n}{h} \cdot \left(\frac{h}{n} - \frac{v}{n}\right) = \binom{n}{h} \cdot \frac{b}{n}. \end{aligned}$$

□

The theorem has received its name from the following classical probability problem, *The Ballot Problem*:

Example 5. *In an election with two candidates, candidate A receives a total of a votes and candidate B receives a total of b votes, where $a > b$. What is the probability that A will be in the lead during the entire counting of votes?*
If we assume that the votes are counted in random order, the answer is given directly by the Ballot theorem:

$$P(\text{A is leading all through the counting}) = \frac{N^{\neq 0}_{a+b}(0, a-b)}{N_{a+b}(0, a-b)} = \frac{a-b}{a+b}.$$

□

Example 6. A variation of the Ballot problem is:
What is the probability, when $a \geq b$, that B never leads during the counting?
We need to know

$$N^{\neq(-1)}_{a+b}(0, a-b) = N^{\neq 0}_{a+b}(1, a+1-b).$$

By the mirroring theorem we have

$$\begin{aligned} N^{\neq 0}_{a+b}(1, a+1-b) &= N_{a+b}(1, a+1-b) - N_{a+b}(-1, a+1-b) \\ &= \binom{a+b}{a} - \binom{a+b}{a+1} = \binom{a+b}{a} \cdot (1 - \frac{b}{a+1}) \\ &= \binom{a+b}{a} \cdot \frac{a+1-b}{a+1} = \frac{a+1-b}{a+1} \cdot N_{a+b}(0, a-b), \end{aligned}$$

so that

$$P(\text{B never leads}) = \frac{a+1-b}{a+1}.$$

In particular we have that, if $a = b$,

$$P(\text{B never leads}) = \frac{1}{a+1}.$$

□

Remark 8. The Ballot problems are purely combinatoric and the answers do not depend on p (and q). The same holds for random walks if we condition with respect to the end point S_n. In fact, we can express the Ballot problems as

$$\begin{aligned} P(S_k > 0, \text{ for } k = 1, \dots, a+b-1 | S_{a+b} = a-b) &= \frac{a-b}{a+b}, \\ P(S_k \geq 0, \text{ for } k = 1, \dots, a+b-1 | S_{a+b} = a-b) &= \frac{a+1-b}{a+1}. \end{aligned}$$

□

4.3 Recurrence

For our next result we need some more notation. As before, we assume that $S_0 = 0$.

$$\begin{aligned} N^{\neq 0}_n &= \#\text{ paths of length } n \text{ with } S_k \neq 0 \text{ for } k = 1, \dots, n, \\ N^{>0}_n &= \#\text{ paths of length } n \text{ with } S_k > 0 \text{ for } k = 1, \dots, n. \end{aligned}$$

Theorem 9. For $n > 0$,

$$N_{2n}^{\neq 0} = N_{2n}(0,0) = \binom{2n}{n}.$$

For the symmetric random walk,

$$\begin{aligned}
(i) \quad & P(S_k \neq 0,\, k = 1, \dots, 2n) = P(S_{2n} = 0),\\
(ii) \quad & P(S_k > 0,\, k = 1, \dots, 2n) = \frac{1}{2} \cdot P(S_{2n} = 0),\\
(iii) \quad & P(S_k \geq 0,\, k = 1, \dots, 2n) = P(S_{2n} = 0).
\end{aligned}$$

Proof: A path that never visits 0 is either positive all the time or negative all the time. By symmetry,

$$N_{2n}^{\neq 0} = 2 \cdot N_{2n}^{>0} = 2 \cdot \sum_{r=1}^{n} N_{2n}^{\neq 0}(0, 2r).$$

With the same argument as in the proof of the mirroring theorem, for $r > 0$,

$$N_{2n}^{\neq 0}(0, 2r) = N_{2n-1}(1, 2r) - N_{2n-1}(-1, 2r) = N_{2n-1}(0, 2r-1) - N_{2n-1}(0, 2r+1),$$

so that

$$\begin{aligned}
\sum_{r=1}^{n} N_{2n}^{\neq 0}(0, 2r) = (N_{2n-1}(0,1) - N_{2n-1}(0,3)) + (N_{2n-1}(0,3) - N_{2n-1}(0,5)) + \dots\\
+ (N_{2n-1}(0, 2n-1) - N_{2n-1}(0, 2n+1))\\
= N_{2n-1}(0,1) - N_{2n-1}(0, 2n+1) = N_{2n-1}(0,1) = \frac{1}{2} \cdot N_{2n}(0,0).
\end{aligned}$$

(i) follows directly from this as, in a symmetric random walk, all paths of length $2n$ have the same probability.
(ii) follows from

$$P(S_1 \neq 0, \dots, S_{2n} \neq 0) = P(S_1 > 0, \dots, S_{2n} > 0) + P(S_1 < 0, \dots, S_{2n} < 0)$$

and that the two cases are equally probable.
To show (iii), we observe that

$$\begin{aligned}
P(S_1 > 0, \dots, S_{2n} > 0) &= P(S_1 = 1, S_2 > 0, \dots, S_{2n} > 0)\\
&= P(S_1 = 1) \cdot P(S_2 > 0, \dots, S_{2n} > 0 | S_1 = 1)\\
&= \frac{1}{2} \cdot P(S_2 - S_1 \geq 0, S_3 - S_1 \geq 0, \dots, S_{2n} - S_1 \geq 0)\\
&= \frac{1}{2} \cdot P(S_1 \geq 0, \dots, S_{2n-1} \geq 0),
\end{aligned}$$

but since $2n - 1$ is odd,

$$S_{2n-1} \geq 0 \Rightarrow S_{2n-1} \geq 1 \Rightarrow S_{2n} \geq 0,$$

so that, using (ii),

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

$\square$

Remark 9. From the above follows that for a symmetric random walk

$$P(\text{never returning to 0}) = \lim_{n\to\infty} P(S_k \neq 0,\ k = 1,\dots,2n) = \lim_{n\to\infty} P(S_{2n} = 0)$$
$$= \lim_{n\to\infty} f_{2n} = 0,\ \text{(see Problem 11).}$$

□

The previous theorem gave an expression for $P(S_1 \neq 0, \dots, S_n \neq 0)$ for a symmetric random walk. In the general case we have

Theorem 10. (i) For $k > 0$,

$$P(S_1 > 0, \dots, S_{n-1} > 0, S_n = k) = \frac{k}{n} \cdot P(S_n = k).$$

(ii) For $k \neq 0$,

$$P(S_1 \neq 0, \dots, S_{n-1} \neq 0, S_n = k) = \frac{|k|}{n} \cdot P(S_n = k).$$

(iii)

$$P(S_1 \neq 0, \dots, S_n \neq 0) = \frac{E(|S_n|)}{n}.$$

Proof: Both (i) and (ii) are trivially true if $P(S_n = k) = 0$. Therefore we may suppose that k is such that

$$P(S_n = k) = N_n(0,k) \cdot p^h \cdot q^v > 0,\ \text{where}$$
$$h = \frac{n+k}{2}\ \text{and}$$
$$v = \frac{n-k}{2}\ \text{are integers.}$$

Then (i) follows immediately from the Ballot theorem after multiplication with $p^h \cdot q^v$ and (ii) follows by symmetry, as

$$N_n^{\neq 0}(0,k) = N_n^{\neq 0}(0,-k)\ \text{and}$$
$$N_n(0,k) = N_n(0,-k).$$

Summing of (ii) over all $k \neq 0$ gives (iii). □

Theorem 9 can also be used to give an alternative expression for the first passage probability $g_{2n} = P(T_0 = 2n)$, where $T_0 =$ “the time of the first revisit to 0” $= \min\{n \geq 1 : S_n = 0\}$.

Theorem 11. For a symmetric random walk,

$$g_{2n} = f_{2n-2} - f_{2n}.$$

Proof: Theorem 9 says that

$$P(T_0 > 2n) = P(S_k \neq 0,\ k = 1,\dots,2n) = P(S_{2n} = 0) = f_{2n},$$

which gives

$$g_{2n} = P(T_0 = 2n) = P(T_0 > 2n-2) - P(T_0 > 2n) = f_{2n-2} - f_{2n}.$$

□

Remark 10. As $f_0 = 1$ and $f_{2n} \to 0$ when $n \to \infty$, from the theorem follows that, for the symmetric random walk,

$$\begin{aligned} P(\text{ever returning to 0}) &= \sum_{n=1}^{\infty} g_{2n} = \sum_{n=1}^{\infty} (f_{2n-2} - f_{2n}) \\ &= (f_0 - f_2) + (f_2 - f_4) + \cdots = f_0 = 1. \end{aligned}$$

□

Another connection between $\{f_n\}$ and $\{g_n\}$ is given by

Theorem 12. For an arbitrary random walk,

$$f_{2n} = \sum_{r=1}^{n} g_{2r} \cdot f_{2n-2r}.$$

Proof: Condition with respect to the time of the first return to 0, T_0.

$$\begin{aligned} f_{2n} = P(S_{2n} = 0) &= \sum_{r=1}^{n} P(T_0 = 2r) \cdot P(S_{2n} = 0 | T_0 = 2r) \\ &= \sum_{r=1}^{n} g_{2r} \cdot P(S_{2n} - S_{2r} = 0 | T_0 = 2r) \\ &= \sum_{r=1}^{n} g_{2r} \cdot P(S_{2n-2r} = 0) = \sum_{r=1}^{n} g_{2r} \cdot f_{2n-2r}. \end{aligned}$$

□

Another useful trick for analyzing random walks is *time reversal*.
As $\{X_k\}$ are i.i.d., the vector $(X_1, X_2, \ldots, X_n)$ has the same distribution as the vector $(X_n, X_{n-1}, \ldots, X_1)$ and thus $(S_1, S_2, \ldots, S_n)$ has the same distribution as $(X_n, X_n + X_{n-1}, \ldots, X_n + \cdots + X_1) = (S_n - S_{n-1}, S_n - S_{n-2}, \ldots, S_n)$.

Let $T_b =$ "time of the first visit in b" $= \min\{n \geq 1 : S_n = b\}$.

Theorem 13. For $b > 0$,

$$P(T_b = n) = \frac{b}{n} \cdot P(S_n = b), \text{ for } n = b, b+1, \ldots$$

Proof: Using time reversal gives

$$\begin{aligned} P(T_b = n) &= P(S_1 < b, S_2 < b, \ldots, S_{n-1} < b, S_n = b) \\ &= P(S_n > S_1, S_n > S_2, \ldots, S_n > S_{n-1}, S_n = b) \\ &= P(S_n - S_{n-1} > 0, \ldots, S_n - S_1 > 0, S_n = b) \\ &= P(S_1 > 0, \ldots, S_{n-1} > 0, S_n = b) = \frac{b}{n} \cdot P(S_n = b), \end{aligned}$$

by Theorem 10 (i). □

Remark 11. With the help of the theorem we can compute

$$P(T_b > n) = \sum_{k=n+1}^{\infty} P(T_b = k) = \sum_{k=n+1}^{\infty} \frac{b}{k} \cdot P(S_k = b)$$

and through that

$$E(T_b) = \sum_{n=0}^{\infty} P(T_b > n).$$

We know from before that $E(T_b) = b \cdot E(T_1)$, so it is enough to compute $E(T_1)$. □

4.4 Maximum

What is the distribution of $M_n = \max(S_0, S_1, \ldots, S_n)$*?*

Lemma 1. For a symmetric random walk,

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

Proof: The first part follows directly from $M_n \geq S_n$. Suppose that $b < r$. By mirroring the end part of the path, from the last visit in r, in the line $y = r$, we see that "# paths of length n with $M_n = r$ and $S_n = b$" equals "# paths of length n with $S_n = 2r - b$"; see Figure 5. The lemma follows since all paths of length n have the same probability for a symmetric random walk. □

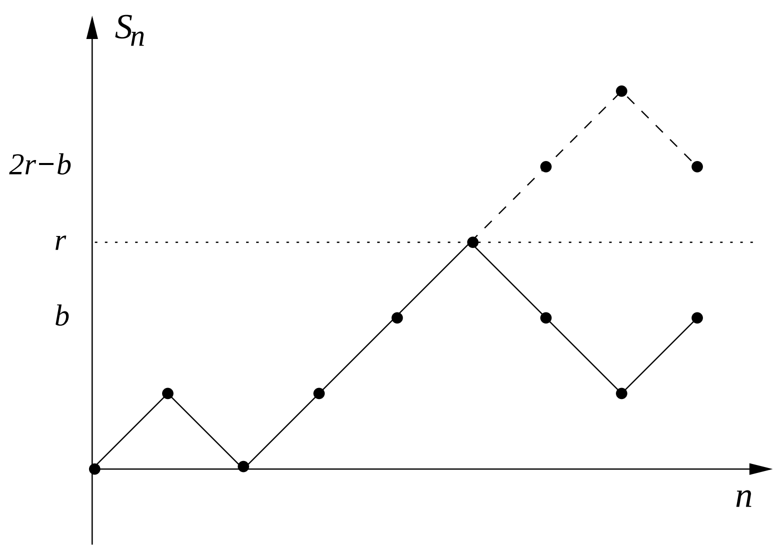

Figure 5: Mirroring for maximum

Theorem 14. For a symmetric random walk, for $r \geq 1$,

$$P(M_n \geq r) = P(S_n = r) + 2P(S_n > r),$$
$$P(M_n = r) = P(S_n = r) + P(S_n = r + 1) = \max(P(S_n = r), P(S_n = r + 1)).$$

Proof: By the lemma,

$$\begin{aligned}
P(M_n \geq r) &= \sum_b P(M_n \geq r, S_n = b) = \sum_{b \geq r} P(S_n = b) + \sum_{b<r} P(S_n = 2r - b) \\
&= P(S_n \geq r) + \sum_{k>r} P(S_n = k) = P(S_n \geq r) + P(S_n > r) \\
&= P(S_n = r) + 2P(S_n > r).
\end{aligned}$$

Further,

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

since only one of $P(S_n = r)$ and $P(S_n = r+1)$ can be non-zero. □

4.5 The Arcsine law

We will show that two stochastic variables with connection to symmetric random walks have the same distribution, the so called *arcsine distribution*.

Consider a symmetric random walk $\{S_n\}$ with $S_0 = 0$.
Let as before $f_n = P(S_n = 0)$.
Define $Y_{2n} = \max(k \leq 2n : S_k = 0)$, which is well defined as $S_0 = 0$, and
$\alpha_{2n}(2k) = P(Y_{2n} = 2k)$. Then,

Theorem 15. For $0 \leq k \leq n$,

$$\alpha_{2n}(2k) = P(Y_{2n} = 2k) = f_{2k} \cdot f_{2n-2k}.$$

Proof:

$$\begin{aligned}
P(Y_{2n} = 2k) &= P(S_{2k} = 0, S_{2k+1} \neq 0, \ldots, S_{2n} \neq 0) \\
&= P(S_{2k} = 0) \cdot P(S_{2k+1} \neq 0, \ldots, S_{2n} \neq 0 | S_{2k} = 0) \\
&= P(S_{2k} = 0) \cdot P(S_1 \neq 0, \ldots, S_{2n-2k} \neq 0) \\
&= P(S_{2k} = 0) \cdot P(S_{2n-2k} = 0) \quad \text{(Theorem 9 (i))} \\
&= f_{2k} \cdot f_{2n-2k}.
\end{aligned}$$

□

Remark 12. The name arcsine distribution is used as

$$P(Y_{2n} \leq 2xn) \sim \frac{2}{\pi} \cdot \arcsin \sqrt{x}, \quad \text{when } n \to \infty.$$

To see this we can use Stirling's formula

$$n! \sim n^n \cdot e^{-n} \sqrt{2\pi n},$$

to show that, for large k,

$$f_{2k} \sim \frac{1}{\sqrt{\pi k}}, \quad \text{(See Problem 11c.)}$$

and that thus

$$\alpha_{2n}(2k) \sim \frac{1}{n} \cdot f(\frac{k}{n}),$$

where

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \quad 0 < x < 1.$$

This gives,

$$P(Y_{2n} \leq 2xn) = \sum_{k \leq xn} \alpha_{2n}(2k) \sim \frac{1}{n} \cdot \sum_{k/n \leq x} f(\frac{k}{n}) \sim \int_0^x f(t)\,dt = \frac{2}{\pi} \cdot \arcsin\sqrt{x}.$$

The density function $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$ is plotted in Figure 6. □

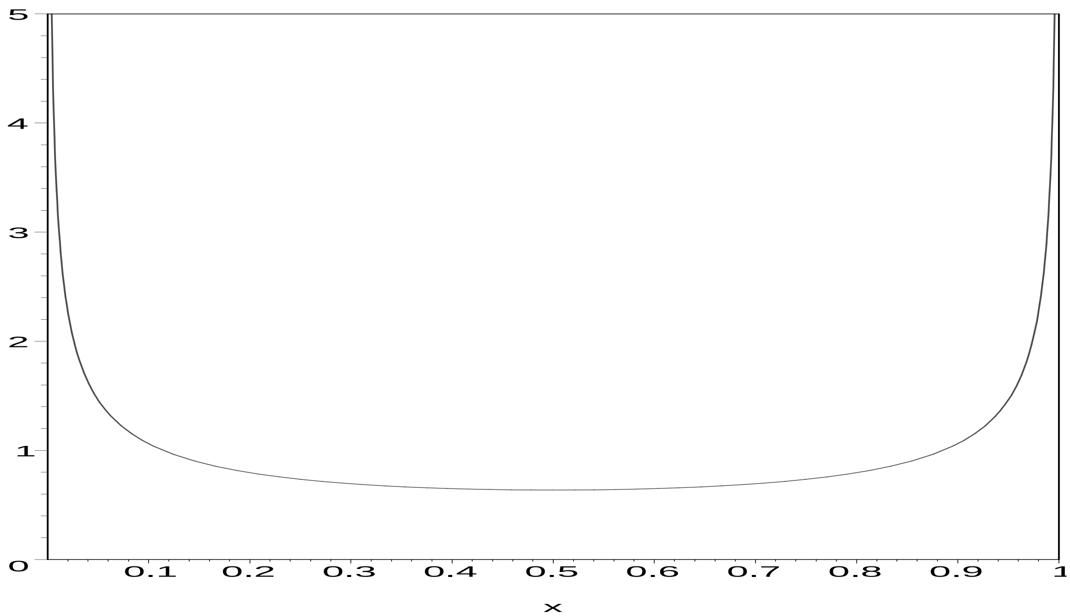

Figure 6: Plot of $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$.

Remark 13. The exact distribution, $\alpha_{2n}(2k)$, is called the *discrete arcsine distribution*. It is, as is the continuous, symmetric around the midpoint, $k = n/2$, where it also takes its minimum. The maximum is attained for $k = 0$ and $k = n$. □

Remark 14. A perhaps surprising property of the symmetric random walk, that is implied by the arcsine law, is that if we have tossed a coin $2n$ times and are interested in when we last had equally many heads and tails, it is most likely that it happened either very recently, or just at the start.
If, for example, $n = 1000$, i.e. we have tossed 2000 times,

$$P(Y_{2000} \leq 200) \sim \frac{2}{\pi} \cdot \arcsin\sqrt{0.1} = 0.205,$$
$$P(Y_{2000} \leq 20) \sim \frac{2}{\pi} \cdot \arcsin\sqrt{0.01} = 0.064.$$

□

There is also an arcsine law for *occupancy times*. We say that the random walk is positive in the time interval $(k, k+1)$ if $S_k > 0$ or $S_{k+1} > 0$. (Natural; see Figure 3.)
Let $Z_n =$ "# positive time intervals between 0 and $2n$". Then Z_{2n} always takes an even value and

Theorem 16. For $0 \leq k \leq n$,

$$P(Z_{2n} = 2k) = \alpha_{2n}(2k) = f_{2k} \cdot f_{2n-2k}.$$

Proof: Let $b_{2n}(2k) = P(Z_{2n} = 2k)$.
We want to show that $b_{2n}(2k) = \alpha_{2n}(2k)$ for all n and $0 \leq k \leq n$.
By Theorem 9 (iii)

$$b_{2n}(2n) = P(S_k \geq 0, k = 1, \ldots, 2n) = f_{2n} = f_{2n} \cdot f_0,$$

so that the theorem holds for $k = n$. By symmetry,

$$b_{2n}(0) = P(S_k \leq 0, k = 1, \ldots, 2n) = f_{2n},$$

so that it also holds for $k = 0$.
It remains to prove that

$$b_{2n}(2k) = \alpha_{2n}(2k)(= f_{2k} \cdot f_{2n-2k}) \quad \text{for all } n \text{ and } 0 < k < n. \tag{6}$$

If $Z_{2n} = 2k$, where $1 \leq k \leq n-1$, we must have $S_{2r} = 0$ for some r, $1 \leq r \leq n-1$, i.e. $T_0 < 2n$. The time up to T_0 is with equal probability spent on the positive and negative side. Conditioning on T_0 gives, for $1 \leq k \leq n-1$,

$$\begin{aligned}
b_{2n}(2k) &= \sum_{r=1}^{n-1} P(T_0 = 2r) \cdot P(Z_{2n} = 2k | T_0 = 2r) \\
&= \sum_{r=1}^{n-1} g_{2r} \cdot \frac{1}{2} \cdot P(Z_{2n-2r} = 2k) + \sum_{r=1}^{n-1} g_{2r} \cdot \frac{1}{2} \cdot P(Z_{2n-2r} = 2k - 2r) \\
&= \frac{1}{2} \cdot \sum_{r=1}^{n-1} g_{2r} \cdot b_{2n-2r}(2k) + \frac{1}{2} \cdot \sum_{r=1}^{n-1} g_{2r} \cdot b_{2n-2r}(2k - 2r).
\end{aligned}$$

Observe that $b_{2n-2r}(2k) = 0$ if $k > n - r$ and that $b_{2n-2r}(2k - 2r) = 0$ if $k < r$, so that

$$b_{2n}(2k) = \frac{1}{2} \cdot \sum_{r=1}^{n-k} g_{2r} \cdot b_{2n-2r}(2k) + \frac{1}{2} \cdot \sum_{r=1}^{k} g_{2r} \cdot b_{2n-2r}(2k - 2r). \tag{7}$$

We will use (7) to show (6) by induction. For $n = 1$ (6) holds trivially. suppose that (6) holds for $n < m$. Then

$$\begin{aligned}
b_{2m}(2k) &= \frac{1}{2} \cdot \sum_{r=1}^{m-k} g_{2r} \cdot b_{2m-2r}(2k) + \frac{1}{2} \cdot \sum_{r=1}^{k} g_{2r} \cdot b_{2m-2r}(2k - 2r) \\
&= \frac{1}{2} \cdot \sum_{r=1}^{m-k} g_{2r} \cdot f_{2k} \cdot f_{2m-2r-2k} + \frac{1}{2} \cdot \sum_{r=1}^{k} g_{2r} \cdot f_{2k-2r} \cdot f_{2m-2k} \\
&= \frac{1}{2} \cdot f_{2k} \cdot \sum_{r=1}^{m-k} g_{2r} \cdot f_{2m-2r-2k} + \frac{1}{2} \cdot f_{2m-2k} \cdot \sum_{r=1}^{k} g_{2r} \cdot f_{2k-2r}.
\end{aligned}$$

By Theorem 12,

$$\sum_{r=1}^{k} g_{2r} \cdot f_{2k-2r} = f_{2k},$$
$$\sum_{r=1}^{m-k} g_{2r} \cdot f_{2(m-k)-2r} = f_{2m-2k},$$

so that

$$b_{2m}(2k) = \frac{1}{2} \cdot f_{2k} \cdot f_{2m-2k} + \frac{1}{2} \cdot f_{2m-2k} \cdot f_{2k} = f_{2k} \cdot f_{2m-2k} = \alpha_{2m}(2k).$$

□

Remark 15. Consider two players, A and B, who are playing a fair game, where both can win one krona from the other with probability 1/2. Many probably interpret The Law of Large Numbers intuitively as that in the long run both players will be in the lead about half of the time. *This is not correct!*
The Arcsine law for occupancy times gives that, after many rounds,

$$P(\text{A leads at least the proportion } x \text{ of the time}) = \frac{2}{\pi} \cdot \arcsin\sqrt{1-x},$$
$$P(\text{A leads at least 80\% of the time}) = \frac{2}{\pi} \cdot \arcsin\sqrt{0.2} = 0.295,$$
$$P(\text{someone leads at least 80\% of the time}) = 2 \cdot \frac{2}{\pi} \cdot \arcsin\sqrt{0.2} = 0.59,$$
$$P(\text{someone leads at least 90\% of the time}) = 2 \cdot \frac{2}{\pi} \cdot \arcsin\sqrt{0.1} = 0.41,$$
$$P(\text{someone leads at least 95\% of the time}) = 2 \cdot \frac{2}{\pi} \cdot \arcsin\sqrt{0.05} = 0.29,$$
$$P(\text{someone leads at least 99\% of the time}) = 2 \cdot \frac{2}{\pi} \cdot \arcsin\sqrt{0.01} = 0.13,$$

□

5 Mixed problems

Problem 13. Consider a random walk with $p < 1/2$, that at time k is in position $a < m$. Calculate the conditional probability that it at time $k+1$ is in position $a+1$ (or in $a-1$) given that it will visit position m in the future. □

Problem 14. Let T_{0a} be defined as in Section 2.2, and $p > 1/2$.
a) Show that

$$\text{Var}(T_{01}) = \frac{4pq}{(p-q)^3}.$$

Hint: Study $E(T_{01}^2)$ and condition.
b) What is Var(T_{0a}) for $a > 0$? □

Problem 15. Consider a player, with probability p of winning a round, who starts with a kronor against an infinitely rich opponent. What is the probability that it takes $a+2k$ rounds before he is ruined? □

Problem 16. Show that there are exactly as many paths (x, y), that end in $(2n + 2, 0)$ and for which $y > 0$ for $0 < x < 2n + 2$, as there are paths that end in $(2n, 0)$ and for which $y \geq 0$ for $0 \leq x \leq 2n$. Show also that this, for a symmetric random walk, implies that

$$P(S_1 \geq 0, \ldots, S_{2n-1} \geq 0, S_{2n} = 0) = 2 \cdot g_{2n+2}. \qquad \square$$

Problem 17. Show that the probability that a symmetric random walk *before* time $2n$ returns exactly r times to 0 is the same as the probability that $S_{2n} = 0$ and that it before that has returned at least r times to 0. $\square$

Problem 18. A particle moves either *two* steps to the right, with probability p, or one step to the left, with probability $q = 1 - p$. Different steps are independent of each other.
a) If it starts in $z > 0$, what is the probability, a_z, that it ever reaches 0?
b) Show that a_1 is the probability that, in a sequence of Bernoulli trials that succeed with probability p, the number of failed trials ever exceeds twice the number of successful trials.
c) Show that, when $p = q$,

$$a_1 = \frac{\sqrt{5} - 1}{2}. \qquad \square$$

Problem 19. Show that, for a symmetric random walk starting in 0, the probability that the first visit in S_{2n} happens in step $2k$ is $P(S_{2k} = 0) \cdot P(S_{2n-2k} = 0)$. $\square$

Problem 20. (Banach's match stick problem)
A person has in each of his two pockets a match box with n matches in each. When he needs a match he randomly chooses one of the boxes, until he encounters an empty box. Let, when this happens, $R =$ # matches in the other box.
a) Calculate $E(R)$.
b) If a) is too difficult; estimate $E(R)$ for $n = 50$ by simulation. $\square$

Problem 21. Let, in a two-dimensional symmetric random walk, starting in the origin, $D_n^2 = x_n^2 + y_n^2$, where (x_n, y_n) is the position of the particle after n steps.
Show that $E(D_n^2) = n$.
Hint: Study $E(D_n^2 - D_{n-1}^2)$. $\square$

Problem 22. Show that a symmetric random walk in d dimensions, with probability 1 will return to an already visited position.
(With probability 1 this happens infinitely many times.)
Hint: In each step the probability of reaching a new point is at most $(2d - 1)/(2d)$. $\square$

6 Literature

A goldmine if you want to read more on random walks is
Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. 1, Third edition, Wiley 1968.
Many of the examples and results are obtained from Feller's book, in particular from Chapter III, but also from Chapter XIV.
Some examples are from
Grimmett, G.R. & Stirzaker, D.R., *Probability and Random Processes*, Second edition, Oxford Science Publications, 1992.
A nice description of many classical probability problems, e.g. random walks, is given in
Blom, G., Holst, L. & Sandell, D., *Problems and Snapshots from the World of Probability*. Springer 1994.