Chapter 1

Random Walk and Discrete Heat Equation

1.1. Simple random walk

We consider one of the basic models for random walk, *simple random walk on the integer lattice* $\mathbb{Z}^d$. At each time step, a random walker makes a random move of length one in one of the lattice directions.

1.1.1. One dimension. We start by studying simple random walk on the integers. At each time unit, a walker flips a fair coin and moves one step to the right or one step to the left depending on whether the coin comes up heads or tails. Let S_n denote the position of the walker at time n. If we assume that the walker starts at x, we can write

$$S_n = x + X_1 + \cdots + X_n$$

where X_j equals ± 1 and represents the change in position between time $j-1$ and time j. More precisely, the increments $X_1, X_2, \ldots$ are independent random variables with $\mathbb{P}\{X_j = 1\} = \mathbb{P}\{X_j = -1\} = 1/2$.

Suppose the walker starts at the origin ($x = 0$). Natural questions to ask are:

- On the average, how far is the walker from the starting point?

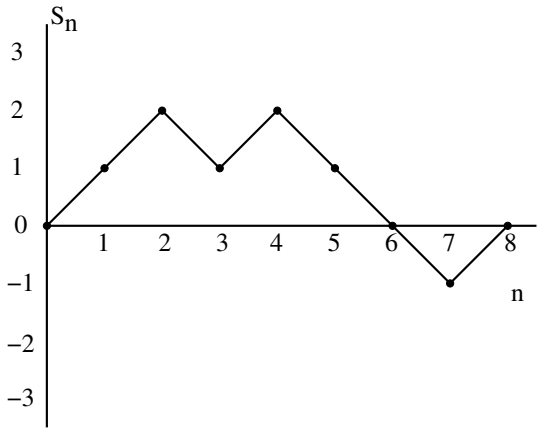

Figure 1. One-dimensional random walk with $x = 0$

- What is the probability that at a particular time the walker is at the origin?
- More generally, what is the probability distribution for the position of the walker?
- Does the random walker keep returning to the origin or does the walker eventually leave forever?

Probabilists use the notation $\mathbb{E}$ for *expectation* (also called *expected value, mean, average value*) defined for discrete random variables by

$$\mathbb{E}[X] = \sum_z z\, \mathbb{P}\{X = z\}.$$

The random walk satisfies $\mathbb{E}[S_n] = 0$ since steps of $+1$ and -1 are equally likely. To compute the average *distance*, one might try to

compute $\mathbb{E}\left[|S_n|\right]$. It turns out to be much easier to compute $\mathbb{E}[S_n^2]$,

$$\begin{aligned}\mathbb{E}[S_n^2] &= \mathbb{E}\left[\left(\sum_{j=1}^{n} X_j\right)^2\right] \\ &= \mathbb{E}\left[\sum_{j=1}^{n}\sum_{k=1}^{n} X_j X_k\right] \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n}\mathbb{E}[X_j X_k] = n + \sum_{j\neq k}\mathbb{E}[X_j X_k].\end{aligned}$$

♢ This calculation uses an important property of average values:

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

The fact that the average of the sum is the sum of the averages for random variables even if the random variables are dependent is easy to prove but can be surprising. For example, if one looks at the rolls of n regular 6-sided dice, the expected value of the sum is $(7/2)\,n$ whether one takes one die and uses that number n times or rolls n different dice and adds the values. In the first case the sum takes on the six possible values $n, 2n, \ldots, 6n$ with probability $1/6$ each while in the second case the probability distribution for the sum is hard to write down explicitly.

If $j \neq k$, there are four possibilities for the (X_j, X_k); for two of them $X_j X_k = 1$ and for two of them $X_j X_k = -1$. Therefore, $\mathbb{E}[X_j X_k] = 0$ for $j \neq k$ and

$$\mathrm{Var}[S_n] = \mathbb{E}[S_n^2] = n.$$

Here Var denotes the variance of a random variable, defined by

$$\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}X)^2\right] = \mathbb{E}[X^2] - (\mathbb{E}X)^2$$

(a simple calculation establishes the second equality). Our calculation illustrates an important fact about variances of sums: if $X_1, \ldots, X_n$ are independent, then

$$\mathrm{Var}[X_1 + \cdots + X_n] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n].$$

♢ The sum rule for expectation and the fact that the cross terms $\mathbb{E}[X_j X_k]$ vanish make it much easier to compute averages of the *square* of a random variable than other powers. In many ways, this is just an analogy of the Pythagorean theorem from geometry: the property $\mathbb{E}[X_j X_k] = 0$, which follows from the fact that the random variables are independent and have mean zero, is the analogue of perpendicularity or orthogonality of vectors.

Finding the probability that the walker is at the origin after n steps is harder than computing $\mathbb{E}[S_n^2]$. However, we can use our computation to make a guess about the size of the probability. Since $\mathbb{E}[S_n^2] = n$, the typical distance away from the origin is of order $\sqrt{n}$. There are about $\sqrt{n}$ integers whose distance is at most $\sqrt{n}$ from the starting point, so one might guess that the probability for being at a particular one should decay like a constant times $n^{-1/2}$. This is indeed the case as we demonstrate by calculating the probability exactly.

It is easy to see that after an odd number of steps the walker is at an odd integer and after an even number of steps the walker is at an even integer. Therefore, $\mathbb{P}\{S_n = x\} = 0$ if $n + x$ is odd. Let us suppose the walker has taken an even number of steps, $2n$. In order for a walker to be back at the origin at time $2n$, the walker must have taken n "+1" steps and n "−1" steps. The number of ways to choose which n steps are $+1$ is $\binom{2n}{n}$ and each particular choice of $2n$ $+1$'s and -1's has probability 2^{-2n} of occurring. Therefore,

$$\mathbb{P}\{S_{2n} = 0\} = \binom{2n}{n} 2^{-2n} = \frac{(2n)!}{n!\, n!} 2^{-2n}.$$

More generally, if the walker is to be at $2j$, there must be $(n + j)$ steps of $+1$ and $(n - j)$ steps of -1. The probabilities for the number of $+1$ steps are given by the binomial distribution with parameters $2n$ and $1/2$,

$$\mathbb{P}\{S_{2n} = 2j\} = \binom{2n}{n+j} 2^{-2n} = \frac{(2n)!}{(n+j)!\,(n-j)!} 2^{-2n}.$$

While these formulas are exact, it is not obvious how to use them because they contain ratios of very large numbers. Trying to understand

the expression on the right-hand side leads to studying the behavior of $n!$ as n gets large. This is the goal of the next section.

1.1.2. Stirling's formula. *Stirling's formula* states that as $n \to \infty$,

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n},$$

where $\sim$ means that the ratio of the two sides tends to 1. We will prove this in the next two subsections. In this subsection we will prove that there is a positive number C_0 such that

$$\lim_{n\to\infty} b_n = C_0, \qquad \text{where} \qquad b_n = \frac{n!}{n^{n+\frac{1}{2}} e^{-n}}, \tag{1.1}$$

and in Section 1.1.3 we show that $C_0 = \sqrt{2\pi}$.

♢ Suppose a_n is a sequence of positive numbers going to infinity and we want to find a positive function $f(n)$ such that $a_n/f(n)$ converges to a positive constant L. Let $b_n = a_n/f(n)$. Then

$$b_n = b_1 \prod_{j=2}^{n} \frac{b_j}{b_{j-1}} = b_1 \prod_{j=2}^{n} [1+\delta_j],$$

where

$$\delta_j = \frac{b_j}{b_{j-1}} - 1,$$

and

$$\lim_{n\to\infty} \log b_n = \log b_1 + \lim_{n\to\infty} \sum_{j=2}^{n} \log[1+\delta_j] = \log b_1 + \sum_{j=2}^{\infty} \log[1+\delta_j],$$

provided that the sum converges. A necessary condition for convergence is that $\delta_n \to 0$. The Taylor's series for the logarithm shows that $|\log[1+\delta_n]| \le c\,|\delta_n|$ for $|\delta_n| \le 1/2$, and hence a sufficient condition for uniform convergence of the sum is that

$$\sum_{n=2}^{\infty} |\delta_n| < \infty.$$

Although this argument proves that the limit exists, it does not determine the value of the limit.

To start, it is easy to check that $b_1 = e$ and if $n \ge 2$, then

$$\frac{b_n}{b_{n-1}} = e \left(\frac{n-1}{n}\right)^{n-\frac{1}{2}} = e \left(1-\frac{1}{n}\right)^{n} \left(1-\frac{1}{n}\right)^{-1/2}. \tag{1.2}$$

Let $\delta_n = (b_n/b_{n-1}) - 1$. We will show that $\sum |\delta_n| < \infty$.

$\diamondsuit$ One of the most important tools for determining limits is Taylor's theorem with remainder, a version of which we now recall. Suppose f is a C^{k+1} function, i.e., a function with $k+1$ derivatives all of which are continuous functions. Let $P_k(x)$ denote the kth-order Taylor series polynomial for f about the origin. Then, for $x > 0$,

$$|f(x) - P_k(x)| \le a_k\, x^{k+1},$$

where

$$a_k = \frac{1}{(k+1)!} \max_{0\le t\le x} |f^{(k+1)}(t)|.$$

A similar estimate is derived for negative x by considering $\tilde{f}(x) = f(-x)$. The Taylor series for the logarithm gives

$$\log(1+u) = u - \frac{u^2}{2} + \frac{u^3}{3} - \cdots,$$

which is valid for $|u| < 1$. In fact, the Taylor series with remainder tells us that for every positive integer k,

$$\log(1+u) = P_k(u) + O(|u|^{k+1}), \tag{1.3}$$

where $P_k(u) = u - (u^2/2) + \cdots + (-1)^{k+1}(u^k/k)$. The $O(|u|^{k+1})$ denotes a term that is bounded by a constant times $|u|^{k+1}$ for small u. For example, there is a constant c_k such that for all $|u| \le 1/2$,

$$|\log(1+u) - P_k(u)| \le c_k\, |u|^{k+1}. \tag{1.4}$$

We will use the $O(\cdot)$ notation as in (1.3) when doing asymptotics — in all cases this will be shorthand for a more precise statement as in (1.4).

We will show that $\delta_n = O(n^{-2})$, i.e., there is a c such that

$$|\delta_n| \le \frac{c}{n^2}.$$

To see this consider $(1-\frac{1}{n})^n$ which we know approaches e^{-1} as n gets large. We use the Taylor series to estimate how fast it converges. We write

$$\begin{aligned}
\log\left(1-\frac{1}{n}\right)^n &= n\log\left(1-\frac{1}{n}\right) \\
&= n\left[-\frac{1}{n} - \frac{1}{2n^2} + O(n^{-3})\right] \\
&= -1 - \frac{1}{2n} + O(n^{-2}),
\end{aligned}$$

and

$$\log\left(1-\frac{1}{n}\right)^{-1/2} = \frac{1}{2n} + O(n^{-2}).$$

By taking logarithms in (1.2) and adding the terms we finish the proof of (1.1). In fact (see Exercise 1.19), we can show that

$$n! = C_0\, n^{n+\frac{1}{2}}\, e^{-n}\left[1+O(n^{-1})\right]. \tag{1.5}$$

1.1.3. Central limit theorem. We now use Stirling's formula to estimate the probability that the random walker is at a certain position. Let S_n be the position of a simple random walker on the integers assuming $S_0 = 0$. For every integer j, we have already seen that the binomial distribution gives us

$$\mathbb{P}\{S_{2n} = 2j\} = \binom{2n}{n+j} 2^{-2n} = \frac{2n!}{(n+j)!(n-j)!}\, 2^{-2n}.$$

Let us assume that $|j| \le n/2$. Then plugging into Stirling's formula and simplifying gives us

$$\begin{aligned}&\mathbb{P}\{S_{2n} = 2j\}\\ &\sim \frac{\sqrt{2}}{C_0}\left(1-\frac{j^2}{n^2}\right)^{-n}\left(1+\frac{j}{n}\right)^{-j}\left(1-\frac{j}{n}\right)^{j}\left(\frac{n}{n^2-j^2}\right)^{1/2}.\end{aligned} \tag{1.6}$$

In fact (if one uses (1.5)), there is a c such that the ratio of the two sides is within distance c/n of 1 (we are assuming $|j| \le n/2$).

What does this look like as n tends to infinity? Let us first consider the case $j = 0$. Then we get that

$$\mathbb{P}\{S_{2n} = 0\} \sim \frac{\sqrt{2}}{C_0\, n^{1/2}}.$$

We now consider j of order $\sqrt{n}$. Note that this confirms our previous heuristic argument that the probability should be like a constant times $n^{-1/2}$, since the typical distance is of order $\sqrt{n}$.

Since we expect S_{2n} to be of order $\sqrt{n}$, let us write an integer j as $j = r\sqrt{n}$. Then the right-hand side of (1.6) becomes

$$\frac{\sqrt{2}}{C_0\sqrt{n}} \left(1 - \frac{r^2}{n}\right)^{-n} \left[\left(1 + \frac{r}{\sqrt{n}}\right)^{-\sqrt{n}}\right]^r \times \left[\left(1 - \frac{r}{\sqrt{n}}\right)^{-\sqrt{n}}\right]^{-r} \left(\frac{1}{1 - (r^2/n)}\right)^{1/2}.$$

♢ We are about to use the well-known limit

$$\left(1 + \frac{a}{n}\right)^n \longrightarrow e^a \qquad n \to \infty.$$

In fact, using the Taylor's series for the logarithm, we get for $n \geq 2a^2$,

$$\log\left(1 + \frac{a}{n}\right)^n = a + O\left(\frac{a^2}{n}\right),$$

which can also be written as

$$\left(1 + \frac{a}{n}\right)^n = e^a \left[1 + O(a^2/n)\right].$$

As $n \to \infty$, the right-hand side of (1.6) is asymptotic to

$$\frac{\sqrt{2}}{C_0\sqrt{n}}\, e^{r^2}\, e^{-r^2}\, e^{-r^2} = \frac{\sqrt{2}}{C_0\sqrt{n}}\, e^{-j^2/n}.$$

For every $a < b$,

$$\lim_{n\to\infty} \mathbb{P}\{a\sqrt{2n} \leq S_{2n} \leq b\sqrt{2n}\} = \lim_{n\to\infty} \sum \frac{\sqrt{2}}{C_0\sqrt{n}}\, e^{-j^2/n}, \tag{1.7}$$

where the sum is over all j with $a\sqrt{2n} \leq 2j \leq b\sqrt{2n}$. The right-hand side is the Riemann sum approximation of an integral where the intervals in the sum have length $\sqrt{2/n}$. Hence, the limit is

$$\int_a^b \frac{1}{C_0}\, e^{-x^2/2}\, dx.$$

This limiting distribution must be a probability distribution, so we can see that

$$\int_{-\infty}^{\infty} \frac{1}{C_0}\, e^{-x^2/2}\, dx = 1.$$

This gives the value $C_0 = \sqrt{2\pi}$ (see Exercise 1.21), and hence Stirling's formula can be written as

$$n! = \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n}\, \left[1 + O(n^{-1})\right].$$

The limit in (1.7) is a statement of the *central limit theorem (CLT)* for the random walk,

$$\lim_{n\to\infty} \mathbb{P}\{a\sqrt{2n} \le S_{2n} \le b\sqrt{2n}\} = \int_a^b \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx.$$

1.1.4. Returns to the origin.

♢ Recall that the sum

$$\sum_{n=1}^{\infty} n^{-a}$$

converges if $a > 1$ and diverges otherwise.

We now consider the number of times that the random walker returns to the origin. Let $J_n = 1\{S_n = 0\}$. Here we use the *indicator function* notation: if E is an event, then 1_E or $1(E)$ is the random variable that takes the value 1 if the event occurs and 0 if it does not occur. The total number of visits to the origin by the random walker is

$$V = \sum_{n=0}^{\infty} J_{2n}.$$

Note that

$$\mathbb{E}[V] = \sum_{n=0}^{\infty} \mathbb{E}[J_{2n}] = \sum_{n=0}^{\infty} \mathbb{P}\{S_{2n} = 0\}.$$

We know that $\mathbb{P}\{S_{2n} = 0\} \sim c/\sqrt{n}$ as $n \to \infty$. Therefore,

$$\mathbb{E}[V] = \infty.$$

It is possible, however, for a random variable to be finite yet have an infinite expectation, so we need to do more work to prove that V is actually infinite.

♢ A well-known random variable with infinite expectation is that obtained from the St. Petersburg Paradox. Suppose you play a game where you flip a coin until you get tails. If you get k heads before flipping tails, then your payoff is 2^k. The probability that you get exactly k heads is the probability of getting

k consecutive heads followed by tails which is 2^{k+1}. Therefore, the expected payoff in this game is

$$2^0 \cdot \frac{1}{2} + 2^1 \cdot \frac{1}{2^2} + 2^2 \cdot \frac{1}{2^3} + \cdots = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots = \infty.$$

Since the expectation is infinite, one should be willing to spend any amount of money in order to play this game once. However, this is clearly not true and here lies the paradox.

Let q be the probability that the random walker ever returns to the origin after time 0. We will show that $q = 1$ by first assuming $q < 1$ and deriving a contradiction. Suppose that $q < 1$. Then we can give the distribution for V. For example, $\mathbb{P}\{V = 1\} = (1 - q)$ since $V = 1$ if and only if the walker never returns after time zero. More generally,

$$\mathbb{P}\{V = k\} = q^{k-1}\,(1 - q), \quad k = 1, 2, \ldots.$$

This tells us that

$$\mathbb{E}[V] = \sum_{k=1}^{\infty} k\,\mathbb{P}\{V = k\} = \sum_{k=1}^{\infty} k\,q^{k-1}\,(1 - q) = \frac{1}{1 - q} < \infty;$$

but we know that $\mathbb{E}[V] = \infty$. Hence, it must be the case that $q = 1$. We have established the following.

Theorem 1.1. *The probability that a (one-dimensional) simple random walker returns to the origin infinitely often is one.*

Note that this also implies that if the random walker starts at $x \neq 0$, then the probability that it will get to the origin is one.

♢ Another way to compute $\mathbb{E}[V]$ in terms of q is to argue that

$$\mathbb{E}[V] = 1 + q\,\mathbb{E}[V].$$

The 1 represents the first visit; q is the probability of returning to the origin; and the key observation is that the expected number of visits after the first visit, given that there is a second visit, is exactly the expected number of visits starting at the origin. Solving this simple equation gives $\mathbb{E}[V] = (1 - q)^{-1}$.

1.1.5. Several dimensions. We now consider a random walker on the d-dimensional integer grid

$$\mathbb{Z}^d = \{(x_1, \ldots, x_d) : x_j \text{ integers}\}.$$

At each time step, the random walker chooses one of its $2d$ nearest neighbors, each with probability $1/2d$, and moves to that site. Again, we let

$$S_n = x + X_1 + \cdots + X_n$$

denote the position of the particle. Here $x, X_1, \ldots, X_n, S_n$ represent points in $\mathbb{Z}^d$, i.e., they are d-dimensional vectors with integer components. The increments $X_1, X_2, \ldots$ are unit vectors with one component of absolute value 1. Note that $X_j \cdot X_j = 1$ and if $j \neq k$, then $X_j \cdot X_k$ equals 1 with probability $1/(2d)$; equals -1 with probability $1/(2d)$; and otherwise equals zero. In particular, $\mathbb{E}[X_j \cdot X_j] = 1$ and $\mathbb{E}[X_j \cdot X_k] = 0$ if $j \neq k$. Suppose $S_0 = 0$. Then $\mathbb{E}[S_n] = 0$, and a calculation as in the one-dimensional case gives

$$\mathbb{E}[|S_n|^2] = \mathbb{E}[S_n \cdot S_n] = \mathbb{E}\left[\left(\sum_{j=1}^{n} X_j\right) \cdot \left(\sum_{j=1}^{n} X_j\right)\right] = n.$$

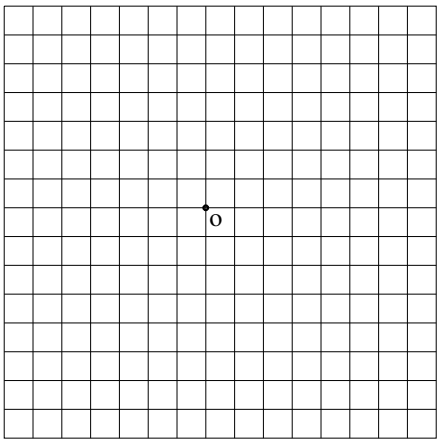

Figure 2. The integer lattice $\mathbb{Z}^2$

What is the probability that we are at the origin after n steps assuming $S_0 = 0$? This is zero if n is odd. If n is even, let us give a heuristic argument. The typical distance from the origin of S_n is of order $\sqrt{n}$. In d dimensions the number of lattice points within distance $\sqrt{n}$ grows like $(\sqrt{n})^d$. Hence, the probability that we choose a particular point should decay like a constant times $n^{-d/2}$.

The combinatorics for justifying this is a little more complicated than in the one-dimensional case, so we will just wave our hands to get the right behavior. In $2n$ steps, we expect that approximately $2n/d$ of them will be taken in each of the d possible directions (e.g., if $d = 2$, we expect about n horizontal and n vertical steps). In order to be at the origin, we need to take an even number of steps in each of the d-directions. The probability of this (Exercise 1.17) is $2^{-(d-1)}$. Given that each of these numbers is even, the probability that each individual component is at the origin is the probability that a one-dimensional walk is at the origin at time $2n/d$ (or, more precisely, an even integer very close to $2n/d$). Using this idea we get the asymptotics

$$\mathbb{P}\{S_{2n} = 0\} \sim \frac{c_d}{n^{d/2}}, \qquad c_d = \frac{d^{d/2}}{\pi^{d/2}\, 2^{d-1}}.$$

The particular value of c_d will not be important to us, but the fact that the exponent of n is $d/2$ is very important.

Consider the expected number of returns to the origin. If V is the number of visits to the origin, then just as in the $d = 1$ case,

$$\mathbb{E}[V] = \sum_{n=0}^{\infty} \mathbb{P}\{S_{2n} = 0\}.$$

Also,

$$\mathbb{E}[V] = \frac{1}{1-q},$$

where $q = q_d$ is the probability that the d-dimensional walk returns to the origin. Since $\mathbb{P}\{S_{2n} = 0\} \sim c/n^{d/2}$,

$$\mathbb{E}[V] = \sum_{n=0}^{\infty} \mathbb{P}\{S_{2n} = 0\} = \begin{cases} < \infty, & d \geq 3, \\ = \infty, & d = 2 \end{cases}.$$

Theorem 1.2. *Suppose S_n is simple random walk in $\mathbb{Z}^d$ with $S_0 = 0$. If $d = 1, 2$, the random walk is* recurrent*, i.e., with probability one it*

returns to the origin infinitely often. If $d \geq 3$, the random walk is transient*, i.e., with probability one it returns to the origin only finitely often. Also,*

$$\mathbb{P}\{S_n \neq 0 \text{ for all } n > 0\} > 0 \text{ if } d \geq 3.$$

1.1.6. Notes about probability. We have already implicitly used some facts about probability. Let us be more explicit about some of the rules of probability. A *sample space* or *probability space* is a set Ω and *events* are a collection of subsets of Ω including $\emptyset$ and Ω. A probability $\mathbb{P}$ is a function from events to $[0,1]$ satisfying $\mathbb{P}(\Omega) = 1$ and the following countable additivity rule:

- If $E_1, E_2, \ldots$ are disjoint (mutually exclusive) events, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(E_n).$$

We do not assume that $\mathbb{P}$ is defined for every subset of Ω, but we do assume that the collection of events is closed under countable unions and "complementation", i.e., if $E_1, E_2, \ldots$ are events so are $\bigcup E_j$ and $\Omega \setminus E_j$.

$\diamondsuit$ The assumptions about probability are exactly the assumptions used in measure theory to define a measure. We will not discuss the difficulties involved in proving such a probability exists. In order to do many things in probability rigorously, one needs to use the theory of Lebesgue integration. We will not worry about this in this book.

We do want to discuss one important lemma that probabilists use all the time. It is very easy, but it has a name. (It is very common for mathematicians to assign names to lemmas that are used frequently even if they are very simple—this way one can refer to them easily.)

Lemma 1.3 (Borel-Cantelli Lemma)**.** *Suppose $E_1, E_2, \ldots$ is a collection of events such that*

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty.$$

Then with probability one at most finitely many of the events occur.

Proof. Let A be the event that infinitely many of $E_1, E_2, \ldots$ occur. For each integer N, $A \subset A_N$ where A_N is the event that at least one of the events $E_N, E_{N+1}, \ldots$ occurs. Then,

$$\mathbb{P}(A) \leq \mathbb{P}(A_N) = \mathbb{P}\left(\bigcup_{n=N}^{\infty} E_n\right) \leq \sum_{n=N}^{\infty} \mathbb{P}(E_n);$$

but $\sum \mathbb{P}(E_n) < \infty$ implies

$$\lim_{N\to\infty} \sum_{n=N}^{\infty} \mathbb{P}(E_n) = 0.$$

Hence, $\mathbb{P}(A) = 0$. □

As an example, consider the simple random walk in $\mathbb{Z}^d, d \geq 3$ and let E_n be the event that $S_n = 0$. Then, the estimates of the previous section show that

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty,$$

and hence with probability one, only finitely many of the events E_n occur. This says that with probability one, the random walk visits the origin only finitely often.

1.2. Boundary value problems

1.2.1. One dimension: Gambler's ruin. Suppose N is a positive integer and a random walker starts at $x \in \{0, 1, \ldots, N\}$. Let S_n denote the position of the walker at time n. Suppose the walker stops when the walker reaches 0 or N. To be more precise, let

$$T = \min\{n : S_n = 0 \text{ or } N\}.$$

Then the position of the walker at time n is given by $\hat{S}_n = S_{n\wedge T}$ where $n \wedge T$ means the minimum of n and T. It is not hard to see that with probability one, $T < \infty$, i.e., eventually the walker will reach 0 or N and then stop. Our goal is to try to figure out which point it stops at. Define the function $F : \{0, \ldots, N\} \to [0, 1]$ by

$$F(x) = \mathbb{P}\{S_T = N \mid S_0 = x\}.$$

$\diamondsuit$ Recall that if V_1, V_2 are events, then $\mathbb{P}(V_1 \mid V_2)$ denotes the conditional probability of V_1 given V_2. It is defined by

$$\mathbb{P}(V_1 \mid V_2) = \frac{\mathbb{P}(V_1 \cap V_2)}{\mathbb{P}(V_2)},$$

assuming $\mathbb{P}(V_2) > 0$.

We can give a gambling interpretation to this by viewing S_n as the number of chips currently held by a gambler who is playing a fair game where at each time duration the player wins or loses one chip. The gambler starts with x chips and plays until he or she has N chips or has gone bankrupt. The chance that the gambler does not go bankrupt before attaining N is $F(x)$. Clearly, $F(0) = 0$ and $F(N) = 1$. Suppose $0 < x < N$. After the first game, the gambler has either $x - 1$ or $x + 1$ chips, and each of these outcomes is equally likely. Therefore,

$$F(x) = \frac{1}{2}\,F(x+1) + \frac{1}{2}\,F(x-1), \quad x = 1, \ldots, N-1. \tag{1.8}$$

One function F that satisfies (1.8) with the boundary conditions $F(0) = 0, F(N) = 1$ is the linear function $F(x) = x/N$. In fact, this is the only solution as we shall now see.

Theorem 1.4. *Suppose a, b are real numbers and N is a positive integer. Then the only function $F : \{0, \ldots, N\} \to \mathbb{R}$ satisfying* (1.8) *with $F(0) = a$ and $F(N) = b$ is the linear function*

$$F_0(x) = a + \frac{x(b-a)}{N}.$$

This is a fairly easy theorem to prove. In fact, we will give several proofs. This is not just to show off how many proofs we can give! It is often useful to give different proofs to the same theorem because it gives us a number of different approaches for trying to prove generalizations. It is immediate that F_0 satisfies the conditions; the real question is one of uniqueness. We must show that F_0 is the *only* such function.

Proof 1. Consider the set $\mathcal{V}$ of all functions $F : \{0, \ldots, N\} \to \mathbb{R}$ that satisfy (1.8). It is easy to check that $\mathcal{V}$ is a vector space, i.e., if

$f, g \in \mathcal{V}$ and c_1, c_2 are real numbers, then $c_1 f + c_2 g \in \mathcal{V}$. In fact, we claim that this vector space has dimension two. To see this, we will give a basis. Let f_1 be the function defined by $f_1(0) = 0, f_1(1) = 1$ and then extended in a unique way to satisfy (1.8). In other words, we define $f_1(x)$ for $x > 1$ by

$$f_1(x) = 2f_1(x-1) - f_1(x-2).$$

It is easy to see that f_1 is the only solution to (1.8) satisfying $f_1(0) = 0, f_1(1) = 1$. We define f_2 similarly with initial conditions $f_2(0) = 1, f_2(1) = 0$. Then $c_1 f_1 + c_2 f_2$ is the unique solution to (1.8) satisfying $f_1(0) = c_2, f_1(1) = c_1$. The set of functions of the form F_0 as a, b vary form a two-dimensional subspace of $\mathcal{V}$ and hence must be all of $\mathcal{V}$.

$\diamondsuit$ The set of all functions $f : \{0, \dots, N\} \to \mathbb{R}$ is essentially the same as $\mathbb{R}^{N+1}$. One can see this by associating to the function f the vector $(f(0), f(1), \dots, f(N))$. The set $\mathcal{V}$ is a subspace of this vector space. Recall to show that a subspace has dimension k, it suffices to find a basis for the subspace with k elements $v_1, \dots, v_k$. To show they form a basis, we need to show that they are linearly independent and that every vector in the subspace is a linear combination of them.

Proof 2. Suppose F is a solution to (1.8). Then for each $0 < x < N$,

$$F(x) \le \max\{F(x-1), F(x+1)\}.$$

Using this we can see that the maximum of F is obtained either at 0 or at N. Similarly, the minimum of F is obtained on $\{0, N\}$. Suppose $F(0) = 0, F(N) = 0$. Then the minimum and the maximum of the function are both 0 which means that $F \equiv 0$. Suppose $F(0) = a, F(N) = b$ and let F_0 be the linear function with these same boundary values. Then $F - F_0$ satisfies (1.8) with boundary value 0, and hence is identically zero. This implies that $F = F_0$.

Proof 3. Consider the equations (1.8) as $N - 1$ linear equations in $N - 1$ unknowns, $F(1), \dots, F(N-1)$. We can write this as

$$\mathbf{A}\mathbf{v} = \mathbf{w},$$

where

$$\mathbf{A} = \begin{bmatrix} -1 & \frac{1}{2} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{2} & -1 & \frac{1}{2} & \cdots & 0 & 0 \\ & & & & \vdots & & \\ 0 & 0 & 0 & 0 & \cdots & -1 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{2} & -1 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} -\frac{F(0)}{2} \\ 0 \\ 0 \\ \vdots \\ 0 \\ -\frac{F(N)}{2} \end{bmatrix}.$$

If we prove that $\mathbf{A}$ is invertible, then the unique solution is $\mathbf{v} = \mathbf{A}^{-1}\mathbf{w}$. To prove invertibility it suffices to show that $\mathbf{A}\mathbf{v} = 0$ has a unique solution and this can be done by an argument as in the previous proof.

Proof 4. Suppose F is a solution to (1.8). Let S_n be the random walk starting at x. We claim that for all n, $\mathbb{E}[F(S_{n\wedge T})] = F(x)$. We will show this by induction. For $n = 0$, $F(S_0) = F(x)$ and hence $\mathbb{E}[F(S_0)] = F(x)$. To do the inductive step, we use a rule for expectation in terms of conditional expectations:

$$\mathbb{E}[F(S_{(n+1)\wedge T})] = \sum_{y=0}^{N} \mathbb{P}\{S_{n\wedge T} = y\}\, \mathbb{E}[F(S_{(n+1)\wedge T}) \mid S_{n\wedge T} = y].$$

If $y = 0$ or $y = N$ and $S_{n\wedge T} = y$, then $S_{(n+1)\wedge T} = y$ and hence $\mathbb{E}[F(S_{(n+1)\wedge T}) \mid S_{n\wedge T} = y] = F(y)$. If $0 < y < x$ and $S_{n\wedge T} = y$, then

$$\mathbb{E}[F(S_{(n+1)\wedge T}) \mid S_{n\wedge T} = y] = \frac{1}{2}\, F(y+1) + \frac{1}{2}\, F(y-1) = F(y).$$

Therefore,

$$\mathbb{E}[F(S_{(n+1)\wedge T})] = \sum_{y=0}^{N} \mathbb{P}\{S_{n\wedge T} = y\}\, F(y) = \mathbb{E}[F(S_{n\wedge T})] = F(x),$$

with the last equality holding by the inductive hypothesis. Therefore,

$$\begin{aligned} F(x) &= \lim_{n\to\infty} \mathbb{E}[F(S_{n\wedge T})] \\ &= \lim_{n\to\infty} \sum_{y=0}^{N} \mathbb{P}\{S_{n\wedge T} = y\}\, F(y) \\ &= \mathbb{P}\{S_T = 0\}\, F(0) + \mathbb{P}\{S_T = N\}\, F(N) \\ &= [1 - \mathbb{P}\{S_T = N\}]\, F(0) + \mathbb{P}\{S_T = N\}\, F(N). \end{aligned}$$

Considering the case $F(0) = 0, F(N) = 1$ gives $\mathbb{P}\{S_T = N \mid S_0 = x\} = x/N$ and for more general boundary conditions,

$$F(x) = F(0) + \frac{x}{N}\,[F(N) - F(0)].$$

One nice thing about the last proof is that it was not necessary to have already guessed the linear functions as solutions. The proof produces these solutions.

1.2.2. Higher dimensions. We will generalize this result to higher dimensions. We replace the interval $\{1, \dots, N\}$ with an arbitrary finite subset A of $\mathbb{Z}^d$. We let ∂A be the *(outer) boundary* of A defined by

$$\partial A = \{z \in \mathbb{Z}^d \setminus A : \operatorname{dist}(z, A) = 1\},$$

and we let $\overline{A} = A \cup \partial A$ be the *"closure"* of A.

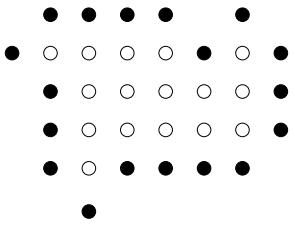

Figure 3. The white dots are A and the black dots are ∂A

$\diamondsuit$ The term closure may seem strange, but in the continuous analogue, A will be an open set, ∂A its topological boundary and $\overline{A} = A \cup \partial A$ its topological closure.

We define the linear operators $\mathbf{Q}, \mathcal{L}$ on functions by

$$\mathbf{Q}F(x) = \frac{1}{2d} \sum_{y \in \mathbb{Z}^d, |x-y|=1} F(y),$$

$$\mathcal{L}F(x) = (\mathbf{Q} - \mathbf{I})F(x) = \frac{1}{2d} \sum_{y \in \mathbb{Z}^d, |x-y|=1} [F(y) - F(x)].$$

The operator $\mathcal{L}$ is often called the *(discrete) Laplacian*. We let S_n be a simple random walk in $\mathbb{Z}^d$. Then we can write

$$\mathcal{L}F(x) = \mathbb{E}[F(S_1) - F(S_0) \mid S_0 = x].$$

We say that F is *(discrete) harmonic* at x if $\mathcal{L}F(x) = 0$; this is an example of a *mean-value property*. The corresponding boundary value problem we will state is sometimes called the *Dirichlet problem* for harmonic functions.

$\diamondsuit$ The term linear operator is often used for a linear function whose domain is a space of functions. In our case, the domain is the space of functions on the finite set A which is isomorphic to $\mathbb{R}^K$ where $K = \#(A)$. In this case a linear operator is the same as a linear transformation from linear algebra. We can think of $\mathbf{Q}$ and $\mathcal{L}$ as $K \times K$ matrices. We can write $\mathbf{Q} = [Q(x,y)]_{x,y\in A}$ where $Q(x,y) = 1/(2d)$ if $|x-y| = 1$ and otherwise $Q(x,y) = 0$. Define $Q_n(x,y)$ by $\mathbf{Q}^n = [Q_n(x,y)]$. Then $Q_n(x,y)$ is the probability that the random walk starts at x, is at site y at time n, and and has not left the set A by time n.

Dirichlet problem for harmonic functions. Given a set $A \subset \mathbb{Z}^d$ and a function $F : \partial A \to \mathbb{R}$ find an extension of F to $\overline{A}$ such that F is harmonic in A, i.e.,

$$\mathcal{L}F(x) = 0 \text{ for all } x \in A. \tag{1.9}$$

For the case $d = 1$ and $A = \{1, \ldots, N-1\}$, we were able to guess the solution and then verify that it is correct. In higher dimensions, it is not so obvious how to give a formula for the solution. We will show that the last proof for $d = 1$ generalizes in a natural way to $d > 1$. We let $T_A = \min\{n \geq 0 : S_n \notin A\}$.

Theorem 1.5. *If $A \subset \mathbb{Z}^d$ is finite, then for every $F : \partial A \to \mathbb{R}$, there is a unique extension of F to $\overline{A}$ that satisfies* (1.9). *It is given by*

$$F_0(x) = \mathbb{E}[F(S_{T_A}) \mid S_0 = x] = \sum_{y\in\partial A} \mathbb{P}\{S_{T_A} = y \mid S_0 = x\}\, F(y).$$

It is not difficult to verify that F_0, as defined above, is a solution to the Dirichlet problem. The problem is to show that it is unique. Suppose F is harmonic on A, $S_0 = x \in \overline{A}$, and let

$$M_n = F(S_{n\wedge T_A}).$$

Then (1.9) can be rewritten as

$$\mathbb{E}[M_{n+1} \mid S_0, \ldots, S_n] = F(S_{n\wedge T_A}) = M_n. \tag{1.10}$$

A process that satisfies $\mathbb{E}[M_{n+1} \mid S_0, \ldots, S_n] = M_n$ is called a *martingale (with respect to the random walk)*. It is easy to see that $F(S_{n\wedge T_A})$ being a martingale is essentially equivalent to F being harmonic on A. It is easy to check that martingales satisfy $\mathbb{E}[M_n] = \mathbb{E}[M_0]$, and hence if $S_0 = x$, then

$$\sum_{y\in\overline{A}} \mathbb{P}\{S_{n\wedge T_A} = y\}\, F(y) = \mathbb{E}[M_n] = M_0 = F(x).$$

An easy argument shows that with probability one $T_A < \infty$. We can take limits and get

$$F(x) = \lim_{n\to\infty} \sum_{y\in\overline{A}} \mathbb{P}\{S_{n\wedge T_A} = y\}\, F(y) = \sum_{y\in\partial A} \mathbb{P}\{S_{T_A} = y\}\, F(y). \tag{1.11}$$

♢ There is no problem interchanging the limit and the sum because it is a finite sum. If A is infinite, one needs more assumptions to justify the exchange of the limit and the sum.

Let us consider this from the perspective of linear algebra. Suppose that A has N elements and ∂A has K elements. The solution of the Dirichlet problem assigns to each function on ∂A (a vector in $\mathbb{R}^K$) a function on A (a vector in $\mathbb{R}^N$). Hence the solution can be considered as a linear function from $\mathbb{R}^K$ to $\mathbb{R}^N$ (the reader should check that this is a linear transformation). Any linear transformation is given by an $N \times K$ matrix. Let us write the matrix for the solution as

$$\mathbf{H}_A = [H_A(x,y)]_{x\in A, y\in\partial A}.$$

Another way of stating (1.11) is to say that

$$H_A(x,y) = \mathbb{P}\left\{S_{T_A} = y \mid S_0 = x\right\}.$$

This matrix is often called the *Poisson kernel*. For a given set A, we can solve the Dirichlet problem for any boundary function in terms of the Poisson kernel.

$\diamondsuit$ Analysts who are not comfortable with probability[1] think of the Poisson kernel only as the matrix for the transformation which takes boundary data to values on the interior. Probabilists also have the interpretation of $H_A(x, y)$ as the probability that the random walk starting at x exits A at y.

What happens in Theorem 1.5 if we allow A to be an infinite set? In this case it is not always true that the solution is unique. Let us consider the one-dimensional example with $A = \{1, 2, 3, \ldots\}$ and $\partial A = \{0\}$. Then for every $c \in \mathbb{R}$, the function $F(x) = cx$ is harmonic in A with boundary value 0 at the origin. Where does our proof break down? This depends on which proof we consider (they all break down!), but let us consider the martingale version. Suppose F is harmonic on A with $F(0) = 0$ and suppose S_n is a simple random walk starting at positive integer x. As before, we let $T = \min\{n \geq 0 : S_n = 0\}$ and $M_n = F(S_{n\wedge T})$. The same argument shows that M_n is a martingale and

$$F(x) = \mathbb{E}[M_n] = \sum_{y=0}^{\infty} F(y)\,\mathbb{P}\{S_{n\wedge T} = y\}.$$

We have shown in a previous section that with probability one $T < \infty$. This implies that $\mathbb{P}\{S_{n\wedge T} = 0\}$ tends to 1, i.e.,

$$\lim_{n\to\infty} \sum_{y>0} \mathbb{P}\{S_{n\wedge T} = y\} = 0.$$

However, if F is unbounded, we cannot conclude from this that

$$\lim_{n\to\infty} \sum_{y>0} F(y)\,\mathbb{P}\{S_{n\wedge T} = y\} = 0.$$

However, we do see from this that there is only one *bounded* function that is harmonic on A with a given boundary value at 0. We state the theorem leaving the details as Exercise 1.7.

---

[1] The politically correct term is stochastically challenged.

Theorem 1.6. *Suppose A is a proper subset of $\mathbb{Z}^d$ such that for all $x \in \mathbb{Z}^d$,*

$$\lim_{n\to\infty} \mathbb{P}\{T_A > n \mid S_0 = x\} = 0.$$

Suppose $F : \partial A \to \mathbb{R}$ is a bounded function. Then there is a unique bounded extension of F to $\overline{A}$ that satisfies (1.9)*. It is given by*

$$F_0(x) = \mathbb{E}[F(S_{T_A}) \mid S_0 = x] = \sum_{y\in\partial A} \mathbb{P}\{S_{T_A} = y \mid S_0 = x\}\, F(y).$$

1.3. Heat equation

We will now introduce a mathematical model for heat flow. Let A be a finite subset of $\mathbb{Z}^d$ with boundary ∂A. We set the temperature at the boundary to be zero at all times and, as an initial condition, set the temperature at $x \in A$ to be $p_n(x)$. At each integer time unit n, the heat at x at time n is spread evenly among its $2d$ nearest neighbors. If one of those neighbors is a boundary point, then the heat that goes to that site is lost forever. A more probabilistic view of this is given by imagining that the temperature in A is controlled by a very large number of "heat particles". These particles perform random walks on A until they leave A at which time they are killed. The temperature at x at time n, $p_n(x)$ is given by the density of particles at x. Either interpretation gives a difference equation for the temperature $p_n(x)$. For $x \in A$, the temperature at x is given by the amount of heat going in from neighboring sites,

$$p_{n+1}(x) = \frac{1}{2d} \sum_{|y-x|=1} p_n(y).$$

If we introduce the notation $\partial_n p_n(x) = p_{n+1}(x) - p_n(x)$, we get the *heat equation*

$$\partial_n p_n(x) = \mathcal{L} p_n(x), \quad x \in A, \tag{1.12}$$

where $\mathcal{L}$ denotes the discrete Laplacian as before. The initial temperature is given as the initial condition

$$p_0(x) = f(x), \quad x \in A. \tag{1.13}$$

We rewrite the boundary condition as

$$p_n(x) = 0, \quad x \in \partial A. \tag{1.14}$$

If $x \in A$ and the initial condition is $f(x) = 1$ and $f(z) = 0$ for $z \neq x$, then

$$p_n(y) = \mathbb{P}\{S_{n \wedge T_A} = y \mid S_0 = x\}.$$

♢ The heat equation is a deterministic (i.e., without randomness) model for heat flow. It can be studied without probability. However, probability adds a layer of richness in terms of movements of individual random particles. This extra view is often useful for understanding the equation.

Given any initial condition f, it is easy to see that there is a unique function p_n satisfying (1.12)–(1.14). Indeed, we just set: $p_n(y) = 0$ for all $n \geq 0$ if $y \in \partial A$; $p_0(x) = f(x)$ if $x \in A$; and for $n > 0$, we define $p_n(x), x \in A$ recursively by (1.12). This tells us that set of functions satisfying (1.12) and (1.14) is a vector space of dimension $\#(A)$. In fact, $\{p_n(x) : x \in A\}$ is the vector $\mathbf{Q}^n f$.

Once we have existence and uniqueness, the problem remains to find the function. For a bounded set A, this is a problem in linear algebra and essentially becomes the question of diagonalizing the matrix $\mathbf{Q}$.

♢ Recall from linear algebra that if $\mathbf{A}$ is a $k \times k$ symmetric matrix with real entries, then we can find k (not necessarily distinct) real eigenvalues

$$\lambda_k \leq \lambda_{k-1} \leq \cdots \leq \lambda_1,$$

and k orthogonal vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ that are eigenvectors,

$$\mathbf{A}\mathbf{v}_j = \lambda_j \, \mathbf{v}_j.$$

(If A is not symmetric, A might not have k linearly independent eigenvectors, some eigenvalues might not be real, and eigenvectors for different eigenvalues are not necessarily orthogonal.)

We will start by considering the case $d = 1$. Let us compute the function p_n for $A = \{1, \ldots, N-1\}$. We start by looking for functions satisfying (1.12) of the form

$$p_n(x) = \lambda^n \, \phi(x). \tag{1.15}$$

If p_n is of this form, then

$$\partial_n p_n(x) = \lambda^{n+1} \, \phi(x) - \lambda^n \, \phi(x) = (\lambda - 1) \, \lambda^n \, \phi(x).$$

This nice form leads us to try to find eigenvalues and eigenfunctions of $\mathbf{Q}$, i.e., to find λ, ϕ such that

$$\mathbf{Q}\phi(x) = \lambda\,\phi(x), \tag{1.16}$$

with $\phi \equiv 0$ on ∂A.

♢ The "algorithmic" way to find the eigenvalues and eigenvectors for a matrix Q is first to find the eigenvalues as the roots of the characteristic polynomial and then to find the corresponding eigenvector for each eigenvalue. Sometimes we can avoid this if we can make good guesses for the eigenvectors. This is what we will do here.

The sum rule for sine,

$$\sin((x \pm 1)\theta) = \sin(x\theta)\cos(\theta) \pm \cos(x\theta)\sin(\theta),$$

tells us that

$$\mathbf{Q}\{\sin(\theta x)\} = \lambda_\theta\,\{\sin(\theta x)\}, \qquad \lambda_\theta = \cos\theta,$$

where $\{\sin(\theta x)\}$ denotes the vector whose component associated to $x \in A$ is $\sin(\theta x)$. If we choose $\theta_j = \pi j/N$, then $\phi_j(x) = \sin(\pi j x/N)$, which satisfies the boundary condition $\phi_j(0) = \phi_j(N) = 0$. Since these are eigenvectors with different eigenvalues for a symmetric matrix $\mathbf{Q}$, we know that they are orthogonal, and hence linearly independent. Hence every function f on A can be written in a unique way as

$$f(x) = \sum_{j=1}^{N-1} c_j\,\sin\left(\frac{\pi j x}{N}\right). \tag{1.17}$$

This sum in terms of trigonometric functions is called a finite *Fourier series*. The solution to the heat equation with initial condition f is

$$p_n(y) = \sum_{j=1}^{N-1} c_j\,\left[\cos\left(\frac{j\pi}{N}\right)\right]^n\,\phi_j(y).$$

Orthogonality of eigenvectors tells us that

$$\sum_{x=1}^{N-1} \sin\left(\frac{\pi j x}{N}\right)\,\sin\left(\frac{\pi k x}{N}\right) = 0 \text{ if } j \neq k.$$

Also,

$$\sum_{x=1}^{N-1} \sin^2\left(\frac{\pi j x}{N}\right) = \frac{N}{2}. \tag{1.18}$$

♢ The Nth roots of unity, $\zeta_1, \ldots, \zeta_N$ are the N complex numbers ζ such that $\zeta^N = 1$. They are given by

$$\zeta_k = \cos\left(\frac{2k\pi}{N}\right) + i\,\sin\left(\frac{2k\pi}{N}\right), \quad j = 1, \ldots, N.$$

The roots of unity are spread evenly about the unit circle in $\mathbb{C}$; in particular,

$$\omega_1 + \omega_2 + \cdots + \omega_N = 0,$$

which implies that

$$\sum_{j=1}^{N} \cos\left(\frac{2k\pi}{N}\right) = \sum_{j=1}^{N} \sin\left(\frac{2k\pi}{N}\right) = 0.$$

The double angle formula for sine gives

$$\begin{aligned}
\sum_{j=1}^{N-1} \sin^2\left(\frac{jx\pi}{N}\right) &= \sum_{j=1}^{N} \sin^2\left(\frac{jx\pi}{N}\right) \\
&= \frac{1}{2}\sum_{j=0}^{N-1}\left[1 - \cos\left(\frac{2jx\pi}{N}\right)\right] \\
&= \frac{N}{2} - \frac{1}{2}\sum_{j=1}^{N} \cos\left(\frac{2jx\pi}{N}\right).
\end{aligned}$$

If x is an integer, the last sum is zero. This gives (1.18).

In particular, if we choose the solution with initial condition $f(x) = 1; f(z) = 0, z \neq x$ we can see that

$$\mathbb{P}\{S_{n\wedge T_A} = y \mid S_0 = x\} = \frac{2}{N}\sum_{j=1}^{N-1} \phi_j(x)\,\left[\cos\left(\frac{j\pi}{N}\right)\right]^n \phi_j(y).$$

It is interesting to see what happens as $n \to \infty$. For large n, the sum is very small but it is dominated by the $j = 1$ and $j = N - 1$

terms for which the eigenvalue has maximal absolute value. These two terms give

$$\frac{2}{N}\cos^n\left(\frac{\pi}{N}\right)\left[\sin\left(\frac{\pi x}{N}\right)\sin\left(\frac{\pi y}{N}\right) + (-1)^n\sin\left(\frac{x\pi(N-1)}{N}\right)\sin\left(\frac{y\pi(N-1)}{N}\right)\right].$$

One can check that

$$\sin\left(\frac{x\pi(N-1)}{N}\right) = (-1)^{x+1}\sin\left(\frac{\pi x}{N}\right),$$

and, hence, if $x, y \in \{1, \ldots, N-1\}$, as $n \to \infty$, then

$$\mathbb{P}\{S_{n\wedge T_A} = y \mid S_0 = x\} \sim \frac{2}{N}\cos^n\left(\frac{\pi}{N}\right)[1 + (-1)^{n+x+y}]\sin\left(\frac{\pi x}{N}\right)\sin\left(\frac{\pi y}{N}\right).$$

For large n, conditioned such that the walker has not left $\{1, \ldots, N-1\}$, the probability that the walker is at y is about $c\sin(\pi y/N)$ assuming that the "parity" is correct ($n + x + y$ is even). Other than the parity, there is no dependence on the starting point x for the limiting distribution. Note that the walker is more likely to be at points toward the "middle" of the interval.

The above example illustrates a technique for finding solutions of the form (1.15) called *separation of variables*. The same idea works for all d although it may not always be possible to give nice expressions for the eigenvalues and eigenvectors. For finite A this is essentially the same as computing powers of a matrix by diagonalization. We summarize here.

Theorem 1.7. *If A is a finite subset of $\mathbb{Z}^d$ with N elements, then we can find N linearly independent functions $\phi_1, \ldots, \phi_N$ that satisfy* (1.16) *with real eigenvalues $\lambda_1, \ldots, \lambda_N$. The solution to* (1.12)–(1.14) *is given by*

$$p_n(x) = \sum_{j=1}^{N} c_j\,\lambda_j^n\,\phi_j(x),$$

where c_j are chosen so that

$$f(x) = \sum_{j=1}^{N} c_j\,\phi_j(x).$$

In fact, the ϕ_j can be chosen to be orthonormal,

$$\langle \phi_j, \phi_k \rangle := \sum_{x \in A} \phi_j(x)\, \phi_k(x) = \delta(k-j).$$

♢ Here we have introduced the *delta function* notation, $\delta(z) = 1$ if $z = 0$ and $\delta(z) = 0$ if $z \neq 0$.

Since $p_n(x) \to 0$ as $n \to \infty$, we know that the eigenvalues have absolute value strictly less than one. We can order the eigenvalues

$$1 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N > -1.$$

We will write $p(x, y; A)$ to be the solution of the heat equation with initial condition equal to one at x and 0 otherwise. In other words,

$$p_n(x, y; A) = \mathbb{P}\{S_n = y, T_A > n \mid S_0 = x\}, \qquad x, y \in A.$$

Then if $\#(A) = N$,

$$p_n(x, y; A) = \sum_{j=1}^{N} c_j(x)\, \lambda_j^n\, \phi_j(y),$$

where $c_j(x)$ have been chosen so that

$$\sum_{j=1}^{N} c_j(x) \phi_j(y) = \delta(y - x).$$

In fact, this tells us that $c_j(x) = \phi_j(x)$. Hence,

$$p_n(x, y; A) = \sum_{j=1}^{N} \lambda_j^n\, \phi_j(x)\, \phi_j(y).$$

Note that the quantity on the right is symmetric in x, y. One can check that the symmetry also follows from the definition of $p_n(x, y; A)$.

The largest eigenvalue λ_1 is often denoted λ_A. We can give a "variational" definition of λ_A as follows. This is really just a theorem about the largest eigenvalue of symmetric matrices.

Theorem 1.8. *If A is a finite subset of $\mathbb{Z}^d$, then λ_A is given by*

$$\lambda_A = \sup_f \frac{\langle \mathbf{Q} f, f \rangle}{\langle f, f \rangle},$$

where the supremum is over all functions f on A, and $\langle \cdot, \cdot \rangle$ denotes the inner product

$$\langle f, g \rangle = \sum_{x \in A} f(x)\, g(x).$$

Proof. If ϕ is an eigenvector with eigenvalue λ_1, then $\mathbf{Q}\phi = \lambda_1 \phi$ and setting $f = \phi$ shows that the supremum is at least as large as λ_1. Conversely, there is an orthogonal basis of eigenfunctions $\phi_1, \ldots, \phi_N$ and we can write any f as

$$f = \sum_{j=1}^{N} c_j\, \phi_j.$$

Then

$$\begin{aligned} \langle \mathbf{Q}f, f \rangle &= \left\langle \mathbf{Q} \sum_{j=1}^{N} c_j\, \phi_j, \sum_{j=1}^{N} c_j\, \phi_j \right\rangle \\ &= \left\langle \sum_{j=1}^{N} c_j \mathbf{Q}\phi_j, \sum_{j=1}^{N} c_j\, \phi_j \right\rangle \\ &= \sum_{j=1} c_j^2\, \lambda_j\, \langle \phi_j, \phi_j \rangle \\ &\le \lambda_1 \sum_{j=1} c_j^2\, \langle \phi_j, \phi_j \rangle = \lambda_1\, \langle f, f \rangle. \end{aligned}$$

The reader should check that the computation above uses the orthogonality of the eigenfunctions and also the fact that $\langle \phi_j, \phi_j \rangle \ge 0$. □

Using this variational formulation, we can see that the eigenfunction for λ_1 can be chosen so that $\phi_1(x) \ge 0$ for each x (since if ϕ_1 took on both positive and negative values, we would have $\langle \mathbf{Q}|\phi_1|, |\phi_1| \rangle > \langle \phi_1, \phi_1 \rangle$). The eigenfunction is unique, i.e., $\lambda_2 < \lambda_1$, provided we put an additional condition on A. We say that a subset A on $\mathbb{Z}^d$ is *connected* if any two points in A are connected by a nearest neighbor path that stays entirely in A. Equivalently, A is connected if for each $x, y \in A$ there exists an n such that $p_n(x, y; A) > 0$. We leave it as Exercise 1.23 to show that this implies that $\lambda_1 > \lambda_2$.

Before stating the final theorem, we need to discuss some parity (even/odd) issues. If $x = (x_1, \ldots, x_d) \in \mathbb{Z}^d$ we let $\mathrm{par}(x) =$

$(-1)^{x_1+\cdots+x_d}$. We call x *even* if $\mathrm{par}(x) = 1$ and otherwise x is odd. If n is a nonnegative integer, then

$$p_n(x, y; A) = 0 \text{ if } (-1)^n \mathrm{par}(x + y) = -1.$$

If $\mathbf{Q}\phi = \lambda\phi$, then $\mathbf{Q}[\mathrm{par}\phi] = -\lambda\mathrm{par}\phi$.

Theorem 1.9. *Suppose A is a finite connected subset of $\mathbb{Z}^d$ with at least two points. Then $\lambda_1 > \lambda_2$, $\lambda_N = -\lambda_1 < \lambda_{N-1}$. The eigenfunction ϕ_1 can be chosen so that $\phi_1(x) > 0$ for all $x \in A$,*

$$\lim_{n\to\infty} \lambda_1^{-n}\, p_n(x, y; A) = [1 + (-1)^n\, \mathrm{par}(x + y)]\, \phi_1(x)\, \phi_1(y).$$

Example 1.10. One set in $\mathbb{Z}^d$ for which we can compute the eigenfunctions and eigenvalues exactly is a d-dimensional rectangle,

$$A = \{(x_1, \ldots, x_d) \in \mathbb{Z}^d : 1 \le x_j \le N_j - 1\}.$$

The eigenfunctions are indexed by $\bar{k} = (k_1, \ldots, k_d) \in A$,

$$\phi_{\bar{k}}(x_1, \ldots, x_d) = \sin\left(\frac{k_1\pi x_1}{N_1}\right) \sin\left(\frac{k_2\pi x_2}{N_2}\right) \cdots \sin\left(\frac{k_d\pi x_d}{N_d}\right),$$

with eigenvalue

$$\lambda_{\bar{k}} = \frac{1}{d}\left[\cos\left(\frac{k_1\pi}{N_1}\right) + \cdots + \cos\left(\frac{k_d\pi}{N_d}\right)\right].$$

1.4. Expected time to escape

1.4.1. One dimension. Let S_n denote a one-dimensional random walk starting at $x \in \{0, \ldots, N\}$ and let T be the first time that the walker reaches $\{0, N\}$. Here we study the expected time to reach 0 or N,

$$e(x) = \mathbb{E}[T \mid S_0 = x].$$

Clearly, $e(0) = e(N) = 0$. Now suppose $x \in \{1, \ldots, N-1\}$. Then the walker takes one step which goes to either $x - 1$ or $x + 1$. Using this we get the relation

$$e(x) = 1 + \frac{1}{2}\,[e(x + 1) + e(x - 1)]\,.$$

Hence, e satisfies

$$e(0) = e(N) = 0, \quad \mathcal{L}e(x) = -1, \quad x = 1, \ldots, N - 1. \tag{1.19}$$