Chapter 1

Introduction

1.1 Basic definitions

We will define the random walks that we consider in this book. We will focus our attention on random walks in \( \mathbb{Z}^d \) that have bounded symmetric increment distributions (although we will occasionally discuss results for wider classes of walks). We will also impose an irreducibility criterion to guarantee that all points in the lattice \( \mathbb{Z}^d \) can be reached.

We start by listing some basic notation. We will use \( x, y, z \) to denote points in the integer lattice \( \mathbb{Z}^d = \{(x_1, \ldots, x_d) : x_i \in \mathbb{Z}\} \). We will use superscripts to denote components, and we will use subscripts to enumerate elements. For example, we can take a sequence of points \( x_1, x_2, \ldots \in \mathbb{Z}^d \) and we can write each one in component form \( x_j = (x_{j1}, \ldots, x_{jd}) \). We will write \( e_i = (0, \ldots, 0, 1, \ldots, 0, 0) \) for the standard basis of unit vectors in \( \mathbb{Z}^d \). The prototypical example is discrete time simple random walk starting at \( x \in \mathbb{Z}^d \). This process can be considered either as a sequence of independent, identically distributed random variables

\[
S_n = x + X_1 + \cdots + X_n
\]

where \( P\{X_j = e_k\} = P\{X_j = -e_k\} = 1/(2d), k = 1, \ldots, d \), or it can be considered as a Markov chain with state space \( \mathbb{Z}^d \) and transition probabilities

\[
P\{S_{n+1} = z \mid S_n = y\} = \frac{1}{2d}, \quad z - y \in \{\pm e_1, \ldots, \pm e_d\}.
\]

We call \( V = \{x_1, \ldots, x_k\} \subset \mathbb{Z}^d \setminus \{0\} \) a (finite) generating set if each element of \( V \) is a \( \{e_1, \ldots, e_d\} \). A \( \{e_1, \ldots, e_d\} \) symmetric, \( k \)-dimensional \( \mathbb{Z}^d \) random walk is given by specifying a \( V = \{x_1, \ldots, x_k\} \subset \mathbb{Z}^d \) and a function \( \varepsilon : V \to \{0, 1\} \) with \( \varepsilon(x_1) + \cdots + \varepsilon(x_k) \leq 1 \). Associated to this is the symmetric \( \{e_1, \ldots, e_d\} \) \( \mathbb{Z}^d \) distribution on \( \mathbb{Z}^d \)

\[
p(x_k) = p(-x_k) = \frac{1}{2} \varepsilon(x_k), \quad p(0) = 1 - \sum_{k \in V} \varepsilon(x).
\]

We will let \( \mathcal{P}_d \) denote the set of such distributions \( p \) on \( \mathbb{Z}^d \) and \( \mathcal{P} = \cup_{d \geq 1} \mathcal{P}_d \). Given \( p \) the corresponding random walk \( S_0 \) can be considered as the time-homogeneous Markov chain with state space \( \mathbb{Z}^d \) with transition probabilities

\[
p(y, z) := P\{S_{n+1} = z \mid S_n = y\} = p(z - y).
\]

We can also write

\[
S_n = S_0 + X_1 + \cdots + X_n
\]

where \( X_1, X_2, \ldots \) are independent random variables, independent of \( S_0 \), with distribution \( p \). (Most of the time we will choose \( S_0 \) to have a trivial distribution.) We will use the phrase \( \mathcal{P} \)-walk or \( \mathcal{P} \)-walk for such a random walk. We will use the term simple random walk for the particular \( p \) with \( p(e_j) = p(-e_j) = 1/(2d) \). Given \( p \in \mathcal{P} \), we write \( p_n \) for the \( n \)-step distribution

\[
p_n(x, y) = P\{S_n = y \mid S_0 = x\}
\]

and \( p_n(0) = p_n(0, x) \). Note that \( p_n(0) \) is the distribution of \( S_1 + \cdots + S_n \) where \( S_0, S_1, \ldots, S_n \) are independent with increment distribution \( p \).

Heuristic note. In many ways the main focus of this book is simple random walk, and a first-time reader might find it useful to consider this example throughout. We have chosen to generalize this slightly, because it does not complicate the arguments much and allows the results to be extended to other examples. One particular example is simple random walk on other regular lattices such as the planar triangular lattice. In Section 1.3, we show how walks on other \( d \)-dimensional lattices are isomorphic to \( p \)-walks on \( \mathbb{Z}^d \).

If \( S_0 \) is a \( \mathcal{P} \)-walk with \( S_0 = 0 \), then \( P\{S_n = 0\} > 0 \) for every even integer \( n \); this follows from the easy estimate \( P\{S_n = 0\} \geq p(z)^n \) for every \( z \in \mathbb{Z}^d \). We will call the
1.1. Basic Definitions

A walk is bipartite if \( p_n(0,0) = 0 \) for every odd \( n \), and we will call it aperiodic otherwise. In the latter case, \( p_n(0,0) > 0 \) for all \( n \) sufficiently large (in fact, for all \( n \geq k \) where \( k \) is the first odd integer with \( p_k(0,0) > 0 \)). Simple random walk is an example of a bipartite walk since \( S_0^1, S_1^2, \ldots \) are always odd for odd \( n \) and even for even \( n \). If \( p \) is bipartite, then we can partition \( \mathbb{Z}^d = \mathbb{Z}_0^d \cup \mathbb{Z}_0^d \) where \( \mathbb{Z}_0^d \) denotes the points that can be reached from the origin in an even number of steps and \( \mathbb{Z}_0^d \) denotes the set of points that can be reached in an odd number of steps. In algebraic language, \( \mathbb{Z}_0^d \) is an additive subgroup of \( \mathbb{Z}^d \) of index 2 and \( \mathbb{Z}_0^d \) is the nontrivial coset. Note that if \( x \in \mathbb{Z}_0^d \), then \( \mathbb{Z}_0^d = x + \mathbb{Z}_0^d \).

Heuristic note. It would suffice and would perhaps be more convenient to restrict our attention to aperiodic walks. Results about bipartite walks can be easily deduced from them. However, since our main example, simple random walk, is bipartite, we have chosen to allow such \( p \).

If \( p \in \mathcal{P} \) and \( j_1, \ldots, j_d \) are nonnegative integers, the \((j_1, \ldots, j_d)\) moment is given by

\[
\mathbb{E}(X_1^{j_1} \cdots X_d^{j_d}) = \sum_{x \in \mathbb{Z}^d} (x_1^{j_1} \cdots x_d^{j_d}) p(x).
\]

We let \( \sigma_d^2 = \sum (x)^2 p(x) \) and we let \( \Gamma \) denote the covariance matrix

\[
\Gamma = \left[ \mathbb{E}(X_i^j X_k^l) \right]_{i,j,k,l \leq d}.
\]

The covariance matrix is symmetric and positive definite. We can find a symmetric positive definite \( \Lambda \) such that \( \Gamma = \Lambda^T \Lambda \) (see Section 8.3). The matrix \( \Gamma \) is invertible. For future use, we define norms \( (\cdot)^* \) by

\[
(\cdot)^* = | \cdot | \Gamma^{-1} | \cdot |, \quad J(x) = d^{-1/2} (\cdot)^*(x).
\]

For simple random walk in \( \mathbb{Z}^d \),

\[
\Gamma = \sigma_d^2 I, \quad J(x) = d^{-1/2} |x|, \quad J(x) = |x|.
\]

We will use \( B_n \) to denote the discrete ball of radius \( n \),

\[
B_n = \{ x \in \mathbb{Z}^d : |x| < n \}
\]

and \( C_n \) for discrete balls using the norm \( J \),

\[
C_n = \{ x \in \mathbb{Z}^d : J(x) < n \} = \{ x \in \mathbb{Z}^d : J^*(x) < d^{-1/2} n \}.
\]

We use \( J \) so that for simple random walk, \( C_n = B_n \). We will write \( R = R_p \) as \( \max \{ |x| : p(x) > 0 \} \) and we will call \( R \) the range of \( p \). The following is very easy, but it is important enough to state as a proposition.

**Proposition 1.1.1** Suppose \( p \in \mathcal{P} \).

(a) There exists an \( \epsilon > 0 \) such that for every unit vector \( u \in \mathbb{R}^d \),

\[
\mathbb{E}(|X_t| u^T u) \geq \epsilon.
\]

(b) If \( j_1, \ldots, j_d \) are nonnegative integers with \( j_1 + \cdots + j_d \) odd, then

\[
\mathbb{E}(X_1^{j_1} \cdots X_d^{j_d}) = 0.
\]

We note for later use that we can construct a random walk with increment distribution \( p \) from a collection of independent one-dimensional simple random walks and an independent multinomial process. To be more precise, let \( V = (x_1, \ldots, x_d) \in \mathcal{G} \) and \( \kappa \) be given. Suppose on the same probability space we have defined \( d \)-independent one-dimensional simple random walks \( S_{n,1}, S_{n,2}, \ldots, S_{n,d} \) and an independent multinomial process \( L_n = (L_{n,1}, \ldots, L_{n,d}) \) with probabilities \( \kappa(x_1), \ldots, \kappa(x_d) \). In other words,

\[
L_n = \sum_{j=1}^d Y_j,
\]

where \( Y_1, Y_2, \ldots \) are independent \( \mathbb{Z} \)-valued random variables with

\[
\mathbb{P}(Y_j = (1,0,\ldots,0)) = \kappa(x_1), \ldots, \mathbb{P}(Y_n = (0,0,\ldots,1)) = \kappa(x_d)
\]

and \( \mathbb{P}(Y_j = (0,0,\ldots,0)) = 1 - [\kappa(x_1) + \cdots + \kappa(x_d)] \). It is easy to verify that the process

\[
S_n := x_1 S_{n,1} + x_2 S_{n,2} + \cdots + x_d S_{n,d}
\]

has the distribution of the random walk with increment distribution \( p \). Essentially what we have done is to split the decision as to where to jump at time \( n \) into two decisions: first, to choose \( x_1, \ldots, x_d \) and then to decide whether to move positively or negatively.

1.2 Continuous-time random walk

It is often more convenient to consider random walks in \( \mathbb{Z}^d \) indexed by positive real times. Given \( V, \epsilon, p \) as in the previous section, the continuous-time random walk with increment distribution \( p \) is the continuous-time Markov chain \( \tilde{S}_t \) with rates \( p \). In other words, for each \( x, y \in \mathbb{Z}^d \),

\[
\mathbb{P} \left( \tilde{S}_{t+\Delta t} = y \mid \tilde{S}_t = x \right) = p(y - x) \Delta t + o(\Delta t), \quad y \neq x.
\]

Let \( \tilde{p}(x,y) = p(y-x) \) and \( \mathbb{P} \left( \tilde{S}_t = x \mid \tilde{S}_0 = x \right) = 1 \). Then we can write the equation above as

\[
\frac{d}{dt} \tilde{p}(x) = \sum_{y \neq x} p(y) \tilde{p}(x - y) - \tilde{p}(x).
\]

There is a very close relationship between the discrete time and continuous time random walks with the same increment distribution. We state this as a proposition which we leave to the reader to verify.
1.2. CONTINUOUS-TIME RANDOM WALK

Proposition 1.2.1 Suppose $S_n$ is a (discrete-time) random walk with increment distribution $p$ and $N_i$ is an independent Poisson process with parameter $1$. Then $\hat{S}_t := S_{N_t}$ has the distribution of a continuous-time random walk with increment distribution $p$.

There are various technical reasons why continuous-time random walks are sometimes easier to handle than discrete-time walks. One is that periodicity does not arise. If $p \in P_d$, then $p(x) > 0$ for every $t > 0$ and $x \in Z^d$. Another advantage can be found in the following proposition which gives an analogous, but stronger, version of (1.2). We leave the proof to the reader.

Proposition 1.2.2 Suppose $p \in P_d$ with generating set $V = \{v_1, \ldots, v_d\}$ and suppose $\hat{S}_t, \ldots, \hat{S}_t$ are independent one-dimensional continuous-time random walks with increment distribution $q_1, \ldots, q_d$ where $q_j(\pm 1) = p(x_j)$. Then

$$\hat{S}_t := x_1 \hat{S}_{t,1} + x_2 \hat{S}_{t,2} + \cdots + x_d \hat{S}_{t,d}$$

has the distribution of a continuous-time random walk with increment distribution $p$.

If $p$ is the increment distribution for simple random walk, we call $\hat{S}_t$ the continuous-time simple random walk in $Z^d$. From the previous proposition, we see that the coordinates of the continuous-time simple random walk are independent of each other as long as the itineraries have a positive probability of occurring.

In fact, we get the following. Suppose $\hat{S}_1, \ldots, \hat{S}_d$ are independent one-dimensional continuous-time simple random walks. Then,

$$\hat{S}_t := (\hat{S}_{t,1}, \ldots, \hat{S}_{t,d})$$

is a continuous time simple random walk in $Z^d$. In particular, if $\hat{S}_0 = 0$, then

$$P(\hat{S}_t = (y_1, \ldots, y_d)) = P(\hat{S}_{t,1} = y_1) \cdots P(\hat{S}_{t,d} = y_d).$$

Remark. To verify that a discrete-time process $S_n$ is a random walk with distribution $p \in P_d$ starting at the origin, it suffices to show for all positive integers $j_1 < j_2 < \cdots < j_k$ and $x_1, \ldots, x_k \in Z^d$

$$P(S_{j_k} = x_{j_k}, \ldots, S_{j_{k-1}} = x_{j_{k-1}}) = p_{j_k}(x_{j_k}) \cdots p_{j_2}(x_{j_2}) \cdots p_{j_1}(x_{j_1}).$$

To verify that a continuous-time process $\hat{S}_t$ is a continuous-time random walk with distribution $p$ starting at the origin, it suffices to show that the paths are right-continuous with probability one, and for all real $t_1 < t_2 < \cdots < t_k$ and $x_1, \ldots, x_k \in Z^d$

$$P(\hat{S}_{t_1} = x_1, \ldots, \hat{S}_{t_k} = x_k) = p_{t_1}(x_1) \cdots p_{t_{k-1}}(x_{k-1}) \cdots p_{t_1}(x_{t_1}).$$

1.3 Other lattices

A lattice $L$ is a discrete additive subgroup of $R^d$. The term discrete means that there is a real neighborhood of the origin whose intersection with $L$ is just the origin. While this book will focus on the lattice $Z^d$, we will show in this section that this also implies results for symmetric, bounded random walks on other lattices. We start by giving a proposition that classifies all lattices.

Proposition 1.3.1 If $L \subseteq \mathbb{R}^d$ is a lattice, then there exists an integer $k \leq d$ and elements $x_1, \ldots, x_k \in L$ that are linearly independent as vectors in $\mathbb{R}^d$ such that

$$L = \{x_1 j_1 + \cdots + x_k j_k : j_1, \ldots, j_k \in \mathbb{Z}\}.$$

In this case we call $L$ a $k$-dimensional lattice.

Proof. Suppose first that $L$ is contained in a one-dimensional subspace of $\mathbb{R}^d$. Choose $x_1 \in L \setminus \{0\}$ with minimal distance from the origin. Clearly $(j x_1 : j \in \mathbb{Z}) \subseteq L$. Also, if $x \in L$, then $j x_1 \leq x < (j+1) x_1$ for some $j \in \mathbb{Z}$, but if $x \neq j x_1$, then $x - j x_1$ would be closer to the origin than $x_1$. Hence $L = \{j x_1 : j \in \mathbb{Z}\}$.

More generally, suppose we have chosen linearly independent $x_1, \ldots, x_k$ such that the following holds: if $L_k$ is the subgroup generated by $x_1, \ldots, x_k$ and $V$ is the real subspace of $\mathbb{R}^d$ generated by the vectors $x_1, \ldots, x_k$, then $L \cap V = L_k$. If $L = L_2$, we stop. Otherwise, let $v_1 \in L \setminus L_k$ and let

$$U = \{tv_1 : v \in V, tv_1 + y_0 \in L \text{ for some } y_0 \in V\} = \{tv_1 : v \in V, tv_1 + x_1 + \cdots + x_k j \in L \text{ for some } x_1, \ldots, x_k \in (0, 1]\}.$$

The second equality uses the fact that $L$ is a subgroup. Using the first description, we can see that $U$ is a subgroup of $\mathbb{R}^d$ although not necessarily contained in $L$. We claim that the second description shows that there is a neighborhood of the origin whose intersection with $U$ is exactly the origin. Indeed, there are only a finite number of lattice points of the form

$$tv_1 + x_1 + \cdots + x_k j$$

with $0 \leq t \leq 1$, and $0 \leq x_1, \ldots, x_k \leq 1$. Hence there is an $\epsilon > 0$ such that there are no such lattice points with $0 < ||v|| \leq \epsilon$. Since $U$ is a one-dimensional lattice, there is a $w \in U$ such that $U = \{tv : v \in L_k\}$. By definition, there exists a $y_1 \in V$ (not unique, but we choose one) such that $x_1 y_1 \colonequals w + y_1 \in L$. Let $L_{k+1}$ be as above using $y_1, \ldots, y_k$. Note that $V_{k+1}$ is also the real subspace generated by $x_1, \ldots, x_{k+1}$. We claim that $L \cap V_{k+1} = L_{k+1}$. Indeed, suppose that $z \in L \cap V_{k+1}$, and write $z = y_k + y_2$ where $y_2 \in V_{k+1}$. Then $y_2 \in U$, and hence $y = y_1$ for some integer $t$. Hence, we can write $z = t x_{k+1} y_2 + y_2 y_2 \in V_{k+1}$. But $z = y \in V_{k+1}$ since $y_2 \in L_{k+1}$, hence $z \in L_{k+1}$.}

Hence if $k \leq d$ and $L$ is a $k$-dimensional lattice in $\mathbb{R}^d$, we can find a linear transformation $A : \mathbb{R}^d \to \mathbb{R}^d$ that is an isomorphism of $L$ onto $\mathbb{Z}^d$. Indeed, we define $A$ by $A(x_j) = e_j$. 


where $x_1, \ldots, x_d$ is a basis for $\mathbb{L}$ as in the proposition. If $S_n$ is a bounded, symmetric, irreducible random walk taking values in $\mathbb{L}$, then $S_n = AS_0$ is a random walk with increment distribution $p \in P_\mathbb{L}$. Hence, results about walks on $\mathbb{Z}^d$ immediately translate to results about walks on $\mathbb{L}$. If $\mathbb{L}$ is a $k$-dimensional lattice in $\mathbb{R}^k$ and $A$ is the corresponding transformation, we call $\det A$ the density of the lattice. The term comes from the fact that as $r \to \infty$, the cardinality of the intersection of the lattice and ball of radius $r$ in $\mathbb{R}^k$ is asymptotically equal to $\det A r^k$ times the volume of the unit ball in $\mathbb{R}^k$. In particular, if $j_1, \ldots, j_k$ are positive integers, then $(j_1 \mathbb{Z}) \times \cdots \times (j_k \mathbb{Z})$ has density $|j_1 \cdots j_k|^k$.

Examples.

- The triangular lattice, considered as a subset of $\mathbb{C} = \mathbb{R}^2$, is the lattice generated by $1$ and $e^{i \sqrt{3} \pi / 3}$:
  \[ \mathbb{L}_T = \{ k_1 + k_2 e^{i \sqrt{3} \pi / 3} : k_1, k_2 \in \mathbb{Z} \} . \]
  Note that $e^{i \sqrt{3} \pi / 3} = e^{i \pi / 3} - 1 \in \mathbb{L}_T$. One also considers this lattice as a graph by asserting that all points of Euclidean norm one are adjacent. In this case, the origin has six nearest neighbors, the six sixth roots of unity. Simple random walk on the triangular lattice is the process that chooses among these six nearest neighbors equally likely. Note that this is a symmetric walk with bounded increments. The matrix
  \[ A = \begin{bmatrix} 1 & -1/\sqrt{3} \\ 0 & 2/\sqrt{3} \end{bmatrix} \]
  maps $\mathbb{L}_T$ to $\mathbb{Z}^2$ sending $\{1, e^{i \pi / 3}, e^{i \pi / 3} \}$ to $\{e_1, e_2, e_1 - e_1\}$. The transformed random walk gives probability $1/6$ to the following vectors: $\pm e_1, \pm e_2, (e_1 - e_1)$. Note that our transformed walk has lost some of the symmetry of the original walk.

- The hexagonal or honeycomb lattice is the dual lattice to the triangular lattice. It can be constructed in a number of ways. One way is to start with the points in the triangular lattice $\mathbb{L}_T$, and then the lattice generated by $1$ and $z_0$. This is considered a graph by stating that two vertices are adjacent if they are at distance $(\sqrt{3}/2)$; each vertex has three nearest neighbors. The simple random walk on the hexagonal lattice is the walk that chooses among these neighbors with equal probability. This is not a random walk in our strict sense because the increment distribution depends on whether one an “even” point (i.e., in $\mathbb{L}_T$) or an odd point (in $z_0 + \mathbb{L}_T$). However, the two-step distribution of this walk is the same as the walk on the triangular lattice with step distribution $p(\pm 1) = p(\pm e^{i \pi / 3}) = p(\pm e^{i \sqrt{3} \pi / 3}) = 1/9, p(0) = 1/3$.

When studying random walks on other lattices $\mathbb{L}$, we can map the walk to another walk on $\mathbb{Z}^d$. However, since this might lose useful symmetries of the walk, it is sometimes better to work on the original lattice. We let $\mathcal{P}(\mathbb{L})$ be the collection of symmetric, bounded random walk distributions. If $p \in \mathcal{P}(\mathbb{L})$ and $A : \mathbb{L} \to \mathbb{Z}^d$ is an isomorphism, we will write $Ap \in \mathcal{P}_d$ for the corresponding increment distribution,
  \[ Ap(y) = p(A^{-1} y) . \]

1.4 Generator

If $f : \mathbb{Z}^d \to \mathbb{R}$ is a function and $x \in \mathbb{Z}^d$, we define the first and second difference operators in the direction $x$ by
  \[ \nabla_x f(y) = f(y + x) - f(y) , \]
  \[ \nabla_x^2 f(y) = \frac{1}{2} f(y + x) + \frac{1}{2} f(y - x) - f(y) . \]
  Note that $\nabla_x^2 = \nabla \nabla_x$. We will sometimes write just $\nabla, \nabla_x^2$ or $\nabla_{e_i}, \nabla_{e_i}^2$. If $p \in \mathcal{P}_d$ with generator set $V$, then the generator $L = L_p$ is defined by
  \[ Lf(y) = \sum_{x \in \mathbb{Z}^d} p(x) \nabla_x f(y) = \sum_{x \in V} \nabla_x f(y) = -f(y) + \sum_{x \in V} p(x) (f(y + x) - f(y)) . \]
  In the case of simple random walk, the generator is often called the discrete Laplacian and we will represent it by $\Delta_\mathbb{L}$.
  \[ \Delta_\mathbb{L} f(y) = \frac{1}{d} \sum_{j=1}^d \nabla_j^2 f(y) . \]

The generator of a random walk is very closely related to the walk. We will write $E, \mathbb{P}$ to denote expectations and probabilities for random walk (both discrete and continuous time) assuming that $S_0 = x$. Then, it is easy to check that
  \[ Lf(y) = \mathbb{E}_x [ f(S_1) - f(y) ] = \frac{d}{dt} \mathbb{E}_{\mathbb{P}_x} [ f(S_t) ] \bigg|_{t=0} . \]
(In the continuous-time case, some restrictions on the growth of $f$ at infinity are needed.) Also, the transition probabilities $p_n(x), \mathbb{P}_x(x)$ satisfy the following “heat equations”:
  \[ p_{n+1}(x) - p_n(x) = L p_n(x), \quad \frac{d}{dt} \mathbb{P}_x(t) = L \mathbb{P}_x(t) . \]
The derivation of these equations uses the symmetry of $p$. For example to derive the first, we write
  \[ p_{n+1}(x) = \sum_{y \in \mathbb{Z}^d} \mathbb{P}[ S_{n+1} = y | S_n = x = y ] = \sum_{y \in \mathbb{Z}^d} p(y) p_n(x - y) = \sum_{y \in \mathbb{Z}^d} p(-y) p_n(x - y) = p_n(x) + L p_n(x) . \]
1.5 Filtrations and strong Markov property

The basic property of a random walk is that increments are independent and identically distributed. It is useful to set up a framework that allows more "information" at a particular time than just the value of the random walk. This will not affect the distribution of the random walk provided that this extra information is independent of the future increments of the walk.

A (discrete-time) filtration \( \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \) is an increasing sequence of \( \sigma \)-algebras. If \( p \in \mathcal{P} \), then we say that \( S_n \) is a random walk with increment distribution \( p \) with respect to \( \{ \mathcal{F}_n \} \) if:

- for each \( n \), \( S_n \) is \( \mathcal{F}_n \)-measurable;
- for each \( n > 0 \), \( S_n - S_{n-1} \) is independent of \( \mathcal{F}_{n-1} \) with distribution \( p \).

Similarly, we define a (right continuous, continuous-time) filtration to be an increasing collection of \( \sigma \)-algebras \( \mathcal{F} \) satisfying \( \mathcal{F}_t = \bigcap_{s \geq t} \mathcal{F}_s \). If \( p \in \mathcal{P} \), then we say that \( \tilde{S} \) is a continuous-time random walk with increment distribution \( p \) with respect to \( \{ \mathcal{F}_t \} \) if:

- for each \( t \), \( \tilde{S}_t \) is \( \mathcal{F}_t \)-measurable;
- for each \( s < t \), \( \tilde{S}_s - \tilde{S}_s \) is independent of \( \mathcal{F}_s \) and \( \mathbb{P}(\tilde{S}_s - \tilde{S}_s = x) = \beta(x) \).

We let \( \mathcal{F}_0 \) denote the \( \sigma \)-algebra generated by the union of the \( \mathcal{F}_t \) for \( t > 0 \).

If \( S_n \) is a random walk with respect to \( \mathcal{F}_n \) and \( T \) is a random variable independent of \( \mathcal{F}_n \), then we can add information about \( T \) to the filtration and still retain the properties of the random walk. We will describe one example of this in detail here; later on, we will do similar adding of information without being explicit. Suppose \( T \) has an exponential distribution with parameter \( \lambda \), i.e., \( \mathbb{P}(T > \lambda) = e^{-\lambda} \). Let \( \mathcal{F}_t \) denote the \( \sigma \)-algebra generated by \( T \) and the events \( \{ T \leq t \} \) for \( t \leq n \). Then \( \{ \mathcal{F}_s \} \) is a filtration, and \( S_n \) is a random walk with respect to \( \mathcal{F}_s \). Also, given \( \mathcal{F}_s \), then on the event \( \{ T > n \} \), the random variable \( T - n \) has an exponential distribution with parameter \( \lambda \). We can do something similar to the continuous-time walk \( \tilde{S}_t \).

We will discuss stopping times and the strong Markov property. We will only do the slightly more difficult continuous-time case to avoid the discrete-time analogue to the reader. If \( \{ \mathcal{F}_t \} \) is a filtration, then a stopping time with respect to \( \{ \mathcal{F}_t \} \) is a \( [0, \infty) \)-valued random variable \( \tau \) such that for each \( t \), \( \{ \tau \leq t \} \in \mathcal{F}_t \). Associated to the stopping time \( \tau \) is an \( \mathcal{F}_\tau \)-algebra consisting of all events \( A \) such that for each \( t \), \( A \cap \{ \tau \leq t \} \in \mathcal{F}_t \). (It is straightforward to check that the set of such \( A \) is a \( \sigma \)-algebra.)

Theorem 1.5.1 (Strong Markov Property) Suppose \( \tilde{S}_t \) is a continuous-time random walk with increment distribution \( p \) with respect to the filtration \( \{ \mathcal{F}_t \} \). Suppose \( \tau \) is a stopping time with respect to the process. Then on the event \( \{ \tau < \infty \} \) the process

\[ \tilde{S}_{t+\tau} = \tilde{S}_\tau, \]

is a continuous-time random walk with increment distribution \( p \) independent of \( \mathcal{F}_\tau \).
Part (b) is done similarly, by letting $\tau$ be the smallest $j$ with $\{\hat{S}_{j+1} - \hat{S}_j \geq b\}$ and writing

$$
\bigcup_{j=1}^{\tau^*} \{ \tau; (\hat{S}_{j+1} - \hat{S}_j, \hat{S}_j - \hat{S}_{j+1} \geq 0) \} \subset \{ |\hat{S}_j| \geq b \}.
$$

**Remark.** The only fact about the distribution $X_1$ that is used in the proof is that it is symmetric about the origin.

### 1.6 Other walks

Although we will focus primarily on $p \in \mathcal{P}$, there are times where we will want to look at more general walks. There are two classes of distributions we will be considering. We let $\mathcal{P}^+_\ell$ denote the set of $p$ which generate aperiodic, irreducible walks supported on $\mathbb{Z}^d$, i.e., the set of $p$ such that for all $x, y \in \mathbb{Z}^d$ there exists an $N$ such that $p_n(x, y) > 0$ for $n \geq N$. We let $\mathcal{P}^+_j \subset \mathcal{P}^+_\ell$ be the collection of $p \in \mathcal{P}^+_\ell$ with mean zero and finite variance. We write $\mathcal{P}^+, \mathcal{P}^j$ for the unions over $d \geq 1$.

Note that under our definition of $\mathcal{P}$, $\mathcal{P}$ is a subset of $\mathcal{P}^j$ since $\mathcal{P}$ contains bipartite walks. However, if $p \in \mathcal{P}$ is aperiodic, then $p \in \mathcal{P}^+$.

### 1.7 A word about constants

Throughout this book we will denote a positive constant that can depend on the dimension $d$ and the increment distribution $p$ but does not depend on any other constants. We write

$$
 f(n, x) = \gamma(n, x) + O(\delta(n)),
$$

to mean that there exists a constant $\gamma$ such that for all $n$,

$$
 |f(n, x) - \gamma(n, x)| \leq c \delta(n),
$$

Similarly, we write

$$
 f(n, x) = \gamma(n, x) + o(\delta(n))
$$

if for every $\varepsilon > 0$ there is an $N$ such that

$$
 |f(n, x) - \gamma(n, x)| \leq \varepsilon \delta(n), \quad n \geq N,
$$

Note that implicit in the definition is the fact that $\gamma, N$ can be chosen uniformly for all $x$. If $f, g$ are positive functions, we will write

$$
 f(n, x) \asymp g(n, x), \quad n \to \infty,
$$

if there exist a $\varepsilon$ (again, independent of $x$) such that for all $n, x$,

$$
 c^{-1} g(n, x) \leq f(n, x) \leq c g(n, x).
$$

### Exercises for Chapter 1

**Exercise 1.1** Show that there are exactly $2^d - 1$ additive subgroups of $\mathbb{Z}^d$ of index 2. Describe them and show that they all arise from some $p \in \mathcal{P}$.

**Exercise 1.2** Show that if $p \in \mathcal{P}_d$, $n$ is a positive integer, and $x \in \mathbb{Z}^d$, then $p_{2n}(0) \geq p_{2n}(x)$.

**Exercise 1.3** Show that if $p \in \mathcal{P}^+_d$, then there exists a finite set $\{x_1, \ldots, x_k\}$ such that:

- $p(x_j) > 0, \quad j = 1, \ldots, k$,
- For every $y \in \mathbb{Z}^d$, there exist (strictly) positive integers $n_1, \ldots, n_k$ with $n_1 x_1 + \cdots + n_k x_k = y$.

Use this to show that there exists $\varepsilon > 0, q \in \mathcal{P}^+_d \cap \mathcal{P}^j_d$ such that $q$ has finite support and $p = c q + (1 - c) q'$.

**Exercise 1.4** Suppose that $S_n = X_1 + \cdots + X_n$ where $X_1, X_2, \ldots$ are independent $\mathbb{R}^d$-valued random variables with mean zero and covariance matrix $\Gamma$. Show that

$$
 M_n := |S_n|^2 - (\varepsilon \Gamma) n
$$

is a martingale.
Chapter 2
Local Central Limit Theorem

2.1 Introduction

If \( X_1, X_2, \ldots \) are independent, identically distributed random variables in \( \mathbb{R} \) with mean zero and variance \( \sigma^2 \), then the central limit theorem (CLT) says that the distribution of

\[
\frac{X_1 + \cdots + X_n}{\sqrt{n}}
\]

approaches that of a normal distribution with mean zero and variance \( \sigma^2 \). In other words, for \( -\infty < r < s < \infty \),

\[
\lim_{n \to \infty} P \left\{ r \leq \frac{X_1 + \cdots + X_n}{\sqrt{n}} \leq s \right\} = \int_r^s \frac{1}{\sqrt{2\pi}\sigma^2} e^{-y^2/(2\sigma^2)} dy.
\]

If \( p \in \mathcal{P}_1 \) is aperiodic with variance \( \sigma^2 \), we can use this to motivate the following nonrigorous approximation:

\[
p_n(k) = P(S_n = k) = P \left\{ \frac{k}{\sqrt{n}} \leq \frac{S_n}{\sqrt{n}} < \frac{k+1}{\sqrt{n}} \right\}
\approx \int_{k/\sqrt{n}}^{(k+1)/\sqrt{n}} \frac{1}{\sqrt{2\pi}\sigma^2} e^{-y^2/(2\sigma^2)} dy \approx \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left\{ -\frac{k^2}{2\sigma^2 n} \right\}.
\]

Similarly, if \( p \in \mathcal{P}_2 \) is bipartite, we can conjecture that

\[
p_n(k) + p_n(k+1) \approx \int_{k/\sqrt{n}}^{(k+2)/\sqrt{n}} \frac{1}{\sqrt{2\pi}\sigma^2} e^{-y^2/(2\sigma^2)} dy \approx \frac{2}{\sqrt{2\pi}\sigma^2} \exp \left\{ -\frac{k^2}{2\sigma^2 n} \right\}.
\]

The local central limit theorem (LCLT) justifies this nonrigorous approximation.

If \( p \in \mathcal{P}_2 \) with covariance matrix \( \Gamma = \Lambda \Lambda^T = \Lambda^2 \), then the normalized sums (2.1) approach a joint normal random variable with covariance matrix \( \Gamma \), i.e., a random variable with density

\[
f(x) = \frac{1}{(2\pi)^{d/2} \lvert \det \Lambda \rvert^{1/2}} e^{-||x||^2/2} = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Gamma}} e^{-\langle x, \Gamma^{-1} x \rangle/2},
\]

(See Section 8.3 for a review of the joint normal distribution.) A similar heuristic argument can be given for \( p_n(x) \). Recall from (1.1) that

\[
f^*(x)^2 = x \cdot \Gamma^2 x.
\]

We will use \( \mathbb{P}_n(x) \) to denote the estimate of \( p_n(x) \) that one obtains by the central limit theorem argument.

\[
\mathbb{P}_n(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Gamma}} e^{-\langle x, \Gamma^{-1} x \rangle/2} = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-\|y\|^2/2} dy,
\]

(2.2)

The second inequality is a straightforward computation, see (8.12). The LCLT states that for large \( n \), \( p_n(x) \) is approximately \( \mathbb{P}_n(x) \). To be more precise, we will say that an aperiodic \( p \) satisfies the weak LCLT if

\[
\lim_{n \to \infty} n^{d/2} \sup_{x \in \mathbb{R}^d} \|p_n(x) - \mathbb{P}_n(x)\| = 0,
\]

A bipartite \( p \) satisfies the weak LCLT if

\[
\lim_{n \to \infty} n^{d/2} \sup_{x \in \mathbb{R}^d} \|p_n(x) + p_{n+1}(x) - 2p_n(x)\| = 0.
\]

We call this a weak LCLT because we have not made any estimate of the error term \( |p_n(x) - \mathbb{P}_n(x)| \) other than it goes to zero faster than \( n^{-d/2} \) uniformly in \( x \). Note that \( \mathbb{P}_n(x) \) is bounded by \( c \exp(-\langle x \rangle^2/\sqrt{n}) \), but \( p_n(x) \) is much smaller for larger \( k \). One can prove the weak LCLT for any mean zero distribution with finite second moment. In this book we are considering symmetric walks with bounded range. For these we will give sharper estimates on the rate of convergence.

We will now state the stronger versions of the LCLT that we will prove. We will use two different approaches, one combinatorial and one using the characteristic function, and they will give different bounds on the error terms. The statements of the LCLT in this section combine the results from both of these methods.

\textbf{Theorem 2.1.1 (Local Central Limit Theorem)} If \( p \in \mathcal{P}_d \) is as defined in (2.1), then there is a \( c < \infty \) such that for all integers \( n > 0 \) and \( x \in \mathbb{Z}^d \),

\[
\|p_n(x) - \mathbb{P}_n(x)\| \leq \frac{c}{n^{d/2}}, \quad \|p_n(x) - \mathbb{P}_n(x)\| \leq \frac{c}{n^{d/2}}.
\]

(3.3)

if \( p \) is aperiodic and

\[
\|p_n(x) + p_{n+1}(x) - 2p_n(x)\| \leq \frac{c}{n^{d/2}}, \quad \|p_n(x) + p_{n+1}(x) - 2p_n(x)\| \leq \frac{c}{n^{d/2}}.
\]

(3.4)
2.1. INTRODUCTION

if $p$ is bipartite, also, if $t > 0$,

$$|\hat{p}_n(x) - \hat{P}_n(x)| \leq \frac{c}{n^{d/2}}, \quad |\hat{p}_n(x) - \hat{P}_n(x)| \leq \frac{c}{n^{d/2}}. \tag{2.6}$$

Moreover, there exists an $\epsilon > 0$ such that if $|x| \leq \epsilon n$, then

$$p_n(x) = \hat{p}_n(x) \exp \left\{ O \left( \frac{1}{n} + \frac{|x|}{n^2} \right) \right\}, \tag{2.7}$$

if $p$ is aperiodic and

$$p_n(x) + p_{n+1}(x) = 2 \hat{p}_n(x) \exp \left\{ O \left( \frac{1}{n} + \frac{|x|}{n^2} \right) \right\}, \tag{2.8}$$

if $p$ is bipartite. Also, if $|x| \leq \epsilon n$,

$$\hat{p}_n(x) = \hat{P}_n(x) \left[ 1 + O \left( \frac{1}{n} \right) \right], \quad |x| \leq \sqrt{n}. \tag{2.9}$$

We will prove this result in a number of steps in the following sections. Before doing so, we will consider what the theorem states. Let us restrict to the discrete time, aperiodic case. For "typical" $x$ with $|x| \leq \sqrt{n}$, $\hat{p}_n(x) \approx n^{-d/2}$. Hence the first equation in (2.3) can be written as

$$p_n(x) = \hat{p}_n(x) \left[ 1 + O \left( \frac{1}{n} \right) \right], \quad |x| \leq \sqrt{n}. \tag{2.3}$$

The error term in that equation is uniform over $x$, but as $|x|$ grows, the ratio between the error term and $p_n(x)$ grows. The second equation in (2.3) is an improvement on the error term for $|x| \geq \sqrt{n}$. An even better improvement for large $x$ is given by (2.7). If $|x| \leq n^{d/2}$, we can use a Taylor series approximation of the exponential to conclude that

$$p_n(x) = \hat{p}_n(x) \left[ 1 + O \left( \frac{n^{d/2}}{n^2} \right) \right]. \tag{2.4}$$

For very atypical $x$, i.e., $|x|$ of order $n^{d/2}$ or larger, Theorem 2.1.1 is not useful. However, in these cases, simple large deviation results as given in the next proposition will suffice.

**Proposition 2.1.2** Suppose $p \in P_d$ and $S_n$ is a $p$-walk starting at the origin.

(a) There exists a $c < \infty$ such that for all $s > 0$ and all $n$,

$$P \{|S_n| \geq s \sqrt{n}\} \leq ce^{-m}. \tag{2.10}$$

(b) For every $\epsilon > 0$, there exist $\rho > 0$, $c_1 < \infty$,

$$P \{|S_n| \geq \epsilon n\} \leq ce^{-m}. \tag{2.11}$$

**Proof.** It suffices to prove the result for one-dimensional walks. See Corollary 8.2.6. \[ \square \]

**Heuristic note.** The statement of the LCLT given here is stronger than is needed for many applications. For example, to determine whether or not the random walk is recurrent or transient, we only need the following corollary. If $p \in P_d$ is aperiodic, then there exist $0 < c_1 < c_2 < \infty$ such that for all $x$, $p_n(x) \leq c_1 n^{-d/2}$ and $n$ sufficiently large and $|x| \leq \sqrt{n}$, $p_n(x) \geq c_2 n^{-d/2}$. The exponent $d/2$ is important to remember and can be understood easily. In $n$ steps, the random walk tends to go distance $\sqrt{n}$. In $Z^d$, there are of order $n^{d/2}$ points within distance $\sqrt{n}$ of the origin. Therefore, the probability of being at a point should be of order $n^{-d/2}$.

There are times when one wants to compare $p_n(x)$ and $p_n(y)$ for nearby $x, y$. We will give estimates that show that this quantity is very close to what one would guess by differentiating $\hat{p}_n$ in the LCLT. If $y \in R$, let $\partial_y \hat{p}_n(x)$ denote the "directional derivative" of $\hat{P}_n(x)$.

$$\partial_y \hat{p}_n(x) = \frac{y \cdot \Gamma^{-1} x}{n} \hat{p}_n(x). \tag{2.12}$$

**Theorem 2.1.3 (LCLT Differece Estimates)** Suppose $p \in P_d$ and $R < \infty$. There exist $c, \epsilon$ such that the following holds. Suppose $t > 0$, $z \in Z^d$ with $|z| \leq \epsilon n$, $y \in Z^d$ with $|y| \leq R$, then

$$\nabla_z p_n(x) = \partial_z \hat{p}_n(x) + O_R \left( \frac{|z|^2}{n^2} \right) \hat{p}_n(x), \tag{2.13}$$

$$\nabla_y p_n(x) = \partial_y \hat{p}_n(x) + O_R \left( \frac{1}{n^{d+2/2}} \right). \tag{2.14}$$

If $p$ is aperiodic, the following also if $n$ is a positive integer

$$\nabla_z p_n(x) = \partial_z \hat{P}_n(x) + O_R \left( \frac{1}{n^{d+2/2}} \right) \hat{P}_n(x), \tag{2.15}$$

$$\nabla_y p_n(x) = \partial_y \hat{P}_n(x) + O_R \left( \frac{1}{n^{d+2/2}} \right). \tag{2.16}$$

The last two inequalities hold for bipartite $p$ provided that $y \in (Z^d)^{c}.$

2.1.1 Other lattices

The LCLT has an easy extension to random walks on more general lattices. Suppose $L$ is a $d$-dimensional lattice and $A : L \rightarrow Z^d$ is an isomorphism. Suppose $p \in P(L)$ is aperiodic and let $p^A$ denote the corresponding distribution in $P_L$, $p^A(Az) = p(x)$, if $p$ has covariance matrix $\Gamma = A \Lambda A^T$, then $p^A$ has covariance matrix $\Gamma_A := (A \Lambda A^T)^T$. Let

$$\hat{p}_n(x) = \hat{p}^A(Az) = \frac{1}{(2\pi n)^{d/2} \sqrt{\det \Gamma_A}} \exp \left\{ -\frac{|x|^2}{2n} \right\}. \tag{2.17}$$
2.2. LCLT — combinatorial approach

We will use an elementary counting approach to prove that
\[
\hat{p}_{\mu}(x) = \mathbb{E}(x) \exp \left\{ O \left( \frac{1}{\sqrt{n}} + \frac{|x|}{n^2} \right) \right\},
\]
and if \( p \) is aperiodic,
\[
p_{\mu}(x) = \mathbb{E}(x) \exp \left\{ O \left( \frac{1}{\sqrt{n}} + \frac{|x|}{n^2} \right) \right\}.
\]

This gives the estimates (2.7) and (2.9) for \( |x| \geq \sqrt{n} \). For \( |x| \leq \sqrt{n} \), these estimates will follow from (2.3) and (2.6) which will be proved in Section 2.3. We will also use this approach to derive the difference estimates (2.13) and (2.11).

We will only prove the result for simple random walk, both discrete and continuous time. The extension to \( p \in \mathcal{P}_d \) is straightforward using (1.2). The arguments are straightforward and relatively elementary, but they do require a lot of calculation. Here is a basic outline:

- Establish the result for one-dimensional random walk by exact counting of paths. Along the way we will prove Stirling’s formula.
- Prove an LCLT for Poisson random variables and use that to derive the result for one-dimensional continuous-time walks. The result for \( d \)-dimensional continuous-time walks follows immediately.

2.2.1 Stirling’s formula and \( 1 \)-d walks

Suppose \( S_n \) is a simple one-dimensional random walk starting at the origin. Determining the distribution of \( S_n \) reduces to an easy counting question. In order for \( X_1 + \cdots + X_n \) to equal \( 2k \), exactly \( n + k \) of the \( X_j \) must equal \( +1 \). Since all \( 2^{n-k} \) sequences of \( \pm 1 \) are equally likely,
\[
p_{\text{RW}}(2k) = \mathbb{P}(S_n = 2k) = 2^{-2n} \binom{2n}{n+k} \frac{(2n)!}{(n+k)!(n-k)!}.
\]

We will use Stirling’s formula, which we now derive, to estimate the factorials. In the proof, we will use some standard estimates about the logarithm. See Section 8.1.2.

Theorem 2.2.1 (Stirling’s formula) As \( n \to \infty \),
\[
n! \sim \sqrt{2\pi n} e^{-n} n^ne^{-n/2}.
\]

In fact,
\[
n! \sim \sqrt{2\pi n} e^{-n} n^ne^{-n/2}.\]

Proof. Let \( b_n = n^{n+(1/2)} e^{-n}/n! \). Then (8.4) and Taylor’s theorem imply
\[
\lim_{n \to \infty} \frac{b_{n+1}}{b_n} = \frac{1}{e} \left( 1 + \frac{1}{n} \right)^n = \left[ 1 - \frac{1}{2n} + \frac{11}{24n^2} + O \left( \frac{1}{n^3} \right) \right] \left[ 1 + \frac{1}{2n} - \frac{1}{8n^2} + O \left( \frac{1}{n^3} \right) \right] = 1 + O \left( \frac{1}{n^2} \right).
\]

Therefore,
\[
\lim_{n \to \infty} \frac{b_{n+1}}{b_n} = \prod_{n=1}^{\infty} \left[ 1 + \frac{1}{12n^2} + O \left( \frac{1}{n^3} \right) \right] = 1 + \frac{1}{12n} + O \left( \frac{1}{n^2} \right).
\]

The second equality is obtained by
\[
\log \prod_{n=1}^{\infty} \left[ 1 + \frac{1}{12n^2} + O \left( \frac{1}{n^3} \right) \right] = \sum_{n=1}^{\infty} \log \left[ 1 + \frac{1}{12n^2} + O \left( \frac{1}{n^3} \right) \right] = 1 + \frac{1}{12n} + O \left( \frac{1}{n^2} \right).
\]
This establishes that
\[
\bar{h}_n = C \left[ 1 - \frac{1}{12n} + O \left( \frac{1}{n^2} \right) \right],
\] (2.18)
for some positive constant \(C\). There are a number of ways to determine the constant \(C\). For example, if \(S_n\) denotes a one-dimensional simple random walk, then
\[
P \{ |S_{2n}| \leq \sqrt{2n} \log n \} = \sum_{k \leq \sqrt{2n} \log n} 4^{-n} \binom{2n}{n+k} = \sum_{k \leq \sqrt{2n} \log n} 4^{-n} \frac{(2n)!}{(n+k)!(n-k)!}.
\]
Using (8.2), we see that as \(n \to \infty\), if \(|k| \leq \sqrt{2n} \log n\),
\[
4^{-n} \frac{(2n)!}{(n+k)!(n-k)!} \sim \sqrt{\frac{\pi}{n}} \left( 1 + \frac{k}{n} \right)^{-n} \left( 1 - \frac{k}{n} \right)^{-n} \sim \sqrt{\frac{\pi}{n}} \frac{1}{\sqrt{n}} e^{-k^2/n}.
\]
Therefore,
\[
\lim_{n \to \infty} P \{ |S_n| \leq \sqrt{2n} \log n \} = \lim_{n \to \infty} \sum_{|k| \leq \sqrt{2n} \log n} \sqrt{\frac{\pi}{n}} \frac{1}{\sqrt{n}} e^{-k^2/n} = \sqrt{\frac{2}{\pi}}.
\]
However, Chebyshev's inequality shows that
\[
P \{ |S_n| \geq \sqrt{2n} \log n \} \leq \frac{\text{Var}[S_{2n}]}{2n \log^2 n} = \frac{1}{\log^2 n} \to 0.
\]
Therefore, \(C = \sqrt{2\pi}\). \(\square\)

**Remark.** By adapting this proof, it is easy to see that one can find \(r = 1/12, r_2, r_3, \ldots\)
such that for each positive integer \(k\),
\[
n! = \sqrt{2\pi n^{n+1/2} e^{-n}} \left[ 1 + \frac{r_1}{n} + \frac{r_2}{n^2} + \cdots + \frac{r_k}{n^k} + O \left( \frac{1}{n^{k+1}} \right) \right].\] (2.19)

We will now prove Theorems 2.1.1 and 2.1.3 in the special case of simple random walk in one dimension by using (2.17) and Stirling's formula. As a warmup, we start with the probability of being at the origin,

**Proposition 2.2.2** For simple random walk in \(\mathbb{Z}\), if \(n\) is a positive integer, then
\[
P \{ S_{2n} = 0 \} = \frac{1}{\sqrt{\pi n \log n}} \left[ 1 - \frac{1}{8n} + O \left( \frac{1}{n^2} \right) \right].
\]

**Proof.** The probability is exactly
\[
2^{-2n} \binom{2n}{n} = \frac{(2n)!}{4^n (n!)^2}.
\]
By plugging into Stirling's formula, we see that the right hand side equals
\[
\frac{1}{\sqrt{\pi n}} \frac{1}{1 + (12n)^{-1} + O(n^{-2})} \sim \frac{1}{\sqrt{\pi n}} \left[ 1 - \frac{1}{8n} + O \left( \frac{1}{n^2} \right) \right].
\]
\(\square\)

In the last proof, we just plugged in to Stirling's formula and evaluated. We will now do the same thing to prove a LCLT for one-dimensional simple random walk. Our proposition gives even sharper bounds on the error than those in (2.2). Note that for one-dimensional simple random walk,
\[
p_{2n}(2k) = \frac{1}{\sqrt{2\pi n \log n}} e^{-k^2/n} \exp \left( -\frac{(2k)^2}{2} \right) = \frac{1}{\sqrt{2\pi n}} e^{-k^2/n}.
\]

**Proposition 2.2.3** For simple random walk in \(\mathbb{Z}\), if \(n\) is a positive integer and \(k\) is an integer with \(|k| \leq n\),
\[
p_{2n}(2k) = P \{ S_{2n} = 2k \} = \frac{1}{\sqrt{2\pi n}} e^{-k^2/n} \exp \left( O \left( \frac{1}{n} + \frac{k^4}{n^3} \right) \right).
\]
In particular, if \(|k| \leq n^{1/4}, \) then
\[
P \{ S_{2n} = 2k \} = \frac{1}{\sqrt{2\pi n}} e^{-k^2/n} \left[ 1 + O \left( \frac{1}{n} + \frac{k^4}{n^3} \right) \right].
\]
The following difference estimates also hold:
\[
p_{2n}(2k+2) - p_{2n}(2k) = \frac{2k}{n} p_{2n}(2k) \left[ 1 + O \left( \frac{|k|}{n} \right) \right],
\]
\[
p_{3n+1}(2k) - p_{2n}(2k) = \left( \frac{k^2}{n^2} - \frac{1}{2n} \right) p_{2n}(2k) \left[ 1 + O \left( \frac{|k|}{n} \right) \right].
\]

**Heuristic note.** While the theorem is stated for all \(|k| \leq n\), it is not a very strong statement when \(k\) is of order \(n\). For example, for \(n/2 \leq |k| \leq n\), we can rewrite the conclusion as
\[
p_{2n}(2k) = \frac{1}{\sqrt{2\pi n}} e^{-k^2/n} e^{O(n)} = e^{O(n)},
\]
which only tells us that there exists \(O\) such that
\[
e^{-\alpha n} \leq p_{2n}(2k) \leq e^{\alpha n}.
\]
In fact, $2\mathcal{P}_n(2k)$ is not a very good approximation of $p_2(2k)$ for large $n$. As an extreme example, note that
\[ p_2(2n) = 4^n, \quad 2\mathcal{P}_n(2n) = \frac{1}{\sqrt{n}} e^{n^2}. \]

**Proof.** If $n/2 \leq |k| \leq n$, the first result is immediate (but also not very informative) using only the estimate $2^{2n} \leq \mathbb{P}(S_n = 2k) \leq 1$. Hence, we may assume that $|k| \leq n/2$. As noted before,
\[ \mathbb{P}(S_n = 2k) = 2^{2n} \left( \frac{2n}{n+k} \right)^{2n} \left( 1 - \frac{k^2}{n^2} \right)^{-1} \left( 1 - \frac{2k}{n+k} \right)^k. \]

If we restrict to $|k| \leq n/2$, we can use Stirling’s formula (Lemma 2.2.1) to see that
\[ \mathbb{P}(S_n = 2k) = \left[ 1 + O \left( \frac{1}{n} \right) \right] \frac{1}{\sqrt{2\pi n}} \left( 1 - \frac{k^2}{n^2} \right)^{-1/2} \left( 1 - \frac{2k}{n+k} \right)^k. \]

The last three terms all approach exponential functions. We need to be careful with the error terms. Using (8.2) we get,
\[ \left( 1 - \frac{k^2}{n^2} \right)^{1/2} = e^{k^2/n} \exp \left\{ O \left( \frac{k^2}{n^2} \right) \right\}, \]
\[ \left( 1 - \frac{2k}{n+k} \right)^k = e^{-k^2/(n+k)} \exp \left\{ \frac{2k^3}{(n+k)^2} + O \left( \frac{k^4}{n^3} \right) \right\}, \]
\[ e^{-k^2/(n+k)} = e^{-k^2/n} \exp \left\{ \frac{2k^3}{n^2} + O \left( \frac{k^4}{n^3} \right) \right\}. \]

Also, using $k^2/n^2 \leq \max\{1/n, (k^4/n^3)\}$, we can see that
\[ \left( 1 - \frac{k^2}{n^2} \right)^{1/2} = \exp \left\{ 1 + O \left( \frac{k^3}{n^3} \right) \right\}. \]

Combining all of this gives the first (and hence the second) equality in the theorem. The last two equalities follow directly from
\[ p_2(2k+2) = \frac{n-k}{n+k+1} p_2(2k) = p_2(2k) \left[ 1 - \frac{2k}{n} + O \left( \frac{k^2+n}{n^2} \right) \right] \]

and
\[ p_2(2n) = \frac{2n}{n+k+1} p_2(2k+2) = p_2(2k) \left[ 1 + \frac{1}{2n} \right] \left( 1 + \frac{k+1}{n} \right)^{-1} \left( 1 + \frac{k+1}{n} \right)^{-1} \]
\[ = p_2(2k) \left[ 1 + \frac{1}{2n} + \frac{k^2+n}{n^2} + O \left( \frac{k^3+n^3}{n^3} \right) \right]. \]

\section*{Chapter 2. Local Central Limit Theorem}

**Remark.** The dominant terms in the difference estimates in Theorem 2.2.3 are exactly what one would guess if one differentiated the function
\[ f(t,x) = \frac{1}{\sqrt{2\pi t}} e^{-t|x|^2} \]

with respect to $x$ and with respect to $t$, respectively.

**Corollary 2.2.4.** If $S_n$ is simple random walk, then for all positive integers $n$ and all $|k| < n$,
\[ \mathbb{P}(S_{n+1} = 2k+1) = \frac{1}{\sqrt{2\pi n}} \exp \left\{ -\frac{(k+1)^2}{2n} \right\} \exp \left\{ O \left( \frac{k^3}{n} \right) \right\}, \]

Hence,
\[ \mathbb{P}(S_{n+1} = 2k+1) = \frac{1}{\sqrt{2\pi n}} \left( e^{-k^2/n} + e^{-k^2/2n} \right) \exp \left\{ O \left( \frac{k^4}{n^3} \right) \right\}. \]

But,
\[ \exp \left\{ -\frac{(k+1)^2}{2n} \right\} = e^{-k^2/n} \left[ 1 - \frac{k}{n} + O \left( \frac{k^3}{n^3} \right) \right], \]
\[ \exp \left\{ -\frac{(k+1)^2}{2n} \right\} = e^{-k^2/n} \left[ 1 - \frac{2k}{n} + O \left( \frac{k^4}{n^3} \right) \right]. \]

Using $k^2/n^2 \leq \max\{1/n, (k^4/n^3)\}$, we get (2.20).

**Remark.** One might think that we should replace $n$ in (2.20) with $n + 1/2$. However,
\[ \frac{1}{n + (1/2)} = \frac{1}{n} \left[ 1 + O \left( \frac{1}{n} \right) \right]. \]

Hence, the error obtained by replacing one with the other is of the same order as the error in the corollary. Hence, the same statement with $n + 1/2$ replacing $n$ is also true.

### 2.2.2 LCLT for Poisson and continuous-time walks

The next proposition establishes the strong LCLT for Poisson random variables. This will be used for comparing discrete-time and continuous-time random walks with the same $p$. If $N_t$ is a Poisson random variable with parameter $t$, then $\mathbb{E}[N_t] = t, \text{Var}[N_t] = t$. The central
2.2. LCLT \hspace{0.5cm} \textbf{COMBINATORIAL APPROACH} \hspace{0.5cm} 23

limit theorem implies that as \( t \to \infty \), the distribution of \((N_t - t)/\sqrt{t}\) approaches that of a
standard normal. Hence, we might conjecture that
\[
P\{N_t = m\} = \mathcal{P}\left\{ \frac{m - t}{\sqrt{t}} \leq \frac{N_t - t}{\sqrt{t}} \leq \frac{m + 1 - t}{\sqrt{t}} \right\}
\approx \int_{(m-\frac{1}{2})/\sqrt{t}}^{(m+\frac{1}{2})/\sqrt{t}} \frac{1}{\sqrt{2\pi t}} e^{-x^2/2} \, dx \approx \frac{1}{\sqrt{2\pi t}} e^{-\frac{m^2}{2t}}.
\]

In the next proposition, we use a straightforward combinatorial argument to justify this
approximation.

\textbf{Proposition 2.2.5} Suppose \( N_t \) is a Poisson random variable with parameter \( t \), and \( m \) is
an integer with \( |m| \leq t/2 \). Then
\[
P\{N_t = m\} = \frac{1}{\sqrt{2\pi t}} e^{-\frac{m^2}{2t}} \exp\left\{ O\left( \frac{1}{\sqrt{t}} + \frac{|m|}{t} \right) \right\}.
\]

Moreover,
\[
P\{N_t = m + 1\} - P\{N_t = m\} = P\{N_t = m\} \left[ 1 - \frac{m - t}{t} + O\left( \frac{(m - t)^2}{t^2} \right) \right], \tag{2.21}
\]

\[
\partial_t P\{N_t = m\} = P\{N_t = m - 1\} - P\{N_t = m\}.
\]

\textbf{Proof.} For notational ease, we will first consider the case where \( t = n \) is an integer, and we
let \( m = n + k \). Let
\[
q(n,k) = P\{N_n = n + k\} = e^{-t} \frac{t^{n+k}}{(n+k)!},
\]
and note the recursion formula
\[
q(n,k) = \frac{n}{n+k} q(n,k-1).
\]

Stirling's formula (Theorem 2.2.1) gives
\[
q(n,0) = \frac{e^{-t} t^n}{\sqrt{2\pi n}} \approx 1 + O\left( \frac{1}{n} \right),
\]
By the recursion formula, if \( k \leq n/2 \),
\[
q(n,k) = q(n,0) \left[ \left( 1 + \frac{1}{n} \right) \left( 1 + \frac{2}{n} \right) \cdots \left( 1 + \frac{k}{n} \right) \right]^{-1},
\]
and,
\[
\log \prod_{j=1}^k \left( 1 + \frac{j}{n} \right) = \sum_{j=1}^k \log \left( 1 + \frac{j}{n} \right).
\]

\section{CHAPTER 2. LOCAL CENTRAL LIMIT THEOREM}

\[
= \sum_{j=1}^k \left( \frac{j}{n} + O\left( \frac{j^2}{n^2} \right) \right) = \frac{k^2}{2n} + O\left( \frac{k^3}{n^2} \right) = \frac{k^2}{2n} + O\left( \frac{1}{n} + \frac{k}{n^2} \right).
\]

The last equality uses the easy estimate \( (k/n) \leq \max\{1/\sqrt{t}, k^3/n^2\} \). Similarly,
\[
q(n, -k) = q(n,0) \left( 1 - \frac{1}{n} \right) \left( 1 - \frac{2}{n} \right) \cdots \left( 1 - \frac{k}{n} \right)
\]
and
\[
\log \prod_{j=1}^k \left( 1 - \frac{j}{n} \right) = \frac{k^2}{2n} + O\left( \frac{1}{n} + \frac{k}{n^2} \right).
\]

The proposition for integer \( n \) follows by exponentiating. For general \( t \), let \( n = \lfloor t \rfloor \) and note that
\[
P\{N_t = n + k\} = P\{N_n = n + k\} e^{-t} \left( 1 + \frac{t - n}{n} \right)^{n+k} = P\{N_n = n + k\} \left( 1 + \frac{t - n}{n} \right)^k \left[ 1 + O\left( \frac{1}{n} \right) \right]
\]
\[
P\{N_t = n + k\} = P\{N_n = n + k\} \left[ 1 + O\left( \frac{k}{n} \right) \right].
\]

Also,
\[
\frac{1}{\sqrt{t}} = \frac{1}{\sqrt{n}} \left[ 1 + O\left( \frac{1}{n} \right) \right], \quad e^{-t} = e^{-n} \exp\left\{ O\left( \frac{k^2}{n^2} \right) \right\}.
\]

The estimate (2.21) follows from
\[
P\{N_t = m + 1\} = \frac{t}{m+1} P\{N_t = m\} = P\{N_t = m\} \left[ 1 - \frac{m - t}{t} + O\left( \frac{m - t^2}{t^2} \right) \right].
\]

The last equality is immediate (and has a natural interpretation in terms of Poisson
processes). \hfill \Box

We will use this to prove Theorem 2.1.1 for one-dimensional, continuous-time simple random
walk.

\textbf{Theorem 2.2.6} If \( \hat{S}_t \) is continuous-time one-dimensional simple random walk, then \( |x| \leq t/2 \),
\[
\hat{p}_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2} \exp\left\{ O\left( \frac{1}{\sqrt{t}} + \frac{|x|^2}{t} \right) \right\}.
\]
2.2. LCLT COMBINATORIAL APPROACH

Proof. We will assume that \( x = 2k \) is even; the odd case is done similarly,

\[
\tilde{p}_n(x) = \sum_{m=0}^n P\{N_t = m\} p_m(2k),
\]

Standard exponential estimates show that for every \( \epsilon > 0 \), there exist \( c, \beta \) such that \( P\{|N_t - t| \geq ct\} \leq e^{-\epsilon t} \). Hence,

\[
\tilde{p}_n(x) = \exp \left\{ O \left( \frac{1}{\sqrt{t}} \right) \right\} \sum_{m=0}^n P\{N_t = m\} p_m(2k),
\]

where here and for the remainder of this proof, we write \( \sum \) alone to denote the sum over all integers \( m \) with \( |m - 2k| \leq ct \). We will specify \( \epsilon \) later in the proof.

In Theorem 2.2.3 we showed that

\[
P_{2m}(2k) = P\{S_{2m} = 2k\} = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(m-t)^2}{2t}} \exp \left\{ O \left( \frac{1}{\sqrt{t}} \right) \right\}.
\]

In Theorem 2.2.5 we showed that

\[
P\{N_t = 2m\} = \frac{1}{\sqrt{2\pi t}} e^{-\frac{m^2}{2t}} \exp \left\{ O \left( \frac{1}{\sqrt{t}} + \frac{m}{t} \right) \right\}.
\]

Also, we have

\[
\frac{1}{\sqrt{2m}} = \frac{1}{\sqrt{t}} \left[ 1 + O \left( \frac{m}{t} \right) \right], \quad \frac{1}{2m} = \frac{1}{t} \left[ 1 + O \left( \frac{m}{t} \right) \right],
\]

which implies

\[
e^{-\frac{(2k)^2}{2t}} = e^{-\frac{(m-t)^2}{2t}} \exp \left\{ O \left( \frac{(k^2 + m-t)^2}{t^2} \right) \right\}.
\]

Combining all of this, we get that the sum in (2.22) can be written as

\[
\frac{1}{\sqrt{2\pi t}} e^{-\frac{(m-t)^2}{2t}} \exp \left\{ O \left( \frac{(m-t)^2}{t^2} \right) \right\} \sum_{m=0}^n \frac{1}{\sqrt{\pi t}} e^{-\frac{m^2}{2t}} \exp \left\{ O \left( \frac{m}{t} \right) \right\}.
\]

We now choose \( \epsilon \) so that \( O\left(\frac{(m-t)^2}{t^2}\right) \leq (m-t)^2/(4t) \) for all \( |m - t| \leq ct \). We will now show that

\[
\sum_{m=0}^n \frac{1}{\sqrt{\pi t}} e^{-\frac{m^2}{2t}} \exp \left\{ O \left( \frac{(m-t)^2}{t^2} \right) \right\} = 1 + O \left( \frac{1}{\sqrt{t}} \right),
\]

which will complete the argument. Since

\[
e^{-\frac{m^2}{2t}} \exp \left\{ O \left( \frac{(m-t)^2}{t^2} \right) \right\} \leq e^{-\frac{m^2}{2t}},
\]

It is easy to see that the sum over \( |m - t| > t^{1/3} \) decays faster than any power of \( t \). For \( |m - t| \leq t^{1/3} \) we write

\[
\exp \left\{ O \left( \frac{(m-t)^2}{t^2} \right) \right\} = 1 + O \left( \frac{1}{\sqrt{t}} \right)
\]

The estimate

\[
\sum_{m=0}^n \frac{1}{\sqrt{\pi t}} e^{-\frac{m^2}{2t}} = O \left( \frac{1}{\sqrt{t}} \right) + \frac{1}{\sqrt{\pi t}} \int_{\infty}^{\infty} \frac{e^{-y^2}}{\sqrt{2\pi t}} dy = O \left( \frac{1}{\sqrt{t}} \right),
\]

is a standard approximation of an integral by a sum. Similarly,

\[
\sum_{m=0}^n \frac{1}{\sqrt{\pi t}} e^{-\frac{(m-t)^2}{2t}} \leq \frac{1}{\sqrt{\pi t}} \int_{\infty}^{\infty} \frac{e^{-y^2}}{\sqrt{2\pi t}} dy = O \left( \frac{1}{\sqrt{t}} \right).
\]

Proposition 2.2.7 Suppose \( p_1, \ldots, p_d \) are positive numbers with \( 0 < p_1 + \cdots + p_d \leq 1 \). Let \( \pi \in \mathcal{P}_d \) be given by \( p(\pi e_1) = p_1/2 \) and \( p(y) = 0 \) for all other \( y \neq 0 \). Then, (2.15) and (2.11) hold. Moreover, for every \( \epsilon > 0 \), the error term can be chosen uniformly over all \( p_1, \ldots, p_d \) with \( \min(p_1, \ldots, p_d) \geq \epsilon \).

Proof. The random walk \( \tilde{S}_t \) can be written as \( \tilde{S}_t = (\tilde{S}_{t1}, \ldots, \tilde{S}_{td}) \), where \( \tilde{S}^1, \ldots, \tilde{S}^d \) are independent, continuous-time simple random walks. Its covariance matrix is the diagonal matrix with entries \( p_1, \ldots, p_d \). The result then follows from Theorem 2.2.6. (Note that it suffices to establish (2.11) for \( e_1, \ldots, e_d \).)

\[\square\]

2.2.3 LCLT for binomial and multinomial

Suppose \( Y_n \) is a binomial random variable with parameters \( n \) and \( q \). Then \( \mathbb{E}[Y_n] = nq \) and \( \mathbb{V}[Y_n] = nq(1 - q) \) and the CLT tells us that \( \frac{\sqrt{n} (Y_n - nq)}{\sqrt{nq(1-q)}} \) approaches a standard normal distribution. This leads us to make the estimate

\[
P\{Y_n = k\} \approx \frac{1}{\sqrt{2\pi q(1-q)n}} \exp \left\{ -\frac{(k - qn)^2}{2q(1-q)n} \right\}.
\]

We can use Stirling's formula to make this rigorous.

Proposition 2.2.8 For every \( 0 < q < 1 \), let \( Y_n \) be a binomial random variable with parameters \( q \) and \( n \). Then for every integer \( k \) with \( |k - nq| < \min(q, 1 - q)/2 \),

\[
P\{Y_n = k\} = \frac{1}{\sqrt{2\pi q(1-q)n}} \exp \left\{ -\frac{(k - qn)^2}{2q(1-q)n} \right\} \exp \left\{ O \left( \frac{1}{\sqrt{n}} + \frac{|k - qn|^2}{n^2} \right) \right\}.
\]

Also,

\[
P\{Y_n = k + 1\} = P\{Y_n = k\} = \mathbb{P}(Y_n = k) \frac{k - qn}{(1-q)n} \left[ 1 + O \left( \frac{(k - qn)^2}{n^2} \right) \right].
\]

Moreover, for every \( \epsilon > 0 \), the \( O(\cdot) \) term can be chosen uniformly for \( \epsilon < q < 1 - \epsilon \).
2.2. LCLT

**Combinatorial Approach**

**Proof.** We fix $\epsilon$ and assume that $\epsilon \leq q \leq 1 - \epsilon$. Error terms in this proof can be bounded uniformly for all $q$ in this range. Let $r = k - q n$. Note that

$$
P(Y_n = k) = \frac{n!}{k!(n-k)!} \left(\frac{q}{n} \right)^k \left(1 - \frac{q}{n}\right)^{n-k} = \frac{n!}{(qn + r)(1 - q)^{n-r} n^r} \left(1 - \frac{q}{n}\right)^{n-r}.
$$

Using Stirling’s formula (Theorem 2.2.1) we see that if $k = q n \pm o(n)$, the right-hand side equals $[1 + O(n^{-1})]$ times

$$
\frac{1}{\sqrt{2\pi}} \left(\frac{r}{qn}\right)^{r/2} \left(1 - \frac{q}{n}\right)^{n-r}.
$$

All we have to do is estimate these terms. We will use (8.2).

Note that

$$
\sqrt{\frac{n}{(qn + r)(1 - q)n - r}} = \frac{1}{\sqrt{r(n-1)}} \left(1 + \frac{r}{n}\right)^{-r/2} \left(1 - \frac{r}{n(1-q)}\right)^{r/2}.
$$

$$
\left(\frac{1 + \frac{r}{n}}{n}\right)^{r} = e^r \exp \left\{-\frac{r^2}{2n} + O\left(\frac{r}{n^2}\right)\right\},
$$

$$
\left(1 - \frac{r}{(1-q)n}\right)^{n-r} = e^{-r} \exp \left\{-\frac{r^2}{2(1-q)n} + O\left(\frac{r}{n^2}\right)\right\},
$$

$$
\left(1 + \frac{r}{n}\right)^{r} = \left(1 + \frac{r}{n}\right)^{r} \exp \left\{-\frac{r^2}{2n} + O\left(\frac{r}{n^2}\right)\right\},
$$

$$
\left(1 - \frac{r}{n(1-q)}\right)^{r} = \left(1 - \frac{r}{n(1-q)}\right)^{r} \exp \left\{-\frac{r^2}{2n} + O\left(\frac{r}{n^2}\right)\right\}.
$$

Combining the last four inequalities gives

$$
\left(1 + \frac{r}{n}\right)^{r} \left(1 - \frac{r}{n}\right)^{n-r} = \exp \left\{-\frac{r^2}{2n} + O\left(\frac{r}{n^2}\right)\right\},
$$

and then combining all the estimates gives (2.23).

For the difference estimate, we write

$$
P(Y_n = k + 1) = P(Y_n = k) \left(1 - \frac{r}{n(1-q)}\right) - O\left(\frac{r^2}{n^2}\right),
$$

where $r = k - q n$, and $q_n = q - q/n$. Then

$$
P(Y_n = k + 1) \leq P(Y_n = k) \left(1 - \frac{r}{q_n}\right) + O\left(\frac{r^2}{n^2}\right).
$$

Moreover, the error term can be chosen uniformly over all $q \geq \epsilon$.

**2.2.4 d-dimensional simple random walk**

Here we establish the result for discrete-time simple random walk in $\mathbb{Z}^d$. In fact, we generalize a bit.

**Proposition 2.2.10** Suppose $p_1, \ldots, p_{d+1}$ are positive numbers with $p_1 + \cdots + p_{d+1} = 1$. Let $\pi \in \mathcal{P}_d$ be given by $p(\pm \mathbf{e}_i) = p_i/2$ and $p(0) = 0$ for all other $y \neq 0$. If $p_0 > 0$, and hence the walk is aperiodic, then

$$
p_n(x) = \mathbb{P}_n(x) \exp \left\{O\left(\frac{1}{n} + \frac{p}{n^2}\right)\right\},
$$

Moreover, for every $\delta > 0$, the error term can be chosen uniformly over all $p_1, \ldots, p_{d+1}$ with $\operatorname{min}(p_1, \ldots, p_{d+1}) \geq \delta$, if $p_{d+1} = p(0) = 0$, and hence the walk is bipartite, then

$$
p_n(x) + p_{n+1}(x) = 2\mathbb{P}_n(x) \exp \left\{O\left(\frac{1}{n} + \frac{p}{n^2}\right)\right\}.
$$

If $\delta > 0$, the error term can be chosen uniformly over all $p_1, \ldots, p_{d+1}$ with $\operatorname{min}(p_1, \ldots, p_{d+1}) \geq \delta$. \[\Box\]
2.2. LCLT  COMBINATORIAL APPROACH

Proof. We will do the case \( p(0) = 0 \). For ease we will assume that \( n \) is even and \( x = (x_1, \ldots, x_d) \) is such that \( x_1 + \cdots + x_d \) is even. Let \( q_m \) denote the probabilities for simple one-dimensional random walk. Let \( L_n = (L_n^1, \ldots, L_n^d) \) denote a multinomial process with parameters \( p_1, \ldots, p_d \). Note that

\[
p_n(x) = \sum_{m_1+\cdots+m_d=n} P(L_n^1 = m_1, \ldots, L_n^d = m_d) q_{m_1}(x^1) \cdots q_{m_d}(x^d), \tag{2.24}
\]

In Theorem 2.2.3 and Corollary 2.2.4 we showed that if \( m + k \) is even

\[
q_{m+n}(k) = P(S_m = k) = \frac{2}{\sqrt{2\pi n}} \exp \left( \frac{k^n}{2n} \right),
\]

Also, \( q_{m+n}(k) = 0 \) if \( m + k \) is odd, Proposition 2.2.9 implies

\[
P(L_n^1 = m_1, \ldots, L_n^d = m_d) = \left( \sum_{j=1}^d \frac{1}{\sqrt{2\pi p_j(1-p_j)n}} \exp \left( -\frac{|m_j - p_j|^2}{2p_j(1-p_j)n} \right) \right) \exp \left( \frac{|P|^2}{n} \right),
\]

where \( r = (m_1 - p_1 n, \ldots, m_d - p_d n) \). We now follow the proof of Theorem 2.2.6 to estimate the sum. We will choose an \( \epsilon > 0 \) and write the sum in (2.24) as

\[
O(e^{-n^n}) + \sum P(L_n^1 = m_1, \ldots, L_n^d = m_d) q_{m_1}(x^1) \cdots q_{m_d}(x^d), \tag{2.25}
\]

where \( \sum \) denotes the sum over \( m_j \) with \( |m_j - p_j n| \leq \epsilon n \). We write

\[
\frac{1}{\sqrt{m_j}} = \frac{1}{\sqrt{p_j n}} \left[ 1 + O \left( \frac{|m_j - p_j n|}{n} \right) \right], \quad \frac{1}{m_j} = \frac{1}{p_j n} \left[ 1 + O \left( \frac{|m_j - p_j n|}{n} \right) \right],
\]

\[
\exp \left( -\frac{|x_j|^2}{2m_j} \right) = \exp \left( -\frac{|x_j|^2}{2p_j n} \right) \exp \left( O \left( \frac{|x_j|^2 |m_j - p_j n|}{n} \right) \right),
\]

\[
\exp \left( \frac{|x_j|^2}{2m_j} \right) = \exp \left( \frac{|x_j|^2}{2p_j n} \right) \exp \left( O \left( \frac{|x_j|^2 |m_j - p_j n|}{n} \right) \right).
\]

We can therefore right the sum in (2.25) as

\[
\frac{2^d}{(2\pi)^{d/2} \prod_{j=1}^d p_j} \exp \left( -\sum_{j=1}^d \frac{|x_j|^2}{2p_j n} \right) \exp \left( O \left( \frac{|P|^2}{n} \right) \right)
\]

times

\[
\frac{1}{2^n} \sum_{j=1}^d \frac{1}{\sqrt{2\pi p_j(1-p_j)n}} \exp \left( -\frac{|m_j - p_j n|^2}{2p_j(1-p_j)n} \right) \exp \left( O \left( \frac{|P|^2}{n} \right) \right).
\]

The fact \( 2^d \) appears in the last expression because we should only be summing over \( m_1, \ldots, m_d \) of the correct parities \( m_1 + \cdots + m_d \) even. If we sum over all \( m_1, \ldots, m_d \), then our answer is off by \( 2^d \). This term cancels the \( 2^d \) in the term above. The sum is then estimated as in Theorem 2.2.6. Note that the dominant term in the sum is a Riemann sum approximation of the integral

\[
\int P \left( \sum_{j=1}^d \frac{1}{\sqrt{2\pi p_j(1-p_j)n}} \exp \left( -\frac{|x|^2}{2p_j(1-p_j)n} \right) \right) \exp \left( -\sum_{j=1}^d \frac{|x_j|^2}{2p_j n} \right) \, dy_1 \cdots dy_d = 1.
\]

Note that

\[
R_n(x) = \frac{1}{(2\pi)^{d/2} \prod_{j=1}^d p_j} \exp \left( -\sum_{j=1}^d \frac{|x_j|^2}{2p_j n} \right).
\]

The difference estimates are done as in Theorem 2.2.6.

2.3 Characteristic Functions and LCLT

2.3.1 Characteristic functions of random variables in \( \mathbb{R}^d \)

One of the most useful tools for studying the distribution of the sums of independent random variables is the characteristic function. If \( X = (X^1, \ldots, X^d) \) is a random variable in \( \mathbb{R}^d \), then its characteristic function \( \phi = \phi_X \) in the function from \( \mathbb{R}^d \) into \( \mathbb{C} \) given by

\[
\phi(\theta) = \mathbb{E}[\exp(i \theta^T X)].
\]

Proposition 2.3.1 Suppose \( X = (X^1, \ldots, X^d) \) is a random variable in \( \mathbb{R}^d \) with characteristic function \( \phi \),

(a) \( \phi \) is a uniformly continuous function with \( \phi(0) = 1 \) and \( |\phi(\theta)| \leq 1 \) for all \( \theta \in \mathbb{R}^d \),

(b) If \( \theta \in \mathbb{R}^d \) then \( \phi_X(\theta) := \phi(\theta) \) is the characteristic function of the one-dimensional random variable \( X \cdot \theta \).

(c) Suppose \( d = 1 \) and \( m \) is a positive integer with \( \mathbb{E}|X|^m \) $\leq 1$. Then \( \phi \) is a C$^m$ function of \( \theta \); in fact,

\[
\phi^{(m)}(s) = m! \mathbb{E}|X|^m e^{isX},
\]

(d) If \( m \) is a positive integer and \( \mathbb{E}|X|^m \) $\leq 1$ and \( |k| = 1 \), then

\[
|\phi(s) - \sum_{j=0}^{m-1} \frac{s^j}{j!} \mathbb{E}|X|^j| | \leq \frac{\mathbb{E}|X|^m}{m!} |k|^m.
\]

(e) If \( X_1, X_2, \ldots, X_n \) are independent random variables in \( \mathbb{R}^d \) with characteristic functions \( \phi_{X_1}, \ldots, \phi_{X_n} \), then \( \phi_{X_1 + \cdots + X_n}(\theta) = \phi_{X_1}(\theta) \cdots \phi_{X_n}(\theta) \).
2.3. CHARACTERISTIC FUNCTIONS AND LCLT

In particular, if $X_1, X_2, \ldots$ are independent, identically distributed with the same distribution as $X$, then the characteristic function of $S_n = X_1 + \cdots + X_n$ is given by

$$
\phi_{S_n}(\theta) = \left[\phi(\theta)\right]^n.
$$

Proof. Uniform continuity comes from the estimate

$$
|\phi(\theta + h) - \phi(\theta)| = |\mathbb{E}[e^{j\theta X} e^{jhX}] - e^{j\theta Xh}| \leq \mathbb{E}[(X - \mathbb{E}[X])^2] \to 0.
$$

The last step uses the dominated convergence theorem. The other parts of (a) and (b) are immediate. Part (c) is derived by differentiating; the condition $\mathbb{E}[|X|^p] < \infty$ is needed to justify the differentiation using the dominated convergence theorem (details omitted). Part (d) follows from (b), (c), and Taylor's theorem with remainder. Part (e) is immediate from the product rule for expectations of independent random variables.

We will write $P_n(\theta)$ for the $n$th order Taylor series approximation of $\phi$ about the origin.

Then the last proposition implies that if $\mathbb{E}[|X|^p] < \infty$, then

$$
\phi(\theta) = P_n(\theta) + o(|\theta|^n), \quad \theta \to 0,
$$

Note that if $\mathbb{E}[X] = 0$ and $\mathbb{E}[|X|^p] < \infty$, then

$$
P_n(\theta) = 1 - \frac{1}{2} \sum_{j=1}^{p} \sum_{k_{j+1}=1}^{d} \mathbb{E}[X^j \theta^k] - \frac{\theta^2}{2} = 1 - \frac{\mathbb{E}[X^2 \theta]^2}{2}.
$$

Here $\Gamma$ denotes the covariance matrix for $X$. Also, if all the third moments of $X$ exist and equal zero, such as is true for increment distributions $p \in \mathcal{P}_d$, $P_n(\theta) = P_2(\theta)$.

2.3.2 Characteristic functions of random variables in $\mathbb{Z}^d$

If $X = (X^1, \ldots, X^d)$ is a $\mathbb{Z}^d$-valued random variable, then its characteristic function has period $2\pi$ in each variable, i.e., if $k_1, \ldots, k_d$ are integers,

$$
\phi(\theta^1, \ldots, \theta^d) = \phi(\theta^1 + 2k_1\pi, \ldots, \theta^d + 2k_d\pi).
$$

The characteristic function determines the distribution of $X$; in fact, the next proposition gives a simple inversion formula. Here, and for the remainder of this section, we will write $db$ for $d\theta^1 \cdots d\theta^d$.

**Proposition 2.3.2** If $X = (X^1, \ldots, X^d)$ is a $\mathbb{Z}^d$-valued random variable with characteristic function $\phi$, then for every $x \in \mathbb{Z}^d$,

$$
P(X = x) = \frac{1}{(2\pi)^d} \int_{\mathbb{Z} \times \mathbb{Z}^d} \phi(\theta) e^{i\theta x} \, d\theta.
$$

32

CHAPTER 2. LOCAL CENTRAL LIMIT THEOREM

**Proof.** Since

$$
\phi(\theta) = \mathbb{E}[e^{i\theta X}] = \sum_{y \in \mathbb{Z}^d} e^{iy\theta} P(X = y),
$$

We get

$$
\int_{\mathbb{Z} \times \mathbb{Z}^d} \phi(\theta) e^{-i\theta x} \, d\theta = \sum_{y \in \mathbb{Z}^d} P(X = y) \int_{\mathbb{Z} \times \mathbb{Z}^d} e^{i(y - x)\theta} \, d\theta.
$$

(The dominated convergence theorem justifies the interchange of sum and integral.) But, if $x, y \in \mathbb{Z}^d$,

$$
\int_{\mathbb{Z} \times \mathbb{Z}^d} e^{i(y - x)\theta} \, d\theta = \begin{cases} (2\pi)^d, & y = x \quad \text{if } x \neq y, \end{cases}
$$

**Corollary 2.3.3** Suppose $X_1, X_2, \ldots$ are independent, identically distributed random variables in $\mathbb{Z}^d$ with characteristic function $\phi$. Let $S_n = X_1 + \cdots + X_n$. Then, for all $z \in \mathbb{Z}^d$,

$$
P(S_n = z) = \frac{1}{(2\pi)^d} \int_{\mathbb{Z} \times \mathbb{Z}^d} \phi(\theta) e^{-i\theta z} \, d\theta.
$$

2.3.3 LCLT — characteristic function approach

In some sense, Corollary 2.3.3 completely solves the problem of determining the distribution of a random walk at a particular time $n$. However, the integral is generally hard to evaluate and estimation of oscillatory integrals is tricky. Fortunately, we can use this corollary as a starting point for deriving the local central limit theorem. Because it is no more difficult, we will consider a wider class of random walks in this chapter. Recall that $\mathcal{P}_d$ denotes the set of increment distributions $p$ in $\mathbb{Z}^d$ with mean zero and finite variance, i.e.,

$$
\sum_{x \in \mathbb{Z}^d} x^T p(x) = 0, \quad \sum_{x \in \mathbb{Z}^d} |x|^4 p(x) < \infty,
$$

and such for all $z \in \mathbb{Z}^d$ there is an $N_z$ such that $p_n(x) > 0$ for $x \geq N_z$. Here, as before, we write $p_n(x)$ for the distribution of $S_n = X_1 + \cdots + X_n$ where $X_1, \ldots, X_n$ are independent with distribution $p$. As before, we let $\Gamma$ denote the covariance matrix which is nonsingular. We also write $S_t$ for a continuous time random walk with rates $p$. We let $\phi$ denote the characteristic function of $p$,

$$
\phi(\theta) = \sum_{x \in \mathbb{Z}^d} e^{i\theta x} p(x).
$$

The characteristic function of $S_n$ is $\phi^n$. The characteristic function of $S_t$ can be calculated easily using the fact that it has the same distribution as $S_N$, where $N_t$ is an independent Poisson process with parameter $t$.

$$
\phi_{S_t}(\theta) = \mathbb{E}[e^{iT\theta S_t}] = \sum_{j=0}^\infty e^{-\frac{t}{j!}} \mathbb{E}[e^{i\theta S_j}] = \sum_{j=0}^\infty e^{-\frac{t}{j!}} \phi(\theta)^j = \exp\{t[\phi(\theta) - 1]\}.
$$
Corollary 2.3.3 gives the formulas
\[ p_n(x) = \frac{1}{(2\pi)^2} \int_{-\pi,\pi} e^{i\theta x} e^{-i\theta^2} d\theta, \quad (2.27) \]
\[ \phi_k(x) = \frac{1}{(2\pi)^2} \int_{-\pi,\pi} e^{i\theta x} e^{-i\theta^2} d\theta. \]
We will prove the estimates for \( p_n(x) \); the arguments for \( \phi_k(x) \) are essentially the same.

**Lemma 2.3.4** Suppose \( p \in \mathcal{P} \).

(a) For every \( \epsilon > 0 \),
\[ \sup \{ |\phi(\theta)| : \theta \in [-\pi, \pi]^d, |\theta| \geq \epsilon \} < 1, \]
(b) There is a \( b > 0 \) such that for all \( \theta \in [-\pi, \pi]^d \),
\[ |\phi(\theta)| \leq 1 - |\theta|^b. \]

**Proof.** By continuity and compactness, to prove (a) it suffices to prove that \( |\phi(\theta)| < 1 \) for all \( \theta \in [-\pi, \pi]^d \setminus \{0\} \). To see this, suppose that \( |\phi(\theta)| = 1 \). Then \( |\phi(\theta)| = 1 \) for all positive integers \( n \). Since,
\[ \phi(\theta) = \sum_{n=0}^{\infty} \frac{1}{(2\pi)^2} \int_{-\pi,\pi} e^{i\theta x} e^{-i\theta^2} d\theta, \]
and for each fixed \( z \), \( p_n(z) > 0 \) for all sufficiently large \( n \), we see that \( e^{i\theta z} = 1 \) for all \( z \in \mathbb{Z}^d \). The only \( \theta \in [-\pi, \pi]^d \) that satisfies this is \( \theta = 0 \). Using (a), it suffices to prove (2.28) in a neighborhood of the origin, and this follows from the second-order Taylor series expansion (2.26).

In order to illustrate the proof of the local central limit theorem using the characteristic function, we will consider the one-dimensional case with \( p(1) = p(-1) = 1/4 \) and \( p(0) = 1/2 \). Note that this increment distribution is the same as the twostep distribution of (1/2 times) the usual simple random walk. The characteristic function for \( p \) is
\[ \phi(\theta) = \frac{1}{2} + \frac{1}{4} e^{i\theta} + \frac{1}{4} e^{-i\theta} = \frac{1}{2} + \frac{1}{2} \cos \theta = 1 - \frac{\theta^2}{4} + O(\theta^4). \]
The inversion formula (2.27) tells us that
\[ p_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\theta x} \phi(\theta) e^{-i\theta^2} d\theta = \frac{1}{2\pi \sqrt{n}} \int_{e^{-i\pi/2}}^{e^{i\pi/2}} e^{i\theta x} e^{-i\theta^2/n} \phi(\sqrt{n}) d\theta, \]
The second equality follows from the substitution \( s = \theta / \sqrt{n} \). We know that
\[ \phi \left( \frac{s}{\sqrt{n}} \right) = 1 - \frac{s^2}{4n} + O \left( \frac{s^4}{n^2} \right) = 1 - \frac{s^2}{4n} + O \left( \frac{s^4}{n^2} \right). \]
We can find \( \delta > 0 \), such that if \( |s| \leq \delta \sqrt{n} \),
\[ \left| \frac{s^4}{n} \right| \leq \frac{n}{2}. \]
Therefore, using (8.2), if \( |s| \leq \delta \sqrt{n} \),
\[ \phi \left( \frac{s}{\sqrt{n}} \right) = \left[ 1 - \frac{s^2}{4n} + O \left( \frac{s^4}{n^2} \right) \right] = e^{i\theta s} e^{i\theta^2/n}, \]
where
\[ |\phi(s/n)| \leq e^{s^2/4} \]
If \( \epsilon = \min(\delta, 1/\sqrt{8}e) \) we also have
\[ |\phi(s/n)| \leq e^{s^2/8}, \quad |s| < \epsilon \sqrt{n}. \]
For \( \epsilon \sqrt{n} < |s| < \sqrt{n} \), (2.28) shows that \( \left| \frac{1}{(2\pi)^2} \int_{-\pi,\pi} e^{i\theta x} e^{-i\theta^2/n} \phi \left( \sqrt{n} \right) d\theta \right| \leq e^{-\beta \theta} \) for some \( \beta > 0 \). Hence, up to an error that is exponentially small in \( n \), \( p_n(x) \) equals
\[ \frac{1}{2\pi \sqrt{n}} \int_{e^{-i\pi/2}}^{e^{i\pi/2}} e^{i\theta x} e^{-i\theta^2/n} \phi(\sqrt{n}) d\theta, \]
We now use
\[ k(x/n) = 1 \leq \left\{ \begin{array}{ll}
\frac{e^{s^2/n}}{e^{s^2/n}}, & |s| \leq n^{1/4} \\
|s| > n^{1/4}
\end{array} \right. \]
to bound the error term as follows:
\[ \left| \frac{1}{2\pi \sqrt{n}} \int_{e^{-i\pi/2}}^{e^{i\pi/2}} e^{i\theta x} e^{-i\theta^2/n} \phi(\sqrt{n}) - 1 \right| ds \leq \frac{e}{\sqrt{n}} \int_{e^{-i\pi/2}}^{e^{i\pi/2}} e^{s^2/4} k(x/n) - 1 | ds, \]
\[ \int_{e^{-i\pi/2}}^{e^{i\pi/2}} e^{s^2/4} k(x/n) - 1 | ds \leq \frac{e}{n} \int_{e^{-i\pi/2}}^{e^{i\pi/2}} \frac{1}{e^{s^2/4} k(x/n) - 1} ds \leq \frac{e}{n} \int_{e^{-i\pi/2}}^{e^{i\pi/2}} e^{s^2/4} k(x/n) - 1 ds \leq \frac{e}{n} \int_{e^{-i\pi/2}}^{e^{i\pi/2}} e^{s^2/8} ds = o(n^{-1}). \]
Hence we have
\[ p_n(x) = O \left( \frac{1}{n^{3/2}} \right) + \frac{1}{2\pi \sqrt{n}} \int_{e^{-i\pi/2}}^{e^{i\pi/2}} e^{i\theta x} e^{-i\theta^2/n} ds. \]
2.3. Characteristic Functions and LCLT

The last term equals $\mathcal{F}_n(x)$, see (2.2), and so we have shown that

$$p_n(x) = \mathcal{F}_n(x) + O \left( \frac{1}{n^{3/2}} \right).$$

We will follow this basic line of proof for theorems in this subsection. Before proceeding, it will be useful to outline the main steps:

- Change variables by $s = \sqrt{n} \theta$. This gives an integral over $[-\sqrt{n}, \sqrt{n}]^d$.
- Approximate the characteristic function by the exponential. The approximations will be valid in a neighborhood of radius $\epsilon \sqrt{n}$ about the origin. Use this to derive the dominant term which is an integral over $\mathbb{R}^d$ that can be computed exactly.
- Estimate the error terms. All of the error terms are bounded from above uniformly over $x \in \mathbb{Z}^d$.

One of the drawbacks of this approach is that the estimates for error terms are the same for all $x \in \mathbb{Z}^d$. These estimates are very good for "typical" $x$ of order $\sqrt{n}$, but for larger $x$ the ratio of the error to $\mathcal{F}_n(x)$ grows. The reader can compare this to (2.16) where the magnitude of the error term decreases as $|x|$ increases. The first equality in (2.3) is a special case of the next theorem.

**Theorem 2.3.5** Suppose $p \in \mathcal{P}'$ with $\mathbb{E}[X_1^3] < \infty$. Then there exist a $c < \infty$ such for all $n, x$,

$$|p_n(x) - \mathcal{F}_n(x)| \leq \frac{c}{n^{3/2}}.	ag{2.29}$$

If $\mathbb{E}[X_1^3] < \infty$ and all the third moments of $X_1$ are zero, then there is a $c$ such that for all $n, x$,

$$|p_n(x) - \mathcal{F}_n(x)| \leq \frac{c}{n^{1/2}}.	ag{2.30}$$

**Proof.** Let $\phi$ denote the characteristic function of $p$. Then we have

$$\phi(\theta) = 1 - \frac{\theta \cdot \Gamma \theta}{2} + O(|\theta|^{3+\epsilon}),$$

where $\alpha = 1$ under the weaker assumption and $\alpha = 2$ under the stronger assumptions. The inversion formula (2.27) gives

$$p_n(x) = \frac{1}{(2\pi)^d} \int_{|\theta| \leq \sqrt{n}} \phi(\theta) e^{i\theta \cdot x} \, d\theta = \frac{1}{(2\pi)^d} \int_{|\theta| \leq \sqrt{n}} \phi \left( \frac{s}{\sqrt{n}} \right) e^{is \cdot x} \, ds,$$

where $s = \sqrt{n} \theta$. We can find an $\epsilon > 0$, such that if $s \leq \epsilon \sqrt{n}$, the following approximation is valid:

$$\phi \left( \frac{s}{\sqrt{n}} \right) = \exp \left\{ n \log \phi \left( \frac{s}{\sqrt{n}} \right) \right\} = e^{-\frac{s^2}{2n}} e^{s \cdot x} \left( \frac{s}{\sqrt{n}} \right)^{3+\epsilon}.$$

Hence, we can conclude that

$$p_n(x) = \frac{1}{(2\pi)^d} \int_{|\theta| \leq \sqrt{n}} \phi \left( \frac{s}{\sqrt{n}} \right) e^{is \cdot x} \, ds.$$

By choosing $\epsilon$ smaller if necessary, we can guarantee that

$$|\phi(s)| \leq e^{-s^2/2}.$$

Lemma 2.3.5 implies that there is a $\beta > 0$, such that $|\phi(\theta)| \leq e^{-\beta |\theta|}$ for $|\theta| \geq \epsilon$. Therefore, up to an error that decays exponentially in $n$, $p_n(x)$ equals

$$\frac{1}{(2\pi)^d} \int_{|\theta| \leq \sqrt{n}} e^{-\frac{s^2}{2n}} e^{is \cdot x} \, ds = \mathcal{F}_n(x).$$

We have already seen (see (2.2)) that the first integral is

$$\frac{1}{(2\pi)^d} \int_{|\theta| \leq \sqrt{n}} e^{-\frac{s^2}{2n}} e^{is \cdot x} \, ds = \mathcal{F}_n(x).$$

The absolute value of the second integral bounded by a constant times

$$\frac{1}{n^{(3+\epsilon)/2}} \int_{|\theta| > \epsilon} |\phi(\theta)|^{3+\epsilon} \, d\theta + \frac{1}{n^{1/2}} \int_{|\theta| > \epsilon} |\phi(\theta)|^{1+\epsilon} \, d\theta = O \left( \frac{1}{n^{1/2}} \right).$$

The estimate (2.14) follows from the first equality in (2.3). However, we will give another proof below that does not require the third moments of the increment distribution to be zero. After that, we prove the theorem for which the second half of (2.3) is a special case.

**Heuristic note.** If $\alpha \neq 0$ and

$$f(n) = n^\alpha + O(n^{\alpha-1}),$$

then

$$f(n+1) - f(n) = [(n+1)^\alpha - n^\alpha] + O((n+1)^{\alpha-1}) = O(n^{\alpha-1}).$$

Clearly, $f(n+1) - f(n) = O(n^{\alpha-1})$, but the fact that we can write about the error terms in $O((n+1)^{\alpha-1}) - O(n^{\alpha-1}) = O(n^{\alpha-1})$. The error term for $f(n+1) - f(n)$ is as large as the dominant term. Hence, an expression such as (2.32) is not sufficient to give good bounds on differences of $f$. If one needs difference estimates, a first approach is to go back to the derivation of (2.32) to see if the difference of the errors can be estimated. This is the approach used in the next theorem.

**Theorem 2.3.6** If $p \in \mathcal{P}'$ with $\mathbb{E}[X_1^d] < \infty$, there is a $c$ such that for all $x, n$ and all $j = 1, \ldots, d$,

$$\nabla p_n(x) = \partial_j \mathcal{F}_n(x) + O \left( \frac{1}{n^{1/2}} \right).$$

CHAPTER 2. LOCAL CENTRAL LIMIT THEOREM

where $|\phi(s)| \leq e^{-s^2/2}$. By choosing $\epsilon$ smaller if necessary, we can guarantee that

$$|\phi(s)| \leq \frac{s \cdot \Gamma s}{2} \quad |\theta| \leq \epsilon \sqrt{n}.$$
2.3. CHARACTERISTIC FUNCTIONS AND LCLT

Proof. Let $y = e_j$. Note that

$$p_n(x + y) - p_n(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi(\theta) e^{iy \cdot \theta} \left[ e^{-it\theta} - 1 \right] \, d\theta = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi(\theta) \left[ e^{\frac{s}{\sqrt{n}}} - 1 \right] \, ds = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi(\theta) \left[ e^{\frac{s}{\sqrt{n}}} - 1 \right] \, ds$$

where $z = z/\sqrt{n}$ and $g(s, n)$ is as in the previous proof. If we choose $\varepsilon$ as in the previous proof we can see that if $\varepsilon$ is less than $1$ in the integral in the last integral to $|\varepsilon| \leq \varepsilon \sqrt{n}$, we only change the value by an exponentially small amount.

To estimate the "error" we note that

$$\left| \int_{\mathbb{R}^d} e^{-\frac{s^2}{n}} e^{is \cdot \theta} O \left( \frac{\|F\|}{n^{1/2}} \right) \, ds \right| \leq \int_{\mathbb{R}^d} \left| e^{-\frac{s^2}{n}} e^{is \cdot \theta} \right| \, ds \leq \sqrt{n} \int_{\mathbb{R}^d} e^{-\frac{s^2}{n}} \, ds = O(n^{-1/2}).$$

(The last estimate is done in the same way as the previous proof.) Therefore,

$$\nabla \phi p_n(x) = O \left( \frac{1}{n^{1+1/2}} \right) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} (iy \cdot s) e^{iy \cdot \theta} e^{is \cdot \theta} \, ds = \partial_j p_n(x) + O \left( \frac{1}{n^{1+1/2}} \right).$$

The last line follows from differentiating (2.31).

Theorem 2.3.7 Suppose $p \in P'$ is aperiodic with $\mathbb{E}|X_1|^p < \infty$. Then, there exist $c < \infty$ such for all $n, x$,

$$|p_n(x) - \mathbb{P}(\theta)| \leq \frac{c}{\mathbb{E} |X_1|^p} \cdot (2, 33)$$

If $\mathbb{E}|X_1|^p < \infty$ and all the third moments of $X_1$ are zero, then there is a $c$ such that

$$|p_n(x) - \mathbb{P}(\theta)| \leq \frac{c}{\mathbb{E} |X_1|^p} \cdot (2, 34)$$

Proof. We write

$$\phi(\theta) = 1 - \frac{\theta \cdot \Gamma \theta}{2} + O \left( |\theta|^p \right),$$

where $\alpha = 1$ under the first assumption and $\alpha = 2$ under the second assumption. We know that

$$p_n(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi(\theta) e^{iy \cdot \theta} \, d\theta.$$
2.3. CHARACTERISTIC FUNCTIONS AND LCLT

Also, \( f_j \equiv 0 \) for odd \( j \) if the odd moments are zero. By choosing \( \epsilon \) smaller if necessary, we can assume that for \( |x| \leq \epsilon \sqrt{n} \),

\[
\frac{f_0(x)}{n^{1/2}} + \frac{f_1(x)}{n} + \ldots + \frac{f_{m-1}(x)}{n^{(m-2)/2}} + g(s,n) \leq \frac{\sigma \Gamma_n}{4}.
\]

For \( |x| \leq n^{1/4} \) we can write

\[
\exp \left\{ \frac{f_0(x)}{n^{1/2}} + \frac{f_1(x)}{n} + \ldots + \frac{f_{m-1}(x)}{n^{(m-2)/2}} + g(s,n) \right\} = \frac{h_0(x)}{n^{1/2}} + \frac{h_1(x)}{n} + \ldots + \frac{h_{m-1}(x)}{n^{(m-2)/2}} + O \left( \frac{|x|^{m-1}}{n^{(m-1)/2}} \right),
\]

for some polynomials \( h_0, \ldots, h_{m-1} \). Again, \( h_k \) has degree at most \( 2k \).

**Theorem 2.3.8** Suppose \( p \in \mathcal{P}' \), \( m \geq 4 \) is a positive integer, and \( \mathbb{E}[|X|^{m+1}] < \infty \). Then we can write

\[
p_n(x) = \mathcal{P}_n(x) + u_3(x/\sqrt{n}) + \ldots + u_{m-1}(x/\sqrt{n}) + O \left( \frac{1}{n^{(m-1)/2}} \right),
\]

where \( u_3, u_4, \ldots \) are smooth functions defined by

\[
u_4(z) = \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} e^{-i\alpha \cdot x} h_4(s) e^{-2\pi^2 |s|^2} \, ds,
\]

and \( h_k \) are the polynomials defined in the previous paragraph. If the odd moments of \( p \) are zero, then the \( u_k \equiv 0 \) for odd \( k \). There exist constants \( c_k \) such that

\[
|u_k(z)| \leq c_k (|z|^{k+1} + 1) \exp \left\{ \frac{-z^k \Gamma_n^{k/2}}{2} \right\},
\]

**Proof.** The proof of (2.35) is similar to the previous propositions and we omit it. To establish (2.36), one does the change of variables \( r = \lambda s \) to get

\[
\int_{\mathbb{R}^3} e^{-i\alpha \cdot x} h_k(s) e^{-2\pi^2 |s|^2} \, ds = \frac{1}{\det \Lambda} \int_{\mathbb{R}^d} g_k(r) e^{-|r|^2/2} e^{i\lambda^\top r} \, dr,
\]

where \( g_k \) is the 2nd-degree polynomial \( g_k(\theta) = h_k(\Lambda \theta) \). By completing the square we see that the quantity on the right equals

\[
\frac{1}{\det \Lambda} e^{-|\lambda|^2/2} \int g(r) \exp \left\{ \frac{1}{2} (ir + \Lambda^{-1} z) \cdot (ir + \Lambda^{-1} z) \right\} \, dr.
\]

The substitution \( s = r - i\Lambda^{-1} z \) converts this integral to

\[
\frac{1}{\det \Lambda} e^{-|s|^2/2} \int g_k(r) e^{-|r|^2/2} \, dr,
\]

Finally,

\[
|g_k(r - i\Lambda^{-1} z)| \leq c (|r|^{k+1} + 1) (|r|^k + 1),
\]

which yields (2.36).

In the theorems in this section we have assumed that \( \mathbb{E}[|X|^3] < \infty \). This is required in order to bound \( |p_n(x) - \mathcal{P}_n(x)| \) in (2.29) by \( O(n^{-3/2}) \). If we only assume \( p \in \mathcal{P}' \), then we can still prove a LCLT but the best that we can do is show that the error is \( o(n^{-3/2}) \).

The proof is similar and we leave it as an exercise (Exercise 2.3). The starting point is the expansion for the characteristic function

\[
\phi(\theta) = 1 - \rho \cdot \Gamma_n \theta^2 + o(\theta), \quad \theta \to 0.
\]

**Corollary 2.3.9** Suppose \( p \in \mathcal{P}' \) is aperodic. Then there is a \( c \) such that for all \( n, x \),

\[
|p_n(x) - \mathcal{P}_n(x)| \leq c \left( \frac{|x|^{3/2} + 1}{n^{3/2}} + \frac{1}{n^{3/2}} \right).
\]

2.4 Some corollaries of the LCLT

Here we will list some applications of the strong LCLT,

**Corollary 2.4.1** (Second Difference Estimates) Suppose \( p \in \mathcal{P}' \) and \( R < \infty \). There exist \( c, \epsilon \) such that the following holds. Suppose \( t \geq 0 \), \( x \in \mathbb{Z}^d \), \( |x| \leq \epsilon n \), \( y \in \mathbb{Z}^d \) with \( |y| \leq R \), then

\[
|\nabla^2_x f(x)| \leq c \frac{|x|^{1/2} + 1}{n^{1/2}}.
\]

If \( p \) is aperiodic, the following also if \( n \) is a positive integer

\[
|\nabla^2_x p_n(x)| \leq c \frac{|x|^{1/2} + 1}{n^{1/2}}.
\]

The last inequality holds for bipartite \( p \) provided that \( y \in (\mathbb{Z}^d)_e \).

**Proof.** This follows immediately from Theorem 2.1.3.

**Remark.** For \( p \in \mathcal{P}' \), one can also prove this result by a proof similar to that in Theorem 2.3.6.

**Proposition 2.4.2** If \( p \in \mathcal{P}' \) with bounded support, there is a \( c \) such that

\[
\sum_{y \in \mathbb{Z}^d} |p_n(x) - p_n(x+y)| \leq c |y|^{n^{1/2}},
\]


2.4. SOME COROLLARIES OF THE LCLT

Proof. By the triangle inequality, it suffices to prove the result for $y = e_j$. Let $\delta = 1/2d$. By Theorem 2.3.6,

$$p_n(z + y) - p_n(z) = \partial_\nu p_n(z) + O\left(\frac{1}{n^{(3+\delta)/2}}\right).$$

Also Corollary 8.2.6 shows that

$$\sum_{k \in \mathbb{Z}^d} \left| \nu_n(z) - \nu_n(z + y) \right| \leq \sum_{k \in \mathbb{Z}^d} \left| \nu_n(z) + \nu_n(z + y) \right| = o(n^{-\delta/2}).$$

But,

$$\sum_{k \in \mathbb{Z}^d} \left| \nu_n(z) - \nu_n(z + y) \right| \leq \sum_{k \in \mathbb{Z}^d} \left| \nu_n(z) - \nu_n(z + y) \right| + \sum_{k \in \mathbb{Z}^d} O\left(\frac{1}{n^{(3+\delta)/2}}\right) \leq O(n^{-\delta/2}) + \sum_{k \in \mathbb{Z}^d} \left| \nu_n(z) - \nu_n(z + y) \right|.$$

We leave it to the reader to verify

$$\sum_{k \in \mathbb{Z}^d} \left| \nu_n(z) - \nu_n(z + y) \right| = O(n^{-\delta/2}).$$

□

In fact, the last proposition holds with much weaker assumptions on the random walk. Recall that $\mathcal{P}^*$ is the set of increment distributions $\nu$ with the property that for each $x \in \mathbb{Z}^d$, there is an $N_x$ such that $\nu_n(x) > 0$ for all $n \geq N_x$.

Proposition 2.4.3 If $\nu \in \mathcal{P}^*$, there is a $c$ such

$$\sum_{k \in \mathbb{Z}^d} \left| \nu_n(z) - \nu_n(z + y) \right| \leq c |y| n^{-\delta/2}.$$

Proof. In Exercise 1.3 it was shown that we can write

$$p = c q + (1 - c)\tilde{q},$$

where $q, \bar{q} \in \mathcal{P}^*, \tilde{q} \not\in \mathcal{P}^*$. By considering the process of first choosing $q$ or $\tilde{q}$ and then doing the jump, we can see that

$$p_n(x) = \sum_{j=0}^n \binom{n}{j} \epsilon^j (1 - \epsilon)^{n-j} \sum_{z \in \mathbb{Z}^d} q_j(x; \tilde{q}) (x - z),$$

(2.39)

Therefore,

$$\sum_{k \in \mathbb{Z}^d} \left| \nu_n(z) - \nu_n(z + y) \right| \leq$$

We split the first sum into the sum over $j < (\epsilon/2)n$ and $j \geq (\epsilon/2)n$. Standard exponential estimates for the binomial (see, for example, (8.10)) give

$$\sum_{j < (\epsilon/2)n} \left(\binom{n}{j} \epsilon^j (1 - \epsilon)^{n-j} \sum_{x \in \mathbb{Z}^d} \left| q_j(x) - q_j(x + y) \right| \right) \leq 2 \sum_{j < (\epsilon/2)n} \left(\binom{n}{j} \epsilon^j (1 - \epsilon)^{n-j} \right) = O(e^{-\alpha n}),$$

for some $\alpha = o(1) > 0$. However,

$$\sum_{j > (\epsilon/2)n} \left(\binom{n}{j} \epsilon^j (1 - \epsilon)^{n-j} \sum_{x \in \mathbb{Z}^d} \left| q_j(x) - q_j(x + y) \right| \right) \leq ce^{-\alpha n/2} \sum_{j > (\epsilon/2)n} \left(\binom{n}{j} \epsilon^j (1 - \epsilon)^{n-j} \sum_{x \in \mathbb{Z}^d} \left| q_j(x) \right| \right) \leq ce^{-\alpha n/2} \leq c n^{-\delta/2}.$$

□

The last proposition has the following useful lemma as a corollary. Since this is essentially a result about Markov chains in general, we leave the proof to the appendix, see Section 8.4.2.

Lemma 2.4.4 Suppose $p \in \mathcal{P}_d$. There is a $c < \infty$ such that if $x, y \in \mathbb{Z}^d$, we can define $S_n, S_n'$ on the same probability space such that: $S_n$ has the distribution of a random walk with increment $p$ with $S_0 = x$; $S_n'$ has the distribution of a random walk with increment $p$ with $S_0 = y$; and such that for all $n$,

$$\mathbb{P}\{S_n = S_n' \text{ for all } m \geq n\} \geq 1 - \frac{c |x - y|}{\sqrt{n}}.$$

Heuristic note. While the proof of this last lemma is a little messy to write out in detail, there really is not a lot of content to it. Suppose that $p, q$ are two probability distributions on $\mathbb{Z}^d$ with

$$\sum_{z \in \mathbb{Z}^d} |p(z) - q(z)| = 2e,$$

Then there is an easy way to define random variables $X, Y$ on the same probability space such that $X$ has distribution $p$, $Y$ has distribution $q$ and $\mathbb{P}\{X \neq Y\} = \epsilon$. Indeed, if we let $f(z) = \min\{p(z), q(z)\}$ we can let the probability space be $\mathbb{Z}^d \times \mathbb{Z}^d$ and define $\mu$ by

$$\mu(z, z') = f(z).$$
2.4. SOME COROLLARIES OF THE LCLT

and for \( x \neq y \),
\[
\mu(x, y) = e^{-1} [ p(x) - f(x) ] [ q(y) - f(y) ].
\]

If we let \( X(x, y) = x, Y(x, y) = y \), it is easy to check that the marginal of \( X \) is \( p \), the marginal of \( Y \) is \( q \) and \( P[X = Y] = 1 - \epsilon \). The more general fact is not much more complicated than this.

**Proposition 2.4.5** Suppose \( p \in P'_d \). There is a \( c < \infty \) such that for all \( n, x \),
\[
p_n(x) \leq \frac{c}{n^{d/2}}. \tag{2.40}
\]

**Proof.** If \( p \in P'_d \) with bounded support this follows immediately from (2.29). For general \( p \in P'_d \), write \( p = e q + (1 - e) q' \) with \( q \in P'_d, q' \in P'_d \) as in the proof of Proposition 2.29. Then \( p_n(x) \) as in (2.30). The sum over \( j < (e/2)n \) is \( O(e^{d/2}n) \) and for \( j \geq (e/2)n \), we have the bound \( q_{j-n}(z) \leq c n^{-d/2} \). \( \square \)

**Exercises for Chapter 2**

**Exercise 2.1** Find \( r_2, r_3 \) in (2.19).

**Exercise 2.2** Let \( S_n \) denote one-dimensional simple random walk. In this exercise we will prove without using Shrink's formula that there exists a constant \( C \) such that
\[
p_{2n}(0) = C \frac{1}{\sqrt{n}} \left[ 1 - \frac{1}{8n} + O \left( \frac{1}{n^2} \right) \right].
\]
a. Show that if \( n \geq 1 \),
\[
p_{n \rightarrow 2n} = \left( 1 + \frac{1}{2n} \right) \left( 1 + \frac{1}{n} \right)^{-1} p_{2n},
\]
b. Let \( b_n = \sqrt{n} p_{2n}(0) \). Show that \( b_n = 1/2 \) for \( n \geq 1 \),
\[
\frac{b_{n+1}}{b_n} = 1 + \frac{1}{8n^2} + O \left( \frac{1}{n^3} \right).
\]
c. Use this to show that \( b_n = \lim b_n \) exists and is positive. Moreover,
\[
b_n = b_n \left[ 1 - \frac{1}{8n} + O \left( \frac{1}{n^2} \right) \right].
\]

**Exercise 2.3** Suppose \( p \in P_d \) is aperiodic. Show that there exists a sequence \( \delta_n \rightarrow 0 \) such that for all \( n, x \),
\[
|p_n(x) - p_n(x)| \leq \frac{\delta_n}{n^{d/2}}. \tag{2.41}
\]

**Exercise 2.4** Suppose \( p \in P_d, \epsilon \in (0, 1) \) and \( \mathbb{E}[|X|^p] < \infty \). Show that the characteristic function has the expansion
\[
\phi(\theta) = 1 - \theta^\epsilon T + O(\theta^{2\epsilon}), \quad \theta \to 0,
\]

Show that the \( \delta_n \) in the previous section can be chosen so that \( n^{d+1} \delta_n \to 0 \).

**Exercise 2.5** Show that if \( p \in P'_d \) there exists a \( c \) such that for all \( x \in \mathbb{Z}^d \) and all positive integers \( n \),
\[
|p_n(x) - p_n(0)| \leq c \frac{|x|}{n^{d+1/2}}.
\]

(Hint: first show the estimate for \( p \in P'_d \) and then use (2.39). Alternatively, one can use Lemma 2.4.4 at time \( n/2 \), the Markov property, and (2.40).)

**Exercise 2.6** Show that Lemma 2.3.4 holds for \( p \in P_d \).

**Exercise 2.7** Suppose \( p \in P'_d \) with \( \mathbb{E}[|X|^p] < \infty \). Show that there is a \( c < \infty \) such that for all \( y \in \mathbb{Z}^d \),
\[
|p_n(0) - p_n(y)| \leq \frac{c |y|}{n^{d+2/2}}.
\]

**Exercise 2.8** Suppose \( p \in P'_d \). Let \( A \subseteq \mathbb{Z}^d \) and
\[
\phi(x) = P^* \{ S_n \in A \ \text{ i.o.} \}.
\]

Show that if \( \phi(x) > 0 \) for some \( x \in \mathbb{Z}^d \), then \( \phi(x) = 1 \) for all \( x \in \mathbb{Z}^d \).
Chapter 3

Approximation by Brownian motion

3.1 Introduction

Suppose $S_n = X_1 + \ldots + X_n$ is a one-dimensional simple random walk. Let us make this into a continuous function by linear interpolation,

$$S_t = S_n + (t - n)[S_{n+1} - S_n], \quad n \leq t \leq n + 1,$$

For fixed integer $n$, the LCLT describes the distribution of $S_n$. A corollary of LCLT is the usual central limit theorem that states that the distribution of $n^{-1/2} S_n$ converges to that of a standard normal random variable. A simple extension of this is the following: suppose $0 < t_1 < t_2 < \ldots < t_k = 1$. Then as $n \to \infty$ the distribution of

$$n^{-1/2} (S_{t_1}, S_{t_2}, \ldots, S_{t_k})$$

converges to that of

$$(Y_1, Y_{t_1}, Y_{t_2} - Y_{t_1}, \ldots, Y_k - Y_{t_{k-1}}),$$

where $Y_1, \ldots, Y_k$ are independent mean zero random variables with variables $t_1, t_2 - t_1, \ldots, t_k - t_{k-1}$, respectively.

The functional central limit theorem (also called the invariance principle or Donsker's theorem) for random walk extends this result to the random function

$$W_t^{(n)} := n^{-1/2} S_{tn}.$$  \hspace{1cm} (3.1)

The functional central limit theorem states roughly that as $N \to \infty$, the distribution of this random function converges to the distribution of a random function $t \to B_t$. From what we know about the simple random walk, here are some properties that would be expected of the random function $B_t$:

- If $s < t$, the distribution of $B_t - B_s$ is $N(0, t - s)$.
- If $0 \leq t_0 < t_1 < \ldots < t_k$, then $B_{t_0}, B_{t_1} - B_{t_0}, \ldots, B_{t_k} - B_{t_{k-1}}$ are independent random variables.

These two properties follow almost immediately from the central limit theorem. The third property is not as obvious.

- The function $t \to B_t$ is continuous.

Although this is not obvious, we can guess this from the heuristic argument:

$$E[B_{t + \Delta t} - B_t] = \Delta t,$$

which indicates that $|B_{t + \Delta t} - B_t|$ should be of order $\Delta t$. A process satisfying these assumptions will be called a Brownian motion (we will define it more precisely in the next section).

There are a number of ways to make rigorous the idea that $W^{(n)}$ approaches a Brownian motion in the limit. For example, if we restrict to $0 \leq t \leq 1$, then $W^{(n)}$ and $B$ are independent random variables taking values in the metric space $C[0, 1]$ with the supremum norm. There is a well understood theory of convergence in distribution of random variables taking values in metric spaces.

We will take a different approach using a method that is often called strong approximation of random walk by Brownian motion. We start by defining a Brownian motion $B$ on a probability space and then define the random walk $S_n$ as a function of the Brownian motion, i.e., for each realization of random function $B_t$ we associate a particular random walk path. We will do this in a way so that the random walk $S_n$ has the distribution of simple random walk. We will then do some estimates to show that there exist positive constants $c, \delta$ such that if $W^{(n)}$ is as defined in (3.1), then for all $r \leq n^{1/4}$,

$$P\{\|B - W^{(n)}\| \geq r N^{-1/4} \sqrt{\log n}\} \leq ce^{-\delta n},$$  \hspace{1cm} (3.2)

where $\| \cdot \|$ denotes the supremum norm on $C[0, 1]$. The convergence in distribution follows from the strong estimate (3.1).

Heuristic note. There is a general approach here that is worth emphasizing. Suppose we have a discrete process and we want to show that it converges after some scaling to a continuous process. A good approach for proving such a result is to first study the conjectured limit process and then to show that the scaled discrete process is a small perturbation of the limit process.

We start by establishing (3.1) for one-dimensional simple random walk using Skorokhod embedding. We then extend this to continuous-time walks and all increment distributions $p \in \mathcal{P}$. The extension will not be difficult; the hard work is done in the one-dimensional case.

We will not handle the general case of $p \in \mathcal{P}$ in this book. One can give strong approximations in this case to show that the random walk approaches Brownian motion. However, the rate of convergence depends on the moment assumptions. In particular, the estimate (3.2) will not hold assuming only mean zero and finite second moment.
3.2 Construction of Brownian motion

A standard (one-dimensional) Brownian motion with respect to a filtration $\mathcal{F}_t$ is a collection of random variables $B_t$, $t \geq 0$ satisfying the following:

(a) $B_0 = 0$;

(b) if $s < t$, then $B_t - B_s$ is an $\mathcal{F}_s$-measurable random variable, independent of $\mathcal{F}_s$, with a $N(0,t-s)$ distribution;

(c) with probability one, $t \mapsto B_t$ is a continuous function.

If the filtration is not given explicitly, then it is assumed to be the natural filtration, $\mathcal{F}_t = \sigma(B_s : 0 \leq s \leq t)$. In this section, we will construct a Brownian motion and derive an important estimate on the oscillations of the Brownian motion.

We will show how to construct a Brownian motion. There are technical difficulties involved in defining a collection of random variables $\{B_t\}$ indexed over an uncountable set. However, if we know a priori that the distribution should be supported on continuous functions, then we know that the random function $t \mapsto B_t$ should be determined by its value on a countable, dense subset of times. This observation leads us to a method of constructing Brownian motion: define the process on a countable set of times and then extend the process to all times by continuity.

Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is any probability space that is large enough to contain a countable collection of independent $N(0,1)$ random variables $\{Z_n\}$ for each we will index by

$$N_{n,k}, \quad n = 0, 1, \ldots, \quad k = 0, 1, \ldots$$

We will use these random variables to define a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$. Let

$$D_n = \left\{ \frac{k}{2^n} : \quad k = 0, 1, \ldots \right\}, \quad D = \bigcup_{n=0}^{\infty} D_n$$

denote the nonnegative dyadic rationals. Our strategy will be as follows:

- define $B_t$ for $t \in D$ satisfying conditions (a) and (b);
- derive an estimate on the oscillation of $B_t$, $t \in D$, that implies that with probability one the paths are uniformly continuous on compact intervals;
- define $B_t$ for other values of $t$ by continuity.

The first step is straightforward using a basic property of normal random variables. Suppose $X$, $Y$ are independent normal random variables, each mean 0 and variance 1/2. Then $Z = X + Y$ is $N(0,1)$. Moreover, the conditional distribution of $X$ given the value of $Z$ is normal with mean $Z/2$ and variance 1/2. This can be checked directly using the density of the normals. Alternatively, one can check that if $Z, \bar{Z}$ are independent $N(0,1)$ random variables then

$$X := \frac{Z}{2} + \frac{\bar{Z}}{2}, \quad Y := \frac{Z}{2} - \frac{\bar{Z}}{2}. \quad (3.3)$$

are independent $N(0,1/2)$ random variables. To verify this, one only needs to note that $(X,Y)$ has a joint normal distribution with $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1/2$, $\mathbb{E}[XY] = 0$. (See Corollary 3.3.1.) This tells us in order to define $X, Y$ we can start with independent random variables $Z, \bar{Z}$ and then use (3.3).

We start by defining $B_t$ for $t \in D_0$. By $B_0 = 0$ and

$$B_t = N_{0,1} + \cdots + N_{0,t},$$

We then continue recursively using (3.3). Suppose $B_t$ has been defined for all $t \in D_n$. Then we define $B_t$ for $t \in D_{n+1} \setminus D_n$ by

$$B_{t+2^{-n}} = B_{t+2^{-n+1}} + \frac{1}{2} \left[ B_{t+2^{-n}} - B_{t+2^{-(n+1)}} \right] + 2^{1/2-n/2} N_{0,1+2^{-n}}.$$

By induction, one can check that for each $n$ the collection of random variables $\mathbb{Z}_{k,n} := B_{k2^{-n}} - B_{(k-1)/2^{-n}}$ are independent, each with a $N(0,2^{-n})$ distribution. Since this is true for each $n$, we can see that (a) and (b) hold (with the natural filtration) provided that we restrict to $t \in D$.

We define the oscillation of $B_t$ (restricted to $t \in D$) by

$$\text{osc}(B_t) = \sup_{s \in D \atop 0 \leq s \leq t} |B_s - B_t|,$$\quad (3.4)

For fixed $\delta, \epsilon$, this is an $\mathcal{F}_t$-measurable random variable. We write $\text{osc}(B_t)$ for $\text{osc}(B_t; \delta, \epsilon)$. We also define a similar random variable that will be easier to estimate.

$$M_n = \max_{0 \leq t \leq 2^n} \sup_{0 \leq s \leq t} |B_{s+n} - B_s|,$$

We note that

$$\text{osc}(B_t; 2^{-n}) \leq 3 M_n.$$\quad (3.4)

To see this, suppose $\delta \leq 2^{-n}$, $0 < s < t < s + \delta \leq 1$, and $|B_s - B_t| \geq \epsilon$. Then there exists a $k$ such that either $k2^{-n} \leq s < (k+1)2^{-n}$ or $(k-1)2^{-n} \leq s < k2^{-n} < t < (k+1)2^{-n}$. In either case, the triangle inequality tells us that $M_n \geq \epsilon \sqrt{2}$. We will prove a proposition that bounds the probability of large values of $\text{osc}(B_t)$, $\text{osc}(B_t; \delta, \epsilon)$. We start with a lemma which gives a similar bound for $M_n$.

**Lemma 3.2.1** For every integer $n$ and every $\epsilon > 0$,

$$\mathbb{P}(M_n > \delta 2^{-n/2}) \leq 4 \sqrt{\frac{a2^n}{\pi \delta}} e^{-\delta^2/2}.$$

**Proof.** Note that

$$\mathbb{P}(M_n > \delta 2^{-n/2}) \leq 2^n \mathbb{P} \left( \sup_{0 \leq s \leq t \leq 2^n} |B_s| > \delta 2^{-n/2} \right) = 2^n \mathbb{P} \left( \sup_{0 \leq s \leq 1} |B_s| > \delta \right).$$
3.2. CONSTRUCTION OF BROWNIAN MOTION

Here the supremums are over \( t \in D \). Also note that

\[
\mathbb{P} \{ \sup \{ B_t : 0 \leq t \leq 1, t \in D \} > \delta \} = \lim_{n \to \infty} \mathbb{P} \{ \max \{ B_{k/n} : k = 1, \ldots, 2^n \} > \delta \} \leq 2 \lim_{n \to \infty} \mathbb{P} \{ \max \{ B_{k/n} - : k = 1, \ldots, 2^n \} > \delta \}.
\]

The reflection principle (see Proposition 1.5.2 and the remark following) shows that

\[
\mathbb{P} \{ \max \{ B_{k/n} - : k = 1, \ldots, 2^n \} > \delta \} \leq 2 \mathbb{P} \{ B_1 > \delta \} = 2 \int_0^\infty \frac{1}{2 \pi} e^{-x^2/2} dx = 2 \int_0^\infty \frac{1}{2 \pi} e^{-x^2/2} dx = 2 \int_0^\infty \frac{1}{\sqrt{2 \pi}} e^{-x^2/2} dx = e^{x^2/2},
\]

\[\square\]

**Proposition 3.2.2** There exists a \( c \) such that for every \( 0 < \delta \leq 1, r \geq 1 \), and positive integer \( T \),

\[
\mathbb{P} \{ \text{osc}(B; \delta, T) > cr \sqrt{\delta \log(1/\delta)} \} \leq c T \delta^r.
\]

**Proof.** It suffices to prove the result for \( T = 1 \) since for general \( T \) we can estimate separately the oscillations over the \( 2^{T-1} \) intervals \([0, 1], [1/2, 3/2], [1, 2], \ldots, [T-1, T]\). Also, it suffices to prove the result for \( \delta \leq 1/4 \). Suppose that \( 2^{-m+1} \leq \delta \leq 2^{-m} \). Using (3.4), we see that

\[
\mathbb{P} \{ \text{osc}(B; \delta) > cr \sqrt{\delta \log(1/\delta)} \} \leq \mathbb{P} \left\{ M_n > \frac{c r^2}{3} 2^{-m} \log n \right\}.
\]

By Lemma 3.2.1, if \( c \) is chosen sufficiently large, the probability on the right-hand side is bounded by a constant times

\[
\exp \left\{ -c \left( \frac{2^{2m}}{18} \right) \log(1/\delta) \right\},
\]

which for \( c \) large enough is bounded by a constant times \( \delta^r \).

**Corollary 3.2.3.** With probability one, for every integer \( T < \infty \), the function \( t \mapsto B_t, t \in D \) is uniformly continuous on \([0, T]\).

**Proof.** Uniform continuity on \([0, T]\) is equivalent to saying that \( \text{osc}(B; 2^{m_n}, T) \to 0 \) as \( n \to \infty \). The previous theorem implies that there is a \( c_1 \) such that

\[
\mathbb{P} \{ \text{osc}(B; 2^{m_n}, T) > c_1 2^{-m_n/2} (\log n) \} \leq c_1 T n^{-1}.
\]

In particular,

\[
\sum_{n=1}^\infty \mathbb{P} \{ \text{osc}(B; 2^{m_n}, T) > c_1 2^{-m_n/2} (\log n) \} < \infty,
\]

which implies by Borel-Cantelli that with probability one \( \text{osc}(B; 2^{m_n}, T) \leq c_1 2^{-m_n/2} (\log n) \) for all \( n \) sufficiently large.

\[\square\]

Given the corollary, we can define \( B_t \) for \( t \notin D \) by continuity, i.e.,

\[
B_t = \lim_{t_n \to t} B_{t_n},
\]

where \( t_n \in D \) with \( t_n \to t \). It is not difficult to show that this satisfies the definition of Brownian motion (we omit the details). Moreover, since \( B_t \) has continuous paths, we can write

\[
\text{osc}(B; \delta, T) = \sup \{ |B_t - B_s| : 0 \leq s, t \leq T; |t - s| > \delta \}.
\]

We restate the estimate and include a fact about scaling of Brownian motion. Note that if \( B_t \) is a standard Brownian motion and \( T > 0 \), then \( Y_t := T^{-1/2} B_{Tt} \) is also a standard Brownian motion.

**Theorem 3.2.4 (Modulus of continuity of Brownian motion)** There is a \( c < \infty \) such that if \( B_t \) is a standard Brownian motion, \( 0 < \delta \leq 1, r \geq 1, T \geq 1 \),

\[
\mathbb{P} \{ \text{osc}(B; \delta) > cr \sqrt{\delta \log(1/\delta)} \} \leq c T \delta^r.
\]

Moreover, if \( T > 0 \), then \( \text{osc}(B; \delta, T) \) has the same distribution as \( T \text{osc}(B, \delta/T) \). In particular, if \( T = 1 \),

\[
\mathbb{P} \{ \text{osc}(B; 1, T) > cr \sqrt{T \log T} \} = \mathbb{P} \{ \text{osc}(B; 1) > cr \sqrt{\log T} \} \leq c T^{-r}.
\]

3.3 Skorokhod embedding

We will now define a procedure that takes a Brownian motion path \( B_t \) and produces a random walk \( S_n \). The idea is very straightforward. Start the Brownian motion and wait until it reaches \( +1 \) or \( -1 \). If it hits \( +1 \) first we let \( S_1 = 1 \); otherwise, we set \( S_1 = -1 \). Now we wait until the next increment of the Brownian motion reaches \( +1 \) or \( -1 \) and we use this value for the increment of the random walk.

To be more precise, let \( B_t \) be a standard one-dimensional Brownian motion, and let

\[\tau = \inf \{ t \geq 0 : |B_t| = 1 \} \]

Symmetry tells us that \( \mathbb{P} \{ \tau = 1 \} = \mathbb{P} \{ \tau = -1 \} = 1/2 \).

**Lemma 3.3.1** \( \mathbb{E}[\tau] = 1 \) and there exists a \( c < \infty \) such that \( \mathbb{E}[e^{\tau r}] < \infty \).

**Proof.** Note that for integer \( n \)

\[
\mathbb{P} \{ \tau > n \} \leq \mathbb{P} \{ \tau > n - 1, |B_n - B_{n-1}| \leq 2 \} = \mathbb{P} \{ \tau > n - 1 \} \mathbb{P} \{ |B_n - B_{n-1}| \leq 2 \}.
\]

50

**CHAPTER 3. APPROXIMATION BY BROWNIAN MOTION**
3.3. SKOROHOD EMBEDDING

which implies for integer \( n \),

\[
P\{r > n\} \leq P\{|B_n - B_{\tau_n}| \leq 2\} = e^{-\rho n},
\]

with \( \rho > 0 \). This implies that \( E[|B_t|] < \infty \) for \( b < \rho \). If \( b \leq t \), then \( E[B_t - t | F_t] = B_t^2 - s \) (Exercise 3.1). This shows that \( B_t^2 - t \) is a continuous martingale. Also,

\[
E[|B_t^2 - t|; r > t] \leq (t+1)P\{r > t\} \rightarrow 0.
\]

Therefore, we can use the optional sampling theorem (Theorem 8.2.8) to conclude that \( E[|B_t^2 - \tau|] = 0 \). Since \( E[|B_t^2|] = 1 \), this implies that \( E[\tau] = 1 \).

More generally, let \( n = 0 \) and

\[
\tau_n = \inf\{t \geq \tau_{n-1} : |B_t - B_{\tau_{n-1}}| = 1\}.
\]

Then \( S_n := B_{\tau_n} \) is a simple one-dimensional random walk. Moreover, if \( \tau_n = \tau_{n-1} \), the random variables \( T_1, T_2, \ldots \) are independent, identically distributed with mean one satisfying \( E[|B_t|] < \infty \) for some \( b > 0 \). As before, we define \( S_n \) for noninteger \( t \) by linear interpolation. Let

\[
\Theta(B;S;n) := \max\{|B_t - S_t| : 0 \leq t \leq n\},
\]

In other words, \( \Theta(B;S;n) \) is the distance between the continuous functions \( B \) and \( S \) in \( C[0,n] \) using the usual supremum norm. If \( j \leq t < j + 1 \), then

\[
|B_t - S_t| \leq |S_j - S_{j+1}| + |B_{j+1} - B_{j+1}| + |B_{j} - S_{j}| \leq 1 + \text{osc}(B;1,n) + |B_t - B_{\tau_n}|.
\]

Hence for integer \( n \),

\[
\Theta(B;S;n) \leq 1 + \text{osc}(B;1,n) + \max\{|B_t - B_{\tau_n}| : j = 1, \ldots, n\}.
\]

We can estimate the probabilities for the second term with (3.5). We will concentrate on the last term. Before we discuss the hard estimates, let us consider how large an error we should expect. Since \( T_1, T_2, \ldots \) are i.i.d. random variables with mean 1 and finite variance, the central limit theorem says roughly that

\[
|\tau_n - n| \approx \sqrt{n}.
\]

Hence we would expect that

\[
|B_n - B_{\tau_n}| \approx \sqrt{|\tau_n - n|} \approx n^{1/4}.
\]

From this reasoning, we can see that we expect \( \Theta(B;S;n) \) to be at least of order \( n^{1/4} \). The next theorem shows that it is unlikely that the actual value is much greater than \( n^{1/4} \).

---

CHAPTER 3. APPROXIMATION BY BROWNIAN MOTION

Theorem 3.3.2 There exist \( 0 < c, a < \infty \) such that for all \( r \leq n^{1/4} \) and all integers \( n \geq 3 \)

\[
P\{(\Theta(B;S;n)) > r \sqrt{n^{1/4}} \sqrt{\log n}\} \leq c e^{-ar}.
\]

Proof. It suffices to prove the theorem for \( r \geq 1 \). Suppose \( 1 \leq r \leq n^{1/4} \). Using (3.6), we see that the event \( \{\Theta(B;S;n) > r \sqrt{n^{1/4}} \sqrt{\log n}\} \) is contained in the union of the two events

\[
\bigcup_{j \leq n} \left\{ \text{osc}(B_t;\sqrt{n};2n) \geq (\sqrt{3}) n^{1/4} \sqrt{\log n} \right\}\text{ and }\bigcup_{j \leq n} \left\{ \text{osc}(B_t;n^{1/4}) \geq (\sqrt{3}) \sqrt{n^{1/2} \log(n^{1/2}/r)} \right\}.
\]

Theorem 3.2.4 gives for \( 1 \leq r \leq n^{1/4} \),

\[
P\{(\text{osc}(B_t;\sqrt{n};2n)) > (\sqrt{3}) n^{1/4} \sqrt{\log n}\} \leq 3P\{(\text{osc}(B_t;\sqrt{n};2n)) > (\sqrt{3}) n^{1/4} \sqrt{\log n}\} \leq 3P\left\{ \text{osc}(B_t;n^{1/4}) > (\sqrt{3}) \sqrt{n^{1/2} \log(n^{1/2}/r)} \right\} \leq ce^{-ar}.
\]

For the second event, consider the martingale

\[
M_j = \tau_j - j.
\]

Using (8.9) on \( M_j \) and \( -M_j \), we get

\[
P\left\{ \max_{1 \leq j \leq n} |\tau_j - j| \geq r \sqrt{n} \right\} \leq c e^{-ar},
\]

\[
(3.7)
\]

---

Extending the Skorokhod approximation to continuous time simple random walk \( \hat{S}_t \) is not difficult although in this case the path \( t \to \hat{S}_t \) is not continuous. Let \( N_t \) be a Poisson process with parameter \( 1 \) defined on the same probability space and independent of the Brownian motion \( B \). Then

\[
\hat{S}_t := S_{N_t}
\]

has the distribution of the continuous-time simple random walk. Since \( N_t - t \) is a martingale and the Poisson distribution has exponential moments, another application of (8.9) shows that for \( r \geq 1 \),

\[
P\left\{ \max_{1 \leq s \leq n} |N_s - s| \geq r \sqrt{n} \right\} \leq c e^{-ar}.
\]

Let

\[
\Theta(B;\hat{S};n) = \sup\{|B_t - \hat{S}_t| : 0 \leq t \leq n\}.
\]

Then the following is proved similarly.
3.4. HIGHER DIMENSIONS

Theorem 3.3.3 There exist \( 0 < c, a < \infty \) such that for all \( 1 \leq r \leq n^{1/4} \) and all positive integers \( n \)

\[
P(\Theta(B, S; n) \geq r n^{1/4} \sqrt{\log n}) \leq ce^{-aw}.
\]

3.4 Higher dimensions

It is not difficult to extend Theorems 3.3.2 and 3.3.3 to \( d > 1 \). A \( d \)-dimensional Brownian motion with covariance matrix \( \Gamma \) with respect to a filtration \( \mathcal{F}_t \) is a collection of random variables \( B_t, t \geq 0 \) satisfying the following:

(a) \( B_0 = 0 \);

(b) if \( s < t \), then \( B_t - B_s \) is an \( \mathcal{F}_s \)-measurable random \( \mathbb{R}^d \)-valued variable, independent of \( \mathcal{F}_s \), whose distribution is joint normal with mean zero and covariance matrix \( (t - s) \Gamma \).

(c) with probability one, \( t \mapsto B_t \) is a continuous function.

Lemma 3.4.1 Suppose \( B^{(1)}, \ldots, B^{(d)} \) are independent one-dimensional standard Brownian motions and \( v_1, \ldots, v_d \in \mathbb{R}^d \). Then

\[
B_t := B^{(1)}_t v_1 + \cdots + B^{(d)}_t v_d
\]

is a Brownian motion in \( \mathbb{R}^d \) with covariance matrix \( \Gamma = AA^T \) where \( A = [v_1 v_2 \ldots v_d] \).

Proof. Straightforward and left to the reader. \( \square \)

In particular, a standard \( d \)-dimensional Brownian motion is of the form

\[
B_t = (B^{(1)}_t, \ldots, B^{(d)}_t)
\]

where \( B^{(1)}, \ldots, B^{(d)} \) are independent one-dimensional Brownian motions. Its covariance matrix is the identity.

The next theorem shows that one can define a \( d \)-dimensional Brownian motion and a \( d \)-dimensional random walk on the same probability space so that their paths are close. Although the proof will use Skorokhod embedding, it is not true that the \( d \)-dimensional random walk is embedded into the \( d \)-dimensional Brownian motion. In fact, it is easy to see that it is impossible to have an embedded walk since if \( d > 1 \), with probability one a \( d \)-dimensional Brownian motion \( B_t \) visits the countable set \( \mathbb{Z}^d \) after time 0 is zero.

Theorem 3.4.2 Let \( p \in \mathcal{P} \) with covariance matrix \( \Gamma \). There exist \( c, a, \Gamma \) and a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) in which are defined a Brownian motion \( B \) with covariance matrix \( \Gamma \); a discrete-time random walk \( S \) with increment distribution \( p \); and a continuous-time random walk \( \tilde{S} \) with increment distribution \( p \) such that for all positive integers \( n \) and all \( 1 \leq r \leq n^{1/4} \),

\[
P(\Theta(B, S; n) \geq r n^{1/4} \sqrt{\log n}) \leq ce^{-aw},
\]

\[
P(\Theta(\tilde{B}, \tilde{S}; n) \geq r n^{1/4} \sqrt{\log n}) \leq ce^{-aw}.
\]

3.5 An alternative formulation

Here we give a slightly different, but equivalent, form of the strong approximation from which we get (32). We will illustrate this in the case of one-dimensional simple random walk. Suppose \( B \) is a standard Brownian motion defined on a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \). For positive integer \( n \), let \( B^{(n)}_t \) denote the Brownian motion

\[
B^{(n)}_t := n^{-1/2} B_{nt}.
\]

Let \( S^{(n)}_t \) denote the simple random walk derived from \( B^{(n)} \) using the Skorokhod embedding. Then we know that for positive integer \( T \),

\[
P \left( \max_{0 \leq t \leq T} |S^{(n)}_t - B^{(n)}_t| \geq cr \left( T n \right)^{1/4} \sqrt{\log(Tn)} \right) \leq ce^{-aw}.
\]

If we let

\[
W^{(n)} := n^{-1/2} S^{(n)}_T,
\]

then this becomes

\[
P \left( \max_{0 \leq t \leq T} |W^{(n)}_t - B^{(n)}_t| \geq cr \left( T n \right)^{1/4} \sqrt{\log(Tn)} \right) \leq ce^{-aw}.
\]

In particular, if \( r = c_1 \log n \) where \( c_1(T) \) is chosen sufficiently large,

\[
P \left( \max_{0 \leq t \leq T} |W^{(n)}_t - B^{(n)}_t| \geq c_1 n^{-1/4} \log^{3/2} n \right) \leq c_1 n^{-2}.
\]
By the Borel–Cantelli lemma, with probability one

\[
\max_{t \in [0,T]} |W_t^{(n)} - B_t| \geq c_n n^{-1/4} \log^{3/2} n
\]

for all \( n \) sufficiently large. In particular, with probability one \( W^{(n)} \) converges to \( B \) in the metric space \( C[0, T] \).

By using a multinomial process (in the discrete-time case) or a Poisson process (in the continuous-time case), we can prove the following.

**Theorem 3.5.1** Suppose \( p \in \mathcal{P}_d \) with covariance matrix \( \Gamma \). There exist \( c < \infty, a > 0 \) and a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) on which are defined \( d\)-dimensional Brownian motions \( B_t \) with covariance matrix \( \Gamma \); an infinite sequence of discrete-time \( p\)-walks \( S^{(1)}, S^{(2)}, \ldots \); and an infinite sequence of continuous time \( p\)-walks \( \tilde{S}^{(1)}, \tilde{S}^{(2)}, \ldots \), such that the following holds for every \( r > 0, T \geq 1 \). Let

\[
\begin{align*}
W_t^{(n)} &= n^{-1/2} S_t^{(n)}, \\
\tilde{W}_t^{(n)} &= n^{-1/2} \tilde{S}_t^{(n)}.
\end{align*}
\]

Then

\[
\begin{align*}
\mathbb{P} \left\{ \max_{t \in [0,T]} |W_t^{(n)} - B_t| \geq c r T^{1/4} n^{-1/4} \log^3(Tn) \right\} &\leq c e^{-an}, \\
\mathbb{P} \left\{ \max_{t \in [0,T]} |\tilde{W}_t^{(n)} - B_t| \geq c r T^{1/4} n^{-1/4} \log^3(Tn) \right\} &\leq c e^{-an}.
\end{align*}
\]

In particular, with probability one \( W^{(n)} \to B \) and \( \tilde{W}^{(n)} \to B \) in the metric space \( C[0, T] \).

**Exercises for Chapter 3**

**Exercise 3.1** Show that if \( B_t \) is a standard Brownian motion with respect to the filtration \( \mathcal{F}_t \) and \( s < t \), then \( \mathbb{E} B_t^2 - t | \mathcal{F}_s = B_s^2 - s \).

**Exercise 3.2** Let \( X \) be an integer-valued random variable with \( \mathbb{P} \{ X = 0 \} = 0 \) and \( \mathbb{E} |X| = 0 \).

(a) Show that there exists numbers \( r_2 \in (0, \infty) \)

\[
r_1 \leq r_2 \leq \cdots \leq r_n \leq r_{n+1} \leq \infty,
\]

such that if \( B_t \) is a standard Brownian motion and

\[
T = \inf \{ t ; B_t \in \mathbb{Z}, t \leq r_n \},
\]

then \( B_T \) has the same distribution as \( X \).

(b) Show that if \( X \) has bounded support, then there exists \( a > 0 \) with \( \mathbb{E} e^{aT} < \infty \).

(c) Show that \( \mathbb{E} T = \mathbb{E} |X^2| \).
Chapter 4

Green’s function

4.1 Recurrence and transience

A random walk \( S_n \) with increment distribution \( p \in \mathcal{P}_d^* \) is called recurrent if \( \mathbb{P}\{S_n = 0 \text{ i.o.}\} = 1 \). If the walk is not recurrent it is called transient. We will also say that \( p \) is recurrent (or transient). It is easy to see using the Markov property that \( p \) is recurrent if and only if for each \( x \in \mathbb{Z}^d \),
\[
\mathbb{P}\{S_n = 0 \text{ for some } n \geq 1\} = 1.
\]

**Theorem 4.1.1** If \( p \in \mathcal{P}_d^* \) with \( d = 1, 2 \), then \( p \) is recurrent. If \( p \in \mathcal{P}_d^* \) with \( d \geq 3 \), then \( p \) is transient. In the latter case,
\[
\mathbb{P}\{S_n \neq 0 \text{ for all } n \geq 1\} = \left[ \sum_{n=0}^{\infty} p_n(0) \right]^{-1}.
\]

**Proof.** Let \( Y = \sum_{n=0}^{\infty} 1\{S_n = 0\} \) denote the number of visits to the origin and note that
\[
\mathbb{E}[Y] = \sum_{n=0}^{\infty} \mathbb{P}\{S_n = 0\} = \sum_{n=0}^{\infty} p_n(0),
\]
If \( p \in \mathcal{P}_d^* \) with \( d = 1, 2 \), the LCLT (see Theorem 2.1.1 and Exercise 2.3) implies that \( p_n(0) \sim c n^{-1/2} \) and the sum is infinite. If \( p \in \mathcal{P}_d^* \) with \( d \geq 3 \), then (2.40) shows that \( p_n(0) \leq c n^{-d/2} \) and hence \( \mathbb{E}[Y] < \infty \).

We will compute \( \mathbb{E}[Y] \) in a different way. Let \( q = \mathbb{P}\{S_n \neq 0 \text{ for all } n \geq 1\} \). Then \( \mathbb{P}\{Y = 1\} = q \), and by the Markov property, \( \mathbb{P}\{Y = j\} = (1 - q)^{j-1} q \). Therefore, if \( q > 0 \),
\[
\mathbb{E}[Y] = \sum_{j=0}^{\infty} j \mathbb{P}\{Y = j\} = \sum_{j=0}^{\infty} j (1 - q)^{j-1} q = \frac{1}{q}.
\]

The quantity \( q = \mathbb{E}[Y]^{-1} \) is often called the escape probability for the random walk.

---

### 4.2 Green’s generating function

If \( p \in \mathcal{P}^* \) and \( x, y \in \mathbb{Z}^d \), we define the **Green’s generating function** to be the power series in \( \xi \):
\[
G(x, y; \xi) = \sum_{n=0}^{\infty} \xi^n \mathbb{P}\{S_n = y - x\}.
\]
Note that the sum is absolutely convergent for \( |\xi| < 1 \). We write just \( G(y; \xi) \) for \( G(0, y; \xi) \).

If \( p \in \mathcal{P} \), then \( G(x; \xi) = G(-x; \xi) \).

The generating function is defined for complex \( \xi \), but there is a very nice interpretation of the sum for positive \( \xi \leq 1 \). Suppose \( T \) is a random variable independent of the random walk \( S \) with a geometric distribution.
\[
\mathbb{P}\{T = j\} = \xi^{j-1} (1 - \xi), \quad j = 1, 2, \ldots,
\]
i.e., \( \mathbb{P}\{T > j\} = \xi^j \). We think of \( T \) as a “killing time” for the walk. At each time \( j \), if the walker has not already been killed, the process is killed with probability \( 1 - \xi \), where the killing is independent of the walk. If the random walk starts at the origin, then the expected number of visits to \( x \) before being killed is given by
\[
\mathbb{E}\left[ \sum_{j \in \mathbb{T}} 1\{S_j = x\} \right] = \mathbb{E}\left[ \sum_{j \geq 1} 1\{S_j = x; T > j\} \right] = \sum_{j \geq 0} \mathbb{P}\{S_j = x; T > j\} = \sum_{j \geq 0} p_j(x) \xi^j = G(x; \xi).
\]

Theorem 4.1.1 can be restated that a random walk is transient if and only if \( G(x; 1) < \infty \), in which case the escape probability is \( G(x, 1)^{-1} \). For a transient random walk, we define the **Green’s function** to be
\[
G(x; y) = G(x, y; 1) = \sum_{n=0}^{\infty} p_n(y - x),
\]
We write \( G(x) = G(0, x) \); if \( p \in \mathcal{P} \), then \( G(x) = G(-x) \). The strong Markov property implies that
\[
G(0, x) = \mathbb{P}\{S_n = x \text{ for some } n \geq 0\} G(0, 0).
\]

Similarly, we can define
\[
\tilde{G}(x, y; \xi) = \int_0^\xi \xi^t p(x, y) \, dt.
\]
For \( \xi \in (0, 1) \) this is the expected amount of time spent at site \( y \) by a continuous-time random walk with increment distribution \( p \) before an independent “killing time” that has an exponential distribution with rate \( -\log(1 - \xi) \). We will now show that if we set \( \xi = 1 \), we get the same Green’s function as the discrete walk, i.e., \( \tilde{G}(x, y; 1) = G(x, y) \).
4.2. **GREEN'S GENERATING FUNCTION**

**Proposition 4.2.1** If \( p \in \mathcal{P}^*_1 \) is transient, then

\[
\int_0^\infty \tilde{\kappa}(x) \, dt = G(x).
\]

**Proof.** Let \( S_n \), denote a discrete-time walk with distribution \( p \); \( N_t \) an independent Poisson process with parameter 1; and let \( \hat{S}_t \) denote the continuous time walk \( \hat{S}_t = S_{N_t} \). Let

\[
Y_x = \sum_{n=0}^\infty 1\{S_n = x\}, \quad \hat{Y}_x = \int_0^\infty 1\{\hat{S}_t = x\} \, dt,
\]

denote the expected amount of time spent at \( x \) by \( S \) and \( \hat{S} \), respectively. Then \( G(x) = \mathbb{E}[Y_x] \). If we let \( T_n = \inf\{t : N_t = n\} \), then we can write

\[
\hat{Y}_x = \sum_{n=0}^\infty 1\{S_n = x\} (T_{n+1} - T_n).
\]

However, independence of \( S \) and \( N \) implies

\[
\mathbb{E}[1\{S_n = x\} (T_{n+1} - T_n)] = \mathbb{P}\{S_n = x\} \mathbb{E}[T_{n+1} - T_n] = \mathbb{P}\{S_n = x\}.
\]

Hence \( \mathbb{E}[\hat{Y}_x] = \mathbb{E}[Y_x] \). \( \Box \)

**Remark.** Suppose \( p \) is the increment distribution of a random walk in \( \mathbb{Z}^d \). For \( \epsilon > 0 \), let \( p_\epsilon \) denote the increment of the "lazy walker" given by

\[
p_\epsilon(x) = \begin{cases} \frac{1-\epsilon}{\epsilon} p(x), & x \neq \theta \\ \epsilon + (1-\epsilon) p(0), & x = \theta \end{cases}
\]

If \( p \) is irreducible but perhaps periodic on \( \mathbb{Z}^d \), then for each \( \epsilon > 0 \), \( p_\epsilon \) is irreducible and aperiodic. Let \( \mathcal{L}, \phi \) denote the generator and characteristic function for \( p \), respectively, then the generator and characteristic function for \( p_\epsilon \) are

\[
\mathcal{L}_\epsilon = (1-\epsilon) \mathcal{L}, \quad \phi(\theta) = \epsilon + (1-\epsilon) \phi(\theta).
\]  

(4.2)

If \( p \) is mean zero and has covariance matrix \( \Gamma \), then \( p_\epsilon \) is mean zero with covariance matrix

\[
\Gamma_\epsilon = (1-\epsilon) \Gamma, \quad \det \Gamma_\epsilon = (1-\epsilon)^d \det \Gamma.
\]  

(4.3)

If \( p \) is transient, let \( G, G_\epsilon \) denote the Green's function for \( p, p_\epsilon \), respectively, then similarly to the last proposition we can see that

\[
G_\epsilon(x) = \frac{1}{1-\epsilon} G(x).
\]  

(4.4)

In a number of our proofs in this book we will assume that the walk is aperiodic; results for periodic walks can then be derived using these relations.

60  **CHAPTER 4. GREEN'S FUNCTION**

If \( n \geq 1 \), let \( f_n(x; y) \) denote the probability that a random walk starting at \( x \) first visits \( y \) at time \( n \) (not counting time \( n = 0 \)), i.e.,

\[
f_n(x, y) = \mathbb{P}^*\{S_n = y; S_1 \neq y, \ldots, S_{n-1} \neq y\} = \mathbb{P}^*\{T_n = n\},
\]

where

\[
\tau_n = \min\{j \geq 1 : S_j = y\}, \quad \tau_n = \min\{j \geq 0 : S_j = y\}.
\]

Let \( f_n(x) = f_n(0, x) \) and note that

\[
\mathbb{P}^*\{\tau_n < \infty\} = \sum_{n=1}^\infty f_n(x, y) = \sum_{n=1}^\infty f_n(y-x) \leq 1.
\]

Define the first visit generating function by

\[
F(x, y; \xi) = F(y-x; \xi) = \sum_{n=1}^\infty \xi^n f(n-x; \xi).
\]

If \( \xi \in \{0,1\} \), then

\[
F(x, y; \xi) = \mathbb{P}^*\{\tau_n < T\xi\},
\]

where \( T\xi \) denotes an independent geometric random variable satisfying \( \mathbb{P}\{T\xi > n\} = \xi^n \).

**Proposition 4.2.2** If \( n \geq 1 \),

\[
p_n(y) = \sum_{j=1}^n f_j(y) p_{n-j}(0).
\]

If \( \xi \in \mathbb{C} \),

\[
G(y; \xi) = \delta(y) + F(y; \xi) G(0; \xi).
\]  

(4.5)

In particular, if \( |F(0, \xi)| < 1 \),

\[
G(0; \xi) = \frac{1}{1 - F(0, \xi)}.
\]  

(4.6)

**Proof.** The first equality follows from

\[
\mathbb{P}^*\{S_n = y\} = \sum_{j=1}^n \mathbb{P}\{T_j = j; S_n - S_j = 0\}.
\]

The second follows from the equality

\[
\sum_{n=1}^\infty p_n(x) \xi^n = \left[ \sum_{n=1}^\infty f_n(x) \xi^n \right] \left[ \sum_{n=0}^\infty p_n(0) \xi^n \right].
\]
4.2. GREEN’S GENERATING FUNCTION

which follows from the first equality. For \( \xi \in (0, 1) \), there is a probabilistic interpretation of (4.5); if \( y \neq 0 \), the expected number of visits to \( y \) is the probability of reaching \( y \) times the expected number of visits to \( y \) given that we reach \( y \). If \( y = 0 \), we have to add an extra 1 to account for \( p_0(y) \).

\[ \square \]

Remark. If \( \xi \in (0, 1) \), the identity (4.6) can be considered as a generalization of (4.1). Note that

\[ F(0; \xi) = \sum_{j=1}^{\infty} P\{\eta_j = j; \tau_\xi > j\} = P\{\eta_0 < \tau_\xi\} \]

represents the probability that a random walk killed at rate \( 1 - \xi \) returns to the origin before being killed. Hence, the probability of no return before killing is

\[ 1 - F(0; \xi) = G(0; \xi)^{-1}. \quad (4.7) \]

The right-hand side is the expected number of visits before killing. If \( p \) is transient, we can plug \( \xi = 1 \) into this expression and get (4.1).

**Proposition 4.2.3** Suppose \( p \) is an increment distribution of a random walk in \( \mathbb{Z}^d \) with characteristic function \( \phi \). Then if \( z \in \mathbb{Z}^d, |z| < 1 \),

\[ G(z; \xi) = \frac{1}{(2\pi)^d} \int_{|\theta| < \xi} \frac{1}{1 - \xi \phi(\theta)} e^{-iz\cdot \theta} d\theta, \]

If \( p \in \mathcal{P}_u \cup \mathcal{P}_d \) with \( d \geq 3 \), this holds for \( \xi = 1, i.e., \)

\[ G(z) = \frac{1}{(2\pi)^d} \int_{|\theta| < 1} \frac{1}{1 - \phi(\theta)} e^{-iz\cdot \theta} d\theta. \]

**Proof.** All of the integrals in this proof will be over \([-\pi, \pi]^d\). The formal calculation is

\[ G(z; \xi) = \sum_{n=0}^{\infty} \xi^n p_n(z) = \sum_{n=0}^{\infty} \xi^n \frac{1}{(2\pi)^d} \int \phi(\theta)^n e^{-iz\cdot \theta} d\theta \]

\[ = \frac{1}{(2\pi)^d} \int \left[ \sum_{n=0}^{\infty} (\xi \phi(\theta))^n \right] e^{-iz\cdot \theta} d\theta \]

\[ = \frac{1}{(2\pi)^d} \int \frac{1}{1 - \xi \phi(\theta)} e^{-iz\cdot \theta} d\theta. \]

The interchange of the sum and the integral is justified by the dominated convergence theorem as we now describe. For each \( N \),

\[ \left| \sum_{n=0}^{N} \xi^n \phi(\theta)^n e^{-iz\cdot \theta} \right| \leq \frac{1}{1 - K|\phi(\theta)|} \]

If \( |K| < 1 \), then the right-hand side is bounded by \( 1/(1 - |K|) \). If \( p \in \mathcal{P}_u \), then (2.28) shows that the right-hand side is bounded by \( C|\phi^2| \) for some \( C \). If \( d \geq 3 \), \( |\phi^2| \) is integrable on \([-\pi, \pi]^d \). If \( p \in \mathcal{P}_d \) is bipartite, we can use (4.2) and (4.4).

Sometimes it is easier to prove results about geometrically killed random walks than walks restricted to a fixed number of steps. This is because stopping time arguments work better. Suppose that \( S_t \) is a random walk; \( \tau \) is a stopping time for the random walk; and \( T \) is an independent geometric random variable. Then on the event \( \{T > \tau\} \) the distribution of \( T - \tau \) given \( S_\tau \) is the same as that of \( T \). This “loss of memory” property for geometric and exponential random variables can be very useful. The next proposition gives an example of a result that is easier to prove for geometrically killed random walks. The result for fixed length random walks can be deduced from the geometrically killed walk result by using Tauberian theorems. Tauberian theorems are the mathematical tool for deducing facts about a sequence from its generating functions. We will only use some simple Tauberian theorems; see Section 8.5.

**Proposition 4.2.4** Suppose \( p \in \mathcal{P}_u \cup \mathcal{P}_d \), \( d = 1, 2 \), Let

\[ q(n) = P\{S_j \neq 0 : j = 1, \ldots, n\}. \]

Then as \( n \to \infty \),

\[ q(n) \sim \begin{cases} \frac{1}{\log n} & d = 1 \smallskip \\ \\ \frac{1}{\log n^{1/2}} & d = 2 \end{cases} \]

where \( r = (2\pi)^{d/2} \sqrt{\det \Gamma} \).

**Proof.** We will assume \( p \in \mathcal{P}_d \); it is not difficult to extend this to bipartite \( p \in \mathcal{P}_d \). We will prove the corresponding facts about the generating functions for \( q(n) \):

\[ \sum_{n=0}^{\infty} \xi^n q(n) \sim \frac{r}{\Gamma(1/2)} \sqrt{1 - \xi}, \quad d = 1, \quad \sum_{n=0}^{\infty} \xi^n q(n) \sim \frac{1}{\xi} \log \left( \frac{1}{1 - \xi} \right)^{1/2}, \quad d = 2. \]

Since the sequence \( q(n) \) is monotone in \( n \), Propositions 8.5.2 and 8.5.3 show that this implies the proposition (recall that \( \Gamma(1/2) = \sqrt{\pi} \)).

Let \( T \) be a geometric random variable with killing rate \( 1 - \xi \). Then (4.7) tells us that

\[ P\{S_j \neq 0 : j = 1, \ldots, T - 1\} = G(0; \xi)^{-1}. \]

Also,

\[ P\{S_j \neq 0 : j = 1, \ldots, T - 1\} = \sum_{n=0}^{\infty} P\{T = n + 1\} q(n) = (1 - \xi) \sum_{n=0}^{\infty} \xi^n q(n). \]
4.3. GREEN'S FUNCTION, TRANSIENT CASE

Using (2.41), we can write

$$G(0; \xi) = \sum_{n=0}^{\infty} \xi^n p_n(0) = \sum_{n=0}^{\infty} \xi^n \left[ \frac{1}{\sigma^2 n^{d/2}} + o \left( \frac{1}{n^{d/2}} \right) \right] \sim \frac{1}{\sigma} F \left( \frac{1}{1 - \xi} \right),$$

where

$$F(s) = \begin{cases} \Gamma(1/2) \sqrt{\pi}, & d = 1 \\ \log s, & d = 2 \end{cases}$$

This gives (4.8) and (4.9).

Corollary 4.2.5 Suppose $S_n$ is a random walk with increment distribution $p \in \mathcal{P}_d$ and let $T = \min(j \geq 1 : S_j = 0)$.

Then $\mathbb{E}[T] = \infty$.

Proof. If $d \geq 3$, then transience implies that $\mathbb{P}\{T = \infty\} > 0$. For $d = 1, 2$, the results comes from the previous proposition that tells us

$$\mathbb{P}\{T > n\} \geq \begin{cases} C_n^{-1/2}, & d = 1 \\ C_n |\log n|^{-1}, & d = 2 \end{cases}.$$

4.3 Green’s function, transient case

In this section, we will assume that $d \geq 3$ and $p \in \mathcal{P}_d$. The Green’s function $G(x, y) = G(y - x)$ is given by

$$G(x) = \sum_{n=0}^{\infty} \mathbb{P}\{S_n = x\} = \mathbb{E} \left[ \sum_{n=0}^{\infty} 1\{S_n = x\} \right].$$

Note that

$$G(x) = 1\{x = 0\} + \sum_y p(x, y) \mathbb{E} \left[ \sum_{n=0}^{\infty} 1\{S_n = 0\} \right] = \delta(x) + \sum_y p(x, y) G(y),$$

where $\delta(\cdot)$ denotes the delta function, $\delta(x) = 1\{x = 0\}$. In other words,

$$G(x) = -\delta(x) = \begin{cases} -1, & x = 0 \\ 0, & x \neq 0 \end{cases}.$$

A simple application of the Markov property gives

$$G(x) = \mathbb{P}\{S < \infty\} G(0).$$

We will now give the asymptotics of the Green’s function as $|x| \to \infty$. Recall that

$$\mathcal{J}(x)^2 = \mathcal{J}(x)^2 = x \cdot \mathcal{J}^2 x; \text{ in particular, } \mathcal{J}(x) \equiv \mathcal{J}(x) \equiv |x|.\)
Choosing \( f(t) = e^{-a t^d} \), we get
\[
\left| n^{-d} e^{-a n} - \int_{n^{-1/2}}^{n^{1/2}} n^{-d} e^{-a t^d} dt \right| \leq \frac{c}{n^{d+1/2}} \left( 1 + \frac{q^2}{n} \right) e^{-a n}, \quad n \geq \sqrt{a}.
\]
(The restriction \( n \geq \sqrt{a} \) is used to guarantee that \( e^{-a t^d}/n \leq e^{-a n} \).) Therefore,
\[
\sum_{n \geq \sqrt{a}} n^{-d} e^{-a n} - \int_{n^{-1/2}}^{n^{1/2}} n^{-d} e^{-a t^d} dt \leq c \sum_{n \geq \sqrt{a}} \frac{n}{n^{d+1/2}} \left( 1 + \frac{q^2}{n} \right) e^{-a n} \leq c \int_{a}^{\infty} \sqrt{t} e^{-a t^d} dt \leq c e^{-a d}.
\]
The last step uses (4.11). It is easy to check that the sum over \( n < \sqrt{a} \) is decays faster than any power of \( a \), \( \Box \)

**Proof of Theorem 4.3.1.** Using Lemma 4.3.2, we have
\[
\sum_{n=1}^{\infty} p_n(x) = \sum_{n=1}^{\infty} \left( \frac{1}{2\pi n^{d+1/2}} \frac{\Gamma(d/2)}{\sqrt{\Gamma(d+1)}} \right) e^{-x^2/(2n)} \leq \frac{\Gamma(d/2)}{\sqrt{\Gamma(d+1)}} \frac{1}{2\pi n^{d+1/2}} x^d + O \left( \frac{1}{|x|^{d+1}} \right).
\]
Hence we only need to prove (4.10). Fix \( \epsilon \) with \( 0 < \epsilon < 4/(d+2) \). A simple estimate shows that
\[
\sum_{n \in \mathbb{F}^+} p_n(x)
\]
as a function of \( x \) decays faster than any power of \( x \). Similarly, using Proposition 2.1.2,
\[
\sum_{n \in \mathbb{F}^-} p_n(x) = o(|x|^{-\epsilon}). \tag{4.12}
\]
For \( n \geq |x|^{d+1} \) we use Corollary 2.3.9 which gives
\[
|p_n(x) - \tilde{p}_n(x)| \leq c \left( \frac{|x|}{\sqrt{n}} \right)^d \left( \frac{1}{n^{d+1/2}} + \frac{1}{n^{d+1/2}} \right), \tag{4.13}
\]
Note that
\[
\sum_{n \in \mathbb{F}^-} n^{-d-1/2} = O \left( |x|^{-d-1/2} \right) = o \left( |x|^{-d} \right),
\]
and
\[
\sum_{n \in \mathbb{F}^-} \left( \frac{|x|}{\sqrt{n}} \right)^d \frac{1}{n^{d+1/2}} \leq c \int_{a}^{\infty} \frac{x^d}{t^{d+1/2}} e^{-a t^d} dt \leq c |x|^{-d},
\]

**Remark.** The error term in this theorem is very small. In order to prove that it is this small we used the sharp estimate (4.13) which uses the fact that the third moments of the increment distribution are zero. If \( p \in \mathbb{P}_d \) with bounded increments but with nonzero third moments, a similar asymptotic expansion exists for about the Green’s function except that the error term is \( O(|x|^{-d-1}) \). We have used bounded increments (or at least the existence of an exponential moment) in an important way in (4.12), the asymptotics could be proved under weaker moment assumptions; however, mean zero, finite variance is not sufficient to conclude that the Green’s function is asymptotic to \( e^{-x^2/(2n)} \) for \( d \geq 4 \). See Exercise 4.4.

**Heuristic note.** Often one does not the full force of these asymptotics. An important thing to remember is that \( \delta(x) = |x|^{-d} \). There are a number of ways to remember the exponent \( 2 - d \). For example, the central limit theorem implies that the random walk should visit order \( R^2 \) points in the ball of radius \( R \). Since there are \( R^d \) points in this ball, the probability that a particular point is visited is of order \( R^{-d} \). In the case of standard \( d \)-dimensional Brownian motion, the Green’s function is proportional to \( |x|^{-d} \). This is the unique (up to multiplicative constant) harmonic, radially symmetric function on \( \mathbb{R}^d \setminus \{0\} \) that goes to zero as \( |x| \to \infty \) (see Exercise 4.9).

**Corollary 4.3.3.** If \( p \in \mathbb{P}_d \), then
\[
\nabla_x G(x) = \frac{C_d}{|x|^{d+1}} + o(|x|^{-d}).
\]
In particular, \( \nabla_x G(x) = O(|x|^{-d+1}) \). Also,
\[
\nabla_x G(x) = O(|x|^{-d}).
\]

**Remark.** We could also prove this corollary with improved error terms by using the difference estimates for the LCLT (see, e.g., Theorem 2.3.6) but we will not need the sharper results in this book. If \( p \in \mathbb{P}_d \) with bounded increments but nonzero third moments, we could prove the corollary but would need to use the LCLT difference estimates.

## 4.4 POTENTIAL KERNEL

### 4.4.1 Two dimensions

If \( p \in \mathbb{P}_2 \), the potential kernel is the function
\[
o(x) = \sum_{n=0}^{\infty} |p_n(0) - p_n(x)| = \lim_{N \to \infty} \sum_{n=0}^{N} p_n(0) - \sum_{n=0}^{N} p_n(x), \tag{4.14}
\]
Exercise 2.5 shows that $|p_n(0) - p_n(x)| \leq c |x|^n \exp(-|x|^2)$, so the first sum converges absolutely. However, since $p_n(0) \to n^\infty$, we cannot write

$$a(x) = \left[ \sum_{n=0}^{\infty} p_n(0) \right] - \left[ \sum_{n=0}^{\infty} p_n(x) \right].$$

If $p \in \mathcal{P}_2$ is bipartite, the potential kernel for $x \in \mathbb{Z}^d$ is defined in the same way. If $x \in \mathbb{Z}^d$, we can define $a(x)$ by the second expression in (4.14). Many authors use the term Green's function for the potential kernel.

**Proposition 4.4.1** If $p \in \mathcal{P}_2$, then $2a(x)$ is the expected number of visits to $x$ by a random walk starting at $x$ before its first visit to the origin.

**Proof.** We delay this until the next section; see (4.23).

**Remark.** Using Proposition 4.4.1, we can see that if $p_1$ is defined as in (4.2), and $a_1$ denotes the potential kernel for $p_1$, then

$$a_1(x) = \frac{1}{1-x} a(x).$$

(4.15)

This will allow us to assume aperiodicity in many of our proofs.

If we write

$$p_{n+1}(x) = \sum_{y \neq 0} p_n(x - y) p(y),$$

we see that

$$a(x) = \sum_{y \neq 0} a(x - y) p(y).$$

More generally, if $p \in \mathcal{P}_2$, we can see that $a(x)$ satisfies: $a(0) = 0$ and

$$\mathcal{L}a(x) = \delta_0(x) = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0. \end{cases}$$

**Proposition 4.4.2** If $p \in \mathcal{P}_2$, then

$$a(x) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1 - e^{-ix\theta}}{1 + \phi(\theta)} d\theta.$$

**Proof.** By the remark above, it suffices to consider $p \in \mathcal{P}_2$. The formal calculation is

$$a(x) = \sum_{n=0}^{\infty} [p_n(0) - p_n(x)] = \sum_{n=0}^{\infty} \left[ \frac{1}{(2\pi)^2} \int_{\mathbb{R}^d} \phi(\theta) \phi^n(\theta) [1 - e^{-ix\theta}] d\theta \right]$$

$$= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^d} \phi^0(\theta) \phi^n(\theta) [1 - e^{-ix\theta}] d\theta$$

$$= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^d} \frac{1 - e^{-ix\theta}}{1 + \phi(\theta)} d\theta,$$

All of the integrals are over $[-\pi, \pi]^2$. To justify the interchange of sum and integral we use (2.28) to obtain the estimate

$$\left| \sum_{n=0}^{\infty} \phi(\theta)^n [1 - e^{-ix\theta}] \right| \leq \frac{||1 - e^{-ix\theta}/1 - \phi(\theta)||^2}{1 - \phi(\theta)} \leq \frac{1}{1 - \phi(\theta)} \leq \frac{1}{1 - \phi(\theta)}.$$

Since $|\theta|^2$ is an integral function in $[-\pi, \pi]^2$, the dominated convergence theorem may be applied.

**Theorem 4.4.3** If $p \in \mathcal{P}_2$, there exists a constant $C = C(p)$ such that $|x| \to \infty$,

$$a(x) = \frac{1}{\pi \sqrt{\det \Gamma}} \log |J^*(x)| + C + O(|x|^2),$$

For simple random walk,

$$a(x) = \frac{2}{\pi} \log |x| + \frac{2\gamma + \log 8}{\pi} + O(|x|^2),$$

where $\gamma$ is Euler's constant.

**Proof.** We will assume that $p$ is aperiodic; the bipartite case is done similarly. We write

$$a(x) = \sum_{n \leq J^*(x)^2} p_n(0) - \sum_{n \leq J^*(x)^2} p_n(x) + \sum_{n > J^*(x)^2} \left[ p_n(0) - p_n(x) \right].$$

We know from (2.30) that

$$p_n(0) = \frac{1}{2\pi \sqrt{\det \Gamma}} \frac{1}{n} + O \left( \frac{1}{n^2} \right).$$

We therefore get

$$\sum_{n \leq J^*(x)^2} p_n(0) = 1 + O(|x|^2) + \sum_{1 \leq n \leq J^*(x)^2} \frac{1}{2\pi \sqrt{\det \Gamma}} \frac{1}{n} + \sum_{n > J^*(x)^2} p_n(0) - \frac{1}{2\pi \sqrt{\det \Gamma}} \frac{1}{n},$$

The last sum is just a constant and

$$\sum_{1 \leq n \leq J^*(x)^2} \frac{1}{n} = 2 \log |J^*(x)| + \gamma + O(|x|^2),$$

where $\gamma$ is Euler's constant (see Lemma 8.1.2). Hence,

$$\sum_{n \leq J^*(x)^2} p_n(0) = \frac{1}{\pi \sqrt{\det \Gamma}} \log |J^*(x)| + d + O(|x|^2)$$

and

$$\sum_{n > J^*(x)^2} p_n(0) = \frac{1}{\pi \sqrt{\det \Gamma}} \log |J^*(x)| + d + O(|x|^2)$$

where $d$ is Euler's constant.
4.4. POTENTIAL KERNEL

for some constant $c$.

Proposition 2.1.2 shows that

$$\sum_{n \leq |x|} p_n(x)$$

decays faster than any power of $|x|$. Corollary 2.3.9 gives

$$\left| p_n(x) - \frac{1}{2\pi n \sqrt{\det I}} e^{-\|x\|^2/2n} \right| \leq c \left[ \frac{e^{-|x|^2/4n}}{n^2} + \frac{1}{n^3} \right].$$

Therefore,

$$\sum_{n \leq \mathcal{J}^*(x)^2} \left| p_n(x) - \frac{1}{2\pi n \sqrt{\det I}} e^{-\|x\|^2/2n} \right| \leq o(|x|^{-2}) + c \sum_{k \leq \mathcal{J}^*(x)^2} \left[ \frac{e^{-|x|^2/2k}}{k^2} + \frac{1}{k^3} \right] \leq c |x|^{-2}.$$

The last estimate is done as in the proof of Theorem 4.3.1. Similarly, to the proof of Lemma 4.3.2 we can see that

$$\sum_{k \leq \mathcal{J}^*(x)^2} \left[ p_n(0) - p_n(x) \right] = \frac{1}{\pi \sqrt{\det I}} \log|\mathcal{J}^*(x)| + C' + O(|x|^{-2}).$$

for some constant $C'$. For $n > \mathcal{J}^*(x)^2$, we use Corollary 2.3.9 and Lemma 4.3.2 again to conclude that

$$\sum_{n > \mathcal{J}^*(x)^2} \left[ p_n(0) - p_n(x) \right] = \int_{\mathcal{J}^*(x)^2} \frac{1}{y} \left[ 1 - e^{-|x|^2/y} \right] dy + O(|x|^{-2}).$$

For simple random walk in two dimensions, it follows that

$$a(x) = \frac{2}{\pi} \log |x| + k + O(|x|^{-2}).$$

for some constant $k$. To determine $k$, we use

$$\phi(\theta^1, \theta^2) = 1 - \frac{1}{2} [\cos \theta^1 + \cos \theta^2].$$

Plugging this into Proposition 4.4.2 and doing the integral (details omitted), we get an exact expression for $a_n = (n, n)$ for integer $n > 0$

$$a_n = \frac{2}{\pi} \log n + \frac{2}{\pi} \log \sqrt{2} + k + O(n^{-2}).$$

Therefore,

$$k = \lim_{n \to \infty} \left[ -\frac{1}{\pi} \log 2 - \frac{2}{\pi} \log n + \frac{4}{\pi} \sum_{j=1}^{n} \frac{1}{2j} - 1 \right].$$

Using Lemma 8.1.2 we can see that as $n \to \infty$,

$$\sum_{j=1}^{n} \frac{1}{2j-1} = \sum_{j=1}^{n} \left[ 1 - \sum_{j=1}^{n} \frac{1}{2j} \right] = \frac{1}{2} \log n + \log 2 + \frac{1}{2} \gamma + o(1),$$

Therefore,

$$k = \frac{3}{2} \log 2 + \frac{3}{2} \gamma.$$

\[\square\]

Remark. Although we have included the exact value of the constant for simple random walk, we will never need to use this value.

**Corollary 4.4.4** If $p \in \mathcal{P}_2$,

$$\nabla_j a(x) = \nabla_j \left[ \frac{1}{\pi \sqrt{\det I}} \log|\mathcal{J}^*(x)| + O(|x|^{-2}) \right].$$

In particular, $\nabla_j a(x) = O(|x|^{-1})$. Also,

$$\nabla_j^2 a(x) = O(|x|^{-2}).$$

4.4.2 One dimension

If $p \in \mathcal{P}_1$, the potential kernel is defined in the same way

$$a(x) = \lim_{N \to \infty} \left[ \sum_{n=0}^{N} p_n(0) - \sum_{n=0}^{N} p_n(x) \right].$$
4.4. POTENTIAL KERNEL

In this case, the convergence is a little more subtle. We will restrict ourselves to the third moment case, \( \mathbb{E}[|X|^3] < \infty \). In this case, it was shown in Exercise 2.7 that \( p_n(0) = p_n(x) = O(n^{-\gamma/2}) \) and hence the sum converges absolutely.

\[
a(x) = \sum_{n=0}^{\infty} |p_n(0) - p_n(x)|.
\]

**Proposition 4.4.5** Suppose \( p \in \mathcal{P}_1 \) with \( \mathbb{E}[|X|^2] < \infty \). Then there is a \( c \) such that for all \( x \),

\[
|a(x)| \leq c \log |x|.
\]

If \( \mathbb{E}[|X|^2] < \infty \) and \( \mathbb{E}[X^2] = 0 \), then there is a \( c \) such that for all \( x \),

\[
|a(x)| \leq c.
\]

**Proof.** The estimate \( |p_n(0) - p_n(x)| = O(n^{-\gamma/2}) \) shows that

\[
\left| \sum_{n=0}^{x^*} [p_n(0) - p_n(x)] \right| \leq c |x| \sum_{n=0}^{x^*} n^{-\gamma/2} \leq c |x|^2.
\]

Therefore,

\[
a(x) = O(|x|^2) + \sum_{n=0}^{x^*} |p_n(0) - p_n(x)|.
\]

Using (2.29) and (2.30), we get

\[
p_n(y) = \frac{1}{\sqrt{2\pi \sigma^2 n}} e^{-\frac{y^2}{2\sigma^2 n}} + O(n^{-\alpha}),
\]

where \( h = 1 \) under the weaker assumptions and \( h = 3/2 \) under the stronger assumptions. Therefore

\[
a(x) = c(x) + \sum_{n=0}^{x^*} \frac{1}{\sqrt{2\pi \sigma^2 n}} \left( 1 - e^{-\frac{1}{2\sigma^2 n}} \right),
\]

where \( c(x) = O(\log |x|) \) under the weaker assumptions and \( c(x) = O(1) \) under the stronger assumptions. Up to a term of order \( O(1) \), the above integral equals

\[
\int_0^{\infty} \frac{1}{\sqrt{2\pi \sigma^2 t}} \left( 1 - e^{-\frac{1}{2\sigma^2 t}} \right) dt = \frac{|b|}{\sigma^2}.
\]

**Theorem 4.4.6** If \( p \in \mathcal{P}_1 \), and \( x > 0 \)

\[
a(x) = \frac{x}{\sigma^2} + \mathbb{E} \left[ a(Y) - \frac{S_T}{\sigma^2} \right],
\]

where \( T = \min\{n : S_n \leq 0\} \). There exists \( \beta > 0 \) such that for \( x > 0 \),

\[
a(x) = \frac{x}{\sigma^2} + C + O(e^{-\beta x}), \quad x \to \infty,
\]

where

\[
C = \lim_{y \to \infty} \mathbb{E} \left[ a(S_T) - \frac{S_T}{\sigma^2} \right].
\]

In particular, for simple random walk, \( a(x) = |x| \).

**Proof.** Assume \( y > x \), let \( T_y = \min\{n : S_n \leq 0 \text{ or } S_n \geq y\} \), and consider the bounded martingale \( S_{\min T_y} \). Then the optional sampling theorem implies that

\[
x = \mathbb{E}[S_y] = \mathbb{E}[S_T] = \mathbb{E}[S_T] = \mathbb{E}[S_T] < T < T_y + \mathbb{E}[S_{T_y} : T > T_y],
\]

If we let \( y \to \infty \), we see that

\[
\lim_{y \to \infty} \mathbb{E}[S_{T_y} : T > T_y] = x - \mathbb{E}[S_T].
\]

Also, since \( \mathbb{E}[S_{T_y} : T_y < T] = y + O(1) \), we can see that

\[
\lim_{y \to \infty} \mathbb{P}(T_y < y) = x - \mathbb{E}[S_T].
\]

We now consider the bounded martingale \( M_n = a(S_{\min T_y}) \). Then the optional sampling theorem implies that

\[
a(x) = \mathbb{E}[M_y] = \mathbb{E}[M_T] = \mathbb{E}[a(S_T) ; T < T_y] + \mathbb{E}[a(S_{T_y}) : T > T_y].
\]

As \( y \to \infty \), \( \mathbb{E}[a(S_T) ; T < T_y] \to \mathbb{E}[a(S_T)] \). Also, as \( y \to \infty \),

\[
\mathbb{E}[a(S_{T_y}) : T > T_y] \sim \mathbb{P}(T_y < y) \left( \frac{y}{\sigma^2} + O(1) \right) \sim x - \mathbb{E}[S_T].
\]

This gives (4.17).

To give the final estimate we will show that there exists a \( \beta \) such that if \( 0 < x < y < \infty \), then

\[
\sum_{j=0}^{\infty} \mathbb{P}(S_T = -j) - \mathbb{P}(S_T = j) = O(e^{-\beta x}).
\]

Let \( \gamma = \min\{n \geq 0 : S_n \leq 0\} \). Irreducibility and aperiodicity of the random walk can be used to see that there is an \( \epsilon > 0 \) such that for all \( x > 0 \), \( \mathbb{P}(\gamma^+ = z) > \epsilon \). Let

\[
f(r) = \gamma(r) = \sup_{r \in \mathbb{R}} \mathbb{P}(S_T = -j) - \mathbb{P}(S_T = j).
\]

Then if \( R \) denotes the range of the walk, we can see that

\[
f(r + 1) \leq (1 - \epsilon) f(r + R).
\]

Iteration of this inequality gives \( f(kR) \leq (1 - \epsilon)^k f(R) \) and this gives (4.18).

**Remark.** The potential kernel in one dimension is not as important as in higher dimensions. For \( d \geq 2 \), we use the fact that the potential kernel is harmonic on \( \mathbb{Z}^d \setminus \{0\} \) and that we know the asymptotics precisely. For \( d = 1 \), many of these same arguments can be done with the function \( f(x) = x \) which is obviously harmonic.
4.5 Fundamental solutions

If \( p \in \mathcal{P} \), the Green’s function \( G \) for \( d \geq 3 \) and the potential kernel \( a \) for \( d = 2 \) are often called the fundamental solution of the generator \( \mathcal{L} \). This is because it satisfies

\[
\mathcal{L} G(x) = -\delta_0(x), \quad \mathcal{L} a(x) = \delta_0(x),
\]

(4.19)

**Remark.** We are using the symmetry of walks in \( \mathcal{P} \). If \( p \in \mathcal{P}_r \) is transient, the Green’s function \( G \) does not satisfy (4.19). Instead it satisfies \( \mathcal{L}^R G(x) = -\delta_0(x) \) where \( \mathcal{L}^R \) denotes the generator of the “backwards random walk” with increment distribution \( p^R(x) = p(-x) \). The function \( f(x) = G(-x) \) satisfies \( \mathcal{L} f(x) = -\delta_0(x) \) and is therefore the fundamental solution of the generator. Similarly, if \( p \in \mathcal{P}_0 \), the fundamental solution of the generator is \( f(x) = a(-x) \).

**Proposition 4.5.1** Suppose \( p \in \mathcal{P}_d \) with \( d \geq 2 \), and \( f : \mathbb{Z}^d \to \mathbb{R} \) is a function satisfying \( f(0) = 0 \), \( f(x) = o(|x|) \) as \( x \to \infty \), and \( \mathcal{L} f(x) = 0 \) for \( x \neq 0 \). Then, there exists \( b \in \mathbb{R} \) such that

\[
f(x) = b G(x) - G(0), \quad d \geq 3,
\]

\[
f(x) = b a(x), \quad d = 2,
\]

**Proof.** See Propositions 6.4.5 and 6.4.7.

**Remark.** The assumption \( f(x) = o(|x|) \) is clearly needed since the function \( f(x^1, \ldots, x^d) = x^1 \) is harmonic.

Assume \( d \geq 3 \), if \( f : \mathbb{Z}^d \to \mathbb{R} \) is a function with finite support we define

\[
Gf(x) = \sum_{y \in \mathbb{Z}^d} G(x,y) f(y).
\]

(4.20)

Note that if \( f \) is supported on \( A \), then \( \mathcal{L} Gf(x) = 0 \) for \( x \not\in A \). Also if \( x \in A \),

\[
\mathcal{L} Gf(x) = -f(x).
\]

In other words \( -G = \mathcal{L}^{-1} \), The Green’s function is often called the inverse of the (negative of the) Laplacian. Similarly, if \( d = 2 \), and \( f \) has finite support, we define

\[
a f(x) = \sum_{y \in \mathbb{Z}^d} a(x,y) f(y).
\]

(4.21)

In this case we get

\[
\mathcal{L} a f(x) = f(x),
\]

i.e., \( a = \mathcal{L}^{-1} \).

4.6 Green’s function for a set

If \( A \subseteq \mathbb{Z}^d \) and \( S \) is a random walk with increment distribution \( p \), let

\[
\tau_A = \min\{ j \geq 1 : S_j \not\in A \}, \quad \tau_A = \min\{ j \geq 0 : S_j \not\in A \}.
\]

If \( A = \mathbb{Z}^d \setminus \{ x \} \), we write just \( \tau_x \), which is consistent with the definition of \( \tau_x \) given earlier in this chapter. Note that \( \tau_A, \tau_A \) agree if \( S_0 \in A \), but are different if \( S_0 \not\in A \). If \( p \) is transient or \( A \) is a proper subset of \( \mathbb{Z}^d \) we define

\[
G_A(x,y) = \mathbb{E}^x \left[ \sum_{n=0}^{\tau_A} 1(S_n = y) \right] = \sum_{n=0}^{\tau_A} p^n(\tau_A < \tau_A).
\]

**Lemma 4.6.1** Suppose \( p \in \mathcal{P}_d \) and \( A \) is a proper subset of \( \mathbb{Z}^d \):

- \( G_A(x,y) \) is defined when \( x \not\in A \).
- \( G_A(x,y) = G_A(y,x) \) for all \( x, y \).
- If \( f(y) = G_A(x,y) \), then \( \mathcal{L} f(x) = -1 \); \( \mathcal{L} f(y) = 0 \) for all \( y \in A \setminus \{x\} \).
- For each \( y \in A \),

\[
G_A(y,y) = \frac{1}{\mathbb{P}(\tau_A < \tau_A)} < \infty.
\]

If \( x, y \in A \), then

\[
G_A(x,y) = \mathbb{P}(\tau_A < \infty) G_A(y,y).
\]

- \( G_A(x,y) = G_A, (0, y - x) \) where \( A = \{ z \in \mathbb{Z}^d : z \in A \} \).

**Proof.** Easy and left to the reader. \( \square \)

The next proposition gives an important relation between the Green’s function for a set and the Green’s function or the potential kernel.

**Proposition 4.6.2** Suppose \( p \in \mathcal{P}_d \), \( A \subseteq \mathbb{Z}^d \), \( x, y \in \mathbb{Z}^d \).

(a) If \( d \geq 3 \),

\[
G_A(x,y) = G(x,y) - \mathbb{E}^x [G(S_{\tau_A}; y) ; \tau_A < \infty] = G(x,y) - \sum_z \mathbb{P}(\tau_A = z \cap G(z,y),
\]

(b) If \( d = 1, 2 \) and \( A \) is finite,

\[
G_A(x,y) = \mathbb{E}^x \left[ a(S_{\tau_A}; y) - a(z,y) \right] = \sum_z \mathbb{P}(\tau_A = z \cap a(z,y) - a(x,y).
\]
4.6. GREEN'S FUNCTION FOR A SET

Proof. The result is trivial if \( x \notin A \). We will assume \( x \in A \) in which case \( \tau_A = \tau_A \).

If \( d \geq 3 \), let \( Y_0 = \sum_{n=0}^{\infty} I(\{S_n = y\}) \) denote the total number of visits to the point \( y \). Then
\[
Y_y = \sum_{n=0}^{\infty} I(\{S_n = y\}) + \sum_{n=\tau_A}^{\infty} I(\{S_n = y\}).
\]
If we assume \( S_0 = x \) and take expectations of both sides, we get
\[
G(x, y) = G_A(x, y) + \mathbb{E}[G(S_{\tau_A}, y)].
\]
The \( d = 1, 2 \) case could be done using a similar approach, but it is easier to use a different argument. If \( S_0 = x \) and \( y \) is any function, then it is easy to check that
\[
M_n = g(S_n) - \sum_{j=0}^{n-1} \mathcal{L}_g(S_j)
\]
is a martingale. We apply this to \( g(z) = a(z, y) \) for which \( \mathcal{L}_g(z) = \delta(z - y) \). Then,
\[
a(z, y) = \mathbb{E}^n[M_0] = \mathbb{E}^n[M_{n^\tau_A}] = \mathbb{E}^n[a(S_{n^\tau_A}, y)] = \mathbb{E}^n\left[\sum_{j=0}^{n^\tau_A} I(S_j = y)\right].
\]
Since \( A \) is finite, the dominated convergence theorem implies that
\[
\lim_{n \to \infty} \mathbb{E}^n[a(S_{n^\tau_A}, y)] = \mathbb{E}^n[a(S_{\tau_A}, y)]. \tag{4.22}
\]
The monotone convergence theorem implies
\[
\lim_{n \to \infty} \sum_{j=0}^{n^\tau_A} I(S_j = y) = \sum_{j=0}^{\tau_A} I(S_j = y) = G_A(x, y).
\]
\[\square\]

The assumption of finiteness of \( A \) was used in (4.22). We will now generalize to all proper subsets \( A \) of \( \mathbb{Z}^d \), \( d = 1, 2 \). Recall that \( B_n = \{x \in \mathbb{Z}^d : |x| < n\} \). Define a function \( F_A \) by
\[
F_A(x) = \lim_{n \to \infty} \frac{\log n}{\pi \sqrt{\det \Gamma}} \mathbb{P}^\tau(\tau_N < \tau_A), \quad d = 2,
\]
\[
F_A(x) = \lim_{n \to \infty} \frac{n}{\sigma^2} \mathbb{P}^\tau(\tau_N < \tau_A), \quad d = 1.
\]
The existence of these limits is established in the next proposition. Note that \( F_A \equiv 0 \) on \( \mathbb{Z}^d \setminus A \).

Proposition 4.6.3 Suppose \( p \in \mathcal{P}_d \), \( d = 1, 2 \) and \( A \) is a proper subset of \( \mathbb{Z}^d \). Then if \( x, y \in \mathbb{Z}^d \),
\[
G_A(x, y) = \mathbb{E}[a(S_{\tau_A}, y) - a(x, y) + F_A(x)].
\]
Proof. The result is trivial if \( x \notin A \). So we will suppose that \( x \in A \). Choose \( n \) such that \( |x|, |y| \leq n \) and let \( A_n = A \cap \{H < n\} \). Using the previous lemma, we know that
\[
G_{A_n}(x, y) = \mathbb{E}[a(S_{\tau_A}, y) - a(x, y)].
\]
Note also that
\[
\mathbb{E}[a(S_{\tau_A}, y)] = \mathbb{E}[a(S_{\tau_A}, y) ; \tau_A < \tau_B] + \mathbb{E}[a(S_{\tau_A}, y) ; \tau_A \geq \tau_B].
\]
The monotone convergence theorem implies that as \( n \to \infty \),
\[
G_{A_n}(x, y) \to G_A(x, y), \quad \mathbb{E}[a(S_{\tau_A}, y) ; \tau_A < \tau_B] \to \mathbb{E}[a(S_{\tau_A}, y)].
\]
Therefore,
\[
\lim_{n \to \infty} \mathbb{E}[a(S_{\tau_A}, y) ; \tau_A > \tau_B] = G_A(x, y) + a(x, y) - \mathbb{E}[a(S_{\tau_A}, y)].
\]
However, \( n \leq |S_{\tau_A}| \leq n + R \) where \( R \) denotes the range of the increment distribution. Hence Theorems 4.4.3 and 4.4.6 show that as \( n \to \infty \),
\[
\mathbb{E}[a(S_{\tau_A}, y) ; \tau_A > \tau_B] \sim \mathbb{P}^\tau(\tau_A > \tau_B) \frac{\log n}{\pi \sqrt{\det \Gamma} / d}, \quad d = 2,
\]
\[
\mathbb{E}[a(S_{\tau_A}, y) ; \tau_A > \tau_B] \sim \mathbb{P}^\tau(\tau_A > \tau_B) \frac{n}{\sigma^2}, \quad d = 1.
\]
\[\square\]

Remark. In the last proof we also established that for \( x, y \in \mathbb{Z}^d \)
\[
F_A(x) = G_A(x, y) + a(x, y) - \mathbb{E}[a(S_{\tau_A}, y)].
\]
This holds for all \( y \); if we choose \( y \in \mathbb{Z}^d \setminus A \), then \( G_A(x, y) = 0 \), and hence we can write
\[
F_A(x) = a(x, y) - \mathbb{E}[a(S_{\tau_A}, y)].
\]
Using this expression it is easy to see that
\[
\mathcal{L}F_A(x) = 0, \quad x \in A.
\]
Also, if \( \mathbb{Z}^d \setminus A \) is finite, \( F_A(x) = a(x) + O_A(1), \quad x \to \infty \).
4.6. GREEN’S FUNCTION FOR A SET

In the particular case $A = \mathbb{Z}^d \setminus \{0\}$, $y = 0$, we see that

$$F_{\mathbb{Z}^d \setminus \{0\}}(x) = a(x).$$

Applying Proposition 4.22, we can see that

$$G_{\mathbb{Z}^d \setminus \{0\}}(x, x) = 2 a(x). \quad (4.23)$$

(Here we are assuming symmetry; in the nonsymmetric case the right-hand side becomes $a(x) + a(−x).$)

The next simple proposition relates Green’s functions to “escape probabilities” from sets. The proof uses as an idea called a last-exit decomposition. Note that the last time a random walk visits a set is a random time that is not a stopping time.

**Proposition 4.6.4 (Last-Exit Decomposition)** Suppose $p \in P_d$ and $x \in \mathbb{Z}^d$. Then,

- If $A'$ is a proper subset of $\mathbb{Z}^d$ with $A \subseteq A'$,
  
  $$P^x(\tau_{A'} < \infty) = \sum_{z \in A} G_{A'}(x, z) \mathbb{P}^x(\tau_A < \tau_{A \setminus A'}).$$

- If $\xi \in (0, 1)$ and $T_\xi$ is a geometric random variable with killing rate $1 - \xi$, then
  
  $$P^x(\tau_A < T_\xi) = \sum_{z \in A} G(x, z; \xi) \mathbb{P}^x(\tau_A \geq T_\xi).$$

- If $d \geq 3$,
  
  $$P^x(\tau_A < \infty) = \sum_{z \in A} G(x, z) \mathbb{P}^x(\tau_A = \infty).$$

**Proof.** We will prove the first; the other two are left as Exercise 4.8. On the event $\{\tau_A < \infty\}$, let $\sigma$ denote the largest $k < \tau_A$ such that $S_k \in A$. Then

$$P^x(\tau_A < \infty) = \sum_{k=0}^\infty \sum_{z \in A} P^x(\sigma = k; S_k = z) = \sum_{z \in A} \sum_{k=0}^\infty P^x(S_k = z; k < \tau_A; S_j \notin A, j = k + 1, \ldots, \tau_A).$$

The Markov property implies that

$$P^x(S_j \notin A, j = k + 1, \ldots, \tau_A \mid S_k = z; k < \tau_A) = P^x(\tau_A < \tau_{A \setminus A}).$$

Therefore,

$$P^x(\tau_A < \infty) = \sum_{z \in A} \sum_{k=0}^\infty P^x(S_k = z; k < \tau_A) \mathbb{P}^x(\tau_A < \tau_{A \setminus A})$$

$$= \sum_{z \in A} G_A(x, z) \mathbb{P}^x(\tau_A < \tau_{A \setminus A}).$$

The next proposition uses a last-exit decomposition to describe the distribution of a random walk conditioned not the return to its starting point before a killing time. The killing time is either geometric or the first exit time from a set.

**Proposition 4.6.5** Suppose $S_n$ is a $p$-walk with $p \in P_d$: $0 \in A \subset \mathbb{Z}^d$; and $\xi \in (0, 1)$. Let $T_\xi$ be a geometric random variable independent of the random walk with killing rate $1 - \xi$. Let

$$\rho = \max\{j \geq 0 : j \leq \tau_A, S_j = 0\}, \quad \rho' = \max\{j \geq 0 : j < T_\xi, S_j = 0\}.$$

- The distribution of $\{S_j : \rho \leq j \leq \tau_A\}$ is the same as the conditional distribution of $\{S_j : 0 \leq j \leq \tau_A\}$, given $\rho = 0$.
- The distribution of $\{S_j : \rho' \leq j < T_\xi\}$ is the same as the conditional distribution of $\{S_j : 0 \leq j \leq T_\xi\}$, given $\rho = 0$.

**Proof.** The usual Markov property shows that for any $j, x_1, x_2, \ldots, x_k$,

$$P\{\rho = j, \tau_A = j + k, S_{j+1} = x_1, \ldots, S_{j+k} = x_k\} = P\{S_j = 0, \tau_A > j\} P\{\tau_A = k, S_1 = x_1, \ldots, S_k = x_k\}.$$

The other equality is done similarly.

**Exercises for Chapter 4**

**Exercise 4.1** Suppose $p \in P_d$ and $S_n$ is a $p$-walk. Suppose $A \subset \mathbb{Z}^d$ and that $P^x(\tau_A = \infty) > 0$ for some $x \in A$. Show that for every $\epsilon > 0$, there is a $y$ with $P^y(\tau_A = \infty) > 1 - \epsilon$.

**Exercise 4.2** Suppose $d = 1$. Show that the only function satisfying the conditions of Proposition 4.5.1 is the zero function.

**Exercise 4.3** Find all radially symmetric functions $f$ in $\mathbb{R}^d \setminus \{0\}$ satisfying $\Delta f(x) = 0$ for all $x \in \mathbb{R}^d \setminus \{0\}$,
Exercise 4.4 For each positive integer $k$ find a $d$ and a $p \in \mathcal{P}_d$ such that $\mathbb{E}[X_1 f] < \infty$ and
\[
\limsup_{n \to \infty} |x| n^d G(x) = \infty.
\]
(Hint: Consider a sequence of points $z_1, z_2, \ldots$ going to infinity and define $P\{X_1 = z_j\} = q_j$. Note that $G(z_j) \geq q_j$. Make a good choice of $z_1, z_2, \ldots$ and $q_1, q_2, \ldots$)

Exercise 4.5 Suppose $X_1, X_2, \ldots$ are independent, identically distributed random variables in $\mathbb{R}$ with mean zero. Let $S_n = X_1 + \ldots + X_n$ denote the corresponding random walk and let
\[
G_n(x) = \sum_{j=0}^{n} P\{S_j = x\}
\]
be the expected number of visits to $x$ in the first $n$ steps of the walk.
(a) Show that $G_n(x) \leq G_n(0)$ for all $n$.
(b) Use the law of large numbers to conclude that for all $\varepsilon > 0$ there is an $N_\varepsilon$ such that for $n \geq N_\varepsilon$,
\[
\sum_{k \leq n} G_k(x) \geq \frac{n}{2}
\]
(c) Show that
\[
G(0) = \lim_{n \to \infty} G_n(0) = \infty
\]
and conclude that the random walk is recurrent.

Exercise 4.6 Let $S_n$ denote simple random walk in $\mathbb{Z}^2$ starting at the origin and let $p = \min\{j \geq 1 : S_j = 0 \}$ or $e_1$. Show that $P\{S_p = 0\} = 1/2$.

Exercise 4.7 Suppose $p \in \mathcal{P}_1$ and let $\Lambda = \{1, 2, \ldots\}$. Show that
\[
F_\Lambda(x) = \frac{x - \mathbb{E}[S_T]}{\sigma^2}
\]
where $T = \min\{j \geq 0 : S_j \leq 0\}$.

Exercise 4.8 Finish the details in Proposition 4.6.4.
Chapter 5

One-dimensional walks

5.1 Gambler's ruin estimate

One of the basic estimates for one-dimensional random walks with zero mean and finite variance goes under the name of gambler’s ruin estimate. We do not need to restrict to integer-valued random walks and it will be useful to consider more general walks. For this section we assume that $X_1, X_2, \ldots$ are independent, identically distributed random variables with $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = \sigma^2 > 0$. We let $S_n = S_0 + X_1 + \cdots + X_n$ be the corresponding random walk. If $r > 0$, we let

$$
\eta_r = \min\{n \geq 0 : S_n \leq 0 \text{ or } S_n \geq r\},
$$

$$
\eta_\infty = \min\{n \geq 0 : S_n \leq 0 \}.
$$

For simple random walk, the gambler’s ruin estimate is particular nice (and is exact).

**Proposition 5.1.1** If $S_n$ is one-dimensional simple random walk and $j < k$ are positive integers, then

$$
P^\mu[S_{\eta_k} = k] = \frac{j}{k}.
$$

**Proof.** Since $M_n := S_{\eta_n}$ is a bounded martingale, the optional sampling theorem implies that

$$
j = \mathbb{E}[M_0] = \mathbb{E}[M_{\eta_n}] = k P^\mu[S_{\eta_n} = k].
$$

**Proposition 5.1.2** If $S_n$ is one-dimensional simple random walk, then for positive integer $n$,

$$
P^\mu[\eta > 2n] = P^\mu(S_{2n} > 0) - P^\mu(S_{2n} < 0) = P(S_{2n} = 0) = \frac{1}{n} + O\left(\frac{1}{n^2}\right).
$$

**Proof.** Symmetry and the Markov property tell us that each $k < 2n$ and each positive integer $x$,

$$
P^\mu(\eta = k; S_{2n} = x) = P^\mu(\eta = k) p_{2n}(x) = P^\mu(\eta = k; S_{2n} = -x).
$$

Therefore,

$$
P^\mu(\eta \leq 2n; S_{2n} = x) = P^\mu(\eta \leq 2n; S_{2n} = -x).
$$

But, $P^\mu(\eta > 2n; S_{2n} = -x) = 0$, Therefore,

$$
P^\mu(\eta > 2n) = \sum_{x \in \mathbb{Z}} |p_{2n}(1, x) - p_{2n}(1, -x)| = p_{2n}(0, 0) = 4^{-n} \left(\frac{2n}{n}\right)^n = \frac{1}{n} + O\left(\frac{1}{n^2}\right).
$$

The basic argument in the proof of Proposition 5.1.1 above extends to more general walks. However, there is a complication arising from the fact that we do not know the exact value of $S_{\eta_n}$. We start by showing that the optional sampling theorem step is valid.

**Lemma 5.1.3** If $X_1, X_2, \ldots$ are i.i.d. random variables in $\mathbb{R}$ with $\mathbb{E}[X_j] = 0$, then for every $r > 0$ and every $x \in \mathbb{R}$,

$$
\mathbb{E}[S_n] = x.
$$

**Proof.** We start by showing that $\mathbb{E}[|S_n|] < \infty$. Fix $x$. There exists an integer $m$ and a $\delta > 0$ such that

$$
P^\mu(|X_1| > m) < (1 - \delta)^{m}.
$$

In particular, $\mathbb{E}[|\eta_n|] < \infty$. By the Markov property,

$$
P^\mu(|S_{\eta_n}| > r; \eta = k) \leq P^\mu(\eta > k - 1; |X_k| \geq y) = P^\mu(\eta > k - 1) P\{|X_k| \geq y\}.
$$

Summing over $k$ gives

$$
P^\mu(|S_{\eta_n}| > r) \leq \mathbb{E}[\eta_n] \mathbb{E}[|X_k| \geq y].
$$

Hence

$$
\mathbb{E}[|S_n|] = \int_0^\infty \mathbb{E}[|S_n| \geq y] \, dy \leq \mathbb{E}[\eta_n] \left[ r + \int_0^\infty \mathbb{P}\{|X_k| \geq y\} \, dy \right]\leq \mathbb{E}[\eta_n] \left[ r + \mathbb{E}[|X_k|] \right] < \infty.
$$

The martingale $M_n := S_{\eta_n}$ is dominated by the integrable random variable $r + |S_n|$. Hence it is a uniformly integrable martingale, and (5.1) follows from the optional sampling theorem (Theorem 3.2.3).
5.1. GAMBLER’S RUIN ESTIMATE

Proposition 5.1.4 For every \( \epsilon > 0 \) and \( K < \infty \), there exist \( 0 < c_1 < c_2 < \infty \) such that if \( \mathbb{P}(|X_1| > K) = 0 \) and \( \mathbb{P}(X_1 \geq \epsilon) \geq \epsilon \), then for all \( 0 < x < r \),
\[
   c_1 \frac{x+1}{r} \leq \mathbb{P}(S_n \geq r) \leq c_2 \frac{x+1}{r}.
\]

Proof. It suffices to show that for \( K < x \leq r/2 \),
\[
   c_1 \frac{x}{r} \leq \mathbb{P}(S_n \geq r) \leq c_2 \frac{x}{r}.
\]

We know that \( \mathbb{E}[S_n] = x \). But,
\[
\mathbb{E}[S_n] \leq \sum_{i=0}^{n} \mathbb{E}[S_{i+1}] \leq \sum_{i=0}^{n} (r + k) \mathbb{P}(S_n \geq r) \leq 4r \mathbb{P}(S_n \geq r),
\]
\[
\mathbb{E}[S_n] \geq \sum_{i=0}^{n} \mathbb{E}[S_{i+1}] \geq \sum_{i=0}^{n} (r - K) \mathbb{P}(S_n \geq r) - x.
\]

One reason to prove the last proposition is to get the following corollary. An important fact is that the constant is uniform over all \( \theta \).

Corollary 5.1.5 Suppose \( p \in \mathcal{P}_e, d \geq 2 \). There is a \( c \) such that for every \( |\theta| = 1 \),
\[
\mathbb{P}(S_j, \theta \geq -r; j = 0, \ldots, n) \leq crn^{-1/2}.
\]

5.1.1 General case

In this subsection we will prove the gambler’s ruin estimate assuming only mean zero and finite variance. While we will not attempt to get the best values for the constants, we do want to show that the constants can be chosen uniformly over a wide class of distributions. In this section we fix \( K < \infty, \delta, \rho > 0 \) and \( 0 < \rho < 1 \), and we let \( \mathcal{A} = \mathcal{A}(K, \delta, \rho) \) the collection of distributions on \( X_1 \) with \( \mathbb{E}[X_1] < \infty \),
\[
\mathbb{E}[X_1^2] = \sigma^2 \leq K^2, \quad \mathbb{P}(X_1 \geq 1) \geq \delta, \quad \inf_{n} \mathbb{P}(S_1, \ldots, S_n \geq -n) \geq \delta, \quad \rho \leq \inf_{n} \mathbb{P}(S_n \leq -n).
\]

It is easy to check that for any mean zero, finite variance random walk \( S_n \) we can find a \( t > 0 \) and some \( K, \delta, \rho > 0 \) such that the estimates above hold for \( tS_n \). Also, if \( p \in \mathcal{P}_e \), then we can find \( t, K, \delta, \rho > 0 \) such that the estimates hold for \( tS_n \) for all \( |\theta| = 1 \).

84

CHAPTER 5. ONE-DIMENSIONAL WALKS

Theorem 5.1.6 (Gambler’s ruin) If \( X_1, X_2, \ldots \) are i.i.d. random variables with mean zero and finite variance, there exist \( 0 < c_1 < c_2 < \infty \) (depending on \( K, \delta, \rho \) but not otherwise on the distribution of \( X_i \)) such that for all \( 0 < x < r \),
\[
   c_1 \frac{x}{r} \leq \mathbb{P}(\eta > r^2) \leq c_2 \frac{x}{r}, \quad c_1 \frac{x}{r} \leq \mathbb{P}(\eta \geq r) \leq c_2 \frac{x}{r}.
\]

We will do this prove in a number of steps. We start with the upper bound.

Lemma 5.1.7 Let
\[
\eta^- = \min\{n > 0 : S_n \leq 0 \text{ or } S_n \geq r\}, \quad \eta^+ = \min\{n > 0 : S_n \leq 0\}.
\]

Then
\[
\mathbb{P}(\eta^+ \geq n) \leq \frac{4K}{\delta \sqrt{n}}, \quad \mathbb{P}(\eta^- < \eta^+) \leq \frac{4K}{\delta \sqrt{n}}.
\]

Proof. Let \( \eta^- = \mathbb{P}(S_1, \ldots, S_n > 0) \). Then
\[
\mathbb{P}(S_1, \ldots, S_n \geq 1) \geq \delta \mathbb{P}(S_0 > 0) = \delta \eta^-.
\]

Let \( J_{\eta^-} \) be the indicator function of the event \( \{S_{n+1}, \ldots, S_n \geq S_n + 1\} \). Let \( M_n = \max\{S_j : 0 \leq j \leq n\} \), and for each \( x \in [m_n, M_n] \), there is at most one \( k \) such that \( S_k \leq x \) and \( S_j > x, k < j \leq n \). On the event \( J_{\eta^-} \), there is an interval of length \( n \) such that for every interval, \( S_k \leq x \) and \( S_j > x, k < j \leq n \). Therefore,
\[
\sum_{k=0}^{n} J_{\eta^-} \leq M_n - m_n.
\]

But \( \mathbb{P}(J_{\eta^-}) \geq \delta \mathbb{P}(\eta^-) \geq \delta \eta^- \). Therefore,
\[
n \delta \eta^- \leq \mathbb{E}[M_n - m_n] \leq 2 \mathbb{E}(\max\{S_j : j \leq n\} \).
\]

Martingale maximal inequalities (Theorem 8.2.3) give
\[
\mathbb{P}(\max\{S_j : j \leq n\} \geq t) \leq \mathbb{E}[S_n^2] \frac{t^2}{K^2} = \frac{t^2}{K^2},
\]

Therefore,
\[
\mathbb{E}[\max\{S_j : j \leq n\}] = \int_0^\infty \mathbb{P}(\max\{S_j : j \leq n\} \geq t) \, dt \leq K \sqrt{n} + \int_{K \sqrt{n}}^\infty K^2 n t^3 \, dt = 2K \sqrt{n}.
\]
5.1. **GAMBLER’S RUIN ESTIMATE**

This gives the first inequality. Conditioned on the event \( \{ \eta^n < \eta^* \} \), the probability that \( \{ \eta^* > n^2 \} \) is at least \( k \), and hence

\[
\delta P \{ \eta^n < \eta^* \} \leq P \{ \eta^* > n^2 \},
\]

which gives the second inequality.

\(\square\)

**Lemma 5.1.8 (Overshoot lemma)** For all \( x > 0 \),

\[
P^* \{ S_n \geq m \} \leq \frac{1}{\rho} \mathbb{E}[X_1^2] \mathbb{I}_n \mathbb{I}_m \mathbb{I}_m \mathbb{I}_m.
\]

Moreover if \( \alpha > 0 \) and \( \mathbb{E}[X_1^{1+\alpha}] < \infty \), then

\[
\mathbb{E}[S_n^\alpha] \leq \frac{\alpha}{\rho} \mathbb{E}[X_1^{1+\alpha}],
\]

**Proof.** Fix \( \epsilon > 0 \). For nonnegative integers \( k \), let

\[
Y_k = \sum_{n=0}^k 1 \{ k \epsilon < S_n \leq (k+1) \epsilon \}
\]

be the number of times the random walk visits \((k \epsilon, (k+1) \epsilon)\) before hitting \((-\infty, 0]\), and let

\[
g(x, k) = \mathbb{P}^*[Y_k] = \sum_{n=0}^\infty P^*[k \epsilon < S_n \leq (k+1) \epsilon; \eta > n].
\]

Note that if \( m, x > 0 \),

\[
P^* \{ S_n \geq m \} = \sum_{n=0}^\infty P^*[S_n \geq m; \eta = n + 1]
\]

\[
= \sum_{n=0}^\infty \sum_{k=n}^\infty P^*[S_n \geq m; \eta = n + 1, k \epsilon < S_n \leq (k+1) \epsilon]
\]

\[
\leq \sum_{k=0}^\infty \sum_{n=0}^\infty P^*[\eta > n; k \epsilon < S_n \leq (k+1) \epsilon; |S_{n+1} - S_n| \geq m + k \epsilon]
\]

\[
= \sum_{k=0}^\infty g(x, k) P[|X_1| \geq m + k \epsilon]
\]

\[
= \sum_{k=0}^\infty g(x, k) \sum_{l=k}^\infty P[m + l \epsilon < |X_1| < m + (l+1) \epsilon]
\]

\[
= \sum_{l=0}^\infty P[m + l \epsilon < |X_1| < m + (l+1) \epsilon] \sum_{k=0}^l g(x, k).
\]

We claim that for all \( x, y \),

\[
\sum_{0 \leq k < x} g(x, k) \leq \frac{y^2}{\rho},
\]

To see this, let

\[
H_y = \max \sum_{k=0}^y g(x, k).
\]

Then for any \( x \leq y \),

\[
g(x, y) \leq y^2 + \mathbb{P}^{*} [ \eta \geq y] \mathbb{E} \sum_{n=0}^{\infty} 1 \{ S_n \leq y; n < \eta \} | \eta \geq y^2 \leq y^2 + (1 - \rho) H_y,
\]

By taking the supremum over \( x \) we get (5.3). We therefore have

\[
P^* \{ S_n \geq m \} \leq \frac{1}{\rho} \sum_{p=0}^\infty P[m + l \epsilon < |X_1| < m + (l+1) \epsilon] (l + \epsilon)^2
\]

\[
\leq \frac{\alpha}{\rho} \mathbb{E}[X_1^2] \mathbb{I}_m \mathbb{I}_m \mathbb{I}_m.
\]

Letting \( \epsilon \to 0 \), we get (5.2).

To get the second estimate, let \( F \) denote the distribution function of \( |X_1| \). Then

\[
\mathbb{E}[S_n^\alpha] = \alpha \int_0^\infty [n^\alpha - 1] P \{ S_n \geq t \} dt
\]

\[
\leq \frac{\alpha}{\rho} \int_0^\infty x^\alpha / \mathbb{E}[X_1^2] \mathbb{I}_m \mathbb{I}_m \mathbb{I}_m \mathbb{I}_m dt
\]

\[
\leq \frac{\alpha}{\rho} \int_0^\infty \mathbb{E}[X_1^{1+\alpha}] \mathbb{I}_m \mathbb{I}_m \mathbb{I}_m \mathbb{I}_m dt
\]

\[
= \frac{\alpha}{\rho} \int_0^\infty x^\alpha / \mathbb{E}[X_1^{1+\alpha}] dF(x) dt
\]

\[
= \frac{\alpha}{\rho} \int_0^\infty \int_0^\infty x^\alpha / \mathbb{E}[X_1^{1+\alpha}] dF(x) dx = \frac{\alpha}{\rho} \mathbb{E}[X_1^{1+\alpha}].
\]

\(\square\)

**Heuristic note.** In short, the lemma can be stated that the overshoot random variable has two fewer moments than the increment distribution. When the starting point is farther away, one might expect that the overshoot would be larger since there is more chance for the walk to take a larger than average step. The next lemma confirms this intuition and shows that one gains one moment if one starts near the origin.
Lemma 5.1.9 (Overshoot lemma II) Let
\[ d = \frac{32 K}{b^2}. \]
Then for all \( 0 < x \leq 1 \),
\[ P^x(|S_n| \geq m) \leq \frac{\rho}{\rho + 2d K^2} P[|X_1| \geq m]. \]
Moreover if \( \alpha > 0 \) and \( E[|X_1|^{1+\alpha}] < \infty \), then
\[ E[|S_n|^\alpha] \leq \frac{\alpha \rho}{\rho + 2d K^2} E[|X_1|^{1+\alpha}]. \]

Proof. The proof proceeds exactly as in Theorem 5.1.8 up to (5.3) which we replace with a stronger estimate that is valid for \( 0 < x \leq 1 \):
\[ \sum_{x \leq k < x} g(x,k) \leq \frac{\rho}{\rho + 2d K^2} \sum_{x \leq k < x} \frac{4K}{b^2 2^{-m}} 2^{2x} \leq \frac{\rho}{\rho + 2d K^2} \frac{16K}{b^2} 2^x. \]
To derive this estimate we estimate
\[ \sum_{x \leq k < x} g(x,k) \]
by estimating the probability of reaching level \( 2^{m+1} \) before dropping below zero times the expected number of visits in this range given that. The first probability is less than \( 4K/(b2^{2m+1}) \) and the conditional expectation, as estimated in (5.3) is less than \( (1-\rho)^{-2} \). Therefore,
\[ \sum_{x \leq k < x} g(x,k) \leq \frac{1}{\rho} \sum_{x \leq k < x} \frac{4K}{b^2 2^{-m}} 2^{2x} \leq \frac{1}{\rho} \frac{16K}{b^2} 2^x. \]
For general \( y \) we write \( 2^{m+1} < y \leq 2^{x+1} \) and this gives (5.4).
Given this, the same argument gives
\[ P^x(|S_n| \geq m) \leq \frac{\rho}{\rho + 2d K^2} P[|X_1| \geq m], \]
and
\[ E[|S_n|^\alpha] = \frac{\alpha}{\rho} \int_0^\infty t^{\alpha-1} P[|S_n| \geq t] dt \leq \frac{\alpha \rho}{\rho + 2d K^2} \frac{16K}{b^2} \int_0^\infty t^{\alpha-1} dt = \frac{\alpha}{\rho} \frac{16K}{b^2} \int_0^\infty t^{\alpha-1} dt. \]

\[ \square \]
5.2. ONE-DIMENSIONAL KILLED WALKS

and hence
\[ \mathbb{E}[S_{n+2}; n \leq S_{n+2} \leq (1 + t_0)n] \geq \frac{1}{2}, \]

which implies
\[ \mathbb{P}^*[n \leq S_{n+2} \leq (1 + t_0)n] \geq \frac{1}{2(1 + t_0)n}. \]

Completion of proof of Theorem 5.1.6. Lemmas 5.1.7 and 5.1.10 prove the result for $0 \leq x \leq 1$. It suffices to prove the result for $1 \leq x \leq r/2$. Clearly the function $x \mapsto \mathbb{P}^*[S_n \geq r]$ is nondecreasing in $x$. Therefore,
\[ \mathbb{P}(S_n \geq r) = \mathbb{P}(S_n \geq x) \mathbb{P}(S_{n+1} \geq r \mid S_n \geq x) \geq \mathbb{P}(S_n \geq x) \mathbb{P}^*[S_n \geq r]. \]

Hence,
\[ \mathbb{P}^*[S_n \geq r] \leq \frac{\mathbb{P}(S_n \geq x)}{\mathbb{P}(S_{n+1} \geq r \mid S_n \geq x)} \leq \frac{4K r}{c_0 x}. \]

For the other direction, we need to find a constant $c_2$ such that $\mathbb{P}[S_n] \leq c_2 x$. Once we have this, we can show as in the previous lemma that
\[ \mathbb{E}[S_{n+2}; S_{n+2} \geq (1 + t_0)n] \leq \frac{c_2 K^2 x}{t}. \]

and hence if $t_0 = 2c_2 K^2$,
\[ \mathbb{E}[S_{n+2}; S_{n+2} \geq (1 + t_0)n] \leq \frac{x}{2}, \]
\[ \mathbb{E}[S_{n+2}; n \leq S_{n+2} \leq (1 + t_0)n] \geq \frac{x}{2}, \]
\[ \mathbb{P}^*[n \leq S_{n+2} \leq (1 + t_0)n] \geq \frac{x}{2(1 + t_0)n}. \]

As we have already shown, this implies
\[ \mathbb{P}^*[n \geq r] \geq \frac{x}{2(1 + t_0)n}. \]

5.2 One-dimensional killed walks

A symmetric defective increment distribution (on $\mathbb{Z}$) is a set of nonnegative numbers $\{p_k : k \in \mathbb{Z}\}$ with $\sum p_k < 1$ and $p_{-k} = p_k$ for all $k$. Given a symmetric defective increment distribution, we have the corresponding symmetric random walk with killing $S_j$. This is a Markov chain with state space $\mathbb{Z} \cup \{\infty\}$, where $\infty$ is an absorbing state, and
\[ \mathbb{P}(S_{j+1} = k + l \mid S_j = k) = p_l, \quad \mathbb{P}(S_{j+1} = \infty \mid S_j = k) = p_\infty. \]

Here $p_n = 1 - \sum p_k$. We let
\[ T = \min\{j : S_j = \infty\} \]
denote the killing time for the random walk. Note that $\mathbb{P}(T = j) = p_\infty(1 - p_\infty)\gamma^j, j \in \{1, 2, \ldots\}$.

Examples.

• Suppose $p_j$ is the increment distribution of a symmetric one-dimensional random walk and $s \in [0, 1)$. Then $p_j = sp_j$ is a defective increment distribution corresponding to the random walk killed with geometric rate $1 - s$. Conversely, if $p_j$ is a symmetric defective increment distribution, and $p_j = p_j/(1 - p_\infty)$, then $p_j$ is an increment distribution of a symmetric one-dimensional random walk (not necessarily aperiodic or irreducible). If we then kill this walk at rate $1 - p_\infty$ we get back $p_j$.

• Suppose $S_j$ is a symmetric random walk in $\mathbb{Z}^d, d \geq 2$ which we write $S_j = (Y_j, Z_j)$ where $Y_j$ is a random walk in $\mathbb{Z}$ and $Z_j$ is a random walk in $\mathbb{Z}^d \setminus \{0\}$. Suppose the random walk is killed at rate $1 - s$ and let $T$ denote the killing time. Let
\[ \tau = \min\{j \geq 1 : Z_j = 0\}. \]

Then set
\[ p_k = \mathbb{P}(Y_r = k ; \tau < T), \]

Note that
\[ p_k = \sum_{j=1}^{\infty} \mathbb{P}(\tau = j ; Y_j = k ; \tau < T) = \sum_{j=1}^{\infty} \mathbb{P}(\tau = j ; Y_j = k) = \mathbb{E}p_\infty^j ; Y_j = k ; \tau < \infty]. \]

If $Z$ is transient, then we can let $s = 1$.

• Suppose $S_j = (Y_j, Z_j)$ is as in the previous example and suppose $A \subset \mathbb{Z} \setminus \{0\}$. Let
\[ \sigma_A = \min\{j : Z_j \in A\}, \quad p_k = \mathbb{P}(Y_r = k ; \tau < \sigma_A). \]

If $\mathbb{P}(Z_j \in A$ for some $j) > 0$, then $\{p_k\}$ is a defective increment distribution.

Given the symmetric defective increment distribution $\{p_k\}$ define the events
\[ V_+ = \{S_j > 0 : j = 1, \ldots, T - 1\}, \quad V_+ = \{S_j \geq 0 : j = 1, \ldots, T - 1\}, \quad V_- = \{S_j < 0 : j = 1, \ldots, T - 1\}, \quad V_- = \{S_j \leq 0 : j = 1, \ldots, T - 1\}. \]
5.2. ONE-DIMENSIONAL KILLED WALKS

Symmetry implies that $P(V_+ = P(V_-; \mathbb{V}_+ = P(V_-))$. Note that
\begin{equation}
P(V_+ \cap V_-) = P(V_+ \cap V_-) = P(T = 1) = p_{\omega},
\end{equation}

Define the defective increment distribution $p_{\omega}$ which is supported on $k = 0,-1,-2,\ldots$ by saying that $p_{\omega}$ is the probability that the first visit to \{-1, -2, -1, 0\} occurs at position $k$ and this occurs before the killing time $T$, i.e.,
\[ p_{\omega} = \sum_{l=1}^{\infty} P(S_j = k ; S_j > 0, l = 1, \ldots, k - 1 ; j < T). \]

Define $p_{k,t}$ similarly and note that $p_{k,t} = p_{\omega}$. Note that
\[ P(V_+) = P(V_+) + p_{\omega} P(V_-), \]
and hence
\begin{equation}
P(V_+) = (1 - p_{\omega}) P(V_-).
\end{equation}

In the next proposition we prove a nontrivial fact,

**Proposition 5.2.1** The events $V_+$ and $V_-$ are independent. In particular,
\begin{equation}
P(V_+) = \sqrt{p_{\omega} (1 - p_{\omega})}.
\end{equation}

**Proof.** Independence is equivalent to the statement $P(V_+ \cap V_-) = P(V_-) P(V_-)$. We will prove the equivalent statement $P(V_+ \cap V_-) = P(V_-) P(V_-)$. Note that $V_+ \cap V_- = \emptyset$, the event that $T > 1$ but no point in \{0, 1, 2, \ldots\} is visited in times \{1, \ldots, T - 1\}. In particular, at least one point in \{-2, -1\} is visited before time $T$.

Let
\[ \rho = \max{k \in \mathbb{Z} : S_j = k \text{ for some } j = 1, \ldots, T - 1}, \]
\[ \eta = \min{j \geq 1 : S_j > 0}, \quad \xi = \max{j \geq 0 : S_j = k ; j < T}. \]

Then,
\[ P(V_+ \cap V_-) = \sum_{k=1}^{\infty} P(\rho = -k) = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} P(\rho = -k ; \xi = j). \]

Note that the event \( \{ \rho = -k ; \xi = j \} \) is the same as the event
\[ \{ S_j = -k ; j < T ; S_i \leq -k, l = 1, \ldots, j - 1 ; S_j = -k, l = j + 1, \ldots, T - 1 \}. \]

Hence,
\[ P(\rho = -k ; \xi = j) = P(\rho = -k ; j < T ; S_i \leq -k, l = 1, \ldots, j - 1) P(\xi), \]
\end{proof}

$\Box$

CHAPTER 5. ONE-DIMENSIONAL WALKS

However, by path reversal and symmetry of the walk, we can see that
\[ P(S_j = -k ; j < T ; S_i \leq -k, l = 1, \ldots, j - 1) = P(\eta = j ; j < T ; S_j = k). \]

But,
\[ \sum_{j=1}^{\infty} \sum_{j=1}^{\infty} P(\eta = j ; j < T ; S_j = k) = P(\eta < T) = P(V_-) = P(V_+). \]

This establishes independence. The equality (5.8) follows from
\[ p_{\omega} = P(V_-) = P(V_-) P(V_+) = P(V_-) P(V_+) (1 - p_{\omega}) = P(V_+) (1 - p_{\omega}). \]

$\Box$

5.2.1 Hitting a half-line

In this section we will give an application of Proposition 5.2.1 to walks in $\mathbb{Z}^d$. Suppose $d \geq 2$ and $S_n$ is a random walk with increment distribution $p \in \mathcal{P}_d$. Let us write $S_n = (Y_n, Z_n)$ where $Y_n$ is a one-dimensional walk and $Z_n$ is a \((d-1)\)-dimensional walk. Let $\Gamma$ denote the covariance matrix for $S_n$ and let $G$ denote the covariance matrix for the \((d-1)\)-dimensional walk $Z_n$. Let $T = \min\{j > 0 : Z_j = 0\}$ be the first time that the random walk returns to the line \((j, x) : \mathbb{Z} \times \mathbb{Z}^d \setminus \{x = 0\}\). Let $T_n$, be the corresponding quantity for the half-line
\[ T_n = \min\{n > 0 : S_n \in \{(j, x) : \mathbb{Z} \times \mathbb{Z}^d \setminus \{x = 0\}\} \}, \]

and let
\[ p_{\omega} = P(\eta = 0) = 0. \]

**Proposition 5.2.2** If $p \in \mathcal{P}_d$, \( d = 2, 3 \), there is an $r > 0$ such that as $n \to \infty$,
\[ P(T_n > n) \sim \begin{cases} n^{-d/4}, & d = 2, \\ n \log(n)^{-d/2}, & d = 3. \end{cases} \]

**Proof.** Let $\sigma = \sigma_t$ denote a geometric random variable, independent of the random walk, with killing rate $1 - \xi$, i.e., $P(\sigma = k) = \xi^k$. Let $\phi_n = P(T_n > n)$. Then,
\[ P(T_n > n) = \sum_{n=1}^{\infty} P(\sigma = n ; T_n > n) = \sum_{n=1}^{\infty} (1 - \xi)^{-n} \phi_n. \]

By Proposition 5.5.2, it suffices to show that $P(T_n > \sigma) \sim c(1 - \xi)^{d/2}$ if $d = 2$ and $P(T_n > \sigma) \sim c(\log(1 - \xi))^{-d/2}$ if $d = 3$. Proposition 5.2.1 tells us that
\[ q(\xi) = \frac{P_{\omega}(\xi)}{1 - p_{\omega}(\xi)}. \]
where \( p_m(\xi) = \mathbb{P}(T > \xi) \) and \( 1 - p_m(\xi) = \mathbb{P}(T \leq \xi; Y_T \neq 0) \). Clearly, as \( \xi \rightarrow 1^- \), 
\( 1 - p_m(\xi) \rightarrow 1 - p_{0,+} \). By using (4.8) and (4.9) on the random walk \( Z_n \), we can see that 
\[
\mathbb{P}(T > \xi) \sim c (1 - \xi)^{1/2}, \quad d = 2,
\]
\[
\mathbb{P}(T > \xi) \sim c \log \left( \frac{1}{1 - \xi} \right), \quad d = 3.
\]

**Remark.** From the proof one can see that the constant \( r \) can be determined in terms of \( \Gamma^* \) and \( p_{0,+} \). We will not need the exact value and the proof is a little easier to follow if we do not try to keep track of this constant.

**Remark.** The proof uses the surprising fact that the events “avoid the positive \( x^+ \)-axis” and “avoid the negative \( x^- \)-axis” are independent up to a multiplicative constant. This idea does not extend to other sets, for example the event “avoid the positive \( x^+ \)-axis” and “avoid the positive \( x^- \)-axis” are not independent up to a multiplicative constant in two dimensions. However, they are in three dimensions.
Chapter 6

Potential Theory

6.1 Introduction

There is a close relationship between random walks with increment distribution \( p \) and functions that are harmonic with respect to the generator \( L = L_p \).

We start by setting some notation. We fix \( p \in \mathcal{P} \), if \( A \subseteq \mathbb{Z}^d \), we let

\[
\partial A = \{ x \in \mathbb{Z}^d \setminus A : \exists y \in A, p(y, x) > 0 \}
\]
denote the (outer) boundary of \( A \) and we let \( \overline{A} = A \cup \partial A \). Our definition of \( \partial A, \overline{A} \) depends on the choice of \( p \); we hope this will not be confusing. In the case of simple random walk,

\[
\partial A = \{ x \in \mathbb{Z}^d \setminus A : |y - x| = 1 \text{ for some } y \in A \}.
\]

Since \( p \) has finite range, if \( A \) is finite, then \( \partial A, \overline{A} \) are finite. The inner boundary of \( A \subseteq \mathbb{Z}^d \) is defined by

\[
\partial_i A = \partial (\mathbb{Z}^d \setminus A) = \{ x \in A : \exists y \in A, p(x, y) > 0 \text{ for some } y \in A \}.
\]

A function \( f : \overline{A} \to \mathbb{R} \) is harmonic (with respect to \( p \)) in \( A \) if \( Lf(y) = 0 \) for every \( y \in A \). Note that we cannot define \( Lf(y) \) for all \( y \in \overline{A} \) unless \( f \) is defined on \( \overline{A} \).

We say that \( A \) is connected (with respect to \( p \)) if for every \( x, y \in A \), there is a finite sequence \( z = z_0, z_1, \ldots, z_k = y \) of points in \( A \) with \( p(z_j, z_{j+1}) > 0, j = 0, \ldots, k \). Our definition of connectedness depends on the choice of \( p \).

Heuristic note. We will be proving a number of results about functions on subsets of \( \mathbb{Z}^d \). These results have continuous analogues. The set \( \mathbb{R}^d \) corresponds to an open set \( D \subseteq \mathbb{R}^2 \) the outer boundary \( \partial A \) corresponds to the usual topological boundary \( \partial D \), and \( \overline{A} \) corresponds to the closure \( \overline{D} = D \cup \partial D \). The term domain is often used for open, connected subsets of \( \mathbb{R}^d \). Finiteness assumptions on \( A \) correspond to boundedness assumptions on \( D \).

6.2 Dirichlet problem

The standard Dirichlet problem for harmonic functions is to find a harmonic function on a region with specified values on the boundary.

**Theorem 6.2.1 (Dirichlet problem 1)** Suppose \( p \in \mathcal{P}_d \) and \( A \subseteq \mathbb{Z}^d \) satisfying \( \mathbb{P}^p(\{ A < \infty \}) = 1 \) for all \( x \in A \). Suppose \( F : \overline{A} \to \mathbb{R} \) is a bounded function. Then there is a unique bounded function \( f : \overline{A} \to \mathbb{R} \) satisfying

\[
L_f(x) = 0, \quad x \in A, \tag{6.1}
\]

\[
f(x) = F(x), \quad x \in \partial A. \tag{6.2}
\]

It is given by

\[
f(x) = \mathbb{E}^p[F(S_n)], \tag{6.3}
\]

**Proof.** Immediate from the definition. \( \square \)
6.2. DIRICHLET PROBLEM

Proof. A simple application of the Markov property shows that $f$ defined by (6.3) satisfies (6.1) and (6.2). Now suppose $f$ is a bounded function satisfying (6.1) and (6.2). Then $M_n = f(S_n)$. $\mathbf{M}_n$ is a bounded martingale. Hence, the optional sampling theorem (Theorem 8.2.3) implies that

$$f(x) = \mathbb{E}^\tau[M_0] = \mathbb{E}^\tau[M_n] = \mathbb{E}^\tau[f(S_n)].$$

Remark. If $A$ is finite, then $\partial A$ is also finite and all functions on $\overline{A}$ are bounded. Hence for each $F$ on $\partial A$, there is a unique function satisfying (6.1) and (6.2). In that case we could have proved this using linear algebra since (6.1) and (6.2) give $\#(A)$ linear equations in $\#(A)$ unknowns. However, algebraic methods do not yield the nice probabilistic form (6.3).

If $A$ is infinite, there may be more than one solution to the Dirichlet problem if we allow unbounded solutions. For example, if $A = \{1, 2, 3, \ldots\}$ and $F(0) = 0$, then there is an infinite number of solutions of the form $f_n(x) = bx$. If $b \neq 0$, $f_n$ is unbounded.

If $d = 1, 2$ and $A$ is a proper subset of $\mathbb{Z}^d$, then we know by recurrence that $P^\tau(f_A < \infty) = 1$ for all $x \in A$.

If $d \geq 3$ and $A \neq \mathbb{Z}^d$, then there are points $x \in A$ with $P^\tau(f_A = \infty) > 0$. The function

$$f(x) = P^\tau(f_A = \infty)$$

is a bounded function satisfying (6.1) and (6.2) with $F = 0$ on $\partial A$. Hence, the condition $P^\tau(f_A < \infty) = 1$ is needed to guarantee uniqueness. However, as Proposition 6.2.6 below shows, all solutions with $F = 0$ on $A$ are multiples of $f$.

Remark. This theorem has a well-known continuous analogue: Suppose $f : \{x\in \mathbb{R}^d : |x| \leq 1\} \rightarrow \mathbb{R}$ is a continuous function with $\Delta f(x) = 0$ for $|x| < 1$. Then

$$f(x) = \mathbb{E}^\tau[f(B_t)],$$

where $B$ is a standard $d$-dimensional Brownian motion and $T$ is the first time that $|B_t| = 1$. If $|x| < 1$, the distribution of $B_t$ given $B_0 = x$ has a density with respect to surface measure on $\{x\} = 1$. This density $h(x, z) = c(1 - |x|^2)/|x - z|^d$ is called the Poisson kernel and we can write

$$f(x) = c \int_{|z| = 1} f(z) \frac{1 - |x|^2}{|x - z|^d} d\sigma(z), \quad (6.4)$$

where $\sigma$ denotes surface measure. To verify this is correct, one can check directly that $f$ as defined above is harmonic in the ball and satisfies the boundary condition on the sphere. Two facts follow almost immediately from this integral formula:

- Derivative estimates. For every $k$, there is a $c = c(k) < \infty$ such that if $f$ is harmonic in the unit ball and $D$ denotes a $k$th order derivative, then $Df(0) \leq c|f|_\infty$.

- Harnack inequality. For every $r < 1$, there is a $c = c_r < \infty$ such that if $f$ is a positive harmonic function on the unit ball, then $f(x) \leq c f(y)$ for $|x|, |y| \leq r$.

An important aspect of these estimates is that the constants do not depend on $f$. We will prove the analogous results for the random walk in Section 6.3.

Proposition 6.2.2 (Dirichlet problem II). Suppose $p \in \mathcal{P}_4$ and $A \subset \mathbb{Z}^d$. Suppose $F : \partial A \rightarrow \mathbb{R}$ is a bounded function. Then the only bounded functions $f : A \rightarrow \mathbb{R}$ satisfying (6.2) and (6.9) are of the form

$$f(x) = \mathbb{E}^\tau(f(S_n); |x| < 0) + b \mathbb{P}^\tau(\tau_A = \infty), \quad (6.5)$$

for some $b \in \mathbb{R}$.

Proof. We may assume that $p$ is aperiodic. We also assume that $P^\tau(f_A = \infty) > 0$ for some $x \in A$; if not, then Theorem 6.2.1 applies. Assume that $f$ is a bounded function satisfying (6.2) and (6.3). Since $M_t := f(S_n \cap A)$ is a martingale, we know that

$$f(x) = \mathbb{E}_x^\tau[f(S_n)] = \mathbb{E}_x^\tau[E(f(S_n); \tau_A < \infty) + \mathbb{E}(f(S_n); \tau_A = \infty).$$

Using Lemma 2.4.4, we can see that for all $x, y$,

$$\lim_{n \to \infty} \mathbb{E}(f(S_n); \tau_A < \infty) = 0,$

Therefore,

$$|f(x) - f(y)| \leq \mathbb{P}^\tau(f_A < \infty) + \mathbb{P}^\tau(f_A = \infty).$$

Let $U = \{x \in \mathbb{Z}^d : \mathbb{P}^\tau(\tau_A = \infty) \geq 1 - \epsilon\}$. Then

$$|f(x) - f(y)| \leq 4 \epsilon |f|_\infty, \quad x, y \in U.$$

Since $P^\tau(f_A = \infty) > 0$ for some $x$, one can easily check (Exercise 4.1) that $U$ is nonempty for each $\epsilon$. Hence, there is a $b$ (which we can think of as $f(\infty)$) such that

$$|f(x) - b| \leq 4 \epsilon |f|_\infty, \quad x \in U.$$

Let $\rho_1$ be the minimum of $\tau_A$ and the smallest $n$ such that $S_n \in U$. Then for every $x \in \mathbb{Z}^d$, the optional sampling theorem implies

$$f(x) = \mathbb{E}(f(S_n); \tau_A \leq \rho_1) + \mathbb{E}(f(S_n); \tau_A > \rho_1),$$

(Here we use the fact that $P^\tau(\tau_A \wedge \rho_1 < \infty) = 1$ which can be verified easily.) Letting $\epsilon \to 0$ and using the dominated convergence theorem, we get (6.5).
6.2. **DIRICHLET PROBLEM**

**Remark.** We can think of (6.5) as being the same as (6.3) except that we have added a point at infinity. We will refer to the constant $b$ in the last proposition as the "boundary value at infinity" and write it as $F(\infty)$. The fact that there is a single boundary value at infinity is closely related to Proposition 6.1.2.

**Definition.** If $p \in \mathcal{P}_d$ and $A \subseteq \mathbb{Z}^d$, then the Poisson kernel is the function $H : \mathbb{T} \times \partial A \to [0, 1]$ defined by

$$H_A(x, y) = P^x(\tau_A < \infty; \tau_A = y).$$

As a slight abuse of notation we will also write

$$H_A(x, \infty) = P^x(\tau_A = \infty).$$

Note that

$$\sum_{y \in \partial A} H_A(x, y) = P^x(\tau_A < \infty).$$

For fixed $y \in \partial A$, $f(x) = H_A(x, y)$ is the unique function on $\mathbb{T}$ that is harmonic on $A$, equals $\delta(x - y)$ on $\partial A$, and tends to 0 as $x$ tends to infinity. We can write (6.3) as

$$f(x) = \mathbb{E}[F(S_x); \tau_A < \infty] + b P^x(\tau_A = \infty) = \sum_{y \in \partial A} H_A(x, y) F(y),$$

and (6.5) as

$$f(x) = \mathbb{E}[F(S_x); \tau_A < \infty] + b P^x(\tau_A = \infty) = \sum_{y \in \partial A} H_A(x, y) F(y),$$

where $F(\infty) = b$.

**Proposition 6.2.3** Suppose $p \in \mathcal{P}_d$ and $A \subseteq \mathbb{Z}^d$. Let $g : A \to \mathbb{R}$ be a function with finite support. Then the function

$$f(x) = \sum_{y \in A} G_A(x, y) g(y) = \mathbb{E}\left[\sum_{j=0}^{\infty} g(S_j)\right],$$

is the unique bounded function on $\mathbb{T}$ that vanishes on $\partial A$ and satisfies

$$\mathcal{L} f(x) = -g(x), \quad x \in A.$$  

**Proof.** Since $g$ has finite support,

$$|f(x)| \leq \sum_{y \in A} G_A(y, y) |g(y)| < \infty,$$

and hence $f$ is bounded. We have already noted that $\mathcal{L} u(x, y) = -\delta(x - y)$ for $x \in A$, and hence $f$ satisfies (6.7). Now suppose $f$ is a bounded function vanishing on $\partial A$ satisfying (6.7). Then, Proposition 6.1.1 implies that

$$M_n := f(S_n) + \sum_{j=0}^{n-1} g(S_j),$$

is a martingale. Note that $|M_n| \leq \|f\| + \|g\|$ where

$$Y = \sum_{j=0}^{\infty} |g(S_j)|,$$

and that

$$\mathbb{E}[Y] = \sum_{y \in \partial A} G_A(x, y) |g(y)| < \infty.$$

Hence $M_n$ is dominated by an integrable random variable and we can use the optional sampling theorem (Theorem 8.2.3) to conclude that

$$f(x) = \mathbb{E}[M_0] = \mathbb{E}[M_{\infty}] = \mathbb{E}\left[\sum_{j=0}^{\infty} g(S_j)\right].$$

\[\square\]

**Remark.** Suppose $A \subseteq \mathbb{Z}^d$ is finite with $\|\cdot\| = m$. Then $[G_A(x, y)]_{x \in \partial A}$ is a symmetric matrix with nonnegative coefficients. We can also consider the $m \times m$ symmetric matrix $[\mathcal{L}^A(x, y)]_{x \in \partial A}$ defined by

$$\mathcal{L}^A(x, y) = p(x, y), \quad x \neq y; \quad \mathcal{L}^A(x, x) = p(x, x) = 1.$$ 

If $g : A \to \mathbb{R}$ and $x \in A$, then $\mathcal{L}^A g(x)$ is the same as $\mathcal{L} g(x)$ where $g$ is extended to $\mathbb{T}$ by setting $g \equiv 0$ on $\partial A$. The last proposition can be rephrased as $\mathcal{L}^A G_A g = -g$, or in other words, $G_A = (-\mathcal{L}^A)^{-1}$.

**Corollary 6.2.4** Suppose $p \in \mathcal{P}_d$ and $A \subseteq \mathbb{Z}^d$ is finite. Let $g : A \to \mathbb{R}^d F : \partial A \to \mathbb{R}$ be given. Then the function

$$f(x) = \mathbb{E}[F(S_x)] + \mathbb{E}\left[\sum_{j=0}^{\infty} g(S_j)\right] = \sum_{y \in \partial A} H_A(x, y) F(z) + \sum_{y \in \partial A} G_A(x, y) g(y),$$

is the unique function on $\mathbb{T}$ that satisfies

$$\mathcal{L} f(x) = -g(x), \quad x \in A,$$

$$f(x) = F(x), \quad x \in \partial A,$$

\[1\] Here we use $\mathcal{L}_x$ to denote $\mathcal{L}$ applied to the variable $x$. 

---

100 **CHAPTER 6. POTENTIAL THEORY**

---
6.3. Difference Estimates and Harnack’s Inequality

Proof. Use the fact that \( h(x) := f(x) - \mathbb{E}^x[F(S_{n_x})] \) satisfies the conditions of the previous proposition.

Corollary 6.2.5 If \( p \in \mathcal{P}_d \) and \( A \subseteq \mathbb{Z}^d \), then the unique bounded function \( f : \mathbb{Z} \to \mathbb{R} \) that satisfies \( \Delta f(x) = -1 \) for \( x \in A \) is \( f(x) = \mathbb{E}^x[S_A] \).

Proposition 6.2.6 Let \( \rho_n = \mathbb{P}_n = \inf\{j \geq 0 : |S_j| \geq n\} \). Then if \( p \in \mathcal{P}_d \) with range \( R \) and \( |x| \leq n \),

\[ [n^2 - |x|^2] \leq (\text{tr} \Gamma) \mathbb{E}^x[\rho_n] \leq [(n + R)^2 - |x|^2]. \]

Proof. In Exercise 1.4 it was shown that \( M_j = (S_{j,\rho_n})^2 - (\text{tr} \Gamma)(j \wedge \rho_n) \) is a martingale. Also, \( \mathbb{E}^x[\rho_n] \leq \infty \) for each \( x \), so \( M_j \) is dominated by the integrable random variable \( n + R + (\text{tr} \Gamma) \rho_n \). Hence,

\[ |x|^2 \leq \mathbb{E}^x[M_0] = \mathbb{E}^x[S_{\rho_n}] = \mathbb{E}^x[(S_{\rho_n})^2] - (\text{tr} \Gamma) \mathbb{E}^x[\rho_n]. \]

However, \( n^2 \leq (n + R)^2 \).

6.3 Difference Estimates and Harnack’s Inequality

In the next two sections we will prove a number of facts about random walk and harmonic functions. The main tools in the proofs are the optional sampling theorem and the estimates for the Green’s function and the potential kernel. The basic idea in many of the proofs is to define a martingale in terms of the Green’s function or potential kernel and then to stop it at a place at which the function is approximately constant. We recall that

\[ B_n = \{ z \in \mathbb{Z}^d : |z| < n \}, \quad C_n = \{ z \in \mathbb{Z}^d : f(z) < n \}. \]

As the next proposition points out, the Green’s function and potential kernel are almost constant on \( \partial C_n \). We recall that Theorems 4.3.1 and 4.4.3 imply that \( x \to \infty \),

\[ G(x) = \frac{C_d}{\mathcal{J}(x)^{-n}} + O\left(\frac{1}{|x|^n}\right), \quad d \geq 3, \]

\[ a(x) = C_2 \log \mathcal{J}(x) + \gamma_2 + O\left(\frac{1}{|x|^n}\right), \quad d = 2. \]

Here \( C_d = \pi \text{det} \Gamma^{-1/2} \) and \( \gamma_2 = C + C_2 \log d^{-1/2} \) where \( C \) is as in Theorem 4.4.3.

Proposition 6.3.1 If \( p \in \mathcal{P}_d, d \geq 3 \) then for \( x \in \partial C_n \cup \partial C_n \),

\[ G(x) = \frac{C_d}{n^{-d}} + O(n^{-\delta}), \quad d \geq 3, \]

\[ a(x) = C_2 \log n + \gamma_2 + O(n^{-\delta}), \quad d = 2, \]

where \( C_d, C_2, \gamma_2 \) are as in (6.9) and (6.10),

\[ | \mathcal{G}_d(0, x) - \mathcal{G}_d(y, x) | \leq \frac{C}{n^\delta}, \]

\[ | \mathcal{G}_d(0, x) - \mathcal{G}_d(y, x) - \mathcal{G}_d(y, -x) | \leq \frac{C}{n^\delta}. \]

Proof. This follows immediately from (6.9) and (6.10) and the estimate

\[ \mathcal{J}(x) = n + O(1), \quad x \in \partial C_n \cup \partial C_n. \]

Note that the error term \( O(n^{-\delta}) \) comes from the estimates

\[ [n + O(1)]^{-n} = n^{-n} + O(n^{-\delta}), \quad \log[n + O(1)] = \log n + O(n^{-\delta}). \]

Proposition 6.3.2 If \( p \in \mathcal{P}_d, d \geq 3 \),

\[ \mathcal{G}_d(0, 0) = G(0, 0) - \frac{C_d}{n^{-d}} + O(n^{-\delta}), \quad d \geq 3, \]

\[ \mathcal{G}_d(0, 0) = C_2 \log n + \gamma_2 + O(n^{-\delta}), \quad d = 2. \]

where \( C_d, \gamma_2 \) are as defined in Proposition 6.3.1.

Proof. Lemma 4.6.2 tells us that

\[ \mathcal{G}_d(0, 0) = G(0, 0) - \mathbb{E}[S_{\rho_n}], \quad d \geq 3, \]

\[ \mathcal{G}_d(0, 0) = \mathbb{E}[S_{\rho_n}], \quad d = 2. \]

Hence this follows immediately from Proposition 6.3.1.

We will now prove difference estimates and a Harnack inequality for harmonic functions. There are different possible approaches to proving these results. One might be to start with the result for Brownian motion and use approximation. We will use a different approach where we start with the known difference estimates for the Green’s function \( G \) and the potential kernel \( a \) and work from there. We start by proving a difference estimate for \( \mathcal{G}_d \). We then use this to prove a result on probabilities that is closely related to the gambler’s ruin estimate for one-dimensional walks.

Lemma 6.3.3 If \( p \in \mathcal{P}_d, d \geq 3 \), then for every \( \epsilon > 0, r < \infty \) there is a \( c > 0 \) such that if \( B_n \subset A \subseteq \mathbb{Z}^d \), then for every \( |x| > \epsilon n \) and every \( |y| \leq r \),

\[ | \mathcal{G}_d(0, x) - \mathcal{G}_d(y, x) | \leq \frac{C}{n^\delta}, \]

\[ | \mathcal{G}_d(0, x) - \mathcal{G}_d(y, x) - \mathcal{G}_d(y, -x) | \leq \frac{C}{n^\delta}. \]
6.3. DIFFERENCE ESTIMATES AND HARNACK’S INEQUALITY

**Proof.** It suffices to prove the result for finite $A$ for we can approximate any $A$ by a finite $A$. Assume that $x \in A$, otherwise the result is trivial. By symmetry $G_A(0, z) = G_A(z, 0)$, $G_A(y, x) = G_A(x, y)$. By Lemma 4.6.2,

$$ G_A(x, 0) - G_A(x, y) = G(x, 0) - G(x, y) - \sum_{z \in A} H_A(x, z) [G(z, 0) - G(z, y)], \quad d \geq 3, $$

$$ G_A(x, 0) - G_A(x, y) = a(x, 0) - a(x, y) - \sum_{z \in A} H_A(x, z) [a(z, 0) - a(z, y)], \quad d = 2. $$

The difference estimates for the Green’s function and the potential kernel (Corollaries 4.3.3 and 4.4.4) give

$$ |G(z) - G(z + y)| \leq c_n |y|^{1-d}, \quad |\Pi G(z) - G(z + y)| = |G(z - y)| \leq c_n |y|^d $$

for $d \geq 3$ and

$$ |a(z) - a(z + y)| \leq c_n n^{-1}, \quad |a(z) - a(z + y) - a(z - y)| \leq c_n n^{-1} $$

for $d = 2$, provided that $|y| \leq r$ and $|z| \geq (c/2)n$. □

The next lemma is very closely related to the one-dimensional gambler’s ruin estimate. This lemma is particularly useful for $x$ on or near the boundary of $C_n$. For $x$ in $C_n \setminus C_{n/2}$ that are away from the boundary, there are sharper estimates. See Lemmas 6.4.1 and 6.4.2.

**Lemma 6.3.4** Suppose $p \in P(A), d \geq 2$. There exist $c_1, c_2$ such that for all $n$ sufficiently large and all $x \in C_n \setminus C_{n/2}$,

$$ \mathbb{P}^x \{ S_{t+1:n} \in C_{n/2} \} \geq c_1 n^{-d}, \quad (6.12) $$

and if $x \in \partial C_n$,

$$ \mathbb{P}^x \{ S_{t+1:n} \in C_{n/2} \} \leq c_1 n^{-d}, \quad (6.13) $$

**Proof.** We will do the proof for $d \geq 3$; the proof for $d = 2$ is identical replacing the Green’s function with the potential kernel. It follows from (6.9) that there exist $R, c$ such that for all $n$ sufficiently large and all $y \in C_{n/2} \cap \partial C_n$,

$$ G(y) - G(z) \geq c n^{-d}. \quad (6.14) $$

By choosing $R$ larger if necessary, we can assume that $\partial C_n \cap C_{n/2} = \emptyset$. By choosing $n$ larger if necessary, we can assume that $\partial C_{n/2} \cap C_{n/4} = \emptyset$.

Suppose that $x \in C_{n/4}$ and let $T = \tau_{C_n \setminus C_{n/2}}$. Applying the optional sampling theorem to the bounded martingale $G(S_{t+T})$, we see that

$$ G(x) = \mathbb{E}^x [G(S_T)] \leq \mathbb{E}^x [G(S_T); S_T \in C_{n/2}] + \max_{z \in C_{n/2}} G(z). $$

Therefore (6.14) implies that

$$ \mathbb{E}^x \{ G(S_T); S_T \in C_{n/2} \} \geq c_n n^{-d}. $$

If $n$ is sufficiently large, then $G(S_T) \not\in C_{n/2}$ and hence (6.9) gives

$$ \mathbb{E}^x \{ G(S_T); S_T \in C_{n/2} \} \leq c_n n^{-d} \mathbb{P}^x \{ \tau_{C_n \setminus C_{n/2}} < \tau_{C_n} \}. $$

This establishes (6.12) for $x \in C_{n/2}$.

To prove (6.12) for other $x$ we note the following fact that follows from our assumptions: there is an $\epsilon > 0$ such that for all $|y| \geq 1$, there is a $y$ with $p(y) \geq \epsilon$ and $J(x+y) \leq J(x) - \epsilon$. It follows that there is a $\delta > 0$ such that for all $n$ sufficiently large and all $x \in C_n$, there is probability at least $\delta$ that a random walk starting at $x$ reaches $C_{n/2}$ before leaving $C_n$.

It suffices to prove (6.13) for $x \in C_n \setminus C_{n/2}$. For such $x$,

$$ G(x) = C_d n^{-d} + O(n^{-d}). $$

Also,

$$ \mathbb{E}^x \{ G(S_T) \mid S_T \in C_{n/2} \} = C_{d/2} n^{-d} + O(n^{-d}), $$

$$ \mathbb{E}^x \{ G(S_T) \mid S_T \in \partial C_n \} = C_d n^{-d} + O(n^{-d}). $$

The optional sampling theorem gives

$$ G(x) = \mathbb{E}^x [G(S_T)] \geq \mathbb{P}^x \{ S_T \in C_{n/2} \} \mathbb{E}^x [G(S_T) \mid S_T \in C_{n/2}] + \mathbb{P}^x \{ S_T \in \partial C_n \} \mathbb{E}^x [G(S_T) \mid S_T \in \partial C_n]. $$

The left-hand side equals $C_d n^{-d} + O(n^{-d})$ and the right-hand side equals

$$ C_d n^{-d} + O(n^{-d}) + C_{d/2} n^{-d} - 1 \cdot n^{-d} \mathbb{P}^x \{ S_T \in C_{n/2} \}. $$

Therefore $\mathbb{P}^x \{ S_T \in C_{n/2} \} \sim O(n^{-1})$. □

**Lemma 6.3.5** If $p \in P(A), d \geq 2$ and $x \in C_n$,

$$ G_{C_n}(0, x) = C_d \left[ J(x)^{d/2} - n^{1-d} \right] + O(|x|^{-d}), \quad d \geq 3, $$

$$ G_{C_n}(0, x) = C_2 \log n - \log J(x) + O(|x|^{-d}), \quad d = 2. $$

In particular, for every $0 < \epsilon < 1/2$, there exist $c_1, c_2$ such that for all $n$ sufficiently large,

$$ c_1 n^{-d} \leq G_{C_n}(y, z) \leq c_1 n^{-d}, \quad y \in C_n, \quad z \in \partial C_{2n} \cup \partial C_{2m}. $$
6.3. DIFFERENCE ESTIMATES AND HARNACK'S INEQUALITY

Proof. Symmetry and Lemma 4.6.2 tell us that
\[ G_{C_n}(0, z) = G_{C_n}(x, 0) = G(x, 0) - \mathbb{P}^x[\mathcal{S}_{S_{R_n}}], \quad d \geq 3, \]
\[ G_{C_n}(0, x) = G_{C_n}(x, 0) = \mathbb{P}^x[\mathcal{S}_{S_{R_n}}] - a(x), \quad d = 2. \]  
(6.15)

Also, (6.9) and (6.10) give
\[ G(x) = C_d \mathbb{J}(x)^{d-d} + O(\|x\|^{d-1}), \quad d \geq 3, \]
\[ a(x) = C_d \log(\mathbb{J}(x)) + \gamma_d + O(\|x\|^{d-1}), \]
and Proposition 6.3.1 implies that
\[ \mathbb{P}^x[\mathcal{S}_{S_{R_n}}] = \frac{C_d}{n^{d-1}} + O(n^{d-1}), \quad d \geq 3, \]
\[ \mathbb{P}^x[\mathcal{S}_{S_{R_n}}] = C_d \log n + \gamma_d + O(n^{d-1}), \quad d = 2. \]
Since \( |x| \leq c_n \), we can write \( O(\|x\|^{d-1}) + O(\|x\|^{d-1}) \leq O(\|x\|^{d-1}). \) To get the final assertion we use the estimate
\[ G_{C_n}(y, x) \leq G_{C_n}(x, y) \leq G_{C_n}(y, y). \]

\[ \square \]

Lemma 6.3.6 If \( p \in P_\delta \), it \( x \in B \subset A \supseteq \mathbb{Z}^d \), \( y \in \partial A \),
\[ H_A(x, y) = \sum_{z \in B} G_A(x, z) \mathbb{P}^z[\mathcal{S}_{S_{R_n}} = y] = \sum_{z \in B} G_A(z, z) \mathbb{P}^z[\mathcal{S}_{S_{R_n}} = z]. \]

In particular,
\[ H_A(x, y) = \sum_{z \in A} G_A(x, z) p(z, y) = \sum_{z \in B \subset A} G_A(x, z) p(z, y). \]

Proof. The first equality follows immediately from a last-exit decomposition (Proposition 4.6.4). The second inequality uses the symmetry of \( p \). The final assertion is the particular case \( B = A \).

\[ \square \]

Lemma 6.3.7 If \( p \in P_\delta \), there exist \( c_1, c_2 \) such that for all \( n \) sufficiently large and all \( x \in \mathcal{C}_{1/n}, y \in \partial \mathcal{C}_n, \)
\[ \frac{c_1}{n^{d-1}} \leq H_{C_n}(x, y) \leq \frac{c_2}{n^{d-1}}. \]

CHAPTER 6. POTENTIAL THEORY

Proof. By the previous lemma,
\[ H_{C_n}(x, y) = \sum_{\omega \in \mathcal{C}_n} G_{C_n}(z, x) \mathbb{P}^z[\mathcal{S}_{S_{R_n}} = z]. \]

Using Lemma 6.3.5 we see that for \( z \in \partial \mathcal{C}_{1/n}, x \in \mathcal{C}_n, G_{C_n}(z, x) \approx n^{-d} \). Also, Lemma 6.3.4 implies that
\[ \sum_{z \in \mathcal{C}_n} \mathbb{P}^z[\mathcal{S}_{S_{R_n}} = z] \approx n^{-1}. \]

\[ \square \]

Theorem 6.3.8 (Difference estimates) If \( p \in P_\delta \) and \( r < \infty \), there exists \( c \) such that the following holds for every \( n \) sufficiently large.

(a) If \( g : \mathbb{R}^d \rightarrow \mathbb{R} \) is harmonic in \( B_n \) and \( |g| \leq r \),
\[ |\nabla g(0)| \leq c \|g\|_\infty n^{-d}, \]  
(6.16)
\[ |\nabla^2 g(0)| \leq c \|g\|_\infty n^{-2}, \]  
(6.17)

(b) If \( f : \overline{B}_n \rightarrow [0, \infty) \) is harmonic in \( B_n \) and \( |f| \leq r \), then
\[ |\nabla f(0)| \leq c f(0) n^{-d}, \]  
(6.18)
\[ |\nabla^2 f(0)| \leq c f(0) n^{-2}. \]  
(6.19)

Proof. Choose \( c > 0 \) such that \( C_{1/n} \subset B_n \). Choose \( n \) sufficiently large so that \( B_n \subset \mathcal{C}_{(n+1)/2m} \) and \( \partial \mathcal{C}_m \cap \mathcal{C}_m = \emptyset \). Let \( H(x, z) = H_{\mathcal{C}_{1/n}}(x, z) \). Then for \( |x| = r \)
\[ g(z) = \sum_{z \in \mathcal{C}_{1/n}} H(z, x) g(x). \]

Hence it suffices to establish (6.18) and (6.19) for \( f(x) = H(x, z) \) (with \( c \) independent of \( n, z \)). Let \( p = \rho_{C_n} = \tau_{\mathcal{C}_{1/n}} G(y, x) = G_{\mathcal{C}_{1/n}}(y, x) \). By Lemma 6.3.6, if \( x \in \mathcal{C}_{(n+1)/2}, \)
\[ f(x) = \sum_{y \in \mathcal{C}_{1/n}} G(y, x) \mathbb{P}^y[\mathcal{S}_{S_{R_n}} = y]. \]

Also, Lemma 6.3.7 shows that \( f(x) \approx n^{-d} \). The estimates (6.18) and (6.19) now follow from Lemma 6.3.3. Note also that we have established
\[ f(z) \leq c f(w), \quad z, w \in \mathcal{C}_{(n+1)/2}. \]  
(6.20)

\[ \square \]
6.4. FURTHER ESTIMATES

Theorem 6.3.9 (Harnack inequality) Suppose \( p \in \mathcal{P}_d \), \( U \subset \mathbb{R}^d \) is open and connected, and \( K \) is a compact subset of \( U \). Then there exist constants \( c = c(K, U, p) < \infty \) and positive integer \( N = N(K, u, p) \) such that if \( n \geq N \),
\[
U_n = \{ x \in \mathbb{R}^d : n^{-1} x \in U \}, \quad K_n = \{ x \in \mathbb{R}^d : n^{-1} x \in K \},
\]
and \( f : \overline{U_n} \to [0, \infty) \) is harmonic in \( U_n \), then
\[
f(x) \leq c f(y), \quad x, y \in K_n.
\]

Proof. Without loss of generality we will assume that \( U \) is bounded. In (6.20) we showed that there exists \( \delta > 0 \), \( c_0 < \infty \) such that
\[
f(x) \leq c_0 f(y) \quad \text{if} \quad |x - y| \leq \delta \operatorname{dist}(x, \partial U_n).
\]

Let us call two points \( z, w \) in \( U \) adjacent if \( |z - w| < (\delta/2) \max\{ \operatorname{dist}(z, \partial U), \operatorname{dist}(w, \partial U) \} \). Let \( p \) denote the graph distance associated to this adjacency; i.e., \( p(z, w) \) is the minimum \( k \) such that there exists a sequence \( z = z_0, z_1, \ldots, z_k = w \) of points in \( U \) such that \( z_j \) is adjacent to \( z_{j+1} \) for \( j = 1, \ldots, k \). Fix \( z \in U \), and let \( V_k = \{ w \in U : p(z, w) \leq k \} \), \( U_n, K \subset \{ x \in \mathbb{R}^d : n^{-1} x \in V_k \} \). For \( k \geq 1 \), \( V_k \) is open and connectedness of \( U \) implies that \( \bigcup V_k = U \). For \( x, y \in U_n, k \), repeated application of (6.21) gives \( f(x) \leq C_k f(y) \). Compactness of \( K \) implies that \( K \subset V_k \) for some finite \( k \), and hence \( K_n \subset U_n, k \).

\[\square\]

6.4 Further estimates

In this section we will collect some more facts about random walks in \( \mathcal{P}_d \) restricted to the set \( C_n \). The first two lemmas are similar to Lemma 6.3.4.

Lemma 6.4.1 If \( p \in \mathcal{P}_d \), \( m < n \), \( T = \tau_{C_n} \), then for \( z \in C_n \setminus C_m \),
\[
\mathbb{P}^z[S_T \in \partial C_n] = \frac{\log m - \log n + O(m^{-1})}{\log m}.
\]

Proof. Let \( q = \mathbb{P}^z[S_T \in \partial C_n] \). The optional sampling theorem applied to the bounded martingale \( M_t = a(S_T) \) gives
\[
a(x) = \mathbb{E}^x[a(S_T)] = \mathbb{E}^x[a(S_T) \mid S_T \in \partial C_n] + q \mathbb{E}^x[a(S_T) \mid S_T \in \partial C_n],
\]
From (6.10) we know that
\[
a(x) = C_d \log |x| + \gamma_2 + O(|x|^{-1}),
\]
\[
\mathbb{E}^x[a(S_T) \mid S_T \in \partial C_n] = C_1 \log m + \gamma_2 + O(m^{-1}),
\]
\[
\mathbb{E}^x[a(S_T) \mid S_T \in \partial C_n] = C_1 \log n + \gamma_2 + O(n^{-1}).
\]
Solving for \( q \) gives the lemma.
\[\square\]

Lemma 6.4.2 If \( p \in \mathcal{P}_d, d \geq 3, T = \tau_{C_n} \), then for \( x \in \mathbb{Z}^d \setminus C_n \),
\[
\mathbb{P}^x[T < \infty] = \left( \frac{m}{\mathcal{J}(x)} \right)^{\frac{2d}{d-2}} \left[ 1 + O(m^{-1}) \right].
\]

Proof. Since \( G(y) \) is a bounded harmonic function on \( \tau_{C_n} \), with \( G(\infty) = 0 \), (6.5) gives
\[
G(x) = \mathbb{E}^x[G(S_T)] ; T < \infty = \mathbb{P}^x[T < \infty] \mathbb{E}^x[G(S_T) \mid T < \infty].
\]
But (6.9) gives
\[
G(x) = C_d \mathcal{J}(x)^{\alpha-1} \left[ 1 + O(|x|^{-2}) \right],
\]
\[
\mathbb{E}^x[G(S_T) \mid T < \infty] = C_d \mathcal{J}(x)^{\alpha-1} \left[ 1 + O(|x|^{-2}) \right].
\]

Suppose \( m \leq n/2 \) and \( x \in C_n \), \( z \in \partial C_n \). Then Theorem 6.3.8 tells us that (for \( n \) sufficiently large)
\[
H_{C_n}(x, z) = H_{C_n}(0, z) \left[ 1 + O\left( \frac{n}{m} \right) \right].
\]

We will use this in the next two propositions to estimate some conditional probabilities.

Proposition 6.4.3 Suppose \( p \in \mathcal{P}_d, d \geq 3, m < n/4 \), and \( C_n \setminus C_m \subset A \subset C_n \). Suppose \( x \in C_n \) with \( \mathbb{P}^x(S_{S_n} \in \partial C_n) > 0 \) and \( z \in \partial C_n \). Then for \( n \) sufficiently large,
\[
\mathbb{P}^x(S_{S_n} = z \mid S_{S_n} \in \partial C_n) = H_{C_n}(0, z) \left[ 1 + O\left( \frac{n}{m} \right) \right].
\]

Proof. It is easy to check that it suffices to verify (6.23) for \( x \in \partial C_{2m} \). Note that (6.22) implies
\[
\mathbb{P}^x(S_{S_n} = z) = H_{C_n}(0, z) \left[ 1 + O\left( \frac{n}{m} \right) \right]
\]
and
\[
\mathbb{P}^x(S_{S_n} = z \mid S_{S_n} \notin \partial C_n) = H_{C_n}(0, z) \left[ 1 + O\left( \frac{n}{m} \right) \right].
\]

Using Lemma 6.4.2, we can see there is a \( c \) such that
\[
\mathbb{P}^x(S_{S_n} \in \partial C_n) \geq c, \quad x \in \partial C_{2m},
\]
\[\square\]

For \( d = 2 \) we get a similar result but with a slightly larger error term.

Proposition 6.4.4 Suppose \( p \in \mathcal{P}_d, m < n/4 \), and \( C_n \setminus C_m \subset A \subset C_n \). Suppose \( x \in C_n \) with \( \mathbb{P}^x(S_{S_n} \in \partial C_n) > 0 \) and \( z \in \partial C_n \). Then for \( n \) sufficiently large,
\[
\mathbb{P}^x(S_{S_n} = z \mid S_{S_n} \in \partial C_n) = H_{C_n}(0, z) \left[ 1 + O\left( \frac{m \log(n/m)}{n} \right) \right].
\]

[6.4.4]
Proof. The proof is essentially the same, except for the last step where Lemma 6.4.1 gives us
\[ P^*\{S_{r_{A}} \in \partial \mathcal{C}_{A}\} \geq \frac{c}{\log(n/m)}, \quad x \in \mathcal{C}_{A}. \]
\[ \square \]

The next proposition is a stronger version of Proposition 6.2.2. Here we show that the boundedness assumption of that proposition can be replaced with an assumption of sublinearity.

Proposition 6.4.5 Suppose \( p \in \mathcal{P}_d \), \( d \geq 3 \) and \( A \subset \mathbb{Z}^d \) with \( \mathbb{Z}^d \setminus \overline{A} \) finite. Suppose \( f : \mathbb{Z}^d \to \mathbb{R} \) is harmonic on \( A \) and satisfies \( f(z) = o(|z|) \) as \( x \to \infty \). Then there exists \( b \in \mathbb{R} \) such that for \( x \in \mathcal{A} \),
\[ f(x) = \mathbb{E}^*[f(S_{r_{A}}) \mid \tau_{A} < \infty] + bP^*\{\tau_{A} = \infty\}. \]
Proof. Without loss of generality, we may assume that \( 0 \notin A \) and \( f \equiv 0 \) on \( \mathbb{Z}^d \setminus \overline{A} \); otherwise, we can consider
\[ f(z) = f(z) - \mathbb{E}^*[f(S_{r_{A}}) \mid \tau_{A} < \infty]. \]
The assumptions imply that there is a sequence of real numbers \( \epsilon_n \) decreasing to 0 such that \( |f(x)| \leq \epsilon_n, n \) for all \( x \in \mathcal{C}_A \) and hence
\[ |f(x) - f(y)| \leq 2 \epsilon_n, \quad x, y \in \partial \mathcal{B}_n. \]
By (6.8), we can see that
\[ 0 = f(0) = \mathbb{E}f(S_{r_{A}}) - \sum_{y \in \mathcal{B}_{n+1}} G_{r_{A}}(0, y) Lf(y), \]
which implies that
\[ \lim_{n \to \infty} \mathbb{E}f(S_{r_{A}}) = b := \sum_{y \in \mathcal{B}_{n+1}} G(0, y) Lf(y). \]
If \( x \in A \) and \( |x| < n \), the optional sampling theorem implies that
\[ f(x) = \mathbb{E}^*[f(S_{r_{A}}) \mid S_{r_{A}} \geq \mathcal{C}_A, \tau_{A} > \epsilon_n]. \]
For \( n \) large, (6.23) implies that
\[ \mathbb{E}^*[f(S_{r_{A}}) \mid \tau_{A} > \epsilon_n] = \mathbb{E}f(S_{r_{A}}) \leq \epsilon_n, \quad n \to \infty \]
Therefore,
\[ f(x) = \lim_{n \to \infty} P^*\{\tau_{A} > \epsilon_n\} \mathbb{E}^*[f(S_{r_{A}}) \mid \tau_{A} > \epsilon_n] \]
\[ = P^*\{\tau_{A} = \infty\} \lim_{n \to \infty} \mathbb{E}f(S_{r_{A}}) = bP^*\{\tau_{A} = \infty\}. \]
\[ \square \]

In Proposition 4.6.3 we proved the following proposition.

Proposition 6.4.6 Suppose \( p \in \mathcal{P}_d \) and \( A \) is a finite subset of \( \mathbb{Z}^d \) containing the origin. Let \( T = T_{A} = \inf\{j \geq 0 : S_{j} \notin A\} \), and
\[ \xi_n = \tau_{A} = \inf\{j \geq 1 : S_{j} \notin \mathcal{C}_{A}\}. \]
Then for each \( x \in \mathbb{Z}^d \) the limit
\[ g_{A}(x) := \lim_{n \to \infty} C_{2}(\log n) P^{*}[\xi_n < T]\]
exists. Moreover, if \( y \in A \),
\[ g_{A}(x) = a(x) = \mathbb{E}^*[a(S_{T} - y)]. \]
Let
\[ \xi_n = \tau_{A} = \inf\{j \geq 1 : S_{j} \notin \mathcal{C}_{A}\}. \]
We will assume that \( x \in \mathcal{C}_{A} \). The optional sampling theorem applied to the martingale \( M_{j} = a(S_{j}, T_{A} - y) \) implies
\[ a(x) = \mathbb{E}a(S_{T} - y) = P^{*}[\xi_n < T] \mathbb{E}[a(S_{T} - y) \mid \xi_n < T] + \mathbb{E}[a(S_{T} - y) \mid \xi_n < T]. \]
As \( n \to \infty \),
\[ \mathbb{E}[a(S_{T} - y) \mid \xi_n < T] \sim C_{2} \log n. \]
Letting \( n \to \infty \), we get the result.
\[ \square \]

Remark. As mentioned before, it follows that the right-hand side of (6.28) is the same for all \( y \in A \). Also, since there exists \( \epsilon \) such that \( C_{1} \mathcal{C}_{A} \subset C_{2} \mathcal{C}_{A} \) we can replace (6.27) with
\[ g_{A}(x) := \lim_{n \to \infty} C_{2}(\log n) P^{*}[\xi_n < T]\]
where \( \xi_n = \inf\{j \geq 1 : |S_{j}| \geq n\} \).

Proposition 6.4.7 Suppose \( p \in \mathcal{P}_d \) and \( A \) is a finite subset of \( \mathbb{Z}^d \) containing the origin. Suppose \( f : \mathbb{Z}^d \to \mathbb{R} \) is harmonic on \( \mathbb{Z}^d \setminus \overline{A} \); vanishes on \( A \); and satisfies \( f(z) = o(|z|) \) as \( |x| \to \infty \). Then \( f = b f_{A} \) for some \( b \in \mathbb{R} \).
Proof. Without loss of generality, assume \( 0 \notin A \) and let \( T_{A} \) be as in the previous proposition. Using (6.8) and (6.11), we get
\[ \mathbb{E}f(S_{r_{A}}) = \sum_{y \in A} G_{r_{A}}(0, y) Lf(y) = C_{2} \log n \sum_{y \in A} Lf(y) + O(1). \]
(Here and below the error terms may depend on \( A \).) Similarly to (6.26) we use (6.24) to see that
\[ \mathbb{E}^*[f(S_{r_{A}}) \mid \xi_n < T] - \mathbb{E}f(S_{r_{A}}) \leq c |x| \log n \sup_{y \in \partial \mathcal{C}_{A}} \|f(y) - f(z)\| \leq c |x| \log n. \]
6.5. **Capacity, Transient Case**

If \( A \) is a finite subset of \( \mathbb{Z}^d \), we let

\[
T_A = \tau_{\mathbb{Z}^d \backslash A}, \quad \overline{T}_A = \tau_{\mathbb{Z}^d \backslash \partial A},
\]

rad \( (A) = \sup \{ |x| : x \in A \} \),

If \( p \in \mathcal{P}_d, d \geq 3 \), let

\[
E_A(x) = P^x(T_A = \infty), \quad g_A(x) = P^x(\overline{T}_A = \infty).
\]

Note that \( E_A(x) = 0 \) if \( x \in A \). If \( d \geq 3 \), then \( g_A \) is the unique function on \( \mathbb{Z}^d \) that is zero on \( A \); harmonic on \( \mathbb{Z}^d \backslash A \); and satisfies \( g_A(x) \sim 1 \) as \( |x| \to \infty \). Moreover, if \( x \in A \), then \( E_A(x) = L g_A(x) \), we define the capacity of a finite set \( A \) by

\[
\text{cap}(A) = \sum_{x \in \mathbb{Z}^d} E_A(x) = \sum_{x \in A} E_A(x) = \sum_{x \in \mathbb{Z}^d} L g_A(x).
\]

This definition may seem unnatural, but as the next proposition shows, \( \text{cap}(A) \) is a measure of how large a set is in terms of the probability that a random walk starting from the set hits every other set.

**Proposition 6.5.1** If \( p \in \mathcal{P}_d, d \geq 3 \), \( A \subset \mathbb{Z}^d \) is finite, then

\[
P^x(T_A < \infty) = \frac{C_d \text{cap}(A)}{|x|^{d-2}} \left[ 1 + O\left( \frac{\text{rad}(A)}{|x|} \right) \right].
\]

**Proof.** By Proposition 4.6.4,

\[
P^x(\overline{T}_A < \infty) = \sum_{y \in \mathbb{Z}^d} G_A(x, y) E_A(y).
\]

For \( y \in A \), \( J(x - y) = J(x) + O(|y|) \). Therefore,

\[
G_A(x, y) = C_d J(x)^{d-2} + O(|y||x|^{d-2}) = \frac{C_d}{J(x)^{d-2}} \left[ 1 + O\left( \frac{\text{rad}(A)}{|x|} \right) \right].
\]

6.6. **Chapter 6. Potential Theory**

**Proposition 6.6.2** If \( p \in \mathcal{P}_d, d \geq 3 \),

\[
\text{cap}(A) = C_d^{-1} n^{d-2} + O(n^{d-2}).
\]

**Proof.** By Proposition 4.6.4,

\[
1 = P(\overline{T}_A < \infty) = \sum_{y \in \mathbb{Z}^d} G_A(0, y) E_A(y).
\]

But for \( y \in \partial \mathbb{C}_n \),

\[
G_A(0, y) = C_d n^{d-2} + O(n^{d-2}).
\]

Recall that \( \xi_n = \inf \{ j \geq 1 : S_j \notin \mathbb{C}_n \} \) and let \( T_{\mathbb{C}_n} = T_A \land \xi_n \). Note that if \( x \in A \subset \mathbb{C}_n \),

\[
P^x(T_A > \xi_n) = \sum_{y \in \mathbb{Z}^d} P^y(S_{T_{\mathbb{C}_n}} = y) = \sum_{y \in \mathbb{Z}^d} P^y(S_{T_{\mathbb{C}_n}} = x).
\]

Therefore,

\[
\sum_{y \in \mathbb{Z}^d} P^y(T_A > \xi_n) = \sum_{y \in \mathbb{Z}^d} P^y(S_{T_{\mathbb{C}_n}} = x) = \sum_{y \in \mathbb{Z}^d} P^y(T_A < \xi_n).
\]

Also,

\[
\text{cap}(A) = \frac{1}{n^{d-2}} \sum_{y \in \mathbb{Z}^d} P^y(T_A > \xi_n) = \sum_{y \in \mathbb{Z}^d} P^y(T_A < \xi_n).
\]

**Proposition 6.6.3** If \( p \in \mathcal{P}_d, d \geq 3 \), and \( A, B \) are finite subsets of \( \mathbb{Z}^d \), then

\[
\text{cap}(A) \leq \text{cap}(A \cup B) + \text{cap}(A \cap B).
\]

**Proof.** Choose \( n \) such that \( A \cup B \subset \mathbb{C}_n \). Then for \( y \in \mathbb{C}_n \),

\[
P^y(T_{A \cup B} < \xi_n) = P^y(T_A < \xi_n \lor T_B < \xi_n) = P^y(T_A < \xi_n) + P^y(T_B < \xi_n) - P^y(T_A < \xi_n, T_B < \xi_n) \leq P^y(T_A < \xi_n) + P^y(T_B < \xi_n) - P^y(T_{A \cap B} < \xi_n).
\]

The proposition follows from (6.29).

**Definition.** If \( p \in \mathcal{P}_d, d \geq 3 \), and \( A \subset \mathbb{Z}^d \) is finite, the harmonic measure of \( A \) (from infinity) is defined by

\[
b_m \mu_A(x) = \frac{E_A(x)}{\text{cap}(A)} \quad x \in A.
\]

Note that \( b_m \mu_A \) is a probability measure supported on \( \partial A \). As the next proposition shows, it is can be considered as the hitting measure of \( A \) by a random walk "started at infinity conditioned to hit \( A \)."
6.5. CAPACITY, TRANSIENT CASE

Proposition 6.5.4 If \( p \in \mathcal{P}_d, d \geq 3, \) and \( A \subset \mathbb{Z}^d \) is finite, then for \( x \in A, \)
\[
\lim_{n \to \infty} P^n \{ S_n = x \mid T_A < \infty \} = \lim_{n \to \infty} \mathbb{P}^x \{ T_A = x \mid T_A < \infty \}.
\]

In fact, if \( A \subset \mathcal{C}_{d/2} \) and \( y \notin \mathcal{C}_n, \)
\[
P^n \{ T_A = x \mid T_A < \infty \} = \sum_{z \in \mathcal{C}_n} G_{\mathbb{Z}^d}(y, z) P^n \{ S_n = z \}.
\]

Proof. If \( A \subset \mathcal{C}_n \) and \( y \notin \mathcal{C}_n, \) the last-exit decomposition (Proposition 4.6.4) gives
\[
P^n \{ T_A = x \mid T_A < \infty \} = \sum_{z \in \mathcal{C}_n} G_{\mathbb{Z}^d}(y, z) P^n \{ S_n = z \}.
\]

But (6.23) gives
\[
P^n \{ S_n = x \mid T_A = \infty \} = P^n \{ S_n = x \mid T_A < \infty \} = \left[ 1 + O \left( \frac{\text{rad}(A)}{n} \right) \right],
\]

Therefore,
\[
P^n \{ T_A = x \mid T_A < \infty \} = \mathbb{E}_x \left[ 1 + O \left( \frac{\text{rad}(A)}{n} \right) \right] \sum_{z \in \mathcal{C}_n} G_{\mathbb{Z}^d}(y, z) H_{\mathbb{Z}^d}(0, z),
\]

and by summing over \( x, \)
\[
P^n \{ T_A < \infty \} = \text{cap}(A) \left[ 1 + O \left( \frac{\text{rad}(A)}{n} \right) \right] \sum_{z \in \mathcal{C}_n} G_{\mathbb{Z}^d}(y, z) H_{\mathbb{Z}^d}(0, z).
\]

Proposition 6.5.5 If \( p \in \mathcal{P}_d, d \geq 3, \) and \( A \subset \mathbb{Z}^d \) is finite, then
\[
\text{cap}(A) = \sup \sum_{x \in A} f(x),
\]

where the sum is over all functions \( f \geq 0 \) supported on \( A \) such that \( Gf \leq 1 \) for all \( y \in \mathbb{Z}^d. \)

Proof. Let \( f(x) = \mathbb{E}_x \mathbb{P} \mathbb{P} \), Note that Proposition 4.6.4 implies that for \( y \in A, \)
\[
1 \geq P^n \{ T_A < \infty \} = \sum_{x \in A} G(y, x) \mathbb{E}_x(x),
\]

Hence \( Gf \leq 1 \) and the supremum is at least as large as \( \text{cap}(A) \). Note also that \( Gf \) is the unique bounded function on \( \mathbb{Z}^d \) that is harmonic on \( \mathbb{Z}^d \setminus A \) equals 1 on \( A \); and approaches 0 at infinity. Suppose \( f \geq 0 \) with \( Gf(y) \leq 1 \) for all \( y \in \mathbb{Z}^d. \) Then \( Gf \) is the unique bounded function on \( \mathbb{Z}^d \) that is harmonic on \( \mathbb{Z}^d \setminus A \); equals \( Gf \leq 1 \) on \( A \); and approaches zero at infinity. By the maximum principle, \( Gf(y) \leq Gf(y) \) for all \( y \). In particular, \( Gf \) is harmonic on \( \mathbb{Z}^d \setminus A \); and nonnegative on \( A \); and approaches zero at infinity.

If \( x, y \in A, \) let
\[
K(x, y) = P^n \{ S_n = x \mid T_A = \infty \}.
\]

Note that \( K(x, y) = K(x, y) \) and
\[
\sum_{y \in A} K(x, y) = 1 - \text{E}_x(A).
\]

If \( h \) is a bounded function on \( \mathbb{Z}^d \) that is harmonic on \( \mathbb{Z}^d \setminus A \), and has \( h(\infty) = 0 \), then \( h(x) = \mathbb{E}_x(h(y)) \mathbb{E}_x(A) \mathbb{E}_x \mathbb{P} \mathbb{P} \), Using this one can easily check that for \( x \in A, \)
\[
\mathcal{L}h(x) = \sum_{y \in A} K(x, y) h(y) - h(x).
\]

Also, if \( h \geq 0, \)
\[
\sum_{x \in A} \sum_{y \in A} K(x, y) h(y) = \sum_{y \in A} h(y) \sum_{x \in A} K(x, y) \geq \sum_{y \in A} h(y) [1 - \text{E}_x(A)] \leq \sum_{y \in A} h(y),
\]

which implies
\[
\sum_{x \in A} \mathcal{L}h(x) = \sum_{x \in A} \mathcal{L}h(x) \leq 0,
\]

Then,
\[
\sum_{x \in A} f(x) = -\sum_{x \in A} \mathcal{L}Gf(x) \leq -\sum_{x \in A} \mathcal{L}Gf(x) - \sum_{x \in A} \mathcal{L}(Gf - f)|x| = \sum_{x \in A} \text{E}_x(x).
\]

Our definition of capacity depends on the random walk \( p \). The next proposition shows that capacities for different \( p \)'s in the same dimension are comparable.

Proposition 6.5.6 Suppose \( p, q \in \mathcal{P}_d, d \geq 3 \) and let \( \text{cap}_p, \text{cap}_q \) denote the corresponding capacities. Then there is a \( \delta = \delta(p, q) > 0 \) such that for all finite \( A \subset \mathbb{Z}^d, \)
\[
\delta \text{cap}_p(A) \leq \text{cap}_q(A) \leq \delta^{-1} \text{cap}_p(A).
\]

\[\square\]
6.5. CAPACITY, TRANSIENT CASE

Proof. It follows from Theorem 4.3.1 that there exists \( \delta \) such that
\[
\delta G_d(x, y) \leq G_d(x, y) \leq \delta^{-1} G_d(x, y),
\]
for all \( x, y \). The proposition then follows immediately from Proposition 6.5.5.

Definition. If \( p \in \mathcal{P}_d \), \( d \geq 3 \), and \( A \subset \mathbb{Z}^d \), then \( A \) is transient if
\[
P\{S_n \in A \mid i.o.\} = 0.
\]
Otherwise, the set is called recurrent.

Lemma 6.5.7 If \( p \in \mathcal{P}_d \), \( d \geq 3 \), and \( A \subset \mathbb{Z}^d \), then a set is recurrent if and only if for every \( x \in \mathbb{Z}^d \)
\[
P^x\{S_n \in A \mid i.o.\} = 1,
\]

Proof. Let \( F(y) = \mathbb{P}^y\{S_n \in A \mid i.o.\} \). Suppose there exists \( y \in A \) and \( \epsilon > 0 \) such that \( F(y) \geq \epsilon \). Then,
\[
\mathbb{E}[F(S_n)] \geq \epsilon.
\]
Then by coupling \( \{X_n\} \) we can see that for all \( x \), \( F(x) = \mathbb{E}[F(S_n)] = \epsilon + \frac{1}{2} |O(n^{1/2})| \), and hence \( F(x) \geq \epsilon \). In particular, for every \( x \) there is an \( N_x \) such that
\[
P^x\{S_n \in A \mid i.o.\} \geq \epsilon/2.
\]
By iterating this we can see for all \( x \)
\[
P^x\{S_n \in A \mid i.o.\} \geq \epsilon/2.
\]
and the lemma follows easily.

Clearly, all finite sets are recurrent. More generally, if \( A \) is a subset such that
\[
\sum_{x \in A} G(x) < \infty, \tag{6.30}
\]
then \( A \) is transient. To see this, let \( S_n \) be a random walk starting at the origin and let \( V \) denote the number of visits to \( A \),
\[
V_A = \sum_{j=0}^{\infty} 1\{S_n \in A\}.
\]
Then \( 6.30 \) implies that \( \mathbb{E}[V_A] < \infty \) which implies that \( \mathbb{P}\{V_A < \infty \} = 1 \). In Exercise 6.2, it is shown that the converse is not true; i.e., there exist transient sets \( A \) with \( \mathbb{E}[V_A] = \infty \).

Lemma 6.5.8 Suppose \( p \in \mathcal{P}_d \), \( d \geq 3 \), and \( A \subset \mathbb{Z}^d \). Then \( A \) is transient if and only if
\[
\sum_{k=1}^{\infty} \mathbb{P}\{T^k < \infty \} < \infty, \tag{6.31}
\]
where \( T^k = T_{p(A \setminus C_A \cup C_A)} \).

Proof. Let \( E_k \) be the event \( \{T^k < \infty \} \). Since the random walk is transient, \( A \) is transient if and only if \( \mathbb{P}\{E_k \mid i.o.\} = 0 \). Hence the Borel-Cantelli Lemma implies that any \( A \) satisfying \( 6.31 \) is transient.

Suppose
\[
\sum_{k=1}^{\infty} \mathbb{P}\{T^k < \infty \} = \infty,
\]
Then either the sum over even \( k \) or the sum over odd \( k \) is infinite. We will assume the former; the argument if the latter holds is almost identical. Let \( B_{k+} = (A_k \setminus A_{k+1}) \cap \{(z^1, \ldots, z^d) : z^1 \geq 0 \} \) and \( B_{k-} = (A_k \setminus A_{k-1}) \cap \{(z^1, \ldots, z^d) : z^1 \leq 0 \} \). Since \( \mathbb{P}\{T^k < \infty \} \leq \mathbb{P}\{T_{B_{k+}} < \infty \} + \mathbb{P}\{T_{B_{k-}} < \infty \} \), we know that either
\[
\sum_{k=1}^{\infty} \mathbb{P}\{T_{B_{k+}} < \infty \} = \infty, \tag{6.32}
\]
or the same equality with \( B_{k+} \) replacing \( B_{k-} \). We will assume \( 6.32 \) holds and write \( a_k = T_{B_{k+}} \). An application of the Harnack principles shows that there is a \( c \) such that for all \( j \neq k \),
\[
\mathbb{P}\{\sigma_j < \infty \mid \sigma_j \land \sigma_k = \infty \} \leq c \mathbb{P}\{\sigma_j < \infty \}.
\]
This implies
\[
\mathbb{P}\{\sigma_j < \infty \mid \sigma_k < \infty \} \leq 2c \mathbb{P}\{\sigma_j < \infty \} \mathbb{P}\{\sigma_k < \infty \}.
\]
Using this and a special form of the Borel-Cantelli Lemma (Corollary 8.6.2) we can see that
\[
\mathbb{P}\{\sigma_j < \infty \mid i.o.\} > 0,
\]
which implies that \( A \) is not transient.

Corollary 6.5.9 (Wiener's test) Suppose \( p \in \mathcal{P}_d \), \( d \geq 3 \), and \( A \subset \mathbb{Z}^d \). Then \( A \) is transient if and only if
\[
\sum_{k=1}^{\infty} 2^{1-\beta d} \text{cap}(A_k) < \infty \tag{6.33}
\]
where \( A_k = A \cap (C_{p'} \setminus C_{p^+}) \). In particular, if \( A \) is transient for some \( p \in \mathcal{P}_d \), then it is transient for all \( p \in \mathcal{P}_d \).
6.5. CAPACITY, TRANSIENT CASE

Proof. By Proposition 6.5.1, we can see that 
P\{T^d < \infty\} \times 2^{d-1}\text{cap}(A_k),
\]

Theorem 6.5.10 Suppose \(d \geq 3\), \(p \in \mathcal{P}_d\), and \(S_n\) is a \(p\)-walk. Let \(A\) be the set of points visited by the random walk,

\[A = S(0, \infty) = \{S_n : n = 0, 1, \ldots\}.
\]

If \(d = 3, 4\), then with probability one \(A\) is a recurrent set. If \(d \geq 5\), then with probability one \(A\) is a transient set.

Proof. Since a set is transient if and only if all its translates are transient, we see that for each \(n\), \(A\) is recurrent if and only if the set

\[\{S_n - S_m : m = n, n + 1, \ldots\}\]

is recurrent. Therefore, the event \(\{A\text{ is recurrent}\}\) is a tail event, and the Kolmogorov 0-1 law implies that \(A\) is either transient or recurrent.

Let \(Y\) denote the random variable that equals the expected number of visits to \(A\) by an independent \(p\)-walk starting at the origin. In other words,

\[Y = \sum_{x \in A} G(x) = \sum_{x \in S_n} 1\{x \in A\} G(x).
\]

Then,

\[\mathbb{E}[Y] = \sum_{x \in S_n} \mathbb{P}\{x \in A\} G(x) = G(x) \sum_{x \in S_n} G(x)\]

Since \(G(x) \propto |x|^{-d}\), we have \(G(x) \propto |x|^{-d}\). By examining the sum, we see that \(\mathbb{E}[Y] < \infty\) for \(d = 3, 4\) and \(\mathbb{E}[Y] < \infty\) for \(d \geq 5\). If \(d \geq 5\), this gives \(Y < \infty\) with probability one which implies that \(A\) is transient with probability one.

We now focus on \(d = 4\). It suffices to show that \(\mathbb{P}\{A\text{ is recurrent}\} > 0\). Let \(S^1, S^2\) be independent \(p\)-walks with increment distribution \(P\) starting at the origin, and let

\[\sigma^d_x = \inf\{n : S^d_x \notin \mathcal{C}_x\}.
\]

Let

\[V^d_x = \{x \in \mathcal{C}_x : S^d_x = x \text{ for some } n \leq \sigma^d_x\}.
\]

Let \(E_x\) be the event \(V^d_x \cap V^d_y \neq \emptyset\). We will show that \(\mathbb{P}\{E_x \mid 0\} > 0\) which implies that \(\sigma^d_x\) with positive probability.

Using Corollary 6.6.2, we can see that to prove this it suffices to show that

\[\sum_{k=1}^{\infty} \mathbb{P}\{E_{k^d}\} = \infty,
\]

and that there exists a constant \(c < \infty\) such that for \(m < k\),

\[\mathbb{P}\{E_{m^d} \cap E_{m^d}\} \leq c\mathbb{P}\{E_{m^d}\} \mathbb{P}\{E_{k^d}\}.
\]

The event \(E_{m^d}\) depends only on the values of \(S^d_x\) with \(\sigma^d_x \leq n \leq \sigma^d_{x+1}\). Hence, the Harnack principle implies \(\mathbb{P}\{E_{m^d} \mid E_{m^d}\} \leq c\mathbb{P}\{E_{m^d}\}\) and hence (6.35) holds. To prove (6.34), let \(J^d(k, x)\) denote the indicator function of the event that \(S^d_x \subset x\) for some \(n \leq \sigma^d_x\). Then,

\[Z^d_k := \sum_{x \in \mathcal{C}_x} J^d(k, x) J^d(k, x).
\]

There exist \(c_1, c_2\) such that if \(x, y \in \mathcal{C}_x \setminus \mathcal{C}_y\),

\[\mathbb{E}[J^d(k, x) J^d(k, y)] \leq c(2^d)^{-1} \mathbb{E}[J^d(k, x) J^d(k, y)] \leq c(2^d)^{-1} \mathbb{E}[J^d(k, x) J^d(k, y)]\]

\[\mathbb{E}[Z^d_k] = \sum_{x \in \mathcal{C}_x} \mathbb{E}[J^d(k, x)] \mathbb{E}[J^d(k, x)] \geq (2^d)^{-1} \mathbb{E}[J^d(k, x) J^d(k, x)]
\]

The latter inequality is obtained by noting that the probability that a random walker hits both \(x\) and \(y\) given that it hits at least one of them is bounded above by the probability that a random walker starting at the origin visits \(y - x\) Therefore,

\[\mathbb{E}[Z^d_k] = \sum_{x \in \mathcal{C}_x} \mathbb{E}[J^d(k, x)] \mathbb{E}[J^d(k, x)] \geq c \sum_{x \in \mathcal{C}_x} (2^d)^{-1} \mathbb{E}[J^d(k, x) J^d(k, x)]
\]

\[\mathbb{E}[Z^d_k] \leq c \sum_{x \in \mathcal{C}_x} (2^d)^{-1} \frac{1}{|1 \setminus \{y : y - x\}|} \leq c
\]

Using the second moment estimate, Lemma 6.4.1, this implies that \(\mathbb{P}\{Z^d_k > 0\} \geq c/k\) which implies (6.34).

Heuristic note. The central limit theorem implies that the number of points in \(B_n\) visited by a random walk is of order \(n^2\). Roughly speaking, we can say that a random walk path is a "two-dimensional set." Asking whether or not this is recurrent is asking whether or not two random two-dimensional sets intersect. Using the example of planes in \(\mathbb{R}^3\), we can see that the critical dimension is four.

6.6. Capacity in two dimensions

The theory of capacity in two dimensions is somewhat similar to that for \(d \geq 3\), but there are significant differences caused by the recurrence of the random walk. If \(p \in \mathcal{P}_2\) and \(0 \in A \subset \mathbb{Z}^2\) is finite, we write

\[g_A(x) = a(x) = \mathbb{E}[a(S^d_x)]\]
Recall that \( g_A \) is the unique function on \( \mathbb{Z}^2 \) that vanishes on \( A \); it is harmonic on \( \mathbb{Z}^2 \setminus A \); and satisfies \( g_A(x) \sim C_2 \log J(x) - C_2 \log |x| \) as \( x \to \infty \). If \( y \in A \), we can also write
\[
g_A(x) = a(x - y) - E_x[a(S_{T_A} = y)].
\]
(This must give the same value by uniqueness.) For the rest of this section we will assume that \( 0 \in A \), but this is only for convenience. The optional sampling theorem implies that if \( x \in \mathcal{C}_n \),
\[
g_A(x) = E_x[g_A(S_{T_A} \wedge \infty)] = P_x'(\xi_n < T_A) | C_2 \log n | [1 + o_A(1)].
\]
In particular,
\[
g_A(x) = \lim_{n \to \infty} C_2 \log n P_x'(\xi_n < T_A).
\] (6.36)
Also, \( a(x) = g_A(x) \) is the unique bounded function on \( \mathbb{Z}^2 \) that is harmonic on \( \mathbb{Z}^2 \setminus A \) and has boundary value \( a \) on \( A \). We define the harmonic measure (from infinity) by
\[
h_m(x) = \lim_{n \to \infty} P_x'[S_{T_A} = x].
\] (6.37)
Since \( P_x'[T_A < \infty] = 1 \), this is the same as \( P_x'[S_{T_A} = x \mid T_A < \infty] \) and hence agrees with the definition of harmonic measure for \( d \geq 3 \). It is not immediately obvious that the limit exists, but the next proposition establishes this fact.

**Proposition 6.6.1** Suppose \( p \in P_2 \) and \( 0 \in A \subset \mathbb{Z}^2 \) is finite. Then the limit in (6.37) exists and equals \( Lg_A(x) \).

**Proof.** We fix \( A \) and let \( r_A = r_A(A) \). Let \( n \) be sufficiently large so that \( A \subset \mathcal{C}_n \). Using (6.24), we have that if \( x \in \mathcal{A}, y \in \partial \mathcal{C}_n \),
\[
P_x'[S_{T_A} = x] = P_x'[\xi_n < T_A] H_{\mathcal{C}_n}(0, y) \left[ 1 + O\left( \frac{r_A \log n}{n} \right) \right].
\]
If \( x \in \mathbb{Z}^2 \setminus \mathcal{C}_n \), the last-exit decomposition (Proposition 4.6.4) gives
\[
P_x'[S_{T_A} = x] = \sum_{y \in \partial \mathcal{C}_n} G_{\mathcal{C}_n}(y, z) P_x'[S_{T_A} = y].
\]
Therefore,
\[
P_x'[S_{T_A} = x] = P_x'(\xi_n < T_A) \left[ 1 + O\left( \frac{r_A \log n}{n} \right) \right] \sum_{y \in \partial \mathcal{C}_n} H_{\mathcal{C}_n}(0, y) G_{\mathcal{C}_n}(y, z).
\] (6.38)
If \( x \in A \), the definition of \( L \), the optional sampling theorem, and the asymptotic expansion of \( g_A \) imply
\[
Lg_A(x) = E_x[g_A(S_{T_A} \wedge \infty)] = E_x[g_A(S_{T_A} \wedge \infty)] = E_x[g_A(S_{T_A} \wedge \infty) \mid \xi_n < T_A] = P_x'(\xi_n < T_A) [C_2 \log n + O_A(1)].
\]
In particular,
\[
\sum_{x \in A} P_x'[\xi_n < T_A] = \sum_{x \in A} \sum_{y \in \partial \mathcal{C}_n} P_x'[S_{T_A} = y] = \sum_{x \in \partial \mathcal{C}_n} \sum_{y \in A} P_x'[S_{T_A} = y] = P_x'[\xi_n < T_A] \left[ 1 + O\left( \frac{r_A \log n}{n} \right) \right] \sum_{y \in \partial \mathcal{C}_n} H_{\mathcal{C}_n}(0, y) G_{\mathcal{C}_n}(y, z).
\]
Combining this with (6.39) gives
\[
\sum_{x \in A} Lg_A(x) = \sum_{x \in \partial \mathcal{C}_n} P_x'[S_{T_A} = x] = 1.
\]
This is a difference between the \( d = 2 \) and \( d \geq 3 \) cases. If \( d \geq 3 \), the sum above grows as \( A \) gets large and gives the capacity of the set. For \( d = 2 \), this quantity is the same for all \( A \).

Note that the summation in (6.38) does not depend on \( x \). If we use the obvious fact
\[
\sum_{x \in A} P_x'[S_{T_A} = x] = 1
\]
we get the proposition. \( \Box \)

We define the **capacity** by
\[
cap(A) = \lim_{y \to A} [a(y) - g_A(y)] = \sum_{x \in A} h_m(x) a(x).
\]
The last proposition establishes the limit. We have the expansion
\[
g_A(x) = C_2 \log J(x) + \gamma_2 - \cap(A) + a_A(1), \quad x \to \infty.
\]
The capacity is translation invariant, \( \cap(A - y) = \cap(A) \) and hence defined for all finite subsets of \( \mathbb{Z}^1 \). Note that singletons sets have capacity zero. Also note that \( a - g_A \) is the unique bounded function on \( \mathbb{Z}^1 \) that is harmonic on \( \mathbb{Z}^2 \setminus A \) and equals \( a \) on \( A \).
Proposition 6.6.2 Suppose \( p \in \mathcal{P}_2 \).

(a) If \( 0 \in A \subset B \subset \mathbb{Z}^2 \) are finite, then \( \varphi_A(x) \geq \varphi_B(x) \) for all \( x \). In particular, \( \text{cap}(A) \leq \text{cap}(B) \).

(b) If \( A, B \subset \mathbb{Z}^2 \) are finite subsets containing the origin, then for all \( x \)

\[
\varphi_{A \cup B}(x) = \max\{\varphi_A(x), \varphi_B(x)\}.
\]

In particular,

\[
\text{cap}(A \cup B) \leq \text{cap}(A) + \text{cap}(B) = \text{cap}(A \cap B).
\]

Proof. The inequity \( \varphi_A(x) \geq \varphi_B(x) \) follows immediately from (6.36). The inequality (6.40) follows from (6.36) and the observation

\[
P^\ast(\mathbb{T}_{A \cup B} \leq \xi_n) = P^\ast(\mathbb{T}_A \leq \xi_n \text{ or } \mathbb{T}_B \leq \xi_n)
\]

\[
= P^\ast(\mathbb{T}_A \leq \xi_n) + P^\ast(\mathbb{T}_B \leq \xi_n) - P^\ast(\mathbb{T}_A \leq \xi_n; \mathbb{T}_B \leq \xi_n)
\]

\[
\leq P^\ast(\mathbb{T}_A \leq \xi_n) + P^\ast(\mathbb{T}_B \leq \xi_n) - P^\ast(\mathbb{T}_{A \cap B} \leq \xi_n),
\]

which implies

\[
P^\ast(\mathbb{T}_{A \cup B} > \xi_n) \geq P^\ast(\mathbb{T}_A > \xi_n) + P^\ast(\mathbb{T}_B > \xi_n) - P^\ast(\mathbb{T}_{A \cap B} > \xi_n).
\]

\[\square\]

We will now derive an analogue of Proposition 6.5.5. If \( A \) is a finite set, let \( a_A \) denote the \( \#(A) \times \#(A) \) symmetric matrix with entries \( a(x, y) \). Let \( a_A \) also denote the operator

\[a_A f(x) = \sum_{y \in A} a(x, y) f(y)\]

which is defined for all functions \( f : A \to \mathbb{R} \) and all \( x \in \mathbb{Z}^2 \). Note that \( z \mapsto a_A f(z) \) is harmonic on \( \mathbb{Z}^2 \setminus A \).

Proposition 6.6.3 Suppose \( p \in \mathcal{P}_2 \) and \( 0 \in A \subset \mathbb{Z}^2 \) is finite. Then

\[
\text{cap}(A) = \left[ \sup_{x \in A} f(x) \right]^{-1},
\]

where the supremum is over all positive functions \( f \) on \( A \) satisfying \( a_A f(x) \leq 1 \) for all \( x \in A \).

If \( A = \{0\} \) is a singleton set, the proposition is trivial since \( a_A f(0) = 0 \) for all \( f \) and hence the supremum is infinity. A natural first guess for other \( A \) (which turns out to be correct) is that the supremum is obtained by a function \( f \) satisfying \( a_A f(x) = 1 \) for all \( x \in A \). If \( \{a_A(x, y)\}_{x \in A} \) is invertible, there is a unique such function that can be written as \( f = a_A^{-1} \mathbf{1} \) (where \( \mathbf{1} \) denotes the vector of all 1s). The main ingredient in the proof of Proposition 6.6.3 is the next lemma that shows this inverse is well defined assuming \( A \) has at least two points.
6.6. CAPACITY IN TWO DIMENSIONS

\[ + G_{C_{0}}(0, 0) \mathbb{P}^* \{ \xi_n < T_A \} + \sum_{y \in A} \mathbb{P}^* \{ S_{r,e,C_{0}} = y \} [G_{C_{0}}(0, 0) - G_{C_{0}}(y, z)]. \]

Letting \( n \to \infty \), this gives

\[ \delta(z - x) = -a(x, z) + \mathcal{L}g_A(z) + \sum_{y \in A} \mathbb{P}^* \{ S_{T_A} = y \} a(y, z). \]

If \( x \in A \), we can use (6.41) to write the previous equation as

\[ \delta(z - x) = \sum_{y \in A} \left[ \mathbb{P}^* \{ S_{T_A} = y \} - \delta(y - x) \frac{\mathcal{L}g_A(y) \mathbb{P}^* \{ S_{T_A} = y \}}{\text{cap}(A)} \right] a(y, z), \]

provided that \( \text{cap}(A) > 0 \).

\[ \square \]

Proof of Proposition 6.6.3 Let \( \hat{f}(x) = \mathcal{L}g_A(x)/\text{cap}(A) \). In (6.41) we showed that

\[ \sum_{y \in A} a(x, y) \hat{f}(y) = 1, \quad x \in A. \]

Suppose \( f \) satisfies the conditions in the statement of the proposition. Let \( h = \hat{f} - af \) which is nonnegative in \( A \). Then, using Lemma 6.6.4,

\[ \sum_{x \in A} \sum_{y \in A} h(x) \hat{f}(y) \geq \sum_{y \in A} \mathbb{P}^* \{ S_{T_A} = y \} h(y) - \sum_{y \in A} h(x) = 0 \]

\[ = \sum_{y \in A} h(y) \sum_{x \in A} \mathbb{P}^* \{ S_{T_A} = x \} - \sum_{x \in A} h(x) = 0. \]

In other words,

\[ \sum_{y \in A} \hat{f}(y) \geq \sum_{y \in A} f(y). \]

\[ \square \]

Proposition 6.6.5 If \( p \in \mathcal{P}_2 \),

\[ \text{cap}(C_{0}) = C_1 \log n + \gamma_2 + O(n^{-1}). \]

Proof. By definition,

\[ g_{C_{0}}(x) = C_2 \log J(x) + \gamma_2 = \text{cap}(C_{0}) + o(1), \quad x \to \infty. \]

But for \( x \not\in C_{0} \),

\[ g_{C_{0}}(x) - \mathbb{E}^* \{ a(S_{T_n}) \} = C_2 \log |x| + \gamma_2 + O(|x|^2) = [C_2 \log n + \gamma_2 + O(n^{-1})]. \]

\[ \square \]

Lemma 6.6.6 If \( p \in \mathcal{P}_1 \), and \( A \subset B \subset \mathbb{Z}^2 \) are finite, then

\[ \text{cap}(A) = \text{cap}(B) - \sum_{y \in B} \text{h}

Proof. \( g_n - g_n \) is a bounded function that is harmonic on \( \mathbb{Z}^2 \setminus B \) with boundary value \( g_A \) on \( B \). Therefore,

\[ \text{cap}(A) = \text{cap}(B) = \lim_{x \to \infty} [g_A(x) - g_B(x)] = \lim_{x \to \infty} \mathbb{E}^* \{ g_A(S_{T_A}) \} = \sum_{y \in B} \text{h}(y) g_A(y). \]

\[ \square \]

Lemma 6.6.7 There exist a \( c_1, c_2 \) such that if \( A \) is a finite subset of \( \mathbb{Z}^2 \) with \( \text{rad}(A) < n \) and satisfying

\[ \# \{ x \in A : k - 1 \leq |x| < k \} \geq 1, \]

for \( k = 1, \ldots, n \), then the following holds,

\( (a) \) if \( x \in \partial C_{n} \),

\[ \mathbb{P}^* \{ T_A < \xi_{n} \} \geq c_1. \]

\( (b) \) \text{cap}(A) = C_2 \log n \leq c_2. \]

Proof. \( (a) \) Choose \( r \in (0,1) \) sufficiently small that \( B_{rn} \subset C_{n} \). Let \( B \) denote a subset of \( A \) contained in \( B_{rn} \) such that

\[ \# \{ x \in A : k - 1 \leq |x| < k \} = 1 \]

for each positive integer \( k \leq rn \). We will prove the estimate for \( B \) which will clearly imply the estimate for \( A \). Let \( V = V_{rn,n} \) denote the number of visits to \( B \) before leaving \( C_{n} \).

\[ V = \sum_{j=0}^{\xi_{n}-1} 1 \{ S_j \in B \} = \sum_{j=0}^{n} \sum_{z \in B} 1 \{ S_j = z : j < \xi_{n} \}. \]

The strong Markov property implies that if \( x \in \partial C_{n} \),

\[ \mathbb{E}^* \{ V \} = \mathbb{P}^* \{ T_B < \xi_{n} \} \mathbb{E}^* \{ V \mid T_B < \xi_{n} \} \leq \mathbb{P}^* \{ T_B < \xi_{n} \} \max_{x \in A} \mathbb{E}^* \{ V \}. \]

Hence, we need only find a \( c_1 \) such that \( \mathbb{E}^* \{ V \} \geq c_1 \mathbb{E}^* \{ V \} \) for all \( x \in \partial C_{n} \), \( z \in B \). Note that \( \#(B) = rn + O(1) \). By (6.11), we can see that \( G_{C_{n}}(x, z) \geq c \) for \( x \in C_{n}, z \in B \). Therefore \( \mathbb{E}^* \{ V \} \geq c n \). If \( z \in B \), there are at most \( 2k + 1 \) points \( w \) in \( B \) satisfying \( |z - w| \leq 2k + 1 \). Using Lemma 6.3.5, we see that

\[ G_{C_{n}}(z, w) \leq C_2 \log n - \log |z - w| + O(1). \]

\[ \square \]
Therefore,
\[ \mathbb{E}[V] \leq \sum_{k=1}^{n} 2C_2 \left[ \log n - \log k + O(1) \right] \leq cn. \]

The last estimate uses the estimate
\[ \sum_{k=1}^{n} \log k \leq \int_{1}^{n+1} \log x \, dx = (n+1) \log(n+1) - (n+1) = n \log n + O(n). \]

(b) There exists an \( r \) such that \( B_r \subseteq C_n \) for all \( n \) and hence
\[ \text{cap}(A) \leq \text{cap}(B_r) \leq \text{cap}(C_n) \leq C_2 \log n + O(1). \]

Hence, we only need to give a lower bound on \( \text{cap}(A) \). By the previous lemma it suffices to give a bound \( g_A(x) \leq c \) for \( x \in \partial C_m \). For \( m > 4n \), let
\[ r_m = r_m m, A = \max_{m \in \mathbb{C}_m} \mathbb{P} \left( \xi_m < T_A \right), \]
\[ r_m^* = r_m m, A = \max_{m \in \mathbb{C}_m} \mathbb{P} \left( \xi_m < T_A \right). \]

Using part (a) and the strong Markov property, we see that there is a \( r < 1 \) such that
\[ r_m \leq r_m^* \rho_m. \] Also, if \( y \in \mathbb{C}_m \),
\[ \mathbb{P} \left( \xi_m < T_A \right) = \mathbb{P} \left( \xi_m < T_{C_m} \right) + \mathbb{P} \left( \xi_m > T_{C_m} \right) - \mathbb{P} \left( \xi_m < T_A \right) \leq \mathbb{P} \left( \xi_m > T_{C_m} \right) + r_m. \]

Lemma 6.4.1 tells us that there is a \( c_3 \) such that for \( y \in \mathbb{C}_m \),
\[ \mathbb{P} \left( \xi_m < T_{C_m} \right) \leq \frac{c_3}{\log m - \log n + O(1)}. \]

Therefore,
\[ g_A(y) = \lim_{m \to \infty} C_1 \left( \log m \right) \mathbb{P} \left( \xi_m < T_A \right) \leq c_2 c_3 \frac{1 - r}{1 + r}. \]

\[ \square \]

**Heuristic note.** A major example of a set satisfying the condition of the theorem is a connected (with respect to simple random walk) subset of \( \mathbb{Z}^2 \) of radius between \( n = 1 \) and \( n \).

In the case of simple random walk, there is another proof of part (a) that basically makes the observation that there is a positive probability that the simple random walk starting on \( \partial C_2 \) makes a closed loop about the origin contained in \( C_0 \). One can justify this rigorously, and bound the probability from below independently of \( n \) by using an approximation by Brownian motion. If the random walk makes a closed loop, then it must intersect any connected set. Unfortunately, it is not easy to modify this argument for nonconnected sets.

### 6.7 Approximating continuous harmonic functions

Discrete harmonic functions appropriately scaled become continuous harmonic functions. We will discuss some quantitative versions of this principle in this chapter. For ease, we will restrict ourselves to \( f \in \mathcal{D} \) whose covariance matrix \( \Gamma \) is a multiple of the identity, but the idea can be extended to the more general case. We let \( U = U_1 = \{ x \in \mathbb{R}^d \mid |x| < 1 \} \) denote the unit ball in \( \mathbb{R}^d \).

**Proposition 6.7.1** If \( f \in \mathcal{D} \) has a covariance matrix that is a multiple of the identity, there is an \( f \in C^\infty \) such that for all \( n \), \( f \) sufficiently large the following holds. Suppose \( f : (n + m)U \to \mathbb{R} \) is a harmonic function with \( \| f \|_1 \leq 1 \). Then there is a function \( \bar{f} \) on \( B_n \) with \( \mathcal{L} f(x) = 0, x \in B_n \) and such that
\[ |f(x) - \bar{f}(x)| \leq \frac{c}{m^2}, \quad x \in B_n. \]

In fact, we can choose
\[ \bar{f}(x) = \mathbb{E} f(S_{x,n}), \]
where \( \xi_n = \min \{ j \geq 0 : |S_j| \geq n \} \).

**Proof.** We will assume that \( m \) is larger than the range of \( f \) so that \( \bar{f} \) as above is well defined. By definition we know that \( \mathcal{L} f(x) = 0, x \in B_n \). We need to establish the estimate. We have already noted that
\[ f(x) = \mathbb{E} \left[ f(S_{x,n}) - \sum_{j=0}^{\xi_n} \mathcal{L} f(S_j) \right] = \bar{f}(x) - \phi(x), \]
where
\[ \phi(x) = \sum_{x \in B_n} G(x, z) \mathcal{L} f(z). \]

We have already noted that there is a \( c_1 \) such that all \( 4 \)th order derivatives of \( f \) at \( x \) are bounded above by \( c(n + m - |x|)^{-4} \). Therefore,
\[ |\phi(x)| \leq c_1 \sum_{j=n}^{\xi_n} \sum_{k=0}^{m} G(x, z_k) \frac{1}{(n + m - k)^4}. \]

We claim that there is a \( c_1 \) such that for all \( x \),
\[ \sum_{k=0}^{m} G(x, z_k) \leq c_1. \]

Indeed, the proof of this is essentially the same as the proof of (5.3). Once we have this, a standard summation by parts gives that \( |\phi(x)| \leq c/m^2 \).
6.8 Eigenvalue of a set

Suppose \( p \in \mathcal{P}_d \) and \( A \subset \mathbb{Z}^d \) is finite and connected with \( \#(A) = m \). The \((\text{first})\) eigenvalue of \( A \) is defined to be the number\(^2\) \( \alpha = \alpha_A \) such that as \( n \to \infty \),

\[
P^*(\tau_A > n) \approx \alpha^n.
\]

We will first consider the case where \( p(x, z) > 0 \). Let \( P^A \) denote the \( m \times m \) matrix \( \{p(x, y)\}_{x, y \in A} \) and, as before, let \( L^A = P^A - I \). Note that \((L^A)^k = \sum \delta(x, y)p_A^k(x, y)\), where \( p_A^k(x, y) = \sum p_A^k(x, y) \).

**Proposition 6.8.1** If \( p \in \mathcal{P}_d \) with \( p(x, z) > 0 \) and \( A \subset \mathbb{Z}^d \) is finite and connected, then there exist numbers \( 0 < \beta = \beta_A < \alpha_A < 1 \) such that if \( x, y \in A \),

\[
p_A^k(x, y) = \alpha^n g_A(z)g_A(y) + O_A(\beta^n),
\]

where \( g_A : A \to \mathbb{R} \) is the unique positive function satisfying

\[
\sum_{x \in A} g_A(x)^2 = 1, \quad P^A g_A(x) = \alpha_d g_A(x), x \in A.
\]

In particular,

\[
P^*(\tau_A > n) = \beta_A(x) \alpha^n + O_A(\beta^n),
\]

where

\[
\beta_A(x) = g_A(x) \sum_{y \in A} g_A(y),
\]

**Proof.** See Proposition 8.4.3; in the notation of that proposition \( u = w = y \). Note that

\[
P^*(\tau_A > n) = \sum_{x \in A} p_A^k(x, y).
\]

We write \( O_A \) to indicate that the implicit constant in the error (as well as \( \beta \)) depends on \( A \).

**Remark.** Although we will assume \( p(x, z) > 0 \) is ergodic, the results in this section can easily be extended to other \( p \). Suppose \( p \in \mathcal{P}_d \) with generator \( L \) and let \( p_\varepsilon(x) = (1 - \varepsilon)p(x) + \varepsilon d(x) \) be the associated lazy walker with generator \( L_\varepsilon = (1 - \varepsilon)L \). Let \( \lambda_{A,\varepsilon} \) denote the first eigenvalue of \( A \) using \( p_\varepsilon \). If \( g_\varepsilon \) is as in the previous proposition, then

\[
(1 - \varepsilon) L g_\varepsilon(x) = \alpha_{A,\varepsilon} g_\varepsilon(x).
\]

In particular, \( g_\varepsilon \) does not depend on \( \varepsilon \) and \( \alpha_{A,\varepsilon} = (1 - \varepsilon) \alpha_A \).

\(^2\)Sometimes the eigenvalues are defined to be \( \lambda = -\log \alpha \) so that \( P^*(\tau_A > n) \approx e^{-\lambda n} \).
6.8. EIGENVALUE OF A SET

The last assertion follows for \( n = m^2 \) by noting that

\[
P\{ S_{m^2} = y \mid \tau_{m^2} > m^2 \} = \frac{\rho_{m^2}(x, y)}{\sum_z \rho_{m^2}(x, z)}
\]

and

\[
\sum_z \rho_{m^2}(x, z) \asymp \rho_m(x) m^{-d} \sum_z \rho_m(z) \asymp \rho_m(x).
\]

For \( n > m^2 \), we can do similarly by conditioning on the walk at time \( n - m^2 \).

\[ \square \]

**Lemma 6.8.3** There exist \( c_1, c_2 \) such that for all sufficiently large \( m \), all \( n \geq m^2 \), and all \( x, y \in \mathcal{C}_m \),

\[
c_1 e^{-\lambda m^2} \rho_m(x) \rho_m(y) \leq m^d \rho_{m^2}(x, y) \leq c_2 e^{-\lambda m^2} \rho_m(x) \rho_m(y).
\]

In particular, there exists \( c_3, c_4 \) such that

\[
c_3 \rho_m(x) \leq m^{1/4} \rho_m(x) \leq c_4 \rho_m(x).
\]

**Proof.** Let

\[
\beta_k = \beta_{km} = m^{-d} \sum_{x \in \mathcal{C}_m} \rho_m(x) P\{ \tau_{m^2} > k m^4 \}.
\]

The previous lemma shows that there exist \( c_1, c_2 \) such that for all positive integers \( k, j \),

\[
c_1 \beta_j \beta_k \leq \beta_{j+k} \leq c_2 \beta_j \beta_k.
\]

Using submultiplicativity (see Corollary 8.7.2), this implies that there is an \( \alpha \) such that

\[
c_3 \alpha e^{-\alpha m} \leq \beta_k \leq c_4 \alpha e^{-\alpha m},
\]

and the definition of the eigenvalue implies that \( \alpha = m^2 \lambda_m \). FINISH THIS PROOF LATER.

It follows from the estimates above that as \( m \to \infty \), \( \lambda_m \asymp m^{-1} \). The next proposition estimates the derivative of \( \lambda_m \) as a function of \( m \).

**Proposition 6.8.4** There exist \( c_1, c_2 \) such that

\[
\frac{c_1}{m^2} \leq \lambda_{m+1} - \lambda_m \leq \frac{c_2}{m^2}.
\]

**Proof.** We let \( k \) denote positive integers. We have shown that as \( m \to \infty \),

\[
P\{ \tau_{m^2} > km^2 \} \asymp e^{-km^3} \lambda_m.
\]

where we emphasize that the implicit constants in the \( \asymp \) notation are independent of \( m \). Let \( j' = j'_{km} \) denote the largest integer \( j \leq k \) such that \( S_{km} \in \mathcal{C}_{km} \setminus \mathcal{C}_{km+1} \) for some \( (j+1)m^2 < n \leq jm^2 \). If there is no such \( j \) we set \( j = k + 1 \), then

\[
P\{ \tau_{m^2} > km^2 \} = P\{ \tau_{m^2} > km^2 \} + \sum_{j=1}^{k} P\{ \tau_{m^2} > km^2 ; j' = j \}.
\]

As \( k \to \infty \),

\[
P\{ \tau_{m^2} > km^2 \} \asymp e^{-km^3} \lambda_m = e^{-km^3} \lambda_m e^{-km^3} \lambda_m \lambda_m.
\]

Below we will show that

\[
P\{ \tau_{m^2} > km^2 ; j' = j \} \asymp m e^{-km^3} \lambda_m e^{-km^3} \lambda_m \lambda_m.
\]

This will imply

\[
\frac{1}{m} \sum_{j=0}^{\infty} e^{-km^3} \lambda_m \lambda_m < 1,
\]

or \( m^2 \lambda_{m+1} - \lambda_m \asymp m^{-1} \). Hence it suffices to prove (6.43), DO THIS LATER.

We will compare \( \lambda_m \) to the corresponding eigenvalue for Brownian motion. Using Theorem 3.4.2, we see that if \( p \in P \) with covariance matrix \( \Gamma \), then we can define a \( p \)-walk \( \mathcal{S} \) and a Brownian motion \( \mathcal{B} \) with covariance matrix \( \Gamma \) on the same probability space so that if \( e_m = \frac{1}{m} \max_{B \in \mathcal{C}_m} |S_m - B| \geq m^{1/2} \log m \),

then \( e_m \to 0 \) faster than any power of \( m \). In particular, for any open set \( D \) containing the origin,

\[
P\{ \text{dist}(S_j, \partial D) > m^{1/2} \log m, 0 \leq j \leq m \log m \} \leq \sum_{B \in \mathcal{C}_m} P\{ B_j \in D, 0 \leq j \leq m \log m \} + e_m,
\]

\[
P\{ \text{dist}(B_j, \partial D) > m^{1/2} \log m, 0 \leq j \leq m \log m \} \leq P\{ S_j \in D, 0 \leq j \leq m \log m \} + e_m.
\]

6.9 Exercises for Chapter 6

**Exercise 6.1** Show that Proposition 6.1.2 holds for \( p \in P^* \).

**Exercise 6.2** In this exercise we construct a transient subset \( A \) of \( \mathbb{Z}^3 \) with

\[
\sum_{x \in A} G(0, y) = \infty.
\]

Here \( G \) denotes the Green's function for simple random walk. Our set will be of the form

\[
A = \bigcup_{k=1} \infty A_k, \quad A_k \{ z \in \mathbb{Z}^3 : |z - 2^k e_1| \leq 2^k e_1 \}.
\]
for some $\epsilon_k \to 0$,

(a) Show that (6.44) holds if and only if $\sum_{k=1}^{\infty} \epsilon_k 2^{2k} = \infty$.

(b) Show that $A$ is transient if and only if $\sum_{k=1}^{\infty} \epsilon_k < \infty$.

(c) Find a transient $A$ satisfying (6.44).

Exercise 6.3 Suppose $p \in \mathcal{P}_d, d \geq 3$. Show that there exists a sequence $k_n \to \infty$ such that if $A \subset \mathbb{Z}^d$ is a finite set with at least $n$ points, then $\text{cap}(A) \geq k_n$. (It suffices to prove the result for simple random walk. Why?)

Exercise 6.4 Show that if $p \in \mathcal{P}_d$ and $r > 0$,

$$\lim_{n \to \infty} [G_{c_n}(0,0) - G_c(0,0)] = 0,$$

Use this and (6.15) to conclude that for all $x, y$,

$$\lim_{r \to \infty} [G_c(0,0) - G_c(x,y)] = a(x,y).$$

Exercise 6.5 Suppose $p \in \mathcal{P}_d$ and $A \subset \mathbb{Z}^d$ is finite. Define

$$Q_A(f,g) = \sum_{x \in A \subset \mathbb{Z}^d} p(x,y) [f(y) - f(x)] [g(y) - g(x)],$$

and $Q_A(f) = Q_A(f,f)$. Let $F : \partial A \to \mathbb{R}$ be given. Show that the supremum of $Q_A(f)$ restricted to the class of $f : A \to \mathbb{R}$ with $f \equiv F$ on $\partial A$ is obtained by the unique harmonic function with boundary value $F$.

Exercise 6.6 Write the two-dimensional integer lattice in complex form, $\mathbb{Z}^2 = \mathbb{Z} + i\mathbb{Z}$ and let $A$ be the upper half plane $A = \{j + ik : k \in \mathbb{Z}, k > 0\}$. Show that for simple random walk

$$G_A(x,y) = a(x,y), \quad x, y \in A,$$

$$H_A(x,j) = \frac{1}{4} \left[ a(x,j+i) - a(x,j+i) \right], \quad x \in A, j \in \mathbb{Z},$$

where $\overline{j + ik} = j - ik$ denotes complex conjugate. Find

$$\lim_{k \to \infty} k H_A(ik,j),$$
Chapter 7

Loop-erased random walk

Throughout this section we fix \( p \in \mathcal{P}_d \).

7.1 Excursions, loops, and \( h \)-processes

7.1.1 Excursion measure

If \( A \subseteq \mathbb{Z}^d \), the boundary Poisson kernel is defined as the function \( H_{\partial A} : \partial A \times \partial A \to [0,1] \) given by

\[ H_{\partial A}(z,w) = P^w\{S_{\tau_A} = w\}. \]

Recall that \( \tau_A = \min\{j \geq 1 : S_j \not\in A\} \). Note that \( H_{\partial A}(z,w) = H_{\partial A}(w,z) \) for \( z,w \in \partial A \). If \( d \geq 3 \), we also define

\[ H_{\partial A}(z,\infty) = H_{\partial A}(\infty,z) = P^w\{\tau_A = \infty\}. \]

For \( d = 2 \), we define \( H_{\partial A}(z,\infty) = H_{\partial A}(\infty,z) = 0 \). Note that if \( z \in \partial A \),

\[ \sum_{w \in \partial A, w = \infty} H_{\partial A}(z,w) = P^z\{S_t \in \overline{A} \}, \]

and if \( \partial A \) is finite and \( d \geq 3 \),

\[ \sum_{w \in \partial A} H_{\partial A}(\infty,z) = \text{cap}(\partial A). \]

If \( \omega = [\omega_0, \omega_1, \ldots, \omega_n] \) is a finite sequence of points in \( \mathbb{Z}^d \) we define \( p(\omega) \) by

\[ p(\omega) = P^w\{S_j = \omega_j, j = 0, \ldots, n\} = \prod_{j=1}^n p(\omega_{j-1}, \omega_j). \]

We will call \( \omega \) a (finite) path (of length \( n \)) if \( p(\omega) > 0 \). We sometimes write \( |\omega| = n \) for the length of the walk. If \( n = 0 \), we call \( \omega \) a trivial path and set \( p(\omega) = 1 \). The points \( \omega_0, \omega_n \) are the endpoints of the path. We call the path a loop if \( \omega_0 = \omega_n \). In particular, the trivial path is a loop. If \( \omega \) is a loop, then \( \omega_0 \) will be called the root of the loop. A loop is called basic if \( \omega_j \neq \omega_0, 1 \leq j \leq n = 1 \). If \( n \geq 1 \), we call \( \omega \) an \( A \)-excursion if \( \omega_0, \omega_n \in \partial A \) and \( \omega_j \in A \) for \( 1 \leq j < n \). The excursion measure on \( A \) is the measure supported on \( A \)-excursions that gives each \( A \)-excursion \( \omega \) measure \( p(\omega) \). The total mass of this measure is

\[ \sum_{\omega \in \partial A} H_{\partial A}(\omega_0,\omega_n). \]

This is finite if \( \partial A \) is finite, but is infinite if \( \partial A \) is infinite.

If \( z, w \in \partial A \) are distinct points with \( H_{\partial A}(z,w) > 0 \), we can consider the conditional measure on \( A \)-excursions from \( z \) to \( w \),

\[ p^z(A) = \frac{p(\omega)}{H_{\partial A}(z,w)}. \]

This is random walk starting at \( z \) conditioned so that the next visit to \( \partial A \) is at \( w \). It is a Markov chain stopped when it reaches state \( w \) with transitions

\[ P^z(x,y) = \frac{p(x,y) H_{\partial A}(y,w)}{\sum_{y'} p(x,y') H_{\partial A}(y',w)}. \]

Here \( H_{\partial A}(\cdot,\cdot) \) denotes the Poisson kernel. If \( x \in A \), the denominator equals \( H_{\partial A}(x,w) \). If \( x = z \), the denominator equals \( H_{\partial A}(z,w) \). This process is called an \( h \)-process and described more in Section 7.1.3.

7.1.2 Loop measure

The loop measure in \( A \) rooted at \( x \) is the measure that assigns weight \( p(\omega) \) to every loop \( \omega = [\omega_0, \omega_1, \ldots, \omega_n] \) rooted at \( x \) with \( \omega_j \in A \) for \( j \geq 1 \). The trivial loop rooted at \( x \) gets weight one even if \( x \) is not in \( A \). We let \( F_A(x) \) denote the total mass of this measure. Then

\[ F_A(x) = 1 + \sum_{w \in \partial A, w \neq x} p(\omega) = \mathbb{P}_x\{T_x > \tau_A\} = 1 - \sum_{\omega \in \mathcal{R}(A)} p(\omega), \quad \text{(7.1)} \]

where the last sum is over all basic loops \( \omega \) in \( A \) of length at least one rooted at \( x \). A standard Markov chain argument shows that

\[ \frac{1}{F_A(x)} = \mathbb{P}_x\{T_x > \tau_A\} = 1 - \sum_{\omega \in \mathcal{R}(A)} p(\omega), \quad \text{(7.1)} \]

Let \( \mathcal{R}(A|x) \) denote the set of basic loops in \( A \) of length at least one rooted at \( x \). A standard Markov chain argument shows that

\[ \frac{1}{F_A(x)} = \mathbb{P}_x\{T_x > \tau_A\} = 1 - \sum_{\omega \in \mathcal{R}(A|x)} p(\omega), \quad \text{(7.1)} \]

where the last sum is over all basic loops \( \omega \) in \( A \) of length at least one rooted at \( x \). If \( |A| \leq 1 \), then

\[ \sum_{n=0}^{\infty} s^n = \frac{1}{1 - s} = e^{-\log(1-s)} = \sum_{k=0}^{\infty} \frac{(\sum_{j=0}^{\infty} s^j)^k}{k!}, \quad \text{(7.2)} \]
7.1. EXCURSIONS, LOOPS, AND \( h \)-PROCESSES

From this we get the relation

The normalized loop measure at \( x \) in \( A \) is the loop measure normalized to be a probability measure, i.e., for each \( A \)-loop rooted at \( x \),

\[
p^x(\omega; x, A) = \frac{p(\omega)}{P_A(x)}
\]

Lemma 7.1.1 Suppose \( d \geq 3 \) or \( A \subsetneq \mathbb{Z}^d \). Suppose \( x \in \mathbb{Z}^d \) and \( S_n \) is a random walk starting at \( x \). Let

\[
k = \max\{j \leq \tau_A : S_j = 0\},
\]

Then the distribution of \([S_0, S_1, \ldots, S_k]\) is that of the normalized loop measure at \( x \) in \( A \).

Proof. This is immediate and left to the reader. If \( d \geq 3 \), it is possible that \( \tau_A = \infty \); however \( k \)'s finite with probability one.

7.1.3 \( h \)-processes

Suppose \( A \subset \mathbb{Z}^d \) is connected and \( h : \mathbb{T} \to [0, \infty) \) satisfies

\[
h(x) > 0, \quad Lh(x) = 0, \quad x \in A.
\]

The \( h \)-process associated to \( h \) is the Markov chain \( Y_j \) with state space \( \mathbb{T} \) and transition probabilities

\[
q(x, y) = p(x, y) \frac{h(y)}{h(x)}, \quad x \in A,
\]

\[
q(x, x) = 1 \text{ if } x \in \partial A \text{ with } h(x) > 0,
\]

\[
q(x, y) = \frac{p(x, y) [h(y) - h(x)]}{Lh(x)} \text{ if } x \in \partial A \text{ with } h(x) = 0.
\]

Note that if \( x \in A \),

\[
\sum_{y \in \mathbb{Z}^d} q(x, y) = \frac{\sum_{y \in \mathbb{Z}^d} p(x, y) h(y)}{h(x)} = 1,
\]

since \( Lh(x) = 0 \). Also if \( x \in \partial A \) with \( h(x) = 0 \), then

\[
\sum_{y \in \mathbb{Z}^d} q(x, y) = \frac{\sum_{y \in \mathbb{Z}^d} p(x, y) h(y)}{Lh(x)} = \frac{\sum_{y \in \mathbb{Z}^d} p(x, y) [h(y) - h(x)]}{Lh(x)} = 1.
\]

Hence \( q \) is a valid transition probability with state space \( \mathbb{T} \).

Suppose \( \omega = [\omega_0, \ldots, \omega_n] \) is a path with \( \omega_0, \ldots, \omega_{n-1} \in A \) and \( \omega_n \in \mathbb{T} \). Then it is easy to check that

\[
P\{Y_n, \ldots, Y_0 = \omega \mid Y_0 = \omega_0\} = p(\omega) \frac{h(\omega_n)}{h(\omega_0)} \quad (7.3)
\]

Here are some important examples.

136

CHAPTER 7. LOOP-ERASED RANDOM WALK

- If \( w \in \partial A \) and we let \( h(z) = H_1(z, w) \), then the \( h \)-process started at \( x \in A \) gives the same measure on paths as that of the random walk conditioned on the event \( \{S_n = w\} \). If \( z \in \partial A \setminus \{w\} \), then \( E h(z) = H_0(z, w) \).

- If \( d \geq 3 \) and \( h(z) = P^x[\mathbb{T} = \infty] \) is strictly positive on \( A \), the associated \( h \)-process started at \( x \in A \) gives the same measure on paths as that of the random walk conditioned on the event \( \{\tau = \infty\} \). Loosely speaking, we call this process random walk in \( A \) conditioned to exit \( A \) “at infinity”.

- If \( d = 2 \) and \( B = \mathbb{Z}^d \setminus A \) is finite, let

\[
h(x) = g_B(x) = \lim_{n \to \infty} C_1 (\log n) P^x[I_n < \tau_A].
\]

This measure can be considered random walk “conditioned to never reach \( B \)”. Note that we cannot simply condition on the event that the path never reaches \( B \) since this is an event of probability zero. Loosely speaking, we can call this process random walk in \( A \) conditioned to exit \( A \) “at infinity”. In the special case \( B = \{0\} \), we have \( g_B(x) = a(x) \) and we call the process random walk conditioned to avoid the origin. As the next lemma shows, this process is transient.

Lemma 7.1.2 Suppose \( x \in \mathbb{Z}^2 \) and \( Y_0, Y_1, \ldots \) is random walk conditioned to avoid the origin starting at \( x \). Suppose \( 0 < r < \|x\| \). Then,

\[
P\{Y_n \in \mathcal{C} \text{ for some } n \geq 0\} = \frac{\log r + O(1)}{\log \|x\|}.
\]

Proof. If \( q \) denotes the probability on the left, then \( (7.3) \) shows that

\[
\min_{y \in \mathbb{T} \setminus \mathcal{C}} a(y) \leq q a(x) \leq \max_{y \in \mathbb{T} \setminus \mathcal{C}} a(y).
\]

Therefore,

\[
q = C_3 \log r + O(1).
\]

7.2 Loop erasure

A self-avoiding walk (SAW) is a path that visits no point more than once, i.e., that satisfies \( \omega_j \neq \omega_k \) for \( 0 \leq j < k \leq n \). Chronological loop-erasure is a way to assign to each path \( \omega = [\omega_0, \ldots, \omega_n] \) a self-avoiding subpath with the same endpoints. To be precise, for any path \( \omega \) we define \( LE(\omega) \) as follows.

- Let \( j_{-1} = -1 \) and \( j_0 = \max\{j \leq n : \omega_j = \omega_0\} \).

- If \( j_n = n \), we stop. Otherwise, we let \( j_1 = \max\{j \leq n : \omega_j = \omega_{j+1}\} \).
7.2. LOOP ERASURE

- Recursively, if \( j_i = n \), we stop. Otherwise, we let \( j_{i+1} = \max\{ j \leq n : \omega_j = \omega_{j+1} \} \).
- Let \( k \) be the smallest \( i \) such that \( j_i = n \). Then
  \[
  \text{LE}(\omega) = [\omega_{j_1}, \omega_{j_2}, \ldots, \omega_{j_k}] = [\omega_{j_{i+1}}, \omega_{j_{i+2}}, \ldots, \omega_{j_{i+k}}].
  \]

Let \( L^i = [\omega_{j_{i+1}}, \ldots, \omega_{j_k}] \). What we have done is divide the path \( \omega \) into a collection of loops, \([L^1, \ldots, L^k]\). This division is uniquely determined by the following conditions:
- The path \( \omega \) is exactly the path \([L^1, \ldots, L^k]\).
- Each \( L^i \) is a loop rooted at \( \omega_{j_k} \).
- For each \( i \geq 1 \), the loop \( L^i \) does not include any of the points \( \{\omega_{j_1}, \omega_{j_2}, \ldots, \omega_{j_k} \} \).

The loop-erased excursion measure on \( A \) is the measure \( \hat{p}_A \) that assigns to each self-avoiding excursion with distinct endpoints the excursion measure of the set of paths whose loop-erasure is that path. In other words, if \( \eta = [\eta_0, \ldots, \eta_k] \) is a self-avoiding \( A \)-excursion, then

\[
\hat{p}_A(\eta) = \sum \hat{p}(\omega),
\]

where the sum is over all \( A \)-excursions \( \omega \) with \( \text{LE}(\omega) = \eta \). Note that if \( z, w \) are distinct points in \( \partial A \), then the loop-erased excursion measure of paths with end points \( z, w \) is \( H_{A}(z, w) \). If \( z \in \partial A \) and \( \omega \) is the trivial path at \( z \), we set \( \hat{p}_A(\omega) = 1 \).

Note that

\[
\text{LE}[\omega_{j_1}, \omega_{j_2}, \ldots, \omega_{j_k}] = \text{LE}(\omega).
\]

In other words, in order to produce a walk with a certain loop-erasure, we can divide the walk into two parts: a beginning loop and a walk that never returns to the starting point. Using this, we can see that if \( k \geq 1 \) and \( \eta = [\eta_0, \eta_1, \ldots, \eta_k] \) is a self-avoiding excursion in \( A \), then

\[
\hat{p}_A(\eta) = \hat{p}(\eta_0, \eta_1) \prod_{j=0}^{k} \hat{p}_{A[\eta_0, \ldots, \eta_j]}(\eta_{j+1}, \ldots, \eta_k).
\]

By continuing, we get for \( 1 \leq i \leq k \),

\[
\hat{p}_A(\eta) = \hat{p}(\eta_0, \ldots, \eta_i) \left[ \prod_{j=0}^{i-1} \hat{p}_{A[\eta_0, \ldots, \eta_j]}(\eta_{j+1}, \ldots, \eta_k) \right] \hat{p}_{A[\eta_0, \ldots, \eta_i]}(\eta_{i+1}, \ldots, \eta_k),
\]

where \( A_j = A \setminus \{ \omega_0, \ldots, \omega_{i-1} \} \). When \( i = k \), \( \hat{p}_{A[\eta_0, \ldots, \eta_k]}(\eta_{k+1}, \ldots, \eta_k) = 1 \).

**Lemma 7.2.1** If \( A \subseteq \mathbb{Z}^d \) and \( \{x_1, \ldots, x_n\} \subseteq \mathbb{Z}^d \), let

\[
F_A(x_1, \ldots, x_n) = \prod_{j=1}^{n} F_A(x_j),
\]

where \( A_j = A \) and \( A_j = A \setminus \{x_1, \ldots, x_{j-1}\} \). Then, for any permutation \( \sigma \) of \( \{1, \ldots, n\} \),

\[
F_A(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = F_A(x_1, \ldots, x_n).
\]

**Proof.** If \( j \) is a transposition (permutations that switch \( j \) and \( j+1 \) and leave all other points fixed) since all permutations can be generated as a product of transpositions, if either \( x_j \) or \( x_{j+1} \) is not in \( A \), this is trivial, so it suffices to show that for any \( A \) and any distinct \( z, y \in A \), that

\[
G_{A[\eta]}(x, x) G_A(y, y) = G_A(x, x) G_{A[\eta]}(y, y).
\]

By splitting the number of returns to \( x \) before leaving \( A \) into those that occur before visiting \( y \) and then after visiting \( y \) we get

\[
G_A(x, x) = G_{A[\eta]}(x, x) + P^x(T_\eta < \tau_A) G_A(x, y) + P^y(T_\eta < \tau_A) G_A(x, x).
\]

Hence,

\[
G_{A[\eta]}(x, x) = [1 - P^x(T_\eta < \tau_A)] P^y(T_\eta < \tau_A) G_A(x, x),
\]

from which (7.5) follows readily.

If \( A \subseteq \mathbb{Z}^d \) and \( B = \{x_1, \ldots, x_n\} \) we define

\[
F_A(B) = \prod_{j=1}^{n} F_A(x_j),
\]

where \( A_1 = A \) and \( A_j = A \setminus \{x_1, \ldots, x_{j-1}\} \). The last lemma shows that this quantity is independent of the ordering of the points. Note that \( F_A(B) = F_A(B \cap A) \). If \( \eta = [\eta_0, \ldots, \eta_k] \) is a self-avoiding path, then we write \( F(\eta) \) for \( F(\eta_0, \ldots, \eta_k) \). We can rewrite (7.4) as

\[
\hat{p}_A(\eta) = \hat{p}(\eta) F(\eta).
\]

In particular if \( \eta^R = [\eta_k, \eta_{k-1}, \ldots, \eta_0] \) is the “reversed” path, then \( \hat{p}_A(\eta^R) = \hat{p}_A(\eta) \).

**Heuristic note.** It is easy to give examples of paths \( \omega \) such that \( \text{LE}(\omega^R) \) is not the same as \( \text{LE}(\omega) \).

The measures \( p \) and \( \hat{p} \) are measures on \( A \)-excursions. Given \( p \) we get \( \hat{p} \) by erasing loops. The next proposition discusses how to get \( p \) from \( \hat{p} \) by adding loops.

**Proposition 7.2.2** Suppose \( d \geq 3 \) or if \( A \subseteq \mathbb{Z}^d \) and \( \eta = [\eta_0, \ldots, \eta_k] \) is a self-avoiding \( A \)-excursion. For each \( j \), let

\[
A_j = A \setminus \{\eta_i : i < j\}.
\]

For \( j = 0, \ldots, n-1 \), let \( \omega^j = [\omega^0, \ldots, \omega^j] \) be chosen independently from the normalized loop measure at \( \eta_j \) in \( A_j \). Let \( \omega \) be the excursion obtained from adding these loops, \( \omega = [\eta_0, \omega^0, \eta_1, \omega^1, \ldots, \eta_{n-1}, \omega^{n-1}, \eta_n] \).

Then the distribution of \( \omega \) is that of \( \eta \) conditional on \( \text{LE}(\omega^R) = \eta^R \).

**Proof.** This follows easily from the definition.
7.3 Unrooted Loop Soup

A fundamental property of loops is that the quantity \( F_n(B) \) does not depend on the order of the points in \( B \). We will describe another way to "add loops" to a set which helps illustrate why this property is independent of the order.

An unrooted loop is an equivalence class of loops under translation of the root, i.e., it is an equivalence class under the equivalence relation

\[
[w_0, w_1, \ldots, w_n] \sim [w_0', w_1', \ldots, w_n'] \sim \cdots \sim [w_{m-1}, w_m, \ldots, w_1, w_0].
\]

We will use \( \mathcal{W} \) to denote the unrooted loop represented by the rooted loop \( \omega \). We define \( p(\mathcal{W}) \) to be \( p(\omega) \); note that \( p(\mathcal{W}) \) is the same for all representatives of \( \mathcal{W} \) so this is well defined. We write \( \mathcal{A} \subset A \) if \( \{w_0, \ldots, w_n\} \subset A \). Let \( \kappa(\mathcal{W}) \) be the largest integer \( m \) that is divisible by \( n \) such that \( w_i = \omega_{i+m(p/m)} \) for all \( k \leq n = (n/m) \). Note that \( \kappa(\mathcal{W}) \) is independent of the choice of representative of \( \mathcal{W} \), and the number of distinct rooted loops of \( \mathcal{W} \) is \( n/\kappa(\mathcal{W}) \).

Definition. The (unrooted) loop soup with intensity \( t \) is a realization \( \mathcal{U}_t \) from a Poisson point process on unrooted loops with measure \( p/\kappa \).

To be more precise, let \( \{N_{\mathcal{W}}\} \) denote a collection of independent Poisson processes indexed by the set of unrooted loops \( \mathcal{W} \). The process \( N_{\mathcal{W}} \) has parameter \( p(\mathcal{W})/\kappa(\mathcal{W}) \). For each \( t > 0 \), let \( \mathcal{U}_t \) denote the set of loops that have been created by time \( t \). The set \( \mathcal{U}_t \) is a multiset (i.e., a set in which elements can appear more than once) in which the loop \( \mathcal{W} \) appears \( N_{\mathcal{W}} \) times. If \( N_{\mathcal{W}} > 0 \), i.e., a jump of this process occurs at time \( t \), we say that the loop \( \mathcal{W} \) is created at time \( t \). Note that with probability one, there is no time at which more than one loop is created.

If \( A \subset \mathbb{Z}^2 \), we let \( \mathcal{U}_t(A) = \{ \mathcal{W} \in \mathcal{U}_t : \mathcal{W} \subset A \} \). Suppose an ordering of \( A \) is given, \( A = \{z_1, z_2, \ldots\} \). Then given \( \mathcal{U}_t(A) \) we can associate rooted loops as follows:

- Given \( \mathcal{W} \) represented by \( \omega = [w_0, \ldots, w_n] \), let \( j \) be the smallest integer such that \( z_j \in \{w_0, \ldots, w_n\} \).
- Let \( \ell_1 < \ell_2 < \cdots < \ell_k \) denote all the indices \( 1 \leq i \leq n \) such that \( w_i = z_j \).
- Let \( Y = Y(\mathcal{W}) \) be a random variable, independent of all other random variables, uniformly distributed on \( \{1, \ldots, k\} \). Then choose the (rooted) loop

\[
[w_{\ell_1}, w_{\ell_2}, \ldots, w_{\ell_1}, w_{\ell_2}, \ldots, w_{\ell_k}, w_{\ell_1}, \ldots, w_{\ell_k}, w_{\ell_1}].
\]

In other words, given an unrooted loop \( \mathcal{W} \), we choose a root for \( \mathcal{W} \) by finding the smallest point in the loop, with smallest defined by the given ordering on \( A \). If this gives more than one choice for the root, we choose randomly.

After doing this process we have a realization \( \mathcal{U}_t(A) \) of rooted loops. Let \( \mathcal{A}_k \) denote the set of all rooted loops \( \omega \) with \( w_0 = z_k \), all of whose points lie in \( A \setminus \{z_1, \ldots, z_k\} \). It is easy to see that we now have an independent collection of Poisson processes \( \{N_{\mathcal{W}}\} \) indexed by rooted loops \( \omega \in \bigcup_k \mathcal{A}_k \). If \( \omega = [w_0, \ldots, w_n] \in \mathcal{A}_k \), then the parameter of the process \( N_{\mathcal{W}} \) is \( p(\omega)/J(\omega) \) where \( J = J(\omega) \) denotes the number of indices \( i \) with \( 1 \leq i \leq n \) and \( w_i = z_k \).

Note that for each \( z \), there are only a finite number of loops in \( \mathcal{U}_t(A) \) rooted at \( z \). Suppose these loops are

\[
\omega^1 = [w^1_0 = z, w^1_1, \ldots, w^1_m = z], \ldots, \omega^r = [w^r_0 = z, w^r_1, \ldots, w^r_m = z].
\]

These are written down in the order that they appeared in the realization. Then we form a single loop at \( z \) by

\[
L_z = L_z(A) = [z, w^1_1, \ldots, w^1_m, w^2_1, \ldots, w^2_m, \ldots, w^r_1, \ldots, w^r_m].
\]

If no loop rooted at \( z \) appears in the realization, then \( L_z \) is the trivial loop.

Proposition 7.3.1 If \( A = \{z_1, z_2, \ldots\} \) and \( t = 1 \), then the loops \( L_{z_1}, L_{z_2}, \ldots \) are independent, with \( L_{z_i} \) chosen from the normalized loop measure in \( A \setminus \{z_1, z_2, \ldots\} \).

Proof. Since the \( \mathcal{A}_k \) are disjoint and the loop \( L_z \) is formed from \( \mathcal{U}_t(A) \cap \mathcal{A}_k \); the loops \( L_{z_1}, L_{z_2}, \ldots \) are independent. Let

\[
\lambda_k = \sum_{\omega \in \mathcal{A}_k} p(\omega) \lambda = \sum_{\mathcal{W} \in \mathcal{U}_t(A)} p(\mathcal{W}) = \sum_{j=1}^{\infty} \sum_{\omega \in \mathcal{A}_k} p(\omega) = \sum_{j=1}^{\infty} \left( \sum_{\mathcal{W} \in \mathcal{A}_k} p(\mathcal{W}) \right) = \log F_{\mathcal{A}_k}(\infty, \ldots, \infty) \cdot (z_k).
\]

Then \( N_{\mathcal{W}} \) is a Poisson process with parameter

\[
\sum_{\mathcal{W} \in \mathcal{A}_k} p(\mathcal{W}) = \sum_{j=1}^{\infty} \frac{1}{j} \sum_{\mathcal{W} \in \mathcal{A}_k} p(\mathcal{W}) = \sum_{j=1}^{\infty} \frac{1}{j} \left( \sum_{\mathcal{W} \in \mathcal{A}_k} p(\mathcal{W}) \right) = \log F_{\mathcal{A}_k}(\infty, \ldots, \infty) \cdot (z_k).
\]

The last equality uses (7.1).

Now suppose that \( \omega \in \mathcal{A}_k \). Let \( \ell_0 = 0 < \ell_1 < \ell_2 < \cdots < \ell_i = n \) denote all of the indices with \( w_i = z_k \). Then for any collection of positive integers \( j_1, \ldots, j_m \) with \( j_1 + \cdots + j_m = l \), we can divide \( \omega \) into \( m \) loops of the form

\[
\omega^r := [w_{\ell_1}, w_{\ell_1+j_1}, \ldots, w_{\ell_1+j_1}, \ldots, w_{\ell_2}, w_{\ell_2+j_2}, \ldots, \ldots, w_{\ell_m}, w_{\ell_m+j_m}].
\]

The probability that the loop soup restricted to \( \mathcal{A}_k \) produces exactly \( r \) loops which are in order \( \omega^1, \ldots, \omega^r \) is

\[
e^{-\lambda} \prod_{r=1}^{\infty} \frac{\lambda^r}{r!} = 1.
\]

We claim that

\[
\sum_{A \in \mathcal{A}_k} \frac{1}{J(\omega)^{J(\omega)}} = 1.
\]

CHAPTER 7. LOOP-ERASED RANDOM WALK
This follows from
\[ \sum_{n=0}^{\infty} a_n = \frac{1}{1-a} = e^{\log(1/(1-a))} = \sum_{k=0}^{\infty} \frac{1}{k!} \left( \sum_{j=0}^{\infty} \frac{a^j}{j!} \right)^k, \]
which is valid for \(|a| < 1\). The left-hand side of (7.7) is the moment generating function for the distribution of 
\(X_0\), whereas the right-hand side is the expected value of \(e^{\lambda X_0}\) for any \(\lambda > 0\).

7.4 Loop-erased random walk (LERW)

If \(A \subseteq \mathbb{Z}^d\) and \(z, w \in \partial A\) are distinct with \(H_{\partial A}(z, w) > 0\), then the loop-erased random walk (LERW) from \(z\) to \(w\) in \(\partial A\) is the probability distribution on self-avoiding paths induced by \(\hat{p}\) restricted to self-avoiding \(A\)-excursions from \(z\) to \(w\). In other words, if \(\eta\) is a self-avoiding \(A\)-excursion from \(z\) to \(w\), then the probability of that \(\eta\) is given by
\[ \hat{p}_A(\eta; z, w) = \frac{\hat{p}(\eta)}{H_{\partial A}(z, w)} = \frac{p(\eta) F_A(\eta)}{H_{\partial A}(z, w)}. \]
The LERW from \(z\) to \(w\) is defined to be the distribution supported on the trivial walk.

If \(n \leq m\) and \(\omega_{\eta} = [\omega_0, \ldots, \omega_m], \omega = [\omega_0, \ldots, \omega_n]\) are two paths we write \(\omega \leq \omega\) if \(\omega_j = \omega_j\) for \(0 \leq j \leq n\). The following proposition follows from (7.4).

**Proposition 7.4.1** Suppose \(z, w\) are distinct and \(\eta = [\eta_0, \ldots, \eta_k]\) is a self-avoiding path with \(\eta_0 = z, \eta_k = w\), and \(\eta_0 \in A\). Then
\[ \sum_{\eta_0} \hat{p}_A(\eta; z, w) = p(\eta) F_A(\eta) H_{\partial A}(\eta_0, w). \]

In particular, if \(\eta_0 \in A\) and \([\hat{S}_0, \ldots, \hat{S}_m]\) denotes the LERW from \(z\) to \(w\) in \(A\), then
\[ \mathbb{P}(\hat{S}_{n+1} = x | [\hat{S}_0, \ldots, \hat{S}_n] = [z, x_1, \ldots, x_m]) = \frac{p(x, z) H_{\partial A}(x_1, \ldots, x_m; x, w)}{\sum_{y \neq z} p(x, y) H_{\partial A}(x_1, \ldots, x_m; y, w)}. \]

The quantity on the right-hand side of the last equation is zero if \(z \not\in (A \setminus \{x_1, \ldots, x_n\}) \cup \{w\}\). Another way to think of the LERW is as a Markov chain taking values in ordered pairs \((V, x)\) where \(V \subseteq \mathbb{Z}^d\) and \(x \subseteq \mathbb{Z}^d \setminus V\). The probability of a transition from \((V, x)\) to \((V \setminus \{y\}, y)\) is the probability that the LERW in \(V\) from \(x\) to \(w\) takes its first step to \(y\) which we have stated is proportional to \(p(x, y) H_V(x, y)\). All other transitions have probability zero. The initial state is \((A, z)\) and the chain is stopped when the second component of the process equals \(w\).

There are a number of similar LERW's.
7.5 Spanning Trees and Wilson's Algorithm

Loop-erasure of walks gives an efficient way to find spanning trees of graphs. We give an example of this in this section. Suppose $A \subseteq \mathbb{Z}^d$ is finite and connected. Define the weighted undirected graph $A_W$ (the $W$ stands for "wired") by:

- The vertices of $A_W$ are $A \cup \{ \partial \}$ where $\partial$ is a single boundary point.
- If $x, y \in A$ are distinct with $p(x, y) > 0$, then $(x, y)$ is an edge with weight $p(x, y)$.
- If $x \in A$ and $z \in \partial A$, there is an an edge between $x$ and $\partial$ with weight $p(x, z)$. Note that there can be more than one edge between $x$ and $\partial$.

A spanning tree $T$ of $A_W$ is a connected subgraph containing all the vertices with no cycles. In other words, it is a subset of the edges of $A_W$ with the property that if $x, y$ are distinct vertices, then there exists a unique path $x = x_0, x_1, \ldots, x_k = y$ with $x_0, \ldots, x_k$ distinct vertices in $A \cup \{ \partial \}$ such that for $j = 1, \ldots, k$, $(x_{j-1}, x_j)$ is an edge of $T$. The number of edges in any spanning tree is $\#(A)$. We define the weight of the tree by

$$p(T) = \prod_{(x, y)} p(x, y).$$

Wilson's algorithm is a way to choose a random spanning tree of $A_W$ using loop-erased walks. We state one form of it here.

- Let $A_0 = A$.
- Order the points in $A$, $x_1, \ldots, x_n$.
- Start a random walk at $x_1$ and stop it when it leaves $A$. Take the loop-erasure of the path producing $\eta = [y_0, \ldots, y_k]$ with $k > 0$. The point $y_k$ is in $\partial A$ and we write it as just $\partial$. Include in the spanning tree the edges $(y_0, y_1), (y_1, y_2), \ldots, (y_{k-1}, y_k)$.
- Let $\hat{A} = A \setminus \{y_1, \ldots, y_n\}$. If $\hat{A} = \emptyset$, stop. Otherwise continue.
- Let $j$ be the smallest index with $x_j \in \hat{A}$. Start a random walk at $x_j$ and stop it when it leaves $\hat{A}$. Take the loop-erasure of the path producing $\eta = [y_0, \ldots, y_k]$ with $k > 0$. If the point $y_k$ is in $\partial \hat{A}$, write it as just $\partial$. Add to the tree the edges $(y_0, y_1), (y_1, y_2), \ldots, (y_{k-1}, y_k)$.
- Let $\hat{A}' = \hat{A} \setminus \{y_1, \ldots, y_m\}$. If $\hat{A}' = \emptyset$, stop. Otherwise, set $\hat{A} = \hat{A}'$ and return to the last step.

It is easy to see that this algorithm produces a spanning tree and hence generates a distribution on spanning trees. The next proposition describes the distribution and shows that it is independent of the ordering of the points $x_1, \ldots, x_n$.

### Proposition 7.5.1
For any spanning tree $T$ of $A_W$, the probability that $T$ is produced is $p(T) F_A(A)$, where $F_A(A)$ is defined as in (7.6).

**Proof.** Suppose $T$ is a spanning tree. Then $T$ produces an ordering of the vertices of $A$, $\{y_1, \ldots, y_n\}$ as follows:

- The unique path from $y_1$ to $\partial$ is given by $\eta^1 = [y_1, \ldots, y_k, \partial]$.
- If $j < n$, let $m = m_i$ be the smallest index $m$ such that $y_m \notin \{y_1, \ldots, y_k\}$. Then $\eta^{i+1} = [y_{k+1}, \ldots, y_{k+m}, \partial]$ is the unique path from $y_m$ to $\partial, y_1, \ldots, y_k$ in $T$.

We have partitioned $T$ into $l$ self-avoiding paths $\eta^1, \ldots, \eta^l$, with disjoint edges. In order for $T$ to be produced from the algorithm, it must be the case that at the first step $\eta^1$ is created, and then $\eta^2$ at the second step, etc. Therefore, the probability that $T$ is created is

$$p(\eta^1) p(\eta^2) \cdots p(\eta^l),$$

where $A^l = A \setminus \{y_1, \ldots, y_k\}$. Using (7.4) and Lemma 7.2.1, we get the proposition. \hfill \square

### Corollary 7.5.2 (Uniform Spanning Tree)
Suppose $p$ is simple random walk in $\mathbb{Z}^d$ and $A \subseteq \mathbb{Z}^d$ is finite and connected. Then Wilson's algorithm chooses a spanning tree for $A_W$ from the uniform measure on spanning trees. The number of spanning trees in $\{2d\}^\#(A) F_A(A)^{\#(A)}$.

**Proof.** If $p$ is simple random walk, then the probability that any particular spanning tree $T$ is $p(T) = (2d)^{\#(A)} F_A(A)^{\#(A)}$. \hfill \square

### 7.6 Fomin's Identity

In this section, we prove a result that relates the probability that the paths of a loop-erased walk and a random walk do not intersect to a determininant involving only the random walk.

### Theorem 7.6.1 (Fomin's identity)
Suppose $A \subseteq \mathbb{Z}^d$ and $z_1, z_2, w_1, w_2 \in \partial A$. For $j = 1, 2$, let

$$q_j = q_j(A; z_1, z_2, w_1, w_2) = \sum_{p(\eta^j)} p(\eta^j)^2,$$

where the sum is over all $A$-excursions $\eta$ in $A$ such that $\eta^1$ has end points $z_1, w_1$; $\eta^2$ has end points $z_2, w_2$; and $LE(\eta^1) \cap \eta^2 \cap A = \emptyset$. Then

$$q_1 = q_2 = \det \begin{bmatrix} H_{3A}(z_1, w_1) & H_{3A}(z_1, w_2) \\ H_{3A}(z_2, w_1) & H_{3A}(z_2, w_2) \end{bmatrix}.$$
7.7 Intersection probabilities for random walk

Proof. Since
\[ H_{A}(z_1, w_1) = \sum \rho^{z_1} \]
where the sum is over \( A \)-excursions \( \omega^{z_1} \) from \( z_1 \) to \( w_1 \), we have
\[
\text{det} \left[ \begin{array}{c} H_{A}(z_1, w_1) \\ H_{A}(z_2, w_1) \\ H_{A}(z_1, w_2) \end{array} \right] = \sum_{z_1} \sum_{w_1} \rho^{z_1} \rho^{w_1} - \sum_{z_2} \sum_{w_2} \rho^{z_2} \rho^{w_2}.
\]
Let \( A^1 \) denote the set of ordered pairs \( (\omega^{z_1}, \omega^{z_2}) \) such that \( \text{LE}(\omega^{z_1}) \cap \omega^{z_2} \cap A \neq \emptyset \) and \( A^2 \) the set of ordered pairs \( (\omega^{z_1}, \omega^{z_2}) \) such that \( \text{LE}(\omega^{z_1}) \cap \omega^{z_2} \cap A = \emptyset \). We will find a one-to-one correspondence between \( A^1 \) and \( A^2 \), \( (\omega^{z_1}, \omega^{z_2}) \leftrightarrow (\Lambda^{z_1}, \Lambda^{z_2}) \) such that \( \rho(\omega^{z_1}) \rho(\omega^{z_2}) = \rho(\Lambda^{z_1}) \rho(\Lambda^{z_2}) \). Once we have done this, we can see that all these terms will cancel in the sum and we get (7.8).

Suppose \( \omega^{z_1} = [z_1, x_1, \ldots, x_m] \) and \( \omega^{z_2} = [z_2, y_1, \ldots, y_n] \) are given and suppose \( \text{LE}(\omega^{z_1}) = [z_1, \hat{x}_1, \ldots, \hat{x}_k, w_1] \). If \( \text{LE}(\omega^{z_1}) \cap \omega^{z_2} \cap A \neq \emptyset \), we can let \( j \) be the smallest positive integer \( j \) such that \( \hat{x}_j \in \{y_1, \ldots, y_n\} \) and \( l \) be the largest integer \( l \) such that \( \hat{x}_l = \hat{x}_j \). We now set
\[
\omega^{z_1} = [z_1, x_1, \ldots, x_k, y_j, x_{k+1}, \ldots, x_m, w_1],
\]
\[
\omega^{z_2} = [z_2, y_1, \ldots, y_n, x_{k+1}, \ldots, x_m, w_1].
\]
One can check that this correspondence works. \( \square \)

### 7.7 Intersection probabilities for random walk

Analysis of the LERW leads naturally to questions about when random walks intersect. In this section we will prove a fundamental inequality. If \( S_n \) is a random walk, we will write
\[
S[n_1, n_2] = \{S_n : n_1 \leq n \leq n_2\}.
\]

**Proposition 7.7.1** If \( p \in P_1 \), there exist \( c_1, c_2 \) such that for all \( n \),
\[
c_1 \phi(n) \leq \mathbb{P}\{S[0, n] \cap S[2n, 3n] \neq \emptyset\} \leq \mathbb{P}\{S[0, n] \cap S[2n, \infty] \neq \emptyset\} \leq c_2 \phi(n),
\]
where
\[
\phi(n) = \begin{cases} 1, & d = 4, \\
\frac{1}{\log n}, & d = 4, \\
\frac{1}{n^{d/2}}, & d > 4. 
\end{cases}
\]

### Proof. The result is trivial in the recurrent case, so we will assume that \( d \geq 3 \). We will assume the walk is aperiodic (only a trivial modification is needed for the bipartite case).

The basic strategy is to consider the number of intersections of the paths,
\[
J_n = \sum_{j=0}^{n} \mathbb{1}\{S_j = S_k\}, \quad K_n = \sum_{j=0}^{n} \mathbb{1}\{S_j = S_k\}.
\]

Note that
\[
\mathbb{P}\{S[0, n] \cap S[2n, 3n] \neq \emptyset\} = \mathbb{P}\{J_n \geq 1\}, \quad \mathbb{P}\{S[0, n] \cap S[2n, \infty] \neq \emptyset\} = \mathbb{P}\{K_n \geq 1\}.
\]

We will derive the following inequalities for \( d \geq 3 \),
\[
c_1 n^{(d-2)/2} \leq \mathbb{E}(J_n) \leq \mathbb{E}(K_n) \leq c_2 n^{(d-2)/2}, \quad (7.9)
\]
\[
\mathbb{E}(J_n^2) \leq \begin{cases} cn, & d = 3, \\
\frac{c}{\log n}, & d = 4, \\
\frac{c n^{(d-2)/2}}, & d \geq 5. 
\end{cases} \quad (7.10)
\]

Once these are established, the lower bound follows by the second moment lemma (Lemma 8.6.1),
\[
\mathbb{P}\{J_n \geq 1\} \geq \frac{\mathbb{E}(J_n)^2}{\mathbb{E}(J_n^2)}.
\]

Let us write \( p(n) \) for \( \mathbb{P}\{S_n = 0\} \). Then,
\[
\mathbb{E}(J_n) = \sum_{j=0}^{n} \sum_{k=0}^{n} p(k-j),
\]
and similarly for \( \mathbb{E}(K_n) \). Since \( p(k-j) \asymp (k-j)^{d/2} \), we get
\[
\mathbb{E}(J_n) \asymp \sum_{j=0}^{n} \sum_{k=0}^{n} \frac{1}{(k-j)^{d/2}} \asymp \sum_{j=0}^{n} \sum_{k=0}^{n} \frac{1}{(k+n)^{d/2}} \asymp \sum_{j=0}^{n} n^{(d-2)/2} \asymp n^{(d-2)/2},
\]
and similarly for \( \mathbb{E}(K_n) \). This gives (7.9). To bound the second moments, note that
\[
\mathbb{E}(J_n^2) \leq \sum_{1 \leq j \leq n} \sum_{1 \leq k \leq n} [\mathbb{P}\{S_j = S_k\} + \mathbb{P}\{S_j = S_k, S_k = S_0\}] + \mathbb{P}\{S_0 = S_k, S_k = S_0\},
\]
If \( 0 \leq i \leq n \) and \( 2n \leq k \leq 3n \), then
\[
\mathbb{P}\{S_j = S_k, S_k = S_0\} \leq \left[ \max_{0 \leq l \leq n} \mathbb{P}\{S_l = x\} \right] \frac{c}{n^{d/2}} \leq \frac{c}{n^{d/2}}.
\]
Therefore,
\[
\mathbb{E}(J_n^2) \leq c_n^2 \sum_{2 \leq k \leq m} \frac{1}{n^{2d/2}(m-k+1)^{2d/2}} \leq c_n^{-d+1} \sum_{0 \leq k \leq m} \frac{1}{(m-k+1)^{2d/2}}.
\]
This yields (7.10).

The upper bound is trivial for \(d = 3\) and for \(d \geq 5\) it follows from (7.9) and the inequality
\[
P\{K_n \geq 1\} \leq \mathbb{E}[K_n].
\]
Assume \(d = 4\). We will consider \(\mathbb{E}[K_n | K_n \geq 1]\). On the event \(\{K_n \geq 1\}\), let \(k\) be the smallest integer \(\geq 2n\) such that \(S_k \in S[0,n]\). Let \(j\) be the smallest index such that \(S_j = S_k\). Then by the Markov property, given \(\{S_0, \ldots, S_j\}\), the expected value of \(K_n\) is
\[
\sum_{j=0}^{n} \sum_{k=j}^{n} \mathbb{P}(S_j = S_k | S_k = S_j) = \sum_{j=0}^{n} G(S_j - S_j).
\]
Define a random variable, depending on \(S_0, \ldots, S_n\),
\[
Y_n = \min \sum_{j=0}^{n} G(S_j - S_j).
\]
For any \(r > 0\), we can see that
\[
\mathbb{E}[K_n | K_n \geq 1, Y_n \geq r \log n] \geq r \log n.
\]
Note that for each \(r\),
\[
P\{Y_n < r \log n\} \leq (n+1) \mathbb{P}\left\{ \sum_{j=0}^{n} G(S_j) < r \log n \right\}.
\]
Using Lemma 7.7.2 below, we can find an \(r\) such that \(P\{Y_n < r \log n\} = o(1/\log n)\). But,
\[
c \geq \mathbb{E}[K_n] \geq \mathbb{P}\{K_n \geq 1; Y_n \geq r \log n\} \mathbb{E}[K_n | K_n \geq 1, Y_n \geq r \log n] \geq \mathbb{P}\{K_n \geq 1; Y_n \geq r \log n\} r \log n.
\]
Therefore,
\[
P\{K_n \geq 1\} \leq \mathbb{P}\{Y_n < r \log n\} + \mathbb{P}\{K_n \geq 1; Y_n \geq r \log n\} \leq \frac{c}{\log n}.
\]
This finishes the proof except for the one lemma that we will now prove. \(\square\)

Lemma 7.7.2 Let \(p \in \mathcal{P}_d\).

(a) For every \(\alpha > 0\), there exist \(c, r\) such that for all \(n\) sufficiently large,
\[
P\left\{ \sum_{j=0}^{n} G(S_j) \leq r \log n \right\} \leq c n^{-\alpha}.
\]

(b) For every \(\alpha > 0\), there exist \(c, r\) such that for all \(n\) sufficiently large,
\[
P\left\{ \sum_{j=0}^{n} G(S_j) \leq r \log n \right\} \leq c n^{-\alpha}.
\]

Proof. It suffices to prove (a) when \(n = 2^k\) for integer \(k\) and we write \(\xi^k = \xi_j\). Since \(G(x) \geq c/|x|^d\), we have
\[
\sum_{j=0}^{n} G(S_j) \geq \sum_{j=0}^{2^k} G(S_j) \geq c \sum_{k=1}^{n} 2^{-kd} (\xi^k - \xi^{k-1}).
\]
The reflection principle (Proposition 1.5.2) and the central limit theorem show that for every \(\epsilon > 0\), there is a \(\delta > 0\) such that if \(n\) is sufficiently large, then \(\mathbb{P}\{\xi^k \leq \delta n^{1/4}\} \leq \epsilon\). Let \(\ell_k\) denote the indicator function of the event \(\{\xi^k \leq \delta n^{1/4}\}\). Then we know that
\[
P(\ell_k = 1 | S_0, \ldots, S_k) \leq \epsilon.
\]
Therefore, \(J_k := \sum_{k} \ell_k\) is stochastically bounded by a binomial random variable with parameters \(l \) and \(\epsilon\). By exponential estimates for binomial random variables (see Lemma 8.2.7), we can find an \(\epsilon\) such that
\[
P\{J_k \geq 1/2\} \leq e^{-\alpha k}.
\]
But on the event \(\{J_k < 1/2\}\) we know that
\[
G(S) \geq c(1/2) \delta \geq r \log n,
\]
where the \(r\) depends on \(\alpha\).

For part (b) we need only note that \(\mathbb{P}\{n < \xi^k\}\) decays faster than any power of \(n\) and
\[
P\left\{ \sum_{j=0}^{n} G(S_j) \leq r \log n \right\} \leq P\left\{ \sum_{j=0}^{n} G(S_j) \leq r \log n^{1/4} \right\} + P\{n < \xi^k\}.
\]
\(\square\)

7.8 Long time behavior

Let \(S_n\) denote a random walk with increment distribution \(p \in \mathcal{P}_d, d \geq 3\). Let \(I_n\) denote the indicator function that \(S_n\) is "erased" in loop-erasing. To be more precise, \(I_n\) is the indicator function of the event
\[
\text{LE}(S_0, \ldots, S_n) \cap \{S_{n+1} = \infty\} = \emptyset.
\]
Chapter 8

Appendix

8.1 Some expansions

8.1.1 Riemann sums

One often approximates sums by integrals. Here we will discuss some standard approximations with sufficient care to discuss the size of the errors.

Suppose \( f : [0, \infty) \to \mathbb{R} \) is a \( C^1 \) function. If \( n \) is a positive integer, Taylor's series shows that for \( |s-n| \leq 1/2 \),

\[
f(s) = f(n) + (s-n) f'(n) + \frac{1}{2} f''(r_s)(s-n)^2,
\]

for some \( |n-r_s| < 1/2 \). Using this we get the following lemma.

**Lemma 8.1.1** If \( f : [0, \infty) \to \mathbb{R} \) is a \( C^1 \) function, and \( b_n \) is defined by

\[
b_n = f(n) - \int_{n-1/2}^{n+1/2} f(s) \, ds,
\]

then

\[
|b_n| \leq \frac{1}{8} \sup_{|s-n| \leq \frac{1}{2}} \left| f''(r) \right| |n - r| \leq \frac{1}{2}.
\]

If the \( b_n \) are absolutely convergent, let

\[
C = \sum_{n=1}^{\infty} b_n, \quad B_n = \sum_{j=n}^{\infty} |b_j|.
\]

Then

\[
\sum_{j=1}^{n} f(j) = \int_{1/2}^{n+1/2} f(s) \, ds + C + O(|B_n|),
\]

Also, for all \( m < n \)

\[
\left| \sum_{j=m}^{n} f(j) - \int_{m+1/2}^{n+1/2} f(s) \, ds \right| \leq B_m.
\]

**Example.** Suppose \( \alpha < 1, \beta \in \mathbb{R} \) and

\[
f(n) = n^n \log^\beta n.
\]

Note that for \( t \geq 2 \),

\[
f''(t) \leq \alpha t^{\alpha-1} \log^\beta t.
\]

Therefore, there is a \( C(\alpha, \beta) \) such that

\[
\sum_{j=1}^{n} n^n \log^\beta n = \int_{1}^{n+1/2} t^n \log^\beta t \, dt + C(\alpha, \beta) + O(n^{n-1} \log^\beta n)
\]

\[
= \int_{2}^{n+1/2} t^n \log^\beta t \, dt + \frac{1}{2} n^n \log^\beta n + C(\alpha, \beta) + O(n^{n-1} \log^\beta n) \tag{8.1}
\]

8.1.2 Logarithm

Here we collect some facts about the logarithm. The starting point will be the series for the logarithm

\[
\log(1+z) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1} z^k}{k}, \quad |z| \leq 1 - \epsilon.
\]

Therefore, if \( r \in (0, 1) \) and \( |k| \leq rt \),

\[
\log \left( 1 + \frac{z}{t} \right) = \frac{z}{t} - \frac{z^2}{2t} + \frac{z^3}{3t^2} + \cdots + (-1)^{k+1} \frac{z^k}{k t^{k-1}} + O \left( \frac{|z|^{k+1}}{k t^k} \right),
\]

\[
\left( 1 + \frac{z}{t} \right)^t = e^z \exp \left( \frac{z}{2t} + \frac{z^2}{3t^2} + \cdots + (-1)^{k+1} \frac{z^k}{k t^{k-1}} + O \left( \frac{|z|^{k+1}}{k t^k} \right) \right). \tag{8.2}
\]

The last equation is valid for \( |z| \leq rt \). If \( |k|/t \) is not too big, we can expand the exponential in a Taylor series. Recall that for fixed \( R < \infty \), we can write

\[
e^z = 1 + z + \frac{z^2}{2} + \cdots + \frac{z^k}{k!} + O(z^{k+1}), \quad |z| \leq R.
\]

Therefore, if \( r \in (0, 1), R < \infty, \quad |k| \leq rt, \quad |z| \leq Rr, \) we can write

\[
\left( 1 + \frac{z}{t} \right)^t = e^z \left[ 1 - \frac{z}{2t} + \frac{8z^2}{24t^2} + \cdots + \frac{b_k(z)}{k t^{k-1}} + O \left( \frac{|z|^{k+1}}{k t^k} \right) \right], \tag{8.3}
\]

where \( b_k \) is a polynomial of degree \( 2k \) and the implicit constant in the \( O(z) \) term depends only on \( r, R \) and \( k \). In particular,

\[
\left( 1 + \frac{z}{n} \right)^n = e \left[ 1 - \frac{1}{2n} + \frac{11}{24n^2} + \cdots + \frac{b_k(z)}{k t^{k-1}} + O \left( \frac{1}{n^{k+1}} \right) \right], \tag{8.4}
\]
Lemma 8.1.2 There exists a constant \( \gamma \) (called Euler's constant) and \( b_1, b_2, \ldots \) such that for every integer \( k \geq 2 \),

\[
\sum_{j=1}^{n} \frac{1}{j} = \log n + \gamma + \frac{1}{2n} + \sum_{j=2}^{n} \frac{b_j}{j} + O \left( \frac{1}{n^{k+1}} \right).
\]

In fact,

\[
\gamma = \int_{1}^{\infty} \left( 1 - e^{-t} \right) \frac{1}{t} dt = \int_{1}^{\infty} e^{-t} \left( 1 - t \right) dt. \quad (8.5)
\]

Proof. We start by writing

\[
\sum_{j=1}^{n} \frac{1}{j} = \log \left( n + \frac{1}{2} \right) + \log 2 + \sum_{j=1}^{n} \beta_j,
\]

where

\[
\beta_j = \frac{1}{j} - \log \left( j + \frac{1}{2} \right) + \log \left( j - \frac{1}{2} \right) = \sum_{k=1}^{\infty} \frac{2}{(2k+1)(2j+2k+1)}.
\]

In particular, \( \beta_j = O(j^{-2}) \), and hence \( \sum \beta_j < \infty \). We can write

\[
\sum_{j=1}^{n} \frac{1}{j} - \log \left( n + \frac{1}{2} \right) - \sum_{j=1}^{n} \beta_j = \log n + \gamma + \sum_{n=1}^{\infty} \frac{(n+1)^{k+1}}{n^{2k+1}} - \sum_{j=1}^{n} \beta_j,
\]

where \( \gamma \) is the constant

\[
\gamma = \log 2 + \sum_{j=1}^{\infty} \beta_j.
\]

We can take as many terms as we wish of this expansion.

We will sketch the proof of (8.5) leaving the details to the reader. By Taylor's series, we know that

\[
\log n = \sum_{j=1}^{n} \left( 1 - \frac{1}{j} \right) \frac{1}{j}.
\]

Therefore,

\[
\gamma = \lim_{n \to \infty} \left[ \sum_{j=1}^{n} \frac{1}{j} \right] - \log n
\]

\[
= \lim_{n \to \infty} \left[ \log n - \sum_{j=1}^{n} \left( 1 - \frac{1}{j} \right) \frac{1}{j} \right] = \lim_{n \to \infty} \left[ \log n - \int_{1}^{n} \left( 1 - e^{-t} \right) \frac{1}{t} dt \right] = \lim_{n \to \infty} \left[ \log n - \int_{1}^{n} (1 - e^{-t}) \frac{1}{t} dt \right].
\]

Lemma 8.1.3 Suppose \( \alpha, \beta \in \mathbb{R} \) and \( m \) is a positive integer. There exist constants \( c_0, c_1, \ldots \) such that if \( k \) is a positive integer and \( n \geq m \),

\[
\prod_{j=m}^{n} \left( 1 - \frac{\alpha}{j} \right) = c_0 \log n + c_1 \log \sqrt{n} + c_2 \log \log n + c_3 + O \left( \frac{1}{n^{1/2}} \right).
\]

Proof. Without loss of generality we assume that \( |\alpha| \leq 2m \); if this does not hold we can factor out the first few terms of the product and then analyze the remaining terms. Note that

\[
\log \prod_{j=m}^{n} \left( 1 - \frac{\alpha}{j} \right) = \sum_{j=m}^{n} \log \left( 1 - \frac{\alpha}{j} \right) = -\sum_{j=m}^{n} \sum_{k=1}^{\infty} \frac{\alpha^k}{kj} = -\sum_{k=1}^{\infty} \sum_{j=m}^{n} \frac{\alpha^k}{kj}.
\]

For the \( l = 1 \) term we have

\[
\sum_{j=m}^{n} \frac{\alpha^1}{j} = -\sum_{j=1}^{m} \frac{\alpha^1}{j} + \log n + \gamma + \frac{1}{2n} + \sum_{k=2}^{\infty} \frac{1}{k} \frac{\alpha^1}{k} + O \left( \frac{1}{n^{1/2}} \right).
\]

All of the other terms can be written in powers of \((1/m)\). Therefore, we can write

\[
\log \prod_{j=m}^{n} \left( 1 - \frac{\alpha}{j} \right) = -\alpha \log n + c_2 \log \log n + c_3 + O \left( \frac{1}{n^{1/2}} \right).
\]

The lemma is then obtained by exponentiating both sides.

If \( \alpha \in \mathbb{R} \), the binomial series gives the expansion

\[
(1 + z)^{\alpha} = \sum_{j=0}^{\infty} r(\alpha, j) z^j + O(|z|^{k+1}), \quad |z| < 1/2,
\]

where the \( O(\cdot) \) term depends on \( \alpha \) and

\[
r(\alpha, j) = \frac{\alpha(\alpha - 1) \cdots (\alpha - (j-1))}{j!}.
\]

If \( \alpha > 0 \), \( |r(\alpha, j)| \leq C_\alpha \). If \( \alpha < 0 \), then

\[
r(\alpha, j) = -\frac{|\alpha|(\alpha + 1) \cdots (\alpha + j - 1)}{j!} = \frac{\alpha^j}{j!} \prod_{k=1}^{j} \left( 1 + \frac{|\alpha|}{k} \right) \times j \cdot 1.
\]

In particular, the series converges absolutely for \( |z| < 1 \).
8.1. SOME EXPANSIONS

Lemma 8.1.4 If $a > 0$, there exist $b(1, a), b(2, a), \ldots$ such that for every positive integer $k$,

$$
a \sum_{j=1}^{\infty} j^{a(m+1)} = n^{\infty} \left[ 1 + \sum_{j=1}^{k} b(j, a) n^{j} + O(n^{a(m+1)}) \right].
$$

Proof. Let $F(s, n) = \sum_{j=1}^{n} j^{s}$. Comparison to an integral shows that $(s - 1) F(s, n) \sim n^{s-1}$ for $s > 1$. Using the binomial expansion, we have

$$
n^{a} = \sum_{j=1}^{\infty} j^{a} = (j + 1)^{a} - 1
$$

$$
= \sum_{j=1}^{\infty} j^{a} \left[ 1 + (1 + 1)^{a} \right]
$$

$$
= \sum_{j=1}^{\infty} j^{a} \left[ O(j^{a+1}) = \sum_{m=1}^{\infty} r(-a, m) j^{m} \right]
$$

$$
= O(j^{a+1}) = \sum_{m=1}^{k} F(a + m, n) r(-a, m)
$$

This gives

$$
a \sum_{j=1}^{\infty} j^{a(m+1)} = a F(a + 1, n) = n^{a} + \sum_{m=2}^{\infty} F(a + m, n) r(-a, m) + O(n^{a(m+1)}).$$

By iterating this with $F(a + 2, n), F(a + 3, n), \ldots$, we get the lemma.

Lemma 8.1.5 If $a > 0$, there exist $c(1, a), c(2, a), \ldots$ such that for every positive integer $k$,

$$
a \sum_{j=1}^{\infty} j^{a(m)} = c_{m} + n^{a} \left[ 1 + \sum_{j=1}^{k} c(j, a) n^{j} + O(n^{a(m+1)}) \right].$$

Proof. Using the binomial expansion, we write

$$
n^{a} - 2^{a} = \sum_{j=2}^{\infty} (j + 1)^{a} - j^{a} = \sum_{j=2}^{\infty} j^{a} \left[ 1 + \frac{1}{j} \right] - 1
$$

$$
= \sum_{j=2}^{\infty} j^{a} \sum_{m=1}^{\infty} \frac{r(a, m)}{j^{m}}
$$

$$
= \sum_{m=1}^{\infty} r(a, m) \sum_{j=2}^{\infty} j^{a(m-1)}$$

(8.6)

We have therefore written

$$
a \sum_{j=1}^{\infty} j^{a(m)} = c_{m} + n^{a} + \sum_{m=2}^{\infty} r(a, m) \sum_{j=2}^{\infty} j^{a(m-1)}.$$

where

$$
c_{m} = 1 - 2^{a} = \sum_{m=2}^{\infty} r(a, m) [\zeta(m - a) - 1].$$

Note that

$$
\left| \sum_{m=k+1}^{\infty} r(a, m) \sum_{j=2}^{\infty} j^{a(m-1)} \right| \leq C(a) \sum_{m=k+1}^{\infty} r(a, m) n^{a(m+1)} = O(n^{a(m+1)}).
$$

For $2 \leq m \leq k + 1$, we can Lemma 8.1.3. This establishes the result for $0 < a < 1$. For integer $a \geq 1$, the result is immediate, and for noninteger $a > 1$ we can use (8.6) and induction.

8.2 Martingales

Martingales are one of the most important concepts in probability. Recall that a filtration $\mathcal{F}_{0} \subseteq \mathcal{F}_{1} \subseteq \cdots$ is an increasing sequence of $\sigma$-algebras.

Definition. A sequence of integrable random variables $M_0, M_1, \ldots$ is called a martingale with respect to the filtration $\{\mathcal{F}_{n}\}$ if each $M_n$ is $\mathcal{F}_n$-measurable and for each $n \leq m$,

$$
\mathbb{E} M_n \mid \mathcal{F}_m = M_m. \quad (8.7)
$$

If (8.7) is replaced with $\mathbb{E} M_n \mid \mathcal{F}_m \geq M_m$, the sequence is called a submartingale. If (8.7) is replaced with $\mathbb{E} M_n \mid \mathcal{F}_m \leq M_m$ the sequence is called a supermartingale.
8.2. MARTINGALES

Using properties of conditional expectation, it is easy to see that to verify (8.7) it suffices to show for each \( n \) that \( \mathbb{E}[M_{n+1} | \mathcal{F}_n] = M_n \). This inequality only needs to hold up to an event of probability zero; in fact, the conditional expectation is only defined up to events of probability zero. If the filtration is not specified, then the assumption is that \( \mathcal{F}_n \) is the \( \sigma \)-algebra generated by \( M_0, \ldots, M_n \). If \( M_0, X_1, X_2, \ldots \) are independent random variables with \( \mathbb{E}[M_0] < \infty \) and \( \mathbb{E}[X_j] = 0 \) for \( j \geq 1 \), and

\[
M_n = M_0 + X_1 + \cdots + X_n,
\]

then \( M_0, M_1, \ldots \) is a martingale. We omit the proof of the next lemma which is the conditional expectation version of Jensen’s inequality.

Lemma 8.2.1 (Jensen’s inequality) If \( X \) is an integrable random variable; \( f : \mathbb{R} \to \mathbb{R} \) is convex with \( \mathbb{E}[|f(X)|] < \infty \); and \( \mathcal{F} \) is a \( \sigma \)-algebra, then \( \mathbb{E}[f(X) | \mathcal{F}] \geq f(\mathbb{E}[X | \mathcal{F}]) \). In particular, if \( M^1, M^2, \ldots \) is a martingale; \( f : \mathbb{R} \to \mathbb{R} \) is convex with \( \mathbb{E}[f(M_n)] < \infty \) for all \( n \); and \( Y_n = f(M_n) \); then \( Y_0, Y_1, \ldots \) is a submartingale.

In particular, if \( M_0, M_1, \ldots \) is a martingale then

- if \( \alpha \geq 1 \), \( Y_n = |M_n|^\alpha \) defines a submartingale;
- if \( \beta \in \mathbb{R} \), then \( Y_n = e^{\beta M_n} \) defines a submartingale.

In both cases, this is assuming that \( \mathbb{E}[Y_n] < \infty \).

8.2.1 Optional Sampling Theorem

As before a stopping time with respect to a filtration \( \{\mathcal{F}_n\} \) is a \( \{0, 1, \ldots \} \cup \{\infty\} \) valued random variable \( T \) such that for each \( n \), \( \{T \leq n\} \in \mathcal{F}_n \). If \( T \) is a stopping time, and \( n \) is a positive integer, then \( T_n := T \wedge n \) is a stopping time satisfying \( T_n \leq n \).

Proposition 8.2.2 Suppose \( M_0, M_1, \ldots \) is a martingale and \( T \) is a stopping time with respect to the filtration \( \mathcal{F}_n \). Then \( Y_0 := M_{T_n} \) is a martingale with respect to \( \mathcal{F}_n \). In particular,

\[
\mathbb{E}[M_n] = \mathbb{E}[M_{T_n}].
\]

Proof. Note that

\[
Y_{n+1} = M_{T_n} \mathbb{1}_{\{T \leq n\}} + M_{n+1} \mathbb{1}_{\{T \geq n+1\}}.
\]

The event \( \{T \geq n+1\} \) is the complement of the event \( \{T \leq n\} \) and hence is \( \mathcal{F}_n \)-measurable. Therefore, by properties of conditional expectation,

\[
\mathbb{E}[M_{n+1} | \mathcal{F}_n] \mathbb{1}_{\{T \geq n+1\}} = \mathbb{E}[M_{n+1} | \mathcal{F}_n] M_n \mathbb{1}_{\{T \geq n+1\}} = M_n \mathbb{1}_{\{T \geq n+1\}} = \mathbb{E}[M_n | \mathcal{F}_n] = \mathbb{E}[Y_n | \mathcal{F}_n] = Y_n.
\]

Theorem 8.2.3 (Optional Sampling Theorem) Suppose \( M_0, M_1, \ldots \) is a martingale and \( T \) is a stopping time with respect to the filtration \( \{\mathcal{F}_n\} \). Suppose that \( P(T < \infty) = 1 \) and \( \mathbb{E}[|M_T|] < \infty \). Suppose also that at least one of the following conditions holds:

- There is a \( K < \infty \) such that \( P(T \leq K) = 1 \),
- There exists an integrable random variable \( Y \) such that for all \( n \), \( |M_{T_n}| \leq Y \),
- \( \lim_{n \to \infty} \mathbb{E}[|M_n| T > n] = 0 \),
- The random variables \( M_0, M_1, \ldots \) are uniformly integrable, i.e., for every \( \epsilon > 0 \) there is a \( K < \infty \) such that for all \( n \),

\[
\mathbb{E}[|M_n| | T > n] > K \gamma < \epsilon.
\]
- There exists an \( \alpha > 1 \) and a \( K < \infty \) such that for all \( n \), \( \mathbb{E}[|M_n|^\alpha] \leq K \).

Then \( \mathbb{E}[M_0] = \mathbb{E}[M_T] \).

Proof. We will consider the conditions in order. The sufficiency of the first follows immediately from Proposition 8.2.2. We know that \( M_{T_n} \to M_T \) with probability one. Proposition 8.2.2 gives \( \mathbb{E}[M_{T_n}] = \mathbb{E}[M_0] \). Hence we need to show that

\[
\lim_{n \to \infty} \mathbb{E}[M_{T_n}] = \mathbb{E}[M_T], \tag{8.8}
\]

If the second condition holds, then this limit is justified by the dominated convergence theorem. Now assume the third condition. Note that

\[
M_T = M_{T_n} \mathbb{1}_{\{T > n\}} + M_n \mathbb{1}_{\{T \leq n\}}.
\]

Since \( P(T > n) \to 0 \) and \( \mathbb{E}[|M_T|] < \infty \), it follows from standard arguments that

\[
\lim_{n \to \infty} \mathbb{E}[M_T \mathbb{1}_{\{T > n\}}] = 0.
\]

Hence if \( \mathbb{E}[M_n \mathbb{1}_{\{T > n\}}] \to 0 \), we have (8.8). Standard exercises show that the fourth implies the third and the fifth condition implies the fourth, so either the fourth or fifth condition is sufficient. In fact, either of these conditions implies \( \mathbb{E}[|M_T|] < \infty \) so we do not need to separately assume that in these cases. \( \square \)
8.2.2 Maximal inequality

Here we prove an important maximal inequality that can be considered a generalization of Chebyshev’s inequality.

**Theorem 8.2.4 (Maximal inequality)** Suppose $M_0, M_1, \ldots$ is a nonnegative submartingale with respect to $\{F_n\}$ and $\lambda > 0$. Then

$$
P \left( \max_{0 \leq n \leq \lambda} M_j \geq \lambda \right) \leq \frac{\mathbb{E}[M_0]}{\lambda}.
$$

**Proof.** Let $T = \min\{j \geq 0 : M_j \geq \lambda\}$. Then

$$
P \left( \max_{0 \leq n \leq \lambda} M_j \geq \lambda \right) = \sum_{j=0}^{\infty} P\{T = j\},
$$

$$
\mathbb{E}[M_0] \geq \mathbb{E}[M_n; T \leq n] = \sum_{j=0}^{\infty} \mathbb{E}[M_n; T = j] \geq \lambda P\{T = j\}.
$$

Since $M_n$ is a submartingale and $\{T = j\}$ is $F_j$-measurable,

$$
\mathbb{E}[M_n; T = j] = \mathbb{E}[\mathbb{E}[M_n | F_j]; T = j] \geq \lambda \mathbb{E}[M_j; T = j] \geq \lambda P\{T = j\}.
$$

Combining these estimates gives the theorem.

By combining Theorem 8.2.4 with Lemma 8.2.1 gives the following theorem as a corollary.

**Theorem 8.2.5 (Martingale maximal inequalities)** Suppose $M_0, M_1, \ldots$ is a martingale with respect to $\{F_n\}$ and $\lambda > 0$. Then if $a \geq 1, b \geq 0$,

$$
P \left( \max_{0 \leq n \leq \lambda} |M_j| \geq \lambda \right) \leq \frac{\mathbb{E}[|M_0|]}{\lambda a},
$$

$$
P \left( \max_{0 \leq n \leq \lambda} M_j \geq \lambda \right) \leq \frac{\mathbb{E}[|M_0|]}{e^b}. 
$$

**Corollary 8.2.6** Let $X_1, X_2, \ldots$ be independent, identically distributed random variables in $\mathbb{R}$ with mean zero, variance $\sigma^2$, and such that for some $\delta > 0$, the moment generating function $\psi(t) = \mathbb{E}[e^{tX}]$ exists for $|t| < \delta$. Let $S_n = X_1 + \ldots + X_n$. Then there is an $N$ such that for all $n \geq N$ and all $a > 0$,

$$
P \left( \max_{0 \leq n \leq \lambda} S_j \geq a \sqrt{n} \right) \leq e^{-a^2 \sigma^2}. 
$$

Moreover, for every $\epsilon > 0$, there exists a $\rho = \rho_\epsilon > 0$ such that for all $n$,

$$
P \left( \max_{0 \leq n \leq \lambda} S_j \geq \epsilon n \right) \leq e^{-\rho n}. 
$$

**8.2.3 Continuous martingales**

A process $M_t$ adapted to a filtration $\mathcal{F}_t$ is called a **continuous martingale** if for each $s < t$, $\mathbb{E}[M_t | M_s] = M_s$ and with probability one the function $t \mapsto M_t$ is continuous. If $M_t$ is a continuous martingale, and $\delta > 0$, then

$$
M^{(\delta)}_n := M_{n\delta}
$$

is a discrete time martingale. Using this, we can extend results about discrete time martingales to continuous martingales. We state one such result here.

---

*Here the word "continuous" refers both to continuous time and continuous paths.*
Theorem 8.2.8 (Optional Sampling Theorem) Suppose $M_t$ is a uniformly integrable continuous martingale and $\tau$ is a stopping time with $\mathbb{P}\{\tau < \infty\} = 1$ and $\mathbb{E}|M_\tau| < \infty$. Suppose that

$$\lim_{t \to \infty} \mathbb{E}|M_t| : \tau > t = 0,$$

Then

$$\mathbb{E}|M_\tau| = \mathbb{E}|M_0|.$$

8.3 Joint normal distributions

A random vector $Z = (Z_1, \ldots, Z_d) \in \mathbb{R}^d$ is said to have a (mean zero) joint normal distribution if there exist independent (one-dimensional) mean zero, variance one normal random variables $N_1, \ldots, N_d$ and scalars $\sigma_{jk}$ such that

$$Z_j = \sigma_{j1} N_1 + \cdots + \sigma_{jm} N_m, \quad j = 1, \ldots, d,$$

or in matrix form

$$Z = AN,$$

where $A = (\sigma_{jk})$ is a $d \times n$ matrix and $Z, N$ are column vectors. Note that

$$\mathbb{E}(Z_j Z_k) = \sum_{m=1}^n \sigma_{jm} \sigma_{km}.$$

In other words, the covariance matrix $\Gamma = \mathbb{E}(Z_j Z_k)$ is the $d \times d$ symmetric matrix

$$\Gamma = AA^T.$$

We say $Z$ has a nondegenerate distribution if $\Gamma$ is invertible.

The characteristic function of $Z$ can be computed easily using the known formula for the characteristic function of $N$, \(\mathbb{E} e^{i\theta N_1} = e^{-\theta^2/2}\),

$$\mathbb{E}\exp\{i\theta \cdot Z\} = \prod_{k=1}^d \mathbb{E}\left\{ \exp\left\{ i \sum_{j=1}^d \theta_j N_j \right\} \right\} = \prod_{k=1}^d \mathbb{E}\left\{ \exp\left\{ i N_k \sum_{j=1}^d \theta_j \sigma_{jk} \right\} \right\} = \prod_{k=1}^d \exp\left\{ -\frac{1}{2} \left( \sum_{j=1}^d \theta_j \sigma_{jk} \right)^2 \right\}.$$

Since the characteristic function determines the distribution, we see that the distribution of $Z$ depends only on $\Gamma$.

The matrix $\Gamma$ is symmetric and nonnegative definite. Hence we can find an orthogonal basis $u_1, \ldots, u_d$ of unit vectors in $\mathbb{R}^d$ that are eigenvectors of $\Gamma$ with nonnegative eigenvalues $\sigma_1, \ldots, \sigma_d$. The random variable $Z = \sqrt{\sigma_1} u_1 + \cdots + \sqrt{\sigma_d} u_d$ has a joint normal distribution with covariance matrix $\Gamma$. In matrix language, we have written $\Gamma = \Lambda \Lambda^T = \Sigma^2$ for a $d \times d$ nonnegative definite symmetric matrix $\Lambda$. The distribution is nondegenerate if and only if all the $\sigma_j$ are strictly positive.

Heuristic note. Although we allow the matrix $A$ to have $n$ columns, what we have shown is that there is a symmetric, positive definite $d \times d$ matrix $\Lambda$ which gives the same distribution. Hence joint normal distribution in $\mathbb{R}^d$ can be described as linear combinations of $d$ independent one-dimensional normals. Moreover, if we choose the correct orthogonal basis for $\mathbb{R}^d$, the components of $Z$ with respect to that basis are independent normals.

If $\Gamma$ is invertible, then $Z$ has a density $f(z^1, \ldots, z^d)$ with respect to Lebesgue measure that can be computed using the inversion formula

$$f(z^1, \ldots, z^d) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2} \mathbb{E}(\exp\{i\theta \cdot Z\})} = \frac{1}{2\pi} \int e^{-\frac{1}{2} \mathbb{E}(\exp\{i\theta \cdot Z\})} d\theta.$$

(Here and for the remainder of this paragraph the integrals are over $\mathbb{R}^d$ and $d\theta$ represents $d^d\theta$.) To evaluate the integral, we start with the substitution $\theta_1 = \Lambda \theta$ which gives

$$\int \exp\left\{ -\frac{1}{2} \theta^T \Lambda \theta \right\} d\theta = \frac{1}{\det \Lambda} \int e^{-\frac{1}{2} \theta^T \Lambda^{-1} \theta} d\theta.$$

By completing the square we see that the right-hand side equals

$$\frac{e^{-\frac{1}{2} \theta^T \Lambda^{-1} \theta}}{\det \Lambda} \int \exp\left\{ \frac{1}{2} \left( \theta_1 - \Lambda^{-1} \theta \right)^T \right\} \exp\left\{ \frac{1}{2} \left( \theta_1 - \Lambda^{-1} \theta \right)^T \right\} d\theta_0,$$

The substitution $\theta_1 = \theta - \Lambda^{-1} \theta$ gives

$$\int \exp\left\{ \frac{1}{2} \left( \theta_1 - \Lambda^{-1} \theta \right)^T \right\} d\theta_1 = \int e^{-\frac{1}{2} \theta_0^T \Lambda \theta_0} (2\pi)^{d/2},$$
Hence, the density of $Z$ is

$$f(z) = \frac{1}{(2\pi)^{d/2} \sqrt{\det T}} e^{-\frac{1}{2} z^{\intercal} T^{-1} z},$$

(8.12)

**Corollary 8.3.1** Suppose $Z = (Z_1, \ldots, Z_d)$ has a mean zero, joint normal distribution such that $\mathbb{E}[Z_i Z_j] = 0$ for all $i \neq j$. Then $Z_1, \ldots, Z_d$ are independent.

**Proof.** Suppose $\mathbb{E}[Z_i Z_j] = 0$ for all $i \neq j$. Then $Z$ has the same distribution as

$$(b_1 N_1, \ldots, b_d N_d),$$

where $b_j = \sqrt{\mathbb{E}[Z_j^2]}$. In this representation, the components are obviously independent. □

**Heuristic note.** If $Z_1, \ldots, Z_d$ are mean zero random variables satisfying $\mathbb{E}[Z_i Z_j] = 0$ for all $i \neq j$, they are called orthogonal. Independence implies orthogonality but the converse is not always true. However, the corollary tells us that the converse is true in the case of joint normal random variables. Orthogonality is often much easier to verify than independence.

### 8.4 Markov chains

A *(time-homogeneous)* Markov chain on a countable state space $D$ is a process $X_n$ taking values in $D$ whose transitions satisfy

$$P\{X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n\} = p(x_n, x_{n+1}),$$

where $p : D \times D \to [0,1]$ is the transition function satisfying $\sum_y p(x, y) = 1$ for each $x$. If $A$ is finite, we call the transition function the transition matrix $P = [p(x, y)]_{x,y \in A}$. The $n$-step transitions are given by the matrix $P^n$. In other words, if $p_n(x, y)$ is defined to be $P\{X_n = y \mid X_0 = x\}$, then

$$p_n(x, y) = \sum_{x_1, \ldots, x_{n-1}} p(x_1, y) p_{n-1}(x_1, x_2) \cdots p_1(x_1, x_2).$$

A Markov chain is called irreducible if for each $x, y \in A$, there exists an $n = n(x,y) \geq 0$ with $p_n(x,y) > 0$. The chain is aperiodic if for each $x$ there is an $N_x$ such that for $n \geq N_x$, $p_n(x,x) > 0$. If $D$ is finite, then the chain is irreducible and aperiodic if and only if there exists an $n$ such that $P^n$ has strictly positive entries.

**Theorem 8.4.1** *(Perron-Frobenius Theorem)* If $P$ is an $m \times m$ matrix such that for some positive integer $n$, $P^n$ has all entries strictly positive, then there exists $\alpha > 0$ and vectors $v, w$, with strictly positive entries such that

$$v P = \alpha v, \quad P w = \alpha w.$$

In particular, if $P$ is the transition matrix for an irreducible aperiodic Markov chain there is a unique invariant probability $\pi$ satisfying

$$\sum_{x \in D} \pi(x) = 1, \quad \pi(x) = \sum_{y \in D} \pi(y) P(y, x).$$

**Proof.** We first assume that $P$ has all strictly positive entries. It suffices to find a right eigenvector, since the left eigenvector can be handled by considering the transpose of $P$. We write $w_i \geq w_j$ if every component of $w_i$ is greater than the corresponding component of $w_j$. Similarly, we write $w_i > w_j$. Let $\mathbf{0}$ denote the zero vector and $e_j$ the vector that is 1 in the $j$th component and all other components are 0. If $w_j \geq 0$, let

$$\lambda_w = \sup \{ \lambda : P \mathbf{w} \leq \lambda \mathbf{w} \}.$$

Clearly $\lambda_w < \infty$ and since $P$ has strictly positive entries, $\lambda_w > 0$ for all $w_j > 0$. Let

$$\alpha = \sup \{ \lambda_w : w_j \geq 0, \sum_{j=1}^m |w_j| = 1 \}.$$

By compactness arguments we can see that there exists a $\omega$ with $w_j \geq 0$, $\sum_{j=1}^m |w_j| = 1$ and $\lambda_w = \alpha$. We claim that $P \mathbf{w} = \alpha \mathbf{w}$. Indeed if $[P \mathbf{w}]_j < \alpha |w_j|$ for some $j$, one can check that there exist positive $\epsilon, r$ such that $P \mathbf{w} + \epsilon e_j \leq (\alpha + \epsilon r) \mathbf{w} + \epsilon e_j$, which contradicts the maximality of $\alpha$. If $v$ is a vector with both positive and negative values, then for each $j$

$$[P \mathbf{v}]_j < [P \mathbf{v}]_j \leq \alpha |v|.$$
8.4. MARKOV CHAINS

Proposition 8.4.2 Suppose \( p : D \times D \to [0, 1] \) is the transition probability for a positive recurrent, irreducible, aperiodic Markov chain on a state space \( D \). Let \( \pi \) denote the invariant probability measure. Suppose that \( k \) is a positive integer such that for all \( x, y \in D \),

\[
\frac{1}{2} \sum_{z \in D} |p(z, y) - p(z, x)| \leq 1 - \epsilon.
\]

Then for all positive integers \( j \) and all \( x \in A \),

\[
\frac{1}{2} \sum_{z \in D} |p(z, y) - \pi(z)| \leq c e^{-\alpha j},
\]

where \( c = (1 - \epsilon)^{-1} \) and \( \epsilon^{\alpha} = (1 - \epsilon)^{1/\beta} \).

Proof. If \( \nu \) is any probability distribution, let us write

\[
\nu_j(x) = \sum_{z \in D} \nu(z)p_j(z, x).
\]

Then (8.13) implies that for every \( \nu \),

\[
\frac{1}{2} \sum_{z \in D} |\nu_j(z) - \pi(z)| \leq 1 - \epsilon.
\]

In other words we can write \( \nu_j = \epsilon \pi + (1 - \epsilon)\nu^{(i)} \) for some probability measure \( \nu^{(i)} \). By iterating (8.13), we can see that for every integer \( i \geq 1 \) we can write \( \nu_{ik} = (1 - \epsilon)^i \nu^{(i)} + (1 - (1 - \epsilon)^i)\pi \) for some probability measure \( \nu^{(i)} \). This establishes the result for \( j = ki \) with \( c = 1 \) for these values of \( j \) and for other \( j \) we find \( i \) with \( \epsilon^i \leq j < (i + 1)\epsilon \).

We will now consider Markov chains restricted to a subset of the original state space. If \( X_n \) is an irreducible, aperiodic Markov chain with state space \( D \) and \( A \) is a finite proper subset of \( D \), we write \( P_A = [p(x, y)]_{x \in A, y \in A} \). Note that \( (P_A)^n = [p^n_A(x, y)]_{x \in A, y \in A} \) where

\[
p^n_A(x, y) = \mathbb{P}\{X_n = y, x \in A \mid X_n = x\}.
\]

We call \( A \) connected and aperiodic (with respect to \( P \)) if for each \( x, y \in A \), there is an \( n \) such that \( p^n_A(x, y) > 0 \). If \( A \) is finite, then \( A \) is connected and aperiodic if and only if there exist an \( n \) such that \( (P_A)^n \) has all entries strictly positive. In this case all of the rows sums of \( P_A \) are less than or equal to one and (since \( A \) is a proper subset) there is at least one row whose sum is strictly less than one.

Suppose \( X_n \) is an irreducible, aperiodic Markov chain with state space \( D \) and \( A \) is a finite, connected, aperiodic proper subset of \( D \). Let \( \alpha \) be as in the Perron-Frobenius Theorem. It is easy to see that \( 0 < \alpha < 1 \). Let \( v, w \) be the corresponding eigenvectors; we will write these as functions,

\[
\sum_{x \in A} v(x)p(x, y) = \alpha v(y), \quad \sum_{x \in A} w(y)p(x, y) = \alpha w(x).
\]

We will normalize the functions so that

\[
\sum_{x \in A} v(x) = 1, \quad \sum_{x \in A} w(x) = 1,
\]

and we let \( \pi = v(x)w(x) \). Let

\[
q^n_A(x, y) = \alpha^{-n} p^n_A(x, y) w(y) / w(x)
\]

Note that

\[
\sum_{y \in A} q^n_A(x, y) = 1
\]

In other words, \( Q^n := [q^n_A(x, y)]_{x \in A} \) is the transition matrix for a Markov chain which we will denote by \( Y_n \). Note that \( (Q^n)^n = [q^n_A(x, y)]_{x \in A} \) where

\[
q^n_A(x, y) = \alpha^{-n} p^n_A(x, y) w(y) / w(x)
\]

and

\[
p^n_A(x, y) = \mathbb{P}\{X_n = y, x \in A \mid X_n = x\}.
\]

From this we see that the chain is irreducible and aperiodic. Since

\[
\sum_{x \in A} \pi(x) q^n_A(x, y) = \sum_{x \in A} v(x) w(x) \alpha^{-n} p^n_A(x, y) w(y) / w(x) = \pi(y).
\]

we see that \( \pi \) is the invariant probability for this chain.

Proposition 8.4.3 Under the assumptions above, there exist \( c, \beta \) such that for all \( n \),

\[
|e^{-\alpha n} p^n_A(y, x) - v(x) w(y)| \leq c e^{-\beta n}.
\]

In particular,

\[
\mathbb{P}\{X_n = y \mid X_n = x\} = w(x) e^{-\alpha n} [1 + O(e^{-\beta n})].
\]

Proof. Consider the Markov chain with transition matrix \( Q^n \). Choose positive integer \( k \) and \( \epsilon > 0 \) such that \( q^n_A(x, y) \geq \epsilon \pi(y) \) for all \( x, y \in A \). Proposition 8.4.2 implies that

\[
|e^{-\alpha n} p^n_A(y, x) - \pi(y)| \leq c e^{-\beta n},
\]

for some \( c, \beta \). Since \( \pi(y) = v(y) w(y) \) and \( q^n_A(x, y) = \alpha^{-n} p^n_A(x, y) w(y) / w(x) \), we get the first assertion. The second assertion follows from the first using \( \sum_{y \in A} v(y) = 1 \) and

\[
\mathbb{P}\{X_n = y \mid X_n = x\} = \sum_{y \in A} p^n_A(x, y).
\]
Heuristic note. The chain $Y_n$ can be considered the chain derived from $X_n$ by conditioning the chain to "stay in $A$ forever". The probability measures $v, \pi$ are both "invariant probability measures", but with different interpretations. Roughly speaking, the three measures $v,w,\pi$ can be described as follows:

- Suppose the chain $X_n$ is observed at a large time $n$ and it known that the chain has stayed in $A$ for all times up to $n$. Then the conditional distribution on $X_n$ given this information approaches $v$.
- For $z \in A$, the probability that the chain stays in $A$ up to time $n$ is asymptotic to $w(x) a^n$.
- Suppose the chain $X_n$ is observed at a large time $n$ and it is known that the chain has stayed in $A$ and will stay in $A$ for all times up to $\infty$ where $\infty \gg n$. Then the conditional distribution on $X_n$ given this information approaches $\pi$. We can think of the first term of the product $v(x)w(x)$ as the conditional probability of being at $x$ given that the walk has stayed in $A$ up to time $n$ and the second part of the product is the conditional probability given this that the walk stays in $A$ for times between $n$ and $\infty$.

### 8.4.1 Symmetric case

When the Markov chain is symmetric ($p(y, z) = p(x, y)$), then $w(x) = c_1 v(x) = c_2 v(x)^2$. The function $g(x) = \sqrt{v(x)}$ can be characterized by the fact that $g$ is strictly positive and satisfies

\[ P A g(x) = a g(x), \quad \sum_{x \in A} g(x)^2 = 1, \]

If $f_1, f_2 : A \to \mathbb{R}_+$, we define the quadratic form $Q = Q_\pi$, by

\[ Q(f_1, f_2) = \frac{1}{2} \sum_{x,y \in A} p(x, y) [f_1(y) - f_1(x)][f_2(y) - f_2(x)], \]

where $f_1, f_2$ are extended to $D$ by setting $f_j \equiv 0$ on $D \setminus A$. The terms in the sum are zero unless one of $x,y$ is in $A$. We write just $Q(f)$ for $Q(f,f)$. Note that

\[
Q(f) = \frac{1}{2} \sum_{x,y \in A} p(x,y) [f(x) - f(y)][f(x) - f(y)] \\
= \sum_{x \in A, y \in A} p(x,y) f(x)[f(x) - f(y)] \\
= \sum_{x \in A} f(x) \sum_{y \in A} p(x,y) [f(x) - f(y)] \\
= \sum_{x \in A} f(x) (I - P) f(x), \quad (8.14)
\]

### Proposition 8.4.4

Under the assumptions above,

\[
1 - \alpha = Q(g) = \inf_{f \in F} Q(f),
\]

where the infimum is over all $f : A \to \mathbb{R}$ with $\sum f(x) = 1$.

**Proof.** To get the first equality, we use (8.14).

\[ Q(g) = \sum_{x \in A} g(x)^2 (1 - P) g(x) = (1 - \alpha) \sum_{x \in A} g(x)^2 = 1 - \alpha. \]

Write $A = \{ x_1, \ldots, x_m \}, D \setminus A = \{ x_{m+1}, \ldots \}$. Then the goal is to minimize the function

\[ F(t_1, \ldots, t_m) = \sum_{x,j,x \in A} p(x, x_j)(t_k - t_j)^2, \]

subject to $\sum t_j^2 = 1$ and $t_j = 0$ for $j > m$. It is easy to see that the minimum is obtained at a local minimum. Using Lagrange multipliers, we set

\[ L(t_1, \ldots, t_m, \lambda) = F(t_1, \ldots, t_m) - \lambda[\sum_{j=1}^m t_j^2 - 1], \]

At the local minimum we must have

\[ \partial_t L(t_1, \ldots, t_m, \lambda) = 0, \quad \partial_t F(t_1, \ldots, t_m, \lambda) = 0, \]

which translates to

\[ 2 \sum_{x \in A} p(x, x_j)(t_k - t_j) = \lambda t_j, \]

From this we see that any $f$ that minimizes must be an eigenvector, $P f = \lambda f$. However, it is also easy to see that any minimizing $f$ must have a constant sign, for otherwise $Q(|f|) < Q(f)$. Since every eigenvector of constant sign is a multiple of $g$ (see the proof of Theorem 8.4.1), this must be the minimizer. \( \square \)

### 8.4.2 Maximal coupling of Markov chains

Here we will describe the maximal coupling of a Markov chain. Suppose that $p : D \times D \to [0,1]$ is the transition probability function for an irreducible, aperiodic Markov chain with countable state space $D$. Assume that $g_0',g_0''$ are two initial probability distributions on $D$. Let $g^n_0$ denote the corresponding density at time $n$ given recursively by

\[
g^n_0(x) = \sum_{x \in D} g^n_0(x,y) p(z,x), \quad (8.14)
\]

\[ g^n_0(x) = \sum_{x \in D} g^n_0(x,y) p(z,x), \quad (8.14)
\]
Let $| \cdot |$ denote the total variation distance,

$$
|g_0^1 - g_0^2| = \frac{1}{2} \sum_{x \in D} |g_0^1(x) - g_0^2(x)| = 1 - \sum_{x \in D} g_0^1(x) \wedge g_0^2(x).
$$

Suppose

$$X_0^1, X_1^1, \ldots, X_0^2, X_1^2, \ldots,$$

are defined on the same probability space such that for each $j$, $\{X_n^j : n = 0, 1, \ldots\}$ has the distribution of the Markov chain with initial distribution $g^j$. Then it is clear that

$$
P(X_0^1 = X_0^2) \leq 1 - |g_0^1 - g_0^2| = \sum_{x \in D} g_0^1(x) \wedge g_0^2(x). \tag{8.15}
$$

The following theorem shows that there is a way to define the chains on the same probability space so that equality is obtained in (8.15). This theorem gives one example of the powerful probabilistic technique called coupling. Coupling refers to the defining of two or more processes on the same probability space in a way so that each individual process has a certain distribution but the joint distribution has some particularly nice properties. Often, as in this case, the two processes are equal except for an event of small probability. This example is also typical in that the definition of the coupled processes is not very natural.

**Theorem 8.4.5** Suppose $p_0^1, g_0^2$ are as defined in the previous paragraph. We can define $(X_n^1, X_n^2)$, $n = 0, 1, 2, \ldots$, on the same probability space such that:

- for each $j$, $X_0^1, X_1^1, \ldots$, has the distribution of the Markov chain with initial distribution $g^j$;
- for each integer $n \geq 0$,

$$
P\{X_n^1 = X_n^2 \text{ for all } m \geq n\} = 1 - |g_0^1 - g_0^2|.
$$

**Heuristic note.** Before doing this proof, we wish to consider an easier problem of defining $(X^1, X^2)$ on the same probability space so that

$$
P(X^1 = X^2) = 1 - |g_0^1 - g_0^2|.
$$

Let $f^j(x) = g_0^j(x) - [g_0^1(x) \wedge g_0^2(x)]; \rho^j(x) = |g_0^1(x) \wedge g_0^2(x)|$ if $g_0^j(x) > 0$, and $\rho^j(x) = 0$ otherwise. Start with independent $(Z^1, Z^2)$ from these distributions. For each $x$, let $Y^j(x)$ be a 0–1 random variable with $P(Y^j(x) = 1) = \rho^j(x)$. Let $J^1 = Y^1(Z^1)$. Then for each $x$,

$$
P(J^1 = 1, Z^1 = x) = P(J^1 = 1, Z^2 = x) = g_0^1(x) \wedge g_0^2(x),
$$

We can then do the following construction.

- Choose a realization of the 0–1 random variable $J$ with $P(J = 0) = |g_0^1 - g_0^2|$. Let $(X^1, X^2)$ be independent random variables independent of $J$, $X$ has the distribution

$$
P(X = x) = P(Z^1 = x \mid J^1 = 1) = \frac{g_1^1(x) + g_1^2(x)}{1 - |g_0^1 - g_0^2|},
$$

and $W$ has the distribution

$$
P(W = x) = P(Z^2 = x \mid J^2 = 0) = \frac{f^2(x)}{|g_0^1 - g_0^2|}.
$$

It is easy to check that this works.

**Proof.** For ease, we will assume that $|g_0^1 - g_0^2| \to 0$ as $n \to \infty$; the adjustment needed if this does not hold is left to the reader. Let us start with $(Z^1, Z^2)$ which are independent Markov chains with the appropriate distribution. Let $f^j(x) = g_0^j(x) - [g_0^1(x) \wedge g_0^2(x)]$ and define $h_0^j$ by $h_0^j(x) = g_0^j(x)$ and for $n > 1$,

$$
h_n^j(x) = \sum_{m \leq n} f_{m+1}(x) p(x, z).
$$

Note that $h_n^{j+1}(x) = h_{n+1}^j(x) - [h_n^j(x) \wedge h_n^j(x)]$, Let

$$
\rho_n^j(x) = \frac{h_n^j(x) \wedge h_n^j(x)}{h_n^j(x)} \text{ if } h_n^j(x) \neq 0.
$$

We set $\rho_n^j(x) = 0$ if $h_n^j(x) = 0$. We let $\{Y^j(n, x) : j = 1, 2; n = 0, 1, 2, \ldots; x \in D\}$ be independent 0–1 random variables, independent of $(Z^1, Z^2)$, with $P(Y^j(n, x) = 1) = \rho_n^j(x)$.

We now define independent 0–1 random variables $J_n^j$ as follows:

- $J_0^j = Y^j(0, Z_0^j)$;
- if $J_n^j = 1$, then $J_{n+1}^j = 1$ for all $m \geq n$;
- if $J_n^j = 0$, then $J_{n+1}^j = Y^j(n+1, Z_{n+1}^j)$.

We claim that

$$
P(J_n^j = 0; Z_n^j = x) = f_n^j(x).
$$

For $n = 0$, this follows immediately from the definition. Also,

$$
P(J_{n+1}^j = 0; Z_{n+1}^j = x) = \sum_{x \in D} P(J_n^j = 0; Z_n^j = z) P(Z_{n+1}^j = x, Y^{j(n+1, x)} = 0 \mid J_n^j = 0; Z_n^j = z).
$$
8.4. MARKOV CHAINS

The random variable \( Y_j(n + 1, x) \) is independent of the Markov chain and the event \( \{ J_m = 0; Z_n = z \} \) depends only on the chain up to time \( n \) and the values of \( \{ Y_j(k, x) \mid k \leq n \}. \) Therefore,
\[
P \{ Z_{n+1} = x, Y_j(n + 1, x) = 0 \mid J_n = 0; Z_n = z \} = p(z, x) [1 - \rho^1_{n+1}(x)],
\]

Therefore, we have the inductive argument
\[
P \{ J_{n+1} = 0; Z_{n+1} = x \} = \sum_{z \in \mathcal{X}} f_n(z) p(z, x) [1 - \rho^1_{n+1}(x)]
\]
\[
= h^1_{n+1}(x) [1 - \rho^1_{n+1}(x)]
\]
\[
= h^1_{n+1}(x) [h^1_{n+1}(x) \wedge h^1_{n+1}(x)] = f^1_{n+1}(x),
\]

which establishes the claim.

Let \( K^1 \) denote the smallest \( n \) such that \( J_n = 1. \) The condition \( \| g^1_n - g^1_0 \| \to 0 \) implies that \( K^1 < \infty \) with probability one. A key fact is that for each \( n \) and each \( x, \)
\[
P \{ K^1 = n; Z_n = x \} = P \{ K^1 = n; Z_n = x \} = h^1_{n+1}(x) \wedge h^1_{n+1}(x).
\]
This is immediate for \( n = 0 \) and for \( n > 0, \)
\[
P \{ K^1 = n + 1; J_{n+1} = x \}
= \sum_{z \in \mathcal{X}} P \{ J_n = 0; Z_n = z \} P \{ Y_j(n + 1, x) = 1; Z_{n+1} = x \mid J_n = 0; Z_n = z \}
= \sum_{z \in \mathcal{X}} f_n(z) p(z, x) \rho^1_n(x)
= h^1_n(x) \rho^1_n(x) = h^1_{n+1}(x) \wedge h^1_{n+1}(x).
\]

The last important observation is that the distribution of \( W_m := X^0_{n+m} \), given the event \( \{ K^1 = n; x_n = x \} \) is that of a Markov chain with transition probability \( p \) starting at \( x. \)

The reader may note that for each \( j, \) the process \( (Z^j, J^j) \) is a time-inhomogeneous Markov chain with transition probabilities
\[
P \{ (Z^j_{n+1}, J^j_{n+1}) = (y, 1) \mid (Z^j_n, J^j_n) = (x, 1) \} = p(x, y),
\]
\[
P \{ (Z^j_{n+1}, J^j_{n+1}) = (y, 0) \mid (Z^j_n, J^j_n) = (x, 0) \} = p(x, y) [1 - \rho^1_{n+1}(y)],
\]
\[
P \{ (Z^j_{n+1}, J^j_{n+1}) = (y, 1) \mid (Z^j_n, J^j_n) = (x, 0) \} = p(x, y) \rho^1_n(y).
\]

The chains \( (Z^1_n, J^1_n) \) and \( (Z^2_n, J^2_n) \) are independent. However, the transition probabilities for these chains depend on both initial distributions and \( p. \)

We are now ready to make our contraction of \( (X_n^0, X_n^1). \)

- Define for each \( (n, x) \) a process \( \{ W^m_{n+m} : m = 0, 1, 2, \ldots, \} \) that has the distribution of the Markov chain with initial point \( x. \) Assume that all these processes are independent.

- Choose \( (n, x) \) according to the probability distribution
\[
h^1_{n+1}(x) \wedge h^1_{n+1}(x) = P \{ K^1 = n; Z_n = x \}.
\]

Set \( J^1_m = 1 \) for \( m \geq n \) and \( J^1_m = 0 \) for \( m < n, \) and \( K^1 = K^1 = n. \) Note that \( K^1 \) is the smallest \( n \) such that \( J^1_m = 1. \)

- Given \( (n, x), \) choose \( X^0_1, \ldots, X^0_n \) from the conditional distribution of the Markov chain with initial distribution \( g^0_n \) conditioned on the event \( \{ K^1 = n; Z_n = x \}. \)

- Given \( (n, x), \) choose \( X^1_1, \ldots, X^1_n \) (conditionally) independent of \( X^0_1, \ldots, X^0_n \) from the conditional distribution of the Markov chain with initial distribution \( g^0_n \) conditioned on the event \( \{ K^1 = n; Z_n = x \}. \)

- Let
\[
X^j_m = W^m_{n+j}, \quad m = n, n + 1, \ldots
\]

The two conditional distributions above are not easy to express explicitly; fortunately, we do not need to do so.

To finish the proof, we need only check that the above construction satisfies the conditions. For fixed \( j, \) the fact that \( X^0_1, X^0_j, \ldots \) has the distribution of the chain with initial distribution \( g^0_n \) is immediate from construction and the earlier observation that the distribution of \( \{ X^0_1, X^0_1, \ldots \} \) given \( \{ K^1 = n; Z_n = x \} \) is that of the Markov chain starting at \( x. \)

Also, the construction immediately gives \( X^0_n = X^1_n \) if \( m \geq K^1 = K^1. \) Also,
\[
P \{ J^1_m = 0 \} = \sum_{x \in \mathcal{X}} f^1_n(x) = |g^1_n - g^1_0|,
\]

\[\square\]

Remark. A quick review of the proof of Theorem 8.4.5 shows that we do not need to assume that the Markov chain is time-homogeneous. However, time-homogeneity makes the notation a little simpler and we only use the result for time-homogeneous chains.

8.5 Some Tauberian theory

Lemma 8.5.1 Suppose \( \alpha > 0, \) Then as \( \xi \to 1^{-}, \)
\[
\sum_{i=1}^{\infty} \xi^i \alpha^i n^{i-\alpha} = \frac{\Gamma(\alpha)}{|1 - \xi|^\alpha} [1 + O(1 - \xi)].
\]

Proof. Let \( \epsilon = 1 - \xi. \) First note that
\[
\sum_{n > n^*} \xi^i \alpha^i n^{i-\alpha} \sum_{n > n^*} [(1 - \epsilon)^{1/\epsilon} n^{i-\alpha}] \leq \sum_{n > n^*} \xi^i n^{i-\alpha},
\]

\[\square\]
and the right-hand side decays faster than every power of $\epsilon$. For $n \leq \epsilon^{-1}$ we can do the asymptotics
\[
\exp\{n \log(1 - \epsilon)\} = \exp\{n(-\epsilon = O(\epsilon^2))\} = e^{\epsilon^2}\Gamma(\epsilon) + O(\epsilon^2)\).
\]
Hence,
\[
\sum_{n=1}^{\infty} \xi^n n^{a-1} = e^\epsilon \sum_{n=1}^{\infty} \xi^n (\epsilon n)^a [1 + O(\epsilon)(\epsilon n)].
\]
Using Riemann sum approximations we see that
\[
\sum_{n=1}^{\infty} \xi^n (\epsilon n)^a = \int_0^\infty e^{\epsilon t} t^{a-1} dt + O(\epsilon) = \Gamma(\alpha) + O(\epsilon),
\]
\[
\sum_{n=1}^{\infty} \xi^n (\epsilon n)^a n^{a-1} = O(1).
\]

\textbf{Proposition 8.5.2} Suppose $u_n$ is a sequence of nonnegative real numbers. If $\alpha > 0$, the following two statements are equivalent:
\[
\sum_{n=1}^{\infty} \xi^n u_n = \frac{\Gamma(\alpha)}{|1 - \xi|^\alpha}, \quad \xi \to 1^-,
\]
\[
\sum_{n=1}^{N} u_n \sim N^{\alpha}, \quad N \to \infty.
\]
Moreover, if the sequence is monotone, either of these statements implies
\[
u_n \sim N^{\alpha - 1}, \quad n \to \infty.
\]

\textbf{Proof.} Let $U_n = \sum_{j=1}^{n} u_j$. Note that
\[
\sum_{n=0}^{\infty} \xi^n u_n = \sum_{n=0}^{\infty} \xi^n |U_n - U_{n-1}| = (1 - \xi) \sum_{n=0}^{\infty} \xi^n U_n + \frac{\Gamma(\alpha + 1)}{1 - \xi} \sum_{n=0}^{\infty} \xi^n u_n.
\]
If (8.17) holds, then
\[
\sum_{n=0}^{\infty} \xi^n u_n \sim (1 - \xi) \sum_{n=0}^{\infty} \xi^n a^{\alpha - 1} n^a = \frac{\Gamma(\alpha + 1)}{1 - \xi} \sum_{n=0}^{\infty} \xi^n u_n.
\]
Now suppose (8.16) holds. We first give an upper bound on $U_n$. Using $1 - \xi = 1/n$, we can see as $n \to \infty$,
\[
U_n \leq n^{\alpha - 1} \left(1 - \frac{1}{n}\right)^a \sum_{j=1}^{n} \left(1 - \frac{1}{n}\right)^j U_j \leq n^{\alpha - 1} \left(1 - \frac{1}{n}\right)^a \sum_{j=1}^{\infty} \left(1 - \frac{1}{n}\right)^j U_j \sim \Gamma(\alpha) n^a.
\]

\textbf{8.6 Second moment method}

\textbf{Lemma 8.6.1} Suppose $X$ is a nonnegative random variable with $\mathbb{E}[X^2] < \infty$ and $0 < r < 1$. Then
\[
\mathbb{P}(X \geq r \mathbb{E}[X]) \geq \frac{(1 - r)^2 \mathbb{E}[X^r]}{\mathbb{E}[X]^2}.
\]

\textbf{Proof.} Without loss of generality, we may assume that $\mathbb{E}[X] = 1$. Since $\mathbb{E}[X; X \leq r] \leq r$, we know that $\mathbb{E}[X; X \geq r] \geq (1 - r)$. Then,
\[
\mathbb{E}[X^r] \geq \mathbb{E}[X^2; X \geq r] = \mathbb{P}(X \geq r) \mathbb{E}[X^2 \mid X \geq r] \geq \mathbb{P}(X \geq r) (\mathbb{E}[X^r \mid X \geq r])^2 = \frac{\mathbb{E}[X; X \geq r]}{\mathbb{P}(X \geq r)} \geq (1 - r)^2 \mathbb{E}[X^r].
\]

\textbf{8.7 Other methods}

\textbf{8.7.1} Convolution of probability distributions

Let $\nu^{(1)}$ denote the measure on $[0, \infty]$ that gives measure $\gamma^{(1)}$ to the point $n/\epsilon$. Then the last estimate shows that the total mass of $\nu^{(1)}$ is uniformly bounded on each compact interval and hence there is a subsequence that converges weakly to a measure $\nu$ that is finite on each compact interval. Using (8.10) we can see that for each $\lambda > 0$,
\[
\int_0^\infty e^{\lambda t} \nu(dt) = \int_0^\infty e^{\lambda t} \nu^{(1)}(dt).
\]
This implies that $\nu$ is $\alpha$-stable. Since the limit is independent of the subsequence, we can conclude that $\nu^{(1)} \to \nu$ and this implies (8.17).

The fact that (8.17) implies the last assertion if $u_n$ is monotone is straightforward using
\[
U_{n+1} - U_n = \alpha^{\alpha - 1} (e(n + 1))^\alpha - (e n)^\alpha \sim \alpha^{\alpha - 1} n \alpha, \quad n \to \infty.
\]
Corollary 8.6.2 Suppose $E_1, E_2, \ldots$ is a collection of events with $\sum \mathbb{P}(E_n) = \infty$. Suppose also that there is a $K < \infty$ such that for all $j \neq k$, $\mathbb{P}(E_j \cap E_k) \leq K \mathbb{P}(E_j) \mathbb{P}(E_k)$. Then
\[ \mathbb{P}(E_k \text{ i.o.}) \geq \frac{1}{K}. \]

Proof. Let $V_n = \sum_{i=1}^n 1_{E_i}$. Then the assumptions imply that
\[ \liminf_{n \to \infty} \mathbb{E}(V_n) = \infty, \]
and
\[ \mathbb{E}(V_n^2) = \sum_{j=1}^n \mathbb{P}(E_j) + \sum_{j \neq k} K \mathbb{P}(E_j) \mathbb{P}(E_k) \leq \mathbb{E}(V_n) + K \mathbb{E}(V_n)^2 = \left[ \frac{1}{\mathbb{E}(V_n)} + K \right] \mathbb{E}(V_n)^2. \]

By Lemma 8.6.1, for every $r > 0$,
\[ \mathbb{P}\{V_n \geq r \mathbb{E}(V_n)\} \geq \frac{(1-r)^2}{K + \mathbb{E}(V_n)^2}. \]

Since $\mathbb{E}(V_n) \to \infty$, this implies
\[ \mathbb{P}\{V_n = \infty\} \geq \frac{(1-r)^2}{K}, \]
and since this holds for every $r > 0$, we get the corollary. \qed

8.7 Subadditivity

Lemma 8.7.1 (Subadditivity lemma) Suppose $f : \{1, 2, \ldots\} \to \mathbb{R}$ is subadditive, i.e., for all $n, m$, $f(n + m) \leq f(n) + f(m)$. Then,
\[ \lim_{n \to \infty} \frac{f(n)}{n} = \inf_{n \to \infty} \frac{f(n)}{n}. \]

Proof. Fix integer $N > 0$. We can write any integer $n$ as $jN + k$ where $j$ is a nonnegative integer and $k \in \{1, \ldots, N\}$. Let $b_n = \max\{f(1), \ldots, f(N)\}$. Then subadditivity implies
\[ \frac{f(n)}{n} \leq \frac{jf(N) + f(k)}{jN} \leq \frac{f(N)}{N} + \frac{b_N}{jN}. \]

Therefore,
\[ \limsup_{n \to \infty} \frac{f(n)}{n} \leq \frac{f(N)}{N}. \]

Since this is true for every $N$, we get the lemma. \qed

Corollary 8.7.2 Suppose $r_n$ is a sequence of positive numbers and $b_1, b_2 > 0$ such that for every $n, m$,
\[ \frac{b_1 b_n r_m}{r_n} \leq r_{n+m} \leq b_2 r_n r_m. \quad (8.20) \]

Then there exists $\alpha > 0$ such that for all $n$,
\[ b_1^{-1} a^n \leq r_n \leq b_2^{-1} a^n. \]

Proof. Let $f(n) = \log r_n + \log b_2$. Then $f$ is subadditive and hence
\[ \lim_{n \to \infty} \frac{f(n)}{n} = \frac{f(n)}{n} := \alpha. \]

This shows that $f(n) \geq \alpha n / b_2$. Similarly, by considering the subadditive function $g(n) = -\log r_n - \log b_1$, we get $f(n) \leq b_1^{-1} a^n$. \hfill \square

Remark. Note that if $r_n$ satisfies (8.20), then so does $\beta r_n$ for each $\beta > 0$. Therefore, we cannot determine the value of $\alpha$ from (8.20).

Exercises for the appendix

Exercise 8.1 Find $f_\delta(\xi), f_\delta(\xi)$ in (8.3).